文章编号: 1672-2892(2011)06-0754-05

区间二型模糊C均值聚类在图像分割中的应用

邱存勇,肖 建

(西南交通大学 电气工程学院, 四川 成都 610031)

摘 要:聚类分析是非监督模式识别的重要分支,模糊 C 均值聚类算法(FCM)是其中的一类经典算法,然而该算法以一型模糊集为基础,无法处理数据集以及算法中的不确定性,为此引入区间二型模糊 C 均值聚类算法(IT2FCM)。二型模糊集处理不确定性的能力强于一型模糊集,基于二型模糊集的 IT2FCM 在处理不确定性时效果优于 FCM 算法。文章以图像分割为应用对象,比较IT2FCM 和 FCM 算法的分割效果,实验证明 IT2FCM 较传统 FCM 有更好的抗噪性。

关键词:区间二型模糊集;均值聚类算法;图像分割

中图分类号: TN919.8; TP391.4 文献标识码: A

An Interval Type-2 Fuzzy C-Means algorithm for image segmentation

QIU Cun-yong, XIAO Jian

(School of Electrical Engineering, Southwest Jiaotong University, Chengdu Sichuan 610031, China)

Abstract: Cluster analysis is an important branch of non-supervision pattern recognition, and Fuzzy C-Means(FCM) algorithm is a classic algorithm in cluster analysis. However, FCM is founded with Type-1 fuzzy sets, which can not handle the uncertainties existing in data and algorithm itself. This paper introduces the Interval Type-2 Fuzzy C-Means(IT2FCM) algorithm, whose core is type-2 fuzzy set that has better performance on handling uncertainties than Type-1 fuzzy set. IT2FCM and FCM are used for image segmentation to compare their segmentation results. The experiment shows that IT2FCM has better performance on suppressing noise and better effects on segmenting images compared with FCM.

Key words: Interval Type-2 Fuzzy set; Fuzzy C-Means algorithm; image segmentation

一型模糊集采用确定的数值作为隶属度以描述模糊集的不确定性,传统的模糊系统基于一型模糊集来构造, 运用相应的模糊逻辑推理和精确化来实现特定系统功能,但一型模糊集在处理实际对象的不确定性时有局限性。 增强系统方法的模糊性可以提高系统处理不确定性的能力,Zadeh 在提出模糊集的概念时曾提及集合的模糊性程 度问题,对一型模糊集合扩展,给出集合中隶属度值的模糊程度,从而使描述的集合有更强的模糊性,这种扩展 的模糊集称为二型模糊集。Mendel J M 在二型模糊集的基础上提出完整的二型模糊系统方法,使二型模糊系统 有更广泛的应用前景^[1]。模糊聚类是模糊理论的重要研究方向,在图像分割中已证明拥有良好的分割效果,其中 模糊 C 均值算法(FCM)是模糊聚类中应用最广泛的方法^[2]。传统 FCM 算法是基于像素分类的图像分割方法,利 用图像的灰度信息进行分类以达到分割效果。然而实际应用中,由于图像存在着各种类型的不确定性,FCM 算 法本身参数的选择也存在不确定性,传统 FCM 算法以一型模糊理论为基础,并不能很好地处理这些不确定性问 题,因此本文引入文献[3]提出的区间二型模糊 C 均值聚类算法 IT2FCM,并将该算法用于图像分割中,算法根 据 FCM 算法参数 m 的选值构建上、下隶属度函数, 解决 m 参数选择时的不确定性问题,实验证明算法较传统 FCM 有更好的图像分割效果。

1 标准 FCM 算法

1.1 标准 FCM 算法

FCM 算法是基于目标函数的模糊聚类方法,即把聚类归结成带约束的非线性规划问题,通过优化求

对应的模糊隶属度矩阵。FCM 算法的目标函数定义为:

$$\begin{cases} J_m(U,V) = \sum_{k=1}^n \sum_{i=1}^\infty (\mu_{ik})^m (d_{ik})^2, & m \in [1,\infty) \\ s.t. \ U \in M_{fc} \end{cases}$$
(1)

$$M_{\rm fc} = \left\{ U \in \mathbb{R}^{cn} \left| \boldsymbol{\mu}_{ik} \in [0,1], \forall_{i,k}; \sum_{i=1}^{c} \boldsymbol{\mu}_{ik} = 1, \forall_{k}; 0 < \sum_{k=1}^{n} \boldsymbol{\mu}_{ik} < n, \forall_{i} \right\}$$
(2)

式中: V表示样本的聚类中心; $d_{ik} = ||x_k - v_i||^2$ 表示第 k个样本到第 i个聚类中心的距离; m称为模糊加权指数, 又称作平滑参数,表示模糊聚类算法的模糊化程度。FCM 的具体步骤如下:

- 1) 给定聚类类别数 c 及参数 m, $2 \le c \le n$, n 是数据总数,设定迭代停止阈值,迭代计数器 l=0;
- 2) 初始化模糊划分矩阵 U, U⁽¹⁾=[µ_{ik}(i, j)⁽¹⁾];
- 3) 由式(3)计算或更新聚类中心 V⁽¹⁾;

$$v_{i}^{(l)} = \sum_{k=1}^{n} (\boldsymbol{\mu}_{ik}^{(l)})^{m} \cdot \boldsymbol{x}_{k} / \sum_{k=1}^{n} (\boldsymbol{\mu}_{ik}^{(l)})^{m}$$
(3)

4) 由式(4)更新模糊划分矩阵 U(l+1);

$$\boldsymbol{\mu}_{ik}^{(l+1)} = 1 / \sum_{j=1}^{c} \left(d_{ik}^{(l)} / d_{jk}^{(l)} \right)^{2/(m-1)} \tag{4}$$

5) 满足 $\|V^{(l+1)} - V^{(l)}\| < \varepsilon \to \infty$ 停止迭代, 否则 l=l+1 转第 3)步。

1.2 m 取值的不确定性

模糊加权指数 m 是 FCM 聚类算法的重要参数,它影响到模糊聚类结果的模糊程度。m 对聚类的影响如图 1 所示。C₁,C₂分别是以 v₁,v₂为聚类中心的划分类,竖线表示 v₁,v₂之间的垂直平分线,也称决策边界,位于边界左 侧(右侧)的数据点归于类 C₁(C₂),而位于边界上的数据点对于类 C₁,C₂有相同的隶属度。图 1(a)所示 m=1,决策 边界为 1 条直线,当受到噪声等不确定因素的影响时,边界附近的数据点很容易产生误划分的情况。图 1(b)为 m 理想取值的情况,图示阴影部分为决策边界,边界的扩展可将不确定数据点归入决策边界中,使其拥有相同的隶 属度,以减小更新聚类中心时不确定因素的影响。而图 1(c)所示 m,决策边界充满全域,模糊化程度达到最大,FCM 算法失去划分特性,此时 U= 1/c。



从图 1 可以看出, *C*₁,*C*₂ 是容量相等的 2 个划分类,即数据集被均分为 2 类。然而实际聚类时很难出现容量 相同的分类情况,如图 2 所示。当 *C*₁,*C*₂容量不同时,*m* 无法调节出理想的决策边界。若能在算法中采用不同的 *m* 值,如图 3 所示,针对不同划分类给定不同的决策边界,必然对模糊聚类的准确性有很大的提升。

2 区间二型模糊聚类算法

以图 1 所示为例,取 v_1 =-1, v_2 =1,数据集 $X \in [-1,1]$,由式(3)、式(4)可得模糊划分矩阵 U,其中数据点到聚 类中心 v_1 的隶属度函数在不同 m下的取值如图 4 所示。IT2FCM 算法采用 2 个不同的 m 取值,以获得 m 变化的不确定域,如图 5 阴影部分。



IT2FCM 算法取 2 个不同 m 值得出上、下隶属度函数,其步骤与 FCM 算法类似:

(

1) 给定聚类类别数 *c* 及参数 m_1, m_2 , 2 \leq *c* \leq *n*, *n* 是数据总数,设定迭代停止阈值 ε ,迭代计数器 *l*=0; 2) 初始化聚类中心 *V*^(*l*);

3) 田式(5)计异或史新上、下模糊划分矩阵
$$U^{\circ}, \underline{U}^{\circ}$$
:

$$\bar{\boldsymbol{\mu}}_{ik}^{(l)} = \begin{cases} 1 / \sum_{j=1}^{c} (d_{ik}^{(l)} / d_{jk}^{(l)})^{2/(m_{1}-1)}, & 1 / \sum_{j=1}^{c} (d_{ik}^{(l)} / d_{jk}^{(l)}) < 1/c \\ 1 / \sum_{j=1}^{c} (d_{ik}^{(l)} / d_{jk}^{(l)})^{2/(m_{2}-1)}, & \text{otherwise} \end{cases}$$

$$\underline{\boldsymbol{\mu}}_{ik}^{(l)} = \begin{cases} 1 / \sum_{j=1}^{c} (d_{ik}^{(l)} / d_{jk}^{(l)})^{2/(m_{2}-1)}, & 1 / \sum_{j=1}^{c} (d_{ik}^{(l)} / d_{jk}^{(l)}) < 1/c \\ 1 / \sum_{j=1}^{c} (d_{ik}^{(l)} / d_{jk}^{(l)})^{2/(m_{1}-1)}, & \text{otherwise} \end{cases}$$

$$\tag{5}$$

更新聚类中心 $V^{(l+1)}$ 。与 FCM 算法不同, IT2FCM 算法的上、下模糊划分矩阵在更新时将得到[v_L, v_R]形式的 区间值聚类中心。求边界 v_L 时,为确保 v_L 取得聚类中心最小值,更新过程中,对数据 $x_k < V^{(l)}$,设对应的隶属度 值为 $\bar{\mu}_{i_k}$,反之则取 $\underline{\mu}_{i_k}$,由式(6)的质心降型方法(Centroid type-reduction)求得 v_L 。

$$\boldsymbol{v}_i = \sum_{k=1}^n \boldsymbol{x}_k \boldsymbol{\mu}_{ik} / \sum_{k=1}^n \boldsymbol{\mu}_{ik}$$
(6)

同理得到最大值 v_R , 解模糊 $V^{(l+1)}=(v_L+v_R)/2$;

- 满足 ||V^(l+1) V^(l)|| < ε 停止迭代, 否则 l=l+1 转第 3)步;
- 5) 对区间二型模糊划分矩阵降型,即 U=(Ū⁽¹⁾ + U⁽¹⁾)/2,最后按最大值解模糊法得到聚类结果。

3 图像分割评价方法

目前图像分割算法并没有一种适用于所有图像的通用算法,本文提及的 FCM 算法亦是如此。由于算法的不通用性,在评判算法时也需要采用不同的评判标准。FCM 算法以及 IT2FCM 算法对灰度图像有较好的分割效果,针对这 2 种算法,本文从 2 个方面对其进行评价。

1) 算法性能刻画

模糊聚类算法是基于模糊划分矩阵的聚类算法,因此划分矩阵的划分系数 Vpc 和划分熵 Vpe是重要的评判指

标^[4],其定义为:

$$V_{\rm pc}(U,c) = \frac{1}{n} \sum_{i=1}^{c} \sum_{k=1}^{n} \mu_{ik}^2$$
(7)

$$V_{\rm pe}(\boldsymbol{U}) = -\frac{1}{n} \left\{ \sum_{i=1}^{c} \sum_{k=1}^{n} \left[\boldsymbol{\mu}_{ik} \log \boldsymbol{\mu}_{ik} \right] \right\}$$
(8)

划分系数越大,划分熵越小,聚类效果越好。然而划分系数与划分熵仅考虑划分矩阵的值,忽略了数据特性,为此,文献[5-6]提出了加入数据特性的评价函数 *V*_{fs}和 *V*_{xb},其函数值越小时聚类效果越好。

$$V_{\rm fs}(U,V;X) = \sum_{i=1}^{c} \sum_{k=1}^{n} \mu_{ik}^{m} \left(\left\| x_{k} - v_{i} \right\|^{2} - \left\| v_{i} - \overline{v} \right\|^{2} \right), \quad \overline{v} = \frac{1}{c} \sum_{i=1}^{c} v_{i}$$
(9)

$$V_{\rm xb}(U) = \frac{\sum_{i=1}^{c} \sum_{k=1}^{n} \mu_{ik}^{2} \|x_{k} - v_{i}\|^{2}}{n\left(\min_{i \neq j} \left\{ \|v_{i} - v_{j}\|^{2} \right\} \right)}$$
(10)

2) 图像分割质量

对图像分割最直观的评价是其错分像素点的个数,但是实际图像难以确定理想的分割效果,无法对比找出 错分的像素点。本文对人工合成图像进行分割,以求出不同算法分割时的错分率 η(错分率=错分点个数/总像 素个数)。

图像分割将图像分成若干内部具有相似特性的区域,可用分割图中各区域内部特性均匀的程度来描述分割图像的质量。由此引入评判函数区域内部均匀性测度(Uniformity Measure, UM)。

$$UM = 1 - \frac{1}{N} \sum_{i=1}^{c} \left\{ \sum_{k \in R_i} \left[x(k) - \frac{1}{A_i} \sum_{k \in R_i} x(k) \right]^2 \right\}$$
(11)

式中: R_i表示划分的第 i 个区域; A_i为该区域像素个数; N为归一化系数,本文中取 N 值为总像素个数。

4 实验与分析

为验证 IT2FCM 算法在图像分割中的有效性,本文分别用 1 张人工合成图像以及实际图像进行测试,并比较 了标准 FCM 算法与 IT2FCM 算法的图像分割效果。由于模糊聚类算法本身无法获取分割的类别数,本文采用读 入图像直方图信息的方法获取聚类数 c 以及初始化聚类中心 V。FCM 算法参数 m 取 2, IT2FCM 算法选择 m₁=1.5,m₂=2.5(不同图像由于各划分类容量不同,相应的参数 m 取值也会不同,当取 m₁=m₂ 时 IT2FCM 算法将还 原为对应 m 值的 FCM 算法)^[7]。

为获取算法的错分率,构造了1幅180×180大小的图像,3种像素灰度值分别为25,125,225,并叠加1个独立的高斯白噪声 N(0,0.006),2种算法的分割效果见图6,评价参数对比见表1。



第 2 幅图像选用图像处理中常用的图像 Cameraman,模糊聚类算法以灰度值为分类标准,因此图像按灰度 不同分成人物、草地、天空 3 类^[8]。为方便观察分割效果,文中对分割后 3 类赋予不同灰度值。分割结果见图 7。 由对人造图像的分割比较可知,当图像加入噪声后,IT2FCM 算法的分割效果明显优于 FCM 算法,而在对 实际图像的处理中,IT2FCM 分割图像的区域均匀性较之 FCM 算法更好。由各评价函数也可看出 IT2FCM 的分 割效果优于 FCM 算法。



Fig.7 Segmentation of the image Cameraman 图 7 Cameraman 图像分割结果

5 结论

本文采用 IT2FCM 算法进行图像分割,以解决传统 FCM 算法无法处理的不确定性问题,相比于 FCM 算法, 参数 m 采用精确值, IT2FCM 引出 m 的不确定域,更好地解决了 m 取值的不确定性,分割结果证明 IT2FCM 对 噪声等不确定性的处理效果优于 FCM 算法。模糊聚类算法在图像分割时仅考虑图像像素点的灰度信息,忽略了 图像像素间的结构关联。为此众多研究对 FCM 算法做出改进,加入图像空间信息,经证明改进后的 FCM 算法 有更好的分割效果^[9]。由于 IT2FCM 算法与 FCM 算法的相似性,FCM 算法的改进方法可以引入到 IT2FCM 算法 中,相信改进的 IT2FCM 算法将会得到更好的分割效果,在后续工作中将作进一步研究。

参考文献:

- [1] Liang Q, Mendel J M. Type-2 Fuzzy Logic Systems Theory Design[J]. IEEE Transactions on Fuzzy Systems, 2000,8(5):535-550.
- [2] Bezdek J C. Pattern Recognition with Fuzzy Objective Function Algorithms[M]. New York: Plenum Press, 1981.
- [3] Cheul Hwang, Frank Chung-Hoon Rhee. Uncertain Fuzzy Clustering: Interval Type-2 Fuzzy Approach to C-Means[J]. IEEE Transactions on Fuzzy Systems, 2007,15(1):107-120.
- [4] Bezdek J C. Cluster Validity with Fuzzy Sets[J]. Journal of Cybernetics, 1974,8(3):58-73.
- [5] Fukuyama Y,Sugeno M. A New Method of Choosing the Number of Clusters for the Fuzzy C-Means Method[C]// Proceedings of 5th Fuzzy Systems Symposium. Japanese:[s.n.], 1989:247-250.
- [6] Xie X L,Beni G. A Validity Measure for Fuzzy Clustering[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1991, 13(8):841–847.
- [7] 宫改云,高新波,伍忠东. FCM 聚类算法中模糊加权指数 m 的优选方法[J]. 模糊系统与数学, 2005, 19(1):143-148.
- [8] 陶剑锋,殷志祥,廖光洪. 基于模糊中值滤波的图像处理方法[J]. 信息与电子工程, 2007,5(5):391-394. (TAO Jianfeng, YIN Zhixiang,LIAO Guanghong. Research of fuzzy medium filtering about image processing[J]. Information and Electronic Engineering, 2005,5(5):391-394.)
- [9] Balafar M A, Ramli A R, Mashohor S, et al. Compare different spatial based Fuzzy-C Mean extensions for MRI Image Segmentation[C]// The 2nd International Conference on Computer and Automation Engineering. Singapore:[s.n.], 2010:609–611.

作者简介:



邱存勇(1986-),男,成都市人,在读博士研 究生,主要研究方向为二型模糊系统、模糊聚 类.email:312655942@163.com. **肖** 建(1950-),男,教授,湖南省衡阳市人, 博士生导师,研究方向为计算机控制系统、鲁棒 控制、电传动系统控制技术等.