

文章编号: 2095-4298(2023)01-0017-05

基于 GEE 的徐州市土地利用分类研究

郭羽羽, 胡召玲*

(江苏师范大学 地理测绘与城乡规划学院, 江苏 徐州 221116)

摘要: 及时、准确地获取土地利用信息, 可为城市发展和生态环境保护提供参考依据。基于谷歌地球引擎(Google earth engine)云平台, 联合 Sentinel-1 SAR 数据、Sentinel-2 MSI 高分辨率光学影像数据、SRTM 高程数据等多源数据构建分类特征集, 利用随机森林算法对徐州市 2021 年的土地利用类型进行分类, 并对分类结果进行精度评价。研究结果表明: 1) 地物的光谱特征, 尤其是归一化水体指数(NDWI)对分类结果贡献较大, 多特征集的分类精度明显高于单一光谱特征; 2) 综合利用地物光谱特征、纹理特征、地形特征和雷达后向散射特征进行随机森林分类, 分类精度最高, 达 93.55%, 水体和耕地的分类效果明显优于裸地和建设用地; 3) 徐州市的主要土地利用类型为耕地和建设用地, 两者面积占比达 92.44%, 水体主要分布于铜山区和新沂市, 林草地和裸地分布较少。

关键词: 谷歌地球引擎(GEE); 多特征; 随机森林算法; 土地利用

中图分类号: K901 **文献标志码:** A **doi:** 10.3969/j.issn.2095-4298.2023.01.004

Land use classification of Xuzhou based on GEE

Guo Yuyu, Hu Zhaoling*

(School of Geography, Geomatics & Planning, Jiangsu Normal University, Xuzhou 221116, Jiangsu, China)

Abstract: Land use information acquired timely and accurately can provide reference for the city's sustainable development and ecological environment protection. Based on the platform of Google earth engine(GEE), Sentinel-1 SAR, Sentinel-2 MSI and SRTM were combined to construct the classification feature set. The random forest algorithm was used to classify the land use in Xuzhou in 2021, and the accuracy of the classification results was evaluated. The results showed that the spectral features, especially the NDWI contributed significantly to the classification results, and the accuracy of multi-feature sets was higher than that of single spectral features. Spectral feature, texture feature, terrain feature and radar backscattering feature were used for random forest classification, with the highest accuracy reaching 93.55%, the classification accuracy of water body and cultivated land was significantly better than that of bare land and construction land. The main land use in Xuzhou was cultivated land and construction land, accounting for 92.44% of the total area, the water body was mainly distributed in Tongshan district and Xinyi city, with less forest and grass land and bare land.

Key words: Google earth engine(GEE); multi-feature; random forest algorithm; land use

土地利用/土地覆被变化(Land use coverage change, LUCC)是地球表面各地物类型及其自然属性和特征的综合体^[1]。作为人类社会活动与自然生态过程交互和链接的纽带, 土地利用可直接反映地球生态对人类活动的影响^[2-3]。准确、实时的土地覆被信息是自然资源评价的重要信息来源, 也是土地管理和环境监测等多学科应用的基础^[4]。遥感数据基于快速、大尺度、长时序、获取方便等优势, 已被广泛应用于土地利用分类研究中。

谷歌地球引擎(Google earth engine, GEE)云

平台是 Google 公司提供的对大量全球尺度地球科学资料进行在线可视化计算分析处理的平台^[5-6]。相对于 ENVI 等传统的影像处理工具, GEE 可以快速、批量处理大尺度的影像, 进行地表水体变化监测^[7]、植被指数分析^[8-9]、全球土地利用覆盖制图^[10]、农作物估产^[11]、全球森林变化监测^[12]等多个领域的研究。已有研究表明, 在长时间序列、大范围的遥感监测研究中, GEE 平台借助云端计算能极大地缩短影像处理时间, 提高工作效率^[13-14]。Sentinel 系列卫星是欧洲空间局(ESA)“哥白尼计划”中的一

收稿日期: 2022-05-19

基金项目: 国家自然科学基金资助项目(52074133), 江苏师范大学研究生科研与实践创新计划项目(2022XK T0074), 江苏高校优势学科建设工程项目

作者简介: 郭羽羽, 女, 硕士研究生, 主要从事遥感信息提取的研究, E-mail: guoyuyu@163.com.

* 通信作者: 胡召玲, 女, 教授, 博士, 主要从事遥感信息分析与应用的研究, E-mail: huzhaoling@jsnu.edu.cn.

项重大工程,该系列卫星包括雷达遥感卫星、光学遥感卫星、气象遥感卫星等,可为全球大尺度、长时间序列的陆地资源观测提供高分数据^[15]。

遥感影像分类是土地利用/土地覆被分类及信息提取的有效方法^[16-17]。近年来,机器学习方法因具有自动学习、信息处理快速的特点,已被成功应用于遥感影像分类领域^[18],其中随机森林^[19]、支持向量机^[20]、神经网络法^[21]等智能方法应用较为成熟,取得了良好的分类效果。随机森林算法具有精度高、速度快等特征,不仅能估算特征变量的重要性,还可以有效处理缺失数据,是一种应用广泛的遥感影像分类方法^[22-23]。目前,学者们利用随机森林算法对土地利用进行了分类研究^[23-25],结果表明,随机森林算法具有较高的分类精度。

因此,本研究基于 GEE 遥感云平台,利用 Sentinel-1 SAR 数据、Sentinel-2 MSI 数据、SRTM 高程数据等多源数据,采用随机森林算法研究徐州市土地利用分类,以期对徐州市生态监测和城市规划提供参考。

表 1 多源数据来源

Tab. 1 The source of multi-source data

数据	年份	空间分辨率/m	数据来源	投影
Sentinel-1 SAR	2021	10	COPERNICUS/S1-GRD	
Sentinel-2 MSI	2021	10, 20, 60	COPERNICUS/S2-HARMONIZED	GCS-WGS-1984
SRTM	—	30	USGS/SRTMGL1-003	

数据预处理包括云量筛选、裁剪、拼接镶嵌等操作。由于大尺度范围内的遥感影像难以保证云量较小(云量小于 20%),使用 GEE 平台的 QA60 产品进行筛选,因此,所有遥感数据投影统一为 GCS-WGS-1984 坐标系统。徐州市范围矢量数据来源于中国科学院资源环境科学数据中心(<http://www.resdc.cn/>),可用于提取研究区范围影像,实现多景影像的合成与裁剪。

2 研究方法

2.1 样本选取

参考《土地利用现状分类》的一级分类标准,并结合徐州市土地覆被现状,将徐州市用地划分为耕地、林草地、水体、建设用地和裸地 5 类土地利用类型。利用高分辨率遥感影像数据进行目视解译选取样本数据,在 5 类土地利用类型中基于随机、均匀的原则,分别选取 179、153、162、180、114 个样本点,选取时尽量对各地物类别进行覆盖,特征地物尽量全面。样本数据分为训练样本和验证样本,其中 70% 的训练样本用于分类器训练,30% 的验证样本用于分类结果的精度验证。

1 研究区概况及数据源

1.1 研究区概况

徐州市地处 116°22' ~ 118°40' E, 33°43' ~ 34°58' N, 位于江苏省西北部,是淮海经济区中心城市。地形以平原为主(平原占全市面积的 90%),海拔较低(30~50 m),地势平坦,整体呈现出西北高、东南低的趋势。徐州市下辖 5 个市辖区(泉山区、鼓楼区、云龙区、贾汪区、铜山区)、3 个县(睢宁县、沛县、丰县)和 2 个县级市(邳州市、新沂市)。区域内气候资源优越,交通发达,是重要的铁路、公路及航空运输枢纽^[26-27]。

1.2 数据源

基于 GEE 云平台进行多源数据的选取,各数据源具体信息见表 1。遥感影像数据为 Sentinel-1 SAR 和 Sentinel-2 MSI Level-1C,均已通过大气校正、辐射校正、正射校正等进行预处理;高程数据由美国航天飞机雷达地形测绘任务(shuttle radar topography mission, SRTM)提供。

2.2 土地利用分类

借助 GEE 云平台编写代码,实时提取遥感影像,并对徐州市土地利用类型进行分类,在线提取土地利用信息,以解决传统遥感影像土地利用信息提取和分类流程耗时费力等问题。参考文献^[14],采用随机森林分类法,具体分类特征集信息见表 2。Sentinel-2 MSI 数据的光谱特征包括波段信息和遥感指数信息;纹理特征主要是基于 NDVI 指数构建灰度共生矩阵^[25,28],用于消除同物异谱的影响;地形特征基于 SRTM 数据获取常用地形因子,以减小地形阴影的影响;后向散射特征基于 Sentinel-1 SAR 数据获取。选取 20 个特征变量进行不同特征的分析,对比多特征组合的分类精度。由于随机森林算法是一种包含多个决策树的分类模型,决策树数量过多会降低模型的效率,过少会降低模型的精度,因此,综合考虑以兼顾模型运行的精度和效率^[23],经过反复试验,最终选择 45 作为决策树个数。

3 徐州市土地利用分类结果与评价

3.1 不同特征组合下的土地利用分类精度

为了比较不同特征对土地利用分类的贡献,基

于 explain()函数得到不同特征的重要性(表 2),可以看出,波段 B_8 、 B_{11} 、 B_{12} 、 B_2 ,遥感指数 NDWI、ND-BI、EVI、NDVI、高程以及后向散射系数 σ_{VH} 对分类的贡献较大,且主要集中于地物光谱特征. NDWI 在土地利用分类中的作用最为突出,与李明洁等^[25]

的结论一致. 原因可能是,徐州市耕地与林草地特征相似,植被指数对这两种土地利用的变化不敏感,而水域面积较大且在光谱特征的变化上更具有区分度,因此,本研究中 NDWI 对土地利用分类的贡献度明显高于其他特征.

表 2 随机森林分类特征集信息及其重要性

Tab. 2 Random forest classification feature set information and their importance

特征类型	特征描述	简称	重要性
光谱特征	蓝波段	B_2	6.09
	绿波段	B_3	4.43
	红波段	B_4	3.60
	近红外波段	B_8	6.87
	短波红外波段	B_{11}	6.80
	短波红外波段	B_{12}	6.12
	归一化植被指数	NDVI	4.82
	增强型植被指数	EVI	5.32
	归一化建筑指数	NDBI	6.81
	归一化水体指数	NDWI	8.60
地形特征	改进的归一化水体指数	MNDWI	3.56
	高程	Elevation	7.85
	坡度	Slope	2.88
	坡向	Aspect	2.64
纹理特征	角二阶矩	Asm	4.17
	相关度	Idm	2.79
	对比度	Contrast	3.00
	熵值	Entropy	3.81
后向散射特征	垂直发射和水平接收后向散射系数	σ_{VH}	6.08
	垂直发射和垂直接收后向散射系数	σ_{VV}	3.77

由表 3 可以看出,仅使用光谱特征进行分类,分类精度相对较低,分别加入纹理特征和地形特征,分类精度分别提高了 5.75%和 5.14%,Kappa 系数增加了 0.07 和 0.06. 光谱、纹理和地形特征组合的分类精度相较于单一光谱特征增加了 6.23%,再加入雷达后向散射特征,精度增加了 7.19%,Kappa 系数增加了 0.10. 可见,综合利用表 2 中所列出的 20 个特征变量构建多维特征集进行分类,分类精度最高. 徐州市土地利用分类结果见图 1.

表 3 不同特征组合的分类精度

Tab. 3 Classification accuracy of different feature combinations

特征组合	分类精度/%	Kappa 系数
光谱特征	86.36	0.81
光谱+纹理特征	92.11	0.88
光谱+地形特征	91.50	0.87
光谱+纹理+地形特征	92.59	0.90
光谱+纹理+地形+后向散射特征	93.55	0.91

3.2 土地利用类型空间分布与面积

由图 1 可以看出,丰县和睢宁县的城市发展中心位于中东部,沛县、铜山区、邳州市的城市发展中心位于东北部,新沂市的城市发展中心位于北部,泉

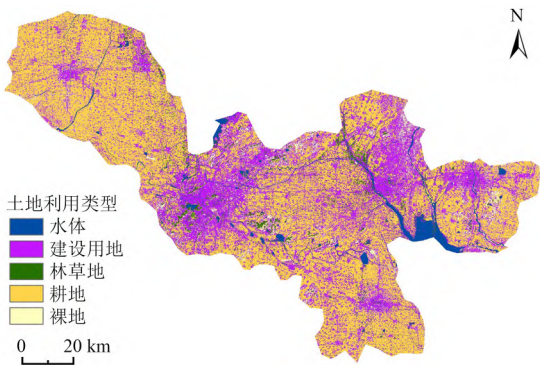


图 1 徐州市土地利用分类结果

Fig. 1 Classification results of land use in Xuzhou

山区、鼓楼区、云龙区、贾汪区的建设用地则围绕区中心向外围扩散. 徐州市以耕地为主,在非主城区,多呈现出村庄与农田毗邻分布的特点. 水体主要以河流、河道、水库、小型湖泊为主. 林地主要集中于云龙湖风景区、吕梁风景区、大洞山景区等的丘陵山体上,草地较少,裸地多分布于林地或城镇周围,城镇周围多为建设中或未完全开发的建设用地.

根据 ArcGIS 得到徐州市各县市区不同土地利用类型的面积(表 4),可以看出,徐州市耕地面积占 63.70%,建设用地面积占 28.74%. 除主城区外,其

表 4 徐州市各土地利用类型的面积
Tab. 4 Land use type area of Xuzhou

区域	面积/km ²					
	林草地	耕地	水体	建设用地	裸地	合计
丰县	3.34	1 153.91	26.68	292.03	4.82	1 480.78
沛县	3.17	895.86	17.99	227.15	3.45	1 147.62
邳州市	9.87	1 235.50	72.92	649.47	36.90	2 004.66
铜山区	107.37	1 594.30	121.44	1 015.27	73.35	2 911.73
新沂市	10.81	925.72	196.41	468.47	25.81	1 627.22
徐州市主城区	15.22	19.19	14.40	139.42	1.69	189.92
睢宁县	9.53	1 301.30	58.70	422.99	31.66	1 824.18
总计	159.31	7 125.78	508.54	3 214.80	177.68	11 186.11

余各县市区的主要土地利用类型均为耕地,其次为建设用地,说明徐州市城镇化发展迅速,但农业仍是不可或缺的产业类型.徐州市域内河流、湖泊、水库众多,水体主要分布于铜山区和新沂市,其中铜山区主要有奎河、吕梁湖等,新沂市主要有沂河、沐河、高塘水库、京杭运河、骆马湖等.徐州市森林资源较为匮乏,多分布于山体附近和城市尚未开发的土地中.

3.3 土地利用分类结果评价

从图 1 中可以看出,徐州市大部分土地利用类型分类正确,特别是采用 10 m 分辨率的 Sentinel-2 MSI 高分影像,细小水体的识别提取更为准确,较少出现不连续的情况.但由于地物的“异物同谱”和“同物异谱”现象,仍有部分地物无法被准确分类,存在错分和误分的情况,如耕地中存在的白色大棚极易与建设用地混淆.为评价分类的结果,结合验证样本计算出徐州市土地利用类型的混淆矩阵(表 5),可以看出,水体和耕地的分类效果较好,生产者精度较高,表明水体和耕地地物分类基本准确,仅存在将小部分其他地物错分为水体和耕地的情况;裸地和建设用地的分类效果略差,比较容易与其他地物混淆,其中部分建设用地会被分类为耕地、水体和裸地;林草地和裸地的用户精度较高,说明其他土地利用类型较少被错分为林草地和裸地.

表 5 土地利用类型分类混淆矩阵

Tab. 5 The confusion matrix of land use type classification

土地利用类型	像元个数						用户精度/%
	林草地	耕地	水体	建设用地	裸地	合计	
林草地	23	0	0	0	0	23	100.00
耕地	2	92	0	6	0	100	92.00
水体	0	2	29	2	0	33	87.88
建设用地	0	2	2	65	2	71	91.55
裸地	0	0	0	0	9	9	100.00
总计	25	96	31	73	11	236	
生产者精度/%	92.00	95.83	93.55	89.04	81.82		

4 讨论与结论

4.1 讨论

随着卫星遥感技术的飞速发展,光学遥感影像由于具有高时空分辨率以及丰富的光谱信息等优势,被广泛应用于各类研究中,而多源遥感数据综合了不同数据源的优势,其使用更加广泛.在多源数据融合基础上构建的随机森林特征集具有更丰富的信息,包括光谱信息、纹理信息、地形信息等,而雷达后向散射特征的加入有助于消除云、雨等天气对遥感影像光谱信息的干扰.因此,综合光学遥感数据、雷达数据、地形数据构建多维分类特征集,丰富的特征信息以及更高的空间分辨率使土地利用分类效果得到优化.另外,不同的特征之间存在一定的重复或信息冗余,通过特征重要性排序及特征组合优选可得到最终的多维特征集,其中光谱特征和雷达特征对分类结果的贡献相对较大,后续应在此基础上对不同特征重要性存在差别的原因进行进一步研究,以减少特征冗余,优化特征组合,从而提高分类精度.

4.2 结论

基于 GEE 平台,联合 Sentinel-1 SAR、Sentinel-2 MSI 遥感影像数据和 SRTM 高程数据,构建地物的光谱特征、纹理特征、地形特征和后向散射特征 4 种分类特征,利用随机森林算法对 2021 年徐州市土地利用类型进行分类研究及精度评价,结果表明:

1)地物的光谱特征对分类结果的影响较大,其中 NDWI 指数对分类结果贡献最大;相较于单一光谱特征,随着纹理特征、地形特征以及雷达后向散射特征的加入,分类精度大幅提高.

2)综合利用地物光谱特征、纹理特征、地形特征和雷达后向散射特征,构建多维特征集进行随机森林分类,分类精度最高,总体精度达 93.55%,Kappa 系数为 0.91.其中水体和耕地的分类效果明显优于裸地和建设用地.

3)徐州市主要土地利用类型为耕地和建设用地,两者占比达到了徐州市总面积的 92.44%,水体集中分布于铜山区和新沂市,林草地和裸地面积占比较少。从建设用地的空间分布可看出各县市区的发展 centers。

参考文献:

- [1] 张磊,吴炳方,李晓松,等. 基于碳收支的中国土地覆被分类系统[J]. 生态学报,2014,34(24):7158.
- [2] 刘纪远,匡文慧,张增祥,等. 20 世纪 80 年代末以来中国土地利用变化的基本特征与空间格局[J]. 地理学报,2014,69(1):3.
- [3] 单勇兵,马晓冬,仇方道. 睢宁县乡村土地利用的格局分析[J]. 江苏师范大学学报(自然科学版),2015,33(1):40.
- [4] 牟晓莉,李贺,黄翀,等. Google Earth Engine 在土地覆被遥感信息提取中的研究进展[J]. 国土资源遥感,2021,33(2):1.
- [5] 戴晓琴,沈瑞淇,王俊,等. GEE 遥感云平台支撑下的河南省土地利用变化检测[J]. 测绘科学技术学报,2021,38(3):287.
- [6] Mutanga O, Kumar L. Google earth engine applications [J]. Remote Sens,2019,11(5):591.
- [7] Xia H,Zhao J Y,Qin Y C,et al. Changes in water surface area during 1989—2017 in the Huai River Basin using landsat data and google earth engine[J]. Remote Sens,2019,11(15):1824.
- [8] 裴杰,牛铮,王力,等. 基于 Google Earth Engine 云平台的植被覆盖度变化长时间序列遥感监测[J]. 中国岩溶,2018,37(4):608.
- [9] 李晶,闫星光,闫萧萧,等. 基于 GEE 云平台的黄河流域植被覆盖度时空变化特征[J]. 煤炭学报,2021,46(5):1439.
- [10] Gong P,Liu H,Zhang M N,et al. Stable classification with limited sample: transferring a 30-m resolution sample set collected in 2015 to mapping 10-m resolution global land cover in 2017[J]. Sci Bull (Beijing),2019,64(6):370.
- [11] Thorp K R,Drajat D. Deep machine learning with Sentinel satellite data to map paddy rice production stages across West Java, Indonesia[J]. Remote Sens Environ,2021,265:112679.
- [12] 邵亚奎,王蕾,朱长明,等. GEE 云平台支持下的西天山森林遥感监测与时空变化分析[J]. 测绘通报,2020(8):13.
- [13] 黄珏,李正茂,张珂,等. 基于 GEE 的中国湖泊浮游植物生物量时空动态分析[J]. 地理学报,2021,76(7):1693.
- [14] 杨丽萍,张静,贡恩军,等. GEE 联合多源数据的西安市土地利用时空格局及驱动力分析[J]. 农业工程学报,2022,38(2):279.
- [15] 杨斌,李丹,王磊,等. 基于 Sentinel-2A 岷江上游地表生物量反演与植被特征分析[J]. 科技导报,2017,35(21):74.
- [16] 范树平,程从坤,刘友兆,等. 中国土地利用/土地覆盖研究综述与展望[J]. 地域研究与开发,2017,36(2):94.
- [17] 杜培军,夏俊士,薛朝辉,等. 高光谱遥感影像分类研究进展[J]. 遥感学报,2016,20(2):236.
- [18] 谷晓天,高小红,马慧娟,等. 复杂地形区土地利用/土地覆被分类机器学习方法比较研究[J]. 遥感技术与应用,2019,34(1):57.
- [19] Zhang X T,Xu J,Chen Y Y,et al. Coastal wetland classification with GF-3 polarimetric SAR imagery by using object-oriented random forest algorithm [J]. Sensors (Basel),2021,21(10):3395.
- [20] Rujoiu-Mare M R,Olariu B,Mihai B A,et al. Land cover classification in Romanian Carpathians and Subcarpathians using multi-date Sentinel-2 remote sensing imagery[J]. Eur J Remote Sens,2017,50(1):496.
- [21] Rajesh S,Nisia T G,Arivazhagan S,et al. Land cover/land use mapping of LISS IV imagery using object-based convolutional neural network with deep features [J]. J Indian Soc Remote Sens,2020,48(1):145.
- [22] Youssef A M,Pourghasemi H R. Landslide susceptibility mapping using machine learning algorithms and comparison of their performance at Abha Basin, Asir Region, Saudi Arabia [J]. Geosci Front,2021,12(2):639.
- [23] 何云,黄翀,李贺,等. 基于 Sentinel-2A 影像特征优选的随机森林土地覆盖分类[J]. 资源科学,2019,41(5):992.
- [24] 陶莉,胡召玲. 利用 Sentinel-2A 数据提取长江中下游丘陵地带农作物种植信息 [J]. 测绘通报,2021(7):39.
- [25] 李明洁,王明常,王凤艳,等. 基于多特征随机森林算法的深圳市土地利用分类[J]. 世界地质,2022,41(3):632.
- [26] 蒋伟萱,高金龙,陈江龙,等. 基于土地利用视角的乡村居业协同多尺度分析——以徐州市为例[J]. 自然资源学报,2020,35(8):2002.
- [27] 李志江,孙姗姗. 徐州市主城区土地利用转型过程及机制[J]. 江苏师范大学学报(自然科学版),2020,38(3):18.
- [28] 侯群群,王飞,严丽. 基于灰度共生矩阵的彩色遥感图像纹理特征提取[J]. 国土资源遥感,2013,25(4):26.

[责任编辑:钟传欣]