

基于 PLS - BPNN 算法的土壤速效磷高光谱回归预测方法

齐海军¹ 李绍稳¹ KARNIELI Arnon² 金 秀¹ 王文才¹

(1. 安徽农业大学信息与计算机学院, 合肥 230036; 2. 内盖夫本·古里安大学雅各布·布劳斯汀沙漠研究所, 斯代博客 84990)

摘要: 土壤速效磷是影响作物生长发育的重要养分指标。光谱分析技术对速效磷的定量预测具有较好的应用前景, 高光谱带宽窄、分辨率高, 但存在数据冗余和共线性等问题。本文针对皖北砂姜黑土 145 个样本开展研究, 提出了利用偏最小二乘回归算法 (PLS - R) 对土壤可见近红外高光谱数据 (400 ~ 1 000 nm) 进行数据降维和特征提取, 根据交叉验证和变量投影重要性分别得到潜在变量和特征波长; 再分别输入 BP 神经网络 (BPNN) 进行训练, 得到回归分析模型对速效磷进行定量预测。结果表明: 与利用全部波长数据建模的预测结果 (校正集和验证集的相对分析误差 M_{RPD} 分别为 10.27 和 2.09) 相比, 利用 9 个特征波长建立的回归模型校正集 M_{RPD} 为 2.66, 预测精度明显降低, 而验证集 M_{RPD} 为 2.05, 近似达到利用全部波长数据建模的预测效果; 利用 5 个潜在变量建立回归模型, 校正集和验证集的 M_{RPD} 分别为 3.10 和 2.29, 其中验证集相对于全部波长建模的预测精度提高了 9.6%。因此, 基于 PLS - BPNN 算法进行回归建模可以有效降低高光谱数据冗余和共线性的影响, 提高模型的泛化能力, 且利用潜在变量进行回归建模能提高模型预测精度。

关键词: 土壤速效磷; 光谱分析; 回归算法; 数据降维; 特征提取

中图分类号: S153.6; O657.3 **文献标识码:** A **文章编号:** 1000-1298(2018)02-0166-07

Prediction Method of Soil Available Phosphorus Using Hyperspectral Data Based on PLS - BPNN

QI Haijun¹ LI Shaowen¹ KARNIELI Arnon² JIN Xiu¹ WANG Wencai¹

(1. College of Information and Computer Science, Anhui Agricultural University, Hefei 230036, China

2. Jacob Blaustein Institutes for Desert Research, Ben - Gurion University of the Negev, Sede Boqer 84990, Israel)

Abstract: Soil available phosphorus (AP) is supposed to be an important nutrient constituent for the growth and development of crops. Hyperspectral analysis has proven to be a rapid and effective means for quantitatively predicting soil AP, which has a good prospect benefit from the narrow bandwidth and the high resolution. However, the existence of multicollinearity and redundant considerably leads to overfitting of the regression model and decrease of the generalization ability. A total of 145 lime concretion black soil samples collected from the Northern Anhui Plain, China, were used as research objects to investigate the prediction performance of the back-propagation neural network (BPNN) based on the partial least square regression (PLS - R) algorithm. The PLS - R was applied to conduct dimensionality reduction and feature selection on the soil visible and near infrared hyperspectral data ranging from 400 ~ 1 000 nm with 339 wavelengths. Five latent variables (LVs) were obtained by the leave one out cross-validation, and nine optimal wavelengths were selected by the variable importance in projection (VIP) scores. The BPNN regression models were built with the input of the five latent variables (LVs - BPNN), the nine optimal wavelengths (VIPs - BPNN), and the whole wavelengths (Ws - BPNN), respectively. The ratio of performance to deviation (M_{RPD}) and the ratio of the interpretable sum squared deviation to the real sum squared deviation ($M_{SSR/SST}$) were selected to evaluate the prediction accuracy and explanatory power of different regression models, respectively. As a result, the prediction accuracies of three BPNN models outperformed the PLS - R model significantly; the VIPs - BPNN model achieved similar performance (M_{RPD} was 2.05, $M_{SSR/SST}$ was 0.79) as the Ws - BPNN

收稿日期: 2017 - 07 - 15 修回日期: 2017 - 10 - 11

基金项目: 农业部引进国际先进农业科学技术计划 (948 计划) 项目 (2015 - Z44、2016 - X34) 和国家留学基金项目 (201608340066)

作者简介: 齐海军 (1991 -), 男, 博士生, 主要从事土壤速效养分高光谱检测研究, E-mail: haijun.qi0418@qq.com

通信作者: 李绍稳 (1962 -), 男, 教授, 博士生导师, 主要从事农业信息技术研究, E-mail: shwli@ahau.edu.cn

model (M_{RPD} was 2.09, $M_{\text{SSR/SST}}$ was 0.85) of the validation set, while the M_{RPD} was decreased obviously from 10.27 (Ws - BPNN) to 2.66 (VIPs - BPNN) of the calibration set; the LVs - BPNN model gained the highest prediction accuracy as M_{RPD} was 2.29 of the validation set, even though the $M_{\text{SSR/SST}}$ was slightly decreased to 0.76. The results illustrated that the PLS - BPNN models could significantly reduce the degree of overfitting and improve the generalization ability; moreover, the LVs - BPNN model could improve the accuracy of predicting soil AP.

Key words: soil available phosphorus; spectral analysis; regression algorithms; data dimensionality reduction; variable selection

0 引言

土壤速效磷是土壤中的重要营养成分,会直接影响作物的生长发育和最终产量^[1]。其精确测量或估算对测土配方施肥、作物高产稳产以及资源节约利用等都具有重要意义。传统的实验室理化测试方法由于耗时、费力、低效等缺陷,已不适应精准农业的发展要求。近红外光谱分析技术具有快速、无损、高效等特点,目前已被广泛应用于土壤成分检测中^[2]。

国内外均有利用室内可见近红外光谱对土壤速效磷进行预测的相关研究报道,且取得了良好效果^[3-7]。高光谱带宽窄,分辨率高,携带大量信息的同时也伴随着多重共线性和数据冗余等问题,从而导致模型过拟合严重,泛化性能较差^[8]。土壤速效磷含量低,且在近红外光谱区域(350 ~ 2 500 nm)没有明显的吸收带,更增大了高光谱模型的预测难度。其成功预测往往是依靠铁氧化物、有机质、粘土矿物和水分等其他成分在光谱中共同响应的结果^[9]。偏最小二乘 PLS 算法虽然能有效克服高光谱数据的共线性问题^[10],但学习能力有限,难以实现速效磷的有效预测。为此, SARATHJITH 等^[11]利用离散小波变换结合支持向量机建立回归模型,对速效磷的预测相对分析误差达到了 2.27。MOUAZEN 等^[12]将 PLS 的潜在变量作为反向传播神经网络 (Back-propagation neural network, BPNN) 的输入,建立新鲜土壤速效磷的回归模型,预测相对分析误差达到了 1.77 ~ 1.94。可以看出,对原始光谱数据进行数据降维或特征提取处理,并结合智能学习算法,能够有效提高模型的预测精度。

本文采用偏最小二乘回归算法 PLS - R 对可见近红外高光谱数据(400 ~ 1 000 nm)分别进行数据降维和特征提取,得到潜在变量和特征波长,再分别作为 BP 神经网络输入以建立土壤速效磷的定量回归模型,通过与全部波长数据建立的 PLS - R 模型和 BPNN 模型进行对比分析,从而得到土壤速效磷的最佳预测模型。

1 试验与方法

1.1 试验概况

野外样本采集试验于 2016 年 5—6 月在皖北平原的蒙城县、埇桥区和怀远县三地开展,土壤类型主要为砂姜黑土,麦玉轮作是该地区的主要种植方式。根据不同的秸秆还田方式、变量施肥梯度和种植密度来选取采样点以增大速效磷的含量差异。采样深度为 0 ~ 20 cm,通过 3 点对角取样进行土壤混合作为一个采样点的样本。经去除石块、秸秆和作物残根后,每份样本约 1.5 kg,共计采样 153 份。土壤封存移送到实验室进行风干处理,随后研磨,过 20 目筛子,得到试验所需的土壤样本粉末。将样本均匀分成 2 份,其中 1 份用于标准理化测试,另 1 份用于土壤光谱采集。速效磷的理化检测使用碳酸氢钠浸提-钼锑抗分光光度法^[13],其中 8 个样本由于检测值过大而被剔除,因此共 145 份样本用于光谱建模分析。

土壤光谱数据使用实验室搭建的室内高光谱成像系统(图 1)采集,该系统主要包括推扫式高光谱成像仪 (OKSI, Torrance, CA, USA)、水平/倾斜二自由度的扫描云台、50 W 稳定卤钨灯、带有 HyperVision 采集软件的戴尔工作站和一套高度可调式支架。土壤样本用培养皿盛放,用钢尺轻轻刮平后放置在垫有黑色吸光布的样品台上。采集试验

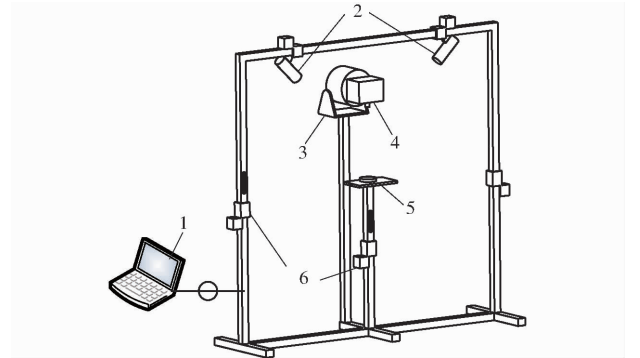


图 1 室内高光谱成像系统

Fig. 1 Indoor hyperspectral imaging system

1. 工作站 2. 卤钨灯光源 3. 扫描云台 4. 高光谱成像仪
5. 样本 6. 支架

在无自然光干扰的暗室环境中进行,光源角度、高度和样本距成像仪高度经多次试验调整优化。每个样本测量前,均采集标准白板(Labsphere, North Sutton, NH, USA)数据用于校正。测量时,每个样本旋转 90°测量 4 次取平均以降低散射影响。高光谱成像系统的输出为包括光谱数据(400 ~ 1 000 nm,分辨率 1.79 nm,共 339 个波长)和空间数据(1 620 像素×841 像素)的“图像立方体”。其中后 100 帧,镜头自动关闭,作为黑板数据。

1.2 光谱处理与变换

由于所有土壤样本均经研磨处理,各样本之间的图像纹理等特性差异较小,且速效磷含量较低,更不足以影响图像特性,故本文仅使用高光谱成像系统得到的光谱数据进行进一步分析研究。

利用 ENVI 软件(Exelis Visual Information Solutions, Boulder, Co, USA)对土壤原始高光谱图像进行黑白板校正,得到反射率图像^[14-15]。手动画出仅包含土壤样本的感兴趣区域(Region of interest, ROI),将 ROI 内的反射率进行平均作为该样本的反射率。由于所有土壤样本位置固定,故此 ROI 可以通用到所有样本。样本较多,此部分功能通过 IDL 编程实现。

光谱反射率曲线在首尾区域有较低信噪比,影响模型预测精度,因此仅使用 420 ~ 960 nm(303 个波长)区域的光谱数据进行进一步分析。经过对比分析常用方法,本文使用以下组合方法对光谱进行预处理与变换:①使用 Savitzky - Golay 卷积平滑算法^[16]对光谱曲线进行平滑去噪。②再使用标准正态变换(Standard normal variate, SNV)来消除固体颗粒大小、表面散射以及光程变化光谱的影响^[17-18]。③建立与应用回归模型之前,分别对光谱矩阵和速效磷浓度向量做均值中心化处理,以保持相同标度^[19]。

1.3 数据回归分析方法

偏最小二乘回归(PLS - R)^[20]算法利用对自变量和因变量系统中的数据信息进行分解和筛选的方式,提取对因变量的解释性最强的综合变量,辨识系统中的有用信息和无关信息,克服变量多重相关性的影响,从而建立适当的预测模型。建模过程中,潜在变量(Latent variable, LV)个数是模型优化的关键,本文采用留一交叉验证的方式,以最小均方误差和赤池信息量准则值为标准确定最佳 LV 个数^[21]。

变量投影重要性(Variable importance in projection, VIP)^[22]的得分可用于识别 PLS - R 模型中的重要波长,具体计算公式为^[23]

$$V_k(a) = p \sum_a w_{ak}^2 (S_{SY_a} / S_{SY_t}) \tag{1}$$

式中 $V_k(a)$ ——在使用 a 个 LV 建模条件下第 k 个自变量的投影重要性的得分

p ——自变量个数

w_{ak} ——对应的权重系数

S_{SY_a} —— a 个 LV 对因变量 y 的解释能力

S_{SY_t} ——全部 LV 使用时对 y 的解释能力

一般认为当 VIP 得分大于 1 时,该自变量对 y 的预测有重要的作用,即特征波长^[23-24]。

反向传播神经网络(BPNN)是一种单向多层感知的前馈式神经网络^[25],由于其强大的学习能力已被广泛应用于土壤光谱回归建模分析中^[12, 26-28]。本文使用高光谱数据全部波长(303 个),PLS - R 潜在变量 LV 和经 VIP 得分筛选的特征波长分别作为输入,构建 3 层神经网络土壤速效磷回归模型。其中隐含层设置 10 个节点。输入层和输出层的传递函数分别为 tan-sigmoid 和 pure-linear。为了降低过拟合现象,本文使用贝叶斯正则化反向传播算法(trainbr)进行模型校正训练。Marquardt 调节参数设置为 10^{-7} ,学习速率设置为 0.2,训练目标设置为 0.01,最大迭代次数设置为 2 000^[29]。

为了进行模型校正与验证,本文使用 Kennard - Stone 算法^[30]按照 7:3 划分校正集和验证集。共 102 份样本用于校正建模,43 份用于独立验证模型泛化效果。模型预测性能使用相对分析误差 M_{RPD} (标准偏差除以均方根误差),作为评价标准;同时利用解释总方差占总方差之比 $M_{SSR/SST}$ 来评价模型的解释能力^[31],较好的模型应有较高的 M_{RPD} 和 $M_{SSR/SST}$ 。根据文献[32],当 $M_{RPD} > 2.0$ 时模型精度可以分为 A 类代表良好的预测能力, $1.4 \leq M_{RPD} \leq 2.0$ 为 B 类代表中等预测能力, $M_{RPD} < 1.4$ 为 C 类代表较差的预测能力。并且 $M_{SSR/SST}$ 大于 0.5,模型才有值得信服的预测能力。2 个指标的计算公式分别为

$$M_{RPD} = \sqrt{m \sum_{i=1}^m (y_i - \bar{y})^2 / \sum_{i=1}^m (f(x_i) - y_i)^2 / (m-1)} \tag{2}$$

$$M_{SSR/SST} = \sum_{i=1}^m (f(x_i) - \bar{y})^2 / \sum_{i=1}^m (y_i - \bar{y})^2 \tag{3}$$

式中 m ——样本个数

y_i ——第 i 个变量的实际测量值

$f(x_i)$ ——第 i 个变量的模型预测值

$\bar{y}$ —— y_i 的平均值

本文使用 Matlab R2016a 进行光谱预处理变换和回归建模分析。

2 结果与分析

2.1 土壤速效磷含量和光谱响应

表 1 为校正集、验证集以及全部样本的速效磷含量统计参数,可以看出速效磷 的含量具有明显的梯度差异,且 3 个样本集的极值和标准差数据接近,说明具有相似的数据范围和分布结构,因此校正集和验证集均可有效的代表整体数据集分布特征^[33]。

表 1 土壤速效磷含量的统计参数

	数量	平均值	标准差	最小值	中值	最大值
校正集	102	18.41	13.49	4.50	12.60	56.10
验证集	43	18.24	15.79	4.40	9.20	56.10
全部样本	145	18.36	14.17	4.40	11.05	56.10

土壤原始反射率光谱如图 2 所示。可以看出由于干燥的砂姜黑土颜色较深呈黄棕色,故反射率整体较低,尤其在蓝光区域(420 ~ 500 nm),在绿光至红光区域出现明显上升趋势(500 ~ 700 nm)。由于砂姜黑土中铁氧化物较多,如针铁矿和赤铁矿等使 420 ~ 480 nm 出现明显吸收波动^[34]。在近红外区域 850 nm 处光谱出现波动,可能是因为土壤中的有机质产生光谱吸收,930 nm 处的波动可能是因为土壤中的羟基官能团或磁赤铁矿吸收所致^[34-36]。

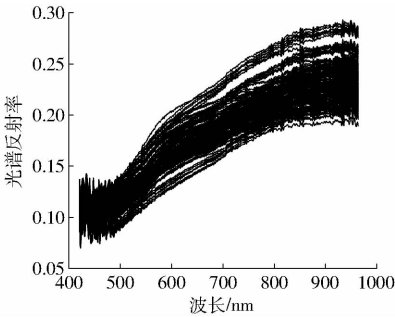


图 2 土壤原始光谱反射率

Fig.2 Spectral reflectance of soil samples

2.2 PLS - R 建模分析

利用 PLS - R 对校正集的光谱数据和土壤速效磷含量建立回归模型,采用留一交叉验证的方式对不同数量 LV 建立的模型进行分析。交叉验证中,

均方误差和赤池信息量准则值随 LV 个数变换的趋势如图 3 所示。可以看出,随着 LV 个数的增加,均方误差和赤池信息量准则值先降低后增高,说明当 LV 个数较少时模型处于欠拟合状态,因此误差较大;而当 LV 个数增加时,模型出现过拟合现象,均方误差会增加,且模型越来越复杂,赤池信息量准则值也会增加。当 LV 个数为 6 时均方误差达到最小值,当 LV 个数为 5 时赤池信息量准则值达到最小值,考虑到模型的复杂程度且 5 个 LV 与 6 个 LV 对应的模型精度差异很小,故选取 5 个 LV 用于建立土壤速效磷的 PLS 回归模型,校正集和验证集的结果如表 2 所示,验证集 43 个样本的土壤速效磷实际理化测量值和模型预测值之间的对比效果如图 4a 所示。验证集 $M_{RPD} = 1.51$, $M_{SSR/SST} = 0.74$,模型精度等级为 B 类,说明模型具有中等预测能力。

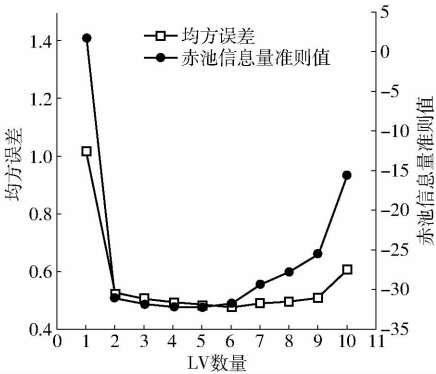


图 3 PLS - R 交叉验证中不同数量 LV 对应的均方误差和赤池信息量准则值

Fig.3 Number of LVs used in PLS - R cross-validation vs mean square error and Akaike information criterion value

使用 5 个 LV 建立的回归模型的 VIP 得分在 420 ~ 960 nm 的分布情况如图 5 所示。当 VIP 值大于 1 时认为该波长对预测速效磷具有重要性。可以看出,420 ~ 480 nm 由于针铁矿和赤铁矿的影响^[34],存在多个波长对速效磷的预测较为重要;540 nm 附近可能是由于土壤颜色或赤铁矿影响^[36-37];680 nm 和 850 nm 附近可能是由于土壤中有机的影响^[34, 38];760 nm 和 950 nm 附近可能是由于残留水分的影响^[39]。文献[40 - 41]等亦发现这些波长对速效磷的预测有重要影响,验证了本研究的正确性。

表 2 不同建模方法的模型性能

Tab.2 Model performance of different methods

建模方法	特征波长数量	校正集		验证集		预测精度分类	备注
		M_{RPD}	$M_{SSR/SST}$	M_{RPD}	$M_{SSR/SST}$		
PLS - R	303	1.62	0.62	1.51	0.74	B	5 LVs
Ws - BPNN	303	10.27	0.92	2.09	0.85	A	
LVs - BPNN	303	3.10	0.86	2.29	0.76	A	5 LVs
VIPs - BPNN	9	2.66	0.80	2.05	0.79	A	VIP 得分大于 1.5

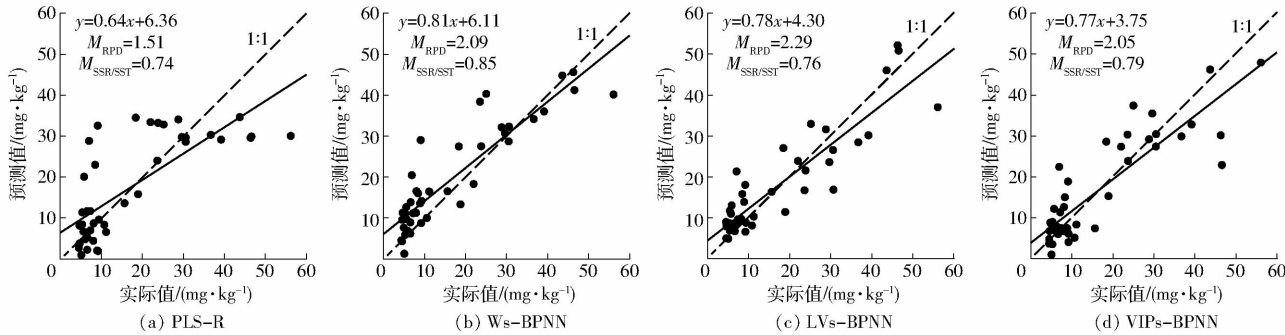


图 4 不同模型验证集样本预测值与实际值的对比

Fig. 4 Predicted *vs* measured values for validation set of different models

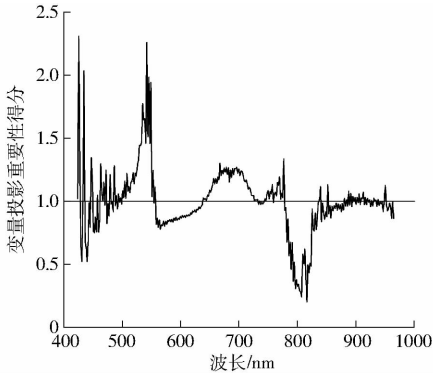


图 5 变量投影重要性得分在光谱区域的分布

Fig. 5 Distribution of VIP scores in spectral region

2.3 PLS - BPNN 回归分析

为了对比 PLS - BPNN 的建模效果,首先使用全部光谱共 303 个波长数据作为 BPNN 的输入建立土壤速效磷的回归模型 Ws - BPNN,校正集和验证集的结果如表 2 所示,验证集 43 个样本的土壤速效磷实际理化测量值和模型预测值之间的对比如图 4b 所示。验证集 $M_{RPD} = 2.09$, $M_{SSR/SST} = 0.85$, 模型精度等级为 A 类,说明模型具有较强的预测能力。

利用 PLS 得到的 5 个 LV 作为 BPNN 的输入建立土壤速效磷的回归模型 LVs - BPNN,校正集和验证集的结果如表 2 所示,验证集 43 个样本的土壤速效磷实际理化测量值和模型预测值之间的对比效果如图 4c 所示。验证集 $M_{RPD} = 2.29$, $M_{SSR/SST} = 0.76$, 模型精度等级也达到 A 类。

当 VIP 得分越大说明该波长在 PLS - R 模型中的重要性越强,特征波长的数量也就越少(图 5)。为了确定最佳特征波长数量,本研究将阈值设为 1 ~ 2,并以 0.1 为梯度递增,分别得到不同数量的特征波长,再分别作为 BPNN 的输入建立土壤速效磷的回归模型,不同的阈值下模型预测结果如图 6 所示。可以看出随着阈值的增加,校正集 M_{RPD} 有总体逐渐降低的趋势,说明模型学习能力随着特征波长数量的减少而降低。验证集 M_{RPD} 有先升高再降低的趋势,说明阈值在 1.1 ~ 1.4 之间选出的特征波长有较

强的共线性,导致模型泛化能力较差;当阈值超过 1.5 后,一些重要的特征波长被移除,导致模型学习能力不足,影响模型的预测精度。因此,选择 1.5 作为 VIP 得分的阈值,共 9 个特征波长作为 BPNN 的输入建立土壤速效磷的回归模型 VIPs - BPNN,校正集和验证集的结果如表 2 所示,验证集 43 个样本的土壤速效磷实际理化测量值和模型预测值之间的对比效果如图 4d 所示。验证集 $M_{RPD} = 2.05$, $M_{SSR/SST} = 0.79$, 模型精度等级也达到 A 类。

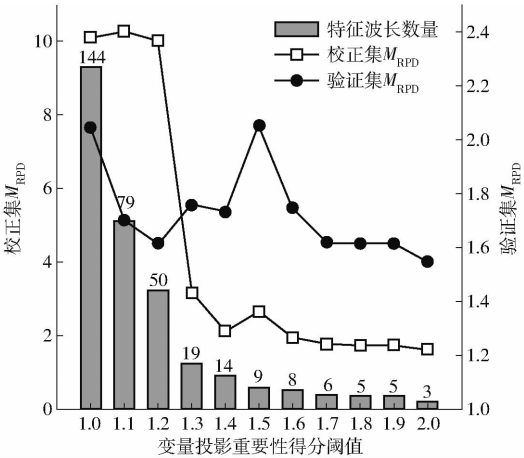


图 6 不同变量投影重要性得分阈值对模型精度的影响

Fig. 6 Effects of different thresholds of variable importance in projection on models' performance

2.4 不同建模方法的对比分析

表 2 为 PLS - R 模型、全部波长直接建立的 Ws - BPNN 模型、潜在变量建立的 LVs - BPNN 模型和利用 VIP 特征波长建立的 VIPs - BPNN 模型在校正集和验证集的结果。图 4 为验证集 43 个样本的土壤速效磷实际理化测量值和各模型预测值之间散点对比效果和一元线性关系。

通过对比可发现,PLS - R 模型预测效果最差,验证集 $M_{RPD} = 1.51$, 预测精度只达到了 B 类标准,因此图 4a 中散点图较为分散,且测量值与预测值的一元回归线更为倾斜,说明模型稳定性不好。3 种 BPNN 模型的预测精度均明显高于 PLS - R 模型,且

均达到了 A 类标准,对土壤速效磷具有较强的预测能力。可以看到 W_s -BPNN 模型校正集 $M_{RPD} = 10.27$, $M_{SSR/SST} = 0.92$,说明全部波长直接建模具有最强的学习能力。但验证集 $M_{RPD} = 2.09$,因此出现严重的过拟合现象,模型泛化性能一般; $M_{SSR/SST} = 0.85$,具有较强的解释能力,故图 4b 中的一元回归线斜率更接近 1。 LV_s -BPNN 模型利用降维后的数据,即潜在变量 LV,构建回归模型,因此解释能力相对于 W_s -BPNN 略微降低,但校正集 $M_{RPD} = 3.10$,验证集 $M_{RPD} = 2.29$,验证集相对于全部波长建模的预测精度提高了 9.6%,说明模型具有更好的泛化能力和预测精度,故图 4c 中散点图更加汇聚在 1:1 线两侧。仅利用 9 个特征波长建立的 VIP_s -BPNN 模型,验证集 $M_{RPD} = 2.05$, $M_{SSR/SST} = 0.79$,近似达到了 W_s -BPNN 模型的预测效果,且模型复杂程度得到了明显改善,计算时间大大缩短。通过综合对比可以看

出, LV_s -BPNN 模型解释能力虽然略差,但预测精度最高,因此更适合用于土壤速效磷的高光谱预测。

3 结束语

应用偏最小二乘回归算法 PLS-R 结合 BP 神经网络实现了对土壤中速效磷的高光谱定量分析。通过利用 PLS-R 分别进行数据降维和特征提取,得到潜在变量 LV 和 VIP 特征波长,再分别作为 BPNN 的输入建立了土壤速效磷的定量回归模型。结果表明,利用 9 个特征波长建立的 BPNN 模型近似达到利用全部波长数据建模的预测效果,且模型更为简洁;利用 5 个潜在变量建立的模型相对于全部波长建模预测精度提高了 9.6%。因此,PLS-BPNN 算法能够有效降低高光谱数据冗余和共线性的影响,且利用潜在变量建立的模型更适合用于土壤速效磷的定量预测。

参 考 文 献

- 1 SHEN J, YUAN L, ZHANG J, et al. Phosphorus dynamics: from soil to plant [J]. *Plant Physiology*, 2011, 156(3): 997-1005.
- 2 BEN-DOR E, BANIN A. Near-infrared analysis as a rapid method to simultaneously evaluate several soil properties [J]. *Soil Science Society of America Journal*, 1995, 59(2): 364-372.
- 3 吴茜, 杨宇虹, 徐照丽, 等. 应用局部神经网络和可见/近红外光谱法估测土壤有效氮磷钾 [J]. *光谱学与光谱分析*, 2014, 34(8): 2102-2105.
- 4 WU Qian, YANG Yuhong, XU Zhaoli, et al. Applying local neural network and visible/near-infrared spectroscopy to estimating available nitrogen, phosphorus and potassium in soil [J]. *Spectroscopy and Spectral Analysis*, 2014, 34(8): 2102-2105. (in Chinese)
- 5 SHAO Y, HE Y. Nitrogen, phosphorus, and potassium prediction in soils, using infrared spectroscopy [J]. *Soil Research*, 2011, 49(2): 166-172.
- 6 贾生尧, 杨祥龙, 李光, 等. 近红外光谱技术结合递归偏最小二乘算法对土壤速效磷与速效钾含量测定研究 [J]. *光谱学与光谱分析*, 2015, 35(9): 2516-2520.
- 7 JIA Shengyao, YANG Xianglong, LI Guang, et al. Quantitatively determination of available phosphorus and available potassium in soil by near infrared spectroscopy combining with recursive partial least squares [J]. *Spectroscopy and Spectral Analysis*, 2015, 35(9): 2516-2520. (in Chinese)
- 8 胡国田, 何东健. 基于直接正交信号校正的土壤磷和钾 VNIR 测定研究 [J/OL]. *农业机械学报*, 2015, 46(7): 139-145. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?file_no=20150721&flag=1. DOI:10.6041/j.issn.1000-1298.2015.07.021.
- 9 HU Guotian, HE Dongjian. Soil phosphorus and potassium estimation using visible-near infrared reflectance spectroscopy with direct orthogonal signal correction [J/OL]. *Transactions of the Chinese Society for Agricultural Machinery*, 2015, 46(7): 139-145. (in Chinese)
- 10 PAZ-KAGAN T, ZAADY E, SALBACH C, et al. Mapping the spectral soil quality index (SSQI) using airborne imaging spectroscopy [J]. *Remote Sensing*, 2015, 7(11): 15748-15781.
- 11 DE JONG S M, PEBESMA E J, LACAZE B. Above-ground biomass assessment of mediterranean forests using airborne imaging spectrometry: the DAIS Payne experiment [J]. *International Journal of Remote Sensing*, 2003, 24(7): 1505-1520.
- 12 STENBERG B, VISCARRA ROSSEL R A, MOUAZEN A M, et al. Visible and near infrared spectroscopy in soil science [J]. *Advances in Agronomy*, 2010, 107: 163-215.
- 13 CHONG I G, JUN C H. Performance of some variable selection methods when multicollinearity is present [J]. *Chemometrics and Intelligent Laboratory Systems*, 2005, 78(1): 103-112.
- 14 SARATHJITH M, DAS B, WANI S, et al. Comparison of data mining approaches for estimating soil nutrient contents using diffuse reflectance spectroscopy [J]. *Current Science*, 2016, 110(6): 1031-1037.
- 15 MOUAZEN A M, KUANG B, DE BAERDEMAEKER J, et al. Comparison among principal component, partial least squares and back propagation neural network analyses for accuracy of measurement of selected soil properties with visible and near infrared spectroscopy [J]. *Geoderma*, 2010, 158(1): 23-31.
- 16 BRAY R, KURTZ L. Determination of total, organic, and available forms of phosphorus in soils [J]. *Soil Science*, 1945,

- 59(1): 39–46.
- 14 STEFFENS M, BUDDENBAUM H. Laboratory imaging spectroscopy of a stagnic Luvisol profile—High resolution soil characterisation, classification and mapping of elemental concentrations [J]. *Geoderma*, 2013, 195: 122–132.
- 15 PEDDLE D R, WHITE H P, SOFFER R J, et al. Reflectance processing of remote sensing spectroradiometer data [J]. *Computers and Geosciences*, 2001, 27(2): 203–213.
- 16 SAVITZKY A, GOLAY M J E. Smoothing and differentiation of data by simplified least squares procedures [J]. *Analytical Chemistry*, 1964, 36(8): 1627–1639.
- 17 HELLAND I S, NÆS T, ISAKSSON T. Related versions of the multiplicative scatter correction method for preprocessing spectroscopic data [J]. *Chemometrics and Intelligent Laboratory Systems*, 1995, 29(2): 233–241.
- 18 GHOLIZADEH A A, BORUVKA L A, SABERIOON M M B, et al. Comparing different data preprocessing methods for monitoring soil heavy metals based on soil spectral features [J]. *Soil and Water Research*, 2015, 10(4): 218–227.
- 19 HSU C W, CHANG C C, LIN C J. A practical guide to support vector classification [J]. *BJU International*, 2008, 101: 1396–1400.
- 20 WOLD S, MARTENS H, WOLD H. The multivariate calibration problem in chemistry solved by the PLS method [C] // *Matrix Pencils Lecture Notes in Mathematics*, (LNN, Volume 973). Springer, Berlin, Heidelberg, 1982: 286–293.
- 21 JI W, SHI Z, HUANG J, et al. In situ measurement of some soil properties in paddy soil using visible and near-infrared spectroscopy [J]. *PLoS ONE*, 2014, 9(8): e105708.
- 22 WOLD S, SJÖSTRÖM M, ERIKSSON L. PLS-regression: a basic tool of chemometrics [J]. *Chemometrics and Intelligent Laboratory Systems*, 2001, 58(2): 109–130.
- 23 ROSSEL R A V, JEON Y S, ODEH I O A, et al. Using a legacy soil sample to develop a mid-IR spectral library [J]. *Australian Journal of Soil Research*, 2008, 46(1): 1–16.
- 24 PAZ-KAGAN T, SHACHAK M, ZAADY E, et al. A spectral soil quality index (SSQI) for characterizing soil function in areas of changed land use [J]. *Geoderma*, 2014, 230: 171–184.
- 25 WIDYANTO M R, NOBUHARA H, KAWAMOTO K, et al. Improving recognition and generalization capability of back-propagation NN using a self-organized network inspired by immune algorithm (SONIA) [J]. *Applied Soft Computing Journal*, 2005, 6(1): 72–84.
- 26 TIAN Y, ZHANG J, YAO X, et al. Laboratory assessment of three quantitative methods for estimating the organic matter content of soils in China based on visible/near-infrared reflectance spectra [J]. *Geoderma*, 2013, 202: 161–170.
- 27 PANG G, WANG T, LIAO J, et al. Quantitative model based on field-derived spectral characteristics to estimate soil salinity in Minqin County, China [J]. *Soil Science Society of America Journal*, 2014, 78(2): 546–555.
- 28 ZHANG C, YE H, LIU F, et al. Determination and visualization of pH values in anaerobic digestion of water hyacinth and rice straw mixtures using hyperspectral imaging with wavelet transform denoising and variable selection [J]. *Sensors*, 2016, 16(2): 244.
- 29 BASHEER I A, HAJMEER M. Artificial neural networks: fundamentals, computing, design, and application [J]. *Journal of Microbiological Methods*, 2000, 43: 3–31.
- 30 KENNARD R W, STONE L A. Computer aided design of experiments [J]. *Technometrics*, 1969, 11(1): 137–148.
- 31 ZHONG P, XU Y, ZHAO Y. Training twin support vector regression via linear programming [J]. *Neural Computing and Applications*, 2012, 21(2): 399–407.
- 32 CHANG C W, LAIRD D A, MAUSBACH M J, et al. Near-infrared reflectance spectroscopy-principal components regression analyses of soil properties [J]. *Soil Science Society of America Journal*, 2001, 65(2): 480–490.
- 33 JIANG Q, CHEN Y, GUO L, et al. Estimating soil organic carbon of cropland soil at different levels of soil moisture using VIS–NIR spectroscopy [J]. *Remote Sensing*, 2016, 8(9): 755.
- 34 ROSSEL R A V, BEHRENS T. Using data mining to model and interpret soil diffuse reflectance spectra [J]. *Geoderma*, 2010, 158(1): 46–54.
- 35 VISCARRA ROSSEL R A, BUI E N, DE CARITAT P, et al. Mapping iron oxides and the color of Australian soil using visible–near-infrared reflectance spectra [J]. *Journal of Geophysical Research: Earth Surface*, 2010, 115(F4): 1–13.
- 36 BAYER A, BACHMANN M, MILLER A, et al. A Comparison of feature-based MLR and PLS regression techniques for the prediction of three soil constituents in a degraded South African ecosystem [J]. *Applied and Environmental Soil Science*, 2012(3): 495–506.
- 37 HUETE A R, ESCADAFAL R. Assessment of biophysical soil properties through spectral decomposition techniques [J]. *Remote Sensing of Environment*, 1991, 35(2–3): 149–159.
- 38 BEN-DOR E, INBAR Y, CHEN Y. The reflectance spectra of organic matter in the visible near-infrared and short wave infrared region (400–2500 nm) during a controlled decomposition process [J]. *Remote Sensing of Environment*, 1997, 61(1): 1–15.
- 39 RODRIGUEZ J M, USTIN S L, RIA O D. Contributions of imaging spectroscopy to improve estimates of evapotranspiration [J]. *Hydrological Processes*, 2011, 25(26): 4069–4081.
- 40 MOUAZEN A M, MALEKI M R, DE BAERDEMAEKER J, et al. On-line measurement of some selected soil properties using a VIS–NIR sensor [J]. *Soil and Tillage Research*, 2007, 93(1): 13–27.
- 41 DANIEL K W, TRIPATHI N K, HONDA K. Artificial neural network analysis of laboratory and in situ spectra for the estimation of macronutrients in soils of Lop Buri (Thailand) [J]. *Australian Journal of Soil Research*, 2003, 41(1): 47–59.