

doi:10.6041/j.issn.1000-1298.2019.05.029

基于梯度提升树算法的夏玉米叶面积指数反演

张宏鸣¹ 刘雯¹ 韩文霆² 刘全中¹ 宋荣杰¹ 侯贵河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: 为了快速、准确、大范围获取大田夏玉米的叶面积指数(Leaf area index, LAI), 基于实地采集的夏玉米 LAI 和株高, 结合同时期的无人机多光谱影像, 选择与夏玉米 LAI 相关性较强的 8 种植被指数以及株高作为反演 LAI 的输入变量, 采用梯度提升树(Gradient boosting decision tree, GBDT)算法建立植被指数及株高与叶面积指数之间的预测模型, 并与支持向量机(Support vector machine, SVM)和随机森林(Random forest, RF)算法建立的模型进行预测精度对比。结果表明, GBDT 算法在 3 个样本组中的 LAI 预测值与实测值 R^2 分别为 0.571 0、0.755 8、0.644 1, 均高于 SVM 算法(0.547 2、0.679 1、0.616 8)和 RF 算法(0.550 5、0.697 3、0.629 5); 对应的 RMSE 分别为 0.002 7、0.001 5、0.001 6, 均低于 SVM 算法(0.211 7、0.152 3、0.159 7)和 RF 算法(0.244 7、0.214 7、0.208 0)。该研究为快速准确的大田夏玉米 LAI 遥感监测提供了技术和方法。

关键词: 夏玉米; 叶面积指数; 多光谱; 梯度提升树; 植被指数

中图分类号: TP79; S127 **文献标识码:** A **文章编号:** 1000-1298(2019)05-0251-09

Inversion of Summer Maize Leaf Area Index Based on Gradient Boosting Decision Tree Algorithm

ZHANG Hongming¹ LIU Wen¹ HAN Wenting² LIU Quanzhong¹ SONG Rongjie¹ HOU Guihe¹

(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: Aiming to obtain the leaf area index (LAI) of large-scale summer maize in a high-efficiency, non-destructive and large-scale manner, and provide a technical reference for remote sensing monitoring of summer maize growth. The research was based on the field-collected summer maize LAI and maize height, as well as combined with multi-spectral data of the same period, eight vegetation indexes and height with strong correlation with summer maize LAI were selected as the input variables of gradient boosting decision tree (GBDT) algorithm model for LAI inversion. The support vector machine (SVM) model and the random forest (RF) model were taken as the reference models, which were used to compare the accuracy of prediction. The results showed that the GBDT algorithm model prediction consequence were better than the other two models among the three sample groups. R^2 of prediction value and measured LAI values of the sample groups 1, 2 and 3 were 0.571 0, 0.755 8 and 0.644 1, respectively, which were higher than those of the SVM models (0.547 2, 0.679 1, 0.616 8) and RF models (0.550 5, 0.697 3, 0.629 5), corresponding root mean square error (RMSE) values were 0.002 7, 0.001 5 and 0.001 6, which were lower than those of the SVM model (0.211 7, 0.152 3 and 0.159 7) and RF model (0.244 7, 0.214 7 and 0.208 0). The research result provided a technical method for fast and accurate monitoring of summer maize LAI remote sensing in the field.

Key words: summer maize; leaf area index; multispectral; gradient boosting decision tree; vegetation index

0 引言

叶面积指数(Leaf area index, LAI)是生态系统

能量流动中极其重要的植被特征^[1]。在 LAI 定义的不断发展中,为了既可以适用针叶植被又可以适用阔叶植被,最终被定义为单位地表面上绿叶表

收稿日期: 2019-01-19 修回日期: 2019-02-28

基金项目: 国家重点研发计划项目(2017YFC0403203)、国家自然科学基金项目(41771315、41301283、41371274)和欧盟地平线 2020 研究与创新计划项目(GA:635750)

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区域土壤侵蚀评价研究, E-mail: zhm@nwsuaf.edu.cn

面积总和的一半^[2]。传统获取 LAI 的方法主要依靠人工实地测量^[3],不仅浪费人力,而且对作物破坏性极大。随着遥感技术的不断发展,获取 LAI 的方法开始转向于遥感反演^[4-6]。目前,遥感数据主要来源于卫星和无人机,但卫星遥感属于高空遥感技术,在获取遥感图像过程中易受到大气因素的干扰^[7]。且卫星遥感的区域面积适合在公顷以上,如果进行小范围的区域研究,其分辨率受到一定的限制。因此,近年来产生了利用无人机搭载多光谱或高光谱相机的方式采集遥感图像,该技术弥补了卫星遥感的不足,提高了图像的地面分辨率,可快速准确获得作物叶面积指数,为农情信息监测提供了有效手段^[8-9]。

中国作为农业大国,玉米是不可或缺的农作物。长期以来,我国在作物的叶面积指数反演方面进行了相关研究^[10]。早期在进行 LAI 反演时都选择用单一植被指数作为输入变量^[11-12],但单一的植被指数存在不同程度饱和性^[13],对于 LAI 的反演受到了一定的限制。之后的研究中发现,红波段和近红外波段的反射率与作物叶片的特征不仅最为密切相关且有别于其他地物^[14],可组合出多种的植被指数,因而是当前用于 LAI 反演的常用波段^[15]。为了可以融合多种与 LAI 相关性较强的植被指数及其他相关数据,衍生出了机器学习的方法。机器学习就单因子训练模型进行了改进,利用多种相关因子的非线性拟合来构建较为精确的模型,使得验证集与实际值之间的预测误差最小,泛化能力最强^[16]。

谢巧云等^[17]利用高光谱数据进行作物 LAI 遥感反演,结果表明,非线性支持向量机模型最适宜用于研究区域冬小麦 LAI 反演;王丽爱等^[18]将支持向量回归和反向传播神经网络算法作为比较模型,证明随机森林模型具有强学习能力和预测能力;张春兰等^[19]基于无人高光谱模型研究了 4 个生育期小麦的 LAI,发现随机森林模型 LAI 的反演精度较高,且适用性较强;HOUBORG 等^[20]发现,利用机器学习的方法,可有效分析和利用数量大、维度高的观测数据。以上研究成果表明,机器学习中的回归模型广泛应用于作物遥感反演,且取得了较好的研究成果^[21-22]。但回归模型都存在小样本数据出现过拟合的问题,且无法判断输入因子中的主要贡献因子。梯度提升树算法 (GBDT) 可以解决以上问题,该算法通过每一棵树学习之前所有树的残差和结果,一步步迭代构建弱学习器,纠正原模型误差,有效提高了预测精度^[23],可作为反演夏玉米 LAI 的有效方法。

本文以 2018 年内蒙古自治区鄂尔多斯市达拉

特旗实验基地为研究区域,采集夏玉米无人机多光谱影像,提取关键生育期 V3~R5(出苗期至初入蜡熟期)的 8 种植被指数,实测 LAI 及玉米株高。将株高和 8 种植被指数作为输入变量,LAI 作为输出变量输入 SVM 算法、RF 算法、GBDT 算法 3 个模型中训练学习,以验证 GBDT 算法反演模型在夏玉米 LAI 反演中的适用性。

1 材料与方法

1.1 研究区概况

实验基地(40°25′46.99″ N, 109°36′35.68″ E)位于内蒙古自治区鄂尔多斯市达拉特旗昭君镇,属于典型的温带大陆性气候,干旱多大风,主要粮食作物为小麦和玉米。实验区所种作物为夏玉米,属于春播中晚熟型,一年一熟。播种时间为 2018 年 5 月中旬,收割时间为 9 月。选取株高为 60 cm,叶片宽度约两指宽(3 cm 左右)时进行田间玉米 LAI 实验。由于当地的降水量无法满足玉米生长期内的需水量,主要的供水方式为喷灌机灌水。据此,将研究区中的④区利用喷灌机作水分胁迫,使之除大气降水外,额外供水量少于其余区域。研究区无人机影像如图 1 所示,实验区域面积约为 1.13 hm²,以喷灌机为中心,分为 5 个扇形区域,每个扇形区域内划分 3 个 4 m×4 m 的实验样方,即研究区总计 15 个小样方。

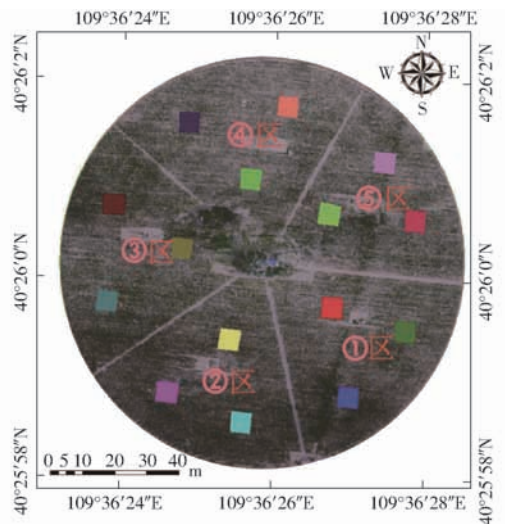


图 1 研究区无人机影像和分区示意图

Fig. 1 UAV image and partition map of study area

1.2 数据的获取与处理

田间夏玉米 LAI 的测量使用 LAI-2200C 型植物冠层仪进行。数据于 2018 年 6 月 26 日开始采集,2018 年 8 月 25 日结束,共 9 次实验,期间覆盖夏玉米的 V3(出苗期)、V6(拔节期)、VT(抽雄期)、R1(吐丝期)、R3(乳熟期)、R5(初入蜡熟期)6 个关键生育期。前 8 次实验数据用于建立模型,最后 1

次实验数据用于模型预测及验证。LAI 采集方式采用 ABBBB,即测量时,取4次冠下B值,一次冠上A值。由于玉米属于行栽作物,测量方式通常采用对角线法:B值取样点位于两垄之间均匀分布,A值取样点避免阳光直射,位于冠层上方。每个小样方重复采集4组数据,取均值。考虑正午时分阳光强烈,测量时需配备遮光帽,以减少视野遮盖帽周围的反射光对数值的影响。夏玉米的株高在测量 LAI 时一同测量同一样方对角线的点,在对角线3点的区域内每一点测量5组数据,取均值。

遥感图像数据采集为六翼无人机 RedEdge 五波段搭载多光谱相机,相机焦距为 5.5 mm,视场角为 47.2°,图像分辨率为 1 280 像素 × 960 像素。相机配备了光强传感器和两个 3 m × 3 m 的灰板。光强传感器可对无人机航拍过程中外界光线的变化对光谱影像造成的影响进行校正,而灰板由于具有固定的反射率,可对航拍影像进行反射率的校正,从而生成反射率影像图,进行植被指数的提取。相机的波段信息以及灰板的反射率如表 1 所示。实验时晴朗无云,平均气温 28.6℃,平均相对湿度 61.96%,平均风速 1.12 m/s,微风。时间在 11:30—14:30。多光谱无人机飞行高度为 70 m,飞行方向为南北方向,航向、旁向重叠度分别为 80% 和 70%。将每次实验按照固定航线拍摄的多张图像,以日期为索引导入到瑞士 Pix4D 公司的 Pix4D mapper 软件中,以实时动态(Real time kinematic, RTK)测量的方法获取地面像控点,导入小图对应的 POS 数据,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,最终生成高清正射多光谱影像。由于拼接预处理后的原始图像包含除研究区域以外很大的区域,为了更加突显遥感影像的作物特征,需在 ENVI 软件中裁剪处理。根据可见光影像中的样方对应裁剪出多光谱影像的 15 块实测区域,每一块实测区域一一对应实验样地的每一小样方。取裁剪后每一小样方的对角线 3 点,生成这 3 点对应的各项植被指数,最终以均值来确定每个小

样方的植被指数。

由于正午时分和黄昏时分叶片的蜷缩程度受温度影响会有一些的区别,且在不同的时间段接收的光谱信息也有很大的差异,因此多光谱影像数据和地面数据采集需在同一天的同一时间段。

1.3 植被指数的提取

植被对于不同波段入射光子的吸收作用和散射作用不同,形成了特殊的光谱响应特征。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和 LAI 具有很好的相关关系^[24]。因此本文借鉴前人研究,选择红光、绿光和近红外波段组合出与 LAI 相关性较强的 8 种植被指数进行 LAI 反演,各计算公式^[25-26]见表 2。

表 2 植被指数计算公式

Tab.2 Formulas of vegetation index

植被指数	计算公式
归一化植被指数 (NDVI)	$NDVI = \frac{R_{Nir} - R_{Red}}{R_{Nir} + R_{Red}}$
比值植被指数 (RVI)	$RVI = \frac{R_{Nir}}{R_{Red}}$
土壤调节植被指数 (SAVI)	$SAVI = 1.5 \frac{R_{Nir} - R_{Red}}{R_{Nir} + R_{Red} + 0.5}$
优化土壤调节植被指数 (OSAVI)	$OSAVI = 1.16 \frac{R_{Nir} - R_{Red}}{R_{Nir} + R_{Red} + 0.16}$
增强性植被指数 2 (EVI2)	$EVI2 = 2.5 \frac{R_{Nir} - R_{Red}}{R_{Nir} + 2.4R_{Red} + 1}$
重归一化植被指数 (RDVI)	$RDVI = \frac{R_{Nir} - R_{Red}}{\sqrt{R_{Nir} + R_{Red}}}$
修正土壤调整植被指数 (MASVI)	$MASVI = \frac{1}{2} [2R_{Nir} + 1 - \sqrt{(2R_{Nir} + 1)^2 - 8(R_{Nir} - R_{Red})}]$
三角植被指数 (TVI)	$TVI = \frac{1}{2} [120(R_{Nir} - R_{Green}) - 200(R_{Red} - R_{Green})]$

注: R_{Nir} 、 R_{Red} 、 R_{Green} 分别为灰板对 RedEdge 相机近红外、红波段、绿波段的平均反射率。

1.4 模型构建

1.4.1 支持向量机

支持向量机(Support vector machine, SVM)中的支持向量回归(Support vector regression, SVR)是早期机器学习中常用的回归模型。SVR 就是寻找一个最优的回归平面,让集合中所有的数据到这个回归平面的距离最近^[13]。本文选择了 SVM 中优化的 LIBSVM 库进行夏玉米的 LAI 预测。LIBSVM^[27]具有操作简单、快速有效和可处理高维空间数据等特点,常用来做分类和回归。支持向量回归与传统的回归模型相比,优点是此算法可容忍基于模型的输

表 1 RedEdge 多光谱相机参数及灰板对其中心波长的反射率

Tab.1 Multispectral camera parameters and reflectivity of gray plate to its center wavelength

波段编号	波段名称	中心波长/nm	波宽/nm	灰板 1 反射率/%	灰板 2 反射率/%
1	蓝	475	20	2.946	58.446
2	绿	560	20	2.945	57.958
3	红	668	10	2.993	57.339
4	近红外	840	40	3.070	56.682
5	红边	717	10	3.010	57.128

出 $f(x)$ 与真实的输出 y 之间有 ε 的偏差,也就是当 $|f(x) - y| > \varepsilon$ 时才计算损失。但 SVM 算法对非线性问题没有通用的解决方案,难以找到合适的核函数。

1.4.2 随机森林

随机森林(Random forest, RF)是决策树的集成算法,通过对大量分类树的汇总以提高模型的预测精度,是取代神经网络等传统机器学习方法的新算法。随机森林算法中包含多个决策树来降低过拟合风险^[28],思想是 Bagging 算法和随机特征选取^[20]。随机森林通过对训练样本重新采样的方法得到不同的训练样本集;在新的训练样本集上分别进行训练学习,由于每个学习器相互独立,所以此类方法更容易并行;最后合并每一个学习器的结果,从而得到最终的学习结果。随机森林的一个缺点是在噪声较大的分类或者回归问题中容易过拟合。

1.4.3 梯度提升树

梯度提升树(Gradient boosting decision tree, GBDT)是集成学习中重要的一种算法,是基于 Booting 算法的一种改进^[29-30]。Booting 算法的工作原理是在初始训练集样本上给每个训练样本赋予相同的权值,在每一次训练之后对于出错的样本进行增加错分点的权值,在经过多次迭代后,生成相应的多个基学习器,之后对于这些基学习器进行组合,最终通过加权或投票得到模型。而梯度提升树回归与分类算法的区别是输入的训练数据是残差,即将上一次的预测结果带入残差中求出本轮的训练数据,而不是损失函数的梯度^[22]。

梯度提升树具有可灵活处理各种数据、预测准确率高、使用健壮的损失函数和对异常值具有很强的鲁棒性等优点,可有效进行回归预测。GBDT 回归算法如下:

(1) 输入训练样本

$$D = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$$

(2) 初始化弱学习器

$$f_0(x) = \arg \min_c \sum_{i=1}^m L(y_i, c)$$

式中 L ——损失函数

c ——样本 y 的均值

(3) 计算负梯度

$$r_{ii} = - \left(\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)} \right)_{f(x) = f_{i-1}(x)}$$

($i = 1, 2, \dots, m; t = 1, 2, \dots, T$)

式中 T ——最大迭代次数

(4) 利用 (x_i, r_{ii}) 拟合一棵分类回归树(Classification and regression tree, CART)中的回归

树,从而得到第 t 棵回归树,其对应的叶子节点区域 $R_{ij} (j = 1, 2, \dots, J)$, J 为回归树 t 的叶子节点的个数。

(5) 计算最佳拟合值

$$c_{ij} = \arg \min_c \sum_{x_i \in R_{ij}} L(y_i, f_{i-1}(x) + c)$$

(6) 更新强学习器

$$f_i(x) = f_{i-1}(x) + \sum_{j=1}^J c_{ij} \quad (x \in R_{ij})$$

(7) 得到强学习器表达式

$$f(x) = f_T(x) = f_0(x) + \sum_{t=1}^T \sum_{j=1}^J c_{tj} \quad (x \in R_{ij})$$

(8) 输出为强学习器。

梯度提升树算法中,所产生的树是回归树而不是分类树,GBDT 的树会累加之前所有树的结果,这种累加的实现只能用 CART 回归树实现。

1.5 模型评价

以决定系数(R^2)和均方根误差(RMSE)来进行模型精度的评价。

为保证模型的有效性和稳定性,本文基于实验分区来划分不重复的训练集和验证集,将样本 3 次分组,重复放入模型训练学习。样本组 1 的训练集为区域①、②、④,验证集为区域③和⑤;样本组 2 的训练集为区域②、③、⑤,验证集为区域①和④;样本组 3 的训练集为区域①、③、⑤,验证集为区域②和④。

2 结果与分析

2.1 植被指数相关性分析

基于前人对于作物 LAI 反演进展的研究,分析光谱特征信息发现作物 LAI 对于红光参数与近红外参数较为敏感,为后期的植被指数的选择提供了方向。本文对 3 组样本中 8 种植被指数和株高与 LAI 进行相关性分析,结果如表 3 所示, LAI 与 8 种植被指数和株高在 $P < 0.01$ 水平呈极显著相关,训练集相关系数均不小于 0.660,验证集相关系数均不小于 0.668。NDVI、OSAVI、RDVI、RVI、SAVI、EVI2、MASVI、TVI 与 LAI 在总样本中相关系数平均为 0.777、0.765、0.751、0.769、0.747、0.743、0.744、0.715($P < 0.01$),因此可选择此 8 种植被指数作为构建 LAI 反演模型的变量。而本实验中新加入的实测株高在 $P < 0.01$ 水平下与 LAI 的相关系数均值也达到了 0.769,说明株高与夏玉米的叶面积指数有着较强的相关性,可以选择与 8 种植被指数一同作为 SVM、RF 和 GBDT 模型的输入变量,进行夏玉米 LAI 的预测研究。

2.2 夏玉米 LAI 反演模型的构建

由于实地采集的样本相对较少,因此选取 3 个区夏玉米生长关键生育期的 9 次实验数据进行反演

表 3 植被指数与 LAI 相关性分析结果

Tab. 3 Correlation analysis result of LAI and vegetation index

参数	样本组 1		样本组 2		样本组 3		全样本
	训练集	验证集	训练集	验证集	训练集	验证集	
株高	0.784**	0.739**	0.722**	0.829**	0.806**	0.710**	0.769**
NDVI	0.826**	0.697**	0.722**	0.849**	0.763**	0.807**	0.777**
OSAVI	0.811**	0.691**	0.712**	0.835**	0.749**	0.799**	0.765**
RDVI	0.795**	0.684**	0.699**	0.821**	0.734**	0.788**	0.751**
RVI	0.807**	0.705**	0.715**	0.848**	0.766**	0.790**	0.769**
SAVI	0.790**	0.681**	0.695**	0.817**	0.730**	0.785**	0.747**
EVI2	0.784**	0.681**	0.691**	0.814**	0.727**	0.780**	0.743**
MASVI	0.784**	0.684**	0.692**	0.815**	0.729**	0.780**	0.744**
TVI	0.749**	0.668**	0.660**	0.788**	0.700**	0.751**	0.715**

注：**表示在 $P < 0.01$ 水平上极显著相关。

模型的训练,8 种植被指数和同期的株高共 9 个因子一同作为训练模型的输入变量,夏玉米 LAI 作为输出变量,分别使用 SVM 算法、RF 算法和 GBDT 算法来构建夏玉米 LAI 反演模型。SVM 算法模型用 LIBSVM 库来实现,核函数选择径向基,其余参数根据网格搜索法来确定最优参数;RF 算法模型根据多次实验,确定决策树的数量为 100,节点分割变量为 3;而本文构建 GBDT 算法模型,用 Python 语言编写回归程序,根据输入样本组的不同,防止出现过拟合现象,需多次实验确定每组训练集的子模型数目 ($n_estimators$)、损失函数 ($loss$)、树的最大深度

(max_depth) 等参数。

以样本组 2 为例说明参数选择过程:调参前,样本组 2 由于选择的树深和子模型数目等过大引起过拟合(图 2a)。调参时,选择最小绝对偏差 (lad) 为损失函数, $n_estimators$ 为 500, max_depth 为 4,步长 ($learning_rate$) 为 0.01,叶节点最小样本 ($min_samples_leaf$) 为 6,作为样本组 2 的模型参数(图 2b)。由图 2 可以看出,在提升迭代次数 ($Boosting\ iterations$) 为 500 时,训练集偏差和验证集偏差 ($Deviance$) 分别最小,500 为最优的提升迭代次数,过大或过小都会造成预测精度的降低。

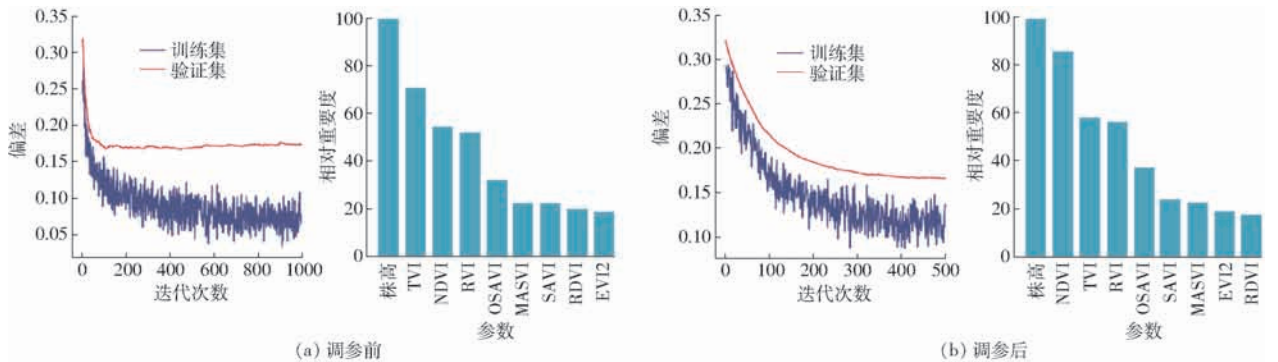


图 2 样本组 2 调参对比

Fig. 2 Contradistinction of parameter adjustment for sample group 2

GBDT 算法模型训练结束后,会出现所使用输入变量的相对重要度 (Relative importance), 便于理解哪些因素对于预测结果有关键影响力。同样也可以判别出几种由红、近红外波段组合出的与 LAI 强相关的相似植被指数中,对于夏玉米 LAI 反演结果的影响力强弱。由图 3 可知,3 组样本中株高在相对重要度中比例占据第一,是影响模型精度的主要因素;而 MASVI 在 3 组样本中相对重要度比例相对较低,即 MASVI 对 LAI 反演模型的影响力较弱;综合 3 组的变量重要性分析,株高对于 LAI 反演模型的结果贡献较大,MASVI 对于 LAI 反演模型的结

果贡献较小。

2.3 夏玉米 LAI 反演模型比较分析

先将 3 组训练集分别以 3 种算法进行训练,不断进行调参和多次迭代训练,得到优化的模型;进而分别将 3 组样本组中独立于训练集的验证集,作为训练模型的最后验证;最后将训练集和验证集模型预测得到的 LAI 与实地测量的 LAI 分别进行散点图分析,拟合成线性的回归线(图 4)。

由图 4 可知,GBDT 算法对于连续值的预测效果较 SVM 算法和 RF 算法好。GBDT 算法在每一样本组中都体现出强大的学习能力,训练集的决定系

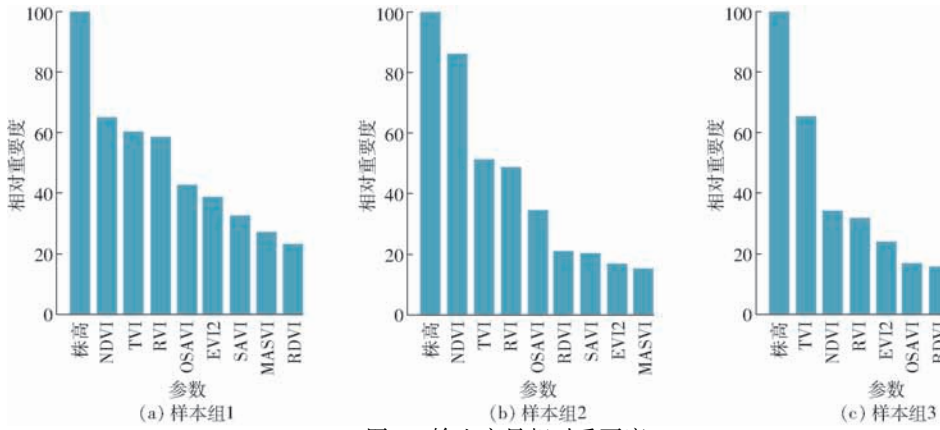


图3 输入变量相对重要度

Fig.3 Relative importance of input variables

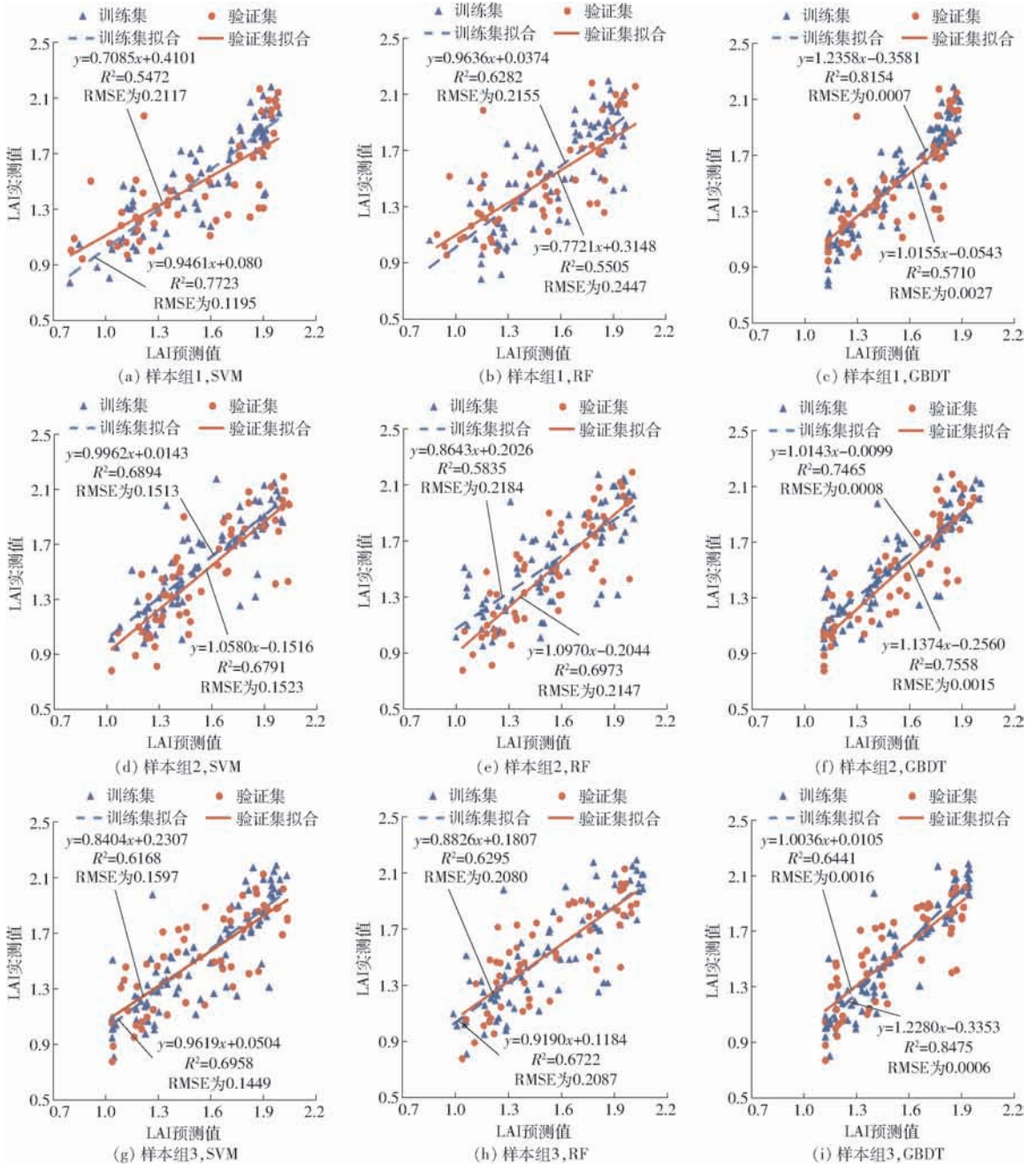


图4 夏玉米 LAI 实测值与预测值关系

Fig.4 Relationship graphs of measured and predicted summer maize LAI

数 R^2 分别为 0.815 4、0.746 5、0.847 5, 对应的 RMSE 为 0.000 7、0.000 8、0.000 6; 验证集的决定系数 R^2 分别是 0.571 0、0.755 8、0.644 1, 对应的 RMSE 为 0.002 7、0.001 5、0.001 6。

2.4 夏玉米 LAI 空间分布

依据 GBDT 算法对 3 个样本组反演结果的分析, 选取效果最佳的样本组 2 模型作为反演模型, 对 8 月 23 日的数据进行研究区内夏玉米 LAI 反演, 该研究区的夏玉米 LAI 空间分布如图 5 所示。

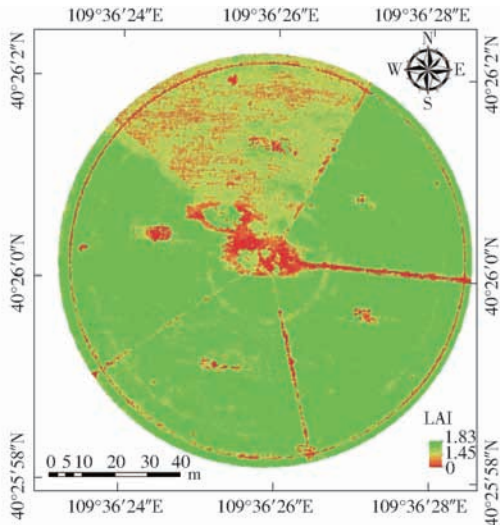


图5 研究区夏玉米 LAI 空间分布图

Fig. 5 Spatial distribution map of summer maize in study area

由图 5 可知, 8 月 23 日的夏玉米 LAI 在①、②、③、⑤区主要集中在 1.8 左右, ④区的 LAI 在 1.4 左右。与 8 月 23 日的实测 LAI 相比较, 总研究区域实测均值为 1.851 4, 实测最大值为 2.190 0, 实测最小值为 1.315 8。5 个区域实测均值分别为 2.092 6、1.795 8、2.087 0、1.406 9、1.888 1。由以上数据可知 LAI 反演数据与实测数据基本相符。整体上①、②、③、⑤区 LAI 较高, 这是因为①、②、③、⑤区喷灌正常, 玉米长势均匀; 研究区④夏玉米 LAI 较低, 主要是因为④区进行了一定的水分胁迫, 导致夏玉米生长较其余 4 个区缓慢, 区域内作物稀疏; 综上, 基于 GBDT 算法模型反演的 LAI 与研究区的 LAI 空间分布相对一致, 进一步证明了该模型的合理性和可靠性。

3 讨论

LAI 反映的是叶片的疏密程度, 随着作物的生长, 高度不断增长, 叶片的疏密程度按照稀疏-稠密-稀疏变化, 高度表现出对于 LAI 的显著影响。本文经过输入变量相关性分析和输出结果相对重要度分

析两方面佐证了株高对 LAI 具有极显著关系, 将遥感数据(植被指数)与实测数据(株高)相结合一同进行 LAI 的反演, 可得到较好的反演效果。因此, LAI 反演中不应只局限于遥感图像提取的植被指数, 还可考虑其他对于 LAI 有影响的因子。

机器学习算法与作物遥感反演密切相关, 一个回归能力很强的机器学习算法模型, 可以融合众多因子(多种植被指数和株高)共同反演 LAI, 大幅度提高结果的精度。本文将 GBDT 算法应用到农作物邻域, 反演夏玉米 LAI。基于模型稳定性的考虑, 分为 3 组样本重复训练, 以 8 种植被指数和株高构建 GBDT 算法反演模型, 并与 SVM 算法和 RF 算法 R^2 、RMSE 相比较, 结果表明 GBDT 算法构建的模型两项指标均高于 SVM 算法的模型, 同样也较同出一派的 RF 算法构建的模型性能有了进一步的提升。在后期的研究中可以将 GBDT 算法应用到玉米叶绿素、玉米生物量等作物相关参数的反演, 以扩大精准农业技术支持的范围。

GBDT 算法的优势在于将若干个弱学习器组合成强学习器, 结果是多棵回归树的累加之和。由此算法构建的模型可以灵活处理各种数据, 不论是本文中的 LAI 连续值, 还是后续研究中作物冠层温度的离散值; 在小样本的回归问题中, GBDT 算法可通过设置不同的损失函数以及在相对较少的调参时间下, 提高反演精度。GBDT 算法不仅减少了 SVR 模型因选择核函数和其他参数造成的时间复杂度的浪费, 而且也解决了 RF 算法对待各输入因子是同一权值, 无法判断其中每一因子的贡献率的问题。因此, GBDT 算法在回归问题中有很强的应用价值。但还需要注意两点: ①利用 GBDT 算法进一步判断了在红波段和近红外波段组成的几种相似植被指数的不同影响度, 突出了算法的优势, 但在精度方面的后续研究中还有进一步提升空间。②GBDT 算法中的基学习器之间存在依赖关系, 一般难以进行并行计算。本文尚未考虑各个基学习器之间的并行操作, 在今后的研究中应着重考虑如何实现部分的并行操作, 进一步提高反演模型的效率。

4 结束语

将机器学习中梯度提升树应用到夏玉米的 LAI 反演中, 并与机器学习中的支持向量机和随机森林算法进行了对比。该算法构建的模型与无人机多光谱图像相结合, 具有较好的反演效果, 为实现大面积、无损夏玉米 LAI 反演和遥感监测作物长势提供了技术支持。

参 考 文 献

- [1] MAIRE G L, MARSDEN C, VERHOEF W, et al. Leaf area index estimation with MODIS reflectance time series and model inversion during full rotations of *Eucalyptus* plantations[J]. *Remote Sensing of Environment*, 2011, 115(2): 586–599.
- [2] CHEN J M, BLACK T A. Defining leaf-area index for non-flat leaves[J]. *Agricultural & Forest Meteorology*, 2010, 15(4): 421–429.
- [3] 刘榕源,王纪华,杨贵军,等. 冬小麦叶面积指数地面测量方法的比较[J]. *农业工程学报*, 2011, 27(3): 220–224.
LIU Rongyuan, WANG Jihua, YANG Guijun, et al. Comparison of ground-based LAI measuring methods on winter wheat[J]. *Transactions of the CSAE*, 2011, 27(3): 220–224. (in Chinese)
- [4] 惠凤鸣,田庆久,金震宇,等. 植被指数与叶面积指数关系研究及定量化分析[J]. *遥感信息*, 2003(2): 10–13.
HUI Fengming, TIAN Qingjiu, JING Zhenyu, et al. Research and quantitative analysis of the correlation between vegetation index and leaf area index[J]. *Remote Sensing Information*, 2003(2): 10–13. (in Chinese)
- [5] 邢丽玮,李小娟,李昂晟,等. 基于高光谱与多光谱植被指数的洪河沼泽植被叶面积指数估算模型对比研究[J]. *湿地科学*, 2013, 11(3): 313–319.
XING Liwei, LI Xiaojuan, LI Angsheng, et al. A comparative study on estimation model for leaf area index of vegetation in marshes in Honghe national nature reserve based on hyperspectral and multispectral vegetation indices[J]. *Wetland Science*, 2013, 11(3): 313–319. (in Chinese)
- [6] 贾玉秋,李冰,程永政,等. 基于GF-1与Landsat-8多光谱遥感影像的玉米LAI反演比较[J]. *农业工程学报*, 2015, 31(9): 173–179.
JIA Yuqiu, LI Bing, CHENG Yongzheng, et al. Comparison between GF-1 images and Landsat-8 images in monitoring maize LAI[J]. *Transactions of the CSAE*, 2015, 31(9): 173–179. (in Chinese)
- [7] 刘照言,马灵玲,唐伶俐. 基于SAIL模型的多角度多光谱遥感叶面积指数反演[J]. *干旱区地理*, 2010, 33(1): 93–98.
LIU Zhaoyan, MA Lingling, TANG Lingli. Inversion of LAI based on SAIL model with multi-angle and multi-spectral remote sensing data[J]. *Arid Land Geography*, 2010, 33(1): 93–98. (in Chinese)
- [8] 王茜,宁纪锋,曹宇翔,等. 基于SIFT算法的无人机遥感图像拼接技术[J]. *吉林大学学报(信息科学版)*, 2017, 35(2): 188–197.
WANG Qian, NING Jifeng, CAO Yuxiang, et al. Matching technologies of UAV remote sensing image based on SIFT[J]. *Journal of Jilin University(Information Science Edition)*, 2017, 35(2): 188–197. (in Chinese)
- [9] 高林,杨贵军,于海洋,等. 基于无人机高光谱遥感的冬小麦叶面积指数反演[J]. *农业工程学报*, 2016, 32(22): 113–120.
GAO Lin, YANG Guijun, YU Haiyang, et al. Retrieving winter wheat leaf area index based on unmanned aerial vehicle hyperspectral remote sensing[J]. *Transactions of the CSAE*, 2016, 32(22): 113–120. (in Chinese)
- [10] 林卉,梁亮,张连蓬,等. 基于支持向量机回归算法的小麦叶面积指数高光谱遥感反演[J]. *农业工程学报*, 2013, 29(11): 139–146.
LIN Hui, LIANG Liang, ZHANG Lianpeng, et al. Wheat leaf area index inversion with hyperspectral remote sensing based on support vector regression algorithm[J]. *Transactions of the CSAE*, 2013, 29(11): 139–146. (in Chinese)
- [11] 李凤秀,张柏,宋开山,等. 玉米叶面积指数与高光谱植被指数关系研究[J]. *遥感技术与应用*, 2007, 22(5): 586–592.
LI Fengxiu, ZHANG Bai, SONG Kaishan, et al. Research and analysis of the correlation between hyperspectral vegetation index and leaf area index[J]. *Remote Sensing Technology & Application*, 2007, 22(5): 586–592. (in Chinese)
- [12] 袁媛,瑚波,武兴厚,等. 夏玉米植被指数与叶面积指数的关系研究[J]. *陕西气象*, 2015(3): 24–28.
YUAN Yuan, HU Bo, WU Xinghou, et al. Research on the relationship between summer corn vegetation index and leaf area index[J]. *Journal of Shanxi Meteorology*, 2015(3): 24–28. (in Chinese)
- [13] 梁栋,管青松,黄文江,等. 基于支持向量机回归的冬小麦叶面积指数遥感反演[J]. *农业工程学报*, 2013, 29(7): 117–123.
LIANG Dong, GUAN Qingsong, HUANG Wenjiang, et al. Remote sensing inversion of leaf area index based on support vector machine regression in winter wheat[J]. *Transactions of the CSAE*, 2013, 29(7): 117–123. (in Chinese)
- [14] 刘洋,刘荣高,陈镜明,等. 叶面积指数遥感反演研究进展与展望[J]. *地球信息科学学报*, 2013, 15(5): 734–743.
LIU Yang, LIU Ronggao, CHEN Jingming, et al. Current status and perspectives of leaf area index retrieval from optical remote sensing data[J]. *Journal of Geo-information Science*, 2013, 15(5): 734–743. (in Chinese)
- [15] DURBHA S S, KING R L, YOUNAN N H. Support vector machines regression for retrieval of leaf area index from multiangle imaging spectroradiometer[J]. *Remote Sensing of Environment*, 2007, 107(1–2): 348–361.
- [16] 夏天,吴文斌,周清波,等. 冬小麦叶面积指数高光谱遥感反演方法对比[J]. *农业工程学报*, 2013, 29(3): 139–147.
XIA Tian, WU Wenbin, ZHOU Qingbo, et al. Comparison of two inversion methods for winter wheat leaf area index based on hyperspectral remote sensing[J]. *Transactions of the CSAE*, 2013, 29(3): 139–147. (in Chinese)
- [17] 谢巧云,黄文江,蔡淑红,等. 冬小麦叶面积指数遥感反演方法比较研究[J]. *光谱学与光谱分析*, 2014, 34(5): 1352–1356.
XIE Qiaoyun, HUANG Wenjiang, CAI Shuhong, et al. Comparative study on remote sensing inversion methods for estimating

- winter wheat leaf area index[J]. *Spectroscopy and Spectral Analysis*, 2014, 34(5): 1352–1356. (in Chinese)
- [18] 王丽爱, 马昌, 周旭东, 等. 基于随机森林回归算法的小麦叶片 SPAD 值遥感估算[J/OL]. *农业机械学报*, 2015, 46(1): 259–265.
WANG Liai, MA Chang, ZHOU Xudong, et al. Estimation of wheat leaf SPAD value using RF algorithmic model and remote sensing data[J/OL]. *Transactions of the Chinese Society for Agricultural Machinery*, 2015, 46(1): 259–265. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20150136&journal_id=jcsam. DOI: 10.6041/j.issn.1000-1298.2015.01.036. (in Chinese)
- [19] 张春兰, 杨贵军, 李贺丽, 等. 基于随机森林算法的冬小麦叶面积指数遥感反演研究[J]. *中国农业科学*, 2018, 51(5): 855–867.
ZHANG Chunlan, YANG Guijun, LI Heli, et al. Remote sensing inversion of leaf area index of winter wheat based on random forest algorithm[J]. *Scientia Agricultura Sinica*, 2018, 51(5): 855–867. (in Chinese)
- [20] HOU BORG R, MCCABE M F. A hybrid training approach for leaf area index estimation via Cubist and random forests machine-learning[J]. *ISPRS Journal of Photogrammetry & Remote Sensing*, 2017, 135: 173–188.
- [21] 岳继博, 杨贵军, 冯海宽. 基于随机森林算法的冬小麦生物量遥感估算模型对比[J]. *农业工程学报*, 2016, 32(18): 175–182.
YUE Jibo, YANG Guijun, FENG Haikuan. Comparative of remote sensing estimation models of winter wheat biomass based on random forest algorithm[J]. *Transactions of the CSAE*, 2016, 32(18): 175–182. (in Chinese)
- [22] 贾洁琼, 刘万青. 基于 GF-1WV 影像和机器学习算法的玉米叶面积指数估算[J]. *中国图象图形学报*, 2018, 23(5): 107–117.
JIA Jieqiong, LIU Wanqing. Estimation of maize leaf area index based on GF-1 WV image and machine learning random algorithm[J]. *Journal of Image & Graphics*, 2018, 23(5): 107–117. (in Chinese)
- [23] FRIEDMAN J H. Greedy function approximation: a gradient boosting machine[J]. *Annals of Statistics*, 2001, 29(5): 1189–1232.
- [24] 刘伟东, 项月琴, 郑兰芬, 等. 高光谱数据与水稻叶面积指数及叶绿素密度的相关分析[J]. *遥感学报*, 2000, 4(4): 279–283.
LIU Weidong, XIANG Yueqin, ZHENG Lanfen, et al. Relationships between rice LAI, CHL D and hyperspectra data[J]. *Journal of Remote Sensing*, 2000, 4(4): 279–283. (in Chinese)
- [25] VASUDEVAN A, KUMAR D A, BHUVANESWARI N S. Precision farming using unmanned aerial and ground vehicles[C]// *Technological Innovations in ICT for Agriculture and Rural Development*. IEEE, 2016.
- [26] HAN J, SONG X, ZHOU Z, et al. Estimating leaf area index of winter oilseed rape using high spatial resolution satellite data [C]// *Fifth International Conference on Agro-geoinformatics*. IEEE, 2016.
- [27] CHANG C C, LIN C J. LIBSVM: a library for support vector machines[J]. *Neural Computation*, 2011, 13(9): 2119–2147.
- [28] CUTLER A, CUTLER D R, STEVENS J R. Random forests[J]. *Machine Learning*, 2004, 45(1): 157–176.
- [29] FRIEDMAN J H. Stochastic gradient boosting[J]. *Computational Statistics & Data Analysis*, 2002, 38(4): 367–378.
- [30] 田嫦丽, 张珣, 潘博, 等. 互联网广告点击率预估模型中特征提取方法的研究与实现[J]. *计算机应用研究*, 2017, 34(2): 334–338.
TIAN Changli, ZHANG Xun, PAN Bo, et al. Research and implementation of feature extraction methods on internet CTR prediction model[J]. *Application Research of Computers*, 2017, 34(2): 334–338. (in Chinese)