Prediction of Carbon Financial Transaction Price Based on CEEMD Denoising and PSO-LSSVM——Take Guangdong Province as an example

Yingshi Cai[1,#,*], Yizhou Yang[2,#], Guangye Yang[3,#]

[1] Economic and Trade College, University of Electronic Science and Technology of China Zhongshan Institute, Zhongshan, Guangdong, 528400, China

[2] Faculty of Economics and Management, Changsha University of Science & Technology, Changsha, Hunan, 410000, China

[3] Surrey International Institute, Dongbei University of Finance & Economics, Dalian, Liaoning, 116025, China

*Corresponding author. Email: caiyingshi0413@163.com

#These authors contributed equally.

Abstract. Climate change has an unprecedented impact on the world, and it is also an important era proposition to be solved in China's development. Carbon emissions trading occupies an important position, which aims to adapt to the global carbon neutrality context and implement sustainable development requirements. The market mechanism of China is to deal with climate change, and carbon emissions trading occupies an important position. Based on the carbon emission trading price data of Guangdong Province from 2015 to 2021, this paper uses the contribution analysis model of complete ensemble empirical mode decomposition (CEEMD), and particle swarm optimization algorithm-least squares support vector machine (PSO-LSSVM) to predict the carbon price of Guangdong Province. Firstly, the carbon emission trading data of Guangdong Province are obtained. Based on CEEMD decomposition data, the data filtering and noise reduction are completed. The PSO-LSSVM model is used for iterative particles and obtains the optimal solution, and each component is predicted respectively. Finally, the results are integrated. The results are shown that: (1) CEEMD data decomposition noise reduction technology is used to improve the prediction accuracy of IMFs ; (2) The optimization of parameters in LSSVM modeling by PSO algorithm helps to select parameters more reasonably and avoids the randomness of artificial selection to some extent. The research results show that the decomposition ensemble prediction model applies to the carbon emission prediction in Guangdong Province. The good generalization ability of the model provides a more accurate prediction scheme for carbon price prediction in China.

Keywords: Carbon price prediction, Complete Ensemble Empirical Mode Decomposition, Particle Swarm Optimization, Least Square SVM.

1. Introduction

Since this century, global warming and excess greenhouse gas emissions have become a major constraint to social and economic development worldwide. In September 2015, President Xi Jinping and President Obama met and signed the Joint Statement on Climate Change between China and the United States dollar, declaring our country planned to establish a national carbon emissions trading system in 2017. In the 19th CPC National Congress report, General Secretary Xi Jinping also proposed actively participating in global environmental governance and implementing emission reduction commitments. China's unified carbon emissions trading market was officially launched in June 2021. The future development potential is enormous. In 2020, the trading volume of the Guangdong carbon market ranks first in the pilot carbon market. Therefore, this paper takes the trading price of carbon emission rights in Guangdong Province as the research object to analyze and support the transformation and upgrading of industrial structure and sustainable economic development in China.

At present, there are few studies on the carbon price prediction. Du Ziping and Liu Fucun analyzed carbon prices based on GA-BP-MIV neural network model[1]. Based on the GA-RS model, Lv Jingye, Yang Hua, Guo Ze, and others decomposed China's s carbon emission rights price[2]. Wei Yu, Zhang Jiahao, and Chen Xiaodan forecast China's carbon emissions trading price based on dynamic model selection (DMS) and dynamic model averaging (DMA)[3]. The construction of China's carbon trading market started late, and there are still loopholes in previous studies on the carbon price prediction. Previous trading price forecasting methods mainly include the time series model, structural model, and other methods such as the neural network. Firstly, the time series model modeling focuses on fluctuations rather than spot prices.

Secondly, the time difference of structural model prediction results is strong, the prediction ability is strong in a certain period, and the prediction deviation may occur in other periods. Third, neural networks and other methods usually only focus on a single time series, making it easy to ignore the impact of other factors. In addition, models such as neural networks and genetic algorithms also lack model decomposition steps, and the accuracy needs to be improved. Although ensemble empirical mode decomposition (EEMD) reduces the mode mixing phenomenon and end effect of intrinsic mode function (IMF) in empirical mode decomposition (EMD), the new mode mixing, spectrum loss, and increase of computation caused by ensemble average also affect the decomposition effect[4]. In this paper, Complete Ensemble Empirical Mode Decomposition (CEEMD) proposed by Torres et al. will be used to solve the problem of missing decomposition steps and the nature of the EEMD model in the process of carbon price analysis and prediction by the neural network in previous papers. The carbon price prediction method is optimized to improve research accuracy and efficiency. After decomposition, each scale residual component added specific Gaussian white noise, and the IMF's were obtained by calculating the unique residual. This method reduces the calculation amount and eliminates modal aliasing and false components[5]. In addition, particle swarm optimization (PSO) is used to find the optimal global solution. Based on the least square support vector machine (LSSVM) regression theory, the regression analysis method is used to solve the equations, which shortens the training time of SVM and improves the speed and accuracy of solving the problem.

This paper first uses the CEEMD model to decompose and screen the carbon price, reduce noise, reduce and calculate costs, and improve data quality. Secondly, PSO-LSSVM is used to iterate the random particles to find the optimal solution, which avoids the result that too many decomposition iterations lead to too high calculation cost. In order to verify the validity of the model, the prediction results of carbon emission trading prices in Guangdong Province are finally obtained.

This paper consists of four chapters, and its chapters are arranged as follows. The second chapter introduces the research model, including the CEEMD and the PSO-LSSVM models. The third chapter completes the case calculation based on the carbon price transaction data of Guangdong Province from March 2015 to March 2021. The fourth chapter describes the above conclusions and research significance.

2. Methodology

2.1 CEEMD

CEEMD is an improved algorithm based on EMD and EEMD. The EMD algorithm is proposed by Huang of NASA, which assumes that a signal is composed of multiple modes, and a mode describes a single vibration state. Any complex time series is decomposed into a series of intrinsic mode functions (IMF) at different frequency scales by EMD, and each IMF reflects the dynamic characteristics of the original signal. The IMF component must meet two conditions: ①The number of poles and zeros is the same or at most one difference. ②The upper and lower envelopes are locally symmetrical about the time axis. Based on the EMD decomposition, it consists of the following steps[6]:

Step1: Add n groups to the original signal in positive and negative pairs

$$\begin{pmatrix} m_1 \\ m_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} x(\mathrm{t}) \\ \mathrm{n}(\mathrm{t}) \end{pmatrix} \tag{1}$$

Where x(t) original signal, n(t) auxiliary noise, and the amplitude can be selected as 0.2 ~ 0.5 times the standard deviation of the original signal or increased with the intensity of the noise.;m_1 and m_2 are signals after adding positive and negative noise, respectively. Thus the number of set signals is 2n.

Step2: EMD decomposition of each signal in a set, Each signal obtains a series of IMF components. The j IMF component of the i signal is $imf_{ij}(t)$;

Step3: Then the mean value of the multi-component combination

$$imf_j(t) = \frac{1}{2n} \sum_{i=1}^{2n} imf_{ij}(t) \tag{2}$$

Where $imf_{ij}(t)$ is the j IMF component of the signal after CEEMD decomposition. This method ensures the completeness of signal decomposition. The modal mixing effect can be well solved, and the computational efficiency is greatly improved.

The framework of the decomposition ensemble prediction model can be represented in Figure 1.

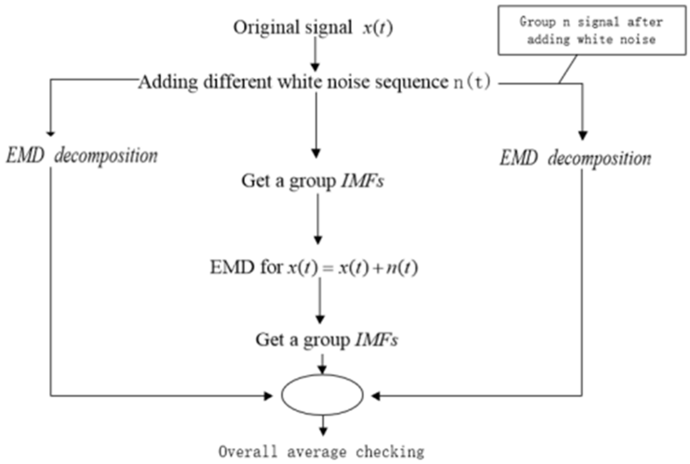

Figure 1 CEEMD model flow chart

2.2 Prediction of Carbon Price Based on PSO Improved LSSVM

The prediction of carbon market price is essentially a regression problem. According to the principle of LSSVM regression, the parameters that need to be selected in the LSSVM algorithm design include the kernel function parameter and the penalty factor r. In order to make PSO and LSSVM better fusion and realize the PSO algorithm to optimize the super parameter of LSSVM, we should focus on the representation of super parameters. In this paper, the Gaussian radial basis function (RBF) is selected as the kernel function of the LSSVM model. RBF kernel function has a strong nonlinear mapping ability. The super parameters that need to be optimized are regularization parameter γ and kernel function parameter σ. Based on the PSO optimization algorithm, the position of each particle represents the possible solution to the objective optimization problem, and a two-dimensional vector is needed to represent the combination of γ and σ. Therefore, in the LSSVM optimization process, the position of the i particle can be expressed as $x_i = (y_i, \sigma_i)$.

According to the construction idea of the appealing model, combined with the actual prediction of Guangdong carbon market price, this paper proposes the PSO-LSSVM algorithm, the algorithm process :

Step1: The survey sample data are cleaned and quantified, the correlation is analyzed, and the influencing factors are determined and normalized to form the sample matrix.

Step2: Determination of training data set and prediction data set

Step3: Set the range of parameters r and σ, initialize population $x(x_1, x_2, x_3, \dots, x_m)$ with m particles by PSO algorithm, where $x_i = \{\gamma_i, \sigma_i\}(i = 1,2,3, \dots, m)$.

Step4: The fitness of the particles is calculated and compared with the training data set. The individual optimal value $P_{best}(i)$ and the optimal global value $G_{best}(i)$ are selected, and the speed and position of the particles are updated.

Step5: Repeat iteration to meet end condition

Step6: The optimal global value $G_{best}(i)$ is output, that is, the optimal hyperparameters γ and σ are assigned to the LSSVM prediction model.

Step7: The training set data is used to train the LSSVM model under the optimal parameters, and the sample data of the prediction set is input to predict the model, and the optimal results are obtained.

3. Case study

3.1 Sample selection and data source

This paper selects the carbon price data of Guangdong Province as an empirical case, and the data come from the most professional carbon market analysis platform in China on the carbon K line(http://k.tanjiaoyi.com/). In order to get more accurate comparison results to verify the accuracy of the CEEMD prediction model in carbon emissions. This paper selects the data from March 2015 to March 2021, using 64 bit Windows processing system using Matlab software for data decomposition and model prediction.

3.2 CEEMD Decomposition Noise Reduction

The original carbon emission data still exist volatility and instability after CEEMD decomposition, and the IMF reflects the short-term interference of carbon price data. The trend of the residual term is consistent with the original data, but compared with the original data, it is smoother. IMFs determine the data fluctuation, which is caused by data instability. The residual term determines the overall trend of the data, which is the main part of the data. Therefore, the accurate prediction of IMFs is conducive to improving the prediction accuracy[7].

CEEMD is used to decompose and denoise the carbon emission data of Guangdong Province from 2015 to 2021. Firstly, a pair of reciprocal positive and negative white noise is added to the source signal as auxiliary noise. Then, three IMF modules and residuals are obtained after CEEMD decomposition. In Figure 2, all IMFs are arranged from high frequency to low frequency and it is shown the change in frequency and amplitude. The last display is the residual item. The frequency and amplitude of all IMFs change with time. Corresponding to the original carbon price series, the IMF function fluctuates violently before 2019, and the amplitude decreases after 2019[8]. Carbon price emissions show an upward trend.

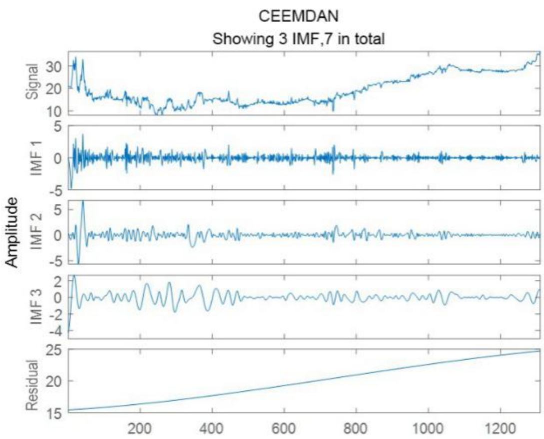

Figure 2 CEEMD decomposition results

3.3 PSO-LSSVM

As shown in Fig. 3, the PSO-LSSVM model is compiled and operated in the Matlab R2021a version environment. In the evolution process of the PSO algorithm, actual value coding is used. Firstly, a swarm of particles is initialized. The PSO algorithm is relatively insensitive to group size, and the population size is set to 8. The initial fitness of each particle is evaluated and recorded as the optimal value. Then, the calculation is carried out according to the set calculation speed and position formula. In the numerical setting, the value of the number of iterations should not be too large or too small. In the experiment, it is found that setting the maximum number of iterations to 3 can meet the fitting requirements; for the learning factor, this paper refers to the previous literature and debugging experience, let the acceleration factor c1=c2=2; according to historical experience and make the inertia weight $\omega \in [0.4,0.95], \gamma \in [0,200], \sigma \in [0,1]$, and the maximum speed $v_{\max}=2.5$. The fitness function is defined as the program comparing the current fitness value with the previous fitness value after the initial calculation. If it is good, it will be updated to find the global optimum of the current particle swarm and repeat until the minimum error or the maximum number of iterations is reached to output the optimized fitting curve to achieve the prediction of the carbon price.

When the PSO-LSSVM model is established, the sample set is the carbon price emission data of 1312 trading days from 2015 to 20217. The closing price data of the previous 917 trading days are used as the training set to train the PSO-LSSVM model, which is using the remaining 394 trading day data as a test set to test the model's prediction ability.

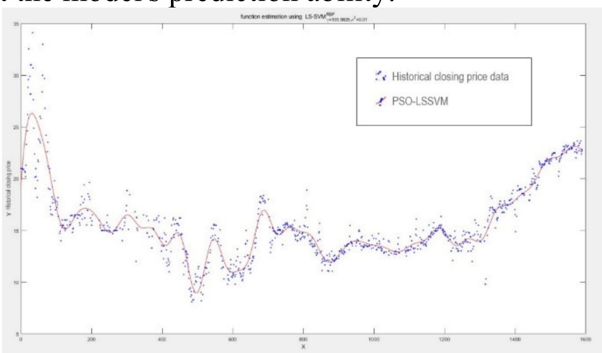

Figure 3 PSO-LSSVM prediction results of test samples

In order to compare, this paper also uses the LSSVM model, which is an algorithm that combines least squares estimation to transform the SVM problem into solving linear equations before optimization to learn and predict the carbon market price in Guangdong Province. In the fitting, the parameter gam is set to 1, and sig2 is set to 5. The prediction results and the corresponding index ranking are shown in Table 1.

Table 1 Evaluation indicators

Model	Evaluation indicators					
	MAPE	ranking	RMSE	ranking	MAE	ranking
PSO-LSSVM	0.071	1	1.22	1	1.061	1
LSSVM	0.083	2	2.30	2	1.720	2

This paper uses root mean square error (RMSE), mean absolute percentage error (MAPE), and mean absolute error (MAE) as the evaluation criteria for model comparison:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2} \tag{3}$$

$$MAPE = \frac{100}{N}\sum_{i=1}^{N}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \tag{4}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|\hat{y}_i - y_i\right| \tag{5}$$

In the formula, $\hat{y}_i$ is the observed value, y_i is the actual value, and N is the number of predicted samples. Obviously, the smaller the values of the three indicators, the higher the prediction accuracy.

The data shows that the PSO-LSSVM model's relevant indicators are better than those of the LSSVM model, and the prediction accuracy is much higher than that of the LSSVM model. The fitting accuracy of the PSO-LSSVM model is 14.46 % higher than that of the LSSVM model on the MAPE index. The fitting accuracy was increased by 46.96 % on the RMSE index. Meanwhile, the fitting accuracy on the MAE index increased by 38.31 %. The LSSVM algorithm is a representative model in the current carbon price research field. By comparing its fitting accuracy, it can be found that the fitting accuracy of the PSO-LSSVM model has been greatly improved. Therefore, it can be seen that the PSO-LSSVM model has relative advantages in carbon price prediction and is a competitive and practical prediction method.

4. Conclusions

Carbon emission is an important index to evaluate environmental quality. As the largest province in China ' s economy, Guangdong Province has brought severe environmental problems such as high pollution, high emission, high energy consumption, and rapid economic development. Accurate prediction of carbon emissions in Guangdong Province can promote industrial structure optimization and realize the goal of low carbon transformation in Guangdong Province. It has particular practical significance and can play an exemplary role in formulating energy conservation and emission reduction policies and optimizing industrial structures in other economically developed provinces. This paper combines the data decomposition method CEEMD and PSO to optimize LSSVM to construct the decomposition integrated prediction model of carbon emission. Firstly, CEMMD decomposes and filters the carbon price, reduces noise, reduces calculation cost, and improves data quality. Secondly, PSO-LSSVM is used to iterate the random particles to find the optimal solution, which avoids the result that too many decomposition iterations lead to too high calculation cost. The conclusions of this paper are as follows :

(1) CEEMD decomposition technology can decompose the carbon emission data into more stable sequence modules and improve prediction accuracy.
(2) The application results of the decomposition-integration prediction model in the emission data of Guangdong Province show its universality in carbon emission prediction.
(3) Using the PSO algorithm to optimize the parameters in LSSVM modeling can make the parameter selection more reasonable and avoid the randomness of artificial selection.
(4) The prediction accuracy of the CEEMD-PSO-LSSVM model is high, indicating that the model has good generalization ability. In practice, the CEEMD-PSO-LSSVM model is used to predict the domestic carbon market, which has great advantages in the case of large changes in the domestic carbon price market. The model provides a more accurate prediction scheme for carbon price prediction in China.
(5) Global warming is one of the most serious problems facing human society today, which is related to human survival and development. Carbon emissions caused by fossil fuel combustion and land-use change are considered the leading cause of global warming. Through the accurate prediction of carbon prices in Guangdong Province, we can promote Guangdong Province to achieve the goal of low carbon transformation and add a force for the country to achieve the goal of carbon neutrality by 2060.
(6) Although the decomposition-integration prediction model in this paper can improve the accuracy of carbon emissions prediction, there are still shortcomings. We fail to fully consider the impact of external factors on carbon emissions, such as the impact of urban policies on carbon emissions. Future research will further consider various influencing factors of carbon emissions.

References

[1] Research on China' s regional carbon emission right price and its influencing factors — empirical analysis based on GA-BP-MIV model Du Ziping and Liu Fucun. Price theory and practice. 2018 (06)

[2] The decomposition research on the influencing factors of China' s carbon emission right price based on GA-RS. Lv, Yang and Guo. Ecological economy. 2019 (11)

[3] Wei Yu, Zhang Jiahao, Chen Xiaodan. Research on the prediction method of carbon emission trading price in China based on DMS and DMA — empirical evidence from Hubei carbon market Systems engineering.

[4] Noise aided signal decomposition method based on complex data empirical mode decomposition.Qujianling, Wang Xiaofei, Gaofeng, Zhou Yuping, Zhang Xiangyu. Physical Journal 2014 (11)

[5] A new denoising method based on CEEMDAN and wavelet adaptive threshold Zhang Jianwen, Liu Yang, Zhang Dapeng, Zhang Huanyu. Electrical measurement and instrument. 2018 (10)

[6] Zhu Help, Wang Ping, Wei Yiming. Multi-scale analysis of influencing factors of carbon market price based on EMD Economic dynamics, 2012 (06) : 92 – 97.

[7] Hu Beibei. Research on price prediction of international carbon market based on multiple factors. Hefei University of Technology, 2017.

[8] Wang Keke. Research on the Optimization of Electricity Spot Market Transactions Considering New Energy. North China Electric Power University (Beijing), 2021.DOI: 10.27140 / d.cnki.ghbbu.2021.000125.