

文章编号:1004-4213(2011)03-0428-5

基于无信息变量和偏最小二乘投影分析的 高光谱散射图像最优波段选择

王爽,黄敏,朱启兵

(江南大学 轻工过程先进控制教育部重点实验室,江苏 无锡 214122)

摘 要:提出了一种无信息变量消除和偏最小二乘投影分析相结合的苹果高光谱散射图像最优波段选择方法.经该算法提取后的波段降为全谱的 26%,将选择后的波段作为输入变量建立了苹果硬度的偏最小二乘预测模型.预测均方根误差由 6.00N 降为 5.73N,相关系数也有所提高,并与遗传算法作了比较.结果表明,该算法能有效消除原光谱矩阵中冗余的信息,且不存在遗传算法中的参量选择随机性等缺点.该算法为高光谱散射图像最优波段选择提供了一个理想的方法.

关键词:无信息变量和偏最小二乘投影算法;高光谱图像技术;波段选择;无损检测

中图分类号:O657.3

文献标识码:A

doi:10.3788/gzxb20114003.0428

0 引言

随着社会的发展,人们对苹果品质的要求越来越高,在选择苹果时对糖度、硬度及内部营养物质等指标也很关注.果实的软硬度是决定其贮藏和加工方式的一项重要指标,用光谱法对水果内部品质检测是一种有效的方法^[1].高光谱图像技术结合了图像技术和光谱技术,既能对待测物体的外观特性检测又能对内部品质进行检测,高光谱的这一特性为水果品质的无损检测提供了可能.近年来国内外有关学者已开始利用该技术展开了对水果内部品质的研究^[2-7].高光谱信息中存在着重叠严重的问题,本文的研究重点就是在存有大量冗余波长信息的光谱中提取有用波长来提高预测模型的性能.

波长选择的方法主要有相关系数法^[8]、回归系数法^[9]、遗传算法^[10]等算法,这些算法各有优缺.无信息变量消除算法(Uninformative Variables Elimination, UVE)^[11],是基于偏最小二乘回归系数建立的波长选择算法,用于消除不提供信息的变量.但有时 UVE 得到的变量依然较多,可从 UVE 得到的变量中进一步选择有效变量.偏最小二乘(Partial Least Squares, PLS)投影分析算法^[12]是一种新兴的波段选择方法,与传统的相关系数比较,其能很好地提高模型的稳健性,并且能很有效地降低建模所需的波长数.

本文利用高光谱图像技术对苹果硬度进行无损检测研究,并用 UVE 结合 PLS 投影分析算法提取的波长作为 PLS 模型的输入变量,建立苹果硬度预测模型.通过与遗传算法等做比较,表明该方法建立的模型不仅简化了,而且预测能力也有一定的提高,为实现对水果的高光谱图像在线检测提供了技术支持.

1 实验材料与方法

1.1 苹果样品

实验样本为‘Golden Delicious’苹果,取自于密歇根州立大学(MSU)Clarksville 园艺实验室和密歇根州的 MSU 园艺教学研究中心.在实验前,苹果储存于一个受控的大气环境(2% O₂ 和 3% CO₂, 0 °C)中大约 5 个月的时间.这些苹果从仓库中取出并且在室温(24 °C)下保存至少 15 h,再进行相关实验.

1.2 样本的高光谱图像采集

高光谱图像数据是用基于光谱仪的高光谱图像系统采集而得到.它是由图像光谱仪(ImSpectorV10, Spectral Imaging Ltd, Oulu, Finland)、高光谱摄像头(Model C4880-21-24A, Hamamatsu Photonics Systems, Bridgewater, NJ, USA),250W 的光纤卤素灯和一套输送装置等部件组成^[13].高光谱摄像头的光谱范围为 400~1 000 nm,光谱分辨率为 4.54 nm,实验数据记录时,取近似值

基金项目:国家自然科学基金(No. 60805014)和中央高校基本科研业务费专项奖金(No. JUSRP20913、No. JUSRP21132)资助

第一作者:王爽(1985-),男,硕士研究生,主要研究方向为高光谱图像技术. Email: wangsg1985@126.com

导师(通讯作者):黄敏(1974-),女,副教授,主要研究方向为农产品无损检测. Email: huangmzqb@163.com

收稿日期:2010-11-01; **修回日期:**2010-11-27

5 nm,空间分辨率 0.20 mm/pixel. 本实验研究共测试了 600 个‘Golden Delicious’苹果,将这 600 个苹果编号并依次放置于高光谱图像采集系统的载物平台上,然后在每个苹果赤道 9 mm 的线性扫描区获得高光谱图像,相机的曝光时间为 0.25 s,在曝光时间内共获得 10 幅图像,然后将这 10 幅图像求平均,对这些平均后的散射图像两边再求平均,然后截取以入射光点为中心、物理尺寸为 20 mm×20 mm 的区域图像.再用平均值法进行特征提取,样本的光谱范围取为 500~1 000 nm,每隔 5 nm 取一个光谱波段值,则每个样本含 101 个光谱波段信息,即每个样本共得到 101 个特征描述.由于每个苹果样本表面曲率的不同会造成光谱波段值测量不准确,需对每个样本的散射图像进行果形修正,具体修正方法可见文献[13].

1.3 样本的硬度测量

从每个苹果中获取高光谱图像之后,再进行标准破坏性试验,测量苹果的硬度.该测试使用一台带有一个直径 11 mm 的 Magness Taylor(MT)钢探针的质地分析仪(model TA. XT2i, Stable Micro Systems, Inc., Surrey, U. K),MT 探针以 2 mm/s 的速度伸入去皮的果实 9 mm.记录下力度/位移曲线上的最大力度值,作为苹果的硬度测量值[4].

2 数据分析

2.1 样本集划分

对经果形校正、平均、平滑校正等处理的 600 个样本集划分. Kennard-Stone 算法[14],是常用的样本集划分方法,它是利用变量空间之间的相对欧氏距离找出样本集中的有代表性的各个边缘样品,直至找出设定的样品数.利用该算法选取 450 个样本作为校正集,剩余 150 个样本作为预测集.苹果样本的硬度测量值统计如表 1.

表 1 苹果硬度测量值统计

Table 1 Statistics of firmness measurements for apples

	Mean/N	Standard deviation/N	Min/N	Max/N
All (n=600)	56.65	12.07	31.22	88.50
Calibration (n=450)	55.55	12.54	31.22	88.50
Prediction (n=150)	59.95	9.85	37.64	77.03

2.2 无信息变量消除算法原理

在 PLS 模型中,光谱矩阵 X 和浓度矩阵 Y 存在关系 $Y = Xb + e$, b 是系数向量, e 是误差向量. UVE 就是把相同于自变量数的随机变量矩阵加入

光谱矩阵中,然后通过交叉验证建立 PLS 模型,得到系数矩阵,分析系数向量 b 的平均值和标准偏差的商 C 的可靠性,其中, $\text{mean}(b)$ 是系数向量 b 的平均值, $s(b)$ 是系数向量 b 的标准偏差, i 是光谱矩阵中的第 i 列向量.

$$C_i = \text{mean}(b_i) / s(b_i)$$

依据 C 的绝对值大小确定是否把第 i 列变量用到最后的 PLS 模型中.具体的算法可见参考文献[11].

2.3 PLS 投影分析算法原理

PLS 投影分析算法是基于 PLS 投影系数的波段选择方法,它的特点是计算了某一波长点光谱数据的变化量与光谱 PLS 回归系数向量在该波长点的投影分量的共同影响.该算法提高了模型的稳定性且较大幅度减少了建模时所需的波长数.具体的算法可见参考文献[12].

3 实验结果与讨论

3.1 样本的高光谱曲线特征

图 1 所示为其中 10 个苹果样本的相对反射光谱图,从图中可看到光谱在 500~730 nm 波动较大,由于叶绿素吸收在 675 nm 处有一波谷,在 730~920 nm 区域内相对反射值比较平稳,在约 970 nm 处由于水分的吸收出现另一波谷.

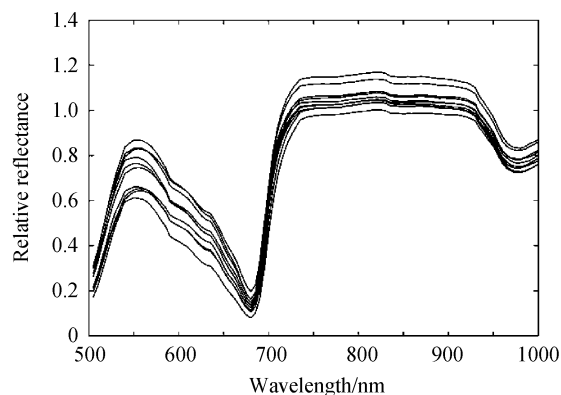


图 1 10 个苹果样本的相对反射光谱

Fig. 1 Spectra of relative reflectance for 10 apples

3.2 UVE 选取有效波长

将处理后的全部波长作为 PLS 模型的输入建立模型.通过比较模型校正相关系数 R_c 、校正均方差 (Root Mean Square Error for Calibration, RMSEC)、交叉验证相关系数 R_{cv} 、交叉验证均方差 (Root Mean Square Error for Cross Validation, RMSECV) 及主因子数 (Factors) 等指标考核校正模型的实用性,然后根据预测相关系数 R_p 、预测均方差 (Root Mean Square Error for Prediction, RMSEP) 指标来确定预测模型的性能优劣.其中主因子数是经过内部交叉验证所得,评价标准是预测

相关系数越高、预测均方差越小,模型的性能就越理想.全波段的模型性能见表2所示,从该表中可以看到由于全波段中部分变量含有无用的或者是不相关的信息,使得预测模型冗余、复杂,预测模型不很理想,有必要对全波段光谱波长进行选择.

用 UVE 算法进行波段初步筛选. UVE 算法选择波长如图2.左侧为101个变量的稳定性C分布图,右侧为 UVE 随机产生的101个变量的稳定性C分布图.其中的两条虚线为变量筛选的阈值上下限虚线以外的变量保留,虚线以内的变量则需全部去掉.经 UVE 变量筛选后得到34个波长,这34个波长的分布大致为:530~605 nm 的黄光、610~640 nm 的橙光、645~770 nm 的红光以及 780~1 000 nm 的近红外范围内.可见有用光谱信息的分布还是比

较广泛的.将得到的这34个波长作为 PLS 的输入变量,得到的预测模型结果见表2,模型的预测相关系数和预测均方根误差都得到了一定的优化.

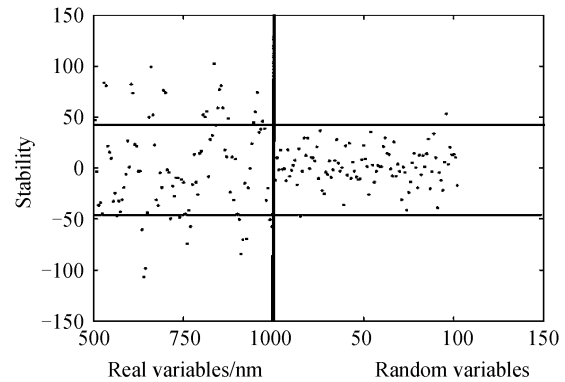


图2 UVE 算法选择波长

Fig. 2 Selecting wavelengths using UVE

表2 不同波段选择方法的 PLS 建模比较

Table 2 The results of PLS modeling after different selection methods

Method	Variables	Factors	R_c	RMSEC/N	R_{cv}	RMSECV/N	R_p	RMSEP/N
All	101	10	0.873	6.05	0.852	6.46	0.791	6.00
UVE	34	9	0.874	6.08	0.858	6.43	0.805	5.85
PLS	44	9	0.884	5.87	0.861	6.37	0.804	5.88
UVE-PLS	27	8	0.870	6.19	0.858	6.44	0.814	5.73
GA	29	9	0.879	5.97	0.864	6.30	0.803	5.87
UVE-GA	25	9	0.868	6.23	0.856	6.47	0.814	5.78

3.3 PLS 投影算法选取有效波长

采用 PLS 投影分析算法进行波段选择.用 PLS 投影分析算法得到的 PLS 投影系数如图3.从 PLS 投影系数图上能够很有效地找出对待测硬度影响较大的波长.从中可以看出这些波长主要分布在橙光、红光和近红外部分. PLS 投影分析法的阈值取为系数最大值的17%,这样就从101个波长中选择了44个作为 PLS 模型的输入,得到的预测模型结果见表2.相比于全波段,经 PLS 投影分析选择波长后的预测模型性能也有所提高.

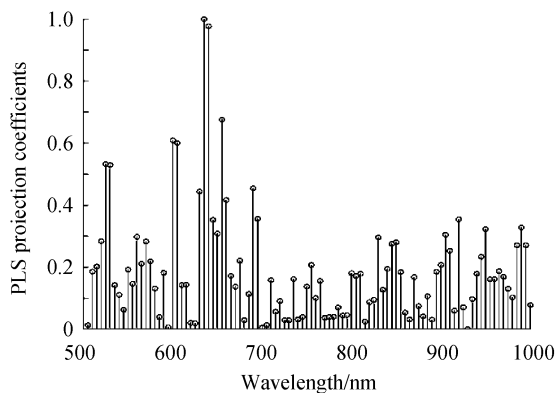


图3 PLS 投影系数

Fig. 3 PLS projection coefficients

3.4 UVE-PLS 投影分析选取有效波长

UVE-PLS 投影分析算法进行波长的再次筛

选.经过 UVE 和 PLS 投影分析后的变量个数是有一定的减少,但仍较多,预测模型的性能还不够理想.在 UVE 算法的基础上再经 PLS 投影分析对波长进一步选择,可得共线性最小的有效波长. UVE-PLS 投影分析的阈值取其系数最大值的20%.从中选出27个波长,这27个波长为(单位:nm):525、530、600、605、630、635、640、650、655、660、695、710、735、755、765、800、830、840、845、850、855、900、905、910、920、945、950,从中可以看出这些波长主要集中在红光和近红外部分.所选的27个波长的 UVE-PLS 投影系数如图4,把这些优选出的波长作为 PLS 模型的输入变量,得到的预测模型结果见表2.

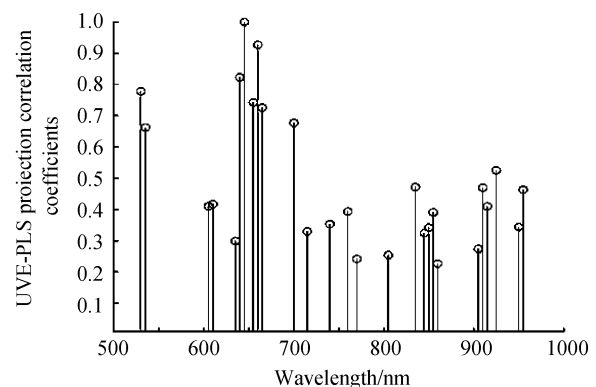


图4 UVE-PLS 投影分析选择的27个波长

Fig. 4 27 wavelengths selected by UVE-PLS projection

可得所建模型的各参量都得到较好的优化,说明用 UVE-PLS 投影分析提取的波长可以代表原始光谱大部分的有效信息。

3.5 与遗传算法波段选择比较

遗传算法(Genetic Algorithm, GA)是传统的应用较为广泛的一种波长选择方法,它借鉴生物界自然选择和遗传机制,利用选择、交换和突变等算子的操作,随着不断的遗传迭代,使目标函数值较优的变量被保留,较差的变量被淘汰,最终达到最优结果。为说明 UVE-PLS 投影算法的优越性,本文将该算法与 GA 和 UVE-GA 算法选择出的波长作比较,遗传算法的参量设定为:种群大小 30,最大繁殖代数 50,交叉概率 0.6,变异概率 0.01。由于遗传算法是一种非导数的随机优化算法,其中的选择、交叉、

变异都带有很强的随机性,且 PLS 因子数的不同选择会使每一次选择的波段不完全相同。本文将 UVE-GA 算法运行 10 次与选择 10 个不同阈值的 UVE-PLS 算法作对比,表 3 列出了经这两种算法后所建立的最优的 5 组 PLS 预测模型的性能参量。从表 3 中可以看出 UVE-PLS 投影分析算法选择波长有略好的优选能力,并且经这两种算法选择的相同波长个数为 20 个,相同率是较高的,但 UVE-PLS 投影分析算法克服了遗传算法在参量选择上存在随机性等弊端。经两种算法波段选择后的预测值和实测值相关图见图 5(a)和(b),比较两个图也可看出经 UVE-PLS 投影分析算法所建的预测模型是等同于 UVE-GA 算法的。

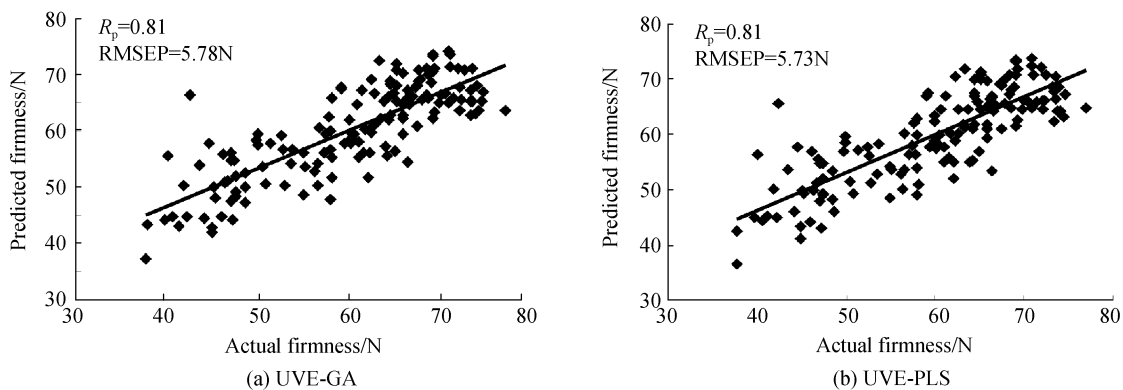


图 5 UVE-GA 及 UVE-PLS 算法预测结果图

Fig. 5 Forecast map of UVE-GA and UVE-PLS

表 3 UVE-PLS 和 UVE-GA 比较

Table 3 Comparison of UVE-PLS and UVE-GA

Group	UVE-PLS				UVE-GA			
	R_c	RMSEC/N	R_p	RMSEP/N	R_c	RMSEC/N	R_p	RMSEP/N
1	0.870	6.19	0.814	5.73	0.878	6.23	0.814	5.78
2	0.871	6.19	0.815	5.74	0.874	6.24	0.812	5.78
3	0.875	6.20	0.803	5.74	0.864	6.26	0.803	5.79
4	0.864	6.20	0.806	5.76	0.865	6.28	0.795	5.81
5	0.865	6.21	0.798	5.78	0.842	6.30	0.777	5.82

4 结论

应用高光谱图像技术结合 UVE-PLS 投影分析算法选取有效波长实现了对苹果硬度的快速无损检测。与传统遗传算法相比,UVE-PLS 投影分析算法不存在相关参量选择上的随机性,也不存在遗传算法的早熟现象,且运行时间较短。可见 UVE-PLS 投影分析算法不失为高光谱散射图像最优波段选择的一种很好的方法。经 UVE-PLS 波长选择后的预测模型,相关系数提高到了 0.814,预测均方根误差由 6.00N 降低到 5.73N,所选变量约减为全部变量的 26%,所建立的预测模型是比较理想的。本文所研究的这一波段选择算法快速、简便,为水果的无损检测

提供了一种新的可行的技术手段。

致谢:对美国农业部 Postharvest Engineering Laboratory 的 Dr. Lu 在实验工作中的指导深表感谢。

参考文献

- [1] LIU Mu-hua, ZHAO Jie-wen, ZHENG Jian-hong, *et al.* Review of hyperspectral imaging in quality and safety inspections of agricultural and poultry products [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2005, **36**(9): 139-143.
刘木华,赵杰文,郑建鸿. 农畜产品品质无损检测中高光谱图像技术的应用进展[J]. *农业机械学报*, 2005, **36**(9): 139-143.
- [2] YANY Ke-ming, CHEN Yun-hao, Guo Da-zhi, *et al.* Information detection and extraction of spectral wheat rust based on stripe hyperspectral image[J]. *Acta Photonica Sinic*, 2008, **37**(1): 145-151.

- 杨可明, 陈云浩, 郭达志, 等. 基于高光谱影像的小麦条锈病光谱信息探测与提取[J]. 光子学报, 2008, **37**(1): 145-151.
- [3] GUO En-you, LIU Mu-hua, ZHAO Jie-wen, *et al.* Nondestructive detection of sugar content on navel orange with hyperspectral imaging[J]. *Transactons of the Chince Society for Agricultural Machinery*, 2008, **39**(5): 91-93.
郭恩有, 刘木华, 赵杰文, 等. 脐橙糖度的高光谱图像无损检测技术[J]. 农业机械学报, 2008, **39**(5): 91-93.
- [4] LU Ren-fu. Nondestructive measurement of firmness and soluble solids content for apple fruit using hyperspectral scattering images[J]. *Sens& Instrumen Food Quality*, 2007, **12**(1): 19-27.
- [5] JUAN Xing, PAL T JANCSOK, *et al.* Detecting bruises on 'Golden Delicious' apples using hyperspectral with multiple wavebands[J]. *Biosystems Engineering*, 2005, **90**(1)2: 7-36.
- [6] HONG Tian-sheng, QIAO Jun, WANG Ning, *et al.* Non-destructive inspection of chinese pear quality based on hyperspectral imaging technique [J]. *Transactions of the CSAE*, 2007, **23**(2): 151-155.
洪添胜, 乔军, Wang Ning, 等. 基于高光谱图像技术的雪花梨品质无损检测[J]. 农业工程学报, 2007, **23**(2): 151-155.
- [7] CHEN Y R, CHAO K, KIM M S. Machine vision technology for agricultural applications[J]. *Computers and Electronics in Agriculture*, 2002, **36**(2): 173-191.
- [8] MIN M, LEE. Determination of significant wavelengths and prediction of nitrogen content for citrus[J]. *Transactions of the ASAE*, 2005, **48**(2): 455-461.
- [9] WU D, HE Y, FENG S. Shortwave near-infrared spectroscopy analysis of major compounds in milk powder and wavelength assignment[J]. *Analytica Chimica Acta*, 2008, **610**(2): 232-242.
- [10] KOUMONSIS V K, KATSARAS C P. A saw-tooth genetic algorithm combining the effects of variable and reinitialization to enhance performance [J]. *IEEE Transactions on Evolutionary Computation*, 2006, **10**(1): 19-28.
- [11] CENTNER V, MASSART D, DENOORS O E, *et al.* Elimination of uninformative variables for multivariate calibration[J]. *Analytical Chemistry*, 1996, **68**(21): 3851-3858.
- [12] DAN Tu-nan, DAI Lian-kui. Spectral wavelength selection based on PLS projection analysis [J]. *Spectroscopy and Spectral Analysis*, 2009, **29**(2): 351-354.
淡图南, 戴连奎. 基于 PLS 投影分析的光谱波段选择方法[J]. 光谱学与光谱分析, 2009, **29**(2): 351-354.
- [13] QIN J, LU R. Measurement of the optical properties of fruits and vegetables using spatially resolved hyperspectral diffuse reflectance imaging technique[J]. *Postharvest Biology and Technoiogy*, 2008, **49**(3): 355-365.
- [14] MACHO S, IUSA R, CALLAO M P, *et al.* Monitoring ethylene content in heterophasic copolymers by near-infrared spectroscopy standardisation of the calibration model [J]. *Analytica Chimica Acta*, 2001, **445**(2): 213-220.

Optimal Wavelength Selection of Hyperspectral Scattering Images Based on UVE-PLS Projection Analysis

WANG Shuang, HUANG Min, ZHU Qi-bing

(Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), Jiangnan University, Wuxi, Jiangsu 214122, China)

Abstract: Partial least squares projection analysis combined with uninformative variable elimination was used to select optimal wavelengths from apple hyperspectral scattering images. After this algorithm, the number of effective wavelengths decreased to 26%. The selected effective wavelengths were set as inputs of partial least squares model. Root mean square error of prediction dropped from 6.00N to 5.73N and correlation coefficient increased a little. The result shows that to select effective wavelengths using the algorithm is feasible. In the parameter selection, there is not such as random defects. It is expected that the algorithm would provide an effective method for optimal wavelengths selection using hyperspectral scattering image technique.

Key words: UVE-PLS projection; Hyperspectral imaging; Wavelength selection; Nondestructive detection