Two-stage grey support vector machine prediction model

Huaping Zhou*, Yue Yuan

Faculty of Computer Science & Engineering Anhui University of Science and Technology, HuaiNan, China

Received 1 October 2014, www.cmnt.lv

Abstract

Two-stage grey support vector machine prediction model (D_2GM-SVM) is put forward by analysing the grey model GM , support vector machine model (SVM) and one-stage grey support vector machine prediction model((DGM-SVM). The prediction accuracy of grey model is improved through two-stage buffer operators D2 to predict the various relevant indexes. At the same time, genetic algorithm is used to find the optimal parameters of the support vector machine model, RBF kernel parameter and penalty parameter, which are the optimal parameters (c, g). Thus, the regression model of the optimal support vector machine is determined. Finally, the final output value is predicted by inputting the predictive value of each index into the support vector machine model. The results show D2GM-SVM has a higher prediction accuracy compared with grey prediction model , BP neural network prediction model and DGM-SVM in this case, and that grey forecasting model combined with the support vector machine model has practical value in solving practical prediction problems.

Keywords: Gary prediction model, Support vector machine, integrated forecasting, Genetic algorithm

1 Introduction

In many areas, computer technology, mathematical methods and theoretical models are used to predict the future state of a given period based on historical data. With a summarization of the extensive literature, there are currently used prediction theories such as multivariate linear (nonlinear) regression prediction[1], trend prediction[2], Markov prediction [3] and grey prediction[4]. Although there're quite a lot of prediction methods, each method predicts from a different point of view with a different scope of application. Under a certain specific application background, the choice of method becomes very difficult, leading to a more serious prediction error.

Aiming at the situation, Bates, J.M. and Granger C.W.J. proposed a new forecasting method – combination forecast [5-8] for the first time in 1969 in their paper. The purpose of combination forecast method was to reduce the prediction error of a single prediction model and improve the accuracy of prediction with a kind of optimal combination. But in the combination forecast model, it was difficult to determine the weights of each single prediction model. So this kind of combination forecast model highlighted significant limitations in practical application. On the other hand, these prediction methods were all based on the known data of the same index. In practical applications, the predicted values were often closely related to many objecttive factors, as the outcome of combined action of various factors. The forecast value affected by objective factors should be considered in prediction.

Support vector machine (SVM) had rapid research and development as early as the end of the last century by Vapnik [9, 10] and other professionals. It had the perfect theoretical basis. Support vector machine was successfully applied to solve the prediction problem.

Literature[11] choses the most relevant parameters of the maximum depth, the maximum tangential stress and the uniaxial compressive strength as the input vector of the support vector machine to classify the long-term rock burst underground caverns. Literature [12] choses the sun set, relative humidity and wind speed as the input vector of the prediction model and gets the support vector machine forecasting model applied to solar power short-term forecast. Literature [13] choses such parameters as the groundwater depth, irrigation water capacity and evaporation capacity and adopt support vector machine model for dynamic prediction soil conductivity.

Literature [14] used support vector machine model to predict the sea surface temperature. Literature [15] adopted support vector machine to predict air quality parameters. Literature [16] combined with fuzzy c – average clustering algorithm, support vector machine model and least square method to forecast the NOx emission of coal burning boiler. The multistage support vector machine model was proposed and applied in the clinical charge distribution prediction of patients with chronic diseases in Literature [17]. The combination model of least squares and support vector machine was proposed to predict the concentration of CO in air quality [18]. The partial least squares and support vector machine were combined to construct the linear and nonlinear model to predict the quantitative structure-activity relationship of the distribution behaviour of the blood flowing to the brain [19].

In these studies, the input values of the support vector machine are either based on the single historical value of one dimensional time series, or are the multi-dimensional matrix of multiple affecting factors related to predictive value. However, either the establishment stage or the prediction stage of the model is based on the situation under which the input matrix is known or can be measured. In practical applications, the predicted values are often closely associated with multiple objective factors and are the outcome of combined action of various factors, so this paper does not discuss the case in which the input is based on time sequence. How shall we predict with the support vector machine when input matrix is still unknown?

* ***Corresponding author's*** e-mail: hpzhou@aust.edu.cn

This paper adopts the two-stage grey prediction model and the support vector machine forecasting model (hereinafter referred to as D_2GM-SVM) to combine a new kind of prediction model. It is different from the traditional combined forecasting model because each model forecasts the results respectively, and then gets final prediction results through certain calculation method based on the respective prediction results. But, in the new prediction model put forward in this paper, the two kinds of forecast models complete respectively different types of forecast. First of all, we predict the future value of the input matrix by grey prediction model. Then we have the value as input of support vector machine. Finally, we forecast the output of the system. The two parts of different predictions have no time order and can be parallel.

2 The Support vector machine

The support vector machine separates these points of training set as far away as possible by constructing the optimal separating hyper plane. If the points of training set are linear inseparable, the points are mapped from two-dimensional space to high-dimensional space. So, nonlinear problem is translated into a linear problem as well.

Set $T=\{(x_1,y_1),...,(x_l,y_l)\}\in(x\times y)^l$ as sample set, among them, $x_i\in x=R^n, y_i\in y=\{-1,+1\}, i=1,...,l$;. The function of the optimal separating hyper plane is as follows:

$$\text{s.t. } y_i[(w\cdot x_i)+b]\geq 1, i=1,...,l \quad (1)$$

After the optimal solution, ω^* and b^*, is obtained according to the formula(1). The form of the decision function is as follows:

$$f(x)=\mathrm{sgn}[(w^*\cdot x)+b^*]. \quad (2)$$

When the points of training set cannot be separated by the separating hyper plane, some error is allowed and recorded as ξ. At this time, in order to approximate linearly separable, the slack variable $\xi_i, \xi_i\geq 0$, as the degree of misclassification. At the same time, to ensure the maximum classification interval, penalty parameter C is introduced as weight to measure the size of the two values. The objective function is rewritten into:

$$\frac{1}{2}\|w\|^2+C\sum_{i=1}^{l}\xi_i \quad (3)$$

To get the solution of objective function, Lagrange function is obtained as follows with Lagrange multipliers $\alpha_i, \alpha_i^*, \eta_i, \eta_i^*$:

$$L=\frac{1}{2}\|w\|^2+C\sum_{i=1}^{l}(\xi_i+\xi_i^*)-\sum_{i=1}^{l}\alpha_i(\varepsilon+\xi_i-y_i+<w\cdot x_i>+b)-\sum_{i=1}^{l}\alpha_i^*(\varepsilon+\xi_i^*+y_i-<w\cdot x_i>-b)-\sum_{i=1}^{l}(\eta_i\xi_i+\eta_i^*\xi_i^*) \quad (4)$$

Solving the derivative of the variable w, b, ξ_i and ξ_i^*, and setting them equal to 0:

$$\begin{cases} w=\sum_{i=1}^{l}(\alpha_i-\alpha_i^*)x_i \\ \sum_{i=1}^{l}(\alpha_i-\alpha_i^*)=0 \\ C-\alpha_i-\eta_i=0 \\ C-\alpha_i^*-\eta_i^*=0 \end{cases} \quad (5)$$

The solution of the original problem is converted into its dual form of the solution:

$$\begin{cases} \min\frac{1}{2}\sum_{i,j=1}^{l}(\alpha_i^*-\alpha_i)(\alpha_j^*-\alpha_j)<x_i\cdot x_j>+\varepsilon\sum_{i=1}^{l}(\alpha_i^*+\alpha_i)-\sum_{i=1}^{l}y_i(\alpha_i^*-\alpha_i) \\ s.t. \qquad \sum_{i=1}^{l}(\alpha_i^*-\alpha_i)=0 \\ \qquad\qquad 0\leq\alpha_i,\alpha_i^*\leq C \end{cases} \quad (6)$$

This is a convex quadratic programming problem. Solving the formula (6), decision equation is obtained:

$$f(x)=\mathrm{sgn}[\sum_{i=1}^{l}y_i\alpha^*<x_i\cdot x>+b^*] \quad (7)$$

The above are the cases of linearly separable or approximately linearly separable. If it is a linear inseparable, kernel function satisfying the Mercer conditions is introduced and it is transformed into solving the linear problem in high dimension space. There are many types of construction methods of the kernel function, in which the radial basis kernel function is widely applied because of its good performance[12-14]. So, the radial basis kernel function is also adopted in this paper, with the decision function after introducing kernel function as follows:

$$f(x)=\mathrm{sgn}[\sum_{i=1}^{l}y_i\alpha^*<x_i\cdot x>+b^*] \ . \quad (8)$$

3 Grey model (GM)

Grey system[20] studies uncertainty problems such as small sample, and poor information under the premise of less information. grey sequence generation digs out a regular data sequence from the disordered raw data samples. As for the shock disturbed system, which itself is influenced by the external environment and the data is irregular and may be changed from time to time, the grey model prediction, will lose its original predictive effect. Then we need to find a data pre-processing method, and the generated data can reflect the system's perturbation. In this paper, buffer operators [21] are adopted to weaken the system disturbance and correct the original data. Thus, the regularity and smoothness of the original data sequence are enhanced and the original data sequence is more suitable for grey prediction theory. Buffer operator acts on the original data sequence for the first time denoted as XD. For the different disturbance system, the forecast effect is not ideal if the first generation sequence through buffer operator is used. Buffer operator effects on the original data sequence, denoted as D. Finally, the prediction data is reverted. The theory basis is listed in the following part:
Set the original data sequence:

$$X^{(0)} = (x_1^{(0)}, x_2^{(0)}, ..., x_n^{(0)}) \ . \tag{9}$$

Buffer operator is denoted as D. After $x_n^{'(0)} = x_n^{(0)}D$, get the new sequence denoted as $X^{'(0)}$:

$$X^{'(0)} = (x_1^{'(0)}, x_2^{'(0)}, ..., x_n^{'(0)}) . \tag{10}$$

Among them:

$$x(k)d^2 = \frac{1}{n-k+1}[x(k)d + x(k+1)d + ... + x(n)d] \\ k = 1, 2, ..., n \ . \tag{11}$$

The new data sequence generated by first-order buffer operator is effected by 1-AGO again and get $X^{'(1)}$:

$$x_k^{'(1)} = \sum_{i=1}^{k} x_k^{'(0)} \qquad k = 1, 2, ..., n . \tag{12}$$

Among them, $X^{'(1)} = (x_1^{'(1)}, x_2^{'(1)}, ..., x_n^{'(1)})$.
Check the smoothness of generating sequence:

$$\rho_k = \frac{x_k^{'(0)}}{x_{k-1}^{'(0)}} = \frac{x_k^{'(0)}}{\sum_{i=1}^{k=1} x_i^{'(0)}} \ . \tag{13}$$

Because the development coefficient, $\alpha(\alpha = \dfrac{\frac{b}{x_{k-1}^{(1)}} - \rho_k}{1+0.5\rho_k})$, depends on the smooth ratio ρ_k of the buffer sequence $X^{'(1)}$ to a certain extent.

After taking the immediate vicinity of the mean of the sequence $X^{'(1)}$, the sequence is:

$$Z(1) = \left(z_2^{(1)}, z_3^{(1)}, ..., z_n^{(1)}\right) . \tag{14}$$

Among them, $z_n^{(1)} = 0.5(x_k^{'(1)} + x_{k-1}^{'(1)}, k = 2, 3, ..., n)$.

The differential equation of $X^{'(1)}$ is as follows:

$$\frac{dx_1^{'(1)}}{dt} + a x_1^{'(1)} = b \ . \tag{15}$$

Obtain the values of a and b by the following formula:

$$[a,b]^T = (B^T B^{-1}) B^T Y \ . \tag{16}$$

Among them:

$$B = \begin{bmatrix} -z_2^{(1)}, 1 \\ -z_3^{(1)}, 1 \\ \vdots \\ -z_n^{(1)}, 1 \end{bmatrix} \quad Y = \begin{bmatrix} x_2^{'(0)} \\ x_3^{'(0)} \\ \\ x_n^{'(0)} \end{bmatrix} . \tag{17}$$

The final predictive model is:

$$\hat{x}_{i+1}^{'(0)} = \hat{x}_{i+1}^{'(1)} - \hat{x}_i^{'(1)} = (1-e)(x_1^{'(0)} - \frac{b}{a})e^{-ak}, t = (1, 2, ..., n) . \tag{18}$$

4 Ggrey support vector machine model (D$_2$GM –SVM)

The overall prediction process is divided into two parts with a use of multistage grey support vector machine model, i.e., the prediction of the input value of support vector machine and the prediction of the final result. The sample data are divided into the input vector X and the output vector Y.

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{pmatrix}, \quad Y = \begin{pmatrix} y_1 \\ y_2 \\ \cdots \\ y_m \end{pmatrix}$$

Among them, m represents the number of the sample. N represents the number of influencing factors related Y. If the values of the input vector, $(x_{i1}, x_{i2}, ..., x_{in})$, are known ,the values of the output vector, $y_i, i = 1, 2, ..., m$, can be determined. In this paper, we discuss that how to predict the vector, $(x_{i1}, x_{i2}, ..., x_{in})$, according to the value of the input sample before i-1 in the case of unknown input values, $(x_{i1}, x_{i2}, ..., x_{in})$. And then the prediction value of $(x_{i1}, x_{i2}, ..., x_{in})$ is used as the input values of support vector machine model. Finally, we predict the value of y_i. As is shown in figure 1, the prediction steps are as follows:

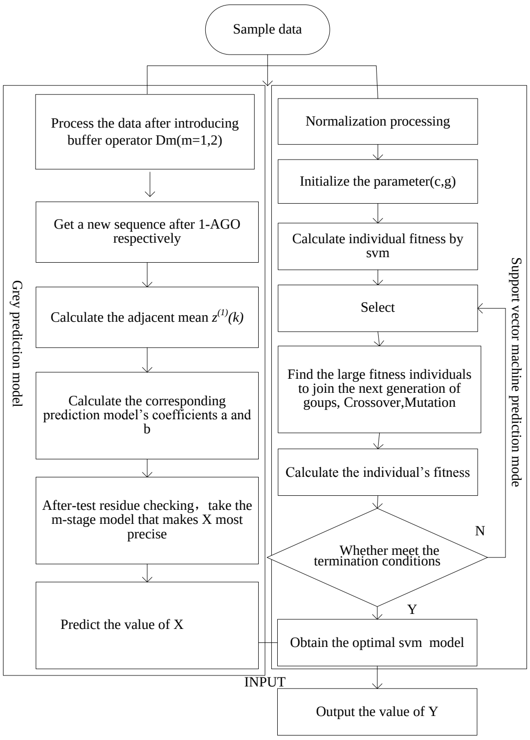

Figure 1. Two-stage gray support vector machine forecasting process

Step 1: Buffer operator D is introduced to pre-process the each component, $X^{(0)} = X_j, j=1,2,...,n$, of the raw data respectively. After buffer operator handling, we obtain n data new sequences denoted as $X_j^{(1)}$, in which, $X_j^{(1)}$ is the new data sequence after affected by buffer operator.

Step 2: Buffer operator D is introduced to pre-process the each component, $X^{(0)} = X_j, j=1,2,...,n$, of the raw data respectively. After buffer operator handling, we obtain n data new sequences denoted as $X_j^{(1)}$, in which, $X_j^{(1)}$ is the new data sequence after affected by buffer operator.

Step 3: The new sequence data, $X_j^{(1)}$, is obtained through 1-AGO of the data sequence $X_j^{(1)}$.

Step 4: The adjacent mean of the data sequence $X_j^{(1,1)}, X_j^{(2,1)},..., X_j^{(t,1)}$ is calculated, which is denoted as z (1)(k).

Step 5: We need to obtain the values of the parameter a and the parameter b of the each stage corresponding model by calculating the differential and derivative of the formula, $x^{(0)}(k)+az^{(1)}(k)=b$, to determine the grey model DGM.

Step 6: The original data samples are normalized.

Step 7: Genetic algorithm (GA) is used to optimize parameters for support vector machine model. Firstly, we are expected to initialize parameters(c,g) of support vector machine.

Step 8: We calculate individual fitness by svm.

Step 9: We select.

Step 10: We find the large fitness individuals to join the next generation of groups, Crossover, Mutation

Step 11: We calculate the individual's fitness and ensure whether its value meets the requirements. And if it does, then take the values of the optimal parameters (c,g) as the support vector machine model parameters. Otherwise the algorithm continues iteration.

Step 12: The predictive value of X in step 6 is identified as input vector of support vector machine model, and ultimately predict the value of Y.

5 The experimental simulation results

The mortality of coal million tons is selected as the predictive value. The data samples are from every coal-producing provinces (regions) in China. Four indicators are selected as the related factors of mortality of coal million tons, which are comprehensive mechanized coal mining rate (%), mechanized tunnelling rate (%), coal mining mechanization rate (%) and raw coal overall efficiency (t/work). More than 200 sample data are selected in the coal-producing provinces (regions) all over the country from 2004 year to 2012 year, including Anhui Province, Henan Province, Shanxi province, Guizhou Province and other 21 provinces (regions). Among them, 100 randomly selected sample data serve as the training sample, 10 are randomly selected as the test sample from the rest of the samples. Finally, the use of the model was established to forecast the mortality of coal million tons of the next two years from2014.

5.1 EACH INDEX PREDICTION BASED ON DGM MODEL

As for each index in each province (area) as a data sequence from 2004 year to 2012 year, D_m(m=1,2) is applied to act on the four indexes' values of the sample data, the data sequences, XD_1,XD_2 is obtained. The values of average residual, P and C [20] of predictive models are as shown in Table1.

As can be seen from Table 1, in the case of raw data, the four indexes' average residual is respectively 0.0724, 0.036, 0.039 and 0.027. In the case of one-stage data XD, the four indexes' average residual is respectively 0.034, 0.024, 0.021 and 0.148. In the case of two-stage data XD_2, the four indexes' average residual is respectively 0.015, 0.052,0.018 and 0.084. In view of mechanized coal mining rate, comparison of X and XD P is with regard to the indicator of mechanized mining rate, and the value of P is from 77% up to 86%. Comparison of XD and XD_2 P is with regard to the indicator of mechanized mining rate, and the value of P is from 86% up to 91%. Comparison of X, XD and XD_2 the value of C is respectively 0.08, 0.073 and 0.018. So the value of D_2GM forecasting model is acquired when forecasting the index, mechanized mining rate. In this way, the prediction models of these four indicators of the mortality of coal million tons are determined. Upon the completion of the model with testing samples respectively, the results are shown in Table 2:

TABLE 1 The value of m of each index

Buffer operator	Mechanized coal mining rate (%)			Mechanized drivage rate (%)			Coal mining mechanization rate (%)			Raw coal overall efficiency (%)		
	The average residual	P	C	The average residual	P	C	The average residual	P	C	The average residual	P	C
X	0,0724	77%	0,08	0,036	71%	0,1	0,039	80%	0,105	0,027	75%	0,076
XD	0,034	86%	0,073	0,024	93%	0,058	0,021	94%	0,052	0,148	70%	0,801
XD2	0,015	91%	0,052	0,018	95%	0,051	0,011	95%	0,043	0,084	81%	0,721

TABLE 2 Model test

	Mechanized coal mining rate (%)			Mechanized drivage rate (%)			Coal mining mechanization rate (%)			Raw coal overall efficiency (%)		
	The average residual	P	C	The average residual	P	C	The average residual	P	C	The average residual	P	C
1	0,047	78%	0,05	0,338	81%	0,13	0,501	81%	0,53	0,021	87%	0,054
2	0,025	80%	0,081	0,476	81%	0,812	0,724	80%	0,068	0,124	76%	0,435
3	0,031	81%	0,071	0,487	82%	0,843	0,705	82%	0,034	0,074	81%	0,467
4	0,032	82%	0,063	0,532	86%	0,902	0,624	80%	0,081	0,038	70%	0,435
5	0,019	82%	0,087	0,093	77%	0,911	0,754	80%	0,068	0,072	78%	0,651
6	0,052	70%	0,077	0,452	83%	0,876	0,712	80%	0,051	0,018	82%	0,425
7	0,017	82%	0,086	0,512	78%	0,831	0,724	87%	0,074	0,075	83%	0,075
8	0,022	78%	0,11	0,432	80%	0,809	0,714	90%	0,084	0,048	87%	0,084
9	0,035	83%	0,099	0,418	89%	0,17	0,712	82%	0,061	0,047	83%	0,071
10	0,051	85%	0,081	0,376	84%	0,105	0,665	83%	0,181	0,042	81%	0,042
Average	0,0331	80%	0,081	0,4116	82%	0,639	0,6835	83%	0,123	0,0559	81%	0,274

What is shown in Table 2:

When the trained D_2GM model is used to predict the four indicators of the mortality of coal million tons, the results are as follows. The average residual errors of four indicators are 0.0331, 0.4116, 0.6853 and 0.0559 respectively. Their average accuracy is 80%, 82%, 83% and 81% respectively. Their variance ratio is 0.081, 0.639, 0.123 and 0.274 respectively. The prediction accuracy can reach more than 80%. The model prediction accuracy is higher, so D_2GM forecast model is feasible and effective in the index of the mortality of coal million tons prediction.

5.2 GA-LSSVM PREDICTION MODEL

Genetic algorithms (GA) is first proposed by John Holland in the 1860s. The intelligent search of genetic algorithm is adopted in the process of parameter selection of support vector machine algorithm in this paper and find the optimal parameter. Looking for the optimal support vector machine model for the sample of the coal mine. By comparing with grid search algorithm, genetic search algorithm can quickly obtain the satisfactory parameters.

SVM model includes the qualitative options and quantitative options. The former includes how to identify specific support vector machine algorithm and Kernel. The latter is support vector parameter selection. LSSVM parameter choice includes: kernel function parameters and the error penalty parameter. The error penalty parameter of different SVM is named different. The name of different kernel function parameters is not the same. However, the role and significance is both the same. For convenience of description herein, penalty parameter and kernel parameters are expressed using γ and σ.

Kernel parameter selection of least squares support vector machine is directly related to the learning performance and generalization ability of least squares support vector machine. The parameter selection methods commonly used are mainly cross-validation method and nuclear calibration. Cross-validation method requires a lot of computing, to determine the optimum parameters. Especially when the number of parameters is large, it will take a lot of time to strike the optimal solution. Nuclear calibration method is related to much knowledge and research of nuclear matrix, so it is more difficult to achieve it. To compensate for the insufficient of existing parameter selection algorithm, the SVM model predicts the gas level with the genetic algorithm. The algorithm is not only able to achieve a global search, and search speed can be guaranteed. The practical application shows: the improved support vector machine parameters selection algorithm based on genetic algorithm can get optimal operating parameters of non-stationary time series and nonlinear prediction model.

It is a proven method of selecting SVM kernel parameters. Optimizing process shown in Figure 2.

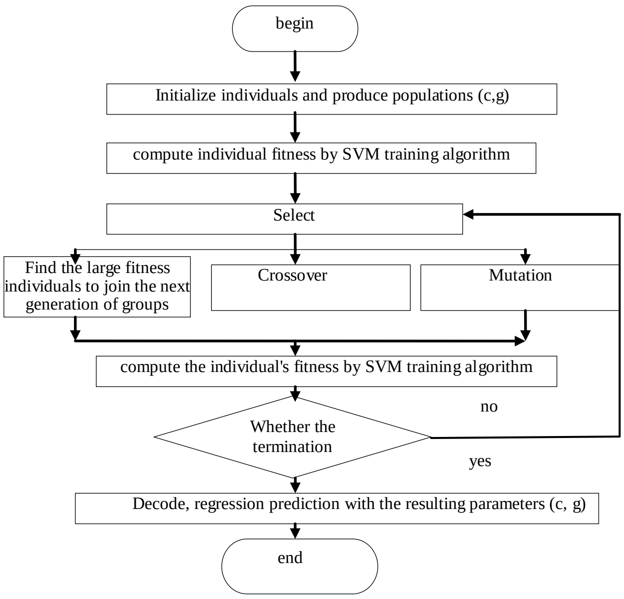

Figure2 Optimization process of GA-SVM

5.3 SVM PREDICTION MODEL

LIBSVM tool is adopted in training and testing SVM prediction model of the mortality of coal million tons. Selection of support vector machine model is mainly about the selection of nuclear parameters and penalty parameters. The common RBF kernel function is adopted in the paper. The particle of swarm optimization (PSO) is used for parameter selection with cross validation method. Mean square error (mse) of the mortality of coal million tons is selected as the fitness function. A search of time and root mean square error of particle of swarm optimization, grid algorithm and genetic algorithm are shown as table 3. Through the comparison of table 3, particle of swarm optimization has minimal root mean square error, and is faster convergence to the optimal solution in this example. In this model, the optimal parameters(c, g) of PSO are (77.654, 0.09541) as the best parameters of support vector machine model.

TABLE 3 Forecast result

Homing algorithm	penalty parameter	Kernel Parameter	Search time *(sec)*	mean square error
GS	2	0.125	13.5194	0.1652
GA	66.325	1.2541	20.201	0.2541
PSO	77.654	0.09541	8.2541	0.0954

The average absolute error of the training samples is 0.09215 when the trained support vector machine model is used to predict the training sample data. The square correlation coefficient is 1.54. The mean square error is 0.0541. In order to validate SVM model, 20 test samples is selected and divided into 2 groups with prediction results shown in Table 4 and Figure 3. In first group, the average relative error of test samples is 0.0292. The maximum relative error is 0.068. The minimum relative error is 0. In second group, the average relative error of test samples is 0.1. The maximum relative error is 0.084219. The minimum relative error is 0.0027.From the forecast results of test samples, the error is smaller and prediction is effective with support vector machine model.

TABLE 4 Test sample prediction error

		1	2	3	4	5	6	7	8	9	10
first group	actual value	1,024	1,654	0,953	0,552	2,412	6,287	8,654	0,541	2,257	6,21
	predicted value	1,004	1,587	0,912	0,547	2,421	6,387	8,657	0,5748	2,1045	6,021
	error	0,02	0,067	0,041	0,005	-0,009	-0,1	-0,003	-0,0338	0,1525	0,189
	relative error	0,020	0,041	0,043	0,009	0,004	0,016	0,000	0,062	0,068	0,030
	accuracy	0,995	0,9691	0,9964	0,5481	0,9802	0,9755	0,969	0,7928	0,9352	0,9596
second group	actual value	3,332	1,237	0,357	0,542	4,257	6,354	3,574	2,745	1,786	4,572
	predicted value	3,341	1,241	0,362	0,245	4,572	6,853	3,875	2,746	1,542	4,875
	error	0,009	0,004	0,005	-0,297	0,315	0,499	0,301	0,001	-0,244	0,303
	relative error	0,00270108	0,00323363	0,014005602	0,54797048	0,073995772	0,078533207	0,084219362	0,000364299	0,136618141	0,066272966
	accuracy	0,985	0,945	0,996	0,574	0,982	0,485	0,687	0,954	0,86	0,88

5.3 METHOD COMPARISON

With a selection of data samples of the mortality of coal mine one million tons from 21 provinces (regions) in 2012, a respective forecast was conducted with the proposed grey support vector machine forecasting model DGM- SVM, GM and BP neural network combination forecast model respectively. The forecast results are shown in table 5. In Table 5, the average relative error is 8.54% with D_2GMSVM model. And the maximum prediction error is 9.85%. The mean square error is 6.54%. The average relative error is 25.48% and the maximum prediction error is 14.27% with GM (1, 1) model. The average relative error is 14.57% with BP neural network model. And the maximum prediction error is 13.54%, the mean square error is 17.57. The average relative error is 9.54% with DGMSVM model. And the maximum prediction error is 10.25%, the mean square error is 7.841.The prediction error of model D_2GM – SVM is least. It is the closest to the axes 0. Through comparison and analysis, the accuracy of forecasting model D_2GM -SVM is superior to the BP neural network combination forecast model and GM (1, 1) model.

TABLE 5 Error value contrast

prediction model	D_2GM-SVM	GM	BP neural network	DGM-SVM
average relative error	0,0854	0,25476	0,14572	0,0954
maximum prediction error	0,0985	0,14265	0,13542	0,1025
mean square error	0,0654	0,09574	0,17567	0,07841

6 Conclusion

Two-stage grey support vector machine forecast model (D_2GM-SVM) is put forward in this paper to avoid the limitations of single prediction model. Each relevant indicator of the predictive value is forecasted with two-stage grey forecasting model (D_2GM), with the following input of the support vector machine. At the same time, the radial basis kernel parameter and the punish parameter of SVM are optimized by the genetic algorithm to get the optimal parameters (c,g) and to determine the best support vector machine model. Finally, the each index value is as the input of SVM model to finish prediction process. A Prediction of the index value with one-stage grey forecasting model and training support vector machine model can be established in parallel. Examples prove that DGM-SVM has the smallest average relative error, prediction error and mean square error, with a higher prediction precision than the single prediction model, such as the original grey model GM (1,1) and BP neural network model. It shows that the proposed one-stage grey support vector machine prediction model DGM-SVM is effective and feasible.

Acknowledgment

This work is supported by Anhui Provincial Natural Science Foundation of Universities Key Projects (Grant No. KJ2010A083) , the Natural Science Foundation of the Anhui Higher Education Institutions of China (KJ2012A099),and is also supported by the National Natural Science Foundation of China (Grant No. 51174257) ,Yong Backbone Teachers of Anhui University of Science and Technology and Colleges and universities in Anhui Province outstanding young talent support plan(ZY289).

References

[1] Steffen Schumann,Lutz-P. Nolte, GuoyanZheng. Comparison of partial least squares regression and principal component regression for pelvic shape prediction. Journal of Biomechanics vol.46, pp:197–199, 2013.
[2] Xiangxi Wen. Xiangru Meng. Zhiqiang Ma.Xiongchun Zhang. The Chaotic Analysis and Trend Prediction on Small Time Scale Network Traffic, ACTA ELECTRONICA SINICA , 40(8):1609-1615, 2012.
[3] AtillaCifter. Forecasting electricity price volatility with the Markov-switching.GARCH model: Evidence from the Nordic electric power market[J] Electric Power Systems Research 102 (2013) :61- 67.
[4] Xiang Changshen,Zhang Linfeng.Grain yield prediction model based on gray theory and markvo. Computer Science,2013,40(2):245-248.
[5] Bates J M,Granger C W J. Combination of forecasts[J].Operations Research Quarterly, 1969, 20(4):451-468.
[6] Bischoff, C. W. (1989). The combination of macroeconomic forecasts[J] Journal of Forecasting, 8, 293–314.
[7] Bunn, D. W. (1989). Forecasting with more than one model. Journal of Forecasting,8, 161–166.
[8] ShujieShen,GangLi, HaiyanSong.Combination forecasts of international tourism demead [J]Annals of Tourism Research, 38(1):72–89, 2011.
[9] V.N.Vapanik,A.J.Chervonenkis. Ordered risk minimization[M].Automation and Remote Control.1974.
[10] V.N.Vapnik,Estimation of dependences based on empiricaldata[M].Springer-Verlag,Berlin,1982.
[11] Zhou Jian, Li Xibing, Shi Xiuzhi. Long-term.Prediction model of rockburst in underground openings usingheuristic algorithms and support vector machines[J]. Safety Science 50 (2012) 629–644.
[12] JianwuZeng, Wei Qiao.Short-term solar power prediction using a support vector machine. Renewable Energy 52 (2013) 118-127.
[13] Xiaoyan Guan, Shaoli Wang, ZhanyiGao, Ye Lv.Dynamic prediction of soil salinization in an irrigation district based onthe support vector machine[J]. Mathematical and Computer Modelling 58 (2013) 719-724.
[14] Isis Didier Lins, MoacyrAraujo, Márcio das ChagasMoura, Marcus André Silva Enrique LópezDroguettPrediction of sea surface temperature in the tropical Atlanticby support vector machines. Computational Statistics and Data Analysis 61 (2013) 187–198.
[15] Guohua Tan, Jianzhuo Yan, Chen Gao, SuhuaYang.Prediction of water quality time series data based on leastsquares support vector machine. Procedia Engineering 31 (2012) 1194 – 1199.
[16] You Lv, Jizhen Liu, Tingting Yang, DeliangZeng.A novel least squares support vector machine ensemble model for NOx emission prediction of a coal-fired boiler. Energy 55 (2013) 319-329.
[17] Wei Zhong, Rick Chow, Jieyue He. Clinical charge profiles prediction for patients diagnosed with chronic diseases using Multi-level Support Vector Machine. Expert Systems with Applications 39 (2012) 1474-1483.
[18] B. Yeganeh, M. Shafie Pour Motlagh, Y. Rashidi, H. Kamalan.Prediction of CO concentrations based on a hybrid Partial Least Square and Support Vector Machine model. Atmospheric Environment 55 (2012) 357-365.
[19] Hassan Golmohammadi, Zahra Dashtbozorgi, William E. AcreeJr.Quantitative structure activity relationship prediction of blood-to -brain partitioning behavior using support vector machine. European Journal of Pharmaceutical Sciences 47(2012) 421-429.
[20] Cheng-Lung Huang, Jian-Fan Dun. A multiple kernel framework for inductive semi-supervised SVM learning [J]. Neurocomputing, 2012,90(1): 46-58.
[21] Dehua Liu, Hui Qian, Guang Dai, Zhihua Zhang.An iterative SVM approach to feature selection and classification in high-dimensional datasets [J].Pattern Recognition,2012,46 (9) :2531-2537

Authors

Zhou Huaping, born in 1979, Henan Province of China

Current position, grades: PhD, associate professor
University studies: School of Computer Science and Engineering, Anhui University of Science and Technology, Huainan, China
Scientific interest: information fusion, data mining, computer applications in the mining industry

Yuan Yue Yuan, born in992, Suzhou city, Anhui Province

Current position, grades: Postgraduate student
University studies: Anhui University Of Science And Technology
Scientific interest: Data mining, Data analysis