Rough Set Based Generalized Fuzzy C-Means Algorithm and Quantitative Indices

Pradipta Maji and Sankar K. Pal, *Fellow, IEEE*

***Abstract*—A generalized hybrid unsupervised learning algorithm, which is termed as rough–fuzzy possibilistic c-means (RFPCM), is proposed in this paper. It comprises a judicious integration of the principles of rough and fuzzy sets. While the concept of lower and upper approximations of rough sets deals with uncertainty, vagueness, and incompleteness in class definition, the membership function of fuzzy sets enables efficient handling of overlapping partitions. It incorporates both probabilistic and possibilistic memberships simultaneously to avoid the problems of noise sensitivity of fuzzy c-means and the coincident clusters of PCM. The concept of crisp lower bound and fuzzy boundary of a class, which is introduced in the RFPCM, enables efficient selection of cluster prototypes. The algorithm is generalized in the sense that all existing variants of c-means algorithms can be derived from the proposed algorithm as a special case. Several quantitative indices are introduced based on rough sets for the evaluation of performance of the proposed c-means algorithm. The effectiveness of the algorithm, along with a comparison with other algorithms, has been demonstrated both qualitatively and quantitatively on a set of real-life data sets.**

***Index Terms*—Clustering, data mining, fuzzy c-means (FCM), pattern recognition, rough sets.**

I. Introduction

CLUSTER analysis is a technique in finding natural groups that are present in data. It divides a given data set into a set of clusters in such a way that two objects from the same cluster are as similar as possible and the objects from different clusters are as dissimilar as possible [1], [2]. Clustering techniques have effectively been applied to pattern recognition, machine learning, biology, medicine, computer vision, communications, remote sensing, etc. A number of clustering algorithms have been proposed to suit different requirements [1], [2].

One of the most widely used prototype-based partitional clustering algorithms is hard c-means (HCM) [1], where each object must be assigned to exactly one cluster. On the other hand, fuzzy c-means (FCM) relaxes this requirement by allowing gradual memberships [3]. In effect, it offers the opportunity to deal with the data that belong to more than one cluster at the same time. It assigns memberships to an object, which are inversely related to the relative distance of the object to cluster prototypes. In addition, it can deal with the uncertainties arising from overlapping cluster boundaries.

Although FCM is a very useful clustering method, the resulting membership values do not always correspond well to the degrees of belonging of the data, and it may be inaccurate in a noisy environment [4]. In real-data analysis, noise and outliers are unavoidable. Hence, to reduce this weakness of the FCM and to produce memberships that have a good explanation of the degrees of belonging for the data, Krishnapuram and Keller [4] proposed a possibilistic approach to clustering which used a possibilistic type of membership function in describing the degree of belonging. However, the possibilistic c-means (PCM) sometimes generates coincident clusters [5]. Recently, the use of both fuzzy (probabilistic) and possibilistic memberships in a clustering algorithm has been proposed in [5] and [6].

Rough-set theory [7] is a new paradigm that is used to deal with uncertainty, vagueness, and incompleteness. It has been applied to fuzzy-rule extraction, reasoning with uncertainty, fuzzy modeling, etc. [7]. In [8], Lingras and West introduced a new clustering method called rough c-means (RCM), which describes a cluster by a prototype (center) and a pair of lower and upper approximations. The lower and upper approximations are different weighted parameters that are used to compute the new centers. Combination of fuzzy and rough sets provides an important direction in reasoning with uncertainty [9]–[11]. Both fuzzy and rough sets provide a mathematical framework to capture uncertainties that are associated with the data [9]. They are complementary in some aspects. Recently, combining both rough and fuzzy sets, Mitra *et al.* [10] proposed a new c-means algorithm, where each cluster is consist of a fuzzy lower approximation and a fuzzy boundary. Each object in lower approximation takes a distinct weight, which is its fuzzy membership value. However, the objects in lower approximation of a cluster should have a similar influence on the corresponding centroid and cluster, and their weights should be independent of other centroids and clusters. Thus, the concept of fuzzy lower approximation, which is introduced in [10], reduces the weights of objects of lower approximation. In effect, it drifts the cluster prototypes from their desired locations. Moreover, it is sensitive to noise and outliers.

In this paper, we propose a generalized hybrid algorithm, which is termed as rough–fuzzy PCM (RFPCM), based on rough and fuzzy sets. While the membership function of the fuzzy sets enables efficient handling of overlapping partitions, the concept of lower and upper approximations of rough sets deals with uncertainty, vagueness, and incompleteness in class definition. The algorithm attempts to exploit the benefits of both probabilistic and possibilistic membership functions.

Manuscript received October 10, 2006; revised March 21, 2007. This paper was recommended by Associate Editor A. Bensaid.

P. Maji is with the Center for Soft Computing Research: A National Facility, Indian Statistical Institute, Kolkata 700 108, India (e-mail: pmaji@isical.ac.in).

S. K. Pal is with the Indian Statistical Institute, Kolkata 700 108, India (e-mail sankar@isical.ac.in).

Digital Object Identifier 10.1109/TSMCB.2007.906578

Integration of probabilistic and possibilistic membership functions avoids the problems of noise sensitivity of the FCM and the coincident clusters of the PCM. Each partition is represented by a set of three parameters, namely, a cluster prototype (centroid), a crisp lower approximation, and a fuzzy boundary. The lower approximation influences the fuzziness of the final partition. The cluster prototype depends on the weighting average of the crisp lower approximation and fuzzy boundary. The algorithm is generalized in a sense that all existing variants of c-means can be derived from the proposed algorithm as a special case. Several quantitative measures are introduced based on rough sets to evaluate the performance of the proposed algorithm. The effectiveness of the proposed algorithm, along with a comparison with the existing c-means algorithms, is demonstrated on a set of benchmark data sets.

The structure of the rest of this paper is as follows. Section II briefly introduces the necessary notions of the FCM and rough sets. In Section III, we describe the RFPCM algorithm based on the theory of rough sets and FCM. Section IV introduces several quantitative performance measures to evaluate the quality of the proposed algorithm. A few case studies and a comparison with other methods are presented in Section V. Concluding remarks are given in Section VI.

II. FCM and Rough Sets

This section presents the basic notions in the theories of FCM and rough sets.

A. Fuzzy C-Means

Let $X = \{x_1, \ldots, x_j, \ldots, x_n\}$ be the set of n objects and $V = \{v_1, \ldots, v_i, \ldots, v_c\}$ be the set of c centroids (means), where $x_j \in \Re^m$, $v_i \in \Re^m$, and $v_i \in X$. The FCM provides a fuzzification of the HCM [3]. It partitions X into c clusters by minimizing the objective function

$$J = \sum_{j=1}^{n}\sum_{i=1}^{c}(\mu_{ij})^{\acute{m}_1}\|x_j - v_i\|^2 \tag{1}$$

where $1 \leq \acute{m}_1 < \infty$ is the fuzzifier, v_i is the ith centroid corresponding to cluster β_i, $\mu_{ij} \in [0,1]$ is the probabilistic membership of the pattern x_j to cluster β_i, and $\|\cdot\|$ is the distance norm, such that

$$v_i = \frac{1}{n_i}\sum_{j=1}^{n}(\mu_{ij})^{\acute{m}_1}x_j, \qquad \text{where } n_i = \sum_{j=1}^{n}(\mu_{ij})^{\acute{m}_1} \tag{2}$$

$$\mu_{ij} = \left(\sum_{k=1}^{c}\left(\frac{d_{ij}}{d_{kj}}\right)^{\frac{2}{\acute{m}_1-1}}\right)^{-1}, \qquad \text{where } d_{ij}^2 = \|x_j - v_i\|^2 \tag{3}$$

subject to $\sum_{i=1}^{c}\mu_{ij} = 1, \forall j$, and $0 < \sum_{j=1}^{n}\mu_{ij} < n, \forall i$.

The process begins by randomly choosing c objects as the centroids of the c clusters. The memberships are calculated based on the relative distance of the object x_j to the centroids $\{v_i\}$ by (3). After computing the memberships of all the objects, the new centroids of the clusters are calculated as per (2). The process stops when the centroids stabilize. That is, the centroids from the previous iteration are identical to those generated in the current iteration.

In the FCM, the memberships of an object are inversely related to the relative distance of the object to the cluster centroids. In effect, it is very sensitive to noise and outliers. In addition, from the standpoint of "compatibility with the centroid," the membership of an object x_j in a cluster β_i should be determined solely by how close it is to the mean (centroid) v_i of the class and should not be coupled with its similarity with respect to other classes.

To alleviate this problem, Krishnapuram and Keller [4] introduced the PCM, where the objective function can be formulated as

$$J = \sum_{i=1}^{c}\sum_{j=1}^{n}(\nu_{ij})^{\acute{m}_2}\|x_j - v_i\|^2 + \sum_{i=1}^{c}\eta_i\sum_{j=1}^{n}(1-\nu_{ij})^{\acute{m}_2} \tag{4}$$

where $1 \leq \acute{m}_2 \leq \infty$ is the fuzzifier, and η_i represents the scale parameter. The membership matrix ν that is generated by the PCM is not a partition matrix in the sense that it does not satisfy the constraint

$$\sum_{i=1}^{c}\nu_{ij} = 1. \tag{5}$$

The update equation of ν_{ij} is given by

$$\nu_{ij} = \frac{1}{1+D}, \qquad \text{where } D = \left\{\frac{\|x_j - v_i\|^2}{\eta_i}\right\}^{1/(\acute{m}_2-1)} \tag{6}$$

subject to $\nu_{ij} \in [0,1], \forall i,j$; $0 < \sum_{j=1}^{n}\nu_{ij} \leq n, \forall i$; and $\max_i \nu_{ij} > 0, \forall j$. The scale parameter η_i represents the zone of influence or size of the cluster β_i. The update equation for η_i is

$$\eta_i = K \cdot \frac{P}{Q} \tag{7}$$

where

$$P = \sum_{j=1}^{n}(\nu_{ij})^{\acute{m}_2}\|x_j - v_i\|^2 \quad Q = \sum_{j=1}^{n}(\nu_{ij})^{\acute{m}_2}.$$

Typically, K is chosen to be one. In each iteration, the updated value of ν_{ij} depends only on the similarity between the object x_j and the centroid v_i. The resulting partition of the data can be interpreted as a possibilistic partition, and the membership values may be interpreted as degrees of possibility of the objects belonging to the classes, i.e., the compatibilities of the objects with the centroids. Updating of the means proceeds exactly the same way as in the case of the FCM.

B. Rough Sets

The theory of rough sets begins with the notion of an approximation space, which is a pair $\langle U, R\rangle$, where U is a nonempty set (the universe of discourse), and R is an equivalence relation on U, i.e., R is reflexive, symmetric, and transitive. The relation

R decomposes the set U into disjoint classes in such a way that two elements x and y are in the same class if and only if (iff) $(x, y) \in R$. Let U/R denote the quotient set of U by the relation R, and $U/R = \{X_1, X_2, \ldots, X_m\}$, where X_i is an equivalence class of R, $i = 1, 2, \ldots, m$. If the two elements x and y in U belong to the same equivalence class $X_i \in U/R$, then we can say that x and y are indistinguishable. The equivalence classes of R and the empty set $\emptyset$ are the elementary sets in the approximation space $\langle U, R\rangle$. Given an arbitrary set $X \in 2^U$, in general, it may not be possible to precisely describe X in $\langle U, R\rangle$. One may characterize X by a pair of lower and upper approximations defined as [7]

$$\underline{R}(X) = \bigcup_{X_i \subseteq X} X_i \quad \text{and} \quad \overline{R}(X) = \bigcup_{X_i \cap X \neq \emptyset} X_i.$$

That is, the lower approximation $\underline{R}(X)$ is the union of all the elementary sets which are subsets of X, and the upper approximation $\overline{R}(X)$ is the union of all the elementary sets which have a nonempty intersection with X. The interval $[\underline{R}(X), \overline{R}(X)]$ is the representation of an ordinary set X in the approximation space $\langle U, R\rangle$ or is simply called the rough set of X. The lower (respectively, upper) approximation $\underline{R}(X)$ [respectively, $\overline{R}(X)$] is interpreted as the collection of those elements of U, which definitely (respectively, possibly) belong to X. Furthermore, a set of X is said to be definable (or exact) in $\langle U, R\rangle$ iff $\underline{R}(X) = \overline{R}(X)$.

In [7], Pawlak discusses two numerical characterizations of imprecision of a subset X in the approximation space $\langle U, R\rangle$: accuracy and roughness. Accuracy of X, which is denoted by $\alpha_R(X)$, is the ratio of the number of objects on its lower approximation to that on its upper approximation, namely

$$\alpha_R(X) = \frac{|\underline{R}(X)|}{|\overline{R}(X)|}.$$

The roughness of X, which is denoted by $\rho_R(X)$, is defined as $\rho_R(X) = 1 - \alpha_R(X)$. Note that the lower the roughness of a subset, the better is its approximation. Furthermore, the following conditions are noted.

1) As $\underline{R}(X) \subseteq X \subseteq \overline{R}(X)$, $0 \leq \rho_R(X) \leq 1$.
2) By convention, when $X = \emptyset$, $\underline{R}(X) = \overline{R}(X) = \emptyset$, and $\rho_R(X) = 0$.
3) $\rho_R(X) = 0$ iff X is definable in $\langle U, R\rangle$.

III. Rough–Fuzzy PCM

Incorporating both fuzzy and rough sets, next, we describe a new c-means, which is termed as RFPCM. The proposed RFPCM adds the concept of fuzzy membership (both probabilistic and possibilistic) of fuzzy sets and the lower and upper approximations of rough sets into c-means algorithm. While the membership of fuzzy sets enables efficient handling of overlapping partitions, the rough sets deal with uncertainty, vagueness, and incompleteness in class definition. Due to the integration of both probabilistic and possibilistic memberships, the RFPCM avoids the problems of noise sensitivity of the FCM and the coincident clusters of the PCM.

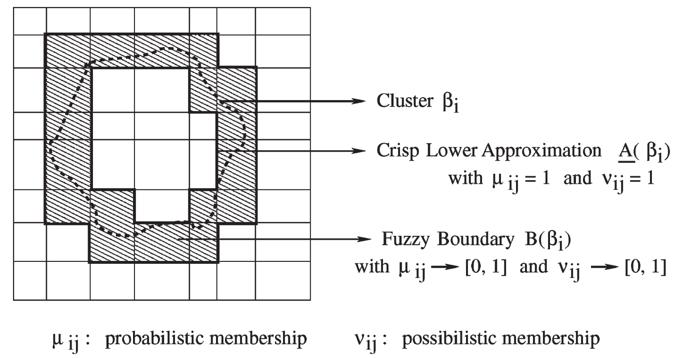


Fig. 1. RFPCM. Cluster β_i is represented by crisp lower bound and fuzzy boundary.

A. Objective Function

Let $\underline{A}(\beta_i)$ and $\overline{A}(\beta_i)$ be the lower and upper approximations of cluster β_i, and let $B(\beta_i) = \{\overline{A}(\beta_i) - \underline{A}(\beta_i)\}$ denote the boundary region of cluster β_i. The proposed RFPCM algorithm partitions a set of n objects into c clusters by minimizing the objective function

$$J_{\mathrm{RFP}} = \begin{cases} w \times \mathcal{A}_1 + \tilde{w} \times \mathcal{B}_1, & \text{if } \underline{A}(\beta_i) \neq \emptyset, B(\beta_i) \neq \emptyset \\ \mathcal{A}_1, & \text{if } \underline{A}(\beta_i) \neq \emptyset, B(\beta_i) = \emptyset \\ \mathcal{B}_1, & \text{if } \underline{A}(\beta_i) = \emptyset, B(\beta_i) \neq \emptyset \end{cases}$$
$$\mathcal{A}_1 = \sum_{i=1}^{c} \sum_{x_j \in \underline{A}(\beta_i)} \left\{a(\mu_{ij})^{\acute{m}_1} + b(\nu_{ij})^{\acute{m}_2}\right\} \|x_j - v_i\|^2 + \sum_{i=1}^{c} \eta_i \sum_{x_j \in \underline{A}(\beta_i)} (1 - \nu_{ij})^{\acute{m}_2}$$
$$\mathcal{B}_1 = \sum_{i=1}^{c} \sum_{x_j \in B(\beta_i)} \left\{a(\mu_{ij})^{\acute{m}_1} + b(\nu_{ij})^{\acute{m}_2}\right\} \|x_j - v_i\|^2 + \sum_{i=1}^{c} \eta_i \sum_{x_j \in B(\beta_i)} (1 - \nu_{ij})^{\acute{m}_2}. \tag{8}$$

The parameters w and $\tilde{w}$ correspond to the relative importance of lower and boundary regions. The constants a and b define the relative importance of probabilistic and possibilistic memberships. Note that μ_{ij} has the same meaning of membership as that in the FCM. Similarly, ν_{ij} has the same interpretation of typicality as in the PCM. Solving (8) with respect to μ_{ij} and ν_{ij}, we get

$$\mu_{ij} = \left(\sum_{k=1}^{c} \left(\frac{d_{ij}}{d_{kj}}\right)^{\frac{2}{\acute{m}_1 - 1}}\right)^{-1}, \qquad \text{where } d_{ij}^2 = \|x_j - v_i\|^2 \tag{9}$$

$$\nu_{ij} = \frac{1}{1 + E}, \qquad \text{where } E = \left\{\frac{b\|x_j - v_i\|^2}{\eta_i}\right\}^{1/(\acute{m}_2 - 1)}. \tag{10}$$

That is, the probabilistic membership μ_{ij} is independent of the constant a, while the constant b has a direct influence on the possibilistic membership ν_{ij}. The scale parameter η_i has the same expression as that in (7).

In the RFPCM, each cluster is represented by a centroid, a crisp lower approximation, and a fuzzy boundary (Fig. 1).

The lower approximation influences the fuzziness of final partition. According to the definitions of lower approximation and boundary region of rough sets, if an object $x_j \in \underline{A}(\beta_i)$, then $x_j \notin \underline{A}(\beta_k), \forall k \neq i$, and $x_j \notin B(\beta_i), \forall i$. That is, the object x_j is definitely contained in β_i. Thus, the weights of the objects in lower approximation of a cluster should be independent of other centroids and clusters and should not be coupled with their similarity with respect to other centroids. In addition, the objects in lower approximation of a cluster should have a similar influence on the corresponding centroid and cluster. Whereas, if $x_j \in B(\beta_i)$, then the object x_j possibly belongs to β_i and potentially belongs to another cluster. Hence, the objects in boundary regions should have a different influence on the centroids and clusters. Therefore, in the RFPCM, the membership values of objects in lower approximation are $\mu_{ij} = \nu_{ij} = 1$, while those in the boundary region are the same as in the FCM [(9)] and PCM [(10)]. In other words, the proposed c-means first partitions the data into two classes—lower approximation and boundary. Only the objects in the boundary are fuzzified. Thus, $\mathcal{A}_1$ reduces to

$$\mathcal{A}_1 = \sum_{i=1}^{c} \sum_{x_j \in \underline{A}(\beta_i)} \|x_j - v_i\|^2$$

and $\mathcal{B}_1$ has the same expression as that in (8).

B. Cluster Prototypes

The new centroid is calculated based on the weighting average of the crisp lower approximation and fuzzy boundary. Computation of the centroid is modified to include the effects of both fuzzy memberships (probabilistic and possibilistic) and lower and upper bounds. The modified centroid calculation for the RFPCM is obtained by solving (8) with respect to v_i

$$v_i^{\mathrm{RFP}} = \begin{cases} w \times \mathcal{C}_1 + \tilde{w} \times \mathcal{D}_1, & \text{if } \underline{A}(\beta_i) \neq \emptyset, B(\beta_i) \neq \emptyset \\ \mathcal{C}_1, & \text{if } \underline{A}(\beta_i) \neq \emptyset, B(\beta_i) = \emptyset \\ \mathcal{D}_1, & \text{if } \underline{A}(\beta_i) = \emptyset, B(\beta_i) \neq \emptyset \end{cases}$$

$$\mathcal{C}_1 = \frac{1}{|\underline{A}(\beta_i)|} \sum_{x_j \in \underline{A}(\beta_i)} x_j \quad (11)$$

where $|\underline{A}(\beta_i)|$ represents the cardinality of $\underline{A}(\beta_i)$ and

$$\mathcal{D}_1 = \frac{1}{n_i} \sum_{x_j \in B(\beta_i)} \left\{ a(\mu_{ij})^{\acute{m}_1} + b(\nu_{ij})^{\acute{m}_2} \right\} x_j$$

where $n_i = \sum_{x_j \in B(\beta_i)} \{a(\mu_{ij})^{\acute{m}_1} + b(\nu_{ij})^{\acute{m}_2}\}$.

Thus, the cluster prototypes (centroids) depend on parameters w and $\tilde{w}$ and constants a and b, and fuzzifiers $\acute{m}_1$ and $\acute{m}_2$ rule their relative influence. This shows that if b is higher than a, the centroids will be more influenced by the possibilistic memberships than the probabilistic memberships. Thus, to reduce the influence of noise and outliers, a bigger value for b than a should be used. The correlated influence of these parameters, constants, and fuzzifiers makes it somewhat difficult to determine their optimal values. The parameter w has an influence on the performance of the RFPCM. Since the objects lying in lower approximation definitely belong to a cluster, they are assigned a higher weight w compared with $\tilde{w}$ of the objects lying in boundary regions. On the other hand, the performance of the proposed c-means significantly reduces when $w \simeq 1.00$. In this case, since the clusters cannot see the objects of the boundary regions, the mobility of the clusters and the centroids reduces. As a result, some centroids get stuck in local optimum. Hence, for the clusters and the centroids to have a greater degree of freedom to move, $0 < \tilde{w} < w < 1$ subject to $w + \tilde{w} = 1$. In addition, $0 < a < 1$, $0 < b < 1$, and $a + b = 1$.

Some limiting properties of the RFPCM algorithm related to the probabilistic and possibilistic memberships are stated next.

$$\lim_{\acute{m}_1 \to 1^+}\{\mu_{ij}\} = \begin{cases} 1, & \text{if } d_{ij} < d_{kj} \quad \forall k \neq I \\ 0, & \text{otherwise} \end{cases} \quad (12)$$

$$\lim_{\acute{m}_2 \to 1^+}\{\nu_{ij}\} = \begin{cases} 1, & \text{if } bd_{ij} < \eta_i \\ \frac{1}{2}, & \text{if } bd_{ij} = \eta_i \\ 0, & \text{if } bd_{ij} > \eta_i \end{cases} \quad (13)$$

$$\lim_{\acute{m}_1 \to \infty}\{\mu_{ij}\} = \frac{1}{c}, \quad \lim_{\acute{m}_2 \to \infty}\{\nu_{ij}\} = \frac{1}{2} \quad (14)$$

$$\lim_{\acute{m}_1, \acute{m}_2 \to \infty}\{v_i\} = \sum_{j=1}^{n} x_j = \bar{v}, \qquad 1 \leq i \leq c \quad (15)$$

$$\nu_{ij} = \begin{cases} 1, & \text{if } b = 0 \\ \frac{1}{2}, & \text{if } bd_{ij} = \eta_{i}. \end{cases} \quad (16)$$

C. Convergence Condition

In this section, we present a mathematical analysis on the convergence property of the proposed c-means algorithm. According to (11), the cluster prototype of the proposed c-means algorithm is calculated based on the weighting average of the crisp lower approximation and fuzzy boundary when both $\underline{A}(\beta_i) \neq \emptyset$ and $B(\beta_i) \neq \emptyset$, i.e.,

$$v_i^{\mathrm{RFP}} = w \times \mathrm{v}_i^{\mathrm{RFP}} + \tilde{w} \times \tilde{\mathrm{v}}_i^{\mathrm{RFP}} \quad (17)$$

where

$$\mathrm{v}_i^{\mathrm{RFP}} = \mathcal{C}_1 = \frac{1}{|\underline{A}(\beta_i)|} \sum_{x_j \in \underline{A}(\beta_i)} x_j \quad (18)$$

$$\tilde{\mathrm{v}}_i^{\mathrm{RFP}} = \mathcal{D}_1 = \frac{1}{n_i} \sum_{x_j \in B(\beta_i)} \left\{ a(\mu_{ij})^{\acute{m}_1} + b(\nu_{ij})^{\acute{m}_2} \right\} x_j$$

where

$$n_i = \sum_{x_j \in B(\beta_i)} \left\{ a(\mu_{ij})^{\acute{m}_1} + b(\nu_{ij})^{\acute{m}_2} \right\}. \quad (19)$$

In the RFPCM, an object may not belong to both lower approximation and boundary region of a cluster. Thus, the convergence of v_i^{RFP} depends on the convergence of $\mathrm{v}_i^{\mathrm{RFP}}$ and $\tilde{\mathrm{v}}_i^{\mathrm{RFP}}$.

Both (18) and (19) can be rewritten as

$$(|\underline{A}(\beta_i)|)\, \mathrm{v}_i^{\mathrm{RFP}} = \sum_{x_j \in \underline{A}(\beta_i)} x_j \quad (20)$$

$$(n_i)\tilde{\mathrm{v}}_i^{\mathrm{RFP}} = \sum_{x_j \in B(\beta_i)} \left\{ a(\mu_{ij})^{\acute{m}_1} + b(\nu_{ij})^{\acute{m}_2} \right\} x_j. \quad (21)$$

Thus, (20) and (21) represent a set of linear equations in terms of $\mathrm{v}_i^{\mathrm{RFP}}$ and $\tilde{\mathrm{v}}_i^{\mathrm{RFP}}$ if both μ_{ij} and ν_{ij} are kept constant. A simple way to analyze the convergence property of the algorithm is to view both (18) and (19) as the Gauss–Seidel iterations for solving the set of linear equations. The Gauss–Seidel algorithm is guaranteed to converge if the matrix representing each equation is diagonally dominant [12]. This is a sufficient condition, not a necessary one. The iteration may or may not converge if the matrix is not diagonally dominant [12]. The matrix corresponding to (18) is given by

$$A = \begin{bmatrix} |\underline{A}(\beta_1)| & 0 & \cdots & \cdots & 0 \\ 0 & |\underline{A}(\beta_2)| & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \cdots & |\underline{A}(\beta_c)| \end{bmatrix} \quad (22)$$

where $|\underline{A}(\beta_i)|$ represents the cardinality of $\underline{A}(\beta_i)$. Similarly, the matrix corresponding to (19) is given by

$$\tilde{\mathrm{A}} = \begin{bmatrix} n_1 & 0 & \cdots & \cdots & 0 \\ 0 & n_2 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \cdots & n_c \end{bmatrix} \quad (23)$$

where $n_i = \sum_{x_j \in B(\beta_i)} \{a(\mu_{ij})^{\acute{m}_1} + b(\nu_{ij})^{\acute{m}_2}\}$ which represents the cardinality of $B(\beta_i)$. For both A and $\tilde{A}$ to be diagonally dominant, we must have

$$|\underline{A}(\beta_i)| > 0 \quad n_i > 0. \quad (24)$$

This is the sufficient condition for matrices A and $\tilde{A}$ to be diagonally dominant. Under this condition, the iteration would converge if (18) and (19) were repetitively applied with μ_{ij} and ν_{ij} which were kept constant. In practice, (9)–(11) are alternatively applied in the iterations. The condition in (24) is still correct according to the convergence theorem of the FCM of Bezdek *et al.* [13] and Yan's convergence analysis of the fuzzy curve-tracing algorithm [14]. Both the matrices A and $\tilde{A}$ are also the Hessian (second-order derivative) of $\mathcal{A}_1$ and $\mathcal{B}_1$ (8) with respect to $\mathrm{v}_i^{\mathrm{RFP}}$ and $\tilde{\mathrm{v}}_i^{\mathrm{RFP}}$, respectively. As both A and $\tilde{A}$ are diagonally dominant, all their eigenvalues are positive. In addition, the Hessian of $\mathcal{B}_1$ with respect to both μ_{ij} and ν_{ij} can easily be shown to be a diagonal matrix and is positive definite. Thus, according to the theorem derived by Bezdek *et al.* [13] and the analysis done by Yan [14], it can be concluded that the proposed algorithm converges, at least along a subsequence, to a local optimum solution as long as the condition in (24) is satisfied. Intuitively, the objective function J_{RFP} [(8)] reduces in all steps corresponding to (9)–(11); therefore, the compound procedure strictly makes the function J_{RFP} descent.

D. Details of the Algorithm

Approximate optimization of J_{RFP} [(8)] by the RFPCM is based on Picard iteration through (9)–(11). The process starts by randomly choosing c objects as the centroids of the c clusters. The probabilistic and possibilistic memberships of all the objects are calculated using (9) and (10). The scale parameters η_i for c clusters are obtained using (7). Let $u_i = (u_{i1}, \ldots, u_{ij}, \ldots, u_{in})$ represent the fuzzy cluster β_i associated with the centroid v_i, and $u_{ij} = \{a\mu_{ij} + b\nu_{ij}\}$. After computing u_{ij} for c clusters and n objects, the values of u_{ij} for each object x_j are sorted, and the difference of two highest memberships of x_j is compared with a threshold value δ. Let u_{ij} and u_{kj} be the highest and second highest memberships of x_j, respectively. If $(u_{ij} - u_{kj}) > \delta$, then $x_j \in \underline{A}(\beta_i)$, as well as $x_j \in \overline{A}(\beta_i)$; otherwise, $x_j \in \overline{A}(\beta_i)$, and $x_j \in \overline{A}(\beta_k)$. After assigning each object in lower approximations or boundary regions of different clusters based on δ, both memberships μ_{ij} and ν_{ij} of the objects are modified. The values of μ_{ij} and ν_{ij} are set to one for the objects in lower approximations, while those in boundary regions are held unchanged. The new centroids of the clusters are calculated as per (11). Hence, the performance of the RFPCM depends on the value of δ, which determines the class labels of all the objects. In other words, the RFPCM partitions the data set into two classes—lower approximation and boundary, which are based on the value of δ. The δ represents the size of granules of rough–fuzzy clustering. In practice, we find that the following definition works well:

$$\delta = \frac{1}{n} \sum_{j=1}^{n} (u_{ij} - u_{kj}). \quad (25)$$

δ represents the average difference of two highest memberships of all the objects in the data set. A good clustering procedure should make the value of δ as high as possible.

E. Generalization of Existing C-Means Algorithms

Here, we describe two derivatives of the RFPCM, namely, rough FCM (RFCM) and rough PCM (RPCM), and prove that the proposed c-means is the generalization of the existing c-means algorithms.

Rough FCM (RFCM): Let $0 < \delta < 1$ and $a = 1$, then (8) and (11) become

$$J_{\mathrm{RF}} = \begin{cases} w \times \mathcal{A}_2 + \tilde{w} \times \mathcal{B}_2, & \text{if } \underline{A}(\beta_i) \neq \emptyset,\ B(\beta_i) \neq \emptyset \\ \mathcal{A}_2, & \text{if } \underline{A}(\beta_i) \neq \emptyset,\ B(\beta_i) = \emptyset \\ \mathcal{B}_2, & \text{if } \underline{A}(\beta_i) = \emptyset,\ B(\beta_i) \neq \emptyset \end{cases}$$

$$\mathcal{A}_2 = \sum_{i=1}^{c} \sum_{x_j \in \underline{A}(\beta_i)} \|x_j - v_i\|^2$$

$$\mathcal{B}_2 = \sum_{i=1}^{c} \sum_{x_j \in B(\beta_i)} (\mu_{ij})^{\acute{m}_1 \|x_j - v_i\|^2} \quad (26)$$

$$v_i^{\mathrm{RF}} = \begin{cases} w \times \mathcal{C}_2 + \tilde{w} \times \mathcal{D}_2, & \text{if } \underline{A}(\beta_i) \neq \emptyset,\ B(\beta_i) \neq \emptyset \\ \mathcal{C}_2, & \text{if } \underline{A}(\beta_i) \neq \emptyset,\ B(\beta_i) = \emptyset \\ \mathcal{D}_2, & \text{if } \underline{A}(\beta_i) = \emptyset,\ B(\beta_i) \neq \emptyset \end{cases}$$

$$\mathcal{C}_2 = \frac{1}{|\underline{A}(\beta_i)|} \sum_{x_j \in \underline{A}(\beta_i)} x_j$$

$$\mathcal{D}_2 = \frac{1}{n_i} \sum_{x_j \in B(\beta_i)} (\mu_{ij})^{\acute{m}_1} x_j \qquad n_i = \sum_{x_j \in B(\beta_i)} (\mu_{ij})^{\acute{m}_1}. \quad (27)$$

That is, if $0 < \delta < 1$ and $a = 1$, the RFPCM boils down to the RFCM, where each cluster is consist of a crisp lower bound and a fuzzy boundary with probabilistic memberships. In [10], a preliminary version of the RFCM has been proposed, where each cluster is consist of a fuzzy lower approximation and a fuzzy boundary. If an object $x_j \in \underline{A}(\beta_i)$, then $\mu_{kj} = \mu_{ij}$ if $k = i$, and $\mu_{kj} = 0$, otherwise. That is, each object $x_j \in \underline{A}(\beta_i)$ takes a distinct weight, which is its fuzzy (probabilistic) membership value. Thus, the weight of the object in lower approximation is inversely related to the relative distance of the object to all cluster prototypes. In fact, the objects in lower approximation of a cluster should have a similar influence on the corresponding centroid and cluster. In addition, their weights should be independent of other centroids and clusters and should not be coupled with their similarity with respect to other clusters. Thus, the concept of fuzzy lower approximation of the RFCM, which is introduced in [10], reduces the weights of objects of lower approximation and effectively drifts the cluster centroids from their desired locations.

Rough PCM (RPCM): Let $0 < \delta < 1$ and $a = 0$, then (8) becomes

$$J_{\mathrm{RP}} = \begin{cases} w \times \mathcal{A}_3 + \tilde{w} \times \mathcal{B}_3, & \text{if } \underline{A}(\beta_i) \neq \emptyset,\ B(\beta_i) \neq \emptyset \\ \mathcal{A}_3, & \text{if } \underline{A}(\beta_i) \neq \emptyset,\ B(\beta_i) = \emptyset \\ \mathcal{B}_3, & \text{if } \underline{A}(\beta_i) = \emptyset,\ B(\beta_i) \neq \emptyset \end{cases}$$

$$\mathcal{A}_3 = \sum_{i=1}^{c} \sum_{x_j \in \underline{A}(\beta_i)} \|x_j - v_i\|^2$$

$$\mathcal{B}_3 = \sum_{i=1}^{c} \sum_{x_j \in B(\beta_i)} (\nu_{ij})^{\acute{m}_2} \|x_j - v_i\|^2 + \sum_{i=1}^{c} \eta_i \sum_{x_j \in B(\beta_i)} (1 - \nu_{ij})^{\acute{m}_2}. \tag{28}$$

Similarly, (11) reduces to

$$v_i^{\mathrm{RP}} = \begin{cases} w \times \mathcal{C}_3 + \tilde{w} \times \mathcal{D}_3, & \text{if } \underline{A}(\beta_i) \neq \emptyset,\ B(\beta_i) \neq \emptyset \\ \mathcal{C}_3, & \text{if } \underline{A}(\beta_i) \neq \emptyset,\ B(\beta_i) = \emptyset \\ \mathcal{D}_3, & \text{if } \underline{A}(\beta_i) = \emptyset,\ B(\beta_i) \neq \emptyset \end{cases}$$

$$\mathcal{C}_3 = \frac{1}{|\underline{A}(\beta_i)|} \sum_{x_j \in \underline{A}(\beta_i)} x_j$$

$$\mathcal{D}_3 = \frac{1}{n_i} \sum_{x_j \in B(\beta_i)} (\nu_{ij})^{\acute{m}_2} x_j, \qquad n_i = \sum_{x_j \in B(\beta_i)} (\nu_{ij})^{\acute{m}_2}. \tag{29}$$

Therefore, if $0 < \delta < 1$ and $a = 0$, the RFPCM reduces to the RPCM, where each cluster is represented by a crisp lower bound and a fuzzy boundary with possibilistic memberships.

Rough C-Means (RCM): If $0 < \delta < 1$, and for any nonzero μ_{ij} and ν_{ij}, if we set $\mu_{ij} = \nu_{ij} = 1$, $\forall i, j$, then (8) and (11) become

$$J_{\mathrm{R}} = \begin{cases} w \times \mathcal{A}_4 + \tilde{w} \times \mathcal{B}_4, & \text{if } \underline{A}(\beta_i) \neq \emptyset,\ B(\beta_i) \neq \emptyset \\ \mathcal{A}_4, & \text{if } \underline{A}(\beta_i) \neq \emptyset,\ B(\beta_i) = \emptyset \\ \mathcal{B}_4, & \text{if } \underline{A}(\beta_i) = \emptyset,\ B(\beta_i) \neq \emptyset \end{cases}$$

$$\mathcal{A}_4 = \sum_{i=1}^{c} \sum_{x_j \in \underline{A}(\beta_i)} \|x_j - v_i\|^2 \quad \mathcal{B}_4 = \sum_{i=1}^{c} \sum_{x_j \in B(\beta_i)} \|x_j - v_i\|^2 \tag{30}$$

$$v_i^{\mathrm{R}} = \begin{cases} w \times \mathcal{C}_4 + \tilde{w} \times \mathcal{D}_4, & \text{if } \underline{A}(\beta_i) \neq \emptyset,\ B(\beta_i) \neq \emptyset \\ \mathcal{C}_4, & \text{if } \underline{A}(\beta_i) \neq \emptyset,\ B(\beta_i) = \emptyset \\ \mathcal{D}_4, & \text{if } \underline{A}(\beta_i) = \emptyset,\ B(\beta_i) \neq \emptyset \end{cases}$$

$$\mathcal{C}_4 = \frac{1}{|\underline{A}(\beta_i)|} \sum_{x_j \in \underline{A}(\beta_i)} x_j \quad \mathcal{D}_4 = \frac{1}{|B(\beta_i)|} \sum_{x_j \in B(\beta_i)} x_j. \tag{31}$$

This is equivalent to RCM of Lingras and West [8]. In the case of RCM, both lower bound and boundary are crisp. Thus, the difference of RCM with RFPCM/RFCM/RPCM is that, while in RFPCM/RFCM/RPCM, each object in the boundary region takes a distinct weight; in RCM, the uniform weight is imposed on all of the objects in the boundary region. In fact, the objects in boundary regions should have a different influence on the centroids (means) and clusters.

Fuzzy PCM (FPCM): If we set $\delta = 1$, then $\underline{A}(\beta_i) = \emptyset$ and $\overline{A}(\beta_i) = B(\beta_i)$. Hence, for $0 < a < 1$, (8) and (11) reduce to

$$J_{\mathrm{FP}} = \sum_{i=1}^{c} \sum_{j=1}^{n} \left\{ a(\mu_{ij})^{\acute{m}_1} + b(\nu_{ij})^{\acute{m}_2} \right\} \|x_j - v_i\|^2 + \sum_{i=1}^{c} \eta_i \sum_{j=1}^{n} (1 - \nu_{ij})^{\acute{m}_2} \tag{32}$$

$$v_i^{\mathrm{FP}} = \frac{1}{n_i} \sum_{j=1}^{n} \left\{ a(\mu_{ij})^{\acute{m}_1} + b(\nu_{ij})^{\acute{m}_2} \right\} x_j \tag{33}$$

where $n_i = \sum_{j=1}^{n} \{a(\mu_{ij})^{\acute{m}_1} + b(\nu_{ij})^{\acute{m}_2}\}$.

Therefore, for $\delta = 1$ and $0 < a < 1$, the RFPCM boils down to the FPCM algorithm of Pal *et al.* [5].

Fuzzy C-Means (FCM): If $\delta = 1$ and $a = 1$, then

$$J_{\mathrm{F}} = \sum_{i=1}^{c} \sum_{j=1}^{n} (\mu_{ij})^{\acute{m}_1} \|x_j - v_i\|^2 \tag{34}$$

$$v_i^{\mathrm{F}} = \frac{1}{n_i} \sum_{j=1}^{n} (\mu_{ij})^{\acute{m}_1} x_j, \qquad \text{where} \quad n_i = \sum_{j=1}^{n} (\mu_{ij})^{\acute{m}_1} \tag{35}$$

which is equivalent to the FCM algorithm proposed in [3].

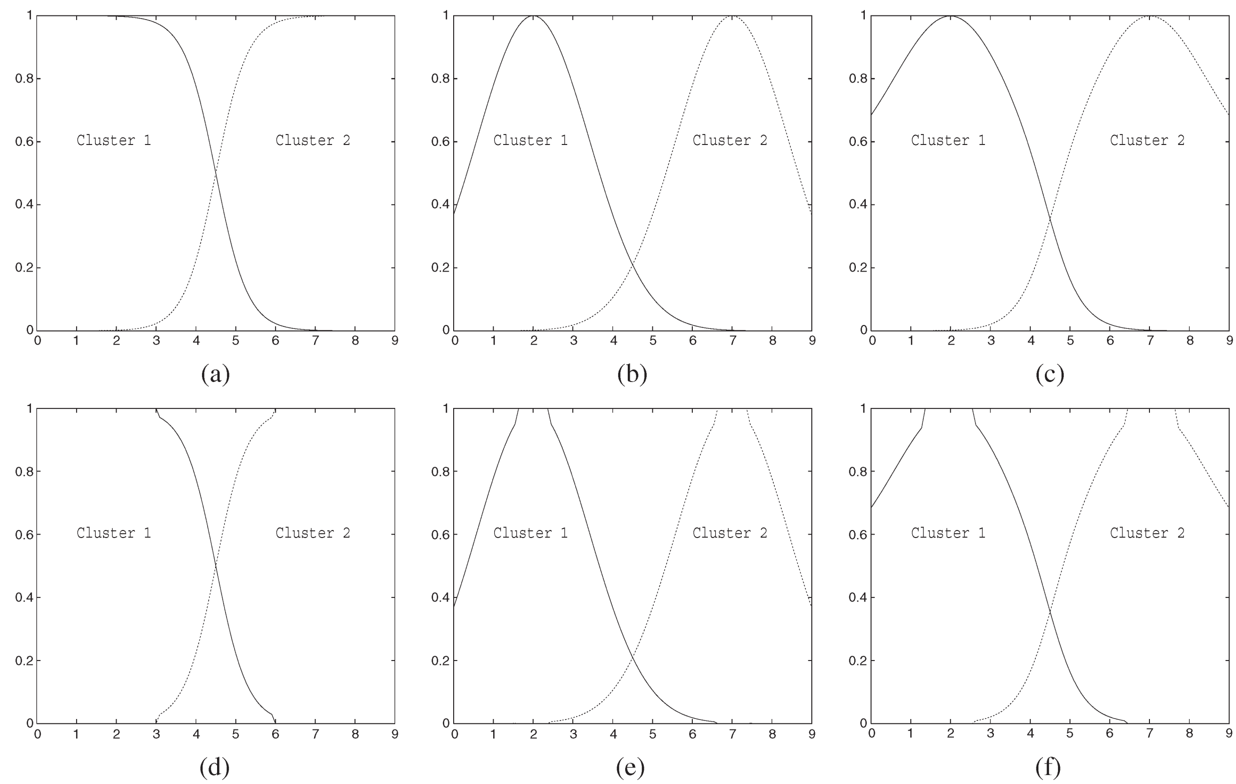


Fig. 2. Plot of memberships for $a = 0.0, 0.5,$ and 1.0, and for $\delta = 0.95$ and 1.00. (a) FCM: $a = 1.00$, $\delta = 1.00$. (b) PCM: $a = 0.00$, $\delta = 1.00$. (c) FPCM: $a = 0.50$, $\delta = 1.00$. (d) RFCM: $a = 1.00$, $\delta = 0.95$. (e) RPCM: $a = 0.00$, $\delta = 0.95$. (f) RFPCM: $a = 0.50$, $\delta = 0.95$.

Possibilistic C-Means (PCM): If $\delta = 1$ and $a = 0$, then

$$J_{\mathrm{P}} = \sum_{i=1}^{c}\sum_{j=1}^{n}(\nu_{ij})^{\acute{m}_2}\|x_j - v_i\|^2 + \sum_{i=1}^{c}\eta_i\sum_{j=1}^{n}(1-\nu_{ij})^{\acute{m}_2} \quad (36)$$

$$v_i^{\mathrm{P}} = \frac{1}{n_i}\sum_{j=1}^{n}(\nu_{ij})^{\acute{m}_2}x_j, \qquad \text{where} \quad n_i = \sum_{j=1}^{n}(\nu_{ij})^{\acute{m}_2}. \quad (37)$$

Thus, for $\delta = 1$ and $a = 0$, the RFPCM reduces to the PCM algorithm introduced in [4].

Hard C-Means (HCM): If $\delta = 0$, then $B(\beta_i) = \emptyset$ and $\overline{A}(\beta_i) = \underline{A}(\beta_i) = \beta_i$. In effect, (8) and (11) become

$$J_{\mathrm{H}} = \sum_{i=1}^{c}\sum_{x_j\in\beta_i}\|x_j - v_i\|^2 \qquad v_i^{\mathrm{H}} = \frac{1}{n_i}\sum_{x_j\in\beta_i}x_j. \quad (38)$$

Hence, for $\delta = 0$, the RFPCM boils down to the HCM.

Fig. 2 shows the membership values of all the objects to two Gaussian-distributed clusters for the values $a = 0.0$, 0.5, and 1.0, and $\delta = 0.95$ and 1.0. The different shape of the memberships, particularly for objects near to the centers and for objects far from the separating line, is apparent. Detailed analysis reported in this section confirms that the proposed RFPCM is the generalization of the existing c-means algorithms. It effectively integrates HCM, FCM, and PCM using the concept of lower and upper approximations of rough sets.

IV. Quantitative Measures

In this section, we present some quantitative indices to evaluate the performance of rough–fuzzy clustering algorithm, incorporating the concepts of rough sets [7].

α Index: It is given by

$$\alpha = \frac{1}{c}\sum_{i=1}^{c}\frac{wA_i}{wA_i + \tilde{w}B_i} \quad (39)$$

where

$$A_i = \sum_{x_j\in\underline{A}(\beta_i)}\left\{a(\mu_{ij})^{\acute{m}_1} + b(\nu_{ij})^{\acute{m}_2}\right\} = |\underline{A}(\beta_i)| \quad (40)$$

$$B_i = \sum_{x_j\in B(\beta_i)}\left\{a(\mu_{ij})^{\acute{m}_1} + b(\nu_{ij})^{\acute{m}_2}\right\}. \quad (41)$$

μ_{ij} and ν_{ij} represent the probabilistic and possibilistic memberships of object x_j in cluster β_i, respectively. Constants a and b define the relative importance of the probabilistic and possibilistic memberships, while parameters w and $\tilde{w}$ correspond to the relative importance of lower and boundary regions.

The α index represents the average accuracy of c clusters. It is the average of the ratio of the number of objects in lower approximation to that in upper approximation of each cluster. In effect, it captures the average degree of completeness of knowledge about all clusters. A good clustering procedure should make all objects as similar to their centroids as possible. The α index increases with an increase in similarity within a cluster. Therefore, for the given data set and c value, the higher the similarity values within the clusters, the higher would be the

α value. The value of α also increases with c. In an extreme case when the number of clusters is maximum, i.e., $c = n$, which is the total number of objects in the data set, the value of $\alpha = 1$. When $\overline{A}(\beta_i) = \underline{A}(\beta_i), \forall i$, i.e., all the clusters $\{\beta_i\}$ are exact or definable, then we have $\alpha = 1$. Whereas, if $\overline{A}(\beta_i) = B(\beta_i), \forall i$, the value of $\alpha = 0$. Thus, $0 \leq \alpha \leq 1$.

ϱ Index: The ϱ index represents the average roughness of c clusters and is defined as follows:

$$\varrho = 1 - \alpha = 1 - \frac{1}{c}\sum_{i=1}^{c} \frac{wA_i}{wA_i + \tilde{w}B_i}. \tag{42}$$

Note that the lower the value of ϱ, the better is the overall cluster approximations. In addition, $0 \leq \varrho \leq 1$. Basically, ϱ index represents the average degree of incompleteness of knowledge about all clusters.

$\alpha^\star$ Index: It can be defined as

$$\alpha^\star = \frac{C}{D} \qquad C = \sum_{i=1}^{c} wA_i \qquad D = \sum_{i=1}^{c}\{wA_i + \tilde{w}B_i\}. \tag{43}$$

The $\alpha^\star$ index represents the accuracy of approximation of all clusters. It captures the exactness of approximate clustering. A good clustering procedure should make the value of $\alpha^\star$ as high as possible. The $\alpha^\star$ index maximizes the exactness of approximate clustering.

γ Index: It is the ratio of the total number of objects in the lower approximations of all clusters to the cardinality of the universe of discourse U and is given by

$$\gamma = \frac{R}{S}, \qquad \text{where } R = \sum_{i=1}^{c} |\underline{A}(\beta_i)| \qquad S = |U| = n. \tag{44}$$

The γ index basically represents the quality of approximation of a clustering algorithm.

V. Performance Analysis

The performance of our three hybrid algorithms, namely, RFCM, RPCM, and RFPCM, is compared extensively with that of different c-means algorithms. The algorithms that are compared are HCM [1], FCM [3], PCM [4], FPCM [5], FPCM of Masulli and Rovetta (FPCM^{MR}) [6], kernel-based HCM (KHCM) [15], kernel-based FCM (KFCM) [16], [17], kernel-based PCM (KPCM) [17], kernel-based FPCM (KFPCM) [18], RCM [8], and RFCM of Mitra *et al.* (RFCM^{MBP}) [10]. All the algorithms are implemented in C language and run in LINUX environment that has a machine configuration of Pentium IV, 3.2 GHz, 1-MB cache, and 1-GB RAM.

To analyze the performance of the proposed algorithms, the experimentation has been done in two parts. In the first part, we have used some benchmark data sets. In the second part, we present the results on segmentation of brain MR images. The major metrics for the evaluation of the performance of different algorithms are the indices proposed in Section IV such as α, ϱ, $\alpha^\star$, and γ, as well as some existing measures like the Davies–Bouldin (DB) and Dunn (D) indexes [19]. The values of δ for RCM, RFCM, RPCM, and RFPCM algorithms have been calculated using (25). The final prototypes of FCM are used to initialize PCM, FPCM, KFCM, KPCM, and KFPCM, while the Gaussian function is used as the kernel.

TABLE I
Performance Analysis on Iris Data Set

Algorithms /Methods	Prototypes $c = 2$			Prototypes $c = 3$		
	DB	Dunn	Time	DB	Dunn	Time
FCM	0.11	12.36	13	0.32	4.32	74
PCM	0.12	9.99	115	-	-	-
FPCM	0.11	11.87	63	0.48	2.63	68
FPCM^{MR}	0.16	9.07	4	0.35	4.36	20
KHCM	0.12	12.08	6	0.32	4.27	18
KFCM	0.11	12.36	22	0.29	4.89	85
KPCM	0.12	9.99	138	-	-	-
KFPCM	0.11	11.44	81	0.46	3.22	99
RCM	0.10	13.49	3	0.22	6.75	16
RFCM^{MBP}	0.11	13.42	17	0.22	6.91	33
RFCM	0.10	13.49	12	0.22	6.94	24
RPCM	0.10	11.67	37	0.22	7.13	59
RFPCM	0.10	13.61	25	0.21	7.75	24

TABLE II
Quantitative Evaluation of Rough–Fuzzy Clustering

Value	Methods	α Index	ϱ Index	$\alpha^\star$ Index	γ Index
$c = 2$	RFCM^{MBP}	0.999991	0.000009	0.999989	0.812500
	RFCM	0.999994	0.000006	0.999994	0.906667
	RPCM	0.999980	0.000020	0.999979	0.506667
	RFPCM	0.999986	0.000014	0.999985	0.686667
$c = 3$	RFCM^{MBP}	0.999971	0.000029	0.999963	0.625000
	RFCM	0.999986	0.000014	0.999988	0.800000
	RPCM	0.999983	0.000017	0.999985	0.553333
	RFPCM	0.999987	0.000013	0.999989	0.766667

A. Benchmark Data Sets

This section demonstrates the performance of different c-means algorithms on some benchmark data sets like Iris, Glass, Ionosphere, Wine, and Wisconsin data sets. All the data sets are downloaded from http://www.ics.uci.edu/~mlearn.

The performance of different c-means algorithms on Iris data set is reported next for $c = 2$ and 3. Several runs have been made with different initializations and different choices of parameters. For the RFPCM, the values of δ are 0.51 and 0.48 for $c = 2$ and 3, respectively, considering $w = 0.99$, $\acute{m}_1 = \acute{m}_2 = 2.0$, and $a = b = 0.5$. The parameters are held constant across all runs.

For $c = 3$, all the c-means algorithms, except PCM and KPCM, generate good prototypes. The final prototypes of FCM are used to initialize PCM, FPCM, KFCM, KPCM, and KFPCM. Even if three initial centroids belong to three different classes, PCM and KPCM generate coincident clusters. That is, two of the three final prototypes are identical in case of PCM and KPCM. Tables I and II depict the best results obtained using different c-means algorithms for $c = 2$ and 3. In Table I, the performance of different algorithms is reported with respect to DB index, D index, and CPU time (in milliseconds). The results reported in Table I establish the fact that, although each c-means algorithm generates good prototypes with lower values of the DB index and higher values of the D index for $c = 2$, the RFPCM provides the best result that has the lowest DB index and highest D index. The results of other versions of rough clustering are quite similar to that of the RFPCM.

TABLE III
PERFORMANCE OF DIFFERENT C-MEANS ALGORITHMS ON GLASS, IONOSPHERE, WINE, AND WISCONSIN BREAST-CANCER DATA SET

Algorithms /Methods	Glass Data Set			Ionosphere Data Set			Wine Data Set			Wisconsin Data Set		
	DB	Dunn	Time	DB	Dunn	Time	DB	Dunn	Time	DB	Dunn	Time
HCM	3.28	0.02	18	2.36	0.58	17	0.20	2.52	11	0.19	5.53	30
FCM	2.52	0.05	333	1.51	0.89	153	0.19	2.78	96	0.17	6.13	298
FPCM	-	-	-	-	-	-	0.45	1.40	204	-	-	-
FPCM^{MR}	2.48	0.11	29	1.52	0.89	24	0.88	1.78	17	0.13	8.29	13
KHCM	2.60	0.03	65	2.30	0.59	74	0.20	2.52	23	0.19	5.42	140
KFCM	2.51	0.05	132	1.51	0.89	360	0.19	2.76	196	0.18	5.84	723
KFPCM	-	-	-	-	-	-	0.48	1.54	459	-	-	-
RCM	1.96	0.37	53	1.01	1.57	27	0.20	3.08	15	0.16	7.85	17
RFCM^{MBP}	1.51	0.13	191	1.09	1.50	98	0.19	3.19	54	0.16	7.84	77
RFCM	1.45	0.49	124	0.99	1.60	96	0.19	3.13	44	0.12	8.82	62
RPCM	1.91	0.31	247	1.09	1.58	297	0.20	3.09	359	0.13	7.95	167
RFPCM	0.96	0.63	76	1.11	1.66	198	0.19	3.87	101	0.13	8.82	98

TABLE IV
QUANTITATIVE ANALYSIS OF ROUGH–FUZZY CLUSTERING ON GLASS, IONOSPHERE, WINE, AND WISCONSIN BREAST-CANCER DATA SET

Algorithms /Methods	Glass Data Set			Ionosphere Data Set			Wine Data Set			Wisconsin Data Set		
	α	$\alpha^\star$	γ	α	$\alpha^\star$	γ	α	$\alpha^\star$	γ	α	$\alpha^\star$	γ
RFCM^{MBP}	0.9891	0.9719	0.5123	0.9918	0.9627	0.6217	0.8387	0.9251	0.5000	0.8977	0.9123	0.6349
RFCM	0.9942	0.9792	0.6250	0.9927	0.9712	0.6590	0.8918	0.9259	0.8275	0.8981	0.9386	0.8175
RPCM	0.9907	0.9775	0.5178	0.9913	0.9681	0.5125	0.8433	0.9306	0.6255	0.8990	0.9207	0.7724
RFPCM	0.9918	0.9804	0.6250	0.9987	0.9701	0.8271	0.9012	0.9258	0.7234	0.9192	0.9188	0.8125

In the Iris data set, since classes 2 and 3 overlap, it may be thought of having two clusters. However, to design a classifier, at least three clusters have to be identified. Thus, for such applications, RFCM, RPCM, and RFPCM will be more useful because they are not sensitive to noise, they can avoid coincident clusters, and their DB- and D-index values are far better than that of other algorithms, as reported in Table I. In addition, the execution time of different rough–fuzzy clustering algorithms is significantly lesser than that of different fuzzy and kernel-based fuzzy clustering algorithms.

Table II compares the performance of different rough–fuzzy clustering algorithms with respect to α, ϱ, $\alpha^\star$, and γ. The proposed RFCM performs better than RFCM^{MBP}. The performance of RFPCM is intermediate between RFCM and RPCM for $c = 2$ and is better over other rough–fuzzy clustering algorithms having $c = 3$. However, it is expected that RFPCM will be more useful as it is not sensitive to noise and outliers and does not produce coincident clusters.

Finally, Tables III and IV present the comparative performance analysis of different c-means algorithms on some other benchmark data sets with respect to DB, D, α, $\alpha^\star$, and γ indices and CPU time (in milliseconds). The following conclusions can be drawn from the results reported in Tables III and IV.

1) The performance of the proposed three algorithms (RFCM, RPCM, and RFPCM) is significantly better than other fuzzy, kernel-based fuzzy, and rough algorithms with respect to DB, D, α, $\alpha^\star$, and γ indices.
2) The execution time required for the different rough clustering algorithms is much lesser compared with the fuzzy and kernel-based fuzzy algorithms.
3) Some of the existing algorithms like PCM, FPCM, KPCM, and KFPCM generate coincident clusters even when they have been initialized with the final prototypes of FCM.
4) The performance of kernel-based algorithms is better than their nonkernel-based counterparts, although they require more time to converge.

TABLE V
PERFORMANCE OF DIFFERENT C-MEANS ALGORITHMS

Data Set	Algorithms	DB	Dunn	β	Time
IMAGE 20497761	HCM	0.16	2.13	12.07	719
	FCM	0.14	2.26	12.92	1813
	FPCM^{MR}	0.13	2.22	12.56	1026
	KHCM	0.16	2.17	12.11	1312
	KFCM	0.14	2.27	12.78	2017
	RCM	0.15	2.31	11.68	806
	RFCM^{MBP}	0.14	2.34	9.99	1052
	RFCM	0.13	2.39	13.06	1033
	RPCM	0.11	2.61	12.98	1179
	RFPCM	0.10	2.73	13.93	1201
IMAGE 20497763	HCM	0.18	1.88	12.02	706
	FCM	0.16	2.02	12.63	1643
	FPCM^{MR}	0.15	2.03	12.41	1132
	KHCM	0.17	2.01	12.10	1294
	KFCM	0.15	2.08	12.71	1983
	RCM	0.15	2.14	12.59	751
	RFCM^{MBP}	0.15	2.08	10.59	1184
	RFCM	0.11	2.12	13.30	1017
	RPCM	0.12	2.71	13.14	1105
	RFPCM	0.10	2.86	13.69	1057

B. Segmentation of Brain MR Images

In this section, we present the results of different c-means algorithms on the segmentation of brain MR images. The experimentation has been done in two parts. In the first part, we have used some real brain MR images. More than 100 brain MR images with different sizes and 16-bit gray levels are tested with different c-means algorithms. All the brain MR images are collected from Advanced Medicare and Research Institute, Salt Lake, Kolkata, India. In the second part, we present the segmentation result on some benchmark images obtained from "BrainWeb: Simulated Brain Database" (http://www.bic.mni.mcgill.ca/brainweb/). The comparative performance of different c-means is reported with respect to DB and D indices, as well as the β index [20].

1) Results on Real Brain MR Images: Table V compares the performance of different c-means algorithms on some brain

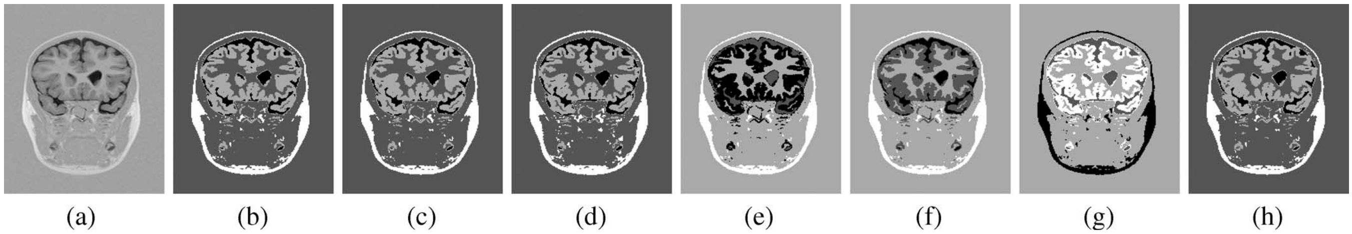


Fig. 3. IMAGE-20497761. Original and segmented versions of different c-means algorithms. (a) Original. (b) KHCM. (c) FCM. (d) RCM. (e) RFCM$^{\mathbf{MBP}}$. (f) RFCM. (g) RPCM. (h) RFPCM.

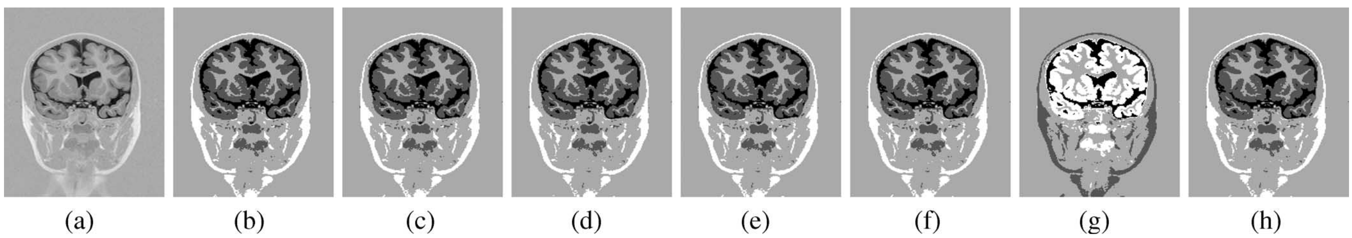


Fig. 4. IMAGE-20497763. Original and segmented versions of different c-means algorithms. (a) Original. (b) HCM. (c) FCM. (d) FPCM$^{\mathbf{MR}}$. (e) RCM. (f) RFCM. (g) RPCM. (h) RFPCM.

MR images with respect to DB, D, and β indices and CPU time, considering $c = 4$ (background, gray matter, white matter, and cerebrospinal fluid). The original images, along with the segmented versions of different c-means, are shown in Figs. 3 and 4. All the results reported in Table V and Figs. 3 and 4 show that, although each c-means algorithm, except PCM, FPCM, KPCM, and KFPCM, generates good segmented images, the values of DB and D, and β indices of the RFPCM are better compared with those of other c-means algorithms.

2) Results on Simulated Brain MR Images: Furthermore, extensive experimentation has been done to evaluate the performance of the RFPCM algorithm on simulated brain MR images. Figs. 5–7 show the original and segmented images obtained using the RFPCM algorithm for different slice thicknesses and noise levels. The noise is calculated relative to the brightest tissue. The results are reported for the three different slice thicknesses of -1, 3, and 5 mm, and the noise varies from 0% to 9%. Finally, Table VI compares the values of DB, D, and β indices of different c-means algorithms for different slice thicknesses and noise levels. All the results shown in Figs. 5–7 and Table VI confirm that the proposed RFPCM algorithm generates good segmented images that are irrespective of the slice thickness and noise level. In addition, the performance of the RFPCM in terms of DB, D, and β indices is significantly better compared with other c-means algorithms.

The following conclusions can be drawn from the results reported in this paper.

1) It is observed that the RFPCM is superior to other c-means algorithms. However, the RFPCM requires higher time compared with the HCM. However, the performance of the RFPCM is significantly higher than the other c-means. The performance of the RFCM and RPCM is intermediate between the RFPCM and FCM/PCM. In addition, the RFCM performs better than the RFCM$^{\mathbf{MBP}}$.
2) The use of rough sets and fuzzy memberships (both probabilistic and possibilistic) adds a small computational load to the HCM algorithm; however, the corresponding integrated methods (RFCM, RPCM, and RFPCM) show a definite increase in the D index and a decrease in the DB index.
3) It is seen that the performance of the RFPCM is intermediate between the RFCM and RPCM with respect to α, ϱ, $\alpha^\star$, and γ. However, the RFPCM will be more useful than the RFCM and RPCM as its prototypes are not sensitive to outliers and can avoid coincident cluster problem.
4) The execution time required for the different rough clustering algorithms is significantly lesser compared with the different fuzzy and kernel-based fuzzy clustering algorithms.
5) The proposed indices, such as α, ϱ, $\alpha^\star$, and γ, based on the theory of rough sets provide good quantitative measures for rough–fuzzy clustering. The values of these indices reflect the quality of clustering.

The best performance of the proposed RFPCM algorithm is achieved because of the following reasons.

1) The concept of crisp lower bound and fuzzy boundary of the proposed algorithm deals with uncertainty, vagueness, and incompleteness in class definition.
2) The membership function of the RFPCM efficiently handles overlapping partitions.
3) The probabilistic and possibilistic memberships of the RFPCM can avoid coincident cluster problem and make the algorithm insensitive to noise and outliers.

In effect, good cluster prototypes are obtained using the RFPCM algorithm with significantly lesser time.

VI. Conclusion

The contribution of this paper lies in the development of a generalized methodology, which judiciously integrates c-means algorithm, rough sets, and probabilistic and possibilistic memberships of fuzzy sets. This formulation is geared toward maximizing the utility of both rough and fuzzy sets with respect to

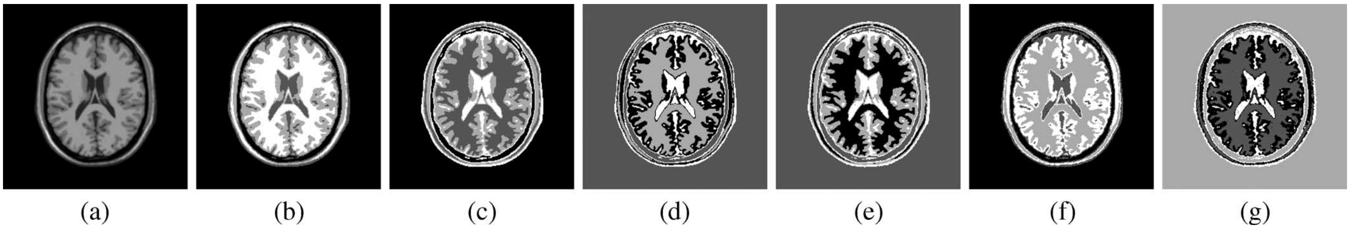


Fig. 5. Slice thickness = 1 mm. Original and segmented versions of RFPCM algorithm for different noise levels. (a) Original. (b) Noise = 0%. (c) Noise = 1%. (d) Noise = 3%. (e) Noise = 5%. (f) Noise = 7%. (g) Noise = 9%.

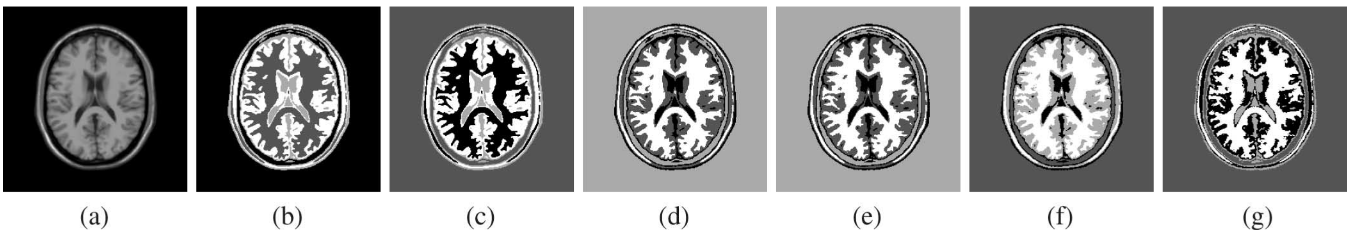


Fig. 6. Slice thickness = 3 mm. Original and segmented versions of RFPCM algorithm for different noise levels. (a) Original. (b) Noise = 0%. (c) Noise = 1%. (d) Noise = 3%. (e) Noise = 5%. (f) Noise = 7%. (g) Noise = 9%.

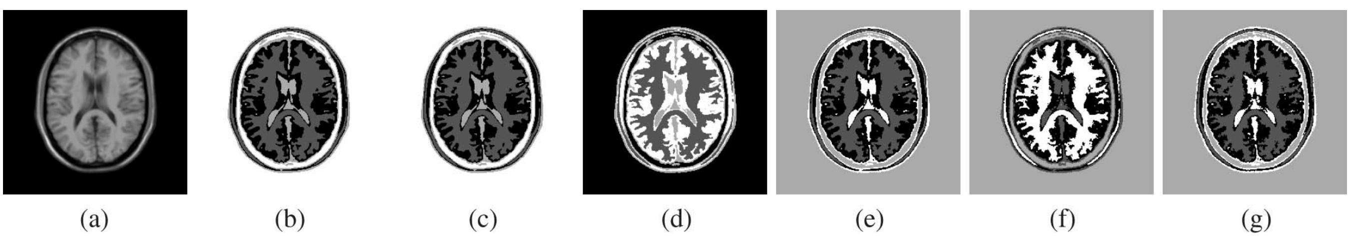


Fig. 7. Slice thickness = 5 mm. Original and segmented versions of RFPCM algorithm for different noise levels. (a) Original. (b) Noise = 0%. (c) Noise = 1%. (d) Noise = 3%. (e) Noise = 5%. (f) Noise = 7%. (g) Noise = 9%.

TABLE VI
VALUES OF DB, D, AND β INDICES FOR SIMULATED BRAIN MRI

Thickness	Algorithms /Methods	DB Index / Noise (%)					Dunn Index / Noise (%)					β Index / Noise (%)				
		0	3	5	7	9	0	3	5	7	9	0	3	5	7	9
1 mm	HCM	0.34	0.35	0.35	0.36	0.38	2.87	3.14	3.05	3.19	2.97	36.2	35.3	34.7	34.4	33.8
	FCM	0.29	0.29	0.31	0.30	0.30	3.63	3.66	3.78	3.92	3.99	41.6	40.4	39.5	37.8	35.9
	FPCM^{MR}	0.27	0.26	0.26	0.28	0.27	3.70	3.55	3.78	3.49	3.96	41.8	40.6	39.8	38.1	36.9
	KHCM	0.30	0.31	0.31	0.30	0.31	3.18	3.29	3.34	3.30	3.37	40.3	38.3	37.0	35.4	34.1
	KFCM	0.29	0.28	0.28	0.29	0.29	3.64	3.68	3.89	4.11	4.01	41.7	40.6	39.4	38.0	36.1
	RCM	0.17	0.18	0.17	0.18	0.19	5.74	5.58	6.30	5.65	7.43	45.0	42.7	42.7	39.5	38.2
	RFCM^{MBP}	0.26	0.25	0.27	0.27	0.27	6.13	6.58	6.91	7.03	6.95	46.1	45.2	43.8	40.1	39.5
	RFCM	0.12	0.12	0.13	0.13	0.14	8.34	8.79	8.88	8.90	8.33	56.9	56.0	53.2	50.3	45.2
	RPCM	0.13	0.13	0.13	0.12	0.12	8.53	8.91	9.07	9.11	8.70	56.5	55.1	52.0	49.4	44.5
	RFPCM	0.12	0.12	0.13	0.13	0.13	8.64	9.12	9.58	9.27	9.27	56.7	55.7	52.8	49.7	44.9
3 mm	HCM	0.41	0.40	0.42	0.42	0.42	2.51	2.56	2.67	2.48	2.70	32.7	29.1	27.8	24.1	22.3
	FCM	0.31	0.30	0.32	0.32	0.32	3.41	3.43	3.52	3.53	3.59	35.9	32.7	32.0	30.5	28.6
	FPCM^{MR}	0.29	0.31	0.31	0.31	0.32	3.49	3.51	3.44	3.41	3.75	35.7	33.1	32.2	31.4	29.5
	KHCM	0.36	0.36	0.37	0.36	0.35	3.01	3.19	3.19	3.23	3.24	34.3	31.0	29.7	28.8	26.1
	KFCM	0.29	0.29	0.29	0.30	0.31	3.41	3.43	3.57	3.59	3.59	36.1	32.9	31.8	30.7	29.5
	RCM	0.23	0.20	0.19	0.19	0.18	3.89	4.92	5.19	5.73	5.47	39.9	40.6	40.9	40.2	39.3
	RFCM^{MBP}	0.31	0.28	0.28	0.28	0.29	6.33	6.57	6.48	6.90	6.77	40.3	38.1	35.9	32.6	31.1
	RFCM	0.14	0.14	0.14	0.14	0.14	7.24	7.39	7.45	7.53	7.47	51.4	50.4	49.9	47.8	45.2
	RPCM	0.13	0.13	0.14	0.15	0.12	7.39	7.17	7.42	7.88	7.60	50.1	50.2	48.7	46.1	44.2
	RFPCM	0.13	0.13	0.13	0.13	0.13	8.42	8.29	8.44	8.89	9.25	50.4	49.7	49.0	46.8	44.5
5 mm	HCM	0.45	0.45	0.45	0.46	0.46	2.14	2.23	2.30	2.27	2.36	26.1	24.0	22.9	22.4	21.8
	FCM	0.32	0.32	0.32	0.30	0.33	3.02	3.09	3.15	3.22	3.28	32.7	30.2	29.9	26.5	27.6
	FPCM^{MR}	0.30	0.31	0.31	0.31	0.31	3.07	3.28	3.28	3.21	3.34	33.0	30.8	29.1	27.6	27.3
	KHCM	0.34	0.33	0.33	0.33	0.33	2.85	2.90	2.89	3.01	3.06	31.3	29.8	29.4	27.3	24.1
	KFCM	0.31	0.31	0.30	0.30	0.30	3.03	3.09	3.18	3.24	3.32	32.7	30.4	30.1	27.2	27.8
	RCM	0.19	0.24	0.21	0.18	0.18	5.34	3.74	5.01	5.95	5.18	40.2	38.3	38.5	40.3	38.0
	RFCM^{MBP}	0.28	0.29	0.29	0.30	0.29	5.31	5.02	4.99	4.71	4.95	43.6	40.2	39.5	39.0	38.8
	RFCM	0.15	0.15	0.15	0.15	0.15	5.76	6.06	5.95	6.03	6.05	48.1	48.3	47.3	46.4	43.0
	RPCM	0.15	0.14	0.14	0.14	0.14	6.12	6.11	6.28	6.23	6.29	45.4	43.8	42.0	41.3	41.0
	RFPCM	0.14	0.14	0.14	0.14	0.14	7.48	8.24	7.89	7.88	7.52	46.6	47.9	45.8	45.0	42.0

knowledge-discovery tasks. Several new measures are defined based on rough sets to evaluate the performance of rough–fuzzy clustering algorithms. Finally, the effectiveness of the proposed algorithm is demonstrated, along with a comparison with other related algorithms, on a set of synthetic as well as real-life data sets.

Although our methodology of the integration of rough sets, fuzzy sets, and c-means algorithm has efficiently been demonstrated for benchmark data sets, the concept can be applied to other unsupervised classification problems. Some of the indices (e.g., α, $\alpha^{\star}$, ϱ, and γ) that are used for the evaluation of the quality of the proposed algorithm may be used in a suitable combination to act as the objective function of an evolutionary algorithm for rough–fuzzy clustering.

ACKNOWLEDGMENT

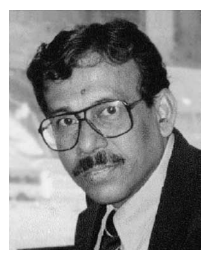
The authors would like to thank the anonymous referees for providing helpful comments and the Department of Science and Technology, Government of India, for funding the Center for Soft Computing Research: A National Facility under its Intensification of Research in High Priority Areas scheme. This paper was done when one of the authors, S. K. Pal, was a J.C. Bose Fellow of the Government of India.

REFERENCES

[1] A. K. Jain and R. C. Dubes, *Algorithms for Clustering Data*. Englewood Cliffs, NJ: Prentice-Hall, 1988.
[2] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification and Scene Analysis*. New York: Wiley, 1999.
[3] J. C. Bezdek, *Pattern Recognition With Fuzzy Objective Function Algorithm*. New York: Plenum, 1981.
[4] R. Krishnapuram and J. M. Keller, "A possibilistic approach to clustering," *IEEE Trans. Fuzzy Syst.*, vol. 1, no. 2, pp. 98–110, May 1993.
[5] N. R. Pal, K. Pal, J. M. Keller, and J. C. Bezdek, "A possibilistic fuzzy c-means clustering algorithm," *IEEE Trans. Fuzzy Syst.*, vol. 13, no. 4, pp. 517–530, Aug. 2005.
[6] F. Masulli and S. Rovetta, "Soft transition from probabilistic to possibilistic fuzzy clustering," *IEEE Trans. Fuzzy Syst.*, vol. 14, no. 4, pp. 516–527, Aug. 2006.
[7] Z. Pawlak, *Rough Sets, Theoretical Aspects of Reasoning About Data*. Dordrecht, The Netherlands: Kluwer, 1991.
[8] P. Lingras and C. West, "Interval set clustering of web users with rough K-means," *J. Intell. Inf. Syst.*, vol. 23, no. 1, pp. 5–16, 2004.
[9] D. Dubois and H. Prade, "Rough fuzzy sets and fuzzy rough sets," *Int. J. Gen. Syst.*, vol. 17, no. 2/3, pp. 191–209, 1990.
[10] S. Mitra, H. Banka, and W. Pedrycz, "Rough–fuzzy collaborative clustering," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 36, no. 4, pp. 795–805, Aug. 2006.
[11] P. Maji and S. K. Pal, "Rough–fuzzy C-medoids algorithm and selection of bio-basis for amino acid sequence analysis," *IEEE Trans. Knowl. Data Eng.*, vol. 19, no. 6, pp. 859–872, Jun. 2007.
[12] G. James, *Modern Engineering Mathematics*. Reading, MA: Addison-Wesley, 1996.
[13] J. 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.*, vol. SMC-17, no. 5, pp. 873–877, Sep./Oct. 1987.
[14] H. Yan, "Convergence condition and efficient implementation of the fuzzy curve-tracing (FCT) algorithm," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 34, no. 1, pp. 210–221, Feb. 2004.
[15] M. Girolami, "Mercer kernel-based clustering in feature space," *IEEE Trans. Neural Netw.*, vol. 13, no. 3, pp. 780–784, May 2002.
[16] S. Miyamoto and D. Suizu, "Fuzzy c-means clustering using kernel functions in support vector machines," *J. Adv. Comput. Intell. Intell. Inf.*, vol. 7, no. 1, pp. 25–30, 2003.
[17] D.-Q. Zhang and S.-C. Chen, "Kernel based fuzzy and possibilistic c-means clustering," in *Proc. ICANN*, Istanbul, Turkey, 2003, pp. 122–125.
[18] X.-H. Wu and J.-J. Zhou, "Possibilistic fuzzy c-means clustering model using kernel methods," in *Proc. Int. Conf. Comput. Intell. Model., Control Autom.*, 2005, pp. 465–470.
[19] J. C. Bezdek and N. R. Pal, "Some new indexes for cluster validity," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 28, no. 3, pp. 301–315, Jun. 1988.
[20] S. K. Pal, A. Ghosh, and B. U. Sankar, "Segmentation of remotely sensed images with fuzzy thresholding, and quantitative evaluation," *Int. J. Remote Sens.*, vol. 21, no. 11, pp. 2269–2300, Jul. 2000.

Pradipta Maji received the B.Sc. (Hons.) degree in physics, the M.Sc. degree in electronics science, and the Ph.D. degree in computer science from Jadavpur University, Kolkata, India, in 1998, 2000, and 2005, respectively.

Currently, he is a Lecturer with the Machine Intelligence Unit, Indian Statistical Institute, Kolkata. He is also with the Center for Soft Computing Research: A National Facility, Indian Statistical Institute. His research interests include pattern recognition, bioinformatics, medical image processing, cellular automata, soft computing, etc. He has published around 40 papers in international journals and conferences. He is also a Reviewer of many international journals.



Sankar K. Pal (M'81–SM'84–F'93) received the Ph.D. degree in radio physics and electronics from the University of Calcutta, Kolkata, India, in 1979 and the Ph.D. degree in electrical engineering, along with the Diploma of the Imperial College, from Imperial College London, University of London, London, U.K., in 1982.

He was with the University of California, Berkeley, and the University of Maryland, College Park, from 1986 to 1987, the NASA Johnson Space Center, Houston, TX, from 1990 to 1992 and in 1994, and the U.S. Naval Research Laboratory, Washington, DC, in 2004. Since 1997, he has been a Distinguished Visitor of the IEEE Computer Society (U.S.) for the Asia-Pacific region, and he has held several visiting positions in Hong Kong and Australian universities. He is currently the Director and a Distinguished Scientist with the Indian Statistical Institute, Kolkata. He is also a J.C. Bose Fellow with the Government of India. He founded the Machine Intelligence Unit and the Center for Soft Computing Research: A National Facility, Indian Statistical Institute. He is a coauthor of 14 books and about 300 research publications in pattern recognition and machine learning, image processing, data mining and Web intelligence, soft computing, neural nets, genetic algorithms, fuzzy sets, rough sets, and bioinformatics.

Dr. Pal is a Fellow of the Academy of Sciences for the Developing World, the International Association for Pattern Recognition, the International Fuzzy Systems Association, and all four national academies for science/engineering in India. He is/has acting/acted as an Associate Editor of the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE (2002–2006), IEEE TRANSACTIONS ON NEURAL NETWORKS (1994–1998 and 2003–2006), *Pattern Recognition Letters*, *International Journal of Pattern Recognition and Artificial Intelligence*, *Neurocomputing* (1995–2005), *Applied Intelligence*, *Information Sciences*, *Fuzzy Sets and Systems*, *Fundamenta Informaticae*, *International Journal of Computational Intelligence and Applications*, and *Proceedings of the Indian National Science Academy-A*. He is a member of the Executive Advisory Editorial Board of the IEEE TRANSACTIONS ON FUZZY SYSTEMS, *International Journal on Image and Graphics*, and *International Journal of Approximate Reasoning*; and a Guest Editor of IEEE COMPUTER. He received the 1990 S. S. Bhatnagar Prize (which is the most coveted award for a scientist in India) and many prestigious awards in India and abroad, including the 1999 G. D. Birla Award, the 1998 Om Bhasin Award, the 1993 Jawaharlal Nehru Fellowship, the 2000 Khwarizmi International Award from the Islamic Republic of Iran, the 2000–2001 Federation of Indian Chamber of Commerce and Industry Award, the 1993 Vikram Sarabhai Research Award, the 1993 NASA Tech Brief Award (USA), the 1994 IEEE TRANSACTIONS ON NEURAL NETWORKS Outstanding Paper Award (USA), the 1995 NASA Patent Application Award (USA), the 2005–2006 Indian Science Congress–P. C. Mahalanobis Birth Centenary Award (Gold Medal) for Lifetime Achievement, the 1997 IETE-R. L. Wadhwa Gold Medal, and the 2001 INSA-S. H. Zaheer Medal.