

doi: 10.11857/j.issn.1674-5124.2018.06.009



GC-MS 指纹图谱结合多元统计方法用于香精质量差异分析

刘秀明¹, 王艳芳², 李超¹, 马明³, 王家俊¹, 段焰青¹

(1. 云南中烟工业有限责任公司技术中心, 昆明, 650231; 2. 科迈恩(北京)科技有限公司, 北京, 100080;
3. 中国烟草总公司郑州烟草研究院, 郑州, 450001)

摘要: 结合多元统计分析比较香精香料存在的质量差异从而对其质量进行控制。提取 A、B、C 3 类香精共 198 批样品的 GC-MS 指纹图谱, 通过主成分分析(PCA)建立不同类样品共有模式, 找出影响该类香精质量稳定性的化学成分; 再对 3 类香精进行 PCA 分析找出造成不同组香精有差异的化学成分; 结合偏最小二乘-判别分析(PLS-DA)方法对香精样品进行分类($P>0.05$), 将 25 批香精样品作为测试集进行基于 PLS-DA 算法的模式识别, 3 类香精识别率达到 100%。该分析方法对香精质量控制提供依据, 针对某一质量波动的成品香精, 可通过对差别成分的调整进行质量控制。

关键词: 香精; GC-MS; 指纹图谱; 多元统计分析; 差异分析

文献标志码: A **文章编号:** 1674-5124(2018)06-0042-06

Quality variance analysis of flavors by GC-MS fingerprint and multivariate statistical

LIU Xiuming¹, WANG Yanfang², LI Chao¹, MA Ming³, WANG Jiajun¹, DUAN Yanqing¹

(1. Technology Center, China Tobacco Yunnan Industrial Co., Ltd., Kunming 650231, China;
2. Chemmind Co., Ltd., Beijing 100080, China; 3. Zhengzhou Tobacco Research
Institute of CNTC, Zhengzhou 450001, China)

Abstract: Multivariate statistical analysis was proposed to compare the quality differences of flavor for controlling of quality. The fingerprint of 198 flavor samples was extracted based on Gas Chromatograph- Mass Spectrometer technology. Type A, B and C flavor of model were built through principle component analysis(PCA), respectively, the chemical compositions affecting the quality stability of the flavor were found; The chemical constituents of different groups of fragrances were identified by PCA analysis. The samples were classified clearly combined with partial least squares discriminant analysis (PLS-DA), then 25 flavor samples were used as validation samples for pattern recognition($P>0.05$), the recognition rate was 100%. This method provide a basis for quality control of flavors, and the quality can be controlled by adjusting the difference components for quality fluctuation flavors.

收稿日期: 2018-03-06; 收到修改稿日期: 2018-04-25

基金项目: 云南中烟工业有限责任公司科技项目(2018CP04); 2015年云南省技术创新人才项目(2016HB009)

作者简介: 刘秀明(1985-), 女, 云南宣威市人, 工程师, 硕士, 主要从事烟用香精香料研究工作。

通信作者: 段焰青(1973-), 男, 云南大理市人, 研究员, 博士, 主要从事烟用香精香料及微生物产香方面研究。

Keywords: flavors; GC-MS; fingerprint; multivariate statistical analysis; variance analysis

0 引言

烟用香精香料是卷烟生产环节所不可缺少的辅料, 配方也是烟草工业企业的核心技术之一, 与卷烟品牌的树立与发展密切相关^[1-2], 因此, 正确识别香精类别和评估其质量稳定性具有重要的意义。香精香料的传统质量检测主要通过折光指数、相对密度、酸度值、挥发分总量等一些物性指标和人工嗅香两个环节来完成^[3]。前者只能从总体反映香精香料的某些特征, 而后者由于受主观意识和环境个体差异影响, 波动较大。传统检测方式可以区分香料大类, 但对香料的较微小差异检测效果不明显, 且要求经验较高, 因此在识别准确性与效率上表现并不尽如人意^[4-5]。本次试验拟从化学组成的角度对香精香料进行定性定量解析, 以增大检测准确率与效率。色谱指纹图谱的应用是一种有效的质量控制方法, 但在色谱指纹图谱分析中, 无论采用相关系数还是矢量夹角余弦来计算相似度, 由于算法本身的问题, 其对数据变化不敏感^[6], 并且采用单一的某个方法很难对指纹图谱提取物这一复杂体系的各成分总含量的相似度进行评价, 若采用化学计量学和多元统计方法, 可对其质量稳定性给出更客观和恰当的评价。

主成分分析 (PCA) 是通过对一组相关变量进行正交旋转等处理, 用维数较少的不相关的新变量来反映原变量携带的大部分信息, 通过分析新变量达到解决问题的一种多元统计方法, 已经广泛应用于许多领域^[7-10]。在香精香料质量控制方面, 如郭紫明等人^[9]采用气相色谱-质谱法和主成分投影分析法用于 10 个批次的烟用香精的分析和质量控制。偏最小二乘-判别分析 (PLS-DA) 是一种新颖的多元线性回归方法, 具有数据降维并提取特征信息的优良性能, 该法以类别编码作为 Y 变量, 是基于 PLS 的判别分析方法^[11-12]。如陈康宁^[13]等人应用热脱附-气质联用技术, 提取烟用香料的指纹图谱, 通过 PLS-DA 方法建立相应的类模型, 用于香料质量等级分类和质量的一致性评价具有良好的实用性。

本文采用 ChemPattern 化学计量学软件, 对香精样品的 GC-MS 指纹图谱进行预处理, 确定其共有模式, 进行 PCA 分析建立各类香精的类模型, 找

出造成各类香精质量波动的化学成分以及 3 类香精有差异的化学成分, 结合 PLS-DA 对样品进行模式识别, 实现对香精质量稳定性的监控, 同时可明确引起香精质量波动的差异性化学成分, 该研究为香精香料调配生产工艺优化提供依据。

1 材料与方法

1.1 仪器和样品

仪器: Thermo Quest TraceGC+Finnigan Trace MS (美国赛默飞世尔公司); 色谱柱: 采用 HP-5MS (美国安捷伦公司) (30 m×0.25 mm, 0.25 μm) 毛细管柱。AB204-S 分析天平 (Mettler-Toledo, 感量 0.0001 g, 美国)。乙酸苯乙酯作内标。ChemPattern 化学计量学与化学指纹图谱系统解决方案软件 2017 版 [科迈恩 (北京) 科技有限公司 (Chemmind Technologies Co., Ltd.)]。

样品: 云南中烟在用的不同批次不同风格特征的香精 (A 样品 63 批, B 样品 57 批和 C 样品 78 批, 共 198 批)。

1.2 仪器条件及色谱分析

载气: 高纯氮; 载气流量 1 mL/min, 恒流; 进样量: 1 μL; 进样口: 温度: 250 °C, 分流比: 10:1, 分流流量: 10 mL/min; 程序升温: 40 °C 保持 3 min, 以 5 °C/min 的速率升至 180 °C, 保持 8 min, 以 10 °C/min 的速率升至 250 °C, 保持 8 min。

质谱条件: 全扫描; 扫描范围: 29 到 445 Da; 溶剂延迟: 14.5 min。

称取一定量的香精样品, 直接采用 GC-MS 分析。

1.3 指纹图谱提取及多元统计分析

采用 ChemPattern 软件, 对 198 批香精样品的 GC-MS 指纹图谱数据进行预处理, 确定其共有模式, 进行 PCA 分析和模式识别研究, 找出造成各类香精质量波动的化学成分以及 3 类香精有差异的化学成分, 分析了同类香精的稳定性及不同类香精的差异性。将 198 个样品随机分为模型训练集 (175 个) 和测试集 (23 个), 建立各类香精的 PLS-DA 分类模型。

2 结果与讨论

2.1 3 类香精共有模式建立

将样品图谱导入 ChemPattern 软件, 样品叠加

图谱如图1所示,可以明显看出各样品间的色谱图差异较大,不同分组的样品间色谱峰有显著差异;同一分组的样品间也存在一定差异。

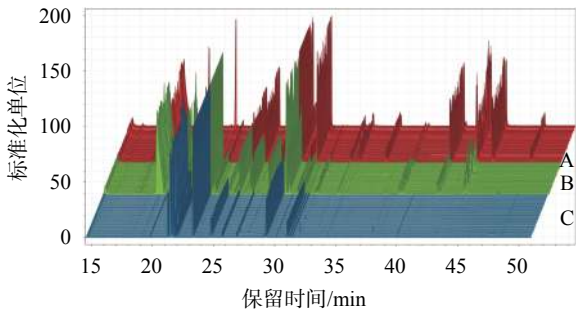


图1 198批香精样品 GC-MS 叠加图谱

色谱指纹图谱的可靠识别不仅与色谱分离条件有关,还与分析软件对色谱信息的提取有关,如何在去除干扰的同时提取足够的信息量对化学计量学分析非常重要。首先对香精样品的 GC-MS 图谱数据进行去背景、积分、剪裁、保留时间校正(色谱峰对齐),以及色谱峰数据规范化处理,将198批样品参与共有模式的生成,共有模式如图2所示,3类香精样品共筛选出21个共有色谱峰。

2.2 相似度分析香精分类情况

利用所建立的共有模式,对样品进行相似度分析。用夹角余弦计算方法得到分析结果如图3所示,结果显示3组样品之间的具有显著差异,但B组中的#73~#75样品与C组中的#145~#147样品存在重叠,总体上可以采用相似度分析方法对3者

的分类加以区分,此外,通过聚类热图也能看出3组样品区分较明显。

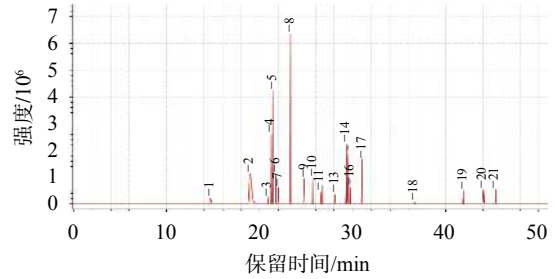


图2 3类香精样品的共有模式

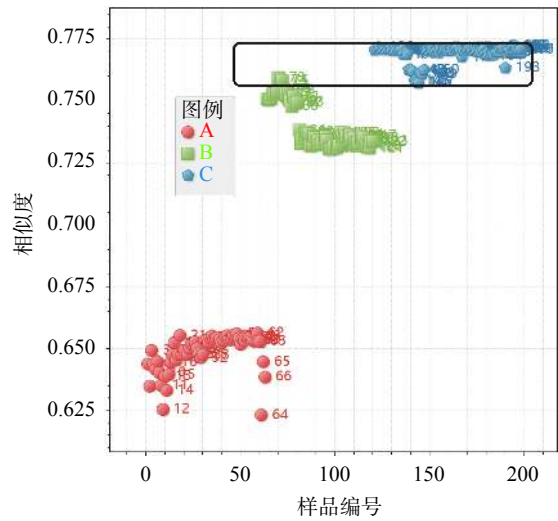


图3 3类香精样品的相似度分析

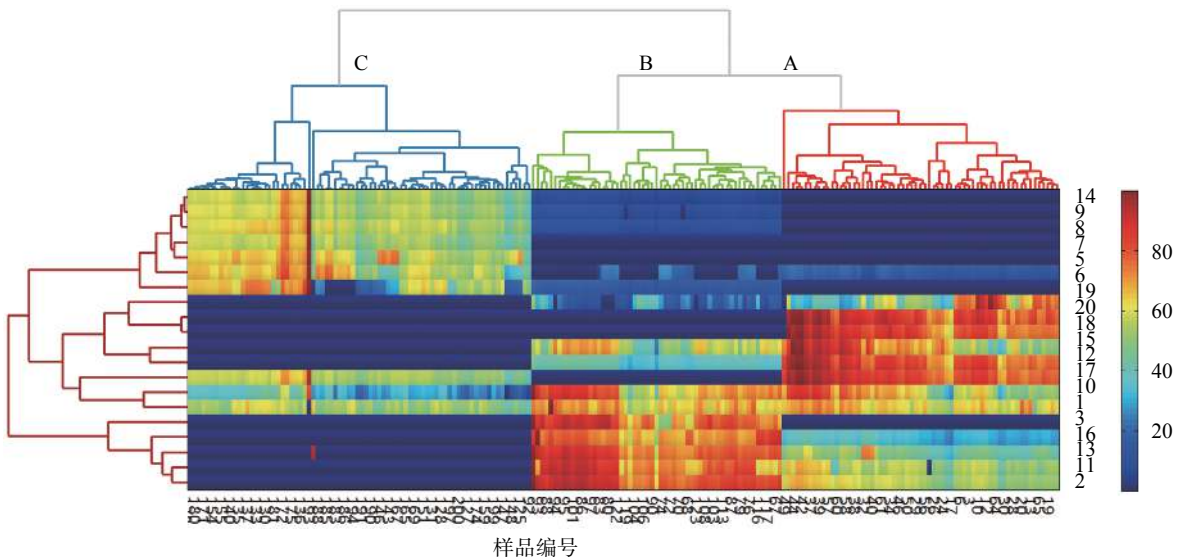


图4 3类香精样品的聚类热图

2.3 多元统计分析香精的差异性

以 A 组样品建立共有模式, 提取出 15 个共有峰, 对 A 组样品进行 PCA 分析, 分析结果如下图 5 中(A)所示, 由图中可看出该成品香精各批次样品间大体上无显著差异, 只是其中有 3 个(#7、#8 和#9)样本偏离较远, 从第一主成分的载荷可以看出, 其中#2、#8、#5 及#13 色谱峰对主成分的影响贡献较大。同样对 B 组样品进行 PCA 分析,

结果如图 5(B)所示, 由图中可看出 B 组样品大部分比较集中, 分散较均匀, 只有#9 和#92 两个样品偏离较大, 说明该组样品质量相对较稳定, 从第一主成分载荷可知, #2、#8 及#15 色谱峰的贡献最大。对 C 组样品进行 PCA 分析的结果如图 5(C)所示, 可看出, C 组可分为 2 个亚组, #158 样品偏离较大, #2、#8 和#5 色谱峰对主成分的影响贡献较大。

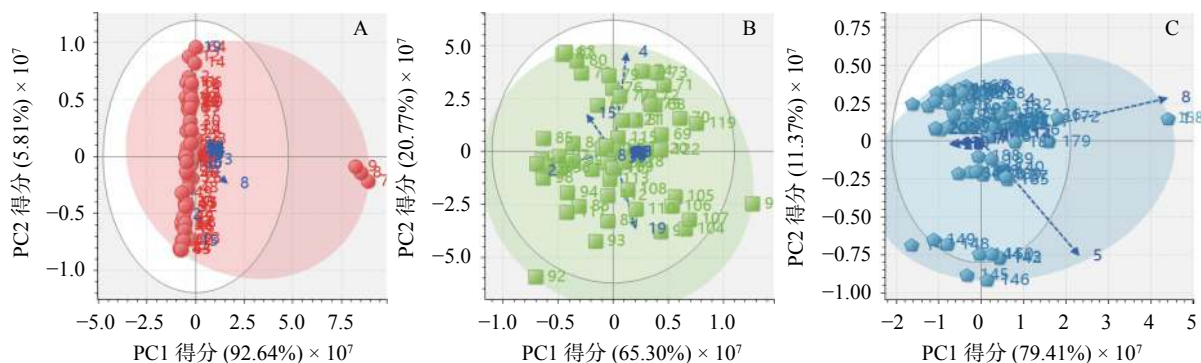


图 5 3 类样品的主成分分析图

对 A、B、C 3 组所有样品进行 PCA 分析, 结果如图 6(左)所示, 可大体上对 3 组样品进行较好地区分。不过与单组分析结果类似, A 组中存在偏离较大的 3 个样品, C 组中 158# 样品也偏离很大。由第一主成分载荷图可以看出 2#、5#、8# 及 17# 色谱峰影响较大, 后期可进一步对色谱峰进行定性鉴别, 以便对特定色谱峰进行归属。对所有样品进行单向多元方差分析 (MANOVA) 如图 6(右)所示,

可以看出可以很好地对各组进行区分, 尽管 PCA 分析显示存在几个偏离较大的样品。另外, 从第一典型相关变量的载荷可以看出#10、#17、#18 及#14 色谱峰影响较大。

2.4 化学模式识别分析香精差异

将 198 个样品随机分为训练集 (175 个) 和测试集 (23 个), 采用 PLS-DA 算法进行模式识别 ($P > 0.05$), 结果如图 7 所示。采用 3 倍交叉验证进行潜

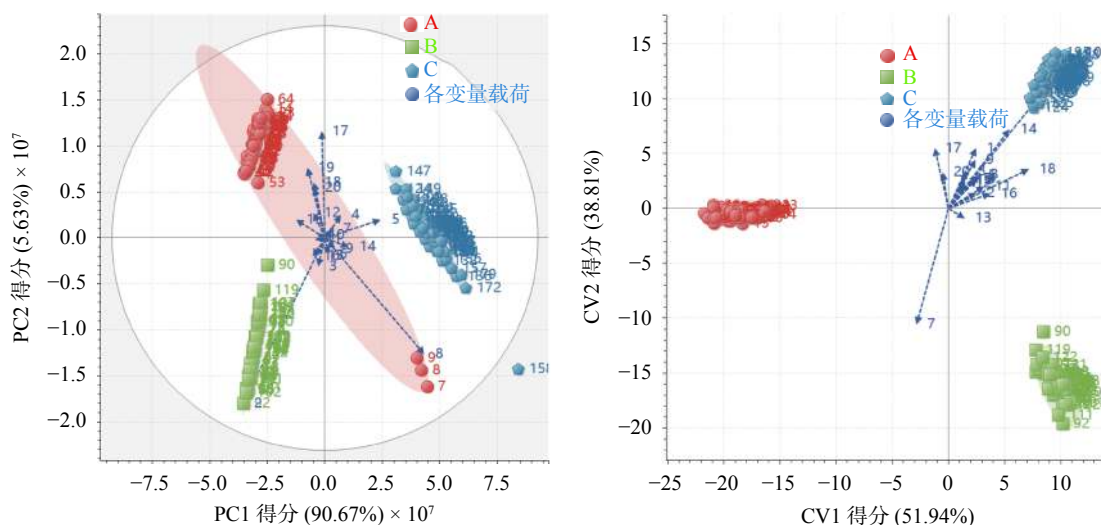


图 6 3 类香精的 PCA 分析图 (左) 及 CCA 分析图 (右)

变量数目 (nLV) 优化, 当潜变量数目为 1 时, 模型的预测率 (Q^2Y) 达到了 100%。模型潜变量数目为 1, 表明所构建的模型简单, 过拟合风险很小。采用测试集对所建模型的预测效果进行评价, 结果如表 1 所示。从表中可以看出, A、B、C 3 类样品拒识

率 FRR(%) 及误识率 FAR(%) 为 0, 即 3 个分组的香精识别准确率达到 100%, 表明所构建的 PLS-DA 能准确判定出不同香精的种类。另外, 从 PLS-DA 类预测概率分布图 (见图 8) 也可以看出所建香精风味模式识别模型是可靠的。

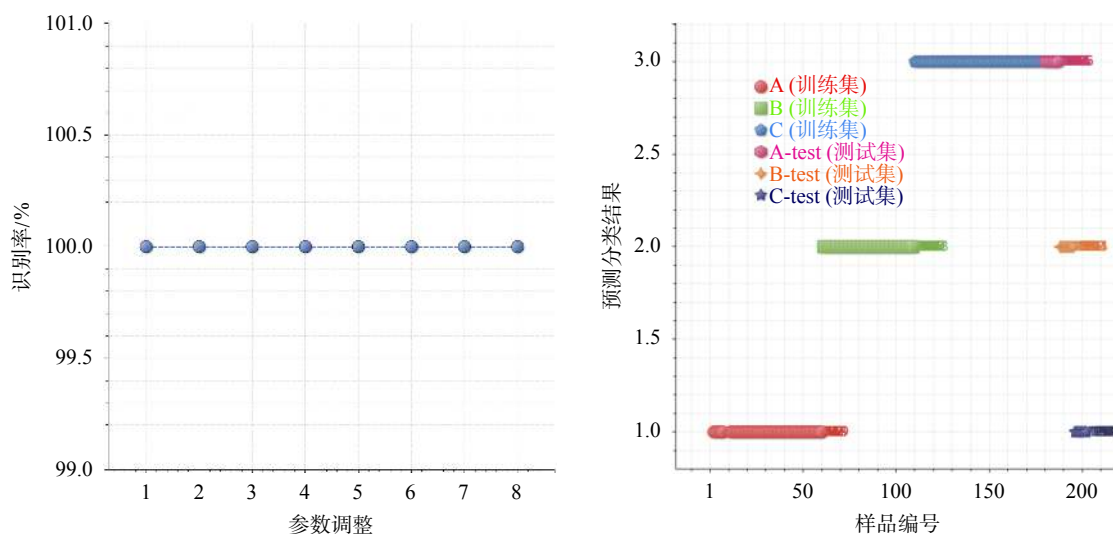


图 7 PLS-DA 模式识别 (左) 及模型的预测率 (右)

表 1 模式识别结果统计

编号	分类名称	归属	总样本	训练集	验证集	测试集	拒识率 FRR/%	误识率 FAR/%
1	A	训练集	56	37	19	0	0	0
2	B	训练集	49	33	16	0	0	0
3	C	训练集	70	47	23	0	0	0
4	Test	测试集	23	0	0	23	0	0
合计			198	117	58	23	0	0

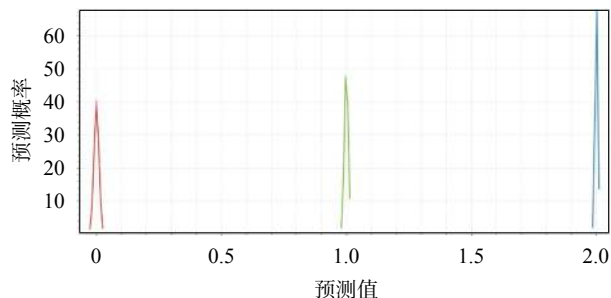


图 8 PLS-DA 分类预测概率分布

3 结束语

香精香料质量控制作为卷烟企业制造链的核心技术之一, 有较强的自主性和技术保密性^[14]。传统

的烟用香精检测方法如密度、折光、香气、色泽、澄清度、溶混度和酸值检测等不能从内在成分上反映香精质量的缺点, 所取得的数据不能反映其全貌, 不能真正体现香精香料的质量稳定性问题。目前国内关于的指纹图谱结合化学计量学分析较多, 这些方法取得数据信息涵盖较为全面, 用于质量控制结果可信度高, 能较好地反映香精香料质量波动情况, 可以准确地控制产品质量的一致性和稳定性^[15-16]。采用 GC-MS 技术对香精样品直接进样检测, 提取其指纹图谱分别建立不同类样品共有模式并进行 PCA 分析, 找出影响该类香精质量稳定性的化学成分; 再对 3 类香精进行 PCA 分析, 找出造成 3 类香精有差异的化学成分; 结合 PLS-DA 分析, 将 23 批

样品作为训练集进行模式识别研究,3个分组的香精识别率达到100%。通过主成分分析、模式识别等发现3类香精样品差异性较大,其中B、C两组样品的组内差异性相对不大,在A类香精中#2、#8、#5和#13 4个峰影响较大,在B类香精中#2、#8和#15峰影响较大,在C类香精中#2、#8和#5峰影响较大,在以后的工作中将对分析找出的差异成分进行定性鉴别和归属,以方便找出可用于区分3类香精的指标性化合物。该分析结果一方面为香精质量控制提供依据,针对某一质量波动的成品香精,可通过对差别成分的调整进行质量控制,另外,该方法实现了对香精质量稳定性进行监控,为香精质量控制提供简单、科学的新方法,具有一定的实用价值。

参考文献

- [1] 斯文,王雨凝,许高燕,等. 烟用香精质量稳定性的评价及统计过程控制[J]. 香料香精化妆品, 2015(1): 28-32.
- [2] 曲国福,孟昭宇,陆舍铭,等. 香料指纹图谱及其在烟用香精质量控制中的应用概述[J]. 烟草科技, 2006(5): 47-50.
- [3] 孙川,桂永发,许永,等. 水蜜桃香精挥发成分的气相色谱/质谱分析[J]. 应用化工, 2007, 36(11): 1138-1140.
- [4] 曲国福,陆舍铭,孟昭宇,等. 色谱指纹图谱在烟用香精质量控制中的应用[J]. 烟草科技, 2007(7): 46-50.
- [5] 卢红兵,孔波,魏维伟,等. GC/MS结合化学计量学用于复杂香精配方中香料成分的定性识别[J]. 烟草科技, 2013(3): 50-53.
- [6] 包锦渊,吴启勋,青海秦. 青海秦艽色谱指纹图谱的灰色关联分析[J]. 上海大学学报(自然科学版), 2006, 12(6): 620-623.
- [7] SONG S Q, ZHANG X M, HAYAT K, et al. Coordinating fingerprint determination of solid-phase microextraction/gas chromatography-mass spectrometry and chemometric methods for quality control of oxidized tallow[J]. Journal of Chromatography A, 2013(1278): 145-152.
- [8] 钟科军,魏万之,郭方遒,等. GC-MS及主成分分析法用于咖啡香精的指纹图谱分析和微差样品的识别[J]. 分析试验室, 2006, 25(8): 119-122.
- [9] 郭紫明,董道竹,李燕春,等. 气相色谱-质谱法和主成分投影分析法用于烟用香精的分析和质量控制[J]. 理化检验-化学分册, 2008, 44: 22-27.
- [10] 黄世杰,陈志燕,蒋宏霖,等. 主成分分析法在烟用香精质量控制中的应用[J]. 烟草科技, 2009(8): 31-34.
- [11] BRERETON R G. Chemometrics for pattern recognition[M]. A John Wiley & Son, Ltd., UK, 2009.
- [12] ERIKSSON L, JOHANSSON E, KETTANEH WOLD N, et al. Multi- and megavariable data analysis (Part I, II)[M]. Sweden: Umetrics AB, 2006. 2009, (8): 31-34.
- [13] 陈康宁,杨燕,李伟,等. 香料的热脱附-气质联用指纹图谱偏最小二乘判别分析与差异表征[J]. 云南大学学报(自然科学版), 2015(1): 117-123.
- [14] 钟坚成,冯毅,雷君虎,等. 烟用香精香料质量控制系统的模型及架构[J]. 计算技术与自动化, 2009, 28(2): 23-27.
- [15] 戴迎雪,朱立军,黎洪利,等. 烟用香精香料质量控制标准体系研究[J]. 安徽农业科学, 2010, 38(31): 17862-17864, 17867.
- [16] 曲国福,孟昭宇,陆舍铭,等. 香料指纹图谱及其在烟用香精质量控制中的应用概述[J]. 烟草科技, 2006, 22(5): 47-50.

(编辑:徐柳)