

文章编号:1001-9014(2006)06-0478-03

## 可见/近红外光谱预测杨梅汁酸度的方法研究

邵咏妮, 何勇

(浙江大学生物系统工程与食品科学学院, 浙江 杭州 310029)

**摘要:**针对可见/近红外光与杨梅汁酸度存在非线性相关的特点,提出了应用偏最小二乘(PLS)法预测线性部分和人工神经网络(ANN)预测非线性部分,结合两种方法综合预测杨梅汁酸度值,通过比较 $r$ , RMSEP,  $Bias$ 的值来检验该方法。其中 PLS 模型用于寻找与杨梅汁酸度值有关的敏感波段,预测杨梅汁酸度的线性部分,将这些敏感波段对应的光谱吸光度值作为人工神经网络的输入,并将杨梅汁酸度的实际测量值减去 PLS 模型校正值,获得的差额部分作为神经网络的输出,建立一个差额神经网络预测杨梅汁酸度的非线性部分。46 个样本用于建模,30 个样本用于预测。结果表明该方法对样本的预测相关系数  $r=0.939$ , RMSEP=0.218,  $Bias=-0.121$ , 好于只使用 PLS 模型的相关系数  $r=0.921$ , RMSEP=0.228,  $Bias=-0.132$ 。

**关键词:**可见/近红外光谱; 偏最小二乘; 杨梅汁; 酸度; 人工神经网络

**中图分类号:**S123; TH744.1 **文献标识码:**A

## METHOD FOR PREDICTING ACIDITY OF BAYBERRY JUICE BY USING VIS/NEAR INFRARED SPECTRA

SHAO Yong-Ni, HE Yong

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

**Abstract:** Aiming at the nonlinear correlation characteristic of visible/near infrared spectra and the corresponding acidity of bayberry juice, one mixed algorithm was presented to predict the acidity of bayberry juice with partial least squares (PLS) and artificial neural network (ANN). The values of correlation coefficient ( $r$ ), the root mean squared error of prediction (RMSEP), and bias were used to estimate the mixed model. PLS was used to find some sensitive spectra related to acidity in juice, and the values of spectral absorbance corresponding to them were regarded as the input neurons of ANN. Remnant values by subtracting standard values and validation values were regarded as the output neurons of ANN. The calibration equation developed from them was used to predict the constituent values for the independent spectra of 30 samples. The results indicate that the observed results by using PLS-ANN ( $r=0.939$ , RMSEP=0.218,  $Bias=-0.121$ ) are better than those obtained by PLS ( $r=0.921$ , RMSEP=0.228,  $Bias=-0.132$ ).

**Key words:** visible/near infrared spectra; partial least squares (PLS); bayberry juice; acidity; artificial neural network (ANN)

### 引言

可见/近红外光谱法是一种快速简便非破坏性的分析方法。它已被广泛应用于定性或定量分析水果的品种鉴别及水果淤伤的检测<sup>[1-3]</sup>等。总体来说光谱技术在各个方面尤其是在食品科学和农业科学上已成为一种重要的检测技术。

杨梅是我国南方著名的特产水果,除日本有少量栽培外,其他国家很少有栽培。杨梅果实色泽艳

丽,酸甜适口,风味独特。营养价值高,是我国重要的经济水果。水果的产后处理、品质鉴别检测一直是农产品加工研究的重要课题。杨梅汁作为一种绿色饮料,其口感、糖分含量、酸度和维生素含量等内部品质,都得到了广泛重视,其中酸度是重要的品质评价指标之一。当前大多测量酸度的仪器都需进行取样,并有一定的消耗,不宜在连续生产线上应用,且易产生人为误差。而采用可见/近红外光谱技术可实现水果和果汁内部品质的无损检测。然而使用近红外光

收稿日期:2005-12-31,修回日期:2006-05-08

Received date: 2005-12-31, revised date: 2006-05-08

基金项目:国家自然科学基金(30671213)、高等学校优秀青年教师教学科研奖励计划(02411)和高等学校博士学科点专项科研基金(20040335034)资助项目

作者简介:邵咏妮(1983-),女,浙江慈溪人,浙江大学生物系统工程与食品科学学院博士生,主要从事数字农业和多光谱检测技术研究。

谱技术的关键在于如何从光谱数据中获得有用的光谱信息。

本文研究的目的是使用可见/近红外光谱技术测量杨梅汁酸度并寻找最佳的数学模型对杨梅汁的酸度进行预测。因此我们提出了一种基于偏最小二乘与BP-神经网络相结合的数据处理模型。

## 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 ~ 1075nm 之间,采样间隔为 1.5nm,分辨率为 3.5 nm,探头视场角为 20°,光谱仪垂直放置于距离样本表面 12cm 处,光源与水平位置成 60° 照射,如图 1 所示。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 V9.2 和 DPS。

4 个品种的杨梅汁(慈溪杨梅,仙居杨梅,丽水杨梅和临海杨梅)采购自超市,共 76 个样本(每个品种 19 个)。样本被放置在密封的内置冰块保持  $4 \pm 1^\circ\text{C}$  温度的玻璃瓶中。所有的样本在可见/近红外光谱分析前都恢复到  $25^\circ\text{C}$  室温。每个样本都被标记并分别放置在容器中。为了减少在整个实验中的操作误差,采用直径为 6.5cm,高度为 1.4cm 的玻璃容器,每个样本在容器中的高度固定为 1cm,待稳定后进行光谱数据的采集,扫描次数设为 30。

杨梅汁酸度的实测值通过使用 PH 计(SJ-4A, Exact instrument Co., Ltd., Shanghai, China) 测得,其测量精度为 0.01。

### 1.2 光谱数据的预处理

为了去除来自高频随机噪音、基线漂移、样本不均匀、光散射等影响。采用 Savitzky-Golay 平滑法,平滑点数为 3,然后采用多元散射校正(MSC)对光谱数据进行处理。为了得到高的信噪比,去除部分噪声信息,选择 400nm ~ 1000nm 波段的数据进行分析处理。

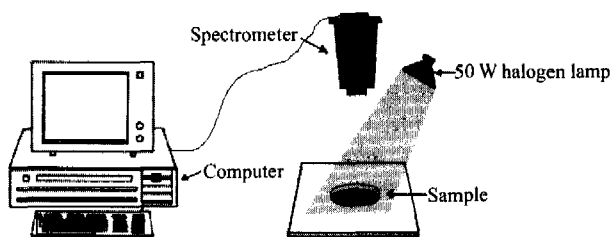


图1 杨梅汁光谱数据的采集  
Fig. 1 Collection for spectral data of bayberry juice

### 1.3 化学计量学方法

偏最小二乘法(PLS)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。

在分析应用中,最常用的人工神经网络<sup>[4,5]</sup>是具有反向学习算法的 BP-神经网络。BP-神经网络通过在学习过程中修改权重系数来减少模型的拟合残差,权重的增加或减少通过梯度下降的规则实现。

## 2 试验结果与分析

### 2.1 杨梅汁酸度的敏感波段

对所有杨梅汁样本建立 PLS 模型,得到各个波段对应的回归系数图,如图 2 所示,685 ~ 695nm, 945 ~ 965nm 为较敏感的波段。由于 685 ~ 695nm 波段属于可见光,在该波段杨梅汁的颜色差异占主导地位,而且在该波段不存在任何的有机酸。因此杨梅汁酸度的敏感波段为 945 ~ 965nm。

### 2.2 偏最小二乘和人工神经网络结合模型

本文提出了一种将偏最小二乘和人工神经网络算法相结合,建立杨梅汁可见/近红外光谱和杨梅汁酸度之间非线性部分模型的方法。因此,整个预测模型算法的建立由线性部分和非线性部分组成。线性部分通过对 46 个杨梅汁样本建立 PLS 校正模型。非线性部分,通过 PLS 模型选择负载较大值所对应的光谱波段,并将这些波段对应的光谱吸光度值作为神经网络的输入部分;对应的杨梅汁的实测值减去校正值得到的差额值作为神经网络的输出部分,建立差额神经网络。

非线性的模型为一个三层的 BP-神经网络, Sigmoid 为变换函数,输入层结点数为 21 个,隐含层结

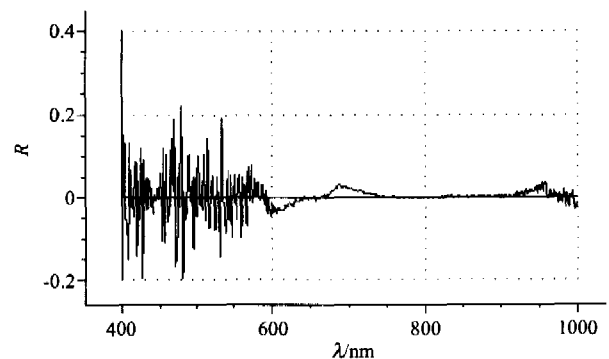


图2 杨梅汁酸度的回归系数与其对应的波段  
Fig. 2 Regression coefficients with corresponding wavelengths for acidity

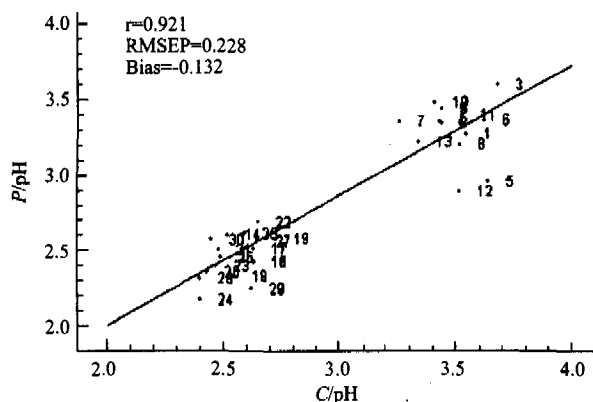


图3 30个未知样本的PLS预测模型

Fig. 3 Prediction results for the unknown 30 samples from the PLS models

点数为6,输出层结点数为1.目标误差设为0.0001,学习速度为0.2,迭代次数为1000.

线性部分的预测,通过所建的PLS模型预测剩余的30个杨梅汁样本的酸度值,如图3所示.非线性部分,将30个样本对应的敏感波段的光谱吸光度值作为BP神经网络的输入,预测其差额值.最后通过由PLS模型得到的预测值和BP神经网络得到的差额值相加得出待预测的30个杨梅汁的酸度值.

该模型通过 $r$ , RMSEP, Bias的值来检验其测量的精确性及误差.计算公式如下

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

$$\text{Bias} = \frac{1}{I_p} \sum_{i=1}^{I_p} (\hat{y}_i - y_i), \quad (2)$$

式中 $\hat{y}$ 表示样本的预测值, $y_i$ 表示样本的实测值, $I_p$ 表示检验的样本数.

图4显示了PLS-ANN模型样本酸度实测值和预测值之间的相关系数 $r=0.939$ , RMSEP=0.218, Bias=-0.121;对应的PLS预测值的相关系数 $r=0.921$ , RMSEP=0.228, Bias=-0.132.结果表明PLS-ANN模型的预测结果要好于PLS.

### 3 结语

通过PLS模型建立46个样本的酸度模型,同时得到酸度的敏感波段,并且将其对应的光谱吸光度作为BP神经网络的输入,通过将实测值和校正

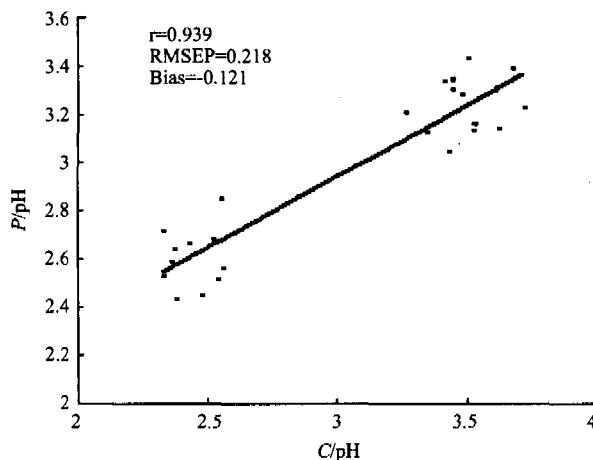


图4 30个未知样本的PLS-ANN预测模型

Fig. 4 Prediction results for the unknown 30 samples from the PLS-ANN models

相减得到的差额作为BP神经网络的输出,建立差额神经网络.30个剩余样本的酸度值由线性部分PLS模型的预测值和非线性部分PLS-ANN模型的差额值相加得到,其实测值和预测值之间的相关系数达到0.939, RMSEP为0.218, Bias为-0.121.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地预测杨梅汁的酸度,结果要好于PLS模型的预测效果.本文提出的基于PLS与ANN相结合的光谱数据处理方法对具有非线性属性的数据预测具有较理想的效果.

### REFERENCES

- [1] HE Yong, LI Xiao-Li, SHAO Yong-Ni. Quantitative analysis of the varieties of apple using near infrared spectroscopy by principal component analysis and BP model [J]. *Lecture Notes in Artificial Intelligence*, 2005, **3809**:1053—1056.
- [2] Bennedsen B S, Peterson D L. Performance of a system for apple surface defect identification in near-infrared images [J]. *Biosystems Engineering*, 2005, **90**(4):419—431.
- [3] HE Yong, LI Xiao-Li. Discrimination of varieties of waxberry using near infrared spectra [J]. *J. Infrared Millim. Waves* (何勇, 李晓丽. 近红外光谱杨梅品种鉴别方法的研究. *红外与毫米波学报*), 2006, **25**(3):192—194.
- [4] LIN San-Hu, ZHU Hong, ZHAO Yi-Gong. Model for sea clutter based on neural network [J]. *J. Infrared Millim. Waves* (林三虎, 朱红, 赵亦工. 基于神经网络的海杂波模型. *红外与毫米波学报*), 2004, **23**(1):55—58.
- [5] HE Yong, SONG Hai-Yan, Pereira A G, et al. A new approach to predict N, P, K and OM content in a loamy mixed soil by using near infrared reflectance spectroscopy [J]. *Lecture Notes in Computer Science*, 2005, **3644**:859—867.