

基于多元分析的食用油油桶差分拉曼光谱的分类研究

刘克新¹, 姜红¹, 段斌², 刘峰²

(1.中国人民公安大学, 北京 100038; 2.南京简智仪器设备有限公司, 南京 210049)

摘要: **目的** 对纵火类案件现场所遗留食用油油桶物证进行快速无损检验及区分, 为公安机关侦查破案提供线索。**方法** 采用最新的差分拉曼光谱技术对 26 个食用油油桶及桶盖, 共计 52 个样品进行检验。首先, 通过传统谱图分析对 2 类样品进行初步分类, 同时, 结合多元分析, 对桶盖进行分层聚类, 对桶身再利用 Gap statistic 算法确定 k 值的基础上 K-means 聚类分析。**结果** 成功将桶盖分为了 3 类, 桶身分为了 3 类, 且对桶盖和桶身所得分类结果都与已知样品所属类别一致。**结论** 此方法可对食用油油桶进行快速无损检验及区分, 为现场遗留有食用油油桶的纵火案件侦破提供一定的帮助。

关键词: 差分拉曼光谱; 食用油; 油桶; 聚类分析; Gap statistic

中图分类号: TB489 文献标识码: A 文章编号: 1001-3563(2022)03-0129-06

DOI: 10.19554/j.cnki.1001-3563.2022.03.016

Classification of Differential Raman Spectra of Edible Oil Barrel Based on Multivariate Analysis

LIU Ke-xin¹, JIANG Hong¹, DUAN Bin², LIU Feng²

(1. People's Public Security University of China, Beijing 100038, China;

2. Nanjing Jianzhi Instrument and Equipment Co., Ltd., Nanjing 210049, China)

ABSTRACT: The work aims to conduct rapid non-destructive inspection and differentiation of the physical evidence of edible oil barrels left on the scene of arson cases, and provide clues for the public security organs to investigate and solve the case. A total of 52 samples of 26 pairs of edible oil barrels and bungs were tested by the latest differential Raman spectroscopy technology. Firstly, the two types of samples were preliminarily classified through traditional spectral analysis. At the same time, combined with multivariate analysis, hierarchical clustering was performed to the bungs and K-means cluster analysis was carried out to the barrels based on the Gap statistic algorithm to determine the k value. The bungs and barrels were successfully divided into 3 categories, respectively. The classification results were consistent with the category of the known samples. Through this method, the rapid non-destructive inspection and differentiation of edible oil barrels can be performed, which provides certain help for the detection of arson cases where edible oil barrels are left on the scene.

KEY WORDS: differential Raman spectroscopy; edible oil barrel; cluster analysis; Gap statistic

收稿日期: 2021-08-23

基金项目: 中国人民公安大学 2021 年度基科费重点项目 (2021JKF212); 南京简智仪器设备有限公司技术合作项目 (20191218)

作者简介: 刘克新 (1998—), 男, 中国人民公安大学硕士生, 主攻微量物证及理化检验。

通信作者: 姜红 (1963—), 女, 中国人民公安大学教授, 主要研究方向为微量物证分析。

当今社会,塑料制品已广泛地应用于人们社会生活的方方面面。利用塑料制品进行犯罪或者辅助犯罪的案件也逐渐增多^[1]。食用油桶由于其易获得性以及廉价性,常被不法分子当作容器以携带各种助燃剂。由于各种因素的影响,塑料油桶常常被犯罪分子或丢弃或遗弃在现场,因此侦查人员进行现场勘查时常能够提取到此类重要物证。通过对塑料油桶的检验鉴定,可以确定油桶底品牌、型号,以缩小侦查范围,为侦查破案提供线索。食用油桶分为桶身和桶盖 2 部分,由于犯罪分子的警惕性,桶身极有可能被犯罪分子带走,因此文中在对桶身检测的同时,对桶盖也进行了检验。桶身的主要材料是聚对苯二甲酸乙二醇酯(PET),桶盖的主要材料是聚乙烯(PE)^[2]。常见的塑料物证检测方法主要有红外光谱法^[3]、X 射线荧光光谱法^[4]、裂解气相色谱-质谱法^[5]等。相对于上述方法,拉曼光谱具有无损、快速、简单和无需样品准备等优点^[6-9],但其本身信号弱,抗干扰能力差,且对于荧光性强的物质检测能力较差。差分拉曼光谱在继承拉曼光谱优点的同时,很好地克服了上述缺点,是一种优良的现场快速检测方法^[10-13]。

文中通过对 26 个不同品牌、型号的食用油油桶桶身和桶盖进行差分拉曼光谱检测,获得 52 组光谱数据,先根据谱图特征峰对样本进行初步分类,但此过程费时费力,为减少人工,实现分类的自动化,快速准确地确定油桶所属品牌和型号,文中利用多元分析的方法,通过 SPSS、Origin、Matlab 软件,对 2 组数据分别进行分层聚类 and K-means 聚类。

1 实验

1.1 样本

生活中收集的 26 种不同种类的食用油油桶桶身及桶盖,见表 1。

1.2 仪器及参数设置

主要仪器:SSR—5000 型便携式差分拉曼光谱仪(南京简智仪器设备有限公司)。

参数设置:选用 2 种激发波长,分别为 784 nm 和 785 nm;激光功率为 440 mW,积分时间为 10 s;连续监测 5 次,连续监测时间间隔为 2 s。

1.3 样品预处理

对收集到的食用油油桶桶身与桶盖分别用无水乙醇进行擦拭,风干。

1.4 重复性检验

随机选取 1#桶盖和 13#桶身,分别在相同条件下,对同一部位,利用差分拉曼光谱仪进行 5 次重复性实验。

表 1 食用油油桶样品表
Tab.1 Samples of edible oil barrel

样品编号	样品名称及规格	产地
1	口福特选大豆油(5 L)	山东省烟台市
2	口福玉米油(4 L)	山东省烟台市
3	口福特香玉米油(5 L)	山东省烟台市
4	口福玉米油(4 L)	山东省烟台市
5	鲁花 5s 一级花生油(2.5 L)	山东省莱阳市
6	鲁花 5s 一级花生油(5 L)	山东省莱阳市
7	鲁花压榨玉米油(5 L)	山东省莱阳市
8	鲁花植物调和油(5 L)	山东省莱阳市
9	长寿花金胚玉米胚芽油(4 L)	山东省滨州市
10	长寿花玉米胚芽油(5 L)	山东省滨州市
11	长寿花玉米油(5 L)	山东省滨州市
12	龙大花生油(5 L)	山东省聊城市
13	龙大一级花生油(5 L)	山东省烟台市
14	胡姬花古法花生油(5 L)	山东省青岛市
15	胡姬花植物调和油(4.5 L)	山东省青岛市
16	金龙鱼大豆油(5 L)	山东省烟台市
17	金龙鱼纯正玉米油(4 L)	山东省烟台市
18	福临门一级大豆油(5 L)	天津市滨海新区
19	福临门大豆油(5 L)	山东省菏泽市
20	福临门家乡味花生油(5 L)	山东省临沂市
21	福临门花生油(6.18 L)	安徽省蚌埠市
22	天下五谷大豆油(5 L)	山东省滨州市
23	香满园大豆油(5 L)	山东省烟台市
24	西王压榨玉米油(5 L)	山东省滨州市
25	还黄牌植物调和油(10 L)	天津市滨海新区
26	胡姬花古法花生油(6.18 L)	山东省青岛市

2 结果与讨论

2.1 重复性分析

根据相同条件下对 1#桶盖和 13#桶身的 5 次重复性实验,分别可以得出 2 种样品的差分拉曼光谱图。由图 1 和图 2 可以看出,2 种样品 5 次实验所得谱图峰数、峰位相同,峰形基本一致,表明差分拉曼光谱仪检验食用油油桶和桶盖重复性良好,可使用其进行检测。

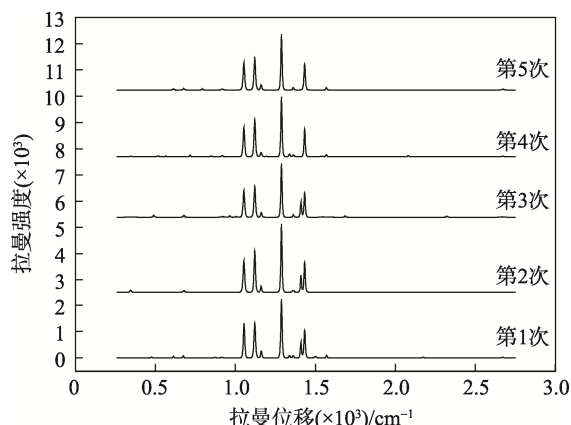


图 1 1[#]桶盖的差分拉曼光谱图
Fig.1 Differential Raman spectra of 1[#] bung

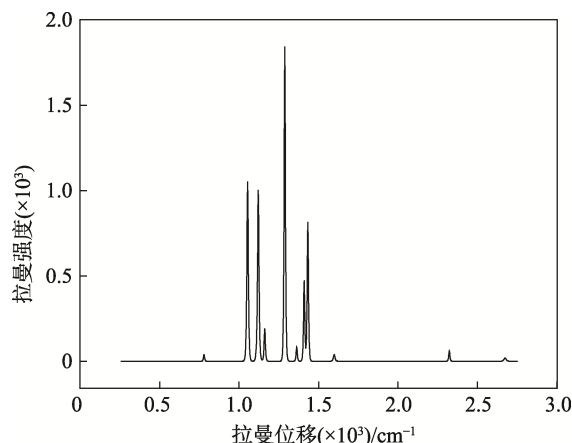


图 3 15[#]桶盖的差分拉曼光谱图
Fig.3 Differential Raman spectra of 15[#] bung

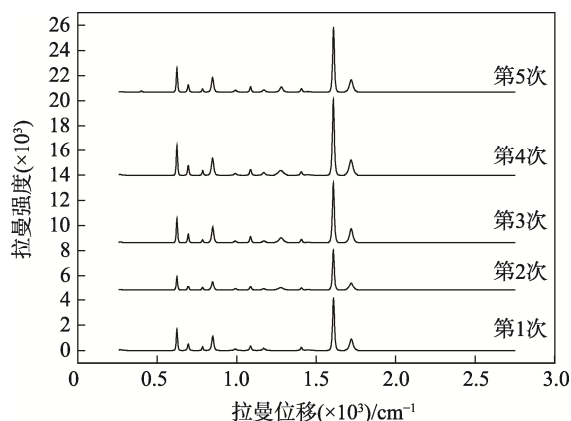


图 2 13[#]桶身的差分拉曼光谱图
Fig.2 Differential Raman spectra of 13[#] barrel

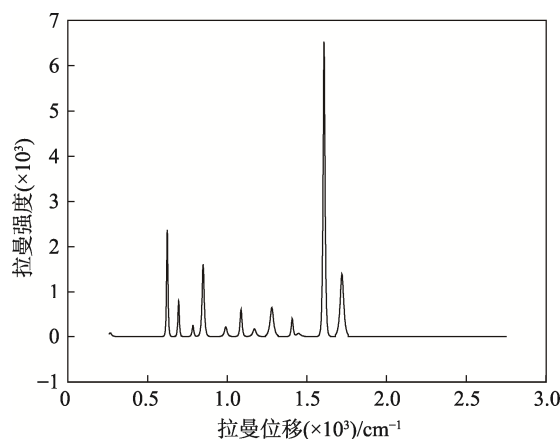


图 4 9[#]桶身的差分拉曼光谱图
Fig.4 Differential Raman spectra of 9[#] barrel

2.2 食用油油桶差分拉曼光谱分析

通过对 26 个样品桶盖与油桶差分拉曼谱图的对比分析,发现桶盖其主要成分都是聚乙烯 (PE),桶身主要成分都是聚对苯二甲酸乙二醇酯 (PET)。

2.2.1 桶盖差分拉曼光谱分析

以 15[#]桶盖为例,其最强峰出现在 1433、1288、1121、1053 cm^{-1} 附近 (见图 3)。其中 1433 cm^{-1} 处振动峰由碳原子与氢原子间非对称弯曲振动引起,1288 cm^{-1} 处振动峰由碳与氢原子间摇摆振动引起,1121 cm^{-1} 处振动峰由碳原子间对称伸缩振动引起,1053 cm^{-1} 处振动峰由碳原子间的非对称伸缩振动引起。

2.2.2 桶身差分拉曼光谱分析

以 9[#]油桶为例,发现其最强峰出现在 1720、1608、1087、849、624 cm^{-1} 附近 (见图 4)。其中 1716 cm^{-1} 处振动峰由苯环内碳原子间非对称伸缩振动引起,1602 cm^{-1} 处振动峰由羧基内碳氧原子间伸缩振动引起,1091 cm^{-1} 处振动峰由苯环与羧基间相连的 3 个碳原子间伸缩振动引起,858 cm^{-1} 处振动峰由苯环与

2 个羧基间相连的碳原子间伸缩振动引起,633 cm^{-1} 处振动峰由苯环内碳原子间对称弯曲振动引起。

2 类样品的谱图中除了具有各自主要成分聚丙烯与聚对苯二甲酸乙二醇酯的特征峰之外,还有其他振动峰,这说明样品中除其主要成分外还存在其他物质。这些物质是塑料在生产过程中为改善其性质或降低其成本而加入的增塑剂、着色剂、填料、热稳定剂等^[7]。根据桶盖谱图中是否存在 1410 cm^{-1} 振动峰,可以将油桶分为 2 类,同时 24[#]样品既在 1410 cm^{-1} 处没有振动峰,且 1410 cm^{-1} 振动峰与 1161 cm^{-1} 振动峰强度之比明显远大于其他样品,需单独归为一类,因此根据桶盖的明显谱图特征可以将油桶分为 3 类,见表 2。根据桶身是否在 1280 cm^{-1} 处存在振动峰,可以将样品分为 2 类,见表 3。

2.3 谱图数据处理

通过谱图以及数据分析,从桶盖光谱信息中提取出 7 组具有代表性的振动峰峰值数据,从油桶光谱信息中提取出 11 组具有代表性的振动峰峰值数据。为消除实验过程中因人为操作所引起的不同样品谱图

表 2 桶盖的分类
Tab.2 Classification of bungs

类别	1410 cm ⁻¹	振动峰强度 (1410 cm ⁻¹ /1161 cm ⁻¹)	样品编号
I	有	>5	1 [#] 、2 [#] 、3 [#] 、4 [#] 、9 [#] 、10 [#] 、11 [#] 、13 [#] 、14 [#] 、15 [#] 、16 [#] 、17 [#] 、20 [#] 、21 [#] 、22 [#] 、25 [#] 、26 [#]
II	无	30 ~ 50	5 [#] 、6 [#] 、7 [#] 、8 [#] 、12 [#] 、18 [#] 、19 [#] 、23 [#]
III	无	196	24 [#]

表 3 桶身的分类
Tab.3 Classification of barrels

类别	1280 cm ⁻¹	样品编号
I	有	1 [#] 、2 [#] 、3 [#] 、4 [#] 、6 [#] 、7 [#] 、8 [#] 、9 [#] 、10 [#] 、11 [#] 、12 [#] 、13 [#] 、14 [#] 、15 [#] 、16 [#] 、17 [#] 、18 [#] 、21 [#] 、23 [#] 、24 [#] 、25 [#] 、26 [#]
II	无	5 [#] 、19 [#] 、20 [#] 、22 [#]

总体强度之间的差异,对桶盖单个样品内部 7 组数据,两两之间求比值,得到 21 组数据。对所提取的油桶数据利用 origin 2019b 采用除以最大值的方法归一化到区间[0,1]。

2.4 桶盖的分层聚类分析

对经过数据处理后的 26 个样品的桶盖数据利用 SPSS 26 进行分层聚类。采用组间质心聚类度量类间距离,选择平方欧氏距离作为计算样品间距离。聚类结果见图 5,当测度距离为 1 时,样本可以分为 3 类,第 1 类包括口福牌食用油 (1[#]、2[#]、3[#]、4[#]、26[#])、长寿花食用油 (9[#]、10[#]、11[#])、龙大食用油 (产地山东烟台 13[#])、胡姬花食用油 (14[#]、15[#])、金龙鱼食用油 (16[#]、17[#])、福临门花生油及古法花生油 (20[#]、

21[#])、天下五谷食用油 (22[#])、海皇牌食用油 (25[#]、26[#]);第 2 类包括鲁花食用油 (5[#]、6[#]、7[#]、8[#])、龙大花生油 (产地山东聊城) (12[#])、福临门大豆油 (18[#]、19[#])、香满园食用油 (23[#]);第 3 类为西王牌食用油 (24[#])。

2.5 桶身的 K-means 聚类分析

对经过标准化处理的油桶数据,利用 K-means 算法进行聚类。K-means 算法的关键点在于 k 值的确定^[14]。在 K-mean 算法确定 k 值的问题里,可以使用肘部法,也可以用 Gap statistic 算法。肘部法则无法自动化地选出最佳的 k 值,且有时并不是那么清晰,因此斯坦福大学的 Robert 等教授提出了 Gap Statistic 方法^[15]。

Gap Statistic 的定义:

$$Gap(K) = E(\log D_k) - \log D_k \quad (1)$$

基本思路:引入参考的测值,这个参考值可以由 Monte Carlo 采样的方法获得。

$$Gap_n(k) = E_n^*(\log(W_k)) - \log(W_k) E_n^* \log(W_k) = (1/B) \sum_{b=1}^B \log(W_{kb}^*) \quad (2)$$

式中: B 为 sampling 的次数。为了修正 MC 带来的误差,通过计算标准差 S_k 来矫正 Gap Statistic。

$$w' = (1/B) \sum_{b=1}^B \log(W_{kb}^*) sd(k) = \sqrt{(1/B) \sum_b (\log W_{kb}^* - w')^2} s_k = \sqrt{\frac{1+B}{B}} sd(k) \quad (3)$$

选择满足 $Gap_k \geq Gap_{k+1} - s_{k+1}$ 的最小的 k 作为最优的聚类个数,即 Gap 取得最大值所对应的 k 值就是最佳簇数^[16]。

设置参考数据集 B 为 1000,最大聚类数 Max 为 10,利用 Matlab 实现代码运行,确定 k 值。由图 6 可以看出,当 k 值选择 3 时 Gap 值最大,为最佳 k

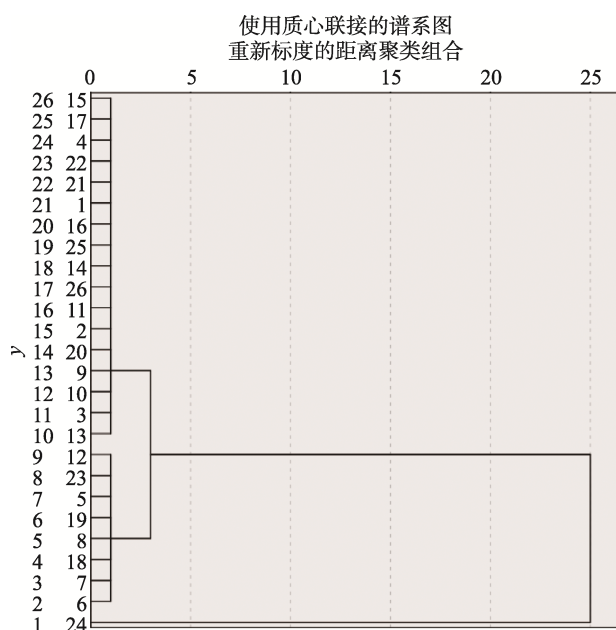


图 5 样品聚类谱系图
Fig.5 Pedigree of sample cluster

值。利用 K-means 聚类分析，设置 k 值为 3，得到聚类结果见表 4。K-means 算法所得样品分类结果与表 3 所得初步分类结果存在极大的趋同性。表 4（K-means 所得分类结果）中的 III 类与表 3（初步分类结果）中 II 类完全相同，表 4 中的 I、II 类是对表 3 中 I 类进一步细分，且符合已知样品类别。

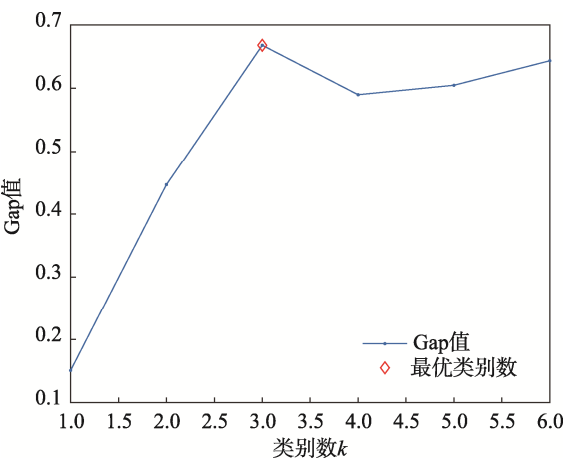


图 6 Gap(k)曲线
Fig.6 Gap(k) curve

表 4 桶身的分类
Tab.4 Classification of barrels

类别	样品编号
I	1 [#] 、2 [#] 、3 [#] 、4 [#] 、8 [#] 、9 [#] 、10 [#] 、12 [#] 、13 [#] 、14 [#] 、16 [#] 、18 [#] 、23 [#] 、24 [#] 、25 [#] 、
II	6 [#] 、7 [#] 、11 [#] 、15 [#] 、17 [#] 、21 [#] 、26 [#]
III	5 [#] 、19 [#] 、20 [#] 、22 [#]

3 结语

文中利用差分拉曼光谱技术，对所收集的 26 个食用油油桶及桶盖，共计 52 个样本分别进行快速无损检验，通过谱图分析对比，将桶盖初步分为 3 类，将桶身初步分为 2 类。结合多元统计学的方法，对谱图信息进行特征提取，利用分层聚类分析对桶盖进行分类，通过 Gap statistic 算法确定 k 值，对桶身利用 K-means 算法进行聚类，两者分类结果均符合已知样品已知类别。能够对不同品牌与型号的食用油油桶进行准确区分。

差分拉曼光谱结合多元分析的方法快速无损，可为现场留有食用油油桶的实际案件中的公安侦查工作确定食用油油桶所属品牌和型号提供一定的参考，提供侦查线索，缩小侦查范围。笔者将进一步收集更多样品，尝试多种其他方法，以期对食用油油桶进行更加科学简便分类。

参考文献：

[1] 姜红, 朱晓晗, 何歆沂, 等. 差分拉曼光谱法对塑料瓶的鉴别分析[J]. 中国测试, 2020, 46(8): 76-79.
JIANG Hong, ZHU Xiao-han, HE Xin-yi, et al. Differential Raman Analysis of Plastic Bottles[J]. China Measurement & Testing Technology, 2020, 46(8): 76-79.

[2] 董聘, 饶之帆, 杨晓云, 等. 几种塑料的拉曼光谱检测[J]. 塑料工业, 2011, 39(6): 67-70.
DONG Kun, RAO Zhi-fan, YANG Xiao-yun, et al. Determination of Several Plastic by Raman Spectroscopy[J]. China Plastics Industry, 2011, 39(6): 67-70.

[3] ANDRADY A L. The Plastic in Microplastics: A Review[J]. Marine Pollution Bulletin, 2017, 119(1): 12-22.

[4] HENDRICKSON E, MINOR E C, SCHREINER K. Microplastic Abundance and Composition in Western Lake Superior as Determined via Microscopy, Pyr-GC/MS, and FTIR[J]. Environmental Science & Technology, 2018, 52(4): 1787-1796.

[5] MAI Lei, BAO Lian-jun, SHI Lei, et al. A Review of Methods for Measuring Microplastics in Aquatic Environments[J]. Environmental Science and Pollution Research, 2018, 25(12): 11319-11332.

[6] TURNER A. Cadmium Pigments in Consumer Products and Their Health Risks[J]. The Science of the Total Environment, 2019, 657: 1409-1418.

[7] 林福华. 拉曼光谱技术在聚合物分析中的应用[J]. 塑料工业, 2018, 46(6): 132-135.
LIN Fu-hua. Application of Raman Spectroscopy in Polymer Analysis[J]. China Plastics Industry, 2018, 46(6): 132-135.

[8] ZHAO Ying, LIN Jun-feng, LIU Jia, et al. Study on the Method of Identifying Waste Plastic Materials Based on Raman Spectroscopy[J]. Spectroscopy and Spectral Analysis, 2021, 41(1): 122-126.

[9] SHCHEBLANOV N S, MANTISI B, UMARI P, et al. Detailed Analysis Of Plastic Shear in the Raman Spectra of SiO₂ Glass[J]. Journal of Non-Crystalline Solubung, 2015, 428: 6-19.

[10] 张进, 姜红, 刘峰, 等. 基于差分拉曼光谱的烟盒外包装薄膜检验研究[J/OL]. 中国测试: 1-9 [2021-03-23]. <http://kns.cnki.net/kcms/detail/51.1714.TB.20200827.1023.002.html>.
ZHANG Jin, JIANG Hong, LIU Feng, et al. Study on Inspection of Cigarette Packaging Film Based on Differential Raman Spectroscopy[J/OL]. China Test: 1-9[2021-03-23]. <http://kns.cnki.net/kcms/detail/51.1714.TB.20200827.1023.002.html>.

- [11] 张进, 姜红, 段斌, 等. 差分拉曼光谱结合化学计量学对口红的检验研究[J/OL]. 激光与光电子学进展: 1-13[2021-03-23]. <http://kns.cnki.net/kcms/detail/31.1690.TN.20210107.1716.006.html>.
ZHANG Jin, JIANG Hong, DUAN Bin, et al. Differential Raman Spectroscopy Combined with Stoichiometry for the Detection of Mouth Red[J/OL]. Laser and Optoelectronics Progress: 1-13[2021-03-23]. <http://kns.cnki.net/kcms/detail/31.1690.TN.20210107.1716.006.html>.
- [12] SHAMEEM K M M, CHOUDHARI K S, BANKAPUR A, et al. A Hybrid LIBS-Raman System Combined with Chemometrics: An Efficient Tool for Plastic Identification and Sorting[J]. Analytical and Bioanalytical Chemistry, 2017, 409(13): 3299-3308.
- [13] KLISIŃSKA-KOPACZ A, ŁYDŹBA-KOPCZYŃSKA B, CZARNECKA M, et al. Raman Spectroscopy as a Powerful Technique for the Identification of Polymers Used in Cast Sculptures from Museum Collections[J]. Journal of Raman Spectroscopy, 2019, 50(2): 213-221.
- [14] LANKERS M. Applications in: Environmental Analytics (Fine Particles)[J]. Physical Sciences Reviews, 2019, 4(6): 16-22.
- [15] 诸葛越. 百面机器学习[M]. 北京: 人民邮电出版社, 2018: 113-115.
ZHUGE Yue. Heterogeneous Machine Learning[M]. Beijing: People's Posts and Telecommunications Press, 2018: 113-115.
- [16] 王俊丰. 基于深度学习的多说话人语音分离技术研究 with 实现[D]. 北京: 中国电子科技集团公司电子科学研究院, 2020: 46-47.
WANG Jun-feng. Research and Implementation of Multi-Speaker Speech Separation Technology Based on Deep Learning[D]. Beijing: Electronic Science Research Institute of China Electronics Science and Technology Group Corporation, 2020: 46-47.