Received March 19, 2018, accepted April 26, 2018, date of publication May 8, 2018, date of current version June 5, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2834348

Rough-Fuzzy Clustering Based on Two-Stage Three-Way Approximations

JIE ZHOU[1,2], ZHIHUI LAI[1,2], CAN GAO[1,2], XIAODONG YUE[3], AND WAIKEUNG WONG[2]

[1]College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China
[2]Institute of Textiles and Clothing, The Hong Kong Polytechnic University, Hong Kong
[3]School of Computer Engineering and Science, Shanghai University, Shanghai 200444, China

Corresponding author: Can Gao (2005gaocan@163.com)

This work was supported in part by the Postdoctoral Science Foundation of China under Grant 2017M612736 and Grant 2017T100645, in part by the Guangdong Natural Science Foundation with the titles, The Study on Knowledge Discovery and Uncertain Reasoning in Multi-Valued Decisions and Rough Sets-Based Knowledge Discovery for Hybrid Labeled Data, in part by the National Natural Science Foundation of China under Grant 61703283, Grant 61573248, Grant 61672358, and Grant 61773328, and in part by the Research Grant of The Hong Kong Polytechnic University under Grant G-YBD9.

ABSTRACT A general framework of rough-fuzzy clustering based on two-stage three-way approximations is presented in this paper. The proposed framework can deal with the uncertainties caused by the membership degree distributions of patterns. In the first stage (macro aspect), three-way approximations with respect to a fixed cluster can be formed from the global observation on data which can capture the data topology well about this cluster. In the second stage (micro aspect), the fuzziness of individual patterns over all clusters can be measured with De Luca and Termini's method, based on which three-way approximations with respect to the whole data set can be generated such that the uncertainties of the locations of individual patterns can be detected. By integrating the approximation region partitions obtained in the two stages, i.e., using the partition results obtained in the second stage to modify the partition results obtained in the first stage, the misled prototype calculations can be verified and the obtained prototypes tend to their natural positions. Comparative experiments on a synthetic data set and some benchmark data sets demonstrate the improved performance of the proposed method.

INDEX TERMS Rough sets, rough-fuzzy clustering, three-way approximations, fuzziness, shadowed sets.

I. INTRODUCTION

Data daily clustering is an unsupervised learning technique. It aims at partitioning a given data set with unlabeled patterns $D = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N\}$, $\mathbf{x}_j \in \Re^M (j = 1, 2, \cdots, N)$, into homogeneous clusters $\{G_1, G_2, \cdots, G_C\}$, such that the patterns in the same cluster will have the highest similarities, and the patterns between different clusters will have the highest dissimilarities. Clustering techniques have been successfully used in many fields, such as text mining, image segmentation, fault diagnosis, power allocation, wireless sensor networks and so on [1]–[4]. The revealed data structures by clustering methods can help human to recognize data and mine knowledge. One of the challenges of clustering techniques is how to deal with the uncertain information implicated in the data sufficiently since there is no prior knowledge beforehand, such as the distributions, densities or topologies of data. In addition, the parameters involved in the clustering models may produce new uncertainties since they affect the results directly, and thus it is challenging to determine the optimal parameters generally.

As one of the famous partitive clustering techniques, fuzzy C-means (FCM) [5] utilizes a partition matrix $U = \{u_{ij}\}$ to reveal the membership degrees of patterns $\{\mathbf{x}_j\}$ belonging to clusters $\{G_i\}$, such that the overlapping structures can be depicted quantitatively. Though the uncertainties arising in the overlapping areas can be well described, the validity of FCM will decrease when dealing with noisy environments. Since the relative distances are involved to compute membership degrees, a distant pattern may belong to a cluster with a higher membership degree, and this pattern will contaminate the prototype calculations [6].

Rough set theory proposed by Pawlak [7], an important methodology for analyzing uncertain, incomplete, imprecision information, has been developed rapidly in the last two decades. The target concept (especially uncertain or fuzzy) can be approximated by a pair of crisp sets, namely the lower

and upper approximations, from the two sides of the concept. By integrating the merits of rough sets, Lingras and West [8] proposed a rough C-means (RCM) clustering method in which all patterns are divided into three approximation regions according to their individual absolute distances, and the prototype calculations are only related to the patterns in the core and boundary regions of this cluster, instead of the whole patterns like in the FCM. Mitra *et al.* [9] further proposed a rough-fuzzy C-means clustering (RFCM) approach, in which individual membership degrees are used instead of individual absolute distances when determining the approximation regions. The key problem in the RCM or RFCM is how to select the separation threshold for partitioning approximation regions. Unreasonable partition thresholds will result in undesired results.

The separation threshold involved in RCM or RFCM is often selected depending on subjective tuning in the available studies [10], [11]. Maji and Pal [12] and Sarkar *et al.* [13] chose this value as the average value and the median of the difference between the highest and second highest membership degrees of all the patterns, respectively. However, the same threshold is adopted for all clusters although the sizes and the densities of clusters may be discrepant. The concept of shadowed sets [14] is an example of three-way, three-valued, or three-region approximations of a fuzzy set and becomes a paradigm of granular computing [15]. It provides an automatic optimization mechanism to determine the separation threshold, and the corresponding fuzzy set can be simplified to three values, namely, 0, 1, and shadows, which can reduce the burden of numeric computations. By incorporating with shadowed sets, Zhou *et al.* [16] introduced a shadowed set-based rough fuzzy clustering approach (SRFCM), in which the partition threshold can be selected automatically, and the three approximation regions of each cluster are obtained independently.

Three-way approximations of fuzzy sets can be interpreted in terms of the positive, negative and boundary regions within the theory of three-way decisions proposed by Yao [17]–[19]. A fundamental issue of three-way approximation of fuzzy sets is the interpretation and determination of separation thresholds. The Pedrycz's optimization model (the principle of uncertainty invariance) [20] is only a specific case to determine the threshold in which the formed shadows need to be interpreted with good semantics. Some other optimization principles to determine the partition thresholds can be specified from several aspects, such as the principle of retaining the total amount of fuzziness of the fuzzy set [21], the principle of minimum distance (including semantic distance) [22] and the principle of least cost [22], [23]. The selected principle should consider the characteristics of practical applications which can be analyzed based on the methodology of three-way decisions.

Under the framework of FCM, the membership degrees of patterns with respect to a fixed cluster $\{u_{ij}\}$ $(j = 1, \cdots, N)$ can be considered as a fuzzy set, and then three-way approximation optimization methods can be used for this cluster independently, no matter what the membership values with respect to other clusters, namely, the fuzziness [21], [24] of individual patterns which can measure the underlying vagueness are not considered. In this case, some abnormal circumstances will happen inevitably, such as the core region of one cluster may have some patterns with higher fuzziness values. Thereafter, the corresponding prototype calculation may be distorted. Under this consideration, the patterns with higher fuzziness values should be wiped off from the core or boundary regions of clusters.

In this study, a rough-fuzzy clustering approach based on two-stage three-way approximations is presented. Specifically, how to partition the approximation regions of each cluster more precisely and guarantee the prototype calculations are not contaminated by uncertain patterns. The main contributions of this paper include: (1) to optimize the partition threshold for each cluster independently based on three-way approximation optimization mechanisms. In this stage, the data topology with respect to a fixed cluster can be captured well from the global observation on data; (2) to capture the fuzziness of each pattern based on individual information entropy, and then partition all patterns into three approximation regions according to their fuzziness values. In this stage, the uncertainties caused by pattern locations can be detected. The patterns with lower fuzziness will locate around the prototypes, and these patterns will be partitioned into the core region with respect to the whole data set D. On the contrary, the patterns with higher fuzziness will locate far away from the prototypes or in the middle of different clusters, and these patterns will be partitioned into the exclusive region with respect to the whole data set D. (3) to modify the partition results obtained in the first stage with respect to a fixed cluster G_i independently by integrating the partition results obtained in the second stage with respect to the whole data set D based on the fuzziness of individual patterns. As a result, the uncertainties implicated in the partition results in the first stage can be reduced and the prototypes can be computed more accurately.

The rest of this paper is organized as follows: the notion of fuzzy C-means as well as some rough set-based partitive clustering methods are briefly reviewed in Section II. The limitations of available shadowed set-based rough-fuzzy clustering are discussed in Section III. The fuzziness of individual patterns based on information entropy is measured in Section IV, based on which the three-way approximations with respect to the whole data set can be generated. In Section V, a new rough-fuzzy clustering approach based on two-stage three-way approximations is introduced. Comparative experimental results are presented in Section VI. Some conclusions are given in Section VII.

II. PRELIMINARIES

In this section, some partitive clustering algorithms will be reviewed, including fuzzy C-means (FCM) [5], rough C-means (RCM) [8], rough-fuzzy C-means (RFCM) [9] and shadowed set-based rough fuzzy C-means [16].

More detailed information about rough sets and shadowed sets can be found in [25]–[29].

A. FUZZY C-MEANS

Suppose a data set with N patterns $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N\}$, that are defined over M-dimensional feature space, i.e., $\mathbf{x}_j \in \Re^M$ $(j = 1, 2, \cdots, N)$, will be grouped into C clusters $G_1, G_2, \cdots, G_C$. The corresponding prototypes for each cluster are denoted as $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_C$, $\mathbf{v}_i \in \Re^M$ $(i = 1, 2, \cdots, C)$. In the fuzzy C-means method, the following objective function will be minimized:

$$J_{FCM}(U, V) = \sum_{i=1}^{C}\sum_{j=1}^{N} u_{ij}^m d_{ij}^2, \tag{1}$$

$u_{ij} \in [0, 1]$ for all i, j, and

$$0 < \sum_{j=1}^{N} u_{ij} \leq N \text{ for all } i = 1, 2, \cdots, C, \tag{2}$$

$$\sum_{i=1}^{C} u_{ij} = 1 \quad \text{for all } j = 1, 2, \cdots, N. \tag{3}$$

Where u_{ij} is a fuzzy membership value that measures the degree of pattern $\mathbf{x}_j$ belonging to the cluster G_i. m $(m > 1)$ denotes the fuzzifier in FCM which controls the shape of membership degrees. In other words, when the value of m is close to 1, it implies a Boolean nature of one cluster, i.e., the memberships are maximally hard. On the other hand, it will result in spike-like membership functions when the value increases. In this case, the memberships are maximally fuzzy, and only the patterns that are located around the cluster centers are assigned 1. d_{ij} denotes the absolute distance between the pattern $\mathbf{x}_j$ and the cluster with the prototype $\mathbf{v}_i$.

Approximate optimization of J_{FCM} by the FCM method is based on the iteration and satisfies the following necessary conditions:

$$u_{ij} = \frac{1}{\sum_{k=1}^{C}\left(\frac{d_{ij}}{d_{kj}}\right)^{\frac{2}{(m-1)}}}, \quad i = 1, 2, \cdots, C \quad \text{and} \quad j = 1, 2, \cdots, N, \tag{4}$$

$$\mathbf{v}_i = \frac{\sum_{j=1}^{N} u_{ij}^m \mathbf{x}_j}{\sum_{j=1}^{N} u_{ij}^m}, \quad i = 1, 2, \cdots, C. \tag{5}$$

If $d_{ij} = 0$, then set $u_{ij} = 1$ and $u_{kj} = 0$ for $\forall k \neq i$. Although FCM is a very useful clustering approach, it is sensitive to noisy environments. The noises or outliers may have higher membership degrees since the relative distances are involved in the constraint (3).

B. ROUGH C-MEANS

By extending the notion of rough approximations, Lingras and West [8] developed a clustering algorithm, called rough C-means(RCM), in which all patterns are partitioned into three regions, namely, core region (lower approximation), boundary region and exclusion region for a fixed cluster. The new prototype calculations are only related to the core and boundary regions, not related to the all data patterns as that of FCM. In this way, the influence on prototype calculations caused by irrelevant patterns with respect to a fixed cluster can be reduced.

Patterns in the core region of one cluster will belong to this cluster certainly and patterns in the boundary region will belong to this cluster possibly, i.e., with vagueness and uncertainty. The rest in the exclusion region will not belong to this cluster definitely. Generally, patterns in the core region are important while patterns in the boundary region have less contribution and patterns in the exclusion region almost have no contribution to update prototypes.

According to RCM, the prototypes $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_C$, $\mathbf{v}_i \in \Re^M$ are computed as follows:

$$\mathbf{v}_i = \begin{cases} w_l A_1 + w_b B_1 & \text{if } \underline{R}G_i \neq \emptyset \wedge R_b G_i \neq \emptyset \\ B_1 & \text{if } \underline{R}G_i = \emptyset \wedge R_b G_i \neq \emptyset \\ A_1 & \text{if } \underline{R}G_i \neq \emptyset \wedge R_b G_i = \emptyset. \end{cases} \tag{6}$$

Where $A_1 = \frac{\sum_{\mathbf{x}_j \in \underline{R}G_i} \mathbf{x}_j}{card(\underline{R}G_i)}$, $B_1 = \frac{\sum_{\mathbf{x}_j \in \underline{R}_b G_i} \mathbf{x}_j}{card(\underline{R}_b G_i)}$ can be considered as the contributions by the crisp core region and Boolean boundary region, respectively. $card(X)$ means the cardinality of set X. $R_b G_i = \bar{R}G_i - RG_i$ denotes the boundary region of cluster G_i, where RG_i and $\bar{R}G_i$ are the lower and upper approximations of cluster G_i with respect to feature set R, respectively. $w_l (0.5 < w_l \leq 1)$ and $w_b = 1 - w_l$ are the weighted values that measure the contributions of the core and boundary regions, respectively. According to the Equation (6), the prototype calculations in rough set-based clustering methods are only related to the patterns in the core and boundary regions of clusters, instead of the whole patterns like that in the FCM.

In order to determine the core region and boundary region of each cluster, Lingras *et al.* utilized the principles as follows:

If $d_{qj} - d_{pj} \leq \Delta$, then $\mathbf{x}_j \in \overline{R}G_p$ and $\mathbf{x}_j \in \overline{R}G_q$. In this case, pattern $\mathbf{x}_j$ cannot be partitioned into the core region of any clusters. Otherwise, $\mathbf{x}_j \in \underline{R}G_p$. d_{pj} is the minimum of $\mathbf{x}_j$ over all clusters and d_{qj} is next to the minimum.

Incorporating with fuzzy membership degrees, Mitra *et al.* [9] proposed the notion of rough-fuzzy C-means (RFCM), in which the absolute distance d_{ij} is replaced by a fuzzy membership degree u_{ij} when dividing patterns into approximation regions. This adjustment enhances the robustness of the clustering to deal with overlapping situations. The prototype calculations are correspondingly modified as follows:

$$\mathbf{v}_i = \begin{cases} w_l A_2 + w_b B_2 & \text{if } \underline{R}G_i \neq \emptyset \wedge R_b G_i \neq \emptyset \\ B_2 & \text{if } \underline{R}G_i = \emptyset \wedge R_b G_i \neq \emptyset \\ A_2 & \text{if } \underline{R}G_i \neq \emptyset \wedge R_b G_i = \emptyset. \end{cases} \tag{7}$$

Where $A_2 = \frac{\sum_{\mathbf{x}_j \in \underline{R}G_i} u_{ij}^m \mathbf{x}_j}{\sum_{\mathbf{x}_j \in \underline{R}G_i} u_{ij}^m}$, $B_2 = \frac{\sum_{\mathbf{x}_j \in \underline{R}_b G_i} u_{ij}^m \mathbf{x}_j}{\sum_{\mathbf{x}_j \in \underline{R}_b G_i} u_{ij}^m}$ can be considered as the contributions by the fuzzy core region and fuzzy boundary region, respectively. The weighted values w_l and w_b have the same meanings as that in RCM. u_{ij} denotes the membership degree of pattern $\mathbf{x}_j$ belonging to the cluster with prototype $\mathbf{v}_i$ which is computed as the same as in FCM.

To determine the approximation regions in the procedures of RFCM, the following principles are exploited. If $u_{pj} - u_{qj} \leq \Delta$, then $\mathbf{x}_j \in \bar{R}G_p$ and $\mathbf{x}_j \in \bar{R}G_q$. In this case, $\mathbf{x}_j$ cannot be divided into the core region of any clusters. Otherwise, $\mathbf{x}_j \in \underline{R}G_p$. u_{pj} is the maximum of $\mathbf{x}_j$ over all clusters and u_{qj} is next to the maximum.

The threshold Δ is crucial in RCM and RFCM which determines the approximation regions of each cluster directly. This value should be selected circumspectly. The smaller the threshold is, the more patterns will be divided into the core regions. In this case, some noise patterns or overlapping patterns may be divided into the core regions at the same time. On the other hand, the larger the threshold is, the more patterns will be divided into the boundary regions. In this case, the representative capabilities of the core regions may be lost. An unreasonable threshold will result in undesired approximation region partitions and then misled the prototype calculations.

C. SHADOWED SET-BASED ROUGH FUZZY C-MEANS

The partition threshold in RCM or RFCM is often selected depending on subjective tuning and kept as a constant for all clusters in all iterations. In this way, the cluster discrepancies cannot be revealed very well, especially for the clusters with different sizes and densities. By integrating shadowed set theory and rough set-based clustering approaches, an improved rough-fuzzy C-means method based on shadowed sets (SRFCM) is presented in [16], in which the determination of approximation regions for each cluster is transferred to an optimization process independently and can be detected automatically during the clustering processes. The principles for determining the approximation regions of each cluster based on shadowed sets can be described as follows.

From the above Steps 1 to 3, the approximation region partition threshold for each cluster is not user-defined beforehand. It can be determined automatically during the iteration procedures and can be optimized for each cluster independently. Thereafter, the prototype calculations can be executed with different forms, such as Equations (6) or (7). The above Steps 1 to 3 provide a framework to partition approximation regions for each cluster which can be adopted for rough set-based clustering approaches. However, shadowed sets are only the one way of three-way approximations of fuzzy sets which keeps the uncertainty invariance. Another optimization principles can also be formed for constructing three-way approximations of a fuzzy set, such as retaining the total amount of fuzziness of the fuzzy set, the principle of minimum distance (including semantic distance) and the principle of least cost. More detailed information can be found in [20]–[23].

Algorithm 1 Shadowed Set-Based Approximation Region Determination for Each Cluster

Step 1: Compute fuzzy partition matrix $\{u_{ij}\}$ using Equation (4);

Step 2: Generate optimal partition threshold α_i for each cluster G_i independently based on shadowed sets:

$$\begin{aligned}\alpha_i &= \min_{\alpha}(V_i) \\ &= \min_{\alpha}\left|\sum_{j:u_{ij}\leq\alpha} u_{ij} + \sum_{j:u_{ij}\geq\max_j(u_{ij})-\alpha}(1-u_{ij})\right. \\ &\quad \left. -card\left(\left\{\mathbf{x}_j|\alpha < u_{ij} < \max_j(u_{ij}) - \alpha\right\}\right)\right|. \quad (8)\end{aligned}$$

Step 3: According to α_i, determine the core region and the boundary region of each cluster G_i:

$$\begin{aligned}RG_i &= \left\{\mathbf{x}_j|u_{ij} \geq \max_j(u_{ij}) - \alpha_i\right\}, \\ R_bG_i &= \left\{\mathbf{x}_j|\alpha_i < u_{ij} < \max_j(u_{ij}) - \alpha_i\right\}. \quad (9)\end{aligned}$$

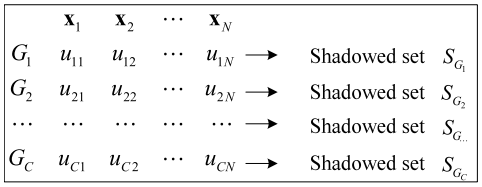

FIGURE 1. The schema of shadowed set construction in the rough-fuzzy clustering.

III. THE LIMITATION OF AVAILABLE SHADOWED SET-BASED ROUGH FUZZY C-MEANS

According to the schema of shadowed set-based rough fuzzy clustering, all patterns will be divided into three approximation regions with respect to a fixed cluster which can be depicted as follows.

In Figure 1, each row (with respect to a fixed cluster G_i) of the partition matrix $\{u_{ij}\}$ will be transferred to a shadowed set independently, and then all patterns will be divided into different approximation regions with respect to this fixed cluster. During this partition procedure, only the membership degrees of patterns belonging to this cluster are considered, the membership degrees of patterns belonging to the other clusters are not involved which may result in some unreasonable situations.

Given a synthetic data set D_{32} with two clusters and some noise data, as shown in Figure 2 and its details can be found in the Appendix.

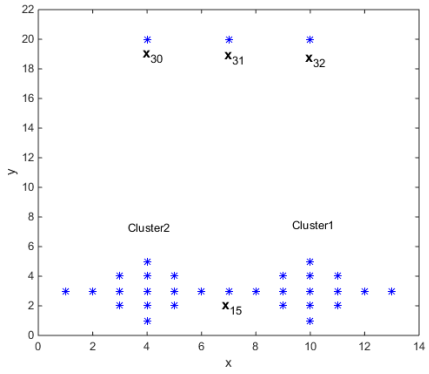

FIGURE 2. Synthetic data set D_{32}.

Intuitively, three patterns $\mathbf{x}_{30}$, $\mathbf{x}_{31}$ and $\mathbf{x}_{32}$ are far away from the cluster centers. The membership degree distributions and approximation region partitions based on shadowed sets with respect to Cluster1 and Cluster2 can be found in Figure 3.

In Figure 3 (a), pattern $\mathbf{x}_{32}$ has a relative higher membership value belonging to Cluster1 (the right cluster) than pattern $\mathbf{x}_{15}$ since the constraints (3) in FCM needs to be satisfied. According to shadowed set-based rough-fuzzy clustering method, pattern $\mathbf{x}_{32}$ will be divided into the core region of Cluster1 (as shown in Figure 3 (b)), which distorts the prototype calculation for Cluster1. Similarly, pattern $\mathbf{x}_{30}$ is divided into the core region of Cluster2 (as shown in Figure 3 (d)), which misleads the prototype calculation for Cluster2. These unreasonable partition situations occur because only the membership degrees with respect to a fixed cluster (each row in the fuzzy partition matrix) are considered independently. In this case, the fuzziness measures of individual patterns are not involved. In other words, the uncertainty of membership degree distribution with respect to a fixed pattern is not considered. If the membership degrees of one pattern over all clusters are equal (like $\mathbf{x}_{15}$ and $\mathbf{x}_{30}$), it means that the fuzziness of this pattern is the highest and this pattern cannot be divided into any clusters certainly. On the contrary, if the membership degrees of one pattern over all clusters are discrepant, the fuzziness of this pattern should be low due to the constraint (3) in FCM, and thus this pattern can be divided into some approximation regions of clusters definitely. Consequently, both the uncertainties caused by the membership degrees in the fuzzy partition matrix with respect to a fixed cluster (the row in the matrix) and a fixed pattern (the column in the matrix) need to be considered as partitioning approximation regions for each cluster.

IV. THE UNCERTAINTY MEASURE FOR THE FUZZINESS OF INDIVIDUAL PATTERNS BASED ON INFORMATION ENTROPY

Since the constraint (3) in FCM needs to be satisfied as computing the fuzzy partition matrix, the fuzziness of one pattern belonging to a specific cluster will be maximal if the membership degrees of this pattern belonging to all clusters are equal. Several measures of fuzziness have been proposed

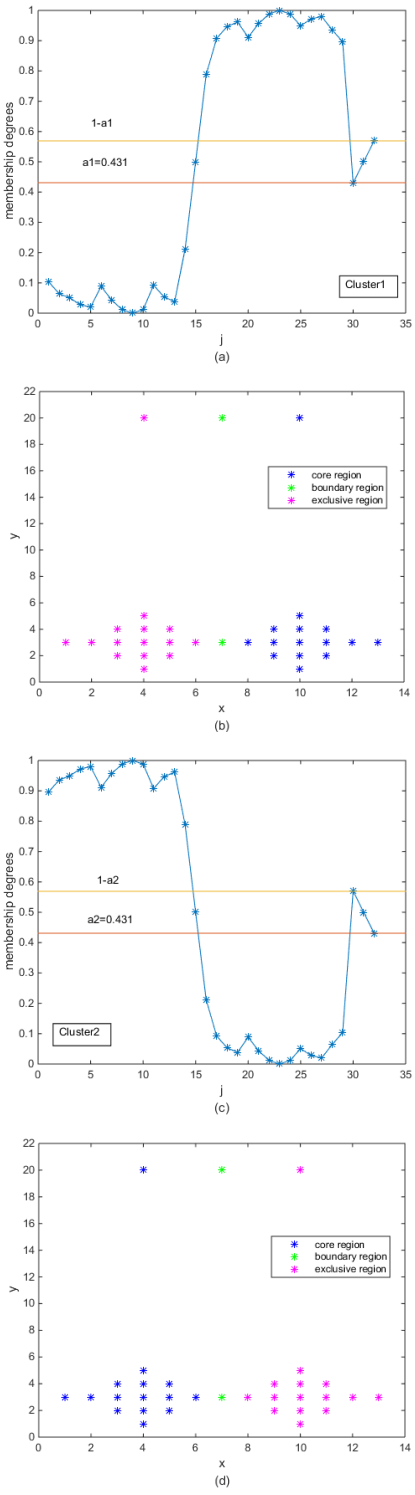

FIGURE 3. The approximation region partitions with respect to a fixed cluster. (a) The membership degrees and optimal separation value based on shadowed sets for Cluster1; (b) The approximation region partitions with respect to Cluster1; (c) The membership degrees and optimal separation value based on shadowed sets for Cluster2; (d) The approximation region partitions with respect to Cluster2.

in the literatures [30]–[32]. In this study, to measure the fuzziness of individual patterns in the clustering iterations, the following information entropy is utilized.

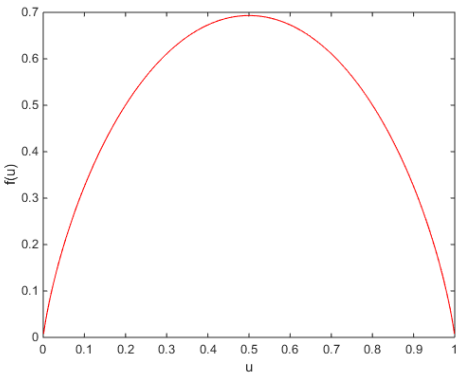

FIGURE 4. The fuzziness measure function.

For $\forall u \in [0, 1]$, De Luca and Termini's measure [33] of fuzziness is defined as :

$$\varphi(u) = \{(u, f(u))\}. \tag{10}$$

Where:

$$f(u) = -u\log(u) - (1-u)\log(1-u). \tag{11}$$

$f(u)$ is a Shannon's information entropy function and its values with respect to u can be described in Figure 4. When $u = 0$ or $u = 1$, the fuzziness of the corresponding fuzzy set equals zero, that means without fuzziness. When $u = 0.5$, the corresponding fuzzy set will have maximum fuzziness.

For each pattern in the fuzzy clustering, its fuzziness over all clusters can be defined with De Luca and Termini's measure according to its membership degrees over all clusters. Specifically, for a fixed pattern $\mathbf{x}_j$ $(j = 1, 2, \cdots, N)$, its fuzziness over all clusters can be formulated as follows:

$$\begin{aligned} IE_j &= \sum_{i=1}^{C} f(u_{ij}) \\ &= -\sum_{i=1}^{C} \left(u_{ij}\log(u_{ij}) + (1-u_{ij})\log(1-u_{ij})\right). \end{aligned} \tag{12}$$

The normalized fuzziness of individual patterns can be further defined as:

$$NIE_j = \frac{IE_j}{\max\limits_{j=1}^{N}\{IE_j\}}. \tag{13}$$

The smaller the value of NIE_j is, the lower the fuzziness of $\mathbf{x}_j$ will have. For a fixed pattern, if its membership degrees over all clusters approach the same value, it will have a maximal fuzziness value. In this case, it is difficult to partition this pattern to a specific cluster. If its membership degrees over all clusters are very different, according to the constraint (3) in FCM, its membership degree belonging to one of the clusters will be large, and the membership degrees belonging to other clusters will be small. In this case, the pattern has a small fuzziness value and can be divided into a specific cluster definitely.

$\{NIE_j\}$ can be considered as a series of uncertainty measure for the fuzziness of individual patterns. Based on the above discussions, the patterns with lower values of NIE_j (lower fuzziness) can be divided into a specific cluster to the greatest extent, and the patterns with higher values of NIE_j (higher fuzziness) cannot be divided into a specific cluster certainly. Under this circumstance, all patterns can be partitioned into three approximation regions according to $\{NIE_j\}$ based on three-way approximation optimization process, namely, patterns with lower fuzziness, patterns with higher fuzziness and the others. They are considered as the core region of the whole data set D (denoted as Pos_{NIE}), the exclusive region of the whole data set D(denoted as Neg_{NIE}), and the boundary region of the whole data set D (denoted as Bnd_{NIE}).

The optimal separation threshold for partitioning $\{NIE_j\}$ can be obtained as follows:

$$\begin{aligned} \alpha_{NIE} &= \min_{\alpha}(V_{NIE}) \\ &= \min_{\alpha}\Bigg|\sum_{j:NIE_j\le\alpha} NIE_j + \sum_{j:NIE_j\ge\max\limits_{j}(NIE_j)-\alpha}\left(1-NIE_j\right) \\ &\quad -card\left(\left\{\mathbf{x}_j|\alpha < NIE_j < \max_{j}\left(NIE_j\right)-\alpha\right\}\right)\Bigg|. \end{aligned} \tag{14}$$

In the Equation (14), Pedrycz's uncertainty invariance principle is utilized to optimize the separation threshold. The other three-way approximation optimization principles can also be used similarly.

According to the value of α_{NIE}, Pos_{NIE}, Bnd_{NIE} and Neg_{NIE} can be formed as follows:

$$Pos_{NIE} = \{\mathbf{x}_j|0 \le NIE_j \le \alpha_{NIE}\}, \tag{15}$$

$$Bnd_{NIE} = \{\mathbf{x}_j|\alpha_{NIE} < NIE_j < \max_{j}\left(NIE_j\right) - \alpha_{NIE}\}, \tag{16}$$

$$Neg_{NIE} = \{\mathbf{x}_j|\max_{j}\left(NIE_j\right) - \alpha_{NIE} \le NIE_j \le 1\}. \tag{17}$$

Where Pos_{NIE} is composed of the patterns with the smaller values of normalized fuzziness. It means that the membership degree distributions of these individual patterns are discrepant which is beneficial for classifying patterns. Neg_{NIE} is composed of the patterns with the larger values of normalized fuzziness. It means that the membership degrees of these patterns over all clusters approach the same.

Taking the synthetic data set D_{32} as an example, the optimal separation threshold is 0.432 according to the Equation (14) and the partitioned three approximation regions with respect to the whole data set is shown in Figure 5.

It can be found from the Figure 5 that three noise patterns and some patterns located in the middle of two clusters are divided into the exclusive region (Neg_{NIE}) with respect to the whole data set. These patterns cannot be partitioned into a specific cluster certainly since their membership degrees belonging to the two clusters approach 0.5. It is easy to observe that the partition results in Figure 5 (b) are different

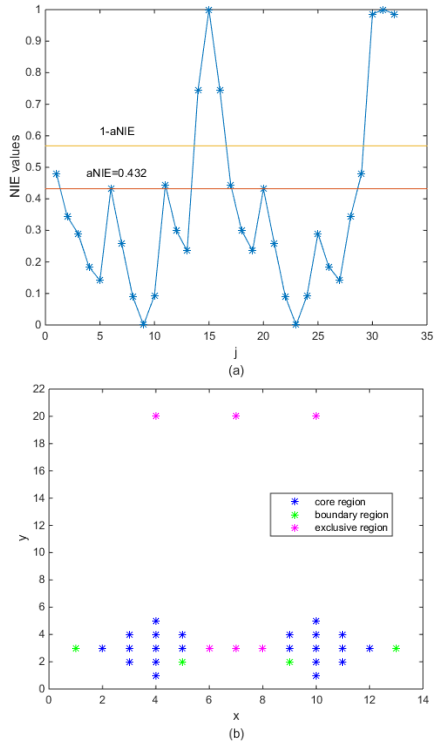

FIGURE 5. The optimal partition threshold for normalized fuzziness values and the corresponding approximation region partitions. (a) The normalized fuzziness values and the optimal separation value; (b) The approximation region partitions based on $\{NIE_j\}$.

with the results in Figure 3 (b) and (d). The partition results obtained based on the fuzziness of individual patterns $\{NIE_j\}$ can be used to adjust or modify the unreasonable partition results obtained for each cluster independently.

V. ROUGH-FUZZY C-MEANS BASED ON TWO-STAGE THREE-WAY APPROXIMATIONS

A. A GENERAL ROUGH-FUZZY C-MEANS CLUSTERING BASED ON TWO-STAGE THREE-WAY APPROXIMA-TIONS

The uncertainties caused by a specific pattern can be analyzed from two sides. One is from the view of three-way approximations with respect to a fixed cluster. The other is from the view of three-way approximations with respect to the whole data set based on the fuzziness of individual patterns. There are intersections between these two perspectives which can be formed as follows:

The item (1, 1) in Table 1 means a pattern belongs to the core region of G_i as well belongs to the Pos_{NIE}. The item $(1, 0)$ means a pattern belongs to the core region of G_i, but belongs to the Bnd_{NIE}. Item $(1, -1)$ means a pattern belongs to the core region of G_i, but belongs to the Neg_{NIE}. The other items in Table 1 can be interpreted similarly. Since the partition results obtained based on the fuzziness of individual patterns $\{NIE_j\}$ can be used to modify the unreasonable partition results with respect to a fixed cluster G_i, and the core region $\underline{R}G_i$ and boundary region R_bG_i are involved for computing the prototype of cluster G_i, only the items with grey shading

TABLE 1. The partitions of individual patterns under two kinds of three-way approximations.

		D		
		Pos_{NIE}	Bnd_{NIE}	Neg_{NIE}
G_i	$\underline{R}G_i$	(1, 1)	(1, 0)	(1, -1)
	R_bG_i	(0, 1)	(0, 0)	(0, -1)
	R_NG_i	(-1, 1)	(-1, 0)	(-1, -1)

in the Table 1 need to be reconsidered. The principles of approximation region modification can be drawn as follows:

P1: If a pattern is with the item $(1, 0)$ or $(1, -1)$, this pattern will be moved from $\underline{R}G_i$ to R_bG_i;

P2: If a pattern is with the item (0, -1), this pattern will be moved from R_bG_i to R_NG_i, and thus this pattern will not be involved for updating prototypes;

P3: The patterns with other items will not be changed.

According to the above principles, the core region and boundary region of each cluster will be adjusted. Some patterns in the core regions of clusters will be moved into the boundary regions of these clusters if they are divided into the boundary region or exclusive region with respect to the whole data set D at the same time. Some patterns in the boundary regions of clusters will be adjusted into the exclusive regions of these clusters if they are divided into the exclusive region with respect to the whole data set D. Consequently, the uncertain patterns partitioned into the core regions or boundary regions with respect to a fixed cluster can be wiped off and the prototypes can be computed more precisely.

Taking the Cluster1 in data set D_{32} as an example. The membership degrees of pattern $\mathbf{x}_{15}$ belonging to Cluster1 and Cluster2 are 0.5 and 0.5, and thus the pattern $\mathbf{x}_{15}$ will be divided into boundary region of Cluster1 according to the Equation (9). However, pattern $\mathbf{x}_{15}$ has the maximum fuzziness, and it is partitioned into *Neg_NIE* according to the Equation (17). Thus the pattern $\mathbf{x}_{15}$ is with the item $(0, -1)$. According to the principle **P2**, pattern $\mathbf{x}_{15}$ is moved from the boundary region of Cluster1 to the exclusive region of this cluster. Thereafter, pattern $\mathbf{x}_{15}$ has no contribution as updating the prototype of Cluster1. The partitions of patterns need to be modified with respect to Cluster1 can be found in Figure 6 (a), and the new approximation region partitions of Cluster1 are shown in Figure 6 (b). The patterns with blue and green circles in Figure 6 (a) are adjusted according to the principles **P1** and **P2**, respectively.

Three-way approximations of membership degrees with respect to a fixed cluster can capture the topology of data from the global observation on data. Three-way approximations of membership degrees based on the fuzziness of individual patterns can detect the uncertainties of pattern locations. The former and the later can be considered as the macro and micro analysis on data respectively. The micro analysis can be used to modify the results obtained by the macro analysis, which is described in Figure 7.

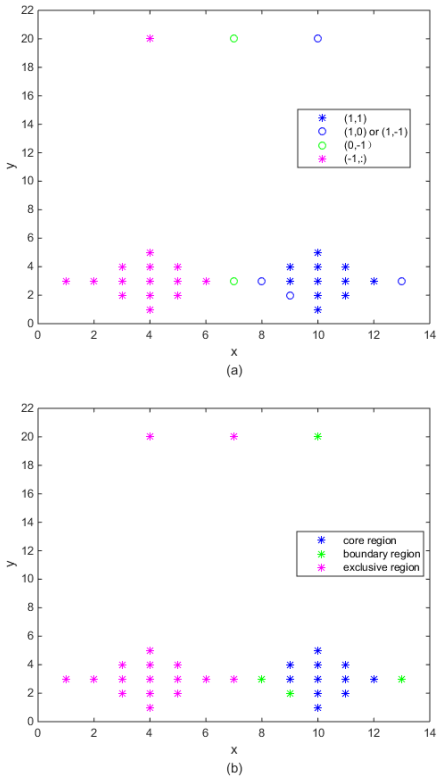

FIGURE 6. The combination of two kinds of approximation region partitions. (a) The partitions of patterns need to be modified with respect to Cluster1; (b) The new approximation region partitions of Cluster1.

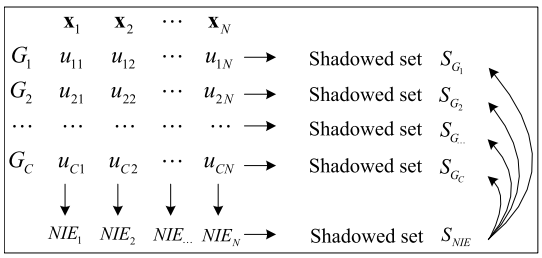

FIGURE 7. Two kinds of three-way approximations of membership degrees.

By integrating the results of two kinds of three-way approximations of membership degrees, a general rough-fuzzy C-means algorithm based on two-stage three-way approximations can be formed as follows:

Steps 2 to 3 can be considered as the first stage of three-way approximations of membership degrees with respect to a fixed cluster. Steps 4 to 5 can be considered as the second stage of three-way approximations of membership degrees with respect to the whole data set based on the fuzziness of individual patterns. Step 6 is the most important step in the algorithm in which the partition results obtained in the two stages are combined, thus the unreasonable partitions of each cluster in the first stage can be modified independently by integrating the partitions results produced in the second stage.

The prototypes obtained by FCM, RFCM, SRFCM and proposed algorithm ARFCM for D_{32} are shown in Table 2.

Algorithm 2 Rough-Fuzzy C-Means Based on Two-Stage Three-Way Approximations (ARFCM)

Step 1: Assign initial prototypes $\mathbf{v}_i (i = 1, 2, \cdots, C)$;
Step 2: Obtain the fuzzy membership degrees u_{ij} $(i = 1, 2, \cdots, Cj = 1, 2, \cdots, N)$;
Step 3: Compute the optimal partition threshold α_i for each cluster$G_i (i = 1, 2, \cdots, C)$, and according to α_i, determine the approximation regions RG_i and R_bG_i for each cluster G_i with respect to u_{ij};
Step 4: Compute the fuzziness of each pattern, i.e., NIE_j, with Equation (13);
Step 5: Compute the optimal partition threshold α_{NIE}, and according to α_{NIE}, determine the approximation regions Pos_{NIE},Bnd_{NIE} and Neg_{NIE};
Step 6: According to the principles **P1** to **P3**, modify RG_i and R_bG_i for each cluster;
Step 7: Update the prototypes $\mathbf{v}_i (i = 1, 2, \cdots, C)$ with Equation (7);
Step 8: Repeat Step 2 to Step 7 until convergence is reached.

TABLE 2. The prototypes obtained by different methods for D_{32}.

	FCM	RFCM	SRFCM	ARFCM
Cluster1	[10.0008, 3.9911]	[10.1145, 3.6665]	[10.0866, 3.8686]	[10.282, 3.1751]
Cluster2	[3.9992, 3.9911]	[3.8855, 3.6665]	[3.9134, 3.8686]	[3.718, 3.1751]
Deviation	1.9822	1.3525	1.7458	0.6639

From Table 2, it can be found that the prototypes obtained by the proposed algorithm ARFCM are better than the other methods. The deviation of the prototypes obtained by ARFCM is minimum. It means that the prototypes in the clustering iterations can be updated more precisely by using the framework of two-stage three-way approximations of membership degrees.

B. COMPLEXITY ANALYSIS

Assume the number of clusters is C, the number of patterns is N, the number of features of each pattern is M, the number of iterations is I, and the number of candidate threshold values is S. The asymptotical time complexity for computing partition matrix is $O\left(C^2NM\right)$ and the computation for selecting partition thresholds for all clusters in the first stage is $O\left(SCN\right)$. The computation for obtaining the fuzziness values of all patterns is $O\left(CN\right)$, and the computation for selecting the partition threshold based on the fuzziness of individual patterns is $O\left(SN\right)$. Subsequently, the computation for dividing approximation regions is $O\left(CN\right)$ and the computation for prototypes is also $O\left(CN\right)$. Thus, the computational complexity of the proposed method can be summarized as $O\left(I\left(C^2NM + SCN + 3CN + SN\right)\right)$. Generally, if $N \gg C$ and $N \gg I$, the asymptotical time complexity of our proposed method approaches to $O\left(NM + SN\right)$. Since no closed-form solution can be drawn for optimizing the partition

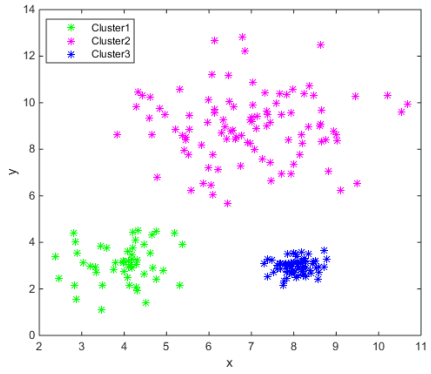

FIGURE 8. Synthetic data set D_{220}.

thresholds, the enumerating method is often exploited. For a practical problem with a big data set, if $N \gg S$ and $N \gg M$, the computational complexity becomes $O(N)$.

VI. EXPERIMENTS

In this section, a synthetic data set and some data sets from UCI repository [34] are used to compare the results obtained by FCM [5], original RFCM [9], SRFCM [16] and the proposed algorithm ARFCM.

The fuzzification coefficient is set as $m = 2$ for all methods which is commonly used in the literatures [9], [10], [12], [16]. The weighted value that measures the importance of core regions is set as $w_l = 0.95$ and kept as a constant for all data sets and all iterative runs. The maximum iteration number is set as 100 and the convergence condition satisfies $\left\| \mathbf{v}_i^{(t+1)} - \mathbf{v}_i^{(t)} \right\| < \varepsilon$ where t is an iterative step, ε is set as 0.001 for all algorithms and $\|\cdot\|$ means the Euclidean distance.

A. SYNTHETIC DATA SET

The synthetic data set D_{220} with a mixture of Gaussian distributions is shown in Figure 8. It has three clusters with 50, 100 and 70 data respectively. The means of three clusters are [4, 3], [7, 9], [8, 3] respectively. The standard deviations of three clusters are 0.5, 2 and 0.1 respectively. It can be found from Figure 8 that the sizes and the densities of three clusters are different. Intuitively, some patterns located in Cluster2 are close to the Cluster1 or Cluster3. It is a challenging work to partition these patterns into their natural groups.

The prototypes and partition labels obtained by the selected clustering methods are shown in Figure 9, and the deviations between the obtained prototypes and the natural means of clusters are given in Table 3.

It can be found that the methods based on the notion of three-way approximations of membership degrees, namely, SRFCM and ARFCM, performs better than FCM and original RFCM. It can be attributed to the approximation region partitions for each cluster which can capture the data topology from the global observation on data. Further, the prototypes obtained by the proposed method ARFCM is better than other available methods, in which only three patterns are classified into wrong clusters, and the deviation value is minimum. The reason is that the mechanism of two-stage three-way approximations in the proposed approach are developed.

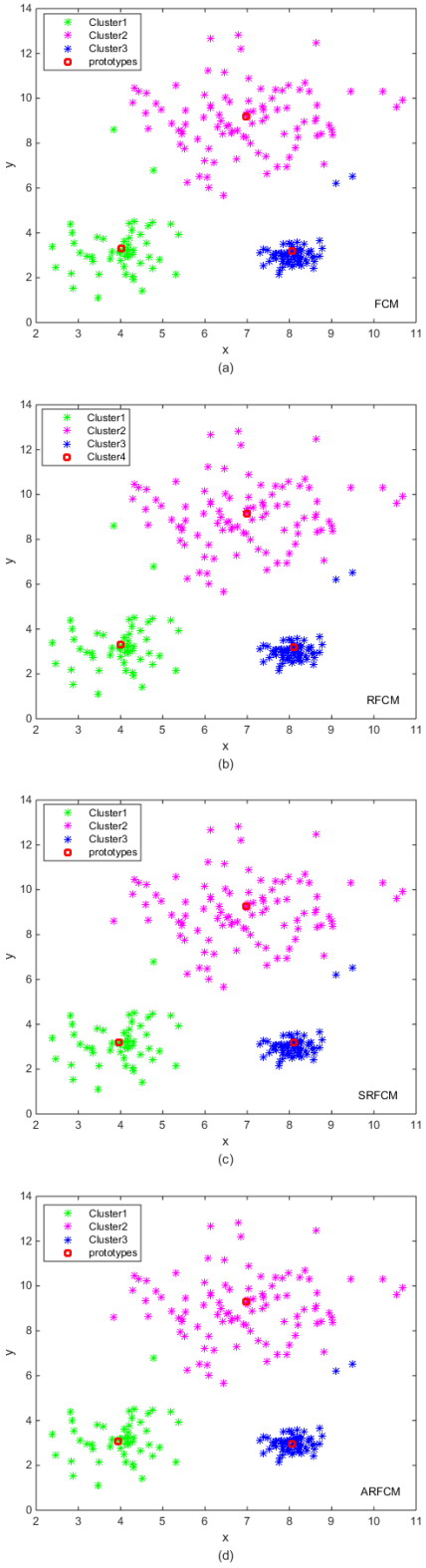

FIGURE 9. The prototypes and partition labels obtained by different clustering methods. (a)FCM; (b)RFCM; (c)SRFCM; (d)ARFCM.

TABLE 3. The prototypes obtained by different clustering methods.

	FCM	RFCM	SRFCM	ARFCM
Cluster1	[8.0552 3.1786]	[8.1012 3.1877]	[8.1025 3.1855]	[8.0643 2.9727]
Cluster2	[4.0205 3.3241]	[4.0032 3.3021]	[3.9595 3.1729]	[3.9332 3.0819]
Cluster3	[6.9795 9.2035]	[6.9867 9.1549]	[6.9732 9.2565]	[6.9656 9.2910]
Deviation	0.7162	0.6708	0.6474	0.4686

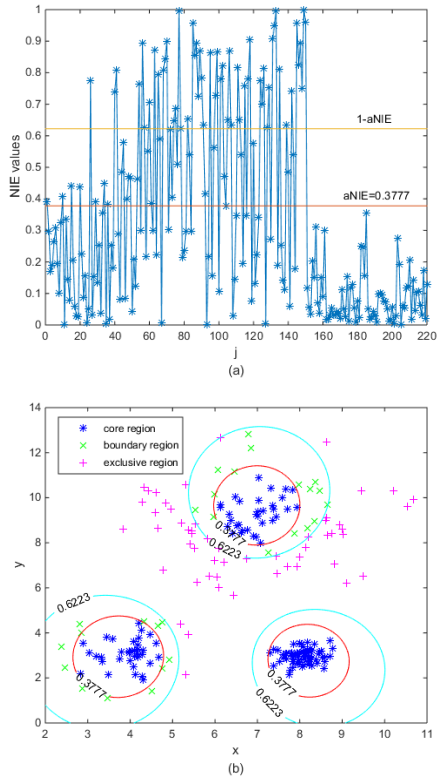

FIGURE 10. The optimal partition threshold based on normalized fuzziness values and corresponding approximation region partitions with respect to D_{220}. (a) The normalized fuzziness values and optimal separation value; (b) The approximation region partitions based on $\{NIE_j\}$.

The approximation region partitions with respect to a fixed cluster are modified by the results of approximation regions generated based on the fuzziness of individual patterns.

The results of approximation region partitions generated based on the fuzziness of individual patterns are shown in Figure 10. Obviously, the important parts of each cluster, i.e., the patterns around the cluster centers, are partitioned into the core region with respect to the whole data set at the same time. The representative capabilities of these patterns to the corresponding clusters are higher than the other patterns. Since the densities of Cluster3 is the highest, all of the patterns in the Cluster3 are divided into the core region although the size of Cluster3 is bigger than the Cluster1. On the contrary, many patterns in the Cluster2 are divided into the exclusive region with respect to the whole data set, since most of them are located in the middle of different clusters.

The approximation region partitions with respect to a fixed cluster can be modified by the results obtained based on the fuzziness of individual patterns, which are illustrated in Figure 11. The core region and boundary region of each cluster are adjusted based on the principles **P1** to **P3** according to the results in Figure 10 (b). Taking Cluster1 as an example, some patterns in Cluster2 are divided into the boundary region of Cluster1, as shown in Figure 11 (a), and these patterns are moved into the exclusive region of Cluster1 after using the modification principle **P2**, as the green circles in Figure 11 (b). In this way, the prototype calculation for Cluster1 cannot be misled by these patterns. Some pattern in Cluster1 that divided into the core region of Cluster1 in the first stage, as shown in Figure 11 (a), are changed into the boundary region of this cluster after using **P1**, as the blue circles in Figure 11 (b). The importance of these patterns will be reduced since these patterns are away from the cluster centers.

TABLE 4. The validity indices obtained by different methods for synthetic data set D_{220}.

	RS	PBM	DB	ACC	NMI	RI
FCM	0.37652	41.80175	0.41699	0.98182	0.92057	0.97393
RFCM	0.37678	41.73214	0.41482	0.98182	0.92057	0.97393
SRFCM	0.42253	45.32285	0.39128	**0.98636**	**0.93702**	**0.97995**
ARFCM	**0.43865**	**47.55655**	**0.38317**	**0.98636**	**0.93702**	**0.97995**

TABLE 5. The comparative validity results of wine.

	Wine					
	RS	PBM	DB	ACC	NMI	RI
FCM	0.14095	84931.64213	2.67309	0.96629	0.87589	0.95429
RFCM	0.28280	176970.24903	1.12176	0.96629	0.87589	0.95429
SRFCM	0.30592	**190916.77963**	1.02567	0.97191	0.89202	0.96134
ARFCM	**0.30664**	186704.12677	**0.87826**	**0.97753**	**0.91085**	**0.96851**

To compare the selected methods comprehensively, some clustering validity indices are used, including relative separation index (RS) [35], PBM-index (PBM) [36], Davies-Bouldin index (DB) [37], normalized mutual information (NMI) [38], [39], rand index (RI) [40] and classification accuracy (ACC) [38]. The smaller the value of DB is, or the higher the values of other validity indices are, the better the clustering methods will be. Since the initialization affects the clustering results directly, each method is executed for ten times and the average index values are generated for comparing. The obtained average values of validity indices for synthetic data set D_{220} are given in Table 4. It can be found that the proposed algorithm ARFCM performs the best over all validity indices.

B. UCI DATA SETS

Eight benchmark data sets from UCI storage [34] are selected for experiments, including Wine, Banknote, Glass, Seeds, Magic Gamma Telescope, Vertebral Column, Anuran Calls and Cryotherapy. The validity indices of experimental results are presented from Table 5 to Table 12.

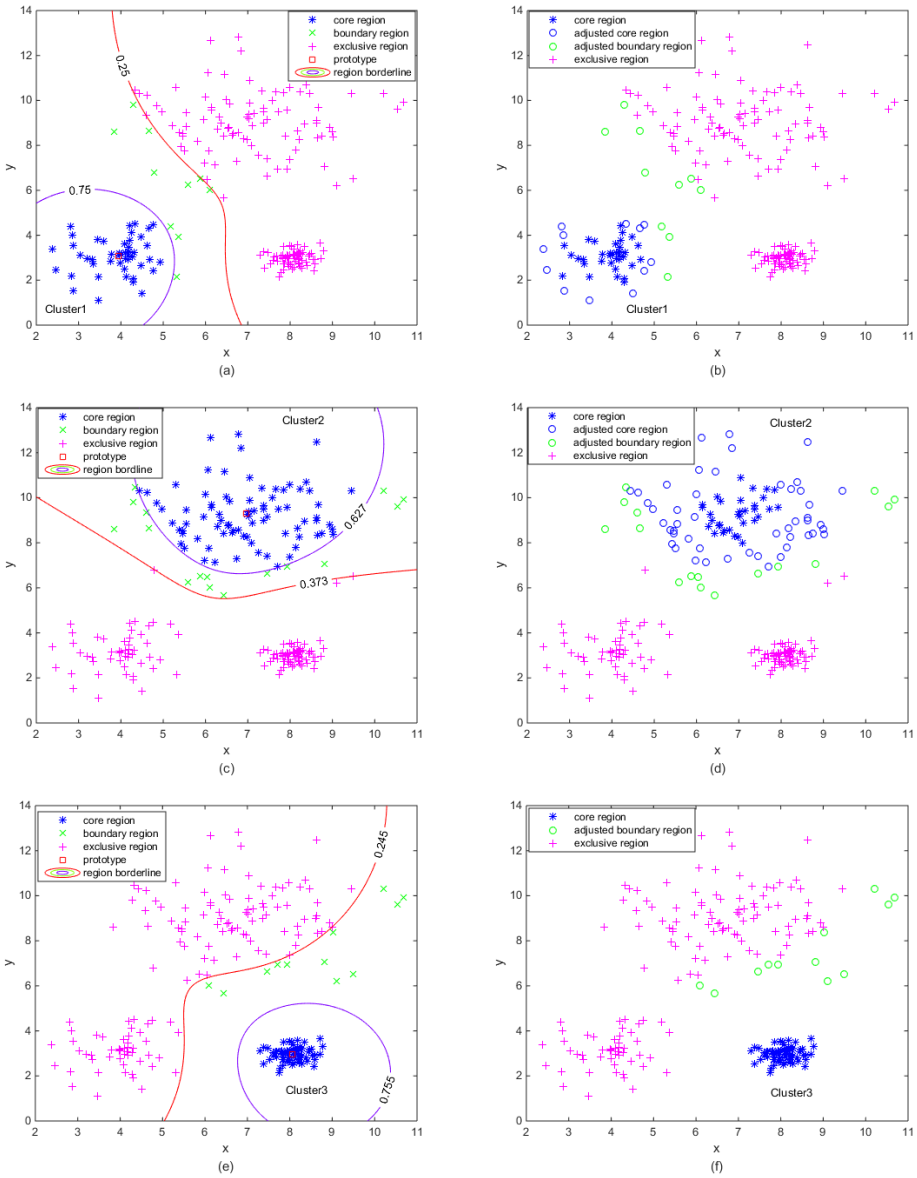

FIGURE 11. The original and modified approximation region partitions with respect to a fixed cluster. (a) The original results of Cluster1; (b) The modified results of Cluster1; (c) The original results of Cluster2; (d) The modified results of Cluster2; (e) The original results of Cluster3; (f) The modified results of Cluster3.

TABLE 6. The comparative validity results of banknote.

Banknote						
	RS	PBM	DB	ACC	NMI	RI
FCM	0.00447	47.24869	1.05044	0.58819	0.02291	0.51520
RFCM	0.00731	76.56448	0.83144	0.58455	0.01895	0.51394
SRFCM	0.00759	79.53363	0.81729	0.58601	0.01914	0.51444
ARFCM	**0.01128**	**120.82433**	**0.66649**	**0.65306**	**0.06243**	**0.54652**

TABLE 7. The comparative validity results of glass.

Glass						
	RS	PBM	DB	ACC	NMI	RI
FCM	0.010150	1.617090	2.009538	0.878505	0.400259	0.785529
RFCM	0.035328	6.136641	0.878090	**0.897196**	0.464818	**0.814664**
SRFCM	0.034990	5.998101	0.886057	**0.897196**	0.464818	**0.814664**
ARFCM	**0.049254**	**7.202634**	**0.759995**	**0.897196**	**0.494357**	**0.814664**

From Table 5 to Table 12, the proposed method ARFCM performs better than other rough-fuzzy clustering methods in terms of the most validity indices. It has the best performance of classification accuracy (ACC) over all selected data sets. The significant improvements of the proposed method can be attributed to the following technologies:

TABLE 8. The comparative validity results of seeds.

Seeds						
	RS	PBM	DB	ACC	NMI	RI
FCM	0.061453	23.597480	0.678368	**0.666667**	0.560473	0.737343
RFCM	0.070855	27.026439	0.627269	0.665873	**0.595031**	**0.748310**
SRFCM	0.074892	28.320327	0.610521	0.661905	0.576533	0.742994
ARFCM	**0.088127**	**37.089974**	**0.541385**	**0.666667**	0.550960	0.734062

TABLE 9. The comparative validity results of magic gamma telescope.

Magic Gamma Telescope						
	RS	PBM	DB	ACC	NMI	RI
FCM	5.15E-05	1779.67110	2.52904	0.52349	0.00026	0.50108
RFCM	1.06E-04	3721.65073	1.76852	0.52986	0.00036	0.50176
SRFCM	1.06E-04	3718.89774	1.76653	0.52823	0.00029	0.50157
ARFCM	**1.22E-04**	**3803.66153**	**1.61398**	**0.54012**	**0.00139**	**0.50319**

TABLE 10. The comparative validity results of vertebral column.

Vertebral Column						
	RS	PBM	DB	ACC	NMI	RI
FCM	0.07466	1677.52779	1.42349	0.48387	0.28880	0.63886
RFCM	0.08706	2461.64798	1.11472	0.50645	0.32524	0.64147
SRFCM	0.10730	**2645.84922**	**1.06612**	0.51828	0.32241	0.64024
ARFCM	**0.12449**	2036.56861	1.25898	**0.62366**	**0.34194**	**0.67147**

TABLE 11. The comparative validity results of anuran calls.

Anuran Calls						
	RS	PBM	DB	ACC	NMI	RI
FCM	0.00214	0.05324	**4.47E+04**	0.56083	0.33405	0.65685
RFCM	0.00221	0.08064	3.07E+05	0.62988	0.34313	0.67763
SRFCM	0.00218	0.08638	1.70E+06	0.61091	0.34611	0.68171
ARFCM	**0.00233**	**0.09300**	6.45E+04	**0.63998**	**0.35196**	**0.68480**

TABLE 12. The comparative validity results of cryotherapy.

Cryotherapy						
	RS	PBM	DB	ACC	NMI	RI
FCM	0.00250	193.03320	5.06025	0.78889	0.25803	0.66317
RFCM	0.01148	839.20385	2.52840	0.77778	0.24372	0.65044
SRFCM	0.00511	405.91552	3.61253	0.77778	0.24372	0.65044
ARFCM	**0.01886**	**1324.41814**	**1.91556**	**0.84444**	**0.37950**	**0.73433**

(1) In the first stage of the proposed algorithm ARFCM, the three-way approximations of membership degrees with respect to a fixed cluster can capture the data topology from the global observation on data, which is considered as macro analysis on data, rather than based on the absolute distances or membership degrees of individual patterns. The patterns in the core regions have the most representative capabilities while the patterns in the boundary regions have less representative abilities and exclusive regions have no contributions as computing the prototypes.

TABLE 13. The synthetic data set D_{32}.

Pattern	x	y	Pattern	x	y	Pattern	x	y
$\mathbf{x}_1$	1	3	$\mathbf{x}_{12}$	5	3	$\mathbf{x}_{23}$	10	4
$\mathbf{x}_2$	2	3	$\mathbf{x}_{13}$	5	4	$\mathbf{x}_{24}$	10	5
$\mathbf{x}_3$	3	2	$\mathbf{x}_{14}$	6	3	$\mathbf{x}_{25}$	11	2
$\mathbf{x}_4$	3	3	$\mathbf{x}_{15}$	7	3	$\mathbf{x}_{26}$	11	3
$\mathbf{x}_5$	3	4	$\mathbf{x}_{16}$	8	3	$\mathbf{x}_{27}$	11	4
$\mathbf{x}_6$	4	1	$\mathbf{x}_{17}$	9	2	$\mathbf{x}_{28}$	12	3
$\mathbf{x}_7$	4	2	$\mathbf{x}_{18}$	9	3	$\mathbf{x}_{29}$	13	3
$\mathbf{x}_8$	4	3	$\mathbf{x}_{19}$	9	4	$\mathbf{x}_{30}$	4	20
$\mathbf{x}_9$	4	4	$\mathbf{x}_{20}$	10	1	$\mathbf{x}_{31}$	7	20
$\mathbf{x}_{10}$	4	5	$\mathbf{x}_{21}$	10	2	$\mathbf{x}_{32}$	10	20
$\mathbf{x}_{11}$	5	2	$\mathbf{x}_{22}$	10	3			

(2) In the second stage of the proposed algorithm ARFCM, the fuzziness of individual patterns over all clusters are measured by an information entropy function based on which three-way approximations with respect to the whole data set can be generated. In this way, the uncertainties of pattern locations can be detected from the micro perspectives.

(3) By integrating the partition results in the two stages, i.e., the partition results obtained in the second stage are used to modify the partition results obtained in the first stage, the unreasonable partition results in the first stage can be verified, for example, some patterns with higher fuzziness that are divided into the core regions of clusters can be adjusted to the boundary regions of these clusters. Consequently, the prototype calculations can be corrected and the obtained prototypes tend to their natural positions.

VII. CONCLUSIONS

The validity of clustering models is directly influenced by dealing with the uncertain information implicated in data. A general framework of rough-fuzzy clustering based on two-stage three-way approximations for membership degrees are presented in this study, in which the data topology with respect to a fixed cluster can be captured from the macro aspect, and the fuzziness of individual patterns over all clusters can be detected from the micro aspect. By integrating the approximation region partition results obtained in the two stages, the representative capabilities of the core and boundary regions will be adjusted according to the characteristics of data, such as the sizes and densities of clusters. The improved performance of the proposed notion is illustrated by comparative experiments. The Pedrycz's optimization principle is adopted in this study to generate the optimal partition thresholds in the two stages. As discussed before, the optimization principles can also be formed from the other aspects, such as the principle of retaining the total amount of fuzziness of

the fuzzy set [21], the principle of minimum distance [22] and the principle of least cost [22], [23]. Comparing different optimization principles and different methods for measuring the fuzziness of individual patterns under the schema of rough-fuzzy clustering are our next works.

APPENDIX

See Table 13.

REFERENCES

[1] C. C. Aggarwal and C. K. Reddy, *Data Clustering: Algorithms and Applications*. Boca Raton, FL, USA: CRC Press, 2013.
[2] J. Hou and B. Xiao, "A data-driven clustering approach for fault diagnosis," *IEEE Access*, vol. 5, pp. 26512–26520, Dec. 2017.
[3] M. S. Ali, H. Tabassum, and E. Hossain, "Dynamic user clustering and power allocation for uplink and downlink non-orthogonal multiple access (NOMA) systems," *IEEE Access*, vol. 4, pp. 6325–6343, Aug. 2016.
[4] D. N. Sandeep and V. Kumar, "Review on clustering, coverage and connectivity in underwater wireless sensor networks: A communication techniques perspective," *IEEE Access*, vol. 5, pp. 11176–11199, Jun. 2017.
[5] J. C. Bezdek, *Pattern Recognition With Fuzzy Objective Function Algorithms*, Norwell, MA, USA: Kluwer, 1981.
[6] R. N. Dave and R. Krishnapuram, "Robust clustering methods: A unified view," *IEEE Trans. Fuzzy Syst.*, vol. 5, no. 2, pp. 270–293, May 1997.
[7] Z. Pawlak, "Rough sets," *Int. J. Comput. Inf. Sci.*, vol. 11, no. 5, pp. 314–356, 1982.
[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] 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.
[10] P. Maji and S. Roy, "Rough-fuzzy clustering and multiresolution image analysis for text-graphics segmentation," *Appl. Soft Comput.*, vol. 30, pp. 705–721, May 2015.
[11] G. Peters, "Some refinements of rough *k*-means clustering," *Pattern Recognit.*, vol. 39, no. 8, pp. 1481–1491, 2006.
[12] P. Maji and S. Paul, "Rough-fuzzy clustering for grouping functionally similar genes from microarray data," *IEEE/ACM Trans. Comput. Biol. Bioinform.*, vol. 10, no. 2, pp. 286–299, Mar./Apr. 2013.
[13] J. P. Sarkar, I. Saha, and U. Maulik, "Rough possibilistic type-2 fuzzy C-means clustering for MR brain image segmentation," *Appl. Soft Comput.*, vol. 46, pp. 527–536, Sep. 2016.
[14] W. Pedrycz, "Shadowed sets: Representing and processing fuzzy sets," *IEEE Trans. Syst. Man. Cybern. B, Cybern.*, vol. 28, no. 1, pp. 103–109, Feb. 1998.
[15] W. Pedrycz, "Granular computing: The emerging paradigm," *J. Uncertain Syst.*, vol. 1, no. 1, pp. 38–61, 2007.
[16] J. Zhou, W. Pedrycz, and D. Miao, "Shadowed sets in the characterization of rough-fuzzy clustering," *Pattern Recognit.*, vol. 44, no. 8, pp. 1738–1749, 2011.
[17] Y. Y. Yao, "An outline of a theory of three-way decisions," in *Rough Sets and Current Trends in Computing* (Lecture Notes in Computer Science), vol. 7413, J. Yao *et al.*, Eds. Berlin, Germany: Springer, 2012, pp. 1–17.
[18] Y. Yao, "Three-way decisions and cognitive computing," *Cognit. Comput.*, vol. 8, no. 4, pp. 543–554, 2016.
[19] Y. Yao, "The superiority of three-way decisions in probabilistic rough set models," *Inf. Sci.*, vol. 181, no. 6, pp. 1080–1096, 2011.
[20] W. Pedrycz, "Interpretation of clusters in the framework of shadowed sets," *Pattern Recognit. Lett.*, vol. 26, no. 15, pp. 2439–2449, 2005.
[21] H. Tahayori, A. Sadeghian, and W. Pedrycz, "Induction of shadowed sets based on the gradual grade of fuzziness," *IEEE Trans. Fuzzy Syst.*, vol. 21, no. 5, pp. 937–949, Oct. 2013.
[22] Y. Yao, S. Wang, and X. Deng, "Constructing shadowed sets and three-way approximations of fuzzy sets," *Inf. Sci.*, vols. 412–413, pp. 132–153, Oct. 2017.
[23] X. Deng and Y. Yao, "Decision-theoretic three-way approximations of fuzzy sets," *Inf. Sci.*, vol. 279, pp. 702–715, Sep. 2014.
[24] G. J. Klir, U. S. Clair, and B. Yuan, *Fuzzy Set Theory: Foundations and Applications*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1997.
[25] Z. Pawlak and A. Skowron, "Rudiments of rough sets," *Inf. Sci.*, vol. 177, pp. 3–27, Jan. 2007.
[26] Z. Pawlak, S. K. M. Wong, and W. Ziarko, "Rough sets: Probabilistic versus deterministic approach," *Int. J. Man-Mach. Stud.*, vol. 29, no. 1, pp. 81–95, 1988.
[27] X. Zhang, D. Miao, C. Liu, and M. Le, "Constructive methods of rough approximation operators and multigranulation rough sets," *Knowl.-Based Syst.*, vol. 91, pp. 114–125, Jan. 2016.
[28] W. Pedrycz, "From fuzzy sets to shadowed sets: Interpretation and computing," *Int. J. Intell. Syst.*, vol. 24, no. 1, pp. 48–61, 2009.
[29] W. Pedrycz and G. Vukovich, "Granular computing with shadowed sets," *Int. J. Intell. Syst.*, vol. 17, no. 2, pp. 173–197, 2002.
[30] B. R. Ebanks, "On measures of fuzziness and their representations," *J. Math. Anal. Appl.*, vol. 97, no. 1, pp. 24–37, 1983.
[31] G. J. Klir and T. A. Folger, *Fuzzy Sets, Uncertainty, and Information*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1988.
[32] J. Knopfmacher, "On measures of fuzziness," *J. Math. Anal. Appl.*, vol. 49, no. 3, pp. 529–534, 1975.
[33] A. De Luca and S. Termini, "A definition of a nonprobabilistic entropy in the setting of fuzzy sets theory," *Inf. Control*, vol. 20, no. 4, pp. 301–312, 1972.
[34] M. Lichman. (2013). UCI machine learning repository. School of Information and Computer Science, University of California, Irvine, CA, USA. [Online]. Available: http://www.ics.uci.edu/ml
[35] M. R. N. Kalhori and M. H. F. Zarandi, "Interval type-2 credibilistic clustering for pattern recognition," *Pattern Recognit.*, vol. 48, no. 11, pp. 3652–3672, 2015.
[36] M. K. Pakhira, S. Bandyopadhyay, and U. Maulik, "Validity index for crisp and fuzzy clusters," *Pattern Recognit.*, vol. 37, no. 3, pp. 487–501, 2004.
[37] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-1, no. 2, pp. 224–227, Apr. 1979.
[38] Q. Gu, C. Ding, and J. Han, "On Trivial solution and scale transfer problems in graph regularized NMF," in *Proc. Int. Joint Conf. Artif. Intell.*, 2011, pp. 1288–1293.
[39] Y. Lei, J. C. Bezdek, J. Chan, N. X. Vinh, S. Romano, and J. Bailey, "Extending information-theoretic validity indices for fuzzy clustering," *IEEE Trans. Fuzzy Syst.*, vol. 25, no. 4, pp. 1013–1018, Aug. 2017.
[40] R. J. G. B. Campello, "A fuzzy extension of the Rand index and other related indexes for clustering and classification assessment," *Pattern Recognit. Lett.*, vol. 28, pp. 833–841, May 2007.

JIE ZHOU received the M.E. degree in computer science and technology from Central South University, China, in 2007, and the Ph.D. degree in pattern recognition and intelligence system from Tongji University, China, in 2011. He is currently a Post-Doctoral Fellow with the College of Computer Science and Software Engineering, Shenzhen University, China. His research interests include rough set theory, data mining, and intelligent systems.

ZHIHUI LAI received the Ph.D. degree in pattern recognition and intelligence system from the Nanjing University of Science and Technology, China, in 2011. He was a Post-Doctoral Fellow with the Bio-Computing Research Center, Shenzhen Graduate School, Harbin Institute of Technology, China. He has been a Research Associate, a Post-Doctoral Fellow, and a Research Fellow with The Hong Kong Polytechnic University since 2010. He has published over 50 papers, including over 20 papers published on the IEEE Transactions. His research interests include machine learning, image processing, and pattern recognition.

CAN GAO received the Ph.D. degree in pattern recognition and intelligence system from Tongji University, China, in 2013. From 2010 to 2011, he was a Visiting Scholar, supported by the China Scholarship Council, with the Department of Electrical and Computer Engineering, University of Alberta, Canada. He has been a Research Associate and a Post-Doctoral Fellow with The Hong Kong Polytechnic University since 2015. He is currently a Post-Doctoral Fellow with the College of Computer Science and Software Engineering, Shenzhen University, China. His research interests include granular computing, machine learning, and computer vision.

XIAODONG YUE received the Ph.D. degree from Tongji University. He was a Research Assistant with Hong Kong Baptist University and also a Research Fellow with the University of Technology Sydney, Sydney and The Hong Kong Polytechnic University. He is currently an Associate Professor with the School of Computer Engineering and Science, Shanghai University. His major research interests include the theories and applications of machine learning and soft computing.

WAIKEUNG WONG received the Ph.D. degree from The Hong Kong Polytechnic University, Hong Kong. He is currently a Professor with The Hong Kong Polytechnic University. He has authored or co-authored over 100 papers in refereed journals and conferences, including the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, the IEEE TRANSACTIONS ON CYBERNETICS, and the IEEE TRANSACTIONS ON IMAGE PROCESSING, among others. His current research interests include pattern recognition and feature extraction.