

基于 DE-GEP 的高光谱遥感图像分类

李 宠^{1,2}, 谷 琼¹, 蔡之华²

(1 湖北文理学院 数学与计算机科学学院, 湖北 襄阳 441053; 2 中国地质大学 计算机学院, 湖北 武汉 430074)

摘 要: 高光谱遥感数据具有波段数目多、数据量庞大等特点. 针对传统方法应用于高光谱图像分类中存在波段选择时计算量大、运行时间长, 以及图像分类精度不高等问题, 首先利用差分演化算法进行波段选择, 有效地降低了信息的冗余和数据的维度, 然后对波段选择后的结果成图, 并对要识别地物的典型区域进行取样, 最后采用基因表达式编程算法构建分类器进行图像分类. 在波段选择中, 与完全搜索的结果相比, 差分演化算法可以在很快的时间里取得了较好的搜索结果, 基因表达式编程在遥感图像分类中, 分类结果优于传统的 KNN 算法.

关键词: 遥感图像; 演化算法; 波段选择; 分类

中图分类号: TP751.1

文献标识码: A

文章编号: 1000-7180(2012)11-0103-04

Hyperspectral Remote Sensing Image Classification Based on DE and GEP

LI Chong^{1,2}, GU Qiong¹, CAI Zhi-hua²

(1 School of Mathematics and Computer Science, Hubei University of Arts and Science, Xiangyang 441053, China;

2 School of Computer, China University of Geosciences, Wuhan 430074, China)

Abstract: Hyperspectral data has the features of multi-band, large amounts of data, etc. For hyperspectral remote sensing image classification, traditional methods spend a long operation time in band selection, and image classification accuracy is not high, so first we use differential evolution(DE) for band selection, effectively reduce the redundancy of information and data dimensions, then show the image based on the result of band selection, sample the typical area of the surface feature to identify, finally use gene expression programming(GEP) to build classifier for image classification. In band selection, compared with full search algorithm, DE obtains relative good results in a fast time, in GEP image classification, the classification results superior to the traditional KNN algorithm.

Key words: remote sensing image; evolutionary algorithms; band selection; classification

1 引言

高光谱波段数众多、信息冗余度高等特点^[1-2], 一些适用于多光谱遥感数据的处理分析技术已不能适应高光谱数据的要求, 因此迫切需要一种高效率、智能化的高光谱遥感数据处理技术, 以降低特征空间维数、避免维数灾难, 提高分类的效率和精度.

从海量数据集中选择较具代表性的特征子集是一个 NP-难问题. 演化算法往往能够有效地解决

NP-难问题, 对结构对象进行操作, 不存在求导和函数连续性的限定; 具有内在的隐并行性和更好的全局寻优能力; 采用概率化的寻优方法, 能自动获取和指导优化的搜索空间, 不需要确定的规则^[3]. 演化算法的这些性质, 已被人们广泛地应用于组合优化、机器学习、信号处理和自适应控制等领域, 它是现代有关智能计算中的关键技术. 在遥感图像处理领域, 已有很多文献资料提出一些较成熟的应用演化算法方法.

收稿日期: 2012-05-03; **修回日期:** 2012-06-08

基金项目: 国家自然科学基金项目(61075063, 61172084); 湖北省自然科学基金项目(2010CDB05201)

针对演化算法能有效解决高光谱波段数众多、信息冗余度高等问题,本文选择遥感图像中的高光谱遥感图像作为处理对象,演化算法作为处理手段,即利用差分演化算法进行遥感图像波段选择,对波段选择后的结果采用基因表达式编程算法进行分类。

2 实验系统及试验区域

波段选择是为了减小计算量和数据的冗余,是分类前的预处理。分类是为了对图像更好的目视解译和识别,从而对地物有更好的理解。本文开发了“基于演化技术的遥感图像处理系统”软件处理遥感图像,处理流程如图 1 所示。

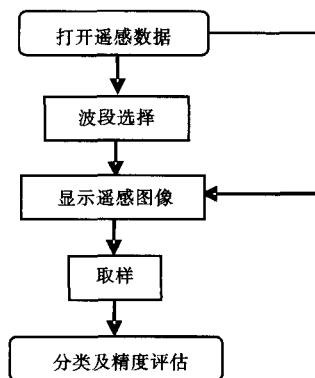


图 1 遥感图像处理流程

本文实验采用 EO-1(Earth Observing-1)卫星搭载的高光谱成像光谱仪 Hyperion^[4] 采集的高光谱数据,利用 FLAASH 2.0 大气校正软件来实现对 Hyperion 高光谱数据大气影响的去除^[5]。Hyperion 高光谱数据共有 242 个波段,像素数为 3400 × 256,本文只选取横坐标为 (1, 256),纵坐标为 (2800, 3400) 的感兴趣区域,大小为 256 × 601。所选的感兴趣区域内地物类别比较多,便于取样和检验算法分类的效果。

3 波段选择

3.1 差分演化算法

差分演化算法(Differential Evolution, DE)是一种新兴的演化计算技术,由 Storn 等人于 1995 年提出。DE 算法通过把种群中两个成员之间的加权差向量加到第三个成员上来产生新参数向量,该操作称为“变异”。然后将变异向量的参数与另外预先确定的目标向量参数按一定规则混合来产生试验向量,该操作称为“交叉”。若试验向量的代价函数比目标向量的代价函数低,试验向量就在下一代中代替

目标向量。最后的操作称为“选择”,种群中所有成员必须被作为目标向量这样操作一次,以便在下一代中出现相同个竞争者。在进化过程中对每一代都评价最佳参数向量,以记录最小化过程,这样利用随机偏差扰动产生新个体的方式可以获得一个有非常好收敛性的结果^[6]。

3.2 差分演化算法的策略选择及参数设定

编码:每个个体(波段子集)有三个基因(波段)^[7],每个基因值都是 [0, 1) 间的实数。以基因值 0.2 为例,假设总波段数为 150,则对应的波段为 $0.2 \times 150 = 30$,即基因值 0.2 对应波段 30。

适应度函数:选用协方差矩阵特征值法对波段进行评价。计算任意三波段组合的协方差矩阵行列式的值并进行由大到小排序,就可选出包含信息量丰富的三波段组合。

变异选用的是 DE/best/1/bin 变异策略:

$$v_{i,G+1} = x_{best,G} + F |x_{r1,G} - x_{r2,G}| \quad (1)$$

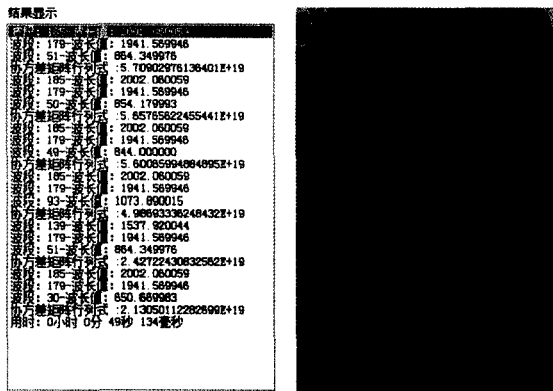
式中 v 为变异向量, x 为种群向量, G 为演化代数,随机选择的序号 r_1, r_2 互不相同, $x_{best,G}$ 为第 G 代种群 x 最优个体。变异算子 $F \in [0, 2]$ 是一个实常数缩放因子,控制偏差变量的放大作用。

差分演化算法采用一对一的选择策略。

参数设置:通过反复试验,设定交叉概率为 0.3,变异概率(缩放因子 F)为 0.5。综合考虑运行时间和收敛速度,种群大小为 30,运行代数为 40 代时搜索到或十分接近最优解。

3.3 波段选择结果分析

图 2 为差分演化算法的一次运行结果及对应协方差行列式最大值的三波段合成的 RGB 彩色图像。



(a) DE 运行结果 (b) 最大协方差合成 RGB

图 2 差分演化算法的运行结果及图像

从图 2 中可以看到,三波段协方差矩阵行列式最大合成的图像的信息量较大,图像色彩丰富,不同

地物间易于区分,并且中间的植被区域(绿色)中的细节纹理也比较清晰,有利于后续的分类和目视解译.这说明了协方差矩阵行列式作为高光谱图像波段选择算法的评价准则的有效性.

作为和差分演化算法的比较,本文选择了完全搜索,顾名思义,即算法计算所有可能的三波段组合,从中选出最优解.完全搜索算法由于效率很低不适用于实际应用,但是却可以作为衡量其它搜索算法的优劣的准则.完全搜索、差分演化算法在相同区域波段选择的结果如表 1 所示.

表 1 两种算法波段选择结果的比较

算法	时间 复杂度	运行 次数	运行 时间	协方差矩 阵行列式
完全搜索	$n(n-1)(n-2)$	1	4h30min	5.80E+19
差分演化	30 * 40	20	1min45s	5.78E+19

完全搜索运行一次即可求得最优解,而差分演化算法由于其算法的随机性,本文取其运行二十次的平均值,这样更接近真实值.从表 1 的结果可看出,完全搜索求出的结果是所有结果中的最优解,但花费的时间难以容忍(大于 4 小时).而差分演化算法在较快的时间里取得了较好的搜索结果.从而验证了差分演化算法在复杂解空间上进行有向随机搜索的方法的有效性.

4 图像分类

4.1 基因表达式编程的分类方法

本文采用 Candida Ferreira 提出的 0/1 舍入阈值的方法^[8]分类.就是当表达式返回值大于某个阈值的时候,就认为该表达式是一个正确分类,也就是属于我们分类的目标类别,否则,不属于此类别.

两类问题分类器的构建:读入训练样本集,评价当前一代群体中每个个体的适应度;第 i 个个体的输出值 $f(x) > 0$,则输入向量 x 属于某类(如河流类); $f(x) < 0$,则 x 不属于此类(非河流类).

个体的适应度函数为:

$$\text{fitness}(i) = \frac{\text{第 } i \text{ 个个体分对的样本数目}}{\text{训练样本总数}} \quad (2)$$

若干代后选择适应度最大的个体作为该类的分类器.

GEP 也可以解决类别多于两个的分类问题,将 n 个不同类别 C 的分类问题划分成 n 个二分类问题,运行 n 次,每次解决一个二分类问题:

1. C_1 vs Not C_1

2. C_2 vs Not C_2

...

n . C_n vs Not C_n

n 次进化的结果产生 n 个分类器,然后将这 n 个分类器结合起来,就得到了最终的分类规则.

4.2 GEP 分类参数设置

先对图像取样,将样本分为训练集和测试集, GEP 利用训练集构建分类器,通过测试集对构建完成的分类器计算其分类精度,检验分类器的有效性.然后用分类器对图像的每一个像素点进行分,分类完成后,输出分类后的图像和分类精度.

通过反复试验,设置 GEP 的参数如表 2 所示.

表 2 GEP 的参数设置

函数集	基因个数	基因长度	种群大小	演化代数	变异概率
+ - * / ^	2	21	40	1000	0.1
转座和插入序列			重组		
IS 转座	RIS 转座	基因转座	单点重组	两点重组	基因重组
0.1	0.1	0.1	0.3	0.3	0.1

实验表明,在进行针对分类的进化计算时,当个体中的函数集仅为数学函数时训练效果反而更好,故本文的函数集选为 $F = \{+, -, *, /, ^\}$,则函数运算符中变量数最多的参数个数 n 为 2. 基因头部长度 h 为 10,则尾部长度 $t = h(n-1) + 1 = 11$,基因长度 = 头部 + 尾部 = 21. 一个染色体的基因个数为 2,即染色体长度为 42.

GEP 的遗传参数比较多,包括变异、转座和插入序列元素、重组,这些概率会直接影响到是否能构造出好的分类器以及算法的收敛速度,表 2 中的这三个参数是通过反复的实验得出的.另外的两个参数种群大小和循环代数都是越大搜索结果越好,但是越大搜索时间就越长.因此,在确定了较好的变异、转座和插入序列元素、重组的概率后,先把循环代数设定的足够大并输出每一代的最优解,在此基础上调整种群大小.因此试验中选定种群大小为 40,这时,再观察收敛情况,基本在 1000 代时 GEP 就能得到比较好的分类器.

4.3 GEP 分类结果分析

作为和 GEP 算法的比较,本文选择了 KNN(K-Nearest Neighbor, K 最近邻)分类算法. KNN 是一种简便高效的算法,它用特征空间中的距离元素表示像元数据和分类类别特征的相似程度,利用训练数据获得了各个类别的特征参数以后,对于一个未知像元,计算它与各个类别特征向量的距离,把未

知像元并到距离最小的类别中去。

取样:通过对当地的土地利用现状的了解结合目视判读,确定试验区域典型地物包括河流、树林、山脊、沙地和坡地,分别选取五种具有代表性的地物类型样本,以不同的颜色表示。

由于 GEP 算法每次分类的结果并不相同,这里运行 10 次,取 10 次中最接近分类精度平均值的分类结果,如图 3 所示。

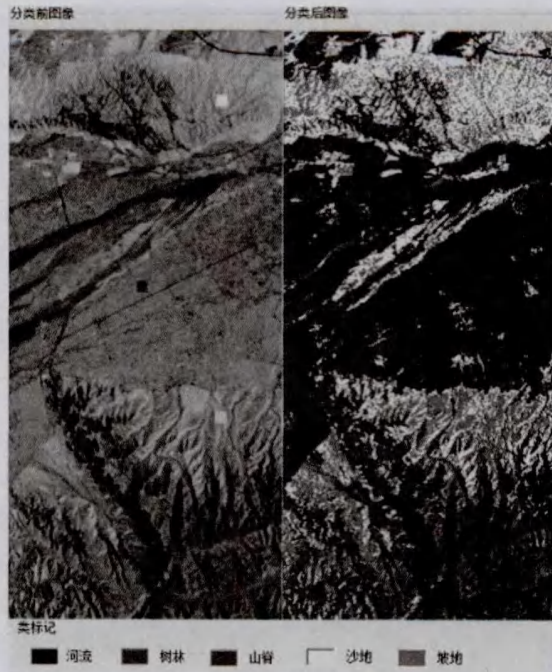


图 3 GEP 算法的分类结果

GEP 算法和 KNN 算法的混淆矩阵分别如表 3、表 4 所示。

表 3 GEP 算法的混淆矩阵

样本类别	河流	树林	山脊	沙地	坡地
河流	37	0	0	0	0
树林	0	39	0	0	0
山脊	0	0	32	4	2
沙地	0	0	2	33	7
坡地	0	0	0	7	32

混淆矩阵用来表示分类的结果,其中,混淆矩阵中对角线上的元素表示被正确分类的样本数目,非对角线上的元素表示被混分的样本数。从两种算法的分类结果图来看,五种地物区分的还是比较清楚的,结合表 3 和表 4,两种算法对河流和树林分类效果较好,采用 KNN 算法时,沙地分类效果较差,沙地错分为坡地比较严重。GEP 算法对沙地和坡地易

混分,但整体效果要比 KNN 算法好。

表 4 KNN 算法的混淆矩阵

样本类别	河流	树林	山脊	沙地	坡地
河流	37	0	0	0	0
树林	0	39	0	0	0
山脊	0	0	33	1	3
沙地	0	0	1	26	6
坡地	0	0	0	17	32

表 5 是两种算法的总体比较,分类精度等于测试集中正确分类的样本个数除以测试集的样本总数。KNN 算法是确定性算法,每次运行的结果相同。GEP 算法是随机性算法,表中 GEP 的运行时间和分类精度是运行 10 次取平均值的结果。从上表可知,GEP 算法在可以接受的时间内(大约 4 分钟)分类精度优于 KNN 算法。

表 5 KNN 算法和 GEP 算法比较

算法	运行时间	运行次数	分类精度(%)
KNN	7.85s	1	85.64
GEP	259.91s	10	87.32

5 结束语

差分演化算法、基因表达式编程两种技术虽然都遵循演化算法的基本原理,但是它们工程应用领域具有不同的功能:差分演化算法主要用于函数优化、搜索寻优,而基因表达式编程则主要用于建模、回归和分类等问题。高光谱数据的特点给处理带来了不便,本文结合差分演化算法和基因表达式编程两者的优点,利用差分演化算法进行波段选择,从而减少了计算量,大大提高了运行时间,在此基础上对波段选择后的结果采用基因表达式编程算法构建分类器进行分类,分类效果较好。由波段选择和分类的实例分析结果表明,演化算法在解决遥感图像处理问题中是可行而且是有效的。

参考文献:

- [1] 孙家炳. 遥感原理与应用[M]. 武汉:武汉大学出版社, 2009:1-20.
- [2] Antonio Plaza, Jon Atli Benediktsson. Recent advances in techniques for hyperspectral image processing[J]. Remote Sensing of Environment, 2009(113): S110-S122.
- [3] Poli R, Langdon W B, McPhee N F. A field guide to genetic programming[M]. Lulu.com, freely available from the internet, 2008. (下转第 111 页)

- wireless technology options and challenges[J]. Journal of Network and Computer Applications, 2012, 35(3): 1140-1150.
- [8] Vetrivelan P, Narayanasamy P, John Charlas J C. A multi-constraint real-time vehicular (MCRV) mobility framework for 4G heterogeneous vehicular Ad-Hoc networks[C]// IMECS2012, Hong Kong, 2012.
- [9] Faezipour M, Nourani M, Saeed A, et al. Progress and challenges in intelligent vehicle area networks[J]. Communications of the ACM, 2012, 55(2): 90-100.
- [10] 白翔宇, 叶新铭, 李军. 感知实时车流信息的 VANET 路由仿真研究[J]. 系统仿真学报, 2012, 24(2): 429-440.
- [11] Bai X Y, Ye X M, Li j, et al. VLS: a map-based vehicle location service for city environments[C]// IEEE ICC '09. Germany; Dresden, 2009
- [12] The network simulator ns 2. [EB/OL][2012-02-10]. <http://www.isi.edu/nsnam/ns/index.html>
- [13] SUMO-Simulation of urban mObility. [EB/OL][2012-01-30]. <http://sumo.sourceforge.net/>

作者简介:



白翔宇 男, (1976—), 博士, 副教授. 研究方向为移动与无线网络、车载自组织网.

(上接第 106 页)

- [4] 周雨霁, 田庆久. EO-1 Hyperion 高光谱数据的质量评价[J]. 地球信息科学, 2008, 10(5): 678-683.
- [5] 张秋爽, 祝民强, 刘碧洪. Hyperion 高光谱遥感数据大气校正方法[J]. 计算机测量与控制, 2010, 18(1): 220-222.
- [6] Das S, Suganthan P N. Differential evolution: a survey of the state-of-the-art[J]. IEEE Transactions on Evolutionary Computation, 2011, 15(1): 4-31.
- [7] 苏红军, 杜培军, 盛业华. 高光谱影像波段选择算法研究[J]. 计算机应用研究, 2008, 25(4): 1093-1096.
- [8] Ferreira C. Gene expression programming: mathematical modeling by an artificial intelligence[M]. Berlin Heidelberg: Springer-Verlag, 2006.

作者简介:

李 宠 男, (1988—), 硕士研究生. 研究方向为遥感图像处理、数据挖掘.