

文章编号:1673-1689(2007)02-0007-05

中红外光谱与模式识别相结合鉴别茶叶种类

冯宇¹, 顾小红², 汤坚^{1,2}, 陈尚卫²

(1. 江南大学 食品科学与安全教育部重点实验室, 江苏 无锡 214122; 2. 江南大学 分析测试中心, 江苏 无锡 214122)

摘要: 采用漫反射傅立叶变换红外(FTIR)光谱法结合主成分分析对苏州“碧螺春”、浙江“龙井”、安徽祁门红茶、福建“铁观音”4种茶叶进行了聚类分析。结果表明,主成分分析结合马氏距离判据的方法对上述4种茶叶可进行鉴别,该法快速、准确,为客观评价茶叶品种提供了一种新的方法。

关键词: 傅立叶变换红外光谱;主成分分析;聚类分析;马氏距离判据;模式识别;茶叶
中图分类号: Q 682 **文献标识码:** A

Discrimination of Tea Varieties by Mid-Infrared Spectroscopy Combined with Pattern Recognition

FENG Yu¹, GU Xiao-hong², TANG Jian^{1,2}, CHENG Shang-wei²

(1. Key Laboratory of Food Science and Safety, Ministry of Education, Southern Yangtze University Wuxi 214122, China; 2. Analysis and Test Centre, Southern Yangtze University, Wuxi 214122, China)

Abstract: Diffusion reflectance FTIR spectroscopy in the mid-IR region combined with principal component analysis have been applied for classification tea from different cultural area, including two green teas (Biluochun from Suzhou and Longjing from Zhejiang), one black tea (Qimen from Anhui) and one semi-fermented tea (Tieguanyin from Fujian). It could be concluded that the four teas could be successfully discriminated by principle component analysis combined with Mahalanobis Distance. As having the fast and accurate characters, it provides a novel way to assess the authenticity of tea varieties.

Key words: FTIR; Principle component analysis; classification analysis; Mahalanobis distance pattern recognition; tea

近年来,红外光谱在食品多组分定量分析^[1,2]、油脂氧化指数测试^[3]、葡萄酒品种鉴别^[4]、咖啡产地识别^[5]等领域都有广泛的应用。我国的孙素琴等将红外光谱和模式识别方法相结合对多种中药材的道地性鉴别进行了研究^[6-8],取得了很好的效

果。茶叶是一个复杂的混合体系,所含化合物的响应信号相互叠加,使得其红外光谱具有难以解析的复杂性。但不同品种的茶叶、不同产地、不同加工工艺和贮存条件,导致茶叶产品在化学组成及含量、各组分的相互配比之间存在差异^[9],其红外光

收稿日期:2006-03-14.

基金项目:国家自然科学基金项目(20436020).

作者简介:冯宇(1980-),男,山西太原人,工学硕士,主要从事食品无损检测方向的研究. Email: f6802528@163.com

谱就具有细微的特征性。借助这种差异,运用模式判别法对其红外光谱进行解析,就可以建立相关的模式识别器,实现茶叶品质的快速识别。作者选择了我国常见的4种茶叶,采用主成分分析法结合马氏距离判据的方法对其红外光谱图进行模式识别,对茶叶品质的红外光谱学研究进行探索。

1 实验器材与方法

1.1 仪器设备和参数设置

采用 Thermo Electron 公司的 Nicolet Nexus470 红外光谱仪,装配 Smart Diffuse Reflection 附件、DTGS 检测器,测定范围 $4\,000\sim 400\text{ cm}^{-1}$,扫描次数 64 次,分辨率 4 cm^{-1} ,动镜速度 12.65 cm/s 。

1.2 样品来源和预处理

样品来源:市售“碧螺春”、“龙井”、祁门红茶、“铁观音”,低温避光储藏。

样品处理^[10]:将茶叶粉碎至 $100\text{ }\mu\text{m}$ 以下^[11],取该茶叶粉末与同样粒度的 KBr 粉末混合均匀。采用 Smart Diffuse Reflection 附件,以同样粒度的 KBr 粉末作为背景,扫描茶叶样品的红外谱图。每份样品扫描 3 次取平均值。实验中对每种茶叶样品采集 70 张光谱图,选择其中信噪比较好的 50 张谱图建模。

1.3 数据处理软件

采用 OMNIC6.2 处理光谱数据,主成分分析和聚类分析采用 TQ Analyst V6 以及 The unscrambler 9.5 化学计量软件。

1.4 主成分分析法

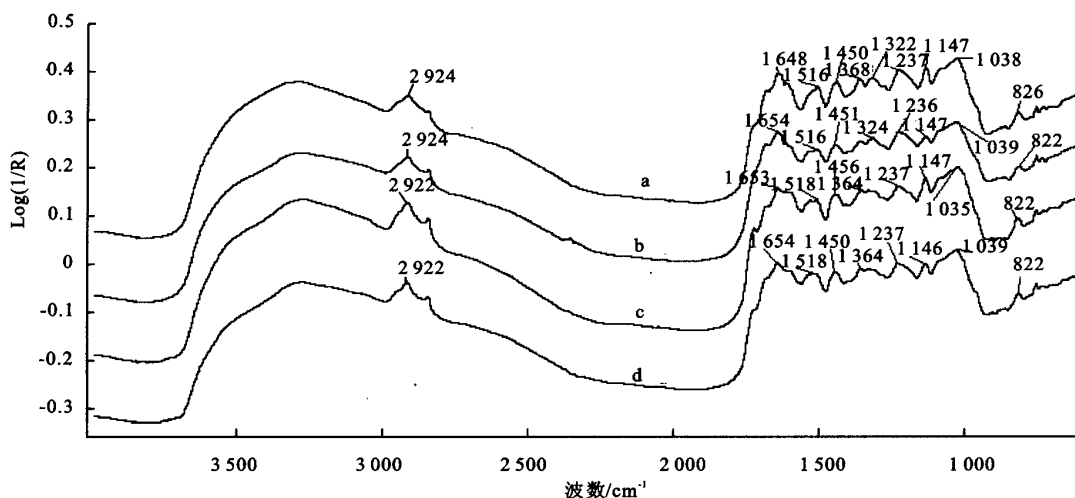
这是一种数据压缩的常用方法,它把原有的各个特征利用线性变化得到一批新的特征,每个特征都是原有特征的函数,但新特征总数少于原有特征数,这样新特征既保留了原有特征的主要信息,又减少了特征个数。本文中为光谱样本,每个样本由多个波长特征描述,这样就构成了一个光谱数据矩阵,通过对新特征的选择,不仅实现了数据的降维,而且还可以在在一定程度上滤除噪音^[12]。

本研究中采用的聚类方法为主成分分析结合马氏距离判据(Mahalanobis Distance)^[13]。即采用光谱矩阵的得分向量来计算马氏距离,可以避免选择过多波长造成过度拟合,同时保留全谱的平均效应。该方法以合格样本建立模型,设定合适的马氏距离判据来实现聚类。

2 结果与讨论

2.1 光谱的预处理

4 种茶叶的漫反射红外光谱如图 1 所示,每种茶叶的光谱图为 50 个样本光谱的平均值。从图中可以看出,光谱的基线都有所倾斜,在建模前选取相对较稳定的光谱较低点 $1\,872\text{ cm}^{-1}$ 处进行单点基线校正。此外,鉴于茶叶的不均一性,以及在样品预处理和测量过程中出现的诸如粒度不均匀、混合不充分、光反射等问题,研究中还将谱图进行了多倍散射校正^[14],从而尽可能地保留茶叶样品结构和组成的信息,最大限度地消除各种误差。



a 碧螺春(Biluochun); b 祁门(Qimen); c 铁观音(Tieguanyin); d 龙井(Longjing)

图1 4种茶叶的红外光谱图

Fig.1 FTIR spectra of four tea

图中 4 种茶叶的光谱图整体轮廓虽然近乎相同,但在指纹区仍然存在一些差异,“碧螺春”的峰形更为尖锐一点。1 147 cm^{-1} 处各茶叶光谱的峰强不同。“铁观音”在 1 653 cm^{-1} 处的峰簇的形状与其它 3 种茶叶有较大差异。总之 1 918 cm^{-1} ~938 cm^{-1} 波段内 4 种茶叶的红外谱图具有一定的特征性。这为它们的模式识别奠定了一定的数学基础。

2.2 光谱波段的选择

为了选择有效光谱差异较大的区域进行聚类,需要对全光谱波段进行筛选。研究中采用下述方法:在整个光谱区域,对全部样本空间进行主成分分析,建立一个主成分模型,然后通过光谱矩阵空间在主成分上的载荷来选取有效的波段建模。上述分析采用 20 个主成分,对全部样本采用留一法交互验证。图 2 就是光谱矩阵在第 1、第 2、第 3 主

成分坐标上的载荷图,前 3 个主成分的贡献率分别为 92.155%,4.005%和 2.447%,前 3 个主成分的累积贡献率为 96.120%。从图中可以看出,除去高频噪音较大的波段,前 2 个主成分上的载荷图 4 000 cm^{-1} ~2 000 cm^{-1} 波段与原茶叶谱图很相似,基本成正或负相关,这是由于该波段主要是样品中羟基的吸收峰,而茶叶中的大多数组分都含有羟基,所以这一波段对于茶叶的聚类贡献不大。而在指纹区的载荷图与原图有明显差异,是由于不同茶叶中各组分的含量及相互比例不同所致,故选取 1 918 cm^{-1} ~938 cm^{-1} 波段建模。此外,从图中可以看出,经过光谱预处理,在全波段内测量误差不是主要变量来源,只在第 3 个主成分的载荷图上可以看到一些,说明了光谱预处理的效果较好。

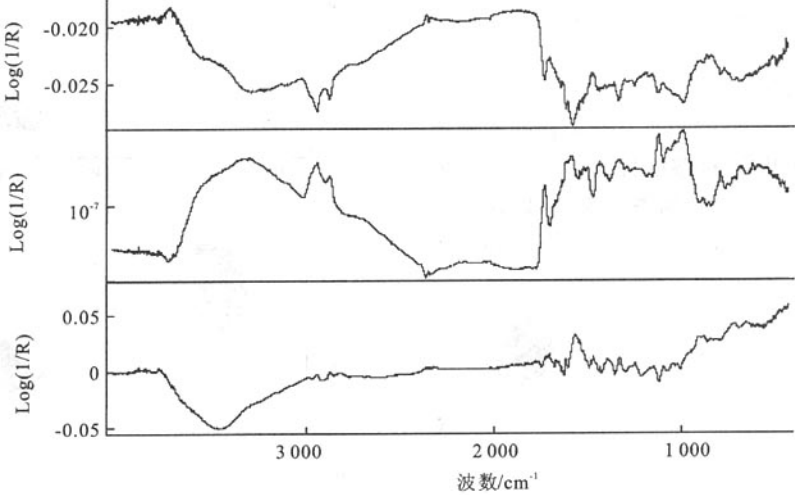
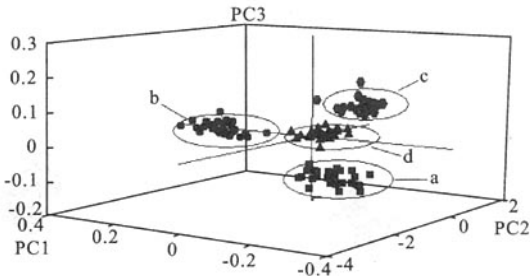


图 2 光谱矩阵在前 3 个主成分上的载荷
Fig. 2 First three PC loadings of spectra matrix

2.3 聚类分析结果

图 3 为 4 种茶叶统一建立主成分分析模型的前 3 个主成分的得分散点。该模型建立时采用上述建模区域的光谱数据,从 200 个样本中随机抽取 50 个作为检验集,由剩下的 150 个样本组成训练集,选用 20 个主成分数,前 3 个主成分的贡献率分别为 96.902 0%、2.575 9%和 1.501 6%,累积贡献率达到 98.979 5%。各茶叶的红外光谱数据经过主成分分析再投影后,显著地聚成一簇,而且各类之间的区域几乎没有重叠。对未知样品,可以由此模型计算其各主成分得分,再根据最大隶属度原则进行判别。

研究中采取的是另一种聚类方法,即每种茶叶分别建立其主成分分析模型。按上述条件对谱图



a 碧螺春 (Biluochun); b 祁门 (Qimen); c 铁观音 (Tieguanyin);
d 龙井 (Longjing)

图 3 4 种茶叶的主成分分析三维图
Fig. 3 3D plot of first versus second versus third PC scores of four tea

进行预处理,采用独立的检验集诊断模型的优劣。研究中在每种茶叶样本中随机抽取 13 个组成检验

集,建模时采用 20 个主成分数,抽取前 3 个主成分建立聚类分析模型,结果如下:

在聚类分析中,一般采用聚类中心之间的距离(类间距)评价聚类效果,而聚类结果好坏的关键则是合适特征的选取。得到类间距后,数值越大,表明类与类之间的差异越明显。4 种不同茶叶的类间距如表 1 所示。

表 1 4 种茶叶样品的类间距
Tab. 1 Inter class—distances of four tea

	祁门红茶	“碧螺春”	“龙井”	“铁观音”
祁门红茶	1	—	—	—
“碧螺春”	17.392 1	—	—	—
“龙井”	23.655	21.619	1	—
“铁观音”	48.866	30.455	15.377	1

下面的组图为 4 种茶叶的聚类结果,图中横竖坐标分别为各样本拟合横竖坐标轴对应的茶叶模型所计算的马氏距离。坐标平面上的两条灰色的正交的线为在 95% 的置信区间上所求得的类别边界。从图中可以看出,4 种茶叶样品的聚类结果较为理想,只要适当选择判别的茶叶模型,则这 4 种茶叶之间可以没有重叠,互不干扰,从而达到满意的判别结果。

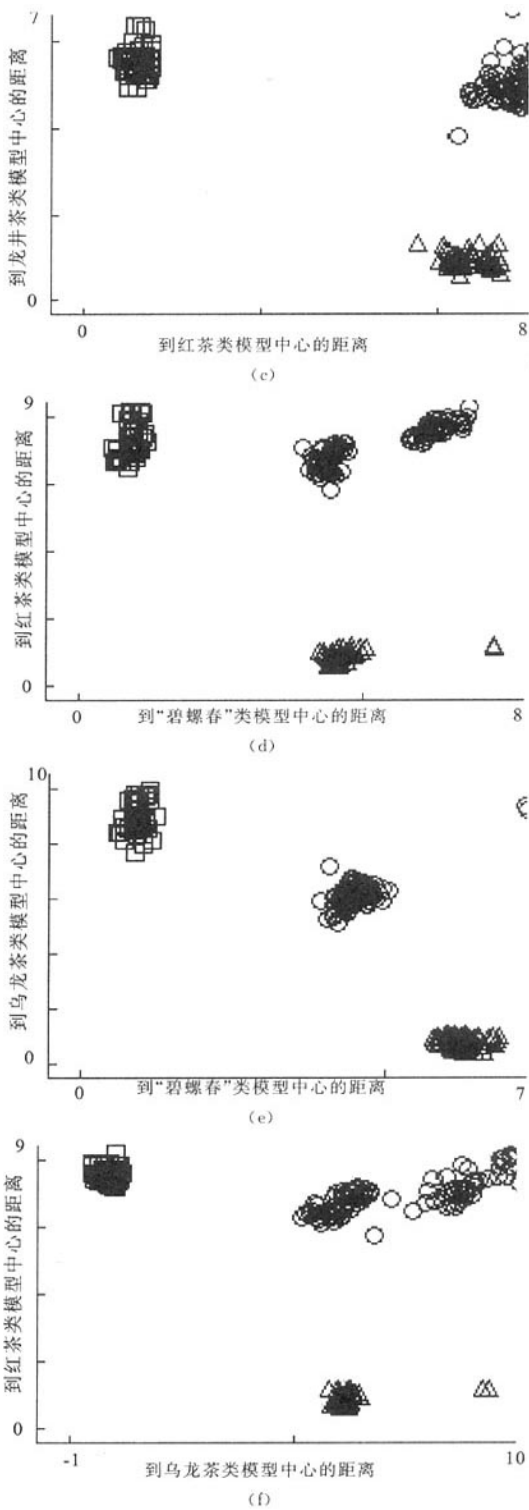
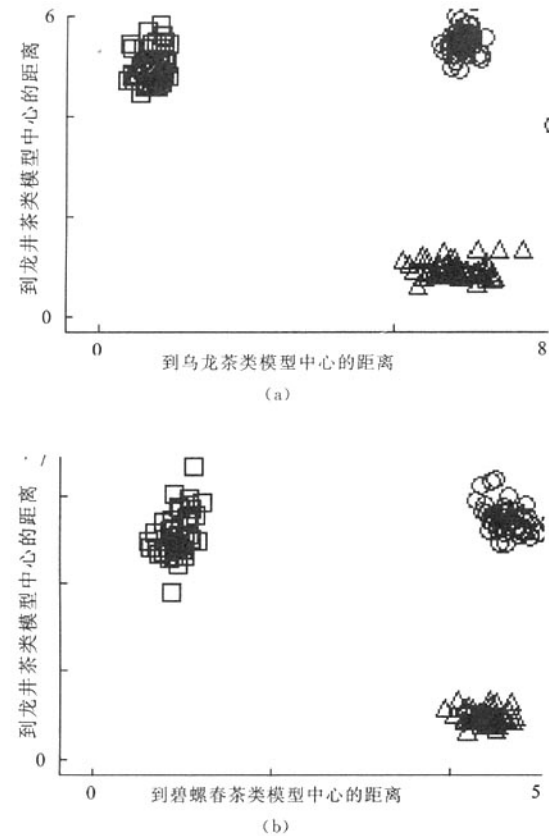


图 4 4 种茶叶的聚类结果
Fig. 4 Classification results of four tea
oolong T “铁观音”主成分分析模型(PCA model of Tieguanyin);
green tea B “碧螺春”主成分分析模型(PCA model of Biluochun);

green tea L“龙井”主成分分析模型(PCA model of Longjing); black tea Q 祁门红茶主成分分析模型(PCA model of Qimen)

3 结 论

红外光谱法制样简便快速,无需繁琐的预处理,给出的是样品的整体信息,因此该方法不仅快速准确公正,而且从一定程度上克服了仪器无法从整体上鉴别茶叶品质的缺点。来自不同产地、不同

加工工艺的4种茶叶,苏州碧螺春、浙江龙井、安徽祁门红茶、福建铁观音的红外光谱图经过多元散射校正的预处理,并应用主成分分析与马氏距离判别方法实现了分类,从模型类型间距以及聚类平面图上看,结果很理想。可以预见,随着采样范围的扩大和采样方法的改进,这种方法提取的样品的共性也将增加,会使得红外光谱结合模式识别来鉴别茶叶品种的方法更为可靠。

参考文献(References):

- [1] José Luís Moreira, Lúcia Santos. Spectroscopic interferences in fourier transform infrared wine analysis[J]. *Analytica Chimica Acta*, 2004, 513: 263—268.
- [2] Luinge H J, Hop E, Lutz E T G. Determination of the fta, protein and lactose content of milk using fourier transform infrared spectrometry[J]. *Analytica Chimica Acta*, 1993, 284: 419—433.
- [3] Ismail A A, van de Voort F R, Emo G. A Rapid FTIR quality control method for fat and moisture determination in butter [J]. *Food Research International*, 1992, 25(3): 193.
- [4] Capron X, Smeyers-Verbeke J, Massart D L. Multivariate determination of the geographical origin of wines form four different countries[J]. *Food Chemistry*, 2007, 101(4): 1585—1597.
- [5] Kemsley E K, Ruault S, Wilson R H. Discrimination between coffea arabica and coffea canephora variant robusta beans using infrared spectroscopy[J]. *Food Chemistry*, 1995, 54: 321—326.
- [6] 孙素琴, 张宣, 秦竹, 等. FTIR 直接鉴别植物生药材[J]. 光谱学与光谱分析, 1999, 19(4): 542.
SUN Su-qin, ZHUANG Xuan, QIN Zhu, et al. Direct Determination of Raw plant drugs by FTIR[J]. *Spectroscopy and Spectral Analysis*, 1999, 19(4): 542(in Chinese).
- [7] 孙素琴, 周群, 郁鉴源. 傅里叶变换拉曼光谱法无损鉴别植物生药材[J]. 分析化学, 2000, 28(2): 211.
SUN Su-qin, ZHOU Qun, YU Jian-yuan. Undamaged Determination of Raw plant Medicinal Drugs with Fourier transform Raman Spectroscopy[J]. *Anal. Chem.*, 2000, 28(2): 211. (in Chinese)
- [8] 孙素琴, 梁曦云, 杨显荣. 36 种灵芝产品傅里叶变换红外光谱快速鉴别研究. 分析化学, 2001, 29(5): 552—554.
SUN Su-qin, LIANG Xi-yun, YANG Xian-rong. A Rapid method for distingwishing the different ganoderma Lucidum products by fourier transform infraned spectroscopy[J]. *Anal. Chem.*, 2001, 29(5): 552—554. (in Chinese)
- [9] 王登良, 张灵枝, 毛明辉, 等. 不同光波晒青对单枞茶品质的影响[J]. 食品与生物技术学报, 2006, 25(2): 57.
WANG Deng-liang, ZHANG Ling-zhi, MAO Ming-hui, et al. Effect of withered with different light-wave bands on gualities of damcong tea[J]. *Journal of Food Science and Biotechnology*, 2006, 25(2): 57. (in Chinese)
- [10] Budinová G, Vlácil D, Mestek O, et al. Application of infrared spectroscopy to the assessment of authenticity of tea[J]. *Talanta*, 1998, 47: 256.
- [11] 林林. 实用傅里叶变换红外光谱学[M]. 北京: 中国环境科学出版社, 1991: 507.
- [12] 何晓群. 多元统计分析[M]. 北京: 中国人民大学出版社, 2004: 136 ~ 137.
- [13] 袁洪福, 陆婉珍. 现代光谱分析中常用的化学计量学方法[J]. 现代科学仪器, 1998, 5: 9.
YUAN Hong-fu, LU Wan-zhen. Modern Scientific Instruments 1998, 5: 9. (in Chinese)
- [14] Svante Wold, Henrik Antti, Fredrik Lindgren et al. Modelling and diagnostics of batch processes and analogous kinetic experiments[J]. *Chemometrics and Intelligent Laboratory Systems*. 1998, (44): 175.

(责任编辑: 杨 萌)