第 59 卷 第 4 期/2022 年 2 月/激光与光电子学进展 研究论文

激光与光电子学进展 先进成像

X射线荧光光谱结合深度学习算法可视化检验食品包装纸

郭琦[1]，姜红[1*]，杨金颉[1]，吴克难[2]，满吉[3]

[1]中国人民公安大学侦查学院，北京 100038；

[2]武汉理工大学计算机科学与技术学院，湖北 武汉 430070；

[3]北京华仪宏盛技术有限公司，北京 100123

摘要 为了实现对案件现场常见食品包装纸的快速分类及认定，提出一种基于X射线荧光光谱(XRF)结合深度学习算法的食品包装纸可视化检验方法。首先，采用XRF检验44个不同来源的食品包装纸样本中的无机元素，并根据主要构成元素的含量，对其进行人工分类和系统聚类分析。其次，分别使用主成分分析和t分布随机邻域嵌入两种降维算法处理数据以检验聚类效果，并实现数据分类可视化。最后，随机选取80%的样本作为训练集构建人工神经网络，并进行相关实验。实验结果表明，所提方法在测试集上的分类正确率为88.9%，可以为未来公安业务实际应用提供参考。

关键词 X射线光学；X射线荧光光谱；系统聚类；主成分分析；t分布随机邻域嵌入；多层前馈神经网络

中图分类号 O657.34 **文献标志码** B **doi**：10.3788/LOP202259.0434001

Visual Inspection of Food Packaging Paper by X-Ray Fluorescence Spectroscopy Combined with Deep Learning Algorithm

Guo Qi[1], Jiang Hong[1*], Yang Jinjie[1], Wu Kenan[2], Man Ji[3]

[1]*Institute of Criminal Investigation, People's Public Security University of China, Beijing* 100038, *China*;

[2]*Institute of Computer Science and Technology, Wuhan University of Technology, Wuhan, Hubei* 430070, *China*;

[3]*Beijing Huayi Honrizon Technology Co., Ltd., Beijing* 100123, *China*

Abstract To quickly classify and identify common food packaging paper at the scene of the case, a visual inspection method of food packaging paper based on X-ray fluorescence spectroscopy (XRF) and deep learning algorithm is proposed. First, the inorganic elements in 44 samples of food packaging paper from different sources were detected via XRF, and artificial classification and cluster analysis were performed based on the content of the main constituent elements. Second, to test the clustering effect and visualize data classification, two-dimensionality reduction algorithms, principal component analysis, and t-distribution random neighborhood embedding are used. Finally, 80% of the samples are randomly selected as the training set to construct the artificial neural network , and relevant experiments are carried out. The experimental results show that classification accuracy of the proposed method on the test set is 88.9%, which can be used as a reference for future practical applications of public security business.

Key words X-ray optics; X-ray fluorescence spectroscopy; hierarchical clustering; principal component analysis; t-distribution random neighborhood embedding; multilayer feedforward neural network

收稿日期：2021-02-05；**修回日期**：2021-03-19；**录用日期**：2021-04-02

基金项目：国家重点研发计划(2017YFC0822004)、中国人民公安大学2019年度基科费重点项目(2021JKF212)

通信作者：*jiangh2001@163.com

1 引 言

纸制品具有降解容易、耐受温度高、环境友好及可回收利用等特点，广泛应用于餐饮业食品包装领域[1]。各类案发现场中也常常能够提取到食品包装纸物证。厂家为了实现利益最大化，在生产以及后续加工食品包装纸的过程中会添加一些助剂，如除菌剂、增塑剂、增白剂、固化剂等[2]。因此，不同品牌和批次的食品包装纸采用的生产工艺以及添加的化学物质含量一般不同，所含元素的种类和含量差异明显，可利用这一点对不同来源的食品包装纸进行区分。

目前常用的纸张物证检验方法包括光谱法、色谱法、等离子体质谱法、仪器联用法等[3]，X射线荧光光谱法（XRF）根据不同元素特征X荧光光谱线和能量强度大小的不同来进行元素的定性与定量分析[4]，且能够同时分析多种元素，在物证鉴定领域中应用较为广泛[5-7]。

本文提出一种基于XRF结合深度学习算法的食品包装纸可视化检验方法。首先，使用能量色散型X射线荧光光谱仪对多个食品包装纸样本进行检验，对检验结果进行系统聚类分析，并制作聚合系数折线图以优选类别数。其次，分别采用线性降维和非线性降维方法约简数据，对比两种降维效果并观察聚类结果，用于反证系统聚类的分类效果。最后，随机选取80%的样本作为训练集构建人工神经网络并进行训练，以实现对未知样本的分类判断[8]。

2 实验部分

2.1 实验仪器及条件

实验仪器为能量色散型X射线荧光光谱仪（X-MET8000，牛津公司），以铑（Rh）为阳极靶。

2.2 实验样本

实验样本为44个不同品牌及来源的食品包装纸样本。

2.3 实验方法

首先，对样本进行前处理，用酒精棉片小心清理表面异物，在清理过程中注意不要破坏样本表面。然后将样本置于仪器上，以10 s为间隔进行检测，检测时间从20 s逐步增加至120 s。实验结果表明，最佳测量时间为60 s。在最优条件下对样本进行10次检验（重现性实验），计算检验结果得到的各元素标准偏差及相对标准偏差均小于5%，表明可以使用此仪器检验食品包装纸[9]。最后对每一个样本平行检验3次取平均值，共检出10余种元素，选择检出比例较高的6种元素作为检验结果，如表1所示。

表1 X射线荧光检验结果

Table 1 X-ray fluorescence detection results

unit: $\mu g \cdot g^{-1}$

Sample No.	Ca	Fe	Sn	Zn	Cl	Ti	Sample No.	Ca	Fe	Sn	Zn	Cl	Ti
1#	1341	336	265	25	972	4315	23#	17091	528	123	0	3046	0
2#	265	185	152	6	4267	0	24#	430	215	137	14	2037	3568
3#	710	280	125	16	642	25	25#	144224	5997	286	156	0	957
4#	22620	352	122	14	2230	0	26#	36102	950	290	48	2467	0
5#	10312	315	75	13	4767	0	27#	57704	1754	288	53	0	142
6#	338277	7237	243	152	27057	7386	28#	86662	1987	303	52	3740	145
7#	974	593	223	15	1610	0	29#	149657	3272	374	67	5898	247
8#	2418	641	191	16	629	0	30#	62564	1139	227	28	754	128
9#	9848	401	316	17	8436	472	31#	81827	1618	430	37	10852	116
10#	9252	313	187	13	2509	988	32#	4417	300	288	16	20487	0
11#	50212	1358	176	47	0	0	33#	1019	417	212	18	2898	28
12#	90821	4351	253	89	0	234	34#	653951	7182	546	1378	9948	821
13#	480	224	199	13	885	0	35#	519	534	93	0	2329	0
14#	48275	1207	220	30	1688	0	36#	348362	1834	261	57	8147	2457
15#	168161	1763	543	60	2121	0	37#	278	231	150	28	5371	0
16#	182537	5715	355	136	2023	249	38#	374	190	111	0	3846	0
17#	58463	2015	266	59	1798	155	39#	87355	988	522	56	4459	0
18#	110658	3077	271	95	0	216	40#	130055	1084	151	33	1650	0
19#	284881	7290	471	259	3050	610	41#	99722	1163	296	37	1839	0
20#	206208	5276	435	114	2608	772	42#	102456	816	165	30	4822	0
21#	127196	3536	309	99	0	306	43#	46252	642	125	15	1682	0
22#	87681	2542	281	105	0	319	44#	29072	1869	75	11	934	4642

3　数据处理与分析

3.1　人工分类

6种元素分别来源于造纸填料$CaCO_3$、TiO_2及性能助剂如防腐剂、阻燃剂、漂白剂、纤维处理助剂等[10-13]。由表1可知，Fe、Ca、Sn元素存在于所有样本中，但Cl、Zn、Ti 3种元素仅存在于部分样本中。根据是否含有Cl、Zn、Ti元素对样本进行人工分类，可将样本划分为5类，如图1所示，图中“+”表示含有该元素，“−”表示不含该元素。每个组别的样本可利用各元素相对含量比值特异性进一步区分[14]。

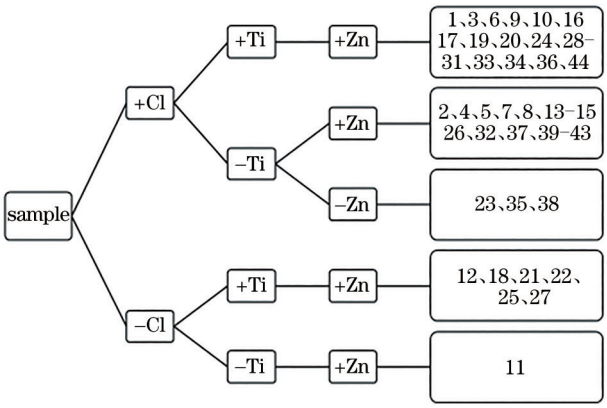

图1　人工分类结果

Fig. 1　Manual classification results

3.2　系统聚类

由于人工分类存在解释不全面、耗费时间长、分类条件受限等缺点，机器分类是一个更好的选择。系统聚类是最常用的机器分类方式，又被称为层次聚类，是利用变量之间的相似性而逐步归簇成类，反映变量或区域之间的内在组合关系的方法[14]。本实验组用组间连接法计算样本的类间距离，采用平方欧氏距离度量个体距离，对z-score标准化处理数据进行系统聚类，聚类结果树状图如图2所示。从图2中可以看出，Ward距离最小时，样本可以被分成10类，随着Ward距离增大，样本类别数递减，迭代停止时合为一类[15]。为优选类别数，以“聚合系数”为纵坐标，“类别数”为横坐标，所绘制的折线图如图3所示。从图中可以看出，当类别数在[4，9]区间内时，折线明显趋缓。优选类别数可为6类或8类。由于存在部分样本单独成类的情况，本实验组将样本类别数设置为8类。

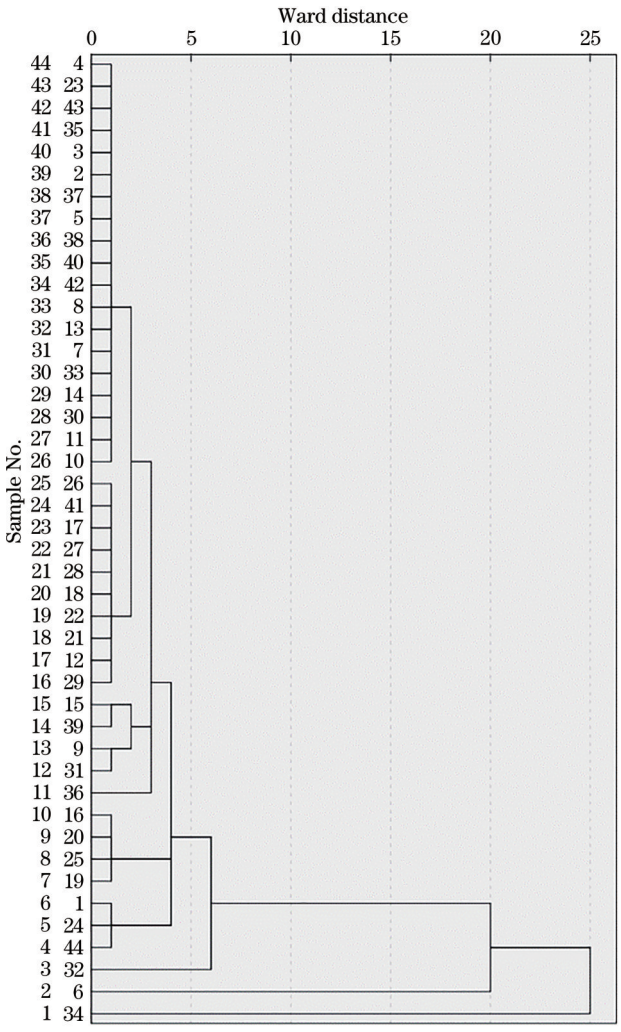

图2　聚类结果树状图

Fig. 2　Tree diagram of clustering results

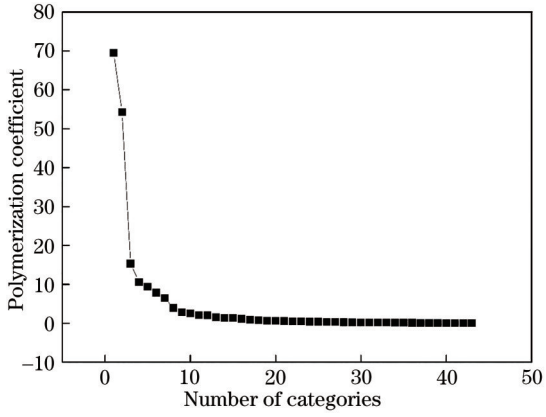

图3　聚合系数折线图

Fig. 3　Line chart of polymerization coefficient

3.3　聚类结果可视化

数据可视化是指通过线性或非线性映射将样本从高维空间映射到低维空间，从而获得高维数据的一个有意义的低维表示的过程[16]，也就是数据降

维。常用方法包括线性降维(投影)和非线性降维(流形学习)两类:线性降维是指用低维数据表现高维数据点之间的线性关系,例如主成分分析(PCA)通过研究原始变量的线性组合如相关矩阵或协方差矩阵,构造若干主成分分量,以达到降维目的;非线性降维旨在发现高维数据集分布的流形空间的内在主要影响变量的规律性[17],非线性降维方法包括局部线性嵌入(LLE)、多维尺度变换(MDS)、t分布随机邻域嵌入(tSNE)等[18]。所采用的tSNE原理是在保持相互之间分布的概率不变的前提下,将高维空间的点对映射到低维空间。高维空间使用高斯分布、低维空间使用t分布将距离转化为概率分布,以联合概率表示点对应的相似度,通过算法优化两个分布之间的距离相对熵(KL散度),得到在低维空间的样本分布[19]。

为检验系统聚类的效果,优选合适的降维方式。对样本数据分别进行主成分分析和tSNE降维。以主成分个数为2和3建立主成分得分图(即2/3D效果图),并和tSNE算法降维效果进行对比,实验结果如图4和图5所示。从图中可以看出:数据降至二维时,主成分分析聚类效果明显,同类样本簇集度高,而tSNE算法降维聚类效果较弱;在三维效果图中,tSNE算法聚类效果明显。原因在于主成分分析存在数据特征缺失大、样本点拥挤等问题,而tSNE算法在低维空间使用的是更偏重长尾分布的t分布,使得在高维空间中距离较小的数据簇在低维空间中的间距拉大,有效解决了主成分分析存在的样本点拥挤问题[18],在集群分离方面效果更优[20]。

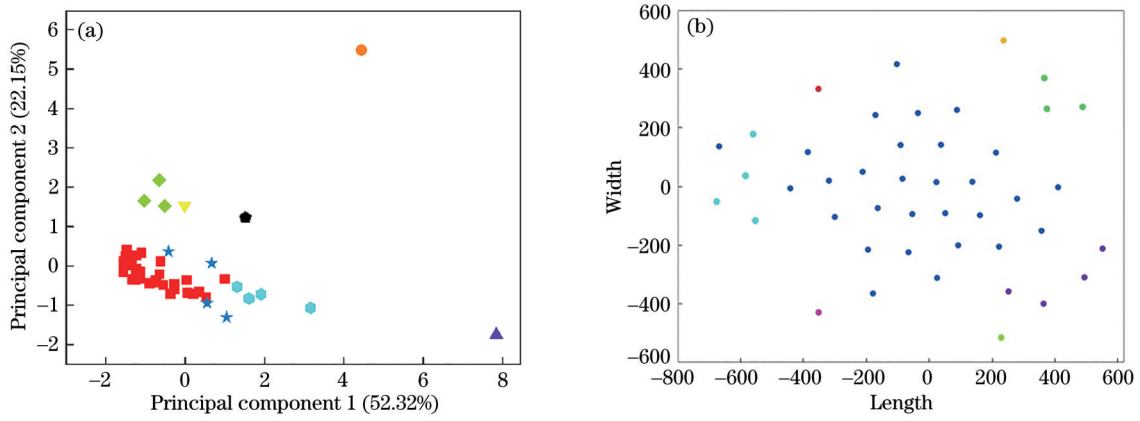

图4　降维算法2D效果图对比。(a) PCA;(b) tSNE

Fig. 4　Comparison of 2D renderings of dimension reduction algorithms. (a) PCA; (b) tSNE

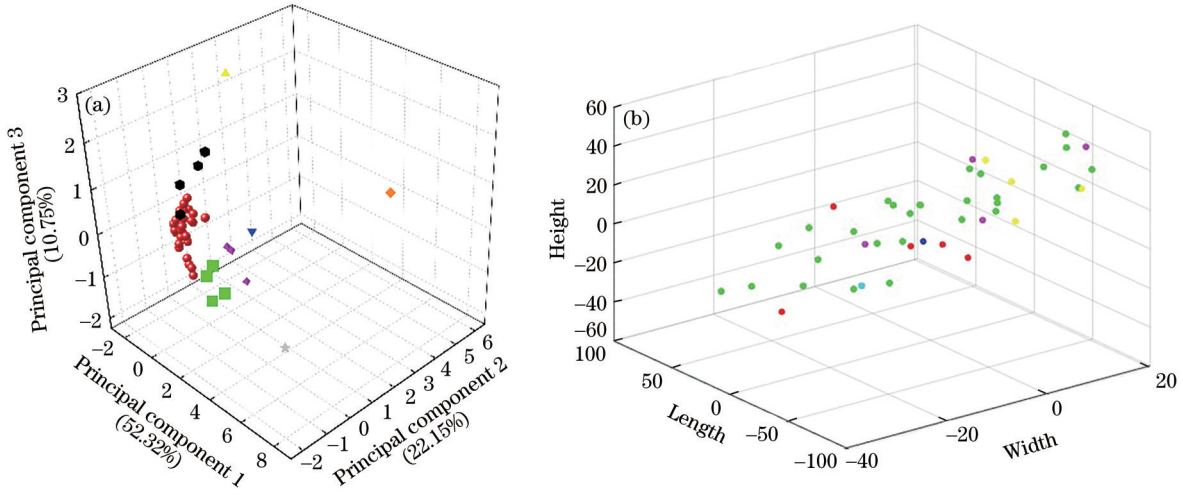

图5　降维算法3D效果图对比。(a) PCA;(b) tSNE

Fig. 5　Comparison of 3D renderings of dimension reduction algorithms. (a) PCA; (b) tSNE

表2为两种算法不同维度效果图的数据损耗值对比,其中主成分分析数据损耗值为1与方差累计贡献率的差值。由于主成分个数为2时,PCA的累计方差贡献率达到74.473%,能够包含样本绝大部

表 2 两种算法数据损耗值对比

Table 2 Comparison of data loss values of two algorithms

Dimension	PCA	tSNE
2D	0.25527	0.336786
3D	0.14772	0.0917815

分信息，相较于tSNE算法时的数据损失量更小，簇集程度更高，在本实验中更适用于与系统聚类结果相互印证。

3.4 神经网络

为对未知样本做到准确的预判，所提方法引入多层前馈神经网络模型，该模型不存在环路或回路，是由输入层、若干个隐藏层和输出层组成的全连接网络[21]。将所有单独成类样本合为同一类命名为“9”后，随机选取总样本数据的80%作为训练集，20%作为测试集。为了验证检验效果，测试集需包含每一类样本，样本量少的类别取一个作为测试样本，样本量大的类别如第6类取多个作为测试样本。设置相关参数，训练集多次将输出结果与真实结果相拟合，最终搭建的神经网络模型[22]中隐藏层内含3个单元，输出层以交叉熵为误差函数。实验结果如表3所示，测试集分类结果正确率为88.9%，其中原属于第9类样本被归为第7类。由于第9类样本实际是单独成类样本的集合，所含样本的特征性区别明显，且第7类和第9类训练样本量较少，训练模型不成熟，导致出现误判。但88.9%的预测正确率表明所建立人工神经网络模型的可行性。

表 3 人工神经网络输出分类结果

Table 3 Artificial neural network outputs classification results

Dataset	Category	Number of output result samples of each category					Accuracy /%
		4	5	6	7	9	
Train	4	3	0	0	0	0	100.0
	5	0	3	0	0	0	100.0
	6	0	0	24	0	0	100.0
	7	0	0	0	2	0	100.0
	9	0	0	0	0	3	100.0
	Overall percentage /%	8.6	8.6	68.6	5.7	8.6	100.0
Test	4	1	0	0	0	0	100.0
	5	0	1	0	0	0	100.0
	6	0	0	5	0	0	100.0
	7	0	0	0	1	0	100.0
	9	0	0	0	1	0	0.0
	Overall percentage /%	11.1	11.1	55.6	22.2	0.0	88.9

4 结 论

所提方法采用X射线荧光光谱法对食品包装纸样本进行定量与定性检验，根据特征元素的种类和含量对44个不同来源的样本进行准确区分。系统聚类后结合深度学习算法，对比讨论了两类降维方法的效果，实现聚类结果可视化，充分挖掘了样本元素含量之间的关系，提高了对包装纸样本分析的科学性和准确性。最后构建人工神经网络模型，实现了对未知数据的预测。

由于实际应用中样本主成分个数可能大于3，无法实现聚类结果可视化，tSNE算法对于集群分离方面效果虽优，但计算数据嵌入花费时间长。基于本研究，下一步可联合使用主成分分析与tSNE算法降维模式，实现短时间达到高维数据的可视化观测；除此之外，考虑到建立较优的人工神经网络应以大量数据为基础，下一步可建立食品包装纸材料的X射线荧光光谱数据库，达到可以存储多种类光谱信息的目的；同时，可结合差分拉曼光谱技术等方法，发展光谱融合技术，建立高速光谱探测分析平台。最终，根据建立的光谱数据库，高效实现未知样本集群分类可视化、有效识别未知样本，为公安业务实践提供识别检测算法。

参 考 文 献

[1] Han Z C. Promotion and application of food packaging paper[J]. Tianjin Paper Making, 2013, 35(3): 22-26.
韩志诚. 食品包装纸的推广和应用[J]. 天津造纸, 2013, 35(3): 22-26.

[2] Song H, Wang T J, Li B, et al. Research progress of detection technology of chemicals in food packaging paper materials[J]. Food Science, 2009, 30(17): 339-344.
宋欢, 王天娇, 李波, 等. 纸质食品包装材料中化学物分析检测技术研究进展[J]. 食品科学, 2009, 30(17): 339-344.

[3] Liu T T, Li J B. A review of the methods of paper inspection in cases[J]. Guangdong Chemical Industry, 2017, 44(20): 115-116.
刘彤彤, 李建邦. 案件中纸张检验方法的研究综述[J]. 广东化工, 2017, 44(20): 115-116.

[4] Li D L, Li Z D. Calibration of energy dispersive X-ray fluorescence spectrometer and evaluation of measurement uncertainty[J]. China Measurement & Test, 2017, 43(S1): 33-36.
李德林, 李志得. 能量色散X射线荧光仪校准及不确定度评定[J]. 中国测试, 2017, 43(S1): 33-36.

[5] Jiang H, Wang X, Xu L L, et al. Difference test of paper ashes by XRF combined with multivariate statistical analysis[J]. Laser Technology, 2021, 45(3): 318-321.
姜红, 王欣, 徐乐乐, 等. X射线荧光光谱结合多元统计学检验纸张灰烬[J]. 激光技术, 2021, 45(3): 318-321.

[6] 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): 223005.
马枭, 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): 223005.

[7] Fu J Z, Jiang H, Li Y, et al. Examination of cigarette ash evidence by XRF combined with chemometrics[J]. Laser & Optoelectronics Progress, 2021, 58(6): 0630003.
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): 0630003.

[8] Xiao K, Tian L J, Wang Z Y. Fast super-resolution fluorescence microscopy imaging with low signal-to-noise ratio based on deep learning[J]. Chinese Journal of Lasers, 2020, 47(10): 1007002.
肖康, 田立君, 王中阳. 基于深度学习的低信噪比下的快速超分辨荧光显微成像[J]. 中国激光, 2020, 47(10): 1007002.

[9] Chen Z, Jiang H, Li C Y, et al. A study on disposable paper cups tested by X-ray fluorescence spectroscopy[J]. China Pulp & Paper Industry, 2018, 39(22): 32-36.
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.

[10] Zhang J D. Effect of fillers on paper properties[J]. Heilongjiang Pulp & Paper, 2017, 45(2): 16-17.
张金顶. 填料对纸张性能的影响[J]. 黑龙江造纸, 2017, 45(2): 16-17.

[11] Chang Y J. Application of flame retardant in paper industry[J]. Heilongjiang Pulp & Paper, 2013, 41(4): 36-38.
常永杰. 阻燃剂在造纸工业上的应用[J]. 黑龙江造纸, 2013, 41(4): 36-38.

[12] Kuang S J. ECF and TCF[J]. China Pulp & Paper, 2005, 24(10): 51-56.
邝仕均. 无元素氯漂白与全无氯漂白[J]. 中国造纸, 2005, 24(10): 51-56.

[13] Shu H Y, Pan L Q, Tu K, et al. Advances in research on antibacterial materials in food packaging [J]. Food Science, 2015, 36(5): 260-265.
束浩渊, 潘磊庆, 屠康, 等. 抗菌材料在食品包装中的研究进展[J]. 食品科学, 2015, 36(5): 260-265.

[14] Wang X, Xi D, Jiang H, et al. Detection of lipstick by XRF combined with cluster analysis[J]. Chemical Research and Application, 2020, 32(10): 1920-1923.
王欣, 习豆, 姜红, 等. X射线荧光光谱法结合聚类分析检验口红[J]. 化学研究与应用, 2020, 32(10): 1920-1923.

[15] Zhang J, Jiang H, Liu F, et al. Differential Raman spectroscopy visualization and rapid identification of shoe sole materials[J]. Laser & Optoelectronics Progress, 2021, 58(8): 0830004.
张进, 姜红, 刘峰, 等. 鞋底材料的差分拉曼光谱可视化快速鉴别[J]. 激光与光电子学进展, 2021, 58(8): 0830004.

[16] Wu X T, Yan D Q. Analysis and research on method of data dimensionality reduction[J]. Application Research of Computers, 2009, 26(8): 2832-2835.
吴晓婷, 闫德勤. 数据降维方法分析与研究[J]. 计算机应用研究, 2009, 26(8): 2832-2835.

[17] Xu R, Jiang F, Yao H X. Overview of manifold learning[J]. CAAI Transactions on Intelligent Systems, 2006, 1(1): 44-51.
徐蓉, 姜峰, 姚鸿勋. 流形学习概述[J]. 智能系统学报, 2006, 1(1): 44-51.

[18] Cheng C, Yang C H. Alzheimer diagnosis model based on t-distributed stochastic neighbor embedding [J]. Journal of Xiamen University (Natural Science),

2017, 56(1): 123-128.
成超, 杨晨晖. 基于t分布随机邻域嵌入的阿尔茨海默症诊断模型[J]. 厦门大学学报(自然科学版), 2017, 56(1): 123-128.

[19] Dong A G, Zhang Q, Liu H C, et al. Hyperspectral image classification based on TSNE and multiscale sparse auto-encoder[J]. Computer Engineering and Applications, 2019, 55(21): 177-182, 219.
董安国, 张倩, 刘洪超, 等. 基于TSNE和多尺度稀疏自编码的高光谱图像分类[J]. 计算机工程与应用, 2019, 55(21): 177-182, 219.

[20] Dimensionality reduction for data visualization: PCA vs TSNE vs UMAP[EB/OL]. (2020-06-02)[2021-02-15].https://www.codercto.com/a/112721.html.

[21] Feng P, Li Y. Semiconductor laser parameter inverse design method based on artificial neural network and particle swarm optimization[J]. Chinese Journal of Lasers, 2019, 46(7): 0701001.
冯佩, 李俣. 基于人工神经网络和粒子群优化的半导体激光器参数反向设计方法[J]. 中国激光, 2019, 46(7): 0701001.

[22] Wang X B, Ma X, Wang X C. Infrared spectral pattern recognition of watercolor pen ink based on artificial neural network[J]. Laser & Optoelectronics Progress, 2020, 57(15): 153005.
王晓宾, 马枭, 王新承. 基于人工神经网络的水彩笔油墨红外光谱模式识别[J]. 激光与光电子学进展, 2020, 57(15): 153005.