

## Research Article

# Intelligent Prediction of Transmission Line Project Cost Based on Least Squares Support Vector Machine Optimized by Particle Swarm Optimization

Tao Yi,<sup>1,2</sup> Hao Zheng ,<sup>1,2</sup> Yu Tian,<sup>1,2</sup> and Jin-peng Liu <sup>1,2</sup>

<sup>1</sup>School of Economics and Management, North China Electric Power University, Beijing 102206, China

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

Correspondence should be addressed to Hao Zheng; 18811358655@163.com and Jin-peng Liu; hbdlljp@163.com

Received 14 November 2018; Accepted 3 December 2018; Published 17 December 2018

Academic Editor: Anna M. Gil-Lafuente

Copyright © 2018 Tao Yi 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.

In order to meet the demand of power supply, the construction of transmission line projects is constantly advancing, and the level of cost control is constantly improving, which puts forward higher requirements for the accuracy of cost prediction. This paper proposes an intelligent cost prediction model based on least squares support vector machine (LSSVM) optimized by particle swarm optimization (PSO). Originally extracting natural, technological, and economic indexes from the perspective of cost composition, principal component analysis (PCA) is used to reduce the dimension of indexes. And PSO is innovatively introduced to optimize the parameters of LSSVM model to obtain the optimal parameters. The obtained principal component data are imported into empirical parameter LSSVM prediction model and the optimized parameter PSO-LSSVM prediction model, respectively, for modeling and prediction, and then comparing the prediction results to analyze the effect of model optimization. The results show that the absolute deviation of the optimized parameter prediction model is less than 9%. And the prediction accuracy of the optimized parameter prediction model is better than that of the empirical parameter model, which can provide a reliable basis for investment decision-making of transmission line projects.

## 1. Introduction

With the rapid economic development, electricity demand has also increased significantly, so the pace of power transmission and transformation project construction is further accelerated, and the investment scale is further expanded. However, if the cost of power transmission and transformation projects can not be effectively controlled, the improvement of cost level will directly affect the economic benefits and construction level of power grid projects. The first task to control the cost of transmission line project effectively is to determine project cost quickly and accurately. Due to the characteristics of large area span, complex construction environment, many unpredictable factors, and nonlinearity of transmission line project, traditional forecasting methods are no longer suitable. 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 transmission line project cost prediction.

In the traditional statistical analysis method, multivariate regression analysis is used to predict the production cost by establishing regression analysis model. Elmousalami HH et al. established a quadratic regression model and determined the key parameters of the model to establish a reliable conceptual cost estimation model for field canal improvement projects (FCIP) [1]. Zhang HB analyzed the cost and elements through the regression analysis of the design parameters, revealed the influence degree of design parameters on the cost and elements of building structure, and obtained the prediction model of the cost and elements of building structure project [2]. However, the regression analysis method is poor in the face of many uncertain factors and less sample data. The number of samples needed by time series analysis method is relatively small, but it can not fully consider the accidental factors which are too numerous and difficult to estimate. Moon S et al. established an autoregressive fractionally integrated moving average (ARFIMA) time series model with long memory characteristics [3]. Moon T et al.

established an interrupted time series forecasting model to predict the construction cost index (CCI) [4]. And based on fuzzy logic, Zhang R et al. improved the prediction accuracy of time series [5].

With the development of artificial intelligence, machine learning and intelligent algorithms have gradually been popular for project cost prediction. Juszczak M introduced the principal component analysis into the neural network to predict the cost of residential buildings in the design stage [6]. Cheng MY et al. used the artificial intelligence method, evolutionary fuzzy neural inference model (EFNIM), to improve the accuracy of cost estimation [7]. Sonmez R proposed a neural network model with bootstrap prediction intervals for range estimation of construction costs [8]. Kim M et al. proposed an approximate cost estimating model for irrigation-type river facility construction at the planning stage, based on case-based reasoning (CBR) with Genetic Algorithms (GA) [9]. Kim S used hybrid analytic hierarchy process (AHP) and case-based reasoning (CBR) to study the cost estimation model in the early stage of Korean highway project [10]. Lesniak A et al. believed that design according to the principle of sustainable development has an impact on construction cost and proposed an approach to estimate the costs of sports field construction using the case-based reasoning method [11]. Rafiei MH et al. proposed a cost estimation model including an unsupervised deep Boltzmann machine (DBM) learning approach and three level back propagation neural network (BPNN) [12]. Juszczak M et al. plotted the relationship between the total cost of construction project and the cost prediction factors of the characteristics of the sports field, and proposed a method for estimating the construction cost of the sports field based on the neural network [13].

With the maturity of forecasting technology, the research on power engineering cost prediction is gradually deepened. Wang FP combines the intelligent learning algorithm with the actual situation of power transmission and transformation projects and proposed a refinement method of control line for power transmission and transformation project cost [14]. Based on Chebyshev inequality, Zhang GL et al. established a prediction model of transmission and transformation project cost, which expanded the dimension of cost analysis [15]. Wang X et al. predicted the static investment of power transmission and transformation project based on Bayesian network, gave the range of probabilistic interval estimation, and provided a reasonable reference for project decision-making [16]. Ling YP et al. proposed a method to predict the construction cost of transmission line based on BP neural network, and established a prediction model of three-layer BP neural network [17]. Lu Y et al. proposed a decomposition-integration cost prediction model and selected different forecasting models to predict different costs considering cost characteristics [18].

Support vector machine (SVM) is based on the statistical learning theory (SLT) established by Vapnik et al. [19]. It has strong generalization ability and outstanding performance in small sample learning and nonlinear problems [20, 21] and is suitable for transmission line project with many unpredictable factors and nonlinear characteristics. Least square

support vector machine (LSSVM) is an improved algorithm of SVM, which can effectively improve the efficiency of machine learning [22]. The LSSVM has been successfully applied to many forecasting problems. Cheng MY et al. established the ELSVM hybrid forecasting model, combined with differential evolution method to optimize the core parameters of LSSVM, and simulated the fluctuation of construction cost index (CCI) of construction project price change [23]. Cong YL et al. proposed a traffic flow forecasting model based on LSSVM with two parameters automatically determined by FOA [24].

The fitting accuracy and generalization ability of LSSVM mainly depend on its two parameters' selection ( $C$  and  $\sigma^2$ ). Therefore, it is very important to use appropriate heuristic algorithm to determine parameter values. General optimization algorithms show low calculation efficiency and poor accuracy [25, 26], and the particle swarm optimization (PSO) algorithm has faster convergence speed and higher accuracy [27], so this paper uses PSO to optimize the parameters of LSSVM.

This paper is organized as follows: Section 2 briefly introduces PCA, LSSVM, and PSO. Section 3 gives the screening and calculation methods of the factors affecting the cost. Section 4 conducts an empirical study to verify the proposed model and Section 5 obtains the conclusion.

## 2. Model Construction

This paper presents the proposal of an approach to the prediction of transmission line project cost which is based on LSSVM. PCA is used to reduce the dimension of the original index to get the corresponding principal component as the input index of the prediction model. And PSO algorithm is used to automatically optimize the parameters of least squares support vector machine to obtain the optimal parameters. The process of model building is shown in Figure 1.

As shown in Figure 1, the construction of the prediction model mainly includes the following four steps:

(1) Select the cost impact indexes and collect the original data of the impact indexes and cost.

(2) PCA is used to reduce the dimension of original data and output new samples. Firstly, the original data are standardized to obtain standardized data; then, the correlation coefficient matrix and its eigenvalues and eigenvectors are calculated based on the standardized data; finally, the variance contribution rate of each principal component is calculated, and the first  $n$  principal components whose variance contribution rate accumulates to the preset value are selected as the new sample output.

(3) Given the empirical parameters of the LSSVM model, an empirical parameter prediction model is constructed.

(4) PSO is used to optimize the parameters of LSSVM. Firstly, the particle velocity and position are initialized, and the current particle fitness is calculated according to the particle velocity and position. The minimum fitness of the current particle is selected as the individual extreme value of the particle, and the minimum global fitness particle is selected as the global extreme value. Then, update the

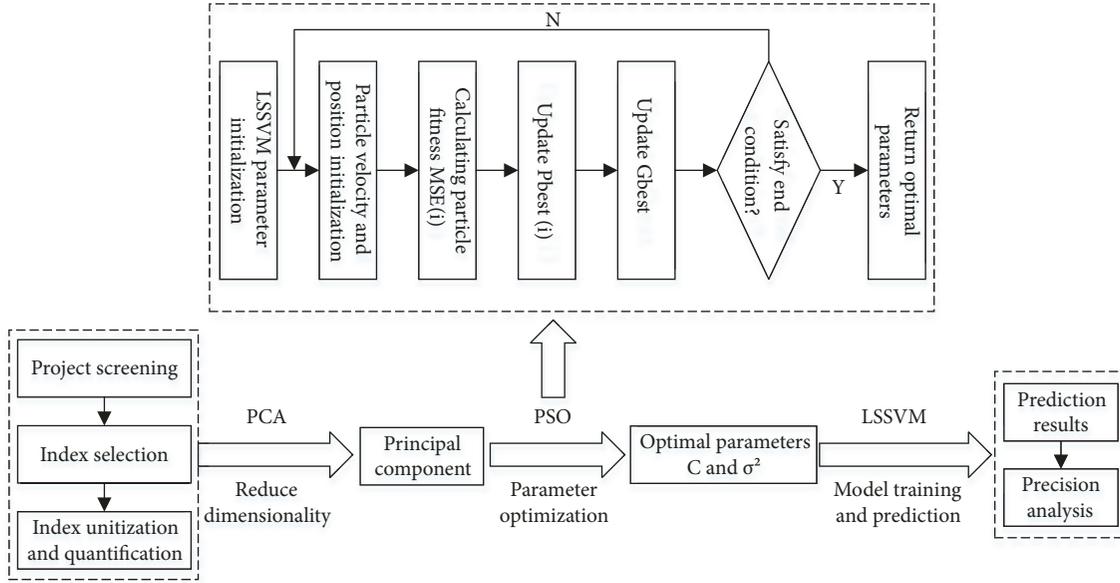


FIGURE 1: Flowchart of transmission line engineering cost prediction model.

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 precision reaches the preset accuracy, that is, the parameters optimized for least squares support vector machine. Finally, an optimized parameter prediction model is constructed according to the optimized parameters.

**2.1. Principal Component Analysis (PCA).** The large number of variables in transmission line project and the complex relationship between them will significantly increase the complexity of the analysis problem. Therefore, this paper uses the principal component analysis method to reduce the dimension of variables, so that the principal components can not only represent the main information of the original variables, but also have no correlation with each other, which will be helpful to the establishment of transmission line project cost prediction model and problem analysis. Principal component analysis (PCA) recombined the  $P$  indicators that had a certain correlation into a new set of interrelated comprehensive indicators to replace the original indicators. The main calculation steps of PCA are as follows:

(1) Standardize the raw data

The original data are standardized according to the following formula:

$$x_{ij}^* = \frac{x_{ij} - \bar{x}_j}{\sqrt{\text{var}(x_j)}}, \quad i = 1, 2, \dots, n; \quad j = 1, 2, \dots, p \quad (1)$$

where  $\bar{x}_j = (1/n) \sum_{i=1}^n x_{ij}$ ,  $\text{var}(x_j) = (1/(n-1)) \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2$  ( $j = 1, 2, \dots, p$ ).

(2) Calculate sample correlation coefficient matrix

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

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ r_{p1} & r_{p2} & \cdots & r_{pp} \end{bmatrix} \quad (2)$$

where  $r_{ij} = \text{cov}(x_i, x_j) / \sqrt{\text{var}(x_i)} \sqrt{\text{var}(x_j)} = \sum_{k=1}^n (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j) / \sqrt{\sum_{k=1}^n (x_{ki} - \bar{x}_i)^2} \sqrt{\sum_{k=1}^n (x_{kj} - \bar{x}_j)^2}$ ,  $n > 1$ .

(3) Calculate the eigenvalues ( $\lambda_1, \lambda_2, \dots, \lambda_p$ ) and corresponding eigenvectors  $a_i = (a_{i1}, a_{i2}, \dots, a_{ip})$ ,  $i = 1, 2, \dots, p$  of the correlation coefficient matrix  $R$

(4) Select the important principal component and get the expression of principal component

Through principal component analysis,  $n$  principal components can be obtained. However, as the variance of each principal component decreases, the amount of information contained therein decreases accordingly. Therefore, in practical analysis, the first  $k$  principal components are usually selected according to the cumulative contribution rate of each principal component (the variance of one principal component accounts for the proportion of the total variance). Generally, the cumulative contribution rate of the first  $k$  principal components is required to be more than 85%.

**2.2. Least Squares Support Vector Machine.** Least squares support vector machine (LSSVM), as an improved algorithm of traditional standard support vector machine, was first proposed by Suykens and Vandewalle [22]. LSSVM inherits a series of excellent characteristics of SVM, such as kernel function, structural risk minimization principle and small sample

size. Based on regularization theory, loss function in SVM is defined as least squares loss function, and equality constraints are used instead of inequality constraints in standard SVM. Through the innovation of SVM objective function setting, the time consuming quadratic programming (QP) problem is transformed into the solving problem of linear equations, which significantly reduces the complexity of the model, reduces the memory consumption in the training process, and greatly improves the solving speed. The following is a brief introduction to the mathematical model of least squares support vector machine.

Given a set of sample vectors of  $n$ -dimensional input and one-dimensional output, the sample under a single technical scheme can be expressed as  $(x_1, y_1)(x_2, y_2), \dots, (x_l, y_l), x_i \in R^n, y_i \in R, i = 1, 2, \dots, l$ . The sample is mapped from the original space to the high-dimensional feature space by a nonlinear mapping  $\varphi(x_i)$ . And the nonlinear estimation function is transformed into the linear estimation function  $f(x) = \omega \cdot \varphi(x) + b$  in the high-dimensional feature space. The weight vector and the offset of the regression function are expressed by  $\omega$  and  $b$  separately. According to the structural risk minimization principle, minimize  $\omega$  and  $b$  as follows:

$$R = \frac{1}{2} \|\omega\|^2 + cR_{emp} \quad (3)$$

where  $\|\omega\|^2$  is used to control the complexity of the model;  $c$  is a regularization parameter to control the penalty degree of the sample exceeding the error;  $R_{emp}$  is the error control function, that is, the insensitive loss function  $\varepsilon$ . LSSVM chooses the square of error  $\xi_i$  as the loss function when optimizing the target, so the optimization problem is as follows:

$$\begin{aligned} \min \quad & \frac{1}{2} \|\omega\|^2 + c \sum_{i=1}^l \xi_i^2 \\ \text{s.t.} \quad & y_i [\omega^T \cdot \varphi(x_i) + b] - 1 + \xi_i = 0 \\ & (i = 1, 2, \dots, l) \end{aligned} \quad (4)$$

where  $\xi_i$  is the relaxation factor. So in order to solve this problem, the Lagrange function is established as follows:

$$\begin{aligned} L(\omega, b, \xi_i, \alpha) = & \frac{1}{2} \|\omega\|^2 + c \sum_{i=1}^l \xi_i^2 \\ & - \sum_{i=1}^l \alpha_i \{y_i [\omega^T \cdot \varphi(x_i) + b] - 1 + \xi_i\} \end{aligned} \quad (5)$$

where  $\alpha_i$  ( $i = 1, 2, \dots, l$ ) is Lagrange multiplier. According to the KKT optimization conditions, the partial derivatives of  $L$  to  $\omega, b, \xi_i$ , and  $\alpha$  are obtained, respectively, and they are equal to 0. The results are as follows:

$$\begin{bmatrix} 0 & Y^T \\ Y & ZZ^T + c^{-1}I \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad (6)$$

Then, an arbitrary symmetric function satisfying Mercer's condition is introduced as the kernel function. The parameters of the kernel function determine the complexity of

the spatial distribution of the sample data and have a great influence on the performance of the model.  $\alpha$  and  $b$  are obtained by least square method. Finally, the decision function of LSSVM regression analysis is obtained as follows:

$$f(x) = \sum_{i=1}^l \alpha_i K(x, x_i) + b \quad (7)$$

**2.3. Particle Swarm Optimization Algorithm.** Particle Swarm Optimization algorithm (PSO) is a global search algorithm, initially inspired by the foraging behavior of birds, and used to solve the global optimal solution of optimization problems [28]. The main idea of PSO algorithm is to initialize the location and velocity of a group of random particles and search for the optimal solution by iteration under certain conditions. The best position of each particle in the search process is defined as the individual extreme value  $P_{best}$ , and the best position of the current population is defined as the global extreme value  $G_{best}$ . In each iteration, the particle updates its speed and position in the next iteration according to the change of the two.

In a  $d$ -dimensional search space, there are  $m$  particles representing possible solutions to the problem  $X = \{X_1, X_2, \dots, X_m\}$ , where  $X_i = \{x_{i1}, x_{i2}, \dots, x_{id}\}$  represents the position of the  $i$  particle,  $d$  is the number of LSSVM parameters, here  $d = 2$ . Individual fitness is expressed by the mean square error generated by each training set sample in LSSVM training, and the fitness function is constructed as follows:

$$\begin{aligned} MSE(\hat{y}_i) &= \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \\ f(x) &= MSE(x) \end{aligned} \quad (8)$$

The velocity of particles in  $D$  dimensional space is defined as  $V_i = \{v_{i1}, v_{i2}, \dots, v_{id}\}$ .  $P_i = \{P_{i1}, P_{i2}, \dots, P_{id}\}$  is the best position  $P_{best}$  (the minimum fitness) that the particle can search for itself.  $P_g = \{P_{g1}, P_{g2}, \dots, P_{gd}\}$  is the best position  $G_{best}$  for the entire population. The velocity and position update of the  $i$  particle is determined according to the following formula:

$$\begin{aligned} v_i^{t+1} &= \omega v_i^t + c_1 \text{rand}() (P_i^t - X_i^t) \\ &+ c_2 \text{rand}() (G_i^t - X_i^t) \\ x_i^{t+1} &= x_i^t + v_i^{t+1} \end{aligned} \quad (9)$$

where  $\omega$  is the weight factor of inertia;  $t$  is the number of iterations;  $c_1$  and  $c_2$  are acceleration factors, representing the step length of particle flying towards its optimal position and overall optimal position;  $\text{rand}()$  is a random number uniformly distributed in the interval  $[0, 1]$ .

If the number of iterations reaches the maximum number of iterations or the precision reaches the preset accuracy, the iteration cycle will be withdrawn and the global optimal parameters will be returned. PSO can be used to optimize kernel function parameters and penalty coefficient in LSSVM

TABLE 1: Influencing indexes of transmission line project cost.

Index type	Influencing indexes	Number
Natural indexes	Topography	X1
	Geology	X2
	Thickness of ice and wind speed	X3
Technical indexes	Single circuit length	X4
	Conductor cross-section	X5
	Conductor consumption	X6
	Tension ratio	X7
	Number of towers	X8
	Tower material consumption	X9
	Earthwork consumption	X10
	Foundation concrete consumption	X11
	Foundation steel consumption	X12
Economic indexes	Price of conductor	X13
	Price of tower material	X14
	Price of foundation steel	X15

model; thus artificial exhaustion can be avoided and better fitting effect can be obtained.

### 3. Selection and Measurement of Cost Prediction Indexes

#### 3.1. Analysis of Factors Affecting Cost and Selection of Indexes.

The transmission line project cost is composed of two parts: ontological project cost and other costs, and other costs only account for about 20% of the project cost. And more than 60% of the other costs are concentrated on the construction site requisition and clearance fees, while the construction site requisition and clearance fees are greatly affected by the region, which is not regular and difficult to accurately predict. Therefore, this paper only studies the ontological cost of transmission line project and does not discuss the changes of other costs. Ontological project cost can be divided into six parts: foundation project cost, tower project cost, grounding project cost, erecting project cost, annex project cost and auxiliary project cost. And the voltage level has a great impact on the cost of transmission line project. The cost of transmission line project with different voltage levels has significant differences. Therefore, this paper only chooses one kind of transmission line project under voltage level for analysis.

The transmission line project consists of six unit projects: foundation project, tower project, grounding project, erecting project, annex project, and auxiliary project. Therefore, this paper originally identifies the influencing factors of the cost from the perspective of the cost of the six unit projects. Firstly, the influencing indexes of each ontological cost module are determined separately, and then the influencing indexes of overall cost are summarized. The process of identifying indexes is shown in Figure 2.

According to the impact indexes identified in Figure 2, they can be classified into three categories: natural, technological, and economic indexes. Among the technical indexes,

because this paper studies the influence indexes of transmission line project cost under single voltage level and the separated number of conductors matches with voltage level, the influence factor of separated number of conductors is eliminated. Among the economic indexes, the price of earthwork and the price of concrete are relatively stable, which have little impact on the cost of transmission line projects, so they are also eliminated. Therefore, the indexes affecting the transmission line project cost are shown in Table 1.

#### 3.2. Index Measurement Method

(1) *Unitization of Gross Index.* In the analysis of the indexes affecting the unit length cost of transmission line project, some gross indexes can not be directly compared because of the different construction scale of the projects. In order to better analyze the impact of these indexes, it is necessary to unit some of the gross indexes as shown in Table 2.

(2) *Quantification of Qualitative Index.* In order to make some qualitative indexes available for the input of prediction model, it is necessary to quantify them according to the factor level of each index. If a project contains multiple levels, weighted average processing should be carried out proportionally. The results of index quantification are shown in Table 3.

## 4. Optimization Model Construction and Experiment Study

4.1. *Project Samples and Data Statistics.* The original data of 78 sets of 110 kV transmission lines in actual settlement in a certain area of China in 2017 are selected as samples. The initial indexes of the model are obtained by quantitative and unit processing of the original data indicators, as shown in Table 4.

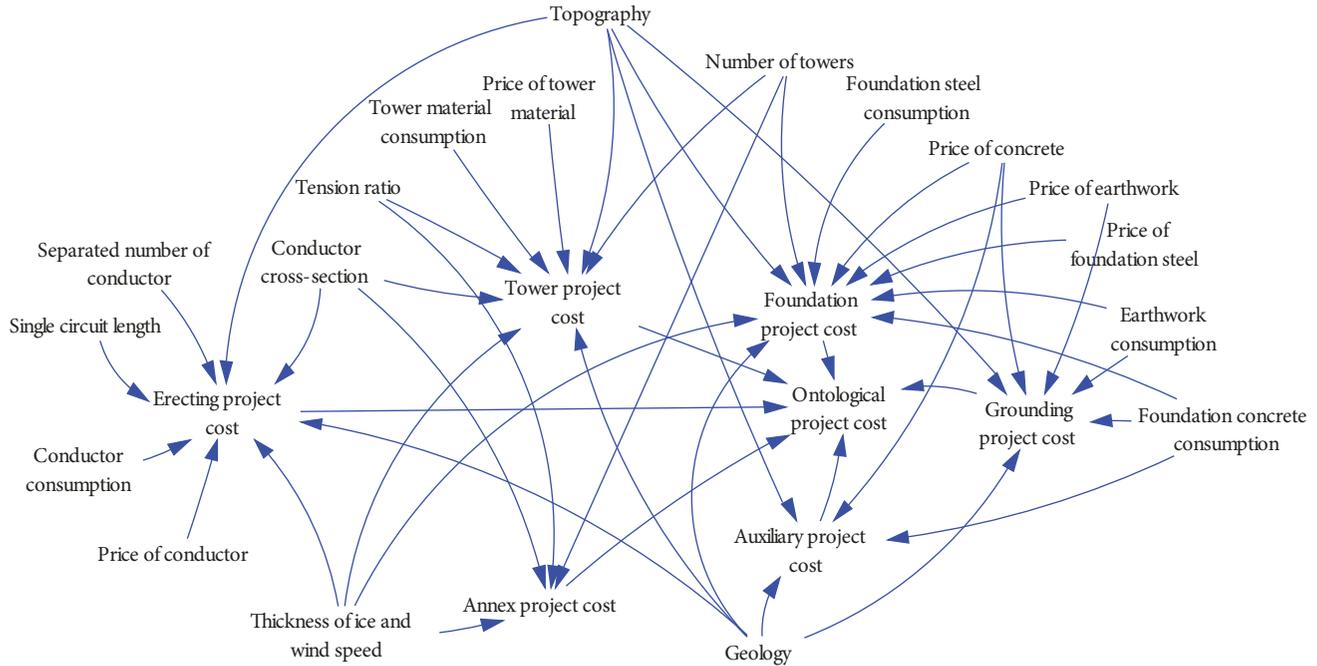


FIGURE 2: Selection of influencing indexes of transmission line project cost.

TABLE 2: Index unitization.

Index	Formula	Unit
Unit length cost	Ontological cost / Single circuit length	10kUSD/km
Unit length tower	Number of towers / Single circuit length	Base/km
Unit tower material consumption	Tower material consumption / Number of towers	t/Base
Unit length conductor consumption	Conductor material consumption / Single circuit length	t/km
Unit foundation steel consumption	Foundation steel consumption / Number of towers	t/Base
Unit foundation concrete consumption	Foundation concrete consumption / Number of towers	m <sup>3</sup> /Base
Unit earthwork consumption	Earthwork consumption / Number of towers	m <sup>3</sup> /Base

In order to eliminate the interaction between indexes and reduce the number of indexes, the principal component analysis (PCA) of the index data of the above 78 sample projects is carried out. Each sample has 15 variables, which constitutes a 78\*15-order matrix:

$$X = \begin{pmatrix} x_{1,1} & \cdots & x_{1,15} \\ \vdots & \ddots & \vdots \\ x_{78,1} & \cdots & x_{78,15} \end{pmatrix} \quad (10)$$

SPSS software was used for calculation, and KMO suitability test and Bartlett sphere test were performed for standardized data. The results showed that KMO value was 0.639 and Bartlett spherical test value was 622.87, which reached a significant level ( $p=0.000 < 0.001$ ), indicating that the data had high efficiency and could be used for principal component analysis.

On the basis of data standardization, the influencing indexes of transmission line project cost are analyzed by principal component analysis. The cumulative contribution of principal component is shown in Table 5. From the table,

the cumulative contribution of variance of the first seven principal components is 86.530%, which is more than 85%. That is to say, the seven principal components can summarize most of the information. Therefore, the first seven principal components are selected as the input set of the transmission line project cost prediction model.

According to the component score coefficient matrix obtained by principal component analysis and the eigenvalues of variables, the principal component coefficient matrix is calculated as shown in Table 6.

The principal components can be calculated by principal component coefficient matrix and standardized variable data. Take the first principal component as an example:

$$\begin{aligned} F_1 = & -0.1297X_1 - 0.0510X_2 - 0.1686X_3 - 0.2460X_4 \\ & + 0.2691X_5 + 0.2278X_6 + 0.3613X_7 \\ & - 0.3620X_8 + 0.3315X_9 + 0.3593X_{10} \\ & + 0.2815X_{11} + 0.3215X_{12} + 0.1932X_{13} \\ & + 0.2157X_{14} - 0.0153X_{15} \end{aligned} \quad (11)$$

TABLE 3: Index quantification.

Topographic coefficient	Flat 1	Hill 2	River network 3	Mire 4	Mountain 5	Alp 6	Desert 7	Steep mountain 8	- -
Geological coefficient	Frozen soil 1	Ordinary soil 2	Hard soil 3	Sand 4	Puddle 5	Mud puddle 6	Drift sand pit 7	Rock blasting 8	Rock artificial 9
Thickness of ice and wind speed coefficient	Ice-free $\leq 29$ 1	Ice-free 31 2	Ice-free 33 3	Ice-free 35 4	Ice-free $\geq 37$ 5	Mild ice 23.5 6	Mild ice 25 7	Mild ice 27 8	Mild ice 29 9
	Mild ice 31 10	Mild ice $\geq 33$ 11	Moderate ice 23.5 12	Moderate ice 25 13	Moderate ice 27 14	Moderate ice 29 15	Moderate ice 31 16	Moderate ice $\geq 33$ 17	- -

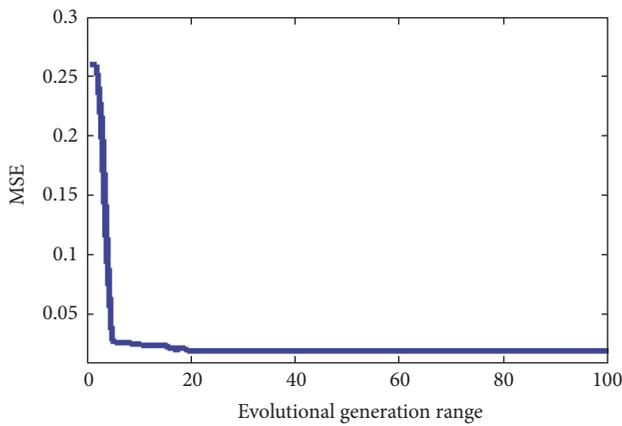


FIGURE 3: Iterative process of parameter optimization. Describe the minimum MSE for each iteration.

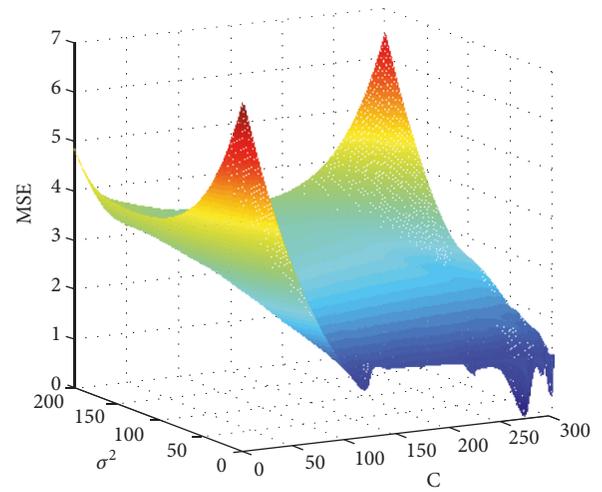


FIGURE 4: Parameter optimization results based on PSO. Best  $C=276.8476$ ,  $\sigma^2 = 0.3412$ .

4.2. *Establishment and Training of Optimized Parameter Model.* Based on the principal component analysis of influencing indexes of transmission line project cost, the calculated seven principal components are used as input parameters of the prediction model, and the unit length cost of transmission line project is used as expected output. 60 groups of sample data were selected as training samples to input LSSVM empirical parameter model and PSO-LSSVM optimization parameter model for model training.

(1) *LSSVM Empirical Parameter Model Training.* A standard LSSVM model is established and then input data into the model for training. Select the empirical parameters  $C=100$  and  $\sigma^2 = 0.4$  as the basic parameters of the LSSVM prediction model: the penalty factor and the kernel function parameter. And select radial basis function as the kernel function  $K(x, x_i)$ :

$$K(x, x_i) = \exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right) \quad (12)$$

(2) *PSO-LSSVM Optimization Parameter Model Training.* The PSO-LSSVM model is used to forecast transmission line

project cost. The number of population is set at 30; the acceleration factor  $c_1$  is 1.5 and  $c_2$  is 1.7; the range of weight factor of inertia  $\omega$  is [0.4, 0.95]; the maximum number of iterations is 300; the radial basis function is also chosen as the kernel function; the range of penalty coefficient  $C$  is [0, 300]; and the range of kernel function parameter  $\sigma^2$  is [0, 200].

The iteration process of PSO is shown in Figure 3, and the result of parameter optimization is shown in Figure 4. The lowest point represents the optimal parameter combination point. The optimal parameters of LSSVM are optimized by PSO algorithm:  $C=276.8476$ ,  $\sigma^2 = 0.3412$ , and the minimum fitness is  $MSE_{min}=0.0188$ .

4.3. *Analysis of Prediction Results.* The other 18 sets of data are selected as test sets to input the model trained above, and the cost prediction analysis of transmission line project is carried out to further verify the rationality of the prediction model. The test samples are substituted into the empirical parameter model and the optimization parameter model, respectively, and the simulation results show that the predicted value of unit length cost is compared with the actual value as shown in Figure 5.

TABLE 4: Original data of transmission line project cost.

Sample	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	Unit length cost (10kUSD/km)
1	0.98	3.98	6.56	20.48	184.98	2.27	0.18	3.15	4.22	57.72	12.78	0.87	1.23	0.60	0.43	6.94
2	1.00	4.53	5.28	11.30	240.00	3.02	0.45	1.77	10.11	153.11	21.75	1.77	1.32	0.67	0.25	8.26
3	1.53	3.85	4.48	22.18	239.98	2.78	0.39	1.51	9.15	119.54	22.74	2.12	1.30	0.66	0.45	5.42
4	1.00	4.00	4.46	13.29	240.00	2.94	0.27	2.26	21.55	108.33	41.44	1.82	1.20	0.61	0.28	8.80
5	1.00	2.00	4.50	20.00	240.00	2.90	0.23	2.35	8.07	81.24	30.52	0.71	1.32	0.58	0.56	10.59
6	1.00	2.56	4.73	8.79	240.00	3.01	0.38	1.82	6.99	64.85	16.77	1.23	1.29	0.62	0.11	5.54
7	1.00	4.34	3.81	21.76	240.00	2.94	0.45	1.93	7.44	161.12	33.39	2.00	1.29	0.65	0.37	6.50
8	1.50	3.55	7.00	71.72	185.00	2.30	0.19	3.37	3.92	60.96	13.09	0.72	1.25	0.62	0.32	6.34
...	...	...	...	...	...	...	...	...	...	...	...	...	.....	...	...	...
78	0.98	2.83	7.98	27.95	239.98	2.88	0.14	2.84	4.76	51.78	12.21	0.51	1.28	0.60	0.31	5.19

TABLE 5: Total variance explained.

Component	Initial Eigenvalues			Extraction Sum of Squared Loading		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5.103	34.019	34.019	5.103	34.019	34.019
2	2.093	13.954	47.973	2.093	13.954	47.973
3	1.936	12.906	60.880	1.936	12.906	60.880
4	1.434	9.557	70.437	1.434	9.557	70.437
5	0.941	6.272	76.709	0.941	6.272	76.709
6	0.807	5.378	82.087	0.807	5.378	82.087
7	0.667	4.443	86.530	0.667	4.443	86.530
8	0.541	3.607	90.137	-	-	-
9	0.388	2.588	92.725	-	-	-
10	0.348	2.319	95.044	-	-	-
11	0.253	1.688	96.732	-	-	-
12	0.212	1.415	98.146	-	-	-
13	0.116	0.775	98.921	-	-	-
14	0.103	0.687	99.608	-	-	-
15	0.059	0.392	100.000	-	-	-

TABLE 6: The coefficient between the principal component and the canonical variables.

	F1	F2	F3	F4	F5	F6	F7
X1	-0.1297	0.5014	0.3511	0.0871	-0.1248	-0.1359	0.0836
X2	-0.0510	0.4369	0.4027	0.2805	0.2123	-0.2092	-0.0392
X3	-0.1686	0.1471	-0.3002	0.2621	0.6817	0.1478	-0.0749
X4	-0.2460	0.2493	0.2965	-0.2906	-0.1893	0.3707	0.0480
X5	0.2691	0.2513	-0.3165	0.3022	-0.1994	0.0933	0.1723
X6	0.2278	0.1963	-0.2009	0.4954	-0.3483	0.0453	0.1816
X7	0.3613	-0.0397	0.0048	-0.0173	0.1011	-0.1750	-0.3387
X8	-0.3620	0.0305	-0.0857	0.1595	0.2706	0.1497	0.1956
X9	0.3315	-0.0967	0.1303	0.0478	0.0762	0.0971	0.4066
X10	0.3593	0.0418	0.2828	-0.0941	0.0824	0.2517	-0.1884
X11	0.2815	-0.1016	0.3283	-0.0001	0.3189	0.1754	0.4340
X12	0.3215	-0.0581	0.0845	-0.0636	0.2444	-0.2249	0.0460
X13	0.1932	0.3201	-0.2440	-0.2722	0.0494	0.6135	-0.1008
X14	0.2157	0.4296	-0.1638	-0.2254	0.1187	-0.2493	-0.2922
X15	-0.0153	0.2407	-0.3007	-0.5079	0.0611	-0.3495	0.5302

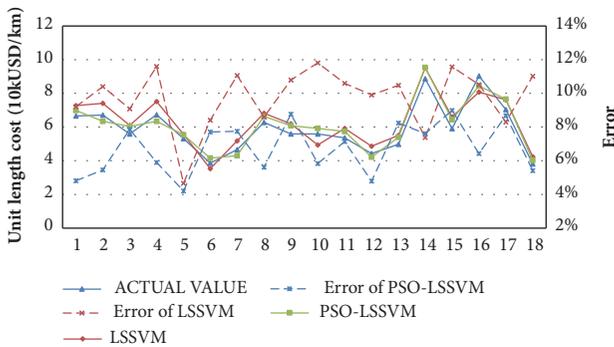


FIGURE 5: Comparison of prediction results of different models.

As shown in Figure 5, the LSSVM empirical parameter model predicts that the average absolute deviation rate between the cost and the real cost is 9.74%; the PSO-LSSVM

optimized parameter model predicts that the average absolute deviation between the cost and the real cost is 6.73%. The prediction accuracy of the optimized parameter model is higher than that of the empirical parameter model, and it can meet the accuracy requirement of project investment decision.

Statistically, the mean absolute percentage error (MAPE) and root mean square error (RMSE) are used to measure the accuracy of the prediction model. The calculation formulas are as follows:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{\hat{y}_i} \right| \quad (13)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (14)$$

TABLE 7: Statistical error measures of prediction methods.

Prediction models	Indexes	
	MAPE	RMSE
LSSVM	9.74%	1.55
PSO-LSSVM	6.73%	1.12

where  $y_i$  represents the predicted value and  $\hat{y}_i$  represents the actual value. The MAPE values and RMSE values of the LSSVM model and the PSO-LSSVM model are calculated, respectively, as shown in Table 7.

Table 7 shows that the MAPE of PSO-LSSVM is 30.90% higher than that of LSSVM, which proves that the optimization parameter model can greatly improve the accuracy of transmission line cost prediction model.

## 5. Conclusions

This paper proposes an intelligent cost prediction model based on LSSVM optimized by PSO. Firstly, PCA is used to screen and reduce the dimension of the transmission line 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 LSSVM model which is mature in theory are given to determine the nonlinear mapping between transmission line 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 transmission line project. And the accuracy of the model is verified by comparing the cost predict value of empirical parameter model, optimized parameter model, and actual value of cost.

Cost control of transmission line 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 transmission line project, and the error is generally high. For example, the neural network model and the regression prediction model can not adapt to the characteristics of fewer samples of transmission line engineering, and the prediction model based on grey theory can not adapt to the characteristics of many influencing factors of transmission line engineering, so the error generated are rarely less than 10%. The average absolute deviation rate of the PSO-LSSVM optimization parameter model is 6.73% compared with the actual value of the cost, and the accuracy of the PSO-LSSVM optimization parameter model is improved greatly compared with other prediction models.

In summary, the PSO-LSSVM cost prediction model proposed in this paper can effectively improve the accuracy of transmission line project cost prediction. This model has strong practical significance and good application effect in transmission line project cost prediction.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

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

## Acknowledgments

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

## References

- [1] H. H. ElMousalami, A. H. Elyamany, and A. H. Ibrahim, "Predicting Conceptual Cost for Field Canal Improvement Projects," *Journal of Construction Engineering and Management*, vol. 144, no. 11, 8 pages, 2018.
- [2] H. B. Zhang, "Research on Structural Cost Analysis at Schematic Design Stage using Multi-variant Regression Analysis," *Engineering Cost Management*, vol. 2, pp. 49–52, 2015.
- [3] S. Moon, S. Chi, and D. Y. Kim, "Predicting Construction Cost Index Using the Autoregressive Fractionally Integrated Moving Average Model," *Journal of Management in Engineering*, vol. 34, no. 2, 2018.
- [4] T. Moon and D. H. Shin, "Forecasting Construction Cost Index Using Interrupted Time-Series," *KSCE Journal of Civil Engineering*, vol. 22, no. 5, pp. 1626–1633, 2018.
- [5] R. Zhang, B. Ashuri, and Y. Deng, "A novel method for forecasting time series based on fuzzy logic and visibility graph," *Advances in Data Analysis and Classification. ADAC*, vol. 11, no. 4, pp. 759–783, 2017.
- [6] M. Juszczak, "Application of PCA-based data compression in the ANN-supported conceptual cost estimation of residential buildings," in *Proceedings of the International Conference on Numerical Analysis and Applied Mathematics 2015*, p. 200007, Rhodes, Greece.
- [7] M.-Y. Cheng, H.-C. Tsai, and W.-S. Hsieh, "Web-based conceptual cost estimates for construction projects using Evolutionary Fuzzy Neural Inference Model," *Automation in Construction*, vol. 18, no. 2, pp. 164–172, 2009.
- [8] R. Sonmez, "Range estimation of construction costs using neural networks with bootstrap prediction intervals," *Expert Systems with Applications*, vol. 38, no. 8, pp. 9913–9917, 2011.
- [9] M. Kim, S. Lee, S. Woo, and D. H. Shin, "Approximate cost estimating model for river facility construction based on case-based reasoning with genetic algorithms," *KSCE Journal of Civil Engineering*, vol. 16, no. 3, pp. 283–292, 2012.
- [10] S. Kim, "Hybrid forecasting system based on case-based reasoning and analytic hierarchy process for cost estimation," *Journal of Civil Engineering and Management*, vol. 19, no. 1, pp. 86–96, 2013.
- [11] A. Leśniak and K. Zima, "Cost Calculation of Construction Projects Including Sustainability Factors Using the Case Based Reasoning (CBR) Method," *Sustainability*, vol. 10, no. 5, p. 1608, 2018.

- [12] M. H. Rafiei and H. Adeli, "Novel Machine-Learning Model for Estimating Construction Costs Considering Economic Variables and Indexes," *Journal of Construction Engineering and Management*, vol. 144, no. 12, p. 9, 2018.
- [13] M. Juszczak, A. Leśniak, and K. Zima, "ANN Based Approach for Estimation of Construction Costs of Sports Fields," *Complexity*, vol. 2018, Article ID 7952434, 11 pages, 2018.
- [14] F. P. Wang, *Subdividing Method and Applied Research of Power Transmission and Transformation Project Cost Control*, Chongqing University, China Chongqing, 2015.
- [15] G. L. Zhang, W. N. Wen, and P. D. Li, "Reasonable Interval Calculation Method for Power Transmission Project Cost Based on Chebyshev's Inequality," *Electric Power Construction*, vol. 35, no. 9, pp. 118–122, 2014.
- [16] X. Wang, P. Y. Geng, and B. B. Yin, "Bayesian Network Based Static Investment Forecast for Transmission and Transformation Projects," *China Power Enterprise Management*, vol. 19, pp. 93–96, 2015.
- [17] Y. P. Ling, P. F. Yan, and C. Z. Han, "BP Neural Network Based Cost Prediction Model for Transmission Projects," *Electric Power*, vol. 45, no. 10, pp. 95–99, 2012.
- [18] Y. Lu, D. Niu, J. Qiu, and W. Liu, "Prediction Technology of Power Transmission and Transformation Project Cost Based on the Decomposition-Integration," *Mathematical Problems in Engineering*, vol. 2015, Article ID 651878, 11 pages, 2015.
- [19] V. N. Vapnik, *Statistical Learning Theory*, John Wiley, New York, 1998.
- [20] C. Aldrich and L. Auret, *Statistical Learning Theory and Kernel-based Methods, Unsupervised Process Monitoring and Fault Diagnosis with Machine Learning Methods*, Springer, London, UK, 2013.
- [21] L. N. Jiang, *Research on the Predict of the Construction Cost Based on Support Vector Machine*, Hebei University of Engineering, China Handan, 2009.
- [22] J. A. K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Processing Letters*, vol. 9, no. 3, pp. 293–300, 1999.
- [23] 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.
- [24] Y. Cong, J. Wang, and X. Li, "Traffic flow forecasting by a least squares support vector machine with a fruit fly optimization algorithm," *Procedia Engineering*, vol. 137, pp. 59–68, 2016.
- [25] Q. Wu, "Hybrid model based on wavelet support vector machine and modified genetic algorithm penalizing Gaussian noises for power load forecasts," *Expert Systems with Applications*, vol. 38, no. 1, pp. 379–385, 2011.
- [26] F.-C. Yuan and C.-H. Lee, "Using least square support vector regression with genetic algorithm to forecast beta systematic risk," *Journal of Computational Science*, vol. 11, pp. 26–33, 2015.
- [27] R. G. Gorjaei, R. Songolzadeh, M. Torkaman, M. Safari, and G. Zargar, "A novel PSO-LSSVM model for predicting liquid rate of two phase flow through wellhead chokes," *Journal of Natural Gas Science and Engineering*, vol. 24, pp. 228–237, 2015.
- [28] R. C. Eberhart and J. Kennedy, "A new optimizer using particle swarm theory," in *Proceedings of the 6th International Symposium on Micromachine and Human Science*, pp. 39–43, Nagoya, Japan, October 1995.




**Hindawi**

Submit your manuscripts at  
[www.hindawi.com](http://www.hindawi.com)

