

## 一种改进隶属度函数的FCM聚类算法

肖满生<sup>a</sup>, 文志诚<sup>b</sup>, 张居武<sup>a</sup>, 汪新凡<sup>b</sup>

(湖南工业大学 a. 科技学院, b. 计算机与通信学院, 湖南 株洲 412008)

**摘要:** 传统模糊  $C$ -均值 (FCM) 算法要求一个样本对于各个聚类的隶属度之和满足归一化条件, 从而导致算法对噪声和孤立点敏感, 对非均衡分布样本的聚类有效性降低. 针对该问题, 提出一种改进模糊隶属度函数约束的 FCM 聚类算法, 通过放松归一化条件, 推导出新的隶属度划分公式, 并在聚类过程中不断进行隶属度修正, 从而达到消除噪声样本、提高聚类有效性的目的. 最后通过实验结果对比验证了改进算法的正确性.

**关键词:** 模糊  $C$ -均值; 隶属度约束; 噪声样本; 有效性

**中图分类号:** TP391

**文献标志码:** A

## An FCM clustering algorithm with improved membership function

XIAO Man-sheng<sup>a</sup>, WEN Zhi-cheng<sup>b</sup>, ZHANG Ju-wu<sup>a</sup>, WAN Xin-fan<sup>b</sup>

(a. College of Science and Technology, b. College of Computer and Communication, Hu'nan University of Technology, Zhuzhou 412008, China. Correspondent: WEN Zhi-cheng, E-mail: zcwen@mail.shu.edu.cn)

**Abstract:** Since the general fuzzy  $C$ -means (FCM) algorithm requires sum of membership satisfying the normalization condition for a sample to each cluster, and thus results algorithm sensitive to noise or outliers and reducing the validity of the clustering on non-equilibrium distribution samples. Therefore, an FCM clustering algorithm with the improved fuzzy membership constraint function is proposed. By relaxing the normalization condition, a new formula of membership division is deduced, and the membership is constantly corrected in the clustering process, so that it will eliminate the noise sample, and improve the validity of clustering. Finally, the comparison of the experimental result verifies the correctness of the improved algorithm.

**Keywords:** fuzzy  $C$ -means; membership constraint; noise sample; validity

### 0 引言

模糊  $C$ -均值 (FCM) 聚类算法是无监督模式识别中的一种重要方法. 该方法通过最小化其目标函数获得样本集的模糊划分或聚类, 已在信息技术和控制决策领域得到了广泛的应用, 如图像处理<sup>[1]</sup>和模式识别<sup>[2]</sup>等. 然而, 传统的 FCM 算法在最小化目标函数的过程中, 要求一个样本对于各个聚类的隶属度之和为 1, 即满足归一化条件  $\sum_{i=1}^c u_{ik} = 1, \forall k$ . 这样一来, 在样本分布不均衡或存在噪声情况下, 会导致不理想的聚类结果. 例如, 当某个噪声样本远离各个聚类中心时, 本来它严格属于每个类的隶属度都很小, 但由于归一化条件的制约, 将使它对各类都具有较大的隶属度, 从而影响最终的迭代结果. 为此, 必须改进传统 FCM 算法的模糊隶属度函数, 以增强聚类结果的有效性.

对于模糊隶属度函数的改进问题, 目前主要集中在如何对其施加某种约束或权重, 如 Hoppner 等<sup>[3]</sup>提出了一种改进模糊划分的聚类算法 IFP-FCM, 通过引入新的隶属度约束函数, 得到了意义更趋明晰的隶属度迭代公式, 但其模糊加权指数  $m$  只能恒为 2; 朱琳等<sup>[4]</sup>在 Hoppner 的基础上, 基于惩罚对手的竞争学习 (RPCL) 思想, 引入隶属度约束函数的惩罚项, 解决了 IFP-FCM 中模糊加权指数的一般化问题, 提高了模糊划分的鲁棒性; Li 等<sup>[5]</sup>提出了一种特征加权算法约束隶属度函数, 并在图像分割中取得了成功的应用. 此外, 贺思艳<sup>[6]</sup>和魏延等<sup>[7]</sup>分别对模糊隶属度进行设计改进, 并通过实验对其正确性进行了验证. 但是, 上述方法都没有跳出隶属度归一化约束条件, 因而对于样本分布不均衡或含有大量噪声的样本集聚类效果不理想, 同时, 如文献 [6-7] 所提出的模糊隶属度函数

**收稿日期:** 2014-11-10; **修回日期:** 2015-01-12.

**基金项目:** 湖南省自然科学基金项目(2015JJ2047, 13JJ9031); 湖南工业大学自然科学基金项目(2014HZX29); 湖南省教育厅项目(12C0074).

**作者简介:** 肖满生(1968—), 男, 教授, 从事智能计算和智能信息处理等研究; 文志诚(1972—), 男, 副教授, 博士, 从事网络安全的研究.

的确定方法,其运算过于复杂,效率不高.而对于样本非均衡分布的数据集,肖满生等<sup>[8]</sup>提出了一种基于原型初始化的样本隶属函数分配方案,通过设计一个高斯形式的隶属度函数,采用指数距离测度方法表达样本隶属度,从而打破了隶属度之和为 1 的归一化约束,有效地解决了非均衡分布样本的聚类问题,但该方法仅适用于二维分布的样本集聚类.对于高维样本,其可行性和有效性尚有待验证.

针对上述问题,本文在国内外研究的基础上,受文献[9-11]的启发,提出一种带有修正值的 FCM 聚类隶属度函数确定方法.该方法通过放松隶属度约束,既能够解决样本分布非均衡以及含有大量噪声样本的聚类问题,同时又可以避免孤立点自成一类的情况,提高了聚类的有效性.

## 1 FCM 算法隶属度函数的改进

### 1.1 放松归一化条件的模糊隶属度

设某一样本集  $X = \{x_1, x_2, \dots, x_n\}$ , 类别数为  $c$ ,  $m(m > 1)$  为模糊加权指数, 又称为平滑因子, 表示样本在模糊类间的分享程度, 当  $m = 1$  表示硬聚类,  $u_{ik}$  为第  $k$  ( $k = 1, 2, \dots, n$ ) 个样本  $x_k$  在第  $i$  ( $i = 1, 2, \dots, c$ ) 类中的模糊隶属度,  $d_{ik}$  为样本  $x_k$  与聚类中心  $v_i$  的欧几里德距离,  $U$  为模糊隶属度集合,  $V$  为聚类中心集合时, FCM 目标函数为

$$J(U, V) = \sum_{i=1}^c \sum_{k=1}^n (u_{ik})^m d_{ik}^2. \quad (1)$$

在最小化该目标函数的过程中, 要求满足任一样本对于各个聚类的隶属度之和为 1 的归一化约束条件, 即

$$\sum_{i=1}^c u_{ik} = 1, \forall k. \quad (2)$$

在此条件约束下, 当某个噪声样本远离各类的聚类中心时, 本来它严格属于各类的隶属度都很小, 但由于归一化约束, 将使它对各类都具有较大的隶属度, 从而影响最终结果. 为此, 人们提出放松归一化约束条件, 只要满足所有样本对于各类的隶属度之和为样本数  $n$  即可, 即

$$\sum_{i=1}^c \sum_{k=1}^n u_{ik} = n, \forall i, k. \quad (3)$$

根据式(3), 采用拉格朗日乘数法最小化目标函数(1), 可求得  $u_{ik}, v_i$ .

由式(1)和(3), 可得

$$J(U, V) = \sum_{i=1}^c \sum_{k=1}^n (u_{ik})^m d_{ik}^2 + \lambda \left( \sum_{i=1}^c \sum_{k=1}^n u_{ik} - n \right), \quad (4)$$

其中  $\lambda$  为任意常数. 最小化  $J(U, V)$  时, 必要条件为对

$u_{lh}$  的一阶导数满足

$$\frac{\partial J(U, V)}{\partial u_{lh}} = m(u_{lh})^{m-1} \cdot d_{lh}^2 - \lambda = 0. \quad (5)$$

此处用  $l = 1, 2, \dots, c$  和  $h = 1, 2, \dots, n$  代替  $i, k$ , 目的是不与  $i, k$  重复. 按式(5)求解  $u_{lh}$ , 得

$$u_{lh} = \left[ \frac{\lambda}{m(d_{lh}^2)} \right]^{1/(m-1)}. \quad (6)$$

将式(6)代入(3)并求解, 即

$$\sum_{k=1}^n \sum_{i=1}^c \left( \frac{\lambda}{m} \right)^{1/(m-1)} \left[ \frac{1}{(d_{ik}^2)} \right]^{1/(m-1)} = n,$$

因而得

$$\left( \frac{\lambda}{m} \right)^{1/(m-1)} = n / \left( \sum_{i=1}^c \sum_{k=1}^n \left[ \frac{1}{d_{ik}^2} \right]^{1/(m-1)} \right)^{-1}. \quad (7)$$

将式(7)代入(6), 得

$$u_{lh} = n / \sum_{i=1}^c \sum_{j=1}^n \left[ \left( \frac{d_{lh}}{d_{ik}} \right)^{2/(m-1)} \right], \quad (8)$$

通常写成

$$u_{ik} = n / \sum_{l=1}^c \sum_{h=1}^n \left[ \left( \frac{d_{ik}}{d_{lh}} \right)^{2/(m-1)} \right]. \quad (9)$$

同理, 将式(1)对  $v_i$  求导, 可得 FCM 的聚类中心

$$\frac{\partial J(U, V)}{\partial v_i} = \frac{\partial}{\partial v_i} \left[ \sum_{k=1}^n \sum_{i=1}^c (u_{ik})^m d_{ik}^2 \right] = 0,$$

得

$$v_i = \left( \sum_{k=1}^n (u_{ik})^m \cdot x_k \right) / \sum_{k=1}^n u_{ik}^m. \quad (10)$$

式(9)和(10)即为放松归一化约束后, FCM 算法的模糊隶属度及聚类中心的迭代公式. 显然, 放松归一化约束后, 聚类中心  $v_i$  的迭代公式没有改变, 但其隶属度已不是通常意义上的隶属度迭代公式, 所求得的值可能会大于 1.

### 1.2 隶属度迭代公式的修正

由上述可知, 当放松隶属度约束归一化条件后, 得到样本的隶属度可能会大于 1, 从而使各样本对类的隶属度相差悬殊, 即某些样本可能对各个类的隶属度很大, 另一些样本可能对类的隶属度很小. 如果在聚类过程中某个样本对某类的隶属度很大, 则最后该类可能只包含这一个样本, 即孤立点(噪声)单独成一类, 这显然不是所希望的结果; 另外, 如果某些样本的隶属度很小, 则在实际聚类过程中迭代终止阈值的选取会非常困难. 因此, 在得到式(9)所示的隶属度函数后, 本文进一步对其进行非线性修正, 使所得到的隶属度不宜太大或太小.

设  $l_{ik}$  表示样本  $x_k$  到聚类中心  $v_i$  的距离, 则修正后的样本隶属度可表示为

$$u'_{ik} = u_{ik} + \left( 1 - \sum_{i=1}^c u_{ik} \right) \cdot \left( 1 - \frac{l_{ik}}{\sum_{i=1}^c l_{ik}} \right)^2. \quad (11)$$

分析如下:

1) 当  $\sum_{i=1}^c u_{ik} < 1$  时, 因其总和小于 1, 样本  $x_k$  相对于各聚类的隶属度都很小, 修正后的隶属度  $u'_{ik}$  增加, 样本越远离类中心, 即  $l_{ik}$  越大, 修正后的  $u'_{ik}$  增加得越少; 反之, 样本离类中心越近,  $u'_{ik}$  增加得越多. 另外, 因为  $(1 - l_{ik} / \sum_{i=1}^c l_{ik})^2 < 1$ , 得  $(1 - \sum_{i=1}^c u_{ik}) \cdot (1 - l_{ik} / \sum_{i=1}^c l_{ik})^2 < 1 - \sum_{i=1}^c u_{ik} < 1 - u_{ik}$ , 故  $u_{ik} + (1 - \sum_{i=1}^c u_{ik}) \cdot (1 - l_{ik} / \sum_{i=1}^c l_{ik})^2 < 1$ , 即  $u'_{ik} < 1$ . 所以修正后的  $u'_{ik}$  大于 0 而小于 1.

2) 当  $\sum_{i=1}^c u_{ik} > 1$  时,  $x_k$  相对于各聚类的隶属度中, 可能有个别隶属度过大, 由于这时  $(1 - \sum_{i=1}^c u_{ik})$  为负值, 修正后的隶属度  $u'_{ik}$  减小, 样本离聚类中心的距离  $l_{ik}$  越近, 修正后的  $u'_{ik}$  减少得越多; 反之,  $u'_{ik}$  减少得越少. 另外, 仿照分析式(1), 同样可以推得修正后的  $u'_{ik}$  值在 0 与 1 之间.

3) 当  $\sum_{i=1}^c u_{ik} = 1$  时, 无需修正, 式(9)中的  $u_{ik}$  等价于传统 FCM 算法中的隶属度公式.

经式(11)修正后, 使得在迭代过程中  $u_{ik}$  不至于太大或太小, 从而能够有效地避免孤立点独成一类的情况出现.

## 2 改进隶属度函数的 FCM 算法在样本聚类中的实现

一个噪声样本, 由于它不属于任何类, 它对各类的隶属度都很小, 满足  $\sum_{i=1}^c u_{ik} < 1$ , 其大小应在该样本对各类隶属度均值的附近, 即

$$u_{ik} \approx \sum_{i=1}^c u_{ik} / c, \quad (12)$$

其中  $c$  为聚类数.

由式(12)可以检测出样本集中的噪声点.

基于上述分析, 改进隶属度函数的 FCM 算法在聚类中的实现过程如下.

**Step 1:** 对于给定样本集  $X = \{x_1, x_2, \dots, x_n\}$ , 初始化聚类数  $c$  ( $2 \leq c \leq n$ ), 隶属度收敛阈值  $\varepsilon$ , 迭代次数  $t = 1$ 、最大迭代次数  $T$  和初始聚类中心集  $V_i$ . 另外, 根据 Bezdek 及 Dunn 所提出的模糊集理论, 在一般情况下, 模糊加权指数值  $m = 2$ .

**Step 2:** 根据式(9)计算样本的隶属度值  $u_{ik}^{(t)}$ .

**Step 3:** 根据式(11)对样本隶属度值  $u_{ik}^{(t)}$  进行修正, 得  $u'_{ik}^{(t)}$ .

**Step 4:** 将修正后的  $u'_{ik}^{(t)}$  代入式(10), 计算各类别的聚类中心  $v_i^{(t+1)}$ .

**Step 5:** 如果  $t > T$  或者  $|u'_{ik}^{(t+1)} - u'_{ik}^{(t)}| < \varepsilon$ , 则算法停止, 否则  $t = t + 1$ , 跳转到 Step 2.

**Step 6:** 根据式(12), 除去样本集中的噪声点, 最后得到各类别的聚类中心  $v_i$  和样本隶属度  $u_{ik}$ .

## 3 实验结果分析

为了验证本文所提出方法的正确性, 分别采用含有噪声点非均衡分布的人造样本数据集、UCI 数据库中的真实数据集以及由纹理库中的纹理合成图像数据集进行实验. 实验环境如下: PC Intel® Pentium® CPU G3220@3.00 GHz 3.00 GHz, RAM 4 GB, OS: Windows 7, Program: Matlab 7.0. 同时, 为便于分析, 在实验过程中采用传统 FCM 算法以及文献[4]中提出的 GIFP-FCM 方法进行对比实验. GIFP-FCM 方法, 即改进模糊划分的 FCM 聚类一般化方法, 其理论推导完善成熟, 且应用较为普遍, 是目前最有说服力的方法之一, 为此将其作为对比算法. 为描述方便, 本文提出的改进隶属度函数的 FCM 算法简称为 IM-FCM, 传统 FCM 算法简称为 G-FCM. 实验中主要评价指标包括: 1) 成功分类率 SCR, 即正确分类的样本数占样本集中样本总数的百分比; 2) 类内样本的互信息 MIC, 有关样本的互信息请参阅文献[12]等, 此处不再细述; 3) 最小化目标函数值 OFV. 在实验 1 中还引入了划分均匀性测度指标 UM, 即划分类内部均匀性测度, 用来评价聚类内部样本分布的均匀程度, 其定义为

$$UM = \frac{1}{n} \sum_{i=1}^c \left\{ \sum_{k \in R_i} \left[ x_k - \frac{1}{A_i} \sum_{k \in R_i} x_k \right]^2 \right\}. \quad (13)$$

其中:  $R_i$  表示第  $i$  个聚类;  $A_i$  为该聚类中样本个数;  $n$  为归一化系数, 在本文可设为样本总数. 由式(13)可以分析, 聚类内部样本分布越均衡, 即样本  $x_k$  与类内其他样本相差越小, UM 值越小, 因此, UM 反映了各聚类内部样本分布均衡性程度.

### 实验 1 人造数据实验.

一个干净无噪声的人造样本数据集如图 1(a) 所示(其中 3 个黑粗点为 IM-FCM 算法实验结果中的聚类中心, 原始数据集中并不存在), 由 1 大 2 小 3 类样本子集构成, 其主要参数如表 1 所示. 以此为基础, 在其中不断增加噪声点来进行实验. 每次增加 30 个, 图 1(b) 为增加 30 个噪声点后的样本集, 最后一次噪声点达 300 个. 包括原来干净样本集在内共进行 11 次针对 UM 值的测定实验, 分别采用 3 种算法(即传统的 G-FCM 算法、改进的 GIFP-FCM 算法以及本文提出的 IM-FCM 算法)进行聚类测定 UM 值时, 每种算法各进行 50 次聚类取平均值, 其中迭代终止阈值(两

迭代隶属度差)  $\varepsilon = 0.01$ , GIFP-FCM 算法中  $\alpha$  按常规取 0.9, 实验结果中 UM 比较曲线如图 2 所示 (其中横轴为增加噪声实验次数序号, 纵轴为 UM 值). 同时, 针对具有 30 个噪声点的人造样本集分别采用上述 3 种算法对 SCR、MIC 以及 OFV 共 3 个指标进行实验测定, 3 种算法的实验参数同 UM 值测定实验, 实验结果如表 2 所示. 另外, 为了更直观地表达聚类结果, 本文所提出的 IM-FCM 算法对含有 30 个噪声点的样本集的实验结果, 其直观图基本上接近图 1(a), 限于篇幅, 其效果直观图不再列出, 其中 3 个黑粗点为实验结果中 3 个聚类中心.

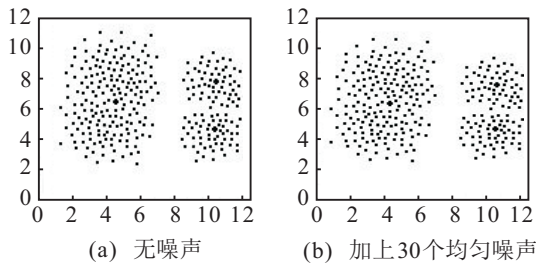


图 1 人造样本数据集

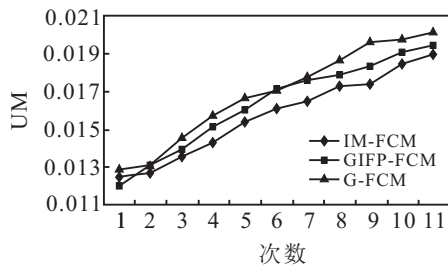


图 2 3 种算法的 UM 值比较曲线

表 1 无噪声人造样本集的主要参数

样本子集	均值坐标	协方差矩阵	样本数
第 1 类	(4.5, 6.2)	18.7 1.5	200
		1.5 19.6	
第 2 类	(10.9, 4.6)	2.3 0.3	50
		-0.3 1.8	
第 3 类	(10.8, 7.6)	2.3 -0.5	50
		-0.5 2.5	

表 2 3 种算法对人造样本集的实验结果

	GIFP-FCM	G-FCM	IM-FCM
SCR/%	77.31	76.42	94.31
MIC	0.8325	0.7831	0.9142
OFV	1364.51	1324.83	1049.53

从图 2 中可以看出, 随着噪声的增加, 3 种算法的 UM 值也不断增加, 表明聚类效果变差, 但对于总体情况, 在噪声样本超过 20% (图 2 第 3 次实验起) 后, 本文所提出的 IM-FCM 算法的 UM 值比其他两种算法所得到的 UM 值明显减小, 表明本文所提出算法的聚类效果好. 另外, 从表 2 中所得到的实验结果看, 本文所提出的 IM-FCM 算法的成功分类率 SCR 以及类内样本的互信息 MIC 值均比其他两种方法所测得的

值大, 而且最小化目标函数比其他两种算法小, 表明本文的 IM-FCM 算法对于含有噪声点样本聚类, 其有效性高.

**实验 2 真实的 UCI 库中 IRIS 数据集实验.**

本次实验采用来自 UCI 样本库<sup>[13]</sup>的 IRIS 数据集进行实验. IRIS 是一个著名的标准测试样本集, 常用来检测聚类效果, 由 150 个样本组成, 分为 3 类 (setosa、versicolor、virginica), 每类 50 个样本, 其中一类完全分离, 另两类有交叉, 因此, 这也是一类非均衡分布的样本, 每一样本包含 petal length、petal width、sepal length、sepal width 共 4 个分量. 3 种方法的实验条件及要求同实验 1, 即每种方法各进行 50 次聚类取平均值, 每次聚类终止条件为两次迭代间隶属度差  $\varepsilon = 0.01$ , GIFP-FCM 算法中  $\alpha$  取通用值 0.9. 表 3 为 3 种算法进行聚类实验的 SCR、MIC、OFV 值对比.

表 3 3 种算法对 IRIS 数据集的实验结果

	GIFP-FCM	G-FCM	IM-FCM
SCR/%	82.31	78.10	86.70
MIC	0.9360	0.7238	0.9480
OFV	158.72	165.61	204.69

从表 3 中可以看出: 对于成功分类率 SCR 以及类内互信息 MIC, 本文提出的算法所获得的结果较好, GIFP-FCM 次之, G-FCM 最差; 而对于最小化目标函数 OFV 值, GIFP-FCM 最优, G-FCM 次之, 本文提出的 IM-FCM 最差. 这是因为 IM-FCM 放松了对隶属度约束条件, 使得  $u_{ik}$  可能大于 1, 在没有噪声样本干扰的情况下, 其目标函数值会随隶属度  $u_{ik}$  的增大而增大, 因而出现了上述情况. 但综合评价 3 种算法, 本文提出的 IM-FCM 在聚类的有效性、正确性方面取得了较优的结果.

**实验 3 纹理合成图像实验.**

实验图像见图 3, 由 5 维纹理合成, 这些纹理来自 Brodatz 纹理库<sup>[14]</sup>中的 D10、D27、D68、D3 和 D56. 该图像大小为  $512 \times 512$ , 灰度为 256 级. 该实验中, 组成纹理图像样本的分布密度大, 样本间的差异较小, 在利用 Gabor 滤波器对该图像进行特征提取后, 3 种算法对图像分割实验所得的指标数据对比如表 4 所示.

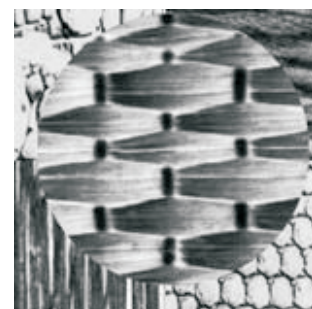


图 3 纹理合成图像

表4 3种算法对纹理合成图像的实验结果

	GIFP-FCM	G-FCM	IM-FCM
SCR/%	91.6	91.1	91.7
MIC	0.9334	0.9240	0.9230
OFV	$1.36 \times 10^5$	$1.89 \times 10^5$	$1.34 \times 101^5$

从表4可看出,3种算法所测得的SCR、MIC两指标值差别不大,都具有较好的聚类效果,仅在OFV指标值上,G-FCM值略大于IM-FCM及GIFP-FCM算法.因此,针对纹理图像这种样本分布密度大、样本间差异较小的数据集的模糊聚类,本文所提出的IM-FCM算法只是稍优于其他两种算法,并没有明显的优势.

综合上述所有实验对比结果,可以表明:在含有大量噪声样本或样本分布不均衡情况下,本文所提出的算法获得了较好的聚类效果;当样本集所含噪声超过一定比例(例如超过20%)时,本文所提出的方法的聚类效果好;而在样本分布相对均衡或不含噪声点的情况下,本文所提出方法的聚类优势并不明显.

## 4 结 论

本文通过放松FCM聚类算法中模糊隶属度约束条件,推导出一种改进的FCM聚类隶属度函数,引用一个基于距离运算的非线性处理对隶属度函数进行修正,以得到很好的解决传统FCM算法中样本分布不均衡,特别是含有大量噪声样本的模糊聚类问题,并通过多次实验验证了所提出方法的有效性.本文的主要创新点在于:一是对放松隶属度约束后的隶属度函数、聚类中心进行了推导,使得求解隶属度函数等有了理论依据;二是分析了基于该隶属度函数的孤立点或噪声点可能会单独成一类的情况,给出了隶属度函数的非线性修正方法,使得迭代后的隶属度值不会因过大或过小而影响聚类结果.需要指出的是,与传统的FCM算法一样,本文所提出的改进隶属度函数的FCM算法仍然对聚类中心十分敏感,尤其是当选取噪声点作为初始聚类中心时,聚类可能会失败,因此,如何在本文研究的基础上,克服初始聚类中心敏感的问题将是下一步的研究目标.

## 参考文献(References)

- [1] Tan Khung Siang, Lim Wei Hong. Novel initialization scheme for fuzzy  $C$ -means algorithm on color image segmentation[J]. Applied Soft Computing, 2013, 13(4): 1832-1852.
- [2] Jiashun Chen, Dechang Pi, Zhipeng Liu. An insensitivity fuzzy  $C$ -means clustering algorithm based on penalty factor[J]. J of Software, 2013, 8(9): 2379-2384.
- [3] Hopper F, Klawonn F. Improved fuzzy partitions for fuzzy regression models[J]. J of Approximate Reasoning, 2003, 33(2): 85-102.

- [4] 朱林, 王士同, 邓赵红. 改进模糊划分的FCM聚类算法的一般化研究[J]. 计算机研究与发展, 2009, 46(5): 814-822.  
(Zhu L, Wang S T, Deng Z H. Research on generalized fuzzy  $C$ -means clustering algorithm with improved fuzzy partitions[J]. J of Computer Research and Development, 2009, 46(5): 814-822.)
- [5] Li J, Gao X B, Jiao L C. A new feature weighted fuzzy clustering algorithm[J]. Acta Electronic Sinica, 2006, 34(1): 412-420.
- [6] 贺思艳, 李鹏, 刘澄玉, 等. 互模糊熵中隶属度函数的改进和影响分析[J]. 山东大学学报: 工学版, 2014, 44(1): 63-68.  
(He S Y, Li P, Liu C Y, et al. Refining of the membership function in cross fuzzy entropy and its influence[J]. J of Shandong University: Engineering Science, 2014, 44(1): 63-68.)
- [7] 魏延, 李晓虹, 邬啸. 后验概率加权的模糊隶属度函数[J]. 重庆大学学报, 2012, 35(8): 127-133.  
(Wei Y, Li X H, Wu X. Design fuzzy membership functions based on the posterior probability weighting[J]. J of Chongqing University, 2012, 35(8): 127-133.)
- [8] 肖满生, 汪新凡, 朱永平. 非均衡原型结构模式模糊类方法研究[J]. 小型微型计算机系统, 2013, 34(4): 868-871.  
(Xiao M S, Wang X F, Zhu Y P. Fuzzy clustering method research based on disequilibrium prototype pattern[J]. J of Chinese Computer Systems, 2013, 34(4): 868-871.)
- [9] Qing Niu, Xinjian Huang. An improved fuzzy  $C$ -means clustering algorithm based on PSO[J]. J of Software, 2011, 6(5): 873-879.
- [10] 李弼程, 彭天强, 彭波, 等. 智能图像处理技术[M]. 北京: 电子工业出版社, 2004: 302-304.  
(Li B C, Peng T Q, Peng B, et al. Intelligent image processing technology[M]. Beijing: Publishing House of Electronics Industry, 2004: 302-304.)
- [11] Dong-hyuck Park, Sang H Lee, Eui-Ho Song, et al. Similarity computation of fuzzy membership function pairs with similarity MEASURE[C]. Proc of the 3rd Int Conf on Intelligent Computing. Qingdao, 2007: 485-492.
- [12] Xiao Mansheng, Xiao Zhe, Liu Zhi. A method of feature automatic selection based on mutual information grouping and clustering[C]. Proc of 2014 Int Conf on Vehicle & Mechanical Engineering and Information Technology. Beijing, 2014: 1613-1618.
- [13] Frank A, Asuncion A. UCI machine learning repository[EB/OL]. [2012-02-01]. <http://archive.ics.uci.edu/ml/datasets.html>.
- [14] Trygve Randen. Brodatz textures[EB/OL]. [2007-08-01]. <http://www.uu.uis.no/~tranden/brodatz.htm>.

(责任编辑: 曹洪武)