

基于光谱技术的芒果糖度酸度无损检测方法研究

虞佳佳, 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 提出了一种用近红外光谱技术结合遗传算法和人工神经网络模型的芒果糖度酸度快速无损检测的新方法。首先用偏最小二乘法计算芒果糖度酸度光谱数据的主成分得分值, 以此获取芒果的近红外指纹图谱, 再结合遗传算法优化人工神经网络技术(GA-BP)进行检测。PLS分析表明, 主因子选取18时对糖度具有较好的聚类作用, 而主因子数17个时对酸度的聚类效果好。选取最佳主因子作为芒果糖度酸度的神经网络的输入, 建立三层GA-BP人工神经网络模型。用135个芒果样本的糖度酸度用来建立遗传算法优化神经网络的芒果糖度酸度检测模型, 对未知的45个芒果样本进行糖度酸度的预测。结果表明, 提出的遗传算法和人工神经网络模型相结合的光谱分析方法具有很好的预测能力, 为芒果糖度酸度检测方法提供了一种新方法。

关键词 可见/近红外光谱; 芒果; 偏最小二乘法; 遗传算法; 人工神经网络

中图分类号: S123; TH744.1 **文献标识码**: A **DOI**: 10.3964/j.issn.1000-0593(2008)12-2839-04

引言

芒果为受欢迎的热带水果, 不仅口感好, 而且营养丰富^[1]。芒果的有机酸含量很高, 特别是谷氨酸的含量高, 有利于机体的新陈代谢, 对口渴咽干、消化不良、晕眩呕吐、咳嗽痰多等病状有一定的食疗帮助。随着人们生活水平的日益提高, 对芒果分级和精深加工的需求也越来越高, 因此研究一种简单、快速、非破坏性的芒果糖度酸度检测技术很有必要。

现代近红外光谱分析技术, 可充分利用全谱段或多波长下的光谱数据进行定性或定量分析^[2]。由于近红外光谱分析技术具有速度快、效率高、成本低、测试重现性好、测量方便等特点, NIRS能够通过漫反射光纤方便地进行光谱的采集, 从而在远距离进行控制, 整个操作过程简单^[3], 已经被越来越多地应用于食品、石油化工、制药等领域。国内外很多学者利用近红外光谱技术检测了芒果苹果成熟度^[4, 5]、鉴别杨梅品种^[6]、评价苹果质量^[7]、咖啡品牌^[8]、橙汁成分^[9]、茶饮料^[10]等。

BP神经网络模型是一个强有力的学习系统^[11], 能够实现输入与输出之间的高度非线性映射。目前使用最多的是多层结构的误差反向传播学习算法(BP)^[12], 已经证明此种模

型可以逼近任何连续的非线性曲线^[13]。但是神经网络训练过程中容易出现局部最优的状况^[14], 并且存在网络拓扑结构难以确定等问题。我们引入了自适应遗传算法^[15]优化拓扑网络的权阈值以防止训练陷入局部最优, 从而提高了BP网络学习效果和增强其预测能力。

应用偏最小二乘法的主成分分析, 可以在不丢失主要光谱信息的前提下选择为数较少的新变量来代替原来较多的变量, 解决了由于谱带的重叠而无法分析的困难。我们提出了应用偏最小二乘法(PLS)和基于误差反向传播算法的多层前馈神经网络相结合的方法, 并通过遗传算法优化神经网络权阈值, 建立了芒果的近红外光谱糖度酸度预测模型。

1 材料与方法

1.1 仪器设备

用美国ASD(analytical spectral device)公司的Handheld Field Spec光谱仪, 其光谱采样间隔(波段宽)1.5 nm, 测定范围325~1 075 nm, 扫描次数30次, 探头视场角为20°。光源是与光谱仪配套的14.5 V卤素灯。光谱数据以ASCLL码形式导出并进行处理。分析软件为ASD View Spec Pro, Unscramble V9.6和Matlab 2006b。

1.2 样品来源及光谱的获取

收稿日期: 2007-05-10, 修订日期: 2008-08-20

基金项目: 国家科技支撑项目(2006BAD10A0403), 国家自然科学基金项目(30671213)和高等学校优秀青年教师教学科研奖励计划项目(02411), 浙江省自然科学基金项目(Y307158)和浙江省教育厅项目(20071064)资助

作者简介: 虞佳佳, 女, 1983年生, 浙江大学生物系统工程与食品科学学院硕士研究生 * 通讯联系人 e-mail: dbao@zju.edu.cn

从超市买来 250 个新鲜的芒果样品。选择大小均匀的几个,避免试验中芒果个体与近红外光谱仪之间的距离剧烈变化。全部样本随机分成建模集和预测集,建模集有 135 个样本,预测集有 45 个样本。对每一个芒果扫描 30 次,从芒果的赤道部位等距正反面测定光谱。

1.3 光谱数据预处理

由于光谱数据在采集时首端产生部分噪声,取 400~1 075 nm 波段的光谱数据进行分析。对光谱数据进行平滑处理、MSC 预处理、SNV 预处理^[16],并进行偏最小二乘法分析。

1.4 遗传算法

遗传算法(GA)是一种模拟自然选择和遗传机制的自适应概率全局寻优算法^[17],可优化拓扑阈值,鲁棒性强。其过程主要包括,编码方式、选择、变异、交叉和解码。

1.5 神经网络

建立了一个三层的人工神经网络结构,其中输入采用 Unscrambler V 9.6 中 PLS 结果的最佳主成分的成分得分值,再将数据导入 Matlab 中,进行建模。

2 结果与分析

2.1 芒果样本的近红外漫反射光谱

芒果的典型近红外光谱曲线如图 1 所示,横坐标为波长,纵坐标为光谱漫反射率。从图 1 中可以看出,不同芒果的光谱形状具有相似性,且其吸收峰位置的差异性也不是很大,可见光部分吸收量相对近红外部分要大,特别是 400~550 nm 的吸收量较大,而 550~770 nm 吸收量小,这与芒果表皮成黄色和胡萝卜素等色素的吸收有关;在 950~1 075 nm 之间吸光度值又有一个小凸起,这是 O—H 和 C—H 吸收带^[18],可能是个体内部物质对光的吸收。应用 ASD View Spec Pro 软件,把同一个芒果正反不同部位的光谱曲线做平均处理,并转换成 ASC II 码,形成反射率矩阵,用 PLS 法对其聚类。

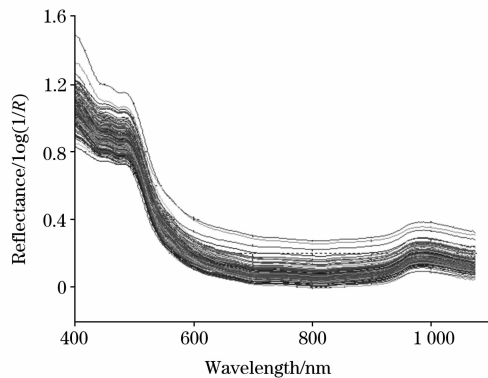


Fig. 1 Near infrared reflectance spectra of mango

2.2 建立芒果糖度酸度预测模型

全波段 325~1 075 nm 共有 750 个点。采用全光谱计算时,计算量大,而且有些区域样品的光谱信息很弱,与样品的组成或性质间缺乏相关关系^[19]。因此,我们通过偏最小二

乘法分析,选取最佳主因子个数的新敏感变量作为输入建立神经网络预测模型。

考虑到信息波段中有可能有非线性信息的存在,采用 BP 神经网络对其建模优化。通过权阈值来调整数据的贡献力,使网络的实际输出值与期望值的误差均方差达到理想值。建模时采用隐含层神经元($2 \times$ 主因子个数+1)个,训练函数采用 Traincglf,误差函数采用 mse,权阈值训练函数采用遗传算法结果,系统误差要求小于 0.000 005,训练次数最高 2 000 次,训练速度 0.2。

从以上分析可知,神经网络中权阈值的设置相当的重要,而遗传算法又可以解决神经网络过程中出现的局部最优的问题,因此本研究采用遗传算法优化神经网络,改善其性能^[20](见图 2)。研究过程中采用变异率 0.9,群落代为 100,目标函数为 $1/E^2$ (即误差平方倒数)。

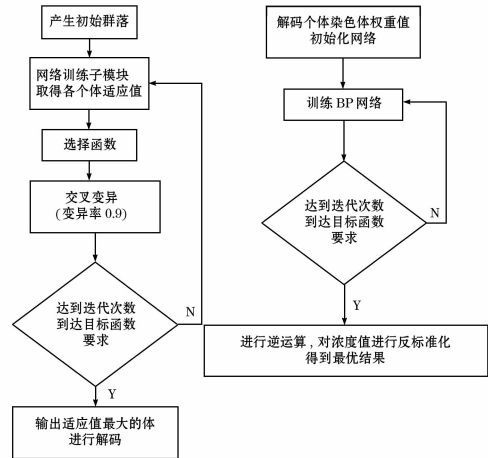


Fig. 2 Flowchart of GA-BP

2.3 神经网络和遗传算法优化神经网络模型结果比较

在 Matlab 中利用 Learndm 函数对神经网络权阈值优化并进行建模,比较图 3 和图 4 可以看出,遗传算法优化神经网络的预测芒果酸度的相关系数达到了 0.999 97,高于直接用 BP 建立的模型,校正集样本的标准偏差 $SEC=0.005 612$ 也较低。从图 5 和图 6 的比较也可以看出,遗传算法优化神经网络的预测芒果糖度的相关系数达到了 1,高于直接用 BP

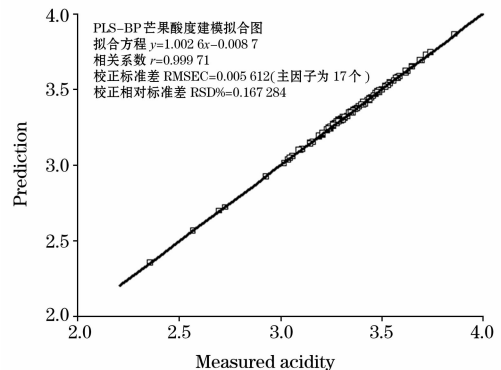


Fig. 3 PLS-BP calibration model for mango's valid acidity of 135 samples

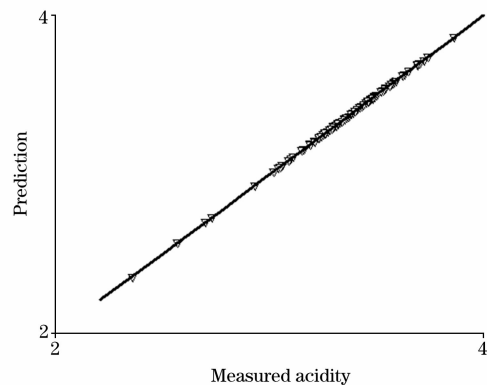


Fig. 4 PLS-GA-BP calibration model for mango's valid acidity of 135 samples

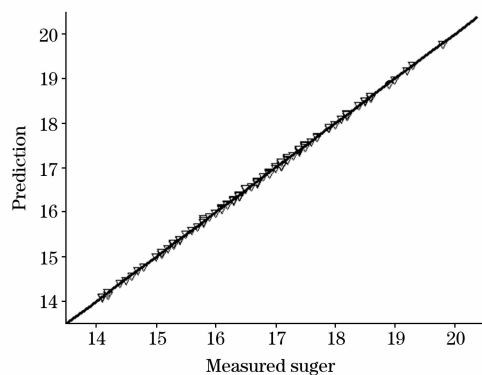


Fig. 5 PLS-BP calibration model for mango's suger content of 135 samples

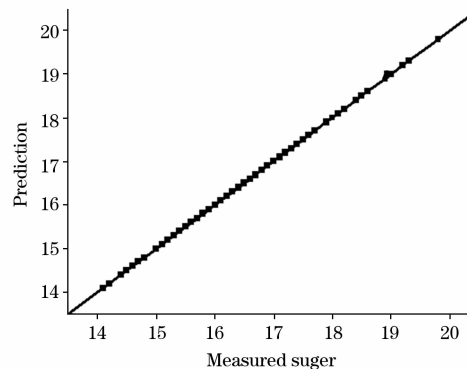


Fig. 6 PLS-GA-BP calibration model for mango's suger content of 135 samples

建立的模型,校正集样本的标准偏差 $SEC = 0.00285$ 也较低。

表 1 和表 2 为用上述两组校正模型对 45 个未知样本的预测结果。从表 1 可以看出 PLS-GA-BP 模型对芒果酸度测量值和预测值之间的相关系数和预测标准分别为 $r = 0.83699$, $SEP = 0.109447$, 对应的 PLS-BP 预测值的相关系数和预测标准差为 $r = 0.788$, $SEP = 0.133312$ 。结果表明 PLS-GA-BP 模型对芒果酸度预测结果要好于 PLS-BP; 从表 2 可以看出 PLS-GA-BP 模型对芒果糖度测量值和预测值之间的相关系数和预测标准差分别为 $r = 0.85409$, $SEP = 0.60934$, 对应的 PLS-BP 预测值的相关系数和预测标准差分别为 $r = 0.75705$, $SEP = 0.864676$ 。表明 PLS-GA-BP 模型对芒果糖度预测结果好于 PLS-BP。

Table 1 Prediction Results for Valid Acidity of PLS-BP and PLSS-GA-BP Models for Unknown Samples

模型类型	预测拟合方程	相关系数/ r	预测标准差	RSD%
PLS-BP	$y = 0.96216x + 0.14675$	0.788	0.133312	4.043301
PLS-GA-BP	$y = 0.91501x - 0.25128$	0.83699	0.109447	3.319447

Table 2 Prediction results for suger content of PLS-BP and PLS-GA-BP models for unknown samples

模型类型	预测拟合方程	相关系数/ r	预测标准差	RSD%
PLS-BP	$y = 0.90113x + 1.62064$	0.75705	0.864676	5.136687
PLS-GA-BP	$y = 0.79337x + 3.28781$	0.85409	0.60934	3.61984

3 结 论

应用可见/近红外光谱技术快速测定了芒果糖度酸度的含量,提出了遗传算法优化神经网络的方法,建立芒果糖度酸度测定的校正模型,提高了预测精度。选择 PRESS 值最小作为遗传算法的优化标准来优化信息值的权阈值。PLS-GA-

BP 模型对芒果酸度的预测结果与测量值之间的相关系数和预测标准差分别为 $r = 0.83699$ 和 $SEP = 0.109447$, 结果要好于 PLS-BP。PLS-GA-BP 模型对芒果糖度的预测结果与测量值之间的相关系数和预测标准差分别为 $r = 0.85409$ 和 $SEP = 0.60934$, 结果好于 PLS-BP。因此,应用光谱技术结合化学计量学方法测定芒果糖度酸度含量是可行的,也为开发无损在线检测仪器提供了参考。

参 考 文 献

- [1] ZHANG Zhen-wen, YAO Qing-qun(张振文,姚庆群). Subtropical Plant Science(亚热带植物科学), 2005, 34(2): 25.
 [2] CHU Xiao-li, YUAN Hong-fu, LU Wan-zhen(褚小立,袁洪福,陆婉珍). Analytical Instrumentation(分析仪器), 2006, (2): 1.

- [3] Wang Li, Lee S C F, Wang Xiao-ru, et al. Food Chemistry, 2006, 95: 529.
- [4] Saranwong Sirinnapa, Sornsrivichai Jinda, Kawano Sumio. Postharvest Biology and Technology, 2004, 31(2): 137.
- [5] Manuela Zude, Bernd Herold, Jean-Michel Roger. Journal of Food Engineering, 2006, 77(2): 254.
- [6] HE Yong, LI Xiao-li(何 勇, 李晓丽). J. Infrared Millim. Waves(红外与毫米波学报), 2006, 25(3): 192.
- [7] Bart M Nicolai, Karen I Theron, et al. Chemometrics and Intelligent Laboratory Systems, 2007, 85(2): 243.
- [8] WANG Yan-yan, HE Yong, SHAO Yong-ni, et al(王艳艳, 何 勇, 邵咏妮, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2007, 27(4): 702.
- [9] Cen Hai-yan, Bao Yi-dan, He Yong, et al. Journal of Food Engineering, 2007, 82(2): 253.
- [10] Li Xiaoli, He Yong, Wu Changqing, et al. Journal of Food Engineering, 2007, 83(3): 316.
- [11] LI Xiao-li, HE Yong, QIU Zheng-jun(李晓丽, 何 勇, 裘正军). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2007, 27(2): 279.
- [12] LU Wan-zhen, YUAN Hong-fu, XU Guang-tong, et al(陆婉珍, 袁洪福, 徐广通, 等). Modern Near-Infrared Spectroscopy Analysis Technology(现代近红外光谱分析技术). Beijing: China Petroleum Press(北京: 中国石油出版社), 2000. 167.
- [13] BAO Yi-dan, WU Yan-ping, HE Yong(鲍一丹, 吴燕平, 何 勇). Journal of Agricultural Mechanization Research(农机化研究), 2004, (3): 162.
- [14] He Yong, Li Xiaoli, Shao Yongni. Lecture Notes in Artificial Intelligence, 2005, 38(9): 1053.
- [15] Wu Xiaojuan, Zhu Xinjian, Cao Guangyi. Journal of Power Sources, 2007, 167(1): 145.
- [16] CHU Xiao-li, YUAN Hong-fu, LU Wan-zhen(褚小立, 袁洪福, 陆婉珍). Progress in Chemistry(化学进展), 2004, 16(4): 528.
- [17] Durand A, Devos O, Ruckebusch C, et al. Analytica Chimica Acta, 2007, 595(1-2): 72.
- [18] YAN Yan-lu, ZHAO Long-lian, HAN Dong-hai(严衍禄, 赵龙莲, 韩东海, 等). Basic and Application of Near Infrared Spectroscopy Analysis(近红外光谱分析基础与应用). Beijing: China Light Industry Press(北京: 中国轻工业出版社), 2005, 369.
- [19] LI Zhi-yong, KUANG Gang-yao, YU Wen-xian, et al(李志勇, 匡纲要, 郁文贤, 等). J. Infrared Millim. Waves(红外与毫米波学报), 2004, 23(4): 286.
- [20] Shen Changyu, Wang Lixia, Li Qian. Journal of Materials Processing Technology, 2007, 183(2-3): 412.

Nondestructive Test on Predicting Sugar Content and Valid Acidity of Mango by Spectroscopy Technology

YU Jia-jia, HE Yong, BAO Yi-dan*

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

Abstract Mango is a kind of popular tropic fruit in the world, and its quality will affect the health of consumers. Unsaturated acid is an important component in mango. So it is very important and necessary to detect the sugar content and valid acidity in mango fast and non-destructively. Visible and short-wave near-infrared reflectance spectroscopy (VIS/SWNIRS) was applied in the present study to predict sugar content and valid acidity of mango. Because of the non-linear information in spectral data characteristics of the pattern were analyzed by neural network optimized by genetic algorithm (GA-BP). Spectral data were compressed by the partial least squares (PLS). The best number of principal components (PCs) was selected according the accumulative reliabilities (AR). PCs could be used to replace the complex spectral data. After some preprocessing and through full cross validation, 17 principal components presenting important information of spectra were confirmed as the best number of principal components for valid acidity, and 18 PCs as best number of principal components for sugar content. Then, these best principal components were taken as the input of GA-BP neural network. One hundred thirty five samples were randomly collected as modeling, and the remaining 45 as samples to check the forecast results by the model. For the sake of testing the GA-BP model, at the same time we took the BP neural network on the same PCs. The quality of the calibration model was evaluated by the correlation coefficients (R) and standard error of calibration (SEC V), and the prediction results were assessed by correlation coefficients (R) and standard error of prediction (SEP). Comparing PLS-BP model with PLS-GA-BP model, the coefficients of determination (R) of 0.788/0.836 99 and standard errors of prediction (SEP) of 0.133 312/0.109 447 were calculated in valid acidity. The sugar content result was calculated by the coefficients of determination (R)=0.757 05/0.854 09 and standard errors of prediction (SEP)=0.864 676/0.609 34. Thus, it is obvious that this model is reliable and practicable. And the PLS-GA-BP model based on the spectroscopy technology is a better pattern to predict sugar content and valid acidity of mango, giving a new method for detecting fruit's sugar content and valid acidity.

Keywords Vis/NIR spectroscopy; Mango; Partial least squares; Genetic algorithm; Neural network

* Corresponding author

(Received May 10, 2007; accepted Aug. 20, 2007)