

Research Article

Prediction Technology of Power Transmission and Transformation Project Cost Based on the Decomposition-Integration

Yan Lu,¹ Dongxiao Niu,¹ Jinpeng Qiu,¹ and Weidong Liu^{2,3}

¹North China Electric Power University, No. 2, Beinong Road, Huilongguan, Changping District, Beijing 102206, China

²Zhejiang Electric Power Corporation, Economic Institute of Technology, No. 1, Nanfu Road, Shangcheng District, Zhejiang 310008, China

³Xi'an Jiaotong University, No. 74, West Road, Yanta District, Xi'an 710049, China

Correspondence should be addressed to Yan Lu; hdluyan@163.com

Received 29 October 2014; Revised 10 March 2015; Accepted 17 March 2015

Academic Editor: Alex Elías-Zúñiga

Copyright © 2015 Yan Lu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Through the analysis of power transmission and transformation project cost, total cost can be decomposed into construction cost, equipment purchase cost, installation cost, and other costs. This paper proposes a decomposition-integration cost prediction model taking a substation project as an example by fully considering the cost characteristics. In decomposition module, the total cost is decomposed into four expenses. In prediction module, different forecasting models are selected to forecast different expense. In integrated module, choose different integration methods to get the predicting results of total cost. The empirical results show that decomposition-integration prediction algorithm has good effect which can effectively predict the cost of power transmission and transformation project and has practical application and popularization value.

1. Introduction

As the infrastructure construction of national economy, good construction and stable operation of power transmission and transformation project not only directly affect the development of the electric power industry, but also have a huge ripple effect on other related industries. With the high-speed development of national economy and rapid growth in electricity demand, electric power construction has accelerated [1]. However, the electric power construction investment is limited. If the project cost control is undeserved, rising cost level will affect the power grid project economic benefits and the level of construction. So how can we determine and control the power transmission and transformation project cost is the main problem to be solved in the power engineering cost management. At present, there are a lot of related researches on electric power engineering cost prediction. The common cost prediction models are multiple regression, nonlinear regression, linear network prediction,

BP network prediction, fuzzy neural network, and so forth. The literature [2] used the gray GM (1,1) method to establish two calculation models, which were used to the estimation budget establishment of the power engineering. The literature [3] established the Markov chain model to accurately forecast the power transmission project cost index. The literature [4, 5] elaborated the key influence factors of the power transmission and transformation project cost, using multiple linear regression method to establish comprehensive cost prediction model. The literature [6, 7] proposed a kind of transmission line engineering cost based on the BP neural network prediction method. The literature [8] estimated the project cost by calculating the similarity degree between the completed projects and projects based on fuzzy mathematics theory. The literature [9, 10] constructed the engineering cost estimation model of support vector machine. Any single forecast model has its own advantages and disadvantages. In order to improve the accuracy of cost forecast, some mixed models appeared in the field of power engineering

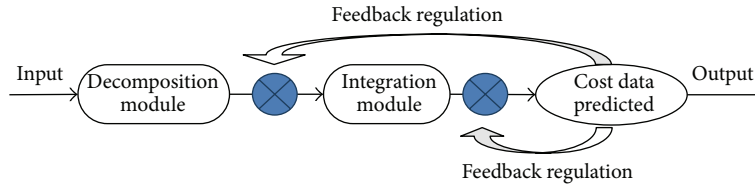


FIGURE 1: Technique flowchart of power transmission and transformation project cost prediction technology based on decomposition-integrated.

cost prediction. The literature [11] proposed a cost estimation method based on gray relational analysis and neural network. The literature [12] established project cost prediction model based on chaos SVM and ARIMA models.

As a whole, domestic and foreign engineering cost prediction methods can be summarized as the following kinds: first one is the method of quota with disadvantages of too long file budgeting time and complex budget work. The second is the engineering analogy method with poor accuracy. The third is the fuzzy mathematics method which estimates the project cost by the similar engineering fuzzy number conversion, while the determination of feature membership and coefficient adjustment is too difficult. The fourth is regression analysis; the disadvantage is that it cannot take enough factors, especially the uncertainty factors, into consideration [13]. The fifth is artificial neural network algorithm which is intelligent and has a strong ability to find the rules, but the training process may face problems of no convergence, being easy to fall into local minimum point or unable to get global optimization, and so forth. The sixth is combination forecast method; at present, the research on the applying of this model in electric power engineering cost is rare.

Based on the above literature analysis, combined with the idea of “decomposition-integration” and the perspective of combination cost forecasting, this paper builds a decomposition-integration power project cost forecasting model to overcome the shortcomings of the various cost prediction models. As for the decomposition module, the power transmission and transformation project cost is decomposed into itemized expenses which can be simply described. As for the prediction module, the main influence factors are identified and the suitable forecasting model is selected according to the characteristics of the itemized expenses. As for the integration module, choose effective integration algorithm to restore prediction system. At last, this paper chooses the cost data of Zhejiang province power transmission and transformation project for the example simulation. The results show that it can predict the power project cost effectively using the decomposition-integration method.

2. Decomposition-Integration Prediction Technology Process

Decomposition-integration of power transmission and transformation project cost prediction technology can be divided into three modules: decomposition module, prediction module, and integration module. And the prediction technology is discussed based on the example of a substation project.

Decomposition Module. According to “Grid Construction Budgeting and Calculating Standard,” the static investment of new substation engineering includes primary and secondary production engineering costs, individual project costs associated with site, compiled years spreads, and other expenses. The main production engineering costs are divided into construction project cost, installation cost, and equipment purchase expense according to the cost types. Therefore, in terms of cost types, substation project costs can be decomposed into construction project cost, installation cost, equipment purchase cost, and other costs [14].

Prediction Module. Cost prediction method is numerous, including multiple linear regression, artificial neural network, and support vector machine. On the basis of cost decomposition, identify, analyze, and filter its influencing factors; then, select the suitable method for each of itemized expenses from the prediction method library to itemized forecast.

Integration Modules. Applying two methods to integrate costs, the first is simple summation, namely, to add the construction costs, installation costs, equipment purchase costs, and other expenses to get the final prediction; the other is the weighted summation, namely, to work out the weight of costs through analyzing the proportion of the four cost items and then calculate the weighted summation.

Feedback Regulation. Compare the prediction results to the actual value with mean absolute error (MAE) and mean square error (MSE), and constantly optimize the analysis and prediction methods according to the feedback adjustment of fitting effect. The training and optimization can be stopped only when the forecast cost data achieve the prediction accuracy. Decomposition-integrated prediction process is shown in Figure 1.

3. Prediction of Subentry Costs of the Substation Project

3.1. Prediction of Construction Cost Based on the MIV-BP Neural Network

3.1.1. Factors Identification. Substation engineering project construction cost can be divided into two main types: primary and secondary production engineering costs, individual project costs associated with site. In the balance sheet, it can be divided into main control building cost, accident oil pool cost, cable channel cost, water supply system cost, earth

leveling cost, walls and doors cost, the air conditioning and ventilation cost, roads cost, drainage cost, fire system cost, special structure cost, afforestation cost, foundation treatment cost, pile work cost, and so forth. Generally, through the analysis of previous substation engineering cost, the proportion of main control building, foundation treatment and pile foundation engineering is the highest. Then comes to the cost of earth leveling, walls and doors, special structure, fire systems, dedicates cable channel, air conditioner and ventilation. The proportion of other individual project costs is the smallest.

Mining the main factors of the major cost of construction cost, more than 20 factors have been identified: there are location, endurance, pollution level, altitude, landform, the number of main transformer sets in current period, substation capacity, total land requisition area of the site, total land requisition area within the wall, total construction area, main control building area, cable length, excavation of field leveling, fill volume, foundation treatment method, path length of entering the substation, cable channel volume, volumes of well, ditch, and pond, weight of steel, weights of steel frame and other material, weight of cement, and so forth [15].

3.1.2. Prediction Model

(1) *Factors Screening Based on MIV.* The original input set has more than 20 factors, which are too many; besides, the information of some factors is overlap. In order to optimize the original variable set, the average impact value (MIV), which was regarded as one of the best indicators for evaluation of variable correlation in the neural network, was used to select the key factors in this paper. MIV is a value that represents the influence on weight of input index to output index. The attribute that MIV is a positive or a negative number embodies the influence direction to the outputs. And the absolute value illustrates the relative influence degree.

The calculation steps are as follows. Generate two samples by increasing or decreasing each index by 10% (the percentage is customizable) on the basis of original training samples. Then, calculate the training results of these two samples separately; the difference between the results can be regarded as the specific index changing's impact on the output value. Calculate the average MIV value of the independent variable for several times in order to reduce the training error. Then, repeat the process for each indicator to calculate the MIV values of all input nodes. At last, sort the MIV value, and key factors can be screened in accordance with the requirements [16].

(2) *Data Preprocessing.* In the process of practical engineering projects, the value of different indicators may vary widely on the order of magnitude as the implications of the original index are not the same. So the data must be normalized. The normalization formula is

$$y = \frac{(1 - y_{\min}) * (x - x_{\min})}{x_{\max} - x_{\min}} - 1. \quad (1)$$

(3) *The BP Neural Network Prediction.* Build a three-layer BP neural network, setting the screened variables as network input and the project construction cost as network output. Confirm the nodes number of hidden layer according to the number of input and the output node. Calculate the predicted output of the neural network by practicing the network training with sample data. Then, compute the network prediction error according to the predicted and desired output. Constantly update weight value and threshold value until the results meet the error precision. At last, save the final weight value and threshold value as the initial weight value and threshold value to predict construction cost of the project [17, 18].

3.2. Prediction of Equipment Purchase Cost Based on Multiple Linear Regression

3.2.1. *Factors Identification.* Equipment cost mainly refers to the purchase of various kinds of electrical equipment in substation project. According to the statistical data, substation engineering costs account for the largest proportion of equipment purchase expense, nearly 50 percent. So the entire project cost fluctuates with the equipment price. The main reason that leads to the variance of equipment purchase expense is the localization degree of the equipment. In recent years, with the development of the domestic technology, constantly improved localization rate of the equipment has made the major equipment purchase expense greatly reduce. But the equipment purchase is still the key link of cost management. In this paper, combined with the statistical data, the equipment purchase expense is analyzed based on the ABC classification method.

Combined with the secondary indexes raised by the experts, major cost in Figure 2 can be decomposed into the main equipment under each secondary index. The factor set of equipment purchase expense generated according to the quantities and price of main equipment is shown in Table 1.

3.2.2. Prediction Model

(1) *Multiple Linear Regression.* Set the dependent variable Y as equipment purchase expense and $x_1, x_2, x_3, \dots, x_n$ as independent variables, representing the main influence factors of equipment purchase expense. The partial correlation coefficients of Y corresponding to $x_1, x_2, x_3, \dots, x_n$ are $\beta_1, \beta_2, \beta_3, \dots, \beta_n$, representing the influence of x_i 's unit change to the size of Y ; ε is random error term [19, 20].

Multiple linear regression equation is

$$Y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_n x_n + \varepsilon. \quad (2)$$

The matrix type is

$$Y = X\beta + \varepsilon. \quad (3)$$

Suppose there are n sets of data $(\hat{y}_i, x_{i1}, x_{i2}, \dots, x_{ip})$ $i = 1, 2, \dots, n$.

The sum of squared residuals is

$$RSS = \sum \hat{\varepsilon}_i^2. \quad (4)$$

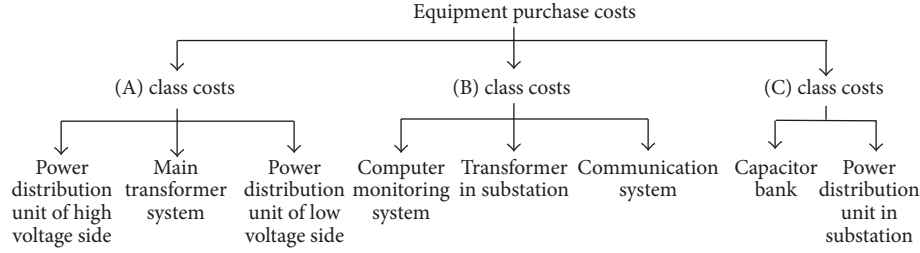


FIGURE 2: Equipment purchase cost analysis based on ABC.

TABLE 1: The main factors that influence the purchase expense recognition equipment.

The first level indicators	The second level indicators	Main influence equipment
Equipment purchase cost	The main transformer system	The main transformer
	Power distribution equipment	High voltage power distribution equipment Low voltage power distribution equipment
	Var Compensator	Capacitor bank
	Control and DC system	The computer monitoring system
	Station power system	Station transformer Station power distribution equipment
	Communications and remote systems	Communications systems

Calculating with the matrix method,

$$\hat{\beta} = (X^T X)^{-1} X^T Y. \quad (5)$$

(2) *Model Test*. As the fitting effect of the obtained multi-variate linear regression model may be not good enough, the equation and parameter must be examined to determine whether these equations are available.

(a) *Test of Goodness Fit*. In multiple linear regression, the multiple determination coefficient R^2 means the influence degree of the explanatory variables within the model to the whole dependent variable. The value of R^2 distributes between 0 and 1. The closer R^2 to 1, the better the fitting degree of regression equations, and the closer of the relationship between the independent variable and dependent variables.

(b) *The Significance Test of Regression Equation*. Significance test of regression equation means to use the F test method to test whether parameter β_i is not equal to 0 significantly. That is, if $F > F_\alpha$, the linear relationship is significant; if $F < F_\alpha$, the linear relationship is not significant. Critical value $F_\alpha(k, n-k-1)$ can be calculated by the given significant level α ; the value of F can be calculated by sample statistics [21].

3.3. Prediction of Installation Cost Based on Grid Search-SVM Method

3.3.1. *Factors Identification*. In the statistical sample, installation cost accounts for about 7% of the total investment, compared with other costs, and the volatility is relatively stable. The influence factors of installation cost can be excavated from the natural factors and technical conditions.

Table 2 shows the two-dimensional division on the production engineering of technical factors.

According to the features of project installation costs, natural factors that affect the construction cost level include endurance and altitude. In summary, the installation construction cost prediction model should take all the technical conditions and natural factors as a factor library into consideration.

3.3.2. Prediction Model

(1) *Data Preprocessing*. Just like the neural network prediction, the original data have to be normalized to make different dimensional data comparable before SVM training.

(2) *The Support Vector Machine (SVM)*. Support vector machine (SVM) is a kind of novel machine learning methods based on the theory of structural risk minimization principle. The basic idea is as follows: if samples are linearly separable, the optimal classification hyperplane of the two categories samples can be found in the original space; if samples are linearly inseparable, bringing in a slack variable, by using the nonlinear mapping method, the input space samples of lower dimensional space can be mapped to higher dimensional space to make it linear. Through these processes, the nonlinear sample can be analyzed in higher dimensional space by using linear algorithm [22].

Most cases are nonlinear, so the lower dimensional space must be mapped to higher dimensional space with nonlinear mapping function $x \rightarrow \varphi(x)$ to make the linear SVR be used in higher dimensional space. Kernel function $k(x_i, x)$ is actually a kind of mapping. To different kernel functions, the corresponding algorithms are not the same; thus the

TABLE 2: Identification of technical conditions of installation cost factors.

The first level indicators	The second level indicators	The main technical factors
The main production engineering	Main transformer system	Current main transformer stations A single main transformer capacity
	Power distribution equipment	(High, medium, and low side) distribution equipment type (High, medium, and low side) number of units breaker
	Var Compensator	Reactor Capacitor Series compensation device
	Control and DC system	Computer monitoring and control system
	Station power system	Station transformer capacity Load voltage form Insulation mode
	Cable	Control cable 1 kv and below power cable
	Ground material	Material type Weight
	Communications and remote systems	Optical transmission equipment Cable type

sample will be mapped to different higher dimensional spaces [23]. In higher dimensional space, applying linear regression method for analysis makes the nonlinear regression problems equal to solve the maximum value of function $w(\alpha, \alpha^*)$ under constraints. In this function,

$$w = \sum_{i=1}^t (\alpha_i - \alpha_i^*) \varphi(x_i) \quad (6)$$

which is finally obtained as

$$f(x) = \sum (\alpha_i - \alpha_i^*) k(x_i, x) + b. \quad (7)$$

(3) *The Grid Search Method.* The grid search method is mainly used for the SVM parameters optimization of c and g . The SVM parameter selection is actually a process of optimization search. Grid search means to split and exhaustively compute the grids within a fixed range. For selected parameters groups c and g , the highest classification accuracy in the sense of cross validation can be found, and the optimal parameters can also be generated through comparing plenty of experiments. This method is simple and intuitive. But the amount of calculation increases with the enlargement of the parameter search range, and the calculation time will be extended [24].

3.4. Prediction of Other Cost Based on PSO-SVM

3.4.1. *Factors Identification.* In other costs, the most uncontrollable factors are the site acquisition costs and clean-up costs at early stage of the project. As the land acquisition costs, construction costs, relocation compensation costs, and other costs have big difference due to different land prices in different regions, the key to cost control is to analyze the station location. Other costs mentioned above include

the following six subcosts: the construction site acquisition cost and clearing costs, management costs, the technical service costs, production preparation cost, and other basic reserve funds cost.

The natural environment factors and technical factors should be mainly considered when doing other costs predicting analysis of substation engineering. Among them, the natural environment includes location, endurance, pollution level, and elevation. Technical factors can be summarized as the number of main transformer in current period, a single main transformer capacity, total land requisition area, land requisition area within the walls, total construction area, main control building area, cable length, the excavation of earth leveling, fill volume, cable channel volume, volume of well, ditch, and pool, weight of steel, weights of steel frame and other material, weight of cement, construction site requisition and compensation cost, the preparatory work of the project cost, pile foundation inspection cost, large transport measures cost, design cost, inspection cost, supervision cost, and production preparation cost.

3.4.2. *Prediction Model.* There are too many factors affecting other costs. The relationships between the factors are complicated. Besides, the effect is not obvious on the trend of cost expenses. So we use principal component analysis (PCA) to reduce the dimension of factors before calculating. And then predict the cost by using the support vector machine with strong generalization ability. Besides, we use the particle swarm algorithm (PSO) to optimize the parameter.

(1) *Data Preprocessing.* After the normalization preprocessing, use PCA to reduce the dimension of original data. PCA's goal is to reduce data dimension as much as possible under the premise of guaranteeing the integrity of data information. The analysis processes are as follows: firstly

calculate the covariance matrix of factor set square, and work out the eigenvector and eigenvalue of covariance matrix; then compute the variance (information) contribution rate of the principal component to select principal components when the cumulative contribution rate meets the condition [25, 26].

(2) *Parameters Optimization Based on PSO Algorithm.* PSO is a kind of swarm intelligence optimization algorithm in the field of computational intelligence which is applied to solve optimization problems. The basic thought is to initialize a random group of particles and then update the iteration to find the optimal parameters. Specific steps are as follows: (a) divide the historical samples into training set and test set; (b) set the values of fold number of the cross validation, evolution algebra, population quantity, and the range of parameters c and g ; (c) calculate the fitness value according to randomly generated parameters c and g ; (d) get the position of the best fitness; (e) reset the search speed and cycle to search for the other optimization parameters c and g ; (f) repeat calculation process and cease the calculation after reaching the maximum range [27–29].

4. Empirical Research

In order to verify the effectiveness of the decomposition-integrated prediction technology in power transmission and transformation project cost prediction, this paper takes Zhejiang province transmission and substation engineering cost as the research object. The historical data table shows that there are 57 new 110 kV substation projects. Randomly choose 50 samples as training samples and the remaining 7 samples as the test validation samples. And set the unit capacity cost as the predicted output. Based on SPSS and MATLAB, the steps of decomposition-integration are shown in Figure 3.

4.1. Prediction of Construction Cost

(1) *The Analysis of the Main Influence Factors Based on MIV.* In order to reduce the MIV calculation error, this paper calculated the average absolute value of 10 calculation results to get the relative influence degree on construction costs of various factors. The calculation results are shown in Table 3.

As can be seen in Table 3, the MIV values of the first nine characteristics are larger, and that indicates the first nine characteristic parameters are the main characteristic parameter values that influence the output structure. So the first nine indicators can be set as new input variables:

Input set = {the number of main transformer sets in current, construction area (main control building), land requisition area (total station), the cable, main transformer capacity, location, steel frame and other materials, fill volume of earth leveling, foundation treatment method}.

(2) *Data Preprocessing.* Preprocessing results of the 9 indexes data selected by the MIV according to the normalization formula are shown in Table 4. And the results show the normalized data range within $(-1, 1)$.

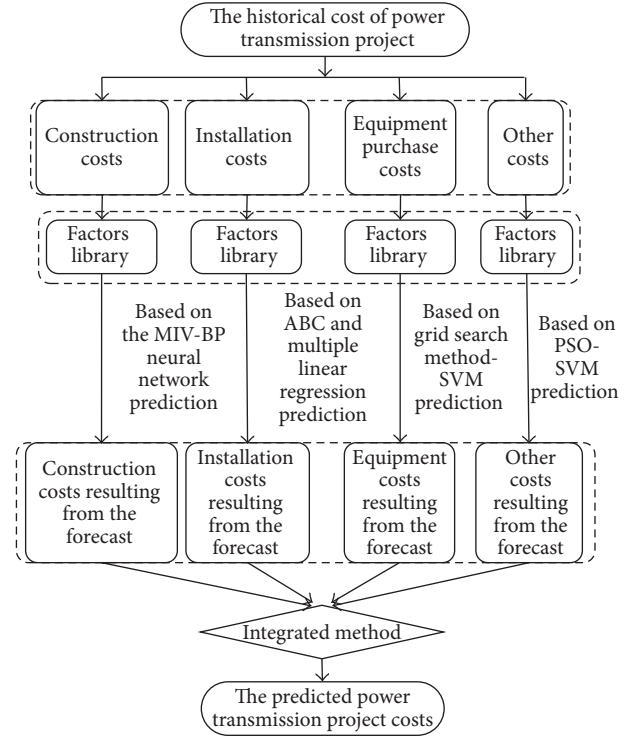


FIGURE 3: Decomposition-integration forecast model.

TABLE 3: The average MIV values of the factors.

Factor	MIV
The number of this period main transformer	3.1
Construction area (main control building)	2.9
Land area (total station)	2.7
Cable	2.7
The main transformer capacity	2.6
Area	2.1
Steel frame and other materials	1.9
Fill volume of site formation	1.7
Foundation treatment method	1.3
Altitude	1
Land area (inside the walls)	0.8
Cable trench	0.7
Pit road length	0.6
Pollution level	0.5
Excavation volume of site	0.5
Volume of the ditch pool	0.3
Construction area (total station)	0.3
Ground endurance	0.2
The weight of steel	0.2
The weight of the cement	0.1

(3) *The Forecast Analysis.* Build a three-layer neural network with nine inputs and one output, and dynamically adjust the hidden node number. The prediction error results are shown in Table 5.

TABLE 4: Part of index sets after being normalized.

Number of main transformer in current period	Construction area (main control building)	Land area (total station)	Cable	The capacity of main transformer	...
-0.98799	-0.93393	-0.99303	-0.93393	-0.69970	...
-0.98718	-0.92949	-0.99474	-0.92949	-0.67949	...
-0.99251	-0.94757	-0.98502	-0.94757	-0.62547	...
-0.99240	-0.95916	-0.98996	-0.95959	-0.61977	...
-0.98918	-0.93514	-0.97838	-0.93514	-0.45946	...
-0.99022	-0.94132	-0.98533	-0.94132	-0.75550	...
-0.99170	-0.94216	-0.98373	-0.94216	-0.58506	...
-0.99286	-0.94286	-0.98571	-0.94286	-0.42857	...
-0.98507	-0.90000	-0.98507	-0.90000	-0.25373	...
-0.99145	-0.95701	-0.98316	-0.95701	-0.57265	...
⋮	⋮	⋮	⋮	⋮	⋮

TABLE 5: The construction cost prediction error.

	1	2	3	4	5	6	7
The actual value	96.78	68.52	80.39	61.22	69.34	68.15	111.96
Predicted values	101.32	63.28	85.34	62.94	63.15	66.19	118.84
Error	4.70%	-7.65%	6.17%	2.80%	-8.94%	-2.88%	6.15%

TABLE 6: Equipment cost prediction error.

	1	2	3	4	5	6	7
The actual value	139.28	131.18	130.84	128.60	135.74	132.49	147.35
Predicted values	143.28	143.95	126.24	115.27	143.25	140.15	137.83
Error	2.88%	9.73%	-3.51%	-10.36%	5.53%	5.78%	-6.46%

4.2. Prediction of Equipment Purchase Cost

(1) *Multiple Linear Regression Equation Solving.* Build the multivariate linear regression model of equipment purchase expense by setting the unit prices of main transformer system as x_1 , high side power distribution device as x_2 , low voltage side power distribution device as x_3 , capacitor banks as x_4 , computer monitoring system as x_5 , station transformer as x_6 , station power distribution equipment as x_7 , and communication system as x_8 as the independent variable. Calculate with SPSS statistical software, and the following equation can be obtained:

$$Y = -80.0609 + 0.5171x_1 - 0.7996x_2 + 26.0658x_3 + 142.3038x_4 - 12.1112x_5 + 2.4942x_6 + 16.7465x_7 + 3.8035x_8. \quad (8)$$

(2) Model Checking

- After calculating, the multiple coefficient of determination is $R^2 = 0.9843$, indicating the model fit better.
- According to the sample to work out the statistics, $F = 8.17$. Through querying the table, $F_{0.05}(8, 41) = 2.18$,

thus proves that the linear relation of the model is obvious.

- Forecast analysis follows. According to the trained model, compute the test set with the regression equation to calculate the prediction error; the results are shown in Table 6.

4.3. Prediction of Installation Cost

(1) *Data Preprocessing.* In order to uniform dimension and eliminate the influence of the numerical value and the unit, before SVM model training, the historical data needs to be preprocessed according to the normalization formula.

(2) Parameters Optimization Based on Grid Search Method

- Firstly, set the variation range of penalty parameter C , RBF, and kernel parameter g . For example, set $c_{\min} = -8$ and $c_{\max} = 8$, which is to find the best parameter c in $[2 \wedge c_{\min}, 2 \wedge c_{\max}]$; set $g_{\min} = -8$, $g_{\max} = 8$, and the variation range of RBF kernel parameter g is $[2 \wedge g_{\min}, 2 \wedge g_{\max}]$, namely, to find the RBF kernel parameter g in this range.

TABLE 7: Installation costs forecast error.

	1	2	3	4	5	6	7
The actual value	20.17	23.34	20.84	23.71	20.00	22.16	34.45
Predicted values	21.95	25.34	22.84	21.54	19.63	20.52	31.64
Error	8.86%	8.59%	9.63%	-9.14%	-1.86%	-7.40%	-8.15%

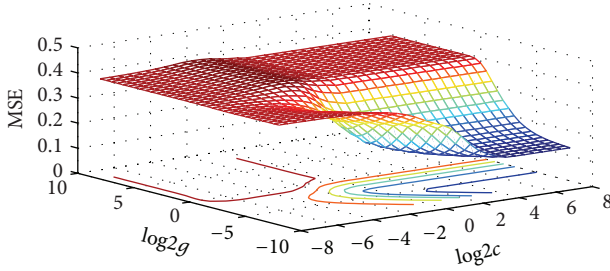


FIGURE 4: Parameter optimization based on grid search method. The SVR parameter selection results (3D view) [grid search method]. Best $c = 32$ $g = 0.0039063$ CVMSE = 0.13206.

- (b) Set the step size of c and g . Take $c_{\text{step}} = 0.5$ and $g_{\text{step}} = 0.5$ as an example, and the range of c in optimization process is $[2 \wedge c_{\min}, 2 \wedge (c_{\min} + c_{\text{step}}), \dots, 2 \wedge c_{\max}]$ and the range of g is $[2 \wedge g_{\min}, 2 \wedge (g_{\min} + g_{\text{step}}), \dots, 2 \wedge g_{\max}]$.
- (c) Consider cross validation. Set the parameters of cross validation process $V = 5$, namely, to handle the training set for 5-fold cross validation process.

Parameter optimization process can be observed through 3D map generated by MATLAB. The highest point represents the optimal parameter combination point, as shown in Figure 4.

After the parameter optimization based on grid search method, the optimal values can be obtained: $c = 32$ and $g = 0.0039063$.

(3) *The Forecast Analysis.* Predict the cost with the trained SVM model after getting the optimal values of penalty parameter c , the RBF, and the kernel parameter g . The installation costs forecast error is shown in Table 7.

4.4. Prediction of Other Costs

(1) Data Preprocessing

(a) *Normalization Process.* In order to make the data processing more convenient, normalize more than 30 original indexes according to the normalization formula.

(b) *PCA Dimensionality Reduction.* Use PCA to reduce the dimension of input parameters as there are too many original indexes. The calculation process generates 8 principal components, and the cumulative contribution rate reaches 97.8175%.

(2) *The Parameter Optimization Based on PSO.* Set the Cross Validation Fold Number $V = 5$, the evolution generation

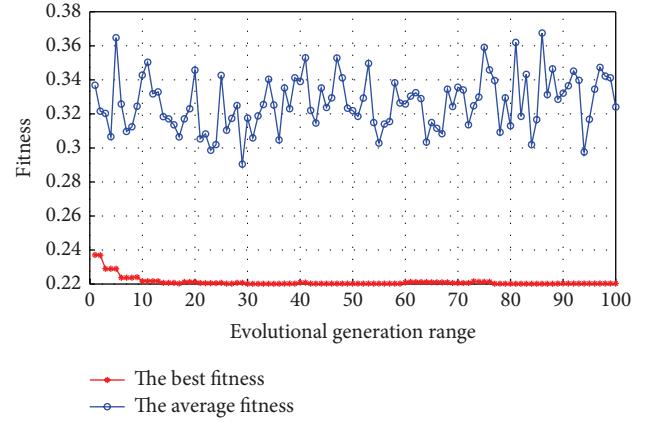


FIGURE 5: PSO parameter optimization. The fit curve MSE [PSO method] (parameter $c_1 = 1.5$, $c_2 = 1.7$, termination generation = 100, and population number pop = 20). Best $c = 2.9811$ $g = 0.01$ CVMSE = 0.22128.

as 100, the population as 20, and the range of the values of parameters c and g as $[0, 100]$.

Optimization process is shown in Figure 5.

After the PSO parameter optimization, get the optimal values of parameters $c = 2.9811$ and $g = 0.01$.

(3) *Forecast and Analysis.* Through PSO parameters optimization, work out the optimal parameters c and g . Then use the trained SVM model to predict the cost. The prediction error is shown in Table 8.

4.5. Prediction of Total Cost

(1) *Simple Summation.* The substation project unit capacity predicted cost generated by module integrating can be obtained by adding the predicted data of test set. The calculated error is shown in Table 9.

(2) *The Weighted Summation.* Set predictive values of the four charges as $x = (x_1, x_2, x_3, x_4)$ and the proportion of each charge as $p = (p_1, p_2, p_3, p_4)$.

According to the weighted summation formula,

$$X = \frac{x_1/p_1 + x_2/p_2 + x_3/p_3 + x_4/p_4}{4}. \quad (9)$$

The statistical analysis of four charges of substation engineering cost shows that the proportions of construction engineering cost, equipment purchase cost, installation cost, and other costs are 30.96%, 43.37%, 7.66%, and 43.37%.

TABLE 8: Other expenses prediction error.

	1	2	3	4	5	6	7
The actual value	37.26	43.93	37.05	43.68	49.47	34.38	68.77
Predicted values	35.79	40.87	36.62	45.54	53.54	37.76	64.85
Error	-3.96%	-6.96%	-1.16%	4.25%	8.23%	9.84%	-5.71%

TABLE 9: Simple integration error.

	1	2	3	4	5	6	7
The actual value	293.48	266.96	269.11	257.21	274.55	257.18	362.52
Predicted values	302.35	273.44	271.05	245.29	279.57	264.63	353.16
Error	3.02%	2.43%	0.72%	-4.64%	1.83%	2.89%	-2.58%

TABLE 10: The error of weighted summation.

	1	2	3	4	5	6	7
The actual value	293.48	266.96	269.11	257.21	274.55	257.18	362.52
Predicted values	285.74	273.52	267.07	250.79	271.95	253.63	368.69
Error	-2.64%	2.46%	-0.76%	-2.50%	-0.95%	-1.38%	1.70%

TABLE 11: Error analysis of the models.

	1	2	3	4	5	6	7
Decomposition-integration model	-2.64%	2.46%	-0.76%	-2.50%	-0.95%	-1.38%	1.70%
BP network	8.42%	-7.52%	7.15%	11.63%	-7.61%	5.38%	12.72%
PSO-SVM	4.72%	8.02%	-3.72%	7.27%	7.07%	-2.92%	8.61%

Then calculate the weighted summation according to these proportions. The weighted summation errors are shown in Table 10.

By comparing the two methods of integration, it can be seen that prediction error of the weighted summation is only a little bigger in samples 2 and 3 than the values of simple summation. And it suggests that the weighted integration can greatly reduce the prediction error and improve the prediction accuracy.

(3) *Comparative Analysis of Models.* Though neural network and support vector machine (SVM) belong to the artificial intelligence algorithms, the two models have obvious difference in prediction accuracy and stability. It can be found in the experiment process that the generalization ability of SVM is better than BP network; the error of previous experiment results of BP network is volatile, so it is difficult for BP network to achieve the global optimal result. It also can be seen that training results of BP network are not stable.

The traditional linear regression analysis is applicable to the models with obvious linear relationship like equipment purchase model and so forth. The experimental results show that, compared with artificial intelligence algorithms such as SVM, linear regression model is simpler and easier to operate, but the error precision depends on the selection of the independent variables. In the model validation, the errors of equipment purchase expense of two samples are higher than 9%, and the fluctuation is bigger than other three charges.

In order to further verify the application effect of decomposition-integration model in cost prediction, BP network and PSO-SVM are compared to the weighted integration model in terms of the total cost prediction error; the results are shown in Table 11.

It can be seen from Table 11 that the sorting of prediction effect accuracy of the three kinds of models is decomposition-integration > PSO-SVM > BP network. The decomposition-integration model considers the problem from multiple perspectives. And it combines the advantages of other models to improve the prediction accuracy of a single model.

5. Conclusion

The total cost is broken into the project construction cost, equipment purchase cost, installation engineering cost, and other costs from the perspective of cost decomposition. Mine and identify the influence factors of their costs, respectively. And establish the adaptive forecast model. Finally integrate the four charges. Through the analysis of the actual forecast for substation construction cost in Zhejiang province, the following conclusions can be drawn:

- (1) The decomposition-integration model is proposed from the angle of multicomponent forecasting. Convert the total cost prediction problem to some child costs prediction. Although the forecasting principle is simple, the advantages of this model are obvious. From engineering point of view, this model can dig

out the main influencing factors and choose the suitable prediction model of every child cost separately, which can be targeted to improve the prediction accuracy of each cost. From the theoretical point of view, this model combines the various forecasting methods to overcome the disadvantages of the single prediction model. Through the module of feedback regulation, it can choose highest accuracy of forecasting model series, which can improve the prediction precision greatly.

- (2) In the prediction module, according to the characteristics of the four child costs, select the suitable prediction model. For equipment purchase expense, there is significant linear relationship between factors and costs. A multivariate linear regression model can be built. While the factors of other three costs are too much and the relationship is complex, the models with strong generalization ability can be chosen, such as neural network and SVM. When optimizing the prediction model, SVM parameters are optimized by algorithm of grid search and PSO. The experiments prove that PSO has great advantage in efficiency, speed, and accuracy.
- (3) In the integration module, we can see that the prediction error is smaller using weighted summation compared with the simple summation from the experiments clearly. The reason is that the results using weighted summation are based on the statistical analysis results of four cost proportions in the total cost. It can add a fixed link compared with the simple summation. Thus, the prediction accuracy can be improved significantly.

In summary, the proposed decomposition-integrated model can improve the prediction accuracy effectively. This model is of a certain significance and good application effect in the power transmission project cost prediction.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was partially supported by the Natural Science Foundation of China (71471059) and the “industry-academics-research” scientific project of “industry-academics-research” coculture postgraduate base in Beijing.

References

- [1] X. Zhang, *Power Engineering Cost Management*, Hefei University of Technology Press, 2011.
- [2] Z. Wang, “Gray GM (1, 1) model in the application of the electric power construction project cost prediction,” *Journal of Inner Mongolia Electric Power Technology*, no. 5, pp. 46–50, 1996.
- [3] D. Wang, Y. Lei, Y. Huang, and Q. Li, “Prediction of comprehensive cost index based on markov chain,” *Power System Technology*, vol. 35, no. 2, pp. 193–197, 2011.
- [4] J. Wang and L. Ding, “Comprehensive prediction models for transmission engineering project cost based on its key influencing factors,” *East China Electric Power*, vol. 36, no. 11, pp. 111–113, 2009.
- [5] H. Guo, “Study on cost estimation of transmission projects based on regression analysis,” *Journal of Shanxi Electric Power*, vol. 36, no. 3, pp. 75–77, 2008.
- [6] Y. Ling, P. Yan, C. Han, and C. Yang, “BP neural network based cost prediction model for transmission projects,” *Electric Power*, vol. 45, no. 10, pp. 95–99, 2012.
- [7] Q. Ying, Y. Tan, and P. Chi, “ANN based power engineering cost forecast,” *Journal of Jiangxi Vocational and Technical College of Electricity*, vol. 24, no. 4, pp. 21–23, 2012.
- [8] F. Yi, *Fuzzy Mathematical-Based Project Cost Estimate on Transmission Line*, North China Electric Power University, 2012.
- [9] G.-J. Peng, J.-H. Yu, J.-T. Wei, and G. Yang, “Cost forecast model for power engineering based on feature extraction and small-sample learning,” *Journal of Chongqing University*, vol. 32, no. 9, pp. 1104–1110, 2009.
- [10] X. Wu, *A Novel Approach Based on SVM to Forecasting the Thermal Power Plant Construction Cost*, North China Electric Power University, Beijing, China, 2009.
- [11] Y. Yang, Y. Wang, X. Fan, and C. Liu, “Cost estimate of power line projects based on grey relational analysis and neural networks,” *Journal of Chongqing University*, vol. 36, no. 11, pp. 15–20, 2013.
- [12] N. Jie and T. Fei, “Combination model for predicting construction cost based on hybrid chaos SVM and ARIMA,” *Journal of Engineering Management*, vol. 27, no. 6, pp. 25–29, 2013.
- [13] N. Dongxiao, *A new method of non-deterministic project plan management research [Dissertation]*, North China Electric Power University, 2001.
- [14] National Development and Reform Commission, *Power Grid Construction Budgeting and Calculation Standard*, National Development and Reform Commission, Beijing, China, 2013.
- [15] Z. Zhenyu, L. Qianlei, Y. Weiyang, and G. Yun, “An evaluation index model of construction cost for 35 kV transmission lines in rural power network,” *Power System Technology*, vol. 32, no. 14, pp. 96–100, 2008.
- [16] F. He, X. Wang, S. Xiao, and X. Li, “Research on risk recognition of real estate projects based on MIV-BP neural network test,” *Operations Research and Management Science*, vol. 22, no. 2, pp. 229–234, 2013.
- [17] J. Wu, L. Zhou, X. Du, and P. Sun, “Junction temperature prediction of IGBT power module based on BP neural network,” *Journal of Electrical Engineering & Technology*, vol. 9, no. 3, pp. 970–977, 2014.
- [18] S. Traore, Y. M. Wang, and W. G. Chung, “Predictive accuracy of backpropagation neural network methodology in evapotranspiration forecasting in Dédougou region, western Burkina Faso,” *Journal of Earth System Science*, vol. 123, no. 2, pp. 307–318, 2014.
- [19] D. Niu, S. Cao, and J. Lu, *The Power Load Forecasting Technique and Its Application*, China Power Press, Beijing, China, 2009.
- [20] S. M. Shahandashti and B. Ashuri, “Forecasting engineering news-record construction cost index using multivariate time series models,” *Journal of Construction Engineering and Management*, vol. 139, no. 9, pp. 1237–1243, 2013.

- [21] H. Wang and J. Meng, "Predictive modeling on multivariate linear regression," *Journal of Beijing University of Aeronautics and Astronautics*, vol. 33, no. 4, pp. 500–504, 2007.
- [22] X. Yan and N. A. Chowdhury, "Mid-term electricity market clearing price forecasting: a multiple SVM approach," *International Journal of Electrical Power & Energy Systems*, vol. 58, pp. 206–214, 2014.
- [23] S.-F. Ding, B.-J. Qi, and H.-Y. Tan, "An overview on theory and algorithm of support vector machines," *Journal of the University of Electronic Science and Technology of China*, vol. 40, no. 1, pp. 2–10, 2011.
- [24] X. Zhang, S. Feng, and C. Wang, "Support vector machine model for predicting sand liquefaction based on grid-search method," *Chinese Journal of Applied Mechanics*, vol. 28, no. 1, pp. 24–28, 2011.
- [25] Z. Songlin, M. Meiqin, and S. Jianhui, "Prediction of wind power based on principal component analysis and artificial neural network," *Power System Technology*, vol. 35, no. 9, pp. 128–132, 2011.
- [26] Z. Sun, J. Pan, Z. Zhang, L. Zhang, and Q. Ding, "Traffic flow forecast based on combining principal component analysis with support vector machine," *Journal of Highway and Transportation Research and Development*, vol. 26, no. 5, pp. 127–131, 2009.
- [27] X. Zhi and L. Lingling, "Forecast of highway traffic volume using PSO-SVM," *Journal of Management Review*, vol. 23, no. 12, pp. 31–38, 2011.
- [28] A. Selakov, D. Cvijetinović, L. Milović, S. Mellon, and D. Bekut, "Hybrid PSO-SVM method for short-term load forecasting during periods with significant temperature variations in city of Burbank," *Applied Soft Computing Journal*, vol. 16, pp. 80–88, 2014.
- [29] W.-H. Steeb, *Nonlinear Workbook: Chaos, Fractals, Cellular Automata, Neural Networks, Genetic Algorithms, Gene Expression*, World Scientific, 2008.

