Transmission Technical Project Investment Forecasting Under the High-dimension Small Sample Conditions

Mianbin Wang 1, Xia Qi1, Pengyun Geng1, Lei An1, Jianjun Wang2,*

1Electric Power Economic Research Institute in northern Hebei Technology, State Grid, Beijing, China
2North China Electric Power University, Beijing, China

*Corresponding author e-mail: wangjianjunhd@126.com

Abstract. The electricity market will realize the separation of transmission and distribution, in the new environment, the main profit model of the power grid company will be dominated by transmission and distribution services, it will require the grid company to control its investment and cost accuracy. As a routine maintenance work to ensure the safe and stable transportation of electric energy, the grid technical transformation project has a great impact of the grid operation cost. Due to the lack of the history data, it is difficult to forecast the project’s investment, and the current investment always relies on experience. It needs to construct a quantitative forecasting model to make scientific and objective predictions of investment in technological transformation projects. In this paper, the principal component analysis (PCA) is employed to analysed the history data, then the main factors are obtained, which are used as input variables into SVM models. Finally, from a case study of a certain grid company, the proposed PCASVM forecasting model has better performance on the grid technology transformation project investment prediction, it confirms the effectiveness of the proposed method.

1. Introduction

With the development of low-carbon economy and green ecological economy in China, in order to meet the energy demand in the future, the State Grid Corporation of China highlights the clean characteristics of electric power, and implements the development strategic direction of electric energy instead of fossil energy. The effective use and development of energy will help to improve people's quality of life. As an important carrier of power transmission channel, power grid plays an important role in ensuring its safe and stable operation. As one of the routine work to ensure the safe operation of power grid, the power grid production and technical transformation project plays a key role in ensuring the safe and stable operation of power grid, improving the quality of power supply and promoting the benign upgrading of power grid assets. With the deepening of power market reform in China, power grid enterprises will shift from the main business of power purchase and sale to the main business of power grid transmission and distribution, and their profit model will change greatly. In the new power market environment, power grid enterprises need to more lean management of their own costs, pay attention to their own transmission benefit and efficiency, and constantly strengthen the competitiveness of power grid enterprises in the power market. This requires power grid enterprises to make reasonable and accurate predictions for investment in power grid technology transformation projects, so as to manage power grid assets more accurately.
The investment environment of power grid project is not only affected by various factors in the power market, but also by national economy and policy factors. In this case, it is a complex problem to find the investment law and predict the investment of power grid[1]. According to the whole process of technical transformation project, some scholars extract the investment control indexes in the investment control process, establish an investment control technology and economic indicator system for substation projects and line projects, calculate the influence degree through examples, and combine trend extrapolation models to predict grid investment trends [2,3]. With the increase of the sample data of power grid investment, some scholars have made investment analysis on the final accounts of the completion of the line project and the substation project. It is concluded that the probability distribution of the line project investment and the substation project investment is similar to the normal distribution. According to the probability distribution, the investment range of the relevant investment can be predicted[4]. The above research has made a macroscopic analysis on the power grid investment and achieved some results, which can predict the trend and range of power grid investment. However, due to the regional characteristics of power grid technical transformation projects, the cost factors of power grid projects in different regions are often quite different, such as the labor costs in the East and West, so it is still necessary to further study the quantitative model for accurately predicting the investment amount of power grid.

The prediction of power grid investment can be summed up as the problem of time series prediction, so some scholars build the prediction model of time series to predict the demand of power grid investment[5,6]. Many scholars have summarized the characteristics of power grid investment prediction, and pointed out that there are many factors affecting the investment of power transmission and transformation projects. In the actual cost of power transmission and transformation projects, different line voltage levels, terrain, outlet method, construction scale and other factors will cause a large deviation in the project cost. Therefore, when forecasting grid investment issues in an area, it is difficult to obtain a large number of data samples. In this case, the prediction of power grid investment belongs to the small sample prediction problem, and use the gray model[7] and support vector machine model[8], which perform well under small sample conditions, to make relevant predictions for grid investment. In order to achieve a more accurate prediction, many scholars predict the power grid investment by building a combination prediction model[9]. When using the combination prediction model to predict the power grid investment, aiming at the problem that it is difficult to assign the weights of various prediction models, the particle swarm optimization method is used to weight the combination prediction model to get a better prediction accuracy[10].

All of the above researches are focused on the prediction of power grid investment, and good results have been achieved. From the above studies, it can be found that the influencing factors of power grid investment are relatively complex. In the study of regional power grid investment, due to the relatively small amount of relevant data, the regional power grid investment belongs to the prediction problem under small samples. In this paper, the investment data of a provincial power grid technological transformation is taken as the research object, and the first choice is to analyze the main factors which influence the investment, and the factors are input SVM models that perform well under small sample conditions for prediction.

2. VMD-SVM forecasting model

2.1. Principle of PCA algorithm

The method of analyzing influencing factors is referred to as factor analysis method, which is a quantitative analysis of the degree of influence between the research object and the relevant influencing factors by using statistical methods. The more commonly used factor analysis is principal component analysis (PCA). PCA method can transform the high-dimensional space into the low-dimensional space for processing by the projection method. Reducing attributes while retaining most of the original attribute information. PCA is based on the variance of the original data, and regarded as the information contribution rate. The steps of applying principal component analysis to attribute screening are as follows:
(1) Selection and Normalize the sample matrix
Select a sample set and construct a sample matrix \( X = x_{ij} \), \( 1 \leq i \leq m \), \( 1 \leq j \leq n \). Where \( x_i = (x_{i1}, x_{i2}, \cdots, x_{im})' \) indicates the number of sample records, indicates the index, and normalizes the matrix.

(2) Calculate the covariance matrix correlation coefficient matrix \( R \)
\[
R_{ij} = \frac{s_{ij}}{\sqrt{s_{ix}} \sqrt{s_{jx}}}; \quad s_{ij} = \frac{1}{n-1} \sum (x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_j)
\]

(3) Calculate the eigenvalues \( \lambda \), arrange them according to their value, and calculate the variance contribution rate and cumulative variance contribution rate of each component according to the eigenvalues. The calculation methods are as follows:
\[
\alpha_i = \frac{\lambda_i}{\sum \lambda_k} \quad (i = 1, 2, \cdots, n) \tag{2}
\]
\[
b_j = \sum_{i=1}^{j} \alpha_i \tag{3}
\]

2.2 Prediction method of support vector machine
Support vector machine inherits the advantages of neural network, and improves the global optimization ability of neural network, so that the prediction value trained with the same sample has a unique value, and performs well in small sample. At present, it is a relatively new prediction method, which is considered to be able to completely replace the intelligent prediction method of neural network in the case of small samples. The basic principles of support vector machine are as follows:

Suppose there is training sample set \( G = \{(x_i, d_i)\}, i = 1 \cdots N \), \( x_i \in \mathbb{R}^n \), \( d_i \in \mathbb{R}^l \). Support vector machine prediction model is to map sample point \( x \) to high-dimensional space \( f \), and construct the following estimation function for linear regression:
\[
y = f(x) = w\psi(x) + b
\]

Its function approximation problem is equivalent to the minimization of the following functions:
\[
R(C) = (C / N) \sum_{i=1}^{N} L_{\epsilon}(d_i, y_i) + ||w||^2 / 2 \tag{5}
\]
\[
L_{\epsilon}(d, y) = \begin{cases} 
0 & |d - y| \leq \epsilon \\
|d - y| - \epsilon & \text{otherwise}
\end{cases}
\]

Where, \( ||w||^2 / 2 \) represents the smoothness of the function, and \( L_{\epsilon}(d, y) \) is called \( \epsilon \) -sensitive loss function.

By introducing two relaxation variables \( \zeta \) and \( \zeta^* \), the above functions can be changed into the following forms:
\[
R(w, \zeta, \zeta^*) = ||w||^2 / 2 + C \sum_{i=1}^{N} (\zeta_i + \zeta_i^*)
\]
s.t.
\[
w\psi(x_i) + b_i - d_i \leq \epsilon + \zeta_i^*, i = 1, 2, \ldots, N
\]
\[
d_i - w\psi(x_i) - b_i \leq \epsilon + \zeta_i, i = 1, 2, \ldots, N
\]
\[
\zeta_i, \zeta_i^* \geq 0, i = 1, 2, \ldots, N
\]

By using Lagrange and Karush Kuhn Tucker condition, we can get that the dual type of the problem is:
The regression function of SVM can be obtained by solving the above problems:

\[ f(x, \beta, \beta') = \sum_{i=1}^{N} (\beta_i - \beta_i')K(x_i, x) + b. \]  

(8)

In the formula, \( K(x_i, x) \) is called kernel function, which needs to meet Mercer condition. Generally, the most commonly used Gaussian kernel function is selected as follows:

\[ K(x_i, x_j) = \exp(-\|x_i - x_j\|^2 / 2\sigma^2) \]  

(9)

3. Empirical analysis

In the experiment, the investment data of a regional power grid technological transformation project in 2003-2017 for 15 years is used, and the normalized original time series is shown in Figure 1.

![Figure 1](image-url)
investment, earthwork investment, and the six attributes contribute more than 95% to the cost technical and economic indicators.

Table 1  PCA calculation results by 95%

| factor                     | Cumulative contribution rate(%) |
|---------------------------|---------------------------------|
| Current scale             | 23.21                           |
| Land Investment           | 57.23                           |
| Investment in power cables| 74.89                           |
| Control cable investment  | 81.32                           |
| Construction investment   | 87.64                           |
| Earthwork investment      | 95.67                           |

According to the six factors obtained, SVM is trained and predicted by using libsvm 3.2.2 software package in MATLAB. In order to compare the prediction accuracy of the algorithm, it is compared with the prediction result obtained by SVM prediction directly using the original sequence, in which the first four prediction points of time series are taken as input variables, the fifth point as the output variable. The relative error $\text{err}_{RE}$ and average absolute error MAPE are used as the specific error analysis indexes. The predicted results and comparison are shown in Table 1.

\[
\text{err}_{RE} = \frac{y^*-y_i}{y_i}
\]

\[
\text{err}_{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{|y^*_i - y_i|}{y_i} \right) \times 100
\]

From the results in Table 2, it can be seen that the prediction accuracy can be effectively improved by using VMD to decompose the original sequence and then further predict. Basically, the relative error values of all prediction points are improved to varying degrees compared with those predicted directly by SVM. From the whole prediction results, the average absolute error (MAPE) of VMD-SVM prediction model is 1.51%, and the MAPE of direct SVM prediction is 2.02%, which also has a greater accuracy improvement in the whole level. The example shows that the investment prediction using VMD-SVM can further control the investment of power grid technical transformation projects. An example shows the validity of the model.

Table 2  final prediction results using support vector machine

| year | investment amount | PCA-SVM | $\text{err}_{RE}$ | SVM | $\text{err}_{RE}$ | MAPE  |
|------|-------------------|---------|-------------------|-----|-------------------|-------|
| 2007 | 0.225             | 0.225   | 0                 | 0.223 | -1.03%            | 0.71% |
| 2008 | 0.232             | 0.234   | 0.86%             | 0.238 | 2.54%             |       |
| 2009 | 0.246             | 0.245   | -0.41%            | 0.252 | 2.54%             |       |
| 2010 | 0.269             | 0.271   | 0.74%             | 0.276 | 2.71%             |       |
| 2011 | 0.293             | 0.292   | -0.34%            | 0.296 | 0.87%             |       |
| 2012 | 0.327             | 0.322   | -1.53%            | 0.332 | 1.43%             |       |
| 2013 | 0.381             | 0.38    | -0.26%            | 0.371 | -2.61%            |       |
| 2014 | 0.439             | 0.437   | -0.46%            | 0.431 | -1.75%            |       |
| 2015 | 0.499             | 0.506   | 1.40%             | 0.514 | 2.97%             |       |
| 2016 | 0.572             | 0.564   | -1.40%            | 0.555 | -2.92%            |       |
| 2017 | 0.654             | 0.651   | -0.46%            | 0.660 | 0.91%             |       |

MAPE 0.71% 2.02%
4. Conclusion
In this paper, the investment prediction of power grid technological transformation project is studied, aiming at the actual situation that the historical data of investment prediction of power grid technological transformation project is less, and there are more influencing factors. In this paper, the SVM model with good prediction performance under the condition of small sample is used for prediction and used PCA to get main input factors. The empirical analysis results show that PCASVM prediction model can better manage the investment cost of power grid technological transformation project. It can improve the work level of power grid investment management.

References
[1] J. van der Geer, J.A.J. Hanraads, R.A. Lupton, The art of writing a scientific article, J. Sci. Commun. 163 (2000) 51-59.
[2] Liwei Ji, Liping Yang, Gaiying Fei, Analysis and prediction of influence factors of power grid project investment, J. Management of Electric Power Enterprises in China, 6 (2016) 85-92.
[3] Xiaomin Liu, Bo Yu, Dongni Wei, Research on the whole process investment control index system and trend extrapolation forecast of power grid technical transformation project, J. Management of Electric Power Enterprises in China, 9 (2016) 49-54.
[4] Xiaomin Liu, Bo Yu, Meiling Wang, Research on the whole process investment control index system and trend extrapolation prediction of power grid technical transformation project -- Based on the principle of system classification and module splicing, J. Project Management Technology, 14 (2016) 89-94.
[5] Yaming Li, Gang Yao, Wei Han, Cost analysis and investment prediction of power grid project based on probability statistics, J. Electric Power Survey and Design, 6 (2017) 76-80.
[6] Huiru Zhao, Lu Yang, Chunjie Li, Xin Ma, Research on prediction of power grid investment demand based on co integration theory and error correction model, J. Power Grid Technology, 35 (2011) 193-198.
[7] Yan Chang, Wu Chen, Gang Zhao, Quantitative analysis technology and model of power grid investment decision, J. Technical Economy, 31 (2012) 56-62.
[8] Baichu Hu, Gang Hu, Zhaohua Hu, Song Qing, Mingwei Li, Chao Peng, Calculation model of power grid infrastructure investment based on Grey Prediction, J. Journal of University of Electronic Science and Technology, 42 (2013) 890-894.
[9] Wei Xie, Wenming Pan, Chenghua Wang, Li Chen, Yang Wang, Jing Ma, Zhiwei Li, Haihong Du, Research on medium and long term investment scale prediction model of provincial power grid based on support vector machine, J. Industrial Technology Economy, 38 (2019) 154-160.
[10] Weidong Liu, Xiaofen Lu, Xiaoyan Yu, Jinpeng Liu, Yongxu Lao, Research on the annual investment portfolio forecast of infrastructure projects of power grid company, J. East China Power, 42 (2014) 174-177.
[11] Weili Hu, Liang Cheng, Qian Ma, Nengping Zhou, Haiyang Wang, Annual investment forecast of power grid infrastructure based on Improved Particle Swarm Optimization, J. Power Grid and Clean Energy, 30 (2014) 52-56
[12] Zhi Liang, Guoqiang Sun, Hucheng Li, etc. Short term load forecasting based on deep belief network optimized by VMD and PSO, J. Power Grid Technology, 42 (2018) 598-606.