

考虑光谱变异性的多光谱植被识别最优特征空间构建

林 怡^{1,2} 厉 朗¹ 宇 洁^{1,2} 高 忱¹ 钟代琪¹ 陈 鑫¹ 杨羽轩¹

(1. 同济大学测绘与地理信息学院, 上海 200092; 2. 同济大学遥感应用研究中心, 上海 200092)

摘要: 在中低分辨率遥感卫星影像上, 植被识别受数据获取条件和不同生长期等因素的影响, 会存在端元光谱变异现象, 导致植被解混误差较大。提出了一种顾及端元光谱变异性的最佳距离遗传算法 (IIDGA), 通过自动特征选择方法减小端元类内差异, 增大类间差异, 构建适用于中等分辨率影像的植被解混最优特征空间, 提高 Landsat 影像的植被识别精度。通过比较传统波段组合、光谱和纹理特征全集与 IIDGA 优选特征的线性解混模型效果, 验证了最优特征选择的重要性。结果显示, 特征选择有助于提升解混精度 (IIDGA 的均方根误差最低, 为 0.180); 同时, 通过比较基于 IID 指数的 Filter 算法、基于标准 GA 的 Wrapper 算法和 IIDGA 在最优特征自动选取方面的性能, 证实了 IIDGA 在平衡精度与效率方面的优势。

关键词: 多光谱遥感; 植被识别; 端元光谱差异; 最佳距离遗传算法; 自动特征选择算法

中图分类号: X87; TP751

文献标识码: A

文章编号: 1000-1298(2024)08-0225-08

OSID:



Optimal Feature Space Construction for Multispectral Vegetation Recognition Considering Endmember Variability

LIN Yi^{1,2} LI Lang¹ YU Jie^{1,2} GAO Chen¹ ZHONG Daiqi¹ CHEN Xin¹ YANG Yuxuan¹

(1. College of Surveying and Geo-informatics, Tongji University, Shanghai 200092, China

2. Research Center of Remote Sensing Technology and Application, Tongji University, Shanghai 200092, China)

Abstract: Due to differences in data acquisition and vegetation growth periods, vegetation recognition on low- and medium-resolution remote sensing imagery widely suffers from endmember variability. The endmember variability directly leads to large vegetation unmixing errors. To increase the vegetation recognition accuracy on the multispectral imagery, an intra-inter distance genetic algorithm (IIDGA) that accounts for the endmember variability was proposed. IIDGA can decrease the intra-distance and increase the inter-distance simultaneously, which enhanced the distinguishability of the endmembers through an automatic feature selection method. An optimal feature space for vegetation unmixing was constructed on the medium resolution imagery to improve the vegetation recognition accuracy based on the Landsat imagery. The importance of optimal feature selection was demonstrated by comparing the linear unmixing model accuracy based on the classical band combination features, the spectral and textural feature set and the proposed IIDGA. The results verified that feature selection was beneficial to improve the unmixing accuracy. The RMSE of IIDGA equalled 0.180 which was the lowest among the three methods. Meanwhile, the IID index-based Filter method, the standard GA-based Wrapper method and the proposed method were compared with their performances in automatic optimal feature selection. The results confirmed the superiority of the IIDGA in trading off accuracy and efficiency.

Key words: multispectral remote sensing; vegetation recognition; spectral endmember variability; intra-inter distance genetic algorithm (IIDGA); automatic feature selection

收稿日期: 2024-04-26 修回日期: 2024-06-06

基金项目: 国家自然科学基金项目(42201376, 41771449)、中央高校基本科研业务费专项资金项目(2022-4-ZD-05, 2023-3-YB-12)和同济大学“中德合作 2.0”培育项目(4300143344/039)

作者简介: 林怡(1970—), 女, 研究员, 博士, 主要从事环境遥感研究, E-mail: linyi@tongji.edu.cn

通信作者: 宇洁(1989—), 女, 讲师, 博士, 主要从事环境遥感研究, E-mail: 2011_jieyu@tongji.edu.cn

0 引言

植被是陆地生态系统的重要组成部分,也是全球气候变化的重要指示器^[1-2]。在植被提取、时空变化特征分析中,中低分辨率遥感影像^[3-4]应用广泛。然而,由于空间分辨率和实际地物分布复杂性的双重限制^[5],此类影像普遍存在混合像元问题。已有研究表明,混合像元制约了地物(如植被、农田)识别和分类精度的提升,通过解混模型获得每个端元(单一地物构成的像元)及其对应的丰度(混合像元中不同端元所占面积百分比),有助于改善地物提取精度^[6]。大气、光照和环境共同作用引起的端元光谱变异性(实际端元光谱与标准端元光谱的差异)是混合像元分解的重要误差来源^[7],在基于遥感的植被识别中,同物异谱(如同种植被在不同时期的光谱曲线)现象增大了端元光谱类内差异,异物同谱(云层阴影与水体光谱)现象减小了端元光谱类间差异,从而加大了特征空间对端元的区分难度。因此,在端元变异性广泛存在的情况下,通过自动特征选择方法,缩小端元类内差异、扩大类间差异,构建最优特征空间,提升解混模型的分类效果对于植被信息的准确提取具有重要意义。

文献[8]将解决端元变异性的方法划分为5类:迭代混合分析类、光谱特征选择类、光谱波段加权类、光谱转换类和光谱传输模型建模类。光谱特征选择类方法因计算量较小、实现方便等原因得以广泛应用,但是该方法主要依赖影像的光谱特性,且无法同时量化端元类内和类间差异。如何结合多类型特征,构建能够同时反映类内和类间差异的定量指标,值得深入研究。在对地物特征的描述过程中,特征过少容易出现欠拟合,致使所选特征难以全面反映研究对象的特点;特征过多则会出现过拟合,限制分类算法的泛化能力^[9]。通过合理的特征选择方法,从原始特征中筛选出相关性高、冗余度小的特征是实现最优特征空间构建的目标。传统特征选择方法主要依赖人工经验,自动化程度低。通过特征优选自动地获取最优特征子集,在提高运算效率的同时,也能更好地保持特征泛化能力^[10]。

常用的特征选择方法^[11]包括:过滤法(Filter算法)、包裹法(Wrapper算法)和嵌入法(Embedded算法)。Filter特征选择算法仅依赖数据集本身,通过对评价函数(如距离、关联性、信息度等)结果排序以获得最优特征子集。由于此类算法评价结果与后续分类算法无关,所以运算速度快、通用性强,但分类效果与特征子集可能有较大偏差。Wrapper算法则以分类准确率作为评价准则,因此获得的特征子

集与分类算法偏差小,但与Filter算法相比,Wrapper算法每次评价都要对分类算法进行训练和测试,计算量大大增加,而且容易产生过拟合现象。Embedded算法将特征选择与模型训练纳入同一过程,通过模型训练后的权重数值表征特征的重要性。尽管Embedded算法较Wrapper算法有效缩短了特征子集的重分类时间,但是对权值相近的特征贡献难以直接量化,因此,其构建的特征空间不一定是最优的。

本文针对端元存在的光谱变异性问题,以端元类内差异最小、类间差异最大化为目标,结合Filter和Wrapper特征选择算法的优点,提出一种最佳距离遗传算法(Intra-inter distance genetic algorithm, IIDGA),自动选择特征,提高光谱-纹理多特征表达的准确性和适应性,构建适用于中等分辨率、多光谱遥感影像的最优特征空间,并以Landsat影像的植被(红树林)识别问题为例,以验证算法的有效性。

1 研究区及数据源

1.1 研究区概况

本文研究区位于广西壮族自治区北海市合浦县沙田半岛。合浦县地处21°0′~21°55′N, 108°51′~109°46′E,森林资源丰富,红树林面积约为2 595.6 hm²。经过实地调查,确定了研究区内的主要地物类型为水体、红树林和裸地。为了更好地保护和修护红树林,广西北海、防城港等地自2000年起开展了大规模的红树林人工种植^[12]。一方面,由于研究区内的红树林处于不同生长阶段,光谱反射特征有所不同,增大了植被类内差异;另一方面,由于红树林生长环境受潮汐影响,周期性地淹没于水体中,减小了植被和水体类间差异,加大了植被识别难度。

1.2 数据源及数据处理

研究选用的GeoEye和Landsat TM多光谱影像如图1所示,其中Landsat影像用于最优特征空间构建的比较实验;Landsat TM影像^[13]由7个波段组成,包括:3个可见光波段(Band1~Band3,分别为蓝

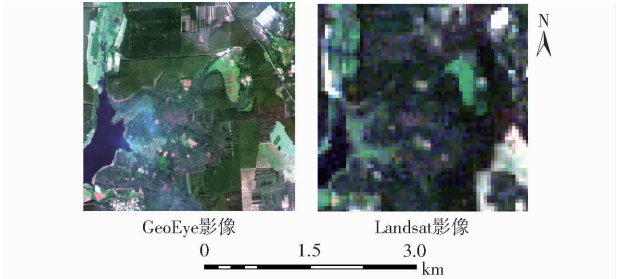


图1 研究区域多光谱遥感影像

Fig. 1 Multispectral imagery of study area

光、绿光 和 红 光 波 段), 波 长 覆 盖 范 围 $0.45 \sim 0.69 \mu\text{m}$; 1 个 近 红 外 波 段 (Band4), 波 长 $0.76 \sim 0.90 \mu\text{m}$; 2 个 短 波 红 外 波 段 (Band5、Band7), 波 长 分 别 为: $1.55 \sim 1.75 \mu\text{m}$ 、 $2.08 \sim 2.35 \mu\text{m}$; 1 个 热 红 外 波 段 (Band6), 波 长 $10.40 \sim 12.50 \mu\text{m}$ 。GeoEye 影 像 因 空 间 分 辨 率 更 高, 通 过 分 类 的 方 法 生 成 参 考 丰 度 图^[14], 用 于 Landsat 影 像 解 混 精 度 评 定。具 体 数 据 如 表 1 所 示。

表 1 实验数据
Tab.1 Description of research data

影像来源	获取时间	空间分辨率/m	用途
Landsat	2009 年 8 月	30 120(热红外)	最优特征空间构建实验
GeoEye	2009 年 10 月	2	精度评定参考影像

本研究中,遥感影像的预处理包括:图像裁剪、大气校正、几何校正。经过预处理后,分别提取影像的光谱和纹理特征,形成初始特征空间。在此基础上,通过特征选择算法优化特征空间,使用典型线性解混方法——全约束最小二乘模型(Fully constrained least squares, FCLS)^[6]对混合像元进行分解,获得每个像元对应不同端元的隶属度(0~1),最终生成丰度图。最后,选取均方根误差(Root mean square error, RMSE)作为精度评价指标,选取运行时间作为效率评价指标对解混结果进行精度评定。

2 研究方法

2.1 顾及端元光谱变异性的最优特征空间构建方法

在最优特征空间的构建中,特征描述和自动特征选择算法是关键。为了减弱光谱变异现象对植被识别精度的不利影响,构建的最优特征空间不仅需要有效区分植被与非植被(类间差异),还要能够识别出处于不同状态的同一类型地物(类内差异)。为此,本研究提出了一种综合类内距离和类间距离的指标用于描述不同特征对植被识别的贡献度,并通过联合 Filter 和 Wrapper 自动特征选择算法,筛选出适用于植被识别的优选特征用于构建最优特征空间。IIDGA 算法总体流程如图 2 所示。

首先,构建 IID 指数,并用于 Filter 算法特征子集的筛选过程。该过程提出了一个综合端元类内和类间差异的距离指数(Intra-inter distance, IID),用于光谱和纹理特征全集的排序,得到 Filter 过程特征子集;其次,设定类内距离最小、类间距离最大的最佳距离准则,将 Filter 阶段计算得到的 IID 指数用于遗传算法(Genetic algorithm, GA)种群初始化,并结合 Wrapper 算法形成联合 IID 结果和 GA 的

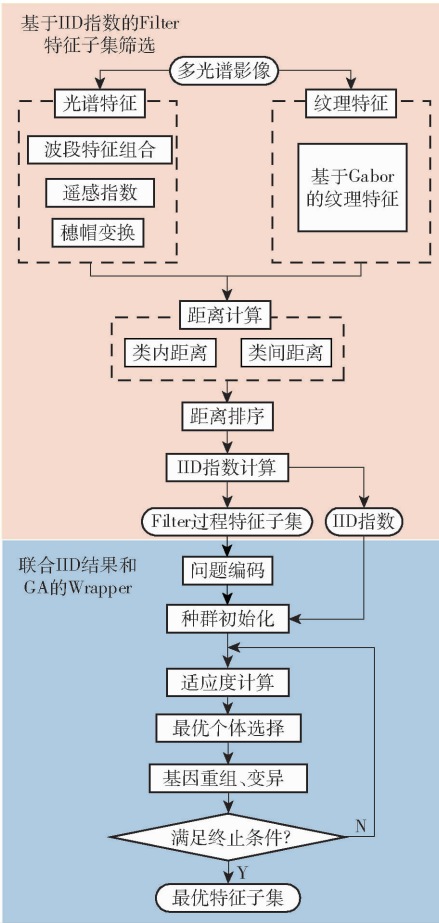


图 2 IIDGA 算法流程图

Fig. 2 Workflow for IIDGA algorithm

Wrapper,用于最优特征空间构建。对于高维数据,Filter 算法不仅可以减小 Wrapper 阶段的搜索特征空间,提高算法运算速度;同时,Filter 算法获得的 IID 结果可表征特征在 Wrapper 阶段被选的概率,为快速、准确地获得 GA 的种群初始值提供依据,更快收敛至最优解。

2.2 光谱-纹理多特征提取

植被除了独特的光谱特性外,空间上由于叶面形状和大小不同,形成鲜明的空间分布特征。因此,在常用光谱特征基础上引入纹理特征,用以描述像素间的空间分布属性,共同构建光谱-纹理特征集。

在基于中低分辨率影像的地物识别研究中,国内外学者大多选取光谱特征构建特征空间^[15]。常用光谱特征包括:遥感影像原始波段或波段组合、遥感指数、波段变换特征等。在波段组合方面,由于端元在不同波段存在光谱数值尺度差异,引入加权变异系数分析法(Weighted coefficient of variation analysis, WCVA)^[14],提取类内差异小、类间差异大的波段组合特征;在遥感指数方面,针对研究区地物分布特性,提取归一化植被指数(Normalized difference vegetation index, NDVI)^[16]和归一化水体指数(Normalized difference water index, NDWI)^[17]。

在波段变换特征方面,提取穗帽变换 (Kauth - thomas transform, K - T)^[18] 的亮度 (Brightness)、绿色度 (Greenness) 和湿度 (Wetness) 分量。

纹理作为物体表面自然属性,能够反映影像同质现象,表征像元空间关系,有效减小同物异谱、同谱异物现象对地物识别的影响,是提高图像分类精度的重要方法之一。Gabor 变换能够同时在空间和频率域的局部达到最优,适用于空间频率 (尺度)、位置及方向选择性的局部结构信息的描述^[19]。因此,本文选择不同尺度和方向的 80 个 Gabor 滤波器用于纹理特征的提取。

2.3 最佳距离遗传算法

IIDGA (图 2) 是一种结合了 Filter 和 GA Wrapper 优势的改进特征选择算法,共包含两部分:基于 IID 指数的 Filter 特征子集筛选;联合 IID 结果和 GA 的 Wrapper。基于 IID 指数的 Filter 特征子集筛选综合考虑了类内和类间距离对特征选择的影响,可以有效限定 Wrapper 算法的特征空间,减小计算量;基于 GA 的 Wrapper 算法利用 IID 指数改进了种群初始化部分,提供了更优的初始种群。

2.3.1 基于 IID 指数的 Filter 特征子集筛选

针对研究区内植被不同生长阶段引起的类内差异增加、采集条件引起的类间差异减小问题,提升分类精度的关键环节在于构建类内和类间差异描述指标,引导自动特征选择过程,获取最优特征子集,增强特征空间的辨识能力。在 Filter 算法评价准则中,距离度量通常用于分离性、差异性或者辨识能力的度量^[20]。常用距离准则包括:欧氏距离、曼哈顿距离、马氏距离、切比雪夫距离等。为避免不同特征间可能存在数值差异问题、排除变量间相关性的影响,本研究选取了马氏距离。综合类内和类间马氏距离构建的指数作为 Filter 特征选择的评价标准。IID 指数计算公式表示为

$$W_{IID} = \frac{\alpha(N - n_{intra}) + (1 - \alpha)n_{inter}}{N} \tag{1}$$

式中 N ——特征总数量
 n_{intra} 、 n_{inter} ——各特征类内马氏距离和类间马氏距离分别按照从小到大和从大到小的方式进行排序后对应的序号

参数 $\alpha \in [0, 1]$, 用来控制类内距离和类间距离对 IID 的重要性。由式 (1) 可知, W_{IID} 表示特征重要性权值计算函数,用于衡量不同特征的重要性, IID 指数越大,表示特征越重要。当存在严重的端元差异时,可增大 α 提高类内距离对特征选择的影响;反之可减小 α 以增加类间距离的影响。

Filter 过程最优特征子集筛选主要是利用 IID 指数作为评价指标,完成对原始特征集的初步筛选,去除冗余信息。具体过程包括:

- (1) 输入 3 类光谱特征和 Gabor 变换获取的纹理特征,分别计算特征集中各特征的类内马氏距离和类间马氏距离。
- (2) 对各特征的类内马氏距离和类间马氏距离,分别按照升序和降序进行排列。
- (3) 确定 α 值,利用式 (1) 计算各特征的 IID 指数。
- (4) 设定 IID 指数阈值,删除冗余特征,输出经 Filter 过程筛选得到的最优特征子集,为后续 Wrapper 过程提供初始搜索空间。

2.3.2 联合 IID 结果和 GA 的 Wrapper 算法

基于 Wrapper 的特征选择方法需要在所有可能的特征子集上进行分类,并根据预先设定的精度评价指标选出最优的子集,计算效率较低^[21]。GA 是一种基于迭代的启发式算法^[22],具有良好的全局搜索能力,可用于大规模数据优化求解^[23]。因此,基于 GA 的 Wrapper 算法可以用于提升 Wrapper 过程的效率。考虑到标准 GA 中随机初始化方式容易陷入局部最优的问题, IID 指数可以为基于 GA 的 Wrapper 算法提供更好的初始种群。

- 联合 IID 结果和 GA 的 Wrapper 过程包括:
- (1) 问题编码。本文选用二进制编码方式。其中,字符串长度即特征个数,字符 1 代表该特征被选中,字符 0 表示未被选中。
 - (2) 种群初始化。标准 GA 是一种随机搜索算法,为了提高算法效率,避免算法陷入局部最优解,本文使用 Filter 阶段产生的 IID 指数来表征特征被选概率,构建初始种群。
 - (3) 适应度函数构建。如何构建适应度函数是 GA 的核心内容之一。在分类问题中,特征选择优化问题主要从分类精度和子集特征数两方面考虑。个体的评价准则为分类精度越高,且特征集中特征数据越少,则适应度函数越大。本文选择的适应度函数^[24]可以表示为

$$A = \beta C_f + (1 - \beta) \left(1 - \frac{\sum_{i=1}^{n_f} f_i}{n_f} \right) \tag{2}$$

式中 C_f ——特征子集 f 的分类精度
 f_i —— f 中第 i 个特征所对应的二进制值
 n_f —— f 对应二进制编码后的长度
 β 为权值,用来调节分类精度和特征数目对最终适应度的影响程度。若需要大幅减少特征空间维度,则减小 β 值;反之,通过增加 β 值来提高分类精度的重要性。

(4)最优个体选择。个体选择以适应度评价为基础,适应度越高,遗传到下一代的概率越大。引入轮盘赌方法^[25]作为选择函数来描述适应度与遗传概率之间关系,则某特征子集被选中概率 P_i 计算公式为

$$P_i = \frac{A(i)}{\sum_{i=1}^{N_c} A(i)}$$

(3)

式中 N_c ——候选特征子集数目
式(3)表明,任一特征子集被选中概率与其适应度成正比。

(5)基因重组和变异。对于二进制编码的染色体,基因重组是随机将染色体对上一个或多个相同位置的编码进行相互对换。目的是通过信息互换、组合来增加群体多样性,实现全局最优搜索。变异是染色体上某一点位上发生基因改变的过程,它决定了 GA 局部搜索能力。对于步骤(4)中被选中的个体,将在此步骤进行基因重组和变异。

(6)停止条件判定。若不满足停止条件,则重复步骤(3)~(5),直至产生最优特征子集。

3 结果与分析

3.1 IIDGA 特征选择结果及精度评定

3.1.1 原始特征空间构建

对预处理后的 Landsat 影像提取光谱和 Gabor 小波纹理特征,构建了 90 维原始特征空间(表 2)。其中,为减小端元差异造成的误差,利用 WCVA 对 Landsat TM 影像的 7 个波段进行筛选,确定了由其中 5 个波段构成波段组合特征。在纹理特征中,选用 Gabor 滤波器提取 80 个纹理特征。由于不同特征的取值范围不同,将原始特征值归一化到[0, 255]。

3.1.2 基于 IIDGA 的最优特征空间

分别计算原始特征空间中各特征的 IID 指数($\alpha=0.5$),按照从大到小排序并取前 40 维特征构建 GA 初始种群(种群个数设置为 30)。根据式(2)计算种群个体的适应度(其中 $\beta=1$),从中选择遗传至下一代的个体。其中参数设置根据先验知识确定。经 IIDGA 获得的最优特征子集如表 3 所示。

为了验证最优特征选择的重要性,比较了传统波段组合、光谱和纹理特征全集与 IIDGA 优选特征的 FCLS 模型解混效果。从解混精度来看,基于光谱-纹理多特征描述的特征空间能够有效提高解混精度。在仅使用光谱特征时,基于波段组合特征构建的特征空间,对应的解混结果 RMSE 为 0.189;综合利用光谱和纹理特征全集构建的特征空间,解混

表 2 原始特征空间构建

Tab.2 Original feature space construction

特征类别	序号	特征
光谱特征	波段组合特征	1 ~ 5 B1 ~ B3, B5, B7 亮度 (K - T Brightness), 绿度 (K - T Greenness), 湿度 (K - T Wetness)
	穗帽变换特征	6 ~ 8 NDVI, NDWI
	遥感指数特征	9 ~ 10 B1_S1_15°, B1_S1_30°, 11 ~ 22 ∴ B1_S1_165°, B1_S1_180° B2_S1_15°, B2_S1_30°, 23 ~ 34 ∴ B2_S1_165°, B2_S1_180° B3_S1_15°, B3_S1_30°, 1 × 1 窗口 (15° 间隔, 提取 (0°, 35 ~ 46 180°] 12 个 方向) B3_S1_165°, B3_S1_180° B5_S1_15°, B5_S1_30°, 47 ~ 58 ∴ B5_S1_165°, B5_S1_180° B7_S1_15°, B7_S1_30°, 纹理特征 59 ~ 70 ∴ B7_S1_165°, B7_S1_180°
		71 ~ 74 B1_S2_45°, B1_S2_90°, B1_S2_135°, B1_S2_180° B2_S2_45°, B2_S2_90°, 2 × 2 窗口 (45° 间隔, 提取 (0°, 79 ~ 82 180°] 4 个 方向) B2_S2_135°, B2_S2_180° B3_S2_45°, B3_S2_90°, B3_S2_135°, B3_S2_180° B5_S2_45°, B5_S2_90°, B5_S2_135°, B5_S2_180° B7_S2_45°, B7_S2_90°, 87 ~ 90 B7_S2_135°, B7_S2_180°

注:BX_SY_Z°表示 X×X 窗口下,Band Y(BY)在 Z°方向上的纹理。下同。

表 3 IIDGA 构建的最优特征子集

Tab.3 Optimal feature subset selected by IIDGA

特征类别	序号	特征
光谱特征	1	B5
	2	B7
	3	K - T Brightness
	4	K - T Greenness
	5	K - T Wetness
	6	NDVI
纹理特征	7	B7_S1_30°
	8	B7_S1_45°
	9	B7_S1_60°
	10	B7_S1_75°
	11	B7_S1_105°
	12	B7_S1_120°
	13	B7_S1_135°
	14	B5_S2_135°
	15	B7_S2_135°

结果 RMSE 为 0.186。从特征数量对解混结果的影响来看,特征数量不足与过多均会造成解混精度的下降。使用 IIDGA 算法对光谱和纹理特征全集进行特征选择后构建的最优特征空间,解混结果 RMSE 为 0.180,明显优于前两种组合。综合而言,不同特征构建的特征空间对解混精度有较大的影响。因此,全面的特征描述与最优特征的筛选是构建最优特征空间的核心环节。

3.2 特征选择算法对比与结果分析

为了进一步验证 IIDGA 自动特征选择算法的优越性,将本文提出的 IIDGA 与基于 IID 指数的 Filter 算法和基于标准 GA 的 Wrapper 算法进行对比。分别利用 3 类算法对原始光谱和纹理全集进行特征选择,构建对应的特征空间,并选择 FCLS 模型解混,最终从解混精度和效率两方面对各算法的识别效果进行对比和分析。其中,IIDGA 特征选择结果如表 3 所示,仅基于 IID 指数的(Filter 算法)特征选择结果和基于标准 GA 的(Wrapper 算法)特征选择结果如表 4 所示。各算法精度和计算时间如表 5 所示。

结合表 3 ~ 5,基于 IID 指数的 Filter 算法在选择过程中仅根据类内-类间距离筛选特征,尽管耗时最短,但精度明显低于另外两个以分类结果作为特征选择标准的 Wrapper 算法。与基于标准 GA 的 Wrapper 算法相比,由于 IIDGA 根据 IID 指数生成了更优的初始种群,显著提高了 GA 的稳定性和运算效率。实验结果表明,IIDGA 较基于标准 GA 的 Wrapper 算法在精度和收敛速度上均有显著提升。因此,证明了本文提出的 IIDGA 综合了 Filter 算法的高效性和 Wrapper 算法高精度的优势,能够准确、

快速地剔除冗余特征,构建最优特征空间。

表 4 基于不同类别特征选择算法的最优特征子集
Tab.4 Optimal feature subsets selected by different types of feature selection methods

特征类别	序号	基于 IID 指数的特征 选择结果 (Filter 算法)	基于标准 GA 的特征选 择结果 (Wrapper 算法)
光谱特征	1	B5	B5
	2	B7	B7
	3	K - T Greenness	K - T Greenness
	4	K - T Wetness	K - T Wetness
	5	NDVI	NDVI
纹理特征	6	B3_S1_120°	B3_S1_120°
	7	B5_S1_120°	B5_S1_135°
	8	B5_S1_150°	B7_S1_30°
	9	B5_S1_165°	B7_S1_60°
	10	B7_S1_120°	B7_S1_105°
	11	B7_S1_135°	B7_S1_120°
	12	B7_S1_150°	B7_S1_135°
	13	B3_S2_135°	B7_S1_150°
	14	B5_S2_135°	B1_S2_135°
	15	B7_S2_135°	B5_S2_135°

表 5 不同特征选择算法性能对比

Tab.5 Performance comparison on different feature selection algorithms

评价指标	基于 IID 指数 的 Filter 算法	基于标准 GA 的 Wrapper 算法	IIDGA
RMSE	0.186	0.185	0.180
耗时/s	1.898	7.500	4.378

不同特征选择算法解混结果如图 3 所示。参考丰度图由 GeoEye 影像生成。每一行的解混结果对应一个端元,每一列对应不同特征选择算法构建特征空间的解混结果。解混结果中,丰度越高,越接近

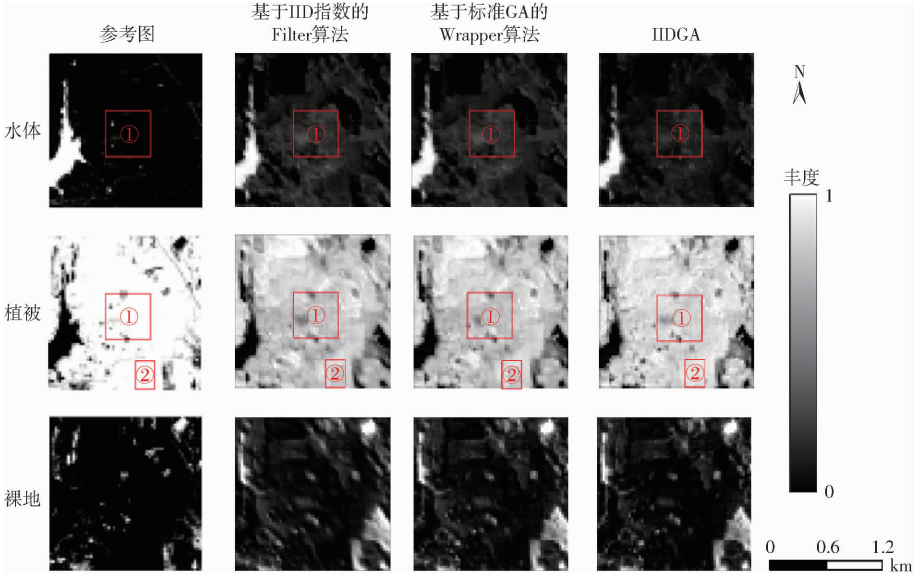


图 3 不同特征选择算法解混结果比较

Fig.3 Comparisons on different feature selection methods on unmixing results

白色(像素值更接近 1);反之,越接近黑色(像素值更接近 0)。

结合水体参考图和植被参考图,区域①内地物以植被(红树林)为主。在水体解混结果中,基于 IID 指数的 Filter 算法、基于标准 GA 的 Wrapper 算法以及 IIDGA 的区域①部分,3 类算法构建的特征空间解混结果中仍然存在水体和植被混合的情况,这是由于红树林通常生长在热带、亚热带的潮间带河口、海岸,受潮汐影响而发生周期性淹没^[26-27]。因此,研究区的水体与植被(红树林)容易发生混合,减小了类间的差异。本文提出的 IIDGA 在区域①内,颜色更暗(即像素值更小),说明 IIDGA 构建的特征空间较基于 IID 指数的 Filter 算法、基于标准 GA 的 Wrapper 算法能够更有效地识别植被;在植被解混结果中,基于 IID 指数的 Filter 算法、基于标准 GA 的 Wrapper 算法以及 IIDGA 的区域①、区域②部分,IIDGA 的解混结果颜色更亮,说明该算法能够更

有效地缩小端元的类内差异,区分不同生长时期的植被。这验证了本文提出方法在植被识别中的有效性。

4 结论

(1)不同特征组合对植被识别精度的影响实验中,IIDGA 表现优异,RMSE 为 0.180,而传统波段组合、光谱和纹理特征全集的 RMSE 分别为 0.189 和 0.186,这表明,顾及端元变异性的 IIDGA 所构建的特征空间是最优的,更有利于植被准确识别。

(2)不同自动特征选择算法实验中,基于 IID 指数的 Filter 算法耗时最短,但 RMSE 仅为 0.186;基于标准 GA 的 Wrapper 算法精度略有提升,但是耗时明显增加。IIDGA 在提升精度的同时减小 Wrapper 算法的计算时间,展现了 IIDGA 在植被识别应用中的优越性,能够很好地平衡精度和效率。

参 考 文 献

[1] 付永硕,张晶,吴兆飞,等.中国植被物候研究进展及展望[J].北京师范大学学报(自然科学版),2022,58(3):424-433.
FU Yongshuo, ZHANG Jing, WU Zhaoifei, et al. Vegetation phenology response to climate change in China[J]. Journal of Beijing Normal University(Natural Science), 2022, 58(3): 424-433. (in Chinese)

[2] XIONG T, DU S, ZHANG H, et al. Satellite observed reversal in trends of spring phenology in the middle-high latitudes of the Northern Hemisphere during the global warming hiatus[J]. Global Change Biology, 2023, 29(8): 2227-2241.

[3] 解毅,王佳楠,刘钰.基于 Sentinel-1/2 数据特征优选的冬小麦种植区识别方法研究[J].农业机械学报,2024,55(2):231-241.
XIE Yi, WANG Jia'nan, LIU Yu. Research on winter wheat planting area identification method based on Sentinel-1/2 data feature optimization[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(2): 231-241. (in Chinese)

[4] 贾凯,陈水森,蒋卫国.粤港澳大湾区红树林长时间序列遥感监测[J].遥感学报,2022,26(6):1096-1111.
JIA Kai, CHEN Shuisen, JIANG Weiguo. Long time-series remote sensing monitoring of mangrove forests in the Guangdong-Hong Kong-Macao Greater Bay Area[J]. National Remote Sensing Bulletin, 2022, 26(6): 1096-1111. (in Chinese)

[5] 杨斌,王斌.光谱解混技术及其应用研究进展[J].激光与光电子学进展,2021,58(16):76-103.
YANG Bin, WANG Bin. Research advances of spectral unmixing technology and its applications[J]. Laser and Optoelectronics Progress, 2021, 58(16): 76-103. (in Chinese)

[6] YU J, CHEN D, LIN Y, et al. Comparison of linear and nonlinear spectral unmixing approaches: a case study with multispectral TM imagery[J]. International Journal of Remote Sensing, 2017, 38(3): 773-795.

[7] 陈晋,马磊,陈学泓,等.混合像元分解技术及其进展[J].遥感学报,2016,20(5):1102-1109.
CHEN Jin, MA Lei, CHEN Xuehong, et al. Research progress of spectral mixture analysis[J]. National Remote Sensing Bulletin, 2016, 20(5): 1102-1109. (in Chinese)

[8] SOMERS B, ASNER G P, TITS L, et al. Endmember variability in spectral mixture analysis: a review[J]. Remote Sensing of Environment, 2011, 115(7): 1603-1616.

[9] LI J, CHENG K, WANG S, et al. Feature selection: a data perspective[J]. ACM Computing Surveys, 2017, 50(6): 1-45.

[10] CHANDRASHEKAR G, SAHIN F. A survey on feature selection methods[J]. Computers and Electrical Engineering, 2014, 40(1): 16-28.

[11] KHAIRE U M, DHANALAKSHMI R. Stability of feature selection algorithm: a review[J]. Journal of King Saud University-Computer and Information Sciences, 2022, 34(4): 1060-1073.

[12] 李瑞利,杨芳,王辉,等.红树林保护与修复标准发展现状及对策[J].北京大学学报(自然科学版),2022,58(5):916-928.
LI Ruili, YANG Fang, WANG Hui, et al. Current development status and countermeasures of mangrove protection and restoration standards[J]. Acta Scientiarum Naturalium Universitatis Pekinensis, 2022, 58(5): 916-928. (in Chinese)

[13] WANG Y, JIANG B, LIANG S, et al. Surface shortwave net radiation estimation from Landsat TM/ETM+ data using four

- machine learning algorithms[J]. *Remote Sensing*, 2019, 11(23): 2847.
- [14] 宇洁, 叶勤, 林怡. 多光谱影像混合像元解混的加权变异系数分析法[J]. *农业机械学报*, 2018, 49(9): 154–159.
YU Jie, YE Qin, LIN Yi. Novel weighted coefficient of variation analysis approach for endmember variability issue in unmixing process of multi-spectral imagery[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2018, 49(9): 154–159. (in Chinese)
- [15] ZHU Z, WOODCOCK C E, ROGAN J, et al. Assessment of spectral, polarimetric, temporal, and spatial dimensions for urban and peri-urban land cover classification using Landsat and SAR data[J]. *Remote Sensing of Environment*, 2012, 117: 72–82.
- [16] LI S, XU L, JING Y, et al. High-quality vegetation index product generation: a review of NDVI time series reconstruction techniques[J]. *International Journal of Applied Earth Observation and Geoinformation*, 2021, 105: 102640.
- [17] CHAI L, JIANG H, CROW W T, et al. Estimating corn canopy water content from normalized difference water index (NDWI): an optimized NDWI-based scheme and its feasibility for retrieving corn VWC[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2021, 59(10): 8168–8181.
- [18] YARBROUGH L D, EASSON G, KUSZMAUL J S. Proposed workflow for improved Kauth–Thomas transform derivations[J]. *Remote Sensing of Environment*, 2012, 124: 810–818.
- [19] 杜晓冬, 滕光辉, TOMAS N, 等. 基于声谱图纹理特征的蛋鸡发声分类识别[J]. *农业机械学报*, 2019, 50(9): 215–220.
DU Xiaodong, TENG Guanghui, TOMAS N, et al. Classification and recognition of laying hens' vocalization based on texture features of spectrogram[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2019, 50(9): 215–220. (in Chinese)
- [20] YAO X, WANG X D, ZHANG Y X, et al. Summary of feature selection algorithms[J]. *Control and Decision*, 2012, 27(2): 161–166.
- [21] 邱云飞, 高华聪. 混合 Filter 与改进自适应 GA 的特征选择方法[J]. *计算机工程与应用*, 2021, 57(11): 95–102.
QIU Yunfei, GAO Huacong. Hybrid Filter and improved adaptive GA for feature selection[J]. *Computer Engineering and Application*, 2021, 57(11): 95–102. (in Chinese)
- [22] KHURMA R A, ALJARAHI I, SHARIEH A, et al. A review of the modification strategies of the nature inspired algorithms for feature selection problem[J]. *Mathematics*, 2022, 10(3): 464.
- [23] LI S, WU H, WAN D, et al. An effective feature selection method for hyperspectral image classification based on genetic algorithm and support vector machine[J]. *Knowledge-Based Systems*, 2011, 24(1): 40–48.
- [24] 卓莉, 郑璟, 王芳, 等. 基于 GA–SVM 封装算法的高光谱数据特征选择[J]. *地理研究*, 2008, 27(3): 493–501.
ZHUO Li, ZHENG Jing, WANG Fang, et al. A genetic algorithm based wrapper feature selection method for classification of hyper spectral data using support vector maching[J]. *Geographical Research*, 2008, 27(3): 493–501. (in Chinese)
- [25] EWEESS A A, AI-QANESS M A A, ABUALIGAH L, et al. Boosting arithmetic optimization algorithm with genetic algorithm operators for feature selection: case study on cox proportional hazards model[J]. *Mathematics*, 2021, 9(18): 2321.
- [26] 李皓宇, 彭逸生, 刘嘉健, 等. 粤东沿海红树林物种组成与群落特征[J]. *生态学报*, 2016, 36(1): 252–260.
LI Haoyu, PENG Yisheng, LIU Jiajian, et al. Current state of mangrove floristic composition and characteristics of communities on the eastern coast of Guangdong Province[J]. *Acta Ecologica Sinica*, 2016, 36(1): 252–260. (in Chinese)
- [27] 吴培强, 任广波, 张程飞, 等. 无人机多光谱和 LiDAR 的红树林精细识别与生物量估算[J]. *遥感学报*, 2022, 26(6): 1169–1181.
WU Peiqiang, REN Guangbo, ZHANG Chengfei, et al. Fine identification and biomass estimation of mangroves based on UAV multispectral and LiDAR[J]. *National Remote Sensing Bulletin*, 2022, 26(6): 1169–1181. (in Chinese)