Computers, Materials & Continua CMC, vol.60, no.2, pp.767-780, 2019

Fuzzy C-Means Algorithm Automatically Determining Optimal Number of Clusters

Ruikang Xing[1, *] and Chenghai Li[1]

Abstract: In clustering analysis, the key to deciding clustering quality is to determine the optimal number of clusters. At present, most clustering algorithms need to give the number of clusters in advance for clustering analysis of the samples. How to gain the correct optimal number of clusters has been an important topic of clustering validation study. By studying and analyzing the FCM algorithm in this study, an accurate and efficient algorithm used to confirm the optimal number of clusters is proposed for the defects of traditional FCM algorithm. For time and clustering accuracy problems of FCM algorithm and relevant algorithms automatically determining the optimal number of clusters, kernel function, AP algorithm and new evaluation indexes were applied to improve the confirmation of complexity and search the scope of traditional fuzzy C-means algorithm, and evaluation of clustering results. Besides, three groups of contrast experiments were designed with different datasets for verification. The results showed that the improved algorithm improves time efficiency and accuracy to certain degree.

Keywords: Fuzzy C-means clustering, affinity propagation (AP) clustering, evaluation index, kernel function.

1 Introduction

As the important technology in data mining field, clustering analysis is widely applied in statistics, decision support, machine learning, pattern recognition, picture processing, spatial database technology and e-commerce, etc. It is a very efficient data analysis method. Classical clustering algorithms mainly include partition-based clustering, hierarchical clustering algorithm, density-based method, grid-based method, model-based method and analysis method based on isolated point, etc. The quality of clustering algorithms greatly influences the final results of clustering process.

Clustering process is an effective grouping of physical or abstract set of objects. The group generated in clustering results is called cluster. Cluster is the set of objects with certain same features in the database. The concrete manifestations include the following: any objects in the cluster have high similarity, while the objects which do not belong to the same cluster have relatively large dissimilarity. The value of similarity and dissimilarity can be calculated according to various attribute values of description objects. Usually, the distance between any objects is the measurement method which is mostly applied.

[1] Air and Missile Defense College, Air Force Engineering University, Xi'an, 710051, China.

[*] Corresponding Author: Ruikang Xing. Email: gm201808@tom.com.

As an important method which is widely applied in data analysis, clustering is used to classify the samples as per the specific standards, with the purpose of maximizing intra-category similarity and minimizing intercategory similarity. In clustering analysis, the key to deciding clustering quality is to determine the optimal number of clusters. At present, most clustering algorithms need to give the number of clusters in advance for clustering analysis of the samples. How to gain the correct optimal number of clusters has been an important topic of clustering validation study. The existing method to determine the optimal number of clusters is mainly the fuzzy C-means (FCM) algorithm.

2 Material and methods

2.1 Improved fuzzy C-means algorithm based on kernel function

2.1.1 Fuzzy C-means algorithm analysis

Clustering analysis aims to classify objects according to their different features, degree of intimacy and similarity. The boundary of relations among things is usually not clear (i.e. fuzzy relation), so the application of fuzzy method for clustering analysis becomes inevitable. Fuzzy clustering analysis has been successfully applied in large-scale data analysis, data mining, picture analysis, pattern recognition, information fusion and so on. And, various fuzzy clustering algorithms appear. Among the numerous fuzzy clustering algorithms, fuzzy C-means (FCM) clustering algorithm [Bezdek (1981)] is most widely and successfully applied. It is a clustering analysis method based on objective function. Membership degree of each object to be classified for all centers of clustering can be gained through optimizing objective function so as to decide the category of classification objects and reach the purpose of automatic classification [Chen, Li and Wang (2006)].

FCM clustering algorithm: the set of objects to be classified is set as:

$$A = \{A_1, A_2, \cdots A_n\} \tag{1}$$

wherein, each object has *m* characteristic indexes, and is set as:

$$A_j = (x_{j1}, x_{j2}, \cdots x_{jm}) \tag{2}$$

Now, the object set A is classified into c categories ($2 \le c \le n$). The matrix which consists of vectors of c centers of clustering is set as:

$$V = \begin{bmatrix} V_1 \\ V_2 \\ \cdots \\ V_c \end{bmatrix} = \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1m} \\ v_{21} & v_{22} & \cdots & v_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ v_{c1} & v_{c1} & \cdots & v_{cm} \end{bmatrix} \tag{3}$$

It is simplified as: $V = (V_1, V_2, \cdots V_c)^T$, in which

$$V_i = \{v_{i1}, v_{i2}, \cdots v_{im}\}, \quad i = 1, 2, \cdots c \tag{4}$$

To gain an optimal fuzzy classification, a best fuzzy classification can be chosen from the fuzzy classification space as per clustering norms. To calculate the appropriate fuzzy classification matrix U and center of clustering V, the objective function:

$$J(U,V)=\sum_{j=1}^{n}\sum_{i=1}^{c}(\mu_{ij})^{q}\parallel A_j-V_i\parallel^2 \tag{5}$$

is made to reach the minimum. Wherein, certain value can be taken for q (generally q=2). $\parallel A_j-V_i\parallel$ represents the distance between object A_j and the vector of i^{th} category of centers of clustering.

Usually, iterative operation is used to figure out the approximate solution of objective function given in Formula (5). The detailed steps are as follows:

Step 1: Choose the number of categories c, $2\le c\le n$; take a primary fuzzy classification matrix $U^{(0)}$ for gradual iteration, $l=0,1,2,\cdots$

Step 2: For $U^{(l)}$, calculate the center of clustering:

$$V^{(l)}=(V_1^{(l)},V_2^{(l)},\cdots V_c^{(l)})^T \tag{6}$$

in which:

$$V_i^{(l)}=\sum_{j=1}^{n}(\mu_{ij}^{(l)})^q A_j \Big/ \sum_{j=1}^{n}(\mu_{ij}^{(l)})^q \tag{7}$$

Step 3: Amend fuzzy classification matrix $U^{(l)}$; when $\forall i, A_j\ne V_i$

$$\mu_{ij}^{(l+1)}=\left[\sum_{k=1}^{c}\left(\frac{\parallel A_j-V_i^{(l)}\parallel}{\parallel A_j-V_k^{(l)}\parallel}\right)^{\frac{2}{q-1}}\right]^{-1} \tag{8}$$

If $\exists k$, $A_j=V_k$

$$\begin{cases}\mu_{ij}^{(l+1)}=1 & i=k\\ \mu_{ij}^{(l+1)}=0 & i\ne k\end{cases} \tag{9}$$

Step 4: Compare $U^{(l)}$ and $U^{(l+1)}$; if precision $\varepsilon>0$, $\max\{|\mu_{ij}^{(l+1)}-\mu_{ij}^{(l)}|\}\le\varepsilon$. $U^{(l+1)}$ and $V^{(l)}$ are the solutions; then, stop iteration. Otherwise, $U^{(l+1)}$和$V^{(l)}$, return to Step 2 for repetition.

The fuzzy classification matrix $U^{(l+1)}$ and the center of clustering $V^{(l)}$ gained from the above algorithm are locally optimal solutions relative to the number of categories c, initial fuzzy classification matrix $U^{(0)}$, ε and q.

Noise in dataset has a great influence on the whole clustering classification process. At present, many algorithms fail to process noise, thus leading to the influence on the dataset classification. Or, noise processing results are unsatisfactory. Noise data processing is too complex or there is no substantive influence of noise reduction. All these lead to some defects of FCM clustering algorithm in practical applications.

2.1.2 Objective function optimization based on kernel function

Kernel function is introduced to enhance optimizing ability of FCM clustering algorithm. It is supposed that the center of clustering v_k^{θ} in high-dimensional space can find primary image in the primary space. Then, the objective function changes to

$$J = \sum_{k=1}^{c}\sum_{i=1}^{n}(\mu_{ki})^m \| \theta(x_i) - \theta(v_k) \|^2 \tag{10}$$

According to Mercer kernel definition,

$$\begin{aligned} d^2(x_i, v_k) &= \| \theta(x_i) - \theta(v_k) \|^2 \\ &= K(x_i, x_i) + K(v_k, v_k) - 2\mathrm{K}(x_i, v_k) \end{aligned} \tag{11}$$

Meanwhile, Gaussian radial basis function ($K(x,x) = 1, \forall x \in X$) is used as the kernel function for simplification. Then, the objective function of improved fuzzy C clustering algorithm can change to

$$\begin{aligned} J &= \sum_{k=1}^{c}\sum_{i=1}^{n}(\mu_{ki})^m K(x_i, x_i) + K(v_k, v_k) - 2K(x_i, v_k) \\ &= \sum_{k=1}^{c}\sum_{i=1}^{n}(\mu_{ki})^m (1 - K(x_i, v_k)) \end{aligned} \tag{12}$$

Lagrange multiplication approach is used to the center of clustering and iterative formula of membership matrix:

$$\begin{aligned} \mu_{ki} &= \frac{(K(x_i, x_i) + K(v_k, v_k) - 2K(x_i, v_k))^{\frac{1}{1-m}}}{\sum_{r=1}^{c}(K(x_i, x_i) + K(v_k, v_k) - 2K(x_i, v_k))^{\frac{1}{1-m}}} \\ &= \frac{(1 - K(x_i, v_k))^{\frac{1}{1-m}}}{\sum_{r=1}^{c}(1 - K(x_i, v_k))^{\frac{1}{1-m}}} \end{aligned} \tag{13}$$

$$v_k = \frac{\sum_{i=1}^{n}(\mu_{ki})^m K(x_i, v_k) x_i}{\sum_{i=1}^{n}(\mu_{ki})^m K(x_i, v_k)} \tag{14}$$

The process of improving fuzzy C clustering algorithm is as follows:

Step 1: Initialize. Give weighted index m and the number of clustering categories $c(2 \le c \le n)$; set the parameter values of the chosen kernel function; set threshold value of iteration stop ε; initialize membership matrix $U^{(0)}$, iteration counter b=0.

Step 2: Work out $K(x_i, v_k)$

Step 3: Update membership matrix $U^{(b)}$.

Step 4: If $\| U^{(b)} - U^{(b+1)} \| < \varepsilon$ ($\| . \| < \varepsilon$ is an appropriate norm), stop updating membership matrix *U*, otherwise, make $U = U + 1$ and turn to Step 2.

2.1.3 Improved FCM algorithm and analysis

The division method of standard FCM clustering algorithm is based on the following criterion: The sum *E* of distance between each data object p and corresponding cluster center is minimum. The computational formula of *E* is

$$E = \sum_{i=1}^{k} \sum_{p \in C_i} d(p, o_i) \tag{15}$$

wherein, o_i is the center of cluster C_i; $d(.)$ is distance function; *E* is the minimum of sum of distance between all data objects p and corresponding cluster centers. When $k = 1$, time complexity of the algorithm is $O(n^2)$. After the kernel function is added, when the i^{th} initial center point $i \in [1, k]$, the computational formula of time complexity *t* of the algorithm is $t = (n+1)(i-1)$.

Time complexity T_i of the i^{th} initial center point is

$$T_i = m(n+1)(i-1) \tag{16}$$

Time complexity $O(T)$ of the algorithm is

$$O(T) = O(\sum_{i=1}^{k} T_i) = O(mk^2 n) \tag{17}$$

Thus, $O(T) = O(n^2)$. In conclusion, after the kernel function is added, the algorithm complexity of FCM clustering algorithm reduces.

2.2 Confirmation of search scope

Since clustering results of FCM clustering algorithm depend on the selection of initial center of clustering, different initial center of clustering will generate different clustering results. Thus, clustering results are unstable. How to determine the optimal clustering according to FCM algorithm is important.

Usually, the basic thought of determining the optimal number of clusters is as follows: for the specific dataset, conform the search scope of number of clusters and operate clustering algorithm to gain the clustering results of different number of clusters; choose appropriate validity indexes to evaluate clustering results, and confirm the optimal number of clusters according to the evaluation result. Thus, the core of confirming the optimal number of clusters is to confirm reasonable search scope of number of clusters and evaluation indexes of clustering effectiveness.

To confirm the search scope of number of clusters $[k_{\min}, k_{\max}]$, $k_{\min}$ and $k_{\max}$ should be confirmed. $k_{\min} = 1$ refers to even distribution of samples, without obvious characteristic difference. The minimum number of clusters in clustering algorithms is usually 2, i.e. $k_{\min} = 2$. There still no explicit theoretical direction about how to confirm $k_{\max}$. The empirical rule that most scholars use is: $k_{\max} \le \sqrt{n}$, which is described in the Literature [Yu and Cheng (2002)]. The conclusion is based on the precondition of uncertainty function $f(x) = x^{-1}$. But the precondition is not the sufficient condition proved by Literature [Yang, Li, Hu et al. (2006)]. The conclusion is deduced based on the precondition that the sample space has fractal geometrical characteristics, and the conclusion have no universality. Besides, sample size and practical category number of all datasets in Literature (Frey and Dueck 2008) also have no such property. Sample size and practical category number of some datasets in Literature [Brusco (2008)] also have no such property. To sum up, $k_{\max} \le \sqrt{n}$ is only an empirical rule, and does not own universality. In this study, AP algorithm proposed by Frey et al. is applied to confirm $k_{\max}$. The algorithm is fast and effective. It has been well applied in multiple fields.

2.2.1 AP clustering algorithm

AP clustering algorithm [Kapp (2007); Xiao and Yu (2008)] is a kind of clustering algorithm based on affinity information propagation. Its purpose is to find out the set of optimal category representatives so that the sum of similarities of all samples to the nearest category is largest. AP algorithm deems all N samples in the dataset as the candidate category representatives and establishes attraction degree information with other samples for each sample. In other words, similarity between any two samples x_i and x_k (when Euclidean distance is applied for measurement, $s(i,k) = -\| x_i - v_k \|^2$) is stored in $N \times N$ similarity matrix. AP algorithm applies $s(i,k)$ to express the appropriateness of sample x_k as sample x_i. It is initially supposed that the possibility of all samples chosen as category representative is same, that is, all $s(k,k)$ are set with the same value *p*. To pick out the appropriate category representative, it is necessary to gather relevant evidence from samples continuously. Thus, AP algorithm introduces two important information quantities parameters: reliability *r* and availability *a*. These two parameters represent different competition purpose $r(i,k)$ points to x_k from x_i. It represents the evidence from x_k, and expresses the appropriateness degree of x_k used as the category representative of x_i, and $a(i,k)$ points to x_i from x_k. It represents the evidence accumulated by x_i, and is used to express the appropriateness of x_i choosing x_k as the category representative. For any sample x_i, the sum of reliability of all samples $r(i,k)$ and availability $a(i,k)$ is calculated. The sample x_k involving the

largest sum is category representative. The iteration process of AP algorithm is the alteration and update process of two information quantities.

To prevent oscillation in the iteration process, AP algorithm introduces the factor λ to prevent oscillation, and the value of λ is between 0 and 1. The update result of $r(i,k)$ and $a(i,k)$ is gained through the weighing of current iteration value and the last iteration result.

2.2.2 AP feasibility analysis

AP algorithm does not give the number of clusters. When the algorithm ends, the number of clusters is determined automatically. For the clustering structure of intra-category compactness and inter-category alienation, AP algorithm can get the accurate clustering result. But for the close clustering structure, the algorithm tends to generate much local clustering. Thus, the number of clusters is generally large and the accurate clustering results cannot be given [Wang, Li, Zhang et al. (2007)]. Because of its fast speed and effectiveness, AP algorithm rather than C-means clustering algorithm is used to complete initial category number screening of dataset. Since the category number searched by AP algorithm is greater than $\sqrt{n}$, the maximum $k_{\max}$ of category number is reduced from n (sample size) to the number of clusters k_{AP} generated by AP algorithm. Compared with $\sqrt{n}$, the scheme involves clustering structure distribution of samples, which is scientific. The experiment also successfully verifies the feasibility of its scheme.

2.3 Confirmation of new evaluation index

At present, many validity evaluation indexes have been proposed to analyze clustering results for FCM algorithm, such as partition coefficient VPC [Bezdek (1974)], partition entropy VPE [Bezdek (1974)], VOS proposed by Kim et al. [Kim, Lee, Lee et al. (2004)] VXB index proposed by Xie et al. [Xie and Beni (1991)], VFS index proposed by Fukuyama et al. [Fukuyama and Sugeno (1989)], VK index pro-posed by Kwon [Kwon (1998)], VCWB index proposed by Rezaee et al. [Rezaee, Lelieveldt and Reiber (1998)], VB index proposed by Boudraa [Boudraa (1999)], VSV index proposed by Kim et al. [Kim, Park and Park (2001)], Wint (Weighted inter-intra) [Boudraa (1999)] and Silhouette [Silhouette (2004)] index. However, due to the defects of these indexes, it is hard for them to judge the clustering results. The clustering validity test effect is not ideal enough. Thus, geometric structure of datasets and clusters with different sizes are fully considered in this study. The specific value of intra-category compactness and inter-category separation degree is combined with clustering membership degree to define a new clustering validity index. Besides, the information of dataset and its geometric structure are fully considered. So, the optimal partition and the optimal number of clusters of fuzzy partition can be accurately confirmed by FCM algorithm. On this basis, a method to confirm the optimal number of clusters of samples is proposed to evaluate the clustering results of AP algorithm and determine the optimal number of clusters of samples.

2.3.1 Definition of compactness Index

Compactness index is used to measure intra-category compactness, and can be expressed with intra-category weighted squared error sum as follows:

$$V(c,U)=(\frac{c+1}{c-1})^{\frac{1}{2}}\sum_{i=1}^{c}\frac{\sum_{j=1}^{N(i)}\mu_{ij}{}^{m}\parallel x_j-V_i\parallel^2}{\omega_i} \tag{18}$$

wherein, $\omega_i=\sum_{j=1}^{N(i)}\mu_{ij}$ is the weight of each category, that is, a different weight corresponds to every different category. The significance of weight represents the support degree of each different category for dataset. $N(i)$ represents the number of data samples included in the i^{th} category.

As c increases, $(\frac{c+1}{c-1})^{\frac{1}{2}}$ and weight ω_i reduce. $\frac{1}{\omega_i}$ inhibits the reduction of measured value through weighing each category.

When the noise point is separately regarded as one category, $\frac{1}{\omega_i}=1$. At this moment, the weight of such category will be larger than other categories. To make compactness index more robust, $(\frac{c+1}{c-1})^{\frac{1}{2}}$ is added for adjustment.

2.3.2 Definition of separation index

Separation index is a method to measure separation degree of two fuzzy sets. Dispersion between two categories is defined as follows: the sample belongs to the minimum of membership degree of these two categories. Separation measurement uses the largest difference of all paired fuzzy clusters. So, similarity between two fuzzy sets F_i and F_j is defined as:

$$S'(F_i,F_j)=\max_{x_k\in X}(\min(\mu_{ik},\mu_{jk})) \tag{19}$$

Separation measurement of given fuzzy partition is

$$\begin{aligned}S(c,U)&=1-\max_{i\neq j}S'(F_i,F_j)\\&=1-\max_{i\neq j}(\max_{x_k\in X}(\min(\mu_{ik},\mu_{jk})))\end{aligned} \tag{20}$$

Then, the boundary of separation index is $0\le S(c,U)\le 1$; when $F_i=F_j$, $S(c,U)=0$.

2.3.3 New evaluation index

Since compactness and separation have different scalar quantity, normalization result can be expressed as:

$$V^N(c,U) = \frac{V(c,U)}{V_{\max}(c,U)} \tag{21}$$

$$S^N(c,U) = \frac{S(c,U)}{S_{\max}(c,U)} \tag{22}$$

These two formulas are effectively combined to get the new evaluation indexes of clustering effect:

$$O^N(c,U) = \frac{V^N(c,U)}{S^N(c,U)} \tag{23}$$

In the new validity indexes, compactness index $V(c,U)$ reflects intra-category total variation, and it is used to express the concentration degree of intra-category samples. When its value is smaller, compactness of category is better. Separation index $S(c,U)$ reflects intra-category total variation, and it is used to express the distance among fuzzy clusters. When separation is larger, the partition result is better. $V(c,U)$ and $S(c,U)$ are combined to reflect the partition features of dataset. When the indexes are smaller, intra-category is more compact, intra-category is more separate and the clustering result is better.

3 Results

To test validity and operation efficiency of the proposed algorithm, three groups of experiments were applied to carry out simulation test of artificial dataset and true dataset, and the algorithms were compared.

Table 1: Datasets and standard number of clusters

Dataset	Data source	Standard number of clusters
Dataset1	Artificial	2
Dataset2	Artificial	4
Dataset3	Artificial	13
Iris	Literature	3
Wine	Literature	3

There are three artificial datasets: Dataset1, Dataset2 and Dataset3. Dataset1 is composed of two two-dimensional Gaussian distribution data with the centers of (0, 0) and (20, 20) respectively. Each category has 400 samples. Dataset2 is composed of four two-dimensional Gaussian distribution data with the centers of (0, 0), (5, 7), (12, 17) and (19, 24) respectively. Each category has 400 samples. Dataset3 is the sample dataset generated artificially at random. The number of samples is 150. The true number of

clusters is 13. The true dataset is composed of UCI true datasets including Iris and Wine datasets. The standard number clusters, data and sources of artificial datasets and UCI true datasets are shown in Tab. 1.

Experiment 1: Validity experiment of kernel-based improved FCM algorithm

IRIS dataset and Wine dataset were chosen as the test samples. FMC clustering algorithm and improved FMC clustering algorithm were simulated. The change trend of objective function with iteration times is shown in the Fig. 1.

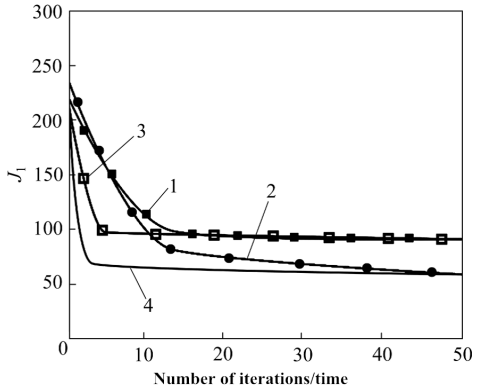


Figure 1: Convergence comparison chart of FCM algorithm based on improved kernel function

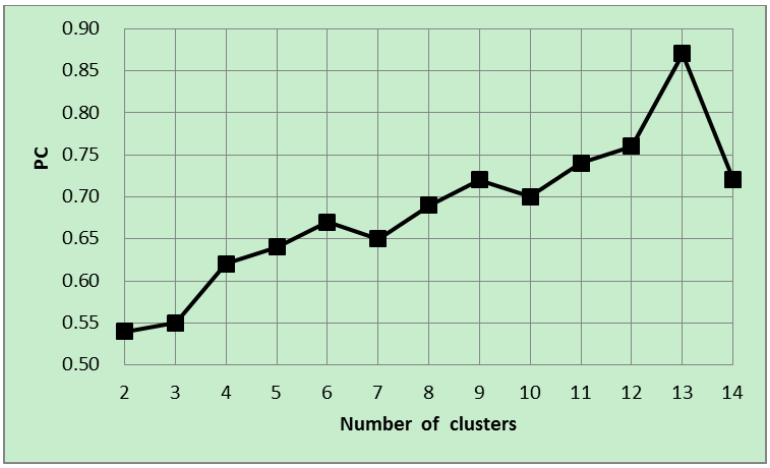
In the Fig. 1, Line 1 represents the first clustering of FCM algorithm; Line 2 represents the second clustering of FCM algorithm; Line 3 represents the first clustering of improved FCM algorithm; Line 4 represents the second clustering of improved FCM algorithm.

According to the figure, the iteration times of improved FCM algorithm is obviously lower than that of traditional FCM algorithm in the process where the target value tends to coincide. Thus, the validity of the algorithm is proved.

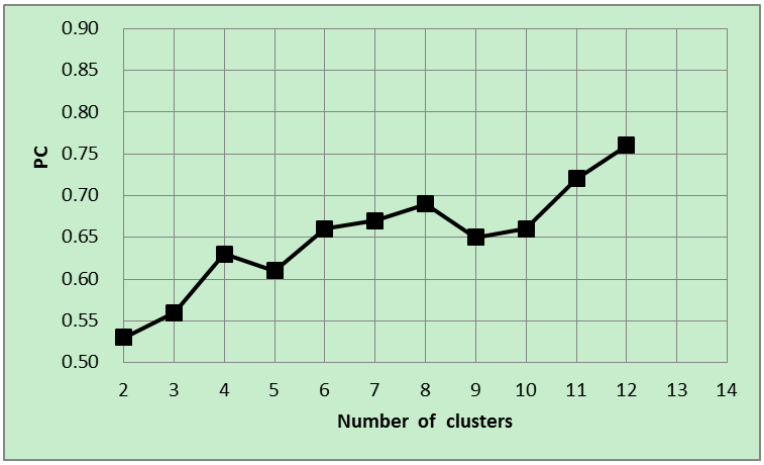
Experiment 2: Validity experiment of AP algorithm determining upper limit of search.

Artificial Dataset3 was chosen asexample. Simulation experiment was carried out for $k_{\max} = k_{ap}$ and $k_{\max} = \sqrt{n}$, respectively.

To eliminate the influence of this algorithm, PC was used as the evaluation index for contrast experiment. The optimal number of clusters is shown in the following figure.



(a) $k_{\max} = k_{ap}$



(b) $k_{\max} = \sqrt{n}$

Figure 2: Comparison chart of optimal number of clusters

It can be easily seen that, since the search scope confirmed by $k_{\max} = \sqrt{n}$ is smaller than the practical optimal number of clusters, the accurate optimal number of clusters cannot be gained. The accurate optimal number of clusters is obtained through AP algorithm.

Experiment 3: Comparison experiment of several validity indexes

Artificial Dataset1 and Dataset2 as well as UCI datasets Iris and Wine were selected to verify several representative evaluation indexes via comparisons. The results are shown in Tab. 2.

Table 2: Comparison of various indicators

Dataset	Optimal number of clusters	PC		PE		XB		ON	
		Number of clusters	Value	Number of clusters	Value	Number of clusters	Value	Number of clusters	Value
Dataset1	2	2	0.6736	2	0.5236	2	0.0437	2	0.7221
Dataset2	4	3	0.7360	2	0.5792	4	0.0469	4	0.8686
Iris	3	3	0.8894	2	0.0633	2	0.0536	3	1.0093
Wine	3	3	0.8722	2	0.1921	3	0.0459	3	0.8831

Judging from the above results, the evaluation indexes proposed in this study can be all converged and get the accurate number of clusters in 4 groups data. They perform better than other indexes. The theoretical research and experimental results indicate that, compared with other indexes and methods, the indexes in this study have better performance and stability.

5 Discussion

The above three groups contrast experiments verify the timeliness of improved algorithm, validity of search scope and the accuracy of evaluation indexes respectively. The result shows that the proposed fuzzy clustering algorithm automatically determining the number of clusters is reliable.

6 Conclusion

Based on the analysis of FCM algorithm, an accurate and efficient algorithm used to confirm the optimal number of clusters is proposed in this study to solve the defects of traditional FCM algorithm. The algorithm is improved in the aspects of reducing algorithm complexity, confirming search scope and constructing clustering validity index. In addition, multiple groups of contrast experiments verify the improvement of algorithm with higher efficiency and accuracy.

Despite some problems exiting in the algorithm, the future researches will be completed to improve the time efficiency, which is caused by mutual application of various algorithms in the process of automatically determining the number of clusters.

Acknowledgement: This research was financially supported by Natural Science Foundation of China (Grant No. 61703426) and Postdoctoral Science Foundation of China (Grant No. 2016M602996).

References

Bezdek, J. C. (1981): *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, New York.

Bezdek, J. C. (1974): Numerical taxonomy with fuzzy sets. *Journal of Mathematical Biology*, no. 11, pp. 57-71.

Bezdek, J. C. (1974): Cluster validity with fuzzy sets. *Journal of Cybernetics*, vol. 3, no. 3, pp. 58-72.

Boudraa, A. O. (1999): Dynamic estimation of number of clusters in data sets. *Electronics Letters*, vol. 35, no. 19, pp. 1606-1608.

Brusco, M. J. (2008): Comment on "clustering by passim messages between data points". *Science*, vol. 319, pp. 726.

Chen, S. L.; Li, J. G.; Wang, X. G. (2006): *Fuzzy Set Theory and Its Application*. Science Press, Beijing.

Chen, L.; Chen, C. L. P.; Lu, M. Z. A. (2011): Multiple-kernel fuzzy c-means algorithm for image segmentation. *IEEE Transactions on Systems Man and Cybernetics Part B-Cybernetics*, vol. 41, no. 5, pp. 1263-1274.

Dodis, Y.; Reyzin, L.; Smith, A. (2004): Fuzzy extractors: how to generate strong keys from biometrics and other noisy data. *Proceedings of the 2004 International Conference on the Theory and Applications of Cryptographic Techniques*, pp. 523-540.

Frey, B. J.; Dueck, D. (2008): Response to comment on "clustering by passing messages between data points". *Science*, vol. 319.

Fukuyama, Y.; Sugeno, M. (1989): A new method of choosing the number of clusters for the fuzzy C-means method. *Proceedings of the 5th Fuzzy Systems Symposium,* pp. 247-250.

Hussein, H. (2013): Joint CFO and time-varying channel estimation by particle filtering in OFDM systems. *3rd International Conference on Communications and Information Technology*, pp. 241-245.

Huang, H. C.; Chuang, Y. Y.; Chen, C. S. (2012): Multiple kernel fuzzy clustering. *IEEE Transactions on Fuzzy Systems*, vol. 20, no. 1, pp. 120-134.

Kwon, S. H. (1998): Cluster validity index for fuzzy clustering. *Electronics Letters*, vol. 34, no. 22, pp. 217.

Kapp, A. V.; Tibshirani, R. (2007): Are clusters found in one dataset present in another dataset? *Biostatistics*, vol. 8, no. 1, pp. 9-31.

Kim, D. W.; Lee, K. H.; Lee, D. (2004): On cluster validity index for estimation of the optimal number of fuzzy clusters. *Pattern Recognition*, vol. 37, no. 10, pp. 2009-2025.

Kim, D. J.; Park, Y. W.; Park, D. J. (2001): A novel validity index for determination of the optimal number of clusters. *IEICE Transactions on Information and Systems*, no. 2, pp. 281-285.

Mok, P. Y.; Huang, H. Q.; Kwok, Y. L.; Au, J. (2012): A robust adaptive clustering analysis method for automatic identification of clusters. *Pattern Recognition*, vol. 45, no. 8, pp. 3017-3033.

Rezaee, M. R.; Lelieveldt, B. P. F.; Reiber, J. H. C. (1998): A new cluster validity index for the fuzzy C-mean. *Pattern Recognition Letters*, vol. 19, no. 3/4, pp. 237-246.

Tan, K. S.; Isa, N. A. M. (2011): Color image segmentation using histogram thresholding-

fuzzy C-means hybrid approach. *Pattern Recognition*, vol. 44, no. 1, pp. 1-15.

Tsai, D. M.; Lin, C. C. (2011): Fuzzy C-means based clustering for linearly and nonlinearly separable data. *Pattern Recognition*, vol. 44, no. 8, pp. 1750 -1760.

Wang, K. J.; Li, J.; Zhang, J. Y.; Tu, C. Y. (2007): Semi-supervised affine propagation clustering. *Computer Engineering*, vol. 33, no. 23, pp. 197-201.

Xiao, Y.; Yu, J. (2008): Semi-supervised clustering based on affinity propagation algorithm. *Journal of Software*, vol. 19, no. 11, pp. 2803-2813.

Xie, X. L.; Beni, G. A. (1991): Validity method for fuzzy clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 8, pp. 841-847.

Yu, J.; Cheng, G. S. (2002): Search range of optimal clustering number in fuzzy clustering method. *Scientific Science: Series E*, vol. 32, no. 2, pp. 274-280.

Yang, S. L.; Li, Y. S.; Hu, X. X.; Pan, R. Y. (2006): Study on k-value optimization problem in K-means algorithm. *Systems Engineering-Theory & Practice*, no. 2, pp. 97-101.

Zhang, F. T.; Sun, Y. X.; Zhang, L.; Geng, M. M.; Li, S. J. (2011): Research on certificateless public key cryptography. *Journal of Software*, vol. 22, no. 6, pp. 1316-1332.