

# 基于近红外光谱技术的大米品种快速鉴别方法

周子立<sup>1,3</sup>, 张 瑜<sup>2</sup>, 何 勇<sup>3\*</sup>, 李晓丽<sup>3</sup>, 邵咏妮<sup>3</sup>

(1. 浙江机电职业技术学院, 杭州 310053; 2. 浙江经济职业技术学院, 杭州 310018;

3. 浙江大学生物系统工程与食品科学学院, 杭州 310029)

**摘 要:**为探索大米无损检测技术, 提出了一种基于可见-近红外光谱技术快速、无损鉴别大米品种的新方法。首先采用主成分分析法对大米品种进行聚类, 然后利用小波变换技术提取光谱特征信息, 把光谱特征信息作为人工神经网络的输入建立品种识别模型, 对大米品种进行鉴别。从每种大米 60 个样本共计 180 个样本中随机抽取 150 个样本(每种 50 个样本)用来建立神经网络模型, 剩下的 30 个大米样本用于预测。品种识别准确率达到 100%。说明所提出的方法具有很好的分类和鉴别作用, 为大米品种鉴别提供了一种新方法。

**关键词:**可见-近红外光谱, 大米, 主成分分析, 小波变换, 人工神经网络

doi: 10.3969/j.issn.1002-6819.2009.08.024

中图分类号: O657.3, S602.3

文献标识码: A

文章编号: 1002-6819(2009)-8-0131-05

周子立, 张 瑜, 何 勇, 等. 基于近红外光谱技术的大米品种快速鉴别方法[J]. 农业工程学报, 2009, 25(8): 131—135.

Zhou Zili, Zhang Yu, He Yong, et al. Method for rapid discrimination of varieties of rice using visible NIR spectroscopy[J]. Transactions of the CSAE, 2009, 25(8): 131—135.(in Chinese with English abstract)

## 0 引 言

大米是中国主要粮食作物之一, 主要产地分布在东北及江南地区。大米的品质检测和品种鉴别一直是大米产后的主要研究课题。从大米的色泽、形状等外部基本特征到蛋白质含量、直链淀粉含量等内部品质, 都得到了广泛的研究。国内外学者已在大米蛋白质含量<sup>[1-2]</sup>、直链淀粉含量<sup>[3-4]</sup>、淀粉纯化工艺<sup>[5]</sup>等方面做过一些有价值研究, 但在大米品种鉴别方面的研究则相对较少。传统的大米品种鉴别多采用化学方法, 既费时又对大米样本具有破坏性, 研究一种简单、快速、无损的大米品种鉴别方法, 具有很高实用性。

可见-近红外光谱技术分析具有成本低、效率高、速度快、检测方便、测试重现性好等优点, 能利用全谱段或多波长下的光谱数据对物质进行定性或定量分析, 已经在石油化工、食品工业等领域被广泛地应用, 国内外也有学者利用可见-近红外光谱技术鉴别咖啡品种<sup>[6]</sup>, 玉米品种<sup>[7]</sup>、瓜品种<sup>[8]</sup>、制动液品牌<sup>[9]</sup>等。但是可见-近红外波谱段具有波谱重叠、信息量大、处理困难等弱点, 判别分析时若直接运用原始光谱数据往往会导致模型稳定性差、精度低, 直接影响分析结果。本文采用主成分分

析法先对大米品种进行聚类分析, 再通过小波变换对波长进行压缩, 并结合 BP 神经网络对大米品种进行鉴别, 提高了鉴别的准确性。一般的研究常用长波(1 300~2 500 nm 波长)近红外光谱技术<sup>[10-11]</sup>, 本文探索利用短波(325~1 075 nm 波长)近红外光谱技术进行大米品种鉴别。

## 1 材料与方法

### 1.1 仪器设备

试验使用美国 ASD (analytical spectral device) 公司的 Handheld Field Spec 光谱仪, 其光谱采样间隔(波段宽) 3 nm, 测定范围 325~1 075 nm, 扫描次数 3 次, 探头视场角为 30°。光源是与光谱仪配套的 14.5 V 卤素灯。光谱数据以 ASCLL 码形式导出进行处理, 分析软件为 ASD View Spec Pro, Unscramble V9.1 和 DPS (data procession system for practical statistics)。

### 1.2 样品来源及光谱的获取

在商店里选购 3 种不同产地的大米, 东北大米、江苏大米、安徽大米。为保证试验结果具有代表性, 每种大米选用 10 个不同生产批次和日期。大米用直径为 65 mm 高度为 14 mm 的圆形透明容器盛装, 装满 1 个容器作为 1 个试验样本。每种大米按不同生产批次和日期均匀选取 60 个样本, 3 种大米共 180 个样本, 样本随机分成预测集和建模集, 预测集有 30 个样本(每个品种 10 个样本), 建模集有 150 个样本(每个品种 50 个样本)。如图 1 所示, 光谱仪经 20 min 预热并进行白板校准后开始测试。卤素灯置于大米样本的侧上方 300 mm 处, 光线经大米样本反射至光谱仪探头, 光谱仪对每一个样本扫描 3 次, 并保存其采样反射光谱平均值。

收稿日期: 2008-07-08 修订日期: 2009-07-14

基金项目: 国家高技术研究发展计划(863 计划)项目(2007AA10Z210); 国家自然科学基金项目(30671213)

作者简介: 周子立(1963—), 男, 高级工程师, 主要从事计算机信息处理及计算机应用技术方面研究。杭州 浙江机电职业技术学院, 310053。

Email: zhouzili88@163.com

\*通信作者: 何 勇(1963—), 男, 教授、博士生导师, 主要从事数字农业与 3S 技术方面的研究。杭州 浙江大学生物系统工程与食品科学学院, 310029。Email: yhe@zju.edu.cn

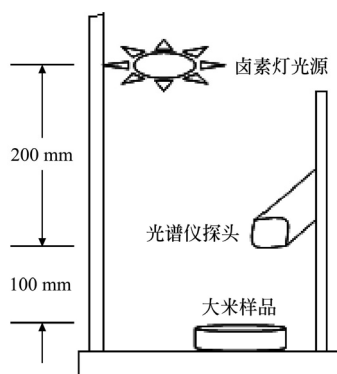


图1 试验样本测试情况

Fig.1 Test of rice samples in the experiment

### 1.3 预处理光谱数据

要消除样本不均匀、基线漂移、高频随机噪音、光散射等影响,必需要对采集到的光谱数据进行预处理。本试验采用移动平均平滑法,先选用平滑窗口尺寸为9,这样可以很好地过滤各种原因所带来的高频噪音,然后再进行多元散射校正处理。由系统误差引起的光谱曲线在首尾端较大的噪声会直接影响试验的准确性,所以去掉前后各75个波段,只研究400~1 000 nm范围内的反射光谱,以提高试验的精度。

### 1.4 人工神经网络模型

人工神经网络是一个由大量简单处理单元(神经元)广泛互连而成的具有自学习、自适应和自组织性的非线性动力学系统,它有着巨量并行性、存储分布性、结构可变性、高度非线性、自学习性及自适应性等特点。BP神经网络模型是一个强有力的学习系统,能够实现输入与输出之间的高度非线性映射<sup>[12]</sup>,在信号和模式识别中的应用非常广泛。它利用实际输出和期望输出之差对网络的多层连接权重由后向前逐层进行校正,而基于BP学习算法的多层前馈网络因具有逼近任意非线性映射<sup>[13]</sup>的功能和很强的学习能力而得到广泛应用。本文在分类和识别的设计中,选用了BP神经网络。

## 2 结果与讨论

### 2.1 样本的可见光-近红外反射光谱

从图2所示的3种大米典型可见光-近红外反射光谱曲线中可以看出,不同品种大米的光谱曲线具有各自的特征性,差异明显。这一差异为大米的不同品种鉴别奠定了数学基础。把3个品种共180个样本做主成分分析,应用主成分分析法对不同品种进行聚类分析。

### 2.2 主成分分析对不同品种大米进行聚类

主成分分析可以降低光谱数据维数,消除众多数据信息中相互重叠的信息部分。通过对原始光谱数据变量进行转换,使数目较少的新变量成为原变量的线性组合,而且,新变量能最大限度的表征原变量的数据结构特征,并最大限度地保留原有信息。对3种大米样本的主成分分析表明,3种大米的前2个主成分累计可信度已达98.819%,仅用前2个主成分就可以表示大米原始光谱的

主要信息<sup>[14]</sup>。图3所示的大米样本主成分1和2的空间得分图对3种大米有较好的聚类作用,可以直观地对不同品种大米进行定性分析。图中横坐标表示每个样本的第一主成分得分值,纵坐标表示每个样本的第二主成分得分值。图3中江苏、安徽、东北3个品种大米明显分布在不同区域,说明主成分1、2对3种大米有较好的聚类作用。从图3中可以看出,江苏大米的60个样本聚合度很好,集中分布在第一象限( $X>0, Y>-0.1$ )。东北大米的60个样本聚合度也很好,主要分布在第二象限( $X<0, Y>-0.1$ )。而安徽大米的60个样本则位于Y轴的下方( $-0.1$ 以下)。

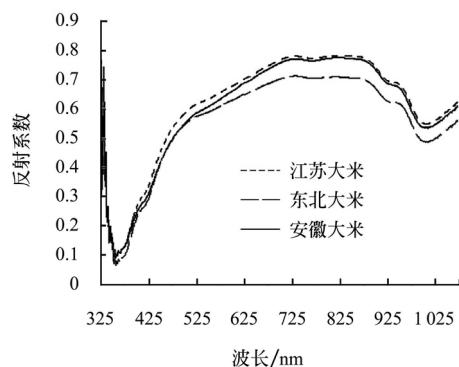


图2 3种大米的典型可见光-近红外光谱反射图

Fig.2 Near infrared reflectance spectra of three different varieties of rice

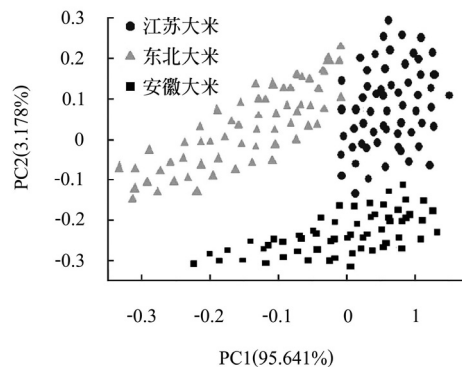


图3 大米样本的主成分1和主成分2的得分图

Fig.3 PCA score plots( PC1×PC2) for rice samples

### 2.3 基于小波分析提取特征信息建立BP品种预测模型

小波变换的本质是信号的时间-尺度分析方法,具有多分辨率分析的特点,而且在时域、频域都具有表征信号局部特征的能力,是20世纪末应用数学界最杰出的成果之一。图像经过小波变换后,其低频部分会保留绝大部分信息和能量。而在图像的轮廓线,突出点等敏感部位,经小波变换后所生成的特征矢量的模值也相对较大。

本文选用的小波为 Daubechies (db2) 正交小波基函数。小波变换先把信号分解成低频 L1 和低频 H1 两部分,而在下一层的分解中,又将上一层的低频 L1 分解成低频 L2 和低频 H2 两部分,以此类推下去,可以得到更深层

次的分解。对图 4 中原始光谱数据进行二级 db2 小波变换, 可得到图 5 图像, 比较图 4 和图 5 可知, 经过五级变换后的低频信号, 大约 1/32 的数据, 包含了原始光谱几乎全部重要的信息。而其他高频信号包含的大量冗余和杂质, 在品种识别中, 起到的作用非常微弱。本文采用小波变换对 180 个样本, 每个样本 600 个光谱数据进行压缩得到新的变量, 新的变量空间从 600 维下降到 21 维, 数据维数大大降低。将样本随机分为预测集和建模集, 预测包含 30 个样本, 建模集包含 150 个样本。把新变量作为神经网络的输入, 品种参数 (1, 2, 3) 作为神经网络的输出分别代表不同的大米品种, 建立品种鉴别模型。各层传递函数选用 S 型 (Sigmoid) 函数, 网络输入层节点数设为 21, 目标误差为 0.00001, 网络指定参数中最小学习速率为 0.1, 设定最大迭代次数为 1 000 次。通过调整隐含层的节点数来优化网络结构<sup>[15-16]</sup>, 经过反复试验得到最佳网络结构为 21 (输入) -12 (隐含) -1 (输出) 3 层 BP 神经网络模型, 对 150 个建模样本的拟合残差为  $9.863 \times 10^{-6}$ , 对未知的 30 个样本进行预测, 预测准确率为 100% (见表 1)。

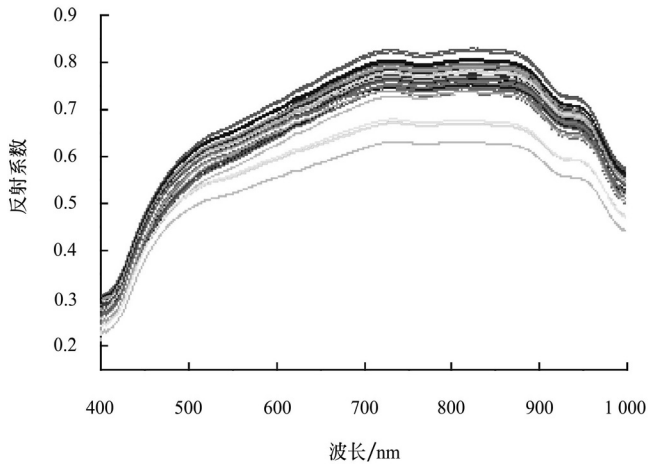


图 4 小波分析前的样本原始光谱

Fig.4 Original spectra of samples before wavelet transform

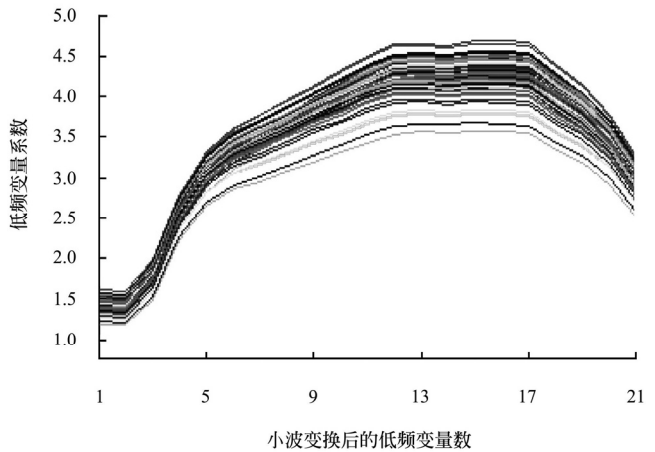


图 5 小波分析提取的特征信息光谱图

Fig.5 Spectra of characteristic information after wavelet transform

表 1 人工神经网络预测 30 个未知大米样本的品种  
Table 1 Prediction results for unknown rice samples by BP model

预测样本 序号	真实值	预测值	预测样本 序号	真实值	预测值
(1)	1	0.99262	(16)	2	1.96587
(2)	1	1.00873	(17)	2	2.00378
(3)	1	0.97656	(18)	2	2.01285
(4)	1	0.98737	(19)	2	2.02346
(5)	1	0.97885	(20)	2	2.00873
(6)	1	1.02376	(21)	3	3.01325
(7)	1	0.99237	(22)	3	2.97365
(8)	1	1.04276	(23)	3	3.00287
(9)	1	1.01267	(24)	3	2.89536
(10)	1	0.96883	(25)	3	3.00878
(11)	2	1.93877	(26)	3	3.01579
(12)	2	1.97827	(27)	3	2.97281
(13)	2	2.01328	(28)	3	2.98763
(14)	2	2.02568	(29)	3	2.89634
(15)	2	1.98639	(30)	3	2.99672

注：1—江苏大米；2—安徽大米；3—东北大米。

3 结 论

- 1) 应用可见-近红外光谱技术可以准确、快速、无损的对大米品种进行鉴别, 所建立的大米品种鉴别模型对大米样本预测效果很好, 识别率达到 100%。
- 2) 组合小波变换、主成分分析、BP 神经网络的模式鉴别方法, 不仅能对不同品种大米进行定性聚类分析, 而且能够定量的预测出未知样本的品种, 方法有效。
- 3) 小波变换得到的 3 种大米特征光谱信息作为人工神经网络的输入, 所建立的品种鉴别模型识别率很高。小波分析不仅能提取 3 种大米敏感的光谱特征信息, 而且能够有效的降低高维空间数据, 与神经网络结合而成的新方法是一种有效的数据挖掘模式和识别方法。

[参 考 文 献]

[1] Peter R Shewry. Improving the protein content and composition of cereal grain[J]. Journal of Cereal Science, 2007, 11(3): 239—250.

[2] 毕京翠, 张文伟, 肖应辉, 等. 应用近红外光谱技术分析稻米蛋白质含量[J]. 作物学报, 2006, 32(5): 709—715. Bi Jingcui, Zhang Wenwei, Xiao Yinghui, et al. Analysis for protein content in rice by NIR spectroscopy technique[J]. Acta Agronmica Sinica, 2006, 32(5): 709—715. (in Chinese with English abstract)

[3] 陆艳婷, 金庆生, 叶胜海, 等. 应用近红外光谱技术快速测稻米的直链淀粉含量[J]. 中国粮食学报, 2007, 22(3): 339—341. Lu Yanting, Jin Qingsheng, Ye Shenghai, et al. Application of NIR spectroscopy for rapid analysis of amylose content in short-grain rice[J]. Journal of the Chinese Cereals and Oils Association, 2007, 22(3): 339—341. (in Chinese with English abstract)

- abstract)
- [4] 刘建学, 吴守一, 方如明. 大米直链淀粉含量的近红外光谱分析[J]. 农业工程学报, 2000, 16(3): 94—96.  
Liu Jianxue, Wu Shouyi, Fang Ruming. Determination of apparent amylose content in rice by Neural Networks based on Near Infrared Spectroscopy[J]. Transactions of the CSAE, 2000, 16(3): 94—96. (in Chinese with English abstract)
- [5] 陈季旺, 刘英, 刘刚, 等. 大米淀粉纯化工艺及其性质的研究[J]. 农业工程学报, 2007, 23(9): 225—228.  
Chen Jiwang, Liu Ying, Liu Gang, et al. Purification and characteristics of rice starch[J]. Transactions of the CSAE, 2007, 23(9): 225—228. (in Chinese with English abstract)
- [6] Esteban-Diez, Gonzalez-Saiz J M, Pizarro C. An evaluation of orthogonal signal correction methods for the characterisation of arabica and robusta coffee varieties by NIRS[J]. Analytica Chimica Acta, 2004, 514: 57—67.
- [7] 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—1808.  
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]. Spectroscopy and Spectral Analysis, 2008, 28(8): 1806—1808. (in Chinese with English abstract)
- [8] Zsolt Seregely, Tamas Deak, Gyorgy Denes Bisztray. Distinguishing melon genotypes using NIR spectroscopy[J]. Chemometrics and Intelligent Laboratory Systems, 2004, 72: 195—203.
- [9] 蒋璐璐, 谈黎虹, 裴正军, 等. 基于可见-近红外光谱的制动液品牌鉴别方法研究[J]. 光谱学与光谱分析, 2008, 28(6): 1296—1299.  
Jiang Lili, Tan Lihong, Qiu Zhengjun, et al. Discrimination of varieties of brake fluid using visual-near infrared spectra[J]. Spectroscopy and Spectral Analysis, 2008, 28(6): 1296—1299. (in Chinese with English abstract)
- [10] 于海燕, 应义斌, 刘燕德. 农产品品质近红外光谱分析结果影响因素研究综述[J]. 农业工程学报, 2005, 21(11): 160—163.  
Yu Haiyan, Ying Yibin, Liu Yande. Research progress of factors affecting the determination of the inner quality of agricultural products by near infrared reflectance spectroscopy[J]. Transactions of the CSAE, 2005, 21(11): 160—163. (in Chinese with English abstract)
- [11] 赵杰文, 张海东, 刘木华. 利用近红外慢反射光谱技术进行苹果糖度无损检测的研究[J]. 农业工程学报, 2005, 21(3): 162—165.  
Zhao Jiewen, Zhang Haidong, Liu Muhua. Non-destructive determination of sugar contents of apples using near infrared diffuse reflectance[J]. Transactions of the CSAE, 2005, 21(3): 162—165. (in Chinese with English abstract)
- [12] 吴春胤, 张文昭, 欧阳庆, 等. 基于 BP 神经网络模型的荔枝树叶面积测定方法[J]. 农业工程学报, 2007, 23(7): 166—169.  
Wu Chunyin, Zhang Wenzhao, Ouyang Qing, et al. BP neural network model for the measurement of the leaf area of litchi[J]. Transactions of the CSAE, 2007, 23(7): 166—169. (in Chinese with English abstract)
- [13] 雷能忠, 王心源, 蒋锦刚, 等. 基于 BP 神经网络插值的土壤全氮空间变异[J]. 农业工程学报, 2008, 24(11): 130—134.  
Lei Nengzhong, Wang Xinyuan, Jiang Jingang, et al. Spatial variability of soil nitrogen by BP neural network interpolation[J]. Transactions of the CSAE, 2007, 23(7): 130—134. (in Chinese with English abstract)
- [14] 岳田利, 彭帮柱, 袁亚宏, 等. 基于主成分分析法的苹果酒香气质量评价模型的构建[J]. 农业工程学报, 2007, 23(6): 223—227.  
Yue Tianli, Peng Bangzhu, Yuan Yahong, et al. Modeling of aroma quality evaluation of cider based on principal component analysis[J]. Transactions of the CSAE, 2007, 23(6): 223—227. (in Chinese with English abstract)
- [15] 逮家辉, 张益波, 张卓勇, 等. 小波变换近红外光谱结合径向基神经网络快速分析异福片[J]. 光谱学与光谱分析, 2008, 28(6): 1264—1268.  
Lu Jiahui, Zhang Yibo, Zhang Zhuoyi, et al. Application of wavelet transform-radial basis function neural network in NIRS for determination of rifampicin and isoniazide tablets[J]. Spectroscopy and Spectral Analysis, 2008, 28(6): 1264—1268. (in Chinese with English abstract)
- [16] 赵琛, 瞿海斌, 程翼宇. 虫草氨基酸的人工神经网络-近红外光谱快速测定方法[J]. 光谱学与光谱分析, 2004, 24(1): 50—53.  
Zhao Chen, Zhai Haibin, Cheng Yiyu. A new approach to the fast measurement of content of amino acids in cordyceps sinensis by ANN-NIR[J]. Spectroscopy and Spectral Analysis, 2004, 24(1): 50—53. (in Chinese with English abstract)

## Method for rapid discrimination of varieties of rice using visible NIR spectroscopy

Zhou Zili<sup>1,3</sup>, Zhang Yu<sup>2</sup>, He Yong<sup>3\*</sup>, Li Xiaoli<sup>3</sup>, Shao Yongni<sup>3</sup>

(1. *Zhejiang Institute of Mechanical and Electrical Engineering, Hangzhou 310053, China;*

2. *Zhejiang Technology Institute of Economy, Hangzhou 310018, China;*

3. *College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310029, China)*

**Abstract:** Based on the visible-near infrared spectroscopy (Vis-NIRS) technology, a new method to discriminate varieties of rice was proposed. First, the clustering of varieties of rice was analyzed by principal component analysis (PCA). Second, characteristics information of spectra were extracted by wavelet transform (WT), which as input sets for artificial neural network (ANN) to discriminate rice varieties of rice. And then a total of 180 (60 in each category) samples of three categories were adopted in this study, with 150 (50 in each category) for training sets and the remaining 30 (10 for each category) for prediction sets. The experimental results show that the identification rate reached 99.3%, which proves that the new method proposed in this study is capable to discriminate the varieties of rice with high accuracy. In addition, it might provide a new method to discriminate rice varieties.

**Key words:** Vis-NIRS, rice, principal component analysis, wavelet transform, artificial neural network