

免疫克隆 SAR 图像分割算法

马文萍 焦李成 尚荣华

(西安电子科技大学智能感知与图像理解教育部重点实验室和智能信息处理研究所 西安 710071)

摘要: 由于存在相干斑噪声的影响,使得常规的图像分割技术应用于 SAR 图像时,效果往往较差。该文提出一种新人工免疫系统 SAR 图像分割算法,算法首先提取每幅图像的纹理特征,并对其进行实数编码,然后通过借鉴生物学免疫系统的抗体克隆选择机理,构造适合于图像分割的克隆算子,以较快的收敛速度优化目标函数。实验结果表明,新算法是一种有效的 SAR 图像区域分割方法。

关键词: 合成孔径雷达; 图像分割; 免疫克隆; 聚类

中图分类号: TN957.52

文献标识码: A

文章编号: 1009-5896(2009)07-1749-04

Immune Clonal SAR Image Segmentation Algorithm

Ma Wen-ping Jiao Li-cheng Shang Rong-hua

(Key Lab of Intelligent Perception and Image Understanding of Ministry of Education of China,
Institute of Intelligent Information Processing, Xidian University, Xi'an 710071, China)

Abstract: Because of the multiplicative nature of the speckle noise in SAR images, it is difficult to solve SAR image segmentation problems using general image segmentation technique. A novel SAR image segmentation algorithm based on an artificial immune system is proposed. After extracting texture features from an image and encoding them with real numbers, it determines partitioning of the feature vectors by optimizing the objective function based on clonal operator. The experimental results show that the novel algorithm is feasible and effective for SAR image segmentation.

Key words: Synthetic Aperture Radar(SAR); Image segmentation; Immune clonal; Clustering

1 引言

合成孔径雷达(Synthetic Aperture Radar, SAR)是一种工作在微波波段的相干成像雷达。它以其高分辨率和全天候、全天时、大面积的数据获取能力而成为全世界各国普遍的对地观测技术^[1]。快速、准确的图像分割,特别是对感兴趣目标的分割,是 SAR 图像解译的关键技术之一^[2]。随着各种数学理论、方法和工具的提出,许多聚类算法被应用于图像分割领域并取得了较为满意的效果^[2-4]。众所周知, SAR 图像中存在大量的斑点噪声,传统的聚类方法对噪声非常敏感,随着遗传算法(Genetic Algorithm, GA)的出现,人们提出了许多基于 GA 的模式识别方法^[5],但 GA 收敛速度较慢,而且还容易出现早熟。

人工免疫系统方法具有认知科学鲜明和作为弱方法适用性强等突出优点,其研究成果涉及到控制、数据处理、优化学习和故障诊断等许多领域,已经

成为继神经网络、模糊逻辑和进化计算后人工智能的又一研究热点^[6]。克隆选择是生物免疫系统理论的重要学说,它是生物体免疫系统自适应抗原刺激的动态过程,而在这一过程中免疫系统所体现出的学习、记忆、抗体多样性等生物特性,正是人工免疫系统所借鉴的。本文结合免疫克隆优化机理,提出一种基于聚类的 SAR 图像分割算法(Immune Clonal Clustering Algorithm, ICCA)。本文将 ICCA, K-均值算法^[7]以及 GAC^[5]分别对真实的 SAR 图像分割,并对结果进行了比较分析。

2 聚类技术

聚类分析是当今飞速发展的数据采掘和探查性数据分析中一个极为重要的技术^[7],通常采用 K-均值算法,它的实现过程可简单描述如下:

步骤1 从聚类数据集 $\{x_1, x_2, \dots, x_n\}$ 中随机选取 K 个点作为初始的聚类中心 z_1, z_2, \dots, z_K ;

步骤2 将点 $x_i, i = 1, 2, \dots, n$ 划分到类 $C_j, j \in \{1, 2, \dots, K\}$, 遵循下列原则:

$$\|x_i - z_j\| < \|x_i - z_p\|, p=1, 2, \dots, K, \text{ 且 } j \neq p \quad (1)$$

即把该点划分到距离其最近的类中;

2008-07-07 收到, 2009-03-17 改回

国家自然科学基金(60703107, 60703108), 国家 863 计划项目(2008AA12Z2475853)和国家 973 计划项目(2006CB705700)资助课题

步骤3 依据下列公式重新计算新的聚类中心 $z_1^*, z_2^*, \dots, z_K^*$:

$$z_i^* = \frac{1}{n_i} \sum_{x_j \in C_i} x_j, \quad i = 1, 2, \dots, K \quad (2)$$

其中 n_i 是类 C_i 中数据点的个数;

步骤4 如果 $z_i^* = z_i, i = 1, 2, \dots, K$, 即每类的聚类中心不再变化, 则返回聚类结果, 否则转步骤2。

目前进化计算已被广泛应用于聚类技术中^[5,8,9], 这些进化聚类方法包括 K-均值算法在处理超大规模分布的数据集上取得很好的成果, 但是在一些复杂数据集以及未知分布特征的数据集上表现稍差。作为智能计算领域的新发展, 免疫克隆优化算法兼顾全局搜索和局部搜索能力, 在搜索中不易陷入局部极值, 具有更好的种群多样性。文献[7]中指出 K-均值算法易陷入局部极值点, 并且它对噪音和孤立点数据非常敏感, 且性能极易受数据分布的影响。因此, 本文结合免疫克隆优化算法的优点, 将其寻优的特性应用到 K-均值算法中, 克服 K-均值算法的缺点, 并成功应用于 SAR 图像分割。

3 基于免疫克隆优化的 SAR 图像分割

3.1 免疫克隆优化

免疫克隆优化算法的基本流程可描述如下:

步骤1 $k:=0$, 初始化抗体种群 $A(0)$, 设定算法参数, 计算初始种群的亲和度;

步骤2 依据亲和度和设定的抗体克隆规模, 对抗体种群 $A(k)$ 进行克隆操作, 得到克隆后的种群 $Y(k)$;

步骤3 以概率 p_m 对 $Y(k)$ 实现免疫基因操作, 得到新的抗体种群 $Z(k)$;

步骤4 对 $Z(k)$ 进行克隆选择操作, 得到新的抗体种群 $A'(k)$;

步骤5 若满足终止条件, 输出 $A'(k)$, 否则, $A(k+1) = A'(k)$, $k:=k+1$ 返回步骤2。

算法中的核心算子: 克隆算子, 具有很强的群体搜索能力, 该算子主要包含3种操作: 克隆操作、免疫基因操作和克隆选择操作。克隆操作有效地实现了搜索空间的扩张; 免疫基因操作包含交叉和变异, 免疫学认为, 亲和度成熟和抗体多样性的产生主要依靠抗体的高频变异, 而非交叉或重组; 克隆选择操作通过局部择优, 有效地压缩了种群的大小。由于克隆算子本身的选择机制具有记忆功能, 因此可以保证算法以概率1收敛到最优解^[6]。我们采用欧式距离作为相似度函数。下面给出了 K 个类别的目标函数 μ :

$$\mu(C_1, C_2, \dots, C_k) = \sum_{i=1}^K \sum_{x_j \in C_i} \|x_j - z_i\| \quad (3)$$

免疫克隆优化算法的任务就是在 N 维聚类空间搜索 K 个最优的聚类中心, 即最小化目标函数 μ 。

3.2 基于免疫克隆优化的 SAR 图像分割

在人工免疫系统中, 抗原一般指问题及其约束, 抗体一般指问题的候选解。针对 SAR 图像分割问题, 抗原代表参与聚类的样本, 每个抗体代表候选的聚类中心。对于 N 维聚类空间, 抗体的长度为 $N \times K$, 采用实数编码方式, 其中第一组 N 个实数表示第一个聚类中心, 下一组 N 个实数表示第二个聚类中心, 依此类推。具体的操作以下举例说明:

例1 假设 $N=2, K=3$, 即聚类空间维数为2, 聚类个数为3。那么抗体(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)。

(1)克隆操作 克隆在免疫学中是指通过无性繁殖可连续传代并形成群体。在人工免疫系统中, 对抗体种群 $B = \{b_1, b_2, b_3, \dots, b_n\}$ 的克隆增殖操作 r_2 定义为^[9]

$$b_1 + b_2 + \dots + b_n \longrightarrow \{b_1^1 + b_1^2 + \dots + b_1^{q_1}\} + \{b_2^1 + b_2^2 + \dots + b_2^{q_2}\} + \dots + \{b_n^1 + b_n^2 + \dots + b_n^{q_n}\} \quad (4)$$

其中 $b_i^j = b_i, i = 1, 2, \dots, n; j = 1, 2, \dots, q_i, q_i \in [1, n_c]$ 为一自适应参数, 也可以设定为一常数, n_c 为设定的克隆比例上限。下面举例说明本文中克隆操作的具体实现方式。

例2 假设抗体种群为

$$(5.1, 3.5, 1.4, 0.2, 4.9, 3.2)$$

$$(2.9, 4.3, 0.8, 5.2, 4.7, 1.6)$$

设克隆比例为2, 那么该抗体经过克隆操作之后变为

$$(5.1, 3.5, 1.4, 0.2, 4.9, 3.2)$$

$$(5.1, 3.5, 1.4, 0.2, 4.9, 3.2)$$

$$(2.9, 4.3, 0.8, 5.2, 4.7, 1.6)$$

$$(2.9, 4.3, 0.8, 5.2, 4.7, 1.6)$$

(2)免疫基因操作 克隆变异操作是对免疫系统学习识别外部模式、抗体基因变异和编辑过程的模拟。本文采用均匀变异策略, 设 $b_i = (b_{i1}, b_{i2}, \dots, b_{im})$ 是群体中的一个个体, 均匀性变异是先个体 b_i 中随机选择一个分量 b_{il} , 然后, 在一个定义区间 $[a_l, b_l]$ 中均匀随机地取一个数 b'_{il} 代替 b_{il} , 以得到变异产生的后代 b'_i , 即 $b'_i = (b_{i1}, \dots, b'_{il}, \dots, b_{im})$ 。本文对于克隆过后的抗体种群以概率 p_m 经历如例3所示的均匀变异过程。

例3 假设抗体(5.1, 3.5, 1.4, 0.2, 4.9, 3.2)以概率 p_m 选择将要进行克隆变异的基因位, 假设选

取的基因位为“3.5”,在事先定义的区间[0,8]中随机地取实数“2.9”以代替基因“3.5”,于是变异过后的抗体变为(5.1, 2.9, 1.4, 0.2, 4.9, 3.2)。

(3)克隆选择操作 克隆选择操作是从抗体各自克隆后的子代中选择优秀的个体,从而形成新的种群,对抗体群 $B(b_1^1, b_1^2, \dots, b_1^{q_1}, b_2^1, b_2^2, \dots, b_2^{q_2}, \dots, b_n^1, b_n^2, \dots, b_n^{q_n})$, 克隆选择操作定义如下

$$\{b_1^1 + b_1^2 + \dots + b_1^{q_1}\} + \{b_2^1 + b_2^2 + \dots + b_2^{q_2}\} + \dots + \{b_n^1 + b_n^2 + \dots + b_n^{q_n}\} \longrightarrow b_1' + b_2' + \dots + b_n' \quad (5)$$

可见,克隆选择操作是克隆操作的逆操作。同一个抗体 b_i 经过克隆后形成的亚群体在经过免疫基因操作后通过克隆选择操作实现局部的亲和度提高。具体如例4所描述:

例4 假设经过免疫基因操作后的亚群体为
(5.1, 2.9, 1.4, 0.2, 4.9, 3.2)
(4.1, 3.5, 1.4, 0.2, 4.9, 3.2)

若经过亲和度计算抗体(4.1, 3.5, 1.4, 0.2, 4.9, 3.2)的亲和度最高,则在该亚群体中抗体(4.1, 3.5, 1.4, 0.2, 4.9, 3.2)的被选择概率最大,则选择该抗体为下一代新的抗体。

(4)亲和度计算 抗体-抗原亲和度反映抗体与抗原之间的结合力。在人工免疫系统中,一般指候选解对问题的适应性度量。本文中的亲和度计算分两个部分完成:首先,根据每一代的聚类中心将所有样本数据划分到不同的类别当中。具体地,遵循式(1)的原则将点 $x_i, i = 1, 2, \dots, n$ 划分到类 $C_j, j \in \{1, 2, \dots, K\}$ 。每一代划分完成之后依据式(2)重新计算新的聚类中心 $z_1^*, z_2^*, \dots, z_K^*$ 。

举例说明如下:

例5 例1中假设的抗体所代表的第一个聚类中心为(5.1, 3.5)。在聚类的过程中将点(4.9, 3.7), (5.9, 3.0)划分到以(5.1, 3.5)为中心的类中,那么该类新的中心为 $((5.1 + 4.9 + 5.9)/3, (3.5 + 3.7 + 3.0)/3) = (5.3, 3.4)$ 新的中心(5.3, 3.4)将会取代旧的中心(5.1, 3.5)。

然后依据式(3)计算 K 个类别的相似度函数 μ , 亲和度函数定义为

$$f = 1/\mu \quad (6)$$

即最大化亲和度函数等价于最小化目标函数 μ 。

4 实验结果

本文用ICCA分别对两幅不同波段的SAR图像进行分割,将实验结果与KM^[7]以及GAC^[5]进行了比较分析。在所有的算法中,聚类数目是已知的。对于图像中的每个像素,依据文献[10]对图像提取基于灰度共生矩阵的3维特征。ICCA的参数设置如下:最大迭代代数500、种群大小为10、变异概率为 $2/(N \times K)$ 、克隆尺度为5;GAC的参数设置如下:最大迭代代数500、种群大小为50、变异概率为0.1。KM算法的最大迭代代数设为500。

图1(a)是位于新墨西哥州Albuquerque城附近的Rio Grande河流的Ku-band SAR图像,分辨率为1m,其中包含3类地物:植被、河流和农作物。图2(a)是一幅无人机载X-band SAR图像,空间分辨率为3m,其中包含4类地物:水体、城区和两种农作物。

对于图像分割的结果,在没有标准数据的情况下,可以用人工目测的方法进行比较,对比图1中1(b), 1(c), 1(d)结果,可以看到:(1)ICCA对河流的划分在3种算法中是最好的;(2)对于植被和农作物的划分,ICCA也比GAC和KM更加清晰。

从图2中3种算法的分割结果来看,采用KM时,错分现象比较严重,如图2(b)中,有一类农作物几乎没有被正确划分,其他3类的错分也很严重;GAC和ICCA的分割结果还比较令人满意,但是如图2(c)所示,GAC对城区以及水体的划分要劣于图2(d)中ICCA的结果。

为了进一步验证算法的有效性,本文还将3种算法进行分割精度的定量比较。对于图1(a)中的Ku-band SAR图像,我们随机选取3个 18×18 的

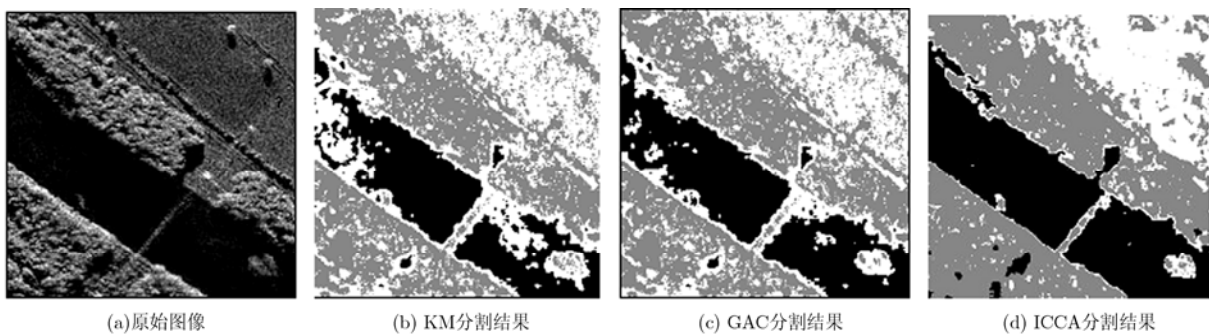


图1 1 m Ku-band SAR 图像分割(256×256)

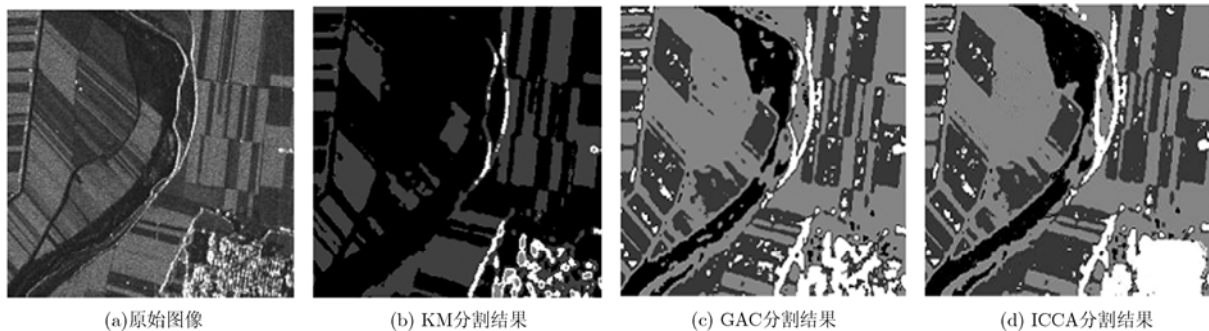


图 2 3 m X-band SAR 图像分割(256×256)

区域, 然后对这些区域用 3 种算法, 得到不同的分割结果, 并计算识别率, 这个概率可以粗略反映各种分割算法性能的优劣。对于图 2(a)中的 X-band SAR 图像, 我们随机选取 4 个 15×15 的区域, 同样得到 3 种算法的识别率, 结果如表 1 所示, 可以看出 ICCA 的分割精度要明显高于其他两种聚类方法。

表 1 3 种分割方法识别率比较(%)

	KM	GAC	ICCA
图 1	83.16	76.49	88.82
图 2	66.40	87.82	91.47

5 结束语

合成孔径雷达已成为国际对地观测领域最重要的前沿技术之一, 而图像分割是实现 SAR 图像信息提取和解译的关键步骤。本文提出了一种免疫克隆 SAR 图像分割方法。在新方法中, 通过借鉴生物学免疫系统的抗体克隆选择机理, 构造适合于人工智能的克隆算子, 利用免疫克隆优化算法搜索最优聚类中心, 新算法最终能以较大的概率获得问题的全局最优解, 且具有较快的收敛速度。通过对真实的 SAR 图像测试, 并与其它两种聚类方法 GAC、KM 相比较, 本文提出的方法都优于其他两种方法的分割效果, 这说明了基于 ICCA 的聚类方法能够很好地处理不同波段及拥有不同地物特征的 SAR 图像分割问题, 并具有较高的精度。

参考文献

- [1] Oliver C and Quegan S. Understanding Synthetic Aperture Radar Images [M]. Norwood: Artech House Inc, 1998: 88-204.
- [2] Dammert P B G, Asken J I H, and Kuhlmann S. Unsupervised segmentation of multitemporal interferometric SAR images[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 1999, 37(9): 2259-2271.
- [3] Dekker R J. Texture analysis and classification of ERS SAR images for map updating of urban areas in the Netherland[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2003, 41(9): 1950-1958.
- [4] Kersten P R, Lee Jong-Sen, and Ainsworth T L. Unsupervised classification of polarimetric Synthetic Aperture Radar images using fuzzy clustering and EM clustering[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2005, 43(3): 519-527.
- [5] Maulik U and Bandyopadhyay S. Genetic algorithm-based clustering technique[J]. *Pattern Recognition*, 2000, 33(9): 1455-1465.
- [6] Gong Mao-guo, Du Hai-feng, and Jiao Li-cheng. Optimal approximation of linear systems by artificial immune response[J]. *Science in China Series F-Information Sciences*, 2006, 49(1): 63-79.
- [7] Jain A K, Murty M N, and Flynn P J. Data, clustering: A review[J]. *ACM Computing Surveys*, 1993, 31(3): 264-323.
- [8] Pan H, Zhu J, and Han D. Genetic algorithms applied to multiclass clustering for gene expression data[J]. *Genomics, Proteomics & Bioinformatics*, 2003, 1(4): 279-287.
- [9] Hall L O, Ozyurt I B, and Bezdek J C. Clustering with a genetically optimized approach[J]. *IEEE Transactions on Evolutionary Computation*, 1999, 3(2): 103-112.
- [10] Clausi D A and Yue B. Comparing Cooccurrence Probabilities and Markov random fields for texture analysis of SAR sea ice imagery[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2004, 42(1): 215-228.

马文萍: 女, 1981 年生, 讲师, 研究方向为自然计算、图像处理。
 焦李成: 男, 1959 年生, 教授, 博士生导师, 主要研究方向为智能信息处理。
 尚荣华: 女, 1979 年生, 讲师, 研究方向为自然计算。