

高光谱和 NSGA2-ELM 算法的粳稻叶片氮素含量反演

冯 帅¹, 曹英丽^{1,2*}, 许童羽^{1,2}, 于丰华^{1,2}, 陈春玲^{1,2}, 赵冬雪¹, 金 彦¹

1. 沈阳农业大学信息与电气工程学院, 辽宁 沈阳 110161

2. 沈阳农业大学辽宁省农业信息化工程技术中心, 辽宁 沈阳 110161

摘 要 为提供一种高效、快速和无损的粳稻叶片氮素含量反演方法,以粳稻小区试验为基础,利用高光谱技术和室内化学实验,获取粳稻分蘖期、拔节期和抽穗期三个生育期共 280 组叶片高光谱数据以及相对应的水稻叶片氮素含量数据,分析不同施氮水平的粳稻叶片光谱特征,采用随机青蛙算法(random_frog)与迭代和保留信息变量算法(IRIV)相结合的方式筛选特征波段,并将任意两个光谱波段随机组合构建差值植被指数 $DSI(R_i, R_j)$ 、比值植被指数 $RSI(R_i, R_j)$ 和归一化植被指数 $NDSI(R_i, R_j)$, 分别将较优的特征波段组合和植被指数组合作为模型输入,构建 BP 神经网络、支持向量机(SVR)和非支配的精英策略遗传算法优化极限学习机(NSGA2-ELM)粳稻叶片氮素含量反演模型,并对模型进行验证分析。结果表明:随着施氮水平的增加,粳稻叶片近红外波段范围反射率逐渐升高,在可见光波段范围反射率逐渐降低。采用 random_frog 与 IRIV 相结合的方式筛选特征波段共得到 8 个特征波段,其中可见光波段 7 个,分别为 414.2, 430.9, 439.6, 447.9, 682.7, 685.4 和 686.3 nm,近红外波段仅有 1 个为 999.1 nm,该方法较好地剔除了干扰信息,大大降低了波段间的共线性。同时从三种植被指数($DSI(R_i, R_j)$, $RSI(R_i, R_j)$, $NDSI(R_i, R_j)$)与粳稻叶片氮素含量的决定系数等势图中可知, $DSI(R_{648.1}, R_{738.1})$, $RSI(R_{532.8}, R_{677.3})$ 和 $NDSI(R_{654.8}, R_{532.9})$ 与叶片氮素含量相关性最好, R^2 分别为 0.811 4, 0.829 7 和 0.816 9。在输入参量不同的建模效果对比分析中,以特征波段组合作为模型输入所构建的模型反演效果略优于植被指数组合, R^2 均大于 0.7, RMSE 均小于 0.57。而在反演模型间的对比分析中,提出的 NSGA2-ELM 反演模型的估测效果要优于 BP 神经网络模型和 SVR 模型,训练集决定系数 R^2 为 0.817 2,均方根误差 RMSE 为 0.355 5,验证集 R^2 为 0.849 7, RMSE 为 0.301 1。鉴于此,random_frog-IRIV 筛选特征波段方法结合 NSGA2-ELM 建模方法在快速检测粳稻叶片氮素含量中具有显著优势,可为粳稻田间精准施肥提供了参考。

关键词 高光谱数据; 叶片氮素含量; 特征波段; 植被指数; 反演模型

中图分类号: O657.3 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2020)08-2584-08

引 言

氮素是农作物生长发育过程中重要的营养成分,实时监测和评估农作物的氮素含量对于农作物田间精准管理和长势预测等均具有十分重要的意义^[1]。

目前,采用高光谱检测法对果蔬^[2-3]和粮食作物^[4-6]的氮素营养诊断已成为国内外学者研究的主要内容。王树文^[7]等研究表明基于主成分分析和相关分析结合多元回归分析模型的差值指数、多变量单波段指数等模型反演效果较好,预测集 R^2 为 0.869, RMSE 为 0.085。刘明博^[8]等采用连续投影

法(SPA)筛选的有效波段、光谱指数 RVI、NDVI 以及全光谱波段构建多种水稻叶片氮素含量反演模型。对比发现,基于 SPA 有效波段构建的模型的估测效果明显优于光谱指数所建,但略差于全光谱波段所建模型。Tian^[9]等通过分析多种高光谱植被指数与水稻叶片氮素含量的定量关系,得出采用绿色比率指数 $SR(R_{553}, R_{537})$ 反演叶片氮素含量具有最佳估测精度。Du^[10]等采用高光谱激光雷达(HSL)技术构建两种积分指数 NOAC 和 RII 反演水稻叶片全氮含量(LNC)。方美红^[11]等采用小波系数构建水稻叶片氮含量反演模型,研究表明该模型有较高估测精度,预测值与估测值的复相关系数高达 0.99,显著优于传统光谱指数反演模型。

收稿日期: 2019-09-17, 修订日期: 2019-12-20

基金项目: 国家“十三五”重点研发计划项目(2016YFD0200600, 2016YFD0200603)资助

作者简介: 冯 帅, 1992 年生, 沈阳农业大学信息与电气工程学院博士研究生 e-mail: int_crazy@163.com

* 通讯联系人 e-mail: caoyingli@syau.edu.cn

水稻生长过程中受氮肥影响较为明显,不同施氮水平下的水稻生长趋势有一定的差异,叶片对自然光的吸收和反射也表现出不同的变化,进而对水稻产量产生较大影响。鉴于此,本研究为实现水稻叶片氮素含量的高效、快速和精准反演,分别采用特征波段组合和光谱植被指数组合作为输入,构建 3 种氮素含量反演模型并对比分析,进而提出基于高光谱的梗稻叶片氮素含量反演方法,以期为梗稻氮素营养诊断和田间精准管理提供科学依据和理论支持。

1 实验部分

1.1 试验设计

梗稻小区试验于 2018 年 6 月至 8 月在辽宁省沈阳市沈河区沈阳农业大学南区水稻试验基地(118°53'E, 38°43'N, 平均海拔 40 m)开展,供试品种为沈稻 9816。共划分 12 个小区,每个试验小区面积为 30 m² (7.61 m×4.20 m),如图 1 所示。人为设置小区试验无肥、低氮、正常和高氮 4 种情况,共设 4 个施氮梯度,分别为 N0(不含氮)、N1(150 kg·hm⁻²)、N2(240 kg·hm⁻²)、N3(330 kg·hm⁻²),每个水平 3 次重复。同时各试验小区之间采取隔离措施,保证小区之间水肥不互相渗透,其他田间管理水平按当地正常水平进行。

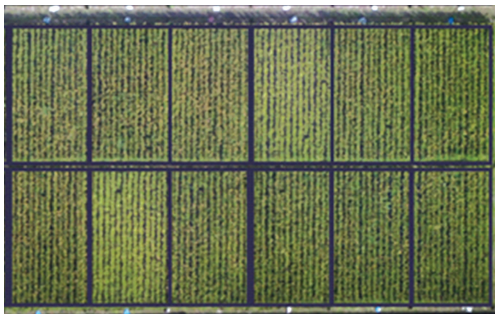


图 1 12 个梗稻小区分布图

Fig. 1 Distribution of 12 rice plots

1.2 叶片高光谱数据采集

采用美国海洋光学公司生产的光纤光谱仪 HR2000+,其光谱范围为 400~1 000 nm,光谱分辨率为 0.45 nm。按标号将磨碎的叶片粉末放置于操作平台,光谱探头紧压在叶片粉末上;测量前通过黑白板校正。通过自带 OceanView 软件完成梗稻叶片高光谱数据的采集。

1.3 叶片氮素含量测定

从每个试验小区采样点获取梗稻不同部位叶片 20 片左右,分别装入自封袋中并标注小区名称和编号,立即带回实验室。在室内,首先对叶片进行洗涤,去除叶片表面灰尘等无用物质,然后在 105 °C 条件下杀青 30 min,并在 70 °C 的烘箱中烘干至恒量,称量粉碎。最后采用凯氏定氮法测定梗稻叶片氮素含量。所采集的数据样本中去粗大误差,最终共得到梗稻叶片氮素含量有效数据 280 组,其概率密度函数如图 2 所示。

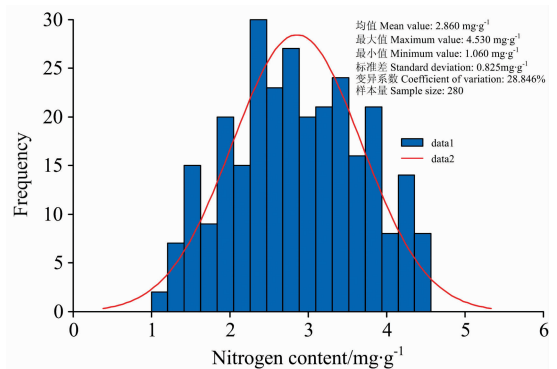


图 2 梗稻 280 组叶片氮素含量的概率密度函数

Fig. 2 Probabilistic density function of nitrogen content in 280 groups of japonica rice leaves

由图 2 可知,280 组梗稻叶片氮素含量数据呈正态分布,均值为 2.860 mg·g⁻¹,最大值为 4.530 mg·g⁻¹,最小值为 1.060 mg·g⁻¹,标准差为 0.825 mg·g⁻¹,变异系数为 28.846%,满足氮素含量反演要求。同时采用 Kennard-Stone 算法(KS)将样本按照训练集与验证集 3:1 的比例进行划分,其氮素含量统计表如表 1 所示。

表 1 训练集与验证集梗稻叶片氮素含量数据统计表

Table 1 Statistical table of nitrogen content in leaves of japonica rice in training set and validation set

数据集	样本量	均值/ (mg·g ⁻¹)	最大值/ (mg·g ⁻¹)	最小值/ (mg·g ⁻¹)	标准差/ (mg·g ⁻¹)	变异 系数/%
训练集	210	2.878	4.530	1.060	0.838	29.117
验证集	70	2.804	4.470	1.550	0.788	28.103

1.4 光谱特征参数的筛选

1.4.1 光谱特征波段筛选方法

高光谱数据具有较高的分辨率,因此在光谱数据中包含了大量的数据信息,但在庞大的数据信息中仍存在一定的冗余信息,给后续的研究带来了一定困扰。筛选特征波段可减少数据冗余,获得对模型起关键性作用的特征波段是不可

或缺的。分别采用随机青蛙法(random frog)与迭代和保留信息变量算法(iteratively retaining informative variables, IRIV)两种方法筛选特征波段。随机青蛙法(random frog)采用蒙特卡洛法通过维度转换进行采样,以每个波段变量的选择频率作为剔除冗余变量的依据,最终得到最佳特征波段。迭代和保留信息变量算法是基于二进制矩阵重排过滤器提出的特征

变量选择算法, 通过多次迭代, 计算包含某一波段和不包含该波段变量时的 RMSECV 平均值, 得到两个均值之差 DMEAN 和曼-惠特尼检验的 P 值, 从而判断波段的重要性, 其判断规则如表 2 所示^[12], 并依据该规则消除无信息和干扰波段变量, 保留强和弱波段变量, 最后通过反向消除获得最佳特征波段。

表 2 迭代和保留信息变量算法判断规则

Table 2 The judgment rule of iteratively retaining informative variables

波段类型	判断规则
强波段变量	DMEAN < 0, $P < 0.05$
弱波段变量	DMEAN < 0, $P > 0.05$
无信息波段变量	DMEAN > 0, $P > 0.05$
干扰波段变量	DMEAN > 0, $P < 0.05$

1.4.2 光谱植被指数的构建

植被指数常常被用来精准监测植被生长状况和生物量信息等植被信息, 同时具有增强植被信息, 减弱非植被信息的作用, 因此选取 3 种对水稻叶片氮素含量具有较好反演效果的高光谱植被指数, 即比值植被指数 (ratio spectral index, RSI)、差值植被指数 (difference spectral index, DSI) 以及归一化植被指数 (normalized difference spectral index, NDSI)。其计算公式如式(1)–式(3)

$$RSI = R_i / R_j \quad (1)$$

$$DSI = R_i - R_j \quad (2)$$

$$NDSI = (R_i - R_j) / (R_i + R_j) \quad (3)$$

式中, R_i 与 R_j 分别表示在 400~1 000 nm 范围内的光谱反射率值。

1.5 反演模型的构建与评价标准

分别将特征波段组合和植被指数作为模型输入, 实测水稻叶片氮素含量作为输出, 构建 BP 神经网络、支持向量机和 NSGA2-ELM 三种反演模型。由于 ELM 模型采用随机生成输入权重和隐含层偏向值, 这将导致 ELM 输出波动大, 模型不稳定。因此提出一种将 NSGA2 算法与 ELM 模型相结合的混合模型算法。具体 NSGA2 优化过程如下:

(1) 根据样本数据集确定 ELM 的网络拓扑结构, 将神经元之间的权重和偏向值构成实数向量, 用以表示种群 M 中的个体。同时随机生成实数向量的初始值构成大小为 N 的第一代父代种群 P 。

(2) 对父代种群进行非支配排序, 并采用传统的遗传算法对父代种群进行选择、交叉和变异操作产生大小为 N 的子代种群 $P1$ 。将种群 P 和 $P1$ 合并为大小为 $2N$ 的种群 B 。

(3) 对种群进行非支配排序, 获得非支配解的前端 F_t , 即为非支配面的 $F1$, $F2$ 和 $F3$, 并计算拥挤度。之后采用精英保留策略筛选最优个体, 即由于子代和父代种群个体均包含在种群 B 中, 则采用非支配排序后的 $F1$ 中的个体为种群 B 中最佳的, 因此现将 $F1$ 全部个体放入新父代种群 $P3$ 中。若 $P3$ 小于 N , 则继续将 $F2$ 中的个体加入种群 $P3$ 中。若 $P3$ 仍小于 N , 则对 $F3$ 进行拥挤度排序, 取 $N - |P3|$ 个种群个

体添加至 $P3$, 直至 $P3$ 种群大小为 N 。然后采用遗传算法对种群 $P3$ 进行选择、交叉和变异产生新的种群 $P4$ 。

(4) 重复 n 次上述计算过程, 达到设定的最大迭代次数则停止迭代, 得到最佳 ELM 的最佳权重和偏向值, 完成优化。

同时为了检测水稻叶片氮素含量反演模型的准确性和可靠性, 采用决定系数 R^2 和均方根误差 RMSE 两个评价指标检测模型的拟合效果和估测能力。其计算公式如式(4)和式(5)

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (4)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (5)$$

式中, y_i 为真实值; \hat{y}_i 为估测值; \bar{y} 为平均值; n 为样本数量。

整体技术路线如图 3 所示。

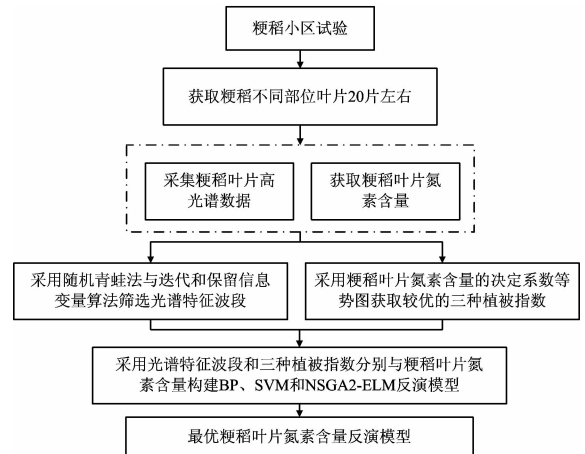


图 3 研究技术路线图

Fig. 3 Research technology roadmap

2 结果与讨论

2.1 水稻叶片氮素含量的光谱分析

由图 4 可知, 在不同施氮水平下的水稻叶片光谱反射率曲线变化趋势大致相同。在可见光波段范围 (400~770 nm) 光谱反射率均呈现出“先升后降”的变化规律。同时, 在绿波段 (550 nm 附近) 出现明显的反射峰, 在红波段 (680 nm 附近) 出现明显的吸收谷。在 680~770 nm 范围内不同施氮水平的光谱反射率曲线变化基本一致, 反射率急剧升高。在近红外波段范围 (>770 nm) 形成较高的反射平台。然而, 在可见光和近红外波段范围, 各个施氮水平的叶片光谱反射率表现出相反的变化规律。在可见光波段范围, 光谱反射率随着施氮水平的升高而减小, 尤其在“红峰”处极其明显, 在近红外波段范围, 光谱反射率随着施氮水平的升高而增大, 尤其在“近红外反射平台”处较为明显。

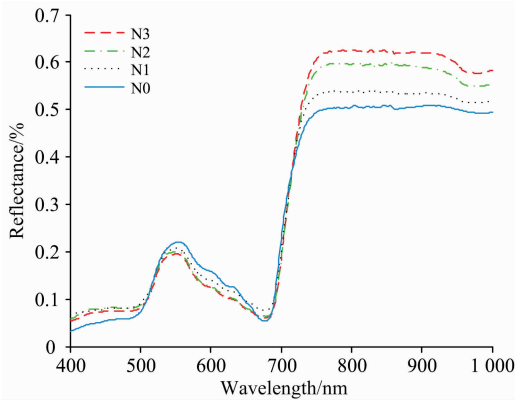


图 4 不同施氮水平下的粳稻叶片平均光谱反射率

N0: 不含氮; N1: 150 kg · hm⁻²;
N2: 240 kg · hm⁻²; N3: 330 kg · hm⁻²

Fig. 4 Average spectral reflectance of japonica rice leaves under different nitrogen levels

under different nitrogen levels

N0: Nitrogen-free; N1: 150 kg · hm⁻²;
N2: 240 kg · hm⁻²; N3: 330 kg · hm⁻²

2.2 光谱特征波段的筛选

鉴于高光谱的波段数较多,且存在无用和干扰波段,对建模的效率和精度产生较大影响。因此首先采用随机青蛙算法(random-frog)初选特征波段,去除冗余信息。确定 random-frog 算法的潜在变量的最大数为 6,初始采样变量数为 1 000,但由于 random-frog 算法以蒙特卡洛法为筛选原理,每次筛选出的特征波段略有差异。因此共运行 random-frog 算法 50 次,最后取结果的平均值作为特征波段的判断依据,其结果如图 5 所示。由图 5 可知,random-frog 算法筛选的结果较好,大部分光谱波段的选样概率较小,小部分光谱波段的选样概率较大。同时,设定筛选阈值为 0.1,选样概率大于 0.1 的波段被认定为特征波段,最终筛选出 23 个波段作为特征波段。

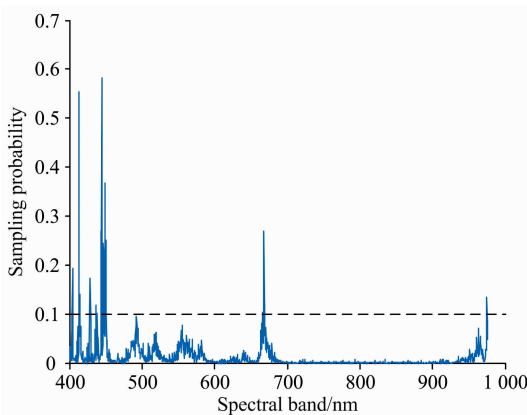


图 5 随机青蛙算法筛选结果图

Fig. 5 The screening result graph of random frog

random-frog 方法筛选得到 23 个特征波段,波段数量较大,且 random-frog 方法在筛选过程具有随机性,每次的筛选结果略有差异,可能仍存在无信息波段。因此,为提高模

型的稳健性和准确性,采用迭代和保留信息变量算法(IRIV)对波段进行进一步筛选,消除冗余信息和无用信息。

通过多次测试确定 IRIV 算法的最大主因子数为 10,交叉验证次数为 5。IRIV 算法筛选特征波段共进行 3 轮,如图 6(a)所示。前 2 轮波段变量个数逐渐减少,从 23 个波段减少到 10 个,其中在第 2 轮迭代过程中基本剔除了无用信息波段和干扰波段,余下波段的 DMEAN 和 P 值如图 6(b)所示,在第 3 轮采用反向消除,最终得到 8 个特征波段,其中弱波段变量仅有 1 个为 439.6 nm,强波段变量仅有 7 个,分别为 414.2, 430.9, 447.9, 682.7, 685.4, 686.3 和 999.1 nm。此时,大部分波段集中于 400~700 nm 之间,即在可见光波段区域之间有 7 个特征波段,仅有 1 个特征波段位于近红外波段区域。从筛选结果来看,虽然各个波段间仍存在少量的共线性,但相比于初次筛选,已较好地降低了波段间的共线性和冗余信息。

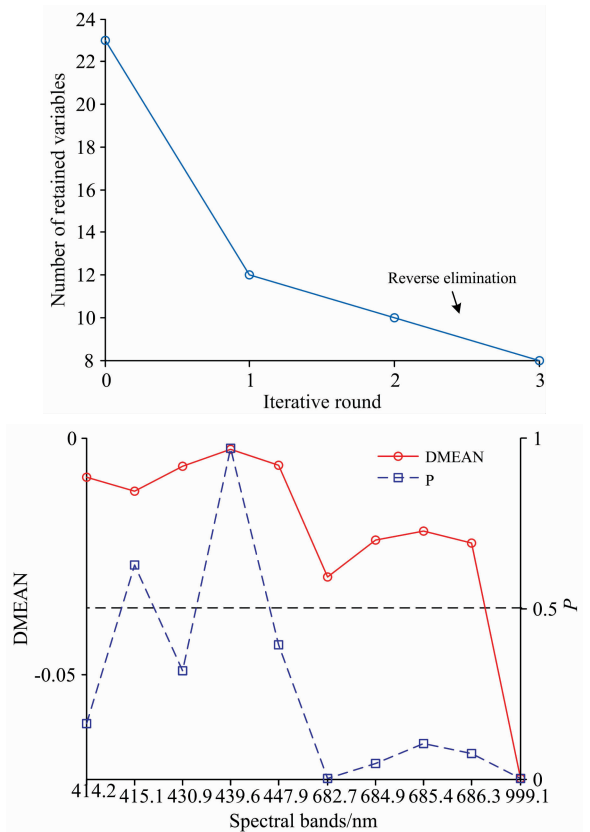


图 6 IRIV 算法筛选过程图

(a): IRIV 迭代保留波段个数; (b): 第 2 轮波段的 DMEAN 和 P 值

Fig. 6 IRIV algorithm screening process diagram

(a): Number of retained variables in iterative round of IRIV;

(b): Dmean and P values in second iterat

2.3 植被指数的筛选

图 7 为 400~1 000 nm 中任意两个光谱波段随机组合构成的 DSI(R_i, R_j), RSI(R_i, R_j)和 NDSI(R_i, R_j)与粳稻叶片氮素含量的决定系数等势图。从决定系数等势图中可知与叶片氮素含量相关性最优的光谱植被指数和相关性最好的波段范围。在图 7(a)中,对于 DSI(R_i, R_j)而言,在 707.3~

852.5 与 625.5~692 nm 的光谱波段组合与叶片氮素含量较好, 决定系数 R^2 均大于 0.7, 其中 DSI($R_{648.1}, R_{738.1}$) 与叶片氮素含量的相关性最好, R^2 为 0.811 4。相比 DSI(R_i, R_j) 而言, RSI(R_i, R_j) 与叶片氮素含量相关性较好的波段范围较窄, 多集中于可见光波段范围[图 7(b)]。其中在 596.1~686.2 和 507.2~607.4 nm 的波段组合相关性最佳, R^2 均大于 0.7, 最好的波段组合构建的 RSI 植被指数为 RSI($R_{532.8}, R_{677.3}$), R^2 为 0.829 7。

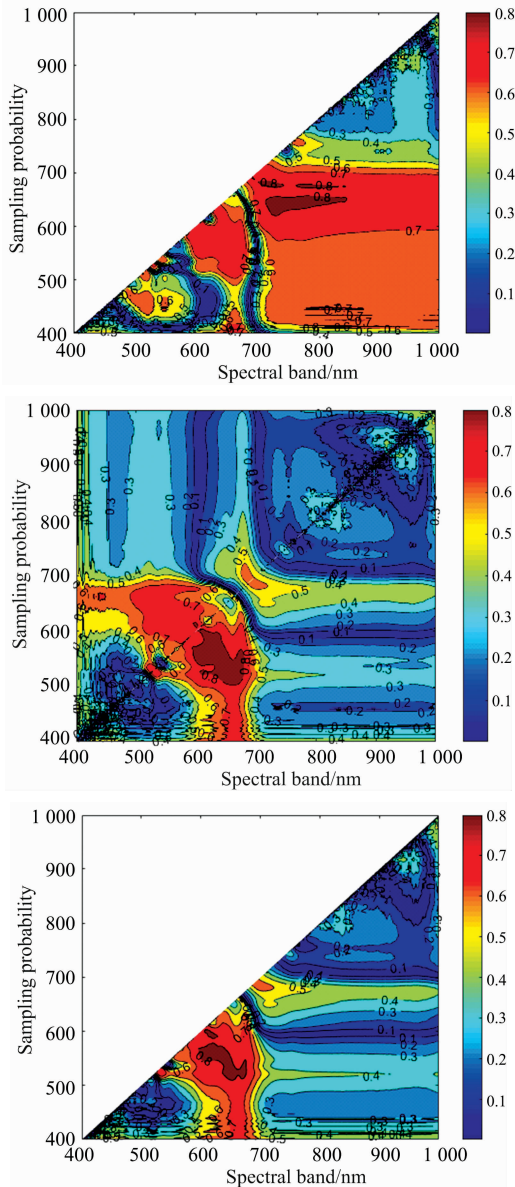


图 7 任意两波段组合构成的 DSI (a), RSI (b) 和 NDSI (c) 与叶片氮素含量的决定系数等势图

Fig. 7 Contour of R^2 between DSI (a), RSI (b), NDSI (c) and leaf nitrogen content in any two bands

图 7(c) 为任意两个波段组合构建的 NDSI(R_i, R_j) 与叶片氮素含量的决定系数等势图。在 601.5~682.7 与 514.1~609.2 nm 的光谱波段组合与氮素含量的相关性较好, R^2 大

于 0.7, 其中 NDSI($R_{654.8}, R_{532.9}$) 与叶片氮素含量相关性最好, R^2 为 0.816 9。

2.4 粳稻叶片氮素含量反演建模与分析

2.4.1 BP 神经网络反演建模

采用 random-frog 和 IRIV 算法筛选的 8 个特征波段和全生长期的 3 种较好的植被指数 (DSI($R_{648.1}, R_{738.1}$), RSI($R_{532.8}, R_{677.3}$) 和 NDSI($R_{654.8}, R_{532.9}$)) 分别作为模型输入, 粳稻叶片氮素含量作为模型输出, 构建 BP 神经网络反演模型。将 Tansig 和 Purelin 法分别设置为隐含层和输出层的传递函数, 将 Trainlm 法设置为训练函数, 训练最大迭代次数为 1 000 次, 学习速率 lr 和训练精度 goal 分别均为 0.1 和 0.01。同时, 通过逐步试验得出隐含层神经节点数为 6 时, 模型估测效果最佳。建模结果如图 8 所示。

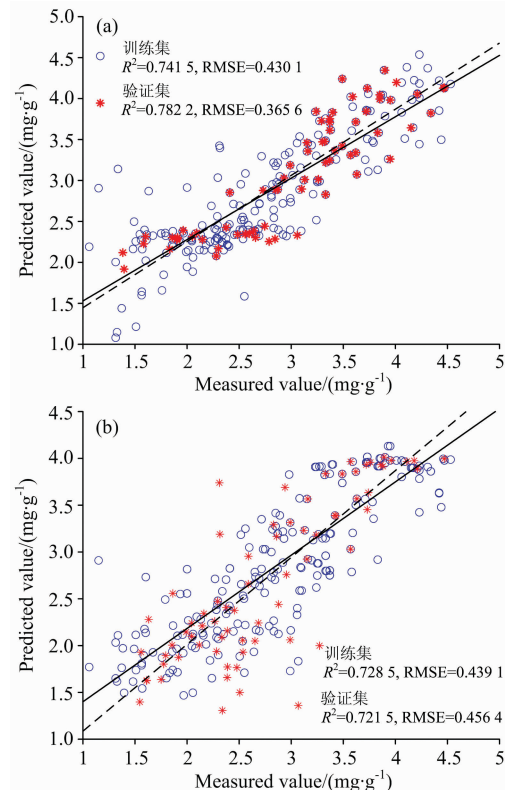


图 8 BP 神经网络建模结果

(a): 特征波段组合; (b): 植被指数组合

Fig. 8 Modeling results of BP neural network

(a): Characteristic band combination;
(b): Vegetation index combination

由图 8 可知, 以 8 个特征波段组合作为输入构建的 BP 神经网络反演模型, 其建模效果要优于植被指数组合。模型训练集的决定系数 R^2 和均方根误差 RMSE 分别为 0.741 5 和 0.430 1, 验证集的 R^2 和 RMSE 分别为 0.782 2 和 0.365 6。由此可见, 采用 8 个特征波段组合作为 BP 神经网络的输入进行建模, 提高了反演粳稻叶片氮素含量的预测能力。

2.4.2 支持向量机反演建模

分别采用特征波段组合和植被指数组合作为支持向量机

(SVR)模型的输入,构建梗稻叶片氮素含量反演模型。将SVR的核函数设置为 Sigmoid,特征波段组合和植被指数组合的惩罚因子 c 分别为 1.3 和 2, Sigmoid 核函数的系数 g 分别为 10.1 和 9, 损失函数值分别为 0.17 和 0.1, SVR 建模结果如图 9 所示。

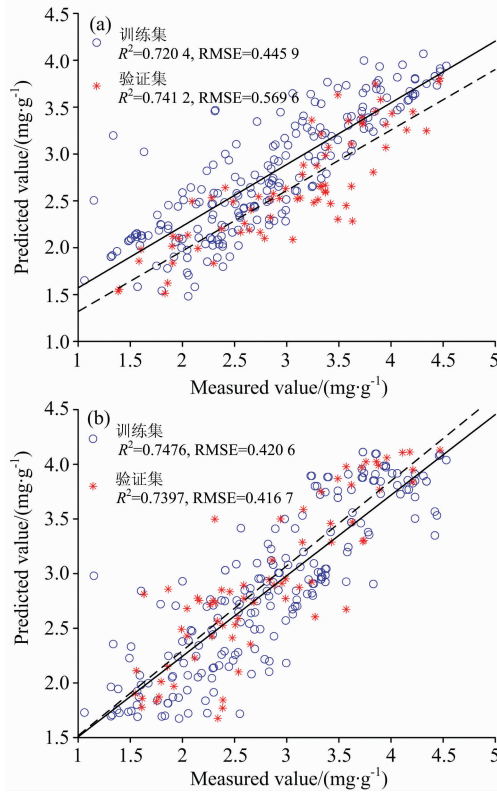


图 9 支持向量机建模结果

(a): 特征波段组合; (b): 植被指数组合

Fig. 9 Modeling results of SVR

(a): Characteristic band combination;

(b): Vegetation index combination

由图 9 比较可知,无论是以特征波段组合还是以植被指数组合作为输入所构建的 SVR 模型,可靠性和估测能力具有一致性。从验证集的建模结果来看,特征波段组合的 SVR 模型的 R^2 略高于植被指数组合,但其 RMSE 略低于植被指数组合。特征波段组合和植被指数组合的训练集 R^2 和 RMSE 分别为: 0.720 4, 0.445 9 和 0.747 6 和 0.420 6, 验证集 R^2 和 RMSE 分别为: 0.741 2, 0.569 6 和 0.739 7 和 0.416 7。

2.4.3 非支配的精英策略遗传算法优化极限学习机反演建模

采用 random-frog 和 IRIV 算法筛选的 8 个特征波段和全生长期的 3 种较好的植被指数 ($DSI(R_{648.1}, R_{738.1})$, $RSI(R_{532.8}, R_{677.3})$ 和 $NDSI(R_{654.8}, R_{532.9})$) 分别作为非支配的精英策略遗传算法优化极限学习机 (NSGA2-ELM) 模型的输入,构建梗稻叶片氮素含量反演模型。将特征波段组合和植被指数组合的 NSGA2-ELM 模型的种群个数均设置为 30, 交叉概率和变异概率分别设置为 0.96, 0.97 和 0.001,

0.001, 最大拥挤距离均为 10 000, 保留最优个体比例均为 0.1, 迭代次数均为 50 次。建模结果如图 10 所示。

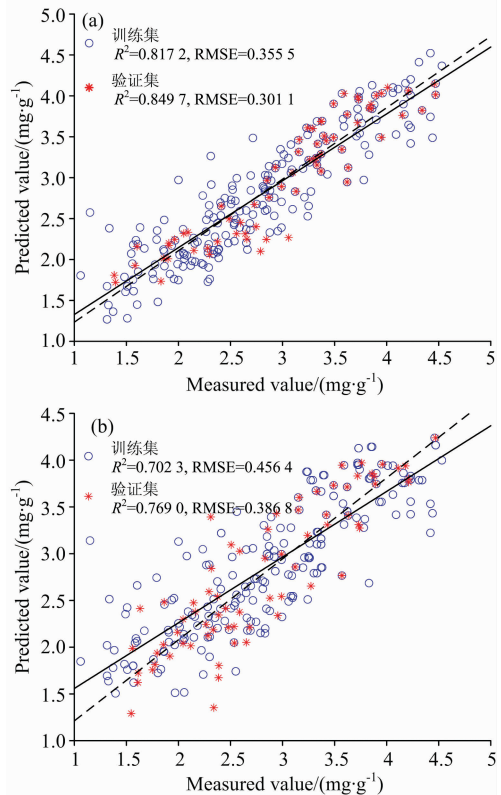


图 10 NSGA2-ELM 建模结果

(a): 特征波段组合; (b): 植被指数组合

Fig. 10 Modeling results of NSGA2-ELM

(a): Characteristic band combination;

(b): Vegetation index combination

由图 10 可知,以特征波段组合作为 NSGA2-ELM 模型的输入构建模型的估测效果明显优于植被指数组合,模型训练集的决定系数 R^2 和均方根误差 RMSE 分别为 0.817 2 和 0.355 5, 验证集的 R^2 和 RMSE 分别为 0.849 7 和 0.301 1。相比 BP 神经网络模型和支持向量机模型, NSGA2-ELM 模型无论是以特征波段组合还是以植被指数组合作为输入,在模型精度和估测能力上都有显著提高,说明采用 NSGA2 算法优化 ELM 模型对梗稻叶片氮素含量的预测具有较大的优势。分析原因在于 NSGA2 算法能够保持种群的多样性,并引入精英策略,不仅能增大样本采集空间,而且能较好地防止最优个体丢失,也避免了类似 BP 神经网络等模型陷入过拟合和局部最优现象,从而使模型具有更好的非线性映射能力和鲁棒性。同时从不同输入类型来看,3 种反演模型多以特征波段组合作为输入,模型估测效果较好。说明使用 random-frog 和 IRIV 筛选出的特征波段具有较好的代表性,包含了梗稻叶片高光谱的大部分特征。因此采用 random-frog 和 IRIV 方法获取的特征波段所建立的反演模型,具有更好的稳定性和估测能力。

3 结 论

以粳稻叶片高光谱数据为数据源,先后采用随机青蛙算法与迭代和保留信息变量算法筛选特征波段,并构建 3 种较优植被指数,分别以特征波段组合和植被指数组合为模型输入,构建 BP 神经网络、支持向量机和非支配的精英策略遗传算法优化极限学习机粳稻叶片氮素含量反演模型并比较其优劣。主要结论如下:

(1)针对高光谱较多的光谱波段,通过采用 random-frog 和 IRIV 算法相结合的方式筛选光谱特征波段,共得到 8 个特征波段,其中可见光波段有 7 个,分别为 414.2, 430.9,

439.6, 447.9, 682.7, 685.4 和 686.3 nm,近红外波段仅有 1 个为 999.1 nm。结果表明能够较好地剔除无用和干扰光谱波段,大大降低了波段间的共线性和冗余信息,提高了建模效率和精度。

(2)分别采用特征波段组合和植被指数组合作为模型输入,构建 3 种粳稻叶片氮素含量反演模型。对比分析得出,NSGA2-ELM 模型的估测效果要优于 BP 神经网络模型和 SVR 模型,这是由于 NSGA2 算法能保持种群的多样性,增大了样本空间,防止最优个体丢失,提高了 ELM 模型的估测精度和鲁棒性。同时,对比不同输入类型建模可知,以特征波段组合作为输入,模型估测效果较好,验证集决定系数 R^2 为 0.849 7,均方根误差 RMSE 为 0.301 1。

References

- [1] SONG Li-juan, YE Wan-jun, ZHENG Yan-yan, et al(宋丽娟,叶万军,郑妍妍,等). China Rice(中国稻米), 2017, (6): 22.
- [2] YANG Fu-qin, FENG Hai-kuan, LI Zhen-hai, et al(杨福芹,冯海宽,李振海,等). Transactions of the Chinese Society for Agricultural Machinery(农业机械学报), 2017, (9): 143.
- [3] LI Lan-tao, REN Tao, WANG Shan-qin, et al(李岚涛,任涛,汪善勤,等). Transactions of the Chinese Society for Agricultural Machinery(农业机械学报), 2017, (3): 226.
- [4] HE Ru-yan, QIAO Xiao-jun, JIANG Jin-bao, et al(何汝艳,乔小军,蒋金豹,等). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2015, (2): 141.
- [5] BAI Li-min, LI Fen-ling, CHANG Qing-rui, et al(白丽敏,李粉玲,常庆瑞,等). Journal of Plant Nutrition and Fertilizers(植物营养与肥料学报), 2018, 24(5): 52.
- [6] LIU Chang-hua, FANG Zheng, CHEN Zhi-chao, et al(刘昌华,方征,陈志超,等). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2018, 34(19): 170.
- [7] WANG Shu-wen, ZHAO Yue, WANG Li-feng, et al(王树文,赵越,王丽凤,等). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2016, 32(20): 187.
- [8] LIU Ming-bo, TANG Yan-lin, LI Xiao-li, et al(刘明博,唐延林,李晓利,等). Infrared and Laser Engineering(红外与激光工程), 2014, 43(4): 1265.
- [9] Tian Y C, Gu K J, Chu X, et al. Plant and Soil, 2014, 376(1-2): 193.
- [10] Lin D, Wei G, Jian Y. Optics & Laser Technology, 2018, 107: 372.
- [11] FANG Mei-hong, LIU Xiang-nan(方美红,刘湘南). Journal of Applied Sciences(应用科学学报), 2010, 28(4): 387.
- [12] Yun Y H, Wang W T, Tan M L, et al. Analytica Chimica Acta, 2014, 807: 36.

Inversion Based on High Spectrum and NSGA2-ELM Algorithm for the Nitrogen Content of Japonica Rice Leaves

FENG Shuai¹, CAO Ying-li^{1,2*}, XU Tong-yu^{1,2}, YU Feng-hua^{1,2}, CHEN Chun-ling^{1,2}, ZHAO Dong-xue¹, JIN Yan¹

1. College of Information and Electrical Engineering, Shenyang Agricultural University, Shenyang 110161, China

2. Liaoning Agricultural Information Technology Center, Shenyang Agricultural University, Shenyang 110161, China

Abstract In order to provide an efficient, rapid and non-destructive Inversion method for the nitrogen content of japonica rice leaves, based on the japonica rice plot experiment, using high spectrum technology and laboratory chemistry experiments to obtain the effective data for three growth periods of japonica rice in the tillering stage, joining stage and heading stage. A total of 280 sets of leaf high spectrum data and corresponding rice leaf nitrogen content data were obtained to analyze the spectral characteristics of japonica rice leaves with different nitrogen treatment levels. The random frog algorithm (Random_frog) is combined with the iteratively retaining informative variables algorithm (IRIV) to screen the feature bands, and any two spectral bands are randomly combined to construct the difference vegetation index $DSI (R_i, R_j)$, ratio vegetation index, $RSI (R_i, R_j)$ and normalized vegetation index $NDSI (R_i, R_j)$, respectively, combine the superior feature band combination and vegetation index as model inputs. Therefore, BP neural network, support vector machine (SVR) and non-dominated elite strategy genetic algorithm optimization limit learning machine (NSGA2-ELM) japonica rice leaf nitrogen content Inversion model were constructed, and the model was verified and analyzed. The results showed that with the increasing level of nitrogen fertilizer treatment, the reflectivity of the japonica rice leaves in the near-infrared range gradually increased, while the reflectance decreased gradually in the visible range. A total of 8 characteristic bands were obtained by combining Random_frog and IRIV. Among them, there are 7 visible light bands, which are 414.2, 430.9, 439.6, 447.9, 682.7, 685.4 and 686.3 nm, respectively. Only one of the near-infrared bands is 999.1 nm. This method effectively eliminates interference information and greatly reduces the collinearity between the bands. At the same time, it can be analyzed from the three vegetation index ($DSI (R_i, R_j)$, $RSI (R_i, R_j)$, $NDSI (R_i, R_j)$) and the determination coefficient of the nitrogen content of the japonica rice leaves. $DSI (R_{648.1}, R_{738.1})$, $RSI (R_{532.8}, R_{677.3})$ and $NDSI (R_{654.8}, R_{532.9})$ have the best correlation with leaf nitrogen content, R^2 are 0.811 4, 0.829 7 and 0.816 9, respectively. In the comparative analysis of the input parameters with different input parameters, the model Inversion with the feature band combination as the model input is slightly better than the vegetation index combination, R^2 is greater than 0.7, and the RMSE is less than 0.57. In the comparative analysis between the Inversion models, the estimation effect of the NSGA2-ELM Inversion model proposed in this paper is significantly better than the BP neural network model and the SVR model. The training set determination coefficient R^2 is 0.817 2, and the root means square error RMSE is 0.355 5. The set R^2 is 0.849 7 and the RMSE is 0.301 1. According to the results of this study, the Random_frog-IRIV screening characteristic band method combined with NSGA2-ELM modeling method has a significant advantage in rapidly detecting the nitrogen content of japonica rice leaves. The research results can provide a theoretical reference for field precision fertilization of japonica rice.

Keywords Hyperspectral data; Leaf nitrogen content; Characteristic band; Vegetation index; Inversion model

(Received Sep. 17, 2019; accepted Dec. 20, 2019)

* Corresponding author