

基于主成分分析的苹果霉心病近红外漫反射光谱判别*

李顺峰 张丽华 刘兴华 李光辉

(西北农林科技大学食品科学与工程学院, 陕西杨凌 712100)

【摘要】 为了探讨利用近红外漫反射光谱技术判别苹果霉心病的可行性,将健康苹果和霉心病苹果的近红外光谱经不同光谱预处理方法处理,将主成分分析提取的主成分作为自变量,对苹果霉心病进行了判别研究。结果表明,光谱经矢量归一化处理后提取到的20个主成分建立的Fisher判别函数判别率最高,性能稳定,建模集正确判别率为89.9%,对检验集正确判别率为87.8%。

关键词: 苹果 霉心病 近红外漫反射光谱 主成分分析

中图分类号: S123; O657.33 **文献标识码:** A **文章编号:** 1000-1298(2011)10-0158-04

Discriminant Analysis of Apple Moldy Core Using Near Infrared Diffuse Reflectance Spectroscopy Based on Principal Component Analysis

Li Shunfeng Zhang Lihua Liu Xinghua Li Guanghui

(College of Food Science and Engineering, Northwest A & F University, Yangling, Shaanxi 712100, China)

Abstract

In order to evaluate the ability of near infrared spectroscopy (NIRS) to detect moldy core in apple fruits, the discriminant efficiency of Fisher function based on several spectra preprocessing methods and principal component analysis (PCA) were investigated. Results showed that the discriminant efficiency of Fisher function established by 20 principal components (PCs) after vector normalization preprocess was the highest, discriminant efficiency for training sample set and test set were 89.9% and 87.8%, respectively.

Key words Apple, Moldy core, Near infrared spectroscopy, Principal component analysis

引言

苹果霉心病又称心腐病,主要危害苹果果实,果实受害从心室开始,逐渐向外扩展霉烂。由于霉心病发生在果实内部(心室周围),如果不把果实切开,很难从外表看出该果实是否被霉心病所侵染,它不仅影响感官品质、降低果实商品价格,而且也可能对食用者造成危害;如果病害果流入消费者手中,对果商信誉、销售市场都会产生不利影响。因此,建立一种无损、可靠的方法来检测苹果霉心病,在现实生产中很有必要。

近红外光谱分析技术是近年来发展起来的一种高效、快速、不破坏或少破坏样品、无公害、多组分同

时分析的现代分析技术,已成功应用于石油、食品、医药等领域^[1-2]。目前,国内外学者运用近红外光谱技术对果实内部病害检测进行了大量研究^[3-10]。但大都采用近红外透射技术,采用近红外漫反射技术对果实内部病害检测的研究较少。

本文将近红外无损检测技术与主成分分析法相结合,对健康苹果和霉心病苹果进行判别研究,并探讨不同光谱预处理方法对判别模型的影响,为近红外漫反射光谱技术判别霉心病苹果提供参考。

1 材料与方法

1.1 材料

供试苹果品种为红富士,于2009年10月15日

收稿日期:2011-04-27 修回日期:2011-05-31

*“十一五”国家科技支撑计划资助项目(2006BAK02A24)

作者简介:李顺峰,博士生,主要从事果蔬保鲜和加工技术研究,E-mail:lishunfeng2000@163.com

通讯作者:刘兴华,教授,博士生导师,主要从事果蔬贮藏及加工技术研究,E-mail:liuxh2830@163.com

采自陕西省扶风县。挑选外观无机械损伤、果径为 7~9 cm 的果实作为试验材料,采集在 $(0 \pm 1)^\circ\text{C}$ 下贮藏时间少于 4 个月的共 367 个苹果的近红外光谱。

1.2 近红外光谱采集

采用傅里叶变换近红外光谱仪(MPA 型, Bruker Optics 公司)采集健康苹果和霉心病苹果的近红外光谱。在每个果实赤道部位平均取 3 个点并标记为光谱采集点,扫描模式为漫反射,检测器为 InGaAs,检测波数范围 $4\,000 \sim 12\,000\text{ cm}^{-1}$,分辨率 8 cm^{-1} ,扫描次数为 64 次,平均每个果实 3 个点的近红外光谱即为每个样本的光谱。光谱采集室温度控制在 25°C 左右。

近红外光谱采集完毕的苹果,全部沿果实赤道处切开,以判断该果实是健康苹果还是霉心病苹果。

1.3 分析方法

1.3.1 主成分分析

主成分分析(principal component analysis, 简称 PCA)是将多个变量通过线性变换以选出较少个数重要变量的一种多元统计方法,又称主分量分析。其基本方法就是将原来众多具有一定相关性的变量,重新组合成一组新的互相无关的综合指标来代替原来的指标;主成分分析的目的就是运用线性变换,将原来的多个指标组合成相互独立的少数几个能充分反映母体信息的指标,从而在不丢掉主要信息的前提下,避开变量之间共线性的问题,便于进一步分析。

1.3.2 Fisher 判别分析

Fisher 判别亦称典型判别,其基本方法就是采用投影,使得每一类别内的方差尽可能小,而不同类别之间的方差尽可能大,即针对 P 维空间中的某点寻找一个能使它降为一维数值的线性函数,然后应用这个线性函数把 P 维空间中的已知类别总体以及未知类别归属的样本都变换为一维数据,再根据其间的亲疏程度判定未知类别的归属。

2 结果与分析

2.1 近红外光谱分析

从图 1 中可以看出,健康苹果和霉心病苹果的近红外光谱曲线形状非常相似,并且健康苹果和霉心病苹果的光谱交叉重合,因此很难从光谱曲线上直接区分某苹果是否为霉心病病果。由于仪器、样本背景、环境条件及其他因素的影响,近红外光谱常出现噪声、谱图基线平移和漂移等现象,为了消除这些不利因素对所建模型的影响,分别采用多元散射校正(MSC)、标准正态变量变换(SNV)、矢量归一

化、最小最大归一化、9 点平滑和一阶导数(9 点平滑)6 种方法对原始光谱进行预处理。

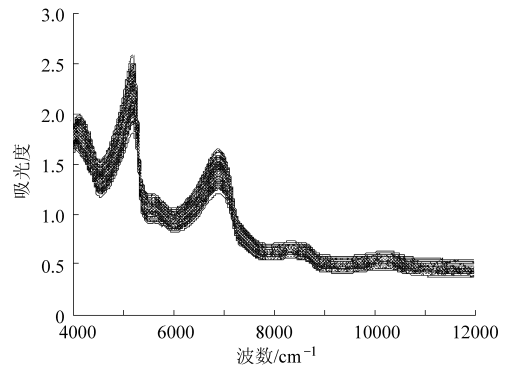


图 1 健康苹果和霉心病苹果的近红外光谱

Fig. 1 NIR spectra of healthy and moldy core apple fruits

2.2 光谱主成分提取

本试验采集的苹果近红外光谱范围为 $4\,000 \sim 12\,000\text{ cm}^{-1}$,共有 2 075 个数据点,构成了 $367 \times 2\,075$ 的光谱矩阵,数据量大,冗余信息多,计算量大。利用 Matlab 7.10 对经不同光谱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,不同预处理后提取的光谱主成分累积贡献率如图 2 所示。从图 2 可以看出,经不同光谱预处理后提取得到的前 20 个主成分除了一阶导数处理仅可以解释 86% 以上的光谱信息外,经其他光谱预处理后的前 20 个主成分均可以解释 99% 以上的光谱信息。由于健康苹果和霉心病苹果的近红外光谱差异很小,若选取主成分数过少可能会造成大的判别误差。为了获得较高的正确判别率并减少计算量,选取前 20 个主成分作为建立 Fisher 判别函数的分析对象。这样光谱数据阵就从原始光谱的 $367 \times 2\,075$ 减少到 367×20 (20 个主成分)。

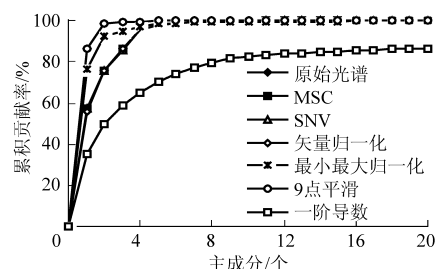


图 2 不同光谱预处理方法提取的前 20 个主成分的累积贡献率

Fig. 2 Cumulative variance plot of the first 20 principal components extracted after different spectra preprocessings

2.3 Fisher 判别函数建立

以数字 1 和 2 分别代表健康苹果和霉心病苹果,分别从健康苹果和霉心病苹果中随机抽取 60 个和 30 个共计 90 个苹果作为检验集,其余 277 个苹

果作为建模集。取经不同光谱预处理后提取的前20个主成分,利用 SPSS 19.0 软件进行 Fisher 判别

分析。对于建模集,不同光谱预处理方法建立的 Fisher 函数判别结果如表 1 所示。

表 1 不同光谱预处理方法结合主成分分析建立的 Fisher 函数判别结果

Tab.1 Discriminant results on training samples of Fisher's function established by combination of different spectra preprocessings with PCA

光谱预处理方法	健康苹果			霉心病苹果			建模集正确判别率/%
	样本数/个	误判数/个	正确判别率/%	样本数/个	误判数/个	正确判别率/%	
原始光谱	190	35	81.6	87	16	81.6	81.6
MSC	190	24	87.4	87	13	85.1	86.6
SNV	190	25	86.8	87	13	85.1	86.3
矢量归一化	190	18	90.5	87	10	88.5	89.9
最小最大归一化	190	48	74.7	87	26	70.1	73.3
9点平滑	190	38	80.0	87	15	82.8	80.9
一阶导数	190	36	81.0	87	14	83.9	81.9

由表 1 可知,原始光谱经不同光谱预处理后建立的判别函数对健康苹果和霉心病苹果的判别效果差异较大。对霉心病苹果原始光谱提取主成分后所建判别函数正确判别率为 81.6%。经矢量归一化预处理所建判别函数的正确判别率最高,达到 89.9%;经最小最大归一化预处理所建判别函数的正确判别率最低,仅为 73.3%;经 MSC、SNV、9 点平滑、一阶导数(9 点平滑)处理所建判别函数正确判别率均低于矢量归一化。

经矢量归一化预处理结合主成分分析所建 Fisher 判别方程为:

健康苹果

$$F_1 = -1.174 - 0.017P_1 - 0.003P_2 - 0.023P_4 + 0.007P_5 - 0.069P_6 - 0.089P_7 - 0.004P_8 - 0.035P_9 + 0.071P_{10} + 0.267P_{11} + 0.751P_{12} + 0.278P_{13} + 0.129P_{14} + 0.055P_{15} - 0.529P_{16} - 0.039P_{17} - 0.254P_{18} - 1.151P_{19} - 1.437P_{20}$$

霉心病苹果

$$F_2 = -1.266 + 0.021P_1 + 0.004P_2 - 0.001P_3 + 0.019P_4 + 0.001P_5 + 0.088P_6 + 0.119P_7 + 0.038P_8 + 0.048P_9 - 0.115P_{10} - 0.270P_{11} - 0.762P_{12} - 0.248P_{13} - 0.167P_{14} + 0.007P_{15} + 0.588P_{16} - 0.030P_{17} + 0.430P_{18} + 1.1P_{19} + 1.382P_{20}$$

式中 P_i ——第 i 个主成分, $i=1,2,\dots,20$

为了检验所建判别函数对外部未知样本的判别效果,将 90 个未参与建模的样本,也即检验集,代入上述 Fisher 判别方程进行验证,比较 F_1 和 F_2 的计算结果,若 $F_1 > F_2$,则被判为组 1(健康苹果);若 $F_1 < F_2$,则被判为组 2(霉心病苹果)。所建判别函数对

检验集判别结果如表 2 所示。表中结果显示,有 6 个健康苹果被错误地判为霉心病苹果,误判率为 10.0%;5 个霉心病苹果被错误判别为健康苹果,误判率为 16.7%;其余的 79 个样本均被正确判别,正确判别率为 87.8%,略低于 Shenderay 等^[10]采用近红外透射模式对苹果霉心病的正确判别率 90.4%(对健康果实、发病面积为 0%~30%、发病面积在 30% 以上的其实正确判别率分别为 92%、86.67%、100%),说明了主成分分析结合 Fisher 判别应用于健康苹果和霉心病苹果判别是可行的。

表 2 矢量归一化结合主成分分析所建 Fisher 判别函数对检验集的判别结果

Tab.2 Discriminant results on test samples of Fisher function established by PCA after vector normalization preprocessing

项目	类别	合计	判别结果	
			健康苹果	霉心病苹果
计数/个	健康苹果	60	54	6
	霉心病苹果	30	5	25
百分率/%	健康苹果	100	90.0	10.0
	霉心病苹果	100	16.7	83.3
正确判别率/%			87.8	

3 结束语

比较了原始光谱、MSC、SNV、矢量归一化、最小最大归一化、9 点平滑、一阶导数(9 点平滑)提取主成分后建立的判别函数对健康苹果和霉心病苹果的判别效果,矢量归一化预处理后所建判别函数的正确判别率最高,经最小最大归一化预处理所建判别函数的正确判别率最低。矢量归一化预处理后提取

主成分所建健康苹果和霉心病苹果的 Fisher 判别函数检验集正确判别率为 87.8%, 验证了近红外漫反射光谱技术应用于健康苹果和霉心病苹果检测的可行性。建模集正确判别率为 89.9%, 对

参 考 文 献

- 1 Fan Q, Wang Y, Sun P, et al. Discrimination of Ephedra plants with diffuse reflectance FT-NIRS and multivariate analysis [J]. *Talanta*, 2010, 80(3):1 245 ~ 1 250.
- 2 李亮, 丁武. 掺有植物性填充物牛奶的近红外光谱判别分析[J]. *光谱学与光谱分析*, 2010, 30(5):1 238 ~ 1 242.
Li Liang, Ding Wu. Discriminant analysis of raw milk adulterated with botanical filling material using near infrared spectroscopy [J]. *Spectroscopy and Spectral Analysis*, 2010, 30(5):1 238 ~ 1 242. (in Chinese)
- 3 Clark C J, McGlone V A, Jordan R B. Detection of brownheart in "Braeburn" apple by transmission NIR spectroscopy [J]. *Postharvest Biology and Technology*, 2003, 28(1):87 ~ 96.
- 4 Upchurch B L, Throop J A, Aneshansley D J. Detecting internal breakdown in apples using interreflectance measurements [J]. *Postharvest Biology and Technology*, 1997, 10(1):15 ~ 19.
- 5 Han D, Tu R, Lu C, et al. Nondestructive detection of brown core in the Chinese pear 'Yali' by transmission visible-NIR spectroscopy [J]. *Food Control*, 2006, 17(8):604 ~ 608.
- 6 周竹, 刘洁, 李小昱, 等. 霉变板栗的近红外光谱和神经网络方法判别[J]. *农业机械学报*, 2009, 40(增刊):109 ~ 112.
Zhou Zhu, Liu Jie, Li Xiaoyu, et al. Discrimination of moldy Chinese chestnut based on artificial neural network and near infrared spectra [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2009, 40(Supp.):109 ~ 112. (in Chinese)
- 7 展慧, 李小昱, 周竹, 等. 基于近红外光谱和机器视觉融合技术的板栗缺陷检测[J]. *农业工程学报*, 2011, 27(2):345 ~ 349.
Zhan Hui, Li Xiaoyu, Zhou Zhu, et al. Detection of chestnut defect based on data fusion of near-infrared spectroscopy and machine vision [J]. *Transactions of the CSAE*, 2011, 27(2):345 ~ 349. (in Chinese)
- 8 韩东海, 刘新鑫, 鲁超, 等. 苹果内部褐变的无损检测研究[J]. *农业机械学报*, 2006, 37(6):86 ~ 88, 93.
Han Donghai, Liu Xinxin, Lu Chao, et al. Study on optical-nondestructive detection of breakdown apples [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2006, 37(6):86 ~ 88, 93. (in Chinese)
- 9 王加华, 孙旭东, 潘璐, 等. 基于可见/近红外能量光谱的苹果褐腐病和水心鉴别[J]. *光谱学与光谱分析*, 2008, 28(9):2 098 ~ 2 102.
Wang Jiahua, Sun Xudong, Pan Lu, et al. Discrimination of brownheart and watercore of apples based on energy spectrum of visible/near infrared transmittance [J]. *Spectroscopy and Spectral Analysis*, 2008, 28(9):2 098 ~ 2 102. (in Chinese)
- 10 Shenderoy C, Shmulevich I, Alchanatis V, et al. NIRS detection of moldy core in apples [J]. *Food and Bioprocess Technology*, 2010, 3(1):79 ~ 86.

(上接第 157 页)

- 8 胡清华, 于达仁, 谢宗霞. 基于邻域粒化和粗糙逼近的数值属性约简[J]. *软件学报*, 2008, 19(3):640 ~ 649.
Hu Qinghua, Yu Daren, Xie Zongxia. Numerical attribute reduction based on neighborhood granulation and rough approximation [J]. *Journal of Software*, 2008, 19(3):640 ~ 649. (in Chinese)
- 9 赵勇, 方宗德, 王侃伟, 等. 邻域粗糙集在轮对踏面缺陷图像特征选择的应用[J]. *计算机测量与控制*, 2008, 16(11):1 730 ~ 1 731.
Zhao Yong, Fang Zongde, Wang Kanwei, et al. Application of neighborhood rough sets in features selection of wheelsets tread defect images [J]. *Computer Measurement & Control*, 2008, 16(11):1 730 ~ 1 731. (in Chinese)
- 10 邓超, 吴军, 万紫娟. 基于支持向量机工具的性能劣化建模方法[J]. *计算机集成制造系统*, 2009, 15(4):685 ~ 689.
Deng Chao, Wu Jun, Wan Zijuan. Performance degradation model based on support vector machine [J]. *Computer Integrated Manufacturing Systems*, 2009, 15(4):685 ~ 689. (in Chinese)