Study on prediction model of construction period of power transmission and transformation project based on support vector machine

Qingmeng Yan

1Ministry of Construction, State Grid Fujian Power Co., Ltd., Fuzhou, Fujian Province, 350003, China

Corresponding author’s e-mail: hbbdcc@126.com

Abstract. With the improvement of people's living standard, the demand for electric energy is increasingly strong. In order to meet the increasing load, the construction scale of power transmission and transformation project of power grid company is increasing, and the difficulty of progress control of power transmission and transformation project is increasing. In order to control the progress of power transmission and transformation projects, it is necessary to make an accurate and reasonable time prediction to make a realistic and feasible schedule. Firstly, this paper analyses the influencing factors of the construction period of power transmission and transformation projects, and determines the key decision factors. Based on the support vector machine method, a prediction model of the construction period of power transmission and transformation projects is established. The prediction accuracy reaches 95.50%, with high accuracy.

1. Introduction

The prediction of construction period of power transmission and transformation project is a systematic project. The selection and treatment of factors and the selection of prediction methods will have a great impact on the accuracy of prediction. In the past work, the scheduling of power transmission and transformation projects is mainly based on experience, and the utilization rate of a large number of project data is relatively low. To solve this problem, we establish a prediction model for the construction period of power transmission and transformation projects based on a number of completed project duration and influencing factor data.

In the aspect of project duration prediction, some scholars have done some work and achieved good results. Hu Wenfa et al. [1] identified the main factors and derived factors that affect the construction period, and used multiple linear regression method to obtain a preliminary model that meets the hypothesis test conditions. After comparison, it was found that the logarithm construction speed as the dependent variable of the construction period model was more accurate. In view of the imperfection of the project duration prediction index and its corresponding prediction method in the existing earned value management index system, Yu Xiaozhong [2]established the project completion duration prediction index, and studied the prediction method of project duration in different situations according to the different time deviation during project inspection and the change of project cost. Yoon Jongsik [3] analyses the factors that affect the reconstruction construction period, makes a comparative analysis of the new housing project and the reconstructed housing project, and has been verified by experts. The analysis conducted by his research institute can provide a reference for the
estimation of the reconstruction construction period in the future. SeokHeonYun et al.[4] first studied the development trend of construction period estimation, and studied the current situation of construction period estimation with the method of fuzzy logic, obtained the determined risk factors and suitable construction period, established the construction period estimation model with fuzzy logic, and finally compared the results based on fuzzy logic with the existing construction duration model.

Support vector machine is a new small sample learning method with solid theoretical basis, which simplifies the usual regression and classification problems. The final decision function is only determined by a small number of support vectors. The complexity of calculation depends on the number of support vectors, rather than the dimension of the sample space, which avoids the "dimension disaster" to some extent. It is feasible, scientific and reasonable to establish the prediction model of electrical period of transmission and transformation by support vector machine method, which has strong practical value.

2. Determination of key factors affecting the construction period of power transmission and transformation projects

2.1. Classification of influencing factors

On the basis of expert investigation, this paper divides the influencing factors of transmission and transformation project construction schedule into five categories: natural factors, management factors, technical factors, policy factors and social factors. Among them, natural factors refer to the factors that affect the progress of the project in the process of survey, design and construction of power transmission and transformation projects, such as the increase of construction difficulty due to terrain and geological conditions, and the shutdown caused by severe weather conditions such as floods and typhoons. Management factors refer to the factors that affect the project progress due to the requirements of the company's management system, or the inadequate management arrangements of the construction unit and the design unit, such as the review of the preliminary design scheme, the construction drawing design, and the commencement schedule. Technical factors refer to the factors that affect the progress of the project due to the constraints of technical capacity and objective conditions during the design and construction of power transmission and transformation engineering, such as the quality of preliminary design and construction drawing design, the working effect of four supplies and one leveling. Policy factors refer to the progress influencing factors faced by communication and negotiation with government departments and handling all kinds of government approval documents and procedures during the development of power transmission and transformation projects, such as relevant land use approval documents of governments at all levels. Social factors refer to the factors that affect the progress of the project due to the different cultural and social environment in different areas during the development of the power transmission and transformation project, such as the development of land acquisition, the obstruction of construction, crossing or adjacent cultural and natural reserves during the construction process, etc.

2.2. Determination of decision factors in construction stage of power transmission and transformation project based on correlation analysis

Correlation analysis is to analyse the signs that do have a connection on the whole, and its main body is to analyse the signs that have causality on the whole. Factor marker is the condition that determines the development of result marker. According to the different responses of result marker to factor marker, it can be divided into two types: function relation and correlation relation. The function relationship is that when the number of factor markers is determined, the number of result markers is also completely determined, which is represented by y=f(x). The value of the factor flag may have several values of the result flag.

The basic steps of simple correlation analysis are as follows:

(1) Draw a scatter diagram. Draw a scatter chart to see if there are regular changes between the two variables.
(2) Select the factor category. According to the type of variables or normal test, choose the appropriate correlation coefficient formula.

(3) Calculate the correlation coefficient. The correlation coefficient \( r \) was calculated to evaluate the correlation degree.

(4) Significance test. Carry out significance test, if \( P < \alpha \) (generally 0.05), it means there is a significant correlation.

(5) Conduct business judgment. Summarize the analysis conclusion, and give the business judgment and business strategy from the business level.

Based on the analysis of the relationship between the commencement section and the construction period of the line project, the commencement time is divided into four sections according to the influence degree of each month of the year on the construction period, which respectively represents the four situations of the influence on the construction period. The starting section and construction period data of 91 completed overhead line projects in 9 cities of a province from 2014 to 2018 were imported into SPSS software. Pearson correlation coefficient was selected to analyse the correlation between the two variables, mean value and standard deviation were selected for descriptive analysis, and the size and dispersion of the data were described to a certain extent. Pearson correlation coefficient, significance test result, mean value and standard deviation calculated by SPSS data analysis software are shown in the figure 1.

| Correlations                                      | Commencement time | Construction period |
|---------------------------------------------------|-------------------|---------------------|
| commencement time Pearson Correlation             | 1                 | 0.223*              |
| Sig. (2-tailed)                                   | 91                | 91                  |
| construction period Pearson Correlation           | 0.223             | 1                   |
| Sig. (2-tailed)                                   | 0.033             |                     |
| N                                                 | 91                | 91                  |

* Correlation is significant at the 0.05 level (2-tailed).

Figure 1. Correlation analysis results of commencement section and construction period.

From the above correlation analysis results, we can see that the significance of the correlation analysis between the commencement section and the construction period is 0.033. In general, we take \( \alpha = 0.05 \). When the significance is less than 0.05, it indicates that there is indeed a correlation between variables. One of the variables can be used as the influencing factor of the other variable when establishing the model as the input variable for analysis. Finally, the key factors influencing the construction period of the overhead line project are the project area, voltage level, starting month, rainfall, line fold length, altitude, terrain, geology and the operation capacity of the construction team.

3. Construction of intelligent prediction simulation model based on support vector machine

3.1. Support vector machine theory

Support vector machine (SVM) can deal with the linear non separability of data by relaxing variables and kernel function technology, so it has unique advantages in nonlinear pattern recognition and can be extended to other machine learning problems such as regression analysis. The support vector machine method is based on the dimension theory of statistical learning theory and the principle of structural risk minimization.

There are three parameters in the SVM regression algorithm, in which the accuracy of \( \varepsilon \) is the control function fitting error; the penalty factor \( C \) is used to adjust the proportion of learning machine confidence range and empirical risk in the determined data subspace, so as to have the best
generalization ability. A small value of C indicates a small penalty for empirical error, a small complexity of SVM, and a large empirical risk value. The function of kernel function parameter $\sigma$ is to affect the complexity of sample data distribution in the high-dimensional feature space. The larger $\sigma$ is, the wider the output response interval of the sample is, the smaller the structural risk of the optimal classification surface is, but the greater the empirical risk is. The smaller $\sigma$ is, the output response interval of the sample is smaller, the narrower, the greater the risk of the optimal classification surface structure, but the empirical risk will be reduced, but it is easy to cause over fitting phenomenon. It can be seen that the selection of SVM parameters, especially the selection of penalty factor C and kernel function parameter $\sigma$, has a great impact on its performance.

3.2. Construction period prediction model of support vector machine for overhead line project

In 2014-2018, among 91 completed overhead line projects in a province, 10 projects were randomly selected as validation samples, and the rest 81 projects were used as training samples to establish the construction period prediction model of support vector machine. The fitting curve of training sample points obtained through MATLAB software is shown in the figure 2.

![Training Sample Point Fitting Curve](image)

Figure 2. Training sample point fitting curve.

In the figure, the blue dot represents the actual construction period value, and the red line is the curve obtained by fitting the blue dot based on the support vector machine method. The sample numbers of each city are concentrated in a certain section of the abscissa axis, so we can see that the actual value of the construction period of the overhead line project in each city is quite different. The process of building simulation prediction model based on support vector machine method is closely related to the fitting curve. After the curve fitting, the prediction of construction period can be realized by certain algorithm.

The comparison between the actual value of the validation sample and the predicted value of the construction period prediction model of the overhead line project based on the support vector machine method is shown in table 1.

| Sample number | Actual duration | Forecast duration | Deviation | Deviation percentage |
|---------------|-----------------|-------------------|-----------|----------------------|
| Sample 1      | 428             | 401.72            | -26.28    | -6.14%               |
| Sample 2      | 426             | 401.90            | -24.10    | -5.66%               |
| Sample 3      | 388             | 400.62            | 12.62     | 3.25%                |
| Sample 4      | 370             | 399.75            | 29.75     | 8.04%                |
| Sample 5      | 432             | 400.80            | -31.20    | -7.22%               |
| Sample | Value 1 | Value 2  | Error | Error Percentage |
|--------|---------|----------|-------|-----------------|
| 6      | 392     | 392.18   | 0.18  | 0.05%           |
| 7      | 373     | 390.60   | 17.60 | 4.72%           |
| 8      | 421     | 404.62   | 6.62  | 1.66%           |
| 9      | 372     | 386.60   | 14.60 | 3.92%           |
| 10     | 428     | 409.54   | -18.46| -4.31%          |

It can be seen from the table and figure that there are obvious differences between the actual value and the predicted value of the construction period of overhead line projects in different cities. The absolute value of the prediction error of the intelligent prediction simulation model based on the support vector machine method for the validation samples in different cities is within 10%, and the total prediction accuracy is 95.50%. The prediction effect is good, which can meet the precise construction period of overhead line projects. Management requirements for quasi pre control.

4. Conclusion
In this paper, from the perspective of grid company to improve the management and control of transmission and transformation project progress, the factors that have a greater impact on the transmission and transformation project are first combed, and the key influencing factors are determined through correlation analysis. Taking the key influencing factors as input variables and the construction period as output variables, a prediction model of transmission and transformation project duration based on support vector machine method is established. The influencing factors determined in this paper cover all aspects of the construction period of power transmission and transformation projects, which are relatively complete and comprehensive. The construction period prediction model based on support vector machine has low requirements for sample quality, high prediction accuracy, innovation and practicability, and can be applied to the schedule making of similar projects.

References
[1] Hu, W., Chang Y., He X. (2018) Influencing factors and prediction model of construction project duration. Journal of Civil Engineering, 2018, 51(02):103-112.
[2] Yu, X. (2009) Research on project duration prediction method based on Earned Value Management. Statistics and decision making, 2009(02):168-170.
[3] Yoon, J., Jung, D., & Yu, I. (2017). Factor Analysis for Development of Construction Period Calculation Model in Apartment House Remodeling. In M. Hajdu & M. E. Skibniewski (Eds.), Creative Construction Conference 2017, Ccc 2017 (Vol. 196, pp. 660-665). Amsterdam: Elsevier Science Bv.
[4] SeokHeonYun, & Paek, J. (2009). A Study on Estimating Construction Period Model based on Fuzzy Logic - Focused on Risk Weight in Project. Journal of the Architectural Institute of Korea Structure & Construction, 25(07), 103-112.
[5] Rather, A. M., Agarvval, A., & Sastry, V. N. (2015). Recurrent neural network and a hybrid model for prediction of stock returns. Expert Systems with Applications, 42(6), 3234-3241.
[6] Cao, J. Y., Fang, Y. L., Liu, Q., Liu, A. M., & IEEE. (2016). Combined Prediction Model of Quantum Genetic Grey Prediction Model and Support Vector Machine. New York: IEEE.
[7] Li, G. H., Zhang, S. L., & Yang, H. (2017). A Deep Learning Prediction Model Based on Extreme-Point Symmetric Mode Decomposition and Cluster Analysis. Mathematical Problems in Engineering, 6.