Research Article

Real-Time Prediction Model of Coal and Gas Outburst

Ru Yandong [ID],[1] Lv Xingfeng,[2] Guo Jikun,[1] Zhang Hongquan,[1] and Chen Lijuan[1]

[1]*College of Electronics and Information Engineering, Heilongjiang University of Science and Technology, Haerbin, Heilongjiang 150027, China*
[2]*College of Computer Science and Technology, Heilongjiang University, Haerbin, Heilongjiang 150080, China*

Correspondence should be addressed to Ru Yandong; 10564006@qq.com

Coal and gas outburst has been one of the main threats to coal mine safety. Accurate coal and gas outburst prediction is the key to avoid accidents. The data is actual and complete by default in the existing prediction model. However, in fact, data missing and abnormal data value often occur, which results in poor prediction performance. Therefore, this paper proposes to use the correlation coefficient to complete the missing data filling in real time for the first time. The abnormal data identification is completed based on the Pauta criterion. Random forest model is used to realize the prediction model. The prediction performance of sensitivity 100%, accuracy 97.5%, and specificity 84.6% were obtained. Experiments show that the model can complete the prediction of coal and gas outburst in real time under the condition of missing data and abnormal data value, which can be used as a new prediction model of coal and gas outburst.

1. Introduction

China is one of the countries with serious coal and gas outburst disasters which pose a great threat to the coal mine safety production [1]. The accidents caused by coal and gas outburst account for 38% of safety accidents in coal mine [2], which is the most dangerous and frequent accident type in coal mine accidents. Before coal and gas outburst, it is of great significance to complete the prediction of coal and gas outburst to ensure the safety of production and protect the life of miners.

The common prediction methods of coal and gas outburst include the index prediction method and mathematical model prediction method [3]. The index prediction method is to detect the values of various indexes and compare them with the standard values of indexes to determine whether they are dangerous for coal and gas outburst. Common indexes include gas content, gas pressure, coal strength coefficient, cuttings index, and comprehensive index [4–6]. There are many influencing factors and large measurement errors in these indexes, which often lead to low prediction accuracy. Therefore, in recent years, researchers pay more attention to mathematical model prediction methods. The method is to select several characteristics that affect coal and gas outburst and use the mathematical model to predict the outburst risk [7–9]. In fact, the existing prediction methods are based on the ideal data to complete the prediction of coal and gas outburst. However, there is still a long distance from the actual application, which is mainly reflected in three aspects. First, the data used by default is complete. In fact, in the process of data transmission and data fusion, data is often missing, resulting in partial or even all data missing. Second, the data used by default has the same value as the real value. In fact, due to the limitations of experimental conditions and experimental process, there may be some differences between the data and the real value. Third, the prediction model is based on the previous data. When the new abnormal data is coming, there is a lack of processing capacity resulting in the inability to complete the prediction in real time. For this reason, this paper proposes a coal and gas outburst prediction model. According to the characteristics of data from the coal mine, the data processing method is given, which can accurately predict coal and gas outburst in practical application. The prediction of coal and gas outburst ensures the safety of miner and property of miners.

2. Real-Time Prediction Model

There are many factors that affect coal and gas outburst. Commonly used factors include gas content, gas pressure, coal strength coefficient, drilling cuttings index, comprehensive index, initial gas release speed, porosity, and coal seam thickness [10]. In this paper, five index parameters are selected including gas content, gas pressure, coal strength coefficient, initial gas release speed, and porosity [11]. The mathematical prediction model is used to predict coal and gas outburst, which is shown in Figure 1. Three kinds of data including gas content, gas pressure, and initial velocity of gas emission collected from the underground are transmitted to the well through the transmission substation. These three kinds of data and coal strength coefficient and porosity of the coal body form the joint dataset. Data processing is used to deal with abnormal data and missing data. The prediction of coal and gas outburst is completed by using the random forest model. In case of coal and gas outburst, the manager shall be informed to take relevant measures at the same time. The audible and visual alarm shall be started to inform the miner to start the relevant risk avoidance equipment. The information shall be transmitted to the underground through the transmission substation. At the same time, the data storage and update are completed, and the new prediction model is trained by using the updated dataset regularly.

3. Experiments

3.1. Data Description. The data is divided into accident data and safe data, in which the safe data is easy to obtain and the data volume is large [12]. However, it is difficult to obtain accident data, even if there are some records of accident data, which are often incomplete, resulting in a small amount of accident data. Therefore, this paper adopts different methods for safe data and accident data. When the safe data is missing, delete the data directly. When the accident data is missing, fill the data according to the existing data. In this paper, five indexes including gas pressure, gas content, initial velocity of gas emission, porosity, and coal strength coefficient, which are closely related to coal and gas outburst, are selected as characteristics. Partial characteristic data is shown in Table 1.

3.2. Data Processing. According to the characteristics of coal mine data, this paper completes the data processing. The processing block diagram is shown in Figure 2. First, check whether there is data missing in the dataset. If there is data missing, use the corresponding processing method to complete the data filling. Secondly, check whether there are abnormal values in the dataset. If there are abnormal values, use the abnormal value processing method to process the data.

3.2.1. Missing Data Processing Method. There are many methods of data filling, but they are based on the existing dataset [13]. As the time goes, the amount of data will

become very large, so it will take a lot of time to realize data filling so as to affect the prediction speed [7]. Therefore, this paper proposes a numerical filling data method based on the correlation of variables which can be considered as real time. For the first time, the correlation between existing data variables is considered, and Pearson correlation coefficient is used to complete the new missing data filling that ensures data integrity. Pearson correlation coefficient can be expressed by numerical value to measure the correlation between two variables, which can be described as follows:

$$\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{E((X - \mu_X)(Y - \mu_Y))}{\sigma_X \sigma_Y}, \quad (1)$$

where $E(X)$ is the mathematical expectation of the variable, σ_X represents the standard deviation of variable X, and μ_X represents the mean value of variable X. In this paper, the correlation matrix of five characteristics is calculated, and the results are shown in Table 2.

When data from a certain feature in the dataset is missing, this article observes the correlation coefficient in the correlation matrix and finds out the maximum value of the correlation coefficient between the characteristics of the data and other characteristics. This article makes use of the correlation coefficient to filling the missing value. In this paper, the calculated values obtained by using the correlation when the five characteristics are missing are listed, respectively. The experimental results are shown in Table 3.

It can be seen from Table 3 that the correlation coefficient matrix method is used to fill the missing data, and the similarity between the filled data value and the actual value is from 83.3% to 98.6%. On the whole, the filled data is very similar to the actual data which ensures higher prediction performance.

3.2.2. Abnormal Data Value Processing Method. Due to the limitation of equipment precision and experimental conditions, noise will inevitably be mixed into the data, which will lead to abnormal data values. In this paper, after processing the missing value of the data, the abnormal value in the data is found according to the Pauta criterion. Pauta criterion is described as follows.

When the value meets the formula,

$$\left| X_i - \overline{X} \right| > 3^* S_x. \quad (2)$$

X_i is considered to be an abnormal, where X_i is the data point, $\overline{x} = (1/n) \sum_{i=1}^n x_i$ is the data mean value, and $S_x = [(1/n) \sum_{i=1}^n (x_i - \overline{x})^2]^{1/2}$ is the standard deviation of data.

In this paper, 5 characteristics of the dataset are detected for the abnormal data value. When detecting the abnormal data value, the missing value filling method mentioned above is used for data correction. In this paper, 530 pieces of accident data and safe data are selected, and the above methods are used to complete the abnormal data identification. The abnormal identification results are shown in Table 4.

It can be seen from Table 4 that there are some abnormal data in the dataset. The main reason is that the accuracy of data acquisition equipment and the limitations of
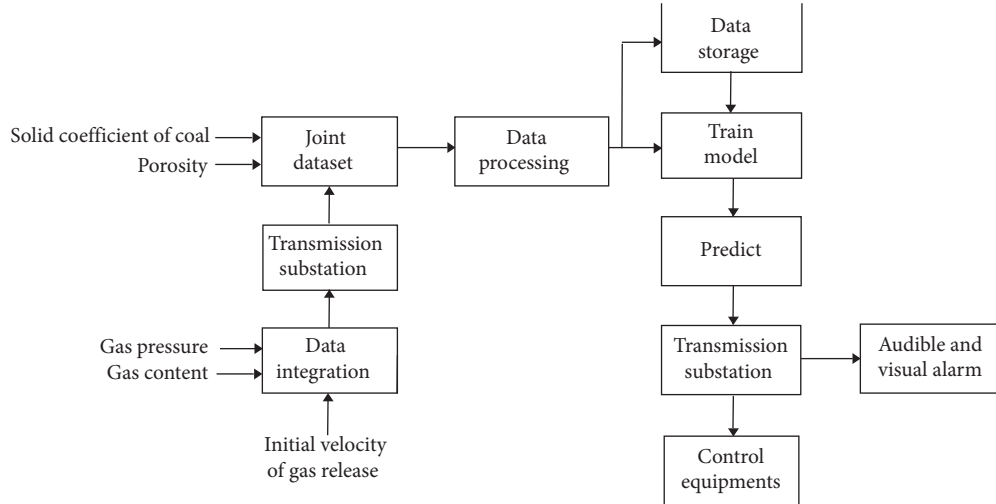
FIGURE 1: Block diagram of coal and gas prediction system.

TABLE 1: Characteristic data (partial).

Gas content	Gas pressure	Porosity	Coal strength coefficient	Initial velocity of gas release	Data class
17.40	1.40	6.20	0.52	10	Safe
16.42	1.60	3.60	0.36	15	Safe
13.62	1.80	4.36	0.39	8	Safe
12.34	0.70	7.36	0.35	13	Safe
11.26	0.39	9.60	0.59	14	Safe
9.54	2.95	3.50	0.24	11	Safe
9.32	4.40	6.30	0.57	10	Safe
8.36	3.10	8.30	0.45	12	Safe
7.55	0.97	4.36	0.52	7	Safe
7.36	2.95	3.60	1.30	7	Safe
8.87	1.20	5.49	1.11	8	Accidental
8.14	0.10	5.23	0.65	11	Accidental
7.80	0.89	5.80	1.30	5	Accidental
7.43	2.10	7.36	1.11	10	Accidental
6.67	0.80	3.56	0.65	6	Accidental
6.55	0.70	8.20	0.98	7	Accidental
5.30	0.65	7.36	0.98	2	Accidental
5.25	0.60	3.60	0.64	5	Accidental
4.38	0.89	8.20	0.64	5	Accidental
3.82	0.46	3.65	0.94	4	Accidental

experimental methods leading to part of the data value deviate from the actual value. Therefore, this part of abnormal signals must be identified and deleted, and the data filling method mentioned above should be used to complete the data filling, so as to ensure the high performance of prediction.

3.3. Realization of Coal and Gas Outburst Prediction.

In this paper, accuracy, sensitivity, and specificity are used as performance evaluation indexes of the model, which are defined as follows:

$$\text{accuracy} = \frac{TP + TN}{TP + FP + FN + TN}, \quad (3)$$

$$\text{sensitivity} = \frac{TP}{TP + FN}, \quad (4)$$

$$\text{specificity} = \frac{TN}{TP + FP}. \quad (5)$$

TP is the number of correctly predicted safe data. FP is the number of wrongly divided safe data into accidental data. FN is the number of wrongly divided accidental data into safe data. TN is the number of correctly predicted accidental data.

In order to achieve high-performance detection and avoid overfitting, this paper uses the random forest model to achieve coal and gas outburst prediction. Random forest (RF) is an efficient integrated classification method composed of a large number of decision trees. According to the characteristics of decision trees, many improvements have been made so that the decision trees constructed are as uncorrelated as possible, thus significantly improving the classification performance of the system. In this paper, the
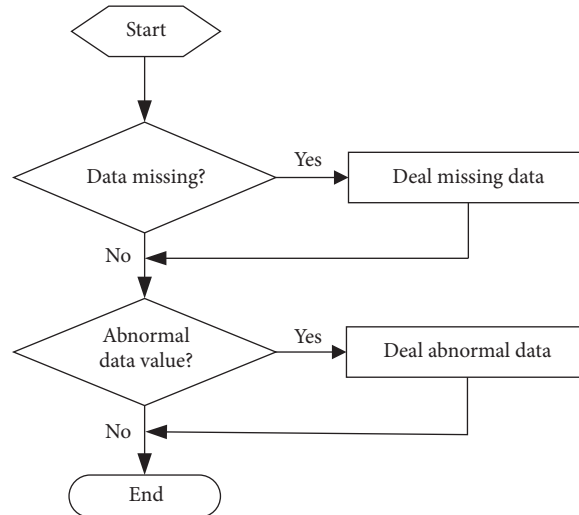
Figure 2: Overall block diagram of data processing.

Table 2: Correlation matrix of characteristics.

Correlation coefficient	Gas content	Gas pressure	Porosity	Coal strength coefficient	Initial velocity of gas release
Gas content	1	0.5518	0.0881	−0.3222	0.6531
Gas pressure	0.5518	1	−0.0005	−0.0787	0.4623
Porosity	0.0881	−0.0005	1	−0.1294	0.0901
Coal strength coefficient	−0.3222	−0.0787	−0.1294	1	−0.4413
Initial velocity of gas release	0.6531	0.4623	0.0901	−0.4413	1

Table 3: Calculation value of feature.

	Gas content	Gas pressure	Porosity	Coal strength coefficient	Initial velocity of gas release
(a) Calculation value of initial velocity of gas emission					
Experimental value	15.36	2	4.60	0.53	Null value
Calculated value	—	—	—	—	10.20
Actual value	15.36	2	4.60	0.53	10.06
Similarity degree	—	—	—	—	98.6%
(b) Calculation value of porosity					
Experimental value	19.72	1.90	Null value	0.15	13
Calculated value	—	—	5.70	—	—
Actual value	19.72	1.90	4.75	0.15	13
Similarity degree	—	—	83.3%	—	—
(c) Calculation value of coal seam firmness coefficient					
Experimental value	16.23	3	9.20	Null value	9
Calculated value	—	—	—	0.18	—
Actual value	16.23	3	9.20	0.16	9
Similarity degree	—	—	—	88.9%	—
(d) Calculation value of gas content					
Experimental value	Null value	1.30	4.60	0.26	13
Calculated value	23.10	—	—	—	—
Actual value	21.68	1.30	4.60	0.26	13
Similarity degree	93.9%	—	—	—	—
(e) Calculation value of gas pressure					
Experimental value	9.32	Null value	4.60	0.26	13
Calculated value	—	2.30	—	—	—
Actual value	9.32	2.10	4.60	0.26	13
Similarity degree	—	91.3%	—	—	—

grid search method is used to optimize the model parameters, the number of optimal decision trees is 500, and the number of random variables is 3. The 10 fold crossvalidation method is used to complete the cross validation, and the R language is used to realize the model. The prediction results are shown in Table 5.

TABLE 4: Abnormal data identification results.

	Gas content	Gas pressure	Porosity	Coal strength coefficient	Initial velocity of gas release
Data volume	530	530	530	530	530
Amount of abnormal data	15	10	1	5	10
Ratio of abnormal data to dataset (%)	2.83	1.89	0.19	0.94	1.89

TABLE 5: Prediction results.

	Safe data (predicted)	Accidental data (predicted)
Safe data (actual)	340	10
Accidental data (actual)	0	55

Through the prediction results in Table 5, we calculated that the model has high performance of sensitivity 100%, accuracy 97.5%, and specificity 84.6%. From the experimental results, we can conclude that the model can completely detect coal and gas outburst with high performance.

4. Conclusion

In this paper, the prediction model of coal and gas outburst is realized. The correlation between the five important characteristics affecting coal and gas outburst is analyzed. This paper completes the filling of missing data according to the correlation between characteristics for the first time which can guarantee to consume very little time. Abnormal values are identified and processed in order to ensure the analyzed data getting closer to the actual data. The processing method of missing data and abnormal value is proved be effective. The random forest model is used to predict coal and gas outburst with high performance. The experiment shows that it can complete the task of real-time prediction of coal and gas outburst with high performance of sensitivity 100%, accuracy 97.5%, and specificity 84.6%. It can be used for safety production of coal mine in practical application.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] L. Yuan and S. Xue, *Theory and Technology of Coal and Gas Outburst Prediction by Coal Seam Gas Content Method*, Science Press, Beijing, China, 2014.

[2] X. Zhao, H. Sun, J. Cao, X. Ning, and Y. Liu, "Applications of online integrated system for coal and gas outburst prediction: A case study of Xinjing mine in Shanxi, China," *Energy Science and Engineering*, vol. 8, no. 2, pp. 1980–1996, 2020.

[3] D. J. Black, "Review of coal and gas outburst in Australian underground coal mines," *International Journal of Mining Ence and Technology*, vol. 29, no. 6, pp. 4–13, 2019.

[4] W. Zhang and Y. Hou, "Study on coal mine gas data denoising based on Ga-LSSVR," *Mining Safety and Environmental Protection*, vol. 44, no. 1, pp. 45–48, 2017.

[5] B. Nie, X. He, and E. Wang, "Research status and development trend of coal and gas outburst prediction technology," *China Safety Science Report*, vol. 13, no. 6, pp. 40–44, 2003.

[6] L. Xinling, Y. Mei, and Y. He, "Prediction of coal and gas outburst grade based on factor analysis and SVM model," in *Proceedings of the 2019 International Conference*, Moscow, Russia, 2019.

[7] Ministry of Coal Industry of the People's Republic of China, *Provisions on Prevention and Control of Coal and Gas Outburst*, Coal Industry Press, Beijing, China, 2009.

[8] Z. Xusheng, M. A. Guolong, and Q. Wei, "Early-warning index system of coal and gas outburst based on fault tree analysis," *Mining Safety & Environmental Protection*, vol. 46, no. 3, 2019.

[9] H. Peng, *Research on Gas Disaster Prediction Method Based on Data Mining and Information fusion*, China University of mining and Technology, Beijing, China, 2013.

[10] Z. Zhao and Y. Tan, "Study on prediction of coal and gas outburst precursor time series based on chaos theory," *Geotechnical Mechanics*, vol. 30, no. 7, pp. 2186–2190, 2009.

[11] Y. Y. Zhang, J. L. Cui, and X. D. Jiao, "Study of the multi-index coupling forecasting model of coal and gas outburst and its application," *Chinese Journal of Engineering*, vol. 11, pp. 45–52, 2018.

[12] Li Sheng, *Study on Pattern Recognition Method for Regional Prediction of Coal and Gas Outburst*, Liaoning University of Engineering and Technology, Liaoning, China, 2004.

[13] J. Mou, H. Liu, Y. Zou, and Q. Li., "A new method to determine the sensitivity of coal and gas outburst prediction index," *Arabian Journal of Geoences*, vol. 13, no. 12, pp. 1–9, 2020.