

李 勇,严煌倩,龙 玲,等. 化学计量学模式识别方法结合近红外光谱用于大米产地溯源分析[J]. 江苏农业科学,2017,45(21):193-195.
doi:10.15889/j.issn.1002-1302.2017.21.054

化学计量学模式识别方法结合近红外光谱 用于大米产地溯源分析

李 勇,严煌倩,龙 玲,余向阳

(江苏省农业科学院农产品质量与安全营养研究所/省部共建国家重点实验室培育基地/江苏省食品质量安全重点实验室,江苏南京 210014)

摘要:对江苏、辽宁、湖北、黑龙江 4 个省份的 169 个大米样品,利用波数测定范围为 $10\,000\sim4\,000\text{ cm}^{-1}$ 的 Thermo Antaris II 傅里叶变换近红外分析仪,采用化学计量学模式识别主成分分析(PCA)和线性判别分析(LDA)方法进行产地溯源分析。结果表明,PCA 方法基于前 2 个主成分可基本区分大米产地,但各类样品有部分重叠;采用 PCA-LDA 法可更有效区分大米产地,利用蒙特卡罗模拟方法随机重复选取训练集和预测集判别 4 个省份的大米产地准确率在 93.00% 以上,识别准确率相对较高。因此,化学计量学模式识别方法结合近红外光谱用于大米产地溯源分析具有一定的可行性和应用价值。

关键词:大米;产地溯源;化学计量学;主成分分析(PCA);线性判别分析(LDA)

中图分类号: TS207.7 **文献标志码:** A **文章编号:** 1002-1302(2017)21-0193-03

地理标志产品具有独特的品质和较高的经济效益,备受消费者青睐。然而,假冒伪劣的地理标志产品极为普遍,严重扰乱市场秩序,损害消费者利益,因此,为整顿市场秩序,急需建立完善的地理标志产品产地溯源分析方法。目前,常用的产地溯源分析方法有近红外光谱(NIR)分析技术、色谱分析技术、DNA 溯源技术、同位素分析技术等,均有一定的应用前景^[1-2],其中,近红外光谱技术作为一种新型的快速检测技术,具有分析速度快、成本低、无损检测、重现性好等特点,常应用于食品、农产品、药品中感兴趣成分的快速定性定量分析^[3-4]。

来源不同产地的地理标志产品其化学成分及各组分含量存在差异,致使近红外光谱也存在差异,而使近红外光谱能够反映出地理标志产品的产地特征信息。近红外光谱属分子吸收光谱,主要反映有机物分子中 C—H、N—H、O—H 等含氢基团振动光谱信息,检测波数范围为 $12\,820\sim4\,000\text{ cm}^{-1}$,由于其光谱吸收带相对较宽,谱带重叠较为严重,利用传统分析方法鉴定产地信息比较困难^[5],而近红外光谱结合化学计量学模式识别方法可更加有效地区分不同产地产品的光谱差异^[6]。目前,常用的化学计量学模式识别方法分为 2 类,即无监督模式识别和有监督模式识别^[5-8]。无监督模式识别法又称为聚类分析法,包括系统聚类法、最小生成树、主成分分析法(PCA)等^[5],类似样本在多维空间中彼此距离相对较近,

而不相似样品的空间距离相对较远,“物以类聚”;有监督模式识别包括距离判别法、线性判别分析(LDA)、偏最小二乘判别分析法(PLS-DA)等,是利用已知类别的样品作为训练集,建立判别模型,再利用模型对未知样品类别进行预测^[5]。

近红外光谱结合化学计量学模式识别方法(SIMCA)已广泛应用于食品、农产品品质鉴定及真伪鉴别中^[9-16]。刘威等利用近红外光谱结合主成分分析及聚类分析研究法国波尔多、我国河北昌黎和沙城葡萄酒产地的溯源分析,准确率达到 88.9%^[17];汤丽华等利用 SIMCA 方法结合近红外光谱区分宁夏回族自治区、甘肃省、青海省、内蒙古自治区、河北省等 8 个产地的枸杞,所建模型识别率达到 80% 以上^[18];周健等采用 PLS 和欧氏距离结合近红外光谱方法实现了茶叶原料的鉴别^[19];钱丽丽等基于 PLS-DA 结合近红外方法对黑龙江省五常、佳木斯、齐齐哈尔、双鸭山、牡丹江 5 个水稻产区进行溯源研究^[20]。本研究利用化学计量学模式识别方法结合近红外光谱对江苏、辽宁、湖北、黑龙江 4 个省份的 169 个大米样品进行产地溯源分析,建立大米产地溯源模型,为大米产地溯源研究提供理论依据。

1 材料与方法

1.1 材料

试验于 2016 年进行,共采集江苏、辽宁、湖北、黑龙江 4 个省份的大米样品 169 个,分别为江苏省泰州市 45 个样品,品种为南粳 46;辽宁省盘锦市 45 个样品,品种为盘锦大米;湖北省荆门市 39 个样品,品种为桥米;黑龙江省五常市 45 个样品,品种为五常香米。大米样品进行脱壳、脱糙,封存,待测。

1.2 数据采集

Thermo Antaris II 傅里叶变换近红外分析仪采用反射积分球模式采集大米近红外光谱,仪器参数为波数范围 $10\,000\sim4\,000\text{ cm}^{-1}$,扫描间隔 3.856 cm^{-1} ,每条光谱包含

收稿日期:2017-06-14

基金项目:国家自然科学基金(编号:31601665);江苏省自然科学基金(编号:BK20160576)。

作者简介:李 勇(1987—),男,黑龙江鹤岗人,博士,助理研究员,主要从事农产品质量安全研究。Tel: (021) 84391229; E-mail: liyong_213@163.com。

通信作者:余向阳,博士,研究员,主要从事农产品质量安全研究。
E-mail: yuxy@jaas.ac.cn。

1 557 个点。每个样品采集 60 次,取平均值作为样品最终光谱;以样品为行变量,以近红外光谱波数为列变量,获得大小为 $169 \times 1\,557$ 的数据阵。

1.3 数据分析

利用 Matlab 软件对近红外光谱数据进行预处理、主成分分析、线性判别分析,相关程序为笔者所在实验室科研人员自编。

2 结果与分析

2.1 数据预处理

由图 1-a 可见,利用近红外漫反射光谱采集大米样品光谱会存在一定的散射和噪声干扰,且同一省份大米样品的光谱差异较大,会对后续数据分析产生很大影响。试验采用标准正态变换(SNV)方法对近红外数据进行预处理,以去除数据中散射和噪声的干扰。由图 1-b 可见,经预处理后,4 个

省份的大米样品光谱均重叠在一起,几乎不能用肉眼来进行区分。

2.2 主成分分析(PCA)

主成分分析是一种常用的数据降维方法,主要是对数据中的原始变量进行线性组合,得到几个正交的成分即主成分,对原数据的协方差阵进行解释,在主成分分析投影图中相近的样品将会聚到一起,而不相似的样品空间距离相对较远^[21]。本试验利用奇异值分解方法对试验数据进行主成分分析,在数据分析前,对数据进行归一化处理。分析结果表明,前 6 个主成分的累计方差贡献率分别为 60.41%、28.03%、4.46%、2.42%、1.51%、0.33%,其中前 5 个主成分的累计方差贡献率为 96.83%,确定主成分数为 5。由图 2 可见,利用前 2 个主成分基本可以区分 4 个省份的大米样品,但部分样品有重叠;利用其他主成分投影时,不同类别样品重叠则更为严重。

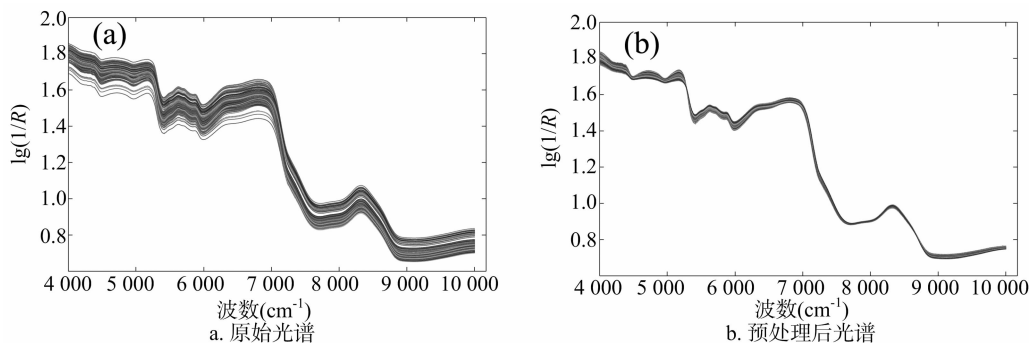


图1 四省份大米样品红外光谱

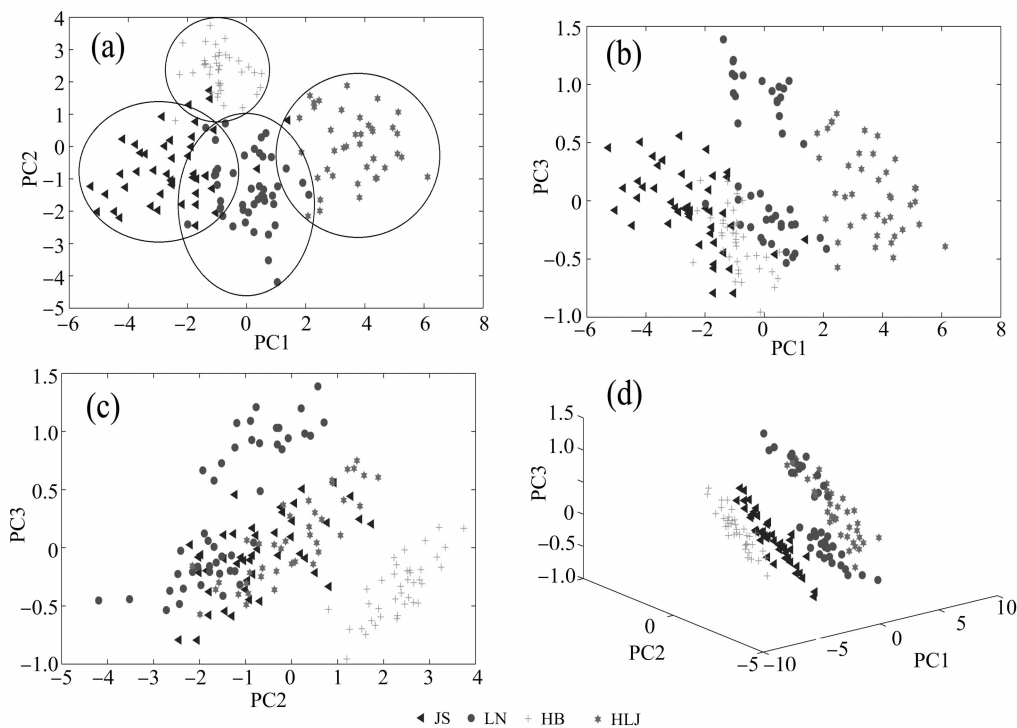


图2 江苏(JS)、辽宁(LN)、湖北(HB)、黑龙江(HLJ)四省份大米样品主成分得分

2.3 线性判别分析(LDA)

线性判别分析是一种有监督学习型的简单线性判别函数

分析方法,其训练集构建函数要求组内方差最小、组间方差最大,再利用构建函数对未知样品类别进行预测^[22]。本试验利

用蒙特卡洛模拟方法,将大米样品随机分为训练集、预测集,利用 LDA 对训练集样品建立大米产地溯源模型,并利用已建立的模型对预测样品进行判定。根据 PCA 结论,该数据的组分数为 5,因此,LDA 对前 5 个主成分进行分析。基于训练集建模,得到 3 个判别函数,即 Function1、Function2、Function3:

$$\text{Function1} = 0.017 \times PC1 + 0.010 \times PC2 + 0.011 \times PC3 - 0.095 \times PC4 + 0.104 \times PC5 + 0.027;$$

$$\text{Function2} = 0.018 \times PC1 + 0.121 \times PC2 - 0.277 \times PC3 + 0.537 \times PC4 - 0.739 \times PC5 + 0.146;$$

$$\text{Function3} = 0.674 \times PC1 - 0.958 \times PC2 + 3.131 \times PC3 - 0.064 \times PC4 + 1.732 \times PC5 + 0.715。$$

式中:Function1 用于区分江苏产地与其他 3 个省份的大米样品,Function2 用于区分辽宁省和湖北、黑龙江 2 个省份的大米样品,Function3 用于区分湖北省和黑龙江省的大米样品。利用蒙特卡洛模拟方法随机重复计算 1 000 次,结果由表 1 可见,在训练集中,识别江苏省大米的准确率为 96.25%,3.75% 的江苏省大米样品错误判为湖北省大米,辽宁、湖北 2 个省份的大米样品识别准确率均为 100.00%,黑龙江省大米识别率为 94.09%,5.91% 的黑龙江省大米样品被判定为江苏省大米,总体而言,预测结果的正确率在 94.00% 以上,训练集结果的准确率相对较高,令人满意;在预测集中,识别江苏、辽宁、湖北、黑龙江 4 个省份大米样品的准确率分别为 94.91%、100.00%、100.00%、93.97%,预测结果的准确率也较高。因此,PCA-LDA 方法可有效区分不同地区的大米。

表 1 PCA-LDA 区分江苏等 4 个省份大米样品的分析结果

集合	省份	样本数 (个)	大米样品识别率(%)			
			江苏省	辽宁省	湖北省	黑龙江省
训练集	江苏省	37	96.25		3.75	
	辽宁省	37		100.00		
	湖北省	32			100.00	
	黑龙江省	33	5.91			94.09
预测集	江苏省	8	94.91		5.09	
	辽宁省	8		100.00		
	湖北省	7			100.00	
	黑龙江省	7	6.03			93.97

3 结论

利用化学计量学主成分分析(PCA)和线性判别分析(LDA)结合近红外光谱,对江苏、辽宁、湖北、黑龙江 4 个省份的 169 个大米样品进行产地溯源分析,结果表明,利用 Thermo Antaris II 傅里叶变换近红外分析仪测定大米近红外光谱会存在散射和噪声的干扰,在采用标准正态变换(SNV)方法去杂、去噪的基础上,PCA 方法基本可区分大米产地,但各类样品有部分重叠,而 PCA-LDA 法可更有效地区分大米产地,利用蒙特卡罗模拟方法重复随机筛选训练集和预测集,识别 4 个省份大米品种的准确率在 93.00% 以上。因此,化学计量学模式识别方法结合红外光谱可对大米产地进行溯源分析,具有一定的可行性和应用价值。

参考文献:

- [1] 魏益民,郭波莉,魏 帅,等. 食品产地溯源及确证技术研究和应用方法探析[J]. 中国农业科学,2012,45(24):5073-5081.
- [2] 张晓焱,苏学素,焦必宁,等. 农产品产地溯源技术研究进展[J]. 食品科学,2010,31(3):271-278.
- [3] 褚小立,刘慧颖,燕泽程. 近红外光谱分析技术实用手册[M]. 北京:机械工业出版社,2016:114-161.
- [4] 褚小立,陆婉珍. 近五年我国近红外光谱分析技术研究与应用进展[J]. 光谱学与光谱分析,2014,34(10):2595-2605.
- [5] 梁逸曾,俞汝勤. 分析化学手册:化学计量学[M]. 北京:化学工业出版社,2000.
- [6] 许 禄. 化学计量学:一些重要方法的原理及应用[M]. 北京:科学出版社,2004:1-25.
- [7] 吴海龙,俞汝勤. 化学多维校正的若干新进展[J]. 化学通报,2011,74(9):771-782.
- [8] Wold S. Chemometrics: what do we mean with it, and what do we want from it? [J]. Chemometrics and Intelligent Laboratory Systems, 1995,3(1):109-115.
- [9] 赵杰文,毕夏坤,林 颢,等. 鸡蛋新鲜度的可见-近红外透射光谱快速识别[J]. 激光与光电子学进展,2013,50(5):209-216.
- [10] Pholpho T, Pathaveerat S, Sirisomboon P. Classification of long an fruit bruising using visible spectroscopy [J]. Journal of Food Engineering, 2011,104(1):169-172.
- [11] 吕 强,汤明杰,赵杰文,等. 近红外光谱预测猕猴桃硬度模型的简化研究[J]. 光谱学与光谱分析,2009,29(7):1768-1771.
- [12] 邵圣枝,陈元林,张永志,等. 稻米中同位素与多元素特征及其产地溯源 PCA-LDA 判别[J]. 核农学报,2015,29(1):119-127.
- [13] 杨春艳,刘 飞,王元忠. 红外光谱结合判别分析对玛咖产地的鉴别[J]. 江苏农业科学,2017,45(5):170-173.
- [14] 顾玉琦,刘瑞婷,寿国忠,等. 应用近红外光谱技术快速鉴别铁皮石斛的产地[J]. 江苏农业科学,2016,44(5):365-368.
- [15] 刘 飞,杨春艳,道永艳,等. 傅里叶变换红外光谱对滇重楼生长年限的鉴别研究[J]. 江苏农业科学,2016,44(2):300-303,304.
- [16] 武小红,孙 俊,武 斌,等. 基于联合区间偏最小二乘判别分析的猪肉近红外光谱定性建模分析[J]. 激光与光电子学进展,2015,52(4):242-247.
- [17] 刘 威,战吉成,董卫东,等. 基于近红外光谱技术的葡萄酒原产地辨识方法[J]. 农业工程学报,2010(增刊1):374-378.
- [18] 汤丽华,刘敦华. 基于近红外光谱技术的枸杞产地溯源研究[J]. 食品科学,2011,32(22):175-178.
- [19] 周 健,成 浩,曾建明,等. 基于近红外的多相偏最小二乘模型组合分析实现茶叶原料品种鉴定与溯源的研究[J]. 光谱学与光谱分析,2010,30(10):2650-2653.
- [20] 钱丽丽,冷候喜,宋雪健,等. 基于 PLS-DA 判别法对黑龙江大米产地溯源的研究[J]. 食品工业,2017,38(1):171-174.
- [21] Brereton G, Brereton. Chemometrics for pattern recognition [M]. New Jersey: John Wiley & Sons Inc, 2009.
- [22] Galdón B R, Rodríguez L H, Mesa D R, et al. Differentiation of potato cultivars experimentally cultivated based on their chemical composition and by applying linear discriminant analysis [J]. Food Chemistry, 2012,133(4):1241-1248.