

改进粒子群优化算法的高光谱波段选择

张 刘¹, 叶 楠¹, 马灵玲², 汪 琪², 吕雪莹¹, 章家保^{1*}

1. 吉林大学仪器科学与电气工程学院, 吉林 长春 130061

2. 中国科学院空天信息创新研究院中国科学院定量遥感信息技术重点实验室, 北京 100094

摘 要 高光谱图像具有数百个连续、狭窄的光谱带, 光谱范围跨越可见光到红外光, 可提供地物的精细光谱属性, 对于地物材质和属性的识别分类具有重要应用价值。针对感兴趣目标选择有限的光谱波段进行传输和处理, 对于提升高光谱数据处理时效性、以及设计面向特定应用的实用化光谱仪都具有重要意义。而如何结合目标特征选择最优波段成为在提升处理效率的同时保证目标识别或分类精度的必然要求。因此如何从数以百计维度的高光谱图像中选择出具有较好分类识别能力的波段子集是急需解决的问题。提出基于改进粒子群优化算法的高光谱波段选择方法, 该方法区别于传统的粒子群优化算法, 引入“概率突跳特性”, 并设定新解的淘汰机制, 将“停滞”的新解进行淘汰, 提高了算法的全局寻优性能。然后基于目标光谱特征采用了最优波段选择的优化目标函数, 通过改进的粒子群优化算法求解目标函数, 并将选定的波段子集反馈到支持向量机(SVM)中执行分类应用。采用两个标准的高光谱数据集(Indian Pines, Salinas)对选择出的波段子集进行分类测试, 结果表明该方法相较于现有方法具有较高的分类精度, 在几种方法中, 传统的粒子群算法筛选出的波段效果最差; 该算法筛选出的波段的分类精度最好, 两个数据集的分类精度分别可以达到 98.141 4% 和 99.084 8%。

关键词 高光谱图像; 波段选择; 粒子群算法; 支持向量机

中图分类号: TP751.1 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2021)10-3194-06

引 言

由于高光谱数据量十分庞大, 受限于通信带宽和处理速度, 往往无法实时传输处理全部数据。对于有实时处理需求的任务, 高光谱载荷通常仅实时上传数个谱段进行快速处理。因此, 如何选择有代表性的谱段, 使得在选出的波段子集上能高精度地完成实时地物分类和目标检测任务, 是高光谱遥感应用中待解决的一大问题^[1-2]。

高光谱图像波段选择的机理可看成是一个非线性函数求解极值点问题。近年来, 许多优化方法应用到高光谱波段选择上, 例如遗传算法^[3-4]、差分算法^[5]、杜鹃搜索^[6]、蚁群算法^[7]、粒子群算法^[8-9]等。Clerc^[10]将混合编码差分算法与粒子群算法结合用于高光谱图像降维。王立国等^[7]利用蚁群算法结合遗传算法进行高光谱波段选择, 认为是一种兼顾时间

效率和精解效率的有效方法。

粒子群优化算法是一种仿生自然现象的元启发式算法, 具有结构简单、参数少易于操作、搜索速度快的优势。Ding^[8]将粒子群算法应用在高光谱波段选择中, 但受限于传统粒子群算法易陷入局部极值和早熟的问题, 所选取的波段不是全局最佳值。为了解决该问题。高鹰等^[11]将模拟退火算法与 PSO 算法结合(SAPSO), 通过模拟退火算法的“概率突跳特性”使传统粒子群算法具有了跳出局部极值的能力, 但其产生新解的邻域仍集中在局部区域内, 降低了对新解的利用率, 同时在针对高维复杂函数全局寻优时, SAPSO 易出现早熟和稳定性差的问题^[12]以致于无法寻找到全局最优值。

针对传统粒子群算法易陷入局部极值和稳定性较差问题, 提出了一种改进粒子群优化算法(MPSO), 在利用模拟退火-粒子群算法中“概率突跳特性”的同时结合“淘汰机制”丰富种群的多样性, 有望提高算法的全局搜索能力和稳定

收稿日期: 2020-10-06, **修订日期:** 2021-02-23

基金项目: 科工局国防基础科研重点突破项目(JCKY2018 * * * * 036), 装备预研领域基金项目(61404140505), 国防基础科研计划项目(JCKY20191110), 国家自然科学基金青年基金项目(41504135), 重点研发计划课题(2017YFC1405402), 中科院先导项目(XDA13030402), 上海航天科技创新基金项目(SAST2018046)资助

作者简介: 张 刘, 1978 年生, 吉林大学仪器科学与电气工程学院教授 e-mail: zhangliu@jlu.edu.cn

* 通讯作者 e-mail: zhangjiabao@jlu.edu.cn

性, 以更加精准、高效地检索到高光谱有效波段。充分考虑高光谱数据中不同地物的类间距离构建了优化目标函数, 采用改进粒子群算法进行全局寻优, 以更精确的选取出波段组合。

1 改进粒子群优化算法

1.1 传统粒子群优化算法

种群由 i 个粒子组成, 每个粒子携带 D 维信息, 粒子当前位置视为优化问题中的一个可行解, 粒子的移动仅与位置和速度相关, 分别用 x_i^t 和 v_i^t 表示粒子的位置和速度

$$x_i^t = (x_{i1}^t, x_{i2}^t, x_{i3}^t, \dots, x_{iD}^t)^T \quad (1)$$

$$x_{id}^t \in [L_d, U_d]$$

其中, L_d 和 U_d 为粒子可行解范围的上限和下限。

$$v_i^t = (v_{i1}^t, v_{i2}^t, v_{i3}^t, \dots, v_{iD}^t)^T \quad (2)$$

$$v_{id}^t \in [v_{\min}, v_{\max,d}]$$

式(2)中, v_{\min} 和 v_{\max} 为粒子搜索速度的上限和下限。

为获取优化的最佳解, 所有粒子需要不断的进行迭代以更新种群的位置和速度, 速度和位置的更新公式如式(3)和式(4)所示

$$v_{id}^{t+1} = \omega v_{id}^t + C_1 R_1 (p_{id}^t - x_{id}^t) + C_2 R_2 (p_{gd}^t - x_{id}^t) \quad (3)$$

$$x_{id}^{t+1} = x_{id}^t + v_{id}^{t+1} \quad (4)$$

式中: ω 为惯性因子用以平衡局部寻优能力和全局寻优能力之间的关系; C_1 和 C_2 为加速度项分别代表个体学习因子和社会学习因子; R_1 和 R_2 随机取值于 $[0, 1]$ 之间的数; v_{id}^{t+1} 为第 d 维的第 i 个粒子在 $t+1$ 时刻的迭代速度; v_{id}^t 为第 d 维的第 i 个粒子的在 t 时刻的迭代速度。

通过粒子的速度更新公式可知, 在搜索过程中所有粒子仅向全局最优值和自身历史最佳值的方向步进, 在迭代后期种群中的粒子趋向于同一位置, 粒子失去活跃度, 最终几乎以“零速度”步进, 算法的全局寻优性能降低。

1.2 改进粒子群优化算法

为弥补传统粒子群优化算法的缺陷, 首先引入模拟退火中的“概率突跳特性”, 在寻优过程中引入的随机因素, 使算法具有跳出局部极值的能力。

由式(3)可知, 粒子在迭代的最初阶段具有较大的速度, 在全局范围内进行大面积搜索以确定最优解的可能范围, 在迭代后期粒子将在小邻域范围开展深度搜索。若每次迭代新解优于当前解, 则新解完全被接受, 反之, 以一定概率接受新解, 以 Metropolis 准则定义解的接收程度, 由式(5)表示。

$$P_{ij}^T = \begin{cases} 1 & FX_{id} \leq FX_{gd} \\ e^{-\frac{FX_{id} - FX_{gd}}{KT}} > \text{rand} & FX_{id} > FX_{gd} \end{cases} \quad (5)$$

但是在迭代后期, 新解的范围仍然集中在相邻区域, 降低了算法对新解的利用率。

为解决此问题, 引入了淘汰机制, 若种群搜索新解的范围一直局限于某个区域内, 则将此解淘汰, 选取其他粒子补充到种群中, 通过扩大新解的搜索范围保持种群的多样性。同时随着解之间差异的缩小接受坏解的概率也随之减小, 以种群中新解与当前解之间的趋近程度作为评判标准, 评判标

准取经验值 1.00×10^{-9} 。若种群新解与当前解整体的趋近程度小于预设值时, 则淘汰新解并在种群中补充新粒子。当新解与当前解之间的数值差异较大时, 保留新解的概率变大, 扩大了搜索范围, 进而提升了算法的全局搜索性能。其过程如图 1 所示。

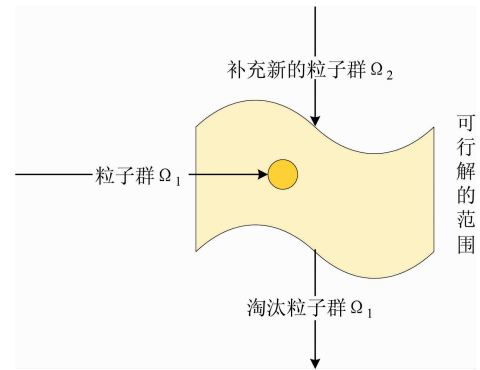


图 1 淘汰机制的过程

Fig. 1 The process of elimination mechanism

当将粒子群算法应用于高光谱波段选择时, 每个粒子代表一种波段的组合数, 粒子的位置和速度则分别代表了波段的组合在种群中的可行解以及向下一个更优的可行解时步进的速度。

高光谱波段选择及验证主要由三部分组成: (1)高光谱图像数据处理模块; (2)波段选择优化模块; (3)地物分类模块。具体流程为: 首先, 在高光谱三维图像的全波段数据中随机选出 n 个波段, 将三维图像数据转换为二维数据, 在 n 个波段中按一定比例随机选取各类地物数据作为波段选择的仿真数据。其次, 经由优化模块进行循环迭代, 得到高光谱最佳组合波段用于表征地物特征。最后, 将筛选出来的波段组合传输到地物分类模块中, 利用已知的数据真实标签值来计算其分类精度, 判定本波段选择方法的有效性。改进粒子群的高光谱波段选取的算法流程如下:

输入: 高光谱图像 X

输出: 筛选出的波段组合数

开始:

(1)设置初始参数: 函数上界 UB, 函数下界 LB, 种群大小 PopSize, 微粒的维数 Dim, 学习因子值 C_1, C_2, R_1, R_2 , 惯性权重初始值 w_{start} , 惯性权重终值 w_{end} , 粒子的最大速度 v_{\max} , 最大迭代次数 MaxIter;

(2)产生初始种群;

(3)采用 JM 距离判断初始最佳波段子集;

循环:

(4)重新计算目标函数值;

(5)保留候选波段子集;

(6)更新粒子的速度和位置;

(7)以一定的概率接受坏解;

(8)如果新解与当前解之间的趋近程度小于预设值, 淘汰新解并在种群中补充新粒子;

(9)得到当前最佳波段子集;

(10)若不满足预设收敛条件返回步骤(4),反之,停止迭代;

(11)输出高光谱图像的最佳波段子集。

结束

2 实验部分

2.1 目标函数的建立

为了提高分类的准确度,构建目标函数应考虑不同地物之间的类间距离,根据两地物之间的均值以及协方差矩阵选择以 Jeffries-Matusita^[13](JM)距离为目标的优化函数,当两类地物的 JM 距离值较大时,则两类地物为同种地物的可能性越小,当两类地物的 JM 距离越小时,则两类地物为同种地物的可能性越大,即两类地物间的特征向量区别越小。

$$D_{ij} = \frac{1}{8}(\text{mean}_i - \text{mean}_j)^T \left(\frac{C_i + C_j}{2} \right)^{-1} (\text{mean}_i - \text{mean}_j) + \frac{1}{2} \ln \frac{\left| \frac{1}{2}(C_i + C_j) \right|}{\sqrt{|C_i|} \sqrt{|C_j|}} \quad (6)$$

$$J_{ij} = [2 \times (1 - e^{-D_{ij}})]^{\frac{1}{2}} \quad (7)$$

$$JM = \sum_{i=1}^{s-1} \sum_{j=i+1}^s J_{ij} \quad (8)$$

式中, mean_i , mean_j 和 C_i , C_j 分别代表高光谱数据中的第 i 类和第 j 类的数据均值和协方差矩阵, s 为高光谱数据中的地物种类。

将目标函数设定为

$$\text{Fitness} = \frac{1}{JM} \quad (9)$$

2.2 数据描述

(1) 第一组高光谱数据场景是由美国加利福尼亚州圣地亚哥的机载可见/红外成像光谱仪(AVIRIS)在印第安纳州西北部的印度松测试点上收集到的数据,场景由 145×145 像素和 224 个光谱通道组成,涵盖 200~2 400 nm 的光谱范围。去除吸水区域波段后,选取 200 个波段用于实验,由印度松(Indian Pines)的第 10, 20, 30 波段合成的高光谱彩色影像如图 2(a)所示。

(2) 第二组高光谱数据是由美国加利福尼亚州圣地亚哥的机载可见/红外成像光谱仪(AVIRIS)在加利福尼亚州萨利纳斯山谷测试点上收集到的数据,由 512×217 个像素和 224 个光谱通道组成。去除吸水区域波段后,选取 204 个波段用于实验,由萨利纳斯(Salinas)的第 10, 20, 30 波段合成的高光谱彩色影像如图 2(b)所示。

2.3 仿真实验

采用 Indian Pines 和 Salinas 两组标准数据开展了改进粒子群波段选择方法的有效性验证,所有实验平台均以 matlab2018a 软件在 Windows10 平台上以 Intel(R) Core(TM) 1.8GHZ 处理器 8GRAM64 位的电脑上运行。将改进粒子群算法(MPSO)与传统粒子群算法(SPSO)、模拟退火-粒子群算法(SAPSO)和遗传-粒子群算法^[14](IEPSO)进行了比较。

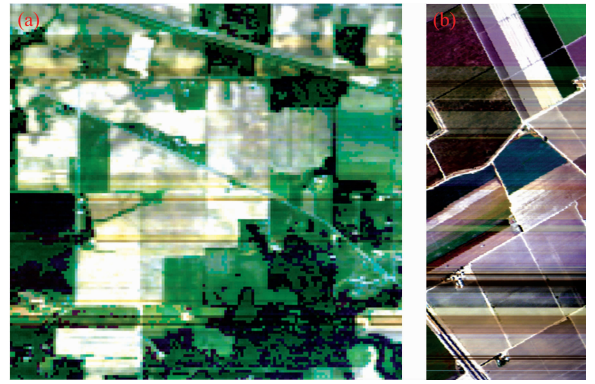


图 2 高光谱彩色影像

(a): Indian Pines; (b): Salinas

Fig. 2 Hyperspectral color images

(a): Indian Pines; (b): Salinas

2.3.1 基于改进粒子群波段选择方法的稳定性仿真测试

为满足目前载荷下传波段大多数为 3 波段的需求,利用两组数据进行实验,以验证改进粒子群算法的适应性和稳定性。数据从 Indian Pines 场景中选择 Grass-pasture-mowed, Oats, Soybean-mintill, Wheat, Buildings-Grass-Trees-Drives 五种地物、Salinas 场景中选择 Broccoli_green_weeds_2, Fallow_rough_plow, Stubble, Grapes_untrained, Corn_senesced_green_weeds 五种地物。

为公平起见,每种算法的仿真数据和测试数据均采用相

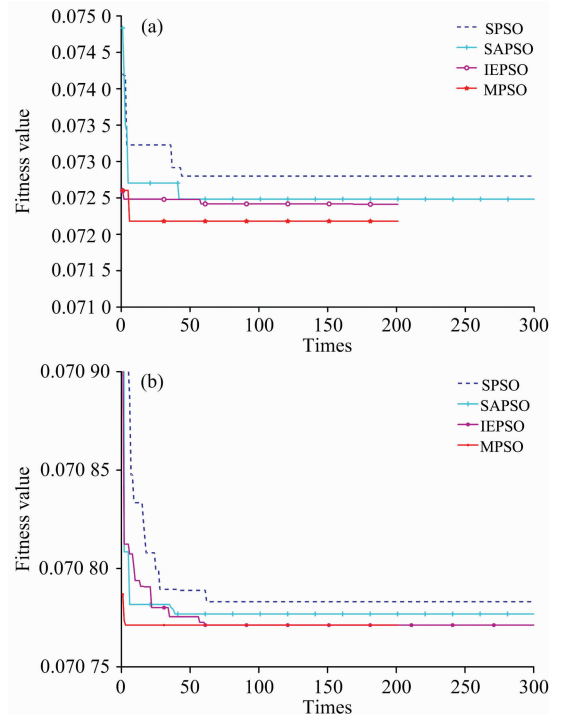


图 3 四种算法目标函数的收敛图

(a): Indian Pines 数据集; (b): Salinas 数据集

Fig. 3 Convergence graphs of objective functions of four algorithms

(a): Indian pines dataset; (b): Salinas dataset

同样本, 分别随机选取它们样本数量的 20% 作为波段选择仿真数据(训练数据), 其余 80% 的样本数量作为后续分类的测试数据。全部运行十次, 四种算法的目标函数值迭代收敛图如图 3 所示。

由图 3 可知, MPSO 算法的迭代次数在 50 次内即可停止收敛, 继承了 SPSO 收敛速度快的优点的, 保留了 SAPSO 算法的深度搜索能力的基础上保证了种群的多样性, 增强了全局寻优能力。MPSO 算法在仿真实验中的收敛精度和收敛速度均高于其他三种算法, 筛选的高光谱的波段具有更高的分类精度。

同时为了验证本算法的稳定性, 分别计算了三种算法运行十次的均值与方差, 如表 1 所示。

表 1 四种算法的均值和方差值

Table 1 Mean and variance of four algorithms

	算法	均值	方差
Indian Pines	SPSO	13.705 1	$4.237 8 \times 10^{-3}$
	SAPSO	13.804 7	$4.595 9 \times 10^{-4}$
	IEPSO	13.814 5	$1.950 3 \times 10^{-4}$
	MPSO	13.840 9	$1.477 8 \times 10^{-4}$
Salinas	SPSO	14.127 3	$2.663 1 \times 10^{-6}$
	SAPSO	14.129 1	$4.573 6 \times 10^{-6}$
	IEPSO	14.129 9	$2.804 8 \times 10^{-18}$
	MPSO	14.130 0	0

由表 1 中的高光谱数据的仿真实验结果可知, MPSO 算法均值在两组数据中的分别可以达到 13.840 9 和 14.130 0, 表明其寻优能力优于其他三种算法; 方差分别为 $1.477 8 \times 10^{-4}$ 和 0, 在四种算法中的方差值最小, 且在第二组数据中其方差值可以达到 0 值, 这就说明了 MPSO 的稳定性很高, 对比其他三种方法具有一定的优势。

2.3.2 基于改进粒子群最佳波段选择分类精度仿真测试

为了验证基于改进粒子群算法筛选出的波段子集的地物分类精度和鲁棒性, 采用相同的训练数据和测试数据, 采用支持向量机^[15](SVM)将实验选出来的波段子集用于地物分类, 同时 SVM 的径向基函数作为支持向量机的核函数, 采用 10 倍交叉验证法选择支持向量机的参数。选取波段选择

中所用到的 20% 仿真数据作为训练样本, 剩下的 80% 数据作为测试样本。

四种方法筛选出的最佳波段子集以及在支持向量机分类器中的性能指标结果分别如表 2 和表 3 所示。

表 2 四种算法选择的波段

Table 2 Selected bands of four algorithms

#	算法	波段组合
Indian Pines	SPSO	114 122 163
	SAPSO	116 120 163
	IEPSO	9 121 163
	MPSO	9 122 163
Salinas	SPSO	79 121 132
	SAPSO	98 139 162
	IEPSO	98 123 132
	MPSO	96 139 162

表 3 数据分类的性能指标

Table 3 Performance index of data classification

#	算法	OIF 指数	Kappa 系数/%	分类精度/%
Indian Pines	SPSO	227.166 7	93.53	97.737 4
	SAPSO	238.302 7	92.35	97.333 3
	IEPSO	281.815 2	93.42	97.697 0
	MPSO	283.404 4	94.70	98.141 4
Salinas	SPSO	800.444 4	98.57	99.000 1
	SAPSO	383.771 1	98.64	99.053 0
	IEPSO	363.817 3	98.63	99.042 4
	MPSO	376.273 8	98.69	99.084 8

通过对两组高光谱数据仿真实验进行分析, 根据不同的波段选择方法, 分别获取了不同的波段组合数的高光谱三波段数据图像, 并对波段选择后的图像进行了 SVM 的分类, 并依次记录了本算法与三个比较算法的 OIF 指数与 Kappa 系数。从分类精度来看, 本文提出的 MPSO 算法均优于其他三种算法。

Indian Pines 和 Salinas 两组高光谱的数据分类图分别如图 4 和图 5 所示。

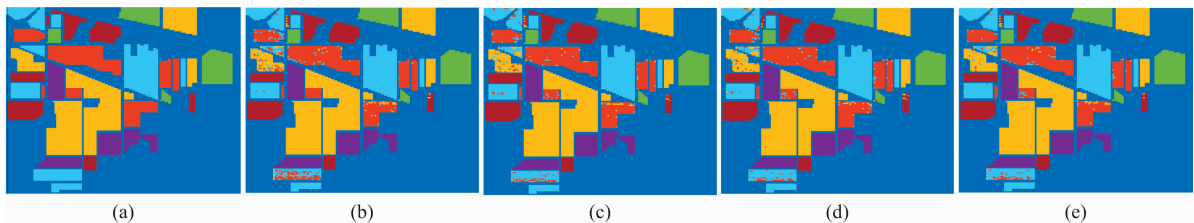


图 4 Indian Pines 分类图

(a): 真值图; (b): SPSO; (c): SAPSO; (d): IEPSO; (e): MPSO

Fig. 4 Classification of Indian Pines

(a): Truth Ground; (b): SPSO; (c): SAPSO; (d): IEPSO; (e): MPSO

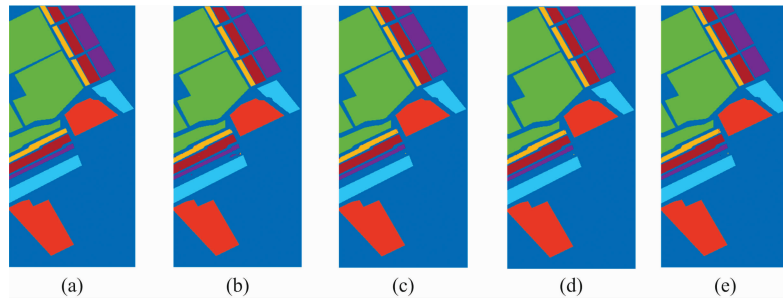


图 5 Salinas 分类图

(a): 真值图; (b): SPSO; (c): SAPSO; (d): IEPSO; (e): MPSO

Fig. 5 Classification of Salinas

(a): Truth Ground; (b): SPSO; (c): SAPSO; (d): IEPSO; (e): MPSO

3 结 论

由于通信带宽和处理速度的限制,往往无法实时传输处理全部数据。高光谱数据的各波段之间具有很强的相关性,通过波段选择算法对原始数据进行降维,在提高了时效性的同时也保证分类精度。从信息处理时效性的角度来看,通过粒子群优化算法来降低数据维数有很大的现实意义。首先利用淘汰机制结合模拟退火-粒子群算法的突跳性优点增强了

对目标函数最优值的探索能力,同时大大的增加了算法的稳定性能,通过对 JM 距离的评判,随后对筛选出的波段组子集进行分类;通过与其他物种算法的对比,显示出本算法的有效性。实验结果表明基于改进粒子群算法的高光谱波段选择方法不仅能满足光谱波段下传能力,且对特定目标实现了高精度分类要求,对于高光谱图像的目标识别和地物分类具有一定的借鉴意义。后期将进一步研究目标函数以及分类器的优化设计,提高数据的分类精度。

References

- [1] Zhang Mingyang, Gong Maoguo, Chan Yongqiang. *Applied Soft Computing*, 2018, 70: 604.
- [2] Long Y, Rivard B, Rogge D, et al. *International Journal of Applied Earth Observation and Geoinformation*, 2019, 79: 35.
- [3] Yang J, Honavar V. *IEEE Intelligent Systems & Their Applications*, 1998, 13(2): 44.
- [4] Wang M, Wu C, Wang L, et al. *Knowledge Based Systems*, 2019, 168(15): 39.
- [5] Ghosh A, Datta A, Ghosh S, et al. *Applied Soft Computing*, 2013, 13(4): 1969.
- [6] Yang X, Deb S. Cuckoo Search *via* Levy Flights. *Proceedings of the World Congresses on Nature and Biologically Inspired Computing*, 2009: 210.
- [7] WANG Li-guo, WEI Fang-jie(王立国, 魏芳洁). *Journal of Image and Graphics(中国图像图形学报)*, 2013, 18(2): 235.
- [8] Ding S, Yuan X, Chen L. *Acta Geodaeticaet Cartographica Sinica*, 2010, 39(3): 257.
- [9] Qi C, Zhou Z, Sun Y, et al. *Neurocomputing*, 2016, 220(12): 181.
- [10] Clerc M, Kennedy J. *IEEE Transactions on Evolutionary Computation*, 2002, 6(1): 58.
- [11] GAO Ying, XIE Sheng-li(高 鹰, 谢胜利). *Computer Engineering and Applications(计算机工程与应用)*, 2004, 40: 47.
- [12] DU Song, ZHOU Jian-yong(杜 松, 周健勇). *Computer Simulation(计算机仿真)*, 2015, 32(12): 218.
- [13] Medjahed S A, AitSaadi T, Benyettou A, et al. *Egyptian Journal of Remote Sensing & Spacecece*, 2016, 19(2): 163.
- [14] Lv X, Wang Y, Deng J, et al. Improved Particle Swarm Optimization Algorithm Based on Last-Eliminated Principle and Enhanced Information Sharing. *Computational Intelligence and Neuroscience*, 2018, doi: 10.1155/2018/5025672.
- [15] Bonah E, Huang X, Yi R, et al. *Infrared Physics & Technology*, 2020, 105: 103220.

Hyperspectral Band Selection Based on Improved Particle Swarm Optimization Algorithm

ZHANG Liu¹, YE Nan¹, MA Ling-ling², WANG Qi², LÜ Xue-ying¹, ZHANG Jia-bao^{1*}

1. College of Instrumentation & Electrical Engineering, Jilin University, Changchun 130061, China

2. Key Laboratory of Quantitative Remote Sensing Information Technology, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China

Abstract Hyperspectral images have hundreds of continuous and narrow spectral bands, spanning visible light to infrared light. They can provide fine spectral properties of ground objects and have important application value for recognizing and classifying ground objects' materials and attributes. It is of great significance to select limited spectral bands for transmission and processing of interested targets, improving the timeliness of hyperspectral data processing and designing practical spectrometers for specific applications. Selecting the optimal band combined with the target features becomes an inevitable requirement to improve the processing efficiency and ensure the accuracy of target recognition or classification. Therefore, selecting the band subset with better classification and recognition ability from hundreds of hyperspectral images is an urgent problem to be solved. This paper proposes a hyperspectral band selection method based on the improved particle swarm optimization algorithm. This method is different from the traditional particle swarm optimization algorithm by introducing the "probability jump characteristic" and setting the elimination mechanism of the new solution to eliminate the "stagnation" new solution, which improves the global optimization performance of the algorithm. Then, based on the spectral characteristics of the target, the objective optimization function of optimal band selection is adopted. The improved particle swarm optimization algorithm is used to solve the objective function, and the selected band subset is fed back to the support vector machine (SVM) for classification application. In this paper, two standard hyperspectral datasets (Indian pines, The experimental results show that the proposed method has higher classification accuracy than the existing methods. Among the several methods, the traditional particle swarm optimization algorithm has the worst effect; the waveband selected by the proposed algorithm has the best classification accuracy, and the classification accuracy of the two data sets can reach 98.141 4% and 99.084 8%, respectively.

Keywords Hyperspectral image; Band selection; Particle swarm optimization algorithm; Support vector machine

(Received Oct. 6, 2020; accepted Feb. 23, 2021)

* Corresponding author