Tech Science Press

Highway Cost Prediction Based on LSSVM Optimized by Intial Parameters

Xueqing Wang[1], Shuang Liu[1,*] and Lejun Zhang[2]

[1]School of Mechanics and Civil Engineering, China University of Mining and Technology-Beijing, Beijing, 100083, China
[2]School of Information Engineering, Yangzhou University, Yangzhou, 225127, China
*Corresponding Author: Shuang Liu. Email: liushuang_0122@163.com

Abstract: The cost of highway is affected by many factors. Its composition and calculation are complicated and have great ambiguity. Calculating the cost of highway according to the traditional highway engineering estimation method is a completely tedious task. Constructing a highway cost prediction model can forecast the value promptly and improve the accuracy of highway engineering cost. This work sorts out and collects 60 sets of measured data of highway engineering; establishes an expressway cost index system based on 10 factors, including main route mileage, roadbed width, roadbed earthwork, and number of bridges; and processes the data through principal component analysis (PCA) and hierarchical cluster analysis. Particle swarm optimization (PSO) is used to obtain the optimal parameter combination of the regularization parameter c and the kernel function width coefficient σ in least squares support vector machine (LSSVM). Results show that the average relative and mean square errors of the PCA-PSO-LSSVM model are 0.79% and 10.01%, respectively. Compared with BP neural networks and unoptimized LSSVM model, the PCA-PSO-LSSVM model has smaller relative errors, better generalization ability, and higher prediction accuracy, thereby providing a new method for highway cost prediction in complex environments.

Keywords: Highway; least squares support vector machine (LSSVM); particle swarm optimization (PSO); principal component analysis (PCA); hierarchical cluster analysis

1 Introduction

According to traditional highway engineering estimation method, calculating its cost is an extremely perplexed task. With the rapid development of mathematical modeling methods and computer technology, experts at home and abroad have studied various mathematical models or computer simulation means for project cost forecasting. Regression analysis methods were commonly used [1] in the early foreign literature and were later combined with other probability analysis model [2]. In recent years, artificial neural network-based cost prediction approaches have become prevalent. Domestic scholars have applied methods, such as fuzzy mathematics [3], grey system theory [4], genetic algorithm [5], system dynamics [6], and big data [7], for the cost prediction of engineering projects.

A large number of documents apply BP neural network [8] for cost prediction. Owing to the slow convergence speed, these documents are liable to fall into a local minimum. Support vector machine (SVM) has excellent learning ability and can be used for small sample size, thereby avoiding structure selection, and the local minima of the neural network. SVM has elicited extensive attention for in-depth study. SVM has several problems. First, its algorithm setting parameters are based on empirical values. Second, its implementation is complicated and difficult. Lastly, it has slow training speed. The least squares support vector machine (LSSVM), as an improved SVM algorithm, inherits a series of excellent features, such as the SVM kernel function, the principle of structural risk minimization, and small sample size. Complex quadratic programming problem is transformed into a simpler linear equation solving problem, which shortens training time and improves solution speed greatly [9].

Particle swarm optimization (PSO) algorithm uses real numbers to find the optimal parameters. The algorithm has strong versatility, fast convergence, and is easier to leap to local optimal information. It has been widely used in parameter optimization. Consequently, the PSO algorithm is used to determine the optimal parameters of LSSVM and improve calculation accuracy [10].

Through preliminary research on the aforementioned algorithms, this work sorts out and collects the data of existing highways, establishes a sample set, processes the samples through hierarchical cluster analysis and principal component analysis (PCA), builds a PCA-PSO-LSSVM [11] highway engineering prediction model, and compares the proposed model with the BP neural network and the unoptimized LSSVM model.

2 Basic Principle of PCA-PSO-LSSVM

2.1 PCA

PCA is an index dimensionality reduction method based on mathematical ideas. It uses the orthogonal transformation in linear programming to reduce the given variables with correlation to a small number of uncorrelated comprehensive variables. These new comprehensive variables carry most of the important information of the original indicators, and the relationship of complex matrix is simplified to achieve the dimensionality reduction of indicators [12]. The specific steps are presented as follows:

Step 1: Select the initial sample. Assuming that population X has n samples $(X_1, X_2, \cdots, X_n)$, and each sample has m-dimensional variables. Thus, the matrix of the observation data is denoted as:

$$X_{m \times n} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix}$$

Step 2: Standardize the original data. The formula is expressed as follows:

$$e_{\cdot j} = \frac{x_{\cdot j} - \hat{x}_{\cdot j}}{S_{\cdot j}} \tag{1}$$

where

$x_{\cdot j}$: j is a random variable;

$\hat{x}_{\cdot j}$: mean of the jth variable;

$S_{\cdot j}$: standard deviation of the jth variable.

Step 3: Calculate the correlation coefficient matrix of $e = (e_1, e_2, \cdots, e_n)^T$ and use $\sum u = \lambda u$ to find the eigenvalue λ_i and its eigenvector u_i. $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_m \geq 0$.

Step 4: Obtain M $(m \leq n)$ principal components by calculation:

$$
\begin{cases}
F_1 = u_{11}X_1 + u_{12}X_2 + \cdots + u_{1n}X_n \\
F_2 = u_{21}X_1 + u_{22}X_2 + \cdots + u_{2n}X_n \\
\quad\quad\quad\quad \vdots \\
F_m = u_{m1}X_1 + u_{m2}X_2 + \cdots + u_{mn}X_n
\end{cases}
\tag{2}
$$

$$
u_{i1}{}^2 + u_{i2}{}^2 + \cdots + u_{in}{}^2 = 1 (i = 1, 2, \cdots, m)
\tag{3}
$$

Step 5: Calculate the principal component contribution rate and cumulative contribution rate. Compute the contribution rate of the ith principal component according to $P_i = \dfrac{\lambda_i}{\sum_{j=1}^{m} \lambda_j}$. The cumulative contribution rate of the first q principal component is $P_i = \dfrac{\sum_{i=1}^{q} \lambda_i}{\sum_{i=1}^{m} \lambda_i}$. When the cumulative contribution rate of the current q principal component reaches over 85%, the first q principal component is used as a new indicator.

2.2 PSO

Kennedy and Eberhart proposed PSO in 1995. This algorithm has the advantages of simplicity, easy implementation, no gradient information, and few parameters. It is particularly suitable for real number optimization problems. It also has a profound intelligent background that is suitable for scientific research, particularly for engineering applications [13]. The main principles are presented as follows:

M particles are found in the D-dimensional space; Particle i position: $x_i = (x_{i1}, x_{i2}, \ldots, x_{iD})$; Particle i velocity: $v_i = (v_{i1}, v_{i2}, \ldots, v_{iD})$, $1 \leq i \leq m$; and the best position in history that particle i has experienced: $p_i = (p_{i1}, p_{i2}, \ldots, p_{iD})$.

$$
v_{im}^{t+1} = \omega v_{im}^t + c_1 rand_1 \left(P_{im}^t - x_{im}^t \right) + c_2 rand_2 \left(P_{gm}^t - x_{im}^t \right)
\tag{4}
$$

$$
x_{im}^{t+1} = x_{im}^t + v_{im}^{t+1}
\tag{5}
$$

where

ω: inertia weight factor;

c_1, c_2: learning factors, usually a value of 2;

$rand$: [0,1] random function of value;

t: number of iterations.

2.3 LSSVM

The main principle of the mathematical model of the LSSVM regression algorithm is presented as follows. The training sample set $D = \{(x_i, y_i), i = 1, 2, \cdots, n\}$, where $x_i \in R^d$ is the ith d-dimensional input vector, and $y_i \in R$ is the predicted value of the corresponding input, is given. Subsequently, the regression function is:

$$
y_i = w\phi(x_i) + b
\tag{6}
$$

where

ω: weight vector;

b: offset.

Different from SVM, LSSVM selects the square of the error ξ_i as the loss function in the optimization objective while changing the constraints into equality constraints. When using the principle of structural risk minimization, the optimization problem becomes:

$$min\frac{1}{2}\parallel\omega\parallel^2 + c\frac{1}{2}\sum_{i=1}^{n}e_i^2$$
$$st. w\phi(x_i) + b + e_i = y_i \tag{7}$$

where

  c: regularization parameters;

  e_i: error vector.

The Lagrangian function is established to solve the above-mentioned problem:

$$L(w,b,e,\xi) = min\frac{1}{2}\parallel\omega\parallel^2 + c\frac{1}{2}\sum_{i=1}^{n}e_i^2 - \sum_{i=1}^{n}\xi_i[w\phi(x_i) + b + e_i - y_i] \tag{8}$$

The optimal solution satisfies the KKT optimization condition, and the partial derivatives of w, b, e, ξ in Eq. (8) are calculated and are equal to zero.

$$\begin{cases}\frac{\partial L}{\partial w} = 0 \rightarrow w = \sum_{i=1}^{n}\xi_i\phi(x_i)\\\frac{\partial L}{\partial b} = 0 \rightarrow \sum_{i=1}^{n}\xi_i = 0\\\frac{\partial L}{\partial e} = 0 \rightarrow \xi_i = \gamma e_i\\\frac{\partial L}{\partial \xi} = 0 \rightarrow w\phi(x_i) + b + e_i - y_i = 0\end{cases} \tag{9}$$

After transforming the above-mentionedconditions using the same solution, variables ω and e are eliminated, and the optimal solution matrix of b and ξ can be obtained.

$$\begin{bmatrix}0\\Y\end{bmatrix} = \begin{bmatrix}0 & Z^T\\Z & K + c^{-1}E\end{bmatrix}\begin{bmatrix}b\\\xi\end{bmatrix} \tag{10}$$

where

  $\xi : \xi = [\xi_1, \xi_2, \cdots, \xi_n]^T$, Lagrange multiplier;

  $Z : Z = [1, 1, \cdots, 1]^T$;

  $Y : Y = (y_1, y_2, \cdots, y_n)^T$;

  $E :$ n-order identity matrix;

  $K : K = K(x_i, x_j) = \phi(x_i)\phi(x_j)$, kernel function matrix.

The final decision function of LSSVM is:

$$y(x) = \sum_{i=1}^{n}\xi_i K(x, x_i) + b \tag{11}$$

The kernel function adopts the Gaussian radial basis kernel function and is expressed as:

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

3 PSO-LSSVM Model Based on PCA

The PSO algorithm is used to determine the optimal solution of the key parameters c and σ of LSSVM and build the PCA-PSO-LSSVM highway engineering cost prediction model. The specific flow chart is shown in Fig. 1.
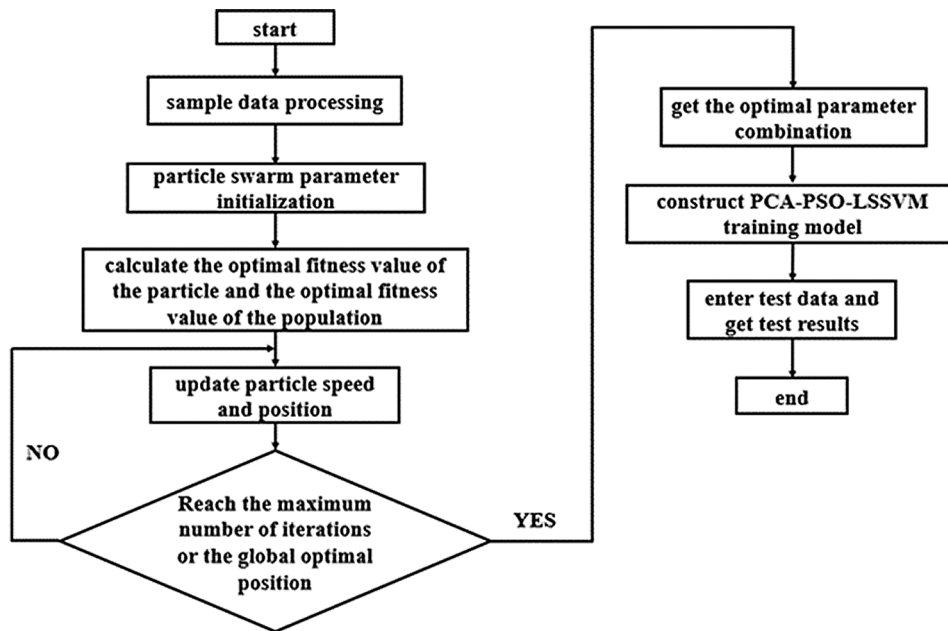

Figure 1: Flow chart of PCA-PSO-LSSVM model implementation

The steps, which are based on the PCA-PSO-LSSVM model, are presented as follows:

Step 1: Sort and collect samples and perform systematic cluster and principal component analyses on the data.

Step 2: Initialize the particle swarm. The regularization parameter c and the kernel function width coefficient σ in the LSSVM model must be optimized. Set the value range of (c, σ) given that the number of particle swarms q, the maximum number of iteration t_{max}, learning factors c_1 and c_2, and inertial weighting factors ω_{max} and ω_{min}. Generate the first-generation particle swarm randomly.

Step 3: Train the generated parameter combinations of each generation c and σ as the parameters of the LSSVM model. Calculate the fitness value of each particle swarm generation through the fitness function, and select the root mean square error (MSE) as the function to evaluate the fitness of the particles.

Step 4: Compare the current fitness value $f(x_i)$ of each particle with the fitness value $f(P_{best_i})$ of the historical optimal position. If $f(x_i) < f(P_{best_i})$, then update $P_{best_i} = x_i$. Compare the fitness value $f(x_i)$ of the optimal position of each particle with the optimal position fitness value $f(G_{best})$ of the entire particle swarm. If $f(x_i) < f(G_{best})$, then update $G_{best} = x_i$. Continue these steps until the optimal solution combination is achieved.

Step 5: Construct the PCA-PSO-LSSVM training model, the fitness graph, and the sample regression curve figure.

Step 6: Input the test sample and obtain the prediction result.

4 Application and Analysis

4.1 Selection of Model Evaluation Indicators

Sorting out and collecting 60 groups of highway data in different regions, the main factors that affect highway project cost, namely, main route mileage X_1/km, subgrade width X_2/m, subgrade earthwork volume $X_3/10^3(\text{m}^3 \cdot \text{km}^{-1})$, number of bridges $X_4/\text{m} \cdot \text{km}^{-1}$, number of interchanges $X_5/\text{road} \cdot \text{km}^{-1}$, number of separated interchanges $X_6/\text{place} \cdot \text{km}^{-1}$, number of tunnels $X_7/\text{m} \cdot \text{km}^{-1}$, pavement form X_8, landform features X_9, and area X_{10}. The predicted value refers to the highway engineering cost per kilometer: $Y/10$ *million yuan*. The pavement form is determined according to different pavement forms, landform characteristics, and the degree of influence of the area on the construction cost of expressway. The values 0.8 and 0.6 represent the asphalt and cement concrete pavements, respectively. The geomorphic features are presented as follows: 0.2 represents plain and hilly area, 0.5 represents heavy hill area, and 0.8 represents mountainous area. Weighted summation is used when different sections of a road have diverse geomorphic features. In the region, China's provinces are divided into I, II, and III taking 0.3, 0.6, and 0.9, respectively.

4.2 Sample Data Processing

First, hierarchical cluster analysis is used to classify the samples, and several projects with higher similarity can be selected to improve prediction accuracy. A total of 60 groups of highway engineering data are standardized in the SPSS software (Tab. 1). The clustering method selects clustering between groups, and the measurement interval uses square European clustering.

Table 1: Standardization of original data of highway construction

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	Y
1	−1.49427	2.05733	−0.57764	−0.33949	6.66921	1.38474	−0.86545	0.388	−1.28160	1.98326	5.115842
2	−1.41891	5.60477	0.73310	3.57918	−0.30064	−1.13625	−0.86724	0.388	−1.10422	−0.49582	11.32167
3	−1.22782	2.06324	0.35260	0.24037	0.28477	1.27352	−0.86622	0.388	−1.28160	−0.49582	3.72445
4	−0.9374	0.29247	0.35261	−0.70322	0.45566	1.01169	−0.86549	0.388	−1.28160	−0.49582	3.62909
5	−0.92144	0.29837	−1.00874	−0.70322	−0.30336	0.53900	−0.86587	0.388	−1.28160	1.98326	3.89426
6	−0.85565	−0.44535	−0.17051	0.95306	−0.30264	−0.60322	3.09739	0.388	1.37916	−0.49582	8.89118
7	−0.77088	−0.44535	0.55095	1.18386	−0.30246	−0.87488	0.70101	0.388	1.11309	−0.49582	8.224901
…	…	…	…	…	…	…	…	…	…	…	…
60	0.40435	−0.44535	1.72603	0.71118	−0.30267	0.15205	0.6807	0.388	0.49224	1.98326	8.78174

After hierarchical cluster analysis, the 10 sets of data (e.g., 1, 2, 43, 15, 29, 23, 28, 27, 36, and 16) were screened out, and the remaining 50 sets of data were standardized to obtain the data in Tab. 2. The characteristic value and cumulative contribution rate of each component were obtained through PCA (Tab. 3). The first 6 factors with a cumulative contribution rate of 85% were selected as the new principal components. The coefficient matrix (Tab. 4) is acquired according to the *coefficient = component load ÷ $\sqrt{eigenvalue}$*. Finally, by using formula $Z_1 = -0.0366X_1 - 0.29835X_2 + 0.20789X_3 + 0.46073X_4 - 0.25574X_5 - 0.32196X_6 + 0.48084X_7 - 0.13289X_8 + 0.47095X_9 + 0.12073X_{10}$ and so on, the input sample matrix is obtained (Tab. 5).

Table 2: Standardization sample data

	X$_1$	X$_2$	X$_3$	X$_4$	X$_5$	X$_6$	X$_7$	X$_8$	X$_9$	X$_{10}$	Y
1	−1.3900	3.5456	0.4081	0.3647	0.8043	2.1664	−0.9197	0.2021	−1.4069	−0.4321	3.72445
2	−1.0321	0.7406	0.4081	−0.8649	1.1534	1.7735	−0.9190	0.2021	−1.4069	−0.4321	3.62909
3	−1.0124	0.7410	−1.0432	−0.8649	−0.3969	1.0641	−0.9194	0.2021	−1.4069	2.2683	3.89426
4	−0.9314	−0.4281	−0.1499	1.234	−0.3542	−0.6502	2.8866	0.2021	1.2678	−0.4321	8.89118
5	−0.8269	−0.4281	0.6199	1.5942	−0.3951	−1.0578	0.5853	0.2021	1.0003	−0.4321	8.224901
6	−0.7716	−0.4281	−0.1570	1.0171	0.3951	−1.1075	1.2213	0.2021	0.9112	−0.4321	9.680698
7	−0.7195	−0.4235	−0.1750	−1.2092	−0.3968	0.3478	−0.9184	0.2021	0.5993	−0.4321	4.39933
8	−0.6843	−0.4281	−0.0703	1.6030	−0.3951	−1.3109	1.4166	0.2021	−1.4069	−0.4321	6.79503
9	−0.3762	−0.4281	0.3579	−1.1465	0.3969	−0.1773	−0.9190	0.2021	−1.4069	−0.4321	4.46452
10	−0.1870	−0.1944	−1.1110	−1.1634	0.3969	−0.4068	−0.9192	0.2021	−1.4069	−0.4321	4.07081
…	…	…	…	…	…	…	…	…	…	…	…
50	0.6215	−0.4281	1.8723	0.9782	−0.3955	0.4834	0.5658	0.2021	0.3763	2.26826	8.78174

Table 3: Eigenvalue, contribution rate, and cumulative contribution rate

Ingredient	Eigenvalues	Contribution rate/%	Cumulative contribution rate/%
1	2.774	27.744	27.744
2	1.520	15.197	42.941
3	1.476	14.756	57.697
4	1.225	12.250	69.947
5	0.811	8.106	78.052
6	0.724	7.243	85.296
7	0.548	5.475	90.771
8	0.450	4.501	95.271
9	0.297	2.970	98.241
10	0.176	1.759	100

Table 4: Coefficient matrix

	1	2	3	4	5	6
X$_1$	−0.03660	0.24245	0.60285	−0.16337	0.53826	−0.15261
X$_2$	−0.29835	−0.14402	−0.37438	0.44428	0.13241	0.00973
X$_3$	0.20789	0.37533	0.22249	0.48797	−0.20796	0.52819
X$_4$	0.46073	−0.02943	−0.03134	0.41220	−0.11068	−0.26665
X$_5$	−0.25574	0.18524	0.44497	0.25020	−0.49556	−0.43301
X$_6$	−0.32196	0.18671	−0.05762	0.47734	0.51664	−0.16687
X$_7$	0.48084	−0.24122	−0.04715	0.18820	0.26038	−0.27967
X$_8$	−0.13289	−0.50348	0.36945	0.17009	0.12506	0.52798
X$_9$	0.47095	−0.03821	0.16807	−0.00255	0.14270	0.04872
X$_{10}$	0.12073	0.62947	−0.27820	−0.11432	0.14287	0.22860

Table 5: Input sample matrix

	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Y
1	−2.84115	−0.24979	−1.85119	3.301447	−0.15993	−0.14825	3.72445
2	−2.54631	0.268266	−0.36882	1.389823	−0.57841	0.011945	3.62909
3	−1.90084	1.007213	−2.08352	−0.35267	0.522654	0.649525	3.89426
4	2.943145	−1.57143	−0.34164	0.636575	0.190138	−0.7438	8.89118
5	2.136404	−0.77696	−0.02976	0.49193	−0.89906	0.264725	8.224901
6	1.986912	−1.19738	−0.19329	−0.03775	−0.51672	−0.1739	9.680698
7	−0.68971	−0.32717	−0.13401	−0.67825	−0.08684	0.743458	4.39933
8	2.515087	−1.25283	−0.10724	0.171007	−0.58139	−0.30968	6.79503
9	−0.39249	−0.14302	0.221569	−0.70136	−0.29185	1.043547	4.46452
10	−1.65322	−0.64778	−0.4021	−1.45658	−0.25661	0.186208	4.07081
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
50	1.586654	2.078496	0.197199	1.037475	0.776276	1.204973	8.78174

4.3 PCA-PSO-LSSVM Prediction Model

The PCA-PSO-LSSVM prediction model is established using the MATLAB2016(a) simulation platform, and the initialization parameters of the prediction model are set as follows: population size $q = 40$, maximum number of iterations $t_{max} = 500$, learning factor $c_1 = 2$, $c_2 = 2$, inertia weight coefficient $\omega_{max} = 0.9$, $\omega_{min} = 0.4$, regularization parameters $c \in [0, 100]$, and kernel function width coefficient $\sigma \in [0, 10000]$. The first 40 groups of the input sample data are applied as the training samples to exercise and learn the PCA-PSO-LSSVM model, and the last 10 groups are utilized as the test samples for prediction. The output is the cost of highway engineering per kilometer/10 million yuan. The fitness curve of the PCA-PSO-LSSVM model is shown in Fig. 2.
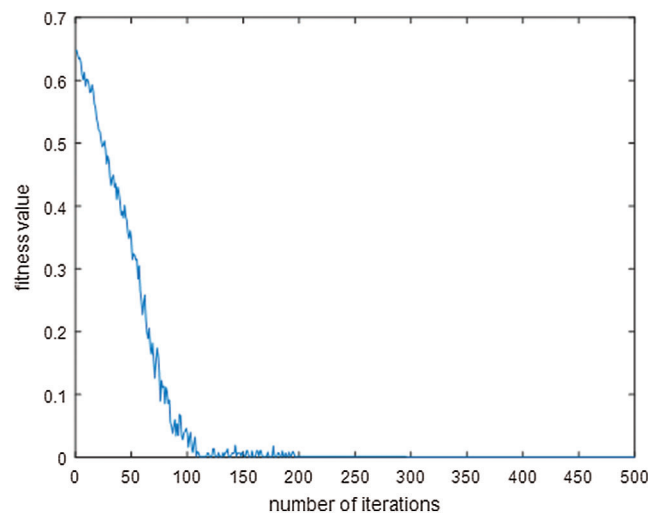

Figure 2: Fitness function diagram

Fig. 2 shows that the fitness curves have reached a stable state when the number of iterations reaches 210. The optimal parameter combination of the prediction model is $(c, \sigma) = (0.0434, 3547.4806)$, and the average relative error of the training sample is $MRE = 0.0017$. The sample regression curve with good fitting effect is shown in Fig. 3.
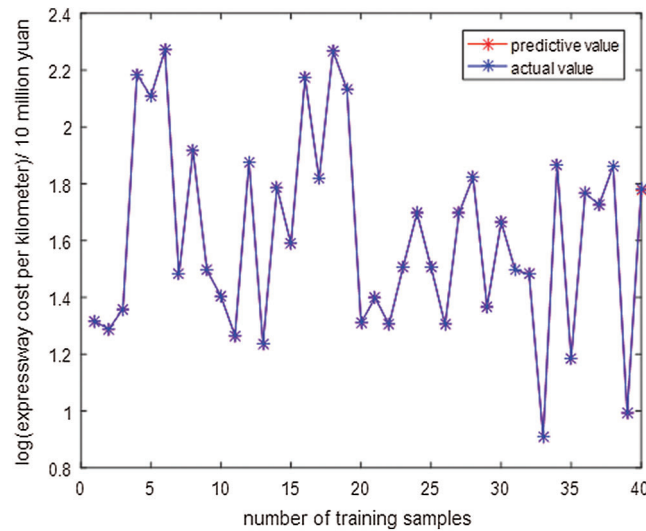

Figure 3: Regression curve of highway engineering cost training sample

4.4 Comparative Analysis with BP neural network and LSSVM model

The regression fitting of the training samples proves that the PCA-PSO-LSSVM model has good learning ability. To verify whether the model also has excellent generalization ability, the prediction is performed by inputting 10 sets of test sample data and by comparing them with the unoptimized LSSVM model and BP neural network model (Fig. 4).
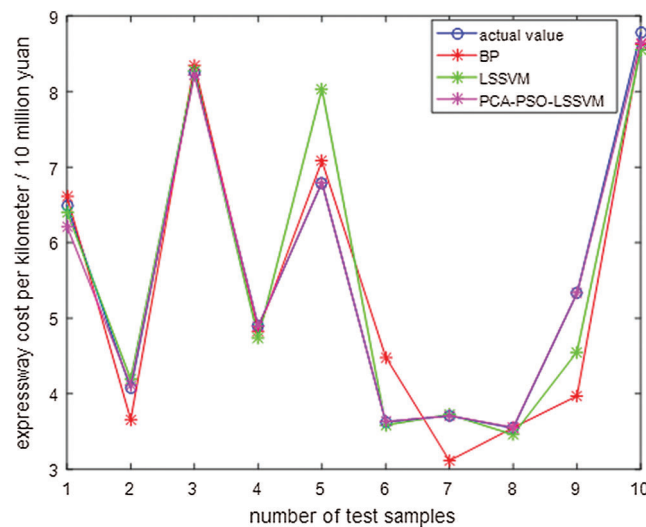

Figure 4: Forecast results of highway engineering cost by different models

Preliminarily, Fig. 4 shows that the effect of the PCA-PSO-LSSVM model prediction is better than those of the BP neural network and the LSSVM model, which have values closest to the actual one. To verify the superiority of the PCA-PSO-LSSVM model more intuitively, the average relative error (MRE) and root mean square relative error (RMSE) are calculated to evaluate the performance of the model (Tabs. 6 and 7, respectively).

Table 6: Comparison of the relative errors of the three prediction models

Output variable	Actual value	PCA-PSO-LSSVM model		LSSVM model		BP neural network	
		Predictive value	Relative error/%	Predictive value	Relative error/%	Predictive value	Relative error/%
Highway cost/ 10 million yuan · km^{-1}	6.49211	6.21329	4.2947	6.41092	1.2506	6.61376	1.8738
	4.06781	4.11126	1.0681	4.19473	3.1202	3.65821	10.0693
	8.260201	8.21326	0.5682	8.26570	0.0665	8.34940	1.0798
	4.89237	4.89725	0.0998	4.74419	3.0287	4.83629	1.1463
	6.79466	6.79510	0.0065	8.03372	18.2358	7.07926	4.1886
	3.62897	3.63017	0.0331	3.58667	1.1657	4.48618	23.6212
	3.70859	3.70905	0.0123	3.72601	0.4698	3.11486	16.0097
	3.54613	3.55405	0.2232	3.45936	2.4470	3.55175	0.1585
	5.3321	5.32938	0.0511	4.55616	14.5522	3.97144	25.5182
	8.78174	8.64644	1.5407	8.55743	2.5543	8.62063	1.8346

Table 7: Comparison of evaluation indexes of the three models

Predictive model	Unit cost	
	MRE	RMSE
BP neural network	8.55%	56.92%
LSSVM model	4.69%	47.35%
PCA-PSO-LSSVM model	0.79%	10.01%

Tabs. 6 and 7 suggest that the accuracy of the BP neural network for highway project cost prediction is poor with an average relative error and root mean square relative error of 8.55% and 56.92%, respectively. The reason is that the BP neural network needs to rely on large sample data, which have poor generalization ability for small sample learning. Meanwhile, the average relative error and root mean square relative error of the unoptimized LSSVM model are 4.69% and 47.35%, which are more accurate than the BP neural network prediction. The PCA-PSO-LSSVM model has an average relative error and root mean square relative error of 0.79% and 10.01%, respectively. Through comparative analysis, the MRE and RMSE of the PCA-PSO-LSSVM model are the smallest. Thus, this model can predict the cost of highway engineering more accurately.

5 Conclusions

Based on the principal component analysis method, the least squares support vector machine prediction model is established. It combined with the PSO algorithm to optimize the regularization parameter c and the kernel function width coefficient σ in LSSVM. Overcome the fact that the traditional LSSVM model determines the parameters through experience, thereby resulting in a lower prediction accuracy.

Through the predictive analysis of highway engineering, the PCA-PSO-LSSVM model has the average relative error of 0.79% and the root mean square relative error of 10.01%. Compared with the BP neural network and the unoptimized LSSVM model, the PCA-PSO-LSSVM model has better learning generalization ability and prediction accuracy.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1]   A. Ashworth and M. Skitmore, "Accuracy in estimating chartered quantity surveyor," London, 1982.

[2]   A. H. Boussahaine and T. M. S. Elhag, "Tender price estimation using ANN methods," *Research Rep. No. 3 School of Architecture*, 1999.

[3]   J. X. Yang and H. Y. Xie, "The application of fuzzy neural network in highway engineering cost estimation," *Journal of China & Foreign Highway*, vol. 27, no. 5, pp. 16–19, 2007.

[4]   H. K. Duan, "Research on highway engineering cost forecast model based on GN-BP," *New Technology and New Process*, no. 3, pp. 28–31, 2017.

[5]   Y. H. Pan, Y. L. Zhang and Y. J. Cai, "Research on highway engineering cost estimation based on GA-BP algorithm," *Journal of Chongqing Jiaotong University (Natural Science Edition)*, vol. 35, no. 2, pp. 141–145, 2016.

[6]   Y. E. Geng, "Analysis of the influencing factors and relationship of highway engineering cost based on system dynamics," *Jiangxi Building Materials*, no. 5, pp. 112–114, 2015.

[7]   C. X. Jiang, "Research on cost control of large real estate companies based on big data," M.S. dissertation, University of Shandong Jianzhu, Jinan, 2015.

[8]   R. Wang, "Determination of influencing factors for road cost prediction based on extended BP network," *Shandong Transportation Science and Technology*, no. 3, pp. 29–31, 2019.

[9]   S. Wang, "Research on construction cost prediction based on particle swarm optimization least square support vector machine," M.S. dissertation, Qingdao University of Science and Technology, Qingdao, 2017.

[10] Z. Liu, B. Xiang, Y. Q. Song, H. Lu and Q. F. Liu, "An improved unsupervised image segmentation method based on multi-objective particle swarm optimization clustering algorithm," *Computers, Materials & Continua*, vol. 58, no. 2, pp. 451–461, 2019.

[11] S. C. Feng, L. S. Shao and W. J. Lu, "Application of PCA-PSO-LSSVM model in gas emission prediction,' *Journal of Liaoning Technical University (Natural Science Edition)*, vol. 38, no. 2, pp. 124–129, 2019.

[12] C. S. Yuan, X. T. Li, Q. M. Jonathan Wu, J. Li and X. M. Sun, "Fingerprint liveness detection from different fingerprint materials using convolutional neural network and principal component analysis," *Computers, Materials & Continua*, vol. 53, no. 4, pp. 357–372, 2017.

[13] Y. Yang, "Establishment of PSO-LSSVM based on distribution network project cost forecast model and its error analysis," *Automation Technology and Application*, vol. 39, no. 2, pp. 98–102, 2020.