红外光谱法结合化学计量学对快递包装纸盒的检验研究

张宇¹, 胡晓光¹, 姜红¹, 陈敏璠², 莫修浩²

(1.中国人民公安大学 侦查学院, 北京 102600; 2.北京鉴知技术有限公司, 北京 100084)

摘要:目的 研究一种红外光谱法与化学计量学相结合的方法,以对现场提取的快递包装纸盒样品进行快速检验分类。方法 利用红外光谱法对 53 个快递包装纸盒样品进行检验,依据其主要填料差异进行分类,并利用系统聚类进行分组。基于该分组,训练随机森林模型、多层感知器判别、Fisher判别3种预测模型,实现对新样品组别的分类预测。结果 53 个快递包装纸盒样品被分为 3 类,而后进一步细分为 9 组,训练得到的 3 种判别模型中的 Fisher 判别预测准确率较高。结论 该检验方法快速、无损、准确,依据化学计量学实现对快递包装纸盒样品的快速检验,为公安机关检验此类物证提供依据。

关键词: 红外光谱法; 快递包装纸盒; 系统聚类; 随机森林模型; 多层感知器判别; Fisher 判别中图分类号: TB487; D918.2 文献标识码: A 文章编号: 1001-3563(2023)21-0279-07

DOI: 10.19554/j.cnki.1001-3563.2023.21.035

Inspection of Express Cartons by Infrared Spectroscopy Combined with Chemometrics

ZHANG Yu¹, HU Xiao-guang¹, JIANG Hong¹, CHEN Min-fan², MO Xiu-hao²

(1. School of Investigation, People's Public Security University of China, Beijing 102600, China; 2. JINSP Company Limited, Beijing 100084, China)

ABSTRACT: The work aims to develop a method combining infrared spectroscopy with chemometrics to achieve rapid inspection and classification of express carton samples on site. 53 samples of express cartons were inspected by infrared spectroscopy and classified according to the differences of their main fillers, and then grouped by clustering method. According to the grouping results, random forest, multi-layered perceptron and Fisher discriminant analysis were established to realize the prediction of new sample groups. 53 express carton samples were divided into 3 categories, and then subdivided into 9 groups, and Fisher discriminant method could achieve more accurate prediction. The inspection method is rapid, non-destructive, accurate, and realizes the rapid inspection of express carton samples based on chemometrics, which provides a basis for public security organs to examine such material evidence.

KEY WORDS: infrared spectrum; express cartoon; clustering; random forest; multi-layered perceptron; Fisher discriminant analysis

在犯罪现场中经常发现各式各样的纸类物证,随着仪器分析灵敏度的提高和普遍的应用,香烟水松纸^[1]、香烟烟盒^[2-3]、一次性纸杯^[4]等非书写用纸在法庭科学发挥出的作用愈发重要。目前,检验纸张的方法主要有红外光谱法^[5]、差分拉曼光谱法^[6]、X射线

荧光光谱法^[7]等。利用光谱法对纸质物证进行分析研究,具有分析速度快、样品无损、样品状态限制少等优点,进一步地将光谱法与化学计量学相结合,建立样品的分类预测模型,可为公安机关现场物证的快速识别分析提供支撑。快递包装纸盒可用来伪装携带犯

收稿日期: 2022-12-16

罪工具,或用来藏匿相关证据等。大量一线处理的案件中,毒品、毒物、易燃易爆物品等危险品便是通过快递包装纸盒来运输。一般情况下,该类快递包装纸盒多为黄色牛皮纸,外表具有一定相似性,从外观难以提取到有用信息,不能充分发挥其作为物证或者线索的潜力。红外光谱法可对未知化合物进行定性,具有应用范围广、不受样品状态限制、样品用量少、检验快速无损等优点,便携式红外光谱仪能够满足现场快速检验此类物证的要求。

本文采用便携式傅里叶变换红外光谱仪对 53 个快递包装纸盒样品进行了检验研究,根据样品的填料成分对其进行分类,而后结合化学计量学对样品含量最多的样品进行分组并建立预测模型,取得了较好的实验结果。

1 实验

1.1 实验样品

不同发货地、不同快递公司、不同规格的样品共53个(部分样品见表1)。

1.2 实验仪器及条件

实验仪器: IT2000 便携式傅里叶变换红外光谱仪,北京鉴知技术有限公司。

实验条件:光谱范围为 $4~000 \sim 500~cm^{-1}$;扫描次数为 8 次;分辨率为 $4~cm^{-1}$; ATR 采样。

1.3 实验方法

分别剪取 5 mm×5 mm 大小的快递包装纸盒样品,将样品置于样品台上,采集红外光谱图。

1.4 重现性实验

在相同条件下,测定 38[#]样品内侧同一部位 10次,得出样品的红外光谱图。在相同条件下,测定 51[#]样品内侧的 5 个不同部位,得出样品的红外光谱图。

2 结果与分析

2.1 重现性实验结果分析

在相同条件下对 38[#]样品内侧同一部位进行 10 次平行测定,发现 38[#]样品的红外光谱图中吸收峰峰数、峰位和峰强基本一致,其红外光谱图见图 1,说明利用便携式傅里叶变换红外光谱仪检验快递包装纸盒样品是可行的。

在相同条件下对 51[#]样品内侧不同部位进行测定,发现其红外光谱图中吸收峰的峰数、峰位和峰强基本一致,其红外光谱图见图 2,说明其内侧表面材料是均匀的。

表 1 快递包装纸盒部分样品 Tab.1 Part of express cartoon samples

编号	名称	规格/mm	来源
2	中通快递	170×95×110	江苏淮安
7	顺丰快递	130×80×90	浙江杭州
13	中通快递	190×105×135	安徽合肥
14	京东物流	295×175×200	北京
19	京东物流	205×140×95	北京
20	天猫超市	275×205×305	北京
22	京东物流	230×170×135	上海
26	京东物流	205×140×80	北京
29	京东物流	300×215×175	北京
32	中通快递	250×145×180	河北沧州
38	韵达快递	135×90×45	广东东莞
39	中通快递	170×95×110	河北邢台
42	京东物流	275×205×100	北京
47	韵达快递	130×80×90	广东广州
51	京东物流	230×170×135	浙江杭州
52	中国邮政	305×220×170	浙江杭州

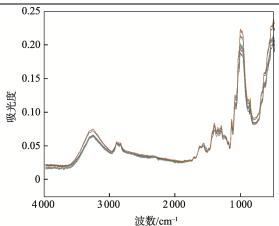


图 1 样品 38[#]重复性实验的红外光谱图 Fig.1 Infrared spectroscopy of reproducible experiments of 38[#] sample

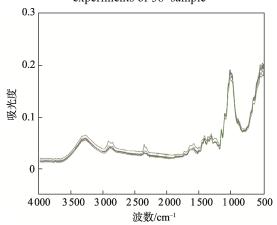


图 2 样品 51[#]重复性实验的红外光谱图 Fig.2 Infrared spectroscopy of reproducible experiments of 51[#] sample

2.2 对不同样品的分类

快递包装纸盒由不同纤维制成的纸板黏合而成。 纸盒的主要成分是纤维素,为改进纸盒的力学性能、 防潮性能, 在生产阶段通常会在纸浆中加入一些填 料,常见的填料有碳酸钙、硫酸钡。碳酸钙可提升纸 盒的平滑度,增强纸盒的实用性;硫酸钡是用于造纸 的填料,可起到提高纸张白度、平滑度等作用,其中 的钡元素作为防伪元素使用[8]。碳酸钙的红外吸收峰 为 1 714、1 425、875、725 cm⁻¹; 硫酸钡的红外吸收 峰为 1 186、1 117、1 081 cm⁻¹,且在 983、610 cm⁻¹ 处有特征吸收峰[9]。红外光谱法是依据分子振动能级 和转动能级的跃迁来分析物质官能团的相关信息,进 而确定物质结构和鉴别不同化合物。根据红外吸收峰 的峰数、峰位和峰强等信息可以确定是否含有某一成 分。根据样品中是否含有碳酸钙和硫酸钡可对样品进 行初步分类。纸盒原料中填料只占据很小比例,99% 以上为纸浆。排除纤维素的红外吸收峰对碳酸钙和硫 酸钡的干扰以及填料之间的相互干扰,选取 875 cm⁻¹ 和 610 cm⁻¹ 分别作为碳酸钙和硫酸钡的特征吸收

峰^[10], 并据此对 53 个样品进行初步分类, 分类结果见表 2。

第I类样品以 13[#]为例, 其在 875 cm⁻¹ 含有明显的 红外吸收峰, 说明其含有填料碳酸钙(见图 3)。第II 类样品以 52[#]为例, 其在 875 cm⁻¹ 没有红外吸收峰, 说明其不含有填料碳酸钙(见图 4)。第III类样品以 7[#]为例, 其在 875 cm⁻¹有明显吸收峰,在 610 cm⁻¹含有较弱吸收峰,说明其含有填料碳酸钙和硫酸钡(见图 5)。通过对红外光谱图的分析,可将 53 个样品分为三大类。

表 2 53 个快递包装纸盒样品的分类结果 Tab.2 Classification results of 53 express carton samples

分类	主要填料成分	编号
I	含碳酸钙	1~5、9~16、18~20、24~27、29~36、38~45、47~51、53
II	不含碳酸钙	6、8、17、28、52
III	含碳酸钙和硫 酸钡	7、21~23、37、46

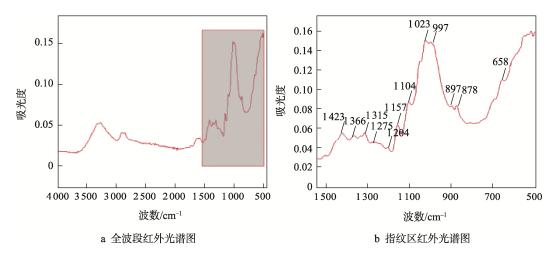


图 3 样品 13[#]的红外光谱图 Fig.3 Infrared spectroscopy of 13[#] sample

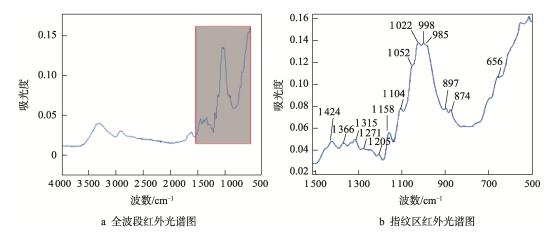
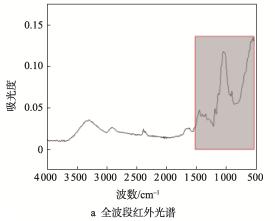


图 4 样品 52[#]的红外光谱图 Fig.4 Infrared spectroscopy of 52[#] sample



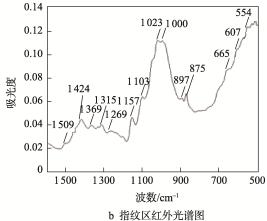


图 5 7[#]样品的红外光谱图 Fig.5 Infrared spectroscopy of 7[#] sample

2.3 系统聚类分析

依据红外光谱图,可将 53 个样品分为三大类,其中第I类样品数量较多,包含 42 个样品。为更好地对这一大类样品进行区分,采用系统聚类法对样品进行处理,聚类变量为样品吸收峰的波数,变量值为特征峰强度^[11]。对数据进行 Z-Score 标准化处理,选择瓦尔德法作为聚类方法,以平方欧氏距离作为样品亲疏程度标准,进行系统聚类分析,得到聚类谱系图,谱系图如图 6 所示。

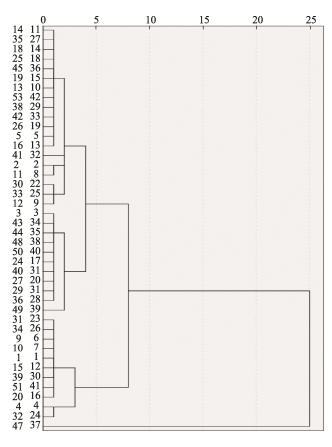


图 6 第I类样品的聚类谱系图 Fig.6 Clustering pedigree of samples in class I

以聚类数为横坐标,以聚合系数为纵坐标,绘制 折线图,如图 7 所示。当类别数为 9 时,折线的下降 趋缓,因此采用系统聚类,将第I类样品分为 9 组是 最为准确合理的。

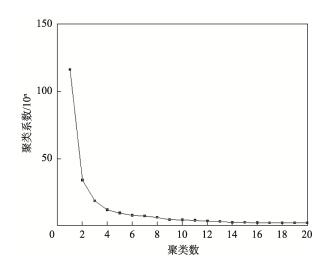


图 7 聚合系数折线图 Fig.7 Line chart of clustering coefficients

2.4 二阶聚类与 K-均值聚类验证

为检验系统聚类结果的准确性,利用二阶聚类与 K-均值聚类对第I类的 42 个快递包装纸盒样品的红外光谱数据进行验证。使用二阶聚类检验第I类样品的 分类成效,证明将第I类样品分为 9 组,其聚类质量良好,聚类质量见图 8;使用 K-均值聚类对第I类样本进行分类,聚类数设置为 9,仅有 4 个样品分组与系统聚类结果不同,这说明在系统聚类分析中将第 I类 42 个样品分为 9 组具有一定的合理性,分组结果见表 3。快递厂商为京东物流,发货地为北京的 14[#]、19[#]、26[#]、42[#]等 4 个样品依据系统聚类也分为了同一组,证明分类具有一定可靠性。

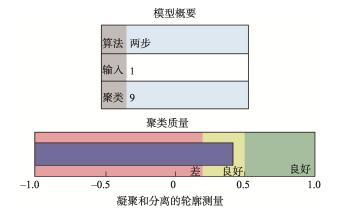


图 8 第 I 类样品聚类质量 Fig.8 Clustering quality of samples in class I

表 3 第 I 类样品分组 Tab.3 Classification results of samples in class I

分组	编号
I-1	1, 5, 9, 10, 31, 34
I-2	2、11
I-3	3、24、27、29、36、40、43、44、48~50
I-4	4、32
I-5	41
I-6	13、14、18、19、25、26、30、35、38、 42、45
I-7	47
I-8	15、16、20、39、51
I-9	12、33、53

3 构建多元判别模型

依据系统聚类的分类结果,建立多元判别模型,以实现对新样品的分类预测。构建随机森林模型、多层感知器判别、Fisher 判别 3 种模型,比较 3 种判别模型的预测准确率,选取适合新样品的判别模型。

3.1 随机森林模型

随机森林模型是一个包含多棵决策树的分类器,不同决策树之间没有关联,是快速分类的监督式聚类的分类方法。利用随机森林模型对第I类 42 个样品进行训练识别,训练集与测试集比例为 2:1,得到分类识别后的混淆矩阵见图 9,判别模型的准确率为 76.5%。

3.2 多层感知器判别

多层感知器判别神经网络,包含输入层、隐层以及输出层三部分结构,不同层之间是全连接的。通过对样品数据中提取的特征值进行线性与非线性组合,能处理复杂的多输入、多输出非线性系统,最终实现

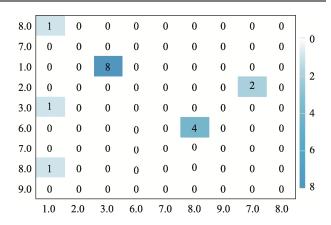


图 9 第I类样品混淆矩阵 Fig.9 Confusion matrix of samples in class I

对样品的有效判别[12]。

对第I类特征峰数据进行建模分析,将 42 个样品按照 7:3 的比例分成训练集与测试集,构建多层感知器判别模型中,得到如图 10 所示的 ROC 曲线。ROC 曲线下的面积越大越符合理想模型。由图 10 可见,该模型能够实现对第I类 42 个样品的有效准确鉴别,判别模型的准确率为88.1%。

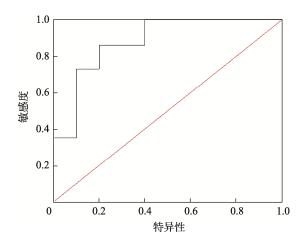


图 10 第 I 类样品多元感知神经网络 模型的 ROC 曲线 Fig.10 ROC curve of multi-Layered perceptron model of samples in class I

3.3 Fisher 判别

Fisher 判别是在方差分析的基础上建立的一种区分各个总体的线性判别法,对多维数据进行投影,降维至低维空间,以实现数据类间距离的尽可能分离,类内距离的尽可能聚拢^[13]。以第I类样品聚类结果为基础,运用 Fisher 判别建立相关模型,以达到对新数据类别变量取值进行预测的目的。利用 Fisher 判别分析,共建立 8 个判别函数。判别函数 1~8 对类别变量取值的预测效果不同。如表 4 所示,判别函数 1~3 特征值均大于 1,累计百分比为 96.7%,其余 5 个判

别函数的解释方差能力较低,仅在0至1.6%之间。 结果说明,前3个判别函数明显优于其他5个判别 函数。

表 4 判别函数特征值 Tab.4 Eigenvalues of discriminant functions

函数	特征值	方差/%	累积/%	正则相关性
1	43.370	75.4	75.4	0.989
2	8.182	14.2	89.7	0.944
3	4.061	7.1	96.7	0.896
4	0.931	1.6	98.4	0.694
5	0.477	0.8	99.2	0.568
6	0.344	0.6	99.8	0.506
7	0.121	0.2	100.0	0.328
8	0.006	0.0	100.0	0.077

为了验证判别函数 1、函数 2 和函数 3 的预测效果,使用 Wilks 统计量(Wilks Lambd)进行反向测度验证,见表 5。该统计量近似服从卡方分布,当概率值小于显著性水平 0.05 时,认为当前判别函数整体的判别能力较强^[14]。因此,可以选择函数 1、函数 2 和函数 3 对新样品进行测量,其判别模型的准确率为 95.2%。

表 5 Wilks 的 Lambda Tab.5 Wilks' Lambda

函数检验	Wilks Lambda	卡方	自由度	显著性
1~8	0.000	268.290	96	0.000
2~8	0.005	156.409	77	0.000
3~8	0.046	91.001	60	0.006
4~8	0.231	43.166	45	0.550
5~8	0.447	23.753	32	0.853
6~8	0.660	12.256	21	0.932
7~8	0.887	3.540	12	0.990
8	0.994	0.176	5	0.999

将建立的随机森林模型、多层感知器判别、Fisher 判别进行比较,随机森林模型准确率为 76.5%,多层感知器判别准确率为 88.1%, Fisher 判别准确率达到 95.2%。Fisher 判别准确率明显高于随机森林模型和 多层感知器判别模型。分析原因可能在于随机森林模型为避免样品拟合而使模型采样具有随机性,但系统聚类是将距离最小的类别进行合并,以将样品分为若干类,导致随机森林模型分类效果不佳;由于样品规模不大且样品相似度高,数据特征趋同,使得 Fisher 判别的分类效果最终优于多层感知器判别模型的分类效果。

4 结语

本文建立了一种利用红外光谱法结合化学计量学快速分析快递包装纸盒的方法,此方法快速、无损、高效。快递包装纸盒的主要成分为纤维素,依据填料成分差异将样品分为三大类。通过系统聚类分析,将样品数量最多的第I类样品分为了9组,同时通过二阶聚类与K-均值聚类验证系统聚类的分类结果。依据系统聚类,建立了不同的判别模型,最终Fisher判别模型能够实现对新样品类别较准确的预测。

参考文献:

- [1] 付钧泽,姜红,陈煜太,等. 傅里叶变换红外光谱法 检验香烟水松纸的研究[J]. 中华纸业,2017,38(20): 43-48
 - FU Jun-ze, JIANG Hong, CHEN Yu-tai, et al. A Study on Test of Tipping Paper with Fourier Transform Infrared Spectroscopy[J]. China Pulp & Paper Industry, 2017, 38(20): 43-48.
- [2] 陆润洲,姜红,满吉,等. X 射线荧光光谱结合差分拉 曼光谱对烟盒的分类研究[J]. 激光与光电子学进展, 2022, 59(5): 19-27.
 - LU Run-zhou, JIANG Hong, MAN Ji, et al. Classification of Cigarette Cases by X-Ray Fluorescence Spectrometry Combined with Differential Raman Spectroscopy[J]. Laser & Optoelectronics Progress, 2022, 59(5): 19-27.
- [3] 姜红, 陆润洲, 段斌, 等. 基于差分拉曼光谱对烟盒物证检验[J]. 包装工程, 2021, 42(21): 79-85.

 JIANG Hong, LU Run-zhou, DUAN Bin, et al. Cigarette
 Case Evidence Based on Differential Raman Spectroscopy[J]. Packaging Engineering, 2021, 42(21): 79-85.
- [4] 陈壮,姜红,李春宇,等. X 射线荧光光谱法检验一次性纸杯的研究[J]. 中华纸业,2018,39(22):32-36.
 CHEN Zhuang, JIANG Hong, LI Chun-yu, et al. A Study on Disposable Paper Cups Tested by X-Ray Fluorescence Spectroscopy[J]. China Pulp & Paper Industry, 2018, 39(22): 32-36.
- [5] 李红,姜红,鞠晨阳,等.利用红外光谱技术检验一次性筷子塑料包装袋[J]. 红外与激光工程,2018,47(7):143-148.
 - LI Hong, JIANG Hong, JU Chen-yang, et al. Inspection of Disposable Chopsticks Plastic Packing Bags by Infrared Spectroscopy[J]. Infrared and Laser Engineering, 2018, 47(7): 143-148.
- [6] 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.
- [7] 姜红, 鞠晨阳, 务瑞杰, 等. 聚类分析法的塑料饮料 瓶光谱分析 [J]. 红外与激光工程, 2018, 47(8): 0823002.
 - JIANG Hong, JU Chen-yang, WU Rui-jie, et al. Spectral Analysis of Plastic Beverage Bottles Based on Cluster Analysis[J]. Infrared and Laser Engineering, 2018, 47(8): 0823002.
- [8] 李玲玲, 柴欣生, 田迎新, 等. 样品预处理结合红外光谱分析鉴定纸张中的硫酸钡[J]. 中国造纸, 2015, 34(4): 39-42.
 - LI Ling-ling, CHAI Xin-sheng, TIAN Ying-xin, et al. A New Sample Pretreatment Method of Infrared Spectrum Analysis for Barium Sulfate Identification in Paper[J]. China Pulp & Paper, 2015, 34(4): 39-42.
- [9] TONG Na, ZHU Chang-jun, ZHANG Chong-hui, et al. Study on Raman Spectra of Aliphatic Polyamide Fibers[J]. Optik, 2016, 127(1): 21-24.
- [10] 朱晓晗,姜红.常用纸张填料的光谱分析[J].中华纸业,2020,41(14):27-30.
 - ZHU Xiao-han, JIANG Hong. Spectral Analysis of Common Paper Fillers[J]. China Pulp & Paper Industry, 2020, 41(14): 27-30.

- [11] 李传召, 姜红, 陈潼, 等. 差分拉曼光谱对透明塑料胶带的分类研究[J]. 上海塑料, 2022, 50(5): 61-64. LI Chuan-zhao, JIANG Hong, CHEN Tong, et al. Classification of Transparent Plastic Tape Based on Differential Raman Spectroscopy[J]. Shanghai Plastics, 2022, 50(5): 61-64.
- [12] 李锦,姜红,杨俊,等. 差分拉曼光谱结合人工神经 网络对药品塑料包装瓶的分类研究[J]. 塑料工业, 2022,50(8): 101-107. LI Jin, JIANG Hong, YANG Jun, et al. Classification of Plastic Pharmaceutical Packaging Bottles Based on Differential Raman Spectroscopy and ANN[J]. China
- [13] ZHANG Jin, JIANG Hong, DUAN Bin, et al. A Rapid and Nondestructive Approach for Forensic Identification of Cigarette Inner Liner Papers Using Shift-Excitation Raman Difference Spectroscopy and Chemometrics[J]. Journal of Forensic Sciences, 2021, 66(6): 2180-2189.

Plastics Industry, 2022, 50(8): 101-107.

- [14] 付钧泽,姜红,李意,等. XRF 结合化学计量学检验 香烟烟灰物证[J]. 激光与光电子学进展,2021,58(6):411-418.
 - FU Jun-ze, JIANG Hong, LI Yi, et al. Examination of Cigarette Ash Evidence by XRF Combined with Chemometrics[J]. Laser & Optoelectronics Progress, 2021, 58(6): 411-418.

责任编辑:曾钰婵