

## LIF 结合 LSTM 神经网络的矿井水源识别

闫鹏程<sup>1,2</sup>, 张孝飞<sup>2\*</sup>, 尚松行<sup>2</sup>, 张超银<sup>2</sup>

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** 矿井水害对煤矿安全生产存在巨大威胁, 所以快速识别矿井突水水源, 对煤矿水灾预警及灾后救援工作开展都有重大意义。激光诱导荧光(LIF)技术具有快速、高效、灵敏度高等特点, 克服了传统水化学方法识别时间长的缺点。循环神经网络(RNN)在解决长序列训练过程中产生的梯度消失、梯度爆炸等问题上存在明显不足, 而特殊变体 RNN 即长短期记忆(LSTM)神经网络很好地弥补了 RNN 的短板及缺陷。提出了将 LIF 技术与 LSTM 算法相结合, 应用在矿井突水水源快速识别中。实验样本采自淮南矿区, 以砂岩水和老空水为原始样本, 并将砂岩水和老空水按照不同比例混合配置成 5 种混合水样, 共 7 种待测水样进行实验。首先采用最大最小值归一化(MinMaxScaler)、平滑滤波(SG)以及标准正态变量变换(SNV)三种预处理方法对原始光谱数据进行预处理, 减少原始光谱数据存在的噪声和干扰信息。之后为防止数据量过大, 维度过高, 将包括原始光谱数据在内的四组数据再进行 LDA 降维至 3 维。最后分别搭建 LSTM 识别模型, 从测试集预测准确率、训练集准确率变化趋势以及训练集损失函数变化趋势三个方面进行比较, 选择最优模型。其中 SG + LDA + LSTM 和 Original + LDA + LSTM 在测试集预测准确率上都能达到 100%, MinMaxScaler + LDA + LSTM 测试集预测准确率为 98.57%, SNV + LDA + LSTM 准确率最低, 只有 87.14%; 在训练集准确率变化趋势表现上, SG + LDA + LSTM 能够保持良好的学习, 很快达到 100%, Original + LDA + LSTM 和 MinMaxScaler + LDA + LSTM 也能达到 100% 的准确率, 但在前几次训练过程中会有准确率下降的情况出现, SNV + LDA + LSTM 训练集准确率在训练次数内并未达到 100%; SG + LDA + LSTM 损失函数变化趋势也具有很好的收敛性和稳定性, Original + LDA + LSTM, MinMaxScaler + LDA + LSTM 以及 SNV + LDA + LSTM 在损失函数变化趋势上表现并不出色。结果表明, 4 组模型中, SG + LDA + LSTM 模型是最适合应用于矿井突水识别, 该方法补充了矿井突水水源识别工作的内容, 为矿井突水识别提供了新的思路。

**关键词** 水源识别; 激光诱导荧光光谱; 预处理; LDA; LSTM

**中图分类号:** O657.3 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2022)10-3091-06

### 引言

煤矿资源对于国家发展和生产起着重要作用, 是重要的不可再生资源 and 基础能源, 能否合理开发与利用煤矿资源, 直接影响经济社会的安全发展与国民经济可持续发展, 而煤矿的安全生产不仅关乎经济发展, 也关乎到人民的生命财产安全<sup>[1]</sup>。矿井水害严重威胁着煤矿的安全生产, 是煤矿生产过程中存在的主要灾害之一<sup>[2-4]</sup>。虽然近些年来矿井采掘技

术的不断提高, 灾害处理能力的不断加强, 已经大幅度减少伤亡人员数量, 但随着矿井采掘深度加深, 开采规模日益加大, 导致水文地质环境也在随之变得复杂, 在不确定的工作环境中, 矿井水害的发生概率也会随之提高, 威胁着矿井工作者的生命财产安全<sup>[5]</sup>。因此快速识别矿井突水水源类型, 对于水害预防工作和灾后救援工作都有着重大意义<sup>[6-7]</sup>。

LIF 技术能够完成非扰动、实时原位测量, 荧光光谱分析具有灵敏度高, 速度快等优点, 近几年在科研中得到了广泛的应用<sup>[8]</sup>。例如, 陈至坤等<sup>[9]</sup>运用 LIF 的光谱特征提取,

收稿日期: 2021-08-04, 修订日期: 2021-11-05

基金项目: 国家重点研发计划重点专项(2018YFC0604503), 安徽省博士后科研经费项目(2019B350), 安徽省自然科学基金青年项目(1808085QE157), 中国煤炭工业协会 2018 年度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: PCYan1988@126.com

\* 通讯作者 e-mail: 1285104634@qq.com

实现了对油类的检测；张大源等<sup>[10]</sup>运用 LIF 技术，实现了对 HCCI 发动机甲醛的定量测量；朱家健等<sup>[11]</sup>将 LIF 技术运用到燃烧诊断的研究。在矿井突水水源识别研究中，LIF 技术也在近几年得到了应用，但仍然有待改进和完善<sup>[12]</sup>。

循环神经网络(RNN)在解决长序列训练过程中产生的梯度消失、梯度爆炸等问题上存在明显不足，而特殊变体 RNN 即长短期记忆(LSTM)神经网络很好地弥补了 RNN 的短板及缺陷。输入门、输出门和遗忘门结构是 LSTM 相较于 RNN，添加的三个门结构<sup>[13]</sup>。LSTM 的关键是细胞状态，它贯穿整个神经网络，用来筛选并保留信息，光谱数据往往含有比较多的复杂冗余的信息，所以 LSTM 非常适合用来处理光谱数据。

鉴于此，提出运用 LIF 技术得到水样的荧光光谱数据，通过 MinMaxScaler, SNV 和 SG 三种方法对光谱数据进行预处理，再将包括原始光谱数据在内的四组数据，通过 LDA 进行降维处理，最后结合 LSTM 神经网络，搭建四种矿井突水水源识别模型，进行对比，选择最优模型。

## 1 实验部分

### 1.1 样本采取

实验样本采自淮南矿区，以砂岩水和老空水为原始样本，并将砂岩水和老空水按照不同比例(7:10, 4:10, 10:10, 10:7, 10:4)混合配制成 5 种混合水样，共 7 种待测水样进行实验，并且每种水样均采集 30 个，共计 210 个水样待测样本。按照不同混合比例，为样本编号 1, 2, 3, 4, 5, 6, 7，将贴好标签的 210 个水样密封存储在遮光玻璃瓶中。

### 1.2 设备和样本光谱数据采集

实验使用的仪器包括北京华源拓达生产的 405 nm 的单模激光器，广州标旗光电生产的浸入式荧光探头(FPB-405-V3)，美国海洋公司生产的微型光纤光谱仪(USB2000+)。实验参数设置：激光功率为 100 mW，积分时间 1 s，采样间距 340~1 020 nm，光谱分辨率为 1 nm。

### 1.3 LSTM 模型搭建

首先对不同样本的光谱数据进行添加标签处理，按照样本采取时的编号，给对应的样本光谱数据添加相应的标签数值(1, 2, 3, 4, 5, 6, 7)。LSTM 神经网络通过训练集与对应的标签形成的映射关系，进行训练学习。调整合适的训练周期以及一次训练抓取的样本数，使用 Adam 优化器调整合适的学习率并对模型其他相关参数进行优化。用 Python3.9 软件建立模型。

### 1.4 模型评价标准

模型优劣的评判是通过测试集预测(Prediction)结果与真实值(Real)比较、训练样本准确率变化趋势(Accuracy)和迭代损失(Loss)来进行模型的评估和对比。测试集预测值与真实值相符越多，则模型识别效果越好；训练集准确率(Accuracy)是通过测试集经过模型输出的预测值和测试机对应的标签真实值对比，是准确个数与总样本数的比值；迭代损失(Loss)属于多分类交叉熵损失，当迭代损失趋于平稳则反应系统训练良好，系统性能更具有优势。

## 2 结果与讨论

### 2.1 样本预处理

如图 1 所示，除原始光谱数据外，实验采用 3 种方法对原始光谱数据进行预处理，分别为 MinMaxScaler、SG 以及

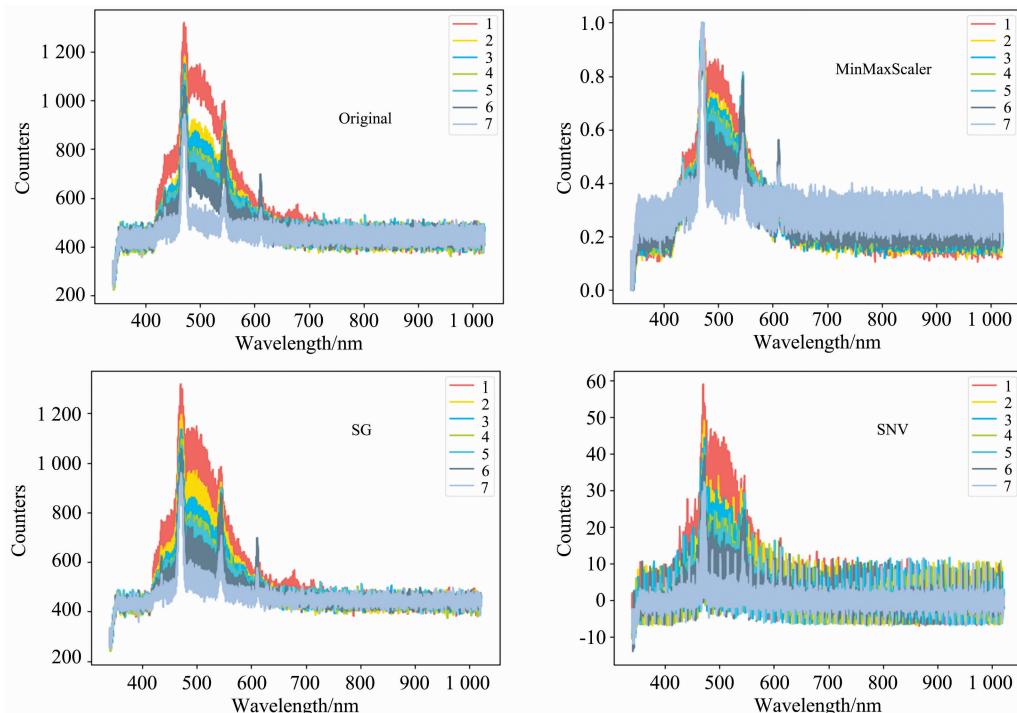


图 1 原始光谱曲线及预处理后的光谱曲线

Fig. 1 Original spectra and spectra pretreated by using different methods

SNV，总共得到 4 组光谱数据。由图 1 可知在 420~650 nm 波段，水样的光谱数据区别比较明显，而在 340~420 以及 650~1 020 nm 波段光谱图像十分相近。

灰岩水所含有有机物比较少，而老空水成分复杂，所以水样荧光光谱图像会随着老空水含量增加，荧光光谱特征逐渐明显。根据水源类的识别研究机理，这是由于不同水源中所含的物质成分以及浓度等特征不同，从而导致在光谱图像中形成了明显的差异。与原始数据比较，SG 和 MinMaxScaler 预处理后的光谱图像都有较好的处理效果，组间间距有所增加或保持良好，SNV 处理后的数据比较冗杂，效果不如 SG 和 MinMaxScaler 理想。

2.2 LDA 降维处理

对原始数据及预处理后的荧光光谱数据进行 LDA 降维处理，降维至 3 维特征数，其降维效果图如图 2 所示。由图 2 可得经 SG 预处理后的 LDA 降维聚类效果最为明显；原始光谱数据及 MinMaxScaler 预处理后的数据进行 LDA 降维后效果相差不大；SNV 预处理后 LDA 降维效果没有其他的效果明显。基于这四种分别搭建 LSTM 识别模型。

2.3 LSTM 识别模型比较

基于四组光谱数据，经 LDA 降维处理后分别搭建 LSTM 水源识别模型。通过对比测试集的预测情况、训练集的准确率变化趋势以及迭代损失函数变化趋势三个方面，多角度考虑，选择最优模型。四种模型的测试集的预测情况如图 3 所示。其中 SNV+LDA 处理后的数据所搭建的 LSTM 模型预测效果最差；MinMaxScaler+LDA 处理后的模型预测准确率比较高；SG+LDA 以及原始光谱数据经 LDA 处理后的两组模型，在测试集的准确率上都有着很好的表现，准确率能够达到 100%。测试集预测准确率如表 1 所示。

表 1 不同模型测试集准确率

Table 1 Accuracies of test set in different models

预处理方法	降维	测试集准确率
Original		100% (70/70)
MinMaxScaler	LDA	98.57% (69/70)
SG		100% (70/70)
SNV		87.14% (61/70)

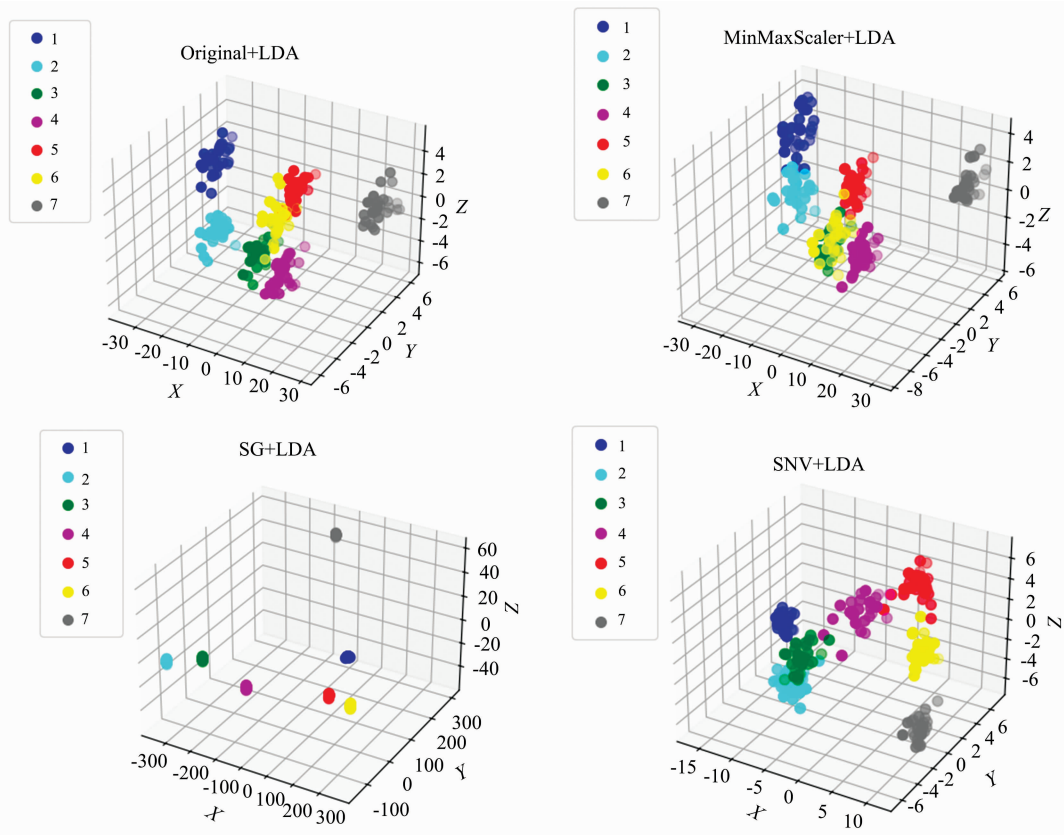


图 2 不同预处理下的 LDA 降维结果图

Fig. 2 Dimension reduction results of LDA under different pretreatments

训练集训练准确率也是很重要的比较依据。如图 4 所示，SNV+LDA 之后的光谱数据在经过 30 次迭代训练之后，训练集训练准确率依旧很难达到 100%，模型训练效果比较差；原始光谱数据经过 LDA 和 MinMaxScaler+LDA 处理后的光谱数据在数次迭代之后训练集训练准确率会有一个

下降过程，说明该模型在前期训练效果比较差，如果训练次数不够，很难达到最佳状态；而 SG+LDA 处理后的光谱数据在训练过程中，训练集训练准确率处于持续上升的趋势，比较少的训练次数就能很快达到 100% 的训练准确率。

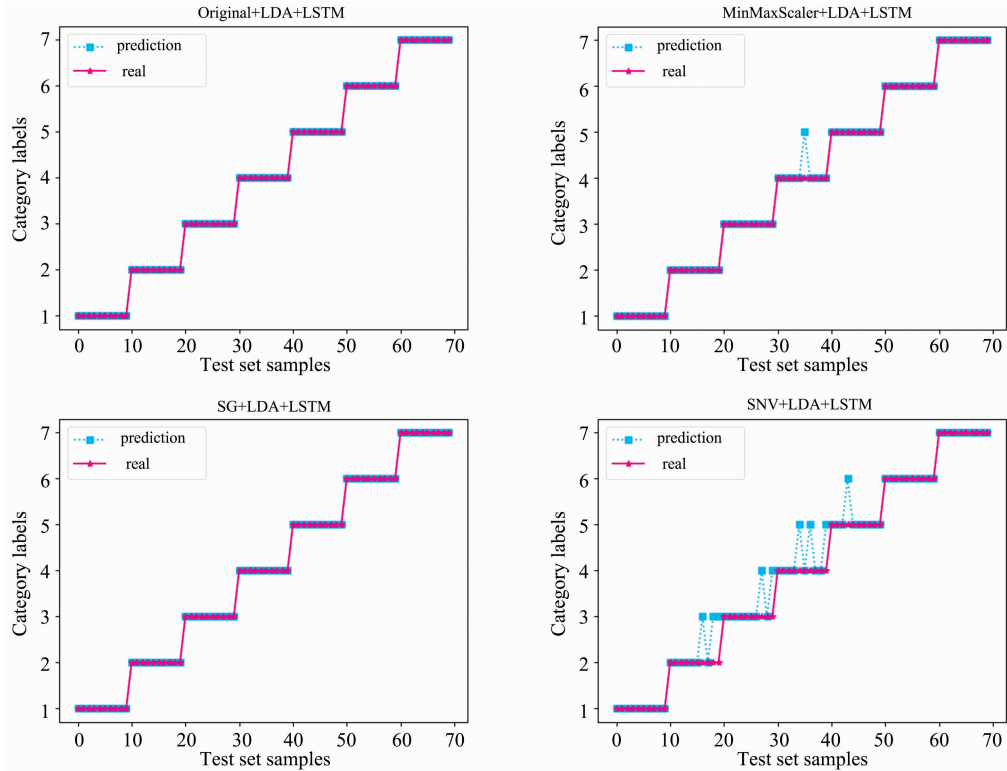


图 3 不同模型下的测试集识别结果

Fig. 3 Recognition results of test sets in different models

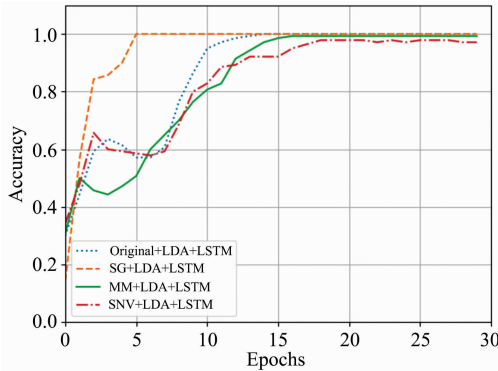


图 4 不同模型的训练过程准确率变化趋势图

Fig. 4 The changing trends of accuracy in training process of different models

损失函数也是评判模型优劣的重要指标之一。如图 5 不同模型训练过程损失函数变化趋势图来看, SG+LDA 处理后的光谱数据建立的模型具有更好的收敛性和稳定性; 原始光谱数据经过 LDA 处理和 MinMaxScaler+LDA 处理后的光谱数据, 两者建立的模型收敛相对较慢, 性能不如前者; SNV+LDA 处理后的数据建立的模型收敛性稳定性最差, 并不具有较好的性能。

综合上述分析, SNV+LDA 处理后的光谱数据建立的 LSTM 识别模型效果最差; 原始光谱数据经过 LDA 降维处理后建立的 LSTM 识别模型和 MinMaxScaler+LDA 处理后的光谱数据建立的 LSTM 识别模型有不错的识别效果和性

能表现; SG+LDA 处理后的光谱数据建立的 LSTM 识别模型在预测准确率、训练过程准确率以及损失函数变化趋势都有更好的表现。

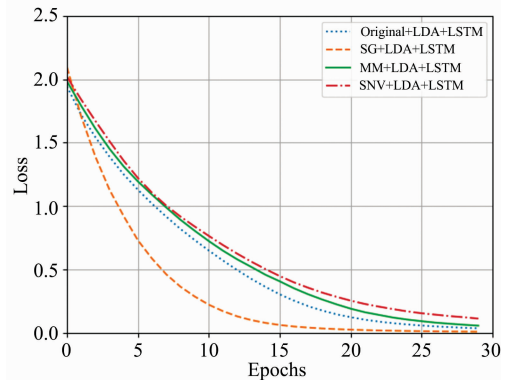


图 5 不同模型的训练过程损失化数变化趋势图

Fig. 5 The changing trends of loss function in training process of different models

### 3 结 论

利用淮南谢桥煤矿的砂岩水和老空水为原始样本, 并将两种水按照不同比例混合成 5 种水样, 共七种水样样本, 对七种水样的激光诱导荧光光谱进行识别分析。经过 MinMaxScaler, SG 以及 SNV 三种方法预处理, 包括原始光谱数据再进行 LDA 降维处理, 搭建四种 LSTM 识别模型。从测试

集预测结果, 训练集训练准确率变化趋势以及训练过程的损失函数变化趋势三个方面进行比较, 选择出最优模型。结果表明, 采用 SG+LDA 处理后的光谱数据搭建的 LSTM 识别模型对水样本识别效果最好, 这是由于三种预处理方法对于光谱数据的作用不同导致的, 其中 MinMaxScaler 主要消

除光谱数据尺度差异过大的影响, SG 用来消除随机噪声并且提高信噪比, 而 SNV 主要是降低分布不均匀带来的影响; 本工作中 SG 预处理方法表现最佳。LSTM 识别模型具有比较好的性能实现对水源的识别, 对人工智能在矿井突水水源识别方面提供了新的探索和改进。

## References

- [ 1 ] WANG Xian-zheng(王显政). Safety in Coal Mines(煤矿安全), 2020, 51(10): 1.
- [ 2 ] YUAN Liang, ZHANG Ping-song(袁 亮, 张平松). Journal of China Coal Society(煤炭学报), 2019, 44(8): 2277.
- [ 3 ] 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.
- [ 4 ] Hu Feng, Zhou Mengran, Yan Pengcheng, et al. IEEE Access, 2019, 7: 107129.
- [ 5 ] 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.
- [ 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 ] LU Xin-pei, WU Fan, LI Jia-yin(卢新培, 吴 帆, 李嘉胤). High Voltage Engineering(高电压技术), 2021, 47(5): 1831.
- [ 9 ] CHEN Zhi-kun, GUO Rui, CHENG Peng-fei(陈至坤, 郭 蕊, 程鹏飞). Laser & Optoelectronics Progress(激光与光电子学进展), 2020, 57(13): 332.
- [10] ZHANG Da-yuan, LI Bo, GAO Qiang, et al(张大源, 李 博, 高 强, 等). Journal of Combustion Science and Technology(燃烧科学与技术), 2019, 25(2): 112.
- [11] ZHU Jia-jian, WAN Ming-gang, WU Ge, et al(朱家健, 万明罡, 吴 戈, 等). Chinese Journal of Lasers(中国激光), 2021, 48(4): 78.
- [12] YAN Peng-cheng, SHANG Song-hang, ZHOU Meng-ran, et al(闫鹏程, 尚松行, 周孟然, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2020, 40(7): 2176.
- [13] LIU Li-jun, LEI Yu, YU Zhen(刘利军, 雷 宇, 余 臻). Aerospace Control(航天控制), 2020, 38(5): 67.

## Research on Mine Water Inrush Identification Based on LIF and LSTM Neural Network

YAN Peng-cheng<sup>1, 2</sup>, ZHANG Xiao-fei<sup>2\*</sup>, SHANG Song-hang<sup>2</sup>, ZHANG Chao-yin<sup>2</sup>

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** Mine water disaster is a great threat to the safety production of a coal mine, so the rapid identification of mine water inrush source is of great significance to the early warning and post-disaster rescue work. Laser-induced fluorescence (LIF) technology has high speed, high efficiency and high sensitivity, which overcomes the shortcomings of long recognition time in traditional hydrochemical methods. Circulating neural network (RNN) has obvious shortcomings in solving the problems of gradient disappearance and gradient explosion in long sequence training, while the special variant RNN, long and short term memory (LSTM) neural network, makes up for the shortcomings of RNN. In this paper, the combination of LIF technology and LSTM algorithm is applied to rapidly identify mine water inrush source. The experimental samples were collected from Huainan Mining Area. The sandstone water and goaf water were taken as the original samples, and the sandstone water and goaf water were mixed into 5 kinds of mixed water samples. According to different proportions, 7 kinds of water samples were experimented. Firstly, MinMaxScaler, SG and SNV were used to preprocess the original spectral data to reduce the noise and interference. After that, to prevent the data from being too large and too high a dimension, the dimension of four groups of data, including the original spectral data, was reduced to 3 dimensions by LDA. Finally, the LSTM recognition models were built respectively, and the optimal model was selected by comparing the prediction accuracy of the test set, the changing trend of the accuracy and the loss function of the training set. Thereinto, SG+LDA+LSTM and Original+LDA+LSTM can reach 100% in the test set prediction accuracy, MinMaxScaler+LDA+LSTM test set prediction accuracy is 98.57%, SNV+LDA+LSTM accuracy is the lowest,

only 87.14%; In terms of the trend of training set accuracy, SG+LDA+LSTM can keep good learning and reach 100% soon. Original+LDA+LSTM and MinMaxScaler+LDA+LSTM can also reach 100% accuracy. However, at the beginning of the training process, the accuracy will decline, and the SNV+LDA+LSTM training set does not reach 100% within the training times; The trend of SG+LDA+LSTM loss function also has good convergence and stability. Original+LDA+LSTM, MinMaxScaler+LDA+LSTM and SNV+LDA+LSTM do not perform well in the trend of loss function. The results show that the SG+LDA+LSTM model is the most suitable for mine water inrush identification among the four models. This method supplements the work of mine water inrush source identification and provides a new idea for mine water inrush identification.

**Keywords** Water source identification; Laser-induced fluorescence spectroscopy; Pretreatment; LDA; LSTM

(Received Aug. 4, 2021; accepted Nov. 5, 2021)

\* Corresponding author

### 撤稿声明

刊登在《光谱学与光谱分析》第 42 卷第 6 期第 1822~1826 页的论文

题目:《高基体样品进样系统-电感耦合等离子体质谱法测定进口铜精矿中稀土元素的含量》

作者:王 斌<sup>1,2</sup>, 郑少锋<sup>2</sup>, 李伟才<sup>2</sup>, 钟康华<sup>2</sup>, 甘久林<sup>1</sup>, 杨中民<sup>1</sup>, 宋武元<sup>3\*</sup>

单位: 1. 发光材料与器件国家重点实验室, 华南理工大学, 广东 广州 510641

2. 国家消费品安全检测重点实验室, 黄埔海关技术中心, 广东 东莞 523070

3. 广州海关技术中心, 广东 广州 510623

doi: 10.3964/j.issn.1000-0593(2022)06-1822-05

经该论文的作者核实后发现文中数据提供有误, 可能引起学术争议, 应此篇论文的全体作者要求, 对此篇论文作撤稿处理。

特此声明!

《光谱学与光谱分析》期刊社

2022 年 9 月 21 日