Pol. J. Environ. Stud. Vol. 28, No. 6 (2019), 4391-4401
DOI: *10.15244/pjoes/94619* ONLINE PUBLICATION DATE: 2019-08-01

Original Research

Predicting and Analyzing CO_2 Emissions Based on an Improved Least Squares Support Vector Machine

Wei Sun[1], Hongyuan Jin[1]*, Xiumei Wang[2]

[1]Department of Business Administration, North China Electric Power University, Baoding, China
[2]Engineering Training Center, North China Electric Power University, Baoding, China

Received: 7 July 2018
Accepted: 26 August 2018

Abstract

Carbon dioxide (CO_2) is the main cause of the greenhouse effect. With the rapid development of the economy in China, CO_2 emissions have increased dramatically. To reduce CO_2 emissions, ensure the sustainability of China's economy and implement the Paris International Convention, it is important to investigate the main factors affecting CO_2 emissions and use those factors to accurately forecast CO_2 emissions. In order to achieve accurate prediction of CO_2, this paper proposes a CO_2 emission prediction model based on principal component analysis (PCA) and particle swarm optimization for least squares support vector machine (PSO-LSSVM). Through data 1990-2016 in Hebei Province of China, this paper identifies 24 influencing factors though the bivariate correlation analysis. After applying PCA to reduce the dimensions of the influencing factors, two principal components were extracted as input variables. Then the parameters of the LSSVM model are obtained by PSO and the forecast model is established. By comparing the prediction results with actual values, it is proved that the prediction error of the PSO-LSSVM prediction model is 0.663%, which is smaller than that of the traditional BPNN and LSSVM models.

Keywords: CO_2 emissions, influencing factors, PCA, PSO-LSSVM model

Introduction

The CO_2 effect brought about by the consumption of fossil energy is becoming more and more obvious. The emission of greenhouse gases has caused serious environmental problems such as global warming, and also has caused huge economic losses. Urban areas and the rapid progress of industrialization and technology are leading to serious air pollution in urban areas. In particular, human beings trying to maintain urban life harm health [1-6]. In Europe, more than two-thirds of the total population lives in cities. Population growth and industrialization have led to air pollution in some cities that reaches levels that threaten human health. This has become one of the most important topics of our day. Human health is affected by all air pollution, but some emissions have more severe atmospheric conditions. In particular, carbon dioxide (CO_2) and other pollutants, which provide global warming, have recently attracted attention; because CO_2 is one of the most searched gases

*e-mail: jin_hongyuan0701@163.com

[7-13]. Recent studies show PM10 and PM2.5 CO_2 air quality indices. There are a lot of studies for carbon emissions, including indoor and outdoor – especially urban areas and parks. This shows that PM2.5 affects human health.

China's primary energy consumption accounts for 23% of world consumption, of which coal consumption accounts for 62%. China is the world's largest energy consumer and CO_2 emitter. In order to solve the problem caused by the greenhouse effect, countries around the world have also made relevant efforts. In 2015, 200 countries agreed the Paris Agreement, clarifying the goals that the global community pursues. Under the agreement, China has pledged that by 2030, carbon emissions per unit of GDP will fall by 60% to 65% from 2005. Research on influencing factors analysis and CO_2 emission prediction will help the government extract the main factors influencing CO_2 emissions and use them to guide CO_2 reduction efforts [14]. In recent years, research on CO_2 emission factors has mainly related to (1) exponential decomposition analysis methods, such as the generalized Fisher Index (GFI) decomposition analysis, logarithmic mean index Divisia decomposition method, and Laspeyres decomposition method [15-17]. Decomposition methods such as LMDI can improve the accuracy of calculations. In the processing data, it has the characteristic of no residual decomposition, simple calculation process and visual decomposition result; (2) structural decomposition analysis methods, including input-output method [18]; and (3) econometric method methods, including Granger causality analysis [19], factor analysis [20], provincial panel data analysis [21], etc. This kind of statistical analysis method, in the general sense, has a wide range of analysis and can handle various situations flexibly, but has many influencing factors. The areas covered by the CO_2 emissions forecasting analysis include data from various departments and regions, such as the industrial sector [22], the transportation sector [23], and the steel industry [24], as well as multi-regional research [25].

The research on CO_2 emissions prediction is mainly on the establishment of different models. The traditional predictive model is computationally complex and prediction accuracy depends on historical data such as the establishment of regression prediction models. Xu and Lin [26] use vector autoregressive models to analyze the influencing factors of industrial CO_2 emissions and show that energy efficiency plays a leading role in CO_2 emissions. Wang et al. [27] adopted a partial least-squares regression model. The results show that the level of urbanization, economic level, and industry share have positive effects on CO_2 emissions. Meng and Niu [28] used a logistic model to predict CO_2 emissions in various industries. Rigoberto et al. [29], based on the extended environmental Kuznets curve (EKC) and environmental logistics curve (ELC) combined with the logistic model, empirical samples of 175 countries, and CO_2 emissions prediction. Zhang et al. [30] used the extended STIRPAT model to analyze and predict CO_2 emissions in Henan Province, as well as the traditional forecasting models such as time series models, Holt-winters models, and smoothing index forecasting models. However, the disadvantage is that it depends heavily on past data and is suitable for the prediction of horizontal data. Due to giving a larger weight in the near future and a smaller proportion in the long-term, it is suitable for short-term forecasting.

In recent years, artificial intelligence technology has been widely applied. Compared with the above models, the artificial intelligence prediction and optimization model has improved the accuracy and calculation speed. For example, Wang and Dang [31] has applied the gray model GM (1, 1) to the prediction of Jiangsu's CO_2 emissions. Although grey prediction requires less sample data and does not require the exploration of internal mechanisms, the model prediction accuracy is not high. Chen and Ye [32] used artificial neural networks (ANN) to predict CO_2 emissions and estimated the amount of CO_2 emitted by the global reservoir. Because the neural network algorithm converges very slowly, it is easy to fall into a local minimum. Thus, Sun et al. [33] proposed the genetic algorithm (GA) optimized back propagation neural network (BPNN) to predict CO_2 emissions, selected multiple influencing factors, and used autocorrelation and partial correlation to analyze CO_2 emissions. The hybrid algorithm improves the speed of convergence. And Wen and Liu [34] use the particle swarm optimization neural network (IPSO-BP) to show that the particle swarm optimizes the BP initial connection weights and thresholds. The algorithm can fully utilize the PSO's global search capability and BP's local search ability. Zhao et al. [35] combined the mixed data sampling regression model with BPNN to predict CO_2 emissions. Literature [36] proposed extreme learning machine (ELM) to overcome the shortcomings of the BP neural network. This method not only reduces the risk of falling into local optimum, but also greatly improves the learning speed and generalization ability of the network. However, when the extreme learning machine's initial function was selected improperly, the extrapolation ability was poor. Therefore, Sun et al. [37] improved the particle swarm optimization limit learning machine, which optimizes the input weights and biases and improves the generalization ability. It also shows that selecting more carbon dioxide emission factors can more fully study CO_2 emissions. The literature [38] proposed a support vector machine (SVM) to avoid local optimization problems. Using an SVM prediction model to study the inevitable link between CO_2 emissions and economic growth, the SVM algorithm is difficult to achieve for large sample data calculation training. There are difficulties in solving the problem of multivariate planning. Therefore, the literature [39] improved LSSVM to study the effects of the three major industries and household consumption on CO_2 emissions. It shows that the classification analysis has higher accuracy than the unclassified prediction.

Based on the above research, it is observed that a number of selected influencing factors and different classifications affect prediction accuracy. In previous studies, decomposition methods or classification method were applied to avoid the problem of including too many influencing factors, which may lead to a decrease in the accuracy of CO_2 emissions forecast. However, the factors of CO_2 emissions are numerous and it is difficult to determine which factor is influential intuitively.

To solve this problem, this paper presents a method based on principal component analysis (PCA) and an improved least squares support vector machine (LSSVM) prediction. The PCA method converts the pre-selected influencing factors for dimension reduction. Through PCA process, fewer components are selected to represent most of the original information as input factors. With fewer inputs, the LSSVM can be solved quickly by PSO and avoid the possibility of PSO procedure falling into a local optimum. Through the combination with PSO-LSSVM, the CO_2 emission prediction model can be established, avoiding the difficulties in the selection of impact factors, strong collinearity between them and low accuracy of prediction results, etc.

Material and Methods

This paper applies the CO_2 emission prediction model of PSO-LSSVM. First, the PCA is applied to screen main factors as input factors, then the PSO is applied to optimize the LSSVM model, and the parameters are trained using historical data to obtain the optimal penalty and nuclear parameters to obtain the final Prediction model.

Principal Component Analysis

Principal component analysis is a statistical method that aims to reduce dimensions and eliminate multi-collinearity, combining multiple correlation variables into one or several variables. Orthogonal transformation transforms a set of p-dimension variables $(X_1, X_2,..., X_p)^T = X$ with correlations into a set of p-dimension linear independent indicators y, which are independent of each other. The relationship between them is $y = a_1X_1 + \text{L} + a_pX_p = a^TX$. Where a is the coefficient vector to be determined, and the goal is to find a. The newly generated comprehensive index y has become the main component. These components should satisfy the following conditions: (1) each principal component is perpendicular to each other; and (2) the sum of the variance of each principal component is equal to the sum of the eigenvalues. In this paper, SPSS 20.0 is used to calculate the principal component whose eigenvalue is greater than 1, so that its cumulative contribution rate exceeds 85%.

Least Squares Support Vector Machine

Support vector machine (SVM) is a supervised learning algorithm that can build prediction models based on regression methods. It maps vectors into a higher dimensional space and solves the maximum interval hyperplane classification problem. Cortes and Vapnik (1995) have a great advantage in solving small samples, nonlinear problems, etc. However, it is difficult to solve large sample data, which consumes a lot of memory and computational time.

In order to solve the above problem, an improved LSSVM based on SVM is proposed by Suykens, (1999). LSSVM proposes a norm in the objective function of the optimization problem so that (1) the original inequality constraints are transformed into equality constraints and different loss functions are applied; (2) turning the optimization problem from quadratic programming to a linear equation solving problem, the complexity of the solution is reduced, and the convergence speed is improved [40]. Thus in this paper, the LSSVM method is applied to build the prediction model.

In the LSSVM model, the given training sample is set as $D = \{(x_n, y_n) \mid n = 1, 2,..., k\}$, $x_n \in R^k$, $y_n \in R$ where x_n is the input variable and y_n is the corresponding output variable. The training set is mapped from the input space to the feature space using the nonlinear mapping φ(·), and then the linear regression is performed in the high-dimensional feature space.

The LSSVM model can be expressed as:

$$y(x) = \omega^T \phi(x) + b \tag{1}$$

…where ω is weight and b is bias. Based on the principle of minimizing structural risks, the optimization problem is expressed as follows:

$$\min J(\omega, l) = \frac{1}{2}\omega^T\omega + \frac{\gamma}{2}\sum_{n=1}^{k} l_n^2 \tag{2}$$

$$s.t. \quad y_n = \omega^T \phi(x_n) + b + l_n, n = 1, 2, \text{L} \ , k \tag{3}$$

…where γ is the regularization parameter and l_n is the error. In order to solve the optimization problem, the Lagrange equation is constructed:

$$L(\omega, b, x, a) = J(\omega, l) - \sum_{n=1}^{k} a_n \left[\omega^T \phi(x_n) + b + l_n - y_n\right] \tag{4}$$

…where $a_n (n = 1,2,..., k)$ are the Lagrange multipliers. According to the KKT (Karush-Khun-Tucker) conditions, find the partial derivative for ω, b, l_n, a_n and make it equal to zero.

$$\begin{cases} \omega = \sum_{n=1}^{k} a_n \phi(x_n) \\ \sum_{n=1}^{k} a_n = 0 \\ a_n = \gamma l_n \\ \omega^T \phi(x_n) + b + l_n - y_n = 0 \end{cases} \tag{5}$$

With simplification after removing both ω and l_n variables, the optimization problem can be simplified to solve linear equations:

$$\begin{bmatrix} b \\ a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} 0 & 1 & \cdots & 1 \\ 1 & K(x_1, x_1) + \frac{1}{\gamma} & \cdots & K(x_1, x_1) \\ \vdots & \vdots & \vdots & \vdots \\ 1 & K(x_1, x_1) & \cdots & K(x_1, x_1) + \frac{1}{\gamma} \end{bmatrix} \tag{6}$$

Then the LSSVM regression model can finally be obtained:

$$y(x) = \sum_{n=1}^{k} a_n K(x, x_n) + b \tag{7}$$

Among them, $K(x, x_n)$ is a positive definite kernel function that satisfies Mercer's theorem. In this paper, a radial basis kernel function with strong generalization ability is applied. Its expression is as follows:

$$K(x, x_n) = e^{-\frac{(x - x_n)}{2\sigma^2}} \tag{8}$$

… where σ^2 is the kernel width of the kernel function.

Particle Swarm Optimization Algorithm

Particle swarm optimization is an intelligent algorithm proposed by Kennedy and Eberhart (1995) [41]. This algorithm is an intelligent optimization algorithm based on inter-group collaboration that observes and simulates behaviors of the birds. Each bird is modeled as an optimal solution to the search space, called "particles." Particle swarm optimization algorithm has good robustness and is easy to converge, but it is easy to fall into the local optimum. It is often used to optimize algorithms such as neural network. This paper applies the PSO algorithm to optimize the parameters of LSSVM.

In order to search for the optimal position, the fitness value is calculated by the fitness function. The current position of each particle is compared with the best position of the particle itself, and the optimal position is selected after iteration. These two optimal extremum are individual optimal and global optimal, respectively.

Assume that the particle population size is M, and the particle performs the flight search optimal solution in the D-dimensional space. The spatial position vector of the i^{th} particle is $x_i^{(t)} = x_{i1}^{(t)}, x_{i2}^{(t)}, ..., x_{id}^{(t)})^T$, and the fitness of each particle's corresponding spatial position is calculated according to the objective function. The i^{th} particle has a spatial displacement speed of $v_i^{(t)} = v_{i1}^{(t)}, v_{i2}^{(t)}, ..., v_{id}^{(t)})^T$,, an individual extremum of $p_i^{(t)} = p_{i1}^{(t)}, p_{i2}^{(t)}, ..., p_{id}^{(t)})^T$,, and a global extremum of $p_g^{(t)} = p_{g1}^{(t)}, p_{g2}^{(t)}, ..., p_{gd}^{(t)})^T$,. According to the particle fitness value, the position and velocity of each generation of particles are iterated under the following equations:

$$\upsilon_{id}^{(t+1)} = \omega \upsilon_{id}^{(t)} + c_1 r_1 \left(p_{id}^{(t)} - x_{id}^{(t)} \right) + c_2 r_2 \left(p_{gd}^{(t)} - x_{id}^{(t)} \right) \tag{9}$$

$$x_{id}^{(t+1)} = x_{id}^{(t)} + \upsilon_{id}^{(t+1)} \tag{10}$$

$$t = 1, 2, ..., n; d = 1, 2, ..., D$$

$$\omega = (\omega_s - \omega_e)(t_m - t) / t_m + \omega_e \tag{11}$$

…where ω is the weight of inertia, where a large value is better for obtaining global optimization; on the contrary, a smaller value is better for the convergence of PSO; c_1 and c_2 are acceleration constants, generally taken between [0, 2]. r_1 and r_2 are randomly distributed between [0, 1]. Additionally, to avoid blind search, $\upsilon_{id} \in [-\upsilon_{max}, \upsilon_{max}]$, $x_{id} \in [-x_{max}, x_{max}]$.

LSSVM Optimized by PSO

The optimization of LSSVM parameters mainly focus on obtaining the parameters C and σ. The steps of the proposed PSO-LSSVM are as follows:

Step 1: Particle group parameter initialization.

Step 2: Calculating particle fitness value.

Step 3: Finding individual extremes and population extremes.

Step 4: Update the particle velocity and position according to equations (10) (11).

Step 5: Check whether the condition is satisfied; if so, go to Step 6; if not, return to Step 2.

Step 6: Substitute the optimization parameters (C, σ) into the LSSVM.

The overall work flow chart for CO_2 emission prediction in this paper is shown in Fig. 1. In Part 1, the bivariate correlation and significance tests were used to study the correlation between the influencing factors and CO_2 emissions. Then the dimension of the preselected influential factors is reduced by using PCA. Part 2 includes the PSO algorithm. The third part is to build the LSSVM for CO_2 emissions prediction.

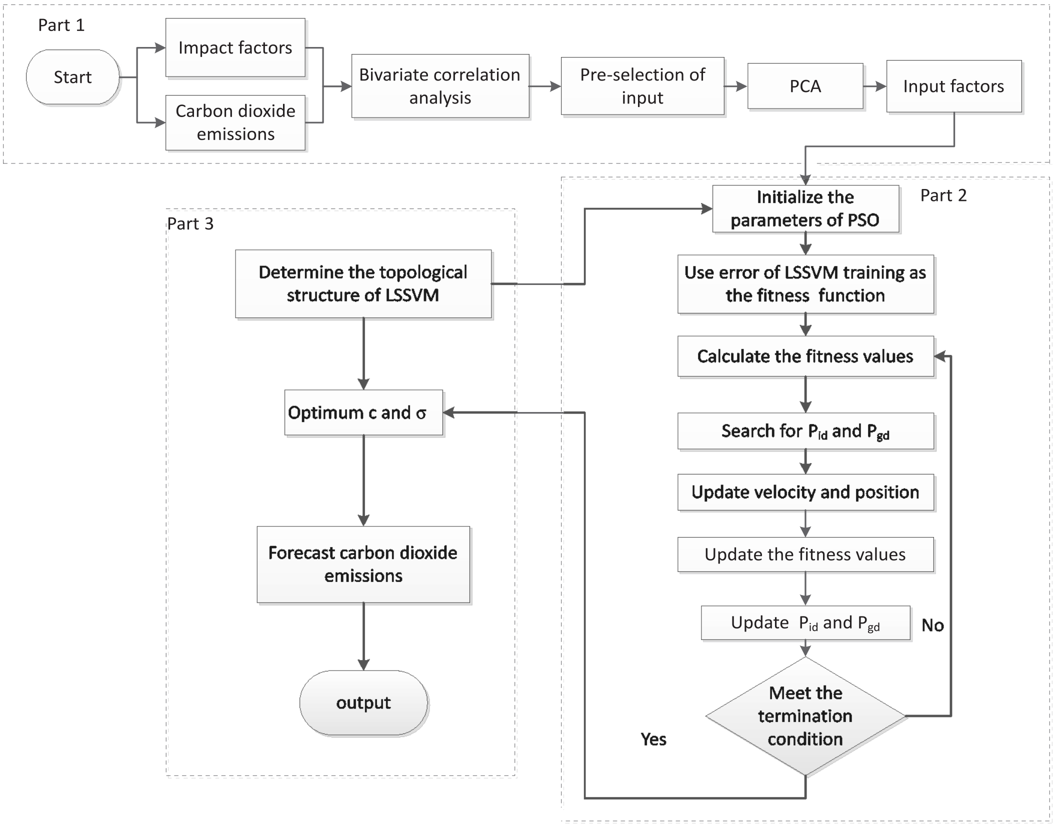

Fig. 1. Flowchart of PSO-LSSVM.

Data Source and Conversion

In order to verify the accuracy of the proposed forecasting model, our paper studies the CO_2 emissions from 1990 to 2016 in Hebei Province. CO_2 emissions are calculated based on consumption of various energy sources, in the Economic Yearbook of Hebei Province, and it is shown in Fig. 2. Then according to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories, the standard coal for energy sources is converted into CO_2 emissions through the energy conversion coefficients in Table 1. To facilitate intuitive understanding, Table 2 lists the four types of energy CO_2 emissions and total CO_2 emissions in Hebei from 1990 to 2016.

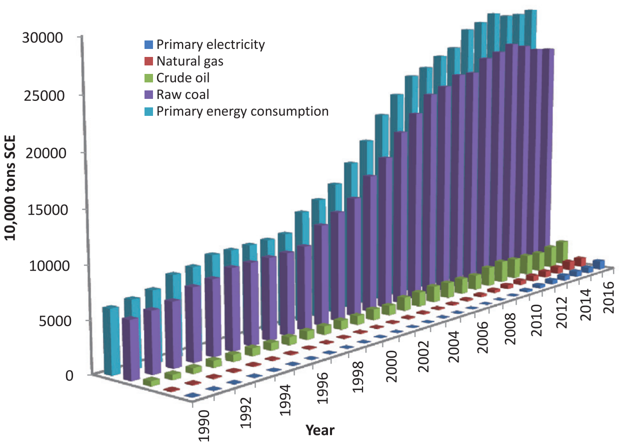

Fig. 2. Primary energy consumption and the source of energy in Hebei during 1990-2016.

As can be seen from Fig. 2, between 1990 and 2016, coal accounts for most of the CO_2 emissions in Hebei, while crude oil and natural gas account for only a small part of CO_2 emissions. In order to respond to the national energy-saving and emission-reduction policy, the proportion of natural gas consumption has been gradually rising. Based on 1990, it has grown by more than 11 times in recent years. The increase in primary electricity was even greater. Using 1990 as a benchmark, there was an increase of 36.44 times in 2016. As China's "coal to electricity" and "coal to gas" implementation, the consumption of coal gradually decreased; on the contrary, the portion of clean energy gradually increased. In order to further study the relationship between CO_2 emissions in Hebei and China's CO_2 emissions, this paper also calculates China's CO_2 emissions for comparison.

Table 1. Coefficients of CO_2 emissions for different energy species.

Type	Coal	Petroleum	Natural gas	Primary Power
C/(t/t)	0.7476	0.5825	0.4435	1.814

Table 2. CO_2 emissions of total energies and main sources during 1990-2016 in Hebei (10,000 tons).

Year	Total emissions	Raw coal	Crude oil	Natural gas	Primary electricity
1990	4501.98	4136.19	282.18	35.85	47.76
1991	4755.83	4385.06	289.15	38.18	43.45
1992	5039.14	4650.20	310.77	40.81	37.37
1993	5785.31	5296.87	386.52	33.47	68.46
1994	5983.63	5522.43	395.41	39.13	26.67
1995	6515.19	6005.10	442.36	37.07	30.66
1996	6553.75	6050.91	429.55	39.25	34.05
1997	6613.53	6100.06	455.67	34.85	22.95
1998	6686.66	6135.34	497.34	35.72	18.27
1999	6858.48	6311.45	491.71	36.61	18.72
2000	8196.27	7611.60	532.81	41.71	10.16
2001	8887.62	8317.62	523.60	37.61	8.80
2002	9816.61	9131.34	636.36	41.61	7.29
2003	11253.50	10610.97	578.32	44.78	19.43
2004	12718.73	11820.14	809.42	57.70	31.47
2005	14573.99	13616.35	860.81	53.66	43.17
2006	15997.19	14923.00	969.90	64.76	39.53
2007	17338.61	16285.14	943.82	71.13	38.51
2008	17866.12	16784.76	944.97	101.11	35.28
2009	18667.91	17579.75	919.48	136.41	32.27
2010	19344.96	17625.44	1182.83	175.47	361.22
2011	20809.10	18699.00	1327.92	206.69	575.49
2012	21466.08	19107.41	1253.21	260.23	845.24
2013	22210.71	19668.86	1247.58	293.38	1000.89
2014	21987.02	19390.25	1192.12	330.29	1074.37
2015	21975.64	19020.20	1368.11	430.22	1157.11
2016	22588.36	18935.38	1497.76	414.91	1740.31

As shown in Fig. 3, by comparing Hebei's CO_2 emissions with that of the whole country, it can be seen that the total emissions of Hebei and China's total emissions continue to grow. It can be seen that Hebei emits at least 6% of China's CO_2 emissions. From 2000 to 2009, the proportion has risen by close to 8%, and the proportion has continued to decrease since 2010. From the figure, before 2000, the growth rates of CO_2 emissions in both China and Hebei increased slowly. After 2000, the growth rates increased significantly. Hebei has a faster rate of increase in CO_2 emissions than the national growth rate. It has only slowed down in the past three years, indicating that Hebei has achieved initial results in conformity with the national emission reduction policies.

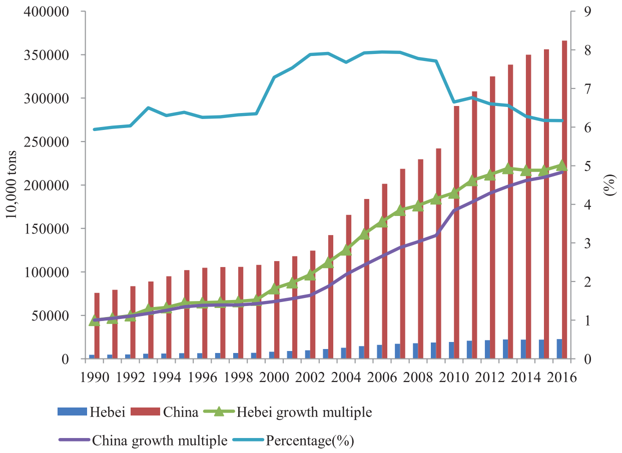

Fig. 3. CO_2 emissions of Hebei Province and China.

The prediction results of CO_2 are related to the selection of influencing factors. In order to better study CO_2 emissions, this paper extracts the main sub-indicators from the general indexes of the Hebei Economic Yearbook. Pre-selected 24 variables, namely, coal consumption, primary sector GDP, secondary industry GDP, tertiary industry GDP, area final consumption, total population, power generation, total export, coverage of railway, coverage of highway, urbanization level (%), total investment in fixed assets of the whole society, consumer price index, transportation possession quantity, fuel and power purchase price index, cement production, urban green areas, added value of construction industry, producer price index, total power of agricultural machinery, consumption of pure fertilizers, total retail sales of consumer goods, finished steel production, and gross output value of agriculture animal husbandry and fisheries.

Pre-selection Factors Analysis

Due to the large number of influencing factors, it is necessary to ensure the rationality of the pre-selection factors. Firstly, the correlation between CO_2 emissions and preselected influencing factors was analyzed. This paper applied SPSS 20.0 to select the Pearson coefficient and the two-sided significance to reflect the correlation. The results are shown in Table 3. The coefficient of correlation of the selected influencing factors was greater than 0.8, except for the producer price index, and the probabilities of the two-sided significance tests were all less than 0.05, satisfying the confidence level of 95%, indicating that there is a significant correlation between CO_2 emissions and influencing factors. Therefore, 24 influencing factors are used as preselected factors. However, the correlation between factors is also great, affecting prediction accuracy. It is important to eliminate multicollinearity between variables using PCA.

Table 3. Correlation and bilateral significance analysis of CO_2.

Factor	Pearson correlation	Factor	Pearson correlation
Coal consumption	0.997**	Consumer Price Index	0.900**
Primary sector GDP	0.971**	Transportation possession quantity	0.935**
Secondary industry GDP	0.972**	Fuel and power purchase price index	0.950**
Tertiary GDP	0.942**	Cement production	0.954**
Area final consumption	0.956**	Urban green areas	0.992**
Total population	0.973**	Added value of construction industry	0.951**
Power generation	0.986**	Producer price index	0.747**
Total export	0.961**	Total power of agricultural machinery	0.904**
Coverage of railway	0.966**	Consumption of pure fertilizers	0.866**
Coverage of highway	0.971**	Total retail sales of consumer goods	0.914**
Urbanization level (%)	0.993**	Finished steel production	0.960**
Total investment in fixed assets of the whole society	0.893**	Gross output value of agriculture animal husbandry and fishery	0.965**

Note: **indicates a significant correlation at the significance level of 0.01.

PCA Analysis

Prior to the principal component analysis, KMO and Bartlett tests were performed on the data applying SPSS 20.0, where the KMO test value was 0.764>0.5, the approximate Chi-Square was 2464.493, the df was 276, and the significance was 0.000<0.05. The results indicate that principal component analysis is effective.

Then reduce the dimensions of the influencing factors. After processing and analysis, as shown in Fig. 4, gravel chart, select the components with eigenvalue greater than 1. The first principal component explains 91.493% of the factors; the second principal component is the auxiliary component, and the cumulative interpretation of more than 97% of the variables. The composition matrix is shown in Table 4. As a result, these two principal components are applied to explain and replace the impact factors as input factors for CO_2 emissions prediction.

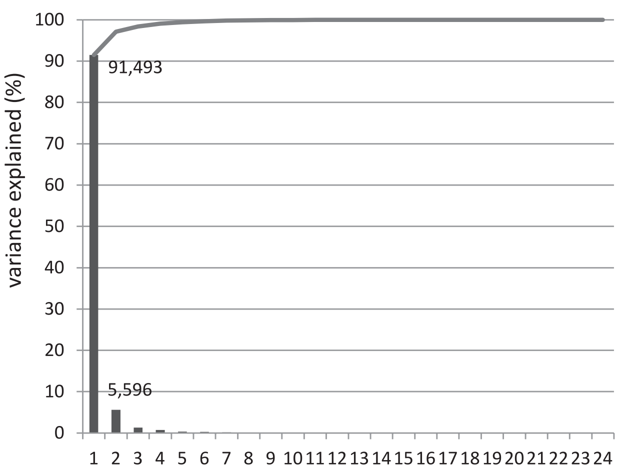

Fig. 4. Scree plot of principal component analysis.

Application of the PSO-LSSVM

In order to verify the accuracy of the prediction model proposed in CO_2 emission prediction, this paper applies the data of Hebei's CO_2 emissions from 1990 to 2016. The PSO-LSSVM model was used to predict the CO_2 emissions in Hebei from 2012 to 2016. The relative error (RE), the mean absolute percentage error (MAPE), and the root mean square error (RMSE) were selected to test and predict the CO_2 emissions. The smaller the error value obtained, the higher the accuracy of the prediction and the better the performance of the prediction model. The error equations are as follows:

$$RE = \left| \frac{y_t - \hat{y}_t}{y_t} \right| \times 100\% \tag{12}$$

$$MAPE = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{y_t - \hat{y}_t}{y_t} \right| \times 100\% \tag{13}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^{n} \left| \frac{y_t - \hat{y}_t}{y_t} \right|^2} \tag{14}$$

Where y_t, $\hat{y}_t$ are the actual and predicted CO_2 emissions during the t-period, and n represents the number of samples of the predicted CO_2 emissions.

In this paper, PSO optimizes the parameters of C and σ in LSSVM. In PSO, max-generation = 100,

Table 4. Component matrix.

Component	PCI	PC2	Component	PCI	PC2
Coal consumption	0.975	0.117	Consumer Price Index	0.934	0.197
Primary sector GDP	0.995	-0.075	Transportation possession quantity	0.964	-0.25
Secondary industry GDP	0.991	-0.101	Fuel and power purchase price index	0.947	0.279
Tertiary industry GDP	0.974	-0.225	Cement production	0.931	0.268
Area final consumption	0.982	-0.186	Urban green areas	0.994	-0.004
Total population	0.988	0.063	Added value of construction industry	0.982	-0.136
Power generation	0.997	-0.039	Producer price index	0.726	0.63
Total export	0.976	-0.104	Total power of agricultural machinery	0.896	0.392
Coverage of railway	0.975	-0.069	Consumption of pure fertilizers	0.875	0.383
Coverage of highway	0.965	-0.048	Total retail sales of consumer goods	0.953	-0.29
Urbanization level (%)	0.987	0.041	Finished steel production	0.979	-0.185
Total investment in fixed assets of the whole society	0.937	-0.334	Gross output value of agriculture animal husbandry and fishery	0.993	-0.08

the size of population = 20, and inertia weight ω = 0.5, C1 = 1.5, C2 = 1.7.

Results and Discussion

The prediction model PSO-LSSVM was implemented by MATLAB 2014a. Based on the data from 1990-2016, the prediction of CO_2 emissions in Hebei was conducted. A total of 26 data were collected, and 21 samples from 1990 to 2011 were selected as the training set. The remaining 5 years of sample was as a test set.

As shown in Fig. 5, the CO_2 emissions forecast for 2012-2016 are predicted by five forecasting models. It can be seen that the PCA/PSO-LSSVM model has the closest predictive value to the actual value.

Fig. 6 shows the relative error of each model for each year. Except for 2016, the relative errors predicted by the PCA/PSO-LSSVM model are less than 0.7%. To further prove the effectiveness of PCA, PCA/LSSVM and LSSVM are also implemented. In contrast, the comparison shows that PCA/LSSVM has higher prediction error accuracy than LSSVM. There is still a large error between the predicted and actual values of the BPNN model, with the maximum relative error reaching 5%. Table 5 clearly lists the mean absolute error percentage (MAPE) and root mean square error (RMSE) values of the five models. The analysis shows that: (1) The PCA/PSO-LSSVM model has the best results in terms of MAPE and RMSE. They are 0.663% and 0.009 respectively. It is already a very accurate result. (2) The application of PCA reduces the multicollinearity between the influencing factors,

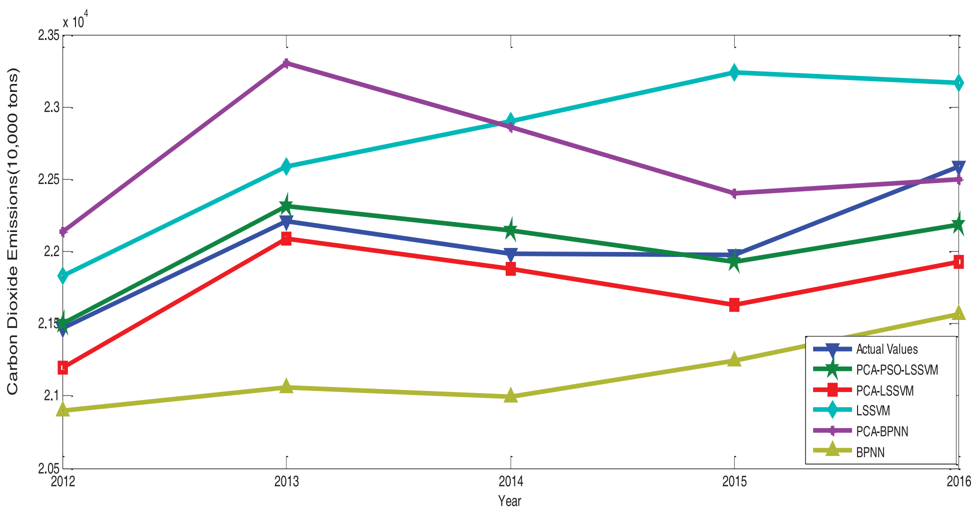

Fig. 5. Fitting curves of the five predictions and actual values.

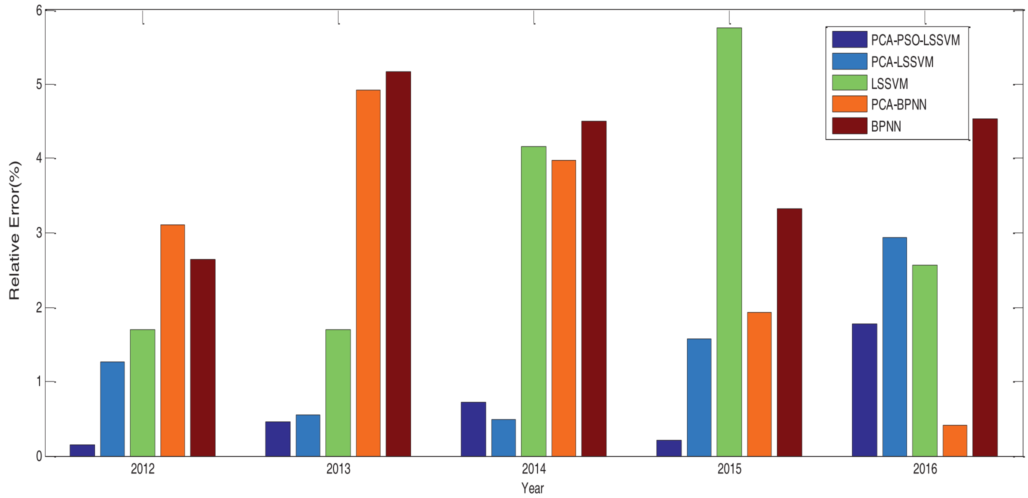

Fig. 6. Comparison of relative errors for 5 forecasting models.

Table 5. Error analysis of compared models.

	PCA/PSO-LSSVM	PCA/LSSVM	LSSVM	PCA/BPNN	BPNN
MAPE (%)	0.663%	1.364%	3.177%	2.873%	4.036%
RMSE	0.009	0.016	0.035	0.033	0.041

simplifies the calculation and improves accuracy. The prediction accuracy of PCA/PSO-LSSVM is higher than that of PCA/LSSVM. The main reason is that PSO optimizes the kernel parameters and regularization parameters of LSSVM and increases the global optimization ability.

Conclusions

This paper uses a least squares support vector machine to establish a CO_2 emission prediction model and uses PSO to solve the parameters. The algorithm not only retains the global optimum searching ability, but also improves PSO convergence. It has certain advantages in dealing with high-dimensional issues. With optimal regularization parameters the CO_2 emissions of Hebei in 2012-2016 are applied to verify the effectiveness of the prediction model. Based on the error values of CO_2 emissions forecast from 2012 to 2016, it can be observed that: (1) PCA is applied and both PSO convergence and LSSVM accuracy are improved; and (2) Compared with other models, the PSO-LSSVM model has higher accuracy. Therefore, the PCA/PSO-LSSVM model proposed in this paper can be effectively applied to CO_2 emissions prediction.

Based on the results, the following suggestions can be proposed for China's future CO_2 emissions reduction: (1) According to the energy consumption data, coal consumption has a high correlation with CO_2 emissions, thus Hebei should respond to the national energy-saving and emission-reduction policies by substituting coal with clean energy and renewable energy. (2) Restricting high-energy-consuming industries, adjusting the structure of the first, second and third industries. Hebei's high-energy-consuming departments, such as steel and cement production, should improve production processes and prohibit high-pollution, poor-quality small workshops to reduce CO_2 emissions during production. (3) The development of public transportation should be encouraged in order to reduce the growth of combustion cars, and the coverage of electric vehicles, hybrid vehicles, and the addition of charging piles should be promoted.

References

1. CETIN M., SEVIK H. Measuring the Impact of Selected Plants on Indoor CO_2 Concentrations. Pol. J. Environ. Stud. **25**(3), 973, **2016**.
2. CETIN M., SEVIK H., SAAT A. Indoor Air Quality: the Samples of Safranbolu Bulak Mencilis Cave. Fresen. Environ. Bull. **26** (10), 5965, **2017**.
3. CETIN M., SEVIK H., ISINKARALAR K. Changes in the particulate matter and CO_2 concentrations based on the time and weather conditions: the case of Kastamonu. Oxid. Commun. **40** (1-II), 477, **2017**.
4. CETIN M. Change in Amount of Chlorophyll in Some Interior Ornamental Plants. Kastamonu University Journal of Engineering and Sciences. **3** (1), 11, **2017**.
5. SEVIK H., CETIN M., KAPUCU O., ARICAK B., CANTURK U. Effects of light on morphologic and stomatal characteristics of Turkish Fir needles (Abies nordmanniana subsp. bornmulleriana mattf.).Fresenius Environmental Bulletin, **26** (11), 6579, **2017**.

6. SEVIK H., CETIN M. Effects of Water Stress on Seed Germination for Select Landscape Plants. Pol. J. Environ. Stud. **24** (2), 689, **2015**.
7. GUNEY K., CETIN M., GUNEY K.B., MELEKOGLU A. The Effects of Some Hormones Applications on Lilium martagon L. Germination and Morpholgical Characters. Pol. J. Environ. Stud. **26** (6), 1, **2017**.
8. CETIN M. Changes in the amount of chlorophyll in some plants of landscape studies. Kastamonu University Journal of Forestry Faculty. **16** (1), 239, **2016**.
9. CETIN M. Chapter 27: Landscape Engineering, Protecting Soil, and Runoff Storm Water, In Tech-Open Science-Open Minds, Book: Advances in Landscape Architecture-Environmental Sciences, 697, **2013**.
10. TURKYILMAZ A., SEVIK H., CETIN M., AHMAIDA SALEH E.A. Changes in Heavy Metal Accumulation Depending on Traffic Density in Some Landscape Plants. Pol. J. Environ. Stud. **27** (5), 2277, **2018**.
11. CETIN M. A Change in the Amount of CO_2 at the Center of the Examination Halls: Case Study of Turkey. Studies on Ethno-Medicine **10** (2), 146, **2016**.
12. CETIN M. Sustainability of urban coastal area management: A case study on Cide. J. Sustain. Forest. **35** (7), 527, **2016**.
13. CETIN M. Determination of bioclimatic comfort areas in landscape planning: A case study of Cide Coastline. Turkish Journal of Agriculture-Food Science and Technology **4** (9), 800, **2016**.
14. SINGH A.S., ZWICKLE A., BRUSKOTTER J.T., WILSON R. The perceived psychological distance of climate change impacts and its influence on support for adaptation policy. Environ. Sci. Policy. **73**, 93, **2017**.
15. FAN F.Y., LEI Y.L. Index Decomposition Analysis on Factors Affecting Energy-Related Carbon Dioxide Emissions from Residential Consumption in Beijing. Math. Probl. Eng. 4963907, **2017**.
16. MENG M., NIU D.X., GAO Q. Decomposition Analysis of Chinese Provincial Economic Growth through Carbon Productivity Analysis. Environ. Prog. Sustain. **33** (1), 250, **2014**.
17. WANG M., FENG C. Decomposing the change in energy consumption in China's nonferrous metal industry: An empirical analysis based on the LMDI method. Renew. Sustain. Energy. Rev. **82**, 2652, **2018**.
18. LIU L.J., LIANG Q.M. Changes to pollutants and carbon emission multipliers in China 2007-2012: An input-output structural decomposition analysis. J. Environ. Manage. **203**, 76, **2017**.
19. WEN L., LI Y. The Causality Relationships between Energy-related CO2 Emissions and its Influencing Factors with Linear and Nonlinear Granger Causality Tests. Pol. J. Environ. Stud. **26** (3), 1313, **2017**.
20. ZHANG J., ZHANG L.Y., DU M., ZHANG W. HUANG X., ZHANG Y.Q., YANG Y.Y., ZHANG J.M., DENG S.H., SHEN F., LI Y.W., XIAO H. Indentifying the major air pollutants base on factor and cluster analysis, a case study in 74 Chinese cities. Atmos. Environ. **144**, 37, **2016**.
21. HAO Y., CHEN H.Y., ZHANG Q.X. Will income inequality affect environmental quality? Analysis based on China's provincial panel data. Ecol. Indic. **67**, 533, **2016**.
22. ZHAO X.R., ZHANG X., SHAO S. Decoupling CO_2 emissions and industrial growth in China over 1993-2013: The role of investment. Energy. Econ. **60**, 275, **2016**.
23. LIANG Y., NIU D.X., WANG H.C., LI Y. Factors Affecting Transportation Sector CO_2 Emissions Growth in China: An LMDI Decomposition Analysis. Sustainability, **9**, 1730, **2017**.
24. WANG X.L., LIN B.Q. How to reduce CO_2 emissions in China's iron and steel industry. Renew. Sust. Energ. Rev. **57**, 1496, **2016**.
25. ZHOU X., ZHANG M., ZHOU M.H., ZHOU M.A comparative study on decoupling relationship and influence factors between China's regional economic development and industrial energy-related carbon emissions. J. Clean. Prod. **142**, 783, **2017**.
26. XU B., LIN B.Q. Reducing carbon dioxide emissions in China's manufacturing industry: a dynamic vector auto regression approach. J. Clean. Prod. **131**, 594, **2016**.
27. WANG Z.H., YIN F.C., ZHANG Y.X., ZHANG X. An empirical research on the influencing factors of regional CO_2 emissions: Evidence from Beijing city, China. Appl. Energ. **100**, 277, **2012**.
28. MENG M., NIU D.X. Modeling CO_2 emissions from fossil fuel combustion using the logistic equation. Energy. **36** (5), 355, **2011**.
29. RIGOBERTO P.S., ANA J.M. Growing green? Forecasting CO_2 emissions with Environmental Kuznets Curves and Logistic Growth Models. Environ. Sci. Pol. **54**, 428, **2015**.
30. ZHANG P.Y., HE J.J., HONG X., ZHANG W. Regional-Level Carbon Emissions Modelling and Scenario Analysis: A STIRPAT Case Study in Henan Province, China. Sustainability. **9** (12), 2342, **2017**.
31. WANG Z., DANG Y.G. Research on carbon emission prediction in Jiangsu Province based on an improved GM (1, 1) model, grey systems and intelligent services. IEEE International Conference on Grey Systems and Intelligent Services (GSIS), 93, **2013**.
32. CHEN Z.H., YE X.Q., HUANG P. Estimating Carbon Dioxide (CO_2) Emissions from Reservoirs Using Artificial Neural Networks. Water, **10** (1), 26, **2018**.
33. SUN W., YE M.Q., XU Y.F. Study of carbon dioxide emissions prediction in Hebei province, China using a BPNN based on GA. J. Renew. Sustain. Energ. **8** (4), 043101, **2016**.
34. WEN L., LIU Y.J. A Research about Beijing's Carbon Emissions Based on the IPSO-BP Model. Environ. Progress. Sustain.**36** (2), 428, **2017**.
35. ZHAO X., HAN M., DING L.L. Forecasting carbon dioxide emissions based on a hybrid of mixed data sampling regression model and back propagation neural network in the USA. Environ. Sci. Pollution. R. **25** (3), 2899, **2018**.
36. MARJANOVIC V., MILOVANCEVIC M., MLADENOVIC I. Prediction of GDP growth rate based on carbon dioxide (CO_2) emissions. J. CO_2 Util. **16**, 212, **2016**.
37. SUN W., WANG C.F., ZHANG C.C. Factor analysis and forecasting of CO_2 emissions in Hebei, using extreme learning machine based on particle swarm optimization. J. Clean. Prod. **162**, 1095, **2017**.
38. MLADENOVIC I., SOKOLOV-MLADENOVIC S., MILOVANCEVIC M., MARKOVIC D., SIMEUNOVIC N. Management and estimation of thermal comfort, carbon dioxide emission and economic growth by support vector machine. Renew. Sustain. Energy. Rev. **64**, 466, **2016**.
39. SUN W., LIU M.H. Prediction and analysis of the three major industries and residential consumption CO_2 emissions based on least squares support vector machine in China. J. Clean. Prod. **122**, 144, **2016**.

40. SUYKENS J.A.K., VANDEWALLE J. Least squares support vector machine classifiers. Neural Process. Lett. **9**, 293, **1999**.
41. KENNEDY J., EBERHART R. C. Particle swarm optimization. In: Proceedings of IEEE International Conference on Neutral Networks (ICNN'95), IV, 1942-1948, Perth, Australia, **1995**.