

微管等离子体电离质谱法结合随机森林模型快速鉴别 纺织品纤维成分

尚宇瀚 吕悦广 孟宪双 吕庆 郭项雨 张庆*

(中国质量检验检测科学研究院, 国家市场监督管理总局消费品质量安全监测与风险评估重点实验室, 北京 100176)

摘 要 为应对传统方法鉴别进口穿着类纺织品纤维成分中存在耗时长、主观性强及环境污染等难题,建立了基于电烙铁灼烧-筛网碰撞-微管等离子体电离质谱法(ESIB-MC- μ TP-MS)和随机森林(RF)算法的快速检测技术。本方法无需样品前处理、操作简便,单次数据采集用时不超过 5 s,适用于纺织品质谱特征数据的快速采集。在优化的条件下,采集了 27 个鉴定纺织品样品的质谱数据,构建包含 181 组数据、4501 个特征维度的数据集,用于训练 RF 预测模型。结果表明,此模型的袋外评分为 0.9762, 5 折交叉验证准确率为 0.9683(± 0.0299),对棉、羊绒/羊毛、聚酯纤维和腈纶的 F1 分数均高于 0.889,分类性能良好。将此模型用于分析 30 个海关抽检穿着类奢侈品样品,发现其中 28 个样品的鉴别成分与标识成分一致; 2 个样品鉴别结果与标识成分一致但置信度较低,可能为虚标成分样品或模型过拟合导致的假阳性。本方法在纺织品质安全监管中具有良好的应用潜力,为纺织品种类快速鉴别提供了一种可行的思路。

关键词 纺织品; 微管等离子体电离质谱法; 随机森林; 纤维成分鉴别

自 2023 年以来,受国际贸易环境变化、本土供应链升级及消费需求结构性调整等多重因素影响,我国穿着类纺织品进口总值持续下滑。海关总署数据显示,2025 年 1~2 月我国服装进口量与去年同期相比下降 27.1%^[1],凸显海外供给端整体收缩态势。面对这一趋势,品牌商可能通过提价维持利润与品牌定位,而价格敏感型消费者可能转向二手交易及非正规代购渠道,从而显著增加遭遇假冒伪劣商品的风险。因此,开发高效精准的纺织品纤维鉴别技术,对强化进口纺织品质量监管与保护消费者权益具有重要的现实意义。

目前,国内相关行业通常使用 GB/T 2910 系列国家标准^[2]以及 FZ/T 01057 系列^[3]和 SN/T 3236—2012^[4]等行业标准方法鉴别纺织品纤维成分。其中,FZ/T 01057 第 2、4、5 部分规定了燃烧法、溶解法和呈色反应法等化学定性分析方法。这些方法操作简单、分析时间较短,但无法准确定量,结果判定依赖检测人员的经验,主观性较强。FZ/T 01057 第 3、6、7、9 部分规定了显微镜法、熔点法、密度梯度法和双折射率法等物理方法,这些方法满足定性分析与定量分析需求,但样品的前处理过程复杂,不适用于现场检测。GB/T 2910 系列标准采用 ISO 1833-1:2006 标准,规定了多种二组分纤维混合物的定量化学分析方法,但样品前处理过程更加耗时,并且需要消耗大量辅助试剂,难以满足绿色分析化学的基本要求^[5]。除了上述方法之外,还有一系列基于现代仪器分析技术的纺织品鉴别方法被开发并投入应用。例如,基于红外和拉曼光谱法的纺织纤维鉴别试验方法在 FZ/T 01057.8 和 SN/T 3236—2012 中被规定为标准方法,而相关方法也被用于分析纤维类痕迹证据成分^[6-7]和鉴别不同来源纺织品种类^[8-9]等。这类方法适用范围广、操作便捷,但容易受到水、二氧化碳和染料分子等干扰,并且提高空间分辨率将显著增加损坏样品的风险。热分析方法被广泛应用于研究纤维热解过程^[10-12],通过监测和比较不同纤维的热变化特征可以实现对纤维种类的鉴别。这类方法具有样品用量小、操作简便和准确度高等优点,但通常依赖于专用精密仪器与稳定的实验室环境,且单次分析用时较长。质谱法以其优良的定性和定量分析能力,在纺

2025-04-22 收稿; 2025-06-20 接受

国家重点研发计划项目(No. 2022YFF0607204)和国家市场监督管理总局工业及消费品质量安全保障项目资助。

* E-mail: njuzhangqing@caiq.org.cn

织品检测与鉴别领域广泛应用^[13]。然而,传统色谱-质谱联用法通常需要复杂的样品前处理过程,耗时较长、环境污染较大。相比之下,原位电离质谱法无需或仅需简单样品前处理,具有快速高效和环境友好等优势,更适合实时原位检测场景。目前,这类技术已实际应用于纺织纤维^[14]、面料和染料^[15]的法医学分析和纤维表面吸附气味分子鉴定^[16]等领域。值得注意的是,目前常用的原位电离质谱法通常需要配备商品化电离装置,成本较高,在一定程度上限制了其推广应用。最近, Lv 等^[17]报道了一种基于微管等离子体的原位电离装置,该装置可使用市售配件自行组装,降低了配置成本;同时,通过引入碰撞筛网提升等离子体电离效率,使其可与不同质量分辨率的质谱仪配合使用,提高了适应性,在纺织品快速鉴别中展现出巨大的应用潜力。

近年来,结合现代仪器分析技术与机器学习对不同样品进行鉴别、溯源及分类的研究日渐增多^[18-20]。在众多算法中,随机森林(Random forest, RF)^[21]在高维数据处理、多分类适应性和小样本鲁棒性等方面优势明显,是与质谱技术联用的理想选择。作为一种集成学习算法,RF 通过构建大量决策树,利用投票机制实现预测,天然适配多类别分类问题。在数据抽样和特征选择环节中,RF 的双重随机性设计能够显著降低单棵决策树的过拟合风险,赋予模型优秀的鲁棒性与泛化能力。此外,RF 算法采用 Bagging 集成策略通过 Bootstrap 重采样生成多样性训练子集,一方面缓解因样本量不足导致的类别分布偏斜问题,有效应对样品总数偏少和类别不平衡等现实约束问题;另一方面,通过聚合多树预测结果,显著抑制质谱数据中的高频随机噪声,并减弱低频基线漂移对判别的系统性干扰。在处理高维度和高稀疏性的质谱数据时,RF 的特征随机选择机制可自动过滤冗余变量,从而缓解维度灾难和提高模型稳定性;同时,还可通过基于 Gini 不纯度下降量的特征重要性评估,识别对分类贡献显著的 m/z 峰,为化学成分解析提供可解释的依据。然而,受限于其集成架构与“黑盒”特性,RF 在模型解释性、计算效率及外推能力方面仍存在一定的局限性。这些问题可通过引入 SHAP/PDP 事后解释、结合并行训练与特征筛选降低计算复杂度和采用增量学习策略等措施进行针对性补偿。

本研究建立了基于电烙铁灼烧-筛网碰撞-微管等离子体电离质谱法(ESIB-MC- μ TP-MS)的纺织品纤维成分快速检测方法,并结合 RF 算法实现了对纺织品样品纤维种类的快速鉴别。本方法采用自制电离装置,具有较强的可定制性和仪器兼容性,实验操作简单,电离过程稳定,数据一致性较好,可以满足监管机构对纺织品进行现场快速检测的需求,为开展数据横向对比和建立跨实验室标准化大数据库等提供了一种可行的技术手段。

1 实验部分

1.1 仪器与试剂

CTP-2000K 低温等离子体实验电源(南京苏曼等离子体科技有限公司); Q Exactive 静电场轨道阱质谱仪,配有 Exactive Tune 质谱仪控制软件和 Xcalibur Qual Browser 数据处理软件(美国 Thermo Scientific 公司); VAPUR 接口(美国 Ionsense 公司); 2652C-50 干式真空泵(美国 Welch 公司); MF5712-N-200 气体质量流量计(美国 Siargo 公司); WSD 71 手持式电烙铁(德国 Weller 公司); Nicolet iS10 傅里叶变换中红外光谱仪,配有 Orbit 附件和 OMNIC 仪器控制软件(美国 Thermo Scientific 公司)。铜质筛网(200 目(孔径 0.074 mm),美国 Ionsense 公司)。其它材料和配件信息见电子版文后支持信息。

Pierce ESI 阴离子校准液(美国 Thermo Scientific 公司); 高纯氦气(99.999%,北京北氧联合气体有限公司); 甲醇(色谱纯,上海安谱实验科技股份有限公司); 无尘纸(美国 Kimberly-Clark 公司)。27 种纺织品标准样品由内蒙古自治区市场监督管理审评查验中心提供并鉴定,具体信息见电子版文后支持信息表 S1。30 种穿着类奢侈品实际样品由上海海关工业品与原材料检测技术中心提供。

1.2 实验方法

1.2.1 MC- μ TP 电离装置的构建

参考文献^[17]的方法,使用购买的相关材料自行搭建 MC- μ TP 电离装置。如图 1 所示,此装置由两个 T 型三通组成,分别作为等离子体发生装置和反应区域。向第一个 T 型 PEEK 三通直管一端插入 30G 平头不锈钢空心针作为微电极,另一端插入熔融石英毛细管作为等离子体约束腔,微电极针尖插入毛细

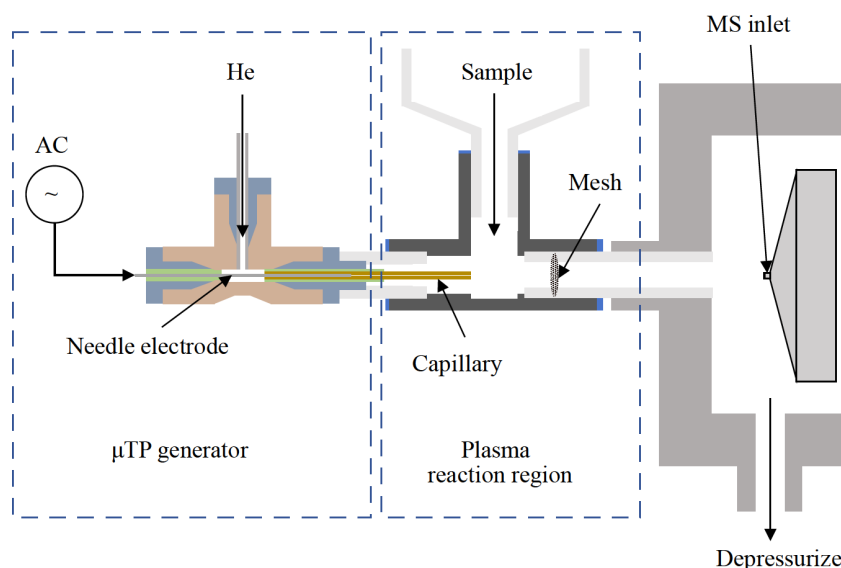


图1 筛网碰撞-微管等离子体(MC- μ TP)电离装置结构示意图

Fig.1 Schematic of structure of mesh collision enhanced microtube plasma ionization (MC- μ TP) device

管中,距毛细管尾端约 5 mm;在支管处接入高纯 He 作为载气,使用气体质量流量计控制流速。将第二个 T 型 PBT 三通直管一端与第一个三通毛细管端相连,另一端使用硬质硅胶管固定 200 目铜质筛网并与 VAPUR 接口相连;在支管处安装集气罩用于吸收样品烟气。在组装过程中,需确保不锈钢针、毛细管和质谱仪离子传输管轴线对齐,避免等离子体偏移。同时,使用规格合适的 PEEK 套管填充微电极、毛细管和 T 型三通之间的空隙,以保证装置的气密性。

1.2.2 样品前处理

使用 ESIB-MC- μ TP-MS 直接分析纺织品时,通常情况下无需复杂的样品前处理过程。本方法对样品形貌和尺寸均无特别要求,只需保证所取样品具有代表性且重量不少于 5 mg。在采集样品数据前,使用无尘纸蘸取甲醇清洁样品表面,清除灰尘与潜在干扰物。

1.2.3 MC- μ TP-MS 法

MC- μ TP 电离装置的工作条件:交流电压(等效值)为 2.0 kV,频率为 10 kHz;载气(He)流速为 0.5 L/min;装置与质谱仪接口负压为 -76 kPa;工作温度为室温。静电场轨道阱高分辨质谱仪的工作条件:离子传输管温度为 320 $^{\circ}\text{C}$, S-lens 电压为 55 V,负离子模式,数据采集使用全扫描模式,扫描范围为 m/z 100~750,自动增益控制目标值为 10^6 ,最大进样时间为 50 ms,每秒钟微扫描 1 次。

使用手持式电烙铁(350 $^{\circ}\text{C}$)灼烧纺织品样品,产生的烟气由 MC- μ TP-MS 分析,单次数据采集时间为 4~5 s,每个标准样品平行采集至少 6 次,实际样品平行采集至少 3 次。每次数据采集完成后,丢弃当前样品灼烧残余,对电烙铁头进行清理,再进行下一个样品数据采集,更换样品用时约 30 s。当质谱总离子流信号强度下降或质谱背景中出现来自样品烟气的离子残留时,应使用甲醇清洗毛细管、筛网与集气罩后再进行实验。

1.2.4 数据预处理

数据预处理脚本、RF 模型的编译与运行环境为 Python 3.10.4,操作系统为 Windows 10 专业版(64 位);计算机处理器为 Intel CoreTM i5-13600KF,机带 RAM 为 32.0 GB。

原始质谱数据预处理的具体步骤如下:(1)提取质量轴范围 100.0~550.0 Da 内的质谱信息(精确质量数和信号强度);(2)使用局部峰检测结合信噪比(S/N)筛选,对导入数据进行降噪,噪声估计窗口宽度设为 50 Da,峰检测最低强度阈值设为 10^5 cps,噪声阈值设为 3 倍信噪比;(3)对保留的信号强度进行最大值归一化,将其缩放至 0~1 范围内;(4)100.0 和 550.0 Da 分别作为分箱起点和终点,0.1 Da 作为分箱步长, ± 0.05 Da 作为分箱半径,建立连续非重叠的全局分箱;(5)将保留的离子信号按质量数分配至各

箱,同一箱内的多个信号强度保留最大值;(6)合并全部样品质谱信息,形成质谱数据集。

1.2.5 RF 分类模型的建立

采用 Python 环境下的 Scikit-learn 机器学习库^[22],构建基于 RF 算法的纺织品种类鉴别模型,用于鉴别纺织品实际样品。在模型训练和实际样品鉴别过程中,对所用数据集进行 Z-score 标准化,并使用主成分分析(Principal component analysis, PCA)降维($n_{\text{component}}=25$)^[23]。用于实际样品鉴别的 RF 模型初始超参数设置如下:决策树数量($n_{\text{estimators}}$)为 100,单棵决策树最大深度(max_depth)限制为 10,在分裂节点时随机选择的特征个数为最大特征数(max_features)以 2 为底的对数(向下取整),节点分裂最小样本数(min_samples_split)设为 2,叶节点最小样本数(min_samples_leaf)为 4。采用自助采样法(Bootstrap=True)生成训练子集,并保留袋外(Out-of-Bag, OOB)数据用于模型性能评估($\text{oob_score}=\text{True}$)。为确保实验可重复性,在整个建模过程中(包括超参数优化、模型训练及验证阶段)固定随机数种子。

1.2.6 红外光谱法测定纺织品纤维成分

采用红外光谱法对 RF 预测置信度较低的样品进行检测,红外光谱仪的工作条件:分辨率为 8 cm^{-1} ,扫描次数为 32,扫描范围为 $4000\sim 400\text{ cm}^{-1}$,测试方式为衰减全反射(ATR),每个样品平行测定 3 次,取其稳定光谱作为该样本的基本光谱。

2 结果与讨论

2.1 适用于纺织品质谱数据采集的 ESIB-MC- μ TP-MS 法

2.1.1 MC- μ TP-MS 参数的优化

为减少每次质谱分析后的信号残留,同时降低所得质谱数据的维度,采用负离子模式对纺织品样品质谱数据进行采集。在此基础上,以特定样品(电子版文后支持信息表 S1,序号 1)的总离子流(Total ion current, TIC)信号强度作为评价指标,分别对 MC- μ TP 电离装置的交流电压(等效值)、VAPUR 接口负压和载气流速 3 个工作参数进行优化。如图 2A 所示,随着交流电压等效值增大,等离子体电离效率逐渐升高;当电压大于 2.0 kV 时, TIC 信号强度缓慢下降;当电压大于 2.5 kV 时,电离装置微电极与质谱仪离子传输管入口之间发生放电,无法产生质谱信号。如图 2B 所示,随着载气流速增大,等离子体电离效率增高;当流速高于 0.5 L/min 时, TIC 变化趋于平稳。如图 2C 所示,随着质谱仪大气压接口处负压增大,

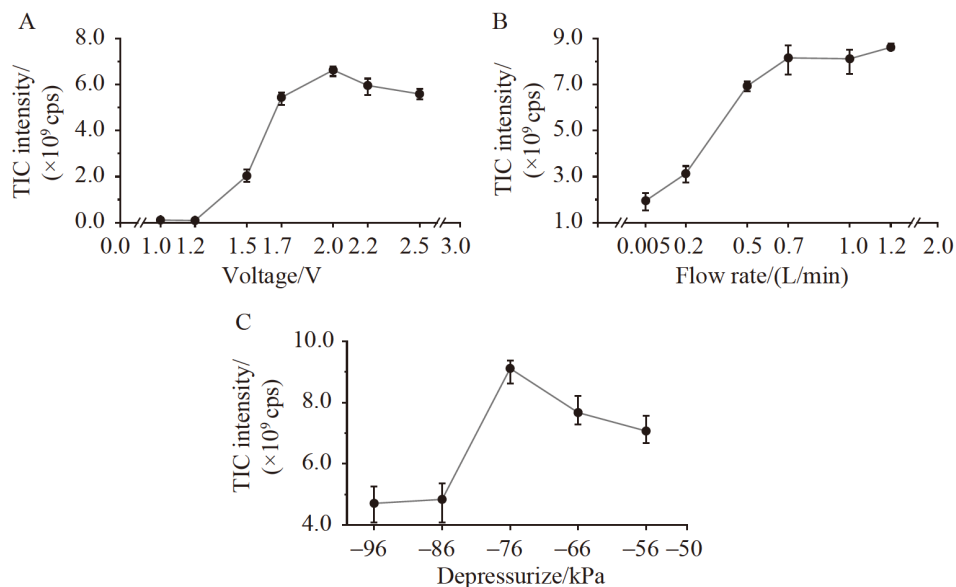


图2 MC- μ TP 电离装置的交流电压(A)、载气流速(B)和 VAPUR 接口负压(C)对总离子流(TIC)信号强度的影响

Fig.2 Effects of alternating current (AC) voltage (A), carrier gas flow rate (B) and negative pressure at VAPUR interface (C) of MC- μ TP ionization device on total ion current (TIC) intensity

TIC 先增大再减小; 当负压为 -76 kPa 时, TIC 达到最大值。值得注意的是, 当负压低于 -86 kPa 时, 尽管 TIC 下降, 但质谱信号残留得到明显抑制。综合考虑 TIC 信号强度、仪器稳定性与检测成本, 选择交流电压(等效值) 2.0 kV、载气流速 0.5 L/min、装置与质谱仪接口负压 -76 kPa 作为后续实验的工作参数。

2.1.2 纺织品质谱数据采集

使用手持式电烙铁灼烧纺织品样品产生烟气, 经 MC- μ TP 装置电离后, 供静电场轨道阱高分辨质谱仪分析。考虑到本研究涉及的棉和羊毛/羊绒的热分解温度分别为 353 和 227 $^{\circ}\text{C}$ ^[24], 聚酯纤维和腈纶的热分解温度分别为 350 和 $200\sim 250$ $^{\circ}\text{C}$ ^[25], 将电烙铁灼烧温度设定为 350 $^{\circ}\text{C}$, 以在获取不同种类纤维热分解特征产物质谱信息的同时, 最大程度抑制烟气信号残留。采集的质谱数据经过预处理后, 合并为质谱数据集(包含 181 组数据, 4501 个变量维度), 用于后续机器学习。经最大值归一化处理的棉、羊毛/羊绒、聚酯纤维和腈纶特征质谱图如图 3 所示, 4 类纤维在碎片离子种类与数量和基峰质荷比等方面均存在较大差异, 表明通过标准样品质谱数据集训练预测模型鉴别未知样品纤维种类具有可行性。

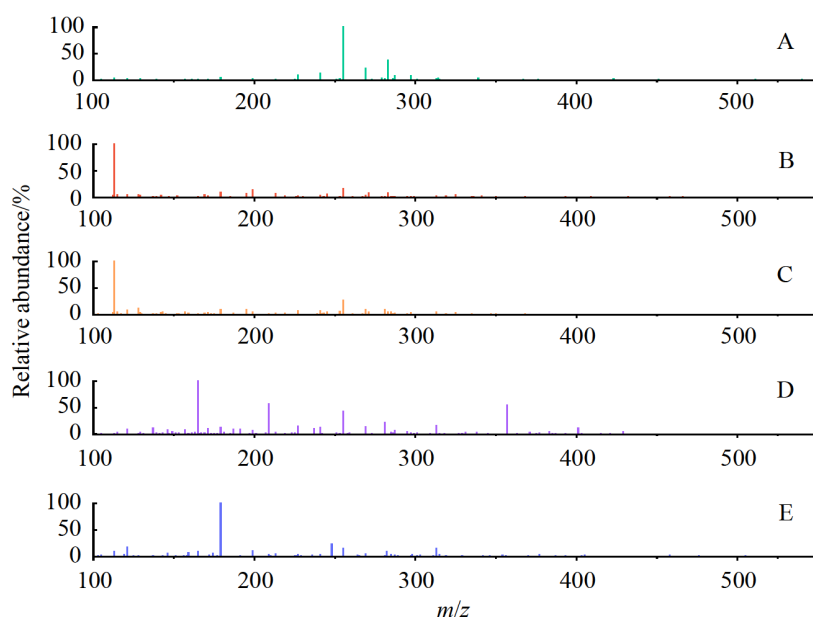


图3 不同纤维成分纺织品标准样品的代表性质谱图(100.0~550.0 Da): (A)棉; (B)羊绒; (C)羊毛; (D)聚酯纤维; (E)腈纶

Fig.3 Representative mass spectra of textile standard samples with different fiber compositions (100.0–550.0 Da): (A) Cotton; (B) Cashmere; (C) Wool; (D) Polyester; (E) Acrylic

2.2 使用无监督分类方法评价纺织品质谱数据集的质量

为评估纺织品质谱数据集的质量并揭示其潜在结构, 采用无监督分类方法进行数据分析、异常值检测与特征有效性验证。首先, 通过 PCA 将纺织品质谱数据集降至 25 维, 以压缩冗余信息、保留核心数据结构, 并降低计算复杂性。图 4A 展示了 PCA 碎石图, 其中, 前 10 个主成分累积解释了 95% 的方差, 表明降维后数据仍能表征原始特征的主要差异。然后, 采用 K 均值聚类算法^[23]对降维后的数据集进行无监督分类, 以评价其数据分群性, 所使用的聚类数(n_{clusters})依据轮廓系数优化为 4。如图 4B 所示, 数据集中各样品点在 PCA 前 3 个主成分的降维空间中成簇分布, 且簇间距离较大, 表明纺织品质谱数据可分性较强。值得注意的是, 羊绒与羊毛的样品点空间分布重叠, 无法独立分簇, 这是由于二者的主要化学成分均为 α -角蛋白(α -Keratin), 仅在氨基酸成分和含量上有较小区别^[26], 导致质谱特征区分度不足。后续实验中将两类样品的分类标签统一为“羊毛/羊绒(Wool/Cashmere)”。最后, 分别采用调整兰德指数(Adjusted rand index, ARI)^[27]、戴维森堡丁指数(Davies-Bouldin index, DBI)^[28]和 Calinski-Harabasz 指数(CHI)^[29]评价无监督分类效果。结果显示, ARI 为 0.833, 表明聚类结果与样品的分组标签高度一致, 数据具有明确的类别结构; DBI 为 0.518, 表明所有簇内样本分布比较紧密, 且不存在明显重叠的簇对;

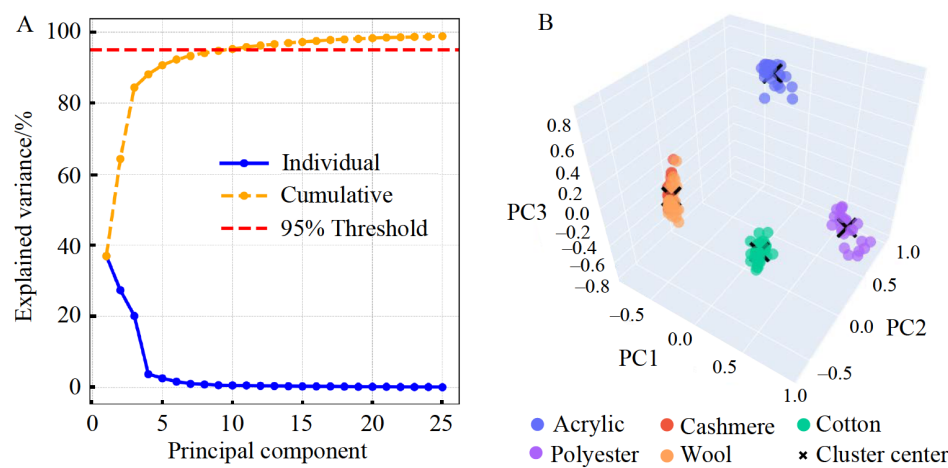


图4 基于纺织品质谱数据集的主成分分析(PCA)碎石图(A)和 K 均值聚类三维可视化图(B)
Fig.4 Principal component analysis (PCA) scree plot (A) and 3D visualizaition (B) of K -means clustering

CHI 为 299.347,表明全局聚类结构清晰,类间差异显著。综上所述,所得纺织品质谱数据集质量可以满足有监督学习要求。

2.3 基于纺织品质谱数据集建立 RF 分类模型

将纺织品质谱数据集按照 7:3 的比例随机拆分为训练集和测试集,用于构建 RF 分类模型,并对其性能进行评价。首先,基于经过 Z-score 标准化和 PCA 降维($n_components=25$)的训练集,通过网格扫描对 RF 模型的超参数进行优化,涉及的 5 个超参数的取值范围与优化结果见表 1。然后分别使用 OOB 数据和测试集,对优化后的模型性能进行评估。结果表明,该模型的 OOB 得分达到 0.9762,表明模型在训练过程中,在随机未被单棵树使用的样本(即袋外样本)上表现良好,反映出模型对训练数据的泛化能力较强,且过拟合风险较低。5 折交叉验证平均准确率为 $0.9683 (\pm 0.0299)$,标准差较小,显示模型在不同数据子集上的性能波动较小,适应能力稳健。如图 5A 所示,模型对 4 种纤维分类的精确率、召回率和 $F1$ 分数均不低于 0.889,表明模型对各类别的判别能力均衡,未出现显著类别偏差。测试集准确率为 0.9636,与训练集性能相近。从测试集的混淆矩阵(图 5B)中可以看出,羊绒/羊毛和腈纶各有 1 个误判,可能是特征区分度不足或样本标注噪声所致;聚酯纤维和腈纶分别仅有 8 个和 9 个测试样本,其高准确率可能存在过拟合风险。

表 1 基于训练集优化随机森林(RF)模型超参数
Table 1 Optimization of hyperparameters for random forest (RF) model based on training dataset

序号 Entry	超参数 Hyperparameter	取值范围 Range of value	优化结果 Optimization result
1	决策树数量 $n_estimators$	100, 200, 500, 1000, 2000	100
2	最大深度 max_depth	None, 10, 20, 30	10
3	最大特征数量 $max_features$	None, sqrt, log2	log2
4	节点分裂最小样本数 $min_samples_split$	2, 5, 10	2
5	叶节点最小样本数 $min_samples_leaf$	1, 2, 4	4

2.4 实际样品鉴别

使用ESIB-MC- μ TP-MS 法采集海关抽检的穿着类奢侈品样品的质谱数据,每个样品平行采集 3 次。此后,合并训练集和测试集,使用优化后的超参数训练 RF 模型,并对样品进行纤维成分鉴别,综合 3 次检

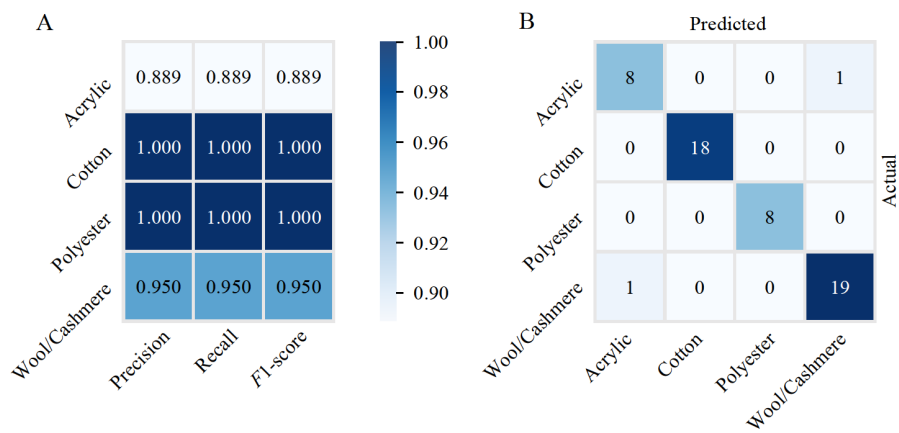


图5 训练的 RF 模型的分类报告 heatmap (A) 和测试集混淆矩阵 (B)

Fig.5 Heatmap of the classification report (A) and confusion matrix of test dataset (B) for trained RF model

测结果最终确定其纤维成分种类。如表 2 所示,对来自 8 个品牌的 30 件穿着类奢侈品鉴别的纤维成分与标识成分基本一致。其中, 20 个样品的预测置信度为 1.00,代表 3 次平行鉴别的结果均与标识成分一致,这些样品主要为棉和羊绒/羊毛材质,表明纺织品质谱数据集中较大的样本量对准确分类具有正面影响。8 个样品置信度为 0.67,主要分布在羊毛、聚酯纤维和腈纶类别,这可能是实际样品质谱数据中存在少量离群值所导致,可以通过增加平行鉴别次数来降低离群值影响,提高置信度。值得注意的是,聚酯纤维和腈纶各 1 个样品(表 2,序号 3 和 27)的置信度仅为 0.33,鉴别结果难以直接采信。

表 2 使用优化的 RF 模型鉴别 30 种穿着类奢侈品样品纤维成分

Table 2 Textile fiber identification of 30 kinds of luxury apparel samples using the optimized RF model

序号 Entry	品牌 Brand	标识成分 Claimed composition	预测分组标签 Predicted group label	置信度 Confidence (n=3)
1	品牌 1 Brand 1	腈纶 Acrylic	腈纶 Acrylic	1.00
2	品牌 1 Brand 1	腈纶 Acrylic	腈纶 Acrylic	0.67
3	品牌 1 Brand 1	腈纶 Acrylic	腈纶 Acrylic	0.33
4	品牌 2 Brand 2	羊毛 Wool	羊毛/羊绒 Wool/Cashmere	1.00
5	品牌 2 Brand 2	聚酯纤维 Polyester	聚酯纤维 Polyester	1.00
6	品牌 2 Brand 2	棉 Cotton	棉 Cotton	1.00
7	品牌 2 Brand 2	聚酯纤维 Polyester	聚酯纤维 Polyester	0.67
8	品牌 3 Brand 3	羊毛 Wool	羊毛/羊绒 Wool/Cashmere	1.00
9	品牌 3 Brand 3	棉 Cotton	棉 Cotton	1.00
10	品牌 3 Brand 3	羊毛 Wool	羊毛/羊绒 Wool/Cashmere	1.00
11	品牌 4 Brand 4	羊毛 Wool	羊毛/羊绒 Wool/Cashmere	0.67
12	品牌 4 Brand 4	羊毛 Wool	羊毛/羊绒 Wool/Cashmere	1.00

续表2 (Continued to Table 2)

序号 Entry	品牌 Brand	标识成分 Claimed composition	预测分组标签 Predicted group label	置信度 Confidence (<i>n</i> =3)
13	品牌 4 Brand 4	羊毛 Wool	羊毛/羊绒 Wool/Cashmere	1.00
14	品牌 4 Brand 4	羊毛 Wool	羊毛/羊绒 Wool/Cashmere	0.67
15	品牌 5 Brand 5	棉 Cotton	棉 Cotton	1.00
16	品牌 5 Brand 5	羊毛 Wool	羊毛/羊绒 Wool/Cashmere	1.00
17	品牌 5 Brand 5	羊毛 Wool	羊毛/羊绒 Wool/Cashmere	1.00
18	品牌 5 Brand 5	羊毛 Wool	羊毛/羊绒 Wool/Cashmere	0.67
19	品牌 6 Brand 6	棉 Cotton	棉 Cotton	1.00
20	品牌 6 Brand 6	棉 Cotton	棉 Cotton	1.00
21	品牌 6 Brand 6	棉 Cotton	棉 Cotton	1.00
22	品牌 6 Brand 6	棉 Cotton	棉 Cotton	1.00
23	品牌 6 Brand 6	棉 Cotton	棉 Cotton	1.00
24	品牌 7 Brand 7	腈纶 Acrylic	腈纶 Acrylic	0.67
25	品牌 7 Brand 7	棉 Cotton	棉 Cotton	1.00
26	品牌 7 Brand 7	聚酯纤维 Polyester	聚酯纤维 Polyester	0.67
27	品牌 7 Brand 7	聚酯纤维 Polyester	聚酯纤维 Polyester	0.33
28	品牌 8 Brand 8	棉 Cotton	棉 Cotton	1.00
29	品牌 8 Brand 8	棉 Cotton	棉 Cotton	1.00
30	品牌 8 Brand 8	棉 Cotton	棉 Cotton	0.67

为了进一步确认这两个低置信度样品的真实属性,验证 RF 预测模型的鉴别准确率,采用红外光谱法^[30]对其进行测定。如图 6A 所示,在腈纶标准样品的红外谱图中,1731 cm⁻¹ 处强吸收带可归属为甲基丙烯酸甲酯基特征吸收,1044 cm⁻¹ 处强吸收带可归属为磺酸基的特征振动。品牌 1 序号 3 样品与腈纶标准样品红外吸收带位置基本一致,由此可判断其材质为腈纶。在图 6B 中,聚酯纤维标准样品谱图中 1721 cm⁻¹ 处的吸收带可归属为 C=O 伸缩振动,1242 cm⁻¹ 处的强吸收带可归属为 C—O 伸缩振动,714 cm⁻¹ 处的吸收带可归属为对位双取代苯环上的 —CH₂ 面内振动。考虑到品牌 7 序号 27 样品与聚酯纤维标准样品谱图基本一致,判断其材质为聚酯纤维。以上结果表明,红外光谱法检测结果与 RF 模型鉴别结果一致,这两个样品实际的纤维成分与标识成分不存在差异。在这种情况下,鉴别置信度低的原因可能是对样品进行平行测定的次数过少,导致个别离群值的影响被放大;也可能是训练集中标签为腈纶或聚酯纤维的样本数量偏少,导致 RF 模型在鉴别相关实际样品材质时出现假阳性。

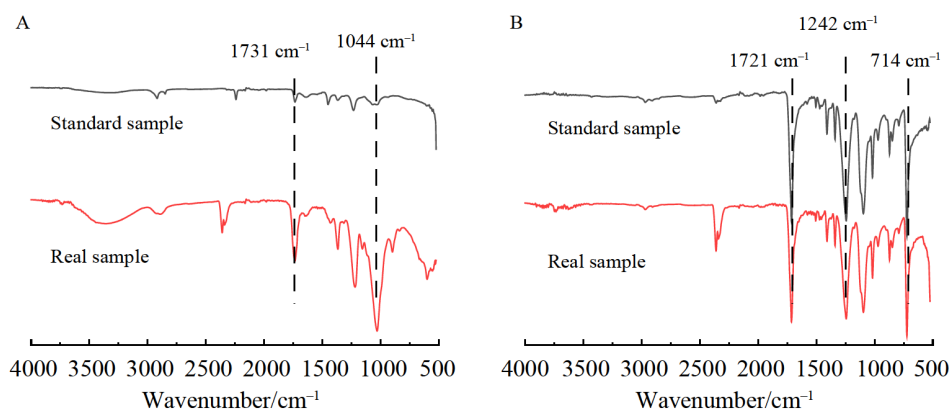


图6 对两个低鉴别置信度样品的傅里叶变换红外光谱分析: (A)品牌 1, 序号 3; (B)品牌 7, 序号 27

Fig.6 Fourier transform infrared spectroscopy (FT-IR) analysis of two low identification confidence samples: (A) Brand 1, entry 3; (B) Brand 7, entry 27

3 结论

采用市售材料自制的原位电离装置和静电场轨道阱质谱仪,建立了ESIB-MC- μ TP-MS法。本方法样品需求量小、无需复杂前处理、操作简便,适用于纺织品纤维成分的微损快速分析。本方法采集的纺织品质谱数据经PCA结合K-均值聚类算法评价质量后,用于训练RF模型,实现了纺织品纤维成分的准确分类。此模型综合性能良好,对4种天然或人造纤维分类的精确率、召回率和F1分数均不低于0.889,测试集准确率为0.9636。将此模型用于鉴别30个穿着类奢侈品样品的纤维成分,其中28个样品的鉴别结果与标识成分一致,基本可以满足对穿着类奢侈品样品纤维成分快速准确鉴别的需求。对鉴别结果与标识成分相同但置信度低的两个样品,进一步采用红外光谱法确认其成分。本方法在海关抽检和产品真实属性鉴别等场景中具有应用潜力,对相关行业具有指导意义。未来将从以下方面优化本方法,以进一步提高采样稳定性、数据可比性和鉴别准确率:(1)参考纤维热分析相关工作结论,优化电烙铁灼烧温度,以获得更高的热分解效率,并提升本方法对不同种类纤维(纺织品和皮革等)的兼容性;(2)进一步扩充纤维质谱数据库,提高RF模型性能,降低过拟合风险;(3)与其它鉴别方法(光谱学和显微成像等)结合,实现结果互相确认,以及建立基于不同来源鉴别结果的综合评价方法;(4)将RF模型特征重要性分析与串联质谱技术相结合,实现对不同种类纤维热分解特征化学产物的识别;(5)结合纤维成分鉴别方法与化学风险物质非靶向筛查技术,从真实属性和安全性两方面综合评价穿着类奢侈品质量。

References

- [1] General Administration of Customs of the People's Republic of China. Commodity Composition Table of Imports and Exports for February 2025 (in RMB). (2025-03-18) [2025-04-01]. <http://www.customs.gov.cn/customs/302249/zfxxgk/2799825/302274/302277/302276/6422378/index.html>.
中华人民共和国海关总署. 2025年2月进出口商品构成表(人民币值). (2025-03-18) [2025-04-01]. <http://www.customs.gov.cn/customs/302249/zfxxgk/2799825/302274/302277/302276/6422378/index.html>.
- [2] GB/T 2910.1—2009. Textiles-Quantitative Chemical Analysis-Part 1: General Principles of Testing. National Standards of the People's Republic of China.
纺织品 定量化学分析 第1部分: 试验通则. 中华人民共和国国家标准. GB/T 2910.1—2009.
- [3] FZ/T 01057.1—2007. Test Method for Identification of Textile Fibers-Part 1: General Introduction. Textiles Industry Standards of the People's Republic of China.
纺织纤维鉴别试验方法 第1部分: 通用说明. 中华人民共和国纺织行业标准. FZ/T 01057.1—2007.
- [4] SN/T 3236—2012. Test Method for Identification of Textile Fibers-Raman Spectrometry. Commodity Inspection Industry Standards of the People's Republic of China.
纺织纤维鉴别试验方法 拉曼光谱法. 中华人民共和国出入境检验检疫行业标准. SN/T 3236—2012.
- [5] GAŁUSZKA A, MIGASZEWSKI Z, NAMIEŚNIK J. TrAC, Trends Anal. Chem., 2013, 50: 78-84.

- [6] SHARMA V, MAHARA M, SHARMA A. *Forensic Chem.*, 2024, 39: 100576.
- [7] ZAPATA F, ORTEGA-OJEDA F E, GARCÍA-RUIZ C. *Spectrochim. Acta, Part A*, 2022, 268: 120695.
- [8] ZHOU C, HAN G, VIA B K, SONG Y, GAO S, JIANG W. *Text. Res. J.*, 2019, 89(17): 3610-3616.
- [9] PEETS P, LEITO I, PELT J, VAHUR S. *Spectrochim. Acta, Part A*, 2017, 173: 175-181.
- [10] CABRALES L, ABIDI N. *J. Therm. Anal. Calorim.*, 2010, 102(2): 485-491.
- [11] XIA Z, YAO C, ZHOU J, YE W, XU W. *Text. Res. J.*, 2016, 86(8): 856-867.
- [12] BASAK S, LAHA A. *Thermal Analysis of Polymeric Materials*. Weinheim, Germany: WILEY-VCH GmbH, 2022: 623-636.
- [13] RUAN Y, MENG X, WANG J, WANG H, YE Q, SHOU Q, MA Q. *J. Anal. Test.*, 2024, 8(4): 518-537.
- [14] COCHRAN K H, BARRY J A, MUDDIMAN D C, HINKS D. *Anal. Chem.*, 2012, 85(2): 831-836.
- [15] LIANG J, FRAZIER J, BENEFIELD V, CHONG N S, ZHANG M. *Anal. Chem.*, 2019, 92(2): 1925-1933.
- [16] YAO L, LAING R M, BREMER P J, SILCOCK P J, LEUS M J. *Text. Res. J.*, 2015, 85(17): 1817-1826.
- [17] LV Y, LI M, LI G, MA Q. *Anal. Chem.*, 2025, 97(4): 2086-2093.
- [18] GAO H, LIN J, JIA X, ZHAO Y, WANG S, BAI H, MA Q. *Talanta*, 2021, 225: 122069.
- [19] QI J, LI Y, ZHANG C, WANG C, WANG J, GUO W, WANG S. *Food Chem.*, 2021, 337: 127779.
- [20] NOVIYANTO A, ABDULLA W H. *J. Food Eng.*, 2020, 265: 109684.
- [21] BREIMAN L. *Mach. Learn.*, 2001, 45(1): 5-32.
- [22] PEDREGOSA F, VAROQUAUX G, GRAMFORT A, MICHEL V, THIRION B, GRISEL O, BLONDEL M, PRETTENHOFER P, WEISS R, DUBOURG V, VANDERPLAS J, PASSOS A, COURNAPEAU D, BRUCHER M, PERROT M, DUCHESNAY É. *J. Mach. Learn. Res.*, 2011, 12: 2825-2830.
- [23] HE Xiao-Qun. *Multivariate Statistical Analysis (Fifth Edition)*. Beijing: China Renmin University Press, 2019: 56, 106.
何晓群. 多元统计分析(第5版). 北京: 中国人民大学出版社, 2019: 56, 106.
- [24] ZHENG Shao-Ming. *China Fiber Insp.*, 2023, (7): 74-76.
郑少明. 中国纤检, 2023, (7): 74-76.
- [25] WANG Chen. *Huaxue Xianwei Jiagong Gongyixue*. Beijing: China Textile Press, 2022: 52, 155.
王琛. 化学纤维加工工艺学. 北京: 中国纺织出版社, 2022: 52, 155.
- [26] BUNSELL A R. *Handbook of Properties of Textile and Technical Fibres (Second Edition)*. Sawston, England: Woodhead Publishing Ltd., 2018: 61, 107.
- [27] HUBERT L, ARABIE P. *J. Classif.*, 1985, 2(1): 193-218.
- [28] DAVIES D L, BOULDIN D W. *IEEE Trans. Pattern Anal. Mach. Intell.*, 1979, PAMI-1(2): 224-227.
- [29] CALIŃSKI T, HARABASZ J. *Commun. Stat.*, 1974, 3(1): 1-27.
- [30] LI Ya-Hui, LI Jun-Fang, GUO Xing-Zhou, XI Guang-Cheng, YANG Hai-Feng. *Stand. Qual. Light Ind.*, 2025, (2): 83-86, 89.
李亚辉, 李俊芳, 郭兴洲, 席广成, 杨海峰. 轻工标准与质量, 2025, (2): 83-86, 89.

Rapid Identification of Textile Fiber Composition Using Microtube Plasma Ionization Mass Spectrometry Combined with Random Forest Algorithm

SHANG Yu-Han, LYU Yue-Guang, MENG Xian-Shuang, LYU Qing,
GUO Xiang-Yu, ZHANG Qing*

(Key Laboratory of Consumer Product Quality Safety Inspection and Risk Assessment of State Administration for
Market Regulation, Chinese Academy of Quality and Inspection & Testing, Beijing 100176, China)

Abstract A rapid and accurate method for textile fiber identification was developed for quality control and consumer protection. This method utilized electric soldering iron burning-mesh collision enhanced microtube plasma ionization mass spectrometry (ESIB-MC- μ TP-MS) to acquire textile fiber MS data and used a random forest (RF) prediction model to identify fiber composition based on these MS data. The MC- μ TP device involved in the method was a homemade low-temperature plasma ionization device constructed using cost-effective and readily available components. The system was applicable for direct analysis of small amount of textile samples without any

complex sample pretreatment processes. Characteristic thermal decomposition products of different fibers were generated via soldering iron burning (350 °C) in ambient atmosphere, and were subsequently analyzed by a mass spectrometer, with each analysis completed within 5 s. Raw MS data underwent noise reduction, normalization, and global binning steps to form a dataset, and its intrinsic class separability was evaluated using principal component analysis (PCA) combined with k -means clustering. Then, the RF model was trained based on the dimensionality-reduced textile fiber dataset. After grid search optimization, this model demonstrated robust performance with a 0.9762 out-of-bag score, a 0.9683 cross-validation accuracy (5-fold), and a 0.9636 test accuracy, supported by precision, recall, and $F1$ -scores exceeding 0.889 for all fiber classes. The method was applied to analysis of 30 luxury apparel samples from eight brands, among which 20 samples achieved 100% prediction confidence, aligning with labeled compositions. The identification result of two low-confidence samples was further confirmed using attenuated total reflection Fourier transform infrared spectroscopy (ATR-FT-IR). The method has been proven to be simple, portable and with minimal sample requirements for on-site customs inspections, providing a viable tool in the fight against counterfeit products, therefore supporting regulatory enforcement and consumer trust in the textile goods market.

Keywords Textiles; Microtube plasma ionization mass spectrometry; Random forest; Fiber composition identification

(Received 2025-04-22; accepted 2025-06-20)

Supported by the National Key Research and Development Program of China (No. 2022YFF0607204) and the Industrial and Consumer Goods Quality Safety Assurance Project of SAMR.