

基于表面增强拉曼光谱的多组分物质分类识别

白鹤轩^{1,2}, 杨峰^{1,2}, 李丹阳^{1,2}, 徐溢^{1,2}, 李顺波^{1,2}, 陈李^{1,2*}¹重庆大学光电工程学院光电技术与系统教育部重点实验室, 重庆 400044;²重庆大学新型微纳器件与系统技术重点学科实验室, 重庆 400044

摘要 为提高混合物质组分识别的准确性、降低建模的复杂度,构建了一种先粗分、后细分的递进式表面增强拉曼光谱检测系统。首先,用一种特征峰判别算法对全部样本进行特征提取并建立粗分模型,以分离单组分和多组分物质。然后,联合归一化与主成分分析法完成光谱特征的自动提取,并建立多输出最小二乘支持向量机细分模型。最后,采用粒子群优化算法进行参数寻优,实现多组分样本成分的精确预测。选用罗丹明 6G、耐尔蓝和结晶紫三种探针分子进行实验,结果表明,特征峰判别算法提取样本特征的正确率达到 99.44%,粗分模型对 90 例盲测样本全部识别正确;细分模型对多组分样本识别的相关系数不小于 0.995,均方根误差不大于 2.67343%。该拉曼检测系统能实现样本的定性、定量检测,为药品等复杂物质的检测提供了一种有效的识别途径。

关键词 表面光学; 表面增强拉曼光谱; 特征提取; 粒子群优化; 最小二乘支持向量机

中图分类号 O657.37

文献标志码 A

doi: 10.3788/AOS202141.2024001

Multi-Component Substance Classification and Recognition Based on Surface-Enhanced Raman Spectroscopy

Bai Hexuan^{1,2}, Yang Feng^{1,2}, Li Danyang^{1,2}, Xu Yi^{1,2}, Li Shunbo^{1,2}, Chen Li^{1,2*}¹Key Laboratory of Optoelectronic Technology & System, Ministry of Education, College of Optoelectronic Engineering, Chongqing University, Chongqing 400044, China;²Key Disciplines Laboratory of Novel Micro-Nano Devices and System Technology, Chongqing University, Chongqing 400044, China

Abstract For better accuracy of mixture component identification and less complexity of modeling, a progressive detection system based on surface-enhanced Raman spectroscopy with rough classification followed by subdivision is constructed in this paper. First, a characteristic peak discrimination algorithm is used to extract features from all samples and establish a rough classification model, according to which substances are classified into single-component and multi-component ones. Then, automatic extraction of spectral characteristics is accomplished through normalization and principal component analysis. A subdivision model is developed upon a multi-output least squares support vector machine. Finally, the particle swarm optimization algorithm is employed to optimize the parameters so as to achieve an accurate prediction of the composition of multi-component samples. Experiments are carried out with Rhodamine 6G, Nile blue and crystal violet as probe molecules. The results show that the characteristic peak discrimination algorithm extracts sample features with an accuracy of 99.44%. The rough classification model correctly identifies all 90 samples in the blind test. Moreover, the correlation coefficient of the subdivision model for multi-component sample identification is not less than 0.995 and the root mean square error is not more than 2.67343%. Enabling both qualitative and quantitative detection of samples, the Raman detection system proposed in this paper provides an effective identification method for future detection of complex substances

收稿日期: 2021-04-08; 修回日期: 2021-05-08; 录用日期: 2021-05-10

基金项目: 国家重点研发计划(2018YFB2002302)、国家自然科学基金(61971074, 62071072)、重庆大学中央高校基本科研业务费专项(2019CDYGYB003)

通信作者: *CL2009@cqu.edu.cn

such as drugs.

Key words surface optics; surface-enhanced Raman spectroscopy; feature extraction; particle swarm optimization; least squares support vector machine

OCIS codes 240.6695; 170.5660; 160.4236

1 引言

表面增强拉曼光谱(SERS)技术具有快速无损检测、灵敏度高等优点,在食品安全^[1-4]、医疗卫生^[5-7]、违禁品检查^[8-10]等领域受到了人们的广泛关注。利用 SERS 进行的物质定性、定量检测中^[11-14],建模方法对于模型的识别准确率有着重要的影响。陈阳等^[15]在基于 SERS 的多环芳烃定量检测中,加入内标物硫氰化钾后提取苾、菲混合溶液各特征峰的峰值强度与内标峰的峰值强度之比作为模型的输入特征,预测结果的平均相对误差不超过 17%。胡潇等^[16]利用 SERS 技术检测茶叶中的农药浓度,结合二维相关光谱法建立了支持向量机(SVM)模型,预测结果的相对误差不超过 6.15%。付兴虎等^[17]在基于拉曼光谱的山羊血清蛋白含量快速无损检测中,提取血清拉曼谱的 10 个特征峰强度进行建模,预测结果的准确率为 99.8%。张燕君等^[18]在检测食用脂肪酸含量的研究中,采集了食用油样本的 9 个关键拉曼特征峰,建模预测结果中饱和、单不饱和、多不饱和三种脂肪酸含量的相关度分别为 0.9343、0.9965 和 0.9943。张冰等^[19]选取混合汽油样本的 33 个特征峰强度,对两种汽油中芳烃、烯烃、苯、甲醇和乙醇的体积分数进行了预测,得到了推广能力较强的预测模型。上述研究中,模型虽然均达到了较好的预测效果,但建模中的特征提取步骤多通过人为选择特征峰,并用特征峰的强度作为模型的输入特征。更换待检物后,就需要重选特征峰,对应的软件程序会随之更改,有悖于开闭原则(OCP)^[20]。特征峰提取的数量及位置受主观因素

的影响较大,没有统一标准;且当物质混合后产生谱峰叠加变形、峰位微移及小峰淹没等现象时,人为选择特征峰的难度变大。此外,当混合物的组分较多时,利用单一模型进行预测建模的复杂度较高。

针对上述问题,本文构建了一种先粗分、后细分的递进式拉曼检测系统。首先,提出了一种特征峰判别算法用于提取特征,建立多分类支持向量机(MC-SVM)粗分模型,完成单组分样本和多组分样本的种类区分;然后,用联合的归一化与主成分分析(N-PCA)法进行光谱特征的自动提取,并结合多输出最小二乘支持向量回归(M-LS-SVR)分析建立组分定量预测的细分模型。最后,采用粒子群优化(PSO)算法进行参数优化^[21],并与传统人工特征提取方法进行了对比实验。

2 实验原理

2.1 实验材料

选取罗丹明 6G(R6G)、耐尔蓝(NB)和结晶紫(CV)三种化学分析纯染色剂(编号依次为 A、B 和 C),分别配制成浓度为 5×10^{-5} , 1×10^{-5} , 1×10^{-6} mol/L 的单组分样本。由于三种样本分子的拉曼散射截面不同,且不同分子与银纳米颗粒的亲合力有一定差异,为避免信号强度差距过大,使用了不同的浓度。用三种单组分样本配置不同体积比的 15 种多组分样本,15 种多组分样本与 3 种单组分样本对应的参数如表 1 所示。将样本在遮光处静置 24 h 后,分别取每种样本 1 mL 与 1 mL 银溶胶经超声振荡混合均匀,形成待检样本。银溶胶的制备方法可参考文献^[22]。

表 1 18 种样本的组分

Table 1 Composition of 18 types of samples

Type	R6G	NB	CV	V(A):V(B):V(C)	Type	R6G	NB	CV	V(A):V(B):V(C)
A	✓			10:0:0	AC		✓	✓	0:3:7
B		✓		0:10:0	AC		✓	✓	0:5:5
C			✓	0:0:10	AC		✓	✓	0:7:3
AB	✓	✓		1:9:0	AC		✓	✓	0:9:1
AB	✓	✓		3:7:0	BC	✓		✓	1:0:9
AB	✓	✓		5:5:0	BC	✓		✓	3:0:7
AB	✓	✓		7:3:0	BC	✓		✓	5:0:5
AB	✓	✓		9:1:0	BC	✓		✓	7:0:3
AC		✓	✓	0:1:9	BC	✓		✓	9:0:1

实验使用 Ocean Optics 的 ID-Raman 光谱仪, 激发光源的波长为 785 nm, 扫描范围为 $200 \sim 2000 \text{ cm}^{-1}$, 激光功率为 11.9 mW, 积分时间为 6 s。开机预热时间为 15 min, 预热后用移液枪每次取 $4 \mu\text{L}$ 样本滴加在铝箔纸上进行光谱扫描, 得到 18 种样本的 SERS, 每种样本取样和测试 15 次, 共得到 270 个二维 SRES 数据矩阵。

2.2 光谱预处理

实验采集的 SERS 不可避免地会受到荧光背景、仪器噪声以及实验环境光线变化的干扰, 为了提取光谱中的有效信息, 需要对原始 SRES 进行一系列预处理。首先, 截取谱段 $500 \sim 1700 \text{ cm}^{-1}$, 去除

其余谱段的数据, 原因是三种物质的特征峰大多集中在 $500 \sim 1700 \text{ cm}^{-1}$ 范围内, 图 1(a) 为截取的 15 条 R6G 的 SERS。其次, 对光谱进行平滑去噪, 以提高信噪比, 在确保光谱形态的前提下, 有效滤除随机噪声, 采用移动窗口平均法 (窗口数为 3) 的处理结果如图 1(b) 所示。然后, 采用自适应迭代加权最小二乘法进行基线校正^[23], 图 1(c) 为计算获得的基线位置及校正后的光谱图, 可以发现, 基线校正后大幅度减少了荧光背景的影响。最后, 对光谱进行归一化处理, 选取谱段内强度最高的特征峰作为最大值值进行归一化, 将样本的拉曼强度映射到 $[0, 1]$ 内, 从而提高模型的训练效率^[24], 处理结果如图 1(d) 所示。

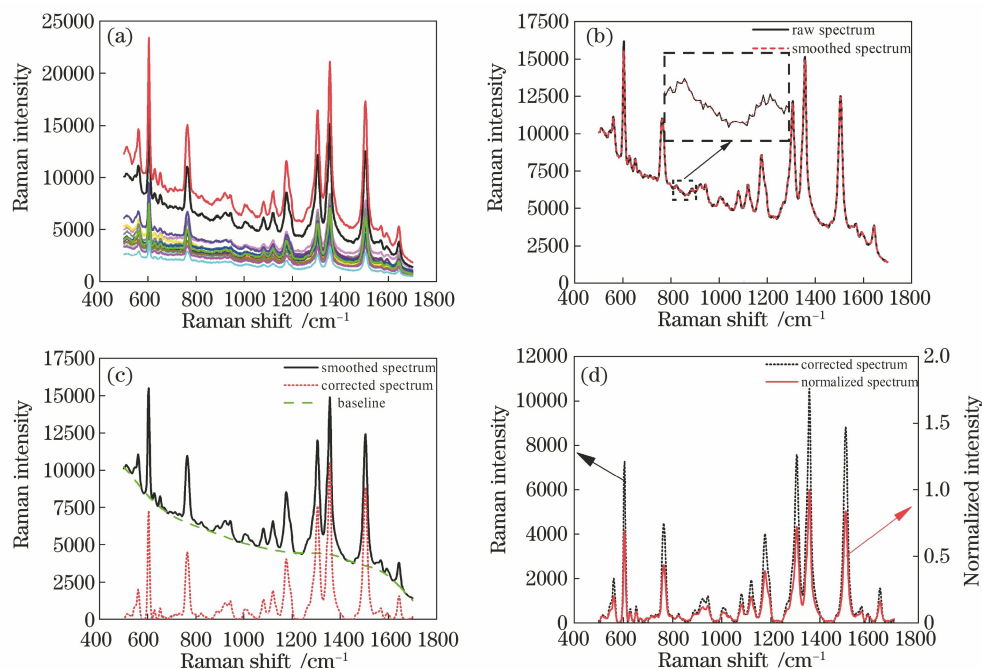


图 1 光谱的预处理结果。(a)光谱的截取;(b)平滑处理;(c)基线校正;(d)归一化处理

Fig. 1 Preprocessing results of the spectrum. (a) Interception of the spectrum; (b) smoothing processing; (c) baseline correction; (d) normalization processing

3 模型建立

3.1 建模和预测流程

模型的建立及预测流程如图 2 所示, 对预处理后的全部样本每种取 $2/3$ 作为训练集, 分别建立粗分模型和细分模型。先用粗分模型区分单组分样本和多组分样本, 再用细分模型对多组分样本的含量进行定量预测。粗分模型的建立步骤: 用特征峰判别算法完成训练集全部样本的特征提取, 给定训练集单组分/多组分的类别标签, 建立 MC-SVM 粗分模型, 进而用该模型对盲测样本进行单组分和多组分鉴定。细分模型的建立步骤: 用 N-PCA 法完成

训练集中多组分样本的特征提取, 给定训练集样本组分含量的标签, 建立 M-LS-SVR 细分模型, 进而对多组分盲测样本的组分进行准确预测。

3.2 粗分模型的建立

3.2.1 特征选择与提取

在单组分、多组分样本的定性鉴别中, 当多组分样本的某一组分含量较低时, 其光谱与单组分样本的光谱相似度很高, 如图 3 所示。其中, 单组分 (NB) 与多组分 ($V(\text{NB}):V(\text{CV})=9:1$) 样本的 SERS 只在 720 cm^{-1} 拉曼位移附近 (虚线圆圈) 有细微差别。因此, 如何有效提取出这类细微差别是需要解决的关键问题。

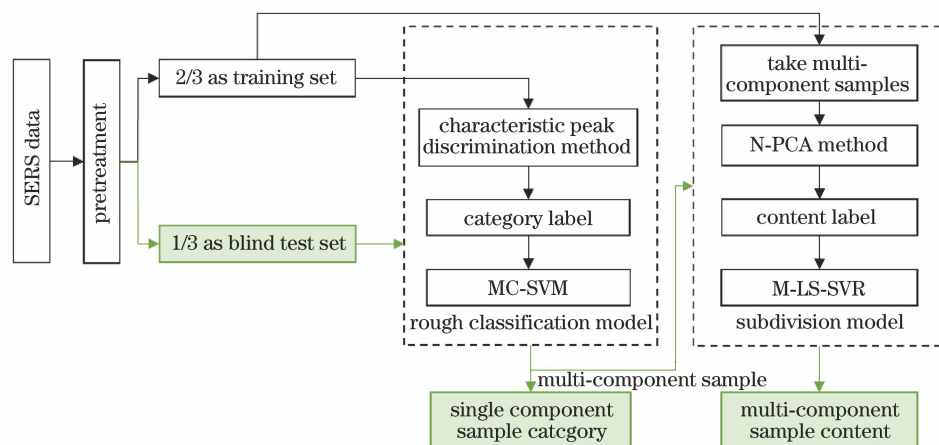


图 2 建模和预测的流程图

Fig. 2 Flow chart of modeling and forecasting

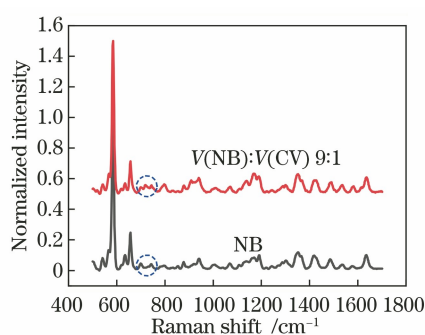


图 3 两条相似的光谱曲线

Fig. 3 Two similar spectral curves

图 4 为 CV、R6G 和 NB 的光谱图,可以发现,在拉曼位移分别为 $583.4, 604.4, 720.4 \text{ cm}^{-1}$ 时, NB、R6G 和 CV 样本光谱存在特征峰,且这 3 个特征峰在 2.1 小节混合实验中受谱峰重叠的影响很小。因此,用这 3 个特征峰构成一个三维特征向量,采用特征峰判别算法分辨单组分样本和多组分样本。若只提取到上述 3 个特征峰中的 1 个,则判定为单组分样本;若提取到 2 个以上的特征峰,则判定为多组分样本。

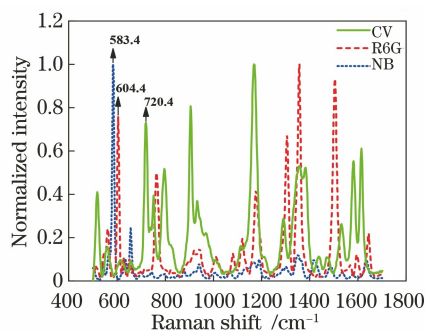


图 4 CV、R6G 和 NB 的光谱图

Fig. 4 Spectra of CV, R6G and NB

在提取特征峰的过程中,常见的极大值寻峰算法是将搜索到的每一个极大值点均认定为拉曼峰,容易受到噪声和干扰峰的影响。 $V(\text{CV}):V(\text{R6G})=1:9$ 的混合溶液和只含 R6G 的溶液 2 种样本经预处理后的光谱如图 5(a) 所示,可以发现,在 CV 溶液的特征峰 720 cm^{-1} 拉曼位移附近,多组分样本中含有 CV,应只检测出 1 个拉曼峰,而只含 R6G 的样本不应检测出拉曼峰。但由于光谱的波动,2 种样本均错误地检测出了 1~3 个峰值点,如图 5(b) 所示。如果通过极大值寻峰,错误率会非常高,而直接设定峰高阈值,又可能出现漏检问题。

为了准确提取出光谱的特征峰,设计的特征峰判别算法如下。

1) 对预处理后的数据再次进行高强度移动窗口平滑处理(窗口数为 5),平滑后的光谱如图 5(c) 所示。可以发现,二次平滑能大幅降低随机噪声的影响,提高特征峰提取的准确性。

2) 将待检测波段中的所有波峰都记为极大值点(LM),对于每个 LM,其两侧均存在若干极小值点,将其左右两侧所有极小值点中的最小值分别记为左侧全局极小值(LGM)和右侧全局极小值(RGM)。对比 LGM 与 RGM 的大小,将两者中数值较大的记作较大的全局最小值(LarMM),计算 LM 与 LarMM 的差值,得到阈值参数 H 。

3) 设置阈值下限 H_0 ,依次确认 $583, 604, 720 \text{ cm}^{-1}$ 拉曼位移附近的波峰是否存在大于阈值参数 H_0 的波峰。存在时记录对应的特征峰位,否则补 0 值代替,将 3 处特征峰位的检测结果分别作为特征向量的第一、第二和第三维度。图 5(d) 为该特征峰判别算法的识别结果。可以发现,包含 CV、

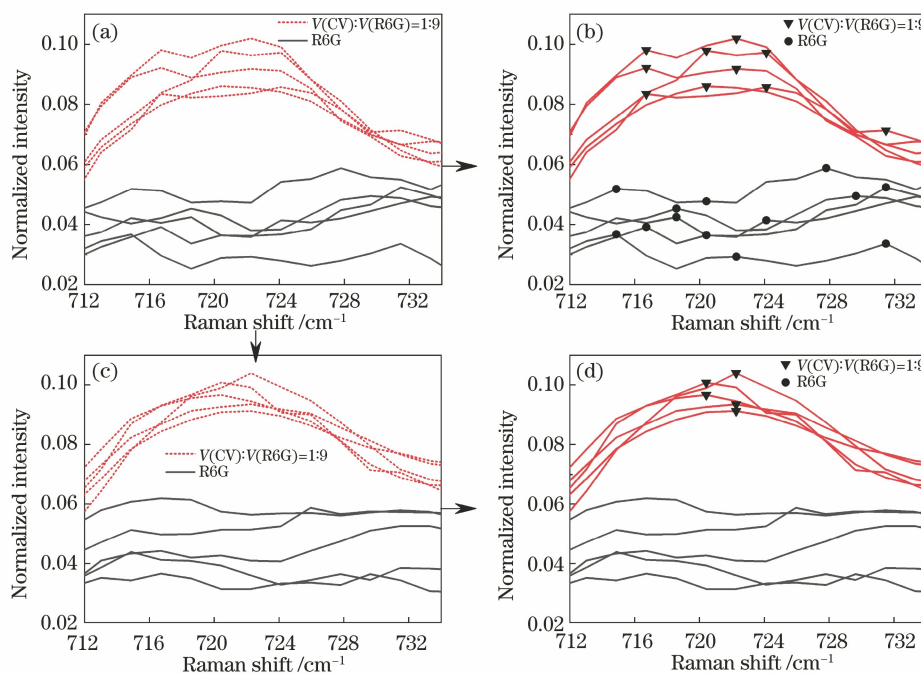


图 5 多组分和单组分溶液的光谱。(a)预处理后的光谱;(b)极大值法寻峰法的结果;
(c)光谱的二次平滑结果;(d)特征峰判别算法的寻峰结果

Fig. 5 Spectra of multi-component and single-component solutions. (a) Preprocessed spectrum; (b) peak-finding result of maximum value method; (c) secondary smoothing result of spectrum; (d) peak-finding result of characteristic peak discrimination algorithm

R6G 的多组分样本在 720 cm^{-1} 附近只检测出了 1 个拉曼峰,而单组分 R6G 样本没有检测出任何峰,与预期结果完全一致。

3.2.2 M-SVM 建模

用上述特征提取算法对随机选取的 180 例训练集样本进行特征提取,聚类结果及点数如图 6 所示。统计发现,180 例训练样本中只有 1 例 R6G、NB 多组分样本的特征被错误提取(箭头所指位置),正确率为 99.44%。由于训练集只有 180 例样本且归属于 6 大类,因此,采用适用于小样本的 MC-SVM 将训练集 180 例样本提取出的特征和已知分类标签共同输入 MC-SVM 中建立了粗分模型。

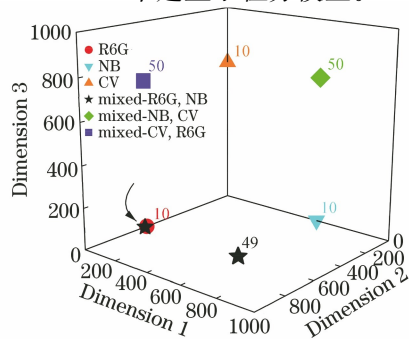


图 6 不同特征的聚类图

Fig. 6 Clustering diagram of different features

3.3 细分模型的建立

用粗分模型对单组分样本和多组分样本进行分类后,需要进一步建立细分模型以精确识别其中多组分样本的含量。

3.3.1 特征的选择与提取

由于拉曼强度与物质浓度正相关,因此,特征峰的相对强度会随混合物比例的变化而变化,但峰位通常不会发生变化。图 7 为 R6G 与 NB 不同体积比混合时峰值强度的变化曲线,可以发现,R6G、NB 混合溶液的配比不同时,峰值强度有明显变化。因此,现有的混合物组分定量 SERS 检测研究报道中,

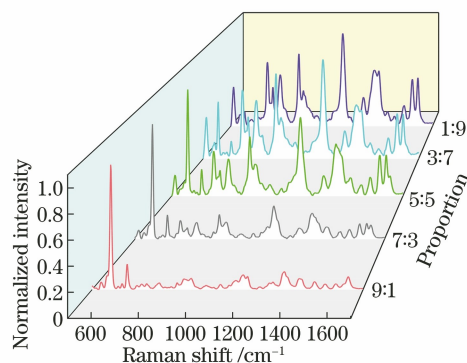


图 7 峰值强度的变化曲线

Fig. 7 Change curve of the peak intensity

多通过人为选定 1 个或多个特征峰,然后提取其归一化强度值作为模型的输入特征。

由于归一化过程中包括比例运算,其结果本身就可以作为样本的特征。但该结果的数据维度较高,不利于建模。为了降低计算复杂度,采用主成分分析(PCA)法对样本的归一化结果(741 维)进行降维,并提取 PCA 累计贡献不低于 99%的所有主成分(PC)作为样本特征。相比人为提取峰值

强度的方法,该方法更简单、可靠。首先,对训练集中 R6G、NB 多组分样本进行 PCA 处理,得到 10 个主成分,其累计贡献率如图 8(a)所示。可以发现,光谱的大部分原始信息(96.975%)被压缩在前 3 个主成分中。前 3 个主成分得分的散点分布及各类样本的置信椭球(置信概率为 95%)如图 8(b)所示,可以发现,体积比不同的 5 种样本已被区分开。

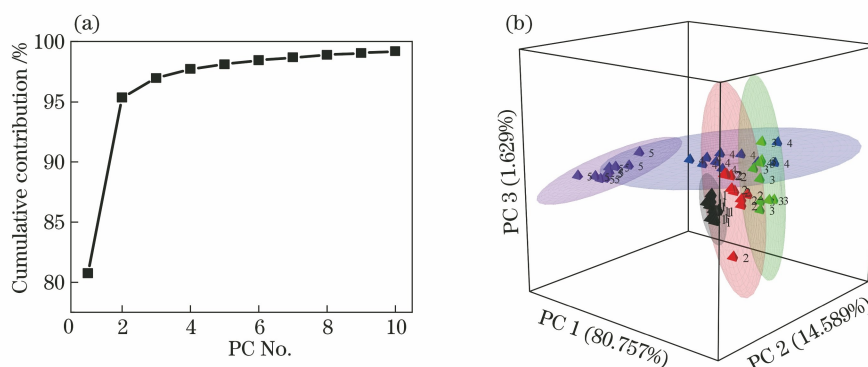


图 8 PCA 的处理结果。(a)主成分的累计贡献图;(b)前 3 个主成分得分的散点图及置信椭球

Fig. 8 Processing results of the PCA. (a) Cumulative contribution graph of principal components; (b) scatter plot and confidence ellipsoid of the scores of the first 3 principal components

同样的,在累计贡献不低于 99%条件下得到多组分样本 NB、CV 的 6 个主成分,CV、R6G 的 11 个主成分,各自的累计贡献率如图 9(a)和图 9(b)所示。其中,前 3 个主成分压缩的原始信息分别为 98.251%

和 95.637%。前 3 个主成分的得分分布如图 10(a)和图 10(b)所示,可以发现,不同比例 NB 和 CV 混合样本的置信椭球存在部分交叠,不同比例 R6G 和 CV 混合样本的置信椭球之间可以明显被区分。

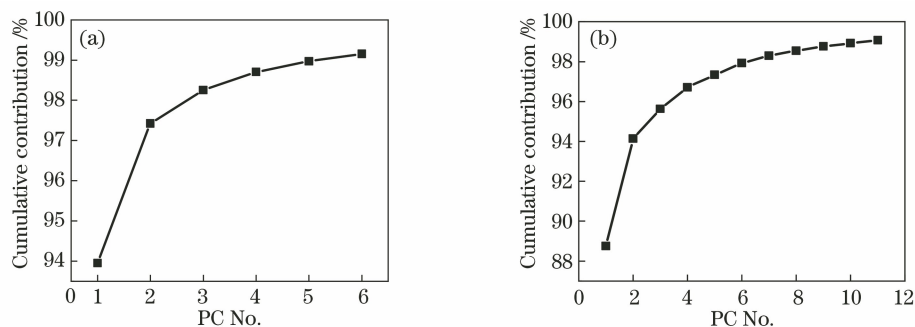


图 9 多组分样本的主成分累计贡献图。(a) NB, CV; (b) CV, R6G

Fig. 9 Cumulative contribution of principal components of multi-component samples. (a) NB, CV; (b) CV, R6G

由置信椭球的交叠情况可知,直接通过样本特征的聚类结果进行混合物组分的定量检测效果不佳,且对于未知组分的混合物,这种无监督方式也无法进行回归分析。因此,有必要进行有监督式学习,即先通过 N-PCA 方法提取训练集多组分样本的特征,给定标签后训练得到最优模型,再利用模型对样本的组分含量进行预测。

3.3.2 M-LS-SVR 建模

M-LS-SVR 作为一种有监督学习方法^[25],通过

核函数完成空间维度的转变,将低维空间中的线性不可分转化为高维空间中的线性可分,有效克服了维数灾难问题。因此,选用 M-LS-SVR 建立细分模型。建模过程中,M-LS-SVR 的参数 γ 和 σ^2 会直接影响模型的预测效果。其中,正则化参数 γ 用来调节模型的经验风险和置信范围的比例,径向基函数(RBF)的半径 σ^2 会影响特征空间中的样本分布复杂度。为了使模型具有良好的预测能力,用 PSO 算法对参数 γ 和 σ^2 进行优化,PSO 算法

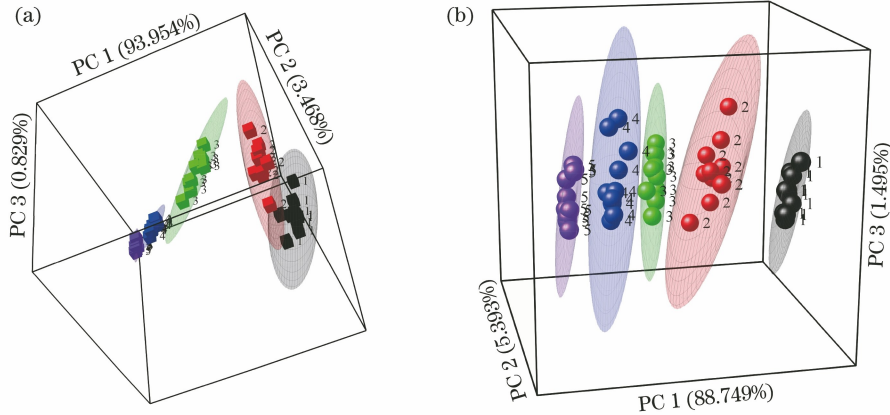


图 10 多组分样本的主成分得分散点图。(a) NB,CV;(b) CV,R6G

Fig. 10 Distribution point diagram of the principal components of multi-component sample. (a) NB,CV; (b) CV,R6G

的优点的是预设参数少,收敛速度快,具体优化步骤如下。

1) 设置算法初始化参数,包括最大迭代次数、种群个数、gam 和 sig2 的取值范围、速度范围、权重、学习因子。

2) 在 gam 和 sig2 取值范围内随机初始化种群位置,在速度范围内随机初始化种群中每个粒子的速度,将每个粒子的当前位置和初始粒子群中最佳位置分别设置为 pBest 和 gBest。

3) 用训练集样本的均方根误差作为适应度值,进行 1 次迭代。当粒子的本代适应度值优于上代自身的适应度值时,进行个体适应度、位置的替换;当粒子的本代适应度值优于上代全体粒子的最优适应

度时,同样进行最优适应度、最优位置的替换。

4) 更新粒子速度和位置,重复迭代过程,满足一定条件或达到最大迭代次数后输出最优参数。用 N-PCA 方法对训练集 150 例多组分样本进行特征提取,给定组分含量标签,将 M-LS-SVR 的参数 gam 和 sig2 设置为 PSO 寻找的最优参数,训练得到细分模型。

4 实验结果与评价

4.1 粗分模型的评价

用粗分模型鉴定盲测集的 90 例样本,结果如图 11 所示。其中,纵坐标表示 6 个样本类。统计发现,盲测集共 90 例样本均被正确检出。

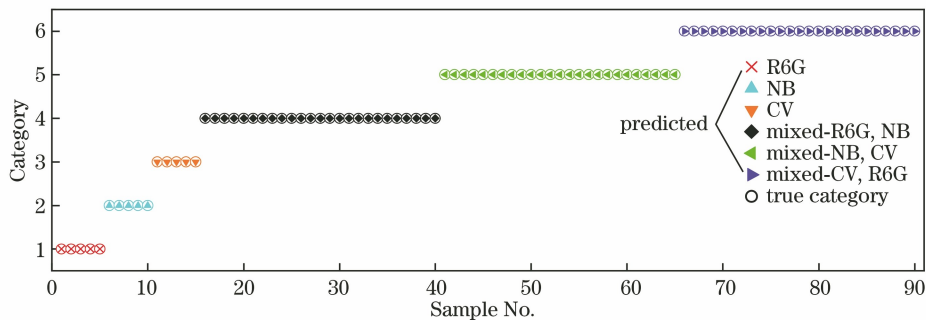


图 11 粗分模型预测结果

Fig. 11 Prediction results of the rough classification model

4.2 细分模型的评价

用细分模型对盲测集中检出的 75 例多组分样本的含量进行预测,结果如图 12 所示,其中,○为真实含量值,×为预测的含量值;5 种含量下 R6G、NB 和 CV 预测结果的平均误差如表 2 所示,统计结果表明,细分模型的最大平均误差为 2.67343%。

将样本的预测值与标准真值进行对比,并用相

关系数(R)和预测均方根误差(RMSEP)对模型的性能进行评价,结果如表 3 所示。其中,相关系数越接近 1,均方根误差越小,表明预测值与真值之间的偏差就越小,建模效果也就越好。RMSEP 和相关系数 R 可表示为

$$X_{RMSEP} = \sqrt{\frac{1}{K} \sum_{i=1}^K (y_i - \hat{y}_i)^2}, \quad (1)$$

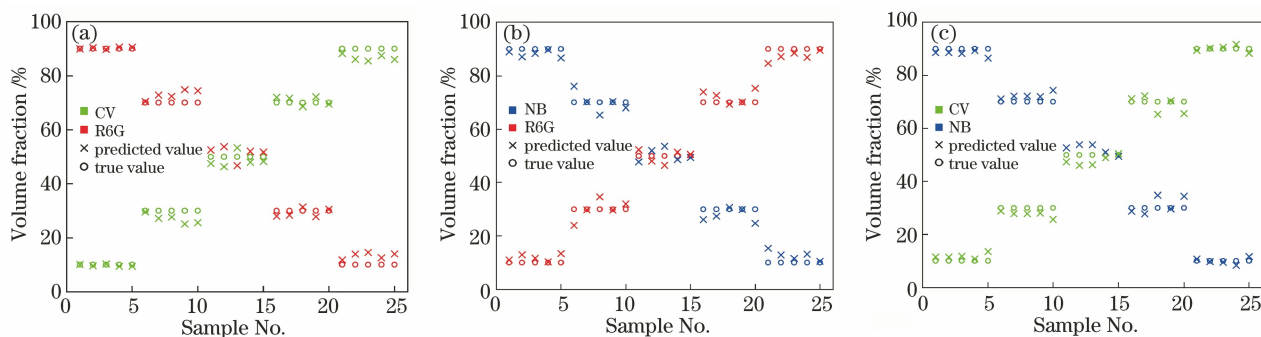


图 12 细分模型预测结果。(a) CV,R6G;(b) NB,R6G;(c) CV,NB

Fig. 12 Prediction results of the subdivision model. (a) CV, R6G; (b) NB, R6G; (c) CV, NB

表 2 细分模型的平均误差

Table 2 Average error of subdivision model

True value	unit: %		
	R6G	NB	CV
10	0.79259	2.59123	1.34607
30	2.67343	0.39853	0.56356
50	1.75400	0.53785	1.20700
70	0.12507	2.61558	1.16878
90	1.67570	1.20206	1.83150

$$R = \frac{\sum_{i=1}^K (y_i - \bar{y}_i)(\hat{y}_i - \bar{\hat{y}}_i)}{\sqrt{\sum_{i=1}^K (y_i - \bar{y}_i)^2} \times \sqrt{\sum_{i=1}^K (\hat{y}_i - \bar{\hat{y}}_i)^2}}, \quad (2)$$

式中, y_i 和 \hat{y}_i 分别为第 i 个样本的标准值和预测值, \bar{y}_i 和 $\bar{\hat{y}}_i$ 分别为标准值和预测值的均值, K 为样

本数。

用人为选取特征峰的方法对样本进行特征提取。以多组分 R6G、NB 样本为例,选择其光谱 8 个关键特征峰 (583.4, 604.4、655.3, 762.6, 1175, 1305, 1357, 1505 cm^{-1}) 的峰值强度作为特征,建立 PSO-M-LSSVR 细分模型并进行评价,结果如表 3 所示。可以发现,相比人为提取特征的方法,本方法用 N-PCA 自动提取的特征建立相同模型后, RMSEP 均有所降低,相关系数均有所提升,这表明本模型的预测准确度较高。R6G、NB 和 CV 真实值 x 与预测值 y 的相关度曲线如图 13(a)、图 13(b) 和图 13(c) 所示,其斜率依次为 0.97714、0.97689 和 0.98806,均接近 1,这表明 N-PCA 法结合 PSO-M-LSSVR 建立的细分模型能更精准地预测多组物质的含量。

表 3 不同特征提取方法的预测结果

Table 3 Prediction results of different feature extraction methods

Method	R6G		NB		CV	
	RMSEP	R	RMSEP	R	RMSEP	R
N-PCA	0.02613	0.99588	0.02635	0.99566	0.02298	0.99677
Artificial selection	0.02663	0.99553	0.02945	0.99455	0.02537	0.99615

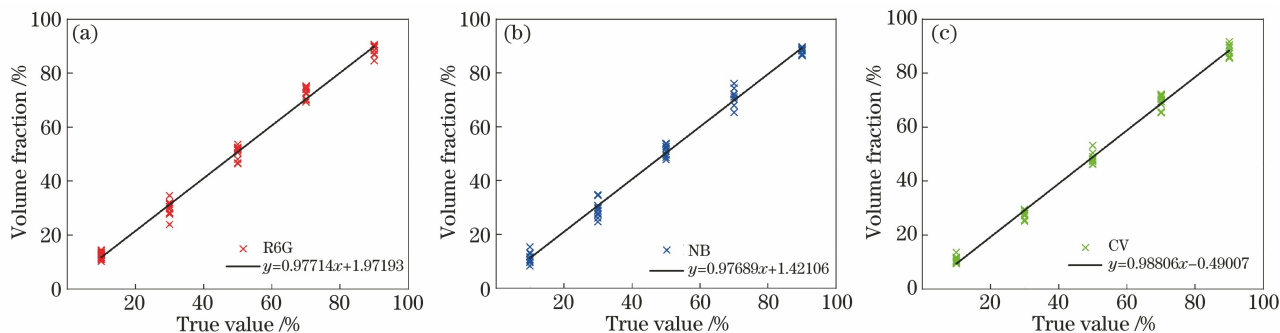


图 13 真实值与预测值的相关度曲线。(a) R6G;(b) NB;(c) CV

Fig. 13 Correlation curve between true value and predicted value. (a) R6G; (b) NB; (c) CV

5 结 论

构建了一种用于物质定性、定量检测的递进式拉曼检测系统,并达到了预期的预测效果。在单组分和多组分物质的定性鉴别中,使用的特征峰判别算法可以准确提取样本的特征,有效弥补了极大值寻峰法的不足;建立的 MC-SVM 粗分模型在物质的定性鉴别上的效果良好,90 例样本全部被正确识别。在多组分物质含量的定量检测中,针对现有人工提取特征的不足,用 N-PCA 法自动提取光谱特征,并建立了用 PSO 算法优化的 M-LS-SVR 细分模型。实验结果表明,本方法测试 R6G、NB 和 CV 的相关系数分别为 0.99588, 0.99566 和 0.99677, 均方根误差分别为 0.02613, 0.02635 和 0.02298, 线性拟合斜率分别为 0.97714, 0.97689 和 0.98806, 最大平均误差为 2.67343%。相比现有人工提取特征方法,本方法在优化特征的同时提高了模型的预测精度,为混合物组分的定量检测提供了一种有效方法。在后续工作中还需以药品、危险液体等复杂混合物作为样本,进一步细化配置梯度,开展更深入的研究。

参 考 文 献

- [1] Nie X M, Chen Z Y, Tian Y P, et al. Rapid detection of trace formaldehyde in food based on surface-enhanced Raman scattering coupled with assembled purge trap [J]. *Food Chemistry*, 2021, 340: 127930.
- [2] Wang H B. Research progress on application of surface enhanced Raman spectroscopy in food detection [J]. *Science and Technology of Food Industry*, 2019, 40(15): 322-329.
王海波. 表面增强拉曼光谱用于食品检测的研究进展 [J]. *食品工业科技*, 2019, 40(15): 322-329.
- [3] Wang Q, Zeng W D, Xia Z P, et al. Recognition of food-borne pathogenic bacteria by Raman spectroscopy based on random forest algorithm [J]. *Chinese Journal of Lasers*, 2021, 48(3): 0311002.
王其, 曾万聃, 夏志平, 等. 基于随机森林算法的食源性致病菌拉曼光谱识别 [J]. *中国激光*, 2021, 48(3): 0311002.
- [4] Liu S S, Li H H, Hassan M M, et al. SERS based artificial peroxidase enzyme regulated multiple signal amplified system for quantitative detection of foodborne pathogens [J]. *Food Control*, 2021, 123: 107733.
- [5] Wang T Y, Wang Y Y, Lin X L, et al. Ultrasensitive quantitative detection of alpha-fetoprotein based on SERS spectroscopy [J]. *Chinese Journal of Lasers*, 2020, 47(2): 0207026.
王廷银, 王运斌, 林学亮, 等. 基于 SERS 光谱技术的甲胎蛋白超灵敏定量检测 [J]. *中国激光*, 2020, 47(2): 0207026.
- [6] Hu D Y, Xu X S, Zhao Z Y, et al. Detecting urine metabolites of bladder cancer by surface-enhanced Raman spectroscopy [J]. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2021, 247: 119108.
- [7] Nargis H F, Nawaz H, Bhatti H N, et al. Comparison of surface enhanced Raman spectroscopy and Raman spectroscopy for the detection of breast cancer based on serum samples [J]. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2021, 246: 119034.
- [8] Dong R L, Li S F, Lin D Y, et al. Progress of the applications of surface-enhanced Raman spectroscopy in illicit drug detection [J]. *Scientia Sinica (Chimica)*, 2021, 51(3): 294-309.
董荣录, 李绍飞, 林东岳, 等. 表面增强拉曼光谱在毒品检测中的应用进展 [J]. *中国科学: 化学*, 2021, 51(3): 294-309.
- [9] Wu J J, Feng Y, Zhang L, et al. Nanocellulose-based surface-enhanced Raman spectroscopy sensor for highly sensitive detection of TNT [J]. *Carbohydrate Polymers*, 2020, 248: 116766.
- [10] Li N, Han S, Zhao H, et al. Study on rapid detection of illegal dye in cinnabar with paper-based SERS substrate [J]. *The Journal of Light Scattering*, 2019, 31(3): 242-248.
李楠, 韩斯琴高娃, 赵航, 等. 应用 SERS 滤纸基底快速检测朱砂中违禁染料的研究 [J]. *光散射学报*, 2019, 31(3): 242-248.
- [11] Zuo Q, Chen Y, Shi C X, et al. Recent developments on quantitative surface-enhanced Raman spectroscopic techniques [J]. *Chinese Journal of Analytical Chemistry*, 2015, 43(11): 1656-1663.
左奇, 陈瑶, 石彩霞, 等. 表面增强拉曼散射光谱定量分析技术的研究进展 [J]. *分析化学*, 2015, 43(11): 1656-1663.
- [12] Lin L. The qualitative and quantitative analysis of pesticide residue in tea using surface-enhanced Raman spectroscopy (SERS) [D]. Nanchang: Jiangxi Agricultural University, 2014.
蔺磊. 表面增强拉曼光谱(SERS)定性定量分析茶叶中农药残留的方法研究 [D]. 南昌: 江西农业大学, 2014.
- [13] Liu C Z, Zhu Q B, Huang M, et al. Identification of components in mixtures based on Raman

- spectroscopy[J]. *Laser & Optoelectronics Progress*, 2019, 56(8): 083004.
- 刘财政, 朱启兵, 黄敏, 等. 基于拉曼光谱的混合物组分识别方法[J]. *激光与光电子学进展*, 2019, 56(8): 083004.
- [14] Ji M Q, Zhu Q B, Huang M, et al. Method for improving identification accuracy of components in mixtures using Raman spectra of known mixtures[J]. *Chinese Journal of Lasers*, 2020, 47(11): 1111001.
- 季明强, 朱启兵, 黄敏, 等. 利用已知混合物拉曼光谱改善混合物成分识别精度的方法[J]. *中国激光*, 2020, 47(11): 1111001.
- [15] Chen Y, Yan X, Zhang X, et al. Surface-enhanced Raman spectroscopy quantitative analysis of polycyclic aromatic hydrocarbons based on support vector machine algorithm[J]. *Chinese Journal of Lasers*, 2019, 46(3): 0311005.
- 陈阳, 严霞, 张旭, 等. 基于支持向量机算法的多环芳烃表面增强拉曼光谱的定量分析[J]. *中国激光*, 2019, 46(3): 0311005.
- [16] 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.
- [17] Fu X H, Zhao F, Wang Z X, et al. Quantitative analysis of goat serum protein content by Raman spectroscopy based on IABC-SVR[J]. *Spectroscopy and Spectral Analysis*, 2021, 41(2): 540-545.
- 付兴虎, 赵飞, 王振兴, 等. 基于 IABC-SVR 算法的拉曼光谱定量分析山羊血清蛋白含量[J]. *光谱学与光谱分析*, 2021, 41(2): 540-545.
- [18] Zhang Y J, Zhang F C, Fu X H, et al. Detection of fatty acid content in mixed oil by Raman spectroscopy based on ABC-SVR algorithm[J]. *Spectroscopy and Spectral Analysis*, 2019, 39(7): 2147-2152.
- 张燕君, 张芳草, 付兴虎, 等. 基于 ABC-SVR 算法的拉曼光谱检测混合油脂脂肪酸含量[J]. *光谱学与光谱分析*, 2019, 39(7): 2147-2152.
- [19] Zhang B, Deng Z Y, Zheng J K, et al. Determination of gasoline composition based on Raman spectroscopy[J]. *Spectroscopy and Spectral Analysis*, 2015, 35(6): 1577-1581.
- 张冰, 邓之银, 郑靖奎, 等. 拉曼光谱技术的汽油组分含量测定[J]. *光谱学与光谱分析*, 2015, 35(6): 1577-1581.
- [20] Meyer B. Object oriented software construction[M]. Beijing: China Machine Press, 1988.
- [21] Yang W, Li Q Q. Survey on particle swarm optimization algorithm[J]. *Engineering Science*, 2004, 6(5): 87-94.
- 杨维, 李歧强. 粒子群优化算法综述[J]. *中国工程科学*, 2004, 6(5): 87-94.
- [22] Yang F, Wen P, Zhang Z Q, et al. Fabrication of flexible surface-enhanced Raman spectroscopy chip[J]. *Chinese Journal of Lasers*, 2021, 48(1): 0113001.
- 杨峰, 文枰, 张志强, 等. 柔性表面增强拉曼光谱芯片制备[J]. *中国激光*, 2021, 48(1): 0113001.
- [23] Zhang Z M, Chen S, Liang Y Z. Baseline correction using adaptive iteratively reweighted penalized least squares[J]. *The Analyst*, 2010, 135(5): 1138-1146.
- [24] Tang R Z. Research on the method of data normalization for improving SVM training efficiency[D]. Jinan: Shandong Normal University, 2017.
- 汤荣志. 数据归一化方法对提升 SVM 训练效率的研究[D]. 济南: 山东师范大学, 2017.
- [25] Gu Y P, Zhao W J, Wu Z S. Least squares support vector machine algorithm[J]. *Journal of Tsinghua University (Science and Technology)*, 2010, 50(7): 1063-1066, 1071.
- 顾燕萍, 赵文杰, 吴占松. 最小二乘支持向量机的算法研究[J]. *清华大学学报(自然科学版)*, 2010, 50(7): 1063-1066, 1071.