

doi: 10.12068/j.issn.1005-3026.2017.10.021

基于可见光-近红外光谱的煤种分类方法

宋亮, 刘善军, 毛亚纯, 李天子

(东北大学资源与土木工程学院, 辽宁沈阳 110819)

摘要: 利用便携式地物光谱仪 SVC HR-1024 对 92 个烟煤和 58 个褐煤样本进行光谱测试, 烟煤和褐煤在可见光-近红外波段光谱特征差异明显, 褐煤的光谱反射率及其斜率均明显高于烟煤。在光谱特征分析的基础上, 利用 MAO 模型法、随机森林法、BP 神经网络法和 ELM 算法进行煤种分类。结果表明: MAO 模型法和随机森林法的分类结果较优。若进行大面积、快速遥感识别时, 对分类时间要求较高, 应选择 MAO 模型法; 若是小面积单一矿区分类, 对分类准确率要求较高, 选择随机森林法较为恰当。

关键词: 烟煤; 褐煤; 可见光-近红外; 遥感; 分类

中图分类号: P 237

文献标志码: A

文章编号: 1005-3026(2017)10-1473-04

Coal Classification Based on Visible and Near-Infrared Spectrum

SONG Liang, LIU Shan-jun, MAO Ya-chun, LI Tian-zi

(School of Resources & Civil Engineering, Northeastern University, Shenyang 110819, China. Corresponding author: LIU Shan-jun, E-mail: liusjdr@126.com)

Abstract: The portable spectrometer SVC HR-1024 was used to carry out spectral tests on the 92 bituminite and 58 lignite coal samples from various coal mines. By comparing their spectral curves, the differences between bituminite and lignite samples can be observed visibly in spectral characteristics. The reflectance of lignite samples is obviously higher than that of bituminite samples, as well as the slope of spectral curves. On the basis of spectral characteristics analysis, the MAO model algorithm, random forests, BP neural networks and ELM-neural network were selected for the classification of bituminite and lignite samples. The results indicated that the MAO model algorithm and random forest algorithm outperform other algorithms on classification. For large-area and rapid recognition by remote sensing, the MAO model algorithm has a great advantage in the classification time. While the random forest algorithm can be used for classification in small mining areas.

Key words: bituminite; lignite; visible and near-infrared; remote sensing; classification

煤炭是我国的主要能源,在国民经济中占有重要地位。对煤炭资源进行探查、及时补充是国民经济可持续发展的保障。近年来,随着遥感技术的快速发展,利用遥感技术进行煤炭资源勘查成为有效方法之一。

文献[1-3]利用 Landsat, SPOT 卫星等多光谱数据,在新疆、西藏、青海、四川、云南等地进行了煤炭资源勘查和预测,发现了许多大型的煤矿床,显示了遥感技术的优越性。Mao 等^[4]基于煤的可见光-近红外光谱特征建立了不同煤体遥感

识别模型,并将该模型应用于 Landsat 卫星数据,结果表明该方法在分类效率和分类准确率上都有明显的优势。

另外,随着机器学习的迅速发展,其在光谱分类方面的应用也很多。程术希等^[5]基于大白菜种子的高光谱数据,验证了 Ada-Boost 算法、极限学习机、随机森林和支持向量机的判别效果。结果认为利用高光谱结合机器学习对大白菜种子进行快速无损分类识别是可行的,为大白菜种子批量化在线监测提供一种新的方法。但是利用机器学

收稿日期: 2016-07-25

基金项目: 国家自然科学基金资助项目(41371437)。

作者简介: 宋亮(1989-),男,山西长治人,东北大学博士研究生;刘善军(1965-),男,河北涿鹿人,东北大学教授,博士生导师。

习在不同煤种的分类上,目前还没有相关应用.

鉴于此,本文首先利用 SVC HR-1024 光谱仪测试了烟煤和褐煤的可见光-近红外光谱,分析褐煤和烟煤的光谱特征及其差异.在此基础上,利用 MAO 模型对不同煤种进行分类,同时在机器学习方面,分别利用随机森林法、BP 神经网络法和 ELM 算法等方法进行分类.最后比较 4 种方法的分类结果,可以根据不同应用需求选择最恰当的分类方法.

1 实验部分

1.1 样品采集

烟煤和褐煤样本均来自于我国主要的产煤地区,如山西省、陕西省、内蒙古自治区等,样本详细情况见表 1.

表 1 样本详细情况列表
Table 1 List of samples

矿区	类型	产地	样本数
霍林河	褐煤	内蒙古自治区	5
伊敏	褐煤	内蒙古自治区	28
白音华	褐煤	内蒙古自治区	25
神东	烟煤	陕西省	9
兖州	烟煤	山东省	8
平朔	烟煤	山西省	5
木里	烟煤	青海省	9
霍州	烟煤	山西省	5
铁法	烟煤	辽宁省	12
蒲河	烟煤	辽宁省	30
鄂尔多斯	烟煤	内蒙古自治区	14

在研究区的不同地点采集 150 个煤样本,其中 8 个烟煤矿区共 92 个样本,3 个褐煤矿区共 58 个样本.

1.2 光谱测试

用美国 Spectra Vista 公司的 SVC HR-1024 便携式地物光谱仪对样品进行光谱测试.该仪器质量约 3 kg,有 1 024 个通道,采样的最小积分时间为 3 ms.测试波段范围为 350 ~ 2 500,350 ~ 1 000,1 000 ~ 1 850 和 1 850 ~ 2 500 nm 的光谱分辨率分别为 3.5,9.5 和 6.5 nm.光谱测试时,选择 4°镜头,每个样品重复测试 2 次,取其反射率平均值^[6].

由于样本数量较多,结合烟煤和褐煤的光谱曲线,选取部分典型样本进行光谱特征分析.图 1 为选择的 6 个烟煤样本和 4 个褐煤样本光谱测试

结果.

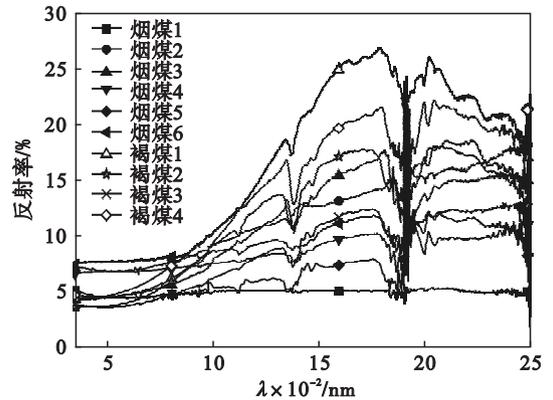


图 1 6 个烟煤样本和 4 个褐煤样本的可见光-近红外光谱线

Fig. 1 Visible and near-infrared spectrum of 6 bituminite and 4 lignite samples

由图 1 可知烟煤和褐煤的主要光谱特征:褐煤的光谱曲线在整个波段变化很大;在可见光波段(350 ~ 600 nm),光谱区曲线变化较平缓;在红光-近红外波段(600 ~ 1 780 nm)反射率快速升高,在近红外波段(2 000 ~ 2 500 nm)反射率缓慢降低.其中,在 1 400 nm 和 1 900 nm 附近有明显的吸收峰,主要是由水气和羟基引起的.

烟煤光谱曲线在整个波段反射率较低,变化也较为平缓.在波段 350 ~ 900 nm 之间反射率变化平缓,在 900 nm 后缓慢升高,在 2 000 nm 后缓慢升高或者基本保持不变.与褐煤相同,在 1 400 nm 和 1 900 nm 两处有水气和羟基引起的吸收峰.

两种煤光谱的主要差别:①在 900 ~ 2 500 nm (尤其在 1 550 ~ 1 750 nm),褐煤的反射率整体高于烟煤.②在 600 ~ 1 780 nm 褐煤的光谱反射率快速升高斜率较大,而烟煤在该波段范围斜率较小.③在近红外波段(2 000 ~ 2 500 nm)褐煤反射率下降,而烟煤大多呈上升趋势或基本保持不变.

2 数据处理与分类

2.1 数据预处理

通过数据分析发现,由光谱仪测试的光谱曲线在 350 ~ 2 500 nm 之间有 973 个波段,各个波段之间存在高度的相关性.采用主成分分析法(principal component analysis, PCA)消除数据之间的信息冗余.

主成分分析法是在不丢失主要光谱信息的前提下,选择数目较少的新变量替代原来较多的变量,从而达到对数据降维、简化模型的目的^[7].为

了减少有效信息的丢失,在本次实验中,选取前 5 个主成分表征原来的光谱数据信息,可代表 99% 以上的信息. 前 5 个主成分累积贡献率如表 2 所示.

表 2 主成分分析的方差贡献率

Table 2 Variance contribution rate of principal component analysis

主成分	PC1	PC2	PC3	PC4	PC5
方差贡献率	50.444	42.754	5.169	0.547	0.216
累积贡献率	50.444	93.198	98.367	98.914	99.130

2.2 数据分类方法

对数据进行预处理后,分别选用 MAO 模型法、随机森林法、BP 神经网络法和 ELM 神经网络法进行分析对比.

Mao 等^[4] 基于煤的可见光-近红外光谱特征建立了煤体遥感识别模型(MAO 模型),其中烟煤的识别模型为

$$\left. \begin{aligned} I_{\text{ndc}} > 0.03, \\ 0.0519 < M_5 < 0.1358. \end{aligned} \right\} \quad (1)$$

式中, I_{ndc} 为归一化煤指数:

$$I_{\text{ndc}} = \frac{M_5 - M_4}{M_5 + M_4}. \quad (2)$$

其中, M_4 和 M_5 分别表示 Landsat 卫星专题绘图仪数据的第 4 波段(0.76 ~ 0.90 μm) 和第 5 波段(1.55 ~ 1.75 μm) 的反射率. I_{ndc} 的大小反映了光谱曲线在 M_4 和 M_5 两个波段反射率的差异.

同理,褐煤的识别模型为

$$\left. \begin{aligned} I_{\text{ndc}} > 0.36, \\ 0.1298 < M_5 < 0.2552. \end{aligned} \right\} \quad (3)$$

随机森林是文献[8-9]提出的一种集成学习算法,其实质是对决策树算法的一种改进.它是利用 bootstrap 重抽样的方法抽取多个样本,并对每个样本进行决策树建模,组合多棵决策树对结果进行预测,并用投票的方法得出最终结果.本文选择的决策树的数目为 100,选择的特征属性值为 2.

BP 神经网络是一种常用的神经网络方法.该方法可以看作是对原始特征空间进行非线性变换,产生一个新的样本空间,使其变换后的特征线性可分,同传统统计方法相比,其分类器与概率分布无关.

极限学习机(extreme learning machine, ELM)算法是一种改进的神经网络算法,该算法随机产生输入层与隐含层之间的连接权值及隐含层神经元的阈值,在训练过程中只需设置隐含层

神经元个数,便可以获得唯一最优解^[10].

3 结果与讨论

本文采用 10 折 10 次交叉验证的方法检验其分类精度^[11]. 将 150 个样本随机平均分成 10 个子样本,轮流用其中的 9 个子样本作为训练样本,剩余的 1 个子样本作为测试样本.循环完 10 次后,所有样本都被预测了 1 次,将 10 次结果的均值作为 1 次 10 折的分类精度评价.为了能够更准确地确定其分类精度,重复上述过程 10 次,将 10 次 10 折的平均分类精度作为最终分类结果.

本文所有算法均由 Matlab R2009(a) 编程实现.实验所用电脑配置为 CPU: Core(TM) - i3, 主频 3.4 GHz, 4.00 GB 内存, Win7 操作系统, GPU 为 NVIDIA GeForce GT 620. 4 种方法的分类准确率和分类时间如表 3 所示.可知 4 种分类方法的运行时间和分类准确率均有所不同,较优的为 MAO 模型法和随机森林分类法.分类准确率从高到低依次为随机森林法、MAO 模型法、BP 神经网络法和 ELM 算法.其中随机森林分类法的分类准确率最高,为 91.2%;从分类时间上看,时间最短的是 MAO 模型法,耗时 0.74 ms, ELM 神经网络和随机森林法次之,耗时最长的是 BP 神经网络法,为 1991.67 ms,是 MAO 模型分类时间的 2690 倍.

因此针对不同的分类需求可以选择恰当的分类方法.若进行大面积、快速分类时,对分类时间要求较高,应选择 MAO 模型法,该方法在分类时间上有极大的优势,同时分类准确率较高,能达到 90%;若进行小面积单一矿区分类的话,应选择随机森林分类法较为恰当,该方法的分类准确率最高,且分类时间较短.

表 3 不同分类方法结果对比

Table 3 Comparison of results for different classification methods

分类方法	分类准确率/%	分类时间/ms
MAO 模型法	90.00	0.74
随机森林法	91.20	74.59
BP 神经网络法	85.66	1991.67
ELM 算法	82.90	37.08

4 结 论

1) 烟煤和褐煤在可见光-近红外波段(特别

是在 900 ~ 2 500 nm 范围内)光谱特征差异明显,褐煤的光谱反射率及斜率均明显高于烟煤。

2) 针对不同煤种光谱数据,利用 MAO 模型法、随机森林法、BP 神经网络法和 ELM 算法进行分类。从分类准确率和分类时间综合考虑,较优的分类方法为 MAO 模型法和随机森林分类法。

3) 若针对大面积、快速分类时,对分类时间要求较高,应选择 MAO 模型法;若是小面积单一矿区分类的话,对分类准确率要求较高时,选择随机森林分类法较为恰当。

参考文献:

- [1] 李新萍,郝多虎,段朋,等. 遥感技术在新疆东部主要含煤区地址解译中的应用[J]. 水土保持研究,2012,19(2): 253-263.
(Li Xin-ping, Hao Duo-hu, Duan Peng, et al. Application of remote sensing technology in the main area of coal geological interpretation in eastern Xinjiang Uygur autonomous region [J]. *Research of Soil and Water Conservation*, 2012, 19(2): 253-263.)
- [2] 郭晓波,管海晏,吴查查. 中巴卫星遥感找煤技术研究[J]. 中国煤炭,2012,38(4):29-31.
(Guo Xiao-bo, Guan Hai-yan, Wu Cha-cha. Research on CBERS remote sensing technique for coal exploration [J]. *China Coal*, 2012, 38(4): 29-31.)
- [3] Tan K L, Wan Y Q, Sun S X, et al. Prospecting for coal in China with remote sensing [J]. *Journal of China University of Mining & Technology*, 2008, 18(4): 0537-0545.
- [4] Mao Y C, Ma B D, Liu S J, et al. Study and validation of a remote sensing model for coal extraction based on reflectance

spectrum features [J]. *Canadian Journal of Remote Sensing*, 2014, 40(5): 327-335.

- [5] 程术希,孔汶汶,张初,等. 高光谱与机器学习结合的大白菜种子品种鉴定[J]. 光谱学与光谱分析,2014,34(9): 2519-2522.
(Cheng Shu-xi, Kong Wen-wen, Zhang Chu, et al. Variety recognition of Chinese cabbage by hyperspectral imaging combined with machine learning [J]. *Spectroscopy and Spectral Analysis*, 2014, 34(9): 2519-2522.)
- [6] 宋亮,刘善军,虞茉莉,等. 基于可见光-近红外和热红外光谱联合分析的煤和矸石分类方法研究[J]. 光谱学与光谱分析,2017,34(9):2519-2522.
(Song Liang, Liu Shan-jun, Yu Mo-li, et al. A classification method based on the combination of visible, near-infrared and thermal infrared spectrum for coal and gangue distinguishment [J]. *Spectroscopy and Spectral Analysis*, 2017, 34(9): 2519-2522.)
- [7] Chang C W, Laird D A, Mausbach M J, et al. Near-infrared reflectance spectroscopy—principal components regression analyses of soil properties [J]. *Soil Science Society of America Journal*, 2001, 6(2): 480-490.
- [8] Breiman L. Random forests [J]. *Machine Learning*, 2001, 45(1): 5-32.
- [9] Daniel R G, Wade A W, Andrew J F, et al. Predicting USCS soil classification from soil property variables using random forest [J]. *Journal of Terramechanics*, 2016, 65(1): 85-92.
- [10] Huang G B, Zhu Q Y, Siew C K. Extreme learning machine: theory and applications [J]. *Neurocomputing*, 2006, 70: 489-501.
- [11] Diamantidis N A, Karlis D, Giakoumakis E A. Unsupervised stratification of cross-validation for accuracy estimation [J]. *Artificial Intelligence*, 2000, 116(1): 1-4.

(上接第 1472 页)

- [12] Luan Q Z, Liu H P, Xiao Z Q. Comparison between algorithms of ortho rectification for remote sensing images [J]. *Remote Sensing Technology Application*, 2007, 22(6): 743-747.

- [13] 杨彬,李旭东,晏磊,等. 航空遥感图像拼接算法及其精度评价[J]. 计算机工程,2014,40(3):253-257.
(Yang Bin, Li Xu-dong, Yan Lei, et al. Mosaic algorithm for aerial remote sensing image and its accuracy evaluation [J]. *Computer Engineering*, 2014, 40(3): 253-257.)