Design of Human Resource Big Data Parallel Classification System Based on Social Information Cognitive Model

Bo Sun and Cai-ming Zhang(✉)

School of Labor Relations and Human Resource, China University of Labor Relations, Beijing 100048, China

Abstract. In order to improve the classification function and operation performance of human resources, a new big data parallel classification system is designed. The parallel processor is installed to optimize the analog-to-digital converter, human resource data storage and wireless communication network. This paper constructs a social information cognitive model, under which human resource data can be obtained in real time, and preprocessed by data cleaning, Chinese word segmentation and stop word elimination. Human resource data features are extracted, and the similarity between the extracted data features and standard data features is calculated to realize the parallel classification function of human resource big data. Through the system test experiment, the conclusion is drawn: compared with the traditional classification system, the recall rate and accuracy rate of the design system are increased by 5.5% and 3.5% respectively, and it has more advantages in classification speed.

Keywords: Social information cognitive model · Human resources · Resource big data · Parallel classification

1 Introduction

Human resource management is the main component of enterprises and institutions, which has a great impact on the business contacts and development prospects of enterprises. In order to adapt to the change and development trend of the current economic structure of enterprises and improve the work efficiency and economic benefits of enterprises, the State applies big data technology to human resources management decision-making in all walks of life, and realizes a more comprehensive, systematic, efficient and realistic decision-making technology by using the software systems and frameworks in the big data platform, so as to enhance the core competitiveness of enterprises and public institutions and meet the long-term development goals of enterprises [1]. Human resource is an important part of enterprises and universities, which is mainly used to manage the information of employees in enterprises and facilitate personnel management. Because of the large number of employees in enterprises, universities and the society, in order to achieve effective utilization of human resources information and

S. Liu and X. Ma (Eds.): ADHIP 2021, LNICST 416, pp. 454–469, 2022.
https://doi.org/10.1007/978-3-030-94551-0_36

build human resources data, enterprises can acquire human resources information in big data, and employees can also upload real-time human resources data to big data to achieve resource sharing.

Due to the large number of social personnel, human resources data is relatively large. In order to achieve the efficient management of human resources big data, the human resources classification system came into being [2, 3]. Data classification is to combine the data with some common attributes or features, and distinguish the data by its category attributes or features. In other words, the same content, the same nature of information and information that requires unified management are gathered together, and the different information and information that needs to be managed separately are distinguished, and then the relationship between each set is determined to form an organized classification system. According to the research status at home and abroad, mature big data classification methods include the data classification system based on fractional-order derivative [4], and the big data classification system based on firefly and simulated annealing [5]. In reference [4], the fractional gradient descent method is proposed, which is an unconstrained optimization algorithm using support vector machine to train the classifier with convex problem. Compared with the classical integer order model, the fractional order model has a significant advantage in speeding up the calculation speed. In this study, in order to study the current situation that these new optimization methods can realize fractional derivative in the classifier algorithm, the problem is qualitatively studied. The main purpose of reference [5] is to introduce the ofs algorithm supported by meta heuristic algorithm of MapReduce paradigm. A new hybrid multi-objective firefly and simulated annealing (hmofsa) algorithm is proposed to select the optimal feature set. Therefore, as the first step, the original big data set is decomposed into sample blocks in the map phase. Hmofsaa algorithm is used to select the selected features from the examples. Part of the results are combined into the final feature vector in the reduce stage, and the kernel support vector machine classifier is used for evaluation. However, the above two traditional classification systems are lack of the construction of information cognitive model, and the acquisition of data to be classified is not comprehensive enough, resulting in the recall and precision of data classification can not reach the ideal effect.

However, the traditional classification system has the problem of low classification efficiency, so a parallel classification system of human resource big data based on social information cognitive model is proposed. Based on the traditional classification system, the system adjusts the classification mode, uses parallel processing mode to replace serial processing mode, and introduces social information cognitive model. Among them, parallelism means that two or more programs are executed in the same time period, with time overlap. The information cognitive model refers to the reaction law of employees' attitude and behavior to the social cues provided by others when they interact with others. Through parallel processing and the application of social information cognitive model, we hope to improve the classification efficiency of human resource big data classification system.

2 Hardware System Design of Human Resource Big Data Parallel Classification

The human resource big data classification system mainly uses text classification algorithm to classify various texts according to their characteristics. The system realizes the classification of large-scale unstructured text. The system has the following design goals: to classify large-scale unstructured human resource big data. The system can realize the function of feature extraction and classification of massive unstructured human resource big data, and the overall time consumption is controlled in an acceptable range. Test the system function and performance. Through the system test, the accuracy of system classification and the overall performance of the system are evaluated, and the improvement scheme is put forward for specific problems. Modular function. The main functional layers of the system are completely independent as different modules, and the loose coupling between each module is maintained as far as possible, so as to optimize, update and secondary development of the system in the future. Based on the overall goal of the system, the basic principles of advanced technology and usability, openness and extensibility, high stability and reliability, integration and fast response are followed in the system design, and the optimal design of the parallel classification system of human resources big data is realized from three aspects: hardware, database and software function.

2.1 Parallel Processor

Parallel processor is the core part of the computer, which plays a very important role in the overall operation of the computer. Parallel processor mainly includes arithmetic logic unit, instruction register, program counter, address calculator, data selector, controller and register group. The overall structure design block diagram of 32-bit parallel processor is shown in Fig. 1.

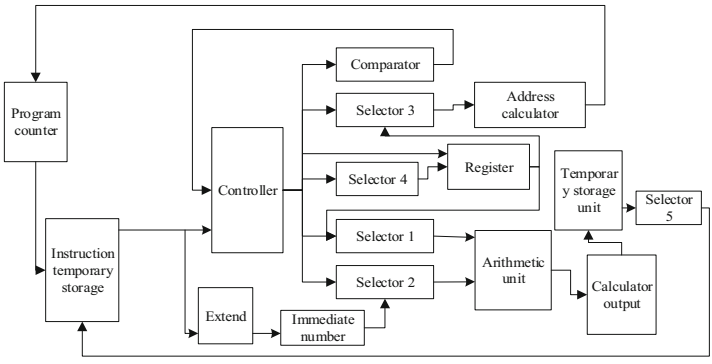

Fig. 1. Block diagram of parallel processor

The core idea of parallel processor pipeline is to divide each instruction into several smaller function segments, and each task is executed in parallel in time, so as to improve

the efficiency of operation. Microprocessors are divided into five stages: instruction fetching stage, instruction decoding stage, execution stage, storage stage and write back stage. The instruction fetching stage includes instruction memory, instruction register, program counter and address calculator. In the value cycle, the computer takes out the data in the corresponding address from the memory according to the value of the program counter and sends it to the instruction register for temporary storage, so as to save the total time. Then, the program counter calculates the address of the next instruction to be executed. The address calculator is also a register whose function is to accept the address of the next instruction to be executed. The instruction decoding stage includes control unit, data selector and register group. During the decoding cycle, the instruction is sent from the instruction register to the decoder in the control unit, which is responsible for decoding the instruction and sending the generated control command to other corresponding parts of the processor. The data selector mux3 controls the output of the address calculator and sends the next address result to the program counter. Register group stores some operation data temporarily in register heap during CPU operation. Execution phase: including arithmetic logic unit and two data selectors. In the execution cycle, the decoding result of decoding phase is taken as the basis. The output data of the data selectors mux1 and MUX2 are sent to the arithmetic logic unit. Alu will select which two 32-bit operands to complete arithmetic, logic, shift and other arithmetic or logic operations. The results of other instructions are branch or jump instructions. After the operation, the output results are temporarily stored in aluout.

2.2 Analog to Digital Converter

The dual-channel analog-to-digital conversion chip ADS5402 of Texas Instruments is selected as the ADC chip in the hardware system. The ADC has a maximum sampling rate of 800MSPS, 12 bit resolution, data output interface mode supports LVDS level standard, and the maximum peak to peak voltage of input analog signal is 1 V. The dual channel design of ADS5402 chip fully ensures the synchronization of IQ two channel signals, and can ensure the signal performance of the whole system to meet the synchronization requirements. The analog-to-digital conversion chip ADS5402 supports LVDS level standards. DDR can be selected as the output signal format, which not only increases the anti-interference ability of D signal, but also greatly improves the reliability and transmission rate of D data. The input signal of ads5402 is differential form. The analog input signal received by the system is generally single ended. Therefore, before the input analog signal is connected to ads5402, the input single ended analog signal needs to be differential transformed and then connected to the analog-to-digital conversion chip. Two back-to-back 1:1 transformers are used to convert the single ended signal into differential signal, and the parasitic capacitance effect caused by the transformer is eliminated. For the two input channels of the ADC, the two circuits in the input conversion circuit are exactly symmetrical and the same.

2.3 Human Resource Data Storage

The total memory capacity of human resource big data parallel classification system is 64 kbit. Using 0.18 μm standard CMOS process, the working voltage of the chip is

3.3 V, the working temperature range is −55 °C–125 °C, and the key design indicators are read time and read threshold. Read time, defined as the time from the beginning of address change to the beginning of output read-out data, expressed as Access_ Time. Considering the requirements of the chip, different levels of reading time are designed in the circuit design process. The fastest reading time index of the chip is 45 ns at room temperature, and the rest are 55 ns/18.2 MHz and 65 ns/16.6 MHz respectively. The circuit parameters corresponding to the latter two different operating frequencies are all given corresponding options in the layout for verification during chip test. When the reading threshold is at least 20K, that is, when the equivalent resistance is below 20K Ω after anti fuse breakdown, it is considered to store "1", otherwise it is considered to store "0". There are 26 peripheral input/output signals, including enable control signal, address signal, data signal, power and ground signal. The internal function module is composed of address decoder, control logic, readout circuit, programming circuit, storage array and bidirectional data port module. The memory structure is shown in Fig. 2.

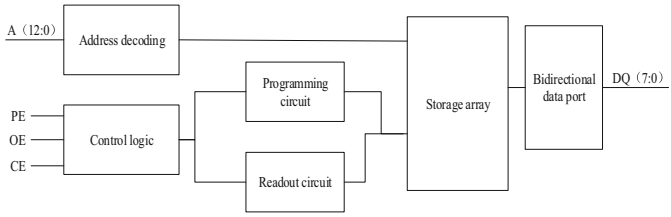

Fig. 2. Memory structure diagram

The 13 address signal lines A (12:0) of the memory are divided into three groups, which respectively control the block, row and column addressing of the memory array. The data signal line D (7:0) of the memory is a bidirectional port, that is, the programming data is input from this port, and the read-out data is output from this port. The number of ports indicates that the storage information width of the designed memory chip is 8 bits.

2.4 Wireless Communication Network

Wireless communication network adopts multi carrier modulation. Its basic principle is to divide high-speed data stream into low-speed data stream and transmit them on orthogonal subcarriers in parallel. In this way, the time dispersion of multipath fading is relatively reduced, and the frequency selective fading channel is transformed into several flat fading subchannels, which greatly reduces inter symbol interference. In wireless communication network, it is assumed that the system bandwidth is B, the number of subcarriers is N, the cyclic prefix is N_g sampling points, and the sampling period is T. If the transmission data of the first symbol period is X_i, then there are multiple carriers in the wireless communication network. Then the baseband equivalent signal can be expressed as:

$$s_l(t) = \sum_{k=0}^{N-1} X_{l,k}\phi_k(t - lT_s) \tag{1}$$

Where T_s is the length of the whole symbol including the cyclic prefix, and $\phi_k(t)$ is the carrier waveform function with frequency kB/N. According to the above communication principle, the human resource data of different regions are transmitted to realize the sharing of human resource data.

2.5 Component Chip Selection

FPGA Chip

FPGA is a kind of ASIC that allows users to configure and program independently. It supports the user to configure the system through the corresponding software. So through the control, we can complete the purpose of the system, and the field programmable gate array can also erase and write many times. Therefore, it greatly shortens the design cycle and R & D cycle of hardware products, and also greatly improves the flexibility of hardware implementation and reduces the cost of product design and R & D. The specific working principle of FPGA chip is based on the look-up table. Under the same hardware circuit conditions, the system related configuration files can be burned into the chip. Then, by querying the contents of the look-up table, we can achieve different logical purposes required by the system. Considering the data speed of ADC high-speed module and high-speed DAC module, XC5VLX330 chip is selected as FPGA chip.

Power Chip

The basic working principle of capacitive switching power supply is to make use of the energy storage characteristics of the capacitor, and store the input energy in the capacitor through the action of high-frequency switch through the controllable switch. When the switch is off, the electric energy is released to the load to provide energy. The output power or voltage is related to the duty cycle. Capacitive switching power supply can be used for step-up and step-down [6]. The internal FET switch array controls the charging and discharging of the fast capacitor in a certain way, so that the input voltage can be doubled or reduced by a certain factor, and the required output voltage can be obtained. Based on the traditional power supply chip, a linear regulator is embedded. The principle of the linear regulator is shown in Fig. 3.

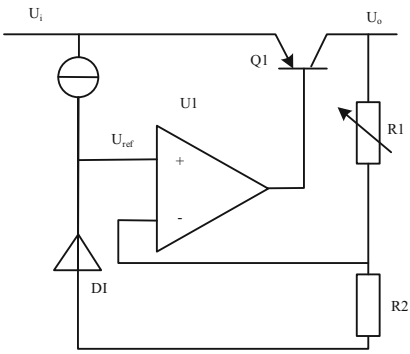

Fig. 3. Working principle of linear regulator

The sampling voltage is added to the non-inverting input of the comparator U1 and compared with the reference voltage U_{ref} added to the inverting input. After the difference between the two is amplified by the amplifier U1, the voltage drop of the series regulator is controlled to stabilize the output voltage. When the output voltage U_o decreases, the difference between the reference voltage and the sampling voltage increases, the driving current of the comparator amplifier increases, and the voltage drop of the series regulator decreases, so that the output voltage increases; If the output voltage U_o exceeds the required setting value, the front drive current of the comparison amplifier output is reduced, so that the output voltage is reduced.

3 Database Design of Human Resource Big Data Parallel Classification System

The system adopts ORACLE10G database, which is a large-scale database supporting JAVA. It is fully competent for the massive sharing of analysis platform data, and ORACLE10G's support for java is also satisfactory for the performance of system statistical queries. In the design of system database, the physical location of data storage should be determined first, then the data structure should be analyzed, and the relevant indexes and logs should be configured. After the above database analysis, the database concept and database table structure are analyzed [7]. The system is managed by SQL Server 2008, which can make the system more secure and reliable. In the human resource big data parallel classification system, the basic information management of employees is mainly to store the information of employees. If the information of employees needs tutors, the database will be updated in real time. The basic information of employees is shown in Table 1.

Table 1. Basic information of employees

Data name	Data description	Data type	Storage length
ID	Employee number (primary key)	Int	10
Name	Employee name	Varchar	20
Sex	Employee gender	Varchar	2
Date	date of birth	Date	8
ID number	ID number	Varchar	18
Place	Native place	Varchar	30
Tel	contact number	Int	20
Department	department	Varchar	30
Post	post	Varchar	20
Wages	base pay	Int	10

(*continued*)

Table 1. (*continued*)

Data name	Data description	Data type	Storage length
Form	Form of employment	Varchar	10
Education	Highest education	Varchar	20
University	University one is graduated from	Varchar	20
Date	Entry date	Date	8
State	On the job status	Varchar	10
Working years	working years	Int	5

In addition, the wage information table mainly stores the wage data information of employees, including the basic wage and performance wage of employees. The storage structure of salary information is shown in Table 2.

Table 2. Wage information

Data name	Data description	Data type
ID	Employee number (primary key)	Varchar(10)
Years	years	Int(10)
Bonus	bonus	Int(10)
Lunch	Lunch allowance	Int(10)
Transportation	Transportation subsidy	Int(10)

4 Software Design of Human Resource Big Data Parallel Classification System

4.1 Constructing the Cognitive Model of Social Information

Due to individual differences, that is, different work, life, learning environment and congenital differences, everyone's cognitive ability is in a relatively stable state. But when the brain is captured by scarcity, the focus will be relatively concentrated and the cognitive ability will be forced to decrease. Considering the changing process of users' cognitive ability. In the research of information retrieval cognitive model, it is necessary to analyze the whole cognitive process of users. In the design of information retrieval system, human intervention can make users' information needs more accurately expressed and interact with the system more tacit, so as to improve the utilization rate of information retrieval system. Figure 4 shows the dynamic change of users' cognitive ability affected by work and task situations.

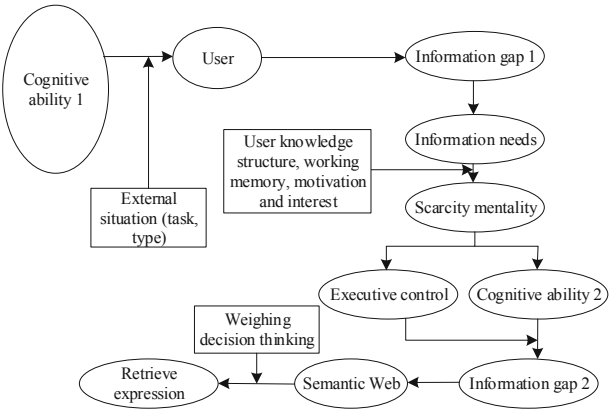

Fig. 4. Dynamic change of user cognitive ability

When users perform information retrieval, there will be an information gap between the user's knowledge structure and the task context. This difference is the sum of the knowledge that users objectively need to solve a certain problem [8]. Usually, when users are placed in the context of work tasks, they will have a scarcity mentality because they have less knowledge than they need. This mentality will capture the user's brain, making the bandwidth available to the user narrower, lacking insight and foresight. At this time, users use the external tool information retrieval system to acquire knowledge and make up for the lack of knowledge state when they solve the problem. When users express their needs to the information retrieval system, due to the capture of users' attention by scarcity mentality, the information needs expressed by users at this time will be less than the objective and actual needs needed to solve the problem. In this state, when the user interacts with the system, the information that the system feeds back to the user can not completely solve the current problems faced by the user. In order to solve this problem, it is necessary to analyze the user's cognitive ability and the ability to express their needs, find its root, and solve the impact of the dynamic change of user's cognitive ability on the search results. From the perspective of scarcity theory, this paper analyzes the dynamic change process of users' cognitive ability, and constructs a social information cognitive model. The model includes three layers: internal cognition layer, trade-off decision-making layer and human-computer interaction layer. The internal cognitive layer analyzes the change process from the user's objective knowledge needs to the user's cognitive needs. The trade-off decision-making layer is a process in which users form semantic network under the influence of weakened executive control and trade-off decision-making thinking. The internal cognitive layer and the trade-off decision-making layer constitute the user's cognitive process. The human-computer interaction layer is a process in which users express their needs in the form of retrieval after cognition and match with the information retrieval system, so as to obtain the required information.

4.2 Real Time Access to Human Resource Data

Multipoint data collection refers to the use of multiple servers at the same time from the same or multiple different data repository to obtain data in the form of web pages. For example, for human resource data, the storage structure and storage format of data in different resource databases are different, but the data will contain the following fields: company name, personnel name, position name, position description, position demand, job search keywords, address, publishing time, age, province, city, county, education background, etc. The collected data must undergo many structural and format transformations before further statistical research. The process of data collection is to directly save the web page in the form of text data locally, and then save it in the database in the form of binary. The purpose of multi-point data acquisition is to quickly collect the data from multiple remote data resource databases to the local database, saving valuable time. In order to dynamically obtain the data needed on the web page and download it in time, we should first set the corresponding conditions to process the data, then query through the trigger time of the database, and the database will immediately compare the relevant data to get the correct data. The data acquisition system processes the relevant pages through DOM technology, and controls the pages according to the corresponding requirements. For example: according to the theory of query data, the condition of the page is analyzed, and the data in the page is initialized according to the customer's needs. These values are processed dynamically in the background according to the needs. At the same time, various events on the page, such as button box link events, are set. According to the requirements of query events, start the query process to query the data effectively. Download the query data and save it to local in the form of web page. If the data is displayed in multiple pages, the corresponding new data can be found according to different page numbers.

4.3 Human Resource Big Data Processing

Human resource big data preprocessing includes three steps: data cleaning, Chinese word segmentation and stop word removal. In the process of data cleaning, we first get the data of the page to be cleaned to judge whether it needs to be cleaned. If there is data to be cleaned, we call the template framework of the page to be cleaned to get the structure description of the page header. Check the header structure and page control module, and output the initial processing results of HR big data after cleaning. Word segmentation is an important step in text preprocessing. In English, there are spaces between words for segmentation, but Chinese is composed of continuous strings without obvious separators, so it is necessary to perform word segmentation on strings [9]. After depunctuation, word segmentation and other operations are performed on all samples in the training text set, the word list of each sample is obtained. The set of the list and column of the obtained words constitutes the text feature dictionary. In order to reduce the data noise caused by the text representation dictionary in the feature vector, improve the ability of text discrimination, and reduce the search range of the representation dictionary in training. To improve the classification accuracy and training efficiency of the classifier, before representing the sample as feature vector, the text representation dictionary will generally implement operations such as "taking root" and "removing

stop words". Stop words usually refer to high-frequency words that appear frequently in various categories of documents at the same time, and are considered to be unfavorable for classification. Generally, they are some prepositions, pronouns, conjunctions, etc. Filtering out the words in the stop-word list from the text representation dictionary is the ultimate goal of de-stopping.

4.4 Extracting Human Resource Data Features

The human resource data feature extraction prepared by the data simulation layer, and the selection of human resource feature words to construct the human resource data feature vector in the language vector model is the core part of the classification process. There are three steps for feature extraction: word frequency statistics, word weight calculation and feature word sorting. Count the frequency of a given word in a given file, that is, calculate the word frequency of the given word. At the same time, count the number of documents in which the word appears in all documents, that is, calculate the inverse word frequency of the word. The calculation formula of word frequency statistics is as follows:

$$\begin{cases} tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \\ idf_i = \lg\left(\frac{D}{\{d:d\in t_i\}}\right) \end{cases} \tag{2}$$

In the above formula, $n_{i,j}$ is the number of times that the word appears in the human resource big data file d_j, and the denominator is the sum of the number of times that all the words appear in the file jd. $|D|$ is the total number of files in the corpus, and $\{d : d \in t_i\}$ is the number of files containing the word t_i. Word weight is used to evaluate the importance of a word in an article. In TF-IDF feature extraction algorithm, word weight is calculated by using the product of tf and idf. As shown in Formula 3.

$$Weight_i = tf_i \times idf_i \tag{3}$$

Taking into account the differences of the domains in the document, the vector normalization process is performed to convert the feature vector into a unit vector, which is convenient for the subsequent similarity calculation.

4.5 Realize the Parallel Classification Function of Human Resource Big Data

The collected human resource big data features are expressed as a central vector, which is obtained by the arithmetic average method or the weighted arithmetic average method. In the process of classification, no training is needed. Given an unknown class of text, the distance between its center vector and the center vector of each known class is calculated. Calculate the center vector of each category of text, the center vector calculation method is generally used to calculate the arithmetic average [10–12]. After the arrival of new text, the center vector is calculated by the same method. Before calculating the center vector, the text preprocessing such as word segmentation and removing stop words is

needed. The similarity between the new text and each class is calculated by the following formula:

$$Sim(d_i, d_j) = \frac{\sum_{k=1}^{m} w_{ik} \times w_{jk}}{\sqrt{\left(\sum_{k=1}^{m} w_{ik}^2\right)\left(\sum_{k=1}^{m} w_{jk}^2\right)}} \tag{4}$$

Where d_j is the center vector of class j, d_i is the new text feature vector, and the dimension of the feature vector is. The similarity between the new text and each known category is calculated, and the category with the highest similarity is regarded as the classification result of the new text.

5 System Test

The design of human resource big data parallel classification system based on social information cognitive model is to provide auxiliary tools for human resource management of enterprises. In the process of system testing, we need to follow the following aspects. The system is an enterprise human resource system. Testers should include developers and users of the system, so that errors can be easily detected and corrected in time. When testing the human resource management system, the probability of finding errors is proportional. In the process of system development, it is inevitable that there will be errors, so we need to test the system. In addition to developers, system testers also need users to test. When designing test cases, we should not only design effective test cases, but also set unreasonable data for testing.

5.1 System Test Environment

According to the basic requirements of developing software information system, the minimum configuration requirements of test environment are determined. Two IBM high-performance servers, application server and database server, need to be installed on the server side. The operating system and database version are Windows Server 2003 and SQL Server 2005 respectively. The hardware configuration of the client part is p43.0, 4G or more memory, 160 g hard disk, and the operating system and communication network are Windows XP and 100M Ethernet LAN respectively.

5.2 Describe the Test Data Set

The data set used for training in the system test is to use the web crawler to crawl the news information in the human resource management website which has been manually edited and labeled. And through the system unstructured simulation preprocessed data set. It includes six categories: basic information of employees, salary, social insurance, enterprise position, labor cost and training work. The size of human resource data of each category is 5 GB. The documents in each category are manually marked to indicate their category. The data set used for big data classification test is the classification

document downloaded from the laboratory. The data comes from the news corpus edited and manually sorted by all human resource management websites in a certain region, corresponding to the six classifications in the training document.

5.3 System Test Process

First of all, the functional test of the system should make the test user, and then test the function of the system, and use the actual user to log in and operate the test. The performance test of the system is to quickly complete the system function when the system is used. System testing is a link before the system test run. System testing is also a relatively important part of the system development process. System testing can ensure the quality of the system. The main methods of system testing include black box testing and white box testing. The system is carried out under the enterprise human resource management system, using black box testing methods. Through the black box test, a black box can be regarded as a running program after it is opened, and the test can be carried out without considering the internal environment. Through the operation of relevant hardware equipment, the main operation interface of the human resource big data parallel classification system is obtained.

In order to form the experimental comparison, the big data system based on fractional derivative, the big data system based on firefly and simulated annealing were used as the experimental control group. The experimental results are compared with those of the system designed in this study. Among them, the traditional human resource data classification system mainly uses the serial classification method to classify and store the input human resource data one by one. The classification system based on multi-classifiers uses multi-classifier equipment on the basis of traditional systems to achieve parallel processing of human resource data to a certain extent.

5.4 System Test Results

This system test experiment mainly conducts specific tests from two aspects: classification function and system performance. The evaluation of the classification results can reflect the quality of the classification algorithm from one side, and at the same time reflect the overall performance level of the classification system. Among them, the system function test evaluates the quality of the system text classification by calculating the recall rate and precision rate. The recall rate refers to the ratio of the number of samples correctly judged by the classifier as the class to the total number of samples belonging to the class. Accuracy refers to the proportion of samples that the classifier judges to be of this class, which actually belong to the class. Calculated as follows:

$$\begin{cases} recall = \frac{A}{A+C} \\ precision = \frac{A}{A+B} \end{cases} \tag{5}$$

In the formula, A represents the number of documents that belong to this category in the original manual classification standard and are also classified into this category by the classifier. B represents the number of data that did not belong to this category in the

original manual classification standard but were classified into this category by the classifier. C represents the number of data that should have been classified into this category but were classified into other categories. After the operation of the three classification systems, the quantitative test results are obtained and substituted into formula 5 to get the test results that can reflect the classification function of the system, as shown in Table 3.

Table 3. System classification function test results

Experiment number		1	2	3	4	5
Big data system based on fractional derivatives	A	4.47	4.54	4.62	4.48	4.55
	B	0.38	0.25	0.27	0.28	0.24
	C	0.15	0.21	0.11	0.24	0.21
Big data system based on firefly and simulated annealing	A	4.82	4.77	4.83	4.85	4.79
	B	0.09	0.11	0.07	0.11	0.12
	C	0.09	0.12	0.1	0.04	0.09
Design of human resource big data parallel classification system	A	4.97	4.96	4.94	4.98	4.95
	B	0.02	0.02	0.03	0.01	0.02
	C	0.01	0.02	0.03	0.01	0.03

Through the statistics of multiple experimental data, it is found that the classification recall rate and accuracy rate of big data system based on fractional derivatives are 94.1% and 96.1%, respectively. The recall and accuracy of the big data system based on firefly and simulated annealing are 98.0% and 98.2% respectively, while the average recall and accuracy of the designed classification system are 99.6%. In addition, the statistical results of system classification time are shown in Table 4.

Table 4. System performance test results

Experiment number	Classification time of big data system based on fractional derivatives/s	Classification time of big data system based on firefly and simulated annealing/s	Design of human resource big data parallel classification system classification time/s
1	2.33	1.57	0.18
2	2.35	1.63	0.53
3	2.17	1.49	0.49
4	2.28	1.52	0.32
5	2.36	1.55	0.26

According to Table 4, the average classification time of the three data classification systems is 2.3 s, 1.6 s and 0.4 s respectively. To sum up, the classification function and operation performance of the designed human resource big data parallel classification system based on social information cognitive model are higher than the two comparison systems, that is, the designed classification system has more application advantages.

6 Conclusion

The human resource big data parallel classification system based on social information cognitive model is easy to operate. Data processing is simple and effective, and its association rules provide an important scientific method for students to understand the development of different echelons of human resources, establish a scientific talent evaluation mechanism, and clarify the age structure of the teaching staff. Of course, the application of data mining technology in the enterprise human resource management system needs to be improved, such as optimizing the algorithm of data mining, data mining from different angles such as multi-layer and multi-dimensional, so as to develop and optimize the allocation of social and enterprise human resources more reasonably and scientifically.

References

1. Ramadhany, S.E., Azwir, H.H.: Developing database system for managing medical insurance claims in the human resource department. PERFORMA Media Ilmiah Teknik Industri **19**(2), 213–230 (2020)
2. Gurmu, A.T., Ongkowijoyo, C.S.: Predicting construction labor productivity based on implementation levels of human resource management practices. J. Constr. Eng. Manage. **146**(3), 04019115.1–04019115.13 (2020)
3. Fu, W., Liu, S., Srivastava, G.: Optimization of big data scheduling in social networks. Entropy **21**(9), 902 (2019)
4. Hapsari, D.P., Utoyo, I., Purnami, S.W.: Unconstrained optimization based fractional order derivative for data classification. J. Phys. Conf. Ser. **1613**(1), 012066 (2020)
5. Gayathri Devi, S., Sabrigiriraj, M.: A hybrid multi-objective firefly and simulated annealing based algorithm for big data classification. Concurrency Comput. Pract. Exp. **31**(14), e4985.1–e4985.12 (2019)
6. Yuan, Y.: Research on the reconstruction of human resource management based on computer big data and artificial intelligence technology. J. Phys. Conf. Ser. **1574**(1), 012075 (2020)
7. Bragina, D., Molodchik, N.: Big data in human resource management. Manage. Pers. Intellect. Res. Russ. **9**(3), 76–80 (2020)
8. Novianti, W., Qadri, F.: Design of information system human resource management. IOP Conf. Ser. Mater. Sci. Eng. **879**, 012126 (2020)
9. Liu, S., Li, Z., Zhang, Y., Cheng, X.: Introduction of key problems in long-distance learning and training. Mob. Netw. Appl. **24**(1), 1–4 (2019). https://doi.org/10.1007/s11036-018-1136-6
10. Devarriya, D., Gulati, C., Mansharamani, V., Sakalle, A., Bhardwaj, A.: Unbalanced breast cancer data classification using novel fitness functions in genetic programming. Exp. Syst. Appl. **140**, 112866 (2020). https://doi.org/10.1016/j.eswa.2019.112866

11. Mousa, S.K., Othman, M.: The impact of green human resource management practices on sustainable performance in healthcare organisations: a conceptual framework. J. Cleaner Prod. **243**, 118595 (2020)
12. Liu, S., Bai, W., Zeng, N., et al.: A fast fractal based compression for MRI images. IEEE Access **7**, 62412–62420 (2019)