

Research Article

Intelligent Prediction of the Construction Cost of Substation Projects Using Support Vector Machine Optimized by Particle Swarm Optimization

Tongyao Lin,^{1,2} Tao Yi,^{1,2} Chao Zhang,^{1,2} and Jinpeng Liu ^{1,2}

¹School of Economics and Management, North China Electric Power University, Beijing 102206, China

²Beijing Key Laboratory of New Energy and Low-Carbon Development, North China Electric Power University, Beijing 102206, China

Correspondence should be addressed to Jinpeng Liu; hbdlljp@163.com

Received 8 July 2019; Revised 2 September 2019; Accepted 6 September 2019; Published 25 September 2019

Academic Editor: Yang Li

Copyright © 2019 Tongyao Lin 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.

To establish and consummate the electric power network, the construction and investment scale of power substation projects is expanding every year. As a capital-technology-intensive project, it has high requirements for power substation project management. Accurate cost forecasting can help to reduce the project cost, improve the investment efficiency, and optimize project management. However, affected by many factors, the construction cost of a power substation project usually presents strong nonlinearity and uncertainty, which make it difficult to accurately forecast the cost. This paper presents a new hybrid substation project cost forecasting method called PCA-PSO-SVM model, which is a support vector machine (SVM) model optimized by a particle swarm optimization (PSO) algorithm with principal component analysis (PCA). In this intelligent prediction model, the PCA method is introduced to reduce the data dimension. Furthermore, the PSO algorithm is used to optimize the model parameters. In the example, 65 sets of substation construction data are input into PCA-PSO-SVM model for construction cost prediction, and the prediction results are compared with other prediction methods to verify the forecasting accuracy. The results show that the MAPE and RMSE of the PCA-PSO-SVM model is 6.21% and 3.62, respectively. And, the prediction accuracy of this model is better than that of other models, which can provide a reliable basis for investment decision-making of substation projects.

1. Introduction

The substations are important parts of the power grid, which connects grids through transformers. With the rapid development of electric power demand, the construction scale of power substation project construction has been gradually expanded. As an important part of the economic evaluation of the power system, the cost of a power substation project has a great effect on the overall economic benefit of the power network. The accurate cost prediction of power substation projects is of great significance, which has positive effects on reducing the project cost of power substation projects, optimizing construction management and rationally formulating investment plans. Therefore, in order to make reasonable investment decisions, it is necessary to

study the intelligent cost prediction model in depth to improve the speed and accuracy of substation project cost prediction.

Scholars worldwide have conducted in-depth research on the cost influence factors of power substation. Niu et al. elaborated that, in terms of the affecting factors of the construction cost, the technical type and voltage level are commonly considered the main factors that affect the cost of power grid projects [1]. Mingjun et al. proposed a model of the asset wall projecting based on stochastic simulation to predict the influence of primary equipment cost on the overall investment of substation [2]. Xu and Moon proposed that the construction cost index (CCI) has been frequently used to predict the cost of construction projects because it reflects a comprehensive trend in the construction costs and

is applicable to the cost prediction of substation engineering projects [3]. As constructed by Weidong et al., through Paasche index analysis or Laspeyres index analysis, the weight of each project cost index can be calculated [4]. Tarmizi et al. discussed that, macroscopically, the level of socioeconomic growth and tax (such as land tax and construction tax) in the environment where the project is located also affects the construction cost index and consequently the project cost level [5].

With the development of artificial intelligence, machine intelligent algorithms, such as neural network model, support vector machine model, and gray prediction model, have gradually been used in project cost prediction. Gulcicek et al. predicted the construction cost of reinforced concrete housing estate buildings by using artificial neural network and multiple regression model and investigated the relationship between cost and influencing factors like earthquake region, soil type, floor area, and the number of stories [6]. Cheng et al. proposed an artificial intelligence approach, the evolutionary fuzzy hybrid neural network (EFHNN), to forecast the estimates of overall and category costs for actual construction projects [7]. Liu et al. built the gray fuzzy predictive model of project costs based on the gray fuzzy predictive theory, which can be used to estimate the budget costs work scheduled of the unfinished project in construction phase [8]. Qin et al. take qualitative and quantitative cost indicators as input sets and single costs as output sets of indicators and use support vector machine (SVM) and LSSVM to predict the cost of 25 residential projects [9]. Cheng et al. established a hybrid intelligence system, named as ELSVM, for modeling construction price variations quantified by the construction cost index (CCI) [10]. Vahdani et al. applied nonlinear regression and back-propagation neural networks (BPNNs) to estimate the conceptual costs in the construction project [11].

Among these methods, support vector machines have unique advantages. SVM is a novel small sample learning method with solid theoretical basis [12]. It basically does not involve probability measure and law of large numbers, which is different from existing statistical methods [13]. In essence, it avoids the traditional process from induction to deduction, achieves the efficient “transduction reasoning” from training samples to prediction samples, and greatly simplifies the usual classification and regression problems [14]. Therefore, these characteristics of support vector machine make it very suitable for solving the problem of power transformation project construction cost prediction.

The quality of input data has a great impact on the accuracy of the prediction model. Therefore, it is necessary to introduce the feature selection method to reduce the dimension of the original data. The commonly used methods of dimension reduction include linear discriminant analysis (LDA), principal component analysis (PCA), and so on [15]. LDA is a supervised machine learning algorithm that uses tag information to select the direction with the best classification performance [16]. While PCA converts a given set of related variables (dimensions) into another set of unrelated variables by linear transformation. This algorithm is an unsupervised learning method, which is more suitable to

be used in combination with other algorithms [17]. Therefore, this paper uses the latter to preprocess the model data.

In order to improve the accuracy of prediction, evolutionary algorithms are usually introduced into the model to optimize the parameters. Guo et al. used genetic algorithm (GA) to optimize the parameter selection methods of SVM in accordance with training data and improved SVM forecast precision [18]. Sun and Sun presents a novel hybrid model based on least squares support vector machine (LSSVM) optimized by cuckoo search (CS). The parameters in LSSVM are fine-tuned by CS to improve its generalization [19]. Yi et al. applied particle swarm optimization (PSO) in least squares support vector machine (LSSVM) to improve the accuracy of the construction cost forecasting [20]. Among these optimization algorithms, particle swarm optimization (PSO) can quickly approach the optimal solution and effectively optimize the model parameters [21, 22]. Due to good exploration capability, particle swarm optimization (PSO) has shown advantages on solving supervised feature selection problems [23]. In this paper, this algorithm is used to optimize the parameters of SVM.

This study extends research on the construction cost prediction of substation projects. The meager literature has revealed that the application of machine learning method in construction cost prediction has significantly improved the accuracy of prediction and effectively optimized the construction management. However, since the cost of substation project is affected by many complex factors, the accuracy of existing forecasting methods in this kind of special project is usually hard to meet the requirements. For example, the neural network model and the regression prediction model cannot adapt to the characteristics of fewer samples of substation project, and the prediction model based on gray theory cannot adapt to the characteristics of many influencing factors of substation project, so the error generated is rarely less than 10%. This study represents one of the first attempts to fill this important void by applying support vector machine optimized by principal component analysis and particle swarm algorithm to forecast the construction cost of substation projects.

This paper is organized as follows: Section 2 briefly introduces SVM, PCA, and PSO. Section 3 analyses the factors affecting the construction cost of substation project by using fishbone diagram. Section 4 conducts an empirical study. Section 5 compares the results to verify the proposed model, and Section 6 obtains the conclusion.

2. Model Principles

2.1. Support Vector Machine. The support vector machine is a type of machine learning analysis model that is widely used in data identification and prediction [24]. The SVM maps the input sample to the high-dimensional feature space F through a nonlinear mapping function $\varphi(x)$ and performs linear regression in F . The regression function of SVM in the high-dimensional feature space is

$$f(x) = \omega \cdot \varphi(x) + b, \quad (1)$$

where ω and b are the normal vector and offset of the regression function, respectively.

The regression can be converted into the following optimization problem:

$$\begin{aligned} \min \quad & J = \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n (\xi_i^* + \xi_i) \\ \text{s.t.} \quad & \begin{cases} y_i - \omega \cdot \varphi(x) - b \leq \varepsilon + \xi_i, \\ \omega \cdot \varphi(x) + b - y_i \leq \varepsilon + \xi_i^*, \\ \xi_i, \xi_i^* \geq 0, \quad i = 1, 2, \dots, n, \end{cases} \end{aligned} \quad (2)$$

where $\|\omega\|$ is the term related to the function complexity, ε is the insensitive loss coefficient, ξ_i, ξ_i^* denotes the relaxation factor, and C denotes the penalty factor.

By introducing Lagrange multipliers, the optimization problem becomes a convex quadratic optimization problem:

$$\begin{aligned} L(\omega, b, \xi, \xi^*, a, a^*, \gamma, \gamma^*) = & \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \\ & - \sum_{i=1}^n a_i (\xi_i + \varepsilon - y_i + f(x_i)) \\ & - \sum_{i=1}^n (\xi_i y_i - \xi_i^* y_i^*), \end{aligned} \quad (3)$$

where a and a^* are Lagrange multipliers.

In order to speed up the solution, equation (3) is converted into dual form, that is,

$$\begin{aligned} W(a, a^*) = & -\frac{1}{2} \sum_{i=1}^n (a_i - a_i^*) (a_j - a_j^*) (\varphi(x_i), \varphi(x_j)) \\ & + \sum_{i=1}^n (a_i - a_i^*) y_i - \sum_{i=1}^n (a_i - a_i^*), \\ \text{s.t.} \quad & \begin{cases} \omega = \sum_{i,j=1}^n (\alpha_i - \alpha_i^*) x_i, \\ \sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0, \\ 0 \leq \alpha_i, \alpha_i^* \leq C. \end{cases} \end{aligned} \quad (4)$$

Kernel function $k(x_i, x)$ is used to replace inner product of vector $(\varphi(x_i), \varphi(x))$ in high-dimensional space to avoid dimension disaster. The regression function of SVM is

$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) k(x_i, x) + b. \quad (5)$$

Studies show that when there is a lack of prior knowledge of the process, the radial basis kernel function has fewer parameters and better performance than other kernel

functions. Therefore, this paper selects radial basis kernel function for SVM construction module, and the radial basis kernel function is defined as follows:

$$k(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right), \quad (6)$$

where σ is the width parameter of the radial basis kernel function.

It can be seen from the SVM modeling process that the learning performance of SVM is closely related to the selection of penalty coefficient C , insensitive coefficient ε , and kernel function parameter σ . In this paper, particle swarm optimization (PSO) is adopted to optimize SVM parameters.

2.2. Principal Component Analysis. Principal component analysis is a mathematical method mainly recombining several indices in the original multidimensional sample to form a new low-dimensional sample of several comprehensive indices [25]. These new indices are arranged in the descending order of variance, and the adopted method is mainly mathematical linear. Generally, the newly generated indicators are called the main components, and each principal component is a linear combination of the original indicators.

It is assumed that there are n sample data and that each sample has p index variables, by which an $n \times p$ data matrix is formed as

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}. \quad (7)$$

Each column (original index) is considered an original variable, and the new variable was obtained by a linear combination of these original variables:

$$\begin{cases} Z_1 = l_{11}X_1 + l_{12}X_2 + \cdots + l_{1p}X_p, \\ Z_2 = l_{21}X_1 + l_{22}X_2 + \cdots + l_{2p}X_p, \\ \vdots \\ Z_m = l_{m1}X_1 + l_{m2}X_2 + \cdots + l_{mp}X_p, \end{cases} \quad (8)$$

where Z_1 has the greatest variance in all linear combinations of X_1, X_2, \dots, X_p ; Z_2 has the greatest variance in all linear combinations of X_1, X_2, \dots, X_p and related to Z_1 . The remainder can be similarly done. Z_m has the greatest variance in all linear combinations of X_1, X_2, \dots, X_p and is not related to Z_1, Z_2, \dots, Z_{m-1} . Finally, Z_1, Z_2, \dots, Z_m can be obtained, which are called the principle component of the original data.

2.3. Particle Swarm Optimization. To improve the prediction accuracy, the particle swarm optimization (PSO) is introduced into the model to optimize the forecasting parameters. The potential optimal solution of the parameter is assumed to be a particle in its range space, the particle has velocity and position properties, and its excellence is

determined by the fitness value [26]. The velocity and position can be updated by the following formula:

$$\begin{aligned} v_{i,d}^{k+1} = & \omega \cdot v_{i,n}^j + c_1 \cdot \text{rand}() \cdot (\text{pbest}_{i,d}^k - x_{i,d}^k) \\ & + c_2 \cdot \text{rand}() \cdot (\text{gbest}_d^k - x_{i,d}^k), \end{aligned} \quad (9)$$

$$x_{i,d}^{k+1} = x_{i,d}^k + v_{i,d}^{k+1}, \quad (10)$$

where pbest is the best position of each individual particle, while gbest is the best position of the whole particle swarm, ω is the inertia weight, c_1 and c_2 are learning factors, $\text{rand}()$ is a random number between 0 and 1, and $v_{i,n}^{j+1}$ and $x_{i,n}^j$ are the velocity and position particles after k iterations.

In the actual application, this algorithm may get into the locally optimum solution in some situations. To solve this problem, the ‘‘catfish effect’’ is introduced into the model [27]. When particles gather in the local optimal result, the catfish operator can stimulate the particle swarm and make it jump out of the local extreme point and find the global most solution. This algorithm is called catfish particle swarm (CFPSO) [28]. CFPSO uses the deviation threshold as the trigger condition and perturbs the global extreme value or individual extreme value through catfish operator. The optimized formula is

$$\begin{aligned} v_{i,n}^{j+1} = & \omega \cdot v_{i,n}^j + c_1 \cdot \text{rand}() \cdot (c_3 \cdot \text{rand}() \cdot \text{pbest}_{i,d}^k - x_{i,d}^k) \\ & + c_2 \cdot \text{rand}() \cdot (c_4 \cdot \text{rand}() \cdot \text{pbest}_{i,d}^k - x_{i,d}^k), \end{aligned} \quad (11)$$

where c_3 is the stimulus intensity of the catfish operator to the optimal position of individual particle, c_4 is the stimulus intensity of catfish operator to the optimal position of particle swarm, and $c_3 \cdot \text{rand}()$ and $c_4 \cdot \text{rand}()$ are catfish operators, which are defined as follows:

$$\begin{aligned} c_3 \cdot \text{rand}() = & \begin{cases} 1, & e_p > e_{0p}, \\ c_3 \cdot \text{rand}(), & e_p \leq e_{0p}, \end{cases} \\ c_4 \cdot \text{rand}() = & \begin{cases} 1, & e_g > e_{0g}, \\ c_4 \cdot \text{rand}(), & e_g \leq e_{0g}. \end{cases} \end{aligned} \quad (12)$$

In formula (9), e_p is the deviation between the current value and the current individual optimal, e_g is the deviation between the current value and the current global optimum, and e_{0p} and e_{0g} are the threshold values of e_p and e_g . According to the definition, if the deviation of the current value is less than the deviation threshold, the catfish operator changes the individual optimal value and global optimal value to avoid the local optimal value.

2.4. PCA-PSO-SVM Model. Based on the above principles, the specific process of the power substation project cost prediction model is built as shown in Figure 1.

As shown in the figure, the procedures of PCA-PSO-SVM are elaborated below:

Step 1: the construction cost data of the power substation should be collected and screened to form a data sample. Then, the principal component analysis is used

to reduce the dimension of the original data, i.e., to recombine a series construction cost that affects the factor data into a new set of mutually irrelevant comprehensive indicators F_1, F_2, \dots, F_n .

Step 2: in this step, several parameters of the particle swarm are generated. At each iteration of the swarm, the particle’s position and velocity are updated. The fitness of the particle should be calculated and compared with the current global optimal fitness to determine whether to update the global extremum. Update the particle’s velocity and position, and repeat the steps of calculating the current particle fitness according to the particle’s velocity and position until the global optimal position is output when the iteration number reaches the maximum iteration number or the iterative process has traversed all coordinates, that is, the parameters optimized for support vector machine. Finally, an optimized parameter prediction model is constructed according to the optimized parameters.

Step 3: in this step, the original data of substation engineering after PCA dimension reduction in step 1 and the optimal parameters obtained by PSO iteration in step 2 are input into the model to establish the PCA-PSO-SVM prediction model. This model can predict the construction cost credibly according to the impact index data of substation projects.

3. Influencing Factors of Cost

The quality of basic data of a substation project significantly affects the accuracy of the construction cost forecast. Therefore, it is necessary to analyze the influencing factors of the construction cost and select the most important influencing factors as the cost prediction indicators.

As shown in Figure 2, the cost of a power substation project generally consists of four parts: construction cost, equipment cost, installation cost, and other costs. Among them, the equipment cost and construction cost more obviously affect the total cost than the other two parts. The construction cost was mainly affected by the substation station building type, building area, concrete consumption, and other factors. The equipment cost is mainly related to the quantity and price of the transformer equipment and circuit breaker equipment in the power substation station. The factors related to the installation cost include the length of the high-voltage cable, middle-voltage cable, low-voltage cable, power cable, and control cable. In addition, other factors such as the requisition area of the construction site, unit land acquisition price, and project planning period management cost affect the overall cost of the power substation project through other costs.

According to the preliminarily selected factors, the selection and quantitative analyses are performed to simplify and process the indicators, remove the highly relevant ones, and classify and number the indicators according to the technical parameters, cost parameters, and external environment levels. The results are shown in Table 1.

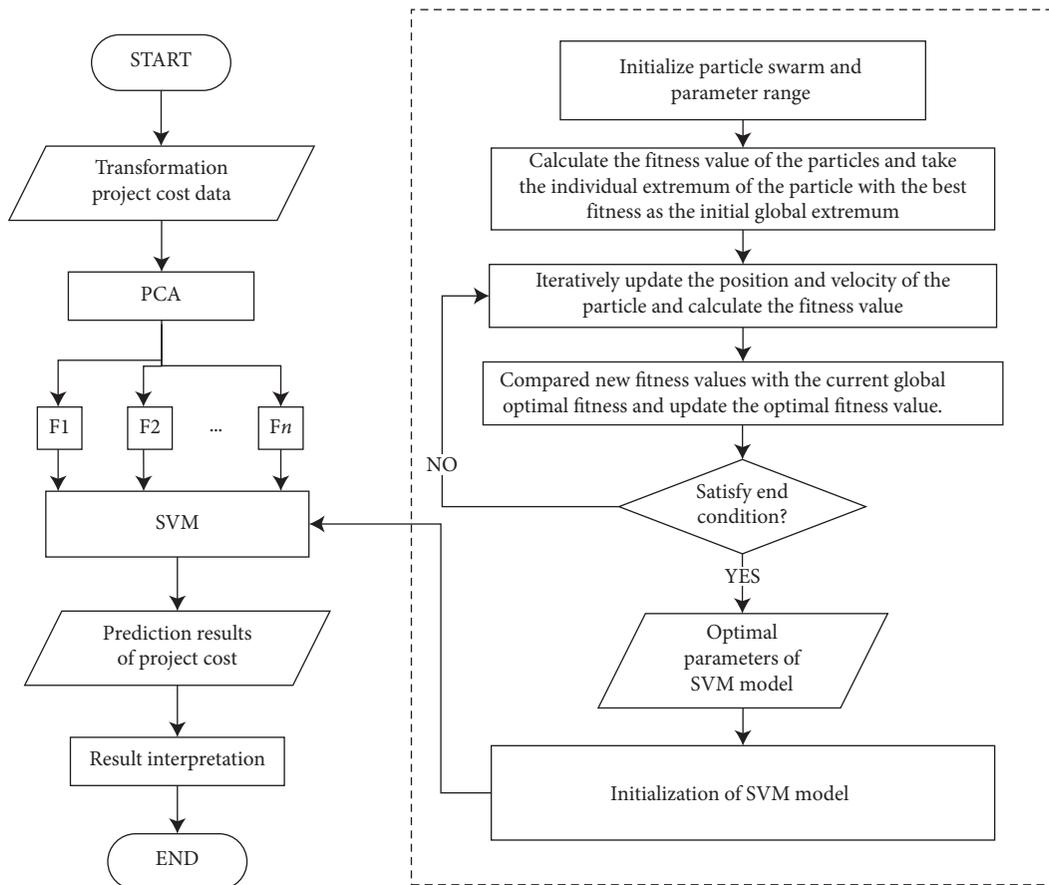


FIGURE 1: Flow chart of the power substation construction cost prediction model based on PCA-PSO-SVM.

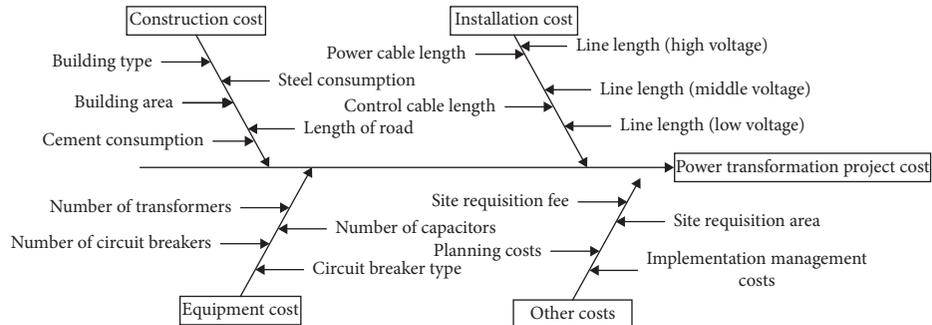


FIGURE 2: Fishbone diagram of the cost-affecting factors.

As the input indices of the cost prediction model of a power substation project, X_1 to X_{16} in Table 1 are called the key influencing factors of the cost. The cost of a power substation project, i.e., the output index of the prediction model, is the unit capacity cost of the power substation project, which is expressed by Y . The above input and output indicators are used to establish the PCA-PSO-SVM model, and the prediction ability of the model is verified by an example as follows.

4. Example Analysis

In this paper, the PCA-PSO-SVM model is applied to forecast the cost of a series of power substation construction projects in China, which were completed in 2017. The sample data

include the cost information of sixteen key indicators. On this basis, this paper establishes a response model between the cost-influencing factors and the cost of the power substation project. The prediction process of the model is shown below.

4.1. Data Preprocessing. This paper identifies and eliminates abnormal data points. The filtered 65 data samples are shown in Table 2.

4.2. Data Grouping. The processed cost data samples are divided into two groups: 55 data are used as model training samples and the other data are used as test samples.

TABLE 1: Sixteen key influencing factors.

Index type	Key influencing factors	Serial number
Technical parameter level	Transformer capacity	X_1
	Number of circuit breaker	X_2
	Number of exit route	X_3
	Number of reactors	X_4
	Number of capacitor	X_5
	Control cable length	X_6
	Power cable length	X_7
	Concrete consumption	X_8
	Steel consumption	X_9
Cost parameter level	Price of transformer	X_{10}
	Price of circuit breaker	X_{11}
	Price of concrete	X_{12}
	Price of steel	X_{13}
External environment	Site acquisition cost	X_{14}
	Site area	X_{15}
	Planning cost	X_{16}

TABLE 2: Cost data of the power substation construction (2017).

Serial number	Unit capacity cost (USD/kVA)	Serial number	Unit capacity cost (USD/kVA)	Serial number	Unit capacity cost (USD/kVA)
1	40.95	23	70.18	45	50.71
2	45.53	24	49.89	46	54.09
3	70.29	25	51.75	47	33.88
4	52.92	26	32.26	48	33.83
5	51.46	27	58.91	49	51.76
6	46.66	28	47.27	50	55.88
7	40.87	29	29.72	51	57.94
8	57.17	30	49.34	52	60.41
9	52.13	31	51.65	53	48.81
10	58.04	32	49.11	54	41.79
11	56.90	33	51.92	55	66.13
12	53.07	34	73.19	56	57.11
13	53.50	35	44.54	57	60.17
14	64.04	36	53.66	58	51.46
15	59.71	37	50.99	59	56.80
16	50.29	38	50.01	60	57.42
17	61.42	39	52.09	61	51.32
18	54.31	40	40.08	62	47.37
19	62.83	41	48.72	63	35.63
20	60.52	42	36.82	64	53.24
21	50.79	43	54.61	65	46.90
22	43.22	44	55.41		

4.3. *Principal Component Analysis.* In this paper, principal component analysis is carried out on the original data, and the analysis results are represented by scree plot, which can help to determine the optimal number of principal components. In Figure 3, the horizontal axis represents the number of principal components and the vertical axis represents the eigenvalue (variance) of the corresponding principal components. Generally, the steeper part of the line is extracted as the main component factor, while the part at the gentle slope has a small variance contribution and contains little information about the original variable, so it is not considered.

In the figure, the variance contribution rate of the first principal component is 0.467 and the cumulative

contribution rate of the first five principal components is 0.916. The principal component, whose cumulative contribution rate is 0.9, basically contains all information of the features. It can also be known from the figure that the turning point of steep slope and gentle slope is between principal component 5 and 6, and the variances of principal components 6 to 10 are small and have little difference, so it is appropriate to extract 5 principal components to build the SVM model.

4.4. *Particle Swarm Optimization.* This paper sets the population number as 30 and sets maximum number of

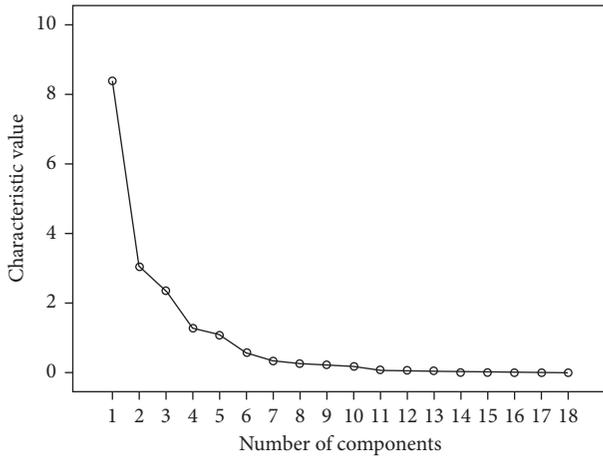


FIGURE 3: Scree plot of the principal component analysis.

iterations as 500. The values of $c_1, c_2, c_3,$ and c_4 are 1.5, 1.7, 1, and 4. The threshold values e_{0g} and e_{0p} are set as 0 and 0.01. The range of inertia weighting factor ω is set as [0.4, 0.95]. The range of the penalty coefficient C is [0, 300], and the range of the kernel parameters σ^2 is [0, 200]. The process of particle swarm optimization is shown in Figure 4. Particle swarm optimization (PSO) is used to traverse all cases within the value range of C and σ^2 and calculate mean-square error (MSE) in all cases. The point with the smallest MSE value, namely, the lowest point in Figure 4, is the value of optimal parameters C and σ^2 for the SVM model. The optimal parameter of the SVM based on particle swarm optimization is $C=292.634$ and $\sigma^2=23.472$.

4.5. Predictive Ability Test. In this paper, 10 grouped test samples were input into the PCA-PSO-SVM model. Table 3 shows the actual unit capacity costs of the ten samples and the unit capacity costs predicted by the model. In order to show the accuracy of model prediction intuitively, the absolute error and relative error of predicted value and actual value are calculated. Absolute error is the absolute value of the difference between the predicted value and the actual value, and relative error is the ratio of the predicted value to the actual value multiplied by 100%, expressed as a percentage.

It can be seen from Table 3 that the sample with the largest prediction error is project VI, with an absolute error of 6.15 and a relative error of 11.98%. The sample with the smallest prediction error is project VII, with an absolute error of 1.11 and a relative error of 2.35%.

5. Comparison of the Forecasting Results

To evaluate the accuracy of the PCA-PSO-SVM prediction model, GM (1, n), SVM, PCA-SVM, and PSO-SVM are selected for comparison.

The GM (1, n) model is a type of gray prediction model, which refers to the prediction of the development and change of the system behavior characteristic values and the

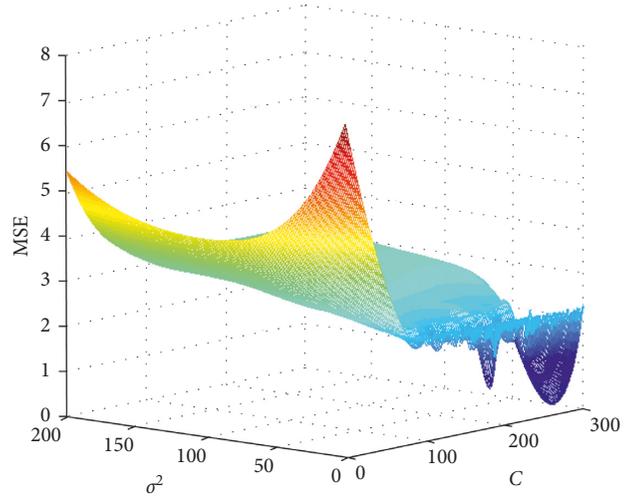


FIGURE 4: Process of particle swarm optimization.

prediction of a system that contains both known information and uncertain information. The prediction principle is to weaken the unknown factors in the gray system and strengthen the influence degree of the known factors. The GM (1, n) model is established to train and predict the same data samples, and the prediction results of the vector machine model are compared. The prediction results are shown in Figure 5(a).

In the previous paper, the support vector machine model has been introduced. The original data screened by the sample engineering through simple outliers are input into the SVM model for the training and prediction, and the output results are compared with the output results of the optimized model. The predicted results are shown in Figure 5(b).

To explore the effect of the PCA in dimensionality reduction, the PCA-SVM model is established for comparison and the prediction results are shown in Figure 5(c).

To explore the effect of the PSO in model parameter optimization, the PSO-SVM model is established for comparison and the prediction results are shown in Figure 5(d).

Finally, the predicted results of the PCA-PSO-SVM model are plotted in Figure 5(e).

In Figure 5, the distance between the predicted result line and the actual cost line in Figure 5(a) (GM (1, n)) and Figure 5(b) (SVM) is obviously larger than that in other figures. In Figure 5(c) (PCA-SVM) and Figure 5(d) (PSO-SVM), the fitting degree of the predicted result line with the actual cost line is significantly higher than that in Figures 5(a) and 5(b), and the distance between each sample point is relatively reduced. In Figure 5(e), the predicted result line of the PCA-PSO-SVM model is most consistent with the actual cost line of the sample project, and the distance among the sample points is minimal.

The comparison shows that the results predicted by a simple GM (1, n) model and the SVM model have large errors with the unit capacity cost of actual projects. The statistical results show that the average relative errors of the two methods are 19.16% and 17.10%. For the PCA-SVM model and PSO-SVM model, with the assistance of

TABLE 3: Forecasting result of the ten test samples.

Project	Actual unit capacity cost (USD/kVA)	Forecasting unit capacity cost (USD/kVA)	Absolute error	Relative error (%)
Project I	57.11	59.17	2.06	3.61
Project II	60.17	56.90	3.28	5.44
Project III	51.46	53.23	1.77	3.43
Project IV	56.80	62.65	5.85	10.30
Project V	57.42	55.59	1.82	3.17
Project VI	51.32	45.17	6.15	11.98
Project VII	47.37	46.26	1.11	2.35
Project VIII	35.63	32.65	2.98	8.37
Project IX	53.24	48.45	4.79	9.00
Project X	46.90	48.96	2.06	4.40

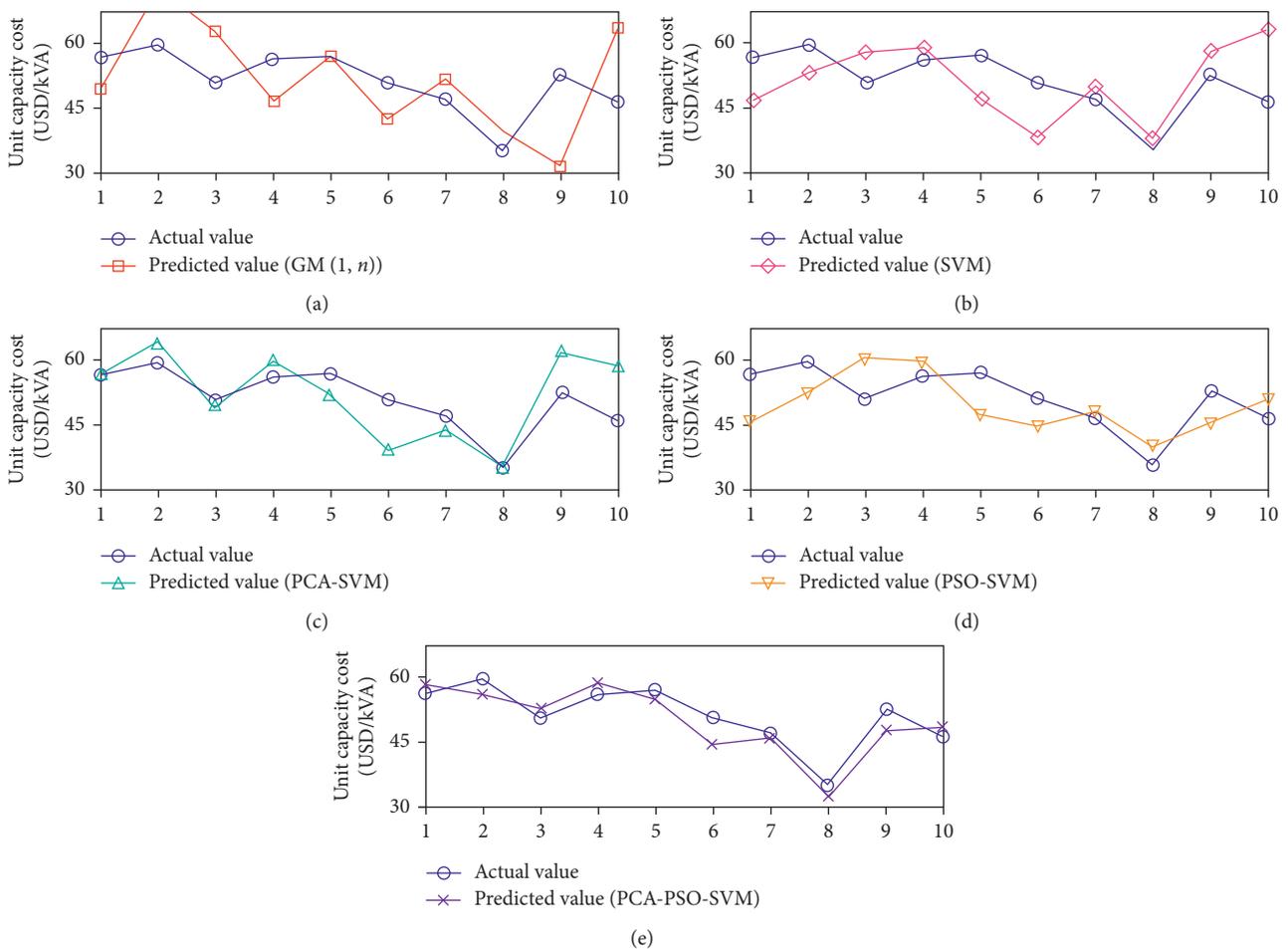


FIGURE 5: Comparison of the actual value and predicted value of different models.

the principal component analysis or particle swarm parameter optimization, the prediction accuracies are 41.17% and 29.4% higher than that of the SVM model. Compared with the above model, the PCA-PSO-SVM model is the closest to the actual unit capacity cost of 10 samples in the training set. The statistical results show that the average relative error is 6.21%, which is

significantly superior to the prediction results of other methods.

To further evaluate the accuracy of each prediction model, various evaluation indicators are selected to analyze the accuracy of the prediction results: the mean absolute percentage error (MAPE) and root-mean-square error (RMSE). The calculation methods are as follows:

TABLE 4: Prediction error of different models.

	GM (1, n)	SVM	PSO- SVM	PCA- SVM	PCA-PSO- SVM
MAPE	19.16%	17.10%	10.21%	11.57%	6.21%
RMSE	11.52	11.73	6.72	6.63	3.62

$$\text{MAPE} = \frac{1}{n} \sum_{k=1}^n \left| \frac{x(k) - \hat{x}(k)}{x(k)} \right| \times 100\%, \quad (13)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{k=1}^n (x(k) - \hat{x}(k))^2},$$

where $x(k)$ is the actual unit capacity cost of the power substation project.

The MAPE and RMSE of the above model are calculated as shown in Table 4. Both the PSO-SVM and PCA-SVM models have smaller errors for the predicted results than the SVM model, which indicates that the principal component analysis and particle swarm optimization algorithm improve the prediction accuracy. Table 4 also shows that both the MAPE and RMSE values of the PCA-PSO-SVM model are minimal. In order to verify the superiority of our proposed model, paired sample t test is applied to analyze the prediction error. The results of pairwise t test showed that the p values of the PCA-PSO-SVM model are 0.006, 0.023, 0.036, and 0.021, respectively, compared with the other four models. The t test results showed that the p values were all less than 0.05, indicating that the result predicted by PCA-PSO-SVM is statistically significantly better than those of the others.

As for the computational cost, according to statistics, the running time of the all these prediction models is in the range of 5–10 seconds. Although PCA-PSO-SVM model has no obvious advantages in running time cost and load cost, its prediction accuracy is obviously better than other methods.

Hence, the SVM model established by using the PSO search algorithm for model parameter optimization after the PCA treatment is the optimal prediction model with the best fitting and prediction ability for the complex response relationship between the unit capacity cost and the influencing factors of power substation projects.

6. Conclusion

Cost control of substation projects directly affects the economic benefits of power grid projects, so it is an urgent problem to predict the cost with high accuracy in the early stage of the project. Although cost prediction models based on various mathematical methods have been proposed in many literatures, they are difficult to apply to substation project and the error is generally high. For example, the neural network model and the regression prediction model cannot adapt to the characteristics of fewer samples of substation project, and the prediction model based on gray theory cannot adapt to the characteristics of many

influencing factors of substation project, so the error generated are rarely less than 10%.

This paper proposes an intelligent cost prediction model based on SVM optimized by PCA and PSO. Firstly, PCA is used to screen and reduce the dimension of the substation project cost data, and the principal components which can basically describe the factors affecting the cost are obtained as the input set of the prediction model. Secondly, the empirical parameters of SVM model which is mature in theory are given to determine the nonlinear mapping between substation project characteristics and cost. Then, the parameters of the model are optimized by PSO to obtain the optimized parameters model, and the model is trained. Finally, the trained model is used to predict the cost of substation project. And, the accuracy of the model is verified by comparing the cost predict value of different forecasting models and actual value of cost. The average relative error of the PCA-PSO-SVM optimization parameter model is less than 7% compared with the actual value of the cost. The MAPE and RMSE are 6.21% and 3.62, which proved that the accuracy of the PCA-PSO-SVM optimization parameter model is improved greatly compared with other prediction models.

In summary, the PCA-PSO-SVM cost prediction model proposed in this paper can effectively improve the accuracy of substation project cost prediction. This model has strong practical significance and good application effect in substation project cost prediction. In the future, more algorithms or models may be introduced into PCA-PSO-SVM substation construction cost prediction model to improve prediction accuracy and adapt to more complex situations.

Data Availability

The data in the example analysis used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (NSFC) (71501071), the Beijing Social Science Fund (16YJC064), and the Fundamental Research Funds for the Central Universities (2017MS059 and 2018ZD14).

References

- [1] D. X. Niu, H. H. Qiao, H. J. Zhai, and X. Lu, "Power grid construction project cost risk factors analysis based on the fishbone diagram theory," *Advanced Materials Research*, vol. 622-623, pp. 1852–1855, 2012.
- [2] L. Mingjun, C. Xing, S. Yiqun et al., "Technological transformation investment prediction of primary equipment of substation based on assets wall," in *Proceedings of the 2015 2nd*

- International Conference on Information Science and Control Engineering (ICISCE)*, April 2015.
- [3] J.-W. Xu and S. Moon, "Stochastic forecast of construction cost index using a cointegrated vector autoregression model," *Journal of Management in Engineering*, vol. 29, no. 1, pp. 10–18, 2013.
 - [4] L. Weidong, S. Huajun, L. Yan, H. Fuyu, and Q. Jinpeng, "Study on grid engineering cost index forecasting based on ARIMA-ES hybrid model," *Management Review*, vol. 28, no. 3, 2016.
 - [5] H. B. Tarmizi, M. Daulay, and I. Muda, "Impact of the economic growth and acquisition of land to the construction cost index in North Sumatra," *IOP Conference Series: Materials Science and Engineering*, vol. 180, 2017.
 - [6] U. Gulcicek, O. Ozkan, M. Gunduz, and I. H. Demir, "Cost assessment of construction projects through neural networks," *Canadian Journal of Civil Engineering*, vol. 40, no. 6, pp. 574–579, 2013.
 - [7] M.-Y. Cheng, H.-C. Tsai, and E. Sudjono, "Conceptual cost estimates using evolutionary fuzzy hybrid neural network for projects in construction industry," *Expert Systems with Applications*, vol. 37, no. 6, pp. 4224–4231, 2010.
 - [8] J. B. Liu, R. Hong, and Z. M. Li, "Model on dynamic control of project costs based on GM (1, 1) for construction enterprises," *Fuzzy Information and Engineering*, vol. 2, 2009.
 - [9] Z. F. Qin, X. L. Lei, D. Zhai, and L. Z. Jin, "Forecasting the costs of residential construction based on support vector machine and least squares-support vector machine," *Journal of Zhejiang University Science A*, vol. 43, pp. 357–363, 2016.
 - [10] M.-Y. Cheng, N.-D. Hoang, and Y.-W. Wu, "Hybrid intelligence approach based on LS-SVM and differential evolution for construction cost index estimation: a Taiwan case study," *Automation in Construction*, vol. 35, pp. 306–313, 2013.
 - [11] B. Vahdani, S. M. Mousavi, M. Mousakhani, M. Sharifi, and H. Hashemi, "A neural network model based on support vector machine for conceptual cost estimation in construction projects," *Journal of Optimization in Industrial Engineering*, vol. 5, no. 10, pp. 11–18, 2012.
 - [12] S. Amari and S. Wu, "Improving support vector machine classifiers by modifying kernel functions," *Neural Networks*, vol. 12, no. 6, pp. 783–789, 1999.
 - [13] Z.-B. Shi, Y. Li, and T. Yu, "Short-term load forecasting based on LS-SVM optimized by bacterial colony chemotaxis algorithm," in *Proceedings of the 2009 International Conference on Information and Multimedia Technology*, Jeju Island, South Korea, December 2009.
 - [14] T. Howley and M. G. Madden, "The genetic kernel support vector machine: description and evaluation," *Artificial Intelligence Review*, vol. 24, no. 3–4, pp. 379–395, 2005.
 - [15] J. Li, K. Cheng, S. Wang et al., "Feature selection: a data perspective," *ACM Computing Surveys*, vol. 50, no. 6, pp. 1–45, 2017.
 - [16] S. Laohakiat, S. Phimolares, and C. Lursinsap, "A clustering algorithm for stream data with lda-based unsupervised localized dimension reduction," *Information Sciences*, vol. 381, pp. 104–123, 2017.
 - [17] K. Yata and M. Aoshima, "Effective PCA for high-dimension, low-sample-size data with noise reduction via geometric representations," *Journal of Multivariate Analysis*, vol. 105, no. 1, pp. 193–215, 2012.
 - [18] X. Guo, D. Li, and A. Zhang, "Improved support vector machine oil price forecast model based on genetic algorithm optimization parameters," *AASRI Procedia*, vol. 1, pp. 525–530, 2012.
 - [19] W. Sun and J. Sun, "Daily PM_{2.5} concentration prediction based on principal component analysis and LSSVM optimized by cuckoo search algorithm," *Journal of Environmental Management*, vol. 188, pp. 144–152, 2016.
 - [20] T. Yi, H. Zheng, Y. Tian, and J.-P. Liu, "Intelligent prediction of transmission line project cost based on least squares support vector machine optimized by particle swarm optimization," *Mathematical Problems in Engineering*, vol. 2018, Article ID 5458696, 11 pages, 2018.
 - [21] Y. J. Zhang, T. Li, G. Y. Na, G. Q. Li, and Y. Li, "Optimized extreme learning machine for power system transient Stability prediction using synchrophasors," *Mathematical Problems in Engineering*, vol. 2015, Article ID 529724, 8 pages, 2015.
 - [22] Y. Li, Y. Li, G. Li, D. Zhao, and C. Chen, "Two-stage multi-objective opf for AC/DC grids with VSC-HVDC: incorporating decisions analysis into optimization process," *Energy*, vol. 147, pp. 286–296, 2018.
 - [23] Y. Zhang, H.-G. Li, Q. Wang, and C. Peng, "A filter-based bare-bone particle swarm optimization algorithm for unsupervised feature selection," *Applied Intelligence*, vol. 49, no. 8, pp. 2889–2898, 2019.
 - [24] A. Chaudhuri and K. De, "Fuzzy support vector machine for bankruptcy prediction," *Applied Soft Computing*, vol. 11, no. 2, pp. 2472–2486, 2011.
 - [25] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," *Journal of the ACM*, vol. 58, no. 3, pp. 1–37, 2011.
 - [26] W. Qi, "A self-adaptive embedded chaotic particle swarm optimization for parameters selection of Wv - SVM.," *Expert Systems with Applications*, vol. 38, no. 1, pp. 184–192, 2011.
 - [27] K. S. Kumar and S. Naveen, "Power system reconfiguration and loss minimization for a distribution systems using "catfish PSO" algorithm," *Frontiers in Energy*, vol. 8, no. 4, pp. 434–442, 2015.
 - [28] L.-Y. Chuang, S.-W. Tsai, and C.-H. Yang, "Improved binary particle swarm optimization using catfish effect for feature selection," *Expert Systems with Applications*, vol. 38, no. 10, pp. 12699–12707, 2011.

