

## 基于红外光谱的城市生活垃圾高值化利用深度分选

胡 斌<sup>1,2</sup>, 付 浩<sup>1</sup>, 王文斌<sup>1</sup>, 张 兵<sup>1,2</sup>, 唐 帆<sup>3\*</sup>, 马善为<sup>1,2</sup>, 陆 强<sup>1,2\*</sup>

1. 华北电力大学新能源学院, 北京 102206
2. 华北电力大学生物质发电成套设备国家工程实验室, 北京 102206
3. 吉林大学人工智能学院, 吉林 长春 130012

**摘 要** 红外光谱分析具有快速、精确度高等优点, 在分类鉴别领域中发挥着重要作用。红外光谱在生活垃圾分类领域的应用主要集中在塑料等可回收垃圾而忽略了对不可回收垃圾的深度分选。现行生活垃圾的四分法分类中, 源头分类得到的其他垃圾中含有多种具有高值化利用潜力的组分, 可分为纤维素类、烯类聚合物、木竹类等。这些垃圾的成分和结构不同, 因此具有不同的红外谱图, 基于其红外谱图特征波段利用机器学习建立相应的分类模型可以将上述几类垃圾从其他垃圾中分选出来。本研究收集了纤维素类、烯类聚合物、木竹类及低值类垃圾, 并采集红外光谱数据共 72 组, 对比分析了预处理方式、降维程度和建模算法对模型分类准确率的影响。利用标准正态变量变换(SNV)、多元散射校正(MSC)、求导(DC)和平滑滤波(Smooth)方法对数据进行预处理, 然后利用主成分分析法(PCA)对预处理后的数据进行降维, 获得  $72 \times 8$  和  $72 \times 5$  的数据集。分别利用概率神经网络(PNN)、广义回归神经网络(GRNN)、支持向量机(SVM)和随机森林(RDF)算法进行建模。分析结果表明, 经 PCA 降维后的数据用于后续建模时, 5 维数据比 8 维数据得到的分类效果更好, 平均准确率上升 2.4%~4.4%。基于 5 维降维数据, DC/Smooth 预处理方法比 SNV 和 MSC 预处理得到的平均准确率更高, 达到了 96.5%; PNN 模型比其他三类模型的平均准确率高 4.2%~6.5%, 可达 98.1%; 针对四类垃圾, 除烯类聚合物的平均判别率只有 93.8%外, 纤维素类及木竹类的平均判别率均在 95%以上, 低值类最高可以到达 100%。验证了红外光谱结合机器学习实现其他垃圾深度分选的可能性及科学性, 为未来开发快捷深度分选设备提供参考。

**关键词** 红外光谱; 机器学习; 其他垃圾; 深度分选; 高值化利用

**中图分类号:** O433.4 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2022)05-1353-08

### 引 言

红外光谱是分析物质成分的有力工具, 广泛应用于食品、化工、现代医学等领域, 具有所需样本量小、不破坏样本、快速、简便、精确度高等优点<sup>[1-2]</sup>。随着计算机科学的发展, 红外光谱与机器学习的联合使用在分类鉴别领域得到了广泛应用<sup>[3-5]</sup>。在垃圾分类研究中, 基于红外光谱检测与机器学习建模的分类鉴别方法主要用于可回收垃圾的精细分选。不同类别塑料垃圾的近红外光谱特征波段具有显著差异, 利用机器学习方法可以有效地识别不同种类的塑料, 进行精细分选, 且目前已有多种近红外光谱塑料分选设备问世<sup>[6]</sup>。然而, 近红外光谱对于深色垃圾的分选仍存在一定的

局限性。近期, 赵冬娥等<sup>[7]</sup>利用高光谱成像和光谱角填图法、Fisher 判别方法提出了纸质、塑料和木质三类可回收垃圾的分类方法, 准确率超过 97%, 为城市可回收类垃圾的高效分选打下了基础。

随着我国城市垃圾源头分类“四分法”的稳步推进, 有害垃圾、厨余垃圾和可回收垃圾得到了有效分离, 剩余的其他垃圾组分十分复杂, 资源化利用难度较大, 其主流处理方法仍为较为粗犷的焚烧发电和卫生填埋。事实上, 其他垃圾的总量十分可观, 通常占生活垃圾总量的 30%<sup>[8]</sup>, 其中含有废纸、废塑料、废橡胶、织物、木竹等多种有机组分, 这些组分可以进一步高值化利用。举例来说, 废纸中主要成分为纤维素, 利用固体酸催化热解纤维素可以高选择性地制备左旋葡聚糖<sup>[9]</sup>; 塑料和化纤织物等烯类聚合物可以经催化热解

收稿日期: 2021-03-01, 修订日期: 2021-07-27

基金项目: 国家自然科学基金优秀青年科学基金项目(51922040), 中国博士后科学基金面上项目(2020M680482)资助

作者简介: 胡 斌, 1992 年生, 华北电力大学新能源学院讲师 e-mail: binhu92@126.com

\* 通讯作者 e-mail: tfan.108@gmail.com; qiangu@mail.ustc.edu.cn

联产碳纳米管和富氢气体<sup>[10]</sup>；木竹类是典型的生物质资源，恰当的预处理后可以热解生产高品质的生物油或生物炭<sup>[11]</sup>。这些组分的红外光谱特征波段差异显著，彼此之间具有较大的区分度<sup>[6, 12-13]</sup>，可以基于红外光谱数据利用机器学习方法建立其他垃圾深度分选模型，然而，相关的分类模型还少有报道。

现阶段，基于红外光谱和机器学习方法从源头分类获得的其他垃圾中深度分选出高值组分的研究还有相当大的发展空间，特别是高效机器学习模型仍有待建立。依据其他垃圾的主要成分和利用手段，可以将其分为：纤维素类、烯类聚合物、木竹类、低值类。本研究将依靠红外谱图和机器学习，建立其他垃圾高值化利用的深度分选模型，为未来城市生活垃圾的自动化分选以及高值化利用提供科学依据。

## 1 实验部分

### 1.1 红外光谱数据获取

采集纤维素类、烯类聚合物、木竹类及低值类四类其他垃圾样本共 18 件。样本选取上尽可能接近实际生产生活中其他垃圾受污染程度，具体实验材料如表 1 所示。四类样本的特征如下：(1)纤维素类，主要成分是纤维素及其衍生物，收集的样本包括打印纸、草纸、一次性纸杯、棉布、烟头等；(2)烯类聚合物，多数为受污染的塑料或人造织物，具有耐腐蚀且难降解的特点，收集的样本包括方便面包袋、食品包装袋、快餐包装纸、奶茶杯、腈纶标签等；(3)木竹类，以自然界植物枝干叶子为主，主要成分为纤维素、半纤维素、木质素等，与纤维素类的主要区别是含有较高比例的木质素，并非仅含有单一的纤维素组分，收集的样本包括竹扇、落叶、干树枝、木质铅笔、一次性筷子等；(4)低值类，无机物含量相较其他类别较高，深度分选后无更多高值化利用方式，主要以卫生填埋或者焚烧处理，收集的样本包括棒骨、陶瓷、贝壳等。

表 1 其他垃圾实验材料  
Table 1 Residual waste materials

高值化类别	材料名称
纤维素类	打印纸、草纸、一次性纸杯、棉布、烟头
烯类聚合物	方便面包袋、食品包装袋、快餐包装纸、奶茶杯、腈纶标签
木竹类	竹扇、落叶、干树枝、木质铅笔、一次性筷子
低值类	棒骨、陶瓷、贝壳

使用美国 PerkinElmer 公司生产的 Spectrum 100N FT-IR 傅里叶变换红外光谱仪，选用衰减全反射红外光谱技术采集光谱数据。每个样本分别选取四个不同特征点，每个特征点采集 6 次光谱信息，取平均值为该点的光谱反射率数据。红外光谱实验背景为空气，光谱范围是 650~4 000 cm<sup>-1</sup> 波段，分辨率为 4 cm<sup>-1</sup>。最终，18 件样本共得到 72 组光谱数据，每组光谱数据为 1×3 351 的一维线性矩阵，整体构成 72×3 351 的光谱反射率数据矩阵，如式(1)所示，其中  $m=$

72，为实验样本总数， $n=3 351$ ，为每个样本对应 650~4 000 cm<sup>-1</sup> 波段的红外光谱反射率数据。

$$\mathbf{X}_{m \times n} = \begin{bmatrix} X_{11} & \cdots & X_{1n} \\ \vdots & \ddots & \vdots \\ X_{m1} & \cdots & X_{mn} \end{bmatrix} \quad (1)$$

### 1.2 数据预处理

为消除实验室光源及仪器发热等干扰因素带来的噪音，分别采用标准正态变量变换 (standard normal variate, SNV)、多元散射校正 (multiplicative scatter correction, MSC)、导数处理联合平滑滤波 (derivative correction/smooth, DC/Smooth) 对原始光谱数据进行预处理。其中，SNV、MSC 主要用于消除在光谱数据采集过程中固体颗粒大小、表面散射及光程变化而产生的影响；DC/Smooth 联合预处理可以有效处理红外光谱数据中的高频噪音和基线平移，提高光谱数据的灵敏度与分辨率<sup>[14]</sup>。

### 1.3 数据降维

对预处理后的光谱数据以主成分分析法<sup>[15]</sup> (principal composition analysis, PCA) 进行降维处理。以预处理光谱反射率数据 72×3 351 矩阵作为输入参数，计算数据矩阵的协方差矩阵，求解其特征值及特征向量，选择其中  $k$  个特征值所对应的特征向量，构成矩阵  $\mathbf{W}_{n \times k}$ 。以式(2)计算得到降维后的数据  $\mathbf{Z}_{m \times k}$ ，其中  $k$  为 PCA 处理后的数据维度。以式(3)和式(4)计算降维后数据贡献度，其中  $Z_i$  为第  $i$  个主成分， $a_1$  为对应样本集标准化矩阵特征值  $\lambda_i$  的特征向量。使  $k$  个主成分的累计贡献度  $\alpha_i$  超过 90%，第  $k+1$  及以后的累计主成分贡献率小于 10%。

$$\mathbf{Z}_{m \times k} = \mathbf{X}_{m \times n} \mathbf{W}_{n \times k} \quad (2)$$

$$Z_i = a_1 x_1 + a_2 x_2 + \cdots + a_k x_k \quad (3)$$

$$\alpha_i = \frac{\lambda_i}{\sum \lambda_i} \quad (4)$$

### 1.4 深度分选模型的建立

为筛选合适的其他垃圾深度分选模型，采用 4 种具有代表性的分类判别方法进行对比：概率神经网络 (probabilistic neural network, PNN)、广义回归神经网络 (general regression neural network, GRNN)、支持向量机 (support vector machine, SVM) 及随机森林 (random decision forests, RDF)。PNN 与 GRNN 鉴别模型具有优秀的非线性映射能力及学习速度，在处理少量样本数时，判别效果很好，处理不稳定数据集时也有较好效果；SVM 模型具有较为良好的泛用性，在面对分类条件复杂时，具有突出的判别能力；RDF 是使用多棵决策树对样本训练并判别的一种分类器，该算法参数选择较少，不需要担心过度拟合，并拥有较强的抗噪声本领。

为弥补样本量少的缺点和提高最终测试结果的可信度，采用留一法交叉验证用于模型的建立。每个模型以准确率 (Accuracy)、均值及标准误差作为评判标准，准确率计算方法如式(5)所示，其中  $TP$  和  $FP$  分别代表测试样本中被正确分类的样本个数与被错误分类的样本个数。

$$\text{Accuracy} = \frac{TP}{TP + FP} \quad (5)$$

## 2 结果与讨论

### 2.1 数据预处理

采集的原始红外光谱如图 1 所示，对比文献<sup>[6, 12-13]</sup>可知，采集的四类垃圾样本的红外数据在各类垃圾中具有典型性，能够较好地涵盖上述四类垃圾的特征红外波段。图 1(a)中，纤维素类样本在 2 800~2 900  $\text{cm}^{-1}$  附近存在两个较强的吸收峰，是甲基、亚甲基中碳氢键的振动吸收峰，1 100~1 200  $\text{cm}^{-1}$  附近存在的强吸收峰，是醚键、羟基中碳氧键的

吸收峰；图 1(b)所示的烯类聚合物样本在 800 和 1 300  $\text{cm}^{-1}$  处是不饱和 C=C 键的吸收峰；图 1(c)中，木竹类样本与纤维素类样本的明显区别在 1 600  $\text{cm}^{-1}$  附近的吸收峰，是其所含木质素组分苯环的吸收峰、3 300  $\text{cm}^{-1}$  附近是羟基及其形成的氢键所对应吸收峰；而图 1(d)所示的低值类样本由于无机物含量较高，除 1 000  $\text{cm}^{-1}$  附近碳氢吸收峰以外，其他吸收峰均相对平缓。由此可见，四类样本具有不同的红外光谱特征，这是分类模型建立的重要前提；此外，机器学习方法还能够基于人为难以鉴别的特征建立精细的分类模型。

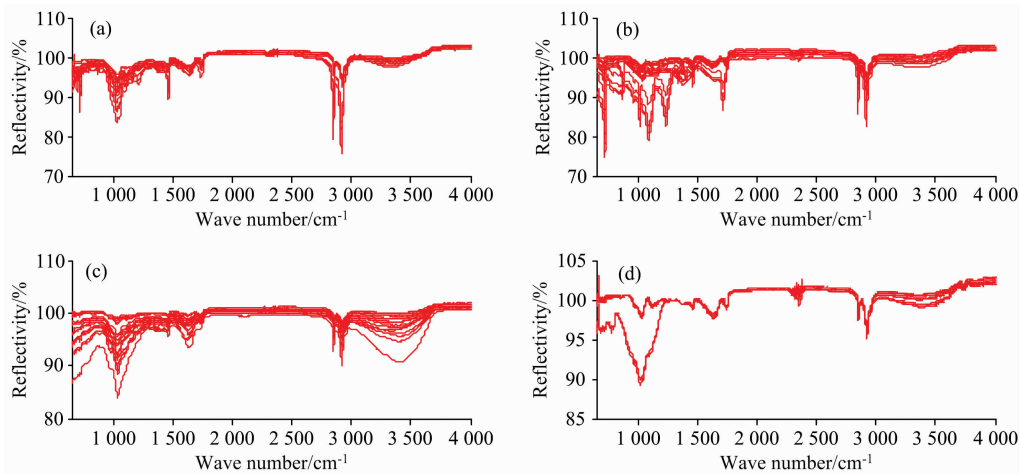


图 1 纤维素类(a)、烯类聚合物(b)、木竹类(c)和低值类(d)的原始红外光谱

Fig. 1 Original infrared spectra of cellulose (a), vinyl polymers (b), woods (c) and low-value wastes (d)

分别采用 SNV, MSC 及 DC/Smooth 对原始红外光谱进行预处理。如图 2 和图 3 所示，经 SNV 和 MSC 预处理后规避掉了许多不必要的杂乱数据，使光谱数据整齐有序。此外，两种预处理方法的效果很接近，因此最终分类模型的预测结果应当也较为接近，这将在后续进一步讨论。如图 4 所

示，经 DC/Smooth 联用预处理后，红外光谱谱图具有明显变化，低值类垃圾的谱图区分度更加明显。对比上述预处理后的光谱数据，可以发现四类垃圾的光谱数据在特征波段仍然具有显著的区别。

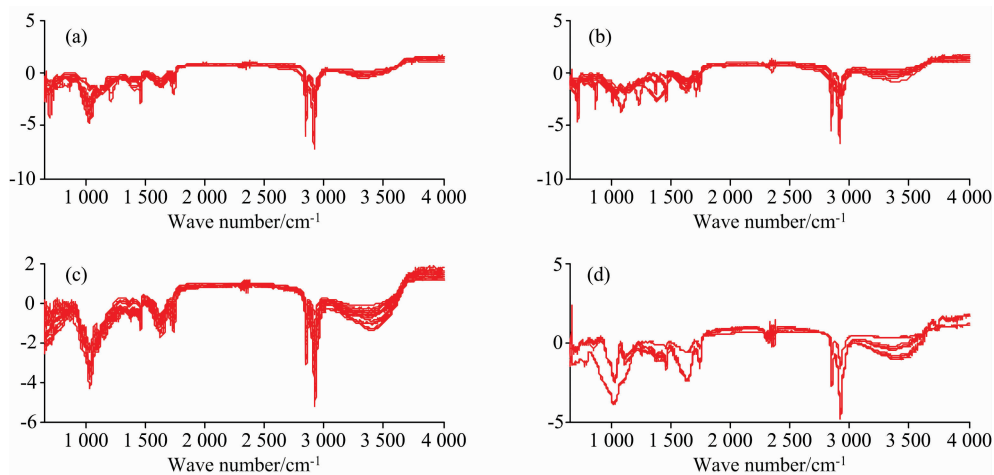


图 2 SNV 预处理后的纤维素类(a)、烯类聚合物(b)、木竹类(c)和低值类(d)红外光谱

Fig. 2 SNV pretreated infrared spectra of cellulose (a), vinyl polymers (b), woods (c) and low-value wastes (d)

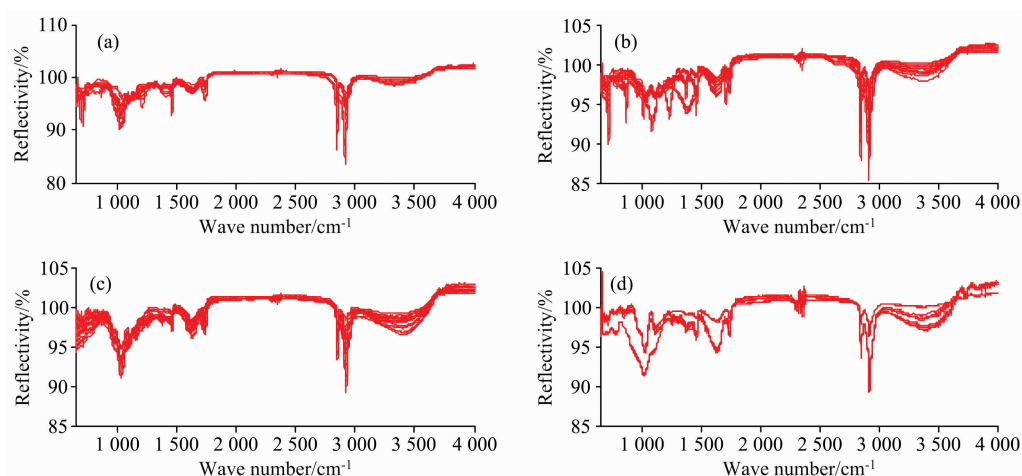


图 3 MCS 预处理后的纤维素类(a)、烯类聚合物(b)、木竹类(c)和低值类(d)红外光谱

Fig. 3 MCS pretreated infrared spectra of cellulose (a), vinyl polymers (b), woods (c) and low-value wastes (d)

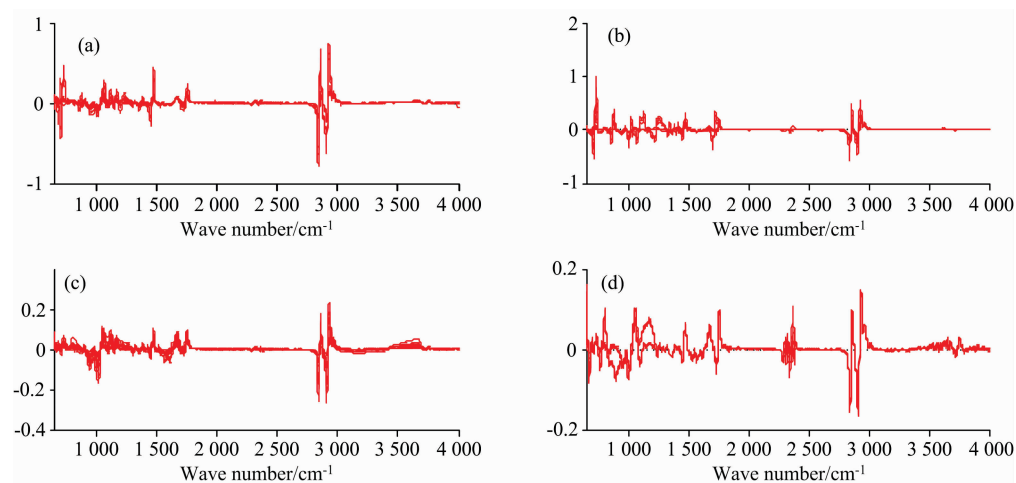


图 4 DC/Smooth 预处理后的纤维素类(a)、烯类聚合物(b)、木竹类(c)和低值类(d)红外光谱

Fig. 4 DC/Smooth pretreated infrared spectra of cellulose (a), vinyl polymers (b), woods (c) and low-value wastes (d)

## 2.2 数据降维

SNV, MSC 及 DC/Smooth 预处理数据经 PCA 降维后得到主成分特征值和方差贡献率, 前 8 维主成分数据列于表 2。

由表 2 可知, 预处理后的数据经过降维后的前 5 维主成

分累计贡献率均达到 90% 以上, 第 8 维及以后主成分的贡献率低于 1%。其中, SNV 预处理后的红外光谱数据经 PCA 降维后得到的 8 维和 5 维数据, 对原始红外光谱数据的贡献率分别达到了 96.5% 和 91.0%。类似地, MSC 预处理后的数据

表 2 SNV, MSC 及 DC/Smooth 预处理数据集经 PCA 处理后主成分的特征值和方差贡献率

Table 2 The principal component eigenvalues and variance contribution of SNV, MSC and DC/smooth datasets

预处理方法	主成分	$Z_1$	$Z_2$	$Z_3$	$Z_4$	$Z_5$	$Z_6$	$Z_7$	$Z_8$
SNV	特征值	208.35	102.04	50.50	39.22	32.71	14.99	6.91	4.15
	方差贡献率/%	43.7	21.4	10.6	8.2	7.1	3.1	1.5	0.9
	累计方差贡献率/%	43.7	65.1	75.7	83.9	91.0	94.1	95.6	96.5
MSC	特征值	861.19	429.45	207.22	161.67	107.20	47.83	36.77	24.15
	方差贡献率/%	44.6	22.2	10.7	8.4	5.6	2.5	1.9	1.2
	累计方差贡献率/%	44.6	66.8	77.5	85.9	91.5	94.0	95.9	97.1
DC/Smooth	特征值	149.73	80.39	16.76	14.34	6.66	4.98	3.28	2.50
	方差贡献率/%	52.2	28.0	5.8	5.0	2.3	1.8	1.1	0.9
	累计方差贡献率/%	52.2	80.2	86.0	91.0	93.3	95.1	96.2	97.1

经 PCA 降维得到  $73 \times 8$  和  $72 \times 5$  的数据集，贡献率分别为 97.1% 和 91.5%；DC/Smooth 预处理后的数据经 PCA 降维得到  $72 \times 8$  和  $72 \times 5$  的数据集，贡献率分别为 97.1% 和 93.3%。对比 5 维和 8 维数据，5 维数据更加简练并可以反映原始数据的大部分信息，而 8 维数据相比于 5 维数据对原始数据的累计贡献率更高(3.8%~5.6%)，后续将分别以 5 维和 8 维数据用于模型建立，进一步探讨数据降维程度对模

型的影响。

由表 2 可以看出，第 1、2 维主成分对于原始数据的贡献率最高，提取以上三种预处理数据降维后第 1、2 主成分的载荷因子，取平均值绘制载荷因子图谱，如图 5 所示。由图可知，在 680, 1 000, 1 200, 1 300, 1 500, 1 650, 2 300, 2 800~2 900 及 3 300  $\text{cm}^{-1}$  等处具有明显振动，说明这些波段在数据中发挥更高的作用。

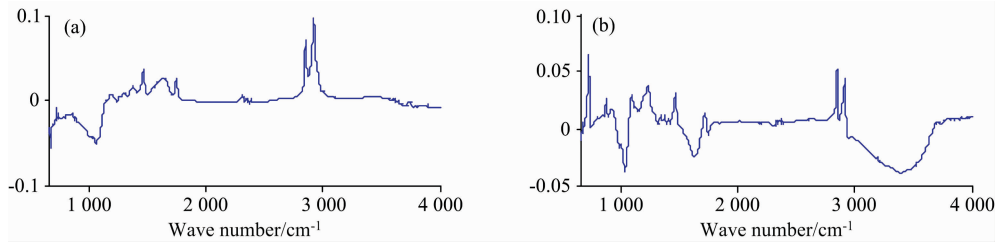


图 5 经 PCA 降维后的第 1(a) 和 2(b) 主成分载荷分析谱图

Fig. 5 The first (a) and second (b) principal component load analysis spectra

### 2.3 深度分选模型判别结果分析

以 PCA 降维后得到的  $72 \times 8$  和  $72 \times 5$  数据集作为输入参数，分别建立 PNN, GRNN, RDF 及 SVM 判别模型，结果分别列于表 3 和表 4。

由表 3 可知，红外光谱数据预处理后所建立的高值化深度分选模型平均准确率接近甚至超过 90%，相较未经预处理的模型平均准确率上升 5.5%~11.2%，其中 SNV, MCS 及 DC/Smooth 三种预处理方式的平均准确率为 89.4%，88.4% 及 94.1%。如图 2 和图 3 所示，SNV 和 MCS 预处理后得到的光谱特征波段相近，导致基于两种预处理方式所建分类模型的鉴别能力相当。DC/Smooth 联合预处理获得的预测平均准确率最高，这是由于导数处理(DC)可有效消除其他背景的干扰，分辨重叠峰，提高分辨率和灵敏度；Smooth 可以防止导数处理的信噪比降低、部分噪声放大。对比表 3 和表 4 中数据发现，不同数据降维程度下，三类预处理方式对应的分类准确率相对关系是一致的。相比之下， $72 \times 5$  数据集得到的预测准确率整体有所提高，SNV, MCS 及 DC/Smooth 三种预处理方式对应的平均准确率分别为 93.8%，92.4% 及 96.5%，准确率提高了 2.4%~4.4%。表 3 中，DC/Smooth 联合预处理的准确率均方根误差较未处理有所上升，而 SNV 和 MSC 预处理后的均方根误差明显降低；对于 5 维数据，SNV 和 MSC 预处理的均方根误差较未预处理有所升高，而 DC/Smooth 联合预处理有所下降。由表 2 可知，三种预处理方式得到的数据经过降维后的前 5 维主成分累计贡献率即达到 90%，而第 6~8 维数据对于原始数据的贡献很低(0.9%~3.1%)，这表明 8 维数据相比于 5 维数据增加了无效数据，尽管对原始数据的贡献率更高，但是预测效果反而不如 5 维数据。针对同一预处理方法所得数据集，对比四类分类模型预测准确率的均方根误差，8 维数据得到的分类模型整体稳定性更高，这是因为 5 维数据去掉了更多的次要信息，从而使得不同建模算法更具区分度，更利于筛选合适的建模算法。

对比表 3 中四类分类模型可以看出，SVM 模型平均准确

表 3 分类模型的准确率对比(基于  $72 \times 8$  数据集)

Table 3 Comparison of classification model accuracies (based on  $72 \times 8$  dataset)

预处理方式	分类准确率/%				均值 /%	均方根误差/%
	PNN	GRNN	RDF	SVM		
未预处理	78.7	85.7	85.7	81.5	82.9	3.4
SNV	88.7	90.1	88.7	90.1	89.4	0.8
MSC	87.3	90.1	87.3	88.7	88.4	1.3
DC/Smooth	95.8	87.5	95.8	97.2	94.1	4.4
均值*	90.6	89.2	90.6	92.0	—	—
均方根误差*	4.6	1.5	4.6	4.6	—	—

注：\* 均值和均方根误差是指预处理后光谱数据所建模型的均值和均方根误差，下同

Note: \* Average values and mean-square errors are based on the pre-processed data, the some below

表 4 分类模型准确率对比(基于  $72 \times 5$  数据集)

Table 4 Comparison of classification model accuracies (based on  $72 \times 5$  dataset)

预处理方式	分类准确率/%				均值 /%	均方根误差/%
	PNN	GRNN	RDF	SVM		
未预处理	77.8	84.7	84.7	87.5	83.6	4.1
SNV	95.8	100.0	88.9	90.3	93.8	5.1
MSC	98.6	88.9	90.2	91.7	92.4	4.3
DC/Smooth	100.0	94.4	95.8	95.8	96.5	2.4
均值*	98.1	94.4	91.6	92.6	—	—
均方根误差*	2.1	5.6	3.6	2.8	—	—

率最高，达到了 92.0%；PNN 模型与 RDF 模型的平均准确率较 SVM 模型低，但整体准确度在生产生活实践接受范围内；GRNN 的平均准确率最低，但均方根误差明显低于其他模型，因此该模型最为稳定，而其他三类模型的稳定性几乎相同。预测准确率最高的是通过 DC/Smooth 联合预处理的 SVM 模型，正确率达到了 97.2%。对比表 3 和表 4 数据可

以发现,数据降维对四种分类模型的平均准确率和均方根误差的影响并不相同。对于 5 维数据, PNN 和 GRNN 模型的分​​类准确率较 8 维数据有明显提升(7.5%和 5.2%),其中通过 DC/Smooth 联合预处理的 PNN 模型及 SNV 预处理的 GRNN 模型的准确率都达到了 100%。此外, 5 维数据的 GRNN 模型的稳定最差, 均方根误差为 5.6%, 和 8 维数据得到的结果完全相反, 这是由于 GRNN 模型对数据维度的敏感性, 维度降低导致数据失真程度变高, 进而导致稳定性的下降。综合来看, 以上四种鉴别模型均具备快速、准确深度分选四类垃圾的潜力, 其中, SVM 和 PNN 模型分别基于 DC/Smooth 预处理的 8 维和 5 维数据获得最高的分类准确率最大值和平均值, 且稳定性相对较好。

由于 DC/Smooth 联合预处理数据得到的预测准确率更高, 进一步基于 DC/Smooth 预处理比较了四类模型的对四类垃圾分类的准确性, 预测结果如表 5 和表 6 所示。由表 5 可知, 依据  $72 \times 8$  的降维数据, 对烯​​类聚合物垃圾的预测准确率最低, 80 个样本中正确预测数只有 72, 平均准确率只有 90.0%, 这是因为烯​​类聚合物种类多, 具有不同的红外光谱特征波段; 同时, 这也导致了烯​​类聚合物预测的均方根误差最高, 稳定性较差。对于纤维素类与木竹类来说, 四种模型共 80 个样本, 正确预测数均为 75, 平均准确率为 95.0%, 均方根误差均为 4.1%, 表明这两类垃圾红外光谱有很多区分度相近的特征波段数据, 分类模型对其分类能力基本相同。四种分类模型对低值类垃圾的分类判别结果最优, 48 个样本只有一个误判, 平均准确率达到 97.9%, 相对其他类别准确率上升 2.9%~7.9%, 而其均方根误差与纤维素类及木竹类相近, 模型稳定性较强, 这是因为低值类垃圾组分中无机物含量较多, 红外光谱特征波段与其他类别分辨率大。对比表 5 与表 6 可知, 四种分类模型对四类垃圾分选的平均准确率由高到低依次是: 低值类, 纤维素类、木竹类及烯​​类聚合物; 5 维数据相比于 8 数据, 平均分类准确率上升 1.3%~2.5%, 其中, 基于 5 维数据低值类分类平均准确率可以达到 100%; 5 维数据对应分类模型的均方根误差相较 8 维数据下降 1.2%~4.2%, 对四类其他垃圾分类判别稳定性有所上升。

表 5 四类其他垃圾分类准确率对比  
(基于  $72 \times 8$  DC/Smooth 预处理数据)

Table 5 Comparison of classification accuracies for the four kinds of residual wastes (based on  $72 \times 8$  DC/Smooth dataset)

垃圾类别	DC/Smooth				均值 /%	均方根误差 /%
	PNN	GRNN	RDF	SVM		
纤维素类	95.0	90.0	100.0	95.0	95.0	4.1
烯​​类聚合物	95.0	80.0	90.0	95.0	90.0	7.1
木竹类	95.0	90.0	95.0	100.0	95.0	4.1
低值类	100.0	91.6	100.0	100.0	97.9	4.2

表 6 四类其他垃圾分类准确率对比  
(基于  $72 \times 5$  DC/Smooth 预处理数据)

Table 6 Comparison of classification accuracies for the four kinds of residual wastes (based on  $72 \times 5$  DC/Smooth dataset)

垃圾类别	DC/Smooth				均值 /%	均方根误差 /%
	PNN	GRNN	RDF	SVM		
纤维素类	100.0	100.0	95.0	95.0	97.5	2.9
烯​​类聚合物	100.0	90.0	95.0	90.0	93.8	4.8
木竹类	100.0	90.0	95.0	100.0	96.3	4.8
低值类	100.0	100.0	100.0	100.0	100.0	0.0

上述结果表明, 基于红外光谱检测和机器学习建立垃圾深度分选模型具有可行性。需要说明的是, 对于预测中出现的误差, 可能是由于实验样本量少, 导致同一类别中红外光谱数据偏差大, 特别是烯​​类聚合物的种类较多, 所以上述模型对烯​​类聚合物的鉴别准确率最低。此外, 由于源头分类结果不一、垃圾受污染程度参差不齐, 导致红外光谱信息波动, 因此在实际应用中还存在较大的不确定性, 这将在后续的模型优化研究中考虑。

### 3 结 论

基于城市生活垃圾“四分法”源头分类, 针对其他垃圾中不同组分的高值化潜力, 将其他垃圾分为纤维素类、烯​​类聚合物、木竹类及低值类, 利用红外光谱和典型分类器建立了其他垃圾深度分选模型, 对比分析了预处理方式、降维程度和建模算法对分类准确率的影响, 主要结论如下:

SNV, MSC 和 DC/Smooth 联合预处理三类预处理方法中, DC/Smooth 联合预处理的数据用于后续建模得到的分类准确率最高、稳定性最强; 经 PCA 降维后的数据用于后续建模时, 5 维数据比 8 维数据得到的分类性能更强, 整体准确率上升 2.4%~4.4%; 基于 5 维降维数据, DC/Smooth 预处理方法比 SNV 和 MSC 预处理得到的平均准确率更高(96.5%), PNN 模型比其他三类模型的平均准确率更高(98.1%), 其中, DC/Smooth 预处理的 PNN 模型及 SNV 预处理的 GRNN 模型的分​​类准确率都达到了 100%, 用来建立其他垃圾深度分选模型效果最为优异。针对四类其他垃圾, 除烯​​类聚合物的平均判别率只有 93.8%以外, 纤维素类、木竹类及低值类的平均分类准确率均在 95%以上, 最高可以达到 100.0%。

通过红外光谱及机器学习实现了其他垃圾高值化利用的深度分选模型的建立。未来开发快捷准确的其他垃圾深度分选技术, 还需要在考虑源头分类结果均一程度和垃圾受污染程度的基础上, 扩大模型学习样本, 同时优化分类算法, 进一步提升模型的分​​类准确率、稳定性和对实际垃圾样本的适应性。

## References

- [1] CHU Xiao-li, CHEN Pu, LI Jing-yan, et al(褚小立, 陈 瀑, 李敬岩, 等). Journal of Instrumental Analysis(分析测试学报), 2020, 39(10): 1181.
- [2] Bec Krzysztof B, Grabska Justyna, Huck Christian W. Analytica Chimica Acta, 2020, 1133: 150.
- [3] LIU Wei-lai, JIA Si-liang, LIU Jing(刘维来, 贾思良, 刘 静). Environmental Engineering(环境工程), 2019, 37(5): 293.
- [4] LIU Yan-de, XIAO Huai-chun, DENG Qing, et al(刘燕德, 肖怀春, 邓 清, 等). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2019, 32(40): 202.
- [5] PAN Xi, LIU Hui, WANG Hao, et al(潘 曦, 刘 辉, 王 昊, 等). Journal of Instrumental Analysis(分析测试学报), 2020, 39(11): 1385.
- [6] YIN Feng-fu, YAN Lei, HAN Qing-xin, et al(尹凤福, 闫 磊, 韩清新, 等). Environmental Engineering(环境工程), 2017, 35(12): 134.
- [7] ZHAO Dong-e, WU Rui, ZHAO Bao-guo, et al(赵冬娥, 吴 瑞, 赵宝国, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2019, 39(3): 921.
- [8] YUAN Jing-zhu, ZHU Tong(袁惊柱, 朱 彤). Energy of China(中国能源), 2018, 40(6): 16.
- [9] QIAN Le, JIANG Li-qun, YUE Yuan-mao, et al(钱 乐, 蒋丽群, 岳元茂, 等). CIESC Journal(化工学报), 2020, 71(12): 5376.
- [10] LAN Mei-chen, SHEN Bo-xiong, WANG Jian-qiao, et al(兰美晨, 沈伯雄, 王建桥, 等). Journal of Fuel Chemistry and Technology(燃料化学学报), 2019, 47(11): 1313.
- [11] Shafaghat H, Sirous-Rezaei P, Daud W M A. Journal of Industrial and Engineering Chemistry, 2016, 35: 268.
- [12] ZHAO Yue-shu, XUE Xiao-ming, SONG Xiao-jiao, et al(赵阅书, 薛晓明, 宋小娇, 等). Journal of Forestry Engineering(林业工程学报), 2019, 4(5): 40.
- [13] WEI Chen-jie, WANG Ji-fen, JI Jia-hua, et al(卫辰洁, 王继芬, 季佳华, 等). Chemical Research and Application(化学研究与应用), 2021, 33(2): 269.
- [14] Saeys W, Do Trong N N, Van Beers R, et al. Postharvest Biology and Technology, 2019, 158: 110981.
- [15] Porizka Pavel, Klus Jakub, Kepes E, et al. Spectrochimica Acta Part B: Atomic Spectroscopy, 2018, 148: 65.

## Research on Deep Sorting Approach Based on Infrared Spectroscopy for High-Value Utilization of Municipal Solid Waste

HU Bin<sup>1,2</sup>, FU Hao<sup>1</sup>, WANG Wen-bin<sup>1</sup>, ZHANG Bing<sup>1,2</sup>, TANG Fan<sup>3\*</sup>, MA Shan-wei<sup>1,2</sup>, LU Qiang<sup>1,2\*</sup>

1. School of New Energy, North China Electric Power University, Beijing 102206, China

2. National Engineering Laboratory for Biomass Power Generation Equipment, North China Electric Power University, Beijing 102206, China

3. School of Artificial Intelligence, Jilin University, Changchun 130012, China

**Abstract** Due to the advantages of high speed and high accuracy, infrared spectra play a vital role in classification and identification. For municipal solid wastes, the application of infrared spectra mainly focuses on recyclable garbage such as plastics, neglecting the deep separation of the non-recyclable wastes. Based on the current “Quartering Method” of municipal solid wastes, the residual wastes contain various high-value potential ingredients that can be sorted into cellulose, vinyl-polymers, and woods. These ingredients have different constituents and structures, so they have different infrared spectra. Therefore, the useful constituents can be further separated from the residual wastes by combing their infrared spectra and the machine learning classification models. This study collected cellulose, vinyl-polymers, woods, and low-value residual wastes, and 72 groups of infrared spectra data were obtained. The influence of data preprocessing, dimension reduction and algorithms on the sorting models’ accuracy was investigated. Infrared spectra data were preprocessed by standard normal variate (SNV), multiplicative scatter correction (MSC), derivative correction (DC), and smooth. Principal component analysis (PCA) was used to reduce the dimension of the preprocessed data, and  $72 \times 8$  and  $72 \times 5$  datasets were obtained. Sorting models were built using probabilistic neural network (PNN), generalized regression neural network (GRNN), support vector machine (SVM), and random forest (RDF) algorithms. As a result, the classification accuracy of 5-Dimensional data was superior to that of 8-Dimensional data, with the average accuracy increasing 2.4%~4.4%. Based on 5-Dimensional data, DC/Smooth preprocessing

achieved the highest average accuracy of 96.5% among the three preprocessing methods. The average accuracy of the PNN model was 4.2%~6.5% higher than the other three sorting models, up to 98.1%. As for the four types of residual wastes, the sorting accuracy for vinyl polymers was 93.8%, it was over 95% for cellulose and woods, and it could be up to 100% for the low-value wastes. This study examined the possibility and scientific potentiality of the combination of infrared spectroscopy and machine learning to achieve the deep sorting of residual wastes, providing a theoretical basis for the future development of fast and accurate deep separation equipment of municipal solid wastes.

**Keywords** Infrared spectra; Machine learning; Residual wastes; Deep sorting; High-value utilization

(Received Mar. 1, 2021; accepted Jul. 27, 2021)

\* Corresponding authors

(上接 1338 页)

### 报告形式

为充分提高会议学术交流的效率,会议将采用“口头报告”和“墙报展示”两种方式进行学术交流。无论是口头报告还是墙报展示,均属大会同等学术交流。为尊重个人意见和便于组委会的安排,请大家在会议注册时,提交“口头报告”或“墙报”的题目。为了鼓励博士、硕士研究生积极参与学术交流活动,本次会议将继续设立“优秀青年论文奖”和“优秀墙报奖”,表彰那些研究水平高、能突出研究内容要点、条理清晰的“口头报告”和“墙报”,大会将给获奖作者颁发优秀论文证书和奖金。同时会议还将邀请国内外知名专家学者就光谱有关学术领域的前沿热点问题作大会报告和主题报告。

主要报告形式有:

1. 大会邀请报告:主要邀请国内外知名专家学者报告光谱分析的前沿技术在各个领域的最新研究进展。
2. 主题邀请报告:本次会议将选择光谱技术的热点应用领域,开设多个专题论坛,邀请在该领域的知名专家作论坛主题报告。
3. 口头报告:由参会代表申请、组委会审核方式确定报告人选。
4. 青年论坛报告:为博士、硕士研究生开设交流平台,并评选“优秀青年论文奖”。
5. 墙报展示:作为本次会议的主要交流和展示形式。会议统一安排墙报讲解时间,希望作者按时到位讲解。

### 重要时间

开通会议注册系统:2022年5月10日

论文截稿日期:2022年6月30日

第二轮会议通知:2022年5月

第三轮会议通知:2022年10月

会议召开期:2022年11月11日—14日

会议组织机构、注册费及缴纳方式、宾馆住宿介绍及住房预定等信息将在2022年5月初在会议主页上发布,请您经常浏览光谱网上会议主页,了解会议筹备情况和会议具体安排。网址:<http://www.sinospectroscopy.org.cn>

### 产品展示

会议热忱邀请国内外仪器厂商参会及展示仪器设备,大会组委会将在本次会议的网站和会议现场提供展出场所,希望各仪器厂商充分利用这次机会展示自己的最新产品。

(下转 1384 页)