

# 基于样本加权的可能性模糊聚类算法

刘 兵,夏士雄,周 勇,韩旭东

(中国矿业大学计算机科学与技术学院,江苏徐州 221116)

**摘 要:** 可能性模糊聚类算法解决了噪音敏感和一致性聚类问题,但算法假定每个待分析样本对聚类的贡献相同,导致离群点或噪声点对算法的干扰较强,算法迭代次数过大.为此,提出一种基于样本加权的可能性模糊聚类算法,新算法具有更快的收敛速度,对标准数据集和人工数据集加噪后的测试结果表明,该算法具有更强的鲁棒性,在有效降低时间复杂度的同时能够取得较好的聚类准确率.

**关键词:** 样本加权; 可能性 C-均值聚类; 可能性模糊聚类

**中图分类号:** TP18      **文献标识码:** A      **文章编号:** 0372-2112 (2012) 02-0371-05

**电子学报 URL:** <http://www.ejournal.org.cn>      **DOI:** 10.3969/j.issn.0372-2112.2012.02.026

## A Sample-Weighted Possibilistic Fuzzy Clustering Algorithm

LIU Bing, XIA Shi-xiong, ZHOU Yong, HAN Xu-dong

(School of Computer Science and Technology, China University of Mining and Technology, Xuzhou, Jiangsu 221116, China)

**Abstract:** The possibilistic fuzzy clustering algorithm overcomes the problem of sensitivity to noises and coincident clusters, but it assumes the contribution of each sample is equal, which leads to strong impact from outliers or noises and too many iterations. For this reason, this paper proposes a novel faster possibilistic fuzzy clustering algorithm based on the sample-weighted idea. The results of the experiments on standard data sets and synthetic data sets show that the sample-weighted algorithm is more robust against noises and outliers and reduces the time complexity effectively, and can obtain good clustering accuracy at the same time.

**Key words:** sample-weighted; possibilistic c-means clustering; possibilistic fuzzy clustering

## 1 引言

模糊 C-均值聚类算法(Fuzzy c-means, FCM)已被广泛地应用于模式识别、数据挖掘、图像处理等领域<sup>[1-5]</sup>,但算法对隶属度进行了归一化约束,从而导致噪声敏感问题. Krishnapuram 和 Keller 提出了可能性 C-均值聚类算法(Possibilistic c-means, PCM)改进 FCM 的不足,通过放弃归一化条件,使用典型值表达样本对类的归属程度构造新的目标函数. PCM 对噪声数据具有一定的鲁棒性,但容易产生一致性聚类<sup>[6]</sup>. 结合 FCM 和 PCM 思想提出的可能性模糊 C-均值聚类算法(PFCM, Possibilistic Fuzzy C-means)<sup>[7]</sup>和 IPCM 算法(Improved Possibilistic C-Means, IPCM)<sup>[8]</sup>,虽然解决了一致性聚类和对噪声敏感问题,但受到噪声和离群点数据干扰时,算法收敛速度较慢. 文献[9]中引入了不确定隶属度的概念,但其限制只取 0 和 1 两个数值,噪声对算法的干扰仍然较大. 上述算法均假定每个待分析样本对聚类的贡献相同,迭代

过程中没有引入样本重要性的权值度量,存在对噪声或离群点抗干扰能力不强的缺陷. 文献[10]中提出了一种样本加权的 WFCM 算法(Weighted Fuzzy C-Means, WFCM),该算法根据样本出现的次数计算样本权值,当每个样本出现次数相同时,算法退化为 FCM 算法,因此算法的应用有一定局限性,鲁棒性不强. 文献[11]中提出了一种基于最大熵原则设计的 Sample-weighted FCM 算法,但该算法对样本权值进行了归一化约束,当离群点或者噪声数据较多时,算法性能会受到较大影响.

为了进一步增强模糊聚类算法的鲁棒性,提出了一种基于样本加权的可能性模糊聚类算法,算法利用了可能性模糊聚类的思想解决了噪声敏感和一致性聚类的问题,同时通过对每个样本点计算权值来减小离群点或噪声对对聚类的影响,加快算法收敛速度. 另外,针对 IPCM 算法存在参数依赖的问题,本文使用样本方差优化参数计算,算法无需进行二次迭代,进一步减少了运行时间.

## 2 样本加权可能性模糊聚类算法

### 2.1 样本权值计算方法

在传统的聚类算法中,所有样本都同等地参与到聚类过程中,这样的方法使得噪声样本很容易影响聚类效果.为了削弱离群点样本对聚类效果的影响,文献[12]引入对样本进行加权的概念,通过为离群点赋予很小的权值,使噪声数据参与聚类过程的程度被降低,并将这种思想应用于PCM算法中,加快了收敛速度,但仍未从根本上解决PCM算法一致性聚类的缺陷.

文献[12]给出了两种计算样本权值的方法,一种方法通过计算每个样本点与周围样本点的距离来刻画样本在整个数据集中的位置.其权值的计算公式如下:

$$\varphi_j = \sum_{k=1}^n \exp(-\alpha \|\mathbf{x}_j - \mathbf{x}_k\|^2) \quad (1)$$

其中  $\alpha$  为常数,  $j = 1, 2, \dots, n$ . 这种方法考虑了样本点相互之间的接近程度,缺点是计算量较大.

另一种计算权值的公式如下:

$$\varphi_j = \sum_{k=1}^c \exp(-\alpha \|\mathbf{x}_j - \mathbf{v}_k\|^2) \quad (2)$$

其中  $\alpha$  为常数,  $j = 1, 2, \dots, n$ . 这种方法考虑了样本对各个中心的接近程度,优点是当聚类中心较准确时,权值更合理,可以融入已有算法的迭代过程中.但缺点是对聚类中心敏感.如果聚类中心与真实的中心偏差较大,得到的样本权值就会有偏差.本文选用式(2)作为计算样本权值的方式.

### 2.2 IPCM 中参数的计算方法

文献[13]对参数  $\eta_i$  的计算公式进行改进,使用样本方差计算  $\eta_i$ ,而且不用进行更新,简化了迭代过程.计算公式如下:

$$\eta_i = \frac{\sigma^2}{m^2 c} \quad (3)$$

其中  $m$  为模糊隶属度的权重指数,  $c$  为分类数,  $\sigma^2$  为样本方差,计算如下:

$$\sigma^2 = \frac{1}{n} \sum_{j=1}^n d^2(\mathbf{x}_j, \bar{x}) \quad \text{其中 } \bar{x} = \frac{1}{n} \sum_{j=1}^n \mathbf{x}_j \quad (4)$$

样本方差衡量了数据集的紧凑程度,运用数据集的紧凑程度和分离程度可以使聚类有效性更好.文献[14]也使用了类似的改进方法.

### 2.3 算法描述

针对 IPCM 算法的缺陷,为增强算法鲁棒性,引入样本加权的思想放松对典型值的限制,同时对参数计算的公式进行改进来减少迭代次数.新的目标函数定义如下:

$$J_{m,p}(U, T, V) = \sum_{i=1}^c \sum_{j=1}^n u_{ij}^m t_{ij}^p d_{ij}^2 + \frac{\sigma^2}{m^2 c} \sum_{i=1}^c \sum_{j=1}^n u_{ij}^m (\varphi_j - t_{ij})^p \quad (5)$$

其中,  $d_{ij}^2 = \|\mathbf{x}_j - \mathbf{v}_i\|^2$ ,  $\varphi_j$  为第  $j$  个样本的权值,计算方法如式(2).  $c$  为聚类数目,  $n$  为样本数目,  $u_{ij}$  为样本  $\mathbf{x}_j$  对第  $i$  类的模糊隶属度,  $t_{ij}$  为样本  $\mathbf{x}_j$  对第  $i$  类的典型值,  $m$  为模糊隶属度的指数,  $p$  为典型值的指数,  $\sigma^2$  为样本方差.在约束条件  $\sum_{i=1}^c u_{ik} = 1 \forall k$  下,该目标函数达到最小值的必要条件为:

$$t_{ij} = \frac{\varphi_j}{1 + \left(\frac{m^2 c d_{ij}^2}{\sigma^2}\right)^{\frac{1}{p-1}}}, \forall i, j \quad (6)$$

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{t_{ij}^{p-1} d_{ij}^2}{t_{kj}^{p-1} d_{kj}^2}\right)^{\frac{1}{m-1}}}, \forall i, j \quad (7)$$

$$v_i = \frac{\sum_{j=1}^n u_{ij}^m t_{ij}^p \mathbf{x}_j}{\sum_{j=1}^n u_{ij}^m t_{ij}^p}, \forall i \quad (8)$$

值得注意的是,在根据式(2)计算样本权值时,需要较为真实的聚类中心,这样才能得到较合理的权值.而在有噪声环境下,FCM的计算出的聚类中心偏差较大,而 IPCM 算法对噪声具有一定的鲁棒性.因此,初始聚类中心可以使用简化的 IPCM 算法计算得到.简化的方法为:使用式(3)计算参数,并且无需二次迭代,即先通过简化的 IPCM 进行很少次的循环迭代,得到近似聚类中心,然后使用式(2)进行加权,最后使用样本加权的迭代公式进行计算.基于样本加权的模糊聚类算法描述如下:

**Step 1** 固定  $c, m, p, \alpha$  的值,  $1 < c < n, 1 < m, p < +\infty, \alpha > 0$ ; 设置循环变量初始值  $L = 1$ , 取得近似聚类中心所需迭代次数  $L_2$ , 以及算法总循环数  $L_{\max}$ , 设置算法停止的阈值  $\epsilon$ ;

**Step 2** 利用式(4)计算  $\sigma^2$  的值,使用式(3)计算参数  $\eta_i$ ;

**Step 3** 运行 FCM 算法,并以其结果作为初始聚类中心  $\mathbf{V}^{(0)}$ ;

**Step 4** 利用初始聚类中心  $\mathbf{V}^{(0)}$ , 运行简化 IPCM 算法获取近似聚类中心;

**Step 5** 使用式(2)计算样本的权值;

**Step 6** 使用样本加权迭代公式进行如下循环,逼近局部最优解:

(1)使用式(6)更新典型值  $\mathbf{T}^{(L)}$ ;

(2)使用式(7)更新隶属度值  $\mathbf{U}^{(L)}$ ;

(3)使用式(8)更新  $\mathbf{V}^{(L)}$ ;

(4)  $L$  增加 1;

直到满足条件  $\|\mathbf{V}^{(L)} - \mathbf{V}^{(L-1)}\| < \epsilon$  或  $L > L_{\max}$ .

算法第 4 步的循环次数由  $L_2$  控制,该步骤的目

的是快速得到一个不精确但很接近真实中心的聚类中心,以便下一步计算样本权值,所需循环次数  $L_2$  一般很小.而算法第 6 步的循环由于使用了样本加权,迭代速度会加快,循环会很快结束.因此算法虽仍经过两次循环,但总迭代次数会明显减少,运行时间也较少.

下面证明与 IPCM 算法相比,本文算法收敛速度更快.

**定理 1** 本文提出的基于样本加权的可能性模糊聚类算法与 IPCM 算法相比,在噪声干扰下,收敛速度更快.

**证明** 令第  $k$  次 IPCM 算法迭代时聚类中心记为  $v_i^{(k)}$ ,样本  $x_j$  与  $v_i^{(k)}$  的距离记为  $(d_{ij}^{(k)})^2 = \|x_j - v_i^{(k)}\|^2$ .第  $k+1$  次 IPCM 算法与本文算法典型值分别记为  $t_{ij}^{(k+1)}$  和  $t'_{ij}^{(k+1)}$ ,若  $x_j$  为样本点,则

$$\lim_{x_j \rightarrow v_i} \varphi_j = \lim_{x_j \rightarrow v_i} \sum_{k=1}^c \exp(-\alpha \|x_j - v_i\|^2) = 1,$$

$$t'_{ij}^{(k+1)} \approx t_{ij}^{(k+1)}$$

因此加权系数不会影响正常数据参与迭代.

若  $x_j$  为距离中心较远的噪音或离群点,则

$$\lim_{x_j \rightarrow \infty} \varphi_j = \lim_{x_j \rightarrow \infty} \sum_{k=1}^c \exp(-\alpha \|x_j - v_i\|^2) = 0,$$

$$\therefore t'_{ij}^{(k+1)} = \varphi_j \left( 1 + \left( \frac{(d_{ij}^{(k)})^2}{\eta_i} \right)^{\frac{1}{p-1}} \right)^{-1} \rightarrow 0,$$

$t'_{ij}^{(k+1)} \ll t_{ij}^{(k+1)}$ ,因此与 IPCM 算法相比,本文算法典型值的收敛范围从  $(0, 1)$  区间缩小为  $(0, \varphi_j)$ ,收敛速度大大加快,此外本文算法优化了参数  $\eta_i$  的计算,进一步减少了算法迭代次数.

## 2.4 算法时间复杂度

本文使用  $N$  表示样本总数,  $C$  表示聚类簇数,  $L$  和  $L'$  分别表示本文算法和 IPCM 算法总的迭代次数.在第 2 步计算方差时的时间复杂度为  $O(N)$ ,运行 FCM 的复杂度为  $O(NCL_1)$  ( $L_1$  为 FCM 算法迭代次数),第 4 步循环部分的时间复杂度为  $O(NCL_2)$  ( $L_2$  为算法第 4 步循环的次数,一般很少),第 5 步计算权值的时间复杂度为  $O(NC)$ ,而第 6 步循环的时间复杂度为  $O(NCL_3)$  ( $L_3$  为算法第 6 步的循环次数,由于使用了样本加权,因此迭代次数会很少),因此算法总的时间复杂度为  $O(N) + O(NCL_1) + O(NCL_2) + O(NC) + O(NCL_3) = O(NCL)$ . IPCM 算法的时间复杂度  $O(NCL')$ .虽然本文算法和 IPCM 算法时间复杂度形式相同,但  $L \ll L'$ ,即本文算法迭代次数较少,收敛速度快,因此本文算法实际耗时较少.

## 3 实验结果

实验中首先对标准 IRIS 数据集和人工生成大样本

数据集加入噪声,然后在噪声数据集上分别运行 FCM、PCM、IPCM、PFCM、WFCM 和本文算法,通过对结果进行分析比较,验证了本文算法的鲁棒性和有效性.

### 3.1 对加噪 IRIS 数据集的测试

我们对 IRIS 数据集增加 40 个随机人工噪声数据,噪声数据的每一维特征都是在  $0.0 \sim 10.0$  之间的随机数.实验条件为:  $\epsilon = 0.00001$ ,最大循环数  $L_{\max} = 200$ ,  $m = 2$ ,  $p = 2$ ,  $a = 1$ ,  $b = 1$ ,  $c = 3$ ,  $\alpha = 1.0/0.4$ ,使用计算机配置为:英特尔酷睿 2 双核 CPU,主频 2.93GHz,内存 2.00GB,利用 Microsoft Visual C++ 6.0 进行仿真实验,结果如表 1 所示.

由表 1 中的数据可知,在噪声干扰下,PCM 依然容易产生一致性聚类,FCM、WFCM 和 PFCM 算法错分数和中心偏差较大,这反映了算法对噪声的鲁棒性较差.只有 IPCM 和本文算法能够同时具有好的错分数和较为准确的聚类中心,说明这两种算法鲁棒性较强.并且在绝大多数情况下,本文算法比 IPCM 算法的运行时间少,并且错分数和中心偏差也较小,因而性能更加良好.

表 1 各算法在加噪 IRIS 数据集上的运行结果

聚类算法	循环次数	总时间(秒)	错分数	中心偏差
FCM	33	0.016	(50)	1.54
PCM	20	0.031	49	1.40
PFCM	22	0.047	36(11)	0.44
IPCM	98	0.135	12(45)	0.11
WFCM	33	0.016	(50)	1.54
本文算法 ( $L_2 = 10$ )	21	0.042	15(15)	0.22
本文算法 ( $L_2 = 30$ )	46	0.070	10(10)	0.06
本文算法 ( $L_2 = 50$ )	62	0.087	11(11)	0.08
本文算法 ( $L_2 = 70$ )	82	0.118	11(11)	0.08
本文算法 ( $L_2 = 90$ )	102	0.134	11(11)	0.08
本文算法 ( $L_2 = 120$ )	102	0.134	11(11)	0.08

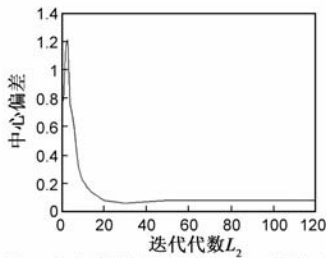
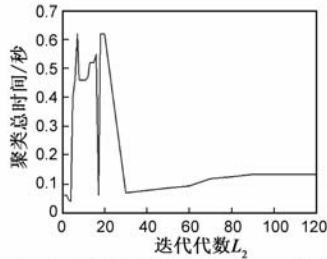
此外,本文算法中迭代次数  $L_2$  的数值会影响近似聚类中心,进而影响样本权值、最终的聚类中心和算法总时间.

图 1 和图 2 分别给出了表 1 中心偏差和算法总时间随  $L_2$  变化的拟合曲线.从图中可以得出如下结论:  $L_2$  越大,近似聚类中心和样本权值计算越准确,最终聚类中心越精确,

中心偏差会越小.但是从图 1 中可以看出,当  $L_2$  达到一定迭代次数后,中心偏差和错分数基本保持不变,但聚类总时间会随着  $L_2$  的增大而增加,  $L_2$  的合理取值应在  $30 \sim 50$  之间.

### 3.2 对加噪大样本人工数据集的测试

为了测试本文算法在较大样本数据集上的运行效果,本文对文献[8]中含有三个类别的正态分布数据集(为方便起见,称为  $X_{2000}$  数据集)进行测试,三个类别的

图1 中心偏差与迭代次数 $L_2$ 的关系图2 聚类总时间与迭代次数 $L_2$ 的关系

正态分布  $N(\mu, \Sigma)$  分别为:

- 类 1:  $N([1.0 \ 0.0], \begin{bmatrix} 0.4 & 0.0 \\ 0.0 & 0.8 \end{bmatrix})$ , 共 500 个点;
- 类 2:  $N([3.0 \ 0.0], \begin{bmatrix} 0.6 & 0.0 \\ 0.0 & 1.2 \end{bmatrix})$ , 共 1000 个点;
- 类 3:  $N([5.0 \ 0.0], \begin{bmatrix} 0.4 & 0.0 \\ 0.0 & 0.8 \end{bmatrix})$ , 共 500 个点.

该数据集为二维数据集, 样本点共 2000 个, 聚类数为 3. 为检验各算法在大样本情况下对噪声数据的鲁棒性, 对  $X_{2000}$  数据集  $[1, 6] \times [-4, 4]$  区域内加入随机均匀的 2000 个噪声点, 如图 3 所示. 各个算法在加噪的  $X_{2000}$  数据集上的聚类结果如表 2 所示, 实验条件为  $\alpha = 1.0/0.4$ , 其余同上.

表 2 的实验结果表明, 在加噪环境下, FCM、WFCM 的性能大大降低, 仅优于出现一致性聚类的 PCM. 图 3 中分别标记了 FCM 和本文算法计算出的聚类中心. PFCM 算法错分数和中心偏差也较大, 只有 IPCM 算法和本文算法能够得到较为精确的聚类结果, 即较少的错分

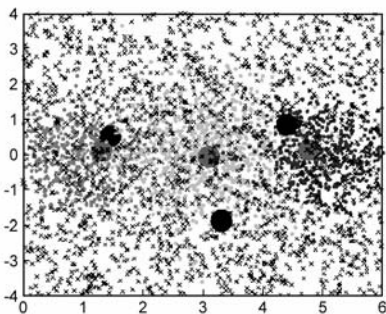


图3 图中黑色实心圆为FCM算法得到的聚类中心,分别为(1.46, 0.52)(3.31, -1.86)(4.40, 0.86),三个灰色实心圆为本文算法在 $L=50$ 时得到的聚类中心,分别为(1.31, 0.10),(3.10,-0.07),(4.77,0.13)

数和较小的中心偏差. 在错分数和中心偏差相当的情况下, 本文算法的运行时间比 IPCM 算法要少得多.

表 2 各算法在加噪的  $X_{2000}$  数据集上的运行结果

聚类算法	循环次数	总时间(秒)	错分数	中心偏差
FCM	359	3.27	(768)	4.25
PCM	38	4.01	1000	8.00
PFCM	172	5.51	523(584)	1.99
IPCM	225	8.25	303(377)	0.14
WFCM	359	4.09	(769)	4.32
本文算法 ( $L_2 = 10$ )	24	4.26	551(551)	1.15
本文算法 ( $L_2 = 30$ )	48	4.61	375(375)	0.34
本文算法 ( $L_2 = 50$ )	65	4.70	321(321)	0.19
本文算法 ( $L_2 = 70$ )	85	5.26	316(316)	0.169
本文算法 ( $L_2 = 90$ )	105	5.69	314(314)	0.165
本文算法 ( $L_2 = 100$ )	115	5.81	314(314)	0.165
本文算法 ( $L_2 = 120$ )	135	6.06	315(315)	0.164

### 3.3 样本权值计算中参数 $\alpha$ 的选择

权值计算式(2)中使用高斯核函数计算样本点与各聚类中心的距离之和, 即使用高斯核函数刻画样本点与各聚类中心的相似度, 相似度越大, 样本权值越大, 反之越小. 因此确定参数  $\alpha$  的问题转化为对高斯核的选取, 根据核矩阵选取问题和谱聚类参数选取问题的等价性<sup>[15]</sup>, 本文利用文献[16]中 Self-Tuning 谱聚类算法选择尺度参数  $\alpha$  的方法, 该算法利用“Local Scaling”局部尺度思想自适应统计每个样本点局部近邻, 进而为每一个样本点确定一个  $\sigma_i$ , 样本点之间的相似度定义为:

$$A_{ij} = \exp\left(\frac{-\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\sigma_i \sigma_j}\right) \quad (9)$$

其中,  $\sigma_i = d(s_i, s_k)$ ,  $s_k$  是样本点  $s_i$  的第  $K$  个近邻. 文献[16]指出, 实验表明一般取  $K$  为 7, 算法效果最佳, 并且在高维数据上实验表现突出.

为验证局部尺度方法的有效性, 对加噪  $X_{2000}$  数据集进行了测试, 实验结果如表 3 所示. 通过对比, 与表 2 取最佳参数  $\alpha = 1.0/0.4$  时的聚类结果基本一致. 这说明在加噪环境下, 使用式(9)自适应确定参数  $\alpha$  来计算样本点与中心点的相似度, 进而计算出样本点权值参与迭代的方法是有效的.

表 3 加噪的  $X_{2000}$  数据集利用“Local Scaling”方法自适应选择参数  $\alpha$  的聚类结果

聚类算法	循环次数	总时间(秒)	错分数	中心偏差
本文算法 ( $L_2 = 10$ )	34	6.82	565(565)	1.2512
本文算法 ( $L_2 = 30$ )	60	7.64	379(379)	0.3688
本文算法 ( $L_2 = 50$ )	82	8.55	332(332)	0.2428
本文算法 ( $L_2 = 70$ )	102	9.38	326(326)	0.1794
本文算法 ( $L_2 = 90$ )	122	10.16	319(321)	0.1742
本文算法 ( $L_2 = 100$ )	132	10.53	319(319)	0.1734

## 4 结论

针对 FCM、PFCM 算法对噪声敏感,PCM 算法易产生一致性聚类和 IPCM 算法运行时间过长等不足,提出了基于样本加权的可能性模糊聚类算法.与其它算法相比,引入样本加权之后的算法对噪声鲁棒性强,收敛速度快,并且具有较少的错分数和较准确的聚类中心.本文算法可以结合特征加权和半监督聚类的思想进一步提高聚类的准确率.

### 参考文献

- [1] 孙吉贵,刘杰,赵连宇.聚类算法研究[J].软件学报,2008,19(1):48-61.  
Sun Ji-gui, Liu Jie, Zhao Lian-yu. Clustering algorithms research[J]. Journal of Software, 2008, 19(1):48-61. (in Chinese)
- [2] S Kirindis, V Chatzis. A robust fuzzy local information c-means clustering algorithm[J]. IEEE Trans Image Process, 2010, 19(5):1328-1337.
- [3] W Cai, S Chen, D Zhang. Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation[J]. Pattern Recognition, 2007, 40(3):825-838.
- [4] 张敏,于剑.基于划分的模糊聚类算法[J].软件学报,2004,15(6):858-868.  
Zhang Min, Yu Jian. Fuzzy partional clustering algorithms[J]. Journal of Software, 2004, 15(6):858-869. (in Chinese)
- [5] 田军委,黄永宣,于亚琳.基于熵约束的快速 FCM 聚类多阈值图像分割算法[J].模式识别与人工智能,2008,21(2):221-226.  
Tian Jun-wei, Huang Yong-xuan, Yu Ya-lin. A fast FCM cluster multi-threshold image segmentation algorithm based on entropy constraint[J]. Pattern Recognition & Artificial Intelligence, 2008, 21(2):221-226. (in Chinese)
- [6] R Krishnapuram, J Keller. A possibilistic approach to clustering[J]. IEEE Trans Fuzzy Systems, 1993, 1(2):98-110.
- [7] N R Pal, K Pal, J C Bezdek. A possibilistic fuzzy c-means clustering algorithm[J]. IEEE Trans Fuzzy Systems, 2005, 13(4):517-530.
- [8] J S Zhang, Y W Leung. Improved possibilistic c-means clustering algorithms[J]. IEEE Trans Fuzzy Systems, 2004, 2(12):209-217.
- [9] 陈健美,陆虎等.一种隶属关系不确定的可能性模糊聚类方法[J].计算机研究与发展,2008,45(9):1486-1492.  
Chen Jian-mei, Lu Hu, et al. A possibility fuzzy clustering algo-

rithm based on the uncertainty membership[J]. Journal of Computer Research and Development, 2008, 45(9):1486-1492. (in Chinese)

- [10] 高新波,李洁.基于加权 FCM 与统计检验指导的多阈值图像自动分割算法[J].电子学报,2004,32(4):661-664.  
Gao Xin-bo, Li Jie. A multi-threshold image segmentation algorithm based on weighting fuzzy c-means clustering and statistical test. Acta Electronica Sinica, 2004, 32(4):661-664. (in Chinese)
- [11] Jian Yu, Miin-Shen Yang, E. Stanley Lee: Sample-weighted clustering methods. Computers & Mathematics with Applications, 2011, 62(5):2200-2208.
- [12] A Schneider. Weighted possibilistic c-means clustering algorithms[J]. IEEE Trans Fuzzy Systems, 2000, 1:176-180.
- [13] M S Yang, K L Wu. Unsupervised possibilistic clustering[J]. Pattern Recognition, 2006, 39(1):5-21.
- [14] 武小红,周建江.可能性模糊 C-均值聚类新算法[J].电子学报,2008,36(10):1996-2000.  
Wu Xiao-hong, Zhou Jian-jiang. A novel possibilistic fuzzy c-means clustering[J]. Acta Electronica Sinica, 2008, 36(10):1996-2000. (in Chinese)
- [15] M Filipponea, F Camastrab, F Masullia, et al. A survey of kernel and spectral methods for clustering[J]. Pattern Recognition, 2008, 41(1):176-190.
- [16] Zelnik Manor L, Perona P. Self-Tuning spectral clustering [A]. Advances in Neural Information Processing Systems (NIPS)[C]. Cambridge, MA: MIT Press, 2004.

### 作者简介



刘 兵 男,1981 年生于河南永城.博士研究生,讲师,研究方向为机器学习、数据挖掘、人工智能.

E-mail: liubing@cumt.edu.cn



夏士雄 男,1961 年生于辽宁抚顺.博士生导师,教授,研究方向为无线传感器网络、智能信息处理.