

自适应差分演化算法在图像监督分类中的应用

吴 佳¹ 蔡之华¹ 金晓文²

(1 中国地质大学(武汉)计算机学院,武汉市鲁磨路 388 号,430074)

(2 中国地质大学(武汉)环境学院,武汉市鲁磨路 388 号,430074)

摘 要:针对传统遥感图像分类算法存在约束条件多、容易陷入局部最优解、分类精度低的缺陷,提出了一种基于自适应差分演化的遥感分类新方法。实验结果表明,基于自适应差分演化的遥感图像分类算法在分类精度上优于传统方法,在收敛速度上优于标准的差分演化分类算法,其分类精度和 Kappa 系数分别达到了 92.66%和 0.901 7。

关键词:自适应;差分演化;遥感图像;监督分类;演化算法
中图分类号:P237.3

目前,差分演化算法(differential evolution algorithm, DE)已在诸多领域中得到应用^[1-4],在遥感图像处理方面,文献[5]提出了一种基于差分演化的无监督图像分类方法,并用美国 4 波段 Lake Tahoe 数据进行了实验分析。但差分演化在遥感图像监督分类方面的研究工作还很少。

在 DE 算法中使用到两个重要的人为设定的参数:变异因子 F 和交叉概率 CR。这两个参数的选取对算法的收敛性和收敛速率有很大的影响。自适应差分演化算法(self-adapting DE, SADE)在 DE 算法基础上引入参数自适应机制,使得 DE 具有自适应的能力,提高了收敛的速率。由于 SADE 算法能有效克服传统分类算法的缺点,因此本文尝试将自适应差分演化应用到遥感图像分类中,提出了一种基于自适应差分演化的遥感图像分类算法。

1 基于 SADE 算法的遥感图像分类

在标准的差分演化中,人工设定的参数有演化的最大代数 MaxGen、初始化种群规模 N 、变异因子 F 和交叉概率 CR 等 4 个。文献[6]通过实验系统分析了各个参数对算法性能的影响。针对 DE 算法中参数选取带来的问题,文献[7]提出了

一种自适应参数调整的机制,实验证明了基于自适应的差分演化算法 SADE 能有效提高 DE 算法的性能。本文提出的基于 SADE 的遥感图像分类的步骤如下。

1) 训练样区的选择。

2) 初始化种群。对于所有样区,其特征值表现为各个波段的灰度值,在确定好种群的大小 N 后,种群中的个体均是通过一定的机制随机产生的。假设有 K 个样本区域,则初始种群 G 中,单个个体 G_i 是由 K 个向量组成的向量组: $G_i = \{g_{ij}, j=1, 2, \dots, k\}$,其中 g_{ij} 为第 i 个个体所属类别的聚类中心向量,其维度为遥感图像的波段数 M 。各个波段取值为该样区中相同波段的最小值和最大值之间的随机数。

3) 变异操作。对第 t 代种群 G^t 中的个体进行变异操作,即从父代种群产生新的变异个体进入中间代。对于父代 G^t 中的任何一个个体 G_i^t , $i=1, 2, \dots, N$,新的变异个体 v_i^{t+1} 为:

$$v_i^{t+1} = g_{\text{best}}^t + F_i^t \times D_{r_1, r_2, r_3, r_4} \quad (1)$$

式中, F_i^t 为第 t 代中种群中第 i 个个体的变异因子,为 $(-1, 1)$ 上的一个均匀分布的随机数; g_{best}^t 为第 t 代种群中的最优个体; D_{r_1, r_2, r_3, r_4} 是进行变异操作的差分个体向量, D_{r_1, r_2, r_3, r_4} 被定义为:

$$D_{r_1, r_2, r_3, r_4} = g_{r_1}^t + g_{r_2}^t - g_{r_3}^t - g_{r_4}^t \quad (2)$$

收稿日期:2012-11-08。

项目来源:国家自然科学基金资助项目(61075063);国家教育部博士点基金资助项目(20090145110007);中国地质大学(武汉)优秀博士论文基金资助项目。

式中, r_1, r_2, r_3, r_4 是从 $\{1, 2, \dots, N\}$ 中随机选取的互不相同的整数, 且不等于 i 。 D_{r_1, r_2, r_3, r_4} 是一个均值为零的随机差分向量, 在此变异操作过程中, 必须保持其均值为零, 因为这样可以最大程度降低不可行解出现的概率。

4) 交叉操作。为了增加演化过程中种群的多样性, 使算法具有搜索到全局最优解的能力, 采用个体交叉操作。对第 t 代种群 G^t 中的第 i 个个体向量 g_i^t 与其对应的变异中间个体向量 v_i^{t+1} 按交叉策略进行交叉产生实验个体向量 u_i^{t+1} 。为保证获得的实验个体向量 u_i^{t+1} 在交叉操作中保持其演化特性, 必须保证 u_i^{t+1} 的维度变量中至少有一维由变异向量 v_i^{t+1} 提供。利用交叉概率因子 CR_i^t , 决定 u_i^{t+1} 中哪些维度变量由变异个体向量 v_i^{t+1} 提供, 哪些由目标个体向量 g_i^t 提供。其变异操作可表示为:

$$u_{ip}^{t+1} = \begin{cases} v_{ip}^{t+1}, \text{rand}(p) \leq CR_i^t \text{ 或 } p = \text{randn}(i) \\ g_{ip}^t, \text{rand}(p) > CR_i^t \text{ 且 } p \neq \text{randn}(i) \end{cases} \quad (3)$$

$(p = 1, 2, \dots, M)$

式中, $\text{rand}(p)$ 是 $[0, 1]$ 之间均匀分布的随机数; p 表示个体向量的维度; CR_i^t 表示第 t 代中第 i 个个体向量的第 p 维, 交叉概率因子在进化的过程中会不断地自适应调整; $\text{randn}(i)$ 表示 $[1, M]$ 之间生成的随机整数, 用以保证实验向量至少有一个维度变量是由变异向量 v_i^{t+1} 提供, 否则交叉操作产生的个体向量与目标向量 g_i^t 相同, 不能生成新的个体。插页 I 彩图 1 为一个 5 维变量的交叉操作的示意图。

5) 适应度函数值的计算。对于第 t 代种群中的个体向量 g_i^t , 计算样本区域中每一个个体像素点对应的个体向量到 g_i^t 中每一维向量的距离, 将其判决为欧氏距离最小的维度向量所属的类别。以此个体向量 g_i^t 为全局最优聚类中心得到样本的分类精度, 作为个体向量 g_i^t 的适应度, 适应度函数的计算过程可表示为:

$$f(g_i^t) = \frac{\sum_{j=1}^{N_{\text{train}}} (J_j^d - J_j + 1.0)}{N_{\text{train}}} \quad (4)$$

式中, J_j^d 是基于聚类中心向量 g_i^t , 样本区域中第 j 个样本的分类结果; J_j 为第 j 个样本的实际类别值。

6) 选择操作。为决定变异个体向量 v_i^{t+1} 是否能取代目标个体向量 g_i^t 成为第 $t+1$ 代种群中的新个体 g_i^{t+1} , DE 算法采用“贪婪”的搜索策略机制, 将个体 v_i^{t+1} 与 g_i^t 进行竞争, 只有当 v_i^{t+1} 的适应度优于目标个体 g_i^t 才被选作子代; 否则, 在

$t+1$ 代中保留个体 g_i^t 。选择操作为:

$$g_i^{t+1} = \begin{cases} u_i^{t+1}, f(u_i^{t+1}) < f(g_i^t) \\ g_i^t, f(u_i^{t+1}) \geq f(g_i^t) \end{cases} \quad (5)$$

式中, $f(g_i^t)$ 为第 t 代中第 i 的个体向量的适应度。

7) 自适应调整参数 F 和 CR 。为了提高算法的收敛性和收敛速度, 增强其稳定性, 引入一种自适应策略, 让 F 和 CR 这两个对算法性能有着很大影响的参数自适应地调整, 其表达式为:

$$F_i^{t+1} = \begin{cases} F_t + \text{rand}_1 \times F_u, \text{若 } \text{rand}_2 < T_1 \\ F_i^t, \text{其他} \end{cases} \quad (6)$$

$$CR_i^{t+1} = \begin{cases} \text{rand}_3, \text{若 } \text{rand}_4 < T_2 \\ CR_i^t, \text{其他} \end{cases} \quad (7)$$

式中, F_i^{t+1} 和 CR_i^{t+1} 分别表示第 t 代种群中第 i 个个体所对应的变异因子和交叉概率; $\text{rand}_j, j \in \{1, 2, 3, 4\}$ 是在 $[0, 1]$ 区间内服从均匀分布的随机数; T_1 与 T_2 时调整控制算法影响因子 F 和 CR 的概率, 在本文实验中都取值为 0.1; F_t 和 F_u 在本文实验中分别取值为 0.1 和 0.9, 可以使得第 $t+1$ 代种群中的第 i 个个体的变异因子 F_i^{t+1} 取区间 $[0.1, 1.0]$ 之间均匀分布的随机数, 交叉概率 CR_i^{t+1} 取区间 $[0, 1]$ 之间均匀分布的随机数。

8) 检验算法终止条件。从步骤 3)~步骤 6) 完成了一次种群的进化。重复此流程, 直到算法演化次数超过预先设定的最大演化代数 MaxGen 或演化过程中得到相同的最优结果次数连续超过给定阈值 T 。

9) 全局最优聚类中心个体向量的获取。演化出来的最优个体 G_{best} , 即对应样区中的第 i 维分量向量 $g_{\text{best}, i}$ 为第 i 个样区的聚类中心。

10) 遥感图像分类。算法对训练样区的样本训练完之后, 计算整幅遥感图像上的每一个像素点的个体向量与最优聚类中心个体每一维中心向量的距离, 并将像素点判决到距离最优的那一维中心向量所属的类别中。

2 实验与分析

2.1 实验数据

实验中采用的数据是 2009-10-05 获取的武汉地区的图像, 所使用的训练数据都是经过野外实地勘测进行标记的数据。插页 I 彩图 2 为原始遥感图像即武汉 TM(7 个波段影像, RGB5、4、3 伪彩色合成图)。

算法中各参数的选择如下: 演化的代数设为 100, 种群大小设为 30, 变异因子为 0.8, 交叉概率

因子设为 0.6。

实验是在内存 4 G, CPU 2.5 GHz 的 PC 机上完成的。为了避免选择待测样本时人为因素的影响,实验采用交叉验证的方法(10 runs of 10-fold cross validation),即用 9/10 的数据进行训练,1/10 的数据进行测试,重复此过程 10 次。利用本文提出的基于 SADE 的分类算法对插页 I 彩图 2 武汉 TM 影像进行分类,得到的分类结果见插页 I 彩图 3(a)。为了验证本文算法的有效性,利用两种被广泛应用的分类方法——最小距离法和最大似然法,对同一幅影像进行分类,得到的分类结果分别见插页 I 彩图 3(b)、图 3(c)。同时,为进一步证明其高效性,将 SADE 的分类算法与标准 DE 的分类算法结果(插页 I 彩图 3(d))进行比较,自适应差分演化方法能快速地收敛到全局最优解。

采用最小距离分类时,植被和房屋错分现象比较严重,如插页 I 彩图 3(b)中,在对图像左下角和右下角房屋和植被交叉比较多的区域进行分类时,许多植被分到了房屋类;采用最大似然法时,在对植被进行分类时出现了较多的错分现象,如插页 I 彩图 3(c)中间的植被在用最大似然法分类时很多被错分成湖泊类,是因为湖泊表面蓝藻物质和植被容易被错分。而采用基于 SADE 的分类算法对图像进行分类时,各种错分现象明

显减少,湖泊、房屋、植被都得到了正确的分类。因此,可以从人工目视判定上说明基于 SADE 的分类算法的分类效果要优于其他的分类方法。

2.2 分类精度比较

为了更进一步验证基于自适应差分演化的遥感图像分类算法的正确性,将本文分类方法与传统最小距离法、最大似然法和标准的 DE 分类方法进行分类精度的定量比较。比较方法采用常用的分类比较方法——混淆矩阵、总精度和 Kappa 系数。表 1 对 4 种分类方法混淆矩阵进行了比较;表 2 对 4 种分类方法的总精度和 Kappa 系数进行了比较。

由表 1 可以看出,最小距离法、最大似然法和标准的差分演化的混淆矩阵对角线元素之和为均小于基于 SADE 的分类算法的混淆矩阵的对角线元素之和,表明基于 SADE 的分类算法正确分类的像元数比传统分类方法有所增加。由表 2 可以看出,总精度分别由传统分类方法的 79.66%, 88.14%, 89.83% 提高到了基于自适应差分演化的分类算法的 92.66%;Kappa 系数也有了较大的提高,由传统分类方法的 0.727 5 提高到了基于自适应差分演化分类算法的 0.901 7。

从插页 I 彩图 4 可以看出,基于自适应差分演化分类算法的全局最优解的收敛能力和达到最优解所需要的代数都优于标准的差分演化分类算法。

表 1 4 种分类方法混淆矩阵的比较

Tab. 1 Comparison for Confusion Matrixes of Four Classification Method Mentioned Above

样本	自适应差分演化分类算法				最小距离分类算法				最大距离分类算法				标准差分演化分类算法			
	长江	湖泊	植被	房屋	长江	湖泊	植被	房屋	长江	湖泊	植被	房屋	长江	湖泊	植被	房屋
长江	41	0	0	0	41	0	0	0	41	0	0	0	41	0	0	0
湖泊	0	46	0	0	0	46	0	0	0	46	0	0	0	46	0	0
植被	0	0	45	7	0	0	34	19	0	0	45	15	0	0	42	9
房屋	0	0	6	32	0	0	17	20	0	0	6	24	0	0	9	30
合计	41	46	51	39	41	46	51	39	41	46	51	39	41	46	51	39
分类精度/(%)	100	100	88.24	82.05	100	100	66.67	51.28	100	100	88.24	38.46	100	100	82.35	76.92

由此可见,基于 SADE 分类算法的分类正确率明显要高于传统的分类方法。这是因为,最小距离法在分类时仅考虑局部特性,常常会陷入局部最优,从而对图像进行错分;最大似然法的前提是每一类的概率密度分布函数都是正态分布,但实际选择的样本数据可能与正态分布偏离较大,因此,影响了最大似然法的分类精度;而基于自适应差分演化的分类算法具有自组织和自学习的能力,从而对样本分布条件没有严格的限制,即使模式空间的分布出现锯齿状情况,该算法也可以根据具体的情况对模式集进行正确的分类。

表 2 4 种分类方法总精度和 Kappa 系数的比较

Tab. 2 Comparison for the Overall Accuracy and Kappa Coefficients of the Four Methods

评价	算法			
	最小距离	最大似然	差分演化	自适应差分演化
精度	79.66%	88.14%	89.83%	92.66%
Kappa 系数	0.727 5	0.840 5	0.863 9	0.901 7

3 结 语

本文尝试将差分演化方法用于遥感图像的分类当中,实验结果表明其取得了较好的效果。并在此基础上,引入自适应的策略,使算法更快速和

稳定地收敛到全局最优解。

参 考 文 献

- [1] Slowik A. Application of an Adaptive Differential Evolution Algorithm with Multiple Trial Vectors to Artificial Neural Network Training [J]. IEEE Transactions on Industrial Electronics, 2010, 58(8): 3 160-3 167
- [2] Abraham A, Das S, Konar A. Document Clustering Using Differential Evolution[C]. IEEE Congress on Evolutionary Computation, Vancouver, Canada, 2006
- [3] Das S, Suganthan P N. Differential Evolution: A Survey of the State-of-the-Art[J]. IEEE Transactions on Evolutionary Computation, 2010, 15(1): 4-31
- [4] Ruzek B, Kvasnicka M. Differential Evolution Algorithm in the Earthquake Hypocenter Location[J]. Pure and Applied Geophysics, 2001, 158(4): 667-693
- [5] Omran M G H, Engelbrecht A P, Salman A. Differential Evolution Methods for Unsupervised Image Classification[C]. IEEE Congress on Evolutionary Computation, Munchen, Germany, 2005
- [6] Gamperle R, Muller S D, Koumoutsakos P. A parameter study for differential evolution[C]. WSEAS International Conference on Advances in Intelligent Systems, Fuzzy Systems, Evolutionary Computation, Interlaken, Switzerland, 2002
- [7] Brest J, Greiner J, Greiner S. et al. Self-Adapting Control Parameters in Differential Evolution: A Comparative Study on Numerical Benchmark Problems[J]. IEEE Transactions on Evolutionary Computation, 2006, 10(6): 646-657

第一作者简介:吴佳,博士生,现主要从事土地利用和覆盖制图以及遥感技术应用研究。

E-mail:wujiawb@126.com, zhcai@cug.edu.cn

Self-Adapting Differential Evolution and Its Application on Remote Sensing Image Supervised Classification

WU Jia¹ CAI Zhihua¹ JIN Xiaowen²

(1 School of Computer, China University of Geosciences, 388 Lumo Road, Wuhan 430074, China)

(2 School of Environmental Studies, China University of Geosciences, 388 Lumo Road, Wuhan 430074, China)

Abstract: A new classification algorithm is proposed for remote sensing imagery based on the self-adapting differential evolution to be waged against the three main disadvantages of traditional classification algorithm for remote sensing image: multiple constraints, easy to fall into local optimal solution, lower classification accuracy. In the new method for supervised classification of multi/hyper-spectral remote sensing image, the globally optimal cluster centers are firstly learned by using the self-adapting differential evolution algorithm, and then the whole remote sensing image can be classified by the cluster centers. The proposed algorithm for classification of remote sensing image is based on the standard differential evolution. The experimental results show that the self-adapting differential evolution clustering algorithm has higher classification accuracy than the traditional classification algorithm of remote sensing image. The classification accuracy and the kappa coefficient can reach 92.66% and 0.901 7, which has some practical application value.

Key words: self-adapting; differential evolution; RS image; supervised classification; evolutionary algorithm

About the first author: WU Jia, Ph.D candidate, majors in land use/cover mapping and application of remote sensing technology.

E-mail: wujiawb@126.com, zhcai@cug.edu.cn

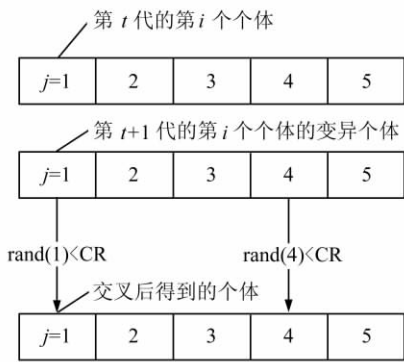


图 1 5 维变量的交叉过程

Fig. 1 Cross-process of 5-dimensional Variables



图 2 武汉 TM 影像 (7 个波段)

Fig. 2 TM Image of Wuhan (7 Bands)

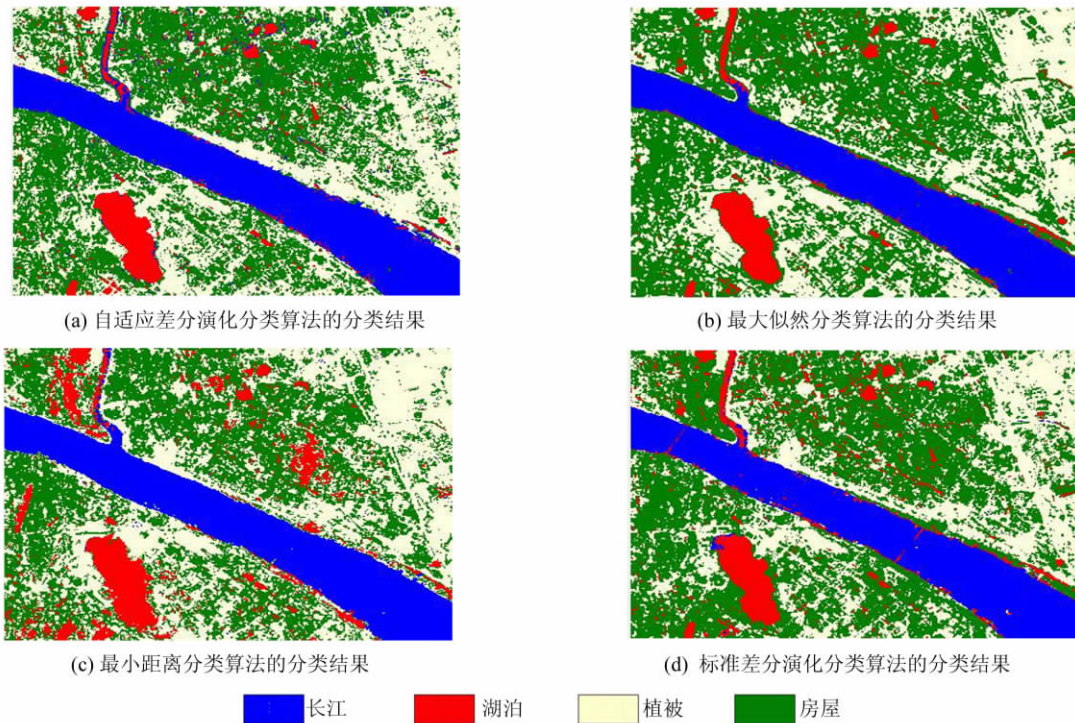


图 3 文中 4 种方法的分类结果图

Fig. 3 Classification Results of Four Methods in This Paper

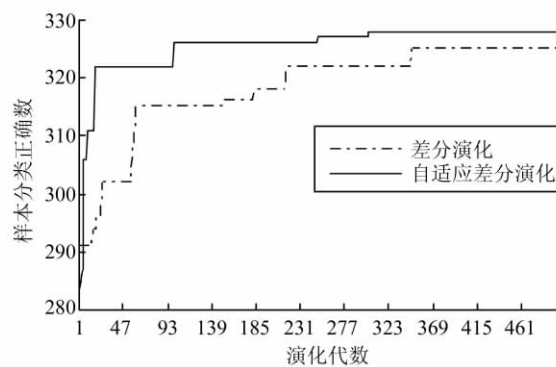


图 4 自适应差分演化与标准差分演化的最优解演化曲线

Fig. 4 Optimal Solution for Adaptive Differential Evolution and Standard Differential Evolution Algorithm