

葡萄浆果糖度可见/近红外光谱检测的研究

吴桂芳^{1,2}, 黄凌霄^{3*}, 何勇^{1*}

1. 浙江大学生物工程与食品科学学院, 浙江 杭州 310029
2. 内蒙古农业大学机电工程学院, 内蒙古 呼和浩特 010018
3. 浙江大学动物科学学院, 浙江 杭州 310029

摘要 针对可见/近红外光谱与水果糖度存在非线性相关的特点,利用漫反射光谱测定方法获取了葡萄浆果的可见/近红外光谱,提出了应用偏最小二乘(PLS)结合人工神经网络(ANN)建立葡萄浆果糖度的预测模型,利用偏最小二乘法(PLS)对原始光谱数据进行处理,得出交叉检验的最佳主因子数为3,并将3个主因子的得分作为三层BP神经网络的输入。通过定标集样本对BP神经网络进行训练,用优化的BP神经网络模型对预测集样本进行预测。PLS-ANN模型对样本的预测模型检验参数 r^2 为0.908, RMSEP为0.112, Bias为0.013,好于只使用PLS模型的预测模型检验参数 r^2 为0.863, RMSEP为0.171, Bias为0.024。结果表明,利用近红外光谱技术无损检测葡萄浆果糖度等内部品质是可行的,为今后进一步分析建立浆果内部品质预测模型奠定了基础。

关键词 可见/近红外光谱;葡萄;浆果;糖度;偏最小二乘;人工神经网络

中图分类号: O433.4 **文献标识码**: A **DOI**: 10.3964/j.issn.1000-0593(2008)09-2090-04

引言

葡萄是世界性的水果,用途很广,除主要用来酿造不同类型的葡萄酒外,还大量用以鲜食,加工成葡萄干、葡萄汁等。葡萄浆果风味优美,是人们最喜爱的水果之一,入世以来,中国葡萄产业首先面对的是“洋葡萄”大量进口的冲击与挑战,与世界发达国家相比,中国葡萄的优质化、标准化生产以及市场运作还处于初级阶段,葡萄浆果的产后处理、品质鉴别检测一直是农产品加工研究的重要课题。葡萄浆果的含糖量是葡萄品质评定的重要指标,尤其在作为酿酒原料时,由于要适应葡萄酒的种类及其酿造工艺,对葡萄浆果的含糖量要严格控制。传统的葡萄浆果含糖量的检测方法是从小穗摘取1~3粒葡萄,取一定数量的浆果以获得250 mL左右的葡萄汁再进行含糖量的测定,这样不仅耗时费力,对葡萄造成损伤,而且由于取样分析时间长,难免造成样品变质而导致较大的人为误差产生。采用可见/近红外光谱技术可实现水果和果汁内部品质的快速无损检测^[1]。

可见/近红外光谱法是一种快速简便非破坏性的分析检测方法。可充分利用全谱或多波长下的光谱数据进行定性或定量分析。由于近红外光谱技术分析具有速度快、效率高、成本低等特点,已经被越来越多地应用于农产品品质分析、烟草行业、石油化工工业分析、药品检测等领域^[2-8]。

应用近红外光谱技术的关键在于如何从光谱数据中获得有用的信息。偏最小二乘法(PLS)能够建立光谱数据和成分之间的相互关系,是一种广泛使用的近红外光谱数据处理方法。然而PLS是用于建立线性模型的,当出现非线性问题时模型的预测结果会受到影响。因此需要在PLS模型的基础上引入非线性的部分,人工神经网络是目前常用的非线性模型。目前应用较广的是一种基于误差反向传播算法的BP神经网络。由于具有高度非线性映射的能力,现已在各领域得到了广泛的应用^[9-12]。

本研究的目的是应用可见/近红外光谱技术测量葡萄浆果糖度,并建立偏最小二乘与BP神经网络相结合的数据分析模型对葡萄浆果糖度进行预测。

收稿日期: 2007-03-16, 修订日期: 2007-06-18

基金项目: 国家科技支撑项目(2006BAD10A04, 2006BAD10A09), 国家自然科学基金项目(30671213)和宁波市重大科技攻关项目(2007C10034)资助

作者简介: 吴桂芳,女,1972年生,浙江大学生物系统工程与食品科学学院讲师

*通讯联系人 e-mail: yhe@zju.edu.cn; lxhuang@zju.edu.cn

1 材料与方法

1.1 仪器设备

实验使用美国 ASD (analytical spectral device) 公司的 Handheld FieldSpec 光谱仪, 其测定范围在 325 ~ 1 075 nm 之间, 采样间隔为 1.5 nm, 探头视场角为 20°, 光谱仪垂直放置于距离样本表面 52 mm 处, 光源与水平位置成 60° 照射。光谱反射率数据先转化成吸收光谱数据, 再转化成 ASCII 码导出。分析软件为 ASD View Spec Pro, Unscramble V9.6 和 DPS V8.01。

葡萄浆果的糖度由 WA Y-2S 数字阿贝折射仪测得, 仪器本身具有温度校正功能, 测量值的单位为 Brix, 测量准确度以折射率表示, $n_D = \pm 0.0002$ 。

1.2 样品来源及光谱的获取

试验用样本均为从水果批发市场购买的京亚葡萄, 选取 60 个葡萄果穗, 并从每个果穗上的穗节部, 穗中部和穗尖部各选一棵葡萄果粒共 180 个样本。全部样本随机分成建模集和预测集, 建模集有 120 个样本, 预测集有 60 个样本。光谱仪经白板校准后进行测试, 对每一个样本扫描 30 次。

1.3 光谱数据的预处理

为了去除来自高频随机噪声、基线漂移、样本不均匀、光散射等影响, 需要进行光谱预处理来消除噪声。采用 Savitzky-Golay 平滑法, 平滑点数为 21, 然后采用多元散射校正 (MSC) 对光谱数据进行处理。所有的预处理过程在 Unscrambler V9.6 软件中进行。为了得到高的信噪比, 去除部分噪声信息, 选择 400 ~ 1 000 nm 波段的数据进行分析处理。

2 试验结果与分析

2.1 葡萄浆果样本的近红外漫反射光谱和葡萄浆果的糖度值

实验所得近红外光谱曲线如图 1 所示。图 1 中横坐标为波长, 纵坐标为光谱吸光度。选择波长范围在 400 ~ 1 000 nm 的光谱, 应用 ASD View Spec Pro 软件, 并转换成 ASCII 码, 形成反射率矩阵, 用偏最小二乘法建立光谱吸光度与糖度的相关关系。

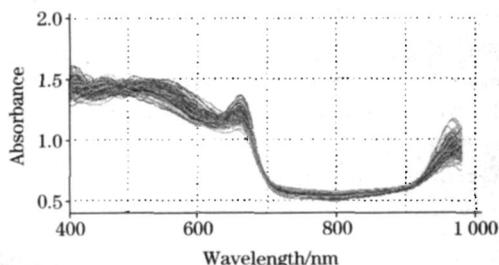


Fig 1 Near infrared reflectance spectra of grape berries

将光谱扫描后的葡萄浆果压成汁, 迅速测量其糖度值, 测试温度为 25 °C, 时间间隔为 $T = 5$ s。葡萄浆果糖度的测

量值如表 1 所示。

Table 1 Statistic data of sugar content in grape berry samples

Samples indexes	Calibration set	Prediction set
	Sugar content (Brix)	
No.	120	60
Mean	12.96	12.99
Range	10 ~ 15.5	10.7 ~ 15.4

2.2 光谱数据与糖度偏最小二乘法的分析

偏最小二乘法本质上是一种基于特征变量的回归方法, 在多元校正中得到了广泛的应用。偏最小二乘法在对量测矩阵进行分解的同时, 对响应矩阵也进行正交分解, 并在这两个矩阵间建立定量关系, 这就是对多个量测响应的预测成为可能。在我们这里, 量测矩阵为光谱矩阵, 量测响应矩阵为糖度矩阵。用偏最小二乘法对光谱矩阵与糖度矩阵进行回归分析, 图 2 为采用偏最小二乘法建立的回归分析模型对 60 个未知样本的预测效果。

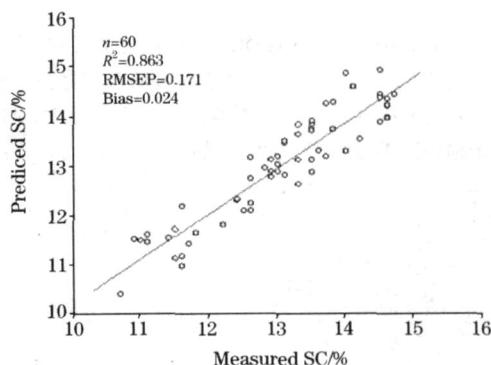


Fig 2 Prediction results for the unknown 60 samples from the PLS models

2.3 偏最小二乘和人工神经网络结合模型

为了提高数据分析模型的预测精度, 我们提出了一种将偏最小二乘和人工神经网络算法相结合, 建立葡萄浆果的可见/近红外光谱和葡萄浆果糖度之间非线性相关模型的方法。由偏最小二乘法对光谱数据交叉检验后得到前 5 个主因子累计贡献率如表 2 所示, 前 3 个主因子的累计贡献率已经达到了 96.4%, 我们将每个样本的光谱数据用前 3 个主成分代替。

Table 2 The accumulative reliabilities of the first 5 principal components

Principal components	Accumulative reliabilities/ %
PC1	88.615
PC2	96.195
PC3	96.430
PC4	96.269
PC5	97.698

将偏最小二乘法分析所得的 3 个最佳主因子作为三层 BP 神经网络的输入, 通过反复训练得到三层最佳网络结构:

3 个输入层节点, 4 个隐含层节点和 1 输出层节点, 目标误差设定为 0.000 1, 限定最大迭代次数为 1 000。采用葡萄浆果 120 个样本用来建模, 利用已建立的偏最小二乘法结合神经网络 (PLS-ANN) 模型对 60 个未知样本进行预测。图 3 为预测结果, 拟合误差为 1.98×10^{-5} 。

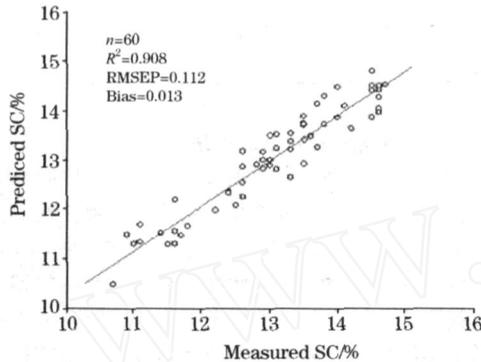


Fig 3 Prediction results for the unknown 60 samples from the PLS-ANN models

2.4 偏最小二乘和神经网络结合模型的评价

对所建立的偏最小二乘和神经网络结合的数据分析模型进行评价, 选取决定系数 r^2 , 预测标准差 $RMSEP$, 偏差 $Bias$ 的值来检验其测量的精确性及误差, 计算公式如下:

$$r^2 = \left[1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - y_m)^2} \right] \times 100\% \quad (1)$$

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

$$Bias = \frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i) \quad (3)$$

式中: \hat{y}_i 为样本的预测值, y_i 为样本的实测值, y_m 为样本真值的平均值, N 为检验的样本数。

图 3 显示了 PLS-ANN 模型样本糖度实测值和预测值之间的 $r^2 = 0.908$, $RMSEP = 0.112$, $Bias = 0.013$; 对应的 PLS 预测值的 $r^2 = 0.863$, $RMSEP = 0.171$, $Bias = 0.024$ 。结果表明 PLS-ANN 模型的预测能力要好于 PLS 模型。

3 结 论

利用 PLS 方法对训练集样本 120 个葡萄浆果进行训练, 建立了葡萄浆果的 PLS 糖度模型, 并得出交叉检验的最佳主因子维数为 3, 将 3 个主因子作为三层 BP 神经网络的输入, 进一步建立了 PLS-ANN 模型, 对验证集样本进行预测。其实测值和预测值之间的决定系数 r^2 达到 0.908, $RMSEP$ 为 0.112, $Bias$ 为 0.013。可以看出应用可见/近红外光谱技术, PLS-ANN 的数据处理模型能够很好的预测葡萄浆果的糖度, 结果要好于 PLS 模型的预测效果。葡萄浆果近红外光谱检测的实验研究表明, 利用近红外光谱技术无损检测浆果糖度等内部品质是可行的, 本文提出的基于 PLS 与 ANN 相结合的光谱数据处理方法提高了定量分析精度, 为今后进一步分析建立草莓、蓝莓、猕猴桃等浆果内部品质预测模型奠定了基础。

参 考 文 献

- [1] HE Pur-chao(贺普超). Study of Grape(葡萄学). Beijing: China Agriculture Press(北京: 中国农业出版社), 1999. 4.
- [2] HE Yong, LI Xiao-li, SHAO Yong-ni(何 勇, 李晓丽, 邵咏妮). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2006, 26(5): 850.
- [3] Ann P, Nico S, Kathleen T, et al. Biosystems Engineering, 2002, 81(3): 305.
- [4] WU Gui-fang, ZHU Deng-sheng, HE Yong(吴桂芳, 朱登胜, 何 勇). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2008, 28(6): 1260.
- [5] Liu Fei, He Yong. Food Research International, 2008, 41(5): 562.
- [6] CHEN Da, WANG Fang, SHAO Xue-guang, et al(陈 达, 王 芳, 邵学广, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2004, 24(6): 672.
- [7] Cen Haiyan, He Yong. Trends in Food Science & Technology, 2007, 18(2): 72.
- [8] He Yong, Li Xiaoli, Deng Xunfei. Journal of Food Engineering, 2007, 79(4): 1238.
- [9] SHAO Yong-ni, HE Yong(邵咏妮, 何 勇). Journal of Infrared and Millimeter Waves(红外与毫米波学报), 2006, 25(6): 478.
- [10] QI Xiao-ming, ZHANG Lu-da, DU Xiao-lin, et al(齐小明, 张陆达, 杜晓林, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2003, 23(5): 870.
- [11] Cen Haiyan, He Yong, Huang Min. Journal of Agricultural and Food Chemistry. 2006, 54(20): 7437.
- [12] Cen Haiyan, Bao Yidan, Huang Min, et al. Applied Optics, 2006, 45(29): 7679.

Research on the Sugar Content Measurement of Grape and Berries by Using Vis/ NIR Spectroscopy Technique

WU Gui-fang^{1,2}, HUANG Ling-xia^{3*}, HE Yong^{1*}

1. College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310029, China

2. College of Mechanical and Electrical Engineering, Inner Mongolia Agricultural University, hohhot 010018, China

3. College of Animal Sciences, Zhejiang University, Hangzhou 310029, China

Abstract Aiming at the nonlinear correlation characteristic of Vis/ NIR spectra and the corresponding sugar content of grape and berries, the Vis/ NIR spectra of grape and berries were obtained by diffusion reflectance. A mixed algorithm was presented to predict sugar content of grape and berries. The original spectral data were processed using partial least squares (PLS), and three best principal factors were selected based on the reliabilities. The scores of these 3 principal factors would be taken as the input of the three-layer back-propagation artificial neural network (BP-ANN). Trained with the samples in calibration collection, the BP-ANN predicted the samples in prediction collection. The values of decision coefficient (r^2), the root mean squared error of prediction (RMSEP), and bias were used to estimate the mixed model. The observed results using PLS-ANN ($r^2 = 0.908$, RMSEP = 0.112 and Bias = 0.013) were better than those obtained by PLS ($r^2 = 0.863$, RMSEP = 0.171, Bias = 0.024). The result indicated that the detection of internal quality of grape and berries such as sugar content by nondestructive determination method was very feasible and laid a solid foundation for setting up the sugar content forecasting model for grape and berries.

Keywords Vis/ NIR Spectroscopy; Grape; Berry; Sugar content; PLS; PLS-ANN

(Received Mar. 16, 2007; accepted Jun. 18, 2007)

*Corresponding author