Anomaly Monitoring System of Enterprise Financial and Economic Information Based on Entropy Clustering

Yu Chen(✉) and Kaili Wang

School of Business, Nantong Institute of Technology, Nantong 226000, China
Cyhsb88@163.com

Abstract. The currently used information anomaly monitoring system has problems of low accuracy and efficiency. In this regard, an abnormal monitoring system of enterprise financial and economic information based on entropy clustering is designed. On the basis of the design of the hardware and control module of the economic information abnormal monitoring system; the crawler tool is used to collect the financial and economic information of the enterprise; the collected data information is cleaned by the mapping operation; the processed data is processed by the abnormal knowledge discovery principle. In feature extraction, abnormal information features are obtained through decision tree; the abnormal information monitoring is realized by using the k-means algorithm of information entropy. The experimental results show that the designed system has an average alarm correct rate of 92.55% and a short response time, which is of practical value.

Keywords: Entropy clustering · Corporate financial economy · Information anomaly · Monitoring system · Data cleaning · k-means

1 Introduction

With the rapid development of modern communication and communication technology and the rapid expansion of the popularization of the Internet, great changes have taken place in the way of information dissemination and reception. Mass information from multiple channels and multiple sources helps people to fully understand events and make correct decisions. However, while enjoying the advantages of multi-channel mass information with multiple dimensions and strong real-time performance, people are also deeply disturbed by it. First, massive amounts of information cannot be obtained and processed by humans alone; secondly, large-scale information from different sources has strong redundancy, making information processing difficult; thirdly, a large amount of noise affects the acquisition of effective information; finally, the efficiency of information center topic judgment is low, the information is difficult to classify correctly. With the continuous deepening of the informatization construction of the financial service industry, the financial industry has become one of the industries with the highest degree of informatization in my country [1]. With the rise of the Internet era, the financial

W. Fu and L. Yun (Eds.): ADHIP 2022, LNICST 468, pp. 217–230, 2023.
https://doi.org/10.1007/978-3-031-28787-9_17

industry has entered the era of information explosion. In addition to paying attention to financial information acquisition channels such as official announcements, professional financial websites, and new financial media platforms, financial institutions and practitioners must also pay attention to financial information sharing platforms such as social networks., it not only consumes a lot of time, but also has low browsing efficiency, and it is easy to miss important information, the news is relatively lagging, it is difficult to respond to risks in time, resulting in financial losses [2]. Perfecting the multi-industry and multi-level financial early warning system, establishing a financial supervision system commensurate with the level of financial development, and enhancing the efficiency of financial supervision are also major issues that must be considered in financial risk prevention. The supervisory department should abandon the traditional post-event supervision method, and use the whole-process monitoring method combining pre-event supervision, in-event supervision and post-event supervision to better prevent financial risks.

Regarding the development of financial business security monitoring information application system, domestic researchers, major banks and system integrators are more active and mature in academic research, system development and system application of financial business security monitoring information application system in the financial industry. However, at present, many domestic financial business security monitoring information application system developers mostly use mobile or China Unicom direct connection. Bank data is landed on mobile or China Unicom, and there are certain security risks in customer financial information [3]. For enterprises, financial and economic information of enterprises is very important for enterprises to guard against financial and economic information. Under the background of the rapid development of modern information technology, the amount of financial and economic information of enterprises is increasing exponentially. By monitoring abnormal financial and economic information of enterprises, it is possible to avoid mistakes in corporate management policies caused by abnormal financial and economic information in the process of operation., or cause corporate financial and economic risks. Therefore, the abnormal monitoring of the financial and economic information of the enterprise is carried out to facilitate the timely adjustment by the business decision-makers of the enterprise, so that the information is more valuable and the information can better serve the business development of the enterprise.

The current methods for abnormal monitoring of information data can be roughly divided into distribution-based methods, graph-based methods and cluster-based methods. The distribution-based anomaly detection method is the earliest detection method [4]. This detection method is based on the assumption that the dataset follows a certain statistical distribution. The statistical knowledge of these statistical distributions is then used to construct mathematical models. Finally, a statistical method is used to detect whether the data conforms to the constructed mathematical model. If there is a significant difference between the detected data and the mathematical model, it is considered that the current monitoring data is an abnormal behavior, and the monitoring of abnormal data is completed. But this method is extremely difficult to estimate or calculate the distribution of the data when the data is in high dimension. The graph-based anomaly detection method [5] mainly maps high-dimensional data to a low-dimensional space through visualization techniques. Outliers appear at special locations in the mapped image. The advantage of using the visualization method is that the possible abnormal points in the data can be found directly and clearly. However, this visualization method requires artificially given five parameters, and human factors will affect the final detection result. The visualization result contains a large number of manual parameters, which requires a lot of manual participation, and the detection results obtained are also different depending on the mapping method of the graph. Therefore, this method has the disadvantages of large workload and uncertainty. Using cluster analysis, the similarity between data can be used to classify unlabeled data objects into different classes without relying on prior knowledge. Cluster-based anomaly detection methods do not require prior knowledge of the data set, and can effectively cluster data and discover anomalies in it. However, this method is prone to the disaster of dimensionality in multi-dimensional situations, and the anomaly detection performance for high-dimensional data is poor. In this regard, this paper designs an abnormal monitoring system for corporate financial and economic information based on entropy clustering. Design the hardware and control module of the abnormal economic information monitoring system, use crawler tools to collect financial and economic information of enterprises, clean the collected data information through mapping operations, and use the principle of abnormal knowledge discovery to characterize the processed data Extraction, obtain abnormal information features through decision tree, and finally use the k-means algorithm of information entropy to realize abnormal information monitoring. Information entropy reflects the stability of the system, and it has been widely used in anomaly detection in recent years. Information entropy is a description of information distribution, which reflects the distribution of uncertainty of random variables in the system. The more stable a system is, the lower the information entropy value is, on the contrary, the more chaotic a system is, the higher the information entropy value is. This paper will design an abnormal financial and economic information monitoring system based on entropy clustering, and reduce the negative impact of abnormal financial and economic information on related businesses by applying the monitoring system. The experimental results show that the average alarm accuracy rate of the system designed in this paper is 92.55%, and the average response time is 10.97s. The accuracy rate is high and the response time is short.

2 Design of Hardware Part of Enterprise Financial and Economic Information Abnormal Monitoring System

This paper optimizes the function module based on the traditional hardware, and improves the system hardware by designing the system control module and communication module. The monitoring system should be extensible for future expansion of network services. The system should accommodate the need to plug in zone controllers and operator terminals anywhere without affecting the normal operation of the existing system. In addition, the monitoring system should be able to ensure long-term stability of various indicators in a long-term stable operation state, thereby reducing system operation and maintenance costs. The selection of the hardware part of the system saves energy as much as possible on the premise of meeting the functional requirements of the design, thereby reducing operating costs. Considering the importance of corporate financial and economic information, the system should be equipped with a permission management function, and users with different permissions can operate different contents, that is, the functions opened to users with different permissions are different. According to the above monitoring system hardware selection requirements, the hardware part of the information anomaly monitoring system designed in this paper adopts a core template of 50 mm * 70 mm, in which the S3C2410A microprocessor is installed, and the MYD-AM335X series development is extended based on the MYC-AM335X core board. The development board has rich peripheral resources including high-speed USB, LCD interface, CAN interface, 485 interface, ADC, SPI, GPMC, LED, 10/100/1000Mbps Ethernet interface, JTAG debugging interface serial port, etc. Figure 1 below is a schematic diagram of the hardware framework of the enterprise financial and economic information anomaly monitoring system.

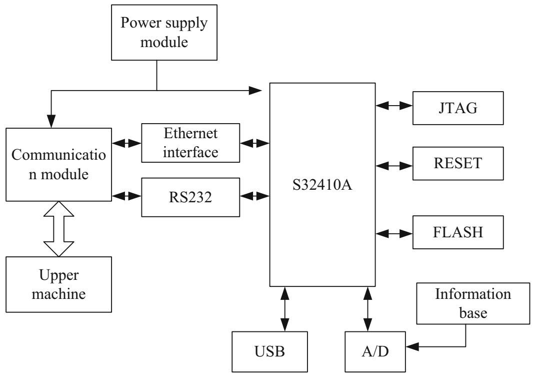

Fig. 1. The hardware frame diagram of the abnormal monitoring system of enterprise financial and economic information

2.1 System Control Module Design

The control module of the hardware part of the system is composed of S3C2410A processor and various peripheral circuits to meet the design requirements. S3C2410A is a low power consumption, high integration, 16/32-bit RISC microprocessor based on ARM920T core, and also adopts AMBA structure. In order to reduce the total system cost, S3C2410A integrates the following resources: 16 KB instruction Cache (cache), 16 KB data Cache, MMU for virtual storage management, LCD controller, NAND Hash controller (supports 4KB code direct boot), SDRAM controller, 3 UART serial ports, 4-channel DMA, 4-channel PWM timer, watchdog counter, 117 multi-function I/O ports, 24 external interrupt sources, RC real-time clock, 8-channel 10-bit ADC and touch screen Interface, USB Host/Device, SD card interface, 2 SPI interfaces, I2S bus interface [6].

Based on the ARM920T core, the S3C2410A processor has built-in Embedded ICE debug module with standard JTAG interface. You can use hardware emulators such as Multi-ICE to perform online real-time simulation debugging through the JTAG interface, or use programming tools such as SJF2410 to program the externally expanded FLASH memory through the JTAG interface. The system hardware microprocessor designed in this paper uses the standard 20-pin JTAG interface to program the test program.

The $3C2410A itself integrates the NAND FLASH controller with the NAND FLASH Boot Loader function, the user can directly expand the large-capacity NAND FLASH memory, and can directly start the user from the NAND FLASH after system reset by configuring the OM [1:0] pins code. The system in this article only uses 64 * 8 bits K9F1209 NAND FLASH memory. After the data has been verified, it can be transferred to the EERAM by the copy scratchpad command to ensure the integrity of the data when changing the memory. The scratchpad is 9 bytes, and the 0th and 1st bytes are the low and high bytes of the temperature code. Bytes 2 and 3 are copies of the temperature encoded low and high bytes, and byte 4 is the configuration register [7].

Among them, the nRTS and nETS signals of UART0 are brought out to support the application of automatic flow control (AFC). In AFC, nRTS is controlled by the receiving condition of the receiver, and nCTS controls the work of the transmitter. Since S3C2410A is a 3.3 V system, the working power supply range is 3.3 V. Call. 5 V MAX3232 for RS. 232 level shifting.

2.2 Communication Module Design

When communicating, first set the communication format, that is, write the D8120 register, and set it to 0C86 under this system condition, that is, the data length is 8 bits, no parity, no start bit and stop bit, baud rate 9600 bps. After modifying the D8120 setting, make sure to power on and off the PLC once. Then use the RS command to transfer data. When processes interact and share data, upstream processes generate new data, and downstream processes consume data. Since each process executes at a different speed, if it is not scheduled, the buffer may fill up quickly, or the same information may be called repeatedly for processing. This problem is generally solved by introducing a circular queue structure. The producer process and the consumer process are located at the head and tail of the queue, respectively, and control the writing of producer data and the reading of consumer data through put operations and get operations, and the rest of the queue is buffer space. The get operation is disabled if the buffer has no data, and the put operation is disabled if the buffer is full. The problem of speed mismatch when processes interact is effectively solved. In this system, the data collection process is responsible for collecting data as a producer, and the data transmission process transmits data as a consumer. The above method is used to complete the process scheduling.

This system uses the RTU mode to transmit the communication protocol, and uses the RS-232C compatible serial interface for communication. The main advantage of this method is that more data can be transmitted than the ASCII method at the same baud rate. When using RTU mode, message transmission begins with a pause interval of at least 3.5 character times. The first field of the transfer is the device address, and the transfer characters that can be used are hexadecimal values. During communication, network devices continuously detect the network bus, including the pause interval, when the first field (address field) is received, each device decodes to determine whether it is sent to its own. After the last transmission character, a pause of at least 3.5 characters is required to mark the end of the message, after which a new message transmission can begin, as shown in Table 1.

Table 1. RTU mode message frame structure

Start Bit	Address	Function code	Data	CRC check	Stop Bit
T1-T4	8 bits	8 bits	N * 8 bits	16 bits	T1-T4
Size is 1 character	The address is 0–247 (decimal)	When a message is sent from the master device to the slave device, the function code field will tell the slave device what actions need to be performed; in response to the objection, the slave device returns a code equivalent to the normal code, but the most important position is logic 1	The data field is composed of two sets of hexadecimal numbers in the range 00…FF	The CRC field is appended at the end of the message, with the low byte followed by the high byte. Therefore, the high-order byte of the CRC is the last byte of the sent message	Least Significant Bit–Most Significant Bit

Using RTU mode, the message includes an error detection field based on the CRC method. The CRC field detects the content of the entire message. The CRC field is two bytes containing a 16-bit binary value. It is calculated by the transmitting device and added to the message. The receiving device recalculates the CRC of the received message and compares it with the value in the received CRC field. If the two values are different, there is an error. The implementation of the CRC algorithm will be described in detail later.

The generation of CRC check bytes is a more critical step, and the process is relatively complicated. The steps are as follows:

(1) Preset a 16-bit CRC register as hexadecimal FFF, that is, all digits are 1.
(2) The lower 8-bit byte of the 16-bit register is XORed with the lower 8-bit of the first byte of the information frame. The result of the operation is placed in a 16-bit register.
(3) Shift the 16 register one bit to the right and fill the high bit with 0.

(4) If the digit shifted to the right (marker bit) is 1, the generator polynomial A001 (1010000000000001) is XORed with this register; if the digit shifted to the right is 0, return (3).
(5) Repeat (3) and (4) until 8 bits are shifted out.
(6) Repeat (3) to (5) until all bytes of the message are XORed with the 16-bit register and shifted 8 times.
(7) The high and low-order bytes of the obtained l6-bit CRC register, that is, 2-byte CRC, are added to the message.

In this paper, the OPB bus is used to connect the hardware-accelerated IP core control logic, because the OPB bus is mainly used for data transmission to external devices. The OPB bus supports multi-bit data, and accepts input from the host when used for peripherals, and performs specified operations, which meets the needs of hardware acceleration. The OPB bus performance is shown in Table 2 below.

Table 2. OPB bus performance table

Serial number	Project	Parameter
1	Data line width	8-bit, 16-bit, 32-bit
2	Address line width	32 bit
3	Architecture	Multiple master/slave devices
4	Data transfer protocol	Single read/write transfer, supports trigger transfer, word, byte, half-word transfer
5	Timing	Synchronize
6	Connect	Multiplex
7	Interconnected	Bus-independent read/write data lines, no tri-state support

The data processing module completes the feature parameter extraction, starts the hardware acceleration module, sets ce to a high level, and the hardware acceleration module starts to work. The three channels of data are processed at the same time, and the respective calculation results prob1, prob2 and prob3 are output. The three output ready signals rdy1, rdy2 and rdy3 are at high level, indicating that the calculation results of the respective components are completed after the operation.

After designing the main core modules of the hardware part of the abnormal information monitoring system for corporate financial and economic information according to the above content, the information entropy clustering algorithm is used to process the collected and acquired corporate financial and economic information data, and the software part of the abnormal information monitoring system is designed to realize Default system functions.

3 Design of Software Part of Enterprise Financial and Economic Information Anomaly Monitoring System Based on Entropy Clustering

3.1 Collection of Corporate Financial and Economic Information

The data acquisition port of the hardware system is connected with the financial and economic information data database interface of the enterprise to obtain the financial and economic information data of the enterprise in a specified time period or a specified business. However, the differences in the scale of enterprises lead to too messy and large amount of financial and economic information data of enterprises. In this paper, crawler tools are used to collect financial and economic information of enterprises.

The distributed crawler tool of this system is based on the Soapy framework, uses an asynchronous non-blocking method to send requests to the target URL, and has a complex structure for the Soapy crawler, does not support distributed crawling, lacks website anti-crawling mechanism response methods, and after downtime The crawler state cannot be recovered and other problems are optimized, the crawler configuration and crawler process are componentized, the crawler configuration parameters are defined by the combination of the preset configuration file and the manual input interface, and the crawler is abstracted into multiple crawler tasks and distributed to multiple machines. Server processing, control of crawlers and extraction of target data through configuration parameters, and solving the problem of website anti-crawling by adding different anti-crawling components, so that new crawlers can be customized simply by modifying configuration files and extracting templates without modifying the crawler logic. Set up a persistent distributed queue of URLs to be crawled to complete the incremental crawling and continuous crawling of the crawler. The crawler task is released through the distributed task queue and assigned to multiple servers for execution to ensure the efficient operation of the crawler and The crawler anomaly is alerted by the crawler monitoring module. The different modules of the system are completely decoupled, which is convenient for further expansion. While ensuring the efficiency and stability of distributed crawling, it is deeply customizable for users. The configuration file is not restricted by the development and running environment, and has a strong generalization suitability. The crawler task scheduling module is responsible for the generation and scheduling of crawler tasks, including crawler task construction unit, task scheduler, task release queue, task execution unit, task execution result queue, task retry queue and task monitoring unit [8].

The crawler task construction unit is used to generate crawler tasks containing various parameters according to the configuration items in the configuration module, which are expressed in the form of key-value and serialized into json strings. Each crawler task has a unique task record. The crawler task construction unit first cannot accept the crawler task parameters by manually inputting the configuration script, finds and parses the static configuration file of the corresponding website through the crawler name, and parses the corresponding running configuration file according to the name of the running configuration file.

The task scheduler obtains the serialized crawler tasks from the crawler task construction unit and the task retry queue, and then detects the task type. For tasks that exceed the maximum number of retries, update their status and send them to the task execution result queue. For common asynchronous tasks and retry tasks within the maximum number of retries, they are directly sent to the associated task release queue. Save periodic tasks and periodically send them to the corresponding task release queue according to the set time interval. After the crawler is used to collect the financial and economic information data of the enterprise, the information data is cleaned and processed.

3.2 Financial and Economic Information Data Cleaning

Since the crawler collects a large amount of financial and economic information and includes many data items, if it is simply written to a text file, the subsequent work such as calculating the information entropy value will require frequent reading and writing operations of the file, which will consume a lot of system resources. The time cost will be huge. The financial and economic information extraction subsystem uses the data flow-data node method to process the collected data in real time, and uses the self-defined analysis plug-in to perform online analysis of the processed data in small batches. Data flow defines the calculation process of data, and data nodes are responsible for executing specific calculation logic, including data cleaning and data standardization nodes. The analysis plug-in is customized according to the business, and provides a plug-in for each analysis business, which is responsible for continuously pulling the unanalyzed data of the business to perform analysis and calculation logic. The plug-ins of different data analysis services are independent of each other. In this paper, the task of extracting financial and economic information is a two-stage task. First, we identify and filter information data that may have abnormal financial and economic information in the information, and then extract information elements from the filtered sentences. If all attributes are used in anomaly detection, it will inevitably increase the amount of computation and consume a lot of system resources, and not all attributes contribute to anomaly detection. Therefore, after the system hardware collects the financial and economic information data of the enterprise, the information is cleaned.

This article will take a mapping operation on the original data and organize it into new clean data. The data selection module queries the required data as the object to be cleaned according to the user's requirements, and fuses the objects to be cleaned by querying set $\{q_1, q_2, \ldots, q_n\}$, and the result obtained is a metadata set. The first step of data cleaning, the form is defined as [9]:

$$CS_{q_1,q_2,\ldots,q_n}() = q_1 \cup q_2 \cup \ldots \cup q_n \tag{1}$$

Because there are many different types of data such as date and time in most business environments, almost all data cleaning implementations must transform these various forms of data into the canonical format of the rule base. For the needs of decision analysis, the data in the data source is generally not stored in a coded form. Before saving into the data source, look up the corresponding text description from the dictionary according to the code, and use the text description to replace the code. Repeatability judgment is mainly completed by three operations: Cartesian product, matching and clustering. In order to improve the efficiency of the data matching operation, we design the Cartesian product operation here, which will only be performed on the data set of the Lai attribute taken out. That is, the data in the data set D output by the data standardization module is subjected to the Cartesian product operation, that is, $D * D$, and the following data pairs can be obtained:

$$D * D = \begin{bmatrix} b & s & b & s \\ b_1 & T & b_2 & T' \\ b_1 & T & b_3 & p \\ b_2 & T' & b_3 & p \end{bmatrix}_{\mathrm{v}} \tag{2}$$

Among them, $b = [b_1, b_2, b_3]$ and $s = \left[T, T', p\right]$ are data sub-items in the dataset. The combination compares the results of matching key attributes, and if and only if the key attributes of an element are all repeated attributes, it can be deduced that the element is a repeated element. After cleaning the repeated financial and economic information data, the information entropy clustering method is used to cluster the financial and economic information to monitor the abnormality in the information data set.

3.3 Clustering of Abnormal Information of Corporate Finance and Economics

The system uses the principle of abnormal knowledge discovery to extract the characteristic data in the financial and economic information of enterprises. Organize the knowledge of corporate financial and economic rules into a rule knowledge tree, which is a directed acyclic tree. The root node in the tree has no parent node, all other nodes have one and only one parent node, and each node can have one or more child nodes, or no child nodes. If a node has no child nodes, it is called a leaf node. Others are called internal nodes. Non-leaf nodes correspond to various possibilities in the value space of each condition attribute in the data set of abnormal financial and economic cases of enterprises. The content of leaf nodes is in addition to the corresponding rules. The conclusion is the attribute value (yes or no) of the decision attribute, including the combination of attribute values reflecting the entire path from the root node to the direct parent node of the leaf node as a condition, and the decision attribute value of the leaf node is the rule of conclusion Whether a valid time-to-live. The mined rule sets are stored and transmitted in the form of decision trees. After receiving the knowledge base subsystem, the initial rule knowledge tree is established by increasing the survival time of the corresponding rules for each leaf node. The rule knowledge tree will be continuously adjusted as the system runs., the number of rule knowledge trees will also increase or decrease. At any time of system operation, the forest composed of all rule knowledge trees constitutes the current rule knowledge base [10]. When a new mining rule comes, adjust and prune the corresponding rule knowledge tree and prune the decision tree corresponding to the new rule set according to whether the new rule set matches the rules in the current rule knowledge base (the premise is the same, the conclusion is the same) or not: The decision leaf node and the rule-related path corresponding to the rule matching the rule in the current rule knowledge base should be pruned, and the TTL value of the corresponding leaf node of the rule knowledge tree where the rule is located should be increased, otherwise it should be retained. For new rules with conflicting (same premise, different conclusions), adjust the decision attribute value of the leaf node of the rule corresponding to the rule tree in the rule knowledge base, and set its TTL value to the initial value max, and then prune the corresponding rule in the decision tree. The leaf nodes and the unique ancestor nodes and paths of the rules, that is, the new rules are discarded for the conflicting rules. The abnormal information attributes in the decision tree are mapped to the feature set, that is, the abnormal information characteristics of the financial economy of the enterprise are obtained.

If the processed enterprise financial and economic information data is $X = \{X_1, X_2, \ldots, X_N\}$, and if the information feature is $f = \{f_1, f_2, \ldots, f_N\}$, it indicates the state of the information data. Assuming that the characteristic attribute f_i of the corporate financial and economic information appears n_i, then the information entropy of the corporate financial and economic information can be defined according to the following formula:

$$H(X) = -\sum_{i=1}^{N} \left(\frac{n_i}{\sum f_i} \right) \log_2 \left(\frac{n_i}{\sum f_i} \right) \tag{3}$$

It can be known from mathematical knowledge that the value range of the information entropy value is $\left(0, \log_2 N\right)$. When the value of $H(X)$ is 0, it represents the maximum centralized distribution of the measurement data set. When the value of $H(X)$ is $\log_2 N$, it represents the maximum scattered distribution of the measurement data set.. Through the change of the information entropy value, it is possible to judge the orderly degree of a piece of enterprise financial and economic information data, and intuitively understand the changes of the characteristics and attributes of the enterprise's financial and economic information.

The information entropy value of the characteristic attributes of the financial and economic information of different abnormal enterprises is used as the classification and determination standard of entropy clustering, and the k-means algorithm combined with information entropy is used to realize abnormal information monitoring according to the following steps.

(1) Calculate the information entropy value of the characteristic attribute of the detection information, and calibrate the type of the detection information according to the information entropy to obtain a detection data set;
(2) The cluster center finally output in the training phase is used as the initial cluster center in the detection phase, and the Euclidean distance is calculated for the detection samples in the detection data set, and cluster analysis is performed;

The Euclidean distance calculation formula is as follows:

$$dis(t, s) = \sqrt{\sum_{k=1}^{n} (t_k - s_k)^2} \tag{4}$$

Among them, t is the financial and economic information data of enterprises that need to be monitored; s is the financial and economic information data of sample enterprises.

(3) Calculate the distance from the detected data to the cluster center, and compare it with the classification threshold. If it is greater than the classification threshold,

the current sample is considered to be new abnormal information, and the K value of the number of clusters is updated plus 1, and the current sample is regarded as the new cluster. If it is less than the classification threshold, the current sample is classified into the nearest cluster for anomaly detection.

(4) If a certain detected sample is classified as abnormal information, start the abnormal response module. Repeat the above process until all samples in the test sample set are tested;
(5) Output the clustering results, and carry out corresponding monitoring and early warning according to the clustering results.

After combining the software part and the hardware part designed above, the design and research of the abnormal monitoring system of enterprise financial and economic information based on entropy clustering is completed in theory.

4 System Test

This section builds an abnormal monitoring system of corporate financial and economic information based on entropy clustering, tests the feasibility and monitoring effect, and evaluates the system from the aspects of comprehensiveness, real-time and accuracy.

4.1 Test Plan Design

First, the hardware test of the monitoring system is carried out, and the hardware part is divided into the circuit schematic diagram and PCB drawing of each module. Carefully check the connection relationship of each pin before submitting the board, pay special attention to the pins without electrical properties, and carefully check whether the pins are missing in the schematic diagram. When welding components, you need to pay attention to the positive and negative poles of the diode, the pin direction of the chip, etc. When the components are welded and the system is not powered on, use a multimeter to measure whether the power line and the ground wire are short-circuited. After all tests are normal, power on the system to check whether the chip is overheating or smoking. If the circuit board is faulty, quickly disconnect the power supply to check the fault; power on the system after the fault is solved.

After the hardware part is tested, the response time of the system and the monitoring and alarm accuracy of abnormal information are the test indicators. By comparing the system designed in this paper with the monitoring system based on multi-source information fusion, we can intuitively evaluate whether the designed system meets the requirements of the enterprise need.

4.2 Evaluation Index

Accuracy is an indicator used to evaluate data. Here, it refers to the proportion of correct results of abnormal monitoring and alarm of enterprise financial and economic information. The definition is as follows:

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \tag{4}$$

In the formula, *TP* represents the true case, *TN* represents the true negative case, *FP* represents the false positive case, and *FN* represents the false negative case.

4.3 Test Results and Analysis

In order to test the effectiveness of the method in this paper, the system performance of the method in this paper and the traditional clustering method are compared and tested by using the artificial enterprise financial and economic information data set with known abnormal information. The comparison results of the system performance test are shown in Table 3 below.

Table 3. Comparison results of system performance tests

Abnormal information content/%	The method of this paper		Clustering method	
	Response time/s	Monitoring alarm accuracy rate/%	Response time/s	Monitoring alarm accuracy rate/%
0.5	10.73	93.2	15.78	89.6
1.0	10.94	90.6	15.98	85.2
1.5	11.12	92.7	16.57	87.3
2.0	11.06	93.3	23.07	75.6
2.5	10.84	92.4	24.3	79.4
5.0	10.95	94.5	25.76	82.7
8.0	11.07	92.6	27.19	86.1
10.0	11.03	91.1	28.07	90.5

It can be seen from the data in Table 3 that the response time of the method in this paper is significantly shorter than that of the clustering method. The average response time of the method in this paper is 10.97 s, and the average response time of the clustering method is 22.09 s. The method in this paper has an average correct rate of 92.55% for monitoring and alarming of abnormal corporate financial and economic information, and the average correct rate for monitoring and alarming of the clustering method is 84.55%.

To sum up, the monitoring accuracy rate of the abnormal monitoring system of enterprise financial and economic information based on entropy clustering designed in this paper is higher than 90%, and the response efficiency of the system is high, which can effectively improve the operation efficiency of enterprises.

5 Concluding Remarks

In the financial market, liquidity risk is the most common risk. The decision-making basis of economic participants is all kinds of information in the market and enterprises.

However, it is extremely complicated and difficult to obtain relevant information about the industry and the market. Risk is an unavoidable uncertainty in the development and operation of an enterprise, and its occurrence is unavoidable. The abnormal monitoring of corporate financial and economic information can reduce the risk of corporate operation caused by abnormal information. To this end, this paper designs an abnormal monitoring system for corporate financial and economic information based on entropy clustering. Through the system performance test, it is verified that the monitoring system can accurately monitor and warn the abnormal information in financial and economic information. Effective use of financial and economic information reduces the negative impact of abnormal information on business operations.However, in the process of monitoring abnormal financial and economic information of enterprises, due to the complexity of the algorithm, the monitoring time did not achieve the expected effect, resulting in a decrease in monitoring efficiency. In the following research, the algorithm will be improved to shorten the calculation time. Improve the monitoring efficiency of abnormal financial and economic information of enterprises.

Fund Project. Jiangsu Province "14th Five-Year" Business Administration Key Construction Discipline Project (SJYH2022-2/285).

References

1. Zhong, H.: The status quo and countermeasures of high-quality financial development in China—research based on international comparison. Southwest Finan. (02), 74–84 (2021)
2. Deng, C., Peng, B.: Research on the relationship between the financialization behavior of entity enterprises and the information disclosure quality. Theor. Pract. Finan. Econ. **42**(03), 110–117 (2021)
3. Shi, Z..-J.., Zhou, X..-J.., Li, K., et al.: Cyberspace security monitoring technology based on multi-source information fusion. Comput. Eng. Des. **41**(12), 3361–3367 (2020)
4. Shi, X., Qiu, R., Ling, Z., et al.: Spatio-temporal correlation analysis of online monitoring data for anomaly detection and location in distribution networks. IEEE Trans. Smart Grid **11**(2), 995–1006 (2019)
5. Zhao, H.-Y., Liu, K., Wang, T.-M., et al.: Simulation of abnormal information acquisition for network text implication relationship recognition. Comput. Simul. **37**(08), 256–260 (2020)
6. Liu, H., Wang, Y., Chen, W.G.: Anomaly detection for condition monitoring data using auxiliary feature vector and density-based clustering. IET Gener. Transm. Distrib. **14**(1), 108–118 (2020)
7. Li, Y.-T., Guo, J., Qi, L., et al.: Density-sensitive fuzzy kernel maximum entropy clustering algorithm. Control Theory Appl. **39**(01), 67–82 (2022)
8. Kim, H., Park, J., Min, K., et al.: Anomaly monitoring framework in lane detection with a generative adversarial network. IEEE Trans. Intell. Transp. Syst. 22(3), 1–13 (2020)
9. Zheng, W., Yao, J., Wang, J., et al.: Segmentation of MRI brain image based on fuzzy entropy clustering and improved particle swarm. Laser J. **42**(01), 98–103 (2021)
10. Wei, L., Xuesongm, H., Ming, Z., et al.: Abnormal behavior detection method for power monitoring system based on fully connected residual neural network. J. Southeast Univ. (Nat. Sci. Edn.) **50**(06), 1062–1068 (2020)