

doi:10.6041/j.issn.1000-1298.2021.07.020

基于 Stacking 集成学习的夏玉米覆盖度估测模型研究

张宏鸣¹ 陈丽君¹ 刘雯¹ 韩文霆² 张姝茵¹ 张凡¹

(1. 西北农林科技大学信息工程学院, 陕西杨凌 712100; 2. 西北农林科技大学机械与电子工程学院, 陕西杨凌 712100)

摘要: 以基于无人机多光谱影像提取的夏玉米植被指数作为特征变量, 利用皮尔森相关系数结合随机森林反向验证权重的方法进行特征选择, 去除冗余特征。以随机森林、梯度提升树、支持向量机和岭回归作为初级学习器, 以岭回归作为次级学习器, 建立基于 Stacking 集成学习的夏玉米覆盖度估测模型, 并通过 5 折交叉验证进一步提升模型泛化能力, 采用随机搜索和网格搜索结合的方法对模型超参数进行优化, 使用 4 种回归指标进行模型精度评价, 并利用次年数据验证其鲁棒性。结果表明, 与单一模型以及决策树、Xgboost、Adaboost、Bagging 集成框架相比, Stacking 集成学习模型具有更高的精度和更强的鲁棒性, R^2 为 0.950 9, 比单一模型平均提升 0.036 9, 比其他集成模型平均提升 0.041 7; Stacking 集成学习模型 RMSE、MAE 和 MAPE 分别为 0.043 2、0.033 0 和 5.01%, 各指标分别比单一模型平均降低 0.013 8、0.013 0 和 2.14 个百分点, 分别比其他集成模型平均降低 0.018 5、0.012 6 和 2.15 个百分点。本研究为夏玉米覆盖度估测提供了新的方法。

关键词: 夏玉米; 植被覆盖度; Stacking 集成学习; 无人机; 多光谱影像

中图分类号: TP391.4

文献标识码: A

文章编号: 1000-1298(2021)07-0195-08

OSID:



Estimation of Summer Corn Fractional Vegetation Coverage Based on Stacking Ensemble Learning

ZHANG Hongming¹ CHEN Lijun¹ LIU Wen¹ HAN Wenting² ZHANG Shuyin¹ ZHANG Fan¹

(1. College of Information Engineering, Northwest A&F University, Yangling, Shaanxi 712100, China

2. College of Mechanical and Electronic Engineering, Northwest A&F University, Yangling, Shaanxi 712100, China)

Abstract: Based on the UAV multi-spectral image, the summer corn vegetation index was extracted as a feature variable, and the Pearson correlation coefficient combined with the random forest algorithm was used to reverse the verification weight method for feature selection and redundant features were removed. Random forest, gradient boosting tree, support vector machine and ridge regression were used as the primary learner, and ridge regression was used as the secondary learner to establish a summer corn coverage estimation model based on Stacking ensemble learning, and 5-fold cross-validation was used to further improve model generalization ability, a combination of random search and grid search was used to optimize model hyper parameters, four regression indicators were used for model accuracy evaluation, and the following year's data was used to verify its robustness. The experimental results showed that compared with a single model and decision tree, Xgboost, Adaboost, and Bagging integrated framework, the Stacking integrated learning model had higher accuracy and stronger robustness. The R^2 was 0.950 9, which was an average improvement of 0.036 9 than that of the single model. Compared with other integrated models, the average increase was 0.041 7; RMSE, MAE and MAPE were 0.043 2, 0.033 0 and 5.01%, respectively, which were 0.013 8, 0.013 0 and 2.14 percentage points lower than that of the single model, and 0.018 5, 0.012 6 and 2.15 percentage points lower than that of the other integrated models. The research result provided a method and effective support for the estimation of summer corn coverage.

Key words: summer corn; fractional vegetation cover; Stacking ensemble learning; unmanned aerial vehicle; multi-spectral image

收稿日期: 2020-08-20 修回日期: 2020-09-27

基金项目: 国家重点研发计划项目(2020YFD1100601)、宁夏智慧农业产业技术协同创新中心项目(2017DC53)、国家自然科学基金项目(41771315)和宁夏回族自治区重点研发计划项目(2017BY067)

作者简介: 张宏鸣(1979—),男,教授,博士,主要从事空间大数据管理与智慧农业研究, E-mail: zhm@nwsuaf.edu.cn

0 引言

玉米是我国重要的粮食作物,为了保证我国玉米产业的快速平稳发展,需要对玉米相关生长参数进行有效监测。植被覆盖度(Fractional vegetation cover, FVC)通常指植被(包括叶、茎、枝)在地面的垂直投影面积占统计区总面积的百分比^[1-2]。植被覆盖度反映植物进行光合作用的面积以及植被生长的茂盛程度,在一定程度上能够体现植被的生长状态和生长趋势^[3-6]。因此,估测玉米覆盖度对指导生产管理和进行农业生态系统评价具有重要意义^[7]。

植被覆盖度测量主要有地表实测法和遥感监测法^[8-9],由于受到人力物力的限制,地表实测法已逐渐被遥感检测法替代^[10-11]。植被覆盖度的遥感估算方法主要包括混合像元分解法、统计模型法和数据挖掘法^[12]。混合像元分解法一般应用于缺乏地面实测数据的情况^[13],主要有线性模型、概率模型、几何光学模型、随机几何模型和模糊分析模型等,其中线性模型应用最为广泛^[14],混合像元分解法要求植被具有较为明显的区域特点,且估测精度不高^[15]。统计模型法主要利用遥感数据的某一波段、波段组合或遥感植被指数与地面 FVC 测量值进行统计分析,建立两者之间的数值模型,可分为线性回归模型和非线性回归模型,前者应用较为广泛^[16-17]。统计模型法简单、容易实现,但只适用于特定区域与特定植被类型,且精度依赖于地面实测数据,对遥感数据空间分辨率要求高,大区域植被覆盖度估算可能失效,具有较大的不确定性和局限性,不宜推广^[12]。数据挖掘法是从大量不完整且含有噪声的随机观测数据中估测潜在信息和知识的过程,目前,决策树、支持向量机和随机森林等非参数回归算法已在植被覆盖度反演中得到应用^[12]。支持向量机在解决小样本、非线性及高维模式识别等问题中表现出很好的泛化能力,近年来也应用于植被覆盖度的估算中^[18];决策树可避免数据冗余,提高计算能力,有效地挖掘数据潜力;与参数回归等方法相比,随机森林模型无需对变量的正态性和独立性等假设条件进行检验,估测效果较好^[19]。

大多数研究使用单一机器学习方法估测植被覆盖度,其中线性回归对非线性以及特征空间大的数据性能表现不足,传统的机器学习模型也容易陷入过拟合。过度依赖一种学习模型很难全面考虑覆盖度变化的特点,从而无法避免估测精度不高、模型鲁棒性不强的问题。

基于以上问题,本文建立基于 Stacking 集成学

习的夏玉米覆盖度估测模型,利用无人机获取夏玉米的多光谱影像,使用随机森林算法进行特征选择,去除冗余特征,集成随机森林、梯度提升树、支持向量机、岭回归等 4 种学习器,并与其他集成学习框架以及单一学习器模型进行对比,以验证模型的有效性。

1 材料与方法

1.1 研究区概况

研究区位于内蒙古自治区鄂尔多斯市达拉特旗昭君镇,地势南高北低,南部为丘陵山区,库布其沙漠沿腹部穿过,北部为肥沃的黄河冲积平原。地块中心地理坐标为(40°25'N, 109°36'E),实验区面积约为 1.13 hm²,属典型的温带大陆性气候,大陆度指数为 75%,干燥少雨,冬寒夏热,昼夜温差大,年均日照时数约 3 000 h,年均气温 6.1~7.1℃,无霜期 135~150 d,年均降水量 240~360 mm。粮食作物主要以小麦、玉米为主,本实验区域种植作物类型为夏玉米,属于春播中晚熟期,一年一熟制,播种时间为 5 月中旬,收获时间为 9 月。由于当前的预期量无法满足玉米生长的需求水量,主要的供水方式为喷灌机灌水。实验研究区无人机可见光影像示意图如图 1 所示,圆形实验区域以喷灌机为中心,平均分为 5 个扇区,并在每个扇区内部等距离处设置 3 个 6 m×6 m 的实验样方(绿色正方形标记),实验区域共计 15 个样方。

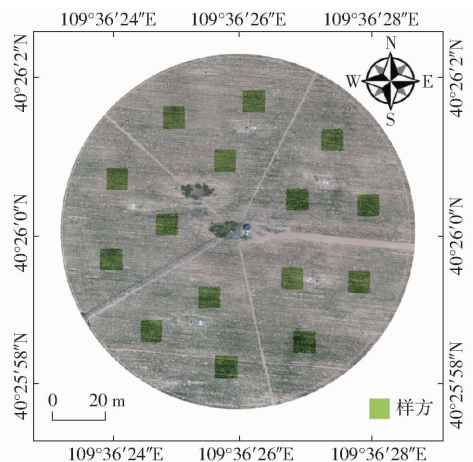


图 1 研究区无人机影像示意图

Fig. 1 UAV image in research area

1.2 数据获取及处理

实验时间为 2018 年 6 月 26 日—8 月 23 日,共采集子空间数据 117 条。遥感数据的采集包括多光谱影像以及在同一时间内低空拍摄的可见光图像。多光谱图像利用六翼无人机搭载五通道多光谱相机 RedEdge 进行拍摄,无人机轴距 900 mm,速度 5 m/s,相

机质量 150 g, 焦距 5.5 mm, 47.2° 水平视场, 5 个谱带对应的中心波长分别为蓝(475 nm)、绿(560 nm)、红(668 nm)、近红外(840 nm)以及红边(717 nm), 其他参数如表 1 所示。

表 1 多光谱相机主要参数

Tab. 1 Main parameters of multispectral camera

参数	数值
质量/g	150
尺寸/(cm × cm × cm)	12.1 × 6.6 × 4.6
电压/V	4.2 ~ 15.8
功耗/W	4(正常)/8(最大)
最快捕获速度/(bit · s ⁻¹)	12

利用大疆精灵 4Pro 型无人机进行低空高清可见光图像采集, 光圈值为 $f/5.6$, 曝光时间为 $1/1\ 600\ s$, 焦距为 9 mm。为避免光照条件不均匀不一致产生的干扰, 实验天气需晴朗无云, 选取太阳光辐射强度稳定的 11:00—13:00, 相机镜头垂直向下进行数据采集。无人机需按照固定航线飞行。其中多光谱图像分辨率为 1 280 像素 × 960 像素, 单次采集数据图像 1 780 幅, 可见光图像分辨率为 5 472 像素 × 3 648 像素, 单次采集数据图像 15 幅, 同时配备光强传感器以及具有固定反射率 3 m × 3 m 的灰板。

将无人机搭载 RedEdge 采集的 TIFF 格式图像, 以日期为索引导入到 Pix4Dmapper 软件中, 导入位置与姿态系统 (Position and orientation system, POS) 数据, 采用 CGCS2000_3_Degree_GK_CM_111E, D_China_2000 坐标系, 导入地面控制点 (Ground control point, GCP) 文件, 样区范围共布设 5 个地面控制点, 使用实时动态定位 (Real-time kinematic, RTK) 进行精度检测。同时用这些点来检测影像集合定位精度, 保证校正影像在作物覆盖度估测中的基本应用需求。在软件中进行初始化处理, 几何校正, 构建三维模型, 提取纹理以及构造地物特征, 根据不同要求设置参数进行图像拼接, 得到的多光谱影像空间分辨率为 4.731 cm。将生成的高清多光谱影像导入 ENVI 软件, 利用不同植被指数的计算公式分别生成对应植被指数的灰度图, 并在 ENVI 中对拼接预处理之后的原始图像进行裁剪, 将根据可见光影像制作的 XML 实测区域像控点文件导入, 每一块实测区域对应实验样地的每一小样方, 取裁剪后每一小样方内全部像素点的各项植被指数, 最终以均值来确定每个小样方对应的各项植被指数^[20-22]。

将同步采集的高清可见光遥感图像作为夏玉米覆盖度参考值, 且在 Photoshop 软件中求算夏玉米覆盖度, 将拍摄的 JPG 格式图像导入软件中, 按不同图像需求人工选择色彩范围并不断调节容差将植株

和背景分离开来, 最终以夏玉米所占像素数与整个小样方的总像素数之比作为该小样方的覆盖度^[23-24]。

1.3 特征变量提取及选择

植被的反射光谱为植被本身、下垫面 (土质、土壤亮度、湿度、颜色等)、阴影和环境等的综合表现, 且其易受大气和季节等的影响^[25]。由于植被和农作物在红光波段强吸收和在近红外波段强反射的特性, 大量的研究证明这两个波段与作物覆盖度有很好的相关关系。因此, 在综合考虑各遥感卫星光谱范围差异的基础上, 结合本研究所使用的遥感影像以及研究区域的特点, 选择红光、蓝光、绿光和近红外波段组合出 9 种植被指数作为特征变量进行植被覆盖度研究。分别为比值植被指数 (RVI)^[26]、归一化差值植被指数 (NDVI)^[27]、土壤调节植被指数 (SAVI)^[28]、改进的土壤调节植被指数 (MSAVI)^[29]、增强植被指数 (EVI)^[30]、优化土壤调节植被指数 (OSAVI)^[31]、重归一化植被指数 (RDVI)^[32]、差值植被指数 (DVI)^[33] 和三角植被指数 (TVI)^[34]。

上述特征集可能存在信息冗余或泛化性较差的问题, 因此需要进行特征选择, 找出与植被覆盖度高度相关的特征变量, 选出能够充分估测植被覆盖度且数目最少的特征变量集。本文综合考虑皮尔森相关系数与随机森林反向验证权重因子, 避免单一线性相关评价的局限性, 获得综合考量重要性指标, 提取出精要项特征, 并对上述筛选方法最终精简效果进行回归验证。

1.4 模型构建

1.4.1 Stacking 集成学习模型

Stacking 由 WOLPERT^[35] 提出, 其实质上是一种串行结构的多层学习系统。Stacking 方法前期使用交叉验证的方式把原始特征转换为二级特征, 然后再对变换得到的二级特征运用元学习器进行常规的训练和拟合。不同于传统的集成框架引导聚集算法 (Bagging) 和提升方法 (Boosting), Stacking 框架是将不同基础学习器组合起来进行模型融合。

训练过程为: ①利用 Stacking 集成学习方法调用不同类型的学习器对数据集进行训练学习。②将各分类器得到的训练结果组成一个新的训练样例作为元分类器的输入。③最终第 2 层模型中元学习器的输出结果为最终的结果输出^[36-37]。

1.4.2 随机森林

随机森林 (Random forest, RF) 是一种监督学习算法, 在每棵决策回归树进行分支时, 从总特征集的子空间中选取最优特征进行分支。该方法保证了各

决策树的独立性和多样性,一定程度上避免了过拟合。

1.4.3 梯度提升树

梯度提升树(Gradient boosting decision tree, GBDT)是基于 Boosting 算法的一种改进。该方法既能处理连续值,也能处理离散值,对各种类型的数据适应能力极强。具有可灵活处理各种数据、估测准确率高、使用损失函数和对异常值具有很强的鲁棒性等优点,可有效进行回归估测。

1.4.4 支持向量机

支持向量机(Support vector machine, SVM)可将低维数据映射到高维空间中,使低维空间的数据在高维空间中形成估计函数,实现回归模型的精度和计算复杂度间的平衡。该方法对于解决小样本分类、回归类问题效果理想,特别适用于时间序列等数据分析挖掘,具有较强的泛化性能。

1.4.5 岭回归

岭回归(Ridge regression, RR)实质上是通过改良最小二乘法,放弃最小二乘法的无偏性产生有偏估计,允许以损失部分信息、降低精度为代价获得更实际更可靠的回归系数。该方法通过损失无偏性,对最小二乘回归进行补充,换取较高的数值稳定性,从而得到较高的计算精度,且建模速度快,不需要复杂的计算过程。

1.4.6 夏玉米覆盖度估测模型

Stacking 集成方法可以集成多种学习器的估测结果。考虑基学习器的预测能力,本文在 Stacking 集成学习模型第 1 层选择学习能力较强和差异度较大的模型作为基学习器,有助于模型整体的预测效果提升。其中,随机森林和梯度提升树分别采用 Bagging 和 Boosting 的集成学习方式,具有出色的学习能力和严谨的数学理论。支持向量机对于解决小样本和非线性以及高维度的回归问题具有优势。岭回归算法因其稳定性强和训练高效等特点,有着良好的实际应用效果。第 2 层元学习器选择泛化能力较强的模型,用于纠正多个学习算法对于训练集的偏置情况,防止过拟合效应出现。综上所述,经多次实验组合,本文最终采用 RF_1、RF_2、GBDT、SVM 和 Ridge 等 5 个模型作为基学习器,利用岭回归作为元学习器进行融合得到集成模型。其中 RF_1 和 RF_2 均使用 RF 算法进行训练,两者仅参数设置不同。通过对多种算法的结合,克服单个模型缺陷,优化岭回归的输入,提升了估测结果。本文提出的 Stacking 集成学习夏玉米覆盖度模型框架(图 2),算法步骤如下:

(1)输入数据集。获取初始数据集,并对数据

进行特征获取及选择,得到数据集 $T = \{(y_i, x_i); i = 1, 2, \dots, m\}$, x_i 为第 i 个样本的特征向量, y_i 为第 i 个样本对应的预测值,并按照 2:1 的比例随机抽取训练集 T_1 和测试集 T_2 , $T = T_1 \cup T_2, T_1 \cap T_2 = \emptyset$ 。

(2)学习并生成新的数据集。设第 1 层选用 n 个基础学习器,并采用 k 折交叉验证来训练第 1 层模型, n 种模型扩展之后生成第 2 层训练集 T'_1 。在基学习器模型进行 k 折交叉验证过程中,对测试集进行 k 次计算,将 k 次计算得到的结果平均, n 种模型扩展之后生成第 2 层测试集 T'_2 。

使用新训练集和测试集构成新数据集, $T' = T'_1 \cup T'_2, T' = \{(y_i, z_{1i}, z_{2i}, \dots, z_{ni}); i = 1, 2, \dots, m\}$, z_{ni} 为第 i 个样本新的特征向量。

(3)将得到的 T'_1 用于训练第 2 层元学习器模型,并用 T'_2 验证 Stacking 模型性能。训练得到最终 Stacking 集成学习模型,并对模型进行参数优化。

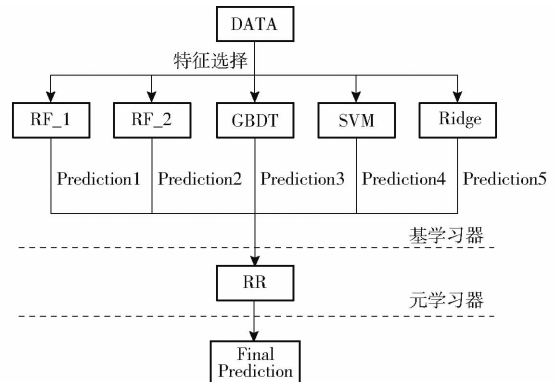


图 2 Stacking 集成学习模型框架图

Fig. 2 Framework of Stacking integrated learning model

采用 RF_1、RF_2、GBDT、SVM 和 Ridge 等 5 个模型作为基学习器,利用岭回归作为元学习器,并进行 5 折交叉验证,步骤(2)中利用每个基础模型(以模型 1 为例)进行 5 折交叉验证的详细过程见图 3,即训练集取其中的 4 折作为训练集,另外一折作为测试集,每次交叉验证都基于测试集训练生成的模型对其进行估测,这部分估测值合成下一层模型的训练集。同时每次交叉验证也对数据集原来的整个测试集进行估测,最后将各部分估测值取算术平均,作为下一层模型测试集。把模型 1 的训练集估测值并列合并得到的矩阵作为训练集,模型 1 的测试集

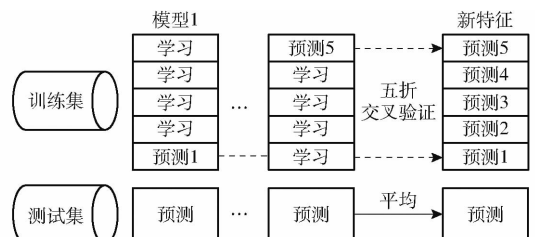


图 3 Stacking 模型步骤 2 详解

Fig. 3 Step 2 of Stacking model

估测值并列合并得到的矩阵作为测试集,代入下一层的模型,再基于它们进一步训练。

1.5 评价指标

采用决定系数(Coefficient of determination, R^2)、均方根误差(Root mean squared error, RMSE)、平均绝对误差(Mean absolute error, MAE)3种常用的评估指标进行模型精度评价,其中 R^2 反映模型拟合优度,越接近1表示所拟合的回归方程越优;RMSE用来衡量模型估测与实际值之间的偏差,越接近0说明模型越准确;MAE能更好地反映估测值误差的实际情况,越接近0说明模型越精确。同时平均绝对百分比误差(Mean absolute percentage error, MAPE)更能直观反映估测结果和真实值之间的差距,越接近0%说明模型越准确,因此选取MAPE来进一步评价模型评估的准确性。

2 结果与讨论

2.1 特征变量的提取及其选择

使用特征变量提取及选择方法,利用波段组合提取9种植被指数。提取测试样本9个特征值,并与相应样方的植被覆盖度进行相关性分析。从皮尔森相关性分析结果(表2)可以看出,与FVC相关性最高的NDVI相关系数为0.953,相关性最低的DVI相关系数为0.904,FVC与9种植被指数相关系数均在0.90以上,且均在 $P < 0.01$ 水平呈极显著相关。但TVI、RVI、DVI与植被覆盖度的相关性与其他特征相比有较明显偏低,因此利用相关系数选择特征集后的结果为选用NDVI、OSAVI、EVI、RDVI、SAVI和MSAVI这6种特征变量。随后基于随机森林模型默认参数进行实验,使用全部特征组合进行训练得分为0.9232,使用以上6种特征组合进行训练得分为0.9224。

表2 测试样本特征变量与FVC相关性分析结果

Tab.2 Correlation analysis results of characteristic variables and FVC of test samples

特征	相关性
NDVI	0.953 **
OSAVI	0.948 **
EVI	0.941 **
RDVI	0.940 **
SAVI	0.938 **
MSAVI	0.934 **
TVI	0.920 **
RVI	0.919 **
DVI	0.904 **

注:**表示在 $P < 0.01$ 水平极显著相关。

随机森林能够度量每个特征的重要性,依据重

要性指标可进一步选择最重要的特征(表3),EVI、SAVI和NDVI重要度在0.20~0.25之间,而其他特征重要度在0.12以下,与前3种植被指数差距较大。因此通过重要性分析选取出重要性依次降低的3种特征组合(EVI、SAVI和NDVI)。基于随机森林模型默认参数进行实验,使用上述3种特征进行训练得分为0.9218。与皮尔森系数法选择的特征组合相比,得分相差仅为0.0006,同时利用重要性排序选择更少的特征变量进行组合训练,得分在0.90以下,与前者相比得分有明显下降。最终选定3种特征组合(EVI、SAVI和NDVI)。

表3 植被指数重要度

Tab.3 Importance of vegetation index

植被指数	重要度
EVI	0.2469
SAVI	0.2132
NDVI	0.2011
RDVI	0.1131
OSAVI	0.0855
MSAVI	0.0606
RVI	0.0579
DVI	0.0114
TVI	0.0076

2.2 模型参数设置

利用随机搜索和网格搜索相结合的方法进行参数优化以提高训练速度和精度。首先利用随机搜索对在较大范围内用大步长进行粗搜索选择出最优点。然后在最优点附近利用较小的步长划分网格进行搜索,从而选择出全局最优解。重复以上步骤,直至网格间距或目标函数变化量小于给定值。优化后各基学习器的主要参数见表4。

表4 各基学习器主要参数

Tab.4 Main parameters of each base learner

学习器	主要参数设置
RF_1	子模型数目:50
	最大特征数量:auto
RF_2	子模型数目:450
	最大特征数量:sqrt
GBDT	子模型数目:350
	最大深度:8
Ridge	学习率:0.0001
	正则化项系数:0.2
SVM	惩罚因子:3

2.3 夏玉米覆盖度估测结果

利用所构建的Stacking集成学习模型对整个研究区的夏玉米覆盖度进行估算(图4)。从整体上来看并不是所有区域长势都均匀,由于每个区域都进

行了不同程度的水分胁迫,影响了夏玉米生长,从而对夏玉米覆盖度也产生了影响。5个区域真实均值分别为0.4127、0.5515、0.5437、0.6386、0.5611,而预测结果也大部分集中在0.36~0.61之间。另外结合夏玉米植被覆盖度估测结果(图4a)和当日实拍无人机可见光影像(图4b)对比发现,估测结果与真实情况基本吻合。

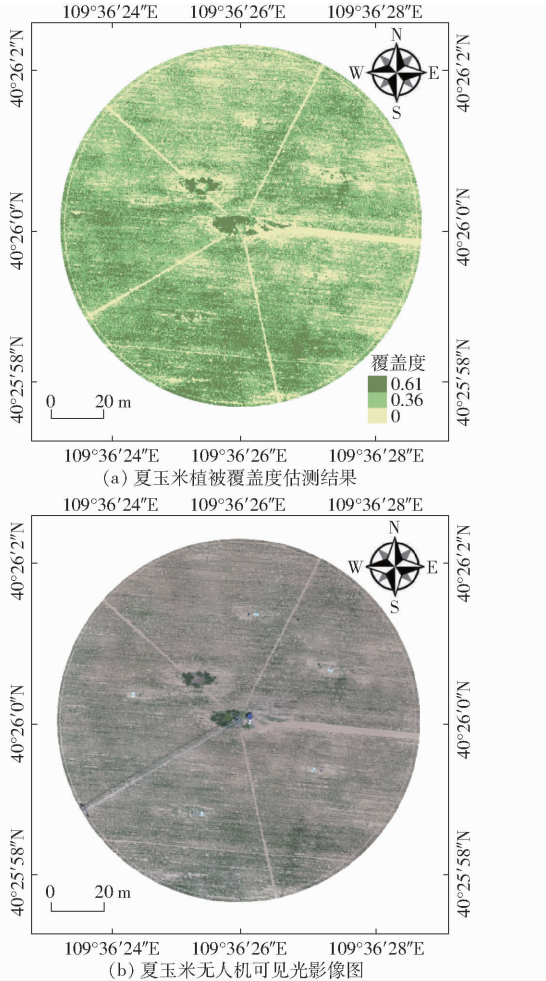


图4 研究区夏玉米 FVC 空间分布图

Fig.4 Spatial distributions of FVC of summer corn in study area

2.4 不同模型结果比较

为了验证模型的性能,将 Stacking 集成学习的夏玉米覆盖度估测模型与单一学习器模型作对比。通过对模型构建与调参得到最优的模型,将测试集分别代入建立好的 RF 模型、GBDT 模型、SVM 模型、岭回归模型、Stacking 集成学习模型中估测夏玉米覆盖度;各算法对夏玉米覆盖度估测的评价指标对比如表5所示。

Stacking 决定系数为 0.9509,对比于 RF、GBDT、SVM 和岭回归分别提高了 0.0291、0.0478、0.0230 和 0.0477,平均提高约 0.0369;Stacking 平均绝对误差为 0.0330,分别降低了 0.0100、0.0154、

表5 单一模型与 Stacking 模型对比

Tab.5 Comparison between the single model and Stacking model

模型	MAPE/%	RMSE	MAE	R^2
RF	6.60	0.0545	0.0430	0.9218
GBDT	7.57	0.0607	0.0484	0.9031
SVM	6.33	0.0523	0.0431	0.9279
岭回归	8.12	0.0606	0.0496	0.9032
Stacking	5.01	0.0432	0.0330	0.9509

0.0191 和 0.0166,平均降低 0.0130;Stacking 均方根误差为 0.0432,分别降低了 0.0113、0.0175、0.0091 和 0.0174,平均降低 0.0138。即 Stacking 集成算法的拟合效果明显好于其他单一模型,具有更高的精度。

Stacking 的平均绝对百分比误差为 5.01%,对比于以上模型分别降低了 1.59、2.56、1.32、3.11 个百分点,平均降低约 2.14 个百分点;即与 4 个单一模型相比,Stacking 集成效果最好。

同时利用其他集成学习框架及常用其他机器学习模型进行训练和测试(表6)。对比于决策树、Xgboost、Adaboost 和 Bagging 模型,Stacking 决定系数分别提高了 0.0426、0.0510、0.0301 和 0.0429,平均提高约 0.0417;Stacking 平均绝对误差分别降低了 0.0119、0.0162、0.0099 和 0.0124,平均降低约 0.0126,Stacking 均方根误差分别降低了 0.0278、0.0185、0.0116 和 0.0159,平均降低约 0.0185。平均绝对百分比误差也分别降低了 2.09、2.73、2.71、2.08 个百分点,平均降低约 2.15 个百分点。即相对于其他集成策略的模型,Stacking 集成学习模型的精度均有较大提升,结果表现较好。

表6 与其他方法比较

Tab.6 Comparison with other methods

模型	MAPE/%	RMSE	MAE	R^2
决策树	7.10	0.0710	0.0449	0.9083
Xgboost	7.74	0.0617	0.0492	0.8999
Adaboost	6.72	0.0548	0.0429	0.9208
Bagging	7.09	0.0591	0.0454	0.9080
Stacking	5.01	0.0432	0.0330	0.9509

为了验证 Stacking 集成学习模型的鲁棒性,利用不同年份采集的其他夏玉米数据集进行训练和测试,实验结果表明,相对于单一学习器,Stacking 集成模型的平均绝对误差平均降低约 0.0101,均方根误差降低约 0.0101,绝对系数提高约 0.0499,平均绝对百分比误差降低约 1.21 个百分点;相对于上述其他模型,Stacking 集成模型的平均绝对误差平均

降低约 0.010 9, 均方根误差降低约 0.012 6, 绝对系数提高约 0.063 8, 平均绝对百分比误差降低约 1.49 个百分点。

3 结论

(1) 利用随机森林算法与皮尔森相关系数相结合的方法简单快速地对特征进行筛选, 一方面说明植被指数在反演植被覆盖度中具有重要作用, 另一方面可考虑将其他的相关信息加入模型特征集合,

以进一步提高模型精度。

(2) 基于 Stacking 集成学习夏玉米覆盖度估测模型结合了多种学习器的优势, 模型精度较高, 仅从 RMSE 来看, 模型误差为 0.043 2, 即准确度可达 95.68%; 仅从 MAE 来看, 模型误差为 0.033, 即准确度可以达到 96.70%; 仅从 MAPE 来看, 模型误差为 5.01%, 即准确度可达 94.99%。且模型拟合程度较好, R^2 达 0.950 9, 该模型集成了多种模型的思想, 进一步提高了模型整体效果。

参 考 文 献

- [1] LEPRIEUR C, KERR Y H, MASTORCHIO S, et al. Monitoring vegetation cover across semi-arid regions: comparison of remote observations from various scales[J]. *International Journal of Remote Sensing*, 2000, 21(2):281-300.
- [2] GITELSON A A, KAUFMAN Y J, STARK R, et al. Novel algorithms for remote estimation of vegetation fraction[J]. *Remote Sensing of Environment*, 2002, 80(1):76-87.
- [3] OWEN T W, CARLSON T N, GILLIES R R. An assessment of satellite remotely-sensed land cover parameters in quantitatively describing the climatic effect of urbanization[J]. *International Journal of Remote Sensing*, 1998, 19(9): 1663-1681.
- [4] CARLSON T N, GILLIES R R, PERRY E M. A method to make use of thermal infrared temperature and NDVI measurements to infer soil water content and fractional vegetation cover[J]. *Remote Sensing Reviews*, 1994, 9(1-2): 161-173.
- [5] 李存军, 赵春江, 刘良云, 等. 红外光谱指数反演大田冬小麦覆盖度及敏感性分析[J]. *农业工程学报*, 2004, 20(5): 159-164.
- LI Cunjun, ZHAO Chunjiang, LIU Liangyun, et al. Retrieval winter wheat ground cover by short-wave infrared indices in field and sensitivity analysis[J]. *Transactions of the CSAE*, 2004, 20(5): 159-164. (in Chinese)
- [6] 赵静, 杨焕波, 兰玉彬, 等. 基于无人机可见光图像的夏季玉米植被覆盖度提取方法[J/OL]. *农业机械学报*, 2019, 50(5): 232-240.
- ZHAO Jing, YANG Huanbo, LAN Yubin, et al. Extraction method of summer corn vegetation coverage based on visible light image of unmanned aerial vehicle [J/OL]. *Transactions of the Chinese Society for Agricultural Machinery*, 2019, 50(5): 232-240. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20190527&journal_id=jcsam. DOI: 10.6041/j.issn.1000-1298.2019.05.027. (in Chinese)
- [7] 杨焕波, 赵静, 兰玉彬, 等. 基于无人机可见光遥感和阈值法的玉米植被覆盖度提取[J]. *福建农林大学学报(自然科学版)*, 2020, 49(2): 283-288.
- YANG Huanbo, ZHAO Jing, LAN Yubin, et al. Vegetation coverage extraction of corn field based on visible light remote sensing and threshold method [J]. *Journal of Fujian Agriculture and Forestry University(Natural Science Edition)*, 2020, 49(2): 283-288. (in Chinese)
- [8] RUNDQUIST B C. The influence of canopy green vegetation fraction on spectral measurements over native tallgrass prairie[J]. *Remote Sensing of Environment*, 2002, 81(1): 129-135.
- [9] SONG W, MU X, RUAN G, et al. Estimating fractional vegetation cover and the vegetation index of bare soil and highly dense vegetation with a physically based method[J]. *International Journal of Applied Earth Observation and Geoinformation*, 2017, 58: 168-176.
- [10] 刘峰, 刘素红, 向阳. 园地植被覆盖度的无人机遥感监测研究[J/OL]. *农业机械学报*, 2014, 45(11): 250-257.
- LIU Feng, LIU Suhong, XIANG Yang. Study on monitoring fractional vegetation cover of garden plots by unmanned aerial vehicles [J/OL]. *Transactions of the Chinese Society for Agricultural Machinery*, 2014, 45(11): 250-257. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20141139&journal_id=jcsam. DOI: 10.6041/j.issn.1000-1298.2014.11.039. (in Chinese)
- [11] 朱学山, 江涛, 魏显虎, 等. 基于遥感卫星的肯尼亚 2000—2013 年植被覆盖度研究[J]. *山东科技大学学报(自然科学版)*, 2016, 35(6): 24-29.
- ZHU Xueshan, JIANG Tao, WEI Xianhu, et al. Study of vegetation coverage during 2000—2013 in Kenya based on remote sensing satellite [J]. *Journal of Shandong University of Science and Technology(Natural Science Edition)*, 2016, 35(6): 24-29. (in Chinese)
- [12] 赵健赟. 地表植被覆盖度遥感估算及其气候效应研究进展[J]. *测绘与空间地理信息*, 2015, 38(8): 77-80, 84.
- ZHAO Jianyun. A review of FVC estimation based on remote sensing and climatic effects [J]. *Surveying and Spatial Geographic Information*, 2015, 38(8): 77-80, 84. (in Chinese)
- [13] 吕长春, 王忠武, 钱少猛. 混合像元分解模型综述[J]. *遥感信息*, 2003, 9(3): 55-58, 60.
- LÜ Changchun, WANG Zhongwu, QIAN Shaomeng. A review of pixel unmixing models[J]. *Remote Sensing Information*, 2003, 9(3): 55-58, 60. (in Chinese)
- [14] 陈彦兵. 基于混合像元分解的鄱阳湖湿地植被覆盖度提取[J]. *科技创新与应用*, 2017(29): 71.
- CHEN Yanbing. Vegetation coverage extraction of Poyang Lake Wetland based on mixed pixel decomposition [J]. *Scientific and Technological Innovation and Application*, 2017(29): 71. (in Chinese)
- [15] 陈向东, 邓江洪. 基于可见光影像的夏季玉米植被覆盖度提取方法研究[J]. *实验技术与管理*, 2019, 36(12): 131-136.
- CHEN Xiangdong, DENG Jianghong. Study on extraction method of summer maize vegetation coverage based on visible light image[J]. *Experimental Technology and Management*, 2019, 36(12): 131-136. (in Chinese)
- [16] 马中刚, 孙华, 王广兴, 等. 基于 Landsat 8-OLI 的荒漠化地区植被覆盖度反演模型研究[J]. *中南林业科技大学学报*, 2016, 36(9): 12-18, 141.
- MA Zhonggang, SUN Hua, WANG Guangxing, et al. Modeling percentage vegetation cover of a desert area using Landsat 8-

- OLI image [J]. *Journal of Central South University of Forestry and Technology*, 2016, 36(9): 12–18, 141. (in Chinese)
- [17] 宋清洁, 崔霞, 张瑶瑶, 等. 基于小型无人机与 MODIS 数据的草地植被覆盖度研究——以甘南州为例[J]. *草业科学*, 2017, 34(1): 40–50.
- SONG Qingjie, CUI Xia, ZHANG Yaoyao, et al. Grassland fractional vegetation cover analysis using small UAVs and MODIS—a case study in Gannan Prefecture[J]. *Pratacultural Science*, 2017, 34(1): 40–50. (in Chinese)
- [18] 蔡宗磊, 包妮沙, 刘善军. 国产高分一号数据估算草地植被覆盖度方法研究——以呼伦贝尔草原露天煤矿区为例[J]. *地理与地理信息科学*, 2017, 33(2): 32–38.
- CAI Zonglei, BAO Nisha, LIU Shanjun. Estimation method of fractional vegetation coverage for grassland based on Chinese GF-1 satellite image; taking Hulun Buir prairie open-pit coal mine as an example [J]. *Geography and Geographic Information Science*, 2017, 33(2): 32–38. (in Chinese)
- [19] 程俊毅, 张显峰, 孙敏, 等. 基于空地协同采样的植被覆盖度随机森林估算方法[J]. *北京大学学报(自然科学版)*, 2020, 56(1): 143–154.
- CHENG Junyi, ZHANG Xianfeng, SUN Min, et al. Random forest model for the estimation of fractional vegetation coverage based on a UAV-ground co-sampling strategy [J]. *Journal of Peking University (Natural Science Edition)*, 2020, 56(1): 143–154. (in Chinese)
- [20] 张宏鸣, 刘雯, 韩文霆, 等. 基于梯度提升树算法的夏玉米叶面积指数反演[J/OL]. *农业机械学报*, 2019, 50(5): 251–259.
- ZHANG Hongming, LIU Wen, HAN Wenting, et al. Inversion of summer maize leaf area index based on gradient boosting decision tree algorithm [J/OL]. *Transactions of the Chinese Society for Agricultural Machinery*, 2019, 50(5): 251–259. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20190529&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2019.05.029. (in Chinese)
- [21] 张宏鸣, 谭紫薇, 韩文霆, 等. 基于无人机遥感的玉米株高提取方法[J/OL]. *农业机械学报*, 2019, 50(5): 241–250.
- ZHANG Hongming, TAN Ziwei, HAN Wenting, et al. Extraction method of maize height based on UAV remote sensing[J/OL]. *Transactions of the Chinese Society for Agricultural Machinery*, 2019, 50(5): 241–250. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20190528&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2019.05.028. (in Chinese)
- [22] 张宏鸣, 王佳佳, 韩文霆, 等. 基于热红外遥感影像的作物冠层温度提取[J/OL]. *农业机械学报*, 2019, 50(4): 203–210.
- ZHANG Hongming, WANG Jiajia, HAN Wenting, et al. Crop canopy temperature extraction based on thermal infrared remote sensing images[J/OL]. *Transactions of the Chinese Society for Agricultural Machinery*, 2019, 50(4): 203–210. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20190423&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2019.04.023. (in Chinese)
- [23] 吉增宝. 用数码相机和 Photoshop 计算植被覆盖度的简易方法[J]. *水土保持应用技术*, 2015(5): 10–11.
- JI Zengbo. A simple method for calculating vegetation coverage with digital photos and Photoshop [J]. *Soil and Water Conservation Application Technology*, 2015(5): 10–11. (in Chinese)
- [24] 虞连玉, 蔡焕杰, 姚付启, 等. 植被指数反演冬小麦植被覆盖度的适用性研究[J/OL]. *农业机械学报*, 2015, 46(1): 231–239.
- YU Lianyu, CAI Huanjie, YAO Fuqi, et al. Applicability of vegetation indices to estimate fractional vegetation coverage[J/OL]. *Transactions of the Chinese Society for Agricultural Machinery*, 2015, 46(1): 231–239. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20150133&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2015.01.033. (in Chinese)
- [25] 高永刚, 徐涵秋. 基于多源遥感影像的多尺度城市植被覆盖度估算[J]. *红外与毫米波学报*, 2017, 36(2): 225–234.
- GAO Yonggang, XU Hanchu. Estimation of multi-scale urban vegetation coverage based on multi-source remote sensing images [J]. *Journal of Infrared and Millimeter Wave*, 2017, 36(2): 225–234. (in Chinese)
- [26] BIRTH G S, MCVEY G R. Measuring the color of growing turf with a reflectance spectrophotometer[J]. *Agronomy Journal*, 1968, 60(6): 640–643.
- [27] ROUSEL J W, HAAS R H, SCHELL J A, et al. Monitoring vegetation systems in the great plains with ERTS[J]. *NASA Special Publication*, 1974(351):309–317.
- [28] HUETE A R. A soil-adjusted vegetation index(SAVI)[J]. *Remote Sensing of Environment*, 1988, 25(3): 295–309.
- [29] QI J, CHEHBOUNI A, HUETE A R, et al. A modified soil adjusted vegetation index[J]. *Remote Sensing of Environment*, 1994, 48(2): 119–126.
- [30] HUETE A, JUSTICE C, VAN LEEUWEN W. MODIS vegetation index(MOD13)[J]. *Algorithm Theoretical Basis Document*, 1999, 3(213): 295–309.
- [31] RONDEAUX G, STEVEN M, BARET F. Optimization of soil-adjusted vegetation indices[J]. *Remote Sensing of Environment*, 1996, 55(2):95–107.
- [32] GILBERT M A, GONZÁLEZ-PIQUERAS J, GARCA-HARO F J, et al. A generalized soil-adjusted vegetation index [J]. *Remote Sensing of Environment*, 2002, 82(2): 303–310.
- [33] JORDAN C F. Derivation of leaf-area index from quality of light on the forest floor[J]. *Ecology*, 1969, 50(4):663–666.
- [34] BROGE N H, LEBLANC E. Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density[J]. *Remote Sensing of Environment*, 2001, 76(2):156–172.
- [35] WOLPERT D H. Stacked generalization[J]. *Neural Networks*, 1992, 5(2):241–259.
- [36] ARIF M, BYUNG-WON O, INGYU L, et al. Spam comments prediction using stacking with ensemble learning[J]. *Journal of Physics: Conference Series*, 2018, 933(1):012012.
- [37] 袁培森, 杨承林, 宋玉红, 等. 基于 Stacking 集成学习的水稻表型组学实体分类研究[J/OL]. *农业机械学报*, 2019, 50(11): 144–152.
- YUAN Peisen, YANG Chenglin, SONG Yuhong, et al. Classification of rice phenomics entities based on Stacking ensemble learning[J/OL]. *Transactions of the Chinese Society for Agricultural Machinery*, 2019, 50(11): 144–152. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20191116&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2019.11.016. (in Chinese)