

网络出版时间:2013-03-27 15:47

网络出版地址: <http://www.cnki.net/kcms/detail/61.1390.S.20130327.1547.006.html>

基于近红外光谱的不同建模方法检测土壤有机质和速效 P 含量的研究

刘雪梅,章海亮

(华东交通大学 土木建筑学院,江西 南昌 330013)

【摘要】 **【目的】**采用近红外光谱和不同建模方法测定土壤中的有机质和速效 P 含量。**【方法】**分别采集江西不同地区的土样 240 个,采集土壤样品的近红外漫反射光谱,以对光谱数据进行主成分分析得到的前 6 个主成分(PCs)和偏最小二乘回归(PLSR)建模得到的 6 个潜在变量(LVs),分别作为反向传播神经网络(BPNN)和偏最小支持向量机(LS-SVM)的输入变量,共建立 6 个模型,分别为主成分回归(PCR)、PLSR、BPNN-PCs、BPNN-LVs、LS-SVM-PCs 和 LS-SVM-LVs,并对这些建模方法预测土壤有机质和速效 P 含量的结果进行评价,从中筛选出最佳模型。**【结果】**在预测土壤有机质和速效 P 含量时,LS-SVM-LVs 模型的预测效果优于 PCR、PLSR、BPNN-PCs、BPNN-LVs 和 LS-SVM-PCs 模型。用 LS-SVM-LVs 模型得到的有机质、速效 P 预测集的决定系数(R^2)和均方差(RMSE)分别为 0.873 4、0.780 1 mg/kg 和 2.92 g/kg、4.97 mg/kg。**【结论】**将近红外漫反射光谱和 LS-SVM、PLSR 相结合可用于测定土壤有机质和速效 P 含量。

【关键词】 建模方法;近红外漫反射光谱;土壤;有机质;速效 P

【中图分类号】 O434.13;S151.9⁺5

【文献标志码】 A

【文章编号】 1671-9387(2013)04-0052-05

Measurement of soil organic matter and phosphorus using different calibration methods based on near-infrared spectroscopy

LIU Xue-mei, ZHANG Hai-liang

(School of Civil Engineering, East China Jiaotong University, Nanchang, Jiangxi 330013, China)

Abstract: **【Objective】** Different calibration methods based on near infrared spectroscopy were used to measure the contents of soil organic matter (OM) and available phosphorus (P). **【Method】** 240 soil samples from the different regions of Jiangxi were collected. Near infrared diffuse reflectance spectroscopy data of soil samples were used for principal component analysis (PCA) to get the first six principal components (PCs), and PLSR model was built to get six latent variables (LVs). PCR, PLSR, BPNN-PCs, BPNN-LVs, LS-SVM-PCs and LS-SVM-LVs modeling methods were built to predict the contents of soil OM and available P. These modeling methods were evaluated and the best model was selected. **【Result】** The results showed that LS-SVM-LVs model outperformed PCR, PLSR, BPNN-PCs, BPNN-LVs and LS-SVM-PCs models. The best predictions were obtained with LS-SVM-LVs model for OM ($R^2=0.873\ 4$ and $RMSE=2.92\ g/kg$) and P ($R^2=0.780\ 1$ and $RMSE=4.97\ mg/kg$). **【Conclusion】** The near-infrared diffuse reflectance spectroscopy based LS-SVM combined with PLSR can be used for the measurement of soil OM and available P.

【收稿日期】 2012-06-25

【基金项目】 江西省科技支撑项目(2009AE01603, 2010EHB02000);铁路环境振动与噪声教育部工程研究中心资助项目

【作者简介】 刘雪梅(1978—),女,江西吉安人,副教授,博士,主要从事环境工程和光谱无损检测研究。

Key words: modeling method; near infrared diffuse reflectance spectroscopy; soil; organic matter; available P

快速获取土壤养分信息是实施精细农业的基础,土壤有机质和速效 P 是鉴别土壤肥力的重要指标,也是植物生长发育所必需的营养元素,快速测定土壤的有机质和速效 P 含量,对于农业和环境等方面的研究都有重要意义^[1]。因此,有必要研究一种快速检测土壤有机质和速效 P 含量的方法。近红外光谱(NIR)分析方法具有快速、高效、无损和适合在线分析等优点,近年来,采用近红外光谱技术获取土壤养分信息,成为国内外学者研究的重点^[2-6]。

常用的估测土壤养分的建模方法有主成分回归(PCR)^[7-8]、偏最小二乘回归(PLSR)^[9-11]、逐步多元回归^[12]、反向传播神经网络(BPNN)^[7,13]和偏最小支持向量机(LS-SVM)^[13-14]等,但这些方法却表现出不同的预测精度,预测结果差异较大。因此,针对某些具体的土壤养分(如有机质和速效 P),选择最优建模方法和建模因子快速测定其含量,可为开发土壤参数便携式测定仪器提供支持。本研究采用近红外光谱技术提取土壤养分信息,在此基础上采用主成分回归(PCR)、偏最小二乘回归(PLSR)、反向传播神经网络(BPNN)和偏最小支持向量机(LS-SVM)建模方法,分析土壤有机质(OM)和速效 P 含量,并对不同建模方法的测定结果进行比较,以期能为土壤参数便携式测定仪器的开发奠定基础。

1 材料与方法

1.1 土壤样本的采集

土壤样本分别为采自江西省安福县寮塘乡汪村的棕壤(褐土)、南昌县昌东镇三联村的砖红土、遂川县枚江乡张塘村的黄土和东江村的水稻土,每个采样点各 60 个土样。由于采集的光谱信号不仅与土壤样品的化学组分有关,而且与样品的颗粒大小、形状、密度等有关,因此为了消除土壤粒径大小对预测有机质和速效 P 含量的影响,所有样本使用研钵磨碎,并分别通过孔径 0.5 mm 的筛子,得到粒径相同的样本。土壤样本采用上海精宏实验设备有限公司 DHG-9070 型电热恒温鼓风干燥箱 90 °C 风干 12 h 以上,用重铬酸钾法测定土壤有机质含量,钼锑抗比色法测定速效磷含量^[15]。

1.2 检测仪器及参数设置

光谱测定使用 ASD FieldSpec HandHeld 光谱仪(LabSpec® Pro Near Infrared Analyzer, Analyti-

cal Spectral Devices, Inc, USA), 测定范围 325 ~ 1 075 nm, 采样光谱间隔: 1 nm, 扫描时间: 100 ms/次, 检测距离 3 cm, 扫描次数: 10 次, 探头视场角为 5°。

1.3 光谱采集和预处理

将土壤样品放入直径为 80 mm、厚度为 15 mm 的玻璃培养皿中,光谱仪置于土样上方距土样表面 12 cm 处,采用漫反射方式进行样品光谱采样,光照入射角为 45°。测试中由于仪器噪声、样本粒径大小引起的散射会影响有效光谱信息的分析和提取,因此一个土壤样本采集 3 次光谱数据,为保证光谱数据具有代表性,将 3 次数据的平均值作为土壤样本最终的光谱数据。图 1 为土壤样品的近红外漫反射光谱。

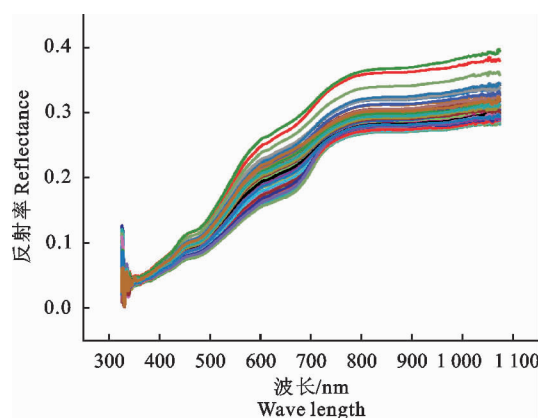


图 1 土壤样品的近红外漫反射光谱

Fig. 1 The diffuse reflection NIRS of soil samples

2 校准模型的建立

2.1 建模集和预测集的划分

由于土壤样品地理分布区域较大,为验证各种建模方法的稳定性,将 240 个样品随机分为 2 组,其中一组为样品的 90%,用于建模集,另一组是剩余的 10%,用于预测集。建模集(90%)和预测集(10%)2 种分组方法随机重复 3 次,以保证模型的稳定性。

2.2 PCR 变量和 PLSR 潜在变量的获取

PCR 是经典的特征抽取和降维技术之一,它可以在不具备任何相关知识背景的情况下对未知样品进行主成分信息提取。采用 PCR 对土壤样品近红外漫反射光谱数据进行分析后,前 3 个主成分样品的得分见图 2。从图 2 可以看出,4 个典型土壤具有

明显的聚类趋势,且前 3 个主成分可以表达原始光谱 97%的信息,其中主成分 1(PC1)的贡献率最大,为 73%;主成分 2(PC2)和主成分 3(PC3) 的贡献率较小,分别为 15%和 9%。鉴于对土壤样品近红外漫反射光谱进行主成分分析(PCA)处理后得到的前 6 个主成分(PCs)几乎可以 100%表达原始光谱信息,且可降低模型复杂度,提高模型运行速度和精度。因此,本研究选用 PCR 处理后得到的前 6 个主成分(PCs)作为 BPNN 和 LS-SVM 的输入变量进行建模。采用 PLSR 对土壤样品近红外漫反射光谱数据进行分析后,得到的 6 个潜在变量(LVs)几乎也可以 100%表达原始光谱信息。因此,本研究选用 PLSR 处理后得到的 6 个潜在变量(LVs)作为 BPNN 和 LS-SVM 的输入变量进行建模。主成分变量和潜在变量的获取都是在分析建模集模型基础上获取的,具体方法参见文献[13]。

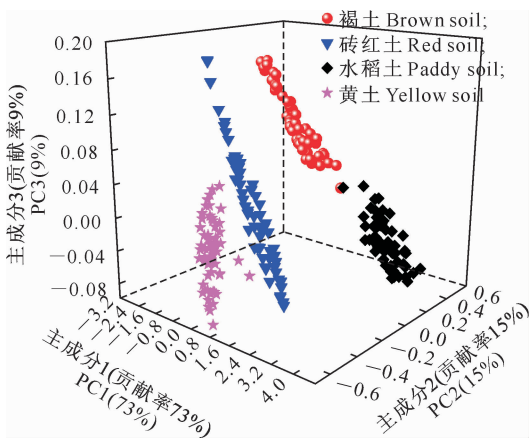


图 2 主成分回归(PCR)处理后不同土壤样品前 3 个主成分的得分

Fig. 2 Scatter plots for scores of the first three principal components

2.3 不同模型的建立方法

2.3.1 PCR 一般对数据进行 PCR 时,采取的是降维的方法,利用线性代数及有关知识,将原来多个指标转化为较少的新指标,并且这些新的指标既彼此互不相关,避免信息的交叉和重叠,又能综合反映原来多个指标的信息^[7]。PCR 的具体分析过程见文献[8]。

2.3.2 PLSR PLSR 方法是光谱数据与变量进行线性回归,其分析过程为:1)建立建模集和预测集样本变量;2)对建模集样本变量与光谱数据进行 PLSR 分析,建立建模集样本变量与光谱数据间的 PLSR 校正模型;3)根据建模集建立的 PLSR 校正模型对预测集样品进行分析。

2.3.3 BPNN-PCs 和 BPNN-LVs 与传统方法相比,BPNN 建模方法具有信息处理能力强,不依赖统计特性的特点,其包括输入层、隐含层、输出层 3 层。本研究中,网络输入层节点数为 7、隐含层节点数为 5,输出层节点数为 1。目标误差为 0.001,网络指定参数中学习速率为 0.2,设定训练迭代次数为 100 次。本研究选用 PCR 处理后得到的前 6 个主成分(PCs)变量以及 PLSR 处理后得到的 6 个潜在变量(LVs)作为 BPNN 的输入变量进行建模,所得模型分别命名为 BPNN-PCs、BPNN-LVs,具体建模过程见文献[13]。

2.3.4 LS-SVM-PCs 和 LS-SVM-LVs LS-SVM 建模方法通过非线性映射函数 $\varphi(x)$ 建立回归模型,具体步骤为:首先将输入变量映射到高维特征空间,然后将优化问题转化为等式约束条件进行分析,一般利用拉格朗日算子求解最优化问题,对各个变量求偏微分^[16]。LS-SVM 需要调节的参数为核参数 σ^2 和惩罚系数 γ 。惩罚系数 γ 主要控制对错分样本惩罚的程度,实现在错分样本的比例与算法复杂度之间的折中。本研究选用 PCR 处理后得到的前 6 个主成分(PCs)变量以及 PLSR 处理后得到的 6 个潜在变量(LVs)作为 LS-SVM 的输入变量进行建模,所得模型分别命名为 LS-SVM-PCs、LS-SVM-LVs。LS-SVM-PCs、LS-SVM-LVs 的详细建模过程见文献[16-17]。

2.4 模型的评价

本研究采用决定系数(R^2)、均方差(RMSE)对模型进行评价, R^2 值偏高、RMSE 值偏小者为优。 R^2 、RMSE 的计算方法参见文献[1]。

3 结果与分析

3.1 建模集和预测集中土样有机质和速效 P 含量

建模集和预测集中土样有机质和速效 P 含量的统计结果见表 1。由表 1 可知,有机质 3 次建模集均值为 36.7~39.0 g/kg,预测集均值为 34.3~37.0 g/kg;速效磷 3 次建模集均值为 18.25~20.12 mg/kg,预测集均值为 16.65~18.10 mg/kg。

3.2 不同建模方法对土样有机质和速效 P 含量的预测结果比较及评价

为了比较不同建模方法对土壤样品有机质和速效 P 含量的预测效果,对 PCR、PLSR、BPNN-PCs、BPNN-LVs、LS-SVM-PCs 和 LS-SVM-LVs 6 种建模方法的所得结果进行评价,结果见表 2。由表 2 可知,在测定土壤样品有机质和速效 P 含量时,无

论是在建模集还是在预测集中,LS-SVM-LVs 的 R^2 值均最高、 $RMSE$ 值均最小,表明在测定土壤有机质和速效 P 时,LS-SVM-LVs 模型的预测效果优

于其他 5 种模型。此外,每种建模方法重复 3 次所得到的标准方差相对原始测量值均很小,表示 6 种建模方法都相对稳定。

表 1 建模集和预测集中土样有机质和速效 P 含量的统计结果

Table 1 OM and P statistics results of calibration and prediction data sets

指标 Index	属性 Property	数量 Number	最小值 Min	最大值 Max	平均值 Mean	标准方差 Standard deviation
有机质/(g·kg ⁻¹) OM	建模集 1 Calibration set 1	200	20.4	49.5	39.0	7.87
	建模集 2 Calibration set 2	200	23.5	45.6	37.5	7.08
	建模集 3 Calibration set 3	200	21.3	46.7	36.7	7.95
速效 P/(mg·kg ⁻¹) Available P	建模集 1 Calibration set 1	205	6.0	28.6	19.3	6.05
	建模集 2 Calibration set 2	205	5.8	29.4	20.1	6.09
	建模集 3 Calibration set 3	205	5.3	30.6	18.3	7.07
有机质/(g·kg ⁻¹) OM	预测集 1 Prediction set 1	20	22.8	48.8	35.7	7.55
	预测集 2 Prediction set 2	20	22.9	47.7	34.3	8.92
	预测集 3 Prediction set 3	20	24.1	47.6	37.0	7.42
速效 P/(mg·kg ⁻¹) Available P	预测集 1 Prediction set 1	22	6.1	25.7	16.7	6.49
	预测集 2 Prediction set 2	22	8.1	27.3	16.7	7.25
	预测集 3 Prediction set 3	22	5.9	29.8	18.1	7.88

表 2 基于不同建模方法预测土样有机质和速效 P 含量的 R^2 、 $RMSE$ 和 SD

Table 2 R^2 , $RMSE$ and SD of OM and available P contents using different model methods

属性 Property	建模方法 Method	有机质 OM				速效 P Available P			
		R^2		$RMSE$		R^2		$RMSE$	
		计算值 Calculate value	标准方差 SD	计算值/ (g·kg ⁻¹) Calculate value	标准方差/ (g·kg ⁻¹) SD	计算值 Calculate value	标准方差 SD	计算值/ (mg·kg ⁻¹) Calculate value	标准方差/ (mg·kg ⁻¹) SD
建模集 Calibration set	PCR	0.838 8	0.002 6	3.60	0.049 4	0.715 8	0.003 0	5.27	0.137 9
	PLSR	0.844 1	0.005 6	3.57	0.049 7	0.727 2	0.010 8	5.15	0.159 5
	BPNN-PCs	0.851 7	0.010 4	3.19	0.174 3	0.734 9	0.007 5	5.09	0.116 4
	BPNN-LVs	0.884 3	0.017 2	3.46	0.052 9	0.781 7	0.015 8	4.58	0.071 2
	LS-SVM-PCs	0.901 4	0.010 7	3.29	0.030 0	0.811 6	0.010 3	4.36	0.054 6
	LS-SVM-LVs	0.901 6	0.005 8	2.83	0.090 1	0.824 6	0.010 4	4.22	0.080 9
预测集 Prediction set	PCR	0.808 1	0.009 3	3.75	0.055 8	0.694 6	0.010 9	6.00	0.067 4
	PLSR	0.829 7	0.008 6	3.66	0.053 7	0.712 1	0.010 3	5.85	0.125 6
	BPNN-PCs	0.848 2	0.017 4	3.29	0.174 3	0.724 1	0.002 8	5.65	0.042 7
	BPNN-LVs	0.869 1	0.016 1	3.47	0.095 3	0.731 2	0.004 9	5.42	0.092 4
	LS-SVM-PCs	0.858 9	0.010 0	3.37	0.159 6	0.754 6	0.010 5	5.09	0.074 6
	LS-SVM-LVs	0.873 4	0.015 2	2.92	0.100 0	0.780 1	0.012 0	4.97	0.088 1

由表 2 还可知,在建模集中,对于有机质的 LS-SVM-LVs 模型,其 R^2 为 0.901 6,优于速效 P 的 LS-SVM-LVs 模型($R^2=0.824\ 6$);在预测集中,对于有机质的 LS-SVM-LVs 模型,其 R^2 为 0.873 4,也优于速效 P 的 LS-SVM-LVs 模型($R^2=0.780\ 1$),表明 LS-SVM-LVs 模型对有机质的预测效果优于速效 P。分析原因主要是因为相对土壤速效 P 而言,土壤有机质含有 C—H+C—H,C—H+

C—C 和 N—H 键组合,其对近红外光谱反应更灵敏;而对速效 P 也有较好的预测效果,可能是因为近红外光通过对土壤中与磷元素有关活性成分和矿物质间接地捕捉,使得采用近红外光谱也能间接反映速效 P 的含量。目前,关于土壤(如速效 P 和速效 K 等)理化性质与近红外光谱预测相关性的机理尚无定论。

从图 3 可以看出,基于 LS-SVM-LVs 模型得到

的有机质和速效 P 含量的预测值与实测值的拟合效果均较好,说明构建的有机质和速效 P 含量的

LS-SVM-LVs 预测模型不存在过拟合和欠拟合现象,具有一定的稳定性和适应性。

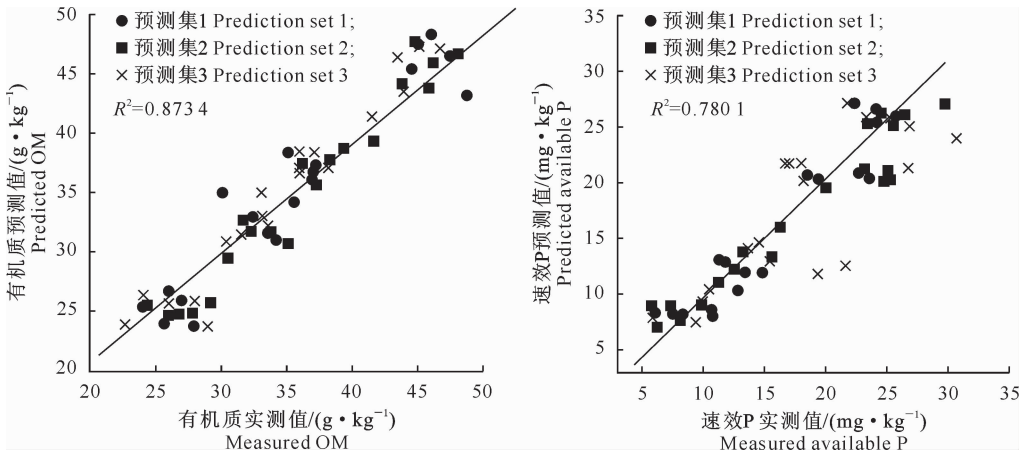


图 3 基于 LS-SVM-LVs 模型所得土壤样品有机质和速效 P 含量的预测值与实测值的拟合结果

Fig. 3 Fitting results of the predicted and measured values of OM and available P based on LS-SVM-LVs model

4 结 论

本研究在采集土壤样品近红外漫反射光谱的基础上,分别采用 PCR、PLSR、BPNN-PCs、BPNN-LVs、LS-SVM-PCs 和 LS-SVM-LVs 6 种建模方法,预测土壤有机质(OM)和速效 P 含量,并对不同建模方法的测定结果进行比较,得出了以下结论:

1)在测定土壤有机质和速效 P 含量时,LS-SVM-LVs 模型优于本研究中的其他 5 种模型,可知采用 PLSR 处理所得潜在变量作为 LS-SVM 建模因子结果较优。

2)用 LS-SVM-LVs 模型预测有机质含量的决定系数(R^2)大于速效 P,说明该模型对有机质的预测效果优于速效 P。

由本研究结果可知,基于近红外光谱测定土壤有机质和速效 P 含量时,建议采用 LS-SVM-LVs 模型。然而,对与近红外光谱无直接关系的土壤性质而言,还需要对其机理进行进一步的研究。本研究结果为将近红外漫反射光谱检测技术应用于土壤其他营养成分的分析提供了参考。

[参考文献]

[1] 岑益郎,宋 韬,何 勇,等. 基于可见/近红外漫反射光谱的土壤有机质含量估算方法研究 [J]. 浙江大学学报:农业与生命科学版,2011,37(3):300-306.
Cen Y L, Song T, He Y, et al. Rapid detection method of soil organic matter contents using visible/near infrared diffuse reflectance spectral data [J]. Journal of Zhejiang University: Agric& Life Sci, 2011, 37(3): 300-306. (in Chinese)
[2] 李 颀,张小超,苑严伟,等. 北京典型耕作土壤养分的近红外

光谱分析 [J]. 农业工程学报, 2012, 28(2): 176-179.
Li J, Zhang X C, Yuan Y W, et al. Analysis of soil nutrient content based on near infrared reflectance spectroscopy in Beijing region [J]. Transactions of the CSAE, 2012, 28(2): 176-179. (in Chinese)
[3] 陈鹏飞,刘良云,王纪华,等. 近红外光谱技术实时测定土壤中总氮及磷含量的初步研究 [J]. 光谱学与光谱分析, 2008, 28(2): 295-298.
Chen P F, Liu L Y, Wang J H, et al. Real-time analysis of soil N and P with near infrared diffuse reflectance spectroscopy [J]. Spectroscopy and Spectral Analysis, 2008, 28(2): 295-298. (in Chinese)
[4] Vohland M, Emmerling C. Determination of total soil organic C and hot water-extractable C from VIS-NIR soil reflectance with partial least squares regression and spectral feature selection techniques [J]. European Journal of Soil Science, 2011, 62(4): 598-606.
[5] Aitkenhead M J, Coull M C, Towers W, et al. Predicting soil chemical composition and other soil parameters from field observations using a neural network [J]. Computers and Electronics in Agriculture, 2012, 82: 108-116.
[6] O'Rourke S M, Holden N M. Determination of soil organic matter and carbon fractions in forest top soils using spectral data acquired from visible-near infrared hyperspectral images [J]. Soil Science Society of America Journal, 2012, 76(2): 586-596.
[7] 张淑娟,王凤花,张海红,等. 基于主成分分析和 BP 神经网络的土壤养分近红外光谱检测 [J]. 山西农业大学学报:自然科学版, 2009, 29(6): 483-487.
Zhang S J, Wang F H, Zhang H H, et al. Near-infrared determination of soil nutrients based on principal component analysis and BP neural network [J]. Jour of Shanxi Agric Univ: Natural Science Edition, 2009, 29(6): 483-487. (in Chinese)