

改进的快速 FCM 图像分割算法^①

Modified Fast Fuzzy C-Means Algorithm For Image Segmentation

郭荣传 叶水生 闵 泉 石海霞 (南昌航空大学 计算机学院 江西 南昌 330063)

摘 要: 针对模糊 C 均值(FCM)聚类算法在图像分割中存在的对初始类中心敏感且迭代过程中计算量大的问题,提出了一种改进的算法。先通过精简数据集,减少算法迭代的时间;再使用样本密度法得到 FCM 分割算法的初始聚类中心,以减少算法收敛所需的迭代次数。实验结果表明,改进后的分割算法较好地解决了类中心的初始化问题,提高了算法的收敛速度和运行速度。

关键词: 样本密度 模糊 C 均值聚类 数据集精简 初始化 图像分割

1 引言

图像分割^[1]是把图像分成各具特性的区域并提取出感兴趣目标的技术和过程,这里的特性可以是像素的灰度、颜色、纹理等。图像分割是模式识别、图像理解、计算机视觉等领域的最重要基础环节。近年来,随着数学理论特别是应用数学理论的飞速发展,人们借助新的数学理论,对图像分割问题进行深入研究,并陆续提出了多种图像分割方法,包括基于直方图阈值的分割方法、基于区域增长的分割方法、基于边缘检测的分割方法、基于模糊聚类的分割方法和基于统计学的分割方法等^[2]。

模糊聚类图像分割算法无需训练样本,是一种无监督的统计方法,它通过迭代执行分类算法来提取各类的特征值。其中应用最为广泛的是模糊 C 均值聚类(FCM)方法,其具有符合人类认知特性、描述简洁明晰、易于实现等特点,但该方法也存在诸多不足,比如聚类数目难以确定,如何确定有效的初始聚类中心和初始隶属度矩阵,迭代容易陷入局部极值,对大数据集聚类时的运算开销太大等等。为了解决 FCM 图像分割方法所存在的问题,许多人对算法进行改进,并提出了新的算法。文献[3]使用图像分块策略来提高算法的分割速度;而文献[4,5]通过对数据集的特征值的量化、合并、聚合,使数据集压缩、减少,从而把对数据集的聚类转化为对特征集的聚类。目前,多数改进的 FCM 分割算法能降低数据集的规模,但很少关注对压缩后的数据集聚

类时如何选取初始聚类中心的问题。

本文继承了数据压缩的思想,通过精简数据集,压缩参与迭代运算的数据量,减少每次迭代过程的运算时间;在此基础上,提出了一种对压缩后的数据集获取初始聚类中心的方法。改进的 FCM 聚类分割算法较好地解决了聚类初始值问题,减少了算法的迭代次数和运行时间,且不影响算法的分割效果。

2 FCM 聚类分割算法

FCM 算法用于图像分割就是根据图像中像素和 C 个聚类中心的每一个中心间的加权隶属度,对目标函数进行迭代优化,得到聚类目标函数 J 最小时的模糊划分矩阵 U 和类中心矩阵 P。目标函数为:

$$J_m(U, P) = \sum_{i=1}^c \sum_{k=1}^n \mu_{ik}^m (d_{ik})^2 \quad (1)$$

其中, U 矩阵的元素 μ_{ik} 表示第 k 个样本属于第 i 类的隶属度,并且满足 $\sum_{i=1}^c \mu_{ik} = 1$; $(d_{ik})^2 = \|x_k - p_i\|^2$; $1 \leq k \leq n$, d_{ik} 表示样本 x_k 到聚类中心 p_i 的距离;

$P = (p_1, p_2, \dots, p_c)$ 表示各类的聚类中心; m 为模糊指数,取值范围为 $1 \leq m \leq \infty$,为使上述目标函数取得最小值,经数学推导得出满足条件的 U 和 P 的迭

^① 基金项目:江西省自然科学基金(0511072)
收稿时间:2008-10-22

代公式如下:

$$p_i = \frac{\sum_{k=1}^n u_{ik}^m x_k}{\sum_{k=1}^n u_{ik}^m}, 1 \leq i \leq c \quad (2)$$

$$u_{ik} = \frac{1}{\sum_{j=1}^c \left(\frac{d_{ik}}{d_{jk}} \right)^{2/(m-1)}} \quad (3)$$

算法的主要实现步骤如下:

① 确定聚类类别数 c 和加权指数 m , 取 d_{ik} 为欧氏距离, 设定迭代停止阈值 ε 为一小正数, 初始化迭代次数 $l=0$ 和模糊分类矩阵 $U^{(0)}$;

② 根据式(3)计算或更新模糊划分矩阵 $U^{(l)}$;

③ 根据式(2), 利用 $U^{(l)}$ 更新 $p^{(l)}$, 得到新的模糊分类矩阵 $p^{(l+1)}$;

④ 若 $\|p^{(l+1)} - p^{(l)}\| \leq \varepsilon$, 迭代停止, 否则置 $l=l+1$, 返回(2), 继续迭代。

待以上 FCM 迭代优化算法收敛以后, 为了进行图像的聚类分割, 使模糊阈值选择描述适合于目标识别任务, 需要将模糊阈值描述去模糊。我们采用最大隶属函数法去模糊, 用表示第 k 个样本点所属的类别, 则有 $c_k = \arg_i \{\max_i (\mu_{ik})\} \forall i, \forall k$ 。

3 FCM聚类分割算法的改进

由于在 FCM 算法中, 每一步迭代都要对整个数据集进行计算, 运算量很大。通过数据集精简, 可减少参与计算的数据, 降低运算量, 加速运算时间。同时, FCM 算法对初始聚类中心十分敏感, 但其初始类中心又是随机选取的, 如果能选择与实际类中心接近的初始类中心, 将减少算法的迭代次数, 缩减聚类时间, 并很快收敛于实际的类中心。综合这两个方面的特点, 本文提出一种改进的 FCM 分割算法。

3.1 精简数据集

在 FCM 聚类分割算法中, 对大数据集 X 进行迭代运算时, 计算量比较大, 算法运行时间过长。如一幅大小为 256×256 的灰度图像共有像素 65536 个, 而灰度等级只有 256 个, 因而有大量像素具有相同的灰度值, 即数据集中存在极大的冗余性。所以, 可以通过数据精简, 使得具有相同特征的向量属于同一个类别, 即对样本的特征向量进行压缩, 使聚类的不同

的样本数目从 n 减少到 $q(n \gg q)$ 。

设数据集 X 中有 q 种特征的向量, 则可以得到新的数据集 $X' = \{x_1', x_2', \dots, x_q'\}$, 每种特征对应的向量个数为 $H = [h_1, h_2, \dots, h_q]$, 即 h_i 为特征 x_i 所对应的特征向量的数目。数据集 X 精简为 X' 以后, 采用随机初始化聚类中心, 用 FCM 算法对数据集 X' 聚类。这种改进的算法记为 DFCM^[6]。

3.2 初始聚类中心的选取

FCM 算法使用随机函数对类中心初始化, 由于初始化的随意性, 使 FCM 算法需要很长的迭代过程才能收敛。所以当所选取的初始类中心非常接近实际的类中心时, FCM 的迭代次数将大大减少, 从而提高算法的收敛速度。因此, 通过样本分块策略, 本文提出了一种新的获取初始聚类中心的方法。

直观上, 类中心应处于所代表类的中心部分, 所有属于该类的样本都在其周围某一领域内。因而在空间上, 类中心所处的位置样本点分布密度较大。同时, 在样本点密度连续的范围内, 应该只有一个类中心, 否则就会出现两个类交错在一起的情况。因此, 初始类中心的选择应该满足两个条件^[7]:

① 类中心所处位置样本点密度较高;

② 类中心之间的距离应尽可能地大。

因此, 可以将样本密度 P 定义为每个灰度值所对应的像素个数。通过对原数据集进行精简, 得到数据集 X' 和特征向量 H , 若灰度图像大小为 256×256 , 灰度等级只有 256 个, 则数据集 $X' = \{0, 1, 2, \dots, 255\}$, H 中存放的是每个灰度值所对应的像素个数, 并且 $P=H$ 。C 表示聚类数目, 根据样本分块策略, 将数据集 X' 平均划分成 C 个模块, 分别从每个模块中找出一个初始聚类中心。为了防止两个具有相似像素个数且灰度值也很接近的样本都被设为聚类中心, 求后续聚类中心时需要考虑类中心之间的距离, 这里采用密度和距离的乘积作为选择的度量。样本间的距离定义为

$$W = D(x'_i, x'_{i+1}) = \sqrt{(x'_i - x'_{i+1})^2 + (P_i - P_{i+1})^2} \quad (4)$$

算法主要步骤如下:

首先, 将第一个模块中样本密度最高的样本点作为第一个初始聚类中心 ($1 \leq j < 256/C$), 即

$$Cmat_1 = \max(P_j) = \max(H_j) \quad (5)$$

其次,找到前一个聚类中心到后续模块中所有样本点的 $P \times W$ 的最大值,将其作为后续聚类中心,即

$$Cmat_i = \max(D(Cmat_{i-1}, x'_j) * P_j) \quad (6)$$

其中, $2 \leq i \leq C$, $(256/C) * (i-1) < j < (256/C) * i$, 由此得到的类中心作为整个数据集的初始聚类中心,用于 FCM 算法聚类。

3.3 改进的 FCM 聚类分割算法

本文从初始化聚类中心减少迭代次数、精简数据集节省迭代时间两个方面对 FCM 算法进行了改进。改进算法的主要思想是:先精简数据集,减少参与运算的数据量;接着对缩减后的数据集采用样本点密度的方法,得到聚类中心;最后对精简后的数据集,以得到的类中心为初始聚类中心,用 FCM 算法进行聚类。这种改进的算法记为 HFCM,其主要实现过程如下:

① 通过数据精简,将数据集 X 压缩为 X' ,得到 q 和 H ;

② 对数据集 X' 采用样本密度法,得到初始类中心;

③ 以上面的类中心,对数据集 X' 进行 FCM 算法聚类,隶属度和类中心的更新为:

$$u_{ik} = \frac{1}{\sum_{j=1}^c \left(\frac{\|x'_i - p_k\|}{\|x'_i - p_j\|} \right)^{2/(m-1)}} \quad (7)$$

$$1 \leq k \leq q$$

$$p_i = \frac{\sum_{k=1}^q h_i u_{ik}^m x'_k}{\sum_{k=1}^q h_i u_{ik}^m}, 1 \leq i \leq c \quad (8)$$

④ 去模糊化处理,采用最大隶属函数法,即 $c_k = \arg_i \{\max(\mu_{ik})\} \forall i, \forall k$ 。

4 实验结果

为检验分割算法的有效性,实验采用两组数据进行算法测试。实验环境为 Intel(R) P4、CPU2.8GHZ、512M 内存的微机,由于篇幅的限制,所有实验结果图都进行了适当的缩小,所有算法均用 Matlab7.0 编

程实现。

首先,采用标准 Lena.bmp 灰度图像进行算法测试。它的大小为 256×256 像素、灰度级为 0-255。为了测试提出的改进的算法的聚类性能,取聚类数 $C=3$,加权指数 $m=2$,用 3 种算法分别重复 5 次聚类分割实验。试验结果如表 1 所示,分割效果对比如图 1 所示。



(a)标准 Lena 图像



(b)FCM 分割结果



(c)DFCM 分割结果



(d)HFCM 分割结果

图 1 三种算法对 Lena 图像的分割结果图

表 1 三种算法的分割实验结果

算法	$J_m(U,P)$ 值	迭代次数	运行时间/s
FCM	3.9258×10^6	55	697.393
DFCM	1.8218×10^5	43	2.772
HFCM	1.8218×10^5	26	1.364

第二组实验数据采用的是标准 cameraman.bmp 灰度图像,它的大小为 256×256 像素、灰度级为 0-255。取聚类数 $C=3$,加权指数 $m=2$,用以上 3 种算法分别重复 5 次聚类分割实验,试验结果如表 2 所示,分割效果对比如图 2 所示。



(a)标准 cameraman 图像



(b)FCM 分割结果



(c)DFCM 分割结果



(d)HFCM 分割结果

图 2 三种算法对 cameraman 图像的分割结果图

表 2 三种算法的分割实验结果

算法	$J_m(U,P)$ 值	迭代次数	运行时间/s
FCM	2.8615×10^6	28	353.875
DFCM	2.2183×10^5	21	1.728
HFCM	2.2183×10^5	10	0.875

从表 1、表 2 可以看出,精简数据集以后,DFCM 算法可以有效的降低算法的运行时间,而再通过样本分块策略,使用样本密度法初始化聚类中心,获得更接近实际的类中心,大大地减少了算法的迭代次数,几乎减少了一半,同时,运行时间也减少了一半;而从图 1、图 2 的分割效果图可以看出,两种改进的算法相对标准 FCM 算法,分割效果有所改善,而本文改进的算法 HFCM 相对于 DFCM 算法,分割效果相对一

致,它们的误差平方和准则函数(U,P)的值相同。所以,实验证明,本文改进的 FCM 算法在提高了算法的运行速度和收敛速度的同时,且不改变算法的分割效果,有一定的实用价值。

5 结束

本文针对 FCM 分割算法对于初始值非常敏感、面对大数据集聚类运算量过大的缺点,通过压缩参与计算的数据集和采用样本密度法获取初始聚类中心这两个方面来优化算法,提出了改进的 FCM 算法。它在减少算法迭代次数的同时,提高了算法的运行速度,并且不影响聚类分割的效果。当然,在使用样本密度法获取初始聚类中心时,如何选取一个更好的样本点密度函数,进一步提高整个算法的运行速度,还有待进一步研究。

参考文献

- 1 章毓晋.图象分割.北京:科学出版社,2001:1-201.
- 2 Cheng HD, Jiang XH, Sun Y, et al. Color image segmentation: advances and prospects. Pattern Recognition, 2001,34(12):2259-2281.
- 3 孙艺峰,王向阳,王春花.一种新的快速模糊 C 均值聚类图像分割算法.小型微型计算机系统, 2008,29(2):320-323.
- 4 丁震,胡钟山,等.FCM 算法用于灰度图像分割的研究.电子学报,1997,25(5):39-43.
- 5 Eschrich S, Ke J W, Hall L O. Fast fuzzy clustering of infrared images. IFSA World Congress and 20th AFIPS International Conference, Joint 9th, Vancouver, BC, Canada, 2001,2:1145-1169.
- 6 陈松生,王蔚.改进的快速模糊 C-均值聚类算法.计算机工程与应用, 2007,43(10):167-190.
- 7 黎俊锋,朱锋锋.基于样本密度的 FCM 改进算法.科学与技术工程, 2007,7(4):636-638.