

## Sensitivity Analysis and Optimization of FCM Initial Clustering Centers

Yong WANG<sup>1a</sup>, Jiao CHENG<sup>1b\*</sup>, Ling ZHANG<sup>1c</sup>, Qiu-Hua Tang<sup>2d</sup>

<sup>1</sup>College of Management, Wuhan University of Science and Technology, Wuhan 430081, China

<sup>2</sup>College of Mechanics and Atomation, Wuhan University of Science and Technology, Wuhan 430081, China

<sup>a</sup>wangyong@wust.edu.cn, <sup>b</sup>346182000@qq.com, <sup>c</sup>472594732@qq.com, <sup>d</sup>467263403@qq.com

**Keywords:** fuzzy C- means clustering, sensitivity analysis, clustering centers move with fuzzy exponent

**Abstract.** Some researches show that fuzzy C- means clustering is sensitive to initial clustering centers. Related variables include cluster number, data characteristics and fuzzy exponent are affect the sensitivity of the algorithm. This paper provided the definition of the sensitivity of the algorithm. Selecting data sets of different characteristics, experimental analyzed the sensitivity of the algorithm to the initial clustering centers under different clustering case, and compared the convergence speed of different initial clustering centers reach the specific clustering distance. The results show that the algorithm is not absolutely sensitive to the initial clustering centers, the sensitivity is related to the number of clustering for the same data set. This paper proposed the clustering centers move with fuzzy exponent algorithm for FCM algorithm by combining the maximum distances product method, and analyzed the validity of the optimization algorithm.

### Introduction

Clustering analysis is an important tool for data preprocessing and important area of artificial intelligence research. Fuzzy clustering gather the entities through establish fuzzy similarity relationship according the characteristics and similarity. Fuzzy C-means Clustering (FCM) has the advantages of fast convergence speed, efficient processing of high-dimensional data, simple calculation and so on. It is widely used in image processing, artificial intelligence, medicine and other fields [1-3]. The FCM algorithm is a partition method based on Picard iteration through necessary conditions for optimizing a weighted sum of squared errors objective function.  $J_1$  is the within-groups sum of squared error (WGSS) objective function of within-class square sum of error, it is mainly used to define hard C-means and hard ISODATA algorithms. In 1973 Dunn first extended  $J_1$  to  $J_2$ , then Bezdek extended  $J_2$  to  $J_m$  by introduced fuzzy exponent  $m$ , and  $m$  ranges from 1 to infinity [4]. The FCM algorithm contains different parameters, and different parameters lead to different clustering results. The input parameters include initial centers, fuzzy exponent, cluster number and data set, and the results are the optimal centers and clustering distance. In general, the objective function  $J_m$  is regarded as the evaluation criterion of the clustering results. However, due to the existence of the local minimum point of the objective function and the greedy property of the algorithm, the algorithm is sensitive to the initial clustering centers, and the clustering results easy to converge to the locally optimal value and fail to reach the global optimum [5]. In 1980 Bezdek presented that every iterate sequence of FCM converge to local minimizer of  $J_m$  by applied Zangwill's theory to FCM iteration [6]. Then the existence of saddle points of the FCM objective function at locations was presented in 1987[4]. The researchers improve and optimize the FCM algorithm considering it is sensitive to the initial clustering centers, and the clustering results easy to converge to the local optimum [7]. Genetic fuzzy C-means (GFCM) algorithm introduces GA algorithm into FCM, the global search performance of GA replace the iterative rules of FCM to solve the sensitive problem [8]. Xu proposed weighted maximum entropy kernel fuzzy clustering algorithm, using the data point distribution density as the weight to optimize the initial clustering centers [9]. Zhang proposed an improved FCM algorithm, designed a simple and fast initial centers selection rule, and improved the objective function according to the

separation characteristics of the clustering centers [10]. Zhou proposed a fuzzy clustering algorithm based on the midpoint density function, combining the midpoint method and the cluster density function to measure the object density [11]. The above researches solve the sensitive problem of FCM algorithm to some extent, but the definition of the sensitivity of initial clustering centers is not clear, and the influence of fuzzy exponent on the clustering results is neglected in the optimization process.

### Fuzzy C-means Clustering

FCM algorithm through the grade of membership to determine the extent of each data point belongs to a clustering in a data set. Giving data set  $X = (x_1, x_2, \dots, x_n)$ , for  $\forall j, 1 \leq j \leq N$ ,  $x_j = \{x_{j1}, x_{j2}, \dots, x_{jk}, \dots, x_{jn}\}$ .  $R^n$  note real n-dimension vector space.  $N$  is the number of data points in sample data set.  $x_{jk}$  ( $k=1, 2, \dots, n$ ) denote the  $k$  th attribute value of the  $j$  th data point. Let  $V = (v_1, v_2, \dots, v_c)$  be a clustering centers vector set, for  $\forall i, 1 \leq i \leq c$ ,  $v_i = \{v_{i1}, v_{i2}, \dots, v_{ik}, \dots, v_{in}\}$ ,  $v_{ik}$  ( $k=1, 2, \dots, n$ ) denote the  $k$  th attribute value of the  $i$  th clustering center. The fuzzy clustering objective function of data set  $X$  as Eq.1:

$$J_m(U, V) = \sum_{i=1}^c \sum_{j=1}^n u_{ij}^m d_{ij}^2 = \sum_{i=1}^c \sum_{j=1}^n u_{ij}^m \|x_j - v_i\|^2. \quad (1)$$

Where,  $v_{ik} \in v_i$ ,  $c$  is cluster number.  $d_{ij}$  denote Euclidean distance from data point  $x_j$  to clustering center  $v_i$ .  $U$  is the membership matrix, and it can be describe the cluster structure of data set  $X$ .  $u_{ij}$  denote the grade of membership of  $x_j$  in the  $i$  th cluster.  $m$  is fuzzy exponent, and it can control the fuzzy degree of clustering.  $J_m(U, V)$  is the square sum of weighted distances between all data points to clustering centers.

The membership satisfies  $\sum_{i=1}^c u_{ij} = 1$ , Lagrange function by make it as the constraint conditions of minimize objective function  $J_m(U, V)$  and the necessary condition of  $J_m(U^*, V^*)$  as follows:

$$F = \sum_{i=1}^c \sum_{j=1}^n u_{ij}^m \|x_j - v_i\|^2 + \lambda \left( \sum_{i=1}^c u_{ij} - 1 \right) \quad (2)$$

$$u_{ij} = 1 / \sum_{q=1}^c \left( \frac{d_{ij}}{d_{iq}} \right)^{\frac{2}{m-1}} \quad (3)$$

$$v_i = \frac{\sum_{j=1}^n u_{ij}^m x_j}{\sum_{j=1}^n u_{ij}^m} \quad (4)$$

The specific steps of FCM clustering algorithm as follow:

Step 1: setting relevant parameters, include data set  $X$ , cluster number  $c$ , fuzzy exponent  $m$ , maximum number of iteration  $gmax$ , and initial clustering centers  $V$ .

Step 2: let  $t=1$ , initialize the membership matrix  $U$  by Eq.3, calculate the objective function  $J_m(U, V)$  by Eq.1.

Step 3: update clustering centers  $V$  by Eq.4.

Step 4: let  $t = t+1 < gmax$ , update membership matrix  $U$  and objective function  $J_m(U, V)$  in iteration  $t$  by Eq.3 and Eq.4, then turn to step 3.

Step 5: when  $t = gmax$ , output the smallest  $J_m(U^*, V^*)$  and the corresponding clustering centers  $V^*$ , which is the optimal cluster results.

## Sensitivity Analysis of FCM Algorithm

### Definition of FCM Algorithm Sensitivity

FCM is based on the minimization of objective function, the algorithm iterative formula and the calculate process imply that the algorithm initializes the clustering state with the initial clustering centers and each iteration move in the direction of the decreasing the objective function value. FCM use the gradient method to search for extreme values in the space. Therefore, the algorithm has a strong local search capability. At the same time, since the search direction of the gradient method always follows the direction of the objective function  $J_m$  decrease, so the initial clustering centers selection is very important, this means that when the initial clustering centers selected near a local minimum point, the algorithm is easy to fall into local minimum point and cannot get the global optimal solution.

The definition of FCM algorithm sensitivity is that different initial clustering centers leads to different clustering results and affecting algorithm converges to the global optimal solution; When reach the given clustering distance, the initial clustering centers affect the convergence speed of the algorithm.

### Experiment Analysis

#### Clustering Data

The three sets of experimental clustering data are shown in Fig.1. The data set 1 are distributed in fuzzy clusters in two-dimensional space as Fig.1 (a), the data set 2 are distributed in fuzzy clusters in three-dimensional space as Fig.1 (b), the data set 3 is randomly distributed in two-dimensional space as Fig.1(c). Each data set contains 5000 data point. In addition, the data set 3 is used to compare the move path of optimal clustering centers with data set 1 in next section, not to participate in the sensitivity experiment.

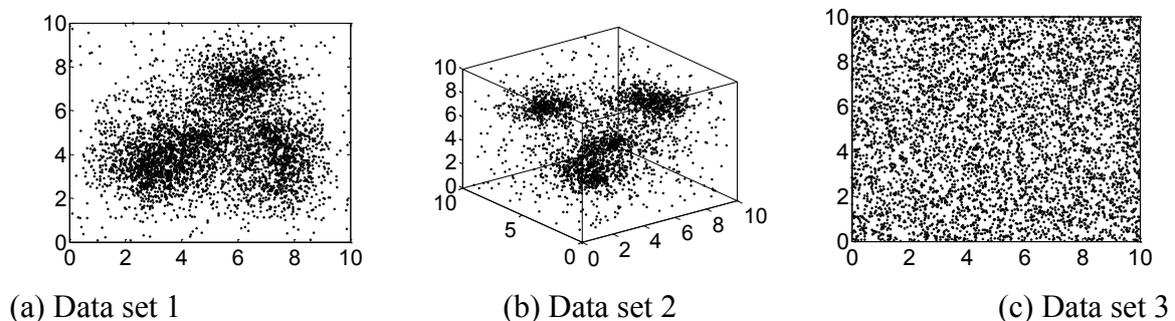


Figure. 1 Clustering Data

#### Sensitivity Experiments of Clustering Results

The experiment include 8 groups, the data set 1 and data set 2 are clustered into multiple classes respectively, the 50 groups of initial clustering centers are randomly generated under different condition of cluster number of each data set. In each experiment, the initial clustering centers changed in turn, and the other parameters unchanged. Compare the difference of the algorithm sensitivity to initial clustering centers in different clustering case. The clustering results include the clustering centers and the value of  $J_m$  which are output at the end of the calculation. Set the iterations number to 500 and the fuzzy exponent to 2, and the clustering results are shown in Table 1 and Table 2. The results reflects the initial clustering centers have no effect on the clustering results when the cluster number is small, and when the cluster number increasing, the clustering results begin to appear in many states, but the probability of convergence to the optimal solution is larger.

The experimental results show that the FCM algorithm is sensitive to the initial clustering centers cannot be generalized. The algorithm sensitivity of initial clustering center is not in the absolute sense, the sensitivity is directly related to the cluster number, when the number is small, the

algorithm is insensitive to the initial clustering centers, the calculation results always converge to the optimal clustering results; when the number increasing, the sensitivity is enhanced, but the probability of convergence to the optimal clustering results is relatively large. Therefore, in the process of clustering, the optimal clustering results can be obtained by performing FCM algorithms repeatedly on different initial clustering centers although the cluster number is large relatively.

Table 1 Experiment results of data set 1

Cluster number	Different results	Value of $J_m$	frequency
3	1	7758.765	50
10	1	1824.379	50
11	3	1643.554	33
		1645.280	12
		1646.210	5
12	4	1486.33	21
		1487.040	19
		1492.872	7
		1494.100	3

Table 2 Experiment results of data set 2

Cluster number	Different results	Value of $J_m$	frequency
3	1	17540.17	50
5	2	9251.080	44
		9274.423	6
8	2	5678.577	46
		5678.590	4
10	4	4468.630	28
		4471.633	18
		4486.631	2
		4503.699	2

### ***Sensitivity Experiments of Convergence Speed***

The experiment takes data set 1 for example, and set cluster number to 3 and convergence accuracy to 7758.77, generating 30 groups of clustering centers randomly. The convergence speed of different initial clustering centers is shown in Fig. 2. The FCM algorithm updates the clustering state by the iterative process, each iteration in the direction of decrease the value of  $J_m$ , the number of iterations of each initial centers reach the same precision is different, which affects the computation time. The results show that the selection of the initial clustering centers will directly affect the convergence speed of the FCM algorithm, although the algorithm is insensitive to the initial centers. The existing research on the optimization of FCM aims to overcome the sensitive problem, and find the global optimal solution, but ignore the computational efficiency of the algorithm. In the cluster process of big data, the sensitivity and the computational efficiency should simultaneously be considered in the optimization of the initial clustering centers.

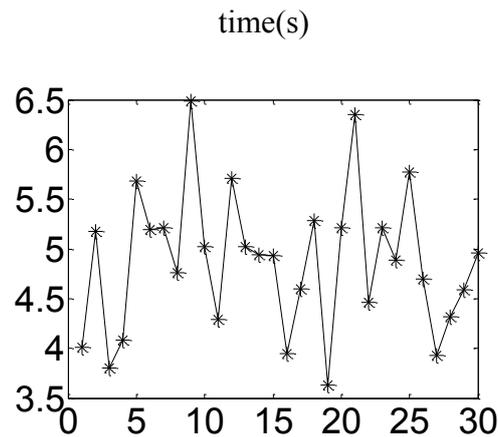


Figure. 2 Convergence speed of different initial clustering centers

## Initial Clustering Centers Optimization

### Maximum Distances Product

When the fuzzy exponent  $m=1$ , the FCM algorithm degrade to HCM algorithm, that is, the K-means algorithm. For the problem of The K-means is sensitive to the initial clustering centers, the maximum and minimum distance method reduce the sensitivity of the algorithm according to the density characteristics of data distribution, it improve the cluster accuracy and convergence rate to some extent[12]. Xiong proposed the maximum distances product method based on the maximum and minimum distance method, the selected clustering centers is more consistent with the data distribution characteristics, and overcomes the shortcomings of the former selected too dense, and can get better clustering results[13]. The selection steps are as follows:

Step 1: setting density parameter  $Minpts$  and  $\epsilon$ , find the high density point set  $D$ . Taking the spatial point  $X_i$  as the center, the radius that containing  $Minpts$  data object is called the density parameter of the object  $X_i$ , denoted by  $\epsilon$ . The larger the  $\epsilon$ , the lower the data density of the region in which the data object is located, on the contrary, the higher the data density of it.

Step 2: in the region, the data object of the highest density is taken as the first clustering center  $Z_1$ , and the high density point farthest from  $Z_1$  is taken as the second clustering center  $Z_2$ .

Step 3: calculating the distance from each remaining data points  $X_i$  to  $Z_1$  and  $Z_2$ , respectively. Taking data point  $X_i$  as the third clustering center  $Z_3$ , the  $Z_3$  satisfies  $\max(d(X_i, Z_1) \times d(X_i, Z_2))$ , and so on. The  $Z_k$  satisfies  $\max(d(X_i, Z_1) \times d(X_i, Z_2) \times \dots \times d(X_i, Z_{k-1}))$ ,  $k$  initial clustering centers can be obtained in turn.

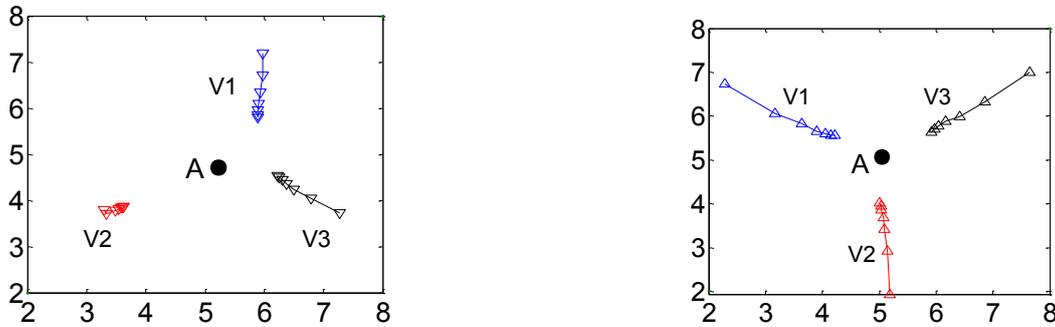
The maximum distances product method combine the data density and distribution characteristics to select the point close to the optimal clustering centers to initialize the clustering centers, which improves the accuracy and convergence rate of K-means clustering. But it is not suitable for FCM algorithm because of the influence of fuzzy exponent. In this paper, the maximum distance product method is adjusted to optimize the initial clustering centers of FCM algorithm in conjunction with the changing rule of the optimal clustering centers with the fuzzy exponent.

### Clustering Centers Move with Fuzzy Exponent

Both the K-means and FCM algorithm are based on the initial clustering centers to initialize partition state of the data set, and update the partition state through the data objects move between clusters in the iteration process, each iteration move in the direction of the decreasing the objective function value. The initial clustering centers initialization method is designed to reduce the possibility that the same data object is divided into different cluster during multiple iterations in the iterative process according to the distribution of the data set, and accelerate the convergence speed. In short, the closer the initial clustering centers to the optimal clustering centers, the more accurate the clustering results and the faster the algorithm convergence speed. Therefore, the selection of the initial clustering centers should be as close as possible to the optimal clustering centers. Clustering

Centers Move with Fuzzy Exponent (CCMFE) based on the maximum distance product method and combines the influence of fuzzy exponent on the optimal clustering centers, the selected initial clustering centers are consistent with the data distribution characteristics and the fuzzy degree.

In the FCM algorithm, the fuzzy exponent  $m$  controls the fuzzy degree of clustering and affects the final clustering results. Taking the data set 1 and data set 3 as examples, the fuzzy exponent range from 1 to 31, the step size is 5. The move path of optimal clustering centers  $V=(V_1, V_2, V_3)$  of the data sets is shown in Fig.3(1) and Fig.3(2), and  $A$  is the mean point of data set.



(a) Optimal clustering centers of data set 1

(b) Optimal clustering centers of data set 2

Figure. 3 Move path of optimal clustering centers

The larger the fuzzy index, the closer the clustering centers is to the mean point of the data set. The relationship between the optimal clustering centers of FCM  $V=(V_1, V_2, V_3, \dots, V_k)$  and K-means  $Z=(Z_1, Z_2, Z_3, \dots, Z_k)$  can be concluded from the Fig.3: the fuzzy exponent  $m$  controls the fuzzy degree of clustering, when the fuzzy exponent increases gradually,  $V_i$  ( $i = 1, 2, \dots, k$ ) moves from  $Z_i$  ( $i = 1, 2, \dots, k$ ) to the mean point  $A$  of data set. The moving distance is affected by two factors: fuzzy exponent and data point density. Firstly, the larger the fuzzy exponent, the larger the moving distance, and the optimal centers moving in deceleration with fuzzy exponent. Secondly, the greater the data point density, the slower the moving speed. Let  $d$  denote the distance from  $V_i$  to  $Z_i$ , then  $d$  can be expressed as Eq. 5:

$$d = a * \log_{1+\frac{1}{\theta}} m \tag{5}$$

In the Eq.5,  $\theta$  ( $\theta > 0$ ) denote search radius, it is the radius that contains  $N/k$  data objects. Considering the differences in the range of different attribute value of different data sets, taking the search radius  $\theta$  as the function base can be better fitted to the curve. Combined with the logarithmic curve characteristics, the base of logarithmic function must be greater than 1, meanwhile, The moving distance  $d$  is proportional to the search radius  $\theta$  and is inversely proportional to the base of logarithmic function, For the same data set, the search object is fixed, the larger the regional data density, the smaller the search radius  $\theta$ , the smaller the moving distance  $d$  of the optimal clustering centers, and the larger the base of logarithmic function; on the contrary, the smaller the regional data density, the larger the search radius  $\theta$ , the larger the moving distance  $d$  of the optimal clustering centers, and the smaller the base of logarithmic function.  $a$  is the adjustment coefficient.

Let  $A=\{a_1, a_2, a_3, \dots, a_n\}$  be the coordinates of the mean points of the data set,  $Z_i=\{z_{i1}, z_{i2}, z_{i3}, \dots, z_{in}\}$  be the coordinates of the  $i$  th clustering center when the fuzzy exponent equal to 1, then the moving direction of FCM optimal clustering center  $V_i$  is vector  $Z_iA$ . Let  $Y_i=\{y_{i1}, y_{i2}, y_{i3}, \dots, y_{in}\}$  be any point on the line  $Z_iA$ , The parametric equation of straight line  $Z_iA$  in  $n$ -dimension space can be obtained:  $y_{i1} = a_1t + z_{i1}, y_{i2} = a_2t + z_{i2}, \dots, y_{in} = a_nt + z_{in}$ . Then the vector equation of straight line can be obtained:  $Y_i = At + Z_i$ .

The main steps of clustering centers move with fuzzy exponent are: first, according to the same steps as the maximum distance product method, the high density point set  $D$  is found according to the density parameter, and the  $k$  initial clustering centers with fuzzy exponent  $m = 1$  are selected; second, the  $i$  th clustering center  $Z_i$  ( $i = 1, 2, \dots, k$ ) moves  $d$  unit length along the direction  $Y_i = At + Z_i$

toward the mean point  $A$  of data set. Then the optimized FCM initial clustering centers  $V=(V_1, V_2, V_3, \dots, V_k)$  is obtained.

## Conclusion

In view of the problem that the definition of initial clustering centers sensitivity is not clear, this paper provides the definition of algorithm sensitivity, and conduct experiment to analyze the sensitivity of the algorithm and compare the convergence speed of different initial clustering centers when reach the given clustering distance. The experimental results show that FCM algorithm is sensitive to the initial clustering centers is not in the absolute sense, the sensitivity is directly related to the cluster number for the same data set, when the number is small, the algorithm is insensitive to the initial clustering centers, the calculation results always converge to the optimal clustering results; when the number increasing, the sensitivity is enhanced, but the probability of convergence to the optimal clustering results is relatively large. For sensitive situation, through analyze the calculation principle and the relationship between the optimal clustering centers of FCM algorithm and K-means algorithm, combined with the effect of fuzzy exponent and the maximum distance product method, the clustering centers move with fuzzy exponent is proposed to optimize the FCM initial clustering centers. Based on the selection principle and validity of the maximum distance product optimization, analyze the effect of fuzzy exponent and data distribution density on the optimal clustering centers. Therefore, the initial clustering centers selected by the optimized algorithm are close to the optimal clustering centers, and conform to the data distribution characteristics and fuzzy degree. In view of the length of this paper, the validity of the convergence speed and clustering results of the optimized FCM algorithm need to be further studied.

## Acknowledgment

This research was supported by the National Social Science Foundation of China under Grant 14BTQ058.

## References

- [1] Kannan S R, Devi R, Ramathilagam S, etc. Effective FCM noise clustering algorithms in medical images, *J. Computers in Biology & Medicine*, 2013, 43 (2) 73-83.
- [2] Shokouhifar M, Jalali A. Optimized sugeno fuzzy clustering algorithm for wireless sensor networks, *J. Engineering Applications and Artificial Intelligence*, 2017, 60 16-25.
- [3] Verma H, Agrawal R K, Sharan A. An improved intuitionistic fuzzy c-means clustering algorithm incorporating local information for brain image segmentation, *J. Applied Soft Computing*, 2016, 46 (C) 543-557.
- [4] Bezdek J C, Hathaway R J, Sabin M J, etc. Convergence theory for fuzzy C-means: Counterexamples and repairs, *J. IEEE Transactions on Systems Man and Cybernetics*, 1987, 17 (5) 873-877.
- [5] WU Chen, XU Youquan, Maximum entropy kernel FCM text clustering algorithm by optimal center choice with point density, *J. computer engineering and design*, 2016,37 (9) 2411-2414.
- [6] Bezdek J C, Hathaway R J, Sabin M J, etc. Convergence theorem for the fuzzy ISODATA clustering algorithms, *J. IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1980, 2 (1) 1-8.
- [7] XIE Liangliang, LIU Jiansheng, ZHU Fan. Fuzzy C-means clustering algorithm based on Grey Wolf optimization, *J. Software Guide*, 2017, 16 (4) 28-30.
- [8] ZHANG Yongku, YIN Lingxue, SUN Jingguang. Fuzzy clustering algorithm based on the improved genetic algorithm, *J. CAAI Transactions on Intelligent Systems*, 2015, 10 (4) 627-634.

- [9] XU Youquan, WU Chen, YANG Xibei. Maximum entropy fuzzy C-means clustering based on sample weighting and initial cluster centers, *J. Computer System & Applications*, 2014, 23 (8) 139-143.
- [10] ZHANG Huizhe, WANG Jian. Improved fuzzy C means clustering algorithm based on selecting initial cluster centers, *J. Computer Science*, 2009, 36 (6) 206-209.
- [11] ZHOU Yueyue, HU Jie, SU Tao. Fuzzy cluster algorithm based on midpoint density function, *J. Journal of Computer Applications*, 2016, 36 (1) 150-153.
- [12] LAI Yuxia, LIU Jianping. Optimization study on initial center of K-means algorithm, *J. Computer Engineering and Applications*, 2008, 44 (10) 147-149.
- [13] XIONG Zhongyang, CHEN Ruotian, ZHANG Ruofang. Effective method for cluster centers' initialization in K-means clustering, *J. Application Research of Computers*, 2011, 28 (11) 4188-4190.