

## 基于 iPLS 的矿井突水激光诱导荧光光谱特征波段筛选

周孟然, 卞凯\*, 胡锋, 来文豪, 闫鹏程

安徽理工大学电气与信息工程学院, 安徽淮南 232001

**摘要** 矿井突水一直威胁着煤矿井下施工人员的安全, 准确且快速识别矿井突水水源类型对于矿井的安全生产起到关键性作用。激光诱导荧光(LIF)光谱技术识别矿井突水水源, 有效避免了常规的水化学法需要测定多种化学参数, 水源识别时间过长的缺点。提出一种间隔偏最小二乘法(iPLS)与粒子群联合支持向量分类算法(PSO-SVC)相结合的方法, iPLS算法常应用于光谱波段优选和模型的回归分析, PSO-SVC则在机器学习领域有着重要的应用, 激光诱导荧光技术具有快速的时间响应、测量精度高等特点, iPLS和PSO-SVC算法运用于光谱图和光谱数据的分析, 进而可以对突水水源类型识别分类。首先, 用淮南矿区采集到的7种(每种水样30组)共210组荧光光谱数据进行实验, 对老空水、灰岩水、灰岩水和老空水不同体积比混合水样的激光诱导荧光光谱图的差异性进行分析。比较了留出法和Kennard-Stone样本划分方法所得到的PSO-SVC模型分类准确率, 采用留出法得到的训练集水样(140组)和测试集水样(70组)作为实验样本。其次, 用iPLS算法将全光谱波段依次按10~25波段区间进行等分, 选取划分区间的RMSECV(交叉验证均方根误差)值小于全光谱波段RMSECV值(阈值)的波段作为特征波段, 结合光谱图对比分析了划分10和14个子区间的建模结果, 发现通过直接观察得到的特征波段与iPLS算法筛选出的特征波段存在误差。最后, 在不进行去噪、降维等预处理条件下, 根据iPLS划分不同区间数的评价指标统计数据, 选取划分11个区间所筛选出具有561个波长点的410.078~478.424和545.078~674.104 nm特征波段范围数据作为PSO-SVC模型的输入, 以iPLS结合PSO-SVC算法筛选出的特征波段与全光谱波段、直接观察得到波段建模准确率相比, 训练集与测试集的分类准确率高达100%, PSO寻优到的最佳惩罚系数 $c$ 为1.3670, 核函数参数 $g$ 为0.5762。从实验结果可以看出, 利用iPLS进行荧光光谱的特征波段筛选是切实可行的, 提取出的特征波段能充分反映出全光谱波段的有效信息, 为激光诱导荧光光谱技术用于矿井突水水源精准在线识别的研究提供了理论依据。

**关键词** 矿井突水; 激光诱导荧光; 间隔偏最小二乘法; 特征波段; 支持向量分类

**中图分类号:** O657.3 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2019)07-2196-06

### 引言

矿井突水是一种重大的煤矿灾害, 不仅会导致人员伤亡, 还会带来巨大的经济损失<sup>[1]</sup>, 随着煤矿开采深度的不断增加, 矿井突水问题变得日益严重<sup>[2]</sup>。怎样才能第一时间查明矿井突水的形成原因, 并且安全、准确地获取突水水源类型, 已经成为了突水治理问题的关键所在。煤矿突水水源识别常用水化学方法进行分析, 通过测定出pH值、电导率、离子浓度等参数, 最后根据这些参数建立水源识别模型, 识

别出不同的突水水源类型<sup>[3]</sup>。一般来说, 在实验过程中, 通过水化学分析法测定水样中的各个参数需要较长的时间(1~2 h), 这样很难搭建在线预警和快速突水水源识别系统, 不利于水害防治。激光诱导荧光光谱技术具有较高的空间分辨能力、快速的时间响应、测量精度高和一定的抗干扰能力等优点, 在生物、医学、环境等领域都有着广泛的应用<sup>[4]</sup>。对于煤矿突水水源识别, 激光诱导荧光光谱分析技术在最近几年已经应用于此领域, 并且取得了一定的成效。如闫鹏程等通过不同的降噪预处理后的LIF技术光谱数据采用主成分分析法建模, 使用独立软模式算法对煤矿突水水源进行精准

收稿日期: 2018-09-09, 修订日期: 2019-02-06

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), 国家自然科学基金项目(51174258), 安徽省青年科学基金项目(1808085QE157)资助

作者简介: 周孟然, 1965年生, 安徽理工大学电气与信息工程学院教授 e-mail: mrzhou8521@163.com

\* 通讯联系人 e-mail: kbian92@163.com

识别。王亚等<sup>[5]</sup>将水样荧光光谱通过正则化非线性特征提取后,采用极限学习机算法建立多分类模型,实现了突水水源的稳定识别。现有的煤矿突水水样荧光光谱分析方法,如降维结合模式识别、模糊 C 均值聚类算法(fuzzy C-means algorithm, FCM)、卷积神经网络(convolutional neural networks, CNN)等都需要全波段光谱数据,此外,依靠主观经验观察筛选波段的定性分析存在一定的误差,没有绝对的说服力,主成分分析等一些特征提取降维方法所提取的主成分并没有实际的物理意义和对输出变量的解释能力。

iPLS 算法是一种光谱特征波段筛选的方法,它能够减少全光谱波段冗余数据量,提高建模效率,消除直接观察进行波段截取的误差。iPLS 已被广泛应用于近红外光谱、紫外光谱、拉曼光谱等特征波长选择,如郭慧娟<sup>[6]</sup>将 iPLS 和 Si-PLS 算法相结合对人体血清胆红素的近红外光谱数据建模,分析了样品的含量。杨鹏程<sup>[7]</sup>提出利用 iPLS 和紫外光谱法,实现了海水硝酸盐最佳建模波长的选取。赵芳<sup>[8]</sup>利用 iPLS 结合拉曼光谱对猪肉皮下脂肪的碘值进行了测定。Giraud<sup>[9]</sup>运用了间隔偏最小二乘判别分析与 RGB 图像结合,自动识别出榛子仁的缺陷。

PSO-SVC 是一种分类优化算法,有研究提出了一种粒子群寻优支持向量机参数的改进算法,提升了高光谱图像的分类精度。PSO-SVC 优化模型具有更优的分类性能与泛化能力,适合突水水源识别模型的建立与研究,本文所采用的 iPLS 算法是直观地对激光诱导荧光光谱波段进行筛选的定量分析,结合 PSO-SVC 算法,免去了去噪、降维的过程,探究了 iPLS 筛选矿井突水水源激光诱导荧光光谱特征波段的可行性。

## 1 原理

### 1.1 iPLS 算法

间隔偏最小二乘法(interval PLS, iPLS)是由 Norgaard 等提出的波段区间选择方法。iPLS 算法主要将数据集分成若干个同等宽度间隔,计算每个子区间的 PLS 模型评价指标,并根据 RMSECV(交叉验证均方根误差)值优选出建模效果比较好的一个或多个波段。iPLS 能够在光谱图上直接对各波段光谱的仿真柱状图参数统计信息进行对比分析,通过直接观察与简单的计算就可以确定对应不同的波段范围和波长点数,物理意义清晰明确<sup>[10]</sup>。

iPLS 对于预测性能进行比较主要基于参数 RMSECV,其次,还评估了其他参数,如相关系数  $r$ 、偏移量 Bias,以确保全面的模型概述。RMSECV,  $r$  和 Bias 可以用式(1)~式(3)来计算

$$\text{RMSECV} = \sqrt{\frac{\sum_{i=1}^n (y_p - y_q)^2}{n}} \quad (1)$$

$$r = \sqrt{1 - \frac{\sum_{i=1}^m (y_p - y_q)^2}{\sum_{i=1}^m (y_q - y_a)^2}} \quad (2)$$

$$\text{Bias} = y_p - y_q \quad (3)$$

式中,  $y_p$  为实验的预测值,  $y_q$  为实验的测量值,  $y_a$  为实验的所有样本值,  $n$  是训练集样本数,  $m$  是测试集样本数, RMSECV 由交叉验证的样本中计算出。

### 1.2 PSO-SVC 算法

支持向量分类算法(support vector classification, SVC)是一种 SVM 用来做分类的算法<sup>[12]</sup>,粒子群算法(particle swarm optimization, PSO)是仿效鸟类捕食行为的一种并行优化的启发式算法<sup>[13]</sup>。在 SVC 训练过程中,选用高斯(RBF)函数作为 SVC 的核函数, SVC 模型性能受惩罚系数  $c$  和核函数参数  $g$  影响,通过 PSO 算法对 SVC 进行参数寻优,寻找到最佳的分类参数  $c$  和  $g$ ,从而获得良好的性能模型。

PSO 优化 SVC 模型的步骤如下:

(1)选择 SVC 模型的训练集和测试集,设定  $c$  和  $g$  的搜索区间;

(2)确定目标函数,初始化种群  $X = (X_1, X_2, \dots, X_n)$  和速度  $v = (v_1, v_2, \dots, v_n)$ ,找到解集的方位;

(3)计算适应度函数,给适应度定标准,变更个体的速度  $v$ ,找到最佳适应度所在方位;

(4)若满足终止条件,则执行步骤(5),否则继续进行步骤(3);

(5)确定并输出最优解 best  $c$ , best  $g$ ;

(6)利用最优解  $c, g$ , SVC 分类模型进行分类识别。

## 2 实验部分

### 2.1 材料

煤矿突水事故中,老空水具有速度快,破坏能力强,静态含量多等特点,危害性要高于其他突水水源,淮南矿区的水文地质特点鲜明,突水灾害治理的对象主要是灰岩水。本实验以灰岩水、老空水、一定体积比例的灰岩水与老空水所组成的混合水为实验对象。实验材料采用 2018 年 4 月 13 日在安徽省淮南矿区采集到的灰岩水、灰岩水和老空水按体积比为 1:3 的混合水(简称“混合水 1”)、灰岩水和老空水按体积比为 1:2 的混合水(简称“混合水 2”)、灰岩水和老空水按体积比为 1:1 的混合水(简称“混合水 3”)、灰岩水和老空水按体积比为 2:1 的混合水(简称“混合水 4”)、灰岩水和老空水按体积比为 3:1 的混合水(简称“混合水 5”)、老空水共七种不同的水样进行实验。其中,每种水样各采集 30 组光谱数据,总共 210 组,收集到的水样需在实验室密封遮光存储。

### 2.2 仪器及荧光光谱采集

煤矿突水水样的荧光光谱数据采集,是在图 1 所示的实验系统下完成的。

实验所选用的仪器为 USB2000+型激光诱导荧光光谱仪(美国 Ocean optics 公司),荧光光谱检测全波段范围为 340.472~1 021.086 nm。激光器使用 100 mW 的 LSR 405 nm 蓝紫光半导体激光器,激光入射波长设定值为 405 nm。激光器与荧光光谱仪通过 SMA905 接头的光纤接口相连接,两块滤光片过滤掉荧光多余信息。为了避免背景光和周围环境因素对荧光光谱采集实验的影响,荧光探头与存放水样的

玻璃器皿都在相同温湿度的暗室中进行。荧光光谱的所有数据都由 Spectra suite 软件采集并记录, 实验中的算法程序均在 Matlab R2016b 环境下完成的, 其中, SVC 算法是利用 LIBSVM 工具箱实现的。

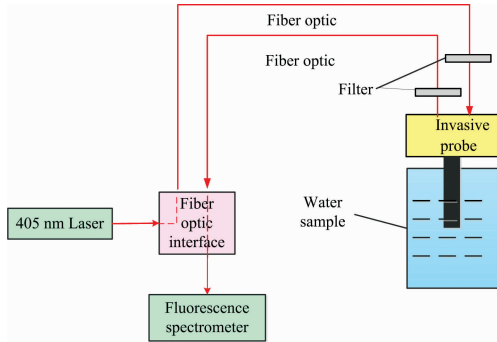


图 1 激光诱导荧光实验系统图

Fig. 1 Laser induced fluorescence experiment system diagram

### 3 结果与讨论

#### 3.1 光谱图分析

为了便于谱图的辨识, 实验所测平滑后的全波段原始荧光光谱图如图 2 所示, 其中, 横坐标为荧光光谱的波长, 纵坐标为荧光强度。自上而下分别是老空水、混合水 1、混合水 2、混合水 3、混合水 4、混合水 5、灰岩水 7 种不同的水样。从图中可以清楚观察到老空水与灰岩水波形差距明显, 此外, 老空水有两个波峰, 灰岩水有一个波峰, 所对应的荧光强度和波峰位置都不相同, 但混合水之间差别不大, 通过观察难以区分混合水样的种类。因此, 需对包含混合水水样的荧光光谱图进行研究, 分析和使用波段优化的分类算法对混合水样进行识别。

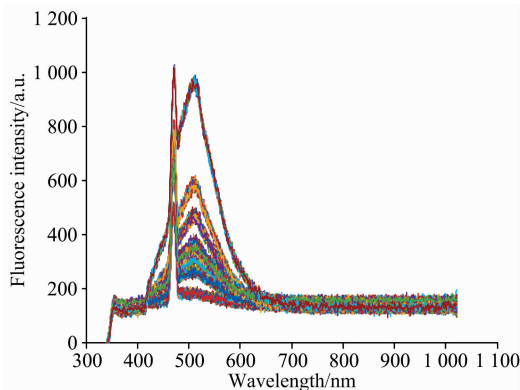


图 2 原始荧光光谱图

Fig. 2 Original fluorescence spectrum

#### 3.2 样本的训练集和测试集划分

首先, 采用 K-S(Kennard-Stone)算法<sup>[13]</sup>按照训练集与测试集 2:1 的比例对七种, 共计 210 个样本的水样进行划分。其次, 采用留出法(hold-out)随机选取 140 个样本, 每种水样各 20 个样本作为训练集, 剩余 70 个样本, 每种水样各 10 个样本作为测试集, 分别将两种不同划分方法的训练集

与测试集作为 PSO-SVC 模型的输入。由表 1 可以看出, 使用 hold-out 划分样本集比 K-S 法的分类准确率高, 后续的实验过程将采用 hold-out 获得的训练集、测试集样本进行。

表 1 不同划分方法准确率结果

Table 1 Accuracy of different division methods

划分方法	最佳 $c$	最佳 $g$	训练集准确率/%	测试集准确率/%
K-S	1.414 2	0.011 0	(140/140)100	(66/70)94.28
hold-out	2.768 2	0.321 9	(140/140)100	(67/70)95.71

#### 3.3 iPLS 波段筛选

通过直接观察图 2, 发现在全光谱波段 340.472 ~ 1 021.086 nm 范围内(共 2 048 个波长点), 老空水和混合水有两个波峰, 分别约在 480 和 520 nm 处, 灰岩水有一个峰, 约在 470 nm 处, 应当选取的特征波段范围为 340~700 nm, 而其他波长范围光谱曲线平缓, 没有特征峰的出现。直接观察得到的特征波段可以用 iPLS 算法来检验其正确性。

现利用 iPLS 波段筛选的方法将全光谱波段依次按 10~25 波段区间进行等分, 并在每一个区间上建立 PLS 回归模型。将全光谱等分成 10 个子区间(每个区间波长点数 206 个)时, 如图 3 所示, 区间上斜体数字表示最佳的 PLS 成分数量, 红色虚线处所对应的 RMSECV 值是用 2 个 PLS 成分建立在全光谱波段模型所得, 称这个值为区间选择的阈值。从每个子区间的 RMSECV 值来看, 第 2, 3, 4 和 5 区间 RMSECV 值明显小于全光谱波段子区间选择的阈值 1.416 7。

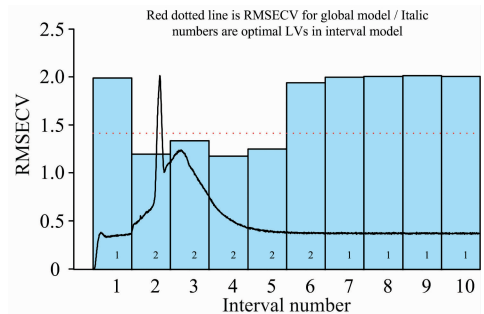


图 3 10 个子区间 iPLS 建模结果

Fig. 3 10 sub interval iPLS modeling result

表 2 iPLS 波段筛选结果

Table 2 iPLS wave bands screening results

区间编号	入选光谱/nm	PLS 成分数	RMSECV	$r$	Bias
全光谱	340.472~1 021.086	2	1.416 7	0.705 7	-0.009 6
第 1 区间	340.472~416.712	1	1.993 2	-0.030 1	0.045 9
第 2 区间	416.712~491.430	2	1.200 1	0.800 2	0.005 4
第 3 区间	419.430~564.552	2	1.340 4	0.745 6	-0.009 1
第 4 区间	564.522~635.884	2	1.176 5	0.809 1	-0.004 7
第 5 区间	635.884~705.412	2	1.250 6	0.780 9	-0.024 2
第 6 区间	705.412~773.004	2	1.944 6	0.313 8	-0.025 1
第 7 区间	773.004~838.555	1	2.001 9	-0.015 2	-0.087 7
第 8 区间	838.555~901.962	1	2.007 2	0.118 9	-0.036 5
第 9 区间	901.962~962.829	1	2.016 7	0.095 3	-0.022 1
第 10 区间	962.829~1 021.086	1	2.008 3	-0.012 9	0.099 5

iPLS 波断筛选结果如表 2 所示, 可见, 416.712~705.412 nm 为筛选出的特征波段, 共 821 个波长点。

将全光谱等分为 14 个子区间(每个区间波长点数 148 个)时, 如图 4 所示, 395.292~502.590 与 554.992~656.428 nm 为筛选出的特征波段, 所对应区间 2, 3, 5 和 6 共有 587 个波长点。与图 3 相比, 不同区间筛选出的区间大体相同, 差别不大, 整体偏向中间区域, 差别主要体现在等分成 14 个区间的 502.590~554.992 nm 波段范围, 该范围

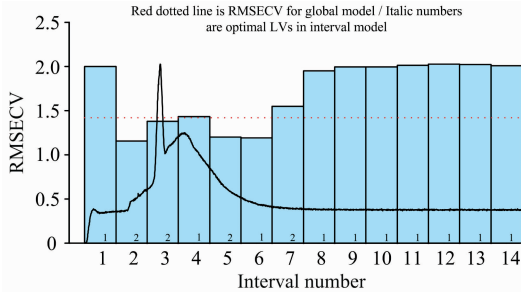


图 4 14 个子区间 iPLS 建模结果

Fig. 4 14 sub interval iPLS modeling result

表 3 iPLS 划分不同区间数的统计结果

Table 3 Statistical results of iPLS division of different interval numbers

iPLS 划分 区间数	特征波段范围/nm	波长点 总数	平均 RMSECV	平均 <i>r</i>
10	416.712~705.412	821	1.241 9	0.783 9
11	410.078~478.424, 545.078~674.104	561	1.181 7	0.792 8
12	404.17~705.747	856	1.274 8	0.768 7
13	399.364~679.518	791	1.261 6	0.767 2
14	395.292~502.590, 554.992~656.428	587	1.230 5	0.780 5
15	391.586~491.791, 540.819~683.234	686	1.255 8	0.778 3
16	388.249~482.042, 528.008~661.88	642	1.221 4	0.779 4
17	385.651~474.440, 560.996~686.27	607	1.269 4	0.762 7
18	383.051~507.977, 548.622~667.320	686	1.262 2	0.761 4
19	380.821~499.354, 537.976~650.965	650	1.285 5	0.757 9
20	417.081~492.151, 565.579~672.071	618	1.263 6	0.748 0
21	413.396~484.933, 554.992~657.110	492	1.262 4	0.760 8
22	410.446~478.424, 545.078~674.104	560	1.298 9	0.747 6
23	407.125~504.387, 567.692~660.177	536	1.296 1	0.738 8
24	404.540~498.634, 559.937~648.914	516	1.282 8	0.754 2
25	401.583~491.430, 550.039~663.921	576	1.300 6	0.741 0

的 RMSECV 值大于阈值, 所选波段波长点数较等分成 10 个区间减少了 234 个, 整体划分更为密集。

iPLS 划分不同区间数的统计结果如表 3 所示, 由于筛选出的特征波段不同, 各统计参数都在改变, 将平均 RMSECV 与平均 *r* 作为最终优选特征波段的评价指标。当 iPLS 划分 11 个区间时, 筛选出特征波段所对应的平均 RMSECV、平均 *r* 分别取到最小值 1.181 7 和最大值 0.792 8, 该波段范围为最终筛选结果。对应波长点数为 561 个, 相比于全波段的 2 048 个波长点, 总数减少了 72.61%, 也减少到直接观察 1 012 个波长点数的一半。较直接观察光谱图筛选出的 340~700 nm 范围特征波段, iPLS 却筛选出 410.078~478.424, 545.078~674.104 nm 两部分特征波段, 说明通过直接观察、主观经验得到的波段存在一定的误差, 可靠性低。

### 3.4 PSO-SVC 模型的建立

将上述等分 11 个区间 iPLS 筛选出的波段数据作为 PSO-SVC 模型输入, 用于最后的突水水源识别分析。PSO 进行参数寻优的适应度曲线如图 5 所示, 在种群数量为 20, 迭代次数为 100 时, 最终 CV 意义下的分类准确率为 97.1429%。分类结果如图 6 所示, 突水水样种类 1~7 分别对应灰岩水、混合水 1、混合水 2、混合水 3、混合水 4、混合水 5、老空水。测试集的 7 种突水水样没有出现偏差, 全部分类正确, 测试集样本分类准确率为 100%。

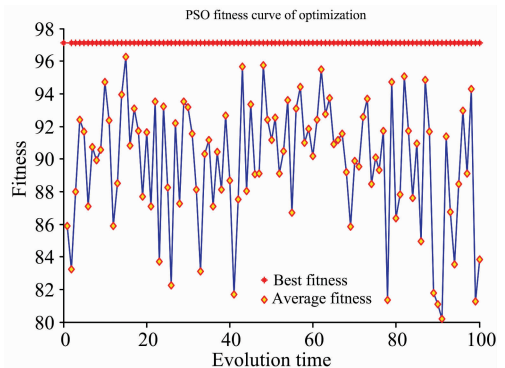


图 5 PSO 参数寻优图

Fig. 5 Optimization diagram of PSO parameters

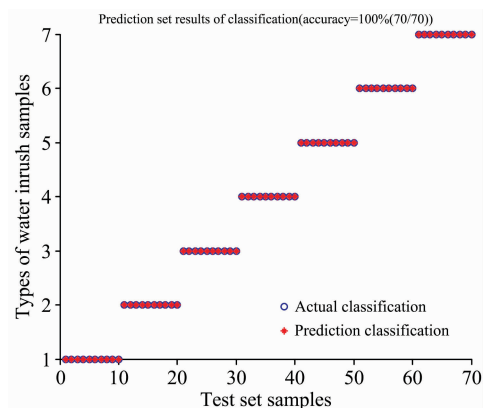


图 6 iPLS 光谱波段筛选分类结果图

Fig. 6 iPLS spectral band screening results classification diagram

为了验证 iPLS 波段筛选对于 PSO-SVC 建模的可靠性, 现将原始光谱全波段、观察筛选的 340~700 nm 范围波段、等分成 11 个区间时筛选出的特征波段, 这几种不同处理方法的分类情况对比如表 5 所示, 对于原始全光谱波段, 错误分类主要集中在混合水 2 和混合水 3, 而观察筛选的波段, 错误分类主要集中在混合水 4 和混合水 5, 原因是灰岩水与

老空水有相似的化学成分, 导致这 2 种水样组成的混合水不易识别。iPLS 算法筛选出特征波段的训练集、测试集分类准确率皆能达到 100%, 与表 5 其他处理方法相比, 最佳惩罚系数  $c$  最小且分类准确率最高, 因此泛化性能也是最理想的。

表 5 分类结果

Table 5 Result of classification

处理方法	参数 $c$	参数 $g$	波长点数	训练集准确率/%	测试集准确率/%
原始全波段	2.768 2	0.321 9	2 048	(140/140)100	(67/70)95.71
观察筛选波段	12.978 4	0.651 3	1 012	(140/140)100	(69/70)98.57
11 个区间波段筛选	1.367 0	0.576 2	561	(140/140)100	(70/70)100

## 4 结 论

对淮南矿区的灰岩水、老空水和按一定体积比混合的混合水原始激光诱导荧光光谱图进行分析, 通过 iPLS 算法筛选出特征波段, 并建立 PSO-SVC 模型, 最后对比了不同处理方法对分类结果的影响。由实验分析过程和结果可以看出:

(1) iPLS 筛选特征波段具有实际物理意义, 且直观、清晰地反映出特征波段在光谱图上的分布情况, 验证了 iPLS 算法对于波段筛选的定量分析比主观观察得到的波段可靠性高, 降低了从光谱图上直接观察获得波段的误差。

(2) 利用 iPLS 对荧光光谱波段进行筛选, 从全波段 2048 个波长点缩减为 561 个, 减少了光谱数据冗余信息的获取, 用较少的波长点使训练模型变得简单, 提升了建模效率和可靠性。

(3) PSO-SVC 对于突水水样建模有良好分类精度, 在不进行去噪和降维的预处理情况下, 结合 iPLS 波段筛选方法使分类准确率达到甚至高于全波段建模, 在某种程度上减少了过拟合, 提升了模型的泛化能力。

(4) 本实验充分论证了 iPLS 结合 PSO-SVC 算法在煤矿突水激光诱导荧光光谱分析的可行性, 既适用于灰岩水和老空水的混合水样的识别, 又适用于其他突水水样的识别及激光诱导荧光光谱分析。

## References

- [1] SUN Yun-jiang, ZUO Jian-ping, LI Yu-bao, et al(孙运江, 左建平, 李玉宝, 等). Rock and Soil Mechanics(岩土力学), 2017, 38(8): 2335.
- [2] FENG Zhi-dong, WANG Tao, GU Qing-hua, et al(冯治东, 王 桃, 顾清华, 等). Systems Engineering-Theory & Practice(系统工程理论与实践), 2017, 37(12): 3289.
- [3] ZHU Le-zhang(朱乐章). China Coal(中国煤炭), 2018, 44(5): 100.
- [4] LI Shao-dan, LIN Yuan-sheng, TAN Si-chao, et al(李少丹, 林原胜, 谭思超, 等). Nuclear Power Engineering(核动力工程), 2014, 35(4): 137.
- [5] WANG Ya, ZHOU Meng-ran, CHEN Rui-yun, et al(王 亚, 周孟然, 陈瑞云, 等). Acta Optica Sinica(光学学报), 2018, 38(7): 367.
- [6] GUO Hui-xian, ZHU Si-qi, LI Yuan-peng, et al(郭慧娴, 朱思祁, 黎远鹏, 等). Journal of Optoelectronics · Laser(光电子·激光), 2016, 27(10): 1136.
- [7] YANG Peng-cheng, DU Jun-lan, CHENG Chang-kuo(杨鹏程, 杜军兰, 程长阔). Marine Environmental Science(海洋环境科学), 2016, 35(6): 943.
- [8] ZHAO Fang, PENG Yan-kun(赵 芳, 彭彦昆). Chinese Journal of Lasers(中国激光), 2017, 44(11): 249.
- [9] Giraud A, Calvini R, Orlandi G, et al. Food Control, 2018, 94: 233.
- [10] XU Long, LU Jian-gang, YANG Qin-min, et al(徐 龙, 卢建刚, 杨秦敏, 等). CIESC Journal(化工学报), 2013, 64(12): 4410.
- [11] SHI Long-qing, TAN Xi-peng, WANG Juan, et al(施龙青, 谭希鹏, 王 娟, 等). Journal of China Coal Society(煤炭学报), 2015, 40(1): 167.
- [12] QI Shao-hua, LIU Qi-yuan, CHEN Jiu-hui, et al(齐少华, 刘启元, 陈九辉, 等). Chinese Journal of Geophysics(地球物理学报), 2016, 59(12): 4544.
- [13] CHEN Yi-yun, ZHAO Rui-ying, QI Tian-ci, et al(陈奕云, 赵瑞瑛, 齐天赐, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2017, 37(7): 2133.

# Selection of Characteristic Wave Bands for Laser Induced Fluorescence Spectra of Mine Water Inrush Based on IPLS

ZHOU Meng-ran, BIAN Kai\*, HU Feng, LAI Wen-hao, YAN Peng-cheng

College of Electrical and Information Engineering, Anhui University of Science and Technology, Huainan 232001, China

**Abstract** Mine water inrush has been threatening the safety of underground construction personnel, so an accurate and rapid identification of mine water inrush source plays a key role in the safe production of the mine. Identification of mine water inrush source by laser induced fluorescence spectroscopy effectively avoids the shortcomings of conventional hydrochemical methods which need to determine a variety of chemical parameters and the identification time is too long. In this paper, a method of interval PLS (iPLS) and particle swarm optimization combined with support vector classification algorithm (PSO-SVC) is proposed. The iPLS algorithm is often used in spectral bands optimization and regression analysis of models, and the PSO-SVC is an important application in the field of machine learning. The laser induced fluorescence spectroscopy (LIF) technology has the characteristics of fast time response and high measurement accuracy, and the iPLS and PSO-SVC algorithms are applied to the analysis of spectral maps and spectral data, and then it can identify and classify water inrush sources. Firstly, The 210 sets of fluorescence spectrum data of 7 kinds (30 groups of each water sample) collected from Huainan mining area were used for experiment, and differences of laser-induced fluorescence spectra of mixed water samples with different volumetric ratios of old-kiln water, limestone water, limestone water and air water were analyzed. The classification accuracy of PSO-SVC model obtained by hold-out and Kennard-Stone partitioning was compared, and the training set water samples (140 groups) and test set water samples (70 groups) obtained by hold-out were used as experimental samples. Secondly, the full spectrum bands were divided into 10~25 bands by using the iPLS algorithm, and the band whose RMSECV (cross validation root mean square error) value is less than RMSECV value (threshold) of full spectrum bands was selected as the characteristic wave bands, and the results of modeling with 10 and 14 sub intervals were compared with spectrogram. It is found that there were errors in the characteristic bands selected by direct observation and the iPLS algorithm. Finally, under the condition of no pretreatment such as denoising and dimension reduction, the statistical data of evaluating indexes for dividing different interval numbers according to iPLS were obtained, and the data of 410.078~478.424 and 545.078~674.104 nm characteristic wave bands with 561 wavelength points selected from 11 regions were used as the input of PSO-SVC model. we compared with full spectrum bands and direct observation bands, and the classification accuracy of the training set and the test set was as high as 100%. The optimal penalty coefficient C of PSO is 1.367, and the kernel function parameter g is 0.576 2. It can be seen from the experimental results that it is feasible to select the characteristic wave bands of the fluorescence spectrum by using iPLS, and the extracted characteristic wave bands can fully reflect the effective information of the full spectrum bands, and it provides a theoretical basis for the application of laser induced fluorescence spectroscopy in the accurate on-line identification of mine water inrush source.

**Keywords** Mine water inrush; Laser-induced fluorescence; Interval PLS; Characteristic wave bands; Support vector classification

(Received Sep. 9, 2018; accepted Feb. 6, 2019)

\* Corresponding author