

doi: 10.6046/zrzyg.2022493

引用格式: 梁锦涛,陈超,张自力,等. 一种融合指数与主成分分量的随机森林遥感图像分类方法[J]. 自然资源遥感,2023,35(3):35-42. (Liang J T, Chen C, Zhang Z L, et al. A random forest - based method integrating indices and principal components for classifying remote sensing images[J]. Remote Sensing for Natural Resources, 2023, 35(3): 35-42.)

一种融合指数与主成分分量的随机森林遥感图像分类方法

梁锦涛¹, 陈超², 张自力³, 刘志松⁴

(1. 浙江海洋大学海洋科学与技术学院, 舟山 316022; 2. 苏州科技大学地理科学与测绘工程学院, 苏州 215009; 3. 浙江省生态环境监测中心(浙江省生态环境监测预警及质控研究重点实验室), 杭州 310012; 4. 浙江海洋大学信息工程学院, 舟山 316022)

摘要: 准确获取土地利用/覆盖(land use/land cover, LULC)信息,对区域空间规划和可持续发展具有重要指导意义。地表形态的复杂性、地物类型的多样性、遥感图像特征的非线性给传统的遥感图像分类方法带来了挑战,且传统方法未充分利用遥感图像所蕴含的丰富信息。文章发展了一种随机森林遥感图像分类方法,融合指数与主成分分量开展 LULC 信息提取。首先,选择研究区影像进行云量筛选、影像中值合成,得到年际遥感影像;其次,计算多种指数,提取主成分分量,将其融入到遥感图像波段堆栈中;然后,构建不同机器学习算法分类器;最后,基于混淆矩阵,使用总体精度与 Kappa 系数对分类结果进行评估。在杭州湾区域的实验结果表明,植被指数、水体指数、建筑物指数与主成分分量的辅助决策能够提高分类的准确性,总体精度和 Kappa 系数分别为 91.42% 和 0.894 2,高于传统随机森林、分类回归树和支持向量机等方法。融合指数和主成分分量的遥感图像分类方法能够准确地提取遥感图像中的地表覆盖特征,得到高精度的土地利用分类结果,为地表精细分类提供方法支持。

关键词: 随机森林; 土地利用/覆盖; 指数; 主成分分析; 精度评价

中图法分类号: TP 79 **文献标志码:** A **文章编号:** 2097-034X(2023)03-0035-08

0 引言

土地利用/土地覆盖(land use/land cover, LULC)作为全球变化最重要基础地理信息之一,是人类活动在地球表面最宏观的表现,也是全球环境变迁研究中的核心议题。准确、高分辨率、高频次获取 LULC 一直是遥感应用的热点领域,对于城市扩展研究、区域规划和人类社会可持续发展具有重要意义^[1]。

随着遥感技术的高速发展,LULC 产品在各种时间和地理尺度上的开发成为可能^[2]。同时,随着分类方法的发展,随机森林(random forest, RF)、分类回归树(classification and regression tree, CART)、支持向量机(support vector machine, SVM)等机器学习算法在 LULC 领域得到更广泛的应用^[3-5]。这些方法为更高精度的 LULC 提取开辟了新的可能性。

近年来,机器学习算法凭借其整合有限训练数据、适应复杂区域的优秀性能,被广泛应用于遥感目标识别及专题信息提取领域^[6]。Wu 等^[7]利用全球覆盖产品和兴趣点数据,基于 RF 算法对城市 LULC 的核心驱动因素进行了研究;Rodriguez - Galiano 等^[8]研究了 RF 分类器在复杂区域土地覆盖分类中的应用性能;武复宇等^[9]研究了改进的多级联 RF 在高光谱影像分类中的性能;Ge 等^[10]评估了 4 种机器学习算法在绿洲-沙漠镶嵌景观 LULC 分类方面的性能;肖湘文等^[11]对比了机器学习算法在北极冰山识别中的精度;董娟等^[12]基于高分辨率影像对比了机器学习算法在珊瑚礁地貌提取中的精度。然而,传统机器学习算法存在单纯依靠原始图像数据的局限性,如何充分利用遥感图像所蕴含的信息提高图像分类精度和信息提取的准确性是值得思考的问题。

为了充分挖掘、利用遥感图像所蕴含的丰富信

收稿日期: 2022-12-26; 修订日期: 2023-03-26

基金项目: 国家自然科学基金项目“人类活动影响下的群岛区域海岸线时空演变机制分析”(编号: 42171311)资助。

第一作者: 梁锦涛(1998-),男,硕士研究生,研究方向为海岸带环境遥感。Email: liangjintao@zjou.edu.cn。

通信作者: 陈超(1982-),男,博士,教授,研究方向为海岸带环境遥感。Email: ayang198206@163.com。

息,更为准确地获取区域 LULC 信息,本研究在传统 RF 方法的基础上,发展了一种融合指数与主成分分量的 RF 遥感图像分类方法。本研究选择杭州湾为研究区开展验证实验,对比了本文方法与传统 RF, CART 和 SVM 这 3 种机器学习算法,结合高分辨率 LULC 产品探究了不同机器学习算法在 LULC 提取方面的性能差异。本研究为 LULC 信息准确获取提供了支持,对区域规划以及可持续发展具有重要意义。

1 研究区概况与数据源

1.1 研究区概况

杭州湾位于我国浙江省东北部,是钱塘江河口

区外海滨部分,湾口南接宁绍平原,北邻上海,东靠舟山群岛,是重要的生态监测区和世界著名的喇叭形潮汐河口^[13-14]。该区地势平坦,江河湖泊多,湿地广阔,泥滩资源丰富,属于亚热带海洋性季风气候^[15]。作为中国东部经济发展的核心区域,杭州湾是沿海开放带、长江经济带、长江三角洲城市群与“一带一路”等多重国家战略的交汇点,影响超过 6 000 万人^[16]。由于受自然条件和人类活动的影响,随着城市化进程的加快,杭州湾区域地表覆盖类型变换频繁且破碎度高,迫切需要高精度的土地利用分类算法准确获取 LULC 信息。本文以环杭州湾城市群为研究对象,研究区包括上海市的 3 个县级行政区和浙江省的 9 个县级行政区,区域总面积约 11 429.80 km²,具体如图 1 所示。

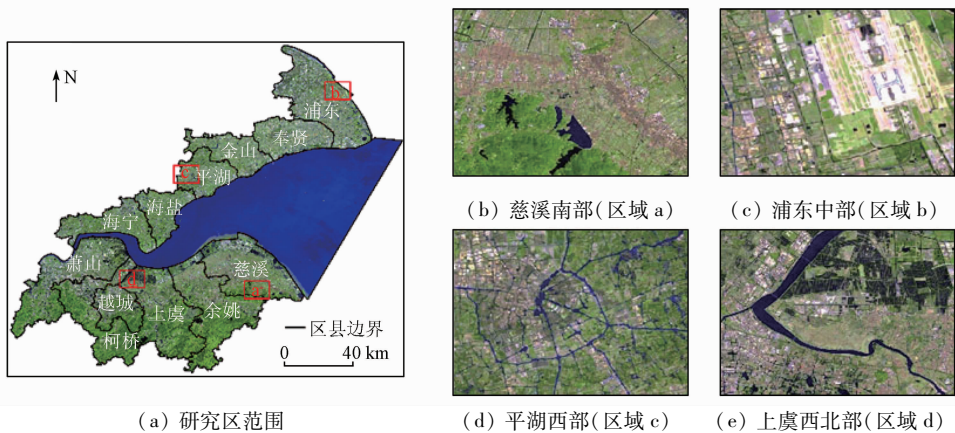


图 1 研究区以及局部对比区域

(Landsat8 OLI B7(R), B6(G), B4(B) 彩色合成)

Fig.1 Location of study area and local comparison area

1.2 数据获取与处理

本研究所用数据为 2020 年的 Landsat8 OLI(operational land imager) 遥感影像。Landsat8 卫星是美国陆地卫星计划的第 8 颗,成像宽度为 185 km × 185 km^[17-19],OLI 是 Landsat8 卫星上搭载的陆地成像仪,包括 30 m 空间分辨率的多光谱影像和 15 m 空间分辨率的全色影像。基于 Google Earth Engine (GEE) 云平台,获取 2020 年覆盖杭州湾区域的地表反射率数据集(LANDSAT/LC08/C01/T1_SR),并使用中值滤波法将 5—12 月份的影像合成为一景。

2 研究方法

本文发展了融合指数和主成分分量的 RF 分类算法,基于 RF 算法开展遥感图像分类研究,提取杭州湾区域 LULC 信息技术流程如图 2 所示,主要包括以下步骤:①选择覆盖研究区的 Landsat 卫星遥感地表反射率数据集(LANDSAT/LC08/C01/T1_

SR),进行云量筛选以及影像中值合成,得到年际遥感影像;②计算归一化植被指数(normalized difference vegetation index, NDVI)、归一化建筑物指数(normalized difference building index, NDBI)和改进归一化差异水体指数(modified normalized difference

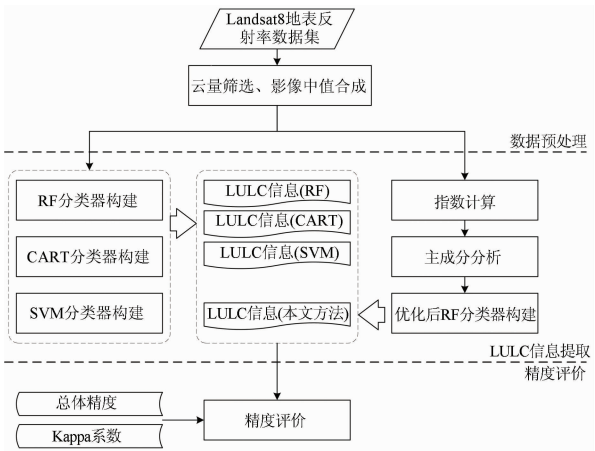


图 2 研究流程

Fig.2 Flowchart of the study

water index, MNDWI),并通过主成分分析(principal component analysis,PCA)提取主成分分量,将其融入到遥感图像波段堆栈中;③利用 RF 分类器开展土地利用分类研究,识别建设用地、林地、水体、耕地、裸地、滩涂等地物类型,获取研究区的 LULC 信息;④基于混淆矩阵,使用总体精度(overall accuracy, OA)和 Kappa 系数对分类结果进行评估,并与传统 RF,CART 和 SVM 结果进行比较。

2.1 数据预处理

基于 GEE 云平台,根据数据质量,使用云掩模操作筛选云量小于 10% 的卫星遥感图像^[20-21]。为了创建年际复合影像,基于像素计算图像堆栈的中值,并将这个中值赋予一景新的影像^[22],得到能够清晰、完整显示研究区地表覆盖信息的遥感图像。

2.2 用于分类的机器学习算法

本研究使用多种机器学习算法提取 LULC 结果。SVM 是一种非参数统计分类器,主要特征是为线性可分数据选择线性决策边界,目标是解决凸二次优化问题,克服局部极值困境,以获得理论上的全局最优解^[23]。CART 是一种基于规则的分类器,它发现并归纳预测变量和目标属性之间的关系,基于训练数据集使用二进制拆分对输入空间进行递归分区,通过评估和修剪生成最优二叉树进行预测^[24]。

然而,SVM 的分类准确性很难用其核函数进行解释,CART 学习能力差,易受噪声数据干扰。相比之下,RF 凭借其更强的泛化能力和学习能力,在遥感图像的分类领域具有更优的表现性^[25]。RF 是一种基于决策树集成的机器学习算法。决策树是一种简单、快捷且无参的分类方法,大多数情况下有较高的分类精度,但是当数据过于复杂或存在噪声时,过拟合现象明显^[26]。RF 分类则是组合多棵决策树对样本的训练结果,最终分类结果由众多决策树投票得到,削弱了过拟合问题,有效提高了分类器的泛化能力^[27]。其基本思路如下:首先,用自助法有放回地随机抽取样本训练集,每次抽取数量约为总量的 2/3,剩余 1/3 用来估计训练内部误差;然后,为抽取样本集生成各自的分类决策树,合并所有决策树,生成一个分类森林;最后,综合所有决策树的结果,通过投票策略完成最终分类^[27],即

$$H(x) = \arg \max_Y \sum_{i=1}^k I[h_i(x) = Y] \quad , \quad (1)$$

式中: $H(x)$ 为组合分类模型; k 为建立的决策树模型个数; h_i 为第 i 个决策树分类模型; Y 为输出变量; $I[\]$ 为示性函数。使用多数投票决策的方式来确定最终的分类结果。

2.3 融合指数与主成分分量的 RF 分类

RF 算法具有高可解释性和可靠性,但由于地物光谱特征同物异谱和异物同谱的复杂性^[28],在处理复杂数据时的性能以及分类精度提升方面,RF 算法还存在着优化的空间。在传统 RF 算法的基础上,本研究引入光谱指数与主成分分量,通过组合光谱波段与辅助特征提高分类器的学习能力,这在一定程度上提高了单个决策树的强度,实现 LULC 信息提取精度的优化。

2.3.1 光谱指数

NDVI,NDBI 和 MNDWI 对植被、建设用地和水体具有良好的表现性。将计算后的指数输入分类影像堆栈中,实现多参数辅助类别决策,提高分类准确性。指数计算公式分别为:

$$NDVI = \frac{NIR - RED}{NIR + RED} \quad , \quad (2)$$

$$NDBI = \frac{SWIR - NIR}{SWIR + NIR} \quad , \quad (3)$$

$$MNDWI = \frac{GREEN - SWIR}{GREEN + SWIR} \quad , \quad (4)$$

式中 NIR , RED , $GREEN$ 和 $SWIR$ 分别为近红外波段、红光波段、绿光波段和短波红外波段反射率。

2.3.2 主成分分量

融合多特征的方法可以极大地丰富遥感数据的信息内容^[29],原本在一种特征空间很难区分的 2 种地物可能在另一特征空间上被分开,但过多的图像信息堆叠会造成信息冗余,降低分类效率及精度,不利于遥感数据的有效利用^[30]。本文运用 PCA 对影像信息进行凝练。PCA 是一种通用的降维和数据分析工具,主要用于机器学习和数据挖掘等重要研究领域^[31]。PCA 的本质是通过线性变换将高维空间中的数据样本投影到低维空间中,同时尽可能保留原始数据特征。运用 PCA 对影像进行处理能够实现特性信息的压缩和图像纹理的增强^[31-32]。图像经 PCA 处理后,前 3 个主成分分量包括了超过 95% 的影像信息,本文将前 3 个主成分分量添加到分类影像堆栈中并应用于 RF 分类器的学习,并以 PCA 影像图为底图进行训练样本点的采集,更好地消除特征向量中各特征之间的相关性,增强 RF 分类器中单个决策树的强度,实现分类精度的提升。

2.3.3 分类方案与样本点

考虑 Landsat 数据的空间分辨率,参照《全国遥感监测土地利用/覆盖分类体系》并结合湾区地物分布特征,本文将研究区分类体系确定为 6 类:建

设用地、林地、水体、耕地、裸地、滩涂,具体方案如表 1 所示。

表 1 分类方案、样本点及图像特征
Tab.1 Classification scheme, sample points and image features

类别	定义	样本点/个	遥感图像特征 (标准假彩色合成)
建设用地	城乡居民点及工矿交通等用地	210	
林地	树木、灌木和其他林地	220	
水体	自然陆地水域、水利设施用地和养殖池塘	130	
耕地	长年可正常耕作的农田	250	
裸地	地表为土壤或岩石,土地基本不被植被覆盖	100	
滩涂	海高低潮位之间的潮汐淹没区	100	

2.4 精度评价

基于混淆矩阵,计算 OA 与 Kappa 系数来评估分类准确性。OA 指被正确分类的像元数目(沿着

混淆矩阵的对角线分布)占像元总数的比例,直接反映正确分类的比例^[33-34]。Kappa 系数则用于检验模型预测结果和实际分类结果是否一致。OA 和 Kappa 系数的计算公式分别为:

$$OA = \frac{P_c}{P_n} \times 100\% \quad , \quad (5)$$

$$Kappa = \frac{N \sum_{i=1}^r x_{ii} - \sum_{i=1}^r (x_{i+} x_{+i})}{N^2 - \sum_{i=1}^r (x_{i+} x_{+i})} \quad , \quad (6)$$

式中: P_c 为正确分类的像元数; P_n 为像元总数; r 为误差矩阵中的行数和列数; x_{ii} 为第 i 行第 i 列的观测值数; x_{i+} 为第 i 行的边际总和; x_{+i} 为第 i 列的边际总和; N 为观察总数。

3 结果与分析

3.1 土地利用分类及精度评价

在 GEE 云平台的支持下,使用 RF,CART,SVM 和本文方法 4 种机器学习算法提取了 2020 年杭州湾地区 LULC 信息,分类结果如图 3 所示。

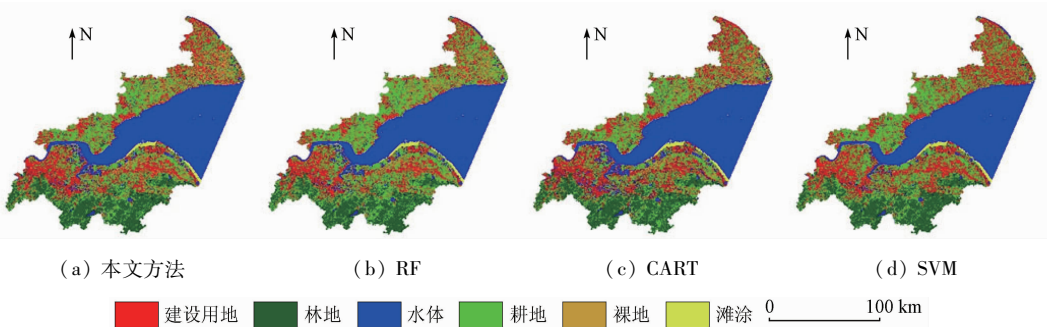


图 3 4 种方法分类结果
Fig.3 Classification results of four methods

基于混淆矩阵,使用 OA 和 Kappa 系数评价分类准确性,结果如表 2 所示。从表 2 中可以看出 RF,CART,SVM 和本文方法 OA 值分别为 88.78%,83.17%,77.23%,91.42%; Kappa 系数分别 0.861 5,0.792 6,0.716 1,0.894 2。相对于 RF,CART 和 SVM,本文方法有较明显的精度提升,分类精度更高,分类结果更为准确。

表 2 4 种机器学习算法分类精度对比
Tab.2 Classification accuracy of four machine learning algorithms

类别	本文方法	RF	CART	SVM
OA/%	91.42	88.78	83.17	77.23
Kappa 系数	0.894 2	0.861 5	0.792 6	0.716 1

3.2 分类结果对比

为了进一步研究不同机器学习算法在具体地物之间分类性能的差异,统计了 4 种分类算法提取的各类地物面积。选择慈溪南部(图 1 (b))、浦东中部(图 1 (c))、平湖西部(图 1 (d))、上虞西北部(图 1 (e))4 个地区对 4 种算法的结果进行可视化对比,具体如表 3 和图 4 所示,此外,与欧洲航天局(European Space Agency,ESA)基于 Sentinel-1 以及 Sentinel-2 制作的 10 m 空间分辨率土地覆盖数据 WorldCover | WORLDCOVER (esa-worldcover.org)(图 4(a),(f),(k),(p))进行对比,增加结果的可信性。

表 3 4 种分类方法各类别面积统计

Tab.3 Area statistics of each category in four classification methods (km²)

类别	建设用地	林地	水体	耕地	裸地	滩涂
ESA	2 851.35	2 884.21	605.99	3 812.34	1 274.23	141.40
本文方法	2 984.76	1 881.52	680.21	4 800.59	1 078.13	161.50
RF	2 545.11	1 776.72	675.00	5 398.44	1 029.98	166.13
CART	2 964.11	1 803.62	1 216.39	4 594.25	848.88	174.46
SVM	3 507.76	1 939.85	438.10	5 539.48	1.68	163.52

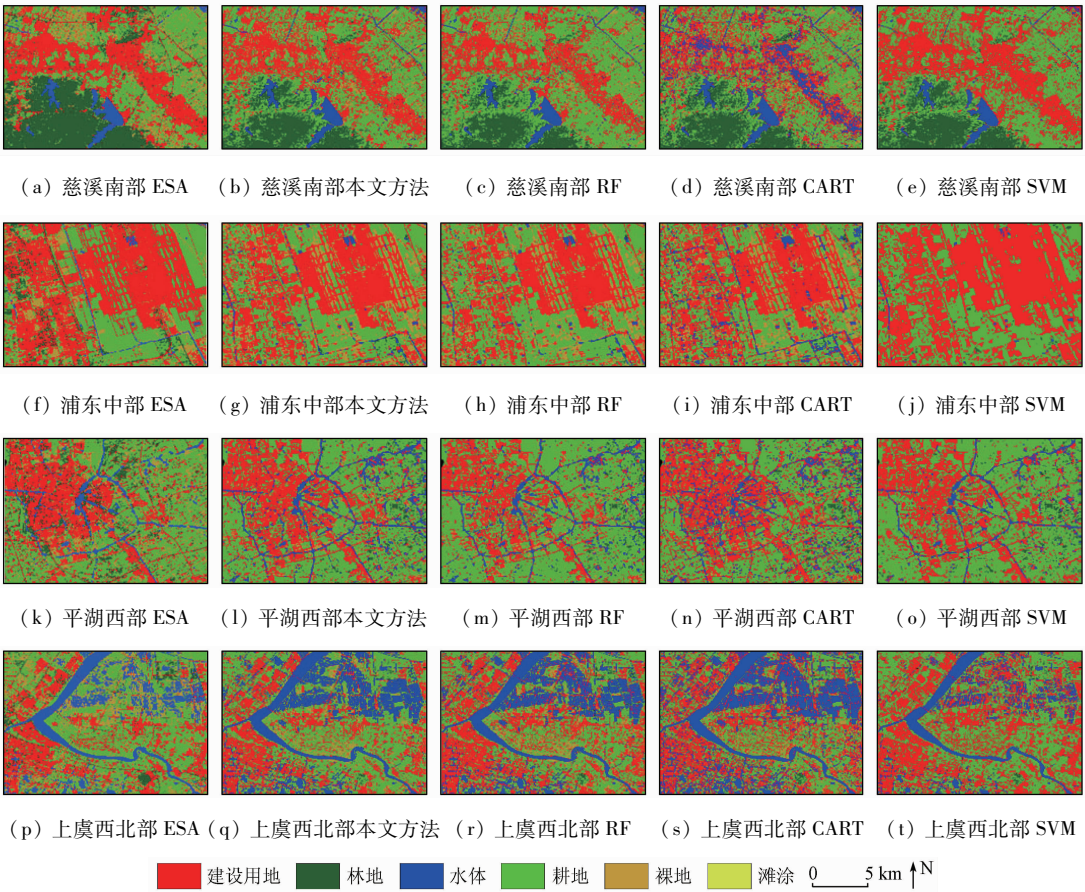


图 4 局部区域分类结果对比

Fig.4 Comparison of classification results in local area

根据表 3 和图 4 并结合 ESA 10 m 分辨率土地覆盖产品分析后发现,在慈溪南部,RF,CART,SVM 和本文方法能够较好区分建设用地与耕地、林地与耕地,但在建设用地与水体之间,CART 发生了错分,基于 CART 的杭州湾水体面积为 1 216.39 km²,远高于 ESA(605.99 km²),本文方法(680.21 km²),RF(675.00 km²)和 SVM(438.10 km²),这在一定程度上降低了 CART 分类的 OA 以及一致性。在浦东中部,SVM 难以区分建设用地与裸地,基于 SVM 的杭州湾裸地面积仅有 1.68 km²,远低于 ESA(1 274.23 km²),本文方法(1 078.13 km²),RF(1 029.98 km²)和 CART(848.88 km²)。在这方面,本文方法与 RF 表现出较好的分离性,精度高于 CART,且本文方法对于部分地物边界的分类更为准确,与 ESA 土地覆盖产品吻合度更高。在平湖西部区分建设用地与水体

混合的城区时,CART 表现出较差的效果,城区出现破碎水体,河道也大部分错分为建设用地。RF,SVM 和本文方法对细小水体分类效果较好。在上虞西北部,本文方法能够更好区分杭州湾水田与耕地,在建设用地与水体、耕地与水体之间的错分低于 RF,CART 和 SVM。

综上所述,植被指数、水体指数、建筑物指数等指数特征与主成分分量的辅助决策在一定程度上提高了遥感图像分类的准确性,本文方法能够获取较为准确的 LULC 信息。

4 结论与讨论

为了提高机器学习分类精度,本研究发展了一种融合指数与主成分分量的 RF 遥感图像分类方

法,选取杭州湾为研究区,结合 ESA 高分辨率 LULC 产品与 RF, CART 和 SVM 3 种机器学习算法进行了对比,探究了不同机器学习算法在海湾地区 LULC 信息提取方面的性能差异。主要结论如下:

1) 发展了融合指数和主成分分量的 RF 遥感图像分类方法, OA 和 Kappa 系数分别为 91.42% 和 0.894 2, 分类精度较高, 分类结果较为准确。

2) 植被指数、水体指数、建筑物指数等指数特征与主成分分量的辅助决策提高了分类的准确性。4 种分类方法在建设用地区与耕地、林地与耕地间的信息提取中都有较高的分离度。本文方法在建设用地区与裸地、建设用地区与水体、耕地与水体间的信息提取中具有更优于 RF, CART 和 SVM 的分离度。

遥感指数和主成分分量的添加在一定程度上提高了分类准确度,但由于混合像素、地物光谱特征复杂性的影响,在 30 m 空间分辨率的 Landsat 影像中部分建设用地区与裸地、林地与耕地的边界依旧较难区分,这在一定程度上影响了训练样本的准确性,从而导致错分现象的发生。验证本文方法在其他空间分辨率遥感图像上的适用性,融合更多遥感特征优化算法将是下一步的研究内容。

参考文献 (References):

- [1] Rawat J S, Kumar M. Monitoring land use/cover change using remote sensing and GIS techniques: A case study of Hawalbagh Block, District Almora, Uttarakhand, India[J]. The Egyptian Journal of Remote Sensing and Space Science, 2015, 18(1): 77–84.
- [2] Yao J, Wu J, Xiao C, et al. The classification method study of crops remote sensing with deep learning, machine learning, and Google Earth Engine[J]. Remote Sensing, 2022, 14(12): 2758.
- [3] 杨耘, 徐丽, 颜佩丽. 条件随机场框架下基于随机森林的城市土地利用/覆盖遥感分类[J]. 国土资源遥感, 2014, 26(4): 51–55. doi:10.6046/gtzyyg. 2014. 04. 09.
Yang Y, Xu L, Yan P L. Urban land use/cover remote sensing classification based on random forest under conditional random field framework[J]. Remote Sensing for Land and Resources, 2014, 26(4): 51–55. doi:10.6046/gtzyyg. 2014. 04. 09.
- [4] 熊华, 刘耀林, 车珊珊, 等. 基于支持向量机的土地利用变化模拟模型[J]. 武汉大学学报(信息科学版), 2009, 34(3): 366–369.
Xiong H, Liu Y L, Che S S, et al. Simulation model of land use change based on support vector machine[J]. Geomatics and Information Science of Wuhan University, 2009, 34(3): 366–369.
- [5] 杨曦光, 黄海军, 严立文, 等. 基于决策树方法的海岛土地利用分类研究[J]. 国土资源遥感, 2012, 24(2): 116–120. doi:10.6046/gtzyyg. 2012. 02. 21.
Yang X G, Huang H J, Yan L W, et al. Research on island land use classification based on decision tree method[J]. Remote Sensing for Land and Resources, 2012, 24(2): 116–120. doi:10.6046/gtzyyg. 2012. 02. 21.
- [6] Bangira T, Alfieri S M, Menenti M, et al. Comparing thresholding with machine learning classifiers for mapping complex water[J]. Remote Sensing, 2019, 11(11): 1351.
- [7] Wu H, Lin A, Xing X, et al. Identifying core driving factors of urban land use change from global land cover products and POI data using the random forest method[J]. International Journal of Applied Earth Observation and Geoinformation, 2021, 103: 102475.
- [8] Rodriguez – Galiano V F, Chimire B, Rogan J, et al. An assessment of the effectiveness of a random forest classifier for land – cover classification[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2012, 67: 93–104.
- [9] 武复宇, 王雪, 丁建伟, 等. 基于改进的多级联随机森林网络高光谱遥感影像分类[J]. 遥感学报, 2020, 24(4): 439–453.
Wu F Y, Wang X, Ding J W, et al. Hyperspectral remote sensing image classification based on improved multi – cascade random forest network[J]. Journal of Remote Sensing, 2020, 24(4): 439–453.
- [10] Ge G, Shi Z, Zhu Y, et al. Land use/cover classification in an arid desert – oasis mosaic landscape of China using remote sensed imagery: Performance assessment of four machine learning algorithms[J]. Global Ecology and Conservation, 2020, 22: e00971.
- [11] 肖湘文, 沈校熠, 柯长青, 等. 基于 Sentinel – 1A 数据的多种机器学习算法识别冰山的比较[J]. 测绘学报, 2020, 49(4): 509–521.
Xiao X W, Shen X Y, Ke C Q, et al. Comparison of multiple machine learning algorithms based on Sentinel – 1A data to identify icebergs[J]. Acta Geodaetica et Cartographica Sinica, 2020, 49(4): 509–521.
- [12] 董娟, 任广波, 胡亚斌, 等. 基于高分辨率遥感的珊瑚礁地貌单元体系构建和分类方法——以 8 波段 Worldview – 2 影像为例[J]. 热带海洋学报, 2020, 39(4): 116–129.
Dong J, Ren G B, Hu Y B, et al. Construction and classification of coral reef geomorphological unit system based on high – resolution remote sensing: Taking 8 – band Worldview – 2 image as an example[J]. Journal of Tropical Oceanography, 2020, 39(4): 116–129.
- [13] 贾明明, 刘殿伟, 王宗明, 等. 面向对象方法和多源遥感数据的杭州湾海岸线提取分析[J]. 地球信息科学学报, 2013, 15(2): 262–269.
Jia M M, Liu D W, Wang Z M, et al. Coastline extraction and analysis of Hangzhou Bay based on object – oriented method and multi – source remote sensing data[J]. Journal of Geo – Information Science, 2013, 15(2): 262–269.
- [14] Li D, Lu D, Wu M, et al. Examining land cover and greenness dynamics in Hangzhou Bay in 1985–2016 using Landsat time – series data[J]. Remote Sensing, 2017, 10(1): 32.
- [15] Liang H, Chen C, Wang K, et al. Long – term spatiotemporal changes in ecosystem services caused by coastal wetland type transformation in China’s Hangzhou Bay[J]. Journal of Marine Science and Engineering, 2022, 10(11): 1781.
- [16] 李加林, 王丽佳. 围填海影响下东海区主要海湾形态时空演变[J]. 地理学报, 2020, 75(1): 126–142.
Li J L, Wang L J. Spatial and temporal evolutions of the major bays in the East China Sea under the influence of reclamation[J]. Acta Geographica Sinica, 2020, 75(1): 126–142.
- [17] Chen H, Chen C, Zhang Z, et al. Changes of the spatial and tempo-

- ral characteristics of land – use landscape patterns using multi – temporal Landsat satellite data: A case study of Zhoushan Island, China[J]. Ocean and Coastal Management, 2021, 213: 105842.
- [18] Chen C, Chen H, Liao W, et al. Dynamic monitoring and analysis of land – use and land – cover change using Landsat multitemporal data in the Zhoushan Archipelago, China[J]. IEEE Access, 2020, 8: 210360 – 210369.
- [19] 许晓聪, 李冰洁, 刘小平, 等. 全球 2000—2015 年 30 m 分辨率逐年土地覆盖制图 [J]. 遥感学报, 2021, 25(9): 1896 – 1916.
- Xu X C, Li B J, Liu X P, et al. Mapping annual global land cover changes at a 30 m resolution from 2000 to 2015[J]. National Remote Sensing Bulletin, 2021, 25(9): 1896 – 1916.
- [20] 邹亚东, 何 亮, 张晓萍, 等. 基于 GEE 数据平台的北洛河流域 1970—2019 年土地利用结构变化特征 [J]. 水土保持通报, 2021, 41(6): 209 – 219.
- Zou Y D, He L, Zhang X P, et al. Characteristics of land use structure change in Beiluo River basin during 1970—2019 based on Google Earth Engine[J]. Bulletin of Soil and Water Conservation, 2021, 41(6): 209 – 219.
- [21] 陈慧欣, 陈 超, 张自力, 等. 一种基于 Google Earth Engine 云平台的潮间带遥感信息提取方法 [J]. 自然资源遥感, 2022, 34(4): 60 – 67. doi:10.6046/zrzyyg. 2022308.
- Chen H X, Chen C, Zhang Z L, et al. A method for extracting intertidal remote sensing information based on Google Earth Engine cloud platform[J]. Remote Sensing for Natural Resources, 2022, 34(4): 60 – 67. doi:10.6046/zrzyyg. 2022308.
- [22] Wang C, Jia M, Chen N, et al. Long – term surface water dynamics analysis based on Landsat imagery and the Google Earth Engine platform: A case study in the middle Yangtze River basin[J]. Remote Sensing, 2018, 10(10): 1635.
- [23] Chachondhia P, Shakya A, Kumar G. Performance evaluation of machine learning algorithms using optical and microwave data for LULC classification[J]. Remote Sensing Applications: Society and Environment, 2021, 23: 100599.
- [24] Li X, Chen D, Duan Y, et al. Understanding land use/land cover dynamics and impacts of human activities in the Mekong Delta over the last 40 years[J]. Global Ecology and Conservation, 2020, 22: e00991.
- [25] Pal M. Random forest classifier for remote sensing classification [J]. International Journal of Remote Sensing, 2005, 26(1): 217 – 222.
- [26] Cutler D R, Edwards Jr T C, Beard K H, et al. Random forests for classification in ecology[J]. Ecology, 2007, 88(11): 2783 – 2792.
- [27] Breiman L. Random forests[J]. Machine Learning, 2001, 45(1): 5 – 32.
- [28] 陈 超, 傅姣琪, 随欣欣, 等. 面向灾后水体遥感信息提取的知识决策树构建及应用 [J]. 遥感学报, 2018, 22(5): 792 – 801.
- Chen C, Fu J Q, Sui X X, et al. Construction and application of knowledge decision tree after a disaster for water body information extraction from remote sensing images[J]. Journal of Remote Sensing, 2018, 22(5): 792 – 801.
- [29] 张 磊, 宫兆宁, 王启为, 等. Sentinel – 2 影像多特征优选的黄河三角洲湿地信息提取 [J]. 遥感学报, 2019, 23(2): 313 – 326.
- Zhang L, Gong Z N, Wang Q W, et al. Wetland mapping of Yellow River Delta wetlands based on multi – feature optimization of Sentinel – 2 images[J]. Journal of Remote Sensing, 2019, 23(2): 313 – 326.
- [30] Wang M, Wan Y, Ye Z, et al. Remote sensing image classification based on the optimal support vector machine and modified binary coded ant colony optimization algorithm[J]. Information Sciences, 2017, 402: 50 – 68.
- [31] Abdi H, Williams L J. Principal component analysis[J]. Wiley Interdisciplinary Reviews: Computational Statistics, 2010, 2(4): 433 – 459.
- [32] Ringnér M. What is principal component analysis? [J]. Nature Biotechnology, 2008, 26(3): 303 – 304.
- [33] 陈 超, 陈慧欣, 陈 东, 等. 舟山群岛海岸线遥感信息提取及时空演变分析[J]. 国土资源遥感, 2021, 33(2): 141 – 152. doi: 10.6046/gtzyyg. 2020248.
- Chen C, Chen H X, Chen D, et al. Coastline extraction and spatial – temporal variations using remote sensing technology in Zhoushan Islands[J]. Remote Sensing of Land and Resources, 2021, 33(2): 141 – 152. doi:10.6046/gtzyyg. 2020248.
- [34] 孙伟伟, 杨 刚, 陈 超, 等. 中国地球观测遥感卫星发展现状及文献分析[J]. 遥感学报, 2020, 24(5): 479 – 510.
- Sun W W, Yang G, Chen C, et al. Development status and literature analysis of China’s earth observation remote sensing satellites[J]. Journal of Remote Sensing, 2020, 24(5): 479 – 510.

A random forest – based method integrating indices and principal components for classifying remote sensing images

LIANG Jintao¹, CHEN Chao², ZHANG Zili³, LIU Zhisong⁴

(1. Marine Science and Technology College, Zhejiang Ocean University, Zhoushan 316022, China; 2. School of Geography Science and Geomatics Engineering, Suzhou University of Science and Technology, Suzhou 215009, China; 3. Zhejiang Ecological and Environmental Monitoring Center (Zhejiang Key Laboratory of Ecological and Environmental Monitoring, Forewarning and Quality Control), Hangzhou 310012, China; 4. School of Information Engineering, Zhejiang Ocean University, Zhoushan 316022, China)

Abstract: Accurate information about land use/land cover (LULC) can provide significant guidance for regional

spatial planning and sustainable development. However, conventional methods for remote sensing image classification are challenging due to complex surface morphologies, diverse surface feature types, and nonlinear features of remote sensing images. Therefore, they fail to fully utilize the rich information in remote sensing images. This study developed a random forest – based classification method for remote sensing images to extract LULC information by integrating indices and principal components. First, the images covering the study area were selected to determine cloud cover and conduct median synthesis of images, obtaining interannual remote sensing images. Then, various calculated indices and the extracted principal components were integrated into the band stacks of remote sensing images. Furthermore, classifiers were constructed using different machine – learning algorithms. Finally, based on a confusion matrix, the classification results were evaluated using overall accuracy and the Kappa coefficient. The experimental results of the Hangzhouwan area show that the decision support based on vegetation, water, building indices, and principal components can improve the classification accuracy, yielding overall accuracy and Kappa coefficient of 91.42% and 0.894 2, respectively, which were higher than those of conventional methods such as random forest, classification and regression tree, and support vector machine. The method for remote sensing image classification proposed in this study, which integrates indices and principal components, can obtain high – accuracy land use classification results by accurately extracting land cover features in remote sensing images. This study will provide method support for fine – scale surface classification.

Keywords: random forest; land use/land cover; index; principal component analysis; accuracy evaluation
(责任编辑: 陈 理)