

基于 GFCMA 算法的医学彩色图像分割

杨帆, 湛海新, 魏冬冬

(国防科学技术大学 电子科学与工程学院, 湖南 长沙 410073)

摘要:医学图像分割是医学图像分析的关键步骤,经典的模糊C-均值聚类算法(FCM)是常用方法,但其依赖于初始聚类中心的选择,通常存在局部收敛的缺陷。通过与遗传算法(GA)结合而成的遗传模糊C-均值聚类算法(GFCMA),采用RGB颜色空间,能够得到全局最优解,并在此基础上实现了医学彩色图像分割和特定目标提取,取得良好分割效果。

关键词:医学彩色图像分割;遗传算法;模糊C-均值聚类

中图分类号:TN911.73

文献标识码:A

文章编号:1673-629X(2007)05-0005-03

Medical Color Image Segmentation Based on Genetic Fuzzy C-Means Clustering Algorithm

YANG Fan, CHEN Hai-xin, WEI Dong-dong

(Electronic Science and Engineering Institute, National University of Defense Techn., Changsha 410073, China)

Abstract: Medical images segmentation is a critical step in medical image analysis. The classic fuzzy C-means clustering algorithm (FCM) is a commonly used method, but depends on the initial cluster centers choice usually, which means the presence of local defects. Through GA, which combines genetic fuzzy C-means clustering algorithm (GFCMA) using RGB color space to be a global optimum. And on this basis it achieves a specific goal and medical color images. The division achieves good results.

Key words: medical color images segmentation; genetic algorithm; fuzzy C-means clustering

0 引言

研究表明,人眼能分辨的灰度级最多只有几十个,而可以分辨的色彩却能达到成千上万种。在利用医学图像诊断过程中,人类已经不满足于看不清细节的灰度图像了,利用彩色图像诊断已经成为一种未来的趋势。医学彩色图像的分割问题是人体组织特征提取、配准、融合及三维重建的基础,也是图像诊断的基础。目前出现的彩色图像分割方法很多,大多在某一个方面或某一个分割标准上有一定的进展,但都不具有通用性。文中针对医学彩色图像的特点,提出了采用优化的遗传和模糊C-聚类混合算法对医学彩色像的分割策略,并通过编写C++程序实验验证了此分割方法的实用性,实验结果表明,达到了很好的分割效果。

1 模糊C-聚类算法(FCM)

因为医学彩色图像的纹理、区域的边界等不确定

的因素,所以其本质上是模糊的,采用模糊方法分割图像是非常合适的。模糊C-聚类算法(Fuzzy C-Means, FCM)是其中的佼佼者之一,FCM最早是由Dunn提出,由Bezdek推广的^[1,2],是一种迭代优化算法。它通过聚类准则函数 J_m 极小化的必要条件之间的Pickard迭代来实现。FCM是一种无监督的图像分割方法,过程完全自动化,无需人工干预,并且能很好地处理噪声、部分体积影响和图像模糊^[3]。FCM非常适合医学多特征的彩色图像分割。算法主要思想为:

设待分析数据集 $X = \{x_1, x_2, \dots, x_N\}$ 分为C类,由于任一数据点 x_j 几乎不可能被严格地划分为某一类,定义它隶属于第i类的程度为 μ_{ij} , μ_{ij} 性质如下:

$$\mu_{ij} \in [0, 1]$$

$$\sum_{i=1}^C \mu_{ij} = 1 \quad \forall j$$

$$0 < \sum_{j=1}^N \mu_{ij} < N \quad \forall i$$

聚类准则函数 J_m 为:

$$J_m(\mu, w) = \sum_{j=1}^N \sum_{i=1}^C (\mu_{ij})^m \|x_j - w_i\|^2$$
$$w = \{w_1, w_2, \dots, w_C\}$$

其中 w_i 表示第i个聚类中心, $i = 1, 2, \dots, C$, $m \in [0, \infty)$

收稿日期:2006-07-14

作者简介:杨帆(1976-),男,陕西岐山人,硕士研究生,研究方向为数字图像处理和模式识别;湛海新,博士,副教授,研究方向为图像图形与模式识别、人工智能、计算机仿真。

∞] 是加权指数, 聚类准则函数表示了各类数据点到相应聚类中心的加权距离平方和。

模糊 C 均值聚类就是求使上式所示的聚类准则函数 J_m 最小的模糊划分矩阵 $\mu = (\mu_{ij})_{C \times N}$ 以及类别中心 c 。具体算法实现要做以下工作:

① 确定类别数 C , 初始化 m 及聚类中心。

② 进行第 k 次迭代, 根据以下两式计算新的隶属度函数和 C 个聚类中心。

$$\mu_{ij} = \frac{1}{\sum_{k=1}^C \left(\frac{d_{ij}}{d_{jk}} \right)^{\frac{2}{m-1}}}$$

$$i = 1, 2, \dots, C; j = 1, 2, \dots, N$$

其中 d 是样本点到聚类中心的欧氏距离。

$$w_i = \frac{\sum_{j=1}^N \mu_{ij}^m x_j}{\sum_{j=1}^N (\mu_{ij})^m} \quad i = 1, 2, \dots, C$$

③ 若 $|J^k - J^{k-1}| \leq \epsilon$ (ϵ 为一可任意小正数, 可以看作是误差精度) 则停止, 否则返回聚类准则函数, 继续迭代^[4]。

2 遗传模糊 C-聚类算法(GFCMA)

遗传算法是建立在生物遗传学机理上的一种高度并行、随机和自适应全局优化算法, 具有简单、鲁棒性和广泛适用性的特点。在寻优过程中, 采用同时处理群体中多个个体的方法, 有效地克服了 FCM 可能出现的局部最优问题和对初始聚类中心的依赖, 获取全局最优。遗传算法的主要特征有^[5]: 初始化群体、选择、交叉、变异。遗传算法具体用于 FCM 算法^[6]的过程如下:

(1) 确定聚类数目 C 。

(2) 生成初始化群体: 随机生成含有 P 个个体的群体, 每个个体代表 C 类中每个类的聚类中心 $\{w_1, w_2, \dots, w_c\}$, w_i 为 N 维向量, 即同时生成由 μ_{ij} 组成的模糊矩阵 μ 。

(3) 选择: 建立个体适应度评价标准, 把父代中的优秀个体(适应度大的)保留到子代, 淘汰适应度小的。

(4) 交叉: 用于优化组合新个体, 使子代继承父代优秀基因。设交叉概率为 P_c , 对每一个个体随机产生一个 $[0, 1]$ 的随机数 γ , 若 $\gamma < P_c$ 则进行交叉操作, 对群体随机选中一个个体, 随机生成一个在 $[1, C]$ 的整数 j , 按如下方式对位置 j 处进行交叉操作, 也就是第 j 类的中心交换。

$$\begin{aligned} \{w_1, w_2, \dots, w_c\} &\rightarrow \{w_1, w_2, \dots, w_j, w_{j+1}, \dots, w_c\} \\ \{w_1, w_2, \dots, w_c\} &\rightarrow \{w_1, w_2, \dots, w_j, w_{j+1}, \dots, w_c\} \end{aligned}$$

(5) 变异: 变异是用来保留群体的多样性。设变异概率为 P_m , 对选中的每一个个体, 对 $\{w_1, w_2, \dots, w_c\}$ 随机选定一个位置, 对 w_j 进行不重复的变异操作, 随机产生新的 w_j 。

3 优化及参数确定

根据图像分割性能评价要求, 好的分割方法应具有自动、快速、精确、自适应性和鲁棒性等特点。基于 GFCMA 的图像分割方法虽然具有了过程自动化、对噪声敏感度低、并行处理、鲁棒性和广泛实用性等优点, 但实际实现仍然存在着耗时、参数确定困难等问题。

针对医学彩色图像的特点, 主要对 GFCMA 做以下优化改进:

(1) 收敛速度。

由于反复迭代过程是非常耗时的, 运算量很大, 故收敛速度很慢, 提高收敛速度是该算法的当务之急, 笔者采取的策略是先利用遗传算法(GA)的全局寻优能力找到最优区域后, 再用 FCM 快速收敛到最优点, 从而提高算法的速度。

(2) 加权系数 m 的确定。

加权系数 m 是聚类准则函数的一个重要参数, 它控制着聚类的模糊性, 即控制着样本在模糊空间的分享程度。Pal 等人从实验中给出了 m 值的最佳选取值, 高新波等人从理论和实验中证明这一结论^[7], m 值选取 2 比较合适。

(3) 适应度评价标准。

采用定义适应度函数来评价父代中的个体是否优秀, 从而决定各种遗传操作, 因为存在着个体适应度越高, 其 $J_m(\mu, w)$ 值越小的反比关系, 故可以定义适应度函数 f 为: $f = \frac{1}{J_m(\mu, w)}$

4 实验及结果

将上述遗传模糊 C-聚类算法用于四幅医学彩色图像分割, 图 1~4 分别为非何杰金氏淋巴瘤(组织细胞型)图像、肝脏炎性病变图像、卵巢内胚瘤图像和肝癌穿刺涂片图像。图像来源于 37° 医学网医学影像图库, 图像大小为 130×130 (像素)。分割后的结果如图 1~4 所示, 取得非常不错的效果。

在利用 GFCMA 算法处理图像过程中, 群体规模设定为 60, $N = 3$ (用 RGB 颜色空间), 交叉概率 $P_c = 0.8$, 变异概率 $P_m = 0.1$, 循环次数为 100 次。图 1~3 分别使用 FCM 和 GFCMA 计算所得的最小均方误差如表 1 所示。针对图 4, 选择不同的初始聚类数, 计算

所得的最小均方误差如表 2 所示。

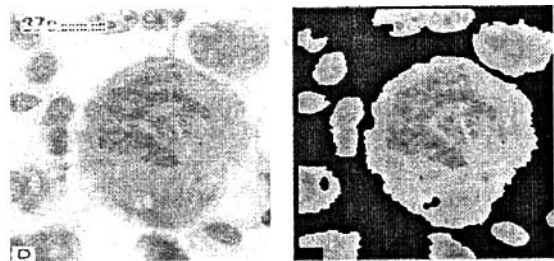


图 1 非何杰金氏淋巴瘤(组织细胞型)
原图及分割后的图像

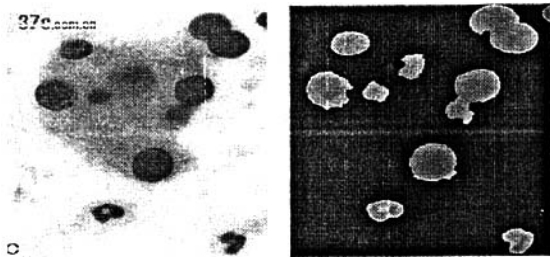


图 2 肝脏炎性病变原图及分割后的图像

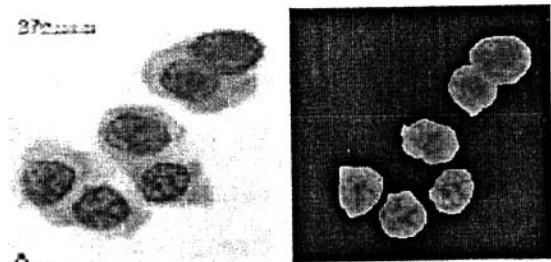


图 3 卵巢内胚瘤原图及分割后的图像



图 4 肝癌穿刺涂片原图及初始聚类数
分别为 4 和 6 时的分割后图像

表 1 三幅医学图像的 FCM 和 GFCMA 算法结果比较

图像	类别数	最小均方误差	
		FCM	GFCMA
1	3	16.386	16.147
2	3	14.534	14.221
3	3	21.783	21.365

表 2 不同初始聚类数的 FCM 和 GFCMA 算法结果比较

初始聚类数	最小均方误差	
	FCM	GFCMA
4	14.674	14.349
5	13.752	13.489
6	11.438	11.246

5 结 论

从分割后的图像来看,FCM 和 GFCMA 算法都达到了对特定目标的分割效果,从分割效果来看,GFCMA 算法明显优于 FCM 算法,因为 GFCMA 算法克服了 FCM 算法局部寻优的缺点,采用了全局寻优策略。在初始聚类数目对分割效果的影响这一问题上,聚类数目增多,分割的精细程度就越高,细节保留的就越多,但计算的复杂度会越大。

参考文献:

[1] Dunn J C. A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well - Separated Clusters[J]. Q. Cyber Net,1974(3):32 - 57.

[2] Bezdek J C. Pattern recognition with fuzzy objective function algorithms[M]. New York:[s. n.],1981.

[3] 章毓晋. 图像分割[M]. 北京:科学出版社,2001.

[4] 李洪兴,汪培庄. 模糊数学[M]. 北京:国防工业出版社,1994.

[5] 史忠值. 知识发现[M]. 北京:清华大学出版社,2002.

[6] 张 维,潘福铮. 一种基于遗传算法的模糊聚类[J]. 湖北大学学报:自然科学版,2002,24(2):101 - 104.

[7] 高新波,裴继红,谢维信. 模糊 C - 均值聚类算法中加权系数 m 的研究[J]. 电子学报,2000,28(4):80 - 83.

(上接第 4 页)

<http://www.cs.umbc.edu/kqnl/papers/kqnl-spec.ps>.

[5] Ilachinski A. Irreducible Semi - Autonomous Adaptive Combat (ISAAC): An Artificial - Life Approach to Land Warfare(U) Center for Naval Analyses[EB/OL]. 1997 - 08. <http://www.nca.orf/isaac/>.

[6] Rossetti R J F, Bordini R H. Using BDI agents to improve driver modelling in a commuter scenario[J]. Transportation Research Part C, 2002,12(3):373 - 398.

[7] 杨克巍,王正元,谭跃进. 基于 DEVS 形式化描述的半自治 Agent 建模研究[C]//CAAI - 10. 北京:北京邮电大学出版社,2003:177 - 182.