

文章编号:1001-9014(2009)04-0272-05

## 基于可见/近红外光谱技术的 黄瓜叶片 SPAD 值检测

刘飞, 王莉, 何勇, 鲍一丹

(浙江大学 生物系统工程与食品科学学院, 浙江 杭州 310029)

**摘要:**为了快速准确检测黄瓜叶片的 SPAD 值,采用可见/近红外光谱技术并结合化学计量学方法建立了黄瓜叶片 SPAD 值校正模型.并用不同建模方法对全波段光谱进行建模,结果表明用最小二乘支持向量机(LSSVM)建模得到的预测效果最好,其相关系数 $r$ 和预测均方根误差 RMSEP 分别为 0.9583 和 0.9732.通过分析黄瓜叶片的光谱反射率与 SPAD 值的相关系数和 PLS 建模回归系数,得到了 531~581nm 和 696~716nm 2 个特征波段以及 556nm、581nm、698nm 和 715nm 4 个特征波长,应用 LSSVM 分别对特征波段和特征波长建模.分析表明,采用特征波段建模,其预测相关系数 $r$ 和预测均方根误差分别为 0.9338 和 1.1370,与全波段建模结果相近,而采用特征波长建模效果稍差.特征波段建模大大减少了建模中的运算量,提高了建模速度,便于相应检测仪器的开发,所以,采用光谱特征波段建模对黄瓜叶片 SPAD 值的检测更为有效.

**关键词:**黄瓜;可见/近红外光谱;最小二乘支持向量机;叶绿素  
中图分类号:S642.2 文献标识码:A

## DETECTION OF SPAD VALUE OF CUCUMBER LEAVES BASED ON VISIBLE/NEAR INFRARED SPECTROSCOPY TECHNIQUE

LIU Fei, WANG Li, HE Yong, BAO Yi-Dan

(College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310029, China)

**Abstract:** For the rapid detection of SPAD value of cucumber leaves, the calibration model of SPAD value was built by using visible and near infrared (Vis/NIR) spectroscopy technique and chemometrics methods. Different calibration methods were used to build the model in the whole wavelength region. The results indicate that the optimal performance is achieved by least squares support vector machine(LSSVM) model, and the correlation coefficient ( $r$ ) and root mean squares error of prediction (RMSEP) are 0.9583 and 0.9732, respectively. Via the analysis of correlation coefficients between the spectral reflectance and SPAD values, and the regression coefficients of partial least squares (PLS), two characteristic wavelength bands (531~581nm and 696~716nm) and four characteristic wavelengths (556, 581, 698 and 715nm) were obtained. LSSVM was used to the aforementioned wavelength bands and wavelengths. The results indicate that the characteristic wavelength bands can achieve a better performance with  $r$  of 0.9338 and RMSEP of 1.1370. The prediction results are similar to the whole wavelength region model, while, the performance of LSSVM model with four characteristic wavelengths was not satisfying. Calibration method of using characteristic wavelength bands can largely reduce the calibration computation, and increase the calibration speed. Hence, it is more effective to use characteristic wavelength bands for the detection of SPAD values of cucumber leaves.

**Key words:** cucumber; Vis/NIR spectroscopy; least squares support vector machine (LSSVM); chlorophyll

### 引言

SPAD(Soil and Plant Analyzer Development)是日本农林水产省农园艺局的“土壤、作物分析仪

器开发”的英文缩写. SPAD 值是一个相对叶绿素含量读数,也称绿色度.在国内外已有试验表明,叶绿素计读数 SPAD 值能预测水稻、棉花及许多园艺作物叶片单位面积的叶绿素含量,可以用 SPAD 值估

收稿日期:2008-07-01,修回日期:2009-01-20

Received date: 2008-07-01, revised date: 2009-01-20

基金项目:国家高技术研究发展计划(863计划)(2007AA10Z210)和国家自然科学基金(30671213)资助项目

作者简介:刘飞(1983-),男,河北定州人,博士生,主要从事光谱与多光谱检测技术研究

通讯作者:鲍一丹(1963-),女,浙江杭州人,副教授,主要从事数字农业及农业装备技术研究,E-mail: ydbao@zju.edu.cn.

计叶片单位重量含氮量,尤其是单位面积含氮量<sup>[1]</sup>.

可见/近红外光谱技术是一种无损的快速检测技术,由于其分析过程中无需使用化学试剂,大大降低了分析成本,一次光谱数据的采集可以同时分析多个指标,加快了信息采集的速度.近年来,随着光谱技术的不断发展,光谱分析法在各个领域都得到了广泛而成功的应用<sup>[2-5]</sup>.

以往对 SPAD 值的获取均采用 SPAD 计,SPAD 计测量时需要手工对叶片进行反复测量,测量速度不高,所以限制了这一指标在植物生长信息分析中的应用.本文的研究目的是建立黄瓜叶片光谱反射率与 SPAD 值的校正模型,这样既可单独预测黄瓜叶片 SPAD 值,也可在分析其它一些指标的同时对叶片的 SPAD 值进行分析,将其作为一个辅助参考指标.

## 1 试验材料与方法

### 1.1 仪器设备

对黄瓜叶片 SPAD 值的测量采用由日本 KONICA MINOLTA 公司生产的 SPAD 502 叶绿素计.实验使用美国 ASD (Analytical Spectral Device) 公司的 Handheld FieldSpec 光谱仪,其光谱采样间隔为 1.5nm,测定范围 325 ~ 1075nm,扫描次数 30 次.分析软件为 ASD View Spec Pro、Unscrambler V9.6、MATLAB7 和 LSSVM 工具包.

### 1.2 试验样本

黄瓜样本采用盆栽种植于智能化人工温室中.首先对黄瓜种子进行育苗,然后将苗移植于塑料盆中,采用 1m<sup>2</sup> 种植空间.将 36 盆植株平均分成 3 组,对其进行梯度施肥.待黄瓜幼苗长至 3~4 片时进行实验.因受人工温室空间、自然光等光源的限制,将每植株叶片悉数采至实验室中,对每植株和每植株叶片都进行编号,每一叶片保存于密闭塑料袋中.然后对每一个叶片样本进行 SPAD 值测量和光谱数据的采集.

### 1.3 化学计量学方法

本研究中用到的化学计量学方法有偏最小二乘法(Partial Least Squares, PLS)、误差反向传播神经网络(Back Propagation Neural Networks, BPNN)和最小二乘支持向量机(Least Squares Support Vector Machine, LSSVM).

支持向量机是一种新型建模方法,它通过结构风险最小化原理来提高泛化能力,较好地提高了学

习机的泛化能力,即有限训练样本得到的决策规则对独立的测试集仍能够得到小的误差.神经网络通过基于梯度迭代的方法进行数据学习,容易陷入局部最小值,支持向量机是通过解决一个二次规划问题,来获得全局最优,而且支持向量机的计算复杂性和质量并不直接依赖于输入数据的维数.因此,SVM 方法可以被用于化学计量学来进行模式识别和函数回归.它比人工神经网络更能控制过拟合,无局部最小点,预测可靠性高且泛化能力更强<sup>[6]</sup>.LSSVM 是 Suykens 提出的一种 SVM 的改进技术,以求解一组线性方程代替经典 SVM 中复杂的二次优化问题,降低了计算的复杂性,并且加快了计算的速度<sup>[7,8]</sup>.

## 2 实验结果与分析

### 2.1 黄瓜叶片样本的光谱曲线

图 1 为黄瓜叶片的光谱曲线( $\lambda$  为波长, $R$  为光谱反射率),由于光谱仪在测量范围的两端有比较大的噪音,所以实验中选用的测量区间为 400 ~ 1000nm.图中的光谱反射率呈现典型的植物光谱特征:在 400 ~ 500nm 与 600 ~ 700nm 之间,蓝光、红光波段的光被叶片中的叶绿素吸收进行光合作用而形成两个比较低的反射区,反射率一般低于 10%;在 550nm 处,即绿色波段部分,叶子的吸收减少,形成一个小的反射峰,反射率大约在 20% 左右,这是植物叶子呈绿色的原因,也是绿色植物所特有的;然后在 700nm 左右,反射率突然上升,形成一个高反射平台,一般可达到 50% ~ 60%.

### 2.2 全波段建模预测

利用全波段 400 ~ 1000nm 建模,建模算法分别采用 BPNN 和 LSSVM,由于建模时样本的输入为有限元,所以首先需要对全波段的光谱数据进

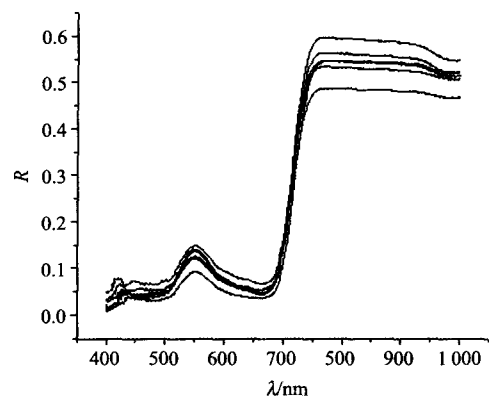


图 1 黄瓜叶片的可见/近红外光谱  
Fig.1 Vis/NIR spectra of cucumber leaves

行特征向量的提取. 采用 PLS 所提取的主因子作为特征向量, 即从原有自变量的样本数据矩阵中提取相互正交的成分, 它们保留了尽量多的与因变量的相关性, 从而在消除原有自变量共线性的同时, 使建立的回归模型仍能充分地反映自变量与因变量之间的相互关系.

在进行 PLS 分析之前, 首先对光谱进行平滑、多元散射校正和变量标准化等预处理. 叶片光谱数据在经过 PLS 处理之后, 可以得到样本残差. 如图 2 所示 (No. 为样本编号,  $R_v$  为样本残差), 残差绝对值  $> 5$  的样本可以视为异常样本, 如图所示 1 号、3 号和 102 号样本为异常样本, 在计算时予以剔除. 原有样本数为 108, 剔除 3 个样本后剩余样本数为 105, 在保证建模集样本数为 75, 预测集样本数为 30 的情况下随机分配建模集和预测集.

采用 PLS 分析方法, 其目的就是获得并选取合适的主因子, 选取主因子可以根据 PLS 模型的残差和累计贡献率两项指标综合考虑. 以全波段建模为例, 选出最优主因子数为 4. PLS 模型参数如表 1 所列, 从表中可以看出当主因子数选为 4 时, 模型的残差最小为 3.480, 而且 4 个主因子的累计贡献率已经达到了 90.095%, 能够反映样本光谱的绝大部分变化. 所以将 4 个主因子作为 BPNN 和 LSSVM 的输入, 以分别建立 SPAD 预测模型.

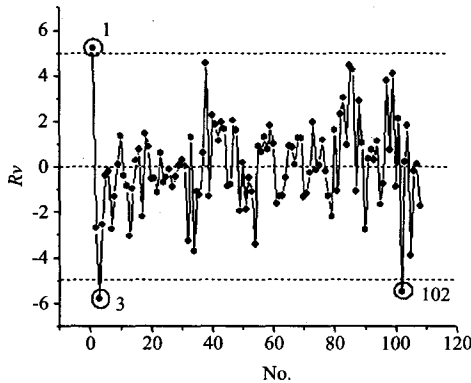


图 2 黄瓜叶片样本残差分析  
Fig.2 Residual variance analysis for cucumber leaves samples

表 1 PLS 模型主要参数  
Table 1 Principal parameters in PLS analysis

主因子数	残差	累计贡献率	主因子数	残差	累计贡献率
1	6.750	68.711%	6	5.178	92.594%
2	4.244	81.877%	7	6.457	93.004%
3	3.548	85.310%	8	7.141	93.437%
4	3.480	90.095%	9	7.689	93.813%
5	4.410	91.561%	10	8.357	94.024%

4 个主因子作为 BP 神经网络的输入建立 3 层 BPNN 模型, 设定隐含层节点数为 3, 最小训练速率为 0.6, 动态参数如 0.6, Sigmoid 参数为 0.9, 允许误差为 0.0001, 最大迭代次数为 1000. 利用 BPNN 获得的黄瓜叶片 SPAD 值预测结果如图 3 所示 ( $C$  为参考值,  $P$  为预测值).

最小二乘支持向量机 LSSVM 也是以 PLS 提取的 4 个主因子作为其输入. 由于 RBF (Radial Basis Function) 核函数作为非线性函数能够减少训练过程中计算的复杂性, 所采用的 RBF 函数作为 LSSVM 的核函数: 超参数  $\gamma$  和 RBF 核函数参数  $\sigma^2$  的优化采用基于交互验证的网格搜索. 利用 LSSVM 获得的黄瓜叶片 SPAD 值预测结果如图 4 所示 ( $C$  为参考值,  $P$  为预测值).

2 个模型的预测结果的指标如表 2 所列, 通过预测均方根误差 RMSEP 和相关系数  $r$  来评判校正模型的优劣, RMSEP 越小相关系数越高, 该模型就越好. 从表中可以看出应用 LSSVM 建模得到的  $r$  和 RMSEP 分别为 0.9583 和 0.9732, 为 2 个模型最优. 说明对黄瓜叶片 SPAD 的预测过程中, LSSVM 要优于传统的线性以及非线性建模方法.

### 2.3 特征波段建模预测

在对作物的光谱分析中, 采用全波段光谱仪以及全波段进行数学建模分析, 不仅花费大, 同时建模所需时间长. 因此, 本实验寻找了黄瓜叶片反射光谱

表 2 不同建模算法的预测结果  
Table 2 Prediction results with different calibration methods

建模算法	校正相关系数	校正均方根误差	预测相关系数	预测均方根误差
BPNN	0.9472	1.1652	0.9457	1.1759
LSSVM	0.9638	0.9719	0.9583	0.9732

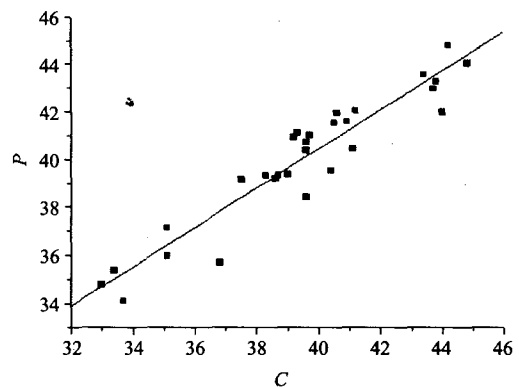


图 3 BP 神经网络对 SPAD 值预测结果  
Fig.3 Prediction results of SPAD values using BPNN model

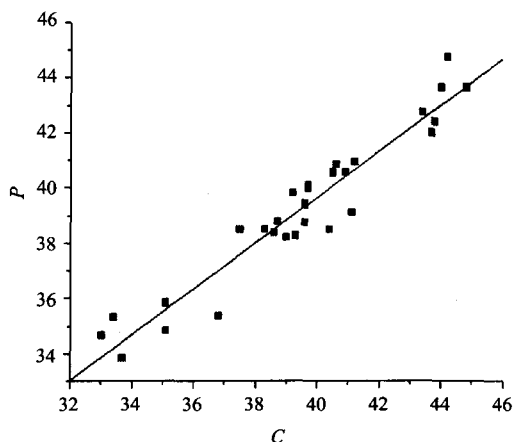


图4 LSSVM 对 SPAD 值预测结果  
Fig.4 Prediction results of SPAD values using LSSVM model

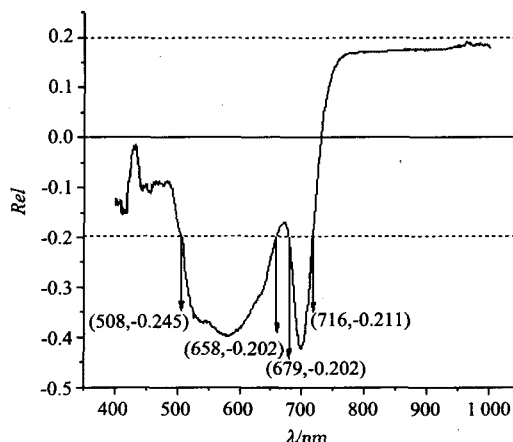


图6 叶片反射率与 SPAD 相关系数图  
Fig.6 Correlation coefficient between reflectance of each wavelength and SPAD values

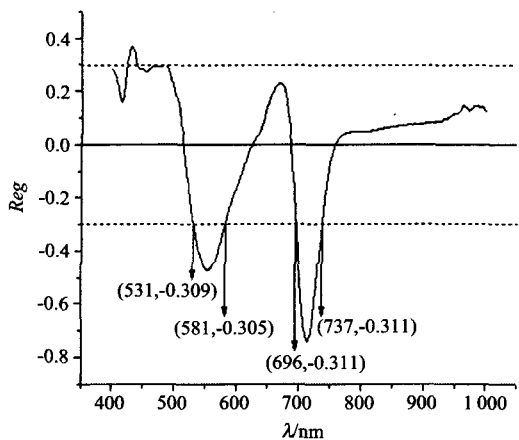


图5 PLS 回归系数图  
Fig.5 Regression coefficient of PLS model

的特征波段,为今后开发简易的仪器提供理论基础.

试验中采用回归系数法和相关系数法相结合,进行特征波段和特征波长的选取.所谓回归系数,它是利用多项式或其它带参数的函数拟合自变量和因变量序列,达到最佳拟合程度时的多项式系数或函数的参数,在多元回归中被用来比较变量间的重要性,可以根据回归系数进行波长选择.此外,多元校正中,还可根据相关系数进行波长的选择.相关系数法是将光谱阵中每个波长对应的反射率向量与待测浓度向量进行相关性计算,得到波长相关系数图,对应相关系数越大其信息相应越多<sup>[9]</sup>.

设定回归系数绝对值 >0.3 的波长为所需的特征波长,如图 5 可见(λ 为波长,Reg 为回归系数),有 3 个光谱区间符合这一设定阈值,结合图 1 黄瓜叶片光谱曲线,得知 430nm 左右的小尖峰与

所研究的叶绿素光谱吸收无关,所以只考虑 531 ~ 581nm 与 696 ~ 737nm 2 个光谱区间.同理,设定相关系数的选择阈值为 0.2,从图 6 可见(λ 为波长,Rel 为回归系数),508 ~ 658nm 与 679 ~ 716nm 光谱区间符合这一阈值.取这几个光谱区间相交的部分为建模时的特征波段,即只用 531 ~ 581nm 和 696 ~ 716nm 范围内的 72 个点进行建模.所选波长也恰在植物叶片所特有的红边和绿峰所对应的波长范围内,说明此实验中波长的选择是合理的.此外,在图 5 和图 6 中回归系数在所选特征波段绝对值取最大时的 2 个波长是 556nm 和 715nm,相关系数在所选特征波段绝对值取最大时的 2 个波长是 581nm 和 698nm.利用这些特征波段和特征波长进行建模预测.建模算法采用 LSSVM.预测结果如表 3 所列.

表 3 列出了分别采用全波段,特征波段和特征波长 3 种建模方式的结果.从表中可以看出采用全波段 400 ~ 1000nm 建模,其相关系数 0.9583 为 3 个模型最高,同时其预测均方根误差 0.9732 为 3 个模型最低.采用 2 个特征波段建模得到预测相关系数

表 3 采用全波段、特征波段和特征波长建模的叶片 SPAD 值预测结果

Table 3 Prediction results of SPAD values with whole wavelength region, characteristic wavelength bands and characteristic wavelengths

波长 (nm)	校正相关系数	校正均方根误差	预测相关系数	预测均方根误差
400 ~ 1000	0.9638	0.9719	0.9583	0.9732
531 ~ 581 696 ~ 716	0.9572	1.1275	0.9338	1.1370
556, 581, 698, 715	0.7692	2.0687	0.7204	2.3036

$r$  为 0.9338, RMSEP 为 1.137, 结果与全波段所建模型的预测结果接近. 说明黄瓜叶片的 SPAD 值与波段 531~581nm 和 696~716nm 有很好的相关性. 采用 4 个特征波长建模得到预测相关系数  $r$  为 0.7204, RMSEP 为 2.3036, 预测结果要明显差于前两种建模方式. 这主要是因为, 在光谱数据采集过程中, 受外界因素的干扰, 光谱曲线会产生一定的漂移, 造成叶绿素含量在一定的波长范围内变化. 综上所述, 全波段建模的预测结果最好, 采用特征波段建模, 其结果稍逊于全波段建模.

### 3 结语

采用便携式光谱仪采集黄瓜叶片样本的可见/近红外光谱, 通过不同的化学计量学方法, 建立光谱反射率同 SPAD 值的关系, 在全波段建模的情况下用 LSSVM 取得了最优结果. 并用回归系数法结合相关系数法对黄瓜叶片光谱进行分析, 针对叶片 SPAD 值获得特征波段和特征波长. 之后, 用 LSSVM 算法对特征波段和特征波长建模. 对黄瓜叶片 SPAD 值的建模结果表明 LSSVM 所建立的全谱预测模型所获得的结果最为精确, 采用特征波段建模的预测结果接近于全波段建模, 而用特征波长得到的预测结果则较差. 其中特征波段建模可以在保证精度的情况下极大减少运算量, 提高了建模速度, 这说明用 LSSVM 和特征波段建模能够对黄瓜叶片的 SPAD 值进行快速准确的测量, 可见/近红外反射光谱法是一种有效的黄瓜叶片 SPAD 值检测方法.

### REFERENCES

- [1] Peng S, Garcia F V, Laza R C, *et al.* Adjustment for specific leaf weight improves chlorophyll meter's estimate of rice leaf nitrogen concentration [J]. *Agronomy Journal*, 1993, **85**(5):987—990.
- [2] LI Xiao-Li, HU Xing-Yue, HE Yong. New approach of discrimination of varieties of juicy peach by near infrared spectra based on PCA and MDA model [J]. *J. Infrared Millim. Waves* (李晓丽, 胡兴越, 何勇. 基于主成分和多类判别分析的可见-近红外光谱水蜜桃品种鉴别新方法. *红外与毫米波学报*), 2006, **25**(6):417—420.
- [3] HE Yong, LI Xiao-Li. Discriminating varieties of waxberry using near infrared spectra [J]. *J. Infrared Millim. Waves* (何勇, 李晓丽. 用近红外光谱鉴别杨梅品种的研究. *红外与毫米波学报*), 2006, **25**(3):192—194.
- [4] Liu F, Zhang F, Jin Z L, *et al.* Determination of acetolactate synthase activity and protein content of oilseed rape (*Brassica napus* L.) leaves using visible/near infrared spectroscopy [J]. *Analytica Chimica Acta*, 2008, **629**(1-2):56—65.
- [5] XU Jing-Ping, ZHANG Bai, SONG Kai-Shan, *et al.* Estimation of Chlorophyll-a concentration in lake Xinmiao based on a semi-analytical model [J]. *J. Infrared Millim. Waves* (徐京萍, 张柏, 宋开山, 等. 基于半分析模型的新庙泡叶绿素 a 浓度反演研究. *红外与毫米波学报*), 2008, **27**(3):197—201.
- [6] WANG Li, HE Yong, LIU Fei, *et al.* Rapid detection of sugar content and pH in beer by using spectroscopy technique combined with support vector machines [J]. *J. Infrared Millim. Waves* (王莉, 何勇, 刘飞, 等. 应用光谱技术和支持向量机分析方法快速检测啤酒糖度和 pH 值. *红外与毫米波学报*), 2008, **27**(1):51—55.
- [7] Suykens J A K, Vanderwalle J. Least squares support vector machine classifiers [J]. *Neural Processing Letters*, 1999, **9**(3):293—300.
- [8] Tang H S, Xue S T, Chen R, *et al.* Online weighted LSVM for hysteretic structural system identification [J]. *Eng. Struct.*, 2006, **28**(12):1728—1735.
- [9] CHU Xiao-Li, YUAN Hong-Fu, LU Wan-Zhen. Progress and application of spectral data pretreatment and wavelength selection methods in NIR analytical technique [J]. *Progress in Chemistry* (褚小立, 袁洪福, 陆婉珍. 近红外分析中光谱预处理及波长选择方法进展与应用. *化学进展*), 2004, **16**(4):528—542.