Investment Forecast of Power Network Infrastructure Project Based on BP Neural Network

Yiquan Gao*, Yang Cao¹ and Zhe Jiang¹

¹Bidding Consulting Service Department, Jiangsu Xingli Construction Group Co., Ltd., Nanjing, Jiangsu, 211100, China

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

Abstract. With the development of China's economy and society, the demand for electric energy is gradually increasing, and large-scale investment in power grid infrastructure projects is being made. Due to the unreasonable investment plan, the investment balance rate of power grid enterprises in recent years is higher, which is not conducive to the scientific management of enterprises. Therefore, it is necessary to make investment prediction for power grid infrastructure projects in order to achieve the rational management and control of power grid enterprises for infrastructure projects. Based on BP neural network, this paper establishes the investment prediction model of power grid infrastructure projects, chooses the factors that have great influence on power grid infrastructure projects to analyse, and reduces the dimension of each factor based on principal component analysis method, which improves the accuracy and efficiency of the prediction. Finally, the effectiveness and advancement of the model are verified by an example of cable line project investment prediction of a power company. The results show that the prediction accuracy of the model is 93.1%, which can provide guidance for the reasonable formulation of the annual investment plan of power grid enterprises.

1. Introduction

The construction of power grid infrastructure projects requires a large amount of capital investment, which results in a large amount of capital occupied by power grid enterprises. If the annual investment plan is unreasonable, it will inevitably face problems such as excessive balance rate and excessive financial expenditure. Reasonable prediction of investment in power grid infrastructure projects is the primary condition to ensure the scientific nature of investment plans. Power grid infrastructure projects mainly refer to power transmission and transformation projects, including power transformation and transmission lines. They have the characteristics of large geographical span, complex construction conditions and strong uncertainty. The accuracy and efficiency of investment prediction for power grid infrastructure projects by traditional forecasting methods are low. In recent years, the development of intelligent forecasting model provides a new method for investment forecasting of power grid infrastructure projects, which can significantly improve the forecasting effect.

At present, scholars in related fields have done a lot of research in the field of investment forecasting, and have obtained some results. Zhang, Yi Bin [1] studied the whole life cycle of power grid investment project, and from the pre-evaluation, mid-evaluation and post-evaluation of the project, established the technical and economic evaluation index system of power grid investment to achieve the complete evaluation of the project. Tang Lichun [2] analyzed the influencing factors of infrastructure investment. Based on the improved GA-PSO algorithm, the V-SVR model of
infrastructure investment prediction was established, and the model was trained under the condition of small samples. Mazurowski [3] discussed two training methods of neural network: classical network propagation and particle swarm optimization. The results show that BP algorithm is better than PSO algorithm for training data imbalance, especially when the sample data is small and the feature quantity is large. Bashir, Z.A. [4] Using adaptive artificial network to predict hourly load demand, in the training stage of neural network, particle swarm optimization algorithm is used to adjust the network weight, and a more reliable prediction model is created. Palmes, PP [5] considers that gradient learning artificial neural network based on evolutionary algorithm is a common method to solve local optimization and design problems of artificial neural network. In order to solve the problems of BP neural network, genetic neural network based on mutation is used instead of BP neural network. Sokolov-Mladenovic, S[6] predicts GDP growth rate based on artificial neural network and back-propagation learning algorithm. It takes import and export of goods and services, trade of goods and services as input variables, and GDP as output variables to establish prediction model. Hossain, A [7] 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. As a model with high prediction accuracy and convenient model construction, BP neural network has been applied more and more widely in the field of economic prediction.

2. Theoretical basis

Based on BP neural network method, this paper constructs the investment prediction model of power grid infrastructure projects. The principal component analysis method is used to reduce the dimension of some original indicators to obtain the corresponding principal components. The principal component analysis results are used as the input indicators of the prediction model. Finally, the predicted results and the expected output of the test set and the predicted output curve of BP neural network are obtained.

2.1. Principal component analysis

Power grid infrastructure projects are affected by natural, technological, economic, environmental and other factors, and there are many influencing factors, and there are often correlations among these factors. Using the original index as the input index of BP neural network will face such problems as high data dimension, poor fitting effect and inaccurate prediction results.

In this paper, the principal component analysis method is used to reduce the dimension of the original data, so that the principal component can represent the vast majority of the information of the original variables and is not related to each other, which is helpful to the establishment of investment prediction model and problem analysis of power grid infrastructure projects. Principal Component Analysis (PCA) combines the original P indicators with a certain correlation linearly and combines them into a new set of independent comprehensive indicators to replace the original indicators. The cumulative contribution rate of information contained in these principal components is more than 85%.

The main calculation steps of principal component analysis are as follows:

1) Standardization of raw data

Suppose the sample observation data matrix is:

\[
X = \begin{bmatrix}
    x_{11} & x_{12} & \cdots & x_{1p} \\
    x_{21} & x_{22} & \cdots & x_{2p} \\
    \vdots & \vdots & \ddots & \vdots \\
    x_{n1} & x_{n2} & \cdots & x_{np}
\end{bmatrix}
\]

Then the original data can be standardized according to the following methods:
j
ij
j
-x= =1,2 , ; 1,2, ,
var( )
x
 =
, ,

Among them,

\[ x^* = \frac{x_j - \overline{x}_j}{\text{var}(x_j)}, \quad i=1,2,\ldots,n; \quad j = 1,2,\ldots,p \]  \hspace{1cm} (2)

\[ \overline{x}_j = \frac{1}{n} \sum_{i=1}^{n} x_{ij}, \quad \text{var}(x_j) = \frac{1}{n-1} \sum_{i=1}^{n} (x_{ij} - \overline{x}_j)^2 \quad (j = 1,2,\ldots,p) \]  \hspace{1cm} (3)

(2) Computation of sample correlation coefficient matrix

Assuming that the original data is still expressed in X after standardization, the correlation coefficients of the data after standardization are as follows:

\[
R = \begin{bmatrix}
    r_{11} & r_{12} & \cdots & r_{1p} \\
    r_{21} & r_{22} & \cdots & r_{2p} \\
    \vdots & \vdots & \ddots & \vdots \\
    r_{p1} & r_{p2} & \cdots & r_{pp}
\end{bmatrix}
\]  \hspace{1cm} (4)

\[ r_{ij} = \frac{\text{cov}(x_i, x_j)}{\sqrt{\text{var}(x_i) \text{var}(x_j)}} = \frac{\frac{1}{n} \sum_{k=1}^{n} (x_{ik} - \overline{x}_i)(x_{kj} - \overline{x}_j)}{\sqrt{\frac{1}{n} \sum_{k=1}^{n} (x_{ik} - \overline{x}_i)^2 \frac{1}{n} \sum_{k=1}^{n} (x_{kj} - \overline{x}_j)^2}}, \quad n > 1. \]  \hspace{1cm} (5)

(3) Calculation of eigenvalues and corresponding eigenvectors of correlation coefficient matrix R

\[ a_i = (a_{i1}, a_{i2}, \cdots a_{ip}), i = 1,2,\ldots,p \]  \hspace{1cm} (6)

(4) Choose the important principal component and write the principal component expression.

P principal component can be obtained by principal component analysis, but because the variance of each principal component decreases, and the information contained in it decreases accordingly, in practical analysis, the first k principal component is usually selected according to the cumulative contribution rate of each principal component (that is, the variance of one principal component accounts for the proportion of the total variance), and the cumulative contribution rate of the first k principal component is more than 85%. The contribution rate of principal component can also be measured by the proportion of a certain eigenvalue to the total eigenvalue. That is:

\[ \text{Contribution rate} = \frac{\lambda_i}{\sum_{i=1}^{p} \lambda_i} \]  \hspace{1cm} (7)

2.2. BP Neural Network

Neural network can be regarded as a generalization of non-linear regression from the statistical point of view. It does not need to make a priori hypothesis through deduction in practical analysis. Moreover, neural network can grasp the relationship between data through learning. Through the study of historical data, it can find out the rules between indicators or variables from complex data. It can gradually approximate the ideal results by repeatedly comparing the observed samples. Among them, BP neural network can approximate any continuous function and has strong non-linear mapping ability. Unlike traditional nonlinear regression, traditional nonlinear regression problems are transformed into linear regression problems through variable transformation. As a supervised learning, BP algorithm adjusts the weights and deviations of the network repeatedly through back propagation...
algorithm in the training process to achieve the maximum approximation between the output vector and the expected vector.

BP neural network is a multi-level feed-forward neural network, which is mainly based on steepest descent method to continuously reverse learning error, adjust the weights and thresholds within the network, and finally achieve the minimum error sum of the whole neural network. Commonly used three-layer neural network model:

(1) input layer
We choose the factors that have a greater impact on the investment of power grid infrastructure projects as input variables. In order to avoid too many input parameters and the correlation between them is large, principal component analysis method is used to reduce the dimension of the input level indicators in this paper.

(2) hidden layer
The number of neurons in the hidden layer represents the degree of nonlinearity between input and output of the neural network, which directly affects the training speed and prediction ability of the model. Too few neurons in the hidden layer will affect the valuable features that the neural network can extract from the input layer, resulting in ineffective training results and poor fault tolerance. However, too many neurons in the hidden layer will also lead to an increase in learning time, and may not be able to estimate the minimum error, so the number of neurons in the hidden layer is usually one more than twice the input layer parameters.

(3) output layer
The output variable of this model is the static investment, and the number of neurons in the output layer is 1.

3. Model Construction and Empirical Analysis
This paper chooses all the historical data of cable line projects completed and settled by a power company from 2014 to 2018 as samples for empirical research.

Power grid infrastructure projects are mainly power transmission and transformation projects. Power transmission and transformation projects generally include power transformation and transmission lines, and transmission lines mainly include overhead lines and cable lines. In recent years, with the improvement of China's urbanization level, land resources are increasingly scarce. Cable line engineering has become the main way of line engineering construction in some urban areas because of its advantages of deep buried underground, low consumption of steel and timber. However, because the investment of cable line project is significantly higher than that of overhead line project, it is more difficult to predict the investment of cable line project. This paper takes cable line project as an example to carry out empirical analysis of investment prediction of power grid infrastructure project.

Thirty-nine samples were randomly selected from the new cable line project of the power company from 2014 to 2018. Thirty-one samples were randomly selected as training samples of the neural network, and the remaining eight samples were used as testing samples of the neural network.

In cable line 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, ratio of open tunnel to undercut tunnel and other factors have great influence on the static investment of cable line engineering. Therefore, these factors are selected as input layer parameters of the neural network. The sample data in this paper are shown in Table 1.
In order to facilitate the investment forecast of cable line engineering, the principal component analysis method is used to compare the cable length, the number of intermediate joints, the number of terminal joints, the unit price of the cable, the unit price of the intermediate joint, the unit price of the terminal joint, the total length of the construction project, the length of the open tunnel, and the darkness. The cross-dimensional processing of the length of the tunnel and the proportion of the tunnel excavation are carried out, and finally three principal components are obtained. These three principal components can express more than 85% of the information of these variables, and the principal component analysis is obtained by SPSS software. The rotation component matrix is shown in Table 2.

Table 2. Principal Component Analysis Rotation Matrix

| Principal Component | 1  | 2  | 3  |
|---------------------|----|----|----|
| Cable length $w_1$  | 0.779 | 0.446 | 0.154 |
| Number of intermediate joints $w_2$ | 0.873 | -0.004 | 0.173 |
| Number of terminal joints $w_3$ | 0.375 | -0.656 | 0.539 |
| Cable unit price $w_4$ | 0.887 | -0.286 | 0.111 |
| Intermediate joint unit price $w_5$ | 0.884 | -0.291 | 0.072 |
| Terminal joint unit price $w_6$ | 0.894 | -0.304 | 0.017 |
| Overall length of construction $w_7$ | 0.756 | 0.604 | 0.040 |
| Open tunnel length $w_8$ | -0.263 | 0.664 | 0.612 |
| Dug tunnel length $w_9$ | 0.839 | 0.449 | -0.114 |
| Excavation tunnel ratio $w_{10}$ | 0.553 | 0.047 | -0.766 |

Take principal component 1 as an example:

$$\text{Principal component 1} = 0.779 \times w_1 + 0.873 \times w_2 + 0.375 \times w_3 + \cdots + 0.839 \times w_9 + 0.553 \times w_{10}$$  \hspace{1cm} (8)

After the principal component analysis method, the input layer variables of the neural network are simplified into five variables: voltage level, cable cross section, principal component 1, principal component 2, and principal component 3.
The number of neurons in the hidden layer is set to 6, and the principal component 1, the principal component 2, and the principal component 3 are continuous variables, and the voltage grade and the cable cross section are categorical variables. A three-layer BP neural network prediction model is established as shown in the figure 1.

![Neural Network Prediction Model](image1)

**Figure 1. Neural Network Prediction Model**

Among the five variables in the input layer, the most important is the principal component 1, the importance is 45%, the least important is the voltage grade, the importance is 3%. The importance of principal component 2, principal component 3 and cable cross section are 30%, 14% and 8% respectively.

The scatter plot consisting of the actual static investment and predicted values of the cable line engineering is shown in the figure 2.

![Scatter plot of predicted results](image2)

**Figure 2. Scatter plot of predicted results**
As can be seen from the above figure, the points on the scatter plot, which takes the actual static investment of cable engineering as abscissa and the predicted value as ordinate, are closely and evenly distributed near the 45 angle bisector. It shows that the deviation between the predicted value and the actual value is small and the predicted effect is good. Expected output of test set and predicted output curve of BP neural network made by MATLAB software are shown in the figure 3.

Figure 3. Scatter plot of predicted results

The expected output curve of the test set is basically consistent with the BP neural network predictive output curve. From this aspect, the validity of BP neural network prediction is also verified. After calculation, the accuracy of BP neural network applied to cable line engineering investment prediction is 93.1%, the deviation is less than 7%, and the precision is high.

4. Conclusion and Outlook

With the continuous expansion of the investment in power grid infrastructure projects, the difficulty of management and control of infrastructure projects in power grid enterprises is gradually increasing. The research on investment forecasting of power grid infrastructure projects will continue to deepen, providing theoretical reference and practical guidance for investment decisions of power grid enterprises. In this paper, an intelligent forecasting model for investment of power grid infrastructure projects based on BP neural network is proposed. The principal component analysis method is used to screen and reduce the dimension of investment data of power grid infrastructure projects. By calculating the investment of a power company's cable line project, the main conclusions are drawn:

(1) The prediction bias of the power grid infrastructure project investment forecasting model based on BP neural network is less than 7%, and the prediction accuracy is high. (2) This paper comprehensively considers various factors affecting the investment of power grid infrastructure projects, and conducts factor selection and dimensionality reduction through principal component analysis methods, which improves the accuracy and efficiency of prediction. (3) This research can provide guidance for the construction of power transmission and transformation projects in areas with scarce land resources and difficult construction, so as to improve the operating efficiency of power grid companies and meet the demand for social electricity.
References
[1] Zhang Y B , Hu W , Zhang J X , et al. (2014). Research on the Investment Evaluation System of Grid Main Network Based on Development Strategy. Applied Mechanics and Materials, 556-562:6391-6394.
[2] Tang, L., Xu, X. (2014). Infrastructure Investment Forecasting: V-SVR Model Based on Improved GA-PSO Algorithms. Technology Economics, 33(02):56-61+70.
[3] Mazurowski, M. A., Habas, P. A., Zurada, J. A., Lo, J. Y., Baker, J. A., & Tourassi, G. D.(2007). Training neural network classifiers for medical decision making: The effects of imbalanced datasets on classification performance. Neural Networks, 21(2-3): 427-436.
[4] Bashir, Z. A., & El-Hawary, M. E. (2009). Applying Wavelets to Short-Term Load Forecasting Using PSO-Based Neural Networks. IEEE Transactions on Power Systems, 24(1): 20-27.
[5] Palmes, P. P., Hayasaka, T., & Usui, S. (2005). Mutation-based genetic neural network. IEEE Transactions on Neural Networks, 16(3): 587-600.
[6] Sokolov-Mladenovic, S., Milovancevic, M., Mladenovic, I., & Alizamir, M. (2016). Economic growth forecasting by artificial neural network with extreme learning machine based on trade, import and export parameters. Computers in Human Behavior, 65: 43-45.
[7] Hossain, A., Zaman, F., Nasser, M., & Islam, M. M. (2009). Comparison of GARCH, Neural Network and Support Vector Machine in Financial Time Series Prediction. In S. Chaudhury, S. Mitra, C. A. Murthy, P. S. Sastry, & S. K. Pal (Eds.). Pattern Recognition and Machine Intelligence, Proceedings, Vol. 5909, pp. 597+-.