# Nonstandard Methods of Fuzzy c-Means Clustering

Sadaaki Miyamoto Department of Risk Engineering School of Systems and Information Engineering University of Tsukuba Ibaraki 305-8573, Japan E-mail: miyamoto@risk.tsukuba.ac.jp

*Abstract*— We overview nonstandard methods of fuzzy *c*-means clustering. The basic framework is the alternate optimization algorithm originally proposed by Dunn and Bezdek, but two types of objective functions considered herein are different from the standard one. The motivation for the use of these functions is twofold; first is regularizing singularities in standard methods and second is to uncover a relation of possibilistic and probabilistic methods. An additional variable of controlling volume size is included as an extension. A method of the kernel trick for obtaining nonlinear cluster boundaries is moreover considered and a simple numerical example is shown.

#### I. INTRODUCTION

Although many variations of fuzzy c-means clustering [1], [2] have been investigated, there are still rooms for further study in the fundamental and methodological aspect in this issue. In this paper we provide a perspective in fuzzy *c*-means methodology by considering nonstandard approaches.

Although most studies in fuzzy c-means use the objective function proposed by Dunn [1] and Bezdek [2] and their variations, there are other types of objective functions for the same purpose of fuzzy c-means clustering.

Namely, the possibility approach by Krishnapuram and Keller [9] is also well-known. Second, the entropy-based approach [12], [15] has been proposed which has many implications and extensions. In fact, it has been shown by Ichihashi *et al.* [7] that an extension of this method encompasses the Gaussian mixture model of the statistical approach [16].

In this paper we overview and propose two objective functions; first is the entropy-based function and second is derived from the one of possibilistic clustering. The two functions can be used for both the ordinary and possibilistic clustering.

A variation of this approach is to incorporate an additional variable for controlling cluster volume sizes. We discuss this variation for the two objective functions.

The second topic in this paper is a method to obtain nonlinear cluster boundaries. As the basic methods of crisp and fuzzy *c*-means produce piecewise linear boundaries between clusters, an ordinary type of their extensions cannot derive strongly nonlinear boundaries. For this purpose the kernel trick used in support vector machines [18] is employed whereby the alternate optimization algorithm of fuzzy *c*-means is transformed into updating dissimilarity values between an object and a cluster center. A typical numerical example of a nonlinear boundary is shown. Throughout this paper we stick to the original idea of the alternate optimization and avoids *ad hoc* generalizations to fixed point iterations.

# II. OBJECTIVE FUNCTIONS OF FUZZY C-MEANS

#### A. Preliminaries

Let the set of objects for clustering be  $X = \{x_1, \ldots, x_n\}$ ; they are points in the *p*-dimensional Euclidean space:  $x_k = (x_k^1, \ldots, x_k^p)^t$ . On the other hand, a cluster *i* is represented by the center  $v_i = (v_i^1, \ldots, v_i^p)$ .

The membership matrix is  $U = (u_{ik})$ , where  $u_{ik}$  is the degree of membership of  $x_k$  to cluster *i*; the sequence of the cluster centers is  $V = (v_1, \ldots, v_c)$ .

The basic alternate optimization algorithm of fuzzy c-means is the iteration of **FC** in the following [2].

#### FC: Basic Fuzzy *c*-Means Algorithm.

#### **FC0.** Set the initial value of $\overline{V}$ .

**FC1.** Solve  $\min_{U \in M} J(U, \overline{V})$  and let  $\overline{U}$  be the optimal solution.

**FC2.** Solve  $\min_{V} J(\overline{U}, V)$  and let  $\overline{V}$  be the optimal solution.

FC3. If the solution  $(\overline{U}, \overline{V})$  is convergent, stop; else go to FC1.

#### End of FC.

The ordinary constraint for U is

$$M = \left\{ (u_{ik}) : u_{ik} \in [0,1], \sum_{i=1}^{c} u_{ik} = 1, \forall k \right\}.$$

As the objective function J(U, V) the following has mainly been considered.

$$J_0(U,V) = \sum_{i=1}^c \sum_{k=1}^n (u_{ik})^m ||x_k - v_i||^2 = \sum_{i=1}^c \sum_{k=1}^n (u_{ik})^m d_{ik},$$

where we put

$$d_{ik} = ||x_k - v_i||^2$$

The optimal solutions of  $u_{ik}$  and  $v_i$  are respectively given by

$$u_{ik} = \left[\sum_{j=1}^{c} \left(\frac{\|x_k - v_i\|^2}{\|x_k - v_j\|^2}\right)^{\frac{1}{m-1}}\right]^{-1}$$
  
$$= \left[\sum_{j=1}^{c} \left(\frac{d_{ik}}{d_{jk}}\right)^{\frac{1}{m-1}}\right]^{-1},$$
  
$$v_i = \frac{\sum_{k=1}^{n} (u_{ik})^m x_k}{\sum_{k=1}^{n} (u_{ik})^m}.$$
  
put  
$$g_{ik}^0 = \frac{1}{d_{ik}^{\frac{1}{m-1}}}$$

When we put

then

$$u_{ik} = \frac{g_{ik}^0}{\sum_{j=1}^c g_{jk}^0} \tag{1}$$

Krishnapuram and Keller [9] proposed the method of possibilistic clustering: the same alternate optimization algorithm **FC** is used in which the constraint M is not employed but nontrivial solution of

$$\sum_{k=1}^{n} u_{ik} > 0, \quad 1 \le i \le \epsilon$$

should be obtained. For this purpose the objective function  $J_0$  cannot be employed since the optimal U is trivial:  $u_{ik} = 0$  for all i and k. Hence a modified objective function

$$J_{pos}(U,V) = \sum_{i=1}^{c} \sum_{k=1}^{n} (u_{ik})^m d_{ik} + \sum_{i=1}^{c} \eta_i \sum_{k=1}^{n} \sum_{i=1}^{c} (1 - u_{ik})^n d_{ik}$$

has been proposed whereby the solution becomes

$$u_{ik} = \frac{1}{1 + \left(\frac{d_{ik}}{\eta_i}\right)^{\frac{1}{m-1}}}$$

while the optimal  $v_i$  remains the same.

The objective function  $J_0$  cannot be used in possibilistic clustering and  $J_{pos}$  is unavailable for the ordinary fuzzy *c*means. Thus the two methods have nothing in common but the alternate optimization algorithm. In contrast, we consider other objective functions usable for the both methods.

### B. Nonstandard objective functions

We consider the following two objective functions. The first is entropy-based and the second is simplified from  $J_{pos}$  by putting m = 2.

$$J_1(U,V) = \sum_{i=1}^c \sum_{k=1}^n \left\{ u_{ik} d_{ik} + \lambda^{-1} u_{ik} \log u_{ik} \right\}$$
(2)

$$J_2(U,V) = \sum_{i=1}^{c} \sum_{k=1}^{n} \left\{ (u_{ik})^2 d_{ik} + \zeta^{-1} (1-u_{ik})^2 \right\}.$$
(3)

Put

$$g_{ik}^1 = e^{-1} \exp(\lambda d_{ik}) \tag{4}$$

$$g_{ik}^2 = \frac{1}{1+\zeta d_{ik}} \tag{5}$$

Then the solution for the possibilistic clustering is

$$u_{ik} = g_i^{\mathsf{I}}$$

when  $J_1$  is used; it is

 $u_{ik} = g_{ik}^2$ 

when  $J_2$  is used.

Consider next the ordinary fuzzy c-means where the constraint M is used. Now the solutions are

$$u_{ik} = \frac{g_{ik}^{1}}{\sum_{j=1}^{c} g_{jk}^{1}} \tag{6}$$

and

$$u_{ik} = \frac{g_{ik}^2}{\sum_{j=1}^c g_{jk}^2} \tag{7}$$

respectively for  $J_1$  and  $J_2$  (cf. also [3]).

We thus observe that  $J_1$  and  $J_2$  are usable for both the ordinary fuzzy *c*-means and the possibilistic clustering. Moreover we notice the above relationship of solutions of  $u_{ik} = g_{ik}^{\ell}$  and  $u_{ik} = g_{ik}^{\ell} / \sum_j g_{jk}^{\ell}$  ( $\ell = 1, 2$ ) between the both methods.

Notice moreover that there are features of 'regularization' in the two objective functions.  $J_0$  adds the entropy term to the objective function of the crisp *c*-means and regularizes, or fuzzifies, the solution *U*. On the other hand,  $J_2$  regularizes the ordinary objective function  $J_0$  by eliminating the singularity in  $g_{ik}^0$ .

#### C. Variable for controlling cluster volume size

Let us consider an extension of the entropy-based method. Although this extension has been discussed by Ichihashi *et al.* [7] in a more general form including the covariance matrix, we do not discuss the covariance matrix here for simplicity, but the use of covariance variable is not difficult [5], [6], [7].

That is, a generalized objective function where an additional variable  $\alpha = (\alpha_1, \dots, \alpha_c)$  for controlling cluster volume sizes is used:

$$J_{1}'(U, V, \alpha) = \sum_{k=1}^{n} \sum_{i=1}^{c} u_{ik} ||x_{k} - v_{i}||^{2} + \lambda^{-1} \sum_{k=1}^{n} \sum_{i=1}^{c} u_{ik} \log \alpha_{i}^{-1} u_{ik}$$
(8)

The constraint for  $\alpha$  is

$$A = \{ \alpha : \sum_{i=1}^{c} \alpha_i = 1, \alpha_i \ge 0, i = 1, \dots, c \}.$$

Then the alternate optimization is as follows.

FC': An Extended Algorithm of Fuzzy *c*-Means. FC'0. Set initial value of  $\overline{V}$ ,  $\overline{\alpha}$ .

- **FC'1.** Solve  $\min_{U \in M} J(U, \overline{V}, \overline{\alpha})$  and let the optimal solution be  $\overline{U}$ .
- FC'2. Solve  $\min_{V} J(\bar{U}, V, \bar{\alpha})$  and let the optimal solution be  $\bar{V}$
- **FC'3.** Solve  $\min_{\alpha \in A} J(\bar{U}, \bar{V}, \alpha)$  and let the optimal solution be  $\bar{\alpha}$ .

FC'4. If the solution  $(\overline{U}, \overline{V}, \overline{\alpha})$  is convergent, stop; else go to FC'1.

# End of FC'.

The optimal solutions are

$$u_{ik} = \frac{\alpha_{i}g_{ik}^{1}}{\sum_{j=1}^{c} \alpha_{j}g_{jk}^{1}}$$
(9)

$$v_i = \frac{\sum_{k=1}^{n} u_{ik} x_k}{\sum_{k=1}^{n} u_{ik}}$$
(10)

$$\alpha_i = \frac{\sum_{k=1}^{k} u_{ik}}{n} \tag{11}$$

The corresponding extension of the second objective function is

$$J_{2}'(U, V, \alpha) = \sum_{i=1}^{c} \sum_{k=1}^{n} \alpha_{i}^{-1} (u_{ik})^{2} d_{ik} + \zeta^{-1} \sum_{i=1}^{c} \sum_{k=1}^{n} \alpha_{i} (1 - \alpha_{i}^{-1} u_{ik})^{2}$$

The optimal solutions of the alternate minimization are

$$u_{ik} = \frac{\alpha_{i}g_{ik}^{2}}{\sum_{j=1}^{c} \alpha_{j}g_{jk}^{2}}$$

$$v_{i} = \frac{\sum_{k=1}^{n} (u_{ik})^{2} x_{k}}{\sum_{k=1}^{n} (u_{ik})^{2}}$$

$$\alpha_{i} = \frac{\left(\sum_{k=1}^{n} \frac{(u_{ik})^{2}}{\zeta g_{ik}^{2}}\right)^{\frac{1}{2}}}{\sum_{j=1}^{c} \left(\sum_{k=1}^{n} \frac{(u_{jk})^{2}}{\zeta g_{jk}^{2}}\right)^{\frac{1}{2}}}$$

# III. NONLINEAR CLUSTER BOUNDARIES

Recently support vector machines have been studied by many researchers [18]. Nonlinear classification technique therein uses the kernel trick, that is, a mapping into a highdimensional feature space of which the functional form is unknown but the inner product has an explicit representation of a kernel function.

We overview the application of the kernel trick to fuzzy *c*-means [14].

Let a mapping defined on the data space into a highdimensional feature space be  $\Phi(x): \mathbf{R}^p \to H$ ; *H* is in general a Hilbert space of which the inner product and the norm are respectively denoted by  $\langle \cdot, \cdot \rangle$  and  $\|\cdot\|_H$ . The explicit form of  $\Phi(x)$  is unknown but the product is given by a kernel function:

$$K(x, y) = \langle \Phi(x), \Phi(y) \rangle.$$

Most important kernel function is the Gaussian kernel

$$K(x,y) = \exp[-const ||x - y||^2]$$

which is used in the numerical example below.

The following objective function is used instead of  $J_1$ .

$$J_1^K(U,V) = \sum_{i=1}^c \sum_{k=1}^n \{u_{ik} \| \Phi(x_k) - v_i \|_H^2 + \lambda^{-1} u_{ik} \log u_{ik} \}$$

For  $J_0$  and  $J_2$ , the corresponding objective functions are analogously defined but we omit the detail.

We proceed to consider the solution in the alternate optimization. The solution for U is given by the same formula of (9) but the center is

$$v_{i} = \frac{\sum_{k=1}^{n} u_{ik} \Phi(x_{k})}{\sum_{k=1}^{n} u_{ik}}.$$
 (12)

which cannot be calculated, since the explicit form of  $\Phi(x_k)$  is unavailable.

We hence substitute (12) into

$$d_{ik} = \|\Phi(x_k) - v_i\|_H^2.$$
(13)

After some manipulation, we have

$$d_{ik} = K_{kk} - \frac{2}{U_i} \sum_{j=1}^n u_{ij} K_{jk} + \frac{1}{U_i^2} \sum_{j=1}^n \sum_{\ell=1}^n u_{ij} u_{i\ell} K_{j\ell}$$
(14)

where

$$U_i = \sum_{k=1}^n u_{ik},$$
  
$$K_{jk} = \langle \Phi(x_j), \Phi(x_k) \rangle$$

Therefore the alternate optimization of **FC** is reduced to the iteration of calculating U by (12) and  $d_{ik}$  by (14) until convergence of U.

The iterative formulas for  $J_0$ ,  $J_2$ ,  $J'_1$ , and  $J'_2$  are derived likewise. We omit the detail.

**Remark:** A crisp *c*-means algorithm using the kernel trick has been proposed by Girolami [4]; a variation of crisp *c*-means algorithm has been studied by Miyamoto *et al.* [13].

#### IV. AN ILLUSTRATIVE EXAMPLE

We discuss a simple and typical illustrative example of a nonlinear cluster boundary. Figure 1 shows a data set classified by the ordinary fuzzy c-means using  $J_0$  with m = 2. The two clusters are shown by  $\Box$  and  $\times$ . Two small circles  $\bigcirc$  show cluster centers. Apparently, the two circular groups recognized by sight cannot be separated by the ordinary fuzzy c-means, since one group is inside the other and hence the cluster boundary should be circular, whereas the crisp and fuzzy c-means produce the Voronoi regions [8] with piecewise linear boundaries in general. It is also clear that such circular boundary cannot be obtained by using extensions such as  $J'_1$  and  $J'_2$ .

Figure 2 has been obtained from the method of the kernel trick using the Gaussian kernel with const = 0.1. The crisp and fuzzy *c*-means methods [14], [13] produced the same clusters.



Fig. 1. Data set of 'a ball and circle' classified by the ordinary fuzzy c-means



Fig. 2. Data set of 'a ball and circle' classified by the kernel-based fuzzy c-means (const = 0.1)

## V. CONCLUSION

We have overviewed two objective functions which is employed for both the fuzzy *c*-means and possibilistic clustering. Moreover an additional variable for controlling cluster volume sizes has been introduced. Such additional variables enable to generate cluster boundaries with quadratic curves, but stronger nonlinearities cannot be handled by the ordinary methods.

Hence the use of the kernel trick has been considered and a new iterative formula has been derived. A typical example of a nonlinear cluster boundary has been shown.

Entropy-based clustering has been considered by several authors both for the fuzzy *c*-means [10], [12], [15], [7] and possibilistic clustering [3]. Moreover this method has close relationships with statistical models [7], [11], [16], [17].

Although recent studies on fuzzy clustering is more focused on applications than theories, there are many rooms for further theoretical development and advanced algorithms. Relationships with neural network techniques should furthermore be investigated.

#### Acknowledgment

This study has partly been supported by the Grant-in-Aid for Scientific Research (B), Japan Society for the Promotion of Science, No.16300065.

#### REFERENCES

- J.C. Dunn, A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters, *J. of Cybernetics*, Vol.3, pp. 32–57, 1974.
- [2] J.C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum, New York, 1981.
- [3] R.N.Davé, R.Krishnapuram, Robust clustering methods: a unified view, IEEE Trans. Fuzzy Syst., Vol.5, No.2, pp. 270–293, 1997.
- [4] M. Girolami, Mercer kernel based clustering in feature space, *IEEE Trans. on Neural Networks*, Vol.13, No.3, pp. 780–784, 2002.
- [5] E.E. Gustafson, W.C. Kessel, Fuzzy clustering with a fuzzy covariance matrix, IEEE CDC, San Diego, California, pp. 761–766, 1979.
- [6] F. Höppner, F. Klawonn, R. Kruse, T. Runkler, *Fuzzy Cluster Analysis*, Wiley, Chichester, 1999.
- [7] H. Ichihashi, K. Honda, N. Tani, Gaussian mixture PDF approximation and fuzzy c-means clustering with entropy regularization, *Proc. of the* 4th Asian Fuzzy System Symposium, May 31-June 3, 2000, Tsukuba, Japan, pp.217–221.
- [8] T. Kohonen, Self-Organization and Associative Memory, Springer-Verlag, Heiderberg, 1989.
- [9] R. Krishnapuram, J.M. Keller, A possibilistic approach to clustering, IEEE Trans. on Fuzzy Syst., Vol.1, No.2, pp. 98–110, 1993.
- [10] R.-P. Li, M. Mukaidono, A maximum entropy approach to fuzzy clustering; Proc. of the 4th IEEE Intern. Conf. on Fuzzy Systems (FUZZ-IEEE/IFES'95), Yokohama, Japan, March 20-24, 1995, pp. 2227–2232, 1995.
- [11] G. McLachlan, D. Peel, Finite Mixture Models, Wiley, New York, 2000.
- [12] S. Miyamoto and M. Mukaidono, Fuzzy c means as a regularization and maximum entropy approach, Proc. of the 7th International Fuzzy Systems Association World Congress (IFSA'97), June 25-30, 1997, Prague, Czech, Vol.II, pp. 86–92, 1997.
- [13] S. Miyamoto, Y. Nakayama, Algorithms of hard *c*-means clustering using kernel functions in support vector machines, *Journal of Advanced Computational Intelligence and Intelligent Informatics*, Vol. 7, No. 1, pp. 19–24, 2003.
- [14] S. Miyamoto, D. Suizu, Fuzzy c-means clustering using kernel functions in support vector machines, *Journal of Advanced Computational Intelligence and Intelligent Informatics*, Vol. 7, No. 1, pp. 25–30, 2003.
- [15] S. Miyamoto, K. Umayahara, Methods in hard and fuzzy clustering, In: Z.Q. Liu and S. Miyamoto, eds., *Soft Computing and Human Centered Machines*, Springer, Tokyo, 2000, pp. 85–129.
- [16] R.A. Redner and H.F. Walker, Mixture densities, maximum likelihood and the EM algorithm, *SIAM Review*, Vol.26, No.2, pp. 195–239, 1984.
- [17] K. Rose, E. Gurewitz, G. Fox, A deterministic annealing approach to clustering, *Pattern Recognition Letters*, Vol.11, pp.589–594, 1990.
- [18] V.N. Vapnik, Statistical Learning Theory, Wiley, New York, 1998.