

doi:10.6041/j.issn.1000-1298.2021.S0.040

基于 CatBoost 算法与图谱特征融合的土壤全氮含量预测

王炜超 杨 玮 崔玉露 周 鹏 王 懂 李民赞

(中国农业大学现代精细农业系统集成研究教育部重点实验室, 北京 100083)

摘要: 针对高光谱技术应用于土壤养分定量检测中忽略彩色图像外部特征与土壤养分的内在关系的问题,结合土壤的光谱信息与图像特征构建一种图谱特征融合的土壤全氮含量预测模型,探究图谱特征融合对于土壤全氮含量的预测能力。通过实验室高光谱成像仪获取土壤样品的高光谱图像,从高光谱图像提取土壤的光谱信息与图像特征。使用无信息变量消除算法(Uninformative variable elimination,UVE)和竞争性自适应重加权采样算法(Competitive adaptive reweighted sampling,CARS)的联合算法对光谱信息进行特征波长的选择,选择后的特征波长作为土壤光谱信息;通过相关性分析选择与土壤全氮含量相关性较高的图像特征。将 CatBoost(Categorical Boosting)算法应用到土壤全氮含量预测中,分别对基于单一光谱信息、单一图像特征和图谱特征融合对土壤全氮含量进行预测并比较。结果表明,UVE-CARS 联合算法选取的特征波长为 942、1 045、1 199、1 305、1 449、1 536、1 600 nm,与含氮基团的倍频吸收相吻合。与土壤全氮含量相关性较高的图像特征为角二阶矩、能量、惯性矩、灰度均值和熵。通过 CatBoost 算法建立的基于单一光谱信息特征波长的模型最终预测土壤全氮含量 R^2 为 0.832 9, RMSE 为 0.203 3 g/kg;基于图像特征建立的模型最终预测土壤全氮含量 R^2 为 0.801 7, RMSE 为 0.219 7 g/kg;基于图谱特征融合建立的模型最终预测土壤全氮含量 R^2 为 0.866 8, RMSE 为 0.160 2 g/kg,预测精度均高于单一光谱特征和单一图像特征的预测精度,与基于单一光谱特征和单一图像特征相比,基于高光谱图谱特征融合的土壤全氮含量预测模型效果较好,为土壤全氮含量的预测研究提出一种新的方法。

关键词: 土壤全氮; 高光谱; CatBoost; 图谱特征融合

中图分类号: S237 文献标识码: A 文章编号: 1000-1298(2021)S0-0316-07

Prediction of Soil Total Nitrogen Based on CatBoost Algorithm and Fusion of Image Spectral Features

WANG Weichao YANG Wei CUI Yulu ZHOU Peng WANG Dong LI Minzan

(Key Laboratory of Modern Precision Agriculture System Integration Research, Ministry of Education, China Agricultural University, Beijing 100083, China)

Abstract: In order to solve the problem that the internal relationship between external features of color images and soil nutrients is ignored when hyperspectral technology is applied to quantitative detection of soil nutrients, a prediction model of soil total nitrogen content based on image and spectral features was constructed by combining the spectral information and image features of soil, and the prediction ability of image and spectral features fusion for soil total nitrogen content was explored. The hyperspectral images of soil samples were obtained by the laboratory hyperspectral imager, and the spectral information and image characteristics of soil were extracted from the hyperspectral images. The characteristic wavelength of spectral information was selected by using a joint algorithm of uninformative variable elimination (UVE) and competitive adaptive reweighted sampling (CARS), and the selected characteristic wavelength was used as soil spectral information. Through correlation analysis, image features with high correlation with soil total nitrogen content were selected. Categorical Boosting (CatBoost) algorithm was applied to the prediction of soil total nitrogen content, and the prediction of soil total nitrogen content based on single

收稿日期: 2021-07-15 修回日期: 2021-09-08
基金项目: 国家重点研发计划项目(2017YFD0201500-2017YFD0201501、2016YFD0700300-2016YFD0700304)和国家自然科学基金项目(31801265)
作者简介: 王炜超(1996—),男,硕士生,主要从事精细农业系统集成研究,E-mail: wwc_cau@163.com
通信作者: 杨玮(1981—),女,副教授,博士生导师,主要从事精细农业系统集成研究,E-mail: cauyw@cau.edu.cn

spectral information, single image feature and map feature fusion was compared. The results showed that the characteristic wavelengths selected by UVE – CARS joint algorithm were 942 nm, 1 045 nm, 1 199 nm, 1 305 nm, 1 449 nm, 1 536 nm and 1 600 nm, which were consistent with the frequency doubling absorption of nitrogen-containing groups. The image features with high correlation with soil total nitrogen content were angle second moment, energy, inertia moment, gray mean and entropy. The model based on the characteristic wavelength of single spectral information established by CatBoost algorithm finally predicted that the total nitrogen content of soil R^2 was 0.832 9 and RMSE was 0.203 3 g/kg, the model based on image features finally predicted that the total nitrogen content of soil R^2 was 0.801 7 and RMSE was 0.219 7 g/kg. And the model based on fusion of image and spectral features finally predicted that the total nitrogen content of the soil R^2 was 0.866 8, and RMSE was 0.160 2 g/kg, the prediction accuracy was higher than that of single spectral feature and single image feature. Compared with the prediction model based on single spectral feature and single image feature, the prediction model based on hyperspectral atlas feature fusion had better effect, which can provide a method for the prediction of soil total nitrogen content.

Key words: soil total nitrogen; hyperspectral; CatBoost; fusion of image spectral features

0 引言

土壤全氮含量对作物的生长、健康以及品质等有着明显的影响,同时对作物生长过程中有机物的形成起到十分关键的作用^[1-3]。目前土壤中全氮含量检测的化学方法已经十分成熟,但传统的测量全氮含量的方法需要大量的时间和繁杂的实验程序;并且存在一定的安全风险,会对环境造成污染等^[4-6]。无法做到土壤全氮含量的实时测量、在线指导大田施肥,难以保证测量的时效性、实用性。

近红外光谱仪可以检测到土壤中含 N 基团的明显光谱特征,并且含 N 基团的敏感波段与其他物质有明显的区分^[7]。但通过光谱仪技术不能获得土壤的纹理特征、空间信息、大小形状特征等,存在一定的缺陷。而高光谱图像不仅包含反映物体内部特征的光谱信息,同样包含物体外部图像特征等直观信息^[8],弥补了近红外光谱仪仅能获得物体光谱信息的缺陷。但诸多学者对高光谱图像的研究局限于反映物体内部特征光谱数据的研究,缺少对表征物体外部特征的可见光图像信息的分析。已有研究表明,土壤中全氮的主要存在形式为有机氮,土壤有机质含量与全氮含量存在高度的线性相关性,可以直接用有机质含量对全氮含量进行反演^[9];此外,土壤中有机质的敏感波长分布在可见光区间之内^[10],因此基于土壤可见光图像进行土壤全氮含量预测有一定的研究意义与探究性。

在以往的研究中,陶培峰等^[11]使用高光谱技术,对土壤养分的多个指标进行预测,运用偏最小二乘回归建立预测模型,全氮含量的预测精度 R^2 达到 0.75;王泽等^[12]使用高光谱技术,对梭梭幼苗植株

冠层的光谱数据的一阶微分变换、对数变换、对数一阶微分变换进行土壤全氮含量的反演,预测精度 R^2 达到 0.937;以上研究表明,高光谱在土壤全氮含量反演中的可行性,但均未考虑反映土壤外部特征的可见光图像与土壤全氮含量的内在关系。将图像特征与光谱信息融合作为预测模型,能够缓解近红外光谱分析技术灵敏性低的缺点,提高模型的鲁棒性和泛化能力^[13]。

本研究拟结合光谱信息与图像特征对土壤中的全氮含量进行预测。对不同全氮含量的土壤高光谱图像进行预处理,选定感兴趣区域。通过无信息变量消除算法(Uninformative variable elimination, UVE)和竞争性自适应重加权采样算法(Competitive adaptive reweighted sampling, CARS)的融合算法选取特征波长,利用光谱特征波长与彩色图像的图像特征进行特征融合,作为模型的输入,采用 CatBoost (Categorical Boosting)算法进行建模,从而对土壤全氮含量进行预测。

1 数据获取

1.1 数据采集

实验使用的土壤样本采集自北京市海淀区中国农业大学上庄实验站,土壤样品为典型的褐色土壤,具有很强的代表性。土壤采集范围为北纬 40.143 436 20°~40.143 957 38°,东经 116.215 015 2°~116.215 406 6°,取深度为除表层土壤后 1~20 cm 的土壤。采样点数为 160 个,所有样品采集时统一编号,避免人为操作造成的误差,装入牛皮纸袋中密封带回实验室。每个样品分为两份,一份用于高光谱图像采集,另一份用于土壤全氮含量测定。

1.2 高光谱图像采集与校正

土壤样品高光谱图像的采集使用 GaiaSorter 高

光谱分选仪(北京卓立汉光有限公司),型号为 Image- λ -N17E-N3 系列地物高光谱成像仪,核心部件包括均匀光源、光谱相机、电控移动平台、计算机及控制软件等,如图 1 所示。光谱波段范围为 400~1 000 nm 和 900~1 700 nm,分别有 728、256 个波段,光谱分辨率为 4 nm,有效狭缝宽度为 2 μ m,相机像素为 320 像素 $\times$ 256 像素,满足实验要求。使用 400~1 000 nm 的镜头采集土壤表面图像,使用 900~1 700 nm 的镜头采集土壤的光谱信息。采集软件为厂家提供的 SpecVIEW 软件。为了采集到较为清晰的图像,需要调节 CCD 相机镜头焦距、镜头与物体之间的距离、扫描速度和曝光时间等。经过多次反复调试,最终选定镜头与物体之间的距离为 25 cm,扫描速度为 9 mm/s,曝光时间为 5 ms。

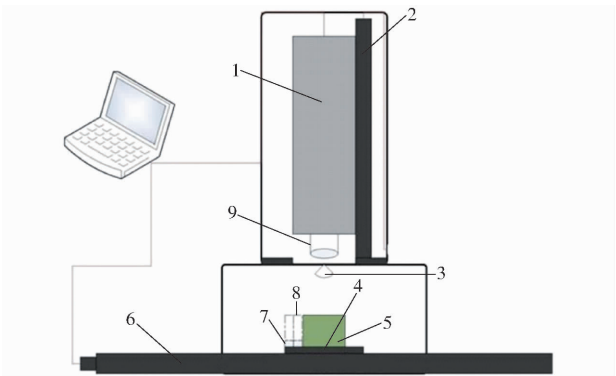


图 1 高光谱成像仪结构示意图

Fig. 1 Structure diagram of hyperspectral imager

1. 光谱相机 2. 工作距离调整机构 3. 漫反射均匀光源 4. 样品台 5. 样品 6. 电控移动平台/传送带 7. 调焦板 8. 校正白板 9. 成像镜头

为了减小光强变化和相机暗电流对图像产生的影响,需对光谱图像进行黑白板校正^[14]。通过黑白校正可以实现对所采集高光谱数据的反射率校正。高光谱相机采集的光谱数据决定于光源光谱、光源强度、镜头透过率、光谱仪的衍射效率、相机的光谱响应效率以及被测物的反射率。而最终得到反射率的高光谱数据需对采集的原始数据进行校正,校正算法为

$$R_{ci} = \frac{S_{ci} - D_{ci}}{W_{ci} - D_{ci}} \tag{1}$$

式中 S_{ci} ——原始样品数据
 D_{ci} ——暗背景数据
 W_{ci} ——白板数据
 R_{ci} ——最终得到的高光谱图像

1.3 全氮含量测定

将获得土壤样本经干燥箱 105℃ 干燥 24 h,手动捏碎大土块并去除杂质后过 60 目筛备用。将过筛后的土壤盛放到杜马斯定氮仪(荷兰 SKALAR

公司)的小坩埚中进行测量,得到土壤全氮含量的真实值,结果如表 1 所示。由表 1 可知,实验所用土壤全氮(TN)含量(质量比)分布在 0.285~2.258 g/kg 之间,样本分布合理,可以用于土壤全氮含量预测模型的建立。

表 1 土壤全氮含量

Tab.1 Data comparison of three groups of total nitrogen content sets

参数	样本数	最小值/ (g·kg ⁻¹)	最大值/ (g·kg ⁻¹)	平均值/ (g·kg ⁻¹)	标准差/ (g·kg ⁻¹)
TN 含量	160	0.285	2.258	1.318	0.460

2 研究方法

2.1 光谱信息处理

将得到的高光谱图像使用 ENVI 5.3 软件进行感兴趣区域的选取,对 100 像素 $\times$ 100 像素的像素点求取平均光谱曲线。为了消除实际操作过程中光的散射、光程畸变和随机噪声等因素对光谱造成的影响^[15],使用 S-G 卷积平滑滤波和移动平均滤波对光谱曲线进行平滑。

对土壤样本的光谱信息采用 UVE 初步筛选特征波长,然后使用 CARS 对 UVE 筛选的波长进行进一步的选择,作为最终的特征波长。UVE 是基于 PLS 回归系数 b 的算法^[16],去除不提供有用信息的变量,优选出包含有用信息较多的变量。CARS 设置采样次数为 50 次,遵循“适者生存”的原则,留下共线性较小且权重较大的变量^[17]。

2.2 土样图像特征分析

结合波谱分离程度和波段相关性的方法^[18]选择波长为 700.69 nm 的红光(R),波长为 546.39 nm 的绿光(G)和波长为 435.59 nm 的蓝光(B)处的灰度图像在 ENVI 5.3 中融合为一副 RGB 图像。 R 、 G 、 B 波长处的灰度图像与融合的彩色图像如图 2 所示,对新的 RGB 图像进行图像特征提取。使用 Matlab 2016a 分别提取 RGB 图像的熵、能量、Hu 矩、角二阶矩、平稳度、相关性、对比度、惯性矩、三阶中心距、原点矩、灰度均值等共 11 个图像特征值。利用 IBM SPSS Statistics 22 软件对 11 个图像特征值与土壤全氮含量进行相关性分析,筛选出与土壤全氮含量相关系数在 0.85 以上的图像特征值进行土壤全氮含量预测。

2.3 全氮含量模型建模方法与评价

对光谱信息与图像特征两组具有不同性质的数据,使用 Z-score 标准化方法进行数据的标准化处理,进行数据归一化处理可以将光谱信息与图像特

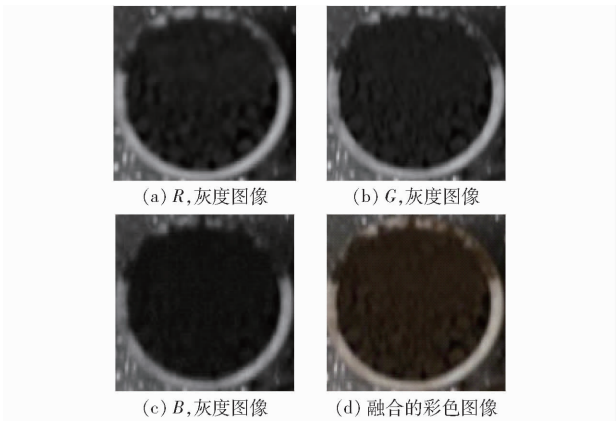


图2 R、G、B 波长处的灰度图像与融合的彩色图像
Fig. 2 Gray level image and fused color image at R, G, B wave length

征进行一定比例的缩放,符合正态分布的标准^[19],图谱特征融合后作为预测模型的输入。

采用 SPXY 算法^[20]对样品集进行划分,设定验证集个数为 30。采用 CatBoost 算法^[21]对全氮含量进行预测。CatBoost 算法是由俄罗斯 Yandex 公司于 2017 年提出的一种梯度提升算法库。CatBoost 对 GBDT 进行了优化,通过使用排序提升的方式,保证数据集均可用于训练、学习,降低了训练过拟合^[22]。CatBoost 算法通过添加先验分布项的方式改进 Greedy TS,可以减少数据集噪声和低频率类别型数据对于数据分布的影响,计算公式为

$$x_{i,k} = \frac{\sum_{j=1}^{p-1} [x_{sj,k} = x_{sp,k}] Y_j + ap}{\sum_{j=1}^{p-1} [x_{sj,k} = x_{sp,k}] + a} \tag{2}$$

式中 p ——添加的先验项
 a ——权重系数

经过多次调试,iterations (最大树数)设置为 100,depth (树深)设置为 6,learning_rate (学习率)设置为 0.03,其他值为默认值。

对所建立模型采用决定系数 R^2 、均方根误差 (RMSE) 进行评价。

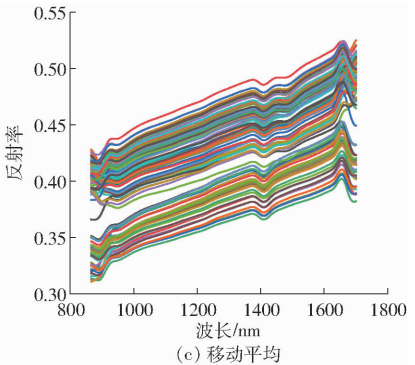
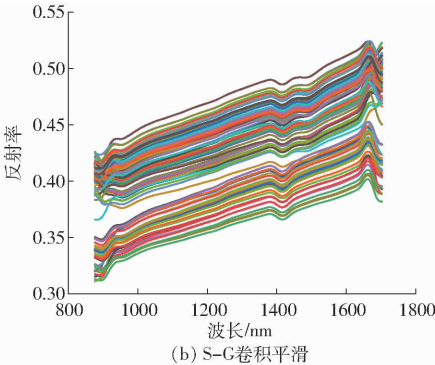
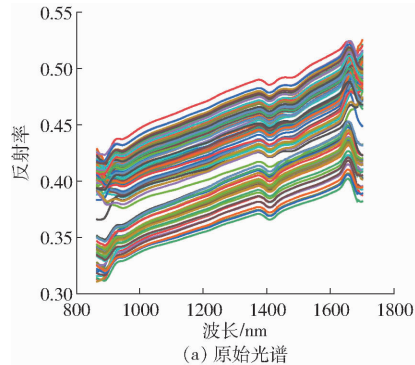


图3 原始光谱曲线与平滑光谱曲线

Fig. 3 Original spectral curves and smooth spectral curves

3 结果与讨论

3.1 基于光谱信息的全氮含量检测

3.1.1 光谱分析

采集的土壤光谱信息如图 3a 所示。从图 3a 可以发现,不同全氮含量的土壤光谱曲线有明显的区别,其中全氮含量较高的土壤光谱曲线的反射率较全氮含量较低的土壤光谱曲线反射率高。因为 N 元素与含 N 基团在可见近红外波段处对光照存在合频与倍频吸收^[23],不同的全氮含量对光照的吸收程度不同。

从图 3a 可以发现,由于操作过程中的误差、光散射的影响,光谱曲线有一定的噪声。使用 S-G 卷积平滑滤波、移动平均滤波两种平滑方法进行滤波,结果如图 3b、3c 所示。从图 3 可以看出,移动平均滤波对光谱原始信息的保留比较好,但平滑效果较差;而 S-G 卷积平滑在光谱曲线平滑与光谱原始信息的保留上效果都比较好,因此选择 S-G 卷积平滑对光谱曲线进行滤波。

3.1.2 特征波长选择

特征变量的选择在光谱定量分析中占有关键地位,它影响到建模的结果及其性能^[24]。对 3.1.1 节所滤波的光谱信息进行特征波长的选择。使用 UVE 和 CARS 联立算法 (UVE-CARS) 从全谱段 254 个变量中选择特征波长。在 UVE 特征波长的选择过程中,把经过滤波的光谱数据和测定的全氮含量作为参数输入模型,通过对不同偏最小二乘法 (PLS) 校准模型预测集均方根误差 (RMSEP) 的变化确定最优的特征波长。RMSEP 最低的波长作为最优的波长。经 UVE 变量筛选后,共保留 101 个波长变量。在 UVE 选择波长的基础上,使用 CARS 方法进行进一步的特征变量筛选。CARS 进行特征变量选择如图 4 所示。

图 4a 中曲线呈递减趋势。图 4b 中校正集均方根误差 (RMSECV) 呈先减小后变大的趋势,说明先

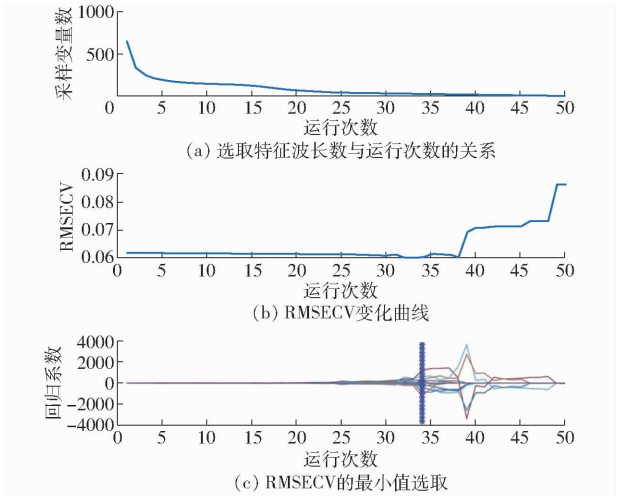


图 4 CARS 方法特征波长选择过程

Fig. 4 Characteristic wavelength selection process of CARS method

剔除了无用的特征变量,后剔除了有用变量。图 4c 中当运行次数为 34 时, RMSECV 最小,此刻选取了最优的波长组合。

最终通过 UVE - CARS 方法选择的特征波长为 942、1 045、1 199、1 305、1 449、1 536、1 600 nm。已有实验表明,光照下土壤中的 N—H 官能团的一级倍频在 1 500 nm 附近、二级倍频在 1 100 nm 附近、三级倍频在 900 nm 附近。本研究选取的 7 个特征波长与 N—H 官能团的倍频吸收相吻合,与土壤全氮含量有着紧密的联系,因此特征波长的选取对土壤全氮含量的预测有着重要的意义。

3.1.3 全氮含量模型建立

对筛选出的 7 个光谱特征波长 (942、1 045、1 199、1 305、1 449、1 536、1 600 nm) 与全谱段作为参数输入 CatBoost 算法,分别构建土壤全氮含量预测模型,结果见图 5。

从图 5 可以看出,特征波长作为输入的预测 R^2 为 0.832 9, RMSE 为 0.203 3 g/kg; 全谱段作为输入的预测 R^2 为 0.856 5, RMSE 为 0.179 9 g/kg。特征波长预测土壤全氮含量的精度比全谱段低,因为全谱中存在大量冗余信息和多重共线性的波段,导致模型的过拟合,影响预测结果。特征波长更能代表含 N 基团对光照吸收的影响,排除了干扰信息和冗余、共线性信息的影响,提高了模型的可信度与实用性、降低了建模时间。因此特征波长进行预测土壤全氮含量比全谱段的预测更有应用价值。

3.2 基于图像特征的全氮含量检测

3.2.1 土样图像特征提取与选择

高光谱图像不仅包含被测物体的光谱信息,同样包括物体的图像信息。 $R、G、B$ 波段处的灰度图像是在单个电磁波频谱内测量每个像素的亮度得到

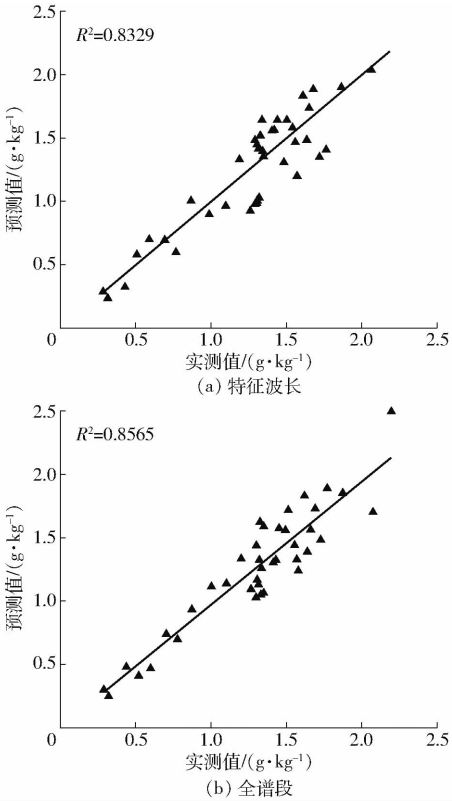


图 5 土壤全氮含量光谱信息预测散点图

Fig. 5 Scatter plots of spectral information prediction

的,而含 N 基团会影响土壤对于光照的吸收,导致相机得到的不同全氮含量土壤图像的灰度分布不同,因此土壤的图像特征与土壤全氮含量存在一定的内在联系。

将 $R、G、B$ 3 个波段下的灰度图像进行融合,形成一幅新的彩色图像。采用灰度共生矩阵的方法^[25]提取图像的纹理特征,并结合图像的 $R、G、B$ 3 个波段的灰度均值,共提取了 11 个图像特征,分别是图像的熵、能量、Hu 矩、角二阶矩、平稳度、相关性、对比度、惯性矩、三阶中心距、原点矩、处灰度均值。

在 SPSS 软件中进行相关性分析,最终得到与土壤全氮含量相关系数在 0.85 以上的 5 个图像特征 (角二阶矩、能量、惯性矩、灰度均值和熵)。

3.2.2 基于土样图像特征的土壤全氮含量模型

将筛选出的 5 个图像特征作为参数输入 CatBoost 算法,构建土壤全氮含量预测模型,结果见图 6。

从图 6 可以看出,基于图像特征建立的预测模型 R^2 为 0.801 7, RMSE 为 0.219 7 g/kg,在一定程度上比基于光谱信息建立的预测模型结果差。因为含 N 基团在可见光区域敏感性较低,而光谱信息在近红外区域,对于 N 元素和含 N 基团的敏感性较高,吸收较明显。从预测结果可以看出,预测的 R^2 虽然低于光谱信息,但 R^2 在 0.80 以上。而且图像特征在一定程度上可以弥补光谱信息的不足,突出图像

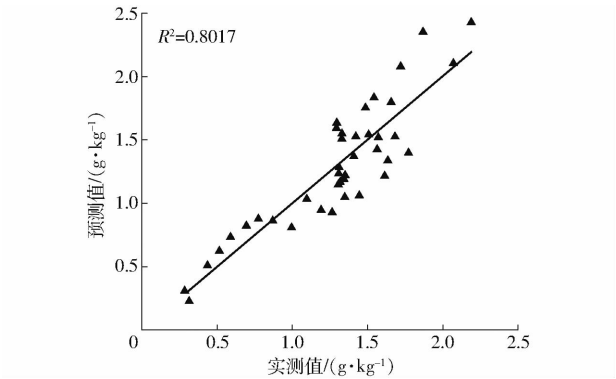


图 6 土壤全氮含量图像特征预测散点图

Fig. 6 Scatter plot of image feature prediction

特征在土壤全氮含量模型中的贡献,因此具有一定的可行性与研究价值。

3.3 基于图谱融合特征的全氮含量检测

3.3.1 数据融合

对采集的土壤光谱信息与图像特征进行数据归一化处理,处理结果如图 7 所示。数据归一化将不同性质的光谱数据与图像特征两组数据经过变换,成为纯量,用于数据的融合。

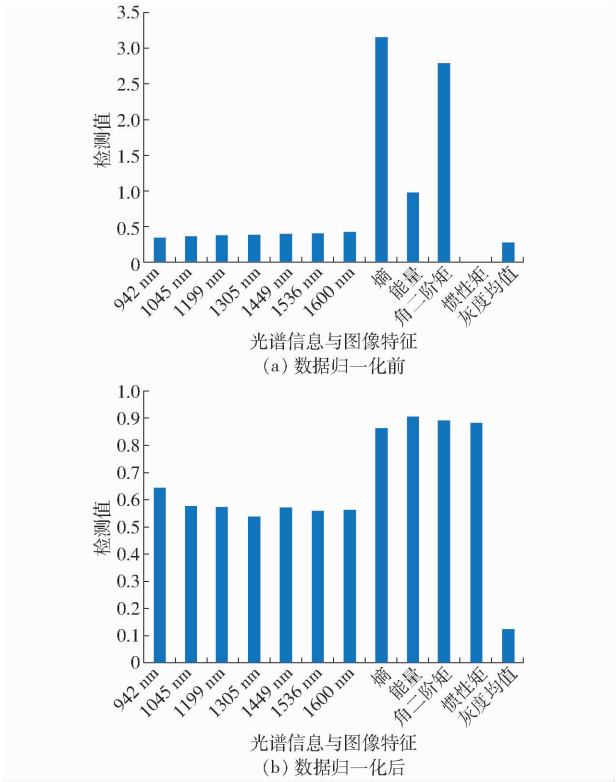


图 7 数据融合对比

Fig. 7 Data fusion comparison

从图 7 可以看出,归一化之前的数据标准差为 1.022,归一化之后的数据标准差为 0.223,说明将光谱数据与图像特征两组不同性质的数据映射到了统一的区间之内。将光谱数据与图像特征进行数据融合不仅可以提高模型的收敛速度,同时可以提高模型的预测精度。

3.3.2 预测模型建立

将图谱融合特征作为参数输入 CatBoost 算法,构建土壤全氮含量预测模型,结果如图 8 所示,不同的参数输入结果见表 2。

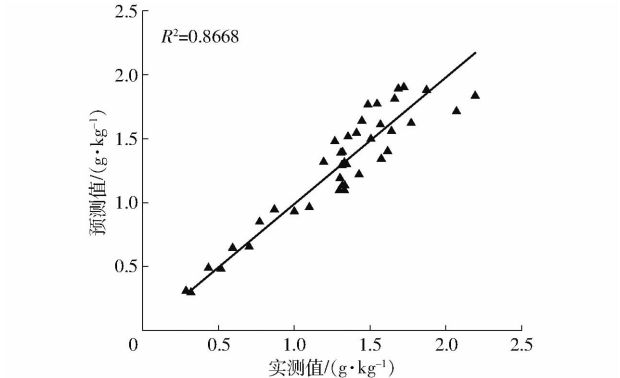


图 8 土壤全氮含量图谱特征融合预测散点图

Fig. 8 Scatter plot of joint prediction of spectral characteristics

表 2 不同输入参数的 CatBoost 算法预测土壤全氮含量结果

Tab.2 Results of CatBoost algorithm with different input parameters for predicting soil total nitrogen content				
输入参数	特征波长	全谱段	图像特征	图谱特征融合
R^2	0.832 9	0.856 5	0.801 7	0.866 8
RMSE/(g·kg ⁻¹)	0.203 3	0.179 9	0.219 7	0.160 2

从表 2 可以看出,基于图谱融合特征建立的预测模型 R^2 为 0.866 8, RMSE 为 0.160 2 g/kg,均在一定程度上比基于单一光谱信息和基于单一图像特征建立的预测模型结果好。说明图谱特征融合在反演土壤全氮含量中更具优势,充分考虑了光谱信息与图像特征的贡献,缓解了光谱信息建模时存在的“同物异谱,同谱异物”现象,提高了模型精度。

本研究选取的图像特征波段来自于可见光 R 、 G 、 B 波段处的灰度图像,没有充分考虑到各个波段处灰度图像的图像特征。在今后的研究中需进一步研究高光谱图像外部特征与土壤全氮含量的关系,建立更有效、精确的预测模型。

4 结论

(1) 基于 UVE - CARS 方法选取的特征波长建模 R^2 为 0.832 9, RMSE 为 0.203 3 g/kg,预测结果精度略低于全谱段建模,但特征波长的选取更能代表含 N 基团在光照吸收中的吸收特征,而且可以滤除全谱信息中存在的大量干扰信息。

(2) 在图谱特征融合建模中, R^2 为 0.866 8, RMSE 为 0.160 2 g/kg。预测结果均优于单一光谱信息建模、单一图像特征建模。

(3) 基于 CatBoost 算法建立的模型最终预测 R^2 均高于 0.80, 表明 CatBoost 算法在土壤全氮含量预测中具有一定的实用价值; 图谱特征融合相对于单

一的基于光谱信息与图像特征对于土壤全氮含量的预测在精度上有一定的提升, 图谱特征融合的处理方法对于土壤全氮含量的预测能力更强。

参 考 文 献

[1] LUO Y Q, ZHAO X Y, WANG T, et al. Characteristics of the plant-root system and its relationships with soil organic carbon and total nitrogen in a degraded sandy grassland[J]. Acta Prataculturae Sinica, 2017, 26(8):200–206.

[2] SONG X D, YANG F, JU B, et al. The influence of the conversion of grassland to cropland on changes in soil organic carbon and total nitrogen stocks in the Songnen plain of Northeast China[J]. CATENA, 2018, 171:588–601.

[3] SAINJU U M, LENNSEN A W, ALLEN B L, et al. Soil total carbon and nitrogen and crop yields after eight years of tillage, crop rotation, and cultural practice [J]. Heliyon, 2017, 3(12):e00481.

[4] 李民赞, 姚向前, 杨玮, 等. 基于卤钨灯光源和多路光纤的土壤全氮含量检测仪研究[J]. 农业机械学报, 2019, 50(11): 169–174.

LI Minzan, YAO Xiangqian, YANG Wei, et al. Design of new portable detector for soil total nitrogen content based on high-power tungsten halogen lamp and “one-six” special optical fiber[J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(11):169–174. (in Chinese)

[5] 周鹏, 杨玮, 李民赞, 等. 基于灰度关联-极限学习机的土壤全氮预测[J]. 农业机械学报, 2017, 48(增刊):271–276.

ZHOU Peng, YANG Wei, LI Minzan, et al. Soil total nitrogen content prediction based on gray correlation-extreme learning machine[J]. Transactions of the Chinese Society for Agricultural Machinery, 2017, 48(Supp.):271–276. (in Chinese)

[6] YAO Xiangqian, YANG Wei, LI Minzan, et al. Prediction of total nitrogen in soil based on random frog leaping wavelet neural network[J]. IFAC PapersOnLine, 2018, 51(17):660–665.

[7] DEBAENE G, NIEDZWIECKI J, PECIO A, et al. Effect of the number of calibration samples on the prediction of several soil properties at the farm-scale[J]. Geoderma: an International Journal of Soil Science, 2014, 214/215:114–125.

[8] 张珏, 田海清, 王轲, 等. 基于高光谱成像技术的羊肉新鲜度预测[J]. 中国农业大学学报, 2020, 25(5):94–103.

ZHANG Yu, TIAN Haiqing, WANG Ke, et al. Nondestructive detection of lamb freshness based on hyperspectral imaging technology[J]. Journal of China Agricultural University, 2020, 25(5):94–103. (in Chinese)

[9] 中科院南京土壤所. 土壤理化分析[M]. 上海: 上海科技出版社, 1981.

[10] NELIDA E, QUINONEZ S, LUIS A, et al. Effects of water, organic matter, and iron forms in mid-IR spectra of soils: assessments from laboratory to satellite-simulated data[J]. Geoderma, 2020, 375:114480.

[11] 陶培峰, 王建华, 李志忠, 等. 基于高光谱的土壤养分含量反演模型研究[J]. 地质与资源, 2020, 29(1):68–75, 84.

TAO Peifeng, WANG Jianhua, LI Zhizhong, et al. Research of soil nutrient content inversion model based on hyperspectral data[J]. Geology and Resources, 2020, 29(1):68–75, 84. (in Chinese)

[12] 王泽, 轩俊伟, 任财, 等. 基于高光谱的梭梭幼苗生境土壤氮素含量估算模型研究[J]. 广东农业科学, 2020, 47(1):68–76.

WANG Ze, XUAN Junwei, REN Cai, et al. Study on estimation model of soil nitrogen content in *Haloxylon ammodendron* seedlings habitats based on hyper-spectral data[J]. Guangdong Agricultural Science, 2020, 47(1):68–76. (in Chinese)

[13] 祖琴, 张水发, 曹阳, 等. 结合光谱图像技术和 SAM 分类法的甘蓝中杂草识别研究[J]. 光谱学与光谱分析, 2015, 35(2):479–485.

ZU Qin, ZHANG Shuifa, CAO Yang, et al. Research on identification of cabbages and weeds combining spectral imaging technology and SAM taxonomy[J]. Spectroscopy and Spectral Analysis, 2015, 35(2):479–485. (in Chinese)

[14] 陈志莉, 刘强, 尹文琦, 等. 高光谱成像砖红壤中石油烃含量的可视化研究[J]. 光谱学与光谱分析, 2018, 38(9):2916–2922.

CHEN Zhili, LIU Qiang, YIN Wenqi, et al. Visualization of petroleum hydrocarbon content in latosol based on hyperspectral imaging technology[J]. Spectroscopy and Spectral Analysis, 2018, 38(9):2916–2922. (in Chinese)

[15] QIN J, CHAO K, KIM M S, et al. Hyperspectral and multispectral imaging for evaluating food safety and quality[J]. Journal of Food Engineering, 2013, 118(2):157–171.

[16] CENTER V, MASSART D L, DE NOORD O E, et al. Elimination of uninformative variables for multivariate calibration [J]. Analytical Chemistry, 1996, 68(21):3851–3858.

[17] LI H, LIANG Y, XU Q, et al. Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration[J]. Analytica Chimica Acta, 2009, 648(1):77–84.

[18] 刘建波. 高光谱遥感图像最佳波段组合的选择[C] // 第十一届全国遥感技术学术交流会论文集, 1999:218–222.

[19] 牛岩, 魏雨露, 刘思琪. 数据标准化方法对 SVM 数据预测结果的影响研究[J]. 西部大开发(土地开发工程研究), 2020, 5(4):18–23.

[20] ROBERTO K H G, MARCIO J C P, EDVAN C S, et al. A method for calibration and validation subset partitioning[J]. Talanta: the International Journal of Pure and Applied Analytical Chemistry, 2005, 67(4):736–740.

[21] WU Lifeng, HUANG Guomin, FAN Junliang, et al. Potential of kernel-based nonlinear extension of Arps decline model and gradient boosting with categorical features support for predicting daily global solar radiation in humid regions[J]. Energy Conversion and Management, 2019, 183:280–295.

[22] TIEN D P, NAOTO Y, JUNSHI X, et al. Comparison of machine learning methods for estimating mangrove above-ground biomass using multiple source remote sensing data in the Red River delta biosphere reserve, Vietnam[J]. Remote Sensing, 2020, 12(8):1334.

[23] STENBERG B, VISCARRA R R A, MOUAZEN A M, et al. Visible and near infrared spectroscopy in soil science[J]. Advances in Agronomy, 2010, 107:163–215.

[24] 杨仁欣, 杨燕, 原晶晶. 高光谱图像的特征提取与特征选择研究[J]. 广西师范学院学报(自然科学版), 2015, 32(2): 39–43.

[25] HARALICK R M, SHANMUGAM K, DINSTEN I. Textural features for image classification[J]. IEEE Transactions on Systems, Man and Cybernetics, 1973, 3(6):768–780.