

Design and Implementation of Massive Information Management System Based on Big Data

Ying Li^{a,*}, Mingchen Shao^b

Admission and Employment Office, Zaozhuang Vocational College of Science & Technology, Tengzhou, Shandong, China

^aly0013@126.com, ^bSmc188@188.com

*corresponding author

Keywords: big data; massive information management; system design; data classification

Abstract: In order to improve the artificial intelligence of mass information management system and carry out the optimization design of mass information management system, a design method of mass information management system is proposed based on big data analysis and Web. The function module of the system includes computer aided management module, big data information processing module, program loading module, data classification module and data output module. The fuzzy clustering method is used to design big data's classification algorithm, and the association rule mining method is used to mine big data and integrate scheduling. The big data processing algorithm is used to load the control instructions through the program loading module, and the software development and design of the mass information management system is realized under the embedded Web environment. The simulation results show that the design of mass information management system has a good mining ability of big data and the intelligence of big data information management has been improved

1. Introduction

With the development and popularization of modern network information technology and digital technology, big data information processing technology is applied in the mass information management system structure. The great investment and extensive application of modern science and technology have gradually become the development trend of modern mass information management. It is very important to fully apply modern network information technology and perfect the mass information management system in the current information management development. With the acceleration of information technology, information technology will be the core technology to promote economic development and social progress^[1]. More and more enterprise managers realize the importance of information management, it is necessary for enterprises to obtain competitive advantage to set up enterprise management information system by using advanced information technology. It has great significance to study the optimal design method of mass information management system, so as to optimize information management and improve big data's ability of information scheduling^[2].

On the basis of database design, the design of mass information management system is taken based on database design. The widely used databases are SQL Server, Oracle and Access. We can determine which database to use according to the function and use scale of the system. The wide application of Oracle database is high security, good stability and cross platform performance. Strong, Oracle database is widely used in most enterprises. The application structure of Oracle database includes C/S structure, B/S structure and distributed system structure, which can meet the requirements of different architecture, and the security performance is high^[3]. The Oracle database is the SQL Server database with high market share in the current data base. Providing direct customer development tools and platforms, providing only the two interfaces of ODBC and DB, data storage and security performance. It is a medium-sized database based on the server side. It can be suitable for the application of large capacity data. In the traditional method, the design of the

embedded framework is used for the design of the mass information management system and the number of the combination. According to the database management technology of database and Access, the development and design of mass information management system are carried out, and some intelligent design results have been achieved. In reference [4], a design method of mass information management system is proposed based on B/S mode and SOA architecture. The embedded ARM addressing technology is used for computer aided information management, and large data information is used. Degree technology improves the intelligence of mass information management. However, as the storage of data is gradually increased, the traditional data storage devices are mainly stored in small computer systems, not only in the storage effect, but also in the quantity of storage, and the intelligence of information management is not good. In this paper, the design method of mass information management system is proposed based on large data analysis and Web. The fuzzy clustering method is used to design the classification algorithm of large data, the association rule mining method is used to carry out large data mining and integrated scheduling. The large data processing algorithm is loaded by the program loading module to implement the control instruction. The software development and design of massive information management system is implemented under the entry Web environment. Finally, simulation experiments are carried out to demonstrate the intelligence of the design system.

2. Overall Design Framework of the System

Firstly, the general design of the mass information management system is analyzed and the function module is analyzed and introduced. The mass information management system is built on the general computer platform, and the mass information management system can manage the mass information on different operating systems^[5]. The adaptive scheduling and data access of mass information management are realized, and the mass information management system is developed in embedded Linux environment. The malloc function is used to apply for the necessary storage space, and the function can be used to input a large amount of information. For the preservation of the program, we can use the knowledge of the linked list in the data structure and insert the student information into the linked list. At the same time, we need to design a new program block about mass information insertion, use pointer knowledge, compare the serial number of new input students, big data analysis and storage function of massive information. In order to make the user better understand their input information, you can design a printf() statement to implement the input information output^[6].

In information browsing module, the browsing of mass information is undoubtedly an output of the whole input information. The most important problem is how to realize the connection between two data links, and build files for users in Web cross-compiling environment. The information sampling module and human-computer interaction platform of the mass information management system are constructed. The system adopts a three-tier architecture to design the software of the mass information management system. Database construction and data information management of massive information management information are carried out in the basic layer^[7]. Mass information management database includes knowledge rule base, mass information management information fusion module, filtering module and association rule mining module. The overall design framework of mass information management system is shown in figure 1.

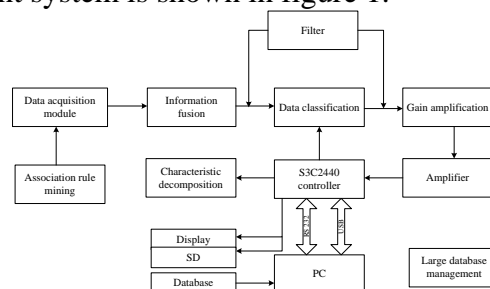


Figure 1. Overall design framework of mass information management system

3. Large Data Information Processing

3.1. Data classification

On the basis of the overall design framework of the system mentioned above, the fuzzy clustering method is used to design big data's classification algorithm. According to the relevant algorithms, the comprehensive trust value is generated and the process of resource data distribution is optimized. Improve UAP cloud computing system to big data processing capacity^[8]. A distributed computing framework is constructed. The distributed computing framework relies on rows, column storage and distributed file storage. The Jaccard coefficients in statistics are often used to calculate the similarity or difference between samples to measure the distance between samples. Sample clustering was determined. For a given set of sample $S = \{S_a, S_b, \dots, S_i, \dots\}$, wherein, $S_a = \{a_1, a_2, \dots\}$, $S_b = \{b_1, b_2, \dots\}$, $S_i = \{i_1, i_2, \dots\}$, the formal definitions are as follows:

Define 1 Similarity, difference formula

1) Single sample-single sample

$$\begin{aligned} \forall (S_i, S_j) \in S : \\ \text{sim}(S_i, S_j) &= \frac{|S_i \cap S_j|}{|S_i \cup S_j|} \\ \text{dis}(S_i, S_j) &= 1 - \text{sim}(S_i, S_j) \end{aligned} \quad (1)$$

2) Single sample-multiple samples

$$\begin{aligned} \forall (S_i, S_{j1}, S_{j2}, \dots) \in S : \\ \text{sim}(S_i, \{S_{j1}, S_{j2}, \dots\}) &= \\ \frac{1}{|\{S_{j1}, S_{j2}, \dots\}|} \sum_{S_j \in \{S_{j1}, S_{j2}, \dots\}} \text{sim}(S_i, S_j) \\ \text{dis}(S_i, \{S_{j1}, S_{j2}, \dots\}) &= 1 - \text{sim}(S_i, \{S_{j1}, S_{j2}, \dots\}) \end{aligned} \quad (2)$$

Define 2 Center points. Given sample clustering feature $C = \{C_1, C_2, \dots, C_i, \dots\}$, where, $\forall C_j \in C \setminus \{C_i\} : \text{dis}(C_i, C \setminus \{C_i\}) < \text{dis}(C_j, C \setminus \{C_j\})$.

Mass information feature selection mainly improves big data classification accuracy rate, so mass information classification accuracy rate (G) as feature optimization, so mass information feature selection problem can be described as:

$$\max_S G(S) \quad (3)$$

The constraints are:

$$\begin{cases} S = \{s_1, s_2, \dots, s_n\} \\ s_i \in \{0, 1\} \\ i = 1, 2, \dots, n \end{cases} \quad (4)$$

Amassive information flow volume model is constructed as:

$$x = \sum_{i=1}^N s_i \Psi_i = \Psi s, \Psi = [\Psi_1, \Psi_2, \dots, \Psi_N] \quad (5)$$

Where, s_i is a resource information request signal. Fuzzy C-means clustering method is used to realize the fuzzy classification and recognition of mass information.

3.2. Association rules mining by big data

The feature selection of mass information is a typical multi-parameter objective optimization problem. It optimizes the feature subset from a set of original features with dimension n, and there are 2^n seed sets^[9]. The set of association rules for mass information is shown in figure 2.

	p_1	p_2	p_4	p_3	p_6	p_5
u_4	1	1	1	0	0	0
u_{12}	1	0	1	1	0	0
u_3	0	0	0	0	1	1
u_7	0	0	0	1	1	0

Figure 2. Association rule vector set of massive information

Fuzzy set is introduced to mining association rules of mass information^[10]. The global updating formula of pheromone of mass information mining is obtained as follows:

$$\begin{cases} \tau_{ij}(n+1) = \rho \cdot \tau_{ij}(n) + \sum_k \Delta \tau_{ij}^k \\ \Delta \tau_{ij}^k = \frac{Q}{F(s_k)} \end{cases} \quad (6)$$

Where, $F(s_k)$ is the fitness value of the feature subset s , introducing quantized weight difference factor α :

$$\alpha = \frac{p(t | C_i) - p(t)}{p(t)} \quad (7)$$

Combined with the average mutual information algorithm, the mining results of massive data association rules are obtained as follows:

$$MI(t) = \sum_{i=1}^r \alpha \times p(C_i) \log \frac{p(t | C_i)}{p(t)} \quad (8)$$

While $p(t | C_i) > p(t)$, $\log \frac{p(t | C_i)}{p(t)} > 0$, $\alpha > 0$, $MI(t, C_i) > 0$, it reflects the degree of positive correlation between feature words and categories. According to the results of association rules mining, big data mining and feature extraction are carried out.

4. Software Development and Design of the System

On the basis of the algorithm design mentioned above, the software development and design of mass information management system are carried out. The functional modules of the system include computer aided management module, big data information processing module, program loading module, etc. Data classification module and data output module. The software development and design of mass information management system is realized in embedded Web environment. Online query and resource scheduling for mass information management in cloud computing environment. Design application layer, application support sublayer, In the network layer, the communication terminal of the mass information management system is based on the medium access layer, combined with the cross-compiler to compile the big data algorithm, and construct the transmission control protocol oriented to the database connection in the transport layer. The information collection and data loading are realized. According to the information fusion processing algorithm in the previous section, the information processing program is loaded by the program loading module, and the knowledge rule base is constructed by using the unified resource locator SIP URI (Unified Resource Identification). The Linux kernel is used to build the GUI of man-machine dialogue module, and the driver is developed in the mode of I/O operation, which realizes the software development and design of mass information management system.

4. Experimental test and analysis

In order to test the application performance of the magnanimous information management system designed in this paper, the experiment test and analysis are carried out, and the embedded ARM core processing program is used to compile and integrate the data of the mass information management system, and the IRcvStran function is called. The test set size of big data is 1000Mbit,

and the bandwidth of data sampling is 10dB. the cross-compiler is designed to load the program of big data information processing algorithm, and the cross-compile interface is shown in figure 3.

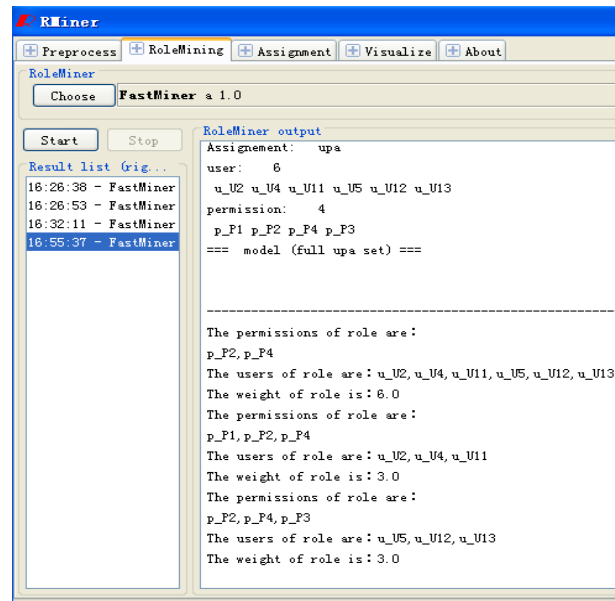


Figure 3. Cross compile interface

The data recall rate of mass information management is tested and the result is shown in figure 4.

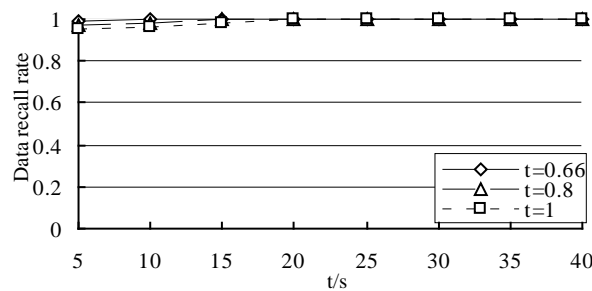


Figure 4. Data recall test

Figure 4 shows that the data recall performance of mass information management based on this method is better. The accurate scheduling rate of mass information management is tested, and the result is shown in figure 5. The analysis figure 5 shows that with the decrease of time window, the accurate scheduling performance of mass information management is improved. It has superior performance of high-volume information management.

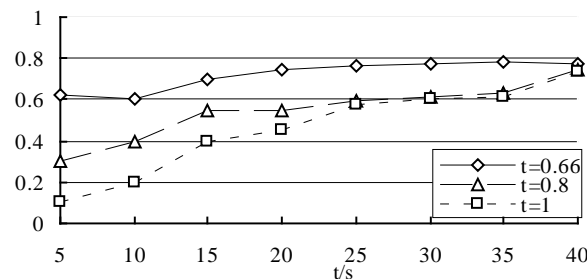


Figure 5. Scheduling performance test

5. Conclusions

In this paper, a design method of mass information management system is proposed based on big data analysis and Web. The function module of the system includes computer aided management

module, big data information processing module, program loading module, data classification module and data output module. The fuzzy clustering method is used to design big data's classification algorithm, and the association rule mining method is used to mine big data and integrate scheduling. The big data processing algorithm is used to load the control instructions through the program loading module, and the software development and design of the mass information management system is realized under the embedded Web environment. The simulation results show that the design of mass information management system has a good mining ability of big data and the intelligence of big data information management has been improved. It has superior performance in massive information management.

References

- [1] Xue S J, Shi W L, Xu X L. A heuristic scheduling algorithm based on PSO in the cloud computing environment [J]. International Journal of u-and e-Service, Science and Technology, 2016, 9(1):349-362.
- [2] Pao W, Lou W, Chen Y, et al. Resource allocation for multiple input multiple output-orthogonal frequency division multiplexing-based space division multiple access systems [J]. IET Communications, 2014, 8(18):3424-3434.
- [3] Arefi M M, Zarei J, Karimi H R. Adaptive output feedback neural network control of uncertain non-affine systems with unknown control direction [J]. J of the Franklin Institute, 2014, 351(8): 4302-4316.
- [4] Xu Y, Tong S, Li Y. Prescribed performance fuzzy adaptive fault-tolerant control of non-linear systems with actuator faults [J]. IET Control Theory and Applications, 2014, 8(6): 420-431.
- [5] Huang X, Wang Z, Li Y, et al. Design of fuzzy state feedback controller for robust stabilization of uncertain fractional-order chaotic systems [J]. Journal of the Franklin Institute, 2015, 351(12): 5480-5493.
- [6] Gu Q, Yuan L, Ning B, et al. A novel classification algorithm for imbalanced datasets based on hybrid resampling strategy [J]. Computer Engineering and Science, 2012, 34(10):128-134.
- [7] Lu Xing-Hua, Chen Pinghua. Traffic Prediction Algorithm in Buffer Based on Recurrence Quantification Union Entropy Feature Reconstruction [J]. Computer Science, 2015, 42(4):68-71.
- [8] MOON J H, PARK J J, KIM D I. Energy signal design and decoding procedure for full-duplex two-way wireless powered relay[C]// URSI Asia-Pacific Radio Science Conference (URSI AP-RASC). 2016: 442-445.
- [9] OKANDEJI A A , KHANDAKER M R A , WONG K K ,et al. Joint transmit power and relay two-way beamforming optimization for energy-harvesting full-duplex communications[C]// IEEE Globecom Workshops (GC Wkshps). 2016: 1-6.
- [10] Zhang R, HO C K. MIMO broadcasting for simultaneous wireless information and power transfer [J]. IEEE Transactions on Wireless Communications, 2013, 12(5): 1989-2001.