基于神经网络的塑料打包带高光谱模式识别

韩林杰¹,姜红^{2*},田陆川¹,赵静远³,刘业林³,牛奕³,张永强³

(1.中国人民公安大学 侦查学院,北京 100038; 2.甘肃警察职业学院 刑事侦查系,兰州 730046;3.江苏双利合谱科技有限公司,江苏 无锡 214000)

摘要:目的 建立一种快速、准确、无损的塑料打包带的检验及分类方法。方法 利用高光谱在波长为 350~990 nm 的条件下采集 52 个不同来源的塑料打包带样品的高光谱数据,并对样品进行 Savitzky-Golay 平滑处理,同时结合主成分分析对样品进行降维。将提取到的主成分进行 K-Means 聚类,以聚类结果为 依据建立径向基函数神经网络(RBFNN)与 BP 神经网络模型(BPNN)。结果 打包带样品的高光谱谱 图在 400~500 nm、600~700 nm 处有较大区别。实验共提取了 5 个初始特征值大于 1 的主成分,可以解 释 96.633%的原始数据。通过 K-means 聚类将塑料打包带样品分为 6 类, Calinski-Harabasz 指数为 28.76, RBFNN 分类准确率为 86.7%; BPNN 分类准确率为 98.1%, BPNN 的分类效果更好。结论 研究表明神 经网络在高光谱谱图分类处理上具有较高的准确度,同时也验证了高光谱在区分检验塑料打包带类物证 的可行性与科学性,为公安机关提供了一种新的检验方法。

关键词: 高光谱; 塑料打包带; 神经网络; 模式识别

中图分类号: TB487; O433.4 文献标志码: A 文章编号: 1001-3563(2024)05-0240-07 DOI: 10.19554/j.cnki.1001-3563.2024.05.029

Hyperspectral Pattern Recognition of Plastic Packaging Tape Based on Neural Network

HAN Linjie¹, JIANG Hong^{2*}, TIAN Luchuan¹, ZHAO Jingyuan³, LIU Yelin³, NIU Yi³, ZHANG Yongqiang³

Investigation Institute, People's Public Security University of China, Beijing 100038, China;
 Criminal Investigation Department, Gansu Police Vocational College, Lanzhou 730046, China;

3. Jiangsu Dualix Spectral Imaging Co., Ltd., Jiangsu Wuxi 214000, China)

ABSTRACT: The work aims to establish a fast, accurate, and non-destructive inspection and classification method for plastic packaging tapes. 52 samples of plastic packaging tape were collected from different sources through hyperspectral data in the wavelength range of 350-990 nm, and the samples were smoothed with Savitzky Golay. Principal component analysis was also used to reduce the dimensionality of the samples. *K*-Means clustering was conducted on the extracted principal components, and a radial basis function neural network (RBFNN) and BP neural network model (BPNN) was established based on the clustering results. There were significant differences in the hyperspectral spectra of the packaged sample at 400-500 nm and 600-700 nm. A total of 5 principal components with initial feature values greater than 1 were extracted in the experiment, which could explain 96.633% of the original data. The plastic packaging tape samples were

收稿日期: 2023-05-10

基金项目:国家重点研发计划项目(2019YFF0303405); 食品药品安全防控山西省重点实验室基金资助(202204010931006) *通信作者

clustered into 6 categories, with a Calinski Harabasz index of 28.76 for *K*-means and a classification accuracy of 86.7% for RBFNN. The classification accuracy of BPNN was 98.1%. BPNN had better classification performance. Research has shown that neural network has high accuracy in the classification and processing of hyperspectral spectra, and it has also verified the feasibility and scientificity of hyperspectral recognition in the detection of plastic packaging tape type evidence, providing a new inspection method for public security organs.

KEY WORDS: hyperspectral; plastic packaging tape; neural network; pattern recognition

塑料打包带在各类案件现场中较为常见,因其具 有质轻、耐磨、耐腐蚀等优势,犯罪分子在包装、运 输凶器等过程中经常使用塑料打包带^[1]。因此,系统 地分析塑料打包带样品可以为侦查人员提供线索,为 缩短侦查时间、缩小侦查范围提供帮助。

塑料打包带常用聚氯乙烯(Polyvinyl Chloride, PVC)、聚乙烯(Polyethylene, PE)、聚丙烯 (Polypropylene, PP)、聚对苯二甲酸乙二醇酯 (Polyethylene Glycol Terephthalate, PET)作为主要 成分^[2],其中以 PP 和 PET 最为常见。为降低成本、 改善性能,生产厂家常在生产过程中加入重质碳酸 钙、滑石粉、高岭土、钙粉等填料和白油(液体石蜡)、 石蜡、邻苯二甲酸二辛酯、低相对分子质量聚乙烯(聚 乙烯蜡)及硬脂酸等助剂。

目前法庭科学领域检测塑料打包带类物证的方 法有:红外光谱法、拉曼光谱法^[3]、X射线荧光光谱 法、差示扫描量热法^[4]、扫描电镜能谱法等。高光谱 成像技术融合了图像技术与光谱信息技术,既可以获 得目标的二维位置信息又可以获得一维的光谱信息, 在光谱分辨率与波段范围上有更好的表现。作为一种 新兴、无损、准确的分析方法,高光谱成像技术已在 食品安全^[5],医学研究^[6],航空航天^[7]等领域有广泛 应用,近些年也在血迹检验^[8-9]、文件检验^[10-11]、生 物物证^[12]等领域有所进展。在塑料打包带生产过程 中,由于主要成分与填料的不同,塑料打包带在高光 谱谱图中会存在一定的差异,这为后续分析检验塑料 打包带样品提供了科学依据。

以往检验塑料打包带的文献通常用谱图峰值与标 准特征峰对比或元素含量对比的方法对其进行分类,如 《拉曼光谱法检验塑料打包带(绳)的研究》^[3]《红外 光谱结合 X 射线光光谱检验塑料打包带(绳)的研 究》^[1]这 2 篇文章,没有使用化学计量学加以分析; 而在《X 射线荧光光谱结合多元统计分析塑料打包带 (绳)》^[2]—文中,作者使用了 Fisher 判别,但相关 系数为 0.993,没有达到 100%的正确率。本实验在上 述方法的基础上进行创新升级,将高光谱与神经网络 相结合对塑料打包带进行分类。

1 实验部分

1.1 实验仪器及条件

主要实验仪器: GaiaFiled Pro-V10E(江苏双利

合谱);测试方式为5×5排列,从上到下依次扫描。

主要实验条件: 光源为卤素灯; 波长为 400~1 000 nm, 分辨率为 1 936×1 456, CPU 为 Intel(R) Core(TM) i5-10400 CPU @ 2.90 GHz, RAM 为 16 G, 硬盘为 256 G。

1.2 实验样本及前处理

不同品牌、不同来源、不同规格的样品共 52 个 (样品表略)。将每个样品剪成长约1 cm,用酒精擦 拭干净,晾干待测。

1.3 谱图采集与预处理

首先将样品按长度为1 cm 进行裁剪,再用酒精 棉球擦拭,在环境光背景下对样品进行扫描测试。采 用 ENVI5.3 光谱图像处理软件处理高光谱图像,采集 不同特征点的光谱曲线,再利用 Savitzky-Golay 平滑 处理提取得到光谱图。

1.4 数据处理方法

1.4.1 主成分分析法(PCA)

主成分分析(PCA)是最常用的线性降维方法, 它通过投影的方式将高维的数据映射到低维的空 间中,同时希望提取的主成分可以最大地表示原始 数据。当数据量庞大且各项数据可能存在线性重叠 的情况下,需要对数据降维再进行处理。在主成分 分析的过程中,当主成分和累计方差百分比过低 时,主成分不足以解释原始光谱数据,因此一般要 求提取的主成分特征值大于 1 且累计方差率高于 85%。

1.4.2 K-means 聚类分析

K-means 算法是经典的无监督聚类算法之一,其 原理是在给定 *K* 个初始类簇中心点的情况下,把每个 样本点分到距离最近的簇中心点所代表的簇中,最终 对划分工作完成后的数据进行迭代运算,从而获得一 个最优的聚类结果。样本点与簇中心点的距离计算方 法采用欧氏距离,具体表达式如式(1)所示。其中, *X*_i表示第*i* 个对象; *C*_j表示第*j* 个聚类中心; *X*_i表示 第*i* 个对象的第*t* 个属性; *C*_{jt}表示第*j* 个聚类中心的 第*t* 个属性。

$$dis(X_{i}, C_{j}) = \sqrt{\sum_{t=1}^{m} (X_{it} - C_{jt})^{2}}$$
(1)

1.4.3 径向基函数神经网络

径向基函数神经网络(RBFNN)是三层前反馈 神经网络,其包括输入层、隐层、输出层。当 RBF 的中心点确定以后,可以确定输入矢量到隐层及隐层 到输出层的映射关系,输入层到隐藏层的神经元之间 的权重全部为1。隐层是使用径向基函数作为激活函 数的神经元,其作用是把向量从低维度映射到高维 度,使得低维度不可分的情况到高纬度变得可分。 RBFNN可以处理系统内难以解析的规律性,具有良 好的泛化能力,并有很快的学习收敛速度,是一种局 部逼近网络^[13-14]。

1.4.4 BP 神经网络

BP 神经网络(BPNN)是一种多层前反馈神经网络。其基本思想是梯度下降法,类似于人类的神经元。 输入层接收数据,输出层输出数据,前一层神经元 连接到下一层神经元,收集上一层神经元传递来的 信息,经过"激活"把值传递给下一层,利用梯度搜 索技术,使网络的实际输出值和期望输出值的误差 均方差为最小。与 RBFNN 不同的是,BPNN 的隐节 点采用输入模式与权值向量的内积作为激活函数的 自变量。此外,各参数(权值、阈值)对网络输出的 影响力相同,且对任何输出都有影响,是一种全局逼 近网络^[15-17]。

1.5 实验流程

实验具体流程如图1所示。

2 结果与讨论

2.1 光谱预处理结果

样品的原始光谱图如图 2 所示。由图 2 可以发现,样品在波长为 400~500 nm,600~700 nm 处有显 著差异,说明样品的主要成分有所区别。为了更好 地增强审视效果,运用 Savitzky-Golay 算法对光谱 数据进行平滑处理,减少噪声干扰,在保留谱图原 始信息的同时可以避免数据冗杂,降低神经网络模型的分类准确度与训练效率。平滑处理的部分结果 如图 3 所示。







图 2 全部样品高光谱的谱图 Fig.2 Hyperspectral spectra of all samples





\cdot 243 \cdot

2.2 主成分分析

高光谱图像因携带有大量的数据信息增强了其 检测能力,但冗余的数据也增加了其分类难度。对光 谱图像进行降维处理的目的在于提取主要特征的同 时减少数据信息量,降低处理信息的时长与难度,同 时增强数据分析的精确度。

本文对 52 个样品进行主成分分析,共提取了 5 个 主成分,其主成分分析结果见表 1。由表 1 可以发现, 样品的累计方差贡献率为 96.633%,降维效果较好。

表 1 样品的主成分分析结果摘要 Tab.1 Summary of principal component analysis results

成分	初始特征值方差			提取载荷平方和			
	总计	方差 百分比	累积/%	总计	方差 百分比	累积/%	
1	25.089	48.249	48.249	25.089	48.249	48.249	
2	13.605	26.164	74.413	13.605	26.164	74.413	
3	5.639	10.845	85.258	5.639	10.845	85.258	
4	4.815	9.260	94.518	4.815	9.260	94.518	
5	1.100	2.115	96.633	1.100	2.115	96.633	

2.3 K-means 聚类分析

为了使聚类效果最好,本文将降维得到的 5 个主 成分作为新的变量带入分析,采用 Calinski-Harabasz 准则判断最佳 K 值。CH 指数的本质是簇间距离与簇 内距离的比值,由于整体计算方式类似于方差计算,又 被称为方差比准则。CH 值越大,代表内部协方差越小, 类别之间协方差越大,可以说明聚类效果更好^[18]。为避 免 K 值设置过大导致原属于同组的样品被细分,使聚 类结果缺乏解释性,本文设置 K 值范围为 2~8,得到 塑料打包带样品不同 K 值对应的 CH 指数,如图 4 所 示。可以发现,当聚类数为 6 时,样品的 CH 值最大, 聚类效果最好,由此设定 K 值为 6。



图 4 不同 K 值下的 Calinski-Harabasz 指数 Fig.4 Calinski Harabasz index at different K values

2.4 神经网络分类

2.4.1 径向基函数神经网络

以 *K*-Means 聚类结果为依据,本次实验按照 8:2 的比例随机分配径向基函数神经网络的训练集与测 试集,其中输入向量为提取到的 5 个主成分,隐藏层 激活函数为 Softmax,输出层向量为 *K*-Means 聚类结 果,具体的分类结果如表 2 所示。由表 2 可以发现, 径向基函数神经网络(RBFNN)对样品的分类准确 率为 86.7%,准确率偏低。

表 2 RBFNN 的分类结果 Tab.2 Classification result of RBFNN

样本	र ेन् र आग							
	头侧	1	2	3	4	5	6	正确百分比/%
	1	4	0	0	0	0	0	100.0
	2	0	6	0	0	0	1	85.7
	3	0	0	3	0	0	0	100.0
训练	4	0	0	0	4	0	0	100.0
	5	0	0	0	0	4	1	80.0
	6	1	0	0	0	1	12	85.7
	总体百分比/%	13.5	16.2	8.1	10.8	13.5	37.8	89.2
	1	2	0	0	0	0	0	100.0
	2	0	4	0	0	0	0	100.0
	3	0	0	4	0	1	0	80.0
检验	4	0	0	0	0	0	0	0.0
	5	0	0	1	0	1	0	50.0
	6	0	0	0	0	0	2	100.0
	总体百分比/%	13.3	26.7	33.3	0.0	13.3	13.3	86.7

2.4.2 BP 神经网络

同样以 *K*-Means 聚类结果为依据,实验按照 8:1:1的比例随机选择 BPNN 的训练集、测试集与 验证集,采用交叉熵作为损失函数,训练算法为量化 共轭梯度法进行训练,其结果如图 5、图 6 所示。当 采用交叉熵作为损失函数时,交叉熵越低,就证明模 型的策略越接近最优策略,在迭代数为 40 时,BPNN 的交叉熵最低,模型训练完成,其具体训练过程如图 5 所示。样品的混淆矩阵(见图 6)表明样品的准确 率为 98.1%,结合 ROC 曲线(见图 7)可以看出,该 训练模型适合对塑料打包带进行分类。选择部分样品 作为测试集测试训练模型的分类正确率,得到的测试 集混淆矩阵(见图 8)表明,该模型对塑料打包带分 类的正确率为 100%。



图 5 BPNN 的训练过程 Fig.5 Training process of BPNN



图 6 样品分类混淆矩阵 Fig.6 Sample classification confusion matrix BPNN 属于全局逼近网络,具有较强的泛化能力、容错能力和高度的自适应能力,可以逼近任意的非线性关系。BPNN的优点在于能够通过学习自动提取并输出数据间的"合理规则",并自适应地将学习内容记忆于网络的权值中,适合时可以将学习成果应用于新知识,并且 BPNN 在局部的或者部分的神经元受到破坏后仍然能正常工作。相较于 RBFNN, BPNN在解决具有相同精度要求问题时的结构要更加简单,运算大量样本时能有效减少运算量。在实验得到低维数据的情况下,采用 BPNN 更为合适。经过验证,BPNN检验样品的准确率更高。综上所述,BPNN可以较好地对塑料打包带进行分类。



图 7 样品 ROC 曲线 Fig.7 ROC curve of sample

输出类	1	1 12.5%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
	2	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	NaN% NaN%
	3	0 0.0%	0 0.0%	3 37.5%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
	4	0 0.0%	0 0.0%	0 0.0%	1 12.5%	0 0.0%	0 0.0%	100% 0.0%
	5	0 0.0%	0 0.0%	0 0.0%	0 0.0%	2 25.0%	0 0.0%	100% 0.0%
	6	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 12.5%	100% 0.0%
		100% 0.0%	NaN% NaN%	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%
		1	2	3		5	6	

图 8 测试集混淆矩阵 Fig.8 Test set confusion matrix 注: NaN%表示非确定数值或数值无法被计算。

3 结语

本实验利用高光谱技术对 52 个塑料打包带进行 了检验,并将样品谱图进行 Savitzky-Golay 平滑处理。 样品在 400~500 nm、600~700 nm 的谱图差异,可以 作为样品分类的主要依据。采用主成分分析法对样品 进行降维,将提取到的5个主成分进行 K-Means 聚类, 比较不同 K 值对应的 CH 指数,选择最大值,确认最 佳聚类数为6,以聚类结果为依据建立径向基函数神 经网络(RBFNN)与 BP 神经网络模型(BPNN)。其 中, BPNN 分类准确率为98.1%,选择部分样品对模 型进行测试,得到测试集的分类正确率为100%,可 以较好地对塑料打包带进行分类。

当今,将神经网络应用于高光谱图像检测与分类 已有不少研究成果见诸报道,但用于公安刑侦中塑料 打包带类物证的检测,属于应用创新,具有较好的应 用价值。本文验证了高光谱在区分检验塑料打包带类 物证的可行性与科学性,该方法操作简便,无损样品, 可以为公安检测工作提供新手段。

参考文献:

[1] 马泉,姜红,杨佳琦.红外光谱结合 X 射线荧光光谱
 检验塑料 打包带(绳)的研究[J].化学研究与应用,
 2019, 31(9): 1643-1648.

MA X, JIANG H, YANG J Q. Research on Inspecting the Plastic Pack Belts(Ropes)by Infrared Spectrometry Combined with X-Ray Fluorescence Spectrometry[J]. Chemical Research and Application, 2019, 31(9): 1643-1648.

 [2] 马泉,姜红,杨佳琦. X 射线荧光光谱结合多元统计 分析塑料打包带(绳)[J]. 激光与光电子学进展, 2019, 56(22): 243-247.

MA X, JIANG H, YANG J Q. Examination of Plastic Pack Belts (Ropes) via X-Ray Fluorescence Spectrometry Combined with Multivariate Statistical Analysis[J]. Laser & Optoelectronics Progress, 2019, 56(22): 243-247.

- [3] 马枭,姜红,杨佳琦,等. 拉曼光谱法检验塑料打包带(绳)的研究[J]. 上海塑料, 2018(4): 29-35.
 MA X, JIANG H, YANG J Q, et al. Research on Inspecting the Plastic Pack Belts (Ropes) by Raman Spectroscopy[J]. Shanghai Plastics, 2018(4): 29-35.
- [4] 冯计民,陶克明,孔晓明. 差示扫描量热法(DSC)对 聚乙烯、聚丙烯制品的检验[J]. 刑事技术, 1992(1):
 1-5.
 FENG J M, TAO K M, KONG X M. Inspection of

Polyethylene and Polypropylene Products by Differential

Scanning Calorimetry[J]. Criminal Technology, 1992(1): 1-5.

- [5] WU D, MENG L, YANG L, et al. Feasibility of Laser-Induced Breakdown Spectroscopy and Hyperspectral Imaging for Rapid Detection of Thiophanate-Methyl Residue on Mulberry[J]. International Journal of Molecular Sciences, 2019, 20(8): 2017- 2017.
- [6] AGGARWAL S L P, PAPAY F A. Applications of Multispectral and Hyperspectral Imaging in Dermatology[J]. Experimental Dermatology, 2022, 31(8): 1128-1135.
- [7] 郝明达, 普运伟, 周家厚, 等. 一种用于预测航空遥 感影像光谱信息的深度学习方法[J]. 遥感信息, 2022, 37(6): 123-129.
 HAO M D, PU Y W, ZHOU J H, et al. A Deep Learning Method for Predicting Spectral Information of Aerial Remote Sensing Images[J]. Remote Sensing Information,
- 2022, 37(6): 123-129.
 [8] 庄园,高树辉,谢菲,等.基于高光谱成像技术鉴别 血痕种属的实验研究[J]. 激光与光电子学进展, 2022, 59(16): 1630001.
 ZHUANG Y, GAO S H, XIE F, et al. Identifying Bloodstain Species Using Hyperspectral Imaging[J].
 Laser & Optoelectronics Progress, 2022, 59(16): 1630001.
- [9] 孙威,刘怀策,刘金坤,等. 基于偏最小二乘法的血 迹陈旧度高光谱预测研究[J]. 应用化工, 2022, 51(1): 273-276.
 SUN W, LIU H C, LIU J K, et al. Age Imaging by Hyperspectral Estimation of Bloodstains Technology Based on Partial Least Squares Regression[J]. Applied
- Chemical Industry, 2022, 51(1): 273-276.
 [10] 李云鹏, 代雪晶, 王猛, 等. 隐性字迹的快速光谱显现与高光谱分类技术研究[J]. 光谱学与光谱分析, 2021, 41(11): 3524-3531.
 LI Y P, DAI X J, WANG M, et al. Study on Rapid Spectral Reappearing and Hyperspectral Classification of Invisible Writing[J]. Spectroscopy and Spectral Analysis, 2021, 41(11): 3524-3531.
- [11] 张倩,陈维娜,郝红光. 高光谱成像技术在文件检验 应用的研究综述[J]. 应用化工, 2020, 49(1): 165-170.
 ZHANG Q, CHEN W N, HAO H G. A Review of Research on Hyperspectral Imaging Technology in Document Inspection Applications[J]. Applied Chemical Industry, 2020, 49(1): 165-170.
- [12] PALLOCCI M, TREGLIA M, PASSALACQUA P, et al. Forensic Applications of Hyperspectral Imaging Technique: A Narrative Review[J]. The Medico-Legal Journal, 2022, 90(4): 216-220.
- [13] 侯伟, 王继芬, 刘怡然. 基于机器学习的人体指甲光

谱模式识别及溯源分析[J]. 激光与光电子学进展, 2022, 59(18): 463-470.

HOU W, WANG J F, LIU Y R. Spectral Pattern Recognition and Traceability Analysis of Human Fingernail Based on Machine Learning[J]. Laser & Optoelectronics Progress, 2022, 59(18): 463-470.

- [14] 潘立剑,陈蔚芳,崔榕芳,等.基于激光诱导击穿光 谱与径向基函数神经网络的铝合金定量分析[J].激光 与光电子学进展,2020,57(19):193002.
 PAN L J, CHEN W F, CUI R F, et al. Quantitative Analysis of Aluminum Alloy Based on Laser-Induced Breakdown Spectroscopy and Radial Basis Function Neural Network[J]. Laser & Optoelectronics Progress, 2020, 57(19): 193002.
 [15] 吴俊,郭大千,李果,等. 基于 CARS-BPNN 的江西
- [15] 关後, 郭大十, 李来, 等. 基于 CARS-BPNN 的江西 省土壤有机碳含量高光谱预测[J]. 中国农业科学, 2022, 55(19): 3738-3750.

WU J, GUO D Q, LI G, et al. Prediction of Soil Organic Carbon Content in Jiangxi Province by Vis-NIR Spectroscopy Based on the CARS-BPNN Model[J]. Scientia Agricultura Sinica, 2022, 55(19): 3738-3750.

- [16] 沈飞龙. 基于机器学习与光谱信息的土壤铁氧化物估 算模型研究[D]. 武汉: 华中农业大学, 2022.
 SHEN F L. Study on Soil Iron Oxide Estimation Model Based on Machine Learning and Spectral Information[D]. Wuhan: Huazhong Agricultural University, 2022.
- [17] 田安红, 付承彪, 熊黑钢, 等. BPNN 对不同人为活动 区域的盐渍土 Na+高光谱估测[J]. 水土保持研究, 2020, 27(2): 364-369.
 TIAN A H, FU C B, XIONG H G, et al. Hyperspectral Estimation of Na+ Ion in Saline Soils in Areas with Different Human Activities Using BPNN Model[J]. Research of Soil and Water Conservation, 2020, 27(2): 364-369.
- [18] WANG C Y, ZHA Q. Measuring Systemic Diversity of Chinese Universities: A Clustering-Method Approach[J]. Quality & Quantity, 2018, 52(3): 1331-1347.