Information Sciences 507 (2020) 553–573

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Multigranulation rough-fuzzy clustering based on shadowed sets

Jie Zhou[a,b,*], Zhihui Lai[a,b], Duoqian Miao[c], Can Gao[a,b], Xiaodong Yue[d]

[a] College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, Guangdong, 518060, China
[b] Institute of Textile and Clothing, The Hong Kong Polytechnic University, Kowloon, Hong Kong
[c] Department of Computer Science and Technology, Tongji University, Shanghai 201804, China
[d] School of Computer Engineering and Science, Shanghai University, Shanghai 200444, China

ARTICLE INFO

Article history:
Received 10 October 2017
Revised 18 May 2018
Accepted 26 May 2018
Available online 30 May 2018

Keywords:
Multigranulation approximation regions
Rough-fuzzy clustering
Shadowed sets
Three-way decisions
Granular computing

ABSTRACT

In this study, a new technique of rough-fuzzy clustering based on multigranulation approximation regions is developed to tackle the uncertainty associated with the fuzzifier parameter m. According to shadowed set theory, the multigranulation approximation regions for each cluster can be formed based on fuzzy membership degrees under the multiple values of fuzzifier parameter with a partially ordered relation. The uncertainty generated by the fuzzifier parameter m can be captured and interpreted through the variations in approximation regions among different levels of granularity, rather than at a single level of granularity under a specific fuzzifier value. An ensemble strategy for updating prototypes is then presented based on the constructed multigranulation approximation regions, in which the prototype calculations that may be spoiled due to the uncertainty caused by a single fuzzifier value can be modified. Finally, a multilevel degranulation mechanism is introduced to evaluate the validity of clustering methods. By integrating the notions of shadowed sets and multigranulation into rough-fuzzy clustering approaches, the overall topology of data can be captured well and the uncertain information implicated in data can be effectively addressed, including the uncertainty generated by fuzzification coefficient, the vagueness arising in boundary regions and overlapping partitions. The essence of the proposed method is illustrated by comparative experiments in terms of several validity indices.

1. Introduction

Clustering is an unsupervised learning technique to find natural groups that are implicated in data. Its main task is to partition unlabeled patterns $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N\}$, $\mathbf{x}_j \in \Re^M (j = 1, 2, \cdots, N)$, into $C (1 < C < N)$ subgroups $\{G_1, G_2, \cdots, G_C\}$ such that the patterns in the same cluster (intra cluster) are expected to have the highest similarities (homogeneities) and the patterns between different clusters (inter cluster) are expected to have the highest dissimilarities (heterogeneities). In this manner, the data structure that is revealed by a clustering method is expected to reflect the natural geometry of the data as much as possible.

* Corresponding author at: Room 519, College of Computer Science and Software Engineering, Shenzhen University, Nanshan District, Shenzhen City, Guangdong Province, China.
E-mail addresses: jie_jpu@163.com (J. Zhou), lai_zhi_hui@163.com (Z. Lai), dqmiao@tongji.edu.cn (D. Miao), 2005gaocan@163.com (C. Gao), yswantfly@shu.edu.cn (X. Yue).

https://doi.org/10.1016/j.ins.2018.05.053
0020-0255/

Hard C-means (HCM) [14] is one of the foremost objective function-based clustering methods in which the degree of each pattern belonging to each cluster is 0 or 1. The validity of HCM will degenerate as dealing with patterns with overlapping areas. Fuzzy clustering, especially fuzzy C-means (FCM) [1], as one extension of HCM, utilizes a partition matrix to evaluate the degree of each pattern belonging to each cluster, so that the overlapping partitions can be described effectively. The main challenge of FCM is the sensitivities to noisy patterns which may contaminate the calculations of the corresponding prototypes and membership degrees.

Based on rough set theory [19], which aims at analyzing data involving uncertain, imprecise or incomplete information, Lingras and West [13] proposed a rough C-means (RCM) clustering method. Each cluster is described not only by a prototype, but also with a pair of lower and upper approximations. Meanwhile, the boundary region is defined as the difference between the lower and upper approximations. The uncertainty and vagueness arising in the boundary region of each cluster can be captured well in RCM. Since no membership degrees are involved, the closeness of patterns to clusters cannot be detected.

Rough sets and fuzzy sets, as two important paradigms of granular computing [24,32,41], are strongly complementary to each other. Incorporating with fuzzy membership degrees, Mitra et al. [17] presented a rough-fuzzy C-means (RFCM) clustering method which integrated the merits of both fuzzy sets and rough sets. The lower and upper approximations are determined according to the membership degrees, rather than the individual absolute distances between a pattern and its neighbors. Maji et al. [15] further proposed a robust rough-fuzzy C-means algorithm that integrated both probabilistic and possibilistic memberships of fuzzy sets, which could handle overlapping clusters in noisy environments as well as the uncertainty and vagueness in cluster definitions due to involving rough sets.

No matter which rough-fuzzy partitive clustering methods will be used, some model parameters are involved: (a) the weighted values that evaluate the contributions of lower approximations and boundary regions when calculating new prototypes; (b) the threshold that determines the lower approximation and boundary region of each cluster; (c) the value of fuzzifier parameter m that controls the shape of memberships. Since the contributions of lower approximations are considered more important than the contributions of boundary regions as computing the prototypes, the weighted value for lower approximations is much higher, and its complementarity is applied for the boundary regions. Meanwhile, through the combinational adjustments of (b) and (c), the influence caused by the weighted values can be reduced. In this case, (b) and (c) will take over the effect of (a).

The threshold, that determines the approximation regions of each cluster, is often selected depending on subjective tuning in the available researches [16,26]. Maji et al. [15] and Sarkar et al. [28] chose this value as the average value and the median of the difference between the highest and second highest fuzzy memberships of all the patterns, respectively. However, the same threshold is employed for all clusters though the sizes and the densities of the clusters may be significantly diverse. Additionally, the approximation regions are partitioned based on the absolute distances or membership degrees of individual patterns, not the global observation on data for a specific cluster, then the topology of data can not be detected well with respect to this cluster. Shadowed sets [21], as a bridge between rough sets and fuzzy sets, provide a new scheme to granular computing. It is an example of three-way, three-valued, or three-region approximations of a fuzzy set [4,6,29] according to the framework of three-way decisions [9,35,39] which is another new paradigm of granular computing and can provide favorable semantical interpretation and generalized optimization methodology for determining partition threshold values [34,36,40]. A shadowed set-based rough-fuzzy C-means (SRFCM) method was introduced in [42] which gave a technique of automatic selection of the partition threshold. No matter which selection methods are used, an unreasonable threshold will result in undesired approximation region partitions, and then the prototype calculations may be spoiled.

The value of fuzzifier parameter m is very important for the updating of prototypes and the corresponding partition matrix. A predefined value of m (mostly $m = 2$) is often used in the rough set and fuzzy set-based clustering methods [15,16,26,28]. However, it is difficult to express the uncertain notion of fuzziness in a given data set using a single fuzzifier value. To manage the uncertainty generated by the fuzzifier parameter m, Hwang et al. [7] proposed interval type-2 fuzzy C-means (TFCM) which extended a pattern set to interval type-2 fuzzy sets using a pair of fuzzifier m_1 and m_2 that created a footprint of uncertainty for the fuzzifier parameter. Linda et al. [12] further revealed a general type-2 fuzzy C-means algorithm via the α-planes representation theorem. However, the values m_1 and m_2 mainly depend on subjective selection or enumeration in the available studies [7,8,12] and the results need more interpretations. Recently, the notion of multigranulation [27,31,37] in granular computing is developed for solving human-centric problems and interpreting the obtained results from the perspective of multiple levels of granularity. This methodology provides a new insight as analyzing the uncertainties generated by the model parameters.

The purpose of this paper is mainly to tackle the uncertainties associated with the two key parameters in rough set and fuzzy set-based clustering approaches, viz., the threshold related to the approximation region partitions and the fuzzification coefficient m. The main objectives of this paper are: (1) to optimize the partition threshold for each cluster based on shadowed sets, which is obtained from the perspective of the global observation on data and will be used as a cornerstone for establishing the multi-levels of granularity for approximation regions; (2) to capture the uncertainty generated by the fuzzifier parameter m in rough-fuzzy clustering methods via the variations in multigranulation approximation regions formed under multiple values of fuzzifier parameter with a partially ordered relation, rather than at a single level of granularity under a specific fuzzifier value; (3) to update the prototypes by combining the intermediate results obtained at different levels of granularity. In this way, the prototypes calculated at a single level can be modified and then tend to their natural

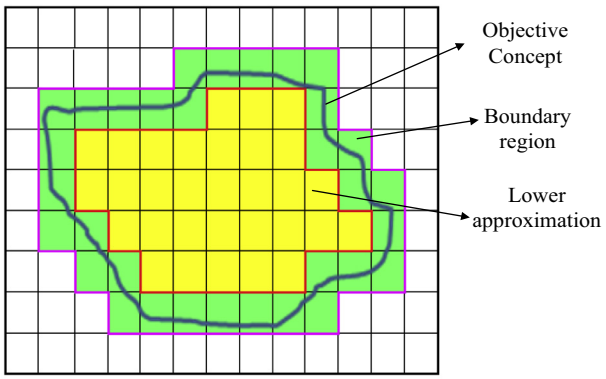

Fig. 1. The lower and upper approximations of a concept.

positions; (4) to develop a multilevel degranulation mechanism according to "granulation-degranulation" philosophy based on which the quality of the clustering model can be evaluated.

By integrating various granular computing technologies, i.e., fuzzy sets, rough sets, shadowed sets and the notion of multigranulation, the uncertain information in data, including overlapping partitions, the vagueness arising in boundary regions and the uncertainty produced by fuzzification coefficient, can be handled sufficiently. Experimental results with the use of synthetic and real-world data illustrate the improved performance of the proposed notion in terms of several validity indices, such as relative separation index (RS) [8], Davies–Bouldin index (DB) [5], Dunns index (Dunn) [2] and PBM-index (PBM) [18] as well as the granulation-degranulation index.

The rest of paper is organized as follows. Some rough set-based partitive clustering methods are reviewed in Section 2. The uncertainty generated by the fuzzifier parameter m is revealed in Section 3. Meanwhile, multigranulation approximate regions are formed based on which a new rough-fuzzy C-means method is introduced. In Section 4, the proposed granulation-degranulation mechanisms based on multiple levels of granularity are explained. Comparative experiments are presented in Section 5. Some conclusions are given in Section 6.

2. Rough set-based partitive clustering methods

In this section, some rough set-based partitive clustering algorithms will be reviewed, which include rough C-means [13], rough-fuzzy C-means [17] and shadowed set-based rough-fuzzy C-means [42]. More detailed information about rough sets and shadowed sets can be found in [19–23,25,29,33,38,40].

2.1. Rough C-means

In rough set theory [19], a concept can be described by a pair of crisp sets, i.e., the lower and upper approximations. The lower approximation is composed of objects that belong to the concept certainly, and the upper approximation is composed of objects that belong to the concept possibly. The lower and upper approximations approach the concept from two sides, viz. inside and outside, as shown in Fig. 1. The union of lower approximation and boundary region constitutes the upper approximation in Fig. 1. Each square in Fig. 1 represents a knowledge granule. The granulation under information in hand is composed of all squares in Fig. 1. If the granulation is refined enough, viz. each square (granule) is small enough, such as a pixel, the objective concept can be certainly depicted by the lower and upper approximations. In this case, the boundary region of this objective concept is eliminated.

Lingras et al. [13] extended the notion of rough approximations to develop a clustering algorithm, called rough C-means (RCM), in which all patterns were divided into three levels, i.e., core level (lower approximation, also called as core region), boundary level and exclusion level for a fixed cluster. The new prototype calculations are only related to the core level and boundary level, instead of all patterns as that in hard C-means (HCM) or fuzzy C-means (FCM). So the useless information can be filtrated and numeric computation can be reduced.

Patterns in the core level will belong to this cluster certainly and patterns in the boundary level will belong to this cluster possibly, i.e., with vagueness and uncertainty. The rest in the exclusion level will not belong to this cluster definitely. Generally, patterns in the core level have the most importance, patterns in the boundary level have less contribution and patterns in the exclusion level almost make no contribution to the updating of prototypes.

A data set with N patterns $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$, $\mathbf{x}_j \in \Re^M (j = 1, 2, \ldots, N)$ is expected to group into C clusters $G_1, G_2, \ldots, G_C$. According to RCM, the corresponding prototypes $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_C$, $\mathbf{v}_i \in \Re^M (i = 1, 2, \ldots, C)$ are renewed by the following principles:

$$\mathbf{v_i} = \begin{cases} w_l A_1 + w_b B_1 \ if \ \underline{R}G_i \neq \emptyset \wedge R_b G_i \neq \emptyset \\ B_1 \ if \ \underline{R}G_i = \emptyset \wedge R_b G_i \neq \emptyset \\ A_1 \ if \ \underline{R}G_i \neq \emptyset \wedge R_b G_i = \emptyset \end{cases} \tag{1}$$

Where $A_1 = \frac{\sum_{\mathbf{x}_j \in \underline{R}G_i} \mathbf{x}_j}{card(\underline{R}G_i)}$, $B_1 = \frac{\sum_{\mathbf{x}_j \in R_b G_i} \mathbf{x}_j}{card(R_b G_i)}$ can be considered as the contributions by the crisp core region and two-valued boundary region, respectively. $card(X)$ means the cardinality of set X. $R_b G_i = \overline{R}G_i - \underline{R}G_i$ denotes the boundary region of cluster G_i, where $\underline{R}G_i$ and $\overline{R}G_i$ 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 which measure the contributions of core region and boundary region, respectively.

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

If $\|\mathbf{x}_j - \mathbf{v}_q\| - \|\mathbf{x}_j - \mathbf{v}_p\| \leq \Delta$, then $\mathbf{x}_j \in \overline{R}G_p$ and $\mathbf{x}_j \in \overline{R}G_q$. In this case, $\mathbf{x}_j$ cannot be partitioned into the core regions of any clusters. Otherwise, $\mathbf{x}_j \in \underline{R}G_p$. Where $\|\mathbf{x}_j - \mathbf{v}_i\|$ denotes the distance between pattern $\mathbf{x}_j$ and prototype $\mathbf{v}_i$. $\|\mathbf{x}_j - \mathbf{v}_p\|$ is the minimum of $\mathbf{x}_j$ over all clusters and $\|\mathbf{x}_j - \mathbf{v}_q\|$ is next to the minimum.

The threshold Δ is crucial for determining the approximation regions of each cluster. The smaller the threshold is, the more objects will belong to the core regions. On the contrary, the larger the threshold is, the more objects will belong the boundary regions. An unreasonable threshold will result in undesired approximation region partitions which may misguide the prototype computations. Additionally, since no membership degrees are involved, the overlapping partitions cannot be effectively described in RCM.

2.2. Rough-fuzzy C-means

Incorporating with fuzzy membership degrees, Mitra et al. [17] proposed the notion of rough-fuzzy C-means (RFCM), in which the absolute distance $\|\mathbf{x}_j - \mathbf{v}_i\|$ was replaced by a fuzzy membership degree u_{ij} when dividing patterns to approximation regions. This adjustment enhances the robustness of the clustering as dealing with overlapping situations. The recomputation of prototypes is correspondingly modified as follows:

$$\mathbf{v}_i = \begin{cases} w_l A_2 + w_b B_2 & if \ \underline{R}G_i \neq \emptyset \wedge R_b G_i \neq \emptyset \\ B_2 & if \ \underline{R}G_i = \emptyset \wedge R_b G_i \neq \emptyset \\ A_2 & if \ \underline{R}G_i \neq \emptyset \wedge R_b G_i = \emptyset \end{cases} \tag{2}$$

Where $A_2 = \frac{\sum_{\mathbf{x}_j \in \underline{R}G_i} {u_{ij}}^m \mathbf{x}_j}{\sum_{\mathbf{x}_j \in \underline{R}G_i} {u_{ij}}^m}$, $B_2 = \frac{\sum_{\mathbf{x}_j \in R_b G_i} {u_{ij}}^m \mathbf{x}_j}{\sum_{\mathbf{x}_j \in R_b G_i} {u_{ij}}^m}$ can be considered as the contributions by the fuzzy core region and fuzzy boundary region, respectively. The weighted values $0.5 < w_l \leq 1$ and $w_b = 1 - w_l$ have the same interpretations as that in RCM. u_{ij} denotes the membership degree of pattern $\mathbf{x}_j$ belonging to the cluster with prototype $\mathbf{v}_i$, and is calculated as the same as in FCM.

$$u_{ij} = \frac{1}{\sum_{k=1}^{C} \left(\frac{\|\mathbf{x}_j - \mathbf{v}_i\|}{\|\mathbf{x}_j - \mathbf{v}_k\|} \right)^{\frac{2}{m-1}}}. \tag{3}$$

$u_{ij} \in [0, 1] (i = 1, 2, \ldots, C, \ j = 1, 2, \ldots, N)$. For $\forall j$, it has $\sum_{i=1}^{C} u_{ij} = 1$ and $\forall i$, it has $0 < \sum_{j=1}^{N} u_{ij} < N$. When one or more of the distances $\|\mathbf{x}_j - \mathbf{v}_i\| = 0$ at any iterations, the singularity in FCM will arise. In this case, assign $u_{ij} = 1$ and $u_{kj} = 0$ for $\forall k \neq i$. The value of u_{ij} in formula (3) depends on the value of fuzzifier parameter m and the relative distances of pattern $\mathbf{x}_j$ with respect to all prototypes. The value of fuzzifier parameter m reflects the shape of fuzzy memberships. When the value is close to 1, it implies a Boolean nature of one cluster, namely, the memberships are maximally hard. In this case, only the memberships of patterns that have equivalent distances to each cluster are fuzzy. On the other hand, it will result in spike-like membership functions when the value increases. In this case, the memberships are maximally fuzzy which means only the memberships of patterns that are located around the cluster centers are assigned 1, otherwise, they are assigned memberships of $1/C$, which can be illustrated in Fig. 2. The uncertainty caused by fuzzifier parameter m will be discussed in Section 3.1 in detail. In order to determine the approximation regions, the following principles are exploited.

If $u_{pj} - u_{qj} \leq \Delta$, then $\mathbf{x}_j \in \overline{R}G_p$ and $\mathbf{x}_j \in \overline{R}G_q$. In this case, $\mathbf{x}_j$ cannot be partitioned into the core regions 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 in RFCM is pivotal for determining the approximation regions of each cluster which is similar to RCM.

2.3. Shadowed set-based rough-fuzzy clustering

The notion of shadowed sets [21–23,25], which is considered as a bridge between rough sets and fuzzy sets, provides a simplification (approximation) technology for fuzzy sets. Zhou and Pedrycz et al. [42] proposed a shadowed set-based rough-fuzzy clustering (SRFCM) method in which the determination of approximation regions for each cluster was transferred to an optimization process and could be detected automatically during the clustering processes. The procedures for determining the approximation regions of each cluster based on shadowed sets are reviewed as follows.

Step1: Compute membership values u_{ij} of each pattern $\mathbf{x}_j$ to each prototype $\mathbf{v}_i$ using formula (3);

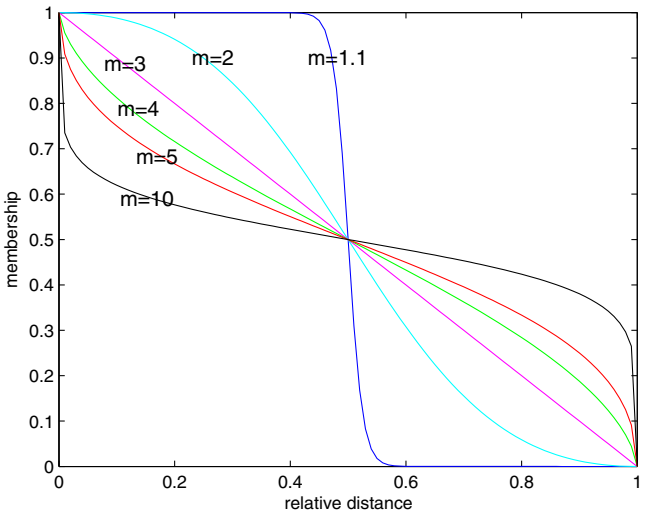

Fig. 2. The relative distances and corresponding membership degrees.

Step2: Based on shadowed sets, compute optimal threshold α_i for each cluster G_i:

$$\alpha_i = \min_{\alpha}(V_i) = \min_{\alpha}\left|\sum_{j:u_{ij}\le\alpha} u_{ij} + \sum_{j:u_{ij}\ge\max_j(u_{ij})-\alpha}(1-u_{ij}) - card\left(\left\{\mathbf{x}_j|\alpha < u_{ij} < \max_j(u_{ij}) - \alpha\right\}\right)\right|. \tag{4}$$

Step3: According to α_i, determine the core region and boundary region of cluster G_i.

$$\underline{R}G_i = \left\{\mathbf{x}_j|u_{ij} \ge \max_j(u_{ij}) - \alpha_i\right\},$$
$$R_bG_i = \left\{\mathbf{x}_j|\alpha_i < u_{ij} < \max_j(u_{ij}) - \alpha_i\right\}. \tag{5}$$

From the above Steps 1–3, the threshold of each cluster is not user-defined beforehand. It can be adjusted automatically in the clustering processes and can be optimized for each cluster independently. The core region and boundary region regarding one cluster can be effectively divided through the global observation on data, rather than depending on individual absolute distances or individual membership values of patterns. Subsequently, the prototypes are updated as the same as in RFCM. The outperformance of SRFCM compared with FCM, RCM and RFCM is demonstrated by some experimental results in [42].

Shadowed sets are a model of three-way approximations of fuzzy sets in which the main fundamental issue is the interpretation and determination of partition thresholds. The Pedrycz's optimization model with formula (4) is 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 within the framework of three-way decision theory which can provide good semantical interpretations, such as the principle of uncertainty invariance, the principle of minimum distance (including semantic distance) and the principle of least cost [6,40]. The selected principle should consider the characteristics of practical applications which can be executed based on the guidelines of three-way decisions.

In rough-fuzzy clustering methods, the value of parameter m controls the shape of fuzzy memberships. The approximation region partitions for a fixed cluster may be varied according to different fuzzifier values. Unfortunately, the uncertainty caused by fuzzification coefficient m, especially the dynamic variations of approximation regions when changing the fuzzifier values, are not discussed in recent rough-fuzzy clustering methods. This value is often user pre-selected as implementing the corresponding algorithms [15–17,28]. An unreasonable value of m may distort the prototype and membership degree computations. To overcome this situation, the uncertainty generated by the fuzzifier parameter needs to be carefully addressed when data information and knowledge are not enough in-hand.

3. Multigranulation rough-fuzzy clustering

3.1. The uncertainty generated by fuzzifier parameter m

As illustrated by Hwang et al. in [7], type-1 fuzzy sets cannot sufficiently manage the uncertainty generated by a single fuzzifier value. For the patterns in a data set which has two clusters, the relationship between their relative distances (also considered as normalized distances) to prototypes and the corresponding membership degrees according to formula (3) is shown in Fig. 2. When the value of m tends to 1, the memberships are most crisp or hard. In this case, only the patterns with relative distances around 0.5 are fuzzy. When the value of m approaches to ∞, the memberships are most fuzzy, and only the patterns closed to the prototypes can be partitioned definitely. Two different fuzzifier values m_1 and m_2 ($1 < m_1 < m_2$, assumed for the whole rest of paper) form a footprint of uncertainty (FOU) for memberships which can be used to manage uncertainty generated by different fuzzifier values from the vertical direction in Fig. 2, i.e., the variations of memberships under the same relative distance. Available type-2 fuzzy clustering methods are mainly presented based on this notion. Perspectively, the description of the uncertainty generated by fuzzifier values can also be drawn from the horizontal direction in Fig. 2, i.e., the variations of relative distances of patterns over all cluster prototypes under a specific membership value.

Suppose a data set with two clusters, according to formula (3), the membership degree of each pattern belonging to each cluster can be considered as a function with respect to their relative distances and the fuzzifier parameter, namely, $u_{ij} = u_{ij}(d, m)$, where d denotes the relative distance of a pattern with respect to one of the clusters. $m > 1$ is the fuzzifier parameter. Denote $u_{ij}(d, m) = g(d, m)$, formula (3) can be rewritten as follows:

$$u_{ij} = g(d, m) = \begin{cases} \dfrac{1}{1 + \left(\frac{d}{1-d}\right)^{\frac{2}{m-1}}} & d \in [0, 1) \\ 0 & d = 1 \end{cases} \tag{6}$$

Obviously, $g(d, m)$ achieves three common values no matter which values of m are used, viz. 0, 0.5 and 1, as illustrated in Fig. 2. According to the formulation of $g(d, m)$, a proposition can be drawn as follows:

Proposition 1. *Suppose* $1 < m_1 < m_2$ *and* $g(d_1, m_1) = g(d_2, m_2)$, *then it has:*

1) If $0 < d_1, d_2 < 0.5$, then $d_2 < d_1$;
2) If $0.5 < d_1, d_2 < 1$, then $d_1 < d_2$.

The detailed proofs can be found in the Appendix. Proposition I reflects the influence generated by different fuzzifier values with respect to the relative distances of patterns. It is an interesting thing that the patterns with the same membership degree to a cluster may have different relative distances to this cluster due to different fuzzifier values.

Given a data set with two clusters G_1 and G_2, taking G_1 as an example, G_2 can be explained similarly. When giving a specific fuzzifier value $m = m_1$ (such as $m_1 = 2$), the optimal threshold $\alpha_1(m_1) = \alpha$ for cluster G_1 under m_1 can be obtained by formula (4) based on the scheme of shadowed sets. The boundary region of cluster G_1 is composed of the patterns which have membership degrees between α and $1 - \alpha$ according to formula (5). In this case, their relative distances to G_1 are in interval $(d_1, 1 - d_1)$, as illustrated in Fig. 3. The core region and the exclusion region of cluster G_1 are composed of the patterns with relative distances in intervals $[0, d_1]$ and $[1 - d_1, 1]$, respectively. The approximation region partitions for cluster G_1 under fuzzifier value m_1 can be shown in Fig. 4(a) schematically. The green and blue regions are the core region and boundary region of cluster G_1, respectively.

If the fuzzifier value m increases from m_1 to m_2 (such as $m_2 = 5$), and the partition threshold for cluster G_1 under m_2 is assigned with the value obtained under m_1 (namely, $\alpha_1(m_2) = \alpha$), the boundary region of cluster G_1 will cover the patterns with the relative distances in interval $(d_2, 1 - d_2)$. Meanwhile, the core region and the exclusion region of cluster G_1 turn to the patterns with relative distances in intervals $[0, d_2]$ and $[1 - d_2, 1]$, respectively. According to Proposition I, it has $(d_1, 1 - d_1) \subset (d_2, 1 - d_2)$. It means that the core region of G_1 under the fuzzifier value m_2 (the yellow region in Fig. 4(b)) are contracted relative to the fuzzifier value m_1 (the green region in Fig. 4(a)) using the same partition threshold. The variation in exclusion region has the same contraction tendency. However, the boundary region of G_1 is extended (from the blue region in Fig. 4(a) to the union of green, blue and gray regions in Fig. 4(b)). In this case, some patterns in the core region that are definitely partitioned into cluster G_1 under m_1 will be partitioned into the boundary region of cluster G_1 under m_2 (the green region in Fig. 4(b)). As pointed out here, one pattern will have different membership values under fuzzifier values m_1 and m_2, based on which it may be divided into different approximation regions with respect to a fixed cluster. Since the membership value of this pattern to a cluster obtained under a specific fuzzifier value m_1 or m_2 is with uncertainties to some degree, the corresponding approximation region partitions are also with some uncertainties. Therefore, by detecting the variations of approximation regions of a fixed cluster between different fuzzifier values m_1 and m_2, the uncertainty caused by the individual fuzzifier value can be captured, such as the green region in Fig. 4(b).

By contrast, if the fuzzifier value is tuned from m_2 to m_1, the core region of G_1 will be extended from the patterns with relative distances in $[0, d_2]$ to the patterns with relative distances in $[0, d_1]$ (from the yellow region in Fig. 4(b) to the green region in Fig. 4(a)), and the boundary region will become narrow. The patterns changing from the core region

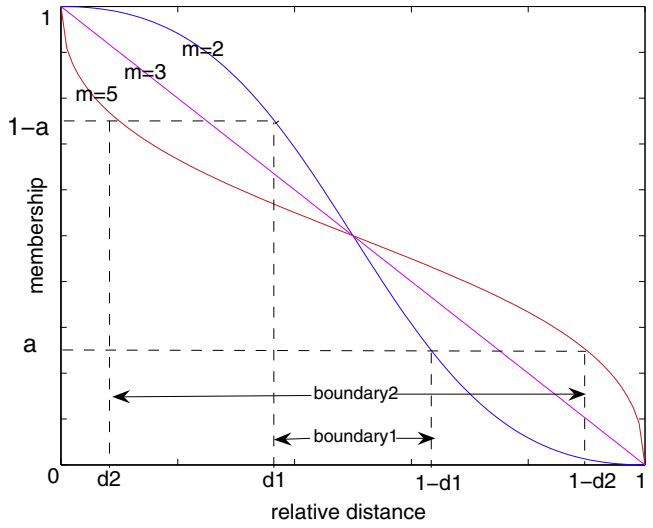

Fig. 3. The variations of approximation regions under different values of m.

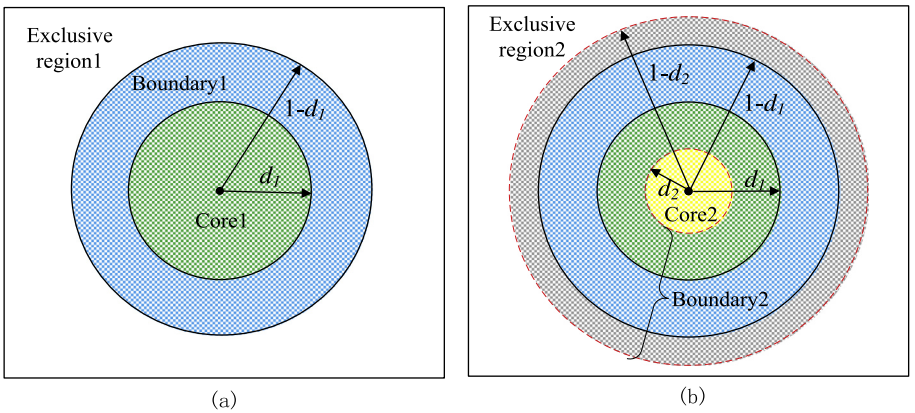

Fig. 4. The variations of approximation regions for a fixed cluster. (a)The approximation region partitions for a fixed cluster under fuzzifier m_1. (b) The approximation region partitions for a fixed cluster under fuzzifier m_1 and m_2 according to the same partition threshold obtained under m_1. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

to the boundary region ($m_1 \to m_2$) or changing from the boundary region to the core region ($m_2 \to m_1$) with respect to a fixed cluster reflect the uncertainty generated by the fuzzifier parameter m. In this manner, the uncertainty related to the fuzzifier parameter m in clustering processes can be captured through the approximation region variations which can form a multigranulation structure.

Taking the data set which has three clusters with normalized distribution (used in [42]) as an example, the core region and boundary region of each cluster under different fuzzifier values are shown in Fig. 5. According to formula (4), the partition thresholds for three clusters under $m = 2$ are 0.33568, 0.31872 and 0.30445, which are fixed for other fuzzifier values.

In Fig. 5, when the fuzzifier value becomes 1.1, the boundary region of each cluster is contracted and the core region of each cluster is extended. In this case, most patterns will be partitioned into the core regions, including some overlapping and noisy patterns. When the value of m becomes larger (3 or 4), the core region of each cluster is contracted and the belt of boundary region of each cluster is broadened. In this case, more patterns will be partitioned into the boundary regions, including some patterns that are partitioned into the core regions or exclusive regions under smaller value $m = 2$ previously. Since the core region of each cluster is contracted gradually as increasing the fuzzifier value according to Proposition I, a multigranulation structure of core regions for each cluster under different fuzzifier values can be established which will be specifically discussed in the next section.

Actually, the fuzzifier value cannot be assigned with a very large value since the core regions of some clusters may become empty, i.e., the representative patterns in the core regions are removed to the boundary regions. In this case, core regions have no representative capacities and the calculated prototypes may be distorted due to the extended boundary regions. The maximum value of fuzzifier parameter should guarantee the nonempty core region for each cluster. Extremely,

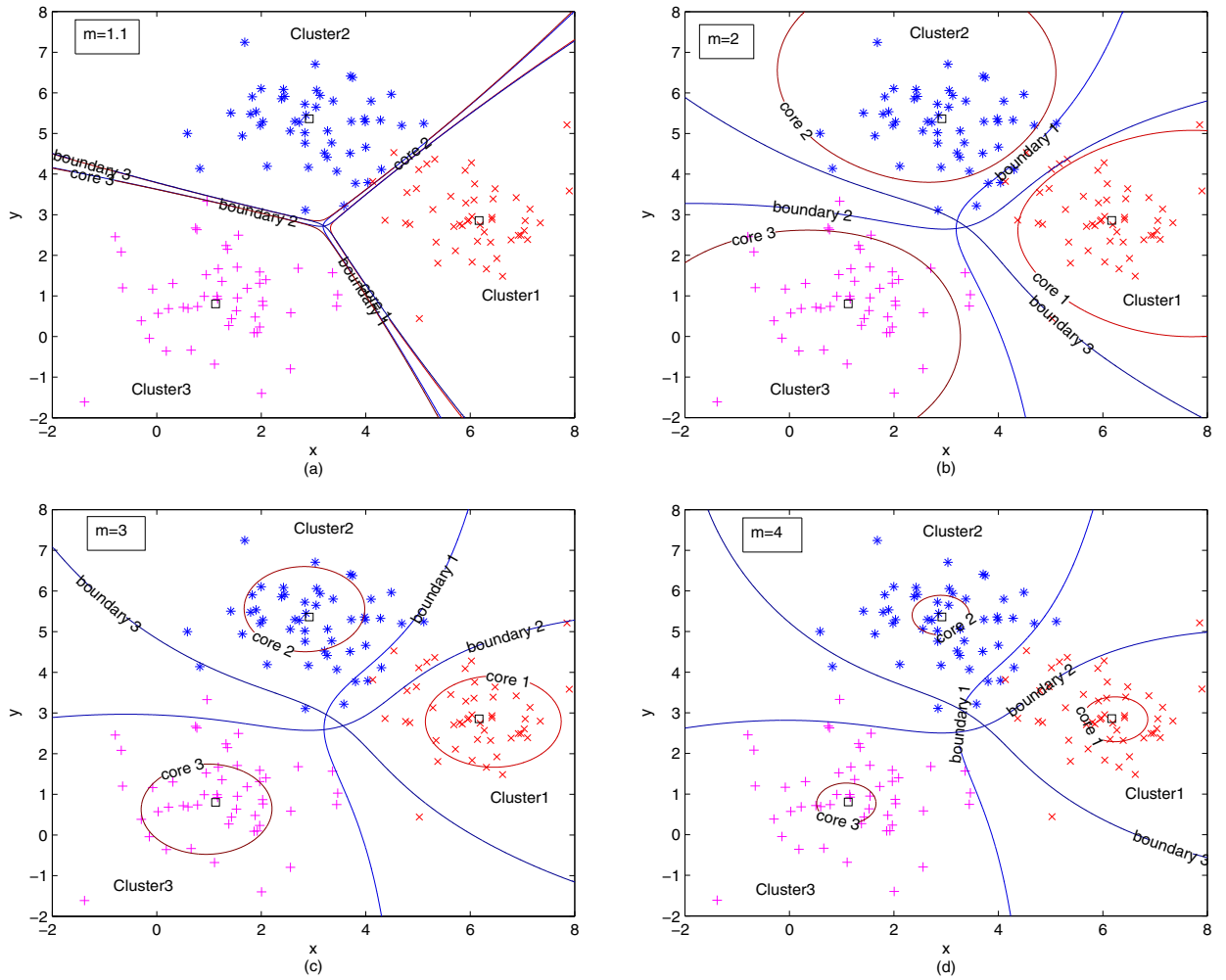

Fig. 5. Approximation regions generated under different values of m according to the same partition threshold of each cluster (a)$m = 1.1$; (b)$m = 2$; (c)$m = 3$; (d)$m = 4$. The black squares in the figures denote the prototypes.

there is only one pattern in the core region of a specific cluster. The selection of value m_2 with respect to the fixed value m_1 should consider the variations in approximation regions which can form multiple levels of granularity of clusters. There will be a balance between the representative capacities of approximation regions and the levels of granulation.

3.2. Multigranulation construction in rough-fuzzy clustering

As shown in Figs. 3 and 5, the core region of each cluster is contracted gradually as increasing the values of fuzzification coefficient with respect to the same partition threshold of this cluster. Assuming a series of fuzzifier values satisfies $1 < m_1 < m_2 < m_3 < \cdots < m_{\max}$, the core regions obtained under the smaller fuzzifier values can be further decomposed, as displayed in Fig. 6.

In Fig. 6, $Core_k$ and $Boundary_k$ denote the core region and boundary region of one cluster under the value of m_k, respectively. $\underline{Boundary_k}$ means the patterns belonging to the core region of this cluster under the value of m_{k-1} and those also belonging to the boundary region of this cluster under the value of m_k (like the green region in Fig. 4(b)). $\underline{Exclusive_k}$ means the patterns belonging to the exclusive region of a cluster under the value of m_{k-1} and those also belonging to the boundary region of this cluster under m_k (like the gray region in Fig. 4(b)). When the fuzzifier value becomes a larger one, the new boundary region of the cluster is formed with three parts, i.e., $\underline{Boundary_k}$, $Boundary_{k-1}$ and $\underline{Exclusive_k}$. Simultaneously, the core region $Core_{k-1}$ at a higher level (like the green region in Fig. 4(a)) is decomposed to a new one $Core_k$ (like the yellow region in Fig. 4(b)) and $\underline{Boundary_k}$ at the adjacent level below.

The approximation regions at higher levels in Fig. 6 can be considered as coarser partitions which may involve some noisy data or overlapping patterns in the core regions. Under this circumstance, the prototype calculations may be distorted

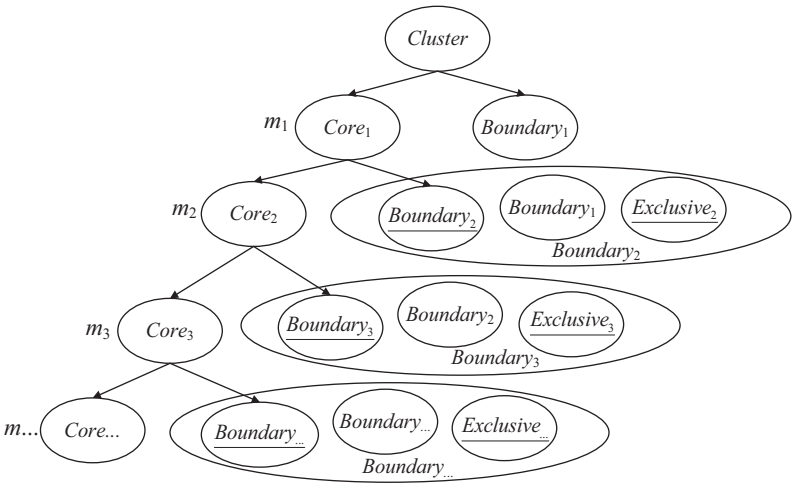

Fig. 6. The multigranulation core regions of a cluster ($1 < m_1 < m_2 < m_3 < \cdots < m_{\max}$).

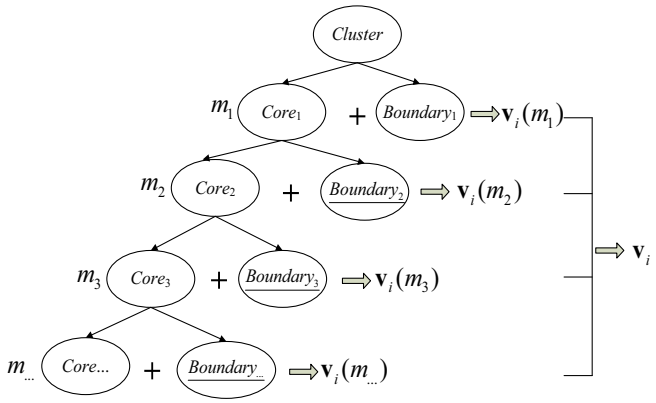

Fig. 7. The prototype calculations based on multigranulation approximation regions ($1 < m_1 < m_2 < m_3 < \cdots < m_{\max}$).

at a single level of granularity. By decomposing the core region of each cluster from a higher level to a lower level, some uncertain patterns in the core region at the higher level will be wiped off from the new core region at the lower level.

The multigranulation approximation regions constructed by increasing the value of fuzzification coefficient can be considered as a "coarse to fine" mechanism, i.e., the core region of each cluster is contracted gradually which tends to the natural centroid of this cluster. On the contrary, the multigranulation approximation regions constructed by decreasing the value of fuzzification coefficient can be considered as a "fine to coarse" mechanism, i.e., the core region of each cluster is extended gradually. Two adjacent levels in the constructed multigranulation structure share some common information, i.e., the approximation regions at one level are obtained depending on the partition results obtained at the other level due to the same partition threshold for each cluster.

3.3. Rough-fuzzy clustering based on multigranulation approximation regions

Based on the constructed multigranulation core regions of each cluster under a series of fuzzifier values with a strict partial order, the corresponding prototypes can be computed at different levels of granularity independently, and the final prototypes can be optimized by these candidate results. The prototype of one cluster computed at a lower level (with a larger value of fuzzifier parameter) can be considered as the modification for the one that is obtained at a higher level (with a smaller value of fuzzifier parameter). This notion is illustrated in Fig. 7, where $\mathbf{v}_i(m_k)$ means the prototype of the ith cluster under the fuzzifier value m_k.

In Fig. 7, only some patterns in the boundary region at a lower level (such as $m = m_k$) in Fig. 6 are selected for computing the prototype, instead of all patterns in the boundary region at this level. Because the patterns in $Boundary_{k-1}$ have been exploited for computing the prototype at the upper level ($m = m_{k-1}$), and the patterns in $\underline{Exclusive_k}$ almost make no contribution to this cluster. $\mathbf{v}_i(m_k)$ can be considered as the correction for $\mathbf{v}_i(m_{k-1})$ due to some uncertain patterns in the $Core_{k-1}$. A generalized framework for calculating the prototypes $\mathbf{v}_i$ ($i = 1, 2, \ldots, C$) based on multigranulation approximation

regions can be described as follows:

$$\mathbf{v}_i = \sum_{k=1}^{k_{\max}} \beta_k \mathbf{v}_i(m_k), \tag{7}$$

where $\beta_k \in [0, 1]$ $(k = 1, 2, \ldots, k_{\max})$ are weighted values, and satisfy $\sum_{k=1}^{k_{\max}} \beta_k = 1$. If $\forall k$, $\beta_k = \frac{1}{k_{\max}}$, it means that the intermediate result obtained at each level of granularity makes the same contribution to renewing the prototype $\mathbf{v}_i$.

The levels in Figs. 6 or 7 cannot be grown unlimited due to the maximum value of fuzzifier parameter which is discussed in Section 3.1. This value can be achieved by the following principle.

Given a fuzzifier value m_1, the corresponding maximum value of fuzzifier parameter $m_{\max}$ satisfies:

$$m_{\max} = \max_{m_k} \left\{ m_k | \forall G_i (i = 1, 2, \cdots, C), \frac{card(\widetilde{\underline{R}}^{m_k} G_i)}{card(\underline{R}^{m_1} G_i)} > \xi \right\}, \tag{8}$$

where $0 \le \xi < 1$. $\underline{R}^{m_1} G_i$ denotes the core region of cluster G_i with respect to m_1 based on shadowed set optimization model. $\widetilde{\underline{R}}^{m_k} G_i$ denotes the core region of G_i with respect to m_k based on the partition threshold obtained under the fuzzifier value m_1. If $\xi = 0$, the core regions of each cluster obtained at all levels will not be empty during the multigranulation construction procedures. In this case, the representative capacities of the core regions at each level of granularity can be guaranteed.

Interval type-2 fuzzy clustering methods often use the FOU (footprint of uncertainty), which is formed between two fuzzifier values m_1 and m_2, to manage the uncertainty generated by the fuzzifier parameter. However, how to select the two fuzzifier values is rarely discussed in the available studies [7,8,12]. Integrating interval type-2 fuzzy sets into rough-fuzzy clustering methods is also not a straightforward task, several issues that need to be addressed, such as type reduction and the uncertainties generated by the selected fuzzifier values. Though a generalized multigranulation framework has been introduced for calculating prototypes, this paper will concentrate on a specific case with two fuzzifier values m_1 and m_2, such that the uncertainty created by a single fuzzifier value can be well captured and iterative computation can be saved evidently.

Given two fuzzifier values m_1 and m_2 $(1 < m_1 < m_2)$, the prototypes under m_1 are computed as follows:

$$\mathbf{v}_i(m_1) = \begin{cases} w_l A_1{}' + w_b B_1{}' & if\ \underline{R}^{m_1} G_i \neq \emptyset \wedge R_b^{m_1} G_i \neq \emptyset \\ B_1{}' & if\ \underline{R}^{m_1} G_i = \emptyset \wedge R_b^{m_1} G_i \neq \emptyset, \\ A_1{}' & if\ \underline{R}^{m_1} G_i \neq \emptyset \wedge R_b^{m_1} G_i = \emptyset \end{cases} \tag{9}$$

where:

$$A_1{}' = \frac{\sum_{\mathbf{x}_j \in \underline{R}^{m_1} G_i} \left(u_{ij}(m_1)\right)^{m_1} \mathbf{x}_j}{\sum_{\mathbf{x}_j \in \underline{R}^{m_1} G_i} \left(u_{ij}(m_1)\right)^{m_1}}, \quad B_1{}' = \frac{\sum_{\mathbf{x}_j \in R_b^{m_1} G_i} \left(u_{ij}(m_1)\right)^{m_1} \mathbf{x}_j}{\sum_{\mathbf{x}_j \in R_b^{m_1} G_i} \left(u_{ij}(m_1)\right)^{m_1}}.$$

$u_{ij}(m_1)$ denotes the membership degree of pattern $\mathbf{x}_j$ belonging to the cluster with prototype $\mathbf{v}_i$ under the fuzzifier value m_1. Similarly, the prototypes under m_2 are obtained as follows:

$$\mathbf{v}_i(m_2) = \begin{cases} w_l A_2{}' + w_b B_2{}' & if\ \widetilde{\underline{R}}^{m_2} G_i \neq \emptyset \wedge \widetilde{R}_b^{m_2} G_i \neq \emptyset \\ B_2{}' & if\ \widetilde{\underline{R}}^{m_2} G_i = \emptyset \wedge \widetilde{R}_b^{m_2} G_i \neq \emptyset, \\ A_2{}' & if\ \widetilde{\underline{R}}^{m_2} G_i \neq \emptyset \wedge \widetilde{R}_b^{m_2} G_i = \emptyset \end{cases} \tag{10}$$

where:

$$A_2{}' = \frac{\sum_{\mathbf{x}_j \in \widetilde{\underline{R}}^{m_2} G_i} \left(u_{ij}(m_2)\right)^{m_2} \mathbf{x}_j}{\sum_{\mathbf{x}_j \in \widetilde{\underline{R}}^{m_2} G_i} \left(u_{ij}(m_2)\right)^{m_2}}, \quad B_2{}' = \frac{\sum_{\mathbf{x}_j \in \widetilde{R}_b^{m_2} G_i} \left(u_{ij}(m_2)\right)^{m_2} \mathbf{x}_j}{\sum_{\mathbf{x}_j \in \widetilde{R}_b^{m_2} G_i} \left(u_{ij}(m_2)\right)^{m_2}}.$$

$u_{ij}(m_2)$ denotes the membership degree of pattern $\mathbf{x}_j$ belonging to the cluster with prototype $\mathbf{v}_i$ under the fuzzifier value m_2. $\widetilde{\underline{R}}^{m_2} G_i$ has the same meaning as in formula (8), i.e., the core region of G_i with respect to m_2 based on the partition threshold obtained under the fuzzifier value m_1. $\widetilde{R}_b^{m_2} G_i = \underline{R}^{m_1} G_i - \widetilde{\underline{R}}^{m_2} G_i = \{\mathbf{x}_j | \mathbf{x}_j \in \underline{R}^{m_1} G_i \wedge \mathbf{x}_j \in R_b^{m_2} G_i\}$ covers the patterns that belong to the core region of G_i under m_1 and those also belong to the boundary region of G_i under m_2 according to the same partition threshold. The prototype of each cluster can be combined by the following principle:

$$\mathbf{v}_i = \beta_1 \mathbf{v}_i(m_1) + \beta_2 \mathbf{v}_i(m_2). \tag{11}$$

$\beta_1, \beta_2 \in [0, 1]$ and $\beta_1 + \beta_2 = 1$, they measure the importance of candidate results obtained under m_1 and m_2, respectively. $\mathbf{v}_i(m_2)$ can be considered as the modification for $\mathbf{v}_i(m_1)$ which may be spoiled by outliers, overlapping areas or the uncertainty generated by a single fuzzifier value. The contributions of the patterns in the core region of one cluster both under m_1 and m_2 will be enhanced, and the contributions of the patterns that removing from the core region under m_1 to the boundary region under m_2 will be reduced. The calculation of $\mathbf{v}_i(m_1)$ with formula (9) is the same as that in RFCM and SRFCM. It is easily seen that the formulations in (2) and (11) can be considered equivalent if $\beta_2 = 0$, namely, the notion of

rough-fuzzy clustering (at a single level) is a specific case of the notion of rough-fuzzy clustering based on multigranulation approximation regions.

The algorithm of multigranulation rough-fuzzy clustering based on shadowed sets can be refined as follows.

Step 5 plays an important role in the MSRFCM algorithm, based on which the embedded multigranulation approximation regions for each cluster can be formed. The proposed multigranulation structure cannot be constructed just by implementing the rough-fuzzy clustering methods with m_1 and m_2 separately, since the Proposition I is not satisfied in this situation.

In order to partition a pattern to one cluster, a hard-partition principle can be formed as follows:

$$u_{ij} = \frac{u_{ij}(m_1) + u_{ij}(m_2)}{2}. \tag{12}$$

If $u_{ij} > u_{kj}$, for $\forall k = 1, 2, \cdots, C$ and $i \neq k$, then $\mathbf{x}_j$ is assigned to the ith cluster G_i.

As discussed in Section 2.3, the optimization mechanism with formula (4) is one aspect to determine the symmetrical threshold in the model of three-way approximations of fuzzy sets based on the principle of uncertainty invariance. An optimization-based framework for constructing three-way approximations of fuzzy sets was investigated by Yao et al. [40], which includes the principle of minimum distance, and the optimization function can be interpreted as minimizing the total errors induced by three-way approximations. The detailed discussions can be found in [6,40]. For each cluster G_i, the corresponding optimization function based on the principle of minimum distance can be described as follows:

$$\eta_i = \min_{\eta}\left(V_i'\right) = \min_{\eta}\left|\sum_{j:u_{ij}\le\frac{\eta}{2}} u_{ij} + \sum_{j:u_{ij}\ge\frac{1+\eta}{2}}\left(1-u_{ij}\right) + \sum_{\frac{\eta}{2}<u_{ij}<\frac{1+\eta}{2}}\left|\eta-u_{ij}\right|\right|. \tag{13}$$

According to η_i, the core region and boundary region of cluster G_i can be determined as follows:

$$\begin{aligned}\underline{R}G_i &= \left\{\mathbf{x}_j | u_{ij} \ge \frac{1+\eta_i}{2}\right\},\\ R_bG_i &= \left\{\mathbf{x}_j | \frac{\eta_i}{2} < u_{ij} < \frac{1+\eta_i}{2}\right\}.\end{aligned} \tag{14}$$

Both of the formulas (4) and (13) are optimization mechanisms for determining the partition threshold. However, they have different semantical interpretations. The former is based on the principle of uncertainty invariance and the later is based on the principle of minimum distance. In this study, to compare the different optimization principles in the three-way approximations of fuzzy sets, we will also use the formulas (13) and (14) instead of formulas (4) and (5) in Steps 4 and 5 in the Algorithm I, and denote the corresponding algorithm as MSRFCMII. The results obtained by MSRFCM and MSRFCMII will be discussed in the experimental section in detail.

3.4. Computational complexity

The proposed algorithms MSRFCM and MSRFCMII definitely have higher complexity than FCM due to the computation time that is required for selecting the optimal partition threshold for each cluster. Assume the number of clusters is C, the number of patterns is N, the number of features of each pattern is M and the number of iterations is I. Since a pair of fuzzifier values are involved, the asymptotical time complexity for computing partition matrix is $O(2C^2NM)$. Assume the number of candidate partition threshold values is S, the computation for selecting partition thresholds with respect to m_1 is $O(SCN)$. The computation for dividing approximation regions under the value of m_2 is $O(CN)$ and the computation for prototypes is $O(2CN)$. Consequently, the computational complexity of the proposed method can be summarized as $O(I(2C^2NM + SCN + CN + 2CN))$, asymptotically, $O(I(C^2NM + SCN))$.

If multiple values of m_2 are used, the proposed algorithm can be implemented repeatedly. Assuming the number of the values of m_2 is K, the total computational complexity becomes $K \times O(I(C^2NM + SCN))$. Generally, $N \gg M$, $N \gg C$ and $N \gg I$, and the maximum value of fuzzifier parameter can be achieved within a finite number of steps according to formula (8), namely $N \gg K$, so the asymptotical time complexity of our proposed method approaches to $O(SN)$. Since no closed-form solution can be drawn for optimizing the partition thresholds, the enumerating methods are often exploited. For a practical problem with a big data set, if $N \gg S$, the computational complexity becomes $O(N)$.

4. Multilevel granulation-degranulation mechanisms

Some cluster validity indices are proposed to determine the true number of clusters as well as evaluating the clustering quality. The available validity indices include fuzzy and non-fuzzy versions. The fuzzy versions are often introduced based on the membership partitions, such as Xie–Beni index (XB) [30], fuzzy rand index (FRI) [3], fuzzy adjusted rand index (FARI) [3], and entropy based indices, such as normalized mutual information index (NMI) [10]. Non-fuzzy versions include relative separation index (RS) [8], Davies–Bouldin index (DB) [5], Dunns index (Dunn) [2], PBM-index(PBM) [18] and so on, which measure the separation and compactness of the clusters based on the obtained prototypes. However, the fuzzy versions of validity indices are sensitive to the selected fuzzifier values. Namely, they will change as increasing or decreasing the

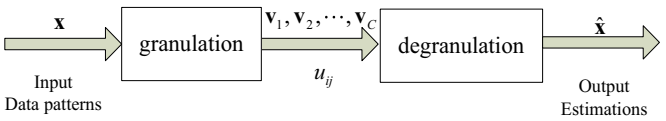

Fig. 8. A schematic view of the granulation-degranulation mechanisms.

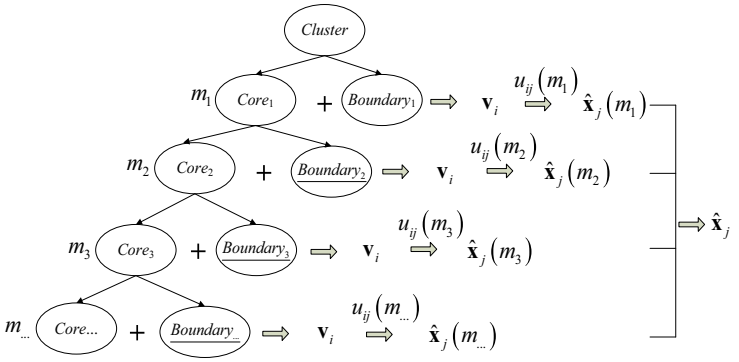

Fig. 9. Degranulation based on multigranulation approximation regions ($1 < m_1 < m_2 < m_3 < \cdots < m_{\max}$).

fuzzifier values under the obtained prototypes. In this case, the effectiveness of the fuzzy validity indices may be lost just by changing the fuzzifier values.

A novel validity index which is based on the "granulation-degranulation" mechanisms was introduced in [42]. The related "granulation-degranulation" mechanisms are schematically presented as in Fig. 8.

Essentially, clustering process can be treated as a granulation mechanism in which the information granules can be established, that are expected to reflect the original data as much as possible, involving prototypes and associated membership degrees. Subsequently, degranulation process is applied to original patterns and reconstructed based on the results of granulation section. Obviously, the results of degranulation, can be considered as the estimations based on the obtained prototypes and corresponding membership degrees, are expected to be close to the original patterns (Fig. 8).

Let $\hat{\mathbf{x}}_j$ denote the estimation of pattern $\mathbf{x}_j$ obtained by the "granulation-degranulation" mechanisms. An overall measure of the validity of the granulation-degranulation mechanisms can be quantified as follows:

$$Q = \sum_{j=1}^{N} \|\mathbf{x}_j - \hat{\mathbf{x}}_j\|^2, \tag{15}$$

where

$$\hat{\mathbf{x}}_j = \frac{\sum_{i=1}^{C} {u_{ij}}^m \mathbf{v}_i}{\sum_{i=1}^{C} {u_{ij}}^m}.$$

Different from the available rough-fuzzy clustering methods, which often use a single fuzzifier value, the proposed multigranulation rough-fuzzy clustering method involves a series of fuzzifier values. The uncertainty generated by the fuzzifier parameter can be captured at different levels of granularity during the granulation procedure. Similarly, the uncertainty generated by the fuzzifier parameter should also be addressed during the degranulation procedure, which is displayed in Fig. 9.

$\hat{\mathbf{x}}_j(m_k)$ denotes the estimation of $\mathbf{x}_j$ under the fuzzifier value m_k. The series of $\hat{\mathbf{x}}_j(m_1), \hat{\mathbf{x}}_j(m_2), \hat{\mathbf{x}}_j(m_3), \ldots, \hat{\mathbf{x}}_j(m_{...})$ can be considered as the estimations of $\mathbf{x}_j$ at different levels of granularity. The final degranulation result of $\mathbf{x}_j$, viz. $\hat{\mathbf{x}}_j$, can be formed as follows:

$$\hat{\mathbf{x}}_j = \sum_{k=1}^{k_{\max}} \beta_k \hat{\mathbf{x}}_j(m_k). \tag{16}$$

β_k has the same interpretation as that in formula (7). For a pair of fuzzifier values m_1 and m_2, the degranulation is formally described as follows:

$$\hat{\mathbf{x}}_j = \beta_1 \hat{\mathbf{x}}_j(m_1) + \beta_2 \hat{\mathbf{x}}_j(m_2), \tag{17}$$

where:

$$\hat{\mathbf{x}}_j(m_1) = \frac{\sum_{i=1}^{C} (u_{ij}(m_1))^{m_1} \mathbf{v}_i}{\sum_{i=1}^{C} (u_{ij}(m_1))^{m_1}}, \quad \hat{\mathbf{x}}_j(m_2) = \frac{\sum_{i=1}^{C} (u_{ij}(m_2))^{m_2} \mathbf{v}_i}{\sum_{i=1}^{C} (u_{ij}(m_2))^{m_2}}.$$

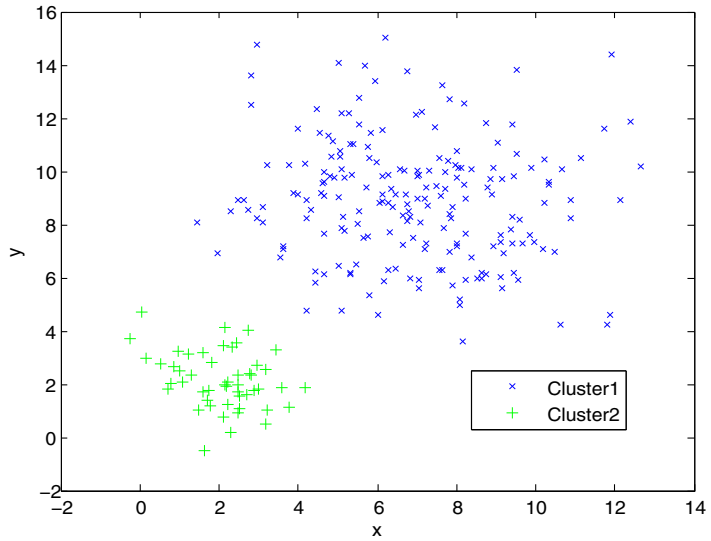

Fig. 10. Synthetic data set.

Table 1
The prototypes, partition thresholds and approximation regions of synthetic data set obtained by MSRFC.

	Prototype		Threshold		$card(\underline{R}^{m_1}G_i)/card(\widetilde{\underline{R}}^{m_2}G_i)/card(\widetilde{R}_b^{m_2}G_i)$	
	Prototype1	Prototype2	Threshold α_1	Threshold α_2	Cluster1	Cluster2
$m_1 = 2, m_2 = 2$	[7.2329, 9.2036]	[2.1387, 2.3345]	0.32589	0.32653	178 / 178 / 0	52 / 52 / 0
$m_1 = 2, m_2 = 3$	[7.259,9.2317]	[2.1619, 2.3836]	0.33517	0.33541	178 / 157 / 21	52 / 50 / 2
$m_1 = 2, m_2 = 4$	[7.2867, 9.3198]	[2.1685, 2.3802]	0.34611	0.34615	178 / 119 / 59	52 / 50 / 2
$m_1 = 2, m_2 = 5$	[7.215, 9.2465]	[2.1487, 2.2542]	0.32545	0.32535	178 / 55 / 123	52 / 41 / 11
$m_1 = 2, m_2 = 6$	[7.1607, 9.2244]	[2.1483, 2.2187]	0.32632	0.32653	179 / 41 / 138	51 / 33 / 18
$m_1 = 2, m_2 = 7$	[7.0998, 9.2593]	[2.2334, 2.1522]	0.3251	0.32565	179 / 19 / 160	51 / 19 / 32
$m_1 = 2, m_2 = 8$	[7.169, 9.2378]	[2.2653, 2.2363]	0.33139	0.33146	178 / 13 / 165	52 / 10 / 42
$m_1 = 2, m_2 = 9$	[7.1508, 9.0748]	[2.2243, 2.2299]	0.31732	0.3172	179 / 4 / 175	51 / 5 / 46
$m_1 = 2, m_2 = 10$	[7.1219, 9.0521]	[2.1887, 2.2348]	0.31034	0.31121	179 / 2 / 177	51 / 1 / 50
$m_1 = 2, m_2 = 11$	[7.0681, 9.0517]	[2.1497, 2.2397]	0.30641	0.30611	179 / 1 / 178	51 / 0 / 51

β_1 and β_2 have the same meaning as that in formula (11). $\hat{\mathbf{x}}_j(m_1)$ and $\hat{\mathbf{x}}_j(m_2)$ can be considered as the lower and upper estimations for $\mathbf{x}_j$ between which the uncertainty generated by the fuzzifier parameter can be detected. Under the same number of clusters, the smaller the value of Q is, the better the established granulation-degranulation model will be.

5. Experimental studies

In this section, some fuzzy clustering algorithms, including FCM [1], TFCM [7], SRFCM [42] and proposed MSRFCM and MSRFCMII, are compared with a synthetic data set and some data sets from UCI repository [11].

5.1. Synthetic data set

The synthetic data set with a mixture of Gaussian distributions is depicted in Fig. 10. It has two clusters with 50 data and 200 data, respectively. The means of *Cluster*1 and *Cluster*2 are $\mu_1 = [7, 9]$ and $\mu_2 = [2, 2]$, respectively.

The results obtained by running FCM, including prototypes and corresponding membership degrees, are utilized as the initial configurations for the implementation of TFCM, SRFCM, MSRFCM and MSRFCMII. The weighted value that evaluates 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. In addition, Euclidean distance is exploited, the maximum iteration number is set as 100 and the convergence condition satisfies $\|\mathbf{v}_i^{(t+1)} - \mathbf{v}_i^{(t)}\| < \varepsilon$ where t is an iterative step and ε is set as 0.0001 for all algorithms.

Taking $m_1 = 2$ as an example and this value maintains the same for all values of m_2 as tuning by one. The other values of m_1 can be explained similarly. The weighted values that measure the importance of results obtained at different levels of granularity are set as $\beta_1 = \beta_2 = 0.5$. The prototypes, partition threshold values and approximation regions obtained by MSRFCM are shown in Table 1.

From Table 1, it can be found that, $\widetilde{\underline{R}}^{m_2}G_i$ is decreasing as increasing the value of m_2. The number of patterns in $\widetilde{R}_b^{m_2}G_i$ is increasing on the contrary. This means that some patterns in the core region of one cluster under value m_1 are divided

Table 2
The prototypes, partition thresholds and approximation regions of synthetic data set obtained by MSRFCMII.

	Prototype		Threshold		$card(\underline{R}^{m_1}G_i)/card(\widetilde{R}^{m_2}G_i)/card(\widetilde{R}_b^{m_2}G_i)$	
	Prototype1	Prototype2	Threshold η_1	Threshold η_2	Cluster1	Cluster2
$m_1 = 2, m_2 = 2$	[7.9146, 9.3381]	[2.2141, 2.5117]	0.799	0.201	94 / 94 / 0	56 / 56 / 0
$m_1 = 2, m_2 = 3$	[7.6870, 9.7244]	[2.2443, 2.5239]	0.822	0.178	80 / 17 /63	56 / 52 / 4
$m_1 = 2, m_2 = 4$	[7.6528, 9.4556]	[22335, 2.5096]	0.829	0.171	83 / 2 / 81	56 / 51 /5
$m_1 = 2, m_2 = 5$	[7.9032, 9.3654]	[2.2316, 2.4999]	0.799	0.201	94 / 0 / 94	56 /50 / 6

Table 3
The comparative validity results of synthetic data set.

	$m = 1.1$ (or $m_1 = 1.1$)					
	RS	PBM	Dunn	Q	DB	*Err*
FCM	0.042434	37.684	0.024977	0.85959	0.56299	6
SRFCM	0.039538	34.938	0.024977	0.87296	0.58801	6
TFCM	0.042434	37.684	0.024977	0.85958	0.563	6
MSRFCM	0.0434	38.944	0.024977	0.83449	0.55071	**5**
MSRFCMII	**0.04476**	**39.783**	**0.024977**	**0.802628**	**0.546299**	6
$TFCM_{avg}$	0.036178	29.016	0.024033	0.889588	0.722434	17.6
$MSRFCM_{avg}$	0.044133	39.512	0.025477	0.826713	0.546916	5.9
$MSRFCII_{avg}$	0.04319	38.475	0.024977	0.824515	0.555960	6
	$m = 2$ (or $m_1 = 2$)					
FCM	0.03942	32.677	0.027555	0.78166	0.65577	15
SRFCM	0.046254	41.366	0.024977	0.77663	0.53388	6
TFCM	0.04961	43.725	0.00617	0.79657	0.52977	8
MSRFCM	0.049893	**44.831**	0.024977	0.76641	**0.51049**	**6**
MSRFCMII	**0.050481**	43.583167	**0.030533**	**0.737285**	0.544338	12
$TFCM_{avg}$	0.047978	41.75422	0.018118	0.799809	0.555521	10.1
$MSRFCM_{avg}$	0.048602	43.60189	0.024977	0.772343	0.518502	6
$MSRFCII_{avg}$	0.051124	43.496778	0.023013	0.742912	0.554643	14.3

into the boundary region of this cluster under value m_2 according to the same partition threshold. The multigranulation approximation regions for each cluster are formed as shown in Fig. 11.

As increasing the value of m_2, the core region of each cluster under m_2, which is obtained based on the same partition threshold optimized under m_1, is contracted towards to the cluster centroid. From Table 1, the core region of *Cluster*2 under m_2 becomes empty when the value of m_2 equals to 11. In this case, the value $m_2 = 10$ can be considered as the maximum value $m_{\max}$ ($\xi = 0$) with respect to the fixed value $m_1 = 2$. It is meaningless to increase the value of m_2 more than 10, since the representative capacity of the core region for *Cluster*2 is eliminated.

Similarly, the prototypes, partition threshold values and approximation regions obtained by MSRFCMII when fixing $m_1 = 2$ and varying m_2 added by 1 are shown in Table 2.

From Table 3, it can be found that the core region of *Cluster*1 under m_2 becomes empty when the value of m_2 equals to 5. It is meaningless to increase the value of m_2 more than 4 based on the algorithm MSRFCMII. In this way, the contraction speed of the core region of *Cluster*1 is faster than the one in MSRFCM, which are shown in Fig. 12.

In order to compare the validity of the proposed algorithms, the recognition rates (or called as classification accuracy) are illustrated in Fig. 13. The value of m_1 is tuned from 1.1 to 10 by one, and the best results obtained by MSRFCM, MSRFCMII and TFCM with fixing m_1 and varying m_2 to the corresponding maximum value by one are selected.

In Fig. 13, the notation m is involved in FCM and SRFCM and the notation m_1 is associated with TFCM, MSRFCM and MSRFCMII. It can be found that the methods TFCM, MSRFCM and MSRFCMII outperform SRFCM and FCM. FCM has the worst performance as increasing the fuzzifier value. MSRFCMII achieves the best performance over all values of m_1, only four patterns are classified incorrectly. This means that the multi-valued fuzzifier-based clustering methods can capture more uncertainties wrapped in the data, especially for the uncertainty generated by a single fuzzifier value.

Though the methods TFCM, MSRFCM and MSRFCMII perform better when dealing with the given synthetic data set, the performance of TFCM is more fluctuant than MSRFCM as fixing m_1 and varying m_2, that is demonstrated in Fig. 14.

From Fig. 14, it can be seen that the method TFCM performs unstably when fixing m_1 and varying m_2. It means different values of m_2 can capture the different degrees of uncertainty generated by a fixed fuzzifier value. However, there is a challenging work to select an optimal value m_2 when giving a fixed value m_1 under the framework of interval type-2 fuzzy clustering methods. Since the performance of the proposed method MSRFCM is more stable, the influence caused by m_2 can be reduced to some extent and the maximum value of m_2 can be supervised according to formula (8).

Moreover, according to Figs. 13 and 14, the algorithm MSRFCMII achieves the best performance in terms of recognition rate over different combinations of fuzzifier values m_1 and m_2. However, it can be found that MSRFCMII is more sensitive to

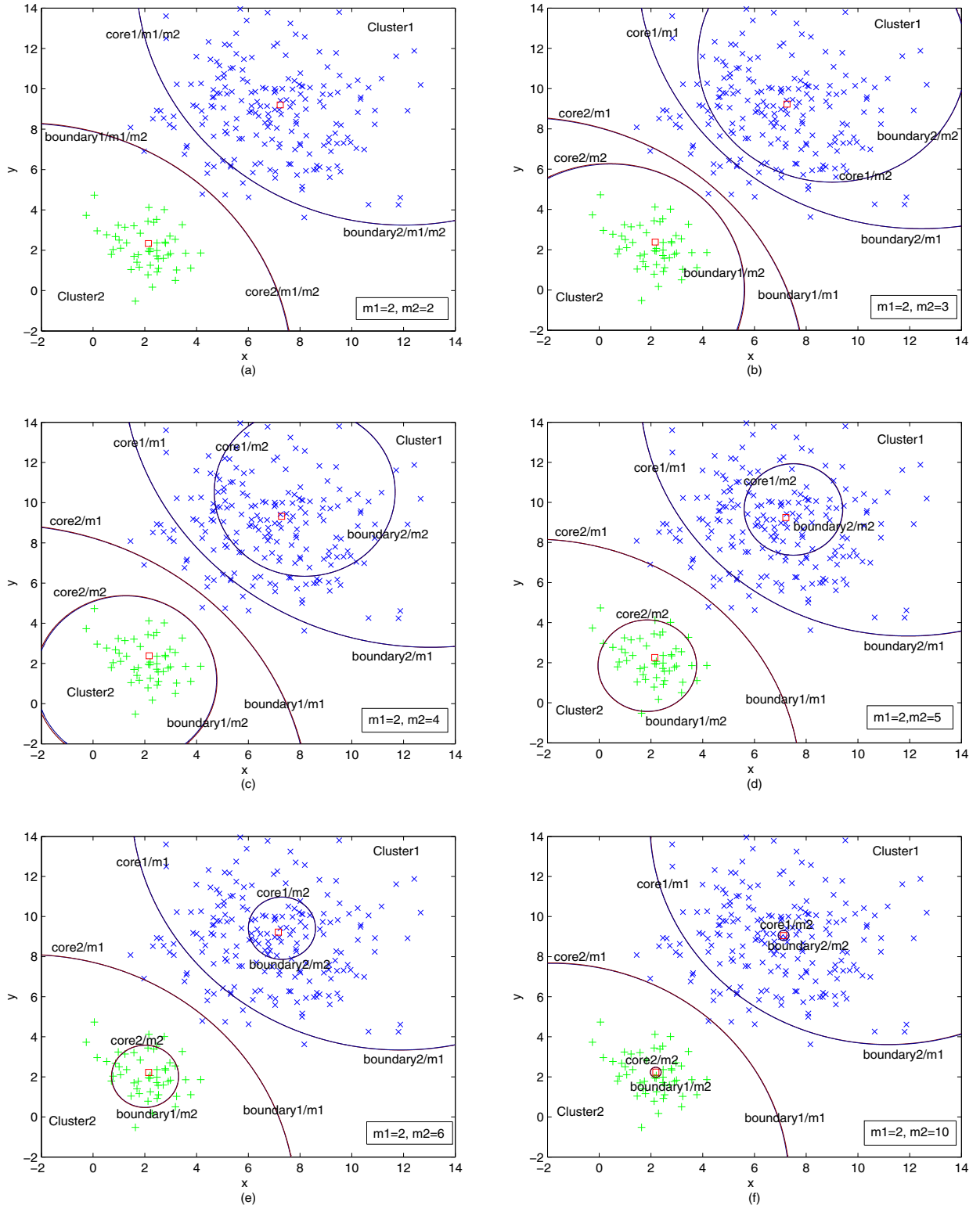

Fig. 11. The approximation region partitions of clusters under different values of m_2 (fixing $m_1 = 2$) with MSRFCM (a)$m_2 = 2$; (b)$m_2 = 3$; (c)$m_2 = 4$; (d)$m_2 = 5$; (e)$m_2 = 6$; (f)$m_2 = 10$. The red squares in the figures denote the prototypes. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

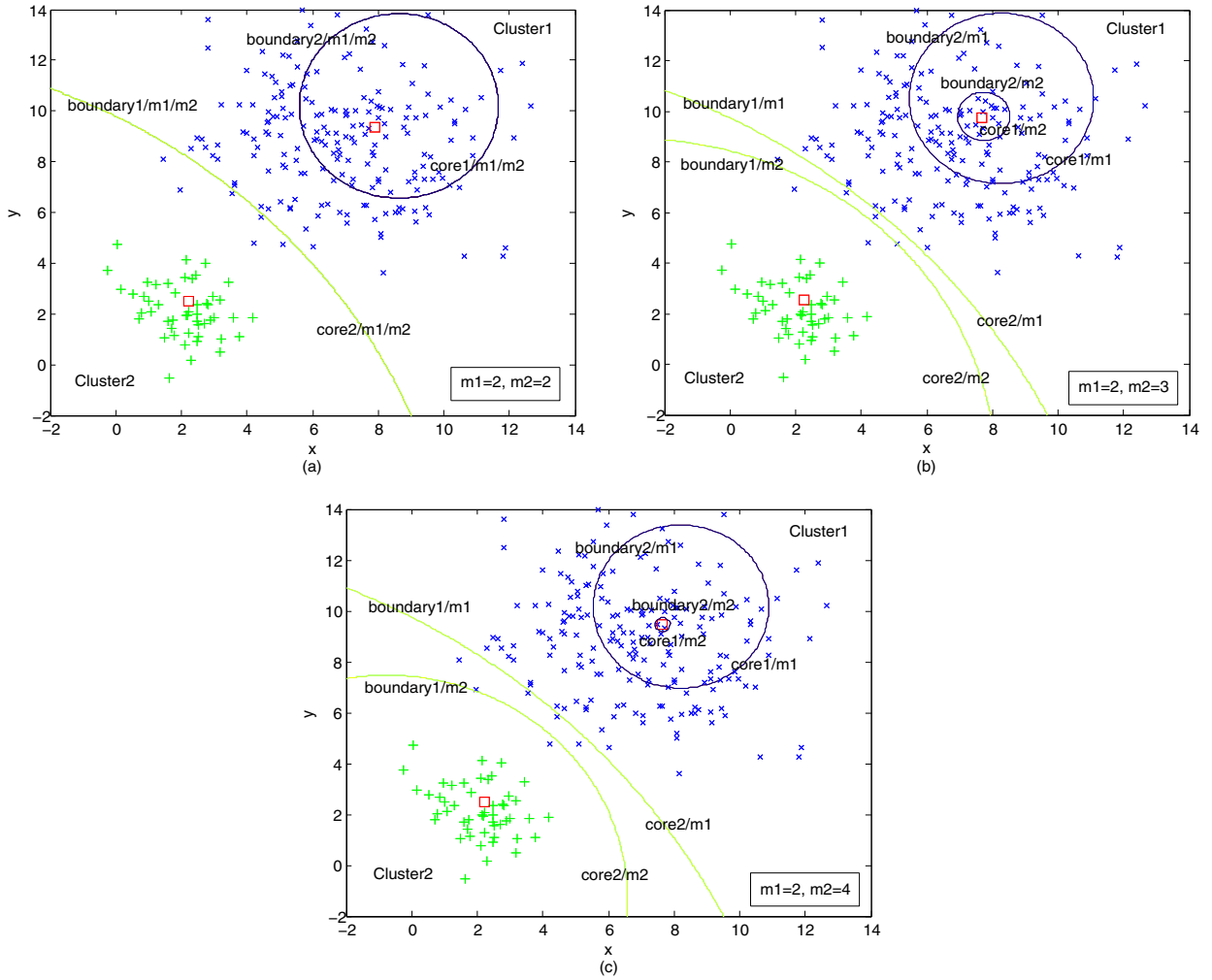

Fig. 12. The approximation region partitions of clusters under different values of m_2 (fixing $m_1 = 2$) with MSRFCMII (a)$m_2 = 2$; (b)$m_2 = 3$; (c)$m_2 = 4$. The red squares in the figures denote the prototypes. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the selected fuzzifier value m_1 than MSRFCM. For the dataset which has clusters with significantly different sizes (as shown in Fig. 10), more patterns will be divided into the boundary region with respect to the bigger cluster according to formulas (13) and (14) compared with MSRFCM. By contrast, the obtained core region of the smaller cluster is almost invariant since the patterns in its boundary region are divided from the bigger cluster. So the initialization of fuzzifier value m_1 is vital for the performance of MSRFCMII. In addition, the maximum value of fuzzifier m_2 can be quickly achieved in MSRFCMII. It means that the core region of one of the clusters tends to empty fast.

In order to compare the performance of the methods used in Fig. 12 comprehensively, some validity indices are utilized, including RS [8], DB [5], Dunn [2] and PBM [18]. The proposed validity index Q in formula (15) and the number of patterns, that are classified into wrong clusters (denoted as *Err*), are also compared. The larger the values of RS, PBM and Dunn as well as the smaller the values of DB, Q and *Err* are, the better the clustering methods will be. The best performance of TFCM, MSRFCM and MSRFCMII are chosen as fixing m_1 and varying m_2, and their average results of the validity indices among all values of m_2, denoted as $TFCM_{avg}$, $MSRFCM_{avg}$ and $MSRFCMII_{avg}$ respectively, are also presented as follows (Table 3).

From Table 2, the proposed multigranulation rough-fuzzy clustering based on shadowed sets has the best performance than the other methods in terms of the validity indices. Subsequently, it can be found that both of the average performance of MSRFCM and MSRFCMII are better than TFCM, the reason is that the results obtained by TFCM are fluctuant, and the results obtained by MSRFCM and MSRFCMII (especially MSRFCM) are far more stable as illustrated in Fig. 14.

Additionally, only $m_1 = 1.1$ and $m_1 = 2$ are selected for comparison. If the fuzzifier value m_1 is selected as a large value, most membership degrees of patterns belonging to clusters will approach to 0.5, thus the partition matrix is very soft (or fuzzy). In this case, the core regions of some clusters may become empty and most patterns will be partitioned into the

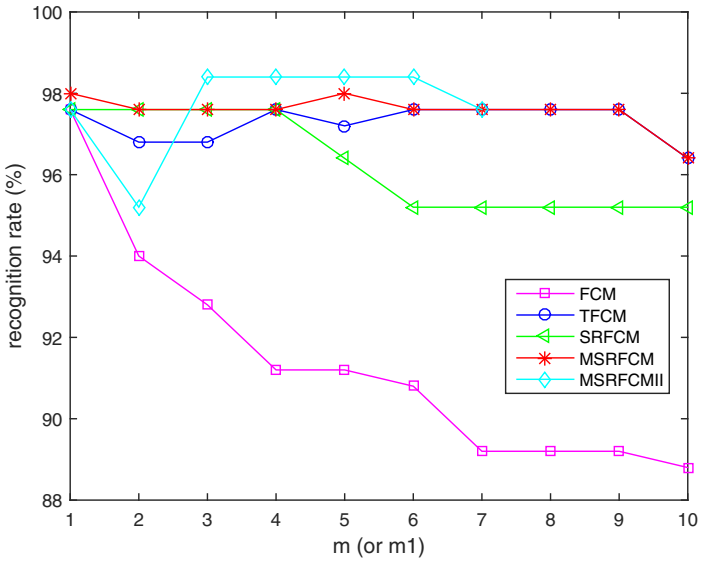

Fig. 13. Recognition rate of the synthetic data set.

Algorithm 1 Multigranulation rough-fuzzy clustering based on shadowed sets (MSRFCM).

Step 1: Assign random initial prototypes $\mathbf{v}_i (i = 1, 2, \cdots, C)$;
Step 2: Compute the partition matrix $u_{ij}(m_1)$ under the fuzzifier value m_1;
Step 3: Compute the partition matrix $u_{ij}(m_2)$ under the fuzzifier value m_2;
Step 4: Compute the optimal partition threshold $\alpha_i(m_1)$ for each cluster $G_i (i = 1, 2, \cdots, C)$ based on shadowed sets with formula (4). And according to $\alpha_i(m_1)$, determine the approximation regions $\underline{R}^{m_1} G_i$ and $R_b^{m_1} G_i$ for each cluster G_i with respect to $u_{ij}(m_1)$;
Step 5: According to $\alpha_i(m_1)$, determine the approximation regions $\widetilde{\underline{R}}^{m_2} G_i$ and $\widetilde{R}_b^{m_2} G_i$ for each cluster G_i with respect to $u_{ij}(m_2)$;
Step 6: Calculate the values of $\mathbf{v}_i(m_1)(i = 1, 2, \cdots, C)$ with formula (9);
Step 7: Calculate the values of $\mathbf{v}_i(m_2)(i = 1, 2, \cdots, C)$ with formula (10);
Step 8: Update the prototypes with formula (11);
Step 9: Repeat Steps 2–8 until convergence is reached.

boundary regions of these clusters. In this case, rough set-based clustering methods cannot obtain good results. So the fuzzifier value m_1 should be initialized with a relatively small value when using rough set and fuzzy set-based clustering methods.

5.2. UCI data sets

Some benchmark data sets from UCI storage are selected for experiments, including Ionosphere, Wine and Iris. The results are presented from Tables 4–6.

From the experimental results presented from Tables 4–6, the following conclusions can be drawn:

(1) The performances of MSRFCM and MSRFCMII are better than SRFCM, as well TFCM performs better than FCM in terms of most validity results. It can be explained that the multiple values of fuzzifier parameter involved in the clustering methods can capture the uncertainty generated by a single fuzzifier value sufficiently. In this way, the variations of the membership degrees under different fuzzifier values can be detected.

(2) The average performances of MSRFCM and MSRFCMII are better than TFCM. The improvement can be attributed to the fact that all patterns are partitioned into three approximation regions with respect to a fixed cluster based on shadowed sets, which helps capture the natural topology of the data. The importance of the patterns in the core regions is enhanced and the contribution of the patterns in the exclusion regions is removed when updating the prototypes. The formed multi-granularities of the core and boundary regions can effectively deal with the vagueness and uncertainty implicated in data from coarse to fine, especially for the uncertain patterns belonging to the core region of one cluster at the higher levels of granularity and those also belonging to the boundary region of this cluster at the lower levels of granularity.

(3) The algorithm MSRFCMII achieves better performance than the one obtained by MSRFCM in terms of most validity indices. It attributes to the optimization mechanism based on minimizing the global total error of three-way approximations for the corresponding fuzzy sets, which can be interpreted from the perspective of minimum distance within the framework

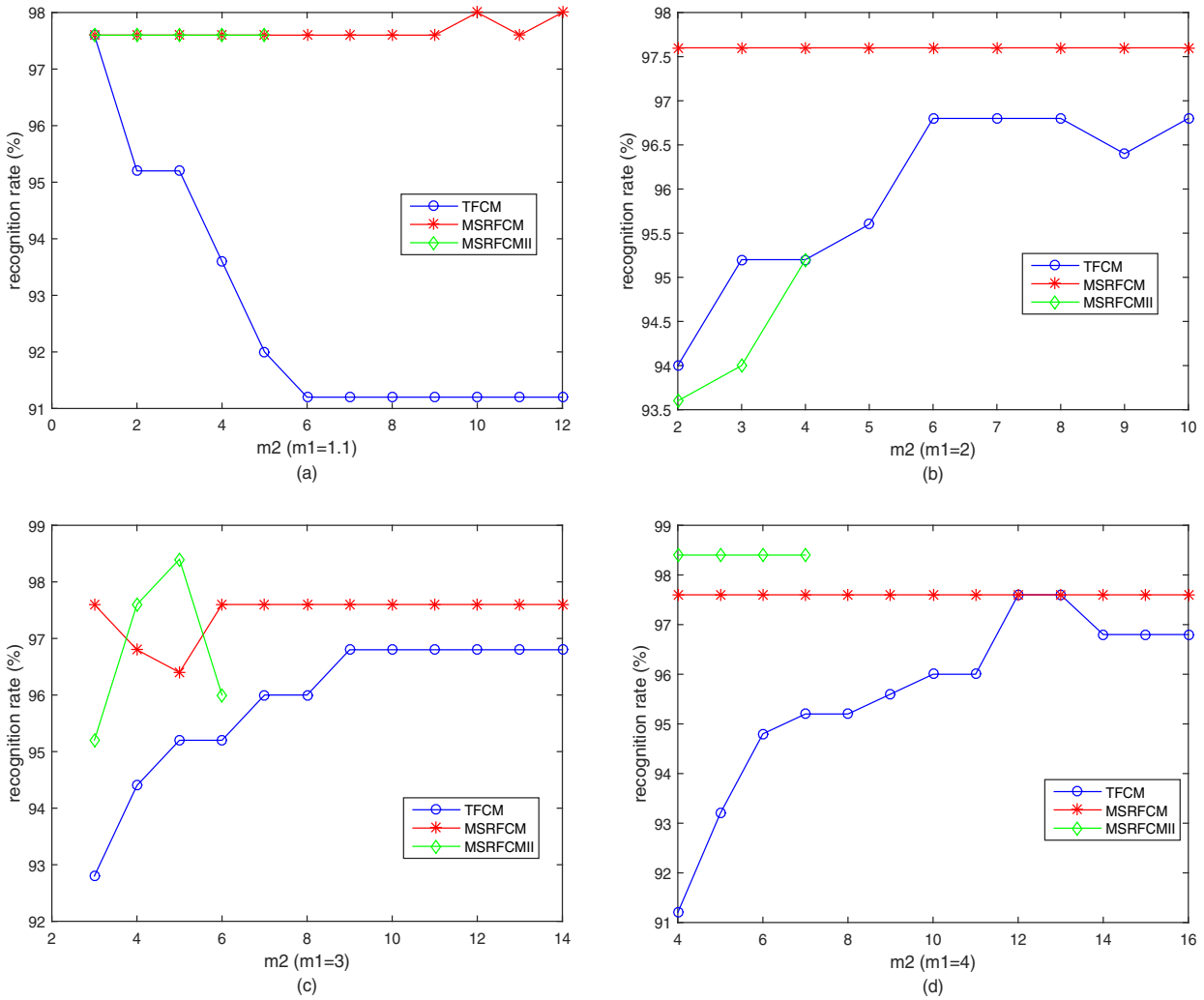

Fig. 14. Recognition rate of TFCM, MSRFCM and MSRFCMII as varying m_2 with respect to a fixed value of m_1. The value of m_2 is terminated until the maximum value $m_{\max}$ $(\xi = 0)$ with respect to the fixed value m_1 is achieved in MSRFCM and MSRFCMII. (a)fixing $m_1 = 1.1$; (b)fixing $m_1 = 2$; (c)fixing $m_1 = 3$; (d)fixing $m_1 = 4$.

Table 4
The comparative validity results of Ionosphere.

	$m = 1.1$ (or $m_1 = 1.1$)					
	RS	PBM	Dunn	Q	DB	*Err*
FCM	0.00713	3.4786	0.071382	**25.535**	1.5357	103
SRFCM	0.00707	3.4382	0.071382	25.547	1.5445	103
TFCM	0.00676	3.2398	0.071382	25.641	1.5958	105
MSRFCM	0.0072	3.608	0.08165	26.055	1.499	**101**
MSRFCMII	**0.00849**	**4.3572**	**0.08165**	25.629	**1.3659**	102
$TFCM_{avg}$	0.00721	3.21231	0.071552	27.5477	1.60287	109
$MSRFCM_{avg}$	0.00777	3.87206	0.072357	25.8191	1.45611	105.4
$MSRFCMII_{avg}$	0.00796	4.00448	0.072784	25.7084	1.43203	103.7
	$m = 2$ (or $m_1 = 2$)					
FCM	0.00452	1.9164	0.08165	27.626	2.0598	105
SRFCM	0.0089	4.3892	0.071382	25.879	1.3805	111
TFCM	0.00535	2.4435	0.08165	26.917	1.8198	103
MSRFCM	0.00865	4.5182	**0.08165**	**26.331**	1.3374	**102**
MSRFCMII	**0.01488**	**6.5158**	0.05469	27.9515	**1.1684**	139
$TFCM_{avg}$	0.00798	3.77363	0.076307	26.5508	1.53657	111.2
$MSRFCM_{avg}$	0.01	4.95189	0.07274	26.3098	1.30217	111.8
$MSRFCMII_{avg}$	0.017084	6.33796	0.028174	27.8597	1.17704	153.9

Table 5
The comparative validity results of Wine.

	$m = 1.1$ (or $m_1 = 1.1$)					
	RS	PBM	Dunn	Q	DB	*Err*
FCM	0.19786	144870	0.0032949	7.0048	1.4274	6
SRFCM	0.18569	138310	0.0032949	7.03	1.4057	7
TFCM	0.19186	136880	0.0032949	6.9942	1.5585	6
MSRFCM	0.21778	162630	0.0032949	**6.8281**	1.1985	6
MSRFCMII	**0.2311**	**174822**	**0.0032949**	6.83922	**1.1007**	**6**
$TFCM_{avg}$	0.20557	141070	0.0032949	7.41532	1.63763	9.1
$MSRFCM_{avg}$	0.205378	152348	0.0032806	7.02404	1.29549	6.1
$MSRFCMII_{avg}$	0.214272	158602	0.0032949	7.00099	1.27334	6
	$m = 2$ (or $m_1 = 2$)					
FCM	0.1249	84934	0.0032949	7.6861	2.6729	6
SRFCM	0.24799	190920	0.0039376	**6.7423**	1.0257	5
TFCM	0.20202	138400	0.0032949	6.9563	1.6974	6
MSRFCM	0.25312	196010	**0.0039376**	6.9197	1.0225	**5**
MSRFCMII	**0.25799**	**198981**	0.0032949	6.9296	**1.02094**	6
$TFCM_{avg}$	0.209305	145250.4	0.0033891	7.13764	1.7777	7.6
$MSRFCM_{avg}$	0.245763	190324	0.0036771	6.94073	1.063729	6.2
$MSRFCMII_{avg}$	0.24747	190789	0.0034734	6.950143	1.062628	6.3

Table 6
The comparative validity results of Iris.

	$m = 1.1$ (or $m_1 = 1.1$)					
	RS	PBM	Dunn	Q	DB	*Err*
FCM	0.41775	20.925	0.045222	0.91031	0.73848	25
SRFCM	0.41426	20.587	0.045222	0.9143	0.76068	25
TFCM	0.41205	20.804	0.045222	0.89894	0.7277	25
MSRFCM	**0.42483**	**21.64**	0.045222	0.87508	0.70327	25
MSRFCMII	0.42021	21.6183	**0.045222**	**0.87479**	**0.67756**	**25**
$TFCM_{avg}$	0.404882	19.4308	0.0590778	0.942424	0.746975	26.5
$MSRFCM_{avg}$	0.417325	21.0611	0.0458811	0.895685	0.719454	24.9
$MSRFCMII_{avg}$	0.421641	21.50186	0.055734	0.891839	0.702759	25.3
	$m = 2$ (or $m_1 = 2$)					
FCM	0.40418	20.395	0.045222	0.89867	0.75115	24
SRFCM	**0.42356**	**21.731**	0.045222	0.83879	0.65972	25
TFCM	0.40801	20.796	0.045222	0.87952	0.72303	24
MSRFCM	0.41112	20.018	0.051813	**0.82688**	0.65574	24
MSRFCMII	0.39298	18.8033	**0.051813**	0.845692	**0.6457**	**22**
$TFCM_{avg}$	0.363399	20.4539	0.0682932	0.895771	0.740238	25.5
$MSRFCM_{avg}$	0.42016	21.4091	0.0458811	0.856106	0.65027	24.2
$MSRFCMII_{avg}$	0.393982	18.9009	0.051813	0.848854	0.647175	22.3

of three-way decisions. Additionally, the contraction speed of the core regions of some clusters in MSRFCMII are faster than MSRFCM as increasing the values of fuzzifier m_2. It means that the new formed core regions under fuzzifier m_2 are closely located around the natural centroids of these clusters, which can improve the prototype calculations.

(4) The performance of FCM and SRFCM are seriously influenced by the selected fuzzifier values as dealing with the same data set. It means the calculated prototypes and corresponding membership degrees may be spoiled by an unreasonable fuzzifier value. Therefore, the uncertainty caused by the fuzzifier parameter needs to be carefully addressed during the algorithm implementations. The best performance of SRFCM is better than FCM among different fuzzifier values which also illustrate the essence that patterns divided into different approximation regions based on shadowed sets can help capture better the overall topology of the data.

6. Conclusions

The management of uncertain information in a data set, such as overlapping patterns, the uncertainty generated by model parameters and the vagueness arising in boundary regions, is crucial for rough set and fuzzy set-based clustering methods. A generalized multigranulation rough-fuzzy clustering method based on shadowed sets is introduced in this paper. All patterns are partitioned into different approximation regions with an optimization process based on shadowed sets which can capture the natural topology of data automatically. The formed multigranulation approximation regions can detect the uncertainty generated by a single fuzzifier value. So that the uncertain information implicated in data can be handled

as much as possible, and the prototype calculations at a single level of granularity can be modified. The improvement of the proposed method is illustrated in terms of several validity indices as compared with other C-means clustering methods, such as fuzzy C-means, type-2 fuzzy C-means and rough-fuzzy C-means.

Shadowed sets is an example of three-way approximations of fuzzy sets according to the framework of three-way decision theory. It can be viewed as a special model of granular computing which can analyze problems from different levels of information granularity. Two kinds of optimization mechanisms are used to form shadows in this paper. The quality of the proposed clustering approach can be practically improved based on the semantical optimization principle, i.e., the principle of minimum distance. The balance between the representative capacity of the model and the size of granules discussed in this paper can be treated as a reference for selecting a suitable level of granularity when solving the problems. This principle can also be employed for establishing multigranulation probabilistic rough set models with three-way decisions which is under our study and will be reported shortly.

Acknowledgments

The authors are grateful to the anonymous referees for their valuable comments and suggestions. This work is supported by the Postdoctoral Science Foundation of China (No. 2017M612736, 2017T100645), 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”, partially supported by the National Natural Science Foundation of China (No. 61703283, 61573248, 61672358, 61773328).

Appendix

The proofs of Proposition I:

Proof. Since $g(d_1, m_1) = g(d_2, m_2)$, it has:

$$\frac{1}{1+(\frac{d_1}{1-d_1})^{\frac{2}{m_1-1}}} = \frac{1}{1+(\frac{d_2}{1-d_2})^{\frac{2}{m_2-1}}} \tag{18}$$

Obviously, it has $(\frac{d_1}{1-d_1})^{\frac{2}{m_1-1}} = (\frac{d_2}{1-d_2})^{\frac{2}{m_2-1}}$. Since $m_1 - 1 > 0$ and $0 < d_1 < 1$, so $(\frac{d_1}{1-d_1})^{\frac{2}{m_1-1}} > 0$ is satisfied. Similarly, $(\frac{d_2}{1-d_2})^{\frac{2}{m_2-1}} > 0$. By using a logarithm function to formula (18), it has:

$$\frac{2}{m_1-1}\log\left(\frac{d_1}{1-d_1}\right) = \frac{2}{m_2-1}\log\left(\frac{d_2}{1-d_2}\right) \tag{19}$$

Since $1 < m_1 < m_2$, so $\frac{2}{m_1-1} > \frac{2}{m_2-1} > 0$, the signs of $\log(\frac{d_1}{1-d_1})$ and $\log(\frac{d_2}{1-d_2})$ must be the same.

(1) If $\log(\frac{d_1}{1-d_1}) < 0$, and $\log(\frac{d_2}{1-d_2}) < 0$, namely, $0 < d_1$, $d_2 < 0.5$. According to (19), it has: $\log(\frac{d_1}{1-d_1}) > \log(\frac{d_2}{1-d_2})$. So $\frac{d_1}{1-d_1} > \frac{d_2}{1-d_2}$. Further, $\frac{1}{\frac{1}{d_1}-1} > \frac{1}{\frac{1}{d_2}-1}$. Consequently, it has $d_1 > d_2$.

(2) If $\log(\frac{d_1}{1-d_1}) > 0$, and $\log(\frac{d_2}{1-d_2}) > 0$, namely, $0.5 < d_1$, $d_2 < 1$. According to (19), it has: $\log(\frac{d_1}{1-d_1}) < \log(\frac{d_2}{1-d_2})$. So $\frac{d_1}{1-d_1} < \frac{d_2}{1-d_2}$. Further, $\frac{1}{\frac{1}{d_1}-1} < \frac{1}{\frac{1}{d_2}-1}$. Consequently, it has $d_1 < d_2$. □

References

[1] J.C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Kluwer Academic Publishers, Norwell, MA, USA, 1981.
[2] J.C. Bezdek, N.R. Pal, Some new indexes of cluster validity, IEEE Trans. Syst. Man Cybern. B 28 (3) (1998) 301–315.
[3] R.J.G.B. Campello, A fuzzy extension of the rand index and other related indexes for clustering and classification assessment, Pattern Recognit. Lett. 28 (7) (2007) 833–841.
[4] D. Ciucci, D. Dubois, A map of dependencies among three-valued logics, Inf. Sci. 250 (2013) 162–177.
[5] D.L. Davies, D.W. Bouldin, A cluster separation measure, IEEE Trans. Pattern Anal. Mach. Intell. 1 (2) (1979) 224–227.
[6] X.F. Deng, Y.Y. Yao, Decision-theoretic three-way approximations of fuzzy sets, Inf. Sci. 279 (2014) 702–715.
[7] C. Hwang, F.C.H. Rhee, Uncertain fuzzy clustering: interval type-2 fuzzy approach to c-means, IEEE Trans. Fuzzy Syst. 15 (1) (2007) 107–120.
[8] M.R.N. Kalhori, M.H.F. Zarandi, Interval type-2 credibilistic clustering for pattern recognition, Pattern Recognit. 48 (11) (2015) 3652–3672.
[9] G.M. Lang, D.Q. Miao, M.J. Cai, Three-way decision approaches to conflict analysis using decision-theoretic rough set theory, Inf. Sci. 406–407 (2017) 185–207.
[10] Y. Lei, J.C. Bezdek, J. Chan, N.X. Vinh, S. Romano, J. Bailey, Extending information-theoretic validity indices for fuzzy clustering, IEEE Trans. Fuzzy Syst. 25 (4) (2016) 1013–1018.
[11] M. Lichman, UCI machine learning repository [http://archive.ics.uci.edu/ml], 2013, Irvine, CA: University of California, School of Information and Computer Science.
[12] O. Linda, M. Manic, General type-2 fuzzy c-means algorithm for uncertain fuzzy clustering, IEEE Trans. Fuzzy Syst. 20 (5) (2012) 883–897.
[13] P. Lingras, C. West, Interval set clustering of web users with rough k-means, J. Intell. Inf. Syst. 23 (1) (2004) 5–16.
[14] J. MacQueen, Some methods for classification and analysis of multivariate observations, in: Proc. Fifth Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, 1967, pp. 281–297.

[15] P. Maji, S.K. Pal, Rough-fuzzy clustering for grouping functionally similar genes from microarray data, IEEE/ACM Trans. Comput. Biol. Bioinf. 10 (2013) 286–299.
[16] P. Maji, S. Roy, Rough-fuzzy clustering and multiresolution image analysis for text-graphics segmentation, Appl. Soft Comput. 30 (2015) 705–721.
[17] S. Mitra, H. Banka, W. Pedrycz, Rough-fuzzy collaborative clustering, IEEE Trans. Syst. Man. Cybern. B 36 (4) (2006) 795–805.
[18] M.K. Pakhira, S. Bandyopadhyay, U. Maulik, Validity index for crisp and fuzzy clusters, Pattern Recognit. 37 (3) (2004) 487–501.
[19] Z. Pawlak, Rough sets, Int. J. Inf. Comput. Sci. 11 (5) (1982) 314–356.
[20] Z. Pawlak, A. Skowron, Rudiments of rough sets, Inf. Sci. 177 (2007) 3–27.
[21] W. Pedrycz, Shadowed sets: representing and processing fuzzy sets, IEEE Trans. Syst. Man. Cybern. B 28 (1998) 103–109.
[22] W. Pedrycz, G. Bukovich, Granular computing with shadowed sets, Int. J. Intell. Syst. 17 (2002) 173–197.
[23] W. Pedrycz, Interpretation of clusters in the framework of shadowed sets, Pattern Recognit. Lett. 26 (2005) 2439–2449.
[24] W. Pedrycz, Granular computing: an emerging paradigm, J. Uncertain Syst. 1 (2007) 38–61.
[25] W. Pedrycz, From fuzzy sets to shadowed sets: interpretation and computing, Int. J. Intell. Syst. 24 (2009) 48–61.
[26] G. Peters, Some refinements of rough k-means clustering, Pattern Recognit. 39 (8) (2006) 1481–1491.
[27] Y.H. Qian, J.Y. Liang, Y.Y. Yao, C.Y. Dang, MGRS: A multi-granulation rough set, Inf. Sci. 180 (2010) 949–970.
[28] J.P. Sarkar, I. Saha, U. Maulik, Rough possibilistic type-2 fuzzy c-means clustering for MR brain image segmentation, Appl. Soft Comput. 46 (C) (2016) 527–536.
[29] H. Tahayori, A. Sadeghian, W. Pedrycz, Induction of shadowed sets based on the gradual grade of fuzziness, IEEE Trans. Fuzzy Syst. 21 (2013) 937–949.
[30] X.L. Xie, G.A. Beni, A validity measure for fuzzy clustering, IEEE Trans. Pattern Anal. Mach. Intell. 13 (13) (1991) 841–847.
[31] X.B. Yang, S.P. Xu, H.L. Dou, X.N. Song, H.L. Yu, J.Y. Yang, Multigranulation rough set: a multiset based strategy, Int. J. Comput. Int. Sys. 10 (2017) 277–292.
[32] Y.Y. Yao, Granular computing: basic issues and possible solutions, in: Proc. Fifth Int. Conf. Inf. Sci., vol. 1, 2000, pp. 186–189.
[33] Y.Y. Yao, Probabilistic rough set approximations, Int. J. Approximation Reasoning 49 (2008) 255–271.
[34] Y.Y. Yao, Three-way decisions with probabilistic rough sets, Inf. Sci. 180 (2010) 341–353.
[35] Y.Y. Yao, The superiority of three-way decisions in probabilistic rough set models, Inf. Sci. 181 (6) (2011) 1080–1096.
[36] Y.Y. Yao, An outline of a theory of three-way decisions, in: Proc. Rough Sets Curr. Trends Comput., LNCS, vol. 7413, 2012, pp. 1–17.
[37] Y.Y. Yao, Y.H. She, Rough set models in multigranulation spaces, Inf. Sci. 327 (2015) 40–56.
[38] Y.Y. Yao, Rough-set concept analysis: Interpreting RS-definable concepts based on ideas from formal concept analysis, Inf. Sci. 346–347 (2016) 442–462.
[39] Y.Y. Yao, Three-way decisions and cognitive computing, Cognit. Comput. 8 (2016) 543–554.
[40] Y.Y. Yao, S. Wang, X.F. Deng, Constructing shadowed sets and three-way approximations of fuzzy sets, Inf. Sci. 412–413 (2017) 132–153.
[41] L.A. Zadeh, Some reflections on soft computing, granular computing and their roles in the conception, design and utilization of information/intelligent systems, Soft Comput. 2 (1998) 23–25.
[42] J. Zhou, W. Pedrycz, D.Q. Miao, Shadowed sets in the characterization of rough-fuzzy clustering, Pattern Recognit. 44 (8) (2011) 1738–1749.