文章编号：0253-2239(2009)02-0537-04

基于可见-近红外反射光谱技术的葡萄品种鉴别方法的研究

曹 芳 吴 迪 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 提出一种利用可见-近红外反射光谱技术快速无损鉴别葡萄品种的新方法。采用主成分分析法对三个葡萄品种的光谱进行聚类分析。结果表明，黑提葡萄能够被区分。进一步采用人工神经网络技术对马奶子和木拉格两种葡萄进行品种鉴别。以前 10 个主成分作为神经网络的输入，品种类型作为神经网络的输出，建立三层 BP 神经网络模型。结果显示，这两个品种的识别准确率达到 98.28 %，结果优于簇类独立软模式(SIMCA)。同时提出葡萄品种鉴别的四个敏感波段：452、493、542 和 668 nm。基于敏感波段光谱的 BP 神经网络预测准确率为 97.41%。说明采用可见-近红外光谱分析技术结合主成分分析和人工神经网络的方法能够快速无损鉴别葡萄的品种，为葡萄品种的鉴别提供了一种新方法。

关键词 光谱学；葡萄品种鉴别；可见-近红外反射光谱；主成分分析；人工神经网络

中图分类号 S123；TH744.1 **文献标识码** A **doi**：10.3788/AOS20092902.0537

Variety Discrimination of Grapes Based on Visible-Near Reflection Infrared Spectroscopy

Cao Fang Wu Di He Yong Bao Yidan

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

Abstract A non-destructive method for discriminating varieties of grapes by visible and near reflection infrared spectroscopy (VIS-NIRS) was developed. The spectral data of three varieties of grape samples were clustered by principal component analysis (PCA). The results indicate that Heiti grape sample can be totally separated from the other two. Mainaizi and Mulage grape samples were discriminated based on back propagation-neural networks (BP-NN) model. The three hidden-layer BP-NN model was built with the first ten PCs as inputs, and the dummy variety numbers of grapes as outputs. The correct answer rate 98.28 % of BP-NN model is achieved, which is better than the one achieved by the soft independent modeling of class analogy(SIMCA) method. Four effective wavelengths for variety discrimination are 453,493,542 and 668 nm. The correct answer rate of BP-NN model based on the spectra of effective wavelengths is 97.41 %. The result indicates that variety discrimination of grapes can be achieved rapidly and non-destructively by using VIS-NIRS with PCA and BP-NN.

Key words spectroscopy; grape recognition; visible-near infrared reflection spectroscopy; principal component analysis; artificial neural network

1 引 言

不同品种的葡萄的口感、品质和营养价值差别较大。随着产后处理和加工技术的发展，葡萄品种的鉴别显得更为重要。由于常用的分析化学方法操作繁琐、成本高、造成样本破坏等不足，研究一种快速、简便、无损的葡萄品种鉴别技术很有必要。

收稿日期：2008-05-05；**收到修改稿日期**：2008-08-02

基金项目：浙江省自然科学基金(Y307158)、宁波科技攻关国际合作项目(2008C10037)和浙江省教育厅(20071064)资助课题。

作者简介：曹 芳(1987－)，女，本科生，主要从事光谱与多光谱检测技术、数字农业方面的研究。E-mail：kathycf919@yahoo.com.cn

导师简介：何 勇(1963－)，男，浙江大学生物系统工程与食品科学学院副院长，博士生导师，主要从事光谱与多光谱检测技术、数字农业方面的研究。E-mail：yhe@zju.edu.cn

近红外光谱(NIRS)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，能够反映有机分子中基团的特征振动信息，其光谱特性与有机物质的类型和含量密切相关。由于其具有效率高、成本低、速度快、测试重现性好、测量方便等优点，已被越来越多地应用于食品工业、石油化工、制药工业等领域[1~9]。

本文采用可见-近红外光谱技术，应用主成分分析和反向传播(BP)神经网络建立不同葡萄品种的可见-近红外光谱鉴别模型，并选择了葡萄品种鉴别的敏感波段。

2 材料和方法

使用美国 Analytical Spectral Device 公司的 Handheld Field Spec 光谱仪，测定范围 325～1075 nm，扫描 30 次，探头视场角为 10°。采用漫反射方式进行样品光谱采样。光源采用与光谱仪配套的 14.5 V 卤素灯。分析软件为 ASD View Spec Pro V2.14，Unscrambler V9.6 和 DPS(Data Procession System for Practical Statistics)。

从超市购买马奶子(Manaizi)葡萄(产地新疆吐鲁番)、木拉格(Mulage)葡萄(产地新疆喀什)、黑提(Heiti)葡萄(产地山东)三种葡萄，共取 439 个样本进行实验，其中马奶子葡萄 197 个，木拉格葡萄 127 个，黑提葡萄 115 个。光谱仪置于葡萄上方，距离葡萄表面 120mm，对每个样本扫描 30 次。

为了消除来自高频随机噪音、基线漂移、样本不均匀、光散射等影响，应用 Unscrambler 软件对光谱数据进行预处理。首先采用平均平滑法，平滑窗口大小为 9，此时能很好地滤除各种因素产生的高频噪音。再进行多元散射校正(Multiplicative Scatter Correction，MSC)处理。

主成分分析(PCA)是多元统计中的一种数据挖掘技术。将数据降维，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，以消除众多信息共存中相互重叠的部分。通过对原始大量光谱变量进行转换，使数目较少的新变量成为原始变量的线性组合。

人工神经网络模型是一个强有力的学习系统，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网络(Back Propagation Networks，BP 神经网络)。采用 DPS 软件进行人工神经网络建模。建立一个 3 层的 BP 神经网络结构，各层的传递函数都采用 BP 神经网络常用的对数 S 型(Sigmoid)函数。以前 10 个主成分作为神经网络的输入，即网络输入层节点数为 10，隐含层节点数由经验公式和实验数据论证确定，输出层节点数为 1(葡萄的不同品种值)。设定目标误差为 0.0001，网络指定参数中学习速率为 0.1，设定训练迭代次数为 1000 次。

3 实验结果与分析

三种葡萄品种的典型近红外光谱曲线如图 1 所示。图 1 中横坐标为波长，范围是 325～1075 nm，纵坐标为光谱漫反射率。从图 1 中可看出，不同品种的葡萄光谱曲线存在差异，并具有一定的特征性和指纹性，为葡萄的品种鉴别提供了数学基础。其中黑提品种的光谱和其它两个品种差距较大，尤其在 520～640 nm 区域能够很好地区分，而马奶子和木拉格的光谱曲线较为接近。由于同一类样本光谱曲线会出现漂移，造成曲线间相互重叠，因此难以通过直接观察光谱曲线进行品种鉴别，需要通过化学计量学方法进一步分析。由于光谱曲线在首端和末端有较大噪音(如图 1)，所以只取 400～1000 nm 波段的光谱进行分析[10]。

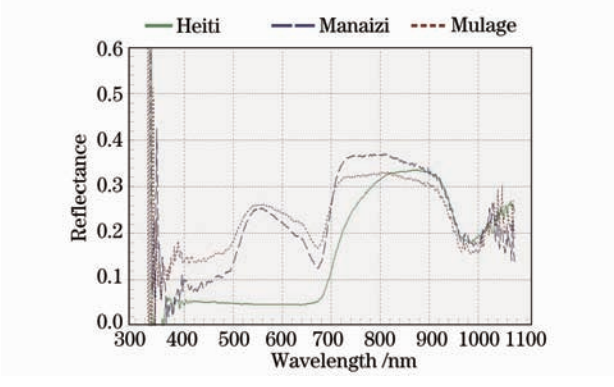

图 1 三种葡萄的典型可见-近红外光谱反射图

Fig. 1 Typical visible and near infrared reflectance spectra of three varieties of grape

对样本进行主成分分析，数据矩阵从原始的 439×601 减少到 439×10(10 个主成分)。主成分的得分能够反映样本间的相似性和独特性，每个样本对应不同主成分有不同得分值。基于样本的得分图能够揭示样本的内部特征和聚类信息。如果把每个样本在第 1 和第 2 个主成分的得分值在图中表达出来，就得到了前 2 个主成分的聚类图(图 2)。图 2 中横坐标表示每个样本的第一主成分得分值，纵坐标表示每个样本的第二主成分得分值。图中三种葡

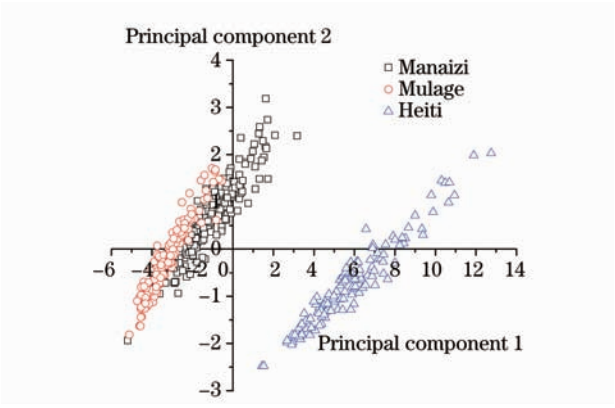

图 2　主成分分析后的 439 个葡萄样本光谱的主成分 1 和主成分 2 的聚类图

Fig. 2　PCA clustering plot(PC1×PC2) for three different varieties of 439 grape samples

葡的分布存在一定的聚类特性，其中黑提葡萄主要分布在纵坐标右侧区域，马奶子和木拉格主要分布在左侧区域。从聚类图中可以将黑提葡萄与其它两种葡萄区分开来。但是马奶子和木拉格两种葡萄存在严重的交叉分布，从聚类图中难以直接区别，为此利用 BP 神经网络分析加以鉴别。

全波段从 400～1000 nm 共有 601 个数据点。如果采用全谱段计算，则输入变量太多增加了建模难度，而且容易导致神经网络的训练时间过长，模型难以收敛。同时，无关的噪声信息也会融入到模型中反而降低了模型的预测精度。为此，通过主成分分析，提取对葡萄品种敏感的新变量作为输入建立神经网络品种鉴别模型。主成分的累计可信度如表 1 所示。计算表明，前 10 个主成分的累计可信度达到 99.88 %，表示这 10 个主成分能够解释原始光谱信息的 99.88 %。

表 1　前 10 个主成分的累计可信度

Table 1　First ten principal components (PCs) and accumulative reliabilities

No. of first PCs	1	2	3	4	5
Accumulative reliabilities /%	90.48	96.37	98.39	99.28	99.54
No. of first PCs	6	7	8	9	10
Accumulative reliabilities /%	99.72	99.80	99.83	99.86	99.88

将建模样本的 10 个主成分作为 BP 神经网络的输入变量建立预测模型，通过调整隐含层的节点数来优化网络结构[11]。选择网络输入节点数为 10，输出节点数为 1(品种值：马奶子葡萄为 1，木拉格葡萄为 2)，分别选取 5、7、9、11、13 作为隐含层节点数进行预测，网络设定训练迭代次数为 1000 次，用两种葡萄共 208 个样本进行建模，对未知的 116 个样本进行预测(见表 2)。对于不同隐含层节点数设置，预测结果准确率均为 98.28%(见表 3)。结果表明，隐含层节点数的不同对本实验预测结果影响并不明显。模型除了把 73 号和 87 号木拉格葡萄样本误判为马奶子葡萄之外，对其它样本的品种判别均正确。此外，为了比较 BP 神经网络所得出的实验结果，利用 SIMCA 方法对相同的建模集和预测集进行品种鉴别，该模型误判了第 27、40、74、88 号样本，所得出的识别正确率为 96.55%。BP 神经网络的结果优于 SIMCA 方法，说明 BP 神经网络能够较好地建立光谱和葡萄品种间的鉴别模型。

表 2　各品种葡萄样本的建模集和预测集样本数

Table 2　Amount of three different varieties of grapes used in calibration and prediction

Variety	Sample number	Calibration set	Prediction set
Manaizi	197	126	71
Mulage	127	82	45

表 3　BP 神经网络模型对未知样本的预测结果

Table 3　Prediction results for unknown samples by BP model

Varieties	Number of samples used for prediction	Number of samples predicted into		Correct answer rate /%
		Manaizi (1)	Mulage (2)	
Manaizi (1)	71	71	0	100
Mulage (2)	45	2	43	95.56
Total	116	73	43	98.28

马奶子和木拉格两种葡萄由于颜色比较相近，当大量样本曲线重叠时，难以直接区分。采用 Discrimination power 对光谱曲线进行分析。结果显示，马奶子和木拉格葡萄品种鉴别的敏感波段为

452 nm、493 nm、542 nm 和 668 nm。由图 1 中也可知，马奶子和木拉格的光谱曲线在这四个波段有较大差异。采用该四个敏感波段的光谱反射率作为输入，品种值作为输出，对相同的建模集和预测集采用 BP 神经网络进行马奶子和木拉格的品种区分。实验结果表明，利用敏感波段进行马奶子和木拉格的品种鉴别，所得到的准确率为 97.41 %。虽然利用敏感波段的品种鉴别率略低，但已能满足实际生产应用需求。同时利用敏感波段进行品种鉴别，能够大大减少计算量，提高鉴别速度。

4 结 论

基于可见-近红外光谱技术建立了快速无损检测葡萄品种的模型。首先对原始光谱数据进行主成分分析，通过主成分聚类图将黑提葡萄与其它两种葡萄区分开来。为减少计算量、提高分析和识别的速度，提取前 10 个主成分，作为 BP 神经网络的输入，对马奶子和木拉格两个品种进行鉴别，识别准确率为 98.28 %。结果表明应用可见-近红外光谱技术能够快速、准确地鉴别葡萄的品种。研究为葡萄品种的快速无损检测提供了一种新方法。论文进一步提出了马奶子和木拉格品种鉴别的四个敏感波段：452 nm、493 nm、542 nm 和 668 nm。基于敏感波段光谱的 BP 神经网络预测准确率为 97.41 %。同时在 542 nm 波段能够将黑提与其它两个葡萄品种区分开来。在未来的研究中将考虑通过滤光片或类似技术采集少数几个敏感波段的光谱，开发简单、快速的葡萄品种鉴别仪器。

参 考 文 献

1 Di Wu, Yong He, Shuijuan Feng. Short-wave near-infrared spectroscopy analysis of major compounds in milk powder and wavelength assignment [J]. *Analytica Chimica Acta*, 2008, **610**(2): 232～242

2 He Yong, Li Xiaoli. Discrimination of varieties of waxberry using near infrared spectra [J]. *J. Infrared Millim. Waves*, 2006, **25**(3): 192～194
何 勇，李晓丽. 近红外光谱杨梅品种鉴别方法的研究[J]. 红外与毫米波学报，2006，**25**(3)：192～194

3 Jia Dongyao, Ding Tianhuai. Novel method of detecting foreign fibers in lint by fiber's infrared absorption characteristic[J]. *J. Infrared Millim. Waves*, 2005, **24**(2): 147～150
郑东耀，丁天怀. 利用纤维红外吸收特性的皮棉杂质检测新方法[J]. 红外与毫米波学报，2005，**24**(2)：147～150

4 Yong He, Shuijuan Feng, Deng Xunfei *et al*.. Study on lossless discrimination of varieties of yogurt using the Visible/NIR-spectroscopy[J]. *Food Research International*, 2006, **39**(6): 645～650

5 Chen Quansheng, Zhao Jiewen, Zhang Haidong *et al*.. Indentification of authenticity of tea with near infrared spectroscopy based on support voctor machine [J]. *Acta Optica Sinica*, 2006, **26**(6): 933～937
陈全胜，赵杰文，张海东 等. 基于支持向量机的近红外光谱鉴别茶叶的真伪[J]. 光学学报，2006，**26**(6)：933～937

6 Zhao Jiewen, Zhang Haidong, Liu Muhua. Preprocessing methods of near-infrared spectra for simplifying prediction model of sugar content of apples [J]. *Acta Optica Sinica*, 2006, **26**(1): 136～140
赵文杰，张海东，刘木华. 简化苹果糖度预测模型的近红外光谱预处理方法[J]. 光学学报，2006，**26**(1)：136～140

7 Shi Youming, Liu Gang, Liu Jianhong *et al*.. Identification of auricularia auricula from different regions by fourier transform infrared spectroscopy [J]. *Acta Optica Sinica*, 2007, **27**(1): 129～132
时有明，刘 刚，刘剑虹 等. 不同产地黑木耳的傅里叶变换红外光谱鉴别[J]. 光学学报，2007，**27**(1)：129～ 132

8 Xu Xiangqun, Wu Liu. Dependence of optical clearing effect on tissue structure [J]. *Chinese J. Lasers*, 2006, **33**(7): 998～1002
徐向群，吴 柳. 不同结构生物组织光透明作用比较[J]. 中国激光，2006，**33**(7)：998～1002

9 Wei Huajiang, Xing Da, Wu Guoyong *et al*.. Using spatially resolved reflectance to measure optical properties of stomach tissue [J]. *Chinese J. Lasers*, 2007, **34**(4): 582～587
魏华江，邢 达，巫国勇 等. 采用空间分辨漫反射测定胃组织光学特性[J]. 中国激光，2007，**34**(4)：582～587

10 Yi Qiu, Shu Xiaozhou, Xu Zhaoan *et al*.. Analysis on the ultra-spectral characteristics of water environmental parameters about lake [J]. *J. Infrared Millim. Waves*, 2004, **23**(6): 427～435
尹 球，疏小舟，徐兆安 等. 湖泊水环境指标的超光谱响应特征分析[J]. 红外与毫米波学报，2004，**23**(6)：427～435

11 Lin Sanhu, Zhu Hong, Zhao Yigong. Model for sea clutter based on neural network [J]. *J. Infrared Millim. Waves*, 2004, **23**(1): 55～58
林三虎，朱 红，赵亦工. 基于神经网络的海杂波模型[J]. 红外与毫米波学报，2004，**23**(1)：55～58