Research on Prediction Model of Cable Line Cost Based on Least Square Support Vector Machine

Bo Yu\textsuperscript{1*}, Ruixin Gou\textsuperscript{1}, Xin Ju\textsuperscript{1} and Dongni Wei\textsuperscript{1}

\textsuperscript{1}Ningxia Electric Power Co., Ltd. Institute of Economy and Technology, Yinchuan, Ningxia, 750000, China

\textsuperscript{*}Corresponding author’s e-mail: hbdlchengxiaobin@126.com

Abstract. With the development of China's economy and society, the total social electricity consumption has risen dramatically, and the investment scale of transmission and transformation projects has increased accordingly. Cable line engineering is a kind of power transmission and transformation engineering, which is buried deep underground, has complex construction conditions and high cost. Compared with substation engineering and overhead line engineering, its cost control is more difficult. In order to facilitate the fine management of cable line cost, this paper analyses the influencing factors of cable project cost based on grey relational analysis theory, uses principal component analysis method to reduce the dimension of influencing factors, and establishes a cable line cost prediction model based on least squares support vector machine method. It can guide the economic evaluation of investment decision of cable line project. The results show that the accuracy of the cost prediction model based on least squares support vector machine method can reach 95.06%. The research in this paper is innovative and practical.

1. Introduction

Power transmission and transformation projects are routine infrastructure projects of power grid enterprises. Especially in recent years, with the increase of residential electricity load, the scale of power transmission and transformation projects has increased year by year, resulting in a large amount of capital occupied by power grid enterprises. Therefore, formulating a reasonable annual investment plan is the key to reduce the investment balance rate and financial expenditure of power grid enterprises. Cable line engineering is a very important form of power transmission and transformation engineering construction in urban areas with large population density and scarce land resources. The reasonable prediction of the cost of cable line engineering is the first condition to ensure the scientific investment of cable line engineering. The accuracy and efficiency of cable engineering prediction by traditional prediction methods are low. The development of artificial intelligence algorithm provides a new method for the cost prediction of cable lines and improves the effect of cost control.

At present, some scholars have done a lot of research on project cost prediction, and achieved remarkable results. Sun Anli [1] aimed at the problem that there are many kinds of projects in the propahse of transmission line project and the deviation of cost estimation is large, and it is difficult to get accurate project cost with a small amount of information, a prediction model of transmission line project cost is proposed based on BP neural network algorithm. Liang Xi [2] considered the fuzziness of decision maker's judgment, the intuitionistic fuzzy AHP method is applied to screen the engineering features which have great influence on the construction cost, and it is used as the input vector of the neural network model to construct the prediction model of the construction cost based on the fuzzy...
neural network. Hossain, A [3] uses GARCH model instead of ARMA model, compares it with standard BP and SVM model, and predicts four international indexes including two Asian stock market indices. Wang, ZW [4] proposed a prediction method of regional electricity consumption based on least squares support vector machine. The error between the prediction model of least squares support vector machine and the prediction model of support vector machine is compared, which can accurately predict power consumption. Ming, Z [5] introduced the basic idea and algorithm of support vector machine, studied the basic principle and specific algorithm description of least squares support vector machine, including the kernel function, and finally studied the application of this algorithm in classification. Based on the grey system prediction model, neural network prediction model and support vector machine prediction model, Gao, S[6] established the linear combination prediction model, neural network combination prediction model and support vector machine combination prediction model. The results show that the support vector machine method is better than the single prediction method, linear prediction method and spirit. The prediction accuracy of the network combination prediction method is higher. He, XS [7] proposed an improved classification algorithm based on least squares support vector machine (LS-SVM) for non-uniform sample classification by using extension discriminant method and introducing correlation. The effect of misclassification cost on classification was also considered.

Through the analysis of the achievements of relevant scholars, we can find that the theory of project cost prediction is relatively mature at this stage, but the research on cable line project cost prediction is less. Least Square Support Vector Machine (LS-SVM) is widely used in various fields as a prediction tool with high prediction accuracy and low sample quality requirements. The least squares support vector machine (LS-SVM) method is applied to the cable line engineering cost prediction, which is innovative and practical.

2. Theoretical method

2.1. Least Squares Support Vector Machine
Least squares support vector machine (LS-SVM) is an improved algorithm of traditional support vector machine (SVM), which inherits a series of excellent characteristics such as support vector machine kernel function, structural risk minimization principle and small sample size. Based on regularization theory, the loss function in support vector machine is defined as least squares loss function, the inequality constraints in support vector machine are transformed into equality constraints, and the problem of solving quadratic programming is transformed into the problem of solving linear equations.

LSSVM introduces an error variable to each sample and adds a regular term of the error variable to the original function in order to solve the problem of some specific points. For non-linear separable training samples, the original samples can be mapped to higher dimensional linear separable space. Compared with the traditional support vector machine method, it reduces the complexity of the model and improves the efficiency of operation.

2.2. Grey relational analysis
The basic idea of grey relational analysis is to judge whether the sequence curves are closely related according to the similarity of their geometric shapes. If the two related factors are closer to each other, the greater the correlation degree between the corresponding sequences is, and vice versa, the smaller the grey correlation degree is, the more consistent the change trend of the two factors is. Grey relational analysis is also applicable to the number of samples and the regularity of samples. Moreover, the calculation is small and convenient, and there will be no discrepancy between quantitative results and qualitative analysis results.
3. Establishment of Cable Line Cost Prediction Model and Case Analysis

3.1. Identification and Screening of Influencing Factors of Cable Line Project Cost

The main factors influencing the cost in cable line engineering are voltage grade, number of cable circuits, length of cable folding, cable cross section, number of cable intermediate joints, number of cable terminal joints, full length of civil engineering, length of open tunnel, length of tunnel excavated under cover, cable unit price, unit price of intermediate joints, unit price of terminal joints, etc. Therefore, the influence mechanism of each factor is as follows: 110 kV and 220 kV voltage levels are the main types of urban cable line projects. The cost of different voltage levels of cable lines differs greatly. The higher the voltage level, the higher the cost. The number of cable loops and the length of cable fold determine the consumption of cable materials. Cable section is mainly 400 mm$^2$ and 800 mm$^2$. According to the size of cable section, the project of cable line is classified, and the cost is quite different under different categories. Cable joints are mainly used to connect cables, including intermediate joints and terminal joints. The cost of joints is relatively high. Generally, the unit cost of joints is between 10,000 yuan and 30,000 yuan. If the number of joints of cable lines is large, the cost will increase greatly. The proportion of cable civil engineering in cable line engineering is smaller than that of cable electrical engineering, but it is still very high. Therefore, the influence of the length of cable civil engineering on the total cost should not be neglected. The cost of tunnel engineering is the main component of the cost of civil engineering, and the larger the proportion of tunnel excavation, the better the effect of cable civil engineering on the total cost. The higher the cost of cable line is. The purchase cost of cable and connector is the main part of the installation equipment cost of cable, so the unit price of cable, the unit price of intermediate connector and the unit price of terminal connector are also the main factors affecting the cost.

Based on the theory of grey relational analysis, this paper analyses the influence degree of each influencing factor on the cost of cable line project. 39 reliable samples are selected from the cable line project completed by a power company in recent years for analysis. The sample data are shown in the table.

| Sample number | Voltage grade | Cable length | Cable cross section | Full length of construction project | Excavation tunnel ratio | Static investment (ten thousand yuan) |
|---------------|---------------|--------------|---------------------|-------------------------------------|------------------------|--------------------------------------|
| 1             | 110           | 12720        | 800                 | 2050.00                             | 100.00                 | 7674.57                              |
| 2             | 110           | 1308         | 800                 | 179.52                              | 82.56                  | 968.68                               |
| ...           |               |              |                     |                                     |                        |                                      |
| 39            | 110           | 1890         | 630                 | 243.10                              | 0.00                   | 795.02                               |

The main analysis steps of grey correlation are as follows:

(1) Determination of Reference Sequence and Comparison Sequence

The static investment of cable engineering is the sum of the static investment of cable electrical and the static investment of cable civil engineering. Voltage grade, cable length, cable cross section, number of cable intermediate joints, number of cable terminal joints, cable unit price, intermediate joint unit price, terminal joint unit price, cable construction project full length, open tunnel length, length of tunnel excavation and proportion of tunnel excavation are selected as comparison series, and these factors are analyzed for electricity. The impact of cable line project cost.

(2) Dimensionless Reference Sequence and Comparative Sequence

In this paper, the mean method is used to deal with the reference sequence and the comparison sequence dimensionless. The average method divides the data of a sequence by the mean of the sequence. Since the average of a sequence of large orders of magnitude is relatively large, the data can be normalized to a nearby area after the average processing.

Taking the static investment as an example, the average static investment of these 39 cable line projects is 163.034 million yuan. The non-dimensional result of the average method is dividing the
static investment value of each project by the average value. The non-dimensional result sequence is 0.47, 0.06, 0.07, 0.13, 0.28, 0.55, 0.17, 0.08, 0.42, 0.03, 0.47, 0.21, 0.10, 0.35, 0.10, 0.42, 0.20, 0.16, 5.11, 2.91, 3.93, 0.62, 1.06, 0.68, 0.54, 1.76, 0.90, 0.13, 0.07, 2.49, 0.36, 3.92, 0.85, 6.70, 1.91, 0.30, 0.42, 0.03, 0.05.

(3) Computation of correlation coefficient
Firstly, the dimensionless comparison sequence is subtracted from the reference sequence to get a set of matrices, and the maximum and minimum values of the matrices are found out. Then the correlation coefficient matrix is calculated. For example, the correlation coefficients of cable length and static investment in these 39 cable line projects are 0.34, 0.75, 0.68, 0.54, 0.26, 0.64, 0.74, 1.21, 0.78, 0.35, 0.60, 0.71, 0.46, 0.71, 0.40, 0.62, 0.65, 3.49, 1.28, 2.31, 0.19, 0.25, 0.13, 0.27, 0.94, 0.09, 0.68, 0.86, 1.27, 2.30, 0.78, 5.08, 1.09, 0.51, 0.39, 0.78, 0.76.

(4) Computation of correlation degree
The correlation coefficients between cable length and static investment of 39 cable line projects are averaged, and the correlation degrees between these 12 factors and static investment are 0.8269, 0.8997, 0.8563, 0.9015, 0.7965, 0.8408, 0.8628, 0.8558, 0.8680, 0.8032, 0.8792 and 0.8161, respectively.

(5) Relevance ranking
After ranking the above 12 correlation degrees, the correlation degree of each factor to static investment is as follows: number of cable intermediate joints > length of cable > length of tunnel excavated under ground > full length of cable construction project > unit price of intermediate joints > section of cable > unit price of terminal joints > unit price of cable > voltage level > proportion of tunnel excavated under ground > length of open tunnel > end of cable.>number of end joints.

According to the results of grey correlation analysis, it can be concluded that the 12 relevant factors mentioned in this paper have a significant impact on the cost of cable line engineering. It can be considered that the correlation degree between each factor and static investment is above 0.8, and the correlation is good. The control of these factors should be taken into account in the management and control of cable line engineering cost. In addition, due to the different degree of influence of different factors, the degree of control of different factors should be different.

3.2. Prediction of Cable Line Cost
Cable line engineering is affected by natural, technological, economic, environmental and other factors. There are many factors that influence cable line engineering. There are often correlations among these factors. If the original index is used as the input index of cost prediction, it will face the problems of high data dimension, poor fitting effect and inaccurate prediction results. Principal Component Analysis (PCA) is a method to reduce the dimension of indicators. The principal components obtained by PCA can represent the vast majority of the information of the original variables and are not related to each other, which is helpful to the establishment of the cable project cost prediction model and problem analysis. Principal Component Analysis (PCA) combines P indices which have some correlation with each other linearly, and combines them into a new set of comprehensive indices which are independent of each other to replace the original indices. The cumulative contribution rate of information contained in these principal components is more than 85%. The main steps of PCA are: (1) standardizing the original data; (2) calculating the sample correlation coefficient matrix; (3) calculating the eigenvalues and corresponding eigenvectors of the correlation coefficient matrix; (4) selecting the important principal components and writing the principal component expression.

Voltage grade and cable section are classified as variables in cable line engineering cost prediction, because the corresponding values of these two factors are discrete, and there are several specific choices in different cable line projects, so these two factors are not considered in PCA. The other 10 factors were analyzed by principal component analysis, and the dimensionality of these 10 factors was reduced by SPSS software. Finally, three principal components were obtained. These three principal
components can express more than 85% of the information of these variables, and the rotation component matrix of principal component analysis was obtained as shown in the table below.

| Principal Component | 1    | 2    | 3    |
|---------------------|------|------|------|
| Cable length w₁      | 0.779| 0.446| 0.154|
| Number of intermediate joints w₂ | 0.873| -0.004| 0.173|
| Number of terminal joints w₃  | 0.375| -0.656| 0.539|
| Cable unit price w₄    | 0.887| -0.286| 0.111|
| Intermediate joint unit price w₅ | 0.884| -0.291| 0.072|
| Terminal joint unit price w₆ | 0.894| -0.304| 0.017|
| Overall length of construction w₇ | 0.756| 0.604| 0.040|
| Open tunnel length w₈  | -0.263| 0.664| 0.612|
| Dug tunnel length w₉  | 0.839| 0.449| -0.114|
| Excavation tunnel ratio w₁₀ | 0.553| 0.047| -0.766|

Take Principal Component 1 as an example:

Principal component 1 = 0.779 × w₁ + 0.873 × w₂ + 0.375 × w₃ + ... + 0.839 × w₆ + 0.553 × w₁₀

After principal component analysis, the input model of cost prediction model is simplified into five variables: voltage level, cable section, principal component 1, principal component 2 and principal component 3.

After determining the input variables of cost prediction, the least squares support vector machine model is used to predict the cost of cable line. The standard LSSVM model is established and the data are substituted for the model training. The basic parameters of LSSVM prediction model are penalty coefficient and empirical parameter C=100, σ²=0.4 and the kernel function is Radial Basis kernel function.

\[ K(x, x_i) = \exp\left(-\frac{||x - x_i||^2}{2\sigma^2}\right) \]

The 39 samples were divided into two groups: training data and test data. The sample number of training data is set to 30, and the sample number of testing data is set to 9. The calculation program is written by MATLAB software, and the fitting curve of training sample points is obtained as shown in the following figure.
Figure 1. Fitting Curve of Training Sample Points

The actual value of the test sample set is compared with the predicted value as shown in the following figure.

Figure 2. Contrast chart between actual and predicted values of test sample set

From the above figure, it can be seen that the deviation between the actual value and the predicted value of the cable line project established by the least square support vector machine method is small, and the percentage of deviation of all test samples is less than 10%. The prediction accuracy is 95.06% after calculation, and the prediction effect is good, which shows the validity of the model.
4. Conclusion
Firstly, this paper identifies the key influencing factors of cable line project cost, uses grey relational degree analysis method to determine the relational degree between each influencing factor and static investment, and uses principal component analysis method to reduce the dimension of 10 factors except voltage level and cable cross-section to get three principal components, so as to make the cost prediction model. The input variable is transformed into five variables, which reduces the complexity of the problem. Then, a cable engineering cost prediction model is established based on least squares support vector machine. The prediction accuracy can reach 95.06%, and the accuracy is high. Therefore, the model established in this paper has strong practicability and innovation, and can be used for cable engineering of power grid company. The work of cost control provides a theoretical basis for reference.

References
[1] Anli Sun, Chun Xi, Hanxi Wu. Research on Prediction Model of Transmission Line Cost Based on BP Neural Network [J]. Modern Electronic Technology, 2018.
[2] Xi Liang, Yu Liu. Forecasting Model of Construction Cost Based on Fuzzy Neural Network [J]. Technical economy, 2017(3).
[3] Hossain A, Zaman F, Nasser M, et al. Comparison of GARCH, Neural Network and Support Vector Machine in Financial Time Series Prediction[C]// Pattern Recognition and Machine Intelligence, Third International Conference, PReMI 2009, New Delhi, India, December 16-20, 2009 Proceedings. DBLP, 2009.
[4] Wang Z, Niu Y. Regional electricity consumption based on least squares support vector machine[C]// International Conference on Machine Vision. International Society for Optics and Photonics, 2013.
[5] Zhao M. Research on Least Squares Support Vector Machines Algorithm[C]// International Industrial Informatics & Computer Engineering Conference. 2015.
[6] Gu L, Wu H Z. Applying a novel decision rule to the sphere-structured support vector machines algorithm[J]. Neural Computing & Applications, 2009, 18(7):675-675.
[7] Tamura H, Tanno K. Midpoint Validation Method for Support Vector Machine with Margin Adjustment Technique[C]// Innovative Computing Information and Control, 2008. ICICIC '08. 3rd International Conference on. IEEE, 2008.