TSINGHUA SCIENCE AND TECHNOLOGY
ISSN 1007-0214 09/21 pp50-53
Volume 11, Number 1, February 2006

Fuzzy Clustering with Novel Separable Criterion*

YIN Zhonghang (尹中航), TANG Yuangang (唐元钢),
SUN Fuchun (孙富春)**, SUN Zengqi (孙增圻)

Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China

Abstract: Fuzzy clustering has been used widely in pattern recognition, image processing, and data analysis. An improved fuzzy clustering algorithm was developed based on the conventional fuzzy *c*-means (FCM) to obtain better quality clustering results. The update equations for the membership and the cluster center are derived from the alternating optimization algorithm. Two fuzzy scattering matrices in the objective function assure the compactness between data points and cluster centers, and also strengthen the separation between cluster centers in terms of a novel separable criterion. The clustering algorithm properties are shown to be an improvement over the FCM method's properties. Numerical simulations show that the clustering algorithm gives more accurate clustering results than the FCM method.

Key words: fuzzy *c*-means (FCM); alternating optimization; fuzzy clustering

Introduction

Fuzzy clustering plays an important role in pattern recognition, image processing, and data analysis. In fuzzy clustering, every point is assigned a membership to represent the degree of belonging to a certain class. The fuzzy *c*-means (FCM) method is one of the best known fuzzy clustering methods[1,2]. In FCM, the objective function is the trace of a within-cluster scatter matrix (specified in Section 1) with spherical clusters obtained by minimizing the objective function with alternative optimization.

Other clustering algorithms such as the Gustafson-Kessel (GK) clustering algorithm[3] and Gath-Geva (GG) clustering algorithm[4] were developed to detect non-spherical structural clusters, but these two algorithms fail to consider the relationships between cluster centers in the objective function. The compatibilities of points with the cluster centers are guaranteed in a possibilistic *c*-means (PCM) method presented by Krishnapuram and Keller[5]. However, their algorithm exhibited bad behavior because there were no links among clusters[6]. The concept of regularization was used by Ozdemir and Akarun[7] in an inter-cluster separation (ICS) algorithm and by Yang et al.[8] in a fuzzy compactness and separation (FCS) algorithm. Unfortunately, the ICS algorithm has different objective functions for different cluster centers with the regulating term only considered as a perturbation. The FCS algorithm has hard kernel boundaries which depend on experiments and all the data points in one kernel cannot be discriminated because they have identical membership values. This paper describes an extended objective function consisting of a fuzzy within-cluster scatter matrix and a new between-cluster centers scattering matrix. The corresponding fuzzy clustering algorithm assures the compactness between data points and cluster centers and also strengthens the separation between cluster centers based on the separation criterion.

Received: 2004-09-15; revised: 2005-01-18

* Supported by the National Excellent Doctoral Dissertation Foundation (No. 200041) and the National Key Basic Research and Development (973) Program of China (No. G2002cb312205)

** To whom correspondence should be addressed.
E-mail: chunsheng@mail.tsinghua.edu.cn; Tel: 86-10-62796858

1 Extended Objective Function

The clustering optimization is based on objective functions so the choice of an appropriate objective function is the key to the success of the cluster analysis. Let $\{z_1, z_2, \ldots, z_n\}$ be a set of n data points represented by p-dimensional feature vectors $z_j = [z_{1j}, \ldots, z_{pj}]^{\mathrm{T}} \in \mathbf{R}^p$. The $p \times n$ data matrix $\boldsymbol{Z}$ has the cluster center matrix $\boldsymbol{V} = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_c]$ $(1 < c < n)$ and the membership matrix $\boldsymbol{U} = [\mu_{ij}]_{c\times n}$, where μ_{ij} is the membership value of $\boldsymbol{z}_j$ belonging to $\boldsymbol{v}_i$. $\hat{\boldsymbol{U}} = [\hat{\mu}_{ik}]_{c\times c}$ represents the weighting matrix, and $\hat{\mu}_{ik}$ is the weighting value between $\boldsymbol{v}_i$ and $\boldsymbol{v}_k$. The fuzzy exponent m is greater than 1[9]. Then, the proposed objective function is

$$\bar{J} = \sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ij}^{m}\left\|\boldsymbol{z}_j - \boldsymbol{v}_i\right\|^2 - \frac{1}{c(c-1)}\sum_{i=1}^{c}\sum_{k=1}^{c}\hat{\mu}_{ik}\left\|\boldsymbol{v}_i - \boldsymbol{v}_k\right\|^2 \quad (1)$$

$$\text{s.t.} \quad \mu_{ij} \in [0,1], \ \sum_{i=1}^{c}\mu_{ij} = 1, \ \forall j,$$

$$0 < \sum_{j=1}^{n}\mu_{ij} < n, \ \forall i \quad (1a)$$

where $\hat{\mu}_{ik}$ is defined as

$$\hat{\mu}_{ik} = \beta\frac{\min\left\{\left\|\bar{\boldsymbol{z}} - \boldsymbol{v}_i\right\|, \left\|\bar{\boldsymbol{z}} - \boldsymbol{v}_k\right\|\right\}}{\max\left\{\left\|\bar{\boldsymbol{z}} - \boldsymbol{v}_i\right\|, \left\|\bar{\boldsymbol{z}} - \boldsymbol{v}_k\right\|\right\}}, \ 0 \leqslant \beta \leqslant 1 \quad (1b)$$

where $\bar{\boldsymbol{z}} = \left(\sum_{i=1}^{n}\boldsymbol{z}_j\right)\Big/ n$ is the data set mass center and β is an adjustable factor. If $\left\|\bar{\boldsymbol{z}} - \boldsymbol{v}_i\right\| = 0$, then $\hat{\mu}_{ik} = \beta$. The properties of the objective function can be analyzed by studying the fuzzy scattering matrices[10]:

Fuzzy within-cluster scatter matrix

$$\boldsymbol{S}_{\mathrm{W}} = \sum_{i=1}^{c}\sum_{j=1}^{n}\left(\mu_{ik}\right)^m\left(\boldsymbol{z}_j - \boldsymbol{v}_i\right)\left(\boldsymbol{z}_j - \boldsymbol{v}_i\right)^{\mathrm{T}} \quad (2)$$

Fuzzy between-cluster scatter matrix

$$\boldsymbol{S}_{\mathrm{B}} = \sum_{i=1}^{c}\left(\sum_{j=1}^{n}\left(\mu_{ij}\right)^m\right)\left(\boldsymbol{v}_i - \bar{\boldsymbol{v}}\right)\left(\boldsymbol{v}_i - \bar{\boldsymbol{v}}\right)^{\mathrm{T}} \quad (3)$$

where $\bar{\boldsymbol{v}} = \left(\sum_{i=1}^{c}\sum_{j=1}^{n}\left(\mu_{ij}\right)^m\boldsymbol{z}_j\right)\Big/\left(\sum_{i=1}^{c}\sum_{j=1}^{n}\left(\mu_{ij}\right)^m\right)$ is the fuzzy extension of the total mean vector.

The new between-cluster scatter matrix

$$\boldsymbol{S}'_{\mathrm{B}} = \sum_{i=1}^{c}\sum_{k=1}^{c}\hat{\mu}_{ik}\left(\boldsymbol{v}_i - \boldsymbol{v}_k\right)\left(\boldsymbol{v}_i - \boldsymbol{v}_k\right)^{\mathrm{T}} \quad (4)$$

can be regarded as a separation criterion representing the weighted distance sum between cluster centers, while the matrix in Eq. (3) denotes the relationships between cluster centers and the fuzzy mean vector. In the standard FCM method, the objective function is

$$J_{\mathrm{FCM}} = \sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ij}^{m}\left\|\boldsymbol{z}_j - \boldsymbol{v}_i\right\|^2 = \mathrm{Tr}\left(\boldsymbol{S}_{\mathrm{W}}\right) \quad (5)$$

Here, the proposed objective function in Eq. (1) can be rewritten as

$$\bar{J} = \mathrm{Tr}\left(\boldsymbol{S}_{\mathrm{W}} - \frac{1}{c(c-1)}\boldsymbol{S}'_{B}\right) \quad (6)$$

The second term in Eq. (6) is a punishing function used to describe the influence of cluster centers on the fuzzy clustering algorithm. Therefore, minimization of the objective function in Eq. (6) is equivalent to minimizing the trace of the fuzzy within-cluster scatter matrix while maximizing the weighted average distance between cluster centers. Consequently, both the compactness and the separation are assured by the fuzzy clustering algorithm.

2 Fuzzy Clustering Algorithm

The goal of the clustering algorithm is to identify the cluster centers and the membership values by solving an optimization problem. Alternating optimization is a popular mathematical tool for the regular objective function-based fuzzy clustering algorithms. The optimal update equations can be obtained using the Lagrange method by setting the partial derivative of the Lagrange with respect to $\boldsymbol{v}_i$ and with respect to μ_{ij} equal to zero. Setting $\partial\bar{J}/\partial\mu_{ij}$ equal to zero gives the update equation for μ_{ij},

$$\mu_{ij} = 1\Bigg/\sum_{s=1}^{c}\left[\left\|\boldsymbol{z}_j - \boldsymbol{v}_i\right\|^2\Big/\left\|\boldsymbol{z}_j - \boldsymbol{v}_s\right\|^2\right]^{(m-1)^{-1}} \quad (7)$$

From $\partial\bar{J}/\partial\boldsymbol{v}_i$ the update equation for $\boldsymbol{v}_i$ is

$$\boldsymbol{v}_i = \frac{\sum_{j=1}^{n}\mu_{ij}^{m}\boldsymbol{z}_j - \frac{1}{c(c-1)}\sum_{k=1}^{c}\hat{\mu}_{ik}\boldsymbol{v}_k}{\sum_{j=1}^{n}\mu_{ij}^{m} - \frac{1}{c(c-1)}\sum_{k=1}^{c}\hat{\mu}_{ik}} \quad (8)$$

Singularities occur in these two update equations when $\left\|\boldsymbol{z}_j - \boldsymbol{v}_i\right\| = 0$ or $\sum_{j=1}^{n}\mu_{ij}^{m} - \frac{1}{c(c-1)}\sum_{k=1}^{c}\hat{\mu}_{ik} = 0$ in any iteration. When this happens, zeros are assigned to the corresponding cluster centers and the memberships are chosen arbitrarily subject to the constraints in Eq. (1a). The fuzzy clustering algorithm can be summarized in the following steps:

Initialization

Given a data set $\boldsymbol{Z}$, set the iteration counter $l=0$ and $m>1$. Then choose the initial number of cluster centers c, the termination criteria $\varepsilon>0$, and the parameter $0<\beta<1$. Randomly initialize $\boldsymbol{U}^{(0)}$ and $\hat{\boldsymbol{U}}^{(0)}$.

Step 1. Calculate the cluster centers $\boldsymbol{v}_i$ using Eq. (8).

Step 2. Calculate the weighting parameters $\hat{\mu}_{ik}$ using Eq. (1b).

Step 3. Calculate the membership values μ_{ij} using Eq. (7).

Increment l **until** $\left\|\boldsymbol{U}^{(l+1)}-\boldsymbol{U}^{(l)}\right\|<\varepsilon$.

Three important properties of the proposed fuzzy clustering algorithm can be compared with the properties of the FCM algorithm.

1) The FCM objective function, Eq. (5), only considers a within-cluster scatter matrix in Eq. (2), while the proposed algorithm, Eq. (6), includes a within-cluster scatter matrix in Eq. (2) and a new between-cluster scatter matrix in Eq. (4). Minimization of Eq. (6) effectively increases both the cluster compactness and the separation.

2) Both algorithms use the alternating optimization algorithm.

3) Both algorithms use the same update equation for the membership μ_{ij}. Comparison of the FCM cluster center update equation $\boldsymbol{v}_i=\sum_{j=1}^{n}\mu_{ij}^{m}\boldsymbol{z}_j \Big/ \sum_{j=1}^{n}\mu_{ij}^{m}$ and the cluster center update Eq. (8) for the proposed algorithm shows that both the numerator and the denominator of Eq. (8) have ad hoc punishing functions, which improve the accuracy of the cluster centers. However, the computational complexity of the proposed algorithm is also increased. Compared with the FCM cluster center update equation, Eq. (8) has additional $2(c+1)$ add operations, $(3+c+3+2c^2)$ product operations, 2 norm operations, and 2 logical operations in a single iteration.

3 Numerical Example

A numerical example using the data set illustrated in Fig. 1 is presented to compare the clustering algorithm and the FCM algorithm. The numerical example includes three groups of data generated randomly from normal distributions around three centers with the standard deviations given in Table 1. The two methods are compared for $m=2$, $\varepsilon=0.000\ 01$, and fixed initial values for the memberships and cluster centers.

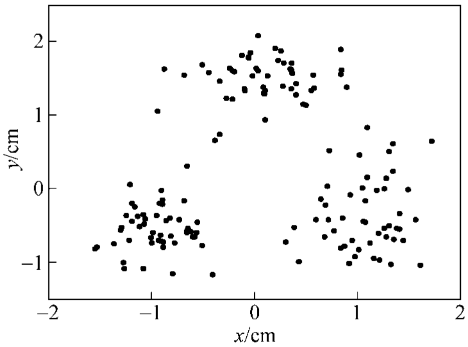

Fig. 1 Two-dimensional data set

Table 1 Data group centers, variances, and sample sized

Group	Center	Variance	Number of samples
1	(−1, −0.5)	(0.1,0.1)	50
2	(1, −0.4)	(0.1,0.2)	50
3	(0,1.5)	(0.2,0.1)	50

The FCM algorithm and the clustering algorithm are compared using the following criteria for the cluster center locations: the mean square error (MSE) of the centers ($\mathrm{MSE}=\sqrt{\left\|\hat{\boldsymbol{v}}-\boldsymbol{v}_{\mathrm{t}}\right\|^2}$, where $\hat{\boldsymbol{v}}$ is the computed center and $\boldsymbol{v}_{\mathrm{t}}$ is the true center), the objective function values (OFV), and the number of iterations (NI). The computed results are listed in Table 2 for various values of β.

The cluster centers found by the clustering algorithm are closer to the true centers (i.e., smaller MSE values) than the centers found by the FCM algorithm when β is not very large, i.e., less than 0.4. Equations (1b) and (8) indicate that the calculated cluster centers may be far from the true centers and have poorer MSE values than the FCM results when β is too large. As shown in the example, when $\beta=0.9$, the MSE value for the second center of the proposed algorithm is larger than the FCM result. For all five values of β, the OFVs

Table 2 Cluster centers, MSE, OFV, and NI obtained from FCM and the proposed clustering algorithm

	Group	FCM	The proposed clustering algorithm				
			$\beta=0.1$	$\beta=0.2$	$\beta=0.3$	$\beta=0.4$	$\beta=0.9$
Centers	1	(−0.9501, −0.5395)	(−0.9522,0.5481)	(−0.9532, −0.5489)	(−0.9542, −0.5496)	(−0.9552,0.5503)	(−0.9401, −0.5540)
	2	(1.0968,0.4345)	(1.0977, −0.4113)	(1.0987, −0.4119)	(1.0998, −0.4125)	(1.1008, −0.4131)	(1.1062, −0.4161)
	3	(0.1143,1.5574)	(0.0961,1.5635)	(0.0962,1.5648)	(0.0962,1.5661)	(0.0962,1.5674)	(0.0964,1.5739)
MSE	1	0.0718	0.0678	0.0677	0.0675	0.0674	0.0671
	2	0.1027	0.0983	0.0994	0.1006	0.1017	0.1074
	3	0.1279	0.1152	0.1159	0.1167	0.1175	0.1215
OFV		29.8230	18.3470	17.8990	17.4497	16.9988	14.7229
NI		11	11	11	11	11	11

from the clustering algorithm were less than the FCM result, which means that different β may produce different clustering results. Therefore, the adjustable factor β is of great importance to the clustering results. Unfortunately, an optimum value of β for all data sets is not possible due to the variations in the data sets. However, various simulations suggest that the clustering algorithm gives better results with β in the range $[0.1,\ 0.4]$. In addition, although the objective function is extended by adding a separation index, the change does not increase the number of iterations.

4 Conclusions

An improved fuzzy clustering algorithm is developed to obtain better quality of fuzzy clustering results. The objective function includes a fuzzy within-cluster scatter matrix and a new between-prototypes scatter matrix. The update equations for the memberships and the cluster centers are derived from the Lagrange method. Comparison of the clustering algorithm and the FCM algorithm shows that the clustering algorithm will increase the cluster compactness and the separation between clusters. Finally, a numerical example shows that the clustering algorithm gives more accurate clustering results than the FCM algorithm for a typical problem.

References

[1] Dunn J C. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *J. Cybernetics,* 1973, **3**(3): 32-57.

[2] Bezdek J C. Pattern Recognition with Fuzzy Objective Function Algorithms. New York: Plenum Press, 1981: 65-70.

[3] Gustafson D E, Kessel W C. Fuzzy clustering with a fuzzy covariance matrix. In: Proceedings of IEEE Conference Decision Control. San Diego, CA, 1979: 761-766.

[4] Gath I, Geva A B. Unsupervised optimal fuzzy clustering. *IEEE Trans. Pattern Anal. Machine Intell.*, 1989, **7**: 773-781.

[5] Krishnapuram R, Keller J M. A possibilistic approach to clustering. *IEEE Trans. Fuzzy Systerms*, 1993, **1**(2): 98-110.

[6] Barni M, Cappellini V, Mecocci A. Comments on “A possibilistic approach to clustering”. *IEEE Trans. Fuzzy Systerms*, 1996, **4**(3): 393-396.

[7] Ozdemir D, Akarun L. Fuzzy algorithms for combined quantization and dithering. *IEEE Trans. Image Processing,* 2001, **10**(6): 923-931.

[8] Yang M S, Wu K L, Yu J. A novel fuzzy clustering algorithm. *Computational Intelligence in Robotics and Automation IEEE International Symposium*, 2003, **2**: 647-652.

[9] Yu J, Cheng Q, Huang H. Analysis of the weighting exponent in the FCM. *IEEE Trans. Systems, Man and Cybernetics, Part B,* 2004, **34**(1): 634-639.

[10] Emami M R, Turksen I B, Goldenberg A A. Development of a systematic methodology of fuzzy logic modeling. *IEEE Trans. Fuzzy Systems,* 1998, **6**(3): 346-361.