

激光与光电子学进展

基于光谱数据融合技术的手帕纸品牌分类

季佳华, 王继芬*, 何欣龙

中国人民公安大学侦查学院, 北京 100038

摘要 光谱数据融合作为一种新兴的光谱分析技术,可以有效结合不同光谱技术的检出特点。本文将红外光谱指纹区与拉曼光谱信号较强波段数据进行融合,对 288 个手帕纸样本开展分类研究。采用线性判别分析对 189 个训练样本建立分类模型,模型分类结果相比于独立使用红外光谱或拉曼光谱具有显著提升的准确率,但区分效果仍然不理想。利用主成分分析提取 35 个主成分进行线性判别分析的准确率可达 100%。提取测试样本的 32 个主成分建模,模型的准确率可达 100%,分类结果理想。光谱融合技术为法庭科学领域手帕纸的分类鉴别提供了一种有效工具。

关键词 光谱学; 光谱数据融合; 手帕纸; 线性判别分析; 红外光谱; 拉曼光谱

中图分类号 O657

文献标志码 A

doi: 10.3788/LOP202158.0330004

Classification of Handkerchief Paper Brand Based on Spectral Data Fusion Technology

Ji Jiahua, Wang Jifen*, He Xinlong

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

Abstract Spectral data fusion, an emerging spectral analysis technology, can effectively combine the detection characteristics of different spectrum technologies. The classification of 288 handkerchief paper samples were studied with data fusion of infrared spectrum fingerprint region and strong signal region of Raman spectrum. A classification model was established on 189 training samples by linear discriminant analysis. Compared with the independent use of infrared spectroscopy or Raman spectroscopy, the accuracy was significantly improved. However, the distinguishing effect was still not ideal. The accuracy of 35 principal components extracted by principal components analysis for linear discriminant analysis can reach 100%. The model accuracy of the 32 principal components extracted from test samples reached 100%, and the classification results were ideal. The spectral fusion technology provides an effective tool for the classification and identification of handkerchief paper in the field of forensic science.

Key words spectroscopy; spectral data fusion; handkerchief paper; linear discriminant analysis; infrared spectrum; Raman spectrum

OCIS codes 300.6170; 300.6340; 300.6450

1 引言

纸巾是生活中不可或缺的清洁工具。手帕纸

主要由木浆、水和各种添加剂制成,其因包装精巧、质地良好等特点而广受大众喜爱。在法庭科学领域,不法分子常常将使用后的手帕纸遗留在犯罪现

收稿日期: 2020-06-24; 修回日期: 2020-07-06; 录用日期: 2020-07-10

基金项目: 中国人民公安大学 2019 年度基本科研业务费项目库重点项目(2019JKF223)

*E-mail: wangjifen58@126.com

场,手帕纸在作为一些生物物证载体的同时,其自身信息也可以为公安机关侦破案件提供线索。因此,如何对手帕纸的来源进行认定成为法庭工作人员研究的焦点问题之一。所查资料显示,目前关于手帕纸的鉴别分类暂无文献报道。

光谱分析作为一种仪器分析方法,不但能高效地检测样本,而且在检测时不需要对样品进行前处理,或者只需要进行简单的处理^[1-3],在化学^[4-6]、物理^[7-9]和医学^[10-12]等领域取得了一定进展。红外光谱和拉曼光谱技术因具有操作简单以及高效率和高通量等特点而被广泛应用。非极性或者具有对称性的基团,如C=C、S—S等,基本没有红外吸收光谱效应,不适合采用红外光谱技术进行检测,但拉曼光谱对此类基团的敏感度很高。类似地,拉曼光谱技术对极性基团(如C=O、N—H等)的灵敏度低,而红外光谱技术常被用于对含有此类基团的分子进行分析。因此,采用红外光谱技术和拉曼光谱技术相结合的方法可以对分子种类复杂的样品进行综合判断^[13-15]。

目前,红外光谱和拉曼光谱技术结合数据挖掘技术的应用很广泛,例如:曹馨艺等^[16]采用拉曼光谱技术与化学计量学相结合的方法对42种花粉样品进行了分类鉴定,科间花粉预测准确率为97.75%,蔷薇科属间花粉预测准确率为90.47%;何欣龙等^[17]基于红外光谱技术对橡胶颗粒建立了分类模型,75个样品的分类准确率为92%;基于光谱融合技术的思想,陈达等^[18]采用离散小波变换等方法融合拉曼光谱和红外光谱信息对掺假奶粉进行了诊断,诊断结果相比于单独对红外或拉曼光谱建模灵敏度显著提高,准确率也提升了几个百分点;徐伟杰等^[19]将近红外光谱和拉曼光谱结合起来对火星表面的相关矿物进行了分类鉴别,并利用累加融合和串联融合的方法分别对分类模型进行考察,结果表明,串联融合后光谱数据的准确率更高;周昆鹏等^[20]将紫外吸收光谱技术和荧光发射光谱技术结合起来对水质化学需氧量进行了检测,检测效果理想,对实践中水质化学需氧量的快速检测具有一定的参考价值;向伶俐等^[21]将近红外光谱技术与中红外光谱技术结合起来对葡萄酒的产地建立判别模型,训练集的平均准确率为87.11%,测试集的平均准确率为90.87%,与单独建模的准确率相比明显提高,说明基于光谱融合对葡萄酒产地的鉴别具有一定的可行性。基于上述文献报道可以发

现,融合数据分析可以根据不同光谱技术和数据挖掘技术的特性引入更多的样品信息,有利于提高数据分析的准确率和有效性,具有一定的参考价值。

为了实现对手帕纸样本的有效分类,本文采集了手帕纸样本的拉曼光谱和红外光谱,将红外光谱的指纹区和拉曼光谱特征信号波段融合,然后对融合后的数据进行标准化处理,并基于线性判别分析和主成分分析对光谱融合数据建立分类模型,期为法庭科学领域中手帕纸的分类研究提供一定的参考和借鉴。

2 实验部分

2.1 实验样本

收集了市场上常见的清风(QF)、维达(WD)、舒洁(SJ)、小宝贝(XBB)、宇柔(YR)、心相印(XXY)、星之恋(XZL)和雪松(XS)8种品牌不同系列的288个手帕纸样本,以及某公安机关提交的两份未知品牌的手帕纸检材。将189个8种品牌的样本作为训练集,建立分类模型并进行优化;将126个心相印品牌不同系列的样本作为测试集,对优化后分类模型的可行性进行考察(其中有27个心相印样本为训练集与测试集共有样本)。利用该未知检材验证分类模型的实际应用价值。

2.2 建模原理

主成分分析(PCA)是一种将复杂的数据结构进行简化的统计学方法^[22],其主要思想是将多个具有相关性的变量转化为几个线性不相关的特征变量,并且得到的特征变量包含了原始变量的大部分信息。令原始数据 A 有 m 个样本,每个样本包含 n

个变量,则有 $A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$,假设该数据只有 p

个主成分,有
$$\begin{cases} P_1 = x_{11}A_1 + x_{12}A_2 + \cdots + x_{1n}A_n \\ P_2 = x_{21}A_1 + x_{22}A_2 + \cdots + x_{2n}A_n \\ \vdots \\ P_p = x_{p1}A_1 + x_{p2}A_2 + \cdots + x_{pn}A_n \end{cases}, \text{令}$$

$X_1 = [x_{11}, x_{12}, \cdots, x_{1n}]$,则有 $P_1 = X_1^T A$,此时向量 X_1 称为第一主成分的载荷。同理可以得出其他主成分。主成分分析的本质就是找到原始变量在 p 个主成分上的载荷,即 x_{ij} 。用线性代数可以证明 x_{ij} 为 A 的 p 个较大特征值对应的特征向量。

线性判别分析^[23](LDA)的思路与PCA相似。PCA的作用是将 n 维数据映射到 p 维上($n > p$),而LDA的作用是将 d 维数据投影到一条直线上,同时

要求数据投影到该直线上的分类最好,即同类别的数据投影点尽可能接近,不同类别的投影点尽可能疏远。在两种类别(w_1 和 w_2)样本数据下,训练样本集 $X=(x_1, \dots, x_N)$,每个样本向量的维度都是 d ,第一种类别(w_1)的样本为 $X_1=(x_{11}, \dots, x_{N1})$,第二种类别(w_2)的样本为 $X_2=\{x_{12}, \dots, x_{N2}\}$ 。要寻找一个投影方向 $W(W$ 向量同样在 d 维上),使投影以后的样本为

$$y_i = W^T X_i, i = 1, 2, \dots, N, \quad (1)$$

式中: N 为正整数。在原样本空间中,类均值向量为

$$m_i = \frac{1}{N_i} \sum_{x_j \in X_j} x_j, i = 1, 2, \dots, N. \quad (2)$$

定义各类的类内离散度矩阵为

$$S_i = \sum_{x_j \in X_j} (x_j - m_i)(x_j - m_i)^T, i = 1, 2, \dots, N, \quad (3)$$

总类内离散度矩阵为

$$S_w = S_1 + S_2, \quad (4)$$

类间离散度矩阵为

$$S_b = (m_1 - m_2)(m_1 - m_2)^T. \quad (5)$$

Fisher 判别准则变为 Rayleigh 商,即

$$\max J_F(w) = \frac{w^T S_b w}{w^T S_w w}. \quad (6)$$

最佳投影方向为

$$w_0 = S_w^{-1} (m_1 - m_2), \quad (7)$$

决策规则为

$$g(x) = w^T x + w_0, \quad (8)$$

则有

$$\begin{cases} g(x) \geq 0, x \in w_1 \\ g(x) < 0, x \in w_2 \end{cases} \quad (9)$$

3 结果与讨论

3.1 谱图分析

采集了手帕纸样本的红外光谱和拉曼光谱数据,如图 1 和图 2 所示。从两个光谱图中可以发现:不同样本的光谱信号在峰形上呈现出高度的一致性,相对峰高的差异较为微弱;在 3350 cm^{-1} 和 1000 cm^{-1} 处,红外光谱出现了两个明显的峰谷,拉曼光谱在 $1400, 1100, 400 \text{ cm}^{-1}$ 处都有明显响应,且无明显差别。横向比对各个谱图中不同样本的光谱信号可以发现,由于手帕纸中的杂质成分较多,谱图存在交叉重叠的现象,因此很难直接根据谱图对样本数据进行分类。纵向比对红外光谱和拉曼光谱谱图可以发现,在 2900 cm^{-1} 和 1400 cm^{-1} 等处出现了此消彼长的现象,这说明了拉曼光谱和红外光谱在检测极性与非极性分子上的互补性特征。通过分析

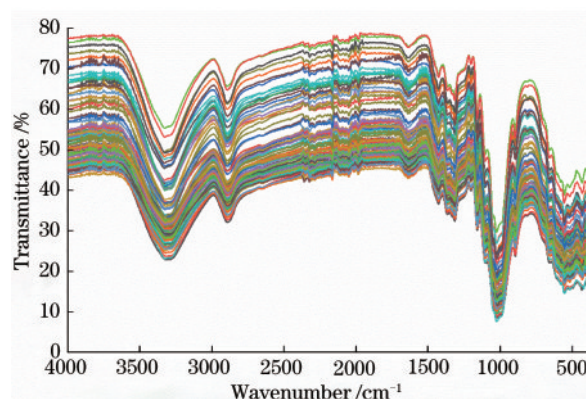


图 1 手帕纸样本的红外光谱

Fig. 1 Infrared spectra of handkerchief paper samples

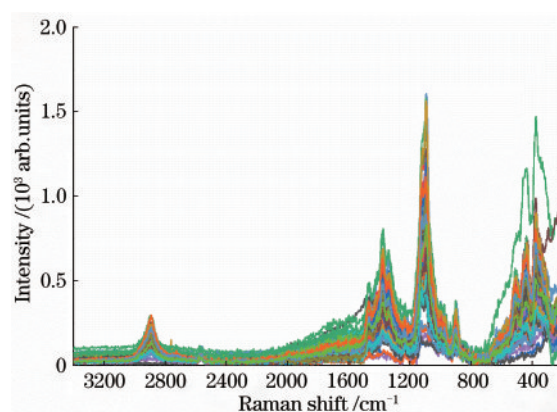


图 2 手帕纸样本的拉曼光谱

Fig. 2 Raman spectra of handkerchief paper samples

谱图可以发现,直接根据谱图对样本的种类进行判别是很难实现的。

3.2 线性判别分析

从红外光谱和拉曼光谱的谱图上看,图像规律呈现高度一致性,无法直接对样本进行分类,因此考虑采用光谱数据挖掘技术对样品进行分析。为了提升对光谱数据的处理速度,本文对红外光谱指纹区($1300 \sim 400 \text{ cm}^{-1}$)和拉曼光谱 $1500 \sim 200 \text{ cm}^{-1}$ 波段的信号进行线性判别分析。指纹区吸收峰的特征性很强,化合物结构上的微小变化会导致这个区域的信号出现明显差异,同时不同化合物的信号强度也不同。从拉曼光谱谱图中可以看出, $3400 \sim 1500 \text{ cm}^{-1}$ 波段的信号趋于平缓,因此选择 $1500 \sim 200 \text{ cm}^{-1}$ 波段进行数据分析。

表 1 是手帕纸在 $1300 \sim 400 \text{ cm}^{-1}$ 波段的红外光谱以及在 $1500 \sim 200 \text{ cm}^{-1}$ 波段的拉曼光谱的线性判别分类结果。准确率作为一种评价指标经常被应用于对分类结果的考察。一般情况下,准确率高,说明分类器越好,但有时准确率高不一定就代

表 1 两种光谱线性判别分析的准确率、召回率和假正率
Table 1 LDA accuracy, recall rate and false positive rate of two spectra

Brand	Infrared spectrum			Raman spectrum		
	Accuracy/%	Recall/%	False positive/%	Accuracy/%	Recall/%	False positive/%
QF	92.1	66.7	5.3	96.8	66.7	0
SJ	85.7	66.7	11.1	98.4	88.9	0
WD	92.1	66.7	2.0	88.9	66.7	5.9
XBB	92.1	55.6	1.9	93.7	77.8	3.7
XS	90.5	83.3	8.8	93.7	83.3	5.3
XXY	92.1	77.8	5.6	90.5	77.8	7.4
XZL	93.7	66.7	3.5	98.4	83.3	0.0
YR	92.1	33.3	1.8	98.4	100	1.8

表分类结果理想。因此,本文同时引入召回率和假正率结合准确率对分类结果进行评价。召回率是对覆盖面的度量,反映了分类器的灵敏度;假正率为被预测为 a 类的非 a 类样本数与非 a 类样本总数的比值,假正率越小,召回率越高,分类结果越好。

从表 1 中不难看出:在红外光谱中,准确率最低的品牌是舒洁,为 85.7%,准确率最高的为星之恋,为 93.7%,8 种品牌的平均召回率为 64.6%,平均假正率为 5.0%;在拉曼光谱中,准确率最低的品牌是维达,为 88.9%,最高的是舒洁、星之恋和宇柔,均为 98.4%,8 种品牌的平均召回率为 80.6%,平均假正率为 3.0%。可以发现,虽然两种光谱的准确率都很高,但召回率偏低,假正率也没有达到期望值。对比红外光谱和拉曼光谱的准确率、召回率和假正率可以发现,拉曼光谱的判别效果要比红外光谱好。结合假正率和召回率可判断出手帕纸样品的两种光谱数据在线性判别模型下的分类结果并不理想,准确率之所以高是因为样本数据中的品牌种类多,非 a 类的样本被判为非 a 类的样本量大,导致准确率增大。

3.3 光谱数据融合

直接对红外光谱和拉曼光谱数据进行线性判别分析无法实现手帕纸样本的精准分类。考虑到红外光谱和拉曼光谱的互补特性,本文将红外光谱在 $1300\sim 400\text{ cm}^{-1}$ 波段的数据与拉曼光谱在 $1500\sim 200\text{ cm}^{-1}$ 波段的数据进行融合。但由于两种光谱数据之间存在较大的数量级差,所以先对融合后的数据进行标准化处理,再对标准化后的数据进行线性判别分析,就可以得到光谱数据融合后的准确率、召回率和假正率。

从表 2 中可以看出,红外光谱与拉曼光谱融合后的数据准确率相比原始独立的数据更高,平均召

回率达到了 84.4%(比拉曼光谱高 3.8 个百分点),平均假正率低至 2.3%(比拉曼光谱低 0.7 个百分点)。召回率升高和假正率降低,说明光谱融合对线性判别分析的结果起到了一定的优化作用,但平均召回率只有 84.4%,最高的是舒洁、小宝贝和心相印品牌,但也只有 88.9%,仍然达不到预期的结果。

表 2 光谱融合后线性判别分析的准确率、召回率和假正率
Table 2 LDA accuracy, recall rate and false positive rate after spectral fusion

Brand	Accuracy/%	Recall/%	False positive/%
QF	98.4	83.3	0
SJ	98.4	88.9	0
WD	92.1	75.0	3.9
XBB	95.2	88.9	3.7
XS	95.2	83.3	3.5
XXY	93.7	88.9	5.6
XZL	98.4	83.3	0
YR	96.8	83.3	1.8

3.4 模型优化

上述为了提升模型的运算速度,选择红外光谱的指纹区和拉曼光谱信号强的波段分别建立线性判别分析的分类模型,结果不理想。考虑到红外光谱和拉曼光谱的互补性特点,对光谱数据进行融合,线性判别分析结果明显改善,但是还需要进行优化处理。红外光谱和拉曼光谱融合后会产生 911 个变量,数据维度高,部分异常数据对结果的判别会有影响。因为主成分分析具有降维、削弱或消除噪声的特点,且不改变原始数据的基本特征,所以本文利用线性判别分析对经主成分分析后的特征变量进行判别分析,以期实现对手帕纸的精准分类。

表3是光谱融合后的911个变量的主成分分析特征方差贡献率。在主成分分析中,特征根与方差贡献率为正相关,即:特征根减小,方差贡献率也会随之减小,对数据整体特征的影响越小。一般把特征根大于1且累计方差贡献率不小于85%的主成

分分数作为原始数据的特征变量。从表3中可以看出,PC1~PC35的特征根都大于1,且累计方差贡献率达到了98.530%,说明这35个特征变量包含了原始数据98.530%的数据信息,可以作为特征变量建立线性判别分析分类模型。

表3 主成分分析的特征方差贡献率

Table 3 Characteristic variance contribution rate of PCA

Component	Eigenvalue	Variance/%	Cumulative variance/%	Component	Eigenvalue	Variance/%	Cumulative variance/%
PC1	437.789	48.162	48.162	PC20	2.176	0.239	96.066
PC2	231.312	25.447	73.608	PC21	2.096	0.231	96.297
PC3	90.879	9.998	83.606	PC22	2.023	0.223	96.520
PC4	26.028	2.863	86.470	PC23	1.918	0.211	96.731
PC5	21.176	2.330	88.799	PC24	1.786	0.196	96.927
PC6	11.269	1.240	90.039	PC25	1.702	0.187	97.114
PC7	8.257	0.908	90.947	PC26	1.587	0.175	97.289
PC8	6.497	0.715	91.662	PC27	1.481	0.163	97.452
PC9	5.023	0.553	92.215	PC28	1.425	0.157	97.609
PC10	4.111	0.452	92.667	PC29	1.382	0.152	97.761
PC11	4.000	0.440	93.107	PC30	1.289	0.142	97.902
PC12	3.757	0.413	93.520	PC31	1.252	0.138	98.040
PC13	3.719	0.409	93.929	PC32	1.183	0.130	98.170
PC14	3.475	0.382	94.311	PC33	1.154	0.127	98.297
PC15	3.178	0.350	94.661	PC34	1.089	0.12	98.417
PC16	3.035	0.334	94.995	PC35	1.024	0.113	98.530
PC17	2.756	0.303	95.298	PC36	0.952	0.105	98.634
PC18	2.507	0.276	95.574	PC37	0.914	0.101	98.735
PC19	2.300	0.253	95.827				

Notes: omit the remaining principle components.

为了考察不同维度下的主成分对线性判别分析分类模型的影响,将PC1~PC35按照顺序依次输入(作为自变量)进行判别分析,得到不同维度下手帕纸样本分类的准确率,如表4所示。

表4是2~34维下各个品牌手帕纸分类的准确率及总体准确率。根据不同维度下各个品牌准确率的变化绘制折线图,如图3所示。从不同品牌手帕纸的分类准确率来看,11维下清风和小宝贝手帕纸的准确率首次达到了100%,13维下心相印手帕纸的准确率首次达到了100%,18维下舒洁和星之恋手帕纸的准确率首次达到了100%,19维下宇柔手帕纸的准确率首次达到了100%,31维下维达手帕纸的准确率首次达到了100%,雪松手帕纸在34维下才被完全检出。这说明在这个分类模型下,清风、小宝贝和心相印手帕纸更容易被检出,其他品牌手帕纸的鉴别在此模型下需要更多的特征输入。

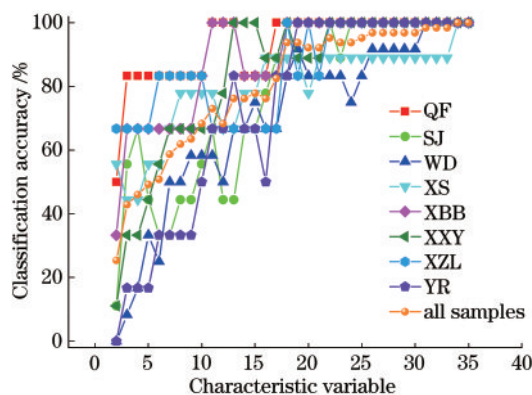


图3 8个品牌手帕纸在不同维度下分类准确率的变化趋势图
Fig. 3 Trend chart of classification accuracy of 8 brands of handkerchief paper in different dimensions

从总体准确率的变化趋势来看,在18维之前的折线比较陡峭,18维之后的折线趋于平缓。这是因为先输入的特征变量的特征值较大,包含了原始变

表 4 线性判别分析在不同维度下的准确率
Table 4 Accuracy of LDA in different dimensions

Dimension	Precision/%								Overall accuracy/%
	QF	SJ	WD	XS	XBB	XXY	XZL	YR	
2	50.0	11.1	0	55.6	33.3	11.1	66.7	0	25.4
3	83.3	55.6	8.3	44.4	66.7	33.3	66.7	16.7	42.9
4	83.3	66.7	16.7	44.4	66.7	33.3	66.7	16.7	46.0
5	83.3	44.4	33.3	55.6	66.7	44.4	66.7	16.7	49.2
6	83.3	33.3	25.0	55.6	66.7	55.6	83.3	33.3	50.8
7	83.3	33.3	50.0	66.7	66.7	66.7	83.3	33.3	58.7
8	83.3	44.4	50.0	77.8	66.7	66.7	83.3	33.3	61.9
9	83.3	44.4	58.3	77.8	66.7	66.7	83.3	33.3	63.5
10	83.3	55.6	58.3	77.8	83.3	66.7	83.3	50.0	68.3
11	100	66.7	58.3	77.8	100	66.7	66.7	66.7	73.0
12	100	44.4	50.0	66.7	100	77.8	66.7	66.7	68.3
13	100	44.4	66.7	66.7	100	100	66.7	83.3	76.2
14	83.3	66.7	66.7	77.8	83.3	100	66.7	66.7	76.2
15	83.3	66.7	75.0	77.8	83.3	100	66.7	66.7	77.8
16	83.3	77.8	66.7	88.9	83.3	88.9	66.7	50.0	76.2
17	100	88.9	66.7	88.9	83.3	88.9	66.7	83.3	82.5
18	100	100	83.3	88.9	100	100	100	83.3	93.7
19	100	100	91.7	88.9	100	88.9	83.3	100	93.7
20	100	100	83.3	77.8	100	88.9	100	100	92.1
21	100	100	83.3	88.9	100	88.9	83.3	100	92.1
22	100	100	83.3	88.9	100	100	100	100	95.2
23	100	88.9	83.3	88.9	100	100	100	100	93.7
24	100	100	75.0	88.9	100	100	100	100	93.7
25	100	100	83.3	88.9	100	100	100	100	95.2
26	100	100	91.7	88.9	100	100	100	100	96.8
27	100	100	91.7	88.9	100	100	100	100	96.8
28	100	100	91.7	88.9	100	100	100	100	96.8
29	100	100	91.7	88.9	100	100	100	100	96.8
30	100	100	91.7	88.9	100	100	100	100	96.8
31	100	100	100	88.9	100	100	100	100	98.4
32	100	100	100	88.9	100	100	100	100	98.4
33	100	100	100	88.9	100	100	100	100	98.4
34	100	100	100	100	100	100	100	100	100

Notes: omit the 35th dimension data.

量的大部分信息,对分类模型影响的相对权重较大,而后输入的特征变量的特征值相对较小,方差贡献率较小,对模型分类结果的影响不大。同时,还可以发现 12 维和 16 维下总体准确率出现了降低的情况,这是由于新特征变量的输入虽然可使一种或多种品牌手帕纸的分类准确率升高,但却会导致另一个或多个品牌手帕纸的分类准确率下降。

随着特征变量的输入,分类的总体准确率呈现

上升趋势,且在 34 维下实现了 100% 的精准分类,相比于原始的 911 个变量的判别分析,计算量缩短为 1/27,分类准确率提升至 100%。这说明采用主成分分析对融合后的光谱数据进行处理可以有效消除噪声和异常数据的干扰,利用主成分分析可对线性判别分析模型实现高效快速优化。

3.5 验证分析

为了检验分类模型的重复性,将心相印手帕纸

14 个系列 126 个样本的红外光谱指纹区与拉曼光谱 1500~200 cm^{-1} 波段融合后的数据进行标准化处理, 然后利用主成分分析对 126 个标准化样本进行降维, 根据降维后得到的主成分建立线性判别分析分类模型。主成分需满足特征根大于 1, 累计方差贡献率大于 85%。按照上述建模思路得到了不同维度下 14 个系列心相印手帕纸的分类准确率, 如图 4 所示。

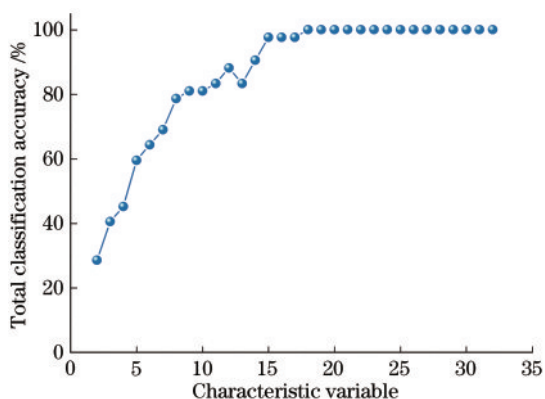


图 4 14 个系列心相印手帕纸在不同维度下的分类准确率
Fig. 4 Classification accuracy of 14 series of Xinxiangyin handkerchiefs in different dimensions

根据主成分分析保留得分的条件, 保留 32 个主成分作为线性判别分析的特征变量, 得到不同维度下线性判别分析分类模型的准确率。从图 4 可以看出, 第 18 维特征变量的输入刚好使分类模型的准确率达到 100%。这说明优化后的分类模型对一个品牌不同系列的手帕纸同样适用, 同时该结果验证了优化后分类模型的有效性。

3.6 案例分析

某公安机关送来两份手帕纸检材, 要求对手帕纸的来源进行认定。采用上述分类模型对手帕纸检材与已知手帕纸进行同一认定, 表 5 是该检材与 8 个品牌手帕纸样本的比对结果。由表 5 可以看出, 检材 1 (JC1) 和检材 2 (JC2) 的品牌分别是心相印和维达。该结果在一定程度上为公安机关侦破案件提供了线索和方向。

表 5 检材与样本的比对结果

Table 5 Comparison results of material and sample

Material	QF	SJ	WD	XS	XBB	XXY	XZL	YR
JC1						✓		
JC2			✓					

基于红外光谱和拉曼光谱数据融合的思路, 8 种品牌手帕纸鉴别分类的效果得到了明显改善, 主成分分析后的降维数据提升了线性判别分析的运

算速度和准确率, 准确率可达 100%。针对心相印品牌 14 个系列的手帕纸样本, 本文以更少的主成分实现了预期的分类结果, 从正面验证了基于光谱数据融合技术对手帕纸种类进行判别分析的优越性, 并说明了该技术在实际中具有重要应用。

4 结 论

本文基于红外光谱与拉曼光谱数据的融合, 对不同品牌和相同品牌不同系列的手帕纸样本进行判别分类, 结合主成分分析的降维优化, 实现了 8 种品牌手帕纸以及心相印手帕纸 14 个系列样本的精确区分, 准确率都为 100%。直接对红外光谱指纹区和拉曼光谱 1500~200 cm^{-1} 波段的数据建立 LDA 分类模型, 分类结果并不理想。这是因为手帕纸的化学成分比较单调, 红外光谱和拉曼光谱响应的特征峰较少。利用红外光谱与拉曼光谱互补的特点对光谱数据进行融合, 融合数据包含了更多的样本特征, 可以有效提高分类模型的准确率。同时, 结合主成分分析, 模型分类结果比较理想。

本文将红外光谱与拉曼光谱数据进行融合, 结合数据挖掘技术有效地对手帕纸样本进行了判别分析, 为手帕纸的来源认定提供了一种新思路。目前, 光谱数据融合技术在法庭科学领域尚无公开报道。该技术对其他物证的鉴定也具有一定的参考价值, 因此下一步拟利用光谱数据融合技术对法庭科学领域常见的物证开展分类鉴别研究, 以期为有效鉴别未知检材提供一种新方法。

参 考 文 献

- [1] He X L, Chen L B, Wang J F, et al. Raman spectroscopy analysis of plastic steel window based on K nearest neighbors algorithm [J]. Laser & Optoelectronics Progress, 2018, 55(5): 053001.
何欣龙, 陈利波, 王继芬, 等. 基于 K 近邻算法的塑钢窗拉曼光谱分析 [J]. 激光与光电子学进展, 2018, 55(5): 053001.
- [2] Zhang T L, Wu S, Tang H S, et al. Progress of chemometrics in laser-induced breakdown spectroscopic analysis [J]. Chinese Journal of Analytical Chemistry, 2015, 43(6): 939-948.
张天龙, 吴珊, 汤宏胜, 等. 化学计量学在激光诱导击穿光谱分析中的研究进展 [J]. 分析化学, 2015, 43(6): 939-948.
- [3] Elbashar Y H, Abd El-Ghany H A. Optical

- spectroscopic analysis of Fe₂O₃ doped CuO containing phosphate glass[J]. *Optical and Quantum Electronics*, 2017, 49(9): 1-13.
- [4] Wang X L, Liu W, Yu Y Y, et al. Operando NMR spectroscopic analysis of proton transfer in heterogeneous photocatalytic reactions [J]. *Nature Communications*, 2016, 7: 11918.
- [5] Zhang Y, Yang J R, Li B, et al. Density, viscosity, and spectroscopic and computational analyses for hydrogen bonding interaction of 1, 2-propylenediamine and ethylene glycol mixtures [J]. *Journal of Molecular Liquids*, 2020, 302: 112443.
- [6] El-Azab A S, Jalaja K, Abdel-Aziz A A M, et al. Spectroscopic analysis (FT-IR, FT-Raman and NMR) and molecular docking study of ethyl 2-(4-oxo-3-phenethyl-3, 4-dihydroquinazolin-2-ylthio) -acetate [J]. *Journal of Molecular Structure*, 2016, 1119: 451-461.
- [7] Finazzi G, Allorent G, Seydoux C, et al. Global spectroscopic analysis to study the regulation of the proton motive force in photosynthetic organisms[J]. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 2018, 1859: e35.
- [8] Bouazizi H, Mabrouk A, Braiek M B, et al. New conjugated organic matrix-carbon nanotube functionalization: DFT modeling and spectroscopic analysis [J]. *Journal of Physics and Chemistry of Solids*, 2020, 136: 109131.
- [9] Yin A, Chen T S. Quantitative fluorescence resonance energy transfer measurement based on spectral unmixing [J]. *Chinese Journal of Lasers*, 2020, 47(2): 0207009.
尹傲, 陈同生. 基于光谱分离的定量荧光共振能量转移检测[J]. *中国激光*, 2020, 47(2): 0207009.
- [10] Huang Y, Zhao N J, Meng D S, et al. Analysis on three-dimensional fluorescence spectra of PAHs in soil using nonsmooth non-negative matrix factorization [J]. *Chinese Journal of Lasers*, 2020, 47(10): 1011002.
黄尧, 赵南京, 孟德硕, 等. 非平滑非负矩阵分解解析土壤多环芳烃三维荧光光谱[J]. *中国激光*, 2020, 47(10): 1011002.
- [11] Krishnamoorthy C, Prakasarao A, Srinivasan V, et al. Monitoring of breast cancer patients under pre and post treated conditions using Raman spectroscopic analysis of blood plasma[J]. *Vibrational Spectroscopy*, 2019, 105: 102982.
- [12] Martel C, Tsutsumi T, Cément V, et al. Diagnosis of idiopathic amyotrophic lateral sclerosis using Fourier-transform infrared spectroscopic analysis of patient-derived skin [J]. *The Analyst*, 2020, 145(10): 3678-3685.
- [13] Daly C A, Streacker L M, Sun Y C, et al. Decomposition of the experimental Raman and infrared spectra of acidic water into proton, special pair, and counterion contributions[J]. *The Journal of Physical Chemistry Letters*, 2017, 8(21): 5246-5252.
- [14] Frost R L, Dickfos M J. Raman and infrared spectroscopic study of the anhydrous carbonate minerals shortite and barytocalcite [J]. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2008, 71(1): 143-146.
- [15] Wojciechowski P M, Michalska D. Theoretical Raman and infrared spectra, and vibrational assignment for para-halogenoanilines: DFT study [J]. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2007, 68(3): 948-955.
- [16] 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.
- [17] He X L, Wang J F, Wu F L, et al. Identification of the infrared spectra of tire rubber based on chemometrics [J]. *Journal of Analytical Science*, 2019, 35(3): 357-361.
何欣龙, 王继芬, 吴福璐, 等. 基于化学计量学的橡胶颗粒红外光谱分析鉴别[J]. *分析科学学报*, 2019, 35(3): 357-361.
- [18] Chen D, Luo W X, Huang Z X, et al. Adulterated milk powder diagnosis method based on multi-spectra fusion [J]. *Nanotechnology and Precision Engineering*, 2017, 15(5): 384-388.
陈达, 骆文欣, 黄志轩, 等. 基于多光谱融合的奶粉掺假诊断方法[J]. *纳米技术与精密工程*, 2017, 15(5): 384-388.
- [19] Xu W J, Wu Z C, Zhu X P, et al. Classification and discrimination of Martian-related minerals using spectral fusion methods [J]. *Spectroscopy and Spectral Analysis*, 2018, 38(6): 1926-1932.
徐伟杰, 武中臣, 朱香平, 等. 基于光谱融合的火星表面相关矿物分类方法研究[J]. *光谱学与光谱分析*, 2018, 38(6): 1926-1932.
- [20] Zhou K P, Bai X F, Bi W H. Detection of chemical

- oxygen demand in water based on multi-spectral fusion of ultraviolet and fluorescence [J]. *Laser & Optoelectronics Progress*, 2018, 55(11): 113003.
- 周昆鹏, 白旭芳, 毕卫红. 基于紫外-荧光多光谱融合的水质化学需氧量检测[J]. *激光与光电子学进展*, 2018, 55(11): 113003.
- [21] Xiang L L, Li M H, Li J M, et al. Determination of wine original regions using information fusion of NIR and MIR spectroscopy [J]. *Spectroscopy and Spectral Analysis*, 2014, 34(10): 2662-2666.
- 向伶俐, 李梦华, 李景明, 等. 近、中红外光谱法融合判定葡萄酒产地[J]. *光谱学与光谱分析*, 2014, 34(10): 2662-2666.
- [22] Lin H M, Du Z F. Some problems in comprehensive evaluation in the principal component analysis [J]. *Statistical Research*, 2013, 30(8): 25-31.
- 林海明, 杜子芳. 主成分分析综合评价应该注意的问题[J]. *统计研究*, 2013, 30(8): 25-31.
- [23] Hou S, Riley C B. Is uncorrelated linear discriminant analysis really a new method? [J]. *Chemometrics and Intelligent Laboratory Systems*, 2015, 142: 49-53.