

## 高光谱成像的煤与矸石分类

李廉洁<sup>1,2</sup>, 樊书祥<sup>2</sup>, 王学文<sup>1</sup>, 李瑞<sup>1</sup>, 文小<sup>1</sup>, 王璐瑶<sup>1</sup>, 李博<sup>1\*</sup>

1. 太原理工大学机械与运载工程学院, 山西太原 030024

2. 北京农业智能装备技术研究中心, 北京 100097

**摘要** 煤与矸石分选是煤矿生产的必要工序, 现有的人工分选与机械分选, 存在效率低, 易造成资源浪费以及环境污染等问题。鉴于可见/近红外高光谱成像具有分析速度快、样品无需预处理、无污染等诸多优势, 旨在探讨基于可见/近红外高光谱成像对黑色背景下块状煤与矸石准确分类的可行性, 并基于特征波长筛选算法简化模型, 为构建多光谱煤与矸石分选系统提供理论参考。首先, 搭建高光谱成像系统并采集山西西铭矿的85个煤样本与83个矸石样本在400~1 000 nm(Vis/NIR)与1 000~2 500 nm(NIR)两个范围内的高光谱图像, 基于图像处理方法去除背景信息, 选取 Vis/NIR 范围内 100×100 像素和 NIR 范围内 50×50 像素区域内的平均光谱作为该样本在对应波段范围的一条光谱, 重复 10 次, 最终在两个波段各获得煤与矸石光谱 850 条和 830 条。其次, 对光谱先后进行 Savitzky-Golay 卷积平滑和标准正态变量变换, 以减少噪音和误差对光谱的影响。基于全波段光谱建立支持向量机(SVM),  $k$  近邻法(KNN), 偏最小二乘判别分析(PLS-DA)三种模型, 每个模型针对预测集的分类准确度均大于 0.95, 结果表明, 基于煤和矸石的光谱信息可将二者区分。随后根据竞争性自适应重加权算法(CARS)和连续投影算法(SPA)选择的特征波长建立简化模型, 综合考虑精度与成本等因素, 在 Vis/NIR 范围内基于 SPA 筛选的 3 个特征波长所建立的 SVM 模型效果最好, 不仅能有效减少波长数量, 还能提高模型分类效果, 对应的灵敏度, 特异度, 准确度分别为 1, 0.965 2, 0.983 3。基于判别模型与样本的平均光谱还可实现煤和矸石的分类可视化。研究结果对开发基于特征波长的低成本煤和矸石多光谱分选系统, 实现煤矸快速、准确的无损检测具有借鉴意义。

**关键词** 高光谱成像; 煤; 矸石; 黑色背景; 无损检测

**中图分类号:** TD94 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2022)04-1250-07

### 引言

煤炭作为我国的主体能源之一, 是国家快速发展的重要支撑, 近年来, 随着环境问题日益突出, 国家大力倡导并推进煤炭的清洁利用。煤矸石是煤矿开采过程中产生的灰色或黑色岩石, 将煤与矸石分离是煤矿生产的必要工序。由于人工分选法与湿选法存在效率低、劳动强度大, 水资源消耗高, 污染环境等问题, 干选法成为煤矸分选的主要研究方向之一。干选法主要有破碎法(冲击破碎, 挤压破碎等), 存在设备寿命较短且普适性差的问题; 射线法<sup>[1]</sup>, 需单独隔离射线源且有辐射风险; 图像识别法<sup>[2]</sup>, 目前仍存在识别结果易受光照、灰尘等环境因素影响的问题。

可见/近红外高光谱成像技术具有分析速度快、无损、

样品无需预处理、无污染, 可同时测定多项指标等诸多优势, 已被广泛应用于各个领域。基于煤与矸石的高光谱数据和多种光谱分析方法, 探究实现煤与矸石的高光谱无损检测, 对实现“绿色开采”具有重要的研究意义。虽然已有很多基于近红外光谱对煤的品质<sup>[3-5]</sup>(固定碳、挥发分、灰分、硫分等)、煤的种类<sup>[6]</sup>以及煤产地<sup>[7]</sup>等问题的探讨, 但基于光谱数据进行煤与矸石分类的研究还较少。宋亮等<sup>[8]</sup>在室外采集样本的光谱数据, 根据样本的可见/近红外光谱反射率和热红外光谱的光谱吸收比率对煤与矸石进行区分。杨恩等<sup>[9]</sup>基于 GRB-KPCA 对样本光谱进行特征提取, 并采用 SVM 区分烟煤与碳质页岩。Mao 等<sup>[10]</sup>同样在室外采集样本光谱数据, 基于 IAM-ELM 算法实现煤与矸石的分类。Hu 等<sup>[11-12]</sup>基于多光谱成像技术挑选单一通道的光谱图像进行图像处理, 从而实现煤与矸石分类。Zou 等<sup>[13]</sup>借助微型光谱仪, 在

收稿日期: 2021-03-14, 修订日期: 2021-05-13

基金项目: 国家自然科学基金项目(51804207, 51875386)资助

作者简介: 李廉洁, 1998 年生, 太原理工大学机械与运载工程学院硕士研究生 e-mail: lilianjie0063@link.tyut.edu.cn

\* 通讯作者 e-mail: libo@tyut.edu.cn

样本正上方 20 mm 处采集光谱数据, 基于 lasso 回归的宽度学习对煤与矸石进行区分。

基于光谱信息对煤的分析研究中, 需对样本进行粉碎、研磨、筛选等预处理, 且未考虑样本的背景颜色, 与实际应用场景不符。通过光谱仪和光纤进行光谱采集, 只能获取样本部分区域的光谱信息, 导致数据所包含的样本信息不够充分。

本研究的具体目标是: (1) 基于高光谱成像系统采集块状煤与矸石样本在黑色背景下的高光谱数据, 提取样本光谱信息并进行预处理后, 探究不同分类模型对煤与矸石分类的可行性; (2) 基于特征波长筛选算法挑选的特征变量建立简化模型, 对比不同简化模型对煤与矸石分类的效果; (3) 选

择适用于开发煤与矸石分类的多光谱成像系统的分类模型, 并对煤与矸石进行分类可视化。

## 1 实验部分

### 1.1 样本

样本为山西太原西铭矿的焦煤以及黑色矸石, 样本总数 168 块(煤 85 块, 矸石 83 块), 样本高度在 15~60 mm 之间, 随机划分校正集和预测集, 其中校正集包含煤和矸石各 60 块, 预测集包含煤 25 块, 矸石 23 块。实验前, 将所有样品置于室内阴暗环境下, 放置至室温。图 1 为部分样本图片, 从图中可看出, 煤与矸石的颜色相似, 肉眼难以将二者区分。

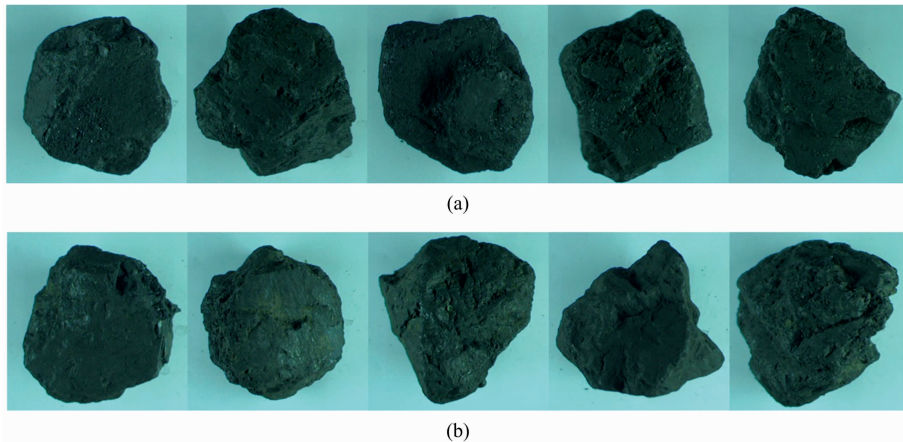


图 1 部分煤样本(a)与矸石样本(b)

Fig. 1 Coal samples (a) and gangue samples (b)

### 1.2 高光谱数据采集及光谱提取

图 2(a)为搭建的高光谱采集系统, 整个系统主要包含两部分:

可见/近红外(visible/near infrared, Vis/NIR)高光谱成像系统, 包括成像范围在 326~1 000 nm 的 Vis/NIR 高光谱成像仪(ImSpector V10E, Spectral Imaging Ltd., Oulu, Finland)、23 mm 的 C 口变焦镜头(OLE23 f-2.4/23 mm, Spectral Imaging Ltd., Oulu, Finland)、像素为 1 004×1 000 的 EMCCD 相机(Luca-R, Andor Technology, Belfast, UK)。

近红外(near infrared, NIR)高光谱成像系统, 包括成像范围在 930~2 548 nm 的 NIR 高光谱成像仪(ImSpector N25E, Spectral Imaging Ltd., Oulu, Finland)、30 mm 的 SPCECIM 口变焦镜头(OLES30 f-2.0/30 mm, Spectral Imaging Ltd., Oulu, Finland)、像素为 320×256 的 CCD 相机(Xeva-2.5-320, Xenics Ltd., Belgium)。

其余附件包括: 一对 150W 的卤素灯(Antefore International Co., Ltd., Taiwan, China), 通过步进电机控制的样本移动平台(EZHR17EN, AllMotion, Inc., USA)。整个采集平台置于暗箱中以降低外界杂散光的影响, 通过计算机(Dell OPTIPLEX 990, Intel (R) Core (TM) i5-2400 CPU at 3.10 GHz)以及配套的专业软件(Isuzu Optics Corp., Taiwan, China)进行动作控制及数据采集。经过调试, 先将

镜头与样本平台表面之间的垂直距离调整至 400 mm, 两光源相距 550 mm 置于镜头两侧, 角度调整为 45°。采集数据时, 将成像光谱仪和相机打开, 30 min 后待设备稳定, 以纯黑色纸板为背景, 将样本置于移动平台上, 分别将曝光时间和平台移动速度设置为 16 ms, 0.73 mm·s<sup>-1</sup>与 2 ms, 42 mm·s<sup>-1</sup>以获取 Vis/NIR 数据以及 NIR 数据。确保数据在同一时间段内获取以提高数据的可靠性。

为了减少照明不均匀以及相机暗电流的影响, 需要对所采集的原始高光谱数据( $I_{\text{origin}}$ )进行黑白校正, 在相同的环境下, 利用反射率接近 100% 的聚四氟乙烯白板采集白参考( $I_{\text{white}}$ ), 再将光源关闭后盖上镜头盖, 采集暗参考( $I_{\text{dark}}$ )。参照式(1)获得校正后的高光谱图像( $I$ )

$$I = \frac{I_{\text{origin}} - I_{\text{dark}}}{I_{\text{white}} - I_{\text{dark}}} \quad (1)$$

为提取样本光谱信息, 先去除数据的背景信息, 针对 Vis/NIR 高光谱图像与 NIR 高光谱图像, 通过处理目标与背景对比度明显的 918 nm, 2 154 nm 的单波段灰度图像获得掩膜图像以去除背景信息。然后进行光谱数据提取, 为了提高模型的适应能力, 利用样本不同区域光谱信息的差异性, 针对获取的 Vis/NIR 和 NIR 高光谱图像, 分别随机选取大小为 100×100 像素、50×50 像素的区域, 提取区域内的平均光谱作为该区域的光谱信息。重复 10 次, 在两个波段各获得煤与矸石光谱 850 条和 830 条。对 Vis/NIR 高光谱图像的

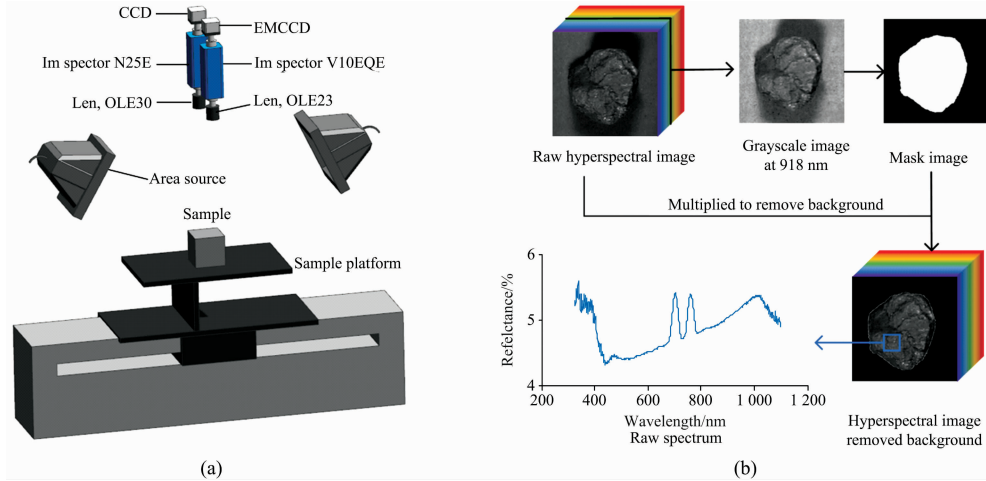


图 2 高光谱采集系统 (a) 以及光谱提取过程 (b)  
Fig. 2 Hyperspectral imaging system (a) and process of spectrum extraction (b)

分析流程如图 2(b)。最终,校正集共 1200 条光谱(煤和矸石各 600 条),预测集共 480 条光谱(煤 250 条,矸石 230 条)。

### 1.3 光谱预处理及特征波长变量的选择

光谱数据中不仅包含样本的化学信息,还包含仪器噪声,杂散光等无关信息,且全波段的光谱数据具有多重共线性和信息冗余性,通过对原始光谱进行预处理和波长选择,不仅可以减少噪声,还能剔除不相关或非线性的变量,减少数据量,简化模型,提高运算速度,一定程度上提高模型的性能,可为构建多光谱分类系统提供理论参考。

通过 Savitzky-Golay(SG)卷积平滑(窗口为 7,拟合一次多项式),消除光谱的随机噪声;由于样本尺寸有一定差异,因此再对数据进行标准正态变量变化(standard normal variate transformation, SNV)以消除光程变化等因素对反射光谱的影响。

竞争性自适应重加权算法(competitive adaptive reweighted sampling, CARS)通过随机采样的方式,在校正中随机选择一部分样本进行偏最小二乘(partial least square, PLS)建模,然后根据指数衰减函数保留回归系数相对较大的波长, $N$ 次重复获得对应的新变量子集,基于此建立 PLS 模型,选择交互验证均方差最小的变量子集作为最佳特征波长变量集。在执行 CARS 特征波长筛选时,设置交叉验证为 10 折,采样次数为 100 次。

连续投影算法(successive projections algorithm, SPA),基于向量空间基变换原理,通过将波长投影到其他波长上,选择投影向量较大的波长子集作为最佳特征波长变量集。在执行 SPA 特征波长筛选时,设置最大特征波长个数为 5。

### 1.4 建模方法及模型评价指标

为获取可靠分类结果,选用支持向量机(support vector machines, SVM), $k$  近邻法( $k$ -nearest neighbor, KNN),偏最小二乘判别分析(partial least squares discriminant analysis, PLS-DA)三种经典的分类算法进行判别模型的建立与对比。基于校正集训练判别模型,基于预测集对模型进行性能评估。将煤作为正类,矸石作为负类,将敏感度(sensitivi-

ty),特异度(specificity),准确率(accuracy)作为模型的评价指标,计算公式如式(2)一式(4)

$$\text{sensitivity} = \frac{TP}{TP + FN} \quad (2)$$

$$\text{specificity} = \frac{TN}{TN + FP} \quad (3)$$

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

式中:TP 为实际为煤,预测也是煤的数量;TN 为实际为矸石,预测为矸石的数量;FP 为实际为矸石,预测为煤的数量;FN 为实际为煤,预测为矸石的数量。

### 1.5 煤和矸石的分类可视化

高光谱数据提供了样本丰富的空间光谱信息,将样本表面所有像素点光谱的平均光谱代入已建立的模型进行判别,通过不同的颜色表示煤和矸石,不但可以检验模型的预测能力,还能更直观的反映样本种类。

所有数据处理均借助 MATLAB R2019a(The Math-Works Inc., MA, USA)进行。

## 2 结果与讨论

### 2.1 反射光谱的曲线特征

由于原始光谱前后端有相当大的随机噪声,对 Vis/NIR 和 NIR 数据,分别取 474~940 nm(600 个变量)和 1 235~2 477 nm(196 个变量)范围内的光谱进行分析,并对光谱进行 SG-SNV 预处理。图 3 为煤与矸石样本在指定范围内的原始光谱,显然煤与矸石的光谱曲线在相同的波段范围内显示出相似的趋势,但也存在一定的差异。

在 474~940 nm 范围,煤和矸石的光谱曲线均在 700 和 760 nm 附近有明显波峰,矸石光谱曲线的斜率大于煤光谱曲线的斜率;在 1 235~2 477 nm 范围,煤和矸石的光谱曲线在 2 315 nm 附近有明显波峰,在 2 347 nm 附近有明显波谷,矸石的反射率整体大于煤的反射率。由于在近红外波段,芳香分子中电子跃迁趋向长波长方向,煤分子的芳构化

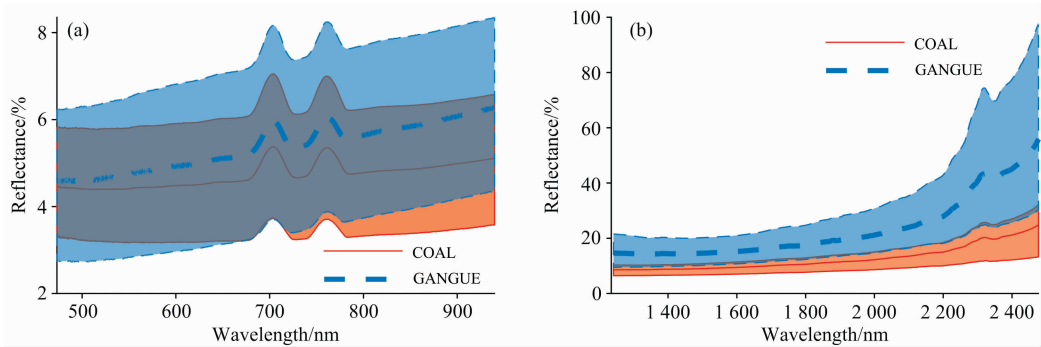


图 3 煤与矸石在 474~940 nm 范围 (a) 和 1 235~2 477 nm 范围 (b) 的原始光谱

Fig. 3 The original spectral curves in the range of 474~940 nm (a) and 1 235~2 477 nm (b)

程度高使其在近红外波段长波方向的光谱吸收系数大，整体反射率较低且反射光谱的斜率小于矸石<sup>[14]</sup>。

2.2 基于全波段光谱的煤与矸石的分类模型

在 Vis/NIR 和 NIR 范围内，基于校正集的全波段光谱训练 SVM, KNN, PLS-DA 模型，预测集的分类结果如表 1。三种全波段分类器的预测结果相同，对应的 sensitivity, spe-

cificity, accuracy 分别为 1, 0.956 5 和 0.979 2；在 NIR 范围，PLS-DA 模型的预测结果最好，sensitivity, specificity, accuracy 分别为 1, 0.987 0 和 0.993 8。由结果可看出，在两个波段范围内，基于全波段光谱的三种分类模型结果均较好，证明基于煤和矸石的光谱信息可将二者区分。但全波段光谱数据的处理速度较慢，故对其进行特征挑选。

表 1 基于全波段光谱的不同分类模型对比

Table 1 Comparison of different classification models based on the full-band spectra

Spectral range	Number of variables	Model	sensitivity	specificity	accuracy
Vis/NIR	600	SVM	1.000 0	0.956 5	0.979 2
		KNN	1.000 0	0.956 5	0.979 2
		PLS-DA	1.000 0	0.956 5	0.979 2
NIR	196	SVM	1.000 0	0.960 9	0.981 3
		KNN	0.956 0	0.956 5	0.956 3
		PLS-DA	1.000 0	0.987 0	0.993 8

2.3 基于特征波长的煤与矸石分类模型

为了消除冗余变量，进一步优化预测模型的性能，提高检测速度，采用 CARS 和 SPA 两种波长挑选算法筛选光谱变量，基于有效变量建立分类模型。

对全波段光谱进行 CARS 特征波长筛选的过程如图 4，在两个波段范围内选择的波长数量随采样次数的增加而减少，波长数量的减少速度先快后慢。在 Vis/NIR 波段，第 85 次采样获得的交互验证均方差最小，选择的变量子集包括 716.94, 717.73, 718.51, 768.82 和 769.61 nm 共 5 个特征波长，占全波段的 0.5%；在 NIR 波段第 95 次采样获得的交互验证均方差最小，选择的变量子集包括 1 247.81, 1 398.11 和 2 186.18 nm 共 3 个特征波长，占全波段的 1.17%。

对全波段光谱进行 SPA 特征波长筛选后的结果如图 5 所示，采用 SPA 筛选特征波长时，通过对比不同特征波长下的预测集样本均方根误差来确定最优特征波长数量。在 Vis/NIR 波段，当特征波长数大于 3 时，预测样本集均方根误差变化不显著，选择了 585.64, 722.43 和 766.46 nm 共 3 个特征波长，占全波段的 0.3%；在 NIR 波段，当特征波长数等于 3 时，预测样本集均方根误差最小，选择了 1 923.48,

2 237.68 和 2 276.36 nm 共 3 个特征波长，占全波段的 1.17%。

基于校正集在 Vis/NIR 和 NIR 范围内的特征波长训练 SVM, KNN, PLS-DA 模型，针对预测集的分类结果如表 2。可以看出，在 Vis/NIR 范围内基于 SPA 筛选的 3 个特征波长所建立的 SVM 模型与在 NIR 范围内基于 CARS 筛选的 3 个特征波长所建立的 KNN 模型效果最好，二者不仅有效的减少了波长数量，还提高了模型的分类效果，对应的 sensitivity, specificity, accuracy 分别为 1.000 0, 0.965 2, 0.983 3 和 0.988 0, 0.991 3, 0.989 6。

近红外光谱仪器的检测器在 Vis/NIR 范围多采用硅(Si)基检测器，成本较低；在 NIR 范围多采用硫化铅(PbS)或碲化镓(InGaAs)检测器，成本较高。因此，综合预测精度以及后续多光谱系统开发成本的考虑，可选择 Vis/NIR 范围内基于 SPA 算法筛选的特征波长所建立的 SVM 模型作为煤与矸石的分类模型。

2.4 煤与矸石的分类可视化

相较于传统光谱，高光谱数据还包含空间光谱信息，这使得煤与矸石的分类可视化具有可行性。选取在 Vis/NIR 范围内的 SPA-SVM 简化模型，基于样本表面所有像素点的平



均光谱, 对样本进行分类和可视化, 用不同的颜色映射煤与矸石, 不仅能够直观地表示样本的类别, 在实际应用中还可

保存可视化图像, 便于系统故障时排查问题, 图 6 为部分煤和矸石样本的分类可视化。

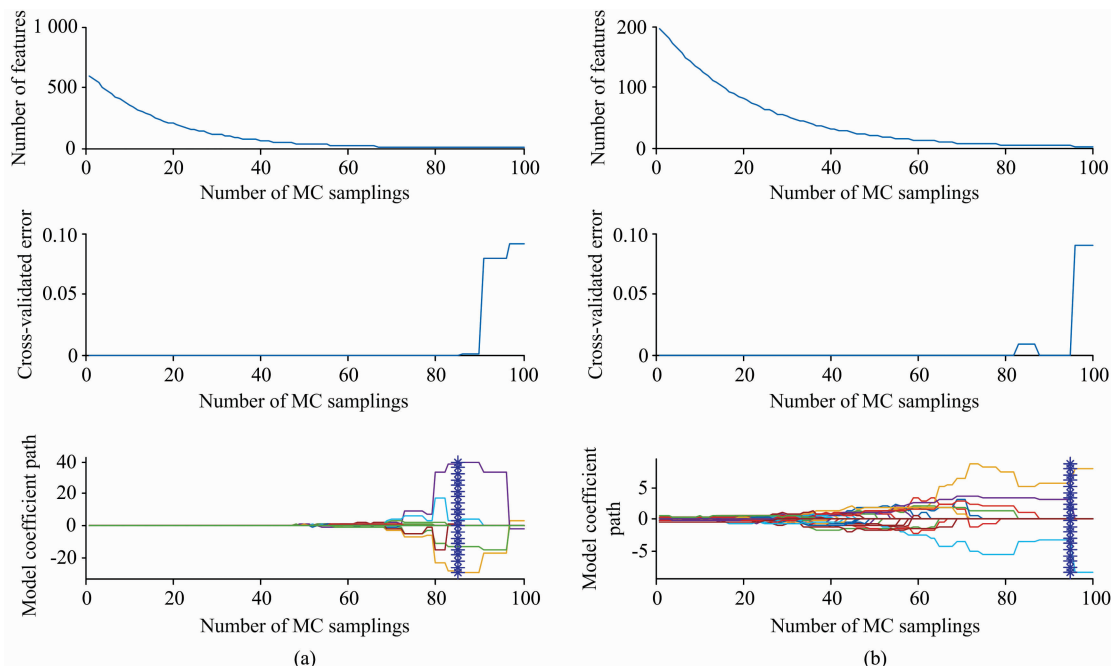


图 4 Vis/NIR (a) 和 NIR (b) 范围内 CARS 波长筛选过程

Fig. 4 The process of variable selection by CARS in the spectral ranges of (a) Vis/NIR and (b) NIR

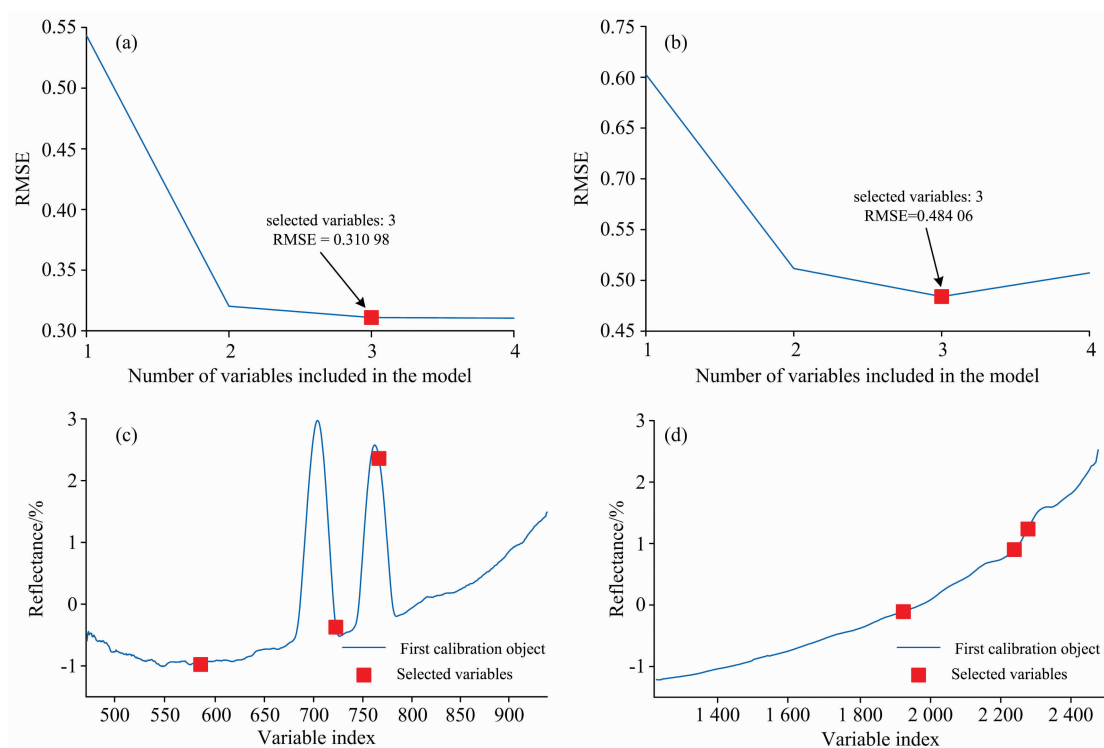


图 5 Vis/NIR (a), (c) 和 NIR (b), (d) 范围内 SPA 波长筛选的结果

Fig. 5 The results of variable selection by SPA over the spectral ranges of (a), (c) Vis/NIR and (b), (d) NIR

表2 基于特征波长的不同分类模型预测结果

Table 2 The prediction results of different classification models based on characteristic wavelengths

Spectral range	Variable selection methods	Number of variables	Model	sensitivity	specificity	accuracy
Vis/NIR	CARS	5	SVM	1.000 0	0.930 4	0.966 7
			KNN	0.976 0	0.930 4	0.954 2
			PLS-DA	0.956 0	0.830 4	0.895 8
	SPA	3	SVM	1.000 0	0.965 2	0.983 3
			KNN	1.000 0	0.956 5	0.979 2
			PLS-DA	0.776 0	0.869 6	0.820 8
Vis/NIR	CARS	3	SVM	0.888 0	0.995 7	0.939 6
			KNN	0.988 0	0.991 3	0.989 6
			PLS-DA	1.000 0	0.965 2	0.983 3
	SPA	3	SVM	0.960 0	0.947 8	0.954 2
			KNN	0.960 0	0.943 5	0.952 1
			PLS-DA	0.932 0	0.917 4	0.925 0

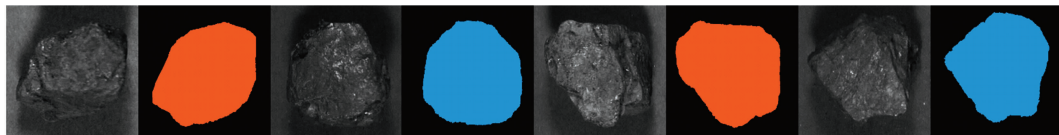


图6 部分样本的灰度图及对应的分类可视化

红色: 煤; 蓝色: 矸石

Fig. 6 Grayscale images of some samples and corresponding classification visualization

red: Coal; blue: Gangue

### 3 结 论

针对黑色背景下块状煤与矸石的准确分类的问题, 提出一种基于高光谱成像技术的煤与矸石的分类及类别可视化方法, 得到以下结论:

(1) 煤和矸石在 Vis/NIR 以及 NIR 范围内的光谱差异明显, 基于光谱信息可将两者区分。

(2) 与基于全波段光谱所建立的判别模型以及其他简化模型相比, Vis/NIR 范围内基于 SPA 筛选的 3 个特征波长所

建立的 SVM 模型与在 NIR 范围内基于 CARS 筛选的 3 个特征波长所建立的 KNN 模型效果最好, 二者不仅有效的减少了波长数量, 还提高了模型的分类效果。对应的 sensitivity, specificity, accuracy 分别为 1.000 0, 0.965 2, 0.983 3 和 0.988 0, 0.991 3, 0.989 6。

(3) 基于精度以及成本等因素考虑, 可选择 Vis/NIR 范围内的基于 SPA 算法筛选的特征波长建立的 SVM 模型用于开发煤与矸石分类的多光谱成像系统。同时可通过样本平均光谱以及分类模型实现样本的可视化, 能够更直观的体现分类结果。

### References

- [ 1 ] Zhang N, Liu C. Scientific Reports, 2018, 8(1): 190.
- [ 2 ] Pu Y, Apel D B, Szmigiel A, et al. Energies, 2019, 12(9): 1735.
- [ 3 ] Wang S H, Zhao Y, Hu R, et al. Chinese Journal of Analytical Chemistry, 2019, 47(4): E19034.
- [ 4 ] Le B T, Xiao D, Mao Y, et al. Infrared Physics & Technology, 2018, 93: 34.
- [ 5 ] Begum N, Maiti A, Chakravarty D, et al. Fuel, 2020, 280: 118676.
- [ 6 ] SONG Liang, LIU Shan-jun, MAO Ya-chun, et al(宋 亮, 刘善军, 毛亚纯, 等). Journal of Northeastern University • Natural Science (东北大学学报 • 自然科学版), 2017, 38(10): 1473.
- [ 7 ] Lei M, Yu X, Li M, et al. Infrared Physics & Technology, 2018, 92: 177.
- [ 8 ] SONG Liang, LIU Shan-jun, YU Mo-li, et al(宋 亮, 刘善军, 虞茉莉, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2017, 37(2): 416.
- [ 9 ] Yang E, Ge S, Wang S, et al. Journal of Spectroscopy, 2018, 2018: ID2754908.
- [ 10 ] Mao Y, Le B T, Xiao D, et al. Optics and Laser Technology, 2019, 114: 10.
- [ 11 ] Hu F, Zhou M, Yan P, et al. IEEE Access, 2019, 10.1109/ACCESS.2019.2955725.

- [12] Lai W, Zhou M, Hu F, et al. *IEEE Access*, 2020, 10, 1109/ACCESS, 2020, 2983740.
- [13] Zou L, Yu X, Li M, et al. *IEEE Transactions on Instrumentation and Measurement*, 2020, 69(10): 8043.
- [14] Cloutis E A. *Fuel*, 2003, 82(18): 2239; 1.

## Classification Method of Coal and Gangue Based on Hyperspectral Imaging Technology

LI Lian-jie<sup>1, 2</sup>, FAN Shu-xiang<sup>2</sup>, WANG Xue-wen<sup>1</sup>, LI Rui<sup>1</sup>, WEN Xiao<sup>1</sup>, WANG Lu-yao<sup>1</sup>, LI Bo<sup>1\*</sup>

1. College of Mechanical and Vehicle Engineering, Taiyuan University of Technology, Taiyuan 030024, China

2. Beijing Research Center of Intelligent Equipment for Agriculture, Beijing 100097, China

**Abstract** The separation of coal and gangue is a crucial step in coal mining, but the existing methods such as manual selection and mechanical separation are ineffective and environmentally hazardous. This study aimed to explore the feasibility of the accurate classification of coal and gangue with black background based on the visible and near-infrared hyperspectral imaging technology, simplify classification models using feature selection methods, and provide a reference for constructing a multispectral system for coal and gangue separation. Hyperspectral imaging technology is a fast and non-destructive detection method without sample pretreatment and environmental contamination. Firstly, a hyperspectral imaging system was developed to collect hyperspectral data of 85 coal samples and 83 gangue samples in the range of 400~1 000 nm (Vis/NIR) and 1 000~2 500 nm (NIR) from the XiMing mine. After removing background information of hyperspectral images, the average spectra in the randomly selected regions of 100 pixel×100 pixels in 400~1 000 nm and 50 pixel×50 pixels in 1 000~2 500 nm were extracted. After repeating 10 times, 850 coal spectra and 830 gangue spectra were obtained in each of the two bands. Savitzky-Golay smoothing and standard normal variate transformation were performed successively to reduce the impact of errors and noise on the spectra. Three models, including support vector machine (SVM), k-nearest neighbor (KNN), partial least squares discrimination analysis (PLS-DA), were established based on full-band spectra. The classification accuracy rate of each model for the prediction set was greater than 0.95, which revealed that coal and gangue could be distinguished by spectral information. Subsequently, competitive adaptive reweighted sampling (CARS) and successive projections algorithm (SPA) were employed to select characteristic wavelengths to simplify models. Considering factors such as accuracy and cost, the SVM model based on the 3 characteristic wavelengths screened by SPA in the Vis/NIR range had the best performance, that not only effectively reduced the number of wavelengths, but also improved the classification capacity and the corresponding sensitivity, specificity, accuracy was: 1.000 0, 0.965 2, 0.983 3, respectively. Based on the discriminant model and the average spectra of the samples, the classification and visualization of coal and gangue can also be realized. The research results have great potential for developing a low-cost and multi-spectral separation system for coal and gangue based on the characteristic wavelengths to achieve fast and accurate non-destructive separation.

**Keywords** Hyperspectral image; Coal; Gangue; Black background; Nondestructive detection

(Received Mar. 14, 2021; accepted May 13, 2021)

\* Corresponding author