Research on Substation Engineering Estimates Based on BIM-DE-RF

Songsong Wang [1], Wenxuan Qiao [1], Lei Wang [1], Zhewei Shen [1], Pengju Yang [2], and Li Bian [2,*]

[1] Hebei Electric Power Engineering Supervision Co., Ltd, Hebei, China
[2] Shanghai Electric Power Design Institute Co., Ltd, Shanghai, China
*Correspondence: 18117590875@189.cn (L.B.)

***Abstract*—Aiming at the problems of heavy workload and large errors in traditional substation engineering estimation methods, an intelligent estimation method for substation engineering based on Building Information Modeling (BIM) combined with a Differential Evolution (DE) algorithm to optimize Random Forest (RF) is proposed. This proposed method uses DE to optimize the RF model's splitting features and decision trees to enhance the model's estimation accuracy. The BIM of the substation project is used to determine engineering quantity information, which serves as the input of the DE-RF model, enabling intelligent cost estimation of the substation project. The results of the example analysis show that the relative error of the proposed cost estimation method for substation engineering based on BIM and DE-RF is below 10%. This accuracy level meets various substation engineering cost estimation scenarios, validating the feasibility and correctness of the proposed model.**

***Keywords*—building information modeling, random forest, optimization, differential evolution**

I. Introduction

Power substation engineering is one of the main components of power grid construction [1, 2]. Conducting cost estimation of power substation engineering during the planning and implementation phases can provide support for optimized decision-making in its construction [3]. In the traditional feasibility study stage of power substation engineering, estimating the project cost requires calculating the project quantity based on the design scheme and construction drawings to obtain the project quantity calculation book. Then, the unit price table is determined based on budgetary quotas, followed by the calculation of equipment and labor costs. Due to the numerous and complex cost categories involved, this traditional pricing method consumes a considerable amount of manpower and resources and requires a certain amount of time. Furthermore, it cannot identify potential problems that may arise during the construction process. Consequently, this approach requires experienced personnel with engineering knowledge [4, 5].

The concept and terminology of Building Information Modeling (BIM) have gained attention in various industries globally. By utilizing BIM technology, similar engineering models can communicate with each other, significantly reducing the design process. Literature [6–8] has achieved fruitful results by incorporating Life Cycle Cost (LCC) into engineering cost management and construction cost, but there is still a gap between research and practical application.

Random Forest (RF) is a powerful learning algorithm that has been widely researched and applied [7–13]. RF was used for data processing and prediction, and simulation results showed that the RF model had a high prediction accuracy [7–10]. Wang and Zhang *et al.* [11], Wang and Yin *et al.* [12], Hu and Si *et al.* [13] applied RF to fault diagnosis and achieved good results. This highlights the strong performance of the RF algorithm in data learning and computing. However, these studies designed the RF model based on experience, and the input data were collected from engineering data. When applied to the preliminary estimate of substation engineering, a significant amount of data collection is required, increasing labor costs.

Based on this, this article combines artificial intelligence theories such as differential evolution algorithm and random forest with BIM technology to conduct preliminary cost estimation research on substation engineering based on nearly five years of substation engineering data in a certain region. The study establishes the BIM-DE-RF substation engineering preliminary cost estimation model, thereby significantly reducing labor costs and improving the intelligent level of substation engineering preliminary cost estimation. BIM technology is used to analyze engineering quantity information and combine it with engineering characteristics to obtain substation engineering information. The obtained substation engineering information is imported into the DE-RF estimation model to quickly obtain the investment scale of the engineering cost.

Section II describes the RF model. Section III describes the DE-RF model. Section IV introduces the application process of BIM in the estimate of substation projects. The case studies will be shown in Section V. The conclusions are given in the last section.

Manuscript received February 6, 2023; revised March 16, 2023; accepted April 11, 2023; published September 6, 2023.

doi: 10.12720/jait.14.5.892-896

II. RF MODEL

A. RF Principle

RF is a comprehensive algorithm proposed to solve the problem of regression prediction and feature classification in data mining. The algorithm is obtained by combining a classification regression tree and bagging [9].

1) Classification and regression tree

Classification and regression tree (CART) is a binary recursive segmentation technique, which divides the current data sample set into two at any other node except leaf nodes [12]. The calculation of CART is shown in Eq. (1).

$$G_D = 1 - \sum_{j=1}^{m} p_i^{\,2} \tag{1}$$

If D is binary divided by attribute A, the corresponding Gini index is shown in Eq. (2).

$$G_{D,A} = \frac{|D_1|}{D} G_{D_1}(D_1) + \frac{|D_2|}{D} G_{D2}(D_2) \tag{2}$$

2) Bagging algorithm

The Bagging algorithm is mainly to improve the prediction accuracy of CART, so it is introduced to combine with CART. The main job of the bagging algorithm is to randomly sample the collected data to obtain a decision tree.

B. RF Model Calculation Process

Fig. 1 illustrates the process of using the RF method for substation engineering cost estimation. By collecting a large amount of substation engineering data and normalizing it, the data is divided into training and testing sets. The training set is used to train the RF model, and then the independent variables in the testing set are inputted into the trained model to obtain the substation engineering cost. The RF model's fitting performance is then compared with the collected substation engineering cost to verify its accuracy.

This process demonstrates an effective approach for substation engineering cost estimation using the RF method, which can improve the accuracy of cost estimation by leveraging large datasets and advanced modeling techniques.

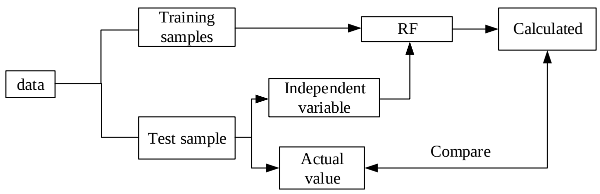

Figure 1. Flow chart of RF-based cost calculation of substation engineering cost.

III. DE-RF MODEL

To improve the performance of the RF model, DE is used to optimize the RF parameters. Based on the DE algorithm, the number of decision trees and splitting features in the random forest is optimized to obtain the optimal solution. The optimal decision tree and splitting features in the random forest are then used to predict and calculate the cost of substation engineering, improving the model's computational accuracy and obtaining more accurate values for substation engineering cost. This approach demonstrates an effective method for optimizing the RF model's parameters, leveraging advanced optimization techniques to improve the accuracy of substation engineering cost estimation.

A. DE

DE is a stochastic, population-based selection tool that deals with continuous spatial domains [14]. It is a parallel direct search method using NP (population size) parameter vectors. A population consists of a real-valued vector of dimension D. The DE optimization process is shown in Fig. 2. If no data is known, the initial population is randomly selected. The optimization process is carried out through three main operations: mutation, cross, and selection [14].

1) Initialization

For each parameter j with a lower bound $\boldsymbol{X}_j^{L}$ and an upper bound $\boldsymbol{X}_j^{U}$, the initial parameter values are usually randomly chosen in the interval [$\boldsymbol{X}_j^{L}$, $\boldsymbol{X}_j^{U}$].

2) Variation

For a given parameter vector $\boldsymbol{V}_{i,G}$, three vectors ($\boldsymbol{V}_{r1,G}$, $\boldsymbol{V}_{r2,G}$, $\boldsymbol{V}_{r3,G}$)are randomly chosen such that their indices *i*, r1, r2, and r3 are different. A donor vector is created by adding the weighted difference between the two vectors to a third vector $\boldsymbol{V}_{i,G+1}$, is shown in Eq. (3).

$$V_{i,G+1} = X_{r1,G} + F(X_{r2,G} - X_{r3,G}) \tag{3}$$

3) Cross

Three parents are selected for crossbreeding, and the child is a perturbation of one of the parents. The trial vector $\boldsymbol{V}_{i,G+1}$ consists of elements of the target vector ($\boldsymbol{Y}$) and elements of the donor vector ($\boldsymbol{X}_{i,G}$).

$$\boldsymbol{U}_{j,i,G+1} = \begin{cases} \boldsymbol{V}_{j,i,G+1} & if\ rand_{j,i} \le CR\ or\ j = I_{rand} \\ \boldsymbol{X}_{j,i,G+1} & if\ rand_{j,i} > CR\ or\ j \ne I_{rand} \end{cases} \tag{4}$$

With $rand_{j,i}$ ~$\boldsymbol{U}(0,1)$, I_{rand} is a random integer from (1,2,...,D), where D is the dimension of the solution i.e. the number of control variables. I_{rand} makes sure $\boldsymbol{V}_{i,G+1} \neq \boldsymbol{V}_{i,G}$.

4) Select

The target vector $\boldsymbol{V}_{i,G}$ is compared with the test vector $\boldsymbol{V}_{i,G+1}$, and the vector with a better fitness value is allowed to enter the next generation. The selection operation in DE can be represented by the following Eq. (5).

$$\boldsymbol{X}_{i,G+1} = \begin{cases} \boldsymbol{U}_{i,G+1} & if\ f(\boldsymbol{U}_{i,G+1}) < f(\boldsymbol{U}_{i,G}) \\ \boldsymbol{X}_{i,G} & \text{otherwise} \end{cases} \tag{5}$$

where, $i \in [1, N_p]$.

B. *Fitness Function of DE-RF*

Fitness represents the strengths and weaknesses of the population in DE. In this paper, the difference between the predicted value and the actual value of each substation project cost is calculated, the obtained difference is squared and then summed, and the reciprocal of the obtained sum is taken as the selection direction of DE optimization, as shown in Eq. (6).

$$E_i = \frac{1}{\sum_{k=0}^{m} (y(k) - s(k))^2} \tag{6}$$

The calculation process of DE-RF substation engineering cost is shown in Fig. 3. First, the collected data is used as the input of the RF model, and the parameters of the RF model are optimized by DE to obtain the optimal RF model, and then the substation engineering cost is calculated.

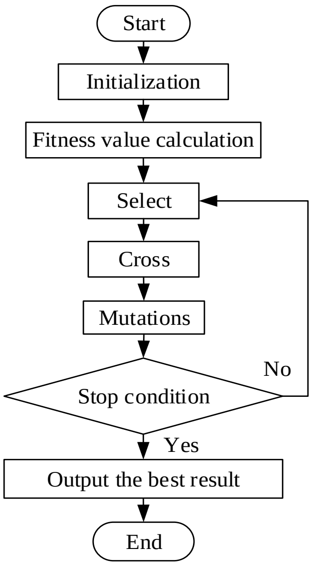

Figure 2. DE optimization flow chart.

IV. Application of BIM in Substation Project Budget

This article introduces BIM technology into substation engineering cost estimation, utilizing the advantage of BIM models to obtain engineering information. BIM is based on digital 3D models and open standards, using professional software to achieve information sharing and full life-cycle management of project construction, improving project implementation efficiency and management level. Combined with the characteristics of the engineering itself, indicators such as voltage level, the number of main transformers, and single unit capacity are determined. These pieces of information are organized as validation samples for the prediction algorithm, which can be used to predict the cost of the engineering project. The route of the BIM-based substation engineering cost estimation method is shown in Fig. 4.

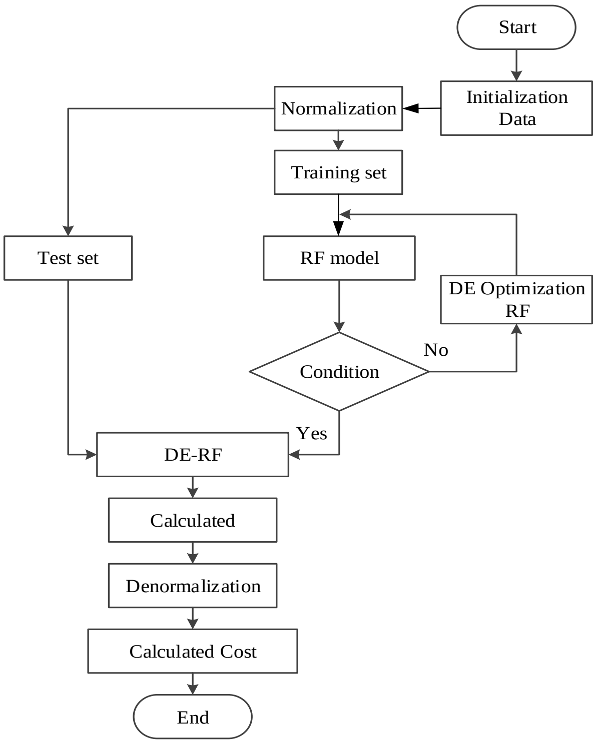

Figure 3. DE-RF model calculation flow chart.

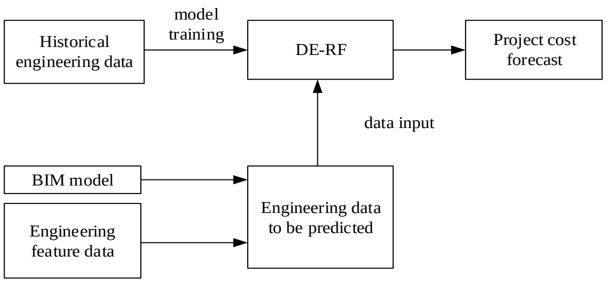

Figure 4. Substation project estimation algorithm based on BIM and DE-RF.

V. Case Analysis

This paper uses BIM technology and DE-RF to predict and analyze the investment scale of substation engineering. Taking a 220kV new substation project in a certain area as an example, the validity and correctness of the proposed method are verified.

A. *Model Input Data*

(1) BIM model data. Using BIM technology to build a substation engineering model can create a digital model that is the same as the actual project. The three-dimensional structure of the project can be viewed intuitively and clearly through the digital model; the corresponding engineering quantity and cost estimate can be calculated conveniently and quickly through the parameters of each module.

(2) Engineering feature data. Engineering feature data refers to the characteristics of the substation project itself in terms of electrical and building area, including voltage level, the number of main transformers, and the capacity of a single unit.

B. Results and Analysis

1) Estimated analysis of investment cost

Using historical data of 132 substation projects in a certain region as the sample, 100 substation projects were randomly selected as the training set, and 32 substation projects were used as the test set, with parameters of the 32nd substation project generated by BIM. The DE-RF model was trained and optimized based on the training set, and the accuracy of the optimized model was tested using the test set. To validate the effectiveness of the proposed method, the same dataset was used as input data for both the DE-RF algorithm and the traditional RF algorithm. The estimated investment costs and error analysis results of the substation projects are shown in Fig. 5.

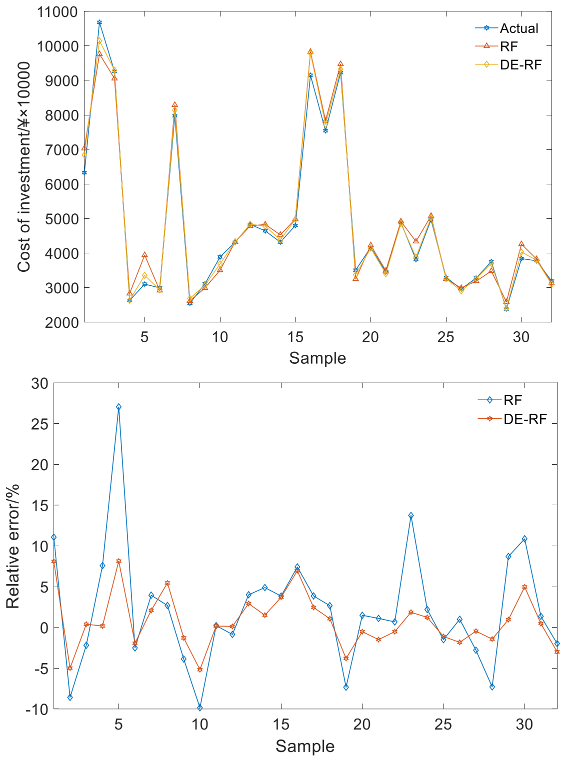

Figure 5. Prediction results of substation project investment cost.

As shown in Fig. 5, the average relative error of substation project cost estimation based on traditional RF is 2.24%, while the average relative deviation of DE-RF is 0.79%, indicating that DE-RF has better calculation accuracy. This is because DE-RF uses DE to optimize the RF characteristic parameters and the number of decision trees, making the algorithm more suitable for the application scenarios of substation project cost estimation and with higher precision. As for the predicted relative error of the 32nd sample, it is less than 10%, which validates the feasibility of using BIM as input parameters for DE-RF.

2) Estimated analysis of itemized costs

The investment in transmission and transformation (T&T) engineering projects is enormous, with numerous influencing factors and complex coordination. The cost estimate of engineering in the feasibility study stage plays a crucial role in optimizing the design and technical solutions and controlling the cost of the entire T&T engineering. For T&T engineering, investment mainly consists of four costs: construction costs, installation costs, equipment procurement costs, and other costs. These four costs are also referred to as the sub-item costs of T&T engineering. That is, the investment in T&T engineering mainly depends on the size of the sub-item costs. To further analyze the accuracy of the model and focus on the sub-item costs, the dependent variables were changed to the four sub-item costs: construction costs, installation costs, equipment procurement costs, and other costs, and analyzed separately. Compared with analyzing a static investment, analyzing sub-item costs can more deeply analyze the cost of engineering. One hundred projects were randomly selected as training samples, and six projects were selected as validation samples, with the parameters of the sixth sample generated by BIM. The sub-item costs of T&T engineering were estimated and the error analysis results are shown in Fig. 6.

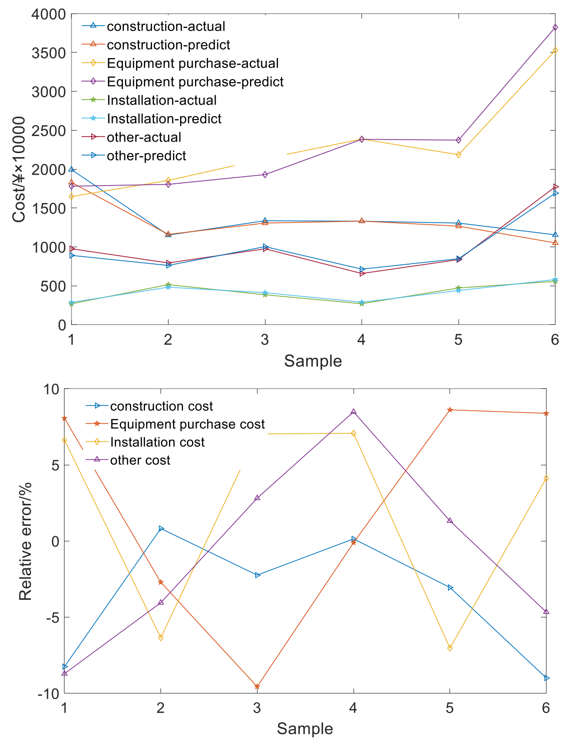

Figure 6. Prediction results of sub-item costs for substation projects.

As shown in Fig. 6, the predicted results of the four sub-item costs, namely, construction costs, equipment procurement costs, installation costs, and other costs, are all below 10% of the actual results of the engineering, verifying the feasibility of the model. The feasibility of using BIM model parameters as inputs for DE-RF in sub-item cost estimates was also validated. This suggests that the model established in the feasibility study stage of T&T engineering can quickly understand the range of sub-item costs, providing a useful reference for the actual engineering. Based on the above analysis, the BIM and DE-RF combined method proposed in this paper for T&T engineering cost estimation can save a lot of manual labor costs, reduce the computational workload, and greatly improve the efficiency of the feasibility study stage of T&T engineering, compared with traditional T&T engineering cost estimation methods.

VI. Conclusion

This paper proposes a BIM and DE-RF-based cost estimation method for substation engineering to improve the accuracy and efficiency of substation engineering cost estimation. The analysis of the case study shows that the DE-RF algorithm has a smaller average relative error in cost estimation than the traditional RF algorithm, indicating higher accuracy. The proposed method takes the BIM model as input, and the resulting substation engineering cost estimation has an error rate below 10%, which has significant practical implications. The next step could be to analyze the factors that affect the accuracy of substation engineering cost estimation by comparing BIM engineering information and actual engineering information, improving the model, and developing engineering application programs.

Parameter

G_D Gini index
p_i Frequency of occurrence of elements of type i
m The categories included in the dataset
F Variation factor the value range is generally [0,1]
$\boldsymbol{D}$ Data set
CR Update probability
U Variable lower limit
L Variable upper limit
$y(k)$ No. k substation project forecast value
$s(k)$ No. k substation project actual value
CR Crossover probability factor, the value range is generally [0,1]

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Under the supervision of Zhewei Shen and Pengju Yang, Songsong Wang formulated the research goals and objectives, designed the methodology, and created and evaluated the models. Wenxuan Qiao and Lei Wang were responsible for preparing the published work, specifically writing the initial draft, critical review, and revision. Li Bian was in charge of the study's oversight and leadership responsibility for the research activity planning and execution, as well as the mentorship of the core team. All authors had approved the final version.

Funding

This research was funded by Contract KY-2022-0004 from State Grid Hebei Electric Power Co., Ltd.

References

[1] Y. Jiang, H. Shan, Y. Shi, and R. Jia, "Design of power transmission and transformation engineering evaluation system based on Fuzzy-AHP," *Electric Power Science and Engineering*, vol. 34, no. 6, pp. 30–36, 2018.
[2] X. Wang, "Research on the construction of big data platform for power transmission and transformation project cost and the application of intelligent analysis management and control," *Electric Power Big Data*, vol. 21, no. 11, pp. 8–14, 2018.
[3] Z. Xiong, J. Liu, and X. Ni, "Research on safety protection measures for UHV substation engineering design and construction connection," *China Electric Power*, vol. 50, no. 10, pp. 52–56, 2017.
[4] W. Li, T. Chen, S. Liu, and R. Hu, "Construction of standard measurement system for power transmission and transformation engineering based on data processing integration," *Electronic Design Engineering*, vol. 26, no. 20, pp. 133–136, 2018.
[5] P. Geng, X. Qi, B. Yin, J. Li, and J. Wang, "Research on the calculation mode and method of basic reserve fee rate for power transmission and transformation projects," *Shaanxi Electric Power*, vol. 43, no. 10, pp. 58–61, 2015.
[6] G. Ren and Q. Zhang, "The application of full life cycle assessment in green buildings in my country," *Shenyang Agricultural University (Social Science Edition)*, vol. 9, no. 5, pp. 773–775, 2007.
[7] H. Li, J. Hua, and B. Zou, "Stock price prediction based on RF-LSTM combination mode," *Journal of Henan University of Technology (Natural Science Edition)*, vol. 41, no. 1, 2022.
[8] T. Zhang, M. Wei, and X. Gao, "Prediction model of bubble dissolution time for selective laser sintering parts based on IPSO-WRF," *Computer Science*, vol. 48, no. S2, pp. 638–643, 2021.
[9] S. Gao, Z. He, Z. Liu, J. Liu, G. Wang, and D. Li, "Application of difference fusion analysis based on machine learning in air quality prediction," *Electronic Measurement Technology*, vol. 44, no. 18, pp. 85–92, 2021.
[10] H. Chen, C. Yang, M. Du, and Y. Zhang, "Credit risk prediction model based on boundary adaptive SMOTE and focal loss function lightGBM," *Computer Applications*, vol. 12, no. 7, pp. 2256–2264, 2021.
[11] T. Wang, L. Zhang, and Y. Wu, "Research on transformer fault diagnosis based on GWO-RF algorithm," *Journal of Physics: Conference Series*, vol. 1952, no. 3, pp. 32–38, 2021.
[12] Q. Wang, Z. Yin, and T. E. (2016). Research on V2G charging pile fault diagnosis based on random forest algorithm. *Electrical Measurement and Instrumentation*. [Online]. Available: http://kns.cnki.net/kcms/detail/23.1202.TH.20210726.0904.002.html (in Chinese)
[13] Q. Hu, X.-S. Si, Q.-H. Zhang, A.-S. Qin, "A rotating machinery fault diagnosis method based on multi-scale dimensionless indicators and random forests," *Mechanical Systems and Signal Processing*, vol. 139, pp. 12–20, 2020.
[14] X. Wang, J. Deng, and Z. Yang, "Parameter optimization strategy of power system controller," *Journal of Electric Machinery and Control*, vol. 24, no. 9, pp. 95–104, 2020.