Early warning and detection of rock burst signal in coal mine

*Tong Fu, Dongping Sheng**

Changzhou Institute of Technology, Changzhou, China

ABSTRACT

With the increase of mining depth, formation pressure gradually increases, and the risk of underground coal and rock dynamic disaster also increases, which seriously affects the safe and efficient mining of coal mines. Rock burst has become one of the most serious disasters threatening the safety production of coal mines in China. The purpose of this study is to analyze the characteristics of electromagnetic radiation and acoustic emission signals to judge the occurrence time of interference signals and precursor signals and predict the probability of precursor signals. This research is of great significance to improve the accuracy and timeliness of the mine safety warning system, which is helpful to reduce the occurrence of rock burst accidents and related losses. The data pre-processing and visual analysis of the interference signal were carried out to preliminatively observe the characteristics of the data, and then the four time-domain features of the data, namely, mean value, variance, root-mean-square and skewness, were extracted. Then the frequency features of the spectral energy and main spectrum components of the EMR (electromagnetic radiation) and AE (acoustic emission) interference signal data were extracted using the fast Fourier transform algorithm. The data is preprocessed, the outliers are processed and arranged in ascending order of time, and then the visual analysis is performed. The precursory data features are extracted by the method of extracting data features in the first work, and it is found that the precursory signals have the characteristics of excessive regional time and random changes. Then, the sliding window method is used to extract the features, and a window of 25 data points is established to calculate the mean value, variance and other features. Since the original data presents strong randomness and noise, the wavelet transform is first used to reduce the noise of the data, so as to make the change of the data smoother, and then the data after noise reduction is analyzed. Then LSTM time series prediction model is used to predict the data of the next 50 time points, and the sliding window is used to extract the features of the data. Finally, the extracted features are analyzed with decision tree model to identify potential precursor signals, and the probability of precursor features appearing at different time points is calculated.

Keywords: Fast Fourier transform, sliding window, decision tree model, wavelet transform, LSTM time series

1 INTRODUCTION

In the process of deep coal mining in China, we are facing the challenge of ground stress, especially the serious threat of rock burst accidents. By collecting electromagnetic radiation and acoustic emission sensor data, it can judge whether there is a rock burst risk in the working face or roadway, and the data can be divided into five categories: normal operation, precursor characteristics, interference signals, sensor line break and working face rest. The construction of accurate mathematical model is very important to analyze monitoring data

and identify the precursor signals of rock burst, which can improve the accuracy of early warning of rock burst events, formulate effective disaster prevention and reduction measures, protect the life safety of miners and reduce economic losses.

Work one: (1.1) A mathematical model is required to analyze the interference signal data in electromagnetic radiation and acoustic emission signals [1]. The characteristics of interference signals (at least three) in electromagnetic radiation and acoustic emission signals are described. (1.2) The model established in (1.1) is used to analyze the data of electromagnetic radiation and acoustic emission signals in a specific period of time, so as to identify the time interval where the interference signals are located and list the time interval where the first five interference signals are located.

Work two: (2.1) It is required to establish a mathematical model to analyze the precursory characteristic signals in the electromagnetic radiation and acoustic emission signals, paying particular attention to the changing trend of the signals, and give the trend characteristics (at least three) of the precursory characteristic data in the electromagnetic radiation and acoustic emission signals [2]. (2.2) A mathematical model is required to be established. According to the characteristics obtained in (2.1), electromagnetic radiation and acoustic emission signal data within a specific time period are analyzed to identify the time interval where the earliest five precursor characteristic signals are located.

Work three: A mathematical model is required to analyze the electromagnetic radiation and acoustic emission signals collected over a discontinuous period of time as provided in Annex 3. Calculate and estimate the probability of precursor characteristic signal appearing at the last moment of each data collection period.

2 RELATED WORK AND ASSUMPTION

2.1 Work one

Work one consists of two works, the first of which analyzes the electromagnetic radiation and acoustic emission signals of interference and gives the characteristics of the interference signals in the electromagnetic radiation and acoustic emission signals. According to the first work, the mean, variance and kurtosis of time domain feature and the spectral energy and main spectrum components of frequency domain feature are selected [3]. For the second work, based on the above 5 data characteristics, The electromagnetic radiation from May 1, 2022 to May 30, 2022 and the time interval of interference signals in acoustic emission signals from April 1, 2022 to May 30, 2022 and from October 10, 2022 to November 10, 2022 were identified by sliding window feature extraction and error estimation interval identification The first 5 interference signals of electromagnetic radiation and acoustic emission are given respectively.

2.2 Work two

Firstly, in order to solve the first work, the sliding window feature extraction method proposed in (1.1) can be used here. However, data need to be extracted in groups during processing here, for example, 25 data are divided into a group to extract features, so that the

change trend of precursor signals can be clearly seen. Then, it is necessary to identify the precursor signal in the given time period of the second minor work according to the characteristics determined by the first minor work. This work identifies the precursor signal in the given time period through the decision tree model, uses the labeled data for training, and compares the data to be identified as the test set to obtain the precursor signal's time period.

2.3 Work three

Work three requires early warning at the first time when the precursor signal occurs, that is, it is necessary to identify the precursor signal as soon as possible, identify the precursor signal on the acoustic wave intensity (AE) and electromagnetic radiation (EMR) data given in Annex 3, and give the probability of the precursor signal appearing in the last time period. First of all, it is necessary to visualize the three types of data in the attachment, analyze the overall changes of the data, and then consider the data pre-processing. Since there are more data given, the data with excessive changes will be denoised by wavelet transform [4]. Then, for attachment 3, it is necessary to extract the change trend feature of the line, and then identify the signal. Therefore, LSTM time series prediction model can be used to predict the original data first, and the number of predictions is 50. Then, the sliding window features are extracted from the predicted data. Finally, the extracted features are identified with the original data in Annex 3 by using the decision tree model, and the precursor signals are identified to find out the probability of precursor signals in the last time period.

2.4 Model assumption

(1) Assume that all acquired sensors have been calibrated to ensure uniform performance and accuracy, so that the collected data is comparable across different sensors;

(2) Assume that the frequency of data acquisition is high enough to capture all important changes and features of the signal to ensure that the time resolution meets the analysis requirements;

(3) It is reasonable to assume that the number of sources of interference signals is limited and has certain rules and its characteristics can be identified;

(3) For those signals that may indicate the occurrence of rock burst, it is assumed that they will show a certain periodicity before the occurrence.

3 MODEL BUILDING AND ANALYZING

3.1 Model establishing and solving for work one

3.1.1 Data preprocessing

After screening the Class C (interference signal data) data, it is found that there are multiple sets of repeated values in the data, that is, multiple signal amplitudes appear in the same time period. First, all the repeated values are found, and then the average value is used as the signal amplitude in this period of time. For example, there are two signal amplitudes of

114.18 and 114 in the time period of 8:49:42 of 2019/1/11 of EMR, and the average value is 114.09 as the signal amplitude in this time period [5].

3.1.2 Data visualization analysis

After data preprocessing, the data is drawn into a time-domain signal diagram, as shown in Fig.1 and Fig.2.

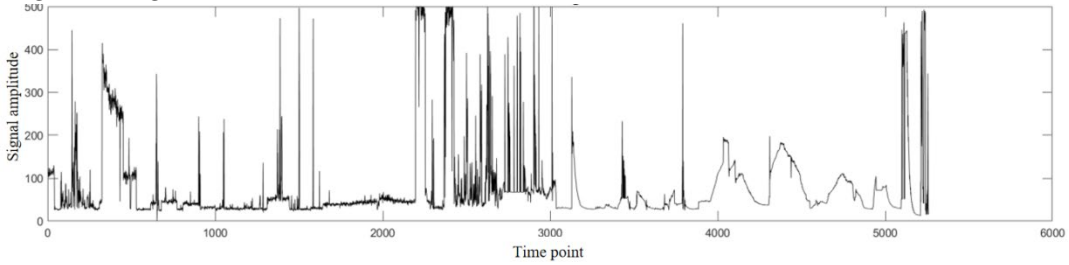


Fig.1 Visualization of EMR interference signal

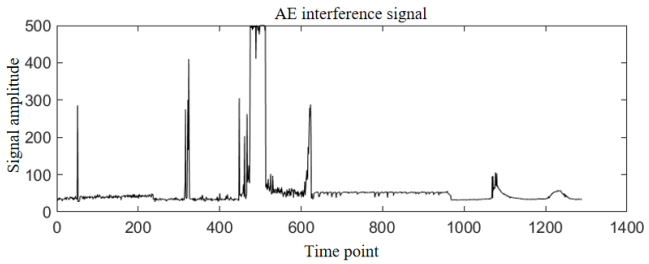


Fig.2 Visualization of AE interference signal

After rigorous analysis of the visual data of EMR and AE interference signals in Fig.1 and Fig.2, the following conclusions are drawn:

The data show that the amplitude of the interference signal fluctuates randomly in the unit range of 100 to 500, and lacks discernable regularity. This indicates that the interference signal may come from a variety of different interference sources, and the interference intensity and change mode of these interference sources are different. In addition, the frequency distribution on the spectrogram is scattered and does not show any obvious periodic structure, which further supports the above hypothesis about multi-source interference.

In the time series data, some peaks and valleys can be observed that may represent abrupt changes in the intensity of interference caused by a particular event. By analyzing the occurrence time of these peaks and valleys, we can try to locate the activity pattern of interference sources and evaluate its potential impact on the performance of the rock burst hazard warning system in deep coal mining [6].

Although there was no apparent periodicity in the data, at certain time intervals the data showed similar features or patterns. This may imply some degree of repetitive behavior on the part of the interference source, or it may be due to some periodic factor in the interference environment. To test this hypothesis, data acquisition over a longer time span and at a higher sampling rate is needed to capture more subtle periodic patterns that may be present.

3.1.3 Time domain and frequency characteristics of interference signals

In the core domain of time domain feature analysis, the mean is the first order origin moment of the data set, which provides a measure of the center position of the data set and is

a direct representation of the equilibrium of the data set. Then, as the second order origin moment of the data set, the variance measures the degree of dispersion of the values in the data set by calculating the average of the square of the difference between the data point and the mean, which is a key indicator to evaluate the stability of the data. In addition, root-mean-square, as a statistic, is the average of the sum of squares of the absolute values of all the values in the dataset, and is often used to describe the fluctuation intensity of the dataset. Finally, skewness, as a third-order origin moment describing the shape of the data distribution, measures the asymmetry of the data distribution by the mean value of the data set, thereby helping to understand the extreme value or skew phenomenon that may exist in the data set [7]. The combined effect of these four statistics makes it possible to comprehensively and deeply analyze and understand the time domain characteristics of interference signals, and then provide scientific basis for subsequent signal processing, interference detection and other tasks.

The calculation formula of the mean value of the interference signal data xi is shown in equation (1) :

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1}$$

The formula for calculating the variance of interference signal data xi is shown in equation (2) :

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \mu)^2 \tag{2}$$

The calculation formula of the root mean square of interference signal data xi is shown in equation (3) :

$$\text{RMS} = \sqrt{\frac{{x_1}^2 + {x_2}^2 + \cdots + {x_n}^2}{n}} \tag{3}$$

According to mean μ and standard deviation (standard deviation symbol), the skewness is calculated by the formula (4) as follows:

$$S = \frac{1}{n}\sum_{i=1}^{n} \left(\frac{x_i - \mu}{\sigma}\right)^3 \tag{4}$$

Table 1 Results of signal characteristics

Time domain feature	Mean value	Variance	Root-mean-square	Skewness
EMR time domain results	77.9585	8229.3701	119.6049	3.1238
AE time domain results	59.9691	6467.6403	100.2943	4.8817

When extracting the frequency domain features of interference signals, it is first recognized that the frequency domain features are the signal characteristics obtained by converting the time domain signal to the frequency domain. These characteristics describe the behavior and properties of the signal at different frequencies, including spectral energy, major spectrum components, frequency distribution, bandwidth, and amplitude and phase of the frequency components. Through Fourier transform, periodic patterns and frequency components of signals that are difficult to observe in the time domain and their relative importance can be clearly displayed in the frequency domain. Next, the spectral energy and main spectrum components of the interference signal are extracted.

Fast Fourier Transform (FFT) is an algorithm used to efficiently compute discrete Fourier

transforms (DFT). Fourier transform is a method of converting a signal from the time domain to the frequency domain, which is used to analyze the components of the signal at different frequencies. The fast Fourier transform algorithm takes advantage of the symmetry and repetition of DFT and reduces the computational complexity from exponential level to approximately linear level, thus greatly improving the computational efficiency [8]. The core idea of FFT algorithm is to decompose the calculation process of DFT into several smaller scale DFT calculations, and use the symmetry and periodicity of frequency domain to reduce the calculation amount. This algorithm can reduce the multiplicative calculation of N point DET from N2 times to n/2$\log_2 N$times. Take the FFT time extraction algorithm of 8 points as an example.

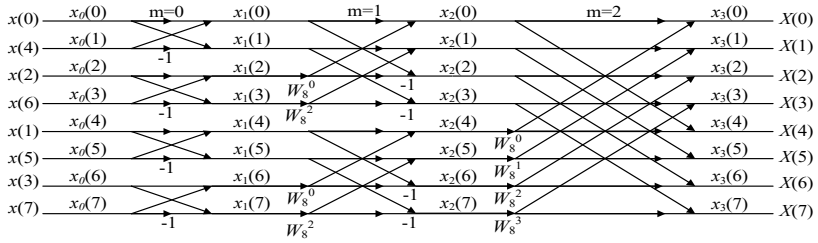


Fig.3 Signal flow diagram of FFT time extraction algorithm with sampling points

The quick step formula for the fast Fourier transform is shown below. The analog signal x(t) continuous Fourier transform (5) is shown:

$$X_{(W)} = \int_{-\infty}^{+\infty} x_{(t)} e^{-j\omega t} dt \tag{5}$$

x(t) becomes x(nT) after T sampling periods. Let x(n) be an ordered long sequence of N points. Its DET is shown in equation (6) :

$$X_{(k)} = \sum_{n=0}^{N-1} x_{(n)} . e^{-\frac{2\pi j}{N} kn} \tag{6}$$

When calculating spectral energy, power spectral density (PSD) is usually used to represent the energy distribution of a signal in the frequency domain. Perform a discrete Fourier transform (DFT): convert a time domain signal to a frequency domain representation. DFT is a common digital signal processing technology that can convert time domain signals into frequency domain signals [9]. The discrete form of DFT is shown in equation (6). Calculate power spectral density (PSD): PSD represents the energy distribution of a signal at different frequencies. For real signals, PSD can be calculated from DFT results. The calculation formula of PSD is shown in equation (7) :

$$S_{xx}[k] = \frac{1}{N} |X[k]|^2 \tag{7}$$

Then by squaring the spectrum data of each frequency point, the energy of the frequency point is obtained. Finally, the energy of all frequency points is added together to obtain the spectral energy of the entire signal. Calculate the amplitude spectrum: The degree spectrum represents the amplitude distribution of the signal in the frequency domain. It can be calculated by DFT results, and the specific calculation formula is shown in equation (8):

$$A[k] = |X[k]| \tag{8}$$

Find major spectral components: Find major peaks or harmonics in the amplitude spectrum. Use a peak detection algorithm or a simple threshold method to determine the location and magnitude of the major spectrum components.

Through the above methods, the frequency domain characteristics of the data are calculated, and the specific values are shown in Table 2:

Table 2 Feature results of EMR and AE frequency domain interference data

Frequency domain	Characteristic spectral energy	Main spectral component	Spectral peak	Spectral density
EMR frequency domain results	75231749.7892	0.00019015	409983.773	14.3053
AE frequency domain results	12965981.8308	0.00077580	77300.177	10.0589

3.1.4 Recognition of interference signals based on sliding window

In data processing and signal analysis, sliding window extraction technology is widely used because of its unique advantages. This technique achieves local feature extraction by sliding a preset size window on the data set and performing a series of feature calculations on the data in the window [10]. This method is particularly effective when dealing with time series data because it can capture local patterns and dynamic features in the data that change over time.

The window size is defined as 50, that is, data features are calculated every 50 data points, and then compared with the data features of the first small work, and the error rate is set to 0.1. The sliding window analysis of the data results in the visualization of the feature results as follows:

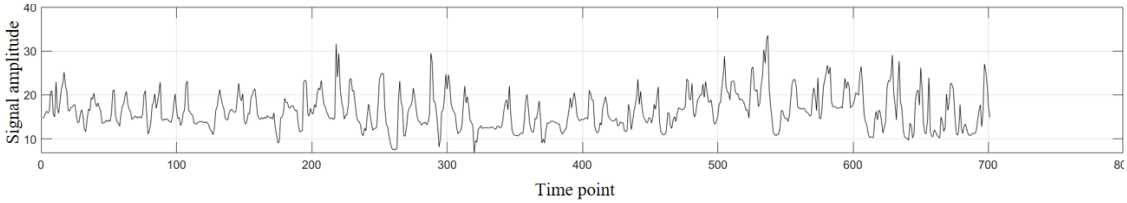


Fig.4 EMR1 mean characteristics

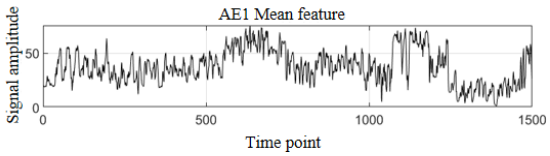


Fig.5 AE1 mean characteristics

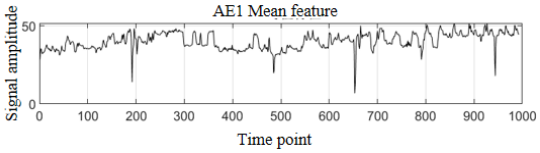


Fig.6 AE2 mean characteristics

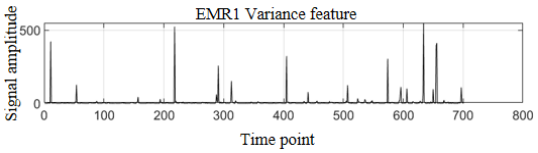


Fig.7 EMR variance characteristics

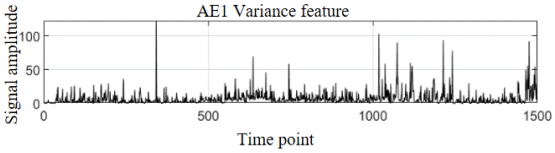


Fig.8 AE1 variance characteristics

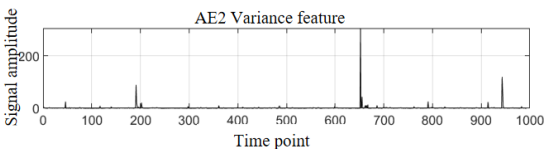


Fig.9 AE2 variance characteristics

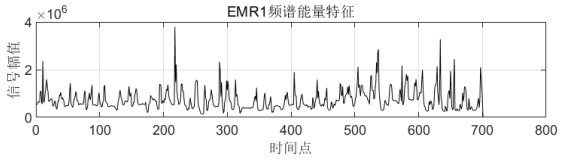
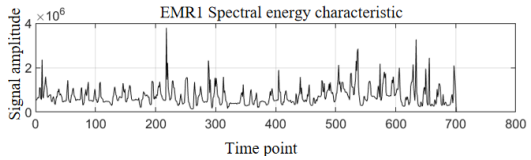


Fig.10 EMR1 spectral energy characteristics



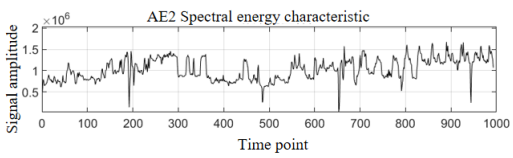


Fig.11 AE1 spectrum energy characteristics *Fig.12 AE2 spectrum energy characteristics*

By analyzing the eigenvalues identified in the sliding window and the initial work one by one, and setting the error threshold as 0.1, the electromagnetic radiation data from May 1, 2022 to May 30, 2022 are sequentially analyzed.The acoustic emission signal data from April 1, 2022 to May 30, 2022 and October 10, 2022 to November 10, 2022 were identified for the time interval of interference signals.

The following are the time intervals of the first five interference signals in electromagnetic radiation and acoustic emission signals, as shown in Table 3 and Table four.

Table 3 Time interval of electromagnetic radiation interference signal

Serial number	Start of time interval	End of time interval
1	2022/5/5 1:16:00	2022/5/5 1:17:49
2	2022/5/6 15:20:39	2022/5/6 15:22:29
3	2022/5/12 23:32:35	2022/5/12 23:34:25
4	2022/5/13 3:29:26	2022/5/13 3:37:36
5	2022/5/24 13:02:30	2022/5/24 13:04:20

Table 4 Time interval of acoustic emission interference signal

Serial number	Start of time interval	End of time interval
1	2022/4/2 21:38:12	2022/4/3 15:20:51
2	2022/4/3 16:08:24	2022/4/3 17:03:16
3	2022/4/3 17:58:09	2022/4/3 19:38:44
4	2022/4/3 21:26:39	2022/4/4 16:37:10
5	2022/4/4 20:25:46	2022/4/4 21:18:48

3.1.5 Result analysis

By analyzing the data in Table 3, we can see that the time span of the obtained data is from May 5, 2022 to May 24, 2022, and the time span of each time interval is different, ranging from about 1 minute to 8 minutes, respectively. By analyzing the data in Table 4, it can be seen that the time span of the obtained data is from April 2, 2022 to April 4, 2022. Compared with the interference signal from electromagnetic radiation, the interference signal from acoustic emission has a more concentrated occurrence time and a longer duration.

3.2 Model establishing and solving for work two

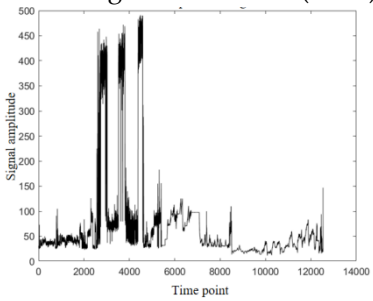
3.2.1 Data preprocessing

In order to solve the first work, firstly, the data in Annex I is screened and cleaned, the precursor signal (Class B signal) is extracted, and the precursor signal itself is analyzed. Process outliers and arrange the data in ascending chronological order. For those groups of different data appearing in the same period of time, the mean value is calculated here, and the data with small differences is calculated and then brought back to the attachment.

3.2.2 Data visualization analysis

After data preprocessing, visualization processing was carried out to analyze the characteristics of the data as a whole. The visualization processing of acoustic intensity (AE)

and electromagnetic radiation (EMR) data was shown in Fig.13 and Fig.14.



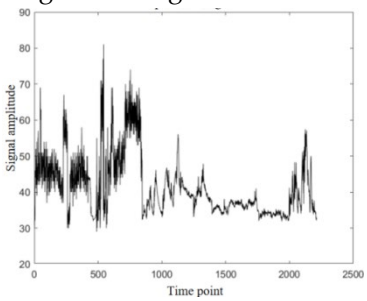


Fig.13 Visualization of AE precursor signals *Fig.14 visualization of EMR precursor signals*

From the analysis of Fig.13 and Fig.14, it can be seen that the transformation of precursor signals has the characteristics of excessive performance and random change in regional time, and also shows randomness and uncertainty in the periodicity of change. In addition, the peaks and transformation intensity shown on the visualization map indicate anomalies in the detection machine data that may be caused by external geological changes in a specific period of time, resulting in the appearance of outliers in the data.

3.2.3 Recognition of interference signals based on sliding window

Firstly, the two types of precursor signals in Annex 1 are extracted. Then, the size of the sliding window is set to 25, and various features such as mean value, variance, kurtosis, spectral energy and spectral density are calculated respectively for the two types of precursor signals in Annex 1. These features are put into the feature matrix 1, that is, 1. Then, the corresponding features of non-precursor signals are extracted and put into matrix 2. Finally, the feature matrix is identified with many data of precursor signals in the time period to be identified, and the test sets of AE and ERM in the three time periods can be obtained. The corresponding images of the obtained features are shown in Fig.15.

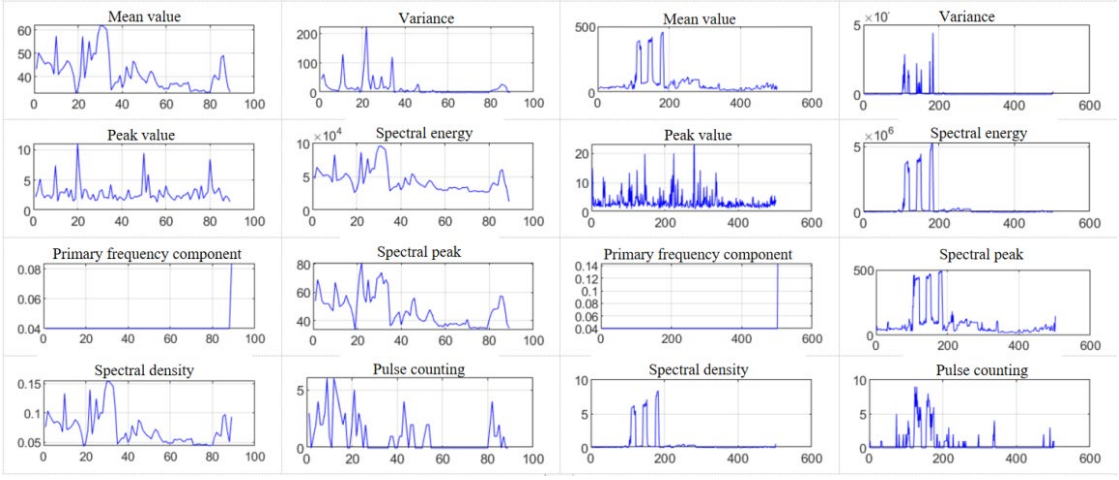


Fig.15 Characteristics of AE and EMR precursor signals

According to these features, precursor signals within the required time period for subsequent identification can be carried out.

3.2.4 Solution based on decision tree model

Decision tree is a kind of non-parametric supervised learning, which can summarize decision rules from a series of data with characteristics and labels, and present these rules in the structure of tree graph to solve classification and regression problems. Decision tree algorithms are easy to understand, applicable to all kinds of data, and have good performance in solving various problems, especially various integrated algorithms with tree model as the core. It is widely used in various industries and fields. The classification decision tree model is a kind of tree structure that describes the classification of instances. The decision tree consists of nodes and directed edges. There are two types of nodes: internal nodes and leaf nodes. An inner node represents a feature or attribute, and a leaf node represents a class. The important parameters of the decision tree model are briefly summarized as shown in equation (9) :

$$Entropy(t) = -\sum_{i=0}^{c-1} p\left(\frac{i}{t}\right)\log_2 p\left(\frac{i}{t}\right) \tag{9}$$

Information entropy is used to measure the degree of order of a system. The greater the information entropy of the system, the more chaotic the system is, that is, the more uncertain the situation is. The greater the entropy, the more chaos. Information gain: Measures the change in entropy, that is, the decrease in the uncertainty of the data after the selection of feature A. The greater the information gain, the stronger the classification capability of this feature, and the priority is given to this feature. Gini coefficient is used to represent the clutter degree of a feature containing attributes, and its calculation formula is shown in equation (10) :

$$Gini(t) = 1 - \sum_{i=0}^{c-1} p\left(\frac{i}{i}\right)^2 \tag{10}$$

Where t represents a given node, i represents any classification of the label, and $p(i/t)$ represents the proportion of the label classification i on the node t.

Note: When using information entropy, sklearn actually calculates information gain based on information entropy, that is, the difference between the information entropy of the parent node and the information entropy of the child node. The confusion matrix and ROC curves obtained by using the decision tree model are shown in Fig.16 and Fig.17.

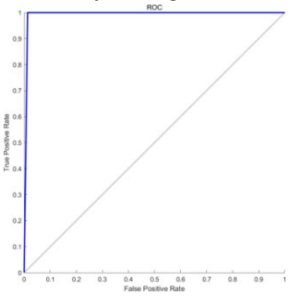


Fig.16 ROC curve of AE

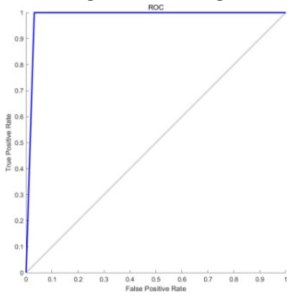
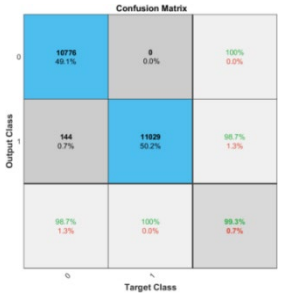


Fig.17 ROC curve of EMR



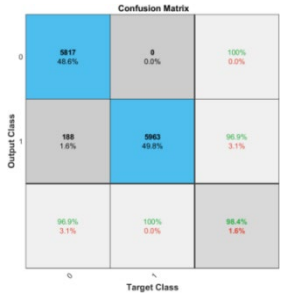
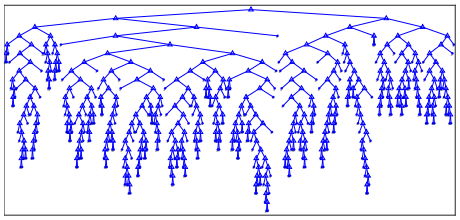


Fig.18 Confusion matrix of AE *Fig.19 Confusion matrix of EMR*

The Fig.18 and Fig.19 shows the confusion matrix and ROC curves of AE and EMR respectively. Through the confusion matrix, various classification indexes can be calculated, such as accuracy rate, recall rate, F1 score and so on. ROC curve is a graphical tool used to evaluate the performance of binary classification models. It shows the performance of classifiers under different thresholds with True Positive Rate (TPR) as the vertical axis and False Positive Rate (FPR) as the horizontal axis. This ROC curve is very close to the upper left corner, indicating that the performance of the model is good and fits the model well. Next, the decision tree model is used for fitting. The structure of the decision tree is shown in Fig.20 and Fig.21:



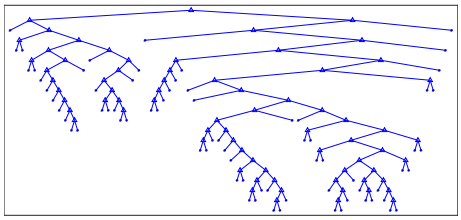


Fig.20 AE decision tree structure *Fig.21 EMR decision tree structure*

The generation process of decision tree is a recursive process, starting from the root node, the optimal features are selected according to the partitioning conditions, until the stopping conditions are met (such as reaching the maximum depth, the number of node samples is less than the threshold, etc.), or the samples of child nodes belong to the same category or the regression is worth stable. Finally, two decision trees are calculated to obtain the evaluation values, as shown in Table 5 and Table 6:

Table 5 Evaluation data of AE decision tree model

Accuracy	Accuracy rate	Recall rate	F1 Score	AUC
98.43%	97%	1	0.98	0.98

Table 6 EMR decision tree model evaluation data

Accuracy	Accuracy rate	Recall rate	F1 Score	AUC
98.43%	99%	1	0.99	0.99

Finally, the earliest five intervals of the precursor signal are shown in Table 7 and Table 8:

Table 7 Time interval of electromagnetic radiation precursor characteristics

Serial number	Start of time interval	End of time interval
1	2020/4/8 1:01:21	2020/4/8 1:11:02
2	2020/4/8 1:40:47	2020/4/8 3:42:48
3	2020/4/8 5:26:25	2020/4/8 5:39:00
4	2020/4/8 6:08:03	2020/4/8 7:56:29
5	2020/4/8 8:04:14	2020/4/8 9:54:03

Table 8 Time interval of acoustic emission precursor characteristics

Serial number	Start of time interval	End of time interval
1	2021/11/1 0:28:11	2021/11/1 0:30:00
2	2021/11/1 0:33:37	2021/11/1 0:42:40
3	2021/11/1 0:44:29	2021/11/1 0:49:55
4	2021/11/1 0:57:10	2021/11/1 1:09:51
5	2021/11/1 1:11:39	2021/11/1 1:13:28

3.2.5 Result analysis

From the analysis of the data in Table 7, we can see that the duration of time interval 2 is the longest, lasting about 2 hours, 2 minutes and 1 second, while the duration of other time intervals is relatively short. From the perspective of time interval, the interval between time intervals is relatively irregular. There are no significant outliers for the duration and time interval, both within reasonable limits. From the analysis of the data in Table 8, it can be seen that the duration of time interval 4 is the longest, lasting about 12 minutes and 41 seconds, while the duration of other time intervals is relatively short. From the perspective of time interval, the interval between time intervals is more irregular, some intervals are shorter, and some intervals are longer.

3.3 Model establishing and solving for work three

3.3.1Data preprocessing

For the noise reduction processing of the original data in Annex 3, the wavelet transform denoising processing is selected here, and the wavelet transform is a signal processing technology for the analysis of signals in the time-frequency domain. It breaks down the signal into components of different scales and frequencies, thus providing a way to capture both local and global features of the signal at the same time. Wavelet transform uses wavelet function to analyze the signal. Wavelet function is a kind of waveform function with present length and bounded energy. The wavelet function can be translated and scaled along the time axis, and has local optimization. The basic steps of wavelet transform can be divided into four stages: decomposition, filtering, downsampling and reconstruction. The following are specific instructions for each stage:

1) Decomposition stage: the original signal is decomposed to obtain the approximate coefficient and detail coefficient; Decomposition is usually achieved using filter banks, with one used to extract the approximation coefficient and the other to extract the detail coefficient; 2) Filtering stage: In the decomposition stage, the original signal passes through a pair of filters, usually called a low-pass filter and a high-pass filter, the low is used to extract approximate components, and the high-pass is used to extract detailed components; 3) Downsampling stage: Downsampling means to reduce the sampling rate of the signal, that is, to reduce the time resolution of the signal; 4) Reconstruction stage: In the reconstruction stage, the approximate coefficient and detail coefficient are used to reconstruct the original signal, and the approximate signal and detail signal are up-sampled, filtered and added by a pair of reconstruction filters, and finally the reconstructed signal is obtained.

The formula of continuous wavelet transformation is shown in equation (11) :

$$\mathrm{CWT}(a,b) = \int_{-\infty}^{+\infty} x_{(t)} \cdot \frac{1}{\sqrt{|a|}} \varphi\left(\frac{t-b}{a}\right) \mathrm{dt} \tag{11}$$

3.3.2 Data visualization analysis

First, visualize the original data and observe the general change trend of the data in the form of line charts. The visualization of some time periods of the two types of data is shown in Fig.22 and Fig.23:

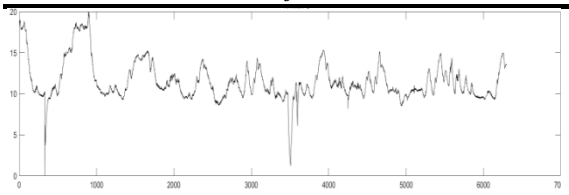


Fig.22 Partial time period visualization of EMR data

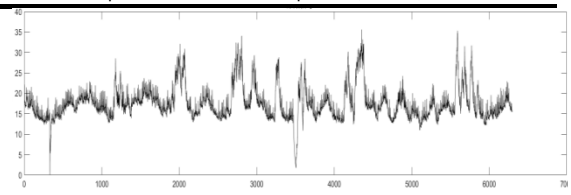


Fig.23 Visualization of AE data in some time periods

According to the analysis of Fig.23 and Fig.24, the randomness of the change trend of the original data is strong, and the internal data contains a large amount of noise, which will affect the accuracy of the final result prediction.

The data visualization after noise reduction is shown in Fig.24 and Fig.25.

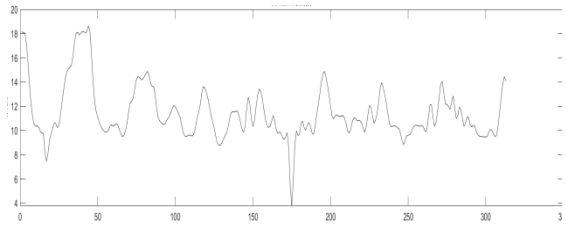


Fig.24 EMR data visualization after noise reduction

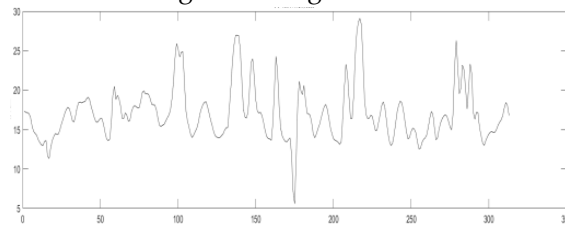


Fig.25 Visualization of AE data after noise reduction

According to the analysis of Fig.24 and 25, after noise reduction, the data changes more smoothly and the noise data is reduced a lot, which is of great help to the following data prediction and can make the final prediction result more accurate.

3.3.3 Predict future new data based on LSTM time series

Since samples are needed to extract features for identification, LSTM time series prediction is performed on the original data, and new predicted data is obtained for feature extraction by sliding window.

A Long Short-Term Memory network (LSTM) is a variant of a recurrent neural network (RNN) that is commonly used to process and predict time series data. The core of the LSTM network is its internal gating mechanism, which consists of three kinds of gates: forget gate, input gate and output gate, and a cell state. These gates control the flow of information, allowing the network to selectively forget, store, and read information to better handle long sequences. The gating mechanism of LSTM network enables it to handle time series data better, capture long-term dependency effectively, and resist the gradient disappearance problem, so it has been widely used in language modeling, machine translation, time series prediction and other tasks. The key of LSTM neural network is the cell state. Its logical architecture is shown in Fig.26, where xt, as the input of the network at the current moment, determines how much is saved to the control unit state ct, and ht represents the output value of the unit control state.

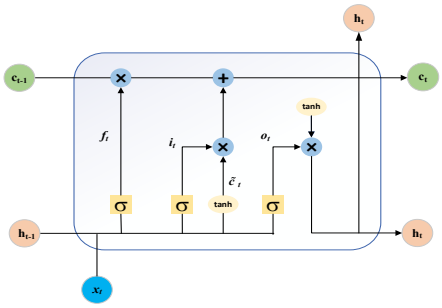


Fig.26 Cell structure of LSTM neurons

The forgetting gate decides to retain or forget the information in the cell state, and saves the required ct-1 state parameters through formula (12); The function of the input gate is opposite to the forgetting gate. The input saved to ct is determined by formulas (13) and (14), and the selection information is put into the cell state; The output gate controls the information output by formula (15), formula (16) and formula (17). Before each output, the information is updated and some information of the old state is discarded. The activation function is equation (18) and equation (19).

$$f_t = \sigma \cdot (W_f \cdot [h_{t-1}, x_t]) + b_f \tag{12}$$

$$i_t = \sigma \cdot (W_i \cdot [h_{t-1}, x_t]) + b_i \tag{13}$$

$$o_t = \sigma \cdot (W_0 \cdot [h_{t-1}, x_t]) + b_o \tag{14}$$

$$\tilde{c}_t = tanh \cdot (W_c \cdot [h_{t-1}, x_t]) + b_c \tag{15}$$

$$h_t = o_t \cdot tanh(c_t) \tag{16}$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \tilde{c}_t \tag{17}$$

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{18}$$

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{19}$$

Where: ft, it and ot are the outputs of the forgetting door, the input door and the output door respectively; ct is the internal function of LSTM neurons. W and b are weight parameter matrix and bias parameter matrix respectively. ht and ct are the internal memory vectors of neurons at time t.

The architecture diagram of the multi-layer stacked LSTM neural network is shown in Fig.27. The neural network architecture shown in Fig.27 is composed of three layers of LSTM hidden layers. A in Fig.represents LSTM neuron cells, and each hidden layer is composed of neuron cell chain as shown in Fig.27.

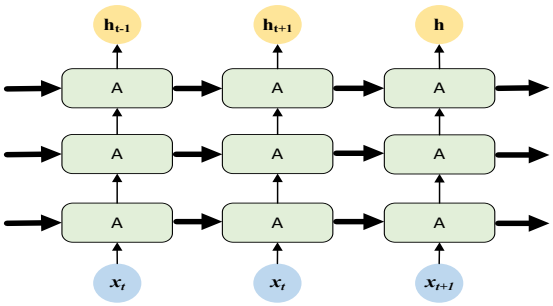


Fig.27 LSTM neural network architecture

After prediction, the two types of data are shown in Fig.28 and Fig.29.

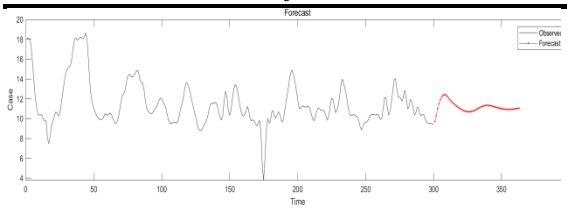


Fig.28 EMR prediction data

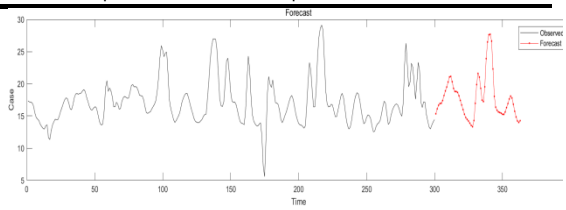


Fig.29 AE prediction data

The above is the forecast data graph of the two types of signals at time period 3, and it can be observed that the predicted results are in good agreement with the change trend of the original data. Prove that the data can use this model.

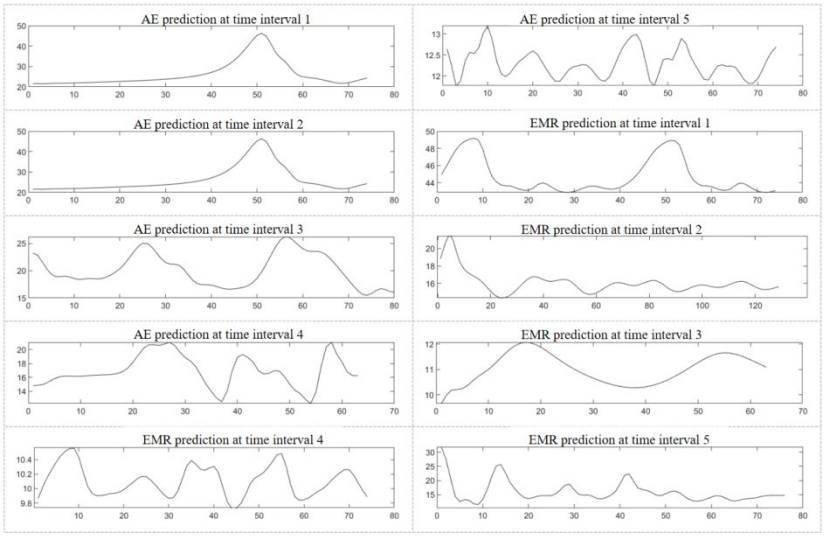


Fig.30 AE and EMR prediction data for each time period

Fig.30 shows the AE and EMR data for the five time periods derived from the force prediction.

3.3.4 Recognition of interference signals based on sliding window

It is the same as the method used in work 2, but here the size of the sliding window needs to be changed to a group of 25 data, so that key features can be better retained and the results obtained by the final identification can be more accurate.

3.4.5 Recognition of precursor signals based on decision tree model

The characteristic matrix extracted from the predicted data is brought into the decision tree model of work 2, and then the position of the precursor signal in Annex 3 of the AE and EMR data can be obtained. Finally, the probability can be obtained by comparing the number of precursor signal data with the total amount of data after the response period. The results are shown in Table 9.

Table 9 Probability of precursor features at the time of data collection

Electromagnetic radiation data	Probability of precursor features	Acoustic emission data	Probability of precursor features
2023-1-24 23:58:36	0.1863	2023-1-24 23:58:36	0.1863
2023-2-11 23:59:20	0.6925	2023-2-11 23:59:20	0.2859
2023-2-26 23:59:27	0.6032	2023-2-26 23:59:27	0.5997
2023-3-10 23:58:14	0.8347	2023-3-10 23:58:14	0.1602
2023-3-30 23:58:13	0.8372	2023-3-30 23:58:13	0.5001

3.4.6 Result analysis

According to the analysis of Table 9, the probability of precursor features produced by electromagnetic radiation has a certain changing trend at different times, sometimes high and sometimes low, which may reflect the instability or periodicity of electromagnetic radiation activities. The probability of precursory features in acoustic emission data also shows a certain variation trend, but it is not exactly the same as that of electromagnetic radiation data. This may indicate that electromagnetic radiation and acoustic emission are correlated in some respects, but there may also be independent influencing factors.

Both sets of data provide some information about electromagnetic radiation and acoustic emission activity, but more in-depth analysis and research is needed to understand the relationship between them and the potential impact on the environment or equipment.

4 CONCLUSION

DFT can convert the time domain signal into a frequency domain representation, so that the frequency component of the signal can be observed more clearly, which is helpful for frequency domain analysis and filtering. DFT can decompose the signal into different frequency components, analyze the spectrum characteristics of the signal, and identify the periodic or periodic components in the signal. DFT can be used to extract the frequency domain features of the signal, such as spectrum peak value, spectrum energy, etc., which helps to analyze and classify the signals.

The sliding window technology has the advantages of real-time, simplicity and flexibility. The sliding window can retain the time order information of the data, so that the extracted features can capture the time correlation of the data, which is conducive to the establishment of time series model or time series analysis. Sliding window can be selected according to the needs of different sizes of Windows and sliding steps, flexible to adapt to the characteristics of different data and analysis needs. By extracting features through sliding window, the local feature information of the data can be obtained, which is conducive to modeling and analyzing the locality of the data.

The rules generated by the decision tree model are similar to the human decision-making process and are easy to understand and explain. The decision tree can be visualized to help users understand how the model works. Decision tree models can be easily scaled to large data sets, and the model is usually trained faster when the amount of data or the number of features is increased. Wavelet transform can perform local analysis of the signal in the time-frequency domain, so as to better capture the local characteristics of the signal. The wavelet transform can achieve multi-resolution analysis by adjusting the scale, so as to observe the signal at different time and frequency scales.

LSTM can process input sequences of indefinite length, so it is suitable for time series data of various lengths. LSTM can effectively capture long-term dependencies in time series data, which is very important for many practical applications. Memory units in LSTM can effectively store and update historical information, which helps to improve the model's understanding and prediction ability of series data.

5 ACKNOWLEDGEMENTS

This work was supported by ministry of education industry-university cooperative education project (Grant No.: 231106441092432) and special research project on teaching reform(Grant No.: 30120300100-23-yb-jgkt03)

REFERENCES

[1] Zhai, J., & Cao, J. (2016). A combined prediction model based on time series ARIMA and BP neural network. *Statistics and Decision Making*, 04: 29-32.

[2] Yang, C. X., Huang, H., & Dai, Y. C. (2017). Research and implementation of key technologies for 3D visualization of distribution networks based on CIM. *Electric Power Science and Engineering*, 33(05): 6-10.

[3] Wu, J. F. (2021). Research on anomaly detection and fault warning technology for in orbit satellites based on LSTM. *National University of Defense Technology*.

[4] Wang, G. N., & Tang, X. P. (2022). Research on rural emergency logistics distribution path based on simulated annealing and floyd optimization algorithm. *Software Engineering*, 25(12): 9-12.

[5] Yang, J., Yang, X. D., & li, L. S. (2022). Optimization of stacker crane path based on genetic simulated annealing algorithm. *Logistics Technology*, 41(08): 119-123.

[6] Zhou,G. Y., Wan, S. P., & Chen.Y. N. (2022). Research on denoising algorithm for phase sensitive time domain reflectometer based on moving variance average algorithm. *Journal of Instruments and Meters*, 43(10): 233-240.

[7] Zhang, J. L., Guo, S. Y., & Ren, C. P. (2024). Research on personal credit rating card model based on logistic regression. *Modern Information Technology*, 8(05): 12-16.

[8] Du, B. Y., Gao, J. H., & Zhang, G. Z. (2024). Seismic prediction method and application of fracture density inversion for shale reservoir in-situ stress. *Petroleum Geophysical Exploration*, 59(2): 279-289.

[9] Zhou, Z. X., Zhen, X. J., & Liang, Y. G. (2024). Study on acoustic emission source location of ancient timber based on wavelet packet transform and cross-correlation. *Shanxi Architecture*, 50(9): 1-5.

[10] Yao, J. W., Li, Y., & Lv, Y. J. (2024). Research on the trend of water quality evolution in drinking water source areas based on water quality index method and M-K Test. *Environmental Science and Management*, 49(04): 43-48.

Copyright © 2024 by the author(s). Published by Sichuan Knowledgeable Intelligent Sciences. This is an open access article under the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) License (https://creativecommons.org/licenses/by-nc/4.0/).