激光与光电子学进展
先进成像

基于拉曼光谱和机器学习的一次性口罩分类识别

刘金坤[1]，李春宇[1*]，吕航[1]，孔维刚[2]，孙威[1]，张格菲[1]

[1]中国人民公安大学侦查学院，北京 100038；

[2]郑州市公安局刑事科学技术研究所，河南 郑州 450000

摘要 一次性口罩的分类识别在法庭科学物证鉴别中具有重要意义，因此，提出了一种基于拉曼光谱和机器学习的一次性口罩分类识别方法。首先，采集来自不同城市、不同厂家生产的 37 种一次性口罩样品的拉曼光谱数据。然后，利用 Savitzky-Golay 平滑和归一化算法对数据进行预处理，通过主成分分析法和拉曼光谱特征峰对照方法划分口罩类别。最后，构建了基于支持向量机(SVM)、贝叶斯判别分析和前反馈(BP)神经网络的一次性口罩分类识别模型。实验结果表明，SVM 模型训练集和测试集的准确率分别为 93.3%和 100.0%，贝叶斯判别分析模型的训练集和测试集准确率均为 100.0%，BP 神经网络模型的训练集、验证集、测试集准确率分别为 93.9%、60.0%、60.0%。因此，将贝叶斯判别分析模型作为口罩分类识别的最佳模型，为法庭科学技术的物证检验提供了一定的借鉴意义。

关键词 光谱学；拉曼光谱；机器学习；数据处理；分类识别

中图分类号 O433 **文献标志码** A **doi**：10.3788/LOP202158.1630004

Classification and Recognition of Disposable Masks Based on Raman Spectroscopy and Machine Learning

Liu Jinkun[1], Li Chunyu[1*], Lü Hang[1], Kong Weigang[2], Sun Wei[1], Zhang Gefei[1]

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

[2] *Institute of Criminal Science and Technology, Zhengzhou Public Security Bureau, Zhengzhou, Henan* 450000, *China*

Abstract In forensic scientific evidence identification, the classification and recognition of disposable masks is critical. Therefore, in this study, a classification and recognition method for disposable masks is proposed using Raman spectroscopy and machine learning. First, the Raman spectra of 37 types of disposable masks are gathered from different cities and manufacturers. Then, the data are processed using Savitzky Golay smoothing and normalization algorithms, and the mask categories are classified using principal component analysis method and the Raman spectrum characteristic peak comparison method. Finally, a disposable mask classification and recognition model is constructed based on support vector machine (SVM), Bayes discriminant analysis, and back propagation (BP) neural network algorithm. The experimental results show that the accuracy rates of the training and tests set of the SVM model are 93.3% and 100.0%, respectively, whereas the those of the training and test sets of the Bayesian discriminant analysis model are both 100.0%. Furthermore, the accuracy rates of the training set, validation set, and test set of the BP neural network model are 93.9%, 60.0%, and 60.0%, respectively. Therefore, as the best model for mask classification and recognition, the Bayes discriminant analysis model provides a certain reference for forensic science and technology.

Key words spectroscopy; Raman spectroscopy; machine learning; data processing; classification and recognition

OCIS codes 300.6170; 300.6340; 300.6450

收稿日期：2020-08-24；**修回日期**：2020-09-12；**录用日期**：2020-09-22

基金项目：国家重点研发计划(2017YFC0822004，2019YFF0303405)、公安部技术研究计划(2019JSYJC21)、中央高校基本科研业务费(2019JKF427)

通信作者：*lichunyu@ppsuc.edu.cn

1 引 言

一次性防护型口罩经常作为重要的物证出现在蓄谋盗窃、杀人、爆炸等犯罪现场，特别是在新冠肺炎疫情暴发期间，口罩使用量的大幅度提升，增加了口罩在犯罪现场出现的可能性，口罩物证研究的重要性也日益凸显。法庭科学中，侦查人员可以对故意佩戴口罩遮挡脸部实施作案行为的犯罪嫌疑人进行视频追踪，也可以对遗留在犯罪现场的口罩物证进行人体脱落细胞、携带灰尘、材质等多方面检验。其中，口罩材质的分类识别可以确定嫌疑人购买口罩的品牌或生产厂家，从而追根溯源，确定犯罪嫌疑人可能出现的位置、缩小侦查范围、提高办案效率。一次性防护型口罩通常有三层，内外两层为非织造布(化学合成纤维)，中间一层为熔喷布，口罩带为绒毛橡筋，鼻梁条为可塑性材料。《日常防护型口罩技术规范 GBT32610-2016》从口罩的耐摩擦色牢度、甲醛含量、pH 值、可分解致癌芳香胺染料、环氧乙烷残留量、吸气阻力等方面提出质量检测标准。《丙纶纺粘/熔喷/纺粘复合无纺布标准》从单位面积质量、断裂强力和断裂伸长率、抗渗水性、透气性方面提出了口罩质量检测方案。但这些方法的实施需要一定的时间周期，会耗费大量的人力、财力，对法庭科学领域中口罩物证的来源确定作用较小，因此，需要探索快速高效的口罩材质分类识别方法，以解决公安机关在实际中面临的问题。

非织造布作为一次性防护型口罩的主要原料，其化学纤维材质主要由丙纶、涤纶、腈纶和氨纶组成。目前，人们已研究了多种纤维材质的鉴定与识别方法，FZ/T01057《纺织纤维鉴别实验方法》将化学溶解、显微镜观察、燃烧等方法作为鉴定纤维的标准，但该标准会损耗检材，且使用的化学物质会产生污染，不符合环境友好型社会发展的理念。此外，红外光谱法是纤维检验的常用方法，张海煊等[1]用红外光谱仪测得锦纶、氨纶的光谱数据，通过建立偏最小二乘法模型实现纤维的快速定量分析。黎海洋等[2]用傅里叶变换衰减全反射红外光谱技术对聚烯烃弹性纤维、聚酯类弹性纤维、二烯类弹性纤维和氨纶进行检验，通过分析其红外特征峰对纤维进行分类。近年来，特种光谱成像技术也逐渐被应用于纤维物证的检验方面，魏子涵等[3]用傅里叶变换红外光谱仪(FTIR)测量出不同种类织物纤维的标准谱图，实现了不同纤维的快速无损鉴别。金肖克等[4]用高光谱成像系统鉴别不同种类的纺织品，证明了高光谱纤维检验的可行性。

科技的进步需要实验仪器、学科领域、数据分析方法等多方面不断创新，拉曼光谱法作为重要的物质分子结构分析方法可与红外光谱法相结合应用在化学新材料鉴别、宝石鉴别、气体鉴别、微量物证鉴别等多个领域[5]，具有无损检验、获取图谱速度快、数据稳定性好等优势，成为化学纤维检验的最佳方法之一。同时，将机器学习建模方法应用在纤维光谱数据分类识别上，可增加实验结果预测的准确性和方便快捷性。因此，本文收集了 37 种不同品牌和种类的一次性防护型口罩，并采集这些口罩的拉曼光谱数据。利用主成分分析(PCA)法对数据进行降维，并建立了支持向量机(SVM)、前反馈(BP)神经网络、贝叶斯判别分析模型，以找到最佳模型，达到对未知类别口罩精确分类和识别的目的。

2 实验原理

实验中使用的一次性防护型口罩中的部分样品参数如表 1 所示。

表 1 一次性防护型口罩的样品参数

Table 1 Sample parameters of the disposable protective masks

Label	Brand	Manufacturer
1	Xuancheng mask	Rizhao Nuohuan Protective Equipment Co., Ltd
2	Huian mask	Zhangjiagang Meibaiqi Trade Co., Ltd
3	ShengWang mask	Xiantao Siqi Protective Equipment Co., Ltd
4	JiaBoNeng mask	Guangdong Dongwan Yian labor protection products Co., Ltd
…	…	…
33	Huian mask	Suzhou Lotte Protective Equipment Co., Ltd
34	ZHICHANG mask	Nanchang of Jiangxi Province
35	Sanbang mask	Foshan Nanhai Weijian Sanbang Protective Equipment Technology Co., Ltd
36	Taibang mask	Yunnan Baiyao Group Co., Ltd
37	3M mask	3M China Ltd

2.1　实验仪器

显微共聚焦拉曼光谱仪（inVia Raman Microscope）由英国雷尼绍（Renishaw）公司生产，配备了波长为 532，633，785 nm 的激光器，倍率为 5×、20×、50×、100×的显微镜，光谱扫描范围为 100～2000 nm，光谱分辨率为 1 cm^{-1}，最低波数为 10 cm^{-1}。

2.2　光谱采集

将口罩外层的非织造布裁剪成尺寸为 0.5 cm×0.5 cm 的小方块，用镊子夹取放置在载玻片上，用双面胶带固定，并用酒精擦拭样品表面，去除污渍、灰尘等，避免杂物干扰。随机选取 20# 样品进行激光器优选实验，激光器的最大功率为 10 mW，曝光时间为 10 s。调整显微镜的倍率，得到不同放大倍率下单根清晰的口罩纤维如图 1 所示，对应的光谱如图 2 所示。可以发现，拉曼位移 100～2000 cm^{-1} 范围内 532 nm 激光器获得的光谱噪声较大，光谱峰值不明显，没有特征峰；633 nm 激光器获得的光谱信号较弱，且在拉曼位移 100～800 cm^{-1} 范围内没有信号出现。785 nm 激光器获得的光谱特征峰峰值明显，强度大且噪声小，因此，实验使用波长为 785 nm 的激光器。

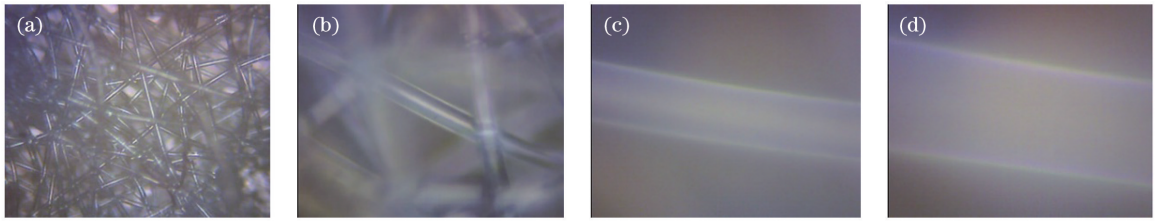

图 1　不同倍率下的口罩外观形态。(a) 5×；(b) 20×；(c) 50×；(d) 100×

Fig. 1　Appearances of masks under different magnifications. (a) 5×; (b) 20×; (c) 50×; (d) 100×

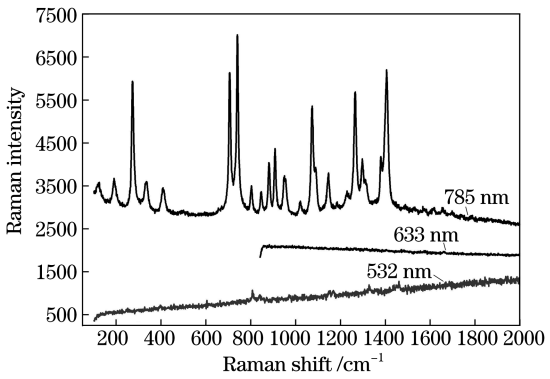

图 2　不同波长激光器得到的口罩光谱

Fig. 2　Mask spectra obtained by different wavelength lasers

为验证拉曼光谱仪的稳定性和光谱重现性，随机选取 10# 样品测量 10 次光谱，结果如图 3(a)所示。可以发现，10 次实验光谱特征峰峰位全部重合，但产生了较小的噪声差异。为验证不同采样点对实验结果的影响，随机选取 18# 样品，在 3 个不同采样点上验证口罩材料的均匀性，结果如图 3(b)所示。可以发现，不同采样点间除噪点、相对拉曼强度稍有差别外，特征峰峰位均相同。这表明在 785 nm 波长激光下测得的拉曼光谱稳定性好，样品材质均匀，在不同位点上的实验结果无明显差异，符合样品实验条件。测量其他样品时，为保证光谱数据的准确性，每个样品测量三次取平均值，对所有样品采集后得到 37 组拉曼图谱。

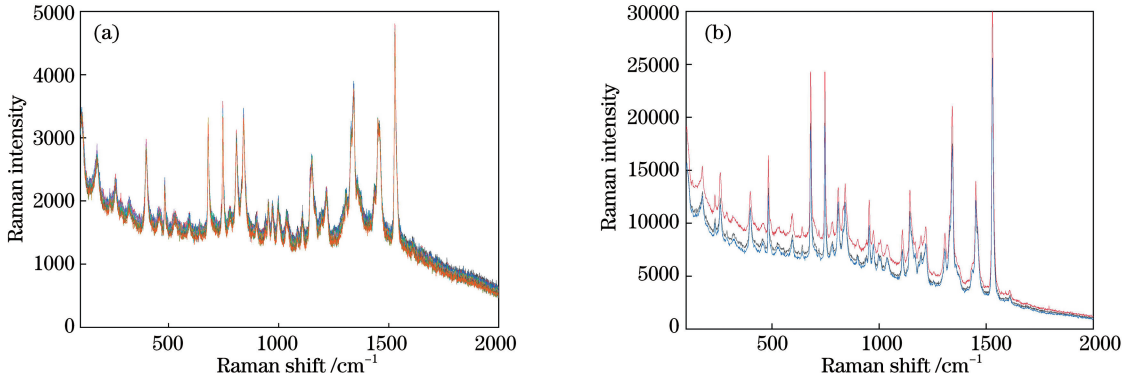

图 3　口罩样品重现性和均匀性的实验结果。(a)重现性实验；(b)均匀性实验

Fig. 3　Experimental results of reproducibility and uniformity of mask samples. (a) Reproducibility experiment; (b) uniformity experiment

2.3 光谱预处理

实验过程中，除光谱特征峰外的噪声主要受仪器性能稳定性、宇宙射线、外界光照及环境温度等因素的影响。光谱预处理可消除这些噪声，常见的光谱预处理方法包括平滑、多元散射矫正（MSC）、标准正态变量变换（SNV）、导数方法。其中，平滑方法可去除光谱白噪声、提高信噪比，解决荧光强度不均的问题，可分为窗口平滑和 Savitzky-Golay（S-G）平滑。S-G 平滑是窗口平滑的改进，因此，实验对口罩光谱数据进行 S-G 平滑处理，可表示为

$$\begin{cases} x_{k,\mathrm{sm}} = \dfrac{1}{H}\sum_{i=-m}^{m} x_{k+i}h_i \\ H = \sum_{i=-m}^{m} h_i \end{cases}, \tag{1}$$

式中，x_k 为输入口罩的光谱数据，$x_{k,\mathrm{sm}}$ 为输出数据，k 为光谱的序号，m 为滤波带宽，h_i 为平滑系数。实验发现，不同样品在相同激光器功率下得到的拉曼光谱相对强度不同，且数量级之间相差较大。为了实现更好的机器学习效果，将拉曼光谱的相对强度归一化在[0，1]之间，可表示为

$$y = \frac{(y_{\max} - y_{\min})(x - x_{\min})}{x_{\max} - x_{\min}}, \tag{2}$$

式中，$x_{\min}$ 和 $x_{\max}$ 分别为原始口罩光谱数据的最小值和最大值，y 为映射范围的参数。当 $y_{\min}=0$，$y_{\max}=1$ 时，可实现拉曼光谱强度在[0，1]区间的归一化。设置滤波带宽 $m=5$，得到预处理后的光谱如图 4 所示。可以发现，在设定拉曼位移下每组光谱数据的特征峰明显，光谱稳定且清晰，可进行下一步数据分析。

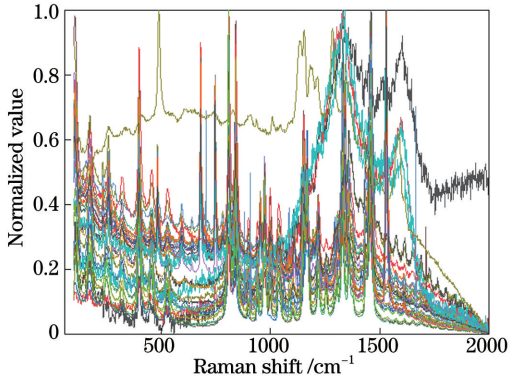

图 4　37 种口罩样品拉曼光谱的处理结果

Fig. 4　Raman spectrum processing results of 37 mask samples

2.4 模型原理

光谱数据的维度较高，实验中的拉曼光谱扫描范围为 100～2000 cm^{-1}，会产生 1900 维数据。高维数据中的噪点、奇异点等信息会影响模型的运行时间和预测精度，因此，需要在降低数据维度的同时保证不丢失重要信息并提取光谱的特征信息。在多元统计分析中，常见的特征提取方法有小波变换法和 PCA 法。文献[6-7]的研究结果表明，PCA 法对光谱数据的降维效果较好，具体步骤如下。

1）计算口罩的光谱数据标准化矩阵 $\boldsymbol{X}$

$$\boldsymbol{X} = \begin{bmatrix} x_{11} & \cdots & x_{1q} \\ \vdots & & \vdots \\ x_{p1} & \cdots & x_{pq} \end{bmatrix}, \tag{3}$$

式中，x_{pq} 为第 p 条口罩样品数据的第 q 维光谱数据。

2）计算相关系数矩阵 $\boldsymbol{P}$

$$\boldsymbol{P} = \begin{bmatrix} r_{11} & \cdots & r_{1q} \\ \vdots & & \vdots \\ r_{p1} & \cdots & r_{pq} \end{bmatrix}, \tag{4}$$

式中，r_{pq} 为任意两组光谱数据的相关系数，$r_{ij}=r_{ji}$，r_{ij} 可表示为

$$r_{ij} = \sum_{k=1}^{n} \frac{(x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)}{\sqrt{\sum_{k=1}^{n} x_{kj} - x_i \sum_{k=1}^{n} (x_{kj} - \bar{x}_j)}}, \tag{5}$$

式中，$\bar{x}_i$、$\bar{x}_j$ 为第 i 条和第 j 条光谱数据的平均值，n 为光谱数据的数量。

3）根据 $|\lambda\boldsymbol{\beta} - \boldsymbol{P}| = 0$ 求出矩阵 $\boldsymbol{P}$ 的特征值 λ_i 和特征向量 $\boldsymbol{\beta}_i$ 并将特征值从大到小排序，最大特征值及其对应的特征向量就是第一主成分的方差和。定义每个特征值在总方差中所占的比例为主成分贡献率 C_i，前 i 个贡献率的累计贡献率为 A_i，可表示为

$$C_i = \frac{\lambda_i}{\sum_{k=1}^{n}\lambda_k}, i = 1, \cdots, n, A_i = \frac{\sum_{k=1}^{i}\lambda_k}{\sum_{k=1}^{n}\lambda_k}, i = 1, \cdots, n。 \tag{6}$$

4）特征向量组成的矩阵 $\boldsymbol{M} = [\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_i]$，则主成分矩阵 $\boldsymbol{Y}$ 可表示为

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{M}。 \tag{7}$$

SVM 模型是一种降低数据结构风险的分类模型，需寻找一个超平面，用距离超平面最近的点组成支持向量，然后通过不断优化支持向量与超平面间的距离实现不同样品的分类。影响模型的主要参数为误差惩罚参数 C，可控制错误分类的样品数，权衡错分样品比例和算法的复杂度。针对非线性问题，SVM 模型还需要选择不同种类的核函数将数据投影到合适的特征空间，核函数可分为多项式核函数、高斯核函数和径向基核函数。其中，径向基核函数

的使用最广泛，其参数 γ 可判定特征空间中向量间的距离，对分类准确率的影响较大[8-11]。

贝叶斯判别分析模型是依据贝叶斯定理和特征条件假设对样品进行分类的模型，可降低错误分类的概率。该模型需要建立关于数据集和对应标签的联合概率分布（先验概率分布），基于先验概率分布得到后验概率分布，当有新的样品数据时，再利用后验概率分布判断其所属类别[11-13]。

BP 神经网络是将每一个样品数据作为输入神经元，用隐藏层数据向量与模型自适应权重矩阵的内积加上可降低误差的偏置项，通过激活函数输出分类概率，进而根据概率确定口罩所属的类别。如果输入未知测试样品数据，训练好的模型就能识别出样品对应的标签。由于权重矩阵是随机生成的，预测结果不准确，会产生损失，因此，需要先计算正向传播的代价函数，然后通过反向传播的梯度下降算法优化权重，降低模型损失，达到准确分类的目的[13-17]。影响模型分类结果的参数有学习率（梯度下降的步长）、隐藏层神经元的个数、激活函数、损失函数及模型迭代次数。

3　实验结果与讨论

3.1　主成分分析法

首先，通过 PCA 法对光谱数据进行降维，并结合拉曼光谱特征峰确定口罩样品的类别，从而将无监督问题转化为有监督问题。然后，搭建多种机器学习模型[17]，通过调参、改变实验条件、多次迭代等方法训练模型并对测试集进行预测。最后，观察模型分类识别的准确率和运行时间，优选出最佳口罩的分类识别方案，具体流程如图 5 所示。

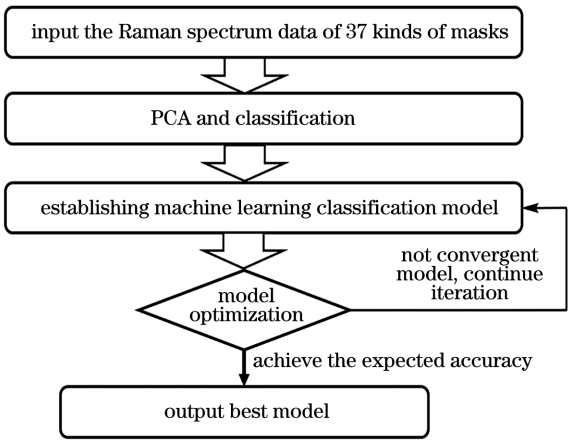

图 5　基于拉曼光谱的口罩分类模型流程图

Fig. 5　Flow chart of the mask classification model based on Raman spectroscopy

表 2 为 PCA 特征方差的贡献率，可以发现，当提取前 31 个主成分（PC）时，累计方差的百分比就能达到 100%。其中，成分 1 的贡献率为 49.02%，成分 2 的贡献率为 36.13%，对于总体数据的可解释性较强。因此，提取前 2 个主成分代表的函数建立二维可视化图，结果如图 6 所示。可以发现，37 个口罩样品大致可分为 6 类。

表 2　PCA 特征方差的贡献率

Table 2　Contribution rate of the PCA characteristic variance

PC	Eigenvalue	Variance /%	Cumulative variance /%
1	18.80335	49.02	49.02
2	13.85901	36.13	85.15
3	2.36197	6.16	91.31
4	1.76937	4.61	95.92
5	1.06630	2.78	98.70
6	0.13423	0.35	99.05
7	0.09534	0.25	99.30
8	0.06265	0.16	99.46
9	0.03321	0.09	99.55
10	0.02942	0.08	99.63
11	0.02400	0.06	99.69
12	0.02108	0.05	99.74
13	0.01752	0.05	99.79
14	0.01311	0.03	99.82
15	0.01148	0.03	99.85
16	0.01001	0.03	99.88
17	0.00800	0.02	99.90
18	0.00729	0.02	99.92
19	0.00525	0.01	99.93
20	0.00471	0.01	99.95
21	0.00387	0.01	99.96
22	0.00341	0.01	99.97
23	0.00250	0.01	99.97
24	0.00202	0.01	99.98
25	0.00168	0.00	99.98
26	0.00142	0.00	99.99
27	0.00116	0.00	99.99
28	0.00090	0.00	99.99
29	0.00084	0.00	99.99
30	0.00063	0.00	99.99
31	0.00060	0.00	100.00

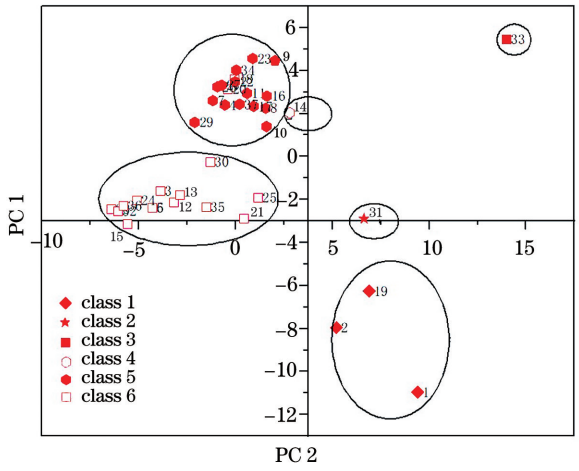

图 6　PCA 的可视化

Fig. 6　Visualization of the PCA

为验证分类结果的可靠性，对比分析了不同口罩拉曼光谱的特征峰，结果如图 7 所示。可以发现，图 7(a)中 14# 样品在 1329，1459 cm^{-1} 附近的特征峰与 8#、10#、16# 样品不同，可将其单独归为 1 类；图 7(b)中 33# 样品在 489，1153，1284，1586 cm^{-1} 附近的特征峰与其他样品不同，31# 样品在 679，746，1525 cm^{-1} 附近的特征峰与其他样品不同，而 1#、2#、19# 样品的特征峰相同，可将 33#，31# 和 1#、2#、19# 样品划分为不同类；图 7(c)中 3#、12#、13#、35# 样品在 397，808，1328，1459 cm^{-1} 附近的特征峰相同，可归为 1 类；图 7(d)中 29# 样品在 680，746，1529 cm^{-1} 附近的特征峰与 21#、25#、30# 样品明显不同，可将 3#、12#、13#、35#，29# 和 21#、25#、30# 样品分别归为 1 类，分类结果与图 6 中的结论一致。

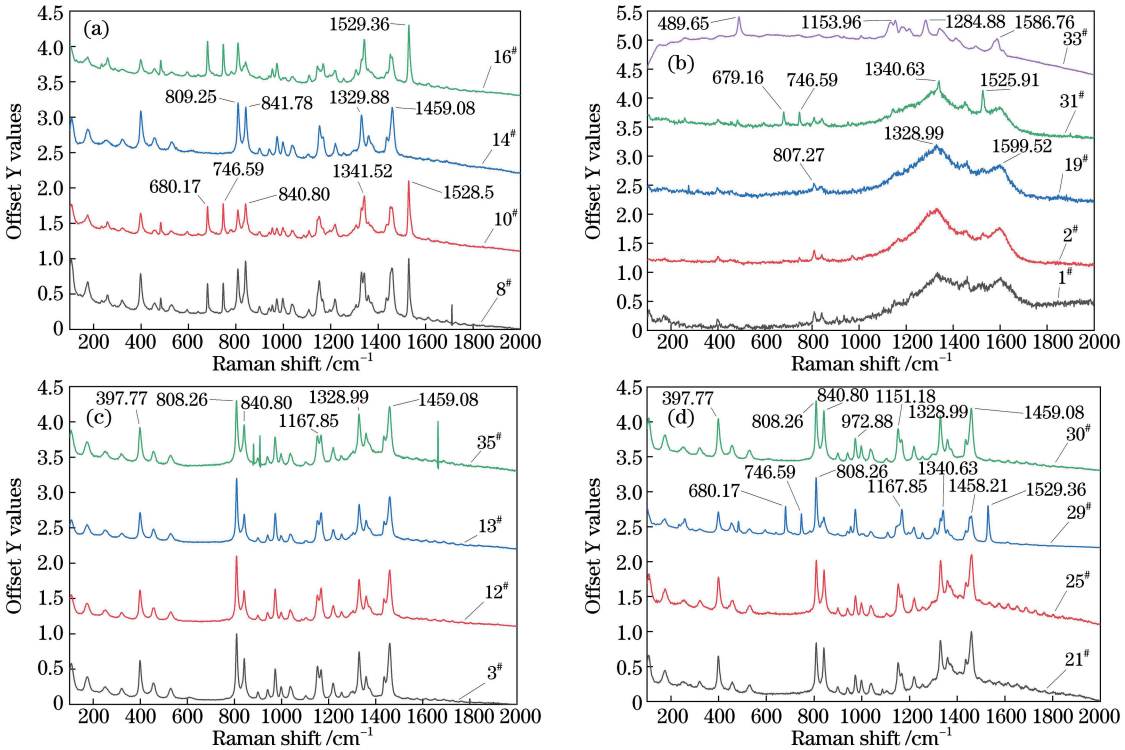

图 7　口罩的拉曼光谱特征峰。(a) 8#、10#、14#、16#；(b) 1#、2#、19#、31#、33#；(c) 3#、12#、13#、35#；(d) 21#、25#、29#、30#

Fig. 7　Characteristic peaks of the Raman spectrum of the mask. (a) 8#, 10#, 14#, 16#; (b) 1#, 2#, 19#, 31#, 33#; (c) 3#, 12#, 13#, 35#; (d) 21#, 25#, 29#, 30#

3.2　机器学习方法的建模

为实现口罩样品的自动分类识别，在输入未知光谱数据时就能判别其种类，将 37 种口罩样品对应唯一类别标签并建立机器学习分类模型，样品对应的标签如表 3 所示。

对于 SVM 模型，实验采用网格寻优法和交叉验证法寻找的最佳参数 $C=1.5157$、$\gamma=0.0078$。将 37 个口罩样品随机划分为两部分，用 30 个样品作为训练模型，用剩余 7 个样品预测分类标签，模型的训练和测试结果如图 8 所示。可以发现，训练集中的 30 个样品有 2 个预测错误，训练准确率为 93.3%；测试集中的 7 个样品全部预测正确，测试准

表 3 基于 PCA 和拉曼光谱的口罩分类结果

Table 3 Mask classification results based on PCA and Raman spectra

Class	Sample number
1	1#、2#、19#
2	31#
3	33#
4	14#
5	4#、7#、8#、9#、10#、11#、16#、17#、18#、20#、22#、23#、26#、27#、29#、34#、37#
6	3#、5#、6#、12#、13#、15#、21#、24#、25#、28#、30#、32#、35#、36#

确率为 100.0%，所用时间为 20 s。表 4 为选用线性核函数、多项式核函数、径向基核函数(RBF)和 Sigmoid 核函数对 SVM 模型预测准确率的影响，可以发现，不同核函数的运行时间和预测准确率均不同。其中，选用 RBF 核函数的运行时间较长，为 20 s，但其训练集和测试集的准确率均高于其他核函数，因此，实验在 SVM 模型中选用 RBF 核函数。

表 4 不同核函数对 SVM 模型的影响

Table 4 Influence of different kernel functions on the SVM model

Kernel function type	Training set accuracy /%	Test set accuracy /%	Run time /s
Linear	50.0	28.6	5
Polynomial	46.5	42.9	10
RBF	93.3	100.0	20
Sigmoid	96.7	85.0	15

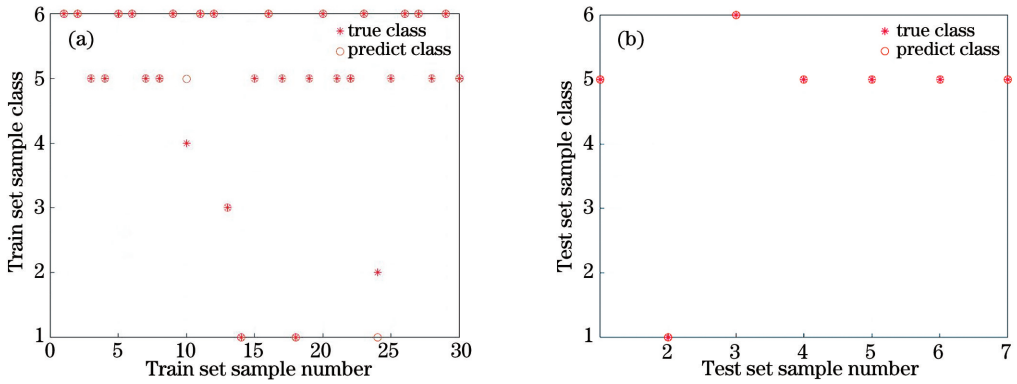

图 8 SVM 模型的预测结果。(a)训练集；(b)测试集

Fig. 8 Prediction results of the SVM model. (a) Train set; (b) test set

对于贝叶斯判别分析模型，实验用 30 个口罩样品作为先验知识并得到判别函数，然后用判别函数验证剩余 7 个口罩所属的类别，结果如表 5 所示。可以发现，贝叶斯判别函数 1 的特征值为 15064.487，方差为 66.2%；函数 2 的特征值为 62222.540，方差为 27.4%，两者的累计方差为 93.6%。对 5 个判别函数进行显著性检验后发现，函数 1、2 的显著性指标(Sig)远小于 0.05，具有统计学意义。

表 5 口罩样品的贝叶斯判别函数

Table 5 Bayesian discriminant function of mask samples

Function	Eigenvalue	Variance /%	Cumulative variance /%	Canonical correlation	Function test	Wilks'lambda	Sig
1	15064.487	66.2	66.2	1	1–5	0.00	0.00
2	6222.540	27.4	93.6	1	2–5	0.00	0.00
3	1303.951	5.7	99.3	1	3–5	0.00	0.00
4	126.670	0.6	99.9	0.996	4–5	0.00	0.00
5	24.217	0.1	100	0.980	5	0.04	0.00

将前 2 个判别函数作为未知样品类别的预测函数，可表示为

$$W_1(x)=27.712x_1-19.636x_2+\cdots-0.461x_{31},\tag{8}$$

$$W_2(x)=-13.792x_1-7.105x_2+\cdots+0.212x_{31},\tag{9}$$

式中，x_i，$i=1,\cdots,31$ 为第 x 个口罩样品的 i 维指标。用建立的判别函数对口罩样品进行训练

和分类预测，结果如图 9 所示。可以发现，30 个训练样品的分类准确率为 100.0%，其余 7 个测试样品的分类准确率为 100.0%，模型运行时间为 10 s。

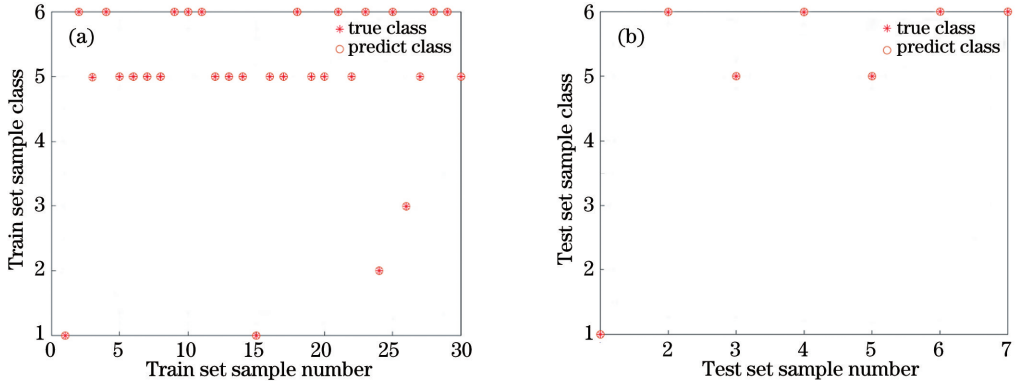

图 9　贝叶斯判别分析模型的预测结果。(a)训练集；(b)测试集

Fig. 9　Prediction results of the Bayesian discriminant analysis model. (a) Train set; (b) test set

对于 BP 神经网络模型，建模时随机选择 20 个样品数据作为训练集，10 个样品数据作为验证集，7 个样品数据作为测试集。设置学习率为 0.001，隐藏层神经元为 8 个，激活函数为 Sigmoid 函数，交叉熵为损失函数，模型迭代 15 次后收敛，得到模型的交叉熵损失函数如图 10 所示。可以发现，该模型验证集和测试集的交叉熵趋于平缓，训练集的交叉熵呈下降趋势，这表明模型在测试集上趋于稳定，未出现过拟合现象。实验得到的最优测试集交叉熵为 0.060186，由混淆矩阵可知，模型训练集的准确率为 93.9%，验证集的准确率为 60.0%，测试集的准确率为 60.0%，运行时间为 5 s。

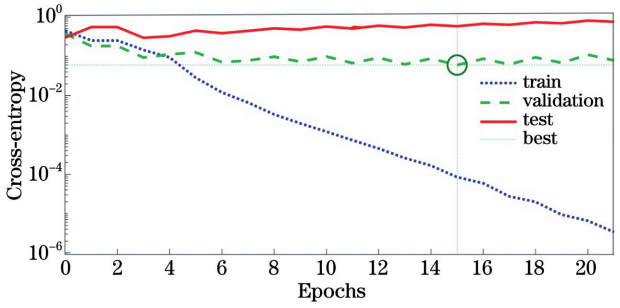

图 10　迭代次数与交叉熵的关系

Fig. 10　Relationship between iteration times and cross entropy

综上所述，贝叶斯判别模型的训练和测试准确率均达到了 100.0%，可作为口罩分类识别的最佳模型；SVM 模型的训练准确率为 93.3%，测试准确率达到 100.0%，存在一定偏差容忍度的情况下可以选择；BP 神经网络的训练、验证和测试准确率都比较低，原因是神经网络适合多维大数据的分析，且实验中的口罩样品数据少，不能满足其权重更新优化的需要，而贝叶斯判别模型和 SVM 模型适合小数据集的分类识别，如果数据量过大，训练速度会很慢，但更适合本课题的研究要求。

4　结　　论

以口罩样品的拉曼光谱为基础，提出了一种基于特征提取和多模型结合优选的口罩种类识别方法。通过对比不同模型的准确率和运行时间，最终确定了基于贝叶斯判别分析的分类模型，模型训练集和测试集的分类准确率均能达到 100.0%且运行时间为 10 s，可满足快速、准确的预测需求。该方案可为从犯罪现场提取的口罩物证种类识别提供借鉴，丰富法庭科学理化检验方法。但实验采集的口罩样品尚处于研究阶段，若想实现任意口罩种类识别，还需增加样品种类，构建更加多样的样品数据集。因此，接下来还将进一步丰富实验样品种类和数量，使分类模型更具实际意义。

参　考　文　献

[1]　Zhang H X, Cao H H, Zhang X, et al. Near infrared quantitative analysis of fiber contents of polyamide/spandex knitted underwear fabric[J]. Knitting Industries, 2020(6): 82-85.
张海煊, 曹海辉, 张续, 等. 锦氨内衣纤维含量近红外光谱法快速定量分析[J]. 针织工业, 2020(6): 82-85.

[2]　Li H Y, Liu W, Cheng X Q, et al. Identification of elastic fibers by infrared spectroscopy, Raman spectroscopy, and pyrolysis gas chromatography-mass spectrometry[J]. Textile Testing and Standard, 2020, 6(1): 13-16.
黎海洋, 刘旺, 程鑫桥, 等. 红外、拉曼光谱和裂解气-质联用技术鉴别弹性纤维[J]. 纺织检测与标准,

2020, 6(1): 13-16.

[3] Wei Z H, Li W X, Du Y J, et al. Establishment and application of fabrics attenuated total reflection Fourier transform infrared spectroscopy spectrum library[J]. Journal of Textile Research, 2019, 40(8): 64-68.
魏子涵, 李文霞, 杜宇君, 等. 织物傅里叶变换衰减全反射红外光谱库的建立及应用[J]. 纺织学报, 2019, 40(8): 64-68.

[4] Jin X K, Tian W, Zhu W J, et al. Qualitative identification of textile chemical composition based on hyperspectral imaging system[J]. Journal of Textile Research, 2018, 39(10): 50-57.
金肖克, 田伟, 朱炜婧, 等. 基于高光谱成像系统的纺织品成分定性鉴别[J]. 纺织学报, 2018, 39(10): 50-57.

[5] Hu W, Ye S, Zhang Y J, et al. Machine learning protocol for surface-enhanced Raman spectroscopy[J]. The Journal of Physical Chemistry Letters, 2019, 10(20): 6026-6031.

[6] Li G W, Gao X H, Xiao N W, et al. Estimation of soil organic matter content based on characteristic variable selection and regression methods[J]. Acta Optica Sinica, 2019, 39(9): 0930002.
李冠稳, 高小红, 肖能文, 等. 特征变量选择和回归方法相结合的土壤有机质含量估算[J]. 光学学报, 2019, 39(9): 0930002.

[7] Xu H D, Lin L L, Li Z, et al. Nephrite origin identification based on Raman spectroscopy and pattern recognition algorithms[J]. Acta Optica Sinica, 2019, 39(3): 0330001.
徐荟迪, 林露璐, 李征, 等. 基于拉曼光谱和模式识别算法的软玉产地鉴别[J]. 光学学报, 2019, 39(3): 0330001.

[8] Li C Y, Liu J K, Jiang H, et al. Identification of X-ray fluorescent spectral paper ashes based on support vector machine algorithm[J]. Laser & Optoelectronics Progress, 2021, 58(3): 0330006.
李春宇, 刘金坤, 姜红, 等. 基于支持向量机算法的X射线荧光光谱纸张灰烬识别研究[J]. 激光与光电子学进展, 2021, 58(3): 0330006.

[9] Sampaio P S, Castanho A, Almeida A S, et al. Identification of rice flour types with near-infrared spectroscopy associated with PLS-DA and SVM methods[J]. European Food Research and Technology, 2020, 246(3): 527-537.

[10] Hu X, Wu R M, Zhu X Y, et al. Fast detection of chlorpyrifos residues in tea via surface-enhanced Raman spectroscopy combined with two-dimensional correlation spectroscopy[J]. Acta Optica Sinica, 2019, 39(7): 0730001.
胡潇, 吴瑞梅, 朱晓宇, 等. 表面增强拉曼光谱结合二维相关谱快速检测茶叶中的毒死蜱残留[J]. 光学学报, 2019, 39(7): 0730001.

[11] He X L, Wang J F, He Y, et al. Infrared spectroscopy identification of plastic steel windows based on Bayes discrimination analysis[J]. Laser Journal, 2019, 40(11): 33-37.
何欣龙, 王继芬, 何亚, 等. Bayes 判别的塑钢窗红外光谱快速识别[J]. 激光杂志, 2019, 40(11): 33-37.

[12] Wen C P, Bai Y Y, Zeng J J, et al. Bayes discriminant analysis method of natural grassland classification[J]. Chinese Journal of Grassland, 2016, 38(3): 50-55.
文畅平, 白银涌, 曾娟娟, 等. 天然草地分类的Bayes判别分析法[J]. 中国草地学报, 2016, 38(3): 50-55.

[13] Zhou S, Shen C Y, Zhang L, et al. Dual-optimized adaptive Kalman filtering algorithm based on BP neural network and variance compensation for laser absorption spectroscopy[J]. Optics Express, 2019, 27(22): 31874-31888.

[14] Ershat A, Baidengsha M, Mamat S, et al. Combined estimation of chlorophyll content in cotton canopy based on hyperspectral parameters and back propagation neural network[J]. Acta Optica Sinica, 2019, 39(9): 0930003.
依尔夏提·阿不来提, 白灯莎·买买提艾力, 买买提·沙吾提, 等. 基于高光谱和BP神经网络的棉花冠层叶绿素含量联合估算[J]. 光学学报, 2019, 39(9): 0930003.

[15] 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.

[16] Cai Y, Su M X, Cai X S. Method for mixed-particle classification based on convolutional neural network[J]. Acta Optica Sinica, 2019, 39(7): 0712002.
蔡杨, 苏明旭, 蔡小舒. 基于卷积神经网络的混合颗粒分类法研究[J]. 光学学报, 2019, 39(7): 0712002.

[17] Liu J X, Du B, Deng Y Q, et al. Terahertz-spectral identification of organic compounds based on differential PCA-SVM method[J]. Chinese Journal of Lasers, 2019, 46(6): 0614039.
刘俊秀, 杜彬, 邓玉强, 等. 基于差分-主成分分析-支持向量机的有机化合物太赫兹吸收光谱识别方法[J]. 中国激光, 2019, 46(6): 0614039.