Journal of Computers Vol. 37 No. 5, October 2026, pp. 287-302
https://doi.org/10.63367/19911599202610370501 8

Research on the Dynamic and Communication Mode of Public Opinion Emotions Based on Text Mining and Sentiment Analysis

Chao-Peng Li, Li-Ning Hao, Jiang Shang*,
Xin-Yuan Wang, Jing-Wen Zhao, and Yan-Ming Zhang

Hebei Mechanical and Electrical Polytechnic,
Xingtai City 054000, Hebei Province, China
{lichaopeng, haolining, shangjiang, wangxinyuan,
zhaojingwen, zhangyanming}@hbjd.edu.cn

Received 10 July 2026; Revised 4 August 2026; Accepted 10 September 2026

Abstract. The swift dissemination of public opinions on university-related events is inevitably associated with topic variations, emotions, and complicated interpersonal interaction among the users. To explore this problem, this study proposes a hybrid framework including text mining, sentiment analysis, and opinion evolution modeling based on fuzzy theory. Data about public opinions will be collected on social platforms like Weibo, Douyin, and Xiaohongshu, then will be subjected to data cleaning, Chinese word segmentation, and stop-word elimination. An optimized version of BERTopic is employed to detect topics by means of domain knowledge mapping, keyword pre-exploration, topic labeling via assistance of large language models and fine-tuning of embedding models. Sentiment tendencies will be further divided into types with help of a model that combines RoBERTa-BiLSTM and interactive attention and biaffine feature fusion. Based on these analyses, a fuzzy opinion-evolution model is built to evaluate the impact of peer trust, opinion leader trust, timely and credible responses on emotions and consensus. A case of university cafeteria-related public opinion event will be adopted for empirical analysis. The findings indicate that the proposed framework can identify theme and emotion variations in campus public opinion. Higher opinion leader trust can speed up inter-group consensus while too much reliance on peer interactions may delay it. Late and low credibility responses are likely to increase public opinion risk.

Keywords: BERTopic, BiLSTM, campus public opinion

1 Introduction

Public opinion has existed since ancient times, without exception in both the East and the West. Whether it is individuals, groups, or society, or governments, media, or third-party organizations, they are directly or indirectly influenced by public opinion, which even changes and influences social public opinion and management decisions. In a sense, public opinion has become a resource, an influence, and even a representation of power or authority, playing an increasingly important and decisive role in modern society. In particular, with the Internet as the "information wholesale market" and "distribution processing center" of public opinion in the context of the digital era, the formation and evolution of online public opinion has been constantly promoted, affecting the adjustment and change of the network social pattern. Therefore, paying attention to public opinion, understanding public opinion, researching public opinion, grasping public opinion, guiding public opinion, researching and exploring the characteristics, relationships, logic, laws, mechanisms, etc. of public opinion, forming a "positive public opinion field", creating a positive public opinion atmosphere, has become a phenomenon that educators, managers, and relevant researchers cannot ignore and an issue that urgently needs special attention. Education is the great plan of the country and the party, and an important cornerstone of national rejuvenation and social progress. It is of decisive significance for improving the comprehensive quality of the people, promoting their all-round development, enhancing the innovation and creativity of the Chinese nation, and realizing the great rejuvenation of the Chinese nation [1]. The attention of the general public and social opinion to education continues to rise, with hot topics and discussions on education emerging one after another. The resulting education network public opinion events involve multiple levels and perspectives such as individuals, families,

* Corresponding Author

schools, and society, with wide coverage, strong resonance effects, and complex triggering points, becoming a highly concerned category of network public opinion. As an important content and component of educational network public opinion, university network public opinion has become the most important sub concept, component, and highest level of educational network public opinion in the digital empowerment network environment. The level of development of higher education is an important indicator of a country's development level and potential. China has built the world's largest higher education system, with a total enrollment of over 44.3 million people. The resonance with scale and quality is the prominent online public opinion of universities. Universities have become high-risk areas for online public opinion, involving various categories and forms of online public opinion events such as policy management, educational achievements, talent cultivation, discipline construction, academic research, innovation and entrepreneurship, teacher development, teacher ethics and conduct, anti-corruption in higher education, campus safety, ideology, and university life. The number of negative public opinion far exceeds the number of positive public opinion, reflecting the accumulated experience, problems, contradictions, and deep-seated issues in the transformation period, labor pains period, and management process of higher education reform, and indirectly reflecting the opinions, views, attitudes, and emotions of the public and media towards these experiences, problems, and contradictions [2].

With the rapid development of information technology, the Internet has gradually become an important medium for the diffusion of public opinion. Especially after the large-scale penetration of new forms of media such as social media, forums and blogs, the efficiency and coverage of information diffusion have been pushed to an unprecedented height. As an indispensable part of the social structure, the internal public opinion of universities is not only a reference indicator for the ideological status of teachers and students on campus, but also deeply related to social stability and progress. Recently, there have been frequent incidents of public opinion in universities, such as academic integrity crisis, campus safety accidents, and conflicts between teachers and students, which have attracted high attention from people from all walks of life. Therefore, how to accurately and efficiently grasp the changes in public opinion and emotions in universities, and implement effective guidance methods, has become a serious problem that urgently needs to be addressed in the governance of higher education and social governance. At present, the research on online public opinion in universities has become an important research field and hot research direction of concern in the academic community. The research space is relatively broad, and the research heat is constantly increasing. While related academic research has achieved some fruitful results, there are also some research shortcomings and problems exposed, such as research paradigms that either lean towards natural sciences or social sciences; The research content focuses more on theory than empirical evidence, and there is not much systematic research; Homogeneity research has more phenomenon of fitting the same model, and less research on heterogeneity analysis; The combination of theoretical research and management practice is not close enough; Emphasize information events over psychological and emotional analysis; Easy to label and classify without considering correlation; The proportion of public opinion research on group emergencies is not high, and there is a lack of research on public opinion management decision analysis. In addition, the research on online public opinion in universities is inevitably influenced by information technology in the digital age, providing new possibilities, paths, and methods for the upgrading and transformation of online public opinion research in universities. It will also become a turning point for improving the decision-making ability of online public opinion management in universities and a fulcrum for leveraging scientific decision-making.

Numerous interests, opinions, viewpoints, attitudes, and emotions are intertwined on online platforms, which to some extent affect and influence the public's cognition and attitudes, and gradually have a profound impact on various management decisions. Especially the negative online public opinion in universities reflects and deeply reflects the common problems, prominent issues, and general contradictions in the external management practices and internal mechanisms and systems of universities. By studying the online public opinion of universities, grasping and analyzing its research context, generation logic, evolution path, influence mechanism, and operation mechanism, optimizing relevant models, and proposing management strategies, it has important theoretical and practical significance [3].

This article focuses on emotional communication and text mining in current university public opinion. The work done is as follows:

1) First, relevant data is obtained from Weibo, Douyin and Xiaohongshu. All of the above applications provide open application program interfaces (APIs), which enable researchers to easily obtain relevant data, including user information, release time, content, interaction and other dimensions. Then, using octopus collector as a data collection tool, the octopus collector can easily convert the data in the web page into structured data according to specific rules [4].

2) After data extraction, it undergoes data cleaning, word segmentation, and stop word processing, and then

uses the BERTopic topic model to mine topics for a certain public opinion event in universities.

3) After extracting key texts and conducting sentiment analysis, this article constructs a fuzzy theory based public opinion evolution model based on the characteristics of public opinion in universities. By using the public opinion evolution model, users can help determine the direction of public opinion. In this process, Stanford University's publicly available social network dataset was used, which can further extract key indicators of mainstream social network sentiment labels.

2 Related Work

In terms of public opinion analysis and sentiment trends, many scholars have made corresponding achievements. Zhen Li, the types of public opinion events in universities are complex and diverse, covering teaching management, academic ethics, campus life, emergencies, student affairs, teacher-student relationships, and campus reactions to social hotspots It is believed that current social media platforms have become an important source of information for identifying public opinion events in universities, but university public opinion analysis still faces unique challenges such as sparse domain terminology, unique platform structures, and diverse event types [5].

Kun Han, combining BERT and TextCNN models, proposes a text classification model that integrates BERT multi-level features (BERT-MLFF TextCNN), and conducts emotional analysis on relevant comment text data in the Tiktok short video platform. The experimental results show that compared with the BERT-MLFF-TextCNN, GloVe TextCNN, and Word2vec TextCNN models, the BERT-MLFF-TextCNN model performs better, with an F1 value of 0.977. This model can effectively identify the emotional tendencies of public opinion on short video platforms. Based on this, the TextRank algorithm is used for topic mining to visualize the polarity of sentiment in public opinion comments, providing decision-making references for relevant departments' public opinion control work [6].

Fan Jiang used the OneKE big model knowledge extraction framework to extract event event relationship pairs based on the Beijing Marathon medal oolong event as the data source, combined with the BERTopic topic model for entity disambiguation and event fusion, and used the Neo4j graph database to visualize the event graph. Based on the event graph query analysis, the evolution path of public opinion was revealed, revealing the dynamic trend of public opinion changes. The results indicate that the event types of online public opinion in sports events include five types: organizational structure, event information, medal information, participants, and public response; The emotional evolution of public opinion events shows a trend of gradually turning from negative to positive under official response measures, but negative feedback from public emotions persists for a certain period of time; The dissemination of public opinion not only has the characteristics of temporal evolution, but also has significant spatial diffusion patterns [7].

Min Liu from Nanchang University, based on BERT The model analyzes the online public opinion on social media, converts the public opinion information into an intuitive fuzzy evaluation matrix, and analyzes the temporal changes of public opinion to calculate the weights of each temporal sequence. Then, the generative big language model is used to simulate the behavior of decision makers, and the news text and online public opinion opinions are provided to the big language model to simulate the opinions of decision makers with different risk preferences. Therefore, a multi-stage public opinion driven group consensus decision-making method is proposed. Finally, the maximum adjustment degree of decision makers is calculated by combining hesitation and trust, and the calculation method of attribute weights and consensus thresholds is given considering the individual differences of decision makers. A dual path feedback model is constructed to adjust the conflict problem between decision makers. Through the analysis of investment selection cases of new energy vehicle enterprises, this method is verified in the group. Feasibility and effectiveness in decision-making, This method has reference value for the analysis of public opinion in this article [8].

Xu Han from China Public Security University aims to study the diffusion mechanism of online public opinion. The research process combines actor network theory and the five forces model of online public opinion. Taking 56 domestic social emergencies from 2021 to 2022 as cases, the fuzzy set qualitative comparative analysis method is used to explore the configuration effect of factors affecting the dissemination of online public opinion in social emergencies. The research results show that a single variable cannot explain the formation of the phenomenon of high popularity online public opinion in social emergencies. In the configuration study, six conditional configurations for generating high popularity online public opinion were identified and summarized into five combination types, namely national identity type, public opinion hype type, high temperature impact

type, accumulation and storage type, and focus shift type. At the same time, the strong role of carrier display power in high popularity online public opinion was observed. Reflecting the characteristics of the driving force of the subject, In addition, other different combinations of conditions can also affect the popularity of online public opinion. Its innovation lies in combining actor network theory with the "Five Forces" model of online public opinion to construct a research framework, and for the first time introducing the condition variable of "overseas anti propaganda" [9].

The research objective of Xiaoning Gao from Zhengzhou University is to accurately and efficiently identify and analyze the key influencing factors and interaction mechanisms of online public opinion risks in universities [10]. Based on the literature research method, the theory of information dissemination process and the "Five Body Theory" are integrated to analyze the influencing factors of online public opinion risks in universities. Then, a system dynamics model is constructed to explore the complex causal relationships of various factors. Sensitivity analysis is conducted on subsystems and their secondary factors to explore the degree of influence of various factors on online public opinion risks in universities. The research results show that the development heat of public opinion events and the ability to control public opinion risks are the factors that have a significant impact on online public opinion risks in universities, and play different degrees of influence in different stages of risk evolution, there are significant differences in the impact of secondary factors of each subsystem on the risk of online public opinion in universities.

3 Data Processing and Thematic Analysis

The data collection platform selected in this paper is Weibo, Xiaohongshu, Tiktok's comment area, etc. From the perspective of data content, the above platform has a large user group, covering users of different ages, genders, regions, occupations and hobbies, which makes the information involved in colleges and universities widely representative and diverse. At the same time, the speed of information dissemination is also very fast, not only through direct forwarding by users, but also through re forwarding by users' followers, forming multi-level dissemination that can quickly spread to a large user group in a short period of time, further expanding the influence of information. At present, the users who use Weibo, Tiktok and Xiaohongshu in colleges and universities are also very active. They can interact with other users through comments, forwarding, likes and other ways. This interaction helps to form a strong public opinion field and promote the development of public opinion. From the perspective of data acquisition, Weibo, Tiktok and Xiaohongshu all provide open application program interfaces (APIs), enabling researchers to easily access relevant data, including user information, release time, content, interaction and other dimensions, providing rich data support for public opinion analysis. The public opinion information related to universities will constantly change over time and events, and this dynamism helps researchers track and analyze the development trend of public opinion [11].

This article uses an octopus collector as a data collection tool. With the help of an octopus collector, data in web pages can be easily converted into structured data according to specific rules. These data can be exported to common file formats such as Excel, CSV, HTML, JSON, and can also be output to a database. The software draws inspiration from human thinking in process configuration and mainly has four operation modes: custom mode, guide mode, intelligent mode, and simple mode, in order to meet the diverse needs of different researchers. Most researchers prefer to use the simplified mode, where users only need to select the corresponding module and set the parameters to easily and quickly obtain the required data, significantly improving the efficiency of data acquisition. The intelligent mode is also highly favored by researchers. Its main operation process is to input the website into a designated location and complete the relevant parameter settings to achieve fully automatic data acquisition.

1) Data cleaning. Although the system can automatically identify and filter out duplicate data generated by repeatedly grabbing page data during the data collection process when exporting data using the octopus collector, in addition, some netizens may repeatedly post the same content at different times, which will also affect the analysis results, so it also needs to be filtered and deleted. Many netizens like to add some emojis and symbols when posting content on the Weibo platform to reinforce their views, as these special characters can interfere with data analysis. The collected data may also contain blank text, link addresses, and some meaningless text, such as "forwarding Weibo", "sharing pictures", and "folding" that may occur when collecting long texts, which can affect the accuracy of topic mining and sentiment analysis and need to be deleted. This article attempts to use a more advanced data processing method - Autoencoder (AE), which is an unsupervised neural network that learns the intrinsic distribution of data through an encoding decoding mechanism. In data cleaning, AE trains

the model with normal samples and uses reconstruction errors to detect and repair anomalies [12]. The basic structure of AE includes an encoder and a decoder. The encoder maps high-dimensional input vectors to low dimensional vectors, and high-dimensional input vectors are represented as:

$$i = \left[i_1, i_2, \cdots, i_n \right] \tag{1}$$

The low latitude vector is represented as:

$$i_{low} = \left[i_{low,1}, i_{low,2}, \cdots, i_{low,m} \right] \tag{2}$$

The mapping process is represented as follows:

$$i_{low} = f\left(a\right)\left(\omega_{encoder} \cdot i + b_{encoder}\right) \tag{3}$$

The decoder reconstructs the low dimensional vector into the original dimensional output, which is represented as follows:

$$o = f\left(a\right)\left(\omega_{decoder} \cdot i_{low} + b_{decoder}\right) \tag{4}$$

In the above equation, $\omega_{encoder}$ represents the encoder weight, $\omega_{decoder}$ represents the decoder weight, $b_{encoder}$ represents the encoder bias, $b_{decoder}$ represents the decoder bias, and $f(a)$ is the activation function. During the training process, the model parameters are optimized by minimizing the difference between the reconstructed data o and the original input i. Mean Squared Error (MSE) is commonly used as the loss function, and its calculation expression is as follows:

$$f\left(a\right) = \frac{\sum_{t=1}^{n} \left\| i_t - o_t \right\|^2}{n} \tag{5}$$

Finally, use the pandas library and regular expression library (re) in Python to complete the above data cleaning.

2) Word segmentation. Word segmentation refers to the process of dividing a continuous sequence of words into individual words, which is crucial for subsequent topic mining and sentiment analysis research. For the collected Chinese data, suitable Chinese word segmentation tools need to be used for segmentation processing, this article calculates the similarity between any two adjacent parts of a sentence based on single word dependency and multi word dependency [13]. The similarity is used to determine whether to split two adjacent segments. The similarity is represented as follows:

$$s\left(a_m, a_n\right) = \frac{f\left(a_m, a_n\right) + f\left(a_n, a_m\right)}{2} \tag{6}$$

In the formula, $s(\ )$ is the similarity function and $f(\ )$ is the dependency function.

This article uses "jieba" segmentation tool. "Jieba" word segmentation is a widely used Python library for Chinese text processing, designed specifically for Chinese text processing. It can automatically recognize and reasonably segment Chinese vocabulary. The use of "jieba" library for word segmentation can handle various complex large-scale texts with fast segmentation speed and high accuracy.

3) Remove stop words. Stop words refer to words that frequently appear in text data but do not contribute significantly to the meaning of the text, such as "de", "de", etc. Removing these words can reduce data noise and improve the accuracy of topic mining and sentiment analysis. Firstly, it is necessary to establish a stop word list that can be customized according to the actual situation. Then, match the vocabulary in the text data with the stop word list, and if the match is successful, remove the vocabulary from the text data. This article adopts the theory of joint entropy as a method for removing stop words. When the average information content of a word

appearing in a sentence and the average information content of sentences containing the word are relatively large [14], it indicates that the word is relatively ordinary. The calculation process is as follows:

$$W(w) = 1 + \frac{1}{\ln(t)} \sum_{i=1}^{t} p_i(w) \ln\left(p_i(w)\right) \tag{7}$$

3.1 Theme Mining

The data collected in the previous chapter is reliable. After preprocessing the collected data, an improved BERTopic topic model is used to mine the topic of a public opinion event in a university. BERTopic is a document clustering based topic modeling tool that combines a Bidirectional Encoder Representation from Transformers (BERT) pre trained model with topic clustering algorithms. It can extract deep features of text through semantic embedding and cluster to generate topics, achieving topic evolution analysis and trend prediction. It has been proven to be more efficient and accurate than mainstream topic clustering tools in processing unstructured text data, deriving common topics and potential semantic structures in text, and other aspects [15].

The architecture of BERTopic tool focuses on clustering and feature engineering, and its application mainly follows four steps: embedding, dimensionality reduction, clustering, and topic representation. The topic clustering effect of BERTopic highly relies on the semantic representation ability of general pre trained models (such as BERT) and the adaptability of basic clustering algorithms (such as HDBSCAN), while lacking a deep understanding of the professional knowledge system in the field of educational technology, resulting in semantic deviations between topic extraction results and the theoretical framework and practical needs of educational technology disciplines. Therefore, this study has made structural improvements to BERTopic from the following aspects:

1) Domain knowledge ontology mapping mechanism. In order to reduce data bias and assist clustering algorithms in capturing semantic information of text more accurately, this study first introduces a word restoration algorithm in the data preprocessing stage to restore words to their dictionary form. Then, to ensure semantic consistency of text input, this study designs a domain keyword mapping table, which equivalently maps synonymous terms and hierarchically maps upper and lower concepts to ensure that clustering algorithms can understand basic concepts of text and adopt standardized and efficient methods for text embedding and dimensionality reduction [16].

2) Pre exploration of data. To improve the accuracy and interpretability of the topic modeling results, this study used the KeyBERT keyword extraction algorithm to conduct topic pre-exploration on the abstract text before conducting cluster analysis [17].

3) By introducing the GPT interface and fully leveraging the advantages of big language models in text analysis, the Prompt project integrates the topic keyword lists, original document information, and zero order clustering candidate keyword information generated by various algorithms to ultimately construct topic names and their keyword representations that are suitable for university public opinion.

4) Fine tuning of the embedding model, focusing on the domain adaptation of the model embedding, using domain annotated corpora to fine tune the BERT embedding model and improve its semantic representation ability for keywords related to university public opinion.

Using the improved BERTopic [18] topic model, the top 10 keywords and their probabilities for each stage and topic of online violence public opinion triggered by a certain event in a university are presented in the form of a table. In vector space, the cosine value of the angle between two vectors can be used to determine whether their directions are consistent. When two vectors are in a completely overlapping state, their angle is 0 degrees, and the cosine value is 1, indicating that the two are extremely similar; If two vectors are perpendicular to each other, with an angle of 90 degrees and a cosine value of 0, it indicates that there is no correlation between these two vectors; When the directions of two vectors are completely opposite, with an angle of 180 degrees and a cosine value of -1, it represents that their properties are completely opposite. If two topic vectors A and B are set, their cosine similarity is calculated by taking the dot product of vectors A and B and taking the product of their magnitudes. This is expressed as:

$$\cos\theta = \frac{X \cdot Y}{\|A\|\|B\|} \tag{8}$$

For the extracted topic text, an improved BiLSTM model is used to perform sentiment analysis on the text topic. The overall process is shown in Fig. 1.

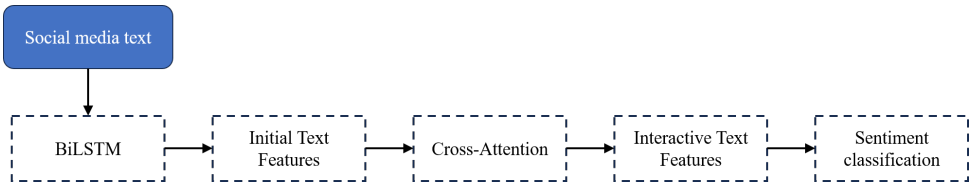


Fig. 1. Overall process of emotion classification

By incorporating dynamic masks into the BiLSTM model and adaptively adjusting the word masking strategy, the ability to extract semantic features has been significantly improved [19].

The extraction process is calculated as follows:

$$\begin{cases} a_i = DM_{Roberta}\left(b_i\right) \\ \overrightarrow{c_i} = \overrightarrow{LSTM}\left(a_i\right) \\ \overleftarrow{c_i} = \overleftarrow{LSTM}\left(a_i\right) \\ c_i = \left[\overrightarrow{c_i}, \overleftarrow{c_i}\right] \\ C = \left[c_1, c_2, \cdots, c_i\right] \end{cases} \tag{9}$$

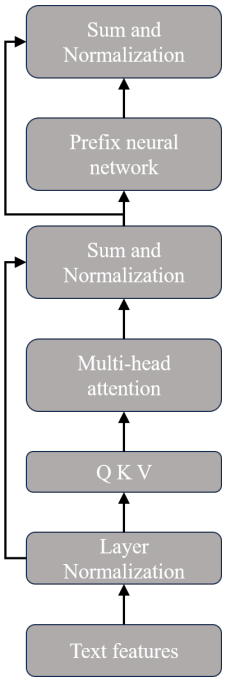


Fig. 2. Principle diagram of fusion process

In the formula, a_i represents the text semantic features extracted by the pre trained model, b represents the text semantic features embedded in the pre trained model, and the range of i is [1, n]. c represents the final state of the feature text semantics. To effectively capture the synergistic effect of language and text on the emotional level and achieve deep integration of data, this paper uses a modal fusion method based on interactive attention mechanism, aiming to promote effective interaction and integration between different modal information. The schematic diagram of the fusion process is shown in Fig. 2.

In the attention mechanism, the feature vectors are represented as query vectors, key values, and representation units Values. Therefore, during feature interaction, the interactive attention feature representation is:

$$f_{attention} = A(Q,K,V) = soft\max\left(\frac{Queries \cdot Keys}{\sqrt{d_k}}\right) \cdot Values \tag{10}$$

Then, the interaction features are obtained through feedforward neural network and layer normalization, and input into the sentiment classification layer:

$$F_{attention} = LN\left(FNN\left(f_{attention}\right)\right) \cdot Values \tag{11}$$

Dependency refers to the relationship of dominance and subordination between words, which is not equivalent and has directional orientation. The two words connected by the dependency relationship are the core word (head) and the dependent word (dependent). Considering that there is also a certain dependency relationship between text and images in social media, Biaffine is used as a classifier instead of MLP classifier to predict the sentiment tendency of fused features, and the relationship between text features and image features is regarded as a dependency relationship. The principle of modal feature fusion based on the double affine mechanism is as follows, where U_m is a tensor, W_m is a matrix, and b_m is a deviation. The expression is as follows:

$$F = F_{attention}U_m + W_m \otimes F_{attention} + b_m \tag{12}$$

The model adopts a cross entropy loss function and minimizes the objective function through regularization. Among them, A represents the number of training samples, B represents the total number of classifications, C represents the actual sentiment labels, and D is the sentiment labels predicted by the output layer. E represents the weight of regularization, and F refers to the set of all trainable parameters.

$$Loss = -\sum_{n=1}^{Number}\sum_{m=1}^{Total} L_{able-i}\log l_{able-i} + \omega\left\|C_{parameter}\right\|^2 \tag{13}$$

3.2 Analysis of the Trend of Public Opinion and Emotion

Based on the publicly available social network dataset from Stanford University, we can further extract key indicators of mainstream social networks (circles), as shown in Table 1.

Table 1. Social network metrics

Indicator	Twitter	Facebook	Google
Number of nodes	81306	4039	107614
Edge number	1768149	88234	13673453
Network diameter	7	8	6
Average path length of the network	4.5	4.7	3
Network clustering coefficient	0.5653	0.6055	0.4901

In the process of public opinion dissemination, in order to ensure that the network has both small world characteristics and scale-free characteristics, the algorithm for generating edges combines two mechanisms: priority connection and neighborhood connection when constructing the network. The improved network model

generates edges for new nodes using a priority connection method based on probability p, and then generates edges using a neighborhood connection method based on probability $1 - p$. According to the above analysis, there are differences in the influence of nodes with different degrees. Therefore, in the neighborhood connection, we set the ratio of edges generated by sub neighborhood connections to edges generated by near neighborhood connections as p. The steps of the improved network evolution algorithm are as follows:

1) Initialize the network, assuming it has x nodes, and then generate a fully connected graph including y_0 nodes;

2) Randomly select node i from $x - y_0$ nodes that have not been added to the network;

3) Connect i to j in a prioritized manner and join the network;

4) In order to increase the clustering coefficient, after executing step 2 once, priority connections are made with probability p, and the neighbors of nodes i and j are given priority connections with probability $p_{i,j}$. The representation of $p_{i,j}$ is as follows:

$$p_{i,j} = \frac{d_j}{\sum d_v} \tag{14}$$

Among them, d_j represents the degree of j's neighbors, $\sum d_v$ is the sum of j's neighbor node degrees. When using $1 - p$ probability for neighborhood connections, the neighbors of i and j perform sub neighborhood connections, and the probabilities are expressed as follows:

$$\frac{p_{i,j}}{p_{i,j}+1}(1-p) \tag{15}$$

When connecting to neighboring domains, the probability is expressed as follows:

$$\frac{1}{p_{i,j}+1}(1-p) \tag{16}$$

5) Return to step 3, set the loop condition for step 3 to prioritize connecting y times and satisfy $y < y_0$;

6) Loop step 2, set the loop condition for step 2 to achieve n nodes.

We consider the student population in university public opinion as ordinary users, who have no specific target opinions. In the classic HK model, ordinary users can only be influenced or not by their neighbors due to the existence of a trust threshold. During the viewpoint update process, ordinary users treat all neighbors within the trust threshold equally, and the viewpoint value at the next moment is the weighted average of the viewpoints of all neighbors within the trust threshold at the current moment. The topology diagram of group viewpoint interaction is shown in Fig. 3.

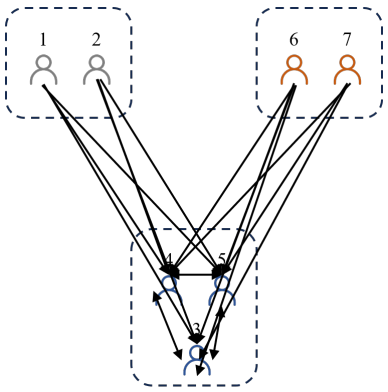


Fig. 3. Topology diagram

For ordinary users, their speech is influenced by two other groups, so their speech is a two-way arrow. The model of differences in viewpoints between groups is represented as follows:

$$\begin{array}{llll} if & Dis_{ij} = \left| n_i(t) - n_j(t) \right| & is & H_i \\ then & \omega_{ij} & is & G_i \end{array} \tag{17}$$

In the formula, H_i and G_i are fuzzy variables, Dis_{ij} represents the distance between individuals' viewpoints, ω_{ij} represents the weight of individual j's influence on individual i. Therefore, this fuzzy rule can be directly translated as if the difference in viewpoints is large, then the influence weight is small, and the membership function $Sub(n)$ of the weight calculates the total area under the membership function curve of the fuzzy set, and finds the geometric center of this area to determine a single and accurate output value. The centroid of the area is represented as:

$$G_{center} = \frac{\int n \cdot Sub(n)\, dn}{\int Sub(n)\, dx} \tag{18}$$

The numerator represents the "weighted" area of the graph formed by the membership function and the axis, where the weight is the membership degree, and the denominator is the total area below the membership function. In the discrete case, the above formula becomes a summation form:

$$G_{center} = \frac{\sum n_i Sub(n_i)}{\sum Sub(n_i)} \tag{19}$$

n_i is the specific value of the input variable, and $Sub(n_i)$ is the membership value at n_i. At this point, deblurring the weights yields:

$$\omega_{ij}^{*} = Blur(\omega_{ij}) \tag{20}$$

The influence between individual opinions after the occurrence of public opinion follows the following reasoning rules:

1) Calculate the difference in viewpoints between individual i and individual j:

$$Dis_{ij} = \left| n_i(t) - n_j(t) \right| = 0.18 \tag{21}$$

2) Blurring the difference in viewpoints, the input viewpoint difference is shown in the above formula, and the total number of interactive groups is set as A, with the number of ordinary users within the trust threshold as B. Then, the fuzzy opinion update model for ordinary users within universities is as follows:

$$Opin(t+1) = \sum_{N_i=1}^{Total(N)} \omega_{ij}(t) Opin(t) + \frac{1}{Total(N)} \sum_{N_i=1}^{Total(N)} Opin(t) + \frac{1}{Total(N_i)} \sum_{1}^{N} Opin(t) \tag{22}$$

$$\omega_{ij}(t) = \frac{\omega^{*}_{ij}(t)}{\sum_{1}^{N} \omega^{*}_{ij}(t)} \tag{23}$$

ω_{ij} represents the weight of updating the opinions of ordinary users.

This chapter first presents a big data preprocessing plan for data collection, which includes three steps of data cleaning, segmentation, and removal of stop words. Then, the processed data is subjected to feature extraction,

and the feature extraction process is improved based on the BERTopic model. By introducing domain knowledge ontology mapping mechanism, data pre exploration mechanism, GPT interface, and embedding model fine-tuning, the accuracy of feature extraction is improved.

4 Experimental Analysis and Result Simulation

The quality of cafeteria meals is one of the main daily concerns for students. Using the method described in this article, we obtained the keywords related to cafeteria meal quality from our school, as shown in Table 2.

Table 2. Theme words for cafeteria food quality

Theme	Top 10 keywords and their probabilities
Theme 1	Canteen price increase (0.1620), unappetizing food (0.1580), small portions (0.1520), crowded cafeteria (0.1480), hygiene issues (0.1420), expensive prices (0.1380), single dishes (0.1340), poor taste (0.1300), long queues (0.1260), poor service (0.1200)
Theme 2	Eating foreign objects (0.1850), mouse heads (0.1780), video exposure (0.1720), food poisoning (0.1680), spoiled ingredients (0.1620), student exposure (0.1580), nausea (0.1520), too outrageous (0.1480), dare not eat (0.1440), exposure (0.1380)
Theme 3	Pointing Mouse to Duck (0.1820) Hot Search (0.1760) Official Response (0.1700) University Cafeteria (0.1650) Food Safety (0.1600) Regulatory Department (0.1550) Logistics Outsourcing (0.1480) Benefit Transmission (0.1420) University Concealment (0.1380) Credibility (0.1320)
Theme 4	Food safety (0.1900) credibility crisis (0.1840) hot search (0.1780) investigation team (0.1720) student rights (0.1660) public opinion supervision (0.1600) meal suspension rectification (0.1540) accountability (0.1480) truth (0.1420) trust collapse (0.1360)
Theme 5	Restore normalcy (0.1750) Long term supervision (0.1700) Teacher and student satisfaction (0.1650) Trust reconstruction (0.1600) System improvement (0.1550) Transparency and openness (0.1500) Normalized supervision (0.1450) Canteen improvement (0.1400) Satisfaction improvement (0.1350) Long term mechanism (0.1300)

This experiment uses Python 3.8 as the development environment and builds the experimental model based on the PyTorch 1.11.0 deep learning platform. In terms of hardware configuration, NVIDIA RTX4060 graphics card with 16GB of video memory is selected for computing acceleration. In the feature extraction stage, text features were obtained through the LiSTM model with the addition of RoBERTa. To verify the improvement of the proposed method on sentiment analysis in topic words, RoBERTa, LiSTM, and RoBERTa+LiSTM were compared, and the comparison results are shown in Fig. 4.

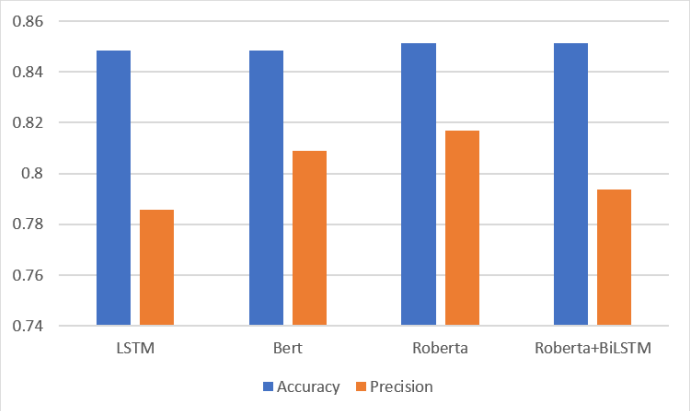


Fig. 4. Comparison of model results

For the analysis of the evolution of public opinion after its occurrence, the number of group members is set to 350, and the initial set of viewpoints is set to:

$$Opin_{group}(0) = \left[Opin_1(0), Opin_2(0), \cdots, Opin_N(0) \right]^T \tag{24}$$

The fixed viewpoint difference membership function is the second set of membership functions, with the benefit parameter set as g = 0.973. To explore the impact of different levels of trust among ordinary students towards different groups on the evolution of public opinion in the development of university public opinion, the trust values among ordinary students are set as 0.4, 0.6, and 0.8, respectively. The simulation results are shown in Fig. 5.

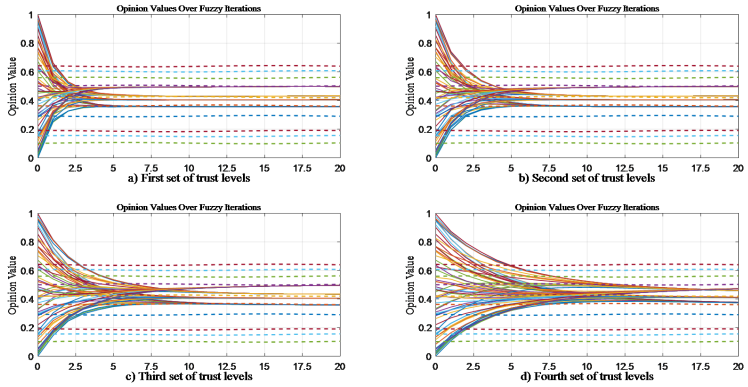


Fig. 5. Schematic diagram of simulation results

According to the three graphs above, the converging ones are ordinary users in the network structure, while the line representing opinion leaders whose other viewpoints remain largely unchanged. As they are only influenced by environmental factors of interest, the interest factors and punishment risks we set are basically balanced, so they can maintain their own viewpoints without major changes As the trust level of ordinary users gradually increases, the convergence speed of the ordinary user group gradually decreases and the convergence time gradually increases. In a sub network of the network structure, ordinary users have a high trust level in their own ordinary users and a low trust level in opinion leaders. This will gradually converge the group opinion of the small network. However, there is not sufficient communication between each small network, because opinion leaders are the opinion orientation of every individual connected to them. Each individual in this group may come from different sub networks. When the trust level of ordinary individuals in the opinion leader's opinion is high, everyone evolves towards the opinion leader's opinion direction and quickly reaches convergence When ordinary individuals have low trust in opinion leaders, they interact within their own sub networks, resulting in a lack of communication between sub networks, which slows down the evolution of group opinions in the overall network.

This Fig. 6 depicts the time course of the interaction among neutral, negative, and positive sentiments and normalized public engagement during the time of evolution of a university-related public opinion event. It can be observed that before the outbreak of the incident, the proportion of neutral sentiment is at its highest point, while both negative sentiment and public interest stay at low levels. The rise of negative sentiment and public interest starts from around the time of 5.5 since that is when the incident begins and the increase in negative sentiment is fast attributed to the rise in public interest. The simultaneous rise symbolizes that the negative emotion plays an important role in early information effectiveness. In terms of the viral diffusion stage of the incident, the level of public interest continues to remain high, while negative emotion has its second wave of oscillation. This means that the consequence of repeated exposure, reposting, and discussion prolongs emotional activation. Positive sentiment stays at a limited level during the crisis and viral diffusion stages, suggesting that there's no enough spontaneous emotional correction in times of negative info dominance.

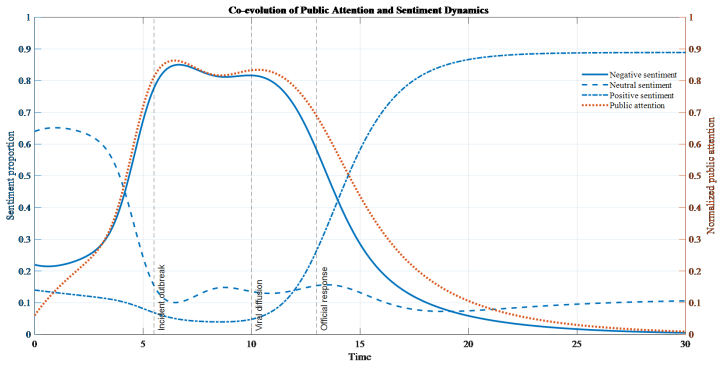


Fig. 6. Co-evolution of public attention and sentiment dynamics

The Fig. 7 displays the interaction effect of timeliness and credibility of responses on the values of the public-opinion risk index. The optimal scenario is characterized by prompt response and high credibility of the response, while the worst scenario involves significant delays and low credibility. In this case, negative messages spread longer without any possibility of correcting them promptly, which causes the overall risk index to go higher. As a result, the peak value of the risk index and the cumulative volume of exposure to dangers rise, since negative moods accumulate longer. On the contrary, having a timely response with high credibility minimizes the public-opinion risk index values, as credible responses lead to the decrease in the level of uncertainty, the impact of rumors, and emotional recovery. Timeliness of responses produces an obvious negative impact on public-opinion risk index values, but this impact depends on the level of credibility. For instance, if credibility is low, making the response faster makes no significant change in the value of the risk index, as timely response that lacks credibility can hardly change people's attitudes.

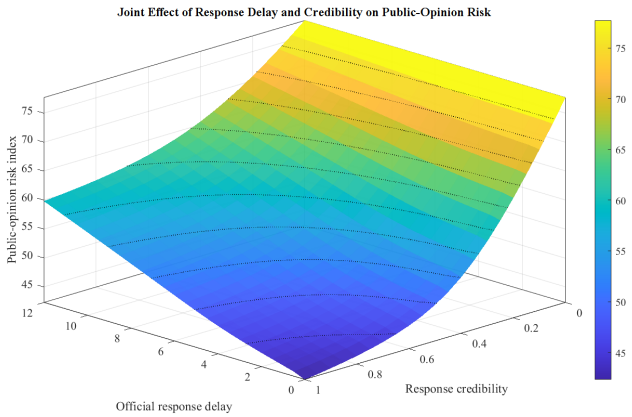


Fig. 7. Joint effect of response delay and credibility on public-opinion risk

Fig. 8 illustrates that trust level among the ordinary and the opinion leaders plays an important role in determining the number of iterations necessary for achieving group consensus. The yellow region at the bottom of the heatmap corresponds to the cases when convergence is either too slow or fails altogether within the simulation period. The occurrences of this region are usually observed in the situations where trust in opinion leaders is extremely low regardless of the level of trust among peers. In this case, ordinary users exchange their opinions only within the specific local communities that they belong to. The rise of trust towards opinion leaders leads to the reduction in time needed for achieving consensus from the yellow/green areas to the blue area. In other words, opinion leaders serve as a reference point for the groups that otherwise remain isolated from each

other in their opinion communities. Peer trust has a more complicated influence. When the trust towards leaders is sufficient, greater peer trust may help local coordination but not hinder the convergence overall.

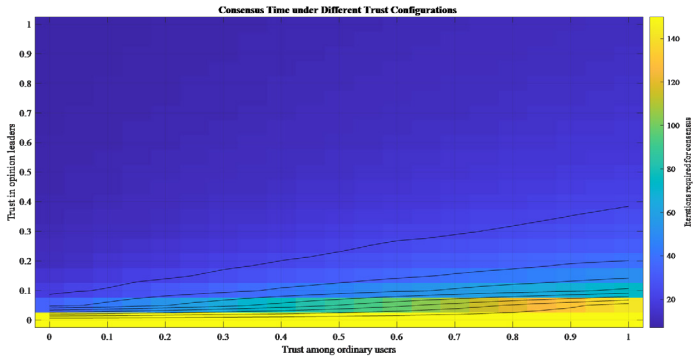


Fig. 8. Consensus time under different trust configurations

5 Conclusion

This article constructs a text extraction model, sentiment analysis model, and sentiment evolution model. Through analysis, scientific methods can be used to monitor the development and direction of public opinion in universities. Therefore, combined with the methods in this article, the following suggestions are given for the governance of public opinion in universities:

1) Effective monitoring work can be carried out during the latent period of public opinion development, using the model in this article to predict trends and make preliminary judgments on netizens' emotions. This can put government agencies in a more proactive position, improve their ability to predict, and take timely response measures. By identifying potential online topics that may attract social attention in advance, relevant departments can have more ample time to make decisions and develop more comprehensive response plans, thereby reducing potential actual damage and image loss. The most crucial aspect of this process is to accurately identify and predict the development trend of online public opinion, and quickly issue warning notices when identifying risks. In order to strengthen the monitoring system of the latency period of online public opinion, the government can start from the following dimensions: updating the understanding of crisis management by online public opinion regulatory departments, enhancing the public's understanding of sudden public events, promoting proactive prevention and control management models, systematically studying public opinion response strategies, building a database of public opinion management strategies and experience in handling sudden events, so that management personnel can establish a comprehensive response mechanism based on different situations, referring to successful cases, including graded response, processing procedures and other contingency plans, improving the effectiveness of public opinion response, and attaching importance to data analysis. The sentence is:, The government or professional public opinion monitoring organization accurately collects and judges all kinds of information, focuses on social platforms with large number of users and great social influence, such as Weibo, WeChat, Tiktok, and closely tracks hot topics on these platforms. In the constantly evolving process of online public opinion environment, relevant institutions urgently need to deploy targeted solutions, strengthen real-time monitoring of various hidden dangers, ensure that possible problems can be identified and properly handled in the first time, and minimize or avoid their negative impacts as much as possible.

2) With the arrival of the outbreak period of online public opinion, the trend prediction model of online public opinion can conduct real-time monitoring and early warning, which facilitates the government to take decisive actions, strengthen supervision of events, and maintain government credibility. This process is often accompanied by many challenges, and the public may have doubts about government actions. When this suspicion accumulates to a certain extent, it will form a concentrated wave of public opinion. If government departments do not handle it properly, it may lead to a loss of control over public opinion and even cause serious social problems. Implementing effective public opinion guidance and control strategies has become a key factor in

resolving the current public opinion crisis. In response to the special situation of high incidence of online public opinion, government departments should establish a rapid response mechanism and take systematic measures to ensure timely handling and effective control of public emergencies. Government departments also need to have keen insight, accurately identify key variables that affect the development of events, and develop corresponding prevention and control plans. They should systematically analyze the essential characteristics of the evolution of the situation and formulate scientific and reasonable response strategies to control the emergency situation. And it is necessary to fully recognize the key value of authoritative information, strengthen the timeliness, authenticity, and credibility of information dissemination, prevent more serious conflicts and contradictions caused by information blockage, and effectively stop rumors

3) In the fading period, a bimodal public opinion sentiment classification model can be used to judge the emotional tendencies of users' generated information, grasp the emotional tendencies of netizens towards the end of the event, and actively carry out online public opinion management work, vigorously promote positive cases, in order to reduce the adverse effects brought by public opinion impact. When the outbreak of online public opinion ends, the government may encounter more serious negative impacts, such as image decline and damage to credibility. In the fading stage of public opinion, it is necessary to strengthen the suppression and elimination of the negative impact of public opinion. The government should increase communication with the public, use effective interaction mechanisms to provide necessary and reasonable compensation to stakeholders affected by public opinion, reshape the organizational image, and improve the government's trust in the public. The government should also engage in profound reflection and summarization, draw lessons from relevant events, improve policies and management measures, and avoid similar incidents from happening again. In addition, the government should encourage and support academic research and professional institutions to conduct long-term tracking and research on online public opinion, use scientific methods to analyze the evolution laws of public opinion, and provide theoretical support and practical guidance for government decision-making. By implementing these measures, the government can effectively respond to the decline of online public opinion and improve the overall level of online public opinion management, contributing to the construction of a harmonious society.

6 Acknowledgement

This paper is based on the research findings of the project Research on the Generation, Evolution and Governance Path of Implicit Public Opinion in Semi-closed Virtual Communities of Universities in Hebei Province.

References

[1] M.-M. Xiang, Z.-X. Ji, L. Zhang, J.-M. Guo, T. Ma, Dual Empathy Evolution Mechanism and Guidance Strategies of Online Public Opinion in Emergent Public Events: A Case Study of the Manner Coffee Shop "Coffee Powder Splashing" Incident, Information Studies: Theory & Application 49(6)(2026) 85-94. https://doi.org/10.16353/j.cnki.1000-7490.2026.06.010

[2] X.-W. Wang, C.-C. Qiu, Y.-T. Wu, Network Group Polarization under the Action of Generative Artificial Intelligence: Formation Mechanism, Influence and Governance, Information Studies: Theory & Application 48(10)(2025) 10-18. https://doi.org/10.16353/j.cnki.1000-7490.2025.10.002

[3] G.-H. Wang, H.-Y. Wang, Y.-S. Wang, A Study on the Spatio-temporal Evolution Path and Driving Mechanisms of Online Public Opinion from a Hypernetwork Perspective: A Case Study of the xx Airlines Aircraft Accident, Journal of Geo-information Science 27(6)(2025) 1462-1477 https://doi.org/10.12082/dqxxkx.2025.250072

[4] S.-Z. Zhao, N. Wei, Y.-Y. Li, P.-F. Gao, X.-H. Gu, Research on Online Public Opinion Prediction Based on Sentiment Analysis and Heat Prediction, Journal of Xi'an Shiyou University (Natural Science Edition) 39(1)(2024) 135-142. https://doi.org/10.3969/j.issn.1673-064X.2024.01.017

[5] Z. Li, Y. Zhang, Z.-C. Li, P. Zhan, L. Chen, Public Opinion Analysis in Universities Based on GNN Multimodal Fusion, Computer Science 53(6)(2026) 30-38. https://doi.org/10.11896/jsjkx.250600158

[6] K. Han, H.-P. Pan, Z.-Y. Liu, Research on Sentiment Analysis of Short Video Network Public Opinion by Integrating BERT Multi-level Features, Journal of Frontiers of Computer Science and Technology 18(4)(2024) 1010-1020. https://doi.org/10.3778/j.issn.1673-9418.2311023

[7] F. Jiang, S.-L. Guo, Construction and Evolution Analysis of a Reasoning Graph for Network Public Opinion Based on Large Language Model-Taking Sports Events as an Example, Information Science 44(4)(2026) 124-134+149. https://doi.org/10.13833/j.issn.1007-7634.2026.04.013

[8] M. Liu, L.-X. Zhang, W.-Y. Ping, F. Huang, Research on Multi-Stage Online Public Opinion-Driven Group Consensus Decision-Making Method Based on Large Language Model, Chinese Journal of Management 22(4)(2025) 750-759. https://doi.org/10.3969/j.issn.1672-884x.2025.04.016

[9] X. Han, M.-K. Li, S.-Y. Dai, C.-H. Zhou, The Configuration of Factors Influencing the Popularity of Network Public Opinion in Social Emergencies-Based on Fuzzy Set Qualitative Comparative Analysis (fsQCA), Information Science 42(12)(2024) 60-69+75. https://doi.org/10.13833/j.issn.1007-7634.2024.12.007

[10] X.-N. Gao, Q. Huang, M.-W. Zhang, R.-L. Geng, C.-W. Bi, Research on Driving Factors Identification and Evolution Mechanism of Network Public Opinion Reversal of Hot Events: Based on the Perspective of System Dynamics Simulation, Information Studies: Theory & Application 49(1)(2026) 115-128. https://doi.org/10.16353/j.cnki.1000-7490.2026.01.014

[11] C. Li, W.-Z. Xu, J.-K. Zheng, Research on associated features and operation mode prediction of air conditioning and window switching behaviors in office buildings based on data mining, Heating Ventilating & Air Conditioning 56(6)(2026) 50-59. https://doi.org/10.19991/j.hvac1971.2026.06.08

[12] Y.-P. Luo, A.-L. He, K.-M. Yu, K. Zhang, F.-Y. Liao, A Structural Damage Early Warning Method Based on Convolutional Neural Networks and Autoencoders, Steel Construction (Chinese & English) 41(6)(2026) 1-9. https://doi.org/10.13206/j.gjgS25092901

[13] J. Zeng, Q. Ma, L. Chen, A Study on the Topic Evolution of Short Video User Comments on Breaking News Based on LDA-BERTopic, Information Science 44(1)(2026) 57-66. https://doi.org/10.13833/j.issn.1007-7634.2026.01.007

[14] H.-F. Yuan, Y.-J. Zhu, Y.-Y. Pan, R.-W. Zhang, Z. Jin, A High-Quality Ocean Observation Profile Datasets Construction Scheme Based on Multi-Source Data Cleaning and Fusion, Frontiers of Data & Computing 8(3)(2026) 68-80. https://doi.org/10.11871/jfdc.issn.2096-742X.2026.03.007

[15] S.-Y. Lu, L. Zhang, X.-L. Li, Q.-J. Xing, D.-P. Duang, H. Lin, Unsupervised word segmentation method in power domain based on BERT, Information Technology (1)(2024) 96-103. https://doi.org/10.13274/j.cnki.hdzj.2024.01.015

[16] A.-Q. Liu, S.-D. Wang, Research on Multi-category Document Classification System Based on TF-IDF and GloVe, New Century Library (10)(2025) 40-46. https://doi.org/10.16810/j.cnki.1672-514X.2025.010.006

[17] Z.-N. He, T.-T. Zhang, Z.-K. He, L.-J. Meng, Topic mining and content analysis of rehabilitation ethics research in China based on BERTopic model, Chinese Journal of Rehabilitation Theory and Practice 32(2)(2026) 132-141. https://doi.org/10.3969/j.issn.1006-9771.2026.02.002

[18] Y.-L. Hu, Y. Chen, Y.-T. He, X.-Q. Gu, The Future of GenAI Education Applications: Topic Modeling Based on the Improved BERTopic Tool, Modern Educational Technology35(7)(2025) 44-53. https://doi.org/10.3969/j.issn.1009-8097.2025.07.005

[19] Y.-H. Ma, H.-M. Wang, H.-Q. Liu, J.-H. Chen, X.-Y. Liu, College Entrance Examination Admission Score Prediction in Colleges and Universities Based on CNN-BiLSTM-ATT Hybrid Model, Modern Information Technology 10(4)(2026) 24-31. https://doi.org/10.19850/j.cnki.2096-4706.2026.04.005