ISSN:2790-1688 DOI: 10.56028/aetr.4.1.444.2023

Design of Engineering Cost Prediction Algorithm Based on Neural Network Model

Yuanduo Sun

OSC Global School, Toronto, Canada

1317968545@qq.com

Abstract. There are major problems in the current project cost management, which is bound to cause huge economic losses to the construction party, the construction party and even the financial institutions. It can be seen that how to reasonably determine the estimate of the early stage of the project to make scientific investment decisions and how to effectively control the project cost is a very important work. Improving the level of project cost prediction is the premise of reasonably determining the project cost, effectively controlling the construction cost, and realizing the lean project cost management, and is the basis of system planning and decision-making. The data in this paper shows that when the number of nodes is small, the acceleration ratio has little difference, and the increasing trend is basically the same. But after the number of nodes increases to 24, the acceleration ratio curve when the data volume is 2000 and 4000 GB has almost no fluctuation, while the acceleration ratio curve when the data volume is 6000 GB continues to show a small upward trend. Through NN (Neural Network), a model can match all existing engineering data samples, and control the error within a certain range. We think this model has a certain prediction function. However, this matching function is not simple linear regression, because any factor affecting the project cost is nonlinear.

Keywords: Neural network model; Project cost; Prediction algorithm.

1. Introduction

Since the reform and opening up, China's construction industry has developed rapidly, a large amount of capital has been used for fixed asset investment, the construction capacity of the construction industry has been continuously improved, the industrial scale has been continuously expanded, a large number of jobs have been created, the development of related industries has been driven, the urban and rural landscape of China has been improved, urbanization has been promoted, and has made great contributions to the growth of China's national economy. There are major problems in the current project cost management, which is bound to cause huge economic losses to the construction party, the construction party and even the financial institutions[1]. It can be seen that how to reasonably determine the estimate of the early stage of the project to make scientific investment decisions and how to effectively control the project cost is a very important work. Cost prediction is one of the important contents of project control and management. Its accuracy directly affects the scientific nature and investment economic effect of project investment decision, capital construction scale determination, project design scheme formulation, etc. Improving the level of project cost prediction is the premise of reasonably determining the project cost, effectively controlling the construction cost, and realizing the lean project cost management, and is the basis of system planning and decision-making[2-3]. Therefore, how to correctly reflect this nonlinear relationship and predict it easily, quickly and accurately is the key to build a highway engineering cost prediction model. The single nature of the project cost and the complexity affected by multiple factors make the cost of the power transmission and transformation project present a variable characteristic, so a set of forecasting tools that can effectively improve the forecasting accuracy is urgently needed[4].

The traditional project cost prediction methods include quota calculation method and empirical value estimation method. The emerging prediction methods analyze the influencing factors of project cost through modeling, and then obtain the prediction results, such as grey prediction method, artificial NN method, support vector machine method, etc. NN is not only a highly

ISSN:2790-1688 DOI: 10.56028/aetr.4.1.444.2023
nonlinear dynamic system, but also an adaptive organizational system. It can mainly be used to describe the intelligent behavior of cognition, decision-making and control. Its central problem is intelligent cognition and simulation. The computer network needs to be based on a large number of data and its logical expression ability is not strong. The focus of fuzzy mathematics prediction is to determine the membership degree, proximity degree and adjustment coefficient of engineering features. The regression analysis method does not take into account the influence of uncertainty factors, the time series method has a high demand for the reliability of sample data, the grey theory rule cannot be used for the prediction of multiple influencing factors, and the Bayes prediction is too subjective[5]. A model built by NN can match all existing engineering data samples and control the error within a certain range, so we think this model has certain prediction function. However, this matching function is not simple linear regression, because any factor affecting the project cost is nonlinear[6].

2. Project cost forecast

2.1 Traditional forecasting method

In different stages of project construction, the construction cost has different forms. In the early stage of the project, the project proposal and feasibility study stage are investment estimation, the design stage is design budget, the construction drawing design stage is budget, the bidding stage is contract price, the completion acceptance stage is settlement price, and the final settlement stage is final settlement price[7]. In this chapter, the traditional forecasting methods of project cost will be summarized, which are divided into qualitative forecasting methods and quantitative forecasting methods[8]. Qualitative prediction methods mainly include: expert meeting method, Delphi method and subjective probability method. The characteristic of qualitative prediction method is to collect experts' opinions and expectations for the project, and the process is cumbersome and the prediction results will inevitably introduce experts' subjective factors.

There is an obvious nonlinear relationship between project cost and its own characteristic factors, which are important factors that can not only express the characteristics of highway engineering, but also reflect the main cost composition of highway engineering. In the traditional cost estimation, the estimation index is often used, and the compilation of the estimation index is representative through screening[9]. The quantitative prediction method is not influenced by subjective factors, and is widely used in the prediction of practical projects. However, for complex construction projects, the original data are complicated and diverse, and the data regularity is poor, so this method has poor ability to deal with the interaction effect and nonlinear relationship between data. Technically, it is in line with the development direction, and the settlement data related to its engineering quantity are comprehensively analyzed and compiled with sufficient quantity and possible reuse.

2.2 Modern intelligent prediction method

NN is a back-propagation learning algorithm for feedforward multilayer. During the calculation process, the connection weights between the artificial neurons that compose the feedforward multilayer network will be continuously modified, and finally the information input into the feedforward multilayer network can be converted into the desired output content. Its purpose is to find a hyperplane to segment the sample. The principle of segmentation is to maximize the interval, and finally transform it into a convex quadratic programming problem to solve. The weight value and membership relationship value are usually determined based on experience or statistics, and their determination is subjective to some extent, but this does not mean that the determination of the weight value and membership function is arbitrary and untested[10]. We can use the above formula to test the accuracy of each known typical project according to the subjectively determined weight value and membership function value, so as to know whether the determination of the weight value and membership function value is accurate and reasonable.

ISSN:2790-1688 DOI: 10.56028/aetr.4.1.444.2023

Due to the different degree of influence of each feature element on the project cost, different weights should be assigned according to the impact of the feature element on the cost. The maximum weight value is 1, and the weight value of other feature elements is the ratio of the maximum weight value. The basic idea of solving the cost prediction model of highway engineering based on NN is to use the extremely strong function discovery ability of NN algorithm to obtain the best fitting function between the cost of many typical highway engineering accumulated in the past and its engineering characteristic factors, and predict the cost of highway engineering based on this function. The process of solving the project cost prediction model is shown in Figure 1.

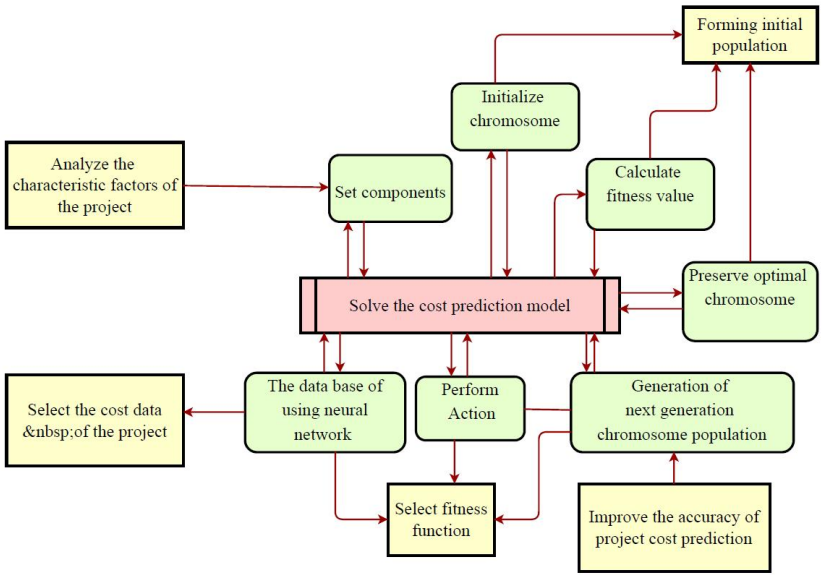

Figure 1 Flow Chart of Cost Prediction Model Solution

The NN prediction model not only retains the accumulation generation method in the grey prediction theory, which can weaken the random components of the original data, but also uses the ability of NN to approximate arbitrary functions to achieve accurate fitting and prediction of data series with arbitrary change laws. The purpose of normalization is to make the sample data more conducive to NN training and reduce the difference between them. Since no proper weight is given to each feature element, it is considered that each feature element has the same contribution to the typical project cost.

However, in fact, the basic treatment of feature elements generally accounts for 26%~37% of the total project cost, which is larger than the proportion of other feature elements in the project cost. The structural characteristics of feature elements also have a very large impact on the cost, while the impact of other elements is smaller. However, due to the neglect of this point, the role of some secondary factors has been amplified, resulting in unreasonable data. Mapping data from low-dimensional space to high-dimensional space effectively avoids "dimension disaster" and has been widely used in many fields such as power forecasting, text classification, etc. According to the principle of project cost composition, as long as the construction and installation costs are known, the cost items of other projects and facilities along the line can be calculated according to the cost standard of other costs, and the whole project cost can be obtained.

3. Application of neural network in the field of engineering cost

3.1 Solve the cost prediction model

The reliability of rapid engineering evaluation method mainly depends on the number of projects in the established engineering database, the typicality of the representative, the selection of characteristic factors, and the rationality of membership and weight. These can be gradually improved in engineering practice, so as to achieve satisfactory rapid evaluation results. According

ISSN:2790-1688 DOI: 10.56028/aetr.4.1.444.2023
to the engineering practice, with the increase of the mileage of the main line, the cost of measures and overheads per unit kilometer decreased, that is, the project cost showed a moderate downward trend with the increase of the mileage of the main line. The number of engineering cost samples is often very large, and some samples show the same or similar construction environment and cost characteristics. We can cluster these samples and get the average attributes of these samples. The traditional method is to make a general estimation by using various forecasting indicators and price limits.

Any kind of estimation based on one or two indicators is difficult to accurately describe the internal technical and economic characteristics of a project, but when calculating the fuzzy relationship coefficient of each project, all characteristic elements are still used for calculation, and its purpose is to comprehensively consider the factors that constitute the project cost, so that the fuzzy relationship coefficient is accurate and undistorted. It can be seen from the calculation results that the error is very small, which shows that this choice is correct. However, many calculation results in Example 1 have large errors, which is because the weight is not considered, so that the role of the main factors is weakened, and the role of many important factors is amplified. Therefore, we turn our research to the analysis of all past data in this paper. As can be seen from Table 1, the relative error between the predicted value and the actual value of the NN prediction model during the training period and the verification period is very small. Table 1 shows that although there is a certain average absolute error between the predicted value and the actual value, their average relative error is very small and the prediction accuracy is high, which further verifies that the prediction accuracy of the model established in this paper is high.

Table 1 Calculation of evaluation indicators

Project	Estimate	Actual value	Error value
1	3790.824	4130.748	339.924
2	3548.7403	3 998.389	449.6487
3	4857.3175	5215.8507	358.5332
4	4284.1308	4976.3198	692.189

You can learn how to classify data by acquiring the feature vector NN. Therefore, it can be said that nodes as hidden layers can be regarded as abstract features. The value of this abstract feature is expressed as follows:

$$L_j^k = \left(w_{ij}^k x_i + b_k\right) \tag{1}$$

We define the output of a neuron in the hidden layer as θ_j^k . Therefore, it also needs nonlinear activation function to divide it. When we add all the different node parameters of all layers, we can get the following matrix functions:

$$F^n = \begin{cases} F^n = M^n \sigma F^{n-1} + B^1 \\ F_1 = M_1 Y_1 + B^2 \end{cases} \tag{2}$$

This formula is the calculation formula of NN forward propagation. In the formula, $B^1 = \left(b_1, b_2, ...b_j\right)$ and $B^2 = \left(b_1, b_2, ...b_h\right)$ represent the offsets of different layers.

So as to guide the next training in the right direction. This paper uses the mean square error function:

$$L = \frac{1}{2N}\left(F(\omega, b) - f(x)\right) \tag{3}$$

Where $F(x)$ represents the calculated value of forward propagation through NN and $f(x)$ represents the actual value of the sample.

From the point of view of rapid valuation method research, the combination of physical quantity quota and price quantity quota established in this paper is difficult to adapt to the changes of labor, materials, machinery and other prices, and the physical quantity quota is the quantitative standard to determine the consumption of sub-projects The model conforms to the engineering practice, and the

ISSN:2790-1688 DOI: 10.56028/aetr.4.1.444.2023
proposed rapid evaluation method of highway engineering is a fast, easy-to-use and accurate method with practical value. Although the variable structure algorithm can be used to discuss any number of layers and compare any number of neurons in each layer, it needs to be limited in consideration of the calculation speed. However, there are some problems in these methods, because for transmission line projects, each project has its own characteristics and will be composed of several key indicators that affect the cost. The eigenvectors formed by these indexes are determined by the specific technology of transmission lines.

3.2 An example of neural network model for cost prediction

Taking the data of a building project in a building database as sample data, the performance of the above NN-based project life cycle cost analysis and prediction algorithm is verified. Because the current quota is not reasonable enough, and the bidding competition for construction is fierce, and the winning bid price is low, the survival and development space of the construction unit is limited, which is not conducive to the healthy development of the construction market in the long run. To control the project cost well, a reasonable quota is the basis. The whole life cycle cost of the project is mainly to reduce the construction cycle cost, ignoring the contents of other stages of the construction cycle. In this chapter, NN model and linear decline model will be used for prediction and analysis, and the results are shown in Figure 2.

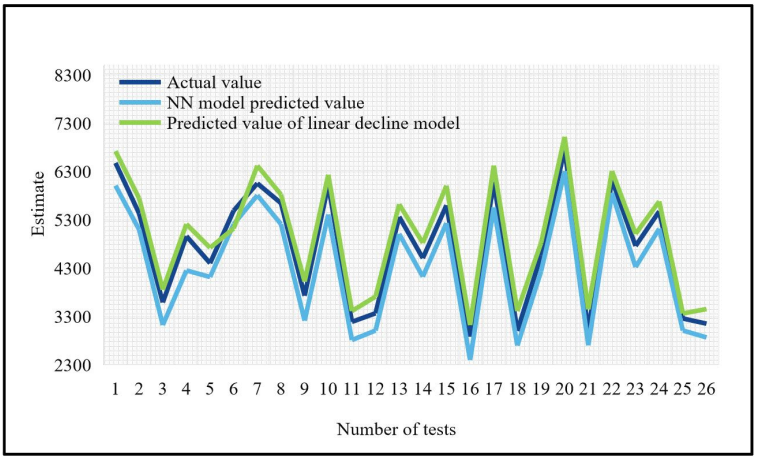

Figure 2 Comparison Chart of Prediction Results of Cost Prediction Model

Figure 2 visually shows that the predicted value of NN cost prediction model is highly consistent with the actual value; The NN model has over-fitting phenomenon, that is, the degree of consistency in training period is high, and the degree of consistency in validation period is low; The consistency of linear regression model is low.

As can be seen from Figure 3, when the number of nodes is small, the acceleration ratio has little difference, and the increasing trend is basically the same. However, after the number of nodes increases to 24, the acceleration ratio curve at the data volume of 2000 and 4000 GB has almost no fluctuation, while the acceleration ratio curve at the data volume of 6000 GB continues to show a slight upward trend. It shows that the above algorithm has a better acceleration ratio when processing a large number of project life-cycle cost data, and has more advantages and efficiency in reducing the project life-cycle cost prediction time. The prediction accuracy of the above algorithm under different data noise levels is shown in Figure 4.

ISSN:2790-1688 DOI: 10.56028/aetr.4.1.444.2023

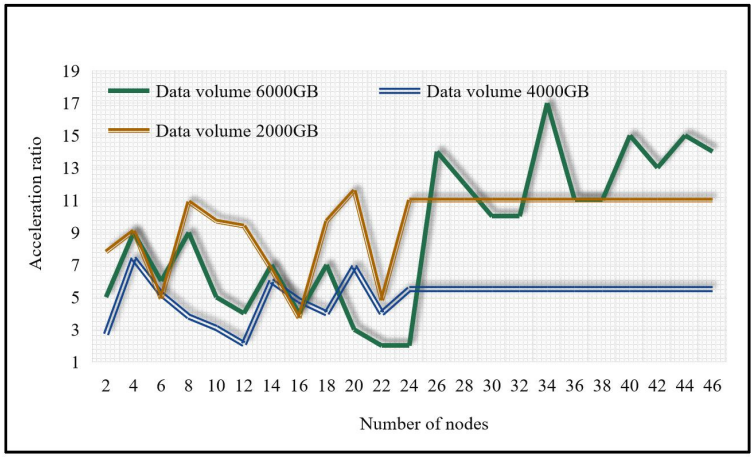

Figure 3 Comparison chart of acceleration ratio under different data amounts

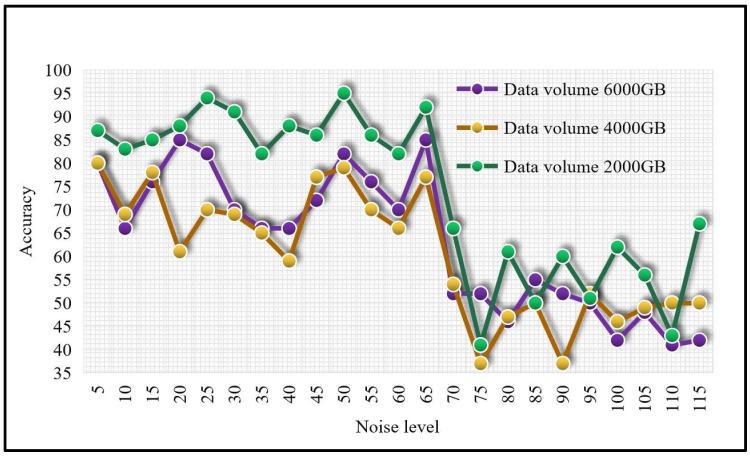

Figure 4 Comparison chart of prediction accuracy under different data noise levels.

As can be seen from Figure 4, when the data volume is 2 000 GB and 6 000 GB, with the increase of data noise, the prediction accuracy rate drops from 94% and 85% to 40% and 37% respectively, and the prediction accuracy rate shows an obvious downward trend. When the data volume is 4000, with the increase of data noise, the prediction accuracy rate drops from 86% to 50%, respectively, which is relatively small. It shows that with the increasing noise of data, the prediction accuracy is higher when the above algorithm is used to deal with large data.

4. Conclusions

The project cost has become a major uncertainty factor in the project construction due to large quantities, complex engineering structure and many influencing factors. Because the scope of use of some quotas is not clearly defined, the definitions of quota items are repeated, even at the same level, there are great differences, and the calculation methods are also different, which makes the projects of the same quality and different prices, and destroys the scientific and authoritative nature of the quota. Therefore, it restricts the bidding and tendering of design and construction across regions and industries, and also weakens the country's macro-control ability to the quota level. This paper maps data from low-dimensional space to high-dimensional space, effectively avoiding "dimension disaster", and has been widely used in many fields such as power forecasting, text classification, etc. According to the principle of project cost composition, as long as the construction and installation costs are known, the cost items of other projects and facilities along the line can be calculated according to the cost standard of other costs, and the whole project cost can be obtained. The data shows that when the number of nodes is small, the acceleration ratio has little difference, and the increasing trend is basically the same. However, after the number of nodes

ISSN:2790-1688 DOI: 10.56028/aetr.4.1.444.2023
increases to 24, the acceleration ratio curve when the data volume is 2000 and 4000 GB has almost no fluctuation, while the acceleration ratio curve when the data volume is 6000 GB continues to show a small upward trend. The model prediction in this paper shows that the prediction accuracy of the two algorithms based on the main characteristic factors is significantly higher than that of the model considering all special factors; And the NN model based on the main characteristic factors can obtain more accurate prediction results.

References

[1] Cai Z Z. Prediction Algorithm for Network Security Situation based on BP Neural Network Optimized by SA-SOA. International Journal of Performability Engineering, vol. 16, no. 8, pp. 50-72, 2020.

[2] Aretoulis G N. Neural network models for actual cost prediction in Greek public highway projects. International Journal of Project Organisation and Management, vol. 11, no. 1, pp. 41-66, 2019.

[3] Ye D. An Algorithm for Construction Project Cost Forecast Based on Particle Swarm Optimization-Guided BP Neural Network. Scientific programming, vol. 45, no. 9, pp. 2021-2045, 2021.

[4] Jun K, Yeon L J. Development of a cost analysis-based defect-prediction system with a type error-weighted deep neural network algorithm. Journal of Computational Design and Engineering, vol. 47, no. 2, pp. 2-18, 2022.

[5] Zhao L, Zhang W, Wang W. Construction Cost Prediction Based on Genetic Algorithm and BIM. International Journal of Pattern Recognition and Artificial Intelligence, vol. 68, no. 18, pp. 30-59, 2019.

[6] Pan Y, Hongtao Y U, Liu S. Neural network-based link prediction algorithm. Chinese Journal of Network and Information Security, vol. 67, no. 25, pp. 46-67, 2018.

[7] Song K, Lv S K, Hu D, et al. Software Defect Prediction Based on Elman Neural Network and Cuckoo Search Algorithm. Mathematical Problems in Engineering, vol. 20, no. 17, pp. 33-52, 2021.

[8] Sun A, Xiang C, Hanxi W U. Study on BP neural network based prediction model for power transmission line project cost. Modern Electronics Technique, vol. 54, no. 11, pp. 28-57, 2018.

[9] Zhong Y, Li H, Chen L. Construction project risk prediction model based on EW-FAHP and one dimensional convolution neural network. PLOS ONE, vol. 65, no. 14, pp. 16-28, 2021.

[10] Phatai G, Chiewchanwattana S, Sunat K. Cultural Algorithm Initializes Weights of Neural Network Model for Annual Electricity Consumption Prediction. International Journal of Advanced Computer Science and Applications, vol. 25, no. 6, pp. 11-19, 2020.