

文章编号: 1672-8785(2012)10-0043-06

基于光谱技术的葡萄活体可溶性固体含量 在线检测研究

吕 刚 杨海清

(浙江工业大学信息工程学院, 浙江 杭州 310032)

摘 要: 快速检测活体水果内部品质对于确定水果最佳采摘时机和果园信息化管理具有重要意义。以南方棚栽葡萄为研究对象, 应用光谱技术对处于生长期的四个葡萄品种的可溶性固体含量 (SSC) 进行现场测试。分别采用偏最小二乘法 (PLS) 回归、潜变量人工神经网络 (LV-ANN) 和潜变量支持向量机 (LV-SVM) 三种方法为光谱建模集建立了 SSC 校正模型。用验证集对模型的预测性能进行了评价。与 PLS 和 LV-ANN 模型相比, LV-SVM 模型的预测性能最佳。实验结果表明, 将光谱技术与 LV-SVM 建模法相结合适用于果园葡萄活体可溶性固体含量无损检测。

关键词: 葡萄; 可溶性固体含量; 在线检测; 光谱分析

中图分类号: S123 **文献标识码:** A **DOI:** 10.3969/j.issn.1672-8785.2012.10.010

On-site Determination of Grape Soluble Solid Content (SSC) Based on Spectroscopic Technology

LV Gang, YANG Hai-qing

(College of Information Engineering, Zhejiang University of Technology, Hangzhou 310032, China)

Abstract: The fast detection of inner quality of living fruit is of importance to the selection of optimal harvest time and to the information management of an orchard. The trellised grapes in the southern part of our country are used as the research object. The soluble solid content (SSC) of four kinds of grapes in growth is detected by using a visible and near infrared spectrophotometer on site. The SSC correction models are established by using Partial Least Square regression (PLS), Latent Variable and Artificial Neural Network (LV-ANN) and Latent Variable and Support Vector Machine (LV-SVM) respectively. The prediction performance of these models is evaluated by using a validation set. Compared with the PLS and LV-ANN models, the LV-SVM model has the best prediction performance. The experimental result shows that the combination of spectroscopy with the LV-SVM modeling is suitable for the nondestructive SSC detection of living grapes in an orchard.

Key words: grape; soluble solid content (SSC); in-field determination; spectroscopic technology

0 引言

葡萄是世界上的主要水果之一。随着人们消费水平的不断提高, 市场上对高品质葡萄的

旺盛需求在客观上要求葡萄种植园能够提供葡萄品质的精准化检测手段。传统的实验室测试方法过程繁琐, 无法实现大量在线无损测试。因此需要一种快速、无损的现场检测手段^[1]。

收稿日期: 2012-08-06

作者简介: 吕刚 (1988-), 男, 浙江永康人, 硕士研究生, 主要研究方向为模式识别技术。

E-mail: gang_lv@hotmail.com

可见-近红外光谱技术用于物质检测和物化性质分析时具有无损、快速、廉价、高效等诸多优点,目前已被广泛应用于食品和农业领域,比如土豆片品质控制^[2]、橄榄油特性测定^[3]、果汁糖分分析^[4]和牛肉成分预测^[5]等。

可溶性固体含量 (Soluble Solid Content, SSC) 是水果内部品质的主要指标之一。国内外对水果可溶性固体含量的光谱测定已有不少的报道^[6-8]。例如, Clark 等人^[6]应用可见-近红外光谱技术对贮藏的猕猴桃的可溶性固体含量进行了建模及预测。McGlone 等人^[7]对储藏前后的皇家嘎啦苹果的可溶性固体含量进行了光谱建模分析。Li 等人^[8]对脐橙的可溶性固体含量进行了光谱测定。以上文献均以采摘后储藏的水果为实验对象,而对生长期葡萄活体的可溶性固体含量进行现场无损检测还未见报道。

本文以浙江省四个棚栽葡萄品种为研究对象,应用可见-近红外光谱技术结合化学计量学方法,建立葡萄活体可溶性固体含量预测模型。研究重点在于通过比较不同预测模型对品种的泛化能力,建立能适用于所有葡萄品种的通用预测模型。

1 材料与方法

1.1 光谱仪与样本测试

光谱测试使用自制手持式光谱仪。该光谱仪内装有美国海洋公司生产的 USB4000 型光谱

模块,其谱段范围为 345.61 ~ 1047.80 nm,间隔为 0.16 nm。光源为 20 W 卤素灯。光线由光纤束传输到探头,探头与样本表面呈直角放置,距离约为 1 cm。光谱数据经 USB 接口与笔记本电脑连接,采用自编的 Visual Basic 程序采集光谱。光谱仪的参数如下:积分时间为 20 ms,扫描次数为 20 次,平滑度为 8。测试设备由 12 V 蓄电池逆变供电。

供试的四个葡萄品种分别为比昂克 (Rosario Bianco)、红地球 (Red Globe)、巨玫瑰 (Muscat Kyoho) 和藤稔 (Fujiminori),由浙江省绍兴市某葡萄种植园提供。在葡萄生长期进行现场光谱测试,时间为 2011 年 7 月 3 日到 8 月 14 日,共 6 周,每周现场测试 1 次。光谱测试后,将测试样本剪下,现场压榨取汁。用美国 HANNA 公司生产的 HI-96801 型便携式数字式折射仪对葡萄汁的可溶性固体含量进行测试记录(单位为 °Brix)。实验共获得 2485 个样本,随机选取 70 % 作为校正集,其余 30 % 作为验证集。表 1 列出了样本统计情况。

1.2 光谱预处理

由于原始光谱两端存在大量噪声,实验选取 475 ~ 930 nm 波段用于数据分析。用 $\log(1/R)$ 将反射光谱转换为吸收光谱,再通过 Savitzky-Golay 平滑算法(2 阶,左右 3 点平滑)对其进行降噪处理。

表 1 四个葡萄品质样本的可溶性固体含量 (°Brix) 统计情况

品种	样本数	值域	校正集		样本数	值域	验证集	
			平均值	标准差			平均值	标准差
比昂克	438	3.4 ~ 17.1	8.33	3.78	188	3.4 ~ 17.1	8.43	3.95
红地球	467	3.9 ~ 16.9	10.12	3.67	202	4.2 ~ 16.7	9.29	3.49
巨玫瑰	432	4.8 ~ 18.3	11.13	3.62	184	4.9 ~ 18.0	10.89	3.69
藤稔	406	2.7 ~ 16.2	8.86	3.71	168	3.5 ~ 16.7	8.02	3.25

1.3 偏最小二乘回归建模法

偏最小二乘回归 (Partial Least Squares Regression, PLS) 是目前化学计量学中最常用的分析方法之一^[9],主要用于探寻因变量矩阵与自

变量矩阵之间的关系。PLS 法能够有效消除光谱波长间的高度共线性。通过对高维光谱数据进行线性变换,用少数潜变量 (Latent Variables, LVs) 代替原谱数据,实现数据压缩。本文选用非

线性迭代偏最小二乘算法 (NIPALS) [10], 设最大潜变量数为 15, 交叉校验段数为 20。根据交叉校验均方根误差 (RMSECV) 最小化原则, 确定 PLS 模型最优潜变量数为 7。

1.4 人工神经网络建模法

人工神经网络 (Artificial Neural Network, ANN) 是一种模仿生物神经网络结构和功能的计算模型。人工神经网络在处理一些复杂问题时常被当作一种优良的建模工具来使用 [11]。本文将 PLS 建模法获得的 7 个潜变量用作人工神经网络的输入数据, 构建潜变量人工神经网络 (LV-ANN) 模型。ANN 训练方法为 Levenberg-Marquardt 法 [12], 隐层神经元数目为 14 个, 最高迭代次数与精度分别为 10000 次和 0.001。

1.5 支持向量机建模法

支持向量机 (Support Vector Machines, SVM) 是一种建立在统计学习理论和结构风险最小化原理基础上的机器学习方法。该方法将向量映射到一个更高维的空间里, 建立方向合适的分隔超平面, 使两个与之平行的超平面之间的距离达到最大化。支持向量机在解决小样本、非线性及高维模式识别问题时具有显著优势 [13]。本文将 PLS 建模法获得的 7 个潜变量用作支持向量机的输入数据, 构建潜变量支持向量机 (LV-SVM) 模型。SVM 类型为 epsilon-SVR, 径向基核函数的两个学习因子 c 和 g 由自适应函数优化确定 [14]。

1.6 模型评价指标

为了对模型性能进行评价, 我们对建模集使用交叉校验决定系数 (R^2) 和均方根误差 (RMSECV); 对验证集使用预测决定系数 (R^2) 和均方根误差 (RMSEP)。计算公式如下:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (2)$$

式中, n 为样本个数, $\hat{y}_i$ 与 y_i 分别为第 i 个样本的模型计算值与参考值, $\bar{y}$ 为参考平均值。

为了测试预测模型对葡萄品种的泛化能力, 先建立单品种预测模型, 再对所有光谱数据建立混合模型, 并比较单品种模型和混合模型的预测性能。数据处理使用 Unscrambler X10.1 和 Matlab v7.5 软件进行。

2 结果与讨论

2.1 样本光谱特征

样本光谱中包含很多的有用信息。图 1 所示为四个葡萄品种的平均吸收光谱曲线。总体上, 不同品种的吸收光谱基本相似, 在 680 nm 处都有一个与叶绿素吸收有关的波峰。在 500 ~ 550 nm 波段, 比昂克品种的光谱曲线呈波谷状, 这与该品种的果皮颜色有关。比昂克品种在成熟时呈绿黄色, 其他葡萄品种均会泛红。

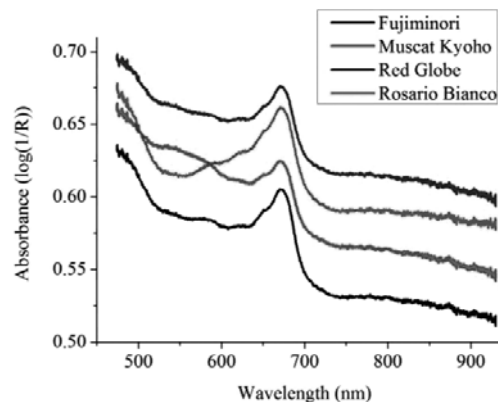


图 1 四个葡萄品种的平均吸收光谱

2.2 PLS 模型

在单品种 PLS 模型中, 红地球品种的交叉校验均方根误差最小 (RMSECV=1.05 °Brix)。其他品种模型的交叉校验性能优劣排序为比昂克 (1.24 °Brix)、藤稔 (1.32 °Brix)、巨玫瑰 (1.53 °Brix)。混合模型的校验误差最大 (1.71 °Brix)。

图 2 所示为单品种和混合模型的 PLS 回归系数曲线。PLS 回归系数表示波长对预测值的贡献大小。回归系数的绝对值越大, 表示该波长对模型的影响越大。可以看出, 在 520 ~ 550 nm 处存在一个明显的波峰, 这可能与样本中的花

青素有关^[15]；介于 650 ~ 700 nm 之间有一处较大的波谷，这可能与葡萄生长期叶绿素 a 含量的变化有关^[16]。

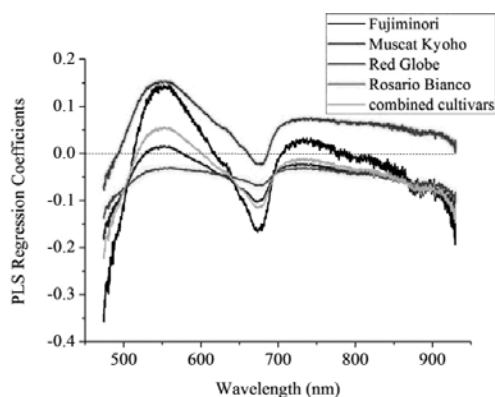


图 2 单品种模型与混合模型的 PLS 回归系数曲线

2.3 LV-ANN 模型

在单品种 LV-ANN 模型中，红地球品种的交叉校验误差最小 (RMSECV=0.60 °Brix)。其他品种模型的交叉校验性能优劣排序为比昂克 (0.81 °Brix)、藤稔 (0.99 °Brix)、巨玫瑰 (1.19 °Brix)。混合模型的校验效果较好 (RMSECV=1.12 °Brix)。

2.4 LV-SVM 模型

在单品种 LV-SVM 模型中，红地球品种的交叉校验误差最小 (RMSECV=0.68 °Brix)。其他品种模型的交叉校验性能优劣排序为比昂克 (0.83 °Brix)、藤稔 (0.89 °Brix)、巨玫瑰 (1.17 °Brix)。混合模型的校验效果较好 (RMSECV=1.07 °Brix)。

2.5 模型预测性能的比较

表 2 列出了 PLS、LV-ANN 和 LV-SVM 单品种模型和相应混合模型的验证集测试结果。可以看出，三种建模法的预测性能都与品种有关，优劣排序为红地球 ($R^2=0.90 \sim 0.92$)、比昂克 ($R^2=0.87 \sim 0.88$)、藤稔 ($R^2=0.82 \sim 0.86$)、巨玫瑰 ($R^2=0.80 \sim 0.82$)。在混合模型中，LV-SVM 模型最好，LV-ANN 模型次之，PLS 模型最差。图 3 所示为混合模型的验证集预测值与测量值之间的拟合关系。单品种 LV-SVM 模型的预测决定系数 (R^2) 分别为 0.87、0.92、0.82 和 0.86，预测均方根误差 (RMSEP) 分别为 1.24、0.91、1.35 和 1.14 °Brix。混合 LV-SVM 模型的预测结果如下： $R^2=0.86$ ，RMSEP=1.36 °Brix。与 PLS 和 LV-ANN 模型相比，LV-SVM 模型在各品种之间以及单品种和混合模型之间的预测性能差别最小，显示出较强的品种泛化能力。

3 总结

本文应用可见-近红外光谱技术对四个葡萄品种的可溶性固体含量进行了现场活体测试，分别建立了 PLS、LV-ANN 和 LV-SVM 三种预测模型。无论是单品种模型还是混合模型，LV-SVM 模型的预测性能都高于 PLS 和 LV-ANN 模型。LV-SVM 模型在各品种之间以及多品种混合光谱上的预测性能差别均最小，品种泛化能力强。实验结果表明，将可见-近红外光谱技术

表 2 不同模型对验证集的预测性能

数据集	样本数	PLS 模型			LV-ANN 模型			LV-SVM 模型		
		LVs ^a	R^2	RMSEP (°Brix)	LVs	R^2	RMSEP (°Brix)	LVs	R^2	RMSEP (°Brix)
比昂克	188	2	0.87	1.43	7	0.88	1.25	7	0.87	1.34
红地球	202	2	0.90	1.14	7	0.91	1.04	7	0.92	0.91
巨玫瑰	184	3	0.80	1.67	7	0.80	1.40	7	0.82	1.35
藤稔	168	5	0.82	1.35	7	0.85	1.20	7	0.86	1.36
混合模型	742	6	0.75	1.89	7	0.84	1.49	7	0.86	1.36

^aLVs: PLS 潜变量个数

与 LV-SVM 建模法相结合, 不仅预测精度高, 而且品种泛化能力强, 可适用于果园葡萄活体的

可溶性固体含量无损检测。

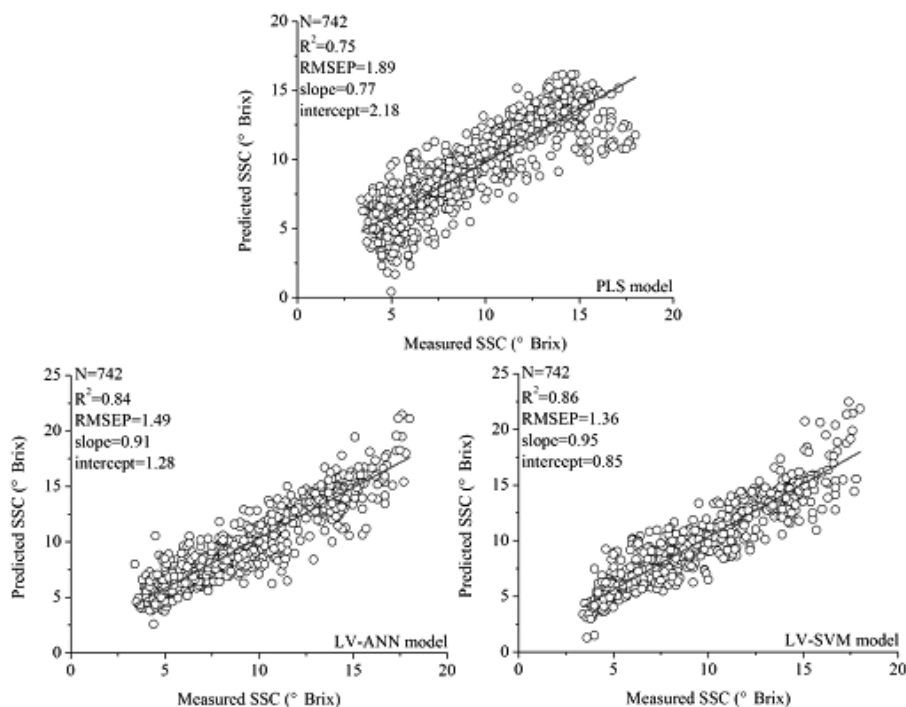


图 3 PLS 模型、LV-ANN 模型和 LV-SVM 模型在验证集上的预测性能

参考文献

- [1] Kemps B, Leon L, Best S, et al. Assessment of the Quality Parameters in Grapes Using VIS/NIR Spectroscopy [J]. *Biosystems Engineering*, 2010, **105**: 507–513.
- [2] Shiroma C, Rodriguez S L. Application of NIR and MIR Spectroscopy in Quality Control of Potato Chips [J]. *Journal of Food Composition and Analysis*, 2009, **22**: 596–605.
- [3] Morales S A, Fernandez C V, Casanova L, et al. Feasibility of NIR Spectroscopy for Non-destructive Characterization of Table Olive Traits [J]. *Journal of Food Engineering*, 2011, **107**: 99–106.
- [4] Rodriguez L E, Fry F S, McLaughlin M A, et al. Rapid Analysis of Sugars in Fruit Juices by FT-NIR Spectroscopy [J]. *Carbohydrate Research*, 2001, **336**: 63–74.
- [5] Tøgersen G, Arnesen J F, Nilsen B N, et al. On-line Prediction of Chemical Composition of Semi-frozen Ground Beef by Non-invasive NIR Spectroscopy [J]. *Meat Science*, 2003, **63**: 515–523.
- [6] Clark C J, Mcglone V A, Silva H N, et al. Prediction of Storage Disorders of Kiwifruit (*Actinidia Chinensis*) Based on Visible-NIR Spectral Characteristics at Harvest [J]. *Postharvest Biology and Technology*, 2004, **32**: 147–158.
- [7] Mcglone V A, Jordan R B, Martinsen P J. Vis/NIR Estimation at Harvest of Pre- and Post-storage Quality Indices for ‘Royal Gala’ Apple [J]. *Postharvest Biology and Technology*, 2002, **25**: 135–144.
- [8] Li J, Xue L, He X W, et al. Visible and Near Infrared Reflectance Spectroscopy for Determining Soluble Solids Content of Navel Orange [C]. *Electric Information and Control Engineering*, 2011: 3780–3786.
- [9] 王凤花, 朱海龙, 戈振扬. 近红外光谱数据建模方法的研究进展 [J]. *农业工程*, 2011, **1**(1): 56–61.
- [10] 肖林, 何大卫. PLS 回归方法及其医学应用 [J]. *中国卫生统计*, 2002, **19**(2): 76–79.
- [11] Poursina M, Tehrani M S, Poursina D. Application of BPANN and Regression for Prediction of Bowing Defect in Roll-forming of Symmetric Channel Section [J]. *International Journal of Material Forming*, 2008, **1**: 17–20.
- [12] 赵弘, 周瑞祥, 林廷圻. 基于 Levenberg-Marquardt 算法的神经网络监督控制 [J]. *西安交通大学学报*, 2002, **36**(5): 523–527.
- [13] 丁世飞, 齐丙娟, 谭红艳. 支持向量机理论与算法研究综述 [J]. *电子科技大学学报*, 2011, **40**(1): 2–8.

- [14] 董国军, 哈力木拉提·买买提, 余辉. 基于 RBF 核的 SVM 核参数优化算法 [J]. *新疆大学学报*, 2009, **26**(3): 355–363.
- [15] Yang H, Kuang B, Mouazen A M. In Situ Determination of Growing Stages and Harvest Time of Tomato (*Lycopersicon Esculentum*) Fruits Using Fiber-Optic Visible-Near-Infrared (Vis-NIR) Spectroscopy [J]. *Applied Spectroscopy*, 2011, **65**(8): 49–56.
- [16] Bertone E, Venturello A, Leari R, et al. Prediction of the Optimum Harvest Time of 'Scarlet' Apples Using DR-UV-Vis and NIR Spectroscopy [J]. *Postharvest Biology and Technology*, 2012, **69**: 15–23.

(上接第 42 页)

染物测量、有毒气体自动识别以及军用化学战剂探测等领域, 受到了人们极大的认同。如果将被动红外遥测技术引入到我国公安禁毒工作的制毒窝点远距离监测任务中, 一定会为战斗在禁毒第一线的公安干警提供巨大的便利, 从而提高禁毒工作的成效。

参考文献

- [1] 郑为建, 胡旭, 苏君红, 等. 远程化学毒气红外探测告警技术进展 [J]. *红外与激光工程*, 2006, **35**(5): 612–617.
- [2] 王建宇, 舒嵘, 刘银年, 等. *成像光谱技术导论* [M]. 北京: 科学出版社, 2011.
- [3] 翁诗甫. *傅立叶变换红外光谱分析* [M]. 北京: 化学工业出版社, 2010.
- [4] NIST Chemistry WebBook [EB/OL]. <http://webbook.nist.gov/chemistry>, 2011.
- [5] 张记龙, 聂宏斌, 王志斌, 等. 化学战剂红外光谱遥测技术现状及发展趋势 [J]. *中北大学学报*, 2008, **29**(3): 265–271.
- [6] Ravreby M. Quantitative Determination of Cocaine and Heroin by Fourier Transform Infrared Spectrophotometry [J]. *J Forensic Sci*, 1987, **32**(1): 20–37.
- [7] Andreas B, Rainer D, Tim J. Detection of Chemical Agents in the Atmosphere by Passive IR Remote Sensing [C]. *SPIE*, 1999, **3856**: 87–93.

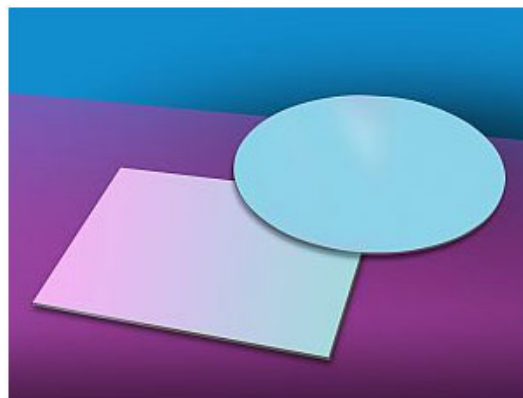
新闻动态 News

美国 DSI 公司推出甚长波红外滤光片

据 www.laserfocusworld.com 网站报道, 位于加利福尼亚州的美国沉积科学公司 (DSI) 最近推出一系列新的甚长波红外滤光片。这些滤光片在 $12 \sim 22 \mu\text{m}$ 波长区具有高的透过率。由于采用该公司的高精度物理气相沉积 (PVD) 工艺加工, 这些坚固的膜层已轻易通过所有环境测试。它们可以在环境温度和低温之间循环使用, 不会发生性能降低现象。

DSI 公司制作的这些甚长波红外薄膜可以应用于各种衬底, 包括 Ge、ZnSe、Si 以及 InSb。它们可以根据需要制成窄带滤光片、宽带滤光

片、长波通滤光片、短波通滤光片以及增透膜等。这些新型甚长波滤光片非常适合在遥感、化学分析、天体物理 / 天文学以及地平仪等敏感器上使用。



□ 高国龙