ISSN : 0974 - 7435 Volume 10 Issue 16

2014

BioTechnology
An Indian Journal

FULL PAPER
BTAIJ, 10(16), 2014 [9063-9070]

The prediction research of SO_2 emissions in thermal power industry

Jianguo Zhou, Sisi Fu*
Department of Economic Management, North China Electric Power University, Baoding 071003, (CHINA)
E-mail : 1019153605@qq.com

ABSTRACT

The prediction of sulfur dioxide (SO_2) emissions in thermal power industry belongs to the small sample, poor information gray system. On this basis, this paper established a combination forecasting model based on support vector machine (SVM) and radial basis function neural network (RBFNN). The weights of the model are obtained by using genetic algorithm (GA). At last, on the basis of the historical data of thermal power industry during 1991 to 2012, it predicts SO_2 emissions of thermal power industry for the next 8 years. The results showed that, the predicted results were accurate by using this model, which is an effective method for the predict of SO_2 emissions of thermal power industry in our country in the future.

KEYWORDS

Sulfur dioxide emissions; Support vector machine (SVM); Radial basis function neural network (RBFNN); Genetic algorithm (GA); Combination forecasting model.

INTRODUCTION

Thermal power industry belongs to heave pollution industry in China, whose SO_2 emissions account for about 50% of the national industrial sector's total emission every year. SO_2 does harm to various aspects of the social life, such as atmospheric environment, human health, construction, water etc. Therefore, it becomes an important issue to control the SO_2 emissions, and predict research of SO_2 emissions is a basic work for the atmospheric pollution control.

At present, in academia, there are a few of statistical methods used for the regional prediction of SO_2 emissions. There are some commonly used methods including gray prediction method[1], regression, neural network methods etc. SO_2 emissions prediction is a complex nonlinear system affected by many factors, and it belongs to the "small simple" and "poor information" gray system. It was found that SVM and RBF neural network have a unique advantage to deal with this kind of problem. Combining these two methods, this paper established an optimal combination forecasting model, the weights of which were obtained by using genetic algorithm (GA). It can not only improve the prediction accuracy, but also provide research foundation for the SO_2 emissions control policy in China.

THEORETICAL BASIS

Support vector machine

Support vector machine (SVM) proposed by Vapnik et al, is based on the VC dimension theory and structural risk minimization principle on the basis of statistical learning theory. SVM show many unique advantages in tackling small sample, nonlinear and high dimensional pattern recognition problems, and it can be applied to other machine learning problems[2]. Since the thermal power industry is short of statistical data and has different trends every year, SVM can be a good solution to the SO_2 emissions prediction.

Radial basis function neural network

Radial basis function (RBF) is a kind of technology conducted in high dimensional space. The basic idea of RBF neural network is that using RBF as the "base" of hidden units constitutes a hidden layer space, in which the input vectors are transformed form low-dimensional model to a high-dimensional space, making the non-linear problem in low-dimensional space can be linearly divided in high-dimensional space. RBF neural network, with the advantages of simple structure, simple training and learning fast convergence, can approximate any nonlinear function[3]. Since SO_2 emissions prediction of thermal power industry, affected by many factors, is a nonlinear system, RBF neural network is an ideal choice for SO_2 emissions prediction.

Combination forecasting model

Combination forecasting method refers to combining the prediction results of varieties of forecasting methods through the establishment of a combined forecasting model. Combination forecasting model is more systematic and more comprehensive than single forecasting model. So it can effectively reduce the influence of random factors to improve the accuracy and stability of prediction[4]. Since SO_2 emissions prediction is a complex nonlinear systems affected by many factors, a single fixed pattern is difficult to accurately describe the actual complexity of prediction[5]. This paper established a combination forecasting model, based on SVM model and RBF neural network model, to improve the prediction accuracy.

THE COMBINATION FORECASTING MODEL OF SO_2 EMISSIONS PREDICTION BASED ON SVM AND RBFNN

Support vector machine forecasting model

This paper sets up a training sample set (x_i, y_i), in which x_i equals to R^n. Construct a linear function:

$$f(x) = w\varphi(x) + b \tag{1}$$

Where w is the weight vector of input space and b is a threshold value. Insensitive loss function (ε) takes the next form:

$$|y_i - f(x_i)| = \begin{cases} 0, |y_i - f(x_i)| < \varepsilon \\ |f(x_i) - y_i| - \varepsilon, others \end{cases} \tag{2}$$

The empirical risk function is defined as:

$$R_{emp} = \sum_{i=1}^{l} |y_i - f(x_i)| \tag{3}$$

The SVM regression algorithm can be expressed as:

$$min\left(\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}|y_i - f(x_i)|\right)$$

$$subject\ to\ \begin{cases} y_i - w\varphi(x) - b \le \varepsilon + \xi \\ w\varphi(x) + b - y_i \le \varepsilon + \xi^* \end{cases} \quad (4)$$

Where ξ and ξ^* respectively represent the upper and lower limits of the training error.

Due to the high dimensionality of the feature space, it's almost impossible to solve the above equation. It can be converted to solve its dual problem by using duality technique and introducing the dot product kernel function $k\,(x_i, x_j)$.

$$max\left[(\alpha,\alpha^*) - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}(\alpha,\alpha^*)^T\,(\alpha,\alpha^*)\langle x_i, x_j\rangle + \sum_{i=1}^{l}(\alpha,\alpha^*)\,y_i - \sum_{i=1}^{l}(\alpha,\alpha^*)\,\varepsilon\right]$$

$$subject\ to\ \sum_{i=1}^{l}(\alpha,\alpha^*) = 0, 0 \le \alpha,\alpha^* \le C \quad (5)$$

So the regression estimation function is:

$$f(x) = \sum_{i=1}^{l}(\alpha,\alpha^*)\,k(x_i, x) + b \quad (6)$$

Radial basis function neural network model

Gaussian function is one of the most commonly used radial basis function in RBF neural network, and its expression is:

$$R_i(x) = exp\left(-\frac{\|x-c_i\|^2}{2\sigma_i^2}\right)\ i=1, 2, \ldots, m \quad (7)$$

Where x is an n-dimension input vector, c is the i-th basis function center, σ_i is the i-th basis function planning factor, $\|x - c_i\|$ is nearly formula norm, m is the number of nodes in the hidden layer.

Set the input of the input layer $X=(x_1, x_2, \ldots, x_j, \ldots, x_n)$, and the actual output $Y=(y_1, y_2, \ldots, y_k, \ldots, y_p)$. The input layer achieves the non-linear mapping from $X \rightarrow R_i(x)$. The output layer implements a linear mapping from $R(X) \rightarrow y_k$. K-th output of output layer neural network is:

$$\widehat{y_k} = \sum_{i=1}^{m} w_{ik}\,R_i(x)\ k = 1, 2, \ldots, p \quad (8)$$

Where m is the number of hidden layer node, p is the output layer node, w is the connection weights of *i-th* hidden layer neurons and *k-th* hidden layer neurons, *R(x)* is the *i-th* hidden layer neuron action function.

When the weights and thresholds of the hidden layer and output layer neurons are determined, the output of the network is determined. So the RBF network learning is the revision process of the weights and threshold of each network layer. This paper selects a newrb function to create an approximate RBF network.

Combination forecasting model

Combining SVM model and RBF neural network model, the combination forecasting model is set up as the following:

$$Y_t = \alpha * T_t + \beta * I_t + \varepsilon \quad (9)$$

Where *Yt* is the predicted value, *Tt* is the SVM model prediction, *It* is the RBF neural network model prediction, *t* is the time serial number, α and β are the weight factor, ε represents the error value caused by other factors, which can be regarded as constant in a certain time.

Constraints for α、β and ε in the above formula is:

$$Subject\ to \begin{cases} \alpha + \beta \le 1; \\ 0 < \alpha, \beta < 1; \\ \varepsilon \ge 0 \end{cases} \quad (10)$$

This paper will use the genetic algorithm to solve these parameters value. It is beneficial to optimizing the combination forecast model, avoid local optimum.

Genetic algorithm

The above combination forecasting model contains several unknown parameters. The problem of determining these parameters' value can be transformed into the optimization problem, which can be solved by genetic algorithm. Both of Y_t and $\widehat{Y_t}$ change as the time *t* varies. When these two curves fit on the n data, the mean square error (mse) is:

$$\delta = \sqrt{\frac{\sum_{t=1}^{n}(\widehat{Y_t}-Y_t)^2}{n}} \tag{11}$$

Obviously, this function value is as small as possible, and set it as a moderate value function. This paper will use the genetic algorithm tool GUI to solve the parameters of combination forecast model.

THE PREDICTION OF SO_2 EMISSIONS IN CHINESE THERMAL POWER INDUSTRY

Affecting factors of thermal power SO_2 emissions

The main factors affecting SO_2 emissions are coal consumption, installed capacity, power generation and GDP. SO_2 emissions and these four factors can be respectively expressed by x_1, x_2, x_3, x_4 and x_5. $X = \{x_2, x_3, x_4, x_5\}$ is the input layer, and SO_2 emission is the output layer.

The raw data for predicting SO_2 emissions

Select the historical data of thermal power industry during 1991 to 2012 as the sample to research the SO_2 emissions (see TABLE 1).

TABLE 1: The original sample data

Year	SO_2 emissions/ (10^4t)	coal consumption/ (10^4t)	installed capacity/ (10^4kw)	power generation/ (10^8kwh)	GDP/ (10^8¥)	Year	SO_2 emissions/ (10^4t)	coal consumption / (10^4t)	Installed capacity/ (10^4kw)	power generation/ (10^8kwh)	GDP/ (10^8¥)
1991	528	30119.1	11359	5524.6	21781.5	2002	750.1	65600	26555	13273.8	120332.7
1992	576	33459.4	12585	6214.7	26923.5	2003	861.9	77976.5	28977	15803.6	135822.8
1993	628.3	36831.3	13802	6838.8	35333.9	2004	994.9	91961.6	32948.3	17955.9	159878.3
1994	685.7	40053.1	14874	7459.2	48197.9	2005	1167.2	103263.5	39138	20473.4	184937.4
1995	717.8	44440.2	16294	8043.2	60793.7	2006	1204.1	118763.9	48382	23696	216314.4
1996	731.6	48808.6	17886	8777.1	71176.6	2007	1147.1	130548.8	55607.4	27229.3	265810.3
1997	789.5	48979	19241	9240.7	78973	2008	1059.9	145246.4	60285.8	27820.2	314045.4
1998	696.8	49489.3	20988	9441	84402.3	2009	933	143967.3	65107.6	29827.8	340902.8
1999	637.4	51163.5	22343	10205.4	89677.1	2010	899.8	154542.5	70967.2	33319.3	401512.8
2000	707.2	55811.2	23754	11141.9	99214.6	2011	901.2	175043	76546	38253.2	472881.6
2001	725.5	57687.9	25314	11767.5	109655.2	2012	883	185548	81917	38554.5	519322

Note: The above datum is from the web site of corresponding bureau statistics.

The prediction of SO_2 emissions

On the basis of the original sample data in TABLE 1, this part will respectively use the SVM model, RBF neural network model and combined forecasting model based on genetic algorithm to predict the SO_2 emissions in Chinese thermal power industry.

(a) Prediction of SVM model

(1) Using SVM model to solve the regression problem, the key is to choose the kernel function and parameters. After comparing, the ε-SVR regression and RBF kernel function are selected. This paper selects the datum from 1991 to 2011 as a training set, and the datum from 1992 to 2012 as a test set. In order to avoid the input vector's magnitude difference affecting the training effect, the datum will be normalized before training. The normalized interval is set to [0,1].

(2) The selection of model parameters c and g. When using the SVM model, the related parameters should be adjusted. Then the model will get more ideal prediction classification accuracy. The main parameters are c and g. Cross-validation can be used to get the optimal parameters in some sense. There are some common cross-validation method, such as Hold-Out Method, K-fold Cross Validation and Leave-One-Out Cross Validation. In this paper, K-fold Cross Validation is used for the optimization process of the related parameters in SVM. Where K takes the experience value 3, the best parameter is that c is equal to 11.3137 and g is equal to 4.

(3) The determination of insensitive loss function ε. When using the optimal parameters (c, g), by changing the value of ε to train the training set, the optimal combination of parameters will be found. In the TABLE 2, there is squared correlation coefficient for each ε.

TABLE 2: Parameter ε and the corresponding squared correlation coefficient

ε	R(%)	ε	R(%)
≥0.5	NaN	0.05	97.78
0.3	91.42	0.01	98.63
0.2	91.27	0.005	98.60
0.1	97.23	0.001	98.50

The results can be seen from TABLE 2 that when c is equal to 11.3137 and g is equal to 4, ε will take 0.01. In this case, squared correlation coefficient reaches a maximum, and mean square error reaches the minimum, at the same time, the imitative effect achieves the best fit.

By the SVM model built above, the prediction results of SO_2 emissions from 1992 to 2012 are shown in TABLE 3.

(b) Prediction of RBF neural network model

On the basis of the original sample data in TABLE 1, this paper selects the datum from 1991 to 2011 as a training set, and the datum from 1992 to 2012 as a test set. The normalized interval is set to [0,1]. Using the RBF neural network training and testing commands in the neural network toolbox of MATLAB 7.8.0, the newrb function is selected to create the RBF neural network.

The call format of newrb function is: [net, tr]=newrb (P, T, GOAL, SPREAD, MN, DF). Where P is input vector, T is the target vector, GOAL is target error, SPREAD is extension constant, MN is the largest number of neurons, DF shows the frequency of the iterative process, and tr is the return value and training records.

Before using RBF neural network model training and testing, set that mn is equal to 22 and df is equal to 1, and then test the value of sp under the condition of that goal is respectively equal to 0, 0.001, 0.01 and 0.1. The standard of selecting sp is that the mean error of the predict value and the actual value is minimum. By series experiments, goal is set to 0.001, sp is set to 1, and the minimum value of MSE is 17.71.

Using the RBF neural network model built above, the prediction results of SO_2 emissions from 1992 to 2012 are shown in TABLE 3.

(c) Prediction of Combination forecasting model

Combining SVM model and RBF neural network model, the formation of SO_2 combination forecast model is as formulas (9).

This research will use the genetic algorithm tool GUI to solve the value of unknown parameters. After 9972 times iterations, the program stopped since the best fitness in several generations didn't get improved and terminated. The best adaptation is 17.4274, and the average modest value is 17.4275. The parameters are obtained: α=0.166, β=0.83, ε=5.22.

The parameter value obtained are brought into the combination forecasting model, using which SO_2 emissions are predicted. The SO_2 emissions predicted results are shown in TABLE 3.

Analysis of predicted results

The prediction results of SO_2 emissions during 1992 to 2012, as well as the relative error of the predicted values and the actual values are shown in TABLE 3. Figure 1 shows the trend of the actual value and the predicted value. According to the analysis in TABLE 3 and Figure 1, this part gets the following conclusion.

(a) The analysis of single forecast result

The mean square errors (mse) of the two single models are respectively calculated: SVM model's mse is 21.3 and RBF neural network model's mse is 17.7. It seems that the predicted value of RBF neural network model has a higher fitness with the actual value, which means it has a higher prediction accuracy.

TABLE 3 shows that the difference between the prediction value of both two single models and actual value are not great. In the SVM model, the relative error of predicted value and actual value are small and are below 2%, except that in the year of 1997 and 1999. In the RBF neural network, the predicted value has a high accuracy, except that the relative errors are more than 4% in the year of 1993, 1997 and 1999.

It can be seen from the graph that the curve trend of RBF neural network predicted value is closer to the actual values'. In comparison, the curve of SVM model predicted value is smooth, and can't show the prominent changes in some individual years.

TABLE 3: Prediction analysis of SO_2 emissions from thermal power industry

Year	Actual value	SVM model		RBF neural network model		GA combination forecasting model	
		Predicted value	Relative error(%)	Predicted value	Relative error(%)	Predicted value	Relative error(%)
1992	576.00	582.78	-1.18	559.95	2.79	566.72	1.61
1993	628.34	628.88	-0.09	602.89	4.05	610.01	2.92
1994	685.67	673.92	1.71	700.79	-2.21	698.75	-1.91
1995	717.84	711.91	0.83	735.16	-2.41	733.58	-2.19
1996	731.57	734.93	-0.46	731.76	-0.02	734.58	-0.41
1997	789.47	724.92	8.18	752.59	4.67	750.21	4.97
1998	696.79	703.71	-0.99	696.73	0.01	700.33	-0.51
1999	637.44	700.56	-9.90	687.07	-7.79	691.78	-8.52
2000	707.20	713.83	-0.94	694.91	1.74	700.49	0.95
2001	725.54	711.52	1.93	701.14	3.36	705.28	2.79
2002	750.09	756.82	-0.90	758.15	-1.08	760.12	-1.34
2003	861.95	855.29	0.77	866.79	-0.56	866.63	-0.54
2004	994.90	1001.42	-0.66	1004.42	-0.96	1005.12	-1.03
2005	1167.20	1152.15	1.29	1152.42	1.27	1152.99	1.22
2006	1204.10	1210.73	-0.55	1210.07	-0.50	1210.56	-0.54
2007	1147.10	1140.07	0.61	1146.41	0.06	1145.99	0.10
2008	1059.90	1053.04	0.65	1061.42	-0.14	1061.00	-0.10
2009	933.00	939.98	-0.75	927.56	0.58	931.13	0.20
2010	899.80	892.95	0.76	905.19	-0.60	904.76	-0.55
2011	901.20	908.28	-0.79	899.34	0.21	902.45	-0.14
2012	883.00	868.34	1.66	874.04	1.01	874.82	0.93

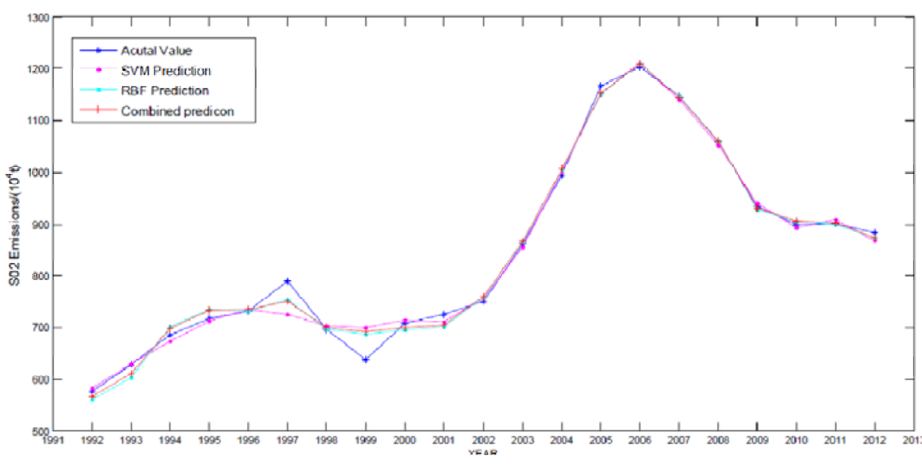

Figure 1: Accuracy comparison of SO_2 emissions prediction from1992 to 2012

(b) The analysis of combination predicted results

The combination forecast model's mse is 17.43, which means it has a higher accuracy than single forecast model. In the combination forecast model, the relative errors are small and below 3%, except that in the year of 1997 and 1999. It is similar to SVM model, but lower than SVM model. It can be seen form Figure 1 that the combination forecast model predicted value has a high inosculation with actual value. It combines the advantages of two single models, whose trend is closer to actual values'.

Informed by the above analysis that, both of two single models have a high accuracy, but the combination forecast model is better than single model. This reflects the scientific and reasonable of combination forecast model based on genetic algorithm, and indicates that it's an effective prediction method and can be used to promote research in each subject area.

Combination forecasting model predictive analysis on SO_2 emissions from 2013to 2020.

Based on the historical data of coal consumption, installed capacity, power generation and GDP, the corresponding data from 2013 to 2020 can be predicted, which can be used to predict the SO_2 emissions. This paper will use gray prediction model GM (1, 1) to predict these variables, the results are shown in TABLE 4.

TABLE 4 : Basic data for SO_2 emissions prediction

Year	Coal consumption /(10^4t)	Installed capacity /(10^4kw)	Power generation /(10^8kwh)	GDP /(10^8¥)	Year	coal consumption /(10^4t)	installed capacity /(10^4kw)	power generation /(10^8kwh)	GDP /(10^8¥)
2013	208206.5	92381.9	44460.7	564486.9	2017	301506.9	139391.7	66359	985783
2014	228399.5	102388.1	49142.5	648911.9	2018	330748.6	154489.7	73346.8	1133217.3
2015	250550.8	113478.1	54317.3	745963.6	2019	362826.3	171223	81070.3	1302701.9
2016	274850.5	125769.3	60037	857530.3	2020	398015.1	189768.7	89607.2	1497534.8

On the basis of the data in TABLE 4, using combination forecasting model can predict the SO_2 emissions from 2013to 2020, the results are shown in TABLE 5.

TABLE 5: SO_2 emissions projections from 2013 to 2020

Year	SVM model	RBF neural network model	GA Combination forecasting model	Year	SVM model	RBF neural network model	GA combination forecasting model
2013	820.4	844.6	842.4	2017	740.7	874.2	853.8
2014	728.9	759.6	756.7	2018	788.9	965.4	937.5
2015	668.3	700.2	697.3	2019	941.6	1043.2	1027.4
2016	688.4	713.7	711.9	2020	1141.8	1138.7	1139.9

The predicted data shows that the trend of SO_2 emissions from the year 2013 to 2020 will decline first and then rise up, and it will have a high emission level. The related predict shows that the SO_2 emissions in 2020 should be around 20 million tons[6,7]. In the last ten years, SO_2 emissions of power industry accounts for 56% of the country's total emissions. Thus, thermal power industries' SO_2 emissions of 2020 are roughly 11.2 million tons, which is similar to the combination forecast. Although it is not precise enough, but it can roughly verify the accuracy and reasonableness of the combined forecasting model.

At present, our country is mainly dominated by thermal power, which will lead to that the power industry will be a major source of SO_2 emissions for a long period time. Therefore, our country should strengthen the monitoring of thermal power industries' SO2 emissions, and strict emission standards as atmospheric pollutants. In addition to improve the reduction processing of waste treatment equipment, the more critical method is to alleviate the thermal power supply pressure by developing hydropower, wind power and nuclear power, which will reduce the fossil fuels consumption, especially coal consumption of electric power industry. Wish the SO_2 emissions of Chinese thermal power industry controlled in a more reasonable level in the future.

CONCLUSIONS

This paper established a combination forecasting model based on support vector machine and radial basis function neural network, which can combine the advantages of the two single prediction models. The weights of the model are obtained by using genetic algorithm, which can improve the forecasting accuracy. On the basis of the historical data of thermal power industry during 1991 to 2012, it respectively uses the SVM model, RBF neural network model and combined forecasting model based on genetic algorithm to predict the SO_2 emissions in Chinese thermal power industry. The prediction result is good, and verifies the accuracy and reasonable of the combination forecasting model. Finally, SO_2 emissions during 2013 to 2020 are forecasted by the trained model. This model is a multi-variable combination forecasting model, the predicted results of which are more accurate. It's an effective method for the prediction of SO_2 emissions of thermal power industry in our country in the future.

REFERENCES

[1] Dang YG, Liu SF, et al.; Research on grey prediction and decision model, Sciences Press **(2009)**.

[2] Ji AB, Sun JP, Pang JH; The research and application of combination forecasting method based on support vector machine, Statistics and Decision, **3 (2005)**.

[3] Shi F, Wang XC, Yu L, et al.; The MATLAB neural network analysis of 30cases, Beijing: Beijing University of Aeronautics and Astronautics Press **(2010)**.

[4] Wang SS, Chen A, Su J, et al.; Combination forecasting model applied in Chinese GDP prediction, Journal of Shandong University (Natural Science), **44(2)**, 56-59 **(2009)**.

[5] Zhou XM; Research on sulfur dioxide emissions prediction of thermal power industry, North China Electric Power University **(2011)**.

[6] Han GG, Liu XY, Liu S, et al.; Feasibility analysis of Chinese SO_2 emissions controlled under 20.7 million ton, Electric Power Technology and Environmental Protection, **27(4)**, 15-20 **(2011)**.

[7] Han GG, Liang P, Han ZY, et al.; Development of SO_2 emissions prediction and countermeasures research of Chinese energy conservation and emissions reduction in 2020, Electric Power Technology and Environmental Protection, **28(2)**, 11-12 **(2012)**.