Available online www.jocpr.com

Journal of Chemical and Pharmaceutical Research, 2014, 6(5):1743-1749

Research Article

ISSN : 0975-7384
CODEN(USA) : JCPRC5

Study on swarm optimization clustering algorithm

Zuo Ying Liu and Xian Wen

Southwest University Chongqing, China

ABSTRACT

This article studies the application of swarm algorithm to fuzzy c-means clustering algorithm. The author proposes the fuzzy kernel hierarchical clustering algorithm based on swarm optimization. This improved algorithm uses the kernel function method and the cut set factor is added to it. The algorithm optimizes the target performance function and uses binary tree split method to cluster data sample what. The experimental results show that the algorithm can effectively overcome the weaknesses of FCM algorithm.

Key words: swarm algorithm; clustering; algorithm optimization; Particle Swarm Optimization algorithm (PSO)

INTRODUCTION

1.1 The concept of clustering algorithms

Clustering refers to the process of dividing the collection consisting of multiple data sample elements into multiple set categories composed of similar sample elements according to certain rules. [1] Clustering analysis means dividing the data elements set reasonably according to certain classification rules in order to determine the category each element belongs to. In different clustering algorithms, Euclidean distance, vector Angle cosine and some other measurement methods are used to describe different similarity function. [2].

1.2 Hard c-means clustering algorithm

Among clustering algorithms, fuzzy clustering algorithm based on the objective function has been widely applied in practice. Fuzzy c clustering algorithm[3] is one fuzzy clustering algorithm which is more mature in theoretical research and more widely used among clustering algorithms based on target function. Fuzzy c clustering algorithm can be achieved by improving the objective function of hard clustering algorithm. Suppose $X = \{x_1, x_2, x_3, \cdots, x_n\}$ represents a set of limited measure data samples of n model in space, $x_k = (x_{k1,} x_{k2,} x_{k3,} \cdots, x_{ks})$ represents the eigenvectors of measure sample data x_k, x_{kj} is the data value of x_k in the j-dimension feature. The clustering analysis on a given sample set X of measurement data is c partition on the sample set X and the process to obtain c clustering results.

Suppose $U = [u_{ik}]_{c \times n}$ is a clustering partition matrix, $V = \{v_1, v_2, v_3, \cdots, v_n\}$ is the clustering center and c is its clustering number, then the objective function of hard c-means clustering can be expressed in the following equation:

$$\begin{cases} J_1(U,V) = \sum_{k=1}^{n} \sum_{i=1}^{c} \mu_{ik} (d_{ik})^2 \\ s.t.\ U \in M_{hc} \end{cases}$$

The hard c partition space corresponding to the data sample set X is expressed in the following equation:

$M_{hc} = \{U \in R^{cn} \mid \mu_{ik} \in \{0,1\}, \forall i,k; \sum_{i=1}^{c} \mu_{ik} = 1, \forall k; 0 < \sum_{k=1}^{n} \mu_{ik} < n, \forall i\}$ In the equation, n represents the number of X elements in the dataset while c is the number of clustering centers（1<c<n） and d_{ik} is the distortion degree between the data sample point x_k and the clustering center V_i and it is measured by distance. $V_i \in R^s (1 \le i \le c)$. μ_{ik} is the membership value of the k data sample point belonging to the i clustering center. The bigger the membership value is, the higher degree the similarity between the data sample point and its clustering center is, the more easily the data sample point is divided into the category determined by the clustering center. The error sum squares between data sample points in all categories and the typical data sample points is expressed as the relation $J_1(U,V)$.

The main procedure of HCM (hard c-means algorithm) is as follows:
The initialization of algorithm: c is the number of clustering classification, 2 ≤ c ≤ n,n represents the number of data in the data sample. Set the iteration stop precisionε >0 and initialize V(0), the iteration counter j=0;
Step 1: Calculate the following update matrix

$$\mu_{ik}^{(j)} = \begin{cases} d_{jk}^{(j)} = \min\limits_{1 \le i \le c} \{d_{jk}^{(j)}\} \\ 0 \quad other \end{cases} \qquad (1-1)$$

Step2: Calculate the update matrix V(j+i);

$$V^{j+1} = \frac{\sum_{k=1}^{n} \mu_{ik}^{(j+1)} . x_k}{\sum_{k=1}^{n} \mu_{ik}^{(j+1)}}, i = 1, 2, 3 \cdots c \qquad (1-2)$$

Step3: if $\|V^{(j)} - V^{(j+1)}\| < \varepsilon$, stop the calculation; or set j=j+1, return to step 1.
The algorithm can also take initializing membership matrix U (0) in clustering algorithm as a condition to start and end. The procedure is similar to the above steps, only that the membership function is used as the criterion function of clustering.

Fuzzy c-means clustering algorithm

In hard c-means clustering algorithm, it can be found that its membership μ_{ik} is 1 or 0. So whether the partition is reasonable or not, each data sample can always be incorporated into a category. The weakness of the algorithm is that it cannot show clearly the relation between data samples and clustering center, and in practice it is hard to find a problem needed to be distinguished so strictly. To deal with this kind of problems more effectively, the concept of fuzzy set is introduced and the clustering algorithm FCM based on object function is proposed.

Fuzzy set theory is to extend the range of membership function values in hard c partition from {0, 1} to the closed interval [0, 1] and deduce the fuzzy c partition, then this clustering partition space is shown as follows:

$$M_{fc} = \{U \in R^{cn} \mid \mu_{ik} \in [0,1], \forall i,k; \sum_{i=1}^{c} \mu_{ik} = 1, \forall k; 0 < \sum_{i=1}^{c} \mu_{ik} < n, \forall i\} \qquad (1\text{-}3)$$

$$\begin{cases} J_m(U,V) = \sum_{k=1}^{n} \sum_{i=1}^{c} \mu_{ik}^{m} (d_{ik})^2 \\ s.t. \quad U \in M_{fc} \end{cases} \qquad (1\text{-}4)$$

Its main idea is adding a fuzzy weighted index m in fuzzy clustering objective function to control the fuzzy degree of matrix U. The bigger m is, the higher fuzzy degree of the target function is. The clustering's object is to get the minimum of the mathematical expression $\min J_m(U,V)$;

$$\min\{J_m(U,V)\} = \min(\sum_{k=1}^{n} \sum_{i=1}^{c} \mu_{ik}^{m} (d_{ik})^2) = \sum_{k=1}^{n} \min\{\sum_{i=1}^{c} \mu_{ik}^{m} (d_{ik})^2\} \qquad (1\text{-}5)$$

Namely, when meeting the equation $\sum_{i=1}^{c} \mu_{ik} = 1$, get the extreme value of the above equation so that the multiplication method Lagrange can be used to get the extreme value of $J_m(U,V)$ when μ_{ik} satisfies the constraint conditions $\sum_{i=1}^{c} \mu_{ik} = 1$;

$$F = \sum_{i=1}^{c} \mu_{ik}{}^{m} (d_{ik})^2 + \lambda(\sum_{i=1}^{c} \mu_{ik} - 1) \quad (1\text{-}6)$$

The conditions for having extreme values are: $\frac{\partial F}{\partial \lambda} = \sum_{i=1}^{c} \mu_{ik} - 1 = 0$ (1-7)

$$\frac{\partial F}{\partial \mu_{jt}} = [m(\mu_{jt})^{m-1}](d_{jt})^2 - \lambda = 0 \quad (1\text{-}8)$$

From the above equations, the following equation can be got: $\mu_{jt} = [\frac{\lambda}{m(d_{jt})^2}]^{\frac{1}{m}}$ (1-9)
Plug the above equation into (1-6), the following equations can be got:

$$\sum_{l=1}^{c} \mu_{lt} = \sum_{l=1}^{c} (\frac{\lambda}{m})^{\frac{1}{m}} [\frac{1}{(d_{lt})^2}]^{\frac{1}{m-1}} = (\frac{\lambda}{m})^{\frac{1}{m-1}} \{\sum_{l=1}^{c} [\frac{1}{(d_{lt})^2}]^{\frac{1}{m-1}} = 1 \quad \lambda \quad (1\text{-}10)$$

$$(\frac{\lambda}{m})^{\frac{1}{m-1}} = \frac{1}{\sum_{l=1}^{c} [\frac{1}{(d_{lt})^2}]^{\frac{1}{m-1}}} \quad 1 \quad (1\text{-}11)$$

In the process of calculating d_{jk}, the value of d_{jk} can be 0, so analyze it under the two cases when d_{jk} is 0 or d_{jk} is not 0. For $\forall k$.

Similarly, the value of V(i) can be got in the same way when the expression is minimum. Set $\frac{\partial J_m(U,V)}{\partial V_i} = 0$
The following equation can be got:

$$V_i = \frac{\sum_{k=1}^{n} (\mu_{ik})^m x_k}{\sum_{k=1}^{n} (\mu_{ik})^m} \quad 1\text{-}12)$$

It can be seen from (1-14) that the object function of fuzzy c-means clustering algorithm is the same as that of hard c-means clustering algorithm when m equals 1. Therefore, fuzzy c-means clustering algorithm is the same as hard c-means clustering algorithm when m equals 1. When $m \to \infty$, the bigger m is, the larger the fuzzy degree of the clustering results in fuzzy clustering algorithm is. The clustering results are the most fuzzy when m is infinite. Matrix U in fuzzy c-means clustering algorithm corresponds to the fuzzy classification of data sample set X.

FCM algorithm is sensitive to the choice of initial value and easily falls into local extremum points. To improve FCM algorithm, fuzzy kernel hierarchical clustering algorithm is proposed in the article.

2 Fuzzy kernel hierarchical clustering algorithm

The improved FCM algorithm proposed in the article is called fuzzy kernel hierarchical clustering algorithm based on particle swarm algorithm. The main improvement is that the algorithm integrates the kernel clustering method to solve the nonlinear inseparable problems. In order to achieve the quickness of clustering, the assembly operator and the binary tree split are added to obtain hierarchy clustering and optimize the performance of target function by using particle swarm algorithm.

2.1 Kernel function method

With the rapid development of support vector machine theory, it has broad application in practice. Kernel function method is a method mapping the data sample from the initial input space Rp to high-dimensional feature space Rq through the use of nonlinear transformation $\phi(\cdot)$ and doing research in the high dimensional feature space [11]. If the relationship between the various elements in the data set in the algorithm is only doing mathematical inner product calculation, the specific mathematical form of $\phi(\cdot)$ need not be known. To get the corresponding nonlinear algorithm in the original input space, only kernel function with Mercer nature is needed to substitute the inner product form in the algorithm. The calculation becomes very simple and convenient through kernel function with Mercer nature

Definition 1 Gram Matrix [12]: set that a given function meets the mathematical relation $k : x^2 \to k$ (in the relation, K equals C or R) and $x_1 x_2, \cdots, x_m \in X$, then m ×m matrix; $K = (k(x_i, x_j))_{ij}$ is called the nuclear

matrix K for $x_1 x_2, \cdots, x_m$.
Definition 2 Positive Definite Matrix[12]

Positive Definite Matrix for a matrix Kij of m ×m, suppose all $c_i \in C$ meet $\sum_{i,j=1}^{m} c_i c_j k_q \geq 0$, then the matrix is Positive Definite Matrix.

Definition 3 Positive Define Kernel: Set X is a nonempty set and has a function defined in $X \times X$, a Positive Definite Gram Matrix is generated for all $m \in N$ m ∈N² and $x_1 x_2, \cdots, x_m \in X$, then the function is called Positive Define Kernel. The equation $(T_k f)(x) = \int_x k((x, x^{'}) f(x^{'})) dx^{'}$ generates operator T_k, then function k is called kernel function of T_k.

Definition 4 Mercer kernel: Set data sample set $x_k \in R^N, k = 1, 2, \cdots, l$, is mapped to feature space H through nonlinear transformation $\phi(\cdot)$ and get the relation $\phi_1(x), \phi_2(x), \cdots, \phi_l(x),$ then input inner product operation between space data samples after mapping to feature space H is expressed with Mercer kernel as $K(x_i, x_j) = (\phi(x_i \cdot x_j))$

All data sample forms a kernel matrix $K_{i,j} = K(x_i, x_j)$, kernel clustering method is using Mercer kernel to achieve the mapping transformation of two spaces. First map sample input space form to the feature space form, then cluster and analyze the samples in the feature space. Any function K can get the characteristic function and characteristic value $(\phi_i(x), \lambda_i)$ of the kernel function K as long as it meets Mercer condition. Its kernel function can be expressed as $K(x, y) = \sum_{i=1}^{N_H} \lambda_i \phi_i(x) \phi_i(y),$ N_H in the equation denotes the dimensions of the feature space. Similarly, $\phi(\cdot)$ can be expressed in the following equation:

$$\phi(x) = (\sqrt{\lambda_1}\phi_1(x), \sqrt{\lambda_2}\phi_2(x), \cdots, \sqrt{\lambda_{N_H}}\phi_{N_H}(x))^T \quad (2\text{-}1)$$

Euclid distance $d_H(x, y)$ can be expressed in the following equation:

$$d_H(x, y) = \sqrt{k(x, x) - 2k(x, y) + k(y, y)} \quad (2\text{-}2)$$

Gaussian kernel function meeting Mercer condition is used;

$$K(x, y) = \exp(-\beta \|x - y\|^2), \beta > 0 \quad (2\text{-}3)$$

Then K (X , X) = 1,the equation(6-20) can be changed into the following equation:

$$d_H(x, y) = \sqrt{2) - 2k(x, y)} \quad (2\text{-}4)$$

2.2 Cut Set factorλ

Mapping the original data sample set to high-dimensional feature space through kernel function and adding trap/Cut Set factor λ to the high-dimensional feature space at the same time solves the ownership problem of data samples when the membership $\mu_{ik}(i = 1, 2, \cdots, c)$ are close to each other, at the same time enhances the convergence capacity of the algorithm and accelerates the convergence speed of the algorithm.

Definition 5 Suppose $A \in X$ denotes a fuzzy set in X, and $\lambda \in [0,1]$. If A meets the mathematical expression $A_\lambda = \{x \mid \mu_A \geq \lambda\}$, then set A is called trap/Cut Set factor λ of A. Set the dividing matrix of $X \subset^{s \times n}$ is expressed $\lambda \in [0,1], 1 \leq i \leq c, Z = [z_{ik}]_{c \times n}$

as $U = [\mu_{ik}]_{c \times n}$ and $z_{ik} = \begin{cases} \mu_{ik} & \mu_{ik} \geq \lambda \\ 0 & \mu_{ik} \leq \lambda \end{cases}, U_k^\lambda = \{\mu_{jk} \mid \mu_{jk} \geq \lambda, j = 1, 2, \cdots, c\}$ generates fuzzy c partition, then it must meet the following conditions: $\sum_{i=1}^{c} \mu_{ik} = 1, 1 \leq k \leq n$

$$w_{ik} = \begin{cases} 1, U_k^\lambda \neq \phi \wedge z_{ik} \in U_k^\lambda \\ 0, U_k^\lambda \neq \phi \wedge z_{ik} \in \overline{U_k^\lambda} \\ \mu_{ik} \quad U_k^\lambda = \phi \end{cases} \quad (2\text{-}5)$$

In the above equation, λ = 0.5 + 1/ ac,a > 0 denotes a positive trap/Cut Set factor. In the general cases, the number of categories of data samples after clustering is $c \geq 2$ .When a = 1, $\lambda \in [0.5,1]$; When a = 2,λ = 0.5 + 1/ 2c.

2.3 Optimization of object function

Convert data sample set from the input space into high-dimensional feature space Rq through nonlinear mapping $\phi(\cdot)$ and use Euclid distance in feature space at the same time, then the expression of objective function of fuzzy kernel clustering can be expressed as:

$$\begin{cases} J_k(U,V) = \sum_{j=1}^{n}\sum_{i=1}^{c} \overline{\mu}_{ji}{}^m d_{kji}(x_i, x_j) \\ \quad = \sum_{j=1}^{n}\sum_{i=1}^{c} \overline{\mu}_{ji}{}^m \|\phi(x_i) - \phi(v_j)\|^2 \\ = \sum_{j=1}^{n}\sum_{i=1}^{c} \overline{\mu}_{ji}{}^m (k(x_i, x_i) - 2k(x_i, y_j) + k(v_j, v_j)) \\ 2 \leq c \leq n, m = 2 \end{cases} \quad (2\text{-}8)$$

Optimize the objective function by using particle swarm algorithm.

Use Gaussian kernel function $K(x,y) = \exp(-\beta\|x-y\|^2), \beta > 0$. Similarly, according to the requirement of the FCM algorithm, the membership function of fuzzy kernel clustering algorithm has to meet the following equation:

$$\overline{\mu}_{ji} = \frac{\left(\frac{1}{d_{jk}(x_i, \overline{v_j})}\right)^{\frac{1}{m-1}}}{\sum_{l=1}^{c}\left(\frac{1}{d_{jk}(x_i, \overline{v_j})}\right)^{\frac{1}{m-1}}} \quad (2\text{-}9)$$

Introduce Cut Set factor λinto high-dimensional space andλ = 0.5 +1/ 2c

$$\overline{z}_{ik} = \begin{cases} \overline{\mu}_{ik} & \overline{\mu}_{ik} \geq \lambda \\ 0 & \overline{\mu}_{ik} < \lambda \end{cases} \quad (2\text{-}10)$$

$$\overline{U}_k^\lambda = \{\overline{\mu}_{jk} \mid \overline{\mu}_{jk} \geq \lambda, j = 1, 2, \cdots, c\} \quad (2\text{-}11)$$

$$\overline{w}_{ik} = \begin{cases} 1, \overline{U}_k^\lambda \neq \phi \wedge \overline{z}_{ik} = \max\{\mu_{ik} \mid \mu_{ik} \in U_k^\lambda\} \\ 0, \overline{U}_k^\lambda \neq \phi \wedge \overline{z}_{ik} = \max\{\mu_{ik} \mid \mu_{ik} \in \overline{U_k^\lambda}\} \\ \overline{\mu}_{ik} \quad \overline{U}_k^\lambda = \phi \end{cases} \quad (2\text{-}12$$

Then the clustering center in high-dimensional feature space is expressed as:

$$\phi(\overline{v}_j) = \frac{\sum_{i=1}^{n} \overline{\mu}_{ji}^m \phi(x_i)}{\sum_{i=1}^{n} \overline{\mu}_{ji}^m} \quad \forall j = 1, 2, \cdots, c \quad (2\text{-}13)$$

After calculating, the following results are got:

$$K(x_i, \overline{v}_j) = \phi(x_i) \cdot \phi(\overline{v}_j) = \frac{\sum_{k=1}^{n} \overline{\mu}_{jk}^m K(x_k, x_i)}{\sum_{k=1}^{n} \overline{\mu}_{jk}^m} \quad (2\text{-}14)$$

$$K(\overline{v}_j, \overline{v}_j) = \phi(\overline{v}_j) \cdot \phi(\overline{v}_j) = \sum_{k=1}^{n}\sum_{l=1}^{n} (\overline{\mu}_{jk})^m (\overline{\mu}_{jl})^m K(x_i, x_l) / (\sum_{l=1}^{n} (\overline{\mu}_{jk})^m)^2 \quad (2\text{-}15)$$

Therefore, the membership function in high-dimensional feature space is expressed as:

$$\overline{\mu}_{ji} = \frac{\left(\frac{1}{d_{kji}(x_i, \overline{v_j})}\right)^{\frac{1}{m-1}}}{\sum_{l=1}^{c}\left(\frac{1}{d_{kji}(x_i, \overline{v_j})}\right)^{\frac{1}{m-1}}} = \frac{(1/k(x_i, x_i) - 2k(x_i, \overline{v}_j) + k(v_j, v_j))^{\frac{1}{m-1}}}{\sum_{j=1}^{c}(1/k(x_i, x_i) - 2k(x_i, \overline{v}_j) + k(v_j, v_j))^{\frac{1}{m-1}}} \quad (2\text{-}16)$$

2.4 Hierarchical Clustering method

Hierarchical Clustering method is a clustering method widely used in practice and its related theory is in constant development. It can be divided into the bottom-up condensing hierarchical clustering method and the top-down split hierarchical clustering method according to the different directions of the decomposition in hierarchical clustering method [13]. The main principle of split hierarchical clustering method is ascribing all the data objects to a category first, then dividing the category into two according to some rules and repeating the same dividing method in the newly-generated categories until the algorithm meets certain end conditions. The algorithm in this article uses the idea of binary tree split algorithm and combines with the improved fuzzy kernel clustering algorithm. In the algorithm, m=2,c=2,and the clustering center of the binary tree which is less than the minimum $D_{v\min}$ or the maximum depth of the binary tree which is greater than the maximum TLmax is used as the end conditions of the algorithm. In order to improve the speed of the algorithm, 5 is set as the maximum depth, merge and adjust the clustering centers according to their distances when the algorithm ends. Figure 2.1 is the demo figure of fuzzy hierarchical clustering algorithm based on particle swarm algorithm.

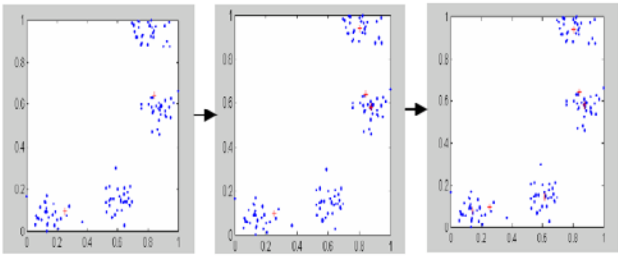

Figure 2.1 Demo Figure of Kernel Hierarchical Clustering Process

2.5 The experimental data and analysis

IRIS data are used to test the different performances of FCM algorithm and fuzzy clustering algorithm. IRIS data are a standard test sample set. IRIS data sample set are composed of 150 data sample points and they represent a four-dimensional data set. Each data sample set is represented respectively by four components Petal Length,Petal Width,Sepal Length and Sepal Width. In the meantime, the whole data sample set is composed of three IRIS categories Setosa,Versicolor and Virginica. Each category is composed of 50 samples.

$K(x, y) = \exp(\quad \beta\, x \quad y)$, $\beta > 0$. The aim of this experiment is using fuzzy kernel hierarchical clustering algorithm based on particle swarm algorithm and FCM to cluster IRIS data set, compare their clustering accuracy and speed. The results are shown as follows:

Table 1 Results of clustering IRIS data

Clustering algorithm	First category	Second category	Third category	Average number of iterations	Time(s)
FCM	0	3	13	24,5	0．1620
IRIS	0	1	1	14．4	0.1036

To overcome the weaknesses of FCM algorithm, the fuzzy kernel hierarchical clustering algorithm based on particle swarm algorithm is proposed. Compared with FCM algorithm, the new algorithm is guided by the principle of maximum membership and it uses the improved particle swarm algorithm to optimize objective function. It not only maintains the advantages of fuzzy clustering but also improves the convergence speed and prevents local optimum. The algorithm applies the idea of kernel function and the hierarchical clustering of binary tree improves the classification speed and accuracy. Experimental results prove that its performance is superior to the latter /FCM algorithm.

CONCLUSION

This article analyzes the weaknesses of Fuzzy C clustering algorithm and proposes an improved clustering algorithm----fuzzy kernel hierarchical clustering algorithm based on particle swarm algorithm. The main improvement is that the algorithm determines clustering number automatically and introduces kernel clustering algorithm. Assembly operator operation is added to the algorithm and the algorithm uses the improved particle swarm optimization (pso) algorithm for global optimization. The experimental results show that the algorithm can effectively overcome the weaknesses of FCM algorithm.

REFERENCES

[1] Gao Xinbo Xie Weixin.. *Chinese Science Bulletin* ,**2010**,11(44):2241-2247
[2]liu xinzhao. Research & Application of Clustering Algorithm.[D].Changsha Uunibersity of Science &Technology,**2013**.
[3]He ling. *research and application of computer* **2007**(1):10-13
[4]Dun J C. *IEEETrans.SMC*,**1974**,4(3):310-313
[5]Bezdek J C .Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press,New York,**1981**
[6]Li R P,Mukaidono M.A maximum-entropy approch to fuzzy clustering. FUZZ-IEEE'95,**1995**.2227-2232.
[7]Bezdek J C. *Pattern Recognition*, **2010**,25(2):211-217.
[9]Bezdek J C, Hathaway R J,et al. *IEEE Trans. PAMI*,2011,17(5):873-877.
[10]Li Zhongwei. *Harbin Engineering University*,**2010**.12(1): 7-9.
[11] Li Ying Zhang Yanning. *Computer engineering and Applications*,**2010** (17):4-
[12]Shengjie et. *Computer engineering and Applications* **2012**,43(31):175-177.