Research on Grid Infrastructure Project Investment Prediction Model Based on Bootstrap Data Expansion and Grey Correlation

Jianjia Zhou [1], Na An [1], Kui Liang [1], Junxian Jiang [2*], Yaxue Mu [2]

E-mail: luotuo868@163.com, 392387440@qq.com, 604783372@qq.com, 331463263@qq.com, 1074202506@qq.com

[1]Leshan Power Supply Company State Sichuan Electric Power Company Leshan 614000;

[2]School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu, 611731

Abstract. Aiming at the problems in the actual construction process of power grid infrastructure projects, such as the lack of time series data, the difficulty in analyzing the correlation between engineering quantity and investment amount, and the deviation of investment budget caused by engineering quantity estimation deviation, this paper studies the power grid infrastructure based on Bootstrap data expansion and gray correlation, Project engineering investment forecasting model. Firstly, the time series distribution of various expenses and engineering costs is analyzed, and the Bootstrap data expansion method is used to construct various cost and expense databases and cost data distribution characteristic models of power grid infrastructure projects. , established a calculation and evaluation model of the correlation between engineering quantity and value of power grid infrastructure projects based on the grey relational analysis method; through the coordinated preparation and adjustment model of the designed plan and budget, a long-term short-term memory neural network-based grid infrastructure project investment prediction model was established. Finally, through the analysis of simulation experiments, it is shown that the model studied can not only improve the accuracy of investment plan and budget management, solve the problem of inconsistency between investment plan and budget and actual project progress, reduce waste and reduce cost; it can also realize the infrastructure construction of power grid enterprises. The project is more standardized and efficient implementation.

Keywords Quantities; Value; Distribution characteristics; Correlation; Synergy to compile

1. Introduction

Power grid construction is an important national economic foundation, with large investment and long cost recovery period. Moreover, power grid operation is significantly affected by the economy. In recent years, with the continuous increase of social electricity demand with economic development, the investment scale of power grid enterprises has also gradually increased. At the same time, affected by the adjustment of the national industrial structure, power grid enterprises are faced with great uncertainty in investment management, including large

BDEIM 2022, December 02-03, Zhengzhou, People's Republic of China

DOI 10.4108/eai.2-12-2022.2328678

fluctuations in corporate benefits, failure of new asset performance to play an effective role, and prominent input-output structure contradictions[1]. Therefore, under the premise of ensuring the safe and stable operation of the power grid, it is necessary to accurately grasp the investment capacity of power grid enterprises, optimize the allocation of investment structure, and improve the economy of power grid construction, so as to realize the rational allocation of resources and sustainable development of enterprises, and steadily improve the investment efficiency of power grids[2]. Problems that power grid enterprises need to solve urgently. The investment plan clarifies the medium and long-term development plan of the enterprise and is the guideline for the sustainable development of the enterprise. Financial budget personnel are familiar with the medium and long-term development goals of the enterprise, which can enhance the rolling nature of budget preparation and avoid the short-sighted behavior of budgeting only for the current period[3]. Aligning the financial budget objectives with the medium and long-term development objectives of the enterprise can promote the sustainable development of the enterprise. Implementing dual management and control of investment plans and financial budgets in power grid infrastructure projects, on the one hand, realizes a comprehensive grasp and clear presentation of the project composition and cost details of all grid infrastructure projects, and enhances the pertinence, practicability and effectiveness of budget management[4]. On the other hand, it promotes the reasonable setting of the project cost pre-control target, strengthens the coordination and integration of the whole process management of the project, and effectively realizes the overall allocation of economic resources of power grid enterprises and the benefit guidance of power grid investment[5].

2. Problem Formulate

There are many difficulties in the analysis of power grid infrastructure. Yunna Wu proposed a cost estimation method for power grid projects based on case reasoning (CBR)[6]. Jinming Li summarized the prediction indicators and selection methods of domestic and foreign scholars for power grid investment demand[7]. Juhua Hong established a CW-MEEM model for post-evaluation of power grid construction projects[8]. Based on AHP, Guoliang Luo established a comprehensive evaluation model for investment in power grid construction projects[9]. Based on AHP and grey relational analysis, Kecheng Li proposed a more accurate evaluation method to evaluate the importance of each index and the correlation between different objects, and realized the quantification of the importance of each evaluation index [10]. In the actual construction process of power grid infrastructure projects, the management methods of projects of different magnitudes are different, and projects of different magnitudes need to be analyzed and processed separately, and the number of project samples is small, which is difficult to directly analyze and process. Therefore, Bootstrap-based data is used. The augmentation method augments the sample size. In addition, due to the large data differences between projects, it is difficult to obtain better results if the project data is directly analyzed. trend, and build a model of project cost time series characteristics. The traditional correlation analysis method is difficult to quantify the direct correlation between the project engineering quantity and the investment amount. This paper adopts the grey correlation analysis method to quantify the correlation between the project engineering amount and the investment amount. Finally, the long short-term memory neural network is used to make rolling predictions on the investment amount, so as to realize its rolling update and adjustment.

3. Project cost time series distribution characteristic model

3.1 Data sampling method based on Bootstrap

Due to the small sample characteristics of the residual value data of power grid infrastructure projects, the parametric modeling method cannot cover all the error distribution characteristics, and the traditional data-driven modeling methods such as the kernel extreme learning machine have large generalization errors, so it is necessary to expand the construction. Model the data sample size, and optimize and correct the data-driven model.

The schematic diagram of the data sampling method based on Bootstrap is shown in Figure 1, and the specific implementation steps are as follows:

(1) Calculate the small sample X density function based on the kernel density method;

(2) Use the Bootstrap method to perform self-sampling operations to generate multiple sub-training samples;

(3) Calculate the sub-training sample density function;

(4) Calculate the small sample density function and the sub-training sample error and variance;

(5) Use the optimization algorithm to obtain the optimal bandwidth parameter and obtain a new density function, so as to obtain the optimized and corrected sample features.

The advantage of using the Bootstrap-based data sampling method is that the collected data can be resampled to expand the sample size of the data set, which can improve the generalization ability of the deep learning modeling method and obtain the overall distribution characteristics of the sample more accurately.

3.2 time series analysis

Time series analysis is a statistical method applied to dynamic data processing of power and power systems [7]. Based on random process theory and mathematical statistics, this method studies the statistical laws followed by random data sequences to solve practical problems. Generally used for system description, system analysis, forecasting the future, etc.

Usually T_t denote the time trend item, S_t denotes the seasonal variation trend item, C_t denotes the cyclic variation trend item, R_t denotes the random disturbance item. Common deterministic time series models have the following types:

(1) additive model

$$y_t = T_t + S_t + C_t + R_t \qquad (1)$$

(2) multiplicative model

$$y_t = T_t \cdot S_t \cdot C_t \cdot R_t \qquad (2)$$

(3) mixed model

$$y_t = T_t \cdot S_t + R_t$$
$$y_t = S_t + T_t \cdot C_t \cdot R_t \tag{3}$$

where y_t is the observation record of the observation target, $E\left(R_t\right)=0$, $E\left(R_t^2\right)=\sigma^2$.

When building the time series distribution characteristic model of grid infrastructure project costs, the operation steps are as follows:

(1) Collect various cost data

The engineering cost data generated at each stage of large and medium infrastructure projects of 35 kV, 110 kV and 220 kV were collected respectively.

(2) Data preprocessing

Perform preprocessing operations on missing data, abnormal data, and data of different magnitudes in the collected data table.

(3) Establish a time series distribution curve

Draw time series curves, analyze whether the original data and their differences have trends, periodicity, etc. Perform correlation analysis on the drawn time series curve, obtain the autocorrelation function, and select a multiplicative model, an additive model or a mixed model according to the characteristics of the data.

(4) Data augmentation based on Bootstrap

Due to the small number of project data samples collected this time, in order to ensure the accuracy of the model, the Bootstrap method is used to resample the original observation information, and then make statistical inferences about the distribution characteristics of the population.

(5) Time series model after expanding data

After the original data is expanded by the Bootstrap method, the distribution model of cost time series characteristics of infrastructure projects of each voltage level is reconstructed.

4. Evaluation Model of Relevance Between Project Quantity and Investment Budget

4.1 Grey relational analysis method based on AHP

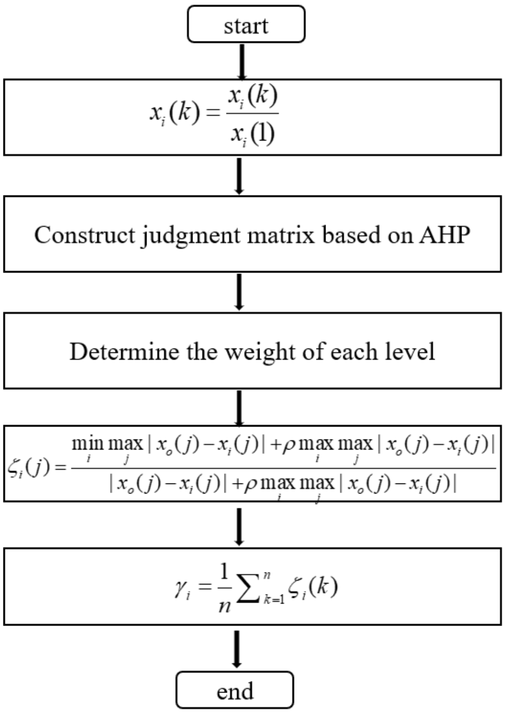

Fig 1 Flow chart of grey relational analysis method based on AHP

Due to the nonlinear coupling relationship between the engineering quantity data and the value data of power grid infrastructure projects, the traditional linear regression fitting method cannot accurately describe the mapping relationship between the two, and there is a large fitting error. A model that can accurately describe the nonlinear relationship between the two; the grey relational analysis method has a higher accuracy of qualitative analysis results, and the AHP method can integrate the influencing factors in the project implementation process into an orderly level of interrelatedness, giving Various influencing factors are matched with a weight, which increases the scientificity of the analysis results [9]. Therefore, the process of obtaining the grey relational analysis method based on AHP is shown in Figure 3. The advantage of this method is that it does not require too much sample size, and does not require a typical distribution law between the engineering quantity and value of power grid infrastructure projects. The grey relational analysis method of AHP can more accurately describe the correlation between the engineering quantity data and the value quantity data of power grid infrastructure projects than the traditional linear regression fitting method.

4.2 Construction of an evaluation model for the correlation between engineering quantity and investment budget

To construct the correlation evaluation model between the engineering quantity and the value quantity of the power grid infrastructure project, the operation steps are as follows:

(1) Data preprocessing

In actual engineering projects, the obtained sample data may contain a large number of missing values and abnormal points caused by manual input errors, which is very unfavorable for the training of algorithm models. Therefore, it is necessary to perform data preprocessing on it to obtain standard and continuous data, which can be used for subsequent data statistics, data mining, etc.

(2) Determine the weights at all levels

As a common analysis method, calculation weights need to be combined with the characteristics of the data to determine the calculation weights of various types of data in practical engineering projects. Common weight analysis methods include factor analysis, principal component analysis, and AHP.

(3) Calculate the correlation coefficient and correlation degree

In this paper, the grey relational method is used to calculate the correlation coefficient between engineering quantity data and value quantity data, and to evaluate the degree of correlation between the two types of data.

(4) Establish a correlation evaluation model

The correlation evaluation model is to complete the comparison of the geometric relationship of the relevant statistical data of the time series in the system time through the quantitative analysis of the development trend of the dynamic process of the infrastructure project, and to obtain the gray correlation degree between the reference sequence and each comparison sequence.

5. Adjustment model based on rolling update

On the basis of the evaluation model of the correlation between the engineering quantity and the investment budget, the subsequent investment budget is predicted by using the occurred engineering quantity and the actual cost. This paper combines Long Short-Term Memory (LSTM) and rolling update adjustment strategy to realize rolling forecast and adjustment of investment budget.

LSTM is a temporal recurrent neural network whose structure is shown in Figure 2. LSTM is also a special recurrent neural network, so it also has a chain structure, but it has a different structure than the repeated modules of the recurrent neural network. It has four neural network layers, and each network layer interacts in a special way, not a single simple neural network layer. As a nonlinear model, LSTM can be used as a complex nonlinear unit to construct larger deep neural networks.

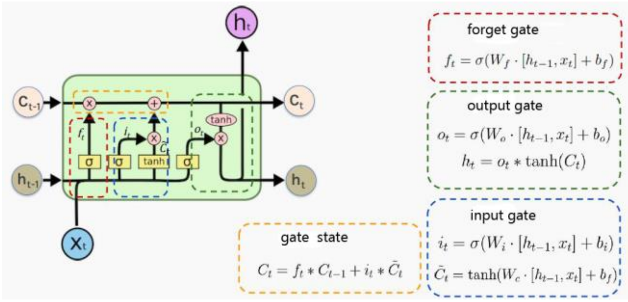

Figure 2 LSTM structure diagram

Considering the business relationship between project construction, investment planning and budget management, integrating the concept of rolling update, according to the actual construction progress and the shortest safe construction period of the project construction branch, the actual construction progress curve is revised in a rolling manner, and then the investment in the remaining period is predicted to be completed in a rolling manner. Measure the amount of adjustment that may be required to issue investment plans and budgets at the beginning of the year, and form a set of feasible quantitative solutions to support investment plans and budget adjustments. The steps to get it are as follows:

(1) Rolling forecast of construction progress plan

The concept of "shortest safety construction period" is introduced [8], and the actual construction progress forecast of the branch is realized according to the actual implementation of the branch project and the shortest safe construction period, and then the rolling forecast of the project construction time is carried out according to the top-down order.

(2) Rolling update of project construction progress

The construction progress rolling update logic is to first divide the construction process into a historical period and a forecast period. For the historical period, the actual monthly construction progress is directly used to replace the update; for the remaining forecast period, according to the rolling update construction progress plan, the monthly construction period will be calculated as a percentage of the remaining total. The proportion of the construction period is used as the forecast value of the monthly construction progress in the remaining period.

(3) Investment completed rolling forecast

According to the rolling forecast results of the construction progress, the forecast value of investment completion is updated rollingly. Its specific forecasting rules: directly use the actual monthly investment completion amount to update the original investment completion forecast value on a rolling basis during the historical period; during the forecast period, according to the quantitative relationship between various costs and construction progress, rolling forecast the investment completion amount in the remaining period.

(4) Budget and plan adjustment calculation

Based on the accumulated data since the start of the project, the historical period and the forecast period are distinguished, the actual situation of historical investment completion is restored, and the planned amount required for the remaining period is forecasted. The total difference

calculated by the two is the recommended plan adjustment amount.

(5) Establish an adjustment model for budgets and plans

Considering the actual application of the project, only predicting the cost of the next month has limited adjustment to the budget. Therefore, LSTM is used to predict the investment in the last quarter of the investment in the first three quarters, and the actual investment completion quota is updated in real time to achieve rolling forecast investment completion. Provide a rough investment reference to provide a basis for budget adjustments, and implement rolling budget updates.

6. Simulation experiment verification

6.1 Experimental data sources and parameters

The data of the power grid infrastructure project in this paper comes from the investment and engineering data of a power company in Sichuan Province from 2016 to 2019. Large-scale projects last from two to three years, and small projects last for several months. The time series distribution characteristic model adopts an additive model, which is decomposed into trend terms and periodic fluctuation terms. LSTM adopts a two-layer hidden layer structure neural network, and the number of nodes in each layer is 8.

6.2 Time series analysis of investment data for power grid infrastructure projects

The time series of the 220 kV project cost is decomposed into trend items and periodic items, and after obtaining the time series model of the sample, the validation set data is retained for model validation.

Figures 3 and 4 show the verification results of the trend item and periodic item of the cost time series distribution characteristic model, respectively. The training and test models of the trend item can match well, but there are some differences at the turning points of the curve. Although the training model and the test model of the periodic item have a certain degree of similarity, the fluctuation of the validation set is large, resulting in a large error between the validation set and the training model. Considering that the disturbance term is included in the annual cycle term, this error is within a reasonable range. Overall, the validation set and the trained model match.

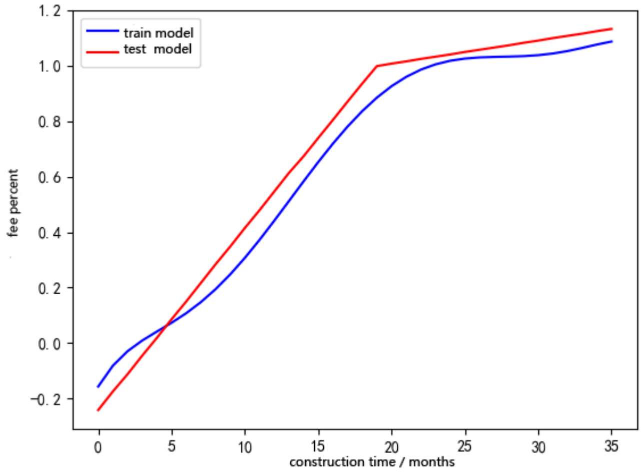

Figure 3 220 kV project cost time series distribution characteristic model trend item

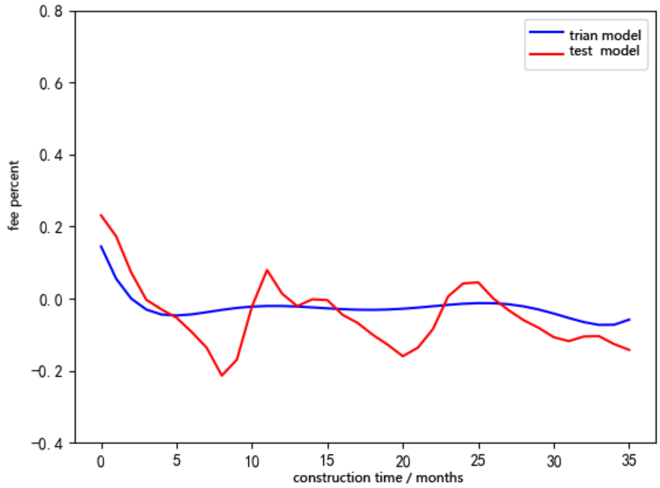

Figure 4 220 kV project cost time series distribution characteristics model period term

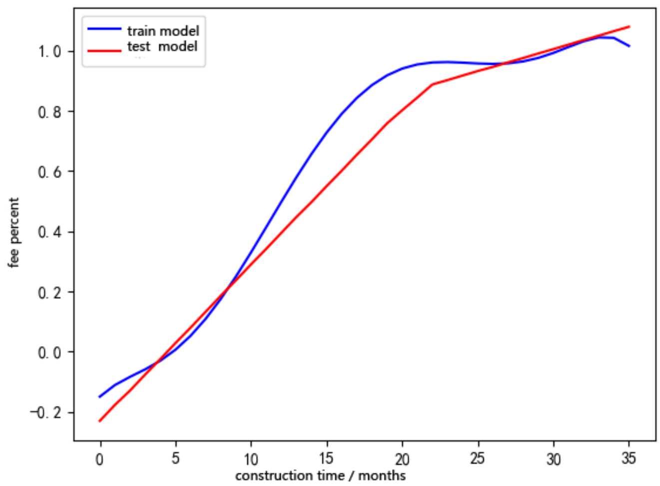

Figure 5 Trend term of time series distribution model for 35kV project

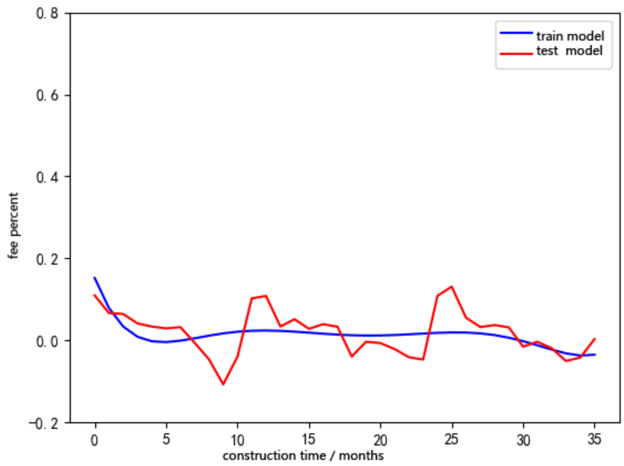

Figure 6 Periodic term of the time series distribution model for 35kV project

In the above figure, the training and testing models of the trend item have a certain error at the turning point of the curve, and the rest have a high degree of similarity. Due to the fluctuation of the period term, there is a certain error between the validation set and the training model, but the trends are basically similar. Overall, the validation set and the training model are highly similar.

6.3 Correlation Analysis of Project Quantity and Investment Data of Power Grid Infrastructure Projects

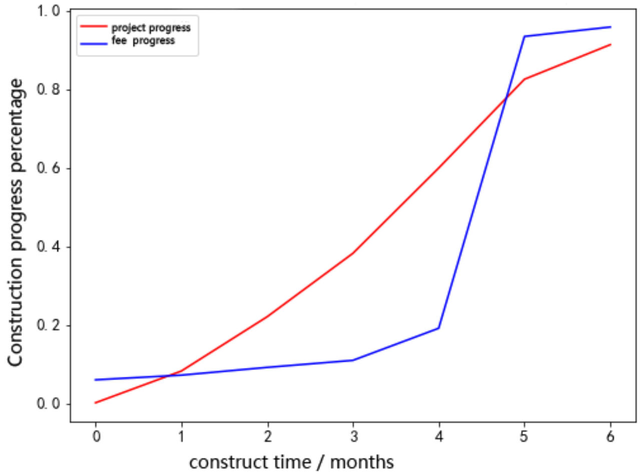

Figure 7 Correlation curve between the value and engineering volume of the 220 kV project

Through the analysis of Figure 7, it can be found that the cost of the project has been generated before the progress of the project, but as the progress of the project continues to advance linearly, the cost does not fluctuate greatly, and it has been stable and increased slowly within the starting range. That is to say, the project has officially started, and while it is steadily advancing, the expenses of the entire project have not yet been in place, or it is in a stagnant state. Until the progress of the entire project progressed to 2/3, the expenses began to increase sharply at this time, directly reaching more than 90% of the total project cost.

From the correlation curve between the value quantity and the engineering quantity, it can be concluded that with the steady progress of the entire engineering project progress, the project cost does not advance synchronously like the trend curve of the engineering project progress, that is, the engineering quantity of the project is more closely related to the value quantity. Low. Ideally, with the steady progress of the project progress, the project value should show the same curve trend.

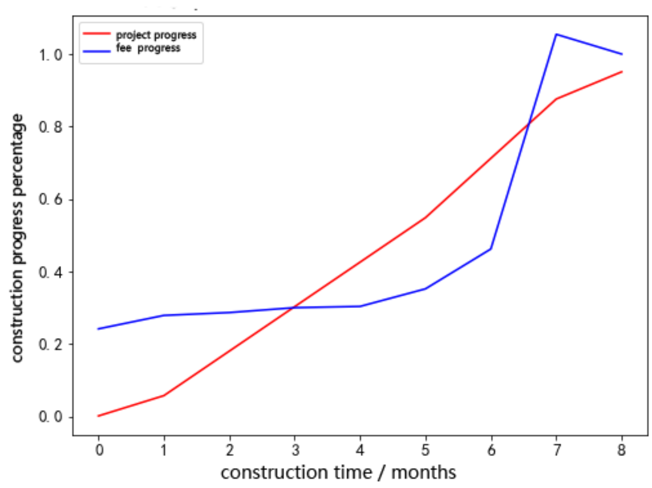

Figure 8 Correlation curve between the value and engineering volume of the 110 kV project

Through the analysis of Figure 8, it can be found that the cost of the project has been generated before the progress of the project. As the progress of the project continues to advance linearly, the cost has not fluctuated greatly, and has been steadily increasing within the starting range. The difference from the 220 kV project is that the cost of the 110 kV project before the start of construction accounts for about 20% of the total cost. Until the progress of the entire project progressed to 2/3, the expenses began to increase sharply at this time, directly reaching more than 90% of the total project cost.

From the correlation curve between the value quantity and the engineering quantity, it can be concluded that with the steady progress of the entire project progress, the project cost does not advance synchronously like the trend curve of the engineering project progress, that is, the project quantity and the value quantity are not related to each other. not tall.

As can be seen from Figure 8, the cost of the project has been generated before the progress of the project. As the progress of the project continues to advance linearly, the cost of the 35kV project is also increasing steadily, ensuring a good linear correlation. At the same time, in the middle and late stage of the project, the project cost will increase sharply to about the total cost, and then it will stabilize in this state.

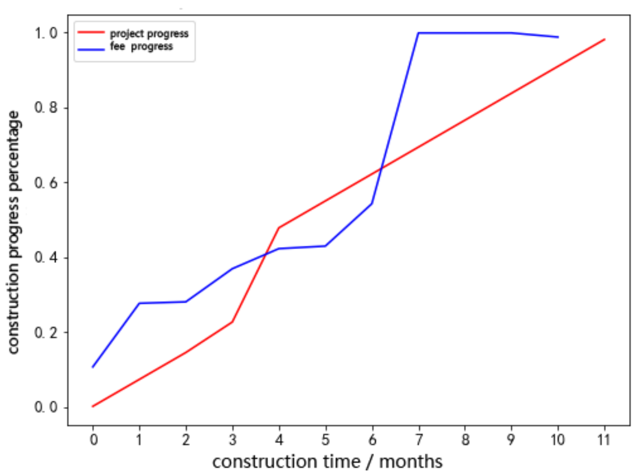

Figure 9 Correlation curve between the value and engineering volume of the 35 kV project

From the correlation curve between the value and the project, it can be concluded that with the steady progress of the entire project, the cost of the project will also be promoted synchronously with the trend curve of the project progress, that is, the project's project and the value are relatively high. Compared with the 220 and 110 kV projects, it can be considered that the degree of correlation between the engineering quantity and the value of the 35 kV project is an ideal state.

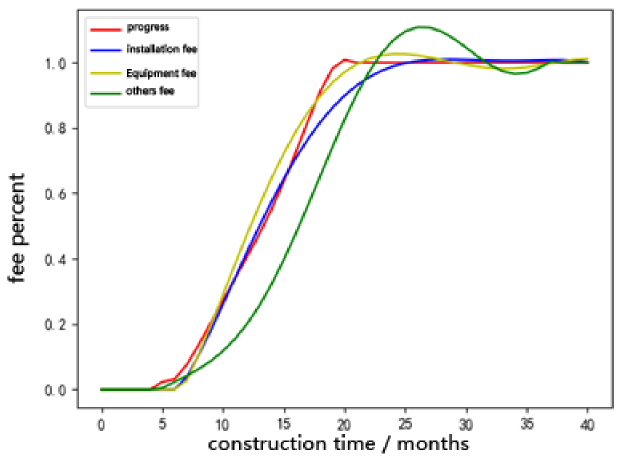

Figure 10 Fitting curve of various costs of 10kV project

Figure 11 Heat map of construction progress and cost correlation

After obtaining the fitting curves of various costs and actual construction progress, the correlation between the curves is obtained through the grey correlation analysis method, as shown in the figure above. In such a sample, construction schedule, installation costs, and equipment costs are all highly correlated, while other costs are less correlated, and while they all have similar upward trends, other costs are relatively lagging and overshoot quantity.

6.4 Investment forecast and result analysis of power grid infrastructure projects

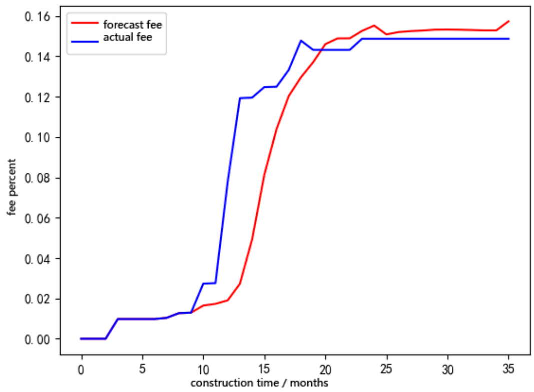

Figure 12 Prediction and actual comparison of other costs of the 110 kV project

As can be seen from Figure 12, due to the sudden increase in the cost of the 110 kV project, there is a certain error between the predicted cost and the actual cost, but in other parts, the deviation between the predicted value and the actual value is small. In addition, due to the particularity of some projects, there is a certain error between the predicted value and the actual value in the verification, but in general, the predicted value is within a certain accuracy range, and effective prediction can be achieved.

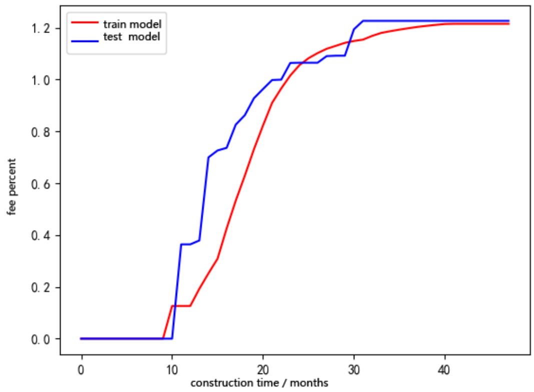

Figure 13 Prediction and actual comparison of other costs of the 35 kV project

As can be seen from Figure 13, although the growth rate of other costs of the 35 kV project is relatively large in a certain period of time, it is not as obvious as the installation cost and equipment cost. Therefore, although there is a small amount of error, the prediction can basically be achieved. Therefore, it can be proved that the built tuning model is valid.

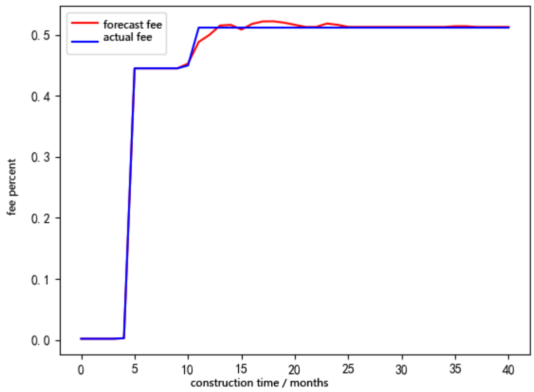

Figure 14 Prediction and actual comparison of other costs of the 10 kV project

Part of the deviation between the actual value and the predicted value is large, mainly concentrated in the period of sudden cost increase, and the model cannot predict the sudden cost increase, which is caused by the management of the project. With the progress of the construction period, the cost entry is linearly advanced in theory, but many projects adopt the method of centralized entry, and the entry has a large delay, which has a great impact on the forecast of the next month's cost, resulting in inaccurate budget adjustment. Not timely.

Although there is a certain error between the prediction and the actual value, the actual value can be tracked in general, but there is a certain delay relative to the actual value. Therefore, it has a good tracking effect and reflects the change trend of the actual value.

7. Conclusion

By constructing a business model for the connection between engineering quantity and value of power grid infrastructure projects, and applying it to the "three-in-one management and control system", the real-time process management and control of the project can be strengthened; Adjust the investment plan and budget plan in time, so that the project quantity and value can be highly matched; it can also help managers to formulate investment plans more accurately, improve the accuracy of investment plans and budget management of power grid enterprises, and realize scientific management of project quantity funds.

References

[1] Bin, ZHU, Yun, et al. Research on Bill of Engineering Quantity and Calculation Standard for Power Grid Marketing Project[J].

[2] Chen D, Lizhong QI, Jin LI, et al. Analysis and Forecasts on Quantity and Economy of Large Angle Steel Popularization in Power Grid Projects During the 12th Five-year-plan Period[J]. Electric Power Construction, 2012, 33(2):79-83.

[3] Yu Z, Meng X. Research on economic evaluation model for distribution network operation based on grey relation method[C]. 2011 4th International Conference on Electric Utility Deregulation and Restructuring and Power Technologies (DRPT). IEEE, 2011.

[4] Yan-Hui L U, Pan C S. Grey Evaluation Model based on AHP and Its Application[J]. Fire Control and Command Control, 2005.

[5] Niu D X, Zhou K, Li YY, et al. Evaluation of Grid Performance Based on Grey AHP Model[J]. Advanced Materials Research, 2011, 341-342: 490-493.

[6] Y. Wu and Z. Huang, "Application of a Case-Based Reasoning Method in Estimating the Power Grid Project Cost," 2008 4th International Conference on Wireless Communications, Networking and Mobile Computing, 2008, pp.1-5, doi:10.1109/WiCom. 2008.1743.

[7] J. Li et al., "A Survey on Investment Demand Assessment Models for Power Grid Infrastructure," in IEEE Access, vol. 9, pp.9048-9054, 2021, doi: 10.1109/ACCESS.2021.3049601.

[8] J. Hong et al. "Establishment and application of the CW-MEEM model for post-evaluation of power grid construction projects" 2022 7th Asia Conference on Power and Electrical Engineering (ACPEE), 2022, pp.1352-1357, doi:10.1109/ ACPEE 53904. 2022. 9783681.

[9] G. Luo and Y. Zhang, "Study on Investment Optimization for Power Grid Construction Projects," 2009 First International Conference on Information Science and Engineering, 2009, pp. 5150-5153, doi:10.1109/ICISE.2009.1126.

[10] Li, KC et al., Comprehensive evaluation index system and weight grey correlation analysis of electric energy substitution project, AIP Conference Proceedings, Vol. 2036, No. 1(2018), DOI:10.1063/1.5075645.