

致谢

岁月如流水，四年的本科生涯即将逝去，在浙大的这些岁月使我成长了许多，受益匪浅。

首先要感谢本文的指导老师陆丽珍，陆老师为人谦和，平易近人，治学严谨，具有极高的科研能力，在大学期间担任了很多门专业课的任课老师，给我展现广阔而丰富的 GIS 世界，从本课题的开始到实施，以及论文各部分的书写都给予了详细而有建设性的指导和建议。

感谢班主任张丰老师，张老师在我的学习和生活方面给予了很多指导和帮助。感谢刘仁义老师、杜震洪老师、林杰老师，老师们丰富的学识和对 GIS 的深刻见解让我对 GIS 有了不同的认识和理解。感谢苏程老师，苏老师本人和他教授的《遥感数字图像处理》都给我留下了深刻的影响。感谢其他学院为 GIS 专业授课的老师，让我们 GIS 的学生提高了跨学科的学习和研究能力。感谢学院的鲍雨欣老师，鲍老师在大学四年里给了我很多关怀，也感谢鲍老师指导我的学业成绩和琐碎生活。

感谢四年中遇到的每一个人。感谢 GIS 班级的每一位同学，大家一起上专业课、一起讨论各种问题，一起研究小组作业的场景仿佛就在昨日。感谢室友鲍俊辰、林昊、陈亦卓、金煜航、余志胜的理解与支持。感谢多次在专业课组队的吴文琪、张思远，我们在一起攻克了许多难题。感谢从中学就一路走来的徐慧琼、王煜、黎留林、赏所坤、罗熊同学，希望大家都有美好的未来。感谢我的父母和其他亲人，你们对我的肯定与支持是我前进的动力。同时也要感谢党和国家，让我有机会来到浙江大学学习和生活。

感谢所有参与论文评审和答辩的老师，感谢老师们给予本文提出的意见。

秦卫付

2021 年 5 月于浙大玉泉

摘要

随着能源和环境问题的日益突出,光伏发电等清洁能源在人类经济和社会的发展中扮演着越来越重要的角色。西北五省(自治区)是我国重要的光伏发电大省,近 10 年光伏电场总规模不断增长。基于遥感大数据开展大尺度的光伏电场空间分布制图研究,可以为西北地区光伏电场选址及布局优化提供数据基础。

本研究以我国西北五省(自治区)为研究区,在 Google Earth Engine(GEE)平台的支持下,基于 Sentinel-2 遥感数据,利用机器学习方法来识别光伏电场。具体包括:首先基于光谱、指数和纹理三类特征,采用分类与回归决策树(CART)和随机森林(RF)算法得到光伏电场初次分类结果,再通过人机交互方式划定光伏电场的大致区域,接着针对大致范围再次采用 CART 和 RT 算法进行二次精分类,并进行简单非迭代聚类(SNIC)超像素分割和形态学方法后处理和精度评价,最后利用距离分析等工具探明西北地区光伏电场的空间分布特征。研究表明:(1) RF 算法在提取光伏电场时效果明显优于 CART 算法;(2) 采用二次分类方法明显提高了分类精度,其分类精度和 Kappa 系数从初次分类的 90.89%和 0.8798 提升到了二次精分类的 97.42% 和 0.9483;(3) 光伏电场面积的计算值和统计值的误差为 16.66%,在可接受的范围内;(4) 光伏电场的空间分布特征为东多西少、北多南少,且约 72%的光伏电场距离沙漠小于 1 千米,其中 44%的光伏电场与荒漠的距离为 0,约 93.6%的光伏电场与主要道路间隔不到 100m,约 77%的光伏电场与城市的距离小于 15 千米。

关键词: 光伏电场; 特征提取; 遥感大数据; 机器学习分类; 空间分布特征

Abstract

As energy and environmental issues become increasingly prominent, clean energy such as photovoltaic (PV) power generation is playing an increasingly important role in the development of human economy and society. As the total scale of PV fields in the five northwestern provinces (autonomous regions), where are China's main PV power generation area, has been increasing in the past 10 years, large-scale research on the spatial distribution of PV fields with remote sensing big data can provide a basis for site selection and arrangement optimization of PV fields.

Taking five provinces (autonomous regions) in northwest China as research area, this study applied machine learning methods to identify photovoltaic electric fields (PVEF) from Sentinel-2 remote sensing images in Google Earth Engine (GEE) platform. Firstly, spectrum, index and texture features of Sentinel-2 were used to extract initial PVEF with classification and regression decision tree (CART) and random forest (RF) algorithms. Secondly, focusing on roughly ranges of PVEF which were delineated according to initial classification results, CART and RT algorithms were utilized again to obtain better accuracy classification results. Thirdly, the simple non-iterative clustering (SNIC) superpixel segmentation and morphological method were employed as post-processing for further improving the classification accuracy. Finally, distance analysis and statistical tools in ArcGIS were adopted to gain spatial distribution characteristics of PVEF in northwest China. The results show that: (1) Comparing with CART, RF algorithm is a more effective s algorithm for extracting PVEF information from remote sensing data; (2) The twice usage of RF and CART can improve the PVEF classification accuracy significantly; (3) The area deviation of PVEF between the classification results and statistical values is 16.66%, which further indicates the effectiveness of our approach of twice usage of classifiers; (4) The spatial distribution characteristics of the PVEF are: PVEF in the east and north are more than which in the west and south, about 72% of the PVEF are less than 1 km away from deserts, 44% of which locate in deserts,

about 93.6% are 100m away from the main road, and about 77% are less than 15 kilometers away from cities.

Keywords: photovoltaic electric field; feature extraction; remote sensing big data; machine learning classification; spatial distribution characteristics

目录

第一部分 毕业论文（设计）

1. 绪论	1
1.1 课题背景	1
1.2 国内外研究现状	3
1.2.1 遥感信息提取现状	3
1.2.2 大尺度遥感制图研究现状	5
1.3 本文研究内容与技术路线	6
1.3.1 研究内容	6
1.3.2 技术路线	6
2. 研究区域与数据	8
2.1 研究区域	8
2.2 研究数据	9
2.2.1 遥感数据	9
2.2.2 其他数据	10
3. 机器学习分类与空间分布特征分析	11
3.1 样本集建立	11
3.2 特征选取	12
3.3 机器学习分类	16
3.4 特征重要性评价与精度评定	17
3.4.1 特征重要性评价	17
3.4.2 精度评定	18
3.5 空间分布特征分析	19
4. 研究结果与分析	19
4.1 初次分类结果	19
4.2 光伏电场空间分布信息	20
4.2.1 光伏电场范围划定结果	20
4.2.2 光伏电场二次精分类结果	22

4.2.3	光伏电场分省制图	25
4.3	特征重要性评价与精度评定结果.....	28
4.3.1	特征重要性评价结果	28
4.3.2	精度评定结果	29
4.4	光伏电场空间分布特征	32
5.	总结与展望.....	34
5.1	总结	34
5.2	不足与展望	35
	参考文献	36
	附录	1
	作者简介	2
	《浙江大学本科生毕业论文（设计）任务书》	
	《浙江大学本科生毕业论文（设计）考核表》	