

基于近红外光谱建立 PE、PP 和 PET 的识别分类模型

张毅民, 王 鹏, 白家瑞, 马冬雅

(天津大学化工学院, 教育部绿色合成与转化重点实验室, 天津 300072)

摘要: 利用近红外高光谱成像仪在 900 ~ 1 700 nm 的范围采集 PE、PP 和 PET 样本的高光谱图像, 并进行黑白校正, 提取感兴趣区域的反射率光谱数据; 利用主成分分析法对提取的数据去噪降维。结果表明, 前 3 个主成分的累计贡献率达 98.89%。把前 3 个主成分的载荷系数对波长作图, 得到了 6 个特征波长; 利用特征波长对应的反射率光谱数据进行判别分析, 并建立了 3 种塑料的识别分类模型; 用预测样本对模型进行检验, 结果显示, 预测样本的识别准确率为 95.24%, 表明该模型可准确有效地对 PE、PP 和 PET 进行识别分类。

关键词: 近红外高光谱; 塑料识别分类; 特征波长; 判别分析; 识别模型

中图分类号: 0657.3

文献标志码: A

文章编号: 0253-4320(2016)03-0182-05

DOI: 10.16606/j.cnki.issn 0253-4320.2016.03.046

Establishment of identification and classification model of PE, PP and PET based on near infrared spectroscopy

ZHANG Yi-min, WANG Peng, BAI Jia-rui, MA Dong-ya

(Key Laboratory for Green Chemical Technology of State Education Ministry, Tianjin University, Tianjin 300072, China)

Abstract: The hyperspectral images of PE, PP and PET samples are acquired using near-infrared hyperspectral imager in the range of 900 - 1 700 nm. Spectral reflectance data in the region of interest are extracted from the hyperspectral images after black and white correction. The denoising and dimensionality reduction of the extracted data is performed using principal component analysis. The results indicate that cumulative contribution rate of the first three principal components can reach up to 98.89%. Six characteristic wavelengths are selected according to the graph of the loadings of the first three principal components versus wavelength. Identification and classification model of these three kinds of plastic is set up on the basis of the discriminant analysis results of the spectral reflectance data which corresponds to the characteristic wavelengths. Finally, the model is validated by predicting samples, and results show forecast accuracy rate of the predicting samples is 95.24%. It turns out that this model can identify PE, PP and PET accurately and effectively.

Key words: near infrared hyperspectral; identification and classification of plastic; characteristics of the wavelength; discriminant analysis; recognition model

由于塑料具有质量轻、强度高、耐腐蚀、不易降解等特性以及相对低价格, 广泛应用于家电及电子产品中, 成为人们日常生产生活中必不可少的一部分。然而, 随着这些塑料制品的退役, 每年都会产生大量的塑料垃圾, 带来严重污染的同时也造成了资源的浪费。目前, 将废混合塑料分类并再生利用是一种可持续发展的解决方案, 既可保护环境, 又能带来可观的经济效益^[1]。因此建立快速、可靠的塑料识别分类模型对废混合塑料的回收再利用具有重要意义。

近红外高光谱图像技术 (near-infrared hyperspectral imaging, NIR-HIS) 是一种图像技术与光谱技术相结合的三维信息获取技术, 包括二维的位置信息和一维的光谱信息, 广泛应用于多种物质的快

速无损检测^[2]。近年来, NIR-HIS 应用于塑料识别领域的研究越来越多。Silvia Serrant 等^[3]用 NIR-HIS 获取聚烯烃样本的光谱数据, 用主成分分析法降低维度, 选择特征波长, 利用偏最小二乘判别分析法对聚乙烯和聚丙烯进行分类, 结果显示, 该流程可用于从生活垃圾中回收聚乙烯和聚丙烯制品过程中的质量和过程控制; G. Bonifazi 等^[4]介绍了基于 NIR-HIS 的聚烯烃回收设备、相关的软件架构和处理算法, 结果表明, 结合多元数据分析方法的 NIR-HIS 是一种快速、无损的检测手段, 可用于聚烯烃回收过程的在线质量和流程控制。

笔者分析了 PE、PP 和 PET 3 种塑料的近红外光谱图像, 简化光谱数据的处理流程, 建立了 3 种塑料的识别分类模型, 为实现废混合塑料的精准在线

识别分类提供理论依据。

1 实验部分

1.1 材料与仪器

塑料样本主要来源于生活中常见的各种塑料制品以及废旧家电、机电及电子产品拆解后得到的各种塑料,包括 PE、PP 和 PET。总共收集到 361 个塑料样本,包括 PE 样本 123 个,PP 样本 132 个,PET 样本 106 个。其中,样本集 256 个,用于建立识别模型,包括 PE 样本 88 个,PP 样本 92 个,PET 样本 76 个;预测集 105 个,用于检验模型,包括 PE 样本 35 个,PP 样本 40 个,PET 样本 30 个。

由于黑色样本的吸收率较强,而透明样本的反射率太强,采集的光谱数据起伏太大,不适合数据分析,因此,在本次实验中,以白色的塑料样本为主。

仪器选用芬兰 SPECIM 公司生产的 SWIR 成像光谱仪,型号为 N17E,光谱范围 900 ~ 1 700 nm,光谱分辨率为 5 nm,图像分辨率为 320 × 256 pixel,光谱仪帧数为 100 fps,光源使用 2 个 200 W 的卤钨灯。

1.2 光谱数据的校正与提取

采样时,曝光时间设定为 30 ms,帧数设定为 30 fps,平台移动速度为 2 cm/s,镜头距离样本的垂直距离为 34 cm。

关闭光源并用不透明的盖子完全盖住相机的镜头获得全黑图像,以消除相机 CCD 传感器暗电流的影响;在采集光谱数据的同时设置标准白板,获得同等条件下的白板光谱图像以消除光线的影响。

计算黑白校正图像的相对反射率 R :

$$R = [(R_0 - B) / (W - B)] \times 100\% \quad (1)$$

其中: R 为黑白校正图像的相对反射率; R_0 为原始高光谱图像; W 为标准白板高光谱图像(反射率为 99%); B 为全黑图像(反射率为 0%)^[5]。

使用 ENVI4.7 软件提取黑白校正过的高光谱图像的光谱反射率数据。选择塑料样本上的感兴趣区域(region of interest, ROI)内的光谱数据,即所选区域内所有像素点的平均反射率光谱。尽量选取能代表塑料样本特征的 ROI,实验中,选取 30 × 30 个像素点^[6]。

实验发现,在 900 ~ 1 000 nm 范围光谱起伏较大,存在噪声干扰,因此提取 1 000 ~ 1 700 nm 波段的光谱数据,共 209 个波段的高光谱数据。

1.3 PE、PP 和 PET 的光谱特征

典型 PE、PP 与 PET 样本在近红外波段 1 000 ~ 1 700 nm 范围的光谱特征如图 1 所示。由图 1 可知,3 种样本在 1 125 ~ 1 225、1 375 ~ 1 425、1 525 ~ 1 575、1 625 ~ 1 675 nm 范围内的反射率光谱图像有着各自明显的特征^[7]。

(上接第 181 页)

参考文献

- [1] Pullela P K, Chiku T, Carvan M J S, *et al.* Fluorescence-based detection thiols in vitro and in vivo using dithiol probes[J]. *Anal-Biochem*, 2006, 352(2): 265 - 273.
- [2] Tang Bo, Xing Yanlong. A rhodamine-based fluorescent probe containing a Se-N bond for detecting thiols and its application in living cells[J]. *J Am Chem Soc*, 2007, 129(38): 11666 - 11667.
- [3] Chusen Huang, Qin Yin. Highly selective fluorescent probe for vicinal-dithiol-containing proteins and in situ imaging in living cells [J]. *Angew Chem Int Ed*, 2011, 50(33): 7551 - 7556.
- [4] Jun Yin, Younghee Kwon. Cyanine-based fluorescent probe for highly selective detection of glutathione in cell cultures and live mouse tissues[J]. *J Am Chem Soc*, 2014, 136(14): 5351 - 5358.
- [5] Haijun Cuia, Hua Chen. A two-photon fluorescent turn-on probe for palladium imaging in living tissues[J]. *Sensors and Actuators B*, 2015, 219: 232 - 237.
- [6] Murat Isşık, Ruslan Guliyev, Safacan Kolemen. Designing an intracellular fluorescent probe for glutathione: Two modulation sites for selective signal transduction[J]. *Org Lett*, 2014, 16(12): 3260 - 3263.

- [7] Jun Yin, Younghee Kwon, abin Kim. Cyanine-based fluorescent probe for highly selective detection of glutathione in cell cultures and live mouse tissues [J]. *J Am Chem Soc*, 2014, 136(14): 5351 - 5358.
- [8] Huan Y F, Fei Q. A novel water-soluble sulfonated porphyrin fluorescence sensor for sensitive assays of H₂O₂ and glucose [J]. *Analyst*, 2015, 140(5): 1655 - 1661.
- [9] 陈欣, 赵燕芳, 李茹. 金属锌卟啉荧光光谱探针与蛋白质的研究与应用[J]. *光谱学与光谱分析*, 2008, 28(10): 247 - 348.
- [10] 罗凯. 席夫碱卟啉的合成及其光谱性能的研究[D]. 长沙: 湖南大学, 2008.
- [11] Fang Liu, Annie Soh Yan Ni. Lipopolysaccharide neutralizing peptide porphyrin conjugates for effective photoinactivation and intracellular imaging of gram-negative bacteria strains [J]. *Bioconjugate Chem*, 2012, 23(8): 1639 - 1647.
- [12] Ste'phane Girouard, Marie-He' le'ne Houle, Alain Grandbois. Synthesis and characterization of dimaleimide fluorogens designed for specific labeling of proteins [J]. *J Am Chem Soc*, 2005, 127(2): 559 - 566.
- [13] Qiuning Lin, Chunyan Bao. Target-Activated coumarin phototriggers specifically switch on fluorescence and photocleavage upon bonding to thiol-bearing protein [J]. *J Am Chem Soc*, 2012, 134(11): 5052 - 5055. ■

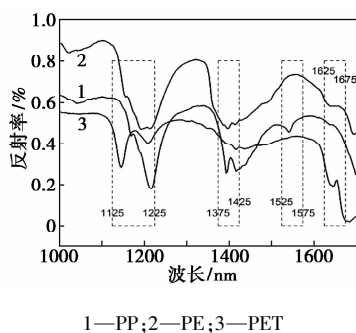


图1 典型PE、PP与PET样本的光谱特征

PE在1 216.64 nm处有明显波谷,1 375 ~ 1 425 nm的范围内有2个波谷和1个波峰,1 541.36 nm处有1个波谷,而PP和PET在这些波段处则无明显特征,因此可以把PE与PP和PET区别开。PET在1 146.54 nm处有明显的波谷,1 645.45 nm处有波谷,1 655.53 nm处有1个小的波峰,而PE和PP在这些波段处无明显特征,因此可以把PET与PE和PP区别开。

1.4 主成分分析

由于提取的光谱波段较多,而且其中有大量冗余的信息,后续的处理流程较为繁琐,不利于下一步数据分析,因此需要对采集到的光谱数据进行降维处理。

主成分分析(principal component analysis, PCA)是一种去除数据相关性、降低维度的多元统计方法,在数学上定义为一种正交线性变换,即将处于高维空间的原始数据信息转换到低维空间,排除原始信息中相关的数据,以少数最显著的特征变量(即主成分)的线性组合代表原始数据中的绝大部分信息^[8-9]。

对提取的光谱数据进行全谱主成分分析。根据主成分分析的结果,将每个主成分与其对应的贡献率作图,结果如图2所示。

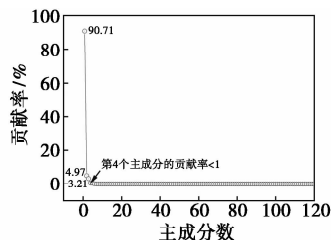


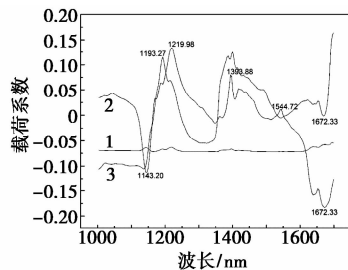
图2 主成分数与其对应的贡献率

由图2可知,前3个主成分的贡献率均大于1,而从第4个开始,其贡献率均小于1。其中,前3个主成分的累积贡献率达到了98.89%,即前3个主成分解释了原始变量的98.89%,代表了原始光谱

数据的绝大部分信息。因此,可确定全部光谱数据主成分分析的主成分数为3个^[10-11]。

1.5 特征波长的选取

由于主成分数确定为3个,因此选择前3个主成分对应的载荷系数对波长作图,结果如图3所示。



1—载荷系数1;2—载荷系数2;3—载荷系数3

图3 前3个主成分对应的载荷系数对波长图

结合图2与图3可知,载荷系数1曲线代表了全部变量信息中约有90.71%的信息相似,所以曲线整体较为平坦,变化较小,但在1 100 ~ 1 250、1 400、1 672 nm附近有小幅波动,说明这些波段处的差异淹没在大量的相似信息中,故表现不太明显。

由图3可知,载荷系数2和3曲线有较明显的波峰和波谷,表明这些波段处的原始信息中有较大的差异性,而且这些波峰与波谷出现的位置对应图1中虚线框的位置。因此,将PCA分析得到的载荷系数对波长作图,取波峰和波谷位置对应的波段作为特征波长,代表原始光谱的众多信息是可行的^[12-13]。

由图3可知,根据载荷系数2选择的特征波长为1 143.20、1 219.98、1 393.88、1 544.72、1 672.33 nm,根据载荷系数3选择的特征波长为1 193.27、1 672.33 nm,因此综合选1 143.20、1 193.27、1 219.98、1 393.88、1 544.72、1 672.33 nm 6个特征波长。

2 识别分类模型的建立

2.1 基于特征波长的判别分析

判别分析是一种应用广泛的多元统计分析方法,他根据原有的分类信息对所研究的对象进行分类,总结判别分类规律,然后根据判别模型判别未知样本的类别,是一个学习和预测的过程^[14-15]。

对本集中256个塑料样本6个特征波长对应的反射率光谱数据进行判别分析,结果如表1所示。

由表1可知,初始分组案例中的分类准确率为100.0%;对分析中的案例进行交叉验证时,其分类准确率为100.0%,说明使用判别分析方法可以对已知3种样本PE、PP和PET的类别进行准确归类。

表1 样本集判别分析分类结果

	分组	成员1	成员2	成员3	合计
计数	1	88	0	0	88
	2	0	92	0	92
	3	0	0	76	76
初始分组的准确率/%	1	100.0	0.0	0.0	100.0
	2	0.0	100.0	0.0	100.0
	3	0.0	0.0	100.0	100.0
计数	1	88	0	0	88
	2	0	92	0	92
	3	0	0	76	76
交叉验证的准确率/%	1	100.0	0.0	0.0	100.0
	2	0.0	100.0	0.0	100.0
	3	0.0	0.0	100.0	100.0

其中,在交叉验证中,每个案例都是按照从该案例以外的所有其他案例派生的函数来分类的。

2.2 判别函数

判别函数是根据已有分类信息建立的判别准则,可用于对未知样本的分类。在判别分析中,得到典型判别式函数的特征值如表2所示,函数系数如表3所示。

表2 典型判别式函数的特征值

函数	特征值	方差/%	累积/%
1	20.371	60.6	60.6
2	13.221	39.4	100.0

表3 典型判别式函数系数表

	函数1	函数2
变量1	11.171	-9.877
变量2	-28.107	-22.160
变量3	83.603	63.535
变量4	-93.512	-42.848
变量5	1.725	19.543
变量6	28.648	-6.808
常量	-1.921	-0.849

由表2可知,因为分组变量(PE、PP和PET)为3类,因此得到2个判别函数。其中,第1判别式函数解释了光谱数据的60.6%,第2判别式函数解释了光谱数据的39.4%,2个函数解释了数据的100%。

由表3可得到2个典型判别函数式:

$$y_1 = 11.171x_1 - 28.107x_2 + 83.603x_3 - 93.512x_4 + 1.725x_5 + 28.648x_6 - 1.921 \quad (2)$$

$$y_2 = -9.877x_1 - 22.160x_2 + 63.535x_3 - 42.848x_4 + 19.543x_5 - 6.808x_6 - 0.849 \quad (3)$$

2.3 模型的建立

PE、PP和PET这3种塑料的识别分类模型包括:判别式函数、识别质心和识别半径。

将样本集中256个样本的6个变量(即特征波长对应的光谱数据) $x_1 \cdots x_6$ 分别带入到判别式函数,以判别式函数1的判别式得分(即 y_1)为横坐标,判别式函数2的判别式得分(即 y_2)为纵坐标,得到样本集的散点图,结果如图4所示。

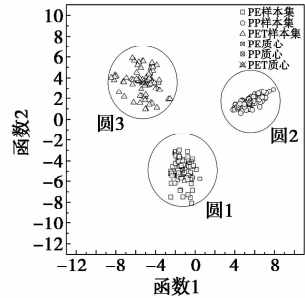


图4 样本集的散点图

由图4可知,同类样本聚集在一起,每1个分组的组质心代表该组中所有点所在位置的坐标,组质心坐标如表4所示。

表4 3种塑料样本在典型判别式函数下的组质心坐标

种类	组质心的横坐标	组质心的纵坐标
1(PE)	-1.264	-4.890
2(PP)	5.590	1.735
3(PET)	-5.304	3.562

其中,每1个分组的识别半径(图4中3个圆的半径)根据分组中距离组质心最远的样本与质心的马氏距离来确定,得到圆1半径为3.5,圆2半径为3,圆3半径为3.5。

2.4 模型的验证

将预测集中的105个样本的6个变量 $x_1 \cdots x_6$ 分别带入到判别函数式中,得到函数1、函数2的值;并以函数1的值为横坐标,函数2的值为纵坐标,并结合图4,得到预测集的散点图,如图5所示。

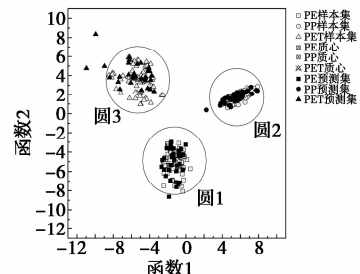


图5 预测集的散点图

在图5中,空心的点代表样本集,实心的点代表预测集。由图5可以看出,每一类预测集样本基本都聚集在相应的质心附近,而且大多数样本均位于各自的识别半径之内。

为了准确验证预测样本的种类,首先分别计算出点 (y_1, y_2) 到质心1、质心2、质心3的距离 d_1 、 d_2 、 d_3 ,然后进行判别。如果 $d_1 \leq 3.5$ 成立,即样本位于圆1之内,则该样本为PE;如果 $d_2 \leq 3$ 成立,即样本位于圆2之内,则该样本为PP;如果 $d_3 \leq 3.5$ 成立,即样本位于圆3之内,则该样本为PET;如果这3个不等式均不成立,则可把该样本归为未知。

经计算可知,105个预测样本中,满足 $d_1 \leq 3.5$ 的样本(即PE样本)有34个,满足 $d_2 \leq 3$ 的样本(即PP样本)有39个,满足 $d_3 \leq 3.5$ 的样本(即PET样本)有27个,3种样本的准确率分别为97.14%、97.5%、90%。该模型的综合识别分类准确率为95.24%,基本满足识别分类的要求,因此可以认为该模型可用来识别分类PE、PP和PET3种塑料。

3 结论

将近红外高光谱数据与化学计量学方法PCA相结合,选取6个特征波长代表全部光谱数据进行分析,减少了冗杂的初始数据量;在数据处理时省去预处理环节,直接对数据进行分析,省去了大量繁琐的计算过程,节省了时间。

根据判别分析的结果,对全部样本进行了准确分类,而且交叉验证正确率为100%,说明该模型可以用于PE、PP和PET的识别分类;使用105个预测样本验证该识别分类模型,结果表明该模型对样本的正确识别率达到95.24%,说明了基于判别分析所建立识别模型的可靠性。

由于各种塑料中的添加剂种类很多,会严重影响光谱数据的采集,因此丰富识别模型中塑料样本的种类与数量可以保证模型的准确性与通用性。随着样本种类的丰富与处理数据量的大幅度减少,该模型可用于在线实时处理提取的光谱校正数据,为快速精准地识别出PE、PP和PET提供理论依据,有利于近红外光谱技术在塑料识别领域的推广与应用。

参考文献

[1] Panda A K, Singh R K, Mishra D K. Thermolysis of waste plastics to liquid fuel: A suitable method for plastic waste management and manufacture of value added products—A world prospective[J]. Re-

newable and Sustainable Energy Reviews, 2010, 14 (1): 233 - 248.

- [2] Kamruzzaman M, Elmasry G, Sun D W, *et al.* Prediction of some quality attributes of lamb meat using near-infrared hyperspectral imaging and multivariate analysis [J]. Analytical Chimica Acta, 2012, 714: 57 - 67.
- [3] Serranti S, Gargiulo A, Bonifazi G. Hyperspectral imaging for process and quality control in recycling plants of polyolefin flakes [J]. Journal of Near Infrared Spectroscopy, 2012, 20 (5): 573 - 581.
- [4] Bonifazi G, D'agostini M, Dall'ava A, *et al.* A new hyperspectral imaging based device for quality control in plastic recycling [C]. Optical Sensors 2013, 2013, 8774: 87741M - 1 - 87741M - 13.
- [5] Elmasry G, Wang N, Vigneault C. Detecting chilling injury in Red Delicious apple using hyperspectral imaging and neural networks [J]. Postharvest Biology and Technology, 2009, 52 (1): 1 - 8.
- [6] Xie C Q, Li X L, Nie P C, *et al.* Application of time series hyperspectral imaging (TS-HSI) for determining water content within tea leaves during drying [J]. Transactions of the Asabe, 2013, 56 (6): 1431 - 1440.
- [7] Serranti S, Gargiulo A, Bonifazi G. Characterization of post-consumer polyolefin wastes by hyperspectral imaging for quality control in recycling processes [J]. Waste Management, 2011, 31 (11): 2217 - 2227.
- [8] Jin J, Tang L, Hruska Z, *et al.* Classification of toxigenic and atoxigenic strains of *Aspergillus flavus* with hyperspectral imaging [J]. Computers and Electronics in Agriculture, 2009, 69 (2): 158 - 164.
- [9] Conde O M, Amado M, Garcia-allende P B, *et al.* Evaluation of PCA dimensionality reduction techniques in imaging spectroscopy for foreign object detection. Algorithms and Technologies for Multispectral [C]. Hyperspectral, and Ultraspectral Imagery X, 2007, 6565: 65650M - 1 - 65650M - 11.
- [10] Williams P, Geladi P, Fox G, *et al.* Maize kernel hardness classification by near infrared (NIR) hyperspectral imaging and multivariate data analysis [J]. Analytica Chimica Acta, 2009, 653 (2): 121 - 130.
- [11] Williams P J, Geladi P, Britz T J, *et al.* Investigation of fungal development in maize kernels using NIR hyperspectral imaging and multivariate data analysis [J]. Journal of Cereal Science, 2012, 55 (3): 272 - 278.
- [12] Elmasry G, Iqbal A, Sun D W, *et al.* Quality classification of cooked, sliced turkey hams using NIR hyperspectral imaging system [J]. Journal of Food Engineering, 2011, 103 (3): 333 - 344.
- [13] Serranti S, Gargiulo A, Bonifazi G. Classification of polyolefins from building and construction waste using NIR hyperspectral imaging system [J]. Resources Conservation and Recycling, 2012, 61: 52 - 58.
- [14] 苏金明, 张莲花, 刘波, 等. MATLAB 工具箱应用 [M]. 4 版. 北京: 电子工业出版社, 2004: 102 - 105.
- [15] Kamruzzaman M, Elmasry G, Sun D W, *et al.* Application of NIR hyperspectral imaging for discrimination of lamb muscles [J]. Journal of Food Engineering, 2011, 104 (3): 332 - 340. ■