第38卷，第2期
2018年2月

光 谱 学 与 光 谱 分 析
Spectroscopy and Spectral Analysis

Vol. 38，No. 2，pp352-357
February，2018

全谱段光谱分析的块状商品煤种类鉴别

任 淯[1，2]，孙雪剑[2*]，戴晓爱[1]，岑 奕[2]，田亚铭[1]，王 楠[2]，张立福[2]

1. 成都理工大学地球科学学院，四川 成都 610059
2. 中国科学院遥感与数字地球研究所，遥感科学国家重点实验室，北京 100101

摘 要 常规的煤炭鉴别方法需进行繁琐的制样过程，且需结合多种化学参数指标进行综合判定，以得到较为准确的分析结果。提出一种基于500～2 350 nm的可见-近红外全谱段光谱分析技术与多层感知器(multilayer perceptron，MLP)分类方法相结合的块状商品煤鉴别方法。该方法具有非接触、无前期制样、无化学分析的优势，可快速高效的获取煤炭的分类信息。采用地物光谱仪采集煤炭原始光谱数据，对噪声过大、影响后续处理的谱段进行删除，剩余部分采用小波阈值去噪法进行噪声去除。将去噪后的数据分成三个数据集：可见-近红外光谱(500～900 nm)数据集、短波红外光谱(1 000～2 350 nm)数据集、全谱段光谱(500～2 350 nm)数据集。对以上三个数据集进行主成分分析，将提取出的25个主成分输入多层感知器分类模型。多层感知器模型由输入层、隐藏层(两层)、softmax分类器构成。对三个数据集进行分类精度的对比，并采用随机森林(random forest，RF)与支持向量机(support vector machine，SVM)两种分类算法进行进一步的验证分析。结果表明：对块状商品煤分类，全谱段光谱分析技术由于数据信息量丰富，能够得到更优的分类效果，在训练样本数为132时，采用MLP分类器的分类精度最高，为98.03%；随机森林与SVM的分类结果验证了全谱段数据集的优越性与普适性。该研究为煤炭的在线分析、便携式煤炭检测仪器的研发提供了可靠的技术支持。

关键词 全谱段；块状商品煤种类鉴别；多层感知器；主成分分析

中图分类号：O657.3 **文献标识码**：A **DOI**：10.3964/j.issn.1000-0593(2018)02-0352-06

引 言

煤炭是一种蕴藏量丰富，分布广泛的化石燃料。为合理有效的利用煤炭资源，煤炭种类鉴别成为后续工业应用的重要一环。在实际加工中，避免繁复的化学检验、免除复杂的制样过程，如何经济、实时、快速的反馈煤炭的类别信息成为亟待解决的问题。

目前较广泛使用的煤炭快速分析方法[1-3]，均只能获取煤样的单项指标。如微波加热法[4]，γ射线法[5]等，该类方法虽然快速直接，但若要进行煤炭种类的鉴别，仍需获取多项指标进行综合判别，鉴别过程复杂，检验费用昂贵。近年来，国内外专家采用近红外光谱分析技术建立分析模型可以安全、无损、快速地得到高精度煤质分析结果[6]。雷萌等[7]应用近红外光谱分析技术和去噪算法对煤炭产地进行鉴别，得到了较高的分类精度。王雅圣等[8]应用近红外光谱技术与置信学习机相结合进行煤种的快速分类。然而，近红外光谱分析过程中，仍需对样本进行破碎、混合、筛分等一系列繁琐的制样过程才能够对煤炭进行鉴别分析。且目前对煤炭光谱的研究尚未涉及到可见光波段，块状商品煤直接进行全谱段的煤炭种类鉴别尚属研究领域的空白。

本方法弥补了块状煤炭在全谱段研究领域的空白，该方法的实用性、可操作性满足于现实需求，真正做到样本的无损快速检测。免除冗余的样本制备过程，便于样本的测试、转移、保存。在现有较为成熟的近红外光谱技术上，添加可见-近红外谱段(500～900 nm)，增加光谱信息量。采用多层感知器分类模型，显著提升了鉴别精度，并采用随机森林与支持向量机的分类方法进行对比分析，验证了全谱段光谱分析技术在块状煤炭的分类中具有的优越性和普适性。

收稿日期：2017-05-14，**修订日期**：2017-10-05

基金项目：国家自然科学基金项目(41501391)资助

作者简介：任 淯，1992年生，成都理工大学地球科学学院硕士研究生 e-mail：renyu_rs@163.com

孙雪剑，1987年生，成都理工大学助理研究员 e-mail：sunxj@radi.ac.cn

任 淯，孙雪剑：并列第一作者 *通讯联系人 e-mail：sunxj@radi.ac.cn

1　实验部分

1.1　样本与光谱采集

实验样本为福建龙岩、云南镇雄、贵州六盘水三地出售的块状无烟煤，三类煤炭的指标参数如下：福建龙岩无烟煤样品的外水为 5.8%，内水为 0.43%，灰分为 13.23%，挥发分为 4.96%，发热量为 6200；云南镇雄无烟煤样品的外水为 6.75%，内水为 0.92%，灰分为 25.32%，挥发分为 8.21%，发热量为 5490；贵州六盘水无烟煤样品的外水 2.44%，内水 1%，灰分 12.06%，挥发分 5.99%，发热量为 7244。

采用美国 Spectral Evolution 公司生产的 PSR-3500 便携式地物波谱仪，光谱覆盖范围为 350～2 500 nm，光谱分辨率最优为 3.5 nm，光源为卤素灯。不对样本进行研磨处理，直接进行光谱获取。

光谱在暗室内采集。将煤炭样本分别放于同一块黑布的同一位置分次测量，固定光源与样本的距离。由于样本为块状样本，为避免镜面反射的发生，PSR 探头与样本表面垂直且选取样本平整表面进行测量。每块样本采集 10 次光谱曲线取均值作为一条实验样本光谱数据。

共采集光谱样本 219 组，并定义分类标签。福建龙岩 69 组，分类标签为 1，贵州六盘水 72 组，分类标签为 2，云南镇雄 78 组，分类标签为 3。

1.2　数据预处理

获取的光谱信息中不仅包含样本自身信息还包含有噪声和其他无关的背景信息，因此消除背景信息和噪声尤为关键。

光谱的首尾部分 350～500 nm，2 350～2 500 nm 与传感器谱段衔接处 900～1 000 nm，1 800～1 950 nm 处的噪声较大，因此将以上部分舍弃。剩余部分采取小波阈值去噪法进行去噪处理，将光谱曲线的分解层数设定为 4 层，阈值处理方法选定软阈值，阈值估计方法设定为启发式阈值选择法，对小波系数进行阈值处理，并进行信号重构，得到预处理后的平均光谱曲线，如图 1 所示。

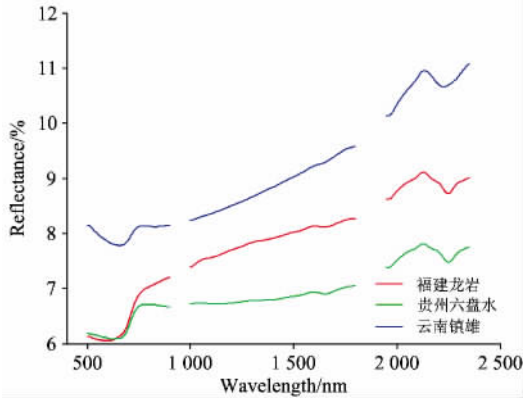

图 1　预处理后的平均光谱曲线

Fig. 1　Average spectral curves after pretreatment

预处理后的数据按照波段范围分成三个新的数据集——可见-近红外光谱数据集(500～900 nm)、短波红外光谱数据集(1 000～2 350 nm)、全谱段光谱数据集(500～2 350 nm)；对三个数据集进行后续的主成分分析及分类对比研究。

1.3　主成分分析

主成分分析(principal component analysis，PCA)是一种能将原始的数据信息尽可能的表达为少数新变量的数据降维方法[9]。对可见-近红外、短波红外、全谱段三个光谱数据集分别进行主成分分析，三者的前三个主成分累积贡献率分别为 98.45%，96.86%，98.01%，基本包含大部分的光谱信息。图 2(a)—(c)分别是可见-近红外光谱数据集、短波红外光谱数据集、全谱段光谱数据集的主成分得分图。

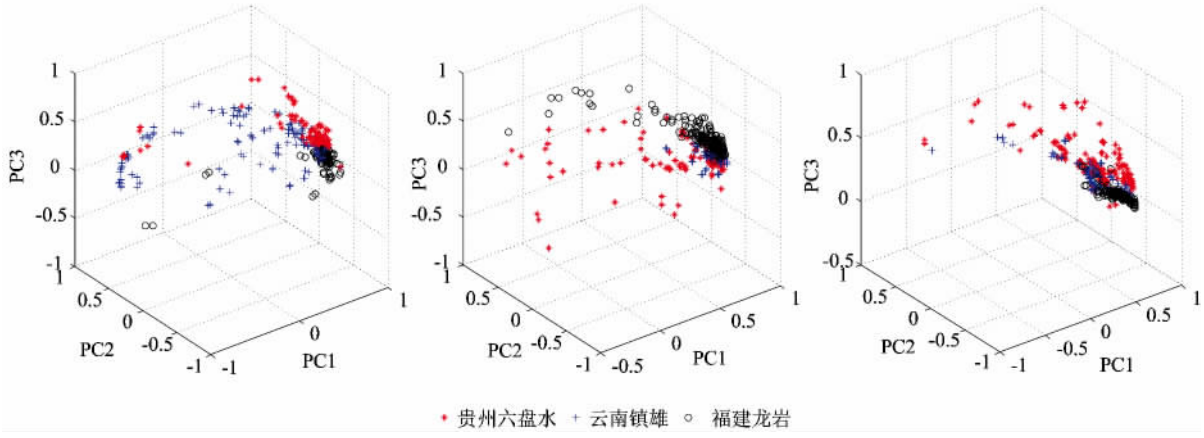

图 2　三个数据集的主成分分析图

(a)：可见-近红外；(b)：短波红外；(c)：全谱段

Fig. 2　Principal component map of three data sets

(a)：Vis-NIRS；(b)：SWIR；(c)：Full spectrum

可见-近红外光谱数据集主成分得分图中，福建龙岩和贵州六盘水的聚类效果较好，云南镇雄分布较为分散；如图 2(b)所示短波红外光谱数据得分图中，云南镇雄的聚类效果较好，但三种煤炭混杂在一起，无法从视觉上明确分离；全

谱段主成分得分图中三种样本的聚类效果均较好，且可分性强，能够从视觉上较清晰的区分出类别。

三个数据集的前 25 个主成分的累积贡献率均达到 99.99%以上，基本包含原始数据的全部信息，故作为分类器的输入特征。

1.4 分类模型构建与模型参数设置

多层感知器神经网络[10]是一种人工神经网络结构，又被称为多层前馈网络。多层感知器网络主要由输入层，一层或多层隐藏层，和输出层所组成。

针对煤炭数据构建多层感知器分类模型，该多层感知器结构由输入层、隐藏层(两层)、softmax 分类器构成。设置网络结构如图 3 所示。

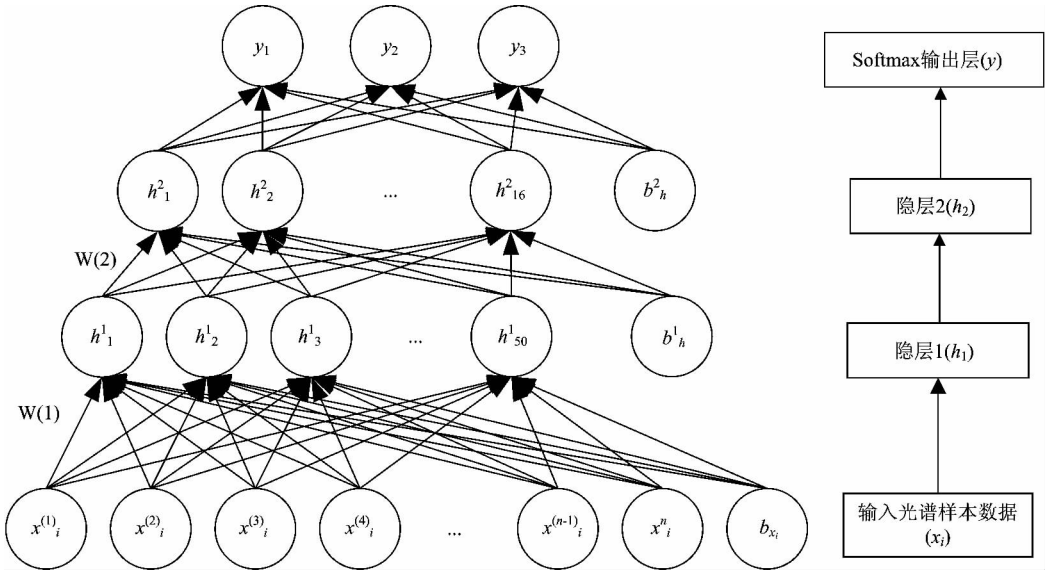

图 3 基于全谱段块状煤炭分类模型示意图

Fig 3 Sketch map of bulk-coal classification model based on full spectrum

多层感知器神经网络有两个隐层，第一个隐层连接输入层，第二个隐层连接输出层。输入模型的光谱数据集为 $X=[x_1, x_2, x_3, \cdots, x_N]$，$x_i \in R^n$。将获取的光谱数据输入模型。

$$f = \sigma(W_{h1}x + b_{h1}) \tag{1}$$

式(1)中，$x \in X$ 为输入，$f \in R^m$ 为第一个隐层的输出，$W_{h1} \in R^{m\times n}$ 为第一个隐层与输入层的连接权重，$b_{h1} \in R^m$ 为第一个隐层的偏置，σ 为非线性变换激活函数 sigmoid，如式(2)所示。

$$\sigma(x) = \text{sigmoid} = \frac{1}{1+e^{-x}} \tag{2}$$

$$g = \sigma(W_{h2}f + b_{h2}) \tag{3}$$

对于式(3)，f 为第一个隐藏层的输出，$g \in R^l$ 为第二个隐层的输出，其中 $W_{h2} \in R^{l\times m}$ 为第二个隐藏层与第一个隐层的连接权重，$b_{h2} \in R^l$ 为第二个隐层的偏置，σ 为非线性激活函数 sigmoid。

$$o = s(W_o g + b_o) \tag{4}$$

$$s(x)_j = \text{soft max} = \frac{e^{x_j}}{\sum_{k=1}^{K} e^{x_k}} \tag{5}$$

对于式(4)，$o \in R^d$ 为输出层的输出，g 为第二个隐层输入，$W_o \in R^{d\times l}$ 为输出层与第二个隐层的连接权重，$b_o \in R^d$ 为输出层的偏置，s 为非线性激活函数 soft max，如式(5)，其中 $N=219$，$n=25$，$m=50$，$l=16$，$d=3$，$j \in \{0, 1, 2\}$，$K=3$。

模型初步构建完毕后，采用反向传播算法作为监督学习方法来训练网络[11]。

模型的目标函数为

$$E = -\sum_x p(x)\log(q(x)) \tag{6}$$

式(6)中，E 为交叉熵损失，$p(x)$ 是真实标签的概率分布，$q(x)$ 为模型预测的概率分布。

模型使用一种自适应的优化算法——“adam”算法进行优化学习，初始学习率设为 $lr=0.001$。其他参数如下：batch_size=64，迭代次数 epoch_number=500(迭代次数)，防止过拟合参数 dropout=0.2。

1.5 对比实验与参数设置

选用随机森林与 SVM 算法作为分类对比实验方法。

随机森林是一个包含多个决策树的分类器，它利用 bootsrap 重抽样方法从原始样本中抽取多个样本，对每个样本建立决策树，然后通过对子决策树的判断结果进行投票，最终得出预测结果[12]。

分类模型建立过程中，为使不同数据集的分类效果最佳，进行决策树个数的最优化设定。最终设定可见-近红外数据集决策树个数 $k=15$，短波红外数据集 $k=10$，全谱段数据集 $k=10$。

支持向量机(SVM)是建立在结构风险最小化原理基础及统计学习理论 VC 维理论的一种典型的小样本学习监督分类方法[13]。

在验证实验参数设置中，核函数选定“线性核函数”。

2　结果与讨论

改变训练样本数量，对三个数据集的输入特征进行 MLP，SVM 和 RF 三种分类方法下的分类实验。分类精度指标采用分类准确率，即正确分类的测试样本除以全部测试样本。实验分类精度结果如表 1 所示，每个分类精度结果值表示 10 次实验结果的平均值。表中第一列为训练样本数。

如图 4 所示，多层感知器模型下，短波红外光谱数据集的分类准确率随训练样本的增加而升高，当训练样本数量达到 110 时，分类准确率达到 0.837 3，随着训练样本的继续增加，分类精度增长速度减慢，虽存在一定波动，但处于较为稳定的状态。短波红外数据集能够对煤炭种类进行区分是因为：煤由有机化合物和无机化合物组成，水和有机物成分均可吸收某些特定的短波红外波长[14]。不同种煤炭的水和有机物成分存在着一定差异。短波红外谱段存在着丰富的区分煤炭类别的信息，因此基于短波红外数据能够对煤炭种类进行区分。

表 1　不同训练样本数量、不同分类方法下的三种波段范围分类精度对比

Table 1　Comparison of classification accuracy of three band ranges for different training sample sizes and different classification methods

训练样本数	可见-近红外谱段			短波红外谱段			全谱段		
	MLP	RF(15)	SVM(linear)	MLP	RF(10)	SVM(linear)	MLP	RF(10)	SVM(linear)
22	0.495 1	0.471 5	0.348 5	0.543 5	0.452 6	0.378 5	0.707 4	0.498 1	0.415 6
44	0.585 9	0.595 4	0.355 8	0.666 1	0.510 8	0.4	0.818	0.613 8	0.5
66	0.698	0.634 1	0.371 6	0.756 3	0.550 2	0.418 5	0.890 5	0.659 2	0.503 8
88	0.732 7	0.674	0.383 4	0.795 6	0.594 5	0.424 9	0.929 7	0.727 1	0.559 1
110	0.821 1	0.73	0.370 7	0.837 3	0.6	0.443 3	0.968 1	0.759 3	0.583 3
132	0.873 2	0.710 8	0.386 7	0.824 5	0.658 3	0.462 5	0.980 3	0.774 2	0.602 5
154	0.890 4	0.786 8	0.389	0.842 4	0.641 8	0.471 4	0.98	0.791 2	0.617 6
176	0.936 3	0.785	0.351 7	0.853 3	0.673 3	0.486 7	0.972 3	0.79	0.62
198	0.927 3	0.8	0.406 7	0.859 3	0.683 3	0.48	0.938 7	0.803 3	0.706 7

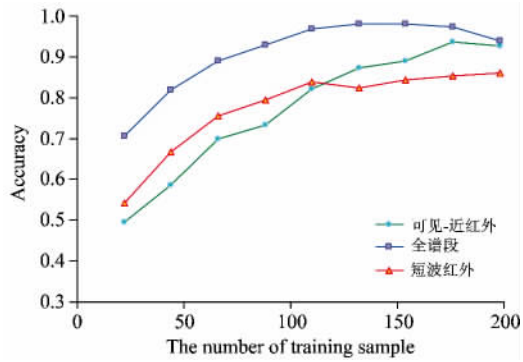

图 4　多层感知器模型下的分类准确率

Fig. 4　Classification accuracy of multilayer perceptron model

煤质的差别与杂质的含量使不同种煤炭在可见-近红外波段的信息存在差异。不同种类的煤炭由于煤化程度不同，在色彩上存在一定差异，随着煤化程度的提高，煤炭颜色会呈现出从褐色到黑色到钢灰色的变化；煤炭光泽等属性又与煤岩组分、成煤物质及变质程度息息相关[15]。这些都是煤对不同可见-近红外谱段的吸收作用的结果。可见-近红外波段含有大量的用于鉴别煤炭的有用信息，由表 1 可知，可见-近红外波段数据集能够得到较好的分类效果，分类精度随着训练样本的增加而上升，在训练样本数达到 176 时，此时分类准确率为 0.936 3，达到最高。

全谱段数据整体上提高了分类的准确率，尤其是训练样本数据量较小的情况下。训练样本数为 22 时，可见-近红外数据集分类准确率为 0.495 1，短波红外数据集为 0.543 5，而全谱段数据集综合两个波段的信息量，在分类精度上有了大幅度的提升，达到了 0.707 4。在训练样本数为 66 时，全谱段数据集已经能够达到 0.890 5 的分类准确率，能较为准确地进行煤炭种类的鉴别；而此时可见-近红外数据集与短波红外数据集的分类准确率分别为 0.698 与 0.756 3。当训练样本数为 132 时，相较于可见-近红外波段的 0.873 2 与短波红外波段的 0.824 5 的分类准确率，全谱段数据集的分类精度有了质的提升，可以达到 0.980 3 的分类准确率。

全谱段数据集在块状煤炭的分类中优势明显，是因为全谱段数据集囊括了可见-近红外数据集与短波红外数据集的全部信息，对目前应用较为广泛的近红外谱段鉴别信息进行了补充，因此能够对煤炭种类进行较为精准的区分。

且由图 4 可以看出，在 MLP 分类器下，可见-近红外数据集的分类精度随训练样本的增加而上升，训练样本为 176 时，分类准确率达到最大为 0.936 3。随着训练样本的继续增大，准确率反而呈现下降的变化趋势；全谱段数据集的分类精度变化趋势与可见-近红外类似，当训练样本为 132 时，准确率最高达到 0.980 3，然而训练样本继续增大时会导致分类精度由于过训练而下降；短波红外数据集的分类精度随训练样本的增大而增大，在训练样本数为 154 时，训练样本的变化趋于平稳，并不会有明显增高。由此可以看出，在 MLP 分类器下，块状煤炭分类时采用有限训练样本即可达到稳定的较高精度。

在对比实验部分，如表 1 所示，采用随机森林分类方法验证时，可见-近红外的分类效果有较好的表现，最高能够达

到 0.8 的分类准确率，此时短波红外波段的最高分类准确率仅为 0.683 3；而全谱段数据集最高分类准确率为 0.803 3，仍有一定的优越性，且在整个训练样本数增高的过程，分类精度均优于可见-近红外与短波红外数据集。采用 SVM 分类方法进行验证时，可见-近红外波段的分类精度较低，最高分类精度仅为 0.406 7；短波红外数据集的最高分类精度也只有 0.48。全谱段的分类精度较两者有显著的提升，最高分类精度能够达到 0.706 7。综上，全谱段基于可见-近红外谱段与短波红外谱段的结合对块状煤炭种类鉴别具有优越性及普适性。

3 结 论

采用地物光谱仪获取三种块状商品煤的光谱数据，进行去噪等预处理；根据波谱范围，分为可见-近红外、短波红外与全谱段三个数据集进行分类研究；经主成分分析，得到前 25 个主成分，输入多层感知器分类模型进行分类，有限的训练样本条件下即可达到稳定的较高分类精度，为有限训练样本下块状煤炭分析的非接触光谱识别提供了可能性；在训练样本数为 132 时，采用 MLP 分类器最高能够达到 98.03%的分类精度。并采用随机森林与 SVM 进行验证，在不同分类方法下，全谱段数据集均具有较好的分类效果，具有一定的普适性。该研究为块状商品煤的无损在线分析提供了新的思路。

References

[1] Hel L, Melnichenko Y B, Mastalerz M, et al. Energy Fuels, 2012, 26(3): 1975.

[2] Zhang Y, Zhang X L, Jia W B, et al. Applied Spectroscopy, 2016, 70(1): 101.

[3] SHAN Qing, ZHANG Xin-lei, ZHANG Yan, et al(单 卿, 张新磊, 张 焱, 等). Journal of Nanjing University of Aeronautics & Astronautics(南京航空航天大学学报), 2015, 47(5): 767.

[4] Tahmasebi A, Yu J, Li X, et al. Fuel Processing Technology, 2011, 92(10): 1821.

[5] CHENG Dong, WEN He, TENG Zhao-sheng, et al(程 栋, 温 和, 滕召胜, 等). Chinese Journal of Scientific Instrument(仪器仪表学报), 2014, 35(10): 2263.

[6] LEI Meng, LI Ming(雷 萌, 李 明). CIESC Journal(化工学报), 2012, 63(12): 3991.

[7] LI Ming, CHEN Fan, LEI Meng, et al(李 明, 陈 凡, 雷 萌, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2016, 36(9): 2793.

[8] WANG Ya-sheng, YANG Meng, LUO Zhi-yuan, et al(王雅圣, 杨 梦, 骆志远, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2016, 36(6): 1685.

[9] CHU Xiao-li(褚小立). Molecular Spectroscopy Analytical Technology Combined with Chemometrics and its Applications(化学计量学方法与分子光谱分析技术). Beijing: Chemical Industry Press(北京: 化学工业出版社), 2011.

[10] Mahmoudi J, Arjomand M A, Rezaei M, et al. Civil Engineering Journal, 2016, 2(1).

[11] Chakraborty M, Ghosh A. Computer Science, 2012, 60(13): 1.

[12] FANG Kuang-nan, WU Jian-bin, ZHU Jian-ping, et al(方匡南, 吴见彬, 朱建平, 等). Statistics & Information Forum(统计与信息论坛), 2011, 26(3): 32.

[13] DING Shi-fei, QI Bing-juan, TAN Hong-yan(丁世飞, 齐丙娟, 谭红艳). Journal of University of Electronic Science and Technology of China(电子科技大学学报), 2011, 40(1): 2.

[14] Wang Y, Yang M, Wei G, et al. Sensors & Actuators B Chemical, 2014, 193(3): 723.

[15] HE Xuan-ming(何选明). Coal Chemistry(煤化学). Beijing: Metallurgical Industry Press(北京: 冶金工业出版社), 2010.

Variety Identification of Bulk Commercial Coal Based on Full-Spectrum Spectroscopy Analytical Technique

REN Yu[1,2], SUN Xue-jian[2*], DAI Xiao-ai[1], CEN Yi[2], TIAN Ya-ming[1], WANG Nan[2], ZHANG Li-fu[2]

1. College of Earth Sciences, Chengdu University of Technology, Chengdu 610059, China
2. The State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100101, China

Abstract To obtain the precise result, complex chemical analysis or complicated sample preparation is needed in universal coal analysis methods. In the paper, a new method to distinguish the type of bulk commercial coal using full spectroscopy which combined visible and near-infrared reflectance spectroscopy (Vis-NIRS) and short-wave infrared reflectance spectroscopy (SWIR) analytical technique and Multilayer Perceptron (MLP) classification method was advanced. The method was non-contact with no sample preparation and no chemical analysis. Besides, the classification information of coal can be quickly and efficiently obtained by this method. In the paper, the band range of original spectral data whose noise was excessive was deleted. The noise of remaining part was denoised by wavelet threshold denoising method. The spectral data pretreated was divided into three data sets: Vis-NIRS data set (500～900 nm), SWIR data set (1 000～2 350 nm) and full-spectrum data set (500～2 350 nm). Principal component analysis (PCA) was adopted in three datasets. The extracted principal components were entered in the MLP classification model. Multilayer perceptron was consist of input layer, hidden layers (two layers), softmax classifier. The contrastive study of classification accuracywas made among the three datasets. Random forest and Support Vector Machine (SVM) was used to verification analysis. The research showed: in the classification research of bulk commercial coal, because of the abundant data information of full-spectrum data, a better classification result can be obtained. When the number of training sample was 132, using the MLP classifier can achieve the highest classification accuracy which was 98.03%. The classification results of random forest and SVM verified the superiority and universality of the full spectrum dataset. The method provides reliable technical support for on-line analysis of coal and development of portable coal detecting instrument.

Keywords Full-spectrum data; Variety identification of bulk commercial coal; Multilayer perceptron; Principal component analysis

(Received May 14, 2017; accepted Oct. 5, 2017)

REN Yu and SUN Xue-jian: joint first authors

*Corresponding author