

基于激光诱导荧光技术的煤矿水源识别研究

闫鹏程^{1,2}, 尚松行², 周孟然², 胡 锋², 刘 瑜¹

1. 安徽理工大学, 深部煤矿采动响应与灾害防控国家重点实验室, 安徽 淮南 232001
2. 安徽理工大学电气与信息工程学院, 安徽 淮南 232001

摘 要 快速准确的识别煤矿含水层水源对于煤矿突水预警及灾后救援意义重大, 针对传统水源识别耗时较长, 不适宜构建在线式预警系统, 提出使用激光诱导荧光技术用于煤矿水源类型识别的方法。利用激光激发待测水样, 获取其荧光光谱, 结合模式识别对水源进行快速辨识。实验采集了淮南矿区谢桥煤矿的两种纯水样本-老空水与砂岩水, 并根据不同混合比配成 5 种混合水样进行实验。首先针对获取的水源荧光光谱中可能会存在的各种噪声及干扰信息, 采用 SG、Normalize、Gapsegment 求导、Detrend 和 MSC 5 种常用的光谱预处理算法对光谱数据进行处理。其次针对荧光光谱数据量过大, 对数据进行 PCA 降维, 作为对比 6 种预处理方式(含原始光谱)主成分数皆取 3, 结果显示 SG 预处理累计贡献度最大, 为 97.26%; 其次是原始光谱, 为 92.38%, Normalize 与 Detrend 累计贡献度相差不大, 分别为 88.04% 和 87.59%, MSC 为 66.41%, Gapsegment 最差, 为 22.65%。最后分别对 PCA 降维后的数据使用线性 LDA 以及非线性 RBF-SVM 模型进行识别对比。使用 LDA 进行建模, SG-PCA-LDA 正确率最高, 达到了 98.86%, 依据建立的 LDA 模型, 对验证集数据进行识别, SG-PCA-LDA 的正确率依然最高, 为 100%。使用 RBF-SVM 进行建模, Original-PCA-RBF-SVM, SG-PCA-RBF-SVM, Normalize-PCA-RBF-SVM 正确率最高, 皆为 97.14%, 依据建立的 RBF-SVM 模型, 对验证集数据进行识别, Original-PCA-RBF-SVM 和 SG-PCA-RBF-SVM 正确率依然最高, 为 97.14%。对比两类模型可以发现, LDA 验证集正确率较建模集有一定的提升, 而 RBF-SVM 验证集正确率较建模集有小幅度降低, 说明 LDA 模型对于此煤矿水源荧光光谱数据的泛化能力较好, 且成功率较高。结果表明, SG-PCA-LDA 模型结合激光诱导荧光技术是一种较佳的应用于本地煤矿水源识别的方法, 且验证了对老空水、砂岩水的纯水样和混合水样识别的可能性, 可以推广到煤矿其他混合水源的识别中。

关键词 激光诱导荧光技术; 水源识别; 煤矿含水层; LDA; RBF-SVM

中图分类号: O657.3 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2020)07-2176-06

引 言

矿井水害是煤矿生产的五大灾害之一, 对煤矿安全威胁巨大^[1-3]。近年来随着国家对煤矿水害的重视, 水害事故无论是从发生次数上, 还是伤亡人数上都有大幅降低。然而随着大多数矿井的采掘深度逐渐加深, 水文地质环境随之愈加复杂, 开采规模与强度也逐渐加大(年产已达 1 500 万 t), 尤其是特厚煤层超大采高(15~20 m)等综放工作面装备的使用, 更是加剧了井下水害的发生几率^[4]。因此如何快速准确的识别煤矿水源, 无论是对于煤矿水灾预警, 还是对于灾后救援都具有重大意义^[5-7]。

针对煤矿水源识别的研究, 国内外专家采用了多种方法, 如 QLT 法^[8]、同位素法^[9-10]、代表离子法^[11]等, 研究较多的又以代表离子法为主。如张淑莹等采用基于独立性权-灰色关联度理论的突水水源判别^[12], 毛志勇等采用基于 KP-CA-MPSO-ELM 的矿井突水水源判别模型^[13], 刘国伟等采用多元统计分析对滨海矿区进行水源识别^[14], 都取得了较好的效果。但是这些代表离子检测的过程过于漫长, 一般实验室检测需要 1 h, 识别精度相对较低, 且识别一般以单一水源为主, 对混合水样识别没有涉及。

针对上述问题, 本课题组提出使用激光诱导荧光光谱技术进行煤矿水源类型识别, 以 405 nm 激光激发待测水样, 获取荧光光谱, 通过 SG、Normalize 等多种预处理方法进行

收稿日期: 2019-11-30, 修订日期: 2020-02-20

基金项目: 国家重点研发计划重点专项(2018YFC0604503), 安徽省自然科学基金青年项目(1808085QE157), 安徽省博士后科研经费资助项目(2019B350), 中国煤炭工业协会科学技术研究指导性计划项目(MTKJ2018-258)资助

作者简介: 闫鹏程, 1988 年生, 安徽理工大学电气与信息工程学院讲师 e-mail: pcyang1988@126.com

处理,并对数据进行 PCA 降维,以简化模型,缩短识别时间,对降维后的数据分别进行线性 LDA 以及非线性 RBF-SVM 模型进行识别对比,为煤矿突水在线预警提供理论依据。

1 实验部分

1.1 材料

水源样本采集地为淮南矿区谢桥煤矿,以全国煤矿水害事故多发水源-老空水,谢桥煤矿常见水源-砂岩水,以及两者按比例混合的 5 种水样(分别为老空水与砂岩水混合比例 10:7, 10:4, 10:10, 7:10, 4:10),共 7 种水样为待测样本,按混合比例以此记为 a, b, c, d, e, f, g。每种水样样本皆采集 30 个,共 210 个,存储于遮光玻璃瓶中。

1.2 仪器及数据采集

激光诱导荧光光谱系统,包括(1)405 nm 单模激光器(北京华源拓达),功率设置 100 mW;(2)浸入式荧光探头(广州标旗光电 FPB-405-V3);(3)微型光纤光谱仪(USB2000+,美国海洋公司),积分时间 1 s,光谱范围 340~1 020 nm,光谱分辨率 1 nm。为避免室内照明对实验的影响,样本放置于暗室中进行激光激发。

数据采集由 SpecSuite 软件进行,建模及数据处理环节在 Matlab2018a 环境下进行。

1.3 数据处理方法

鉴于检测水样样本时可能会出现的高频、背景噪声等,对获取的荧光光谱数据进行数据预处理。所使用的方法包含 SG、Normalize、Gapsegment 求导、Detrend、MSC,通过后续对比,以获取最佳预处理方法。

水样光谱冗余度及维度较高,尤其是 5 种混合水样谱线混杂,因此需要采用降维算法进行数据特征提取,以除去噪声与冗余特征,达到后续识别的快速性。主成分分析算法

(PCA)在数据压缩,消除冗余等领域有着广泛的使用。PCA 的主体内容是把一个 n 维的矩阵通过正交变换投影到 k 维上,投影后的变量即为主成分。

识别模型采用不同方法,即针对线性数据的线性判别式分析(LDA),以及针对非线性数据的径向基核函数支持向量机(RBF-SVM)分别进行识别,以对比获取最佳识别模型。

2 结果与讨论

2.1 样本选择

随机使用每种样本的 5/6(25 个)作为建模集,剩余 1/6(5 个)作为验证集,即建模集样本 175 个,验证集样本 35 个。

2.2 样本预处理

使用 SG、Normalize、Gapsegment 求导、Detrend、MSC 对原始光谱进行处理,加上原始光谱(Original),共获得 6 组光谱数据,如图 1 所示。由图可以看出,水样光谱数据在 340~420 及 700~1 020 nm 波段趋于一致,主要区别处在 420~700 nm 波段,且波峰点皆集中在此波段内。随着老空水比例的增加,荧光光谱强度有所增强,这是因为灰岩水所处地层较深,导致有机物含量较少,其内部成分主要以无机离子为主,而老空水属于人为活动区域,其水中所含成分较为复杂,因此荧光光谱特征较为明显。不同水样的荧光光谱差异明显,其根本原因是水中所含物质成分、浓度的不同导致的,这也是进行煤矿水源类型识别研究的理论基础。

对比预处理前后光谱图可以发现,使用 SG 预处理后的光谱组间间距有所增加,Normalize 和 Detrend 预处理后的光谱数据较原始数据组间间距下降,Gapsegment 和 MSC 预处理后的组间间距最不明显,其中又以 Gapsegment 处理后的效果最差,这是由于在求导的过程中对部分噪声进行了放大导致的。

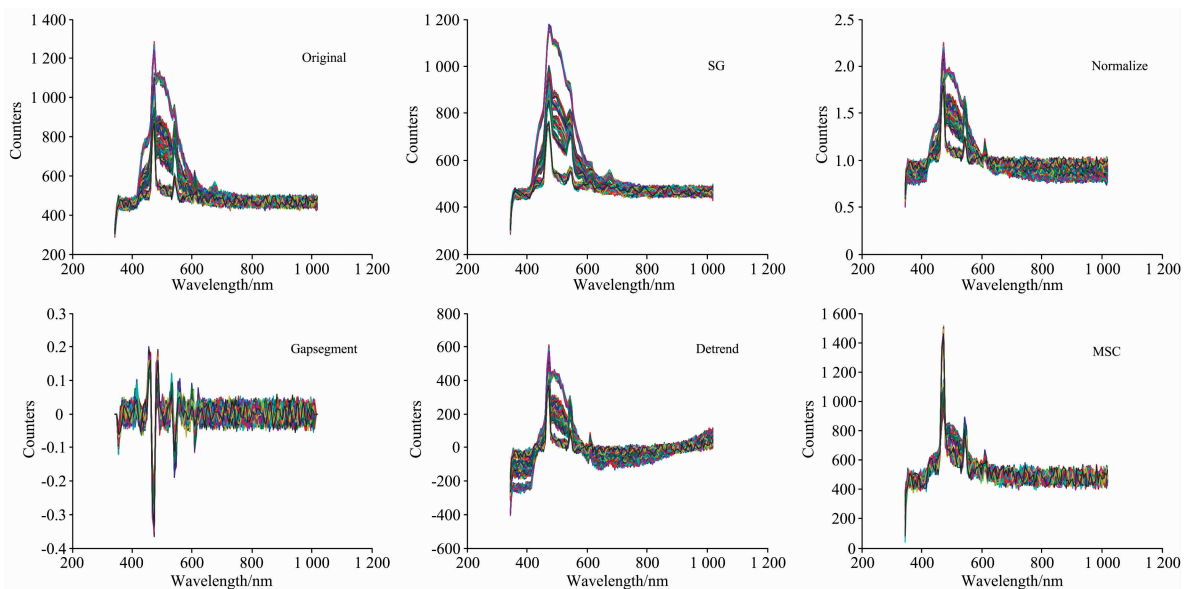


图 1 原始光谱曲线及预处理后的光谱曲线

Fig. 1 Original spectrum and spectrum of different pretreatments

2.3 PCA 降维

对所有水样样本进行 PCA 降维, 作为对比 6 种预处理方式主成分数皆取 3, 其累计贡献度如图 2 所示, 由图可以看出 SG 处理后的数据在主成分为 3 的时候累计贡献度最大, 为 97.26%; 其次是原始光谱, 为 92.38%, Normalize 与 Detrend 累计贡献度相差不大, 分别为 88.04% 和 87.59%; MSC 为 66.41%, Gapsegment 最差, 为 22.65%。由图 3 可以看出 7 种水样在主成分数为 3 时候的得分三维立体分布情况, 在图 3(a), (b), (c) 和 (e) 中, 7 种水样聚类明显, 而图 3(d) 和 (f) 中, 聚类效果较差。因此选择主成分数为 3 进行后续的建模识别。

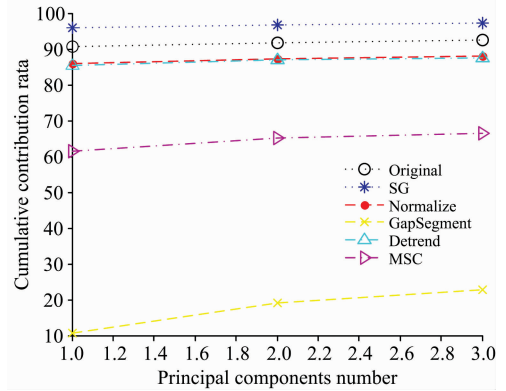


图 2 累计贡献度

Fig. 2 The cumulative contribution

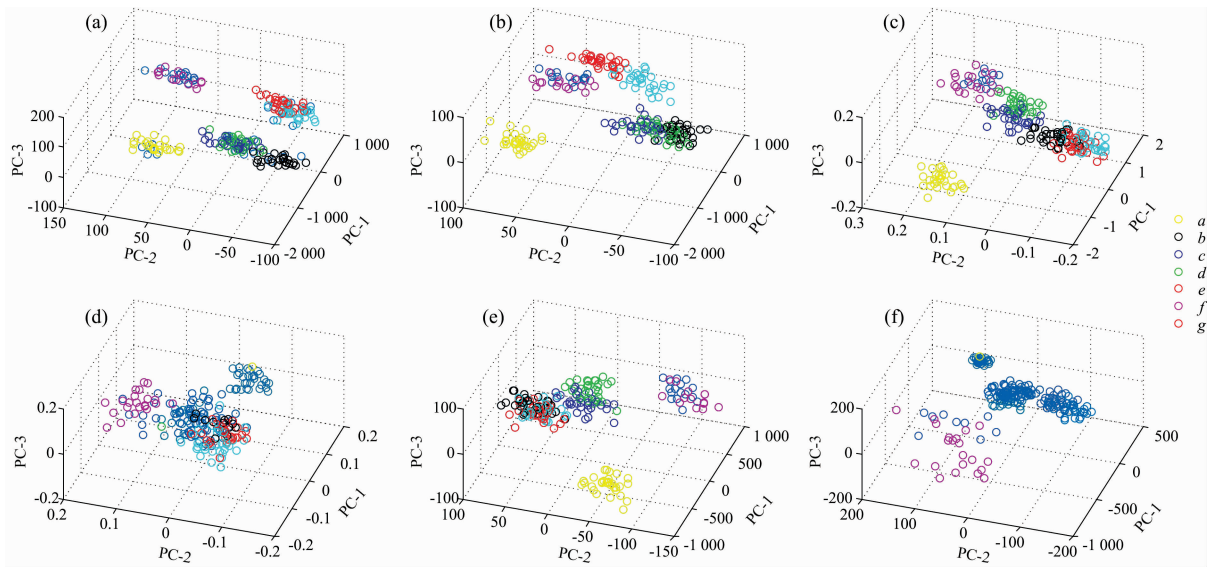


图 3 不同预处理方法得分分布图

Fig. 3 Scores distribution of different pretreatments

(a): Original; (b): SG; (c): Normalize; (d): Gapsegment; (e): Detrend; (f): MSC

2.4 LDA 对煤矿水源的识别模型

对原始光谱及预处理后的 6 种光谱进行 PCA 降维得到的主成分 LDA 建模进行对比, 结果如表 1 所示。总体来看 SG-PCA-LDA 的建模集效果最好, 总体正确率为 98.86% (173/175); 其次为 Original-PCA-LDA, 总体正确率 98.29% (172/175); Detrend-PCA-LDA 总体正确率 97.71% (171/

175); Normalize-PCA-LDA 总体正确率 97.71% (171/175); MSC-PCA-LDA 总体正确率 93.71% (164/175); Gapsegment-PCA-LDA 总体正确率 73.71% (129/175)。出错部分集中于 c 和 d 两种水的互相误判, 而对于 a 和 g 两种纯水样水的识别没有出现错误。

表 1 不同预处理方法 LDA 建模集识别结果混淆矩阵

Table 1 Confusion matrix of modeling set with LDA recognition by different pretreatments

	Original							SG							
	a	b	c	d	e	f	g	a	b	c	d	e	f	g	
a	25	0	0	0	0	0	0	a	25	0	0	0	0	0	0
b	0	25	0	0	0	0	0	b	0	25	0	0	0	0	0
c	0	0	23	1	0	0	0	c	0	0	23	0	0	0	0
d	0	0	2	24	0	0	0	d	0	0	2	25	0	0	0
e	0	0	0	0	25	0	0	e	0	0	0	0	25	0	0
f	0	0	0	0	0	25	0	f	0	0	0	0	0	25	0
g	0	0	0	0	0	0	25	g	0	0	0	0	0	0	25

续表 1

	Normalize							Gapsegment						
	a	b	c	d	e	f	g	a	b	c	d	e	f	g
a	25	0	0	0	0	0	0	a	25	0	0	0	0	0
b	0	25	0	0	0	0	0	b	0	17	1	3	5	2
c	0	0	22	1	0	0	0	c	0	1	15	7	0	0
d	0	0	3	24	0	0	0	d	0	5	8	15	0	0
e	0	0	0	0	25	0	0	e	0	2	0	0	10	1
f	0	0	0	0	0	25	0	f	0	0	0	0	10	22
g	0	0	0	0	0	0	25	g	0	0	1	0	0	0

	Detrend							MSC						
	a	b	c	d	e	f	g	a	b	c	d	e	f	g
a	25	0	0	0	0	0	0	a	25	0	0	0	0	0
b	0	25	0	0	0	0	0	b	0	25	2	0	0	0
c	0	0	22	0	0	0	0	c	0	0	18	4	0	0
d	0	0	3	25	0	0	0	d	0	0	5	21	0	0
e	0	0	1	0	25	0	0	e	0	0	0	0	25	0
f	0	0	0	0	0	25	0	f	0	0	0	0	0	25
g	0	0	0	0	0	0	25	g	0	0	0	0	0	0

验证集识别结果如表 2 所示, 从表 2 可以看出除 Gapsegment-PCA-LDA 识别模型外, 其他识别模型整体表现良好, 识别正确率皆高于 90%, 最佳的仍是 SG-PCA-LDA 模型, 识别率 100%。出错部分仍以 c 和 d 两种水的互相误判为主, 对于 a 和 g 两种纯水水样的识别仍然没有出现错误。

表 2 LDA 验证集识别结果

Table 2 Recognition of verification set with LDA

	预处理方法	正确率
	验证集	Original
SG		100%(35/35)
Normalize		97.14%(34/35)
Gapsegment		74.29%(26/35)
Detrend		94.29%(33/35)
MSC		91.43%(32/35)

由上述数据可以看出, 煤矿水源的荧光光谱在经过预处理、PCA 降维后, 可以使用 LDA 进行识别, 且 SG-PCA-LDA 效果较佳, 作为对比后续将进行煤矿水源荧光光谱的非线性 RBF-SVM 模型识别。

2.5 RBF-SVM 对煤矿水源的识别模型

对原始光谱及预处理后的 6 种光谱进行 PCA 降维得到的主成分 RBF-SVM 建模进行对比, 结果如表 3 所示。出错部分集中于 c 和 d 两种水的互相误判, 而对于 a 和 g 两种纯水水样的识别没有出现错误。

验证集识别结果如图 4 所示。图 4(a)为 Original-PCA-RBF-SVM 识别模型, 有 1 个 b 类水被判定为 c 类, 正确率 97.14%(34/35); 图 4(b)为 SG-PCA-RBF-SVM 识别模型, 有 1 个 b 类水被判定为 c 类, 正确率 97.14%(34/35); 图 4(c)为 Normalize-PCA-RBF-SVM 识别模型, 有 1 个 b 类水被判定为 c 类, 1 个 e 类水被判定为 f 类, 3 个 f 类被判定为 e 类, 正确率 85.71%(30/35); 图 4(d)为 Gapsegment-PCA-

RBF-SVM 识别模型, 有 1 个 b 类水被判定为 c 类, 1 个 b 类水被判定为 e 类, 1 个 d 类水被判定为 g 类, 2 个 e 类被判定为 f 类, 2 个 f 类被判定为 e 类, 正确率 74.29%(26/35); 图 4(e)为 Detrend-PCA-RBF-SVM 识别模型, 有 1 个 b 类水被判定为 c 类, 1 个 e 类水被判定为 f 类, 2 个 f 类被判定为 e 类, 正确率 88.57%(31/35); 图 4(f)为 MSC-PCA-RBF-SVM 识别模型, 有 2 个 c 类水被判定为 b 类, 2 个 d 类水被判定为 c 类, 2 个 f 类被判定为 e 类, 正确率 82.86%(29/35)。出错部分相对 LDA 模型在种类上有所增加, 但是对于 a 和 g 两种纯水水样的识别仍然没有出现错误。

表 3 RBF-SVM 建模集结果

Table 3 Recognition of modeling set with RBF-SVM

	预处理方法	正确率
	建模集	Original
SG		97.17%(34/35)
Normalize		97.14%(34/35)
Gapsegment		68.57%(24/35)
Detrend		94.29%(33/35)
MSC		91.43%(32/35)

表 3 可以看出, 煤矿水源的荧光光谱在经过预处理、PCA 降维后, 可以使用 RBF-SVM 进行识别, 且 SG-PCA-RBF-SVM 效果较佳。

3 结 论

实验对淮南矿区谢桥煤矿的老空水、砂岩水以及 5 种混合水样的激光诱导荧光光谱进行了识别分析, 选取了不同预处理方法及不同识别模型, 得到结论如下:

(1) 煤矿水源在物质成分及浓度上的差异反应在激光诱导荧光光谱上也会出现较大差异, 因此可以通过激光诱导荧

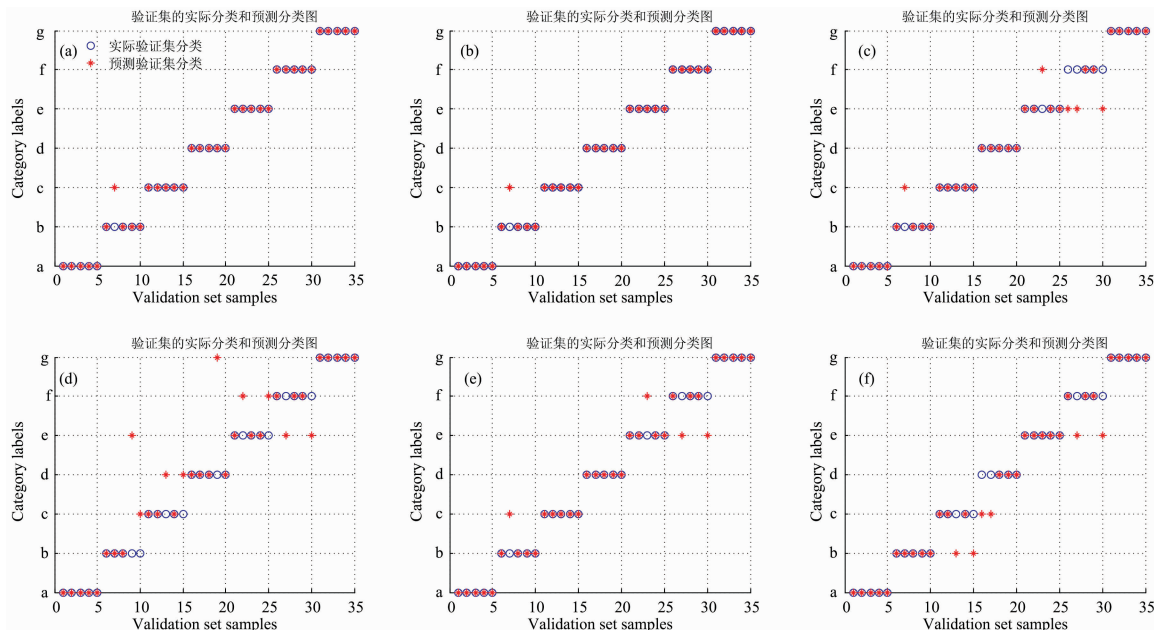


图 4 RBF-SVM 的验证集识别结果

Fig. 4 Recognition of verification set with RBF-SVM

(a): Original; (b): SG; (c): Normalize; (d): Gapsegment; (e): Detrend; (f): MSC

光技术进行水源的有效识别。

(2)SG 预处理是被选择的几种预处理方式中最适宜本地水样预处理的方法, 结合 PCA 降维得到的 3 个主成分, 并分别利用 LDA 与 RBF-SVM 进行识别, 验证集分别到达了 100% 与 97.14% 的正确率; 对比同种预处理方式不同识别模型的结果可以发现, 使用 LDA 进行建模的识别结果平均正确率高于使用 RBF-SVM 进行建模。且对比 RBF-SVM 的识别数据可以发现, 验证集较建模集正确率有所下降, 说明该算法在此应用中的泛化能力较差, 因此 SG-PCA-LDA 模型

结合激光诱导荧光技术是一种较佳的应用于本地煤矿水源识别的方法, 且因为使用了数据降维, 可以大幅度降低后期识别模型的运算复杂度, 节省时间, 对于煤矿水源的快速识别意义重大。

(3)实验验证了 SG-PCA-LDA 模型结合激光诱导荧光技术用于煤矿水源识别的可行性, 对老空水、砂岩水的纯水样和混合水样进行了识别, 可以推广到煤矿其他混合水源的识别中。

References

[1] WU Qiang, XU Hua, ZHAO Ying-wang, et al(武 强, 徐 华, 赵颖旺, 等). Journal of China Coal Society(煤炭学报), 2018, 43(10): 2661.

[2] Hu Feng, Zhou Mengran, Yan Pengcheng, et al. IEEE Access, 2019, 7: 107129.

[3] WU Jin-gang, MAO Jun-rui, CHAI Pei(吴金刚, 毛俊睿, 柴 沛). Safety in Coal Mines(煤矿安全), 2019, 50(10): 239.

[4] LIU Shou-qiang, WU Qiang, ZENG Yi-fan(刘守强, 武 强, 曾一凡). Coal Engineering(煤炭工程), 2019, 51(3): 1.

[5] YAN Peng-cheng, ZHOU Meng-ran, LIU Qi-meng, et al(闫鹏程, 周孟然, 刘启蒙, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2016, 36(1): 243.

[6] SUN Ji-ping, JIN Chun-hai(孙继平, 靳春海). Industry and Mine Automation(工矿自动化), 2019, 45(4): 1.

[7] WANG Tian-tian, JIN De-wu, LIU Ji, et al(王甜甜, 靳德武, 刘 基, 等). Journal of China Coal Society(煤炭学报), 2019, 44(9): 2840.

[8] WANG Yang, ZUO Wen-zhe, WANG Bin-hai, et al(汪 洋, 左文喆, 王斌海, 等). Modern Mining(现代矿业), 2018, 34(1): 69.

[9] TIAN Xiu-rong, WEI Fang, WEI Tian(田秀荣, 魏 芳, 魏 甜). Coal Geology of China(中国煤炭地质), 2015, 27(12): 53.

[10] Guan Zilong, Jia Zhifeng, Zhao Zhiqiang, et al. Journal of Earth System Science, 2019, 128(7): 200.

[11] PENG Cheng, LIU Yong-tao, YOU Wen-qiang, et al(彭 程, 刘永涛, 尤文强, 等). China Mining Magazine(中国矿业), 2019, 28(7): 176.

[12] ZHANG Shu-ying, HU You-biao, XING Shi-ping(张淑莹, 胡友彪, 邢世平). Hydrogeology & Engineering Geology(水文地质工程地质), 2018, 45(6): 36.

[13] MAO Zhi-yong, HUANG Chun-juan, LU Shi-chang, et al(毛志勇, 黄春娟, 路世昌, 等). China Safety Science Journal(中国安全科学学

报), 2018, 28(8): 111.

- [14] LIU Guo-wei, MA Feng-shan, GUO Jie, et al(刘国伟, 马凤山, 郭捷, 等). Gold Science and Technology(黄金科学技术), 2019, 27(2): 207.

Research on Identification of Coal Mine Water Source Based on Laser Induced Fluorescence Technology

YAN Peng-cheng^{1, 2}, SHANG Song-hang², ZHOU Meng-ran², HU Feng², LIU Yu¹

1. State Key Laboratory of Mining Response and Disaster Prevention and Control in Deep Coal Mine, Anhui University of Science and Technology, Huainan 232001, China
2. College of Electrical and Information Engineering, Anhui University of Science and Technology, Huainan 232001, China

Abstract The rapid and accurate identification of coal mine aquifer water source is of great significance for coal mine water inrush warning and post-disaster rescue. It takes a long time for water source identification with the traditional method, and it is not suitable to construct an online early warning system. A method of using laser induced fluorescence technology to identify the type of coal mine water source is proposed. The laser is used to excite the water sample. Then the fluorescence spectrum is obtained, with pattern recognition the water source can be rapidly identified. Two kinds of water samples-goaf water and sandstone water of Xieqiao Coal Mine in Huainan Mining Area were collected, and five mixed water samples were prepared according to different mixing ratios. Firstly, according to the various noise and interference information that may exist in the obtained water source fluorescence spectrum, the spectral data were pretreated by SG, Normalize, Gapsegment derivation, Detrend and MSC. Secondly, PCA was used to reduce the dimension of fluorescence spectral data due to a large amount of data. As a comparison of the six pretreatment methods (including the original spectrum), the number of principal components was taken by 3, and the results showed that the cumulative contribution of SG pretreatment is the largest, which was 97.26%. The second was the original spectrum, which was 92.38%. The cumulative contribution of Normalize and Detrend were not much different, which were 88.04% and 87.59%, MSC was 66.41%, and Gapsegment was the worst with 22.65%. Finally, the linear model of LDA and nonlinear model of RBF-SVM were used to identified and compared with the data of reduced dimension by PCA. Using LDA for modeling, SG-PCA-LDA had the highest accuracy rate, which reached 98.86%. According to the LDA model established, the verification set data were identified, and the accuracy rate of SG-PCA-LDA was still the highest with 100%. Using RBF-SVM for modeling, Original-PCA-RBF-SVM, SG-PCA-RBF-SVM, and Normalize-PCA-RBF-SVM had the highest accuracy rate, both of which was 97.14%. Based on the RBF-SVM model established, verification set data were identified, and the accuracy rate of Original-PCA-RBF-SVM and SG-PCA-RBF-SVM was still the highest, which is 97.14%. It can be found that the accuracy rate of the LDA verification set was improved which compared with the modeling set, and the accuracy rate of the RBF-SVM verification set was slightly lower than the modeling set, which showed that LDA model had better generalization ability and higher accuracy rate for fluorescence spectral data of this coal mine water. The results showed that the SG-PCA-LDA model combined with laser induced fluorescence technology is a better method for local coal mine water source identification, and it verified the possibility of identification for goaf water, sandstone water and mixed water, which can be extended to identify other mixed water sources of coal mines.

Keywords Laser induced fluorescence technology; Water source identification; Coal mine aquifer; LDA; RBF-SVM

(Received Nov. 30, 2019; accepted Feb. 20, 2020)