基于划分的模糊聚类算法*

张 敏, 于 剑+

(北京交通大学 计算机与信息技术学院,北京 100044)

Fuzzy Partitional Clustering Algorithms

ZHANG Min, YU Jian+

(School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China)

+ Corresponding author: Phn: +86-10-51688055, E-mail: jianyu@center.njtu.edu.cn, http://cit.njtu.edu.cn

Received 2003-07-16; Accepted 2003-11-11

Zhang M, Yu J. Fuzzy partitional clustering algorithms. ***Journal of Software*****, 2004,15(6):858~869.**
http://www.jos.org.cn/1000-9825/15/858.htm

Abstract: Fuzzy partitional clustering algorithms are widely used in pattern recognition field. Until now, more and more research results on them have been developed in the literature. In order to study these algorithms systematically and deeply, they are reviewed in this paper based on c-means algorithm, from metrics, entropy, and constraints on membership function or cluster centers. Moreover, the advantages and disadvantages of the typical fuzzy partitional algorithms are discussed. It is pointed out that the standard FCM algorithm is robust to the scaling transformation of dataset, while others are sensitive to such transformation. Such conclusion is experimentally verified when implementing the standard FCM and the maximum entropy clustering algorithm. Finally, the problems existing in these algorithms and the prospects of the fuzzy partitional algorithms are discussed.

Key words: partitional clustering; C-means; weighting exponent; entropy; membership function

摘 要: 在众多聚类算法中,基于划分的模糊聚类算法是模式识别中最常用的算法类型之一.至今,文献中仍不断有关于基于划分的模糊聚类算法的研究成果出现.为了能更为系统和深入地了解这些聚类算法及其性质,本文从改变度量方式、改变约束条件、在目标函数中引入熵以及考虑对聚类中心进行约束等几个方面,对在 C-均值算法的基础上得到的基于划分的模糊聚类算法作了综述和评价,对各典型算法的优缺点进行了实验比较分析.指出标准 FCM 算法被广泛应用的原因之一是它对数据的比例变化具有鲁棒性,而其他类似的算法对这种比例变化却很敏感,并以极大熵方法为例进行了比较实验.最后总结了基于划分的模糊聚类算法普遍存在的问题及其发展前景.

关键词: 划分聚类;C 均值;权重指数;熵;隶属度函数

中图法分类号: TP18 **文献标识码**: A

聚类就是将物理或抽象的对象,按照对象间的相似性进行区分和分类的过程.在这一过程中没有教师指导,

---

* Supported by the National Natural Science Foundation of China under Grant No.60303014 (国家自然科学基金); the Scientific Key Project of Ministry of Education of China under Grant No.02031 (教育部科学技术研究重点项目)

作者简介: 张敏(1979－),女,江西南昌人,硕士生,主要研究领域为数据挖掘,聚类分析;于剑(1969－),博士,副教授,主要研究领域为计算智能,模式识别,数据挖掘.

因此是一种无监督的分类.聚类所生成的簇是一组数据对象的集合,在同一簇中的对象之间具有较高的相似度,而不同簇中对象的差别较大.聚类分析已经被广泛地应用到许多领域中,包括模式识别、数据分析、图像处理以及市场研究等.在商务上,聚类能够帮助市场分析人员从客户基本库中发现不同的客户群;在生物学上,聚类用于推导植物和动物的分类,对基因进行分类;聚类也能对 Web 上的文档进行分类,以发现信息,等等.

聚类的方法可以分为基于划分的方法、基于分层的方法、基于密度的方法和基于网格的方法.其中,基于划分的聚类算法在模式识别里是最常用的聚类算法类型,本文主要是针对此类算法进行讨论.基于划分的聚类方法有时也叫做基于目标函数的聚类算法.本文假设聚类算法的目标函数都是可微的,算法处理的都是数值型数据,而且除了初始化以外,没有采用抽样技术.

传统的划分方法是一种硬划分,是把每个待处理的对象严格地划分到某个类中.硬划分方法的典型代表是 C 均值算法.在这个算法中,隶属度不是 1 就是 0,而现实中大多数的对象并没有严格的属性,这种硬划分并不能真正地反应对象和类的实际关系,因此,人们就提出了要对待处理的对象进行软划分.Zedeh 提出的模糊集理论为软划分提供了有力的分析工具,人们开始用模糊的方法来处理聚类问题,基于划分的模糊聚类的研究也由此开始.基于划分的模糊聚类分析建立了样本类属的不确定性的描述,能够比较客观地反映现实世界.

模糊聚类属于模式识别中的无监督学习,它不需要训练样本,可以直接通过机器学习达到自动分类的目的.模式识别中最关键的技术就是特征提取,模糊聚类不但能从原始数据中提取特征,而且还能对特征进行优化选择和降维;在提取特征之后,模糊聚类还可以提供最近邻原型分类器,以及进行空间划分和模糊规则的提取,帮助构造基于模糊 IF-THEN 规则的分类器;在物体识别和线条检测中,模糊聚类可以用于原始的数据上,也可用于变换域中.在模式识别的一些具体应用领域中,模糊聚类也取得了较好的结果,比如,汉字识别的字符预分类、语音识别中的分类和匹配等.

本文从改变度量方式、改变隶属度约束条件、在目标函数中引入熵以及加入对聚类原型的约束条件等几个方面对文献中现有的基于模糊划分的聚类算法进行了分类,至于其他聚类算法,有兴趣的读者可以参考文献[1,2].

本文的第 1 节介绍了基于硬划分的典型代表聚类算法 C-均值算法.第 2 节按照实现模糊化的两种思路对基于划分的模糊聚类算法作了综述和分析.第 3 节对各典型算法的优缺点进行了实验比较分析.最后总结了基于划分的模糊聚类算法中普遍存在的问题以及未来的发展.

1 基于硬划分的聚类算法——C-均值聚类算法

基于前面的讨论,我们可以作如下假设:设 $X=\{x_1,x_2,...,x_n\}\subset R^s$ 是一个数据集,$u=\{u_{ik}\}_{c\times n}\in M_{fcn}$ 是一个隶属度矩阵,$v=\{v_1,v_2,...,v_c\}$ 是 c 个聚类中心,$v_i\in R^s, 2\le c<n$.

C-均值算法把 n 个向量 $x_i(i=1,2,\ldots,n)$分成 C 个簇 $G_i(i=1,2,...,c)$,并求得每个簇的聚类中心,使得簇内方差的和达到最小 $J(u,v)=\sum_{k=1}^{n}\sum_{i=1}^{c}u_{ik}\|x_k-v_i\|^2$,其中$\sum_{i=1}^{c}u_{ik}=1, u_{ik}\in\{0,1\}$.C-均值聚类算法的基本步骤如下:

(1) 初始化:给出初始聚类中心 $v^{(0)}=\{v_1^{(0)},v_2^{(0)},...,v_c^{(0)}\}, l=0$, l 为迭代次数,最大迭代次数为 T,阈值为ε.

(2) 用下列公式更新 $u_{ik}^{(l+1)}$:

$$u_{ik}^{(l+1)}=\begin{cases}1, \text{如果} i=\arg\min\{\|x_k-v_i^{(l)}\|\}\\ 0, \text{否则}\end{cases} \tag{1}$$

(3) 用下列公式更新 $v_i^{(l+1)}$:

$$v_i^{(l+1)}=\frac{\sum_{k=1}^{n}u_{ik}^{(l+1)}x_k}{\sum_{k=1}^{n}u_{ik}^{(l+1)}} \tag{2}$$

如果 $\max_i\|v_i^{(l+1)}-v_i^{(l)}\|<\varepsilon$ 或者 $l>T$,则停止;否则,$l=l+1$,转至步骤(2).

C-均值算法思想简单,实现容易,收敛快,运行速度快,内存消耗小,能有效地处理大数据集,是目前最常用的

聚类算法之一.但是,C-均值算法也有不少缺点,如采用的是硬划分,每类由类中心代表,使用欧氏度量,每个数据点的影响一样,没有考虑噪音数据的影响,也缺少对类中心的约束等.文献中针对这些缺点提出了各种改进措施.本文主要综述基于模糊集理论的各种对于 C-均值算法的改进算法.

2 基于划分的模糊聚类算法

C-均值算法的隶属度要么是 1,要么是 0,这不能反映数据点与类中心的实际关系.为了处理这个问题,人们引入了模糊集的概念.使用模糊数学理论的聚类算法被称为模糊聚类算法.自 1969 年 Ruspini 首先提出第 1 个解析的模糊聚类算法以来,已经有很多人提出了许多的模糊聚类算法.基于模糊划分的模糊聚类算法,其主要思想是将经典划分的定义模糊化,文献中主要有两种比较成功的思路来实现这种模糊化,一是在 C-均值算法的目标函数中引入隶属度函数的权重指数,二是在 C-均值算法目标函数中引入信息熵.

2.1 引入隶属度函数的权重指数

在众多的模糊聚类算法中,应用最广泛而且较成功的是 1974 年由 Dunn 提出并由 Bezdek 加以推广的模糊 C-均值(fuzzy C-means,简称 FCM)算法.同样,FCM 算法是把 n 个向量 $x_i(i=1,2,...,n)$ 分成 C 个模糊簇,并求得每个簇的聚类中心,使目标函数达到最小,$1<m<+\infty$,FCM 的目标函数定义为

$$J_m(u,v)=\sum_{k=1}^{n}\sum_{i=1}^{c}(u_{ik})^m d(x_k,v_i) \tag{3}$$

这里,$\sum_{i=1}^{c}u_{ik}=1, u_{ik}\in(0,1)$, $\forall k, d(x_k,v_i)=\|x_k-v_i\|^2$.与 C 均值算法不同的是,在目标函数中增加了模糊权重指数 m.为使目标函数达到最小,聚类中心和隶属度的更新如下:

$$v_i=\frac{\sum_{k=1}^{n}u_{ik}^m x_k}{\sum_{k=1}^{n}u_{ik}^m}\ (i=1,2,\ldots,c) \tag{4a}$$

$$u_{ik}=\frac{1}{\sum_{j=1}^{c}\left(\frac{d_{ik}}{d_{jk}}\right)^{1/(m-1)}}\ (i=1,2,...,c;k=1,2,...,n) \tag{4b}$$

FCM 算法计算简单而且运算速度快,具有比较直观的几何意义,但是与 C 均值算法一样,只用类中心来表示类,这样只适合于发现球状类型的簇.在很多情况下,算法对噪音数据敏感.Bezdek 等人已经证明了 FCM 算法只能保证收敛到公式(4a)和(4b)的不动点,不能保证收敛到目标函数的极小值点,有时会收敛到目标函数的鞍点.FCM 算法与 C 均值算法相比,引入了权重指数 m,而 m 选择的好坏直接影响聚类的效果.由于 FCM 算法存在种种缺陷,许多人对 FCM 算法进行改进,并提出了新的模糊聚类算法.

2.1.1 改变度量方式

在噪音环境下,采用欧氏度量的许多聚类方法有时不够稳定,而且对于算法的初值,类的形状、大小都过于敏感.改变度量方式可以部分地控制这些问题.沿着这一思路,文献中已经提出了许多 FCM 型的算法,见表 1.当然,度量方式的改变也会使得迭代过程中聚类中心、隶属度的调整以及各算法的参数都有所不同.但与 FCM 算法类似,这些调整大多是通过 Lagrange 乘法算子求得的.由于受篇幅的限制,各算法的聚类中心、隶属度和参数的调整方法就不一一列出了,可以参考相关文献.

由于通过改变度量方式得到的目标函数形式 $J=\sum_{k=1}^{n}\sum_{i=1}^{c}(u_{ik})^m d(x_k,v_i)$ 和隶属度的约束条件与 FCM 算法形式一样,在此不一一重复.我们在表 1 中总结了文献中部分改变度量方式的 FCM 型算法的度量方式.

Table 1 The FCM algorithms with different metrics

表 1 通过改变度量方式由 FCM 得到的模糊聚类算法

Algorithm	Metric	Note
Gustanfson-Kessel[3]	$d(x_k,v_i)=(x_k-v_i)A_i^{-1}(x_k-v_i)^T$	A_i is a fuzzy covariance matrix
AFCM[4]	$d(x_k,v_i)=1-\exp(-\beta\|x_k-v_i\|^2)$	$\beta>0$, choice of β depends on data
PFCM[5]	$d(x_k,v_i)=\|x_k-v_i\|^2-w\ln\xi_i$	Adding a penalty term $w\ln\xi_i$, which satisfies the conditions in Remark 1
CFCM[6]	$d(x_k,v_i)=\|x_k-v_i\|^2+\tau\tanh\zeta_i$	Adding a compensated term $\tau\tanh\zeta_i$, which satisfies Remark 2
PIM[7]	$d(x_k,v_i)=\|x_k-v_i\|^2-w$	$w\geq 0$ is called index term
FGcM[8]	$d(x_k,v_i)=\|x_k-v_i\|^2+[\lambda(m-1)]^{-1}$	$\lambda>0,m>1$
Lp-norm[9]	$d(x_k,v_i)=\|x_k-v_i\|_p^p=\sum_{j=1}^{s}\|x_{k,j}-v_{i,j}\|^p$	Using BEA algorithm during iteration, more details in Ref.[9,10,52].
FCS[11]	$d(x_k,v_i)=(\|x_k-v_i\|-r_i)^2$	Remark 3
FCSS[12]	$d(x_k,v_i)=(\|x_k-v_i\|^2-r_i^2)^2$	Remark 4
KS[13]	$d(x_k,v_i)=\|x_k-v_i\|^2-r_i^2$	r_i is a constant
Fuzzy compactness and separation[14]	$d(x_k,v_i)=\|x_k-v_i\|^2-\eta_i\|v_i-\bar{x}\|^2$	Remark 5
NC[15]	$d(x_k,v_{c+1})=\delta^2$	Remark 6
RCP[16]	$d(x_k,v_i)=\rho_i(\|x_k-v_i\|^2)$	Remark 7

注 1: $\forall i,\xi_i\geq 0$,且 $\sum_{i=1}^{c}\xi_i=1$, and $w\geq 0$.

注 2: $\forall i,\zeta_i\geq 0$,且 $\sum_{i=1}^{c}\zeta_i=1$, and $\tau\geq 0$.

注 3:可以发现球型和椭球型的簇,采用了牛顿法,增加了计算复杂度.

注 4:同样可以用来发现球型和椭球型的簇,在计算复杂性上较 FCS 有所改进.有关细节问题见文献[12].

注 5: $\eta_i\geq 0$, $\bar{x}=\frac{\sum_{k=1}^{n}x_k}{n}$.

注 6:定义第 c+1 个类为噪音类,数据点距离噪音类的类中心的距离定义为常数δ,数据集中数据点隶属于第 c+1 个类的隶属度定义为 $u_{*k}=1-\sum_{i=1}^{c}u_{ik}$,因此,也可以将此算法归入基于隶属度条件改变的 FCM 型算法.有关细节可见文献[15,53].

注 7:应用一个损失函数 ρ 在平方距离之上可以降低孤立点的影响,当$\rho(d)=d$时,就是标准的FCM算法,文献[16]主要研究了一个分段连续损失函数对 FCM 算法的性能造成的影响.

2.1.2 改变隶属度约束条件

从上面我们可以看到,C-均值和 FCM 算法都有 $\sum_{i=1}^{c}u_{ik}=1$ 的归一化约束条件.这一条件假定每个数据点的影响力是相同的,显然这与实际情况并不总是一致.如果放松对隶属度函数的要求,则将导致新的聚类算法.针对这一思路,文献中提出了许多新的算法.表 2 总结了文献中通过改变隶属度的约束条件,由 FCM 得到的算法.算法中隶属度、聚类中心以及相关参数的调整可见相关的文献(以下算法中 $d_{ik}=\|x_k-v_i\|^2$,如果没有进行特殊说明,$0\leq u_{ik},t_{ik}\geq 0$) .

Table 2 The fuzzy clustering algorithms induced by changing the constraints on membership function in the FCM

表 2 通过改变隶属度约束条件由 FCM 得到的模糊聚类算法

Algorithm	Objective function	Note
Lee's algorithm[17]	$J=\sum_{k=1}^{n}\sum_{i=1}^{c}(u_{ik})^m d_{ik}$	$\sum_{i=1}^{c}\sum_{k=1}^{n}u_{ik}=n$
Conditional FCM[18]	$J=\sum_{k=1}^{n}\sum_{i=1}^{c}(u_{ik})^m d_{ik}$	$\sum_{i=1}^{c}u_{ik}=a_k\geq 0$
PCM(1)[20]	$J=\sum_{i=1}^{c}\sum_{k=1}^{n}t_{ik}^m d_{ik}+\sum_{i=1}^{c}\eta_i\sum_{k=1}^{n}(1-t_{ik})^m$	Remark 8
WPCM[21]	$J=\sum_{i=1}^{c}\sum_{k=1}^{n}u_{ik}^m d_{ik}+\sum_{i=1}^{c}\eta_i\sum_{k=1}^{n}(\varphi_k-u_{ik})^m$	Remark 9
PCM(2)[22]	$J=\sum_{i=1}^{c}\sum_{k=1}^{n}u_{ik}d_{ik}+\sum_{i=1}^{c}\eta_i\sum_{k=1}^{n}(u_{ik}\log u_{ik}-u_{ik})$	The entropy of the *ith* cluster is $\sum_{k=1}^{n}(u_{ik}\log u_{ik}-u_{ik})$
PGcM[9]	$J=\sum\sum u_{ik}^m d_{ik}+\frac{1}{\lambda(m-1)}\sum_{i=1}^{c}\sum_{k=1}^{n}(u_{ik}^m-u_{ik})-\frac{1}{\lambda}\sum_{i=1}^{c}\sum_{k=1}^{n}u_{ik}$	If $m\rightarrow 1$, the objective function of PGcM is like that of PCM
GFCM[23]	$J=\sum_{k=1}^{n}\sum_{i=1}^{c}(u_{ik})^m d_{ik}$	Remark 10
FPCM[24]	$J=\sum_{k=1}^{n}\sum_{i=1}^{c}(u_{ik}^m+t_{ik}^\eta)d_{ik}$	Remark 11
PFPCM[25]	$J=\sum_{k=1}^{n}\sum_{i=1}^{c}(u_{ik}^m+t_{ik}^m)d_{ik}-w\sum_{i=1}^{c}\sum_{k=1}^{n}(u_{ik}^m\ln\alpha_i+t_{ik}^m\ln\beta_k)$	Remark 12
CFPCM[25]	$J=\sum_{k=1}^{n}\sum_{i=1}^{c}(u_{ik}^m+t_{ik}^\eta)d_{ik}+\tau\sum_{k=1}^{n}\sum_{i=1}^{c}(u_{ik}^m\tanh(\alpha_i)+t_{ik}^\eta\tanh(\beta_k))$	Remark 13
IPCM(1)[56]	$J=\sum_{i=1}^{c}\sum_{k=1}^{n}u_{ik}^{m_f}t_{ik}^m d_{ik}+\sum_{i=1}^{c}\eta_i\sum_{k=1}^{n}u_{ik}^{m_f}(1-t_{ik})^m$	Remark 14
IPCM(2)[56]	$J=\sum_{i=1}^{c}\sum_{k=1}^{n}u_{ik}^{m_f}t_{ik}d_{ik}+\sum_{i=1}^{c}\eta_i\sum_{k=1}^{n}u_{ik}^{m_f}(t_{ik}\log t_{ik}-t_{ik}+1)$	Remark 14

注 8:$0\leq t_{ik}\leq 1,\forall i\forall k,t_{ik}$表示点 x_k 在第 i 个类内的典型性或 x_k 属于第 i 个类的概率,它只依赖于 x_k 与 v_i 的距离而与其他类中心的位置无关,细节请见文献[14,19,20].

注 9:$0\leq u_{ik}\leq\varphi_k,\forall i\forall k$,若 $\varphi_k=1,\forall k$,WPCM 就和 PCM(1)相对应,可以认为,WPCM 是 PCM(1)的一个特例.

注 10:隶属度的约束条件 $\forall i,\forall k,u_{ik}\geq 0,\forall k,F(u_{1k},u_{2k},\ldots,u_{ck})=1$ 且 $\forall\alpha,F(\alpha u_{1k},\alpha u_{2k},\ldots,\alpha u_{ck})=\alpha$,FCM 算法是 GFCM 算法的一个特例,有关算法的情况见文献[23].

注 11:$u_{ik}\geq 0,\sum_{i=1}^{c}u_{ik}=1$; $t_{ik}\geq 0$, $\sum_{k=1}^{n}t_{ik}=1$.关于 u_{ik} 和 t_{ik} 的具体解释,可见文献[24].

注 12:$u_{ik}\geq 0,\sum_{i=1}^{c}u_{ik}=1$; $\forall i,\alpha_i\geq 0,\sum_{i=1}^{c}\alpha_i=1$; $t_{ik}\geq 0$, $\sum_{k=1}^{n}t_{ik}=1$, $\forall k,\beta_k\geq 0,\sum_{k=1}^{n}\beta_k=1$, $w>0$.

注 13:$u_{ik}\geq 0,\sum_{i=1}^{c}u_{ik}=1$; $\forall i,\alpha_i\geq 0,\sum_{i=1}^{c}\alpha_i=1$; $t_{ik}\geq 0$, $\sum_{k=1}^{n}t_{ik}=1$, $\forall k,\beta_k\geq 0,\sum_{k=1}^{n}\beta_k=1$, $\tau>0$.

注 14:$u_{ik}\geq 0,\sum_{i=1}^{c}u_{ik}=1$, $m_f>1$,IPCM(1)(2)的提出是为了消除 PCM(1) (2)和 FPCM 的缺点,见文献[56].

2.2 引入熵

在 C 均值算法的目标函数中,引入熵的概念,则可以得到最大熵意义下的模糊聚类算法.表 3 总结了目前文献中这类算法.在这些算法的目标函数中,大部分取消了隶属度的权重指数这一参数,但也同时引入了一个新的参数 Lagrange 乘子,隶属度函数依然要满足归一化条件(其中 $d_{ik}=\|x_k-v_i\|^2$).类似地,表 3 中算法的隶属度和聚类中心的迭代公式可由 Lagrange 乘子法得到(但表 3 中最后两个算法除外),在此不再一一赘述,有兴趣的读者可

以参考相关的文献.

Table 3 The fuzzy partitional clustering algorithms when adding entropy term into the objective function of C-means

表 3 在 C 均值的目标函数中引入熵得到的算法

Algorithms	Objective function	Note
MEC[26]	$J=\sum_{k=1}^{n}\sum_{i=1}^{c}u_{ik}d_{ik}+\lambda^{-1}\sum_{k=1}^{n}\sum_{i=1}^{c}u_{ik}\log u_{ik}$	If λ>0, the objective function of DA has the same form
MECA[27]	$J=\alpha\sum_{k=1}^{n}\sum_{i=1}^{c}u_{ik}\log u_{ik}+\frac{1-\alpha}{n}\sum_{k=1}^{n}\sum_{i=1}^{c}u_{ik}d_{ik}$	Remark 15
A fuzzy explanation of EM clustering algorithm[28]	$J=\sum_{k=1}^{n}\sum_{i=1}^{c}u_{ik}\left(d_{ik}+\log\lvert A_i\rvert\right)+\lambda\sum_{k=1}^{n}\sum_{i=1}^{c}u_{ik}\log\frac{u_{ik}}{\pi_i}$	K-L entropy ;$\forall i,\pi_i\geq 0,\ \sum_{i=1}^{c}\pi_i=1$ A_i is a fuzzy covariance matrix
Fuzzy clustering based on Fermi-Dirac statitics[29]	$J=\sum_{k=1}^{n}\sum_{i=1}^{c}u_{ik}d_{ik}+\lambda\sum_{k=1}^{n}\sum_{i=1}^{c}\left[u_{ik}\log u_{ik}+(1-u_{ik})\log(1-u_{ik})\right]$	Remark 16
FBACN[30]	$J=\sum_{k=1}^{n}\sum_{i=1}^{c}u_{ik}^{m}d_{ik}+\lambda^{-1}\sum_{k=1}^{n}\sum_{i=1}^{c}u_{ik}\log u_{ik}$	Consider Shannon's entropy without deleting the weighting exponent, see Remark 17

注 15: $0<\alpha<1$,若设 $\lambda=\frac{1-\alpha}{n\alpha}$,则目标函数与 MEC 的相同.

注 16:Fermi-Dirac 统计下的聚类算法由于缺乏聚类中心和隶属函数的闭式解,已经不能用简单的两步迭代方法来实现聚类中心和隶属函数的更新.

注 17:FBACN 算法的目标函数同样得不到聚类中心和隶属函数的闭式解,发明者使用神经网络去求解 FBACN 算法的目标函数的极值,技术细节见文献[30].

2.3 类中心的约束

注意到模糊 C 均值算法和 C 均值算法都没有考虑聚类中心的约束,如果考虑这一情形,将会得到一些新的模糊聚类算法.现在,就我们所知,文献中至少有 3 种模糊算法是按照这一思路来设计的.下面,我们将依时间的先后顺序逐一简述这 3 种算法.

为了应用决定退火聚类的思想来解决旅行商问题,Rose 提出了考虑类中心之间的距离,为此提出了如下目标函数:

$$J=\sum_{k=1}^{n}\sum_{i=1}^{c}u_{ik}d_{ik}+\beta\left(\lVert v_c-v_1\rVert^2+\sum_{i=1}^{c-1}\lVert v_{i+1}-v_i\rVert^2\right)+\lambda^{-1}\sum_{k=1}^{n}\sum_{i=1}^{c}u_{ik}\log u_{ik},\ \forall i,k,u_{ik}\geq 0,\sum_{i=1}^{c}u_{ik}=1,\text{且 }\lambda,\beta>0.$$

应用 Lagrange 乘子法,可以得到类中心的迭代公式.有兴趣的读者可以自己动手试一试,也可见文献[31~33].

2001 年,Ozdemir 提出了 Inter-cluster separation (ICS)[34],在模糊 C 均值算法的目标函数中加入一个分裂项,形成如下的目标函数:

$$J=\frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{c}\left[(u_{ik})^m d_{ik}-\frac{\gamma}{c}\sum_{t=1}^{c}d(v_i,v_t)\right],\ \forall i,k,u_{ik}\geq 0,\sum_{i=1}^{c}u_{ik}=1.$$

最近,2002 年 Timm 和 Kruse 为了避免产生一致的类中心,在 PCM 算法的目标函数中也加入类中心的排斥项,得到如下的目标函数:

$$J=\sum_{k=1}^{n}\sum_{i=1}^{c}u_{ik}^{m}d_{ik}+\sum_{i=1}^{c}\eta_i\sum_{k=1}^{n}(1-u_{ik})^m+\sum_{i=1}^{c}\gamma_i\sum_{i\neq t}[d(v_i,v_t)]^{-1},\ \forall i,k,u_{ik}\geq 0,\sum_{i=1}^{c}u_{ik}<1,\ \eta_i>0,\gamma_i\geq 0.$$

在同一篇文章里,Timm 和 Kruse 也提出了其他类中心的排斥项形式,可见参考文献[35,36].

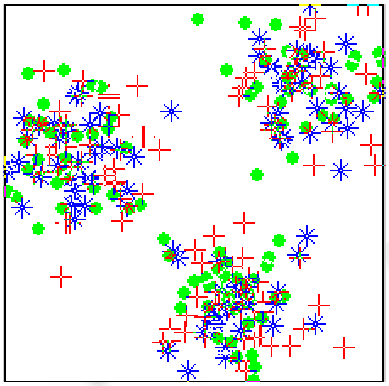
3 算法的比较分析

从上面我们所作的总结来看,基于划分的模糊聚类算法大致可分为 FCM 型、PCM 型、引入熵型和类中心约束型.下面我们针对每类算法的优缺点进一步地进行分析比较.

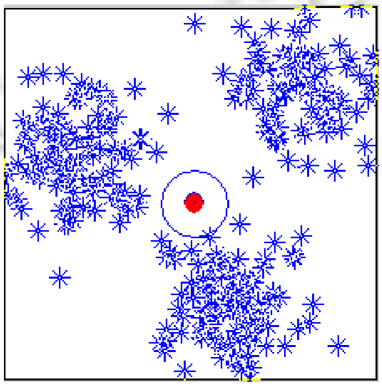
从表 1 中可以看出,改变度量方式得到的 FCM 型算法可以分为线性改变和非线性改变两种.例如,PFCM,CFCM 等算法都是对度量方式进行线性改变得到的新算法,这种线性改变并没有增加算法的复杂度,对 FCM 算法参数选择的理论以及聚类有效性评判的一些标准仍然适用于这些算法.而像 AFCM 等算法是通过对度量方式进行非线性变化而得到的算法.这种改变针对某些应用,从某种程度上改善了聚类结果,能发现更多具有非球型的类,但是算法变得更加复杂了,有些算法的迭代过程都发生了改变,而对这些算法的参数选择和聚类有效性的评价,目前还缺乏理论上的支持.但是,容易证明这种改变度量方式得到的聚类算法在数据尺度发生改变的时候,即数据扩大或缩小一定倍数的时候,会比较敏感,可能会导致很糟糕的聚类结果,所以这些算法的适用性并不如 FCM 算法广泛.

FCM 算法及改变度量方式得到的聚类算法都要求隶属度归一,这种约束的最大缺陷就是当数据中包含噪音和孤立点的时候,聚类效果不好,很可能这些噪音和孤立点都被赋予较大的隶属度,因此出现了很多通过改变隶属度的约束条件而得到的算法.其典型的代表就是 PCM,它本质上是一种穷举型搜索算法,见文献[14].与 FCM 算法相比,PCM 算法能较好地处理噪音,弥补了 FCM 受噪音影响大的缺陷.但是,PCM 需要一个好的初始划分,以提供准确的聚类,这是一个比较难解决的问题.Krishnapuram 和 Keller 建议把 FCM 算法得出的结果作为 PCM 算法的初始划分,而且建议 PCM 算法执行两次,每次参数的设置都不同,见文献[19,20].我们对 PCM 算法在不同的数据集上进行测试,其中权重指数 m=2,误差取为 0.00001,迭代次数为 100,无论指定类别个数是多少,PCM 算法都以较大的概率产生了一致的类中心.产生一致的类中心的原因是类之间不存在相互联系,也就是说,类是相互独立的,目标函数的最小化是通过对每个类相互独立地操作完成的,而且即使有一个很好的初始划分,PCM 也仍然会丢失数据中隐藏的一些结构.Krishnapuram 和 Keller 还指出,导致一致的类中心是由于受到权重指数 m 的影响,权重指数 m 在 FCM 和 PCM 中的解释是不同的,FCM 中 m 的增加表示数据集中的所有点在类间的模糊性都增加了,而在 PCM 中 m 的增加表示数据集中的所有点完全属于一个指定的类的概率增加了.有人建议对 PCM 的目标函数进行改进,消除 m 的影响,因而之后出现了许多新型的 PCM 算法,例如 PCM(2).在很多相关文献[14,22]中都有针对新型的 PCM 算法的一些介绍和理论推导.

图 1 给出了 PCM(1)在随机产生的数据集上的聚类结果.如图 1(a)所示的数据集有 3 个类中心:(0.6,0.2),(0.2,0.6),(0.8,0.8).根据算法 PCM(1)的初始隶属度将数据分成 3 类,并用不同的表示标出.图 1(b)为聚类后,对于如图 1(a)所示的数据集算法 PCM(1)输出一个重合的类中心,所有的数据点都分派给此类中心.



(a)



(b)

Fig.1 The dataset and the cluster result of PCM (1)

图 1 PCM(1)的数据集和聚类结果

引入熵得到的算法与 FCM 算法相比,有更清晰简洁的数学形式和物理含义.FCM 算法需要对数据与类中心重合的情形进行特殊的处理,引入熵得到的算法却消除了这种需要特殊处理的情形.当把极大熵原则运用在最小化均方误差的时候,很自然地就能得到极大熵聚类算法(MEC),它是引入熵方法中的一个典型代表.通过对 FCM 算法和 MEC 算法的迭代过程中隶属度和类中心的更新情况进行分析,我们发现,在 FCM 算法的迭代中式

(4a)、式(4b),当数据尺度扩大或缩小一定的倍数时,$\frac{d_{ik}}{d_{jk}}$ 并不会随着这种变化而变化,因而隶属度是不会发生改变的,这样,类中心 v_i 和数据 x_k 是呈线性变化的,类中心随数据扩大和缩小的倍数也相应地扩大和缩小相同的倍数.而 MEC 算法的迭代公式如下:

$$v_i = \frac{\sum_{k=1}^{n} x_k u_{ik}}{\sum_{k=1}^{n} u_{ik}} \tag{5a}$$

$$u_{ik} = \frac{e^{-\lambda d_{ik}}}{\sum_{j=1}^{c} e^{-\lambda d_{jk}}} \tag{5b}$$

隶属度会随着数据的扩大或缩小发生变化,类中心 v_i 和数据 x_k 并不满足线性关系.也就是说,在数据扩大或缩小的时候,MEC 算法得到的聚类中心受这种变化的影响非常大,而 FCM 算法的聚类中心对于这种变化却具有鲁棒性.为了验证这一点,我们把 MEC 算法和 FCM 算法分别应用在 IRIS 数据上(IRIS 是来自于 UCI 的四维数据 http://www.ics.uci.edu/~mlearn/MLRepository.html).为方便起见,我们取 IRIS 的第 1 维和第 2 维数据进行测试,算法迭代次数取为 100,误差取为 0.000 01,类别个数取为 3,FCM 算法中 $m=2$,MEC 中 $\lambda=2$.表 4 列出了我们的实验结果.

Table 4 The outcomes of FCM and MEC on rescaling dataset: Cluster centers

表 4 FCM 和 MEC 算法在数据进行比例变换后得到的类中心

Algorithm	*data*	10**data*	100**data*	0.1**data*
FCM	(5.834 9, 2.765 8)	(58.348 9, 27.657 8)	(583.488 7, 276.578 1)	(0.583 5, 0.276 6)
	(4.976 6, 3.336 4)	(49.766 1, 33.364 0)	(497.661 4, 333.640 4)	(0.497 7, 0.333 6)
	(6.816 2, 3.071 2)	(68.161 6, 30.711 8)	(681.615 6, 307.118 4)	(0.681 6, 0.307 1)
MEC	(5.825 3, 2.872 5)	(57.736 3, 26.924 7)	(577.358 5, 269.245 3)	(0.584 3, 0.305 4)
	(5.093 7, 3.250 8)	(50.060 0, 34.180 0)	(500.600 0, 341.800 0)	(0.584 3, 0.305 4)
	(6.722 2, 3.021 8)	(68.128 2, 30.744 9)	(681.276 6, 307.446 8)	(0.584 3, 0.305 4)

从表 4 可以看出,FCM 算法的类中心随数据扩大和缩小的倍数也相应地扩大和缩小相同的倍数,而 MEC 算法得到的类中心的变化却不服从这个规律,这与我们的分析是一致的.引入熵的大多数算法迭代过程中类中心和隶属度的更新与 MEC 是大致相同的,所以它们得到的类中心对数据的扩大或缩小同样会比较敏感.

在实验的过程中我们还发现,MEC 算法在 IRIS 数据缩小到一定的时候,会得到一致的类中心,而 λ 和这种尺度的变化密切相关.

从表 5 可以看出,在数据集缩小 10 倍、100 倍和 1 000 倍的时候,MEC 算法得到了非常糟糕的结果,即得到一致的聚类中心,此时可以认为聚类算法失败.λ 的选取是和这种数据的尺度变化密切相关的,可见 λ 选取的好坏对聚类结果的影响非常大.而对 FCM 而言,只要 m 选取合适,不论数据比例如何变换,都不会导致聚类算法的失败.

为了在一定程度上避免一致类中心的产生,有人提出了在目标函数中增加对类中心的约束,可以使得类中心互相远离.作为该类算法的典型代表,我们对 ICS 算法进行了较为深入的研究.ICS 是针对图像应用提出来的,目的是使类中心彼此远离,从而获得一个更宽的颜色空间,提高颜色的对比度.它是一种将图像处理中的量化技术和抖动进行结合的算法,在图像分割上取得了不错的结果.FCM 只是一种单纯的图像量化技术,所产生的颜色空间的跨度不如 ICS 大,而恰恰是由于 ICS 产生较大的颜色空间,使得图像在经过抖动处理之后让人眼感觉到更多颜色的存在.另外,ICS 算法的一个关键问题是参数 γ 的选择.文献[34]中给出了参数选择的一些经验.

Table 5 The outcomes of MEC algorithm with different λ on down-scaling dataset

表 5 λ取不同值时,MEC 算法在进行缩小变化的数据集上得到的结果

Data transformation	λ=1	λ=100	λ=1e+4	λ=1e+8
data	(5.512 4, 3.076 4) (5.512 2 , 3.076 4) (6.511 2, 3.008 8)	(5.774 4, 0.692 8) (5.006 0, 3.417 9) (6.813 5, 3.074 7)	(5.800 0, 2.700 0) (5.003 9, 3.400 0) (6.823 9, 3.078 3)	(5.800 0, 2.700 0) (5.003 9, 3.400 0) (6.823 9, 3.078 3)
0.1**data*	(0.584 3, 0.305 4) (0.584 3, 0.305 4) (0.584 3, 0.305 4)	(0.551 2, 0.307 6) (0.551 2, 0.307 6) (0.651 1, 0.300 9)	(0.577 4, 0.269 3) (0.500 6, 0.341 8) (0.681 4, 0.307 5)	(0.577 4, 0.269 2) (0.500 6, 0.341 8) (0.681 3, 0.307 4)
0.01**data*	(0.058 4, 0.030 5) (0.058 4, 0.030 5) (0.058 4, 0.030 5)	(0.058 4, 0.030 5) (0.058 4, 0.030 5) (0.058 4, 0.030 5)	(0.058 4, 0.030 5) (0.058 4, 0.030 5) (0.058 4, 0.030 5)	(0.058 0, 0.027 0) (0.050 0, 0.034 0) (0.068 2, 0.030 8)
0.001**data*	(0.005 8, 0.003 1) (0.005 8, 0.003 1) (0.005 8, 0.003 1)	(0.005 8, 0.003 1) (0.005 8, 0.003 1) (0.005 8, 0.003 1)	(0.005 8, 0.003 1) (0.005 8, 0.003 1) (0.005 8, 0.003 1)	(0.005 8, 0.002 7) (0.005 0, 0.003 4) (0.006 8, 0.003 1)

4 基于划分的模糊聚类算法存在的问题及其发展前景

在本文中,我们总结了文献中出现的基于划分的模糊聚类算法,对这些算法进行了综述和分析比较,并指出了标准 FCM 算法对数据的比例变化具有鲁棒性,而其他的算法对这种变化非常敏感.这些算法的提出其基本出发点都是为了改进 C 均值算法的性能,扩大 C 均值算法的应用领域.确实,在一定程度上,文献中提出的算法实现了设计者的部分良好愿望,但是我们也遗憾地看到,基于划分的模糊聚类算法存在着很多需要解决的问题.这些问题也是我们正在从事的研究方向:

(1) 与 C 均值算法比较,它们都引入了新的参数,有的引入的参数还比较多.不仅如此,理论和实验都已经证明,上述算法新引入参数的不同取值会直接影响算法的性能,例如:模糊 C 均值算法引入了加权指数 m.就这一问题,我们已经进行了深入研究,得到了部分算法实用的参数选择理论成果,见文献[38~40,54,55].

(2) 基于划分的模糊聚类算法如此之多,如何根据实际情况选取一个合适的算法呢?更简单一点,同一个算法,由于采用度量方式的不同,聚类结果也不尽相同,但是如何根据实际情况,选取最为合适的度量方式呢？这与不同聚类算法的性能评估有关.相对于参数选择问题,这似乎是一个更难的问题.目前文献中对这个问题的研究基本上是基于经验比较,缺少严格的理论支持.

(3) 目前,对于已提出的聚类算法的收敛性、解的稳定性问题,进行深入理论研究的相对较少.文献中对于 C 均值聚类算法,FCM 算法的研究较为深入,它们的收敛性都已经被证明,参见文献[44~46,48],极大熵聚类算法的收敛性也已得到证明,参见文献[47],至于其他算法的收敛性则很少涉及.关于解的稳定性问题,FCM 算法已经有人研究过,参见文献[49~51],对于其他算法的稳定性问题我们已经得到了部分结果,参见文献[37,47].

考虑到现存的基于划分的模糊聚类算法存在这些问题,以及目前文献对这些问题的研究都比较零散,如果能够提出一个统一的基于划分的模糊聚类算法模型,统一研究算法的性质,对于解决基于划分的聚类算法存在的上述问题,甚至提出新的聚类算法,都将是一件十分有意义的事情.容易想象,随着人类社会的进步,会出现有别于以前的新需求,因此我们可以预测,针对不同应用的具体需要,今后还会有新的基于划分的模糊聚类算法提出来,这些新出现的算法,其具体的出发点也许别出心裁,但是,同样会遇到上述问题.如果我们能在理论上对上述问题作出回答,自然是最好的了,如果不能,从设计合理的实验中总结出有价值的结论也是很有意义的.而这些问题的早日解决,将会促进基于划分的模糊聚类算法的进一步成熟和更广泛的应用.

References:

[1] Han I, Kamber M. Data Mining: Concepts and Techniques. Berlin: Morgan Kaufmann Publishers, 2000. 335~389.

[2] Yang MS. A survey of fuzzy clustering. Mathl. Comput. Modelling, 1993,18(11):1~16.

[3] Gustafson DE, Kessel WC. Fuzzy clustering with a fuzzy covariance matrix. In: Proc. of the IEEE CDC. San Diego: IEEE, 1979. 761~766.

[4] Wu KL, Yang MS. Alternative c-means clustering algorithms. Pattern Recognition, 2002,35(10):2267~2278.

[5] Yang MS. On a class of fuzzy classification maximum likelihood procedures. Fuzzy Sets and Systems, 1993,57(3):365~375.

[6] Lin JS. Fuzzy clustering using a compensated fuzzy hopfield network. Neural Processing Letters, 1999,10(1):35~48.
[7] Ozdemir D, Akarun L. A fuzzy algorithm for color quantization of images. Pattern Recognition, 2002,35(8):1785~1791.
[8] Menard M, Courboulay V, Dardignac P. Possibilistic and probabilistic fuzzy clustering: Unification within the framework of the non-extensive thermostatistics. Pattern Recognition, 2003,36(6):1325~1342.
[9] Hathaway RJ, Bezdek JC, Hu YK. Generalized fuzzy c-means clustering strategies using L_p norm distances. IEEE Trans. on Fuzzy Systems, 2000,8(5):576~582.
[10] Bobrowski L, Bezdek JC. C-means clustering with the l_1 and l_∞ norms. IEEE Trans. on SMC, 1991,21(3):545~554.
[11] Dave RN, Bhaswan K. Adaptive fuzzy c-shells clustering and detection of ellipses. IEEE Trans. on Neural Networks, 1992,3(5): 643~662.
[12] Krishnapuram R, Nasraoui O, Frigui H. The fuzzy c spherical shells algorithm: A new approach. IEEE Trans. on Neuaral Network, 1992,3(5):663~671.
[13] Kaymak U, Setne M. Fuzzy clustering with volume prototypes and adaptive cluster merging. IEEE Trans. on Fuzzy Systems, 2002, 10(6):706~712.
[14] Yang MS, Wu KL, Yu J. A novel fuzzy clustering algorithm. In: Proc. of the 2003 IEEE Int'l Symp. on Computational Intelligence in Robotics and Automation. Kobe: IEEE, 2003. 647~652.
[15] Dave RN. Characterization and detection of noise in clustering. Pattern Recognition Letters, 1991,12(11):657~664.
[16] Frigui H, Krishnapu R. A robust algorithm for automatic extraction of an unknown number of clusters from noisy data. Pattern Recognition Letters, 1996,17(12):1223~1232.
[17] Cherkassky V, Mulier F. Learning From Data Concepts, Theory, and Methods. John Wiley & Sons, Inc., 1998. 413~417.
[18] Pedrycz W. Conditional fuzzy c-means. Pattern Recognition Letters, 1996,17(6):625~632.
[19] Barni M, Cappellini V, Mecocci A. Comments on 'A possibilistic approach to clustering'. IEEE Trans. on Fuzzy Systems, 1996,4(3):393~396.
[20] Krishnapuram R, Keller JM. A possibilistic approach to clustering. IEEE Trans. on Fuzzy Systems, 1993,1(2):98~110.
[21] Schneider A. Weighted possibilistic clustering algorithms. In: Proc. of the 9th IEEE Int'l Conf. on Fuzzy Systems. Texas：IEEE, 2000,1:176 ~180.
[22] Krishnapuram R, Keller JM. The possiblistic means algorithms: Insights and recommendation. IEEE Trans. on Fuzzy Systems, 1996,4(3):98~110.
[23] Karayiannis NB. Generalized fuzzy c-means algorithms. In: Proc. of the 5th IEEE Int'l Conf. on Fuzzy Systems, Vol 2. New Orleans: IEEE, 1996. 1036~1042.
[24] Pal RN, Pal K, Bezedek JC. A mixed c-means clustering model. In: Proc. of the 6th IEEE Int'l Conf. on Fuzzy Systems, Vol 1. Barcelona：IEEE, 1997. 11~21.
[25] Liu SH, Lin JS. Vector quantization in DCT domain using fuzzy possibilistic c-means based on penalized and compensated constraints. Pattern Recognition, 2002,35(10):2201~2211.
[26] Li RP, Mukaidon M. A maximum entropy approach to fuzzy clustering. In: Proc. of the 4th IEEE Int'l Conf. on Fuzzy System. Yokohama: IEEE, 1995. 2227~2232.
[27] Karayiannis NB. MECA: Maximum entropy clustering algorithm. IEEE World Congress on Computational Intelligence, Vol 1. In: Proc. of the 3rd IEEE Conf. on Fuzzy Systems, Vol 2. Orlando: IEEE, 1994. 630~635.
[28] Ichihashi H, Miyagishi K, Honda K. Fuzzy C-means clustering with regularization by K-L information. In: Proc. of the 10th IEEE Int'l Conf. on Fuzzy Systems, Vol 2. Melbourne: IEEE, 2001. 924~927.
[29] Yasuda M, Furuhashi T, Matsuzaki M, Okuma S. A study on statistical mechanical characeteristics of fuzzy clustering. In: Proc. of IEEE Int'l Conf. on Systems, Man and Cybemetics, Vol 4. Tucson: IEEE, 2001. 2415~2420.
[30] Wei C, Fahn C. The multisynapse neural network and its application to fuzzy clustering. IEEE Trans. on Neural Networks, 2002, 13(3):600~618.
[31] Rose K. Deterministic annealing for clustering, compression, classification, regression and Related optimization problems. Proc. of the IEEE, 1998,86(11):2210~2239.
[32] Rose K, Gurewitz E, Fox GC. Constrained clustering as optimization method. IEEE Trans. on Pattern Analysis and Machine Intelligence, 1993,15(8):785~794.

[33] Yasuda M, Furnuhashi T, Okuma S. Fuzzy clustering using deterministic annealing method and its statistical mechanical characteristics. In: Proc. of the 10th IEEE Int'l Conf. on Fuzzy Systems, Vol 2. Melbourne: IEEE. 2001. 797~800.

[34] Ozdemir D, Akarun L. Fuzzy algorithms for combined quantization and dithering. IEEE Trans. on Image Processing, 2001,10(6): 923~931.

[35] Timm H, Kruse R. A modification to improve possibilistic fuzzy cluster analysis. In: Proc. of the 2002 IEEE Int'l Conf. on Fuzzy Systems, Vol 2. Honululu: IEEE, 2002. 1460~1465.

[36] Timm H, Borgelt C, Dorring C, Kruse R. Fuzzy cluster analysis with cluster repulsion. In: Proc. of the Eruopean Symp. on Intelligent Technologies, Tenerife. 2001. CD-ROM

[37] Yu J, Yang MS. A study on generalized FCM. RSFDGrC Lecture Notes in Computer Science, 2003,2639:390~393.

[38] Yu J, Huang HK. A new weighting fuzzy c-means algorithms. In: Proc. of the 12th IEEE Int'l Conf. on Fuzzy Systems, Vol 2. St. Louis: IEEE, 2003. 896~901.

[39] Yu J. On the Fuzziness index of the FCM algorithm. Chinese Journal of Computers, 2003,26(8):974~981 (in Chinese with English abstract).

[40] Yu J, Cheng QS. A note on the weighting exponent *m* of FCM algorithm. Acta Electronica Sinica, 2003,31(3):478~480 (in Chinese with English abstract).

[41] Gao XB, Xie WX. Advances in theory and application of fuzzy clustering. Chinese Science Bulletin, 1999,44(22):961~969 (in Chinese with English abstract).

[42] Jain AK, Mutty MN, Flynn PJ. Data clustering: A review. ACM Computing Surveys, 1999,31(3):264~323.

[43] Yu J. Cluster validity and its application [Ph.D. Thesis]. Beijing: School of Mathematic Science, Peking University, 2000 (in Chinese with English abstract).

[44] Selim SZ, Ismail MA. K-means-type algorithms: A generalized convergence theorem and characterization of local optimality. IEEE Trans. on PAMI, 1984,6(1):81~86.

[45] Bezdek JC, Hathaway RJ, Sabin MJ, Tucker W. Convergence theory for fuzzy c-means: Counter-examples and repairs. IEEE Trans. on Systems, Man, and Cybernetics-part B: Cybernetics, 1987,17(5):873~877.

[46] Bezedek JC. A convergence theorem for the fuzzy ISODATA clustering algorithms. IEEE Trans. on PAMI, 1980,2(1):1~7.

[47] Yu J, Shi HB, Huang HK. Counterexamples to convergence theorem of maximum entropy clustering algorithm. Science in China, Series F, 2003,46(5):321~326.

[48] Miyamoto S, Agusta Y. Algorithms for L_1 and L_p fuzzy c-means and their convergence. In: Proc. of the IFCS'96. 1996. 295~302.

[49] Yu J, Huang HK, Tian SF. An efficient optimality test for the fuzzy c-means algorithm. In: Proc. of the 2002 IEEE Int'l Conf. on Fuzzy Systems. Honululu: IEEE, 2002,1(2):98~103.

[50] Wei W, Mendel JM. Optimality tests for the fuzzy c-means algorithm. Pattern Recognition, 1994,27(11):1567~1573.

[51] Selim SZ, Ismail MA. On the local optimality of the fuzzy ISODATA clustering algorithm. IEEE Trans. on PAMI, 1996,8(2): 284~288.

[52] Jajuga K. L_1 norm based fuzzy clustering. Fuzzy Sets and Systems, 1991,39(1):43~50.

[53] Dave RN, Krishnapuram R. Robust clustering methods: A unified view. IEEE Trans. on Fuzzy Systems, 1997,5(2):270~293.

[54] Yu J, Cheng QS, Huang HK. Analysis of the weighting exponent in the FCM. IEEE Transactions on Systems, Man and Cybernetics-part B: Cybernetics, 2004, 34(1):634~639.

[55] Yu J. General c-means clustering model and its applications, In: Proc. of the 2003 IEEE Computer Society Conference on CVPR, Vol 2. Madison: IEEE, 2003. 122~127.

[56] Zhang JS, Leung YW. Improved possibilistic c-means clustering algorithms. IEEE Trans. on Fuzzy Systems, 2004,12(2):209~217.

附中文参考文献:

[39] 于剑,论模糊 C 均值算法的模糊指标.计算机学报,2003,26(8):974~981.

[40] 于剑,程乾生.关于 FCM 算法中的权重指数 *m* 的一点注记.电子学报,2003,31(3):478~480.

[41] 高新波,谢维新.模糊聚类理论发展及应用的研究发展.科学通报,1999,44(22):961~969.

[43] 于剑.聚类有效性及其应用[博士学位论文].北京:北京大学数学科学学院,2000.