

# 基于阴影集的粗糙模糊可能性 C 均值聚类算法

汪海良 余 堃 周明天

(电子科技大学计算机科学与工程学院 成都 610054)

**摘 要** 相对于硬聚类算法,软聚类算法可以更好地表示具有不精确边界的类簇。粗糙集和模糊集均是用于描述不确定数据的有效的数学工具,二者互为补充。研究人员已经将粗糙集和模糊集的概念相结合,并应用到聚类算法中,提出了粗糙模糊可能性 C 均值聚类算法。而文中通过引入阴影集,有效地解决了粗糙模糊可能性 C 均值聚类算法中的阈值选择问题。

**关键词** 粗糙集,阴影集,FCM,PCM

**中图法分类号** TP18 **文献标识码** A

## Shadowed Sets-based Rough Fuzzy Possibilistic C-means Clustering

WANG Hai-liang SHE Kun ZHOU Ming-tian

(School of Computer Science & Engineering, University of Electronic Science & Technology of China, Chengdu 610054, China)

**Abstract** It has been shown that soft clustering is advantageous over hard clustering in describing clusters without crisp boundaries. Both rough sets and fuzzy sets are effective mathematical tools in handling uncertainty. As claimed in many studies, they are complementary. The theories of rough sets and fuzzy sets have been integrated into clustering algorithms, such as rough fuzzy possibilistic C-means clustering (RFPCM). In this study, we introduced the shadowed sets optimization theory and proposed an objective method to select the threshold in RFPCM.

**Keywords** Rough sets, Shadow sets, FCM, PCM

## 1 引言

聚类算法是一种重要的非监督学习算法。通常,聚类算法可以描述为如下过程:给定一个有限数据集,将其中对象划分到不同的类簇中,使同一类簇中的对象尽可能的相似而不同类簇中的对象差异尽可能大。聚类分析在图像分割、机器视觉、数据挖掘、模式识别、特征提取等领域均有广泛的应用。

K 均值算法(k-means)<sup>[1]</sup>是一种典型的划分式聚类算法。在 k-means 算法中,每个对象均被严格地划分到某一个类簇中,每个类簇均具有精确的边界。但是在实际的应用中,不同的类簇之间往往存在互相重叠的部分。为了更好地描述这种具有不精确边界的类簇,模糊集的概念被引入到了聚类算法中。

1965 年,Zadeh 提出了模糊集的概念<sup>[2]</sup>。在模糊集中,对象可以部分地属于一个集合,并使用隶属度来表示对象属于该集合的程度。基于模糊集的聚类算法可以有效地表示一个对象同时属于多个类簇的情况。到目前为止,已经有多种基于模糊集的聚类算法被提出,包括模糊 C 均值聚类算法(fuzzy c-means, FCM)<sup>[3,4]</sup>和基于可能性的模糊 C 均值聚类算法(possibilistic fuzzy c-means, PCM)<sup>[5]</sup>。

粗糙集<sup>[6]</sup>是另一种处理不精确、不确定和不一致数据的数学工具。Lignras 等人将 K 均值聚类算法和粗糙集相结合,提出了基于粗糙集的 C 均值聚类算法(rough c-means,

RCM)<sup>[7]</sup>。RCM 算法利用粗糙集中上下近似的概念来表示类簇之间不精确的边界。Peters 讨论了 RCM 算法中的离群点的表示<sup>[8]</sup>,并且提出了针对 RCM 算法的若干改进意见<sup>[9]</sup>。

研究表明,粗糙集和模糊集在处理不精确数据方面互为补充。研究人员正致力于实现粗糙聚类和模糊聚类的结合和统一,目前已经提出了多种粗糙模糊聚类算法<sup>[10-12]</sup>。

阴影集是模糊集的一种新的表示形式<sup>[13]</sup>。阴影集通过引入阈值  $\alpha$  将模糊集映射到三值逻辑,从而实现了模糊集和粗糙集概念的统一。利用阴影集可以简化 FCM 算法的复杂度<sup>[14]</sup>,也可以实现粗糙聚类和模糊聚类算法的统一<sup>[15,16]</sup>。

本文提出了一种基于阴影集的粗糙模糊可能性 C 均值聚类算法,其利用粗糙集表示类簇之间的重叠部分,利用模糊集表示类簇的不精确边界,同时通过引入阴影集来解决粗糙聚类过程中的阈值选择问题。

## 2 背景知识

本节简要回顾了基于模糊集、粗糙集和阴影集的各种典型划分式聚类算法。

### 2.1 模糊聚类算法

在模糊聚类算法中,一个对象可以同时属于多个类簇,使用隶属度函数来表示一个对象属于某个类簇的程度。模糊聚类的结果通常表示为  $U=(\mu_{ij})_{k \times n}$  形式的隶属度矩阵,其中元素  $\mu_{ij}$  表示对象  $x_j$  隶属于类簇  $C_i$  的程度,并满足  $0 \leq \mu_{ij} \leq 1$ 。

到稿日期:2012-03-18 返修日期:2012-07-15

汪海良(1976-),男,博士,主要研究方向为粗糙集理论、决策分析。

根据隶属度  $\mu_{ij}$  的计算方式的不同,可以得到不同的模糊均值聚类算法。

给定一个数据集  $X = \{x_1, x_2, \dots, x_n\}$  和类簇的数目  $k$ , FCM 算法<sup>[4]</sup>将  $n$  个对象分别划分到  $k$  个模糊子集中,并使目标函数(1)达到最小:

$$J_m(\mu, c) = \sum_{i=1}^k \sum_{j=1}^n (\mu_{ij})^m \|x_j - c_i\|^2 \quad (1)$$

式中,  $c_i$  表示第  $i$  个模糊子集的中心,其计算方式如式(2)所示:

$$c_i = \frac{\sum_{j=1}^n (\mu_{ij})^m x_j}{\sum_{j=1}^n (\mu_{ij})^m}, i=1, 2, \dots, k \quad (2)$$

式中,  $m$  是大于 1 的常数,用来控制 FCM 算法的模糊程度。

$\|x_j - c_i\|$  表示对象  $x_j$  到模糊子集  $C_i$  中心点  $c_i$  之间的距离,  $\mu_{ij}$  表示对象  $x_j$  隶属于模糊子集  $C_i$  的程度,其计算方式如式(3)所示:

$$\mu_{ij} = \frac{\|x_j - c_i\|^{-\frac{2}{m-1}}}{\sum_{l=1}^k \|x_j - c_l\|^{-\frac{2}{m-1}}} \quad (3)$$

显然,在 FCM 算法中,隶属度  $\mu_{ij}$  具有如式(4)所示的性质:

$$\sum_{i=1}^k \mu_{ij} = 1 \quad (4)$$

因此在 FCM 算法中,隶属度  $\mu_{ij}$  也可以解释为对象  $x_j$  隶属于模糊类簇  $C_i$  的概率或  $x_j$  在不同类簇之间的共享程度,因此 FCM 算法又被称为基于概率(probability)的模糊 C 均值聚类算法。

然而,并不是在所有的应用中,隶属度都能用概率模型来解释。为此, Krishnapuram 和 Keller 提出了基于可能性的 PCM 聚类算法(possibilistic c-means)。

PCM 算法修改了隶属度  $\mu_{ij}$  的计算方式。对象  $x_j$  隶属于  $C_i$  的程度仅由  $x_j$  到  $C_i$  的中心点的距离决定,而与  $x_j$  到其他类簇的中心点的距离无关。

由于在 PCM 算法中,隶属度不再满足同一对象对所有类簇的隶属度的和为 1 的约束,为了避免得到平凡解(所有的  $\mu_{ij}$  均等于 0),PCM 在式(1)的基础上增加了约束项,在 PCM 算法中需要最小化的目标函数如式(5)所示:

$$J_m(\mu, c) = \sum_{i=1}^k \sum_{j=1}^n (\mu_{ij})^m \|x_j - c_i\|^2 + \sum_{i=1}^k \eta_i \sum_{j=1}^n (1 - \mu_{ij})^m \quad (5)$$

在 PCM 算法中,中心点  $c_i$  的计算方式与 FCM 算法相同,但隶属度  $\mu_{ij}$  的计算方式如式(6)所示:

$$\mu_{ij} = \frac{1}{1 + (\frac{\|x_j - c_i\|^2}{\eta_i})^{\frac{1}{m-1}}} \quad (6)$$

式中,  $m$  为大于 1 的常数,而参数  $\eta_i$  决定了对象与类簇中心之间的绝对距离如何转换成相对的隶属程度,一种常用的  $\eta_i$  的计算方式如式(7)所示:

$$\eta_i = K \frac{\sum_{j=1}^n \mu_{ij}^m \|x_j - c_i\|^2}{\sum_{j=1}^n \mu_{ij}^m} \quad (7)$$

式中,  $K$  为常数,通常  $K=1$ 。

然而人们在实验中发现,PCM 算法往往容易产生中心重叠的类簇。究其原因,问题在于 PCM 算法中的约束条件太弱。为了克服这一缺陷,研究人员将 FCM 算法和 PCM 算法

相结合,提出了 FPCM 算法<sup>[17]</sup>和 PFCM 算法<sup>[18]</sup>。

## 2.2 粗糙聚类算法

Polawak 提出的粗糙集模型<sup>[6]</sup>是一种用来处理不精确知识的数学方法。粗糙集利用集合的边界来表示数据的不精确性。设  $U$  代表对象的全集,  $R \subseteq U \times U$  代表定义在  $U$  上的不可分关系。给定对象  $x$ ,所有与  $x$  不可区分的对象构成了  $x$  相对于  $R$  的等价类,记为  $[x]_R$ 。对  $U$  的任一子集  $X$ ,其相对于关系  $R$  的粗糙近似定义为:

$$\text{下近似: } \underline{R}(X) = \{x \in X | [x]_R \subseteq X\}$$

$$\text{上近似: } \overline{R}(X) = \{x \in X | [x]_R \cap X \neq \emptyset\}$$

$$\text{边界: } BNP(X) = \overline{R}(X) - \underline{R}(X)$$

粗糙下近似代表了所有必然属于  $X$  的元素的集合,上近似代表了所有可能属于  $X$  的元素的集合,上下近似之差的边界则反映了  $X$  相对于  $R$  的不确定性。

Lingras 和 West 将粗糙集应用到 C 均值算法中,提出了基于粗糙集的 C 均值算法(RCM)<sup>[7]</sup>。在 RCM 算法中,每个类簇都表示为一个中心点和一对上下近似,下近似中包含明确属于该类簇的对象,而上近似中包含可能属于该类簇的对象。在 RCM 算法中,每个对象满足如下性质:

1) 一个对象最多只能属于一个类簇的下近似;

2) 如果对象  $x_j$  属于一个类簇的下近似,那么它一定属于该类簇的上近似;

3) 如果对象  $x_j$  不属于任何一个类簇的下近似,那么它一定属于多个类簇的上近似。

RCM 算法在执行过程中,需要事先设置 3 个参数:上、下近似的权重参数  $w_l$  和  $w_u$  以及阈值参数  $\delta$ 。

其中  $w_l$  和  $w_u$  分别代表了在计算类簇中心时,上、下近似中的对象对中心位置的影响权重。在 RCM 算法,类簇的中心位置的计算方式如式(8)所示:

$$c_i = \begin{cases} w_l \times \frac{\sum_{x \in \underline{C}_i} x}{|\underline{C}_i|} + w_u \times \frac{\sum_{x \in \overline{C}_i} x}{|\overline{C}_i|}, & \underline{C}_i \neq \overline{C}_i \\ w_l \times \frac{\sum_{x \in \underline{C}_i} x}{|\underline{C}_i|}, & \text{otherwise} \end{cases} \quad (8)$$

式中,  $\underline{C}_i$  表示第  $i$  个类簇的下近似,  $\overline{C}_i$  表示第  $i$  个类簇的上近似,而  $B_i = \overline{C}_i - \underline{C}_i$ 。

在聚类过程中,首先将对象到不同类簇中心点的距离差与阈值  $\delta$  进行比较,如果差值小于  $\delta$ ,表示该对象同时隶属于多个类簇,应该被划分到类簇的上近似中,反之则应被划分到中心点距其最近的类簇的下近似中。

## 2.3 基于阴影集的粗糙模糊聚类

粗糙聚类算法中权重和阈值参数的选择对聚类结果有直接的影响,目前研究人员已经提出了多种参数选择算法<sup>[10,20,21]</sup>。而将基于阴影集的粗糙模糊聚类算法引入阴影集的概念,来解决阈值选择的问题。

阴影集是模糊集的一种新的表示形式<sup>[13]</sup>。阴影集中只保留了最重要的模糊信息,简化了模糊集表示。

阴影集引入阈值  $\alpha$  来决定保留哪些模糊信息。给定模糊集  $f(x)$ ,如果  $f(x) > 1 - \alpha$ ,则令  $f(x) = 1$ ;如果  $f(x) < \alpha$ ,则令  $f(x) = 0$ 。在阴影集中,只有隶属度在区间  $[\alpha, 1 - \alpha]$  内的元素保留了模糊性,而其余的元素都被明确地分类,因此阴影集相当于将模糊集  $X$  映射到三值逻辑:  $x: f(x) \rightarrow 0, 1, (0, 1)$ 。

一个典型的阴影集如图 1 所示。

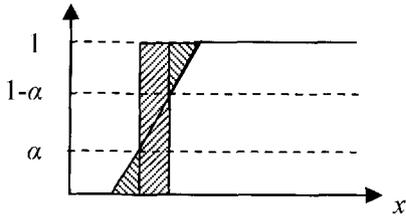


图 1 阴影集

对于连续的隶属度函数,  $\alpha$  的取值应满足使式(9)取得最小值:

$$V = \left| \int_{-\infty}^{\alpha} f(x) dx + \int_{1-\alpha}^{+\infty} (1-f(x)) dx - \int_{\alpha}^{1-\alpha} dx \right| \quad (9)$$

而对于离散的隶属度函数,  $\alpha$  的取值应满足使式(10)取得最小值:

$$V = \left| \sum_{i, f(x_i) < \alpha} f(x_i) + \sum_{i, f(x_i) > 1-\alpha} (1-f(x_i)) - \text{card}\{x_i \in X \mid \alpha < f(x) < 1-\alpha\} \right| \quad (10)$$

式中,  $|\cdot|$  代表取绝对值运算, 而  $\text{card}\{\cdot\}$  代表集合中对象的数量。

阴影集中的三值逻辑和粗糙集在概念上有相通之处, 隶属度为 0 的对象表示一定不属于  $X$  的成员, 对应于粗糙集中的负域; 隶属度为 1 的对象表示一定属于  $X$  的成员, 对应于粗糙集的下近似; 而隶属度属于区间  $[0, 1]$  的对象则对应于粗糙集中的边界区域(上下近似之差)。

最初, Yu 等人将阴影集用于改进模糊聚类算法, 提出了基于阴影集的模糊聚类算法(fuzzy c-means clustering based on shadow sets, S-FCM)<sup>[14]</sup>。Mitra 等人提出了阴影集 C 均值聚类(shadowed c-means, SCM)<sup>[15]</sup>, SCM 算法是模糊聚类和粗糙聚类的结合体。相对于传统的模糊聚类算法, 阴影集聚类只保留了必要的模糊性, 因此减少了计算量, 而相对于粗糙集聚类, 阴影集聚类可以自动地选择阈值参数的大小。

Zhou 等人提出了基于阴影集的粗糙模糊 C 均值聚类(shadowed set-based rough-fuzzy c-means, S-RFCM)<sup>[16]</sup>。S-RFCM 利用阴影集来决定阈值, 为每个类簇定义了相对近似区域, 最后还提出了基于粒化和反粒化(granulation-degranulation)机制的有效性验证指数。

利用阴影集确定阈值参数的过程如下: 首先利用 FCM 算法求对象隶属于每个类簇的模糊隶属度函数, 然后利用公式(10)对每个模糊类簇  $C_i$  求阴影集阈值  $\alpha_i$ , 凡是隶属度大于  $\alpha_i$  的对象归类到  $C_i$  的上近似  $\bar{C}_i$ , 而隶属度大于  $1-\alpha_i$  的对象则归类到  $C_i$  的下近似  $\underline{C}_i$ 。

### 3 基于阴影集的粗糙模糊可能性 C 均值聚类算法

文献[18]中提出了一种结合 FCM 算法和 PCM 算法的模糊可能性 C 均值聚类算法(possibilistic fuzzy c-means, PF-CM)。该算法在聚类过程中, 同时计算一个对象隶属于某个类簇的概率(probability)和典型性(typicality), 并引入参数  $a$ ,  $b$  来控制 FCM 和 PCM 在聚类过程中所占的重要程度。文献[12]中将 PFCM 算法和粗糙聚类相结合, 提出了一种基于粗糙集的通用模糊聚类算法(rough-fuzzy possibilistic c-means, RFPCM)。

本文在此基础之上, 引入阴影集来实现粗糙聚类中阈值

的自动选择, 提出了基于阴影集的粗糙模糊可能性 C 均值聚类算法(shadowed sets-based rough fuzzy possibilistic c-means, S-RFPCM)。

在 S-RFPCM 算法中, 对象隶属于某个类簇的隶属度既包含了对象属于该类簇的概率, 又包含了属于该类簇的典型程度, 其计算方式如式(11)所示:

$$\mu_{ij} = a(u_{ij})^{m_1} + b(t_{ij})^{m_2} \quad (11)$$

式中,  $a, b, m_1, m_2$  均为常数, 且满足  $a > 0, b > 0, m_1 > 1, m_2 > 1$ 。  $u_{ij}, t_{ij}$  的计算方式如下:

$$u_{ij} = \frac{d_{ij}^{\frac{-2}{m_1-1}}}{\sum_{j=1}^K d_{ij}^{\frac{-2}{m_1-1}}}, t_{ij} = \frac{1}{1 + (d_{ij}^2)_{m_2-1}^{-1}}$$

式中,  $d_{ij}$  代表对象  $x_j$  到类簇中心  $c_i$  的欧几里得距离( $d_{ij} = \|x_j - c_i\|$ )。而  $\eta_j$  的计算方式如下:

$$\eta_j = K \frac{\sum_{i=1}^n t_{ij}^2 d_{ij}^2}{\sum_{j=1}^n t_{ij}^m}$$

式中,  $K$  为常数, 通常  $K=1$ 。

S-RFPCM 算法需要最小化的目标函数如式(12)所示:

$$J_{S-RFPCM} = \begin{cases} w_l \times A + w_u \times B, & \underline{C}_i \neq \bar{C}_i \\ A, & \text{otherwise} \end{cases} \quad (12)$$

式中,  $w_l$  和  $w_u$  分别代表了下近似和边界中的对象在聚类过程中所占的不同权重, 通常有  $0 < w_l < 1, 0 < w_u < 1$ , 并且满足  $w_l + w_u = 1$ 。

式(11)中,  $A$  和  $B$  分别代表了下近似和边界中的对象, 其计算方式如下:

$$A = \sum_{i=1}^k \sum_{x_j \in \underline{C}_i} \{a(u_{ij})^{m_1} + b(t_{ij})^{m_2}\} d_{ij}^2 + \sum_{i=1}^k \eta_j \sum_{x_j \in \underline{C}_i} (1-t_{ij})^{m_2}$$

$$B = \sum_{i=1}^k \sum_{x_j \in \bar{C}_i} \{a(u_{ij})^{m_1} + b(t_{ij})^{m_2}\} d_{ij}^2 + \sum_{i=1}^k \eta_j \sum_{x_j \in \bar{C}_i} (1-t_{ij})^{m_2}$$

最后, 定义类簇  $C_i$  的中心的计算方式如式(13)所示:

$$c_i = \begin{cases} w_l \times C + w_u \times D, & \underline{C}_i \neq \bar{C}_i \\ C, & \text{otherwise} \end{cases} \quad (13)$$

式中

$$C = \frac{1}{|\underline{C}_i|} \sum_{x_j \in \underline{C}_i} x_j$$

$$D = \frac{\sum_{x_j \in (\bar{C}_i - \underline{C}_i)} \{a(u_{ij})^{m_1} + b(t_{ij})^{m_2}\} x_j}{\sum_{x_j \in (\bar{C}_i - \underline{C}_i)} \{a(u_{ij})^{m_1} + b(t_{ij})^{m_2}\}}$$

S-RFPCM 算法的流程简述如下:

- 1) 利用 FCM 算法选择  $k$  个初始聚类中心;
- 2) 利用式(11)计算对象的隶属度矩阵;
- 3) 对每个模糊类簇, 利用阴影集求阈值  $\alpha_i$ , 使式(9)取得最小值;
- 4) 利用式(12)计算目标函数的值;
- 5) 如果目标函数收敛或满足其他终止条件, 结束迭代过程, 否则利用式(13)重新计算每个类簇的新的中心并重复执行步骤 2)~4)。

### 4 实验分析

本节通过试验对 S-RFPCM 算法的性能进行分析, 并和 RFPCM 算法<sup>[12]</sup>的结果进行比较。我们选择 Iris, Wine, Ionosphere 3 个常用数据集作为基准数据集。所有的基准数据集

均从 UCI 的机器学习数据集存储库中下载。

我们将从两个方面对 S-RPFCM 算法的性能进行考察,即聚类的正确率和粗糙率,其定义分别如式(14)、式(15)所示。

粗糙聚类的正确率:

$$\chi = \frac{\sum_{i=1}^k |C_i \cap C_i^*|}{\sum_{i=1}^k |C_i|} \quad (14)$$

式中,  $C_i$  表示第  $i$  个粗糙类簇的下近似,  $C_i^*$  表示实际归属第  $i$  类的数据点的集合。

粗糙聚类的粗糙率:

$$\gamma = 1 - \frac{\sum_{i=1}^k \sum_{x_j \in C_i} u_{ij}}{\sum_{i=1}^k \sum_{x_j \in C_i} \bar{u}_{ij}} \quad (15)$$

式中,  $u_{ij}$  表示对象  $x_j$  隶属于粗糙类簇  $i$  的下近似的隶属度;  $\bar{u}_{ij}$  则表示隶属于上近似的隶属度。

对粗糙聚类算法来讲,其结果的正确率越高越好,而粗糙率越低越好。

为了排除初始设置的选择对聚类结果的影响,每种聚类算法需在基准数据集上重复运行 20 次,最后得到该算法的平均聚类正确性和粗糙率。S-RPFCM 算法和 RPFCM 算法均使用如下初参数:  $m_f=2; m_p=2; a=0.5; b=0.5; (\omega_l, \omega_r) = (0.75, 0.25)$ , 但 S-RPFCM 利用阴影集求取粗糙聚类的阈值,而 RPFCM 算法利用式(16)确定阈值:

$$\delta = \frac{1}{n} \sum_{j=1}^n (u_{ij} - \bar{u}_{ij}) \quad (16)$$

实验结果如表 1、表 2 所列。

表 1 聚类结果的正确率

算法	Iris	Wine	Ionosphere
RPFCM	0.6136	0.6395	0.7337
S-RPFCM	0.8495	0.6500	0.7083

表 2 聚类结果的粗糙率

算法	Iris	Wine	Ionosphere
RPFCM	0.4321	0.4414	0.5528
S-RPFCM	0.2373	0.3610	0.3228

从实验结果可知,在 Iris 和 Wine 数据集上, S-RPFCM 算法得到的聚类结果具有更高的正确率和更小的粗糙率。这表明 S-RPFCM 算法将更多的对象正确地划分到下近似中,其下近似更好地反映了隶属于该类型的对象的真实特征,因此聚类结果更加准确。而在 Ionosphere 数据集上,虽然 RPFCM 算法得到的聚类结果的正确率稍高于 S-RPFCM 算法,但其粗糙率也高于 S-RPFCM 算法。这说明更多的对象被划分到类簇之间的边界,从而导致下近似中仅包含少量对象,难以表示该类型的真实特性。这往往会使聚类模型在未知数据上的泛化能力降低。因此,我们认为 S-RPFCM 算法在 Ionosphere 数据集上的结果也优于 RPFCM 算法。实验表明,引入阴影集可以有效地解决粗糙模糊聚类算法中的阈值选择问题,使 S-RPFCM 算法相对于 RPFCM 具有更好的性能。

**结束语** 本文提出了一种新的基于阴影集的粗糙模糊可能性 C 均值聚类算法 S-RPFCM。S-RPFCM 在聚类过程中结合了粗糙集和模糊集的概念,并通过引入阴影集解决了粗糙聚类中的阈值自动选择问题。实验表明, S-RPFCM 算法相

对于 RPFCM 具有更好的性能。

## 参考文献

- [1] Hartigan J A, Wong M A. Algorithm AS136: A K-Means Clustering Algorithm[J]. Applied Statistics, 1979, 28: 100-108
- [2] Zadeh L A. Fuzzy sets[J]. Inform. Control, 1965, 8: 338-353
- [3] Dunn J C. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters[J]. J. Cybernet, 1974(3): 32-57
- [4] Bezdek J C. Pattern Recognition With Fuzzy Objective Function Algorithm[M]. New York: Plenum, 1981
- [5] Krishnapuram R, Keller J M. A possibilistic approach to clustering[J]. IEEE Trans. Fuzzy Syst., 1993, 1(2): 98-110
- [6] Pawlak Z. Rough sets[J]. International Journal of Computer and Information Sciences, 1982(11): 341-356
- [7] Lingras P, West J. Interval set clustering of web users with rough k-means[J]. Journal of Intelligent Information Systems, 2004, 23(1): 5-16
- [8] Peters G. Outliers in rough k-means clustering[C] // LNCS. 2005, 3776: 702-707
- [9] Peters G. Some refinement of rough k-means clustering[J]. Pattern Recognition, 2006, 39: 1481-1479
- [10] Mitra S, Banka H, Pedrycz W. Rough-fuzzy collaborative clustering[J]. IEEE Trans. Syst., Man, Cybern. B, 2006, 36(4): 795-805
- [11] Maji P, Pal S K. RFCM: A Hybrid Clustering Algorithm Using Rough and Fuzzy Sets[J]. Fundamenta Informaticae, 2007, 80(4): 477-498
- [12] Maji P, Pal S K. Rough Set Based Generalized Fuzzy C-Means Algorithm and Quantitative Indices[J]. IEEE Trans. Systems, Man, and Cybernetics, 2007, 37(6): 1529-1540
- [13] Pedrycz W. Shadowed Sets: Representing and Processing Fuzzy Sets[J]. IEEE Trans. Syst., Man, Cybern. B, 1998, 28(1): 103-109
- [14] Yu F S, Wu F M, Sun Q Y. Improving Fuzzy C-Means with Shadow Set [C] // Proc. 4th Int. Conf. Fuzzy Systems and Knowledge Discovery. 2007: 387-391
- [15] Mitra S, Pedrycz W, Barmanc B. Shadowed c-means: Integrating fuzzy and rough clustering [J]. Pattern Recognition, 2010, 43(4): 1282-1291
- [16] Zhou J, Pedrycz W, Miao D Q. Shadowed sets in the characterization of rough-fuzzy clustering[J]. Pattern Recognition, 2011, 44(8): 1738-1749
- [17] Pal N R, Pal K, Bezdek J C. A mixed c-means clustering model [C] // IEEE Int. Conf. Fuzzy Systems. Spain, 1997: 11-21
- [18] Pal N R, Pal K, Keller J M, et al. Bezdek: A Possibilistic Fuzzy c-Means Clustering Algorithm IEEE Trans [J]. Fuzzy syst., 2005, 13(4): 517-530
- [19] Zhou T, Zhang Y N, Lu H L. Rough k-means Cluster with Adaptive Parameters[C] // Proceedings of the Sixth International Conference on Machine Learning and Cybernetics. Hong Kong, August 2007: 3063-3068
- [20] Mitra S. An evolutionary rough partitive clustering[J]. Pattern Recognition Letters, 2004, 25: 1439-1449
- [21] Małyszko D, Stepaniuk J. Rough Entropy Based K-Means Clustering[C] // Sakai H, Chakraborty M, Hassanien A, et al., eds. Rough Sets, Fuzzy Sets, Data Mining and Granular Computing, LNCS 5908. Springer-Verlag, 2009: 406-413