

FTIR-ATR 指纹图谱的主成分分析-马氏距离法 应用于烟用香精质量控制

王家俊¹, 邱启杨¹, 刘 巍²

1. 红河卷烟总厂产品部, 云南 弥勒 652300

2. 云南烟草科学研究院烟草化学研究室, 云南 昆明 650021

摘 要 在配制过程正常的条件下, 分批采集 150 个三种正常烟用香精的傅里叶变换衰减全反射红外指纹图谱, 并对原始数据进行均值中心化处理, 同时对其光谱信号进行二阶导数基线校正和 Karl Norris 降噪处理, 应用化学计量学中的主成分分析-马氏距离法分类建模, 然后对烟用香精配制质量进行多元统计过程控制(MSPC), 结果表明, 将马氏距离上限的控制范围(UCL)设定在该类平均马氏距离的 $+3\sigma$ 范围之内, 均获得了正确的监控预报。

关键词 傅里叶变换衰减全反射红外指纹图谱; 主成分分析-马氏距离分类法; 多元统计过程控制; 烟用香精

中图分类号: O657.3 文献标识码: A 文章编号: 1000-0593(2007)05-0895-04

引 言

烟用香精是一种复杂的多组分化学体系, 一般认为, 烟用香精改善烟草制品的香吃味质量, 是烟草和香精各组分协同作用的整体效应, 所以, 整体性控制烟用香精的质量有着重要的意义。通常是依照 YC/T 145-1998 行业系列标准方法测定酸值、相对密度、折光指数和溶混度等理化指标来监控质量, 一是检测步骤烦琐, 涉及较多仪器, 二是采用有限的理化指标表征品质, 信息单一, 存在一定的局限性, 满足不了现场快速质控的要求。近些年来, 傅里叶变换衰减全反射红外光谱(FTIR-ATR: Fourier transform infrared-attenuated total reflection)分析技术获得了广泛的应用^[1,2], 但采用其指纹图谱结合模式识别技术用于香精质量控制还少见报道。由于香精的红外光谱是组成它们的所有组分的红外光谱的叠加, 不同特性的香精, 其组分或是配比上的差异信息, 最终在指纹图谱中都得到整体表征, 而对指纹图谱的解析, 必须应用化学计量学中的多变量分析技术, 其中主成分分析-马氏(Mahalanobis)距离法是一种比较实用的模式分类方法, 采用该法对烟用香精的 FTIR-ATR 指纹图谱进行分类建模, 并应用类模型对烟用香精配制的过程质量进行多元统计过程控制^[3,4](MSPC: multivariate statistical process control), 实

现的不是观测烟用香精各组分的特性, 而是对香精所含不同组分, 不同量比化合物整体所体现的宏观特性的监控。本实验在配制过程稳定的情况下, 分批次采集 150 个三种正常烟用香精的 FTIR-ATR 指纹图谱, 并应用该法建立相应的类模型。然后, 对不同批次的烟用香精进行模式识别, 将马氏距离的控制上限设定在该类平均马氏距离 $+3\sigma$ 范围之内, 获得了正确的监控预报。又因 ATR 采样技术操作简单、快速无损, 故该法应用于烟用香精配制过程质量监控具有较好的实用性。

1 基本原理

对任意一个样本的原始光谱量测数据 x_i 可用矢量表示为: $x_i = [x_{i1}, x_{i2}, \dots, x_{in}]$ 。那么, 对不同种类的 m 个样本构成的样本集, 就可表达为 $m \times n$ 阶的光谱数据矩阵 $x_{m \times n}$, 在矩阵 $x_{m \times n}$ 中, 既包含样本性质的信息, 也包含样本特征或变量的信息。从数学角度来看, 样本集可由特征坐标轴构成的 n 维特征空间中的点集来表示。一个样本对应一个点, n 个特征包含了样本性质的整体信息, 在多变量分析中, 采用主成分分析-马氏距离分类法^[5,6]建立不同种类样本的类模型, 可表征在样本集中样本与样本之间、样本与类之间的隐含关系。对矩阵 $x_{m \times n}$ 进行主成分分析得到在主成分 f 下的得

收稿日期: 2006-02-16, 修订日期: 2006-06-28

基金项目: 红河卷烟厂科研经费资助

作者简介: 王家俊, 1962 年生, 云南红河卷烟厂产品中心工程师

e-mail: honghe@vip.163.com

分矩阵 $T_{m \times f}$ 和 $P_{f \times n}$ 载荷矩阵, 其中, 得分矩阵反映了样本与样本之间的关系。通过样本集的得分矩阵 T , 可计算样本集的马氏距离, 即

$$MD_i = [(t_i - \bar{t}) M^{-1} (t_i - \bar{t})]^{1/2}$$

式中: $\bar{t} = \frac{1}{m} \sum_i t_i$ 为平均得分矩阵, M^{-1} 为马氏矩阵 M 的逆矩阵, $M = \frac{T^T T}{m - 1}$, t_i 为样本集中 i 样本的得分矢量, MD_i 为 i 样本的马氏距离。在构建类模型时, 样本集中点的分布是“物以类聚”的, 具有相似性质或是相似特征的样本在模式空间中处于相近的位置, 聚集为一类, 特征的集合就构成了模式。将类模型应用于模式分类, 便可通过马氏距离来反映样本点与该类的聚集程度, 马氏距离越小, 样本性质或特征的相似程度就越高; 马氏距离越大, 相似程度就越低, 或是归属其它类别。显然, 通过确定适当的马氏距离阈值, 就可应用马氏距离对配制过程中香精的整体宏观特性进行监控。

2 实验部分

2.1 主要仪器及工作参数设置

Nicolet NEXUS 670 FTIR 光谱仪, DTGS-KBr 检测器; ATR 采样附件, ZnSe 晶体, 入射角 45° ; TQ Analyst V6.2 数据处理软件(美国 Thermo Nicolet 公司)。仪器的主要工作参数为光谱扫描范围: $4\,000 \sim 650\text{ cm}^{-1}$; 分辨率: 4 cm^{-1} ; 扫描次数: 64。

2.2 采集光谱数据

在常温下, 开机预热光谱仪 2 h, 分别取在配制过程稳定条件下的三种不同批次的正常烟用香精 A 类和 B 类各 60

个, C 类 30 个。应用 ATR 采样附件采集烟用香精样本的 FTIR-ATR 光谱数据。

2.3 建立类模型及应用

应用 TQ Analyst V6.2 数据处理软件, 对数据进行均值中心化(Mean centering)预处理, 同时进行二阶求导校正基线和 Karl Norris 滤波降噪^[7]等处理, 应用主成分分析-马氏距离法建立正常烟用香精 A, B, C 的类模型。然后, 应用所建的类模型对烟用香精配制质量进行多元统计过程控制(MSPC), 并将监控结果进行数据可视化表征。

3 结果与讨论

3.1 光谱重构与类模型建立

所获取的红外指纹图谱是香精所有组分的叠加谱, 见图 1(a), 信号质量时常会受样品颜色和仪器硬件性能等因素的影响, 存在明显的基线飘移和噪音, 本实验采用了二阶求导结合 Karl Norris 滤波器重构光谱, 消除基线飘移干扰, 提高信噪比。在实际操作中, Karl Norris 滤波器窗口宽度的选择具有一定的经验性, 一般来说, 较大的段长(Segment length)会造成信号失真, 较小的段间距(Gap between segments), 降噪效果不佳。在本实验中, 就所使用的仪器和量测对象, 选择段长 19, 段间距 2, 获得了较好的分类效果。同时通过二阶导数重构指纹图谱, 图谱的指纹表征也更为明显, 见图 1(b), 可以看出, 香精 C 与香精 A 和 B 的指纹特征差异比香精 A 和 B 的差异明显, 所以, C 与 A 和 B 的分类效果比 A 与 B 的分类效果理想, 见图 2(a)~图 2(c)。在类模型中类内样本点的平均马氏距离与类之间的马氏距离见表 1。

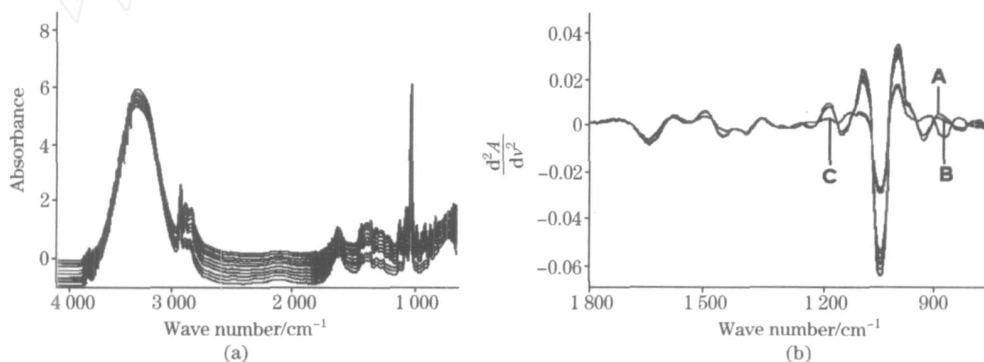


Fig 1 FTIR-ATR spectra of tobacco flavour (a) and their 2nd derivative spectra(b)

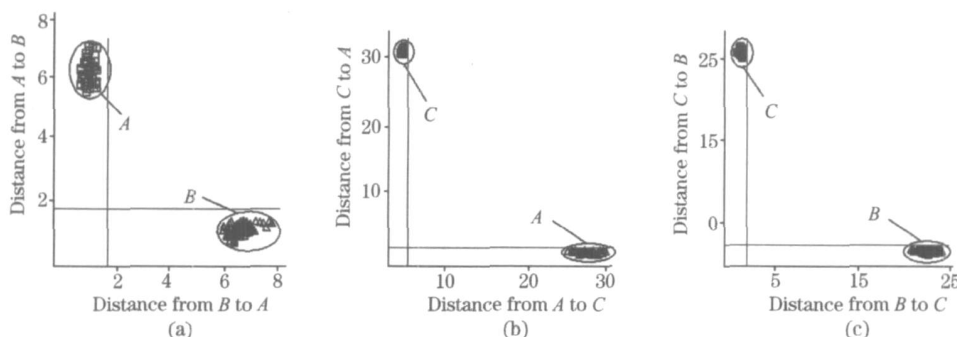


Fig 2 Class plot of A, B and C from classification of principal component analysis-Mahalanobis distance

Table 1 Mahalanobis distances of A, B and C and their interclass distances ($m_A = 60$, $m_B = 60$, $m_C = 30$)

类	类内马氏距离统计					类之间的马氏距离		
	平均值	最小值	最大值	标准差	数据分布	到 A 的距离	到 B 的距离	到 C 的距离
A	0.970 9	0.649 4	1.278 6	0.176 9	Normal distribution	0	6.215 9	29.451 4
B	0.979 1	0.538 9	1.272 1	0.170 4		6.544	0	22.655 4
C	0.958 3	0.548 3	1.190 6	0.152 0		30.624 3	25.518 1	0

3.2 MSPC 图的建立与应用

MSPC 技术应用于过程质量监控,主要是建立多元统计模型,将大量过程变量映射到由少量具有代表性的特征隐变量定义的低维空间,避免变量间多重相关性造成的信息重叠,从而使过程监控简化,并采用 Hotelling T^2 , Q 等统计量对服从正态分布的特征信号建立 MSPC 控制图,对过程进行监控,见文献[3, 4]所作的综述。然而,本实验直接应用以上所建的多元统计模型——类模型监测过程样本所得到的马氏距离建立 MSPC 控制图,通过实际应用对比,将马氏距离的控制上限设定为: $UCL = \mu + 3\sigma$,对香精配制进行监控,所得到的香精质量特性均能满足生产要求。同时也减少了 MSPC 控制图的使用数量。以监控 B 类烟用香精为例说明,图 3 为监控 32 个不同批次配制的 B 类烟用香精的 MSPC 控制图,其中配制过程出现异常(组分比例错配)的 B022 和 B029 两个样本(马氏距离分别为 1.74 和 1.62)均获得了正确预报。

3.3 小结

通过 ATR 采样技术获取红外指纹图谱,操作简捷,样

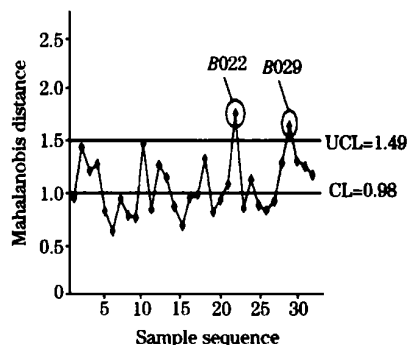


Fig 3 3 σ -control plot of Mahalanobis distance for B class

品前处理无损,采用主成分分析-马氏距离法对红外指纹图谱分类建模,并应用于烟用香精配制质量的多元统计过程控制(MSPC),方法简单,可实现对烟用香精整体宏观特性的监控。本文用 FTIR-ATR 方法测定烟用香精,得到一些有实用意义的成果。国内还有把这种方法(FTIR-ATR)应用于其他方面的工作^[6]。

参 考 文 献

- [1] Workman Jr Jerome J. Applied Spectroscopy Reviews, 1999, 34(1 & 2): 1.
- [2] XU Lin, WANG Nai-yan, BA Shu-hong, et al(徐琳,王乃岩,霸书红,等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2004, 24(3): 317.
- [3] Wise B M, Gallagher N B. Journal of Process Control, 1996, 6(6): 329.
- [4] MacGregor J F, Kourti T. Control Engineering Practice, 1995, 3(3): 403.
- [5] XU Lu(许禄). Chemometrics: Principles and Applications of Important Methods(化学计量学:一些重要方法的原理及应用). Beijing: Science Press(北京:科学出版社), 2004. 24.
- [6] ZHANG Yao-ting, FANG Kai-tai(张尧庭,方开泰). Introduction to Multivariate Statistical Analysis(多元统计分析引论). Beijing: Science Press(北京:科学出版社), 2003. 123.
- [7] Norris K H, Williams P C. Cereal Chem., 1984, 61: 158.
- [8] XU Yi-zhuang, ZHAO Ying, XU Zhi, et al(徐怡庄,赵莹,徐智,等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2005, 25(11): 1775.

Quality Assessment of Tobacco Flavor by Classification of Principal Component Analysis-Mahalanobis Distance Combined with FTIR-ATR Fingerprint

WANG Jia-jun¹, QIU Qi-yang¹, LIU Wei²

1. Products Research Centre, Honghe Cigarette Factory, Mile 652300, China

2. Tobacco Chemistry Laboratory, Yunnan Tobacco Research Institute, Kunming 650021, China

Abstract In normal condition, 150 Fourier transform infrared-attenuated total reflection spectra of three kinds of tobacco flavor were measured. Mean centering and Karl Norris second derivative filter were employed for treating the spectra data. Whereafter, class models of tobacco flavor were established with classification of principal component analysis-Mahalanobis distance. Class models have been applied to multivariate statistical process control (MSPC) for the quality assessment of tobacco flavor in site. The upper control limit (UCL) was based on this kind of average Mahalanobis distance within $+3\sigma$. The results of monitoring are correct.

Keywords FTIR-ATR fingerprint; Classification of principal component analysis-Mahalanobis distance; Multivariate statistical process control; Tobacco flavor

(Received Feb. 16, 2006; accepted Jun. 28, 2006)

《光谱学与光谱分析》对来稿英文摘要的要求

来稿英文摘要不符合下列要求者, 本刊要求作者重写, 这可能要推迟论文发表的时间。

1. 请用符合语法的英文, 要求言简意明、确切地论述文章的主要内容, 突出创新之处。
2. 应拥有与论文同等量的主要信息, 包括四个要素, 即研究目的、方法、结果、结论。其中后两个要素最重要。有时一个句子即可包含前两个要素, 例如“用某种改进的 ICP-AES 测量了鱼池水样的痕量铅”。但有些情况下, 英文摘要可包括研究工作的主要对象和范围, 以及具有情报价值的其他重要信息。在结果部分最好有定量数据, 如检测限、相对标准偏差等; 结论部分最好指出方法或结果的优点和意义。
3. 句型力求简单, 尽量采用被动式, 通常应有 20 个左右意义完整、语句顺畅的句子。英语词数以 300 至 400 个为宜, 不能太短; 也不要太长。用计算机单面隔行打印。
4. 摘要不应有引言中出现的内容, 换言之, 摘要中必须写进的内容应尽量避免在引言中出现。摘要也不要对论文内容作解释和评论, 不得简单重复题名中已有的信息; 不用非公知公用的符号和术语; 不用引文, 除非该论文证实或否定了他人已发表的论文。缩略语、略称、代号, 除相邻专业的读者也能清楚地理理解外, 在首次出现时必须加以说明, 例如用括号写出全称。