

doi:10.6041/j.issn.1000-1298.2022.01.025

基于 Google Earth Engine 的黄土高原覆膜农田遥感识别

郑文慧^{1,2} 王润红^{1,2} 曹银轩^{1,2} 靳 宁³ 冯 浩^{2,4} 何建强^{1,2}

(1. 西北农林科技大学旱区农业水土工程教育部重点实验室, 陕西杨凌 712100;

2. 西北农林科技大学中国旱区节水农业研究院, 陕西杨凌 712100;

3. 山西能源学院资源与环境工程系, 晋中 030600; 4. 中国科学院水利部水土保持研究所, 陕西杨凌 712100)

摘要: 为了建立覆膜农田遥感识别技术体系,本研究选取甘肃省定西市安定区团结镇作为黄土高原地膜覆盖旱作农业代表性区域,基于 Google Earth Engine 云平台和 Landsat-8 反射率数据,采用特征重要性分析优选纹理特征,利用参数优化后的随机森林算法提取覆膜农田区域并选出最佳特征组合方案,最后通过对比随机森林、支持向量机、决策树和最小距离分类 4 种算法的分类结果来评价不同分类算法的性能。结果表明:优化关键参数后的随机森林算法能够显著提高遥感影像的分类精度;单一特征方案中,基于光谱特征的分类精度最高,且加入指数和纹理特征可提高总体识别精度;利用随机森林特征重要性分析选取的优选纹理特征分类性能优于全部纹理特征,基于“光谱+指数+优选纹理”特征方案的识别结果最佳,总体精度和 Kappa 系数达 95.05% 和 0.94;与支持向量机、决策树和最小距离分类相比,随机森林优势明显,总体精度分别高 3.10、7.74、50.78 个百分点。本研究实现了对地形复杂地区覆膜农田空间分布较为精准的识别。

关键词: Google Earth Engine; 覆膜农田; 遥感识别; 随机森林; 特征选择; 黄土高原

中图分类号: S127

文献标识码: A

文章编号: 1000-1298(2022)01-0224-11

OSID:



Remote Sensing Recognition of Plastic-film-mulched Farmlands on Loess Plateau Based on Google Earth Engine

ZHENG Wenhui^{1,2} WANG Runhong^{1,2} CAO Yinxuan^{1,2} JIN Ning³ FENG Hao^{2,4} HE Jianqiang^{1,2}

(1. Key Laboratory of Agricultural Soil and Water Engineering in Arid and Semiarid Area, Ministry of Education, Northwest A&F University, Yangling, Shaanxi 712100, China

2. Institute of Water-saving Agriculture in Arid Areas of China, Northwest A&F University, Yangling, Shaanxi 712100, China

3. Department of Resources and Environment, Shanxi Institute of Energy, Jinzhong 030600, China

4. Institute of Water and Soil Conservation, Chinese Academy of Sciences and Ministry of Water Resources, Yangling, Shaanxi 712100, China)

Abstract: Plastic-film-mulching has made an outstanding contribution to agricultural production and food security in China, but also caused many serious environmental problems. It is very important to quickly and accurately obtain the spatial distribution information of plastic-mulched farmlands. In order to establish a framework for remote sensing recognition of plastic-film-mulched farmland, the Tuanjie Town of Dingxi City in Gansu Province was chosen as the research area, which was a typical dry farming agricultural area with heavy plastic film application on the Loess Plateau. Based on the Google Earth Engine, Landsat-8 reflectance data was used to analyze the importance of different features and select the optimal textural features. Then, the random forest (RF) algorithm with optimized parameters was used to extract the plastic-film-mulched farmland area and select the best feature combination. Finally, based on the best feature combination, the performance of RF algorithm was evaluated through comparison between the classification results based on the other algorithms of support vector machines

收稿日期: 2020-12-19 修回日期: 2021-01-29

基金项目: 国家自然科学基金项目(52079115)、陕西省重点研发计划项目(2019ZDLNY07-03)、陕西省气象局秦岭和黄土高原生态环境气象重点实验室开放研究基金项目(2019Z-5)、西北农林科技大学人才专项资金项目(千人计划项目)和高等学校学科创新引智计划(111计划)项目(B12007)

作者简介: 郑文慧(1996—),女,博士生,主要从事农业生态系统模拟研究,E-mail: zwh9623@nwfufu.edu.cn

通信作者: 何建强(1977—),男,教授,博士生导师,主要从事农业生态系统模拟研究,E-mail: jianqiang_he@nwsuaf.edu.cn

(SVM), decision tree (DT), and minimum distance classifier (MDC), respectively. The results showed that the optimized parameters of RF algorithm could greatly improve the classification accuracy. Among the schemes based on single kind of features, the accuracy of scheme based on spectral features was the highest. The addition of index and textural features could also improve the overall identification accuracy to some extent. The performance of the selected optimal texture features was better than that of all texture features. The recognition result based on the combination of ‘spectral + index + optimal textural features’ was the best, whose overall accuracy and Kappa coefficient were 95.05% and 0.94, respectively. The overall accuracy of RF algorithm was 3.10 percentage points, 7.74 percentage points and 50.78 percentage points higher than the algorithms of SVM, DT and MDC, respectively, which proved the RF algorithm had some obvious advantages in recognition of plastic-film-mulched farmlands. The research realized an accurate identification of plastic-film-mulched farmlands in areas with complex terrain features in China. The results can provide theory and technology supports for the studies related to spatial variations and sustainability production with plastic film mulching in the near future.

Key words: Google Earth Engine; plastic-film-mulched farmland; remote sensing recognition; random forest; feature selection; Loess Plateau

0 引言

黄土高原位于我国西北地区中部,是典型的旱作农业区,在国家粮食安全体系中具有举足轻重的作用^[1]。黄土高原蒸发强烈,干旱少雨且年内降雨分布不均匀^[2],同时田间管理水平落后,导致水资源严重匮乏,限制了黄土高原农业的发展^[3-4]。因此,采取有效的田间管理措施提高水资源利用效率是旱区农业可持续发展的必然趋势。地膜覆盖技术具有节水、保墒和增产的性能^[5],因此在黄土高原大田作物生产中被广泛应用^[6]。然而,随着地膜用量的逐渐增加和覆盖面积的不断扩大,地膜易破碎、难降解、难回收的短板日益显露,生态环境问题日益严重^[7-8]。精准获取覆膜农田的时空分布信息有助于发挥地膜覆盖技术的积极作用、缓解生态环境压力,也有助于相关部门对地膜的使用和分布进行监管和调度,提高对地膜污染的防治能力。

随着科学技术的深入发展,应用遥感技术快速提取覆膜农田信息成为可能。遥感技术具有信息获取快、观测范围大等特点,能够宏观地把握地物空间分布概况^[9-11],从而解决了大范围内覆膜农田监测的难题。近年来,覆膜农田的监测和识别逐渐发展为农业遥感领域的研究热点^[12-15]。目前,我国大多数相关研究地区主要集中在新疆等地膜连片分布地区,该地区的地膜识别相对容易。此外,已有研究多采用支持向量机和决策树等单一算法进行分类,较少采用多分类器集成算法。

谷歌地球引擎(Google Earth Engine,GEE)是专门用于卫星影像及其它空间数据解译运算的开源智能云平台^[16],它将谷歌最先进的云计算能力和存储能力用于处理各类热点问题,为遥感工作者及其他公众用户提供了便利^[17-19]。近年来,GEE 平台逐

渐被广泛应用于农田面积提取等研究^[20-21]。GEE 平台改变了传统遥感软件处理数据的定式,解决了大尺度应用研究中数据收集难、解译效率低的弊端,从而为黄土高原覆膜农田的识别提供了可能。

本研究以甘肃省定西市安定区团结镇作为黄土高原覆膜农田的典型区域,借助 GEE 平台,利用分辨率 30 m 的陆地卫星地表反射率数据(USGS Landsat 8 Surface Reflectance Tier 1),采用特征选择和随机森林算法对该地区的覆膜农田进行提取,构建基于 GEE 平台的黄土高原覆膜农田识别方法框架,以研究随机森林算法的关键参数对遥感影像地物分类结果的影响。通过分析不同特征组合下的遥感影像分类结果,确定识别黄土高原覆膜农田的最优特征组合方案;通过比较不同的分类算法,验证随机森林算法在黄土高原覆膜农田识别中的有效性。

1 材料和方法

1.1 研究区概况

本研究选择黄土高原和西秦岭山地交汇处的甘肃省定西市安定区团结镇作为研究区域(图 1)。该地区地势南高北低,山脉南北走向,属温带大陆性季风气候,总面积约 134.43 km²,年平均气温 6.3℃,有效积温 2 300℃,年平均降雨量 430 mm,无霜期 140 d。统计数据显示,截至 2019 年底,团结镇总耕地面积达 53.11 km²,约占全镇总面积的 2/5^[22]。该地区属于典型的黄土高原地膜覆盖种植区,地膜覆盖马铃薯种植面积占总种植面积的 90% 以上。

1.2 数据来源与处理

1.2.1 Landsat-8 卫星数据

本研究所需 Landsat-8 卫星数据来源于 GEE 平台(<https://earthengine.google.com>)的陆地卫星地表反射率数据(USGS Landsat 8 Surface

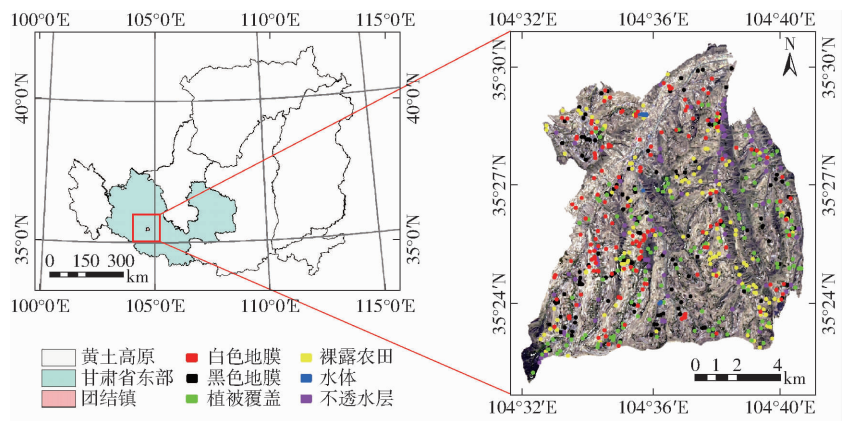


图1 研究区域位置图和地面样本点分布示意图

Fig. 1 Location schematic of study area and distributions of land surface sample points

Reflectance Tier 1), 空间分辨率为 30 m, 时间分辨率为 16 d。该数据已进行过大气校正, 消除了大气和光照等因素所造成的辐射误差。

黄土高原主要的覆膜作物为春玉米和马铃薯, 一般于 4 月下旬至 5 月上旬播种^[23-24]。在 5 月上中旬时, 冬小麦正处于拔节抽穗期, 夏玉米还未种植, 春玉米和马铃薯刚刚播种, 此时地膜受作物叶片的干扰程度小, 容易获取光谱信息。因此, 5 月上中旬是解译研究区地膜覆盖范围的最佳时相。首先利用时间和空间的过滤函数获取团结镇上空 2018 年 5 月 11 日的 Landsat-8 影像, 轨道号为 130/035; 然后对该影像进行去云处理, 以减少云层遮挡的影响; 最后使用团结镇的矢量图进行影像裁剪。

1.2.2 地面样本数据

团结镇内的地物类型主要包含白色地膜、黑色地膜、植被覆盖、裸露农田、水体和不透水层 (包括建筑、道路、山体、废弃用地等) 共 6 种土地覆盖类型。基于地球大数据科学工程数据共享服务系统 (<http://data.casearth.cn/>) 提供的全球分辨率 30 m 精细地表覆盖数据和 Google Earth 高分辨率影像, 采集了团结镇 6 种不同土地覆盖类型共计 648 个矩形样本, 样本尺寸为 2 像素×2 像素, 并在 GEE 平台中随机分为训练数据和验证数据两类, 两类数据的数量基本一致 (表 1)。

表 1 研究区地物分类体系及样本数量

Tab. 1 Classification scheme and sample number of surface features in study area

地物类型	训练样本数	验证样本数	样本总计
白色地膜	78	76	154
黑色地膜	62	80	142
植被覆盖	71	61	132
裸露农田	67	51	118
水体	7	6	13
不透水层	40	49	89
总计	325	323	648

1.3 基于 GEE 平台的覆膜农田识别

1.3.1 基于 GEE 平台的覆膜农田识别框架

基于 GEE 平台提取团结镇覆膜农田区域的具体步骤 (图 2) 为: ①对获取的 Landsat-8 影像数据进行预处理, 并提取影像的光谱特征、指数特征和纹理特征。②使用提取的遥感识别特征和训练数据选取随机森林 (Random forest, RF) 算法的关键参数。③利用参数优化后的 RF 算法进行纹理特征重要性分析, 获得优选纹理特征。④基于提取的遥感识别特征和优选纹理特征制定 6 种不同的分类方案, 利用参数优化后的 RF 算法进行不同方案下的地物分类并从中提取出覆膜农田区域, 通过混淆矩阵方法和验证数据对分类结果进行精度评价, 确定最佳分类方案。⑤使用最佳分类方案和同一地面样本数据分别利用支持向量机 (Support vector machines,

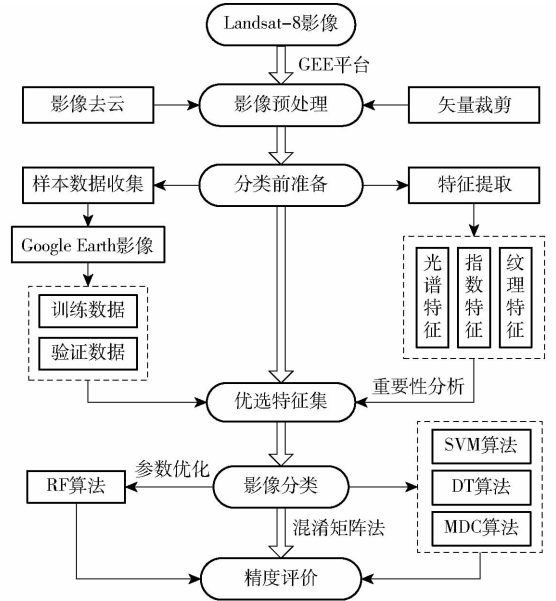


图2 基于 GEE 平台提取研究区覆膜农田区域的技术流程图

Fig. 2 Technical flowchart of extraction of plastic-film-mulched farmland based on Google Earth Engine platform in study area

SVM)、决策树 (Decision tree, DT) 和最小距离分类 (Minimum distance classifier, MDC) 算法进行地物分类, 并进行精度验证, 通过与 RF 分类精度进行比较和 McNemar's 检验, 评价 RF 算法的分类性能。

1.3.2 遥感识别特征构建

本研究基于 Landsat-8 地表反射率数据提取了光谱、指数和纹理 3 组特征进行地膜识别。其中, 光谱特征包括遥感影像可见光、近红外和短波红外 7 个波段的反射率数据。指数特征包括归一化植被指数^[25]、归一化水体指数^[26]和归一化建筑指数^[27]。纹理特征选用灰度共生矩阵 (Gray level co-occurrence matrix, GLCM) 来提取。GLCM 是由 HARACLICK 等^[28]提出的通过计算影像灰度级之间联合条件概率密度来提取纹理特征的一种统计学方法。GEE 平台提供了基于 GLCM 快速计算纹理特征的函数, 可在短时间内同时导出 18 种不同纹理特征。若将 18 种纹理测度全部用于机器学习分类必然会产生冗余, 因此本研究选取最为常见的和平均、方差、对比度、异质性、逆差矩、熵、角二阶矩和相关性 8 种纹理指标来区分不同地物类型的空间结构差异。由于光谱特征中考虑了影像的 7 个波段, 因此选用的纹理特征共包含 56 (8 × 7) 个波段信息。最终共提取了 66 个光谱特征、指数特征以及纹理特征, 其中纹理特征占全部遥感特征的 80% 以上。

为了进一步提高覆膜农田的识别精度和算法运行效率, 研究利用 RF 算法对 56 个纹理特征进行重要性分析, 然后按照重要性排序, 逐一添加纹理特征开展覆膜农田识别工作, 最后通过评价分类的生产者精度、用户精度、总体精度和平均精度 (白膜和黑膜的生产者精度与用户精度的平均值) 来选取最优纹理特征组合, 从而实现基于最优特征集下的覆膜农田识别。

1.3.3 随机森林关键参数对分类结果的影响

RF 算法在一定程度上可以避免过拟合^[29-31], 对噪声和异常值有较好的容忍性, 表现出很多普通机器学习算法未有的独特优势^[30,32]。

RF 算法可在 GEE 平台上实现, 且在构建 RF 算法时仅需对两个最为关键的参数进行优化, 即决策树数量 T 和节点分裂特征个数 M , 其他参数保持默认值。一般地, 增加决策树数量 T 可有效减小算法的泛化误差, 但同时降低计算效率; 节点分裂特征个数 M 是单棵决策树分类性能的决定性因素, 并对树之间的相关性产生影响^[33]。

RF 算法中参数 T 取值上限一般设置为 1 000, M 默认为特征数的平方根 (本文为 $\sqrt{66} \approx 8$)。因此本研究在构建 RF 算法时将参数 T 的取值范围设为

1 ~ 1 000, 参数 M 分别取 2、4、6、8 和 10, 基于训练数据遍历所有参数组合方式下的 RF 地物分类, 利用袋外误差作为评价依据并综合考虑算法的运算效率和稳定性, 最终确定参数 T 和 M 的最优取值。

1.3.4 团结镇覆膜农田的识别

为了评价不同特征组合在覆膜农田识别中的表现, 本研究制定了 6 组不同的分类方案, 具体为: 光谱特征 (方案 S); 指数特征 (方案 I); 优选纹理特征 (方案 T1); 光谱 + 指数特征 (方案 S + I); 光谱 + 指数 + 优选纹理特征 (方案 S + I + T1); 光谱 + 指数 + 全部纹理特征 (方案 S + I + T2)。利用参数优化后的 RF 算法和 325 个训练数据进行不同特征组合下的地物分类, 并从分类结果中提取出覆膜农田区域, 对比分析不同分类方案的分类精度, 从而选取最佳分类方案。

为了进一步比较不同分类算法在覆膜农田识别中的有效性, 本研究基于最佳分类方案, 使用同一训练和验证样本, 分别采用 SVM、DT 和 MDC 算法提取覆膜农田区域, 依据分类精度和 McNemar's 检验结果与 RF 算法提取结果作对比分析, 从而评价不同算法的分类性能。SVM 算法是以统计学为理论基础, 以结构风险最小化为原则的机器学习方法^[34], 核函数是影响其分类性能的决定性因素之一, 本研究选用线性核函数进行分类, 该核函数已被证明在覆膜农田识别中效果最佳^[13]。DT 算法是一种依据分割阈值对遥感影像的像元进行归类的非参数监督分类方法^[35], 包括 ID3、C4.5 和 CART 共 3 种常见类型, 本研究选用其中的 CART 算法执行地物分类。MDC 算法是通过计算未知类别向量与训练样本向量中心点的距离, 将非样本像元归类到距离最短类别中的分类方法^[36], 本研究采用欧氏距离计算距离。

1.3.5 识别精度验证

采用混淆矩阵法对分类结果进行精度评价, 包括生产者精度 (Producer accuracy, PA)、用户精度 (User accuracy, UA)、总体精度 (Overall accuracy, OA) 和 Kappa 系数 4 个评价指标^[37-38]。

此外, 利用 McNemar's 检验来检测不同算法间分类精度的差异, 该检验是基于两个算法的混淆矩阵而计算的非参数检验, 以 Z 值来评判算法间差异的显著性^[39]。现有研究表明, 若 $|Z| > 1.96$, 则表明在 0.05 检验水平下两种算法间的分类精度差异性显著。

2 结果与分析

2.1 随机森林关键参数对分类结果的影响

决策树数量 T 和节点分裂特征个数 M 对 RF 算

法的袋外误差均有明显的影响(图3)。随着 T 的增大,袋外误差开始降低,尤其是当 $T < 100$ 时,此时因还未形成“森林”,袋外误差明显降低;当 $T \geq 100$ 时,袋外误差趋于稳定。节点分裂特征个数 M 的增加也可以有效降低袋外误差,但相比参数 T 的影响较小,且参数 M 对算法的作用也与 T 的取值有关。当 $T < 100$ 时, M 的增加能够引起袋外误差大幅度降低,而当 $T > 100$ 时, M 的变化对袋外误差的影响不明显。当 $M < 6$ 时,袋外误差降低相对明显;在 $M \geq 6$ 时误差变化幅度很小,尤其在 $T \geq 100$ 后基本趋于稳定。因此,兼顾 RF 算法的稳定、精度与效率,本研究选取 $T = 600$ 、 $M = 8$ (即默认值)作为 RF 算法的输入参数用于下一步最优特征子集的筛选。

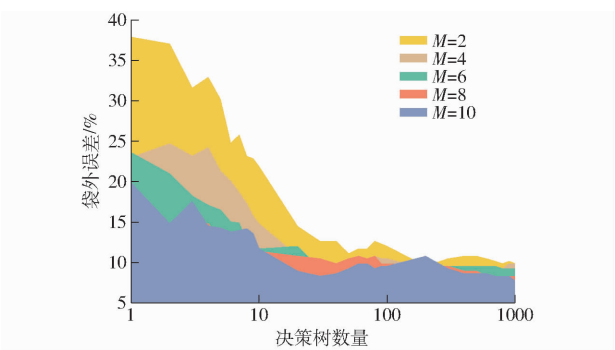
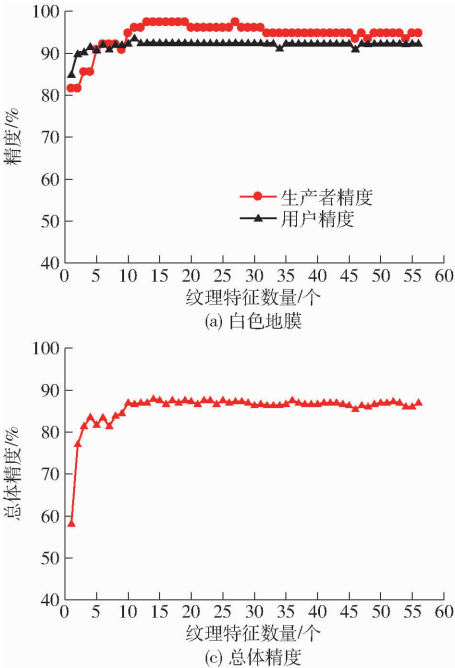


图3 随机森林算法中决策树数量 T 和节点分裂特征个数 M 与袋外误差的关系

Fig.3 Relationships between number of trees (T) or variables per split (M) and error of out-of-bag dataset in random forest algorithm

2.2 最优纹理特征组合的选取

白膜覆盖农田识别的生产者精度和用户精度波



动较小,且随纹理特征数量的增加均呈缓慢上升趋势,在特征数量达到 12 时,两种精度趋于平稳并保持在 90% 以上(图 4a)。黑膜覆盖农田识别的生产者精度和用户精度的变化幅度较大,当特征数量小于 4 时,生产者精度直线增长,随后转变为小区间内浮动,精度保持在 85% 以上;而用户精度则在特征数量增加到 2 时,由 53.13% 直接上升到 82.35%,随后保持在 85% 上下波动(图 4b)。特征数量小于 10 时,总体精度呈现整体上升趋势,之后随着特征数量的增加而逐渐趋于平缓,在特征数量为 14 和 15 时达到较大值,其值依次为 87.93% 和 87.62% (图 4c)。根据平均精度,即白膜和黑膜的生产者精度与用户精度的平均值(图 4d),可知该值在特征数量为 15、23 和 36 时较高,分别为 90.66%、90.79% 和 90.86%。

总体而言,本研究的 RF 算法运行过程中,使用的特征数量越多,则运算时间越长,工作效率越低,而通过重要性分析和精度评价筛选特征,将贡献度和重要性较小的变量予以剔除,进而以较少的特征变量来确保较高的分类精度,降低变量维度。因此,本研究确定前 15 个纹理特征作为最优纹理特征组合。

2.3 不同特征组合下随机森林覆膜农田识别结果

根据不同特征组合下随机森林地物分类结果的空间分布(图 5),可以看出方案 I(图 5b)与方案 S+I(图 5d)的分类结果中不透水层的分布范围大于其他方法,并且已经超出了 Google Earth 高分辨率影像中目视解译的不透水层范围,存在一定程度

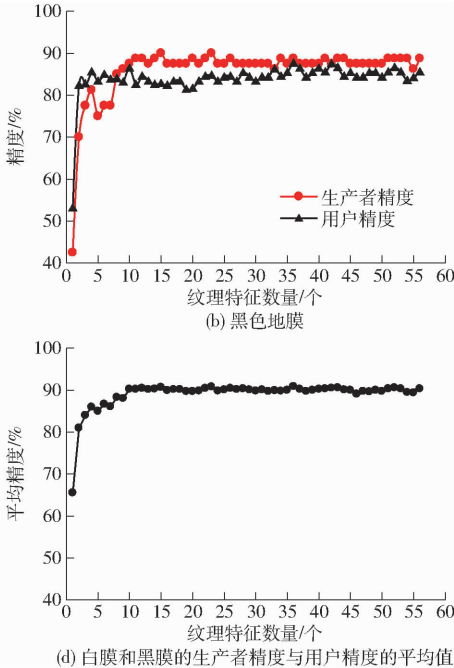


图4 不同纹理特征数量与识别精度之间的关系曲线

Fig.4 Relationships between number of textural features and recognition accuracy

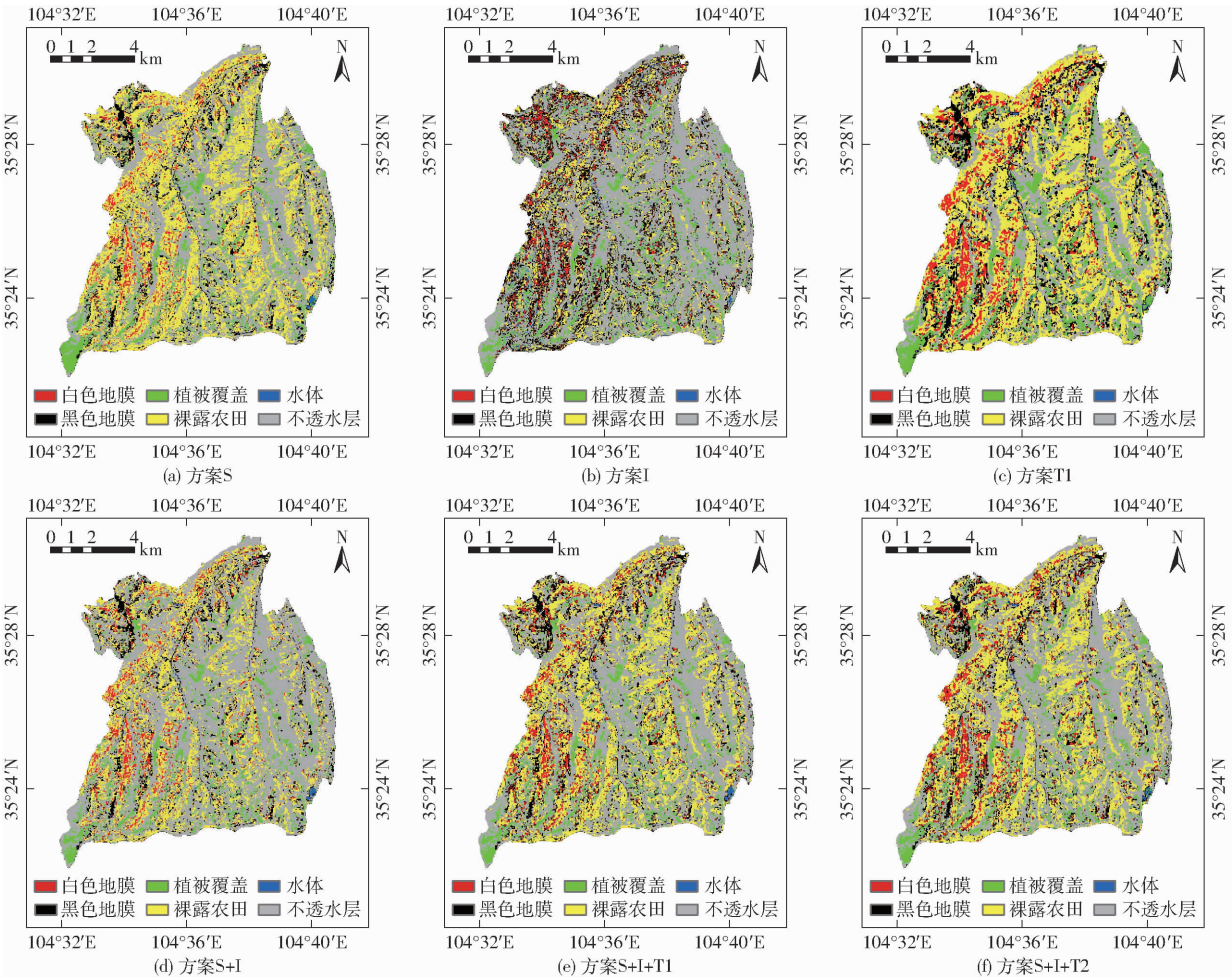


图 5 基于随机森林算法和不同特征组合下的研究区地物分类结果

Fig. 5 Classification results of land covers in study area based on random forest algorithm and different feature combinations

的不合理性。方案 T1(图 5c)获取的分类结果中农田区域约占研究区的 1/2,这与统计数据(约 2/5)不相符,因此仅依靠优选纹理特征的识别结果也不可靠。而方案 S(图 5a)、方案 S + I + T1(图 5e)以及方案 S + I + T2(图 5f)得到的分类结果中各类地物的空间分布较为合理,也与 Google Earth 高分辨率影像大致相符。

根据不同分类方案下随机森林覆膜农田识别结果的空间分布(图 6),可以看出覆膜农田集中分布在研究区的北部和西南部,在中部、东部较为分散。在对 Google Earth 高分辨率影像进行目视解译的过程中发现,研究区存在大量白色地膜覆盖区域,且主要集中在西部地区。然而方案 S(图 6a)与方案 S + I(图 6d)获取的覆膜农田分布范围较小且分散,方案 I(图 6b)得到的覆膜农田分布中白膜和黑膜分布范围差异较大,同时白色地膜分布信息缺失较多,因此均存在严重的错分漏分现象。方案 T1(图 6c)、方案 S + I + T1(图 6e)以及方案 S + I + T2(图 6f)得到的分类结果差异不明显,提取的覆膜农田结构均较为明显,空间分布基本符合目视解译的结果,错分

漏分现象较轻。

对研究区内随机选中区域在不同方案下的分类结果的细节进行对比分析(图 7),可见与对应的 Google Earth 实景影像和 Landsat - 8 OLI 标准真彩色合成影像相比,6 种方案的分类结果在像元上均存在不同程度的误差。其中,方案 S(图 7c)、方案 S + I(图 7f)和方案 S + I + T1(图 7g)的分类结果与实际情况吻合程度相对较高,像元误差相对较小。

2.4 不同特征组合下随机森林分类精度对比

根据混淆矩阵计算的不同特征组合下的地物分类精度(表 2),可知 6 种不同特征组合下地物分类的总体精度均超过 80%。其中方案 S、方案 I 和方案 T1 这 3 种单一特征方法的总体精度分别为 92.57%、80.50% 和 87.62%, Kappa 系数分别为 0.91、0.76 和 0.84,表明光谱特征在影像分类过程中发挥着重要作用。与方案 S 相比,方案 S + I 中添加指数特征可有效提升分类精度,总体精度和 Kappa 系数分别提高 1.24 个百分点、0.01。此外,在“光谱 + 指数”特征的基础上加入纹理特征也可改善分类精度,与方案 S + I 相比,方案 S + I + T1 和

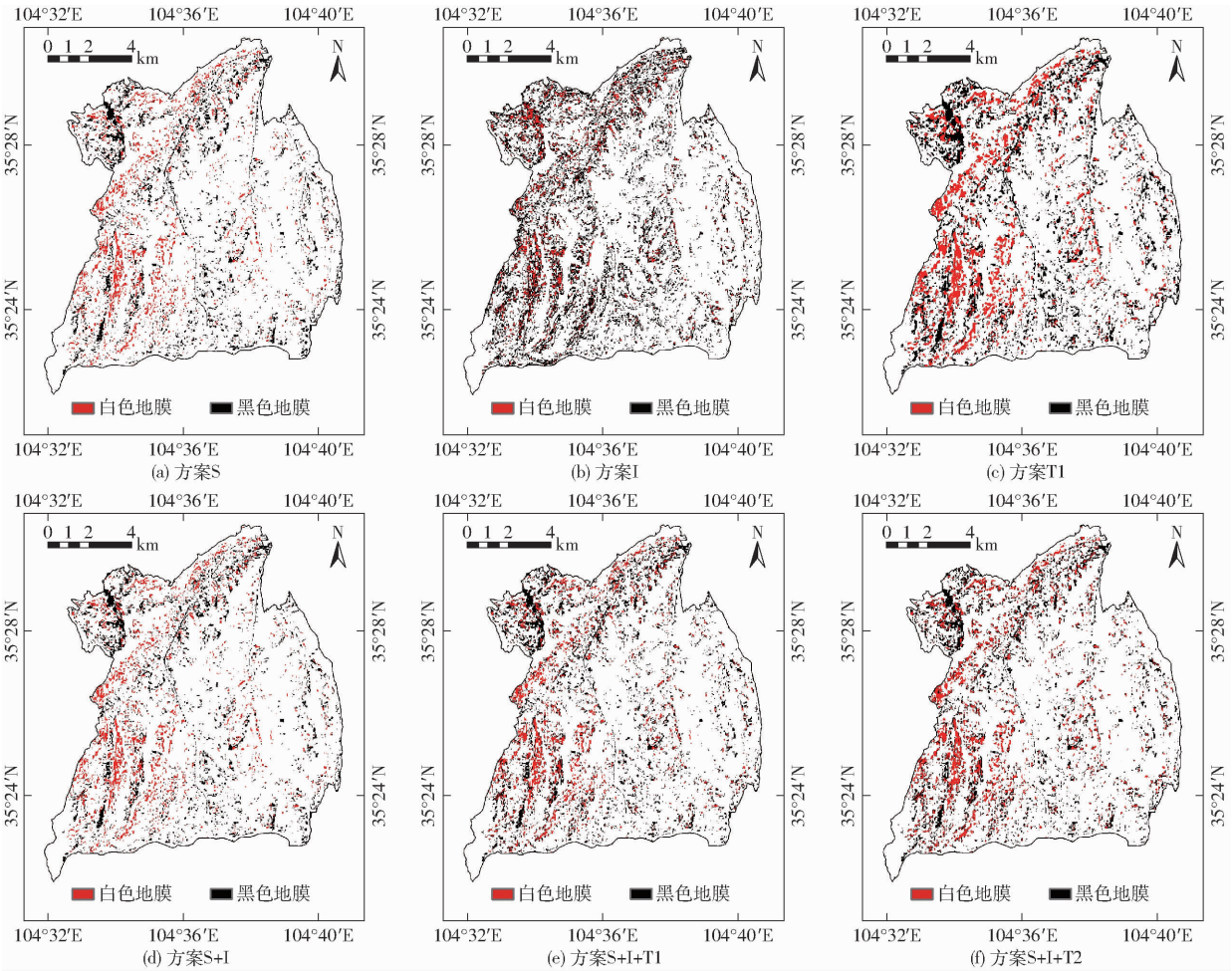


图 6 基于随机森林算法和不同特征组合下的研究区覆膜农田空间分布

Fig.6 Spatial distributions of plastic-film-mulched farmlands in study area based on random forest algorithm and different feature combinations

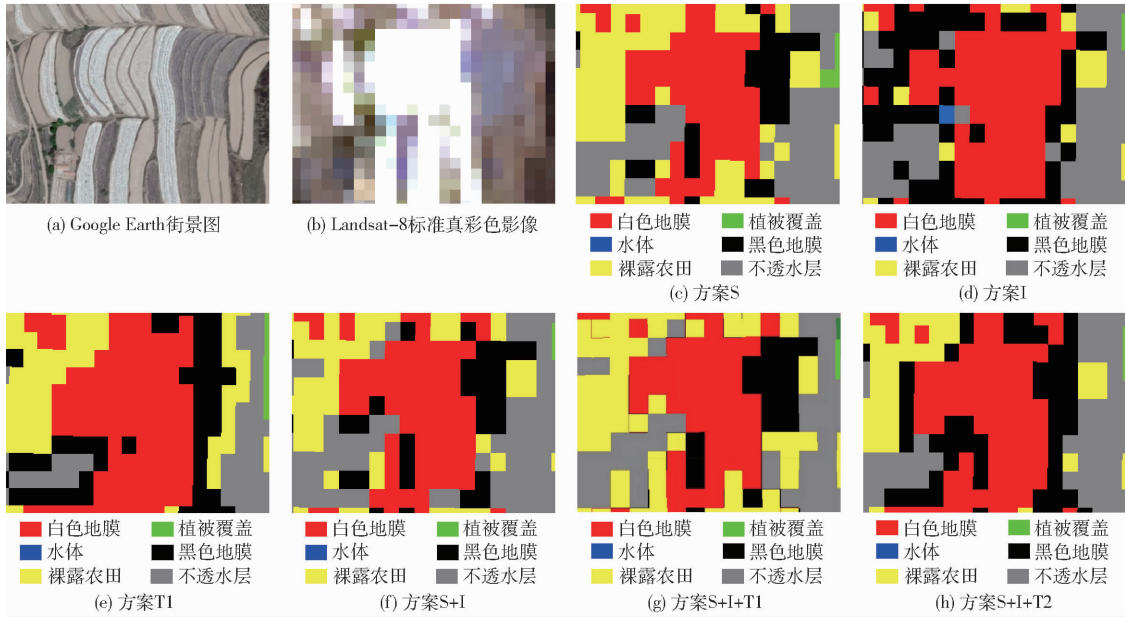


图 7 研究区某一随机覆膜农田区域在不同特征组合下的分类结果对比

Fig.7 Comparisons of classification results of plastic-film-mulched farmlands at random site in study area based on different feature combinations

方案 S + I + T2 的总体精度分别提高了 1.24、0.62 个百分点,Kappa 系数分别增加了 0.02 和 0.01。单从覆膜农田的分类精度上来看,无论是生产者精度还是用户精度,白膜都远高于黑膜,表明白膜的可分

表 2 基于不同特征组合和随机森林算法的研究区地物分类精度

Tab.2 Classification accuracies of land covers based on different feature combinations and random forest algorithm in study area

地物类型	方案 S		方案 I		方案 T1		方案 S + I		方案 S + I + T1		方案 S + I + T2	
	PA/%	UA/%	PA/%	UA/%	PA/%	UA/%	PA/%	UA/%	PA/%	UA/%	PA/%	UA/%
白色地膜	98.68	96.15	78.95	81.08	97.37	92.50	98.68	96.15	98.68	97.40	97.37	97.37
黑色地膜	86.25	95.83	57.50	67.65	90.00	82.76	83.75	94.37	92.50	91.36	91.25	92.41
植被覆盖	98.36	95.24	100	100	95.08	86.57	100	100	100	100	100	98.39
裸露农田	98.04	86.21	92.16	81.03	90.20	88.46	98.04	92.59	96.08	96.08	96.08	92.45
水体	100	100	66.67	80.00	100	100	100	100	100	100	100	100
不透水层	79.59	84.78	85.71	73.68	55.10	87.10	89.80	83.02	85.71	89.36	85.71	89.36
OA/%	92.57		80.50		87.62		93.81		95.05		94.43	
Kappa 系数	0.91		0.76		0.84		0.92		0.94		0.93	

离性优于黑膜。对不同特征组合下白膜的分类精度进行分析,可知方案 S、方案 S + I 以及方案 S + I + T1 的生产者精度最高为 98.68%;而用户精度在方案 S + I + T1 时达到最大值 97.40%。同样分析黑膜的分类精度,发现方案 S + I + T1 下生产者精度最高为 92.50%,方案 S 的用户精度最高为 95.83%。综合考虑分类性能和工作效率,确定本研究的最佳分类方案为 S + I + T1,即光谱 + 指数 + 优选纹理特征。

2.5 不同机器学习算法对覆膜农田识别精度的影响

在特征组合 S + I + T1 下,与 RF 算法(图 8a)相比,SVM(图 8b)、DT(图 8c)和 MDC(图 8d)算法的

分类结果差异明显,尤其是 DT 和 MDC 算法提取的覆膜农田没有从整体上呈现出其空间分布特征。进一步对比不同算法的分类精度(表 3),可知 RF 算法精度明显高于 SVM、DT 和 MDC 算法,总体精度分别高 3.10、7.74、50.78 个百分点,Kappa 系数分别高 0.04、0.10 和 0.62。由于 RF、SVM 和 DT 均为非参数监督分类方法,只有 MDC 属于参数监督分类方法,这表明非参数监督方法相比参数监督方法更适合用于地貌复杂地区的地物分类。单从覆膜农田的分类精度来看,在 RF 算法下,白膜的生产者精度和用户精度达到最大值,分别为 98.68% 和 97.40%;黑膜的生产者精度和用户精度也达到最大值,分别为 92.50% 和 91.36%。

表 3 方案 S + I + T1 下基于不同机器学习算法的研究区地物分类精度比较

Tab.3 Classification accuracies based on different classification algorithms of scheme S + I + T1 in study area

地物类型	RF		SVM		DT		MDC	
	PA/%	UA/%	PA/%	UA/%	PA/%	UA/%	PA/%	UA/%
白色地膜	98.68	97.40	98.68	92.59	97.37	96.10	47.37	81.82
黑色地膜	92.50	91.36	90.00	85.71	75.00	86.96	33.75	52.94
植被覆盖	100	100	98.36	98.36	93.44	100	32.79	41.67
裸露农田	96.08	96.08	92.16	97.92	86.27	84.62	72.55	38.95
水体	100	100	100	100	100	85.71	100	37.50
不透水层	85.71	89.36	75.51	86.05	83.67	67.21	34.69	24.64
OA/%	95.05		91.95		87.31		44.27	
Kappa 系数	0.94		0.90		0.84		0.32	

此外,根据 RF、SVM、DT 和 MDC 算法分类结果的 McNemar's 检验,可知 RF 算法与 SVM、DT、MDC 算法之间的 Z 值分别为 2.89、5.00 和 12.88,结果均大于 1.96,表明在 0.05 检验水平下 RF 算法与 SVM、DT、MDC 算法的分类精度差异性显著。综上所述,4 种机器学习算法中 RF 算法在地物分类和覆膜农田识别上优势明显,能够有效地提高识别精度和工作效率。

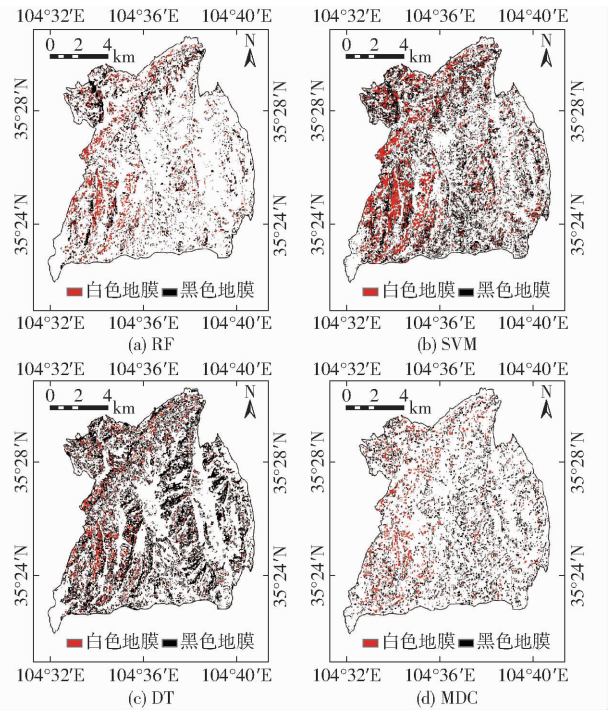


图 8 方案 S + I + T1 下基于不同机器学习算法的研究区覆膜农田空间分布

Fig.8 Spatial distributions of plastic-film-mulched farmlands based on scheme S + I + T1 and different classification algorithms in study area

3 讨论

本研究提出了基于 GEE 平台快速提取覆膜农田的方法,获得了甘肃省定西市安定区团结镇的覆膜农田空间分布情况,识别精度和效率比已有相关研究有所提升。LU 等^[40]基于 Landsat 系列数据开展新疆地区的覆膜农田识别工作,总体精度和 Kappa 系数分别达到 97.82% 和 0.97,由于新疆地区覆膜农田几乎呈连片分布状,在一定程度上降低了工作难度。哈斯图亚^[41]选择了同样位于黄土高原的宁夏固原地区作为研究区进行覆膜农田提取,总体精度和 Kappa 系数为 90.67% 和 0.87。本文地物分类的总体精度和 Kappa 系数分别为 95.05% 和 0.94,实现了地膜分布相对破碎地区的精准识别,可为地貌复杂地区应用农业遥感技术提供参考。此外,研究同时考虑了白膜和黑膜覆盖农田,并成功对其分别进行提取,这也弥补了 XIONG 等^[42]在其研究结果分析中提出的未考虑研究区中的黑膜而导致精度降低这一缺陷。

目前大多数研究对于随机森林算法关键参数的选取多参考经验值,如哈斯图亚^[41]、张鹏等^[43]根据经验将决策树数量 T 设为 500、节点分裂特征个数 M 取输入特征数的平方根;侯蒙京等^[44]参考大量研究将参数 T 设为 1 000、参数 M 保持默认,这在一定程度上限制了 RF 算法高精度优势的发挥。本研究首先根据参数 T 和 M 的经验值确定了它们的取值范围,然后遍历所有参数组合来执行 RF 算法分类,最终选择运算效率可接受范围内使袋外误差最低的参数值,作为下一步研究中 RF 算法的输入参数。结果表明,在利用 RF 算法进行影像分类时,参数 T 和 M 均对袋外误差有明显影响,且参数 T 的影响大于参数 M 。当参数 M 一定时,并非参数 T 越大,袋外误差就越小,因此通过不断提高参数 T 的值来降低袋外误差往往是无法实现的。这与 RODRIGUEZ-GALIANO 等^[32]的研究结果相近。本研究中,参数 T 取 600、参数 M 保持默认值的组合可在运算效率承受范围内使袋外误差最低。这表明 RF 算法关键参数的优化是提高覆膜农田识别精度的有效途径。

遥感识别特征决定了各类地物之间的相似性和差异性,是判读识别各类地物的依据。本研究基于 RF 算法对纹理特征进行重要性分析,按照重要性排序逐一添加纹理特征并进行分类精度验证,确定前 15 个特征为优选纹理特征,并结合光谱和指数特征制定了 6 套方案进行覆膜农田识别。结果表明,基于“光谱 + 指数 + 优选纹理”特征的总体精度和 Kappa 系数最高,分别为 95.05% 和 0.94。这与朱

秀芳等^[45]的研究结果相近,该研究认为纹理特征越多并不代表分类精度越高,分类前进行特征优选是必要工作。然而该研究仅依靠优选纹理特征进行分类,总体精度和 Kappa 系数分别为 94.84% 和 0.89,并提出可通过增加指数特征来改善分类精度。本研究针对其提出的不足之处加以改进,明显提高了覆膜农田的识别精度。

作为目前机器学习领域的研究热点,随机森林被广泛应用于地物分类研究中。本研究基于“光谱 + 指数 + 优选纹理”特征评价不同算法的分类性能时发现,无论是从生产者精度、用户精度、总体精度和 Kappa 系数还是 McNemar's 检验结果来看,随机森林算法均显著优于其他算法,这与哈斯图亚^[41]、侯蒙京等^[44]的研究结果相符。沙先丽^[12]基于 Landsat 系列数据,利用决策树模型和多特征结合方法提取新疆地区的覆膜农田信息,但是在研究区地膜覆盖连片、地物特征明显的前提下分类的总体精度和 Kappa 系数仅为 91.35% 和 0.89。究其原因,在所用数据和特征子集相同的情况下,关键因素可能在于所采用的算法。这也再次印证了随机森林算法在覆膜农田识别中的有效性。

本研究主要考虑的是春玉米、马铃薯等春播作物地膜覆盖种植区域的单一时相识别,尽管其种植面积占全年覆膜农田面积的 90% 左右,但仍然缺失其他季节的覆膜农田信息。在其他季节,研究区主要的覆膜作物为蔬菜,而蔬菜的播种时间几乎取决于农户本身,不确定因素很多,因此对覆膜蔬菜种植区域进行提取难度较大,有待后续进一步深入研究。此外,本研究仅考虑了基于像元的遥感影像分类方法,而裴欢等^[46]的研究表明面向对象分类方法为土地利用/覆盖信息提取提供了新的有效途径,该方法在覆膜农田识别中的应用潜力还需要进一步深入探讨。

4 结论

(1) 确定随机森林算法关键参数的最佳取值能够大幅提高遥感影像的分类精度。在非关键参数保持默认的条件下,在一定阈值范围内存在决策树数量 T 和节点分裂特征个数 M 的最佳参数组合,能够在运算效率可承受范围内使袋外误差最小。因此,在利用随机森林算法实现影像分类之前,对其关键参数进行了优化,以获取更为可靠的分类结果。

(2) 基于优选纹理特征的分类性能优于基于全部纹理特征的分类,且相比其他方案,基于“光谱 + 指数 + 优选纹理”特征的识别结果最佳。基于“光谱 + 指数 + 优选纹理”特征分类的总体精度达

95.05%，比基于单一特征方案的总体精度高 2.48 ~ 14.55 个百分点，同时也比“光谱 + 指数”特征、“光谱 + 指数 + 全部纹理”特征方案分别高 1.24、0.62 个百分点。因此，在实际应用中建议采用优选纹理特征和多特征相结合的方法进行覆膜农田识别。

(3) 随机森林算法的总体精度比支持向量机、

决策树和最小距离分类算法分别高 3.10、7.74、50.78 个百分点。结合 McNemar’s 检验结果，随机森林算法与其它 3 种算法间的 Z 值分别为 2.89、5.00 和 12.88，表明随机森林算法与其它算法间的分类精度存在显著性差异。因此，随机森林算法是较适于覆膜农田识别的方法。

参 考 文 献

[1] 林文. 地膜和秸秆覆盖对黄土高原旱作农田土壤水库与作物产量的影响[D]. 杨凌: 西北农林科技大学, 2017.
LIN Wen. Influence of plastic mulch and straw mulch on soil reservoir and crop yield on the Loess Plateau[D]. Yangling: Northwest A&F University, 2017. (in Chinese)

[2] 逯亚杰. 黄土高原地区气候生产潜力估算及其对气候变化的响应[D]. 杨凌: 西北农林科技大学, 2016.
LU Yajie. Evaluating the climate potential productivity and its response to climate change in the Loess Plateau[D]. Yangling: Northwest A&F University, 2016. (in Chinese)

[3] AUSTIN A T. Has water limited our imagination for aridland biogeochemistry? [J]. Trends in Ecology and Evolution, 2011, 26(5): 229 – 235.

[4] ZAND-PARSA S, SEPASKHAH A, RONAGHI A. Development and evaluation of integrated water and nitrogen model for maize [J]. Agricultural Water Management, 2006, 81(3): 227 – 256.

[5] YU Yueyuan, TURNER N C, GONG Yanhong, et al. Benefits and limitations to straw-and plastic-film mulch on maize yield and water use efficiency: a meta-analysis across hydrothermal gradients[J]. European Journal of Agronomy, 2018, 99(3): 138 – 147.

[6] LI Fengmin, WANG Jun, XU Jinzhang, et al. Productivity and soil response to plastic film mulching durations for spring wheat on entisols in the semiarid Loess Plateau of China[J]. Soil and Tillage Research, 2004, 78(1): 9 – 20.

[7] HAN Juan, JIA Zhikuan, WU Wei, et al. Modeling impacts of film mulching on rainfed crop yield in Northern China with DNDIC[J]. Field Crops Research, 2014, 155: 202 – 212.

[8] LIU Enke, HE Wenqing, YAN Changrong. ‘White revolution’ to ‘white pollution’—agricultural plastic film mulch in China [J]. Environmental Research Letters, 2014, 9(9): 091001.

[9] 陈仲新, 任建强, 唐华俊, 等. 农业遥感研究应用进展与展望[J]. 遥感学报, 2016, 20(5): 748 – 767.
CHEN Zhongxin, REN Jianqiang, TANG Huajun, et al. Progress and perspectives on agricultural remote sensing research and applications in China[J]. Journal of Remote Sensing, 2016, 20(5): 748 – 767. (in Chinese)

[10] MOREAU S, BOSSENO R, GU Xingfa, et al. Assessing the biomass dynamics of Andean bofedal and totora high-protein wetland grasses from NOAA/AVHRR[J]. Remote Sensing of Environment, 2003, 85(4): 516 – 529.

[11] OU Cong, YANG Jianyu, DU Zhenrong, et al. Long-term mapping of a greenhouse in a typical protected agricultural region using landsat imagery and the Google Earth Engine[J]. Remote Sensing, 2020, 12(1): 55.

[12] 沙先丽. 地膜农田遥感信息提取及覆膜地表温度反演[D]. 杭州: 浙江大学, 2012.
SHA Xianli. A study on extraction of plastic-mulched landcover and on retrieval of mulched land surface temperature using remote sensing data[D]. Hangzhou: Zhejiang University, 2012. (in Chinese)

[13] HASITUYA, CHEN Zhongxin, WANG Limin, et al. Monitoring plastic-mulched farmland by Landsat – 8 OLI imagery using spectral and textural features[J]. Remote Sensing, 2016, 8(4): 353.

[14] 宁纪锋, 倪静, 何宜家, 等. 基于卷积注意力的无人机多光谱遥感影像地膜农田识别[J]. 农业机械学报, 2021, 52(9): 213 – 220.
NING Jifeng, NI Jing, HE Yijia, et al. Convolutional attention based plastic mulching farmland identification via UAV multispectral remote sensing image[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(9): 213 – 220. (in Chinese)

[15] YANG Yanmin, ZHOU Xinyao, YANG Yonghui, et al. Evaluating water-saving efficiency of plastic mulching in Northwest China using remote sensing and SEBAL[J]. Agricultural Water Management, 2018, 209(9): 240 – 248.

[16] GORELICK N, HANCHER M, DIXON M, et al. Google Earth Engine: planetary-scale geospatial analysis for everyone[J]. Remote Sensing of Environment, 2017, 202: 18 – 27.

[17] DONG Jinwei, XIAO Xiangming, MENARGUEZ M A, et al. Mapping paddy rice planting area in northeastern Asia with Landsat 8 images, phenology-based algorithm and Google Earth Engine[J]. Remote Sensing of Environment, 2016, 185: 142 – 154.

[18] PEKEL J F, COTTAM A, GORELICK N, et al. High-resolution mapping of global surface water and its long-term changes [J]. Nature, 2016, 540: 418 – 422.

[19] SAZIB N, MLADENOVA I, BOLTEN J. Leveraging the google earth engine for drought assessment using global soil moisture data[J]. Remote Sensing, 2018, 10(8): 1265.

[20] TELUGUNTLA P, THENKABAIL P, OLIPHANT A, et al. A 30-m landsat-derived cropland extent product of Australia and China using random forest machine learning algorithm on Google Earth Engine cloud computing platform[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2018, 144(38): 325 – 340.

[21] 何昭欣, 张森, 吴炳方, 等. Google Earth Engine 支持下的江苏省夏收作物遥感提取[J]. 地球信息科学学报, 2019, 21(5): 752 – 766.

- HE Zhaoxin, ZHANG Miao, WU Bingfang, et al. Extraction of summer crop in Jiangsu based on Google Earth Engine[J]. *Journal of Geo-information Science*, 2019, 21(5): 752–766. (in Chinese)
- [22] 甘肃省统计局, 国家统计局甘肃调查总队. 甘肃发展年鉴[M]. 北京: 中国统计出版社, 2019.
- [23] LIN Wen, LIU Wenzhao, ZHOU Shanshan, et al. Influence of plastic film mulch on maize water use efficiency in the Loess Plateau of China[J]. *Agricultural Water Management*, 2019, 224: 105710.
- [24] HOU Xianqing, LI Rong, HE Wenshou, et al. Effects of planting density on potato growth, yield, and water use efficiency during years with variable rainfall on the Loess Plateau, China[J]. *Agricultural Water Management*, 2020, 230: 105982.
- [25] TUCKER C J. Red and photographic infrared linear combinations for monitoring vegetation [J]. *Remote Sensing of Environment*, 1979, 8(2): 127–150.
- [26] MCFEETERS S K. The use of the normalized difference water index (NDWI) in the delineation of open water features[J]. *International Journal of Remote Sensing*, 1996, 17(7): 1425–1432.
- [27] ZHA Yong, JAY G, NI Shaoxiang. Use of normalized difference built-up index in automatically mapping urban areas from TM imagery[J]. *International Journal of Remote Sensing*, 2003, 24(3): 583–594.
- [28] HARALICK R M, SHANMUGAM K, DINSTEN I H. Textural features for image classification[J]. *IEEE Transactions on Systems, Man, and Cybernetics*, 1973, 3(6): 610–621.
- [29] BREIMAN L. Random forests[J]. *Machine Learning*, 2001, 45(1): 5–32.
- [30] GAO Tian, ZHU Jiaojun, ZHENG Xiao, et al. Mapping spatial distribution of larch plantations from multi-seasonal Landsat-8 OLI imagery and multi-scale textures using random forests[J]. *Remote Sensing*, 2015, 7(2): 1702–1720.
- [31] BELGIU M, DRĂGUT L. Random forest in remote sensing: a review of applications and future directions[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2016, 114: 24–31.
- [32] RODRIGUEZ-GALIANO V F, GHIMIRE 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.
- [33] 刘杰, 刘吉凯, 安晶晶, 等. 基于时序 Landsat 8 OLI 多特征与随机森林算法的作物精细分类研究[J]. *干旱地区农业研究*, 2020, 38(3): 281–288, 298.
- LIU Jie, LIU Jikai, AN Jingjing, et al. Precise crop classification based on multi-features from time-series Landsat 8 OLI images and random forest algorithm [J]. *Agricultural Research in the Arid Areas*, 2020, 38(3): 281–288, 298. (in Chinese)
- [34] MOUNTRAKIS G, IM J, OGOLE C. Support vector machines in remote sensing: a review [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2011, 66(3): 247–259.
- [35] BREIMAN L, FRIEDMAN J, OLSHEN R, et al. *Classification and regression trees*[M]. New York: Chapman & Hall/CRC, 1984.
- [36] 张若琳, 万力, 张发旺, 等. 土地利用遥感分类方法研究进展[J]. *南水北调与水利科技*, 2006, 4(2): 39–42.
- ZHANG Ruolin, WAN Li, ZHANG Fawang, et al. Development of research on remote sensing classification technology for land use and cover[J]. *South-to-North Water Transfers and Water Science & Technology*, 2006, 4(2): 39–42. (in Chinese)
- [37] CONGALTON R G. A review of assessing the accuracy of classifications of remotely sensed data[J]. *Remote Sensing of Environment*, 1991, 37(1): 35–46.
- [38] 赵英时. 遥感应用分析原理与方法[M]. 北京: 科学出版社, 2013.
- [39] FOODY G M. Classification accuracy comparison: hypothesis tests and the use of confidence intervals in evaluations of difference, equivalence and non-inferiority[J]. *Remote Sensing of Environment*, 2009, 113(8): 1658–1663.
- [40] LU Lizhen, DI Liping, YE Yanmei. A decision-tree classifier for extracting transparent plastic-mulched landcover from Landsat-5 TM images[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2014, 7(11): 4548–4558.
- [41] 哈斯图亚. 基于多源数据的地膜覆盖农田遥感识别研究[D]. 北京: 中国农业科学院, 2017.
- HASITUYA. Mapping plastic-mulched farmland with multi-source remote sensing data [D]. Beijing: Chinese Academy of Agricultural Sciences, 2017. (in Chinese)
- [42] XIONG Yuankang, ZHANG Qingling, CHEN Xi, et al. Large scale agricultural plastic mulch detecting and monitoring with multi-source remote sensing data: a case study in Xinjiang, China[J]. *Remote Sensing*, 2019, 11(18): 2088.
- [43] 张鹏, 胡守庚. 地块尺度的复杂种植区作物遥感精细分类[J]. *农业工程学报*, 2019, 35(20): 125–134.
- ZHANG Peng, HU Shougeng. Fine crop classification by remote sensing in complex planting areas based on field parcel[J]. *Transactions of the CSAE*, 2019, 35(20): 125–134. (in Chinese)
- [44] 侯蒙京, 殷建鹏, 葛静, 等. 基于随机森林的高寒湿地地区土地覆盖遥感分类方法[J]. *农业机械学报*, 2020, 51(7): 220–227.
- HOU Mengjing, YIN Jianpeng, GE Jing, et al. Landcover remote sensing classification method of alpine wetland region based on random forest algorithm[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2020, 51(7): 220–227. (in Chinese)
- [45] 朱秀芳, 李石波, 肖国峰. 基于无人机遥感影像的覆膜农田面积及分布提取方法[J]. *农业工程学报*, 2019, 35(4): 106–113.
- ZHU Xiufang, LI Shibo, XIAO Guofeng. Method on extraction of area and distribution of plastic-mulched farmland based on UAV images[J]. *Transactions of the CSAE*, 2019, 35(4): 106–113. (in Chinese)
- [46] 裴欢, 孙天娇, 王晓妍. 基于 Landsat 8 OLI 影像纹理特征的面向对象土地利用/覆盖分类[J]. *农业工程学报*, 2018, 34(2): 248–255.
- PEI Huan, SUN Tianjiao, WANG Xiaoyan. Object-oriented land use/cover classification based on texture features of Landsat 8 OLI image[J]. *Transactions of the CSAE*, 2018, 34(2): 248–255. (in Chinese)