Design of Public Opinion Research and Judgment System for Network Hot Events Based on Data Mining Technology

Qing-mei Cao[1(✉)] and Hui-fang Guo[2]

[1] Department of Computer Technology and Information Management, Vocational and Technical College of Inner Mongolia Agricultural University, Baotou 014100, China
[2] School of Transportation and Municipal Engineering, Inner Mongolia Vocational and Technical College of Architecture, Hohhot 010010, China

Abstract. The traditional network hot event public opinion research and judgment system in the face of excessive operating pressure, there are a large number of blank database connections, leading to the quality of the system running worse and worse. In order to solve this problem, this paper proposes the design of network hot event public opinion research and judgment system based on data mining technology. In hardware design, the serial communication module is designed. With the support of embedded processor, it is combined with the designed Ethernet transceiver to realize real-time communication; In the software design, it collects public opinion data, uses data mining technology to find network hot events, calculates the public opinion level of hot events, and designs corresponding early warning function to realize the judgment of public opinion. So far, the overall design of the system is completed. The experimental results show that: in the case of increasing operating pressure, the designed public opinion research and judgment system based on data mining has high operating efficiency, good public opinion recognition ability, and high operating quality.

Keywords: Data mining · Network hot events · Public opinion collection · Public opinion research and judgment

1 Introduction

Born in the 1960s, the Internet has been applied far beyond its original intention in today's society. With the rapid development of the Internet, it has become an important part of people's life, work and study and an important carrier of the spread and development of human civilization [1, 2]. As the main carrier of network communication, the Internet has penetrated into economy, politics, culture, social life and other aspects, changing people's way of communication and thinking [3].

S. Liu and X. Ma (Eds.): ADHIP 2021, LNICST 416, pp. 656–671, 2022.
https://doi.org/10.1007/978-3-030-94551-0_51

Due to the characteristics of Instant Internet communication, uncontrollability of communication content, relative equality of discourse power and long-term retention of information, Internet communication is completely different from traditional communication mode. First of all, in terms of communication speed, the traditional communication mode has to go through the audit of each editorial level before releasing an information, but it is different on the Internet, It saves the cumbersome audit of traditional communication mode and greatly improves the speed of communication; Secondly, in terms of communicators, the communicators of traditional communication mode are only a few traditional media, and there are a large number of blogs, various types of websites, forums, various chat rooms and communication tools on the Internet, which can be used as communicators; Secondly, in the aspect of equal right of discourse, in the traditional media environment, only a few media and individual groups have the right to speak, and the information spread is spread after screening. In the Internet environment, everyone can say that all kinds of information can appear in front of the Internet people at the same time; Finally, in terms of information retention, traditional media, such as radio and television, disappear after a while, and few people go to read the previous materials for newspapers and magazines. On the Internet, information will be retained for a long time. Even if the problem is solved, negative information will be left on the Internet [4–7]. Therefore, in the Internet era, Internet communication can have a huge impact on economy, politics, culture, social life and other aspects [8].

With the continuous development of the Internet, public opinion plays a more and more important role in government administrative departments or industries. At present, there are certain developments in the theory and software products of network public opinion at home and abroad.

There are two main parts involved in network public opinion, one is web information mining, the other is data collection, that is, web crawler [9, 10]. Web information mining, network public opinion is the mapping of social public opinion in the Internet space, is a direct reflection of social public opinion, the main data of network public opinion comes from the web. The other part is web crawler, which is more based on text content and can realize the transparency of public opinion data [11]. But in the face of the above two forms, the traditional public opinion research and judgment system is mainly for the collection and analysis of various unstructured data, such as text data, audio and video data, graphics, image data and other data fusion multimedia data. In the practical application, there are some disadvantages, such as a large number of blank requests and blank database connections in the application process, which lead to limited utilization of resources and poor quality of system operation.

Therefore, this paper proposes the design of network hot event public opinion research and judgment system based on data mining technology to solve the problems existing in the traditional public opinion research and judgment system. From the aspects of serial communication module, Ethernet transceiver, public opinion acquisition module, hot event discovery function and public opinion information early warning, the public opinion research and judgment system of network hot events is optimized and designed by using data mining technology, and the effectiveness of the design system is verified by experiments.

2 Hardware Design of Network Hot Event Public Opinion Research System Based on Data Mining Technology

2.1 Design of the Serial Communication Module

In the system design, serial communication refers to RS-232 communication, and serial communication can be divided into synchronous communication and asynchronous communication. In the process of data transmission of asynchronous communication system, receiver clock and transmitter clock are not synchronized, and are transmitted in independent byte mode [12]. Now most microprocessors provide UART interface. The program can receive and receive data only by operating the relevant registers. Full duplex communication is generally supported. After the level conversion, RS-232 serial communication with PC can be carried out, or chip level communication can be directly conducted with other processors. Hardware driver module is always closely related to hardware, and different hardware will have different influence on program processing [13, 14]. In the serial communication module of the design system, the USART hardware interface of ATmega128 is used.

ATmega128 has 2 USART serial interfaces. It is an enhanced UART interface. Besides the traditional UART interface, it also has some more reliable and flexible features, such as supporting high-speed asynchronous communication, independent high-precision baud rate generator, etc.; In the stop bit detection, AVR is optimized, which makes the detection error smaller.

For data receiving and receiving, interrupt driven mode is adopted to receive or send data in interrupt service program, and the upper program will be further processed. For data receiving, in interrupt mode, a byte received will produce an interrupt. Reads bytes from the data register in the interrupt service program and clears the receive interrupt flag. At this time, the received bytes can be processed directly in ISR or put them in the buffer for background tasks. In the project development, if the data package is relatively simple, this kind of processing is very simple [15]; If the packet is complex, the ISR is quite complex. When the background task is processed, the software level is particularly clear, and the interrupt is only responsible for receiving, while the background task is responsible for parsing the data content. However, the size of its buffer should be

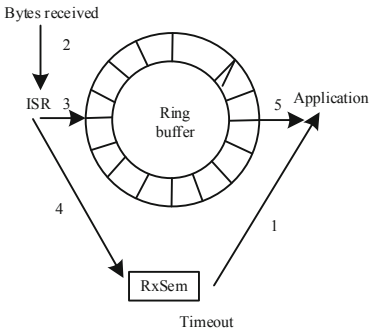


Fig. 1. Serial buffer receiving process with Signal Volume

determined reasonably to meet the worst-case scheduling situation, otherwise data will be lost. The response to input data depends on the execution speed of background tasks. If the frequency of execution is fast, the response time is fast. For real-time kernel like IBC/OS-II, the speed of processing input data is almost as fast as that of ISR only receiving data without processing [16]. Therefore, in the management of the ring buffer, a semaphore rxsem is added, as shown in Fig. 1.

The application task waits for the arrival of the semaphore (1). After receiving a byte, an interrupt is generated. The ISR reads the byte (2) from the data register and stores it in the ring buffer (3). Then the ISR sends a semaphore to tell the task that it has received a byte (4). After the application gets the semaphore, it fetches the byte from the ring buffer and performs the required task (5).

The whole process of data receiving is as follows: the application program calls commgetchar() or other functions to read data from the buffer. Generally, the timeout value should be specified to prevent the application task from being suspended permanently. When a semaphore is received, the required processing is performed. After receiving a byte, an interrupt is generated to enter the interrupt service program. Call the middle layer function commputrxchar() to put the data into the ring buffer. If the receive buffer is full, the data will be discarded [17, 18]. Therefore, when setting the buffer size, the processing speed of the task and the maximum amount of data actually received should be considered to prevent data loss. When the byte is put into the buffer, commputrxchar () decides when to send a signal to the received semaphore according to the actual needs. Is it sent every time, or when a specific character is encountered.

Compared with data receiving, data sending is simpler. Because the receiving interrupt will happen at any time, it is completely asynchronous, sometimes it will involve complex packets, and the receiver will have a special task to process the data. When sending data, generally the data to be sent can be packaged and sent directly, and the background still adopts buffer and interrupt drive mode.

2.2 Ethernet Transceiver Design

In order to ensure the reliable and timely communication of the public opinion research and judgment system, the embedded processor is used as the core processor of the system to support the normal communication of the system. STM32F407 is a 32-bit RISC chip based on ARM Cortex M4F. Its main frequency is 168 mhzo STM32F407. It has two APB enhanced I/O and peripheral buses, three AHB buses and a 32-bit AHB bus matrix, as shown in Fig. 2.

STM32F407 is a high-performance ARM chip. Its main features are as high as 1 Megabyte flash memory, flexible static memory controller, high-performance AD converter, general DMA controller and 17 timers. With the support of the processor, the Ethernet transceiver is designed.

Each node in the system communicates through Ethernet. Because the embedded processor STM32F407 has Ethernet controller, it only needs one Ethernet PHY chip. The nodes in the Internet of things use LAN8720 as the Ethernet PHY layer chip [19].

LAN8720 is an IEEE 802.3–2005 standard transceiver. Embedded processor STM32F407 communicates with it through standard RMII interface. LAN8720 supports auto negotiation mode, which can be used to determine the optimal network speed

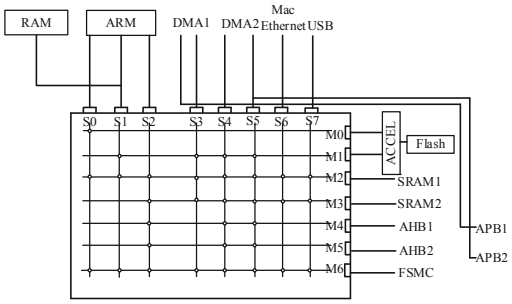


Fig. 2. STM32F407 Bus architecture diagram

and single and duplex communication modes. LAN8720 supports the operation of IEEE 802.3–2005 specific registers. This kind of operation does not need to access registers. The initial configuration can be selected through the configuration pin, and the function of transceiver can be further defined by using the register configuration option [20, 21]. All digital interface pins of LAN8720 can withstand 3.6 V. The LAN8720 uses an integrated 3.3 V to 1.2 V linear regulator, which can be selectively disabled.

LAN8720 is connected with the main control chip through RMII interface. RMII interface (reduced MII interface) is a simplified MII interface. It has two modes of MAC and PHY, and the nodes of Internet of things use PHY mode. The definition of RMII interface of LAN8720 chip in PHY mode is shown in Table 1.

Table 1. RMII interface definition in PHY mode

Signal name	Number	Direction	Describe
CLK_REF	1	Input	Reference clock
TXD[0:1]	2	Input	Transmit data
TX_EN	1	Input	Send enable
RXD[0:1]	2	Output	Receive data
RX_ER	1	Output	Acceptance error
CSR_DV	1	Output	Conflict detection
MDC	1	Input	Management clock
MDIO	1	I/O	management data

By combining the designed Ethernet transceiver with the serial communication module [22, 23], the internal and external communication of the public opinion research and judgment system can be realized, and the real-time public opinion data can be obtained. With the support of this technology, the software part of the system is designed.

3 Software Design of Network Hot Event Public Opinion Research System Based on Data Mining Technology

3.1 Design of the Public Opinion Collection Module

Public opinion collection is the crawling and storage of all media data and information on the Internet, including text content and other content. Media forms include forum, news, microblog, etc. Public opinion information must be timely, accurate and complete, and the information and attributes of the sender should be classified, counted and analyzed, which requires that the design of public opinion collection module should consider the characteristics of comprehensiveness, accuracy and stability. Therefore, the design of public opinion system should consider the storage space and database efficiency.

Data comprehensiveness includes that the content of data collection must include all URLs and keywords of all preset website pages, and cover large websites with large traffic and rich content, and the website content and keywords can be expanded. The timeliness of public opinion collection is to ensure the final update status of public opinion information. Considering the storage space and data validity of public opinion, the system only collects and saves information within one month. Data stability is to consider the collection efficiency and stable operation of the collection system, without process crash, downtime and other special circumstances. The architecture of data acquisition module is shown in Fig. 3.

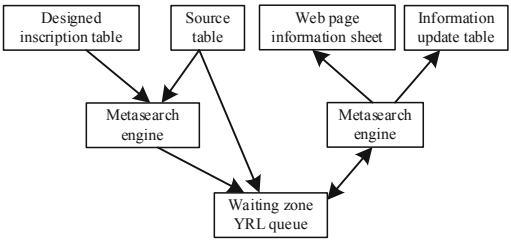


Fig. 3. Architecture of the public opinion data collection module

Public opinion data collection adopts three-level thread structure: the rule making and operation of search engine and web crawler are classified as one level thread, including the definition of website, crawler rules and keywords. The search engine collects data by keyword judgment, filters and discards the data, which is a secondary thread. Its main work is to complete the data collection through the current mainstream search engine. The data collected by the search engine is analyzed and stored locally as a three-level thread. Its main work is to store and write the processing results to the database. After collecting public opinion data, the hot event discovery function is designed to further analyze public opinion.

3.2 Design of Hotspot Event Function

Hot spot discovery is to discover the influential public opinion events in a certain period of time by automatically grabbing the information and identifying the topic of the target

website set by users. At present, the commonly used target websites are mainstream news websites, large-scale commercial websites and popular BBS.

The traditional network public opinion analysis system of public opinion discovery basically adopts two methods: first, manually input keywords to search, the second method is based on the word frequency statistics of web page. The first method has low statistical efficiency, manual input of subjective interference factors, it is not easy to grasp the public opinion information in time: the statistical accuracy based on the word frequency of the web page is low, and the theme of the web page is not clear, which will lead to the distortion of the relevant statistical data. The design of the software part of the system is based on the original public opinion discovery method, and designs a new hot public opinion discovery function.

Through the web crawler module to capture the required information of news web pages, multiple network data sources are continuously monitored, and then the web page parser starts to clean up the web page. The main function is to remove the "noise" in the web page, and retain the web page links, title, time, first-class title, and related document fields, so as to complete the data cleaning. Through the word segmentation module, the feature words of the standardized document are segmented. Randomly select a part of web pages, use them as test samples for feature extraction to obtain a group of feature words, then the feature extraction module evaluates the feature words of the sample web pages, and extracts the common feature vectors of the test samples. The TFIDF value of the feature vector is used to express the text topic, and the VSM construction module establishes the vector space model. The hot spot discovery module clusters the text relation matrix to find new hot spots.

The similarity between the report and the topic is calculated, and the maximum similarity and the topic with the maximum similarity are recorded to determine the topic closest to the current report; The theme is expressed by several feature words with the highest comprehensive weight in all the news within the theme itself; The similarity between news report and topic is calculated by Formula 1. If the calculated value is greater than the set closed value, it is considered that the new topic is included in the calculated page. The calculation formula is as follows:

$$score(x) = 1 - max\left\{1 - \frac{k}{m}\right\} \times sim(\vec{x}, \vec{c}_i) \tag{1}$$

$$sim(x, c) = \frac{\sum_{j=1}^{M} a_{jx} \times a_{jc}}{\sqrt{\left(\sum_{j=1}^{M} a_{jx}^2\right) \times \left(\sum_{j=1}^{M} a_{jc}^2\right)}} \tag{2}$$

$$a(r, \vec{w}) = \frac{rf(r, \vec{w}) \times \log\left(N/n_i + 0.01\right)}{\sqrt{\sum_{r \in \vec{w}} \left[rf(r, \vec{w}) \times \log\left(N/n_i + 0.01\right)\right]^2}} \tag{3}$$

Where $a(r, \vec{w})$ represents the weight of word r in text w, $rf(r, \vec{w})$ refers to the word frequency of word r in text w, N represents the total number of training text, n_i represents the number of r text in the training text set, $sim(x, c)$ represents the similarity of the new file x for a certain time cluster set, a_{jx} represents the weight of word j in cluster c, M

table shows the total number of words in the file set, and x represents the information of new file, c_i is the i cluster in the time interval, i is the number of files contained in the time interval, and k represents the number of files increased between the latest file collection time in cluster c_i and the arrival time of the new file x. In the setting threshold, as long as score is greater than the set value, the new file is considered as a new topic.

Cycle the above process. After a class processes a fixed number of new topics, it compares the topics in the class. If the similarity of the two topics is greater than the merge closed value, it merges them. Two clustering similarity methods can be used to calculate the similarity between topics

$$Sim(P_1, P_2) = \frac{\sum_{b_i \in P_1} \sum_{b_j \in P_2} sim(b_i, b_j)}{|P_1| \cdot |P_2|} \tag{5}$$

Where P_1 and P_2 are the monitored news topics, b_i and b_j are the news reports in P_1 and P_2 respectively, and $|P_1|$ and $|P_2|$ are the number of news reports in the two topics respectively. When a class processes a fixed number of new topics determined by users, the text information in each topic is eliminated, and the similarity between news reports and the topic is recalculated; When the similarity is lower than the closed value of clustering, or does not meet the restrictions (such as whether the report is within 30 days or not), the news report will be eliminated; Then, the internal representation and its weight are recalculated; Finally, the monitoring results are output according to the needs of users. For all the current topics in the category, combined with the time characteristics of the topic and the quantity characteristics of the text information in the topic, several topics with the highest score are selected from all the categories as the hottest information of the category.

3.3 Warning Design of Public Opinion Information

The degree of netizens' attention to a specific network public opinion event is the popularity of network public opinion, which can be used to measure the development stage of a network public opinion event, and also can be used to investigate the attitude and behavior of the participants in the network public opinion event. The popularity of Internet public opinion depends on the appeal of public opinion to post and follow a news or event. In the system design, the average number of follow-up posts in a certain period of time is used to express the appeal and the degree of attention of the theme post.

$$NH = \left[\frac{g_i}{n_i} + \sum_{j \in u_i} \frac{\sum\limits_{k \in v_j} \frac{g_k}{n_k}}{m_j} \right] \tag{6}$$

Where g_i is the number of post nodes, n_i is the number of posts of post node i, u_i is the reply set of post node i, v_i is the reply set of post node j, m_j is the number of user posts of post node j, T is the delay, and N is the number of replies of current post node. The larger the value calculated by the formula, the higher the degree of attention of the post corresponding to the news or event. It can not directly determine the severity of

public opinion, but it can be used as a reference for the classification and early warning of public opinion events.

According to the severity and level of network public opinion (very serious, serious, general and slight), systematic early warning can be carried out, corresponding to red warning, orange warning, yellow warning and green warning respectively. In addition to the popularity of Internet public opinion, the severity of the classification is also related to the spread speed, scope and political tendency of public opinion. Generally speaking, public opinion against the party, the state and ethnic conflicts can be classified as very serious public opinion. For the early warning parameter setting of public opinion, the system needs to carry out the classification and weight score calculation of comment posts, and the calculation method is shown in formula 7.

$$NW = \frac{2n_1 + n_2 + 0.5n_3}{N} \tag{7}$$

Where, N represents the number of Posts related to public opinion, n_1, n_2 and n_3 represent the number of Posts related to level 1, level 2 and level 3 public opinion. Finally, the value calculated by formula 7 is used for early warning of public opinion. The score between 1–1.5 indicates serious public opinion, the score between 0.5–1 indicates general public opinion, and the score below 0.5 indicates slight public opinion. So far, the design of network hot event public opinion research and judgment system based on data mining technology is completed.

4 Experimental Research on Public Opinion Research System of Network Hot Events Based on Data Mining Technology

4.1 Experimental Data Preparation

The main carriers of online public opinion are news websites, blogs, forums, post bars, etc. (Baidu Post Bar, zhonghua.com, Sohu Community, Sina forum, Tianya community, people.com, xinhua.com, cnews.com, Chinanet, sina.com, sohu.com, Netease and Tencent), Several representative news websites and forums are selected as shown in Table 2. The selected websites are very representative, with top click through rate and rich in network public opinion.

Table 2. Network public opinion data collection sources

Serial number	Website name	Home page URL address
1	Baidu Post Bar	http://tieba.baidu.com/
2	China.com	http://club.china.com/
3	Sohu Community	http://club.sohu.com/
4	Sina Forum	http://bbs.sina.com.cn/
5	Tianya community	http://www.tianya.cn/

(*continued*)

Table 2. (*continued*)

Serial number	Website name	Home page URL address
6	People's daily	http://www.people.com.cn/
7	Xinhuanet	http://www.xinhuanet.cn/
8	China News Network	http://www.chinanews.com.cn/
9	China.com	http://www.china.com.cn/
10	Sina	http://www.sina.com.cn/
11	Sohu	http://www.sohu.com/
12	Netease	http://www.163.com/
13	Tencent	http://www.qq.com
……	……	……

Supported by the data sources in Table 2, several groups of comparative experiments are designed to verify the actual performance of the proposed public opinion research and judgment system compared with the public opinion research and judgment system based on information awareness in reference [2] and the public opinion research and judgment system based on SNA in reference [7]. In order to ensure the scientific and fair experiment.

The comparative experiment is designed according to the application requirements of the system, and the experimental results are analyzed.

4.2 Application Performance Performance Analysis

In the experiment, two traditional public opinion research and judgment systems are introduced. Under the same experimental environment, the application performance of the system under pressurized conditions is tested. The concurrent users for each system are set to 300 and the number of experiments is 50. The experimental results are shown in Table 3.

Table 3. Experimental results of performance performance of different public opinion system

Public opinion research and judgment system	Project	Unit	Result value
Public opinion research and judgment system of reference [2]	Total duration	Hours: Minutes: seconds	1:19:37
	Maximum number of running Vuser		180
	Total throughput	Byte	156872253
	Average throughput	Byte	37547
	Total hits	Second	16643
	Average hits per second	Second	3983

(*continued*)

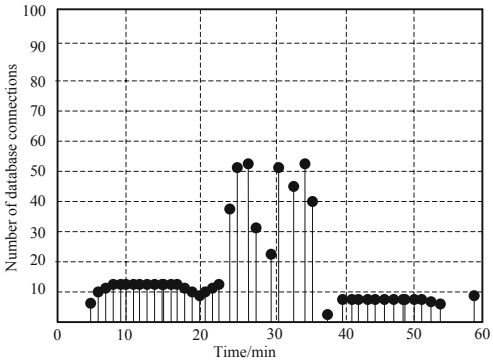
Table 3. (*continued*)

Public opinion research and judgment system	Project	Unit	Result value
Public opinion research and judgment system of reference [7]	Total duration	Hours: Minutes: seconds	00:56:49
	Maximum number of running Vuser		180
	Total throughput	Byte	128379926
	Average throughput	Byte/s	28136
	Total hits	Second	15348
	Average hits per second	Second	3048
Public opinion research and judgment system of this paper	Total duration	Hours: Minutes: seconds	00:36:15
	Maximum number of running Vuser		240
	Total throughput	Byte	276584359
	Average throughput	Byte	92491
	Total hits	Second	18231
	Average hits per second	Second	4134

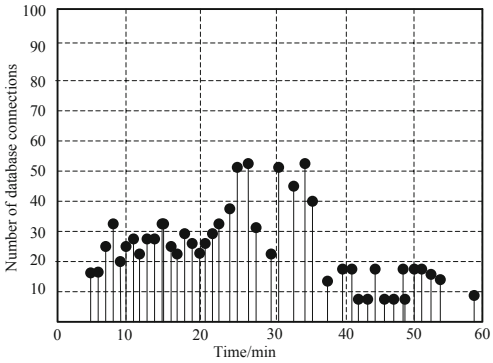
According to the data in Table 3, the experimental data of the public opinion research and judgment system in reference [2] shows that with the increasing pressure, the number of users who send out requests has reached 180, the overall response time of the system has decreased, the average throughput is low, the system can not respond to customer requests in time, nor can it meet the needs of users, and the overall running time is long. The experimental data of the public opinion research and judgment system in reference [7] shows that the running time is reduced by a part compared with the previous experimental result, and the total number of clicks is similar to the previous result, but the overall response time of the system is also reduced, and the average throughput is lower, which can not meet the user's request; Compared with the previous two groups of experimental results, the designed public opinion research and judgment system has a short duration, an average throughput of 92491 bytes, which is much higher than the previous two groups of results, and the number of users processed reaches 240, indicating that the system can maintain a better response level and application performance in the case of increased pressure, and can meet various user requests.

4.3 Experimental Analysis of System Middleware Resources Use

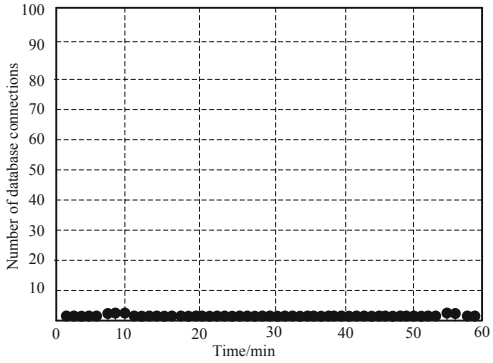
When the concurrent users of the system reach a high level, the database connection pool and thread pool will run out of resources, which indicates that the resource allocation is unreasonable and idle connections exist when the system uses middleware resources. To solve this problem, the middleware resource usage experiment is designed. The number of concurrent users is set to 100, and the idle connection of the system database is observed. The specific experimental results are shown in Fig. 4.



(a) Experimental results of public opinion research and judgment system in reference [2]



(b) Experimental results of public opinion research and judgment system in reference [7]



(c) Experimental results of public opinion research and judgment system in this paper

Fig. 4. Experimental results of middleware resources of different public opinion research system

According to the experimental results of each public opinion research and judgment system in Fig. 4, when the system is faced with the concurrent access of 100 users, the number of idle connections of the two traditional public opinion research and judgment systems is high and low, up to 50, which indicates that there are a lot of idle connections in the process of system operation, resulting in excessive consumption of resources and reducing the operation efficiency of the system; In contrast, in the effective experimental time, the number of idle connections in the database of the designed public opinion research and judgment system is always relatively low, approaching 0, which indicates that the system has high efficiency, and there is no excessive consumption of resources. The main reason is that the design system uses data mining technology to compare the two topics in a class after a fixed number of new topics are processed. If the similarity of the two topics is greater than the combined closed value, they will be combined to improve the clustering similarity and further optimize the resource utilization. To sum up, the design of public opinion research and judgment system based on data mining middleware resource use effect is good, high efficiency.

4.4 Experimental Analysis of Public Opinion Identification

The public opinion recognition experiment mainly evaluates the public opinion recognition module in the system, and evaluates the performance of the system by using the missing rate, error rate and normalized recognition cost. The calculation formula of missing rate and error rate of public opinion is defined as follows:

$$miss_i = \frac{p_i}{S_i} \tag{8}$$

$$F_i = \frac{e_i}{E_i} \tag{9}$$

Based on the above formula, the average miss rate, average error rate and normalized identification cost of the system are calculated.

$$R_{miss_i} = \sum\nolimits_i \frac{miss_i}{N} \tag{10}$$

$$R_{F_i} = \sum\nolimits_i \frac{F_i}{N} \tag{11}$$

$$C_{norm} = \frac{C_{miss} \cdot R_{miss} \cdot R_t + C_{miss} \cdot R_F \cdot R_{-t}}{N} \tag{12}$$

Where $miss_i$ is the missing rate, p_i is the number of unrecognized texts related to public opinion i, S_i is the total number of texts related to public opinion i, F_i is the error rate, e_i is the number of recognized texts not related to public opinion i, E_i is the total number of texts not related to public opinion i, R_{miss} is the average missing rate, R_F is the average error rate, C_{norm} is the normalized recognition cost, C_{miss} is the missing cost, C_{miss} is the cost of the error, R_t is the prior probability of the current event public opinion, $R_{-t} = 1 - R_t$. After the above calculation, the evaluation results of different

Table 4. Public opinion identification experimental results of different research and evaluation systems

	Number of network hot events	Missing rate	Error rate	Normalized cost
Public opinion research and judgment system of reference [2]	20	0.351	0.05632	0.0711
	50	0.409	0.07936	0.0793
	80	0.482	0.08413	0.0886
	100	0.594	0.11237	0.0922
Public opinion research and judgment system of reference [7]	20	0.692	0.09547	0.0118
	50	0.749	0.09846	0.0512
	80	0.811	0.10114	0.0831
	100	0.829	0.10236	0.1057
Public opinion research and judgment system of this paper	20	0.062	0.00006	0.0007
	50	0.084	0.00011	0.0015
	80	0.092	0.00015	0.0019
	100	0.109	0.00018	0.0024

public opinion research and judgment systems are obtained, and the specific contents are shown in Table 4.

According to the data in Table 4, when the two traditional public opinion research and judgment systems face different number of hot events, the missed check rate and error rate are increasing, and the normalization cost is increased. Compared with the previous two experimental results, the designed public opinion research and judgment system error rate and missed check rate based on data mining technology are also increasing, but the increase is not obvious, The values are always low and the normalization overhead is low. According to the analysis of the above two experimental results, the design of the network hot event public opinion research and judgment system based on data mining technology has high efficiency, accurate public opinion recognition and good overall application performance, which is superior to the traditional two public opinion research and judgment systems. The reason for this result is that the public opinion research and judgment system of network hot events based on data mining technology optimizes the public opinion collection module, crawls and stores all media data and information on the Internet, including text content and other content, and optimizes the performance of public opinion recognition.

5 Conclusion and Prospect

Due to the increasing impact of the Internet on people's lives, the increasing number of Internet users, and the more complex and changeable network environment, the

establishment of an effective network public opinion monitoring system is of great significance for coping with network emergencies, maintaining social stability, and helping the government make democratic and scientific decisions.

Based on the characteristics of rapid formation, rapid change, large amount of information and great social impact of network public opinion, as well as the requirements of current government work for network public opinion monitoring, combined with the research results of data mining, this paper proposes to establish a public opinion research and judgment system based on Data Mining, and studies the key technologies in it, It provides a method for further semantic analysis of network public opinion information for public opinion control. After the experimental demonstration, the practical application ability of the system proposed in this paper has been proved, which shows that the design of the system effectively solves the problems existing in the traditional system, and realizes the more effective control of network public opinion.

However, in the research process, there are still some shortcomings, such as the supervision of network hot events, including text form, picture form, video form, etc. in this paper, there is no one by one verification of multiple forms. Therefore, in the follow-up development, research and analysis will be carried out from this aspect to further improve the public opinion research and judgment system.

Fund Projects. Science and Technology Plan Project of Inner Mongolia Autonomous Region Project Description: Research and Development and Application of Key Technologies of Smart Agriculture in Inner Mongolia Yellow River Region (2020GG0033).

References

1. Wei, D., Wei, X.: Design of network experimental teaching evaluation system based on data mining. Modern Electron. Tech. **43**(03), 142–145+149 (2020)
2. Zhang, S., Wang, L., Lou, G.: Research on network public opinion study and early-warning system based on information awareness. Inf. Stud. Theory Appl. **43**(12), 149–155 (2020)
3. Xu, D.: Research on the judgment system of network public opinion on the emergency initiated by the major epidemic based on spatio-temporal big data. Modern Inf. **40**(04), 23–30+81 (2020)
4. Liu, J., Shi, M., Liu, C.: Semantic graph of net citizens' information perception on network group emergencies. Inf. Stud. Theory Appl. **42**(02), 158–163 (2019)
5. Shi, J., Guo, J.: Public Opinions on "The US-China Trade War": features and strategies based on the study of Youtube platform. J. Intell. **38**(08), 105–112+135 (2019)
6. Liu, J., Li, L.: Analysis of intelligent early warning mechanism of network public opinion in the background of big data. J. Intell. **38**(12), 92–97+183 (2019)
7. Cheng, J., Zhang, S., Ji, Q.: Analysis of micro-blog public opinion periodic propagation network in social hotspot events based on SNA—an empirical study on "Yu Huan". Manage. Rev. **31**(03), 295–304 (2019)
8. Gao, P., Li, J., Liu, S.: An introduction to key technology in artificial intelligence and big data driven e-Learning and e-Education. Mob. Networks Appl., 1–4 (2021)
9. Chen, Z., Wang, J.: Analysis of network public opinion events based on bayesian network. Inf. Sci. **38**(04), 51–56+69 (2020)
10. Zhang, C., Ma, X., Zhou, Y., et al.: Analysis of public opinion evolution in COVID-19 pandemic from a perspective of sentiment variation. J. Geo-Inf. Sci. **23**(02), 341–350 (2021)

11. Liu, S., Liu, D., Srivastava, G., et al.: Overview and methods of correlation filter algorithms in object tracking. Complex Intell. Syst. **3**, 1–23 (2020)
12. Li, Z., Huang, S.: Analysing on network public opinion based on LDA topic model. J. Syst. Sci. Math. Sci. **40**(03), 434–447 (2020)
13. Liu, S., Liu, X., Wang, S., Muhammad, K.: Fuzzy-aided solution for out-of-view challenge in visual tracking under IoT assisted complex environment. Neural Comput. Appl. **33**(4), 1055–1065 (2021)
14. Lei, L., Wu, Y., Zhang, Y., et al.: Time+user dual attention based sentiment prediction for multiple social network texts with time series. IEEE Access **PP**(99), 1 (2019)
15. Peng, L.J., Shao, X.G., Huang, W.M.: Research on the early-warning model of network public opinion of major emergencies. IEEE Access **PP**(99), 1 (2021)
16. Gu, Q., Li, L.: Research on process optimization and unified system content in big data network information sharing environment. J. Phys. Conf. Ser. **1952**(3), 032039 (9pp) (2021)
17. Li, K., Jing, M., Tao, X., et al.: Research on online management system of network ideological and political education of college students. Int. J. Electr. Eng. Educ. (2021), 002072092098370
18. Gao, J., Niu, S., Li, H., et al.: The reliability research of V2V based on sample judgment and balanced modulation. Int. J. Internet Protoc. Technol. **12**(4), 229–235 (2019)
19. Xie, X., Ma, Z., Ye, J., et al.: Research and development of sound quality in portable testing and evaluation system based on self-adaptive neural network. Appl. Acoustics **154**, 138–147 (2019)
20. Yue, H.A., Lw, A., Tz, A.: Area definition and public opinion research of natural disaster based on micro-blog data - ScienceDirect. Procedia Comput. Sci. **162**, 614–622 (2019)
21. Durso, G., Haws, K.L., Way, B.M.: Drug influences on consumer judgments: emerging insights and research opportunities from the intersection of pharmacology and psychology. Mark. Lett. **31**, 19–23 (2020)
22. Graf, H., Menter, M.: Public research and the quality of inventions: the role and impact of entrepreneurial universities and regional network embeddedness. Small Bus. Econ., 1–18 (2021)
23. Chen, M.Y., Liao, C.H., Lughofer, E.D., et al.: Informetrics on social network mining: research, policy and practice challenges. Libr. Hi Tech. **38**(2), 273–275 (2020)