frontiers | Frontiers in Earth Science

TYPE Original Research
PUBLISHED 18 September 2025
DOI 10.3389/feart.2025.1674305

Check for updates

OPEN ACCESS

EDITED BY
Jie Dou,
China University of Geosciences
Wuhan, China

REVIEWED BY
Mohammad Azarafza,
University of Tabriz, Iran
Chunde Piao,
China University of Mining and
Technology, China

*CORRESPONDENCE
Desheng Cao,
cds@ncist.edu.cn
Keshun Wei,
keshun@uibe.edu.cn

RECEIVED 27 July 2025
ACCEPTED 01 September 2025
PUBLISHED 18 September 2025

CITATION
Cheng G, Wu Y, Cao D, Wei K, Wang Y and Wu Y (2025) Comprehensive analysis and application of geological disaster information leveraging topic modeling and sentiment mining.
Front. Earth Sci. 13:1674305.
doi: 10.3389/feart.2025.1674305

COPYRIGHT
© 2025 Cheng, Wu, Cao, Wei, Wang and Wu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Comprehensive analysis and application of geological disaster information leveraging topic modeling and sentiment mining

Gang Cheng[1,2,3], Yaxi Wu[1], Desheng Cao[1]*, Keshun Wei[4]*, Ye Wang[1] and Yongfei Wu[1]

[1]School of Computer Science and Engineering, North China Institute of Science and Technology, Beijing, China, [2]School of Earth Sciences and Engineering, Nanjing University, Nanjing, China, [3]Hebei Cangzhou Groundwater and Land Subsidence National Observation and Research Station, Cangzhou, China, [4]University of International Business and Economics, Beijing, China

Introduction: In recent years, with the rapid advancement of urbanization in China and the successive implementation of major national strategies such as the Belt and Road Initiative, the Sichuan–Xizang Railway, and the South-to-North Water Diversion Project, the potential risks and losses from geological disasters have continued to rise. Secondary disasters—including landslides, mudslides, and barrier lakes triggered by earthquakes—have significantly intensified the overall impact, posing severe challenges to disaster monitoring, early warning, emergency response, recovery, and reconstruction efforts. In this context, how to leverage new information technologies to achieve in-depth mining and application of geological disaster data has become a critical issue in disaster risk reduction and sustainable crisis management.

Methods: This study focuses on topic modeling and sentiment analysis of disaster-related data, using geological disasters in China as a background. First, it reviews the recent advances in topic modeling and sentiment analysis techniques. Then, based on data characteristics and applicability, two major social media platforms—Weibo (Sina Weibo) and Rednote (Xiaohongshu)—are selected as primary data sources. The advantages of the LDA topic model (e.g., its unlabeled and multi-topic capabilities) and the lightweight processing efficiency of the SnowNLP sentiment analysis algorithm are discussed. As a case study, the "1•07" earthquake in Xigaze, Tibet, in 2025 is analyzed. The LDA model is used to conduct multi-topic classification and clustering visualization of Weibo disaster topic data. Combined with the SnowNLP sentiment analysis algorithm, the phased sentiment evolution judgment application is carried out using the 6-month Rednote comment data.

Results: The LDA model effectively extracts geological disaster-related themes—such as emergency response and post-disaster recovery—and that sentiment analysis technology can reveal phase-based patterns in public emotions.

Conclusion: These findings provide scientific support for geological disaster emergency management and public opinion guidance. The research also expands the application potential of topic modeling and sentiment analysis in

the field of geological disasters and offers a direction for future integration and optimization of multimodal social media data.

KEYWORDS

geological disaster, data mining, topic model, sentiment analysis, visual analysis

1 Introduction

China is situated at the junction of the Eurasian Plate, the Indian Ocean Plate, and the Pacific Plate, spanning the two most active seismic belts in the world: the Circum-Pacific and the Mediterranean-Himalayan belts. Its geological and tectonic environment is highly complex. In recent years, with the continuous advancement of urbanization in China, human engineering activities (such as the storage of large reservoirs and the exploitation of deep mineral resources) have been increasing day by day, as well as the intensive development of high-rise buildings and underground Spaces in cities, which have changed the propagation characteristics of seismic waves and amplified the vibration effects caused by earthquakes. As a result, under the dual influence of geological conditions and human factors, earthquake disasters in China have shown a regional tendency to occur frequently. Taking 2024 as an example, there were 1,066 earthquakes of magnitude 3.0 or above in China (Figure 1), primarily located in Xinjiang, Tibet, Sichuan, and Taiwan. Among them, the largest earthquake was the 7.3-magnitude earthquake in the sea area of Hualien County, Taiwan, on April 3rd. The frequent earthquakes mentioned above pose a continuous and significant threat to people's lives, property, and the national territory's ecological environment.

For geological disasters, the traditional method of obtaining disaster information primarily relies on on-site monitoring in disaster-stricken areas and real-time statistics provided by government departments. The time cycle for collecting, processing, and transmitting disaster information is long, with low efficiency, and problems such as the missed transmission of key information and time lags (Zhu et al., 2012; Du et al., 2020) occur, making it challenging to meet the high timeliness requirements of modern disaster emergency response. With the deep integration of information technology and multimedia technology, mainstream social media such as Weibo and Rednote have become essential channels for people to express emotions and obtain information in their daily lives. People are increasingly enthusiastic about expressing their views and understanding of unexpected events around them on social platforms. This not only promotes fundamental changes in the pattern of event communication (Zou and Yang, 2019; Cao and Liu, 2019) but also broadens the channels and speeds up the dissemination of various disaster information. Meanwhile, due to the unstructured nature of social media data, its high noise, heterogeneity, and semantic complexity pose significant challenges in extracting high-value and expected disaster information from it. Utilizing the correct data analysis methods to extract valuable information from multi-platform disaster data has become a key issue in the field of disaster management. The Topic Model, as a typical unsupervised text mining method, has been widely applied in the field of natural language processing, especially demonstrating excellent processing capabilities in tasks such as topic discovery and text classification (Murshed et al., 2023; Tu and Yang, 2021; Hananto et al., 2022). It can identify the focus of attention at different stages of events. Table 1 shows the research and application status of the topic model in the past 10 years. As for the field of geological disaster prevention and control, Nanehkaran et al. (2021) utilized the Fuzzy Logic-based Multi-Criteria Decision-Making (MCDM) method. Combining climate, geomorphology, tectonic earthquakes, geological hydrology, and human activity factors, the susceptibility of landslides in the Tabriz area of Iran was analyzed, and a distribution map of the high-risk regions was drawn. The area was divided into five sensitive grades, providing a scientific basis for the prevention and control of regional landslide disasters Nanehkaran et al. (2022). utilized a multi-layer perceptron (MLP) artificial neural network, combined with traditional and specific triggering factors as well as historical landslide data, to predict the rockfall susceptibility in Alborz Province, Iran. The results show that the northern part of the area is the highest-risk area. The MLP model is superior to the Support Vector Machine (SVM), Decision Tree (DT), and Random Forest (RF) classifiers, providing a solid foundation for regional landslide risk management. Cemiloglu et al. (2023) evaluated the landslide susceptibility in Maragheh County, Iran, based on a logistic regression model. They selected multiple topographic, climatic, geological, and human activity factors, combined with historical landslide data, to establish a model, and determined that the area as a whole is in a medium to high-risk zone. The experimental verification results show that the AUC value is 0.769, and the LR model has high reliability. At the same time,

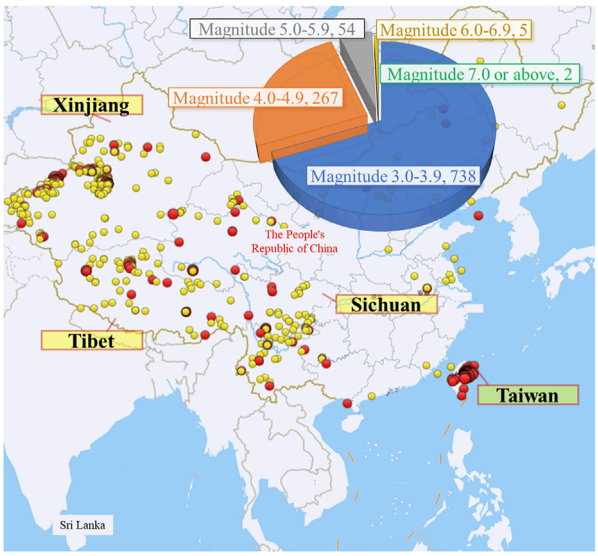

FIGURE 1
Distribution of earthquakes above magnitude 3.0 in China in 2024.

as disaster emergency management not only focuses on disaster information mining and emergency response, but also needs to take into account the long-term evolution of public opinion, public psychological recovery and post-disaster reconstruction strategies, the use of emotion analysis technology can fully understand the process of public emotional changes, determine the key points of emotional turning, and provide a scientific basis for precise rescue decisions. Therefore, an increasing number of researchers are integrating topic models and sentiment analysis techniques to apply them in the field of disaster management, conducting correlation analysis of disaster big data and public opinion control, and providing new research paradigms and technical support for emergency response, recovery, and reconstruction of various disasters (Luo et al., 2021).

Amiresmaili et al. (2021) demonstrated through empirical research that social media is primarily utilized for four key dimensions of disaster situation awareness, emergency information dissemination, disaster response coordination, and disaster data retrieval in earthquake disasters. They revealed that the disaster processing mode based on social media exhibits obvious cross-regional universality. The research results can serve as a reference for optimizing the national disaster management system. However, its research mainly relies on literature reviews and lacks empirical data support from specific cases, resulting in limited methodological depth. Wang (2021) constructed a disaster data sample database by collecting a large number of landslide and debris flow disaster data, studied the characteristics of high-dimensional disaster data, and carried out multidimensional visual displays of the results based on different data characteristics. At the same time, the resolution optimization of visualization results was carried out by using methods such as topic clustering and nonlinear variation. The research results play a crucial guiding role in clarifying the distribution characteristics of disasters and the entire process of occurrence, development, and evolution. But this research focuses more on the display of data features and has insufficient integration with the actual situation of disaster emergency management. Its application value still needs to be expanded. Eligüzel et al. (2023) systematically analyzed disaster-related social media content based on the Twitter data of the 2020 earthquake event off the coast of Izmir, Turkey, by comprehensively using Latent Semantic Analysis (LSA), Word2Vec word embedding model and Latent Dirichlet allocation (LDA) topic modelling method. The research results provide necessary data support for government departments to formulate post-disaster psychological intervention and economic aid policies. Ghaly and Laksito, (2023) used Non-negative Matrix Factorization (NMF) technology to model and analyze the Indonesian natural disaster text dataset. They confirmed that this method could significantly enhance disaster monitoring and early warning capabilities, providing an essential reference for innovating text analysis methods in disaster management and formulating precision disaster reduction strategies. Nevertheless, research remains at the level of technical verification and has not been closely integrated with disaster emergency decision-making and practical needs. Ruan et al. (2022) utilized a GPU-accelerated Poisson Dirichlet Mixture model (GPU-PDMM) to compare and analyze the public discussion content on Twitter and Reddit platforms following the Ridgecrest earthquake, effectively revealing the cognitive differences among various social media user groups regarding the same disaster event. Xing et al. (2019) integrated spatiotemporal analysis and semantic mining technology to conduct a multidimensional analysis of earthquake-related information on the Weibo platform, accurately describing the spatial distribution characteristics of disaster impact and the spatiotemporal evolution law of emergency information. However, its research only focused on the single platform of Weibo and lacked cross-platform data integration and verification. To address the limitations of text sentiment analysis, Hassan et al. (2022) developed a deep visual sentiment analyzer based on a Convolutional Neural Network (CNN) and transfer learning, enabling the research of visual sentiment analysis in disaster images by extracting visual object and scene-level feature information from the pictures. Contreras et al. (2022) used the Twitter data of L′Aquila earthquake to conduct unsupervised sentiment analysis research based on machine learning algorithm, and analyzed that about 33.1% of the tweets had negative polarity, 29.3% had neutral polarity, 28.7% had positive polarity, and 8.9% had irrelevant topics. Thus, the reconstruction and public opinion of the disaster area 10 years after the earthquake can be effectively evaluated. The studies mentioned above have all provided methodological guidance and practical examples for the mining, analysis, and application of disaster data on social media, thereby promoting the development of intelligent, efficient, and precise disaster information analysis systems. Based on this, the effective utilization of geological disaster information has been achieved. However, the research primarily focuses on single-platform data and single-functional applications, such as disaster situation identification or emergency management. Few scholars have conducted long-term research on the thematic development trends and emotional evolution characteristics by integrating the specific geological disaster features and data from multiple platforms, and there is a lack of an overall analysis of the entire life cycle of disasters (from early warning, emergency response, to recovery and reconstruction). Therefore, based on the current research status of geological disasters and the topic data of Weibo, a mainstream social media platform in China, as well as the user comment data of Rednote for 6 months, this paper summarizes and analyzes the application of topic models and sentiment analysis techniques in the field of disaster information mining. To provide new research ideas and practical references for the development of topic models and sentiment analysis in the field of disaster emergency management. The main contributions of this article are as follows:

1. This paper systematically reviews the research progress and application status of topic models over the past decade, summarizes the development process of the LDA model and its potential applications in the field of social media data processing and analysis, and provides technical references for disaster information mining.
2. Based on the differentiated data characteristics of Weibo and Rednote platforms, a multi-platform collaborative analysis framework was proposed. Combined with the LDA topic model and SnowNLP sentiment analysis technology, the topic classification of earthquake disaster data and the sustainable tracking of sentiment evolution were achieved, thereby optimizing the domain adaptability of sentiment analysis.

TABLE 1 Research and application status of topic models in the past 10 Years.

Time	Researchers	Research contents	Point of innovation	Application fields
2015	Das et al. (2015)	GLDA model with multivariate Gaussian distribution	The Linear Discriminant Analysis (LDA) parameterized representation is improved to improve the performance of out-of-vocabulary word processing	Text modeling
	Cao et al. (2015)	Neural Topic Model (NTM)	Combined the "word-document" relationship to solve the problem of single topic distribution and initialization sensitivity	Text analysis
2016	Li et al. (2016)	Generalized Polya-Urn-Dirichlet Polynomial Mixture (GPU-DMM)	The GPU model is introduced into DMM to enhance semantic relevance	Short Text Processing
	Yin et al. (2016)	Spatiotemporal LDA (STLDA)	Enhance the ability of regional interest reasoning	Point of interest recommendation
2017	Yao et al. (2017)	Knowledge Graph Embedding LDA (KGE-LDA)	Embed knowledge graphs to enhance semantic coherence	Semantic enhancement
	Kowald et al. (2017)	Cognitive Heuristic Tag Recommendation Method	Consider the effect of time on personal/social tagging	Tag recommendation
2018	Jacobi et al. (2018)	LDA analyzes news content and trends	Combine the topic model to achieve the analysis of news information content	News dissemination
2019	Shao et al. (2019)	Emotion-aware Multimodal Topic Model (SMTM)	Combining multimodal data to achieve personalized recommendations	Travel Recommendation
	Shi et al. (2019)	Sparse RNN Topic Model (SRTM)	Combine sparsity with an RNN structure	Dynamic Text Modelling
2020	Wang and Yang (2020)	Topic Attention Model (TAM)	An attention mechanism is used to replace words to represent topics and reduce modelling complexity	Document modeling
	Shi et al. (2020)	User-based Aggregated Topic Model (UATM)	Study user preferences and intention distributions	Intelligent product recommendation
	Yang et al. (2020)	System decision doctor recommendation model	Take into account the patient's preferences and opinions comprehensively	Medical decision making
	Liu et al. (2020)	Cross-Site LDA (C-LDA)	Modelling Cross-site User-Generated Content	Cross-platform content analysis
2021	Meng and Xiong (2021)	Hybrid Physician Recommendation Model	Based on the online medical platform, doctors are recommended according to the needs of patients	Medical recommendation
	Toubia (2021)	Innovation literature research topic model	Provide assistance for literature writing	Academic Writing
	Peng et al. (2021)	Emotion and Behavior Topic Model (SBTM)	Quickly obtain the relevant text of participants 'concerns	Content analysis
	Liu et al. (2021)	Adversarial Cross-media Retrieval Based on Semantic Similarity (SSACR)	Cross-modal retrieval is achieved by negative training of neural networks	Cross-media modeling
2022	Ji et al. (2022)	Social Cycle Aware Topic Model (SPATM)	Distinguish between user interests and social preferences	Personalized venue recommendation
2023	Li et al. (2023)	Multi-view Scholar Clustering Topic Model (MSCT)	The clustering was carried out by integrating scholars' interests and internal and external dual-view information	Scholar Recommendations

(Continued on the following page)

TABLE 1 (*Continued*) Research and application status of topic models in the past 10 Years.

Time	Researchers	Research contents	Point of innovation	Application fields
2024	Pavithra and Savitha (2024)	Performance comparison of LDA, HDP, NMF, BERTOPIC, and DTM models	The topic classification model is combined with the topic evaluation of academic papers	Evaluation of Academic Papers
	Koltcov et al. (2024)	Evaluating topic model stability and interpretability	An optimized topic model based on Granular Sampling of Word Embedding Vector (GLDAW) is proposed	Text topic classification
2025	Park et al. (2025)	Time series forecasting	Topic modelling was combined with deep learning to predict blockchain topic trends	Blockchain Topic Prediction
	Cheng et al. (2024)	The LDA topic model is used in geological disasters	The topic model is combined with geological disaster management and assessment	Disaster management

3. Taking the "1·07" earthquake in Xigaze, Tibet, in 2025 as a case, through topic clustering, time series analysis, and heat map visualization, the phased characteristics of disaster emergency response, rescue dispatch, and public opinion evolution were revealed, providing data-driven decision support for disaster management.
4. Innovatively integrating spatiotemporal dimensions with emotional computing, it analyzed the periodic fluctuation patterns of public sentiment, proposed key time nodes and intervention strategies for public opinion monitoring. It expanded the depth and breadth of disaster public opinion research.

The rest of this article is organized as follows: Section Two elaborates on the characteristics of social media data and the principles of LDA and SnowNLP methods; The third section presents in detail the topic classification and emotional evolution analysis experiments of the earthquake cases in Tibet. The fourth section summarizes the research results and outlines future directions for multimodal data fusion and model optimization.

2 Data and methods

2.1 Data evaluation

Weibo and Rednote, as two mainstream platforms in the Chinese social media landscape (Figures 2, 3), offer significant advantages over other platforms. Weibo pioneered the public square mode of "hot search list + topic tag", and Rednote created the vertical community of "interest tag + personalized recommendation", which jointly realized the accurate contact and transmission of information content between strangers. It is a representative social media platform in China. However, there are significant differences in data characteristics and research applicability. The Weibo platform utilizes highly structured topic text data as its primary information source, and its platform users generally adhere to standardized Hashtag behavior. According to different topic tag categories, an official hierarchical classification system covering 30 vertical fields is constructed. It provides the topic data with a clear direction and domain attribution characteristics. This type of structural feature offers an ideal basis for scientific research on topic classification. In contrast, the Rednote platform primarily uses short images, texts, and videos as content forms, and its topic tags are less frequently used and less standardized. Coupled with the relatively weak official classification system, the effect of traditional text topic classification methods is limited, resulting in fuzzy topic boundaries and overlapping topic fields, which makes it difficult to establish clear topic divisions and domain attributions. However, the unique review collection ecosystem of this platform shows significant research value. Its user comment data volume is enormous and interactive, and the density of emotional words is significantly higher than that of Weibo topic text, as determined by the Chi-square test (Li et al., 2022). Therefore, the emotional expression is more direct and abundant, which is perfectly suitable for sentiment analysis research. In summary, the two mainstream social platforms exhibit strong complementarity in data characteristics, providing a diversified range of research materials and methodological options for social media research. This paper takes the January 7 earthquake disaster event in Tibet as the research object and carries out macro-topic mining of the Tibet earthquake disaster on the Weibo platform. Moreover, this paper utilizes the Rednote comment data to conduct micro-emotional computing and user psychology research following the earthquake, aiming to obtain more comprehensive disaster emergency response plans and analysis of public opinion evolution.

2.2 LDA topic classification model

In 2003, Blei et al. (2003) proposed a three-level unsupervised probabilistic topic model, LDA, consisting of "document-topic-word" based on the research of PLSA (Probability Latent Semantic Analysis) model (Figure 4). LDA uses the Bag-of-Words model to represent documents, where each document is represented as a word frequency vector. At the same time, the Dirichlet prior distribution is introduced as a model parameter to enhance the model's generalization ability. The model structure is divided into two parts: the document-topic Level and the topic-word Level. The probability distribution is used to describe the relationship between the Document topics and the composition of the Topic

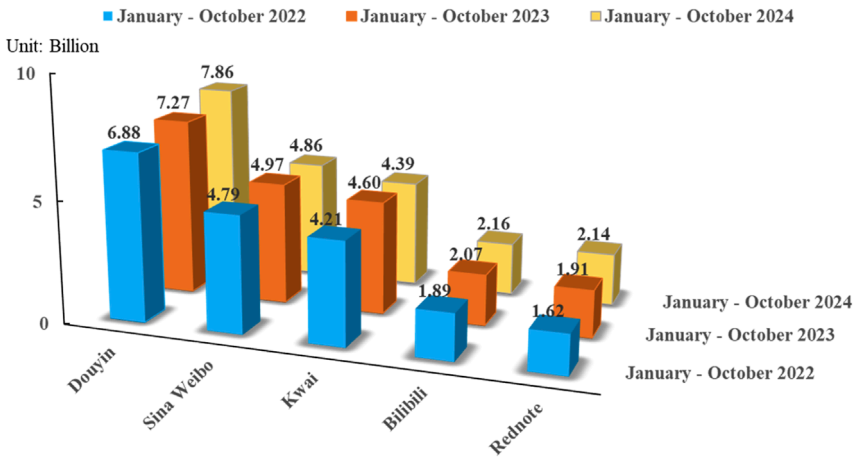

FIGURE 2
The average monthly growth of active users on mainstream social media.

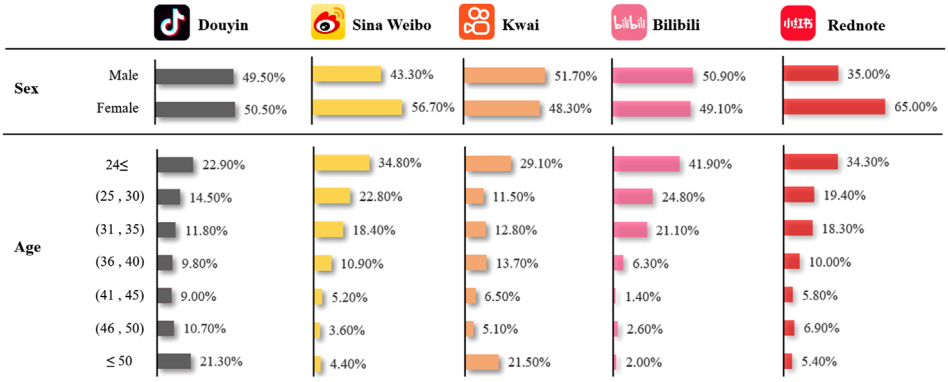

FIGURE 3
User portraits of mainstream social media platforms in October 2024.

words. The classic LDA model has significant advantages in topic classification, including: it does not need to pre-label the document category, which is suitable for dealing with massive unlabeled text and perfectly fits the characteristics of social media data; Compared with BERTopic model (Wang et al., 2024), LDA topics are presented by word probability distribution, which is more interpretable and convenient for subsequent topic data classification. LDA allows multi-topic attribution of documents (Huang et al., 2024), which is more suitable for social attribute texts. In summary, the LDA model is selected to classify social media disaster topic data, thereby obtaining analysis results closer to the actual application scenario. Through indepth research on topic classification, the LDA model has also been extended and expanded. Figure 5 shows the key nodes in the development of the LDA model.

2.3 SnowNLP sentiment analysis algorithm

SnowNLP, as a mainstream natural language processing tool library, is primarily used to calculate text sentiment scores, distinguishing between positive and negative sentiment of words, and performing a preliminary analysis of sentiment orientation. In the sentiment analysis module, the Naive Bayes classifier (Zhang, 2020) is used to predict the sentiment polarity (positive/negative). The main steps are as follows: Firstly, the text is transformed into structured data using word segmentation and feature extraction technology. Secondly, the Naive Bayes classifier is employed to train the sentiment model, with probability smoothing and sentiment dictionary enhancement integrated to process unknown words and negative structures. Finally, the sentiment score, ranging from 0 to 1, is output, and the positive/negative tendency is classified using a simple threshold, which is suitable for short text analysis, such as social media reviews and product reviews. SnowNLP has the following advantages: with “lightweight” as its core, it strikes a balance between Chinese adaptation, development efficiency, and interpretability, providing a simple and efficient practical tool for small and medium-sized Chinese sentiment analysis in general fields. It supports custom domain dictionary injection and can add disaster terms such as “Aftershock” and “Barrier Lake”. It is suitable for vertical scenarios such as e-commerce, social media, and

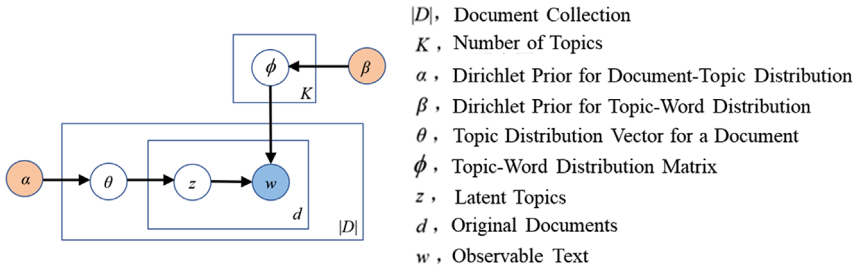

FIGURE 4
LDA model.

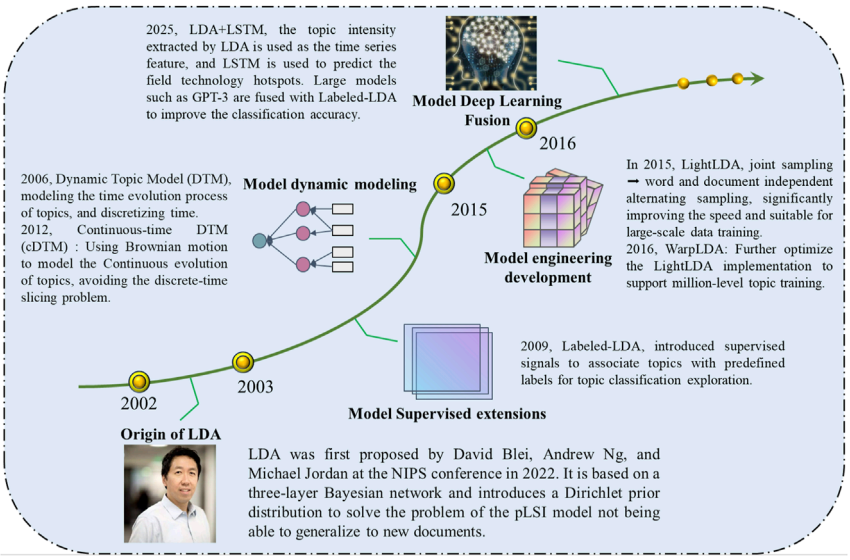

FIGURE 5
Development process of the LDA model.

public opinion during disasters. It has low computational resource requirements, as well as high-cost performance in sentiment analysis experiments.

3 Case study experiment

Xigaze, Tibet, is located near the suture zone of the Yarlung Zangbo River on the southern edge of the Qinghai-Tibet Plateau. It is one of the most active seismic tectonic belts in the world. It is affected by tectonic factors such as the thrust fault (MFT) of the central front of the Himalayas and the detachment system of southern Tibet (STD). Crustal stress continuously accumulates and can only be released through the rupture of the crust's brittle layer in the end. Therefore, on 7 January 2025, a magnitude 6.8 earthquake struck Xigaze, Tibet, resulting in 126 deaths, damage to 27,248 houses, and the collapse of 3,612 houses, as shown in Figure 6.

For this major earthquake disaster event, Weibo topic discussion and Rednote user comments were selected as research samples. Through LDA topic model construction and emotional polarity calculation, real-time monitoring of earthquake disasters and accurate research and judgment of public opinion situations were achieved, providing data support and management direction for the emergency management department. The specific experimental steps are shown in Figure 7.

3.1 Data acquisition and preprocessing

3.1.1 Media data acquisition

The data collection process in this paper is mainly divided into two steps: 1) based on the web crawler technology, the related data information is crawled from Weibo and Rednote with Tibet earthquake-related words as keywords. Because the geological disaster data information on Weibo exhibits dynamic characteristics, real-time monitoring of data is achieved by automatically crawling every hour. The keywords ("Tibet", "Xigaze", "Earthquake", "Earthquake in Tibet") for data retrieval are continually modified to ensure data integrity, thereby obtaining a more comprehensive dataset of geological disaster data. Finally, 32,114 topic data of Weibo and 13,395 comment data of Rednote are crawled; 2) Sort the data according to the release time in the Weibo data set, focus on the

disaster data within 72 h after the disaster, especially strengthen the real-time retrieval and analysis of the Weibo data within 24 h after the disaster, to ensure the timeliness of the data. Figure 8 illustrates the specific collection effect of disaster social data.

3.1.2 Data language difference processing

1. Identification of platform data discrepancies

In view of the differences in semantics, syntax and discourse style between Weibo and Rednote, this paper proposes a research path of "differential targeted identification - hierarchical language normalization - platform style collaborative adaptation" to achieve effective comparison of cross-platform data. The main differences between Weibo and Rednote are shown in Table 2.

2. Hierarchical normalization processing

To achieve a unified analysis across platforms, the language differences are normalized hierarchically in the data preprocessing stage, which mainly includes the following three aspects:

1. Format and redundancy cleaning: Unified data structure as " (timestamp, core text, user type)". Platform identifiers such as "@user, forwarding//comment" were removed from Weibo data, and only "topic tag + body" was retained. For the Rednote data, the image/video links and emoticons were removed, and only the text content was retained. In terms of the time dimension, the Weibo data focused on 72 h after the disaster (the focus was 24 h), sorted by "hours". The Rednote data covers 6 months after the disaster, sorted by "day", and unified into "YYYY-MM-DD HH:MM: SS" format. Through keyword matching and manual verification, advertisements and irrelevant content were removed, and finally, 32,114 Weibo topic data and 13,395 Rednote comment data were retained.
2. Syntactic normalization: using the Jieba word segmentation tool, combining the characteristics of different platforms to formulate word segmentation strategies. Weibo data uses "exact patterns" to match canonical terms. The data of Rednote adopts "full mode + colloquial rules", such as splitting "broken defense" into "broken defense" and "broken defense". At the same time, a double-layer stop word list of "general + disaster domain" is constructed to remove redundant words (such as "forward" in Weibo and "I feel" in Rednote). In addition, the common ellipses in the Rednote were completed (such as "there is no water and no electricity" was completed as "there is no water and no electricity in the earthquake area"), and the punctuation marks were unified. The brevity of the Weibo remains the same (e.g., "Initiate level III response").
3. Semantic normalization: a "disaster domain terminology mapping table" (Table 3) is constructed to unify the concept expression of different platforms. For example, the common expressions of "earthquake center" and "epicenter" are unified as "epicenter", and "national rescue department and government disaster relief department" are unified as "emergency management department". At the same time, a custom dictionary containing professional terms such as "aftershock, quake lake, relief materials" is imported in the word segmentation stage to ensure the consistency of semantic recognition.

3.2 Classification of data topics related to disasters on weibo

The Tibet earthquake disaster data crawled from the Weibo platform exhibits apparent characteristics of unlabeled, multi-topic, and structured data. LDA has the advantages of unsupervised learning ability, dimensionality reduction of high-dimensional sparse text, and the ability to mine the probability distribution of potential topics. Therefore, this paper selects the LDA model to classify the issues of the Tibet earthquake disaster data obtained from the Weibo platform.

3.2.1 Topic number selection

When using the LDA topic model to carry out topic classification research, the optimal number of topics K is often determined by the index of perplexity. In the research of text topic mining based on the LDA model, determining the optimal number of topics K is a key problem. In research, the perplexity index is usually used as the model evaluation standard (Ankner et al., 2024), and the generalization performance of the subject model is quantified by calculating the prediction effect of the model on the data. The calculation formula of the perplexity degree is shown in Formula 1:

$$Perplexity(D_{test}) = exp\left(-\frac{\sum_{d=1}^{M} logp(w_d)}{\sum_{d=1}^{M} N_d}\right) \quad (1)$$

D_{test} is the test set, M is the number of documents, w_d is the word sequence of document D in the test set, N_d is the total number of words in document D, and $p\ (w_d)$ is the generation probability of document D by the LDA model. When the model is well fitted and has strong generalization ability, $p\ (w_d)$is close to 1, and -log$p\ (w_d)$ is close to 0, and the degree of perplexity is the lowest.

During the model experiment, set the number of topics to [1:27] with an interval of 1, calculate the perplexity degree and model score corresponding to each topic, and comprehensively evaluate the optimal number of topics. As shown in Figure 9, the lowest point of the perplexity degree curve and the highest point of the model score are for the number of topics 23. At this time, the generalization effect of the model is the best, so the best performance K value is 23. Therefore, in the process of parameter selection, the perplexity of the topic is used as the main evaluation index, supplemented by the Model Score for comparison. Under the premise of ensuring the interpretability of the topic, the main parameter Settings of the experiment are determined as follows: Number of topics (num_topics): K is 23, $\alpha = 50/K$, $\beta = 0.01$; The number of iterations is set to 1000 to ensure the convergence of the model under the sample size. Number of passes (training rounds): Set to 20 to balance computational efficiency and model stability; alpha and eta (hyperparameters): By adopting the "auto" mode, the model undergoes adaptive optimization during the training process, avoiding deviations caused by manual Settings. Random seed (random_state): Fixed at 42 to ensure the repeatability of the experiment.

3.2.2 Topic feature extraction and classification of disaster data

The topic model was trained using the optimal number of topics (K) and specific parameters, including the learning method,

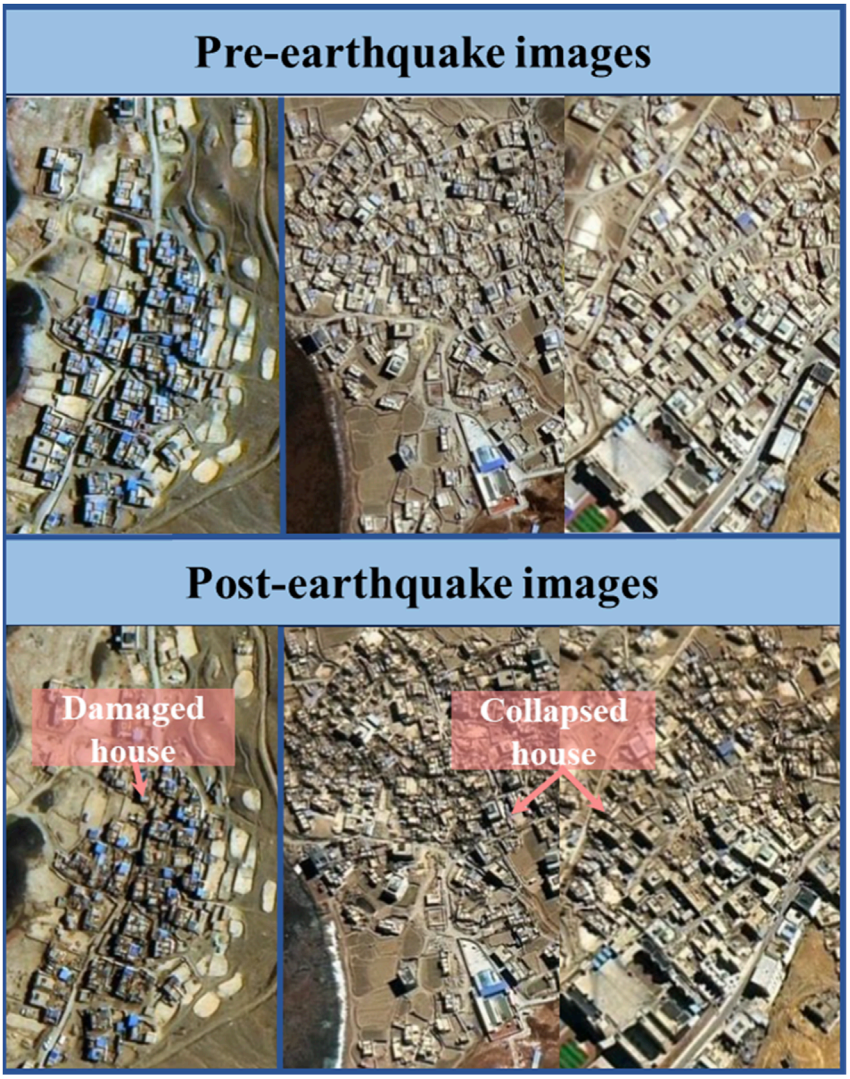

FIGURE 6
Comparison of ground cover changes between pre-earthquake and post-earthquake.Satellite imagery from the Chinese Jilin-1 satellite, acquired on 5 December 2024 (pre-earthquake) and 7 January 2025 (post-earthquake).

learning rate attenuation, and document-topic, before obtaining the topic-word probability distribution of the "Tibet 1·07 Earthquake" disaster, as shown in Table 4. The effect diagram of the 23 topic word clouds is shown in Figure 10.

3.2.3 Visual analysis of topic classification

To enhance the efficiency of data analysis and facilitate a deeper discussion of the results, this paper utilizes interactive visual analysis through the PyLDAvis library. PyLDAvis, as a professional topic model visualization tool, constructs a web-based dynamic visualization interface by analyzing the implicit semantic structure of the LDA model, as shown in Figure 11. The interface includes: 1) the left bubble map space mapping, the bubble represents the topic, the area reflects the distribution proportion of the topic in the data, and the bubble spacing represents the semantic similarity between topics; 2) The bar graph of word frequency on the right shows the 30 feature words with the highest relevance to the selected topic and their frequency distribution. Due to the high semantic overlap among the 23 topics obtained from the initial modeling, to avoid interpretation redundancy, this paper clusters the topics based on the distance matrix between adjacent bubbles provided by PyLDAvis. It combines keyword semantics with manual interpretation of typical documents to integrate the topics with similar content into five macro topics. To optimize the efficiency and systematicness of the subsequent overall analysis. T (1,13,16) can be combined into a new disaster topic (DT1) based on the semantic similarity of the topic (bubble spacing); T (17,19,20,21,22,23) is combined as DT2; T (4,8,10,11,12,14,15,18) is combined as DT3; T (7,9) is combined as DT4; T (2,3,5,6) is combined as DT5.

In the word cloud graph, words are displayed in different sizes and colors. The higher the frequency of words, the larger the font in the graph. According to the analysis of the cloud chart of disaster topic words (Figure 12), the topic DT5 is disaster

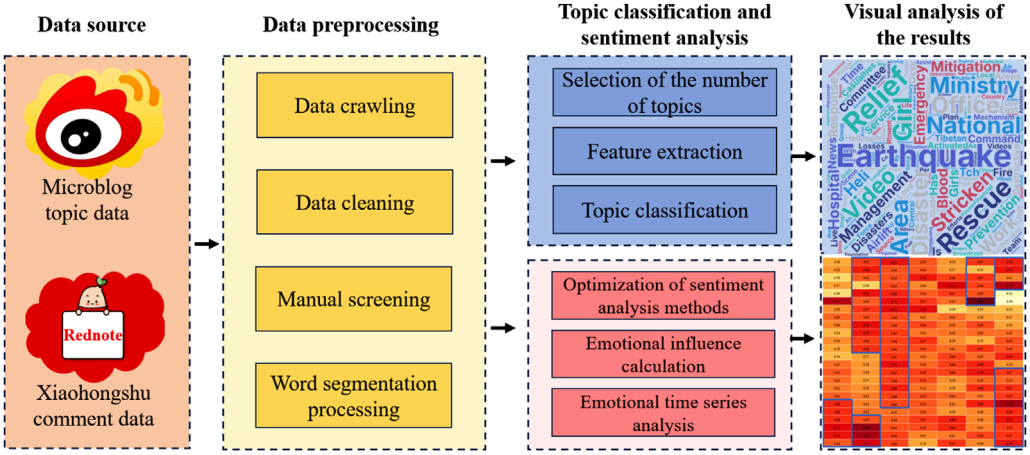

FIGURE 7
Experimental analysis of Tibet seismic events.

weibo_id	weibo_bid	User	Release time	Weibo content
2117508734	P8rpD1jPH	China Meteorolo	2025-1-7 10:00	# A magnitude 6.9 earthquake occurred ne
2097116807	P8rpD1jU5	Weather in Fosha	2025-1-7 10:00	【Earthquake Magnitude Knowledge】 I
1378417172	P8rpCCS6s	Chinese Red Cro	2025-1-7 10:00	# Xiaohong Science Popularization # How
2106558073	P8rpCF46I	Henan Fire Depa	2025-1-7 10:00	# Xizang Earthquake # Automatic determi
2127460165	P8rpD1jxb	China Daily Wel	2025-1-7 10:00	# Xizang Earthquake # [A Magnitude 6.8
1821499183	P8rpD1iXQ	Dalian Daily	2025-1-7 10:00	# Earthquake News # [# 6.8-magnitude Ea
5567800203	P8rpC84LI	Aria_24601	2025-1-7 9:59	# Xizang Earthquake # Safe and Sound??
6598272072	P8rpBbx3g	Chika Momohara	2025-1-7 9:59	# Xizang Earthquake # Hope everyone is s
2987073672	P8rpAihDs	Peaceful Shehong	2025-1-7 9:59	# Xizang Earthquake # [Share and Save! H
6534512380	P8rpzrg0t	I'm not Mu Bai	2025-1-7 9:59	@Chinese meteorological enthusiasts: The
6160152775	P8rpywUS7	Ordinary Ah Yu	2025-1-7 9:59	# Earthquake in Xizang ## Earthquake Wa
2472068187	P8rpyy0DV	Yuzi Peach	2025-1-7 9:59	# Xizang Earthquake # May everyone be s
5721822734	P8rpxnaCg	A breeze blew pa	2025-1-7 9:59	# Earthquake in Xizang # Earthquakes hav
7343885958	P8rpxkY7v	I like you so muc	2025-1-7 9:59	# Xizang Earthquake # Hope for peace. M
5873646112	P8rpx4t9O	Langya Editor	2025-1-7 9:59	【Attention! Wish you peace! # There are
7789013535	P8rpv86fB	Pearls don't fall f	2025-1-7 9:59	# Xizang earthquake # Oh my god, a magn
1241210440	P8rpufYii	Thin people shou	2025-1-7 9:59	# Xizang Earthquake # Terrifying?? . May
7403302991	P8rprhymf	Unknown Knife	2025-1-7 9:59	# Earthquake in Xizang # wish you peace!
7876208105	P8rprhy2F	Tashi Nuyoah	2025-1-7 9:59	# Xizang Earthquake # Is everyone safe ar
7536318306	P8rpqhHDC	Cat lover mo	2025-1-7 9:59	# 6.8-magnitude earthquake in Xigaze, Xiz
5417441659	P8rpqgBTZ	Zha Jin 007	2025-1-7 9:59	# Earthquake in Xizang #?
1192119933	P8rppCMqi	Beijiang News	2025-1-7 9:59	# Consecutive Earthquakes in Xigaze, Xiz
2649732781	P8rpp6RC1	Yangxin Media C	2025-1-7 9:59	# Consecutive Earthquakes in Xigaze, Xiz

（a）Weibo topic data

Time	User	Rednote comments
2025-01-28 12:49:14	Youdaoplaceholder0 xiao	I bring it in to hand in my homework. The author should not mind, right? [Giggling R][Giggling R]
2025-02-07 17:29:30	MD- Yuning	Sure, that's fine too [chuckle R]
2025-02-09 02:23:09	Little orange cat	@Meet at seven o 'clock
2025-02-08 19:24:40	Small sweet potato 653CF9D3	I'm going to copy it too. [Shy R]
2025-01-08 19:04:51	Shu Zhi Clothing	Why was the reception of relief supplies stopped? Is it enough?
2025-01-09 09:58:39	Chubby and fair	Is what's going on true? Since I haven't seen any reports, they are all posted by self-media. May I ask if they are true

（b）Rednote comment data

FIGURE 8
Example of extracting social data related to the "Tibet Earthquak". **(a)** Extraction result of Weibo topic data, (8 January 2025 – 12 January 2025). **(b)** Extraction result of Rednote comment data (1 January 2025 – 30 June 2025).

dispatch, which focuses on the means of disposal at the initial stage of the disaster. The words "work in the earthquake area", "press conference", "rescue" and other words are large, indicating that after the disaster, the country should actively respond to the disaster rescue work at the first time, allocate rescue forces, quickly start the disaster rescue action to strive for the golden time of rescue, form a "full chain, three-dimensional, intelligent" emergency response mechanism, provide timely life assistance and basic life security for the affected people, and minimize casualties and property losses; DT2 refers to on-site rescue. The topic mainly includes words such as "house", "ruins", "fire protection" and "people", indicating that the losses caused by disasters are mainly caused by house collapse and mass casualties, and the rescue force is primarily the fire department, which provides essential decision-making guidance and Implementation reference for subsequent rescue deployment and emergency treatment; DT3 is post disaster recovery. The topic involves various social relief forces and materials, such as "foundation", "milk tea", "group", etc. through the integration and analysis of the topic data, we can fully understand the material situation in the disaster area, quickly respond to the diversified needs after the disaster, realize the precise rescue and efficient allocation of resources in the disaster area, and explore a modern disaster relief model of "technology + social collaboration"; DT4 is a disaster notification. The topic is the collection and release of basic disaster relief information such as "earthquake source", "east longitude" and "news", so that the public can timely and accurately obtain the on-site disaster information, understand the disaster development situation and rescue progress, control the development of public opinion, maintain social stability, and build a disaster relief environment of "government media public" mutual trust; DT1 is a post disaster response. The topic focuses on the recovery and reconstruction after the earthquake disaster. The words "country", "emergency

TABLE 2 Shows the differences in semantics, syntax and discourse style between Weibo and Rednote.

Dimension of difference	Weibo data characteristics	Rednote data characteristics
Semantic	(1) With "# hashtag #+ short text" as the core structure, the official classification system provides 30 vertical fields with clear topic boundaries; (2) Include "@user, forwarding//comment" and other platform specific identifiers; (3) The data is mainly real-time, focusing on the fragmented information in a short time after the disaster	(1) In the main form of "short text + image/video", the frequency of hashtags is low and non-standard, there is no unified classification, and the topic boundary is fuzzy; (2) contains a large number of emoticons and modal words; (3) The data is more interactive, and most of the comments are long-term (such as 6 months) after the disaster, and the emotional expression is more coherent
Syntactic	(1) The text is more formal, mostly official circulations or media reports; (2) Complete sentence structure and proper use of punctuation; (3) less redundant information and clear core content	(1) The text is colloquial, with common daily expressions; (2) The sentence structure is loose, with ellipses sentences and network buzzwords frequently appearing; (3) More redundant content, including subjective expressions
Discourse Style	(1) The usage specification of domain terms; (2) The concept expression is relatively unified	(1) Domain terms are mostly replaced by colloquial forms; (2) The same concept is expressed in various forms

TABLE 3 Examples of disaster-domain terminology mapping.

Standard term (Weibo)	Rednote Variants	Normalization strategy
Epicenter	Earthquake center, Seismic source	Unified as "Epicenter"
Ministry of Emergency Management	National rescue department, government disaster relief	Unified as "Ministry of Emergency Management"
Post-disaster reconstruction	Rebuilding houses, restoring life, rebuilding homes	Unified as "post-disaster reconstruction"

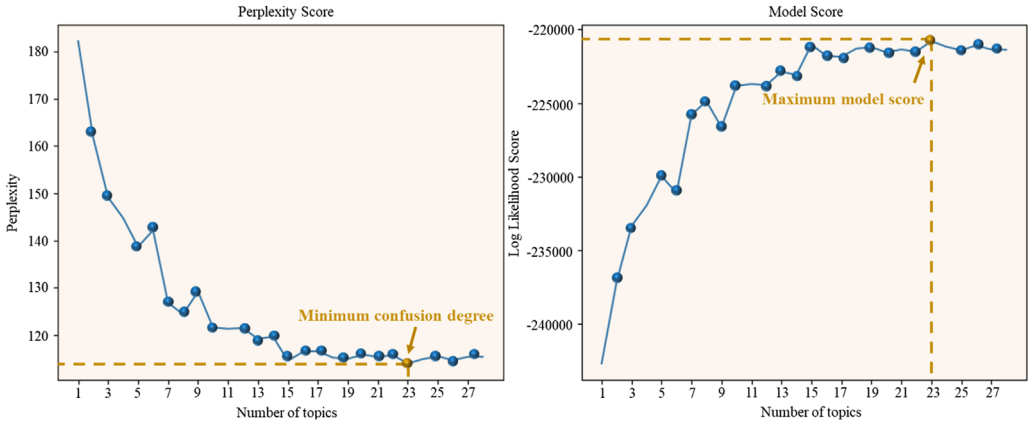

FIGURE 9
Perplexity curve and topic coherence score for the 2025 Xigaze (Tibet) earthquake dataset.

management department", "People's Daily", "early warning" and other words account for a relatively high proportion, which fully reflects the government's active organization and development of post-disaster reconstruction, and timely sending earthquake aftershock warning information to the masses. While restoring normal production and life in the affected areas as soon as possible, we will do our best to minimize the impact of aftershocks on secondary disasters in the disaster areas.

3.3 Overall data trend analysis of earthquake disasters

In the process of earthquake disaster emergency response, disaster monitoring and situation assessment are the core work of the emergency management department. Based on the theory of spatiotemporal data analysis, this study selects the 1·07 earthquake disaster in Tibet as a typical case, constructs a disaster information

TABLE 4 Topic-word probability distribution (Top 10 words).

Topic 1		Topic 2		Topic 3		···	Topic 23	
Emergency	0.2698	Rescue	0.1479	Work	0.1295	···	Rescue	0.1665
Country	0.1382	Service Desk	0.1221	Rescue	0.1212	···	Fire protection	0.1379
Disaster	0.0888	Disaster area	0.1200	Disaster relief	0.0610	···	Personnel	0.1288
Management Department	0.0724	Public welfare	0.1002	Earthquake region	0.0531	···	Team	0.0461
Hospital	0.0402	Pray for blessings	0.0537	Tent	0.0470	···	Reporter	0.0288
Disaster relief	0.0361	Energy	0.0513	Dingri	0.0445	···	Power	0.0276
Office	0.0284	Materials	0.0501	The masses	0.0409	···	House	0.0268
Start	0.0283	Get through the hard times)	0.0495	Region	0.0372	···	Scene	0.0254
Girl	0.0260	Power	0.0444	Temperature	0.0358	···	Brigade	0.0232
Rescue	0.0226	Era	0.0425	Scene	0.0330	···	Headquarters	0.0219

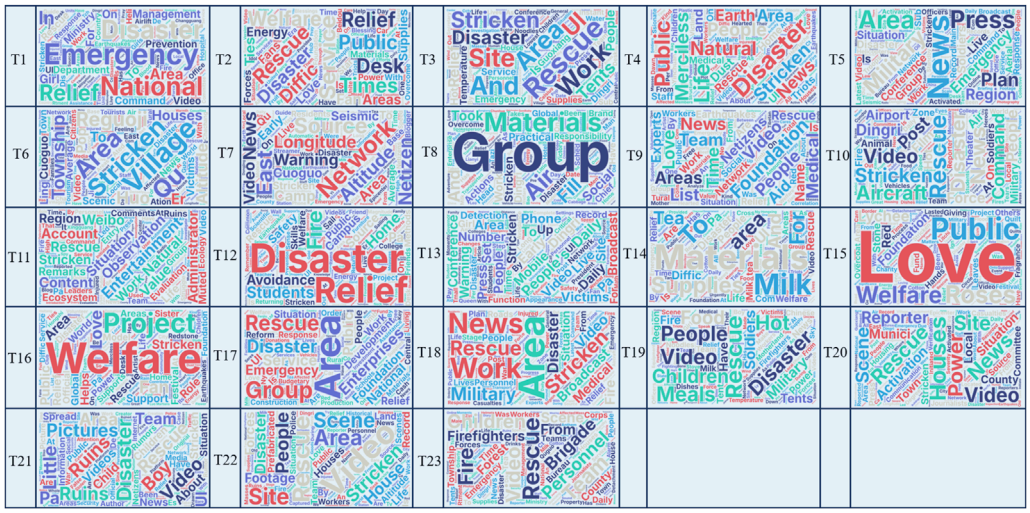

FIGURE 10
Word cloud analysis of "Tibet Earthquake" discourse topics.

dataset with the time granularity of days, and analyzes the release time of relevant Weibo data to analyze the overall trend of disaster data, as shown in Figure 13.

Stage 1 (2025-01-07~2025-01-08): In the early stage of the earthquake disaster, the disaster is sudden and unpredictable, causing widespread panic among the public. Affected people tend to seek help by releasing information through Weibo, resulting in a rapid growth in the amount of information on the platform. According to the life cycle theory of disaster management, this stage corresponds to a typical "emergency response period." 1) It is necessary to start the disaster rescue mechanism quickly; 2) Effective implementation of psychological intervention strategies to stabilize the masses' emotions; 3) Establish efficient information communication channels. The efficiency of emergency response at this stage will directly affect the overall effect of the follow-up earthquake relief work.

Stage 2 (2025-01-08~2025-01-09): In the middle stage of earthquake disaster evolution, with the full launch of the national emergency rescue system and the systematic intervention of professional rescue forces, the disaster situation has been significantly controlled. The mode of disaster information dissemination on social media platforms has undergone structural changes, and the frequency of emergency information released by individuals through these platforms, such as Weibo, has shown a downward trend. This stage belongs to the "emergency rescue period", and such changes are due to: 1) the government-led emergency rescue system has achieved full coverage of the affected areas; 2) Effective operation of the infrastructure repair and material

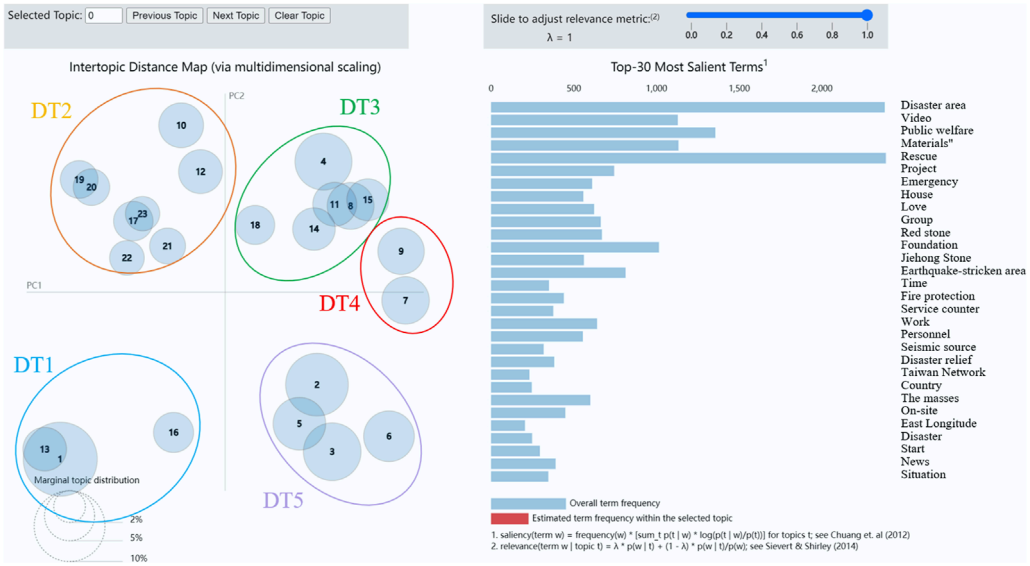

FIGURE 11
Visualization of the topic-word distribution results.

FIGURE 12
Cloud map of new topic terms for disasters. (Data source: Weibo topic dataset related to the 2025 Xigaze earthquake).

allocation mechanism; 3) improvement of the disaster information disclosure system. At the same time, the security level of the affected people has been substantially improved, which marks the gradual transition of disaster response from the emergency rescue phase to the recovery and reconstruction phase.

Stage 3 (2025-01-09~2025-01-10): In the late stage of earthquake rescue, the rescue forces of social organizations and the national professional rescue system form an institutionalized cooperation mechanism to maximize the rescue efficiency through resource integration and information sharing. With the gradual restoration of order in the disaster area and the improvement of the information disclosure mechanism, the psychological stress response of the affected people showed a trend of mitigation, and the discussion heat and information flow of topics related to the Tibet earthquake on

FIGURE 13
Statistics chart of the daily number of Weibo posts about the "Tibet Earthquake".

the Weibo platform returned to the baseline level. According to the disaster management cycle theory, this stage has achieved a smooth transition from emergency rescue to recovery and reconstruction, marking the late stage of emergency response with "system recovery and capacity reconstruction" as its core.

To sum up, this experiment visually displayed the disaster situation, emergency measures, post-disaster recovery and rescue work of Tibet's 1·07 earthquake disaster by classifying the topics of Weibo's topic data and visualizing the word cloud results. It analyzed the different characteristics of disasters and the differences in disaster responses during the initial, middle, and later stages of the earthquake disaster. The research results provided scientific guidance for follow-up rescue forces and resource scheduling, as well as for post-disaster recovery and reconstruction work, serving as a reference for emergency rescue and disposal of similar disasters and accidents.

3.4 Sentiment analysis research

Sentiment analysis is the process of analyzing, processing, summarizing and reasoning subjective texts with emotional color (Zhao et al., 2010). Sentiment analysis based on social media review data can effectively quantify the public sentiment tendency. By constructing the emotional time evolution model, it can accurately identify the development trend of public opinion, which is very important for disaster emergency management and public opinion control. Because the comment data crawled from the Rednote platform has the characteristics of short text, colloquial, multi-form and obscure emotional expression, the hybrid sentiment analysis method of "SnowNLP + Dictionary" has the generalization ability of combining the statistical model with the domain dictionary and the high-precision recognition ability for the emotional words in a specific domain. Therefore, this paper employs a hybrid sentiment analysis method to analyze disaster event comment data obtained from the Rednote platform.

3.4.1 Hybrid sentiment analysis method

To enhance the effectiveness of emotion analysis and align with the practical application scenario of earthquake disasters, a mixed emotion analysis method is employed, combining a user-defined dictionary with the SnowNLP emotion analysis algorithm. This approach optimizes and analyzes the results of single-emotion analysis.

1. The SnowNLP pre-trained model established based on the general corpus has a relatively low accuracy rate and obvious regional language differences in the sentiment analysis task in the field of earthquake disasters. Therefore, the experiment does not rely entirely on the default sentiment analysis of SnowNLP. Instead, it combines a user-defined sentiment dictionary (Table 5). It expands SnowNLP by adding words related to the disaster domain, including colloquial expressions (such as "safety", "casualties") and specific regional terms (such as "Xigaze", "Tibet"). To reduce the influence of language and regional differences on the analysis results, thereby making the sentiment analysis more in line with the actual context of earthquake disasters.
2. The emotion score is the basic index for judging the emotion polarity, and the calculation formula is shown in formula 2. The default SnowNLP evaluation score is 0.5, which serves as the boundary, and the classification effect is poor. To verify the rationality of the threshold, 500 pieces of data were randomly selected for manual sentiment annotation. After comparing the classification results under different thresholds (Figure 14), it was found that the default threshold of 0.5 is prone to judging neutral texts as positive, resulting in a higher proportion of positive emotions; When the threshold is adjusted to 0.65 (positive) and 0.35 (negative), the consistency between the model and manual annotation is the highest, and positive and negative emotions can be better distinguished. Therefore, the experiment adopted 0.65/0.35 as the emotion threshold for the emotion classification experiment.

TABLE 5 Part custom sentiment words.

Positive words		Negative words		Neutral words	
Come on	Safe and sound	Collapse	Casualties	Earthquake	Epicenter
Prayer	Bless	Danger	Ruins	Aftershock	Magnitude
Hope	Persist	Tragedy	Fear	Report	Location
Unity	Salute	Pain	Mourning	Warning	Time
Strong	Reconstruction	Unfortunate	Panic	Monitoring	Depth
Restore	Rebirth	Worry	Terrifying	Data	Expert
United as one	Rebuild the homeland	Cut off water and electricity	Traffic disruption	Altitude	Planning
One side is in trouble	Support from all directions	Communication interruption	Lifeline is interrupted	Section	Resettlement
…	…	…	…	…	…

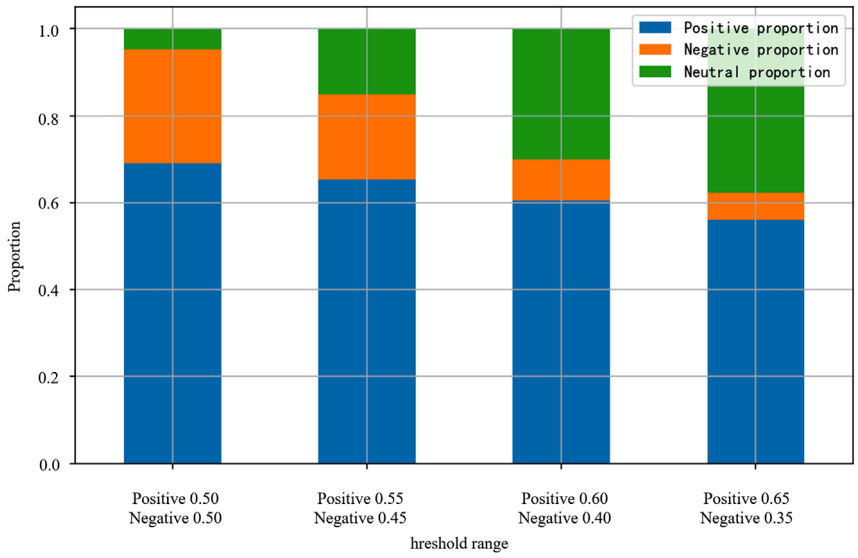

FIGURE 14
Comparison of threshold classification results.

The experimental improvement evaluation standard is: positive >0.65 to reduce the misjudgment rate; negative <0.35, to improve the severity of discrimination and further enhance the robustness of emotion classification. The overall emotional distribution of the Rednote earthquake disaster comment data is shown in Figure 15, where positive emotional words accounted for 56.1%, while negative emotional words accounted for only 6.3%. This distribution characteristic indicates that, during the 1·07 earthquake disaster event in Tibet, the public as a whole exhibited an optimistic attitude, and the development of public opinion was generally favourable, with significant positive characteristics.

$$S = \sigma\left(\sum_{i=1}^{n} logP(wi|pos) - \sum_{i=1}^{n} logP(wi|neg) + log\frac{P(pos)}{P(neg)}\right) \quad (2)$$

S: Emotional score, range [0,1], the closer to 1, the more positive, and the closer to 0, the more negative; *P* (*pos*): the prior probability of the positive category in the training data; *P* (*neg*): the prior probability of the negative category in the training data; *wi*: the *i*th word after text segmentation; *P* (*wi*|*pos*): conditional probability of word *wi* in positive category; *P* (*wi*|*neg*): conditional probability of word *wi* in negative category; $\sum_{i=1}^{n} logP$: sum the logarithmic probabilities of all words in the text.

3. Combining snownlp emotion score (snownlp_score) and dict_score (dict_score), the weighted average (0.7*snownlp_score+0.3*dict_score) is used to obtain the final emotion score, which improves the accuracy of emotion analysis and domain professionalism.

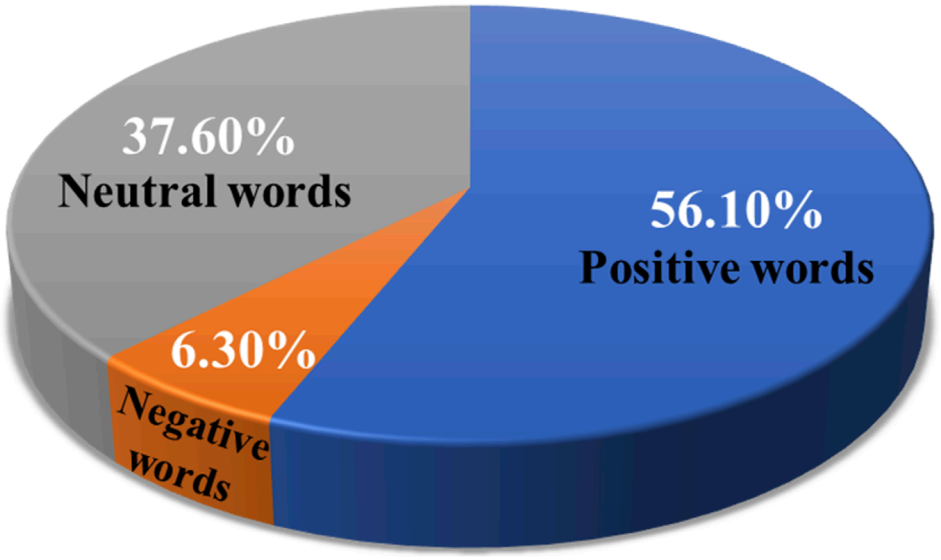

FIGURE 15
Sentiment Distribution of "Tibet Earthquake" data.

3.4.2 Sentiment time series analysis

To deeply explore the dynamic evolution characteristics of public sentiment, this experiment conducts time series analysis and Research on emotional data, investigating the temporal change law of social public sentiment in disaster events, and then reveals the phased characteristics of public sentiment evolution.

1. Sentiment score time series analysis:

Based on the big data of social public opinion from January to June 2025, the emotional score time series analysis was completed by using dual time dimension analysis and multi-angle aggregation technology, as shown in Figure 16. The division of temporal stages for sentiment evolution followed the disaster management lifecycle framework, which includes emergency response, rescue operations, and recovery/reconstruction. This division was informed by both official disaster response reports and inflection points in public opinion volume and sentiment distribution. The study found that after the 1·07 earthquake in Tibet, social emotions showed obvious stage characteristics: in the emergency response period (January and February) after the disaster, although facing major disasters, the emotional changes were relatively stable, and the overall social people showed a significant positive emotional tendency (the average emotional score was higher than 0.65). At this time, the positive guidance of public opinion is mainly attributed to the timely start of the rapid response mechanism of the national emergency management department, the scientific dispatch of rescue forces, and the disaster notification system with open and transparent information, which has formed a benign disaster relief environment dominated by the government, coordinated by society and participated by the public. During the post-disaster recovery period (March to May), emotional fluctuations were noticeable; however, the overall average emotional score remained higher than 0.35. Social emotions remained stable, reflecting the public's continued recognition of post-disaster reconstruction; However, in April, there was an obvious low emotional value (the lowest value was 0.2), showing a negative attitude. According to the analysis of the source of public opinion, this phenomenon was due to the dispute over emergency resource allocation caused by the video of "The police escorted the ambulance by motorcycle" on April 8. This phenomenon highlights the optimization space for emergency resource allocation in specific situations. When formulating rescue plans for the same type of disaster events, more attention should be paid to the balance between resource utilization efficiency and social perception. Overall, the social-emotional evolution of the earthquake in Tibet effectively reflects the maturity of the national emergency management system and provides an essential reference for enhancing the public opinion response strategy for major disasters.

2. Sentiment score heat analysis:

Using the social public opinion data during the earthquake in Tibet, an hour week two-dimensional emotional heat map (as shown in Figure 17) was constructed to deeply explore the temporal and spatial distribution of public emotion fluctuations, in which red represents>0.75 for high positive emotions and yellow represents<0.5 for negative emotions.

3.4.2.1 Week length dimension

Mid-week peak: the fluctuation of public sentiment shows obvious cyclical characteristics. A significant emotional peak is formed on Tuesday (0.68 ± 0.07) and Wednesday (0.67 ± 0.05), which is closely related to the positive progress of relief supplies in place and the construction of temporary resettlement sites in the middle period (the third to fifth days) after the disaster. The duration of the peak is highly consistent with the "golden 72 h" window for rescue, suggesting a key sensitive period for post-disaster emotional recovery. Weekend rebound: the emotional score on Sunday (0.64 ± 0.10) increased by 7% compared with Friday (0.60 ± 0.08), revealing the "leisure effect" of public attention to disasters - during the rest days, people have more information contact time and cognitive

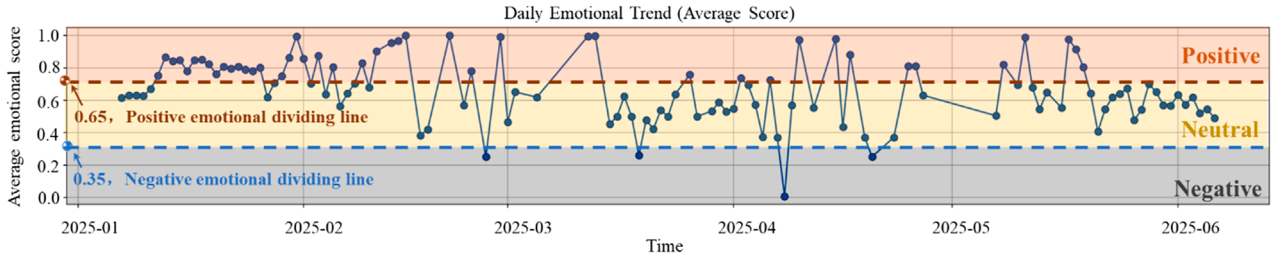

FIGURE 16
Analysis of the Emotional Time Series of "Tibet Earthquake" data.

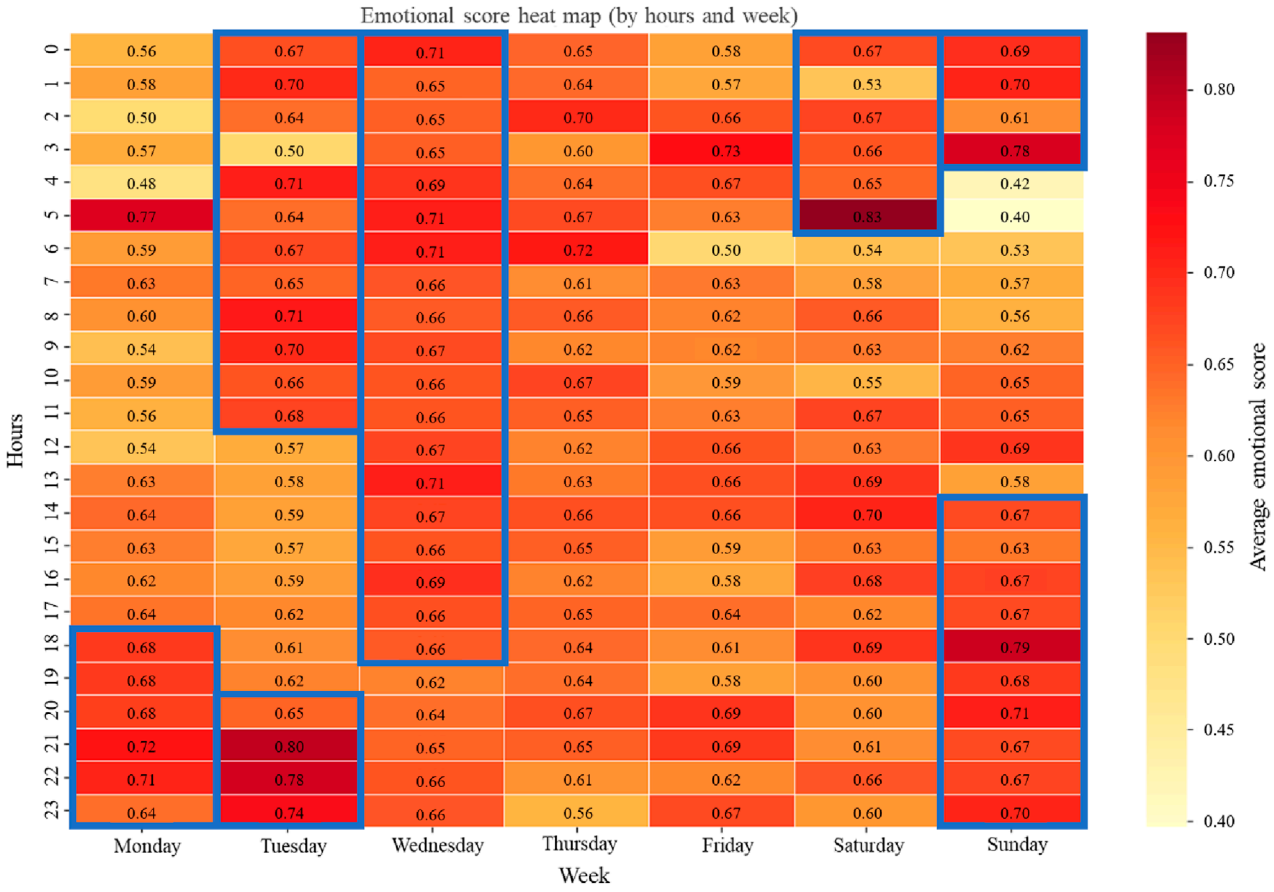

FIGURE 17
Emotional Thermal Analysis of the "Tibet Earthquake" data.

resources to deal with disaster reconstruction information, which promotes the systematic deviation of emotional evaluation.

3.4.2.2 Hour short dimension

Emotional fluctuations showed a significant "two peaks and one valley" circadian rhythm. Morning and evening bimodal: As social users browse content late at night, they may only forward, comment, or search the next morning. The emotional information exposed at night, after sleep-dependent memory, is given a more substantial emotional weight the next day, and for other reasons. The first emotional peak (0.69 ± 0.04) appeared in the information precipitation period at 0-6, which was consistent with the delayed feedback effect of social media information dissemination at night. The second peak appeared in the social activity period (0.67 ± 0.05) from 18:00 to 23:00, which was the "golden time" of public catalytic behavior. The attention of social media users was the most concentrated, interaction was the most active, communication efficiency was the highest, and emotional interaction was the most frequent, which was the most critical time node in sentiment analysis research. The peak value (0.90) occurred at 18:00 on Tuesday,

which was attributed to the fact that the disaster relief special report broadcast by CCTV's "Topics in Focus" generated widespread positive resonance, and the emotional score increased rapidly. Midday emotional trough: A significant decline in the emotional index (0.58 ± 0.12) was observed at 12-14, which corresponded to the stage of cognitive fatigue and information overload. Public emotional expression tended to be negative, with an average score for emotion being low.

3.4.2.3 Improvement of public opinion research

Based on the above cyclical characteristics of emotional fluctuation, when analyzing the social public opinion fluctuation of similar disaster data in the future, high-quality comments on Tuesday, Wednesday and weekend evenings (18o-23 o'clock) should be taken as a priority in the data collection stage to provide more representative sample data for sentiment analysis and effectively improve the efficiency of sentiment analysis; In view of the persistent negative emotion fluctuations on Thursday and at noon, the algorithm is used to identify and timely push the disaster notification, rescue progress and other information released by the authority department, strengthen positive and positive guidance, curb the spread of negative emotions, and further control the development of public opinion.

4 Implementation of the LDA-SnowNLP composite framework

To implement the LDA-SnowNLP combined framework in the real-time early warning system and break through the constraints of latency, computational efficiency and data quality in large-scale deployment, the four-layer architecture of "data access → preprocessing → cluster computing → early warning closed loop" is carried out (Figure 18). The core logic is as follows:

1. Data access layer

Distributed nodes are used in collaboration with Kafka to collect multi-source data. Anti-crawling is avoided through government affairs APIs and dynamic IP pools to ensure that the crawling delay is no more than 5 min and the frequency is once every 10 min. Weibo focuses on the 72 h after the disaster, while Rednote covers data from the past 6 months.

2. Preprocessing layer

Stream cleaning, deduplication (SimHash repetition rate ≤5%), word segmentation, and stop word filtering are implemented based on Spark Streaming. Combined with the rumor base and glossary, the data reliability is improved to more than 95%.

3. Cluster computing layer

Incremental LDA: Reuse pre-trained topics, process 100,000 pieces of data within no more than 5 min, and trigger a warning when the topic proportion suddenly increases by 30%.

Optimize SnowNLP: Integrate basic sentiment score with dictionary matching score, process 100,000 entries within no more than 3 min, and trigger a warning if the sentiment score is less than 0.35 and lasts for 1 h.

The GPU + Spark cluster supports end-to-end processing of millions of data points within no more than 8 min.

4. Early warning closed-loop layer

Build a theme-emotion association matrix and output three-level warnings of red (≤5 s), yellow (≤10 s), and green (≤30 s). The stability of the system is guaranteed through a triple closed loop of delay, efficiency, and quality: timeout optimization scheduling, dynamic GPU expansion, and reverse tuning when the sampling verification accuracy is lower than 90%.

5 Conclusion

To quickly and sustainably obtain disaster information and its evolution trend throughout the entire process of disaster occurrence and development, and to deeply explore the social impact mechanism of the 1·07 earthquake disaster in Tibet, this research summarizes the research status of topic models over the past 10 years. It conducted retrieval analysis, topic classification, emotion analysis, and result visualization research on the 1·07 disaster in Tibet through multidisciplinary research methods, combining Weibo topic data and Rednote review data. By accurately extracting five clustering topics with "emergency, rescue, materials, source and news" as the primary keywords, this paper deeply analyzes the disaster situation, emergency measures, disaster information release, rescue direction and reconstruction strategies in the area where the earthquake occurred in Tibet, and further uses the comment data related to the disaster events on the Rednote platform to calculate the emotional score by using the mixed sentiment analysis method of "SnowNLP + Dictionary". Through the time series analysis of 6-month emotional scores, this paper continuously reveals the characteristics of social people's emotional response to the January 7 earthquake in Tibet and the evolution trend of public opinion. The research results not only expand the technical path of earthquake and geological disaster prevention but also innovatively apply emotional computing to the field of disaster public opinion monitoring, providing a scientific basis and a practical paradigm for improving the geological disaster emergency management system and enhancing social public opinion management capabilities. It has significant academic value and practical significance for promoting the modernization of geological disaster risk management and public security governance in the future.

In the future, to comprehensively improve the ability of earthquake disaster prevention, monitoring and emergency response, and establish an efficient and accurate disaster prevention and control system, it is necessary to strengthen technology research and application in the following aspects:

1. Application of multimodal data: This study only relies on the text data of Weibo and Rednote platforms, and thus may have problems such as platform bias and short text sparsity. Future research can further integrate multimodal data such as geographic marker images, videos, and audio to construct a cross-modal joint analysis framework. The main challenges lie in: the heterogeneity of data preprocessing, the complexity of cross-modal feature alignment, the incompleteness of spatiotemporal information, and the scalability of large-scale

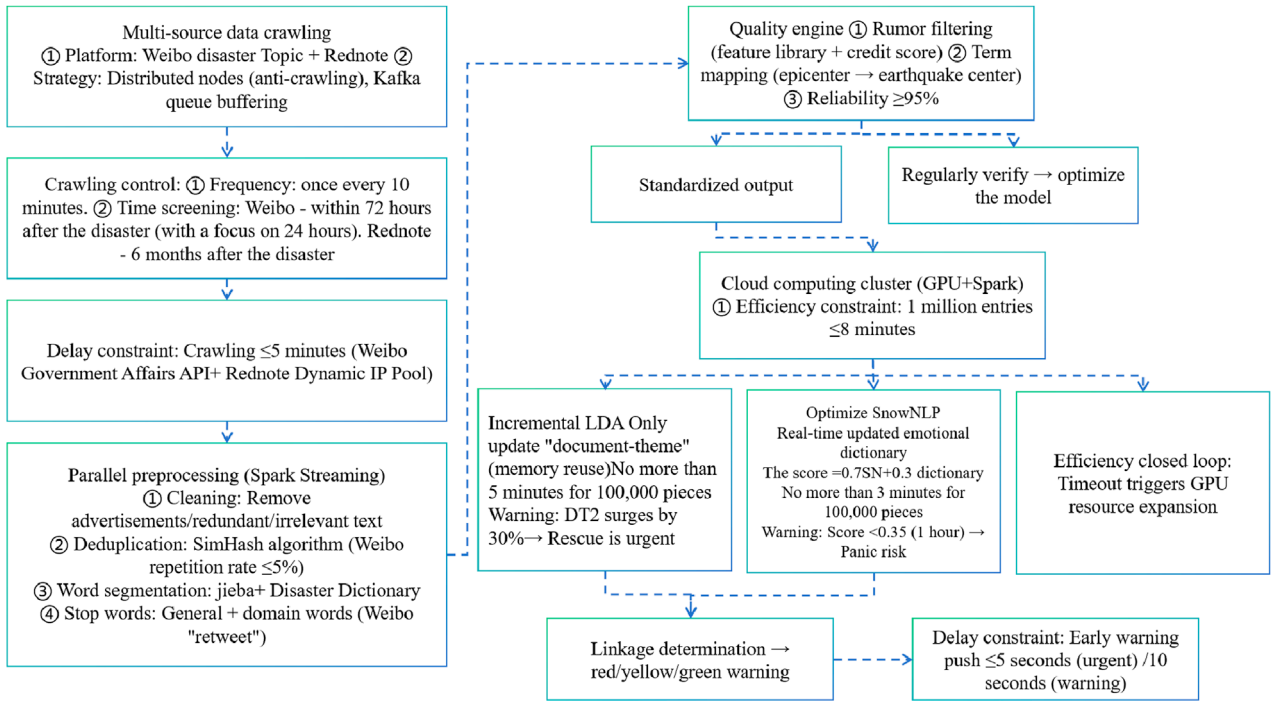

FIGURE 18
Shows the hierarchical architecture and constraint closed-loop process of the LDA-SNOW NLP real-time early warning system under large-scale deployment.

real-time processing. Solving these problems will help drive disaster monitoring from a single text analysis to a new paradigm of multi-modal intelligent recognition.

2. Model improvement: sentiment analysis research only uses the SnowNLP sentiment analysis algorithm with high cost-performance. Although the demand for computing resources is low and the speed is high, there are performance bottlenecks in terms of analysis accuracy, processing complex semantic relationships, and handling large-scale data applications. Future research will introduce a pre-training language model based on the transformer architecture, enabling a leap from surface emotion recognition to deep semantic understanding, and expanding the application depth of research in disaster public opinion monitoring, emergency decision support, and other fields.
3. Limitations of a single case: This study focuses solely on the Tibet earthquake, and its conclusions may be constrained by the suddenness of earthquakes and the region's unique geographical and cultural characteristics, which differ markedly from gradual disasters such as floods and typhoons. Future research should build a multi-disaster case library and conduct cross-case comparisons to generalize findings and enhance applicability.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

GC: Conceptualization, Methodology, Writing – original draft. YaW: Writing – original draft, Data curation, Software. DC: Writing – review and editing, Supervision, Formal Analysis. KW: Investigation, Writing – review and editing, Validation. YeW: Investigation, Writing – review and editing, Software. YoW: Writing – review and editing, Software.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This research was funded by the National Natural Science Foundation of China (42377200); the Natural Science Foundation of Hebei Province, China (D2025508013); the Fundamental Research Funds for the Central Universities (3142025033); the Open Fund of Hebei Cangzhou Groundwater and Land Subsidence National Observation and Research Station (No. CGLOS-2025-05); the Central Government Guided Local Science and Technology Development Fund (226Z5404G).

Acknowledgments

The author would like to thank Shi Lei for his valuable suggestions on optimizing the structure of the article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Amiresmaili, M., Zolala, F., Nekoei-Moghadam, M., Salavatian, S., Chashmyazdan, M., Soltani, A., et al. (2021). Role of social media in earthquake: a systematic review. *Iran. Red Crescent Med. J.* 23 (5), e447. doi:10.32592/ircmj.2021.23.5.447

Ankner, Z., Blakeney, C., Sreenivasan, K., Marion, M., and Paul, M. (2024). Perplexed by perplexity: perplexity-Based data pruning with small reference models. *arXiv Prepr. arxiv:2405.20541*. doi:10.48550/arXiv.2405.20541

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *J. Mach. Learn. Res.* 3, 993–1022. doi:10.5555/944919.944937

Cao, Z. Y., and Liu, X. M. (2019). Socialized marketing integration of magazines in the Mobile internet environment. *Publ. J.* 27 (6), 87. doi:10.13363/j.publishingjournal.2019.06.030

Cao, Z., Li, S., Liu, Y., and Ji, H. (2015). A novel neural topic model and its supervised extension. *Proc. AAAI Conf. Artif. Intell.* 29 (1), 2210–2216. doi:10.1609/aaai.v29i1.9499

Cemiloglu, A., Zhu, L., Mohammednour, A. B., Azarafza, M., and Nanehkaran, Y. A. (2023). Landslide susceptibility assessment for maragheh county, Iran, using the logistic regression algorithm. *Land* 12 (7), 1397. doi:10.3390/land12071397

Cheng, G., You, Q., Li, G., Li, Y., Yang, D., Wu, J., et al. (2024). Research on the application of topic models based on geological disaster information mining. *Information* 15 (12), 795. doi:10.3390/info15120795

Contreras, D., Wilkinson, S., Balan, N., and James, P. (2022). Assessing post-disaster recovery using sentiment analysis: the case of l'Aquila, Italy. *Earthq. Spectra* 38 (1), 81–108. doi:10.1177/87552930211036486

Das, R., Zaheer, M., and Dyer, C. (2015). Gaussian LDA for topic models with word embeddings. *Assoc. Comput. Linguistics 7th Int. Jt. Conf. Nat. Lang. Process.* 1, 795–804. doi:10.3115/v1/P15-1077

Du, Z. Q., Li, Y., Zhang, Y. T., Tan, Y. Q., and Zhao, W. H. (2020). Knowledge graph construction method on natural disaster emergency. *Geomatics Inf. Sci. Wuhan Univ.* 45 (9), 1344–1355. doi:10.13203/j.whugis20200047

Eligüzel, N., Çetinkaya, C., and Dereli, T. (2023). Comparative analysis with topic modeling and word embedding methods after the aegean sea earthquake on Twitter. *Evol. Syst.* 14 (2), 245–261. doi:10.1007/s12530-022-09450-4

Ghaly, M. Z., and Laksito, A. D. (2023). "Topic modeling of natural disasters in Indonesia using NMF," in *2023 eighth international conference on informatics and computing*, 1–6. doi:10.1109/ICIC60109.2023.10382064

Hananto, V. R., Serdült, U., and Kryssanov, V. (2022). A text segmentation method for automated annotation of online customer reviews, based on topic modeling. *Appl. Sci.* 12 (7), 3412. doi:10.3390/app12073412

Hassan, S. Z., Ahmad, K., Hicks, S., Halvorsen, P., Al-Fuqaha, A., Conci, N., et al. (2022). Visual sentiment analysis from disaster images in social media. *Sensors* 22 (10), 3628. doi:10.3390/s22103628

Huang, Z. Y., Mo, G. Q., and Yu, K. M. (2024). General text matching based on topic model. *Comput. Appl. Softw.* 41 (5), 310–318. doi:10.3969/j.issn.1000-386x.2024.05.045

Jacobi, C., Van, A. W., and Welbers, K. (2018). Quantitative analysis of large amounts of journalistic texts using topic modelling. *Digit. Journal.* 4 (1), 89–106. doi:10.4324/9781315115047-7

Ji, W., Meng, X., and Zhang, Y. (2022). SPATM: a social period-aware topic model for personalized venue recommendation. *IEEE Trans. Knowl. Data Eng.* 34 (8), 3997–4010. doi:10.1109/TKDE.2020.3029070

Koltcov, S., Surkov, A., Filippov, V., and Ignatenko, V. (2024). Topic models with elements of neural networks: investigation of stability, coherence, and determining the optimal number of topics. *PeerJ Comput. Sci.* 10, e1758. doi:10.7717/peerj-cs.1758

Kowald, D., Pujari, S. C., and Lex, E. (2017). "Temporal effects on hashtag reuse in Twitter: a cognitive-inspired hashtag recommendation approach," in *Proceedings of the 26th international conference on world wide web*, 1401–1410. doi:10.1145/3038912.3052605

Li, C., Wang, H., Zhang, Z., Sun, A., and Ma, Z. (2016). "Topic modeling for short texts with auxiliary word embeddings," in *Proceedings of the 39th international ACM SIGIR conference on research and development in information retrieval*, 165–174. doi:10.1145/2911451.2911499

Li, M., Bao, L., Hu, Y., Cheng, S., Hu, X. B., and Gao, Y. (2022). Implementation of random number online detection method based on chi square test. *Microelectronics* 52 (3), 388–392. doi:10.13911/j.cnki.1004-3365.210329

Li, A., Li, Y., Shao, Y., and Liu, B. (2023). Multiview scholar clustering with dynamic interest tracking. *IEEE Trans. Knowl. Data Eng.* 35 (9), 9671–9684. doi:10.1109/TKDE.2023.3248221

Liu, B., Zhang, P., Lu, T., and Gu, N. (2020). A reliable cross-site user generated content modeling method based on topic model. *Knowledge-Based Syst.* 209, 106435. doi:10.1016/j.knosys.2020.106435

Liu, C., Du, J. P., and Zhou, N. (2021). A cross media search method for social networks based on adversarial learning and semantic similarity. *Sci. China Inf. Sci.* 51 (5), 779–794. doi:10.1360/ssi-2019-0120

Luo, J., Wang, L. H., Tu, S. S., Song, G., and Han, Y. (2021). Analysis of public sentiment tendency in sudden meteorological disasters based on LSTM-BLS. *J. Nanjing Univ. Inf. Sci. and Technol.* 13 (4), 477–483. doi:10.13878/j.cnki.jnuist.2021.04.014

Meng, Q., and Xiong, H. (2021). A doctor recommendation based on graph computing and LDA topic model. *Int. J. Comput. Intell. Syst.* 14 (1), 808–817. doi:10.2991/ijcis.d.210205.002

Murshed, B. A. H., Mallappa, S., Abawajy, J., Saif, M. A. N., Al-Ariki, H. D. E., and Abdulwahab, H. M. (2023). Short text topic modelling approaches in the context of big data: taxonomy, survey, and analysis. *Artif. Intell. Rev.* 56, 5133–5260. doi:10.1007/s10462-022-10254-w

Nanehkaran, Y. A., Mao, Y. M., Azarafza, M., and Kockar, M. (2021). Fuzzy-based multiple decision method for landslide susceptibility and hazard assessment: a case study of tabriz, Iran. *Geomechanics Eng.* 24 (5), 407–418. doi:10.12989/gae.2021.24.5.407

Nanehkaran, Y. A., Zhu, L. C., Chen, J. D., Mohammad, A., and Mao, Y. M. (2022). Application of artificial neural networks and geographic information system to provide hazard susceptibility maps for rockfall failures. *Environ. Earth Sci.* 81 (19), 475. doi:10.1007/s12665-022-10603-6

Park, Y., Lim, S., Gu, C., Syafiandini, A. F., and Song, M. (2025). Forecasting topic trends of blockchain utilizing topic modeling and deep learning-based time-series prediction on different document types. *J. Inf.* 19 (2), 101639. doi:10.1016/j.joi.2025.101639

Pavithra, C. B., and Savitha, J. (2024). Topic modeling for evolving textual data using LDA HDP NMF BERTOPIC and DTM with a focus on research papers. *J. Technol. Inf.* 5 (2), 53–63. doi:10.37802/joti.v5i2.618

Peng, X., Xu, Q., and Gan, W. (2021). SBTM: a joint sentiment and behaviour topic model for online course discussion forums. *J. Inf. Sci.* 47 (4), 517–532. doi:10.1177/0165551520917120

Ruan, T., Kong, Q., McBride, S. K., Sethjiwala, A., and Lv, Q. (2022). Cross-platform analysis of public responses to the 2019 ridgecrest earthquake sequence on Twitter and Reddit. *Sci. Rep.* 12 (1), 1634. doi:10.1038/s41598-022-05359-9

Shao, X., Tang, G., and Bao, B. K. (2019). Personalized travel recommendation based on sentiment-aware multimodal topic model. *IEEE Access* 7, 113043–113052. doi:10.1109/ACCESS.2019.2935155

Shi, L., Du, J. P., Liang, M. Y., and Kou, F. F. (2019). SRTM: a sparse RNN-Topic model for discovering bursty topics in big data of social networks. *J. Inf. Sci. and Eng.* 35 (4), 749–767. doi:10.6688/JISE.201907_35(4).0003

Shi, L., Song, G., Cheng, G., and Liu, X. (2020). A user-based aggregation topic model for understanding user's preference and intention in social network. *Neurocomputing* 413, 1–13. doi:10.1016/j.neucom.2020.06.099

Toubia, O. (2021). A poisson factorization topic model for the study of creative documents (and their summaries). *J. Mark. Res.* 58 (6), 1142–1158. doi:10.1177/0022243720943209

Tu, S., and Yang, B. (2021). Research on sentiment classification of microblog short text based on topic clustering. *J. Phys. Conf. Ser.* 1827 (1), 012160. doi:10.1088/1742-6596/1827/1/012160

Wang, X. M. (2021). *Feature research and visual analysis of high-dimensional landslide and debris flow disaster data*. Beijing: Beijing Jiaotong University. doi:10.26944/d.cnki.gbfju.2021.000857

Wang, X., and Yang, Y. (2020). Neural topic model with attention for supervised learning. *Int. Conf. Artif. Intell. Statistics* 9, 1147–1156. Available online at: https://proceedings.mlr.press/v108/wang20c.html

Wang, Z., Chen, J., Chen, J., and Chen, H. (2024). Identifying interdisciplinary topics and their evolution based on BERTopic. *Scientometrics* 129 (11), 7359–7384. doi:10.1007/s11192-023-04776-5

Xing, Z., Su, X., Liu, J., Su, W., and Zhang, X. (2019). Spatiotemporal change analysis of earthquake emergency information based on microblog data: a case study of the 8.8 jiuzhaigou earthquake. *ISPRS Int. J. Geo-Information.* 8 (8), 359. doi:10.3390/ijgi8080359

Yang, Y., Hu, J., Liu, Y., and Chen, X. (2020). Doctor recommendation based on an intuitionistic normal cloud model considering patient preferences. *Cogn. Comput.* 12, 460–478. doi:10.1007/s12559-018-9616-3

Yao, L., Zhang, Y., Wei, B., Zhe, J., Chen, Q., et al. (2017). Incorporating knowledge graph embeddings into topic modeling. *Proc. AAAI Conf. Artif. Intell.* 31 (1), 3119–3126. doi:10.1609/aaai.v31i1.10951

Yin, H., Zhou, X., Cui, B., Wang, H., Zheng, K., and Nguyen, Q. V. H. (2016). Adapting to user interest drift for poi recommendation. *IEEE Trans. Knowl. Data Eng.* 28 (10), 2566–2581. doi:10.1109/TKDE.2016.2580511

Zhang, X. L. (2020). *Research on naive bayes classifiers and its improved algorithms*. Qingdao: Shandong University of Science and Technology. doi:10.27275/d.cnki.gsdku.2020.001131

Zhao, Y. Y., Qin, B., and Liu, T. (2010). Sentiment analysis: sentiment analysis. *J. Softw.* 21 (8), 1834–1848. doi:10.3724/sp.j.1001.2010.03832

Zhu, J. X., Zhang, L. Z., Zhou, X. Y., Liang, G. L., Wang, G., Cai, Z. Z., et al. (2012). Application of entropy-based grey model in geological hazard assessment-a case study of qingchuan county, Sichuan Province. *J. Catastrophology* 27 (1), 78–82. doi:10.3969/j.issn.1000-811X.2012.01.016

Zou, H., and Yang, Y. F. (2019). Transmutation, shaping and cultivation of college students' values in we media context. *J. South China Univ. Technol. Soc. Sci. Ed.* 21 (4), 118–124. doi:10.19366/j.cnki.1009-055X.2019.04.013