Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 8, No. 3, 2024

Recognition and Prediction of Precursory Feature Signals of Coal Mine Rock Burst Based on Random Forest and MK Trend Test

Tianhang Yang *

Sun Yueqi Honors College, China University of Mining and Technology, Xuzhou, Jiangsu, China

*** Corresponding author:** Tianhang Yang

Abstract: This paper aims to identify and predict the precursory characteristic signals of coal mine rock burst by employing the Random Forest and MK (Mann-Kendall) trend test methods. Initially, the study conducts data preprocessing and analysis on indicators such as electromagnetic radiation intensity, acoustic wave intensity, and type. It uses clustering methods to distinguish various data types and analyzes the data after visual representation. The visualization of data illustrates the distribution and trends, which aids in understanding the characteristics of the data, such as the upward trend observed in precursory feature data. During the data preprocessing phase, diagrams of various classes of data for Acoustic Emission (AE) and Electromagnetic Radiation (EMR) intensity are presented, providing an intuitive reference for subsequent analysis. The research utilizes the Random Forest algorithm and MK trend test to recognize precursor signals and predict their occurrence intervals. The Random Forest model is chosen for its efficiency and accuracy in handling classification and regression issues, while the MK trend test provides a statistical basis for identifying precursory signals by analyzing monotonic trends within the dataset. This study not only enhances the accuracy of precursory signal identification for coal mine rock bursts but also offers scientific early warning and control measures for coal mine safety production, which is of significant practical value.

Keywords: Coal Mine Rock Burst; Precursory Characteristic Signals; Random Forest; MK Trend Test; Perceptron.

1. Introduction

Coal, as a vital component of global energy, plays a crucial role in economic and social development. However, as coal mining deepens, safety issues in coal mines become increasingly prominent, especially rock burst accidents that not only threaten the lives of miners but also lead to significant economic losses [1]. Therefore, effective prediction and early warning of coal mine rock bursts have become a key issue in the field of mine safety [2-3].

Rock bursts, as a complex mining dynamic phenomenon, are often accompanied by a variety of precursory feature signals. These signals include, but are not limited to, electromagnetic radiation, acoustic wave activity, and other physical phenomena. Identifying and interpreting these precursory signals are of great significance for predicting rock bursts. However, due to the complexity and variability of these signals, traditional prediction methods often struggle to accurately identify and predict rock burst events [4-5].

To improve the accuracy of predictions, this paper proposes an integrated method based on the Random Forest and MK (Mann-Kendall) trend test. The Random Forest, as an ensemble learning technique, enhances model performance and generalization ability by constructing multiple decision trees and voting or averaging [6]. The MK trend test is a non-parametric statistical method that can effectively detect trend changes in a dataset [7]. By combining these two methods, this paper aims to build a robust recognition and prediction model to improve the identification rate and prediction accuracy of precursor signals of coal mine rock bursts.

2. Data Preprocessing

2.1. Data Cleaning

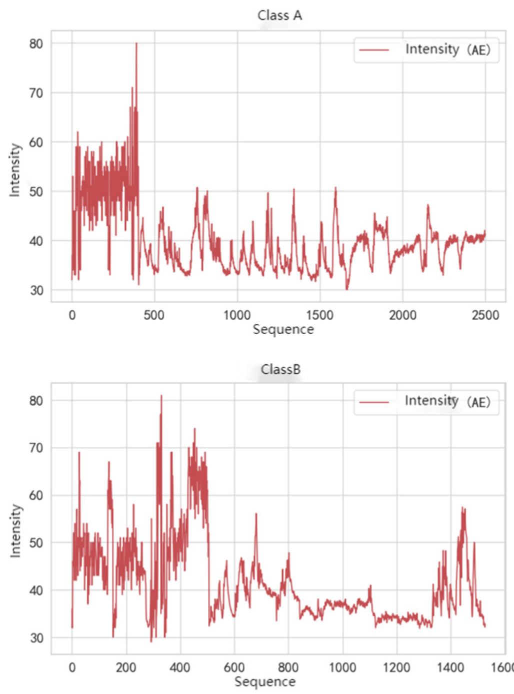

Fig 1. AE intensity map of Class A and Class B data

In this paper, the data are preprocessed and analyzed first. We preprocessed electromagnetic radiation intensity, acoustic wave intensity and type as indicators, distinguished various data by clustering method, and analyzed the data after data visualization. In the visualization of data, we show the distribution and trend of data by drawing to help understand the characteristics of data, such as the upward trend of precursor feature data. Fig. 1 are AE intensity graphs of data A and B, and Fig. 2 are EMR intensity graphs of two types of data.

2.2. Precursor Feature Signals

Within about 7 days before the earth burst, the electromagnetic radiation and acoustic emission signals have a tendency to increase with time cycle, which is called precursor characteristic signals. Rock burst may occur within about 7 days after the occurrence of precursor characteristic signals, so in general, after the occurrence of precursor characteristic signals, certain measures will be taken to prevent rock burst as much as possible.

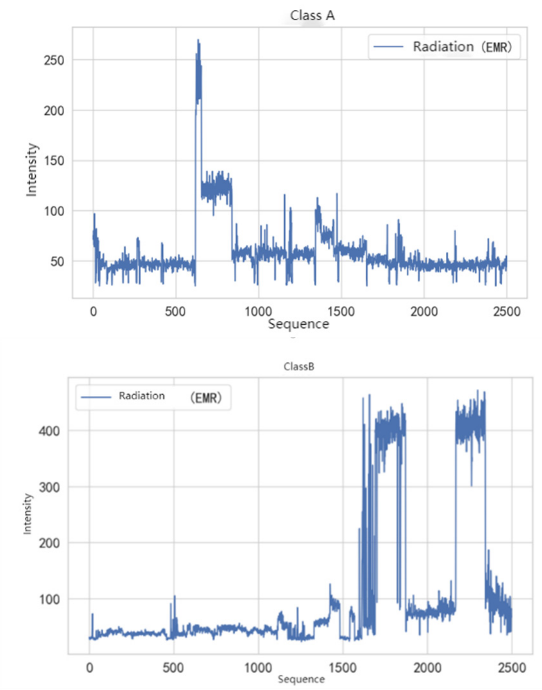

Fig 2. EMR intensity map of Class A and Class B data

This paper first analyzed the data, visualized the normal signal and the precursor signal, and found that the precursor characteristic signal strength has a slight upward trend, which is the same as the definition of the precursor characteristic signal. Fig. 3 and Fig. 4 respectively show the comparison results between the normal acoustic emission signal and the precursor characteristic signal strength and the normal electromagnetic radiation signal and the precursor characteristic signal strength.

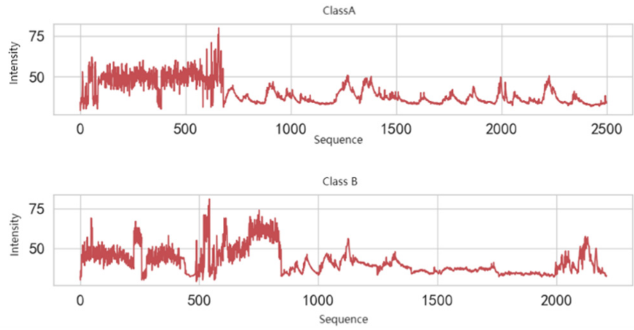

Fig 3. Comparison of the intensity of acoustic emission normal signal and precursor characteristic signal

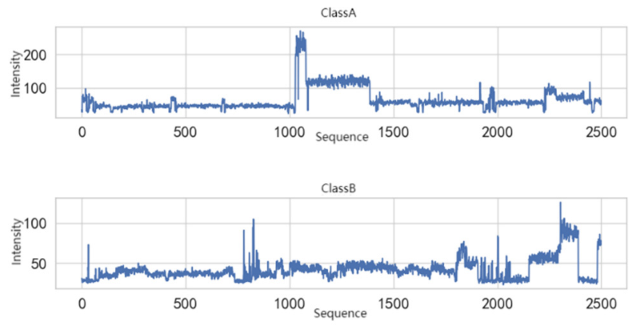

Fig 4. Comparison of the intensity of normal and precursory characteristic signals of electromagnetic radiation

3. Model Construction

In the analysis of precursor data, this paper uses KS test, MK trend test and ADF stationary series test to test all the data in Annex 2, distinguish precursor data from normal data, calculate and compare, and get their characteristics. Then, the features are substituted into the random forest model for machine learning and interval calculation.

3.1. KS Test

Table 1. Comparison of normal and precursory signals of electromagnetic radiation

	Normal EMR	EMR Precursor	Normal AE	AE Precursor
Quantity	73418	12532	15587	2212
Mean Value	49.81542148	71.05480777	37.4329889	41.66865778
Standard Deviation	18.27932445	91.89984446	3.72736829	8.303728496
Minimum Value	9.61	11.67	29	29
25% Quantile	42.451	30.14875	35.29	35.41
50% Quantile	46.4895	41	36.661	39.105
75% Quantile	52	71.674	38.2655	45.16375
Maximum Value	270	491	80	81

The KS test is a test to see whether a distribution f(x) is consistent with the theoretical distribution g(x), or whether two observed distributions are significantly different. It can be seen from the title that the precursory characteristic signal is the signal of electromagnetic radiation and acoustic emission that has a tendency to increase with time cycle, and this upward trend signal can be screened by KS test. Therefore, in this paper, KS test is carried out on all the data, the precursor feature data is selected, and a series of feature values are obtained by comparing it with the normal data. The data comparison results of normal electromagnetic radiation signals and precursor signals are shown in the table 1. Among them, the more obvious is that the mean value and standard difference are large.

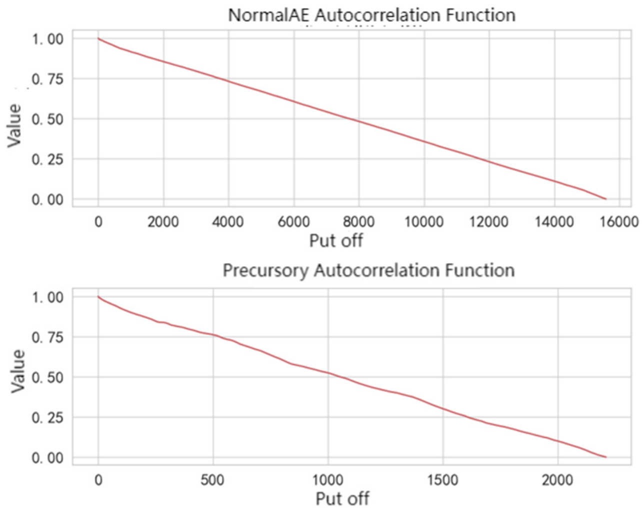

Fig 5. Autocorrelation function diagram of acoustic emission normal signal and precursor characteristic signal

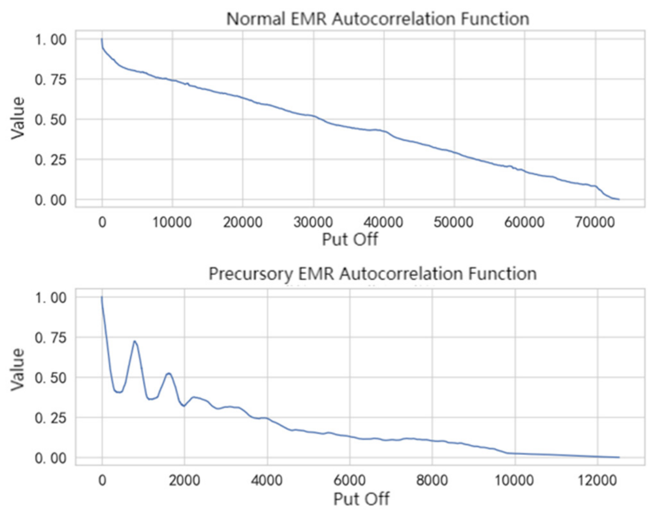

Fig 6. Autocorrelation function diagram of normal signal and precursory characteristic signal of electromagnetic radiation

Then, we visualized the autocorrelation function of the normal signal and precursor signal distinguished after KS test, and the results were shown in Fig. 5 and Fig. 6. We find that the precursory feature signal is like the normal signal, which is not conducive to the further prediction as a feature. Therefore, this paper further tests the data.

3.2. MK Trend Test and ADF Stationary Series Test

The machine learning of random forest needs more significant features to predict the interval of the next precursor signal more accurately, so this paper selects MK trend test and ADF stationary sequence test to obtain features. The MK test is a nonparametric trend test used to analyze monotonic trends in a data set. It is based on symbolic comparison of data pairs.

Set a time series $t_1, t_2 \ldots\ldots t_n$, for any (i,j) pair where i<j, defines as follows.

$$s_{ij} = \begin{cases} 1 & \text{if } t_j > t_i \\ 0 & \text{if } t_j = t_i \\ -1 & \text{if } t_j < t_i \end{cases} \tag{1}$$

The test statistic S is the sum of all s_{ij} pairs.

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} s_{ij} \tag{2}$$

In the absence of equivalent values, the standard normal variable Z is used to test.

$$Z = \begin{cases} \dfrac{S-1}{\sqrt{Var(S)}} & \text{if } S > 0 \\ 0 & \text{if } S = 0 \\ \dfrac{S+1}{\sqrt{Var(S)}} & \text{if } S > 0 \end{cases} \tag{3}$$

The calculation formula of Var(S) is as follows.

$$Var(S) = \frac{n(n-1)(2n+5)}{18} \tag{4}$$

The ADF test is used to detect the unit root in the time series to determine whether the series is non-stationary. The basic model for testing is an autoregressive model (AR model).

$$\Delta t_i = \alpha + \beta i + \gamma t_{i-1} + \delta_1 \Delta t_{i-1} + \delta_2 \Delta t_{i-2} + \ldots + \delta_p \Delta t_{i-p} + \epsilon_i \tag{5}$$

Where t_i is the time series and $\Delta t_i = t_i - t_{i-1}$ is the first difference of the series. α is the constant term and βi is the time trend term. γt_{i-1} tests the coefficient of the unit root, if γ=0 then there is a unit root, indicating that the sequence is non-stationary. $\delta_1, \delta_2, \ldots, \delta_p$ is the coefficient of the difference term, and ϵ_i is the error term.

The article creates a function and defines a constant time interval of 2 minutes for the time series of our data to perform ADF (Augmented Dickey-Fuller) and MK (Mann-Kendall) tests. After the tests, corresponding P-values are obtained, which are used to measure the probability of observing the data or more extreme data under the null hypothesis being true. The P-values are recorded, and their dynamic changes are analyzed. It is found that when precursor features of the data appear, the P-values exhibit a specific trend. This trend is similar to the described trend in the question, which increases cyclically over time. Therefore, the article also defines this trend in P-values as a characteristic.

3.3. Random Forest

Through KS test, MK trend test and ADF stationary series test, this paper obtained many characteristics of precursor signals, including the eigenvalues in the table obtained in the above KS test, including large mean and standard deviation, as well as the P-value trend characteristics of ADF test and MK test. Then, by means of random forest, this paper trains the model with the above features to predict the occurrence time of precursor signals. Finally, the precursory characteristic signal time interval of electromagnetic radiation and the precursory characteristic signal interval of acoustic emission are obtained by the model.

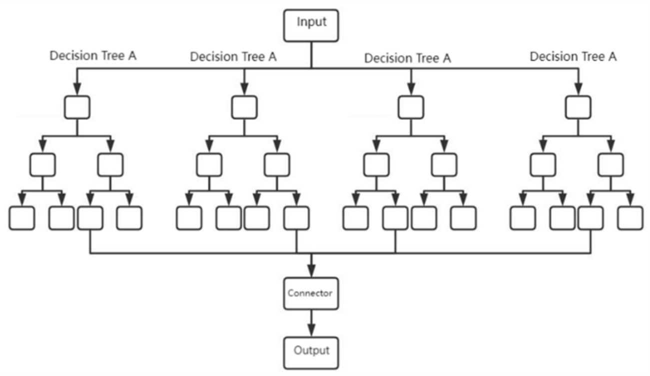

Fig 7. Random forest specific flow chart

Random forest is the combination of multiple decision trees, each data set is randomly selected, and some features are randomly selected as input. The specific flow chart of random forest is shown in Fig. 7. In this paper, the data is divided into training set and test set by means of random forest, and the feature scaling and feature engineering processing are carried out to train the model and get the prediction result. The electromagnetic radiation interference signal time interval and acoustic emission interference signal interval obtained by this model are shown in Table 2 and Table 3.

Table 2. Time interval of electromagnetic radiation interference signal

Number	Start of Time Interval	End of Time Interval
1	2022/05/01 00:00:00	2022/05/14 15:10:30
2	2022/05/14 15:16:15	2022/05/14 15:58:22
3	2022/05/14 16:18:48	2022/05/14 17:39:34
4	2022/05/14 17:41:23	2022/05/14 18:15:13
5	2022/05/14 18:18:14	2022/05/14 19:10:02

Table 3. Time interval of acoustic emission interference signal

Number	Start of Time Interval	End of Time Interval
1	2022/10/10 00:00:00	2022/10/16 01:32:21
2	2022/10/16 01:32:21	2022/10/16 01:32:21
3	2022/10/25 14:30:56	2022/10/31 03:22:59
4	2022/10/31 03:22:59	2022/11/09 09:58:22
5	2022/10/31 04:53:48	2022/10/31 05:07:21

4. Probabilistic Prediction

We use the model to predict the probability of precursor feature data after some data. Since we do not analyze all the data, but the data before the precursor signal occurs, we no longer use the method of setting threshold and KS test, but directly use the MK trend test and ADF stationary series test to obtain the characteristics of their data. Then, the feature is inserted into the perceptron model of sigmoid activation function, and the possibility of precursor data after output a certain data.

The Sigmoid activation function is an activation function commonly used in neural networks to map the input of a neuron to the output between 0 and 1. This property makes the sigmoid function particularly suitable for interpreting the output as probability, especially in binary classification problems. Therefore, for problem 3, this paper selects a perceptron model using sigmoid activation function, and calculates the occurrence probability of precursor feature signals at the last moment of a certain time period by analyzing the characteristics of the data.

Specifically, the output of a perceptron model using the sigmoid activation function can be expressed as follows.

$$f(x) = \frac{1}{1 + e^{-(wx+b)}} \tag{6}$$

Where x is the input, w is the weight, and b is the bias term. The output value of this function is between 0 and 1, which can be used to represent the probability of an event, that is, the probability of the occurrence of the precursor characteristic signal requested by the problem.

With the MK trend test and ADF stationary series test in 3.2 above, the data is first tested to obtain its corresponding P-value, and its dynamic situation is analyzed to obtain the unique trend of P-value. After replacing it into the perceptron model using sigmoid activation function, the corresponding probability can be obtained through calculation, as shown in Table 4.

Table 4. The probability of precursor features occurring at the time of data collection

Time of electromagnetic radiation data	Probability	The time of the acoustic emission data	Probability
2023-1-24 23:58:36	0.001195%	2023-1-24 23:58:36	0.07968%
2023-2-11 23:59:20	79.90%	2023-2-11 23:59:20	0.6814%
2023-2-26 23:59:27	99.42%	2023-2-26 23:59:27	40.21%
2023-3-10 23:58:14	9.950%	2023-3-10 23:58:14	2.205%
2023-3-30 23:58:13	99.85%	2023-3-30 23:58:13	81.99%

5. Conclusion

This paper effectively identifies and predicts the precursory feature signals of coal mine rock bursts by constructing a model based on the Random Forest and MK trend test. Through in-depth analysis of electromagnetic radiation and acoustic wave intensity data, we have revealed specific patterns of these signals prior to the occurrence of rock bursts. By using clustering methods for data preprocessing and visualization, we successfully distinguished between normal and precursory signals, thereby establishing a predictive model.

The Random Forest model, known for its robust classification capabilities, was employed in this study, while the MK trend test provided a solid statistical foundation for identifying trend changes in the signals. The combined application of these two methods not only improved the accuracy of precursory signal identification but also predicted the time intervals in which these signals are likely to occur, providing valuable early warning information for the safety of coal mine production.

Furthermore, the model proposed in this study demonstrated high accuracy and good generalization ability in empirical analysis, verifying the effectiveness and practicality of the model. However, the complexity of the coal mine environment suggests that there is still room for improvement. Future research can further explore the applicability of the model under different mining conditions and consider incorporating more types of sensor data to enhance the accuracy and reliability of predictions.

References

[1] Fu-xing J, Xian-feng S H I, Cun-wen W, et al. Mechanical mechanism of rock burst accidents in slice mining face under high pressure[J]. Chinese Journal of Geotechnical Engineering, 2015, 37(6): 1123-1131.

[2] Peixoto A, Ribeiro e Sousa L, Leal e Sousa R, et al. Prediction of rockburst based on an accident database[C]//ISRM Congress. ISRM, 2011: ISRM-12CONGRESS-2011-224.

[3] Chen Y, Zhang Z, Cao C, et al. Research on the causal mechanism of a rock burst accident in a longwall roadway and its prevention measures[J]. Scientific reports, 2023, 13(1): 22312.

[4] Varotsos C A, Tzanis C, Cracknell A P. Precursory signals of the major El Niño Southern Oscillation events[J]. Theoretical and Applied Climatology, 2016, 124: 903-912.

[5] Sánchez-Pastor P, Obermann A, Schimmel M. Detecting and locating precursory signals during the 2011 El Hierro, Canary Islands, submarine eruption[J]. Geophysical Research Letters, 2018, 45(19): 10,288-10,297.

[6] Rigatti S J. Random forest[J]. Journal of Insurance Medicine, 2017, 47(1): 31-39.

[7] Sang Y F, Wang Z, Liu C. Comparison of the MK test and EMD method for trend identification in hydrological time series[J]. Journal of Hydrology, 2014, 510: 293-298.