

不同核函数支持向量机和可见-近红外光谱的 多种植被叶片生化组分估算

陈方圆^{1,2}, 周鑫^{1,2}, 陈奕云^{1,2}, 王奕涵³, 刘会增^{4,5}, 王俊杰^{5,6}, 邬国锋^{1,5,6*}

1. 武汉大学资源与环境科学学院, 湖北 武汉 430079
2. 武汉大学教育部地理信息系统重点实验室, 湖北 武汉 430079
3. 湖北省测绘工程院, 湖北 武汉 430074
4. 香港浸会大学地理系, 中国 香港
5. 深圳大学国家测绘地理信息局海岸带地理环境监测重点实验室及深圳市空间信息智能感知与服务重点实验室, 广东 深圳 518060
6. 深圳大学生命与海洋科学学院, 广东 深圳 518060

摘要 氮、磷、钾元素是植物有机质的重要生化组分, 准确估算其含量对监测管理植被的新陈代谢和健康状况具有重要意义。可见-近红外光谱结合多种建模方法已被用于植被生化参数的监测, 其中支持向量机回归方法被证明能够较好拟合反射光谱和植被生化参数之间的非线性关系, 而选取适当的核函数是其成功的关键。以宜兴地区水稻、玉米、芝麻、大豆、茶叶、草地、乔木和灌木等八种植被叶片样本为研究对象, 分析比较基于径向基核函数、多项式核函数和S形核函数的支持向量回归模型估算叶片氮、磷、钾元素含量的能力。利用一阶微分变换、标准正态变量变换和反对数变换对叶片可见-近红外光谱进行预处理, 运用 bootstrapping 法生成 1 000 组校正集和验证集, 分别建立基于三种核函数的支持向量回归估算模型, 以决定系数 (R^2) 和相对分析误差 (RPD) 的均值作为评价指标。结果显示, 结合一阶微分和反对数变换光谱, 采用径向基核函数模型对氮、钾元素估算精度最高(氮: 平均 $R^2 = 0.64$, 平均 RPD = 1.67; 钾: 平均 $R^2 = 0.56$, 平均 RPD = 1.48), 结合一阶微分变换光谱, 采用径向基核函数模型对磷元素估算精度最高(磷: 平均 $R^2 = 0.68$, 平均 RPD = 1.73)。研究表明, 结合不同预处理的可见-近红外光谱, 基于径向基核函数的支持向量回归模型具有较好的估算多种植被叶片生化组分含量的潜力。

关键词 核函数; 支持向量机; 可见-近红外光谱; 生化组分

中图分类号: S127 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2019)02-0428-07

引言

植被有机质控制着植被的新陈代谢过程和健康状况, 例如营养循环、光合作用、植物生产力和凋落物分解速率等。氮(N), 磷(P)和钾(K)元素是植被有机质的三种重要生化组分, 快速掌握这三种元素的含量(c_N , c_P 和 c_K), 对于监测植被生长过程中的新陈代谢过程和健康状况至关重要^[1]。

传统化学分析手段测定植被氮、磷、钾元素含量, 具有较高的准确性, 但是过程复杂费时, 且对植株本身具有破坏性, 同时, 传统方法手段也无法提供这些元素含量的时空特

征信息。而遥感技术具有精细尺度下数种植被生化参数的估算潜力^[1], 植被可见光和近红外(visible and near-infrared, VNIR)波段的反射率光谱包含了纤维素、蛋白质、木质素以及其他参数的综合光谱信息, 可被考虑用于估算氮、磷、钾元素的含量。许多研究发现植被氮元素含量与叶片尺度的VNIR光谱相关性密切^[2-3], 也有研究表明VNIR光谱与植被磷、钾元素含量之间存在相关关系^[4], 因此, 在精细时空尺度下, 植被的VNIR反射率光谱具备相对快速、低成本量化解分析植被营养水平的潜力。

植被的光谱反射率与生化组分含量之间存在非线性关系, 尤其在近红外波段区间, 因而需要探究非线性模型来解

收稿日期: 2018-05-23, 修订日期: 2018-09-30

基金项目: 2017 年国家重点研发计划(2017YFC0506206), 深圳市科技创新委员会基础研究学科布局项目(JCYJ20151117105543692)资助

作者简介: 陈方圆, 1988 年生, 武汉大学资源与环境科学学院博士研究生 e-mail: fangyuanchen@whu.edu.cn

* 通讯联系人 e-mail: guofeng.wu@szu.edu.cn

决问题。支持向量机(support vector machine, SVM)是一种新的基于统计学习理论的机器学习方法,其利用结构风险最小化原则避免过拟合问题,在最小化经验风险下所得结果优于传统的神经网络算法,而且 SVM 在小样本、高维度数据情况下具有优异的建模能力^[5]。若干研究已证明了支持向量机回归(support vector machine regression, SVR)在估算植被生化参数方面的潜力,例如 Wang 等^[6]运用 SVR 模型和 VNIR 光谱对小麦叶片中的氮元素进行量化研究, Yao 等^[7]利用包括 SVR 模型在内的多种算法进行小麦氮元素含量的估算, Zhai 等^[8]基于 VNIR 光谱,结合 SVR 模型和偏最小二乘回归(partial least squares regression, PLSR)方法估算植被叶片的氮、磷、钾元素含量,结果表明结合室内 VNIR 光谱测量的 SVR 方法具备估测植被生化组分含量的可行性。

SVR 模型中的核函数可将目标空间映射到更高维的特征空间,将目标空间内自变量和因变量之间的非线性关系转化为特征空间中的线性关系进行处理,常用的核函数包括线性(linear)核函数,径向基(radial basis function, RBF)核函数,多项式(polynomial)核函数和 S 形(sigmoid)核函数等^[9]。核函数的选择是构建 SVR 模型的关键,决定了模型的复杂度和表现。若干研究对比分析了选用不同核函数建模的效果;刘文杰等^[9]利用四种核函数对玉米种子的高光谱数据进行分析,结果表明径向基核函数表现最优, S 形核函数表现最差; Yang 等^[10]通过对比分析径向基核函数、多项式核函数和 ANOVA 核函数估算水稻叶面积指数的能力,发现多项式核函数预测结果最优。可见,对于不同的样本种类、不同

的生化参数,核函数的估算能力不尽相同。当前关注不同核函数估测多种植被生化参数表现的研究还不多,而对比不同核函数的估测表现,对于提高植被生化参数估算精度颇具意义。

以江苏省宜兴地区的多种植被为研究对象,基于室内 VNIR 光谱,比较采用径向基核函数、多项式核函数和 S 形核函数的 SVR 模型在估算多种植被叶片氮、磷、钾元素含量方面的有效性和稳健性,加深我们对利用 VNIR 光谱和 SVR 模型进行植被生化参数反演的理解。

1 实验部分

1.1 研究区域及样本采集

宜兴市地处江苏省南部(31°07'—31°37'N, 119°31'—120°03'E),属亚热带季风区,四季分明,年平均气温 16℃,年平均降水量 1 498.8 mm。该市南部为低山丘陵,北部为平原低洼,土地肥沃,主要植被种类包括北部的水稻、小麦、茶叶等农作物和南部的毛竹等林木。

2010 年 8 月 11 日—14 日采集了包括水稻(14 个)、玉米(13 个)、芝麻(12 个)、大豆(11 个)、茶叶(11 个)、草地(13 个)、乔木(10 个)和灌木(11 个)在内的八种植被共计 95 个叶片样本,每类植被中的样本点均随机选取(图 1),采样时记录下每个样本点的地理坐标,并将剪下的新鲜植被叶片用塑料袋保存好,当天送回实验室进行室内分析。

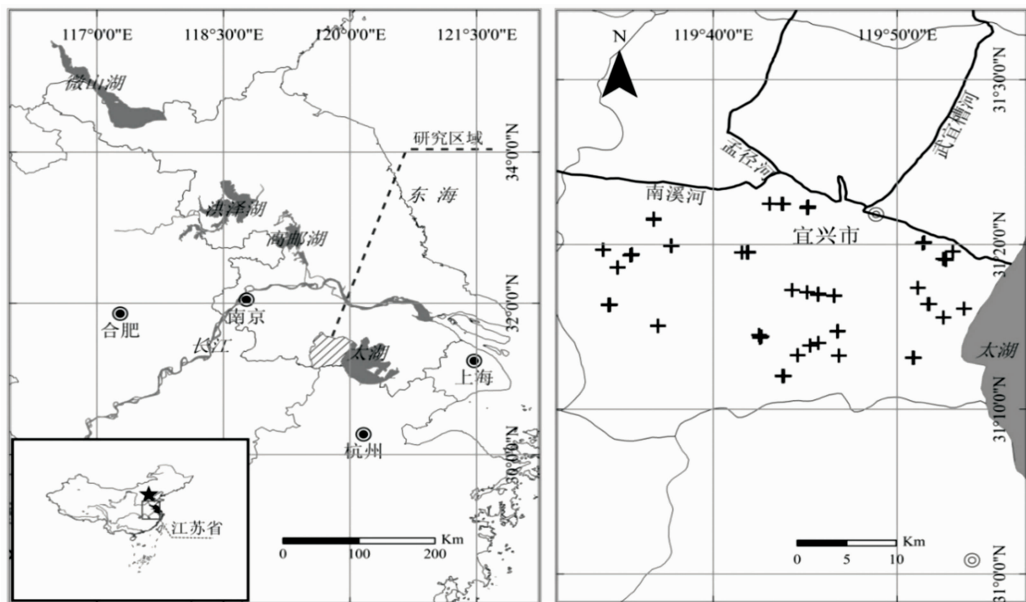


图 1 研究区域(a)和采样点(+)分布图(b)

Fig. 1 Maps showing the Yixing region in China (dashed area) (a) and the distribution of sampling sites (+) (b)

1.2 室内光谱测量与分析

采用 ASD FieldSpec3 地物光谱仪测量植被叶片反射光谱。光谱范围为 350~2 500 nm,其中 350~1 000 nm 之间采样间隔为 1.4 nm, 1 000~2 500 nm 之间为 2 nm。光谱测量在暗室进行,将当天采集的新鲜叶片样本均匀铺盖在无反射

的黑色托盘内,采用 50 W 卤素灯作唯一光源,光源距离样本中心 30 cm,入射角为 15°,光谱仪探头位于样本垂直上方 15 cm,其视场角为 25°。测试前先对仪器进行标准白板校正,测试过程中每测六个样本便重新校正一次,确保测量稳定。每个样本采集 10 条光谱曲线,进行算术平均,得到样本

点的光谱反射率数据。

植被叶片样本经过光谱测量后,分别进行叶片氮、磷、钾元素含量的化学分析。其中叶片中氮元素的含量采用凯氏定氮法测定,磷元素的含量采用钼锑抗比色法测定,钾元素的含量采用火焰光度计法测定^[8]。

1.3 数据预处理

对分析得到的氮、磷、钾元素含量数据分别进行异常点检测,当某一样本落在样本集学生化残差 ± 2.5 范围之外且大于两倍平均杠杆值时,将其视作异常点予以剔除,以此确保模型的准确度。经计算,氮元素样本集中剔除了两个异常点,磷元素样本集中剔除了两个异常点,钾元素样本集中剔除了两个异常点。

去除叶片反射光谱中噪声较大的 350~399 和 2 401~2 500 nm 边缘波段,选取 400~2 400 nm 波段以 10 nm 为间隔进行重采样,得到 401~2 391 nm 波段原始反射光谱。在结合前人研究的基础之上,采用三种光谱预处理方法消减背景噪声,首先对原始反射光谱进行一阶微分变换(first derivative, FD),随后对一阶微分变换光谱分别采用标准正态变量变换(standard normal variate, SNV)和倒数的对数变换(logarithmic reciprocal transformation, Log(1/R))。以上光谱预处理运用 Parles V3.0 软件完成。

1.4 方法

1.4.1 支持向量回归及核函数

支持向量模型可表述为

$$f(x) = \sum_{i=1}^n a_i y_i K(x_i, x) - b, 0 \leq a_i \leq C \quad (1)$$

式中 $K(x_i, x)$ 为核函数, x_i 为输入向量(本研究中为光谱反射率数据), x 是用于计算更高维特征空间的数据项, y_i 为输出向量(本研究中为生化组分含量), a_i 为拉格朗日因子矩阵, n 为样本数量, b 为残差, C 为正则化参数,用以控制超出误差的样本的惩罚程度。核函数可将非线性关系映射到高维空间中,构建线性回归进行处理,本研究分别采用了多项式核函数[式(2)],径向基核函数[式(3)]和 S 形核函数[式(4)]

$$K(x_i, x) = (k \langle x_i, x \rangle + c)^d \quad (2)$$

$$K(x_i, x) = \exp\left(-\frac{\|x_i, x\|^2}{2\sigma^2}\right) \quad (3)$$

$$K(x_i, x) = \tanh(k' \langle x_i, x \rangle + \vartheta) \quad (4)$$

其中 $c, d, k, \sigma, \vartheta$ 和 k' 等核函数中的参数由优化算法确定^[11]。

1.4.2 模型建立与验证

从样本集中随机选取 70%(65 个)样本作为校正集,针对每一种核函数,在 Matlab2010b 环境下利用 LIBSVM 工具箱进行支持向量回归模型的校正^[12]。三种核函数中的参数采用网格参数寻优法和五折交叉验证确定,取交叉验证过程中均方根误差(root mean square error, RMSE)最小时所对应的各参数为最优值^[8]。剩余 30%(28 个)样本作为验证集,以检验校正所得 SVR 模型的表现,模型精度评价指标采用决定系数(determination coefficient, R^2)和相对分析误差(ratio of performance to standard deviate, RPD)。上述评价指标中,

R^2 越大,模型预测效果越好,此外,当 RPD 在 1.4~2 时,表示模型有一定的预测能力,在 2~2.5 时表示模型预测能力良好,大于 2.5 时则表明模型有很好的预测能力。

为减少随机抽样建模所导致的误差,采用 bootstrapping 方法,对原始 93 个样本进行有放回地抽样,获得 1000 组校正集和验证集,将上述模型校正和验证流程重复 1 000 次,并计算出 1 000 次模型校正和验证的 R^2 均值与 RPD 均值,用以评价基于不同核函数的模型表现。

2 结果与讨论

2.1 叶片生化组分含量与反射光谱

植被叶片样本生化组分含量统计如表 1 所示,不同种类植被叶片的氮、磷、钾元素含量差异较大,可达 10 倍以上。八种植被叶片的平均反射光谱曲线总体形态比较一致,反射率高低存在差异,如图 2 所示。其中在 550 nm 附近因叶绿素的弱吸收而形成反射峰,700 至 750 nm 区间内出现反射率增长最快的“红边”(red edge),780~1 300 nm 近红外波段区间反射率较高,主要是叶片细胞组织反射造成,1 450 和 1 950 nm 附近的反射谷则主要是由水分的吸收引起。

表 1 植被叶片样本氮、磷、钾元素含量统计

Table 1 Descriptive statistics of the contents of nitrogen (c_N), phosphorus (c_P) and potassium (c_K) of plant leaf samples

	样本数	均值	标准差	最小值	中值	最大值
$c_N/\%$	93	2.406 8	0.976 2	0.319 8	2.494 8	5.099 8
$c_P/\%$	93	0.169 2	0.124 3	0.000 7	0.161 2	0.583 5
$c_K/\%$	93	1.701 8	0.811 7	0.465 8	1.623 6	4.481 5

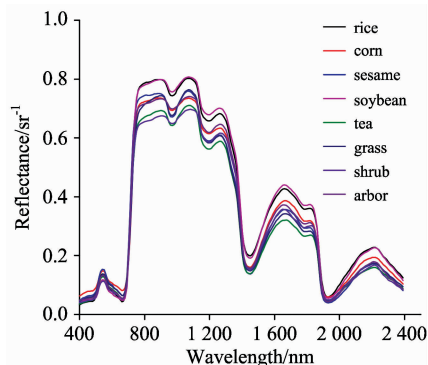


图 2 八种植被叶片平均反射光谱曲线

Fig. 2 The average original reflectance spectra of leaf samples from eight kinds of plants

2.2 光谱预处理

对植被叶片氮、磷、钾元素含量与预处理光谱及原始光谱的相关关系进行分析,其结果如图 3 所示。叶片原始光谱与各生化组分含量之间的相关系数都比较小(c_N : -0.22~0.20, c_P : -0.21~0.19, c_K : -0.37~0.12),经预处理后的光谱与三种含量之间的相关系数都有显著增大,某些波段

达到 0.05 显著性水平。而三种预处理方法对相关关系的提升作用相当,与一阶微分变换相比(最大相关系数分别为 c_N :

0.54, c_P : 0.63 和 c_K : -0.66), 进一步采用 SNV 变换或 $\text{Log}(1/R)$ 变换仅略微提高相关关系(<0.03)。

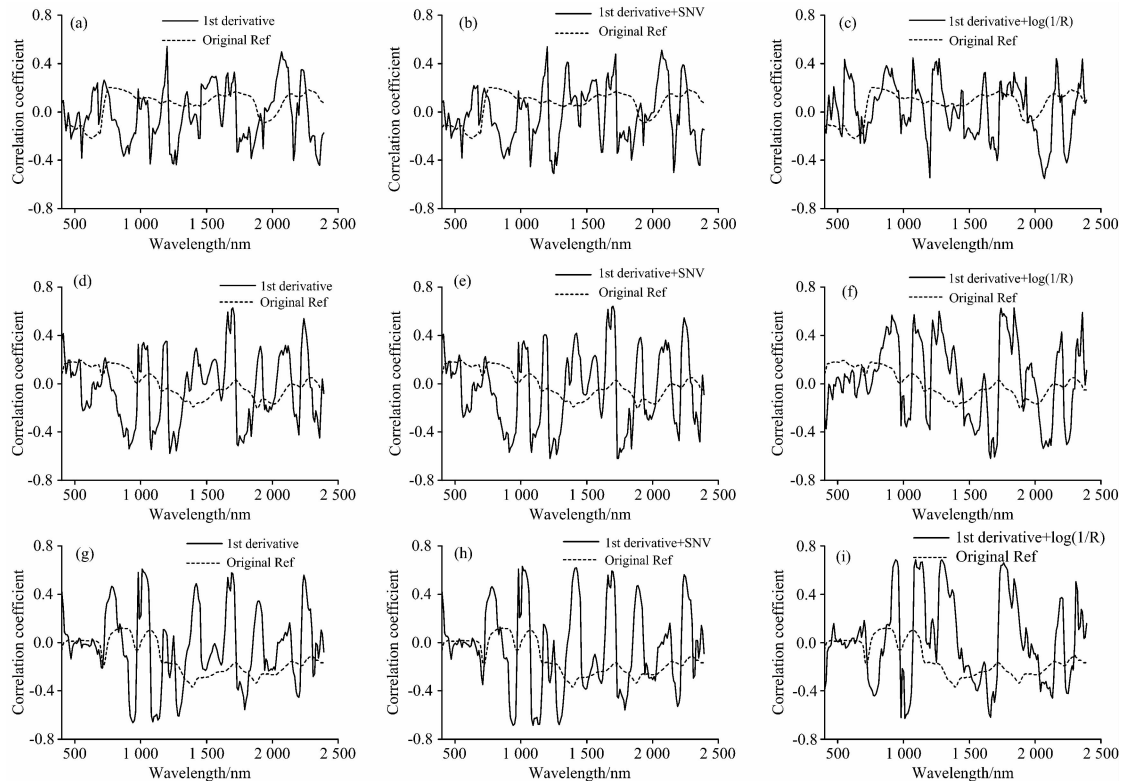


图 3 植被叶片 c_N , c_P , c_K 与预处理光谱的相关性

注: 虚线代表生化组分含量与原始光谱的相关性

(a): 一阶微分变换与 c_N ; (b): 一阶微分和 SNV 变换与 c_N ; (c): 一阶微分和 $\text{Log}(1/R)$ 变换与 c_N ; (d): 一阶微分变换与 c_P ; (e): 一阶微分和 SNV 变换与 c_P ; (f): 一阶微分和 $\text{Log}(1/R)$ 变换与 c_P ; (g): 一阶微分变换与 c_K ; (h): 一阶微分和 SNV 变换与 c_K ; (i): 一阶微分和 $\text{Log}(1/R)$ 变换与 c_K

Fig. 3 Correlations of the contents of nitrogen (c_N), phosphorus (c_P) and potassium (c_K) against the original reflectance (dash line) and their derived values with different pre-processing methods

Note: the dashed lines refer to the correlations between the biochemical components and original reflectance spectra

(a): First derivative; (b): first derivative plus standard normal variate (SNV); (c): First derivative plus absorbance transformation ($\text{Log}(1/R)$) for c_N ; (d): First derivative; (e): First derivative plus SNV; (f): First derivative plus $\text{Log}(1/R)$ for c_P ; (g): First derivative; (h): First derivative plus SNV; (i): First derivative plus $\text{Log}(1/R)$ for c_K

2.3 基于不同核函数的模型结果比较

利用三种不同预处理方法和三种核函数, 建立支持向量机回归模型估算植被叶片氮、磷、钾元素含量, 其结果如表 2 所示。总体上看, 基于 RBF 核函数模型的精度最佳, 多项式核函数模型结果次之, 而 S 形核函数模型效果最差。模型校正结果显示, 基于一阶微分变换的 RBF 核函数模型对三种组分含量的反演结果最优; 模型验证结果表明, 基于一阶微分和 $\text{Log}(1/R)$ 变换的 RBF 核函数模型对氮、钾元素含量的反演精度最佳(c_N : 平均 $R^2=0.64$, 平均 $\text{RPD}=1.67$; c_K : 平均 $R^2=0.56$, 平均 $\text{RPD}=1.48$), 而基于一阶微分变换的 RBF 核函数模型则是磷元素的最佳估算模型(平均 $R^2=0.68$, 平均 $\text{RPD}=1.73$)。分别选出验证过程中 R^2 最高的模型, 其对生化组分含量的预测值与实测值在 0.05 显著性水平下均存在强相关性(c_N : $R^2=0.85$, $\text{RPD}=2.63$; c_P : $R^2=0.87$, $\text{RPD}=2.80$; c_K : $R^2=0.83$, $\text{RPD}=2.30$), 如图 4 所

示。

基于 RBF 核函数的 SVR 模型对植被叶片氮、磷、钾元素含量的预测优于 S 形核函数和多项式核函数模型。植被叶片生化组分含量与其 VNIR 光谱之间的关系通常是非线性的, 且会受到叶片生化组分特征及实验条件的影响。RBF 核函数(也称高斯核函数)能够实现高维空间的非线性映射, 即使在样本分布未知情况下, 其旋转对称性可确保不造成大的偏差^[5], 而 S 形核函数在某些参数条件下不满足核函数半正定性要求, 其应用会受到限制^[13]。此外, 与多项式核函数和 S 形核函数相比, RBF 核函数的待优化参数较少, 模型复杂度更低, 稳健性更好。综合他人研究^[6]和本实验结果, 基于 VNIR 光谱和支持向量机进行植被叶片生化组分含量估算时, 可优先考虑 RBF 核函数。

传统的估算植被生化参数方法包括多元线性回归(MLR)和偏最小二乘回归(PLSR)。多元线性回归法一般选

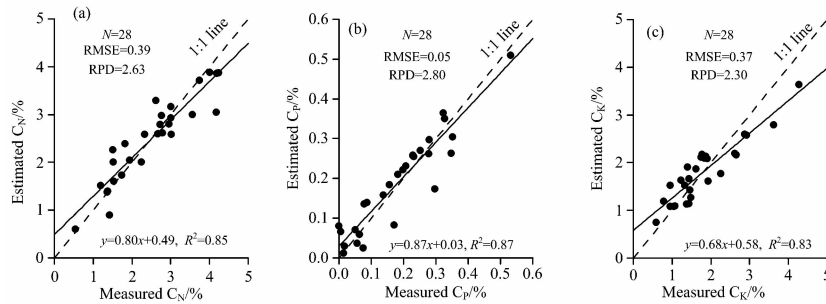


图 4 验证过程中 R^2 最高模型的预测表现

(a): c_N ; (b): c_P ; (c): c_K

Fig. 4 Performance of the best models for estimating the contents of nitrogen (c_N) (a), phosphorus (c_P) (b) and potassium (c_K) (c) with the validation dataset

表 2 基于径向基核函数、S 形核函数和多项式核函数的植被叶片 c_N , c_P , c_K 估算 SVR 模型结果

Table 2 Performances of support vector regression models with radial basis function kernel, sigmoid kernel and polynomial kernel for estimating the contents of nitrogen (c_N), phosphorus (c_P) and potassium (c_K) of plant leaves

	FD			FD+SNV			FD+Log(1/R)		
	RBF ^a	Sig ^b	Poly ^c	RBF	Sig	Poly	RBF	Sig	Poly
c_N									
R^2 *	0.93	0.84	0.88	0.92	0.80	0.89	0.91	0.80	0.89
RPD*	7.80	2.97	3.62	5.78	2.78	4.23	4.19	2.66	3.40
R_V^2 **	0.58	0.56	0.58	0.60	0.58	0.61	0.64	0.59	0.60
RPD _V **	1.51	1.47	1.51	1.55	1.50	1.56	1.67	1.52	1.59
c_P									
R^2	0.99	0.83	0.83	0.89	0.72	0.75	0.94	0.78	0.85
RPD	32.97	3.04	4.30	6.17	2.05	2.18	15.95	2.59	3.54
R_V^2	0.68	0.53	0.50	0.54	0.51	0.52	0.57	0.54	0.56
RPD _V	1.73	1.40	1.34	1.44	1.34	1.39	1.47	1.40	1.44
c_K									
R^2	0.94	0.75	0.78	0.89	0.73	0.75	0.87	0.69	0.76
RPD	11.04	2.11	2.43	6.17	2.05	2.18	4.35	1.97	2.26
R_V^2	0.55	0.50	0.52	0.54	0.51	0.52	0.56	0.48	0.52
RPD _V	1.46	1.37	1.39	1.44	1.38	1.39	1.48	1.34	1.39

^a RBF: 径向基核函数; ^b Sig: S 形核函数; ^c Poly: 多项式核函数; * R^2 : 校正集平均决定系数; * RPD: 校正集平均相对分析误差; ** R_V^2 : 验证集平均决定系数; ** RPD_V: 验证集平均相对分析误差

择若干个波段参与运算,但波段选择没有考虑植被的吸收特征,容易忽略与目标参数密切相关的光谱信息;偏最小二乘回归是多元线性回归与主成分分析(PCA)的组合,克服了波段选择和多元共线性的不足,但与多元线性回归同属线性方法,对非线性关系的处理能力有限^[6]。本工作所用的支持向量机回归方法,通过合适的核函数将非线性关系映射到高维空间,可有效处理变量间非线性关系;若干研究对比了 PLSR 和 SVR 对非线性关系数据的建模预测能力,发现 SVR 表现更佳^[6,14]。人工神经网络(ANN)法在处理非线性问题时同样有较好的效果^[15],但其存在模型物理意义不明、

模型过拟合等不足^[7]。因此,在利用植被叶片光谱进行生化组分参数估算时,基于 RBF 核函数的支持向量机回归模型具有很大潜力。

基于 RBF 核函数的支持向量机回归模型,针对八种不同植被叶片的光谱数据,估算氮、磷、钾等三种元素含量的平均精度为 0.64,0.68 和 0.56,最佳精度可达 0.85,0.87 和 0.83。Yao 等^[7]利用支持向量机方法估算小麦叶片氮含量的精度可达 0.78, Axelsson 等^[14]结合各种支持向量机方法基于景观尺度的红树林高光谱数据反演其叶片的氮、磷、钾、钙等化学元素的含量,发现只有氮元素的反演结果令人满意($R^2=0.67$)。与这些研究相比,本工作所得模型精度适中,稳健性较好,表明建立一个针对混合植被类型的多生化参数估算模型是可以实现的。然而本研究采用的是基于植被叶片尺度的室内光谱测量,对模型在不同季节,不同尺度,不同光谱数据时的适用性,还需进一步验证;同时,对于其他潜在的光谱预处理方法和核函数模型,也有待进一步探究比较。

3 结 论

利用八种植被叶片的可见-近红外光谱数据,结合不同光谱预处理方法,构建了基于径向基核函数、多项式核函数和 S 形核函数的支持向量回归模型,并比较这三种不同核函数的模型估算叶片氮、磷、钾元素含量的表现。得到结论如下:(1)利用一阶微分变换、标准正态变量变换及倒数的对数变换等方法对植被叶片的可见-近红外光谱进行预处理,可有效降低光谱噪声,显著提高模型的估算精度。(2)基于径向基核函数的支持向量回归模型估算精度最好,基于多项式核函数所得结果次之,基于 S 形核函数建立模型估算精度最差。利用径向基核函数建立的模型估算氮、磷、钾元素含量的决定系数和相对分析误差分别为 0.64,0.68,0.56 和 1.67,1.73,1.48。(3)针对多种类别混合组成的植被叶片样本,基于径向基核函数的支持向量回归模型预测精度尚可,表明其具有一定的适用性和稳健性,后续工作将进一步利用不同季节、不同尺度下的数据对模型进行验证。

References

- [1] Mahajan G, Sahoo R, Pandey R, et al. Precision Agriculture, 2014, 15(5): 499.
- [2] Tang R, Chen X, Li C. Applied Spectroscopy, 2018, 72(5): 740.
- [3] Cui L, Fei T, Qi Q, et al. International Journal of Remote Sensing, 2013, 34(5): 1866.
- [4] Sanches I, Tuohy M, Hedley M, et al. International Journal of Remote Sensing, 2013, 34(1): 276.
- [5] Vapnik V N. The Nature of Statistics Learning Theory. 2nd ed. Springer-Verlag New York Inc. , 2000.
- [6] Wang L, Zhou X, Zhu X, et al. Computers and Electronics in Agriculture, 2017, 140: 327.
- [7] Yao X, Huang Y, Shang G, et al. Remote Sensing, 2015, 7(11): 14939.
- [8] Zhai Y, Cui L, Zhou X, et al. International Journal of Remote Sensing, 2013, 34(7): 2502.
- [9] LIU Wen-jie, LI Wei-jun, LI Hao-guang, et al(刘文杰, 李卫军, 李浩光, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2017, 37(7): 2024.
- [10] Yang X, Huang J, Wu Y, et al. Science China Life Sciences, 2011, 54(3): 272.
- [11] Smola A, Schölkopf B. Statistics and Computing, 2004, 14(3): 199.
- [12] Chang C C, Lin C J. ACM Transactions on Intelligent Systems and Technology (TIST), 2011, 2(3): Article 27 (doi: 10.1145/1961189.1961199).
- [13] Zhou X, Ma Y. Communications in Statistics-Simulation and Computation, 2013, 42(10): 2175.
- [14] Axelsson C, Skidmore A, Schlerf M, et al. International Journal of Remote Sensing, 2013, 34(5): 1724.
- [15] Xu S, Zhao Y, Wang M, et al. Geoderma, 2018, 310: 29.

Estimating Biochemical Component Contents of Diverse Plant Leaves with Different Kernel Based Support Vector Regression Models and VNIR Spectroscopy

CHEN Fang-yuan^{1, 2}, ZHOU Xin^{1, 2}, CHEN Yi-yun^{1, 2}, WANG Yi-han³, LIU Hui-zeng^{4, 5}, WANG Jun-jie^{5, 6},
WU Guo-feng^{1, 5, 6*}

1. School of Resource and Environmental Sciences, Wuhan University, Wuhan 430079, China

2. Key Laboratory of Geographic Information System of the Ministry of Education, Wuhan University, Wuhan 430079, China

3. Surveying and Mapping Engineering Institute of Hubei Province, Wuhan 430074, China

4. Department of Geography, Hong Kong Baptist University, Hong Kong, China

5. Key Laboratory for Geo-Environmental Monitoring of Coastal Zone of the National Administration of Surveying, Mapping and GeoInformation & Shenzhen Key Laboratory of Spatial Smart Sensing and Services, Shenzhen University, Shenzhen 518060, China

6. College of Life Sciences and Oceanography, Shenzhen University, Shenzhen 518060, China

Abstract Nitrogen (N), phosphorus (P) and potassium (K) are important biochemical components of plant organic matters, and estimating their contents are useful for monitoring plant metabolism processes and health. Visible and near-infrared (VNIR) spectroscopy has been applied to monitor plant biochemical parameters with many modeling methods, in which support vector machine (SVM) has been proved to be a potential approach for modeling the nonlinear relationships between the reflectance spectra and biochemical parameters of plant organic matters, and the successful application of SVM relies on the proper selection of kernels. This study aimed to compare the performances of radial basis function (RBF), polynomial and sigmoid kernels based support vector machine regression (SVR) models in estimating the contents of nitrogen (c_N), phosphorus (c_P) and potassium (c_K) of diverse plant leaves using laboratory-based VNIR spectroscopy. The c_N , c_P , c_K and VNIR reflectance of leaf samples in eight plant species (rice, corn, sesame, soybean, tea, grass, shrub and arbor) were measured in laboratory. Three transformation methods, namely the first derivative (FD), standard normal variate (SNV) and logarithmic reciprocal transformation ($\text{Log}(1/R)$) were used for spectral transformation. The SVR models using three aforementioned kernels were calibrated and validated with 1 000 bootstrap sample datasets. The average determination coefficients (R^2) as well as ratio of performance to standard deviate (RPD) were calculated to compare the performances of three different kernels. The results showed that, the RBF kernel

based SVR model with FD and absorbance transformation obtained the best accuracy for c_N and c_K estimations (c_N : mean $R^2 = 0.64$, mean RPD=1.67; c_K : mean $R^2 = 0.56$, mean RPD=1.48), and the RBF kernel based SVR model with FD transformation obtained the best accuracy for c_P estimations (c_P : mean $R^2 = 0.68$, mean RPD=1.73). The study indicated that RBF kernel based SVR model has great potential in estimating biochemical component contents of diverse plant leaves with VNIR spectroscopy.

Keywords Kernel function; Support vector machine; VNIR spectroscopy; Biochemical content

* Corresponding author

(Received May 23, 2018; accepted Sep. 30, 2018)

悼念 Garry Horlick, (1944—2018)



我们非常沉痛地宣布, 分析光谱学领域的巨人 Garry Horlick 于 2018 年 11 月 1 日去世。Horlick 以他的科学洞察力以及科学出版物和会议报告的清晰性和有效性而闻名于世。他在光谱学和光谱化学分析的各个领域作出了开创性的贡献。Garry Horlick 1944 年 4 月 15 日出生于萨斯喀彻温省雷吉纳, 在艾伯塔省卡尔加里上学。1965 年, 他在艾伯塔大学获得学士学位, 1970 年与霍华德·马尔姆斯塔特合作获得伊利诺伊大学厄本纳-香槟分校的博士学位。1969 年, 他以助理教授的身份加入艾伯塔大学化学系。从 1996 年到 2001 年, 他迅速晋升为艾伯塔大学系主任。

他的研究兴趣是分析光谱学的一般领域, 重点是开发新的元素和分子分析方法和仪器。Garry Horlick 是第一个使用当时新近介绍的傅里叶变换红外光谱(FT-IR)方法的人之一, 他是个研究人员, 自己制造了迈克尔逊干涉仪。这是一项艰巨的任务, 要求他不仅从头开始设计仪器, 甚至研磨和抛光氯化钠光学元件。这个早期的仪器在他早期在阿尔伯特罕改进和应用, 使其能够在紫外和可见光谱区域使用。后来, Garry Horlick 率先将线性传感器(光电二极管阵列)应用于原子光谱化学测量, 这项工作比该领域的商业发展早了 20 年。

在他的实验室中, 基于光电二极管阵列和 FT-IR 光谱学以及声光可调滤波器, 开发了许多新的元素分析系统, 这些系统可以构成全电子光谱仪系统的基础。

通过傅里叶变换, Garry Horlick 自然而然地进入了相关方法、信号处理以及计算机在化学实验室中的应用。他是最早采用微处理器并将其纳入日常实验室测量的人之一。他和他的学生制定了自己的软件频谱存储和文件搜索。

在他后来的职业生涯中, Garry Horlick 处理了 ICP 光谱学中无法解释的仪器现象。在这些研究开始时, 那些研究 ICP 发射光谱的人对基质效应的起源和性质存在相当大的混淆。例如, 当在样品溶液中加入易电离元素(EIE)如钾时, 发现一些信号增强, 而另一些则报告抑郁。Gary 和他的小组表明, 在存在 EIE 的情况下, 明显的差异是由于分析物发射强度的下移和轻微损失引起的。结果, 那些在等离子中看到高发射的人看到了 lossin 信号, 而那些在较低观察位置的人则经历了增强。Horlick 也是电感耦合等离子体质谱(ICP-MS)的早期采用者, 并且发表了大量的关键论文, 阐明了该方法中元素间和基质效应的起源。其中一项研究探讨了 ICP 中心气体流量的影响, 并产生了可能是分析光谱化学中最密集的数据集。这些地块因其形状而被称为“霍利克山脉”。

Garry Horlick 于 1985 年获得应用光谱学会的 Megers 奖, 1987 年获得加拿大化学研究所的 Fisher 科学讲座奖。1989 年, 他获得了应用光谱学会的莱斯特·W·斯特罗克奖, 并于 1990 年当选为加拿大皇家学会会员。1996 年, 他获得美国化学学会分析化学分会的光谱化学分析奖, 2003 年获得匹兹堡光谱学会的匹兹堡光谱学奖, 以表彰他对新方法和系统前向分析的发展和表征的毕生贡献。他于 1990 年和 1992 年获得了《Spectrochimica Acta B: 原子光谱学最佳论文奖》。为了花更多的时间和家人在一起, Garry Horlick 在 2006 年退休了。他对摄影非常感兴趣, 并热情地向不列颠哥伦比亚省维多利亚市的游客展示他收集的本地野生动物照片。

加里是一位绅士和学者。他以诚实和富有加拿大特色的热情发表了科学和评论。他的朋友和同事们会想念他, 但最重要的是, 他的妻子莫琳、他的两个孩子凯伦、劳拉、他的孙子卡兰和泰莎会想念他。

在他的职业生涯中, 曾指导过许多本科生、研究生和博士后研究员。他的实验室对于大多数到北美的科学参观者来说是“必须”停留的地方, 并吸引了广泛的合作者和同事。

Gary Hieftje and Mike Blades

(孟昭红摘自黄本立私人通讯)