

Retraction

Retracted: University Media Content Detection and Classification Based on Information Fusion Algorithm

Advances in Multimedia

Received 15 August 2023; Accepted 15 August 2023; Published 16 August 2023

Copyright © 2023 Advances in Multimedia. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] S. Zhang, Q. Yu, T. Yang, and K. Peng, "University Media Content Detection and Classification Based on Information Fusion Algorithm," *Advances in Multimedia*, vol. 2022, Article ID 9655452, 14 pages, 2022.

Research Article

University Media Content Detection and Classification Based on Information Fusion Algorithm

Shuntao Zhang,¹ Qinglan Yu,² Tianming Yang ,¹ and Kai Peng³

¹CPC Publicity Department, North China Electric Power University, Beijing, China

²School of Foreign Languages, North China Electric Power University, Beijing, China

³Big Data Strategy Research Institute, Guangdong University of Technology, Guangzhou, China

Correspondence should be addressed to Tianming Yang; dsjyhdz@ncepu.edu.cn

Received 18 August 2022; Revised 31 August 2022; Accepted 6 September 2022; Published 7 October 2022

Academic Editor: Tao Zhou

Copyright © 2022 Shuntao Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to further solve the problems in promoting the classification of media content in colleges and universities, the effective analysis and understanding of multimedia data content can be better realized based on the characteristics of multimedia data in colleges and universities, combining with the characteristics of rich information, large differences in performance, and large amount of large-scale data. This essay mainly introduces the technology of university media content detection and classification based on information fusion algorithm and focuses on the application of university multimedia content detection, analysis, and understanding, to explore the image discrimination auxiliary attribute feature learning and content association prediction and classification. A benchmark model for media content detection and classification is constructed. Through the model test, it is found that the *F1* value of the model is more than 70%, the check rate is more than 80%, and the recall rate is more than 50%. On this basis, a content detection system based on campus network is constructed.

1. Introduction

With the rapid development of the Internet, the influence of the Internet on the contemporary people is increasing. Especially for teenagers, the Internet has brought great influence on their thoughts and life. For example, websites such as violence, pornography, and gambling under online media are contacted and browsed by more and more teenage users, which is not conducive to the establishment of correct values among teenagers. For colleges and universities, in the context of the rapid development of the Internet, how to effectively monitor the campus media network information content and timely find the bad information and content in the college media network is a problem that must be paid attention to in the field of education today [1]. The rapid development of the network makes the development of the campus network is also rising rapidly. Campus media network has a wide range of users, including office users, teaching users, and students. How to better realize the detection and classification of media content with the help of technical

optimization under the condition of multiple users is a problem that universities must attach importance to [2]. Therefore, this essay tries to use information fusion algorithm to better realize the detection and scientific classification of media content.

2. Literature Review

Shu et al. proposed a new weakly supervised depth matrix decomposition algorithm for weakly supervised label association images provided by users to learn the potential representation of data [3]. In this method, the latent image representation and markup representation hidden in the latent subspace are revealed through the collaborative study of weakly supervised markup information, visual structure, and semantic structure. It can naturally embed new images into subspaces, combine semantic and visual structures, and learn idiomatic subspaces without overfitting noise, incomplete or subjective labels. In addition, the method can deal with noisy, incomplete or subjective labels and

noisy or redundant visual features. Xu et al. proposed the introduction of attribute-based classification, in which objects are identified based on high-level descriptions of semantic attributes (such as object color or shape). The system classifies objects from a list of high-level semantically meaningful attributes called attributes. Attributes act as an intermediate layer in the classifier cascade, enabling the system to identify object classes that do not see a single training example [4]. Because the recognition of each attribute is beyond the current specific learning task, attribute classifiers can be independently prelearned, for example, from existing image datasets unrelated to the current task, new classes can then be detected based on their attribute representation without the need for a new training phase. Representation-based semantic attributes can describe visual content well and improve the performance of visual understanding. Therefore, it is very important to explore the semantic properties of visual understanding, especially fine-grained visual understanding. Plaza-Del-Arco et al. proposed an interactive method to obtain the important and distinguishing attributes of manual continuous annotation. Local attributes with distinguishing and semantic meaning are discovered from image data sets using only fine-grained category labels and object boundary box annotations [5]. A potential conditional random field model is used to discover candidate attributes that are detectable and differentiated, and then a recommendation system is used to select attributes that may have semantic significance. Human interactions are used to provide semantic names for discovered properties. Design features that we learn based on known attributes are not necessarily significant and can automatically and efficiently construct visual recognition to distinguish unknown categories. Alsagri and Ykhlef believed that the description model required fine-grained image information, so they used the object detection model to segment the subgraph in the image and transform the subgraph into local features, so as to provide more detailed image features [6]. Verdoliva proposed fine-tuning the neural network by using attribute triplet loss and proposed a feature generation learning framework based on CNN to solve the generalized zero-sample learning task [7].

3. Principles of Content-Based Audio Classification and Retrieval

3.1. System Framework. Content-based audio classification and retrieval system (CBRA) is a kind of information service system between information user and audio database. Figure 1 shows the system framework of THE CBRA system [8]. In audio retrieval, we need to go through the key steps of feature extraction, audio segmentation, audio recognition, and classification and index retrieval. The system includes two parts: audio database generation module and user query and browse module.

The audio data is clustered by feature, and the clustering information is packed into the part of clustering parameter library.

3.1.1. Extraction of Nonpresentation Attribute Information. Used to deal with the user's regular query, this module extracts two major attributes: general file attributes, including the full file name, file size, editing time; and audio encoding attributes, including encoding format, playback time, number of channels, sampling rate, and sampling bits.

3.1.2. Feature Extraction. Feature extraction is one of the core functions of the system [9]. Every time a piece of audio data is added to the audio library, its audio features should be extracted. Analyze the value of each feature, so that it can be segmented and classified and then added into the feature database. Feature extraction is also commonly used in audio retrieval, and the query feature vector is determined by combining attribute values. For example, when a user submits a sample, its features must be extracted before similarity calculation can be carried out.

3.1.3. Audio Segmentation. Using the corresponding audio features in the feature database, the current long-time audio stream is segmented so that the segmented audio segment contains a single type.

3.1.4. Audio Classification. Audio Classification is the key functions of audio retrieval system. The segmented audio segments are automatically classified and classified into different predefined semantic classes. In this step, the segmented audio physical units can be roughly divided [10]. For example, the segmented audio can be classified as mute, music, voice, environment sound, etc.; and an event or a person can also be finely classified, such as "explosion" event and "speech" event.

3.2. Audio Feature Analysis and Expression

3.2.1. Audio Signal Digitization and Preprocessing. Audio signal is a one-dimensional analog signal with continuous change of time and amplitude. Although its forms are various, the first step of processing with modern information technology is digital processing and preprocessing of the signal. The digital processing of audio signal is mainly to process it and turn it into digital signal with discrete time and amplitude: generally including amplification and gain control, antialiasing filtering, sampling, A/D conversion, and coding (generally PCM code). As shown in Figure 2, the digitized audio signal is actually a time-varying signal [11]. Preprocessing generally includes endpoint detection, preweighting, and frame adding window.

After prefiltering and sampling, the signal only becomes discrete signal in time, but still keeps continuous characteristic in amplitude. Therefore, we quantized its amplitude to make it a digital signal with discreteness in both time and amplitude, that is, converted into binary digital code by A/D converter [12]. A quantizer is to divide the amplitude of the whole signal into several finite intervals, and the sample points falling into the same interval are represented by the same amplitude, which becomes the quantization value, generally expressed in binary. Quantization inevitably produces

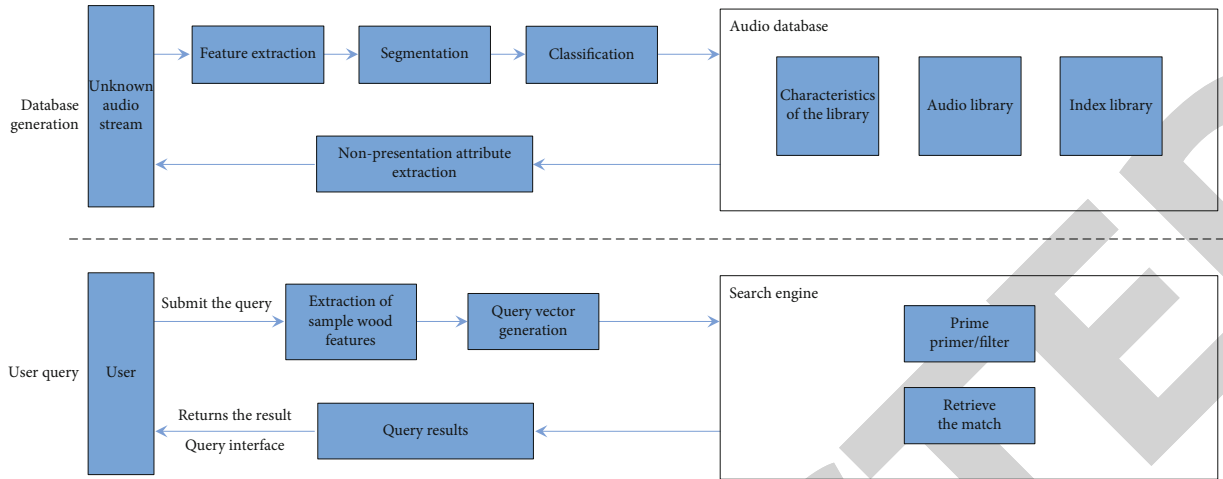


FIGURE 1: System framework.

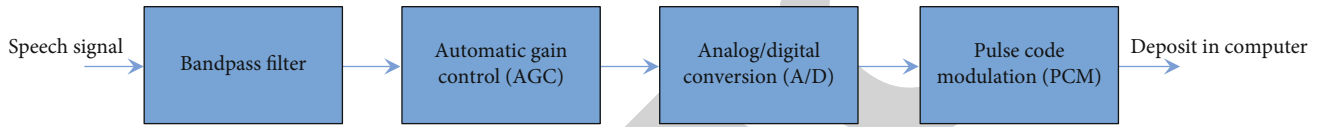


FIGURE 2: Block diagram of audio signal digitalization process.

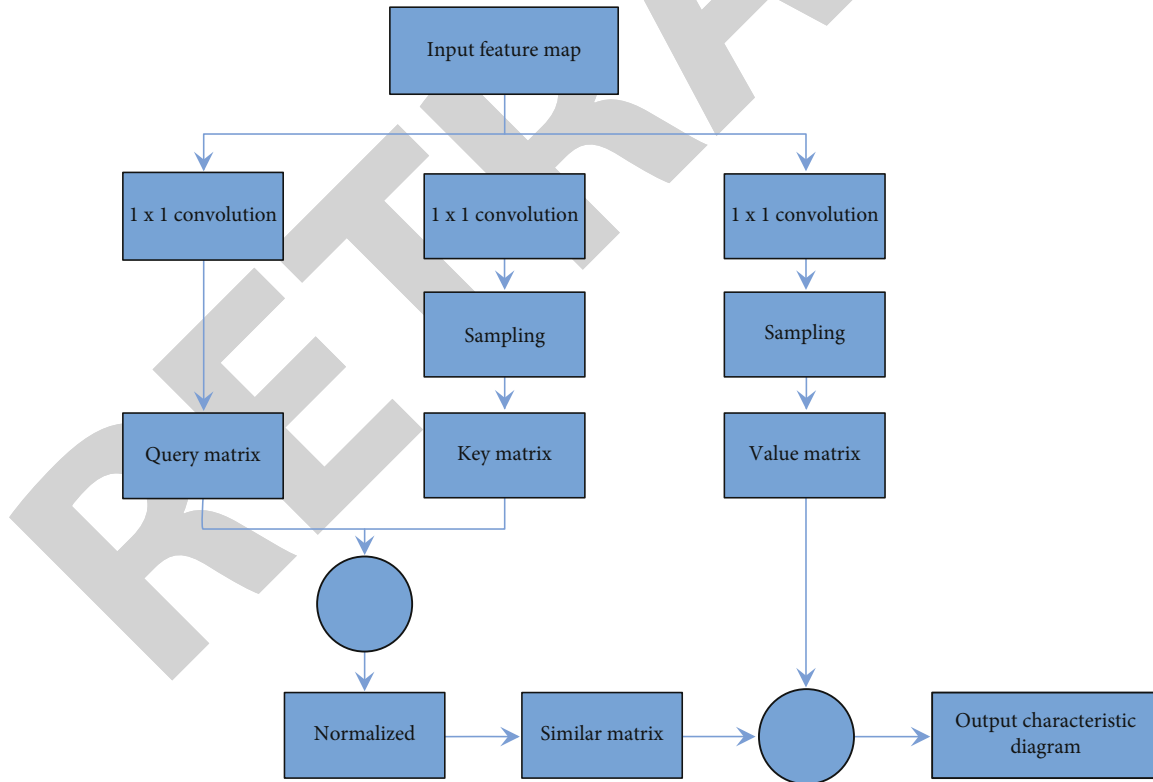


FIGURE 3: Schematic diagram of the pyramid non-local module (APNB) proposed in this essay.

quantization error, which is defined as

$$e(n) = \hat{x}(n) - x(n), \quad (1)$$

where $e(n)$ is called quantization error or quantization noise, $\hat{x}(n)$ is the quantized sampling value, namely, the quantizer output value, and $x(n)$ is the unquantized sampling value, namely, the quantizer input value.

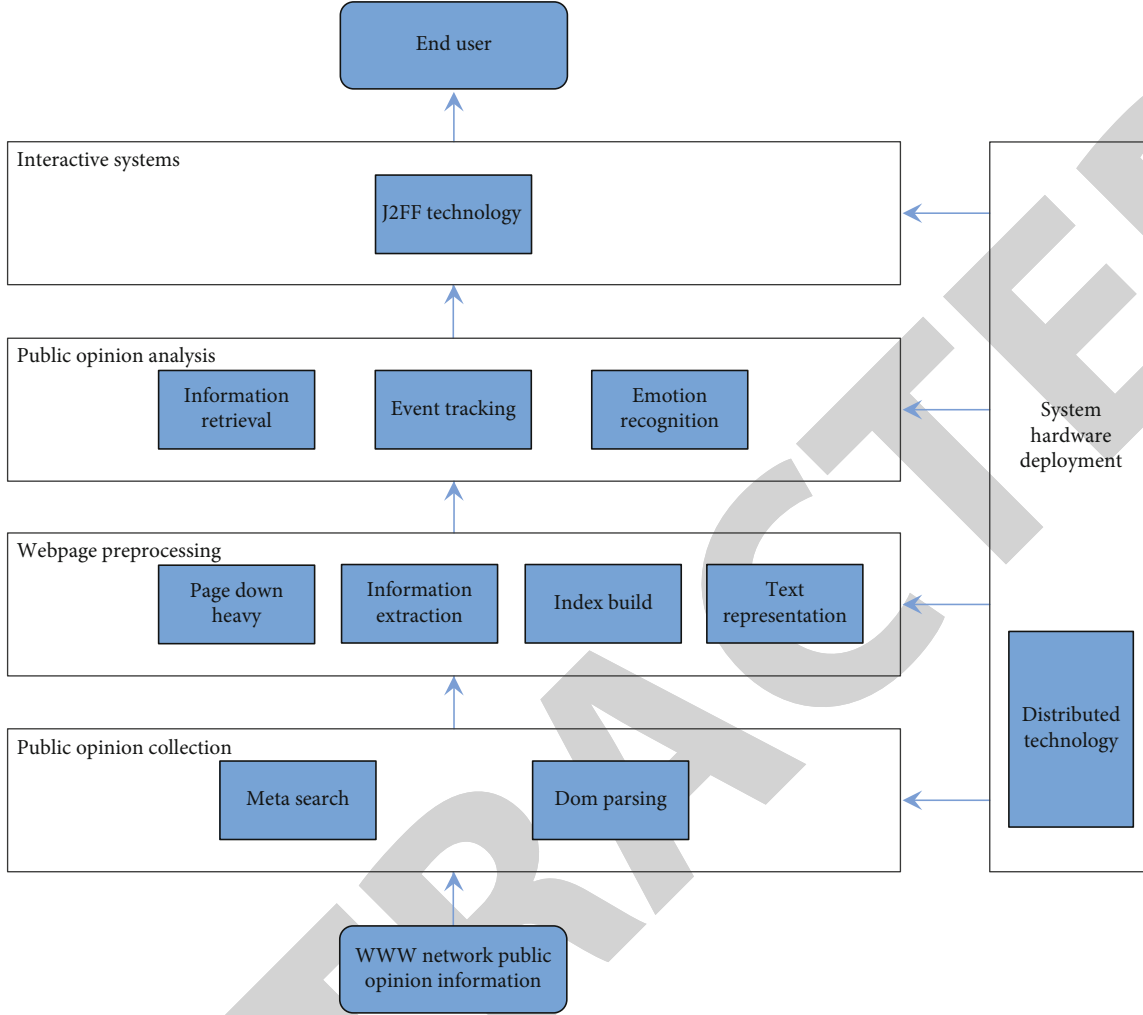


FIGURE 4: Framework of public opinion monitoring system.

Assuming that σ_x^2 represents the variance of the input audio signal sequence, $\hat{x}(n)$ represents the peak value of the audio signal, $x(n)$ represents the variance of the noise sequence, and the quantized SNR can be expressed as

$$\text{SNR}(\text{dB}) = 10 \log \left(\frac{\sigma_x^2}{\sigma_e^2} \right) = 6.02B + 4.77 - 20 \log \left(\frac{x_{\max}}{\sigma_x} \right). \quad (2)$$

If the amplitude of the audio signal follows the Laplace distribution, the probability of the amplitude of the audio signal exceeding $4\sigma_x$ is very small, only 0.35%. Therefore, x_{\max} can be set as $4\sigma_x$ [13]. In this case, the above equation can be changed into

$$\text{SNR}(\text{dB}) = 6.02B - 7.2. \quad (3)$$

3.2.2. Analysis and Expression of Audio Features. The energy of audio signal changes obviously with time, and its short time energy analysis gives a suitable description method to reflect these amplitude changes. Short-time average energy

refers to the average energy of signal gathering at sampling points in an audio frame, which can better reflect the variation of audio signal amplitude with time. It is assumed that the audio signal is divided into M audio frames after sampling, each frame contains N sampling points, and the frame shift is half of the frame length [14]. The short-term average energy is defined as

$$E_m = \frac{1}{N} \sum [x(n)w(n-m)]^2, \quad (4)$$

where E_m represents the average energy of the m_{th} audio frame signal; $x(n)$ represents the value of the n_{th} sampling point in the m_{th} audio frame; $w(n)$ is the window function.

The calculation formula of short-term zero crossing rate is as follows:

$$Z_m = \frac{1}{2} \sum_m |\text{sgn}[x(n)] - \text{sgn}[x(n-1)]| w(n-m). \quad (5)$$

Z_m represents the short-time zero crossing rate of the m

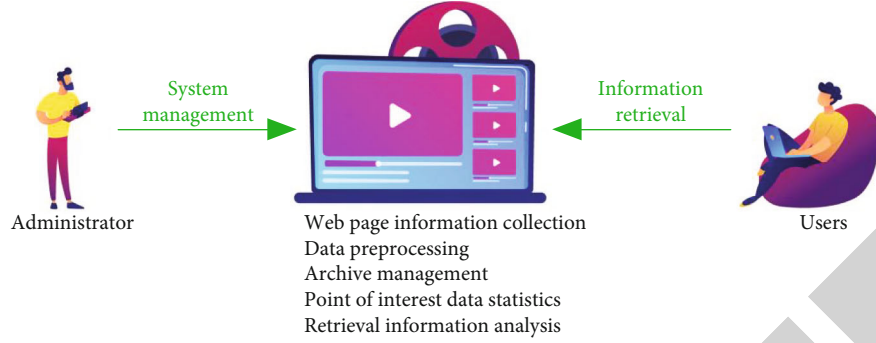


FIGURE 5: Top-level use case diagram of system functions.

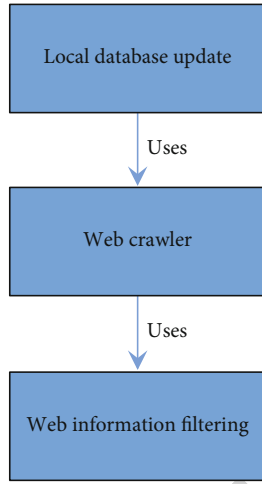


FIGURE 6: Use case diagram of web page information collection function.

m th audio frame; $x(n)$ represents the value of the n th sampling point in the m th audio frame; $w(n)$ is the window function. $\text{sgn}(\bullet)$ is a symbolic function defined as follows:

$$\text{sgn}[x(n)] = \begin{cases} 1, & x(n) \geq 1, \\ -1, & x(n) < 1. \end{cases} \quad (6)$$

Short-time autocorrelation function is obtained by windowing the signal on the basis of autocorrelation function, namely,

$$R_n(k) = \sum_{m=n}^{n+N-k-1} x_w m x_w(m+k). \quad (7)$$

The autocorrelation function provides a way to obtain the period of a periodic signal: its autocorrelation function can reach its maximum value on an integer multiple of the periodicity of the periodic signal. That is, the period can be estimated from the position of the first maximum value of the autocorrelation function without considering the start time of the signal [15].

Short-time autocorrelation function is an important parameter in time domain analysis of audio signal, but the

calculation of autocorrelation function is very large because of the long time required for multiplication operation. In order to avoid multiplication, another parameter which has a similar effect with autocorrelation function, namely, short-time mean amplitude difference function, can be used. If the signal is a completely periodic signal, (let the period be N_p); then, the amplitude at the sample points separated by integer multiples of the period is equal, and the difference is 0 [16]. That is

$$d(n) = x(n) - x(n+k) = 0(k = 0, \pm N_p, \pm 2N_p, \dots). \quad (8)$$

For an actual audio signal, d_n is small, although not equal to zero. These minima will occur at integer multiples of the period, for which the short-time mean amplitude difference function can be defined:

$$F_n(k) = \sum_{m=0}^{N-k-1} |x_n m - x_n(m+k)|. \quad (9)$$

Obviously, if $x(n)$ is periodic within the window value range, then $F_n(k)$ will have a minimum value at $(k = 0, N_p, 2N_p, \dots)$. In contrast to $R_n(k)$, $F_n(k)$ has valleys rather than peaks at various integral multiples of the period.

The frequency center is the central frequency of the Fourier transform, which is an indicator to measure the brightness of sound. The calculation formula is as follows:

$$\text{FC} = \frac{\int_0^{\bar{w}} |F(w)|^2 w dw}{E}. \quad (10)$$

E is the frequency energy, and its calculation formula is

$$E = \int_0^{\bar{w}} |F(w)|^2 dw. \quad (11)$$

The calculation formula \bar{w} is as follows:

$$\bar{w} = \frac{fs}{2}, \quad (12)$$

where fs is sampling frequency.

TABLE 1: Collected web page information content name of the collected content.

Range of collection	A basic description
Key work of colleges and universities	The main work of the college is to reflect the information of mainstream media such as the official website of the college, as well as the mainstream public opinion of netizens and the corresponding comments.
College policies and measures	National college reforms and policies implemented as well as online public opinion in mainstream media; different environment, knowledge, and background of users produce different network public opinion praise or criticism.
Significant events	The events with great influence in a certain period of time, as well as the information comments and netizens' release.
The emergency	Large-scale traumatic emergencies and the information dissemination and evaluation caused by them.
The people's livelihood related	Major application distribution platforms release, comment, and spread the vital interests of people's livelihood, forming a large number of online public opinion information resources.

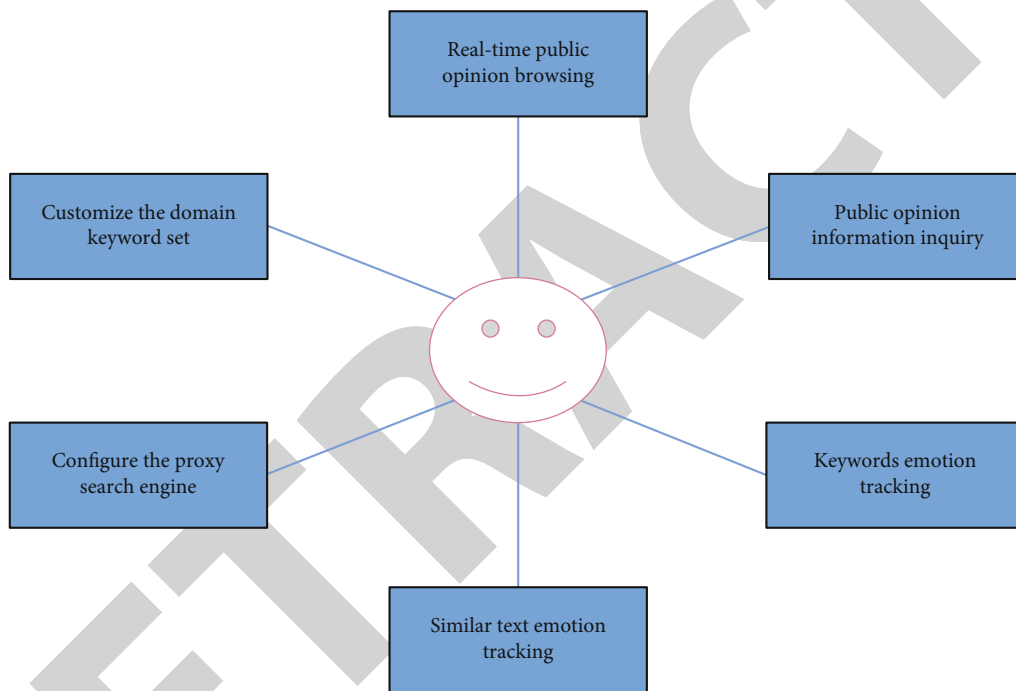


FIGURE 7: Use case diagram of public opinion monitoring system.

TABLE 2: Original data table obtained by crawler.

The column name	The data type	Explanatory notes
Id	Int	The table's primary key
URL	Varchar (200)	The URL of the opinion article
Source	Varchar (200)	The source of opinion articles
Theme	Varchar (200)	Subject of public opinion
Author	Varchar (200)	Author of public opinion
Content	Text	The body of public opinion
Date	DateTime	Publication time of public opinion
Comment_num	Int	Comment number
Link_num	Int	Forwarding number

TABLE 3: Total table of data sets.

The column name	The data type	Explanatory notes
Id	Int	The table's primary key
URL	Varchar (200)	The URL of the opinion article
Source	Varchar (200)	The source of opinion articles
Theme	Varchar (200)	Subject of public opinion
Author	Varchar (200)	Author of public opinion
Date	DateTime	Publication time of public opinion
Keyword	Varchar (200)	Key words of public opinion
Comment_num	Int	Comment number
Link_num	Int	Forwarding number
Element	Varchar (200)	Elements of public opinion
IsNegative	Double	Negative or not
IsPublk	Int	Whether the public opinion
Class_media	Varchar (200)	Classified information

TABLE 4: Keyword table (record all keywords).

The column name	The data type	Explanatory notes
Id	Int	The table's primary key
Keyword	Varchar (200)	Keywords
Class_media	Varchar (200)	Category of belonging

TABLE 5: Keywords of elements (record all key words of elements).

The column name	The data type	Explanatory notes
Id	Int	The table's primary key
Keyword	Varchar (200)	Keywords

TABLE 6: Classification Information table (record classification name).

The column name	The data type	Explanatory notes
Id	Int	The table's primary key
Name	Varchar (200)	Category name
Super	Varchar (200)	Category of belonging

TABLE 7: URL table (record the corresponding information of URL and website name).

The column name	The data type	Explanatory notes
Id	Int	The table's primary key
URL	Varchar (200)	URL
Name	Varchar (200)	Corresponding website name
Type	Varchar (200)	Site type

Bandwidth is an indicator to measure the range of audio frequency and is calculated as follows:

$$BW = \sqrt{\frac{\int_0^{\bar{w}} (w - FC)^2 |F(w)|^2 dw}{E}}, \quad (13)$$

where FC is the frequency center. The cepstrum of the audio signal can be obtained by taking the logarithm of the modulus after Fourier transform of the signal and then calculating the inverse Fourier transform [17]. In practical application, the linear prediction cepstrum coefficient is obtained by the following recursion based on the relationship between it and the linear prediction coefficient:

$$C_1 = a_1, \quad (14)$$

$$C_n = a_n + \sum_{j=1}^{n-1} \frac{i}{n} a_{n-j} C_j, \quad 2 \leq n \leq p, \quad (15)$$

$$C_n = \sum_{j=1}^p \frac{i}{n} a_{n-j} C_j, \quad n > p. \quad (16)$$

3.3. Features of Content-Based Audio Retrieval

- (1) *Extracting Information Clues from Media Content.* Content-based retrieval breaks through the limitation of traditional keyword-based retrieval. Audio is retrieved according to the inherent characteristics of audio itself rather than the external attributes or keywords manually marked, making the retrieval closer to media objects [18]. Its core idea is to analyze the structure and semantics of audio through certain computer processing and establish their structured organization and index, so that the "disorderly" audio becomes "orderly," which is conducive to users' retrieval and browsing.
- (2) *Similarity Retrieval.* This is an important feature of content-based audio retrieval. For audio, the content is imprecise, and the sensory and expressive inconsistencies greatly increase the difficulty of processing. Therefore, content-based audio classification can only be a kind of similarity classification, abandoning the traditional exact matching and avoiding the uncertainty caused by the traditional retrieval

TABLE 8: Temporary data table (record the corresponding thesaurus information after word segmentation).

The column name	The data type	Explanatory notes
Id	Int	The table's primary key
Raw_id	Int	idId of the corresponding tb_raw entry
Keyword	Text	Formatted text after word segmentation
Type	Varchar (200)	Site type

TABLE 9: Core keywords (record the core keywords concerned by the website).

The column name	The data type	Explanatory notes
Id	Int	The table's primary key
Name	Varchar (200)	Filter keywords throughout the site

methods, but the results often appear false detection and omission

- (3) *Fast Retrieval of Large Databases (Sets)*. For a large number, a wide variety of multimedia databases. It can realize the rapid retrieval and positioning of multimedia information
- (4) As a multimedia technology, it has strong interactivity, that is, users can participate in the retrieval process [19]

4. Context Information Fusion Algorithm in Scene Semantic Parsing

The basic idea of this method is shown in Figure 3: this essay tries to introduce a sampling module to reduce the spatial dimension of the key matrix and value matrix. The method proposed in this essay has achieved excellent performance on three challenging semantic segmentation datasets Cityscapes, ADE20K, and PASCAL Context [20]. In terms of time and space efficiency, APNB runs about 6 times faster on GPU than NB for 256×128 input feature map and occupies about 28 times less GPU memory space.

Assume that the input characteristic diagram of NB is denoted as

$$X \in R^{C \times H \times W}, \quad (17)$$

where C, W, H , respectively, represent the number of channels, width, and height of the feature graph. NB firstly adopt three parallel 1×1 convolution operations to transform X into three different feature graphs, called query feature graphs $\varphi \in R^{C \times H \times W}$, key feature graphs $\theta \in R^{C \times H \times W}$, and value feature graphs $\gamma \in R^{C \times H \times W}$, respectively. The above

transformation process can be expressed as

$$\varphi = W_{\varphi}(\chi), \theta = W_{\theta}(\chi), \gamma = W_{\gamma}(\chi), \quad (18)$$

where C is the number of channels in the feature graph after transformation. Then, NB flattens the query feature graph, key feature graph, and value feature graph along the spatial dimension and transforms them from dimension $R^{C \times H \times W}$ to $R^{C \times H}$. Where $N = H \cdot W$ represents the number of pixels of all spatial positions in the feature graph [21]. Then, the similarity matrix $V \in R^{N \times N}$ can be calculated by matrix multiplication.

$$V = f(\varphi^T \times \theta), \quad (19)$$

where f represents the regularization function, and the usual choice is the Softmax function, which guarantees that the sum of each row of the similarity matrix is 1. After obtaining the similarity matrix V , NB will further fuse the long-range context information of the whole feature graph through matrix multiplication:

$$O = V \times \gamma^T. \quad (20)$$

For each spatial position in feature graph O , its value is the weighted sum of the features of all spatial positions in value feature graph, so the long-range context information is effectively integrated [22]. In order to avoid information loss and facilitate gradient propagation and optimization of network, in NB, feature graph O and input feature X , which are integrated with long-range context information, are generally fused together by addition or splicing:

$$Y = W_o(O^T) + X \text{ or } T = \text{cat}(W_o(O^T), X). \quad (21)$$

W_o is also a 1×1 convolution layer, which mainly plays two roles. On the one hand, it can restore the number of channels represented by the feature O from \hat{C} to C to keep consistent with the input feature X . On the other hand, the convolution layer W_o can be used as a weighting factor to adjust the importance of context feature O and input feature X to the final output feature graph Y .

5. Design of Content Detection and Analysis System of University Campus Network Based on Network Public Opinion

5.1. System Architecture Design. The purpose of the campus network public opinion detection and monitoring work is to detect and deal with public opinion in time. Therefore, for the monitoring of public opinion on related websites, it is necessary to build a complete set of internal monitoring mechanism for public opinion information in colleges and universities. In the era of big data, the monitoring of public opinion in schools must be supported by a strong data collection and analysis technology platform. The system mainly consists of processing system, analysis system, collection system, and report system. Only with the cooperation of each

TABLE 10: User table (record user name, password, and type information).

The column name	The data type	Explanatory notes
Id	Int	The table's primary key
Username	Varchar (45)	User login name
Password	Varchar (45)	The user password
Type	Int	Type 1. Common user 2. The administrator

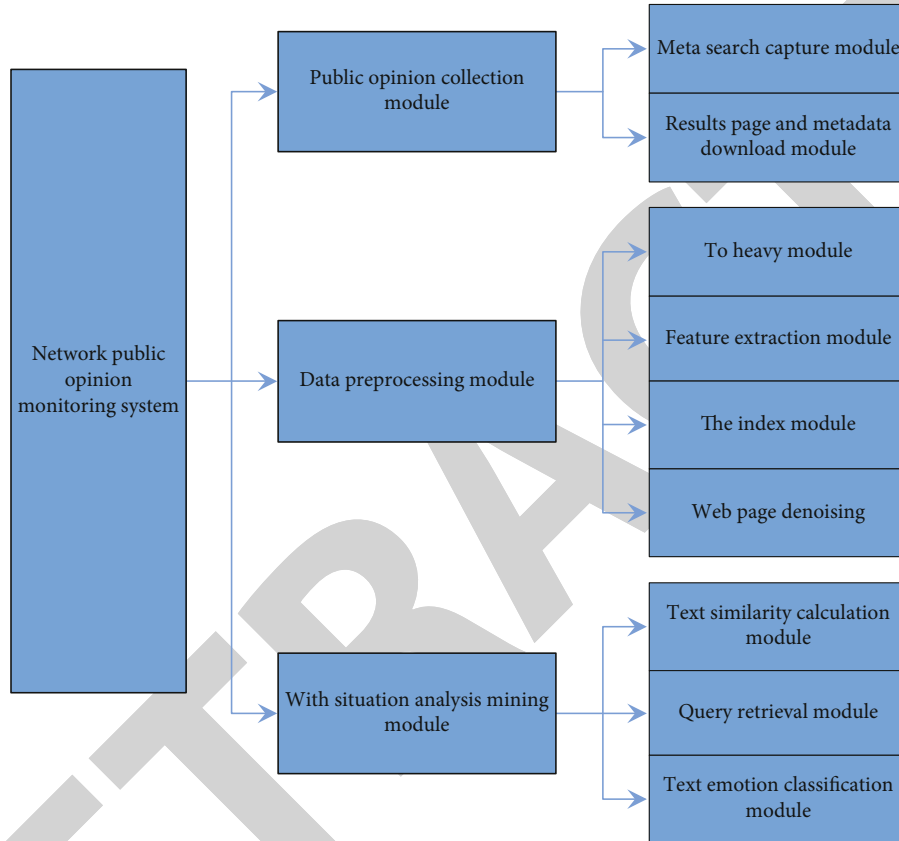


FIGURE 8: Functional module division of public opinion monitoring system.

module can public opinion retrieval, event trend analysis, and corpus collection be completed. Here, corpus collection focuses on news websites and media with strong interactivity with a large number of netizens' comments [23]. For the collection of the above media public opinion corpus, it can use metasearch method to obtain the latest information, can use event tracking and trend analysis technology to grasp the topic emotion direction, and can take the form of graphs and tables to show the results of public opinion analysis. Figure 4 shows the system technical architecture.

5.2. System Case Test. For school network public opinion detection, the relevant monitoring content can also be divided into three different words, such as subject, object, and emotional tendency, and then the words are dynamically combined and matched to generate relevant public opinion topics, and then use professional big data to monitor analytical technology, multidimensional collection of relevant information. Different users, its needs reflect

personalized characteristics. In order to meet the demand, the retrieval function needs to be further improved. The top layer mainly uses the content in Figure 5, which mainly includes information search, information statistics, information collection, information analysis and classification, data processing, data storage to the archives, and unified system management.

5.2.1. Subsystem Function Use Case Analysis. Figure 6 shows web page information collection. The web crawler technology is applied to obtain more accurate web page information, which can effectively eliminate the information irrelevant to their own needs, so that the content of the database can be updated in time.

Generally collected information comes from the Internet, but there is no way to directly collect and extract information on the Internet. Even personalized search cannot be implemented [24]. Therefore, it is necessary to conduct a detailed investigation on the user's own preferences and

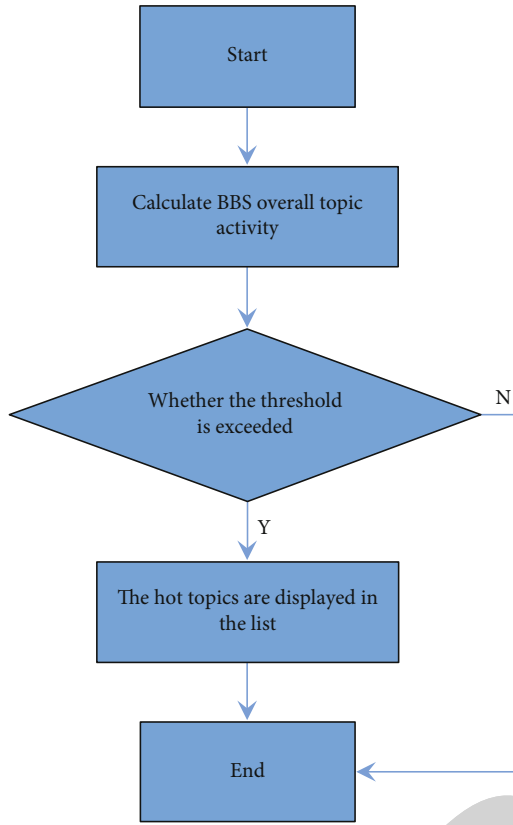


FIGURE 9: Flow of public opinion analysis module.

background and divide the browsing scope of the content of the web terminal. In combination with the specific situation of public opinion analysis, the scope of information collection in the web terminal is obtained. For details, please refer to Table 1.

In the supervision of online public opinion, the following two functional requirements should be realized: (1) mining public opinion information; (2) collect public opinion content. For end users, it can configure public opinion collection services in the field. B/S method is adopted, and it can have the function of follow-up tracking of public opinion information and emotion discovery. The following Figure 7 is the use case diagram of the public opinion monitoring system.

5.2.2. *System Database Design.* According to the above business operating procedures, database design concepts, and system module construction, the mining and analysis of university campus media can be realized. This database needs to contain the following Tables 2–10.

5.2.3. *System Module Division.* This public opinion supervision and management tool can effectively help public opinion management departments to quickly distinguish and analyze information. The collection, analysis, classification, and other contents of the media content in colleges and universities are assembled to build a relatively complete system. The system can complete intelligent network information collection, retrieval, public opinion release, emotion track-

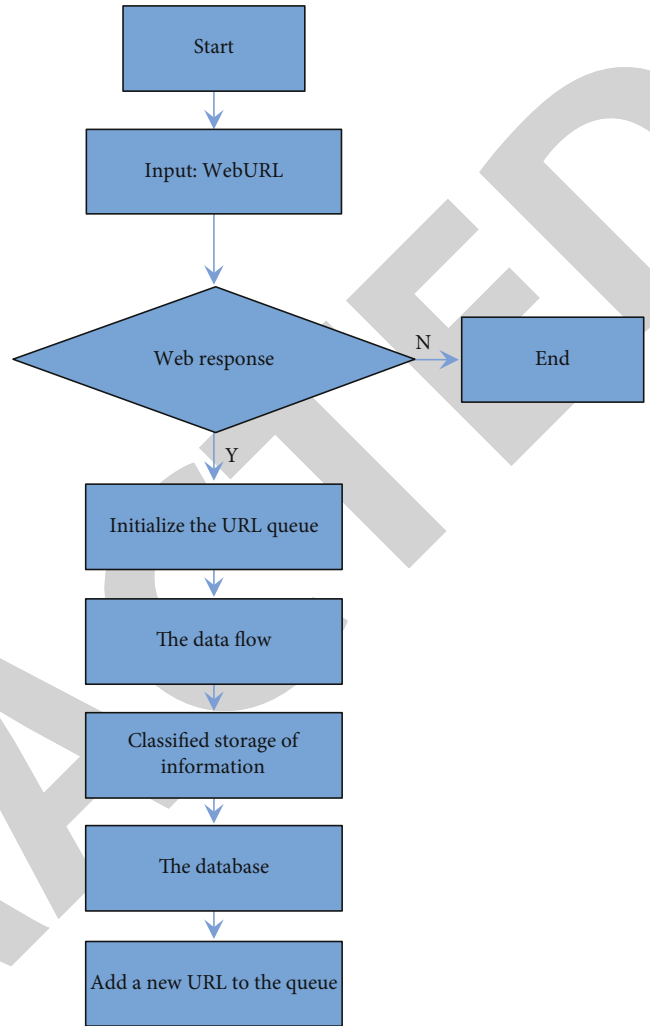


FIGURE 10: Flow of public opinion collection module.

ing, statistical reports, and other functions, as shown in Figure 8.

According to the content of the above figure, it can be known that the entire network public opinion monitoring platform involves a series of modules such as collection, mining, analysis, and processing of media content in colleges and universities, as shown in Figures 9–12.

The collection system of media content in colleges and universities is mainly to be able to discover and collect new media content in time. The module also contains two sub-modules: metasearch capture module and results page and metadata download module. The former takes the returned URL and aggregates the keywords into the search set. The latter is responsible for extracting metadata of query results, downloading, and saving snapshots of result pages, and the data preprocessing module mainly cleans and standardizes data [25]. Data preprocessing module mainly includes the elimination of repeated data, feature module, and index module for the web page. In the analysis and mining module, the positive and negative information of the whole public opinion content is deeply mined, and there are three

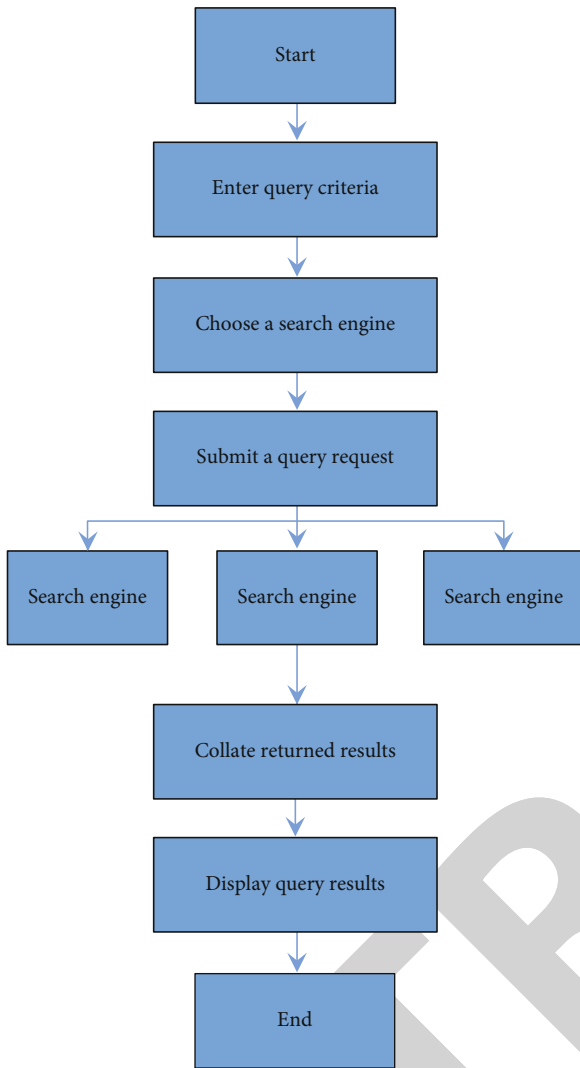


FIGURE 11: Yuan search module process.

small modules mainly including text content, emotional content, and text similarity statistics.

5.2.4. *System Deployment Diagram.* As for the public opinion detection system structure on campus network of colleges and universities, it mainly depends on the Internet to collect and analyze network information. Therefore, the system can be roughly divided into two parts, namely, front-end acquisition and back-end analysis. In addition, a special isolation device must be used between front-end acquisition and back-end analysis to ensure the security performance of back-end analysis system. Public opinion monitoring system mainly consists of three parts, namely, front-end collection, network isolation, and background analysis. To be specific:

(1) The front-end acquisition

The general collection server can use firewall to carry out the collection process of Internet information data. Moreover, front-end acquisition can not only receive processing instructions from the back-end analysis platform but also

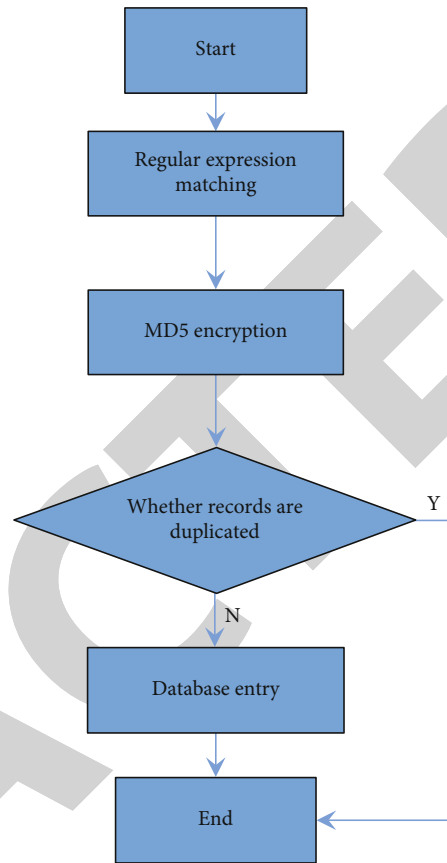


FIGURE 12: Process of module removal.

safely transfer data to the back-end processing platform in the form of files.

(2) Safe isolation

The main task of the security isolation device is to isolate and distinguish the front-end and back-end platforms. In this way, the information and instructions of the front-end and back-end platforms can be correctly transmitted to ensure the overall data security.

(3) Back-end analysis and processing

The back-end analysis and processing platform is mainly composed of data analyzer, data processor, switching equipment, terminal, content operation processor, and other related equipment. It can categorize and store data and balance each process under load.

(4) Feature extraction module

An article is composed of words, words, paragraphs, and chapters. In some articles, some features can be selected to represent the content of the article. As the words in the four parts are the most basic ideographic units, they can reflect the content characteristics of the text. If single words can show the characteristics of an article; then, they can represent some spatial vectors accordingly. In the process of feature extraction, all the words can be concentrated, and its

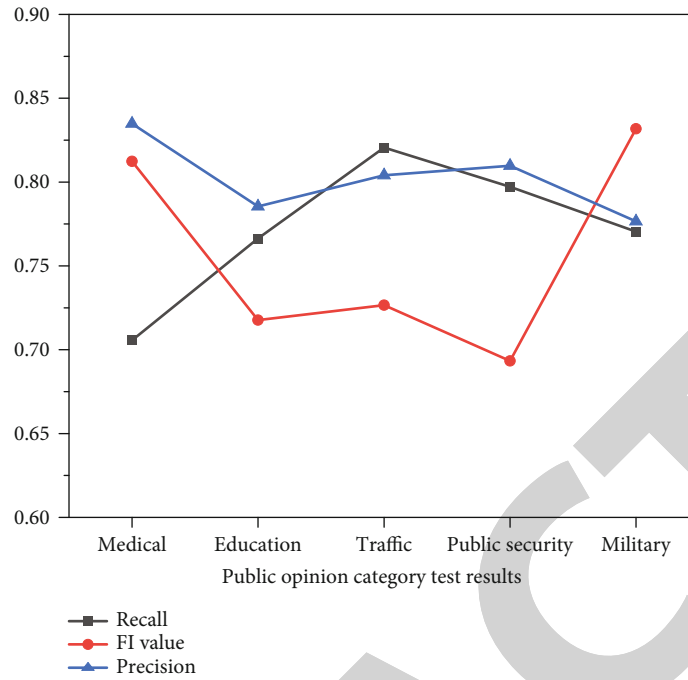


FIGURE 13: Public opinion category test results.

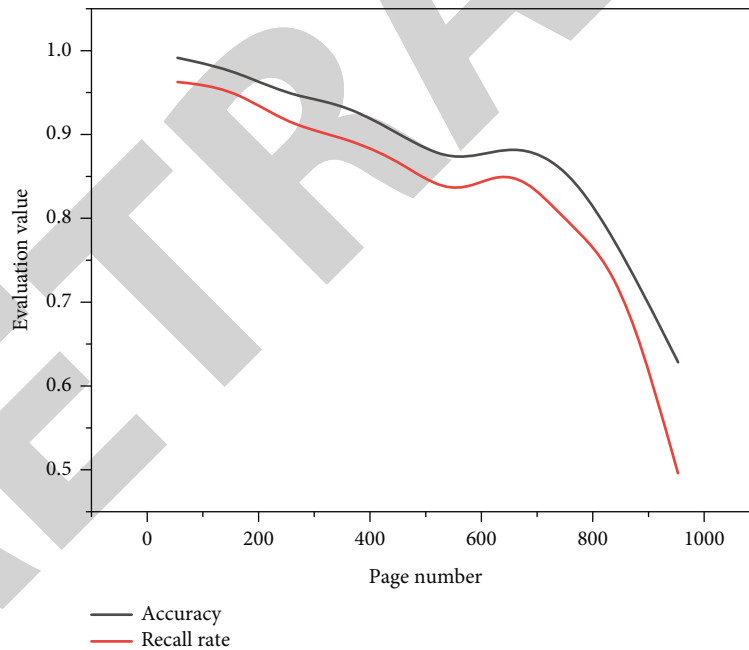


FIGURE 14: Performance test results of text classification based on the number of web pages.

principle is the expression method of the words in the article. It has a high similarity with a certain record of structured database. In a certain document, it can be said that feature vector can reflect the content characteristics of the document. Use multidimensional or one-dimensional web data to display text content and information, so that each content in the data table represents a feature and then forms a feature set. Each line can represent the functionality of the page, so the entire line is the statistics of the page. Based on

the representation method of TF-IDF vector, two-dimensional data table is formed. The distinguishing values in dictionaries include feature set and column set. For each column, there may be hundreds of thousands of columns in the whole column set. Each line stores page word information corresponding to the feature set. For each word in the feature set, if the word does not form a set in the web interface, the value is 0. If the number of occurrences on the page is K , the final value is K . In this case, the

construction of a two-dimensional table belongs to the statistics of the words in the collection of web pages. Also note that if the above method is used, it will represent the frequency of words on the page. Subset characteristics have been determined before mining, if the whole system reduces its dimension with $DF > 50$. In this case, we need to use rough set instead of attribute reduction.

5.3. System Test Module. This chapter mainly studies the source of college network public opinion in the whole system, carries on the statistics of the emotional tendency, and also carries on the experimental research for the scope of the whole public opinion. After that, the web crawler was used to timely capture public opinion and repeatedly test the algorithm. Figure 13 shows the random capture of 500 web pages in five industries, including medical care, transportation, education, public security, and military. The above web pages are clearly divided into statistics, and statistics are made for the corresponding industry types. As can be seen from the following figure, the recall probability of these five industry categories is higher than 70%, and the recall rate is greater than 80%, among which $F1$ value is greater than 70%. For $F1$ values, you can categorically classify them as response classes. The overall effect of this kind of medical and military industry is better, and the overall classification of transportation and education industry has basically the same effect. The overall effect of public security is relatively low, and the final reason is related to the data of training samples collected. The training sample content of public security can be adjusted again, so as to confirm the more suitable initial clustering center and confirm the accuracy of clustering again.

In order to analyze and verify the public opinion of the search system and determine its performance, Figure 14 clearly presents the classification based on the number of pages and text of the web page, so as to analyze the results and finally evaluate the corresponding accuracy rate and recall rate in terms of performance. Recall rate is also called weight check rate. In the figure above, random webpage extraction is carried out. It can be seen that with the increase of the number of webpage, the recall rate and accuracy rate of webpage classification also decrease. The reason is that the classification algorithm has not been considered in depth, and a large amount of web page information has been ignored, which is also a problem to be further studied and analyzed in the next step.

6. Conclusion

Times with the irreversible trend of rapid development, in people's work and life, the Internet is playing an increasingly important role. Compared with traditional media, the Internet has more prominent features of convenience and timeliness. On the Internet, everyone can surf the Internet in an infinite virtual space. From the characteristics of the network, it has a certain concealment. Many people prefer to choose this mode when expressing their opinions, and the concept of online public opinion is derived from this background. Due to the rapid updating and upgrading of the

Internet, online public opinion has gradually replaced other media as the main way for people to spread public opinions. This makes all countries attach more importance to online public opinion, and China attaches more importance to this issue. College departments focus on online public opinion factors to conduct research and take this as the basis for improving management methods. It can be judged that the trend of social public opinion is largely affected by online public opinion, so it is necessary to strengthen dynamic management of online public opinion and increase attention. Based on network public opinion, a content detection and analysis system for college campus network is established. Especially for colleges and universities, it is better to strengthen the campus network public opinion monitoring, content detection, and classification, to provide students with a healthier and more perfect campus network environment support.

Data Availability

The dataset can be accessed upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] H. Zhang, S. Qian, Q. Fang, and C. Xu, "Multi-modal meta multi-task learning for social media rumor detection," *IEEE Transactions on Multimedia*, vol. 24, pp. 1449–1459, 2021.
- [2] S. Kaur, S. Singh, and S. Kaushal, "Abusive content detection in online user-generated data: a survey," *Procedia Computer Science*, vol. 189, no. 7, pp. 274–281, 2021.
- [3] K. Shu, D. Mahudeswaran, S. Wang, D. Lee, and H. Liu, "Fake-newsnet: a data repository with news content, social context, and spatiotemporal information for studying fake news on social media," *Big Data*, vol. 8, no. 3, pp. 171–188, 2020.
- [4] K. Xu, F. Wang, H. Wang, and B. Yang, "Detecting fake news over online social media via domain reputations and content understanding," *Tsinghua Science & Technology*, vol. 25, no. 1, pp. 20–27, 2020.
- [5] F. M. Plaza-Del-Arco, M. D. Molina-Gonzalez, L. A. Urena-Lopez, and M. T. Martin-Valdivia, "A multi-task learning approach to hate speech detection leveraging sentiment analysis," *Access*, vol. 9, pp. 112478–112489, 2021.
- [6] H. S. Alsagri and M. Ykhlef, "Machine learning-based approach for depression detection in twitter using content and activity features," *IEICE Transactions on Information and Systems*, vol. E103.D, no. 8, pp. 1825–1832, 2020.
- [7] L. Verdoliva, "Media forensics and deepfakes: an overview," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 5, pp. 910–932, 2020.
- [8] S. Ni, J. Li, and H. Y. Kao, "Mvan: multi-view attention networks for fake news detection on social media," *Access*, vol. 9, pp. 106907–106917, 2021.
- [9] Q. Guan, G. Wei, L. Wang, and Y. Song, "A novel feature points tracking algorithm in terms of imu-aided information fusion," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 8, pp. 5304–5313, 2020.

- [10] E. Guo, V. Jagota, M. Makhatha, and P. Kumar, "Study on fault identification of mechanical dynamic nonlinear transmission system," *Nonlinear Engineering*, vol. 10, no. 1, pp. 518–525, 2021.
- [11] Y. Fang, L. Yu, and S. Fei, "An improved moving tracking algorithm with multiple information fusion based on 3d sensors," *Access*, vol. 8, pp. 142295–142302, 2020.
- [12] Z. He and W. Zhang, "High resolution information reserved anchor-free detection algorithm," *Journal of Computer-Aided Design & Computer Graphics*, vol. 33, no. 4, pp. 580–589, 2021.
- [13] Y. Qiu, C. Sun, and J. Tang, "Seismic attribute fusion approach using optimized fastica-based blind source separation algorithm," *Geophysical Prospecting for Petroleum*, vol. 57, no. 5, pp. 733–743, 2022.
- [14] J. Chen, J. Liu, X. Liu, X. Xiaoyi, and F. Zhong, "Decomposition of toluene with a combined plasma photolysis (CPP) reactor: influence of UV irradiation and byproduct analysis," *Plasma Chemistry and Plasma Processing*, vol. 41, no. 1, pp. 409–420, 2021.
- [15] S. Dong, Y. Quan, W. Feng, G. Dauphin, and M. Xing, "A pixel cluster cnn and spectral-spatial fusion algorithm for hyperspectral image classification with small-size training samples," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 4101–4114, 2021.
- [16] L. Yi, S. Ji, L. Ren, R. Su, and Y. Liang, "A nonlinear feature fusion-based rating prediction algorithm in heterogeneous network," *IEEE Transactions on Computational Social Systems*, vol. 8, no. 3, pp. 728–736, 2021.
- [17] N. Chilamkurti, "A secure, energy- and sla-efficient (sese) e-healthcare framework for quickest data transmission using cyber-physical system," *Sensors*, vol. 19, no. 9, p. 2119, 2019.
- [18] D. Li, X. Jia, and J. Zhao, "A novel hybrid fusion algorithm for low-cost gps/ins integrated navigation system during gps outages," *Access*, vol. 8, pp. 53984–53996, 2020.
- [19] A. K. Bhandari, A. Ghosh, and I. V. Kumar, "A local contrast fusion based 3d otsu algorithm for multilevel image segmentation," *IEEE/CAA Journal of Automatica Sinica*, vol. 7, no. 1, pp. 203–216, 2020.
- [20] R. Huang, S. Zhang, W. Zhang, and X. Yang, "Progress of zinc oxide-based nanocomposites in the textile industry," *IET Collaborative Intelligent Manufacturing*, vol. 3, no. 3, pp. 281–289, 2021.
- [21] S. Zhang, Y. Wang, P. Wan, J. Zhuang, and Y. Li, "Clustering algorithm-based data fusion scheme for robust cooperative spectrum sensing," *Access*, vol. 8, pp. 5777–5786, 2020.
- [22] J. Qu, Y. Li, Q. Du, and H. Xia, "Hyperspectral and panchromatic image fusion via adaptive tensor and multi-scale retinex algorithm," *Access*, vol. 8, pp. 30522–30532, 2020.
- [23] K. Li, Y. Gong, and Z. Ren, "A fatigue driving detection algorithm based on facial multi-feature fusion," *Access*, vol. 8, pp. 101244–101259, 2020.
- [24] P. Ajay, B. Nagaraj, and J. Jaya, "Bi-level energy optimization model in smart integrated engineering systems using WSN," *Energy Reports*, vol. 8, pp. 2490–2495, 2022.
- [25] Y. Zhuang and Z. Fang, "Smartphone zombie context awareness at crossroads: a multi-source information fusion approach," *Access*, vol. 8, pp. 101963–101977, 2020.