

基于混沌杜鹃搜索算法的高光谱影像波段选择

胡梦林¹, 万幼川¹, 王明威¹, 高贤君², 高 雄³

(1 武汉大学 遥感信息工程学院, 湖北 武汉 430079; 2 长江大学 地球科学学院, 湖北 武汉 430100; 3 中国冶金地质总局地球物理勘察院, 河北 保定 071051)

摘 要: 针对高光谱遥感影像的波段降维问题, 结合杜鹃搜索算法良好的全局寻优能力和混沌映射局部寻优能力强的特点, 提出一种基于混沌杜鹃搜索算法的高光谱影像波段选择方法, 并对 HYDICE 影像进行仿真实验. 同时进行该算法与遗传算法、粒子群算法和基本杜鹃搜索算法的对比实验, 实验结果表明该算法搜索能力更强, 最终所选波段子集的分类精度更高.

关键词: 高光谱影像; 波段选择; 杜鹃搜索算法; 混沌映射

中图分类号: TP751

文献标识码: A

文章编号: 1000—7180(2018)04—0124—06

Band Selection Based on Chaotic Cuckoo Search Algorithm for Hyperspectral Image

HU Meng-lin¹, WAN You-chuan¹, WANG Ming-wei¹,
GAO Xian-jun², GAO Xiong³

(1 College of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China;
2 College of Geoscience, Yangtze University, Wuhan 430100, China; 3 Geophysical Exploration
Academy of China Metallurgical Geology Bureau, Baoding 071051, China)

Abstract: A band selection technique based on chaotic cuckoo search algorithm is proposed in this paper, which combined with cuckoo search algorithm and chaotic map to solve the problem of dimension reduction of hyperspectral image. As a common used evolutionary algorithm, cuckoo search algorithm has the better global optimization ability, and chaotic map can improve the local optimization ability of evolutionary algorithm. Hence, a HYDICE image is utilized to test the efficiency of the proposed method. experimental results demonstrate that the search ability of chaotic cuckoo search algorithm is better than genetic algorithm, particle swarm optimization algorithm, and standard cuckoo search algorithm. Meanwhile, the classification accuracy could be increased by using the selected band subset.

Key words: hyperspectral image; band selection; cuckoo search algorithm; chaotic map

1 引言

高光谱遥感影像丰富的空间、辐射和光谱三重信息使其应用越来越广泛, 但波段数多、波段间相关度高且数据量庞大的问题也给影像的分类和目标识别带来了较大的困难. 因此在保证目标地物识别率

的情况下, 对高光谱影像进行降维处理则变得必要^[1]. 高光谱影像的降维方法主要包括波段提取和波段选择^[2-3]. 基于波段提取的方法在降维过程中会改变高光谱影像数据的原始特性. 而波段选择则是从原始波段中选择出具有代表意义的波段子集, 降低高光谱影像数据维度的同时不会改变原始影像的

收稿日期: 2017—06—23; **修回日期:** 2017—07—28

基金项目: 国家重点专项(2016YFC0600210); 测绘遥感信息工程国家重点实验室开放基金资助项目(13R04); 长江大学青年基金资助项目(2016cq04)

物理意义以及光谱特性,因而被广泛采用。

波段选择,本质上是一种复杂的组合优化问题,可采用群集智能算法进行求解^[4]。杜鹃搜索算法,作为一种新型元启发式群智能优化算法,具有参数少、全局搜索能力强易于实现和扩展的特点^[5]。Medjahed等^[6]采用杜鹃搜索算法分别对 Pavia University、Indian Pines 和 Salinas Valley 三幅高光谱影像数据进行波段选择,取得了较高的分类精度,同时实验结果表明算法具有很好的鲁棒性。然而杜鹃搜索算法作为一种全局搜索算法存在易陷入局部极值的问题^[7],针对这一问题,本文将混沌映射与杜鹃搜索算法进行结合。混沌映射具有伪随机性、遍历性、规律性和分形性的特点,是一种收敛速度快、搜索能力较强的方法,其伪随机性可使算法具有避免陷入局部极值的特点,其分形性使整体和局部具有自相似性,可对算法进行局部扰动。因此,本文提出一种基于混沌杜鹃搜索算法的高光谱影像波段选择方法,利用杜鹃搜索算法进行全局搜索,对质量较优解进行混沌扰动,通过混沌映射收敛速度快和避免陷入局部极值的特点,提高算法寻优能力,改善算法的性能。

2 混沌杜鹃搜索算法

2.1 杜鹃搜索算法

杜鹃搜索算法是由 Xin-She Yang 和 Suash Deb 提出的一种模拟杜鹃鸟寄生育雏行为的新型元启发式群智能优化算法^[5]。该算法源于杜鹃鸟的巢寄生繁殖方式和列维飞行(Lévy Flights)机制。巢寄生是指杜鹃鸟在繁殖期间将鸟蛋产在其他宿主鸟的鸟巢中,并由宿主鸟代替孵化和育雏的行为。列维飞行是杜鹃鸟在进行宿主鸟巢搜索时的随机游走过程,该搜索过程中步长随机,以小步长为主偶尔较大步长的行走交替进行,大步长可增加种群多样性、避免陷入局部最优,小步长可在群体中收敛到全局最优,是一种最为理想的搜索方式。

杜鹃鸟寻窝产卵过程满足3个理想化规则:(1)每只杜鹃鸟每次只产一枚鸟蛋,并随机选择宿主鸟巢放置;(2)只有孵化杜鹃鸟蛋最好的宿主鸟巢才会保留至下一代;(3)可选择的宿主鸟巢数量是确定的,宿主鸟发现外来鸟蛋的概率为 p_a 。规则3中宿主鸟一旦发现外来鸟蛋则推出外来鸟蛋或放弃这个鸟巢建筑新鸟巢。为简化过程,规则3近似为原鸟巢被新鸟巢代替的概率为 p_a 。

基于上述理想化规则,杜鹃鸟寻巢的位置更新

如下式:

$$x_i(t+1) = x_i(t) + \alpha \oplus \text{Levy}(\lambda) \quad (1)$$

式中, $x_i(t)$ 表示第 i 只杜鹃鸟第 t 次迭代的鸟巢位置, $\oplus$ 表示点对点乘法, α 表示步长,且 $\alpha > 0$,一般情况下取 $\alpha = 1$ 。 λ 为指数参数, $\text{Levy}(\lambda)$ 表示随机搜索路径,遵从式(2):

$$\text{Levy} \sim u = t^{-\lambda}, (1 < \lambda \leq 3) \quad (2)$$

式(2)是带有重尾的概率分布,通常采用 Mantegna^[8] 提出的模拟 Lévy 飞行路径的计算公式:

$$s = \mu/|\nu|^{1/\beta}, (0 < \beta < 2) \quad (3)$$

式中, s 即为飞行路径 $\text{Levy}(\lambda)$; 参数 β 一般取 1.5; 参数 μ, ν 服从正态分布。

2.2 混沌映射

混沌来源于非线性动力学系统,它是由确定性系统产生的一种运动状态,主要具有如下特征。(1)伪随机性。混沌的运动状态貌似随机,但每一时刻的状态是受前一时刻确定而出现,并非随机出现,因此混沌的随机性是具有确定性的。(2)遍历性。混沌能在有限区域内按自身规律不重复地遍历所有状态。(3)规律性。混沌的规律性体现在确定性方程的迭代。(4)初值的敏感依赖性。它是混沌的关键要素之一,初始条件的微小差异,在经过长期演化后也会使系统彻底偏离原来演化方向。(5)分形性。分形性是一种具有自相似性的现象,即其组成部分与整体相似。混沌的整体和局部之间也具有相似性。

基于以上性质,已有专家学者将混沌映射与粒子群算法等结合从而将其引入最优化问题^[9],结果较基本粒子群算法优化性能更好。本文将混沌映射与杜鹃搜索算法结合以解决高光谱影像波段选择的优化问题。混沌优化的基本思想是将混沌状态引入到优化变量中,是一个映射过程,优化变量的规律性伪随机遍历搜索比毫无规律的随机搜索更具优越性。

Logistic 映射作为一种有效且易操作的混沌系统而被广泛应用^[10],因此本文采用 Logistic 映射产生混沌变量,将其运动范围映射到鸟巢位置的变化范围,然后利用混沌变量对质量较优的鸟巢位置进行扰动,可改善杜鹃搜索算法易陷入局部极值的问题。Logistic 映射公式如下:

$$x_{i+1} = \mu \cdot x_i(1 - x_i) \quad (4)$$

式中, $x_i \in [0, 1]$ 表示映射过程中优化变量归一化后的混沌变量,当 $\mu = 4$ 且 $x_0 \notin \{0, 0.25, 0.5, 0.75, 1\}$ 时处于完全混沌状态。

3 基于混沌杜鹃搜索算法的波段选择

3.1 算法的编码形式

杜鹃搜索算法的提出最初是用于解决连续空间的优化问题,混沌变量的运动范围也为连续空间,但高光光谱影像的波段选择问题却属于离散组合优化问题的范畴,因此本文将杜鹃鸟搜索时宿主鸟巢的位置作为波段选择的最优解,对鸟巢位置采用两部分编码形式,分别为实数连续空间编码 X 和二进制离散编码 BX . X 用于算法的搜索, BX 中鸟巢位置的每一维分量分别用 1 或 0 表示,代表该维波段是否被选中. 实数编码的位置维分量 x_{id} 通过 Sigmoid 函数计算得到离散编码中分量 bx_{id} 取值为 0 或 1 的概率,转换公式如下:

$$\text{Sigmoid}(x_{id}) = 1 / (1 + \exp(-x_{id})) \quad (5)$$

$$bx_{id} = \begin{cases} 1, \text{rand}() < \text{Sigmoid}(x_{id}) \\ 0, \text{rand}() \geq \text{Sigmoid}(x_{id}) \end{cases} \quad (6)$$

式(5)为 Sigmoid 函数的定义公式,式(6)中 $\text{rand}()$ 为区间 $[0,1]$ 之间的随机数.

3.2 目标函数

高光光谱影像的降维目的在于选择尽可能少的波段,获取更高的分类精度,因此本文综合考虑分类精度和所选波段数选用目标函数如式(7).

$$F(bx_i) = \text{Accuracy}(bx_i) / (1 + \theta \cdot n(bx_i)) \quad (7)$$

$$\text{Accuracy}(bx_i) = F_P / (F_P + F_N) \quad (8)$$

式(7)中, $n(bx_i)$ 为第 i 个鸟巢离散位置中所选的高光谱影像波段个数; θ 为波段个数的权重参数,一般取 0.01. 式(8)为根据 bx_i 中所选波段子集进行分类所得的分类精度; F_P 为正确分类象元数; F_N 为错误分类象元数.

3.3 基于混沌杜鹃搜索算法的波段选择流程

本文采用混沌杜鹃搜索算法对高光光谱遥感影像进行波段选择实验,将波段选择问题抽象成杜鹃鸟巢寄生繁殖时对宿主鸟巢位置进行搜索的过程. 迭代时鸟巢位置质量的判断以目标函数式(7)为准. 鸟巢位置的搜索与更新过程如下: (1) 设置算法所需参数,产生初始鸟巢位置; (2) 计算鸟巢位置目标函数值,选取部分质量较优的位置根据式(4)进行混沌扰动,对剩余位置根据式(1)~(3)进行杜鹃搜索算法的全局更新; (3) 比较现有鸟巢位置和上一代鸟巢位置,若现有位置更好,则将其更新为当前鸟巢位置; (4) 将随机数 $r \in [0,1]$ 与鸟巢主人发现外来卵的概率 p_a 进行比较,若 $r > p_a$ 则随机产生新位置作为鸟巢位置; (5) 计算鸟巢位置的目标函数值,并记录当前最优鸟巢位置. 当鸟巢位置更新后判断循环

终止条件,若终止则算法运行结束,输出最优鸟巢位置,即所选波段子集,并对该波段子集组成的影像进行分类和精度评价,反之继续鸟巢位置的更新.

4 实验与结果分析

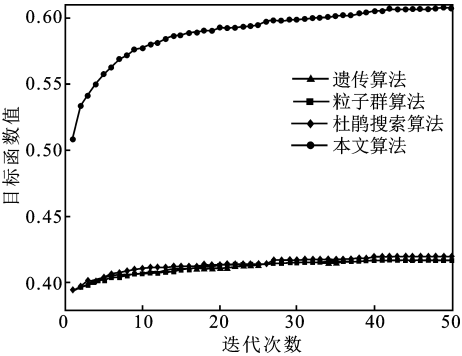
本实验采用超光谱图像收集实验仪器 HYDICE 拍摄于美国华盛顿地区的数据进行算法的仿真实验. 该数据的波长范围为 $0.4 \sim 2.4 \mu\text{m}$, 光谱分辨率为 10 nm, 空间分辨率为 2.8 m, 去除噪声波段和水吸收波段后,共有 191 个波段. 结合该地区的地物真实报告,根据本文所选区实际情况,选取七类典型地物进行样本点的提取,并设置训练样本/测试样本比例为 50%/50% 进行分类实验. 整个实验在 Intel Core i3, 3.20 GHz CPU, 内存为 2 GB 的 PC 机上进行,算法采用 Matlab R2014a 编程实现.

为比较本文混沌杜鹃搜索算法的有效性,选择遗传算法、粒子群算法和杜鹃搜索算法进行对比实验. 不同算法所需参数: 本文算法中鸟巢规模 $n = 20$, 混沌扰动的鸟巢个数 $m = 5$, 发现外来卵概率 $p_a = 0.25$; 遗传算法中交叉概率 $p_c = 0.9$, 变异概率 $p_m = 0.05$; 粒子群算法中感知因子 $c_1 = 2$, 社会因子 $c_2 = 2$, 惯性权重 $\omega \in [0.9, 0.4]$ 线性递减; 杜鹃搜索算法参数同本文算法,所有算法中种群规模 $\text{psize} = n = 20$, 终止迭代次数 $N = 50$. 以下实验中算法均独立运行 30 次,结果取其均值表示.

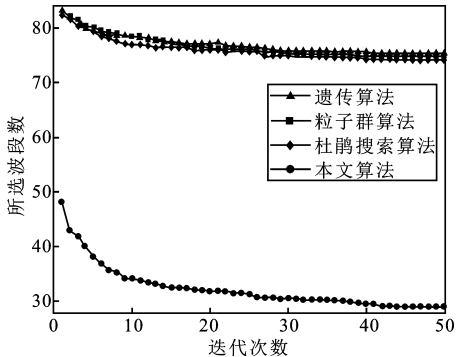
4.1 不固定波段数的高光谱影像波段选择

本实验所选 HYDICE 影像包含 191 个波段,维度较高,易造成维数灾难,因此采用混沌杜鹃搜索算法进行不固定数目的波段选择,并与其他算法进行对比. 图 1 为目标函数值与所选波段数随迭代次数变化的曲线图,图中遗传算法、粒子群算法和杜鹃搜索算法的变化曲线相近,本文算法较其他三种算法目标函数值更高,所选波段数也更少,同时本文算法曲线的变化幅度较其他三种算法更大,表明本文算法的搜索范围更大,搜索能力更强. 表 1 所示为不同算法迭代终止时目标函数值、总体分类精度、所选波段数和运行时间的比较,从表中数据可看出本文算法较其他三种算法优化性能更好,虽然比遗传算法、粒子群算法花费的运行时间要长,但目标函数值和总体分类精度更高,所选波段数更少;与杜鹃搜索算法相比不仅运行时间更短,目标函数值和总体分类精度分别高出 0.187 6 和 5.242 0%,所选波段数也减少了 60%,同时所得值方差更小,表明本文算法不仅提高了杜鹃搜索算法的运行效率和搜索能力,

还改善了杜鹃搜索算法中 Lévy 飞行搜索导致算法难以稳定收敛到最优解的问题。



(a) 目标函数值随迭代次数的变化曲线图



(b) 所选波段数随迭代次数的变化曲线图

图 1 不同算法中目标函数与所选波段数随迭代次数的变化曲线图

表 1 不同算法迭代终止时的比较

	目标函数值		总体分类精度/%		所选波段数		运行时间/ s
	均值	方差	均值	方差	均值	方差	
遗传算法	0.417 6	0.007 5	73.178 1	0.500 3	75.27	2.66	284.924 2
粒子群算法	0.416 9	0.007 8	72.885 8	0.604 0	74.87	2.78	309.968 3
杜鹃搜索算法	0.419 8	0.008 6	73.123 3	0.654 7	74.23	6.08	486.714 7
本文算法	0.607 4	0.007 7	78.365 3	0.535 3	29.03	1.69	424.175 7

图 2 中目标函数值随波段数的增加呈先升后降的趋势,这与波段维度造成的 hughes 现象所一致.图中虚线表明当波段数为 3 时,目标函数值最大,由此确定所选波段数以进行下一步实验.在确定所选波段数的基础上,分别使用不同算法进行波段选择,并从算法最终所选波段子集的目标函数值、总体分类精度、信息量和相关性方面进行对比分析.其中信息量由波段子集的最佳指数因子(Optimum Index Factor, OIF)值表示,相关性由波段子集中波段间的平均相关性表示。

图 3 为确定所选波段数为 3 时目标函数值随迭代次数的变化曲线,图中可看出本文算法的目标函数值较其他三种算法更高.表 2 所示为不同算法迭代终止时参数的比较,与遗传算法和粒子群算法相比,本文算法运行时间虽更长,但平均单个粒子单次迭代时间仅

4.2 固定波段数的高光谱影像波段选择

不固定波段数的实验中本文算法所选波段数均值为 29.03,目标函数值最大时对应波段数为 26,该波段数较原始 191 个波段减少了近 85%的数据量,但在实际应用中 26 个波段仍属多光谱范围,其数据量仍然较大,需要进一步进行实验.考虑到高光谱影像的波段数造成的 hughes 现象,以不固定波段数的实验结果为基础,缩小目标波段数范围为 1~26,然后在该范围内分别固定波段数为 1,2,⋯,26 进行本文算法的波段选择实验,根据目标函数值的变化从而确定最终所选波段数,实验结果如图 2 所示。

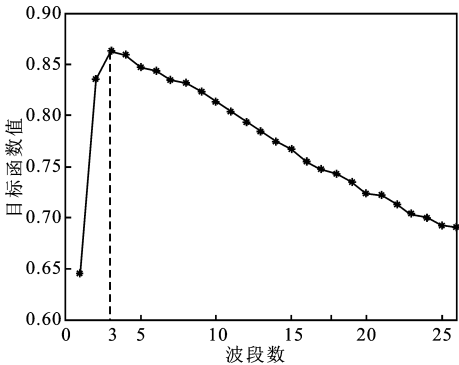


图 2 目标函数值随波段数的变化曲线图

比遗传算法多用 0.09 s,且所得目标函数值和总体分类精度更高;与杜鹃搜索算法相比,本文算法运行效率更高,同时目标函数值和总体分类精度也高出了 0.006 6 和 0.684 9%。

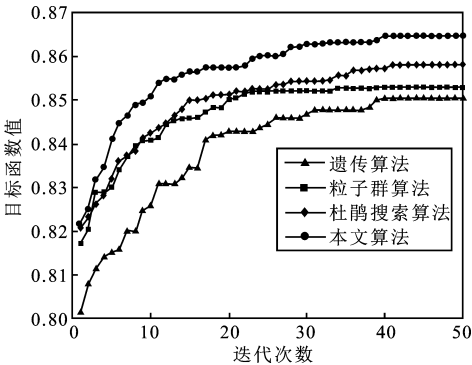


图 3 目标函数值随迭代次数的变化曲线图

表 3 所示为不同算法最终所选波段子集进行分类所得的相关信息,从表中数据可看出相比于遗传算法、粒子群算法和基本杜鹃搜索算法,本文算法最终所选波段子集的目标函数值更高,总体分类精度更好,且仅有本文算法 OIF 值高于 3.10×10^3 ,平均相关性低于 0.57,即本文算法性能更好. 为更直观反映本文算法与其他算法性能的对比,将各算法所选波段子集的分类结果示于图 4. 从影像上可以看出,图 4 中本文算法分类结果较遗传算法、粒子群算法和杜鹃搜索算法分类结果更接近原始影像,且更加完整,如图中方块标记区域所示,本文算法错分象元数目更少,边界更清晰.

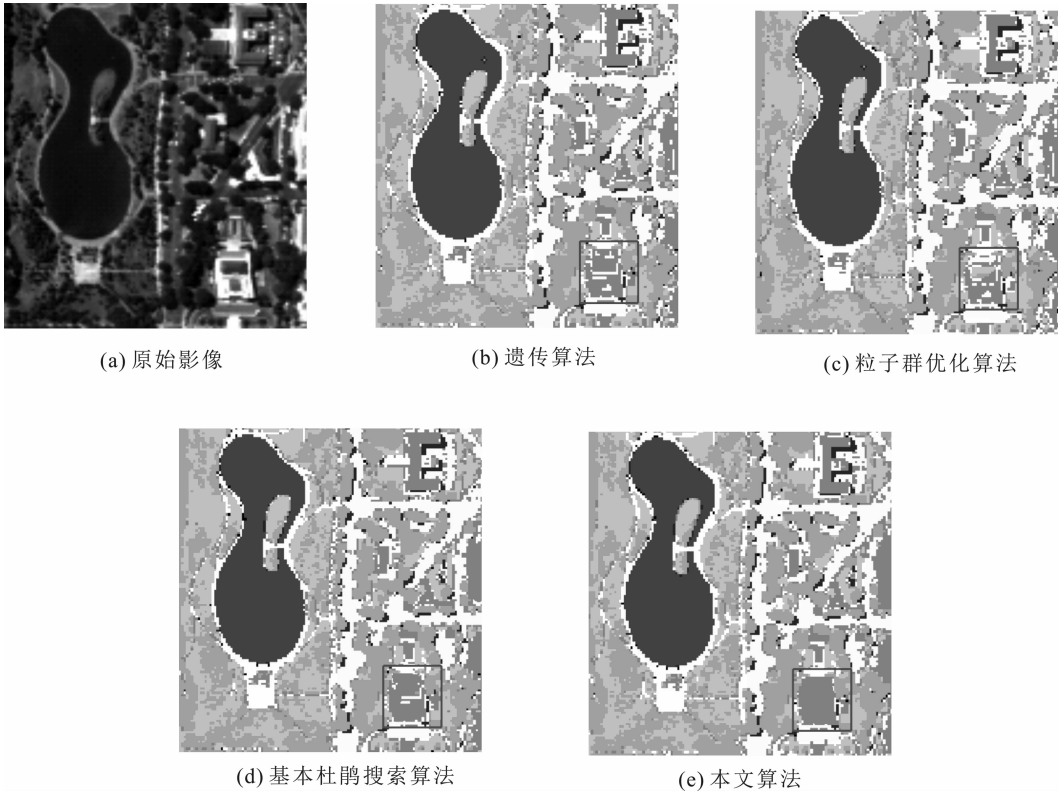


图 4 原始影像和不同算法分类图

5 结束语

针对高光谱遥感影像波段维数较高,易产生“维度灾害”的问题,本文结合杜鹃搜索算法较优的全局搜索能力,混沌映射收敛速度快且避免陷入局部极值的特点,提出了一种基于混沌杜鹃搜索算法的波段选择方法. 并分别利用遗传算法、粒子群算法和基本杜鹃搜索算法对高光谱影像进行了仿真对比实验,结果表明本文算法较其他三种算法搜索能力更强,收敛性能更好,且最终所选波段子集的分类精度

表 2 不同算法迭代终止时的比较

	目标函数值		总体分类精度/%		运行时间/s
	均值	方差	均值	方差	
遗传算法	0.850 6	0.009 3	87.607 3	0.956 4	211.404 0
粒子群算法	0.853 0	0.009 2	87.853 9	0.947 2	227.992 3
杜鹃搜索算法	0.858 2	0.011 3	88.392 7	1.163 4	368.867 5
本文算法	0.864 8	0.009 2	89.077 6	0.948 3	297.826 7

表 3 不同算法所选波段子集的对应信息

	所选波段	目标函数值	总体分类精度/%	信息量OIF ($\times 10^3$)	平均相关性
遗传算法	(39,55,66)	0.869 8	89.582 3	3.074 6	0.613 0
粒子群算法	(40,55,73)	0.856 5	88.224 0	2.557 2	0.578 4
杜鹃搜索算法	(39,56,67)	0.872 5	89.861 9	3.094 0	0.621 9
本文算法	(39,54,65)	0.877 8	90.414 7	3.140 5	0.565 3

更高,更加满足实际需求.

参考文献:

[1] Arslan O, AKYÜREK Ö, KAYA S. A comparative analysis of classification methods for hyperspectral images generated with conventional dimension reduction methods[J]. Turkish Journal of Electrical Engineering & Computer Sciences, 2017, 25(1): 58-72.

[2] Datta A, Ghosh S, Ghosh A. Unsupervised band extraction for hyperspectral images using clustering and

- kernel principal component analysis[J]. International Journal of Remote Sensing, 2017, 38(3): 850-873.
- [3] Cao X, Li X, Li Z, et al. Hyperspectral band selection with objective image quality assessment[J]. International Journal of Remote Sensing, 2017, 38(12): 3656-3668.
- [4] Xu M, Sun Q, He Z, et al. Band selection for hyperspectral images based on particle swarm optimization and differential evolution algorithms with hybrid encoding[J]. Journal of Computational Methods in Sciences and Engineering, 2016, 16(3): 629-640.
- [5] Yang X S, Deb S. Cuckoo search via Lévy flights [C]//Nature & Biologically Inspired Computing, 2009. IEEE, 2009: 210-214.
- [6] Medjahed S A, Saadi T A, Benyettou A, et al. Binary cuckoo search algorithm for band selection in hyperspectral image classification[J]. IAENG International Journal of Computer Science, 2015, 42(3): 183-191.
- [7] Civicioglu P, Besdok E. A conceptual comparison of the Cuckoo-search, particle swarm optimization, differential evolution and artificial bee colony algorithms [J]. Artificial intelligence review, 2013, 39(4): 1-32.
- [8] Mantegna R N. Fast, accurate algorithm for numerical simulation of Levy stable stochastic processes [J]. Physical Review E, 1994, 49(5): 4677.
- [9] 吴超, 吴一全. 基于混沌粒子群优化投影寻踪的高光谱图像目标检测[J]. 光学学报, 2011, 31(12): 63-70.
- [10] 陈志刚, 梁涤青, 邓小鸿, 等. Logistic 混沌映射性能分析与改进[J]. 电子与信息学报, 2016, 38(6): 1547-1551.

作者简介:

胡梦林 女,(1992-),硕士研究生. 研究方向为遥感图像处理、人工智能. E-mail: humenglin@whu.edu.cn.

万幼川 男,(1960-),教授,博士生导师. 研究方向为遥感数据处理与应用、摄影测量与 LIDAR 数据处理、地理信息系统开发与环境信息系统.

王明威 男,(1989-),博士研究生. 研究方向为图像处理、人工智能.

高贤君 女,(1986-),讲师. 研究方向为遥感影像处理.

高 雄 男,(1986-),工程师. 研究方向为机器学习.

(上接第 123 页)

- [6] 郭强, 张晓萌, 朱若函. 基于粒子群优化的模糊神经网络垂直切换算法[J]. 微电子学与计算机, 2015, 32(8):12-18.
- [7] 孔令斌, 王军选, 孙长印. 多网络共存下基于贝叶斯的垂直切换算法[J]. 电讯技术, 2016, 56(2): 122-127.
- [8] Lte. 3GPP TR 25. 942, Radio frequency (RF) system scenarios (Release 10) [S]. Lte, 2011.
- [9] Srinivasan Roshni, Zhuang Jeff, Jalloul Louay , et al. IEEE802. 16 m-08/004r5, IEEE 802. 16m evaluation methodology document [EB/OL]. [2017-03-28]. [http. //www. ieee802. org](http://www.ieee802.org).
- [10] IEEE. IEEE 802. 21-2008, IEEE standard for local and metro-politan area networks-part 21: media independent handover[S]. IEEE, 2008.
- [11] Kim Younghyun, Pack Sangheon, Kang Chung Gu, et al. An enhanced information server for seamless vertical handover in IEEE 802. 21 MIH networks[J]. Computer Networks, 2011, 55(1):147-158.
- [12] 张吉军. 模糊层次分析法[J]. 模糊系统与数学, 2000, 14(2):80-88.
- [13] 陶剑骅, 武向农, 蒋祥鹏. 基于多属性决策的异构网络选择算法比较研究[J]. 电视技术, 2015, 39(17): 79-83.

作者简介:

杨明极 男,(1971-),博士,教授. 研究方向为移动通信系统.

吴 烨(通讯作者) 男,(1993-),硕士研究生. 研究方向为无线通信技术. E-mail:wuye930825@163.com.

范华峰 男,(1976-),工程师. 研究方向为无线控制与通信系统.