

Research Article

The Use of Intelligent Sensing Algorithm for Internet of Things and Hash Spatial Information Location Technology

Shaobo Wang ¹ and Yujia Liu^{1,2}

¹State Grid Info & Telecom Group, China Gridcom Corporation, LTD, Beijing, 102299 Beijing, China

²School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, 100049 Beijing, China

Correspondence should be addressed to Shaobo Wang; victorwhy@emails.bjut.edu.cn

Received 22 October 2021; Accepted 27 November 2021; Published 29 December 2021

Academic Editor: Vinoth Babu Kumaravelu

Copyright © 2021 Shaobo Wang and Yujia Liu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This exploration is aimed at quickly obtaining the spatial position information of microseismic focal points and increasing the accuracy of microseismic rapid positioning, to take timely corresponding measures. A microseismic focal point location system completely different from the traditional microseismic location method is proposed. The search engine technology is introduced into the system, which can locate the microseismic focal point quickly and accurately. First, the propagation characteristics of microseismic signals in coal and rock layers are analyzed, and the focal position information is obtained. However, the collected microseismic signal of the coal mine contains noise, so it is denoised at first. Then, a waveform database is established for the denoised waveform data and focal point position. The structure and mathematical model of the location-sensitive hash (LSH) based on P stable distribution are introduced and improved, and the optimized algorithm multiprobe LSH is obtained. The microseismic location model is established according to the characteristics of microseismic data. The values of three parameters, hash table number, hash function family dimension, and interval size, are determined. The experimental data of the parameters of the search engine algorithm are analyzed. The results show that when the number of hash tables is 6, the dimension k of the hash function family is 14, and the interval size W is 8000, the retrieval time reaches a relatively small value, the recall rate reaches a large value, and the proportion of retrieved candidates is large; the parameters of the search engine algorithm of the measured coal mine microseismic data are analyzed. It is obtained that when the number of hash tables is 4, the dimension k of the hash function family is 6, and the interval size W is 500, the retrieval time reaches a relatively small value, the recall rate obtains a large value, and the proportion of retrieved candidates is large. The contents studied are of great significance to the evaluation of destructive mine earthquakes and impact risk.

1. Introduction

In recent years, with the rapid progress of the Internet of things (IoT) technology, intelligent sensing algorithm, and computer vision, accurate and fast search of large-scale spatial data has become a major problem perplexing researchers worldwide. The key point of searching large-scale high-dimensional spatial data is to convert the data points. In this way, the distance between the data points and the query can be approximately regarded as the distance between compact codes based on efficient calculation, so as to shorten the time required for retrieval.

The advantage of using this kind of retrieval technology is that it can find large-scale data accurately without compressing it [1, 2]. Combining this technology with microseismic focal point location can realize the balance between microseismic focal point calculation efficiency and positioning accuracy, providing a feasible scheme for fast microseismic focal point location. The application of the location-sensitive hash (LSH) algorithm to the establishment of the microseismic search engine can quickly find and interpret the waveform data most similar to the input microseismic waveform in the built complete database. The spatial position of the microseismic focal point can be accurately located

[3–5] through this method. LSH is involved in computing problems in multiple fields, such as computer vision, language computing statistics, drug design, and text clustering. Therefore, massive scholars have studied the specific problems in its application. Some researchers verify the entropy-based LSH by querying the adjacent data points of randomly generated disturbance points. Finally, the data result of the query point and its set will be returned. This method can reduce the number of hash tables, improve the utilization of hash tables, and save storage space [6]. LSH is also widely used in the fields of audio and image search in China. However, its main application is to detect the repeatability and approximation of web pages to reduce the cost of web page indexing.

Today, LSH technology has encountered a bottleneck problem. Traditional LSH technology usually needs massive hash tables to ensure precision and recall. However, too many hash tables will lead to too much memory, causing some difficulties in storing indexes [7, 8]. Thereby, the structure of the LSH algorithm is designed based on P stable distribution, and the multiprobe LSH algorithm is improved, which improves the accuracy of microseismic rapid location and increases the recall rate. Therefore, it is of great significance to study the application of the approximate proximity algorithm of LSH in microseismic focal point location.

2. Materials and Methods

2.1. Principle of Microseismic Monitoring Technology

2.1.1. Generation Principle of Microseismic Signal in Coal Mine Strata [9–11]. Microseismic refers to the destruction of rock structures under the influence of mining behavior, which produces a kind of elastic energy. When the generated energy continues to accumulate and reaches a limit, it will be released and propagated in the form of an elastic wave. Figure 1 is its schematic diagram.

Mine microseism is determined by the damage stress of rock mass and the accumulation ability of rock mass to elastic properties. There is a positive correlation between the stress causing damage to the rock mass and the strength of the rock mass. The greater the stress required for damage to the rock mass is, the greater the strength of the rock mass is. The Mohr-Coulomb criterion considers the failure criterion of matrix rock block and the weak structural plane of a broken body under compressive shear stress, expressed as

$$\tau = \sigma \tan \phi + C, \quad (1)$$

where τ, σ are shear stress and normal stress on the failure surface of rock mass, ϕ is the angle when friction is caused inside the rock mass, $\tan \phi$ is the internal friction coefficient, and C is cohesion force in rock mass.

Figure 2 displays the causes of microseismic in coal and rock mass.

2.1.2. Propagation of Microseismic Signal in Coal Mine Strata. The vibration wave generated by the coal and rock stratum propagates in the form of an elastic stress wave

between coal and rock mass. Therefore, the theoretical basis for deriving the propagation equation of vibration wave in coal and rock stratum medium is the relevant theory of stress wave [12–14]. The specific derivation process is as follows.

It is assumed that there is a force shown in equation (2) in any microunit $dv = dx dy dz$ of coal and rock mass:

$$F_i (F_x = F' x r dv, F_y = F' y r dv, F_z = F' z r dv). \quad (2)$$

$u_i(u_x, u_y, u_z)$ represents displacement. The equation reads

$$u_i = \text{grad } \Phi + \text{rot } \Psi, \quad (3)$$

where Φ, Ψ are the scalar potential and vector potential. In addition to the force shown in equation (2), there are the surface force and inertial force in the microunit of coal and rock mass. Equation (4) can represent the density component of inertial force:

$$\rho \frac{\partial^2 u_x}{\partial t^2}, \rho \frac{\partial^2 u_y}{\partial t^2}, \rho \frac{\partial^2 u_z}{\partial t^2}. \quad (4)$$

Differential equation (5) when the microunit in coal and rock mass moves can be obtained as follows:

$$\begin{cases} \frac{\partial \sigma_x}{\partial x} + \frac{\partial \tau_{yx}}{\partial y} + \frac{\partial \tau_{zx}}{\partial z} + F_x = \rho \frac{\partial^2 u_x}{\partial t^2}, \\ \frac{\partial \sigma_y}{\partial x} + \frac{\partial \tau_{zy}}{\partial y} + \frac{\partial \tau_{xy}}{\partial z} + F_y = \rho \frac{\partial^2 u_y}{\partial t^2}, \\ \frac{\partial \sigma_z}{\partial x} + \frac{\partial \tau_{xz}}{\partial y} + \frac{\partial \tau_{yz}}{\partial z} + F_z = \rho \frac{\partial^2 u_z}{\partial t^2}. \end{cases} \quad (5)$$

Equations (6)–(9) are geometric and physical equations in elasticity.

Geometric equations:

$$\begin{cases} \varepsilon_x = \frac{\partial u}{\partial x}, \\ \varepsilon_y = \frac{\partial u}{\partial y}, \\ \varepsilon_z = \frac{\partial \omega}{\partial y}, \end{cases} \quad (6)$$

$$\begin{cases} \gamma_{xy} = \frac{\partial v}{\partial x} + \frac{\partial v}{\partial x}, \\ \gamma_{zx} = \frac{\partial u}{\partial z} + \frac{\partial \omega}{\partial x}, \\ \gamma_{yz} = \frac{\partial v}{\partial z} + \frac{\partial \omega}{\partial y}. \end{cases} \quad (7)$$

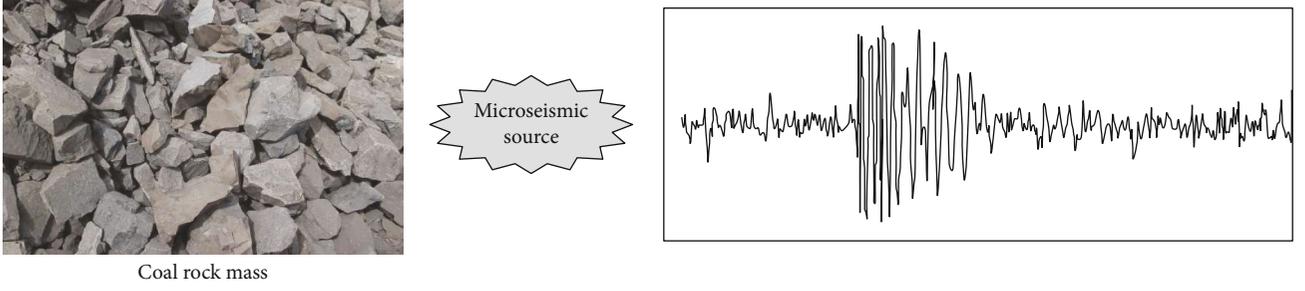


FIGURE 1: Schematic diagram of microseismic signal propagation.

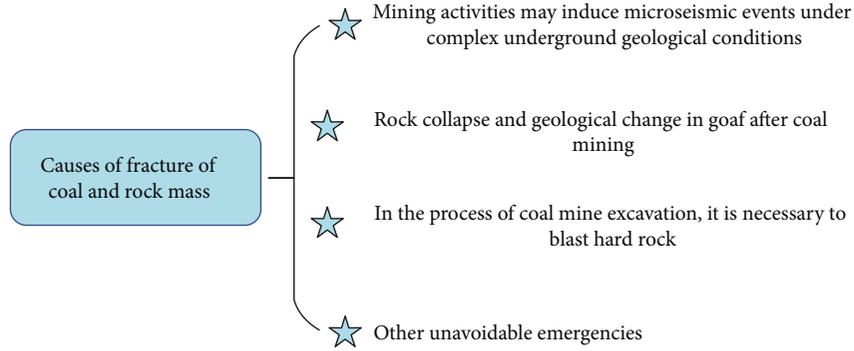


FIGURE 2: Causes of microseismic in coal and rock mass.

Physical equations:

$$\begin{cases} \varepsilon_x = \frac{1}{E} (\sigma_x - \mu(\sigma_y + \sigma_z)), \\ \varepsilon_y = \frac{1}{E} (\sigma_y - \mu(\sigma_x + \sigma_z)), \\ \varepsilon_z = \frac{1}{E} (\sigma_z - \mu(\sigma_x + \sigma_y)), \end{cases} \quad (8)$$

$$\begin{cases} \gamma_{xy} = \frac{2(1+\mu)}{E} \tau_{xy}, \\ \gamma_{yz} = \frac{2(1+\mu)}{E} \tau_{yz}, \\ \gamma_{zx} = \frac{2(1+\mu)}{E} \tau_{zx}. \end{cases} \quad (9)$$

Equations (6)–(9) are brought into equation (5), so that the differential equation of elastic wave propagation in coal and rock stratum is obtained as

$$\begin{cases} (\lambda + \mu) \frac{\partial e}{\partial x} + \mu \nabla^2 u_x + F_x = \rho \frac{\partial^2 u_x}{\partial t^2}, \\ (\lambda + \mu) \frac{\partial e}{\partial y} + \mu \nabla^2 u_y + F_y = \rho \frac{\partial^2 u_y}{\partial t^2}, \\ (\lambda + \mu) \frac{\partial e}{\partial z} + \mu \nabla^2 u_z + F_z = \rho \frac{\partial^2 u_z}{\partial t^2}, \end{cases} \quad (10)$$

$$\lambda = \frac{Ev}{2(1+\nu)(1-2\nu)}, \quad (11)$$

$$\mu = \frac{E}{2(1+\nu)}, \quad (12)$$

$$\nabla^2 = \frac{\partial}{\partial x^2} + \frac{\partial}{\partial y^2} + \frac{\partial}{\partial z^2}, \quad (13)$$

$$e = \frac{\partial}{\partial x} + \frac{\partial}{\partial y} + \frac{\partial}{\partial z}, \quad (14)$$

where λ is constant, ν is Poisson's ratio, E is Young's modulus, and μ is shear modulus.

If $F_i = 0$, equation (15) can be obtained by deriving equation (10):

$$\begin{cases} \frac{\partial^2 u_x}{\partial t^2} = \frac{\lambda + 2\mu}{\rho} \nabla^2 u_x, \\ \frac{\partial^2 u_y}{\partial t^2} = \frac{\lambda + 2\mu}{\rho} \nabla^2 u_y, \\ \frac{\partial^2 u_z}{\partial t^2} = \frac{\lambda + 2\mu}{\rho} \nabla^2 u_z. \end{cases} \quad (15)$$

If the rotor can replace u_i , equation (16) can be obtained:

$$\frac{\partial^2 (\text{rot} u_i)}{\partial t^2} = \frac{\mu}{\rho} \nabla^2 \text{rot} u_i. \quad (16)$$

Equations (15) and (16) can, respectively, describe the deformation caused by microseism on the volume and position of coal and rock stratum. They can be written as

$$\frac{\partial^2 \Phi}{\partial t^2} = \frac{\lambda + 2\mu}{\rho} \nabla^2 \Phi = V_p^2 \nabla^2 \Phi, \quad (17)$$

$$\frac{\partial^2 \psi}{\partial t^2} = \frac{\mu}{\rho} \nabla^2 \psi = V_s^2 \nabla^2 \psi. \quad (18)$$

The above equations display that the force propagating in coal and rock stratum will propagate in two different forms: shear wave (S wave) and longitudinal wave (P wave). P wave propagates faster than S wave in coal and rock strata [15–17]. Its velocity expression reads

$$V_p = \sqrt{\frac{E(1-\nu)}{\rho(1+\nu)(1-2\nu)}}. \quad (19)$$

The propagation direction of the S wave differs from that of the P wave. It is perpendicular to the vibration direction of coal and rock stratum. The velocity expression of the S wave reads

$$V_s = \sqrt{\frac{E}{2\rho(1+\nu)}}, \quad (20)$$

where ρ is density of coal and rock stratum and ν is Poisson's ratio.

2.2. Principle and Method of Microseismic Monitoring and Positioning. Figure 3 shows the distribution of focal points and sensors.

If the microseismic focal point in Figure 3 is $X(x_0, y_0, z_0, t_0)$ and the sensor is represented by $T_i(x_i, y_i, z_i, t_i)$, x, y, z represent the spatial position and t is time [18–21]. Unlike S wave, picking up P wave is relatively simple. Equation (21) can be used to locate focal points at different time points:

$$v(t_i - t_0) = \sqrt{(x_i - x_0)^2 + (y_i - y_0)^2 + (z_i - z_0)^2}, \quad (21)$$

where all parameters except x_0, y_0, z_0, t_0 are known.

One of the methods to locate the microseismic focal point is the simplex method, as shown in Figure 4.

Equation (22) is the functional expression of the method:

$$w = \sum_{i=1}^n \left| t_i - \frac{\sqrt{(x_i - x_0)^2 + (y_i - y_0)^2 + (z_i - z_0)^2}}{v} - t_0 \right|, \quad (22)$$

where i is the number of microseismic sensors is v is set velocity of vibration wave.

It is assumed that $A, B, C,$ and D in the left figure of Figure 4 are the occurrence points of microseisms. First, the w values of the four points are calculated by equation (22), and the point with the largest function value among the four points is found and removed according to the calcu-

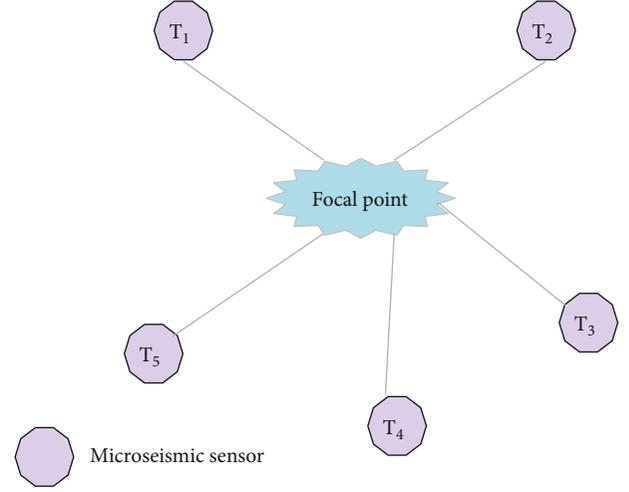


FIGURE 3: Distribution of focal point and sensor.

lation results. It is assumed that the maximum value of the function is point A . After the removal of point A and some operations such as shrinking and stretching the graph, vertex S is selected to finally form a new tetrahedron $SBCD$. Then, the above operations are repeated until the limit of the objective function or the constraint of the number of iterations is reached, and the process is over. At this time, the minimum point of the function value is the occurrence point of microseisms [22, 23].

2.3. Introduction of the Microseismic Monitoring System. The proposed microseismic focal point location method completely differs from the original location method. The advantage of building a search engine based on the LSH algorithm is that iterative calculation is not required when determining the focal point position, so as to reduce the damage caused by rockburst to rock mass [24].

Microseismic focal point location technology mainly determines the spatial position of the focal point through the vibration wave released by coal and rock mass. It can detect mine seismic activities in real time, determine the spatial position of the focal point, and calculate microseismic energy to judge the microseismic strength of coal and rock strata [25–27]. Figure 5 is a schematic diagram of the microseismic monitoring system.

There are 10 geophones in the mining area studied. Figure 6 shows their spatial coordinates.

Mainly, the underground geophone is powered by the ground acquisition station. Each geophone will receive 37–42 V safety voltage from the acquisition station, and the acquisition station contains a fault indicator light, which can monitor the faults of underground lines in real time. The downhole environment is quite complex, so it is essential to protect the geophone sensor. Figure 7 is a schematic diagram of its protection device.

2.4. Scheme Design of Ensemble Empirical Mode Decomposition (EEMD) Wavelet Threshold Denoising Algorithm for Coal Mine Microseismic Waveform. The mining process is often affected by other construction, resulting

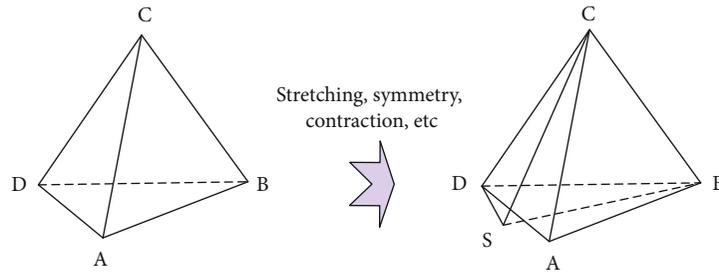


FIGURE 4: Iterative tetrahedron diagram of the simplex method.

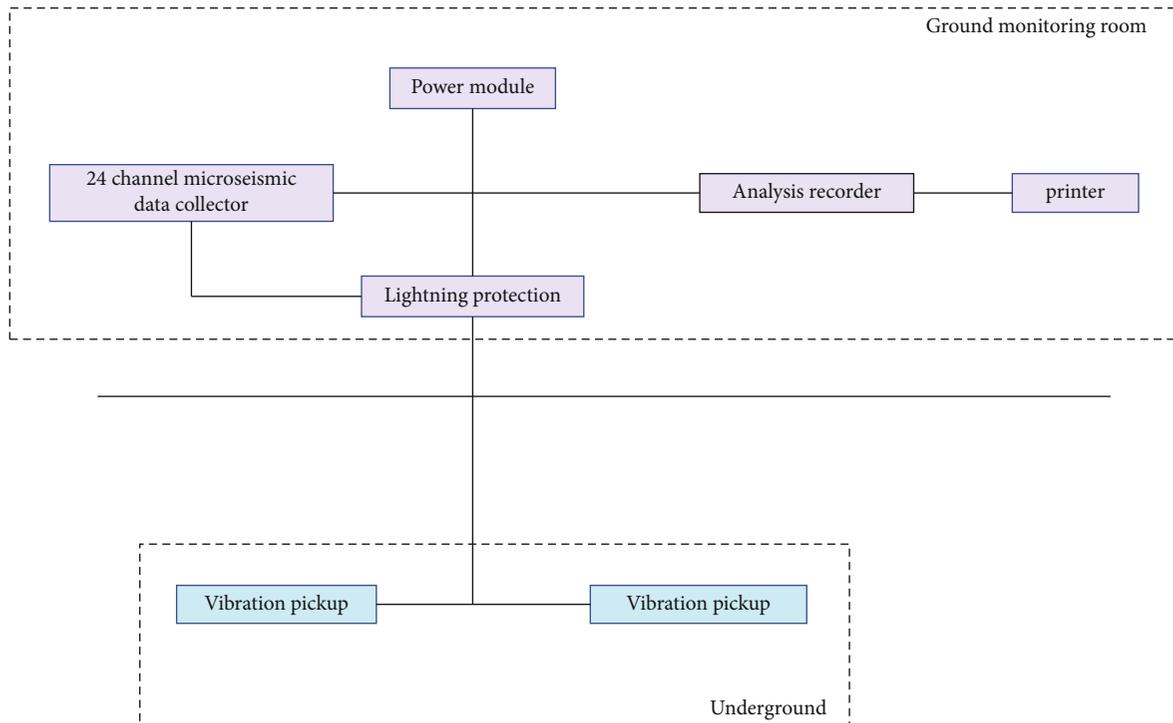


FIGURE 5: Schematic diagram of the microseismic monitoring system.

in a lot of noise in the signals collected by the microseismic focal point monitoring system. Therefore, it is essential to denoise the collected microseismic signals. However, the traditional denoising methods are not suitable for the nonlinear, nonstationary, and non-Gaussian signals generated by microseisms. Empirical mode decomposition (EMD) can process such a signal because of its adaptive nature, but its disadvantage is that there will be mode aliasing. Thereby, the EMD algorithm is improved to EEMD, which has all the advantages of the EMD algorithm. EEMD can decompose the collected signal according to its own characteristics without setting the basis function [28, 29]. Figure 8 is the algorithm flow chart.

2.4.1. Selection of EEMD Algorithm Parameters. The parameters of the EEMD algorithm mainly include the amplitude of white noise and the average number of integration. The amplitude of white noise is mainly selected by experience. The data difference may affect the reliability of the results,

and the sensitivity of signals with different frequencies to noise is also different. The signal fluctuation will become stronger when the noise is too large. At this time, it is beneficial to decompose the high-frequency signal, but it will destroy the arrangement of extreme points of low-frequency components, resulting in mode aliasing; however, when the added noise is small, mode aliasing of low-frequency signals will be prevented. Since there is a wide frequency band in the high-frequency signal, the noise does not have a strong pulling ability for the high-frequency signal, resulting in multiple high-frequency signals mixed in multiple IMF, and mode aliasing also occurs. The standard deviation of the added noise amplitude is set to 0.2 according to relevant research. The amplitude shall be appropriately reduced when there are many high-frequency signals in the processed signals. Otherwise, the amplitude shall be increased.

Relevant literature review shows that when the noise is decomposed based on the EEMD algorithm, the average

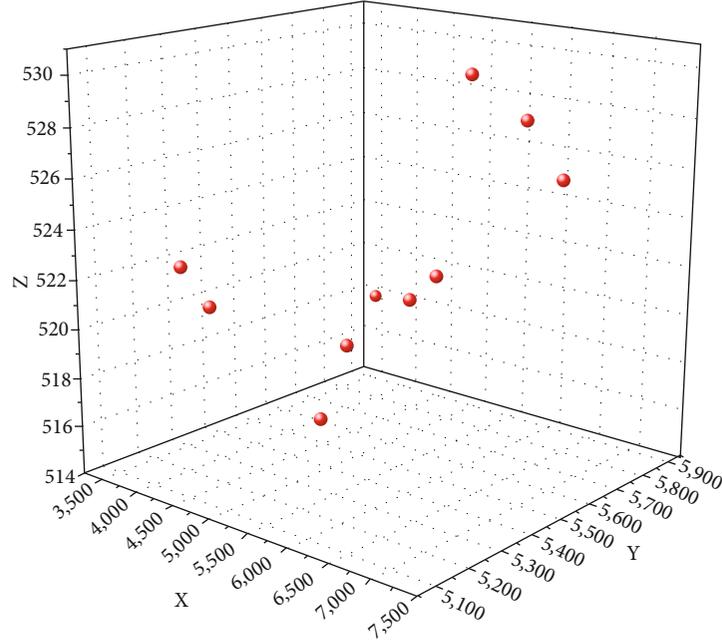


FIGURE 6: Coordinate diagram of geophones.

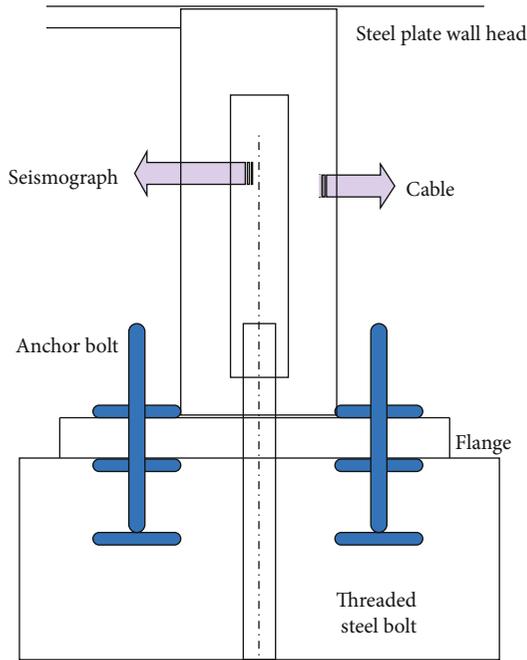


FIGURE 7: Protection device diagram of geophone sensor.

number of integrations shall meet

$$\varepsilon = \frac{\alpha}{\sqrt{N}}, \quad (23)$$

where N is average times of integration, α is the amplitude standard deviation ratio between the added white noise and the signal to be processed, and ε is maximum relative error. After massive experiments, relevant scholars believe

that the value range of 100-300 can give full play to the advantages of the EEMD algorithm.

2.4.2. Wavelet Threshold Denoising Based on EEMD. The high-frequency IMF is not directly proposed but denoised by the wavelet threshold algorithm based on EEMD to retain some effective information in the high-frequency signal [30, 31]. Before denoising the high-frequency signal, the dividing point of high-frequency and low-frequency IMF needs to be determined by using the continuous mean square error criterion. First, the reconstructed signal is set by

$$\tilde{y}_k = \sum_{j=k}^n \text{IMF}_j(t) + r_n(t), \quad k = 1, 2, \dots, n-1, \quad (24)$$

where $\text{IMF}_j(t)$ is the j th order IMF component after denoising, N is the total IMF component, and $r_n(t)$ is the residual IMF component.

Equation (25) can be obtained according to the continuous mean square error criterion:

$$\text{CMSE}(\tilde{y}_k, \tilde{y}_{k+1}) = \frac{1}{N} \sum_{i=1}^N [\tilde{y}_k(t_i) - \tilde{y}_{k+1}(t_i)]^2 = \frac{1}{N} \sum_{i=1}^N [\text{IMF}_k(t_i)]^2. \quad (25)$$

In (25), N is signal length and $\text{IMF}_j(t_i)$ is the k th order IMF component after denoising.

The minimum value of the objective function is obtained through equation (26) to obtain the dividing point of high-

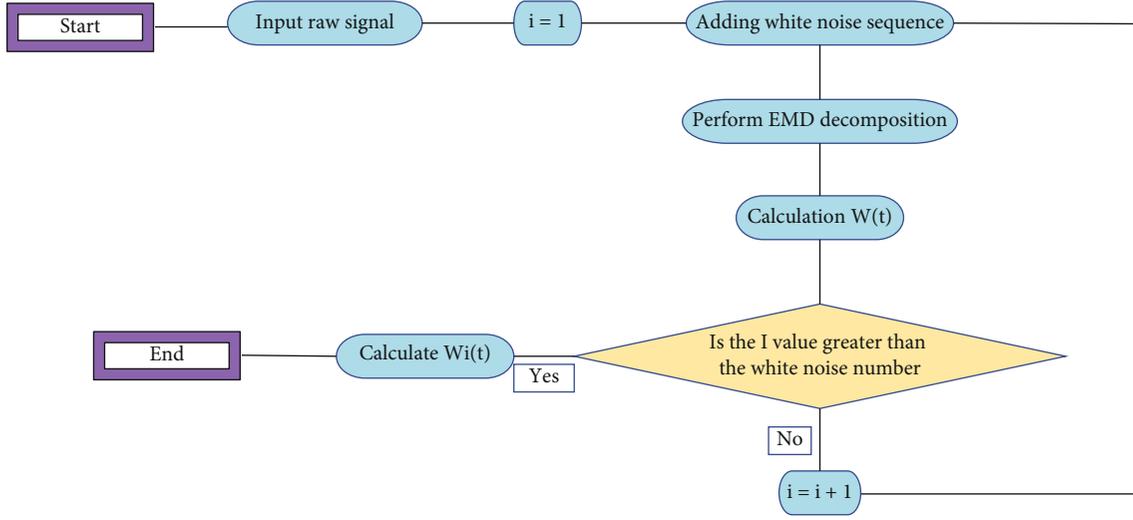


FIGURE 8: Flow chart of EEMD algorithm.

and low-frequency IMF:

$$j_s = \arg \min_{1 \leq k \leq n-1} [\text{CMSE}(\tilde{y}_k, \tilde{y}_{k+1})]. \quad (26)$$

The IMF component before the dividing point, that is, $\text{IMF}'_j (j = 1, 2, \dots, m)$, is obtained after denoising the obtained dividing point and the high-frequency IMF before the dividing point by the wavelet threshold algorithm. Then, the remaining low-frequency components, $\text{IMF}'_j (j = 1, 2, \dots, m)$, and res are reconstructed to obtain the denoised signal, expressed as

$$x'(t) = \sum_{j=1}^m \text{IMF}'_j(t) + \sum_{j=4}^n \text{IMF}_j(t) + r_n(t). \quad (27)$$

Figure 9 is the algorithm flow chart of denoising based on the EEMD wavelet threshold.

2.5. The Framework of Microseismic Search Engine

2.5.1. Theoretical Framework. The microseismic search engine can accurately and quickly determine the microseismic focal point location. The signal waveform generated by microseism changes with time, which is similar to sound signals [32]. In the determination of the focal point position, it is assumed that a waveform very similar to the input signal is found in the waveform database; then, the found waveform microseismic location is the place of this microseismic occurrence. Figure 10 shows the steps of microseismic search.

2.5.2. Fast Retrieval Algorithm. Finding the waveform similar to the input microseismic waveform from the microseismic database is a crucial step of the microseismic search engine. In short, it is to search for the similarity of the microseismic waveform. At present, the local sensitive hash algorithm has been applied in image and audio retrieval, text retrieval, and

other aspects. It can effectively detect high-dimensional data with a short time and high accuracy.

Based on the local sensitive hash function, the local sensitive hash algorithm defines the function mapped from S^n to B set as F . Any hash function f is selected from the hash function family, and any two points a and b are analyzed to make $f(a) = f(b)$.

It is assumed that points a and b meet the conditions in

$$\begin{aligned} \text{If } \|a - b\| \leq S, \text{ then } \Pr_F[f(a) = f(b)] &\geq P_1, \\ \text{If } \|a - b\| \leq cS, \text{ then } \Pr_F[f(a) = f(b)] &\leq P_2. \end{aligned} \quad (28)$$

In general, $c > 1, P_1 > P_2$.

2.6. The Local Sensitive Hash Algorithm Based on P Stable Distribution

2.6.1. P Stable Distribution. U is assumed to be a distribution in the set of real numbers R . If there is an integer not less than 0, any real number S_1, S_2, \dots, S_n and random variable T_1, T_2, \dots, T_n are independently and identically distributed in U , the distributions of variables $\sum_i s_i T_i$ and $(\sum_i |s_i|^p)^{1/p} T$ are consistent, and T is a random variable subject to U . U satisfies P stable distribution.

When P values are 1 and 2, they are 1 stable distribution and 2 stable distribution, respectively. Equation (29) is its probability density function:

$$\begin{aligned} c(x) &= \frac{1}{\pi} \frac{1}{1+x^2}, \\ g(x) &= \frac{1}{\sqrt{2\pi}} e^{-x^2/2}. \end{aligned} \quad (29)$$

2.6.2. Algorithm Description. P stable distribution first needs to define the hash function, as shown in

$$h(m) = \frac{s \cdot m + K}{W}, \quad (30)$$

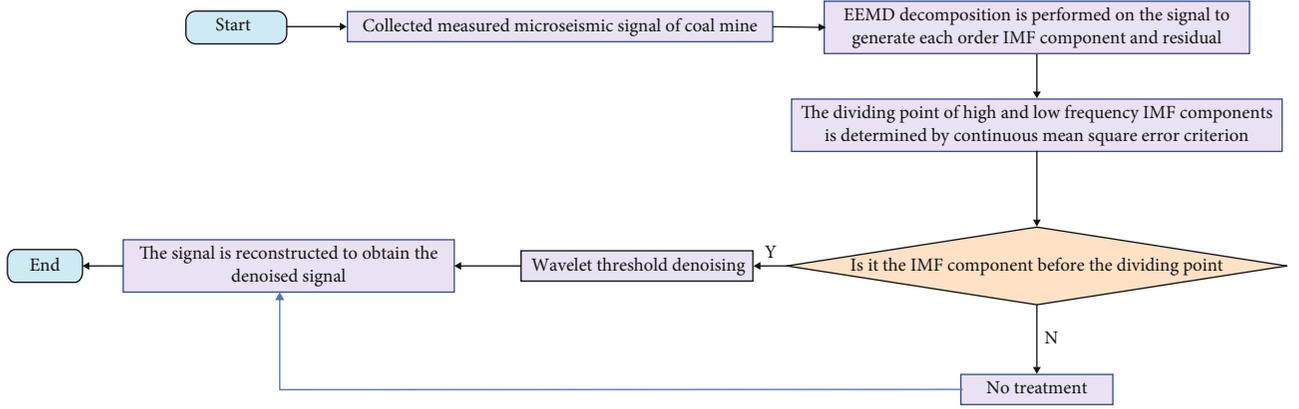


FIGURE 9: Flow chart of the denoising process.

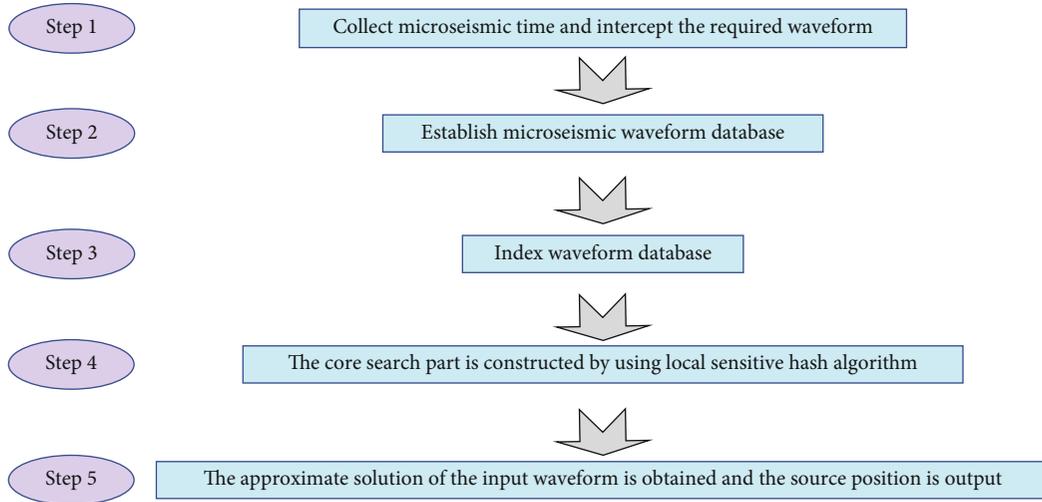


FIGURE 10: Steps of microseismic search engine.

where s is the vector, subject to standard normal distribution; m is the vector formed by dataset; K is a random number; and W is segment length on the straight line.

For any two vectors m_1 and m_2 , if the value of $\|m_1 - m_2\|$ is small, the two vectors are very likely to collide. If the distance between two vectors is far, the two vectors will not collide. The distance between two high-dimensional data can be kept unchanged by equally dividing and labeling the real number axis.

k functions are selected from the hash function h_1, h_2, \dots, h_n to form the function $g(x)$, expressed as

$$g(x) = (h_1(x), h_2(x), \dots, h_k(x)). \quad (31)$$

L functions $g(x)$ form $g_1(x), \dots, g_L(x)$. Any function can map a dataset to a hash table. The vector $S_i = (S_1, S_2, \dots, S_d)$ ($i = 1, 2, \dots, k$) is used as a projection vector to map the point m_i in the dataset to the hash bucket $g(m_i) = (h_1(m_i), h_2(m_i), \dots, h_k(m_i))$. The projection values of points in the same hash bucket are the same. The specific process of the algorithm based on P stable distribution is as follows.

L, k, K, W, m and dataset dimension d have been given. $L * kd$ -dimensional vectors $L * k$ are randomly generated

to map the dataset, and equation (32) is obtained:

$$\begin{aligned} g_1(m_i) &= (h_{11}(m_i), h_{12}(m_i), \dots, h_{1k}(m_i)) (i = 1, \dots, k) \\ &\dots \\ g_L(m_i) &= (h_{L1}(m_i), h_{L2}(m_i), \dots, h_{Lk}(m_i)) (i = 1, \dots, k). \end{aligned} \quad (32)$$

The query points are mapped to the random space to obtain

$$\{x | h_i(q) = h_i(m), i = 1, 2, \dots, L\}. \quad (33)$$

Approximate points similar to query points are obtained.

2.6.3. Local Sensitive Hash Algorithm Based on Multisearch. Figure 11 shows the steps of the multisearch local sensitive hash algorithm.

Equation (34) is a hash function based on P stable distribution:

$$h_{s_{ij}, K_{ij}}(m) = \left\lfloor \frac{s_{ij} \cdot m + K_{ij}}{W} \right\rfloor. \quad (34)$$

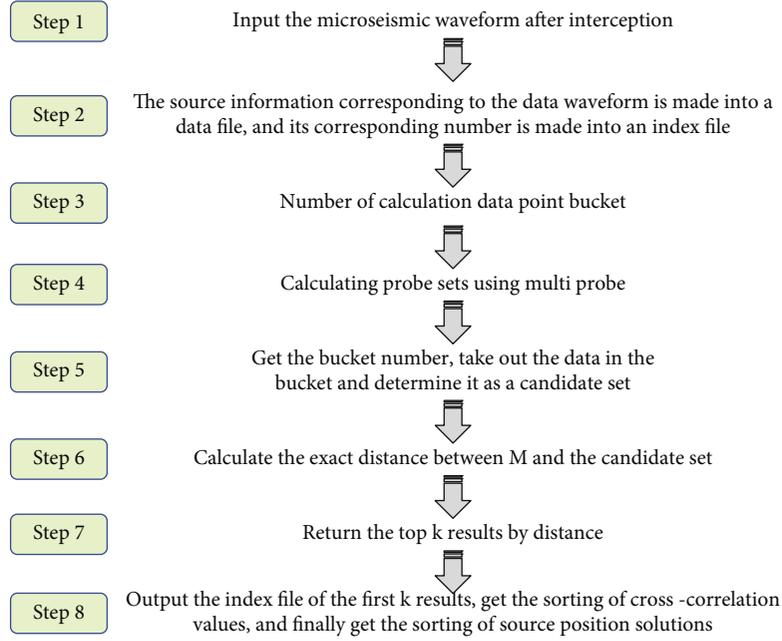


FIGURE 11: Specific steps of microseismic search engine.

This function indicates that m is projected in the direction of S_{ij} and divided into the slot of W width.

2.6.4. *LSH Microseismic Search Engine Model Based on Multisearch.* Figure 11 shows the specific steps of the LSH microseismic search engine based on multisearch.

In this experiment, ten groups of microseismic events are selected as the parameter set of the LSH algorithm. Each set of parameters is tested 10 times, and the average value is calculated to ensure the accuracy of the parameter results. When the search engine algorithm is used to query the input data, the query results must meet the corresponding conditions. The two indicators of recall and precision can judge the quality of query results. The calculation equations of recall rate and precision read

$$\text{recall} = \frac{|A(a) \cap T(a)|}{|A(a)|}, \quad (35)$$

$$\text{precision} = \frac{|A(a) \cap T(a)|}{|T(a)|}, \quad (36)$$

where a is the query point, $T(a)$ is the set of query results, and $A(a)$ is the set of results satisfying the conditions.

The performance of the LSH algorithm largely depends on three parameters: dimension k of hash function family, number L of hash tables, and interval size L . Therefore, the above three parameters need to be determined first. Equation (37) displays the probability of collision between the two vectors after hash:

$$p(u) = \Pr_{a,b}[h_{a,b}(p) = h_{a,b}(q)] = \int_0^w \frac{1}{u} f_s\left(\frac{t}{u}\right) \left(1 - \frac{t}{w}\right) dt. \quad (37)$$

k functions are selected optionally from the hash function family and mapped, represented by

$$g(x) = (h_1(x), \dots, h_k(x)). \quad (38)$$

L hash functions are selected optionally from the k -dimensional hash function family. Equation (39) is the probability of collision of vectors after projection:

$$q(u, k, L) = 1 - [1 - p(u)^k]^L. \quad (39)$$

The purpose of the waveform query is to obtain all waveforms with a Euclidean distance less than the query range u_0 from the database. After the collision probability and density function of the vector are obtained, the calculation equations of recall rate and precision read

$$R(k, L) = \frac{\int_0^{N_0} f(u) q(u, k, L) du}{\int_0^{N_0} f(u) du}, \quad (40)$$

$$P(k, L) = \frac{\int_0^{N_0} f(u) q(u, k, L) du}{\int_0^{\infty} f(u) q(u, k, L) du}. \quad (41)$$

An important index to evaluate the retrieval process is the retrieval time. While ensuring the recall rate and precision, the shorter the time spent in the retrieval process is, the better the retrieval efficiency is. When the three parameters k , L , and W remain unchanged, the number of hash tables will exert a certain impact on the retrieval time, so the selected number of hash tables L should be appropriate.

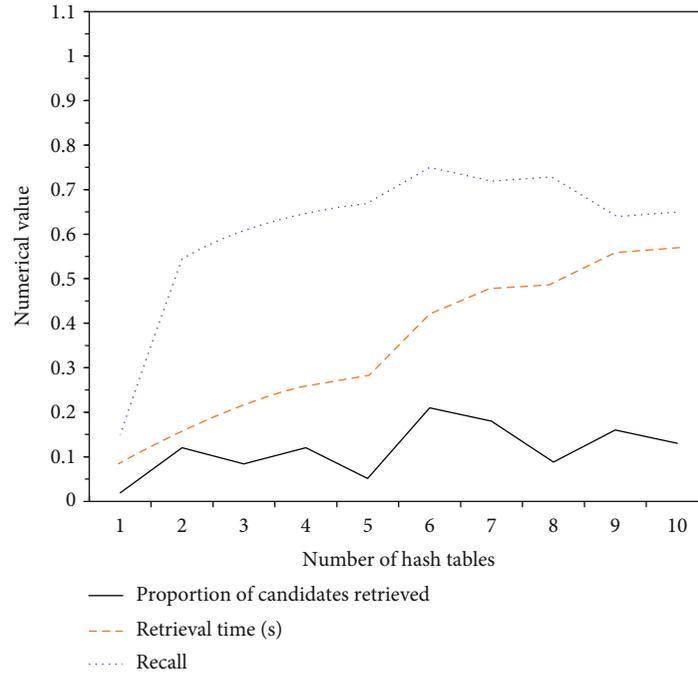


FIGURE 12: Operation diagram of the selection results of the hash table.

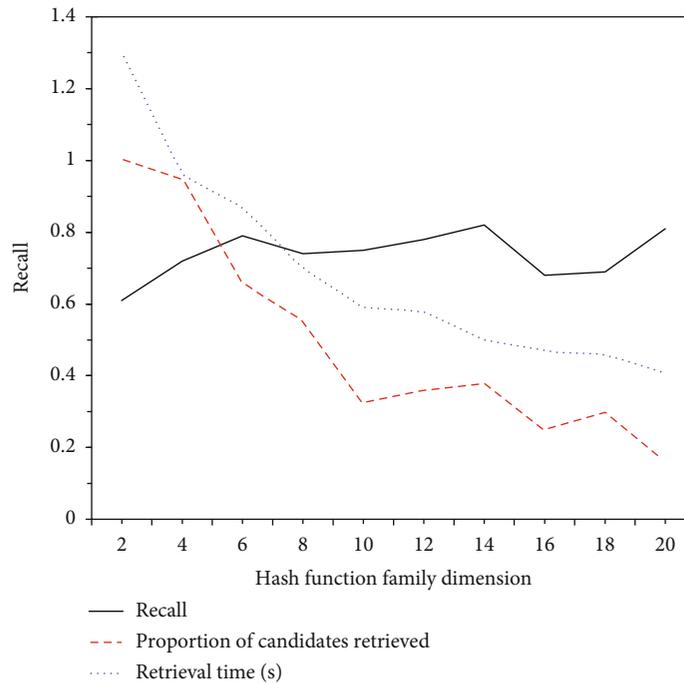


FIGURE 13: Operation diagram of dimension selection results of hash function family.

3. Results

3.1. *Parameter Analysis of LSH Microseismic Search Engine Model Based on Multisearch.* The number of hash tables L , dimension k of hash function family, and interval size are analyzed in detail.

The number of hash tables will have a certain impact on the search accuracy and time. The larger the value is,

the longer the search time is. In the case of limited memory, the value is generally 1-10. The dimension of the hash bucket is closely related to the dispersion of the dataset after projection. The larger the k value is, the fewer the points that enter the same hash bucket through projection are. The value is generally 2-20. The interval size is determined by the module length of the data point. The larger the W value is, the more concentrated the points in the

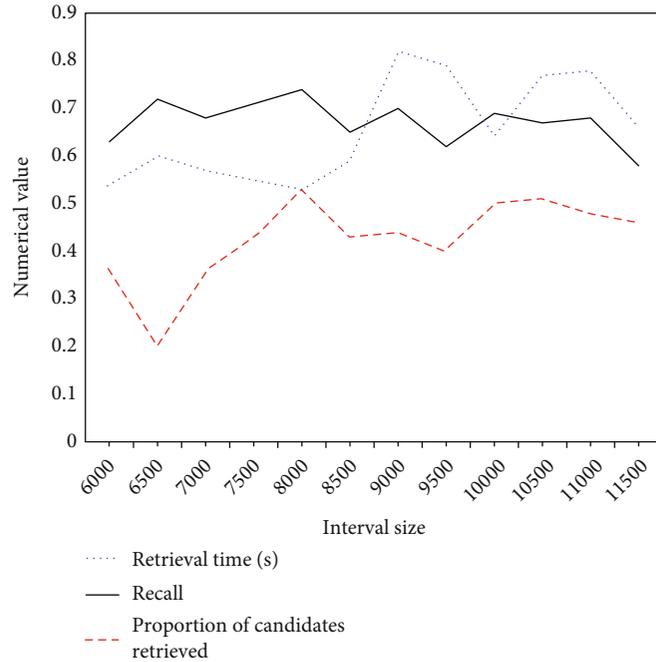


FIGURE 14: Operation diagram of interval size selection results.

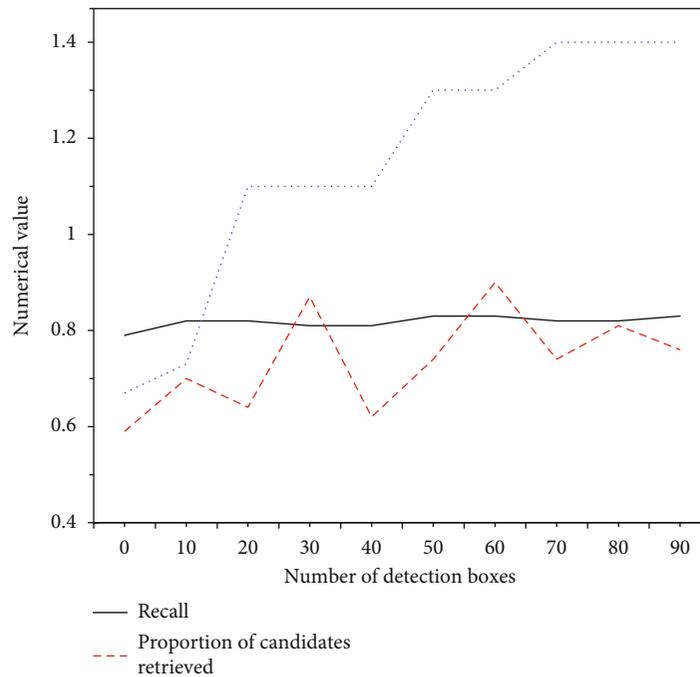


FIGURE 15: Operation diagram of T selection results.

dataset after projection are, and the longer the search time is. The value of W needs to be determined through multiple experiments.

The number of hash tables is set to 1-10. All hash tables are tested 10 times, and the average value is calculated to obtain the recall rate. Figure 12 is a curve of the time required for searching and the proportion of candidates retrieved with the number of hash tables.

Figure 12 displays that the more the hash tables are, the longer the retrieval time is. Overall, when the number of hash tables is 6, the proportion of candidates retrieved, retrieval time, and recall rate are better. The hash function family dimension is set to 2-20. All hash function group dimensions are tested 10 times, and the average value is calculated to obtain the recall rate. Figure 13 is the change curve of the time required for searching and the

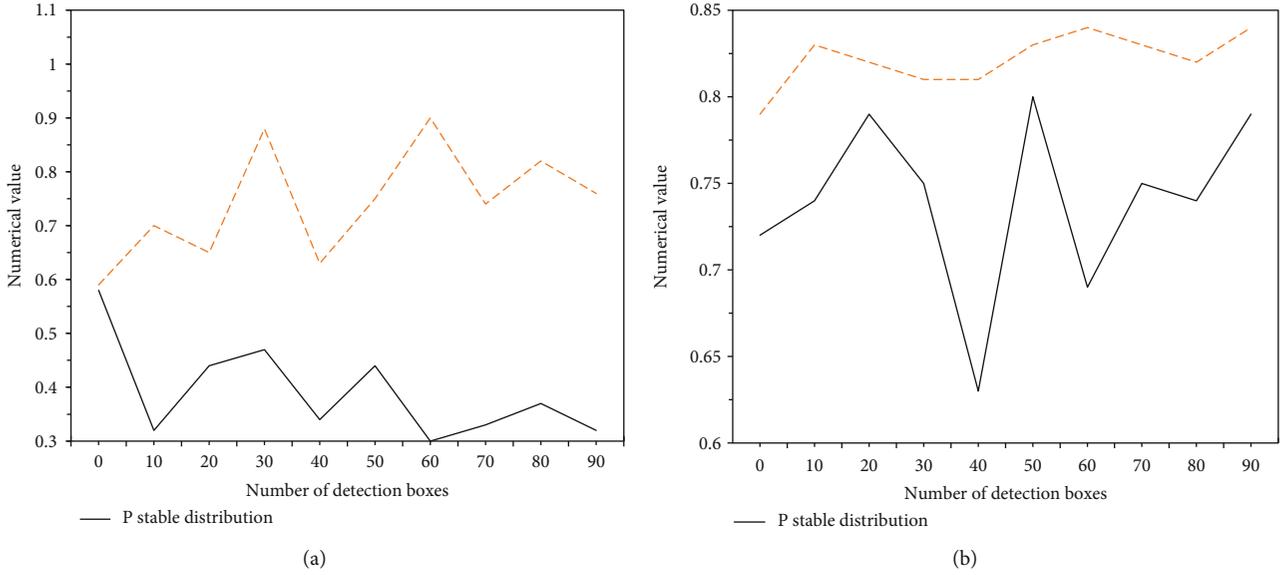


FIGURE 16: Comparison of operation results of multiprobe LSH and P stable distribution: (a) comparison of the proportion of candidates retrieved; (b) comparison of recall rate.

proportion of candidates retrieved with the dimension of the hash function family.

Figure 13 shows that the larger the dimension of the hash function family is, the shorter the time required for data search is. The comprehensive comparison of the above data shows that when the hash function family dimension is 14, the proportion of candidates retrieved, retrieval time, and recall rate are better. The value range of interval size W is set to 6000-11500. All interval sizes are tested 10 times, and the average value is calculated to obtain the recall rate. Figure 14 is a change curve of the time required for searching and the proportion of candidates retrieved with the interval size.

Figure 14 shows that when the interval size is 8000, the proportion of candidates retrieved, retrieval time, and recall rate are better. The above analysis suggests that when the number of hash tables, hash function family dimension, and interval size are 6, 14, and 8000, respectively, the time spent in data retrieval is short, the recall rate is high, and the proportion of candidates retrieved is large.

Next, when the number of detection boxes is 1-90, the changes of recall rate, the time required to search, and the proportion of candidates retrieved with the number of detection boxes are analyzed. Figure 15 shows the results.

Figure 15 shows that it will take a longer time to retrieve the data when the number of detection boxes increases. When the number of detection boxes is 10, the retrieval time is 0.73021 s. When the number of detection boxes increases to 20, the retrieval time has been greater than 1 s, so it is decided that the number of detection boxes is 10. At this time, the recall rate and the proportion of search candidates are 0.825 and 0.6967, respectively.

The changes in recall rate and the proportion of retrieved candidates with the number of detection boxes

based on P stable distribution and multiprobe LSH are analyzed. Figure 16 shows the results.

Figure 16 shows that the proportion of candidates retrieved by multiprobe LSH is significantly better than that of P stable distribution, and the proportion of candidates retrieved is about twice higher than that of P stable distribution; multiprobe LSH also has obvious advantages on the recall rate.

3.2. Parameter Analysis of Search Engine Algorithm for Measured Coal Mine Microseismic Data. Because there are some differences between the experimental microseismic database and the measured microseismic database, it is still necessary to determine the parameter value of LSH before searching the measured coal mine microseismic data.

The training set selected contains 296 groups of microseismic events. Ten groups of microseismic events are selected as the parameter test set of the LSH algorithm. Each group of parameter sets is tested 10 times, and its average value is calculated to ensure the accuracy of parameter results.

Consistent with the method of determining the parameters in the previous section, the number of hash tables, the dimension of the hash function family, and the size of the interval are analyzed. Figure 17 shows the results.

Figure 17 shows that when the search engine algorithm parameters of the measured coal mine microseismic data are selected, the retrieval time, recall rate, and the proportion of retrieved candidates can get better values when the number of hash tables, the dimension of hash function family, and the interval size are 4, 600, and 500, respectively.

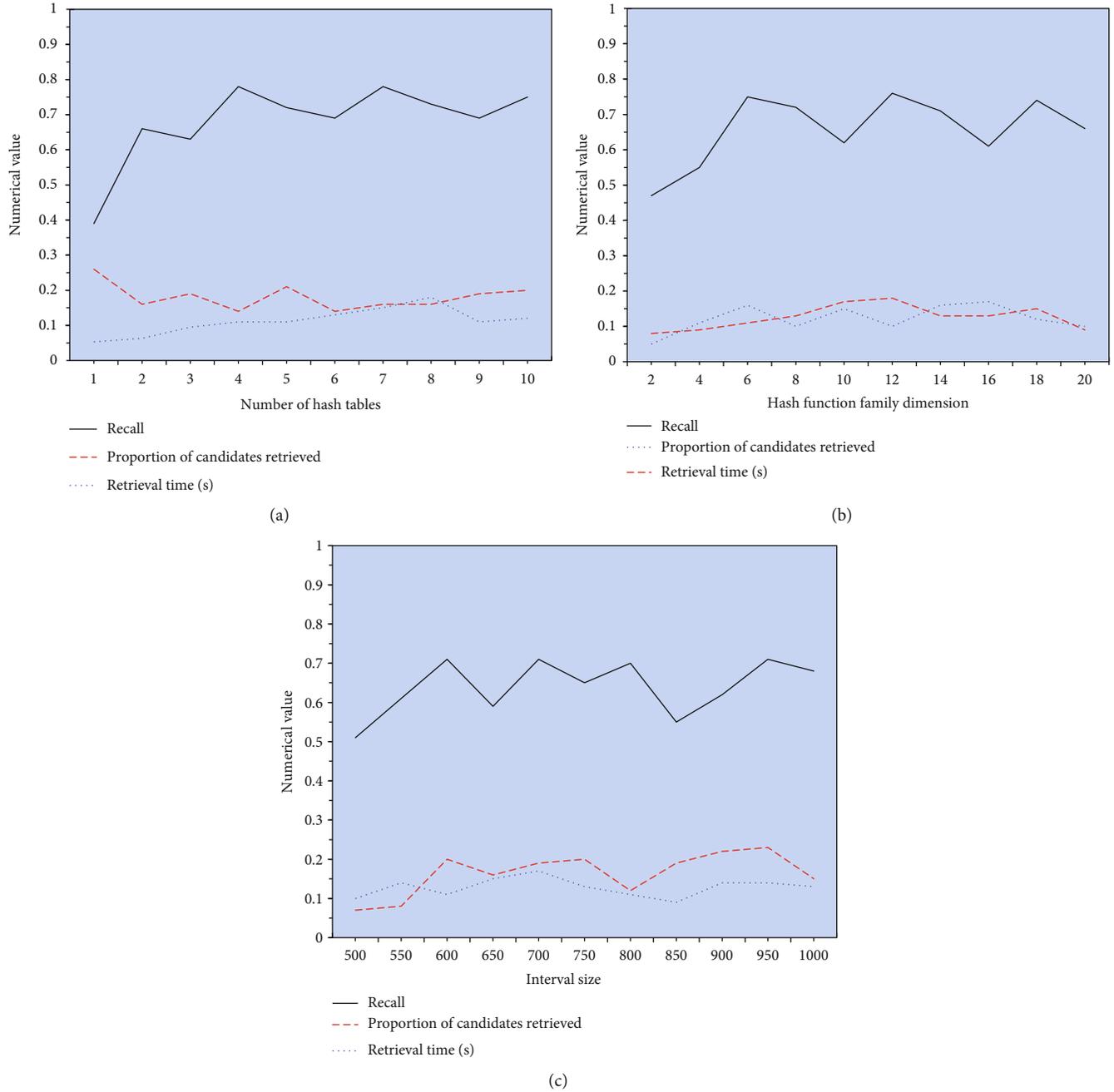


FIGURE 17: Running diagram of results: (a) selection result of the number of hash tables; (b) selection result of the hash function family dimension; (c) selection result of the interval size.

4. Conclusions

As one of the crucial problems in microseismic monitoring technology, the location of the microseismic focal point is the basis of the study of focal point distribution and mechanism. The accurate location of the microseismic focal point is of great significance for mine earthquake monitoring and early warning and impact risk assessment with strong destructiveness such as rockburst. Based on the stress wave method, the signal of the microseismic model is collected, the collected waveform is intercepted, and then, the waveform database is constructed. The search engine location of

the constructed database is conducted based on the LSH of P stable distribution. To solve the problems of inaccurate location information and low recall rate of retrieval results, a microseismic search engine model is constructed. This exploration is aimed at completing the following work. EEMD combined with wavelet denoising is used to denoise the coal mine microseismic signal. Then, a waveform database is established for the denoised waveform data and focal point location. The structure and mathematical model of LSH based on P stable distribution are introduced and improved, and the optimized algorithm, multiprobe LSH is obtained. The microseismic positioning model is established

according to the characteristics of microseismic data. The effects of the number of hash tables, the dimension of hash function family, and interval size on the LSH microseismic positioning model are discussed, and the above parameters are determined. The results show that the microseismic search engine based on multiprobe LSH has higher recall and more accurate positioning results.

Although LSH has achieved good results in the fast and accurate positioning of the microseismic search engine, many deficiencies still need to be further improved. (1) Due to the limited experimental conditions, there are fewer experimental models and fewer measured microseismic data of coal mines. More microseismic waveforms will be added to the database, so that LSH has greater applicability in the focal point location. (2) LSH parameters include the number of hash tables, hash function family dimension, and interval size. The parameters here are determined by multiple experiments. In future research work, appropriate algorithms can be used to adaptively adjust these parameters to improve the generalization ability of the model. (3) Although the application of data-oriented multiprobe LSH in microseismic search engines has achieved good results, the dataset information is not fully used. Therefore, in future research, it is necessary to increase the capacity of the dataset, use multiple PCA, make full use of the dataset information, and maintain the positioning accuracy.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] W. U. Zhao, "Research on the application of internet of things technology to digital museum construction," *Acta Geoscientica Sinica*, vol. 38, no. 2, pp. 293–298, 2017.
- [2] Y. Chen, S. Zhao, and Y. Zhou, "Research on intelligent agricultural planting system based on internet of things technology," *Journal of Computer and Communications*, vol. 6, no. 6, pp. 54–60, 2018.
- [3] A. Barbu, Y. She, L. Ding, and G. Gramajo, "Feature selection with annealing for computer vision and big data learning," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 39, no. 2, pp. 272–286, 2017.
- [4] A. Ioannidou, E. Chatzilari, S. Nikolopoulos, and I. Kompatsiaris, "Deep learning advances in computer vision with 3D data," *ACM Computing Surveys*, vol. 50, no. 2, p. 1, 2017.
- [5] J. Wldchen and P. Mder, "Plant species identification using computer vision techniques: a systematic literature review," *Archives of Computational Methods in Engineering*, vol. 25, no. 2, pp. 507–543, 2018.
- [6] W. Shao, R. Xiao, J. Huang, H. Liu, and X. du, "FJLT-FLSH: more efficient fly locality-sensitive hashing algorithm via FJLT for WMSN IoT search," *Internet of Things Journal*, vol. 6, no. 4, pp. 7122–7136, 2019.
- [7] X. Zhang, C. Wei, and Z. Song, "Fast locally weighted PLS modeling for large-scale industrial processes," *Industrial & Engineering Chemistry Research*, vol. 59, no. 47, pp. 20779–20786, 2020.
- [8] H. Zhu, B. Liang, and T. Zhang, "Fast access for star catalog based on locality-sensitive hashing," *Xibei Gongye Daxue Xuebao/Journal of Northwestern Polytechnical University*, vol. 36, no. 5, pp. 988–994, 2018.
- [9] H. Zhang, C. Ma, V. Pazzi, T. Li, and N. Casagli, "Deep convolutional neural network for microseismic signal detection and classification," *Pure and Applied Geophysics*, vol. 177, no. 12, pp. 5781–5797, 2020.
- [10] J. Akram, "An application of waveform denoising for microseismic data using polarization–linearity and time–frequency thresholding," *Geophysical Prospecting*, vol. 66, no. 5, pp. 872–893, 2018.
- [11] B. Li, N. Li, E. Wang, X. Li, Y. Niu, and X. Zhang, "Characteristics of coal mining microseismic and blasting signals at Qianqiu coal mine," *Environmental Earth Sciences*, vol. 76, no. 21, 2017.
- [12] A. Mb, B. Bh, and C. Rm, "Coal mine roof rating (CMRR), rock mass rating (RMR) and strata control: Carborough Downs Mine, Bowen Basin, Australia," *International Journal of Mining Science and Technology*, vol. 30, no. 2, pp. 225–234, 2020.
- [13] D. S. Yantek, L. Yan, P. T. Bissert, and M. D. Klein, "Effects of mine strata thermal behavior and mine initial temperatures on mobile refuge alternative temperature," *Mining Engineering*, vol. 69, no. 4, pp. 41–48, 2017.
- [14] D. Zhao and W. Qiang, "An approach to predict the height of fractured water-conducting zone of coal roof strata using random forest regression," *Scientific Reports*, vol. 8, no. 1, 2018.
- [15] P. Antolin, D. Schmit, T. Pereira, B. de Pontieu, and I. de Moortel, "Transverse wave induced Kelvin-Helmholtz rolls in spicules," *Astrophysical Journal*, vol. 856, no. 1, p. 44, 2018.
- [16] X. Ji, Y. Fan, Q. Xiao, and P. Cai, "Passive wireless measurement of tension for overhead transmission line based on surface transverse wave," *Iet Generation Transmission & Distribution*, vol. 12, no. 8, pp. 1866–1871, 2017.
- [17] X. Song, "Attosecond interference induced by Coulomb-field-driven transverse backward-scattering electron wave packets," *Physical Review A*, vol. 95, no. 3, article 033426, 2017.
- [18] M. Tannarana, P. Pataniya, G. K. Solanki, C. K. Zankat, K. D. Patel, and V. M. Pathak, "Humidity sensor based on Sb_{0.1}Sn_{0.9}Se₂ ternary alloy for human breath monitoring and touchless positioning interface," *The European Physical Journal Plus*, vol. 134, no. 5, pp. 1–5, 2019.
- [19] S. A. Okawa, "Displacement monitoring performance of relative positioning and precise point positioning (PPP) methods using simulation apparatus," *Advances in Space Research*, vol. 63, no. 5, pp. 1697–1707, 2019.
- [20] D. Lyu, F. Zeng, X. Ouyang, and H. Zhang, "Real-time clock comparison and monitoring with multi-GNSS precise point positioning: GPS, GLONASS and Galileo," *Advances in Space Research*, vol. 65, no. 1, pp. 560–571, 2020.
- [21] L. Malheiros, D. Georges, and L. X. Cardoso, "Fall detection system and body positioning with heart rate monitoring," *IEEE Latin America Transactions*, vol. 15, no. 6, pp. 1021–1026, 2017.

- [22] D. Salookolaei and S. H. Nasser, "A dual simplex method for grey linear programming problems based on duality results," *Grey Systems: Theory and Application*, vol. 10, no. 2, pp. 145–157, 2020.
- [23] K. Okawa, "Self-localization estimation for mobile robot based on map-matching using downhill simplex method," *Journal of Robotics and Mechatronics*, vol. 31, no. 2, pp. 212–220, 2019.
- [24] M. Hammoudeh, "A wireless sensor network border monitoring system: deployment issues and routing protocols," *IEEE Sensors Journal*, vol. 17, no. 8, pp. 2572–2582, 2017.
- [25] J. Ni, L. Yao, J. Zhang, W. Cao, Y. Zhu, and X. Tai, "Development of an unmanned aerial vehicle-borne crop-growth monitoring system," *Sensors*, vol. 17, no. 3, p. 502, 2017.
- [26] S. R. Madeti and S. N. Singh, "Monitoring system for photovoltaic plants: a review," *Renewable and Sustainable Energy Reviews*, vol. 67, pp. 1180–1207, 2017.
- [27] H. Morishima, J. Washio, J. Kitamura, Y. Shinohara, T. Takahashi, and N. Takahashi, "Real-time monitoring system for evaluating the acid-producing activity of oral squamous cell carcinoma cells at different environmental pH," *Scientific Reports*, vol. 7, no. 1, p. 10092, 2017.
- [28] V. Nourani and A. Partoviyan, "Hybrid denoising-jittering data pre-processing approach to enhance multi-step-ahead rainfall-runoff modeling," *Stochastic Environmental Research and Risk Assessment*, vol. 32, no. 2, pp. 545–562, 2018.
- [29] X. Zhong, Y. Dai, Y. Dai, and T. Jin, "Study on processing of wavelet speech denoising in speech recognition system," *International Journal of Speech Technology*, vol. 21, no. 3, pp. 563–569, 2018.
- [30] S. Y. Jung, Y. J. Chyung, and P. W. Kim, "Kernel design for real-time denoising implementation in low-resolution images," *Journal of Real-Time Image Processing*, vol. 16, no. 1, pp. 31–47, 2019.
- [31] O. T. Grozdi and S. T. Jovii, "Whispered speech recognition using deep denoising autoencoder," *Engineering Applications of Artificial Intelligence*, vol. 59, no. 12, pp. 2313–2322, 2017.
- [32] W. Habib, T. Sarwar, A. M. Siddiqui, and I. Touqir, "Wavelet denoising of multiframe optical coherence tomography data using similarity measures," *IET Image Processing*, vol. 11, no. 1, pp. 64–79, 2017.