

基于近红外光谱分析技术的食品包装塑料的定性分析

田 静¹, 王晓娟², 齐文良¹, 梁振楠¹, 陈 斌^{1*}

(1. 江苏大学 食品与生物工程学院, 江苏 镇江 212013; 2. 宁波海关技术中心, 浙江 宁波 315048)

摘要: 该文基于近红外漫反射光谱分析技术对食品包装材料聚乙烯、聚丙烯进行定性判别试验研究, 选取不同波段范围、采用不同光谱预处理方法, 使用主成分分析法(Principal component analysis, PCA)结合SIMCA、贝叶斯判别、K-近邻3种模式识别方法建立定性预测模型, 并根据正确识别率比较了各模型预测性能。结果表明: 使用SIMCA方法、贝叶斯判别、K-近邻3种方法建立的定性校正模型均在1 050~1 550 nm波长范围内效果较好; 采用矢量归一化、标准正态变量变换、中心化、滑动均值滤波、多项式平滑滤波、一阶微分6种光谱预处理方法和上述3种模式识别方法对塑料样品近红外光谱进行了数据处理, 其中在1 050~1 550 nm范围内, 主成分因子数为3, 采用原始光谱建立的K-近邻定性校正模型较优, 对样品校正集和预测集的正确识别率均为100%。可为食品包装材料聚乙烯、聚丙烯的快速鉴别研究提供参考。

关键词: 近红外光谱; 塑料; 主成分分析; 正确识别率

中图分类号: O657.33 **文献标识码:** A **文章编号:** 1004-4957(2020)11-1416-05

Research on Food Packaging Plastics Based on Near Infrared Spectroscopy

TIAN Jing¹, WANG Xiao-juan², QI Wen-liang¹, LIANG Zhen-nan¹, CHEN Bin^{1*}

(1. School of Food and Biological Engineering, Jiangsu University, Zhenjiang 212013, China; 2. Ningbo Customs Technology Center, Ningbo 315048, China)

Abstract: Based on near infrared diffuse reflectance spectroscopy for the qualitative discrimination on food packaging materials polyethylene and polypropylene using different spectral pretreatment methods with different wavebands, principal component analysis(PCA) combined with three pattern recognition methods of SIMCA, Bayes discriminant and K-Nearest neighbor were adopted to establish three qualitative prediction models, and the prediction performances of the models were compared according to their correct recognition rates, in order to select the best model. Results showed that the qualitative correction models established by SIMCA, Bayes discriminant analysis and K-Nearest neighbor are better in the wavelength range of 1 050-1 550 nm. Six types of spectral preprocessing methods, i. e vector normalization, standard normal variable transformation, Centralization, moving average filtering, Savitzky-Golay filtering and first order differential combined with three pattern recognition methods of SIMCA, Bayes discrimination and K-Nearest neighbors were used to process the near-infrared spectra of plastic samples. In the range of 1 050-1 550 nm, the principal component factor was 3. The qualitative correction model by K-Nearest neighbor using the original spectrum was the best, whose correct recognition rates for the sample's calibration set and prediction set were both 100%. It could provide a reference for the rapid identification of polyethylene and polypropylene as food packaging materials.

Key words: near infrared spectroscopy; plastic; principal component analysis; correct recognition rate

随着塑料化工技术的迅速发展, 塑料制品在生活中的应用愈发广泛, 食品和药品会直接接触到大量塑料制品, 因此塑料制品的安全性将会直接影响消费者身体健康^[1-3]。聚乙烯(Polyethylene, PE)和聚丙烯(Polypropylene, PP)是最广泛应用于食品、药品直接接触的塑料制品, 但由于PE和PP的耐热性和拉伸强度等性质相差较大, 故两者的适用场合和成本差异较大。不法厂商为节约成本, 使用不符合国家标准的塑料制品作为食品和药品的内包装材料, 严重威胁消费者健康, 因此对食用级塑料包装

收稿日期: 2020-06-13; 修回日期: 2020-07-24

基金项目: 国家质量监督检验检疫总局科技计划项目(2017IK275)

* 通讯作者: 陈 斌, 博士, 教授, 研究方向: 食品无损检测技术及装备, E-mail: nep@ujs.edu.cn

材料的鉴别尤为重要。

目前,对塑料制品种类的判别方法有燃烧法、溶解法、静电法、光学法、热分析法等^[4-6],但这些方法在时间、工序及成本上存在很多不足。近红外光谱技术具有快速、绿色、无损、成本低等优点,被广泛应用于聚合材料以及农产品的品质检测^[7]。近红外光谱区是含氢基团(C—H、N—H、O—H)的倍频与合频吸收区,利用该区域的吸收信息可进行含氢基团物质的定性定量检测^[8],将近红外光谱与化学计量学方法结合建立定性定量模型,可快速判别塑料原料的种类和制品品质^[9-10]。本研究基于近红外漫反射光谱分析技术,对食品和药品的外包装常用塑料材料PE和PP进行定性判别,通过选取不同波长范围、不同光谱预处理方法,在主成分分析的基础上,采用多种模式识别算法,建立了定性校正模型,并通过比较各个模型性能的预测效果,优选出最佳预测模型,以期为塑料及塑料制品材质的种类快速鉴别提供参考。

1 实验部分

1.1 实验材料

塑料样品(包括PP和PE)共100个,由宁波市检验检疫局提供。其中,PP新生料17个,PP再生料15个,PE新生料36个,PE再生料32个,样品分类情况见表1。

表1 塑料样品类别个数统计表
Table 1 Statistical table of plastic sample categories

Sample	New material	Renewable material	Total	Correction set			Forecast set			Total number of sample
				New material	Renewable material	Subtotal	New material	Renewable material	Subtotal	
PP material	17	15	32	12	11	23	5	4	9	100
PE material	36	32	68	25	22	47	11	10	21	

1.2 仪器与光谱采集

S450近红外光谱分析仪(上海棱光技术有限公司),仪器内置聚四氟乙烯参比模块,可自动完成参比校正和波长监控,并配有聚四氟乙烯积分球漫反射系统和样品旋转装置,适合粉末状、颗粒状样品的测量。

在室温(25℃)条件下,将装有塑料样品的样品杯置于采集窗口进行光谱采集(波长范围900~2500nm),取每个塑料样品扫描3次的平均光谱进行后续数据分析,共采集100个样品的近红外光谱。

1.3 样品集划分与光谱预处理

塑料样品采用Kennard-Stone(K-S)方法进行样品集划分,校正集与预测集比例为70:30。

在采集塑料样品光谱图的过程中,为减少塑料样品密度与形态不同以及表面特征不均匀等因素的影响,并降低干扰因素产生的噪声,通常需对光谱进行预处理,以提高定性校正模型的预测精确性。采用矢量归一化、标准正态变量变换(Standard normal variate transformation, SNV)、中心化、滑动均值滤波(Moving average filtering, MAF)、多项式平滑滤波(Savitzky-Golay filtering, SGF)、一阶微分方法对光谱预处理^[11-16]。

1.4 数据分析及评价指标

使用江苏大学近红外研究小组自主开发的NIRSA4.5数据处理软件进行光谱预处理和定性校正模型的建立。

采用主成分分析结合簇类的独立软模式方法(Soft independent modeling of class analogy, SIMCA)、贝叶斯判别(Bayes discriminant analysis, BDA)、K-近邻(K-Nearest neighbor, KNN)3种方法建立两类塑料样品的定性校正模型。主成分分析是通过线性变换来提取近红外光谱中主要特征分量的数据降维方法, SIMCA方法是一种建立在主成分分析基础上的模式识别方法,根据模型的拟合结果对其判别归类。贝叶斯判别是通过计算总体各自出现的概率大小,对未知类别样品进行归类,将未知样品归属于出现概率最大的一类。K-近邻是直接利用已有的带有类别标记的样本集(称为训练样本集)分类得

分样本(训练样本集中观察到的邻近样本)。正确识别率(CRR,%)是用于评价定性校正模型精度的指标,CRR越高(越接近100%)表明模型的判别精度越高,具体计算公式为: $CRR = m_1/m_2 \times 100\%$,式中 m_1 为正确识别样品数, m_2 为样品总数。

2 结果与讨论

2.1 塑料样品的近红外漫反射光谱

2类4种PP与PE塑料样品的近红外漫反射原始光谱图如图1所示。由图可见,PE和PP塑料的新生料近红外光谱图无太大差异,均在1213 nm和1394 nm处有较强吸收峰,这可能由样品中C—H基团、C—H₂基团的倍频和合频振动造成。但两类塑料的再生料与新生料在1000~1800 nm波段范围内的原始光谱吸光度相差较大,可能由于再生料因反复使用夹带了一定的杂质。因两类塑料的吸收峰形和位置均较为相似,无法通过近红外图谱直观鉴别,需将样品的光谱数据结合化学计量学方法进行分析及判别。

2.2 塑料样品近红外光谱主成分的提取

在主成分提取中,累计贡献率能够反映前 n 个主成分对原始数据的信息表征能力,塑料样品近红外光谱的前10个主成分(PC₁~PC₁₀)的累计贡献率趋势图见图2,图中可见前3个主成分(PC₁、PC₂、PC₃)的累计贡献率为99.22%,已接近100%,表明前3个主成分已可表征大部分光谱信息,因此数据分析均选取塑料样品前3个主成分数建立各种判别定性校正模型。

2.3 3种分类定性校正模型的精度分析

2.3.1 波段对定性校正模型判别精度的影响

不同的波长范围所含有的塑料样品属性信息不同,建立的定性校正模型判别精度也不同。塑料样品原始光谱选取不同波长区域的3种分类方法建立的模型预测结果见表2。由表中数据可见,在不同波长区间内,3种方法建立的定性校正模型均在1050~1550 nm波长范围内的效果较好,此区域内SIMCA方法校正集和预测集的正确识别率分别为91.4%和86.7%,贝叶斯判别的校正集和预测集的正确识别率分别为74.3%和63.3%,K-近邻方法的校正集和预测集的正确识别率分别为100%和100%。

表2 原始光谱不同波长区间的3种分类定性模型分类错误的样品数统计分析

Table 2 Statistical analysis of the number of wrong samples of three classification qualitative models with different wavelength intervals in the original spectrum

Pattern recognition method	Wavelength range (nm)	Principal component contribution rate(%)	Number of misclassified samples (PCS)		CRR(%)	
			Correction set	Forecast set	Correction set	Forecast set
SIMCA	1 050 ~ 1 550	99.83	6	4	91.4	86.7
	1 000 ~ 1 700	99.76	10	5	85.7	83.3
	1 000 ~ 2 250	99.63	9	5	87.1	83.3
	1 000 ~ 2 500	99.22	11	4	84.3	86.7
Bayes	1 050 ~ 1 550	99.83	18	11	74.3	63.3

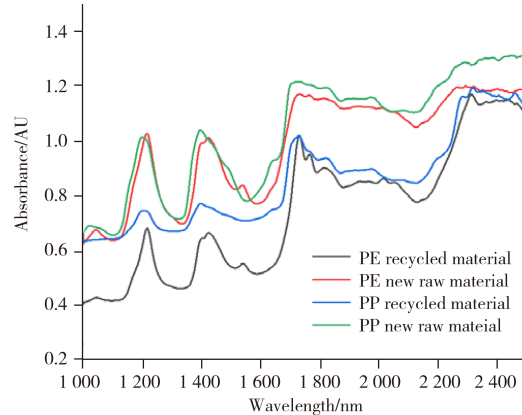


图1 塑料样品的原始近红外光谱图

Fig. 1 Raw NIR spectra of plastic samples

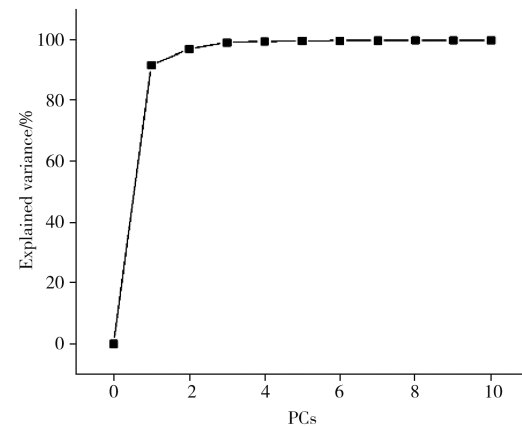


图2 原始光谱前10个主成分的累计贡献率趋势图

Fig. 2 Cumulative contribution rate trend of the first 10 principal components extracted from the original spectrum

(续表2)

Pattern recognition method	Wavelength range (nm)	Principal component contribution rate(%)	Number of misclassified samples(PCS)		CRR(%)	
			Correction set	Forecast set	Correction set	Forecast set
KNN	1 000 ~ 1 700	99.76	21	8	70.0	73.3
	1 000 ~ 2 250	99.63	40	20	42.9	33.3
	1 000 ~ 2 500	99.22	21	10	70.0	66.7
	1 050 ~ 1 550	99.83	0	0	100.0	100.0
	1 000 ~ 1 700	99.76	0	11	100.0	63.3
	1 000 ~ 2 250	99.63	0	11	100.0	63.3
	1 000 ~ 2 500	99.22	0	6	100.0	80.0

2.3.2 预处理方法对定性校正模型判别精度的影响 在优化得出的最优波长区域内, 使用 SIMCA 方法和贝叶斯判别 2 种识别方法结合不同预处理方法建立定性校正模型, 比较不同光谱预处理方法对定性校正模型判别精度的影响, 进一步提高模型的判别精度。表 3 为 1 050 ~ 1 550 nm 波段范围采用不同光谱预处理方法的上述 2 种方法定性校正模型的统计分析结果。由表中数据可见, SIMCA 方法结合 SNV + 5 点 MAF 平滑、SNV + 5 点 SGF 平滑或者 SNV + 5 点 SGF 平滑预处理方法预测模型判别精度最好, 预测模型精度得以显著提高, 样品校正集和预测集的正确识别率分别为 95.7% 和 93.3%, 表明该预处理方法可提取样品近红外光谱中的有效信息, 减少了噪音干扰, 提高了预测模型精度; 贝叶斯判别结合 SNV + 5 点 MAF 平滑光谱预处理的预测模型判别精度最好, 其中样品校正集和预测集的正确识别率分别为 97.1% 和 93.3%。

表3 不同预处理方法的 2 种分类定性模型的分错误样品数统计分析

Table 3 Statistical analysis of the number of wrong samples of two classification qualitative models with different pretreatment methods

Pattern recognition method	Spectral data type	Principal component contribution rate(%)	Number of misclassified samples(PCS)		CRR(%)	
			Correction set	Forecast set	Correction set	Forecast set
SIMCA	Original spectrum	99.83	6	4	91.4	86.7
	SNV + 5 point MAF smoothing	91.67	3	2	95.7	93.3
	SNV + 5 point SGF smoothing	91.64	3	2	95.7	93.3
	SNV + 5 point first order differential	84.33	5	3	92.9	90.0
	Normalized + 5 point MAF smoothing	97.93	4	2	94.3	93.3
	Normalized + 5 point SGF smoothing	97.94	4	2	94.3	93.3
	Centralized + 5 point MAF smoothing	99.83	3	2	95.7	93.3
Bayes	Original spectrum	99.83	18	11	74.3	63.3
	SNV + 5 point MAF smoothing	91.67	2	2	97.1	93.3
	SNV + 5 point SGF smoothing	91.64	6	2	91.4	93.3
	SNV + 5 point first order differential	84.33	2	3	97.1	90.0
	Normalized + 5 point MAF smoothing	97.93	8	4	88.6	86.7
	Normalized + 5 point SGF smoothing	97.94	8	4	88.6	86.7
	Centralized + 5 point MAF smoothing	99.83	17	12	75.7	60.0

2.4 最优定性校正模型分类错误的样品数统计

表 4 为 3 种分类方法在不同波段范围、不同预处理方法下得到的最优定性校正模型各类样品分类错误的样品数统计。由表 4 数据可见, 在 1 050 ~ 1 550 nm 波段范围内, SIMCA 方法采用 SNV + 5 点 MAF 平滑或者 SNV + 5 点 SGF 平滑光谱预处理方法的预测模型对于 PP 新生料与再生料样品区分效果较好; 贝叶斯判别采用 SNV + 5 点 MAF 平滑预处理方法的预测模型对于 PP 新生料与再生料区分效果较好; K - 近邻方法采用原始光谱建立的预测模型对 2 类 4 种 PP 与 PE 塑料样品的分类效果最好。

表 4 3 种分类方法的最优定性校正模型分类错误的样品数统计
Table 4 Statistics of the number of misclassified samples of the optimal qualitative correction model under the three classification methods

Pattern recognition method	Spectral data type	Number of misclassified samples (PCS)							
		Correction set				Forecast Set			
		PP new raw material	PP recycled material	PE new raw material	PE recycled material	PP new raw material	PP recycled material	PE new raw material	PE recycled material
SIMCA	SNV +5 point MAF smoothing	0	0	1	2	0	0	2	0
	SNV +5 point SGF smoothing	1	0	1	1	0	0	1	1
Bayes	SNV +5 point MAF smoothing	1	0	0	1	0	0	2	0
KNN	Original spectrum	0	0	0	0	0	0	0	0

3 结 论

本研究采用主成分分析法结合 3 种不同模式识别方法(SIMCA、贝叶斯判别、K-近邻)建立了 PP 和 PE 定性校正模型,结果显示:3 种模式识别方法在 1 050 ~ 1 550 nm 波长内建立的定性模型判别精度最好;采用矢量归一化、标准正态变量变换、中心化、滑动均值滤波、多项式平滑滤波、一阶微分 6 种光谱预处理方法和上述 3 种模式识别方法对塑料样品近红外光谱进行数据处理发现,在 1 050 ~ 1 550 nm 范围内未经光谱预处理,前 3 个主成分空间下的 K-近邻定性校正模型判别精度最好,样品校正集和预测集的正确识别率均接近 100%。该检测方法具有结果准确、耗时短、无损等优点,可为食品包装材料聚乙烯、聚丙烯的快速鉴别提供新方法。

参考文献:

- [1] Zhao M J, Zhou W. *Plast. Sci. Technol.* (赵敏娟, 周伟. 塑料科技), **2015**, 43(9): 92-95.
- [2] Pholisa D, Marena M, Louwrens H, Paul J W. *Food Control*, **2020**, 110: 110056.
- [3] Qian L L, Song X J, Lei Y B, Wang Y, Zuo F, Zhang L Y. *Food Ind.* (钱丽丽, 宋雪健, 类彦波, 王月, 左锋, 张丽媛. 食品工业), **2018**, 39(6): 257-261.
- [4] Zou X B, Feng T, Zheng K Y, Shi J Y, Huang X W, Sun Y. *Spectrosc. Spectral Anal.* (邹小波, 封韬, 郑开逸, 石吉勇, 黄晓玮, 孙悦. 光谱学与光谱分析), **2019**, 39(5): 1445-1450.
- [5] Bai J, Li J P, Zou H. *Food Sci.* (白京, 李家鹏, 邹昊. 食品科学), **2019**, 40(8): 287-292.
- [6] Lü J, Wang D Y. *China Mod. Med.* (吕晋, 王黛莹. 中国当代医药), **2018**, 25(33): 40-43, 58.
- [7] Wang J J. *Research on Rapid Detection Technology for Food Packaging Plastic Based on NIR*. Yiwu: China Jiliang University(王晶晶. 基于近红外光谱的食品包装用塑料快速检测技术的研究. 义乌: 中国计量学院), **2013**.
- [8] Ma G Z, Yang X L, Wang W C, Chen Y X. *Yunnan Chem. Technol.* (马公喆, 杨晓丽, 汪文超, 陈云秀. 云南化工), **2015**, 42(1): 45-47.
- [9] Hou Y L, Xie L G, Zhao X, Sun H M. *Chin. J. Pharm.* (侯玉磊, 谢兰桂, 赵霞, 孙会敏. 中国医药工业杂志), **2015**, 46(6): 614-619.
- [10] Chu X L, Shi Y Y, Chen P, Li J Y, Xu Y P. *J. Instrum. Anal.* (褚小立, 史云颖, 陈瀑, 李敬岩, 许育鹏. 分析测试学报), **2019**, 38(5): 603-611.
- [11] Shen X, Zheng X, Song Z Q, He D P, Qi P S. *Computer and Computing Technologies in Agriculture VI—Proceedings of 6th IFIP WG 5.14 International Conference(CCTA2012) Part I*, **2012**: 28-35.
- [12] Liang L, Fang G G, Wu T, Cui H H, Zhang X M, Zhao Z Y. *J. Instrum. Anal.* (梁龙, 房桂干, 吴珽, 崔宏辉, 张新民, 赵振义. 分析测试学报), **2016**, 35(1): 101-106.
- [13] Zhang C X, Ma Y. *Appl. Chem. Ind.* (张翠侠, 马钺. 应用化工), **2016**, 45(9): 1795-1797, 1802.
- [14] Long C F. *Digital Technol. Appl.* (龙草芳. 数字技术与应用), **2017**, (4): 145-146, 150.
- [15] Elizabeth N W, Irene N W, Carolien B, Nghia-Do-Trong N, Peter K K, Wouter S, Daniel N S, Marc H. *J. Food Eng.*, **2020**, 284: 106981.
- [16] Zhang Y M, Wang P, Bai J R, Ma D Y. *Mod. Chem. Ind.* (张毅民, 王鹏, 白家瑞, 马冬雅. 现代化工), **2016**, 36(3): 182-186.

(责任编辑:周启动)