引文格式：王茂山, 何敏, 侣文娜, 等. 基于组合优选特征的极化SAR滨海湿地[J]. 地理空间信息, 2023, 21(6): 63-67. doi:10. 3969/j.issn. 1672-4623.2023.06.016

基于组合优选特征的极化SAR滨海湿地

王茂山[1]，何　敏[1]，侣文娜[1]，蔡　炜[1]
（1. 河海大学 地球科学与工程学院，江苏 南京 211100）

摘　要：滨海湿地的复杂性为准确获取湿地地物类别带来了挑战，不加区分地利用从湿地数据提取的特征参数会导致计算效率低、信息冗余的问题，以合适的特征选择方法选取特征子集是提升湿地分类精度的关键。针对此问题提出了基于组合优选特征的分类方法，将顺序后向选择法和粒子群优化法分别与可分性指数结合筛选优化特征子集，并利用支持向量机对湿地ALOS-2全极化SAR影像进行分类实验，比较2种搜索策略驱动下特征选择方法获取的优选特征的湿地分类精度。结果表明：组合优选特征方法适用于极化SAR数据滨海湿地分类，基于优选特征子集的湿地分类除提升运行速度外，提高了3%~6%的总体分类精度；相比顺序后向选择法，以粒子群优化算法为搜索算法筛选出的特征子集在湿地分类上具有更好的表现，其分类总体精度达到95.20%，Kappa系数达到0.944 8。

关键词：极化SAR；特征提取；湿地分类
中图分类号：P208　　**文献标志码：**B　　**文章编号：**1672-4623（2023）06-0063-05

Polarimetric SAR Coastal Wetland Feature Classification Based on Combined Optimization Features

WANG Maoshan[1], HE Min[1], SI Wenna[1], CAI Wei[1]
(1. School of Earth Sciences and Engineering, Hohai University, Nanjing 211100, China)

Abstract: The complexity of coastal wetlands brings challenges to the accurate acquisition of wetland feature types. Indiscriminate use of feature parameters extracted from wetland data will lead to problems of low computational efficiency and redundant information. Selecting feature subsets with appropriate feature selection methods is the key to improving wetland classification accuracy. In order to solve this problem, we proposed a classification method based on combined optimization feature, which combined the sequential backward selection method and the particle swarm optimization method with the separability index to select optimization feature subsets respectively, used the support vector machine to classify the wetland ALOS-2 fully polarized SAR image, and compared the wetland classification accuracy of preferred features obtained by the feature selection method driven by the two search strategies. The results show that the combined optimization feature method is suitable for the classification of coastal wetlands based on polarimetric SAR data, and the wetland classification based on optimization feature subsets improves the overall classification accuracy by 3%~6% in addition to improving the running speed. Compared with the sequential backward selection method, the feature subset selected by the particle swarm optimization algorithm has better performance in wetland classification, its overall classification accuracy reaches 95.20%, and the Kappa coefficient reaches 0.944 8.

Key words: polarimetric SAR, feature extraction, wetland classification

顺序后向选择法（SBS）是一种常用的序列搜索算法，在总特征集合下设定评价函数进行迭代，在迭代过程中依次对特征进行剔除直到评价函数符合要求。顺序后向选择法时间复杂度低，其缺陷在于被剔除的特征不能重新加入，所以容易陷入局部最优。随机搜索迭代时每次选择的特征子集具有随机性，具有防止陷入局部最优的优点，在特征选择领域被广泛采用。随机搜索算法有受人类遗传现象启发而设计的[1-2]，也有受自然界动物某种行为启发构建的[3]，有学者对二者进行了融合以提升算法的性能[4]。粒子群优化算法（PSO）是一种经典的随机搜索算法，相较其他随机搜索算法其参数少，更易编程实现，但相较序列搜素算法，其计算速率略有不足。选择合适的特征搜索算法能提升特征子集的质量，但目前围绕搜索算法评估的研究仍然较少，分析不同搜索策略的表现是有待解决的问题。

针对以上情况，为降低极化信息冗余，分析搜索算法对特征选择的影响，有效解决滨海湿地的特征选择问题，提升极化SAR数据应用到湿地分类中的分类能力，本文以美国圣巴勃罗湾国家自然保护区的

收稿日期：2022-01-24；修回日期：2022-04-01。
第一作者简介：王茂山（1997—），硕士研究生，研究方向为极化SAR图像分类、极化SAR地物分类，E-mail：ricksanchez@163.com。

ALOS-2全极化数据为实验数据，引入了可分性指数进行特征初选，同时采用顺序后向选择法以及粒子群优化算法2种搜索算法进行特征选择，利用SVM分类器对筛选出的特征子集进行湿地地物分类及精度评价，探究适用于滨海湿地分类的特征选择方法，为后续基于极化SAR的湿地研究提供参考。

1 研究方法

1.1 特征选择方法

1.1.1 可分性指数（SI）

根据特征选择与学习器的结合方式，特征选择方法主要分为过滤式（Filter）、封装式（Wrapper），本文选用的是封装式方法。封装式方法将特征选择过程与学习算法结合，利用学习器在特征子集上的预测精度评价所选特征的优良，针对特定的学习器具有更好效果，但由于每构建出一组特征子集就要构造学习器进行评价，导致封装式方法运行时间通常较长，进行特征初选能改善这一缺陷。由于在不同特征下地物类型之间的区分度具有差异性，引入可分性指数（SI）表示一种特征下的类间差异性，通过可分性指数对特征集合进行特征初选，提升特征选择运行效率。SI计算公式如下：

$$I_{p,q}^{(i)}=\frac{\left|\mu_p^i-\mu_q^i\right|}{\sigma_p^i+\sigma_q^i} \tag{1}$$

式中，i为特征；p、q为类别；μ_p^i、μ_q^i和σ_p^i、σ_q^i为p类、q类i特征的均值和标准差。可分性指数值$I_{p,q}^{(i)}$越小，表明在i特征内，p类和q类的类间差异性小，即该特征对分类效果提升不大，可以从特征集合中去除。通过对可分性指数这一指标设置适当的阈值，可以筛选出一批地物区分性强的特征，组成初选特征集合作为后续特征选择方法的输入。

1.1.2 顺序后向选择法（SBS）

顺序后向选择法是一种自顶而下的序列搜索方法，其思想是从所有特征开始逐一剔除不被选中的特征，每次剔除的特征都使得剩余的特征在预设的评价准则下达到最优。顺序后向选择法的具体步骤如下：

1）输入特征集合$F=\{f_1,f_2,f_3,\cdots f_m\}$，$m$为特征数量；

2）对$i=1,2,3,\cdots m$，去除特征f_i，计算剩余特征所组成的特征集合$F_i^{'}=\{f_1,f_2,f_3,\cdots f_{i-1},f_{i+1},\cdots,f_m\}$的总体分类精度$O(F_i^{'})$；

3）对于精度集合$O=\{O(F_1^{'}),O(F_2^{'}),O(F_3^{'}),\cdots,O(F_m^{'})\}$，将最大整体分类精度$O(F_k^{'})$所对应的特征$f_k$从特征集合$F$中剔除；

4）重复1）~3），直到剩余特征数量$m^{'}$等于预设值。

1.1.3 粒子群优化算法（PSO）

粒子群优化算法[6]是一种随机搜索算法，通过模拟鸟群的觅食行为初始化一群随机粒子，粒子具有自身的位置和速度决定飞行的方向和距离，同时还具有一个被优化函数决定的适应值。粒子们追随当前的最优粒子在解空间中进行搜索，每次迭代粒子都通过跟踪个体极值和最优极值来更新自己的速度和位置，从而得到全局最优解。在D维目标搜索空间中，由种群数为n的粒子组成粒子群，其中，第i个粒子在第d维的位置为x_{id}，其飞行速度为v_{id}，该粒子当前搜索到的最优位置为p_{id}，整个粒子群当前的最优位置为p_{gd}。速度与位置更新公式如下：

$$v_{id+1}=w\cdot v_{id}+c_1\times r\times(p_{id}-x_{id})+c_2\times r\times(p_{gd}-x_{id}) \tag{2}$$

$$x_{id+1}=x_{id}+v_{id+1} \tag{3}$$

式中，w为惯性权重；r为0~1之间的随机数；c_1和c_2为加速系数。

针对属于离散优化问题的特征选择，本文采用二进制粒子群算法（BPSO）[7]，该算法对每个粒子进行二进制编码，粒子速度代表粒子的每个位置赋值为1的概率，因此需要通过转换函数将速度转换到[0，1]区间。使用sigmoid函数进行转换的粒子更新公式如下：

$$x_{id}=\begin{cases}1, r\leqslant s(v_{id})\\0, \text{otherwise}\end{cases} \tag{4}$$

式中，$s(v_{id})$为位置x_{id}，取值为1的概率。

1.2 湿地分类过程

1.2.1 特征提取和初选

依次进行数据格式转换，辐射定标，提取相干矩阵，多视处理，地理编码，滤波。以4∶6为视数比进行多视处理，采用3×3窗口的Refined Lee滤波进行图像去噪处理。利用19种极化分解算法，对相干矩阵共提取了86个极化特征参数组成特征集合。根据研究区域的地物类型并参照同时期的光学影像选取合适数量的地物样本并划分训练集和测试集，对训练集数据计算可分性指数后以合适的阈值参数筛选特征组成特征初选子集。

1.2.2　特征选择

以特征初选子集为输入，分别以PSO和SBS为搜索算法进行特征选择，将SVM[8]分类精度作为评价准则对特征子集进行性能比较。将训练集数据划分为特征训练集和特征测试集，以特征训练集数据根据搜索算法筛选出备选特征子集。计算特征子集在特征测试集上的总体分类精度并以此为依据评价子集质量。2种特征选择算法都运行10次并取获得最高总体分类精度的备选特征子集作为优选特征子集。

1.2.3　SVM分类

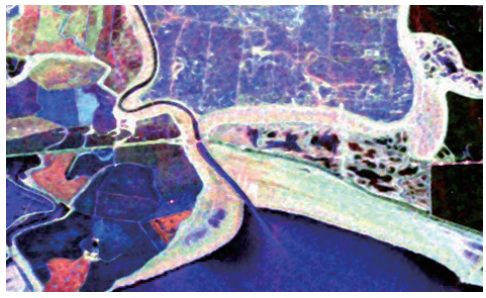
将总特征集合、基于可分性指数计算出的特征初选集合、基于PSO和基于SBS的特征选择计算出的优选特征子集分别进行SVM分类并计算测试集的总体分类精度。以全部特征集合、特征初选集合、2种搜索策略下的优选特征子集为特征向量，以训练集样本建立SVM分类模型其中采用径向基函数（RBF）作为SVM核函数。用训练好的SVM模型对SAR图像进行分类，以测试集的总体分类精度为标准评估特征子集质量，对湿地分类的效果和精度进行定量分析。

2　研究结果与分析

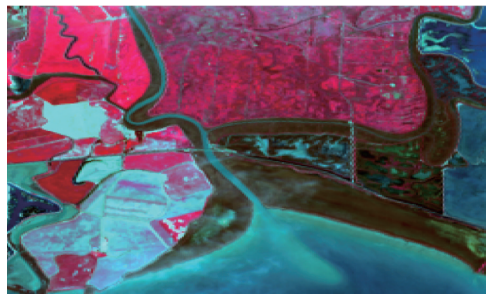
2.1　研究区域与数据源

选择美国圣巴勃罗湾国家自然保护区为研究区域，面积约53.4 km^2，保护区内涵盖了开阔水域、泥滩、潮汐沼泽、河口以及季节性的湿地等多种地形。

本文选用的实验数据为ALOS-2采集的PALSAR全极化SAR数据，研究区域影像的成像时间为2015年3月24日，影像尺寸为631像素×821像素，影像分辨率为10 m。图1a为实验区域的Pauli-RGB彩色合成图。依据保护区的特点，将研究区域内的地物分为库塘、浅海水域、河流湿地、潮间盐水湿地、潮间淤泥湿地、森林沼泽、藓类沼泽和草本沼泽8类。此外以同时期Landsat8卫星影像作为参考，图1b为Landsat8近红外、红、绿波段合成的标准假彩色影像，地物色彩鲜明，有利于进行植被信息观察。



a　Pauli-RGB彩色图像



b　Landsat8假彩色图像

图1　圣巴勃罗湾国家自然保护区影像

2.2　特征选择结果分析

利用公式（1）对所选的8类地物进行可分性指数计算，通过分析发现，anisotropy12、derd、entropy、lnb、p1、p2、SE这8个特征在某一类地物的可分性指数接近甚至超过2的同时，另有一类地物的可分性指数也具有接近1.5的水平，说明这8个特征对湿地地物具有较强的区分能力。库塘在anisotropy12、asymetry、AY3_Odd、Lnb、P2、WAH2_T11、Y3_Odd等6个特征下可分性指数达到或超过1.5，所以用这6种特征参数提取库塘信息能获取较好的效果；在anisotropy12、ArN_Odd、entropy、N_tau特征下，浅海海域与其他地物具有较为明显的区分，因此这些特征是浅海海域的代表特征；ArN_Vol、AY3_Vol、AY4_Vol、F_Vol、S4_Vol、VZ3_Vol、Y3_Vol、Y4_Vol等极化特征对河流湿地的可分性指数达到2，可以用于区分河流湿地与其他地物；ArN_Dbl、AY3_Dbl、AY4_Dbl、F_Dbl、S4_Dbl、S6_Dbl、Y3_Dbl、Y4_Dbl等特征对潮间盐水湿地的敏感性远高于对其他地物的敏感性，因此可以用这些特征将潮间盐水湿地提取出来；针对潮间淤泥沼泽，选用alpha、N_mod特征区分度较好；选择ArN_Odd、entropy、F_Odd、P3、PF、PH、RVI、S4_Odd等特征可以用于区分森林沼泽；相比其他地物类别，草本沼泽和藓类沼泽没有区分度较高的特征，二者的可分性指数均处于1之下，选用anisotropy12、F2_Grd、F_Odd、HAA_T22等特征可以对二者进行区分。

此外还可以观察到，特征与地物散射机制具有关联性，因此二者在多个Odd特征下可分性指数较大，这说明与其他地物类别相比，二者的面散射分量更小。

对于8类地物在86种特征下计算出的可分性指数，设定0.9为SI数值阈值，设定3为SI数量阈值，即对于在一种特征下，如果少于3类地物的可分性指

数大于0.9，就将这一特征去除。经过可分性指数筛选后保留了34个特征，将其组成初选特征集合，分别进行基于PSO和基于SBS的特征选择过程。为保证2种搜索策略对比的合理性，以基于PSO的特征选择方法计算出的优选特征数量为标准，对基于SBS的特征选择的特征保留数量进行设置。进行特征选择后保留的特征子集。

2.3 分类精度分析

为检验特征选择方法的有效性，比较SBS算法和PSO算法对特征选择结果的影响，以支持向量机作为分类器，以不同特征子集为支持向量进行极化SAR分类和精度评价。将全部86维特征组成的特征集合记为特征集合ALL；将基于可分性指数筛选出的34维特征集合，记为特征集合SI；将基于SBS的特征选择方法直接筛选ALL保留的19维特征集合记为特征集合ALL+SBS；将基于PSO的特征选择方法直接筛选ALL保留的19维特征集合记为特征集合ALL+PSO；将通过可分性指数结合基于SBS的特征选择方法筛选保留的13维特征集合记为特征集合SI+SBS；将基于可分性指数结合基于PSO的特征选择方法筛选保留的13维特征集合记为特征集合SI+PSO。

从图2看出，基于特征集合ALL的湿地分类结果存在潮间淤泥湿地、草本沼泽与浅海海域混分的情况，同时椭圆区域内河流湿地的地块破碎，并存在将潮间淤泥湿地错分为浅海海域的情况，与实际情况不相符合。从图3可以观察到，利用特征集合SI的分类效果并没有改善浅海海域错分为潮间淤泥湿地、草本沼泽的情况，此外还出现了圆形区域内中草本沼泽和藓类沼泽混分的情况，矩形框内部分潮间盐水湿地错分为浅海海域。图4中可以看出，基于特征集合SI+SBS的分类结果有效改善了浅海海域错分为潮间淤泥湿地的情况，椭圆区域内草本沼泽和苔藓沼泽混分情况也较大程度的减轻，森林沼泽的边缘更为完整平滑，分类效果得到一定程度的提升。在图5中可以观察到，基于特征集合SI+PSO的分类结果除了改善了浅海海域内的混分、草本沼泽内苔藓沼泽的错分情况外，多数区域的边缘更加平滑，如矩形实线框内森林沼泽，此外圆形区域内河流湿地被完整区分出来，矩形虚线区域内的潮间盐水湿地的几何轮廓与实际情况更加符合。从图6中可以观察到，利用特征集合ALL+SBS的分类效果一定程度上改善了浅海海域内的混分，但没有改善草本沼泽内苔藓沼泽的错分，此外出现了圆形区域内中河流湿地与潮间淤泥湿地混分的情况。从图7中可以观察到，利用特征集合ALL+SBS的分类效果一定程度上改善了浅海海域内的混分，但没有改善草本沼泽内苔藓沼泽的错分，此外出现了圆形区域内中河流湿地与潮间淤泥湿地混分的情况。

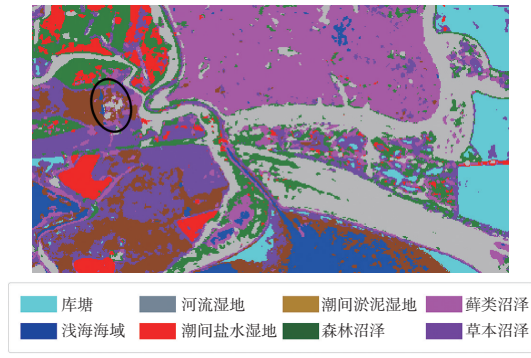


图2 基于特征集合ALL的湿地分类结果

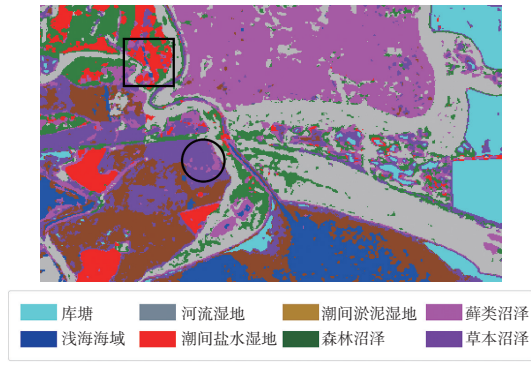


图3 基于特征集合SI的湿地分类结果

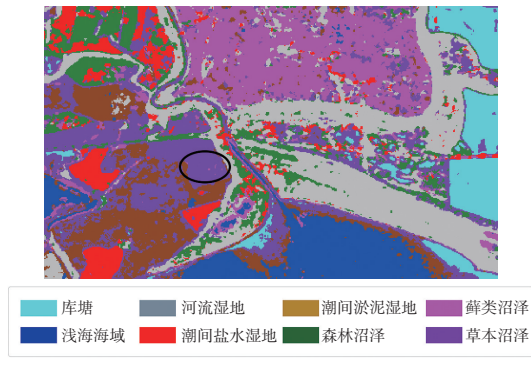


图4 基于特征集合SI+SBS的湿地分类结果

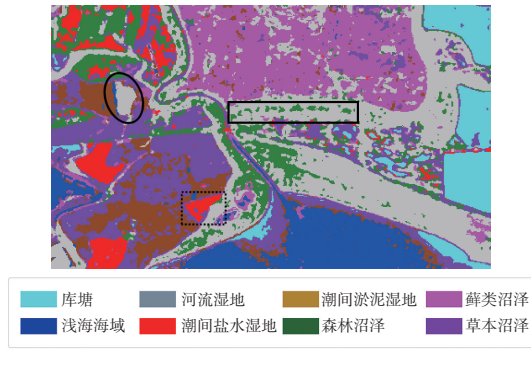


图5 基于特征集合SI+PSO的湿地分类结果

表1是不同特征集合进行分类的精度指标，可以

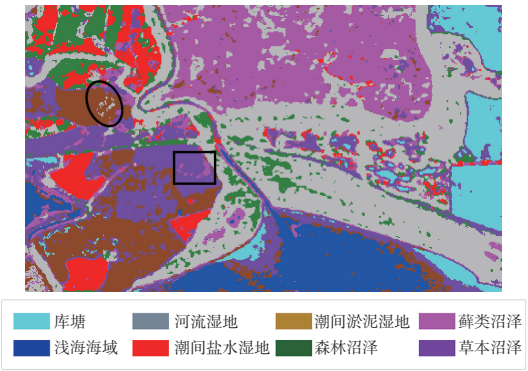


图6　基于特征集合ALL+SBS的湿地分类结果

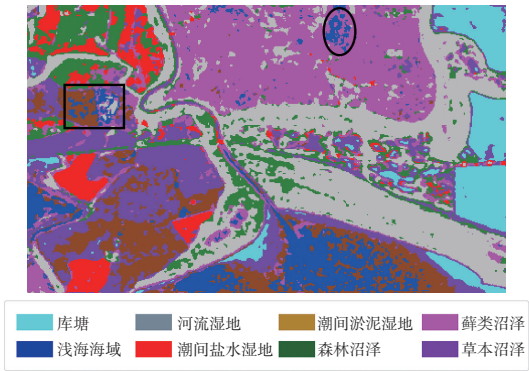


图7　基于特征集合ALL+PSO的湿地分类结果

明显看出经过特征选择获取的优选特征子集不仅提升了计算效率，同时提高了分类精度，适用于湿地分类。基于SBS和基于PSO的特征选择方法都减少了分类所需的计算时间，并在一定程度上改善了分类效果。使用SBS特征选择方法获取的特征子集对分类精度有小幅提升，而基于PSO特征选择方法获取的特征子集总体分类精度达到95.20%，Kappa系数达到0.944 8。

表1　不同特征选择算法分类效果对比表

特征集合	ALL	SI	ALL+SBS	ALL+PSO	SI+ SBS	SI+PSO
特征维度	86	34	19	19	13	13
OA/%	89.07	78.67	88.94	87.92	92.53	95.20
Kappa	0.873 8	0.753 8	0.871 5	0.856 1	0.913 9	0.944 8
运行时间/s	32.228 7	19.254 3	12.650 6	13.320 9	4.397 9	2. 983 4

3　结　论

本文选择美国圣巴勃罗湾国家自然保护区的ALOS-2全极化SAR数据作为湿地分类数据，通过可分性指数结合特征选择方法对不同极化参数进行特征优选，比较顺序后向选择法和粒子群优化算法2种搜索策略的特征选择结果在SVM下的分类精度，得出如下结论：①不同极化方法获得的极化参数对湿地典型地物的敏感程度具有差异；②结合可分性指数初选的特征选择方法能够降低特征维度；③本文选用的基于顺序后向选择法和基于粒子群优化算法的特征选择过程都能获得优选特征子集，优选特征子集最大能提高6%左右的分类精度；相较顺序后向选择法，基于粒子群优化算法为搜索策略的特征选择方法可以获得质量更好的特征子集，更适宜极化SAR滨海湿地分类。

参考文献

[1]　辛宇，童孟军，华宇婷. 一种基于最优特征选择改进的遗传算法[J]. 传感技术学报，2018，31(11):1 747-1 752

[2]　高鹏毅，陈传波，张葵，等. 一种使用多Filter初始化GA种群的混合特征选择模型[J]. 小型微型计算机系统，2012，33(11):2 379-2 384

[3]　王金杰，李炜. 混合互信息和粒子群算法的多目标特征选择方法[J]. 计算机科学与探索，2020，14(1):83-95

[4]　张进，丁胜，李波. 改进的基于粒子群优化的支持向量机特征选择和参数联合优化算法[J]. 计算机应用，2016，36(5):1 330-1 335

[5]　李平，徐新，董浩，等. 利用可分性指数的极化SAR图像特征选择与多层SVM分类[J]. 计算机应用，2018，38(1):132-136

[6]　肖艳，姜琦刚，王斌，等. 基于ReliefF和PSO混合特征选择的面向对象土地利用分类[J]. 农业工程学报，2016，32(4):211-216

[7]　Kennedy J，Eberhar R. A Discrete Binary Version of the Particle Swarm Algorithm[C]// Proceeding of 1997 IEEE International Conference on Systems，Man，and Cybernetics. Computational Cybernetics and Simulation，Orlando，FL，USA:IEEE，1997:4 104-4 109

[8]　陈媛媛，何秀凤，王静. 基于SVM的极化SAR沿海滩涂分类[J]. 地理空间信息，2015，13(6):150-153