Mathl. Comput. Modelling Vol. 18, No. 11, pp. 1–16, 1993
Printed in Great Britain.
0895-7177/93 \$6.00 + 0.00

A Survey of Fuzzy Clustering

M.-S. YANG
Department of Mathematics
Chung Yuan Christian University
Chungli, Taiwan 32023, R.O.C.

(Received and accepted October 1993)

Abstract—This paper is a survey of fuzzy set theory applied in cluster analysis. These fuzzy clustering algorithms have been widely studied and applied in a variety of substantive areas. They also become the major techniques in cluster analysis. In this paper, we give a survey of fuzzy clustering in three categories. The first category is the fuzzy clustering based on fuzzy relation. The second one is the fuzzy clustering based on objective function. Finally, we give an overview of a nonparametric classifier. That is the fuzzy generalized k-nearest neighbor rule.

Keywords—Cluster analysis, Fuzzy clustering, Fuzzy c-partitions, Fuzzy relation, Fuzzy c-means, Fuzzy generalized k-nearest neighbor rule, Cluster validity.

1. INTRODUCTION

Cluster analysis is one of the major techniques in pattern recognition. It is an approach to unsupervised learning. The importance of clustering in various areas such as taxonomy, medicine, geology, business, engineering systems and image processing, etc., is well documented in the literature. See, for example, [1–3]. The conventional (hard) clustering methods restrict that each point of the data set belongs to exactly one cluster. Fuzzy set theory proposed by Zadeh [4] in 1965 gave an idea of uncertainty of belonging which was described by a membership function. The use of fuzzy sets provides imprecise class membership information. Applications of fuzzy set theory in cluster analysis were early proposed in the work of Bellman, Kalaba and Zadeh [5] and Ruspini [6]. These papers open the door of research in fuzzy clustering. Now fuzzy clustering has been widely studied and applied in a variety of substantive areas. See, for example, [7,8]. These methods become the important tools to cluster analysis.

The purpose of this paper is briefly to survey the fuzzy set theory applied in cluster analysis. We will focus on three categories: fuzzy clustering based on fuzzy relation, fuzzy clustering based on objective functions, and the fuzzy generalized k-nearest neighbor rule—one of the powerful nonparametric classifiers. We mention that this paper is a nonexhaustive survey of fuzzy set theory applied in cluster analysis. We may have overlooked the papers which might be important in fuzzy clustering.

In Section 2, we give a brief survey of fuzzy c-partitions, (fuzzy) similarity relation, and fuzzy clustering based on fuzzy relation. Section 3 covers a big part of fuzzy clustering based on the objective functions. In the literature of fuzzy clustering, the fuzzy c-means (FCM) clustering algorithms defined by Dunn [9] and generated by Bezdek [7] are the well-known and powerful methods in cluster analysis. Several variations and generalizations of the FCM shall be discussed in this section. Finally, Section 4 contains an overview of a famous nonparametric classifier—k-nearest neighbor rule in the fuzzy version.

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX

2. FUZZY CLUSTERING BASED ON FUZZY RELATION

Let X be a subset of an s-dimensional Euclidean space $\mathbb{R}^s$ with its ordinary Euclidean norm $\|\cdot\|$ and let c be a positive integer bigger than one. A partition of X into c clusters can be represented by mutually disjoint sets $B_1, \dots, B_c$ such that $B_1 \cup \dots \cup B_c = X$ or equivalently by the indicator function $\mu_1, \dots, \mu_c$ such that $\mu_i(x) = 1$ if $x \in B_i$ and $\mu_i(x) = 0$ if $x \notin B_i$ for all $x \in X$ and all $i = 1, \dots, c$. In this case, X is said to have a hard c-partition by the indicator functions $\mu = (\mu_1, \dots, \mu_c)$.

The fuzzy set, first proposed by Zadeh [4] in 1965, is an extension to allow $\mu_i(x)$ to be a function (called a membership function) assuming values in the interval [0,1]. Ruspini [6] introduced a fuzzy c-partition $\mu = (\mu_1, \dots, \mu_c)$ by the extension to allow $\mu_i(x)$ to be functions assuming values in the interval [0,1] such that $\mu_1(x) + \dots + \mu_c(x) = 1$ since he first applied the fuzzy set in cluster analysis.

A (hard) relation r in X is defined to be a function $r : X \times X \to \{0,1\}$ where x and y in X are said to have relation if $r(x,y) = 1$. A (hard) relation r in X is called an equivalence relation if and only if for all $x, y \in X$,

(1) $R(x,x) = 1$ (reflexivity),
(2) $r(x,y) = r(y,x)$ (symmetry), and
(3) $r(x,z) = r(y,z) = 1$ for some $z \in X \Longrightarrow r(x,y) = 1$ (transitivity).

Zadeh [4] defined a fuzzy relation r in X by an extension to allow the values of r in the interval [0,1], where $r(x,y)$ denotes the strength of the relation between x and y. In Zadeh [10], he defined a similarity relation S in X if and only if S is a fuzzy relation and for all $x, y, z \in X$, $S(x,x) = 1$ (reflexivity), $S(x,y) = S(y,x)$ (symmetry), and $S(x,y) \geq \vee_{z \in X}(S(x,z) \wedge S(y,z))$ (transitivity) where $\vee$ and $\wedge$ denote max and min. It is obvious that the similarity relation S is essentially a generalization of the equivalence relation.

Let us consider a finite data set $X = \{x_1, \dots, x_n\}$ of $\mathbb{R}^s$. Denote $\mu_i(x_j)$ by μ_{ij}, $i = 1, \dots, c$, $j = 1, \dots, n$, and $r(x_j, x_k)$ by r_{jk}, $j, k = 1, \dots, n$. Let V_{cn} be the usual vector space of real $c \times n$ matrices and let u_{ij} be the ij^{th} element of $U \in V_{c \times n}$. Following Bezdek [7], we define the following:

$$M_c = \left\{ U \in V_{cn} \mid \mu_{ij} \in \{0,1\}\ \forall i,j; \qquad \sum_{i=1}^{c} \mu_{ik} = 1\ \forall k; \qquad 0 < \sum_{j=1}^{n} \mu_{ij}\ \forall i \right\}.$$

Then M_c is exactly a (nondegenerate) hard c-partitions space for the finite data set X. Define $A \leq B$ if and only if $a_{ij} \leq b_{ij}\ \forall i,j$ where $A = [a_{ij}]$ and $B = [b_{ij}] \in V_{nn}$. Define $R \circ R = \left[r'_{ij}\right] \in V_{nn}$ with $r'_{ij} = \vee_{k=1}^{n} (r_{ik} \wedge r_{kj})$. Let

$$R_n = \left\{ R \in V_{nn} \mid r_{ij} \in \{0,1\}\ \forall i,j; \qquad I \leq R; \qquad R = R^T; \qquad R = R \circ R \right\}.$$

Then R_n is the set of all equivalence relations in X. For any $U = [\mu_{ij}] \in M_c$, let the relation matrix $R = [r_{jk}]$ in V_{nn} be defined by

$$r_{jk} = \begin{cases} 1, & \text{if } \mu_{ij} = \mu_{ik} = 1 \text{ for some } i, \\ 0, & \text{otherwise.} \end{cases}$$

Then R is obviously an equivalence relation corresponding to the hard c-partitions U, since it satisfies reflexivity, symmetry, and transitivity. That is, for any $U \in M_c$ there is a relation matrix R in R_n such that R is an equivalence relation corresponding to U.

For a fuzzy extension of M_c and R_n, let

$$M_{fc} = \left\{ V \in V_{cn} \mid \mu_{ij} \in [0,1]\ \forall i,j; \qquad \sum_{i=1}^{c} \mu_{ij} = 1\ \forall j; \qquad \sum_{j=1}^{n} \mu_{ij} > 0\ \forall i \right\},$$

and

$$R_{fn} = \{R \in V_{nn} \mid r_{ij} \in [0,1]\ \forall i,j; \qquad I \le R; \qquad R = R^T \text{ and } R \ge R \circ R\}.$$

Then M_{fc} is a (nondegenerate) fuzzy c-partitions space for X, and R_{fn} is the set of all similarity relations in X. Note that Bezdek and Harris [11] showed that $M_c \subset M_{co} \subset M_{fc}$ and that M_{fc} is the convex hull of M_{co} where M_{co} is the set of matrices obtained by relaxing the last condition of M_c to $\sum_{j=1}^{n} \mu_{ij} \ge 0\ \forall i$.

Above we have described the so-called hard c-partitions and fuzzy c-partitions, also the equivalence relation and the similarity relation. These are main representations in cluster analysis since cluster analysis is just to partition a data set into c clusters for which there is always an equivalence relation correspondence to the partition. Next, we shall focus on the fuzzy clustering methods based on the fuzzy relation. These can be found in [10,12–16].

Fuzzy clustering, based on fuzzy relations, was first proposed by Tamura *et al.* [12]. They proposed an N-step procedure by using the composition of fuzzy relations beginning with a reflexive and symmetric fuzzy relation R in X. They showed that there is an n such that $I \le R \le R^2 \le \cdots \le R^n = R^{n+1} = \cdots = R^\infty$ when X is a finite data set. Then R^n is used to define an equivalence relation R_λ by the rule $R_\lambda(x,y) = 1 \iff \lambda \le R^n(x,y)$. In fact, R^n is a similarity relation. Consequently, we can partition the data set into some clusters by the equivalence relation R_λ. For $0 \le \lambda_k \le \cdots \le \lambda_2 \le \lambda_1 \le 1$, one can obtain a corresponding k-level hierarchy of clusters, $D_i = \{\text{equivalence clusters of } R_{\lambda_i} \text{ in } X\}$, $i = 1, \dots, k$ where D_i refines D_j for $i < j$. This hierarchy of clusters is just a single linkage hierarchy which is a well-known hard clustering method. Dunn [14] proposed a method of computing R^n based on Prim's minimal spanning tree algorithm. Here, we give a simple example to explain it.

EXAMPLE. Let

$$R = \begin{bmatrix} 1 & & & \\ 0.4 & 1 & & \\ 0.8 & 0.6 & 1 & \\ 0.3 & 0 & 0 & 1 \end{bmatrix},$$

$$R^2 = \begin{bmatrix} 1 & & & \\ 0.6 & 1 & & \\ 0.8 & 0.6 & 1 & \\ 0.3 & 0.3 & 0.3 & 1 \end{bmatrix} = R^3 = \begin{bmatrix} 1 & & & \\ 0.6 & 1 & & \\ 0.8 & 0.6 & 1 & \\ 0.3 & 0.3 & 0.3 & 1 \end{bmatrix}.$$

Therefore,

$$\begin{aligned} \lambda = 0.29 &\Longrightarrow \{1,2,3,4\}, \\ \lambda = 0.59 &\Longrightarrow \{1,2,3\} \cup \{4\}, \\ \lambda = 0.79 &\Longrightarrow \{1,3\} \cup \{2\} \cup \{4\}, \\ \lambda = 0.81 &\Longrightarrow \{1\} \cup \{2\} \cup \{3\} \cup \{4\}. \end{aligned}$$

3. FUZZY CLUSTERING BASED ON OBJECTIVE FUNCTIONS

In Section 2, we introduced fuzzy clustering based on fuzzy relation. This is one type of fuzzy clustering. But these methods are eventually the novel methods and nothing more than the single linkage hierarchy of (hard) agglomerative hierarchical clustering methods. Nothing comes out fuzzy there. This is why there is not so much research and very few results in this type of fuzzy clustering. If we add the concept of fuzzy c-partitions in clustering, the situation shall be totally changed. Fuzzy charateristics could be represented by these fuzzy c-partitions. In this section, we give a survey of this kind of fuzzy clustering.

Let a distance (or dissimilarity) function d be given, where d is defined on a finite data set $X = \{x_1, \dots, x_n\}$ by $d : X \times X \to \mathbb{R}^+$ with $d(x,x) = 0$ and $d(x,y) = d(y,x)$ for all $x, y \in X$. Let

$d_{ij} = d(x_i, x_j)$ and $D = [d_{ij}] \in V_{nn}$. Ruspini [6] first proposed the fuzzy clustering based on an objective function of $J_R(U)$ to produce an optimal fuzzy c-partition U in M_{fcn}, where

$$J_R(U) = \sum_{k=1}^{n} \sum_{j=1}^{n} \left(\sigma \sum_{i=1}^{c} \left(\mu_{ij} - \mu_{ik}\right)^2 - d_{jk}^2 \right)^2$$

with $\sigma \in \mathbb{R}$. He derived the necessary conditions for a minimizer U^* of $J_R(U)$. Ruspini [17,18] continued making more numerical experiments. Since J_R is quite complicated and difficult to interpret, the Ruspini method was not so successful but he actually opened the door for further research, especially since he first put the idea of fuzzy c-partitions in cluster analysis. Notice that the only given data in Ruspini's objective function $J_R(U)$ is the dissimilarity matrix D, but not the data set X even though the element of D could be obtained from the data set X.

Based on the given dissimilarity matrix D, Roubens [19] proposed the objective function J_{RO} of which the quality index of the clustering is the weighted sum of the total dissimilarity, where

$$J_{RO}(U) = \sum_{j=1}^{n} \sum_{k=1}^{n} \sum_{i=1}^{c} \left(\mu_{ij}^2 \mu_{ik}^2\right) d_{jk}.$$

Even though Roubens' objective function J_{RO} is much more intuitive and simple than Ruspini's J_R, the proposed numerical algorithm to optimize J_{RO} is unstable. Libert and Roubens [20] gave a modification of it, but unless the substructure of D is quite distinct, useful results may not be obtained.

To overcome these problems in J_{RO}, Windham [21] proposed a modified objective function J_{AP}

$$J_{AP}(U, B) = \sum_{j=1}^{n} \sum_{k=1}^{n} \sum_{i=1}^{c} \left(\mu_{ij}^2 b_{ik}^2\right) d_{jk},$$

where $B = [b_{ik}] \in V_{cn}$ with $b_{ik} \in [0,1]$ and $\sum_{k=1}^{n} b_{ik} = 1$. Such b_{ik}'s are called prototype weights. Windham [21] presented the necessary conditions for minimizers U^* and B^* of $J_{AP}(U, B)$ and also discussed numerical convergence, its convergence rate, and computer storage. The Windham procedure is called the assignment-prototype (AP) algorithm.

Hathaway *et al.* [22] and Bezdek *et al.* [23] gave a new type of objective function J_{RFCM}, called the relational fuzzy c-mean (RFCM), where

$$J_{RFCM}(U) = \sum_{i=1}^{c} \frac{\left(\sum_{j=1}^{n} \sum_{k=1}^{n} \mu_{ij}^m \mu_{ik}^m d_{jk}\right)}{2\left(\sum_{l=1}^{n} \mu_{il}\right)^m}$$

with $m \in [1, \infty)$. They gave an iterative algorithm by using a variation of the coordinate decent method described in [24]. The objective function J_{RFCM} seems to be hardly interpreted. In fact, it is just a simple variation of the well-known fuzzy clustering algorithms, called the fuzzy c-means (FCM). The FCM becomes the major part of fuzzy clustering. We will give a big part of survey next. The numerical comparison between AP and RFCM algorithms was reported in [25].

The dissimilarity data matrix D is usually obtained indirectly and with difficultly. In general, the only given data may be the finite data set $X = \{x_1, \dots, x_n\}$. For a given data set $X = \{x_1, \dots, x_n\}$, Dunn [9] first gave a fuzzy generalization of the conventional (hard) c-means (more popularly known as k-means) by combining the idea of Ruspini's fuzzy c-partitions. He gave the objective function $J_D(U, \underset{\sim}{a})$, where

$$J_D(U, \underset{\sim}{a}) = \sum_{j=1}^{n} \sum_{i=1}^{c} \mu_{ij}^2 \left\|x_j - a_i\right\|^2$$

with $U \in M_{fcn}$ and $\underline{a} = (a_1, \dots, a_c) \in (\mathbb{R}^s)^c$ called the cluster centers. Bezdek [7] generalized $J_D(U, \underline{a})$ to $J_{FCM}(U, \underline{a})$ with

$$J_{FCM}(U, \underline{a}) = \sum_{j=1}^{n} \sum_{i=1}^{c} \mu_{ij}^m \left\| x_j - a_i \right\|^2$$

for which $m \in [1, \infty)$ represents the degree of fuzziness. The necessary conditions for a minimizer $(U^*, \underline{a}^*)$ of $J_{FCM}(U, \underline{a})$ are

$$a_i^* = \frac{\sum_{j=1}^{n} \left(\mu_{ij}^*\right)^m x_j}{\sum_{j=1}^{n} \left(\mu_{ij}^*\right)^m}, \quad \text{and} \quad \mu_{ij}^* = \left(\sum_{k=1}^{c} \frac{\left\| x_j - a_i \right\|^{2/(m-1)}}{\left\| x_j - a_k \right\|^{2/(m-1)}} \right)^{-1}, \qquad \begin{array}{l} i = 1, \dots, c, \\ j = 1, \cdots, n. \end{array}$$

The iterative algorithms for computing minimizers of $J_{FCM}(U, \underline{a})$ with these necessary conditions are called the fuzzy c-means (FCM) clustering algorithms. These FCM clustering procedures defined by Dunn [9] and generalized by Bezdek [7] have been widely studied and applied in a variety of substantive areas. There are so many papers in the literature that concern themselves with some aspects of the theory and applications of FCM. We give some references by the following categories. Interested readers may refer some of these:

(a) Numerical Theorems [26–37]
(b) Stochastic Theorems [34,38–41]
(c) Methodologies [9,37,42–51]
(d) Applications
  (1) Image Processing [45,52–58]
  (2) Engineering Systems [59–63]
  (3) Parameter estimation [38,64–68]
  (4) Others [69–74].

Here we are interested in describing a variety of generalizations of FCM. These include six generalizations and two variations.

(A) GENERALIZATION 1

To add the effect of different cluster shapes, Gustafson and Kessel [47] generalized the FCM objective function with fuzzy covariance matrices, where

$$J_{FCM_A}(U, \underline{a}, A) = \sum_{j=1}^{n} \sum_{i=1}^{c} \mu_{ij}^m \left\| x_j - a_i \right\|_{A_i}^2$$

with $U \in M_{fcn}$, $\underline{a} = (a_1, \dots, a_c) \in (\mathbb{R}^s)^c$ and $A = (A_1, \dots, A_c)$ for which A_i are positive definite $s \times s$ matrices with $\det(A_i) = \rho_i$ being fixed and $\left\| x_j - a_i \right\|_{A_i}^2 = (x_j - a_i)^T A_i (x_j - a_i)$. The iterative algorithms use necessary conditions for a minimizer $(U, \underline{a}, A)$ of $J_{FCM_A}(U, \underline{a}, A)$ as follows:

$$\begin{aligned} a_i &= \frac{\sum_{j=1}^{n} \mu_{ij}^m x_j}{\sum_{j=1}^{n} \mu_{ij}^m}, \\ \mu_{ij} &= \left(\sum_{k=1}^{c} \frac{\left\| x_j - a_i \right\|_{A_i}^{2/(m-1)}}{\left\| x_j - a_k \right\|_{A_k}^{2/(m-1)}} \right)^{-1} \qquad \text{and} \\ A_i &= \left(\rho_i \det(S_i)\right)^{1/s} S_i^{-1}, \qquad \begin{array}{l} i = 1, \dots, c, \\ j = 1, \dots, n. \end{array} \end{aligned}$$

where $S_i = \sum_{j=1}^{n} \mu_{ij}^m (x_j - a_i)(x_j - a_i)^T$.

(B) GENERALIZATION 2

$J_{FCM}(U, \underset{\sim}{a})$ is well used to cluster the data set which has the hyper-spherical shapes. Gustafson and Kessel [47] generalized it to $J_{FCM_A}(U, \underset{\sim}{a}, A)$ for improving the ability to detect different cluster shapes (especially in different hyper-ellipsoidal shapes) in the given data set. Another attempt to improve the ability of $J_{FCM}(U, \underset{\sim}{a})$ to detect nonhyper-ellipsoidally shaped substructures was proposed by Bezdek *et al.* [42,43], called the fuzzy c-varieties (FCV) clustering algorithms.

Let us denote the linear variety of dimension $r (0 \le r < s)$ in $\mathbb{R}^s$ through the point a and spanned by the vectors $\{b_1, \dots, b_r\}$ by

$$V_r(a; b_1, \dots, b_r) = \left\{ y \in \mathbb{R}^s \mid y = a + \sum_{k=1}^{r} t_k b_k, \qquad t_k \in \mathbb{R} \right\}.$$

If $r = 0$, $V_0 = \{a\}$ is a point in $\mathbb{R}^s$; if $r = 1$, $V_1 = L(a; b) = \{y \in \mathbb{R}^s \mid y = a + tb,\ t \in \mathbb{R}\}$ is a line in $\mathbb{R}^s$; if $r = 2, V_2$ is a plane in $\mathbb{R}^s$; if $r \ge 3$, V_r is a hyperplane in $\mathbb{R}^s$; and $V_s = \mathbb{R}^s$. In general, the orthogonal distance (in the A-norm on $\mathbb{R}^s$ for which A is a positive definite $s \times s$ matrix) from $x \in \mathbb{R}^s$ to V_r, where the $\{b_i\}$ is an orthonormal basis for V_r, is

$$D_A(x, V_r) = \min_{y \in V_r} \{d_A(x, y)\} = \left(\left\| x - a \right\|_A^2 - \sum_{k=1}^{r} \left(\langle x - a, b_k \rangle_A \right)^2 \right)^{1/2},$$

$$\text{where } \langle x, y \rangle_A = x^T A y, \quad \left\| x \right\|_A^2 = \langle x, x \rangle_A \quad \text{and}$$

$$d_A(x, y) = \left\| x - y \right\|_A = \left((x - y)^T A (x - y) \right)^{1/2}.$$

Let $\underset{\sim}{a} = (a_1, \dots, a_c)$ and $\underset{\sim}{b}_k = (b_{1k}, \dots, b_{ck})$, $k = 1, \dots, r$. Consider the objective function $J_{FCV}(U, \underset{\sim}{a}, \underset{\sim}{b}_1, \dots, \underset{\sim}{b}_r)$ which is the total weighted sum of square orthogonal errors between the data set $X = \{x_1, \dots, x_n\}$ and c r-dimensional linear varieties, where

$$J_{FCV}\left(U, \underset{\sim}{a}, \underset{\sim}{b}_1, \dots, \underset{\sim}{b}_r\right) = \sum_{j=1}^{n} \sum_{i=1}^{c} \mu_{ij}^m D_{ij}^2, \quad \text{with}$$

$$D_{ij} = D_A(x_j, V_{ri}),$$

$$V_{ri} = \left\{ y \in \mathbb{R}^s \mid y = a_i + \sum_{k=1}^{r} t_k b_{ik},\ t_k \in \mathbb{R} \right\}, \qquad i = 1, \dots, c, \quad j = 1, \cdots, n, \text{ and } \quad m \in [1, \infty).$$

The FCV clustering algorithms are iterations through the necessary conditions for minimizers of J_{FCV} as follows:

$$a_i = \frac{\sum_{j=1}^{n} \mu_{ij}^m x_j}{\sum_{j=1}^{n} \mu_{ij}^m},$$

$$b_{ik} = A^{-1/2} s_{ik},$$

where s_{ik} is the k^{th} unit eigenvector of the fuzzy scatter matrix S_i with

$$S_i = A^{1/2} \left(\sum_{j=1}^{n} \mu_{ij}^m (x_j - a_i)(x_j - a_i)^T \right) A^{1/2}$$

corresponding to its k^{th} largest eigenvalue, and

$$\mu_{ij} = \left(\sum_{k=1}^{c} \frac{(D_{ij})^{2/(m-1)}}{(D_{kj})^{2/(m-1)}} \right)^{-1}, \qquad i = 1, \dots, c, \quad j = 1, \dots, n, \quad \text{and} \quad k = 1, \dots, r.$$

(C) GENERALIZATION 3

Consider any probability distribution function G on $\mathbb{R}^s$. Let $X = \{x_1, \ldots, x_n\}$ be a random sample of size n from G and let G_n be the empirical distribution which puts mass $1/n$ at each and every point in the sample $\{x_1, \ldots, x_n\}$. Then the FCM clustering procedures are to choose fuzzy c-partitions $\mu = (\mu_1, \ldots, \mu_c)$ and cluster center $\underline{a}$ to minimize the objective function

$$\begin{aligned} J_{GFCM}(G_n, \mu, \underline{a}) &= \frac{1}{n} J_{FCM}(U, \underline{a}) \\ &= \frac{1}{n} \sum_{j=1}^{n} \sum_{i=1}^{c} \mu_i^m\,(x_j) \left\|x_j - a_i\right\|^2 \\ &= \int \sum_{i=1}^{c} \mu_i^m(x) \left\|x - a_i\right\|^2 G_n\,(dx). \end{aligned}$$

Let us replace G_n by G. Then we generalized the FCM with respect to G by the minimization of $J_{GFCM}\,(G, \mu, \underline{a})$,

$$J_{GFCM}(G, \mu, \underline{a}) = \int \sum_{i=1}^{c} \mu_i^m(x) \left\|x - a_i\right\|^2 G\,(dx).$$

Consider a particular fuzzy c-partition of $\underline{a}$ in $(\mathbb{R}^s)^c$ of the form

$$\mu_i(x, \underline{a}) = \left(\sum_{j=1}^{c} \frac{\left\|x - a_i\right\|^{2/(m-1)}}{\left\|x - a_j\right\|^{2/(m-1)}} \right)^{-1}, \qquad i = 1, \ldots, c.$$

Define $\mu(\underline{a}) = \big(\mu_1(x, \underline{a}), \cdots, \mu_c(x, \underline{a})\big)$ and define that

$$\begin{aligned} L_m(G, \underline{a}) &= J_{GFCM}(G, \mu(\underline{a}), \underline{a}) \\ &= \int \sum_{i=1}^{c} \left(\sum_{j=1}^{c} \frac{\left\|x - a_i\right\|^{2/(m-1)}}{\left\|x - a_j\right\|^{2/(m-1)}} \right)^{-m} \left\|x - a_i\right\|^2 G\,(dx) \\ &= \int \left(\sum_{i=1}^{c} \left\|x - a_i\right\|^{-2/(m-1)} \right)^{1-m} G\,(dx). \end{aligned}$$

Then, for any fuzzy c-partition $\mu = \big(\mu_1(x), \ldots, \mu_c(x)\big)$ and $\underline{a} \in (\mathbb{R}^s)^c$,

$$J_{GFCM}\big(G, \mu(\underline{a}), \underline{a}\big) \leq J_{GFCM}\big(G, \mu, \underline{a}\big).$$

The reason for the last inequality is that

$$\int \sum_{i=1}^{c} \mu_i^m(x, \underline{a}) \left\|x - a_i\right\|^2 G\,(dx) \leq \int \sum_{i=1}^{c} \mu_i^m(x) \left\|x - a_i\right\|^2 G\,(dx),$$

which can be seen from the simple fact that

$$\sum_{i=1}^{c} \left(\sum_{j=1}^{c} \frac{y_i^{1/(m-1)}}{y_j^{1/(m-1)}} \right)^{-m} y_i \leq \sum_{i=1}^{c} p_i^m\, y_i$$

for $p_k \geq 0$, $y_k > 0$, $k = 1, \ldots, c$, $p_1 + \cdots + p_c = 1$. To claim the simple fact, we may take the gradient of the Lagrangian

$$h(p_1, \ldots, p_c, \lambda) = \sum_{i=1}^{c} p_i^m\, y_i - \lambda \left(\sum_{i=1}^{c} p_i - 1 \right)$$

and also check that the Hessian matrix of h with respect to $p_1, \ldots, p_c$ will be positive definite.

Let $(\mu^*, \underset{\sim}{a}^*)$ be a minimizer of $J_{GFCM}(G, \mu, \underset{\sim}{a})$ and let $\underset{\sim}{b}^*$ be a minimizer of $L_m(G, \underset{\sim}{a})$. Then,

$$\begin{aligned} J_{GFCM}(G, \mu^*, \underset{\sim}{a}^*) &\leq J_{GFCM}(G, \mu(\underset{\sim}{b}^*), \underset{\sim}{b}^*) \\ &= L_m(G, \underset{\sim}{b}^*) \\ &\leq L_m(G, \underset{\sim}{a}^*) \\ &= J_{GFCM}(G, \mu(\underset{\sim}{a}^*), \underset{\sim}{a}^*) \\ &\leq J_{GFCM}(G, \mu^*, \underset{\sim}{a}^*). \end{aligned}$$

That is, $J_{GFCM}(G, \mu^*, \underset{\sim}{a}^*) = L_m(G, \underset{\sim}{b}^*)$. Therefore, we have that solving the minimization problem of $J_{GFCM}(G, \mu, \underset{\sim}{a})$ over μ and $\underset{\sim}{a}$ is equivalent to solving the minimization problem of $L_m(G, \underset{\sim}{a})$ over $\underset{\sim}{a}$ with the specified fuzzy c-partitions $\mu(\underset{\sim}{a})$. Based on this reduced objective function $L_m(G, \underset{\sim}{a})$, Yang *et al.* [39–41] created the existence and stochastic convergence properties of the FCM clustering procedures. They derived the iterative algorithms for computing the minimizer of $L_m(G, \underset{\sim}{a})$ as follows:

Set $\underset{\sim}{a}_0 = (a_{10}, \dots, a_{c0})$. For $k \geq 1$, let $\underset{\sim}{a}_k = (a_{1k}, \dots, a_{ck})$, where for $i = 1, \dots, c$,

$$a_{ik} = \frac{\int x \mu_i^m(x, \underset{\sim}{a}_{k-1})\, G(dx)}{\int \mu_i^m(x, \underset{\sim}{a}_{k-1})\, G(dx)}, \quad \text{and}$$

$$\mu_i(x, \underset{\sim}{a}_{k-1}) = \left(\sum_{j=1}^{c} \frac{\left\| x - a_{i(k-1)} \right\|^{2/(m-1)}}{\left\| x - a_{j(k-1)} \right\|^{2/(m-1)}} \right)^{-1}.$$

They extended FCM algorithms with respect to G_n to FCM algorithms with respect to any probability distribution function G, and also showed that the optimal cluster centers should be the fixed points of these generalized FCM algorithms. The numerical convergence properties of these generalized FCM algorithms had been created in [36]. These include the global convergence, the local convergence, and its rate of convergence.

(D) GENERALIZATION 4

Boundary detection is an important step in pattern recognition and image processing. There are various methods for detecting line boundaries; see [75]. When the boundaries are curved, the detection becomes more complicated and difficult. In this case, the Hough transform (HT) (see [76,77]) is the only major technique available. Although the HT is a powerful technique for detecting curve boundaries like circles and ellipses, it requires large computer storage and high computing time. The fuzzy c-shells (FCS) clustering algorithms, proposed by Dave [45,46], are fine new techniques for detecting curve boundaries, especially circular and elliptical. These FCS algorithms have major advantages over the HT in the areas of computer storage requirements and computational requirements.

The FCS clustering algorithms are used to choose a fuzzy c-partition $\mu = (\mu_1, \dots, \mu_c)$ and cluster centers $\underset{\sim}{a} = (a_1, \dots, a_c)$ and radius $r = (r_1, \dots, r_c)$ to minimize the objective function $J_{FCS}(\mu, \underset{\sim}{a}, r)$, which is the weighted sum of squared distance of data from a hyper-ellipsoidal shell prototype

$$J_{FCS}(\mu, \underset{\sim}{a}, r) = \sum_{j=1}^{n} \sum_{i=1}^{c} \mu_i^m(x_j) \left(\left\| x_j - a_i \right\|_{A_i} - r_i \right)^2$$

among all $\underset{\sim}{a} \in (\mathbb{R}^s)^c$, $\mu \in \mathbb{R}^s$, and $r \in (\mathbb{R}^+)^c$ with $m \in [1, \infty)$. The necessary conditions for a minimizer $(\mu^*, \underline{a}^*, r^*)$ of J_{FCS} are the following equations satisfied for $\underline{a}^*$ and r^*:

$$\sum_{j=1}^{n} (\mu_i^*(x_j))^m \left(1 - \frac{r_i^*}{\|x_j - a_i^*\|_{A_i}}\right)(x_j - a_i^*) = 0,$$

$$\sum_{j=1}^{n} (\mu_i^*(x_j))^m \left(\|x_j - a_i^*\|_{A_i} - r_i^*\right) = 0$$

$$\text{with} \quad \mu_i^*(x_j) = \left(\sum_{k=1}^{c} \frac{\left(\|x_j - a_i^*\|_{A_i} - r_i^*\right)^{2/(m-1)}}{\left(\|x_j - a_k^*\|_{A_i} - r_i^*\right)^{2/(m-1)}}\right)^{-1}, \qquad \begin{array}{l} i = 1, \dots, c, \\ j = 1, \dots, n. \end{array}$$

The FCS clustering algorithms are iterations through these necessary conditions. Dave *et al.* (see [45,53]) applied these FCS's in digital image processing. Bezdek and Hathaway [78] investigated its numerical convergence properties. Krishnapuram *et al.* [79] gave a new approach to the FCS.

(E) GENERALIZATION 5

Trauwaert *et al.* [51] derived a generalization of the FCM based on the maximum likelihood principle for the c members of mutivariate normal distributions. Consider a log likelihood function $J_{FML}(\mu, \underline{a}, W)$ of a random sample $X = \{x_1, \dots, x_n\}$ from c numbers of mutivariate normals $N(a_i, W_i)$, $i = 1, \dots, c$, where

$$J_{FML}(\mu, \underline{a}, W) = -\frac{1}{2} \sum_{j=1}^{n} \sum_{i=1}^{c} \left(\mu_{ij}^{m_1} (x_j - a_i)^T W_i^{-1} (x_j - a_i) + \mu_{ij}^{m_2} \ln (2\pi)^s + \mu_{ij}^{m_3} \ln |W_i|\right),$$

in which $m_1, m_2, m_3 \in [1, \infty)$. To derive iterative algorithms for a maximizer of $J_{FML}(\mu, \underline{a}, W)$, they further limited to the case where m_1, m_2, and m_3 are either 1 or 2. They claimed that only the case of $m_1 = 2$ and $m_2 = m_3 = 1$ appeared as an acceptable basis for a membership recognition algorithm. In the case of $m_1 = 2$ and $m_2 = m_3 = 1$, Trauwaert *et al.* [51] gave the iterative algorithm as follows:

$$a_i = \frac{\sum_{j=1}^{n} \mu_{ij}^2 x_j}{\sum_{j=1}^{n} \mu_{ij}^2},$$

$$W_i = \frac{\sum_{j=1}^{n} \mu_{ij}^2 (x_j - a_i)(x_j - a_i)^T}{\sum_{j=1}^{n} \mu_{ij}}, \quad \text{and}$$

$$\mu_{ij} = \frac{1/B_{ij}}{\sum_{k=1}^{c} (1/B_{kj})} + \frac{1}{B_{ij}} \left(\frac{\sum_{k=1}^{c} (A_{kj}/B_{kj})}{\sum_{k=1}^{c} (1/B_{kj})} - A_{ij}\right), \quad \text{where}$$

$$B_{ij} = (x_j - a_i)^T W_i^{-1} (x_j - a_i), \qquad A_{ij} = \frac{1}{2} \ln |W_i|, \qquad \begin{array}{l} i = 1, \dots, c, \\ j = 1, \dots, n. \end{array}$$

This general algorithm is called "fuzzy product" since it is an optimization of the product of fuzzy determinants. In the case of $A_i = I$ for all i the algorithm becomes the FCM.

(F) GENERALIZATION 6

Mixtures of distributions have been used extensively as models in a wide variety of important practical situations where data can be viewed as arising from two or more subpopulations mixed in varied proportions. Classification maximum likelihood (CML) procedure is a remarkable mixture maximum likelihood approach to clustering, see [80]. Let the population of interest be known or be assumed to consist of c different subpopulations, or classes, where the density of an observation x from class i is $f_i(x; \theta)$, $i = 1, \dots, c$ for some parameter θ. Let the proportion of individuals in the population which are from class i be denoted by α_i with $\alpha_i \in (0,1)$ and $\sum_{i=1}^{c} \alpha_i = 1$. Let $X = \{x_1, \ldots, x_n\}$ be a random sample of size n from the population. The CML procedure is the optimization problem by choosing a hard c-partition μ^* and a proportion α^* and an estimate θ^* to maximize the log likelihood $B_1(\mu, \alpha, \theta)$ of X, where

$$\begin{aligned} B_1(\mu, \alpha, \theta) &= \sum_{i=1}^{c} \sum_{j \in I_i} \ln \alpha_i f_i(x_j; \theta) \\ &= \sum_{j=1}^{n} \sum_{k=1}^{c} \mu_i(x_j) \ln \alpha_i f_i(x_j; \theta), \end{aligned}$$

in which $\mu_i(x) \in \{0,1\}$ and $\mu_1(x) + \cdots + \mu_c(x) = 1$ for all $x \in X$. Yang [37] gave a new type of objective function $B_{m,w}(\mu, \alpha, \theta)$, where

$$B_{m,w}(\mu, \alpha, \theta) = \sum_{j=1}^{n} \sum_{i=1}^{c} \mu_i^m(x_j) \ln f_i(x_j; \theta) + w \sum_{j=1}^{n} \sum_{i=1}^{c} \mu_i^m(x_j) \ln \alpha_i,$$

in which $\mu_i(x) \in [0,1]$ and $\mu_1(x) + \cdots + \mu_c(x) = 1$ for all $x \in X$ with the fixed constants $m \in [1, \infty)$ and $w \geq 0$. Thus, $B_{m,w}(\mu, \alpha, \theta)$, called the fuzzy CML objective function, is a fuzzy extension of $B_1(\mu, \alpha, \theta)$. Based on $B_{m,w}(\mu, \alpha, \theta)$, we can derive a variety of hard and fuzzy clustering algorithms.

Now we consider the fuzzy CML objective function $B_{m,w}(\mu, \alpha, \theta)$ with the multivariate normal subpopulation densities $N(a_i, W_i)$, $i = 1, \ldots, c$. Then, we have that

$$\begin{aligned} &B_{m,w}(\mu, \alpha, \theta) \\ &= -\frac{1}{2} \sum_{j=1}^{n} \sum_{i=1}^{c} \left(\ln (2\pi)^s \mu_i^{m_1}(x_j) + \mu_i^{m_2}(x_j) \ln |W_i| + \mu_i^{m_3}(x_j)(x_j - a_i)^T W_i^{-1}(x_j - a_i) \right) \\ &\quad + w \sum_{j=1}^{n} \sum_{i=1}^{c} \mu_i^m(x_j) \ln \alpha_i \\ &= J_{FML}(\mu, \underset{\sim}{a}, W) + w \sum_{j=1}^{n} \sum_{i=1}^{c} \mu_i^m(x_j) \ln \alpha_i. \end{aligned}$$

Based on $B_{m,w}(\mu, \underset{\sim}{a}, W)$, we extend the fuzzy clustering algorithms of Trauwaert *et al.* [51] to a more general setting by adding a penalty term $\left(w \sum_{j=1}^{n} \sum_{i=1}^{c} \mu_i^m(x_j) \ln \alpha_i\right)$. The construction of these penalized types of fuzzy clustering algorithms is similar to that of [51]. In the special case of $W_i = I$, $i = 1, \ldots, c$. We have that

$$\begin{aligned} J_{PFCM}(\mu, \underset{\sim}{a}, \alpha) &= \sum_{j=1}^{n} \sum_{i=1}^{c} \mu_i^m(x_j) \|x_j - a_i\|^2 - w \sum_{j=1}^{n} \sum_{i=1}^{c} \mu_i^m(x_j) \ln \alpha_i \\ &= J_{FCM}(\mu, \underset{\sim}{a}) - w \sum_{j=1}^{n} \sum_{i=1}^{c} \mu_i^m(x_j) \ln \alpha_i. \end{aligned}$$

Since J_{PFCM} just adds the penalty term $\left(-w \sum_{j=1}^{n} \sum_{i=1}^{c} \mu_i^*(x_j) \ln \alpha_i\right)$ to J_{FCM}, J_{PFCM} is called the penalized FCM objective function. We get the following necessary conditions for a minimizer $(\mu, \underset{\sim}{a}, \alpha)$ of J_{PFCM}:

$$a_i = \frac{\sum_{j=1}^{n} \mu_i^m(x_j) x_j}{\sum_{j=1}^{n} \mu_i^m(x_j)},$$

$$\alpha_i = \frac{\sum_{j=1}^{n} \mu_i^m(x_j)}{\sum_{j=1}^{n} \sum_{i=1}^{c} \mu_i^m(x_j)}, \quad \text{and}$$

$$\mu_i(x_j) = \left(\sum_{k=1}^{c} \frac{\left(\left\| x_j - a_i \right\|^2 - w \ln \alpha_i \right)^{1/(m-1)}}{\left(\left\| x_j - a_k \right\|^2 - w \ln \alpha_k \right)^{1/(m-1)}} \right)^{-1}, \qquad i = 1, \cdots, c, \quad j = 1, \cdots, n.$$

Thus, the penalized FCM clustering algorithms for computing the minimizer of $J_{PFCM}\,(\mu, \underset{\sim}{a}, \alpha)$ are iterations through these necessary conditions. By comparison between the penalized FCM and the FCM, the difference is the penalty term. This penalized FCM has been used in parameter estimation of the normal mixtures. The numerical comparison between the penalized FCM and EM algorithms for parameter estimation of the normal mixtures has been studied in [68].

(G) Other Variations

From (A) to (F), we give a survey of six generalizations of the FCM. These generalizations are all based on the Euclidean norm $\| \cdot \|$ (i.e., L_2-norm). In fact, we can extend these to L_p-norm with

$$J_{FCM_p}\,(U, \underset{\sim}{a}) = \sum_{j=1}^{n} \sum_{i=1}^{c} \mu_{ij}^m \, d_{ij}$$

where $d_{ij} = \left(\sum_{k=1}^{s} \left| x_{jk} - a_{ik} \right|^p \right)^{1/p}$ is the L_p-norm of $(x_j - a_i)$. When $p = 2$ it is just the FCM. In Mathematics and its applications, there are two extreme L_p-norms most concerned. These are L_1-norm and L_∞-norm. Thus,

$$J_{FCM_1}(U, \underset{\sim}{a}) = \sum_{j=1}^{n} \sum_{i=1}^{c} \mu_{ij}^m \, d_{ij} \qquad \text{with } d_{ij} = \sum_{k=1}^{s} \left| s_{jk} - a_{ik} \right| \quad \text{and}$$

$$J_{FCM_\infty}(U, \underset{\sim}{a}) = \sum_{j=1}^{n} \sum_{i=1}^{c} \mu_{ij}^m \, d_{ij} \qquad \text{with } d_{ij} = \max_{k=1,\ldots,s} \left| s_{jk} - a_{ik} \right|.$$

We cannot find the necessary conditions to guide numerical solutions of these nonlinear constrained objective functions J_{FCM_1} and J_{FCM_∞}. Bobrowski and Bezdek [44] gave a method to find approximate minimizers of J_{FCM_1} and J_{FCM_∞} by using a basis exchange algorithm used for discriminate analysis with the perception criterion function by Bobrowski and Niemiro [81]. Jajuga [49] also gave a method to find approximate minimizers of J_{FCM_1} based on a method for regression analysis with the L_1-norm by Bloomfield and Steiger [82]. Jajuga [49] gave another simple method to find approximate minimizers of J_{FCM_1} which we shall describe next.

Recall that

$$J_{FCM_1}(U, \underset{\sim}{a}) = \sum_{j=1}^{n} \sum_{i=1}^{c} \mu_{ij}^m \, d_{ij} \qquad \text{with } d_{ij} = \sum_{k=1}^{s} \left| x_{jk} - a_{ik} \right|.$$

Then U may be a minimizer of J_{FCM_1} for $\underset{\sim}{a}$ being fixed only if

$$\mu_{ij} = \left(\sum_{k=1}^{c} \frac{\left(d_{ij} \right)^{1/(m-1)}}{\left(d_{kj} \right)^{1/(m-1)}} \right)^{-1}.$$

For U being fixed, let $w_{ij} = \dfrac{\mu_{ij}^m}{\left| x_{jk} - a_{ik} \right|}$. Then,

$$J_{FCM_1}(w, \underset{\sim}{a}) = \sum_{i=1}^{c} \sum_{k=1}^{s} \sum_{j=1}^{n} w_{ij} \left(x_{jk} - a_{ik} \right)^2.$$

The a_{ik} may be a minimizer of $J_{FCM_1}(w, \underset{\sim}{a})$ for w_{ij} being fixed only if

$$a_{ij} = \frac{\sum_{j=1}^{n} w_{ij} x_{jk}}{\sum_{j=1}^{n} w_{ij}}, \qquad \begin{array}{l} i = 1, \dots, c, \\ k = 1, \cdots, s. \end{array}$$

Jajuga [49] suggested an algorithm for finding approximate minimizers of J_{FCM_1} based on the following steps:

$$w_{ij} = \frac{\mu_{ij}^{m}}{\left|x_{jk} - a_{ik}\right|},$$

$$a_{ik} = \frac{\sum_{j=1}^{n} w_{ij} x_{jk}}{\sum_{j=1}^{n} w_{ij}}, \quad \text{and}$$

$$\mu_{ij} = \left(\sum_{k=1}^{c} \frac{\left(d_{ij}\right)^{1/(m-1)}}{\left(d_{kj}\right)^{1/(m-1)}} \right)^{-1}, \qquad d_{ij} = \sum_{k=1}^{x} \left|x_{jk} - a_{ik}\right|, \qquad \text{where} \quad \begin{array}{l} i = 1, \dots, c, \\ j = 1, \dots, n, \\ k = 1, \dots, s. \end{array}$$

4. FUZZY GENERAL k-NEAREST NEIGHBOR RULE

In the conventional nearest neighbor pattern classification problem, X represents the set of patterns and θ represents the labeling variable of c categories. We assume that a set of n correctly classified samples (x_1, θ_1), $(x_2, \theta_2), \dots, (x_n, \theta_n)$ is given and follows some distribution $F(x, \theta)$, where the $x_i's$ take values in a metric space (X, d) and the $\theta_i's$ take values in the set $\{1, 2, \dots, c\}$. A new pair (x_0, θ_0) is given, where only the measurement x_0 is observable by the statistician, and it is desired to estimate θ_0 by utilizing the information contained in the set of correctly classified points. We shall call $x'_n \in \{x_1, x_2, \dots, x_n\}$ a nearest neighbor to x if $d(x'_n, x) = \min_{i=1,2,\dots,n} d(x_i, x)$. The nearest neighbor rule (NNR) has been a well-known decision rule and perhaps the simplest nonparametric decision rule. The NNR, first proposed by Fix and Hodges [83], assigns x to the category θ'_n of its nearest neighbor x'_n.

Cover and Hart [84] proved that the large sample NN risk R is bounded above by twice the Bayes probability of error R^* under the $0-1$ loss function L, where L counts an error whenever a mistake in classification is made. Cover [85] extended the finite-action (classification) problem of Cover and Hart [84] to the infinite-action (estimation) problem. He proved that the NN risk R is still bounded above by twice the Bayes risk for both metric and squared-error loss function. Moreover, Wagner [86] considered the posterior probability of error $L_n = P\left[\theta'_n \neq \theta_0 \mid (x_1, \theta_1), \dots, (x_n, \theta_n)\right]$. Here L_n is a random variable that is a function of the observed samples and, furthermore, $EL_n = P\left[\theta'_n \neq \theta_0\right]$, and he proved that under some conditions, L_n converges to R with probability one as n tends to ∞. This obviously extended the results of Cover and Hart [84]. There are so many papers about k-NNR in the literature. The k-NNR is a supervised approach to pattern recognition. Although k-NNR is not a clustering algorithm, it is always used to cluster an unlabelled data set through the rule. In fact, Wong and Lane [87] combined clustering with k-NNR and developed the so-called k-NN clustering procedure.

In 1982, Duin [88] first proposed strategies and examples in the use of continuous labeling variable. Jóźwik [89] derived a fuzzy k-NNR in use of Duin's idea and fuzzy sets. Bezdek *et al.* [90] generalized k-NNR in conjunction with the FCM clustering. Keller *et al.* [91] proposed fuzzy k-NNR which assigns class membership to a sample vector rather than assigning the vector to a particular class. Moreover, Béreau and Dubuisson [92] gave a fuzzy extended k-NNR by choosing a particular exponential weight function. But in a sequence of study in the fuzzy version of k-NNR, there is no result about convergence properties. Recently, Yang and Chen [93] described a fuzzy generalized k-NN algorithm which is a unified approach to a variety of fuzzy

k-NNR's and then created the stochastic convergence property of the fuzzy generalized NNR, that is, the strong consistency of posterior risk of the fuzzy generalized NNR. Moreover, Yang and Chen [94] gave its rate of convergence. They showed that the rate of convergence of posterior risk of the fuzzy generalized NNR is exponentially fast. Since the conventional k-NNR is a special case of the fuzzy generalized k-NNR, the strong consistency in [93] extends the results of Cover and Hart [84], Cover [85], and Wagner [86].

The fuzzy generalized k-NNR had been shown that it could improve the performance rate in most instances. A lot of researchers gave numerical examples and their applications. The results in [93,94] just give us the theoretic foundation of the fuzzy generalized k-NNR. These also confirm the value of their applications.

5. CONCLUSIONS

We have already reviewed numerous fuzzy clustering algorithms. But it is necessary to pre-assume the number c of clusters for all these algorithms. In general, the number c should be unknown. Therefore, the method to find optimal c is very important. This kind of problem is usually called cluster validity.

If we use the objective function J_{FCM} (or J_{FCV}, J_{FCS}, etc.) as a criterion for cluster validity, it is clear that J_{FCM} must decrease monotonically with c increasing. So, in some sense, if the sample of size n is really grouped into $\hat{c}$ compact, well-separated clusters, one would expect to see J_{FCM} decrease rapidly until $c = \hat{c}$, but it shall decrease much more slowly thereafter until it reaches zero at $c = n$. This argument has been advanced for hierarchical clustering procedures. But it is not suitable for clustering procedures based on the objective function. An effectively formal way to proceed is to devise some validity criteria such as a cluster separation measure or to use other techniques such as bootstrap technique. We do not give a survey here. But interested readers may be directed towards references [95–103].

Finally, we conclude that the fuzzy clustering algorithms have obtained great success in a variety of substantive areas. Our survey may give a good extensive view of researchers who have concerns with applications in cluster analysis, or even encourage readers in the applied mathematics community to try to use these techniques of fuzzy clustering.

REFERENCES

1. R.O. Duda and P.E. Hart, *Pattern Classification and Scene Analysis*, Wiley, New York, (1973).
2. P.A. Devijver and J. Kitttler, *Pattern Recognition: A Statistical Approach*, Prentice-Hall, Englewood Cliffs, NJ, (1982).
3. A.K. Jain and R.C. Dubes, *Algorithms for Clustering Data*, Prentice-Hall, Englewood Cliffs, NJ, (1988).
4. L.A. Zadeh, Fuzzy sets, *Inform. and Control* **8**, 338–353 (1965).
5. R. Bellman, R. Kalaba, and L.A. Zadeh, Abstraction and pattern classification, *J. Math. Anal. and Appl.* **2**, 581–586 (1966).
6. E.H. Ruspini, A new approach to clustering, *Inform. and Control* **15**, 22–32 (1969).
7. J.C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York, (1981).
8. S.K. Pal and D.D. Majumdar, *Fuzzy Mathematical Approach to Pattern Recognition*, Wiley (Halsted), New York, (1986).
9. J.C. Dunn, A fuzzy relative of the ISODATA process and its use in detecting compact, well-separated clusters, *J. Cybernetics* **3**, 32–57 (1974).
10. L.A. Zadeh, Similaritiy relations and fuzzy orderings, *Inform. Science*, 177–200 (1971).
11. J.C. Bezdek and J.D. Harris, Convex decompositions of fuzzy partitions, *J. Math. Anal. and Appl.* **67** (2), 490–512 (1980).
12. S. Tamura, S. Higuchi and K. Tanaka, Pattern classificaiton based on fuzzy relations, *IEEE Trans. Syst. Man Cybern.* **1** (1), 61–66 (1971).
13. A. Kandel and L. Yelowitz, Fuzzy chains, *IEEE Trans. Syst. Man Cybern.*, 472–475 (1974).
14. J.C. Dunn, A graph theoretic analysis for pattern classification via Tamura's fuzzy relation, *IEEE Trans. Syst. Man Cybern.* **4** (3), 310–313 (1974).
15. J.C. Bezdek and J.D. Harris, Fuzzy relations and partitions: An axiomatic basis for clustering, *Fuzzy Sets and Systems* **1**, 111–127 (1978).

16. B.N. Chatterji, Character recognition using fuzzy similarity relations, In *Approximate Reasoning in Decision Analysis*, (Edited by M.M. Gupta and E. Sanchez), pp. 131–137, Amsterdam, New York, (1982).
17. E.H. Ruspini, Numerical methods for fuzzy clustering, *Information Science* **2**, 319–350 (1970).
18. E.H. Ruspini, New experimental results in fuzzy clustering, *Inform. Sciences* **6**, 273–284 (1973).
19. M. Roubens, Pattern classification problems with fuzzy sets, *Fuzzy Sets and Systems* **1**, 239–253 (1978).
20. G. Libert and M. Roubens, Non-metric fuzzy clustering algorithms and their cluster validity, In *Fuzzy Information and Decision Processes*, (Edited by M. Gupta and E. Sanchez), pp. 417–425, Elsevier, New York, (1982).
21. M.P. Windham, Numerical classification of proximity data with assignment measures, *J. Classification* **2**, 157–172 (1985).
22. R.J. Hathaway, J.W. Davenport and J.C. Bezdek, Relational duals of the c-means algorithms, *Pattern Recognition* **22**, 205–212 (1989).
23. J.C. Bezdek and R.J. Hathaway, Dual object-relation clustering models, *Int. J. General Systems* **16**, 385–396 (1990).
24. J.C. Bezdek, R.J. Hathaway, R.E. Howard, C.E. Wilson and M.P. Windham, Local convergence analysis of a grouped variable version of coordinate descent, *J. Optimization Theory and Applications* **54** (3), 471–477 (1986).
25. J.C. Bezdek, R.J. Hathaway and M.P. Windham, Numerical comparison of the RFCM and AP algorithms for clustering relational data, *Pattern Recognition* **24** (8), 783–791 (1991).
26. J.C. Bezdek, A convergence theorem for the fuzzy ISODATA clustering algorithms, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2** (1), 1–8 (1980).
27. J.C. Bezdek, R.J. Hathaway, M.J. Sabin and W.T. Tucker, Convergence theory for fuzzy c-means: Counterexamples and repairs, *IEEE Trans. Syst. Man Cybern.* **17** (5), 873–877 (1987).
28. R. Cannon, J. Dave and J.C. Bezdek, Efficient implementation of the fuzzy c-means clustering algorithms, *IEEE Trans. Pattern Anal. machine Intell.* **8** (2), 248–255 (1986).
29. R.J. Hathaway and J.C. Bezdek, Local convergence of the fuzzy c-means algorithms, *Pattern Recognition* **19**, 477–480 (1986).
30. M.A. Ismail and S.A Selim, On the local optimality of the fuzzy ISODATA clustering algorithm, *IEEE Trans. Pattern Anal. Machine Intell.* **8** (2), 284–288 (1986).
31. M.A. Ismail and S.A Selim, Fuzzy c-means: Optimality of solutions and effective termination of the algorithm, *Pattern Recognition* **19** (6), 481–485 (1986).
32. M.S. Kamel and S.Z. Selim, A thresholded fuzzy c-means algorithm for semi-fuzzy clustering, *Pattern Recognition* **24** (9), 825–833 (1991).
33. T. Kim, J.C. Bezdek and R. Hathaway, Optimality test for fixed points of the fuzzy c-means algorithm, *Pattern Recognition* **21** (6), 651–663 (1988).
34. M.J. Sabin, Convergence and consistency of fuzzy c-means/ISODATA algorithms, *IEEE Trans. Pattern Anal. Machine Intell.* **9**, 661–668 (1987).
35. S.Z. Selim and M.S. Kamel, On the mathematical and numerical properties of the fuzzy c-means algorithm, *Fuzzy Sets and Systems* **49**, 181–191 (1992).
36. M.S. Yang, Convergence properties of the generalized fuzzy c-means clustering algorithms, *Computers Math. Applic.* **25** (12), 3–11 (1993).
37. M.S. Yang, On a class of fuzzy classification maximum likelihood procedures, *Fuzzy Sets and Systems* **57** (3), 365–375 (1993).
38. R.J. Hathaway and J.C. Bezdek, On the asymptotic properties of fuzzy c-means cluster prototypes as estimators of mixture subpopulations centers, *Commum. Statist.—Theory and Method* **15** (2), 505–513 (1986).
39. M.S. Yang and K.F. Yu, On stochastic convergence theorems for the fuzzy c-means clustering procedure, *Int. J. General Systems* **16**, 397–411 (1990).
40. M.S. Yang and K.F. Yu, On existence and strong consistency of a class of fuzzy c-means clustering procedures, *Cybernetics and Systems* **23**, 583–602 (1992).
41. M.S. Yang, On asymptotic normality of a class of fuzzy c-means clustering procedures, (submitted).
42. J.C. Bezdek, C. Coray, R. Gunderson and J. Watson, Detection and characterization of cluster substructure: I. Linear structure: Fuzzy c-lines, *SIAM J. Appl. Math.* **40** (2), 339–357 (1981).
43. J.C. Bezdek, C. Coray, R. Gunderson and J. Watson, Detection and characterization of cluster substructure: II. Fuzzy c-varieties and convex combinations thereof, *SIAM J. Appl. Math.* **40** (2), 358–372 (1981).
44. L. Bobrowski and J.C. Bezdek, C-means with l_1 and l_∞ norms, *IEEE Trans. Syst. Man Cybern.* **21** (3), 545–554 (1991).
45. N.R. Dave, Fuzzy shell-clustering and applications to circle detection in digital images, *Int. J. General Systems* **16** (4), 343–355 (1990).
46. R.N. Dave, Generalized fuzzy c-shells clustering and detection of circular and elliptical boundaries, *Pattern Recognition* **25** (7), 713–721 (1992).
47. R.W. Gunderson, An adaptive FCV clustering algorithm, *Int. J. Man Machine Studies* **19** (1), 97–104 (1983).
48. D. Gustafson and W. Kessel, Fuzzy clustering with a fuzzy covariance matrix, *Proc. IEEE, San Diego*, pp. 761–766, (1978).
49. K. Jajuga, L_1-norm based fuzzy clustering, *Fuzzy Sets and System* **39**, 43–50 (1991).

50. W. Pedrycz, Algorithms of fuzzy clustering with partial supervision, *Patt. Recog. Letters* **3**, 13–20 (1985).
51. E. Trauwaert, L. Kaufman and P. Rousseeuw, Fuzzy clustering algorithms based on the maximum likelihood principle, *Fuzzy Sets and Systems* **42**, 213–227 (1991).
52. R. Cannon, J. Dave, J.C. Bezdek and M. Trivedi, Segmentation of a thematic mapper image using the fuzzy *c*-means clustering algorithm, *IEEE Trans. Geo. & Remote Sensing* **24** (3), 400–408 (1986).
53. R.N. Dave and K. Bhaswan, Adaptive fuzzy *c*-shells clustering and detection of ellipses, *IEEE Trans. on Neural Networks* **3** (5), 643–662 (1992).
54. L.O. Hall, A.M. Bensaid, L.P. Clarke, R.P. Velthuizen, M.S. Silbiger and J.C. Bezdek, A comparison of neural network and fuzzy clustering techniques in segmenting magnetic resonance images of the brain, *IEEE Trans. on Neural Networks* **3** (5), 672–682 (1992).
55. T.L. Huntsberger and M. Descalzi, Color edge detection, *Pattern Recog. Letters* **3**, 205–209 (1985).
56. T.L. Huntsberger, C.L. Jacobs and R.L. Cannon, Iterative fuzzy image segmentation, *Pattern Recog.* **18** (2), 131–138 (1985).
57. J.M. Keller, D. Subhangkasen, K. Unklesbay and N. Unklesbay, An approximate reasoning technique for recognition in color images of beef steaks, *Int. J. General Systems* **16**, 331–342 (1990).
58. M. Trivedi and J.C. Bezdek, Low level segmentation of aerial images with fuzzy clustering, *IEEE Trans. Syst. Man Cybern.* **16** (4), 589–597 (1986).
59. I. Anderson, J. Bezdek and R. Dave, Polygonal shape descriptions of plane boundaries, In *Systems Science and Science*, Vol. 1, pp. 295–301, SGSR Press, Louisville, (1982).
60. J.C. Bezdek and I. Anderson, Curvature and tangential deflection of discrete arcs, *IEEE Trans. Pattern Anal. Machine Itell.* **6** (1), 27–40 (1984).
61. J.C. Bezdek and I. Anderson, An application of the *c*-varieties clustering algorithms to polygonal curve fitting, *IEEE Trans. Syst. Man Cybern.* **15** (5), 637–641 (1985).
62. K.P. Chan and Y.S. Cheung, Clustering of cluster, *Pattern Recognition* **25**, 211–217 (1992).
63. T.L. Huntsberger and P. Ajjimarangsee, Parallel self-organizing feature maps for unsupervised pattern recognition, *Int. J. General Systems* **16**, 357–372 (1990).
64. J.C. Bezdek and J.C. Dunn, Optimal fuzzy partitions: A heuristic for estimating the parameters in a mixture of normal distributions, *IEEE Transactions on Computers* **24** (8), 835–838 (1975).
65. J.C. Bezdek, R.J. Hathaway and V.J. Huggins, Parametric estimation for normal mixtures, *Pattern Recognition Letters* **3**, 79–84 (1985).
66. J.W. Davenport, J.C. Bezdek and R.J. Hathaway, Parameter estimation for finite mixture distributions, *Computers Math. Applic.* **15** (10), 819–828 (1988).
67. S.K. Pal and S. Mitra, Fuzzy dynamic clustering algorithm, *Pattern Recognition Letter* **11**, 525–535 (1990).
68. M.S. Yang and C.F. Su, On parameter estimation for the normal mixtures based on fuzzy clustering algorithms, (submitted).
69. J.C. Bezdek, Numerical taxonomy with fuzzy sets, *J. Math. Bio.* **1** (1), 57–71 (1974).
70. J.C. Bezdek and P. Castelaz, Prototype classification and feature selection with fuzzy sets, *IEEE Trans. Syst. Man Cybern.* **7** (2), 87–92 (1977).
71. J.C. Bezdek and W.A. Fordon, The application of fuzzy set theory to medical diagnosis, In *Advances in Fuzzy Set Theory and Applications*, pp. 445–461, North-Holland, Amsterdam, (1979).
72. G. Granath, Application of fuzzy clustering and fuzzy classification to evaluate provenance of glacial till, *J. Math Geo.* **16** (3), 283–301 (1984).
73. A.B. Mcbratney and A.W. Moore, Application of fuzzy sets to climatic classification, *Agricultural and Forest Meteor* **35**, 165–185 (1985).
74. C. Windham, M.P. Windham, B. Wyse and G. Hansen, Cluster analysis to improve food classification within commodity groups, *J. Amer. Diet. Assoc.* **85** (10), 1306–1315 (1985).
75. D.H. Ballard and C.M. Brown, *Computer Vision*, Prentice-Hall, Englewood Cliffs, NJ, (1982).
76. J. Illingworth and J. Kittler, A survey of Hough transforms, *Computer Vision, Graphics and Image Processing*, 87–116 (1988).
77. E.R. Davies, Finding ellipses using the generalized Hough transform, *Pattern Recognition Lett.* **9**, 87–96 (1989).
78. J.C. Bezdek and R.J. Hathaway, Numerical convergence and interpretation of the fuzzy c-shells clustering algorithm, *IEEE Trans. on Neural Networks* **3** (5), 787–793 (1992).
79. R. Krishnapuram, O. Nasraoui and H. Frigui, The fuzzy *c* spherical shells algorithm: A new approach, *IEEE Trans on Neural Networks* **3** (5), 663–671 (1992).
80. G.J. McLachlan and K.E. Basfork, *Mixture Models: Inference and Applicaitons to Clustering*, Marcel Dekker, NY, (1988).
81. L. Bobrowski and W. Niemiro, A method of synthesis of linear discriminant function in the case of nonseparability, *Pattern Recognition* **17** (2), 205–210 (1984).
82. P. Bloomfield and W.L. Steiger, *Least Absolute Deviations*, Birkhäuser, Boston, MA, (1983).
83. E. Fix and J.L. Hodges, Jr., Discriminatory analysis, nonparametric discrimination, consistency properties, Project 21-49-004, Report No.4, USAF School of Aviation Medicine, Randolph Field, TX, (1951).
84. T.M. Cover and P.E. Hart, Nearest neighbor pattern classification, *IEEE Trans. Inform. Theory* **13**, 21–27 (1967).
85. T.M. Cover, Estimation by the nearest neighbor rule, *IEEE Trans. Inform. Theory* **14**, 50–55 (1968).

86. T.J. Wagner, Convergence of the nearest neighbor rule, *IEEE Trans. Information Theory* **17**, 566–571 (1971).
87. M.A. Wong and T. Lane, A k^{th}-nearest neighbor clustering procedure, *J. R. Statist. Soc.* **B 45** (3), 362–368 (1983).
88. R.P.W. Duin, The use of continuous variables for labelling objects, *Patt. Recog. Letters* **1**, 15–20 (1982).
89. A. Jóźwik, A learning scheme for a fuzzy k-NN rule, *Pattern Recognition Lett.* **1**, 287–289 (1983).
90. J.C. Bezdek, S. Chuah and D. Leep, Generalized k-nearest neighbor rules, *Fuzzy Sets and Systems* **18**, 237–256 (1986).
91. J.M. Keller, M.R. Gray and J.A. Givens, A fuzzy k-nearest neighbor algorithm, *IEEE Trans. Syst. Man Cybern.* **15** (4), 580–585 (1985).
92. M. Béreau and B. Dubuisson, A fuzzy extended k-nearest neighbors rule, *Fuzzy Sets and Systems* **44**, 17–32 (1991).
93. M.S. Yang and C.T. Chen, On strong consistency of the fuzzy generalized nearest neighbor rule, *Fuzzy Sets and Systems* **60** (1993) (to appear).
94. M.S. Yang and C.T. Chen, Convergence rate of the fuzzy generalized nearest neighbor rule, *Computers Math. Applic.* (to appear).
95. J.C. Bezdek, Cluster validity with fuzzy sets, *J. Cybernetics* **3**, 58–73 (1974).
96. R. Dubes and A.K. Jain, Validity studies in clustering methodologies, *Pattern Recognition* **11**, 235–254 (1979).
97. A.K. Jain and J.V. Moreau, Bootstrap technique in cluster analysis, *Pattern Recognition* **20** (5), 547–568 (1987).
98. F.F. Rivera, E.L. Zapata and J.M. Carazo, Cluster validity based on the hard tendency of the fuzzy classification, *Pattern Recognition Letters* **11**, 7–12 (1990).
99. M. Roubens, Fuzzy clustering algorithms and their cluster validity, *Eur. J. Oper. Research* **10**, 294–301 (1982).
100. E. Trauwaert, On the meaning of Dunn's partition coefficient for fuzzy cluster, *Fuzzy Sets and Systems* **25**, 217–242 (1988).
101. M.P. Windham, Cluster validity for the fuzzy c-means clustering algorithm, *IEEE Trans. Pattern Anal. Machine Intell.* **4** (4), 357–363 (1982).
102. M.A. Wong, A bootstrap testing procedure for investigating the number of subpopulations, *J. Statist. Comput. Simul.* **22**, 99–112 (1985).
103. X.L. Xie and G. Beni, A validity measure for fuzzy clustering, *IEEE Trans. Pattern Anal. Machine Intell.* **13**, 841–847 (1991).