

## Rough possibilistic C-means clustering based on multigranulation approximation regions and shadowed sets



Jie Zhou<sup>a,b</sup>, Zhihui Lai<sup>a,b</sup>, Can Gao<sup>\*,a,b</sup>, Duoqian Miao<sup>c</sup>, Xiaodong Yue<sup>d</sup>

<sup>a</sup> College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, Guangdong 518060, China

<sup>b</sup> Institute of Textile and Clothing, The Hong Kong Polytechnic University, Kowloon, Hong Kong

<sup>c</sup> Department of Computer Science and Technology, Tongji University, Shanghai 201804, China

<sup>d</sup> School of Computer Engineering and Science, Shanghai University, Shanghai 200444, China

### ARTICLE INFO

#### Keywords:

Shadowed sets

Rough sets

Granular computing

Possibilistic C-means

Multigranulation approximation regions

### ABSTRACT

The management of uncertain information in a data set is crucial for clustering models. In this study, we present a rough possibilistic C-means clustering approach based on multigranulation approximation regions and shadowed sets, which can handle the uncertainties implicated in data and generated by the model parameters simultaneously. In particular, all patterns are first partitioned into three approximation regions with respect to a fixed cluster according to their possibilistic membership degrees based on shadowed set theory, which can help capture the natural topology of the data, especially when dealing with outliers and noisy data. The multigranulation approximation regions of each cluster can then be formed under a series of fuzzifier values, where the uncertainty caused by a specific fuzzifier value can be detected based on variations in the approximation regions with different levels of granularity. We also introduce a framework for updating prototypes based on ensemble strategies to attenuate the distortions due to iteration during clustering procedures. Finally, an adaptive mechanism is developed for dynamically adjusting scale parameters based on the notion of the maximal compatible regions of clusters. By integrating various granular computing techniques, i.e., rough sets, fuzzy sets, shadowed sets, and the notion of multigranulation, the uncertainties implicated in data and produced by model parameters can be adequately addressed, and the possibilistic membership values involved make the method sufficiently robust to deal with noisy environments. The improved performance of the proposed approach was demonstrated in experiments based on the comparisons with other available fuzzy and possibilistic clustering methods.

### 1. Introduction

Clustering is a data analysis method for finding natural groups implicated in data, which has been successfully applied in fields such as medical sciences, image segmentation, text mining, and network security [1]. The main task in clustering involves dividing an unlabeled data set  $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ ,  $\mathbf{x}_j \in \mathcal{R}^M$ , into  $C(1 < C < N)$  subgroups  $\{G_1, G_2, \dots, G_C\}$  such that high intra-cluster similarities and low inter-cluster similarities can be obtained. The structural topology of the data is not available; thus, one of the main challenges that affect clustering techniques is their capacity to deal with uncertain information in data, such as overlapping partitions, outliers, noisy data, uncertain data distributions, and the uncertainties caused by model parameters.

Fuzzy clustering, especially fuzzy C-means (FCM) [2], utilizes a partition matrix  $U = \{\mu_{ij}\}$  to evaluate the relative degree of each pattern

belonging to each cluster, so the overlapping partitions can be effectively described. The main drawback of FCM is the sensitivities to noisy patterns that may contaminate the calculations of the corresponding prototypes and membership degrees. However, it is indeed true that the techniques used in engineering and scientific application scenarios need to be robust, namely have the ability to tolerate noises and outliers [3].

Possibilistic approaches to clustering, especially possibilistic C-means (PCM) [4], have been proposed to address the problems associated with the constraints on membership degrees used in FCM. PCM uses a possibilistic partition to measure the absolute degree of typicality for each pattern in each cluster. The patterns far from the prototypes belong to clusters with very small possibilistic memberships, so PCM is more robust than FCM when handling data in noisy environments. The objective function of PCM can be considered as a collection of  $C$ -independent sub-objective functions. If the initializations of the iteration

\* Corresponding author at: College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, Guangdong 518060, China.

E-mail addresses: [jie\\_jpu@163.com](mailto:jie_jpu@163.com) (J. Zhou), [lai\\_zhi\\_hui@163.com](mailto:lai_zhi_hui@163.com) (Z. Lai), [2005gaocan@163.com](mailto:2005gaocan@163.com) (C. Gao), [dqmiao@tongji.edu.cn](mailto:dqmiao@tongji.edu.cn) (D. Miao), [yswantfly@shu.edu.cn](mailto:yswantfly@shu.edu.cn) (X. Yue).

<https://doi.org/10.1016/j.knosys.2018.07.007>

Received 14 February 2018; Received in revised form 3 July 2018; Accepted 4 July 2018

Available online 05 July 2018

0950-7051/ © 2018 Elsevier B.V. All rights reserved.

implementations are not adequate, the performance of PCM tends toward coincidental clusters [5]. To overcome this problem, Pal et al. [6] proposed a possibilistic-fuzzy C-means model (PFCM) that integrates both the typicality values and fuzzy membership degrees in clustering processes. Zhang et al. [7] presented an improved PCM algorithm that also involves both possibilistic and fuzzy memberships. Xenaki et al. [8,9] introduced sparse PCM to deal with closely located clusters. Yu et al. [10] proposed cutset-type possibilistic clustering (CPCM) where a cluster core can be generated from a  $\beta$ -cutset for each cluster. By integrating the advantages of rough set theory [11] and the notion of typicality, Maji et al. [12] developed rough-fuzzy PCM where whole patterns are divided into three approximation regions with respect to a fixed cluster, and the calculations of the prototypes are only related to the core and boundary regions of clusters instead of all data. Some new parameters are employed in these methods and the performance of the corresponding models is directly influenced by these new parameters. However, the uncertainty caused by the model parameters was rarely discussed in previous studies.

There are two important types of parameters in PCM, i.e., the scale parameters  $\{\gamma_i\}$  and the fuzzifier value  $m$ . Krishnapuram and Keller [5] explained the significance of these two parameters in detail. In most previous studies, the scale parameters  $\{\gamma_i\}$  were considered equal for all clusters and they were kept unchanged when updating the typicality values during iterations. Using this method, the ability of PCM will drastically decrease when dealing with closely located clusters that have significantly different variances or different sizes. Xenaki et al. [13] presented a novel adaptive possibilistic clustering algorithm where the scale parameters  $\{\gamma_i\}$  are adapted as the algorithm evolves and expressed in terms of the mean absolute deviation from the mean for the patterns that are most compatible with the clusters. It was demonstrated that the adaptive adjustment of the scale parameters is more suitable for determining cluster structures with different sizes and densities during the execution of the algorithm.

The value of the fuzzifier  $m$  has a major impact on the prototype and partition matrix calculations in FCM and in PCM. A predefined value of  $m$  is often used in possibilistic clustering methods [14,15]. However, it is difficult to express the uncertain notion of fuzziness in a given data set using only one fuzzifier value. Thus, to manage the uncertainty generated by the fuzzifier  $m$ , Hwang and Rhee [16] proposed interval type-2 FCM, which extends a pattern set to interval type-2 fuzzy sets using a pair of fuzzifier values,  $m_1$  and  $m_2$ , which creates a footprint of uncertainty caused by the fuzzifier parameter. Rubio et al. [17] extended PCM algorithms by using type-2 fuzzy logic techniques. However, the values of  $m_1$  and  $m_2$  depend mainly on subjective selection or enumeration in previously proposed methods [18,19] and the results obtained require further interpretation.

Granular computing is an emerging computing paradigm for information processing, which aims to solve computational intelligence problems by simulating human cognitive behaviors [20–22]. In particular, the notion of multigranulation [23–28] in granular computing is often employed for solving human-centric problems. Truong et al. [29] described a PCM algorithm based on granular computing. Li et al. [30] presented a hierarchical cluster ensemble model based on knowledge granulation, which can handle cluster ensemble problems by applying ensemble learning for knowledge granulation. Fujita et al. [31] introduced the first method to systematically integrate state of the art solutions of granular computing into different phases of resilience analysis for critical infrastructures, where they proposed a challenging real domain to contextualize some recent granular computing results. Jing et al. [32] developed an incremental feature selection method with a multi-granulation view to process large-scale data sets, where the knowledge granularity can be formed based on sub-decision systems from the perspective of multigranulation. Xu et al. [33] proposed two types of generalized multigranulation double-quantitative decision-theoretic rough sets, which provide the theoretical foundation for making decisions and extending generalized multigranulation rough set

models. The multigranulation method can interpret the results obtained from the perspective of multiple levels of granularity, thereby providing new insights by analyzing the uncertainty generated by the fuzzifier parameter.

In this study, we focus mainly on rough set-based possibilistic clustering approaches. In particular, we consider how to address the uncertainties implicated in data and generated by model parameters, including the threshold for partitioning approximation regions, the fuzzifier  $m$ , and scale parameters, with the current techniques of granular computing. The main objectives of this study are: 1) to optimize the partition threshold for each cluster based on shadowed sets [34] according to possibilistic membership degrees, which are used as the basis for establishing the multi-levels of granularity for approximation regions and to make the proposed approach sufficiently robust to deal with noisy environments; 2) to capture the uncertainty generated by the fuzzifier  $m$  in possibilistic clustering methods by detecting the variations in multigranulation approximation regions formed with multiple values of the fuzzifier parameter using a partially ordered relation, rather than at a single level of granularity for a specific fuzzifier value; and 3) to automatically adjust the scale parameters  $\{\gamma_i\}$  according to the maximal compatible regions of clusters obtained based on shadowed sets, which can reflect the zone of influence for each cluster well. After rationally resolving the model parameters in the rough set-based possibilistic clustering method, the prototypes can be obtained using ensemble strategies by combining the candidate results produced at the different levels of the granularity. In this manner, the prototypes calculated at a single level can be modified and they tend toward their natural positions.

The main contributions of this study are as follows. 1) The advantages of several granular computing techniques are integrated, including fuzzy sets, rough sets, and shadowed sets, so the uncertain information in the data, such as overlapping partitions, the vagueness arising in boundary regions, and the uncertainties produced by model parameters, can be handled adequately. 2) The proposed method is sufficiently robust to deal with noisy environments due to the typicality values associated with possibilistic clustering techniques. Our experimental results obtained using synthetic and real-world data demonstrated the superior performance of the proposed approach based on the comparisons with other available fuzzy and possibilistic clustering methods.

The remainder of this paper is organized as follows. In Section 2, we provide brief descriptions of FCM and PCM, as well as reviewing some rough set-based partitive clustering methods. In Section 3, we establish a generalized framework for the shadowed set-based rough possibilistic clustering approach. In Section 4, we discuss the uncertainty generated by the fuzzifier  $m$  in detail. Furthermore, multigranulation approximation regions are formed, which are used as the basis for introducing and analyzing a novel rough PCM method. The results of comparative experiments are presented in Section 5. We give our conclusions in Section 6.

## 2. Preliminaries

In this section, we review some partitive clustering algorithms, including FCM [2], PCM [4,5], and rough-fuzzy C-means (RFCM) [12,35,36]. More detailed information about rough sets can be found in previous studies [11,37,38].

### 2.1. FCM

Suppose that  $N$  patterns  $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ , defined over an  $M$ -dimensional feature space, i.e.,  $\mathbf{x}_j \in \mathcal{R}^M (j = 1, 2, \dots, N)$ , are grouped into  $C$  clusters  $G_1, G_2, \dots, G_C$ . The corresponding prototypes for each cluster are denoted as  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_C, \mathbf{v}_i \in \mathcal{R}^M (i = 1, 2, \dots, C)$ . In the FCM method, the following objective function is minimized.

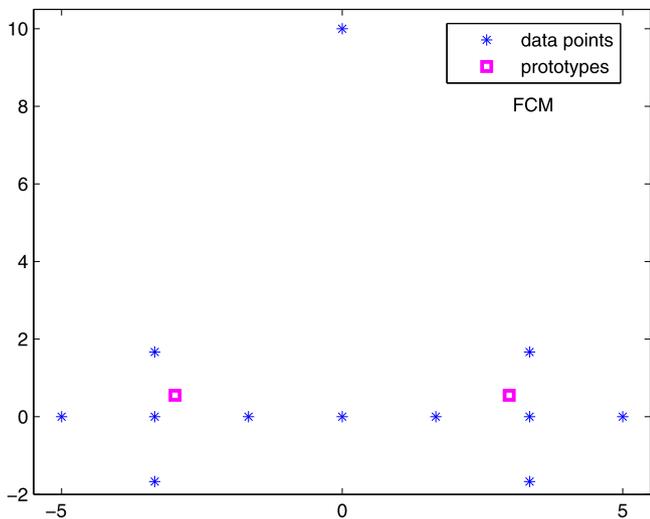


Fig. 1. Dataset  $D_{12}$  and prototypes obtained by FCM.

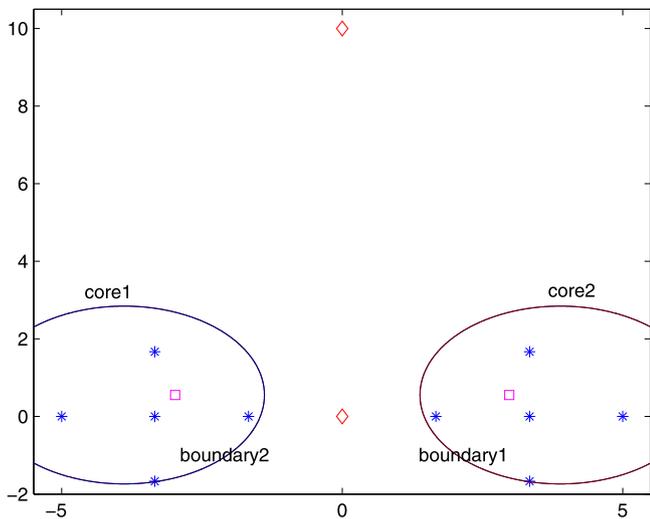


Fig. 2. Approximation partitions for  $D_{12}$  based on shadowed sets according to fuzzy memberships.

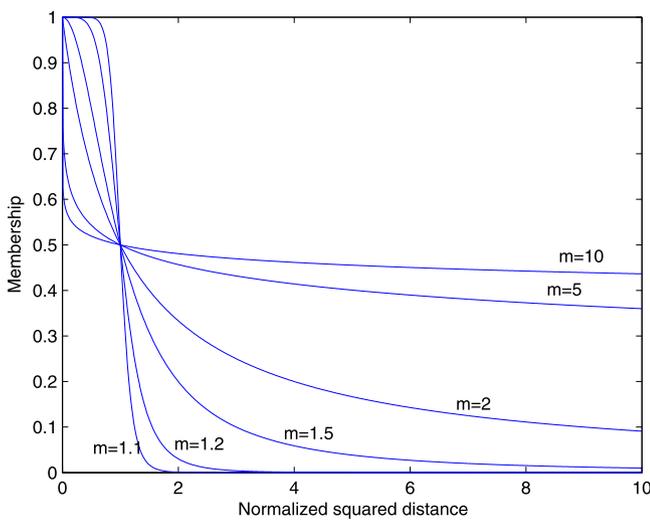


Fig. 3. Possibilistic membership degrees under different fuzzifier values.

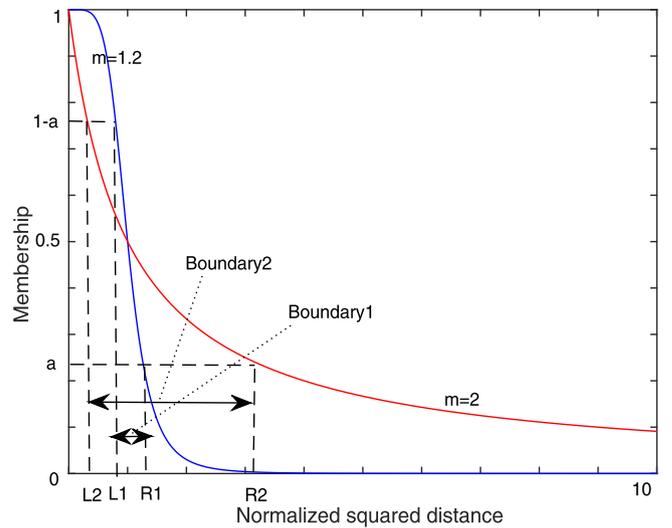


Fig. 4. The approximation region distributions between different fuzzifier values with respect to a fixed cluster according to possibilistic membership degrees.

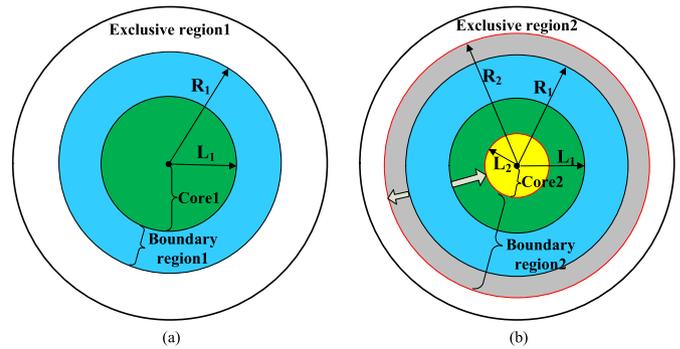


Fig. 5. The variation of approximation regions for a fixed cluster based on different possibilistic fuzzifier values. (a) The approximation region partitions for a fixed cluster under fuzzifier  $m_{p_1}$ . (b) The approximation region partitions for a fixed cluster under fuzzifier  $m_{p_1}$  and  $m_{p_2}$  according to the same partition threshold obtained under  $m_{p_1}$ .

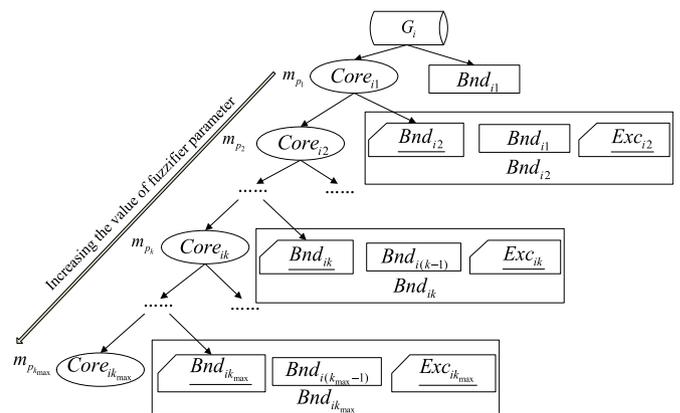


Fig. 6. The construction of multigranulation approximation regions of cluster  $G_i$ .

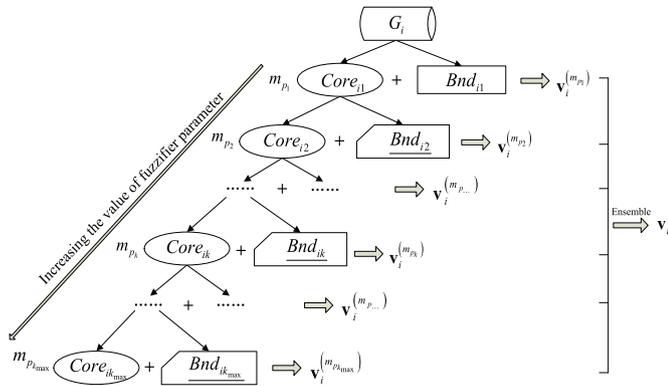


Fig. 7. The prototype calculations based on multigranulation approximation regions for cluster  $G_i$ .

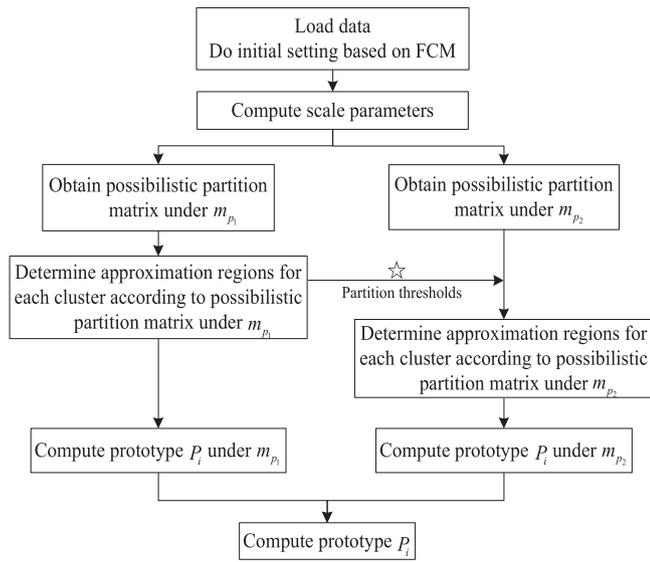


Fig. 8. The diagram of computing prototypes with two fuzzifier values.

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

$$s. t. \quad \mu_{ij} \in [0, 1] \text{ for all } i, j, \text{ and } 0 < \sum_{j=1}^N \mu_{ij} < N \text{ for all } i = 1, 2, \dots, C; \tag{2}$$

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

$\mu_{ij}$  is a fuzzy membership value that measures the degree to which pattern  $x_j$  belongs to the cluster  $G_i$ .  $m_f (m_f > 1)$  denotes the fuzzifier in FCM, which controls the shape of membership degrees, i.e., when the value is close to 1, this implies the Boolean nature of one cluster. In addition, it will yield spike-like membership functions when the value increases.  $d_{ij}$  denotes the distance between the pattern  $x_j$  and the cluster with the prototype  $v_i$ . In this study, we employ the weighted Euclidean distance in Eq. (4) to eliminate the influence of significantly different ranges of individual features.  $\delta_k$  is the deviation of the  $k$ th feature.

$$d_{ij} = \sqrt{\sum_{k=1}^M \frac{(x_{jk} - v_{ik})^2}{\delta_k^2}} \tag{4}$$

The approximate optimization of  $J_{FCM}$  using the FCM method is based on iteration and it satisfies the following necessary conditions:

$$\mu_{ij} = \frac{1}{\sum_{k=1}^C \left(\frac{d_{ij}}{d_{kj}}\right)^{\frac{2}{m_f-1}}}, \quad i = 1, 2, \dots, C \text{ and } j = 1, 2, \dots, N; \tag{5}$$

$$v_i = \frac{\sum_{j=1}^N \mu_{ij}^{m_f} x_j}{\sum_{j=1}^N \mu_{ij}^{m_f}}, \quad i = 1, 2, \dots, C. \tag{6}$$

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

### 2.2. PCM

To overcome the limitations of FCM, Krishnapuram and Keller [4]

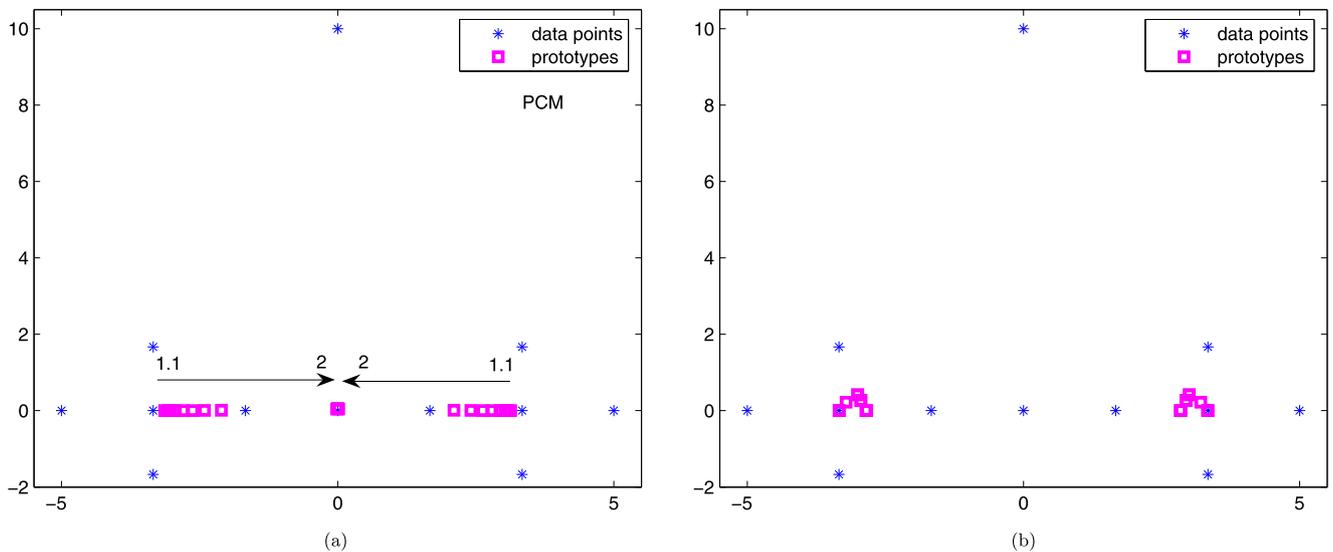


Fig. 9. The obtained prototypes as increasing the fuzzifier value (a) PCM; (b) MSPCM, fixing  $m_{p2} = 2$ .

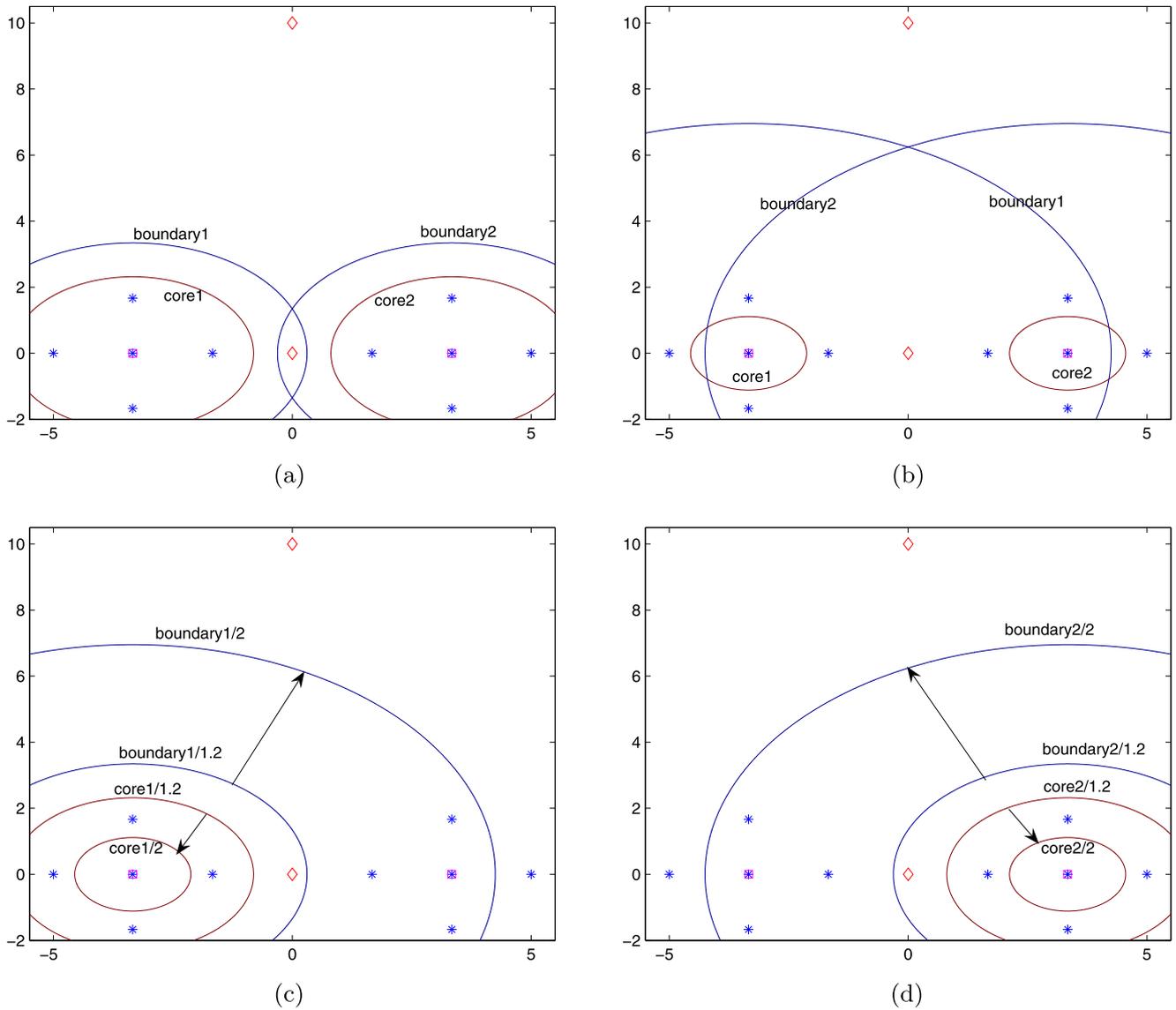


Fig. 10. The partitions of approximation regions obtained by MSPCM. (a) under  $m_{p1} = 1.2$ ; (b) under  $m_{p2} = 2$ ; (c) The variations of approximation regions with respect to the Cluster 1; (d) The variations of approximation regions with respect to the Cluster 2. The pink squares denote the obtained prototypes.

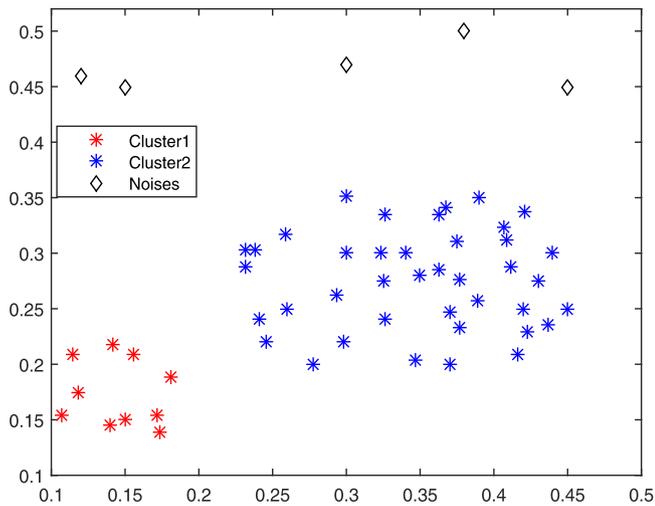


Fig. 11. Synthetic dataset I.

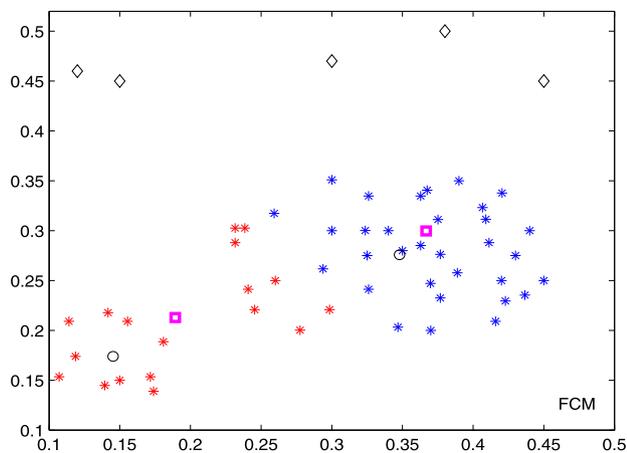
relaxed constraint (3) in FCM, and proposed a PCM model. The objective function of PCM that needs to be minimized is described as follows:

$$J_{PCM}(U, V) = \sum_{i=1}^C \sum_{j=1}^N u_{ij}^{m_p} d_{ij}^2 + \sum_{i=1}^C \gamma_i \sum_{j=1}^N (1 - u_{ij})^{m_p}; \tag{7}$$

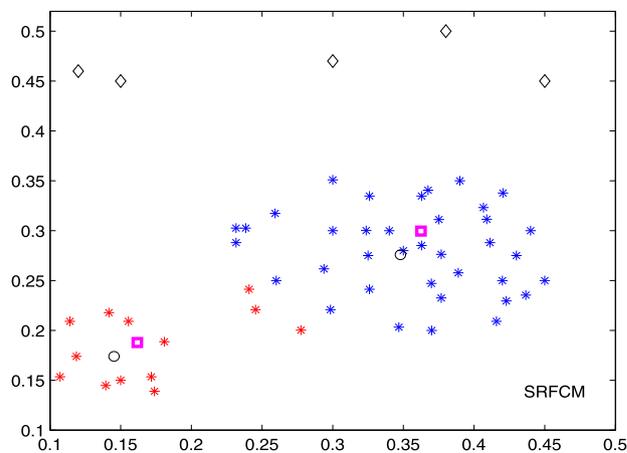
$$s. t. \quad u_{ij} \in [0, 1] \text{ for all } i, j, \text{ and } 0 < \sum_{j=1}^N u_{ij} < N \text{ for all } i = 1, 2, \dots, C; \tag{8}$$

$$\max_i(u_{ij}) > 0 \text{ for all } j = 1, 2, \dots, N, \tag{9}$$

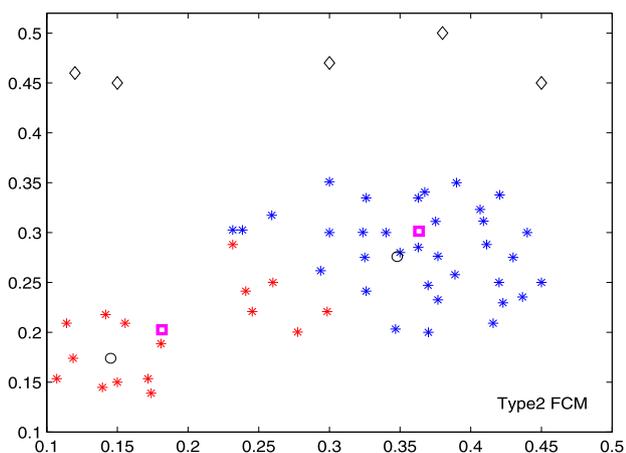
where  $d_{ij}$  has the same meaning as that in FCM,  $u_{ij}$  is the possibilistic membership value that measures the typicality for pattern  $x_j$  in cluster  $G_i$ ,  $m_p(m_p > 1)$  is the fuzzifier in PCM, and  $\gamma_i$  is a scale parameter representing the zone of influence or the size of the  $i$ -th cluster. The approximate optimization of  $J_{PCM}$  using the PCM method is also based



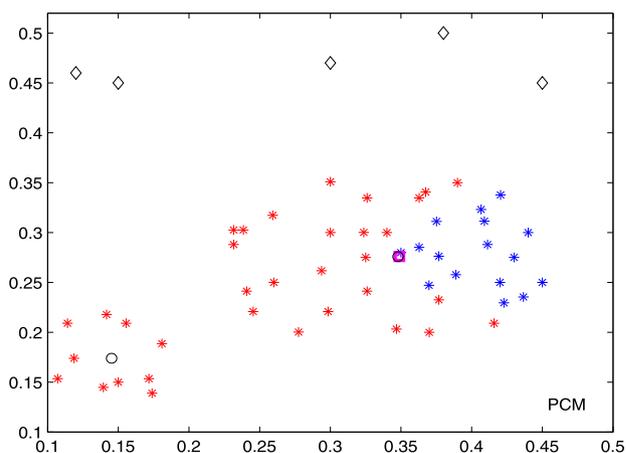
(a) FCM



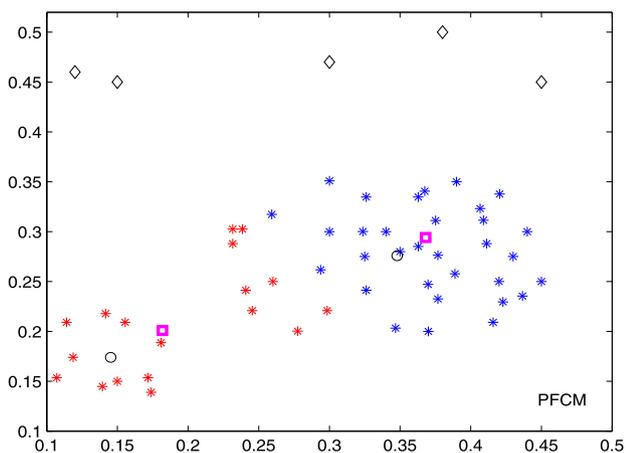
(b) SRFCM



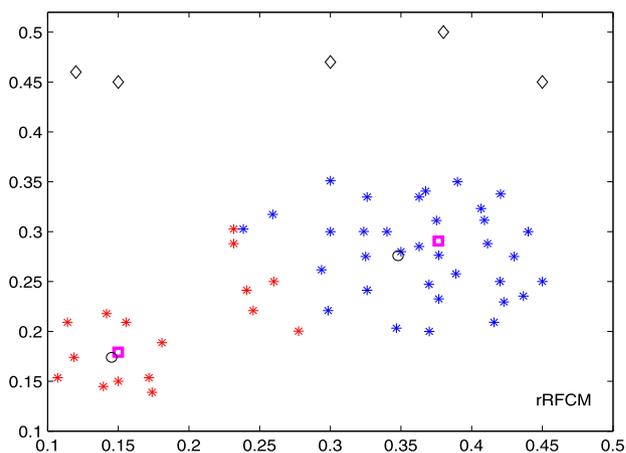
(c) Type2 FCM



(d) PCM

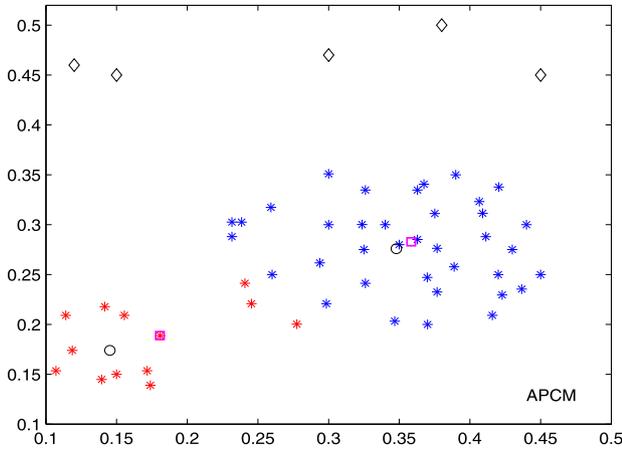


(e) PFCM

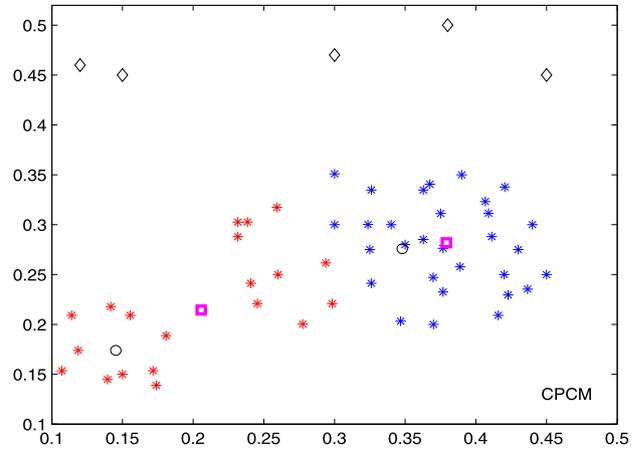


(f) rRFCM

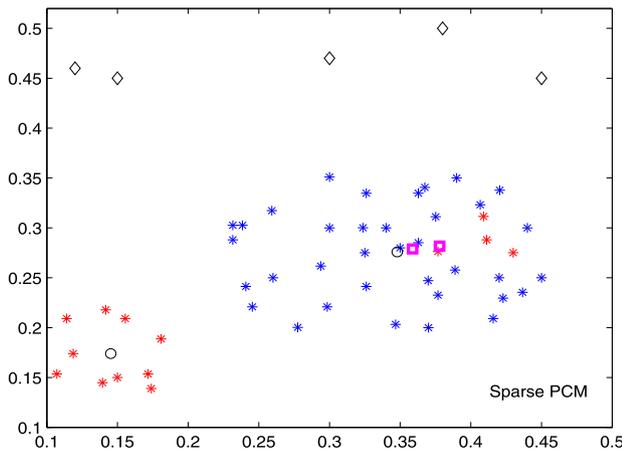
**Fig. 12.** Synthetic dataset I and clustering results. The black circles and pink squares mean the ground centroids of clusters and obtained prototypes, respectively. The patterns with color red and blue denote that patterns are classified into Clusters 1 and 2, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



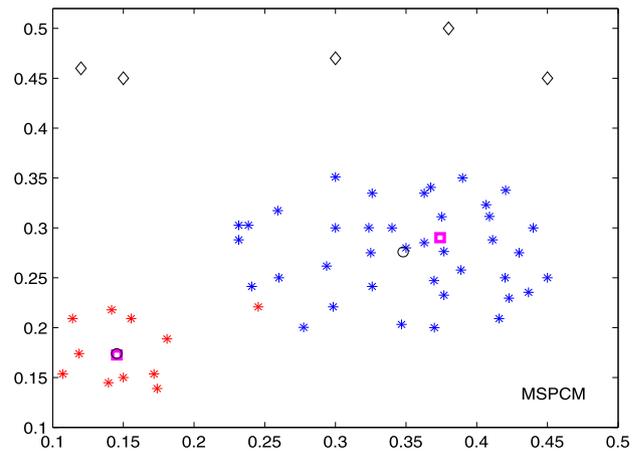
(g) APCM



(h) CPCM



(i) Sparse PCM



(j) MSPCM

Fig. 12. (continued)

on iteration strategies and it satisfies the following necessary conditions:

$$u_{ij} = \frac{1}{1 + \left(\frac{d_{ij}^2}{\gamma_i}\right)^{m_p-1}}, \quad i = 1, 2, \dots, C \text{ and } j = 1, 2, \dots, N; \quad (10)$$

$$\mathbf{v}_i = \frac{\sum_{j=1}^N u_{ij}^{m_p} \mathbf{x}_j}{\sum_{j=1}^N u_{ij}^{m_p}}, \quad i = 1, 2, \dots, C. \quad (11)$$

The scale parameters  $\{\gamma_i\}(i = 1, 2, \dots, C)$  are suggested by the following computation:

$$\gamma_i = \kappa \frac{\sum_{j=1}^N \mu_{ij}^{m_f} d_{ij}^2}{\sum_{j=1}^N \mu_{ij}^{m_f}}, \quad i = 1, 2, \dots, C. \quad (12)$$

The value of  $\kappa$  is a constant and it is kept as one.  $\{\mu_{ij}\}(i = 1, 2, \dots, C, j = 1, 2, \dots, N)$  are the terminal results obtained by FCM. Relative distances are involved in the computation of the fuzzy membership degrees in FCM, whereas the calculation of the possibilistic membership degree or typicality for a pattern in a cluster is only associated with the absolute distance  $d_{ij}$  in PCM. Thus, a pattern with a greater distance to a cluster will have a lower typicality with respect to this cluster. By contrast, a pattern with a closer distance to a cluster will have a higher typicality with respect to this cluster. In this manner, the

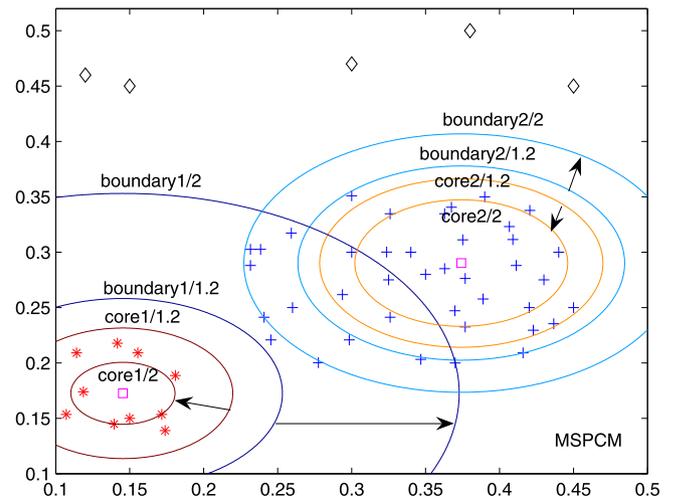


Fig. 13. The partitions of approximation regions generated by MSPCM, where “core1/1.2” and “boundary1/1.2” mean the borderline of core region and boundary region of Cluster 1 obtained under fuzzifier value 1.2, respectively.

noise or outliers will be assigned very lower typicality values, so they will make almost no contribution to the updating of prototypes during the iteration procedures.

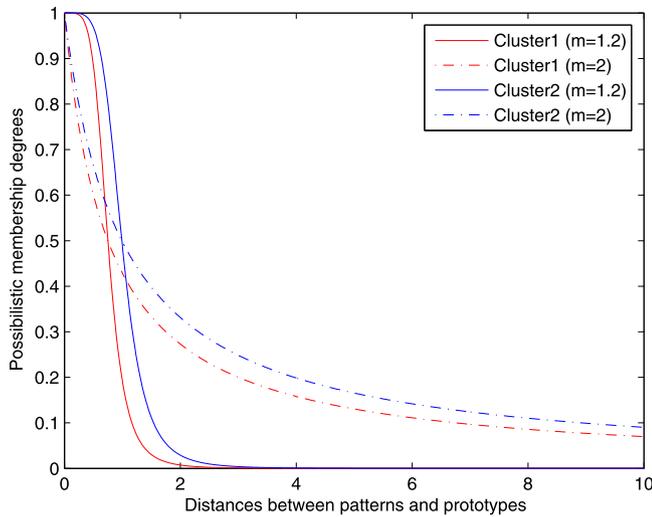


Fig. 14. The possibilistic membership degrees under different fuzzifier values for synthetic data set I .

2.3. RFCM

In FCM and PCM, the prototype calculations are related to all patterns. Intuitively, the patterns around a prototype are more important for updating this prototype and the patterns far away from this prototype will make no contribution or even the opposite contribution to the evolution of the prototype. Lingras et al. [39] extended the concept of rough sets to develop a clustering algorithm called rough C-means (RCM) where all the patterns can be divided into three approximation regions with respect to a fixed cluster, i.e., core, boundary, and exclusion regions. The new prototype calculations are only related to the core and boundary regions, and not all patterns, as found in FCM or PCM. Thus, the useless information can be filtered out and the number of numeric computations can be reduced. By incorporating the fuzzy membership degrees, Mitra et al. [35] proposed the notion of RFCM, where the absolute distances  $\{d_{ij}\}$  are replaced by fuzzy membership degrees  $\{\mu_{ij}\}$  when dividing patterns into approximation regions. This adjustment enhances the robustness of the clustering method when dealing with overlapping situations.

In rough set-based clustering methods, the patterns in the core region will certainly belong to this cluster. The patterns in the boundary region may possibly belong to this cluster, i.e., with vagueness and uncertainty. Other patterns in the exclusion region will definitely not belong to this cluster. The contribution of each approximation region to

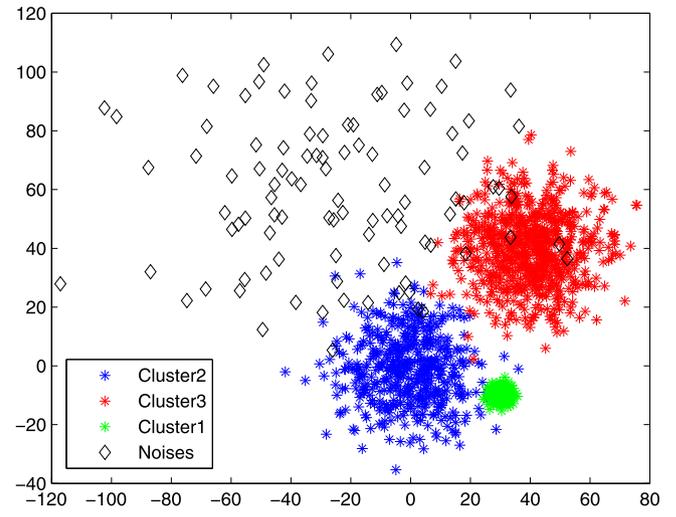


Fig. 16. Synthetic dataset II.

the iterative updating of the prototype is greater diversity.

According to the notion of RFCM, the prototypes  $v_1, v_2, \dots, v_C, v_i \in \mathcal{R}^M$ , are renewed based on the following principles:

$$v_i = \begin{cases} w_l A_i + w_b B_i & \text{if } \underline{R}^{(m_f)} G_i \neq \emptyset \wedge R_b^{(m_f)} G_i \neq \emptyset \\ B_i & \text{if } \underline{R}^{(m_f)} G_i = \emptyset \wedge R_b^{(m_f)} G_i \neq \emptyset \\ A_i & \text{if } \underline{R}^{(m_f)} G_i \neq \emptyset \wedge R_b^{(m_f)} G_i = \emptyset \end{cases} \quad (13)$$

where  $A_i = \frac{\sum_{x_j \in \underline{R}^{(m_f)} G_i} \mu_{ij}^{m_f} x_j}{\sum_{x_j \in \underline{R}^{(m_f)} G_i} \mu_{ij}^{m_f}}$ ,  $B_i = \frac{\sum_{x_j \in R_b^{(m_f)} G_i} \mu_{ij}^{m_f} x_j}{\sum_{x_j \in R_b^{(m_f)} G_i} \mu_{ij}^{m_f}}$  can be considered as

the contributions from the fuzzy core region and fuzzy boundary region, respectively.  $R_b^{(m_f)} G_i = \overline{R}^{(m_f)} G_i - \underline{R}^{(m_f)} G_i$  denotes the boundary region of cluster  $G_i$  with respect to the feature set  $R$  according to the fuzzy membership degrees.  $\underline{R}^{(m_f)} G_i$  and  $\overline{R}^{(m_f)} G_i$  are the lower and upper approximations of cluster  $G_i$  under the fuzzifier value  $m_f$ , respectively. The lower approximation  $\underline{R}^{(m_f)} G_i$  is also considered as the core region of cluster  $G_i$ .  $w_l (0.5 < w_l \leq 1)$  and  $w_b = 1 - w_l$  are the weighted values that measure the contributions of the core region and boundary region, respectively.  $\{\mu_{ij}\}$  are calculated in the same manner as those in FCM.

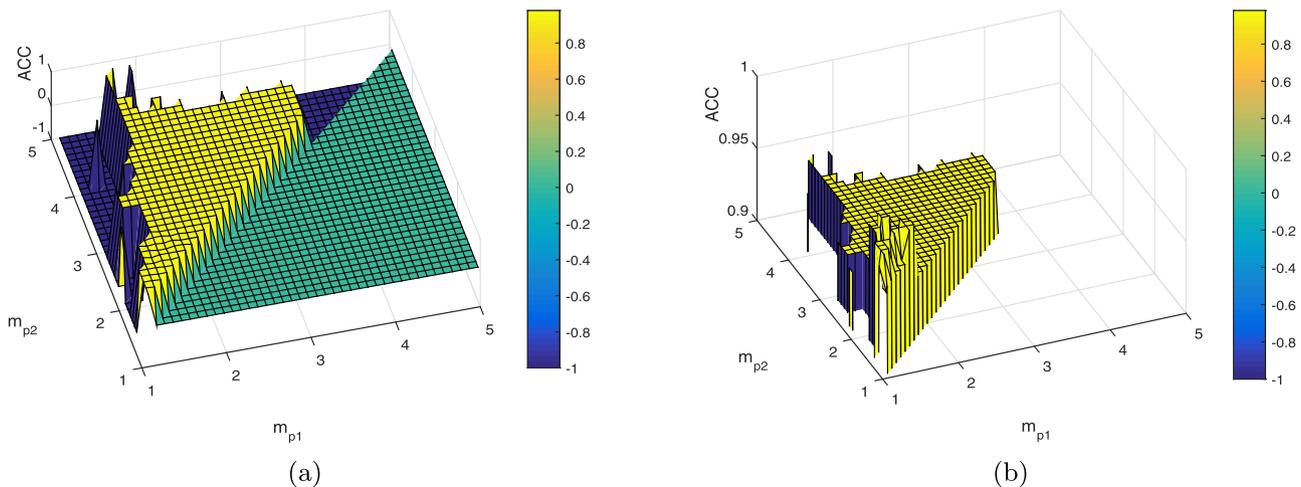


Fig. 15. ACC index for synthetic dataset I. (a) ACC index with respect to different values of  $m_{p1}$  and  $m_{p2}$ ; (b) The local profile of ACC index.

To determine the core and boundary regions of each cluster, the following principles are utilized.

If  $\mu_{pj} - \mu_{qj} \leq \Delta$ , then  $\mathbf{x}_j \in \bar{R}^{(m_f)}G_p$  and  $\mathbf{x}_j \in \bar{R}^{(m_f)}G_q$ . In this case,  $\mathbf{x}_j$  cannot be partitioned into the core region of any clusters; otherwise,  $\mathbf{x}_j \in \underline{R}^{(m_f)}G_p$ .  $\mu_{pj}$  is the maximum of  $\mathbf{x}_j$  over all clusters and  $\mu_{qj}$  is next to the maximum.

The threshold  $\Delta$  is crucial for determining the approximation regions of each cluster. More objects will be divided into the core regions when the threshold is smaller. By contrast, more objects will belong to the boundary regions when the threshold is larger. An unreasonable threshold will yield inaccurate approximation regions and then guide the prototype updating process incorrectly.

### 3. General framework for rough PCM based on shadowed sets

In this section, we introduce a framework for rough PCM clustering based on shadowed sets, where the advantages of rough set-based clustering methods and PCM are integrated. More detailed descriptions of shadowed sets can be found in previous studies [40–42].

#### 3.1. Limitations of shadowed set-based rough-fuzzy clustering approaches

In rough set-based clustering approaches, i.e., RCM and RFCM, the threshold  $\Delta$  that determines the approximation regions directly affects the prototype updating process. This value is often selected by subjective tuning and kept as a constant for all clusters during all iterations, which fails to determine the data structures well, especially for clusters with different sizes and densities. Zhou and Pedrycz [43] integrated shadowed set theory and rough set-based clustering approaches to develop an improved rough-fuzzy C-means method called SRFCM where the determination of the approximation regions for each cluster is transformed into an optimization process so they can be detected automatically during the clustering processes. The principles for determining the approximation regions of each cluster based on shadowed sets in SRFCM are described as follows.

Step 1: Compute the fuzzy membership values  $\{\mu_{ij}\}$  with Eq. (5).

Step 2: Based on the shadowed sets, compute the optimal threshold  $\alpha_i^{(m_f)}$  for each cluster  $G_i$ :

$$\alpha_i^{(m_f)} = \min_{\alpha} (V_i) = \min_{\alpha} \left[ \sum_{j:\mu_{ij} \leq \alpha} \mu_{ij} + \sum_{j:\mu_{ij} \geq \max(\mu_{ij}) - \alpha} (1 - \mu_{ij}) - \text{card} \left( \left\{ \mathbf{x}_j \mid \alpha < \mu_{ij} < \max(\mu_{ij}) - \alpha \right\} \right) \right]. \tag{14}$$

Step 3: According to  $\alpha_i^{(m_f)}$ , determine the core region (lower approximation) and the boundary region of cluster  $G_i$ :

$$\begin{aligned} \underline{R}^{(m_f)}G_i &= \left\{ \mathbf{x}_j \mid \mu_{ij} \geq \max_j(\mu_{ij}) - \alpha_i^{(m_f)} \right\}, \\ \bar{R}^{(m_f)}G_i &= \left\{ \mathbf{x}_j \mid \alpha_i^{(m_f)} < \mu_{ij} < \max_j(\mu_{ij}) - \alpha_i^{(m_f)} \right\}, \end{aligned} \tag{15}$$

where  $\alpha_i^{(m_f)} \in \left[ \min_j(\mu_{ij}), \frac{\min(\mu_{ij}) + \max(\mu_{ij})}{2} \right)$ . In Steps 1–3, the approximation region partition threshold of each cluster is not defined in advance by the user, but instead it can be adjusted automatically in the clustering processes and optimized for each cluster independently.

The shadowed set-based rough fuzzy clustering method performs better than FCM, RCM, and RFCM [43], but it also has a limitation when dealing with noisy environments. According to Eq. (15), each pattern will be divided into the core region or boundary region of at

least one cluster due to constraint (3) in FCM. For the shadowed sets induced from an available fuzzy set, according to Eqs. (14) and (15), if the membership degree of a pattern belonging to a cluster is 0.5, this pattern will inevitably belong to the boundary region of this cluster. However, outliers and some noisy patterns should not be divided into any core region and boundary region over all clusters. Thus, they should belong to the exclusion region over all clusters, which is called the absolute exclusion region. This concept is illustrated with the data set used in [6] and shown in Table 1.

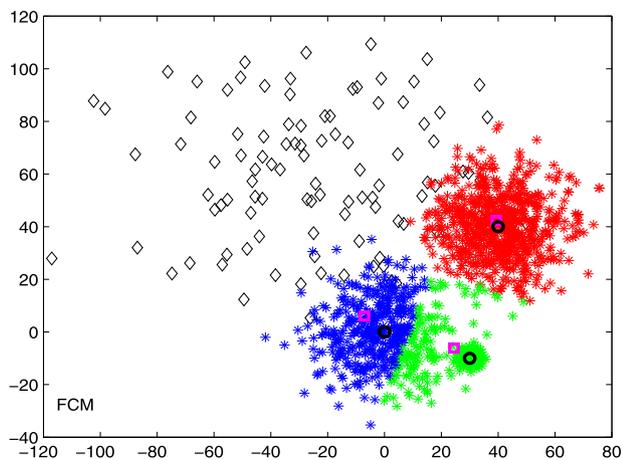
The prototypes obtained by FCM are shown in Fig. 1 and the corresponding membership degrees are presented in Table 1 (columns 4 and 5). Patterns  $\mathbf{x}_{11}$  and  $\mathbf{x}_{12}$  have membership degrees of 0.5 with respect to each cluster because both are equidistant from the prototypes, although the pattern  $\mathbf{x}_{12}$  is far away from the prototypes, which can be attributed to constraint (3) in FCM, i.e., relative distances are involved. According to (15), the approximation region partition of each cluster is presented in Table 1 (columns 6 and 7) and shown in Fig. 2. The values of 1, 0, and  $-1$  in column 6 (column 7) denote belonging to the core, boundary, and exclusive regions of Cluster 1 (Cluster 2), respectively. Pattern  $\mathbf{x}_{12}$  belongs to the boundary region of each cluster and it contributes to the computation of the prototypes, which results in undesired prototypes, as shown in Fig. 2. Intuitively, pattern  $\mathbf{x}_{12}$  should be partitioned into the absolute exclusion region over all clusters. Thus, this implies that pattern  $\mathbf{x}_{12}$  should not be partitioned into the core regions and boundary regions of any clusters. In addition, it is better to partition pattern  $\mathbf{x}_{11}$  into the boundary region of each cluster to some extent.

#### 3.2. Generalized framework for rough PCM based on shadowed sets

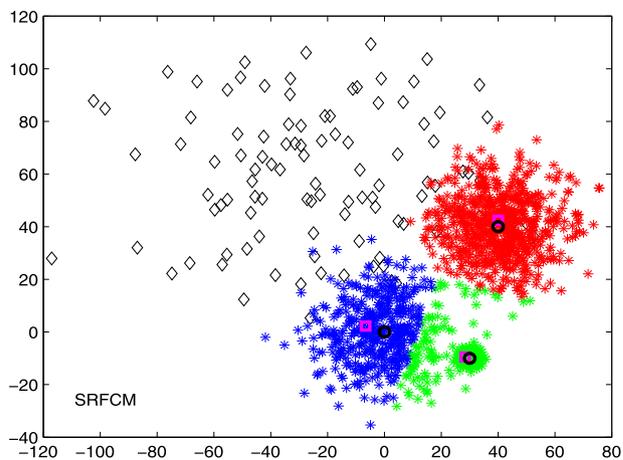
PCM is more robust than FCM when dealing with noisy environments. The objective function of PCM can be considered as a collection of  $C$ -independent sub-objective functions. The possibilistic membership degrees (typicality values) of patterns with respect to a fixed cluster can be considered as an independent solution. Thus, according to the shadowed set optimization mechanism, the possibilistic membership degrees of all patterns obtained for a fixed cluster can be divided into three levels, i.e., sufficiently high, sufficiently low, and the shadows, based on which three approximation regions can be formed with respect to this cluster. By integrating the notions of shadowed sets and PCM, we present a generalized framework for rough PCM based on shadowed sets in Algorithm 1.

In contrast to the SRFCM algorithm reviewed in Section 3.1, the possibilistic membership degrees  $\{u_{ij}\}$  are utilized instead of the fuzzy membership degrees  $\{\mu_{ij}\}$  in Algorithm 1. The generalized framework of rough PCM based on shadowed sets integrates the advantages of PCM, rough sets, and shadowed sets. The typicality values make the algorithm sufficiently robust to deal with noisy environments, the approximation regions deduced by rough set theory allow the algorithm to capture the topologies of the data, and the shadowed sets provide an optimal separate threshold for dividing approximation regions based on global observations of the data.

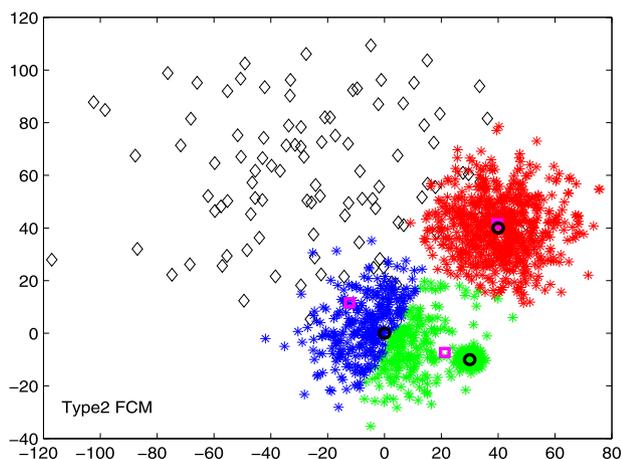
Pedrycz’s optimization model with formula (16) is a specific method used for determining the partition threshold based on the mechanism of uncertainty invariance, i.e., uncertainty relocation. Other optimization principles can be specified in different ways within the framework of three-way decision theory [44,45], such as the principle of minimum distance and the principle of least cost [42,46]. The principle selected should consider the characteristics of practical applications, which can be analyzed based on the three-way decisions method. The selection of different optimization principles for constructing shadowed sets is not the major focus of this study, but instead we employ Pedrycz’s optimization strategy for designing a symmetrical partition threshold.



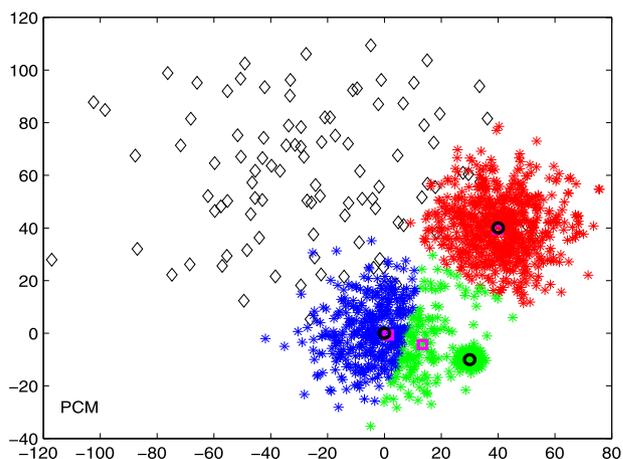
(a) FCM



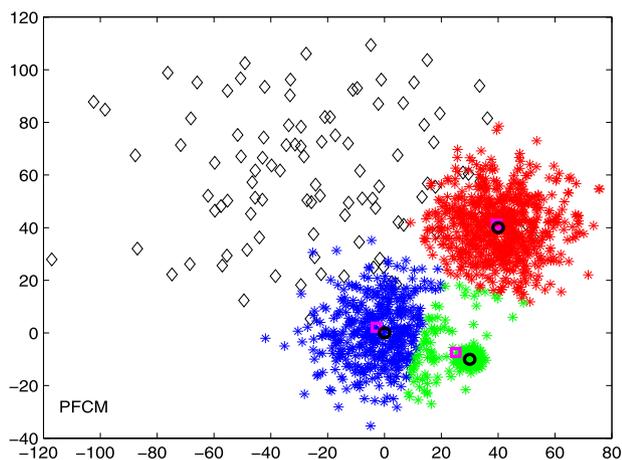
(b) SRFCM



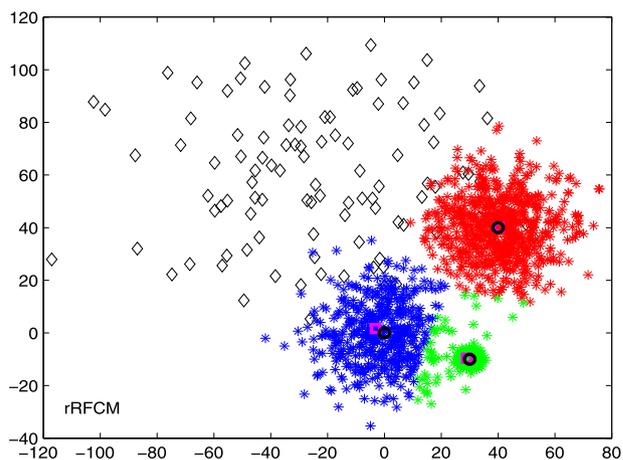
(c) Type2 FCM



(d) PCM

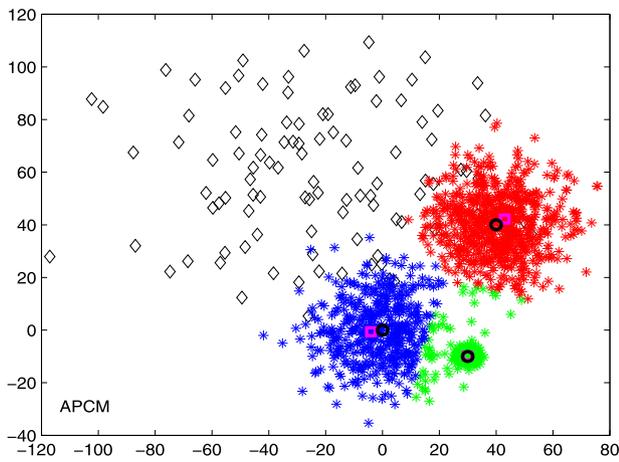


(e) PFCM

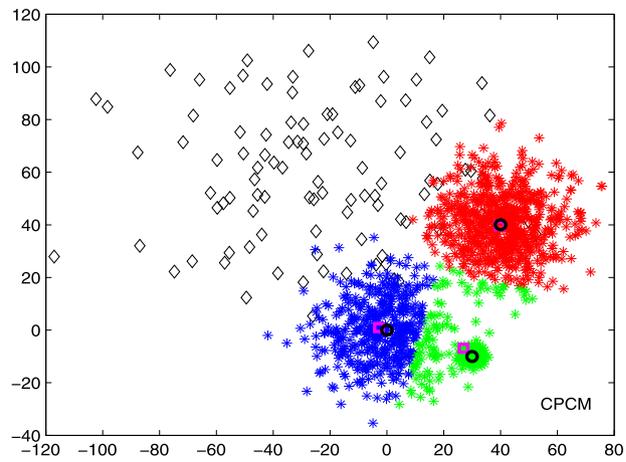


(f) rRFCM

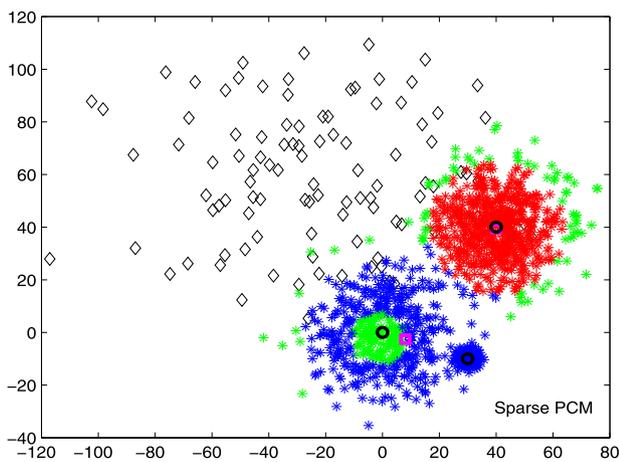
Fig. 17. Synthetic dataset II and clustering results. The black circles denote the means of Gaussian distributions. The pink squares denote the obtained prototypes. The patterns with color green, blue and red denote that patterns are classified to Clusters 1–3, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



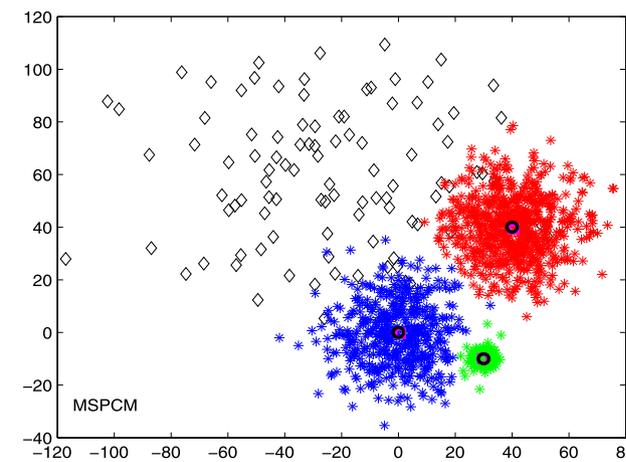
(g) APCM



(h) CPCM



(i) Sparse PCM



(j) MSPCM

Fig. 17. (continued)

Two issues need to be addressed in Algorithm 1 before its execution: selecting the fuzzifier value  $m_p$  and determining the scale parameters  $\{\gamma_i\}$ . We discuss the potential problems caused by these issues and their solutions in the following section.

#### 4. Rough possibilistic clustering approach based on multigranulation approximation regions and shadowed sets

##### 4.1. Uncertainty generated by the fuzzifier $m_p$

The fuzzifier  $m_p$  determines the rate of decay for the possibilistic membership values. The relationships between memberships and the normalized squared distance ( $d_{ij}^2/\gamma_i$ ) are shown in Fig. 3. As the value of  $m_p$  increases, the possibilistic memberships will tend toward 0.5. In this case, irrespective of the distance between a pattern and a cluster, the membership values over all clusters tend to be the same because no restrictions need to be satisfied, such as constraint (3) in FCM.

As shown by Hwang and Rhee in [16], type-1 fuzzy sets cannot adequately manage the uncertainty generated by a single fuzzifier value, and thus type-2 fuzzy sets are employed to capture the uncertainty caused by a specific fuzzifier value in FCM. Similarly, two different fuzzifier values  $m_{p_1}$  and  $m_{p_2}$  ( $1 < m_{p_1} < m_{p_2}$ , is assumed in the rest of this study) form a footprint of uncertainty for possibilistic memberships from the vertical direction in Fig. 3, i.e., the variations in memberships associated with the same normalized squared distance. The description of the uncertainty generated by the fuzzifier values can

also be drawn from the horizontal direction in Fig. 3, i.e., the variations in the normalized squared distances of patterns under a fixed possibilistic membership value. An important proposition can be made as follows.

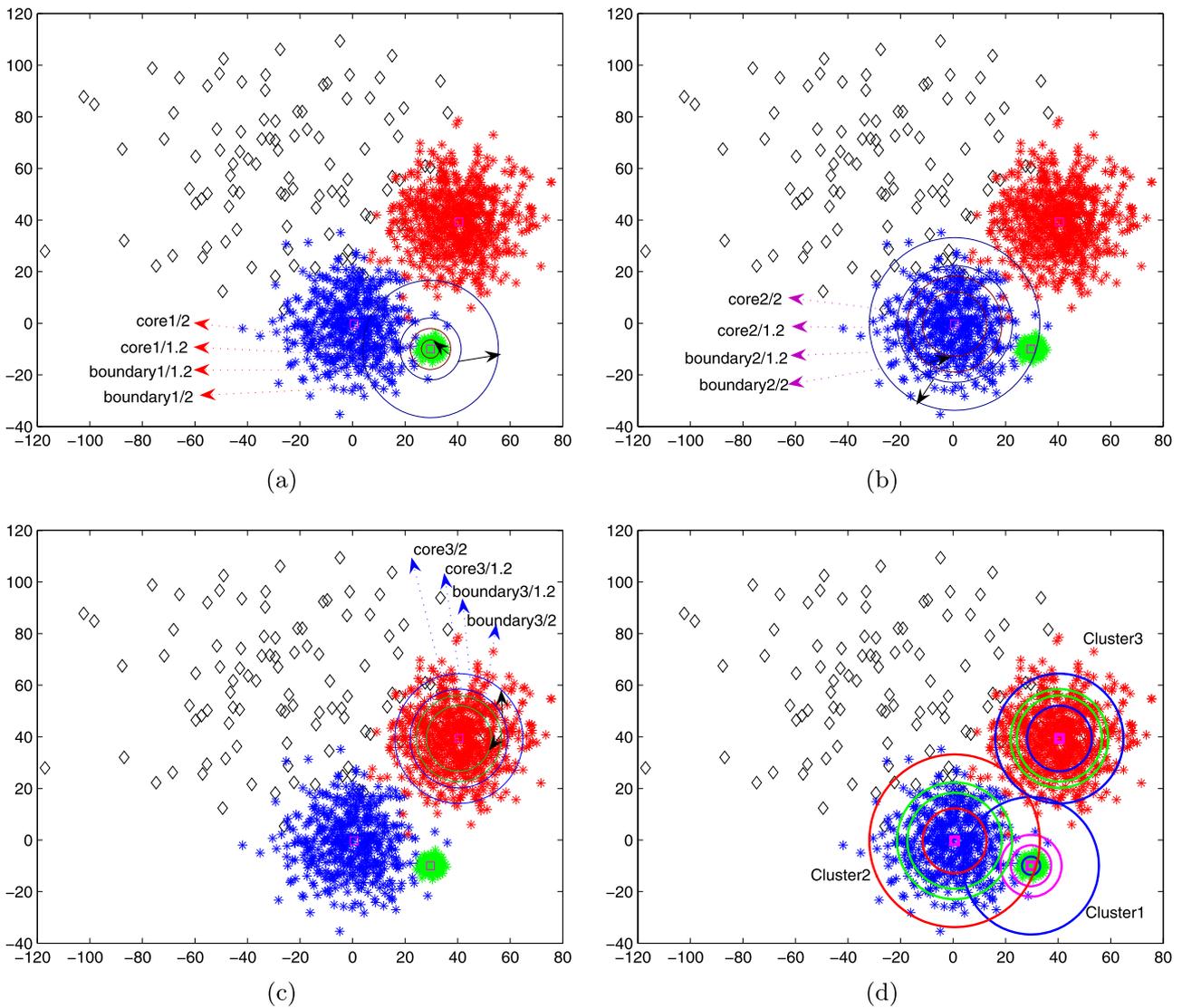
**Proposition 1.** Set  $x = \frac{d_{ij}^2}{\gamma_i}$ ,  $f_{PCM}(x, m) = \frac{1}{1 + x^{m-1}}$ . If we suppose that  $1 < m_{p_1} < m_{p_2}$  and  $f_{PCM}(x_1, m_{p_1}) = f_{PCM}(x_2, m_{p_2})$ , then we have the following:

- 1) If  $0 < x_1, x_2 < 1$ , then  $x_2 < x_1$ ;
- 2) If  $x_1, x_2 > 1$ , then  $x_1 < x_2$ .

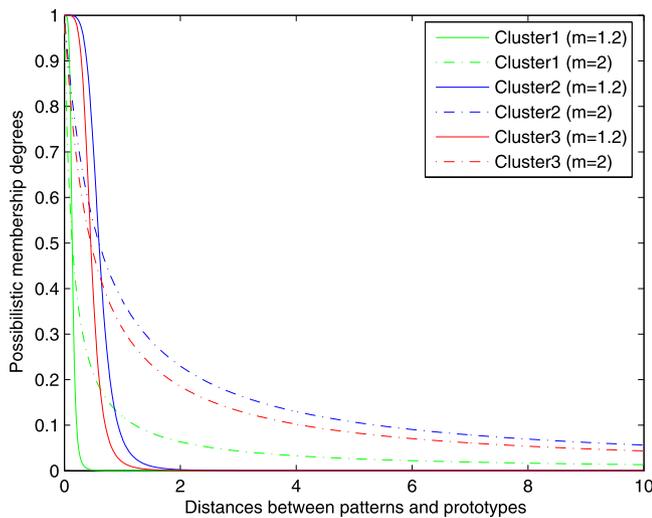
The detailed proofs can be found in Appendix.

Proposition 1 reflects the influence generated by different fuzzifier values with respect to the normalized squared distances of patterns. It can be found that patterns with the same possibilistic membership degree with respect to a cluster may be different distances from this cluster due to the different fuzzifier values.

Suppose a data set with two clusters  $G_1$  and  $G_2$ , if we take  $G_1$  as an example, then  $G_2$  can be explained in a similar manner. When giving a specific fuzzifier value  $m_p = m_{p_1}$  (such as  $m_{p_1} = 1.2$ ), the optimal threshold  $\alpha_1^{(m_{p_1})}$  for cluster  $G_1$  under  $m_{p_1}$  can be obtained by formula (16) based on the scheme of shadowed sets according to the possibilistic membership degrees, such as  $\alpha_1^{(m_{p_1})} = \alpha$ . The boundary region of cluster  $G_1$  comprises the patterns with membership degrees between  $\alpha_1^{(m_{p_1})}$  and  $1 - \alpha_1^{(m_{p_1})}$  according to formula (17). In this case, their normalized



**Fig. 18.** The partitions of approximation regions obtained by MSPCM, where “core1/1.2” and “boundary1/1.2” mean the borderline of core region and boundary region of Cluster 1 obtained under fuzzifier value 1.2, respectively. (a) The partitions of approximation regions of Cluster 1; (b) The partitions of approximation regions of Cluster 2; (c) The partitions of approximation regions of Cluster 3; (d) The partitions of approximation regions for three clusters.



**Fig. 19.** The variations of possibilistic membership degrees under different fuzzifier values in MSPCM.

**Table 1**  
Synthetic dataset  $D_{12}$ .

	Data		FCM		Approximation region	
	x	y	$\mu_{1j}$	$\mu_{2j}$	Cluster 1	Cluster 2
$x_1$	-5.00	0.00	<b>0.93486</b>	0.065136	1	-1
$x_2$	-3.34	1.67	<b>0.96239</b>	0.037611	1	-1
$x_3$	-3.34	0.00	<b>0.98781</b>	0.012191	1	-1
$x_4$	-3.34	-1.67	<b>0.88377</b>	0.11623	1	-1
$x_5$	-1.67	0.00	<b>0.91375</b>	0.086247	1	-1
$x_6$	1.67	0.00	0.086245	<b>0.91375</b>	-1	1
$x_7$	3.34	1.67	0.037612	<b>0.96239</b>	-1	1
$x_8$	3.34	0.00	0.012191	<b>0.98781</b>	-1	1
$x_9$	3.34	-1.67	0.11623	<b>0.88377</b>	-1	1
$x_{10}$	5.00	0.00	0.065136	<b>0.93486</b>	-1	1
$x_{11}$	0.00	0.00	0.5	0.5	0	0
$x_{12}$	0.00	10.00	0.5	0.5	0	0

**Input:** Data set  $\{\mathbf{x}_j | j = 1, 2, \dots, N\}$ , number of clusters  $C$ .

**Output:** Prototypes  $\{\mathbf{v}_i | i = 1, 2, \dots, C\}$ .

- 1: Initialize the prototypes  $\mathbf{v}_i (i = 1, 2, \dots, C)$  based on the results obtained by FCM;
- 2: **repeat**
- 3: Compute the possibilistic membership values  $\{u_{ij}\}$  with equation (10);
- 4: Based on the shadowed sets, compute the optimal threshold  $\alpha_i^{(mp)}$  for each cluster  $G_i$ ;

$$\alpha_i^{(mp)} = \min_{\alpha} (Y_i) = \min_{\alpha} \left| \sum_{j: u_{ij} \leq \alpha} u_{ij} + \sum_{j: u_{ij} \geq \max_j(u_{ij}) - \alpha} (1 - u_{ij}) \right|$$

$$-card \left( \left\{ \mathbf{x}_j | \alpha < u_{ij} < \max_j(u_{ij}) - \alpha \right\} \right); \tag{16}$$

- 5: According to  $\alpha_i^{(mp)}$ , determine the core region and boundary region of cluster  $G_i$  based on the possibilistic membership

degrees:

$$\underline{R}^{(mp)} G_i = \left\{ \mathbf{x}_j | u_{ij} \geq \max_j(u_{ij}) - \alpha_i^{(mp)} \right\};$$

$$\underline{R}_b^{(mp)} G_i = \left\{ \mathbf{x}_j | \alpha_i^{(mp)} < u_{ij} < \max_j(u_{ij}) - \alpha_i^{(mp)} \right\}; \tag{17}$$

- 6: Update the prototypes:

$$\mathbf{v}_i = \begin{cases} w_1 \times A_2 + w_b \times B_2 & \text{if } \underline{R}^{(mp)} G_i \neq \emptyset \wedge \underline{R}_b^{(mp)} G_i \neq \emptyset \\ A_2 & \text{if } \underline{R}^{(mp)} G_i \neq \emptyset \wedge \underline{R}_b^{(mp)} G_i = \emptyset \\ B_2 & \text{if } \underline{R}^{(mp)} G_i = \emptyset \wedge \underline{R}_b^{(mp)} G_i \neq \emptyset \end{cases} \tag{18}$$

where  $A_2 = \frac{\sum_{\mathbf{x}_j \in \underline{R}^{(mp)} G_i} u_{ij}^{mp} \mathbf{x}_j}{\sum_{\mathbf{x}_j \in \underline{R}^{(mp)} G_i} u_{ij}^{mp}}$ ,  $B_2 = \frac{\sum_{\mathbf{x}_j \in \underline{R}_b^{(mp)} G_i} u_{ij}^{mp} \mathbf{x}_j}{\sum_{\mathbf{x}_j \in \underline{R}_b^{(mp)} G_i} u_{ij}^{mp}}$ ;

- 7: **until** convergence is reached.

**Algorithm 1.** Generalized framework for rough PCM based on shadowed sets.

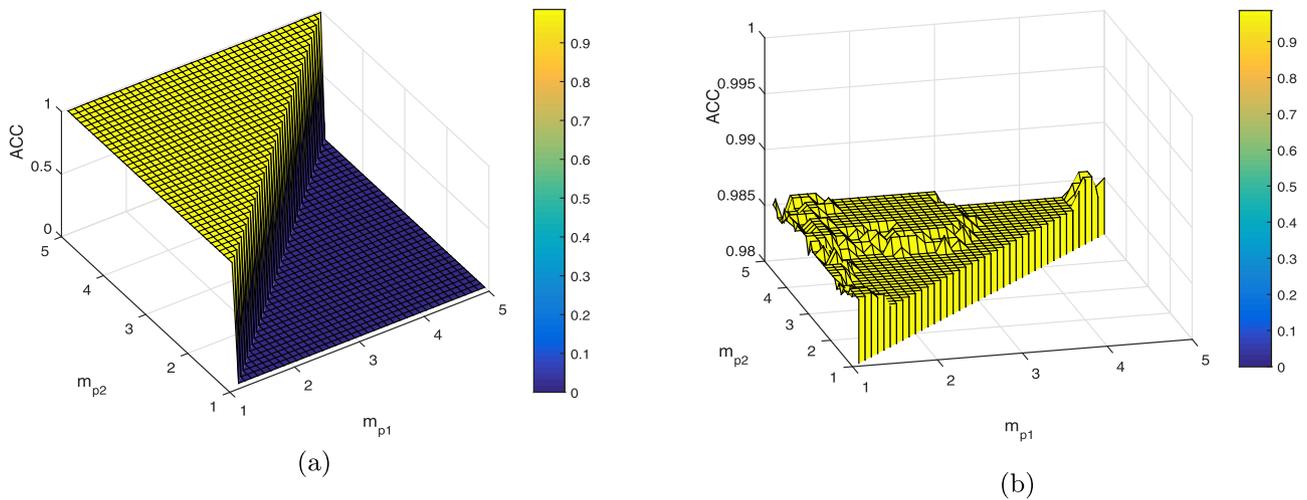


Fig. 20. ACC index for synthetic dataset II. (a) ACC index with respect to different values of  $m_{p1}$  and  $m_{p2}$ ; (b) The local profile of ACC index.

squared distances with respect to the prototype of  $G_1$  are in the interval  $(L_1, R_1)$ , as illustrated in Fig. 4. The core region and the exclusion region of cluster  $G_1$  comprise the patterns with normalized squared distances in the intervals  $[0, L_1]$  and  $[R_1, +\infty)$ , respectively. The approximation region partitions for cluster  $G_1$  are represented schematically in Fig. 5(a).

If the fuzzifier value  $m_p$  increases from  $m_{p1}$  to  $m_{p2}$  (such as  $m_{p2} = 2$ ), and the partition threshold for cluster  $G_1$  under  $m_{p2}$  equals the value obtained under  $m_{p1}$ , i.e., set  $\alpha_1^{(m_{p2})} = \alpha_1^{(m_{p1})} = \alpha$ , then the core region and the exclusion region of cluster  $G_1$  change into patterns with normalized squared distances in the intervals  $[0, L_2]$  and  $[R_2, +\infty)$ , respectively. In addition, the boundary region of cluster  $G_1$  will cover the patterns with normalized squared distances in the interval  $(L_2, R_2)$ . According to Proposition I, we have  $(L_1, R_1) \subset (L_2, R_2)$ , which means that the core region of  $G_1$  under the fuzzifier value  $m_{p1}$  (the green region in Fig. 5(a)) is contracted relative to the fuzzifier value  $m_{p2}$  using the same partition threshold (the yellow region in Fig. 5(b)). The variation in the exclusion region has the same tendency to contract. Moreover, the boundary region of  $G_1$  is extended (as shown in Fig. 5, from the blue region in (a) to the union of the green, blue, and gray regions in (b)). In this case, some patterns in the core region under  $m_{p1}$  that are definitely partitioned into cluster  $G_1$  will be partitioned into the boundary region of cluster  $G_1$  under  $m_{p2}$ , as shown by the green region in Fig. 5(b).

On the other hand, if the fuzzifier value is tuned from  $m_{p2}$  to  $m_{p1}$ ,

core region of  $G_1$  will be extended from the patterns with normalized squared distances in the interval  $[0, L_2]$  to patterns with normalized squared distances in the interval  $[0, L_1]$ , and the boundary region will become narrow. The patterns that change from the core region to the boundary region ( $m_{p1} \rightarrow m_{p2}$ ) or that change from the boundary region to the core region ( $m_{p2} \rightarrow m_{p1}$ ) with respect to a fixed cluster reflect the uncertainty generated by a single value of the fuzzifier  $m_p$ . In this manner, the uncertainty related to the fuzzifier  $m_p$  in the possibilistic clustering processes can be captured quantitatively based on the variations in the approximation regions, which can form an embedded structure according to Proposition I.

#### 4.2. Construction of multigranulation approximation regions in rough possibilistic clustering

As shown in Fig. 4, the core region of each cluster is contracted gradually as the value of the fuzzifier  $m_p$  increases under the same partition threshold for this cluster. If we assume that a series of fuzzifier values satisfy  $1 < m_{p1} < m_{p2} < \dots < m_{pk} < \dots < m_{pmax}$ , then the core regions obtained under the smaller fuzzifier values can be decomposed further, which is similar to the multigranulation approximation regions formed based on fuzzy membership degrees in [24], as shown in Fig. 6.

In Fig. 6, there are four important types of patterns that are formed in the approximation region decomposition processes, which can be described as follows:

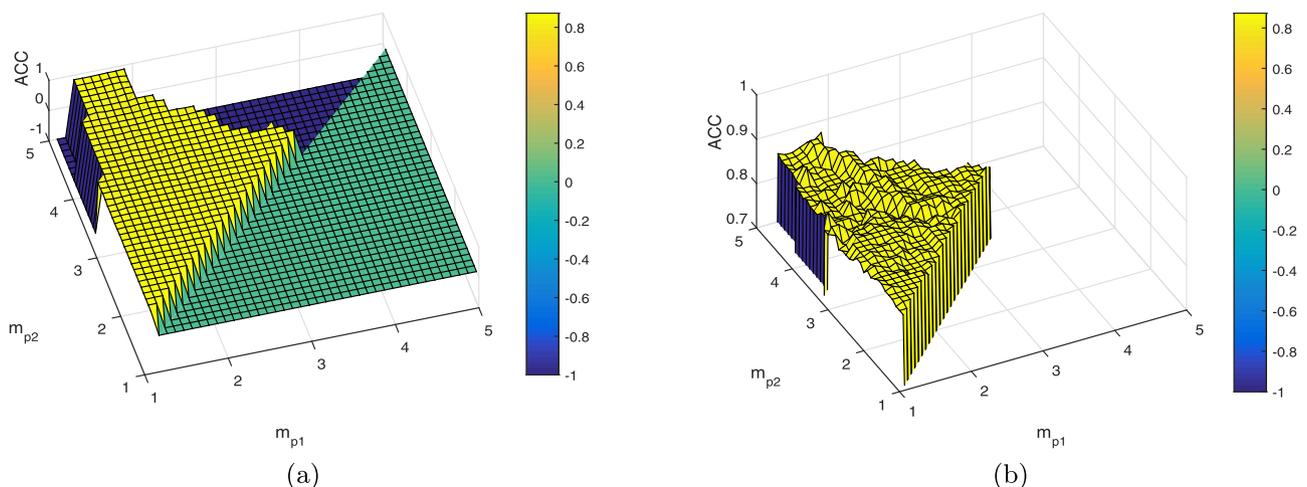


Fig. 21. ACC index for Iris. (a) ACC index with respect to different values of  $m_{p1}$  and  $m_{p2}$ ; (b) The local profile of ACC index.

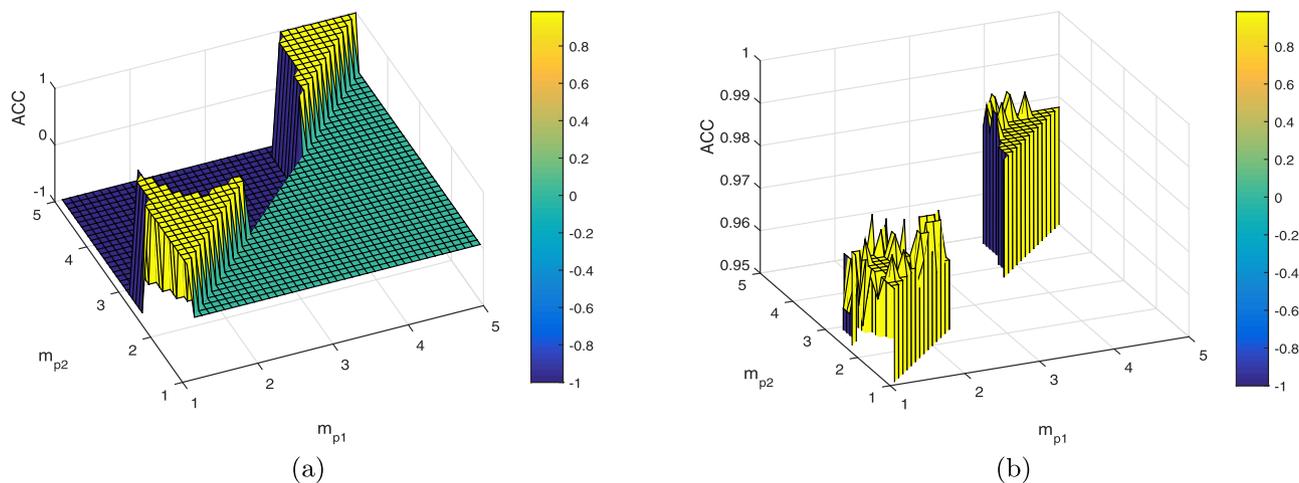


Fig. 22. ACC index for Wine. (a) ACC index with respect to different values of  $m_{p1}$  and  $m_{p2}$ ; (b) The local profile of ACC index.

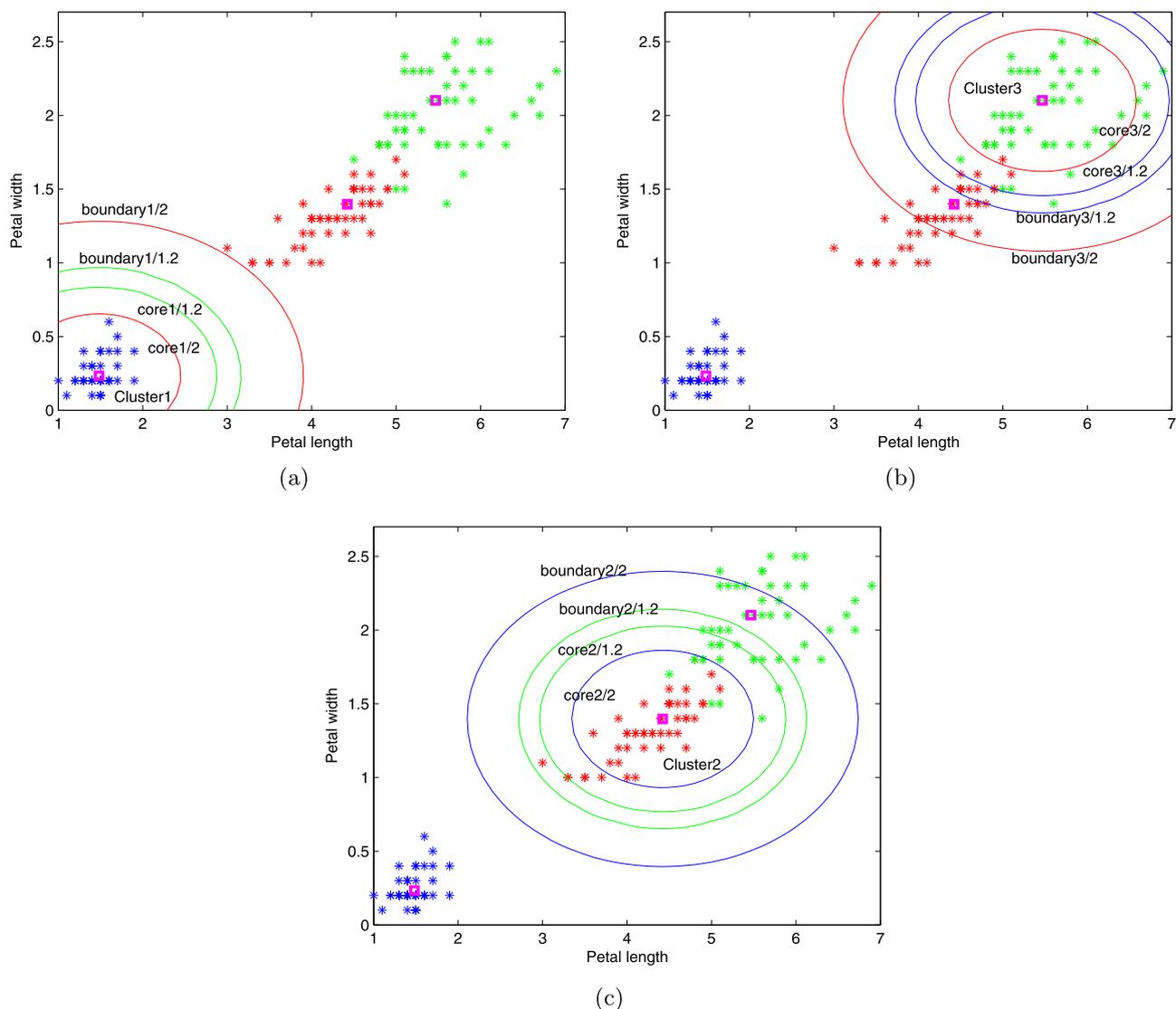


Fig. 23. The partitions of approximation regions obtained by MSPCM for Iris based on the features petal length and petal width. (a) The partitions of approximation regions of Cluster 1; (b) The partitions of approximation regions of Cluster 2; (c) The partitions of approximation regions of Cluster 3.

- (1)  $Core_{ik}$  denotes the core region of cluster  $G_i$  under the value of  $m_{p_k}$ ;
- (2)  $Bnd_{ik}$  denotes the boundary region of cluster  $G_i$  under the value of  $m_{p_k}$ ;
- (3)  $\underline{Bnd}_{ik}$  denotes the patterns that belong to the core region of cluster  $G_i$  under the value of  $m_{p_{(k-1)}}$  and that also belong to the boundary region of this cluster under the value of  $m_{p_k}$ , such as the green region in Fig. 5(b);
- (4)  $\underline{Exc}_{ik}$  denotes the patterns that belong to the exclusive region of cluster  $G_i$  under the value of  $m_{p_{(k-1)}}$  and that also belong to the boundary region of this cluster under  $m_{p_k}$ , such as the gray region in Fig. 5(b).

When the fuzzifier value tends to become larger, a new boundary region of the cluster is formed with three parts, i.e.,  $\underline{Bnd}_{ik}$ ,  $Bnd_{i(k-1)}$ , and  $\underline{Exc}_{ik}$ . The core region  $Core_{i(k-1)}$  at a higher level is also decomposed into a new one,  $Core_{ik}$ , with  $\underline{Bnd}_{ik}$  at the adjacent level below.

The approximation regions at the higher levels in Fig. 6 can be considered as coarser partitions, which may involve some noisy data or overlapping patterns in the core regions. In this case, the prototype calculations may be distorted 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 eliminated from the newly formed core region at the lower level.

The construction of multigranulation approximation regions by increasing the fuzzifier value  $m_p$  can be considered as a coarse to fine mechanism. In this case, the core region of each cluster is contracted gradually and it tends toward the natural centroid of each cluster. By contrast, the construction of multigranulation approximation regions by decreasing the fuzzifier value can be considered as a fine to coarse mechanism, i.e., the core region of each cluster will be extended gradually.

#### 4.3. Ensemble strategy for updating prototypes based on multigranulation approximation regions

According to the multigranulation core regions formed for each cluster, the corresponding prototypes can be computed independently at different levels of granularity and the final prototypes can be optimized using these candidate results. The prototype of one cluster computed at a lower level can be considered as a modification of that obtained at a higher level. This process is illustrated in Fig. 7, where  $\mathbf{v}_i^{(m_{p_k})}$  is the candidate prototype obtained for the  $i$ -th cluster  $G_i$  under the fuzzifier value  $m_{p_k}$ .

In Fig. 7, only some of the patterns in the boundary region at a lower level (such as  $m = m_{p_k}$ ) are selected for computing the prototype rather than all patterns in the boundary region at this level because the patterns in  $Bnd_{i(k-1)}$  are employed for computing the prototype at the upper level ( $m = m_{p_{(k-1)}}$ ) and the patterns in  $\underline{Exc}_{ik}$  make almost no contribution to this cluster.  $\mathbf{v}_i^{(m_{p_k})}$  can be considered as the correction for  $\mathbf{v}_i^{(m_{p_{(k-1)}})}$  because some uncertain patterns may be involved in  $Core_{i(k-1)}$ . A generalized method for calculating the prototypes  $\mathbf{v}_i$  ( $i = 1, 2, \dots, C$ ) with ensemble strategies based on multigranulation approximation regions in rough-possibilistic clustering processes is described as follows:

$$\mathbf{v}_i = \sum_{k=1}^{k_{\max}} \varphi_k \mathbf{v}_i^{(m_{p_k})}, \quad (19)$$

where  $\varphi_k \in [0, 1]$  ( $k = 1, 2, \dots, k_{\max}$ ) are weighted values,  $\sum_{k=1}^{k_{\max}} \varphi_k = 1$ , and  $\{\varphi_k\}$  measure the importance of the candidate results obtained at different levels of granularity. If the values of  $\varphi_k$  are equal for all  $k$ , then the candidate results obtained at each level of granularity make the same contribution to the updating of the prototype  $\mathbf{v}_i$ .

We have introduced a generalized multigranulation framework for calculating prototypes, but we focus on a specific case with two fuzzifier values  $m_{p_1}$  and  $m_{p_2}$ , such that the uncertainty caused by a single

fuzzifier value can be captured well and the number of iterative computations is not greatly increased. The corresponding diagram for computing the prototypes is illustrated in Fig. 8.

The process with denoted by ‘☆’ in Fig. 8 is very import for computing prototypes with ensemble strategies because without this process, Proposition 1 will not be satisfied and the multigranulation approximation regions of each cluster cannot be formed.

In particular, given two fuzzifier values  $m_{p_1}$  and  $m_{p_2}$  ( $1 < m_{p_1} < m_{p_2}$ ), the prototypes under  $m_{p_1}$  are computed as follows:

$$\mathbf{v}_i^{(m_{p_1})} = \begin{cases} w_l A_3 + w_b B_3 & \text{if } \underline{R}^{(m_{p_1})} G_i \neq \emptyset \wedge R_b^{(m_{p_1})} G_i \neq \emptyset \\ B_3 & \text{if } \underline{R}^{(m_{p_1})} G_i = \emptyset \wedge R_b^{(m_{p_1})} G_i \neq \emptyset, \\ A_3 & \text{if } \underline{R}^{(m_{p_1})} G_i \neq \emptyset \wedge R_b^{(m_{p_1})} G_i = \emptyset \end{cases} \quad (20)$$

where  $A_3 = \frac{\sum_{\mathbf{x}_j \in \underline{R}^{(m_{p_1})} G_i} (u_{ij}^{(m_{p_1})})^{m_{p_1}} \mathbf{x}_j}{\sum_{\mathbf{x}_j \in \underline{R}^{(m_{p_1})} G_i} (u_{ij}^{(m_{p_1})})^{m_{p_1}}}$ ,  $B_3 = \frac{\sum_{\mathbf{x}_j \in R_b^{(m_{p_1})} G_i} (u_{ij}^{(m_{p_1})})^{m_{p_1}} \mathbf{x}_j}{\sum_{\mathbf{x}_j \in R_b^{(m_{p_1})} G_i} (u_{ij}^{(m_{p_1})})^{m_{p_1}}}$ .  $A_3$  and  $B_3$  can be considered as the contributions by the possibilistic core region and possibilistic boundary region, respectively.  $u_{ij}^{(m_{p_1})}$  denotes the possibilistic membership degree of pattern  $\mathbf{x}_j$  belonging to the cluster with prototype  $\mathbf{v}_i$  under the fuzzifier value  $m_{p_1}$ .

Similarly, the prototypes under  $m_{p_2}$  are obtained as follows:

$$\mathbf{v}_i^{(m_{p_2})} = \begin{cases} w_l A'_3 + w_b B'_3 & \text{if } \underline{R}^{(m_{p_2})} G_i \neq \emptyset \wedge \tilde{R}_b^{(m_{p_2})} G_i \neq \emptyset \\ B'_3 & \text{if } \underline{R}^{(m_{p_2})} G_i = \emptyset \wedge \tilde{R}_b^{(m_{p_2})} G_i \neq \emptyset, \\ A'_3 & \text{if } \underline{R}^{(m_{p_2})} G_i \neq \emptyset \wedge \tilde{R}_b^{(m_{p_2})} G_i = \emptyset \end{cases} \quad (21)$$

where  $A'_3 = \frac{\sum_{\mathbf{x}_j \in \underline{R}^{(m_{p_2})} G_i} (u_{ij}^{(m_{p_2})})^{m_{p_2}} \mathbf{x}_j}{\sum_{\mathbf{x}_j \in \underline{R}^{(m_{p_2})} G_i} (u_{ij}^{(m_{p_2})})^{m_{p_2}}}$ ,  $B'_3 = \frac{\sum_{\mathbf{x}_j \in \tilde{R}_b^{(m_{p_2})} G_i} (u_{ij}^{(m_{p_2})})^{m_{p_2}} \mathbf{x}_j}{\sum_{\mathbf{x}_j \in \tilde{R}_b^{(m_{p_2})} G_i} (u_{ij}^{(m_{p_2})})^{m_{p_2}}}$ .  $u_{ij}^{(m_{p_2})}$  denotes the possibilistic membership degree of pattern  $\mathbf{x}_j$  belonging to the cluster with prototype  $\mathbf{v}_i$  under the fuzzifier value  $m_{p_2}$ .  $\tilde{R}_b^{(m_{p_2})} G_i = \underline{R}^{(m_{p_1})} G_i - \underline{R}^{(m_{p_2})} G_i = \{\mathbf{x}_j | \mathbf{x}_j \in \underline{R}^{(m_{p_1})} G_i \wedge \mathbf{x}_j \in R_b^{(m_{p_2})} G_i\}$  (i.e.,  $\underline{Bnd}_{i2}$  in Figs. 6 and 7), denotes the patterns that belong to the core region of  $G_i$  under  $m_{p_1}$  and that also belong to the boundary region of  $G_2$  under  $m_{p_2}$  based on the same partition threshold. According to Proposition 1,  $\underline{R}^{(m_{p_2})} G_i \subseteq \underline{R}^{(m_{p_1})} G_i$ , which means that some uncertain patterns partitioned into the core region of  $G_i$  under  $m_{p_1}$  can be captured in  $\tilde{R}_b^{(m_{p_2})} G_i$  and their importance for the computation of the prototype of cluster  $G_i$  needs to be reduced.

The prototype of each cluster can be combined by the following principle:

$$\mathbf{v}_i = \varphi_1 \mathbf{v}_i^{(m_{p_1})} + \varphi_2 \mathbf{v}_i^{(m_{p_2})}, \quad (22)$$

where  $\mathbf{v}_i^{(m_{p_2})}$  can be considered as the modification of  $\mathbf{v}_i^{(m_{p_1})}$ . The contributions of the patterns in the core regions under both  $m_{p_1}$  and  $m_{p_2}$  will be enhanced (the yellow region in Fig. 5(b)), and the contributions of the patterns that are removed from the core region under  $m_{p_1}$  to the boundary region under  $m_{p_2}$  will be reduced (the green region in Fig. 5(b)). The weighted values  $\varphi_1$  and  $\varphi_2$  measure the importance of the candidate results obtained under  $m_{p_1}$  and  $m_{p_2}$ , respectively. Clearly, the functions in the original PCM, PFCM, and RFCM are performed at a single level, which differs from the proposed clustering method based on multi-levels of granularity.

#### 4.4. Adaptive adjustment of scale parameters $\{\gamma_i\}$

The scale parameters  $\{\gamma_i\}$  determine the distances at which the membership degrees become 0.5, i.e., they determine the influence zone of each cluster. A pattern  $\mathbf{x}_j$  will have more (little) influence on the prototype calculation for cluster  $G_i$  if the distance  $d_{ij}^2$  is small (large) compared with  $\gamma_i$ . Xenaki et al. [13] observed that the fixed scale parameters cannot reflect the diversities among all clusters, especially closely located clusters with significantly different variances, which

- Input:** Data set  $\{x_j\}(j = 1, 2, \dots, N)$ , number of clusters  $C$ .  
**Output:** Prototypes  $\{v_i\}(i = 1, 2, \dots, C)$ .
- 1: Initialize prototypes  $v_i (i = 1, 2, \dots, C)$  based on the results obtained by FCM;
  - 2: **repeat**
  - 3: Obtain the fuzzy membership degrees  $\{\mu_{ij}\} (i = 1, 2, \dots, C, j = 1, 2, \dots, N)$ ;
  - 4: Compute the optimal partition threshold  $\alpha_i^{(m_i)}$  for each cluster  $G_i (i = 1, 2, \dots, C)$  based on the shadowed sets. According to  $\alpha_i^{(m_i)}$ , determine the approximation regions  $\underline{R}^{(m_i)}G_i$  and  $\overline{R}^{(m_i)}G_i$  for each cluster  $G_i$  with respect to  $\{u_{ij}\}$ ;
  - 5: Compute the scale parameters  $\{y_i\}$  with equation (23);
  - 6: Obtain the partition matrices  $\{u_{ij}^{(m_{p_1})}\}$  and  $\{u_{ij}^{(m_{p_2})}\}$  under the fuzzifier values  $m_{p_1}$  and  $m_{p_2}$ , respectively;
  - 7: Compute the optimal partition threshold  $\alpha_i^{(m_{p_1})}$  for each cluster  $G_i (i = 1, 2, \dots, C)$  based on the shadowed sets. According to  $\alpha_i^{(m_{p_1})}$ , determine the approximation regions  $\underline{R}^{(m_{p_1})}G_i$  and  $\overline{R}^{(m_{p_1})}G_i$  for each cluster  $G_i$  with respect to  $\{u_{ij}^{(m_{p_1})}\}$ ;
  - 8: According to  $\alpha_i^{(m_{p_1})}$ , determine the approximation regions  $\underline{R}^{(m_{p_2})}G_i$  and  $\overline{R}^{(m_{p_2})}G_i$  for each cluster  $G_i$  with respect to  $\{u_{ij}^{(m_{p_2})}\}$ ;
  - 9: Calculate the values of  $v_i^{(m_{p_1})}$  and  $v_i^{(m_{p_2})} (i = 1, 2, \dots, C)$  with equations (20) and (21), respectively;
  - 10: Update the prototypes with equation (20);
  - 11: **until** convergence is reached.

Algorithm 2. MSPCM.

may lead the algorithm to generate coincident clusters. Therefore, the scale parameters need to be dynamically adapted based on the properties of clusters during the iteration procedures, including the size and density of each cluster. The computation of the scale parameter for a specific cluster only needs to consider the patterns that are relevant to this cluster, i.e., those that possibly belong to this cluster to the greatest extent, instead of the whole data set because irrelevant patterns may distort the computation of this scale parameter.

Based on the shadowed sets, the scale parameters can be obtained as follows:

$$\gamma_i = \frac{\sum_{x_j \in \underline{R}^{(m_f)}G_i} \left(\frac{m_f}{|G_i \cup R_b^{(m_f)}G_i}\right) \mu_{ij}^{m_f} d_{ij}^2}{\sum_{x_j \in \underline{R}^{(m_f)}G_i} \left(\frac{m_f}{|G_i \cup R_b^{(m_f)}G_i}\right) \mu_{ij}^{m_f}}, \tag{23}$$

where  $\underline{R}^{(m_f)}G_i$  and  $R_b^{(m_f)}G_i$  are the core region and boundary region of cluster  $G_i$ , respectively, which are calculated with Eq. (15). The union of  $\underline{R}^{(m_f)}G_i$  and  $R_b^{(m_f)}G_i$ , i.e., the upper approximation of cluster  $G_i$ , can be considered as the maximal compatible region of cluster  $G_i$  where the patterns possibly belong to this cluster. As mentioned earlier, the fuzzy membership degrees are involved in the partitions of the approximation regions and the calculations of the scale parameters in Eq. (23), but not the possibilistic membership degrees because more patterns would be considered relevant to the fixed cluster to the maximum extent. In this manner, the influence zone of each cluster can be controlled to the greatest extent according to the structural properties of data.

4.5. Rough PCM based on multigranulation approximation regions and shadowed sets (MSPCM)

According to the solutions proposed for dealing with the partition threshold, fuzzifier value, and scale parameters, the MSPCM is summarized in Algorithm 2.

In Algorithm 2, Steps 3–5 correspond to the adaptive adjustment of the scale parameters. Steps 6–10 implement the ensemble strategy for updating prototypes where multigranulation approximation regions are formed based on two fuzzifier values.

If we consider the data set  $D_{12}$  as an example, the prototypes obtained by PCM and the proposed MSPCM algorithm are shown in Table 2, and their deviations from the ground centroids  $[-3.34, 0]$  and  $[3.34, 0]$  are also given. To illustrate the influence of the uncertainties generated by the fuzzifier values, the fuzzifier  $m_p$  is increased in steps of 0.1 from 1.1 to 2 for PCM, and the fuzzifier  $m_{p_2} = 2$  is fixed as the value of the fuzzifier  $m_{p_1}$  increases by adding 0.1 for MSPCM.

Table 2 shows that the prototypes obtained by PCM tend toward coincidental clusters as the fuzzifier value increases from 1.1 to 2. The deviations of the prototypes obtained are also large compared with the ground centroids. The prototypes obtained by MSPCM are more stable compared with those produced by PCM. All prototypes obtained by MSPCM are located close to the ground centroids and the average deviation is far lower than that obtained by PCM. The variations in the prototype locations are shown in Fig. 9.

According to the results obtained by MSPCM shown in Table 2, given  $m_{p_1} = 1.2$  and  $m_{p_2} = 2$ , the prototypes obtained are almost equal to the ground centroids. The approximation region partitions obtained under fuzzifier values of  $m_{p_1}$  and  $m_{p_2}$  are shown in Figs. 10(a) and (b), respectively. According to the same partition threshold obtained under  $m_{p_1}$ , the data pattern  $x_{12}$  is divided into the absolute exclusion region for clusters under both  $m_{p_1}$  and  $m_{p_2}$ . This situation is more reasonable compared with the results in Fig. 2 because the distances between  $x_{12}$  and Cluster 1 or Cluster 2 are further away than those for the other patterns. Thus, pattern  $x_{12}$  can be considered as an outlier and it makes no contribution to the renewal of the prototypes according to equations (20)–(22). For a fixed cluster, the core region under  $m_{p_2}$  according to the same partition threshold optimized under  $m_{p_1}$  is contracted toward the natural centroid of this cluster. Moreover, the corresponding

**Table 2**  
The prototypes obtained by PCM and MSPCM for  $D_{12}$ .

$m_p(m_{p_1})$	PCM		MSPCM( $m_{p_2} = 2$ )	
	Prototypes	Deviation	Prototypes	Deviation
1.1	[−3.1335 0][3.1335 0]	0.4130	[−3.339 0][3.339 0]	0.002
1.2	[−3.076 0][3.076 0]	0.5280	[−3.3387 0][3.3387 0.000006]	0.0026
1.3	[−3.03 0.000024][3.03 0.000024]	0.6200	[−3.3389 0.000003][3.3389 0.000003]	0.0022
1.4	[−2.9379 0.000251][2.9379 0.000251]	0.8042	[−2.8529 0][2.8529 0]	0.9742
1.5	[−2.8013 0.001081][2.8013 0.001081]	1.0774	[−2.8362 0.000001][2.8362 0.000001]	1.0076
1.6	[−2.6266 0.002935][2.6266 0.002935]	1.4268	[−2.948 0.26492][2.948 0.26492]	0.9462
1.7	[−2.4115 0.006072][2.4115 0.006072]	1.8570	[−2.9404 0.26787][2.9404 0.26787]	0.9622
1.8	[−2.106 0.010870][2.106 0.010870]	2.4681	[−3.2116 0.22197][3.2116 0.22197]	0.5129
1.9	[−0.01933 0.040750][0.01933 0.040750]	6.6418	[−3.0026 0.41687][3.0026 0.41687]	1.0726
2	[−0.00061 0.048706][0.00061 0.048706]	6.6795	[−3.0003 0.41326][3.0003 0.41326]	1.0699
AVG	/	<b>2.2516</b>	/	<b>0.65524</b>

boundary region under  $m_{p_2}$  is extended compared with that obtained under  $m_{p_1}$ , as shown in Figs. 10(c) and (d).

Furthermore, the boundary region is extended under larger fuzzifier values, and more patterns that are partitioned into an exclusion region under a lower fuzzifier value will be partitioned into the boundary region under a larger fuzzifier value. However, their contributions will not be considered when updating the prototypes according to equations (20) and (21), such as  $x_6, x_7, x_8,$  and  $x_9$  for Cluster 1. Only the contributions of patterns in the core region partitioned under  $m_{p_1}$  are subdivided into two parts under  $m_{p_2}$ . In particular, the contributions of patterns in the core region under  $m_{p_1}$  and in the core region under  $m_{p_2}$  will be enhanced, such as  $x_3$  for Cluster 1. The contributions of patterns in the core region under  $m_{p_1}$  and in the boundary region under  $m_{p_2}$  will be reduced, such as  $x_1, x_2, x_4,$  and  $x_5$  for Cluster 1. The variations in the approximation regions of Cluster 2 can be explained in a similar manner.

4.6. Time complexity analysis

The proposed algorithm MSPCM clearly has greater complexity than original PCM because of the computation time required to select the optimal partition threshold for each cluster. We assume that the number of clusters is  $C$ , the number of patterns is  $N$ , the number of features for each pattern is  $M$ , the number of iterations is  $I$ , and the number of candidate partition threshold values is  $S$ .

The time complexity for computing the fuzzy partition matrix in step 3 is  $O(C^2NM)$  and that for computing the optimal partition threshold in step 4 is  $O(SCN)$ . Thus, the time complexity for computing the scale parameters in step 5 is summarized as  $O(C^2NM+SCN)$ . A pair of fuzzifier values are involved so the asymptotical time complexity for computing the possibilistic partition matrix in step 6 is  $O(2CNM)$ . The time complexity of the computation for selecting partition thresholds with respect to  $m_{p_1}$  in step 7 is  $O(SCN)$ . The time complexity for partitioning the approximation regions in steps 7 and 8 is  $O(CN)$ , and that for computing the prototypes in step 9 is  $O(2CN)$ . Thus, the time

**Table 3**  
The properties of selected clustering methods.

	Fuzzy memberships	Possibilistic memberships	Involving approximation region partitions	Adaptive adjustment of scale parameters	Involving multiple fuzzifier values
FCM	Y	N	N	/	N
SRFCM	Y	N	Y	/	N
Type2 FCM	Y	N	N	/	Y
PCM	N	Y	N	N	N
PFCM	Y	Y	N	N	N
rRFCM	Y	Y	Y	N	N
APCM	N	Y	N	Y	N
CPCM	N	Y	Y	Y	N
Sparse PCM	Y	Y	N	N	N
MSPCM	Y	Y	Y	Y	Y

complexity of the proposed method can be summarized as  $O(I(C^2NM + SCN + 2CNM + SCN + CN + 2CN))$ , asymptotically,  $O(I(C^2NM + (S + M)CN))$ . In general,  $N \gg C$ , so the asymptotical time complexity of the proposed method approaches  $O(I(S + 2M)N)$ . No closed-form solution can be proposed for optimizing the partition thresholds, and thus enumeration methods are often exploited. For a practical problem with a large data set, if  $N \gg S$ , then the time complexity becomes  $O(INM)$ .

5. Experimental studies

We used two synthetic data sets, each of which was assigned with some noisy data, and data sets from the UCI repository [47] to compare the performance of various fuzzy clustering methods, i.e., FCM [2], Type2 FCM [16], and SRFCM [43], and possibilistic clustering algorithms, i.e., PCM [4], PFCM [6], rRFCM [14,36], APCM [13], CPCM [10], sparse PCM [9], and the proposed MSPCM method.

The fuzzification coefficient for the fuzzy clustering methods was set as  $m_f = 2$ , as generally used in most previous studies. According to the experimental performance of the selected methods, the fuzzifier parameter values for possibilistic clustering methods were set as  $m_p = 1.2$ ,  $m_{p_1} = 1.2$ , and  $m_{p_2} = 2$ . Details regarding the reasons for selecting  $m_{p_1}$  and  $m_{p_2}$  are given in the following. The results obtained by running FCM, including obtaining the prototypes and corresponding membership degrees, were employed as the initial configurations for the implementation of the selected methods. The weighted value used to evaluate the importance of core regions was set as  $w_i = 0.95$ , and it was kept as a constant for all the data sets and iterative runs. The maximum iteration number was set as 100 and the convergence condition satisfied  $\|v_i(t + 1) - v_i(t)\| < \epsilon$ , where  $t$  is an iterative step, and  $\epsilon$  was set as 0.001 for all the algorithms.

The properties of the selected methods are summarized in Table 3, where the notations “Y”, “N”, and “/” denote yes, no, and not applicable, respectively.

Five strategies can be employed in fuzzy or possibilistic clustering methods. Only one strategy or a combination of only some strategies are involved in the previous clustering studies, such as only one strategy is employed in FCM or PCM. Different strategies can handle the different types of uncertain factors when clustering data. To handle the uncertain information adequately, more strategies need to be integrated. Clearly, all the five strategies are used in MSPCM which make the proposed method robust to uncertain environments.

### 5.1. Synthetic data set I

This synthetic data set contained 50 patterns with five noisy items, and two clusters with 10 and 40 data items. The sizes of the two clusters differed greatly, as shown in Fig. 11. The prototypes and final classification results obtained by each clustering method are shown in Fig. 12.

According to Fig. 12, the PCM method obtained coincident clusters, i.e., the independent sub-objective functions for two clusters were optimized at the same point. However, the PCM method could not separate the data set as well as other methods. Rough set-based clustering methods, i.e., SRFCM, rRFCM, and MSPCM, produced better prototypes because the approximation region partitions could capture the topology of the data based on a global observation associated with a fixed cluster.

The methods based on the adaptive adjustment of scale parameters, i.e., APCM and MSPCM, achieve better performance in terms of ACC because dynamic adjustments of the scale parameters are better at reflecting the characteristics of clusters. In addition, the shadowed set-based clustering approaches, i.e., SRFCM and MSPCM, obtained better prototypes and ACC results. Only one pattern was misclassified by MSPCM, and the corresponding prototypes obtained had the best locations, especially for the small cluster.

The partitions of the approximation regions obtained by the proposed MSPCM method are shown in Fig. 13, and the possibilistic membership degrees in Fig. 14. The core region (boundary region) under  $m_{p_2} = 2$  was contracted (extended) compared with the core region (boundary region) under  $m_{p_1} = 1.2$  according to the same threshold, which is guaranteed by Proposition I. The patterns divided into core regions under  $m_{p_1}$  and those divided into boundary regions under  $m_{p_2}$  reflected the uncertainty generated by a specific fuzzifier value. Thus, the uncertainty generated by a specific fuzzifier value could be captured by detecting the variations in the approximation regions. Furthermore, we found that the approximation region partitions for two clusters were not approximately symmetrical, i.e., the borderline of the boundary region for one cluster was not the borderline of the core region for the other cluster, in a similar manner to the results shown in Fig. 2. This difference was attributable to the dynamic changes in the scale parameters and their corresponding effects on the possibilistic membership calculations.

Validity indices were employed to compare the selected methods, i.e., the normalized mutual information (NMI) [48,49], rand index (RI) [50], ACC [48], and the difference between the prototypes obtained and the ground centroids of clusters (MD). The clustering methods were better when the values of NMI, ACC, and RI were higher, or when the value of MD was lower.

The validity indices were influenced by different values of  $m_{p_1}$  and  $m_{p_2}$ . When  $m_{p_1}$  was assigned a larger value, the possibilistic memberships were around 0.5, as shown in Fig. 3. In this case, most of the patterns, including the patterns in the overlapping areas, noisy data, or outliers, were partitioned into the boundary regions of the clusters. The corresponding core regions of the clusters could be empty and the core regions had no representative capabilities, and then they were not beneficial for updating the prototypes. Thus,  $m_{p_1}$  should be set at a relatively low value. In addition, as the value of  $m_{p_2}$  increased, the core region of each cluster contracted according to Proposition I, and the core regions of some clusters were inevitably empty with a larger value of  $m_{p_2}$ ; thus, the representative capabilities of these core regions were eliminated. Therefore, the value of  $m_{p_2}$  could not be increased greatly

with respect to a fixed value of  $m_{p_1}$ .

Next, we consider the ACC index as an example, where the results in terms of the NMI index and RI index were similar. The values of the ACC index with different values of  $m_{p_1}$  and  $m_{p_2}$  are presented in Fig. 15. The ACC index was assigned a value of 0 when the value of  $m_{p_2}$  exceeded  $m_{p_1}$ , and a value of -1 if the core regions of at least one cluster became empty as the values of  $m_{p_1}$  or  $m_{p_2}$  increased. We found that the ACC index tended to -1 with larger values of  $m_{p_1}$ , or for larger values of  $m_{p_2}$  with respect to a fixed value of  $m_{p_1}$ . Values around  $m_{p_1} = 1.2$  and  $m_{p_2} = 2$  achieved the best performance, as shown in Fig. 15(b).

The validity indices obtained using the selected methods and synthetic data set I are given in Table 4.

Table 4 demonstrates that the proposed MSPCM method performed far better than PCM. MSPCM performed better than the other fuzzy clustering methods and possibilistic clustering methods because of the approximation region partition mechanism based on shadowed sets and the adaptive adjustment of the scale parameters involved in the clustering procedures.

### 5.2. Synthetic data set II

The synthetic data set employed with a mixture of Gaussian distributions is shown in Fig. 16. Synthetic data set II comprised three clusters with 200, 500, and 700 data items, where the means of the three clusters were  $v_1 = [30, -10]$ ,  $v_2 = [0, 0]$ , and  $v_3 = [40, 40]$ , respectively. The standard deviations of these three clusters were 2, 12 and 12, respectively. The data set contained 100 noisy data with a Gaussian distribution, where the mean and standard deviation were  $[-20, 60]$  and 25, respectively. Clearly, Cluster 1 was very close to Cluster 2, and Cluster 1 was far smaller than Clusters 2 and 3. Some patterns in Clusters 2 and 3 also overlapped. Thus, it was difficult to separate the patterns in synthetic data set II with partitive clustering methods, especially in a noisy environment.

The prototypes and classification labels obtained using the fuzzy clustering and possibilistic clustering methods are shown in Fig. 17.

According to Fig. 17, the proposed MSPCM method performed better compared with the other fuzzy and possibilistic clustering approaches in terms of the prototype locations and final classification labels. The shadowed set-based method SRFCM performed better with synthetic data set I, but it obtained poor results with synthetic data set II. As shown in Table 3, different techniques (strategies) are employed in different clustering methods. It is not suitable to use only one strategy or a combination of only some strategies when analyzing a complex data set, such as the data in Fig. 16. The proposed MSPCM method performed far better than other clustering methods with synthetic data set II because it integrates multiple techniques.

The partitions of the approximation regions for each cluster under the MSPCM schema are shown in Fig. 18.

According to Fig. 18, most of the noisy patterns were assigned to the absolute exclusion region for all the clusters. Thus, their effects on the calculation of the prototypes were eliminated, which was the biggest

**Table 4**  
Comparative validity indices for synthetic dataset I.

	NMI	ACC	RI	MD
FCM	0.43871	0.84	0.72571	0.0891
SRFCM	0.67063	0.94	0.8849	0.049291
Type2FCM	0.51221	0.88	0.78449	0.0759
PCM	0.15695	0.52	0.49061	0.2287
PFCM	0.43871	0.84	0.72571	0.0727
rRFCM	0.51221	0.88	0.78449	0.0391
APCM	0.67063	0.94	0.8849	0.051064
CPCM	0.380332	0.8	0.673469	0.1046
Sparse PCM	0.608927	0.92	0.849796	0.2675
MSPCM	<b>0.84372</b>	<b>0.98</b>	<b>0.96</b>	<b>0.031189</b>

difference compared with the fuzzy clustering methods and possibilistic clustering methods. For Cluster 1, as the fuzzifier value increased from  $m_{p_1}$  to  $m_{p_2}$ , some of the patterns assigned to the boundary region or core region of Cluster 2 under  $m_{p_1}$  were assigned to the boundary region of Cluster 1 under  $m_{p_2}$ . However, these patterns made no contribution to the calculation of the prototype for Cluster 1 according to Eq. (21) because they did not belong to the core region of Cluster 1 under  $m_{p_1}$ . By contrast, the core region of Cluster 1 under  $m_{p_2}$  contracted toward the mean of the distribution for Cluster 1, and the contributions of these patterns were enhanced. The same situations also occurred for Clusters 2 and 3. The variations in the membership degrees under different fuzzifier values with MSPCM are shown in Fig. 19. We found that the possibilistic membership degrees associated with Cluster 1 decreased dramatically with respect to the normalized squared distances when the fuzzifier value increased from  $m_{p_1}$  to  $m_{p_2}$ , thereby indicating that the influence zone or the size of Cluster 1 was clearly smaller than that of the other two clusters.

The ACC indices obtained with different values of  $m_{p_1}$  and  $m_{p_2}$  are shown in Fig. 20. For synthetic data set II, all core regions of the three clusters were not empty under all combinations of  $m_{p_1}$  and  $m_{p_2}$  when varied from 1–5. Thus, the ACC index values could not be assigned to –1. However, the best performance was achieved with the smaller values of  $m_{p_1}$ , where they were located in the interval (1,1.5], as shown in Fig. 20(b).

The validity indices obtained using different methods for synthetic data set II are given in Table 5.

Table 5 demonstrates that the proposed MSPCM method had the best performance in terms of the validity indices, with the highest ACC and differences between the prototypes obtained, and the natural means of the distributions were the smallest.

### 5.3. UCI data sets

Twelve benchmark data sets from the UCI repository [47] were selected for our experiments. The details of these data sets are given in Table 6.

The performance of the proposed approach is influenced by different combinations of the values of  $m_{p_1}$  and  $m_{p_2}$ , so we tested each data set using different values of  $m_{p_1}$  and  $m_{p_2}$ , where both varied from 1–5 with a step size of 0.1. We employed the ACC index to assess the results. In addition to the results shown in Figs. 15 and 20, two other typical situations for the test results are presented in Figs. 21 and 22, where were obtained based on the Iris data set and the Wine data set, respectively. We found that these two situations were very different. For the Iris data set, the ACC index values were equal to –1 with larger values of  $m_{p_1}$ . However, the ACC index could be obtained normally only at the two ends of the distributions for the combinations of the values of  $m_{p_1}$  and  $m_{p_2}$  for the Wine data set. After integrating the results obtained based on all the data sets, as shown in Figs. 15 and 20–22, the reasonable values for  $m_{p_1}$  and  $m_{p_2}$  were determined in intervals of (1,2] and [1.5,3], respectively. According to the performance of the proposed

**Table 5**  
Comparative validity indices for synthetic dataset II.

	NMI	ACC	RI	MD
FCM	0.76308	0.89929	0.90771	18.409
SRFCM	0.79768	0.92786	0.92855	11.0992
Type2 FCM	0.72796	0.84571	0.88054	27.6731
PCM	0.74746	0.88857	0.89856	19.4941
PFCM	0.80845	0.93643	0.93508	10.5097
rRFCM	0.85449	0.95857	0.95567	4.8792
APCM	0.84364	0.95643	0.95202	8.0528
CPCM	0.775118	0.921429	0.91849	7.7452
Sparse PCM	0.573384	0.688571	0.796155	32.0046
MSPCM	<b>0.9269</b>	<b>0.98429</b>	<b>0.97954</b>	<b>1.7865</b>

**Table 6**  
The information of selected datasets from UCI repository.

No.	Name	Patterns	Features	Clusters
1	Iris	150	4	3
2	Wine	178	13	3
3	Balance	625	4	3
4	Liver	345	6	2
5	Fertility	100	9	2
6	ILPD	583	10	2
7	Flowmeter A	87	37	2
8	Flowmeter B	92	52	3
9	Flowmeter C	181	44	4
10	Magic	19,020	10	2
11	Connect-4	67,557	42	3
12	Healthy Order	22,646	8	4

algorithm,  $m_{p_1}$  was assigned as 1.2 and  $m_{p_2}$  as 2 for all the data sets in the following comparative experiments. The natural centroids of the data distributions were not available, so it would have been meaningless to compare the MD index. The other indices for the experimental results are presented in Tables 7–12.

According to Tables 7–12, the proposed MSPCM method performed better than the other fuzzy and possibilistic clustering methods in terms of most of the validity indices. MSPCM performed best in terms of the ACC values with all the selected data sets. In particular, MSPCM obtained significant improvements in terms of the ACC index with the Balance, Fertility, ILPD, and Flowmeter data sets because the following techniques are employed in the proposed method.

- (1) By integrating rough set theory, all patterns obtained in the proposed approach are partitioned into three approximation regions based on the shadowed sets, which helps to capture the natural topology of the data. For a specific cluster, the approximation region partitions are completed based on global observations of the data set, rather than depending on the absolute distances or membership degrees of individual patterns, and thus the data distribution structures are reflected well. The importance of the patterns in the core regions is enhanced and the contributions of the patterns in the exclusion regions are eliminated when updating the prototypes. More importantly, the possibilistic membership values are very small for noisy data or outliers, so they are partitioned into the absolute exclusion region for all clusters to the greatest extent. Thus, they make no contributions to the updating of the prototypes. If we consider the Iris data set as an example, the approximation region partitions for each cluster are as shown in Fig. 23. All patterns in Cluster 1 were partitioned into the core region of this cluster under fuzzifier values of 1.2 and 2, as shown in Fig. 23(a). The contributions of the patterns obtained from other clusters were eliminated when computing the prototype for Cluster 1, which is the main difference compared with the FCM and PCM methods.

**Table 7**  
The comparative validity results of Iris and Wine.

	Iris			Wine		
	NMI	ACC	RI	NMI	ACC	RI
FCM	0.66593	0.84	0.83678	0.87589	0.96629	0.95429
SRFCM	0.65949	0.83333	0.83221	0.89202	0.97191	0.96134
Type2FCM	0.65949	0.83333	0.83221	0.87589	0.96629	0.95429
PCM	0.67284	0.84667	0.84152	0.61954	0.82022	0.79471
PFCM	0.65884	0.83333	0.83221	0.82078	0.94944	0.93182
rRFCM	0.66593	0.84	0.83678	0.43401	0.7191	0.72812
APCM	<b>0.72207</b>	0.84	0.83678	0.31216	0.66292	0.68539
CPCM	0.67284	0.84667	0.84152	0.89259	0.97191	0.96204
Sparse PCM	0.58790	0.63333	0.71141	0.61902	0.84831	0.81375
MSPCM	0.7037	<b>0.86667</b>	<b>0.85682</b>	<b>0.90876</b>	<b>0.97753</b>	<b>0.96915</b>

**Table 8**  
The comparative validity results of Balance and Liver.

	Balance			Liver		
	NMI	ACC	RI	NMI	ACC	RI
FCM	0.11516	0.5264	0.59251	0.008072	0.48696	0.49889
SRFCM	0.14952	0.5952	<b>0.65914</b>	8.45E-04	0.55072	0.50371
Type2 FCM	0.1077	0.504	0.58417	4.19E-07	0.55362	0.50431
PCM	0.008835	0.4624	0.43262	0.007112	0.48116	0.49926
PFCM	0.008835	0.4624	0.43262	5.04E-04	0.51014	0.49875
rRFCM	0.056842	0.4624	0.55637	3.56E-05	0.51884	0.49926
APCM	N/A	N/A	N/A	<b>0.01407</b>	0.57101	0.50866
CPCM	0.06919	0.47200	0.56353	0.000626	0.518841	0.499259
Sparse PCM	0.00712	0.46240	0.50012	6.77E-16	0.5757576	0.509828
MSPCM	<b>0.15895</b>	<b>0.696</b>	0.60997	6.79E-16	<b>0.57971</b>	<b>0.51129</b>

Notes: “N/A” means the number of prototypes obtained by APCM is less than the number of clusters given in the dataset.

**Table 9**  
The comparative validity results of fertility and ILPD.

	Fertility			ILPD		
	NMI	ACC	RI	NMI	ACC	RI
FCM	0.011321	0.52	0.49576	0.025447	0.58491	0.51358
SRFCM	0.008519	0.5	0.49495	0.022738	0.58491	0.51358
Type2 FCM	0.011321	0.52	0.49576	<b>0.03396</b>	0.58834	0.51477
PCM	0.028427	0.54	0.49818	0.009592	0.5506	0.50427
PFCM	<b>0.03548</b>	0.51	0.49515	0.015628	0.57976	0.51189
rRFCM	0.005061	0.53	0.49677	0.018804	0.5952	0.5173
APCM	N/A	N/A	N/A	N/A	N/A	N/A
CPCM	0.02	0.50	0.49495	0.01221	0.56947	0.50881
Sparse PCM	0.00737	0.58	0.50788	0.00775	0.67865	0.56285
MSPCM	0.026422	<b>0.65</b>	<b>0.5404</b>	0.001115	<b>0.71355</b>	<b>0.59051</b>

Notes: “N/A” means the number of prototypes obtained by APCM is less than the number of clusters given in the dataset.

Similarly, the patterns belonging to Cluster 1 made no contributions to the computation of the prototypes for Clusters 2 and 3. In addition, most of the patterns in the overlapping areas between Clusters 2 and 3 were partitioned into the boundary regions of these two clusters with a fuzzifier value of 1.2. Thus, these overlapping patterns made smaller contributions to the computation of the prototypes for Clusters 2 and 3.

- (2) Multiple fuzzifier parameter values are considered in the proposed method, which can adequately capture the uncertainty generated by a single fuzzifier value. Using this method, the variations in the membership degrees under different fuzzifier values can be detected. After increasing the fuzzifier value to a larger one, the multi-granularities of the core region and boundary region can effectively deal with the vagueness and uncertainty implicated in data from coarse to fine levels, especially for the uncertain patterns that belong to the core region of a cluster at higher levels of granularity

**Table 10**  
The comparative validity results of Flowmeters A and B.

	Flowmeter A			Flowmeter B		
	NMI	ACC	RI	NMI	ACC	RI
FCM	0.001337	0.528736	0.495857	0.381074	0.576087	0.680841
SRFCM	0.002799	0.540230	0.497461	0.264730	0.456522	0.490683
Type2FCM	0.001337	0.528736	0.495857	0.279688	0.467391	0.501672
PCM	0.000448	0.528736	0.495857	0.295398	0.554348	0.523172
PFCM	0.000448	0.528736	0.495857	0.308401	0.565217	0.534878
rRFCM	0.000448	0.528736	0.495857	0.481331	0.576087	<b>0.716197</b>
APCM	6.85E-16	0.597701	0.513499	4.56E-16	0.532609	0.387721
CPCM	2.19E-05	0.517241	0.494787	0.408615	0.663043	0.688963
Sparse PCM	0.001359	0.505747	0.494253	0.119643	0.413043	0.400860
MSPCM	<b>0.006722</b>	<b>0.597701</b>	<b>0.513499</b>	<b>0.504121</b>	<b>0.717391</b>	0.678691

and those that belong to the boundary region of this cluster at lower levels of granularity.

For the two overlapping clusters in the Iris data set, such as Cluster 3, five patterns that belonged to Cluster 2 were partitioned into the core region of Cluster 3 under a fuzzifier value of 1.2. The prototype computed for Cluster 3 was distorted by these five patterns when only a single fuzzifier value of 1.2 was used. After increasing the fuzzifier value to 2, the number of misclassified patterns in the core region of Cluster 3 was reduced, as shown in Fig. 23(b), where only one pattern that belonged to Cluster 2 was partitioned into the core region of Cluster 3. Using this method, the uncertainty caused by the incorrectly partitioned patterns could be detected and their influence on the computation of the prototypes was reduced. A similar situation based on the analysis of Cluster 2 is shown in Fig. 23(c).

- (3) Dynamic adjustment of the scale parameters is employed in both the APCM and MSPCM methods, but the performance of the APCM algorithm is sensitive to the model parameters selected, and thus it could not generate the true number of clusters for all data sets considered, such as the Balance, Fertility, ILPD, and Healthy Order data sets. This limitation does not affect the proposed method MSPCM, where adaptive adjustments of the scale parameters are executed based on the maximal compatible region of each cluster. Thus, MSPCM is more stable when dealing with real-world data.

## 6. Conclusions

The performance of a model developed for data analysis is influenced directly by uncertain information, such as overlapping patterns, noise, or outliers, as well as the uncertainties generated by the model parameters. In this study, by integrating several granular computing techniques, including rough sets, fuzzy sets, and shadowed sets, we proposed a novel MSPCM algorithm, which considers the potential

**Table 11**  
The comparative validity results of Flowmeter C and Magic.

	Flowmeter C			Magic		
	NMI	ACC	RI	NMI	ACC	RI
FCM	0.201851	0.392265	0.591099	0.000258	0.523502	0.501078
SRFCM	0.196491	0.392265	0.589503	0.000395	0.529916	0.501764
Type2FCM	0.215968	0.392265	0.594659	0.000960	0.539222	0.503051
PCM	0.115706	0.375691	0.503867	0.000291	0.536488	0.502637
PFCM	0.105415	0.370166	0.514549	0.000021	0.523607	0.501088
rRFCM	0.238327	0.480663	0.648680	<b>0.001179</b>	0.551788	0.505338
APCM	0.272447	0.475138	0.641805	7.12E-16	0.648370	0.544003
CPCM	0.268145	0.469613	0.647821	0.001040	0.536593	0.502652
Sparse PCM	0.002461	0.303867	0.292449	7.12E-16	0.648370	0.544003
MSPCM	<b>0.300893</b>	<b>0.491713</b>	<b>0.650645</b>	7.12E-16	<b>0.648370</b>	<b>0.544003</b>

**Table 12**  
The comparative validity results of Connect-4 and Healthy Order.

	Connect-4			Healthy order		
	NMI	ACC	RI	NMI	ACC	RI
FCM	0.000519	0.400299	0.497218	0.040440	0.374989	0.351472
SRFCM	<b>0.002945</b>	0.498305	<b>0.512411</b>	0.049518	0.383511	0.349811
Type2FCM	0.000441	0.403807	0.497413	0.043426	0.373620	0.351238
PCM	0.000386	0.425951	0.497503	0.072052	0.443257	0.406381
PFCM	0.000442	0.425004	0.497398	0.090591	0.361786	0.351587
rRFCM	0.001489	0.354915	0.500180	0.043254	0.346242	0.349046
APCM	5.47E-16	0.658303	0.503101	N/A	N/A	N/A
CPCM	0.002050	0.370724	0.500784	0.045508	0.375563	0.356650
Sparse PCM	0.000484	0.536169	0.498737	<b>0.093668</b>	0.468383	0.391802
MSPCM	5.47E-16	<b>0.658303</b>	0.503101	0.062441	<b>0.560276</b>	<b>0.443700</b>

problems caused by the three key parameters and obtains the corresponding solutions. The threshold for partitioning approximation regions is optimized according to global observations of the data based on shadowed sets, which can capture the structural topology well. Multiple fuzzifier values can produce a multigranulation structure for approximation regions and quantitatively detect the uncertainty generated by a single fuzzifier value. Dynamically adjusting the scale parameters based on the maximal compatible region of each cluster can reflect the characteristics of clusters adaptively, such as the sizes and densities of the clusters. Moreover, the typicality values can deal with noisy environments effectively. The improvements obtained by the proposed method were demonstrated in terms of several validity indices based on comparisons with other fuzzy and possibilistic clustering methods. A specific case with two fuzzifier values was only considered in this study, and thus the statistical properties of models using multiple (more than two) fuzzifier values require further study. In addition, the effectiveness of using different optimization principles to form shadowed sets in the proposed approach will be investigated in our future research.

**Acknowledgements**

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

**Appendix**

The proofs of Proposition I

**Proof.** Since  $f_{PCM}(x_1, m_{p_1}) = f_{PCM}(x_2, m_{p_2}) > 0$ , it has  $x_1^{\frac{1}{m_{p_1}-1}} = x_2^{\frac{1}{m_{p_2}-1}}$ .

By using a logarithmic function, it has:  $\log\left(x_1^{\frac{1}{m_{p_1}-1}}\right) = \log\left(x_2^{\frac{1}{m_{p_2}-1}}\right)$ .

Further, it has:  $\frac{1}{(m_{p_1}-1)} \log(x_1) = \frac{1}{(m_{p_2}-1)} \log(x_2)$ . So the signs of  $\log(x_1)$  and  $\log(x_2)$  must be the same.

(1) Since  $m_{p_1} - 1 > 0$ ,  $m_{p_2} - 1 > 0$  and  $0 < x_1, x_2 < 1$ , it has  $\frac{1}{(m_{p_1}-1)} \log(x_1) < 0$  and  $\frac{1}{(m_{p_2}-1)} \log(x_2) < 0$ . Due to  $\frac{1}{m_{p_1}-1} > \frac{1}{m_{p_2}-1} > 0$ , such that  $\log(x_2) < \log(x_1) < 0$ , viz,  $x_2 < x_1$ .

(2) Since  $m_{p_1} - 1 > 0$ ,  $m_{p_2} - 1 > 0$  and  $x_1, x_2 > 1$ , it has  $\frac{1}{(m_{p_1}-1)} \log(x_1) > 0$  and  $\frac{1}{(m_{p_2}-1)} \log(x_2) > 0$ . Due to  $\frac{1}{m_{p_1}-1} > \frac{1}{m_{p_2}-1} > 0$ , such that  $0 < \log(x_1) < \log(x_2)$ , viz,  $x_1 < x_2$ . □

**References**

- [1] C.C. Aggarwal, C.K. Reddy, Data Clustering: Algorithms and Applications, Chapman & Hall/CRC Press, 2013.
- [2] J.C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Kluwer Academic Publishers, Norwell, MA, USA, 1981.
- [3] R.N. Dave, R. Krishnapuram, Robust clustering methods: a unified view, IEEE Trans. Fuzzy Syst. 5 (2) (1997) 270–293.
- [4] R. Krishnapuram, J.M. Keller, A possibilistic approach to clustering, IEEE Trans. Fuzzy Syst. 1 (1993) 98–110.
- [5] R. Krishnapuram, J.M. Keller, The possibilistic c-means algorithm: insights and recommendations, IEEE Trans. Fuzzy Syst. 4 (3) (1996) 385–393.
- [6] N.R. Pal, K. Pal, J.M. Keller, J. C. Bezdek, A possibilistic fuzzy c-means clustering algorithm, IEEE Trans. Fuzzy Syst. 13 (4) (2005) 517–530.
- [7] J.S. Zhang, Y.W. Leung, Improved possibilistic c-means clustering algorithms, IEEE

- Trans. Fuzzy Syst. 12 (2) (2004) 209–217.
- [8] S.D. Xenaki, K.D. Koutroumbas, A.A. Rontogiannis, Sparsity-aware possibilistic clustering algorithms, *IEEE Trans. Fuzzy Syst.* 24 (6) (2016) 1611–1626.
- [9] K.D. Koutroumbas, S.D. Xenaki, A.A. Rontogiannis, On the convergence of the sparse possibilistic c-means algorithm, *IEEE Trans. Fuzzy Syst.* 26 (1) (2018) 324–337.
- [10] H.Y. Yu, J.L. Fan, Cutset-type possibilistic c-means clustering algorithm, *Appl. Soft Comput.* 64 (2018) 401–422.
- [11] Z. Pawlak, Rough sets, *Int. J. Comput. Inf. Sci.* 11 (5) (1982) 314–356.
- [12] P. Maji, S.K. Pal, Rough set based generalized fuzzy c-means algorithm and quantitative indices, *IEEE Trans. Syst. Man. Cybern. B* 37 (6) (2007) 1529–1540.
- [13] S.D. Xenaki, K.D. Koutroumbas, A.A. Rontogiannis, A novel adaptive possibilistic clustering algorithm, *IEEE Trans. Fuzzy Syst.* 24 (4) (2016) 791–810.
- [14] P. Maji, S. Roy, Rough-fuzzy clustering and multiresolution image analysis for text-graphics segmentation, *Appl. Soft Comput.* 30 (2015) 705–721.
- [15] J.P. Sarker, 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.
- [16] 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.
- [17] E. Rubio, O. Castillo, P. Melin, Interval type-2 fuzzy possibilistic c-means clustering algorithm, in: L.A. Zadeh, et al. (Ed.), *Recent Developments and New Direction in Soft-Computing Foundations and Applications*, Studies in Fuzziness and Soft Computing, 2016, pp. 185–194.
- [18] M.R.N. Kalhori, M.H.F. Zarandi, Interval type-2 credibilistic clustering for pattern recognition, *Pattern Recognit.* 48 (11) (2015) 3652–3672.
- [19] E. Rubio, O. Castillo, F. Valdez, et al., An extension of the fuzzy possibilistic clustering algorithm using type-2 fuzzy logic techniques, *Adv. Fuzzy Syst.* 2017 (1) (2017) 1–23.
- [20] Y.Y. Yao, Granular computing: basic issues and possible solutions, *Proc. of the 5th Joint Conf. Inf. Sci.* (1999), pp. 186–189.
- [21] S. Salehi, A. Selamat, H. Fujita, Systematic mapping study on granular computing, *Knowl. Based Syst.* 80 (2015) 78–97.
- [22] W.H. Xu, W.T. Li, Granular computing approach to two-way learning based on formal concept analysis in fuzzy datasets, *IEEE Trans. Cybern.* 46 (2) (2016) 366–379.
- [23] B.Z. Sun, W.M. Ma, Y.H. Qian, Multigranulation fuzzy rough set over two universes and its application to decision making, *Knowl. Based Syst.* 123 (2017) 61–74.
- [24] J. Zhou, Z.H. Lai, D.Q. Miao, et al., Multigranulation rough-fuzzy clustering based on shadowed sets, *Inf. Sci.* (2018), <https://doi.org/10.1016/j.ins.2018.05.053>.
- [25] X.B. Yang, S.P. Xu, H.L. Dou, et al., Multigranulation rough set: a multiset based strategy, *Int. J. Comput. Intell. Syst.* 10 (2017) 277–292.
- [26] X.H. Zhang, D.Q. Miao, C.H. Liu, et al., Constructive methods of rough approximation operators and multigranulation rough sets, *Knowl. Based Syst.* 91 (2016) 114–125.
- [27] W.H. Xu, J.H. Yu, A novel approach to information fusion in multi-source datasets: a granular computing viewpoint, *Inf. Sci.* 378 (2017) 410–423.
- [28] X. Yang, T.R. Li, H. Fujita, D. Liu, Y.Y. Yao, A unified model of sequential three-way decisions and multilevel incremental processing, *Knowl. Based Syst.* 134 (2017) 172–188.
- [29] H.Q. Truong, L.T. Ngo, W. Pedrycz, Granular fuzzy possibilistic c-means clustering approach to DNA microarray problem, *Knowl. Based Syst.* 133 (2017) 53–65.
- [30] J. Hu, T.R. Li, H.J. Wang, H. Fujita, Hierarchical cluster ensemble model based on knowledge granulation, *Knowl. Based Syst.* 91 (2016) 179–188.
- [31] H. Fujita, A. Gaeta, V. Loia, F. Orciuoli, Resilience analysis of critical infrastructures: a cognitive approach based on granular computing, *IEEE Trans. Cybern.* (2018), <https://doi.org/10.1109/TCYB.2018.2815178>.
- [32] Y.G. Jing, T.R. Li, H. Fujita, Z. Yu, B. Wang, An incremental attribute reduction approach based on knowledge granularity with a multi-granulation view, *Inf. Sci.* 411 (2017) 23–38.
- [33] W.H. Xu, Y.T. Guo, Generalized multigranulation double-quantitative decision-theoretic rough set, *Knowl. Based Syst.* 105 (1) (2016) 190–205.
- [34] W. Pedrycz, Shadowed sets: representing and processing fuzzy sets, *IEEE Trans. Syst. Man. Cybern. B* 28 (1998) 103–109.
- [35] S. Mitra, H. Banka, W. Pedrycz, Rough-fuzzy collaborative clustering, *IEEE Trans. Syst. Man. Cybern. B* 36 (4) (2006) 795–805.
- [36] 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.
- [37] 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.
- [38] C. Gao, Z.H. Lai, J. Zhou, et al., Maximum decision entropy-based attribute reduction in decision-theoretic rough set model, *Knowl. Based Syst.* 143 (2018) 179–191.
- [39] P. Lingras, C. West, Interval set clustering of web users with rough k-means, *J. Intell. Inf. Syst.* 23 (1) (2004) 5–16.
- [40] W. Pedrycz, From fuzzy sets to shadowed sets: interpretation and computing, *Int. J. Intell. Syst.* 24 (2009) 48–61.
- [41] P. Grzegorzewski, Fuzzy number approximation via shadowed sets, *Inf. Sci.* 225 (2013) 35–46.
- [42] 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.
- [43] J. Zhou, W. Pedrycz, D.Q. Miao, Shadowed sets in the characterization of rough-fuzzy clustering, *Pattern Recognit.* 44 (8) (2011) 1738–1749.
- [44] Y.Y. Yao, An outline of a theory of three-way decisions, *Proc. Rough Sets Curr. Trends Comput., LNCS*, vol. 7413, (2012), pp. 1–17.
- [45] Y.Y. Yao, Three-way decisions and cognitive computing, *Cognit. Comput.* 8 (2016) 543–554.
- [46] X.F. Deng, Y.Y. Yao, Decision-theoretic three-way approximations of fuzzy sets, *Inf. Sci.* 279 (2014) 702–715.
- [47] M. Lichman, UCI machine learning repository, Irvine, CA: University of California, School of Information and Computer Science, (2013). [<http://archive.ics.uci.edu/ml>]
- [48] Q.Q. Gu, C. Ding, J.W. Han, On trivial solution and scale transfer problems in graph regularized NMF, *Int. Joint Conf. Artif. Intell.* (2011), pp. 1288–1293.
- [49] Y. Lei, J.C. Bezdek, J. Chan, et al., Extending information-theoretic validity indices for fuzzy clustering, *IEEE Trans. Fuzzy Syst.* 25 (4) (2016) 1013–1018.
- [50] 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.