IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON NOVEL LEARNING APPLICATIONS AND SERVICES FOR SMART CAMPUS

Received November 9, 2018, accepted November 19, 2018, date of publication November 28, 2018,
date of current version December 27, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2883799

Real-Time Monitoring of Smart Campus and Construction of Weibo Public Opinion Platform

FENG NAN[1], YINA SUO[2], XUEYONG JIA[3], YUYING WU[3], AND SHUAIJIE SHAN[3]

[1]College of Electrical Engineering, North China University of Science and Technology, Tangshan 063000, China
[2]Information Engineering Institute, North China University of Science and Technology, Tangshan 063000, China
[3]Key Laboratory of Engineering Calculation in Tangshan City, North China University of Science and Technology, Tangshan 063000, China

Corresponding author: Yina Suo (suoyina203@126.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 51674121 and in part by the Natural Science Foundation of Hebei under Grant E2017209178.

ABSTRACT With the development and penetration of information technology, the ''fourth media''-the network media is coming, and the ''grievance crisis'' is increasingly happening. The effective monitoring of online public opinion information becomes a problem. For each university, how to understand the sensation of teachers and students in real time in the era of informationization entering the intelligent campus has become an urgent problem for all colleges and universities. This paper first analyzes the characteristics of Weibo's communication and the process of microblogging's public opinion formation, the mechanism of communication, and so on. For the acquisition of this information, the reptile method was used to obtain the required data, which was beneficial to the real-time sensation in the smart campus. Then, based on the improved single-pass algorithm to analyze the new characteristics of microblog propagation information and the public opinion sent by college students, and compared the improved single-pass algorithm with the effect of single-pass algorithm, the improved algorithm had the recall rate was high, the false detection rate was low, and the running time was short. Finally, taking the lyrics of college teachers and students as the research object, unified modeling language modeling was used to analyze the public opinion monitoring platform of the smart campus, and the campus public opinion of several major events was monitored and analyzed. This provided a reference for the effective, intelligent, and real-time detection of Weibo public opinion in colleges and universities.

INDEX TERMS Public opinion monitoring, single-pass algorithm, UML, smart campus, Weibo.

I. INTRODUCTION

With the popularity of the mobile Internet, a large number of social media such as Weibo and WeChat have sprung up, and the proportion of college students using smartphones is basically 100%. Because the mobile Internet not only has the characteristics of openness, inclusiveness, immediacy, fairness and low cost of the traditional Internet, it is more convenient, which makes the users of teachers and students in colleges and universities grow, and the network rapidly develops into the fourth media [1].

In recent years, many experts and scholars have conducted in-depth discussions and research on smart campuses. Huang [2] proposed five basic characteristics and definitions of a smart campus, and proposed five key technologies to support the construction of a smart campus. On the basis of analyzing the transformation of digital campus to smart campus, Chen and Liu [3] elaborated on the integration of current information technology and smart campus technology bearer and application, and put forward strategies and suggestions for developing smart campus. On the basis of analyzing the smart campus and its connotation features and key technologies, Wang [4] designed the overall architecture model of smart campus construction, and analyzed the typical application of smart campus from many aspects.

Because the Single-Pass clustering algorithm has obvious advantages of being simple, intuitive and easy to understand, the algorithm is used by many domestic and foreign scholars in the microblog topic detection module. Based on the classical Single-Pass clustering algorithm, Ye [5] screens the more representative keywords based on the feature weights of the

2169-3536 © 2018 IEEE. Translations and content mining are permitted for academic research only.
Personal use is also permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

central vector. Before using the improved Single-Pass algorithm to detect the microblog topic, pre-process the isolated points in the text to make the topic more accurate. Seo [6] use the Single-Pass algorithm to detect events. Although this method is fast and computationally simple, its execution results are highly dependent on the order of processing data. On the basis of in-depth analysis of a large number of news reports and corpora, Fang [7] improved Single-Pass for its shortcomings in timeliness and accuracy of text clustering.

The research on Weibo public opinion monitoring platform has attracted more and more scholars' attention. Buzzlogic's "Buzz Logic Insights" service helped companies discover, attract and evaluated industry influenced by providing high-impact, comprehensive, multi-angle sensational dynamic analysis of blogs [8]. Diakopoulos used a timeline-based graphical approach on the Twitter dataset to conduct a perspective analysis of microblogging in Obama's inauguration and other events [9]. Yahia *et al.* designed a system "MAQSA" to analyzed news events microblogging. The system combined relevant microblogs with news reports for analysis of relevant news events, in order to promote journalists to timely grasp the reaction of public opinion to news reports, and expand the attention of journalists [10]. Qu *et al.* [11] studied the effects of earthquake-related microblogging on natural disasters. Cao *et al.* designed and implemented a system for automatically collecting news microblogs to automatically generate user comments for other news sites. Based on Sina Weibo, Fu Yulin designed a microblog summary system that displays in time axis visualization. Chen Xianfeng designed and implemented a microblog public opinion analysis system based on Sina Weibo. The system used Weibo API to collect microblog data and used the topic model to analyze microblog text clustering [12].

This paper taked the teachers and students of colleges and universities as the research object, analyzed the new characteristics of microblogging information and the sentiment of microblogs by using the improved Single-Pass algorithm, and then used UML language to model and analyze the detection platform of smart campus.

Section II division of this paper introduced the smart campus in the information age. Section III introduced in detail the overall architecture model of the smart campus microblogging public opinion monitoring platform. Section IV introduced the concept of microblogging lyrics, the characteristics of communication, and the use of college students' microblogging lyrics formation process, communication characteristics and internal mechanism to analyze and disseminate mechanisms. Section V analyzed the information collection function and information analysis function requirements of the public opinion monitoring platform. Section VI studied several key technologies (data preprocessing, text modeling, text clustering, etc.) of the intelligent campus public opinion monitoring platform, and used the improved Single-Pass algorithm in the microblog topic detection. Section VI used UML language to model and analyze the smart campus public opinion detection platform. Section VIII analyzed the public opinion obtained by the public opinion monitoring system of the smart campus.

II. INFORMATIZATION ENTERS THE ERA OF SMART CAMPUS

The smart campus is a new stage in the construction of information technology in colleges and universities. Its characteristics include high-speed ubiquity of the Internet, Internet of Things, seamless interconnection and environment awareness, widespread popularization and application of intelligent terminals, open collaborative learning and research environment, and working environment. Smart campus is closely integrated with new Internet technologies, which using cloud computing, virtualization and the Internet of Things to change the forms of interaction, knowledge transfer and resource sharing among campus users. The school's teaching, research, management and resources, applications are integrated to improve the flexibility, comprehensiveness and ease of use of the service, thus achieving a new model of campus intelligent service and management [13].

From digital campus to smart campus, we need to integrate new service concepts, management services and information sharing mechanisms to optimize processes and improve management. Based on the application requirements and design principles of smart campus functions, the overall construction of the smart campus is shown in Figure 1.

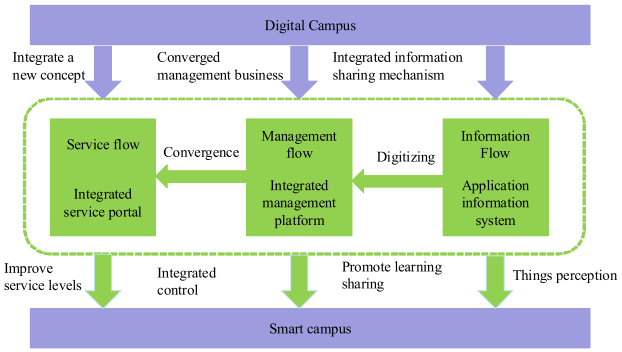


FIGURE 1. The process of building a smart campus.

The connotation and ultimate purpose of the smart campus is service. Based on the digital campus, the Smart Campus pursues a high degree of integration of resources and applications, emphasizing the provision of personalized services for users and more intelligent and humanized services.

III. THE OVERALL ARCHITECTURE MODEL OF THE SMART COMPUS MICROBLOGGING PUBLIC OPINION MONITORING PLATFORM

Based on the reference to the wisdom campus planning and construction model of many universities in China, we designed the overall loose model of smart campus construction, as shown in Figure 2. From the bottom to the top, the model is divided into five levels: the perception layer, the network layer, the data layer, the application layer

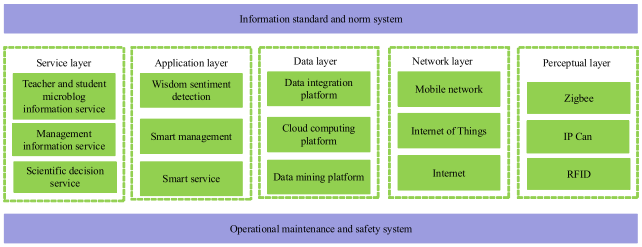


FIGURE 2. Smart campus public opinion monitoring platform architecture model.

and the service layer,and the supplemented by information standards and normative systems, operation and maintenance and security systems. These two guarantee systems guarantee the standard construction and operation and maintenance of the smart campus microblog public opinion monitoring platform [14], [15].

A. PERCEPTUAL LAYER

In the smart campus, through various application systems, the perception layer has a comprehensive perception of the campus personnel, equipment and other environments, including the perception of objects and objects, the perception of people and things, and the real-time perception, capture and transmission of information between systems. The platform gains information by sensing microblogging software.

B. NETWORK LAYER

The main function of the network layer is to realize the interconnection and interoperability of various networks such as mobile network, Internet of Things, and campus network, and realize the comprehensive interconnection, intercommunication and interaction between people, and people, things and things, and people and things in the nuclear garden. Provide high-speed, ubiquitous network conditions for anytime, anywhere, on-demand applications to enhance microblogging's access to information and real-time services.

C. DATA LAYER

The data layer is responsible for the full integration of collected information, data retrieval and intelligent analysis. Relying on multi-source, massive unstructured and structured data precipitated in smart campuses, all of this data is mined through Hive-Hadoop clusters, and the results of data analysis and processing are stored in a dedicated database for using by systems and users. The smart campus aims to realize personalized service, and objectively requires the mining of the actual needs of campus users. The microblogs associates with teachers and students in colleges and universities contain a lot of personal information about teachers and students, hot topics and other content, which provides conditions for improving user experience, improving service processes and improving service quality through data mining.

D. APPLICATION LAYER

The application layer mainly provides personalized services and intelligent decision-making services. Through typical business systems such as educational management system, personnel management system, asset management system, and new business systems such as sensing systems and social network systems, various application systems are highly integrated to build open display information. Finally, the application service for the real-time microblogging public opinion situation of teachers and students is realized.

E. SERVICE LAYER

Through the integration of various services, the smart campus is no longer an isolated application system in front of users, but a unified and friendly use of the portal interface, an integrated information service platform (practice information portal). The unified information portal provides a unified access portal and portal interface to provide personalized information services for users with different authorization roles. Users only need to access the personalized campus portal to query, interact and agree on various information resources. At the same time, the information service method improves management efficiency and management level, and helps to monitor service quality and improve service capabilities.

F. INFORMATION STANDARD AND NORM SYSTEM

The information standard and normative system determines the standards and norms of information gathering, information processing, information exchange and other processes, standardizes the data structure of the application system, meets the requirements of information construction, and lays a foundation for data fusion and service integration.

G. OPERATIONAL MAINTENANCE AND SAFETY SYSTEM

Operational maintenance and safety system are important guarantee for the normal operation of smart campuses. Security in a smart campus involves four aspects: physical security, network security, data security, and content security. Physical security includes device security, environmental security, disaster recovery backup, and media security. Network security includes risk assessment, security testing, data backup, tracking auditing, security protection, and more. Data security mainly involves database security, digital signatures, authentication technologies, and so on. Content security mainly includes data mining, privacy protection, and information filtering.

IV. AN OVERVIEW AND ANALYSIS OF COLLEGE STUDENTS' WEIBO PUBLIC OPINION

A. WEIBO LYRIC CONCEPT

For Weibo public opinion, different scholars have given different understandings of Weibo public opinion through different research directions. For example, Weibo's lyrics are based on facts, and the views of the netizens' attitudes have

spread their opinions and expressions in the communication interaction after the spread of Weibo. Its performance form highlights the subjectivity of the netizens and is directly posted on Weibo without verification and filtering by the gatekeeper. Weibo sensation is a concentrated expression of the public opinion of the public opinion on the carrier of Weibo. Weibo users have certain influences and inclinations on the emotions, attitudes and opinions caused by the problems related to all aspects of their social life.

B. THE SPREADING MECHANISM OF COLLEGE STUDENT' WEIBO LYRICISM

College students' Weibo lyrics refer to the sum of the information, the emotions, attitudes, opinions and opinions expressed by the participants in the main media for college students, which are interested in the new media Weibo. It is a concrete manifestation of Weibo's sensation in the special group of college students, and it has a different form of communication than the general network. The microblogging sensation of college students has become the main channel for gathering public opinion and the source of information for the formation of public sentiment. The discussion and mastery of its communication mechanism can help to carry out ideological education for college students, monitor and guide the microblogging sensation of college students. With regard to the dissemination mechanism of college students' microblogging sentiment, we can analyze the formation process, communication characteristics and internal mechanism of college students' microblogging lyrics [16].

1) THE FORMATION PROCESS OF COLLEGE STUDENTS' MICROBLOGGING PUBLIC OPINION

There are often hot topics on Weibo. Take Sina Weibo as an example. On the right side of its Weibo web interface, real-time hot topics are posted every day. Different microblogging owners arbitrarily hold different attitudes on hot topics or events, some raise doubts, some speak fiercely, some comprehensively analyze problems, put forward rational opinions, and some avoid lightness, and can't distinguish between primary and secondary. The attitudes they speak and the nature of their language are different. In the self-media environment, ''everyone is a media,'' when college students are interested in a hot event, they will constantly ''forward'' and ''comment'' to prepare and lay the foundation for the formation of Weibo's lyrics. When college students quickly gather the sudden onlookers to become the collective expression of opinions, it constitutes the microblogging sensation of college students in emergencies.

2) THE SPREADING CHARACTERISTICS OF COLLEGE STUDENTS' WEIBO LYRICISM

College students are active people using Weibo, often exchange ideas and communicate ideas on Weibo. After in-depth exchange of ideas and collision of views, Weibo public opinion can be more comprehensively reflect the positions and values of college students. College students' microblogging lyrics are important parts of online public opinion. Due to its communication characteristics and special subjects, college students' microblog lyrics have both commonality and individuality compared with online public opinion. Weibo lyrics belong to an important section of the network lyrics. On the basis of inheriting the characteristics of the network, the arbitrariness and freedom of Weibo's lyrics are more significant. However, it also has new characteristics of communication. The situation is shown in Figure 3.

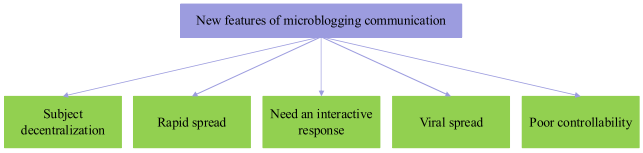


FIGURE 3. New features of Weibo's communication.

3) THE INTERNAL MECHANISM OF COLLEGE STUDENTS' WEIBO LYRICISM

College student is an important part of the youth group, and they are in the period of emotional ups and downs. They are bloody, have the ability to distinguish information, are not strong in self-control, and are easily impulsive. They think that the problem is not comprehensive enough, and they only look at things according to their own understanding, and they are easy to go to extremes. This kind of psychological instability makes college students lack objective and rational analysis of social problems, and is easily influenced by negative, inaccurate and rumored information on the Internet. Because information dissemination has a ''first cause effect,'' ''When people first come into contact with something or someone, they will be deeply impressed. The first input of the first impression has a strong influence on their later cognition. This also has a stronger effect on the overall impression of things than the information obtained in the future, and even more affects my understanding of new information.

V. ANALYSIS OF PUBLIC OPINION MONITORING NEEDS

The Weibo public opinion monitoring platform model of the Smart Campus will be studied in the following directions:

A. INFORMATION COLLECTION FUNCTION

In many cases, the access to the website requires the user to log in before accessing the resource. To obtain the information of the microblog page through the web crawler, it is necessary to implement the login through simulation. For the simulation of Sina Weibo, mainly using Fiddler for packet capture analysis, the Http Client is used for simulated login. The focus of the simulated login is to obtain the page jump and save the coolie.

The information collection system researches and implements a multi-threaded web crawler for Weibo by studying the open source crawler framework Nutch, Crawler4j and Web Collector crawler kernel, combining the mechanism

of Nutch kernel + plug-in and the good user interface of Crawler4j.

Information collection mainly consists of two parts, page capture and information storage. The page crawling is performed on the Weibo page by focusing on the web crawler, and the jsoup open source framework is used for page parsing, and the obtained information is saved to the database to prepare for the next information analysis.

B. INFORMATION ANALYSIS FUNCTION

Data preprocessing is a very crucial step in data mining and the most basic one. Because for the original data, because the data source is not single, the format is not uniform, the noise data, the existence of invalid data, will seriously affect the effect of clustering. This is especially true for Weibo data. In Weibo data, there is a large amount of useless data for the extraction of hot topics, similar to advertising microblogging, personal microblogging, etc., before the keyword extraction, the need to filter data .

Information Analysis This section also includes two parts, data preprocessing and data analysis. The web crawler will process the captured webpage information through transcoding and segmentation processing, and then perform data denoising, word segmentation, stop word filtering, and advertisement word filtering through the data preprocessing module. The data after data preprocessing can continue to analyze the data.

1) ANALYSIS OF MICROBLOG ELEMENTS

With Sina Weibo as the research platform, the information in Weibo can be divided into two categories: user information and Weibo items. Each piece of information contains many detailed information. The specific information categories and contents are shown in Table 1.

TABLE 1. Weibo information categories and content.

User Info	User identification information	User name, user ID, etc.
Weibo item information	Basic user information	User city, personal description, etc.
	User activity information	Number of followers, fans and Weibo and list
	User identification information	Publish Weibo user, user ID
	Weibo item text	Post microblog content
	Weibo item attribute information	Release time, channel, forwarding and comments
	Weibo item reference information	Reference original Weibo item information

2) ANALYSIS OF MICROBLOGGING PUBLIC OPINION MONITORING FACTORS

The network public opinion caused by campus emergencies is directly related to campus security issues, so it is necessary to carefully analyze and judge the relevant information of emergencies. The occurrence and development of emergency network grievances is generally reflected by the identification of certain variables. These variables are crucial for monitoring the current development of emergency network grievances. The external manifestation of network public opinion transmission is the change of the amount of network information, so the degree of influence of network public opinion transmission can be described according to the change of the number of network information. Through the stable relationship between the amount of network information and the degree of influence of the network public opinion, it can describe the changing process of the degree of network public opinion. With the change of information volume, the network public opinion presents different situations. The changes of these situational factors determine the degree of influence of public opinion. Therefore, it is possible to find the outbreak point of public opinion by monitoring the change of information volume, and provide a basis for the pre-emptive processing of emergency network public opinion [17].

The authentication category is a visual display of Weibo users' social influence and authority in the Weibo network environment, and is divided into different certification levels according to their influence. In Sina Weibo, there are three certification categories: ordinary users, Weibo Daren and real-name certified large V users. The potential influence and credibility of these three categories on public opinion transmission are in turn enhanced. The number of microblogs will affect the user's communication energy. The more microblogs are published, the more active the users are, and the more likely they become opinion leaders, speeding up the speed of public opinion.

This paper studies the microblogging network with emergency events as the starting point. Different from the previous online social network, the strategy of crawling microblog data is divided into four steps:

1) Through the microblog search page, enter keywords to obtain the original microblog related to the incident;
2) Crawl the original Weibo forwarding and comments;
3) Crawl the personal information of the users involved in it;
4) Crawling user relationships refers to the user's concerns.

Therefore, the overall demand framework of the system is shown in Figures 4 and 5:

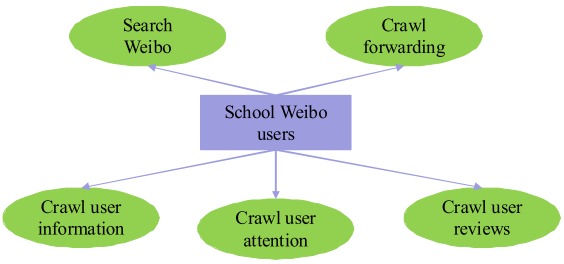


FIGURE 4. Reptile system analysis.

The public opinion monitoring system in the smart campus is obtained by crawling the data and stored in the database,

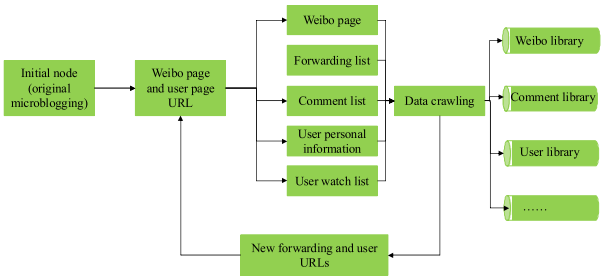


FIGURE 5. Crawler program logic diagram.

and then through the processing and analysis of the data, and the final lyric results are displayed by the public opinion monitoring platform of the smart campus.

C. DATABASE SYSTEM ARCHITECTURE

The database system is divided into three modules: data service agent layer, database layer, Memcached data cache layer, and its system architecture diagram is shown in figure 6. The main function of the data service proxy layer is to receive the client's application (request) and successfully send the returned result of the system to the client. The master database and the slave database (Slave) together form the database layer (DB Layer), which is composed of the My SQL cluster of the master-slave architecture. The main database manages the corresponding slave database, and uses the master-slave mode to read and write data. Although the read operation can only be performed from the database, the pressure on the primary database is still relieved to some extent, so that the load of the system reaches an effective equilibrium state. Memcached data cache layer belongs to the memory of the system. Its main function is to store the object obtained by query. The speed of reading and writing data from memory is much faster than the speed of reading and writing data directly from the database. Therefore, the read and write efficiency of the system will be significantly improved.

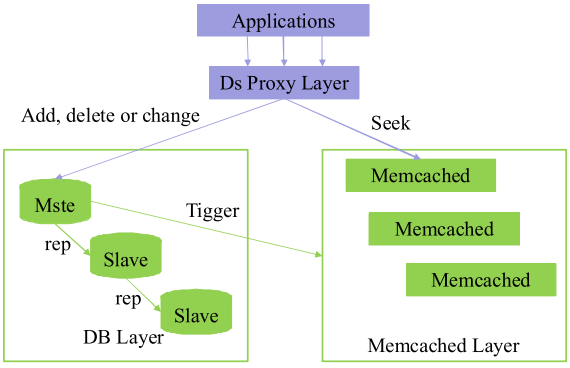


FIGURE 6. System module information interaction.

The three functional templates of this system are interconnected and cooperate with each other. The information interaction between the modules is shown in figure 7.

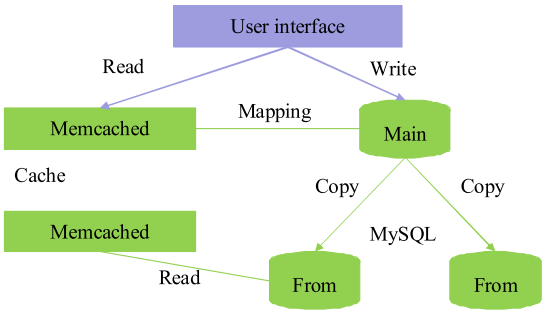


FIGURE 7. Crawler program logic diagram.

VI. APPLICATION OF IMPROVED SINGLE-PASS ALGORITHM IN MICROBLOG TOPIC DETECTION

This paper studies several key technologies (data preprocessing, text modeling, text clustering, etc.) of the intelligent campus public opinion monitoring platform. The specific workflow of microblog topic detection: Firstly, based on Sina Weibo open platform crawler method to collect microblog information; secondly, pre-processing the collected data, mainly including filtering spam, word segmentation and part-of-speech tagging, etc. Then, the microblog text is modeled, including text representation, text feature extraction, feature weighting, etc., on the basis of which the text similarity is calculated; finally, the appropriate clustering algorithm is selected to merge the topics.

A. MICROBLOG DATA PREPROCESSING

Selecting valuable and research-oriented information (such as Weibo information content) from the original microblog data collected is the primary task of data preprocessing. However, in order to make the information easy to analyze and process, it is necessary to cut the extracted information to lay the foundation for the subsequent topic detection. In this chapter, the data preprocessing is divided into three steps: filtering noise information, word segmentation and part-of-speech tagging, and removing stop words for the characteristics of microblog data.

1) FILTERING NOISE INFORMATION

Due to the grassroots, autonomy, and repetitive nature of Weibo text content, there is very little information in Weibo that is too complicated, has a high repetition rate, and is truly practical. Therefore, how to filter out the useless data contained in Weibo plays a decisive role in the later research.

By mining the original microblog data, selecting the corresponding processing method according to their own special format to exclude filtering out some information that may be useless for the research, this is the main purpose of filtering the noise information. For the preprocessing rules of microblog content other than microblogging content, the noise filtering of microblog data is roughly divided into two categories. It is to filter the text content of the microblog itself and filter some other information including the microblog user outside the microblog content.

a: TEXT CONTENT FOR WEIBO

Weibo contains a lot of information similar to the format of ''#topic name#'' and ''@user,'' and I believe everyone will not be unfamiliar with it. In Chinese Weibo, ''#topic name#'' is generally a set of keywords given by the Weibo platform, and there is a considerable human factor as a topic mark. For this purpose, the field of the ''#topic name#'' format is directly removed and other text content is retained. When we want to forward a favorite Weibo, we usually use ''@+username.'' However, the content related to this type of format is generally irrelevant to the topic, which will cause adverse effects on the subsequent topic detection, so it needs to be removed.

b: FOR MICROBLOGGING

i: LINEAR THRESHOLD MODEL

The linear threshold model assumes that the node has two states at a given point in time, active and inactive. Initially, a group of nodes are in an active state, and each active node will synchronously affect the inactive nodes connected to it, making it active. In the linear threshold model, each inactive node v has an activation threshold of $F \in [0, 1]$. v is activated when the sum of the influences of all neighbor nodes of v is greater than the threshold F. All active neighbor nodes of v can participate in activation v and can generate multiple activation actions. In the microblog information dissemination, the information publisher (including the forwarded) is regarded as an active node, and the information receiver is regarded as an inactive node. It can be seen that the independent cascading model is centered on the information publisher, and the linear threshold model is based on the information receiver.

ii: FILTER MEANINGLESS MICROBLOGGING INFORMATION

By setting the threshold F, for some people in Weibo for profit or even to achieve a certain commercial purpose, using some zombie accounts or advertising accounts to publish a large number of duplicate Weibo information, it is obvious that such information is of no research value. In this case, in order to prevent a large number of zombie accounts from appearing, users who listen to less than a certain threshold 0.5 are generally ignored.

2) PARTICIPLE AND PART OF SPEECH

This paper uses the ICTCLAS (Institute of Computing Technology, Chinese Lexical Analysis System) word segmentation system of the Institute of Computing Technology of the Chinese Academy of Sciences for word segmentation and part-of-speech tagging. The system has good word segmentation and supports special word classes such as name recognition, place name recognition, name recognition, new word recognition, etc. More importantly, it also supports part-of-speech tagging, and the part of speech is used to assist word frequency statistics.

3) REMOVING STOP WORDS

Words that have no distinction or ability to describe are stop words (such as ''bar,'' ''and,'' ''have,'' etc.). By definition, these words have no practical meaning for the processing of text. These words usually appear with similar probabilities in different types of documents, they are mostly virtual words, and they also increase the dimension of the text feature vector, so these words must be removed.

B. WEIBO TEXT MODELING

1) TEXT REPRESENTATION

Computers don't understand unstructured Chinese text as humans do, so if you want to use a computer to manipulate the text, you need some way to turn it into a form that the computer can recognize. To this end, people have explored several representation models such as language model, Boolean model and vector space to deal with this problem. In this paper, we use a flexible, convenient and widely used Vector Space Model [19].

Before modeling the text, give a brief introduction to the definition of feature items and feature item weights. As an inseparable unit in VSM, feature items require not too many numbers, certain identification of text content, ability to distinguish text from other text, and so on. Only when these requirements are met can they be called a qualified ''feature item.'' The feature item weight is used to indicate the ''component size'' of a feature in the text. It's like in a class group, the squad leader may be heavier as the grassroots leader and organizer in the class. Similarly, in a document, the larger a certain feature component, that is, the greater the weight, the more the amount of information contained in the text can be represented.

The VSM modeling process is: set to the i-th document in the d_i document set D, and process each document into the form of the set $\{a_1, a_2, \cdots, a_n\}$. Among them, a_i represents the i-th feature item, and each of the a_i is given a weight which can reflect its role in the document according to a certain feature-giving rule-w_i. So in the vector space model, we can write the first document as $d_i = \{a_{i1}, w_{i1}; a_{i2}, w_{i2}; \cdots a_{im}, w_{im}\}$, as shown in Figure 8 for the vector space model of the text.

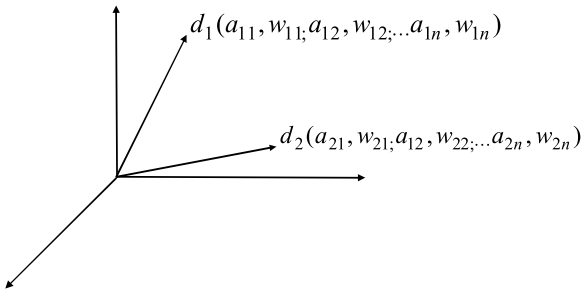


FIGURE 8. Text vector space model.

The feature words obtained after the feature selection are used to vectorize the text, and the acceleration ratios when the

TABLE 2. Text vectorization processing time and speedup.

Number of nodes	Time (s)	Speedup ratio
Click	447.9	-
1	491.7	0.911
2	286.1	1.582
3	242.2	2.03

number of nodes is different are calculated. The experimental results are shown in Table 2.

2) TEXT FEATURE EXTRACTION

In machine learning, there are characteristics and satisfaction conditions: the frequency in the counter sample is relatively high and the frequency in the positive sample is relatively low or the frequency in the counter sample is relatively low and the frequency in the positive example is relatively high. The information gain method can solve this problem very well. The formula for the information gain G(w) is shown in (1):

$$\begin{aligned} G(w) = & \sum_{i=1}^{k} p(c_i) \log p(c_i) \\ & - p(w) \sum_{i=1}^{k} p(c_i/w) \log p(c_i/w) \\ & - p(\bar{w}) \sum_{i=1}^{k} p(c_i/\bar{w}) \log p(c_i/\bar{w}) \end{aligned} \tag{1}$$

Among them, $p(w)$ is the probability that word w appears, $p(\bar{w})$ is the probability that w does not appear, $p(c_i)$ is the probability of c-type training text, k means there is k-type training text, $p(c_i/w)$ is the probability that w belongs to c_i, and $p(c_i/\bar{w})$ means w does not appear. And the probability of belonging to c_i.

Information gain refers to the amount of contribution of a feature to classification when text classification is performed. Many complex entropy theories and mathematical theories are covered in the process of information gain calculation. Therefore, information gain is estimated based on entropy. For a given text data, the information gain value of each feature item is calculated according to the information gain method, and the features are sorted in descending order according to the level of the value. From the above analysis of the information gain, the higher the information gain value of a certain feature item, the greater the dedication in the text and the more influence on the classification.

3) EATURE ITEM WEIGHTS

In this paper, the TF-IDF weight calculation method is used to calculate the feature weight of Weibo public opinion on the smart campus [20]. The main ideas of the algorithm are:

A word presents a particularly high frequency TF in a piece of text, but presents a particularly low frequency TF in other texts. TF of TF-IDF is how many times a word appears in text d. IDF is the inverse document frequency, which mainly indicates the distinguishing power of a word on a certain category, that is, if a certain feature exists in less documents. Then the higher the IDF value, the more obvious the discriminating power. Its calculation formula is as follows (2):

$$w_{ij} = tf_{ij} \times idf_{ij} \tag{2}$$

Where idf_{ij} represents the inverse document frequency of feature t_{ij} and tf_{ij} represents the frequency of feature t_{ij} in the j-th text D_j. There are several ways to calculate the inverse document frequency. The calculation formula used in this paper is shown in (3):

$$idf_{ij} = \log(\frac{N}{n_{ij}}) + 0.01 \tag{3}$$

Where N represents the total number of text collections and n_{ij} represents the number of texts containing feature t_{ij}. Therefore, the normalized TF-IDF function calculation formula is as shown in (4):

$$w_{ij} = \frac{tf_{ij} \times \log(\frac{N}{n_{ij}} + 0.1)}{\sqrt{\sum_{j=1}^{M} (tf_{ij} \times \log(\frac{N}{n_{ij}} + 0.01))^2}} \tag{4}$$

Where, tf_{ij} represents the frequency of the jth word in the i-th document, and w_{ij} represents the weight of the jth word in the i-th document.

C. TEXT SIMILARITY CALCULATION

After the text is modeled by the above operations, the computer can recognize the text. However, how to judge whether the content expressed by the text belongs to a topic, we need to calculate the similarity between the texts, and then classify the documents with higher similarity under the same topic.

Based on the vector space model established above, the most widely used text similarity calculation method, cosine similarity calculation, is selected based on the topic detection application field. It is simple and easy to implement, so this method is the best choice for similarity calculations. In the vector space model (VSM), [19], [21] the similarity between D_m and D_n is represented by $Sim(D_m, D_n)$, and the operation formula is as shown in (5):

$$Sim(D_{m,}D_{n,}) = \cos\theta = \frac{\sum_{i=1}^{N} w_{mi} \times w_{ni}}{\sqrt{(\sum_{i=1}^{N} w_{mi}^2)(\sum_{i=1}^{N} w_{ni}^2)}} \tag{5}$$

Among them, W_{mi}, W_{ni} are the weights of the i-th feature items of documents D_m and D_n, respectively, N represents the total number of two vector feature items, and 1 <= i <= N.

D. IMPROVED SINGLE-PASS CLUSTERING ALGORITHM

1) INTRODUCTION TO TEXT CLUSTERING ALGORITHM AND IMPROVEMENT OF SINGLE-PASS ALGORITHM

In order to achieve better clustering effect in the process of microblog topic detection, an improved single-Pass clustering microblog topic detection method is proposed [22]. The algorithm improves the selection of the initial clustering center and how to improve the efficiency of the algorithm, and then achieves a better clustering effect on the public opinion in the smart campus.

a: INITIALIZE THE CLUSTER CENTER

In this paper, the clustering center is randomly selected for the Single-Pass algorithm, and the disadvantages of low computational efficiency are optimized. First, the neighborhood radius eps and the minimum density threshold Min Pts are set, and the number of all documents T in the neighborhood radius eps of each document d is calculated. If $T > MinPts$, the document d is set as the initial cluster centroid.

b: IMPROVEMENT OF ALGORITHM EFFICIENCY

In this paper, when the new document is compared with the existing topic, only the new document is compared with the data with the largest topic similarity value, and the maximum similarity value is determined by the recursive algorithm with simple operation and high efficiency. If the maximum value is still smaller than the similarity threshold θ, the text does not need to be compared with the topic centroid in turn, avoiding a lot of cumbersome comparison and calculation, and the algorithm efficiency is greatly improved.

The specific implementation steps of the improved algorithm in this paper are as follows:

Input: D—microblogging document collection

Eps—radius

Min Pts—Minimize the given point as the minimum of the initial cluster center within the radius eps

Number of domain points

θ−similarity threshold

Output: target cluster collection

① For all documents D_j entered, calculate the number of all documents in each neighborhood within the neighborhood radius eps T_j. If $T_j > MinPts$, the document D_j is set as the initial cluster center, otherwise D_j is added to the unprocessed set F;

② Determine whether all the documents entered have been processed, and if so, perform step ③, otherwise continue to perform the steps ①;

③ Calculate the similarity between the document F_i in the document set F and the centroid of each topic in the initial cluster C, and use the recursive algorithm to find the maximum similarity;

④ Comparing the maximum similarity value calculated by step ③ with the similarity threshold θ. If the maximum similarity value is smaller than θ, a new topic cluster is created, otherwise D_j and F_i are grouped into one class;

⑤ Determine whether all the documents in the document set F have been processed. If not, the process returns to step ③ to continue the loop execution, otherwise the clustering result is output.

2) EXPERIMENTAL DATA AND EVALUATION INDICATORS

This paper uses the data collected by the crawled Sina Weibo platform to randomly select and use the improved Single-Pass cluster proposed in this paper to conduct related experiments on topic detection. In this experiment, the neighborhood radius eps is 0.39, the minimum density threshold Min Pts is 4, and the similarity threshold is usually between 0.6 and 0.9.

This article randomly captured 8017 Weibo data from 6 hot topics. In order to conduct the test more effectively, the data was preprocessed first, and 400 relatively good Weibos were selected in each topic, and 2400 pieces of Weibo information were shared for experimental testing.

In this paper, the performance of the algorithm is tested by the commonly used evaluation index recall rate and false detection rate. The recall rate and false detection rate of topic i (i = 1, 2, 3, . . . n) are defined as follows:

$$\text{Recall rate} = \frac{\text{Correctly detected number of microblogs related to topic i}}{\text{Total number of Weibo related to topic i}}$$

$$\text{Error rate} = \frac{\text{The number of microblogs detected that are not related to topic i}}{\text{Total number of Weibos not related to topic i}}$$

From the formulas of these two evaluation indicators, the higher the recall rate and the lower the false detection rate, the better the clustering effect of the algorithm. Therefore, the goal of this paper is to maximize the recall rate and try to ensure that the false positive rate is small enough.

3) EXPERIMENTAL RESULTS AND ANALYSIS

In this paper, the classic Single-Pass clustering algorithm is tested simultaneously with the improved algorithm to verify whether the improved algorithm has significant effects. The comparison of the topic recall rate between Single-Pass clustering and the improved Single-Pass algorithm is shown in Figure 9. It can be concluded from the figure that the recall rate of the improved Single-Pass algorithm is significantly improved.

As shown in Figure 10, the comparison of the false detection rate of the topic obtained by Single-Pass clustering and the improved Single-Pass algorithm is presented. The analysis of the graph shows that the false detection rate of the improved Single-Pass algorithm is significantly reduced.

The Single-Pass clustering is compared with the running time of the improved Single-Pass algorithm. The situation is shown in Table 3. It can be seen that the improved Single-Pass algorithm reduces the computation time by half compared to the Single-Pass algorithm.

The improved algorithm in this paper is significantly lower in time consumption than the classic Single-Pass

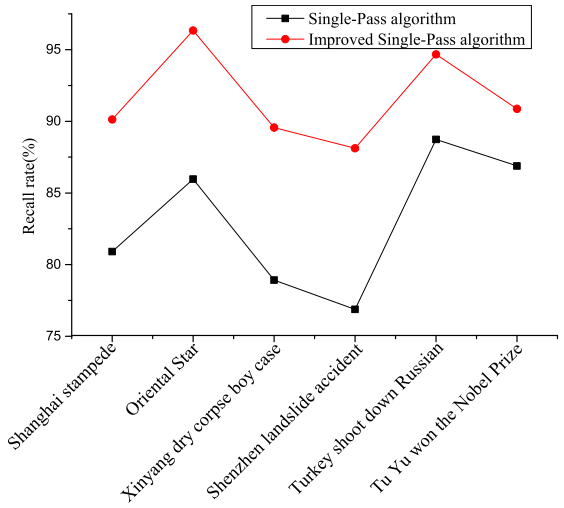


FIGURE 9. Comparison of topic recall rate between Single-Pass clustering and improved Single-Pass algorithm.

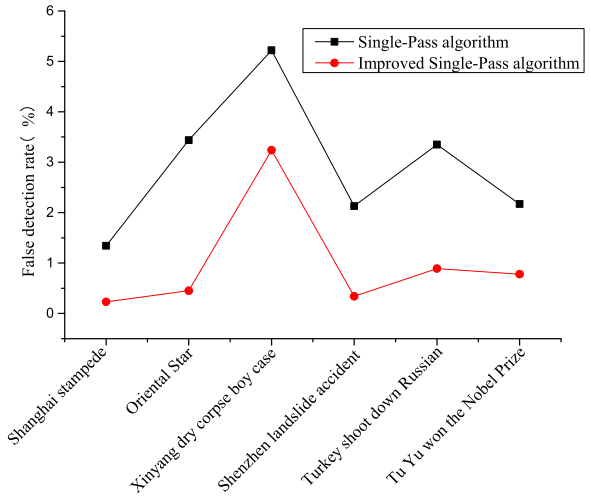


FIGURE 10. Comparison of topic false detection rate between Single-Pass clustering and improved Single-Pass algorithm.

TABLE 3. Comparison of runing time of two algorithms.

Serial number	Algorithm	The time spent
1	Single-Pass algorithm	57.6s
2	Improve Single-Pass algorithm	26.4s

clustering algorithm. When the new document in the improved algorithm is compared with the existing topic centroid, the strategy of simple operation and high efficiency is used to determine the maximum similarity. Thereby, the calculation time can be reduced, and the operation efficiency of the algorithm is greatly improved.

VII. UML MODELING ANALYSIS OF SMART CANPUS PUBLIC OPINION DETECTION PLATFORM

A. OVERVIEW OF UNIFIED MODELING OF SMART CAMPUS PUBLIC OPINION DETECTION PLATFORM

In the software requirements analysis phase, user analysis can be captured through use case analysis. The software requirements analysis phase is mainly concerned with the main concepts in the problem domain, such as objects, classes, abstractions, etc., and needs to identify the relationships between classes and classes, and describe them with class diagrams.

In this paper, in the process of software requirements analysis, the use of dynamic models to describe the need for collaboration between classes. In the software analysis phase, classes that define the technical details in the software system are not considered, and only objects of the problem domain are modeled. The programming implementation requirements function is an independent phase whose main task is to transform the classes from the design phase into actual code in an object-oriented programming language to perform the requirements analysis. In order to verify the design scheme of the smart campus public opinion detection platform and realize the intelligent processing and dynamic management of the automatic collection and research of public opinion information, the system needs to be modeled and analyzed. The public opinion information processing model abstractly describes various aspects involved in the automatic collection, classification, automatic judgment, alarm, and public opinion reports of public opinion information. Choosing the right modeling tools and modeling perspectives is key to modeling the network public opinion information management business process. Based on the actual needs of the smart campus public opinion monitoring platform, this paper selects the Unified Modeling Language as the modeling representation language, and uses the modeling tool to establish the reference model for public opinion information management, which is used for the requirements analysis, system design and system implementation of the platform construction.

B. ANALYSIS OF USE CASE OF SMART CAMPUS PUBLIC OPINION DETECTION PLATFORM

The goal of UML static modeling is to accurately describe the static characteristics of the model's internal characteristics and their interrelationships. Static modeling can be used to describe the organization and structure of the system. The Unified Modeling Language describes the static structure of a model through package diagrams, class diagrams, and structure diagrams [23]–[25].

Through the analysis of the functional requirements of the smart campus microblogging public opinion monitoring platform, it is known that system analysts and system architects understand the designed system. A very popular method of capturing system requirements is to use a use case-based approach. Use cases are primarily used to capture the high-level user functional requirements of the system and are a key technology in application development. But use cases cannot be used to capture internal functional requirements, nor can they be used to capture non-functional requirements, because use cases are an informal and inaccurate modeling technique that is the basic structure that builds our system. It is not only the unit that defines the demand, but also the unit that

estimates our workload. In the modeling process of the smart campus public opinion monitoring platform, the first step is to identify the participants, the second step is to analyze the use cases, and the third step is to create a use case diagram. The smart campus public opinion monitoring platform participants and use case diagram are shown in Figure 11.

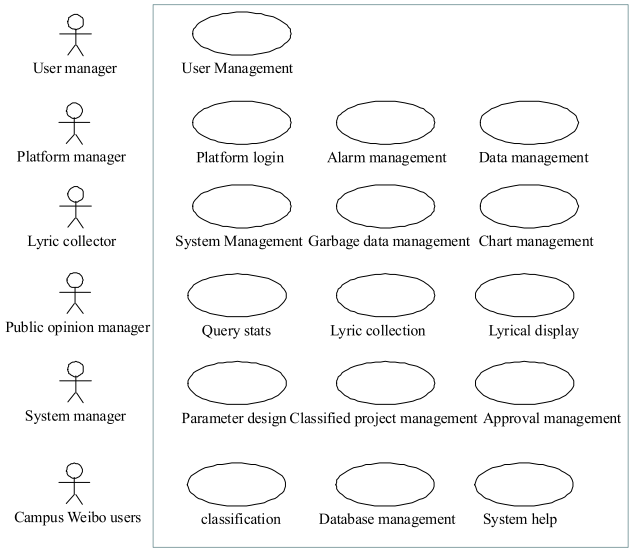


FIGURE 11. Use case diagram of public opinion monitoring platform.

The top-level use case diagram of the smart campus public opinion monitoring platform is shown in Figure 12.

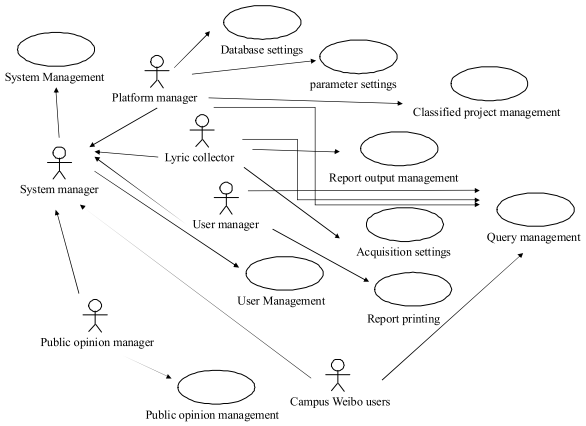


FIGURE 12. Top-level use case diagram of public opinion monitoring platform.

The connection between the model elements in Figure 10 indicates that there is an association relationship between them, which is the User Case model of the system layer of the smart campus public opinion monitoring platform. It only contains the most basic User Case model and is a high-level abstraction of the system. In the development process, as the understanding of the system continues to deepen, the model can be refined from the top down, and a more detailed User Case model is evolved as shown in Figure 13 and Figure 14.

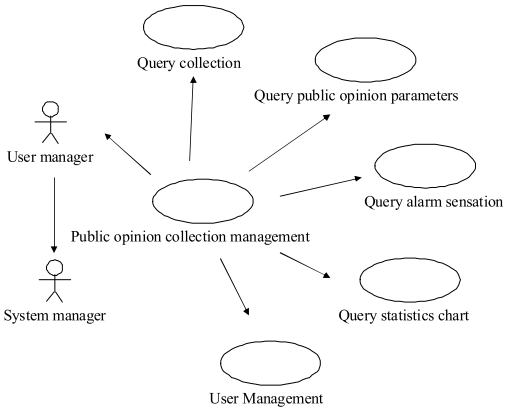


FIGURE 13. On the machine query client B/S use case diagram.

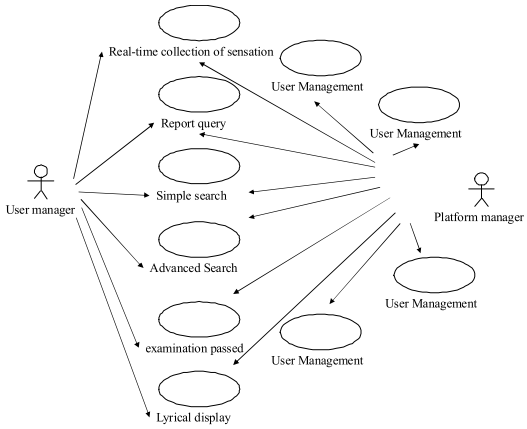


FIGURE 14. User manager query client use case diagram.

In the above use case, when describing the use case, this article uses the case of the public opinion parameter setting management case as an example to describe the use case.

Use case name: collection parameter query, modification management

Description of the use case: the query of the acquisition parameters, the modification of the acquisition parameters of the query are displayed, and the modification can be made.

Participants: system administrators, collectors

Precondition: Participant login into the system

Postcondition: If the use case is successful, the acquired acquisition parameter information is displayed.

Basic operation process:

① User login system

② The user clicks on the menu display area to set the parameters, and selects ‘‘Acquisition Parameter Settings’’ under the ‘‘Parameter Settings’’ menu.

③ If the user is a member of the collection group or the director management group, the system displays all the item list information. If it is a member of another user group, the system prompts ‘‘read-only permission, cannot modify the operation.’’

④ The user enters the project information to be modified and clicks the ‘‘Save’’ button to modify the data.

⑤ The system automatically saves the changes made by the user to the original parameter table and re-transmits the relevant data.
⑥ The user browses information such as keyword information and websites.
⑦ The user quits the system.

C. NETWORK STRUCTURE DESIGN OF SMART CAMPUS PUBLIC OPINION DETECTION PLATFORM

Through the analysis of the use cases by the previous participants, the main program modules involved in the system include key parts such as the control main program, the user manager query program and the platform manager management program. The use of the B/S structure can meet the requirements of each participant, but it is not guaranteed for the main control program in terms of stability and timeliness. The C/S result has great advantages in the main control program, especially the acquisition end, and can quickly handle the transient processing requirements such as public opinion collection and automatic classification. However, for the query of platform managers and user managers, the network burden is undoubtedly increased, which seriously affects the public opinion data collection of the main control program.

Therefore, the B/S structure based on the public network Internet worker is used to query the public opinion data to meet the needs of the user manager query and the platform manager query. The C/S structure of the "fat server" is adapted to the real-time public data collection requirements of the main control program, which not only ensures the system speed but also the superiority of the query, and solves the problem that the client causes the access bottleneck. The hybrid architecture with B/S and C/S can complement each other and make up for the shortcomings of the B/S structure. It is a widely used security solution with good reliability in LAN. The advantages of the B/S architecture are mainly reflected in the fact that the business logic and performance logic are implemented on the Web Server. The changes in the business logic only affect and do not affect the normal operation of other client browsers. This advantage just makes up for the shortcomings of the C/S architecture style, and the disadvantage of the B/S architecture is that the client browser does not communicate directly with the database. B/S needs to use Web Server as a transit station to connect with the database, and the advantages of C/S architecture style just make up for the shortcomings of B/S architecture style.

According to the above analysis, the structure of the smart campus public opinion detection platform is shown in Figure 15 below. A three-layer structure of a presentation layer, a data access layer, and a business logic layer is adopted. The presentation layer provides a user interface and other work buttons for providing human-computer interaction with the user to provide business logic and other constraints for the business processing data layer for data storage.

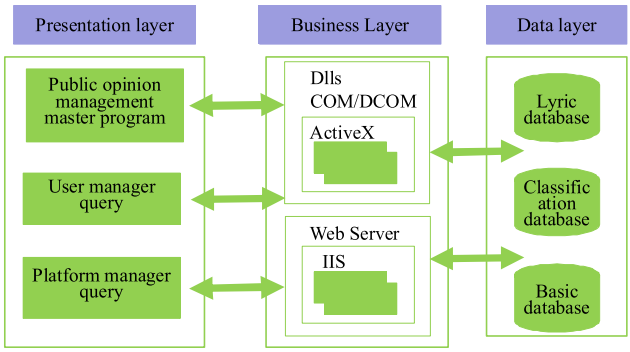


FIGURE 15. Network architecture.

VIII. RESULT ANALYSIS

This paper selects a total of 458.913 million Weibo data in several time periods to establish a test data set. Firstly, the test and analysis of 317,412 data of Sina Weibo from 4.15 days to 4.23 days were carried out. Through the tracking of hot topics and keyword filtering and filtering of sensitive words, it was found that there were several topics with high attention. The basic situation is shown in Figure 16.

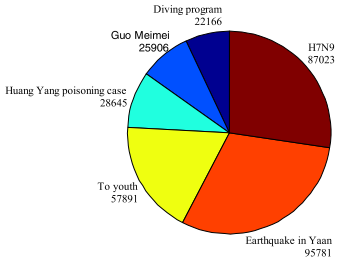


FIGURE 16. Hot topic concerns.

The total attention of Fudan University's "Huangyang Poisoning Case" event reached 28,645. On April 15th, three Weibo users published Weibo information on the matter, and quickly attracted widespread attention. On April 20th, the microblog attention of this topic reached 7903, after which the attention of this topic gradually decreased. The change law of "Yangyang poisoning case" attention degree is shown in Figure 17:

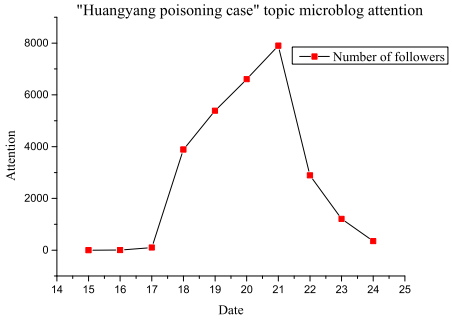


FIGURE 17. "Yellow case poisoning case" topic microblog attention.

In this paper, through the processing and analysis of the captured data of 92456 of Tencent Weibo, the improved

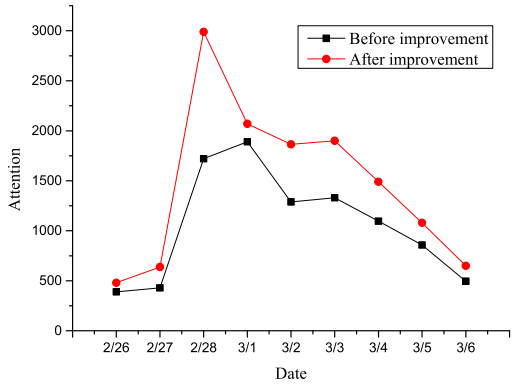


FIGURE 18. Comparison of the focus of the topic of the restaurant's microblogging hot topic.

database with Memcached as the cache layer is applied to the public opinion monitoring platform of the smart campus and the traditional database is used for comparison experiments. The comparison chart of the hot topic of the ''canteen meal'' microblogging topic is shown in Figure 18. It can be observed that the data collected by the improved data management method is larger than the data collected by the traditional method, and it proves that such data management methods still have a certain positive effect, and the data recall rate of the database is improved to some extent. and query efficiency. In addition, it is better to see the situation of the trend of Weibo attention to ''canteen meals.''

On May 27th, the microblogging public opinion about the ''campus incident'' was detected in the public opinion monitoring system. A total of 49 relevant microblog records were displayed in the Weibo public opinion monitoring system. The publishing time of its microblog is shown in Figure 19.

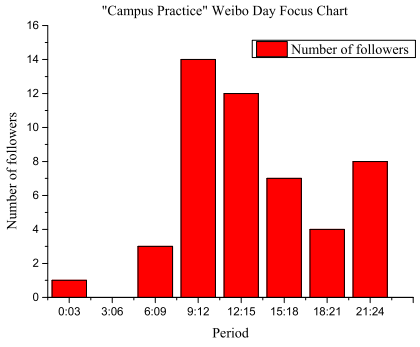


FIGURE 19. "Campus Practice" Weibo day focus trend chart.

By counting the release time of the ''Campus Events'' Weibo on May 27, it was found that the law shown in Figure 16 was presented. Through observation, the trend of attention of Weibo is basically in line with the trend of the human body clock. 9:00 to 15:00 is the most energetic and most active period of time, so the number and attention of Weibo in this period is the highest in one day. Therefore, the monitoring of Weibo's public opinion can increase the crawling and analysis of Weibo data during this period.

IX. CONCLUSION

This paper focused on the real-time platform construction of microblogging in the smart campus. Firstly, it analyzed the characteristics of microblogging communication, and the use of microblogging and information dissemination mechanism by college students, which paved the way for the platform construction. According to the microblog obtained by the crawler, the improved Single-Pass algorithm will be used to analyze the public opinion data, and the effect obtained by the Single-Pass algorithm and the improved Single-Pass algorithm will be compared. And the improved Single-Pass algorithm had the advantages of high recall rate, low false detection rate and short running time. With the four aspects of innovation wisdom,open wisdom,fusion wisdom and intelligent wisdom of the smart campus, UML is used to model and analyze the smart campus microblogging public opinion monitoring platform, and several hot topics were randomly selected for analysis. This paper discussed the innovation, openness and security of campus public opinion monitoring in platform design. The construction of this platform played an important role in the effective understanding of all aspects of students in colleges and universities.

REFERENCES

[1] T. S. Zhang and D. Wang, ''The new problem of 'Chaoyang industry'—Analysis of the dilemma faced by the media network public opinion analysis industry,'' *Business Culture*, no. 2, p. 346, 2012.

[2] R. H. Huang *et al.*, ''Smart campus: The inevitable trend of digital campus development,'' *Open Educ. Res.*, vol. 18, no. 4, pp. 12–17, 2012.

[3] P. Chen and Z. Liu, ''Research on IoT infrastructure of smart campus,'' *J. Wuhan Univ., Sci. Ed.*, vol. 58, no. S1, pp. 141–146, 2012.

[4] Y. Wang, ''The overall architecture model and typical application analysis of smart campus construction,'' *China's Electrific. Educ.*, no. 9, pp. 88–92 and 99, 2014.

[5] S. R. Ye *et al.*, ''Weibo topic detection method combining isolated point preprocessing and single-pass clustering,'' *Appl. Res. Comput.*, vol. 33, no. 8, pp. 2294–2297, 2016.

[6] Y.-W. Seo and K. Sycara, ''Text clustering for topic detection,'' Robot. Inst., Carnegie Mellon Univ., Pittsburgh, PA, USA, Tech. Rep. CMU-RI-TR-04-03, 2004.

[7] X. X. Fang and Y. Q. Lv, ''Based on the improved single-pass network lyric topic discovery research,'' *Comput. Digit. Eng.*, vol. 42, no. 7, pp. 1233–1237, 2014.

[8] K. Niederhoffer, R. Mooth, D. Wiesenfeld, and J. Gordon, ''The origin and impact of CPG new-product buzz: Emerging trends and implications,'' *J. Advertising Res.*, vol. 47, no. 4, pp. 420–426, 2007.

[9] L. Kang, H. L. Du, H. Zhang, and W. Li Ma, ''Systematic research on the application of steel slag resources under the background of big data,'' *Complexity*, vol. 2018, Oct. 2018, Art. no. 6703908.

[10] S. Amer-Yahia *et al.*, ''MAQSA: A system for social analytics on news,'' in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, 2012, pp. 653–656.

[11] Y. Qu, C. Huang, P. Zhang, and J. Zhang, ''Microblogging after a major disaster in China: A case study of the 2010 Yushu earthquake,'' in *Proc. ACM Conf. Comput. Supported Cooper. Work.*, 2011, pp. 25–34.

[12] X. Cao, K. Chen, R. Long, G. Zheng, and Y. Yu, ''News comments generation via mining microblogs,'' in *Proc. 21st Int. Conf. World Wide Web*, 2012, pp. 471–472.

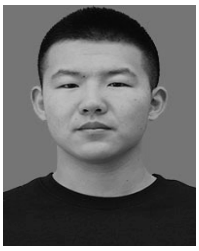
[13] A.-M. Yang, X.-L. Yang, J.-C. Chang, B. Bai, F.-B. Kong, and Q.-B. Ran, ''Research on a fusion scheme of cellular network and wireless sensor for cyber physical social systems,'' *IEEE Access*, vol. 6, pp. 18786–18794, 2018.

[14] Y. Huang and X. Li , ''Technical architecture based on network smart campus and its implementation,'' *Southeast Acad.*, no. 6, pp. 309–316, 2012.

[15] H. T. Wang, H. Chen, and X. P. Zhang, "Discussion on the construction ideas and design schemes of comprehensive university smart campus," *Data Commun.*, no. 2, pp. 1–4, 2018.
[16] Y. Han, J. Li, X.-L. Yang, W.-X. Liu, and Y.-Z. Zhang, "Dynamic prediction research of silicon content in hot metal driven by big data in blast furnace smelting process under Hadoop cloud platform," *Complexity*, vol. 2018, Oct. 2018, Art. no. 8079697.
[17] F.-C. Liu, Y.-L. Liu, D.-H. Jin, X.-Y. Jia, and T.-T Wang, "Research on workshop-based positioning technology based on Internet of Things in big data background," *Complexity*, vol. 2018, Oct. 2018, Art. no. 7875460.
[18] J. Wu, L. Zhang, S. Yin, H. Wang, G. Wang, and J. Yuan, "Differential diagnosis model of hypocellular myelodysplastic syndrome and aplastic anemia based on the medical big data platform," *Complexity*, vol. 2018, Nov. 2018, Art. no. 4824350.
[19] A. Majid, M. Khan, N. Iqbal, M. A. Jan, and M. Khan, "Application of parallel vector space model for large-scale dna sequence analysis," *J. Grid Comput.*, pp. 1–12, 2018.
[20] S. Jiang, M. Lian, C. Lu, Q. Gu, S. Ruan, and X. Xie, "Ensemble prediction algorithm of anomaly monitoring based on big data analysis platform of open-pit mine slope," *Complexity*, vol. 2018, Aug. 2018, Art. no. 1048756.
[21] H. Fu, Z. Li, Z. Liu, and Z. Wang, "Research on big data digging of hot topics about recycled water use on micro-blog based on particle swarm optimization," *Sustainability*, vol. 10, no. 7, p. 2488, 2018.
[22] S. H. Pan, K. Y. Wang, and S. F. Wang, "On-line detection method for agricultural product safety incident based on improved single-pass," *Sci. Technol. Eng.*, vol. 18, no. 13, pp. 98–103, 2018.
[23] H. Jiang and L. Wang, "Research on CORBA-based network public opinion monitoring system," *Consum. Electron.*, no. 10, pp. 182–183, 2014.
[24] A. Yang, Y. Li, F. Kong, G. Wang, and E. Chen, "Security control redundancy allocation technology and security keys based on Internet of Things," *IEEE Access*, vol. 6, pp. 50187–50196, 2018.
[25] P. P. Geng, X. Q. Ding,Z. Tao, X. W. Xu, and X. D. Wang, "Design and implementation of a universal IoT data analysis platform," *Comput. Appl. Softw.*, vol. 30, no. 11, pp. 115–118, 2013.

FENG NAN received the Ph.D. degree from the School of Civil and Resource Engineering, University of Science and Technology Beijing, China, in 2017. He is currently the Associate Dean with the College of Electrical Engineering, North China University of Science and Technology, Tangshan, Hebei. His research interests involve mining, industrial economics research, numerical calculation, big data, and mathematical modeling.

YINA SUO was born in 1980. She received the master's degree from the School of North China University of Science and Technology in 2012. She is currently a Lecturer with the North China University of Science and Technology. Her interest is the research of multimedia data compression and the development of mobile games under Android system.



XUEYONG JIA was born in Xingtai, Hebei, China, in 1998. He is currently pursuing the degree with the North China University of Science and Technology, majoring in electrical engineering and automation. His main researches are automatic control, data mining, writing, and typesetting.

YUYING WU was born in Tangshan, Hebei, China, in 1997. She is currently pursuing the degree in energy and power engineering with the North China University of Technology. She has studied several programming languages and is interested in big data and machine learning algorithms.

SHUAIJIE SHAN was born in Hebi, Henan, China, in 1997. He received the degree in measurement and control technology and instrumentation from the College of Electrical, North China University of Science and Technology. His research interests include data mining, educational technology and automatic technology.