

## 近红外光谱分析技术在茶叶鉴别中的应用研究

赵杰文<sup>1</sup>, 陈全胜<sup>1</sup>, 张海东<sup>1, 2</sup>, 刘木华<sup>1, 3</sup>

1 江苏大学食品与生物工程学院, 江苏 镇江 212013

2 云南农业大学工程技术学院, 云南 昆明 650201

3 江西农业大学工学院, 江西 南昌 330045

**摘要** 茶叶快速准确鉴别方法研究是当前茶叶行业亟待解决的一个重要课题。该研究采用近红外光谱结合主成分-马氏距离模式识别方法鉴别了龙井、碧螺春、毛峰和铁观音4种中国名茶。研究表明,在6500~5300 cm<sup>-1</sup>波数范围内的光谱,通过MSC预处理方法,用8个主成分建立的模型最好,模型对校正集样本和预测集样本的鉴别率分别达到98.75%和95%。该研究为快速准确鉴别茶叶提供了一种新思路。

**主题词** 近红外光谱; 主成分分析; 马氏距离; 茶叶识别

**中图分类号:** O657.3 **文献标识码:** A **文章编号:** 1000-0593(2006)09-1601-04

### 引言

中国是茶叶的故乡,盛产多种多样品种的茶叶。目前中国的茶叶市场相对混乱,特别在名优茶市场,以次充好以假乱真的现象比较严重,这既损害了消费者的利益,也不利于中国茶叶品牌的保护。研究快速、准确的鉴别方法,对于维护中国茶叶品牌,提高茶叶品质有着直接的现实意义。

传统的茶叶鉴别方法是感官评定法和化学方法。感官评定的结果受人为因素和外界环境的干扰很大,影响到结果的客观性;化学方法虽然能够准确地鉴别茶叶,但是繁琐的步骤和昂贵的费用使它不能应用到茶叶的快速鉴别上。近红外光谱(NIR)分析具有速度快、成本低以及重现性好等优点。国内外学者先后利用近红外光谱方法定性和定量地分析了茶叶中蛋白质、咖啡碱、氨基酸、多酚类以及水分的含量<sup>[1-6]</sup>,但是近红外光谱方法在茶叶鉴别上的应用研究还很少。近红外光谱结合模式识别的方法在石油<sup>[7]</sup>和中草药<sup>[8,9]</sup>的鉴别和分类上得到了许多成功地应用,本研究尝试了将此方法应用到茶叶的快速鉴别中,希望这种方法能成为一种茶叶快速鉴别的新技术得到广泛地应用。

## 1 材料与方法

### 1.1 实验材料

实验所用的材料是龙井、碧螺春、毛峰和铁观音4种中

国名茶。为使取样均匀,试验前先将每一个品种的茶叶分别用咖啡粉碎机粉碎过40目筛,然后在每一个品种的茶叶中,按照四分法原则,随机称取5g作为一个样本。每种品种的茶叶分别各选20个样本作为校正集,10个样本作为预测集。实验中所用茶叶均购自安徽省茶叶进出口公司,出产日期都在2004年5~7月。样本材料的名称,产地和样本数量见表1。

Table 1 Origin and number of samples

名称	产地	校正集数量	预测集数量
龙井	浙江	20	10
碧螺春	江苏	20	10
毛峰	安徽	20	10
铁观音	福建	20	10
总计		80	40

### 1.2 光谱采集

实验所用的近红外检测系统主要是近红外光谱仪(Nexus 670 FTIR, 美国 Nicolet 公司)。扫描范围:4000~11000 cm<sup>-1</sup>;扫描次数:64次;分辨率:4 cm<sup>-1</sup>。实验时,保持室内的温度和湿度基本一致,将样本倒入样品杯中,充分压实。每一个样本在不同时间,不同位置分别采集4次,取4次采集的平均值作为该样本的原始光谱。

### 1.3 分析方法<sup>[10,11]</sup>

本研究中采用的分析方法是主成分分析结合马氏距离的

收稿日期:2005-06-16, 修订日期:2005-09-28

基金项目:国家高技术“863”计划(2002AA248051)和国家自然科学基金(30370813)资助项目

作者简介:赵杰文,1945年生,江苏大学生物与环境工程学院教授

模式识别方法。具体步骤如下:

利用下列公式(1)对校正集中的光谱矩阵  $X_{n \times m}$  进行主成分降维, 得到主成分得分矩阵  $T_{n \times f}$  和载荷矩阵  $P_{m \times f}$ ; 按公式(2)与(3)将得分矩阵  $T$  进行中心化处理; 按公式(4)计算校正集中样本中的协方差矩阵  $M$ ; 最后由公式(5)计算校正集中样本的马氏距离。

$$T_{n \times f} = X_{n \times m} \times P_{m \times f} \quad (1)$$

在公式(1)中  $n$  为校正集中的样本数;  $m$  为变量数;  $f$  为主成分数

$$\bar{T}_i, : j = \bar{T}_j = \frac{\sum_{i=1}^n T_{ij}}{n} \quad (2)$$

在公式(2)中  $i = 1, 2, \dots, n; j = 1, 2, \dots, f$

$$T_c = T - \bar{T} \quad (3)$$

$$M = \frac{T_c^T T_c}{n-1} \quad (4)$$

$$MD_i = \sqrt{(t_i - \bar{T}) M^{-1} (t_i - \bar{T})} \quad (5)$$

在公式(3~5)中  $\bar{T}$  为  $n$  个校正集中样本的平均得分矩阵;  $T_c$  为  $T$  的均值中心化矩阵;  $t_i$  为校正集中样本  $i$  的光谱得分向量;  $MD_i$  为校正集中样本  $i$  的马氏距离;  $M$  为校正集中样本中的协方差矩阵。

根据校正集中样本的马氏距离和误差允许的范围, 确定判别未知样本属于第  $k$  类样本群体 ( $G_k$ ) 的域值 ( $D_k$ )。对于未知样本的光谱矢量  $x_{unk}$ , 利用校正集中光谱的载荷矩阵  $P$ , 计算其得分向量  $t_{unk}$ , 再由  $t_{unk}$  计算其马氏距离  $D_{unk}$ 。判别规则如下

当时:  $D_{unk} < D_k$  时:  $x_{unk} \in G_k$ ;

当时:  $D_{unk} > D_k$  时:  $x_{unk} \notin G_k$ ;

当时:  $D_{unk} = D_k$  时: 重新选定  $D_k$ 。

所有的数据分析都是基于 TQ Analysis V6(Nicolet 近红外系统自带)和 Matlab V6.5 的软件平台。

## 2 结果与讨论

### 2.1 波长范围的选择

图 1 是 4 种茶叶的原始光谱图 (a) 和一阶导数光谱图 (b)。从图 1(a) 中可以看出原始光谱在波数为  $5155 \text{ cm}^{-1}$  (波长  $1940 \text{ nm}$ ) 的附近有一个明显的吸收峰, 从图 1(b) 中可以看一阶导数光谱在波数为  $5155 \text{ cm}^{-1}$  (波长  $1940 \text{ nm}$ ) 和  $6944 \text{ cm}^{-1}$  (波长  $1440 \text{ nm}$ ) 的附近有明显的波动。因为纯水中的 O—H 伸缩振动的一级基频区在  $1440 \text{ nm}$  附近, 它的一个合频区在  $1940 \text{ nm}$  附近, 在这两个波长附近是水分吸收的敏感区, 从图 1 中可以看出在这两个区域, 水分对茶叶的近红外图谱的影响还是很大的。本研究所用的样本都是干茶(一般干茶中的水分含量在 5% 左右), 为了减少水分的影响, 选择光谱波长范围尽量避开水分吸收峰的特征波长区。本研究有比较地选用了各段的波长进行了分析, 结果显示选用  $6500 \sim 5300 \text{ cm}^{-1}$  (波长  $1538 \sim 1887 \text{ nm}$ ) 范围内的光谱数据既避开了水分的影响, 且取得了较好的实验结果。

### 2.2 模型的建立

虽然茶叶样本是经过粉碎后过 40 目筛得到的, 但是样本颗粒的大小仍然会对光的漫反射有一定影响; 同时样本的密度也影响了光在样品中的传播。因此, 需要对样本的原始光谱数据进行预处理。不同的预处理方法对所建的模型有一定的影响, 本研究对多元散射校正 (MSC)、标准归一化 (SNV)、一阶导数和二阶导数等几种预处理方法进行了比较。

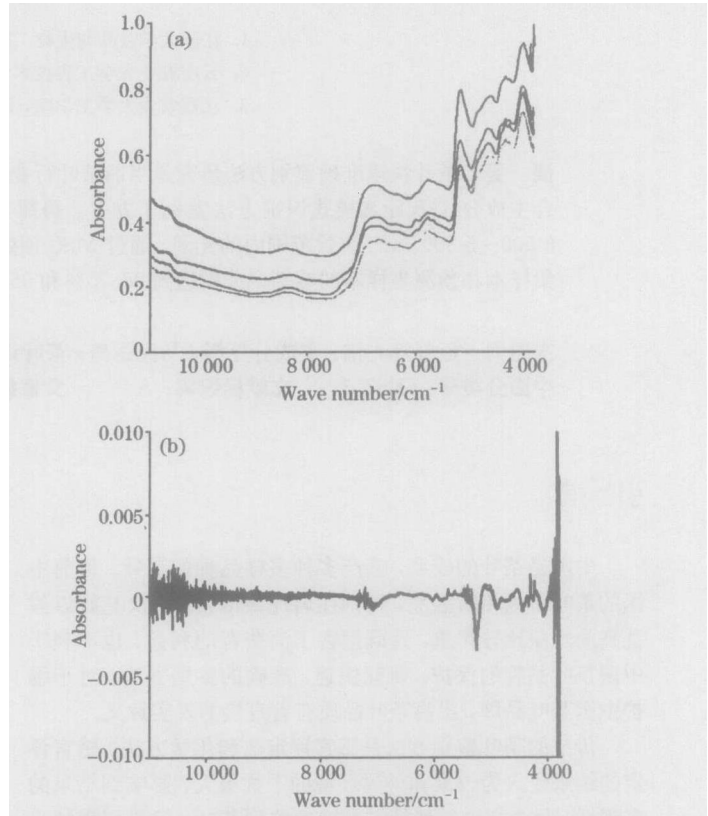


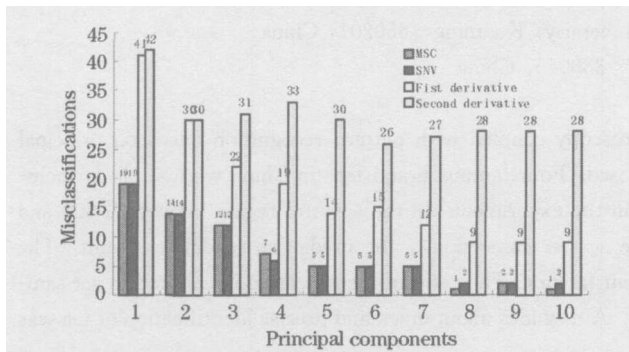
Fig 1 The original NIR spectra (a) and first derivative NIR spectra (b) of four different teas

由 1.3 节中的分析方法可以知道校正过程中主成分数的多少对模型的建立也有一定的影响。我们将误判数作为衡量模型优劣的一个指标, 首先筛选出校正集中误判数较少的几种模型, 用这几种模型分别对预测集中的样本进行预测, 比较预测集中的误判数, 最后, 校正集和预测集中误判数都相对较少的模型即为最佳模型。

图 2 表示了基于不同的预处理方法, 不同的主成分数时校正集的误判数。由图 2 可以看出用多元散射校正 (MSC) 略好于标准归一化 (SNV) 的预处理方法, 两者都明显优于一阶导数和二阶导数预处理方法。主成分数在 8~10 范围误判概率最小, 并且变化不大。所以本研究采用 MSC 预处理方法, 分别选用了 8~10 个主成分建立模型对预测集样本进行预测, 预测的结果表明选用 8 个主成分建立的模型预测效果最好, 预测集中只有两个样本发生误判。但用 9 个主成分时预测集中出现 4 个误判, 用 10 个主成分时预测集中出现 5 个误判。

**Table 2 The results of calibration and prediction**

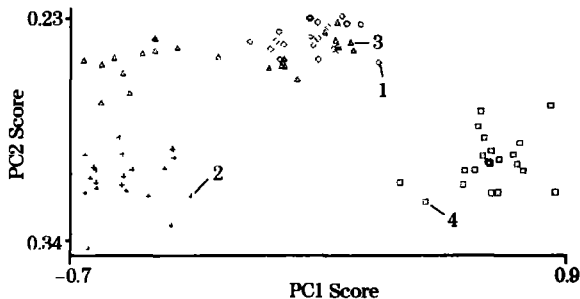
	校正集	预测集
样本数	80	40
鉴别率 / %	98.75	95.00
误判的情况	1 个龙井样本误判成碧螺春样本	2 个龙井样本误判成碧螺春样本



**Fig 2 The relationship between misclassifications and principal components in calibration set with four preprocessing methods**

**2.3 校正和预测的结果**

根据以上的分析, 选用 6 500 到 5 300  $\text{cm}^{-1}$  范围内的光谱数据, 经过多元散射校正 (MSC) 预处理, 选用 8 个主成分数建立预测模型最佳。模型校正和预测结果见表 2。从校正和预测的结果看, 在校正集中有 1 个龙井样本误判为碧螺春, 在预测集中, 有 2 个龙井样本误判为碧螺春。产生这种结果的原因从图 3 也可以看出, 图 3 是校正集光谱矩阵第一



**Fig 3 Score cluster plot using first and second principal components (PCs)**

1: Longjing; 2: Maofeng; 3: Biluochun; 4: Tieguanynin

主成分和第二成分的得分图, 它表明校正集中的样本点在该二维平面上的投影。在图 3 中, 碧螺春样本在该二维平面的投影点比较分散, 它们与龙井样本的投影点之间有相互交叉的区域。也就造成了将少数龙井样本误判为碧螺春的结果。

**3 结论与展望**

本研究利用近红外光谱结合模式识别的方法鉴别了龙井、碧螺春、毛峰和铁观音 4 种中国名茶, 取得了满意的结果, 实验证明了近红外光谱在茶叶鉴别上的可行性。在本研究中模型的建立基本是利用了线性模式识别方法, 结果不一定能做到 100% 的正确, 所以在以后的工作中还可以尝试利用其他非线性的模式识别方法来完善模型, 使结果更准确。

**参 考 文 献**

[ 1 ] Yan S H, Meurens M, Dufour J P. Proceedings of the 2<sup>nd</sup> International NIRS Conference, Tsukuba, Japan, 1990. 166.  
 [ 2 ] Goto T, Mukai T, Horie H, et al. Proceedings of the International Symposium on Tea Science, Shizuoka, Japan, 1991. 81.  
 [ 3 ] Ikegagy, K. Japan Agricultural Research Quarterly, 1990, 24(1): 49.  
 [ 4 ] Lupaert J, Zhang M H, Massart D L. Analytica Chimica Acta, 2003, 487(2): 303.  
 [ 5 ] Zhang M H, Lupaert J, Fernandez J A, et al. Talanta, 2004, 62(1): 25.  
 [ 6 ] Schulz H, Engelhardt U H, Wengent A, et al. J. Agric Food Chem., 1999, 47: 5064.  
 [ 7 ] WANG Li, ZHUO Lin, HE Ying, et al(王 丽, 卓 林, 何 鹰, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2004, 24(12): 1537.  
 [ 8 ] ZHOU Qun, SUN Su-qin, LEUNG H+win(周 群, 孙素琴, 梁曦云). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2003, 23(3): 509.  
 [ 9 ] Candolfi A, De Maesschalck R, Massart D L, et al. Journal of Pharmaceutical and Biomedical Analysis, 1999, 19(6): 923.  
 [ 10 ] LU Wan-zhen, YUAN Hong-fu, XU Guang-tong, et al(陆婉珍, 袁洪福, 徐广通, 等). The Modern Analysis Technique for Near-Infrared Spectra(现代近红外光谱分析技术). Beijing: Chinese Oil and Chemical Industry Press(北京: 中国石化出版社), 2000.  
 [ 11 ] XU Lu, SHAO Xue-guang(许 禄, 邵学广). Methods of Chemometrics(化学计量学方法). Beijing: Science Press(北京: 科学出版社), 2004.

# Study on the Identification of Tea Using Near Infrared Reflectance Spectroscopy

ZHAO Jie-wen<sup>1</sup>, CHEN Quan-sheng<sup>1</sup>, ZHANG Hai-dong<sup>1, 2</sup>, LIU Mu-hua<sup>1, 3</sup>

1. School of Food and Biological Engineering, Jiangsu University, Zhenjiang 212013, China

2. Faculty of Engineering and Technology, Yunnan Agricultural University, Kunming 650201, China

3. Engineering College, Jiangxi Agricultural University, Nanchang 330045, China

**Abstract** A rapid tea identification method by near infrared spectroscopy coupled with pattern recognition based on principal components analysis and Mahalanobis' distance technique was proposed. Four famous brand teas in China were studied, including Longjing tea, Biluochun tea, Maofeng tea and Tieguanyin tea in the experiment. In the spectral region between 6 500 and 5 300  $\text{cm}^{-1}$ , through preprocessing method of MSC(multiplicative scatter correction), the prediction model was built. The result showed that the model was the best with 8 principal component factors. The rates of identification in calibration set samples and prediction set samples were 98.75% and 95%, respectively. A new idea about quick and precise identification of tea was offered.

**Keywords** Near-infrared spectroscopy; Principal components analysis; Mahalanobis' distance; Tea identification

(Received Jun. 16, 2005; accepted Sep. 28, 2005)

中国科技核心期刊  
CODEN: YACEEK

《岩矿测试》

ISSN 0254-5357  
CN 11-2131/TD

欢迎订阅 欢迎投稿 承接广告

《岩矿测试》杂志是中国地质学会岩矿测试专业委员会和国家地质实验测试中心共同主办的分析测试技术科技期刊。国际标准刊号:ISSN 0254-5357; 国际刊名代码 CODEN: YACEEK; 国内统一刊号 CN11-2131/TD。

《岩矿测试》的宗旨是突出服务于地球科学和地质找矿事业以及促进岩矿测试技术的发展;根据国家地质工作的重点由单一资源向资源环境并重的转变,《岩矿测试》的内容有所拓宽,主要报道国内与分析科学、资源环境、地球科学相关的新技术、新方法、新理论和新设备等研究成果、动态、评述及相关实践经验。

《岩矿测试》1982年创刊,为国内外公开发行人刊物。近年来刊物地位不断提高,是中国科技核心期刊,中国期刊方阵双效期刊。目前被国内外多家检索数据库、文摘收录。曾先后被评为国家、原地矿部、北京市、中国科协的优秀科技刊物。适合于地质、冶金、环保、石油、化工、煤炭等部门从事分析测试的科技工作者及大专院校分析化学、环境资源、地球科学等相关专业的师生阅读。

为顺应国家地质工作的根本转变,拓宽地质分析领域,扩大信息容量,缩短发稿周期,适应科技发展和读者的需求,自2007年起,《岩矿测试》由季刊变更为双月刊,大16开版本,80页/期,逢双月出版;国内邮发代号2-313;国际书店发行代号BM4089;广告经营许可证:京西工商广字第0227号;定价10.00元/本,全年60.00元。漏订的读者可直接与编辑部联系。

《岩矿测试》编辑部地址:北京西城区百万庄大街26号国家地质实验测试中心

邮政编码:100037 电话:010-68999562 传真:010-68999563

E-mail: ykcs\_zazhi@163.com