International Journal of Multiphysics
Volume 18, No. 3, 2024
ISSN: 1750-9548

Prediction Model of Coal Spontaneous Combustion Based on PCA-GWO-SVM

Junwen Zheng*
School of Business Administration, Liaoning Technical University, Huludao 125000, Liaoning, China
*Corresponding Author.

Abstract

To effectively prevent coal spontaneous combustion disasters, the paper proposes coal spontaneous combustion prediction model based on PCA-GWO-SVM. The principal component analysis (PCA) method is used for attribute approximation and reduction of feature indicators with correlation. The grey wolf optimization algorithm (GWO) algorithm is introduced to optimally select the penalty parameter C and kernel parameter g of support vector machines (SVM). Taking 40 sets of historical data of coal spontaneous combustion in Xuandong No.2 coal mine as the research object, 30 sets of data are selected as training samples and the remaining 10 as prediction samples, the PCA-GWO-SVM model is trained and tested, and the predicted results were compared with those of Fisher and back propagation neural network (BPNN) models. The results show that the PCA method can eliminate the interaction between indicators to reduce the complexity of the model. The GWO algorithm can effectively improve the learning ability of the SVM algorithm. The coal spontaneous combustion prediction model based on PCA-GWO-SVM has higher prediction accuracy and good stability.

Keywords: Coal spontaneous combustion, principal component analysis (PCA), grey wolf optimization algorithm (GWO), support vector machines (SVM), prediction model.

1. Introduction

As the world's largest carbon emitter and the most complete industrial categories' country, China is in the stage of rapid development of urbanization and faces multiple challenges, such as economic transformation, environmental protection, and coping with climate change [1]. It is more necessary to guarantee national energy security and build a modern energy system as an essential way to achieve the 'double carbon' goal and accelerate the construction of a vital energy country [2]. Coal is the primary energy source in China's disposable energy structure because China's energy structure is characterized by coal-rich, oil-poor, and gas-poor [3,4]. However, spontaneous coal combustion, as one of the five major natural disasters in coal mines, seriously restricts the safe and sustainable development of coal mines. Spontaneous coal combustion threatens mine workers' lives and further leads to substantial economic losses. In order to effectively prevent the occurrence of spontaneous coal combustion disasters, the key is to make timely and accurate predictions of the spontaneous combustion of coal.

In the field of coal spontaneous combustion research, the traditional coal spontaneous combustion prediction methods include establishing gas and temperature mathematical models, cluster analysis, gas correlation analysis, thermodynamic monitoring method [5-10]. Furthermore, some scholars proposed corresponding prediction methods by using statistical analysis theory and data mining algorithms. These methods were successfully applied to the prediction of spontaneous coal combustion. For example, Zhou and Qi [11] and Li et al. [12] conducted a study on the propensity to spontaneous combustion and residual coal. They established a fuzzy comprehensive judgment and grey correlation model. The predictions of the model were in general agreement with the actual results. To address the spontaneous coal combustion in the coal mine goaf, scholar Zhao et al. [13] designed a highly integrated real-time on-line analysis and detection system for multiple indicator gases using tunable semiconductor laser absorption spectroscopy and multiplexed phase-locked technology. The system provided an adequate data guarantee for coal mine safety production and early warning. Luo [14], Gao et al. [15] and Bian et al. [16] used existing natural predictorion indicators to establish a black propagation (BP) neural network prediction model to determine the likelihood of coal spontaneous combustion in the future. Zhong and Sun [17] combined particle swarm algorithms and extreme learning machines to establish a model. The model was successfully applied to the problem of predicting spontaneous coal combustion. Wen and Yu [18] used kernel principal component analysis for nonlinear extraction of coal spontaneous combustion indicators. He combined with Fisher's algorithm to establish a model for coal spontaneous combustion prediction with high prediction accuracy. Ding [19] used imbalance data as a prerequisite to judge the possibility of coal spontaneous combustion by using an improved Gaussian model for the study of gas emission. The model had strong applicability. The methods proposed by the above scholars have all contributed to the development of the field of coal spontaneous combustion prediction with high practical value. But there are still some limitations, such as the fuzzy comprehensive evaluation method to determine the indicator weights through the expert evaluation method. It carried a certain degree of subjectivity. The BP neural network model tends to fall into local optima because of the slow convergence speed. Fisher and SVM discriminant analysis rely too much on the original sample data. There are some drawbacks in the processing of nonlinear and high-dimensional data. It is difficult to determine the parameters in the process of application. In addition, most methods do not consider the redundancy of information between the predictors of spontaneous coal combustion. It may lead to some misjudgment problems.

Based on the above analysis, the paper uses PCA to extract the information on coal spontaneous combustion prediction indicators to eliminate the redundant information between indicators. Then, it introduces the GWO to optimize the parameter pairs (C, g) of SVM to improve the learning ability and generalization ability of SVM. Then, the GWO Algorithm is introduced to optimize the parameter pairs (C, g) of SVM to improve the learning ability and generalization ability of SVM. Finally, the PCA-GWO-SVM model is established to predict the spontaneous combustion of coal. Examples are also used to demonstrate the predictive validity of the model with solid generalization.

2. PCA-GWO-SVM theoretical model

2.1 PCA method

The PCA method is a mathematical approach to the dimensionality reduction of data [20]. Its basic idea is to try to replace the original indicators by recombining multiple indicators with a certain degree of correlation into a set of a smaller number of uncorrelated composite indicators through a linear transformation. Under the premise of ensuring the minimum loss of information, the primary information of the indicators is extracted using the PCA method. It can effectively eliminate the correlation between the indicators and achieve the purpose of dimensionality reduction. The mathematical model and specific calculation steps of PCA are as follows:

There are data sets $X = [x_1, x_2, \cdots, x_n]$ n groups of data, each corresponding to m eigenvalues.

(1) Data standardization: $Z_{ij} = \frac{x_{ij} - \overline{x_j}}{\sqrt{\sigma(x_j)}}$, where $\overline{x_j}$ is the column characteristic mean, $\sigma(x_j)$ is the standard deviation.

(2) Calculated out covariance matrix V with normalized matrix M.

(3) Compute the variance matrix V eigenvalues and eigenvectors: λ, α.

(4) The cumulative contributions corresponding to the first P principal components were calculated and derived: $\Phi(p) = \sum_{i=1}^{p} \lambda_i / \sum_{i=1}^{m} \lambda_i$.

(5) The variables corresponding to the downscaling of the P principal components Y: $Y = U^T \cdot X$.

Where $U = [u_1, u_2, \cdots u_p]; Y = (y_{ij})_{n \times p} = [y_1, y_2, \cdots y_p]^T$.

After dimensional reduction, p-dimensional data Y is transformed into m-dimensional data $\acute{X} = U \cdot Y$.

2.2 GWO algorithm

In 2014, Mirjalili proposed a GWO by mimicking the Grey wolf's racial leadership and hunting mechanisms. The social hierarchy of grey wolves needs to be modeled during the design of the GWO. The grey wolves with the optimal, sub-optimal, and third-best adaptation values in the grey wolf population are denoted as α, β, and δ. Respectively, the rest of the individuals are referred to as ω, where α denotes the alpha wolf, β and δ are the assisting wolves. The rest of the individual wolves, ω, follow the first three in prey encirclement, hunting, and attacking.

1) The mathematical description of the process of encircling prey is as follows:

$$D = \left| C_g X_P(t) - X(t) \right| \tag{1}$$

$$X_p(t+1) = X_p(t) - AgD \tag{2}$$

$$A = 2 \cdot \alpha \cdot r_1 - a \tag{3}$$

$$C = 2 \cdot r_2 \tag{4}$$

$$a = 2 - 2t/T \tag{5}$$

Where t denotes the current iteration number. T denotes the maximum iteration number. A and C denote the synergy coefficient vectors. X_p denotes the global optimal solution vector, i.e., the location of the prey. X denotes the potential solution vector, i.e., the location of the wolves. a denotes the convergence factor. r1, r2 denote the stochastic vectors with values ranging in the interval of [0, 1].

2) The mathematical description of the prey hunting process is as follows:

$$D_\alpha = \left| C_1 \cdot X_\alpha - X(t) \right| \tag{6}$$

$$D_\beta = \left| C_2 \cdot X_\beta - X(\mathrm{t}) \right| \tag{7}$$

$$D_\delta = \left| \mathrm{C}_3 \cdot X_\delta - X(\mathrm{t}) \right| \tag{8}$$

$$X_1 = X_\alpha - A_1 \cdot D_\alpha \tag{9}$$

$$X_2 = X_\beta - A_2 \cdot D_\beta \tag{10}$$

$$X_3 = X_\delta - A_3 \cdot D_\delta \tag{11}$$

$$X(t+1) = (X_1 + X_2 + X_3)/3 \tag{12}$$

3) The attack hunting process, i.e., the GWO algorithm to obtain the optimal solution. It is mainly realized by decreasing a-value from 2 to 0. The corresponding A-value obtains any value in the range of [-2a, 2a]. When |A| ≦ 1, the grey wolf group launches an attack on the prey centrally, equivalent to a local search. When |A| > 1, the grey wolf group disperses and re-searches for other local optimal solutions, equivalent to a global search.

2.3 GWO-SVM model fundamentals

Support vector machine [21] is a kind of neural network model based on the theory of structural risk minimization and VC dimension. It is suitable for solving small-sample, high-dimensional nonlinear problems. Its basic idea is to transform a nonlinear problem under low-dimensional space into a linear problem in high-dimensional space by nonlinearly transforming through the kernel function. Then classify the samples linearly in this high-dimensional space, so as to achieve the low-dimensional nonlinear classification under the input space.

2.4 Process design of the PCA-GWO-SVM model

Based on the above study, the PCA-GWO-SVM prediction model is proposed. The specific computational flow is shown in Figure 1.

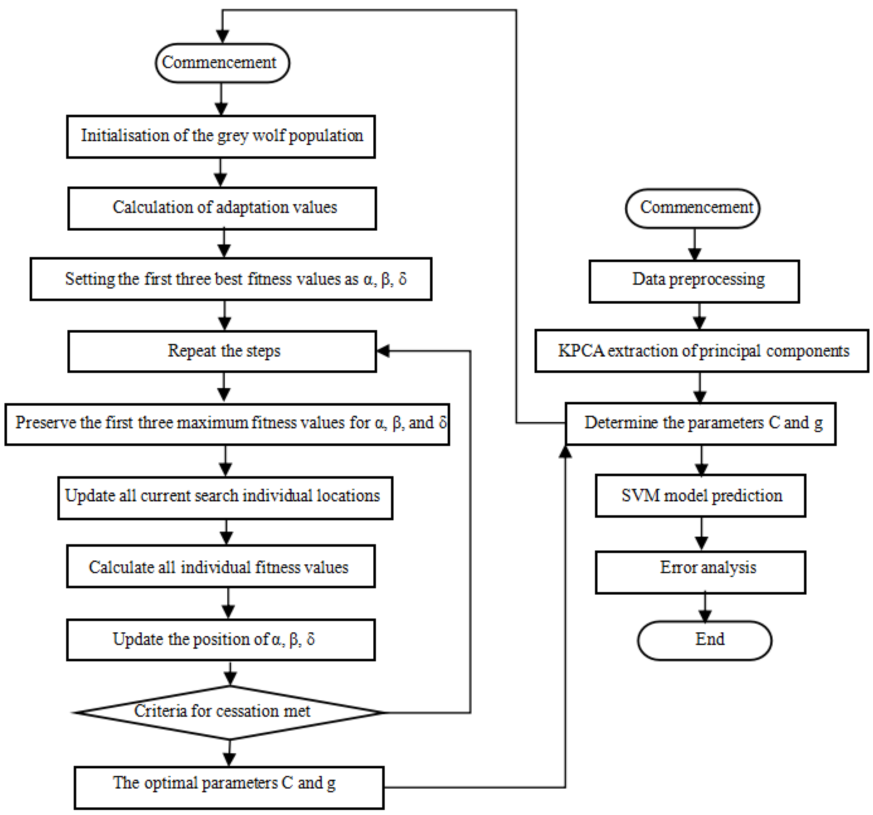

Figure 1 The specific calculation flow of PCA-GWO-SVM model

3. Establishment and Application of Coal Spontaneous Combustion Prediction Models

3.1 Selection of characteristic indicators

The spontaneous combustion of coal is a complex physicochemical process resulting from the joint action of many factors. Coal piled up in a fragmented state comes into contact with air, and an oxidation reaction occurs to generate heat. When the heat accumulates and can not be released, the coal temperature exceeds the ignition point, ultimately leading to the spontaneous combustion of coal. Therefore, temperature and oxygen are two critical indicators that predict the spontaneous combustion of coal. Coal produces many gases during oxidation, including CO, CO_2, alkanes, olefins, and other significant gases. There are considerable differences in gas generation in different spontaneous combustion states. Therefore, these gas compositions can be used as characteristic indicators for predicting coal spontaneous combustion. Considering the influence of detecting the spontaneous coal combustion gases in the mining area by using gas drainage boreholes. The parameters of the boreholes are chosen as the prediction indexes.

Based on the above study and referring to the literature [22]. The measure sample data of coal spontaneous combustion in Xuandong No.2 coal mine was used as the research object. Selection of eight influencing factors as characteristic indicators for coal spontaneous combustion prediction. They are temperature (x1/°C), volume fraction of O_2 (x2/%), volume fraction of CO (x3/%), volume fraction of CO_2 (x4/%), volume fraction of CH4 (x5/%), hole distance from the mining face (x6/m), horizontal depth of the hole into the open area (x7/m), and hole height (x8/m). The degree of spontaneous combustion of coal was classified into two categories: non-spontaneous (0) and spontaneous (1). Sample data (partial) are shown in Table 1.

Table 1 The original sample data.

Serial number	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	Degree of spontaneous combustion
1	20.3	19.8214	0.0004	0.7436	0.3371	94	5	42.8	0
2	46.9	11.2365	0.0096	0.8069	21.2323	144	19	44.3	0
…	…								…
30	36.8	14.3695	0.0075	1.0039	15.2323	214	73	52.1	0
31	23.8	14.9632	0.0065	0.5211	13.8653	155	16	46.9	0
…	…								…
39	125.4	6.0596	0.0152	0.2845	40.1236	96	18	44.3	1
40	107.4	14.6676	0.0069	0.5507	32.5821	150	18	36.4	1

3.2 PCA processing of data

Sample standardisation. In order to eliminate the influence of the differences in the scale between the indicators on the predicted results of the model, according to Eq: $x_{ij} = (a_{ij} - a_{min})/(a_{max} - a_{min})$. The raw samples are standardized using SPSS 19.0 statistical software, allowing the raw data to be regularised within the [0, 1] interval.

1)Correlation analysis. Pearson correlation analysis is carried out on the characteristic indicators using SPSS19.0 software. To eliminate the redundant information between the indicators and improve the predictive validity. The analysis of the correlation coefficient matrix (see Table 2) shows an apparent correlation between the characteristic indicators of spontaneous coal combustion. There is information overlap. Therefore, it is necessary to perform PCA on the indicator data.

2)Principal components wer are extracted using PCA. Principal component analysis is performed on the data through SPSS 19.0 software. In order to reduce the loss of information, the first four principal components are extracted. They are denoted as F1, F2, F3 and F4, respectively. The overall cumulative variance contribution is 85.464 percent. According to the variable coefficient vector matrix of PCA. The expression of the relationship between the extracted principal components and the standardized original variables $X = [x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8]$ is:

$$F_1 = 0.215x_1 - 0.227x_2 + 0.203x_3 - 0.165x_4 + 0.249x_5 - 0.173x_6 + 0.030x_7 + 0.115x_8 \quad (13)$$

$$F_2 = -0.156x_1 - 0.019x_2 + 0.143x_3 + 0.220x_4 + 0.039x_5 + 0.222x_6 + 0.539x_7 + 0.425x_8 \quad (14)$$

$$F_3 = 0.372x_1 + 0.011x_2 + 0.469x_3 + 0.573x_4 + 0.184x_5 + 0.449x_6 - 0.165x_7 - 0.359x_8 \quad (15)$$

$$F_4 = 0.411x_1 + 0.568x_2 - 0.208x_3 - 0.519x_4 + 0.284x_5 + 0.682x_6 - 0.260x_7 + 0.462x_8 \quad (16)$$

The four principal component scores are calculated based on each principal component expression. The results are (partially) shown in Table 2.

Table 2 Correlation coefficient matrix.

Serial number	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_1	1	-	-	-	-	-	-	-

x_2	-0.542	1	-	-	-	-	-	-
x_3	0.557	-0.596	1	-	-	-	-	-
x_4	-0.426	0.419	-0.159	1	-	-	-	-
x_5	0.783	-0.689	0.688	-0.463	1	-	-	-
x_6	-0.383	0.547	-0.277	0.520	-0.426	1	-	-
x_7	-0.148	-0.157	0.182	0.146	0.117	0.139	1	-
x_8	0.097	-0.266	0.322	-0.175	0.398	-0.104	0.512	1

3.3 Predictive modelling

Based on the 40 sets of coal spontaneous combustion sample data from Xuan Dong No. 2 coal mine provided by literature [20]. The study selects 30 samples for training. The remaining 10 were used as prediction samples. The four principal components extracted by PCA are used as inputs to the GWO-SVM prediction model. The type of coal spontaneous combustion degree is used as output. MATLAB R2016 b software is used to write the corresponding procedures to establish the GWO-SVM model for coal spontaneous combustion prediction with the help of the LIBSVM toolbox.

3.4 Tests and analyses of predictive models

The trained PCA-GWO-SVM coal spontaneous combustion prediction model is used to predict the remaining 10 samples. The prediction results are consistent with the actual results. From the experimental results, the model has a good prediction ability for coal spontaneous combustion. In order to verify whether the PCA-GWO-SVM prediction model is superior to other intelligent algorithms in the prediction of coal spontaneous combustion. It is compared with the Fisher model and the BPNN model processed by PCA. The prediction results are shown in Table 3. As shown from the Table 3, the prediction accuracy of the PCA-GWO-SVM model is better than the other two models.

Table 3 Comparison of discrimination results.

Serial number	F_1	F_2	F_3	F_4	Actual result	GWO-SVM Prediction results	Fisher Projected results	BPNN Projected results
31	-1.6744	0.3378	-0.7394	-0.1622	0	0	0	0
32	1.7479	-0.7213	-0.4046	0.2034	1	1	1	1
33	1.6030	-0.2912	-0.6273	0.3712	1	1	**0***	1
34	2.1283	1.9696	2.6361	1.1483	1	1	1	1
35	-1.4572	-0.3727	-0.1166	-0.2334	0	0	0	0
36	-1.2560	0.0622	-0.253	-0.0820	0	0	0	0
37	0.1251	-0.5406	0.8469	0.6127	1	1	1	1
38	2.4207	-0.2324	-0.3034	-0.3247	1	1	1	1
39	2.8892	-0.3551	0.9928	-0.0992	1	1	1	**0***
40	0.0565	-0.6052	0.9574	0.5684	1	1	1	1

In comparing the accuracy of the three models, it is essential to consider that the coupling phenomenon between a single set of data and the model may be unconvincing. So again, 6 sets of predictive sample data are randomly selected, each containing 10 predictive samples. The same three models are used to predict the six data groups. The results are given in the form of accuracy and comparison. The results are shown in Figure 2.

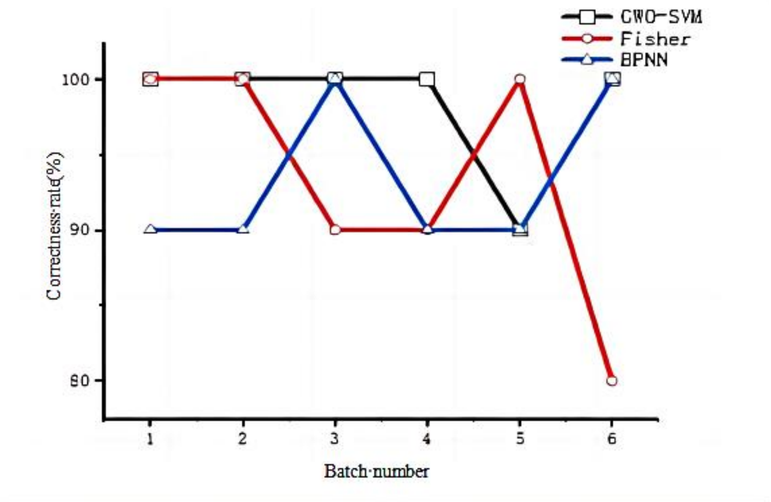

Figure 2 Model accuracy comparison chart

As can be seen from Figure 2, relative to the other two models, the GWO-SVM model maintains an accuracy of 90% and above in the prediction of six groups of randomly selected data. It shows good stability relative to Fisher and BPNN models. It shows that the model is more suitable for the problem of predicting spontaneous combustion in coal mines under multiple indicators and high dimensions.

In order to verify the optimization effect of the GWO algorithm, 40 samples in Table 1 are selected and written into the Python program for training. The best fitness curves of the optimized GWO-SVM model and the SVM model are shown in Figure 3 and Figure 4. It can be seen that the GWO algorithm speeds up the capture of the global optimal solution after optimizing penalty parameter C and kernel parameter g in the SVM algorithm. It is more adaptive than the SVM model.

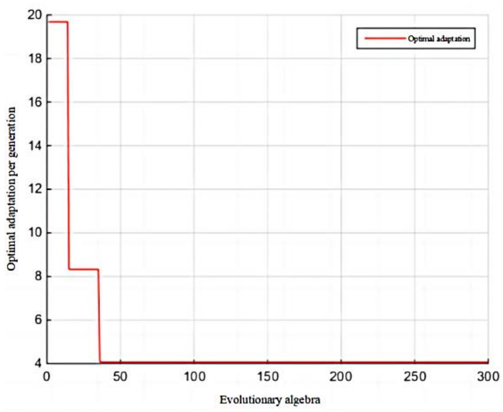

Figure 3 GWO-SVM Best fitness curve

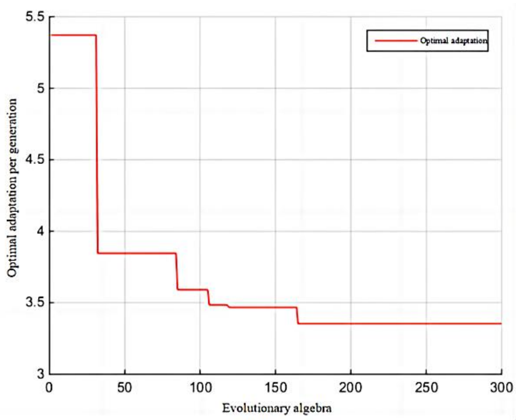

Figure 4 SVM Best fitness curve

4. Conclusion

Based on the SVM algorithm, the PCA-GWO-SVM coal spontaneous combustion prediction model is constructed. The model has high stability, the evaluation results are scientific to a certain extent, both the operation rate and the prediction accuracy are greatly improved, and the prediction of coal spontaneous combustion can be better completed. The conclusion are as follows:

1) The article uses the PCA method to attribute approximate reduction of coal spontaneous combustion characteristic indicators to eliminate the information overlap between indicators. It reduces the complexity of the model while achieving the purpose of dimensionality reduction and improves the algorithm operation efficiency.

2) Statistical theory and machine learning algorithms are applied to the coal spontaneous combustion prediction problem. The penalty parameter C and kernel parameter g of SVM are optimized using the GWO optimization algorithm. It improves the ability of the SVM model to find the global optimal solution with the prediction accuracy and stability of the model. The GWO-SVM model has a higher discriminative accuracy compared to Fisher and BPNN models. The prediction results are consistent with the actual results and have a more robust learning ability.

The data used for the spontaneous combustion prediction of coal are limited. The paper can not consider all the influencing factors. Moreover, only two cases of non-spontaneous combustion and spontaneous combustion are considered. The other states are analyzed and processed. Therefore, in future research on spontaneous combustion prediction, it is necessary to analyze the influencing factors of coal spontaneous combustion and study the coal in different states. In order to provide more reliable predictive information and enhance the applicability of the model.

References

[1] Zhang Xiliang, Huang Xiaodan, Zhang Da, et al. Research on the path and policy of energy economy transition under the goal of carbon neutrality. Journal of Management Word,2022,38(01):35-66.

[2] Liu Huajun, Shi Yin, Guo Lixiang, et al. China's Energy Revolution in the New Era: History, Achievements and Prospects. Journal of Management Word,2022,38(07):6-24

[3] Wu Qiang, Tu Kun, Zeng Yifan, et al. Discussion on the main problems and countermeasures faced by building an upgraded version of China's main energy (coal). Journal of China Coal Society,2019,44(06):1625-1636.

[4] Han Jianguo. Analysis of the path of "soft landing" of energy structure adjustment: developing clean coal utilization, solving the energy dilemma, and practicing the energy revolution. Journal of Management Word,2016,(02):3-7.

[5] Ren Wanxing, Guo Qing, Shi Jingtai, et al. Construction of early warning indicators for coal spontaneous combustion based on statistical characteristics of marker gases. Journal of China Coal Society,2021,46(06):1747-1758.

[6] Chen Liang, Wang Enyuan, Feng Junjun, et al. A Dynamic Gas Emission Prediction Model at the Heading Face and its Engineering Application. Journal of Natural Gas Science and Engineering.2016

[7] Guo Qing, Ren Wanxing, Lu Wei, et al. Evolution of spontaneous coal combustion gases and hazardous area delineation in a mining face. Journal of Coal Science,2023,48(S2):647-656.

[8] Tan Bo, Shao Zhuangzhuang, Guo Yan, et al. Research on coal spontaneous combustion classification early warning based on index gas correlation analysis. Chinese Safety Science Journal,2021,31(02):33-39.

[9] Jiang Meng, Feng Wenbin, Gao Hui,et al. Carbon monoxide sensor for monitoring thermal and dynamic disasters in coal mines. Journal of China Coal Society:1-6[2021-08-30].

[10] Patrick Booth, Heidi Brown, Jan Nemcik, et al. Spatial context in the calculation of gas emissions for underground coal mines. International Journal of Mining Science and Technical.2017

[11] Zhou Xinhua, and Qi Qingjie. Fuzzy comprehensive evaluation of the spontaneous combustion tendency of coal. Coal Geology &Exploration,2011,39(06):16-19+23.

[12] Li Yanchang, Li Fei, Hao Chaoyu. Based on AHP-TOPSIS Open-pit Coal Residual Coal Spontaneous Combustion Comparative Analysis. Mining Safety and Environmental Protection,2019,46(02):111-114+118.

[13] Zhao Xiaohu, Sun Pengshuai, Yang Juan, et al. It is applied to the online monitoring system of gas volume fraction of coal spontaneous combustion index. Journal of China Coal Society,2021,46(S1):319-327.

[14] Luo Dayong. Research on coal spontaneous combustion temperature prediction method based on BP neural network. Coal Technology,2020,39(09):111-112.

[15] Gao Ping, Bian Bing, Liu Zhijian. Forecasting System of Spontaneous Combustion of Coal Based on BP Neural Network. Coal Technology,2014,33(09):60-62.

[16] Bian Bing, Liu Zhijian, Gao Ping. Studies on Indicator Gases of Coal Based on BP Neural Network Forecasting System. Journal of Hebei United University(Natural Science Edition),2014,36(03):8-11.

[17] Zhong Weiqing, Sun Jianbiao. Application of ELM in prediction of coal spontaneous Combustion in caving zone. Journal of Liaoning Technical University(Natural Science),2016,35(06);581-585.

[18] Wen Tingxin, Yu Fengyu. Research on Prediction of Coal Spontaneous Combustion Based on KPCA－Fisher Discriminant Analysis. Mining Safety & Environmental Protection,2018,45(02):49-53+58.

[19] Dingwen Dong. Mine Gas Emission Prediction based on Gaussian Process Model. Procedia Engineering.2012

[20] Yan Xin, Zhu Yonghao, Tu Naiwei, et al. Prediction of coal and gas outburst in working face based on- PCA and weighted Bayes. Computer Engineering, 2021,47(08):315-320.

[21] Wen Tingxin, Sun Hongjuan, Xu Bo, et al. RS-SVM model for predicting underground goaf collapse risk. China Safety Science Journal,2015,25(10);16-21.
[22] Cai Ling. Application of rough sets and support vector machines in spontaneous combustion prediction of goafs. Fu Xin: Liaoning Technical University,2011.