

霉变板栗的近红外光谱和神经网络方法判别

周竹¹ 刘洁¹ 李小昱¹ 李培武² 王为¹ 展慧¹

(1. 华中农业大学工程技术学院, 武汉 430070; 2. 中国农业科学院油料作物研究所, 武汉 430062)

【摘要】 利用近红外光谱检测了带壳板栗的品质。在波数为 12 000~4 000 cm^{-1} 范围内采用近红外漫反射法采集了合格板栗和霉变板栗的光谱, 用 6 种光谱预处理方法分析数据, 比较了板栗近红外光谱在不同预处理方法下所建模型的识别率。试验结果表明经矢量归一化预处理所建模型识别效果最好, 对预测集中的合格板栗、表面霉变板栗、内部霉变板栗的预测正确率分别为 94.74%、94.44%、92.31%。

关键词: 板栗 近红外光谱 BP 神经网络 预处理 主成分提取

中图分类号: O657.33; S123 **文献标识码:** A

Discrimination of Moldy Chinese Chestnut Based on Artificial Neural Network and Near Infrared Spectra

Zhou Zhu¹ Liu Jie¹ Li Xiaoyu¹ Li Peiwu² Wang Wei¹ Zhan Hui¹

(1. College of Engineering & Technology, Huazhong Agricultural University, Wuhan 430070, China

2. Oil Crops Research Institute, Chinese Academy of Agricultural Sciences, Wuhan 430062, China)

Abstract

The nondestructive discrimination of the shelled chestnuts was studied with near infrared spectra, which could provide a new method for quality detection of other shelled agricultural products. 178 chestnut samples were prepared and the diffuse spectral reflectance of the samples were collected in the wave number range of 12 000~4 000 cm^{-1} . First, six preprocessing methods including smooth, vector normalization, min-max normalization, standard normal variate transformation, multiplication scattering correction and first derivative were used to improve the original spectrum. Then, principal component analysis was applied to compress thousands of spectral data into several variables and to collect spectral information. The principal components extracted by PCA were employed as the inputs of the BP neural networks. Effects of the six preprocessing methods on the models based on BP neural network were compared. The results show that prediction precision varied to different preprocessing methods. The optimum network structure of 7-4-1 was obtained after vector normalization method. Discriminating rate of qualified chestnut, surface moldy chestnut and internal moldy chestnut in prediction reached 94.74%, 94.44% and 92.31%, respectively.

Key words Chinese chestnut, Near infrared spectra, BP neural network, Preprocessing, Principal component analysis

引言

板栗营养丰富, 采收后极易霉烂变质, 食用不当会引发安全问题。目前, 剔除霉变板栗主要采用人

工分选或盐水浮选, 分选效率低、错误率高。研究一种快速、有效且适用的霉变板栗检测方法, 不仅能及时分检霉变板栗, 而且为板栗综合品质的快速无损检测和分级打下基础。

收稿日期: 2009-07-10 修回日期: 2009-07-29

作者简介: 周竹, 博士生, 主要从事智能化检测与控制技术研究, E-mail: zhouzhugexy@webmail.hzau.edu.cn

通讯作者: 李小昱, 教授, 博士生导师, 主要从事智能化检测与控制技术研究, E-mail: lixiaoyu@mail.hzau.edu.cn

近红外光谱技术是利用全谱或部分波段的光谱数据进行定性或定量分析的无损检测技术,已广泛应用于农业、食品、石油化工等领域。目前,国外学者利用近红外光谱技术开展了开心果、无花果等的缺陷识别、花生霉变识别等研究^[1~3],然而研究的物料均未带壳。国内鲜有对带壳坚果品质的近红外光谱鉴别研究,更未见应用近红外光谱进行带壳霉变板栗识别的研究。利用近红外光谱识别霉变板栗,因板栗外壳而增加识别的难度,需要选择合适的光谱预处理方法和建模方法。

BP神经网络(back-propagation neural network)能够实现输入与输出之间的高度非线性映射,是目前应用较多的一种目标分类器。Levenberg-Marquardt(简称L-M)优化算法稳定性好,收敛速度快。主成分分析在保留主要光谱信息的前提下选择较少的新变量来代替较多的旧变量,能够解决由于谱带的重叠而无法分析的困难^[4]。本研究利用近红外光谱技术结合主成分分析和BP神经网络等方法,建立识别带壳霉变板栗的快速检测方法。

1 材料与方法

1.1 试验材料及设备

试验板栗产于湖北麻城,单个栗重范围8.50~20.41 g,购买后按中国商业行业标准SB/T 10192—1993相关规定贮藏。样品包括两类:合格板栗和霉变板栗,其判别依GH/T 1029—2002规定。霉变板栗以表面霉变和内部霉变加以区分。内部霉变板栗通过光谱采集后去壳确定。

试验设备包括德国Bruker公司FT-2 NIR近红外光谱仪(VECTOR 33型),配备镀金积分球、样品旋转器、12 mm石英杯、PbS检测器和OPUS分析软件。

1.2 板栗光谱采集

光谱采集方式为漫反射。光谱测定条件为:光谱扫描范围12 000~4 000 cm^{-1} ,镀金漫反射体作背景,分辨率为8 cm^{-1} ,扫描次数为64。板栗光谱采集时,将完整带壳板栗果实平置于光源上方并完全遮蔽光源。每个板栗及其对应光谱视为一个样本,每周定期测量30~40个样本,共测量6周,最终共获得178个样本,其中合格板栗69个,霉变板栗109个。

1.3 基于L-M算法的神经网络模型

L-M算法是一种利用标准的数值优化技术的快速算法,它弥补了标准BP算法及其众多改进算法的不足,在高斯牛顿法和梯度下降法之间进行平滑调和,能够在远离最小值时逐渐切换到高斯牛顿

法。既具有高斯牛顿法的局部收敛性,又有梯度下降法的全局特性^[5]。由于L-M算法利用了近似的二阶导数信息,它比梯度法快得多。基于L-M算法的BP神经网络已被广泛应用^[6~8],其原理可参考文献^[9]。

本研究构建的基于L-M算法的三层BP神经网络模型,根据输入层和输出层的特点,隐含层的传递函数选用双曲正切S型传递函数tansig,输出层选用logsig函数。网络输出层节点数为1,目标误差为0.01,学习速率为0.01,训练迭代次数为1 000次,其余参数均为默认值。程序在Matlab 7.0上调试完成。

2 结果与分析

2.1 板栗近红外光谱

板栗样本原始光谱如图1所示,横坐标为波数,单位为 cm^{-1} ,光谱数据范围从12 000~4 000 cm^{-1} ,纵坐标为吸光度。从图中可以看出合格板栗和霉变板栗光谱相互重叠,很难直接从形态上加以区分。分别使用平滑9点(smooth)、矢量归一化(vector normalization,简称VN)、最大最小归一化(min-max normalization,简称MMN)、标准正态变量变换(standard normal variate transformation,简称SNV)、多元散射校正(multiplication scattering correction,简称MSC)、一阶导17点(first derivative,简称FD)等6种方法对原始光谱进行预处理。由于系统和样本误差,板栗光谱曲线在高频段噪声较大,经过比较最终选取10 000~4 000 cm^{-1} 范围光谱用于分析。

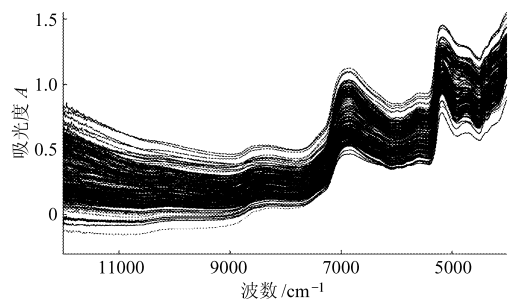


图1 板栗样本原始光谱

Fig.1 Original spectrum of Chinese chestnut samples

2.2 光谱主成分提取

利用BP神经网络对目标进行分类时,若前端输入太多的样本特征量,则会降低网络的训练速度与效率,严重时会导致网络不收敛^[10]。本研究用于分析的板栗近红外光谱范围为10 000~4 000 cm^{-1} ,光谱数据点数为1 556个,构成了178(样本总数)×1 556的光谱矩阵,数据量大,冗余信息多。利用Matlab 7.0对经不同预处理后的光谱提取主成分,

其累积贡献率能够反映前 n 个主成分对原数据信息的表征能力。表 1 列出了经过计算后,不同预处理方法保留的主成分累积贡献率,它们均能表征原

始数据的信息,从而有效地降低了板栗光谱矩阵的数据量。

表 1 不同预处理方法主成分分析累积贡献率

Tab.1 Principal component analysis accumulative contribution rate of different preprocessing techniques %

预处理方法	主成分									
	1	2	3	4	5	6	7	8	9	10
smooth	88.83	98.74	99.39	99.81	—	—	—	—	—	—
SNV	49.22	88.06	95.82	97.39	98.40	—	—	—	—	—
MSC	49.25	88.32	96.08	97.60	98.58	—	—	—	—	—
MMN	57.29	84.51	96.62	98.03	99.33	—	—	—	—	—
VN	51.14	85.33	93.01	97.23	98.62	99.08	99.47	—	—	—
FD	54.32	72.14	81.88	87.49	91.35	93.21	99.42	95.10	95.62	96.09

注:—为未选

2.3 霉变板栗 BP 神经网络识别模型的建立

将样本随机分为建模集和预测集:建模集中共有样本 128 个,其中合格板栗、表面霉变板栗和内部霉变板栗分别为 50 个、48 个和 30 个;预测集中共有样本 50 个,合格板栗、表面霉变板栗和内部霉变板栗样本个数分别为 19 个、18 个和 13 个。

将不同预处理后的主成分分别作为 BP 神经网络的输入建立识别模型,用数字 0、1 来分别表示合格板栗、霉变板栗。设定偏差阈值为 ± 0.1 ,若真值与预测值之差在 ± 0.1 范围内,则判定识别正确,反之识别错误。隐含层神经元节点数用“试凑法”确定,通过调整隐含层的节点数来优化网络结构,以识别率最高为评价指标,经过反复试验得到 6 种预处理方式下的最优网络结构的隐含层节点数如表 2 所示。

表 2 不同预处理后所建神经网络模型隐含层节点数

Tab.2 Artificial neural network model hidden layer nodes of different preprocessing techniques

预处理方法	smooth	SNV	MSC	MMN	VN	FD
隐含层节点数	14	4	10	10	4	4

对建模集中 128 个样本进行学习,经过 1 000 次迭代之后,拟合残差已满足要求。使用预测集样本进行预测,预测结果如表 3 所示。从表 3 中可以看出,使用平滑 9 点(smooth)、标准正态变量变换(SNV)、多元散射校正(MSC)、最大最小归一化(MMN)、矢量归一化(VN)、一阶导 17 点(FD)分别预处理所建模型对合格板栗、内部霉变板栗的识别率差异较大。经矢量归一化(VN)预处理所建模型对合格板栗、内部霉变的识别率最高,识别正确率分别达到 94.74%、92.31%。经最大最小归一化(MMN)预

处理所建模型对合格板栗识别正确率最低,为 68.42%。使用平滑 9 点(smooth)预处理所建模型对内部霉变板栗识别正确率最低,仅为 53.85%。6 种预处理方法所建模型对表面霉变板栗识别准确率相差不大,误判个数控制在 1 个左右,识别正确率能达到 94.44% 以上。

表 3 不同预处理方法下所建模型识别正确率比较

Tab.3 Models identification correct rates of different preprocessing techniques %

预处理方法	预测集			建模集		
	合格板栗	霉变板栗	内部霉变	合格板栗	霉变板栗	内部霉变
	表面霉变	内部霉变	表面霉变	内部霉变	表面霉变	内部霉变
smooth	89.47	94.44	53.85	96.00	100	96.67
SNV	89.47	94.44	76.92	98.00	97.92	93.33
MSC	78.95	94.44	84.62	98.00	100	100
MMN	68.42	100	76.92	98.00	100	100
VN	94.74	94.44	92.31	98.00	100	100
FD	84.21	94.44	84.62	98.00	100	100

综上所述,比较 6 种预处理方法所建 BP 神经网络识别模型,经矢量归一化(VN)预处理的识别模型效果最好。该模型对预测集中合格板栗识别正确率达到 94.74%,对表面霉变板栗、内部霉变板栗误判均为 1 个,则霉变板栗识别的正确率分别为 94.44%、92.31%。

3 结束语

针对带壳霉变板栗不易识别的特点,利用近红外漫反射技术和人工神经网络方法无损识别霉变板栗,比较了平滑 9 点(smooth)、矢量归一化(VN)、最大最小归一化(MMN)、标准正态变量变换(SNV)、

多元散射校正(MSC)、一阶导17点(FD)等6种预处理方法所建BP神经网络模型对板栗识别的影响。不同光谱预处理方法所建模型对霉变板栗识别的正确率差异较大,经矢量归一化(VN)预处理建立的7-4-1网络结构的BP模型识别效果最好,对预

测集中合格板栗、表面霉变板栗、内部霉变板栗的预测正确率分别为94.74%、94.44%、92.31%。本项研究提出的利用近红外光谱技术识别霉变板栗,为板栗综合品质的快速无损检测和分选打下了基础,也为其他坚果物料的研究提供了新的思路。

参 考 文 献

- Haff R P, Pearson T. Spectral band selection for optical sorting of pistachio nut defects[J]. Transactions of the ASABE, 2006, 49(4): 1105~1113.
- Burks C S, Dowell F E, Xie F. Measuring fig quality using near-infrared spectroscopy[J]. Journal of Stored Products Research, 2000, 36(3): 289~296.
- Hirano S, Okawara N, Narazaki S. Near infra red detection of internally moldy nuts[J]. Biosci. Biotechnol. Biochem., 1998, 62(1): 102~107.
- 何勇, 李晓丽, 邵永妮. 基于主成分分析和神经网络的近红外光谱苹果种鉴别方法研究[J]. 光谱学与光谱分析, 2006, 26(5): 850~853.
He Yong, Li Xiaoli, Shao Yongni. Discrimination of varieties of apple using near infrared spectra based on principal component analysis and artificial neural network model[J]. Spectra Spectroscopy and Spectral Analysis, 2006, 26(5): 850~853. (in Chinese)
- 曹小平, 胡昌华, 郑志强, 等. 基于LMBP神经网络的惯性器件故障预报方法研究[J]. 电光与控制, 2005, 12(6): 38~41.
Cao Xiaoping, Hu Changhua, Zheng Zhiqiang, et al. Fault prediction for inertial device based on LMBP neural network[J]. Electronics Optics & Control, 2005, 12(6): 38~41. (in Chinese)
- 赖树雨, 李向明, 段涛. 基于Gabor变换和LMBP神经网络的车牌汉字字符识别[J]. 湖南工业大学学报, 2008, 22(5): 94~96.
Lai Shuyu, Li Xiangming, Duan Tao. Recognition of license plate Chinese character based on Gabor filters and BP neural networks[J]. Journal of Hunan University of Technology, 2008, 22(5): 94~96. (in Chinese)
- 陈建, 陈晓, 李伟, 等. 基于近红外光谱技术和人工神经网络的玉米品种鉴别方法研究[J]. 光谱学与光谱分析, 2008, 28(8): 1806~1809.
Chen Jian, Chen Xiao, Li Wei, et al. Study on discrimination of corn seed based on near-infrared spectra and artificial neural network model[J]. Spectra Spectroscopy and Spectral Analysis, 2008, 28(8): 1806~1809. (in Chinese)
- 王博雷, 马孝义, 郝晶晶. 基于L-M优化算法的喷头射程神经网络预测模型[J]. 农业机械学报, 2008, 39(5): 35~40.
Wang Bolei, Ma Xiaoyi, Hao Jingjing. Prediction of sprinkler nozzle range based on L-M optimized algorithm[J]. Transactions of the Chinese Society for Agricultural Machinery, 2008, 39(5): 35~40. (in Chinese)
- 王■松, 褚福磊, 何永勇, 等. BP神经网络快速收敛算法研究[J]. 农业机械学报, 2004, 35(6): 182~184.
Wang Yunsong, Chu Fulei, He Yongyong, et al. New and rapidly converging BP neural network algorithm[J]. Transactions of the Chinese Society for Agricultural Machinery, 2004, 35(6): 182~184. (in Chinese)
- 李军梅, 胡以华, 陶小红. 基于主成分分析与BP神经网络的识别方法研究[J]. 红外与激光工程, 2005, 34(6): 719~723.
Li Junmei, Hu Yihua, Tao Xiaohong. Recognition method based on principal component analysis and back-propagation neural network[J]. Infrared and Laser Engineering, 2005, 34(6): 719~723. (in Chinese)