

文章编号:1001-9014(2006)03-0192-03

用近红外光谱鉴别杨梅品种的研究

何勇, 李晓丽

(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快速无损鉴别杨梅品种的新方法,首先用主成分分析法对典型的四个杨梅品种进行聚类分析,获取杨梅的近红外指纹图谱,再结合人工神经网络技术进行品种鉴别.主成分分析表明,以主成分1和2对样本的得分值做出的得分图,对不同种类杨梅具有较好的聚类作用,可以定性分析杨梅种类.利用主成分分析可以把原始波长变量压缩成能代表原始变量的少数相互正交的主成分,用这些新变量作为神经网络的输入,建立3层BP人工神经网络模型.四个杨梅品种共100个样本用来建立神经网络品种鉴别模型,对未知的20个样本进行预测,结果表明,品种识别准确率达到95%.说明综合主成分分析和人工神经网络的方法具有很好的分类和鉴别作用,为杨梅的品种鉴别提供了一种新方法.

关键词:近红外光谱;杨梅;主成分分析;人工神经网络;聚类

中图分类号:S123; TH744.1 **文献标识码:**A

DISCRIMINATING VARIETIES OF WAXBERRY USING NEAR INFRARED SPECTRA

HE Yong, LI Xiao-Li

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

Abstract: A new nondestructive method for discriminating varieties of waxberry by visible and near infrared spectroscopy (Vis/NIRS) was developed. First, the spectral data were analyzed by principal component analysis (PCA) for varieties clustering. Then diagnostic information was obtained from original spectra, these informations were used for pattern recognition based on ANN model. The score plot provided the reasonable clustering of the varieties of waxberry. Small quantities of principal components from PCA were used as inputs of a back propagation neural network (BPNN) with one hidden layer. 100 samples were selected randomly from four varieties, then they were used to build BPNN model. This model had been used to predict the varieties of 20 unknown samples. The recognition rate of 95% was achieved. This model is reliable and practicable. So this method could offer a new approach to the fast discriminating varieties of waxberry.

Key words: near infrared spectra; waxberry; principal component analysis; artificial neural network; clustering

引言

中国是盛产杨梅的少数几个国家之一.杨梅果实柔软极不耐贮藏和运输,不同品种之间口感、大小、品质等方面差别较大,随着杨梅产后处理和加工技术的发展,品种鉴别显得越来越重要.所以研究一种简单、快速、非破坏的杨梅品种鉴别技术很有必要.我们以光谱技术为基础研究杨梅品种的快速鉴别方法.

现代光谱分析技术,可充分利用全波段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,已经被越来越多地应用于食品工业、制药工业等领域.有国内外很多学者研究利用近红外光谱技术区别物质品种如咖啡品种^[1]、小麦品种^[2]、苹果品种^[3]和道地山药^[4]等.以上研究都表明光谱技术具有对农产品进行品种识别的能力.但是大部分的品种鉴别只给出了定性结果,没有建立定量鉴别模型.即使某些研究中建立了定量鉴别模型,但是模型的正确识别率也不高,如小麦品种鉴别模型对于两个品种未知样本的识别率只有81%,82%^[2],对道地山药进行鉴别的正确识别率为70%^[4].

BP神经网络模型是一个强有力的学习系统,能

收稿日期:2005-11-20,修回日期:2006-03-20

Received date: 2005-11-20, revised date: 2006-03-20

基金项目:国家自然科学基金项目(30270773);教育部高等学校优秀青年教师教学科研奖励计划(02411)

作者简介:何勇(1963-),男,浙江宁波人,浙江大学生物系统工程与食品科学学院博导,博士,主要从事数字农业和多光谱检测技术研究.

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)。主成分分析是多元统计中的一种数据挖掘技术。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。我们提出了应用主成分分析(PCA)和 BP 神经网络相结合的方法,建立不同品种杨梅的近红外光谱鉴别模型。

1 材料与方 法

1.1 仪器 设备

实验使用美国 ASD 公司的 Handheld Field Spec 光谱仪,其光谱采样间隔 1.5nm,测定范围 325 ~ 1075nm,扫描次数 30 次,分辨率 3.5 nm,探头视角为 20 度。光源是与光谱仪配套的 14.5V 卤素灯。用该光谱仪自带的软件 ASD View Spec Pro 来观察数据曲线。

1.2 样品来源及光谱的获取

从超市买来荸荠梅、下坑梅、临海早大梅、东魁梅等四种杨梅,分别各取 30 个,共计 120 个样本进行实验(见表 1)。光谱仪置于杨梅的上方,距离杨梅表面 120mm,对每一个样本扫描 30 次。

1.3 光谱数据预处理

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,应用 Unscrambler 分析软件进行光谱预处理。先用平均平滑法,选用平滑窗口为 9,此时能很好滤除各种因素产生的高频噪音,再进行 MSC (Multiplicative Scatter Correction) 处理。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 ~ 1000nm 波段的光谱用于分析^[5]。

1.4 人工神经网络模型

建立了一个 3 层的人工神经网络结构^[6],各层传递函数都用 S 型函数。网络输入层节点数为 20,经多次实验确定隐含层节点数为 12,输出层节点数为 1。目标误差为 0.0001,网络指定参数中学习速率为 0.2,设定训练迭代次数为 1000 次。采用 DPS (data procession system for practical statistics) 软件进行

人工神经网络建模。

2 试验结果与分析

2.1 杨梅样本的近红外漫反射光谱

四种杨梅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 ~ 1075nm,纵坐标为光谱漫反射率。从图 1 中可以看出,不同品种杨梅的光谱曲线有明显区别,并具有一定的特征性和指纹性,这一差异为杨梅的不同品种鉴别奠定了数学基础。选择波长范围在 400 ~ 1000nm 的光谱,应用 ASD View Spec Pro 软件,把同一个杨梅 3 个不同部位的光谱曲线做平均处理,并转换成 ASCII 码,形成反射率矩阵,并用主成分分析法对其聚类分析。

2.2 主成分分析对不同品种杨梅进行聚类

主成分分析的目的是将数据降维,以消除众多信息共存中相互重叠的信息部分。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变量的线性组合,而且,新变量能最大限度的表征原变量的数据结构特征。对样本进行主成分分析,数据矩阵从原始的 120×600 减少到 120×20 (20 个主成分)。每个样本对应每个主成分有一个得分值,主成分的得分能够反映样本间的相似性和独特性。基于样本的主成分得分图能够揭示样本的内部特征和聚类信息。如果把每个样本的第 1 和第 2 个主成分的得分值在图中表达出来,就得到了前两个主成分的二维的得分图(如图 2 所示)。对四种杨梅共 120 个样本进行主成分分析聚类。分析表明前 2 个主成分对四种杨梅有较好的聚类作用,能直观地对不同品种杨梅进行定性分析。

图 2 表示 120 个建模样本的主成分 1、2 得分图,图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 2 中临海早大梅、荸荠梅、东魁梅明显分成 3 类,但是荸

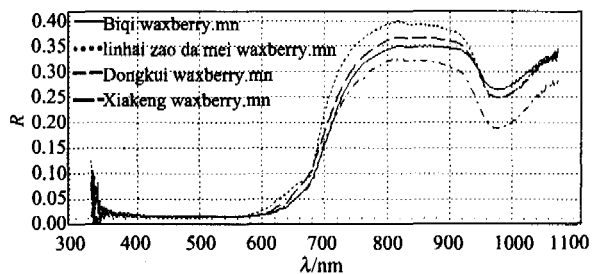


图 1 四种杨梅的典型近红外光谱反射图

Fig. 1 Near infrared reflectance spectra of four typical waxberry varieties

表 1 实验样品的品种情况

Table 1 Varieties of waxberry in the experiment

品 种	产地	样本数目	颜色
荸荠梅 (Biqi waxberry)	浙江慈溪	30 个	紫红色
下坑梅 (Xiakeng waxberry)	浙江丽水	30 个	红色
临海早大梅 (Lin Hai Zaodamei waxberry)	浙江临海	30 个	紫红色
东魁 (Dongkui waxberry)	浙江仙居	30 个	红色

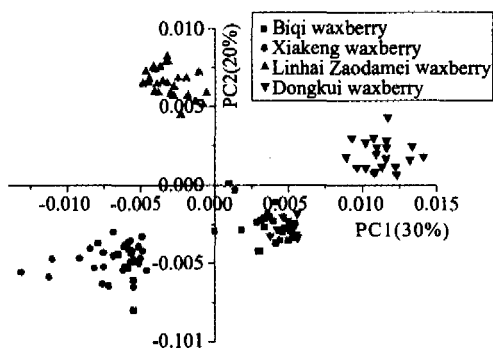


图2 120个杨梅的主成分1和主成分2的得分图
Fig.2 PCA scores plots(PC1PC2) for 120 waxberry samples

荠梅与下坑梅和东魁梅有部分重叠在一起. 说明主成分1、2对4种杨梅有较好的聚类作用, 尤其是对临海早大梅、荸荠梅、东魁梅3种杨梅有很好的聚类作用. 因此, 我们基于主成分1、2对于120个样本的得分图, 可对120个杨梅样本的品种进行定性分析. 从图2中可以看出, 临海早大梅的30个样本聚合度较好, 紧密的分布在图2的第一象限, 其它样本均处于第一象限之外. 下坑梅的聚合度也很好, 所有30个样本全部位于第三象限. 东魁杨梅大部分样本位于第二象限, 但是有一些样本落在了第四象限. 荸荠杨梅的聚合度不如前面几种, 大部分位于第四象限, 但是有部分样本落在了第三象限与下坑梅重叠在一起. 于是我们可以发现主成分1、2对四个品种杨梅虽有较好的聚类作用, 但是荸荠梅、下坑梅和东魁梅有一些交叉, 不能完全区分四个品种. 为此, 我们在主成分分析的基础上, 结合人工神经网络建立杨梅品种的鉴别模型.

2.3 基于神经网络建立杨梅品种鉴别模型

全波段从400~1000nm共有600个数据点, 但是, 直接把全光谱数据输入神经网络建立模型时, 容易导致神经网络的训练时间过长, 模型难以收敛, 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. 所以通过主成分分析, 提取对于杨梅品种敏感的新变量作为输入建立神经网络品

种鉴别模型. 计算表明, 前20个主成分的累计贡献率达到98.9%, 这20个新变量能代表绝大部分原变量包含的信息.

将全部样本随机分成建模集和预测集, 建模集有100个样本, 预测集有20个样本. 把这20主成分作为BP神经网络的输入变量, 品种作为输出变量(分别以1、2、3和4代表荸荠梅、下坑梅、临海早大梅、东魁梅), 建立杨梅品种鉴别模型. 通过调整隐含层的节点数来优化网络结构^[7], 经过反复试验得到最佳网络结构为20(输入)-12(隐含)-1(输出)的3层BP神经网络模型, 对100个建模样本的拟合残差为 1.508458×10^{-6} , 对未知的20个样本进行预测, 预测结果如表2所示. 模型除了把一个荸荠梅样本误判为下坑梅之外, 对其它样本品种的判断均正确, 预测准确率为95%.

3 结论

提出了一种基于近红外光谱的杨梅品种无损快速鉴别方法. 通过实验获得了四个典型杨梅品种的光谱特征, 利用这些光谱特征结合主成分分析人工神经网络方法建立了杨梅品种鉴别模型. 实验结果表明, 该模型的预测效果很好, 识别率达到95%. 说明运用近红外光谱技术可以快速、准确、无损的对杨梅品种进行鉴别. 本文用于杨梅品种分析的光谱在320-1075nm范围内, 说明该波长范围是对杨梅品种敏感的特征波段. 我们提出的主成分结合人工神经网络方法用于鉴别杨梅品种是成功的, 为红外光谱品种鉴别提供了一种新的分析方法.

REFERENCES

[1] Esteban-Diez I, Gonzalez-Saiz J M, Pizarro C. An evaluation of orthogonal signal correction methods for the characterisation of arabica and robusta coffee varieties by NIRS [J]. *Analytica. Chimica. Acta*, 2004, 514(1):57-67.
[2] Haluk Utku. Application of the feature selection method to discriminate digitized wheat varieties [J]. *Journal of Food Engineering*, 1999, 46(3):211-216.
[3] HE Yong, LI Xiao-Li, SHAO Yong-Ni. Quantitative analysis of the varieties of apple using near infrared spectroscopy

表2 人工神经网络预测20个未知样本的杨梅品种
Table 2 The prediction results for unknown samples by BP-ANN model

Sample No.	Standard Value	Calculated value	Sample No.	Standard Value	Calculated value	Sample No.	Standard Value	Calculated value	Sample No.	Standard Value	Calculated value
(1)	1	1.04758	(6)	2	2.19956	(11)	3	2.91051	(16)	4	3.99921
(2)	1	2.41633	(7)	2	1.99140	(12)	3	3.02290	(17)	4	3.99916
(3)	1	1.00218	(8)	2	2.05164	(13)	3	2.92354	(18)	4	3.99609
(4)	1	1.01220	(9)	2	1.73715	(14)	3	2.99921	(19)	4	3.99885
(5)	1	1.04973	(10)	2	1.97255	(15)	3	2.92418	(20)	4	3.99612

Note: (1)-(5), Biqi waxberry; (6)-(10), Xiakeng waxberry; (11)-(15) Linhai Zaodamei waxberry; (16)-(20), Dongkui waxberry

- B66**,205209.
- [5] Mlayah A, Carles R, Landa G, *et al.* Raman study of longitudinal optical phonon-plasmon coupling and disorder effects in heavily Be-doped GaAs. [J]. *J. Appl. Phys.*, 1991, **69**:4064—4070.
- [6] Irmer G, Wenzel M, Monecke J. Light scattering by a multicomponent plasmon coupled with longitudinal-optical phonons; Raman spectra of p-type GaAs:Zn [J]. *Phys. Rev.*, 1997, **B56**:9524—9538.
- [7] DENG Jia-Jun, ZHAO Jian-hua, JIANG Chun-ping, *et al.* Effect of annealing on structural and magnetic properties of a thick (Ga, Mn)As layer [J]. *Chin. Phys. Lett.*, 2005, **22**:466—468.
- [8] LI Zhi-Feng, LU Wei. Raman Spectroscopy study on LO phonon-plasmon coupled mode in GaN thin films [J]. *J. Infrared Millim. Waves* (李志锋, 陆卫. 氮化镓薄膜中 LO 声子-等离子体激元耦合模拉曼光谱研究. *红外与毫米波学报*) 2003, **22**(1):8—12.
- [9] SHEN Xue-Chu. *Spectroscopy and Optical Properties of Semiconductors* [M]. Second Edition. Beijing: Scientific and Technology Press (沈学础. 半导体光谱和光学性质. 第二版. 北京: 科学出版社) 2002.
- [10] Seong M. J, Chun S H, Cheong H M, *et al.* Spectroscopic determination of hole density in the ferromagnetic semiconductor $\text{Ga}_{1-x}\text{Mn}_x\text{As}$ [J]. *Phys. Rev.*, 2002, **B66**, 033202.
- [11] ZHANG Guang-Yan, LANG Guo-Xiang, WANG Yu-Fang. *Lattice Vibration Spectroscopy* [M]. Second edition. Beijing: Higher Education Press (张光寅, 蓝国祥, 王玉芳. 晶格振动光谱学. 第二版. 北京: 高等教育出版社), 2001.
- [12] MA Bao-Shan, FAN Zai-Li, SU Fu-Hai, *et al.* Study on the Raman spectra of GaP nanorods synthesized within carbon nanotube templates [J]. *J. Infrared Millim. Waves* (马宝珊, 方再利, 苏付海, 等. 纳米碳管模板法制取的 GaP 纳米棒拉曼光谱研究. *红外与毫米波学报*) 2003, **22**(1):1—7.
- [13] JIANG De-Sheng, LIU Jian-Ping, YANG Hui. Investigations on optical properties of AlGaInN epilayers grown by MOCVD [J]. *J. Infrared Millim. Waves* (江德生, 刘建平, 杨辉. MOCVD 生长 AlGaInN 外延层的光学性质研究. *红外与毫米波学报*), 2005, **24**(3):193—197.
- [14] FENG Qian, WANG Feng-Xiang, HAO Yue. Comparative analysis of characteristics of GaN and GaN: Mg films grown by MOCVD [J]. *J. Infrared Millim. Waves* (冯倩, 王峰祥, 郝跃. MOCVD 生长 GaN 和 GaN: Mg 薄膜的对比研究. *红外与毫米波报*), 2004, **23**(3):201—204.

(上接第 194 页)

- by principal component analysis and BP model [J]. *Lecture Notes in Artificial Intelligence*, 2005, **3809**:1053—1056.
- [4] SUN Su-Qin, TANG Jun-Ming, YUAN Zi-Min, *et al.* Discrimination of trueborn tuber dioscoreae by fingerprint infrared spectra and principle component analysis [J]. *Spectroscopy and Spectral Analysis* (孙素琴, 汤俊明, 袁子民, 等. 道地山药红外指纹图谱和聚类分析的鉴别研究. *光谱学与光谱分析*), 2003, **23**(2):258—260.
- [5] YIN Qiu, SU Xiao-Zhou, XU Zhao-An, *et al.* Analysis on the ultra-spectral characteristics of water environmental parameters about lake [J]. *J. Infrared Millim. Waves* (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. *红外与毫米波学报*), 2004, **23**(6):427—435.
- [6] HE Yong, ZHANG Yun, XIANG Li-Guo. Study of application model on BP neural network optimized by fuzzy clustering [J]. *Lecture Notes in Artificial Intelligence*, 2005, **3789**:712—720.
- [7] Lin Sao-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.