

基于拉曼光谱技术鉴别新陈大米的方法研究

赵迎^{1,2}, 李明², 王小龙², 李小佳^{2*}

1. 钢铁研究总院, 北京 100081

2. 钢研纳克检测技术股份有限公司, 北京 100094

摘要 基于拉曼光谱检测技术结合化学判别方法, 建立新陈大米拉曼光谱判别模型; 建立适当的样品预处理方法, 确保样品制备的均一性, 使用拉曼光谱仪对新陈大米共计60组样品进行检测, 在785 nm波长激光激发下, 获取样品200~2400 cm⁻¹的拉曼光谱信息; 对原始拉曼光谱进行基线校正、平滑、滤波等处理。利用主成分分析法(PCA)对拉曼光谱进行降维处理及粗分类鉴别; 基于偏最小二乘分析法(PLS), 建立新陈大米快速鉴别模型, 该模型对建模训练集鉴别正确率为100%, 模型验证集鉴别正确率为95%。结果表明: 该模型判断新陈大米是可行的, 为大米新陈度的快速判别提供了一种新的方法。

关键词 拉曼; 大米; 主成分分析; 偏最小二乘判别法

中图分类号: TS212.7 **文献标识码**: A **DOI**: 10.3964/j.issn.1000-0593(2019)05-1468-04

引言

稻谷是我国最主要的粮食作物之一, 产量约占全国粮食总产量的一半, 占商品粮一半以上的份额。作为季节性作物, 稻谷需要储存以满足全年需求。稻谷在存储过程中, 受环境因素的影响, 营养成分会发生变化, 这一过程称为大米的陈化。陈化了的大米损失了原有的营养成分且食用品质下降, 甚至会产生有害的毒素。但是受利益诱惑, 一些不法商家贩卖陈化粮的现象屡禁不止, 市场上不时出现陈化粮、染色大米、漂白大米等违法现象。为此国家相关部门曾下发“关于进一步加强陈化粮监管等工作的通知”, 大米的质量监控已经刻不容缓, 如何快速准确的判别大米的陈化度显得尤为重要^[1-3]。

目前我国对稻谷的质量非常重视, 并制定了《稻谷存储品质判定规则》。规则中规定了采用气味、色泽、脂肪酸值和品尝评分值作为稻谷品质的判定依据。但是这些判定依据对应的鉴别方法不仅耗时、花销大、对操作者的专业水平要求较高, 无法满足大米流通环节的快速检测需求^[4-6]。在存储过程中, 受温度、湿度等环境因素的影响, 稻谷的组成成分被氧化从而发生改变。拉曼光谱与红外光谱作为表征物质结构的快检常用手段, 其中, 拉曼光谱具有高灵敏度、快速检测、无损检测且适应性强等优点, 且可对引起大米陈化的理

化特性如脂肪酸、蛋白变性等进行表征。目前, 拉曼技术已广泛应用于制药、法医学、宝石鉴定、石油化工、食品等领域^[1]; 结合表面增强拉曼技术, 拉曼光谱被广泛应用与食品安全领域。表面增强技术虽然极大限度的提高了检测灵敏度, 但样品的前处理相对复杂; 因此, 本实验采用拉曼光谱技术直接探测, 结合化学计量学方法, 建立了大米陈化度的判别模型, 为新陈大米的快速判别及市场监控提供了可能^[7-9]。

1 实验部分

1.1 样品

湖北某粮食库选取库存的新陈稻谷(湘晚籼)各30份, 其中陈化粮稻谷储存超过三年, 新稻为当年新产稻谷, 且新陈稻谷经过国标《稻谷存储品质判定规则 GB/T 20569—2006》方法判定验证, 共60份相同品种的稻谷样品, 用自封袋密闭盛装。

1.2 仪器

便携式拉曼光谱仪(AGILITY 美国 Bayspec), 光谱采集范围: 100~2700 cm⁻¹, 分辨率 6 cm⁻¹, 样品杯 7 mm, Matlab2013a。

1.3 拉曼光谱采集

实验前需对拉曼光谱仪预热 30 min, 为提高样品的代表

收稿日期: 2016-07-20, 修订日期: 2017-03-09

基金项目: 国家重大科学仪器设备开发专项(2017TFF0108900)资助

作者简介: 赵迎, 1989年生, 钢研纳克检测技术股份有限公司光学工程师 e-mail: zhaoying@ncschina.com

* 通讯联系人 e-mail: lixiaojia@ncschina.com

性,对稻谷进行脱壳、打粉、过筛处理。将均匀的大米粉填满样品杯,尽量确保样品表面平整,用直边工装将多余的大米粉刮去,并用平整表面的工装将样品表面压平。将样品杯放入密闭的样品室,选取最优的激光器参数、积分时间及聚焦位置。每份样品随机检测八个位置。

2 结果与讨论

2.1 光谱预处理

在拉曼检测过程中,不可避免的会受到荧光干扰及仪器自身的噪声干扰,采集的原始拉曼光谱如图 1 所示。若不采取合适的数据处理方法,会对分析结果产生很大影响。为提高模型的鉴别准确性,减少噪声信号干扰,需对光谱数据进行合理的预处理。对拉曼光谱进行背景扣除、基线校正、平滑滤波等处理,最终得到拉曼光谱图如图 2 所示。

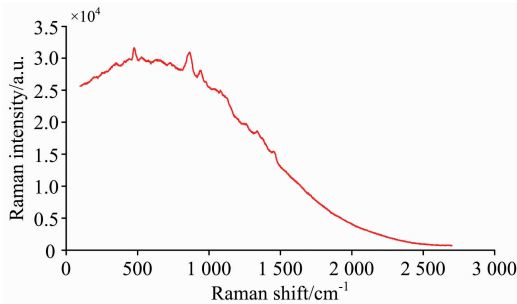


图 1 大米原始拉曼光谱

Fig. 1 Raw Raman spectra of rice sample

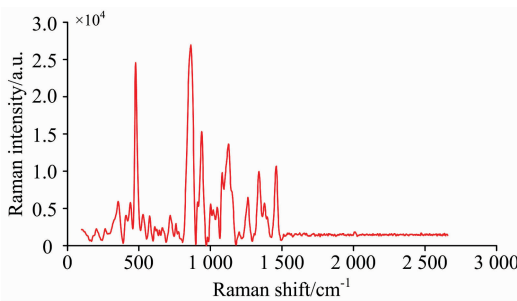


图 2 基线校正处理后大米拉曼光谱

Fig. 2 Raw spectrum after baseline correction

2.2 光谱分析及波段选择

新陈大米拉曼峰出现在 477, 866, 940, 1 083, 1 127, 1 262, 1 336, 1 378, 1 460 cm^{-1} 处,不同峰位代表不同的化学键及振动形式,如图 3 所示。其中,477 cm^{-1} 为淀粉的主链特征峰;866 和 1 262 cm^{-1} 为 CH_2 摇摆振动;940, 1 083 和 1 127 cm^{-1} 为 C—C 键伸缩振动;1 336 cm^{-1} 为 C—H 平面形变振动;1 460 cm^{-1} 为 CH_3 和 CH_2 形变振动,这些化学键及振动形式与样品的保存年份及陈化度息息相关^[10]。

在相同测量条件下,得到新陈大米的拉曼光谱如图 3。新陈大米的拉曼波形重合性高,简单的峰强判断法难以准确鉴别大米的新陈度。但特征波段对应大米中淀粉、蛋白质与脂肪等成分的含量,特征波段差异主要反映了不同年份大米

的营养成分差异。特征峰主要集中在 280~1 500 cm^{-1} 波段,因此模型建立选取该波段。

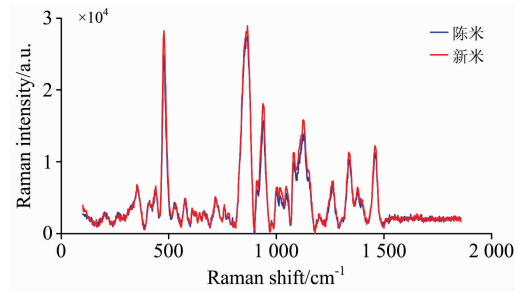


图 3 新陈大米的拉曼光谱

Fig. 3 Raman spectra of two samples which have different storage time

2.3 主成分判别分析

对测得样品的拉曼光谱,进行主成分分析(principal component analysis),降维后结果如图 4,样品在得分图中的分布特征反映样品的新陈度的不同。如图 4 所示,新米主要集中在图上侧,而陈米主要集中在图下侧,新陈大米在得分图中的分布具有规律性,可实现新陈大米的初步鉴别。

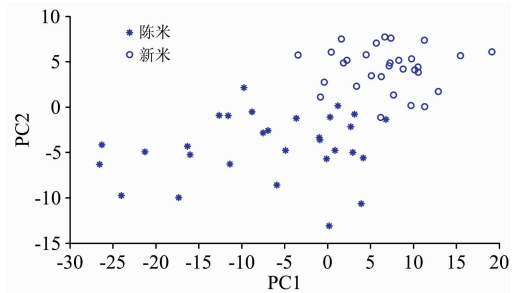


图 4 两种不同新陈度大米的主成分分析

Fig. 4 Principal component analysis score plots of two samples which have different storage time

2.4 预测模型优化

2.4.1 异常数据剔除

在实验检测中,每份样品分别采集八次拉曼光谱以提高光谱数据的代表性,需对每份样品数据中异常数据进行剔除,选取 477 和 866 cm^{-1} 两处对应的峰强 K1 和 K2 建立坐标轴,如图 5 所示,没有落入圆内的点对应的拉曼光谱视为异常数据需剔除。

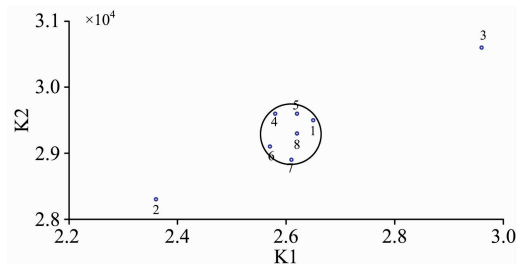


图 5 强度图

Fig. 5 Intensity map

2.4.2 最佳主因子数确定

最佳主因子数与模型及预测准确度息息相关,若主因子数过低则难以准确的反应样品的特征,光谱数据中的某些有用特征便不能体现在模型中,模型鉴别的准确度会大大降低;但主因子数过高又会增大模型的噪声,即使训练样本集结果较好,验证样本集的结果也会受到不良影响,从而导致模型预测效果不佳。本文采用预测残差平方和判断方法,在光谱波段即预处理确定后,采用预测残差平方和来确定最佳因子数。结果如图 6 所示,从图中可看出,最佳主因子数为 8。

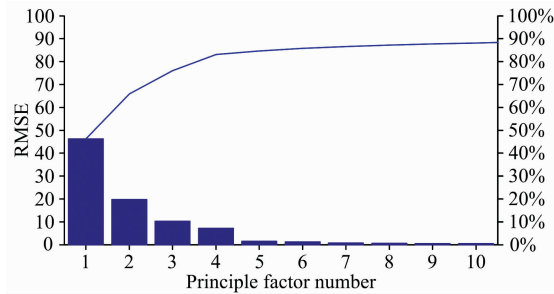


图 6 不同主因子数的预测残差平方和

Fig. 6 RMSE of rice with different principle factor numbers

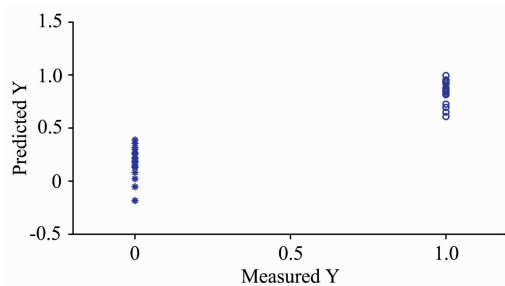


图 7 新鲜大米和陈米的拉曼鉴别模型

Fig. 7 Raman model of fresh and stale rice

2.5 新鲜大米鉴别模型的建立及可靠性检验

选取新鲜大米共 60 份,其中 2/3 样品用于模型建立训练,1/3 样品用于验证模型可靠性。在相同测量条件下测得新鲜大米的拉曼光谱数据,用偏最小二乘法 PLS 对新鲜大米进行模型建立,最佳因子数取值 8。将新米赋值为 1,陈米赋值为 0,建立新鲜大米鉴别模型,模型计算数值大于 0.5 视为新米,小于 0.5 视为陈米,结果如图 7。

对建立的模型准确性进行验证,将验证样本集输入模型得到结果如图 8 所示。结果表明:验证样品集中,仅有 1 个新米判别错误,模型的综合判断率为 95%。由此可见,该模型的预测准确性较好,拉曼光谱结合化学计量学对新鲜大米具有较好的判别预测能力。

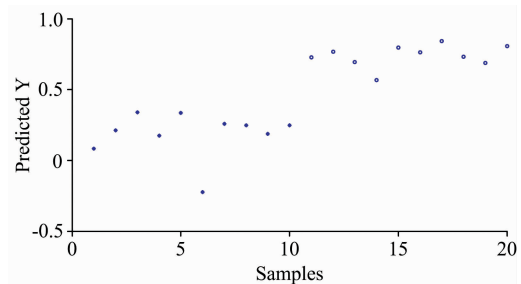


图 8 新鲜大米和陈米的预测结果

Fig. 8 Predictive result of fresh and stale rice

3 结论

(1) 通过对大米进行前处理,结合拉曼光谱技术及化学计量学,建立新鲜大米的拉曼鉴别模型。结果显示,该模型判别效果良好,验证模型的样品中,仅有 1 个新米判别错误,综合的判断率为 95%。说明拉曼光谱技术结合化学计量学适用于大米的新陈度检测,本方法对水分不敏感,且具有快速、无污染、易操作等优点,适合推广使用。

(2) 实验使用的大米样品数量有限,且建模过程中未考虑大米品种等因素。因此,在今后推广过程中,需进一步加大样本量,并建立不同品种大米的区分,以提高模型的预测准确性。

References

- [1] HOU Cai-yun, LI Hui-yuan, SHANG Yan-fen, et al(侯彩云,李慧园,尚艳芬,等). Journal of the Chinese Cereals and Oils Association (中国粮油学报), 2003, 18(4): 80.
- [2] ZHANG Yu-rong, ZHOU Xian-qing, YANG Lan-lan(张玉荣,周显青,杨兰兰). Journal of the Chinese Cereals and Oils Association(中国粮油学报), 2009, 24(8): 155.
- [3] ZHOU Xian-qing, ZHANG Yu-rong(周显青,张玉荣). Cereal & Feed Industry(粮食与饲料工业), 2011, (5): 37.
- [4] RUI Chuang, LIU Ying, SUN Jian-ping(芮闯,刘莹,孙建平). Food Science and Technology(食品科技), 2012, 37(3): 164.
- [5] HAN Xiao-zhen, GUO Zheng-ye, KANG Yan, et al(韩孝联,郭正也,康燕,等). Acta Optica Sinica(光学学报), 2015, 35(1): 130003.
- [6] SUN Juan, ZHANG Hui, WANG Li, et al(孙娟,张晖,王立,等). Food & Machinery(食品与机械), 2016, 32(1): 41.
- [7] HUANG Ya-wei, ZHANG Ling, WANG Ruo-lan(黄亚伟,张令,王若兰). Modern Food Science and Technology(现代食品科技), 2014, 30(12): 149.
- [8] SUN Lu, CHEN Bin, GAO Rui-chang, et al(孙璐,陈斌,高瑞昌,等). Journal of Chinese Institute of Food Science and Technology (中国食品学报), 2012, 12(12): 113.

- [9] ZHAO Xiao-rong, ZHANG Lan-tian, WANG Li-jun, et al(赵肖蓉, 张兰天, 王利军, 等). The Food Industry(食品工业), 2012, 33(12): 158.
- [10] Lee H, Cho B K, Kim M S, et al. Sensors and Actuators B: Chemical, 2013, 185: 694.

Study on Rapid Discrimination of Fresh and Stale Rice Based on Raman Spectroscopy

ZHAO Ying^{1,2}, LI Ming², WANG Xiao-long², LI Xiao-jia^{2*}

1. Central Iron & Steel Research Institute, Beijing 100081, China

2. NCS Testing Technology Co., Ltd., Beijing 100094, China

Abstract A prediction discrimination model of fresh and stale rice was established based on Raman Spectroscopy and Chemometrics. The pretreatment processes have to be employed before the experiment. A total of 60 samples were put in the special box. The samples were measured by 785nm Raman spectrometer, which can collect spectral range of 200~2 400 cm^{-1} . Smoothing, baseline correction were conducted to process the raw spectra. Principal Component Analysis (PCA) was employed to reduce dimension analysis of full-wave band of fresh and stale rice, and it could classify the samples preliminarily. The discrimination model was developed with Partial Least Squares (PLS). The correct classification rates in the training set and prediction set were 100% and 95%, respectively. The results in this research indicated it is a quickly useful method to discriminate between fresh and stale rice.

Keywords Rice; Raman spectroscopy; PCA; PLS

(Received Jul. 20, 2016; accepted Mar. 9, 2017)

* Corresponding author