Intelligent Integration of Diversified Retirement Information Based on Feature Weighting

Ye Wang(✉) and Yuliang Zhang

Wuxi Vocational Institute of Commerce, Wuxi 214153, China
wanglin45347@163.com

Abstract. When carrying out intelligent integration of retirement information, the existing data processing and analysis architecture can no longer meet the current requirements for the storage and processing of massive text data, which reduces the efficiency and accuracy of intelligent integration of diversified retirement information. Therefore, a feature-weighted intelligent integration method of multi-element retirement information is proposed. Use the packet capture mechanism to collect diversified retirement information samples, and complete the pre-processing of initial retirement information through information filtering, normalization and other steps. Intelligent calculation and distribution of diversified retirement information weights, extraction of mutual information, information gain and other characteristics of diversified retirement information, use of feature weighting algorithm to determine the type of retirement information, and complete the intelligent integration of diversified retirement information. Through the performance test experiment, it is concluded that compared with the traditional integration method, the integrity coefficient of the retirement information integration result obtained by the optimization design method has increased by 2.4%, and the information redundancy coefficient has been effectively controlled. It is applied to the retrieval of retirement information, effectively improving the retrieval speed of information.

Keywords: Feature Weighting · Diversified Information · Retirement Information · Intelligent Integration

1 Introduction

Retirement means that according to the relevant regulations of the state, a worker quits his job due to old age or disability due to work or illness, completely losing the ability to work. On March 12, 2021, in the "14th Five-Year Plan" and the outline of the 2035 long-term goals announced, it is clearly stated that the statutory retirement age should be gradually delayed in accordance with the principles of "small-step adjustment, flexible implementation, classified advancement, and overall consideration". In April 2021, the Ministry of Human Resources and Social Security and the Ministry of Finance issued the "Notice on Adjusting the Basic Pension for Retirees in 2021". The overall adjustment level is 4.5% of the monthly per capita basic pension for retirees in 2020. In order to

B. Wang et al. (Eds.): ICMTEL 2023, LNICST 533, pp. 151–167, 2024.
https://doi.org/10.1007/978-3-031-50574-4_11

improve the efficiency of retirement work, it is necessary to verify the relevant information of retirees in advance [1]. Today's society is in the era of digitization, intelligence, and networking. Information technology is increasingly integrated into people's daily lives. Productivity has undergone a qualitative innovation, realizing the digital storage and management of retirement information. For this purpose, an intelligent integration method for retirement information is designed.

Information integration is to realize the serialization, sharing and transfer of information resources under the leadership of certain organizations according to the development trend of information technology, thus realizing the management process of optimizing the allocation of information resources, broadening the application fields of information resources, and maximizing the value of information. From the perspective of the elements covered by the information system, there are many reasons for information integration, which can be viewed from different perspectives. Under the influence of various advanced technologies, the public's access to retirement information has become more diversified and richer, and the use of retirement information resources has also increased. Due to the complex content and large number of retirement information resources, it is difficult to achieve efficient use of resources in a short period of time. Therefore, only by scientifically integrating retirement information resources, clarifying the subject content and information items, and doing a good job of summarizing and sorting out, can the utilization efficiency of retirement information resources be improved. Using advanced technologies such as big data, Internet of Things, Internet and other advanced technologies and related equipment to carry out information resource integration work, and actively build a retirement information resource with technological attributes as a whole, will help improve the collection, arrangement, protection and utilization of retirement information resources. At the same time, after completing the integration of retirement information resources, we should actively purchase modern equipment, build a retirement information resource library, promote the intelligent management of retirement information, and continuously demonstrate the effect of resource sharing and sharing.

At present, more mature information intelligent integration methods include: information integration method based on support vector machine, information integration method based on wavelet decomposition and information integration method based on deep neural network. Among them, information integration method based on support vector machine takes information as processing object, and is based on VC dimension theory of statistical learning theory and structural risk minimization principle, According to the limited sample information, we should seek the best compromise between the complexity of the model and the learning ability, determine the integration method between data, and complete the task of information integration. The information integration method based on wavelet decomposition decomposes the initially collected information to improve the sufficiency of the signal integration results. In addition, the information integration method based on deep neural network uses deep neural network for feature extraction, and integrates information at the feature level. With the continuous development of information technology, the diversified retirement information data sets have reached TB and PB, or even larger. The existing data processing and analysis architecture can no longer meet the application requirements of the current massive text data storage and processing analysis. The intelligent integration efficiency and accuracy of diversified retirement information are reduced.

Feature weighting refers to the process of assigning a certain weight value to each vector feature in the vector space, and the size of this weight value depends on the ability of feature items to express text. By effectively weighting the feature items, the feature items that can well represent the text are given higher weight values, and the feature items that have less ability to distinguish categories are given smaller weight values. Not only the distribution state of the original feature set in the vector space is effectively improved, but also the influence of those noise feature items can be suppressed well. Feature weighting algorithms include: Boolean weight, entropy weight, word frequency weight, inverse document frequency and TF-IDF function.

Based on the above contents, the intelligent integration method of diversified retirement information based on feature weighting is studied, and the feature weighting technology is applied to the optimization design of the intelligent integration method of diversified retirement information, and the diversified retirement information samples are collected and preprocessed by using the data packet capture mechanism. The type of retirement information is determined by feature weighting algorithm, and the intelligent integration of diversified retirement information is completed. The integrity coefficient of the integration result of retirement information obtained by the design method in this paper is improved by 2.4%, and the information redundancy coefficient is effectively controlled. When it is applied to the retrieval of retirement information, the integration effect of retirement information can be improved, and the retrieval and retrieval performance of retirement information can be effectively improved.

2 Intelligent Integration Method of Diversified Retirement Information

The core processing flow of the optimally designed intelligent integration method of diversified retirement information is shown in Fig. 1.

According to the different stages of information integration and integration methods, the integration level is divided into three levels: information level integration, feature level integration and decision level integration. The information level is to directly process the observation information of diversified retirement information and the observed characteristic values of the target, and then estimate, predict and judge the state of the target according to these characteristic values. As the lowest level integration, information level integration requires that the information to be processed comes from the same sensor. Given the huge amount of information to be processed at the bottom, as well as the characteristics of information discontinuity and jump, information level integration is urgently required to have high error prevention and correction capabilities [2]. At the same time, these harsh conditions will inevitably lead to the decline of the anti-interference ability of the whole system. Therefore, in order to ensure that multiple sensors observe the target at the same time, their processing information needs to be the original observation information. The main advantage of information-level integration is its high integration accuracy. As the lowest level of integration, it provides necessary intermediate integration results for advanced information integration to improve overall

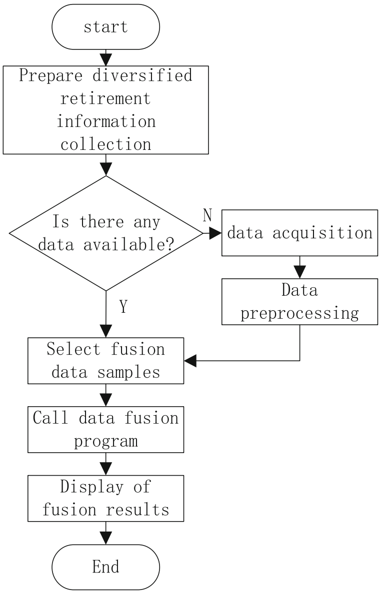


Fig. 1. Flow chart of intelligent integration of diversified retirement information

operational efficiency. Functional level integration is an intermediate level integration. This level of data integration is usually divided into: the integration of the characteristic values of the observed events and the integration of the characteristic states of the observed events. This integration level is in the middle of the integration level. This level of integration requires each sensor in the system to observe a target, and to sort out the corresponding eigenvalues according to the target's own characteristics, and then the eigenvalues are sent to the intermediate link of data integration for further integration. The process also judges the target based on the properties of the target's eigenvalues. Decision level integration requires that the target eigenvalue to be processed must be an intermediate decision result, that is, the processed data in this part is the data that has been properly processed by the sensors in the system. Here, it is necessary to continue to correlate and combine the integration results at a higher level, and finally get the final result of the level integration at this level. Due to the large number of times of consolidation at this level, the amount of indirect result data using the upper level is large and complex, which will lead to low accuracy of the consolidation results. However, this level of integration itself relies on the results of low-level integration as the integration data, reducing the requirements for each other's traffic and the underlying receiver, so as to achieve strong processing capacity and the ability to integrate and adapt to complex environments. Combine the above three levels to complete the integration and processing of diversified retirement information.

2.1 Collect Diversified Retirement Information

In the environment of retirement information registration and management system, data packet capture technology is used to collect diversified retirement information samples. A packet capture mechanism consists of three main parts: the first is the bottom-level packet capture mechanism for a specific operating system, the second is the top-level user program interface, and the last is the packet filtering mechanism. The implementation of the underlying packet capture mechanism may be different due to different operating systems, but the form is not very different. Usually, the transmission path of the data packet is the network card, the device driver layer, the data link layer, the network layer, the transport layer, and finally it is processed by the application program [3]. For the user, the user program only needs to call the corresponding processing function to obtain the required data packets, because the packet capture mechanism provides a unified interface for it, and the captured data packets are processed according to the user's requirements. Filter, and finally pass only the data packets that meet the conditions to the user program. The collection process of diversified retirement information is shown in Fig. 2.

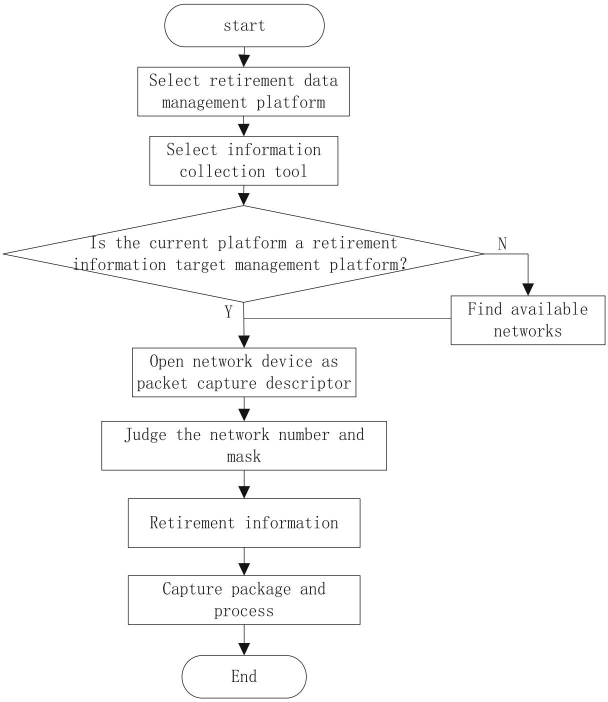


Fig. 2. Flowchart of Diversified Retirement Information Collection

First, specify the network device interface to be captured. In Linux, etho represents the first network card, and lo represents the loopback interface of the host, both of which are optional. Then initialize pcap to tell it which device it is capturing packets from. You can capture packets from multiple devices by using multiple file handles. Next, set and compile the filter rules. If you only want to capture specific transport packets, such as HTTP protocol packets, you must create a rule set, use it to filter out the required packets, and then enter the execution loop of the main program that captures the packets.In this phase, each time pcap receives a packet, it calls another defined function. This function can have any desired function, such as analyzing the data packet and printing out the result to the user, or saving the result as a file. Or do nothing and pcap will work until all the desired packets have been received before stopping. Finally, when the required data is captured, it needs to be closed and returned [4]. Since the information in the retirement information registration and management system will be updated according to the retiree's procedures, it is necessary to update the diversified retirement information in real time during the actual collection process. The update process can be expressed as:

$$x_{\text{retire}}(t) = x_{\text{retire}}(t-1) + x_{\text{add}} - x_{\text{complete}} \tag{1}$$

In the formula, $x_{\text{retire}}(t-1)$ and $x_{\text{retire}}(t)$ represent the information collection samples at the previous time and the information update results at the current time, respectively. x_{add} and x_{complete} correspond to the added retirement information and completed retirement information. Complete the collection of initial diversified retirement information according to the above process.

2.2 Preprocessing of Diversified Retirement Information

In the capture process, in order to ensure that the buffer is not overflowed by invalid data, it is necessary to perform coarse filtering first, then perform another filtering, and finally perform packet analysis. When filtering is complete, packet disassembly must be performed because the header of the packet is the most critical part of the packet. The principle of decomposition is to extract the data of each field in the packet structure according to the data structure specified by the protocol. Use the Berkeley packet filter to filter the initial diversified retirement information data samples. The processing principle is shown in Fig. 3.

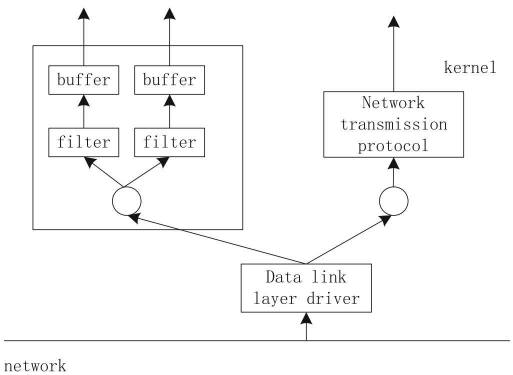


Fig. 3. Schematic diagram of filtering diversified retirement information

After receiving a packet, the network card driver usually submits it to the protocol stack of the system. When a process uses BPF for network monitoring, the driver will first call BPF and copy a copy of data to the BPF filter. The filter will decide whether to receive this packet according to user-defined rules [5]. Then judge whether the data packet is sent to the local machine. If not, the driver returns from the interrupt and continues to receive data; If it is sent to the local computer, the network card driver will submit it to the protocol stack of the system and return it. After the initial information filtering process is completed, the collected retirement information is quantified using the probabilistic reasoning model. The quantitative representation of retirement information is as follows:

$$\chi(\mathrm{T}, \mathrm{I}) = \sum \lg \frac{P_i(1 - Q_i)}{Q_i(1 - P_i)} \tag{2}$$

In the formula, T and I represent the collected retirement information samples and the urgency of retirement, respectively. The calculation formulas of variables P_i and Q_i are as follows:

$$\begin{cases} P_i = \frac{M_i}{M} \\ Q_i = \frac{N_{\text{text},i} - M_i}{N_{\text{text}} - M} \end{cases} \tag{3}$$

In Formula 3, variables M_i and M respectively represent the number of information with characteristic items and high urgency. $N_{\text{text},i}$ and N_{text} correspond to the number of retirement information with characteristic items and the total number of retirement information collected [6]. In addition, we also need to process retirement information by word segmentation and de stop words, and finally put the processing results of retirement information into Formula 4 to achieve the normalization of all diversified information.

$$\chi_{\text{normalization}} = \frac{\chi_{\max} - \chi}{\chi_{\max} - \chi_{\min}} \tag{4}$$

In the above formula, parameters $\chi_{\max}$ and $\chi_{\min}$ are the maximum and minimum values of diversified retirement information respectively, and χ and $\chi_{\text{normalization}}$ are the samples of return information before and after normalization. After the above process, complete the preprocessing of all diversified retirement information collected.

2.3 Intelligent Calculation and Allocation of Diversified Retirement Information Weights

The weight value of diversified retirement information is extracted from three aspects: Boolean weight, word frequency weight, and word frequency inverse document frequency weight. The Boolean weight uses the mutual information relationship between rows and columns as a measure, and is defined as follows:

$$\omega_b(y_j) = P(y_j) \sum_{x_i \in X} \lg \frac{P(x_i y_j)}{P(x_i)} \tag{5}$$

In the formula, $P()$ is the probability solution function, x_i and y_j represent the sample in line i and j of the joint probability matrix respectively, and the solution result $\omega_b(y_j)$

represents the weight value of diversified retirement information χ_i. The word frequency in the word frequency weight represents the number of times the feature item is contained in the text. The word frequency is used to measure the differentiated contribution rate of the feature item to the category [7]. Therefore, if the feature item appears more times in the information set, its word frequency weight will be larger. The calculation formula is as follows:

$$\omega_c = TF_i \tag{6}$$

In the formula, TF_i is the number of target words contained in the retirement information set. In addition, the calculation of the word frequency inverse document frequency weight is divided into two parts, namely the word frequency and the inverse document frequency. The inverse document frequency is a measure of the popularity of a feature word. The number of documents for the item, and then logarithmically obtain the final value [8]. If the feature word only exists in individual information sets, it means that the higher the concentration of the feature word, the higher its contribution rate to the document category. The calculation formula of the word frequency inverse document frequency weight is as follows:

$$\omega_w = TF_i \times \log_2\left(\frac{N_{tui}}{num_i}\right) \tag{7}$$

Among them, N_{tui} and num_i respectively represent the number of training retirement information and the number of information that the feature words appear in the retirement information. For the convenience of calculation, normalization is usually required. The formula for calculating the frequency weight of the word frequency inverse document can be converted into:

$$\omega_w = \frac{TF_i \times \log_2\left(\frac{N_{tui}}{num_i} + 0.01\right)}{\sqrt{\sum\left(TF_i \times \log_2\left(\frac{N_{tui}}{num_i} + 0.01\right)\right)^2}} \tag{8}$$

In the actual intelligent calculation process of the weight of diversified retirement information, the influence of the position and length of the feature word on the value of the weight can be considered, and the weight calculation result can be updated based on the calculation result. The final update calculation result of the weight of the diversified retirement information is as follows:

$$\omega = \alpha \cdot \rho(\omega_b \times \omega_c \times \omega_w) \tag{9}$$

where, α and ρ represent the length and location parameters of the feature words respectively. The calculation results of Formula 5, Formula 6 and Formula 8 are substituted into Formula 9, and the output results are the calculation results of the weight of diversified retirement information. On this basis, use the linear regression theory to allocate the weight of retirement information. This method needs to establish a multiple linear regression model on the training data, minimize the gap between the estimated score of diversified retirement information and the actual correlation score through the least square method, and obtain a set of coefficients of the model as the weight of each member system [9]. In the process of establishing the model, it is required that the training data contain effective information scores and information correlation evaluation. In this linear model, the score provided by the system is the estimated score of the document. According to the given relevance evaluation information, we can accurately know whether a piece of information is relevant to the query. This strategy trains a specific fusion model by linearly combining the estimated scores of information in multiple retrieval systems to ensure the overall relevance of diversified retirement information.

2.4 Extracting Diversified Retirement Information Features

The idea of feature selection is to calculate the evaluation value of all feature words according to the evaluation function. The feature term is filtered by a set threshold, so as to reduce the feature dimension. The extraction process of diversified retirement information features is shown in Fig. 4.

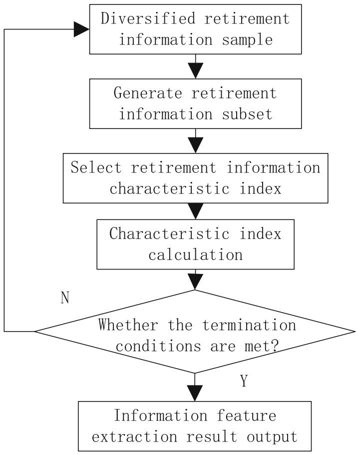


Fig. 4. Flow chart of feature extraction of diversified retirement information

The extracted diversified retirement information features include mutual information, information gain, GINI χ2 Statistics, chi square statistics and other feature vectors, in which the information gain is the proportion of the information of the feature item in the text, and the information gain method is a calculation method based on the feature item in the information entropy. It is defined as the amount of information that a feature item can provide for the whole classification. The calculation formula of information gain is:

$$\begin{cases} \tau_{\mathrm{IG}} = \phi_1 + \phi_2 \\ \phi_1 = \mathrm{P(t)} \sum_{i=1}^{M} P(C_i t) \log \dfrac{P(C_i t)}{P(C_i)} \\ \phi_2 = P(\bar{t}) \sum_{i=1}^{M} P\left(C_i \bar{t}\right) \log \dfrac{P\left(C_i \bar{t}\right)}{p(C_i)} \end{cases} \tag{10}$$

In the formula, t is the feature item, $P(C_i)$ and $P(C_it)$ are the probability value of the occurrence of category i and the probability that the text belongs to C_i when t appears in the text. The $\chi 2$ statistic is a relatively common feature selection method. It mainly selects those items that are strongly associated with each class. The extraction results of the $\chi 2$ statistic feature are:

$$\tau_{\chi^2} = \frac{N_{tui} \times (\lambda_A \lambda_D - \lambda_C \lambda_B)^2}{(\lambda_A + \lambda_C) \times (\lambda_B + \lambda_D) \times (\lambda_A + \lambda_B) \times (\lambda_C + \lambda_D)} \tag{11}$$

In the formula, λ_A, λ_B, λ_C and λ_D respectively represent the number of occurrences of item t and category c in the retirement information set, the number of occurrences of item t but category c in the retirement information set, the number of occurrences of item t but category c in the retirement information set, and the number of occurrences of both item t and category c. If t and c are independent, the value of the $\chi 2$ statistic feature will be 0. For each class in the labeled dataset, the value of the $\chi 2$ statistic feature between each item and that class is calculated. In addition, the GINI index is a method of measuring impurity. Its main idea is to use all attribute values in the N records as candidate division points, and then calculate the GINI index of each candidate, that is, the best division point corresponds to Generates the smallest GINI indicator value point [10]. The original intention of the GINI indicator is to measure the impurity of an attribute in text classification. The smaller the impurity, the better the attribute. Then the extraction result of the GINI feature index is:

$$\tau_{Gini} = 1 - \sum [P(i|t)]^2 \tag{12}$$

The extraction results of diversified retirement information features such as chi square statistics and mutual information are obtained according to the above methods.

2.5 Determine Retirement Information Types Using Feature Weighting Algorithm

A feature weighting algorithm is used to weight the calculated retirement information weight and the extracted retirement information features, and with the support of the information classifier, determine the type of retirement information. The execution principle of the feature weighting algorithm is as follows:

$$\gamma = \tau_{com} \cdot \omega \tag{13}$$

In the formula, τ_{com} is the fusion result of the extracted features of diversified retirement information, and ω is the weight of retirement information. The specific value of this variable can be determined according to Formula 9. Input the retirement information feature weighting result γ into the classifier. Randomly select any information sample in the retirement information as the information classification center, and use Formula 14 to measure the similarity between the random retirement information and the information sample.

$$s = \frac{\sum_{k=1}^{n} \left(\gamma_{ik} \times \gamma_{jk}\right)}{\sqrt{\sum_{k=1}^{n} \gamma_{ik}^{2} \times \sum_{k=1}^{n} \gamma_{jk}^{2}}} \tag{14}$$

Substitute the calculation result of formula 13 into formula 14 to obtain the measurement result of similarity. The retirement information whose similarity measure result is higher than the threshold s_0 is divided into the same category as the classification center. According to the above process, the classification and processing of all the collected diversified retirement information is realized.

2.6 Realize the Intelligent Integration of Diversified Retirement Information

According to the feature extraction results and weight assignment results of diversified retirement information, the data objects form a cluster tree, and each node is a cluster formed by data objects. Hierarchical clustering methods can be divided into cohesive and split ones according to whether the clustering tree is formed from the bottom up or from the top down. Each data object is regarded as a cluster at first, and then similar clusters are merged successively until all data objects are in a cluster or a certain termination condition is met. Each data object in the diversified retirement information set is regarded as a cluster, and the similarity between any two clusters is calculated using formula 14. Select the two clusters with the largest similarity, and combine the two clusters into a new cluster. The update result can be expressed as:

$$Z_{\text{update}} = Z_i \cup Z_j \tag{15}$$

The above operations are performed in a loop until merged into a cluster or a certain termination condition is met. Divide hierarchical clustering method: It is the opposite of agglomerative hierarchical clustering method, which starts by placing all data objects in a cluster, and then gradually splits the clusters into smaller clusters until each data object becomes a separate cluster or reach a certain termination condition. Finally, the intelligently integrated diversified retirement information is output in the form of directory topology to complete the intelligent integration of diversified retirement information.

3 Experimental Analysis of Information Integration Performance Test

In order to test the advantages of the intelligent integration method of diversified retirement information based on feature weighting of optimal design in terms of integration performance, experiments are set up in the way of comparative testing, and the results of information integration are applied to the search of retirement information. By comparing with traditional methods, the performance advantages of the optimal design method are reflected.

3.1 Prepare a Sample of Retirement Information

This experiment involves multiple data sets, which are the information of retirees in different regions. The data sets are all from the regional personnel resource management system. The preparation of some data sets is shown in Table 1.

Table 1. Retirement information sample preparation

Retirement Area	Number of retirees	Retirement Information Proficiency/GB
A1	365	15.6
A2	577	27.2
A3	283	13.8
A4	496	26.5
A5	434	25.8
A6	502	26.4
A7	378	18.9

The retirement information initially prepared includes the name, age, length of service, social security payment, retirement time, retirement amount, etc. of retirees. The corresponding information is extracted in the management system environment, and the necessary preprocessing process is carried out. Only the plain text information in the web page is extracted, and the anchor text information that will cause noise is deleted. Chinese documents are converted to BIG5 encoding format, removing stop words, words with only one word and words with a frequency of less than 4.

3.2 Configure the Experimental Environment

The experimental program uses the Visual Studio2005 development environment and uses the C# language to write. The test machine is a Dell server with a Xeon® CPU E5405 @2.00 GHz and 4 GB of RAM. The operating system uses Windows Server 2003.

3.3 Describe the Experimental Process of Integrated Performance Testing

In the configured experimental environment, realize the conversion between the intelligent integration method of diversified retirement information based on feature weighting and the running program, input the prepared diversified retirement information sample into the information integration program, and obtain the intelligent integration result of retirement information, as shown in Fig. 5.

In order to reflect the performance advantages of the optimization design method, the traditional information integration method based on support vector machine and the information integration method based on wavelet decomposition are set as the experimental comparison methods in the experiment, and the integration method is developed

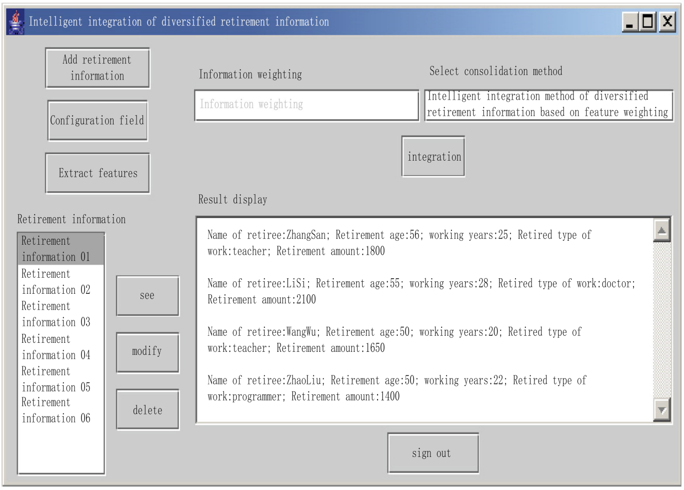


Fig. 5. Results of intelligent integration of diversified retirement information

in the same experimental environment. The same information samples are processed, and the output integration result of retirement information is shown in Fig. 6.

According to the above methods, the development of intelligent integration methods for diversified retirement information will be realized, and the integrated data will be extracted and counted.

3.4 Set the Information Integration Performance Test Indicators

In order to verify the integration performance and application performance of the intelligent integration method of diversified retirement information based on feature weighting, the integration data integrity coefficient, redundancy coefficient and retrieval delay of retirement information are set as the quantitative test indicators of the experiment. The test results of the integrity coefficient are as follows:

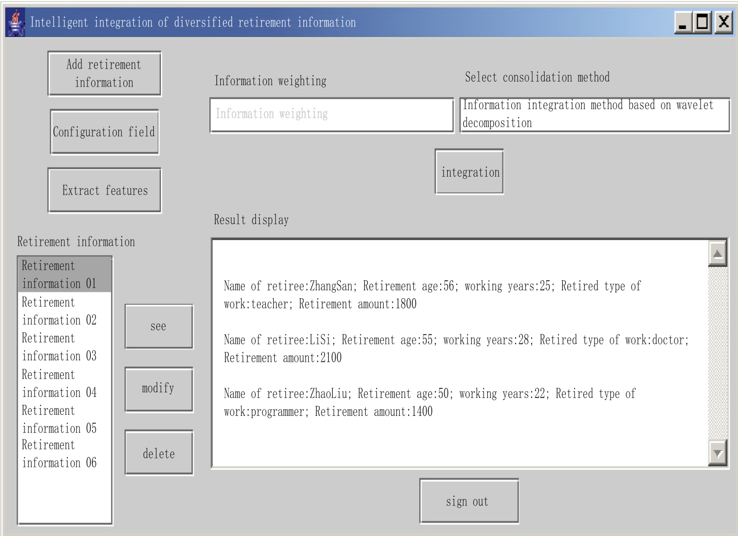
$$\zeta_{\text{complete}} = \frac{N_{\text{integration}}}{N_{\text{sample}}} \times 100\% \tag{16}$$

In the formula, $N_{\text{integration}}$ and N_{sample} represent the integrated retirement information and prepared retirement information respectively. In addition, the numerical results of redundancy coefficient are:

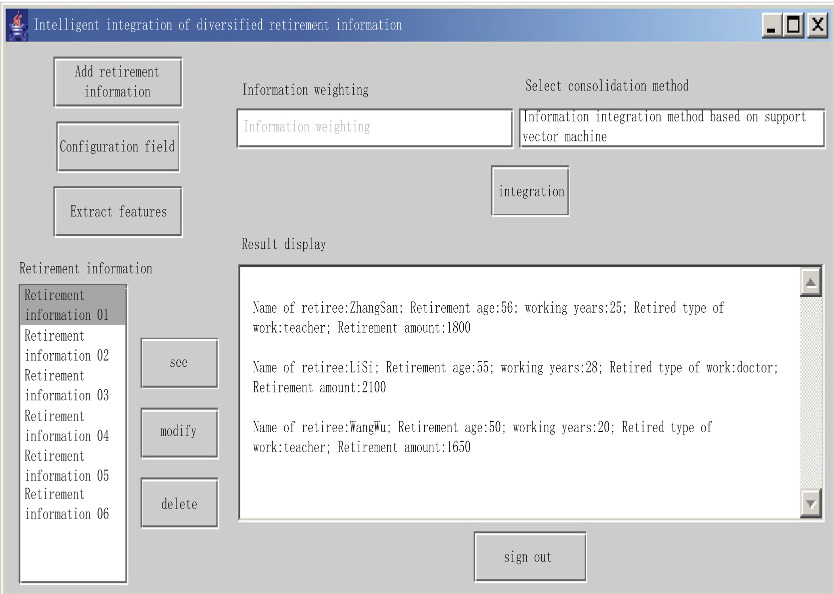
$$\zeta_{\text{redundancy}} = \frac{N_{\text{redundancy}}}{2N_{\text{sample}}} \times 100\% \tag{17}$$

The variable $N_{\text{redundancy}}$ is the amount of redundant information in the retirement information integration result. In addition, the retrieval delay of retirement information is used to reflect the application performance of the retirement information integration method. The test results of this indicator can be expressed as:

$$\Delta t_{\text{retrieval}} = t_{\text{input}} - t_{out} \tag{18}$$



(a) Information integration results based on support vector machines



(b) Information integration results based on wavelet decomposition

Fig. 6. Results of intelligent integration of diversified retirement information

In Formula 18, t_{input} and t_{out} correspond to the start time and output time of the retirement information retrieval task. The final calculation shows that the greater the integrity coefficient of the integrated data and the smaller the redundancy coefficient, the better the intelligent integration performance of diversified retirement information, the smaller the retrieval delay of retirement information, and the faster the retrieval speed, the better the application performance of the intelligent integration method of diversified retirement information.

3.5 Analysis of Information Integration Performance Test Results

Through the statistics and analysis of relevant data, the test comparison results of the integrated data integrity coefficient obtained by the intelligent integration method of diversified retirement information are obtained, as shown in Table 2.

Table 2. Retirement information integration complete coefficient test data table

Number of experiments	The output data volume of the information integration method based on support vector machine/GB	Output data volume of information integration method based on wavelet decomposition/GB	Output data volume of intelligent integration method of diversified retirement information based on feature weighting/GB
1	15.2	15.1	15.5
2	26.4	26.6	27.1
3	13.4	13.3	13.6
4	25.7	25.7	26.3
5	25.0	25.1	25.5
6	25.5	25.3	26.2
7	18.1	18.2	18.8

By substituting the data in Table 2 into formula 16, the average integrity coefficient of the two traditional integration methods is 96.8%, and the average integrity coefficient of the optimal design method is 99.2%. In addition, through the calculation of formula 17, the test results of the intelligent integration redundancy coefficient of retirement information are obtained, as shown in Fig. 7.

It can be seen intuitively from Fig. 7 that compared with the traditional integration method, the retirement information integration method obtained by the optimal design method has a lower information redundancy coefficient, that is, the integration performance is better. Applying the intelligent integration results of diversified retirement information to the retrieval of retirement information, through the calculation of formula 18, the test results reflecting the application performance of the integration method are obtained, as shown in Fig. 8.

It can be seen from Fig. 8 that the intelligent integration method of diversified retirement information based on feature weighting optimized design has shorter retrieval delay and faster retrieval speed.

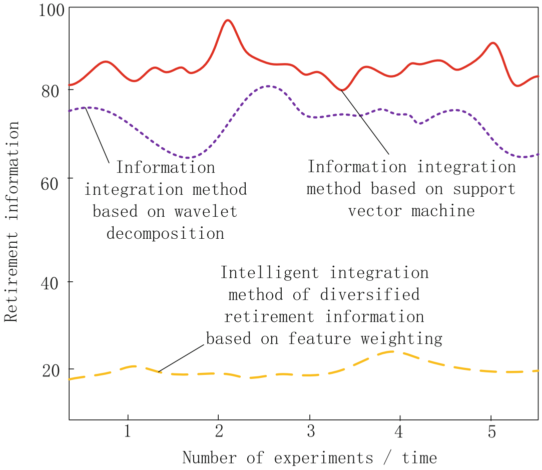


Fig. 7. The test results of the redundancy coefficient of intelligent integration of retirement information

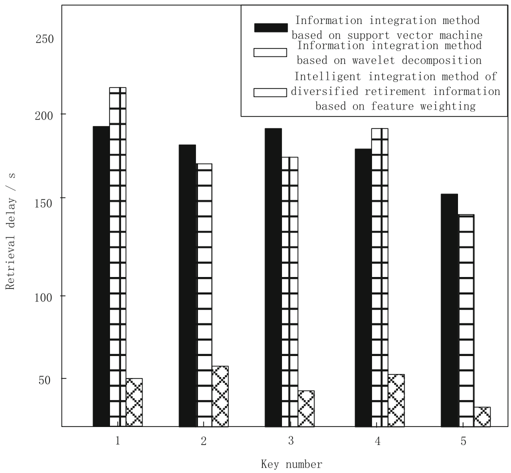


Fig. 8. Application performance test results of retirement information integration method

4 Conclusion

Due to the increase of retirees and the gradual increase of retirement information resources, the traditional information management and integration work can no longer meet the needs of retirement information. The application of feature weighting technology has changed the integration method of diversified retirement information, and created a better sharing platform for it. The experimental results show that the integrity coefficient of the retirement information integration result obtained by this design method is improved by 2.4%, the information redundancy coefficient is effectively controlled, and the information retrieval speed is effectively improved.

References

1. Cai, Y., Li, Z., Yu, X.: The model design of government affairs archives integration and utilization based on enterprise user portrait. Archives Sci. Study **2**, 125–131 (2021)
2. Wan, R., Li, T.: Research on the integration technology of government information resources in standard geographical name and address. Bull. Sur. Mapp. **4,** 136–140 (2021)
3. Huang, S.: Multimedia network information integration and management technology in the era of big data. J. Beijing Inst. Graph. Commun. **29**(8), 146–148 (2021)
4. Li, L., Dong, J., Zuo, D., et al.: SLA-aware and energy-efficient VM consolidation in cloud data centers using host state 3rd-Order Markov chain model. Chin. J. Electron. **29**(6), 1207–1217 (2020)
5. Aoih, A., Dp, B.: Why and when does financial information affect retirement planning intentions and which consumers are more likely to act on them? - ScienceDirect. J. Bus. Res. **117**, 411–431 (2020)
6. Palmer, J.C., Chung, Y., Park, Y., et al.: Affectivity and riskiness of retirement investment decisions. Pers. Rev. **49**(9), 2093–2110 (2020)
7. Srivastava, G., Lin, C.W., Pirouz, M., et al.: A pre-large weighted-fusion system of sensed high-utility patterns. IEEE Sensors J. **PP**(99), 1–1 (2020)
8. Huang, G., Qu, H.: Data visualization and data fusion on the visual performance of illustration. J. Intell. Fuzzy Syst. **39**(6), 8795–8803 (2020)
9. Bahrami, S., Dornaika, F., Bosaghzadeh, A.: Joint auto-weighted graph fusion and scalable semi-supervised learning. Inform. Fusion **66**(1), 213–228 (2021)
10. Yin, C., Wang, P., Wang, C., et al.: Multi area resource information search integration simulation for industrial product design. Comput. Simul. **36**(6), 438–441,454 (2019)