

Research on the Prediction Model of Material Cost Based on Data Mining

Liu Shenyang^{*1}, Gao Qi¹, Li Zhen¹, Li Si¹ and Li Zhiwei^{1,2}

¹Department of Equipment Command & Management of Mechanical Engineering College, Shijiazhuang, Hebei, 050003, P.R. China; ²Maintenance Depot of Shijiazhuang Flying College, Shijiazhuang, Hebei, 050003, P.R. China

Abstract: Material cost prediction should be based on the scientific mathematical models so that the influence of subjective factors on the quota and other indicators of decomposition can be reduced. This paper analyzes the particle swarm optimization (PSO) algorithm to optimize the parameters of support vector machine and establishes the prediction model of material cost after preprocessing the actual data and uses the support vector regression (SVR) machine to carry out data mining. In the forecasting process, the total cost of material is first predicted and the predicted results are then adjusted with the actual value, and finally, the relative errors are tested. The result indicates that the forecasting effect is fulfilled.

Keywords: Data mining, optimization algorithm, particle swarm prediction of material cost, support vector machine.

1. INTRODUCTION

The prediction of material cost refers to utilization of special methods to estimate and predict the level of material cost on the basis of historical data and relevant information. Its characteristic feature is the prediction of the future on the basis of history and prediction of unknown level on the basis of known information. The renewal of statistical data and a series of characteristics of material production of enterprises determine that the sequence formed by material cost data generally has the non-stationarity, nonlinearity and the point of abrupt change. The support vector machine can register the minimization of structural risk and the error rate on testing data (namely generalization error rate) of learning machines by taking the sum of training error rate and a dependent item as boundary. A specific characteristic of the support vector machine is that it does not utilize the field having internal problems, which provides a good generalization performance in situations of pattern classification.

This paper carries out data mining to establish the prediction model of material cost according to the learning method of support vector regression machine. This method is a type of relational schema between the spatial pattern of learning input and functional mapping of learning output and researchers generally consider the function set in this type of mapping relation as a learning machine. It starts with the research and observation of the data (namely the sample). Researchers adopt some rules that cannot be obtained from principle in the current situations and meanwhile utilize these rules to analyze the data obtained. Thus, after reaching the point of value prediction, they conduct the decision making and value estimation processes.

2. PSO-SVR MODELING ALGORITHM

In order to estimate the optimal computation of the support vector regression machine (SVR) parameter, the advanced optimization algorithm is introduced to SVR algorithm which is a hotspot in the support vector machine (SVM) field [1, 2].

SVM is an effective new method that can be utilized to conduct research on data mining. The three types of kernel functions are shown in Table 1.

Table 1. The three types of kernel functions.

Type of Kernel Function	Formula
Multinomial Function	$k(x, z) = \exp[(x \cdot x_i) + 1]^q$
RBF Function	$k(x, z) = \exp\left\{-\frac{ x - x_i ^2}{\sigma^2}\right\}$
Sigmoid Function	$k(x, z) = \tanh[v(x, x_i) + c]$

In this paper, the model of support vector regression machine adopted the RBF function as a kernel function.

At present, the genetic algorithm is most widely used in optimization of algorithms but its operation such as choice, cross, variation, etc., is more complicated and its rate of convergence and precision is certainly limited when it conducts high dimension samples. Comparatively, the particle swarm optimization algorithm is a type of optimization algorithm which is based on group parallel and global search and is simpler than other algorithms and its parameter setting is relatively less and does not provide much solution for the problems. In addition, this algorithm exhibits faster rate of convergence and stronger global searching ability. Hence, the model is easy to operate and

*Address correspondence to this author at the Department of Equipment Command & Management of Mechanical Engineering College, Shijiazhuang, Hebei, 050003, P.R. China; Tel: 15127134769; E-mail: 1037279749@qq.com

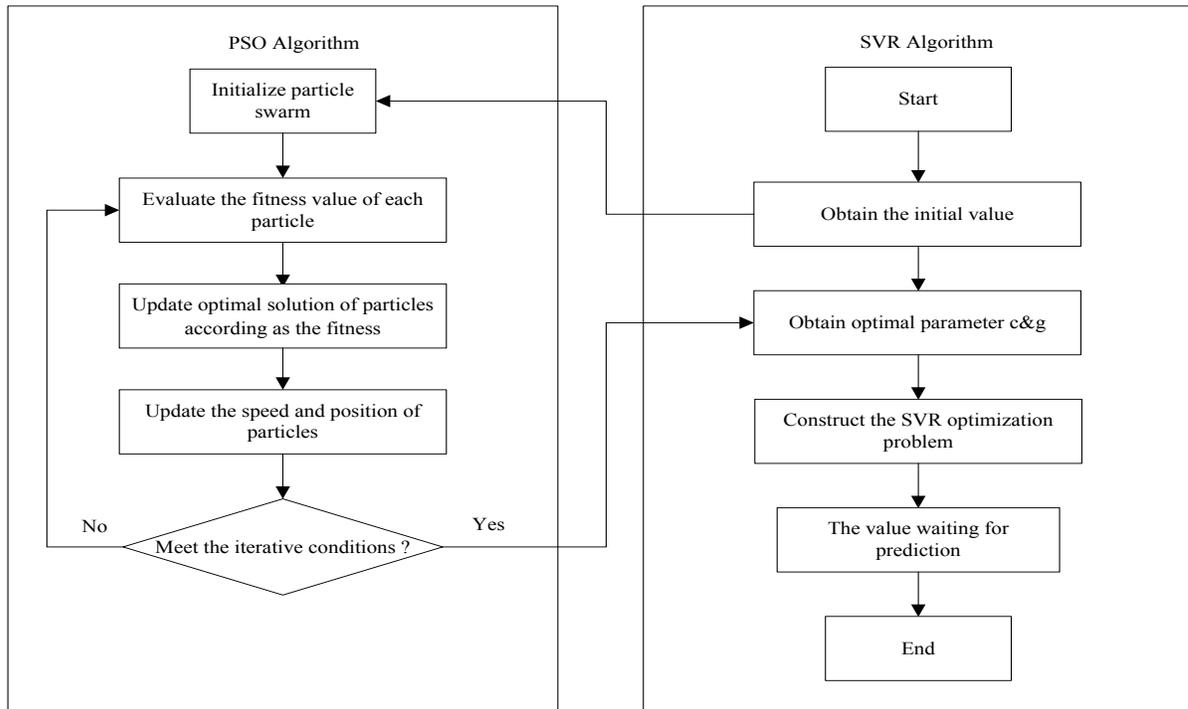


Fig. (1). Flow chart of PSO-SVR algorithm.

can easily be generalized [3-5]. In this paper, the global optimization search is conducted for the sum of parameters in the algorithm model of support vector regression machine and the optimal penalty coefficient and kernel parameter obtained from search are taken as the parameters of the final model.

2.1. Parameters of Selection Optimization

Optimizing the parameters g and c of the kernel function proposed simultaneously and carrying out automatic selection of these parameters can provide a reference idea for solving the parameter optimization problems of the support vector machine. This paper mainly optimized the kernel parameters g and c of support vector regression machine (namely penalty factor), whereby, it can be popularized in other parameter optimizing problems of SVR.

2.2. Fitness Function Design of Particle Swarm Optimization

The selection of appropriate fitness function is very important for the particle swarm optimization algorithm. The target of selection features is to try to use a small number of features to obtain the same or better classifying effect when designing the fitness function. Therefore, the number of selected features should be considered in the evaluation of a fitness function. Specifically, the sample with fewer features will have high fitness if the accuracy rate of two feature samples is the same [6, 7].

2.3. Operating Steps for Optimizing SVR Parameter by Using PSO Algorithm

The flow chart of PSO-SVR algorithm is shown in Fig. (1).

1. The particle swarms g and c are initialized first including the confirmation of the group size and setting of the location and speed of each particle. Then after presetting the inertia weight of PSO, the maximum iterations are set.
2. The optimal solution of each particle is set as the current location of particles. The fitness value of each particle is calculated and the optimal solution of particles with maximum fitness is taken as the optimal solution of the current group.
3. The position coordinate and speed of particles are updated.
4. The fitness function is used to evaluate the fitness of particles.
5. For each particle, its current fitness value is compared with the optimal solution of the current individual, and if the former is better, the optimal value of the individual shall be replaced by the fitness value.
6. For each particle, its current fitness value is compared with the optimal solution of the current group, and if the former is better, it can be regarded as the current global optimum.
7. The process is terminated if the iterative conditions are met or else it is skipped on to Step 3.
8. The test sample is predicted by using the well-trained support vector regression machine.
9. The optimal parameter combination (namely penalty coefficient c and kernel parameter g) is used to substitute into the programmed support vector regression machine and then the data of test sample is established for model prediction.

3. INPUT AND OUTPUT OF MODEL

In order to avoid the influence to the model caused by different dimensions of each data and improve the training speed, the matrix of data is subjected to normalization processing and the input and output data

is limited within [0, 1]. The normalized mapping adopted in this paper is as follows:

$$y = (y_{\max} - y_{\min}) \times \frac{x - x_{\min}}{x_{\max} - x_{\min}} + y_{\min}$$

Where, x is the original data, and x_{\max} and x_{\min} are the maximum and minimum values of the original data respectively. y_{\min} and y_{\max} are range parameters of the mapping. Accordingly, the needed predicting value is obtained from negated normalization of data after the prediction is finished.

4. TRAINING AND TEST OF MODEL

This paper adopted the algorithm of particle swarm optimization to obtain the optimal path and the relative optimal value [8].

Table 2. Material cost and actual data of correlative factors.

No	Material Cost	Material Unit Price	Equipment Runtime	Degree of Mechanization
	(Yuan)	(Yuan)	(Hour)	%
1	396252	9900	1104	79.5
2	494461	10000	1082	80.1
3	392997	10000	1003	81.3
4	409790	11548	1104	78.4
5	437172	10000	1167	77.5
6	487557	10003	1007	83.2
7	430350	10337	1120	80.5
8	386098	10000	1040	77.5
9	405947	10641	1050	79.4
10	394035	10090	1110	82.5
11	407125	10000	1077	83.2
12	447905	10002	1118	81.7
13	385545	10102	1190	80.5
14	414508	10380	1251	84.1
15	387878	10450	1200	82.5
16	377847	10500	1091	83.6
17	468784	13517	1057	80.3
18	362679	10233	950	79.8
19	474944	10503	1210	81.2
20	465280	10670	1047	82.4
21	385689	10500	1253	80.3
22	428563	11102	1210	84.6
23	451182	10680	1041	85.4
24	444048	10600	1108	82.5
25	379868	10363	1110	83.4
26	483312	11780	1032	84.5
27	396275	10855	1124	82.9
28	412365	10085	1109	83.7
29	480553	95000	1234	80.4
30	391849	10439	1052	82.4

The steps for the algorithm of particle swarm optimization are as follows:

1. The parameter optimization for the established SVR prediction model is conducted by using PSO and also some parameters of PSO algorithm are initialized.
2. Selection of the learning factors: the local searching ability of parameter is 1.5, belonging to [0, 2] and the global searching ability of parameter is 1.7, belonging to [0, 2].
3. The size of particle swarm is 20. The maximum and minimum of penalty coefficients c are 100 and 0, respectively. The maximum and minimum of kernel parameter g are also 100 and 0, respectively. The maximum number of iterations is 200.

In the modeling process, two processing programs (svmtrain and svmpredict) are mainly used. In the training process, the input parameters should be adjusted repeatedly to obtain the optimal results.

5. APPLICATION OF MODEL

The time series for research in this paper can be obtained based on all indicators data. Through the time series, a series of prediction values according to time sequence can be obtained and different time series can also be obtained based on different objects of study or problems. For prediction of material cost, this paper adopts the data series of the total material cost on account of months as the historical data to predict the future value of a certain variable. The material cost and actual data of correlative factors are shown in Table 2.

Since the data has different dimensions under different influence factors, it can have different economic meanings and explanations. In order to avoid the different influence degrees to data caused by different dimensions, and for considering the difference existing between variable data and unit set in this paper, the normalization processing is conducted on the data. In the prediction process of total cost of material, 30 sets of data are extracted from the historical data which are true and effective. The data is between 0 and 1 after normalization processing. In this way, the significant influence caused by different dimensions of data can be avoided in estimation process of the support vector regression machine. For evaluating the effectiveness and source of data, firstly the data is divided into training data and test data. The former 25 data are chosen as the training set, and latter 5 data as test set. The management of material cost not only controls the total material cost but also controls the key components among it.

5.1. Parameter Optimization

According to the steps of particle swarm optimization parameters, the optimal parameters of support vector regression machine are obtained for predicting the total cost of material: $c=43.22$, $g=0.01$. The minimum mean square error of cross validation in parameter selection is $MSE=0.0102$ gives better optimization effect.

5.2. Training and Prediction

The model training and prediction of support vector regression machine are conducted by using the optimal parameters obtained from particle swarm optimization with the correlation coefficient being $R=98.63\%$. The fitting between original data and prediction data is carried out after negated normalization. The fitting results of material cost prediction are shown in Fig. (2).

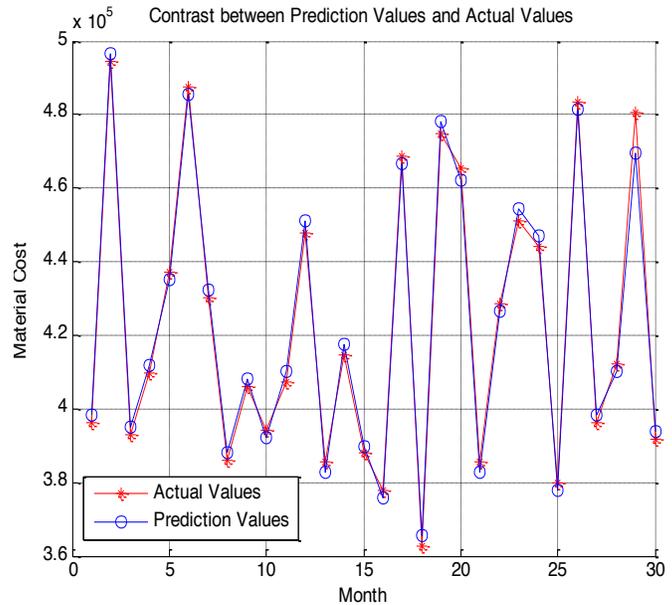


Fig. (2). Fitting results of material cost prediction.

5.3. Model Evaluation

For the purpose of illustrating that the support vector regression machine of particle swarm optimization parameters has excellent performance in the prediction of material cost, the relative errors of each data are calculated. The prediction effect is better when the relative error of prediction is basically within 0.005. The relative errors of material cost prediction are shown in Fig. (3).

The results are calculated by means of calculating the same data in combination of BP neural network and multiple regression method. BP neural network is a three-layer structure and the unit number of hidden layer can be set to 6 through a contrast test. Material cost prediction results by using three methods are shown in Table 3. It is observed that the maximum relative error in prediction model of support vector regression machine established in the thesis is 2.32%, the maximum relative error of BP neural network is 5.52%, and the maximum relative error of multiple regressions is 14.93%. The results from three prediction models show that the best effect in particle swarm optimization algorithm is of the prediction model of support vector regression machine.

Table 3. Material cost prediction results by using different prediction methods.

Model	No	Actual Value	Prediction Value	Relative Error
PSO-SVR	28	412365	410330	-0.49%
	29	480553	469400	-2.32%
	30	391849	393770	0.49%
BP Neural Network	28	412365	394780	-4.26%
	29	480553	459830	-4.31%
	30	391849	413490	5.52%
Multiple Regression	28	412365	473930	14.93%
	29	480553	453780	-5.57%
	30	391849	425630	8.62%

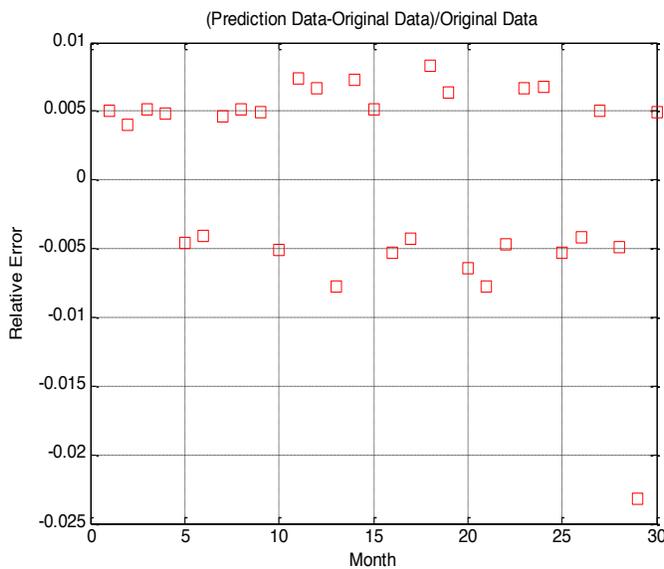


Fig. (3). Relative errors of material cost prediction.

CONCLUSION

The performance optimization of support vector machine mainly reflects the optimization of parameters, and in this respect, the particle swarm optimization algorithm possesses the characteristic of faster rate of convergence and stronger global searching ability. Therefore, this paper established the prediction model of the support vector regression machine in particle swarm optimization algorithm and described the modeling algorithm, input and output, training and test in detail. The paper has also provided the determining standard and establishment method of parameters as well while effectively estimating the prediction of material cost with higher accuracy.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

Declared none.

REFERENCES

- [1] W. Bao, Q. Hu, and C. Wang, "Support vector machine based on multi-granulations", *Journal of Nanjing University (Natural Sciences)*, vol. 49, pp. 637-643, 2013.
- [2] S. Pan, W. Wang, and H. Guo, "An incremental support vector machine approach based on probability density distribution", *Journal of Nanjing University (Natural Sciences)*, vol. 49, pp. 603-610, 2013.
- [3] J. Liu, X. Chen, Q. Liu, and J. Sun, "LS-SVM based on improved PSO for prediction of satellite clock error", *Journal of Astronautics*, vol. 34, pp. 1509-1515, 2013.
- [4] L. Peng, R. Niu, Y. Zhao, and Q. Deng, "Prediction of landslide displacement based on KPCA and PSO-SVR", *Geomatics and Information Science of Wuhan University*, vol. 38, pp. 148-152, 2013.
- [5] Y. Cheng, X. Liang, and Y. Huang, "Improved quantum particle swarm optimization based on good-point set", *Journal of Central South University (Science and Technology)*, vol. 44, pp. 1409-1414, 2013.
- [6] S. Gu, J. Wei, and D. Pi, "Particle swarm optimization-fuzzy support vector machine based prediction of spacecraft parameters", *Journal of Astronautics*, vol. 35, pp. 1270-1276, 2014.
- [7] Z. Chen, Y. Bo, P. Wu, G. Song, and W. Duan, "Novel particle filtering based on adaptive particle swarm optimization and its application", *Journal of Applied Sciences Electronics and Information Engineering*, vol. 31, pp. 285-293, 2013.
- [8] L. Shang, and Q. Tan, "Emergency classification based on support vector machine", *Journal of Industrial Engineering /Engineering Management*, vol. 28, pp. 119-123, 2014.