Research on Data Mining Algorithm Based on Cloud Platform

Yadong Zhu
Nanjing Engineering Branch, Jiangsu Union Technical Institute, Nanjing, 210019, China

Abstract: The control and management of plenty of traditional home furnishing take a long time. For achieving the remote control and uniform management of home equipment, according to the characteristics of complex network, intelligence home furnishing data mining should be systematically studied. Therefore, a intelligence home furnishing data mining method based on K-means clustering of genetic algorithm is proposed. The method is utilized to collect data from cloud platform, and extract features combined with the collected data. Then, K-means clustering method of genetic algorithm method is adopted to cluster extracted intelligence home furnishing data to complete the research on the data mining algorithm of cloud platform. Experimental results show that this method can effectively reduce the time of data processing.

Keywords: Cloud platform; Data acquisition; Data feature extraction; Data mining

1. Introduction

With the rapid development of Internet, intelligence home furnishing data network has become an integral part of people's life [1]. How to dig out the valuable information from complex intelligence home furnishing data has become a new challenge. In order to promote the development of the company, it is necessary to dig out useful intelligence home furnishing data from the intelligence home furnishing data set, otherwise, massive intelligence home furnishing data becomes the burden of the company [2-3]. Faced with the challenges of rich intelligence home furnishing data and lack of knowledge, it is urgent to use intelligence home furnishing data analysis tools to mine useful knowledge, relationships and rules from massive intelligence home furnishing data sets, thus bringing greater value for development [4]. Intelligence home furnishing data mining technology is of great significance to the development of all walks of life. Intelligence home furnishing data mining refers to acquisition of interesting knowledge from intelligence home furnishing data, which is effective, novel and potentially useful, and demonstration for users through an understanding model [5-6]. Intelligence home furnishing data mining technology is not only including retrieval faced to some specific intelligence home furnishing databases, but also intelligence home furnishing data statistics, analysis and reasoning, and offering the solution for solving the existing problems [7-8]. intelligence home furnishing data mining algorithms mainly include classification, clustering, association rules and prediction, etc. [9]. Classification refers to extraction a model from a intelligence home furnishing data set, and the model is used to partition intelligence home furnishing data into different types [10]. Clustering refers to the division of intelligence home furnishing data based on the similarity of intelligence home furnishing data in the case of unclear classification [11-12].

A network intelligence home furnishing data mining algorithm based on association principle is commonly used [13]. The experimental results show that the network intelligence home furnishing data mining algorithm based on association principle can effectively improve the accuracy of intelligence home furnishing data mining and reduce memory usage, but there is a problem of slow running speed. Hu et al. put forward a method of intelligence home furnishing data mining research based on intelligence home furnishing data density increment space in the network [14]. This method carries on the analysis to the intelligence home furnishing data density in the network, and has high accuracy of intelligence home furnishing data mining, but intelligence home furnishing data processing takes a lot of time. Besides, a intelligence home furnishing data mining method based on improved ant colony algorithm is described [15]. The ant colony algorithm development process as well as the research situation are described in the paper, and analysis of the parameters of ant colony algorithm is helpful to the analysis of experimental results, and improves the accuracy of intelligence home furnishing data mining. But this method is unable to deal with individual intelligence home furnishing data.

To solve the above problems, a intelligence home furnishing data mining method based on K-means clustering of genetic algorithm is proposed. Experimental results show that the proposed method can effectively improve the accuracy of intelligence home furnishing data mining and reduce the time for intelligence home furnishing data processing.

2. Intelligence Home Furnishing Data Mining Based on Cloud Platform

The intelligence home furnishing data of cloud platform is collected using the method of movable node, and the collected intelligence home furnishing data is employed to extract features via kernel principal component analysis method. Afterwards, extracted features are clustered with K-means clustering method of genetic algorithm. The intelligence home furnishing data clustering results completes the research of the intelligence home furnishing data mining algorithm for cloud platform. The concrete steps are as follows:

2.1. Intelligence Home Furnishing Data Acquisition based on Movable Nodes

The operation range of movable node can be regarded as a disk and the intelligence home furnishing data acquisition node is the center. The first task is to find the coordinate set, and the following conditions need to be fulfilled:

(1) The number of intelligence home furnishing data coordinates is as small as possible;

(2) The disk which has radius r covers the area.

When intelligence home furnishing data disks are arranged, in order to avoid intelligence home furnishing data gaps, the adjacent intelligence home furnishing data need to be overlapped. In order to reduce the number of intelligence home furnishing data disks, the smaller the overlap of the intelligence home furnishing data disks is better. The theorem can be used to illustrate.

Theorem 1 when overlapping intelligence home furnishing data disks form triangles, the area of overlap is minimum.

Proof: assume that overlap areas of intelligence home furnishing data disks are S_{AB}, S_{AC}, S_{BC}, and the overlap area S is:

$$S = S_{AB} + S_{AC} + S_{BC}$$
$$= \frac{1}{2} r^2 (a - \sin a) + \frac{1}{2} r^2 (g - \sin g) + \frac{1}{2} r^2 (b - \sin b) \quad (1)$$

In equation (1), a represents the length of the intelligence home furnishing data disk, g represents the height of the intelligence home furnishing data disk, and b represents the width of the intelligence home furnishing data disk.

Since the intelligence home furnishing data disks overlap to form polygons, thus:

$$za + (p - a) + 2b + (p - b) + 2g + (p - g) = 4p \quad (2)$$

It can be concluded that:

$$A + b + g = p \quad (3)$$

The intelligence home furnishing data extreme value of formula (1) while satisfying the formula (2) need to be obtained. According to the multiplier method, the minimum value of S can be found when $a = b = g = p/3$. At this point, 3 data disks form the equilateral triangle of data, and the length is $\sqrt{3}r$.

The position of intelligence home furnishing data disk satisfying the theorem 1 has a plurality of sets, wherein one set of intelligence home furnishing data is:

$$\left\{\left(\left(m_1 + \frac{1}{2}\right)\sqrt{3r}, \left(3m_2 + \frac{1}{2}\right)r\right)\right\} \cup \left\{\left(m_3\sqrt{3r}, (3m_4 + 2)r\right)\right\} \quad (4)$$

In formula (4), m_1, m_2, m_3, and m_4 represent nonnegative integers.

After the intelligence home furnishing data disk is determined, a intelligence home furnishing data acquisition point is selected in the intelligence home furnishing data disk. When determining the location, the energy consumption of intelligence home furnishing data is taken into account.

When a node transmits intelligence home furnishing data, it is represented by formula (5):

$$P_r = P_t \left(\frac{l}{4p\,d}\right)^2 \frac{G_r G_t}{L} \quad (5)$$

The formula (5), G_r is the gain of intelligence home furnishing data receiver, G_t is the intelligence home furnishing data transmitting gain, P_r is the power of intelligence home furnishing data receiver, P_t is the intelligence home furnishing data transmitting power, d represents the distance between intelligence home furnishing data node and destination node, L denotes the intelligence home furnishing data loss factor, l is the intelligence home furnishing data wavelength. For the specific network, suppression of intelligence home furnishing data node is relatively ideal, formula (5) can be simplified to:

$$P_t = Kd^2 \quad (6)$$

In formula (6), $K = L(4p)^2 P_r / G_r G_t l^2$ is considered a intelligence home furnishing data constant.

If there are n intelligence home furnishing data nodes in the intelligence home furnishing data disk, in the process of intelligence home furnishing data acquisition, the transmitting power of the intelligence home furnishing data node is P:

$$P = \sum_{i=1}^{n} P_t(t) = K \sum_{i=1}^{n} (x_s - x_i)^2 + (y_s - y_i)^2 \quad (7)$$

In formula (7), (x_i, y_i) represents the location of i intelligence home furnishing data nodes, and (x_s, y_s) represents the location of the intelligence home furnishing data acquisition point.

It can be seen from formula (7) that in order to lower the intelligence home furnishing data transmitting power, a reasonable intelligence home furnishing data acquisition point is needed.

Theorem 2 assumes that there are n points in intelligence home furnishing data and the intelligence home furnishing data coordinate is (x_i, y_i), $1 \le i \le n$. The intelligence home furnishing data node is $\left(\frac{1}{n}\Sigma x_i, \frac{1}{n}\Sigma y_i\right)$.

Proof: if the intelligence home furnishing data coordinate of the center is (x_c, y_c), the coordinates of the intelligence home furnishing data points are (x_s, y_s), and the distance between each intelligence home furnishing data node is d_i, then the distance square sum f_s of all intelligence home furnishing data is expressed as:

$$f_s = \sum_{i=1}^{n} d_i = \sum_{i=1}^{n} (x_s - x_i)^2 + (y_s - y_i)^2 \tag{8}$$

Theorem 2 points out the ideal location of intelligence home furnishing data nodes, but some intelligence home furnishing data nodes will not be collected, thus the theorem 3 is brought up.

Theorem 3 assumes that the horizontal and vertical coordinates of maximum and minimum intelligence home furnishing data in the intelligence home furnishing data disk are $x_{\max}, x_{\min}, y_{\max}, y_{\min}$, and (x_s, y_s) represents the average of all intelligence home furnishing data nodes. When the formula (9) is satisfied, the distance of intelligence home furnishing data node (x_s, y_s) is smaller than the radius r.

$$(x_{\max} - x_{\min})^2 + (y_{\max} - y_{\min})^2 \le r^2 \tag{9}$$

Proof: $\begin{cases} x_{\min} \le x_s \le x_{\max} \\ y_{\min} \le y_s \le y_{\max} \end{cases}$ thus, (x_s, y_s) represents the distance of intelligence home furnishing data node.

As explained by theorem 3, when the intelligence home furnishing data node of disk meets the formula (9), the intelligence home furnishing data node is further analyzed. Since the radius of the intelligence home furnishing data disk is r, then:

$$\begin{cases} x_{\max} - r \le x_s \le x_{\min} + r \\ y_{\max} - r \le y_s \le y_{\min} + r \end{cases} \tag{10}$$

That is, the range of (x_s, y_s) does not exceed the rectangular area, assuming that lengths of the rectangle are l_1, l_2. It can be represented as:

$$\begin{cases} l_1 = x_{\min} + r - (x_{\max} - r) = 2r - (x_{\max} - x_{\min}) \\ l_2 = y_{\min} + r - (y_{\max} - r) = 2r - (y_{\max} - y_{\min}) \end{cases} \tag{11}$$

From the formula (11), we can see that when the intelligence home furnishing data nodes are more centralized, l_1 and l_2 are relatively large, and the area S_R of the rectangle is large. Thus, the intelligence home furnishing data collection of the cloud platform is completed.

2.2. Intelligence Home Furnishing Data Feature Extraction Based on Kernel Principal Component Analysis

After the intelligence home furnishing data is collected, the kernel principal component analysis is used to extract the intelligence home furnishing data features. Assuming that $\{x_1, x_2, x_3, \mathbf{L}, x_n\}$ is the original input intelligence home furnishing data, and the kernel space is determined by the kernel function $k(x, y)$. The input intelligence home furnishing data is mapped into the intelligence home furnishing data kernel space, so that the mapping corresponding to the kernel function $k(x, y)$ is f, and the mapped intelligence home furnishing data is $\{f(x_1), f(x_2), \mathbf{L}, f(x_n)\}$. Assuming that the intelligence home furnishing data after mapping satisfies the mean, $\sum_{i=1}^{n} \Phi(x_i) = 0$. The covariance matrix of the intelligence home furnishing data samples is established to obtain the eigenvalues.

The covariance matrix of the intelligence home furnishing data samples is represented as:

$$\mathrm{cov} = \frac{1}{n} \sum_{i=1}^{n} \Phi(x_i) \Phi^T(x_i) \tag{12}$$

In formula (12), Φ represents the sample number of the intelligence home furnishing data, T represents the eigenvalues of the covariance matrix of the intelligence home furnishing data sample, and n represents the eigenvectors of the covariance matrix, and the formula (13) is solved:

$$\mathrm{cov} * b = l b \tag{13}$$

Both two sides of the formula (13) are multiplied by the kernel space sample $\Phi(x_k)$, and the formula (14) is obtained:

$$\Phi(x_k) * \mathrm{cov} * b = l \Phi(x_k) b \quad k = 1, 2, \mathbf{L}, n \tag{14}$$

From the related theory of the reproducing intelligence home furnishing data kernel space, we can see that any vector in the intelligence home furnishing data kernel space can be represented by the intelligence home furnishing data base vector $\Phi(x_1), \Phi(x_1), \mathbf{L}, \Phi(x_n)$, that is, $b = \sum_{i=1}^{n} a_i \Phi(x_i)$, which is put into the formula (15) to get:

$$\Phi(x_k)*\mathrm{cov}*\left(\sum_{i=1}^{n} a_i\Phi(x_i)\right) = l\Phi(x_k)\left(\sum_{i=1}^{n} a_i\Phi(x_i)\right) \quad k = 1,2,\mathbf{L},n \tag{15}$$

Combining with formula (12), the formula (15) is algebraically calculated, and the formula (16) can be acquired:

$$\frac{1}{n}\sum_{i=1}^{n} a_i\Phi(x_k)\cdot\sum_{j=1}^{n}\Phi(x_i)(\Phi(x_i))\cdot l\Phi(x_j) = l\sum_{i=1}^{n} a_i(\Phi(x_i)) \quad k = 1,2,\mathbf{L},n \tag{16}$$

The operations of the spatial points of the intelligence home furnishing data kernel can be computed by the numerical value of the corresponding kernel function.

The intelligence home furnishing data matrix $K = \left[\mathbf{p}\ \Phi(x_i),\Phi(x_j)\ \mathbf{f}\right]$ is defined, and K is substituted into formula (16) to simplify:

$$n l K a = K^2 a \tag{17}$$

Let $l' = nl$, the intelligence home furnishing data in formula (17) can be solved by solving the intelligence home furnishing data eigenvalue in formula (18):

$$Ka = l'a \tag{18}$$

Assuming that the eigenvector corresponding to the maximum eigenvalue of the intelligence home furnishing data is represented as $a_1, a_2, \mathbf{L}, a_m$, whereby the attribute of i -th dimension of the arbitrary intelligence home furnishing data test point $\Phi(x)$ is denoted as:

$$x(i) = \sum_{j=1}^{n} a_k(j)\,\mathbf{p}\ \Phi(x_j),\Phi(x)\ \mathbf{f} \tag{19}$$

With formula (19), we can calculate the result of intelligence home furnishing data dimensionality reduction, which can change the intelligence home furnishing data set matrix and intelligence home furnishing data of input intelligence home furnishing data space into low-dimensional intelligence home furnishing data space.

The original mean of samples is subtracted from each sample, and the intelligence home furnishing data is processed to satisfy the mean value of zero.

$$\bar{\Phi}(x_i) = \Phi(x_i) - \frac{1}{n}\Sigma_j\Phi(x_j) \tag{20}$$

Then the intelligence home furnishing data element (i,j) in the intelligence home furnishing data kernel matrix is changed to:

$$K'(i,j) = \left\langle \bar{\Phi}(x_i),\bar{\Phi}(x_j)\right\rangle = \left\langle \Phi(x_i) - \frac{1}{n}\Sigma_k\Phi(x_k), \Phi(x_j) - \frac{1}{n}\Sigma_k\Phi(x_k)\right\rangle = (K - L_n\cdot K - K\cdot L_n)(i,j) \tag{21}$$

In formula (21), L_n stands for intelligence home furnishing data element.

Assuming that the core matrix formula (22) can be calculated by using m sample intelligence home furnishing data:

$$\bar{K} = \left[\ker nel(x_i, x_j)\right] \tag{22}$$

In formula (22), $\ker nel$ represents the intelligence home furnishing data kernel function, and $\bar{K}$ takes the mean value of the intelligence home furnishing data in the intelligence home furnishing data kernel space as 0. This method is identical to the formula (21). The transformed intelligence home furnishing data mean is expressed by formula (23):

$$\bar{K}n = \bar{K} - UNIT * \bar{K} - \bar{K}\cdot UNIT \tag{23}$$

In formula (23), $UNIT$ represents the unit matrix of intelligence home furnishing data.

Calculating the eigenvalue of the following intelligence home furnishing data:

$$\bar{K}n * g = lg \tag{24}$$

In order to test and train the intelligence home furnishing data set conveniently, the kernel matrix is expressed as:

$$K = \left[\ker nel(u,v)\right]_{n\times m} \tag{25}$$

In the formula (25), n represents the total samples of intelligence home furnishing data set, m is the number of samples of $\bar{K}$, u is intelligence home furnishing data in the set, v is one of intelligence home furnishing data set of $\bar{K}$, the following formula is obtained by centralizing intelligence home furnishing data:

$$K = K\_TRAIN\cdot UNIT\cdot\bar{K} - K\cdot UNIT \tag{26}$$

In formula (26), $UNIT\_TRAIN$ represents the intelligence home furnishing data matrix.

For the test intelligence home furnishing data set, the kernel matrix K_test of test intelligence home furnishing data can be calculated using the same method:

$$K\_test = \left[\ker nel(u',v')\right]_{test} \tag{27}$$

The transformed intelligence home furnishing data kernel matrix after intelligence home furnishing data centralization is:

$$K\_test\_new = K\_test - UNIT\_TEST\cdot\bar{K} \tag{28}$$

In formula (28), $UNIT\_TEST$ represents the intelligence home furnishing data matrix of $test\_number$, $test\_number$ is the number of intelligence home furnishing data samples in the intelligence home furnishing data set, and the above processes are combined to complete the feature extraction of the intelligence home furnishing data.

2.3. Intelligence Home Furnishing Data Clustering of K-means based on Genetic Algorithm

Intelligence home furnishing data after feature extraction is classified using K-means clustering of genetic algorithm. The genetic algorithm and K-means clustering algorithm are combined. The genetic algorithm is applied to optimize the function, and the convergence speed of genetic algorithm is increased with crossover and mutation factor, thus, the optimal intelligence home furnishing data clustering results is obtained.

Encoding operation of intelligence home furnishing data clustering floating-point is: based on the number of clusters H , one of the chromosomes is acted as a string composed of H intelligence home furnishing data clusters, for example, clustering operation is applied to D dimensional intelligence home furnishing data samples to get H clusters, the chromosome structure is expressed as:

$$Z=\{(x_{11},x_{12},\ldots,x_{1d}),(x_{k1},x_{k2},\ldots,x_{kd})\} \quad (29)$$

In formula (29), Z means that the chromosome is made up of $k\times d$ floating point codes. The encoding method based on intelligence home furnishing data clustering floating points is relatively simple, and the calculation speed is fast because the problem of encoding repetition and decoding is avoided.

The fitness function is used to evaluate the fitness of the intelligence home furnishing data population, and distinguish the individual intelligence home furnishing data in the population. The individual fitness in the population is proportional to the survival probability. The individual fitness degree is higher, the greater the chance of survival. K-means clustering algorithm is mainly to find out the minimum division of G intelligence home furnishing data objective function:

$$G=\sum_{j=1}^{c}\sum_{k=1}^{n_j}\left\|x_k^{(j)}-m_j\right\|^2 \quad (30)$$

In formula (30), $m_j\,(j=1,2,\ldots c)$ represents the center of each intelligence home furnishing data cluster, and x_k represents the intelligence home furnishing data sample.

In the process of using genetic algorithm, chromosome code need to be clustered using population, intelligence home furnishing data object function G is used to calculate the distance between points in the cluster and center of cluster. The objective function can be applied to judge intelligence home furnishing data clustering quality, if intelligence home furnishing data object function G is smaller, the intelligence home furnishing data clustering effect is better

Based on the above analysis, the minimum solution of the G is found in the space of the search target function in the genetic algorithm, hence, the fitness function is constructed according to the intelligence home furnishing data objective function:

$$fitness=\frac{1}{G} \quad (31)$$

The selection of the genetic algorithm is based on the individual fitness of the population. The best individual is selected from the parent individual for inheriting, the selection operator depends on the selection probability, and the selection probability of the individual X_i can be defined as:

$$P(X_i)=\frac{finess(X_i)}{\sum_{J=1}^{N} finess(X_i)},i=1,2,\ldots,N \quad (32)$$

In formula (31), the greater the $P(X_i)$ value, the greater the probability that individual X_i is selected to inherit into the next generation.

In the process of genetic manipulation, selection of crossover probability P_m and mutation probability P_c influence the results of genetic algorithm. For cross operator, when crossover probability is large, the individual generation speed is prone to be faster, and vice versa. For mutation operator, when the probability of mutation is small, it is tend to lose novelty, when it is large, the genetic algorithm is void and turn into search algorithm.

In view of the problems existing in the operation of crossover and mutation, the adaptive crossover and mutation operation are adopted, and P_c and P_m can be adjusted according to different situations. Adaptive adjustment of P_c and P_m can be expressed as:

$$P_c=\begin{cases}k_1(f_{\max}-f')/f_{\max}-f_{avg}, & f'\ge f_{avg}\\ k_2, & f'<f_{avf}\end{cases}$$

$$P_m=\begin{cases}k_3(f_{\max}-f)/f_{\max}-f_{avg}, & f\ge f_{avg}\\ k_4, & f'<f_{avf}\end{cases} \quad (33)$$

In formula (33), $f_{\max}$ represents the fitness value in the intelligence home furnishing data population, f_{avg} represents the average fitness value of the intelligence home furnishing data population, and f' represents the fitness value of the crossover individual, and f represents fitness value of the mutation individual. The values of k_1, k_2, k_3, k_4 is in the range of $(0,1)$, if no clear definition of k_1, k_2, k_3, k_4, the value of k_1, k_2, k_3, k_4 will be determined preliminarily. By influencing P_c and P_m to compare intelligence home furnishing data optimization goal and evolving generations, P_c and P_m of less evolution times are better, the corresponding k_1, k_2, k_3, k_4 are more reasonable.

K-means clustering is an algorithm of great local search, and genetic algorithm is advertised for good global search. Two algorithms are combined, using the genetic algorithm to search global intelligence home furnishing data center to find the optimal cluster center and K-means clustering algorithm to perform intelligence home furnishing data clustering. The concrete process is as below:

(1) According to the genetic algorithm, the chromosome of strongest fitness in the intelligence home furnishing data population is obtained, and the clustering center is decoded.

(2) According to the criterion function, the distance between the intelligence home furnishing data cluster point and the cluster center is calculated.

(3) Find out the population of shortest distance to complete the intelligence home furnishing data clustering.

The concrete data clustering flow chart is shown in figure 1:

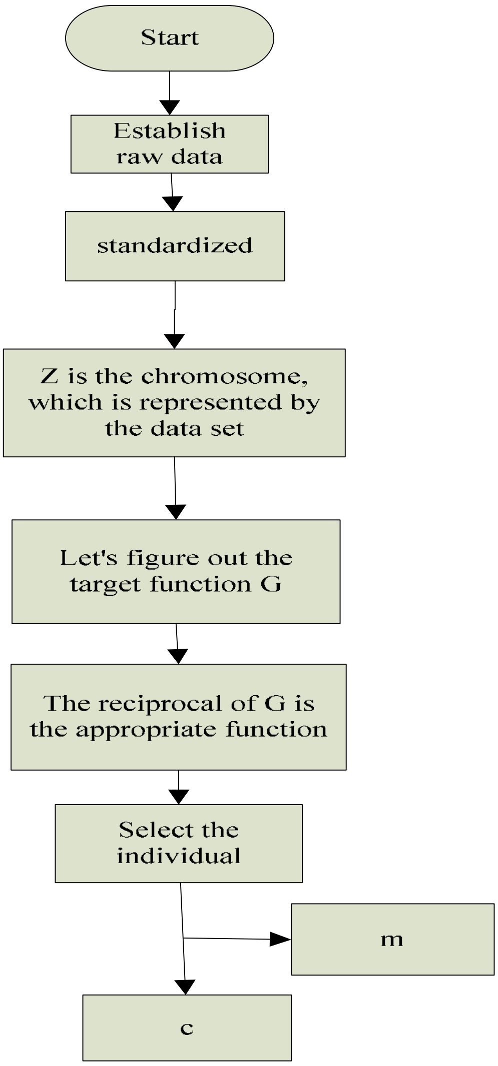

Figure 1. data classification flowchart

3. Experimental Results and Analysis

In order to verify the validity and accuracy of the mining method based on K-means clustering and genetic algorithm, a simulation experiment was performed. The intelligence home furnishing data mining experimental platform is built in the matlab8.0 environment.

The experimental images are shown in figure 2:

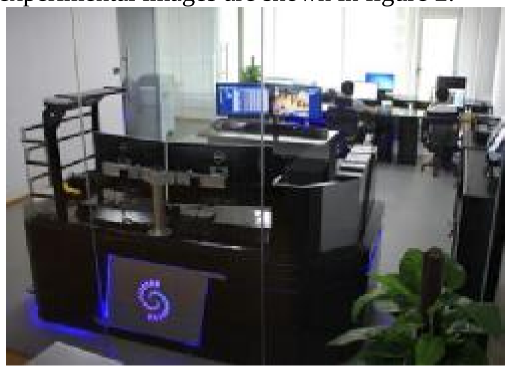

figure 2 .experimental images

The experimental intelligence home furnishing data from KDD cup2017 intelligence home furnishing data set, the intelligence home furnishing data set includes 2 million intelligence home furnishing data. The basic parameters of genetic algorithm: population size is $N = 70$, encoding length is $l = 30$, the maximum evolution generation is $G = 900$, crossover probability is $P_{c1} = 0.8$, $P_{c2} = 0.5$. Table 1 shows two clustering algorithms with 40 times running.

In order to verify search speed of proposed method, table 1 has the maximum and minimum and average evolution generation of two clustering algorithms, denoted as $G_{\max}$, $G_{\min}$, G_{ave}. On the basis of the analysis of experimental intelligence home furnishing data, evolution generation obtained using the proposed method is minimum, the convergence is the fastest. The accuracy of searching optimal solution of intelligence home furnishing data mining is important for optimization. In table 1, $f_{\max}$, $f_{\min}$, f_{ave}, represents the maximum and minimum and average optimal value, the method proposed in this paper has the highest precision for function optimization. The experiment of 40 operations is performed using the proposed method, the number of times meets the condition is divided by 40 to represent the success rate Su_{rate}, which shows the robustness of the algorithm. For function Su_{rate}, it can be seen that the method proposed in this paper is more reliable.

Table 1. Comparison of Optimal Evolution Generation of Different Algorithm Functions

Methods	The proposed method	Standard genetic algorithm

G_{max}	65	900
G_{min}	9	10
G_{ave}	30.7	542.2
Su_{rate}	100%	56%
f_{max}	$9.502e^{-5}$	0.5582
f_{min}	$1.23e^{-10}$	$3.58e^{-5}$
f_{ave}	$2.7048e^{-5}$	0.01407

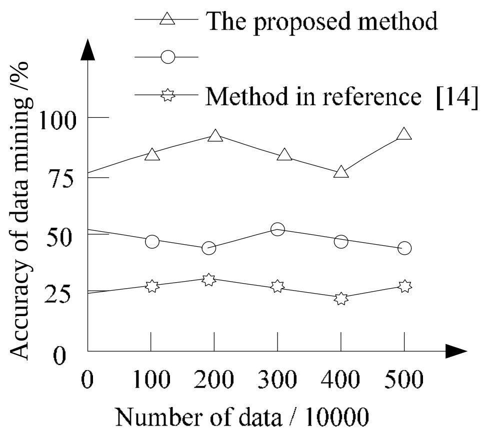

Figure 3. Comparison of Intelligence Home Furnishing Data Mining Accuracy of Different Methods

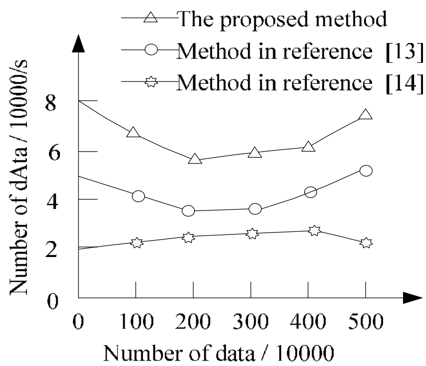

Figure 4. comparison of data mining time in different methods

The analysis of Figure 3 shows that the accuracy of the proposed method is much higher than methods based on association principle and intelligence home furnishing data density increment space [13, 14]. Because the proposed method is the combination of genetic algorithm and K-means clustering algorithm. The genetic algorithm is utilized to optimize the function, improve the convergence speed of genetic algorithm with crossover and mutation factor, so as to obtain the best intelligence home furnishing data clustering results. It can be seen that the feasibility of intelligence home furnishing data mining based on genetic algorithm and K-means clustering is high, and it can solve the time-consuming problem of intelligence home furnishing data processing effectively.

The analysis of Figure 4 shows that the time of the proposed method is much lower than methods based on association principle and intelligence home furnishing data density increment space [13, 14]. Because the proposed method is the combination of genetic algorithm and K-means clustering algorithm. The genetic algorithm is utilized to optimize the function, improve the convergence speed of genetic algorithm with crossover and mutation factor, so as to obtain the best intelligence home furnishing data clustering results. It can be seen that the feasibility of intelligence home furnishing data mining based on genetic algorithm and K-means clustering is short, and it can solve the time-consuming problem of intelligence home furnishing data processing effectively.

4. Conclusion

The current method is based on the relationship between intelligence home furnishing data sets to determine the principle of association of intelligence home furnishing data sets, and the intelligence home furnishing data mining factor and relative error are introduced to improve the accuracy of intelligence home furnishing data mining algorithm, but this kinds of method are time-consuming. Therefore, a intelligence home furnishing data mining method based on K-means clustering and genetic algorithm is proposed. Experimental results show that this method can effectively process intelligence home furnishing data, reduce the time of intelligence home furnishing data processing and improve the accuracy of intelligence home furnishing data mining, and has a wide range of practical value.

Acknowledgment

Jiangsu Education Science 12th Five-Year plan topic (No. B-b/2015/03/067); Jiangsu Provincial Department of Education: general topic of Union College: design and application of Desktop Virtualization (No. B-2016-09-010).

References

[1] S C Saha, C Lei, J C Patterson. Detecting the financial statement fraud: The analysis of the differences between data mining techniques and experts' judgments[J]. Knowledge-Based Systems, 2015, 89(9):459-470.

[2] A A Razi-Kazemi, M Vakilian, K Niayesh, et al. Data Mining of Online Diagnosed Waveforms for Probabilistic Condition Assessment of SF$_{6$ Circuit Breakers[J]. IEEE Transactions on Power Delivery, 2015, 30(3):1354-1362.

[3] O U Osokogu, F Fregonese, C Ferrajolo, et al. Pediatric Drug Safety Signal Detection: A New Drug–Event Reference Set for Performance Testing of Data-Mining Methods and Systems[J]. Drug Safety, 2015, 38(2):207-217.

[4] K Yoshimura, T Okanoue, H Ebise, et al. Identification of novel noninvasive markers for diagnosing nonalcoholic steatohepatitis and related fibrosis by data mining[J]. Hepatology, 2016, 63(2):462-473.

[5] N Settouti, H Aourag. A Study of the Physical and Mechanical Properties of Lutetium Compared with Those of Transition Metals: A Data Mining Approach[J]. JOM, 2015, 67(1):87-93.

[6] Bui D T, Ho T C, Pradhan B, et al. GIS-based modeling of rainfall-induced landslides using data mining-based functional trees classifier with AdaBoost, Bagging, and MultiBoost ensemble frameworks[J]. Environmental Earth Sciences, 2016, 75(14):1-22.

[7] Pattanapairoj S, Silsirivanit A, Muisuk K, et al. Improve discrimination power of serum markers for diagnosis of cholangiocarcinoma using data mining-based approach.[J]. Clinical Biochemistry, 2015, 48(10–11):668-673.

[8] Kelley B P, Klochko C, Halabi S, et al.Datafish Multiphase Data Mining Technique to Match Multiple Mutually Inclusive Independent Variables in Large PACS Databases[J]. Journal of Digital Imaging, 2016, 29(3):331-336.

[9] Elsabahy M, Wooley K L. Data Mining as a Guide for the Construction of Cross-Linked Nanoparticles with Low Immunotoxicity via Control of Polymer Chemistry and Supramolecular Assembly[J]. Accounts of Chemical Research, 2015, 48(6):1620-30.

[10] Lu A T, Austin E, Bonner A, et al. Applications of machine learning and data mining methods to detect associations of rare and common variants with complex traits.[J]. Genetic Epidemiology, 2015, 38(S1):S81-S85.

[11] RUAN Mengli. Simulation of Aircraft Fault Data Mining Algorithm Based on Mew GEP[J]. Computer Simulation, 2015 , 32 (6) :92-95.

[12] SUN Hongbin.Design and implementation of sports performance management system based on data mining[J]. Electronic Design Engineering, 2016 , 24 (5) :74-77.

[13] LIANG Fenglan.Agricultural Products Quality Fluctuation Traceability Method Based on Data Mining[J].Science and Technology and Engineering, 2017 , 17 (3) :268-272.

[14] HU Wenhong, SUN Xinxin. Study of Data Mining Technology in Application of City Waterlogging Based on Time Series[J].Bulletin of Science and Technology, 2016 , 32 (6) :229-231.

[15] REN Gaoju, Wang Hongwei . Design and Implementation of University Management Information System Based on intelligence home furnishing data Mining[J].Computer Measurement & Control, 2016 , 24 (10) :255-258.