Prediction of Calorific Value for Coal Gangue Based on the Machine Learning Algorithm

Xiangbin Gao

Xinjiang Xinneng Group Company Limited, Urumqi Electric Power Construction and Commissioning Institute, Urumqi 830000, China

gxb0802ts@163.com

Abstract. In order to solve the problems of inconvenience and high cost of coal gangue calorific value measurement, the machine learning algorithm is adopted and applied to the prediction of practical engineering. Firstly, the industrial analysis components and calorific value of coal gangue in Xinjiang are obtained by experiment following the Standard GB/T213-2008 and GB/T212-2008 in China. Then, the experimental methods and experimental data used are analyzed to establish a data set of coal gangue. Secondly, a nonlinear prediction model of coal gangue calorific value combined with the particle swarm optimization algorithm is proposed. The prediction model is trained and predicted based on the experimental data. Finally, the accuracy and reliability of the above model are verified by comparing the deviation between actual measurement values and prediction values. The research results show that calorific value prediction model for coal gangue based on PSO-SVR proposed in this study has high prediction accuracy and convergence speed, which can provide scientific basis for energy utilization of coal gangue.

Keywords: Coal Gangue, Calorific Value Prediction, Support Vector Regression, Particle Swarm Optimization, Machine Learning Algorithm.

1 Introduction

Due to the acceleration of industrialization and the continuous increase of energy demand, more and more attention has been paid to the utilization potential of coal gangue as an energy resource [1-2]. However, the measurement of coal gangue's calorific value and production capacity, as important indicators to evaluate its comprehensive utilization value and energy potential, has always been facing problems such as inconvenience in practical measurement and high cost, which limits the scale and efficiency of coal gangue resource utilization.

With the advancement of artificial intelligence, machine learning technologies have been intensively utilized owing to its ability to predict functions through extensive sample learning [3-4]. Therefore, many scholars have constructed many prediction models for the calorific value of various fuels such as coal based on experimental data. By predicting the relevant sample calorific value from limited experimental data, a more

A. E. Abomohra et al. (eds.), *Proceedings of the 2023 9th International Conference on Advances in Energy Resources and Environment Engineering (ICAESEE 2023)*, Atlantis Highlights in Engineering 29,
https://doi.org/10.2991/978-94-6463-415-0_5

convenient and effective way is provided for the determination of fuel calorific value. Mason [5] and Given [6] began research on predicting the coal calorific value in the 1980s, and established a multiple linear regression model based on results of industrial analysis and element analysis. Many researchers have used multiple linear regression analysis to establish predictive models for coal calorific value through moisture and ash content [7-9].

According to the non-linear relationship between the coal calorific value and industrial analysis components, especially the non-linear discrete relationship between moisture, volatile matter and other components in coal and the calorific value [10]. Many researchers have employed different methods to construct non-linear prediction models for the coal calorific value. Nonlinear prediction modeling were carried out by artificial neural network (ANN) methods [11]. There are also studies using support vector regression (SVR) to model and predict the coal calorific value of different regions. The research results indicate that the performance of the nonlinear prediction model is much better than that of linear prediction model [12-13].

Since calorific value of solid waste coal gangue in mining is small and distributed relatively dispersed, hence the previous studies mainly focusd on the calorific value of coal and the identification and separation gangue of from coal. And few manuscripts are published on calorific value experiments and prediction for coal gangue. Gao et al. [14] used a mixed kernel function support vector regression method to predict the calorific value of coal gangue, but the construction of its kernel function was complex. As a result, it is necessary to find a simple and accurate predicting method for the calorific value of coal gangue. This study is based on machine learning algorithms to forecast coal gangue's calorific value. Firstly, the calorific value and industrial analysis components of coal gangue in Xinjiang were measured. Through analysis of experimental data, the distribution law of calorific value of coal gangue in Xinjiang was summarized. Then, a sample database containing 1500 sets of industrial analysis components and calorific value for coal gangue was established. Secondly, three nonlinear prediction models, SVR, random forest (RF) and back propagation neural network (BPNN), were used to construct a coal gangue calorific value prediction model combined with particle swarm optimization (PSO) algorithm. The optimal values of key parameters in the prediction model were given. Finally, appropriate evaluation indicators were used to compare and analyze the performance of different prediction models. The results show that the PSO-SVR coal gangue calorific value prediction model has high prediction accuracy and fast convergence speed. Obviously, the use of machine learning technology to predict the calorific value of fuel resources can reduce testing costs and improve testing efficiency, providing a basis for the analysis and calibration of fuel calorific value, as well as comprehensive utilization of resources.

2 Analysis of Coal Gangue Composition and Calorific Value

2.1 Overview of Experiment

In this study, we used the raw samples of coal gangue to conduct experimental research on the industrial analysis components such as moisture, ash, and volatile matter, as well

as the calorific value. A total of 1500 sets of test samples were obtained, and a database of industrial analysis indicators and calorific values of Xinjiang coal gangue was established, including: air drying base moisture (M_{ad}), air drying base ash (A_{ad}), air drying base volatiles (V_{ad}), air drying base fixed carbon (FC_{ad}), and air drying base bomb calorific value ($Q_{b,ad}$). The experimental equipment used included a 5E-C5500 oxygen bomb calorimeter, an LC-213 forced air drying box, an SDMF300 intelligent muffle furnace, and a BSA124S electronic balance.

2.2 Measurement Method for the Calorific Value of Coal Gangue

Currently, there is no standard or specialized measurement method for gaguing the calorific value of coal gangue. In industry, the measurement method of coal calorific value is often used as a reference. The way for measuring the calorific value of coal is often used as a reference. The actual calorific value calibration mainly follows the "Approach for Determination of the Calorific Value of Coal" (GB/T213-2008), which is measured directly using an oxygen bomb method, or calculated by the composition analysis and the industrial analysis method of coal. However, the direct detection method is not only susceptible to instrument and human factors, but also has problems such as high detection costs and long analysis [15].

The Basic Principle of Measurement.

The basic principle of measurement for the energy value is the first law of thermodynamics. The amount of coal gangue sample is placed in a sealed container (oxygen bomb) with sufficient oxygen for complete combustion. The heat generated by the combustion is absorbed by a certain amount of water (i.e., the inner cylinder water) and the calorimeter system around the oxygen bomb, and the temperature rise of the water is directly proportional to the heat released by the combustion of the sample.

Calculation of Calorific Value.

The first law of thermodynamics states that energy is conserved during the transfer and conversion of different forms, and heat can be transferred from one object to another while the total energy remains constant. Therefore, the equation for calorific value by using an oxygen bomb calorimeter is as follows:

$$m_1 Q_v + m_2 Q_{vb} + \Delta W_F Q_F = (C_{H_2O} W_{H_2O} + W_1) \cdot \Delta T \tag{1}$$

Where, m_1 represents the mass of the substance to be measured, Q_v represents the calorific value per unit mass of the substance to be measured, m_2 represents the mass of benzoic acid, Q_{vb} represents the calorific value per unit mass of benzoic acid, ΔW_F represents the mass of the ignition wire burned, Q_F represents the combustion heat of the ignition wire. ΔT represents the temperature change of the system before and after

the sample burns, W_1 represents the water equivalent of the instrument, C_{H_2O} represents the specific heat capacity of water, W_{H_2O} represents the mass of water.

2.3 Industrial Analysis and Measurement Methods for Coal Gangue

In the research on the forecasting model of coal calorific value, the main indicators of coal quality are the industrial analysis component. Therefore, this study carried out industrial analysis of coal gangue, mainly referring to the "Industrial Analysis Method of Coal" (GB/T212-2008).

Air Drying Base Moisture M_{ad} (%).

In this experiment, the air drying method was used. A specified quantity of coal gangue was weighed and put in a forced air drying box at (105~110) ℃. The sample was dried to a constant mass in the air stream. The moisture content was calculated from the mass loss of the coal gangue sample. Moisture content can be computed as follows:

$$M_{ad} = \frac{m_1}{m} \times 100 \tag{2}$$

Where, m represents the mass of the sample (g), m_1 represents the mass lost after drying (g).

Air Drying Base Ash A_{ad} (%).

Slow combustion method was used in this experiment. A specified quantity of coal gangue was weighed and put in a muffle furnace, heated at a certain rate to (815±10) ℃, and slowly ashed and burned to a constant mass. The ash content was determined by the mass fraction of the residual mass to the total mass of the test sample. Ash content can be computed as follows:

$$A_{ad} = \frac{m_1}{m} \times 100 \tag{3}$$

Where, m represents the mass of the air-dried coal sample (g), m_1 represents the mass of the residue after burning (g).

Air Drying Base Volatiles V_{ad} (%).

The determination of volatile content involves placing a certain mass of coal gangue in a special crucible that is isolated from the air. Under certain high temperature conditions, the sample is heated for a length of time, and gaseous products (excluding moisture content) resulting from the decomposition of the coal gangue are the volatile content of the coal gangue. The coal gangue sample is weighed and placed in a covered

porcelain crucible. At a temperature of (900±10) °C, the sample is heated in isolation from the air for 7 minutes. The volatile content of the coal gangue is determined by subtracting the moisture content from the mass fraction of the reduced mass to the total mass of the sample. Volatile content can be computed as follows:

$$V_{\mathrm{ad}} = \frac{m_1}{m} \times 100 - M_{\mathrm{ad}} \tag{4}$$

Where, m denotes the mass of the coal sample (g), m_1 represents the mass reduction of the coal sample after heating (g).

Air Drying Base Fixed Carbon FC_{ad} (%).
According to the elemental analysis of coal industry, fixed carbon can be obtained as follows:

$$FC_{\mathrm{ad}} = 100 - M_{\mathrm{ad}} - A_{\mathrm{ad}} - V_{\mathrm{ad}} \tag{5}$$

2.4 Analysis of Experimental Results

In industrial production, different coal-based waste recycling technologies are often used based on the calorific value. The coal gangue is classified as high-calorific coal gangue (with the calorific value of 6270-12550 kJ/kg), medium-calorific coal gangue (with the calorific value of 2090-6270 kJ/kg), and low-calorific coal gangue (with the calorific value of <2090 kJ/kg). High calorific coal gangue is often utilized for power plants, while medium and low calorific of that are mostly used in the production of building materials [2]. Obviously, compared to the coal, coal gangue is a low-calorific solid waste during mining.

The experimental results of 1500 coal gangue exemplars are displayed in Fig. 1. The horizontal coordinate is $Q_{\mathrm{b,ad}}$ (MJ/kg), and the vertical coordinate is the industrial component analysis of coal gangue, including M_{ad}, A_{ad}, V_{ad} and FC_{ad} (%).

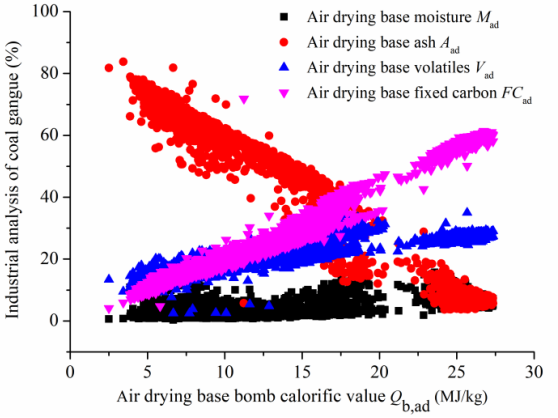

Fig. 1. Distribution of coal gangue database.

It was shown in Fig. 1 that the measured and collected coal gangue calorific values are distributed between 2MJ/kg-28MJ/kg. Most of the coal gangue calorific values are concentrated between 4MJ/kg-18MJ/kg (i.e., 4000kJ/kg-18000kJ/kg), and a small portion is distributed near 25MJ/kg. The number of coal gangue samples with calorific values between 20MJ/kg-24MJ/kg is almost negligible.

To avoid mixing other ore solid materials into the coal gangue during mining, leading to the calorific value of coal gangue that is not consistent with reality, these abnormal data were screened out and removed. Considering the low possibility of coal gangue having comparable calorific value to coal, in this study, only data sets with coal gangue calorific values less than 18MJ/kg were selected for model training and learning. Similarly, there are a small number of sample data in the coal gangue calorific value dataset that are less than 2MJ/kg.

To guarantee the reasonable validity of the inputs for prediction model, experimental data with coal gangue calorific values distributed between 2MJ/kg and 18MJ/kg were selected. Fig. 2 presents the coal gangue calorific value and industrial analysis database established by our research institute. To better observe the distribution of the coal gangue database, Fig. 3 also provides the distribution and proportion of coal gangue samples with calorific values in different intervals.

It can be obtained from Fig. 2 and Fig. 3 that the calorific values of coal gangue mainly between 4MJ/kg-12MJ/kg, accounting for a high proportion of 87.3%. The proportions of 4MJ/kg-6MJ/kg, 6MJ/kg-8MJ/kg, and 8MJ/kg-10MJ/kg are 23.8%, 40.1% and 23.5%, respectively. There are fewer distributions in the ranges of 2MJ/kg-4MJ/kg and 12MJ/kg-18MJ/kg. The proportions of 2MJ/kg-4MJ/kg and 16MJ/kg-18MJ/kg are both only 0.3%, while the proportions of 12MJ/kg-16MJ/kg is 3.4%. Obviously, the calorific values of the established coal gangue database mainly distribute in the range of 4MJ/kg-12MJ/kg, which belong to high-calorific and medium-calorific gangue suitable for power generation. Therefore, accurately predicting the calorific values of coal gangue in this range will directly improve the utilization rate and value of related resources.

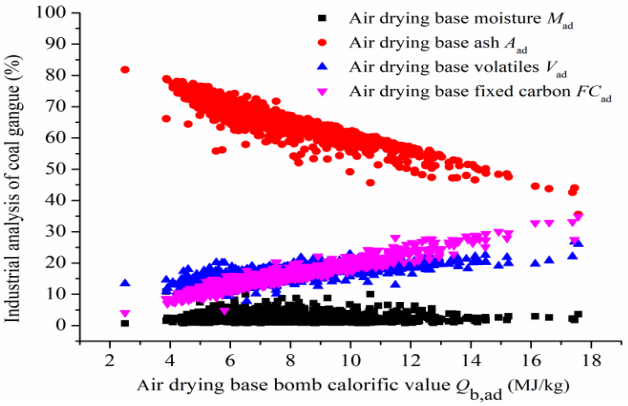

Fig. 2. Prediction model database.

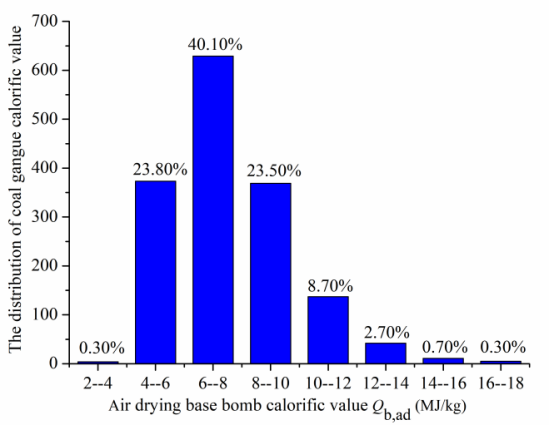

Fig. 3. Distribution of calorific value.

3 Prediction Model of Calorific Value of Coal Gangue

3.1 Support Vector Regression Model

SVR is a machine learning algorithm for solving regression problems [16]. This algorithm performs linear regression on sample data by searching for the optimal hyperplane in the feature space. For a given sample set $S=\{(x_1,y_1),(x_2,y_2),\cdots(x_n,y_n)\}, y_i \in \mathrm{R}$, it is mapped to a multi-dimensional feature space by a nonlinear transformation function $f(x)$. The relationship between input data x and output data $f(x)$ is as follow:

$$f(x)=w^T\phi(x)+b \tag{6}$$

Assuming that there can be a maximum deviation of ε between $f(x)$ and y, we can construct a width of 2ε interval centered at $f(x)$, extending ε on both sides. If the sample data falls within this interval, the forecasted result is regarded correct. SVR can be equivalently transformed into a problem of solving the optimization.

$$\min_{w,b,\xi^*} R_\varepsilon = \frac{1}{2}w^T w + C\sum_{i=1}^{n}(\xi_i^* + \xi_i) \tag{7}$$

Where, C represents the Penalty factor, ξ_i^* and ξ_i represent relaxation factor.

3.2 Random Forest Prediction Model

RF is an extensive version of decision trees. It solves prediction problem by combining multiple models, as shown in Fig. 4. RF constructs multiple decision trees and combines

them for prediction [17]. The basic idea is to create a new training set Q_1 by sampling from the original dataset. Then, from Q_1, we randomly select a subset Q_2 of n samples and k features. A decision tree is built by using Q_2. After each iteration, the sampled samples are put back into the dataset and the process is repeated to create a forest of m decision trees. The final prediction result of RF is obtained by averaging the predictions of all decision trees in the forest. RF has the advantages of randomly selecting samples and features, effectively avoiding overfitting, and having a simple implementation process with strong noise resistance and stable model performance. However, it has a longer training time compared to other models.

3.3 BP Neural Network Prediction Model

BPNN is a type of feed-forward neural network with hidden layers for error back-propagation training [18]. The core idea of the BPNN is to use gradient search techniques to minimize the mean squared error between the predicted and actual outputs of the neural network, as shown in Fig. 5.

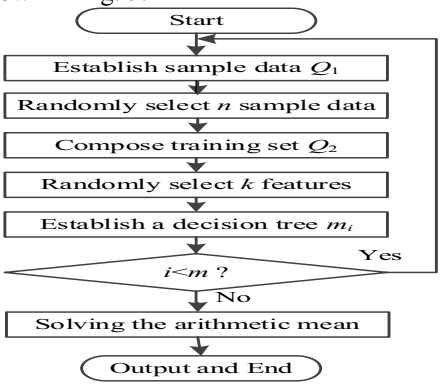

Fig. 4. Flowchart of random forest.

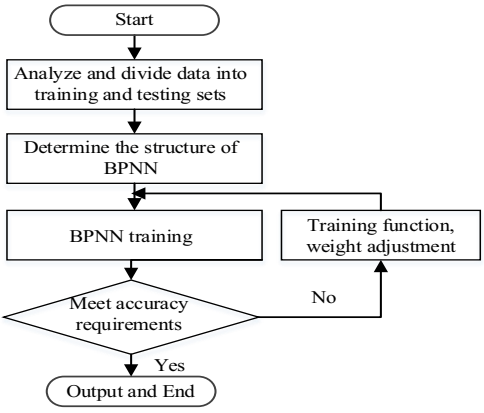

Fig. 5. Flowchart of random forest.

From the flowchart of Fig. 5, it can be seen that the structure of each layer is a combination of numerous neurons, and neurons in adjacent layers can be connected, while neurons in the same layer are not connected. The number of the input layer, the hidden layer and the output layer are *N*, *L* and *M*, respectively. The state of each neuron is only influenced by the result of signal processing from the neurons in the previous layer. If the outputs of the output layer is not up to the expected level after judgment, it will become a backward propagation. The BPNN has efficient computational capabilities, can reflect many complex mapping relationships, and has comprehensive recognition capabilities.

3.4 Particle Swarm Optimization Algorithm

There are key parameters that affect the prediction accuracy in various prediction algorithms. For the purpose of further promote the performance of the coal gangue calorific value forecasting model, the PSO algorithm is adopted. The mean square error (MSE) of the outputs of the presented model are employed as the objective function of the optimization method to carry out a large-scale search for the optimal key parameters of the prediction model.

$$\mathrm{MSE} = \frac{1}{n} \times \sum_{i=1}^{n} (q_i - y_i)^2 \tag{8}$$

Where, *n* denotes the total scale of training samples, *y* stands for the output value of the training samples, and *q* suggests the target value of the output.

The PSO algorithm was initially proposed by Kennedy as powerful and widely used swarm intelligence optimization algorithm to solve optimization problems [19]. It's found that the members of a bird flock always share information about the location of food while foraging, which accelerates the act of seeking out food. The answer for an optimization problem is seen as a bird in the search space, and the bird changes its location constantly to find the location of food during its flight.

In this algorithm, each particle has a memory and can track its best position from the previous iteration, including the individual best position X_{ipbest} and the global best position X_{gbest}. There are *n* particles in a *d*-dimensional search space. The position and velocity of the $i(i = 1, 2, \cdots, n)$ particle at the *t*-th iteration are represented by vectors $\mathbf{X}_i^t = (x_{i,1}^t, x_{i,2}^t, \cdots, x_{i,d}^t)$ and $\mathbf{V}_i^t = (v_{i,1}^t, v_{i,2}^t, \cdots, v_{i,d}^t)$, respectively. The update method for the *i*-th particle at the (*t*+ *1*)-th iteration is as follows:

$$\mathbf{V}_i^{t+1} = \omega \mathbf{V}_i^t + c_1 r_1 (\mathbf{X}_{ipbest}^t - \mathbf{X}_i^t) + c_2 r_2 (\mathbf{X}_{gbest}^t - \mathbf{X}_i^t) \tag{9}$$

$$\mathbf{X}_i^{t+1} = \mathbf{X}_i^t + \mathbf{V}_i^{t+1} \tag{10}$$

Where, ω is inertia weight factor, c_1 and c_2 are learning factors. r_1 and r_2 are random variables subject to uniform distribution $U(0,1)$.

To solve a minimization optimization problem, the best positions $\mathbf{P}_{best}$ and $\mathbf{g}_{best}$ experienced by particles and the population are determined by Eq.(11) and Eq.(12), respectively.

$$\mathbf{P}_{best}(t+1)=\begin{cases}\mathbf{P}_{best}(t), & \text{if } f(\mathbf{P}_{best}(t))<f(x_i(t+1))\\ x_i(t+1), & \text{if } f(x_i(t+1))<f(\mathbf{P}_{best}(t))\end{cases} \tag{11}$$

$$\mathbf{g}_{best}(t+1)=\min\left\{\mathbf{p}_1(t+1),\mathbf{p}_2(t+1),\cdots\mathbf{p}_n(t+1)\right\} \tag{12}$$

Where, $f(\cdot)$ represents the fitness function for each particle, which is also regard as the objective function for the optimization problem.

4 Results and Analysis

4.1 Evaluation Index of Prediction Results

In this study, three evaluation metrics are elected to evaluate the prediction effectiveness of the forecasting model, including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Coefficient of Determination (R^2). The calculation equation has been provided in reference [13]. The small values of MAE and RMSE indicate that the predicted values are closer to the actual sample values. The Coefficient of Determination (R^2) is used to evaluate the degree of fit between the predicted values and the actual sample values. The closer its value is to 1, the closer the predicted values are to the actual values.

4.2 Initial Parameter Setting of Algorithm

The key parameters of various prediction model algorithms directly affect the accuracy of the prediction results. In this study, PSO algorithm was introduced to optimize the key parameters of the prediction model. Similarly, the optimization algorithm also has some key parameters, including population size and maximum iteration times, which need to be determined in the initial stage. The optimization range or values are shown in Table 1.

Table 1. Optimization range or values of key parameters for prediction models.

Prediction models	Key parameter	Optimization range/Values
SVR	Penalty factor C	[100,10000]
	Kernel function coefficient σ	[0.0001,1]
RF	Number of decision trees n_estimators	[10,1000]
	Maximum depth of decision tree max_depth	[50,1000]
BPNN	Penalty coefficient alpha	[0.0001,1]

	Maximum number of iterations max_iter	[50,500]
	Number of hidden layer nodes hidden_layer_sizes	[10,10,10,10]
	Population size	30
PSO	Maximum number of iterations	250
	Inertia weight factor ω	0.8
	Learning factors c_1 / c_2	0.2/0.2

4.3 Comparison and Discussion of Prediction Performance

This study used the experimental coal gangue industrial analysis and calorific value database as the dataset. The industrial analysis of Mad, A_{ad}, V_{ad} and FC_{ad} were used as input, and $Q_{b,ad}$ was used as the output prediction value. Three nonlinear prediction models, SVR, RF, and BPNN, combined with PSO algorithm were used to forecast the calorific value of coal gangue. 1500 samples were divided into training set and test set based on the ratio of 8:2, and cross validation was carried out on the prediction results. To reduce the randomness of the optimization process, each prediction model was independently run 20 times, and the minimum mean squared error of the model output was used as the objective function for optimization. The capability of the prediction models established in this study was compared and analyzed. Table 2 gives the best optimization results of different prediction models. Fig. 6 and Fig. 7 show the change curve of fitness value with iteration times and the coal gangue calorific value prediction results corresponding to the best optimization results. At the same time, the average value of each index of 20 prediction results was calculated to obtain the comprehensive prediction index of the model, as shown in Table 3.

Table 2. Comparison of Best Optimization Results for Predictive Models (Optimal).

Methods	E_{MAE}	E_{RMSE}	R^2	Optimization parameter values
PSO-SVR	0.2617	0.3569	0.9706	(2.53e+02, 3.17273285e-02)
PSO-RF	0.3024	0.4247	0.9585	(100, 843)
PSO-BPNN	0.2661	0.3622	0.9698	(7.8276858e-01, 499)

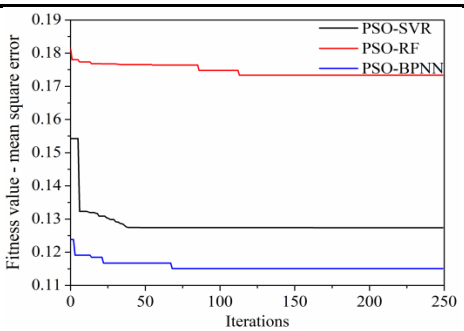

Fig. 6. Fitness value with iteration.

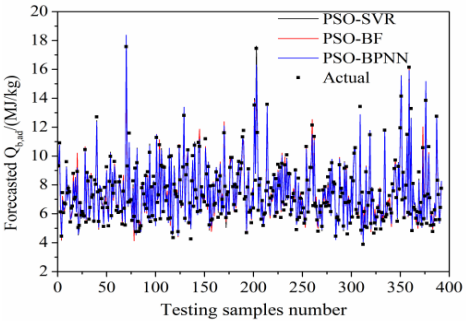

Fig. 7. Prediction results of calorific value.

Table 3. Comparison of Average Optimization Results of Prediction Models (20 Times).

Methods	E_{MAE}	E_{RMSE}	R^2
PSO-SVR	0.2625±6.93E-03	0.3586±1.49E-02	0.9704±2.50E-03
PSO-RF	0.3072±6.73E-03	0.4337±1.46E-02	0.9566±2.96E-03
PSO-BPNN	0.2707±6.22E-02	0.3755±1.62E-02	0.9675±2.84E-03

According to the results shown as Table 2, Fig. 6 and Fig. 7, it is known that among the best optimization result obtained from 20 times, the average absolute error and RMSE of the prediction models based on SVR and BPNN are relatively small, and R^2 is relatively large, indicating higher prediction accuracy. Among them, PSO-SVR model converged after about 40 iterations, and its convergence speed is clearly superior to other prediction models. Therefore, it is verified that the PSO-SVR model can achieve high prediction accuracy and fast convergence speed to predict the calorific value of coal gangue. However, the prediction performance of the PSO-RF model is lower and its convergence time is longer, which suggest that the model is unsuitable for predicting the calorific value of coal gangue.

Based on analysis of the comprehensive results of 20 times in Table 3, it can be seen that among the various prediction models used in the prediction of coal gangue calorific value, the prediction accuracy of the PSO-SVR model is the highest, slightly higher than that of the PSO-BPNN model, and the accuracy of the PSO-RF model is relatively low. Considering both the prediction accuracy and convergence speed, the PSO-SVR prediction model is appreciate for predicting the calorific value of coal gangue.

5 Conclusion

This paper conducted industrial analysis and calorific value measurements on coal gangue from major mining areas in Xinjiang using experimental methods, constructed a database of 1500 sets of coal gangue industrial analysis indicators and calorific values, and summarized and analyzed the distribution patterns of coal gangue calorific values. Using M_{ad}, A_{ad}, V_{ad} and FC_{ad} from industrial analysis as input data, and $Q_{b,ad}$ as the output prediction value, three nonlinear prediction models of SVR, RF and BPNN were adopted. With the deployment of the PSO algorithm to optimize the critical parameters

of the models, a prediction model for coal gangue calorific value was established. Through a comparison of several constructed coal gangue calorific value prediction models, it was found that both the SVR and BPNN prediction model have high accuracy, while the RF model has lower prediction accuracy and a longer convergence time. Among the constructed prediction models, the proposed model has the fastest convergence speed and high accuracy, making it suitable for predicting coal gangue calorific value. In summary, the coal gangue calorific value prediction model proposed in this study, which combines parameter optimization with machine learning algorithms, can effectively predict and analyze coal gangue calorific value. This approach is more efficient, convenient, and cost-effective than traditional experimental calorific value testing methods, providing a new way of thinking for the determination and analysis of fuel calorific values such as coal gangue.

Funding

This work was supported by the Science and Technology Project of State Grid Corporation of China (grant No. SGXJXN00TSJS2200141).

References

1. Yang, C. J.: Research and thinking about coal gangue backfilling technology. Mining Safety & Environmental Protection 1, 104–108 (2020).
2. Du, H.: Research on coal gangue disposal and comprehensive utilization. China Energy and Environmental Protection 3, 139–145 (2022).
3. Kumar, B., Sinha, A., Chakrabarti, S., et al.: A fast learning algorithm for one-class slab support vector machines. Knowledge-Based Systems 228(7), 107267 (2021).
4. Linardatos, P., Papastefanopoulos, V., Kotsiantis, S.: Explainable AI: A review of machine learning interpretability methods. Entropy (Basel, Switzerland) 23(1), 18–63 (2021).
5. Gandhi, M.: Formulas for calculating the calorific value of coal and coal chars: Development, tests, and uses. Fuel Processing Technology 7(1), 11–22 (1983).
6. Given, P.H., Weldon, D., Zoeller, J.H.: Calculation of calorific values of coals from ultimate analyses: theoretical basis and geochemical implications. Fuel 65(6), 849–854(1984).
7. Chehreh, C.S., Makaremi, S.: Explaining the relationship between common coal analyses and Afghan coal parameters using statistical modeling methods. Fuel Processing Technology 110, 79–85(2013).
8. Tie, C.: Estimation of calorific value of coal based on commercial coal test data. Shandong Coal Science and Technology 5, 207–209(2022).
9. Wang, H. X., Chen, Z. Y., Wang, Y. H., et al.: Prediction model of Guizhou coal calorific value and industrial analysis. Clean Coal Technology S01, 112–115(2020).
10. Patel, S. U., Kumar, B. J., Badhe, Y. P., et al.: Estimation of gross calorific value of coals using artificial neural networks. Fuel 86(3), 334–344(2007).
11. Kavšek, D., Bednárová, A., Biro, M., et al.: Characterization of Slovenian coal and estimation of coal heating value based on proximate analysis using regression and artificial neural networks. Central European Journal of Chemistry 11(9), 1481–1491(2013).

12. Jiang, W. H., Wei, H. Q., Qu, T. Z., et al.: Predication of the Calorific value for fuel coal based on the support vector regression machine with parameters optimized by genetic algorithm. Thermal Power Generation 03, 14–19(2011).
13. Li, D. H., Li, Q. K., Wang, W. C., et al.: Prediction of coal calorific value based on MIV characteristic variable selection and PSO-BP neural network. Coal Engineering 52(11), 154–160(2020).
14. Gao, X. B., Jia, B., Li, G., et al.: Calorific value forecasting of coal gangue with hybrid kernel function–support vector regression and genetic algorithm. Energies 15, 6718(2022).
15. Yang, Y. L., Wu, Q. Y.: The study of determination method of calorific value for the low heat value coal. Coal and Chemical Industry 42(10), 155–160(2019).
16. Liu, Y., Wang, L., Gu, K.: A support vector regression (SVR)-based method for dynamic load identification using heterogeneous responses under interval uncertainties. Applied Soft Computing 110, 107599(2021).
17. Peter, D. F. I.: A random forest approach to improve estimates of tributary nutrient loading. Water Research 1(248), 120876(2024).
18. Ma, Q. Z., Yue, C. P., Yu, M., et al.: Research on fault diagnosis strategy of air-conditioning system based on signal demodulation and BPNN-PCA. International Journal of Refrigeration 2(158), 124–134(2024).
19. Liu, S.B., Xue, P., Lu, J. Y., et al.: Fitting analysis and research of measured data of SAW yarn tension sensor based on PSO–SVR model. Ultrasonics 116, 106511(2021).