Application of carbon emissions prediction using least squares support vector machine based on grid search

WEI SUN, YI LIANG, YANFENG XU
Department of Economy Management
North China Electric Power University
CHINA
bdsunwei@126.com, liangyibd@126.com, ncepuxyf@163.com

Abstract:-The rapid development of the global economy has brought a high-speed increase in carbon emissions, which is the primary cause of the greenhouse effect, thus how to effectively alleviate the economic and ecological threats caused by greenhouse effect, as well as studies about carbon emissions mechanism and related fields are all of great urgency. This paper puts forward a new method for carbon emissions prediction--least squares support vector machine based on grid search, where grid search is used to optimize the regularization parameter and kernel width parameter in least squares support vector machine. Then choose population, per capita GDP, energy consumption per unit GDP, urbanization rate, the proportion of coal consumption and proportion of value added services these six influencing factors as independent variables, for the aim to predict China's annual carbon emissions. The results show that the mean absolute percentage error of this new method is 1.61%, superior to that of BP neural network model, illustrating the proposed can be effectively used to predict the carbon emissions.

Key-Words: -Carbon emissions prediction; Least squares support vector machine; Grid search optimization

1 Introduction

Carbon emissions is an important cause of the more and more severe greenhouse effect. Therefore, whether the fossil energy is increasing presenting the trend of supply shortage, or greenhouse effect brings ecological threat and economic loss, these require all countries to accelerate the research on the mechanism and influencing factors of carbon emissions. Currently, the research of carbon emissions prediction model can be divided into two modes: One is the direct construction mode, namely construct Kaya equation[1], IPAT[2] or STIRPAT model[3] to predict carbon emissions based on the relationship between carbon emissions and its influencing factors. The other one is construction mode, namely construct input-output[4], MARKAL-MACRO[5], CGE[6] and other hybrid energy economy model to predict carbon emissions based on the relationship between carbon emissions and macro-economy, energy consumption,and sector technology and so on. Mixed construction mode has diversified research purpose. Carbon emissions is usually calculated by the application of carbon emissions coefficient based on energy consumption forecast, which requires a lot of technical data of each department. This paper selects direct construction mode to predict carbon emissions in China.

In recent years, with the quickly change of computer change and the rise of artificial intelligence, prediction methods have entered the stage of using intelligent forecasting techniques. BP neural network[7] has been widely used in the prediction field, it has good nonlinearity and self-learning ability, but often appears the disadvantages of easy to shock, slow convergence, easy to fall into local minimum value and so on. support vector machine(SVM) [8,9] is a very powerful machine learning method on the basis of statistical learning theory.It has strong generalization ability, and can solve the small sample, nonlinear problems such as short-term load forecasting. Least squares support vector machine(LSSVM) [10] is an extension of standard SVM. It transforms quadratic optimization problem with inequality constraints in original space into linear system with equation constraints in feature space by nonlinear mapping, which improves the speed of solving problem and accuracy of convergence. Different selections of parameters in SVM have a great impact on the fitting accuracy and generalization of the model. This paper select grid research to optimize the parameters in LSSVM, and establish LSSVM prediction model of carbon emissions based on grid research. Through case studies, it has proved that the model has high prediction accuracy.

2 LSSVM based on grid research

2.1 LSSVM

LSSVM is an improvement of SVM. It changes the inequality constraints of traditional SVM into equality constraints,and it uses square error sum loss function as experience loss of training set, so that it can transform quadratic programming problems into linear equation problems, which improves the speed of solving problems and accuracy of convergence.

N samples and their values are expressed as $D=\{(x_i,y_i)|\,i=1,2,\cdots,l\}$, $x_i \in R^l$, $y_i \in R$,

Which x_i is input data, y_i is output data. The function estimation problem in ω weight space is described as follows:

$$\min \quad \frac{1}{2}\omega^T\omega+\frac{1}{2}C\sum_{i=1}^{l}\xi_i^{\,2} \tag{1}$$

$$s.t. \quad y_i=\omega^T\varphi(x_i)+b+\xi_i\,,\ i=1,2,\cdots,l \tag{2}$$

Here $\varphi(\cdot)$ is nonlinear mapping function, namely the training data are mapped into a high-demensional linear feature space by this function.C is regularization parameter, $\xi \in R$ is error variable, b is deviation. Lagrange function is defined according to the Eq.(1) and Eq.(2):

$$L(\omega,b,x,a)=\frac{1}{2}\omega^T\omega+C\sum_{i=1}^{l}\xi_i^{\,2} -\sum_{i=1}^{l}a_i\left[\omega^T\varphi(x_i)+b+\xi_i-y_i\right] \tag{3}$$

Where a_i is Lagrange multiplier.

According to Karush-Khun-Tucker (KKT) conditions, calculate the partial derivatives on ω, b, ξ_i, a ,and make them zero. We can get

$$\begin{cases}\omega=\sum_{i=1}^{l}a_i\varphi(x_i)\\ \sum_{i=1}^{l}a_i=0\\ a_i=C\zeta_i\\ \omega^T\varphi(x_i)+b+\zeta_i-y_i=0\end{cases} \tag{4}$$

According to Eq.(4),the optimization problem is changed into solving the linear equation.

$$\begin{bmatrix}b\\a_1\\\vdots\\a_l\end{bmatrix}=\begin{bmatrix}0 & 1 & \cdots & 1\\ 1 & K(x_1,x_1)+\frac{1}{C} & \cdots & K(x_1,x_l)\\ 1 & \vdots & \vdots & \vdots\\ 1 & K(x_1,x_l) & \cdots & K(x_l,x_l)+\frac{1}{C}\end{bmatrix} \times\begin{bmatrix}0\\y_1\\\vdots\\y_l\end{bmatrix} \tag{5}$$

Finally the LSSVM model is as follows:

$$y(x)=\sum_{i=1}^{l}a_iK(x,x_i)+b \tag{6}$$

Here $K(x_i, x_j) = \varphi(x_i)^T \cdot \varphi(x_j)$ is symmetric function which meets Mercer conditions, usually referred as kernel function.The kernel function in this paper uses Gaussian(RBF) kernel

$$K(x, x_i) = e^{-\frac{(x-x_i)}{2\sigma^2}} \quad (7)$$

Where σ^2 is the kernel width of kernel function.

For the RBF kernel function LSSVM, parameters include regularization parameter C and kernel width σ. Parameter C is compromise between structural risk and sample error. The value of the parameter C is associated with tolerable error. The larger C allows smaller errors, the smaller C allows larger errors. The kernel width σ is related to the input spatial extent or width of learning samples. The bigger the sample input space is, the greater the value is. Conversely, the less the sample input space is,the smaller the value is.

2.2 Parameter optimization based on grid search

(1) Cross validation(CV)

CV is a statistical method used to verity the classifier performance.Its basic idea is to divide original data into two parts in some sense, ones are used as training set, the other are used as validation set. First use training set to train the classifier, then use validation set to test the trained model, the index of evaluation of the classifier performance is gotten. Usually people use K-fold CV: Divide the raw data into K groups(usually average), each subset of the data will respectively be used as validation set once, the rest of K-1 subsets of the data are used as training sets. Then we can get K models. The average of classification accuracy of K models' validation sets acts as the classifier performance index under K-CV. K is generally more than or equal to 2, which is selected from 3 in actual operation, only when the amount of original data set is small, it's taken 2.K-CV can effectively avoid over-learning and less-learning, and its final result is also more persuasive.

(2) Grid search

The basic principle of grid search is to make parameters to be optimized divide grid in a certain range and take their values through all points in the grid. For the given parameters, select K-CV method to obtain the accuracy of classification of the training set verification in this set of parameters. Finally, take it as the best set of parameters which makes the accuracy of classification of the training set verification highest.

3 Determination of carbon emissions influencing factors

STIRPAT model studies environment random effects caused by population,wealth and technology based on regression methods,in which the factors such as population, wealth and technology can be extended. Considering relevant factors of the model, select the six influencing factors population, per capita GDP, unit GDP, urbanization rate, coal consumption proportion, service industry increased value proportion as independent variables of carbon emissions prediction.Among them, population and urbanization rate respectively reflect the influence of total population and the urban population(or urbanization), per capita GDP and service industry increased value proportion respectively reflect the influence of per capita wealth and service industry wealth(or economic service), unit GDP and coal consumption proportion respectively reflect the influence of energy conservation and the clean use of technology.

4 LSSVM carbon emissions prediction model based on grid search

As we know the sample made up of N years carbon emissions influencing factors and carbon emissions data (x_i, y_i), $i = 1,2,\cdots,n$, construct carbon emissions GS-LSSVM model.

(1) Made normalization processing for the independent variables and the dependent variables of training samples in accordance with the following equations:

$$x_{il} = \frac{x_{il} - \min\limits_{i=1,2,\cdots n} x_{il}}{\max\limits_{i=1,2,\cdots n} x_{il} - \min\limits_{i=1,2,\cdots n} x_{il}},$$

$$i = 1,2,\cdots n, l = 1,2,\cdots,6 \quad (8)$$

$$y_i = \frac{y_i - \min\limits_{i=1,2,\cdots n} y_i}{\max\limits_{i=1,2,\cdots n} y_i - \min\limits_{i=1,2,\cdots n} y_i},$$

$$i = 1,2,\cdots n \quad (9)$$

So that all the data are in [0,1].

(2) Select m samples as training samples,other n-m samples as testing samples. For training samples, select Gaussian (RBF) Kernel to construct LSSVM model as Eq.(7) shows. For regularization parameter C and kernel width σ of LSSVM, determine them by grid search and cross validation: The ranges of $\log_2 C$ and $\log_2 \sigma$ are divided into several grids,then divide all samples into k groups,namely apply k to carry out cross validation. Fix the parameter (C,σ) on the grid,then take $k-1$ groups as training samples in order and put them into the model to obtain optimal solution and regression function. Put the rest set of sample into regression function to get fitted values,then calculate the errors between them and actual values, so that mean square errors(MSE) of n samples of all k groups. MSE is calculated as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(R_i - P_i)^2 \quad (10)$$

Where R_i is actual value, P_i is search value.

(3) Seek the parameter whose MSE is minimum through all parameters (C,σ) in the grid. The parameter (C^*,σ^*) which makes the error minimal is the optimal one.

(4) Put the optimal parameter C^*, σ^* and training samples into LSSVM model so that the trained carbon emissions prediction model is obtained. Then put the testing samples into the trained model to predict and get the predictive value $\hat{y}_i$.

(5) Make normalization processing for influencing factors x_i in testing samples, then put them into trained carbon emissions prediction model to obtain output results y_i^0 and make anti-normalization for them:

$$\hat{y}_i = y_i^0 \cdot \left(\max_{i=1,2,\cdots,n} y_i - \min_{i=1,2,\cdots,n} y_i\right) + \min_{i=1,2,\cdots,n} y_i \quad (11)$$

The predicted value of carbon emissions can be obtained.

5 Case Study

Using the proposed model, this paper selects China's energy consumption and related data from 1980 to 2012 for analysis.

(1) Data collection

Obtain the total energy consumption and component percentage of different kinds of energy over the years of China, which are recorded according to the standard coal, from "China statistical yearbook 2013", then convert these data into absolute number of different kinds of energy consumption, as shown in Table 1.

Table 1 The total energy consumption and compositions

Year	Total energy consumption/ 10000 tons of SCE	Energy consumption of different kinds/10000 tons of SCE			
		Coal	Petroleum	Natural gas	Hydro power, nuclear power, wind power
1980	60275	43518.55	12476.93	1868.53	2411.00
1981	59447	43039.63	14029.49	1605.07	772.81
1982	62067	45619.25	14027.14	1551.68	868.94
1983	66040	48935.64	14528.80	1584.96	990.60
1984	70904	52965.29	15386.17	1559.89	992.66
1985	76682	58124.96	13112.62	1687.00	3757.42
1986	80850	61284.30	13906.20	1859.55	3799.95
1987	86632	66013.58	14727.44	1819.27	4071.70
1988	92997	70863.71	15809.49	1952.94	4370.86
1989	96934	73669.84	16575.71	1938.68	4749.77
1990	98703	75211.69	16384.70	2072.76	5033.85
1991	103783	78978.86	17746.89	2075.66	4981.58
1992	109170	82641.69	19104.75	2074.23	5349.33
1993	115993	86646.77	21110.73	2203.87	6031.64
1994	122737	92052.75	21356.24	2332.00	6996.01
1995	131176	97857.30	22955.80	2361.17	8001.74
1996	135192	99366.12	25280.90	2433.46	8111.52
1997	135909	97039.03	27725.44	2446.36	8698.18
1998	136184	96554.46	28326.27	2451.31	8851.96
1999	140569	99241.71	30222.34	2811.38	8293.57
2000	145531	100707.45	32307.88	3201.68	9313.98
2001	150406	102727.30	32788.51	3609.74	11280.45
2002	159431	108413.08	35553.11	3826.34	11638.46
2003	183792	128286.82	38963.90	4594.80	11946.48
2004	213456	148351.92	45466.13	5336.40	14301.55
2005	235997	167085.88	46727.41	6135.92	16047.80
2006	258676	183918.64	49924.47	7501.60	17331.29
2007	280508	199441.19	52735.50	9256.76	19074.54
2008	291448	204887.94	53334.98	10783.58	22441.50
2009	306647	215879.49	54889.81	11959.23	23918.47
2010	324939	220958.52	61738.41	14297.32	27944.75
2011	348002	238033.37	64728.37	17400.10	27840.16
2012	361732	240913.51	68005.62	18810.06	34002.81

Compute standard coal data into corresponding carbon emissions, using different carbon emissions coefficients according to the type of the original energy. And the coefficients are selected from the EIA, updated fastest for now, as shown in Table 2.

Table 2 Carbon emissions coefficients of different energies

Energy type	Coal	Petroleum	Natural gas	Hydro power, nuclear power
Carbon conversion coefficient /(t(C)/t)	0.702	0.478	0.389	0

The specific data of carbon emissions calculated by carbon conversion coefficients and six influencing factors mentioned above (population, per capita GDP, energy consumption per unit GDP, urbanization rate, the proportion of coal consumption and proportion of value added services) are listed in Table 3.

Table 3 Carbon emissions and its influencing factors

Year	carbon emissions/ 10000 tons	Population/ 100 million	Per capita GDP/ RMB	Energy consumption per unit GDP/Tce per 10000 yuan	Urbanization rate/%	Proportion of coal consumption/ %	Proportion of value added services/%
1980	37240.85	9.87	463	13.26	19.39	72.2	21.60
1981	37544.29	10.01	492	12.15	20.16	72.7	22.01
1982	39333.29	10.17	528	11.66	21.13	73.7	21.85
1983	41914.14	10.3	583	11.08	21.62	74.2	22.44
1984	45143.02	10.44	695	9.84	23.01	75.3	24.78
1985	47727.80	10.59	858	8.51	23.71	75.8	28.67
1986	50392.11	10.75	963	7.87	24.52	75.8	29.14
1987	54088.95	10.93	1112	7.18	25.32	76.2	29.64
1988	58062.96	11.1	1366	6.18	25.81	76.1	30.51
1989	60393.57	11.27	1519	5.70	26.21	76.1	32.06
1990	61436.79	11.43	1644	5.29	26.41	76.2	31.54
1991	64733.61	11.58	1893	4.76	26.94	76.1	33.69
1992	67953.41	11.72	2311	4.05	27.46	75.7	34.76
1993	71774.26	11.85	2998	3.28	27.99	74.7	33.72
1994	75736.46	11.99	4044	2.55	28.51	75.0	33.57
1995	80587.19	12.11	5046	2.16	29.04	74.6	32.86
1996	82785.90	12.24	5846	1.90	30.48	73.5	32.77
1997	82325.79	12.36	6420	1.72	31.94	71.4	34.17
1998	82274.75	12.48	6796	1.61	33.35	70.9	36.23
1999	85207.59	12.58	7159	1.57	34.78	70.6	37.77
2000	87385.25	12.67	7858	1.47	36.22	69.2	39.02
2001	89191.66	12.76	8622	1.37	37.66	68.3	40.46
2002	94588.82	12.85	9398	1.32	39.09	68.0	41.47
2003	110469.47	12.92	10542	1.35	40.53	69.8	41.23
2004	127951.72	13.00	12336	1.34	41.76	69.5	40.38
2005	142016.86	13.08	14185	1.28	42.99	70.8	40.51
2006	155892.90	13.14	16500	1.20	44.34	71.1	40.94
2007	168816.17	13.21	20169	1.06	45.89	71.1	41.89

2008	173520.27	13.28	23708	0.93	46.99	70.3	41.82
2009	182436.87	13.35	25608	0.90	48.34	70.4	43.43
2010	190185.50	13.41	30015	0.81	49.95	68.0	43.11
2011	204808.23	13.47	35198	0.74	51.27	68.4	42.96
2012	208945.08	13.54	38420	0.70	52.57	66.6	44.63

(2) parameter optimization

According to the build steps of GS - LSSVM model for carbon emissions, what to do first is to do normalization pretreatment upon 32 sample data. Select the former 20 sets of data as the training samples, the remaining 12 sets of data as the test ones. The range of parameter $\log_2 C$ and $\log_2 \sigma$ is set between $[-10,10]$, and mesh width is 0.4. Five-fold cross validation is carried out on the training samples. Then the contour map and 3D view plot of grid search process are respectively shown in Fig. 1 and Fig. 2.

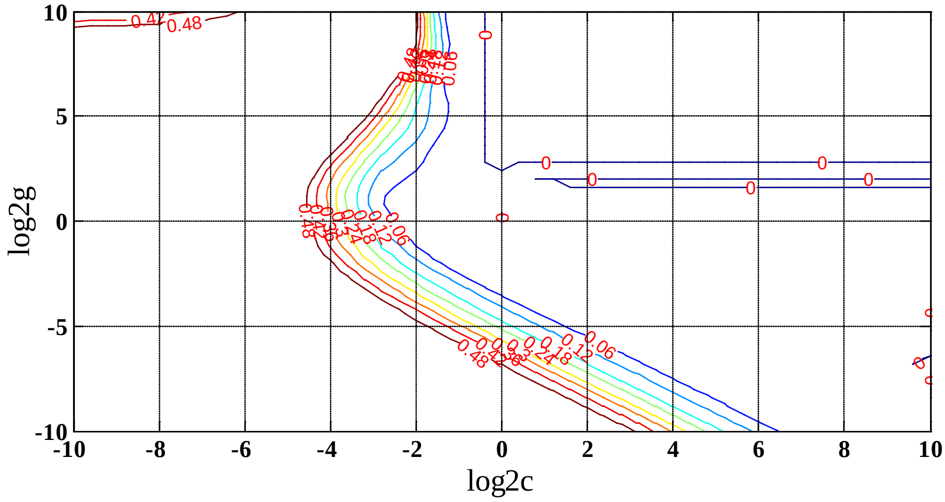

Fig. 1 Contour map of Grid search

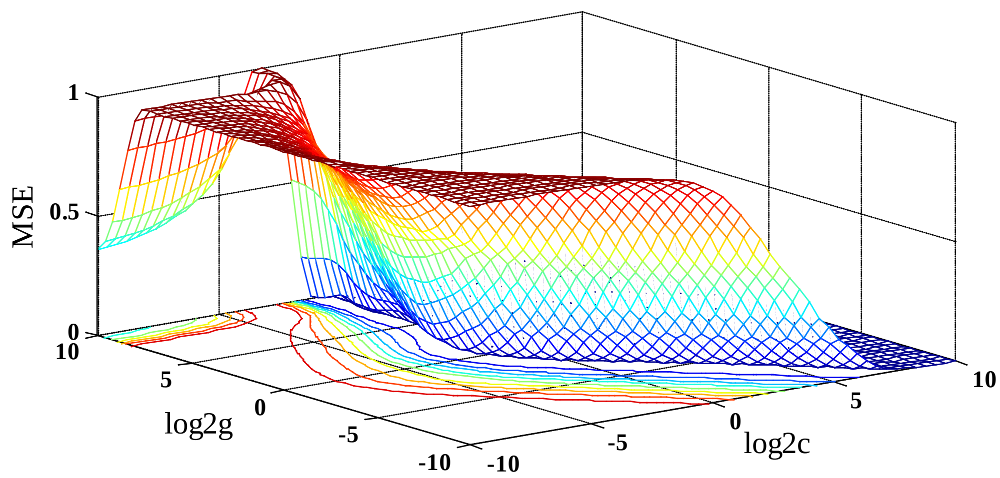

Fig. 2 3D view plot of Grid search

Therefore, the best value of regularization parameter C is 1.32 and kernel width parameter σ is 0.76, and mean square error of cross-validation is 0.00035138.

(3) Training and Testing

Set the parameters of LSSVM: $C = 1.32$ and $\sigma = 0.76$. And 20 sets of training samples are trained by LSSVM, the training results as shown in Fig.3.

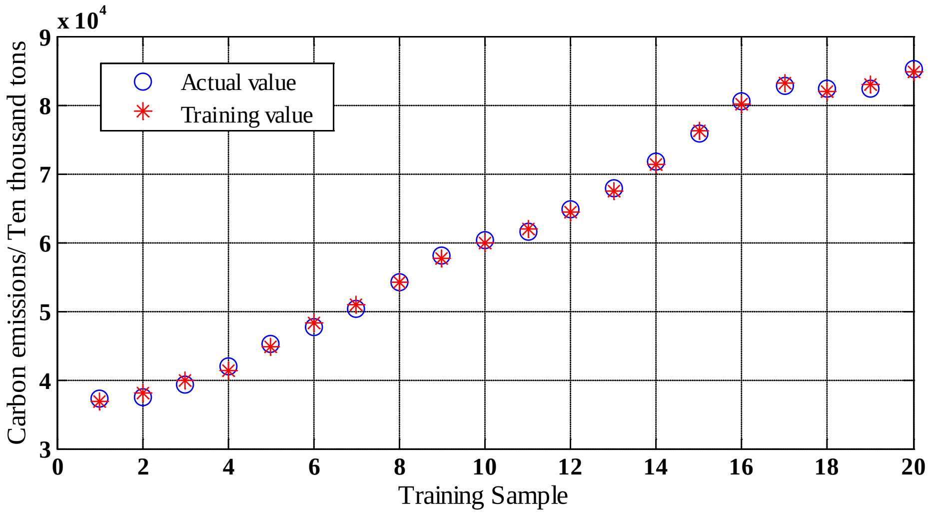

Fig. 3 Training results of LSSVM

As can be seen from Fig. 3, among 20 sets of training samples there are 15 relative errors of training results are less than 1%, the rest 5 are also maintained within 2%, which can describe the training result is satisfactory.

The remaining 12 sets of test samples are used to predict by being substituted into the trained LSSVM model. In order to reflect the GS-LSSVM model has higher accuracy and stronger generalization ability more objectively, the widely used BP neural network prediction model is introduced in this paper as a comparison to the same test. So parameters of BP model are set in Table 4.

Table 4 Parameter setting of BP neural network

Parameters	Value
The number of hidden layer	1
Hidden layer nodes	6
Maximum number of training iterations	1000
Error accuracy	0.0001
Learning rate	0.5

This paper introduces three kinds of error calculation method for analysis including Eq.(10).

$$RE_i = \frac{P_i - R_i}{R_i} \times 100\% \tag{12}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{P_i - R_i}{R_i}\right| \times 100\% \tag{13}$$

Where P_i is the predicted value, and R_i is the actual value.

The test results are shown in Table 5.

Table 5 Test results

Test sample number	Actual value	GS-LSSVM		BPNN	
		Output value	RE (%)	Output value	RE (%)
1	87385.25	87682.20	0.34	87887.08	0.57
2	89191.66	90371.11	1.32	91350.27	2.42
3	94588.82	96171.65	1.67	97262.32	2.83
4	110469.47	114187.74	3.37	115849.29	4.87
5	127951.72	123677.01	-3.34	122068.25	-4.60
6	142016.86	140335.09	-1.18	143500.09	1.04
7	155892.90	152683.28	-2.06	157083.59	0.76
8	168816.17	167076.87	-1.03	167669.55	-0.68
9	173520.27	177902.43	2.53	170823.94	-1.55
10	182436.87	184145.38	0.94	176277.80	-3.38
11	190185.50	190568.94	0.20	179721.78	-5.50
12	204808.23	201993.98	-1.37	181795.94	-11.24
MSE		6869000		63734000	
MAPE(%)		1.61		3.29	

(4) Analysis of results

As can be seen from Table 5, relative errors (RE) of the GS-LSSVM model are controlled within 3.5% for all 12 sets of test samples, among which the minimum is 0.20% and the maximum 3.37% while the smallest RE of BP neural network model is 0.57%, and the biggest is 11.24%. On the whole, there are eight sample points from GS-LSSVM whose prediction error is smaller than that of BP model. Besides, mean absolute percentage error (MAPE) of GS-LSSVM is 1.61, less than 3.29 obtained from BP, which can explain GS-LSSVM has a higher prediction precision. In addition, mean squared error (MSE) of GS-LSSVM is 6869000, far below that of BP, 63734000, indicating the degree of error floating of GS-LSSVM is lower than that of BP, i.e. it has a better generalization ability.

6 Conclusions

For the problem of carbon emissions forecast, this paper introduces the current widely used intelligent algorithms and proposes a new method--least squares support vector machine based on grid search optimization algorithm(GS-LSSVM), in which grid search is used to optimize regularization parameter and kernel width parameter of LSSVM. The prediction model contains population, per capita GDP, energy consumption per unit GDP, urbanization rate, the proportion of coal consumption and proportion of value added services these six factors affecting carbon emissions. Through the prediction for China's annual carbon emissions, it is proved that this model has a good performance in terms of prediction accuracy and generalization ability, superior to the BP neural network forecasting model. Therefore it can be effectively applied to the prediction of the carbon emissions.

References:

[1] Juan Antonio Duro, Weighting vectors and international inequality changes in environmental indicators: An analysis of CO2 per capita emissions and Kaya factors, *Energy Economics*, Vol. 39, 2013, pp.122-127.

[2] Janis Brizga, Kuishuang Feng, Klaus Hubacek, Drivers of CO2 emissions in the former Soviet Union: A country level IPAT analysis from 1990 to 2010, *Energy*, Vol.59, No.15, 2013, pp. 743-753.

[3] Ping Wang, Wanshui Wu, Bangzhu Zhu,

Yiming Wei, Examining the impact factors of energy-related CO2 emissions using the STIRPAT model in Guangdong Province, China, *Applied Energy*, Vol.106, 2013, pp. 65-71.

[4] G.Q. Chen, Shan Guo, Ling Shao, J.S. Li, Zhan-Ming Chen, Three-scale input-output modeling for urban economy: carbon emissions by Beijing 2007, *Communications in Nonlinear Science and Numerical Simulation*, Vol.18, No.9, 2013, pp. 2493-2506.

[5] Wenying Chen, The costs of mitigating carbon emissionss in China: findings from China MARKAL-MACRO modeling, *Energy Policy*, Vol.33, No.7, 2005, pp. 885-896.

[6] Panida Thepkhun, Bundit Limmeechokchai, Shinichiro Fujimori, Toshihiko Masui, Ram M. Shrestha, Thailand's Low-Carbon Scenario 2050: The AIM/CGE analyses of CO2 mitigation measures, *Energy Policy*, Vol.62, 2013, pp. 561-572.

[7] Yanxi Zhang, Xiangdong Gao, Seiji Katayama, Weld appearance prediction with BP neural network improved by genetic algorithm during disk laser welding, *Journal of Manufacturing Systems*, Vol.34, 2015, pp. 53-59.

[8] Vapnik V. *The nature of statistic learning theory*. New York: Springer-Verlag;1995.

[9] Vapnik V. *Statistical learning theory*. New York: Wiley; 1998.

[10] Wei Sun, Yi Liang, Research of least squares support vector regression based on differential evolution algorithm in short-term load forecasting model, *Journal of Renewable and Sustainable Energy*, Vol.6, No.5, 2014, pp. 053137.