

X 射线荧光光谱结合多元统计分析塑料打包带(绳)

马枭¹, 姜红^{1,2*}, 杨佳琦¹

¹ 中国人民公安大学刑事科学技术学院, 北京 100038;

² 证据科学教育部重点实验室(中国政法大学), 北京 100088

摘要 采用 X 射线荧光光谱法对 32 个塑料打包带(绳)样本进行检验, 在 80 s 的采集时间下, 所得元素种类含量最为稳定。结合多元统计方法构建函数模型, 将离差平方和法作为聚类方法, 进行系统聚类时将平方欧氏距离作为测量区间以描述样本间的亲疏程度, 将 32 个塑料打包带(绳)样本分为 4 类。通过样本之间的相关性和判别分析检验系统聚类的正确性和可靠性, 挖掘出样本元素含量之间的内在联系。通过系统聚类的方法将 32 个塑料打包带(绳)样本进行分类, 促进了模式识别技术在理化检验中的应用。

关键词 光谱学; X 射线荧光光谱法; 系统聚类; 判别分析; 塑料打包带(绳)

中图分类号 O657.34

文献标识码 A

doi: 10.3788/LOP56.223005

Examination of Plastic Pack Belts (Ropes) via X-Ray Fluorescence Spectrometry Combined with Multivariate Statistical Analysis

Ma Xiao¹, Jiang Hong^{1,2*}, Yang Jiaqi¹

¹ School of Forensic Science, People's Public Security University of China, Beijing 100038, China;

² Key Laboratory of Evidence Science (China University of Political Science and Law),
Ministry of Education, Beijing 100088, China

Abstract Herein, 32 plastic pack belt (rope) samples were examined via X-ray fluorescence spectrometry. At the acquisition time of 80 s, the elemental composition was most stable. The function model was constructed by combining the multivariate statistical method. The ward's method was selected as the clustering method. The square Euclidean distance was used as the measurement interval to describe the degree of affinity between the samples for hierarchical clustering. The 32 plastic pack belt (rope) samples were divided into four categories. The accuracy and reliability of the hierarchical clustering were tested via correlation between the samples and discriminant analysis, and the internal relationship between the contents of the sample elements was determined. In this experiment, 32 samples of plastic pack belts (ropes) were classified using the hierarchical clustering method, which promoted the application of pattern recognition technology for physical and chemical examination.

Key words spectroscopy; X-ray fluorescence spectrometry; hierarchical clustering method; discriminant analysis; plastic pack belts (ropes)

OCIS codes 300.6560; 100.5010; 160.5470

1 引言

塑料打包带(绳)具有耐腐蚀、强度高、成本低等诸多优点, 被广泛使用于日常生活中。在公安司法鉴定工作中, 塑料打包带(绳)出现的次数逐年增长, 如能通过系统性的数据处理构建检验数学模型, 将

有利于锁定侦查方向, 缩小侦查范围, 推动公安工作的开展。

对塑料物证的检验分析一直是法庭科学研究的重点。目前, 对塑料物证的检验方法主要有 X 射线荧光光谱法、红外光谱法、拉曼光谱法、裂解气相色谱法和扫描电镜/能谱法等。拉曼光谱法、红外光谱

收稿日期: 2019-04-29; 修回日期: 2019-05-17; 录用日期: 2019-05-28

基金项目: 证据科学教育部重点实验室(中国政法大学)开放基金资助课题(2018KFKT08)

* E-mail: jiangh2001@163.com

法和裂解气相色谱法主要是对样本的有机物成分进行检验^[1],但裂解气相色谱法会损坏检材。扫描电镜/能谱法能检测元素的种类及含量,但灵敏度较低。因此,本实验采用高灵敏度的 X 射线荧光光谱法对样本所含元素种类和含量进行检验^[2]。在前期的工作中,笔者已使用拉曼光谱法对塑料打包带(绳)进行了研究,并取得了一定的成果^[3]。

本实验采用 X-MET8000 Optimum 手持式 X 射线荧光光谱仪对不同来源的 32 个塑料打包带(绳)样本进行检验,通过系统聚类的方法对样本进行分类,并验证了其合理性与准确性,取得了较好的实验结果。

2 理论基础

2.1 X 射线荧光光谱法

X 射线荧光光谱法具有分析时间短、分析元素广、工作曲线线性范围宽、对检材无损且光谱干扰少等优点,被广泛应用于材料、钢铁、化工、食品等领域^[4]。通过测量一系列由样本辐射出的特征 X 射线(荧光 X 射线)的波长,即能确定元素的种类;将测得的谱线强度与标准样本进行比较,即可确定该元素的含量,由此建立了 X 射线荧光光谱(XRF)分析法。

2.2 系统聚类

系统聚类是将样本进行分类的一种统计方法,它根据样本数据计算样本之间的距离,将距离较近的样本归为同一类^[5],再在第一步的基础上,计算新类与其他样本的距离,重复计算归类,直至所有样本合并为一类^[6],最终样本被逐级分散开来。不同类别样本的距离相对较远。

2.3 显著性 Sig 值

在统计学概念中,显著性 Sig 值(P 值)是指在一个概率模型中,统计摘要与实际观测数据相同(或前者比后者更大)这一事件发生的概率^[7]。换言之,是检验零假设成立或表现更严重的可能性。如果 P 值很小,说明原假设情况发生的概率很小,若发生了,根据小概率原理,就有理由拒绝原假设。 P 值越小,拒绝原假设的理由就越充分。如果 $0.01 < P < 0.05$,则差异有统计学意义;如果 $P < 0.01$,则差异有高度统计学意义^[8]。

2.4 Fisher 判别函数

判别函数是指各个类别的判别区域确定后,可以用一些函数来表示和鉴别某个特征矢量属于哪个类别。在 Fisher 准则下,根据均值向量有显著差异

的已知总体的子样观测值,建立线性判别函数,把全部样品判入已知总体,适用于样本的识别与分类^[9]。

3 实验部分

3.1 实验仪器

本文选用的实验仪器为 X-MET8000 Optimum 手持式 X 射线荧光光谱仪(英国牛津公司生产),其电流为 60 mA,电压为 40 kV。

3.2 样本前处理

截取面积为 $2\text{ cm} \times 1.5\text{ cm}$ 的样本并进行编号,取光滑平整面用无水乙醇棉球擦拭,晾干后待测。

3.3 测量时间的优选

随机选取 14[#] 样本,进行测量时间的优选。理论分析认为,测量时间并不会影响元素含量的测量结果。但在实验过程中却发现,不同的测量时间下元素的含量处于波动状态,测量时间对元素含量的精度会产生影响。这可能是由于在较短的测量时间内,仪器采集点过少,实验中峰值的出现是多个测量数据累积的结果,如果采集点较少,就会导致峰值参数不够准确^[10]。前期实验已得出最优的测量时间范围,最终选定测量时间为 50,60,70,80,90,100 s。

3.4 仪器重现性检验

随机选取 25[#] 样本,在最优测量时间条件下对其进行检验,测量 4 次取平均值,分别计算不同元素的相对标准偏差(RSD),其计算公式为

$$R_{SD} = \frac{S}{\bar{x}} \times 100\% = \frac{\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}}{\bar{x}} \times 100\%, \quad (1)$$

式中: R_{SD} 为相对标准偏差; S 为标准偏差; \bar{x} 为平均值; n 为计算值的个数。

4 结果与讨论

4.1 测量时间的优选结果

随机选取 14[#] 样本,进行测量时间的优选,测量结果见表 1。

对测定结果进行分析后可知,当测量时间选定为 80 s 时,测定的元素种类、含量最稳定,有利于开展实验,故确定最优测量时间为 80 s。

4.2 重现性实验分析

X 射线荧光光谱仪重现性实验结果见表 2,通过对元素含量数据的分析可知,同一样本经过 4 次测量,元素的种类、含量相对稳定。相对标准偏差是

表1 测量时间的优选结果

Table 1 Optimal results of measurement time

Duration /s	Mass fraction of element /10 ⁻⁶					
	Ca	Fe	Sn	Zn	Hg	Cl
50	62472	322	86	29	14	0
60	60653	395	59	58	13	0
70	77274	412	79	46	14	0
80	77638	422	76	46	14	0
90	63311	331	77	36	9	0
100	72440	384	60	44	11	0

用于检验检测工作中分析结果精度的一个变动系数,表2中各元素的相对标准偏差均小于10%,表明在该条件下实验仪器效果稳定,实验结果准确可靠。

表2 相对标准偏差计算结果

Table 2 Calculation results of relative standard deviation

Measurement No.	Mass fraction /%				
	Ca	Fe	Sn	Zn	Cl
1	6253	523	65	58	1532
2	6206	530	55	55	1297
3	6368	543	56	54	1455
4	6802	578	63	57	1506
Average value	6407.25	543.5	59.75	56	1447.5
Standard deviation	271.8227	24.4472	4.9917	1.8257	105.3075
Relative standard deviation /%	4.24	4.50	8.35	3.26	7.28

表3 6种塑料打包带(绳)样本相关性计算结果

Table 3 Correlation analysis results of six plastic pack belt (rope) samples

Sample No.	Parameter	5 [#]	7 [#]	10 [#]	15 [#]	20 [#]	28 [#]
5 [#]	Pearson coefficient	1					
	Sig. (2-tailed)						
7 [#]	Pearson coefficient	0.985**	1				
	Sig. (2-tailed)	0					
10 [#]	Pearson coefficient	0.892**	0.878**	1			
	Sig. (2-tailed)	0	0				
15 [#]	Pearson coefficient	0.737**	0.683*	0.813**	1		
	Sig. (2-tailed)	0.010	0.020	0.002			
20 [#]	Pearson coefficient	0.884**	0.862**	0.992**	0.881**	1	
	Sig. (2-tailed)	0	0.001	0	0		
28 [#]	Pearson coefficient	0.891**	0.872**	0.998**	0.851**	0.998**	1
	Sig. (2-tailed)	0	0	0	0.001	0	

Note: ** Correlation is significant at the 0.01 level (2-tailed), and * Correlation is significant at the 0.05 level (2-tailed).

根据表3可知:除15[#]样本和7[#]样本之外,所有样本之间的Sig值均小于0.01,表示差异具有高度统计学意义;15[#]样本和7[#]样本之间的Sig值满

4.3 系统聚类

本实验借助SPSS24.0统计分析软件对X射线荧光光谱测量结果进行统计分析。选择离差平方和(Ward)法作为聚类方法,采用平方欧氏距离作为测量区间以描述样本间的亲疏程度^[11],进行系统聚类分析,聚类结果树状图见图1。

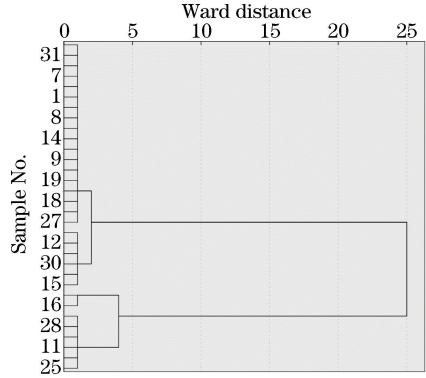


图1 32种塑料打包带(绳)样本的聚类分析结果
Fig. 1 Cluster analysis results of 32 plastic pack belt (rope) samples

当Ward距离最小时,32个样本被分为了4类;当Ward距离为5时,32个样本被分为了2类。当凝聚到某个程度,聚类之间的距离都大于阈值25时,就停止凝聚,即所有个体归为一类。

为考察系统聚类的准确性和合理性,以不同的连接距离作为依据,抽取5[#]、7[#]、10[#]、15[#]、20[#]、28[#]这6个样本进行计算验证,并考察其显著性Sig值和Pearson相关系数。计算结果见表3。

足0.01<P<0.05,表示两样本之间的差异具有统计学意义,符合检验要求;10[#]与20[#]、10[#]与28[#]的Pearson相关系数分别为0.992、0.998,表明两者之

间相关性十分强,但7[#]与15[#]、7[#]与20[#]的Pearson相关系数分别为0.683、0.862,表明两者之间相关性较弱。由此可以将5[#]与7[#]归为一类,10[#]、20[#]和28[#]归为一类,15[#]单独归为一类。

通过对显著性Sig值和Pearson相关系数的分析可知,使用Ward法作为聚类方法,平方欧氏距离作为测量区间的效果最好。当Ward距离最小时,将样本划分为4个类别较为合理。

4.4 判别分析

为进一步保证系统聚类结果的准确可靠,借助SPSS24.0统计分析软件中的Fisher判别分析法对初始分组中的样本进行判别分析,判别函数特征值见表4。

表4 判别函数特征值表

Table 4 Eigenvalue table of discriminant function

Function No.	Eigenvalue	Variance /%	Cumulation /%	Canonical correlation
1	73.171 ^a	96.3	96.3	0.993
2	2.263 ^a	3.0	99.3	0.833
3	0.544 ^a	0.7	100.0	0.059

Note: a represents the first 3 canonical discriminant functions were used in the analysis.

表4展示了判别函数的特征值及其累积贡献率。在分类过程中,软件建立了3个判别函数,其中判别函数1的特征值的方差贡献率高达96.3%,而判别函数2和3的方差贡献率却仅为3.0%和0.7%,这表明判别函数1所携带的信息远大于判别函数2和3。判别函数1可作为区分样本的主要判别依据。

由于判别函数1的方差贡献率远超过其他判别函数,为确保判别分析结果准确可靠,进一步考察其结构矩阵,见表5。

表5 判别函数结构矩阵

Table 5 Structural matrix of discriminant function

Element	Function 1	Function 2	Function 3
Ca	-0.797*	-0.099	-0.148
Fe	-0.092*	0.054	-0.089
Cl	-0.034	0.359*	-0.121
Zn	-0.092	-0.195*	-0.051
Ti	-0.038	-0.24	0.422*
Ba	0.002	-0.002	-0.106*
Sn	0.039	-0.004	-0.052*

Note: * represents largest absolute correlation between each variable and any discriminant function.

由判别函数结构矩阵可知,判别函数1主要与Ca元素的含量相关。在塑料加工过程中,除了使用

合成树脂作为主要原料外,为了满足各种商业、工业的需求,通常还会科学合理地添加无机粉体填料,其中碳酸钙是使用最为广泛且用量最大的无机粉体材料,占到了使用的无机粉体材料总量的70%(质量分数)以上^[12]。在塑料中添加碳酸钙除了可以增强塑料的硬度外,还可以改善塑料的加工性能和制品性能^[13]。此外,为了达到改善高分子材料的稳定性,延长产品寿命,扩大其应用范围等目的,通常会根据用途的不同,添加不同种类的助剂(如稳定剂、着色剂、阻燃剂等),从而引入了Fe、Cl、Zn、Sn等元素。由此可知,在判别分析中变量因子的选取较为恰当,判别结果准确可靠。

判别函数1与函数2之间典型相关性系数分别为0.993和0.833,表示4组样本在函数1和函数2这两个维度上具有显著差异。由此选择函数1和函数2作为函数判别轴建立联合分布图,见图2。

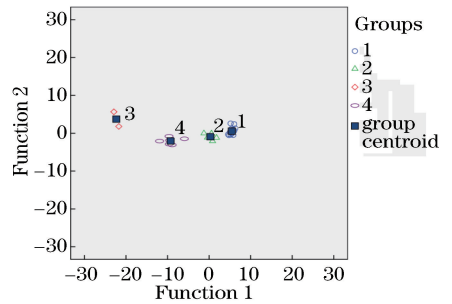


图2 建立在判别函数上的联合分布图

Fig. 2 Joint distribution map based on discriminant function

联合分布图显示,4组样本的组质心被明显区分开来,特别是在函数1的维度上,4组样本之间的差异性显著,表示函数1对样本之间的区分效果更好。如需检验新的未知样本的分组情况,只需输入样本位置,就能直观地在联合分布图上找到与之距离最近的分组质心,从而完成对未知样本的准确分类。

综上所述,将32种塑料打包带(绳)进行聚类分析后分为4类,见表6。

表6 32种样本系统聚类结果

Table 6 Hierarchical clustering results of 32 plastic pack belt (rope) samples

Classification	Sample No.
I	1 [#] , 2 [#] , 5 [#] , 7 [#] , 8 [#] , 9 [#] , 13 [#] , 14 [#] , 17 [#] , 18 [#] , 19 [#] , 22 [#] , 24 [#] , 26 [#] , 27 [#] , 29 [#] , 31 [#] , 32 [#]
II	4 [#] , 6 [#] , 12 [#] , 15 [#] , 21 [#] , 30 [#]
III	3 [#] , 16 [#]
IV	10 [#] , 11 [#] , 20 [#] , 23 [#] , 25 [#] , 28 [#]

5 结 论

结合 X 射线荧光光谱法和多元统计分析,对塑料打包带(绳)进行了系统性分析,确定了聚类方法和测量区间,通过系统聚类将 32 个样本初步分为 4 类,随后通过计算样本之间的相关性确定聚类方法是否可取。最后使用判别分析,确定各组之间组质心完全分离开来,判别情况良好,确保了系统聚类效果明显。借助多元统计的方法,挖掘了各样本元素含量之间的内在关联,成功建立了判别函数模型,达到了对样本快速分类的目的。

本文有助于公安司法鉴定合理有效的展开,促进了模式识别技术在理化检验中的应用。接下来的研究工作主要是进一步训练函数模型,强化核心算法,以实现对未知样本的自动化识别与定性定量分类。

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] Yi L T, Liu Z G, Chen M, *et al.* A new automatic analysis method for X-ray fluorescence spectrometric qualitative analysis [J]. *Laser & Optoelectronics Progress*, 2014, 51(7): 073001.
易龙涛, 刘志国, 陈曼, 等. 一种 X 射线荧光光谱数据批处理新方法的研究[J]. *激光与光电子学进展*, 2014, 51(7): 073001.
- [3] 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.
马泉, 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.
- [4] Chen P R, Li J H, Deng B. Modern instrument analytical experiment and techniques [M]. Beijing: Tsinghua University Press, 2006: 50-53.
陈培榕, 李景虹, 邓勃. 现代仪器分析实验与技术[M]. 北京: 清华大学出版社, 2006: 50-53.
- [5] Tian B. Hierarchical clustering method and its research about application[J]. *Yinshan Academic Journal(Natural Science Edition)*, 2014, 28(2): 11-16.
田兵. 系统聚类法及其应用研究[J]. *阴山学刊(自然科学版)*, 2014, 28(2): 11-16.
- [6] He X L, Wang J F, Liu T F, *et al.* Examination the front bumpers based on Raman spectroscopy and hierarchical clustering [J]. *The Journal of Light Scattering*, 2018, 30(2): 168-173.
何欣龙, 王继芬, 刘腾飞, 等. 拉曼光谱结合系统聚类法检验汽车前保险杠[J]. *光散射学报*, 2018, 30(2): 168-173.
- [7] Wasserstein R L, Lazar N A. The ASA statement on p-values: context, process, and purpose [J]. *The American Statistician*, 2016, 70(2): 129-133.
- [8] Zhang H C, Xu J P. Modern psychology and educational statistics [M]. Beijing: Beijing Normal University Press, 2011.
张厚燊, 徐建平. 现代心理与教育统计学[M]. 北京: 北京师范大学出版社, 2011.
- [9] Li Y P, Huang F R, Xian R Y, *et al.* Discrimination analysis of adulterating olive oil with soybean oil by attenuated total reflectance-Fourier transform infrared spectrometry combined with pattern recognition methods [J]. *Laser & Optoelectronics Progress*, 2015, 52(3): 033001.
黎远鹏, 黄富荣, 洗瑞仪, 等. 衰减全反射傅里叶变换红外光谱法结合模式识别分析橄榄油中掺杂大豆油的研究[J]. *激光与光电子学进展*, 2015, 52(3): 033001.
- [10] Jiang H, Zhang L, Man J, *et al.* Analysis of black plastic bag by XRF [J]. *Journal of Liaoning Normal University(Natural Science Edition)*, 2017, 40(1): 69-73.
姜红, 张璐, 满吉, 等. X 射线荧光光谱法检验黑色塑料袋的研究[J]. *辽宁师范大学学报(自然科学版)*, 2017, 40(1): 69-73.
- [11] He X L, Wang J F, Liu W H, *et al.* Discrimination and classification the plastic steel window based on Raman spectroscopy and cluster analysis [J]. *Chemical Research and Application*, 2017, 29(9): 1387-1392.
何欣龙, 王继芬, 刘文浩, 等. 拉曼光谱结合聚类分析法区分检验塑钢窗[J]. *化学研究与应用*, 2017, 29(9): 1387-1392.
- [12] Liu Y J. Progress on the application of calcium carbonate in plastics [J]. *Inorganic Chemicals Industry*, 2008, 40(3): 11-13.
刘英俊. 碳酸钙在塑料中应用进展[J]. *无机盐工业*, 2008, 40(3): 11-13.
- [13] Zhang H. Production of calcium carbonate and its application in plastics [J]. *Rubber & Plastics Resources Utilization*, 2010(3): 16-20.
张亨. 碳酸钙的生产及其在塑料中的应用研究[J]. *橡塑资源利用*, 2010(3): 16-20.