World Journal of Engineering Research
ISSN: 2959-9873
DOI: https://doi.org/10.61784/wjer3024

OPTIMIZATION OF COAL MINE ROCKBURST EARLY WARNING SYSTEM

JiaQi Wu[1*], YunMin Tian[2], TianLe Xiong[1], JunYao Hou[3], YunFeng Luo[3], Hao Chen[3]
[1]*Reading Academy, Nanjing University of Information Science and Technology, Nanjing 210044, Jiangsu, China.*
[2]*School of Atmosphere Sciences, Nanjing University of Information Science and Technology, Nanjing 210044, Jiangsu, China.*
[3]*Waterford Institute, Nanjing University of Information Science and Technology, Nanjing 210044, Jaingsu, China.*
Corresponding Author: Jiaqi Wu, Email: 202283100009@nuist.edu.cn

Abstract: As the main energy and important industrial raw materials, coal plays a vital role. With the deep development of coal mining, the risk of underground coal and rock dynamic disasters is rising, which seriously threatens the safety of coal mining. In this paper, the interference signals and precursory characteristic signals in acoustic emission (AE) and electromagnetic radiation (EMR) signals are analyzed. A multi classification model based on the fine KNN model is established to classify the jamming signal data in three different intervals. ARIMA model is used to summarize and analyze the trend characteristics of precursory characteristic signals. The method of random forest classification model is used to classify and identify the time interval of the precursor signal. And calculate the probability of precursory characteristic data at a specific time.
Keywords: ARIMA model; Refined k-nearest neighbor algorithm; Random forest classification model; Non-linear classification

1 INTRODUCTION

In the process of coal mine production, monitoring and early warning of rock burst and effective prevention and control are still scientific and technological problems to be solved. By monitoring the change trend of acoustic emission (AE) and electromagnetic radiation (EMR) signals, we can determine whether there is a risk of rock burst in the working face or roadway. By dividing the electromagnetic radiation and acoustic emission data into different categories, such as normal working data, precursory characteristic data, interference signal data, sensor disconnection data and working face rest data, the potential rockburst risk can be better identified. Therefore, the analysis and early warning of these monitoring data is of great significance to reduce the occurrence of coal mine accidents[1].

Dou et al.[2] advanced the theoretical understanding of rockburst mechanisms by analyzing the interaction between dynamic (seismic) and static (tectonic) loads. Their research revealed that high static loads in deep mining exacerbate rockburst risk, with microseismic increments from mining-induced tremors acting as critical precursors. Wang[3] introduced a locally weighted C4.5 decision tree algorithm for rockburst risk prediction, achieving 100% accuracy on testing datasets from the Yanshitai coal mine in China. By discretizing continuous attributes via the minimum description length principle and applying 10-fold cross-validation, this method outperformed traditional C4.5 models, which yielded only 71.43% accuracy. Qi Hegang[4] integrated numerical modeling (FLAC3D) with reinforcement learning to simulate stress redistribution during mining, enabling dynamic adjustment of warning thresholds. This approach, validated in the Datong coalfield, reduced false alarms by 30% compared to static threshold systems. Concurrently, human-machine interfaces (HMIs) are evolving to incorporate augmented reality (AR) overlays, providing miners with real-time hazard maps and evacuation routes—a feature tested in the Austar mine post-2014 rockburst reforms[5].

After data preprocessing, outliers are eliminated and the missing values are filled by k-nearest neighbor algorithm. Then the electromagnetic radiation (EMR) and acoustic emission (AE) signals with interference signals are analyzed in three different dimensions through the data in the data table: the external characteristics, internal characteristics and time characteristics of the interference signal distribution. Firstly, according to the external characteristics, the average value, variance, median and extreme value of the transmitted signal are obtained to distinguish the numerical characteristics of the interference signal and other signals. Secondly, according to the time characteristics, the main time period of interference signal is obtained through investigation and analysis. Finally, by drawing the time series distribution map and establishing the nonlinear image analysis model, the size changes of electromagnetic radiation and acoustic emission signals and the corresponding signal type changes in different time periods are analyzed, and the method of identifying the internal characteristics of interference signals and the size changes of electromagnetic radiation and acoustic emission signals is further optimized. This paper analyzes the time series of the corresponding precursor signal sequence, and establishes an appropriate ARIMA model through white noise test and ADF test, so as to summarize and analyze the remaining trend characteristics of the precursor signal over time.

2 PRELIMINARY

2.1 KNN Algorithm

KNN algorithm is a commonly used machine learning algorithm. Its core idea is based on the nearest neighbor principle. It can classify or regression predict by finding K training data points nearest to the test sample [6]. At the same time, when dealing with the problem of interference signal recognition, it is very important to choose the appropriate distance measurement method for the accuracy of the algorithm. Euclidean distance can effectively evaluate the similarity between samples, so as to achieve accurate signal classification and recognition. Therefore, when dealing with such problems based on high-precision KNN algorithm, Euclidean distance is selected as the distance measurement method. The K-Nearest Neighbors (KNN) algorithm is a supervised learning method used for both classification and regression tasks. It operates on the principle of feature similarity, where the prediction for a new data point is based on the labels or values of its K closest neighbors in the training dataset. The algorithm calculates distances (commonly Euclidean, Manhattan, or Hamming) between the new data point and all training examples, selects the K nearest ones, and determines the prediction through majority voting (for classification) or averaging (for regression). KNN is appreciated for its simplicity and intuitive approach, making it suitable for various applications such as credit rating evaluations, political election forecasting, and pattern recognition. However, its performance can be sensitive to the choice of K and the scale of the data.

2.2 ARIMA Model

ARIMA (*P, D, q*) model, fully known as autoregressive integrated moving average model, is a statistical model used to analyze and predict time series data. ARIMA model changes the time series data into a stationary series, and then uses the autoregressive (AR) and moving average (MA) parts of the series to fit and predict the model. It is suitable for non seasonal time series data with trend or seasonality [7]. ARIMA model is composed of three main parts: autoregressive term (AR), difference item (*I*), and moving average term (MA), which are represented by three parameters: *P, D*, and *Q*. The general form of the model is ARIMA (*P, D, q*). The Autoregressive Integrated Moving Average (ARIMA) model is a statistical tool for time series analysis and forecasting. It integrates three components: autoregression (AR), which uses past observations to predict future values; differencing (I), which transforms non-stationary time series into stationary ones by subtracting previous values; and moving average (MA), which incorporates past forecast errors into the prediction. The model is denoted as ARIMA(p, d, q), where p is the order of the autoregressive component, d is the degree of differencing, and q is the order of the moving average component. ARIMA is particularly effective for time series data with trends or seasonality and is extensively used in economics, finance, and inventory management to forecast future values based on historical patterns.

2.3 Random Forest

Random forest is an algorithm based on the idea of ensemble learning. It builds bagging ensemble based on decision tree, further introduces random attribute selection in its training process, and finally makes the decision trees of random forest independent of each other. By inputting new samples, each decision tree of the forest can be judged and classified separately to obtain their own classification results, and finally vote to determine the final random forest classification results [8]. In this process, the feature importance value can be retained. The Random Forest algorithm is an ensemble learning method that combines multiple decision trees to enhance predictive accuracy and model robustness. By utilizing bootstrap aggregation (bagging), it generates diverse training datasets through random sampling with replacement, and each decision tree is trained on a subset of these data. During the tree-building process, a random selection of features is employed at each node split, further increasing the diversity among trees. For classification tasks, the final prediction is determined by majority voting across all trees, while regression tasks use the average of all tree predictions. This approach effectively reduces overfitting and improves generalization performance. Random Forest is widely applied in various domains, including credit risk assessment, medical diagnosis, and recommendation systems, due to its ability to handle high-dimensional data and its resistance to overfitting.

2.4 Notations

The symbols used in the paper are listed in Table 1.

Table 1 Notations

Symbols	Notations
P	Forecast data
X	Original dataset data as opposed to forecast data
Gini	Purity measurement
pri	Probability of correct classification of the ith node
imp	Feature importance function
I	Decision tree set established during algorithm execution

3 ELECTROMAGNETIC RADIATION AND ACOUSTIC EMISSION SIGNALS

If the amount of data is large and the proportion of outliers is small, we can consider deleting outliers to improve the stability and accuracy of the model. And the characteristics of outliers are very obvious and easy to identify: in d/e type data, if the data fluctuates greatly between several sample points (this paper sets a reasonable adjacent fluctuation range of 45%-150%), it is set as an outlier and the outliers are processed[9]. In order to get the complete data set, this paper uses the k-nearest neighbors algorithm [10] to fill the gap value. According to the similarity between samples, the k-nearest neighbor method uses the eigenvalues of the nearest K samples to fill in the missing values. For the values containing the vacancy due to deletion, the data in Annex I is huge and extremely dense, which perfectly meets the numerical characteristic requirements of the k-nearest neighbor algorithm for adjacent data. In this paper, the external characteristics of the electromagnetic radiation and acoustic intensity in the above problems are tested respectively. It is found that the standard deviation and mean value of class C data are significantly different from other data in the electromagnetic radiation and acoustic intensity, so it can be used as the feature selection standard of class C data. By analyzing the approximate time of C-type data (interference signal), we can get its time characteristics: the interference signal distribution caused by electromagnetic radiation and sound wave intensity is concentrated in January to July every year, with certain periodic characteristics. By observing and analyzing the time series distribution map, this paper found that the electromagnetic radiation and acoustic intensity of class C data reached the peak at almost the same time, indicating that there is a strong internal relationship between electromagnetic radiation and acoustic intensity in class C data. Electromagnetic radiation and sound wave intensity will fluctuate greatly in the presence of interference signals (Class C signals). Therefore, this paper describes the internal characteristics of interference signals: EMR and AE produce violent oscillation.

3.1 Classification Forecast

This paper finds that the overall data (a, B, C, D/E) has the following characteristics:
1.Data features with relatively obvious separation boundaries.Data features with certain local properties, that is, similar samples will gather together in the feature space.
2.Almost perfectly meet the requirements of using KNN algorithm model in this paper.
According to the proportion of interference signal finally counted, Table 2 and 3 obtained the interval range.

Table 2 Time Interval of Electromagnetic Radiation Interference Signal

S/N time	interval start	interval end
1	2022-5-1 0:01:12	2022-5-1 13:53:24
2	2022-5-1 23:58:53	2022-5-2 16:17:30
3	2022-5-2 18:31:00	2022-5-3 6:29:41
4	2022-5-3 20:25:32	2022-5-4 7:05:44
5	2022-5-4 21:27:23	2022-5-5 6:25:07

Table 3 Time Interval of Acoustic Emission Interference Signal

S/N time	interval start	interval end
1	2022-4-1 0:00:11	2022-4-1 10:20:18
2	2022-4-1 11:38:56	2022-4-2 8:24:23
3	2022-4-9 3:47:37	2022-4-9 21:06:36
4	2022-4-10 1:55:35	2022-4-10 9:05:24
5	2022-4-11 1:56:47	2022-4-11 9:12:02

3.2 ARIMA Forecast

By observing and analyzing the time series diagram, it can be found that when the precursor characteristic signal appears, the electromagnetic radiation signal intensity will gradually increase or intermittently increase, and when the rock burst occurs, the electromagnetic radiation signal intensity reaches the highest value, and then decreases sharply in a short time. The intensity of acoustic emission signal will gradually increase, and when rock burst occurs, the intensity of acoustic emission signal will reach the highest value, and then in a short time, the intensity of acoustic emission signal will sharply reduce [3]. The overall trend is characterized by cyclic increase. The time series diagram of the occurrence of the overall precursory characteristics shows a periodic repeating pattern, which indicates that there is a periodic trend when the precursory characteristic signals appear; In the same period of time, the precursory characteristic signal has an obvious growth or decline trend, reflecting a certain trend of violent fluctuations. According to these characteristics, this paper intends to establish ARIMA model to analyze the trend characteristics of precursory characteristic signals.

According to the precursory characteristic signal data of EMR, this paper uses t-value test and finds that the p value of the test statistic is 0.0000, which is less than the significance level of 0.01. Therefore, the original hypothesis is rejected and the alternative hypothesis is accepted, indicating that it is a non-stationary sequence with fluctuation.

According to the precursory characteristic signal data of AE, this paper also uses t-value test, and finds that the p value of the test statistic is also 0.0000, which is less than the significance level of 0.01. Therefore, the original hypothesis is rejected and the alternative hypothesis is accepted, indicating that it is a non-stationary sequence with fluctuation. The maximum lag point of ACF autocorrelation function graph is used to roughly judge the Q value. The p value is determined by the maximum lag point of PACF partial autocorrelation function graph. However, the correctness of the parameters obtained in this way is low. In order to ensure the correctness of the parameters, this paper next needs to carry out model estimation to obtain the values of P and Q.

By comparing the BIC values under different differential orders, the parameter value that can minimize the BIC is selected. In this comparison, it is found that when the autoregressive term P=0, the order of the moving average term q=4, that is, the BIC value reaches the minimum. Therefore, this paper chooses to establish ARIMA (0,1,4) model.

After the model is established, the residual is tested by white noise. If the residual is white noise, it indicates that the selected model can fully identify the law of time series data, that is, the model is acceptable; If the residual is not white noise, it means that the sequence may have a certain pattern, structure or correlation, and does not have pure randomness, which may be useful for data analysis and prediction. This means that other types of information and associations in the data can be explored. From the results, the p value of the Ljung box test of the two groups of data is less than 0.01, which means that there is a significant autocorrelation in the residuals, rejecting the original assumption that the residuals are white noise. This shows that the model can still be further optimized.

4 RANDOM FORESTS

This paper found that the prediction probability of the precursor characteristic signal reached more than 80%, which met the extraction standard of the time series of subsequent precursor characteristic signals, and then obtained the time interval of electromagnetic radiation precursor characteristics and acoustic emission precursor characteristics, as shown in Table 4 and 5.

Table 4 Characteristic Time Interval of Electromagnetic Radiation Precursor for RF

S/N time	interval start	interval end
1	2020-4-8 2:23:05	2020-4-11 10:06:07
2	2020-4-22 21:41:27	2020-4-27 12:33:47
3	2020-5-23 10:03:33	2020-6-5 5:21:55
4	2021-12-15 3:47:11	2021-12-20 23:59:11
5	2021-11-24 5:47:11	2021-11-30 17:04:02

Table 5 Characteristic Time Interval of Acoustic Emission Precursors for RF

S/N time	interval start	interval end
1	2021-11-1 0:01:01	2021-11-2 17:00:13
2	2021-11-25 20:59:12	2021-11-30 8:25:06
3	2021-12-3 10:10:06	2021-12-9 19:14:11
4	2021-12-12 6:21:47	2021-12-16 17:02:55
5	2022-1-1 5:59:07	2022-1-14 7:48:56

Based on the analysis and detection of ARIMA (0,1,4) model and random forest model, this paper found that ARIMA model had excellent fitting effect (the goodness of fit r was as high as 0.96), but because the original hypothesis was rejected in ADF test, that is, the residual did not meet the white noise sequence, this model could not predict and estimate the target sequence well. Therefore, this paper uses the random forest classification model to predict and classify the target sequence and get the corresponding classification data results. Because through the analysis of ARIMA model and random forest model, this paper confirms that the characteristics of precursor characteristic signals have a certain persistence, that is, they are aggregated and distributed in a certain period of time series and the time is about 7 days (about 1000 samples).

Firstly, by using the frequency estimation method to calculate the proportion of the number of precursor characteristic signals in the first 1000 sample points of the target time point to the total 1000 samples, the corresponding probability value is obtained. Next, according to these probability values, the occurrence probability of precursory characteristic data in the target time period is evaluated. This probability value can be regarded as the probability of precursory characteristic data at the last moment of each time period. The comparison is shown in Table 6.

Table 6 Probability of Occurrence of Precursory Characteristics at the Time of Data Collection

Time of electromagnetic radiation data	Time of probabilistic acoustic emission	Data	Probability
2023-1-24 23:58:36	0.07715	2023-1-24 23:58:36	0.05299
2023-2-11 23:59:20	0.57242	2023-2-11 23:59:20	0.51245
2023-2-26 23:59:27	0.51605	2023-2-26 23:59:27	0.48765
2023-3-10 23:58:14	0.55637	2023-3-10 23:58:14	0.55601
2023-3-30 23:58:13	0.51187	2023-3-30 23:58:13	0.54237

In the construction of the model, the model absorbs and processes a large amount of data, which has strong stability. The model is suitable for the prediction of rockburst indexes in the future, and integrates the advantages of various classification models. It has been tested for many times and found that its fitting and prediction effect is good and has strong universality; For example, the ARIMA (0,1,4) model introduced in this paper has a high goodness of fit for precursor signals, and can achieve high accuracy and good prediction effect. The ADF detection of ARIMA model found that it was not white noise, but failed to show the structure and characteristics of its residual sequence.

5 CONCLUSION

This paper presents a comprehensive study on the optimization of coal mine rockburst early warning systems, focusing on the analysis and processing of electromagnetic radiation (EMR) and acoustic emission (AE) signals. The research primarily addresses three key objectives: firstly, to identify and classify interference signals within EMR and AE data; secondly, to develop mathematical models for precisely locating precursor characteristic signals and determining significant trend features; and thirdly, to establish a probabilistic model for predicting the occurrence of precursor signals at specific time intervals. The paper underscores the effectiveness of the proposed methodologies. The integration of advanced data preprocessing and KNN modeling demonstrates proficiency in interference signal identification and classification.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Song C H, Lu C P, Liu J R. Moment Tensor and Stress Field Inversions of Mining-Induced Seismicity in A Thick-Hard Roof Zone. Rock Mechanics and Rock Engineering, 2024, 57(3): 2267-2287.
[2] Watson J, Canbulat I, Zhang C, et al. Energies Within Rock Mass and the Associated Dynamic Rock Failures. Rock Mechanics and Rock Engineering, 2025, 58(5): 4935-4958.
[3] Wojtecki U, Krawiec K, Wikaa M, et al. An attempt to determine the cause of the strong tremor responsible for a rockburst in a hard coal mine based on numerical modeling and spectral parameters. Geology, Geophysics & Environment, 2024, 50(4).
[4] Wang K, Xie T, Mei L I, et al. A surrogate model for the rapid prediction of rockburst risk based on numerical samples and random forest classifier. Journal of Tsinghua University (Science and Technology), 2024, 64(7): 1203 -1214.
[5] Kuzniar K, Tatara T, Zajac M. Experimental and numerical assessment of soil-structure interaction effects in the case of mine-induced vibrations. IOP Publishing Ltd, 2024.
[6] Eremin M O, Chirkov A O, Pazhin A, et al. Finite-Difference Analysis of Influence of Borehole Diameter and Spacing on Reduction in Rockburst Potential of Burst-Prone Coal Seams. Mining, 2024, 4(4).
[7] Jiahao S, Wenjie W, Lianku X. Predicting Short-Term Rockburst Using RF–CRITIC and Improved Cloud Model. Natural resources research, 2024, 33(1): 471-494.
[8] Dai J, Gong F, Huang D, et al. Quantitative evaluation method of rockburst prevention effect for anchoring rock masses around deep-buried tunnels. Tunnelling and Underground Space Technology incorporating Trenchless Technology Research, 2025, 156.
[9] Wojtecki U, Iwaszenko S, Apel D B, et al. Use of machine learning algorithms to assess the state of rockburst hazard in underground coal mine openings. 2022.

[10] Wang Q, Ma T, Yang S, et al. Intelligent rockburst level prediction model based on swarm intelligence optimization and multi-strategy learner soft voting hybrid ensemble. Geomechanics & Geophysics for Geo-Energy & Geo-Resources, 2025, 11(1).