Pattern Recognition 43 (2010) 1282–1291

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.de/locate/pr

Shadowed c-means: Integrating fuzzy and rough clustering

Sushmita Mitra [a,*], Witold Pedrycz [b], Bishal Barman [c]

[a] Machine Intelligence Unit, Indian Statistical Institute, Kolkata 700 108, India
[b] Electrical and Computer Engineering Department, University of Alberta, Edmonton, Canada T6G 2G7
[c] Electrical Engineering Department, S.V. National Institute of Technology, Surat 395 007, India

ARTICLE INFO

Article history:
Received 29 December 2008
Received in revised form
17 September 2009
Accepted 29 September 2009

Keywords:
Shadowed sets
c-Means algorithm
Three-valued logic
Cluster validity index
Fuzzy sets
Rough sets

ABSTRACT

A new method of partitive clustering is developed in the framework of shadowed sets. The core and exclusion regions of the generated shadowed partitions result in a reduction in computations as compared to conventional fuzzy clustering. Unlike rough clustering, here the choice of threshold parameter is fully automated. The number of clusters is optimized in terms of various validity indices. It is observed that shadowed clustering can efficiently handle overlapping among clusters as well as model uncertainty in class boundaries. The algorithm is robust in the presence of outliers. A comparative study is made with related partitive approaches. Experimental results on synthetic as well as real data sets demonstrate the superiority of the proposed approach.

1. Introduction

The goal of clustering is to partition N patterns into c desired clusters with high *intra-class* similarity and low *inter-class* similarity while optimizing an objective function. It pertains to unsupervised learning, when data with class labels are not available. In the partitive approach one needs to provide the desired number of clusters c. The traditional c-means algorithm [18] represents each such cluster by its center of gravity. There exist several other variations, like partitioning around medoids (PAM) [5] and c-modes clustering [4], differing mainly in the manner of choosing their cluster prototypes.

Soft computing is a consortium of methodologies that works synergistically and provides flexible information processing capability for handling real life ambiguous situations [20]. Its aim is to exploit the tolerance for imprecision, uncertainty, approximate reasoning, and partial truth in order to achieve tractability, robustness, and low-cost solutions. The main constituents of soft computing, at this juncture, include fuzzy logic, neural networks, genetic algorithms and rough sets. The use of soft computing in clustering large data has been reported in literature [9]. Fuzzy sets and rough sets were incorporated in the c-means framework to develop the fuzzy c-means (FCM) [1], rough c-means (RCM) [7,10,15] and rough–fuzzy c-means (RFCM) [10,8] algorithms, respectively. While membership in FCM enables efficient handling of overlapping partitions, the rough sets [12] deal with uncertainty, vagueness and incompleteness of data in terms of upper and lower approximations.

Questions have been raised about the modeling of a vague phenomenon with precise numeric membership values. In order to disambiguate and capture the essence of a distribution, recently the concept of shadowed sets has been proposed in the literature [13]. It shares a close relationship with Lukasiewicz's concept of three-valued logic [16], as it tends to partition the distribution into three distinct zones, viz., core, shadowed and exclusion zones.

Refinement to the existing fuzzy and rough clustering approaches, based on the concept of shadowed sets, is proposed in this article. Shadowed clustering serves as a conceptual and algorithmic bridge between the FCM and RCM, thereby incorporating the generic merits of both these approaches. Originally, shadowed sets were introduced to characterize fuzzy sets. In this sense they were induced by fuzzy sets, and hence they were algorithmically implied by fuzzy sets. If used as stand-alone constructs in clustering, they are different from fuzzy sets as they introduce (0,1) intervals to denote only those points for which we have no information regarding belongingness. This uncertainty among patterns lying in the shadowed region is efficiently handled in terms of membership. On the other hand, the contrast between the core and the exclusion zones is enhanced; thereby reducing computation in these regions which are modeled as $\{1,0\}$. While in fuzzy sets we use membership values [1, 0] for the entire region, in shadowed sets the membership calculations are restricted to the shadowed region only. Analogously, in rough sets, it corresponds to regions outside the lower approximation and within the upper approximation. However, in rough sets we have an added burden of

* Corresponding author. Tel.: +91 33 2575 3102; fax: +91 33 2578 3357.
E-mail addresses: sushmita@isical.ac.in (S. Mitra), pedrycz@ece.ualberta.ca (W. Pedrycz), bishalbarman@gmail.com (B. Barman).

0031-3203/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2009.09.029

several user-defined parameters. The issue of automated threshold selection in shadowed sets is analyzed in this context.

There is a strong motivating factor behind the use of shadowed sets in pattern recognition. The three-valued logic along with the detailed algorithmic setting, showing how shadowed sets are constructed, are crucial to the design of classifiers and interpretation of classification results obtained in this manner. In addition to the binary yes–no classification results, we are also provided with a region in the feature space with the 'don't know' quantification. This description is crucial to the prudent generation of the class assignment. In the third, 'don't know' scenario, the classifier is self-flagging—which means that in such cases we are at position to examine the existing evidence about the pattern under discussion before proceeding with making decision about assigning class label.

The salient features of the algorithm include: (i) reduction in the number of user-defined parameters and (ii) robustness to presence of outliers. The Davies–Bouldin [2], Xie–Beni [19] and Silhouette [17] clustering validity indices are employed to generate the optimal number of clusters c. The superior comparative performance of the proposed algorithm is established on four sets of real and synthetic data.

The rest of the paper is organized as follows. Section 2 outlines the existing partitive algorithms in the soft computing framework. Validity indices, to optimize the number of clusters, are also described. Section 3 presents the central idea of shadowed sets along with the proposed shadowed c-means clustering algorithm. Comparative experimental results on synthetic as well as real data demonstrate, in Section 4, the effectiveness of the proposed clustering method. The article is concluded in Section 5 with a discussion on the applicability of the obtained results.

2. Soft partitive clustering and validation

In this section we describe a few soft computing-based partitive clustering algorithms. These include the fuzzy c-means (FCM), rough c-means (RCM) and rough–fuzzy c-means (RFCM) algorithms. The objective is to contrast the essence of each of these algorithms in a unified fashion.

The conventional c-means algorithm [18] proceeds by partitioning N objects $\mathbf{x}_k$ into c non-empty subsets. During each partition, the centroids or means of clusters are computed as

$$\mathbf{v}_i = \frac{\sum_{\mathbf{x}_k \in U_i} \mathbf{x}_k}{|c_i|}, \tag{1}$$

where $|c_i|$ is the number of objects in cluster U_i. The process is repeated until convergence, i.e., there are no more new assignments of objects to the clusters.

2.1. Fuzzy c-means (FCM)

This is an extension of the c-means algorithm, as proposed by Bezdek [1], in the sense that we allow for partial membership of patterns to clusters. It partitions a set of N patterns $\{\mathbf{x}_k\}$ into c clusters by minimizing the objective function

$$\mathcal{J} = \sum_{k=1}^{N} \sum_{i=1}^{c} (u_{ik})^m \|\mathbf{x}_k - \mathbf{v}_i\|^2, \tag{2}$$

where $m > 1$ is the fuzzifier, $u_{ik} \in [0,1]$ is the membership of the k th pattern to cluster center $\mathbf{v}_i$, and $\|.\|$ is the Euclidean distance, such that

$$\mathbf{v}_i = \frac{\sum_{k=1}^{N} (u_{ik})^m \mathbf{x}_k}{\sum_{k=1}^{N} (u_{ik})^m} \tag{3}$$

and

$$u_{ik} = \frac{1}{\sum_{j=1}^{c} \left(\frac{d_{ik}}{d_{jk}}\right)^{2/m-1}} \tag{4}$$

$\forall i$, with $d_{ik} = \|\mathbf{x}_k - \mathbf{v}_i\|^2$, subject to $\sum_{i=1}^{c} u_{ik} = 1$, $\forall k$, and $0 < \sum_{k=1}^{N} u_{ik} < N$. The object assignment and mean computation are repeated until $|u_{ik}(t) - u_{ik}(t-1)| < \varepsilon$, at iteration t. Note that for $u_{ik} \in \{0,1\}$ the objective function of Eq. (2) boils down to the hard c-means case, whereby a *winner-take-all* strategy is applied in place of membership values in Eq. (3).

2.2. Rough c-means (RCM)

In rough sets [12] we approximate a *rough* (imprecise) concept by a pair of *exact* concepts, called the lower and upper approximations. The lower approximation is the set of objects definitely belonging to the vague concept, whereas the upper approximation is the set of objects possibly belonging to the same. Fig. 1 provides a schematic diagram of a rough set X within the upper and lower approximations, consisting of granules coming from the rectangular grid.

In RCM, the concept of c-means is extended by viewing each cluster as an interval or rough set [7] X. It is characterized by the lower and upper approximations $\underline{B}X$ and $\overline{B}X$, respectively, with the following properties: (i) an object $\mathbf{x}_k$ can be part of at most *one* lower approximation; (ii) if $\mathbf{x}_k \in \underline{B}X$ of cluster X, then simultaneously $\mathbf{x}_k \in \overline{B}X$; and (iii) if $\mathbf{x}_k$ is not a part of any lower approximation, then it belongs to two or more upper approximations. This permits overlaps between clusters.

The right hand side of Eq. (1) is split into two parts. Since the patterns lying in the lower approximation definitely belong to a rough cluster, they are assigned a higher weight that is controlled by parameter w_{low}. The patterns lying in the upper approximation are assigned a relatively lower weight, controlled by parameter w_{up} during computation. The centroid of cluster U_i is determined [10,15] as

$$\mathbf{v}_i = \begin{cases} w_{low} \dfrac{\sum_{\mathbf{x}_k \in \underline{B}U_i} \mathbf{x}_k}{|\underline{B}U_i|} + w_{up} \dfrac{\sum_{\mathbf{x}_k \in (\overline{B}U_i - \underline{B}U_i)} \mathbf{x}_k}{|\overline{B}U_i - \underline{B}U_i|} & \text{if } \underline{B}U_i \neq \emptyset \wedge \overline{B}U_i - \underline{B}U_i \neq \emptyset, \\ \dfrac{\sum_{\mathbf{x}_k \in (\overline{B}U_i - \underline{B}U_i)} \mathbf{x}_k}{|\overline{B}U_i - \underline{B}U_i|} & \text{if } \underline{B}U_i = \emptyset \wedge \overline{B}U_i - \underline{B}U_i \neq \emptyset, \\ \dfrac{\sum_{\mathbf{x}_k \in \underline{B}U_i} \mathbf{x}_k}{|\underline{B}U_i|} & \text{otherwise,} \end{cases} \tag{5}$$

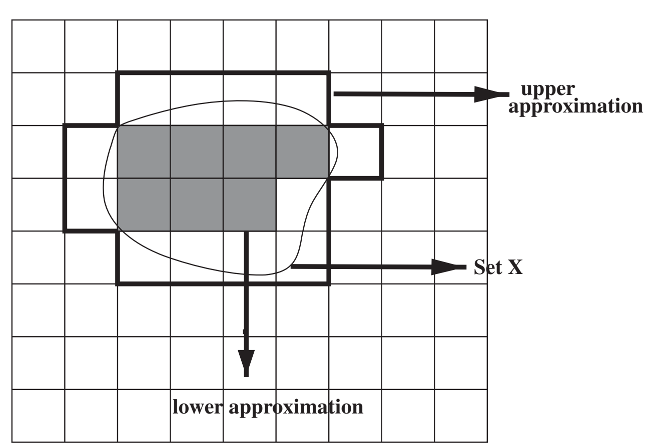

Fig. 1. Lower and upper approximations in a rough set.

where the parameters w_{low} and w_{up} correspond to the relative importance of the lower and upper approximations, respectively. Here $|\underline{B}U_i|$ indicates the number of patterns in the lower approximation of cluster U_i, while $|\overline{B}U_i - \underline{B}U_i|$ is the number of patterns in the rough boundary lying between the two approximations. Let d_{ik} be minimum and d_{jk} be next to minimum.

If $d_{jk} - d_{ik}$ is less than some *threshold*

then $\mathbf{x}_k \in \overline{B}U_i$ and $\mathbf{x}_k \in \overline{B}U_j$ and $\mathbf{x}_k$ cannot be a member of any lower approximation,

else $\mathbf{x}_k \in \underline{B}U_i$ such that distance d_{ik} is minimum over the c clusters.

The mean computation is repeated until there are no more new assignments of objects. Expression (5) transforms to (1) when the lower approximation is equal to the upper approximation, implying an empty boundary region.

It is observed that the performance of the algorithm is dependent on the choice of w_{low}, w_{up} and *threshold*. The parameter *threshold* measures the relative distance of an object $\mathbf{x}_k$ from a pair of clusters having centroids $\mathbf{v}_i$ and $\mathbf{v}_j$. The parameter w_{low} controls the importance of the objects lying within the lower approximation of a cluster in determining its centroid. Hence an optimal selection of these parameters is an issue of reasonable interest. We allowed $w_{up} = 1 - w_{low}$, $0.5 < w_{low} < 1$ and $0 < threshold < 0.5$.

2.3. Rough–fuzzy c-means (RFCM)

A rough–fuzzy c-means algorithm, involving an integration of fuzzy and rough sets, has been developed [10]. This allows one to incorporate fuzzy membership value u_{ik} of a sample $\mathbf{x}_k$ to a cluster mean $\mathbf{v}_i$, relative to all other means $\mathbf{v}_j \forall j \neq i$, instead of the absolute individual distance d_{ik} from the centroid. This sort of relativistic measure, in terms of Eqs. (3) and (4), enhances the robustness of the clustering with respect to different choices of parameters. The major steps of the algorithm are provided below.

1. Assign initial means $\mathbf{v}_i$ for the c clusters.
2. Compute u_{ik} by Eq. (4) for c clusters and N data objects.
3. Assign each data object (pattern) $\mathbf{x}_k$ to the lower approximation $\underline{B}U_i$ or upper approximation $\overline{B}U_i$, $\overline{B}U_j$ of cluster pairs U_i, U_j by computing the difference in its membership $u_{ik} - u_{jk}$ to cluster centroid pairs $\mathbf{v}_i$ and $\mathbf{v}_j$. Distances d_{ik}, $\forall i$, are normalized in [0, 1].
4. Let u_{ik} be maximum and u_{jk} be the next to maximum.
   If $u_{ik} - u_{jk}$ is less than some *threshold*
   then $\mathbf{x}_k \in \overline{B}U_i$ and $\mathbf{x}_k \in \overline{B}U_j$ and $\mathbf{x}_k$ cannot be a member of any lower approximation,
   else $\mathbf{x}_k \in \underline{B}U_i$ such that membership u_{ik} is maximum over the c clusters.
5. Compute new mean for each cluster U_i, incorporating (3) and (4) into (5), as

$$\mathbf{v}_i = \begin{cases} w_{low}\dfrac{\sum_{\mathbf{x}_k \in \underline{B}U_i} u_{ik}^m \mathbf{x}_k}{\sum_{\mathbf{x}_k \in \underline{B}U_i} u_{ik}^m} + w_{up}\dfrac{\sum_{\mathbf{x}_k \in (\overline{B}U_i - \underline{B}U_i)} u_{ik}^m \mathbf{x}_k}{\sum_{\mathbf{x}_k \in (\overline{B}U_i - \underline{B}U_i)} u_{ik}^m} & \text{if } \underline{B}U_i \neq \emptyset \wedge \overline{B}U_i - \underline{B}U_i \neq \emptyset, \\ \dfrac{\sum_{\mathbf{x}_k \in (\overline{B}U_i - \underline{B}U_i)} u_{ik}^m \mathbf{x}_k}{\sum_{\mathbf{x}_k \in (\overline{B}U_i - \underline{B}U_i)} u_{ik}^m} & \text{if } \underline{B}U_i = \emptyset \wedge \overline{B}U_i - \underline{B}U_i \neq \emptyset, \\ \dfrac{\sum_{\mathbf{x}_k \in \underline{B}U_i} u_{ik}^m \mathbf{x}_k}{\sum_{\mathbf{x}_k \in \underline{B}U_i} u_{ik}^m} & \text{otherwise}. \end{cases} \tag{6}$$

6. **Repeat** Steps 2–5 **until** convergence, i.e., there are no more new assignments.

As in the case of RCM, we use $w_{up} = 1 - w_{low}$, $0.5 < w_{low} < 1$, $m = 2$, and $0 < threshold < 0.5$. Analogous to (5), here the terms on the right hand side indicate computations in the non-empty lower approximation and boundary region, respectively. Additionally, the membership concept is incorporated in the rough–fuzzy formalism.

3. Shadowed c-means

Extending the concept of c-means algorithm, a novel clustering algorithm called shadowed c-means (SCM) is proposed here. Shadowed sets come as an interesting formal construct in Granular Computing in the sense that they help 'localize' uncertainty of membership grades in some selected regions of the universe of discourse and elevate or reduce membership grades in some other regions. In this sense, we arrive at a construct in which the resulting uncertainty is reflected in the form of the unit interval. The underlying algorithm offers a well-defined way of forming a shadowed set on a basis of the detailed numeric membership grades of fuzzy sets. It is able to efficiently handle overlaps among clusters, while modeling uncertainty at the boundaries. The algorithm is robust to the presence of outliers. Cluster validity indices are used to determine the optimal number of clusters.

At the onset we present the central idea of shadowed set theory [13,3] and its inherent three-valued logic [16]. The optimization equation to automatically determine the threshold parameter is delineated. Various cluster validity indices, like Davies–Bouldin, Xie–Beni and Silhouette, are also outlined. A few analytical questions are finally discussed, regarding the optimum threshold computation.

3.1. Shadowed sets

Conventional uncertainty models like fuzzy sets tend to capture vagueness exclusively through membership values. This poses a dilemma of excessive precision in describing imprecise phenomenon. The notion of shadowed sets tries to solve this problem of selecting the optimum level of resolution in precision.

The motivation behind the debate on excessive precision of fuzzy sets is the conceptual shortcoming associated with precise numeric values of membership used to describe vague concepts. While there is hardly any difficulty in assigning membership grades close to 1 or 0, a lot of uncertainty prevails during the assignment of membership grade of 0.5. Based on this central idea, Pedrycz [13] developed the concept of shadowed sets to improve the observability and interpretability of vague phenomenon.

Consider a fuzzy set $\mathbb{J}$ as depicted in Fig. 2. We attempt to modulate the membership values (MVs) on the lines of three-valued logic by elevating and reducing some MVs and balancing the uncertainty thus introduced. The elevation, reduction and balance of uncertainty is quite radical. We try to disambiguate the concept represented by the original fuzzy set by promoting a few of the MVs to one and reducing a few other MVs to zero. Such enhancement of contrast reduces the number of computations as compared to the fuzzy framework. In order to maintain the overall level of vagueness, some other region is defined as the zone of uncertainty. Provision is made so that this particular area of the universe of discourse has intermediate membership values on a unit interval between [0, 1], but left undefined. Rather than a single value, the entire unit interval can be marked as a non-numeric model of membership grade. Note that to induce a shadowed set, a fuzzy set must accept a specific threshold level. Effectively, this transforms the domain of discourse into clearly marked zones of vagueness. This mapping is called a shadowed

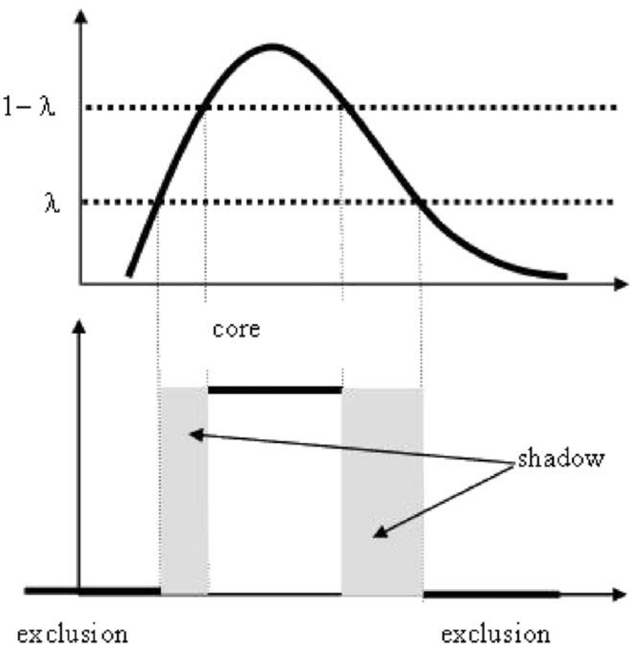

Fig. 2. The fuzzy set J inducing a shadowed set via a threshold.

set, and is defined as

$$\mathbb{J} : \mathbf{X} \rightarrow \{0, 1, [0, 1]\}.$$

Here elements with grade equal to one constitute the core, while the elements with $\mathbb{J}(x) = [0, 1]$ lie in the shadow of the mapping; the rest form the exclusion. These zones are clearly demarcated in Fig. 2. Originally, shadowed sets were introduced to characterize fuzzy sets.

The most appealing question concerns the computation of the threshold. Pedrycz proposed an optimization based on balance of vagueness. Reduction of some MVs to zero and elevation to one should be compensated by marked indecision in the other zones, or increased uncertainty in MVs in the form of a unit interval [0, 1] over particular ranges of $\mathbb{J}$. A particular threshold λ is selected for the quantification process and is expressed in terms of the relationship

$$\mathcal{O}(\lambda) = \left| \int_{-\infty}^{L_1} J(x)\,dx + \int_{L_2}^{\infty} (1 - J(x))\,dx - \int_{L_1}^{L_2} dx \right|, \tag{7}$$

where $\lambda \in (0, \frac{1}{2})$ such that $\mathcal{O}(\lambda) = 0$. The three terms on the right hand side of Eq. (7) correspond to regions A_1, A_2 and A_3 in Fig. 3. The parameters L_1 and L_2 denote the boundaries in the integral, delineating the regions in the figure where the membership values are below the threshold λ and above the threshold $1 - \lambda$.

Shadowed sets reveal interesting relationships with rough sets. Although conceptually similar, we must remember that the mathematical foundation of rough sets is very different. In rough sets, approximation spaces are defined in advance and the equivalent classes are kept fixed. On the other hand, in shadowed sets the class assignment is dynamic.

In the discrete domain, Eq. (7) gets transformed to

$$\mathcal{O}(\lambda_i) = \left| \sum_{\mathbf{x_k} | u_{ik} \le \lambda_i} u_{ik} + \sum_{\mathbf{x_k} | u_{ik} \ge u_{i_{\max}} - \lambda_i} (u_{i_{\max}} - \lambda_i) - \mathrm{card}\{\mathbf{x_k} | \lambda_i < u_{ik} < (u_{i_{\max}} - \lambda_i)\} \right|, \tag{8}$$

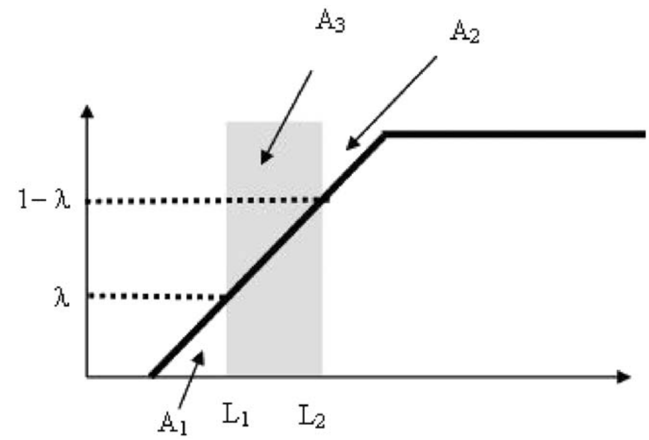

Fig. 3. Computing the threshold λ via an optimization: $A1 + A2 = \mathrm{card}\{A3\}$.

such that

$$\lambda_i = \lambda_{opt} = \underset{\lambda_i}{\arg\min}\, \mathcal{O}(\lambda_i), \tag{9}$$

where u_{ik}, $u_{i_{\min}}$ and $u_{i_{\max}}$ denote, respectively, the discrete, the lowest and the highest membership values to the i th class. Computation of the threshold λ for common membership functions, such as the triangular and Gaussian, has been reported in the literature [14]. The minima of $\mathcal{O}(\lambda)$ leads to the evaluation of λ_{opt}.

3.2. Shadowed clustering

Based on the concept of shadowed sets, we develop here the shadowed *c*-means clustering algorithm or *SCM*. The quantization of the MVs into core, shadowed and exclusion region permit reduced computational complexity [13]. We believe that the elements corresponding to the core should not have any fuzzy weight factor in terms of its MV. In other words, unlike uniform computation of MVs as in FCM, here the MV should be unity for core patterns while calculating the centroid. The elements corresponding to the shadowed region lie in the zone of uncertainty, and are treated as in FCM. However, the members of the exclusion region are incorporated in a slightly different manner. Here the fuzzy weight factor for the exclusion is designed to have the fuzzifier raised to itself, in the form of a double exponential. The centroid for the i th class is evaluated as

$$\mathbf{v_i} = \frac{\sum_{\mathbf{x}_k | u_{ik} \ge (u_{i_{\max}} - \lambda_i)} \mathbf{x}_k + \sum_{\mathbf{x}_k | \lambda_i < u_{ik} < (u_{i_{\max}} - \lambda_i)} (u_{ik})^m \mathbf{x}_k + \sum_{\mathbf{x}_k | u_{ik} \le \lambda_i} (u_{ik})^{m^m} \mathbf{x}_k}{\phi_i + \eta_i + \psi_i}, \tag{10}$$

where

$$\phi_i = \mathrm{card}\ \{\mathbf{x}_k | u_{ik} \ge (u_{i_{\max}} - \lambda_i)\}, \tag{11}$$

$$\eta_i = \sum_{\mathbf{x}_k | \lambda_i < u_{ik} < (u_{i_{\max}} - \lambda_i)} (u_{ik})^m, \tag{12}$$

$$\psi_i = \sum_{\mathbf{x}_k | u_{ik} \le \lambda_i} (u_{ik})^{m^m}, \tag{13}$$

and λ_i is the corresponding threshold. This arrangement causes a much wider dispersion and a very low bias factor for elements which can generally be considered outside the class under discussion or most definitely, the exclusion members. This prevents the mean from getting drifted from its true value. It also minimizes the effect of noise and outliers. The threshold to induce the core, shadowed and exclusion region is automatically calculated through a functional optimization using Eq. (8).

The mean in Eq. (10) basically tries to first get a coarse idea regarding the cluster prototype (using the first term in the numerator and denominator, respectively) and then proceeds to tune and refine this value using data from the shadowed and exclusion region. This enables a better estimation of the actual cluster prototypes. The major steps of the algorithm are outlined below.

1. Assign initial means, $\mathbf{v_i}$, $i=1,\ldots,c$. Choose values for fuzzifier, m and $t_{\max}$. Set iteration counter $t=1$.
2. **Repeat** Steps 3–5 by incrementing t **until** no new assignment is made and $t<t_{\max}$.
3. Compute u_{ik} by Eq. (4) for c clusters and N data objects.
4. Compute threshold λ_i for the i th class, in terms of Eq. (8).
5. Update mean, $\mathbf{v_i}$, using Eq. (10).

The range of feasible values of λ_i for the i th class could be taken as $[u_{i_{\min}}, u_{i_{\min}}+u_{i_{\max}}/2]$.

3.3. Validity indices

Partitive clustering algorithms typically require pre-specification of the number of clusters. Hence the results are dependent on the choice of c. However, there exist validity indices to evaluate the goodness of clustering, corresponding to a given value of c. In this article we compute the optimal number of clusters c_0 in terms of the Davies–Bouldin cluster validity index [2], Xie–Beni index [19] and Silhouette index [17].

1. The Davies–Bouldin index is a function of the ratio of the sum of within-cluster distance to between-cluster separation. The optimal clustering, for $c=c_0$, minimizes

$$DB=\frac{1}{c}\sum_{k=1}^{c}\max_{l\neq k}\left\{\frac{d_w(U_k)+d_w(U_l)}{d_b(U_k,U_l)}\right\} \tag{14}$$

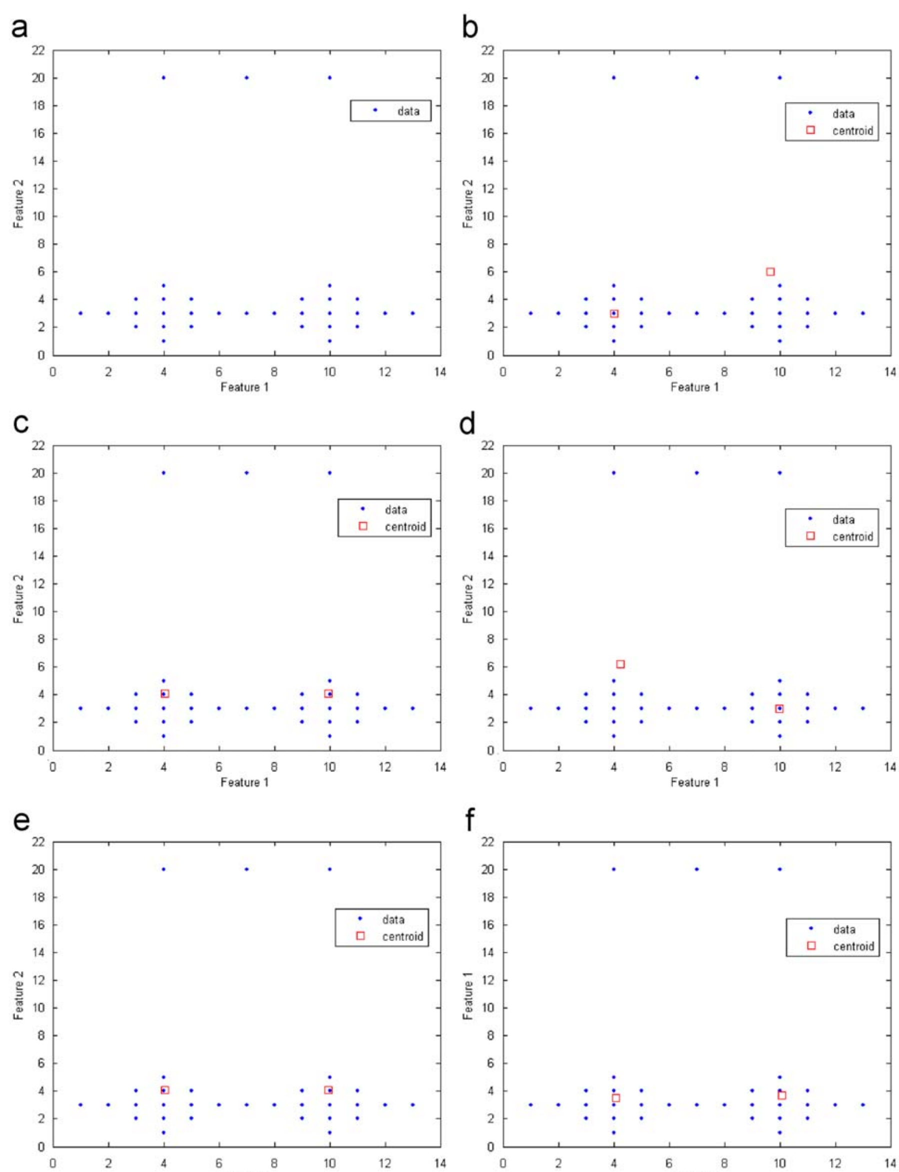

Fig. 4. Synthetic data $X32$: (a) original, and after clustering with; (b) HCM; (c) FCM; (d) RCM; (e) RFCM; and (f) SCM algorithms.

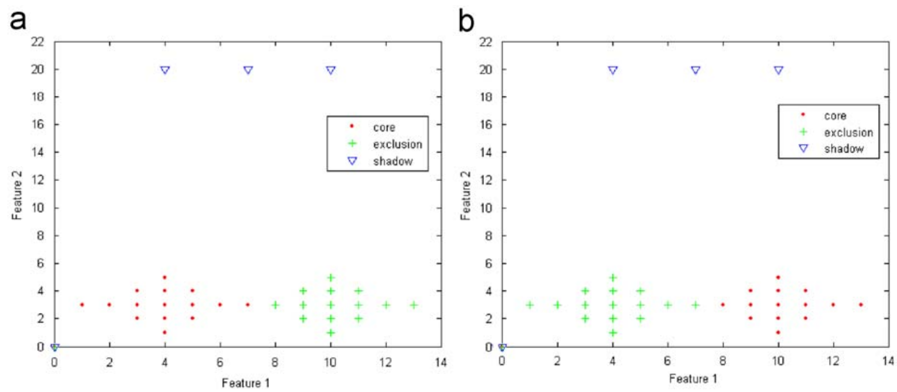

Fig. 5. Global view of exclusion and shadow regions, with the core as: (a) cluster 1 and (b) cluster 2, using SCM on $X32$.

for $1 \leq k, l \leq c$. In this process, the within-cluster distance $d_w(U_k)$ is minimized and the between-cluster separation $d_b(U_k, U_l)$ is maximized. The distance can be chosen as the traditional Euclidean metric for numeric features.

2. The Xie–Beni index [19] presents a fuzzy-validity criterion based on a validity function which identifies overall compact and separate fuzzy c-partitions. This function depends upon the data set, geometric distance measure, distance between cluster centroids and on the fuzzy partition, irrespective of any fuzzy algorithm used.
We define χ as a fuzzy clustering validity function

$$\chi = \frac{\sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^2 \|\mathbf{v}_i - \mathbf{x}_j\|^2}{N \min_{i,j} \|\mathbf{v}_i - \mathbf{v}_j\|^2}. \tag{15}$$

In case of FCM and RFCM algorithms, with $m = 2$, (15) reduces to

$$S = \frac{\mathcal{J}}{N \times d_{\min}}, \tag{16}$$

where $\mathcal{J}$ is the fuzzy objective function of (2) and $d_{\min} = \min_{i,j} \|\mathbf{v}_i - \mathbf{v}_j\|^2$. The more separate the clusters, the larger $d_{\min}$ and the smaller χ. Thus the smallest χ, corresponding to $c = c_0$, is indeed indicative of a valid optimal partition.

3. The Silhouette statistic [17], though computationally more intensive, is another way of estimating the number of clusters in a distribution. The Silhouette index, S, computes for each point a width depending on its membership in any cluster. This silhouette width is then an average over all observations. This is expressed as

$$S_k = \frac{1}{N} \sum_{i=1}^{N} \frac{b_i - a_i}{\max(a_i, b_i)}, \tag{17}$$

where N is the total number of points, a_i is the average distance between pattern $\mathbf{x}_i$ and all other points in its own cluster, and b_i is the minimum of the average dissimilarities between $\mathbf{x}_i$ and patterns in other clusters. Finally, the global silhouette index, S, of the clustering is given by

$$S = \frac{1}{c} \sum_{k=1}^{c} S_k. \tag{18}$$

The partition with highest S is taken to be optimal.

3.4. Salient features

The proposed shadowed c-means (*SCM*) clustering algorithm evaluates the centroids in a manner so as to minimize the influence of outliers as well as patterns having minimal typicality with the concept under consideration. Some of the characteristics of SCM, that make it unique as compared to the better-known FCM, RCM and RFCM algorithms, could be identified as follows:

Table 1
Cluster validity indices, for two clusters, on synthetic data $X32$.

Index	HCM	FCM	RCM	RFCM	SCM
Davies–Bouldin	0.6550	0.5973	0.6034	0.5973	**0.5611**
Xie–Beni	0.5210	0.4560	0.4572	0.4477	**0.3996**
Silhouette	−0.6087	−0.2233	−0.2411	−0.2384	−**0.2013**

1. For elements having MV above the calculated threshold, the algorithm does not attempt fuzzification as in FCM. This philosophy provides a much stronger bias to core members and prevents the most-likelihood estimate for a given cluster from drifting away. This is expected to reduce the computational burden.
2. A comparison with the RCM and RFCM algorithms exhibits the absence of external user-defined parameters, such as w_{low}, w_{upper} and *threshold*, in SCM clustering. This completely eliminates the idea of tuning an algorithm before its actual execution. The removal of this initial trial and error factor makes SCM more robust as well as insensitive to the fluctuations in the incoming data.
3. The radical elevation and reduction of the MVs to 1 and 0, respectively, results in a marked *contrast enhancement* in the observability of the incoming data. During the clustering process, this helps in filtering out a lot of insignificant information in the initial stages of the iteration, thus focusing on the ambiguous boundary region and thereby gaining in terms of the quality of the results.

The proposed way of forming of shadowed sets dwells upon the original FCM method and provides a clear way of delineation between core, shadow and exclusion range. It is of significant help in the interpretation of clustering results. It also makes it clear as whether to proceed with further processing of data or just make decisions on the basis of the already reported results. Like any other mean-based partitive clustering approach (say, HCM and FCM), the results of SCM are dependent on the initial choice of means. Therefore, an average over several runs needs to be computed. Analogous to HCM and FCM, algorithm SCM has limitations with non-convex partitions. Moreover, when the clusters are well-separated and disjoint the HCM is likely to perform more efficiently.

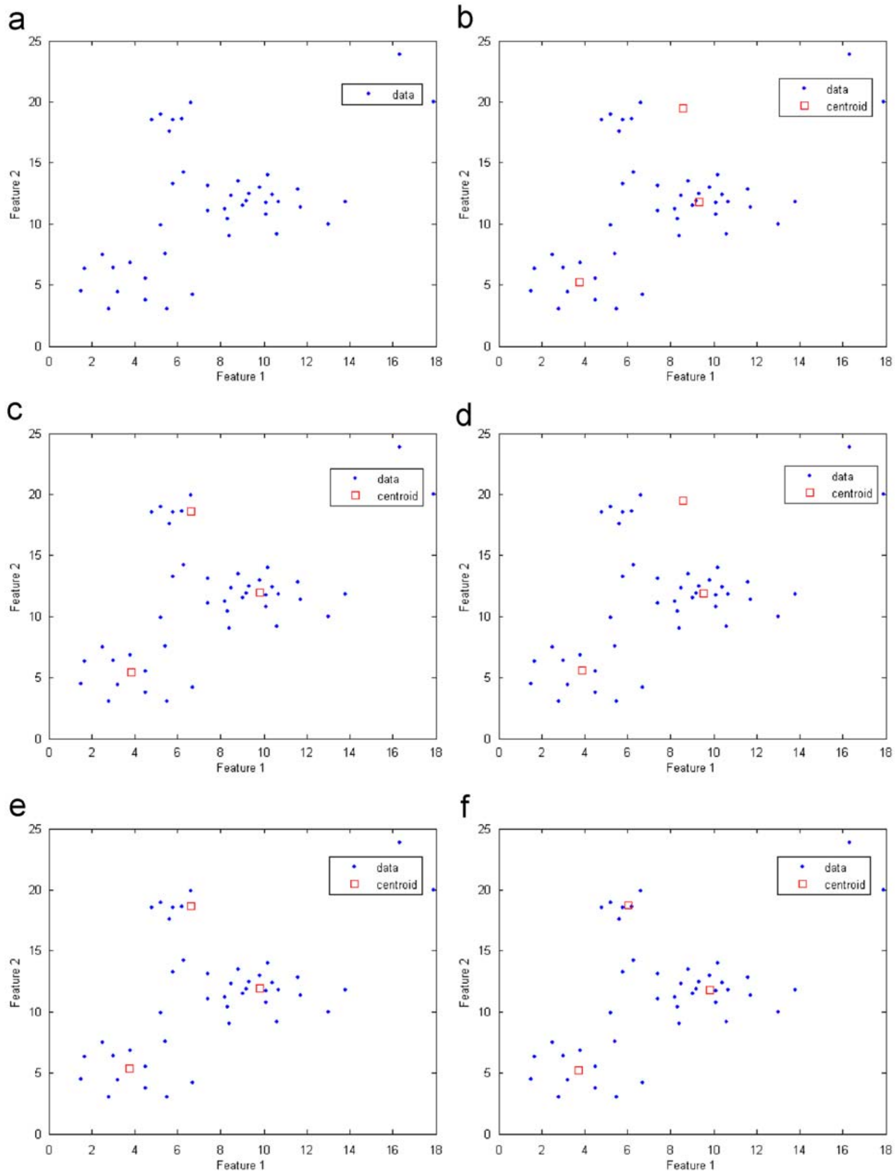

Fig. 6. Synthetic data $X45$: (a) original, and after clustering with; (b) HCM; (c) FCM; (d) RCM; (e) RFCM; and (f) SCM algorithms for $c=3$.

SCM, to some extent like the possibilistic approach to clustering [6], tries to minimize the influence of the position of the centroids of other clusters while calculating the mean $\mathbf{v_i}$ for a given cluster. It represents the extent to which a data point belongs to a particular cluster, and has an exponentially decreasing nature as the point drifts further away from the prototype.

There are several essential parameters of the clustering method; however, the choice of their values does not cause any substantial hesitation. As usual, the number of clusters is very much implied by the nature of the problem at hand. Here with the shadowed sets involved, one can anticipate that the upper bound of the number of clusters could be guided by the overlap criterion of the cores of the clusters, i.e., the number of clusters should not exceed a critical value at which any two of the cores of the clusters overlap. The fuzzification coefficient (m) could be optimized, however, it is common to assume a fixed value of 2.0. The choice of $m=2.0$ is associated with the form of the membership functions of the generated clusters.

In the following section we present a comparative study with HCM, FCM, RCM and RFCM to experimentally validate these claims.

Table 2
Cluster validity indices for synthetic data $X45$.

Index	c	HCM	FCM	RCM	RFCM	SCM
	2	0.4238	0.4154	0.3521	0.4097	0.4302
Davies	3	0.3343	0.3878	0.3418	0.3878	**0.2964**
Bouldin	4	0.5994	0.6829	0.4246	0.6268	0.3754
	5	0.4974	0.7662	0.4976	0.5409	0.6603
	2	0.1483	0.1730	0.1443	0.1730	0.1461
Xie	3	0.1153	0.1204	0.1230	0.1174	**0.1042**
Beni	4	0.1619	0.5340	0.5810	0.4202	0.4639
	5	0.4094	0.6860	0.5950	0.4912	0.4310
	2	−0.5888	−0.4902	−0.4908	−0.4902	−0.5523
Silhouette	3	−0.3812	−0.3666	−0.3897	−0.3812	−**0.3442**
Index	4	−0.8505	−0.7139	−0.7239	−0.6562	−0.6646
	5	−0.7791	−0.7252	−0.8364	−0.7950	−0.8097

4. Experimental results

The different clustering algorithms were implemented on four sets of synthetic and real data, for a detailed comparative analysis of the partitions—both qualitatively and quantitatively.

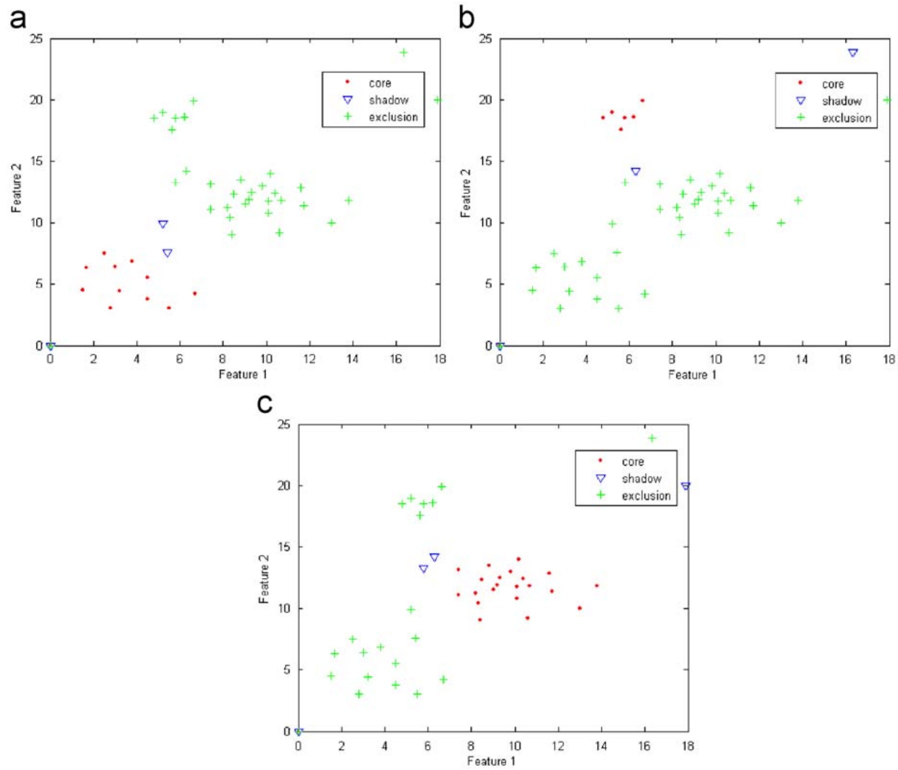

Fig. 7. Global view of exclusion and shadow regions, with the core being: (a) cluster 1, (b) cluster 2 and (c) cluster 3, using SCM on $X45$.

The two sets of synthetic data, viz., $X32$ and $X45$, consist of 32 and 45 points, and are depicted in Figs. 4(a) and 6(a), respectively.

The speech data *Vowel* [11] is a set of 871 vowel sounds from the Indian Telegu language, obtained by the utterance of three male speakers in the age group of 30–35 years, in a Consonant–Vowel–Consonant context. The three input features correspond to the first, second and third vowel format frequencies obtained through spectrum analysis of the speech data. Fig. 8(a) shows the six highly overlapped vowel classes ∂, a, i, u, e, o, marked with symbols 'star', 'plus', 'decagon', 'circle', 'upper triangle' and 'cross', respectively. The *Iris* data from the UCI Machine Learning Repository[1] is a four-dimensional data set containing 50 samples, each of three types of Iris flower. The results, along with the cluster validity indices, are presented here.

4.1. Synthetic data

There are two clusters in $X32$, with three outlier patterns as marked on the upper part of the scatter plots in Fig. 4. These have been purposely inserted in order to test the ability of the algorithms to resist any bias during the estimation of cluster prototypes. The centroids are marked by rectangles on the figure.

It can be easily seen that noise or outlier has maximum effect on HCM [Fig. 4(b)] and RCM [Fig. 4(d)], while FCM [Fig. 4(c)] and RFCM [Fig. 4(e)] show reasonable improvement. Not surprisingly, SCM [Fig. 4(f)] generates the best estimation of the centroids in this scenario. Fig. 5 exhibits the detailed mechanism through which the quantification of the membership values result in core, shadowed and exclusion regions in the proposed clustering algorithm. Here each cluster is considered, in turn, as a core, and its corresponding regions of exclusion and shadow are marked. It should be noted that in both the cases, elements belonging to one cluster have been readily labeled as exclusion members for the other cluster, even though they were closer as compared to the shadowed region members (located at the top of the figure). Unquestionably, the shadowed label is established only upon elements which possess sufficient ambiguity to prevent their belonging to any exclusive partition, irrespective of their proximity to the prototype.

All the cluster validity indices concurred to generate two optimal partitions for data $X32$. Table 1 provides the three index values for the different clustering algorithms. It is observed that the last column, corresponding to SCM, provides the best results—viz. minimum *DB* and χ, and maximum *S*.

The second set $X45$, of Fig. 6(a), has basically three clusters as far as visual perception is concerned. Although the clusters are not as well separated as in $X32$, yet the results follow a similar behavior. Table 2 demonstrates that all the validity indices concurred in generating three optimal partitions. Moreover, the entries in bold indicate the overall best performance for SCM partitions (smallest *DB* and χ, and largest *S*).

Thereafter, we used $c=3$ for comparing the partitioning results in case of the different algorithms. FCM [Fig. 6(c)] and RFCM [Fig. 6(e)] showed marked quality improvement in handling the overlap of the boundary between the clusters, as compared to HCM [Fig. 6(b)] and RCM [Fig. 6(d)]. However, all of them were afflicted by the outlier located towards the top right side of the scatter plots. It is interesting to observe from Fig. 6(f), that SCM prevented the centroid of the upper cluster from drifting towards the noise. The core, shadowed and exclusion regions are also plotted in Fig. 7, for the three clusters. Considering any of the clusters as the core, in turn, leads to an automatic clubbing of the other two into its exclusion zone. The shadow region encompasses the most ambiguous patterns.

[1] http://www.ics.uci.edu/~mlearn

4.2. Real data

The boundaries portrayed in the scatter plot of *Vowel*, as observed from Fig. 8(a), are quite fuzzy. The partitioning produced by the algorithms HCM, FCM, RCM and RFCM are depicted in parts (b)–(e) of the figure, while part (f) corresponds to the proposed SCM. The validity indices in Table 3 demonstrate the best results with SCM for $c=6$. This corresponds to the actual number of vowel categories under consideration. For example, in case of HCM, FCM and RCM, we observe that *DB* is indicative of incorrect optimization at five partitions. On the other hand, SCM provides better modeling of the uncertainty in the overlapped data in all cases.

Table 4 presents the results for *Iris* data. One of the three clusters in this data is well separated from the rest, while the remaining two clusters are overlapped. All three indices concur in indicating three optimal partitions for SCM, thereby tallying with the actual number of flower types.

5. Discussion and conclusions

A novel partitive shadowed clustering technique has been described in the hierarchy of the general *c*-means algorithms. The main objective was to reduce the effect of external parameters on the performance of the algorithm. Refinements to the existing fuzzy and rough clustering algorithms in the framework of shadowed sets were proposed, culminating in the shadowed *c*-means algorithm. Faster convergence and better management of uncertainty resulted in a more realistic modeling of the data. SCM demonstrated that it is possible to filter out irrelevant information while remaining in the fuzzy framework, thereby resulting in an unbiased estimation of the prototypes. The performance of SCM was also found to be robust in the presence of outliers.

It is worth noting that the advantage of SCM comes with the interpretation abilities offered by this clustering technique. The three-valued quantification of the resulting structure of clusters helps us easily identify regions (and patterns) which may require further attention while pointing at the core structure and patterns that arise with high values of typicality with respect to the detected clusters.

Future work aims to focus on the mathematical basis of the existence of a minima in the optimization process of the threshold and on the possible existence of two or more extremum points. This work along with the proof of convergence of the algorithm is being currently investigated.

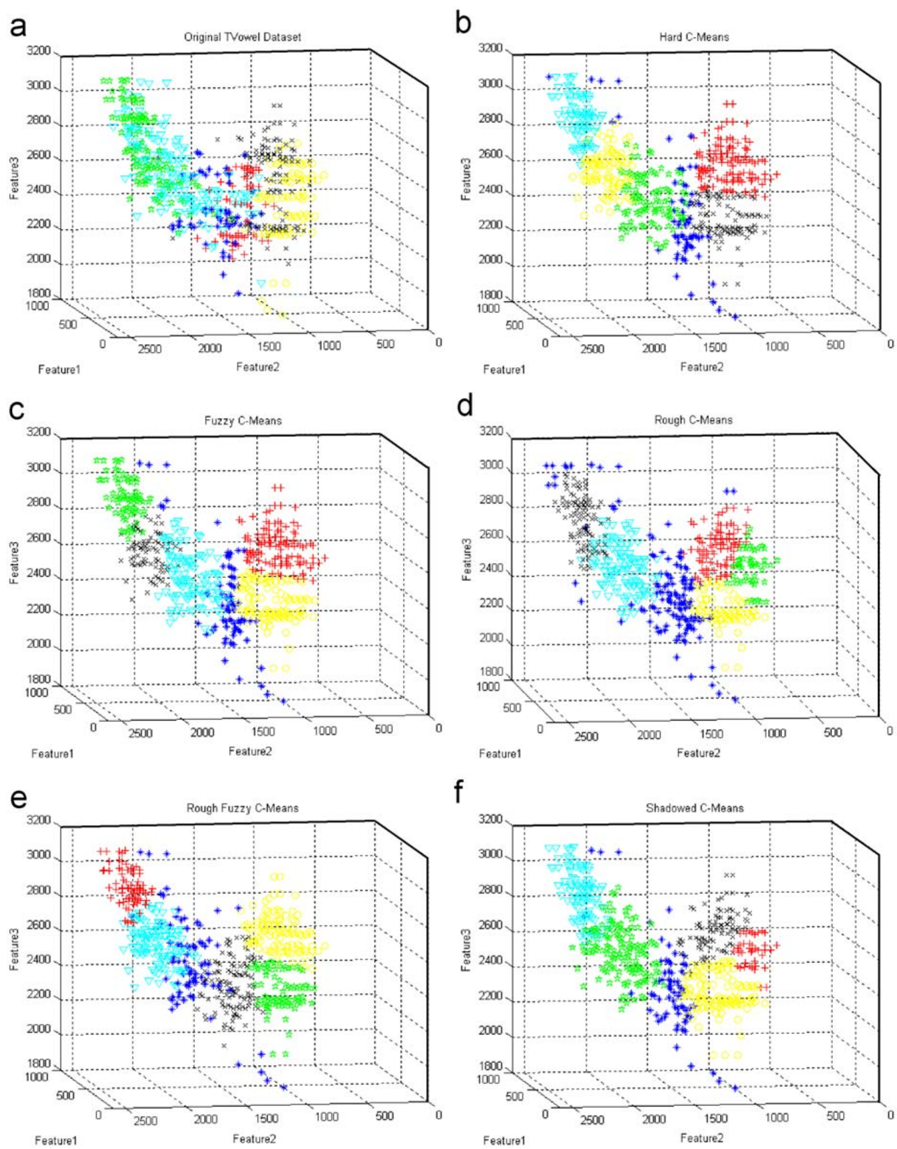

Fig. 8. Speech data *Vowel*: (a) original, and after clustering with; (b) HCM; (c) FCM; (d) RCM; (e) RFCM; and (f) SCM algorithms for $c=6$.

Table 3
Cluster validity indices on speech data, *Vowel*.

Index	c	HCM	FCM	RCM	RFCM	SCM
Davies	5	0.0659	0.0696	0.0659	0.0721	0.0676
Bouldin	6	0.0724	0.0744	0.0703	0.0673	**0.0650**
	7	0.0862	0.0799	0.0764	0.0744	0.0691
Xie	5	0.2074	0.2378	0.2503	0.2142	0.3268
Beni	6	0.1665	0.1893	0.1795	0.1625	**0.1496**
	7	0.1774	0.1913	0.2049	0.1598	0.3466
Silhouette	5	−0.8438	−0.8507	−0.8523	−0.8462	−0.8491
Index	6	−0.8541	−0.8410	−0.8467	−0.8157	−**0.7574**
	7	−0.8867	−0.8851	−0.8661	−0.8932	−0.8214

Table 4
Cluster validity indices on *Iris* data.

Index	c	HCM	FCM	RCM	RFCM	SCM
Davies	2	0.4501	0.4451	0.4335	0.4451	0.4374
Bouldin	3	0.7880	0.8198	0.7892	0.4129	**0.3781**
Xie	2	0.0559	0.0542	0.0522	0.0542	0.0530
Beni	3	0.1256	0.1371	0.1261	0.0345	**0.0222**
Silhouette	2	−0.1966	−0.3201	−0.1966	−0.1960	−0.3201
Index	3	−0.6180	−0.6155	−0.6180	−0.5902	−**0.2680**

Acknowledgments

Authors are grateful to Dr. P. Maji for his helpful discussions. The second author would like to thank the Indian National Academy of Engineering (INAE) for a fellowship to carry out this work at the Center for Soft Computing Research, Indian Statistical Institute, Kolkata, India.

References

[1] J.C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, New York, 1981.
[2] J.C. Bezdek, N.R. Pal, Some new indexes for cluster validity, IEEE Transactions on Systems, Man, and Cybernetics—Part B 28 (1998) 301–315.
[3] G. Cattaneo, D. Ciucci, An algebraic approach to shadowed sets, Electronic Notes in Theoretical Computer Science 82 (2003) 64–75.
[4] Z. Huang, Extensions to the *k*-means algorithm for clustering large data sets with categorical values, Data Mining and Knowledge Discovery 2 (1998) 283–304.
[5] L. Kaufman, P.J. Rousseeuw, Finding Groups in Data: An Introduction to Cluster Analysis, Wiley, New York, 1990.
[6] R. Krishnapuram, J.M. Keller, A possibilistic approach to clustering, IEEE Transactions on Fuzzy Systems 1 (1993) 98–110.
[7] P. Lingras, C. West, Interval set clustering of Web users with rough *k*-means, Journal of Intelligent Information Systems 23 (2004) 5–16.
[8] P. Maji, S.K. Pal, RFCM: a hybrid clustering algorithm using rough and fuzzy sets, Fundamenta Informaticae 79 (2007) 1–22.
[9] S. Mitra, T. Acharya, Data Mining: Multimedia, Soft Computing, and Bioinformatics, Wiley, New York, 2003.
[10] S. Mitra, H. Banka, W. Pedrycz, Rough–fuzzy collaborative clustering, IEEE Transactions on Systems, Man, and Cybernetics—Part B 36 (2006) 795–805.
[11] S.K. Pal, D. Dutta Majumder, Fuzzy sets and decision making approaches in vowel and speaker recognition, IEEE Transactions on Systems, Man, and Cybernetics 7 (1977) 625–629.
[12] Z. Pawlak, Rough Sets, Theoretical Aspects of Reasoning about Data, Kluwer Academic, Dordrecht, 1991.
[13] W. Pedrycz, Shadowed sets: representing and processing fuzzy sets, IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics 28 (1998) 103–109.
[14] W. Pedrycz, Interpretation of clusters in the framework of shadowed sets, Pattern Recognition Letters 26 (2005) 2439–2449.
[15] G. Peters, Some refinements of rough *k*-means clustering, Pattern Recognition 39 (2006) 1481–1491.
[16] N. Rescher, Many-Valued Logic, McGraw-Hill, New York, 1969.
[17] P. Rousseeuw, Silhouette: a graphical aid to the interpretation and validation of cluster analysis, Journal of Computational and Applied Mathematics 20 (1987) 53–65.
[18] J.T. Tou, R.C. Gonzalez, Pattern Recognition Principles, Addison-Wesley, London, 1974.
[19] X.L. Xie, G. Beni, A validity measure for fuzzy clustering, IEEE Transactions on Pattern Analysis and Machine Intelligence 13 (1991) 841–847.
[20] L.A. Zadeh, Fuzzy logic, neural networks, and soft computing, Communications of the ACM 37 (1994) 77–84.

About the Author—SUSHMITA MITRA is a Professor at the Machine Intelligence Unit, Indian Statistical Institute, Kolkata. From 1992 to 1994 she was in the RWTH, Aachen, Germany as a DAAD Fellow. She was a Visiting Professor in the Computer Science Departments of the University of Alberta, Edmonton, Canada in 2004, 2007, Meiji University, Japan in 1999, 2004, 2005, 2007, and Aalborg University Esbjerg, Denmark in 2002 and 2003. Dr. Mitra received the National Talent Search Scholarship (1978–1983) from NCERT, India, the University Gold Medal from University of Calcutta in 1988, the IEEE TNN Outstanding Paper Award in 1994 for her pioneering work in neuro-fuzzy computing, and the CIMPA-INRIA-UNESCO Fellowship in 1996.

She is an author of books like 'Neuro-Fuzzy Pattern Recognition: Methods in Soft Computing' (1999) and 'Data Mining: Multimedia, Soft Computing, and Bioinformatics' (2003) [John Wiley], and 'Introduction to Machine Learning and Bioinformatics' (2008) [CRC Press], beside four edited books. Dr. Mitra has guest edited special issues of several journals, is an Associate Editor of 'Neurocomputing' and a Founding Associate Editor of 'Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery (WIRE DMKD)'. She has more than 100 research publications in referred international journals. According to the Science Citation Index (SCI), two of her papers have been ranked 3rd and 15th in the list of Top-cited papers in Engineering Science from India during 1992–2001.

Dr. Mitra is a Senior Member of IEEE, and Fellow of Indian Academy of Engineering and The National Academy of Sciences. She served in the capacity of Program Chair, Tutorial Chair, and as member of program committees of many international conferences. Her current research interests include data mining, pattern recognition, soft computing, image processing, and Bioinformatics.

About the Author—WITOLD PEDRYCZ is a Professor and Canada Research Chair (CRC) in the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Canada. He is also with the Systems Research Institute of the Polish Academy of Sciences.

He is actively pursuing research in Computational Intelligence, fuzzy modeling, knowledge discovery and data mining, fuzzy control including fuzzy controllers, pattern recognition, knowledge-based neural networks, relational computation, bioinformatics, and Software Engineering. He has published numerous papers in this area. He is also an author of nine research monographs covering various aspects of Computational Intelligence and Software Engineering.

Witold Pedrycz has been a member of numerous program committees of conferences in the area of fuzzy sets and neurocomputing. He currently serves as an Associate Editor of IEEE Transactions on Systems, Man and Cybernetics, IEEE Transactions on Neural Networks, and IEEE Transactions on Fuzzy Systems. Dr. Pedrycz is an Editor-in-Chief of Information Sciences. He is a President of IFSA and NAFIPS. Dr. Pedrycz is a Fellow of the IEEE.

About the Author—BISHAL BARMAN received his B. Tech. degree in Computer Science from National Institute of Technology, Surat, India, in 2008. He is currently pursuing his MS. in University of Texas at Austin, USA.