

林 萍,陈永明. 利用可见近红外光谱技术快速鉴别大米品种[J]. 江苏农业科学,2015,43(12):320-323.
doi:10.15889/j.issn.1002-1302.2015.12.101

利用可见近红外光谱技术快速鉴别大米品种

林 萍,陈永明

(盐城工学院电气工程学院,江苏盐城 224051)

摘要:提出了 1 种应用近红外光谱技术快速无损鉴别大米品种的新方法。采用近红外光谱仪获取 3 种大米的光谱吸收特征曲线,运用遗传算法抽取 15 个特征波长,并对 15 个特征波长运用偏最小二乘法进行模式特征分析;经过交叉验证法判别提取主成分,完成特征提取后,将 7 种主成分作为神经网络的输入变量,建立了 3 层误差逆传播(back propagation, BP)神经网络。结果表明,对 30 个未知样本进行预测,预测相对偏差均 $< 5\%$,预测结果准确率达 100%;遗传算法结合偏最小二乘法进行聚类比单独使用偏最小二乘法对大量原始光谱数据进行聚类分析的效果好, BP 神经网络预测的精确度也大大提高。该方法能快速无损地检测大米品种,同时为其他有机物品种鉴别提供了新方法。

关键词:遗传算法;偏最小二乘法;BP 神经网络;大米;可见近红外光谱

中图分类号:S126;TH744.1 **文献标志码:**A **文章编号:**1002-1302(2015)12-0320-03

大米是中国人日常生活所必需的主食,中国 60% 以上的人口以大米为主食。随着城镇居民收入的提高,人们对日常食用大米的品质要求也越来越高,大米不但要口感好、味道香,还要富含营养。如今市场上可供选择的大米品种繁多,其中泰国香米以其独特的口感和香味得到广大消费者的喜爱;粒粒珍硒牌香米有机硒含量较高,是普通大米的数十倍,被认为具有较高的营养价值;东北长粒香米属于生态米,无农药耕种,且富含氨基酸、蛋白质及其他许多种有益人体健康的微量元素。这些大米各有特色,因此对不同产地、不同品种的大米加以鉴别区分显得十分必要。

本研究提出了 1 种快速简便无损的分析方法对不同品种的大米进行鉴别,即可见-近红外光谱法。可见-近红外光谱是波长在 $0.380 \sim 2.526 \mu\text{m}$ 之间的电磁波,已广泛应用于各种定性或定量分析^[1],如石油、化工、食品、医药等行业。国内外很多学者利用近红外光谱技术进行物质品种鉴别,如大黄品种^[2]、中药材^[3]、咖啡品种^[4]等,但其在大米品种鉴别方面的应用还鲜有报道。

遗传算法是一种用计算机模拟生物自然进化过程搜索最优解的方法^[5],是通过类比达尔文的进化论过程而产生的一种概率搜索算法,利用某种编码技术作用于被称为染色体的数串,并模拟由这些数串组成的群体的进化过程。遗传算法通过运用简单的编码技术和繁殖机构来演绎复杂的现象,进

而应对较为棘手的问题;另外遗传算法对求解问题无需进行连续性、可导性、局部单峰等假设,没有约束于特定的求解空间,这些优点使其在有机物质光谱分析中表现出了较大的应用潜力。偏最小二乘法(partial least squares, PLS)是一种目前应用较广泛的定量分析的化学计量学方法,可以实现全谱或部分谱数据的分析。人工神经网络通过模拟人脑神经细胞的功能和构型来实现复杂的非线性映射,其中提出的误差逆传播(back propagation, BP)神经网络通过采用误差反向传播方法表现出较好的预测性能而得到了推广应用^[6]。本研究提出了基于遗传算法选择波长,并与 BP 神经网络以及偏最小二乘法相结合的判别技术,研究用近红外光谱技术对大米进行品种鉴别。对 30 个未知样本进行预测,准确率为 100%,30 个样本的预测相对偏差均 $< 5\%$,说明该方法对于大米具有很好的分类和鉴别能力。

1 材料与方法

1.1 仪器设备

采用美国 ASD (Analytical Spectral Device) 公司的 Hand-held FieldSpec 光谱仪,其测定范围在 $0.325 \sim 1.075 \mu\text{m}$,光谱采样间隔为 $1.5 \times 10^{-3} \mu\text{m}$,扫描次数 30 次,探头视场角为 20° 。光源采用与光谱仪配套的 14.5 V 卤素灯。光谱数据经 ASD View Spec Pro 软件转化为 ASCII 码形式,再由分析软件 Unscrambler V9.7、Matlab 7.5 及 DPS^[7]对数据进行分析处理。

1.2 样品来源及光谱的获取

从超市买来 3 种不同产地的大米:东北长粒香生态米(产地:吉林长白山生态保护区)、口口牌泰国香米(产地:泰国)、粒粒珍硒牌御贡香米(产地:黑龙江),共 120 个样本(每个品种 40 个)。各种大米样本均用直径 12 cm、高 1 cm 的培养皿盛装。为了减小试验误差,保证被测物体与仪器等距,将每个培养皿装满作为 1 次试验用测试样本。每类品种各制作出 40 个试验样本,共计 120 个试验样本。所有测试用样本被随机划分为建模集、预报集,建模集包括 90 个测试样本(每类品种 30 个),预报集有 30 个测试样本(每类品种 10 个)。

收稿日期:2015-06-01

基金项目:国家自然科学基金(编号:31501221);江苏省自然科学基金(编号:BK20140467);江苏省高校自然科学研究面上项目(编号:13KJB210006);盐城市农业科技指导性计划(编号:YKN2014009、YKN2014010);盐城工学院人才引进项目(编号:KJC2014007、KJC2014006);盐城工学院校级培育项目(编号:XKY2013013、XKY2014056、XKY2014055)。

作者简介:林 萍(1982—),女,江苏盐城人,博士,讲师,主要从事农产品质量检测的研究。E-mail:binglvcha007@126.com。

通信作者:陈永明,博士,讲师,主要从事农业智能系统的研究。E-mail:billrange@126.com。

大米样本光谱数据采集前先用白板对光谱仪进行校正。光谱仪探头最前端放置于距大米表面 10 cm 处的位置上,光谱仪扫描的次数设置为 30 次,最后计算所有获得的光谱曲线的平均值作为样本特征光谱曲线。

1.3 光谱数据预处理

为了消除光谱曲线受到高频随机噪声、扫描样本位置排布不均匀、光散射、基线漂移等影响,需要进行光谱预处理来消除噪声。先对数据进行 Savitzky Golay 一阶微分处理,平滑点数设为 6,一阶微分可以去除与波长无关的漂移^[8]。再采用 Moving Average 平滑法,平滑点数为 9,平滑处理可以提高分析信号的信噪比,很好地滤除各种因素产生的高频噪声^[9];最后对数据进行 MSC 处理^[10],多元散射校正可以去除近红外漫反射光谱中样品的镜面反射及不均匀性造成的噪声,消除漫反射光谱的基线及光谱的不重复性,处理后的光谱数据作为遗传算法的输入。

1.4 遗传算法工作原理

本研究采用遗传算法(genetic algorithm,GA)进行特征波长提取,遗传算法特征波长提取过程分为以下 5 个步骤。(1)根据用户设定的成员数产生第 1 代群体,每条染色体的长度由优化参数的个数决定。(2)染色体的适应度度量通过将每条染色体中的参数值代入目标函数中,并求解出函数值。(3)通过目标函数值转换,得到染色体的适应能力强弱指标。(4)得到新一代个体。随机地从母染色体中选出 2 条进行交叉,从而产生下一代的新个体。在选择母代染色体的过程中,适应性强的染色体被选中的概率高。(5)变异。根据一定概率对染色体上的基因进行变异,产生一些新的个体。整个算法将在步骤(2)和步骤(5)之间循环重复^[5]。

1.5 偏最小二乘法

作为一种多元统计方法,偏最小二乘法能够建立光谱数据和主成分之间的关系模型,常用于近红外光谱数据处理^[11]。它分别对原始光谱矩阵 X 和浓度矩阵 Y 进行相应的分解,消除无用信息。

步骤 1:对原始光谱矩阵 X 和浓度矩阵 Y 的分解模型分别为:

$$X = TP + E; \quad (1)$$

$$Y = UQ + F. \quad (2)$$

式中: T 、 U 分别对应于原始光谱矩阵 X 、浓度矩阵 Y 的评分矩阵; P 、 Q 分别对应于原始光谱矩阵 X 、浓度矩阵 Y 的负载矩阵; E 、 F 分别对应于原始光谱矩阵 X 、浓度矩阵 Y 的剩余残差矩阵。

步骤 2:将 T 、 U 作线性回归:

$$U = TB. \quad (3)$$

式中: B 为关联系数矩阵。

在预测时,首先根据光谱矩阵 X 的载荷矩阵 P 求出未知样品光谱矩阵的得分 $T_{\text{未知}}$,然后求出浓度预测值 $Y_{\text{未知}}$:

$$Y_{\text{未知}} = T_{\text{未知}} BQ. \quad (4)$$

1.6 BP 神经网络模型

偏最小二乘法适合处理线性模型,但是当输入变量和输出变量不完全呈线性关系时,可能会给分析结果带来一定的偏差,因此需要考虑在线性模型里引进非线性项。针对非线性问题,目前应用最多的是 BP 人工神经网络。BP 神经网络

通常采用由非线性变换神经单元组成的一种前馈型结构形式,神经元由 3 个层次组成,分别为输入层、输出层、隐含层,该结构具有良好的非线性映射逼近和泛化能力。数据由输入层输入,通过标准化及加权处理传输到隐含层,隐含层经过权值、阈值和激励函数运算后,由输出层计算出预测值,并将预测值与期望值进行比较,若存在误差,则进行误差反向传播,并对权值和阈值进行调整,逐步使 BP 神经网络输出值与期望值趋于一致^[12]。

本研究建立了 1 个 3 层 BP 神经网络。所有的样本被随机地分成建模集、预测集,其中 90 个样本为建模集样本,30 个样本为预测集样本。网络输入层、隐含层、输出层节点数分别为 7、7、1 个,其中输入层的 7 个节点来自 PLS 分析得到的主成分。最小训练速率为 0.1 步长,设定训练迭代次数为 1 000 次,对输入样本进行标准化处理。

2 结果与分析

2.1 不同品种大米的近红外光谱图分析

图 1 中横坐标为波长,范围 0.325 ~ 1.075 μm ;纵坐标为吸光度。可以看出,3 种大米的吸光度曲线没有明显差别,很难直接从光谱图中将它们区别开。波谱在可见光部分曲线较为平坦,红外光谱区域吸收率有所上升。应用 ASD View Spec Pro 软件,把大米的光谱曲线作平均处理,并转换成 ASC II 码,形成吸收率矩阵,并运用遗传算法提取特征波长。

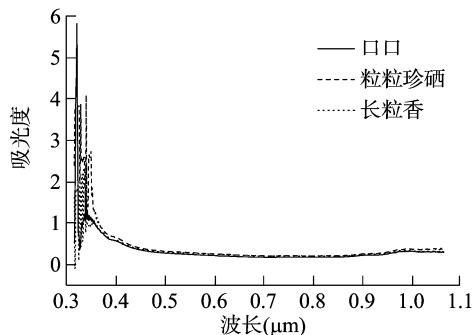


图1 3类不同品种大米的典型近红外光谱吸收结果

2.2 不同品牌大米的聚类分析

对 3 种大米—东北长粒香生态米、口口牌泰国香米、粒粒珍晒牌御贡香米(各 30 个,共 90 个)样本进行主成分分析,将获得的主成分(principal component,PC)对应的特征值由大到小进行排列,根据相应特征值所占的比重确定主成分的可信度,将前 N 个主成分对应的可信度进行累加,即获得了前 N 个主成分的累计可信度。如表 1 所示,本研究列出了试验获得的前 7 个主成分的累计可信度,可见前 7 个主成分的可信度已达 98.1%,因此用前 7 个主成分就可以表示原近红外光谱的主要信息。

表 1 前 7 个主成分的累计可信度

主成分	累计可信度(%)
PC1	14.3
PC2	50.8
PC3	88.0
PC4	92.3
PC5	94.7
PC6	97.0
PC7	98.1

图 2 表示 90 个建模样本的主成分 1、2、3 得分,可以看出,东北长粒香生态米与口口牌泰国香米,粒粒珍硒牌御贡香米可以明显区别开来,而口口牌泰国香米与粒粒珍硒牌御贡香米则有部分重叠,说明主成分 1、2、3 对 3 种大米有一定的聚类作用。

2.3 GA 抽取特征波长

遗传算法控制参数设定:初始群体 100,遗传迭代次数 50,交叉概率 0.8,变异概率 0.1。将预处理后的光谱数据用 GA 进行波长筛选,抽取出东北长粒香生态米、口口牌泰国香米与粒粒珍硒牌御贡香米特征波长各 5 个,共 15 个特征波长作为 PLS 输入变量。图 3 显示了 GA 抽取的 15 个敏感特征波长,虚线对应的横坐标即为 GA 抽取的特征波长。

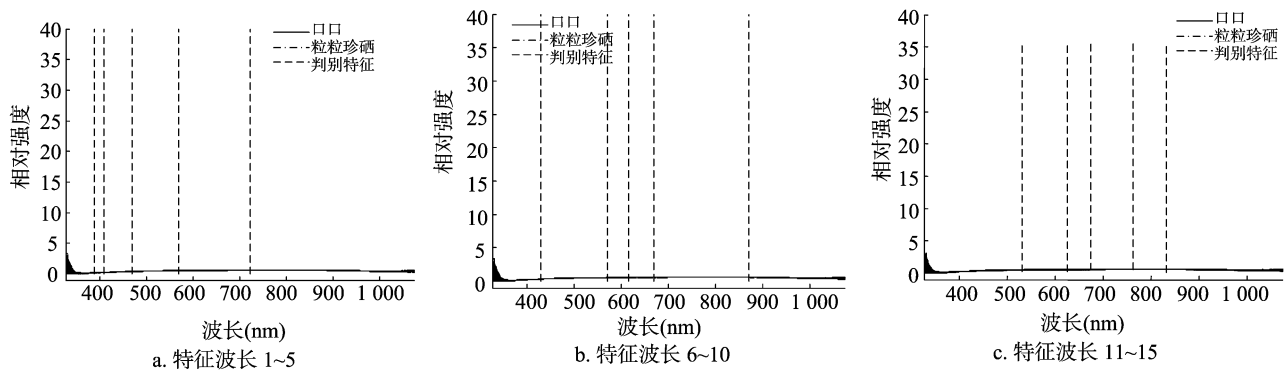


图3 GA 抽取的 15 个敏感特征波长

2.4 基于特征变量建立 BP 神经网络模型

试验采集到大米光谱数据波段范围在 0.325 ~ 1.075 μm 之间,如果将全波段的光谱数据作为 BP 神经网络的输入,这将加重其计算量,影响其学习的性能和效率;另外,由于获取的部分波段样品的光谱信号较弱,从而使得获取的光谱曲线与样品的浓度之间缺乏相关性,因此本研究使用主成分分析法来消除大米光谱矩阵中无关信息,利用获得的前 7 个主要成分来表达大米特征,然后将这 7 个主成分特征变量作为 BP 神经网络的输入变量,即网络输入层、隐含层、输出层节点数分别为 7、7、1 个。训练集、预测集样本数分别为 90、30 个。对 30 个未知样本的预测结果见表 2。可以看出,预测样本 1 至 10 号真实值为 1,预测样本 11 至 20 号真实值为 2,预测样本 21 至 30 号真实值为 3,依次代表口口牌泰国香米、粒粒珍硒牌御贡香米、东北长粒香生态米这 3 个品种。拟和残差为 0.000 6,30 个样本的预测相对偏差均 < 5%,即预测的正确率达到了 100%。

3 结论

应用近红外光谱技术建立了大米类别鉴别的模型,该模型的预测效果较好,对未知样品的预测相对误差均在 < 5%,识别率达到 100%。说明运用近红外光谱可以快速、准确地对大米类别进行鉴别。采用遗传算法进行特征波长提取,并用偏最小二乘法进行主成分分析,最后结合 BP 神经网络进行预测,大大提高了识别的精确度。采用从 PLS 分析中得到的特征变量作为 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,用遗传算法、偏最小二乘法结

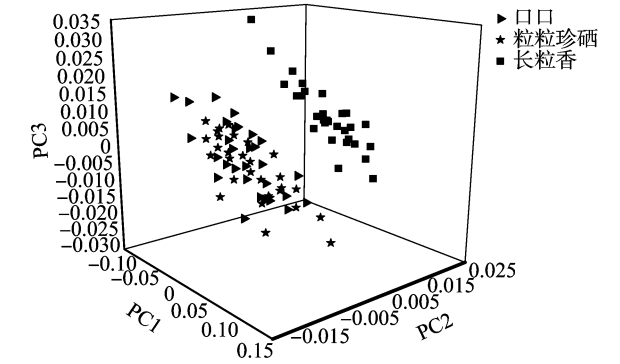


图2 90个样本的主成分1、2、3的得分

表 2 BP 神经网络模型对未知样本的预测结果

样本号	预测值	真实值	相对偏差(%)	样本号	预测值	真实值	相对偏差(%)
1	1.000 0	1	0	16	2.022 5	2	1.12
2	1.000 0	1	0	17	2.018 2	2	0.91
3	1.021 9	1	2.19	18	2.011 3	2	0.56
4	1.000 0	1	0	19	1.997 3	2	0.13
5	1.016 0	1	1.60	20	2.993 6	2	0.21
6	1.000 0	1	0	21	2.973 3	3	0.89
7	1.001 5	1	0.15	22	2.981 8	3	0.60
8	1.000 0	1	0	23	2.995 9	3	0.13
9	1.040 8	1	4.08	24	2.997 1	3	0.09
10	1.000 1	1	0.01	25	2.963 0	3	1.23
11	2.001 1	2	0.05	26	2.962 8	3	1.24
12	1.982 5	2	0.87	27	2.989 7	3	0.34
13	2.034 5	2	1.72	28	2.995 0	3	0.16
14	1.913 3	2	4.33	29	2.985 4	3	0.48
15	2.009 8	2	0.49	30	2.991 2	3	0.29

注:真实值为 1、2、3 分别代表口口牌(泰国)、粒粒珍硒牌(黑龙江)、长粒香牌(东北)。

合 BP 神经网络的模式识别和用近红外光谱技术研究大米品种鉴别是可行的。

参考文献:

[1] 李晓文,李萍萍,袁俊杰. 基于可见/近红外光谱的鸡腿菇总糖含量测定方法[J]. 江苏农业科学,2013,41(9):279-281.
[2] 汤彦丰,侯占忠,郝青川,等. 径向基函数神经网络和近红外光谱法鉴别大黄的应用[J]. 湖北农业科学,2014,53(2):423-425.

马 敏,刘嘉宝,陈兰兰. 苯酚-硫酸法测定多糖含量显色条件的优化与改进[J]. 江苏农业科学,2015,43(12):323-324.
doi:10.15889/j.issn.1002-1302.2015.12.102

苯酚-硫酸法测定多糖含量显色条件的优化与改进

马 敏,刘嘉宝,陈兰兰

(青海民族大学,青海西宁 810007)

摘要:在单因素试验基础上,选择对多糖含量测定结果产生显著影响的4个显色因素:6%苯酚用量、浓硫酸用量、反应温度、水浴显色时间进行正交试验,获得传统苯酚-硫酸法测定多糖含量的最佳显色条件是1 mL 6%苯酚溶液、7 mL 浓硫酸、4℃反应温度、30 min 水浴显色。改进后的多糖含量测定的显色条件即按照1:7的比例直接配制苯酚-硫酸显色液,在4℃条件下加入样品溶液,在90℃水浴进行30 min 显色后,于490 nm处测定吸光度,计算多糖含量。通过传统最佳显色方法的验证试验和改进后的显色方法测定多糖含量的结果比较,表明改进后的方法具有操作简便、精密度高、准确度高、实用性强的优势。

关键词:苯酚-硫酸法;多糖测定;显色条件

中图分类号: O657.32 **文献标志码:** A **文章编号:** 1002-1302(2015)12-0323-02

多糖是一种天然高分子化合物,根据其功能不同可分为活性多糖和膳食纤维2类。活性多糖类化合物作为绝大多数天然产物中的主要降血糖活性成分,具有比重大、作用温和、毒性反应低、作用强的特点,同时兼有降血脂和抗血小板聚集等生理功能和活性^[1]。因此,从天然产物中筛选、发现、分析、测定多糖类化合物引起了研究者高度重视。目前,多糖含量测定的传统方法有苯酚-硫酸法、蒽酮-硫酸法、DNS法,但皆存在操作步骤繁琐、耗时长、测定结果重现性和准确度较差的问题^[2]。因此,本研究将在获得最佳显色条件的基础上,改进显色剂的加入方式,从而获得测定多糖含量的新方法。

1 仪器与试剂

葡萄糖(AR,天津市大茂化学试剂)、硫酸(AR,白银市良友化学试剂有限公司)、苯酚(AR,天津市大茂化学试剂)、电子分析天平(FA2204,上海安亭电子)、722E型可见分光光度计(上海光谱仪器有限公司)、DZKW-4电子恒温水浴锅(北京中兴伟业仪器有限公司)。

收稿日期:2014-11-18

基金项目:青海民族大学青年基金(编号:2014XJZ09)。

作者简介:马 敏(1984—),女,河南,硕士,讲师,主要从事分析化学研究。E-mail:mamin273670@126.com。

2 方法与结果

2.1 苯酚液的配制

苯酚100.00 g,加铝片0.10 g和碳酸氢钠0.06 g,常压蒸馏,收集182℃馏分。称取该馏分5.00 g,置100.00 mL容量瓶中,加新鲜蒸馏水稀释至刻度,摇匀,过滤至棕色试剂瓶中,得6%苯酚液,置冰箱中冷藏,待用。

2.2 葡萄糖对照品溶液的制备

105℃干燥至恒质量的葡萄糖0.100 8 mg,溶解后转移至1 000 mL棕色容量瓶中,加水稀释至刻度,摇匀,得浓度0.100 8 mg/mL葡萄糖对照品溶液。

2.3 测定波长的选择

量取0.100 8 mg/mL葡萄糖对照品溶液3.00 mL,置于25.00 mL具塞比色管中,加蒸馏水1.00 mL,混匀后,加入新配制的6%苯酚液2.00 mL,混匀,迅速滴加浓硫酸7.00 mL,控制在微放热或不放热为宜,90℃水浴加热30 min,取出至冷水中冷却至室温。另取蒸馏水同法平行操作作为空白对照,于400~600 nm范围内扫描获得吸收曲线,结果表明在488 nm波长处有最大吸收(图1),故以490 nm作为测定波长^[3]。

2.4 葡萄糖标准曲线的绘制

精确移取0.15 mg/mL葡萄糖标准溶液0.40、0.50、

学,2010,38(3):342-346。

[9]王伟明,董大明,郑文刚,等. 梨果糖浓度近红外漫反射光谱检测的预处理方法研究[J]. 光谱学与光谱分析,2013,33(2):359-362。

[10]罗曦芸,杜一平,沈美华,等. 红外光谱在纤维质文物材料鉴别中的应用研究[J]. 光谱学与光谱分析,2015,35(1):60-64。

[11]邹孝恒,郝中骐,易荣兴,等. 基于遗传算法和偏最小二乘法的土壤激光诱导击穿光谱定量分析研究[J]. 分析化学,2015,43(2):181-186。

[12]韩瑞珍,宋 韬,何 勇. 基于可见/近红外光谱的土壤有机质含量预测[J]. 中国科学:信息科学,2010,40(增刊1):111-116。

[3]孙仁爽,金哲雄,张哲鹏,等. 牦牛儿苗科11种中药材红外光谱鉴定及聚类分析[J]. 光谱学与光谱分析,2013,33(2):371-375。

[4]鲍一丹,陈 纳,何 勇,等. 近红外高光谱成像技术快速鉴别国产咖啡豆品种[J]. 光学精密工程,2015,23(2):349-355。

[5]祝诗平,王一鸣,张小超,等. 基于遗传算法的近红外光谱谱区选择方法[J]. 农业机械学报,2004,35(5):152-156。

[6]吴燕萍,鲍一丹,何 勇. 基于BP神经网络的黄豆含水率无损检测分析[J]. 农机化研究,2007(2):126-129。

[7]李秀昌,韩曦英,孙 健. 利用DPS数据处理系统进行均匀试验设计与分析[J]. 中国卫生统计,2010,27(2):201,203。

[8]谢 军,潘 涛,陈洁梅,等. 血糖近红外光谱分析的 Savitzky-Golay 平滑模式与偏最小二乘法因子数的联合优选[J]. 分析化