

基于面向对象随机森林算法的主被动遥感协同实现农作物自动识别

孙博学 吴永强 张静宜 马 玥

吉林建筑大学测绘与勘查工程学院 吉林长春 130118

摘要:以 Sentinel-1/2 哨兵卫星获取的雷达和多光谱遥感影像为数据源,提取多时相光谱反射和微波后向散射特征变量,基于面向对象的随机森林算法对吉林省扶余市水稻、玉米和大豆分类,并与支持向量机算法比较,综合评估基于多源遥感数据的面向对象随机森林算法对大范围农作物自动识别的适用性。结果表明:利用多时相多源遥感数据能够有效提高作物分类精度。面向对象随机森林算法的分类总体精度为 87.32%,高于支持向量机算法 9.82%。利用面向对象随机森林算法结合多时相主被动遥感特征变量能够准确地自动识别大范围农作物,可为作物种植面积调查与产量估值提供重要的支撑。

关键词:哨兵卫星;主被动遥感;面向对象随机森林算法;农作物自动识别

掌握农作物的种植分布状况,对于宏观指导农业生产、作物估产及农情监测具有重要意义^[1-3]。遥感技术能够快速实时监测地表信息,为识别作物提供高效快捷的技术手段。遥感光学传感器易受云雨天气影响,较难获取清晰的影像图,不同作物在可见光—红外波段下特征相似,识别精度较低。微波传感器发射的微波能够穿透云层,不易受天气影响,可获得清晰的雷达影像数据,用以表征不同作物的后向散射特征。但由于微波传感器成像方式不同,作物影像特征不易识别,仅利用微波影像分类作物存在较大困难^[4-5]。应用主被动遥感技术协同实现农作物自动识别对分类精度的提高具有重要意义。

分类方法的选择在提高作物识别精度方面至关重要。面向对象分类法将同质像元组成的对象作为基本分类单元,能够有效保留地物的空间结构、形状和纹理等特征,降低独立像元噪声的影响,有助于提高分类精度^[6-7]。随机森林算法是目前应用最广泛的机器学习算法,模型泛化能力强、稳定性高,但在多时相多源遥感数据作物识别方面仍需进一步验证分析^[8]。

以 Sentinel-1/2 哨兵卫星获取的雷达和多光谱遥感影像为数据源,提取多时相光谱反射和微波后向散射特征变量,基于面向对象随机森林算法对吉林省扶余市水稻、玉米和大豆自动分类,并与支持向量机算法比较,综合评估基于多源遥感数据的面向对象随机森林算法对大范围农作物自动识别的适用性。

一、研究区概况和数据源

(一)研究区概况

研究区位于中国吉林省扶余市,主要包括大林子镇、更新乡、新万发镇、弓棚子镇、五家站镇和肖家乡共六个镇

(乡)级行政区,地理坐标为 44°52′~45°25′N, 125°24′~125°59′E,年平均气温 4.5℃,平均降水 145.8mm,区内作物主要包括水稻、玉米和大豆。

(二)数据源

采用 2020 年 4—10 月逐月的 Sentinel-1/2 遥感影像数据。Sentinel-1 雷达 C 波段数据为 GRD 1 级产品, VV 和 VH 两种极化方式。Sentinel-2 多光谱数据为 LIC 级产品,包括可见光-近红外到短波红外共 13 个波段。利用 SNAP 软件对 Sentinel-1/2 数据预处理。Sentinel-1 数据进行辐射定标, Refined Lee 滤波和地形校正处理。Sentinel-2 数据借助 Sen2cor 功能模块进行辐射定标和大气校正。将预处理后的 Sentinel-1/2 数据利用 ENVI 5.3 软件进行空间配准,空间分辨率统一为 10m。

二、分类方法与精度评价

(一)面向对象分类法

利用 eCognition 软件的多尺度分割功能将待分类影像分割成同质性最小的对象单元,需设置三个重要参数,即尺度参数、形状参数和紧致度参数^[9-10]。尺度参数的大小决定分割对象的面积大小,直接影响分类效率和精度。形状参数和紧致度参数可以进一步优化对象的边缘特征,使其更加准确地表达地物细节信息。通过多次试验,最终确定的尺度参数、形状参数和紧致度参数分别为 80、0.8 和 0.2。

(二)随机森林和支持向量机算法

随机森林算法(RF)是由多个 CART 决策树组合构成的集成学习算法。RF 采用 bootstrap 抽样技术从原始数据集中抽取 N 个训练数据集生成 N 个 CART 决策树,在每个 CART 决策树生长过程中,从全部 M 个特征变量中随机抽

选 m 个 ($m \leq M$), 根据 Gini 系数最小原则选出最优属性进行内部节点分支, 最后采用投票法综合 N 个决策树的预测结果确定最终的样本类别^[11]。本文设置决策树的个数 N 为 1000, 节点分裂的特征变量数为总特征变量数的开平方值。

支持向量机算法 (SVM) 是一种基于统计学习理论的机器学习算法, 通过选择相应的非线性映射函数, 将输入向量从原空间映射到高维特征空间, 采用结构风险最小化准则, 在最小化样本误差的同时缩小模型泛化误差的上界, 提高模型的泛化能力^[12]。本文设置 SVM 的惩罚系数 C 值设为 1, 选用 RBF 核函数。

(三) 精度评价

综合 2020 年野外地面调查样点数据和 Google Earth 高分辨率影像, 随机选取 560 个验证点, 计算混淆矩阵, 采用总体精度、Kappa 系数、制图精度和用户精度对分类结果进行精度评价。

三、农作物分类遥感影像特征分析

(一) 多源遥感数据特征分析

利用 Sentinel-1 的 VH 和 VV 极化数据分别计算 VV/VH 比值极化指数和 $(VV-VH)/(VV+VH)$ 归一化极化指数。利用 Sentinel-2 的多光谱反射率数据计算 NDVI 归一化植被指数、DVI 差值植被指数和 RVI 比值植被指数。借助 ENVI 5.3 软件制作水稻、玉米和大豆作物的 ROI 感兴趣区, 获取不同作物的后向散射、光谱反射率和多种指数的均值, 分析作物在不同生长时间下的遥感数据特点。经

过分析可知, 对于 Sentinel-1 数据, 三种作物随季节变化有不同的生长状态, 微波后向散射系数也随之变化。在 VH 和 VV 极化模式下, 水稻的后向散射系数最低, 大豆和玉米的后向散射系数较接近。对于 VV/VH 和 $(VV-VH)/(VV+VH)$ 两种极化指数, 水稻、玉米和大豆的曲线特征相似, 但经过数学比值运算后, 使三种作物的后向散射系数差异更加显著, 有利于区分不同的作物。对于 Sentinel-2 数据, 三种作物在 4~6 月播种至出苗的过程中, 光谱反射率逐渐升高, 6~8 月生长期中, 反射率逐渐降低, 9~10 月完全成熟至收获后, 反射率逐渐升高。在播种至成熟的过程中, 水稻的光谱反射率最低, 大豆和玉米的反射率较高, 且二者的反射率值较为接近。在红边波段 (B5-B7, B8A) 处, 三种作物的光谱反射率差异较可见光 (B1-B4) 和短波红外波段 (B11-B12) 处显著。三种作物从播种至收获, 植被指数数值先升高后降低, 在 7~9 月三种作物的指数值差异最为显著。

(二) 分类特征变量选取

针对 4 个雷达后向散射特征变量和 15 个光谱反射特征变量, 将三种作物类型两两对比, 筛选出特征值差异最显著的月份, 用于后续参与分类模型的构建。以 VV 特征变量为例, 大豆和水稻在 5 月后向散射系数差值最大, 大豆和玉米在 10 月后向散射系数差值最大, 水稻和玉米在 5 月后向散射系数差值最大, 对于 VV 特征变量, 最终筛选 5 月和 10 月两个月份的特征变量数据用于后续分类。分类特征变量筛选情况如表 1 所示。

表 1 分类特征变量筛选

卫星数据	特征变量(月份)	特征变量总数
Sentinel-1	VV(5,10), VH(6,10), VV/VH(5), (VV-VH)/(VV+VH)(5)	6
Sentinel-2	B1(5,6), B2(4,6), B3(5,6), B4(6,10), B5(6,7), B6(6,8), B7(6,8), B8(6,8), B8A(6,8), B9(6,8), B11(5,7), B12(5,7), NDVI(6,10), DVI(8,9), RVI(7,8)	30

四、分类实验及精度评价

为了明确多时相、多源遥感数据的综合应用对农作物种植结构分类精度的影响, 针对不同的特征变量数据集, 共设定 5 组实验, 即单一时相多光谱特征数据集 (OPTS)、单一时相雷达特征数据集 (SARS)、多时相多光谱特征数据集 (OPTM)、多时相雷达特征数据集 (SARM) 和多时相多源特征数据集 (OPTM-SARM)。OPTS 组选取作物光谱反射率差异显著的 7 月 Sentinel-2 多波段数据为特征变量, SARS 组选取作物后向散射系数差异显著的 6 月 Sentinel-1 两种极化数据为特征变量。选用 RF 和 SVM 算法进行作物分类。

(一) 特征变量对分类精度的影响

由表 2, 多时相模型 (OPTM、SARM) 的总体精度和 Kappa 系数高于单一时相模型 (OPTS、SARS), 多时相模型比单一时相模型的整体精度可提升 3.39% 以上, 最高可达 13.93%。选取水稻、玉米和大豆在生长周期中反射或后向散射差异显著的多时相数据参与分类, 能够增强不同作物

间的影像特征差异, 有利于提高识别精度。从单一数据源模型 (OPTM、SARM) 到多源数据模型 (OPTM-SARM) 的总体精度显著增加, 最高可提升 9.82%, 综合利用光学和雷达影像能够丰富农作物的影像特征, 弥补利用单一光学遥感数据或微波遥感数据的不足, 可有效提高具有相似特点的作物识别精度。

表 2 不同特征组合的分类总体精度和 Kappa 系数

实验模型	RF		SVM	
	总体精度/%	Kappa 系数	总体精度/%	Kappa 系数
SARS	71.43	0.6196	60.36	0.4731
OPTS	76.07	0.6812	73.57	0.6480
SARM	77.50	0.7009	74.29	0.6579
OPTM	82.50	0.7669	76.96	0.6935
OPTM-SARM	87.32	0.8311	77.50	0.7007

(二) 分类算法对分类精度的影响

由表 3, 利用 RF 和 SVM 算法结合多时相多源特征数

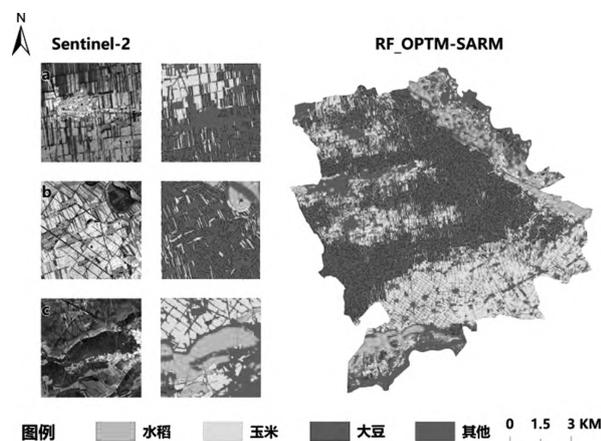
据集对水稻、玉米和大豆分类,并计算混淆矩阵。RF 算法的总体精度和 Kappa 系数为 87.32% 和 0.8311,较 SVM 算法高 9.82% 和 0.1304。RF 算法中水稻、玉米和大豆的制图精度较 SVM 算法高 6%、4.29% 和 25.72%,用户精度较 SVM 算法高 0.15%、11.08% 和 8.9%。通过对比分析,RF 算法的模型性能较好,泛化能力较强,能够高效地处理高维数据,更高精度地精细化提取农作物种植结构信息,适用于大范围农作物的种植结构制图。

表 3 不同分类算法的精度比较

类型	RF		SVM	
	制图精度/%	用户精度/%	制图精度/%	用户精度/%
水稻	92.67	97.89	86.67	97.74
玉米	80.00	89.60	75.71	78.52
大豆	82.86	86.57	57.14	77.67
其他	93.85	76.73	90.77	62.43
	总体精度/%: 87.32	Kappa: 0.8311	总体精度/%: 77.50	Kappa: 0.7007

五、分类结果

利用基于面向对象随机森林算法自动绘制研究区农作物种植结构图,获得的分类结果比较准确,较少出现错分和漏分现象,能够准确划分不同农作物的地块边界。面向对象分类法有效改善了像素分类法中独立像元引起的“椒盐”现象,最大程度地保留了地物间的空间邻域关系,可视化效果较好。



基于面向对象随机森林算法的农作物种植结构分类图

结语

本文利用面向对象随机森林算法结合 Sentinel-1 微波数据和 Sentinel-2 多光谱数据实现了对吉林省扶余市水稻、玉米和大豆大宗农作物种植结构的自动识别。多时相多源遥感数据的应用极大程度丰富了地物的遥感影像特征,增强了不同作物的影像特征差异,有利于提高分类精度。面向对象随机森林方法可以高效并行处理高维数据,有效挖掘特征信息,获得的农作物种植结构图精度较高,可视化效果较好,可以充分保留地物空间邻域特征。协同

利用主被动遥感技术能够高效准确地实现大范围的农作物种植结构监测,为农业生产与管理提供有力的科学依据和数据支撑。

参考文献:

- [1] 潘晓卉,张平宇,刘世薇.东北地区主要粮食作物种植结构时空变化分析[J].土壤与作物,2018,7(3):303-311.
 - [2] 代海涛.吉林省大豆生产发展现状及对策分析[J].大豆科技,2009,000(002):51-53.
 - [3] 李冰,梁燕华,李丹丹,等.多时相 GF-1 卫星 PMS 影像提取农作物种植结构[J].中国农业资源与区划,2017,38(9):56-62.
 - [4] 朱凤敏,吴迪,杨佳琪.基于 Sentinel-1B SAR 数据的农作物分类方法研究[J].测绘与空间地理信息,2020,43(5):105-108.
 - [5] 韩冰冰,陈圣波,曾庆鸿,等.基于 J-M 距离的多时相 Sentinel-1 农作物分类[J].科学技术与工程,2020,20(17):6977-6982.
 - [6] 毕恺艺,牛锋,黄妮,等.基于 Sentinel-2A 时序数据和面向对象决策树方法的植被识别[J].地理与地理信息科学,2017,33(5):16-20+27+127.
 - [7] 贾明明,王宗明,张柏,等.综合环境卫星与 MODIS 数据的面向对象土地覆盖分类方法[J].武汉大学学报(信息科学版),2014,39(3):305-310.
 - [8] 王利民,刘佳,杨玲波,等.随机森林方法在玉米-大豆精细识别中的应用[J].作物学报,2018,44(4):569-580.
 - [9] 杜保佳,张晶,王宗明,等.应用 Sentinel-2A NDVI 时间序列和面向对象决策树方法的农作物分类[J].地球信息科学学报,2019,21(5):740-751.
 - [10] 张澜,王妮,朱冰雪,等.基于无人机高分辨率影像的农作物分类研究[J].江西农业学报,2022,34(02):178-187.
 - [11] 王德军,姜琦刚,李远华,等.基于 Sentinel-2A/B 时序数据与随机森林算法的农耕地土地利用分类[J].国土资源遥感,2020,32(4):236-243.
 - [12] 张东彦,戴震,徐新刚,等.基于时序 Sentinel-2 影像的现代农业园区作物分类研究[J].红外与激光工程,2021,50(05):262-272.
- 基金项目:** 吉林省教育厅科学技术研究项目(JJKH 20210269KJ);吉林省大学生创新创业训练计划项目(202110191108)
- 作者简介:** 孙博学(2001—),男,汉族,天津人,本科,研究方向:环境遥感和遥感图像分类。