

融入辅助数据集的面向对象土地利用分类研究*

李坤玉, 王雪梅, 李锐, 李顿

新疆师范大学地理科学与旅游学院 / 新疆干旱区湖泊环境与资源实验室, 新疆 乌鲁木齐 830054

摘要: 土地利用分类结果对国土空间的管理至关重要。为提高土地利用分类结果的准确性, 本文以博湖县为研究区, 使用Sentinel-2A影像提取光谱特征, 并结合雷达、光谱指数、土壤和地形特征构建6个面向对象的土地利用分类模型, 使用简单非迭代聚类(SNIC)算法和随机森林(RF)算法对影像进行分割和分类, 得出模型的分类精度以及特征重要性排序, 最后使用分类回归树(CART)算法验证辅助数据集对提高分类精度的影响。结果表明: 使用SNIC算法分割影像时, 分别设置种子大小为17、紧凑度为0时, 该研究区影像分割效果最好。基于RF分类算法, 在只使用光谱信息进行分类时分类精度最低, 加入雷达、光谱指数、土壤和地形特征中任何一个辅助数据集均可提高土地利用的分类精度, 其中地形特征对提高分类精度的效果更显著, 加入所有辅助数据集时分类精度达到最高, OA=92.34%, Kappa系数=0.91。使用CART算法进行分类有效性验证得出, 基于RF算法的分类效果优于CART算法。基于遥感云平台的SNIC分割算法, 融入辅助数据集进行面向对象分类, 为提高土地利用分类精度提供参考。

关键词: 土地利用分类; 辅助数据集; SNIC分割; 面向对象; 随机森林; Sentinel-2A影像

中图分类号: P237 **文献标志码:** A **文章编号:** 2097-0137(2024)01-0034-11

The object-oriented land use classification incorporating auxiliary data sets

LI Kunyu, WANG Xuemei, LI Rui, LI Dun

College of Geographic Science and Tourism, Xinjiang Normal University / Key Laboratory of Lake Environment and Resources in Arid Region of Xinjiang, Urumqi 830054, China

Abstract: Land use classification is critical to the management of land space. To improve the accuracy of land use classification, this study takes Bohu County as the research area, uses Sentinel-2A images to extract spectral features, and combines radar, spectral index, soil, and terrain features to construct six object-oriented land use classification models. We then use a simple non-iterative clustering algorithm and random forest algorithm to segment and classify the images and obtain the classification accuracy and feature importance ranking of the model. In the final step, we use the classification regression tree algorithm to verify the influence of the auxiliary dataset on the improvement of the classification accuracy. The results show that when using the SNIC algorithm to segment the images, with seed size 17 and compactness 0, the image segmentation effect in this study area is the best. The classification accuracy is the lowest when only spectral information is used, and adding any auxiliary dataset of radar, spectral index, soil, and terrain features can improve the classification accuracy of land use. Among those auxiliary datasets, the effect of terrain features on improving classification accuracy is more significant, and the classification accuracy reaches the highest when all auxiliary datasets are added, with

* 收稿日期: 2023-05-19

录用日期: 2023-05-30

网络首发日期: 2023-10-31

基金项目: 新疆维吾尔自治区自然科学基金(2020D01A79); 国家自然科学基金(41561051)

作者简介: 李坤玉(1999年生), 女; 研究方向: 资源环境遥感; E-mail: kunyu_li@163.com

通信作者: 王雪梅(1976年生), 女; 研究方向: 干旱区资源环境遥感技术应用; E-mail: wangxm_1225@sina.com

OA=92.34% and Kappa coefficient=0.91. The classification validity is verified using the categorical regression tree algorithm, it shows that the classification effect based on the random forest algorithm is better than that of the categorical regression tree algorithm. The SNIC segmentation algorithm based on the remote sensing cloud platform is integrated into an auxiliary data set for object-oriented classification, which provide a reference for improving the accuracy of land use classification.

Key words: land use classification; auxiliary datasets; SNIC segmentation; object-oriented; random forest; sentinel-2A image

土地利用信息在国土空间规划、自然资源保护以及气候变化研究等领域起着至关重要的作用, 为可持续发展工作的开展提供信息保障(朱永森等, 2017)。利用遥感技术不仅可以大规模地调查土地利用现状, 又能够监测和获取土地利用变化等信息(舒弥等, 2022)。随着遥感技术的快速发展, 遥感影像数据被广泛地用于土地利用分类的研究。欧空局于2015年发布的Sentinel-2数据具有高时空分辨率以及丰富的光谱信息, 被广泛用于土地利用分类研究中。但由于影像存在同物异谱和同谱异物现象, 仅利用卫星影像的光谱信息进行土地利用分类很难达到较高的精度(陈媛媛等, 2022)。随着大量的辅助数据开放可用, 许多研究试图通过结合卫星数据和辅助数据来提高土地利用分类的准确性(Tomáš et al., 2018)。将辅助特征与遥感数据相结合以提高区域和全球尺度的分类精度已经成为40年来人们关注的话题(Phiri et al., 2017)。李恒凯等(2021)在探讨加入地形特征和雷达特征对南方丘陵山地土地利用分类信息提取的作用时得出, 加入雷达特征和地形特征均有助于提高分类精度, 其中加入地形特征更有助于耕地和园地的提取; 张来红等(2023)对Sentinel-2和Landsat数据从光谱指数特征、纹理特征、地形特征3个方面构建多维特征集进行土地利用分类, 得出融入光谱指数特征、地形特征能够有效提高土地利用分类模型的精度。Hurskainen et al.(2019)在研究免费开放的地理信息数据对面向对象土地利用分类精度的影响时得出, 地形特征、土壤特征可显著提高土地利用分类精度。

虽然辅助特征可以提高土地利用分类的准确性, 但是可用的免费开放的地理空间数据有限, 需要从多个来源请求获得。谷歌地球引擎(GEE, Google Earth Engine)提供了一个强大的地理空间计算分析遥感云平台, 并可以直接调用卫星影像和各种类型的地理空间数据集。使用GEE可进行影

像分割, 其主要支持3种遥感影像分割算法, 包括K-means、G-means和SNIC算法。其中, SNIC算法占用内存小, 运行速度最快, 而且可以通过设置参数来控制分割效果, 该算法已成为国内外影像分割的热点算法(Tassi et al., 2020)。刘通等(2022)、毛丽君等(2021)等使用SNIC分割算法均得到了较好的分割效果, 然后结合RF算法进行分类后均得到了较高分类精度, 有效地提高了土地利用分类结果的准确性。以往许多利用辅助特征提高高分辨率卫星图像分类精度的研究是基于像素的分类上, 而较少关注基于对象的分类(Franklin, 2018)。同时, 国内在研究辅助数据对分类精度的影响中, 研究土壤特征对提高分类精度有效性的实验较少。因此, 本研究基于GEE平台调用Sentinel-2高分辨率影像, 利用SNIC分割算法对影像进行图像分割, 结合雷达特征、光谱指数特征、土壤特征以及地形特征, 使用RF算法进行面向对象的土地利用分类, 得出最优辅助数据集, 然后进行特征重要性评价, 得出贡献率最高的辅助特征, 最后使用CART算法验证结论, 使得研究结果更具可靠性, 为后续更好地进行土地利用分类工作提供技术参考。

1 研究区与数据

1.1 研究区概况

博湖县位于新疆维吾尔自治区中部, 经纬度 $41^{\circ}33' \sim 42^{\circ}14'N$, $86^{\circ}19' \sim 87^{\circ}26'E$, 隶属于新疆巴音郭楞蒙古自治州, 天山南麓, 区域海拔为1 037~2 529 m, 南北两端地势较高, 呈蝶状谷地, 因其境内有中国最大的内陆淡水湖博斯腾湖而得名。该研究区气候温和湿润, 光照充足, 年均气温 $9.3^{\circ}C$, 年降水量为79.6 mm, 年均相对湿度为62%, 有“湿岛效应”, 适合农业以及畜牧业的发展。博斯腾湖分布在该县东北部, 湖泊面积较大, 占博湖县总面积的43.2%, 为当地产业发展提供丰

富的水资源;耕地资源丰富,种植作物主要有小麦、玉米、棉花、辣椒等;分布在博斯腾湖西岸的芦苇区是全国4大苇区之一;区域南部未利用地集中,有少量片状耕地。博湖县土地利用类型多样,具有典型代表性(图1)。

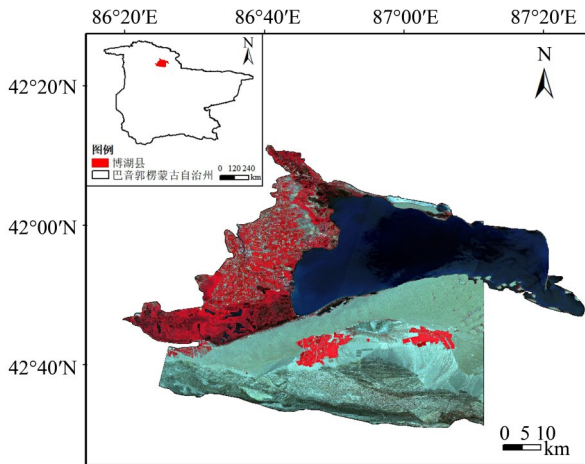


图1 研究区位置示意图及其假彩色影像

Fig. 1 Sketch map for the study area and false color imagery

1.2 数据源及预处理

高分辨率影像对于区域较小的土地利用分类研究是最佳数据源。本文采用多光谱高分辨率影像 Sentinel-2 数据作为主要遥感信息进行分类。通过 GEE 平台加载 2022 年 7 月 20~23 日云层覆盖率小于 7% 的 Sentinel-2 L2A 图像,选择 4 个空间分辨率为 10 m 和 6 个空间分辨率为 20 m 的 10 个波段作为原始影像数据,并对其进行拼接、裁剪以及云掩膜等处理生成博湖县无云影像。本文结合了卫

星数据和辅助数据集来提高土地利用分类的精度和准确性,使用的分类特征包括光谱、雷达、光谱指数、土壤和地形特征(如表 1)。

光谱特征是在 Sentinel-2 影像中选取 10 个波段及其主成分的中值作为分类特征,包括蓝、绿、红 3 个可见光波段(B2-B4)、近红外波段(B8)、3 个红边波段(B5-B7)、Narrow NIR 波段(B8A)以及 2 个短波红外波段(B11、B12)。有研究表明,波段的主成分(PC, principal component)可以提高土地利用的分类精度(Pareeth et al., 2019)。雷达特征是从 Sentinel-1 影像获取的 VV、VH 两个极化波段。光谱指数特征是在 GEE 平台使用 Sentinel-2 影像波段计算得出 8 个指数的中值,包括归一化植被指数(NDVI, normalized difference vegetation index)、归一化水体指数(NDWI, normalized difference water index)、建筑用地指数(ABI, index-based built-up index)、比值植被指数(RVI, ratio vegetation index)、土壤调节植被指数(SAVI, soil-adjusted vegetation index)、差值植被指数(DVI, difference vegetation index)、裸土指数(BSI, bare soil index)和红边位置指数(REP, red edge position index),光谱指数计算公式如表 2 所示。土壤特征是在 GEE 平台调用 OpenLandMap 提供的 250 m 分辨率的表层土壤有机碳含量、土壤含水量、粘土含量、土壤容重、土壤 pH 值 5 个特征,并对其进行重采样。地形特征是利用 GEE 平台调用 30 m 分辨率的 DEM 数据计算得出,包括高程、坡度和坡向数据,该数据集需重采样为 10 m。

表 1 土地利用分类特征

Table 1 The features in the land use classification

项目	特征数量	特征描述	数据来源
光谱	20	B2-B8、B8A、B11、B12 波段及其 10 个主成分的中值	Sentinel-2
雷达	2	VV、VH	Sentinel-1
光谱指数	8	NDVI、NDWI、ABI、RVI、SAVI、DVI、BSI、REP 的中值	Sentinel-2
土壤	5	表层土壤有机碳含量、含水量、粘土含量、容重、pH 值	OpenLandMap
地形	3	高程、坡度、坡向	SRTM

2 研究方法

2.1 分类体系建立及样本点选取

参照国家土地利用分类标准《GB/T21010-2017》,结合 Google Earth 影像,充分了解研究区景观地貌特征,综合考虑研究区地物分布的整体

性以及遥感影像的可分性,将研究区土地利用类型分为 8 种类型:种植耕地、已收获耕地、林地、草地、芦苇湿地、建设用地、水域以及未利用地。通过实地调查以及参考高空间分辨率的 Google Earth 影像和 Sentinel-2 影像,结合欧空局提供的

10 m 分辨率的土地利用分类数据进行目视解译, 采集具有代表性的730个样本点, 其中包括种植耕地132个、已收获耕地83个、林地75个、草地49个、芦苇湿地103个、建设用地100个、水域69个以及未利用地119个, 分别从每个地类中随机选取

70%的样本点作为训练样本, 选取30%的样本点作为验证样本。这8种土地利用类型的分类体系、影像细节特征以及每种地类训练样本和验证样本数量如表3所示。

表2 光谱指数及其计算公式
Table 2 Spectral indices and calculation formula

指数名称	计算公式	参考文献
NDVI	$(B8-B4)/(B8+B4)$	(陈媛媛等, 2022)
NDWI	$(B3-B8)/(B3+B8)$	(胡云峰等, 2018)
IBI	$(B8/(B8+B4))+(B3/(B3+B11))$	(何炳伟等, 2017)
RVI	$B8/B4$	(潘磊等, 2020)
SAVI	$1.5 \times (B8-B4)/(B8+B4+0.5)$	(潘磊等, 2020)
DVI	$B8-B4$	(潘磊等, 2020)
BSI	$((B4+B11)-(B8+B2))/((B4+B11)+(B8+B2))$	(刘通等, 2022)
REP	$705+35 \times (0.5 \times (B7+B4)-B5)/(B6-B5)$	(陈媛媛等, 2022)

表3 土地利用分类体系和影像特征
Table 3 Land-use classification systems and imagery features

类型	假彩色影像	假彩色影像特征	训练样本	验证样本
种植耕地		红色, 边界规则呈矩形	92	40
已收获耕地		灰黄色, 边界规则呈矩形	58	25
林地		深红色, 边界不规则	53	22
草地		浅红色, 边界不规则	34	15
芦苇湿地		暗红或黑红色, 边界不规则, 主要分布于湖泊周围	72	31
建设用地		青色或灰色, 分布规则	70	30
水域		深蓝色, 边界明显	48	21
未利用地		青白或黄白色, 形状不规则	83	36

2.2 SNIC分割算法

简单非迭代聚类(SNIC, simple non-iterative clustering)是基于简单线性迭代聚类(SLIC, simple linear iterative clustering)算法改进的最新的超像素分割算法(Achanta et al., 2017), 改变了SLIC算法占用内存大、运算速度慢的特点, 可以更快地生成超像素(岳巍等, 2022)。SNIC算法是用

户通过设置5个参数控制分割结果, 包括影像、紧凑度、连通性、邻域尺寸和种子大小。参数设置基于重复迭代, 并结合视觉进行评价。影像是指参与分割的影像。紧凑度是指分割后对象的规整程度, 数值越大, 每个对象的形状越规整, 越接近于正方形。连通性是指像元的邻接性, 参数一般设置为4或8, 表示4连通或8连通。邻域尺寸表

示邻域范围,通常设置为256。种子大小是指聚类中心的间隔,数值越大,种子点之间的间隔越大,分割得到的对象越大。其中紧凑度和种子大小对结果的影响比较大,需要根据地物的实际情况进行设置。

2.3 土地利用分类模型

为了比较融入不同辅助数据集对土地利用分类精度以及结果准确性的影响,本文构建了6种融入不同辅助数据集的土地利用分类模型,如表4所

示。在6种分类模型中,M1模型是基于对象的光谱特征分类模型;M2模型除了光谱特征以外,还结合了雷达特征;M3模型则综合了光谱特征和光谱指数特征进行土地利用分类;M4模型是在光谱特征的基础上结合了土壤特征进行影像分类;M5模型则是基于对象的光谱特征与地形特征的分类;M6模型融入了光谱特征、雷达特征、光谱指数特征、土壤特征和地形特征这5种辅助数据集进行研究区土地利用分类。

表4 土地利用分类模型

Table 4 Models used in this study for LULC classification

模型名称	分类模型	辅助数据集
M1	基于对象的光谱特征分类模型	光谱特征
M2	基于对象的光谱特征+雷达特征分类模型	雷达特征
M3	基于对象的光谱特征+光谱指数分类模型	光谱指数特征
M4	基于对象的光谱特征+土壤特征分类模型	土壤特征
M5	基于对象的光谱特征+地形特征分类模型	地形特征
M6	基于对象的全特征分类模型	所有特征

2.4 土地利用分类方法

随机森林(RF, random forest)是Breiman(2001)提出的一种机器学习方法,是一种基于决策树的分类器。随机森林分类器的构建主要涉及两个方面:随机选择样本和随机选择特征。使用Bootstrap重抽样方法从原始样本中有放回地抽取大约2/3的样本用于对当前决策树模型进行训练,抽取剩余的样本作为对随机森林模型进行分类性能评估的袋外数据(OOB, out-of-bag),计算该模型的预测错误率,该错误率被称为袋外误差。基于RF分类器进行分类的基本思想是先使用Bootstrap重抽样方法从原始训练集中抽取 m 个样本,每个样本的容量与原始训练集一致,然后给 m 个样本建立 m 个决策树模型,得出 m 种分类结果,最后通过投票得到最终的分类或预测结果。有大量研究表明,RF算法在土地利用分类中具有较高的分类精度(匡开新等,2023),同时削弱了过拟合现象,有效地提高了分类器的泛化能力。本研究使用RF算法进行土地利用分类研究。

分类回归树(CART, classification and regression tree)是一种以二叉树为逻辑结构,用于完成线性回归任务的决策树,使用预定义的阈值工作。该决策树使用基尼系数划分属性,一个数据集的纯度可用基尼系数来衡量,基尼系数越小,数据

集的纯度越高(刘睿等,2021)。本文使用CART算法来验证结论的通用性和可靠性。

2.5 精度评价

为了评价分类的结果,往往需要使用验证样本进行检验。生产者精度(PA, producer's accuracy)、用户精度(UA, user's accuracy)、总体精度(OA, overall accuracy)和Kappa系数是最常用的检验指标。在面向对象分类中,PA指分类器将整个影像的所有对象正确分为某类的对象数与该类真实参考总数的比率,UA指正确分到某类的对象总数与分类器将整个影像的对象分为该类的对象总数比率,OA指被正确分类的对象总和与总对象数的比率,Kappa系数是用于一致性检验的指标,用来衡量分类的效果,其计算是基于混淆矩阵的(吴静波,2018),Kappa取值在-1~1之间,一般>0,

$$\text{Kappa} = \frac{N \sum_{i=j=1}^n X_{ij} - \sum_{i=j=1}^n (X_{i+} X_{+j})}{N^2 - \sum_{i=j=1}^n X_{i+} X_{+j}}$$

其中 N 为用于精度评价的对象数; X_{ij} 为混淆矩阵中第 i 行第 j 列的对象数; X_{i+} 为第 i 行的总对象数, X_{+j} 为第 j 列的总对象数。

2.6 特征重要性评价

使用随机森林进行特征重要性评估的基本思想是计算每个特征在随机森林中的每棵树中作出的贡献,求其平均值,最后比较特征之间的贡献

大小, 贡献值通常用袋外数据错误率来衡量(吴静波等, 2018)。在随机森林算法中某个特征 X 重要性的计算方法如下: 首先对于随机森林中的每一颗决策树, 使用相应的 OOB 数据来计算它的袋外数据误差, 记为 OOB_{err1} ; 其次随机地对袋外数据 OOB 所有样本的特征 X 加入噪声干扰, 再计算它的袋外数据误差, 记为 OOB_{err2} ; 最后假设随机森林中有 N_{tree} 棵树, 计算特征 X 的重要性

$$X_{重要性} = \frac{\sum(OOB_{err2} - OOB_{err1})}{N_{tree}}$$

若给某个特征随机加入噪声之后, 袋外数据的准确率大幅度降低, 则说明这个特征对于样本的分类结果影响很大, 表明它的重要程度比较高。

3 结果与分析

3.1 分割参数确定

面向对象土地利用分类效果的好坏与影像分割结果有密切关系。SNIC 分割算法需要设置 5 个参数控制图像分割效果, 连通性与邻域尺度分别设置为 8、256。使用控制变量法, 对分割结果影响较大的紧凑度和种子大小选取 6 组参数进

行影像分割实验, 并对分割结果进行目视判读, 确定合适取值。图 2(a-f) 表示使用 SNIC 分割算法对影像进行分割的局部结果, 其中图 2(a-d) 的紧凑度取值为 0, 种子大小取值分别为 10、13、17 和 20, 图 2(e-f) 的种子大小取值为 17, 紧凑度取值分别为 1 和 2。对比图 2(a-b) 可以发现种植耕地、林地、建设用地、水域存在过分割现象; 图 2(c) 建设用地分割效果较好, 种植耕地、林地、水域过分割现象有所减缓, 整体分割效果较好; 图 2(d) 中建设用地存在欠分割现象。将图 2(c) 与图 2(e-f) 对比可以发现, 图 2(e-f) 边界贴合度不高。从图 2(c) 可以看出, 种植耕地、水域仍有少量过分割现象的存在, 过度分割会产生大量的对象, 通常分割对象越多, 地物分割越精细, 分类精度越高, 与此同时需要更多的计算时间, 本文中使用的 GEE 云平台, 它可以为数据处理工作提供强大的数据分析平台, 因此不必考虑计算时间成本。虽然图像分割结果中存在少量过分割现象, 但是对研究目标和结论影响不大。综上所述, 分割效果最好的是图 2(c), 最终选取的图像分割参数为: 种子大小为 17, 紧凑度为 0。

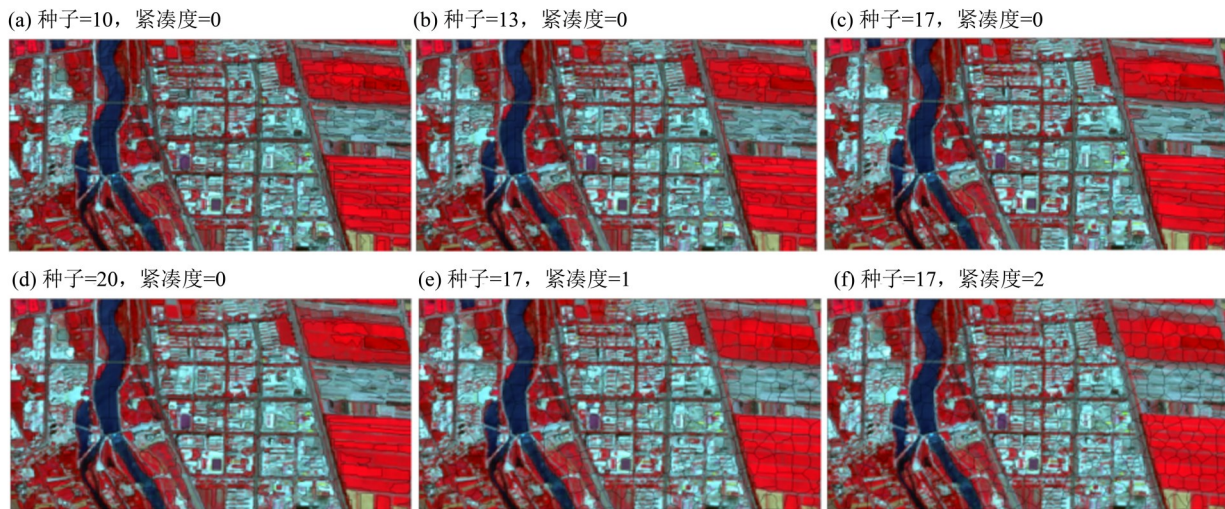


图 2 局部区域不同分割参数的图像分割结果

Fig. 2 Image segmentation results with different segmentation parameters in the local area

3.2 融入不同辅助数据集分类结果比较

基于最优分割参数的选择结果, 融入各类辅助数据集, 进行 2022 年 7 月新疆博湖县土地利用分类, 得到不同模型的土地利用分类结果(图 3)。在 6 个分类模型中, 精度最低的是基于模型 M1 的分类结果, 总体分类精度 $OA=84.82\%$, $Kappa=0.82$, 精度最高的是基于模型 M6 的分类结果, 分

类精度 $OA=92.34\%$, $Kappa=0.91$ 。

由图 3 可知, 模型 M1~M3 的分类效果没有模型 M4~M6 的分类准确性高, 与表 5 的分类精度结果一致。整体上, M1~M3 有明显的错分现象, M1 中将研究区南部少量未利用地错分为已收获耕地和建设用地, 博斯腾湖西岸部分芦苇湿地错分为种植耕地; M2 和 M3 在研究区的南部未利用地均

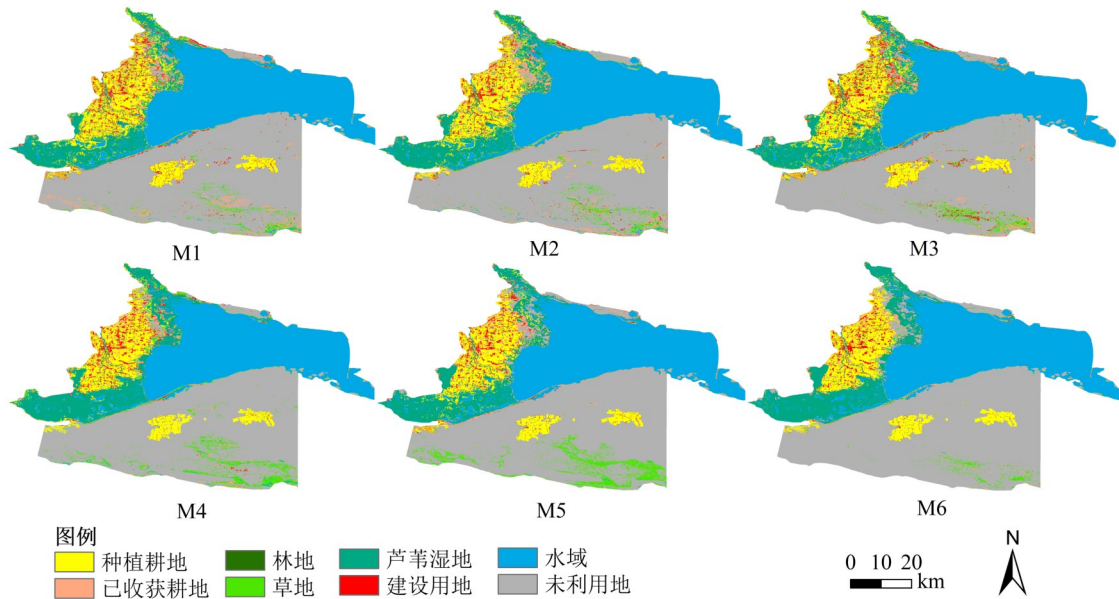


图3 基于6个分类模型的分类结果展示

Fig. 3 Display of classification results based on 6 classification models

表5 不同分类模型的分类精度结果

Table 5 Accuracy results for different classification models

模型		种植耕地	已收获耕地	林地	草地	芦苇湿地	建设用地	水域	未利用地	OA/%	Kappa系数
M1	PA/%	86.96	75.00	81.82	78.57	76.92	82.35	100.00	91.18	84.82	0.82
	UA/%	88.89	72.00	78.26	73.33	86.96	93.33	96.00	81.58		
M2	PA/%	86.05	79.17	90.91	86.36	80.00	86.21	90.00	97.06	87.21	0.85
	UA/%	90.24	82.61	100.00	100.00	95.24	92.59	100.00	66.00		
M3	PA/%	94.29	90.00	85.71	83.33	89.29	84.38	91.67	91.18	89.15	0.87
	UA/%	80.49	75.00	100.00	93.75	100.00	100.00	100.00	79.49		
M4	PA/%	81.58	81.82	92.31	88.24	96.43	96.30	100.00	91.43	90.45	0.89
	UA/%	88.57	81.82	92.31	88.24	87.10	96.30	95.00	94.12		
M5	PA/%	93.75	95.65	80.00	100.00	89.19	90.91	100.00	91.50	91.44	0.90
	UA/%	78.95	75.86	94.12	100.00	97.06	96.77	100.00	97.22		
M6	PA/%	91.18	78.95	100.00	100.00	96.77	82.76	92.00	96.97	92.34	0.91
	UA/%	88.57	83.33	92.31	100.00	100.00	88.89	100.00	88.89		

有错分为建设用地的现象。M4~M6整体分类效果较好,破碎现象较少,与博湖县土地用地整体情况相符。说明仅使用Sentinel-2影像的光谱特征进行土地利用分类时,分类的OA和Kappa系数最低,分别为84.82%和0.82;当加入雷达、光谱指数、土壤和地形特征后,OA均有所提高,加入地形特征后分类精度提高最为明显,OA提高了6.22%,Kappa系数提高了0.08,说明加入这4种辅助数据集中的任意一个特征数据集均有助于提高分类的整体准确性,其中地形特征对提高分类的

整体精度起着最重要的作用。融入所有辅助数据集的模型M6分类精度最高,OA和Kappa系数分别为92.34%和0.91,各类地物均能够得到较好的区分,较其余5个分类模型来说分类效果最佳。

选择所含地物类型较为丰富的3个典型区域,对比分析不同地物类型在不同分类模型中的分类结果(图4)以及结合Sentinel-2影像可以发现,对博湖县进行土地利用分类的误差主要来源于种植耕地、林地、草地与芦苇湿地的误分以及建设用地和未利用地的误分,这主要是由于其光谱特征

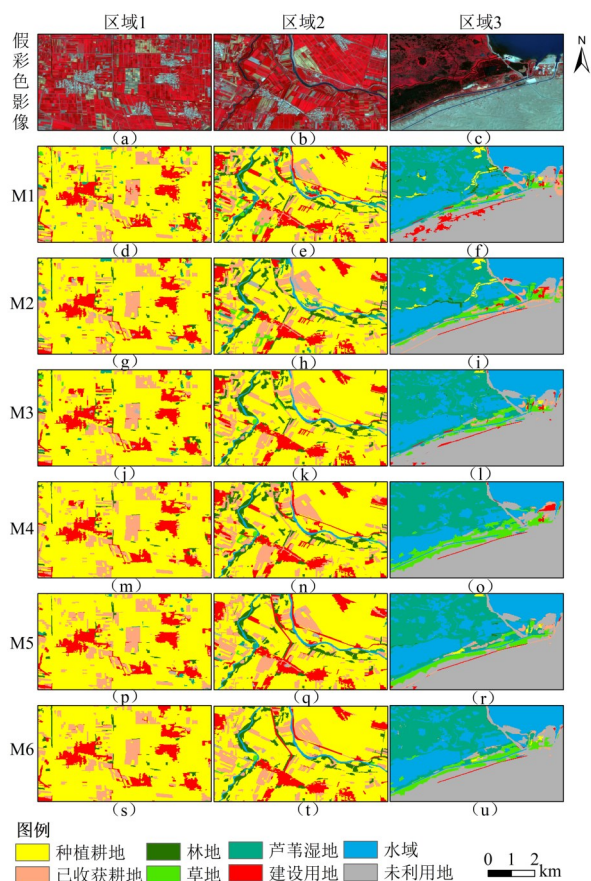


图4 不同分类模型的典型区域分析

Fig. 4 Typical area analysis of different classification models

相似, 会出现异物同谱或同谱异物的现象, 在这种情况下可以借助光谱以外的辅助特征, 有助于丰富分类所需的有用信息, 从而提高土地利用分类的准确性。

基于6种分类模型的8种地类的PA和UA如表5所示。模型M2加入雷达特征后, 林地、草地、芦苇湿地的分类精度均显著提高, 这是因为微波比光波能更深地穿透植物, 对植被散射体的形状、结构较为敏感, 可以更好地区分植被类型。同时也可以看出建设用地和未利用地的分类精度均有所提高, 说明雷达特征也有利于区分建设用地和未利用地。模型M3加入了4个植被指数、1个红边指数、1个水体指数、1个建筑指数和1个裸土指数, 与模型M1相比, 所有植被类型的土地利用分类精度均得到提高, 其中芦苇湿地精度提高最多, PA提高了12.37%, UA提高了13.04%, 说明植被特征以及红边指数特征对于区分不同植被类型起着重要作用。模型M4加入了土壤特征, 由于耕地的土壤特征相似, 所以不同耕地类型不易区分, 因此种植耕地与已收获耕地的分类精度相对

于其他地物来说较低; 其他地物的土壤属性区分较为明显, 所以较于M1分类精度均有所提高。模型M5加入地形特征后, 各地类的分类精度有明显提升, 种植耕地和草地, 以及建设用地和未利用地的区分更为容易, 这与博湖县地形南北高、中间低有关, 草地和未利用地主要分布在南部海拔较高的地区, 种植耕地和建设用地主要分布在中西部平原区。对表5中模型M6与M1分类结果进行比较可以看出, 除水域以外, 每种地物的土地利用分类精度均有较大程度的提高。对比图4中局部区域的分类结果图, 也可以看出模型M6的分类效果相较于其他模型也达到最佳, 消除了种植耕地与芦苇湿地、种植耕地与林地、建设用地和未利用地的误分, 分类细节也与遥感影像更为契合。水域分类效果均较好, 在6个模型中其分类精度均在90%以上。

3.3 特征重要性评价

基于RF算法融入所有辅助数据集的土地利用分类精度最高, 使用基于RF算法的袋外误差来衡量每种特征的重要性值, 图5为本文38个特征的重要性排序。

由图5可以看出, 对分类结果影响最大的是高程特征, 西北地区复杂的地形特征对于土地利用类型的识别至关重要。VH极化特征的重要性居于第2位, 该区植被类型较多, 能更深穿透植物的微波可以更好地识别不同的植被类型。紧接着即为土壤特征包括表层土壤有机碳含量以及表层土壤含水量, 西北干旱区不同地类的土壤特征存在较大差异, 因此依据不同土壤特征区分地物类型是有效的。其次就为波段B10的主成分中值和波段B5的主成分中值, Sentinel-2影像波段重要性最大的是B8波段, 因为近红外波段对叶绿素的反射效果好, 在标准假彩色影像中植被呈现红色, 植被越密集, 红色越深, 植被类型不同其密集度也不同, 可更好地识别该区域不同的植被类型。光谱指数特征重要性最大的是RVI指数, 因为该指数可以更好地反应植被生长的健康状况。坡向特征对土地利用分类的贡献率最低, 说明该区域坡向变化不大。IBI指数的贡献率位于倒数第2位, 因为该区域的建筑用地较少, 而植被、水域和裸地面积较大, 相对于植被光谱指数、水体指数和裸地指数来说, 建筑用地指数对于该区域的土地利用分类结果贡献率较小。由此可看出, 特征重要性

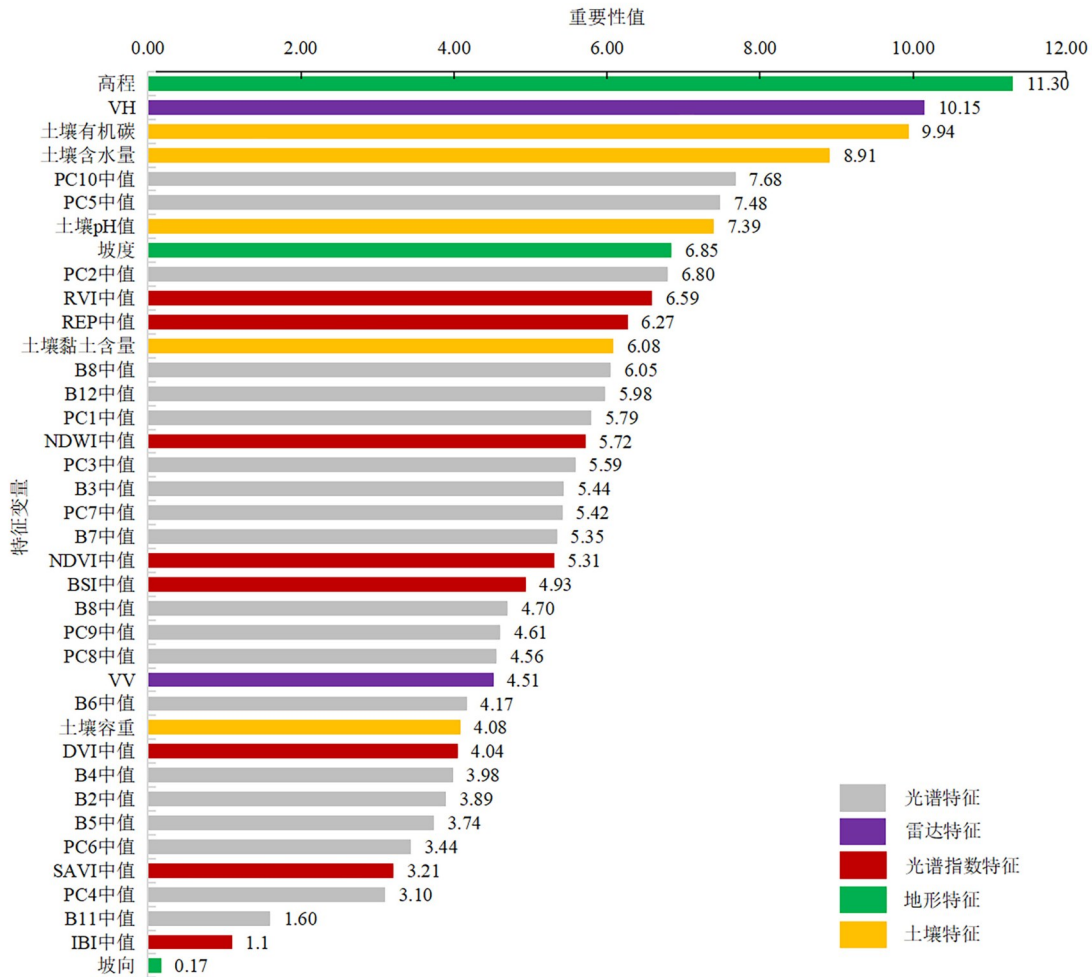


图5 特征重要性评价

Fig. 5 Feature importance evaluation

的大小主要与研究区域的地物分布特征有关。

3.4 分类方法验证

从以上分析可以得出，采用面向对象的RF算法进行土地利用分类时，将光谱数据与辅助数据结合可以有效地提高土地利用分类精度，结合雷达、光谱指数、土壤、地形特征分类精度依次提高，当融入所有辅助数据集时分类精度达到峰值。使用CART算法验证结论的通用性和可靠性。在对比实验中，基于CART的分类模型所使用的特征与基于RF的分类模型相同。基于RF和CART算法分类的OA和Kappa系数如图6所示，图6(a)为基于RF和CART算法分类的OA，图6(b)为基于RF和CART算法分类的Kappa系数。基于CART算法的分类结果表明，随着辅助特征数量的增加，OA和Kappa系数均有所提高，融入雷达、光谱指数、土壤、地形以及所有辅助特征的分类精度依次提高，与基于RF算法的分类结果规律一致，通过对比基

于两种算法的分类精度可以看出，基于RF算法的分类精度更高。

4 讨论

面向对象的最优辅助数据集的选取可提高分类精度的同时避免了基于像素分类时出现的“盐胡椒”现象，使分类结果不仅具有较高的分类精度，视觉上也更美观，强化了分类结果中每个地物的整体性，为下一步开展基于土地利用分类结果的应用型研究提供更好的分类效果。研究基于遥感云平台对Sentinel-2影像使用SNIC分割算法进行影像分割，并结合RF分类器进行土地利用分类，OA>84%，Kappa>0.82，说明分类精度均较高。Djerriri et al.(2020)证明了基于SNIC聚类结合RF分类器对Sentinel-2图像进行分类的面向对象分类方法具有较大优势，可以得到较好分类效果，说明选用的分类方法可靠性较高。分类时使用光

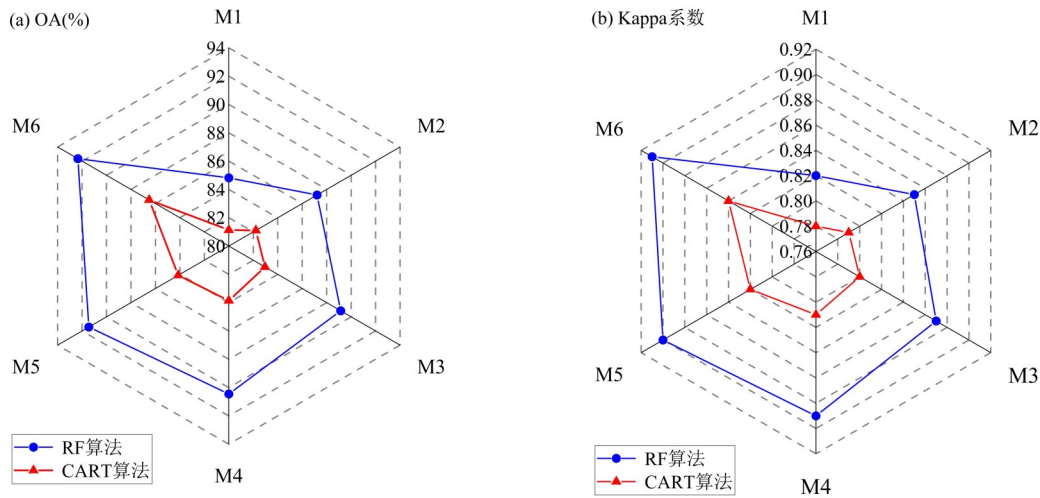


图6 不同分类模型基于RF、CART算法分类的分类精度比较

Fig. 6 Comparison of classification accuracy of different classification models based on RF and CART algorithms

谱信息并融入雷达、光谱指数、土壤和地形特征均可以提高分类精度, 其中地形特征对提高分类精度效果最显著。Qu et al.(2021)结合辅助数据集进行土地利用分类, 得出在面向对象分类中地形特征对提高分类精度的影响最大, 本文也得出该结论。除了地形特征, 土壤特征对提高土地利用分类效果也有较大贡献, 国内融入土壤特征提高土地利用分类精度的研究较少, 而国外较多且认为土壤特征有利于提高分类精度(Ouma et al., 2022), 因此研究中加入了土壤特征, 探究其对提高土地利用分类精度的影响, 与已有研究结论一致。研究中使用的辅助数据集表现出在景观异质性地区提高土地利用分类精度的重要潜力, 下一步工作考虑研究探索更多的辅助数据集对分类结果的影响, 如纹理特征、物候特征、水源距离等, 并在不同的地形条件下研究辅助特征最优组合的异同。

5 结 论

本文借助GEE平台使用SNIC分割算法, 融入辅助数据集构建了基于RF算法的面向对象土地利用分类的6个分类模型, 对博湖县的土地利用进行分类研究, 得出以下结论:

1) 使用GEE平台的SNIC分割算法, 通过控

制变量法、参数迭代设置以及目视判读法得出最优分割参数, 对分割结果影响较大的两个参数分别设置为种子大小为17、紧凑度为0, 该分割参数对该研究区影像分割效果最好, 有利于进行下一步的分类研究。

2) 辅助特征包括雷达、光谱指数、土壤、地形均可以提高土地利用分类精度, 将Sentinel-2影像的波段特征数据与这些辅助数据相结合进行土地利用分类, 有效地提高了土地利用分类结果的准确性, 当融入所有辅助特征时分类精度达到最高, OA=92.34%, Kappa=0.91, 比仅使用光谱特征进行分类的结果OA提高了7.52%, Kappa系数提高了0.09。当只使用一种辅助数据集时, 地形特征对提高土地利用分类精度的效果最好, 土壤特征也起着重要作用。对所有辅助特征进行重要性评价时, 高程特征对提高分类精度的贡献最大。

3) 使用RF算法进行实验, 并使用CART算法验证实验结论, 均可得出融入雷达、光谱指数、土壤、地形特征都可以提高土地利用分类精度, 并且分类精度依次提高, 融入所有辅助数据集时分类精度达到最高。通过对比基于两种算法的分类精度可以看出, 基于RF算法的分类精度更高, 为基于土地利用分类结果进行应用型研究提供高精度的分类参考方法。

参考文献:

- 陈媛媛,雷鸣,王泽远,等,2022.基于 Sentinel 卫星影像的土地利用类型提取——以丽水市莲都区为例[J].森林工程,38(2):54-61.
- 胡云锋,商令杰,王召海,等,2018.GEE 平台和 CART 方法的北京市土地解译[J].测绘科学,43(4):87-93.
- 何炳伟,赵伟,李爱农,等,2017.基于 Landsat 8 遥感影像的新旧城区热环境特征对比研究——以成都市为例[J].遥感技术与应用,32(6):1141-1150.
- 匡开新,杨英宝,高永年,等,2023.协同 Sentinel 合成孔径雷达和光学影像多特征的不透水面随机森林提取方法[J].遥感技术与应用,38(2):422-431.
- 李恒凯,王利娟,肖松松,等,2021.基于多源数据的南方丘陵山地土地利用随机森林分类[J].农业工程学报,37(7):244-251.
- 刘通,任鸿瑞,2022.GEE 平台下利用物候特征进行面向对象的水稻种植分布提取[J].农业工程学报,38(12):189-196.
- 刘睿,王志勇,高瑞,2021.时序 SAR 影像的干旱地区土地利用分类应用[J].测绘科学,46(10):90-97.
- 毛丽君,李明诗,2023.GEE 环境下联合 Sentinel 主被动遥感数据的国家公园土地覆盖分类[J].武汉大学学报(信息科学版),48(5):756-764.
- 潘磊,孙玉军,王轶夫,等,2020.基于 Sentinel-1 和 Sentinel-2 数据的杉木林地上生物量估算[J].南京林业大学学报(自然科学版),44(3):149-156.
- 舒弥,杜世宏,2022.国土调查遥感 40 年进展与挑战[J].地球信息科学学报,24(4):597-616.
- 吴静波,2018.盐碱地信息提取和变化检测方法应用与比较[D].银川:宁夏大学.
- 吴静波,汪西原,2018.宁夏石嘴山地区盐碱地变化的目标级检测[J].陕西师范大学学报(自然科学版),46(2):104-109.
- 岳巍,李世明,李增元,等,2022.基于多时相 Sentinel-2 影像和 SNIC 分割算法的优势树种识别[J].林业科学,58(9):60-69.
- 张来红,秦婷婷,泽仁卓格,等,2023.基于 GEE 和多维特征集的锡林浩特露天矿区近 30a 土地利用分类[J].金属矿山,(3):234-241.
- 朱永森,曾永年,张猛,2017.基于 HJ 卫星数据与面向对象分类的土地利用/覆盖信息提取[J].农业工程学报,33(14):258-265.
- ACHANTA R, SUSSTRUNK S, 2017. Superpixels and polygons using simple non-iterative clustering [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Honolulu:4651-4660.
- BREIMAN L, 2001. Random forests [J]. Mach Learn, 45(1):5-32.
- DJERRIRI K, SAFIA A, ADJOUJ R, 2020. Object-based classification of sentinel-2 imagery using compact texture unit descriptors through google earth engine [C]//2020 Mediterranean and Middle-East geoscience and remote sensing symposium (M2GARSS). Tunisia:105-108.
- FRANKLIN S E, 2018. Pixel- and object-based multispectral classification of forest tree species from small unmanned aerial vehicles [J]. J Unmanned Veh Sys, 6(8):195-211.
- HURSKAINEN P, ADHIKARI H, SILJANDER M, et al, 2019. Auxiliary datasets improve accuracy of object-based land use/land cover classification in heterogeneous savanna landscapes [J]. Remote Sens Environ, 233:111354.
- OUMA Y, NKWAE B, MOALAFHI D, et al, 2022. Comparison of machine learning classifiers for multitemporal and multisensor mapping of urban lulc features [J]. Int Arch Photogramm Remote Sens Spatial Inf Sci, 43(B3):681-689.
- PAREETH S, KARIMI P, SHAFIEI M, et al, 2019. Mapping agricultural landuse patterns from time series of Landsat 8 using random forest based hierarchial approach. [J]. Remote Sens, 11(5):601.
- PHIRI D, MORGENROTH J, 2017. Developments in Landsat land cover classification methods: A review [J]. Remote Sens, 9(9):967.
- QU L, CHEN Z, LI M, et al, 2021. Accuracy improvements to pixel-based and object-based LULC classification with auxiliary datasets from Google Earth engine [J]. Remote Sens, 13(3):453.
- TASSI A, VIZZARI M, 2020. Object-oriented LULC classification in google earth engine combining SNIC, GLCM, and machine learning algorithms [J]. Remote Sens, 12(22):3776.
- TOMÁŠ K, DAVID M, JAN K, et al, 2018. Selecting appropriate variables for detecting grassland to cropland changes using high resolution satellite data [J]. Peer J, 6: e5487.

(责任编辑 秦社彩)