SCIENTIFIC OASIS

Journal of Intelligent Decision Making and Granular Computing

Journal homepage: www. jidmag.org
ISSN: 3042-3759

Research on Mine Disaster Risk Monitoring Based on Topic Model Retrieval Technology

Ziyi Wang[1], Gang Cheng[1, 2*], Yaxi Wu[1], Yujie Nie[1], Liu Yang[1], Jingjing Xia[1]

[1] School of Computer Science and Engineering North China Institute of Science and Technology, Beijing 101601, China
[2] School of Earth Sciences and Engineering Nanjing University, Nanjing, Jiangsu 210023, China

ARTICLE INFO

Article history:
Received 11 October 2025
Received in revised form 28 December 2025
Accepted 2 January 2026
Available online 9 January 2026

Keywords:
Topic model retrieval technology; Mine disasters risk monitoring; LDA model

ABSTRACT

As the main energy source in China, coal holds an irreplaceable position in ensuring the country's energy security and supporting economic and social development. However, as the intensity and depth of mining continue to increase, especially in mines with complex geological conditions, improper mining processes and strength design can easily trigger various geological disasters. In the context of the digital intelligence era, the efficient acquisition, rapid management, and precise retrieval of multimodal information provide key technical support for cross-modal information retrieval and visual analysis applications. Based on this, this manuscript proposes a mining disaster risk monitoring method that integrates topic model retrieval technology. First, analysing the retrieval principles and adaptation mechanisms of the topic model in the context of mining engineering disaster scenarios; Secondly, based on topic model retrieval technology, visualising disaster accidents and conducting heat analysis. Thereby, carrying out dynamic regulation and decision-making for mine disaster risks; Finally, based on the above dual research findings, achieving precise monitoring and reliable prediction of various types of mining disaster risks.

1. Introduction

China, the globe's foremost energy producer and user, has for decades treated coal as the dominant fuel in its energy mix. During 2024, China's coal yielded 4.78 billion metric tons of standard coal, accounting for 95.9% of the total primary energy production. During the same period, coal consumption was 3.17 billion tons of standard coal, accounting for 53.2% of total energy consumption, continuing to hold the top position in the energy structure [1]. Therefore, as the main energy source in China, coal remains indispensable for safeguarding the nation's energy security and supporting economic and social development [2, 3]. In the past decade, although China's energy mix has undergone steady upgrading, with the share of coal use progressively declining in total energy consumption has generally shown a slow downward trend. But both raw coal production and

* *Corresponding author.*
E-mail address: chenggang@ncist.edu.cn

https://doi.org/10.31181/jidmgc11202629

consumption have continued to grow [4] (Figure 1). According to relevant statistics, by 2050, coal is projected to continue supplying roughly half of the China's total primary energy demand.

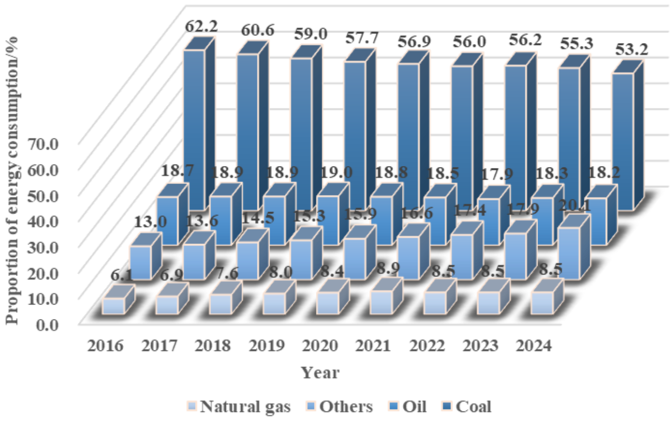


Fig. 1. Structure of energy consumption in China, 2015-2024

As extraction proceeds to greater depths and under higher stress, particularly in geologically intricate conditions, if the mining method and mining intensity are not properly designed, such conditions readily trigger a spectrum of mining-induced geohazards— tunnel deformation, mine water inrush, ground deformation, and induced landslides (Figure 2). The occurrence of the above-mentioned disasters will result in significant casualties and property losses, and pose potential pollution risks to the groundwater environment in the mining area [5], seriously threatening mine safety production and ecological-environmental safeguarding within the extraction zone [6-9]. Thus, comprehensively utilizing new technologies for retrieving multi-source, multi-hazard data and carrying out dynamic monitoring of mine disaster risks is important to securing safe operations in mining areas and protecting the geological environment. It is a crucial demand for China's current mine safety and disaster prevention and mitigation efforts.

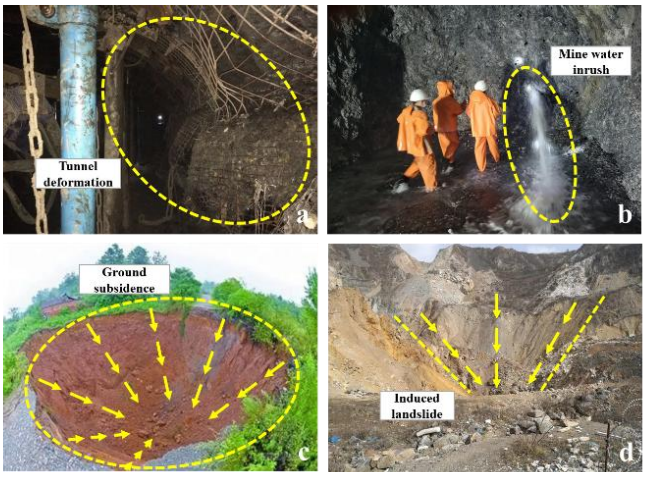


Fig. 2. Geological disasters and underground incidents stemming from overburden deformation

In recent years, as China has tightened oversight of safe mining and rolled out pervasive digital and smart technologies for extraction and hazard control, both the frequency and lethality of mining-related incidents have been markedly curbed. Against the backdrop of the era of digital intelligence, the efficient acquisition, rapid management, and precise retrieval of multimodal information provide key technical support for cross-modal information retrieval and visual analysis applications[10]. It continuously promotes in-depth research on multimodal information retrieval technology by scholars at home and abroad, and expands its application into the study of mine disaster prevention and control. Table 1 provides an overview of the development stages and characteristics of multimodal information retrieval technology.

Table 1
Development Characteristics of Multimodal Information Retrieval Technology [11, 12]

Time nodes	Core features of technological development	Application connections in the field of mining disasters	Key supporting technologies
1960s-1970s	1. At this stage, research on multimodal technology is just beginning, mainly focusing on multi-source data fusion, feature analysis, and similarity retrieval; 2. The scope of research is gradually expanding to the fields of speech recognition and synthesis.	1. The technology is in the early accumulation stage and has not been directly applied to mine disaster prevention and control; 2. The underlying logic of data fusion and feature analysis provides a theoretical reference for the subsequent processing of multi-source mine monitoring data.	1. Traditional signal processing techniques; 2. Early machine learning algorithms.
Early 21st century-2010	1. In this stage, big data and cloud computing technologies have emerged, breaking through the computational bottleneck of large-scale data analysis and complex model training; 2. Multimodal technology has achieved leapfrog development, with the capability to initially handle high-dimensional data.	1. Technology began to extend into the field of engineering; 2. Pilot studies were conducted, starting with the use of the technology for preliminary retrieval of monitoring data for single-type mine disasters such as surrounding rock strain.	1. Distributed cloud computing architecture; 2. Classic machine learning models (SVM, K-Means).
2010-Present	1. At this stage, the technology is becoming increasingly mature, forming a technical system for high-dimensional heterogeneous data adaptation, cross-modal precise matching, and real-time retrieval and analysis; 2. Algorithms such as sparse representation and locality-sensitive hashing are deeply integrated with multimodal technology, improving retrieval efficiency and robustness.	1. With its advantage in adapting to high-dimensional data, it has gained wide attention in the field of mine disaster data retrieval; 2. It is applied to disaster risk monitoring and similar case matching for tunnel deformation, mine water inrush, surface subsidence, and induced landslides, becoming a key technological support for mine disaster prevention and mitigation.	1. Deep learning algorithms; 2. SR-DL; 3. LSH.

Note: Support Vector Machine (SVM), K-Means Clustering Algorithm (K-Means), Sparse Representation and Dictionary Learning Method (SR-DL method), Locality-Sensitive Hashing (LSH)

2. Mine Disaster Data and Retrieval Technology

2.1 Disaster Data

Mine disaster data refers to the collection of structured and unstructured information that can be collected, stored, analyzed, and applied, encompassing the entire process related to the formation, occurrence, development, evolution, and prevention of various mine disasters throughout the entire lifecycle of mine exploration, construction, production, and closure. Through mining, modeling, and validating full-cycle monitoring data of mine disasters, the mechanisms of disaster formation, evolution patterns, and influencing factors can be revealed. This will enhance the precise management of disaster risks, the timeliness of early warnings, and the scientific nature of emergency response, achieving the greatest possible reduction in the occurrence and impact of mining disasters.

Journal of Intelligent Decision Making and Granular Computing
Volume 2, Issue 1 (2026) 1-12

Mining disasters have typical characteristics such as high frequency, multiple types, wide impact, and secondary chain effects. And data on such disasters exhibit multi-dimensional, spatiotemporal, and dynamic evolutionary characteristics, which resulted in the traditional data retrieval paradigm facing application bottlenecks such as 'incomplete search, inaccurate retrieval, and slow response' during use. The above data characteristics indirectly determine that the in-depth processing and value extraction of disaster information require the introduction of interdisciplinary approaches and advanced data retrieval technologies to achieve a high-fidelity representation of disaster scenarios and efficient reuse of data resources. Thereby supporting a synergistic improvement in the accuracy, timeliness, and robustness of mine disaster prediction, early warning, and emergency response. To this end, the manuscript systematically reviews the adaptation principles and application paths of three retrieval methods, which are Convolutional Neural Networks (CNN), LSH, and SR-DL methods in the field of mining disasters, aiming to provide a methodological reference for the efficient retrieval and intelligent application of big data in mining disasters.

2.2 Retrieval Technology

2.2.1 CNN

CNN, as a classic feedforward architecture in deep learning, focuses on convolution operations and can quickly uncover deep patterns or structures in images, text, and other data. And it can enable automatic extraction of key features from large-scale data, which can significantly improve the efficiency and accuracy of data retrieval. Figure 3 shows the development history of convolutional networks.

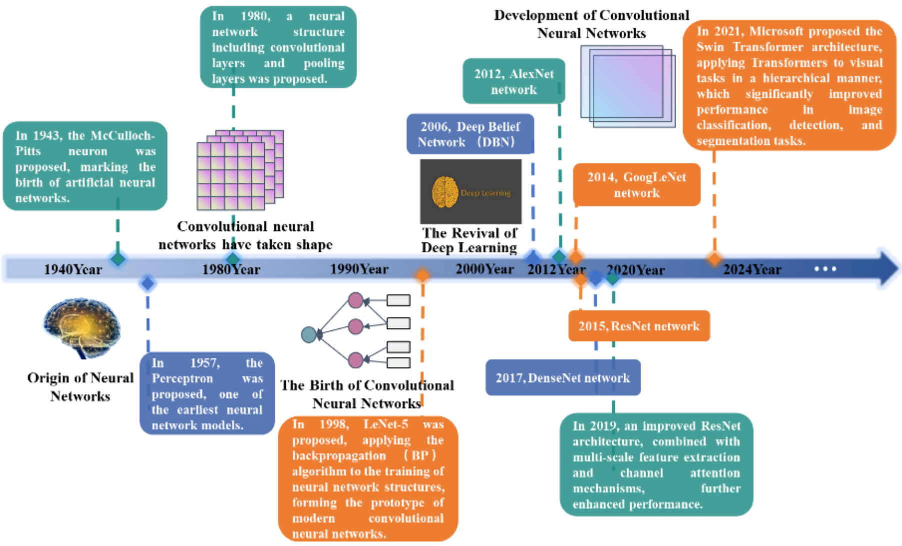


Fig. 3. Development history of convolutional networks

In image retrieval, CNN can accurately recognize different visual modalities such as objects, scenes, and faces. In text retrieval, it can efficiently identify high-level semantic features such as keywords, topics, and syntactic structures. These advantages give it significant strengths in the deep representation of complex heterogeneous data and the mining of hidden correlations. CNN perform local perception and weight-sharing operations on input data through their unique convolutional layers, its classification accuracy higher than most early shallow neural network models. Convolution operation serves as the basis for feature extraction, and its operation formula is: $N = (W - F + 2P)/S + 1$, where W is the input, F is the convolution kernel, P is the padding, and S is the stride. When performing convolution output, the convolution output formula as Eq (1):

$$Y[i,j,k] = \sigma(\sum_{a=0}^{H-1}\sum_{b=0}^{W-1}\sum_{c=0}^{C-1}X[i+a,j+b,c]\cdot W[a,b,c,k]+b[k]) \quad (1)$$

where, $Y[i,j,k]$ is the value at position (i,j) of channel k in the output feature map; σ is the activation function, usually ReLU (Rectified Linear Unit) or another nonlinear activation function; H and W are the height and width of the convolution kernel; C is the number of channels in the input feature map; $X[i+a,j+b,c]$ is the value at position $(i+a,j+b)$ of channel c in the input feature map; $W[a,b,c,k]$ are the parameters of the convolution kernel; $b[k]$ is the bias term. In the convolution operation, after the input feature map undergoes element-wise multiplication and accumulation with the convolution kernel, a nonlinear mapping is completed through an activation function. Thereby, it generates the response values at each spatial position in the output feature map and completes the convolution output.

Based on CNN, feature extraction can be performed, and mapping can be completed in the feature space. This enables CNN to achieve good performance in image retrieval. Therefore, CNN serves a pivotal function in the processing of various types of geological disaster monitoring data. With the exponential increase in disaster data volume, and CNN mostly applied to static spatial data, a single frameworks now fall short of coping with the volatile and intricate requirements posed by disaster-related data processing. Thus, there is an urgent need to carry out data processing and analysis by introducing multimodal fusion methods. Against this background, LSH and SR-DL method have been widely applied in data retrieval and processing due to their ability to handle multimodal data and their simplicity and flexibility.

2.2.2 LSH

The LSH is an advanced optimization of the hashing method. Compared to the hashing method, this approach can effectively ensure that similar data have the same or similar hash values in the hash space. It addresses the challenge of similarity search for high-dimensional mining disaster data, such as distributed fiber optic strain monitoring data. Unlike the 'random mapping' of traditional hashing, for the similarity retrieval needs of distributed fiber optic strain monitoring data in mining disasters, LSH designs specific hash functions such as cosine distance hashing and Euclidean distance hashing [13-15]. This allows disaster data with similar features to be mapped into the identical hash bucket alongside high probability, significantly reducing the number of retrieval comparisons. The implementation process of this method is shown in Figure 4.

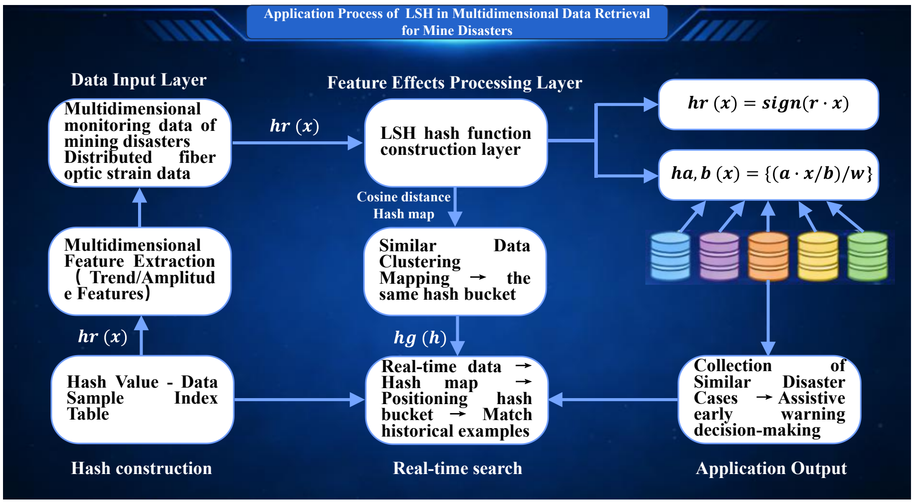


Fig. 4. Application Process of LSH in Multidimensional Data Retrieval for Mine Disasters

Cosine distance (adapted for directional similarity in high-dimensional data) is employed to quantify the angular congruence of a pair of high-dimensional vectors. It is suitable for monitoring scenarios in mine data, such as changes in surrounding rock strain over different time periods, where the 'trends are similar but the magnitudes differ.' For high-dimensional monitoring data vectors $x, y \in R^d$, randomly generate a hyperplane normal vector $r \in R^d$ (with vector elements following the standard normal distribution N (0, 1)), The hash function as Eq (2):

$$h_r(x) = sign(r \cdot x) = \begin{cases} 1, & r.x \geq 0 \\ -1, & r.x < 0 \end{cases} \tag{2}$$

where $r.x$ represents the dot product of vectors, and $sign(\cdot)$ is the sign function.

If the angle between two vectors x, y is θ, then the probability that the vectors map to the same hash value as Eq (3):

$$P_r[h_r(x) = h_r(y)] = 1 - \frac{\theta}{\pi} \tag{3}$$

where the smaller θ is, the more similar the data are, the higher the probability that the hash values will be the same, and the easier it is to map the clustering of similar data.

Therefore, in practical applications, it is possible to quickly match cases similar to real-time monitoring data from a vast amount of historical disaster data, thereby assisting disaster warning decisions.

Euclidean distance (suitable for high-dimensional data with similar numerical values) is used to measure the numerical differences between two high-dimensional vectors. It applies to monitoring scenarios in mining data where measurements in the same area, such as gas concentration monitoring sequences, have 'similar magnitude and trends'. For high-dimensional monitoring data vector $x \in R^d$, construct a linear hash function as Eq (4):

$$h_{a,b}(x) = \left\lfloor \frac{a \cdot x + b}{w} \right\rfloor \tag{4}$$

where, $a \in R^d$ is a random vector following the standard normal distribution N (0, 1); $b \in [0, w]$ is a random offset, following a uniform distribution; w is the bucket width parameter, which needs to be adjusted according to the data distribution; $\lfloor \cdot \rfloor$ denotes the floor function. The smaller the Euclidean distance $\|x - y\|_2$ between vectors x, y, the higher the probability that $h_{a,b}(x) = h_{a,b}(y)$, increasing the probability that akin data entries will be mapped to an identical hash bucket.

2.2.3 SR-DL method

The SR-DL method addresses the inherent characteristics of mining disaster data, which are multidimensional, highly noisy, and multimodally heterogeneous. By constructing a sparse dictionary that is deeply adapted to the disaster data, it achieves efficient compression and accurate retrieval of the target data. Given that mine monitoring data generally contains sensor noise and environmental interference signals, sparse representation can significantly improve the robustness of the retrieval process by eliminating redundant noise and preserving the core features of disasters. At the same time, it can effectively integrate distributed fiber optic strain monitoring data, formation lithology parameters, and in-situ stress distribution data to construct a multimodal joint sparse dictionary, thereby addressing the challenge of unified retrieval of heterogeneous data. In addition, this method can simultaneously perform disaster information feature extraction while achieving data dimensionality reduction and compression, without the need for additional dimensionality reduction preprocessing steps. Thereby it can boost system-wide data-processing performance substantially. The implementation process of the method is shown in Figure 5.

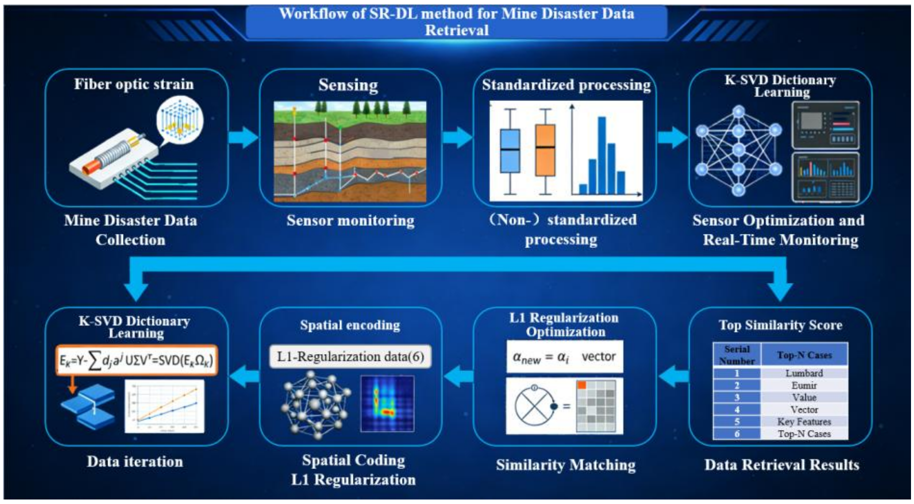


Fig. 5. Workflow of SR-DL method for Mine Disaster Data Retrieval

When using this method in practical monitoring, historical disaster data is usually used as training samples. Through classical dictionary learning algorithms such as K-SVD, an overcomplete sparse dictionary is trained, where the atoms in the dictionary correspond to typical features of mining disaster data. For example, the sudden change characteristics of strain before the roof collapse and the concentration fluctuation characteristics of gas outbursts, the K-SVD dictionary update iteration as Eq (5):

$$E_k = Y - \sum_{j\neq k} d_j a^j U\Sigma V^T = SVD(E_k\Omega_k) d_k = u_1,\quad a^k = \sigma_1 v_1^T \tag{5}$$

where, SVD stands for Singular Value Decomposition; E_k is the reconstruction error matrix after removing the k atom; Ω_k is the index matrix of nonzero coefficients; d_k is the dictionary atom; u_1, σ_1, v_1 are the left singular vector, the largest singular value, and the right singular vector from the SVD, respectively; a^k is the row vector of encoding coefficients corresponding to the dictionary atom.

For the retrieved fiber optic strain monitoring data $y \in R^d$, it is transformed into a coded vector containing only a few non-zero coefficients through sparse representation on a predefined dictionary—the non-zero coefficients correspond to atoms. By further calculating the similarity between this encoded vector and the encoded vectors of historical disaster data, rapid matching of similar historical disaster cases can be achieved. This provides an efficient solution for mine disaster retrieval tasks. The optimization objective of sparse coding (L1 regularization, noise resistance) is given by Eq (6):

$$min_\alpha \|\alpha\|_1 s.t. \quad \|y - D\alpha\|_2^2 \leq \varepsilon \tag{6}$$

where, $\|\alpha\|_1$ is the L1 norm, which promotes sparsity in the coding vector; ε is the reconstruction error tolerance, used to filter out interference caused by sensor noise.

The similarity matching as Eq (7):

$$sim(\alpha_{new},\ \alpha_i) = \frac{\alpha_{new}\cdot\alpha_i}{\|\alpha_{new}\|_2\|\alpha_i\|_2} \tag{7}$$

where, α_{new} is the encoding vector of the new data; $\{\alpha_i\}_{i=1}^n$ is the set of encoding vectors of historical disaster data. The closer the calculated similarity value is to 1, the greater the alignment between the disaster profiles of the two datasets.

3. Case Analysis of Mine Disaster Incidents

The particularly serious collapse accident of the Alxa Xinjing open-pit coal mine in Inner Mongolia Autonomous Region of China occurred on February 22, 2023 (Figure 6). The investigation results of the accident showed that in the early morning of the day of the accident, a small area of landslide had occurred on the west side and the top of the accident location. And it generates precursors of collapse, such as the expansion of cracks on the slope and at the bottom, and dust emission. In addition, the continuous high-intensity stripping operations at the bottom of the mining area have continuously reduced the stability of the slope. This has left the slope in an unstable state. The stripping and mining disturbance, combined with the over-boundary dumping, further aggravates the development of faults and joint fissures. Ultimately, it triggered a large-scale landslide and collapse of the side-hill rock mass along the fault and joint planes, and caused significant casualties and enormous economic losses.

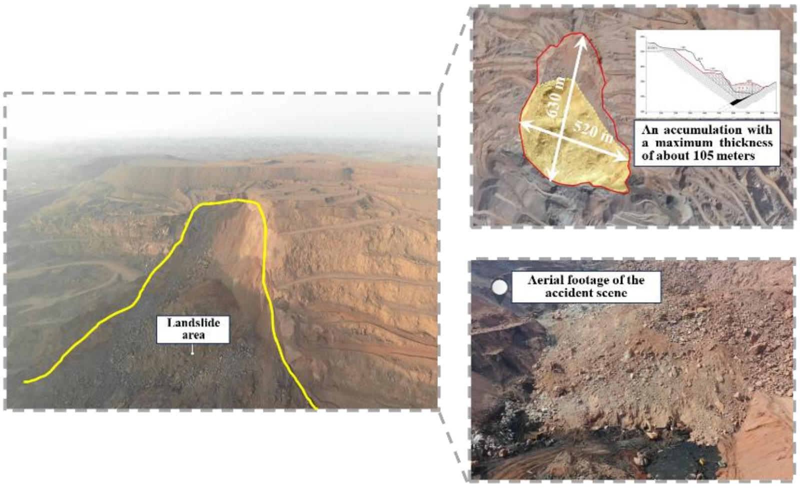


Fig. 6. Scene of the open-pit coal mine collapse in Alxa Left Banner, Alxa League, Inner Mongolia

3.1 Research Process of Collapse Disaster Data

Taking "Inner Mongolia coal mine collapse" as the keyword, the manuscript uses crawler technology to obtain a total of 5829 disaster text data from Sina Weibo about the "2·22" open-pit coal mine collapse accident in Alxa, Inner Mongolia, in 2023. Combined with the LDA (late Dirichlet allocation) model (Figure 7) [16-18], the text data is subject classified and visually displayed. Aiming to provide real-time dynamic information support for mine disaster risk monitoring. The research process covers four modules: data collection and preprocessing, word frequency statistics, LDA model classification and word cloud result visualization. In the data collection phase, the hash method is selected to achieve rapid retrieval of similar content in Weibo, which significantly improves the efficiency of data collection.

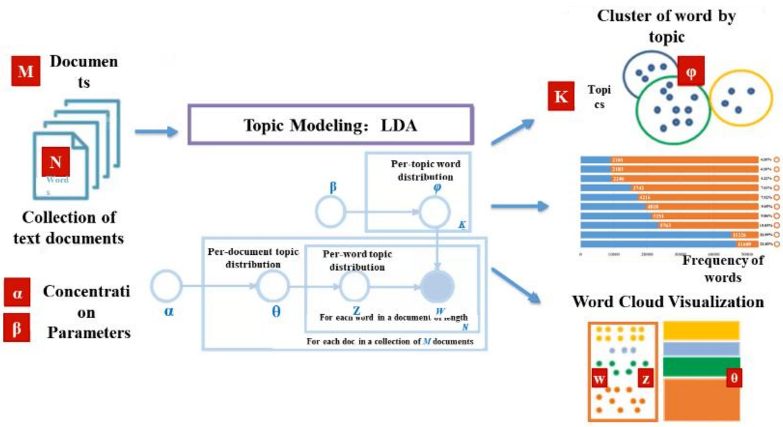


Fig. 7. Data processing flow diagram based on the LDA model

Given that Weibo data is dynamic and real-time, set it to be automatically crawled every 3 hours, with keywords and crawling intervals dynamically updated to ensure the timeliness of the collected data. In response to the common issue of multi-source heterogeneous differences in mine disaster data, the manuscript sequentially performs deduplication, missing value completion, and anomaly removal on the collected data to ensure its uniqueness and consistency. Subsequently, the Weibo data was processed with Chinese word segmentation, meaningless word filtering, and noise reduction [19]. The processed data can provide high reliability input for the follow-up LDA topic model and fast semantic clues for disaster research and rescue decision-making.

After completing the data preprocessing, LDA topic model was employed to extract topic features and perform clustering on the text corpus. This process aimed to uncover latent topic information and categorized it into three topics: disaster site, disaster cause analysis, and disaster relief within disaster events. Concurrently, making use of word cloud visualization technology to map the three-dimensional distribution of topics-words-heat distribution [20]. This reveals the public's focus of attention and the path of information dissemination, providing data support for emergency command and rescue decision-making in mining disasters (Figure 8).

3.2 Analysis of Research Results on Collapse Disaster Data

3.2.1 Visual analysis

A word cloud uses words as its basic units and employs visual encoding in which the font size is proportional to word frequency, mapping high-dimensional word frequency vectors to low-dimensional visual weights, thereby making thematic features instantly readable. Consequently, following the topic classification based on the LDA model, it needs to extract and conduct frequency statistics on the high-probability terms under each topic. Subsequently, the wordcloud engine is invoked to generate topic-specific word clouds. In these word clouds, the character size is monotonically positively correlated with the frequency of word occurrence, which means font size scales directly with term frequency. In these word clouds, the character size is monotonically positively correlated with the frequency of word occurrence, which means font size scales directly with term frequency. This reflects the greater weight of the term within the topic. This method intuitively presents the core semantics of each topic and the focus of public attention. It also provides technical support for analyzing the propagation patterns of disaster-related public opinion and the evolution trends of hotspots.

Topic 1
Disaster site

Topic 2
Disaster Cause Analysis

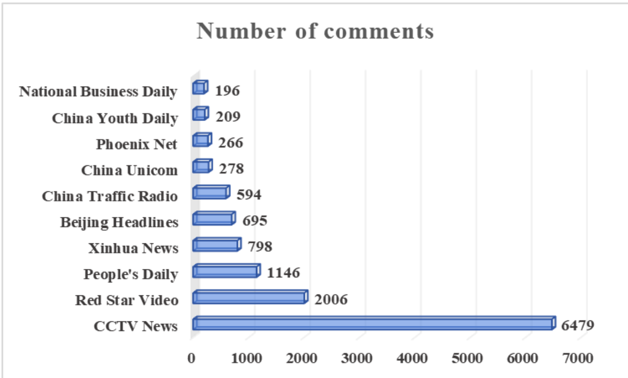
Topic 3
Disaster relief

Fig. 8. Disaster keyword cloud map

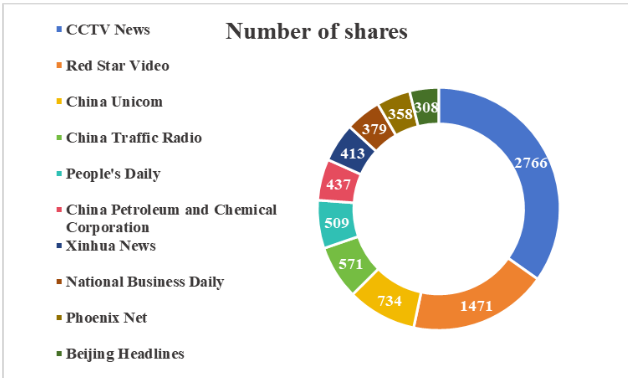
As shown in Figure 8, in Theme 1, keywords such as 'dust,' 'subsidence,' and 'slip' are relatively larger in font size. This indicates that the topic mainly focuses on a direct description of the conditions at the disaster site following the collapse at the Alxa open-pit coal mine. In Topic 2, keywords such as 'cause,' 'investigation,' and 'press conference' are larger in font, indicating that this topic focuses on investigating the causes and analyzing events after a disaster. In Topic 3, keywords such as 'protection,' 'guidance,' and 'disposal' are highlighted, reflecting that this topic focuses on emergency rescue and response actions at disaster sites.

3.2.2 Heat Analysis

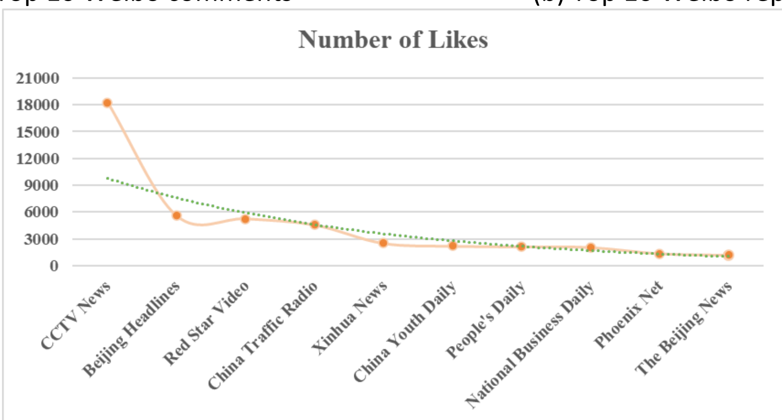
To further analyze the heat distribution of the 'Inner Mongolia coal mine collapse' disaster text data on the Weibo platform. The manuscript uses the number of comments, shares, and likes as proxy variables for heat, and conducts a comparative analysis of the top 10 heat data (Figure 9). Statistical data indicate that the top accounts for the three metrics are highly homogeneous, all centred around authoritative media such as 'CCTV News,' 'Red Star Video,' and 'People's Daily.' Among them, a single post by CCTV News reached peak numbers in comments (6,479), shares (2,766), and likes (18,219), making it the primary channel for information dissemination. From this, it can be seen that China's official mainstream media, leveraging high credibility and a large fan base, has significantly increased the public's understanding of disaster events. The above heat distribution not only quantifies the allocation of public attention, but also provides a calculable and reusable basis for decision-making in media selection for guiding public opinion during disasters, mobilizing post-disaster relief, and disseminating disaster prevention and mitigation knowledge.



(a) Top 10 Weibo comments



(b) Top 10 Weibo reposts



(c) Top 10 Weibo likes

Fig. 9. Weibo disaster data heat distribution

In summary, word cloud visualization and heat analysis can transform the intensity of disasters, emergency measures, and rescue trajectories of coal mine collapses into a semantic map that is easy to interpret. This can provide a quantitative basis for the spatiotemporal optimization of rescue forces and resources, thereby supporting post-disaster recovery and reconstruction of various mining engineering accidents. This approach can also be applied to the rapid assessment and emergency decision-making of other similar disasters, continuously contributing to the upgrade of mining engineering disaster risk management toward a data-driven and agile response model.

3. Conclusions

The manuscript focuses on the key issues of mine disaster risk monitoring, and based on topic model retrieval technology, it conducts research on the data retrieval of 'Inner Mongolia coal mine collapse' accidents through result visualization and keyword popularity analysis. Based on the retrieval results, dynamically monitor and precisely assess the risk of mining disaster accidents, and reach the following conclusions:

(1) Systematically analyzed the development history and feature overview of topic model retrieval technology, and focused on reviewing the principles and application processes of retrieval methods such as CNN, LSH, and SR-DL method;

(2) Based on crawler technology, relevant data on the '2·22' open-pit coal mine collapse accident in Alxa, Inner Mongolia, in 2023 were retrieved and analyzed. By taking the evolution process of disaster-related topics as the entry point, the distribution of the collapse accident disaster topics, development trends, and disaster effects was obtained, providing a scientific reference for the implementation of dynamic regulation strategies for mine disaster risks.

(3) With the widespread application of technologies such as the Internet of Things, large models, and digital twins in the field of mining engineering, in the future, multi-source technologies including AI video intelligent monitoring can be integrated to build large models for disaster accident prediction from aspects such as risk point identification, risk level establishment, and disaster severity assessment, thereby achieving dynamic monitoring and precise evaluation of mining disaster risks. At the same time, massive accumulated fiber optic sensing data and online reviews from social media can be further utilized to continuously optimize the accuracy of model evaluations. Based on this, precise monitoring and reliable prediction of mining disaster risks can be achieved. It provides a scientific basis for safe mine operations and the prevention and control of disaster incidents.

Acknowledgement

This research was funded by the National Natural Science Foundation of China (42377200); the Natural Science Foundation of Hebei Province, China (D2025508013); Graduate Student Innovation Ability Training Funding Project of Hebei Province, China (CXZZSS2025139); the Funds for the Key Laboratory of Earth Fissures Geological Disaster, Ministry of Natural Resources (JSDDY-HJ-D-2024-006)

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] National Bureau of Statistics of China. (2024). Statistical Communique of the People's Republic of China on the National Economic and Social Development.
[2] Wu, M., Wang, H., Zhang, P., & Qin, Z. (2024). Characteristics of Global Industrial Development in 2023 and Outlook of Its Future. Petroleum Science and Technology Forum, 43(2), 15–21.
[3] Zhao, H. (2023). Characteristics and trends of the international energy security situation. International Petroleum Economics, 31(8), 16–24.
[4] National Bureau of Statistics of China. (2015-2024). Statistical Communique of the People's Republic of China on the National Economic and Social Development. People's Daily, 2016-2025.2.

[5] Ding, H., Cheng, T., Huang, Y., Yin, Z., & Lu, P. (2025). Screening of Characteristic Pollutants in Groundwater of High-sulfur Coal Mines in China. Geological Journal of China Universities, 31(1), 48–57. https://doi.org/10.16108/j.issn1006-7493.2024086

[6] Sun, R., Wang, J., Zhou, Y., Shang, X., & Leung, C. (2025). Stability analysis of a compressed air energy storage cavern transformed from a horseshoe-shaped roadway in an abandoned coal mine. Deep Underground Science and Engineering, 4, 562–581. https://doi.org/10.1002/dug2.70041

[7] Piao, C., Yin, Y., He, Z., Du, W., & Wei, G. (2025). Research on transparency of coal mine geological conditions based on distributed fiber-optic sensing technology. Deep Underground Science and Engineering, 4(2), 255–263. https://doi.org/10.1002/dug2.12134

[8] Cheng, G., Wang, Z., Shi, B., Zhu, H., Li, G., Zhang, P., & Wei, G. (2022). Research progress of DFOS in safety mining monitoring of mines. Journal of China Coal Society, 47(8), 2923–2949. https://doi.org/10.13225/j.cnki.jccs.2022.0569

[9] Al-Fakih, A., Abdulraheem, A., & Kaka, S. (2024). Application of machine learning and deep learning in geothermal resource development: Trends and perspectives. Deep Underground Science and Engineering, 3(3), 286–301. https://doi.org/10.1002/dug2.12098

[10] Zhang, L., & Wang, Z. (2023). Flow Pattern Recognition Method of Gas-Liquid Two-Phase Flow Based on Multiple Empirical Mode Decomposition and Convolution Neural Network. Acta Metrologica Sinica, 44(1), 73–79.

[11] Li, H., Zhang, Y., & Yang, W. (2023). Gas explosion early warning method in coal mines by intelligent mining system and multivariate data analysis. PloS one, 18(11), e0293814. https://doi.org/10.1371/journal.pone.0293814

[12] Jiang, C., Zhu, S., Hu, H., An, S., Su, W., Chen, X., Li, C., & Zheng, L. (2022). Deep learning model based on big data for water source discrimination in an underground multiaquifer coal mine. Bulletin of Engineering Geology and the Environment, 81, 26. https://doi.org/10.1007/s10064-021-02535-5

[13] Hussain, W., Merigó, J. M., Raza, M. R., & Gao, H. (2022). A new QoS prediction model using hybrid IOWA-ANFIS with fuzzy C-means, subtractive clustering and grid partitioning. Information Sciences, 584, 280–300. https://doi.org/10.1016/j.ins.2021.10.054

[14] Dai, H., Zhu, M., & Gui, X. (2024). LSH Models in Federated Recommendation. Applied Sciences, 14(11), 4423. https://doi.org/10.3390/app14114423

[15] Huang, W., Li, Q., & Meng, S. (2024). KG2Rec: LSH-CF recommendation method based on knowledge graph for cloud services. Wireless Networks, 30(5), 3483–3494. https://doi.org/10.1007/s11276-020-02387-z

[16] Agarwal, S., Chugh, P., Singh, A., Sakinala, V., Mukherjee, A., Prasad, B., Dagli, C., & Zou, Y. (2025). Application of natural language processing and machine learning for analyzing mining accident reports and automating the process of root cause analysis. International Journal of Coal Science & Technology, 12(1), 91. https://doi.org/10.1007/s40789-025-00822-0

[17] Lin, R., Zhou, Z., You, S., Raghuveer, R., & C.-C., J. (2024). Geometrical Interpretation and Design of Multilayer Perceptrons. IEEE Transactions on Neural Networks and Learning Systems, 35(2), 2545–2559. https://doi.org/10.1109/TNNLS.2022.3190364

[18] Wang, Y., Yao, X., Yu, G., Zhang, Y., Tao, Z., & Zhao, J. (2025). Analysis and Management of Safety Hazards in Intelligent Mines Based on Multidimensional Data Mining. Mining Research and Development, 45(10), 173–181. https://doi.org/10.13827/j.cnki.kyyk.2025.10.019

[19] Li, S., You, M., Li, D., & Liu, J. (2022). Identifying coal mine safety production risk factors by employing text mining and Bayesian network techniques. Process Safety and Environmental Protection, 162, 1067–1081. https://doi.org/10.1016/j.psep.2022.04.054

[20] Guo, D., Li, G., Hu, N., & Hou, J. (2022). Big data analysis and visualization of potential hazardous risks of the mine based on text mining. Chinese Journal of Engineering, 44(3), 328–338. https://doi.org/10.13374/j.issn2095-9389.2020.10.23.004