

RF-CARS 结合 LIF 光谱用于矿井涌水的预测评估

卞凯, 周孟然*, 胡锋, 来文豪, 闫鹏程, 宋红萍, 戴荣英, 胡天羽

安徽理工大学电气与信息工程学院, 安徽淮南 232001

摘要 快速且准确识别矿井涌水水源对于防范煤矿水灾事故有着重大的研究意义。利用激光诱导荧光(LIF)光谱融合智能分类算法进行矿井涌水水源识别打破了传统水化学方法耗时过长等不足, 具有灵敏度高、响应速度快等特点; 然而目前这些已使用的算法仅能依靠分类准确率定性判别不同矿井涌水水样的种类。把随机森林(RF)算法和竞争性自适应重加权(CARS)算法相结合, 基于激光诱导荧光的水样荧光光谱数据建立偏最小二乘回归(PLSR)模型来预测不同矿井涌水的含量, 实现水样的定量评估。首先, 采集300组以老空水为基础混入不同含量砂岩水的矿井涌水样本, 将采集到的水样按4:1比例随机划分成校正集和预测集, 校正集共240组用于建立回归模型, 预测集共60组用于预测不同水样, 搭建激光诱导荧光涌水光谱系统完成光谱数据的获取并生成荧光光谱图。然后分别通过S-G卷积平滑法和Lowess平滑法对原始荧光光谱进行去噪处理, 发现处理后的荧光光谱较原始光谱更为分散, 适合光谱分析, 对比了两种去噪方法的预测精度, 选择Lowess平滑法作为最终去噪方法。接着采用RF算法约简去噪后属性重要度较低的光谱属性, 依据最优回归模型的性能选择约简出的223个属性再用于CARS算法的二次属性精简, 根据CARS算法采样过程中交叉验证均方根误差值最小原则选择出的77个属性光谱数据建立PLSR模型。最后与全光谱、其他变量选择方法、不同回归模型相比, RF-CARS算法属性精简效果最好, 较全光谱建模, 属性由2048个减少到77个, 模型预测集判定系数 R_{pre}^2 由0.9914增长到0.9967, 预测均方根误差RMSEP由0.0294减少到0.0183, 预测精度得到提升, 其余评估指标也相对较好。实验结果表明, RF-CARS结合激光诱导荧光光谱可快速、精准预测矿井涌水, 精简出的光谱属性用来建立回归模型, 为实现矿井涌水含量的实时定量评估提供了一定的理论保障。

关键词 激光诱导荧光光谱; 回归模型; 随机森林; 竞争性自适应重加权; 矿井涌水

中图分类号: O657.3 **文献标识码**: A **DOI**: 10.3964/j.issn.1000-0593(2020)07-2170-06

引言

涌水灾害已成为煤矿五大灾害中影响矿井安全作业的第二大灾害^[1]。矿井涌水不仅需要现场早期预警, 还需要准确判断涌水水源的类型^[2], 这有利于及时掌握涌水水源信息并采取必要的治理手段, 降低灾害所带来的伤亡事故率与重大经济损失。煤矿井下涌水水源识别和研究的方法主要有离子浓度法、同位素分析法、微量元素法等^[3], 这些方法虽然取得了一定的识别效果, 但存在分析时间长、判别准确性不高、操作过程复杂等特点^[4], 难以根据其动态变化迅速提供

可靠的信息, 限制了矿井涌水的实时在线分析, 达不到预警效果。

为了解决现有涌水水源识别方法的不足, 激光诱导荧光(laser-induced fluorescence, LIF)被用于矿井水害研究领域, 并结合机器学习和深度学习实现了矿井涌水水源类型的快速且准确识别, 取得了良好的鉴别效果; 如何晨阳^[5]等采取主成分分析法将小波变换去噪的突水水样荧光光谱数据进行降维处理, 利用K最近邻算法进行水样的识别, 达到了极高的分类准确率。Hu^[6]等利用激光诱导荧光技术结合深度学习方法, 提出了一维卷积神经网络方法用于自动识别九种矿井突水水样, 在不进行复杂的预处理情况下实现了突水水

收稿日期: 2019-07-16, 修订日期: 2019-11-08

基金项目: 国家“十二五”科技支撑计划重点项目(2013BAK06B01), 国家安全生产重大事故防治关键技术科技项目(anhui-0001-2016AQ), 国家重点研发计划项目(2018YFC0604503), 国家自然科学基金项目(51174258), 安徽省自然科学基金青年项目(1808085QE157), 安徽省科技重大专项项目(201903Q07020013)资助

作者简介: 卞凯, 1992年生, 安徽理工大学电气与信息工程学院博士研究生 e-mail: kbian92@163.com

* 通讯联系人 e-mail: mrzhou8521@163.com

样的快速、精准识别。然而目前这些利用激光诱导荧光对矿井涌水进行识别的机器学习和深度学习方法所建立的都只是分类识别模型,只能定性的判别涌水的种类,而不能预测和精准评估不同种类水样的含量以实现定量分析,且各属性因素之间的相关程度未能进行有效分析,部分训练模型和调参过程也较为复杂。

随机森林(random forest, RF)是一种不仅可用于分类结果分析,还能根据属性重要度进行特征选择,降低模型的复杂度的高效算法,如文泽波等^[7]利用 RF 特征选择算法提取出烟雾特征并结合支持向量机实现了视频烟雾的检测。Brokamp^[8]等使用 RF 模型对城市空气中的细颗粒物的浓度进行了预测。Lefkovits^[9]等提出了将 RF 特征选择算法应用于提升脑肿瘤图像分割的判别模型。竞争性自适应重加权算法(competitive adaptive reweighted sampling, CARS)是通过自适应重加权采样技术消去无用信息的一种变量选择方法,能提升运算和建模效率,如 Li^[10]等利用 CARS 结合偏最小二乘线性判别分析用于高果糖玉米糖浆和麦芽糖浆掺假蜂蜜的检测。刘珊珊^[11]等提出 CARS 用于激光诱导击穿光谱来确定猪饲料中铜元素的含量,Wang^[12]等采用 CARS 方法结合近红外光谱技术对大豆秸秆的生物含量进行快速评价与分析。

偏最小二乘回归(partial least squares regression, PLSR)是一种新型多元统计回归分析算法,可以解决多属性之间的高度相关性,避免回归模型估计失真。本文通过联合 RF 与 CARS 方法对荧光光谱属性进行精简,利用精简后的光谱属性建立 PLSR 模型进行水样预测,为矿井涌水激光诱导荧光光谱的预测定量评估提供理论依据。

1 实验部分

1.1 设备和仪器

用如图 1 所示的激光诱导荧光涌水光谱系统完成光谱数据的采集,该系统主要由激光器、光谱仪、荧光探头、光纤和上位机组成。选用波长为 405 nm 的蓝紫光半导体激光器(北京华源拓达激光技术有限公司),最大输出功率为 120 mW,光谱仪选用 USB2000+ 微型光纤光谱仪(美国 Ocean optics 公司),内含高灵敏度 2 048 像素的线性 CCD 阵列(型号 ILX511,日本索尼公司),设定其光谱检测范围为

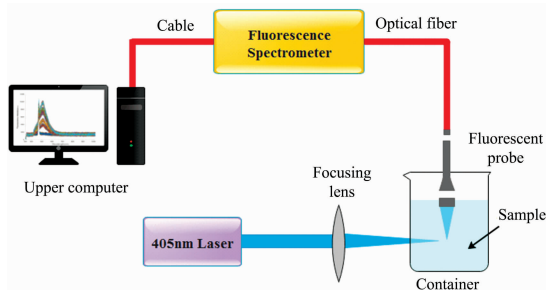


图 1 激光诱导荧光涌水光谱系统

Fig. 1 Laser-induced fluorescence spectroscopy system for water inflow

340~1 021 nm,分辨率为 0.5 nm,积分时间为 1 s/1 000 nm,浸入式微型荧光探头(型号 FPB-405-V3,广东科思凯公司)可插入待测水样获取荧光信号。为了避免其他光源对荧光光谱获取的干扰,测量在避光的暗室中进行,测量将探头垂直浸入水样并确保探头侵入透明容器的高度始终保持一致。使用计算机上的 Spectra Suite 软件进行记录收集所有样品的荧光光谱数据,算法仿真则是在 Matlab R2016b 和 Origin 2017 环境下运行。

1.2 材料和样本

矿井水害约 80%是由老空水引起的,老空水较其他涌水水源有着极强的破坏性,砂岩水害有着持续时间长、温度高等特点;将老空水、砂岩水、老空水与砂岩水的混合水作为研究对象进行实验,水样在 2019 年 3 月采集于安徽省淮南市顾桥煤矿。

由于老空水危害性之大,其含量严重关系到水害的防治工作,则以老空水为基础分别混入不同含量的砂岩水,第一组水样为老空水含量占总量的 0%(纯砂岩水),第二组水样为老空水含量占总量的 50%、第三组水样为老空水含量占总量的 67%、第四组水样为老空水含量占总量的 75%、第五组水样为老空水含量占总量的 80%、第六组水样为老空水含量占总量的 100%(纯老空水)。

将现场采集到的水样立即进行密封和遮光处理并带回实验室储存,以保障实验所采集数据的真实性与可靠性,每种水样各采集 50 组,共 300 组光谱数据作为实验样本。

1.3 光谱的预处理

光谱数据在传输过程中会受到设备、外界环境、操作不当等因素的干扰而存在大量的噪声信号,这些干扰信号与有用的光谱数据信息叠加在一起会严重影响实验结果^[13],为了避免噪声干扰,需要对原始荧光光谱进行滤波去噪处理。分别对原始光谱采用 Savitzky-Golay 卷积平滑法(Savitzky-Golay smoothing, S-G)、局部加权回归散点平滑法(locally weighted scatterplot smoothing, Lowess)进行去噪处理,根据选定回归模型的评估指标对比原始光谱和去噪后光谱的预测能力,选择合适的去噪方法。

1.4 RF-CARS 联合 PLSR 用于矿井涌水预测评估

随机森林是一种基于模型聚合思想用于解决分类和回归问题的算法,由 Breiman 于 2001 年提出^[14],可以在不增加计算复杂度的情况下,对变量有着较好的解释作用;RF 算法进行属性约简主要是删去重要度较低的属性,当加入随机噪声后,袋外数据准确率无明显变化,说明这个特征对于样本的预测结果影响不大,进而说明重要程度较低,需将其删去以保留重要度较高的属性。

RF 算法递归属性约简的步骤如下:

(1) 计算每个属性的重要度 W ,并按重要度依次递减顺序降序排列

属性重要度

$$W = \sum (\text{Error}_2 - \text{Error}_1) / N \quad (1)$$

式(1)中, N 为随机森林树的棵树, Error_2 代表加入噪声干扰的属性袋外数据误差, Error_1 代表属性的袋外数据误差。

(2) 剔除重要度最低的属性, 剩余的属性组成新的属性集合。

(3) 利用新组成的属性集合构建回归预测模型;

(4) 重复步骤(1)和(2), 比较每次递归所建 PLSR 模型的预测精度;

(5) 选出预测精度最高, 评估指标最好的一组属性集合。

竞争性自适应重加权算法是基于达尔文自然选择方式提出的^[15], 通过蒙特卡洛采样建立回归模型并计算变量回归系数的绝对值权重, 去掉权重小的波长点, 利用交互验证选出交叉验证均方根误差 (root mean square error of cross validation, RMSECV) 最小的子集, 可有效寻找出最优变量组合。

CARS 算法进行属性精简的步骤如下:

(1) 蒙特卡洛采样, 随机抽取一定比例样品作为校正集建立 PLSR 模型。

(2) 计算变量回归系数的绝对值权重 w , 评价属性的有用性

$$\text{权重: } w = |\alpha_i| / \sum_i^q |\alpha_i|, i = 1, 2, \dots, q \quad (2)$$

式(2)中, α_i 为第 i 个属性的重要度, q 为属性个数。

(3) 指数衰减函数用来去除贡献度 α 较小的属性。

(4) 采用 ARS 采样技术提取出新的属性集合 X , 建立 PLSR 模型, 并计算 RMSECV。

(5) n 次采样后, 挑选出 RMSECV 最小的集合为最优属性子集。

最终采用判定系数 R^2 、残差平方和 RSS、校正均方根误差 RMSEC、预测均方根误差 RMSEP、平均绝对误差 MAE 作为根据 PLSR 建立回归模型的评估指标, 模型 R^2 越高, RSS, RMSEC, RMSEP, MAE 越小表明模型具有良好的预测精度和性能。

2 结果与讨论

2.1 原始光谱数据获取

利用激光诱导荧光涌水光谱系统采集光谱数据, 呈现出的水样原始荧光光谱如图 2 所示, 不同比例的老空水受激光照射时, 其中的荧光物质会吸收光能, 并释放能量产生荧光, 形成荧光光谱, 在 400~600 nm 之间出现峰值; 六组水样的光谱分布主要分为三个部分, 最上层是老空水光谱, 最下层是砂岩水光谱, 由于这两种水样的化学成分和荧光物质浓度不同, 导致其光谱形状与波峰数量有很大差异, 容易区分, 中间部分为老空水分别混入不同含量砂岩水的四组水样, 这四组水样化学成分接近, 所呈现的光谱难以进行准确的辨别, 因此, 需要借助机器学习回归算法对涌水样本进行精确分析。

2.2 光谱预处理

分别对原始光谱采用 S-G 卷积平滑和 Lowess 平滑方法进行去噪预处理, 都采用 3 个窗口进行平滑, 如图 3 所示。较原始荧光光谱, 由于第四、五两组水样老空水含量相近, 光谱依然存在部分重叠情况, 但其他重叠部分更为分散, 整

体水样更容易区分, 说明经 S-G 卷积平滑和 Lowess 平滑去噪预处理后的涌水荧光光谱更适合光谱分析。

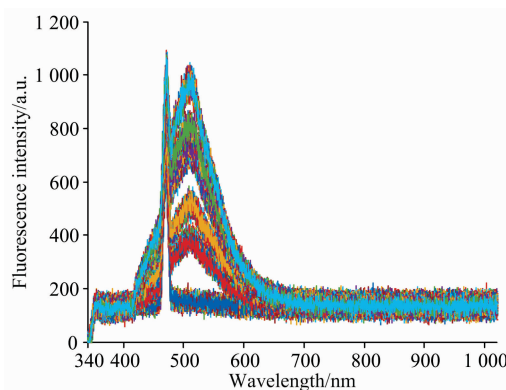


图 2 水样原始荧光光谱

Fig. 2 Original fluorescence spectra of water sample

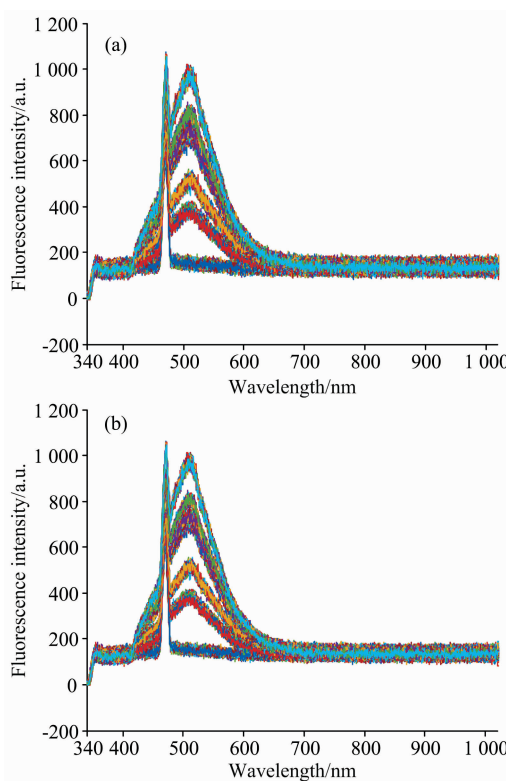


图 3 预处理后的荧光光谱

(a): S-G 卷积平滑; (b): Lowess 平滑

Fig. 3 Fluorescence spectra after pretreatment

(a): Savitzky-Golay smoothing; (b): Lowess smoothing

2.3 RF-CARS 属性精简

原始荧光光谱数据共 2 048 个属性, 每个属性都含有不同光谱信息, 不同属性对于光谱分析的重要性程度存在明显差异, 非关键且重要度低的属性将会影响涌水水样回归模型的建立, 达不到较好的预测效果; 利用 RF 属性约简方法删除重要度低的属性, 提升建模的效率和预测能力。随机将 300 组样本以 4:1 的比例划分, 240 组样本 (每组 40 个样

本)划分为校正集,剩余的 60 组(每组 10 个样本)作为预测集,先对原始光谱及两种去噪方法分别建立 PLSR 模型,再用 RF 分别对其进行初次属性约简;设置树的棵数 n_{trees} 为 200,如表 1 所示,经 Lowess 平滑法去噪后,模型的预测效果最好,且经过初次约简的属性整体重要性程度 W 较高且相对稳定,最终选用 Lowess 平滑法去噪的光谱数据进行研究。

表 1 不同去噪方法预测结果

Table 1 Prediction results of different denoising methods

去噪方法	R^2	RSS	RMSEP	MAE	W 平均值	W 标准差
原始光谱	0.990 1	0.058 6	0.031 4	0.027 0	0.015 8	0.041 2
S-G	0.991 2	0.052 7	0.029 7	0.024 7	0.016 9	0.041 0
Lowess	0.991 4	0.051 7	0.029 4	0.024 3	0.017 8	0.040 0

属性初次约简的属性重要度分布情况如图 4 所示,可以看出在 2 048 个属性中,大部分属性(共 1 662 个)的重要度为 0,主要分布在波峰两侧平缓的光谱范围区间内,这些属性对水样的预测结果没有任何影响,属于非关键光谱数据信息,可以将其删去,其中部分属性由于平滑处理消除了干扰的噪声信号,使其具有了重要度,保留剩下的 386 个属性作为初次约简后的属性。

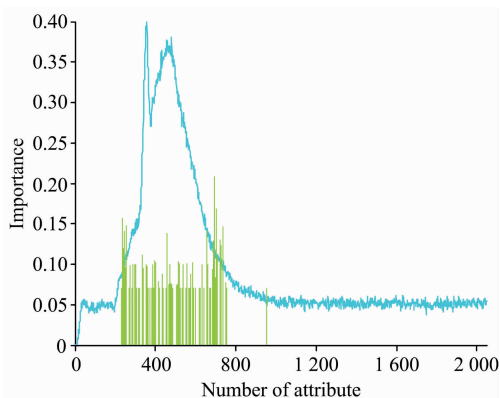


图 4 属性重要度分布

Fig. 4 Distribution of attribute importance

利用 RF 算法继续对初次约简出的 386 个属性进行递归约简,删去重要度为 0 的不相关属性,根据 PLSR 建立回归模型,递归循环 16 次的 RF 属性约简结果如表 2 所示,随着递归次数不断增加,关键属性个数逐渐递减且所对应的预测评估指标 R^2 和 RSS 也在发生变化,当递归次数达到 6 次时,回归模型的 R^2 达到最大, RSS 达到最小,预测效果最好,之后关键属性个数和整体预测精度都逐渐趋于稳定,则约简出的 223 个光谱属性作为 RF 算法的最终约简结果。

约简后的光谱属性数量明显减少,模型预测精度得到提升,为了达到精准评估的要求,需进一步精简光谱属性,将 RF 算法约简后的 223 个光谱属性采用 CARS 算法进行二次约简。

设定蒙特卡洛采样次数为 200,属性精简情况如图 5 所示,图 5(a)中表示属性精简过程中被选中属性的变化趋势,

随着采样次数的逐渐增加,由于属性的粗选和精选过程,被选属性的选择速度逐渐减小,由图 5(b)看出采样过程中, RMSECV 值整体变化趋势是先减小后增大,与矿井涌水预测评估无关的荧光光谱信息在 RMSECV 值减小过程中被剔除,对照图 5(c),当对应蓝色标注位置采样 46 次时, RMSECV 值达到最小值 0.021 1,有用的光谱信息则在之后的采样过程中被消去而降低了模型的预测能力, CARS 算法最终精简出了 77 个有用的光谱属性。

表 2 RF 属性约简结果

Table 2 Result of attribute reduction in RF

递归次数	属性个数	R^2	RSS	递归次数	属性个数	R^2	RSS
1	386	0.993 8	0.037 4	9	196	0.994 0	0.036 1
2	305	0.993 9	0.037 2	10	190	0.993 8	0.037 2
3	271	0.993 9	0.036 7	11	187	0.993 9	0.036 6
4	251	0.993 9	0.036 6	12	184	0.993 8	0.037 1
5	231	0.993 7	0.038 0	13	182	0.993 8	0.037 2
6	223	0.994 1	0.036 0	14	178	0.993 6	0.038 6
7	211	0.994 0	0.037 0	15	175	0.993 5	0.039 1
8	204	0.993 8	0.037 7	16	172	0.993 5	0.038 8

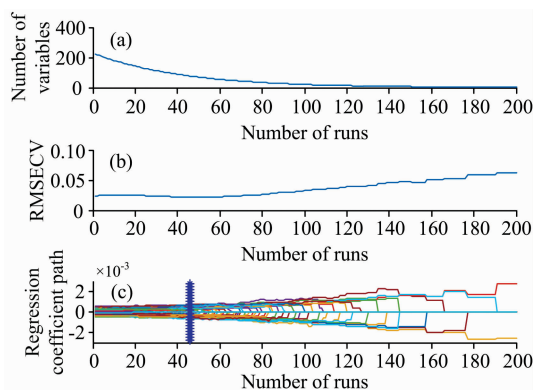


图 5 CARS 属性精简情况

Fig. 5 Attribute reduction results of CARS

2.4 PLSR 模型验证

对 RF-CARS 算法精简出的 77 个光谱属性数据建立 PLSR 模型来最终预测评估不同含量的矿井涌水。校正集、预测集的预测结果分别为图 6 和图 7 所示,可以看出校正集中老空水不同含量占比的预测值与真实值之间的判定系数 R^2_{cal} 为 0.997 5,预测集的预测值与真实值之间的判定系数 R^2_{pre} 为 0.996 7,均介于 0.8~1.0 之间,呈极强正相关性,模型预测性能优异,且预测样本能够很好的拟合在回归直线上。

为了验证 RF-CARS 算法结合 PLSR 模型用于矿井涌水预测的评估效果与可靠性,还需将与全属性分别使用 PLSR 和间隔偏最小二乘法(interval partial least squares, iPLS)建模,精简后的属性采用 PLSR、支持向量回归(support vector regression, SVR)、主成分回归(principle component regression, PCR)建模方法进行纵向和横向全面对比,对比结果如

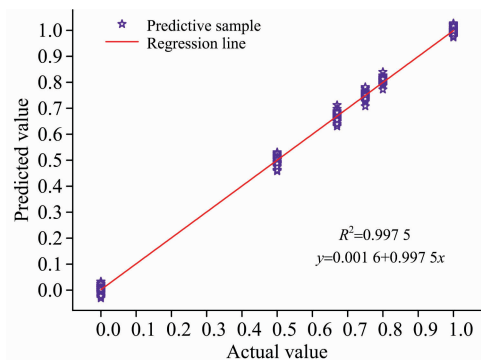


图 6 校正集预测结果

Fig. 6 Prediction results of calibration set

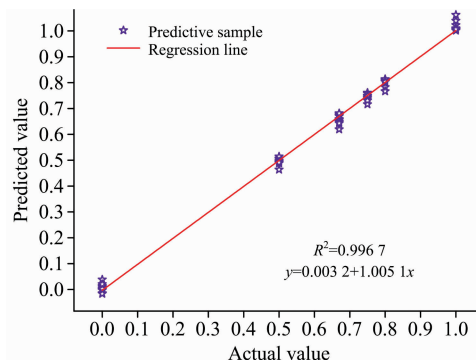


图 7 预测集预测结果

Fig. 7 Prediction results of prediction set

表 3 所示, 全光谱属性先后经 RF 约简和 RF-CARS 精简后属性的 PLSR 模型预测结果皆好于全光谱属性, 属性个数分别减少至仅原来的 10.89% 和 3.76%, 且较全属性和基于 RF 的 PLSR 模型, RF-CARS 方法精简出属性建模的校正集判定系数 R_{cal}^2 、预测集判定系数 R_{pre}^2 最高, 其他评估指标达到最低, 模型预测评估性能最好。较全属性 PLSR 模型, iPLS

建模属性个数减少至原来的 8.4%, 仅校正集预测精度有所提升, 预测集预测精度基本不变, 较基于 RF-CARS 的 PLSR 模型, 虽然基于 RF-CARS 的 SVR 模型校正集达到了很高的预测精度, 但预测集预测精度较低, 采用 PCR 建模, 则校正集和预测集的预测精度都相对较低。

表 3 预测结果对比

Table 3 Comparisons of predicted results

建模方法	属性个数	校正集				预测集			
		R_{cal}^2	RSS	RMSEC	MAE	R_{pre}^2	RSS	RMSEP	MAE
全属性-PLSR	2 048	0.993 0	0.166 6	0.026 4	0.021 3	0.991 4	0.051 7	0.029 4	0.027 0
RF-PLSR	223	0.997 2	0.066 9	0.016 7	0.013 4	0.994 1	0.036 0	0.025 1	0.020 3
RF-CARS-PLSR	77	0.997 5	0.060 1	0.015 8	0.012 7	0.996 7	0.019 9	0.018 3	0.013 5
全属性-iPLS	172	0.994 3	0.135 3	0.023 8	0.018 7	0.991 3	0.052 6	0.029 7	0.024 7
RF-CARS-SVR	77	0.999 8	0.016 5	0.004 4	0.004 2	0.987 8	0.069 7	0.038 3	0.023 3
RF-CARS-PCR	77	0.992 1	0.187 0	0.028 0	0.022 4	0.994 0	0.036 9	0.025 3	0.021 3

3 结 论

提出了一种基于激光诱导荧光光谱结合 RF-CARS 属性精简方法对不同含量的矿井涌水进行预测评估, 通过 Lowess 平滑预处理能有效降低噪声影响, 经 RF-CARS 精简后的光谱属性建立 PLSR 模型的校正集判定系数 R_{cal}^2 为 0.997 5, 预测集判定系数 R_{pre}^2 为 0.996 7, 预测性能均高于全属性建立的 PLSR 模型, RMSEC 和 RMSEP 分别为 0.015 8

和 0.018 3, 模型预测误差很小, 同时与 SVR 和 PCR 模型及 iPLS 属性选择方法比较, RF-CARS 算法的属性精简能力和模型性能要好于 iPLS 算法, 精简出的 77 个光谱属性仅占全属性的 10.89%, 有效降低了模型的复杂程度, 减少了参数的繁琐调节过程, 增强了各光谱属性之间的联系, 避免无用属性对预测结果的干扰, 且精简后属性所建立的 PLSR 模型预测性能要好于 SVR 和 PCR 模型。通过实验分析得出利用 RF-CARS 结合激光诱导荧光技术用于矿井涌水预测评估是切实可行的, 可做到矿井涌水含量的精准预测和定量分析。

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): 5.
- [2] ZHANG Hao, YAO Duo-shan, LU Hai-feng, et al(张好, 姚多喜, 鲁海峰, 等). Geological Journal of China Universities(高校地质学报), 2017, 23(2): 366.
- [3] LIU Guo-wang, CHANG Hao-yu, GUO Jun-zhong(刘国旺, 常浩宇, 郭均中). Coal Science and Technology(煤炭科学技术), 2017, 45(8): 223.
- [4] Hu F, Zhou M, Yan P, et al. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2019, 219: 367.
- [5] HE Chen-yang, ZHOU Meng-ran, YAN Peng-cheng(何晨阳, 周孟然, 闫鹏程). Spectroscopy and Spectral Analysis(光谱学与光谱分析)

- 析), 2016, 36(7): 2234.
- [6] Hu F, Zhou M, Yan P, et al. RSC Advances, 2019, 9(14): 7673.
- [7] WEN Ze-bo, KANG Yu, CAO Yang, et al(文泽波, 康宇, 曹洋, 等). Journal of University of Science and Technology of China(中国科技大学学报), 2017, 47(8): 653.
- [8] Brokamp C, Jandarov R, Hossain M, et al. Environmental Science & Technology, 2018, 52(7): 181.
- [9] Lefkovits László, Lefkovits Szidónia, Emerich S, et al. Proc. SPIE, 2017, 10341: 1034117.
- [10] Li S, Zhang X, Shan Y, et al. Food Chemistry, 2017, 218: 231.
- [11] LIU Shan-shan, ZHANG Jun, LIN Si-han, et al(刘珊珊, 张俊, 林思寒, 等). Laser & Optoelectronics Progress(激光与光电子学进展), 2018, (2): 463.
- [12] Wang Y, Jiang F, Gupta B B, et al. IEEE Access, 2018, 6: 5290.
- [13] LIU Wen-xia(刘文霞). Computer Simulation(计算机仿真), 2016, 33(5): 192.
- [14] Matin S S, Hower J C, Farahzadi L, et al. International Journal of Mineral Processing, 2016, 155: 140.
- [15] Chen W, Zou J, Wan F, et al. AIP Advances, 2018, 8(3): 035204(doi: 10.1063/1.5012685).

RF-CARS Combined with LIF Spectroscopy for Prediction and Assessment of Mine Water Inflow

BIAN Kai, ZHOU Meng-ran*, HU Feng, LAI Wen-hao, YAN Peng-cheng, SONG Hong-ping, DAI Rong-ying, HU Tian-yu
College of Electrical and Information Engineering, Anhui University of Science and Technology, Huainan 232001, China

Abstract Quick and accurate identification of mine water inflow has important research significance for preventing coal mine flood accidents, the laser-induced fluorescence (LIF) spectroscopy is used to integrate with the intelligent classification algorithm to identify the mine water inflow, it breaks the shortcomings of traditional water chemistry methods, such as long time consuming, etc., and has the characteristics of high sensitivity and fast response. However, these currently used algorithms can only rely on the classification accuracy to qualitatively discriminate the types of water samples from different mine water inflow. This paper combines the random forest algorithm with the competitive adaptive weighting algorithm (RF-CARS), the partial least squares regression (PLSR) model based on fluorescence spectrum data from the laser-induced fluorescence was used to predict the water inflow in different mines and to achieve quantitative assessment of water samples. Firstly, 300 sets of mine water inflow samples mixed with different sandstone waters based on goaf water were collected, and the collected water samples were randomly divided into the calibration set and the prediction set according to the ratio of 4 : 1, a total of 240 sets of calibration sets were used to establish a regression model, a total of 60 sets of prediction sets were used to predict different water samples, and a laser-induced fluorescence inflow spectroscopy system was built to complete the acquisition of spectral data and generated a fluorescence spectrum. Then the original fluorescence spectrum was denoised by S-G convolution smoothing method and Lowess smoothing method, and it was found that the processed fluorescence spectrum was more dispersed than the original spectrum, which was suitable for spectral analysis, the prediction accuracy of two denoising methods were compared, the Lowess was chosen as the final denoising method. Then, the RF algorithm was used to reduce the spectral attributes with low attribute importance after denoising, according to the performance of the optimal regression model, the 223 reduced attributes were selected and then it was used for the secondary attribute reduction of the CARS algorithm. The PLSR model was established based on 77 spectral attribute data selected according to the principle of minimum cross validation root mean square error in the sampling process of CARS algorithm. Finally, we compared with the full spectrum, other variable selection methods, and different regression models, the RF-CARS algorithm had the best streamlining effect, and the total spectral modeling attribute was reduced from 2048 to 77, the model prediction set determination coefficient R_{pre}^2 increased from 0.9914 to 0.9967, the predicted root mean square error RMSEP decreased from 0.0294 to 0.0183, the prediction accuracy was improved, and the remaining evaluation indicators were relatively good. The experimental results show that the RF-CARS combined with laser induced fluorescence technology can quickly and accurately predict mine water inflow, the simplified spectral attributes are used to establish regression model, which provides a theoretical guarantee for real-time quantitative evaluation of mine water inflow.

Keywords Laser-induced fluorescence spectroscopy; Regression model; Random forest; Competitive adaptive reweighted sampling; Mine water inflow

* Corresponding author

(Received Jul. 16, 2019; accepted Nov. 8, 2019)