第25卷第5期
2006年10月

红外与毫米波学报
J. Infrared Millim. Waves

Vol. 25, No. 5
October, 2006

文章编号:1001-9014(2006)05-0342-03

基于可见—近红外光谱技术的家蚕蚕种鉴别方法的研究

黄 敏[1], 何 勇[1], 黄凌霞[2], 楼程富[2]

(1. 浙江大学生物系统工程与食品科学学院,浙江 杭州 310029;2. 浙江大学动物科学学院,浙江 杭州 310029)

摘要:提出了一种结合主成分分析和人工神经网络技术的可见-红外光谱家蚕蚕种快速鉴别新方法.主成分分析法用于家蚕蚕种品种的聚类分析及主成分的提取.从主成分1和2对所有建模样本的得分图可以看出,主成分分析法对不同种类家蚕蚕种具有较好的聚类作用,可以定性分析家蚕蚕种品种.提取了6个能解释原始光谱的大部分信息的主成分,作为BP神经网络的输入,建立了三层BP人工神经网络模型.选取了4个典型的家蚕蚕种品种,共120个样本,其中随机选取了100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100%.说明该方法具有很好的分类和鉴别作用,为家蚕蚕种的品种鉴别提供了一种新的途径.

关 键 词:近红外光谱;蚕种;主成分分析;人工神经网络;聚类

中图分类号:S881; TH744.1 **文献标识码**:A

DISCRIMINATION OF VARIETIES OF SILKWORM EGG BASED ON VISIBLE-NEAR INFRARED SPECTRA

HUANG Min[1], HE Yong[1], HUANG Lin-Xia[2], LOU Cheng-Fu[2]

(1. College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310029, China;
2. College of Animal Science, Zhejiang University, Hangzhou 310029, China)

Abstract: A new method which was based on principal component analysis (PCA) and artificial neural network (ANN) was developed to discriminate the varieties of silkworm eggs nondestructively by visible and near infrared spectroscopy (Vis/NIRS). Principal component analysis (PCA) was used to analyze the clustering of silkworm egg samples, and offered the principal components of silkworm egg samples. The score plots of first and second components show that PCA can provide the reasonable clustering of the varieties of silkworm eggs, and can be used to analyze the silkworm eggs varieties qualitatively. The scores of the first 6 principal components computed by PCA were applied as the inputs of a back propagation neural network with one hidden layer. 100 samples from four varieties were selected randomly to build BP-ANN model, and then the model was used to predict the varieties of 20 unknown samples. The discrimination rate of 100% was achieved. It indicates that this model is reliable and practicable. So this model can offer a new approach to the fast discrimination of varieties of silkworm egg.

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

引言

家蚕约有500多个生物学和经济学形状各不相同的品种,主要可以归纳为中系、日系、欧系、热带及亚热带系统4类.由于家蚕品种的多样性,在蚕业生产中经常需要对品种进行鉴别.家蚕品种通常采取形态学特征鉴别,这要求鉴别人员有丰富的经验,全面掌握家蚕特有的形态学特征知识,但是并非所有家蚕品种均具有明显的、易于辨别的形态学特征,因而,此方法存有鉴别准确率不高的缺点,不便在蚕业生产中推广使用.所以研究一种简单、快速、非破坏的家蚕品种鉴别技术是很有必要的,本文以光谱技术为基础研究家蚕蚕种的快速无损鉴别.

现代可见-近红外光谱分析技术,可充分利用

收稿日期:2006-04-10,**修回日期**:2006-06-25 **Received date**: 2006-04-10, **revised date**: 2006-06-25

基金项目:国家自然科学基金项目(30270773);高等学校博士学科点专项科研基金资助课题(20040335034);高等学校优秀青年教师教学科研奖励计划(02411);浙江省自然科学基金人才基金资助项目(RC02067)

作者简介:黄敏,(1982-),男,浙江永康人,博士,主要研究方向:精细农业、光谱学.

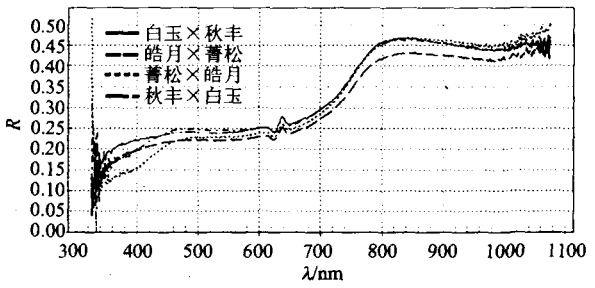

图 1　四种不同品种蚕种的近红外光谱图

Fig. 1　Near infrared reflectance spectroscopy of four different varieties of silkworm eggs

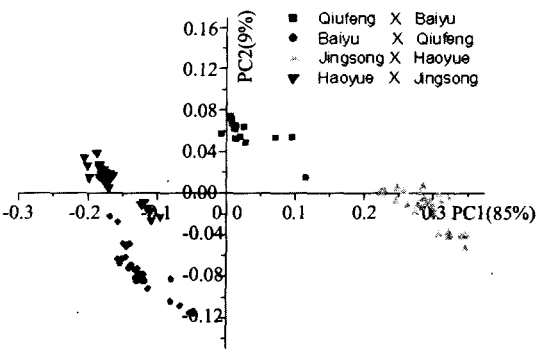

图 2　120 个家蚕蚕种的主成分 1 和主成分 2 的得分图

Fig. 2　PCA scores plots(PC1 PC2) for silkworm egg samples

多波长下的光谱数据进行定性或定量分析. 由于可见－近红外光谱技术分析具有速度快、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、石油化工、制药工业等领域[1~4].

BP 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学习算法(BP)，并且已经证明此种模型可以逼近任何连续的非线性曲线. 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. 本研究采用可见和近红外光谱技术，选用主成分分析(PCA)和基于误差反向传播算法(Back Propagation, BP)多层前馈神经网络建立不同品种家蚕蚕种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(Analytical Spectral Device)公司的 Handheld FieldSpec 光谱仪，其光谱采样间隔(波段宽)1.5nm，测定范围 325 ~ 1 075nm，扫描次数 30 次，分辨率 3.5nm，探头视场角为 20 度. 分析软件为 ASD View Spec Pro V2.14，Unscramble V9.2 和 DPS(data procession system for practical statistics).

1.2　样品来源与光谱数据采集

实验选用秋丰 × 白玉、白玉 × 秋丰、菁松 × 皓月、皓月 × 菁松等 4 种典型的家蚕蚕种样本. 每个品种各取 30 个样本. 蚕种均用直径是 120mm，高 10mm 的培养皿盛装，样品高度为 5mm. 一个培养皿的蚕种为一个样本. 全部样本随机分为建模集和预测集，建模集 100 个样本，预测集有 20 个样本. 光谱仪置于蚕种样本的上方，距蚕种表面 120mm，探头视场角为 20 度，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. 采用平均平滑法进行光谱预处理，选用平滑窗口大小为 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 2 所示，所以只取 400 ~ 1000nm 波段的光谱用于分析[5,6].

2　实验结果与分析

2.1　光谱图谱分析

四个品种蚕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，不同品种蚕种的光谱图有明显区别，并具有一定的特征性和指纹性. 去除光谱前端和末端噪声比较大的区域，选择波长范围在 400 ~ 1000nm 的光谱，应用 ASD View Spec Pro 软件，用主成分分析法对其聚类.

2.2　主成分分析对不同品种家蚕蚕种进行分析

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，并不丢失信息. 对 4 类家蚕蚕种共 120 个样本进行主成分分析聚类. 分析表明前 2 个主成分对 4 种蚕种有一定的聚类作用，能对不同品种家蚕蚕种进行定性分析.

图 2 表示 120 个建模样本的主成分 1、2 得分

表1 前6个主成分的累计可信度
Table 1 Prins and reliabilities

主成分 Principal component	PC1	PC2	PC3	PC4	PC5	PC6
累计可信度 Reliabilities	80.13%	88.30%	93.14%	95.21%	97.13%	99.67%

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中秋丰×白玉(Qiufeng × Baiyu)、白玉×秋丰(Baiyu × Qiufeng)、菁松×皓月(Jingsong × Haoyue)、皓月×菁松(Haoyue × Jingsong)4个蚕种明显分成4类,说明主成分1、2对四种家蚕蚕种有较好的聚类作用.从图2可以看出,品种为菁松×皓月的30个样本聚合度较好,紧密分布在Y轴右方并聚合于x轴的附近,而其他90个样本大部分在Y轴的左方.品种为白玉×秋丰的30个样本聚合度较好,紧密的分布在图2的第三象限,其它样本大部分处于第三象限之外.品种为秋丰×白玉和皓月×菁松的家蚕蚕种样本聚合度不如另外两种,但这两个品种的家蚕蚕种也没有和其他两个品种重叠,能够相对区分开来.

2.3 基于神经网络定量分析建立家蚕品种鉴别模型

波段从325~1075nm共有750个点,但是,采用全谱段计算时,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.所以通过主成分分析,选取对于家蚕蚕种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.累计可信度表示主成分对原始变量的解释程度.前6个主成分的累计可信度已达99.67%,表示这6个主成分能够解释原始波长变量的99.67%.主成分分析是一种非常有效的数据挖掘方法,它把原来的600个波长变量压缩成了彼此正交的6个新变量,这是6个新变量彼此间是互不影响的,而且能代表绝大部分原变量包含的信息.

把这6个主成分作为BP神经网络的输入变量建立鉴别模型,通过调整隐含层的节点数来优化网络结构[7].因而确定网络输入层节点数为6,经多次实验确定最佳隐含层节点数为6,输出层节点数为1(品种值).网络设定训练迭代次数为1000次.对输入样本进行标准化处理.对100个建模样本的拟合残差为9.96×10^{-5},对未知的20个样本进行预测,预测结果如表2所示.20个样本的预测相对偏差均在5%以下,调整后未知样本的品种识别正确率达到了100%.

3 结论

应用主成分分析方法和BP神经网络相结合,对实验中采集的4个家蚕蚕种的可见-近红外光谱数据进行了处理,建立了家蚕蚕种品种鉴别的模型,该模型的预测效果好,对未知样品的预测相对误差均在5%以下,品种识别率达到100%.说明运用可见-近红外光谱技术可以快速、准确的对家蚕蚕种品种进行鉴别.本文采用主成分分析方法,对原始的大量光谱数据进行降维处理,提取了6个能很好反映家蚕蚕种性状的主成分,作为BP神经网络的输入,不但减少了神经网络的计算量,加快了训练速率,而且因为去除了光谱干扰信息,也提高了预测的正确率.因此,用主成分分析方法结合BP神经网络的模式识别和光谱技术研究家蚕蚕种品种鉴别是可行的,为家蚕品种的快速无损检测提供了一种新的方法.

(下转359页)

表2 BP神经网络模型对未知样本的预测结果
Table 2 Prediction results for unknown samples by BP model

样本号	预测值	品种值	相对偏差(%)	样本号	预测值	品种值	相对偏差(%)
1	1.0163	1	1.630	11	2.9076	3	-3.080
2	1.0017	1	0.170	12	2.9671	3	-1.096
3	1.0193	1	1.930	13	3.0034	3	0.113
4	1.0285	1	2.850	14	2.9913	3	-0.290
5	1.0199	1	1.990	15	3.0052	3	0.173
6	1.9625	2	-1.875	16	3.9886	4	-0.285
7	1.9928	2	-0.360	17	3.9759	4	-0.602
8	1.9807	2	-0.965	18	3.9889	4	-0.277
9	2.0884	2	4.420	19	3.9855	4	-0.362
10	2.0254	2	1.270	20	3.9948	4	-0.130

Note: 品种值1-秋丰×白玉;2-白玉×秋丰;3-菁松×皓月;4-皓月×菁松

的关系好于荧光峰高度（$R_{max\ red}/R_{580}$ 和 $R_{max\ red}/R_{575}$），这是由于查干湖悬浮物较高，不同波段光谱对悬浮物响应不同造成的.

REFERENCES

[1] 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—430.

[2] Duan H T, Zhang B, Song K S *et al.* Hyperspectral monitoring model of eutrophication in Lake Nanhu, Changchun [J]. *Journal of Lake Science* (段洪涛,张柏,宋开山,等. 长春市南湖富营养化程度高光谱遥感监测模型研究. **湖泊科学**), 2005,**17**(3):282—288.

[3] YIN Qiu, GONG Cai-Lan, KUANG Ding-Bo, *et al.* Method of satellite remote sensing of lake water quality and its applica-tions [J]. *J. Infrared Millim. Waves* (尹球,巩彩兰,匡定波,等. 湖泊水质卫星遥感方法及其应用. **红外与毫米波学报**), 2005,**24**(3):198—202.

[4] Koponen S, Pulliainen J, Kallio K, *et al.* Lake water quality classification with airborne hyperspectral spectrometer and simulated MERIS data [J]. *Remote Sensing of Environment*, 2002,**79**(1):51—59.

[5] Gitelson A. The peak near 700 nm on reflectance spectra of algae and water: relationships of its magnitude and position with chlorophyll concentration[J], *Int. J. Remote Sensing*, 1992,**13**(17): 3367—3373.

[6] Neville R A, Gower J F R. Passive remote sensing of phytoplankton via chlorophyll-a fluorescence [J]. *Journal of Geophysical Research*, 1977,**82**:3487—3493.

[7] Gower J F R. Observations of in situ fluorescence of chlorophyll-a in Saavich Inlet [J]. *Bouridary-Layer Meteorology*, 1980,**18**:235—248.

[8] Zhao D Z, Zhang F S, Du F, *et al.* Interpretation of sun-induced fluorescence peak of chlorophyll a on reflectance spectrum of algal waters [J]. *Journal of Remote Sensing* (赵冬至,张丰收,杜飞,等. 不同藻类水体太阳激发的叶绿素荧光峰(SICF)特性研究. **遥感学报**), 2005,**9**(3):265—270.

(上接 344 页)

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] 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.

[3] JIA Dong-Yao, DING Tian-Hua. Novel method of detecting foreign fibers in lint by fiber's infrared absorption characteristic [J]. *J. Infrared Millim. Waves* (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. **红外与毫米波学报**),2005, **24**(2): 147 - 150.

[4] HE Yong, FENG Shui-Juan, DENG Xun-Fei. *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] QI Xiao-Ming, ZHANG Lu-Da, DU Xiao-Lin, *et al.* Quantitative analysis using NIR by building PLS-BP model[J]. *Spectroscopy and Spectral Analysis* (齐小明,张录达,杜晓林等. PLS—BP 法近红外光谱定量分析研究. **光谱学与光谱分析**), 2003,**23**(5): 870—872.

[6] 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.

[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.