

鞋底材料的差分拉曼光谱可视化快速鉴别

张进¹, 姜红^{1*}, 刘峰², 段斌²

¹中国人民公安大学侦查学院, 北京 100038;

²南京简智仪器设备有限公司, 江苏 南京 210049

摘要 利用差分拉曼光谱技术和决策树模型,建立了常见鞋底材料种类鉴别的分类模型,为鞋底材料种类鉴别提供了一种新方法。通过前期谱图预处理后,根据样本主要成分的不同对样本进行分类,然后利用主成分分析和层次聚类分析验证分类结果;基于分类结果建立分类树模型,最终对 51 个样本实现了整体分类正确率为 98.0% 的识别分类,经交叉验证后正确率达 84.3%。这表明,利用差分拉曼光谱技术和决策树模型可以实现对鞋底样本光谱较为准确的识别与分类,为其他物证的区分鉴别提供了参考。

关键词 光谱学; 鞋底材料; 差分拉曼光谱; 决策树; 鉴别

中图分类号 TQ322

文献标志码 B

doi: 10.3788/LOP202158.0830004

Differential Raman Spectroscopy Visualization and Rapid Identification of Shoe Sole Materials

Zhang Jin¹, Jiang Hong^{1*}, Liu Feng², Duan Bin²

¹School of Investigation, People's Public Security University of China, Beijing 100038, China;

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

Abstract Using differential Raman spectroscopy and decision tree model, a classification model for identifying common shoe sole materials is established. This is a new method for identifying shoe sole materials. After preprocessing the pre-spectrum, the samples are classified according to the main components of the shoe sole, and then principal component and hierarchical clustering analyses are used to verify the classification results. Based on the classification results, a classification tree model is established, and finally 51 samples are recognized and classified with overall classification accuracy of 98.0% and the accuracy is 84.3% after cross-validation. These results show that using differential Raman spectroscopy and decision tree model can realize a more accurate identification and classification of shoe sole sample spectrum, providing a certain reference for the differentiation and identification of other physical evidence.

Key words spectroscopy; shoe sole material; differential Raman spectroscopy; decision tree; identification

OCIS codes 300.6170; 300.6450; 300.6330

1 引言

鞋底材料的种类鉴定是法庭科学领域中一项重要的研究课题。鞋底是案件现场中的一种常见物证,通过对提取到的鞋底物证进行检验,可以为侦查

提供线索,为破案、起诉以及法庭判决提供有力证据。

常见的鞋底材料包括乙烯-醋酸乙烯共聚物(EVA)、天然橡胶(NR)、聚氯乙烯(PVC)、顺丁橡胶(BR)等。目前,检验鞋底材料的主要方法有红外

收稿日期: 2020-08-28; 修回日期: 2020-09-16; 录用日期: 2020-09-24

基金项目: 中国人民公安大学 2019 年度基科费重点项目(2019JKF222)、南京简智仪器设备有限公司技术合作项目(20191218)

* E-mail: jiangh2001@163.com

光谱法^[1-5]、X 射线荧光光谱法^[6]和扫描电镜/能谱法^[7]等。差分拉曼光谱技术是一种新型的拉曼技术,它利用两个有轻微偏移的激发激光收集两张不同的拉曼光谱,基于后向传播(BP)神经网络技术,通过迭代最优解的方式,同时完成差分、对齐和基线校正。该技术可以准确分离差分信号和基线偏差,提取出纯净的拉曼光谱,克服了一般拉曼光谱技术受荧光干扰的问题,有效提高了系统的信噪比^[8]。唐左军等^[9]利用差分拉曼光谱技术采集了琥珀和柯巴树脂的差分拉曼谱图,然后采用关联度较高的“特征峰组”的方式进行差分拉曼光谱的解谱判定,为琥珀和柯巴树脂的鉴别提供了新方法。决策树模型是一种基本的分类与回归方法,与传统的回归和分类方法相比,决策树更接近人的决策模式。李志豪等^[10]利用激光拉曼光谱分析技术对两类易制毒化学品和易燃易爆化学品展开研究,他们采用决策树算法对样本数据进行处理,建立分类模型,并将分类结果与其他分类模型的结果进行比较,为验证一线执法、检查过程中所使用的拉曼光谱设备的准确性提供了参考。

本实验利用差分拉曼光谱技术采集样本的光谱数据,结合主成分分析、层次聚类分析和决策树模型对样本进行建模区分,以期为法庭科学领域鞋底材料的快速鉴别提供参考。

2 实验部分

2.1 实验样本

本实验收集了 51 个不同品牌、不同系列的白色运动鞋鞋底,用于后续的谱图采集及模型训练。

2.2 实验仪器及参数设置

实验仪器:便携式差分拉曼光谱仪(SEDERS Portable-Base)。

数据采集:选用两种激发波长,分别为 784 nm 和 785 nm;测试波数范围为 280~2600 cm^{-1} ,激光功率设置为 220 mW,积分时间设置为 10 s。

仪器校准:在合适的环境条件下,按照说明书要求对仪器进行预热稳定,并对位移准确度、位移重复性、光谱强度重复性、信噪比、检出限以及测量时间分别进行校准。

2.3 谱图预处理

为了有效分离差分信号和基线偏差,利用 BP 神经网络技术不断迭代寻找最优解,同时完成差分、对齐与基线校正,然后结合差分拉曼解调和去噪算法,提取出实验可用的纯净拉曼光谱谱图。

3 结果与讨论

3.1 差分拉曼光谱的组合谱图

通过解析谱图,根据样本的主要成分将 51 个样本分为 4 类,分类结果如表 1 所示。

表 1 51 个样本的分类结果

Table 1 Classification results of 51 samples

Category	Main ingredient	Sample serial number
1	EVA	1 [#] , 3 [#] , 8 [#] , 9 [#] , 10 [#] , 11 [#] , 12 [#] , 14 [#] , 16 [#] , 17 [#] , 18 [#] , 20 [#] , 21 [#] , 22 [#] , 23 [#] , 24 [#] , 26 [#] , 27 [#] , 28 [#] , 29 [#] , 31 [#] , 32 [#] , 33 [#] , 35 [#] , 39 [#] , 43 [#] , 44 [#] , 45 [#] , 50 [#]
2	NR	2 [#] , 4 [#] , 7 [#] , 13 [#] , 19 [#] , 25 [#] , 30 [#] , 36 [#] , 46 [#] , 51 [#]
3	PVC	5 [#] , 34 [#] , 40 [#] , 41 [#] , 42 [#] , 49 [#]
4	BR	6 [#] , 15 [#] , 37 [#] , 38 [#] , 47 [#] , 48 [#]

4 类样本的鞋底材料分别是 EVA、NR、PVC 和 BR。4 类样本的标准光谱与实际测试样本的光谱如图 1 所示,4 类样本的振动基团信息^[11-14]如表 2 所示。由图 1 可以看出,实际测试样本的谱图与标准光谱存在差异。这是因为在鞋底生产中,生产企业会添加不同种类、配比、来源和生产工艺的填料,用以提升产品性能和降低成本。虽然传统解析谱图的方法比较直观,但却不能避免人为带来的误差;因此,本实验从多角度对谱图数据进行深度挖掘,利用主成分分析结合聚类分析对样本进行分类,对表 1

的分类结果进行验证,考察分类的准确性。

3.2 系统聚类分析

主成分分析是一种常用的数据降维手段,它将多个具有相关性的变量转化成少数几个没有线性关系的综合变量,而这个综合变量又能涵盖原来多个变量的绝大多数信息。由于拉曼光谱数据量较大,冗长信息较多^[15-17],为避免维数灾难的影响,本文借助主成分分析对差分拉曼光谱数据进行降维处理。主成分分析一般提取特征值大于 1 且累计方差贡献率大于 85% 的主成分。由于前 26 个主成分的

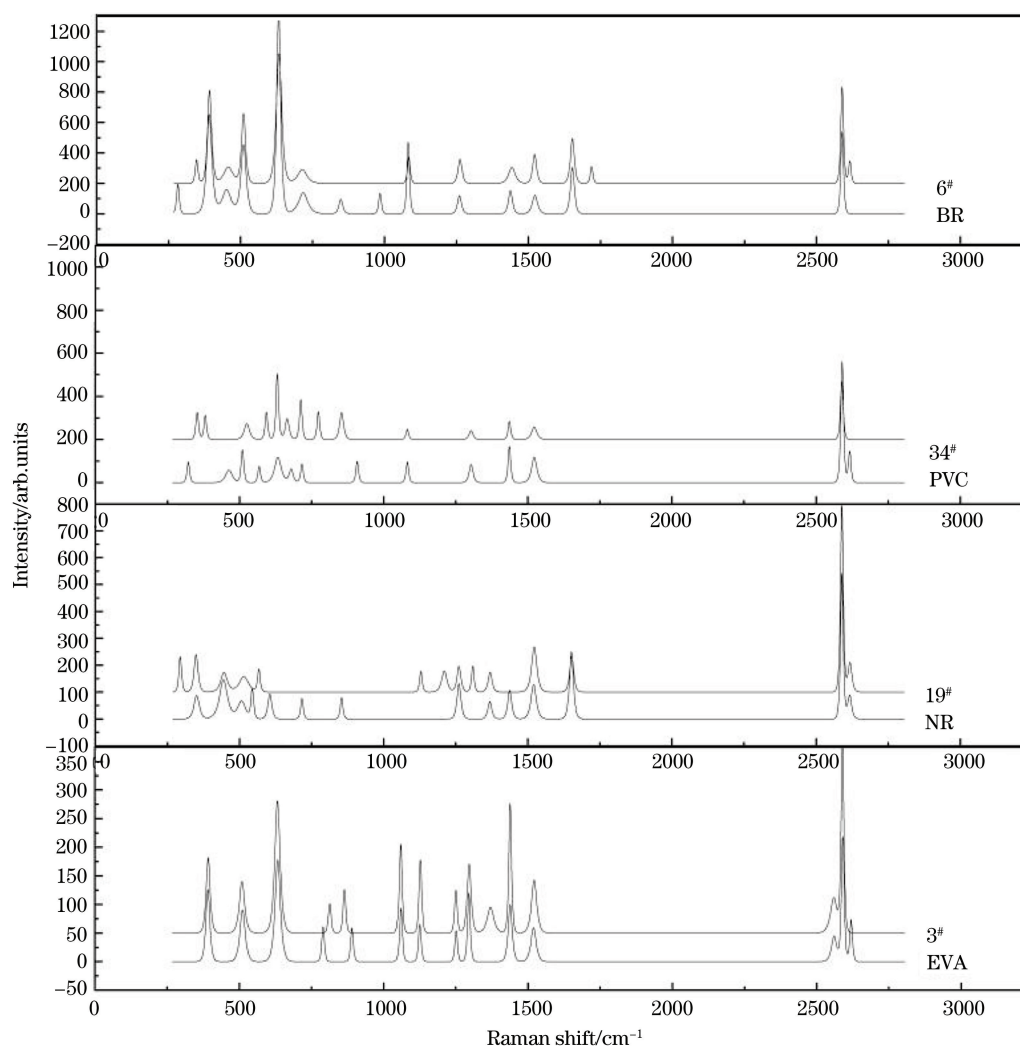


图 1 4 类样本的标准光谱与实际测试样本的光谱。(a)BR;(b)PVC;(c)NR;(d)EVA

Fig. 1 Standard spectra of four types of samples and actual test sample spectra. (a)BR;(b)PVC;(c)NR;(d)EVA

表 2 4 类样本的振动基团信息

Table 2 Vibration group information of four types of samples

Ingredient	Peak/cm ⁻¹	Vibration attribution
EVA	633	-C=O stretching vibration
	814,1059,1127	C-C stretching vibration
	1296	-CH ₂ non-planar rocking vibration
	1370	-CH ₃ symmetrical deformation vibration
	1435	-C-H shear vibration in CH ₂
NR	294,349,516	C-C-C stretching vibration
	1210	C-C stretching vibration
	1261	-CH ₂ in-plane deformation
	1369	-CH ₃ and -CH ₂ in-plane deformation
	1652	Trans C=C double bond stretching vibration
PVC	634,680	C-Cl stretching vibration
	1082	C-C stretching vibration
	1303	C-H bond asymmetric swing vibration
	1436	C-H bond bending vibration characteristic peak
BR	282,452	C-C-C stretching vibration
	985	C-C stretching vibration
	1437	-C-H stretching vibration in CH ₂
	1652	Trans C=C stretching vibration

特征值均大于 1,且累计贡献率为 85.61%,因此本文提取了前 26 个主成分用于后续模型的构建。

系统聚类分析是一种数据归约技术,属于无监督的学习方法,是将对象集划分为若干类别的过程。将降维后的数据进行系统聚类分析,聚类方法选用 Ward 法,区间测量选用平方欧氏距离,聚类结果树状图如图 2 所示。

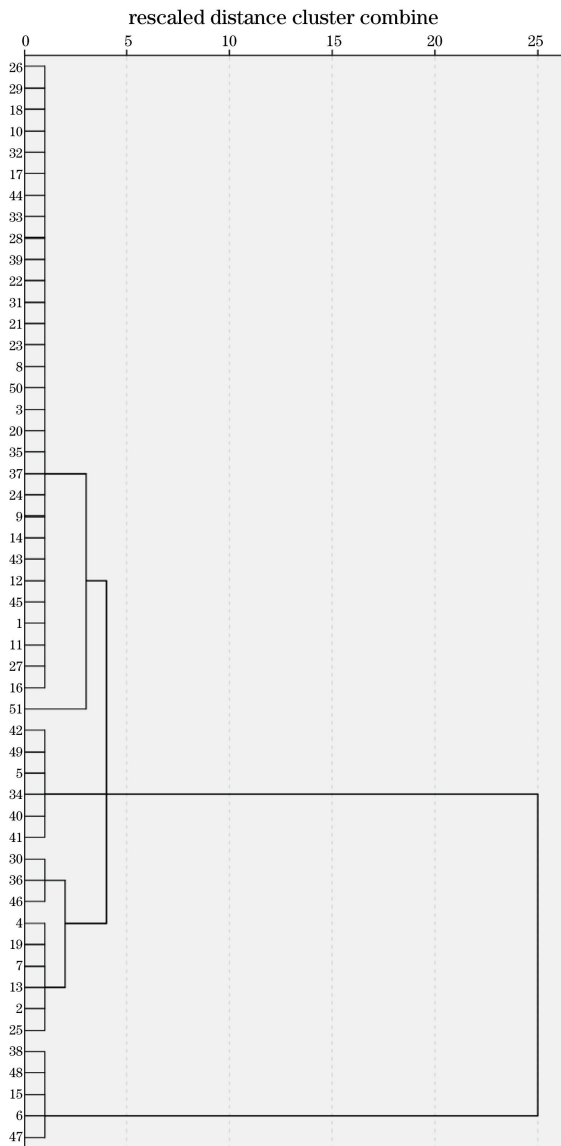


图 2 聚类树状图

Fig. 2 Clustering dendrogram

从图 2 可以直观地看出不同样本的划分分类情况。当类间距离为最小值时,样本被分为 6 类;当类间距离为 2 时,样本被分为 5 类;随着类间距离增大,样本被划分的类数逐渐减少,当类间距离为 25 时,所有样本被归为一类;当类间距离为 3 时,样本

被分为 4 类,分类结果与 2.1 节的分类结果基本一致,仅有两个样本(37# 样本和 51# 样本)被分到了错误的类别,总体分类正确率为 96.08%,分类结果理想。

3.3 CART 决策树模型

决策树是数据挖掘中一种简单、经典且被广泛应用的分类技术,是使用树结构算法对数据进行分类预测的方法^[18-19]。CART (Classification and Regression Tree)是由 Breiman 等于 1984 年提出的一种决策树算法,是利用二叉树不断地对由测试变量和目标变量构成的集合进行循环二分的决策树结构^[20]。相较于 C4.5 和卡方自动交互检测法 (CHAID),CART 决策树无需建立复杂的非线性模型,基于信息增益就可以以决策树图的形式直观地做出决策分类^[21]。

本实验采用 SPSS 25.0 软件对降维后的光谱数据进行建模仿真,选用 CART 作为生长法,将降维后提取的 26 个主成分作为变量,51 个样本作为训练集,建立决策树模型,得到了 4 种鞋底样本的决策树判定结果(如图 3 所示)、决策树模型的索引图(如图 4 所示)和 4 种鞋底样本的决策树分类结果(见表 3)。

表 3 4 种鞋底样本的决策树分类结果

Tab. 3 Decision tree classification results of four types of shoe sole samples

Practical testing	1	2	3	4	Classification accuracy/%
1	29	0	0	0	100
2	0	10	0	0	100
3	0	0	6	0	100
4	0	0	1	5	83.3
Overall classification accuracy/%	56.9	19.6	13.7	9.8	98.0

如图 3 所示,本文得出的是 1 棵 4 层决策树,树中一共包含 10 个树节点与 6 个终端节点,其中根节点包含 51 个样本(EVA 类 29 个,NR 类 10 个,PVC 类 6 个,BR 类 6 个)。主成分 1、主成分 3 和主成分 17 为该决策树模型的判定节点,如主成分 1 是第 1 个最佳判定节点,就以此分成二叉树;当数值 ≤ -0.056085 时,生成节点 1,在这个类别中有 90.6% 的样本被分到了 EVA 类,有 6.2% 的样本被分到了 PVC 类,3.1% 的样本被分到了 BR 类。

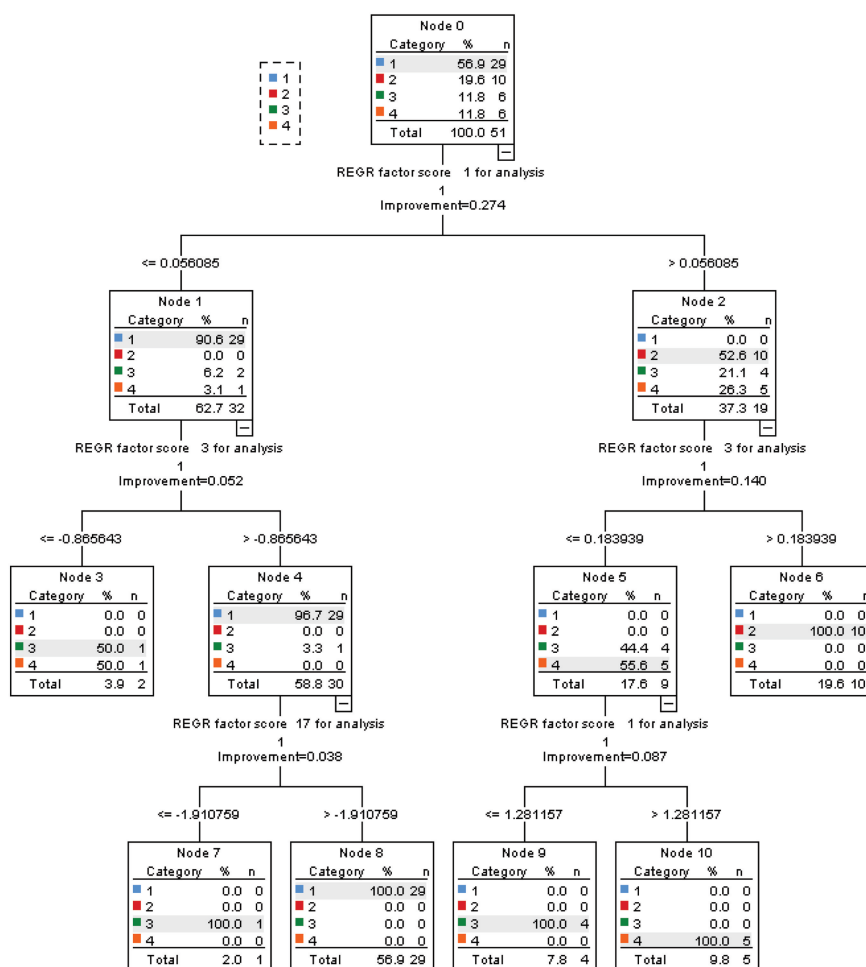


图 3 样本的 CART 分类树模型

Fig. 3 CART classification tree model of the sample

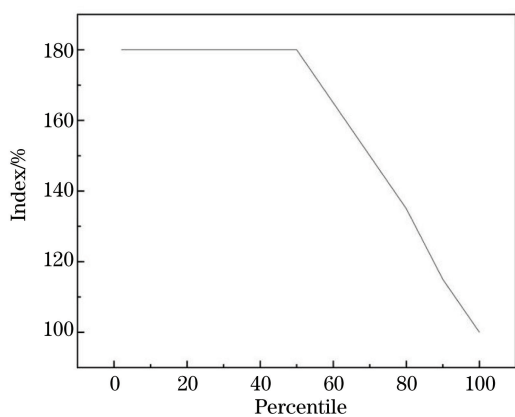


图 4 决策树模型目标类别为 EVA 的索引图

Fig. 4 Index graph of the decision tree model whose target category is EVA

图 4 是决策树模型目标类别为 EVA 的索引图。索引是目标分类的响应百分比除以总样本中该类的响应百分比,索引值表示所在节点观测目标分类百分比与期望目标分类百分比相差的程度^[22]。从图 4 中可以看出,曲线索引值从 100% 轴以上开

始,先沿移动方向保持水平,然后快速下降。该曲线走势表明该决策树模型是一个良好的模型。

表 3 是 51 个样本的分类结果,除第 4 类中有 1 个样本被错分到了第 3 类,其他样本均被分到了正确的类别中,整体分类正确率为 98.0%,经交叉验证后正确率为 84.3%,实验结果理想。

4 结 论

本实验利用差分拉曼光谱技术和决策树模型建立了一种新的分类鉴别鞋底材料的方法,该方法不仅实现了量少、快速和无损检验的目的,而且借助数学模型展开模式识别,实现了对鞋底样本更为合理的鉴别。模型对样本的总体分类正确率达 98.0%,经交叉验证后正确率达 84.3%,实验结果较为理想。后续将进一步针对性收集样本,扩充训练样本容量,并尝试使用感知神经网络等更多的模式识别方法,对样本数据进行进一步挖掘,以期为法庭科学领域中鞋底物证的区分鉴别提供新思路。

参 考 文 献

- [1] Lin X K, Lin O W, Hu X H, et al. Research on ATR-FTIR rapid identification method of sole materials[J]. *China Measurement & Test*, 2017, 43(8): 45-49.
林先凯, 林欧文, 胡秀红, 等. ATR-FTIR 快速识别鞋底材料的方法研究[J]. *中国测试*, 2017, 43(8): 45-49.
- [2] Wang J F, Gao C F, Xu B Q, et al. Rapid identification of sole materials with mid-infrared visible spectroscopy[J]. *China Plastics*, 2019, 33(8): 101-105.
王继芬, 高春芳, 徐佰祺, 等. 鞋底材料的中红外光谱可视化快速鉴别[J]. *中国塑料*, 2019, 33(8): 101-105.
- [3] Wang D, Jiang H. Correlation analysis between leather shoes materials and Fourier transform infrared spectroscopy[J]. *Leather Science and Engineering*, 2019, 29(5): 50-54.
王丹, 姜红. 鞋底材料种类与 FTIR 关联分析[J]. *皮革科学与工程*, 2019, 29(5): 50-54.
- [4] Hu H P, Gan C F. Analysis on the identification of sole materials[J]. *Textile Testing and Standard*, 2017, 3(5): 11-13.
胡浩平, 甘长凤. 鞋底材质鉴别方法分析[J]. *纺织检测与标准*, 2017, 3(5): 11-13.
- [5] Wu Q H, Zhang Z Y. Analysis of sole materials by Fourier transform infrared spectroscopy[J]. *Analysis and Testing Technology and Instruments*, 2017, 23(1): 24-28.
吴清华, 张振宇. 鞋底材料的傅立叶变换红外光谱类型分析[J]. *分析测试技术与仪器*, 2017, 23(1): 24-28.
- [6] Jiang H, Fan Y, Wang J G, et al. Study on the test of rubber soles by X-ray fluorescent spectrum[J]. *Infrared and Laser Engineering*, 2017, 46(10): 326-331.
姜红, 范焯, 王嘉庚, 等. X 射线荧光光谱法检验橡胶鞋底的研究[J]. *红外与激光工程*, 2017, 46(10): 326-331.
- [7] Wu R J, Jiang H. A study on discrimination of plastic slipper by SEM/EDS[J]. *Shanghai Plastics*, 2017(1): 18-22.
务瑞杰, 姜红. 扫描电镜/能谱法检验塑料拖鞋[J]. *上海塑料*, 2017(1): 18-22.
- [8] Fang G, Yin L, Liu F, et al. Application research of fluorescence suppression based on differential Raman technique[J]. *Laser Technology*, 2019, 43(3): 359-362.
方刚, 殷磊, 刘峰, 等. 基于差分喇曼技术在抑制荧光中的应用研究[J]. *激光技术*, 2019, 43(3): 359-362.
- [9] Tang Z J, Zhang M M, Lu T J, et al. Application of portable differential Raman spectroscopy in the identification of amber [C]//China International Jewelry Conference. [S.l.:s.n.], 2019: 426-430.
唐左军, 张苗苗, 陆太进, 等. 便携式差分拉曼光谱技术在琥珀鉴定中的应用[C]//中国国际珠宝首饰学术交流会议论文集. [出版地不详: 出版者不详], 2019: 426-430.
- [10] Li Z H, Shen J, Bian R H, et al. Accuracy comparison of the machine learning algorithm used to Raman real sample collection in the front line of public security [J]. *Spectroscopy and Spectral Analysis*, 2019, 39(7): 2171-2175.
李志豪, 沈俊, 边瑞华, 等. 机器学习算法用于公安一线拉曼实际样本采样学习及其准确度比较[J]. *光谱学与光谱分析*, 2019, 39(7): 2171-2175.
- [11] Han Z R, Huang W Q, Wang H X, et al. The identification of the natural rubber with different labels by Raman spectra [J]. *Journal of Guiyang University (Natural Sciences)*, 2008, 3(4): 41-43.
韩志嵘, 黄伟其, 王海旭, 等. 不同标号天然橡胶的拉曼光谱鉴定[J]. *贵阳学院学报(自然科学版)*, 2008, 3(4): 41-43.
- [12] He Y J, Jing X B, Wang F S. Micro-structure determination of polybutadienes by Raman spectroscopy [J]. *Chinese Journal of Analytical Chemistry*, 1988, 16(12): 1065-1069.
何永钜, 景遐斌, 王佛松. 测定聚丁二烯微结构的拉曼光谱方法 [J]. *分析化学*, 1988, 16(12): 1065-1069.
- [13] Dong K, Rao Z F, Yang X Y, et al. Determination of several plastic by Raman spectroscopy[J]. *China Plastics Industry*, 2011, 39(6): 67-70.
董鸱, 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.
- [14] Wang Z G, Wang C H, Sun S Q, et al. Fourier transform Raman spectroscopy inspection of rubber [J]. *Journal of Chinese the People's Public Security University (Natural Science)*, 2002, 8(3): 14-17.
王志国, 汪聪慧, 孙素琴, 等. 橡胶的傅立叶变换拉曼光谱法检验[J]. *公安大学学报(自然科学版)*, 2002, 8(3): 14-17.
- [15] Cao X Y, Jin S Z, Hou B, et al. Pollen detection and

- classification method via Raman spectroscopy [J]. *Laser & Optoelectronics Progress*, 2020, 57(13): 133001.
- 曹馨艺, 金尚忠, 侯彬, 等. 基于拉曼光谱的花粉检测及分类方法 [J]. *激光与光电子学进展*, 2020, 57(13): 133001.
- [16] Song H S, Ma L Z, Wang Y F, et al. Recognition of formaldehyde, methanol based on PCA-BP neural network [J]. *Laser & Optoelectronics Progress*, 2020, 57(7): 071201.
- 宋海声, 麻林召, 王一帆, 等. 基于 PCA-BP 神经网络对甲醛和甲醇的识别研究 [J]. *激光与光电子学进展*, 2020, 57(7): 071201.
- [17] Ma Y J, Ma Y T, Chen J H. Vehicle recognition based on multi-layer features of convolutional neural network and support vector machine [J]. *Laser & Optoelectronics Progress*, 2019, 56(14): 141001.
- 马永杰, 马芸婷, 陈佳辉. 结合卷积神经网络多层特征和支持向量机的车辆识别 [J]. *激光与光电子学进展*, 2019, 56(14): 141001.
- [18] Ji Y P, Yang Y H, Huang S J. Identification of wool and cashmere based on decision tree algorithm [J]. *Journal of Textile Research*, 2013, 34(6): 16-20.
- 季益萍, 杨云辉, 黄少君. 基于决策树算法的羊绒与羊毛纤维鉴别 [J]. *纺织学报*, 2013, 34(6): 16-20.
- [19] Zhang M M, Liu Y A, Song P. Applications of partial connection clustering algorithm and random forest algorithm in radar signal sorting [J]. *Laser & Optoelectronics Progress*, 2019, 56(6): 062804.
- 张萌萌, 刘以安, 宋萍. 偏联系数聚类和随机森林算法在雷达信号分选中的应用 [J]. *激光与光电子学进展*, 2019, 56(6): 062804.
- [20] Liu J G, Li H, Sun D F, et al. Land use/cover decision tree classification fusing multi-temporal and multi-spectral of MODIS [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2010, 26(10): 312-318, 389.
- 刘建光, 李红, 孙丹峰, 等. MODIS 土地利用/覆被多时相多光谱决策树分类 [J]. *农业工程学报*, 2010, 26(10): 312-318, 389.
- [21] Chen C Z, Zhang H. Detection of bolt tightness abnormality based on SMOTE and CART decision tree theory [J]. *Mechanical Engineer*, 2019(11): 1-4.
- 陈长征, 张赫. 基于 SMOTE 和 CART 决策树的螺栓打紧质量异常检测 [J]. *机械工程师*, 2019(11): 1-4.
- [22] Xie L H, Shang T. SPSS statistical analysis and data mining [M]. Beijing: Publishing House of Electronics Industry, 2012: 232-235.
- 谢龙汉, 尚涛. SPSS 统计分析与数据挖掘 [M]. 北京: 电子工业出版社, 2012: 232-235.