

基于太赫兹时域光谱技术的煤岩识别方法研究

苗曙光¹, 邵丹^{1*}, 刘忠育^{2,3}, 樊强¹, 李素文¹, 丁思杰^{2,3}

1. 淮北师范大学物理与电子信息学院, 安徽 淮北 235000

2. 中国矿业大学信息与控制工程学院, 江苏 徐州 221116

3. 中国矿业大学物联网(感知矿山)研究中心, 江苏 徐州 221116

摘要 煤岩识别一直是制约煤矿无人化开采的关键问题之一。传统的人工采煤因为工作环境极其复杂, 很难精准地找到煤岩的分界面, 容易造成欠切割或过切割现象。太赫兹光谱技术作为一种无损探测技术, 能够反映出被测物体的物理和化学信息, 可以成为研究煤岩识别的有效方法。采用太赫兹时域光谱技术与多元统计法—聚类分析(CA)和主成分分析(PCA)相结合的方法来识别不同种类的煤岩。通过透射式太赫兹光谱仪获得六种煤岩样品的太赫兹光谱, 对其进行FFT等一系列数学计算可以得到各种样品的折射率、吸收系数以及介电常数。计算结果表明不同种类的煤岩在折射率、吸收系数上都存在差异。分析各类煤炭样品的折射率和吸收系数与样品的各组成成分含量之间的关系, 可以发现碳含量是影响其样品折射率大小的因素之一, 灰分含量是影响其样品吸收系数大小的因素之一。聚类分析中两类样品的欧氏距离与主成分分析中的第一主成分(PC1)得分都能反映煤岩样品之间的相似性和相异性, 并且CA与PCA的结果保持一致。分别将各类样品在0.5~2.5 THz频率范围内的折射率、吸收系数与CA和PCA结合, 组成太赫兹数据与煤岩之间的模型。分析表明: 根据不同样品之间的相似性, 两种模型中六种煤岩样品均被分为两类; 在各种样品的吸收系数与CA-PCA组成的模型中, 四种煤炭被聚集在一起, 并且石英砂岩(GSR-4)具有很好的独特性: 石英砂岩拥有最小的PC1得分值以及石英砂岩与第二类之间的欧氏距离最大, 为219.03。由此可见采用太赫兹技术与多元统计方法结合, 可以实现煤岩的准确识别, 识别准确率可以达到100%。

关键词 太赫兹时域光谱; 煤岩识别; 主成分分析; 聚类分析

中图分类号: O441.6 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2022)06-1755-06

引言

《煤炭工业发展“十三五”规划》指出: 煤炭工业是关系国家经济的命脉和能源安全的重要基础产业。煤炭是我国的基础能源和重要原料, 占我国化石能源资源的90%。近几年, 煤炭在一次能源消费中的比重逐渐减少, 但煤炭的主要能源地位不会变化。在煤矿开采的过程中, 为了更最大限度地使用资源和实现采煤机的自动截割, 人们希望采煤机能够沿着煤岩的分界面进行切割, 然而采煤工作环境复杂, 仅仅靠采煤工人的耳听目测难以实现对煤岩界面的准确判别, 所以研究煤炭自动识别技术变得非常有意义, 该问题也一直是国内外的研究热点。

近年来, 太赫兹光谱技术作为一项在线检测技术, 在农

业、医学、食品安全、航天等领域拥有巨大的应用前景^[1-3]。太赫兹(Terahertz, THz)波通常指的是频率在0.1~10 THz(波长在0.03~3 mm)之间的电磁波, 其波段在微波和红外光之间, 属于远红外波段^[4]。太赫兹光谱含有丰富的样本信息, 通过研究该光谱可以获得样品的折射率、吸收系数等数据。在太赫兹频段下不同种类煤质的响应存在差异, 且煤炭的几种成分(氢、挥发分、碳、灰分等)含量影响着太赫兹常数^[5-7]。本文引入两种多元统计分析法—主成分分析(principal component analysis, PCA)与聚类分析(cluster-analysis, CA), 再次处理所获得的太赫兹数据, 并把所得到第一主成分(the first principal component, PC1)得分与欧氏距离联系起来, 由此确定煤炭之间的相似性以及煤岩之间的差异性, 以此实现分类。

收稿日期: 2021-06-05, **修订日期:** 2021-08-02

基金项目: 国家自然科学基金项目(52074273), 安徽省高校自然科学基金重点项目(KJ2020A0027)资助

作者简介: 苗曙光, 1983年生, 淮北师范大学物理与电子信息学院讲师 e-mail: msgmeu@126.com

* 通讯作者 e-mail: 191780618@qq.com

1 实验部分

1.1 材料

太赫兹时域光谱(Terahertz time-domain spectroscopy, THz-TDS)系统通常分为两种:反射式系统和透射式系统。本实验采用的是 AdvantestTAS7400SU 透射式太赫兹光谱系统,该系统频谱范围 0.5~7 THz,每个谱的扫描时间约为 200 ms,频率分辨率为 7.6 GHz,其结构示意图如图 1 所示。该系统主要由飞秒激光器、延迟线、THz 发射器、样品测试区、THz 探测器组成,由飞秒激光器发射出的光束被分光器分为两束光。一束光经过多次反射作用到太赫兹发射器上使其产生太赫兹脉冲,太赫兹脉冲通过透镜和 M4 后聚焦在样品上。另一束光经过延迟区后与携带样品信息的太赫兹脉冲相遇在 THz 探测器上,产生光电流。延迟线可以消除飞秒激光与 THz 脉冲之间的时间延迟,使系统能得到完整的时域波形^[8-9]。

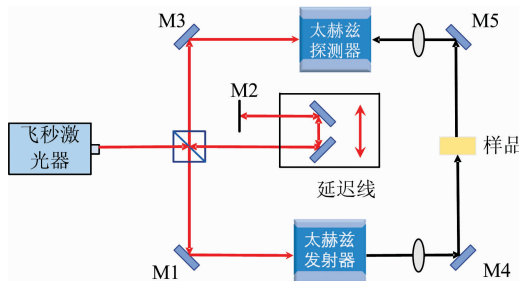


图 1 太赫兹时域光谱系统基本结构
Fig. 1 The basic structure of terahertz time domain spectroscopy system

实验中用于制作样品的六种煤岩粉末:ZBM111C(烟煤)、ZBM111A(烟煤)、ZBM102(烟煤)、ZBM097A(无烟煤)、ZBM110B(煤矸石)、GSR-4(石英砂岩)均来自于国家标准物质中心网,其物理特性和化学成分见表 1。为了防止样品在测量的过程中因厚度太小而破碎,在使用压片机对煤岩粉末压片之前加入太赫兹波对其吸收率几乎为零的聚乙烯粉末,将煤岩粉末与聚乙烯按照 1:2 的比例掺加,然后将其倒入压片机磨具中,用 10 MPa 的压强压置 5 min,最后得到厚度为 1.1 mm 的样品。为了消除环境温度和湿度对测量结果的影响,实验在室温(24 °C)下进行,光路部分通入干燥气体,

表 1 六种煤岩标准物质的编码及对应的性质参数
Table 1 Code of six coal-rock standard materials and corresponding property parameters

样品编码	全硫/%	灰分/%	挥发分/%	碳/%	氢/%	氮/%
ZBM111C	0.92	8.00	31.29	77.14	4.59	1.26
ZBM111A	0.81	9.62	33.64	74.16	4.44	1.38
ZBM102	1.26	25.88	30.43	57.82	3.40	1.02
ZBM097A	1.03	11.87	8.99	79.96	3.31	1.12
ZBM110B	1.75	85.72	9.12	5.80	1.03	0.27
GSR-4	/	/	33	0.06	/	21

保持相对湿度低于 1%,并且每种样品重复测量三次取平均值作为样品信号。

1.2 原理

生活中许多材料以及电介质的振动能级的大小都落在太赫兹波段范围内。所以当处于太赫兹频段的电磁波照射在样品上时,被测样品会吸收其光子能量,从而产生含有被测样品物理信息的太赫兹光谱。将太赫兹光谱仪输出的太赫兹光谱进行快速傅里叶变换(fast Fourier transform, FFT),然后通过公式计算出样品的折射率、消光系数和吸收系数^[10],计算公式如式(1)~式(3)

$$n(\omega) = \frac{\varphi(\omega)c}{\omega d} + 1 \quad (1)$$

$$k(\omega) = \frac{c}{\omega d} \ln \left\{ \frac{4n(\omega)}{\rho(\omega)[n^2(\omega) + 1]^2} \right\} \quad (2)$$

$$\alpha(\omega) = \frac{2\omega k(\omega)}{c} \quad (3)$$

式(1)~式(3)中, $\varphi(\omega)$ 为测试样品信号与参考信号的相位差; $\rho(\omega)$ 为测试样品信号与参考信号幅度的比值; d 为样品厚度,单位为 m; c 为光速,单位为 $\text{m} \cdot \text{s}^{-1}$; ω 为角频率,单位为 $\text{rad} \cdot \text{s}^{-1}$ 。

本文使用聚类分析和主成分分析分别处理上述计算所得的折射率等太赫兹数据,得到的结果可以体现各种煤岩样本之间的相似性。聚类分析是在没有给定划分类别的情况下,根据样本相似度进行样本分组的一种方法,是一种非监督的学习算法。该算法是基于“物以类聚”的原理,根据不同种类样品之间的内部相似性将样品分为多个类别的数学分析方法^[11]。在分析过程中,根据样品数据之间的特征,计算各种样品之间的欧氏距离,将欧氏距离最小的两种样品组成一个新的类,然后继续与其他样品进行计算,以此类推,最终将所有样品合为一类。同一类的样品具有很强的相似性,不是同一类的样品就具有很大的差异性。主成分分析是一种统计方法,通过正交变换将一组可能存在相关性的变量转换为一组线性不相关的变量,转换后的这组变量叫主成分^[12]。主成分包含的信息不会重复并且第一个主成分包含最大的信息量。通过绘制第一主成分的得分直方图,可得到样本之间的相互关系,然后对样品做出分类。聚类分析和主成分分析都是通过逻辑运算,建立简单的模型来提取原始数据中无法直接表达的重要信息,两种方法都可以简化复杂的样本数据。

2 结果与讨论

2.1 太赫兹时域光谱分析

将实验样品按照实验原理的方法进行测量,得出几种煤岩样品的太赫兹时域光谱数据,其波形如图 2(a)所示。分析图 2(a),可以看出不同样品的太赫兹光谱有着不同的延迟和振幅。其中 ZBM111A 的主峰出现的最早,在 17.69 ps 时出现,振幅为 0.089 V;而 ZBM097A 相对于 ZBM111A 就拥有最长的时间延迟,主峰出现时间为 18.09 ps,振幅为 0.082 V。太赫兹脉冲穿过被测物体时的波速与光程因各样品的折射率的不同而不同,所以各种样品出现主峰的时间也不相同。图 2(b)是几种样品的功率随频率的变化趋势,可以看出

在频率 0.5~3.5 THz 的范围内脉冲透过 GSR-4 时能量消耗最大，其在 ZBM111C, ZBM111A, ZBM102, ZBM097A, ZBM110B 中的能量消耗逐渐减小。结合图 2(a)和(b)可知，六类样品的时域光谱波形表现出不同的延迟，并因为 THz

脉冲在各类样品中能量的消耗有所不同，导致各类样品的振幅出现了不同程度的衰减，这也表明使用太赫兹时域光谱技术识别煤岩是可行的。

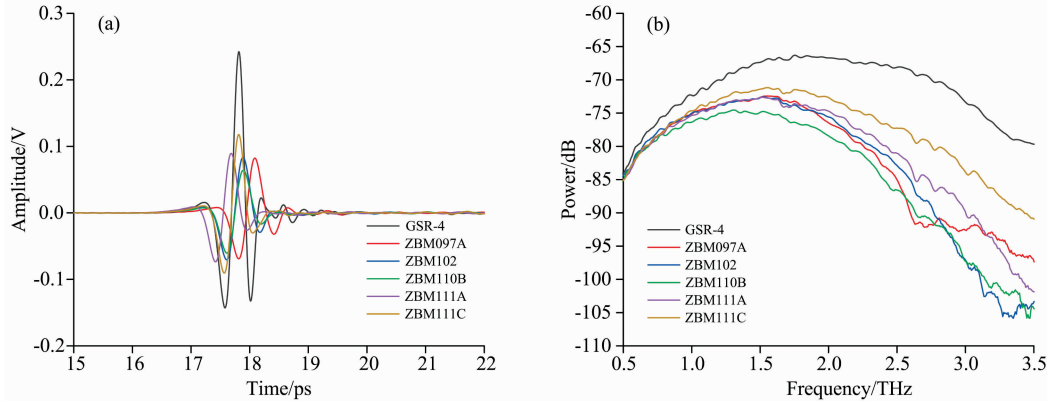


图 2 六种样品的时域谱(a)及功率谱(b)

Fig. 2 Time-domain spectra (a) and power spectra (b) of six samples

将样品的时域光谱数据代入计算式(1)和式(3)分别可以得到样品的折射率和吸收系数，如图 3(a)和(b)所示。由图 3(a)可知六种样品在 0.5~3.5 THz 波段内的折射率有着明显的差异，其中 ZBM110B 的折射率最大，ZBM097A 的折射率以微弱的差距小于 ZBM110B；而 ZBM111A 与 ZBM111C 的折射率几乎相同，ZBM102 的折射率最小。根据表 1 各种煤炭样品的物理性质可以得到 ZBM097A 的碳含量最高，为 79.96%；其次 ZBM111A 与 ZBM111C 的碳含量相近，分别为 74.16%和 77.14%；ZBM102 的碳含量最低，为 57.82%。几种煤炭样品中的碳含量的排列顺序与样品的折射率大小的排列顺序完全一致，而其他成分的含量没有此类的排序，说明煤炭样品中碳含量的大小影响折射率的大小。样品 ZBM097A 与样品 ZBM111C 的折射率的差异相对较大，但两者的碳含量却很接近，相对差为 2.82%，而两者的灰分与挥发分差值较大。由此可以看出，煤炭样品的折射率是由样品所含有的成分综合决定的，其中碳含量是主要因素之一。

图 3(b)表示六种样品在 0.5~3.5 THz 频率内吸收系数的波形。按照折射率的分析方法，将样品的吸收系数从大到小排列起来：样品 ZBM110B 的吸收系数最大，ZBM102 次

之，ZBM097A, ZBM111A 和 ZBM111C 三者的吸收系数以微小的差异递减、GSR-4 拥有最小的吸收系数。为了确定煤炭中哪些成分会影响吸收系数，将表格 1 中样品的物理性质做出排列，即发现样品中灰分含量的大小顺序与样品吸收系数的大小顺序完全相同。这就表明样品中灰分的含量影响着煤炭的太赫兹吸收系数。

样品的介电常数可由样品的消光系数和折射率求得，其计算公式如式(4)^[13]

$$\hat{\epsilon} = \epsilon_r + i\epsilon_i, \epsilon_r = n^2 - k^2, \epsilon_i = 2nk \quad (4)$$

式(4)中， ϵ_r 是介电常数的实部， ϵ_i 是介电常数的虚部。

根据式(4)得出六种样品的介电常数波形，如图 3(c)所示。由图可以看出，样品的介电常数与样品的折射率随着频率的变化有着相同的趋势，ZBM110B 的介电常数高于其他样品的介电常数，GSR-4 的介电常数仅高于 ZBM102 的介电常数。由式(4)还能看出当电磁波达到太赫兹频段时，样品的折射率影响着介电常数，所以决定折射率大小的因素也影响着介电常数的大小，即碳含量是影响样品介电常数大小的主要因素之一。

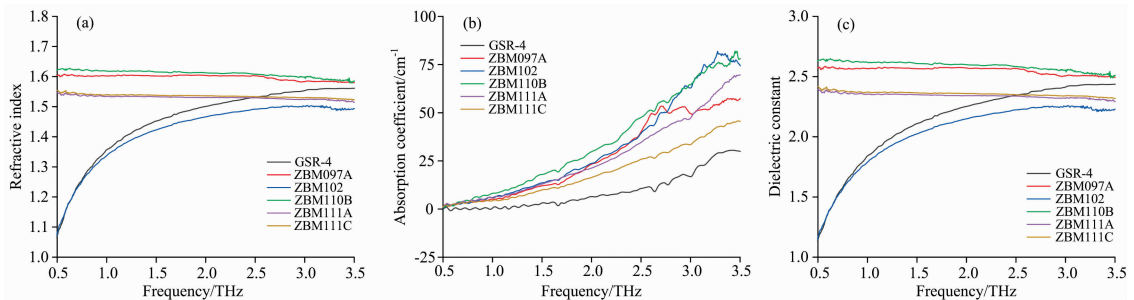


图 3 六类样品的折射率(a)、吸收系数(b)以及介电常数(c)

Fig. 3 Refractive index (a), absorption coefficient (b) and dielectric constant (c) of six samples

2.2 煤岩分类

采用聚类分析和主成分分析的方法对不同种类的煤岩进行分类。在聚类分析中, 首先将每个样品各自分为一类, 计算各类样品之间的距离, 随后按规定将距离最近的两个样品

归为一个新的类, 然后计算该新类与其他样品之间的距离, 重复将两个最近的类合并, 最后将所有样品都归为一类。文中六类样品经过五步聚类, 获得反映六种样品相似性、相异性的树状图。

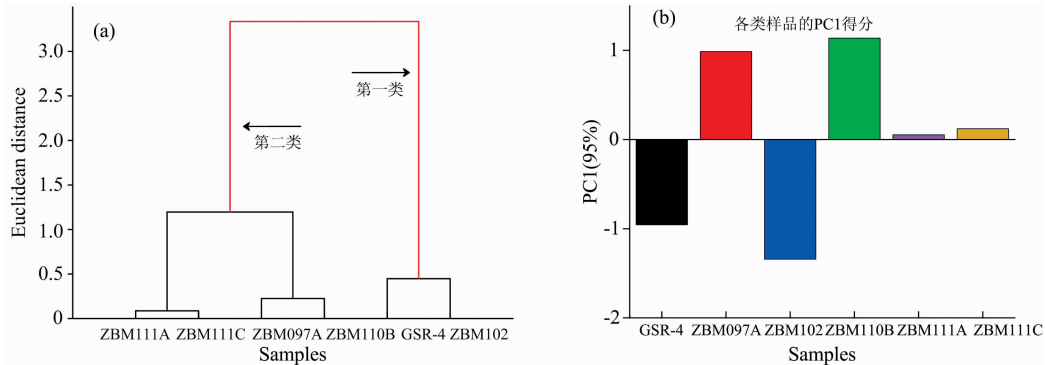


图 4 基于折射率的树状图(a)与 PC1 得分直方图(b)

Fig. 4 Euclidean distance dendrogram (a) and PC1 score histogram (b) with the input of refractive index

图 4(a)是将各类样品在 0.5~2.5 THz 频率范围内的折射率数据作为输入集, 运用聚类分析法得到基于折射率谱的树状图。从图中可以看出, 六种样品被分为两类: 第一类由样品 GSR-4 和 ZBM102 组成, 其余样本分为第二类。在六种样品中 ZBM111A 和 ZBM111C 之间的欧氏距离最小, 为 0.086, 其二者拥有最高的相似性, 从图 3(a)中也可看出两者对应的折射率曲线几乎完全重合; ZBM097A 与 ZBM110B 组成的新类次之, GSR-4 与 ZBM102 随后。同时, 由树状图可知四种煤炭样品被分为三类: ZBM111A 和 ZBM111C, ZBM097A, ZBM102, 对应各样品的碳含量, 正好验证了碳含量是影响其样品折射率的因素之一。

对样品的折射率数据进行主成分分析, 并保留从太赫兹折射率中提取出的前三个主成分。前三个主成分对原数据的贡献率达到了 100%, 其中第一主成分的贡献率为 95%。图 4(b)为各类样品的折射率 PC1 得分直方图。在主成分分析中, 各主成分的重要性能用相应的贡献率来描述, 贡献率越大, 主成分反映的信息就越多。两个样本之间的相似性与样本的 PC1 得分有关, 两样本的 PC1 得分差值越小, 样本就越相似; 反之, 样本的 PC1 得分差值越大, 差异也就越大。结合

CA 与 PCA, 对比图 4 中折射率的树状图和 PC1 直方图, 可以清楚地发现 ZBM111A 与 ZBM111C 的 PC1 偏差最小, 所以在树状图中两者的欧氏距离最小, 被分为一类。其次 ZBM097A 和 ZBM110B 之间、GSR-4 和 ZBM102 之间的 PC1 得分偏差与欧氏距离都逐渐增大。由此看出树状图表现出样品的相似性与由 PC1 得分表现出的样品相似性是一致的。

同理, 将各类样品在 0.5~2.5 THz 频率范围内的吸收系数作为输入数据, 分别采用 CA 和 PCA 方法分析, 得到基于吸收系数谱的树状图和其 PC1 得分直方图, 如图 5 所示。由图 5(a)可知六种样品被分为两类, GSR-4 被单独分为第一类, 其他 5 种样品作为第二类: GSR-4 与第二类间的欧氏距离最大, 为 219.03; 第二类中, ZBM097A 与 ZBM102 之间的距离最小为 22.4, 其二者组成的新类与 ZBM111A 的欧氏距离略有增大, ZBM111C 和 ZBM110B 也依次与上一步的新类聚集。由此可以看出基于吸收系数的 CA 分析可以明确地将四种煤炭聚集在一起, 并鉴别出 ZBM110B 与 GSR-4。根据此聚类分析结果, 观察各种煤炭样品中灰分的含量可以验证样品中灰分的含量影响着煤炭的太赫兹吸收系数。

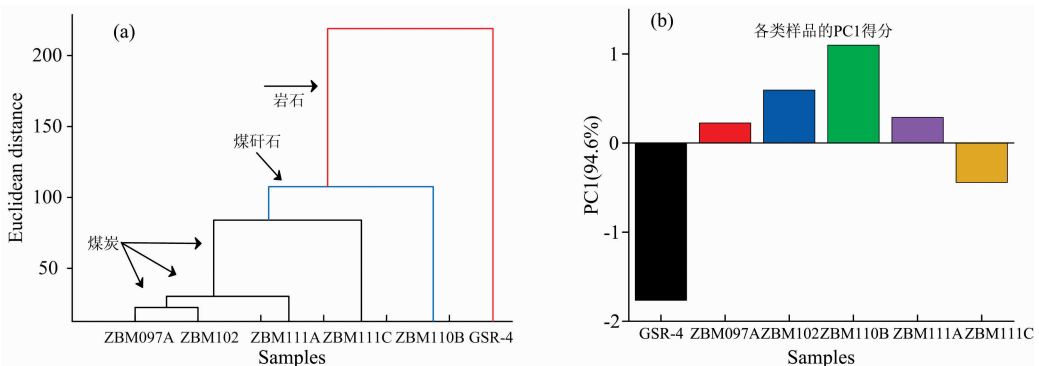


图 5 基于吸收系数的树状图(a)与 PC1 得分直方图(b)

Fig. 5 Euclidean distance dendrogram (a) and PC1 score histogram (b) with the input of absorption coefficient

对样品的吸收系数进行主成分分析, 其中前三个主成分对原数据的贡献率达到了 99.2%, 第一主成分的贡献率为 94.6%。图 5(b)中 GSR-4 的 PC1 得分与其他样品差距较大, 对比图 5(a)和(b), 可以发现不论是在 CA 分析还是 PCA 分析中 GSR-4 都具有独特性; ZBM110B 拥有最大的 PC1 正得分值, ZBM097A 和 ZBM102 之间的 PC1 得分很近, 所以二者具有一定的相似性, 这些结论都与 CA 分析相对应。由此可知, 以吸收系数作为输入数据, 结合 CA 与 PCA 分析模型, 可以很好的反映出四种煤炭之间的相似性, 并将四种样品聚在一起; 同时分析结果也表现出 ZBM110B 和 GSR-4 与煤炭之间的差异性, 并分离出二者, 达到了煤岩分类的效果。

不同种类的煤岩样品在太赫兹波段下的响应存在差异。采用 CA 和 PCA 方法对煤岩的相似性进行分类, 结果表明欧氏距离和 PC1 得分与不同样品的某些性质有较好的匹配性。同时由于煤岩的折射率和吸收系数光谱存在差异, 所以六种样品的折射率和吸收系数经过 CA-PCA 模型处理后的分类结果不同。但该统计方法也很好的提高了鉴别效果, 为鉴别不同种类的煤岩提供了更为直接有效的方法。基于上述结果, 可以将不同地区的煤炭和岩石通过太赫兹时域光谱技

术进行测量, 然后进行 CA-PCA 聚类, 构建丰富的数据库作为其地区的煤岩鉴别标准。因此, 太赫兹时域光谱技术是一种快速鉴别煤岩的方法, 可用于煤岩识别及煤炭种类的分类。

3 结 论

采用太赫兹技术加上多元统计方法对煤岩进行鉴别是可行的。在 0.5~2.5 THz 的太赫兹范围内, 各类样品无论是折射率还是吸收系数都存在一些差异, 然后经过 CA 和 PCA 方法处理后这些差异会得到一定的放大, 这样更有助于实现对不同种类煤岩样品的识别。结果表明 CA 树状图中的欧氏距离与 PCA 直方图中的 PC1 得分相对应, 都反映了不同样品之间的相似性与差异性, 并同时验证了样品中碳含量是影响煤炭样品折射率的因素之一, 灰分的含量是影响煤炭样品吸收系数的因素之一。但由于我国煤岩种类丰富, 因此还需要进一步研究更多种类的煤岩样本, 建立一个庞大的基于太赫兹光谱下的煤岩数据库, 使其成为快速检测煤岩类别的重要依据。

References

- [1] LI Hui, QI Li-jun, ZHANG Jian-hua(李 慧, 祁力钧, 张建华). Transactions of the Chinese Society for Agricultural Machinery(农业机械学报), 2012, 43(9): 184.
- [2] YANG Shuai, ZUO Jian, ZHANG Cun-lin(杨 帅, 左 剑, 张存林). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2016, 36(12): 3870.
- [3] CHEN Yang, TAN Zuo-jun, XIE Jing, et al(陈 阳, 谭佐军, 谢 静, 等). Science and Technology of Food Industry(食品工业科技), 2014, 35(14): 49.
- [4] YAO Jian-quan(姚建铨). Journal of Chongqing University of Posts and Telecommunications • Natural Science Edition(重庆邮电大学学报 • 自然科学版), 2010, 22(6): 703.
- [5] XU Chang-hong, TENG Xue-ming, ZHAO Hui, et al(许长虹, 滕学明, 赵 卉, 等). Modern Scientific Instruments(现代科学仪器), 2013, (4): 228.
- [6] TENG Xue-ming, ZHAO Kun, ZHAO Hui, et al(滕学明, 赵 昆, 赵 卉, 等). Modern Scientific Instruments(现代科学仪器), 2011(6): 19.
- [7] WANG Xin, ZHAO Duan, HU Ke-xiang, et al(王 昕, 赵 端, 胡克想, 等). Journal of China Coal Society(煤炭学报), 2018, 43(4): 1146.
- [8] LI Peng-peng, ZHANG Yuan, GE Hong-yi, et al(李鹏鹏, 张 元, 葛宏义, 等). Science and Technology of Food Industry(食品工业科技), 2017, 38(3): 372.
- [9] YU Xian-shu, GAO Lei, LU Gui-wu(于宪书, 高 磊, 卢贵武). Lubrication Engineering(润滑与密封), 2016, 41(12): 26.
- [10] WANG Xin, ZHAO Duan, DING En-jie(王 昕, 赵 端, 丁恩杰). Coal Mining Technology(煤矿开采), 2018, 23(1): 13.
- [11] ZHAN Hong-lei, WANG Yu-xia, WANG Xue-song, et al(詹洪磊, 王玉霞, 王雪松, 等). Journal of Terahertz Science and Electronic Information Technology(太赫兹科学与电子信息学报), 2016, 14(1): 26.
- [12] Zhan H, Xi J, Zhao K, et al. Food Control, 2016, 67: 114.
- [13] LIU Ling-yu, YANG Chuan-fa, ZHANG Xian-sheng, et al(刘陵玉, 杨传法, 张献生, 等). Journal of China Coal Society(煤炭学报), 2016, 41(2): 497.

Study on Coal-Rock Identification Method Based on Terahertz Time-Domain Spectroscopy

MIAO Shu-guang¹, SHAO Dan^{1*}, LIU Zhong-yu^{2,3}, FAN Qiang¹, LI Su-wen¹, DING En-jie^{2,3}

1. School of Physics and Electronic Information, Huaibei Normal University, Huaibei 235000, China

2. School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, China

3. Internet of Things Perception Mine Research Center, China University of Mining and Technology, Xuzhou 221116, China

Abstract Coal-rock identification is one of the key problems restricting unmanned coal mining. Because of the extremely complicated working environment, the traditional manual coal mining is difficult to find the interface of coal and rock accurately, which is easy to cause the phenomenon of undercutting or overcutting. As a non-destructive detection method, Terahertz spectroscopy can reflect the physical and chemical information of the object under test and be an effective method to study the identification of coal and rock. In this paper, the terahertz time-domain spectroscopy and multivariate statistical method-cluster analysis (CA) and principal component analysis (PCA) are used to identify different types of coal and rock. The THz spectra of six coal and rock samples are obtained by transmission terahertz spectrometer. FFT and other mathematical calculations can obtain various samples' refractive index, absorption coefficient and dielectric constant. The results show differences in the refractive index and absorption coefficient of different types of coal and rock. By analyzing the relationship between the refractive index and absorption coefficient of various coal samples and the content of each component of the samples, it can be found that carbon content is one of the factors affecting the refractive index of the samples, and ash content is one of the factors affecting the absorption coefficient of the samples. The Euclidean distance of two kinds of samples in cluster analysis and the score of PC1 in principal component analysis can reflect the similarity and dissimilarity between coal and rock samples, and the results of CA and PCA are consistent. The refractive index and absorption coefficient of various samples in the 0.5~2.5 THz frequency range are combined with CA and PCA to form a model between terahertz data and coal and rock. According to the analysis, the six types of coal samples in the two models can be divided into two types based on the similarity between different samples. In the CA-PCA model with the absorption coefficient of various samples adopted, four kinds of coal are clustered together. Moreover, quartz sandstone (GSR-4) has a unique characteristic: quartz sandstone has the smallest PC1 score value, and the Euclidean distance between quartz sandstone and the second type is the largest, up to 219.03. It can be seen that the combination of terahertz technology and multivariate statistical method can realize the accurate identification of coal and rock, and the recognition accuracy can reach 100%.

Keywords Terahertz time-domain spectroscopy; Coal-rock identification; Principal component analysis; Cluster analysis

(Received Jun. 5, 2021; accepted Aug. 2, 2021)

* Corresponding author