

基于粗糙集与差分免疫模糊聚类算法的图像分割^{*}

马文萍, 黄媛媛, 李豪, 李晓婷, 焦李成

(智能感知与图像理解教育部重点实验室(西安电子科技大学), 陕西 西安 710071)

通讯作者: 马文萍, E-mail: wpma@mail.xidian.edu.cn

摘要: 提出了基于粗糙集模糊聚类与差分免疫克隆聚类的图像分割算法. 该算法在差分免疫克隆聚类算法的基础上, 通过引入粗糙集模糊聚类, 将差分免疫克隆聚类算法中的硬聚类变成模糊聚类, 从而获得更丰富的聚类信息. 具体来说, 由于粗糙集的优势是处理不确定的数据, 因此, 加入粗糙集模糊聚类后更有利于算法解决不确定性问题. 通过对 9 幅图像分割实验结果与 4 种算法的对比, 验证了该算法在聚类性能稳定性方面的优越性, 结果还同时证明了该算法具有更高的分割正确率和更好的分割结果.

关键词: 粗糙集; 差分免疫克隆; 图像分割

中图法分类号: TP391

中文引用格式: 马文萍, 黄媛媛, 李豪, 李晓婷, 焦李成. 基于粗糙集与差分免疫模糊聚类算法的图像分割. 软件学报, 2014, 25(11): 2675-2689. <http://www.jos.org.cn/1000-9825/4519.htm>

英文引用格式: Ma WP, Huang YY, Li H, Li XT, Jiao LC. Image segmentation based on rough set and differential immune fuzzy clustering algorithm. Ruan Jian Xue Bao/Journal of Software, 2014, 25(11): 2675-2689 (in Chinese). <http://www.jos.org.cn/1000-9825/4519.htm>

Image Segmentation Based on Rough Set and Differential Immune Fuzzy Clustering Algorithm

MA Wen-Ping, HUANG Yuan-Yuan, LI Hao, LI Xiao-Ting, JIAO Li-Cheng

(Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education of China (Xidian University), Xi'an 710071, China)

Corresponding author: MA Wen-Ping, E-mail: wpma@mail.xidian.edu.cn

Abstract: In this paper, a new method based on rough-fuzzy set and differential immune clone clustering algorithm (DICCA) for image segmentation is proposed. By replacing hard clustering with fuzzy clustering through incorporating rough-fuzzy set into DICCA, this algorithm can obtain more abundant clustering information. Specially, as the advantage of rough set is processing uncertain data, the proposed algorithm is more conducive to solve the uncertainty problem. In experiments, nine images are used for segmentation and four algorithms are chosen for comparison to validate the performance in the clustering stability. The experimental results show that the algorithm has higher segmentation accuracy and better segmentation results.

Key words: rough set; differential immune clone; image segmentation

图像领域的处理按照其处理问题和抽象程度的差异分为 3 个层次: 图像处理、图像理解和图像分析. 图像工程技术是将这 3 方面进行有机结合的技术. 图像处理是在像素级上进行的, 属于最初步的操作. 图像处理的方法包括图像降噪、图像编码和图像分割. 图像分割作为图像处理中的关键步骤, 没有图像分割, 图像理解和分析就无法进行, 然而很多方面的因素影响着图像理论的发展. 图像分割的应用也十分广泛, 例如变化检测、目标测量、目标识别等. 由于图像分割是否准备充分对后续的研究有很大影响, 所以这个问题也具有较大的研究意义.

* 基金项目: 国家自然科学基金(61203303, 61202176, 61272279)

收稿时间: 2013-03-15; 修改时间: 2013-07-09; 定稿时间: 2013-11-11

而图像分割问题实际上可以看成是一个优化问题,寻找最优的区域平滑图像来匹配原始图像.在实际操作中,通常通过一种优化方法来寻找平滑区域的灰度均值,在聚类方法中,这样的灰度均值称为聚类中心.基于这个角度,本文主要涉及优化技术和聚类两大课题.优化技术一直受到人们的广泛重视,它是以数学为基础,用于求解各种工程问题优化的应用技术.现实世界中的很多工程问题都具有一定的约束性、非线性、复杂性,且建模较为困难,因此,寻求一种有效且通用性较好的优化算法,已成为该领域一个重要的研究目标和引人注目的研究方向^[1].

20世纪80年代以来,一类新颖的优化算法——进化算法(evolutionary algorithm)在优化技术领域引起了国内外学者的广泛重视.进化算法是一类借鉴生物界自然选择和自然遗传机制的随机搜索算法,它以达尔文进化论的“物竞天择、适者生存”作为算法的进化规则,并结合孟德尔的遗传变异理论,将生物进化过程中的繁殖、变异、竞争和选择引入到算法中.进化算法的基本思想是使用人类进化和模拟生物的方法来解决复杂问题.它从初始种群出发,采用优胜劣汰、适者生存的自然法则选择相对优秀个体,并对这些个体进行交叉、变异而产生新一代种群,对种群进行逐代进化,直到满足终止条件为止^[2,3].

进化算法因适用于求解高度复杂的非线性问题而得到广泛应用,特别是使用传统优化算法无能为力的优化问题,采用进化算法往往可以获得性能非常好的解.进化算法是一种鲁棒性较好的方法,能够在不同环境下解决各种不同问题,而且多数情况下都能得到比较满意的有效解.它用某种编码方案对整个参数空间进行编码,从一个包含多个个体的初始种群开始搜索,而不是从某个单一的初始点开始搜索,因此,得到符合要求的解的概率较大.搜索过程中主要利用目标函数值的信息,不必用到与具体问题有关的特殊知识等其他信息,因而进化算法具有广泛的通用性.进化算法作为一种优秀的优化技术,虽然在各个工程领域得到了普遍应用,但在应用和发展的同时也暴露出一些亟待解决的问题,如算法的收敛性会严重影响优化结果等,因此,对进化算法进行更深入的研究具有一定的现实意义^[4].

聚类算法包括硬聚类和软聚类算法.类似于 K 均值的传统聚类方法属于硬聚类,他们对每一个对象的划分,不是属于这一类就是属于那一类,没有中间量;而有中间量的聚类方法叫做模糊聚类^[5],这个中间量我们称为模糊隶属度.在现实生活中,大多数的对象没有固定划分的界限,没有一定属于某一类.Zadeh^[6]提出的模糊集理论对于模糊聚类来说是一种有力的解决问题的方法,人们也逐渐运用模糊聚类分析的方法来处理问题.

Ruspini^[7]是最早提出模糊聚类的.在模糊聚类这个概念基础上,许多学者提出了很多聚类方法,比较典型的有:基于模糊等价关系的传递闭包方法^[8]、基于相似性和模糊关系的方法^[9]、基于模糊图论最大树的方法^[10]以及基于数据集的动态规划、凸分解和难以辨识关系等.实际中,上面所说的几种算法,由于本身所具有的限制性,不太适合在大量的数据上进行聚类,因此会很难用于对时间和实用性要求比较高的地方,其应用范围受到局限,所以对于这方面的学习研究也就较少.

模糊聚类的理论与生产实践相互推动,相互促进,共同发展.在模糊聚类的发展过程中,它在许多地方都得到了大量应用,而且得到了让人欣喜的利益和效果.它的应用范围涉及到图像处理、水质分析、食品分类、天气预报、医学诊断、参数估计、神经网络的训练、时间序列的预测、矢量量化编码中的码书设计、通信系统中的信道均衡等领域^[5].

基于上述论述,本文重点叙述一种基于粗糙集与差分免疫模糊聚类算法(rough set and differential immune fuzzy clustering algorithm,简称 RDIFC).首先介绍粗糙集与模糊聚类的理论基础;然后,对基于粗糙集与差分免疫模糊聚类算法的图像分割进行详细描述;最后,给出与该算法相关的仿真实验、结果分析以及对本项研究的总结与展望.

1 粗糙集与模糊聚类理论基础

1.1 粗糙集理论

粗糙集理论源于一个概念——近似空间——一个关系系统(或二元组).

对 (U, R) 一个组合,其中, U 是由研究对象组成的非空有限集合, U 上的一个等价关系为 R . R 具有自反性、对

称性和传递性.按照一定的规则,集合 U 被关系 R 分解成不相交的类,即,对于 x 和 y 这两个元素,当且仅当 $(x,y) \in R$ 时,判定 x 和 y 属于同一类.此处,关系 R 下 U 的商集用 U/R 表示,并有 $U/R = \{X_1, X_2, \dots, X_m\}$, 其中, X_i 是 R 的一个等价类, $i=1, 2, \dots, m$. 如果在 U 中两个元素 x 和 y 属于相同的等价类 $X_i \in U/R$, 那么可认为 x 和 y 是不可区分的.此外, R 的等价类和空集 \emptyset 是近似空间 $\langle U, R \rangle$ 中的初等集.给定一个任意集合 $X \in 2^U$, 一般来说,它不能在 $\langle U, R \rangle$ 中精确描述,但此时,它可以通过一个下近似和上近似对来描述,即

$$\underline{R}(X) = \bigcup_{X_i \subseteq X} X_i \text{ and } \bar{R}(X) = \bigcup_{X_i \cap X \neq \emptyset} X_i \quad (1)$$

具体来讲, X 子集的所有初等集是由下近似 $\underline{R}(X)$ 的联合,而上近似 $\bar{R}(X)$ 是与有 X 交集的所有初等集的联合.近似空间 $\langle U, R \rangle$ 中,一个普通集合 X 用区间 $[\underline{R}(X), \bar{R}(X)]$ 表示,称为 X 的粗糙集.我们可以将下近似 $\underline{R}(X)$ 理解为那些明确属于 X 的 U 中的元素集合,而上近似 $\bar{R}(X)$ 为可能属于 X 的 U 中的元素集合.再者,当且仅当 $\underline{R}(X) = \bar{R}(X)$ 时,我们认为 X 在 $\langle U, R \rangle$ 中是可定义的.

关于粗糙集, Pawlak 提出了一些基本的性质,本文仅仅描述 3 条由这些基本性质推理出来相关性质^[10].

对于一个划分 $U/R = \{X_1, X_2, \dots, X_m\}$, 虽然缺少足够的知识,不能准确定义其中的 $X_i, i=1, 2, \dots, m$, 但可以使用现有信息定义 X_i 的下近似区域 $\underline{R}(X_i)$ 和 X_i 的上近似区域 $\bar{R}(X_i)$, 从而以此来描述 X_i . X_i 的下近似区域 $\underline{R}(X_i)$ 和上近似区域 $\bar{R}(X_i)$ 满足以下性质:

- (1) $\emptyset \subseteq \underline{R}(X_i) \subseteq \bar{R}(X_i) \subseteq U$;
- (2) $\underline{R}(X_i) \cap \underline{R}(X_j) = \emptyset, i \neq j$;
- (3) 一个元素 x 不属于任何一个下近似区域 $\Leftrightarrow x$ 一定属于 2 个或多个上近似区域.

这 3 条性质不是相互独立的.本文后续内容涉及到了以上的性质.

1.2 模糊聚类算法思想

传统的聚类分析是一种硬划分.它把每个待辨识的对象严格地划分到某个类中,这种分类的类别界限是分明的.而实际上,大多数对象并没有严格的属性,适合进行软划分.因此,模糊聚类与硬聚类相比,能够获得更丰富的聚类信息.

最为大家所熟悉的一种聚类算法是模糊 C 均值聚类算法(fuzzy C -means, 简称 FCM)^[11].该算法主要通过优化一个目标函数来获得最终的聚类结果.这里,令 $X = \{x_1, \dots, x_j, \dots, x_n\}$ 表示 n 个目标的集, $V = \{v_1, \dots, v_i, \dots, v_c\}$ 表示 c 个聚类中心的集合,其目标函数表示如下:

$$J = \sum_{j=1}^n \sum_{i=1}^c (u_{ij})^m \|x_j - v_i\|^2 \quad (2)$$

这里, u_{ij} 称为模糊隶属度函数.

采用负梯度流方法优化上述方程,即可分别得到模糊隶属度函数和聚类中心的更新公式.根据更新公式进行不断的迭代,最终使得聚类中心达到稳定,并通过对模糊度函数解模糊,获得聚类结果.

然而在该算法中,目标与类中心的距离越大,其概率性隶属度就越小,造成 FCM 对噪声相对敏感. Krishnapuram 和 Keller^[12]提出了可能性 C 均值(possibilistic C -means, 简称 PCM)算法,能够很好地解决 FCM 中对噪声敏感的问题,缺点是容易有重合类别.已有文献提出了将 PCM 和 FCM 相结合的算法,即将两种隶属度联合起来使用^[13,14].

1.3 粗糙集模糊聚类算法思想

粗糙集^[15]理论能够处理不精确数据,因此,将模糊聚类方法与粗糙集理论相结合的聚类分析方法逐渐成为一个重要的研究方向,并且获得了广泛的应用^[16].1982 年, Lingras 等人提出了一个粗糙 C 均值(rough C -means, 简称 RCM)聚类算法^[17],通过一个类中心、一个上近似(upper approximation)和下近似(lower approximation)对来描述一个类.其中,上近似和下近似用两个不同加权参数来计算新的聚类中心.在文献^[17]中,模糊集和粗糙集理论的结合,为处理不确定性问题提供了一个重要的方向.

模糊集和粗糙集在某些方面的互补性为解决与数据相关的不确定性问题提供了数学框架^[18].近年来,Mitra 等人^[19]将粗糙集和模糊集结合,并提出了一种新的 C 均值聚类算法,其中,一个类由一个模糊的下近似区域和边缘区域构成.下近似区域中的每一个对象都有一个与之对应的模糊隶属值,即权重.对于相应的聚类中心和类,其下近似区域内的每一个对象应该具有相似的影响作用,并且其权重应该与其他聚类中心和类无关.这样,根据下近似区域的模糊概念得出,该区域内对象的权重会减小.事实上,这会使得聚类中心与期望的位置有一定偏差.此外,该算法对噪声和异常值敏感.

之后,Maji 等人^[20]提出了一个一般化的模糊 C 均值算法(rough-fuzzy PCM,简称 RFPCM),在 C 均值算法中增加了模糊集中的模糊隶属度和粗糙集中的上下近似两个概念.其中,模糊隶属度包括概率性隶属度和可能性隶属度,其应用能够有效地处理相互重合的类;粗糙集能够处理类中知识的不确定性和不完整性.因此,RFPCM 可以有效地避免 FCM 对噪声敏感的弊端,同时弥补 PCM 对重合类无法确定的不足.

这里,我们详细阐述 RFPCM 的算法框架.首先,令 $X=\{x_1, \dots, x_j, \dots, x_n\}$ 表示 n 个目标的集, $V=\{v_1, \dots, v_i, \dots, v_c\}$ 表示 c 个聚类中心的集合, β_i 表示第 i 个聚类区域;其次,令 $\bar{A}(\beta_i)$ 和 $\underline{A}(\beta_i)$ 表示 β_i 的上下近似, $B(\beta_i) = \{\bar{A}(\beta_i) - \underline{A}(\beta_i)\}$ 表示 β_i 的边界区域.通过优化下面的函数,该算法将 X 分为 c 类:

$$J_{RFP} = \begin{cases} \omega \times A_i + (1 - \omega) \times B_i, & \text{if } \underline{A}(\beta_i) \neq \emptyset, B(\beta_i) \neq \emptyset \\ A_i, & \text{if } \underline{A}(\beta_i) \neq \emptyset, B(\beta_i) = \emptyset, \\ B_i, & \text{if } \underline{A}(\beta_i) = \emptyset, B(\beta_i) \neq \emptyset \end{cases}$$

$$A_i = \sum_{j=1}^c \sum_{x_j \in \underline{A}(\beta_i)} \{a(\mu_{ij})^{m_1} + b(v_{ij})^{m_2}\} \|x_j - v_i\|^2 + \sum_{j=1}^c \eta_i \sum_{x_j \in \underline{A}(\beta_i)} (1 - v_{ij})^{m_2},$$

$$B_i = \sum_{j=1}^c \sum_{x_j \in B(\beta_i)} \{\alpha(\mu_{ij})^{m_1} + \gamma(v_{ij})^{m_2}\} \|x_j - v_i\|^2 + \sum_{j=1}^c \eta_i \sum_{x_j \in B(\beta_i)} (1 - v_{ij})^{m_2}.$$

这里, m_1 和 m_2 是模糊度,通常取 2. ω 为下近似的相对重要性,常量 α 和 γ 分别定义了概率性隶属度 μ_{ij} 和可能性隶属度 v_{ij} 的相对重要性.对于比例参数 η_i ,可通过以下公式来计算:

$$\eta_i = \frac{\sum_{j=1}^n (v_{ij})^{m_2} \|x_j - v_i\|^2}{\sum_{j=1}^n (v_{ij})^{m_2}} \quad (3)$$

采用梯度流方法求解上述方程,从而得到概率性隶属度 μ_{ij} 、可能性隶属度 v_{ij} 以及聚类中心的更新公式.通过不断地迭代,最终得到目标函数的最优解.需要指出的是:对于一个像素 x_j ,其总的隶属度函数 $u_{mj} = \alpha\mu_{ij} + \gamma v_{ij}$.我们在其隶属度 $X = \{x_{1j}, \dots, x_{ij}, \dots, x_{cj}\}$ 中选取最大的值 u_{mj} 和第二大的值 u_{nj} ,取二者之差,如果该差值大于预设的阈值,则该素既属于 $\underline{A}(\beta_m)$ 也属于 $\bar{A}(\beta_m)$.

然而,梯度流方法在迭代寻找最优解的同时,很容易陷入局部最优,使得聚类结果不稳定.故,本文将采用一种典型的进化算法——差分免疫克隆算法,通过全局搜索来优化上述问题,具体实现细节见第 2 节.

2 基于粗糙集与差分免疫模糊聚类算法的图像分割

2.1 差分免疫克隆算法与粗糙集模糊聚类

现实世界中的诸多问题都可以归结为一个优化问题.要解决一个优化问题,首先要明确其目标函数,然后对问题进行建模,再采用某种优化算法来优化该目标函数.例如,聚类问题就可以归结为一个优化问题,设计一个聚类评价准则作为目标函数,采用优化算法的目的就是通过不断优化该目标函数,使聚类结果达到最优.为了有效地解决这些优化问题,人们已经提出了各种有效的优化算法,免疫克隆选择算法及差分进化算法就是其中两种比较优秀的算法.

图像分割是计算机视觉和模式识别中关键技术之一.图像分割的结果就是将图像分成若干个部分,每部分代表图像中不一样的特征,并把同一部分像素标记为同一个值.因为图像中的每一个点都对应着一个像素点,一

个像素点就是一个数据,数据的划分即图像的分割,因此,图像分割在本质上就是一个对数据的聚类问题。

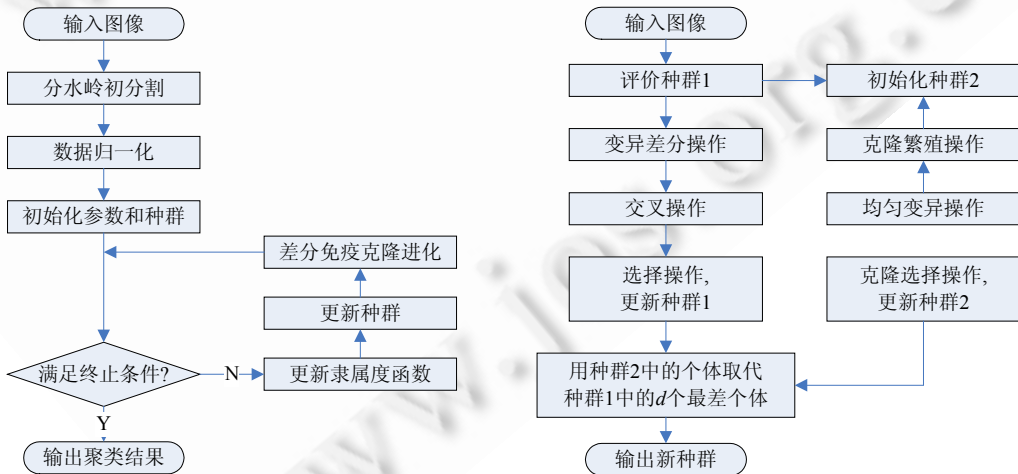
免疫克隆选择算法与差分进化算法在解决优化问题时具有不同的特性:免疫克隆选择算法本质上固有并行性和搜索变化的随机性,不易陷入局部极值;差分进化算法则具有较好的全局收敛性和鲁棒性,非常适合求解各种数值最优化问题.基于这两种不同的算法,马文萍等人提出了一种差分免疫克隆选择算法(differential immune clone clustering algorithm,简称 DICCA)^[21].

差分免疫克隆算法主要是基于差分进化及免疫克隆,通过克隆繁殖、差分变异、交叉及克隆选择操作来进化种群,在进化过程中加入局部搜索机制来提高算法的收敛性.具体过程是:首先,通过基于类中心的编码方式随机生成初始种群1,评价种群(即,计算每个抗体的适应度),选择最优的5个抗体作为种群2;其次,对种群1执行差分变异、差分交叉操作生成变异交叉后的抗体,对种群2执行克隆繁殖、均匀变异、克隆选择操作后形成抗体子种群;最后,用子种群取代种群1中最差的5个个体,作为下一代进化的新抗体种群^[22-24].

模糊聚类和硬聚类相比较,能够获得更丰富的聚类信息,而粗糙集聚类算法善于处理不精确的数据,因此,我们将前面已讨论过的进化算法与粗糙集模糊聚类理论相结合,在 DICCA 聚类算法中引入粗糙集模糊聚类的思想,并将其应用于图形分割领域中,以达到目标识别和分析的目的。

2.2 粗糙集与差分免疫模糊聚类算法描述

将粗糙集与差分免疫模糊聚类算法(RDIFC)运用到图像分割时,首先读入图像,再对图像进行特征提取和分水岭初分割操作,接下来对数据归一化,判断是否满足终止条件,如果不满足则更新隶属度函数,进而更新聚类中心,获得新的种群,新的种群通过差分免疫克隆算法进化.本文算法的流程如图 1(a)所示,差分免疫克隆算法的流程如图 1(b)所示.首先,对初始种群1评价,在种群1的个体中选出最优的5个个体作为种群2,对种群1进行差分变异操作、交叉操作与差分选择操作更新种群1,对种群2进行克隆繁殖操作、均匀变异操作和克隆选择操作更新种群2,用种群2中的个体取代种群1中的最差的d个个体,最终用替代后的种群作为下一代进化的种群.如此迭代,直到满足终止条件时,输出最终的聚类结果.下面将详述具体步骤。



(a) 本文算法流程图

(b) 差分免疫克隆进化算法流程图

Fig.1 Flowcharts of proposed algorithm and DICCA

图1 本文算法及差分免疫克隆进化算法流程图

2.3 数据归一化

为了便于处理待聚类的种群,此处的种群通过上文提到的基于聚类中心编码方式生成的初始化种群,将所有的待聚类数据按比例归一化到闭区间[0,1]中.归一化的过程可以视为一种形式的线性映射.此处采用的映射

函数为

$$x' = -1 + 2(x - \min) / (\max - \min) \quad (4)$$

其中, x' 为将原始待聚类数据 x 映射后得到的结果, 而 \min, \max 分别为所有原始待聚类数据的最小值与最大值. 归一化后的种群将成为后续过程所要处理的目标.

2.4 初始化参数

需要预先给定的参数: 最大迭代次数 T 、最小迭代精度 ε 、下近似的相对重要性 ω 、概率性隶属度的相对重要性 a 、种群规模 P 、交叉概率以及变异概率. 将归一化后的种群作为聚类数据, 从聚类数据中随机选取 $S(S=10)$ 个样本初始化种群. 初始化参数的具体数值在下文仿真实验中详述.

2.5 编码方式

编码的实质是在问题的解空间与算法的搜索空间之间建立一个映射. 编码方案的选择对算法的性能、效率等有很大的影响. 这里, 通过编码将聚类类别数 c 和比列参数 η 嵌入到个体基因当中. 具体来说, DICCA 采用基于聚类中心的编码方式, 种群中每个个体代表候选的聚类中心, 对于 N 维向量的聚类空间, 个体的长度为 $N \times K$, K 为聚类类别数. 采用实数编码方式, 其中, 第 1 组 N 个实数表示第 1 个聚类中心, 第 2 组 N 个实数表示第 2 个聚类中心, 依此类推, 每一组包含上半部分和下半部分. 具体操作说明如下:

例 1: 假设 $N=2, K=3$, 即, 聚类空间维数为 2, 聚类类别数为 3, 则个体 $(5.1, 3.5, 1.4, 0.2, 4.9, 3.2)$ 代表聚类中心 $(5.1, 3.5), (1.4, 0.2), (4.9, 3.2)$.

通过更新概率性隶属度函数和可能性隶属度函数, 从而更新聚类中心, 得到新的种群, 新的种群经过进化, 在全局上搜索最优解. 基于这种交叉的搜索策略提出的算法能够获得稳定的聚类效果, 有效地解决原有粗糙集模糊聚类迭代规则容易陷入局部最优的问题.

2.6 适应度函数

将进化算法用于聚类问题, 就是要寻求合适的聚类评价指标和设计适应度函数, 从而指导进化算法迭代进化方向. 常用的评价聚类划分结果的指标有: 均方误差最小化的 MSE 指标、PBM 指标、DB 指标、XB 指标等, 本文采用最常用的 MSE 指标, 具体描述如下:

(1) 根据样本点与各个聚类中心欧氏距离最近原则获得各个聚类中心的划分集合 $\omega=0.9$, 对每个集合 $\alpha=0.9$ 中的所有样本点求平均值, 得到新的聚类中心为

$$z_i^* = \frac{1}{n_i} \sum_{x_j \in O_i} x_j, \quad i=1, \dots, K, \quad j=1, \dots, n \quad (5)$$

其中, n_i 为 O_i 中样本点的个数.

举例说明如下:

例 2: 例 1 中假设的个体所代表的第 1 个聚类中心为 $(5.1, 3.5)$. 在聚类过程中, 将点 $(4.9, 3.7), (5.9, 3.0)$ 按欧氏距离最近原则划分到以 $(5.1, 3.5)$ 为中心的类中, 那么该类新的中心为 $(5.3, 3.4)$, 它将取代旧的聚类中心 $(5.1, 3.5)$.

(2) 根据公式(3)计算出的新的聚类中心, 得到适应度函数为

$$f = 1 / \left(1 + \sum_{i=1}^K \sum_{x_j \in C_i} \|x_j - z_i^*\| \right) \quad (6)$$

其中, $\|\cdot\|$ 为求取欧氏距离的运算符.

进化算法的目标就是寻找最优的聚类划分结果, 使公式(4)最大化.

2.7 操作算子

2.7.1 差分变异操作

我们首先将当前所有抗体子种群中的每个抗体作为一个目标向量, 利用种群中抗体之间的差异信息得到一个变异向量, 然后, 随机重组得到一个实验向量. 如果该实验向量优于其对应的目标向量, 则用该实验向量代

替目标向量.对所有抗体子种群中的所有抗体执行差分变异交叉操作后,我们就得到变异后的抗体子种群.具体实施方式如下:

对于当前所有子种群中的每个抗体,在差分进化算法中成为目标向量^[25],记为

$$\{X_{i,G} = (x_{i,G}^1, \dots, x_{i,G}^D) \mid i = 1, \dots, NP\},$$

其中, NP 为种群规模,按下式计算后产生变异向量:

$$V_{i,G} = X_{i,G} + F_i \times (X_{best,G}^p - X_{i,G}) + F_i \times (X_{r_1',G} - \tilde{X}_{r_2',G}) \quad (7)$$

其中, F 为变异算子,是一个实常数因子,控制偏差变量的放大作用; $X_{best,G}^p$ 是从当前种群中前 $p\%$ 的优秀个体中随机选择的, $p \in (0,1)$; $X_{r_1',G}$ 是从当前种群中随机选择的, r_1' 是 $[1, NP]$ 之间均匀分布的随机数;与 $X_{r_1',G}$ 不同, $\tilde{X}_{r_2',G}$ 是从当前群体与集合 A 中的并集中随机选出的,其中,集合 A 中存储的是在种群进化过程中被变异向量取代的目标向量.采用该变异策略,有助于增加种群的多样性,从而减少早熟收敛.

在差分进化算法中,如果变异算子 F 选择不当,会严重影响算法的性能. DICCA 中没有采用固定的变异算子,对于当前群体中的每个目标向量 $X_{i,G}$,变异算法 F_i 是依据一个柯西分布独立产生的.采用柯西分布的原因是:柯西分布可以增加变异算子的多样性,从而避免贪婪搜索中的早熟收敛现象.按下式计算产生变异算子^[26]:

$$F_i = randcauchy(\mu_F, 0.1) \quad (8)$$

其中, μ_F 为柯西分布的位置参数,算法中取值为 0.5; 0.1 为柯西分布的尺度参数.在每次迭代后,按下式进行更新:

$$\mu_F = (1-c) \times \mu_F + c \times mean_L(S_F) \quad (9)$$

其中, $c \in (0,1)$, 为一个固定的常数,在算法中取值为 0.1; S_F 为当前代能够成功地进入下一代进化的实验向量对应的变异算子集合, $mean_L(S_F)$ 值由下式计算获得:

$$mean_L(S_F) = \frac{\sum_{F \in S_F} F^2}{\sum_{F \in S_F} F} \quad (10)$$

2.7.2 差分交叉操作

变异结束后,针对每一个目标向量 $X_{i,G}$ 以及它所对应的变异向量 $V_{i,G}$ 进行交叉操作,产生一个实验向量 $U_{i,G}$. 一般情况下采用二项式交叉,定义如下:

$$u_{i,G}^j = \begin{cases} v_{i,G}^j, & \text{if } (rand_j[0,1] \leq CR) \text{ or } (j = j_{rand}) \\ x_{i,G}^j, & \text{if otherwise, } j = 1, 2, \dots, D \end{cases} \quad (11)$$

其中, CR 为交叉概率; $rand_j[0,1]$ 为 $[0,1]$ 随机数产生器的第 j 个估计值; j_{rand} 是 $[1, D]$ 之间均匀分布的随机整数,它的引入,是为了确保实验向量中至少有 1 个参数与它所对应的目标向量不同.

DICCA 中的交叉概率 CR 值不是固定的,它是由服从正态分布 $N(CRmean, 0.15)$ 的随机数确定^[21]. 首先,产生一个服从正态分布 $N(0.5, 0.15)$ 的随机数序列;然后,针对当前克隆种群中的每个抗体,从该正态分布中随机为其选择一个交叉概率.每代结束后,按下式对 $CRmean$ 值进行更新:

$$CRmean = (1-c) \times CRmean + c \times mean(S_{CR}) \quad (12)$$

其中, c 为 $[0,1]$ 之间的随机数, $mean(S_{CR})$ 为所有成功进入下一代的实验向量所对应的 CR 值的均值.实验结果表明,这种方法要优于传统的采用固定交叉概率的差分进化算法.

交叉结束后,对所有的实验向量进行评价,即,计算每个实验向量的目标函数值,将该目标函数值与其对应目标向量的函数值进行比较,如果实验向量的目标函数值优于目标向量,则该实验向量将取代目标向量,完成该操作后,所有抗体子种群中的抗体已完成变异操作,其中,只有某些抗体被更新.

2.7.3 克隆增殖操作

在免疫学中,克隆增殖是指通过无性繁殖可连续传代并形成群体.在人工免疫响应模型中,对抗体种群 $A = \{a_1, a_2, \dots, a_n\}$ 的克隆增殖操作 T_c^c 定义为

$$T_c^c(A) = T_c^c(a_1 + a_2 + \dots + a_n) = T_c^c(a_1) + T_c^c(a_2) + \dots + T_c^c(a_n) \quad (13)$$

其中,

$$T_c^c(a_i) = \{a_i^1 + a_i^2 + \dots + a_i^{q_i}\}, a_i^j = a_i, j = 1, 2, \dots, q_i \quad (14)$$

q_i 为抗体 a_i 的克隆规模. 经过克隆增值操作的抗体种群定义如下:

$$A' = T_c^c(A) = \{a_1^1, a_1^2, \dots, a_1^{q_1}, a_2^1, a_2^2, \dots, a_2^{q_2}, \dots, a_n^1, a_n^2, \dots, a_n^{q_n}\} \quad (15)$$

DICCA 算法中采用比例克隆. 比例克隆是指具有较大亲和度的抗体被更多次地复制, 即, 具有较大亲和度值的个体有较大的 q_i . 算法中, q_i 的值按下式计算:

$$q_i = \left\lfloor nc \times \frac{\text{aff}(a_i, A)}{\sum_{j=1}^{NP} \text{aff}(a_j, A)} \right\rfloor \quad (16)$$

其中, nc 是克隆种群大小的期望值, nc 取 25; $\text{aff}(a_i, A)$ 为抗体种群 A 中抗体 a_i 的亲合度值. 经过克隆繁殖后, 得到每个父代抗体对应的抗体子种群.

2.7.4 均匀变异操作

DICCA 中采用均匀变异策略, 设 v 为待变异的基因位值, 则变异后的 v' 值按下式计算:

$$v' = \begin{cases} v + (\varepsilon + \sigma) \times v, & \text{if } \text{rand}(0,1) < 0.5 \\ v - (\varepsilon + \sigma) \times v, & \text{if otherwise} \end{cases} \quad (17)$$

其中, ε 和 σ 为 (0,1) 之间均匀分布的随机数. 经过克隆变异操作后的种群可记为

$$A'' = T_c^m(A') = \{(a_1^1)', (a_1^2)', \dots, (a_1^{q_1})', (a_2^1)', (a_2^2)', \dots, (a_2^{q_2})', \dots, (a_n^1)', (a_n^2)', \dots, (a_n^{q_n})'\} \quad (18)$$

均匀变异操作结束后, 按照下文介绍的克隆选择操作, 从父代抗体种群与变异后的克隆子种群中选择优秀的抗体进入下一代进化操作.

2.7.5 克隆选择操作

克隆选择操作是从抗体各自克隆增值后的子代中选择优秀个体形成新的种群, 是一个无性选择过程. 对经过变异操作的抗体种群的克隆选择操作 T_c^s 定义如下:

$$T_c^s(\{a_1 + (a_1^1)' + (a_1^2)' + \dots + (a_1^{q_1})'\} + \{a_2 + (a_2^1)' + (a_2^2)' + \dots + (a_2^{q_2})'\} + \dots + \{a_n + (a_n^1)' + (a_n^2)' + \dots + (a_n^{q_n})'\}) = a_1' + a_2' + \dots + a_n' \quad (19)$$

其中, $T_c^s(\{a_i + (a_i^1)' + (a_i^2)' + \dots + (a_i^{q_i})'\}) = a_i'$, 抗体 a_i' 在子种群 $\{a_i, (a_i^1)', (a_i^2)', \dots, (a_i^{q_i})'\}$ 中具有最高亲合度. 经过克隆选择操作, 得到下一代种群.

2.8 终止准则

粗糙集与差分免疫模糊聚类算法的终止条件是: 在种群迭代的过程中, 当达到算法最大迭代次数, 或者连续 4 代的适应度值都小于最小迭代的精度时, 算法停止迭代, 输出最终聚类结果, 算法终止.

3 仿真实验

3.1 评价指标

本节将 RDIFC 算法应用在图像分割上, 以实验结果来评价此算法的有效性和可行性, 并与其他 4 种算法进行对比实验, 从视觉评价方面和分割正确率两个方面来检测算法的性能.

其中, 视觉评价主要从错分情况、区域一致性、边缘保持、细节信息保留等对分割后的图像结果进行评价, 而分割正确率指的是正确分割的像素点个数总量占总像素点个数的比例.

3.2 实验内容与结果分析

为了进一步证明 RDIFC 在图像分割上的有效性, 我们选择了 9 幅图像进行实验, 其中包括 5 幅纹理图像(如图 2 所示)、2 幅自然图像(如图 3 所示)和 2 幅 SAR 图像(如图 4 所示). 我们选择了 4 种聚类算法与 RDIFC 算法进行对比, 分别是 GARFPCM(基于粗糙集和 GA 的模糊聚类算法)^[27]、MISA(密母算法)^[28]、DICCA(差分免

疫克隆算法)^[21]和 IMCA(免疫密母聚类算法)^[29].

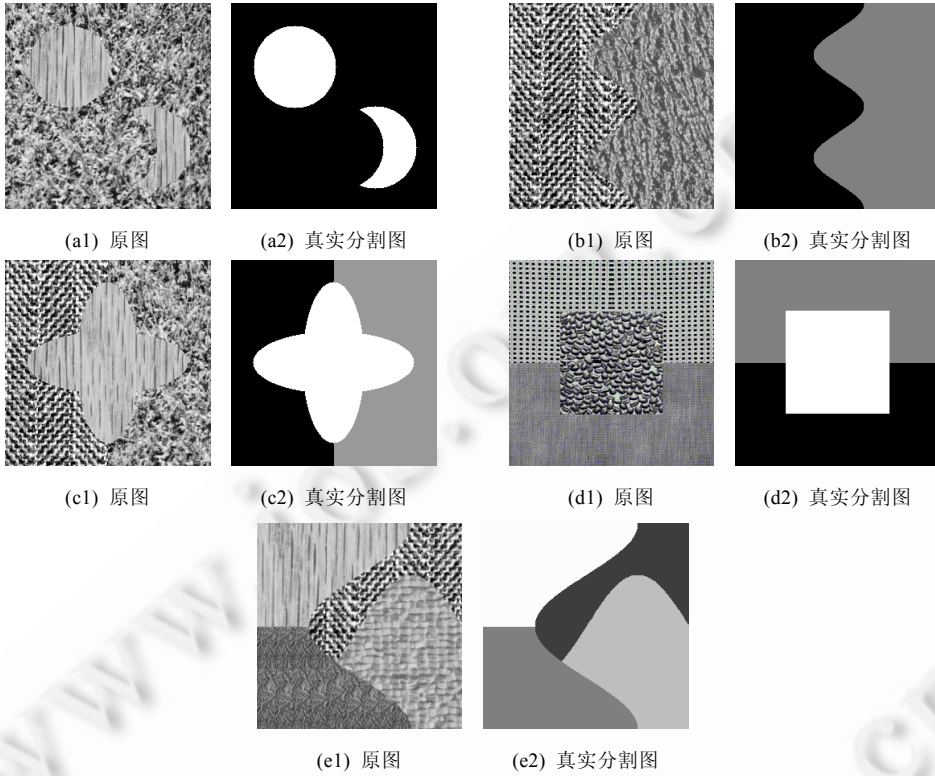


Fig.2 Five original images of artificial texture and real segmentation images

图 2 5 幅人工纹理图像原图及真实分割图



Fig.3 Two original nature images

图 3 两幅自然图像原图

Fig.4 Two original SAR images

图 4 两幅 SAR 图像原图

在每个数据集上运行聚类过程 20 次,实验中算法参数设置如下:

- RDIFC:种群规模 10,最小迭代精度 $\epsilon=10^{-6}$,两个相对重要性系数 $\omega=0.9, \alpha=0.9$,交叉概率 0.5,变异概率 0.5,最大迭代次数 100.
- GARFCM:种群规模 10,最小迭代精度 $\epsilon=10^{-6}$,两个相对重要性系数 $\omega=0.9, \alpha=0.9$,交叉概率 0.8,变异概率 0.11,最大迭代次数 100.
- MISA:种群规模 30,交叉概率 0.85,变异概率 0.05,最大迭代次数 100.
- DICCA:种群规模 25,交叉概率 0.5,变异概率 0.5,最大迭代次数 100.
- IMCA:种群规模 30,交叉概率 0.85,变异概率 0.05,最大迭代次数 100.

分割算法的第 1 阶段首先需要对图像进行分水岭初分割,以将图像划分成不相互重叠的图像块,并将此块作为后续聚类过程的处理对象.图 5 给出了 5 幅人工纹理图像经过初分割后的结果,图 6 给出了自然图像经过初分割后的结果,图 7 给出了 SAR 图像经过初分割后的结果.从这些图像的初分割结果上可以大致看出:在微观上,这些小的图像块基本上是按照图像中的纹理、像素等图像的基本元素划分出来的;在宏观上,则是按照图像中不同的目标区域划分出来.

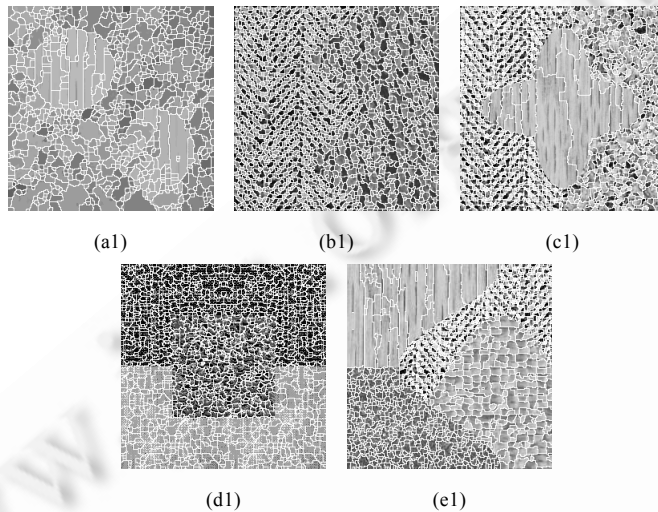


Fig.5 Five artificial texture initial segmentation images

图 5 5 幅人工纹理图像初分割图

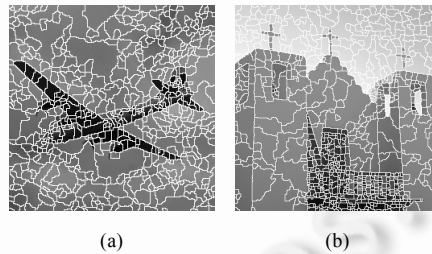


Fig.6 Two nature initial segmentation images

图 6 两幅自然图像初分割图

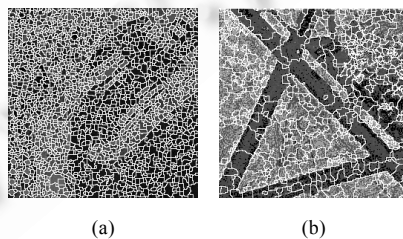


Fig.7 Two SAR initial segmentation images

图 7 两幅 SAR 图像初分割图

图 8~图 16 给出了 RDIFC 与 4 个基于聚类的对比分割算法 MISA, DICCA, IMCA 和 GARFPCM 在 9 幅图上的分割结果(图中显示的是从 20 次重复实验中选择的平均结果图).

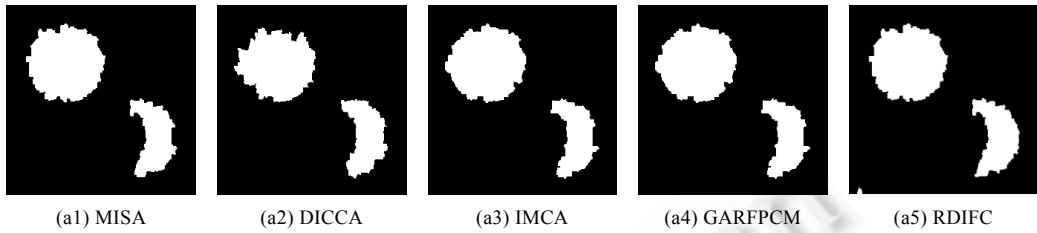


Fig.8 Five algorithm results in artificial texture image (a1)
图 8 5 种算法在人工纹理图像(a1)上的结果



Fig.9 Five algorithm results in artificial texture image (b1)
图 9 5 种算法在人工纹理图像(b1)上的结果

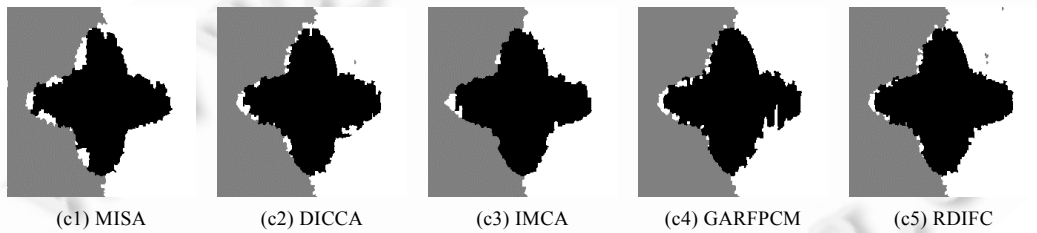


Fig.10 Five algorithm results in artificial texture image (c1)
图 10 5 种算法在人工纹理图像(c1)上的结果

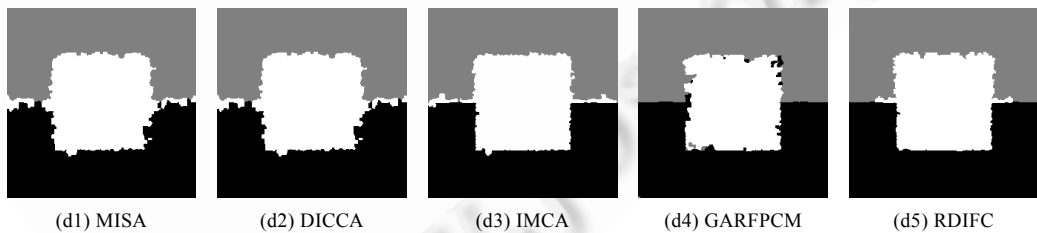


Fig.11 Five algorithm results in artificial texture image (d1)
图 11 5 种算法在人工纹理图像(d1)上的结果

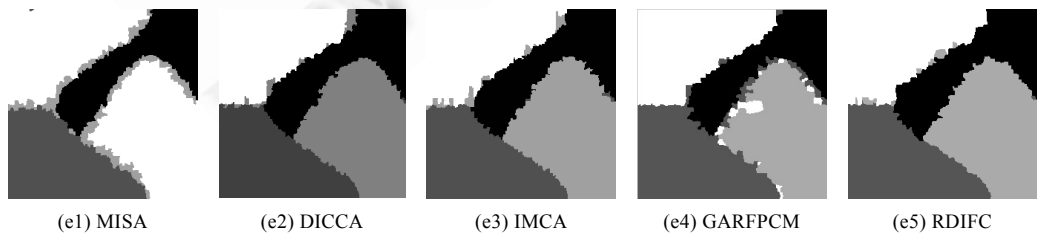


Fig.12 Five algorithm results in artificial texture image (e1)
图 12 5 种算法在人工纹理图像(e1)上的结果

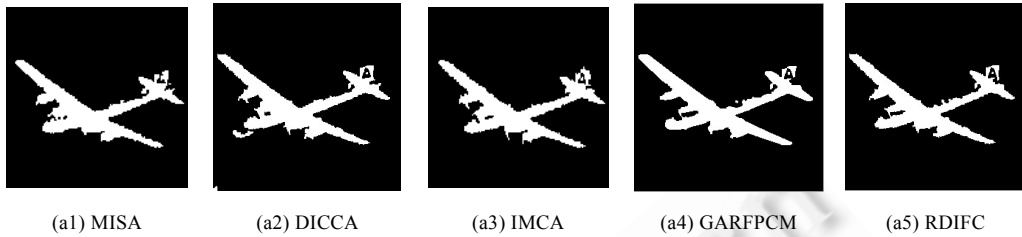


Fig.13 Five algorithm results in nature image (a)

图 13 5 种算法在自然图像(a)上的结果



Fig.14 Five algorithm results in nature image (b)

图 14 5 种算法在自然图像(b)上的结果



Fig.15 Five algorithm results in SAR image (a)

图 15 5 种算法在 SAR 图像(a)上的结果

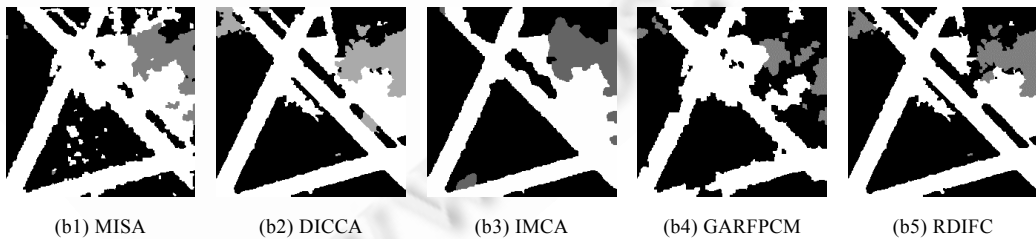


Fig.16 Five algorithm results in SAR image (b)

图 16 5 种算法在 SAR 图像(b)上的结果

下面从视觉角度和统计学角度分析 5 种算法在人工纹理图像(a1,b1,c1,d1,e1)上的表现。

从图 8 可以看到:5 幅图都将两类区域分割了出来,从视觉效果上看并无明显差异,但是除了 RDIFC 算法以外,其他 4 种算法在右下角类似于月亮形状的区域分割出来结果都有一个缺口,而 RDIFC 的边缘保持较好且比较光滑。从图 9 可以看到:由于其一致性较好,纹理比较简单,5 种算法的分割结果从视觉上看不出明显差异,但从细节信息来看,RDIFC 和 GARFPCM 在边缘保持上比其他 3 种算法要好。从图 10 可以看到:MISA 算法在左半部分边缘处有几大块错分,并且边缘不光滑,5 种算法在边缘衔接处都有错分的情况;相比之下,RDIFC 错分位置较

少,并且它的边缘保持较好.从图 11 可以看到:MISA 和 DICCA 在边缘衔接处都有大块的错分情况,IMCA 与这两种算法相比,边缘衔接较好;GARFPCM 算法在左、右两部分边缘分割最好,但是在中间区域有错分,而且边缘不光滑;RDIFC 在中间区域边缘保持较好而且最光滑,并且局域一致性较好.从图 12 可以看到:MISA 有明显错分的类别,将左上角区域和右下角区域分成了一类,它只分出了 3 类;GARFPCM 在右下角区域边缘衔接处有明显的错分;与其他算法相比,RDIFC 边缘保持较好.

关于以上 5 幅人工纹理图像的实验,表 1 从统计学的角度给出了 20 次运行的平均结果.这些结果是通过真实分割结果计算得到的.从表 1 中的数据可以看出:对于 a1 图,MISA 的正确率最低,DICCA,IMCA 和 GARFPCM 的正确率相当,RDIFC 获得了最高的平均分割正确率;对于 b1 图,MISA 的正确率最低,DICCA 和 IMCA 的正确率持平,GARFPCM 与 RDIFC 正确率相当,RDIFC 获得了最高的正确率;对于 c1 图,RDIFC 获得了最高的平均正确率;对于 d1 图,RDIFC 与基于 MISA,DICCA,IMCA 和 GARFPCM 聚类的分割算法相比,正确率分别提高了 2%,1.44%,0.36% 和 0.64%;对于 e1 图,MISA 的正确率最低,RDIFC 获得了最高的平均正确率,与基于 MISA,DICCA,IMCA 和 GARFPCM 聚类的分割算法相比,正确率分别提高了 26.19%,0.58%,0.26% 和 1.61%.在这 5 幅人工纹理图像分割实验中,RDIFC 都取得了最高的平均正确率.

Table 1 Statistical data of 20 times run in the artificial texture image

表 1 在人工纹理图像上 20 次运行的统计数据

| 人工纹理图像 | 算法平均分割正确率 | | | | |
|--------|-----------|---------|---------|---------|---------|
| | MISA | DICCA | IMCA | GARFPCM | RDIFC |
| a1 | 0.976 1 | 0.980 1 | 0.980 3 | 0.980 1 | 0.981 1 |
| b1 | 0.980 4 | 0.986 5 | 0.986 9 | 0.988 4 | 0.988 5 |
| c1 | 0.960 4 | 0.967 8 | 0.971 4 | 0.964 5 | 0.972 5 |
| d1 | 0.962 5 | 0.968 1 | 0.978 9 | 0.976 1 | 0.982 5 |
| e1 | 0.708 3 | 0.964 4 | 0.967 6 | 0.954 1 | 0.970 2 |

关于自然图像的实验,图 13 给出了 5 种分割算法在自然图像(a)上的对比分割结果.原图像上有飞机和天空两个区域,对于 MISA,飞机的螺旋桨没有分出来,并且飞机尾部上面的字母分割不完全,细节信息不够完善;DICCA 在机头前部有错分,应该分为天空的部分与飞机划为一类;GARFPCM 在机身靠近尾部的地方有错分,并且左下角有错分点;RDIFC 相对于 IMCA 来说,机翼的区域边缘保持较好.图 14 给出了 5 种分割算法在自然图像(b)上的对比分割结果.原图像上有教堂、楼梯部分和天空 3 个区域,基于 DICCA 算法将教堂和楼梯部分分割出来,但是在右上角天空区域有很大一块的错分,将天空和教堂归为一类;MISA 中教堂上面中间的那个十字架分割有断层,并且在右上角天空区域也有错分;GARFPCM 上半部分有明显的错分现象,但是楼梯部分的栅栏处分割效果好;RDIFC 与 IMCA 相比,在楼梯部分的右部 IMCA 有一小块错分,而 RDIFC 没有,并且在楼梯部分栅栏上,RDIFC 比 IMCA 细节分割效果好.因此,从区域一致性和边缘保持方面来说,IMCA 的结果比以上 4 种算法的结果要好.

关于 SAR 图像的实验,图 15 给出了 5 种分割算法在 SAR 图像(a)上的对比分割结果.原图像上有水域和陆地两个区域,MISA 在左边区域有两块错分的斑点,并且边缘不光滑;DICCA 在左边陆地部分有一块错分的斑点,右上角水域和陆地衔接处的细节信息没有被分割出来;GARFPCM 右侧水域部分有一块属于陆地的区域错分为水域,而 RDIFC 在这一块分割较好;对于 IMCA 算法,上半部分水域部分有错分,将一部分水域错分为陆地部分,而 RDIFC 在这一部分边缘保持和区域一致性都较好.图 16 给出了 5 种分割算法在自然图像(b)上的对比分割结果.原图像上有机场跑道、平地 and 建筑物 3 个区域,MISA 跑道中间的三角区域有许多斑点噪声,右上角区域一致性不好;GARFPCM 有明显的错分;DICCA 左上角有一块将平地错分为建筑物,在中间部分有一块错分情况,将平地错分为机场跑道,右半部分有一块将平地错分为建筑物;IMCA 左下角有一小块错分;RDIFC 不仅将右部属于平地的区域完全分割了出来,并且中间三角区域的一致性较好.

4 总结与展望

本文提出了粗糙集与差分免疫模糊聚类算法(RDIFC).该算法在差分免疫克隆算法的基础上,结合了粗糙集模糊聚类的思想,利用粗糙集模糊聚类善于处理不确定数据的优势,对 9 幅图像进行了分割实验,并与 MISA, DICCA,IMCA 和 GARFPCM 这 4 种算法进行了对比实验.实验结果验证了本算法能够获得更高的正确率和更好的分割结果,说明本文算法可以有效地解决图像分割问题.采用模糊聚类的思想,能够比硬聚类获得更丰富的聚类信息.因为加入了模糊集的思想,模糊聚类算法比 DICCA 和 IMCA 进化算法更善于处理不确定的数据.然而,RDIFC 和 GARFPCM 都需要多个参数才能保证算法的正常使用,比如,下区域的相对重要性、概率隶属度的相对重要性等.在实验中,这些参数设置得是否合适,对分割结果的影响是很大的.因此,如何能够保证设置一个可以使分割结果最优的参数,是值得我们去深入讨论的问题,也是日后研究的一个重点.

References:

- [1] Wang L. Intelligent Optimization Algorithm and Its Application. Beijing: Tsinghua University Press, 2001 (in Chinese).
- [2] Song XF, Kang JL, Wang H. Development and application of evolutionary algorithm. Modern electronic technology, 2006,29(20): 66–68 (in Chinese with English abstract).
- [3] Jiao LC, Gong MG, Wang S, Hou B, Liu F, Zhang XR, Zhou WD. Natural Computation, Machine Learning and Image Understanding Frontier. Xi'an: Xi'an University Press, 2008 (in Chinese).
- [4] Kuang Y, Xiong JC. Evolutionary algorithms and evolutionary theory. Journal of Xinyang Normal University, 2000,13(4):486–489 (in Chinese with English abstract).
- [5] Gao XB, Xie WX. Fuzzy clustering theory development and application research. Chinese Science Bulletin, 1999,44(21): 2241–2251 (in Chinese with English abstract).
- [6] Zadeh LA. Fuzzy set. Information and Control, 1965,8(3):338–353. [doi: 10.1016/S0019-9958(65)90241-X]
- [7] Ruspini EH. A new approach to clustering. Information and Control, 1969,15(1):22–32. [doi: 10.1016/S0019-9958(69)90591-9]
- [8] Le K. Fuzzy relation compositions and pattern recognition. Information Sciences, 1996,89(2):107–130. [doi: 10.1016/0020-0255(95)00231-6]
- [9] Tamura S, Higuchi S, Tanaka K. Pattern classification based on fuzzy relations. IEEE Trans. on Systems, Man, and Cybernetics, 1971,1(1):61–66. [doi: 10.1109/TSMC.1971.5408605]
- [10] Wu ZY, Leahy R. An optimal graph theoretic approach to data clustering: Theory and its application to image segmentation. IEEE Trans. on Pattern Analysis and Machine Intelligence, 1993,15(11):1101–1113. [doi: 10.1109/34.244673]
- [11] Bezdek JC. Pattern Recognition with Fuzzy Objective Function Algorithm. New York: Springer-Verlag, 1981.
- [12] Krishnapuram R, Keller JM. A possibilistic approach to clustering. IEEE Trans. on Fuzzy Systems, 1993,1(2):98–110. [doi: 10.1109/9/91.227387]
- [13] Pal NR, Pal K, Keller JM, Bezdek JC. A possibilistic fuzzy c-means clustering algorithm. IEEE Trans. on Fuzzy Systems, 2005, 13(4):517–530. [doi: 10.1109/TFUZZ.2004.840099]
- [14] Masulli F, Rovetta S. Soft transition from probabilistic to possibilistic fuzzy clustering. IEEE Trans. on Fuzzy Systems, 2006,14(4): 516–527. [doi: 10.1109/TFUZZ.2006.876740]
- [15] Feng Z. A K -means clustering algorithm based on rough set. Computer Engineering and Application, 2006,(20):141–146 (in Chinese with English abstract).
- [16] Maji P, Pal SK. Rough-Fuzzy C -medoids algorithm and selection of bio-basis for amino acid sequence analysis. IEEE Trans. on Knowledge and Data Engineering, 2007,19(6):859–872. [doi: 10.1109/TKDE.2007.190609]
- [17] Lingras P, West C. Interval set clustering of Web users with rough K -means. Journal of Intelligent Information Systems, 2004, 23(1):5–16. [doi: 10.1023/B:JIIS.0000029668.88665.1a]
- [18] Dubois D, Prade H. Rough fuzzy sets and fuzzy rough sets. Int'l Journal of General Systems, 1990,17(2):191–209.
- [19] Mitra S, Banka H, Pedrycz W. Rough-Fuzzy collaborative clustering. IEEE Trans. on Systems, Man, and Cybernetics, Part B: Cybernetics, 2006,36(4):795–805. [doi: 10.1109/TSMCB.2005.863371]
- [20] Maji P, Pal SK. Rough set based generalized fuzzy C -means algorithm and quantitative indices. IEEE Trans. on Systems, Man, and Cybernetics, Part B: Cybernetics, 2007,37(6):1529–1540. [doi: 10.1109/TSMCB.2007.906578]

- [21] Ma WP, Ti FF, Li CL, Jiao LC. Image segmentation based on differential immune clone clustering algorithm. *Int'l Journal of Intelligent Computing and Cybernetics*, 2013,6(1):83–102.
- [22] Du HF, Jiao LC, Wang SA. Clonal operator and antibody clone algorithms. In: *Proc. of the 1st Int'l Conf. on Machine Learning and Cybernetics*. 2002. 506–510. [doi: 10.1109/ICMLC.2002.1176807]
- [23] Storn R, Price K. Differential evolution—A simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, 1997,11(4):341–359. [doi: 10.1023/A:1008202821328]
- [24] Qin AK, Huang VL, Suganthan PN. Differential evolution algorithm with strategy adaptation for global numerical optimization. *IEEE Trans. on Evolutionary Computation*, 2009,13(2):398–417. [doi: 10.1109/TEVC.2008.927706]
- [25] Nguyen QH, Ong YS, Lim MH. A probabilistic memetic framework. *IEEE Trans. on Evolutionary Computation*, 2009,13(3):604–623. [doi: 10.1109/TEVC.2008.2009460]
- [26] Maulik U, Bandyopadhyay S. Genetic algorithms-based clustering technique. *Pattern Recognition*, 2000,33(9):1455–1465. [doi: 10.1016/S0031-3203(99)00137-5]
- [27] Ma WP, Ge XH, Jiao LC. An improved generalized fuzzy *C*-means clustering algorithm based on GA. In: Zhang Y, *et al.*, eds. *Proc. of the Intelligent Science and Intelligent Data Engineering*. LNCS 7202, Berlin, Heidelberg: Springer-Verlag, 2012. 599–606. [doi: 10.1007/978-3-642-31919-8_76]
- [28] Hu YS, Chen ZH. Clustering analysis based on hybrid genetic algorithm. *Pattern Recognition and Artificial Intelligence*, 2001, 14(3):352–355 (in Chinese with English abstract).
- [29] Ma WP, Huang YY, Li CL, Liu J. Image segmentation based on a hybrid immune memetic algorithm. In: *Proc. of the IEEE Congress on Evolutionary Computation*. Piscataway: IEEE, 2012. 1–8. [doi: 10.1109/CEC.2012.6256422]

附中文参考文献:

- [1] 王凌. 智能优化算法及应用. 北京: 清华大学出版社, 2001.
- [2] 宋晓峰, 亢金龙, 王宏. 进化算法的发展与应用. *现代电子技术*, 2006, 29(20): 66–68.
- [3] 焦李成, 公茂果, 王爽, 侯彪, 刘芳, 张向荣, 周伟达. 自然计算、机器学习与图像理解前沿. 西安: 西安电子科技大学出版社, 2008.
- [4] 况颐, 熊吉春. 进化算法及进化理论初探. *信阳师范学院学报*, 2000, 13(4): 486–489.
- [5] 高新波, 谢维信. 模糊聚类理论发展及应用的研究进展. *科学通报*, 1999, 44(21): 2241–2251.
- [15] 冯征. 一种基于粗糙集的 *K*-Means 聚类算法. *计算机工程与应用*, 2006, (20): 141–146.
- [28] 胡玉锁, 陈宗海. 基于混合遗传算法的聚类分析. *模式识别与人工智能*, 2001, 14(3): 352–355.



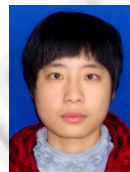
马文萍(1981—),女,陕西铜川人,博士,副教授,CCF 会员,主要研究领域为自然计算,智能图像处理.
E-mail: wpma@mail.xidian.edu.cn



黄媛媛(1988—),女,硕士,主要研究领域为智能图像处理.
E-mail: hdoudou521@126.com



李豪(1989—),男,硕士,主要研究领域为智能图像处理.
E-mail: lihaolifan@163.com



李晓婷(1989—),女,硕士生,主要研究领域为智能图像处理.
E-mail: bujizhizhoutop@163.com



焦李成(1959—),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为自然计算,数据挖掘,图像处理,智能信息处理.
E-mail: jlc1023@163.com