User Browsing Data Mining Method of E-commerce Platform based on Genetic Algorithm

Yadong Zhu
Nanjing Engineering Branch, Jiangsu Union Technical Institute, Nanjing, 210000, China

Abstract: In order to shorten the processing time of massive e-commerce platform users browsing data and improve the processing speed of e-commerce platform users browsing data, it is necessary to study the data mining algorithm. Therefore, this paper proposes a k-means clustering data mining method based on genetic algorithm. Firstly, the data of browsing data of e-commerce platform users are collected, and the data features are extracted. Finally, the K-means clustering method of genetic algorithm is used to cluster the user browsing data of e-commerce platform to complete the data mining analysis. The results show that the method can effectively reduce the processing time of browsing data of e-commerce platform users, and can complete data mining in at least 2 seconds, and the accuracy is high, up to 97%, which has certain application value.

Keywords: E-commerce platform; Data acquisition; Data feature extraction; Data mining

1. Introduction

With the rapid development of the Internet, e-commerce platform provides more and more convenient shopping experience for people. In this case, how to mine valuable information from the complex e-commerce platform user browsing data has become a new challenge [1]. In order to promote the development of its own development, it is necessary to mine useful user browsing data of e-commerce platform from the user browsing data set [2, 3]. However, e-commerce platform users browsing data is extremely rich, so it is urgent to use data analysis tools to mine useful knowledge, relationships and rules from the massive user browsing data of e-commerce platform, so as to bring greater development value. Therefore, data mining technology is of great significance to the development of electronic commerce industry [4]. E-commerce platform user browsing data mining refers to the acquisition of interested knowledge from the data browsed by users of e-commerce platform, which is effective, novel and potentially useful, and is demonstrated to users through an understanding model [5, 6]. This technology not only includes the retrieval of database browsing for specific e-commerce platform users, but also includes the statistics, analysis and reasoning of these data, so as to provide solutions to the existing problems such as message push [7, 8]. At present, e-commerce platform user browsing data mining algorithms mainly include classification, clustering, association rules and prediction [9]. Classification refers to extracting a model from browsing data set to classify user browsing data of e-commerce platform [10]. Clustering refers to dividing the e-commerce platform users' browsing data according to the similarity when the classification is not clear [11, 12].

Nowadays, data mining technology has been widely used in various fields, and many scholars have analyzed it. For example, literature [13] uses data mining to detect the fluctuation of agricultural product quality characteristics, which improves the disadvantages of inaccurate traceability and low efficiency of agricultural product quality characteristic fluctuation; literature [14] applies data mining technology based on time series to urban water logging disaster related fields, so as to provide certain guidance for urban flood control; literature [15] aims to improve university management The ability of information management is analyzed by using data mining technology, and the software development and design of university management information system are carried out, so that the improved system has better reliability and human-computer interaction. Although the above data mining methods have some advantages, they also have the disadvantages of too long mining time and low accuracy.

In order to solve the above problems, this paper proposes a data mining method based on K-means clustering of genetic algorithm. Firstly, the data of browsing data of e-commerce platform users are collected, and the data features are extracted. Finally, the K-means clustering method of genetic algorithm is used to cluster the user browsing data of e-commerce platform to complete the data mining analysis. The experimental results show that the method can effectively improve the accuracy of data mining and reduce the data processing time.

2. E-commerce Platform Users Browsing Data Mining based on Cloud Platform

Mobile node method is used to collect user browsing data of e-commerce platform, and kernel principal component analysis method is used to extract features of the collected data. Then, the K-means clustering method of genetic algorithm is used to cluster the extracted features, so as to complete the research of e-commerce platform user browsing data mining algorithm. The specific steps are as follows:

2.1. E-commerce platform users browsing data acquisition based on movable nodes

The operation range of movable node can be regarded as a disk and the E-commerce platform users browsing data acquisition node is the center. The first task is to find the coordinate set, and the following conditions need to be fulfilled:

A. The number of E-commerce platform users browsing data coordinates is as small as possible;

B. The disk which has radius r covers the area.

When users of e-commerce platform browse the data disk arrangement, in order to avoid the blank of browsing data, it is necessary to overlap the adjacent data. In order to reduce the number of data disks, the smaller the overlap of data disks, the better. This theorem can be used to illustrate.

Theorem 1 when overlapping E-commerce platform users browsing data disks form triangles, the area of overlap is minimum.

Proof: assume that overlap areas of E-commerce platform users browsing data disks are S_{AB}, S_{AC}, S_{BC}, and the overlap area S is:

$$S = S_{AB} + S_{AC} + S_{BC} = \frac{1}{2}r^2(\alpha - \sin\alpha) + \frac{1}{2}r^2(\gamma - \sin\gamma) + \frac{1}{2}r^2(\beta - \sin\beta) \quad (1)$$

In equation (1), α represents the length of the E-commerce platform users browsing data disk, γ represents the height of the E-commerce platform users browsing data disk, and β represents the width of the E-commerce platform users browsing data disk.

Since the E-commerce platform users browsing data disks overlap to form polygons, thus:

$$z\alpha + (\pi - \alpha) + 2\beta + (\pi - \beta) + 2\gamma + (\pi - \gamma) = 4\pi \quad (2)$$

It can be concluded that:

$$A + \beta + \gamma = \pi \quad (3)$$

The E-commerce platform users browsing 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 $\alpha = \beta = \gamma = \pi / 3$. At this point, 3 data disks form the equilateral triangle of data, and the length is $\sqrt{3}r$.

There are several sets for users to browse data disk in e-commerce platform satisfying theorem 1, among which one set 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.

E-commerce platform users browse the data disk to confirm, and then select the electronic data collection point. Data energy consumption needs to be considered when determining the location.

When a node transmits E-commerce platform users browsing data, it is represented by formula (5):

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

The formula (5), G_r is the gain of E-commerce platform users browsing data receiver, G_t is the E-commerce platform users browsing data transmitting gain, P_r is the power of E-commerce platform users browsing data receiver, P_t is the E-commerce platform users browsing data transmitting power, d represents the distance between E-commerce platform users browsing data node and destination node, L denotes the E-commerce platform users browsing data loss factor, λ is the E-commerce platform users browsing data wavelength. For the specific network, suppression of E-commerce platform users browsing data node is relatively ideal, formula (5) can be simplified to:

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

In formula (6), $K = L(4\pi)^2 P_r / G_r G_t \lambda^2$ is considered a E-commerce platform users browsing data constant.

If there are n E-commerce platform users browsing data nodes in the E-commerce platform users browsing data disk, in the process of E-commerce platform users browsing data acquisition, the transmitting power of the E-commerce platform users browsing 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 E-commerce platform users browsing data nodes, and (x_s, y_s) represents the location of the E-commerce platform users browsing data acquisition point.

It can be seen from formula (7) that in order to lower the E-commerce platform users browsing data transmitting power, a reasonable E-commerce platform users browsing data acquisition point is needed.

Theorem 2 assumes that there are n points in E-commerce platform users browsing data and the E-commerce platform users browsing data coordinate is

(x_i, y_i), $1 \le i \le n$. The E-commerce platform users browsing data node is $\left(\frac{1}{n}\Sigma x_i, \frac{1}{n}\Sigma y_i\right)$.

Proof: if the E-commerce platform users browsing data coordinate of the center is (x_c, y_c), the coordinates of the E-commerce platform users browsing data points are (x_s, y_s), and the distance between each E-commerce platform users browsing data node is d_i, then the distance square sum f_s of all E-commerce platform users browsing 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 e-commerce platform users browsing data nodes, but some users of nodes will not be collected, so Theorem 3 is proposed.

Theorem 3 assumes that the horizontal and vertical coordinates of maximum and minimum E-commerce platform users browsing data in the E-commerce platform users browsing data disk are $x_{\max}, x_{\min}, y_{\max}, y_{\min}$, and (x_s, y_s) represents the average of all E-commerce platform users browsing data nodes. When the formula (9) is satisfied, the distance of E-commerce platform users browsing 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 E-commerce platform users browsing data node.

As explained by theorem 3, when the E-commerce platform users browsing data node of disk meets the formula (9), the E-commerce platform users browsing data node is further analyzed. Since the radius of the E-commerce platform users browsing 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 E-commerce platform users browsing 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 E-commerce platform users browsing data collection of the cloud platform is completed.

2.2. E-commerce platform users browsing data feature extraction based on kernel principal component analysis

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

The covariance matrix of the E-commerce platform users browsing 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 E-commerce platform users browsing data, T represents the eigenvalues of the covariance matrix of the E-commerce platform users browsing data sample, and n represents the eigenvectors of the covariance matrix, and the formula (13) is solved:

$$\mathrm{cov} * \beta = \lambda \beta \tag{13}$$

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

$$\Phi(x_k) * \mathrm{cov} * \beta = \lambda \Phi(x_k) \beta \quad k = 1, 2, \cdots, n \tag{14}$$

From the related theory of the reproducing E-commerce platform users browsing data kernel space, we can see that any vector in the E-commerce platform users browsing data kernel space can be represented by the E-commerce platform users browsing data base vector $\Phi(x_1), \Phi(x_1), \cdots, \Phi(x_n)$, that is, $\beta = \sum_{i=1}^{n} \alpha_i \Phi(x_i)$, which is put into the formula (15) to get:

$$\Phi(x_k) * \mathrm{cov} * \left(\sum_{i=1}^{n} \alpha_i \Phi(x_i) \right) = \lambda \Phi(x_k) \left(\sum_{i=1}^{n} \alpha_i \Phi(x_i) \right) \quad k = 1, 2, \cdots, n \tag{15}$$

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

$$\frac{1}{n} \sum_{i=1}^{n} \alpha_i \Phi(x_k) \cdot \sum_{j=1}^{n} \Phi(x_i) (\Phi(x_i)) \cdot \lambda \Phi(x_j) = \lambda \sum_{i=1}^{n} \alpha_i (\Phi(x_i)) \quad k = 1, 2, \cdots, n \tag{16}$$

The operations of the spatial points of the E-commerce platform users browsing data kernel can be computed by the numerical value of the corresponding kernel function.

The E-commerce platform users browsing data matrix $K=\left[\prec\Phi(x_i),\Phi(x_j)\succ\right]$ is defined, and K is substituted into formula (16) to simplify:

$$n\lambda K\alpha = K^2\alpha \tag{17}$$

Let $\lambda' = n\lambda$, the E-commerce platform users browsing data in formula (17) can be solved by solving the E-commerce platform users browsing data eigenvalue in formula (18):

$$K\alpha = \lambda'\alpha \tag{18}$$

Assuming that the eigenvector corresponding to the maximum eigenvalue of the E-commerce platform users browsing data is represented as $\alpha_1,\alpha_2,\cdots,\alpha_m$, whereby the attribute of i_{-th} dimension of the arbitrary E-commerce platform users browsing data test point $\Phi(x)$ is denoted as:

$$x(i)=\sum_{j=1}^{n}\alpha_k(j)\prec\Phi(x_j),\Phi(x)\succ \tag{19}$$

Formula (19) can be used to calculate the result of dimension reduction of e-commerce platform users' browsing data, which can convert the data set matrix and the input e-commerce platform user browsing data space into data space.

The original mean value of each sample is subtracted from each sample, and the browsing data is processed to meet the zero mean value.

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

Then the E-commerce platform users browsing data element (i,j) in the E-commerce platform users browsing data kernel matrix is changed to:

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

In formula (21), L_n stands for E-commerce platform users browsing data element.

Assuming that the core matrix formula (22) can be calculated by using m sample E-commerce platform users browsing data:

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

In formula (22), $\mathrm{ker}\,nel$ represents the E-commerce platform users browsing data kernel function, and $\bar{K}$ takes the mean value of the E-commerce platform users browsing data in the E-commerce platform users browsing data kernel space as 0. This method is identical to the formula (21). The transformed E-commerce platform users browsing 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 E-commerce platform users browsing data.

Calculating the eigenvalue of the following E-commerce platform users browsing data:

$$\bar{K}n*\gamma=\lambda\gamma \tag{24}$$

In order to test and train the E-commerce platform users browsing data set conveniently, the kernel matrix is expressed as:

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

In the formula (25), n represents the total samples of E-commerce platform users browsing data set, m is the number of samples of $\bar{K}$, u is E-commerce platform users browsing data in the set, v is one of E-commerce platform users browsing data set of $\bar{K}$, the following formula is obtained by centralizing E-commerce platform users browsing data:

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

In formula (26), $UNIT\_TRAIN$ represents the E-commerce platform users browsing data matrix.

For the test E-commerce platform users browsing data set, the kernel matrix K_test of test E-commerce platform users browsing data can be calculated using the same method:

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

The transformed E-commerce platform users browsing data kernel matrix after E-commerce platform users browsing data centralization is:

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

In formula (28), $UNIT\_TEST$ represents the E-commerce platform users browsing data matrix of $test\_number$, $test\_number$ is the number of E-commerce platform users browsing data samples in the E-commerce platform users browsing data set, and the above processes are combined to complete the feature extraction of the E-commerce platform users browsing data.

2.3. E-commerce platform users browsing data clustering of k-means based on genetic algorithm

E-commerce platform users browsing 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 E-commerce platform users browsing data clustering results is obtained.

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

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

In formula (29), Z means that the chromosome is made up of $k\times d$ floating point codes. The encoding method based on E-commerce platform users browsing 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 E-commerce platform users browsing data population, and distinguish the individual E-commerce platform users browsing 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 E-commerce platform users browsing data objective function:

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

In formula (30), $m_j(j=1,2,\cdots c)$ represents the center of each E-commerce platform users browsing data cluster, and x_k represents the E-commerce platform users browsing data sample.

In the process of using genetic algorithm, chromosome code need to be clustered using population, E-commerce platform users browsing 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 E-commerce platform users browsing data clustering quality, if E-commerce platform users browsing data object function G is smaller, the E-commerce platform users browsing 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 E-commerce platform users browsing data objective function:

$$fitness=\frac{1}{G} \tag{31}$$

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:

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

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:

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

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

In formula (33), $f_{\max}$ represents the fitness value in the E-commerce platform users browsing data population, f_{avg} represents the average fitness value of the E-commerce platform users browsing 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 E-commerce platform users browsing 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 E-commerce platform users browsing data center to find the optimal cluster center and K-means clustering algorithm to perform E-commerce platform users browsing data clustering. The concrete process is as below:

A. According to the genetic algorithm, the chromosome of strongest fitness in the E-commerce platform users browsing data population is obtained, and the clustering center is decoded.

B. According to the criterion function, the distance between the E-commerce platform users browsing data cluster point and the cluster center is calculated.

C. Find out the population of shortest distance to complete the E-commerce platform users browsing data clustering.

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

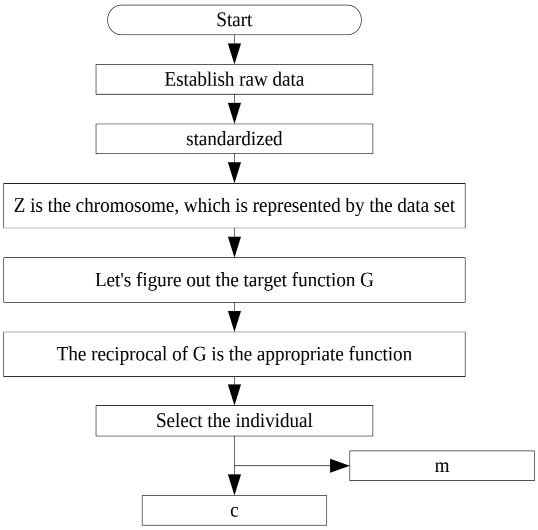

Figure 1. Data classification flowchart

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 E-commerce platform users browsing data mining experimental platform is built in the matlab8.0 environment.
The experimental images are shown in figure 2:

Figure 2. Experimental images

The user browsing data of experimental e-commerce platform comes from a certain e-commerce platform user browsing data set, including 2 million related browsing 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 E-commerce platform users browsing data, evolution generation obtained using the proposed method is minimum, the convergence is the fastest. The accuracy of searching optimal solution of E-commerce platform users browsing 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

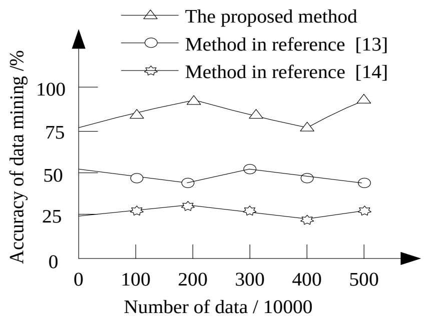

Figure 3. Comparison of E-commerce platform users browsing data mining accuracy of different methods

It can be seen from the analysis in Fig. 3 that the accuracy of this method is much higher than that of the method based on association principle and the method of browsing the incremental space of data density of e-commerce platform users, that is, the method in references [13] and [14], and has certain feasibility.

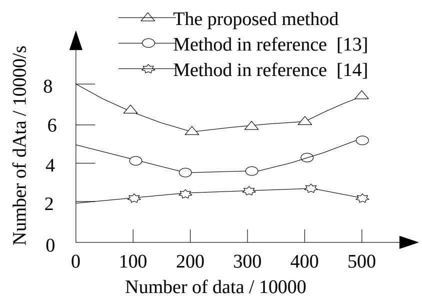

Figure 4. Comparison of data mining time in different methods

The analysis of Figure 4 shows that the time of the proposed method is much lower than methods based on association principle and E-commerce platform users browsing 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 E-commerce platform users browsing data clustering results. It can be seen that the feasibility of E-commerce platform users browsing data mining based on genetic algorithm and K-means clustering is short, and it can solve the time-consuming problem of E-commerce platform users browsing data processing effectively.

4. Conclusion

The current method is to determine the association principle of e-commerce platform user browsing data set according to the relationship between e-commerce platform user browsing data set, and introduce the mining factor and relative error of e-commerce platform user browsing data to improve the accuracy of e-commerce platform user browsing data. However, this method is time-consuming. Therefore, this paper proposes a data mining method based on K-means clustering and genetic algorithm. The experimental results show that the method can effectively deal with the browsing data of e-commerce platform users, reduce the processing time of browsing data, improve the accuracy of data mining, and has a wide range of practical value

5. Acknowledgments

A. Jiangsu Education Science 12th Five-Year plan topic (B-b/2015/03/067).
B. Jiangsu Provincial Department of Education: general topic of Union College: design and application of Desktop Virtualization (B-2016-09-010).

References

[1] Chi-Chen Lin, Ann Chiu, Shaio Yan Huang, David C. Yen. Detecting the financial statement fraud: The analysis of the differences between data mining techniques and experts' judgments. Knowledge-Based Systems. 2015, 89(9), 459-470.

[2] Razi-Kazemi A A, Vakilian M, Niayesh K, et al. Data mining of online diagnosed waveforms for probabilistic condition assessment of SF$_{6$ Circuit Breakers. IEEE Transactions on Power Delivery. 2015, 30(3), 1354-1362.

[3] Osokogu O U, Fregonese F, Ferrajolo C, et al. Pediatric drug safety signal detection: a new drug–event reference set for performance testing of data-mining methods and systems. Drug Safety. 2015, 38(2), 207-217.

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

[5] Settouti N, Aourag H. A study of the physical and mechanical properties of lutetium compared with those of transition metals: a data mining approach. 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. 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. 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. 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. 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. Genetic Epidemiology. 2015, 38(S1), S81-S85.

[11] Ruan Mengli. Simulation of aircraft fault data mining algorithm based on Mew GE. Computer Simulation. 2015 , 32(6), 92-95.

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

[13] Liang Fenglan. Agricultural products quality fluctuation traceability method based on data mining. 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. 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. Computer Measurement & Control. 2016, 24(10), 255-258.