

# 基于遗传算法的聚类分析在 CT 图像分割中的应用

朱玲利 李吉桂 鲍苏苏  
(华南师范大学计算机学院 广州 510631)

**摘要** 针对 CT 医学图像的特点, 本文将遗传算法与聚类分析两种工具相结合, 应用到医学 CT 图像分割中。对 K 均值聚类做了简要分析和评论, 在此基础上将遗传算法引入聚类分析中, 利用遗传算法搜索的随机性和并行性, 从而克服了 K 均值聚类的局部性和对初始聚类中心的敏感性; 并且可以根据分割的要求, 合理地调整聚类时的特征向量和权重。试验结果表明上述方法是可行的, 达到了较好的分割效果。

**关键词** 图像分割, K 均值聚类, 遗传算法

## Application of Clustering Analysis Based on Genetic Algorithm in CT Image Segmentation

ZHU Ling-Li LI Ji-Gui BAO Su-Su

**Abstract** Aiming at the characteristic of medical images, this paper integrates genetic algorithm with clustering analysis and applies in medical CT image segmentation. K-means clustering is introduced and remarked firstly. On the basis of systematic analysis of current algorithms, genetic algorithm is inducted into clustering analysis to solve the locality and the sensitiveness of the initial condition of K-means clustering by searching randomly and in parallelism. Besides, according to different request, feature vector and weight factors are adjusted rationally. The example shows that the method is feasible, and good segmentation results have got.

**Keywords** Image segmentation, K-means clustering, Genetic algorithm

## 1 引言

医学图像分割技术是医学图像处理和体视化中的关键技术。实现医学图像的三维重建的首要前提就是对图像数据进行正确、合理的分割, 从中提取出感兴趣的器官、组织或病变体, 从而实现这些被提取出的器官、组织或病变体的三维重建, 达到辅助治疗与手术规划的目的。在医学图像的获取过程中, 由于影像设备中各电子器件的随机扰动和周围环境的影响, 使图像多少含有噪声和失真, 这就给医学图像准确、快速的分割带来了一定的难度, 影响了组织的分割与提取。利用传统的一些分割算法对图像进行分割处理时, 因图像本身复杂性和噪声的影响很难达到好的分割效果, 并且也不能实现自动分割的目的<sup>[1]</sup>。

K 均值聚类方法<sup>[2]</sup>是一种常用的分类方法, 该方法虽然能准确地对样本进行分类, 但大量的试验表明, 在样本集不理想的情况下可能导致不好的聚类效果。遗传算法<sup>[3]</sup>(GA)是目前比较理想的优化方法, 它模拟自然进化中的优胜劣汰、适者生存的原理来进行学习和寻优, 可以解决大空间、非线性、全局寻优等复杂问题。本文是将遗传算法与聚类算法这两种功能强大的工具结合起来, 从而在特征空间内对像素进行遗传聚类, 当然, 对表征图像特点的特征向量可以根据分割的要求进行调整和扩充, 并且可以通过对各个特征向量的权重的动态调整来控制各类信息在聚类时所起作用的大小, 从而达到了很好的分割效果。

## 2 遗传聚类分割的实现

### 2.1 K 均值聚类算法简介

K 均值(或者 ISODATA)聚类算法是一种最普遍的不断

迭代调整  $C$  个聚类质心的算法, 其中思想就是最小化总的类内距离(各类中样本到对应质心的距离的总和)。其数学描述为: 设模式样本集为  $X = \{X_1, X_2, \dots, X_n\}$ ;  $C(2 \leq C \leq n)$  是要将数据样本分成的类别的数目;  $X_i$  为  $d$  维模式向量, 聚类问题就是要找到一个划分  $C = \{C_1, C_2, \dots, C_m\}$ , 满足:

$$\begin{aligned} X &= \bigcup_{i=1}^m C_i \\ C_i &\neq \Phi (i=1, 2, \dots, m) \\ C_i \cap C_j &= \Phi, (i, j=1, 2, \dots, m; i \neq j) \end{aligned}$$

并且使得类内距离和:

$$E = \sum_{j=1}^C \sum_{X_i \in C_j} \|x_i - m_j\|^2 \text{ 达到最小} \quad (1)$$

该算法执行的步骤: 先初始化  $C$  个聚类中心, 然后将数据集中的每一个  $X_i$  分配到与之相距最近的那个质心  $M_j$  代表的聚类中去, 计算分配后新的质心  $M_j$  (常用聚类均值点计算) 以及  $E$ , 这样不断迭代, 当算法收敛时, 理论上就得到了各类的聚类中心, 从而完成了聚类划分。尽管 K 均值聚类有很高的搜索能力, 但它是一种局部搜索算法, 对聚类中心的初始值十分敏感, 如果初始选择不当, 它会收敛到局部极小值, 并且所需的计算量也很大。

### 2.2 面向 CT 图像的引入遗传算法的 K 均值聚类

由于人体组织和器官的复杂性, 要对 CT 图像进行准确目标提取, 必须有一个准确的分类, 为此我们引入遗传算法在图像特征空间内进行聚类, 达到全局优化分类的目的。

采用遗传算法进行聚类分析, 其主要步骤: 1) 个体编码; 2) 初始种群生成; 3) 适应度函数设计; 4) 遗传算子设计; 5) 控制参数设定。下面分别从这几个方面对引入遗传算法的 K 均值聚类算法进行描述。

#### 2.2.1 个体编码

遗传算法中编码机制<sup>[4]</sup>的不同对算法本身的影响很大,由于聚类的本身目的就是杂乱无序的数据按照彼此间相同或相似的程度进行“物以类聚”,因此可以根据图像灰度信息和空间位置信息进行编码,当然根据分割的目的可以对这些特征信息进行扩充(比如梯度等),为了处理的方便,本文的特征向量为5维  $S = \{R, G, B, X, Y\}$ ,各分量为像素的  $R, G, B$  值和坐标位置,它们对应的权重分别设为 0.8 和 0.2,当然可以根据需要动态的调整不同特征向量所占的权重。我们采用基于聚类中心的整数编码方案。若有  $n$  个聚类中心,则一条染色体就可以表示为:

$$R = Z_1, Z_2, \dots, Z_n$$

的形式(下面是对应的数据结构),则每条染色体即为长  $5 \times n$  的整数串,这样每条染色体就代表了一种聚类结果,通过不断的调整,最后就可以得到最优解。

```
//每个基因或聚类中心的结构
typedef struct{
    BYTE R;
    BYTE G;
    BYTE B;
    long y;
    long x;
}GENE;
GENE Pop[POPSIZE][MAXCLASS]; //存放种群的结构数组
#define POPSIZE 50 //种群规模(可以根据需要调整)
#define MAXCLASS 6 //最大类群(可以根据需要调整)
当聚类中心确定后,对样本的划分可按下面的最紧邻准则进行
若  $X_i, Z_j$  满足下式:
```

$$\|X_i - Z_j\| = \min \|X_i - Z_k\| \quad (k=1, 2, \dots, n)$$

则称  $X_i$  属于第  $j$  类。

### 2.2.2 初始化种群

随机产生特征向量范围内的向量作为初始聚类中心,从而形成第一代种群。也可以根据图像对应的灰度直方图,选出  $C$  个聚类中心,后者寻优速度更快,收敛到最优个体的时间短。一般来说,初始群体素质都很差,遗传算法的任务就是从这些群体出发,择优劣汰,最后得出非常优秀的群体就是问题的解。

### 2.2.3 适应度函数设计

每个个体以式(1)所算出的  $E$  为目标函数,  $E$  越小,个体的适应值越高,只需优化使其到达最小就好了。

### 2.2.4 遗传算子设计

(1)选择算子:求出每个个体的适应度,并将其按适应度从大到小的顺序排列,将群体中 10% 的个体直接进入交配池,剩下部分按轮盘赌的比例选择法进行选择,这样可以保证下一代的群体绝不比上一代的最佳群体差。

(2)交叉算子:选择出个体随机两两配对后,对于每一个个体以交叉概率进行交叉操作。按照传统的遗传算法,交叉是把两个染色体基因的一段进行交叉,但是考虑到图像处理特殊性,上述交叉方法的结果不太理想,为此选用窗口交叉方法进行操作<sup>[5]</sup>,就是把两个染色体基因中的多段同时进行交叉。

(3)变异算子:随着遗传算法应用的不断扩展,变异操作也越来越多的加以利用。文<sup>[5]</sup>采用在传统的变异操作基础上,再根据变异位相邻像素的平均值来最后决定该变异位是否改变的方法。一些研究表明十进制编码对于变异操作的种群稳定性比二进制编码要好<sup>[6]</sup>。在试验中,我们采用十进制编码,考虑到图像数据相邻像素相关性强的特点,如下定义变异操作:

$$g_{mn} = \frac{1}{m \times n} \sum_{m=t}^{M-t} \sum_{n=t}^{N-t} g_{ij} \quad m=t, t+1, \dots, M-t; n=t, t+1, \dots, N-t$$

$$t \quad (2)$$

上式中,  $g_{mn}$  表示进行变异操作的基因,  $m, n$  分别为该基因在染色体二维矩阵( $M \times N$ )中的行、列位置。  $T$  为常数,可以根据需要设为不同值。上式表示,当基因  $g_{mn}$  变异操作时,构造以  $g_{mn}$  为中心,以  $2t$  为边长的矩形,用该矩形内所有的平均值替换  $g_{mn}$  的值。由式(2)中的结构可知,处于染色体边缘的基因不能进行变异操作,但是考虑到分割的目标一般都处于图像中央,所以忽略这些基因对图像分割影响不大。

### 2.2.5 控制参数的设定

群体规模  $M$  影响遗传算法的最终结果及其执行效率,在试验中我们取为 50,交叉概率  $P_c$  和变异概率  $P_m$  分别取为 0.2 和 0.01,当然我们可以根据分割的对象对这些参数进行适当的调整,以达到更好的分割效果。

### 2.2.6 算法的步骤和总体算法实现流程

#### (1)算法步骤

- 1)输入控制参数:种群规模  $M$ ,交叉概率  $P_c$ ,变异概率  $P_m$ ,最大代数 MaxGen;
- 2)初始化种群:随机生成初始化种群  $G_1$ ,并计算群体中个体的适应度、群体总适应度;
- 3)用适应度函数评价父代个体,并排序,保留最优染色体;
- 4)应用遗传算子(即选择、交叉、变异)产生新一代群体  $G_t$ ;
- 5)判断是否满足收敛条件,满足则输出结果,退出;否则,转入步骤 3);

#### (2)算法实现流程

```
BOOL GeneticClustering::GeneticClusteringSg()
{int G=0;
 while(G<MaxGen)
 {Distribute(); //按最近邻准则,将像素分配到各聚类中心
 AdjustIllegalChrom(); //对非法染色体进行调整
 SortSelection(); //排序选择
 CrossOver(); //交叉操作
 Mutation(); //变异操作
 G++;
 }
 //按进化后的最优个体对图象进行分割
 return GenClusSegment(SelectBestMember());
}
```

## 3 试验结果与分析

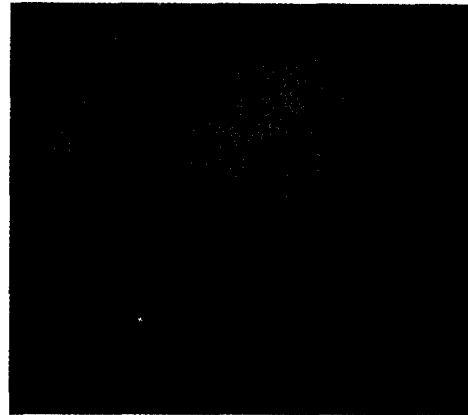
为了验证本文算法效果,取医学图像为  $512 \times 512$  的 CT 图像,如下图所示(人体肝脏 CT 扫描图像)。考虑到计算效率的因素,我们在试验中均匀取约为总像素的 25% 为样本(隔行隔列取像素),进行聚类分割,当聚类数为 6,交叉概率  $P_c = 0.2$ ,变异概率为  $P_m = 0.01$  时在不同进化代数下的分割结果如右边图所示。由试验结果可以看出它能够将肝脏内部不同物质进行清楚分类,随着进化代数的增加,我们可以看到其分割的效果是不断改善的,大约到 150 代以后分割基本上就稳定了,并且我们分别用二维最大熵法和直方图双峰法也进行了分割,从图中的比较可以看出,他们对复杂图像的分割效果不太好,本文算法的分割结果明显比较好,图像中不同的物质被合理地进行分类,并且突出出来,为后续的目标提取做好准备。可以设想将分割结果应用到医学图像的三维重建中,效果会比较好。

**结论** 本文针对医学图像的特点,设计了一种聚类分析的图像分割算法,并且将遗传算法引入聚类,利用遗传算法的并行性和随机搜索性,在一定程度上克服 K 均值聚类搜索的

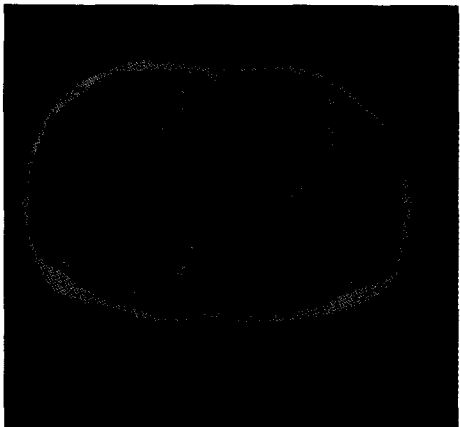
局部性和对聚类中心初值的敏感性,从而可以一次将图像分



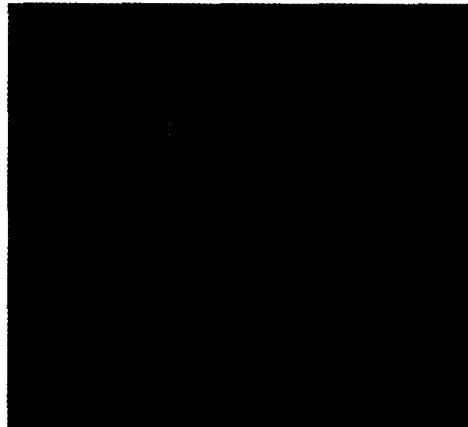
原始图像



本文算法结果 (g=80)



二维最大熵法结果



本文算法结果 (g=120)



直方图双峰法结果



本文算法结果 (g=150)

割成各个区域(多类),方便而有效。当对不同要求、不同应用背景或不同性质的图像进行分割时,只需对分割图像的特征向量进行扩充和改变,特征的权重根据图像的性质再加以自适应的调整,因此具有一定的通用性。从实现结果可以看出,分割效果较好,有助于后续的目标提取。

### 参考文献

- 1 王合. 医学图像边缘检测技术的研究[M];[学位论文]. 上海交通大学,1996
- 2 Marques J P. de Sa. Pattern Recognition Concepts, Methods and

Applications

- 3 Kristinsson K, Dumont G A. System identification and control using genetic algorithms [J]. IEEE Trans Syst Man, 1992, 22(5): 1033~1046
- 4 Goldberg D E. Genetic Algorithms in Search, Optimization & Machine learning [M]. Addison-Wesley Publishing, 1989
- 5 陈国良,等. 遗传算法及其应用[M]. 北京:人民邮电出版社, 1996
- 6 张晓绩,方浩,戴冠中. 遗传算法的编码机制研究. 信息与控制, 1999, 26(2): 134~139