

改进 BP 神经网络算法对煤矿水源的分类研究

闫鹏程^{1,2}, 尚松行^{2*}, 张超银², 张孝飞²

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 煤矿安全对煤炭工业的健康持续发展至关重要, 而煤矿水灾又是煤矿事故的重大隐患, 因此煤矿水源数据的处理对于预防矿井突水事故具有重要意义。实验在激光器的辅助下利用激光诱导荧光技术获取 7 种水源的数据信息, 设定激光发射功率为 100 mW, 向被测水体发射波长 405 nm 激光, 获取实验水样 210 组的荧光光谱数据, 为了剔除光谱在采集过程受到的荧光背景、检测器噪声以及功率波动等影响, 利用 SG 平滑、多元散射矫正(MSC)预处理对数据进行降噪以及提高光谱特异性, 由于初始数据运算量过大并对数据压缩、消除冗余和数据噪音, 利用主成分分析(PCA)分别对 7 种水样进行建模降维处理, 从而得到小数据并且保持原有信息的数据特征。为了识别煤矿水源的突水类型, 对于降维后的数据利用粒子群算法(PSO)优化 BP 神经网络, PSO 算法通过对新粒子的适应度值和个体极值、群体极值适应度值的比较更新个体极值和群体极值的位置, 将最优初始权值和阈值赋予 BP 神经网络, 从而对待测水样的种类进行预测分析。普通的 PSO 优化 BP 神经网络, 容易出现早熟收敛, 故在改进的 PSO 算法中引入变异因子来提高模型寻找更优解的可能性。实验证明: SG, MSC 以及 Original 三种预处理方式中, SG 算法表现良好, 提高了模型的相关性。在 SG 预处理的前提下, BP 的决定系数 R^2 为 0.984 5, 平均相对误差 MRE 7.39%, 均方根误差为 7.25%; PSO-BP 的决定系数 R^2 为 0.999 8, 平均相对误差 MRE 0.17%, 均方根误差 0.08%; IPSO-BP 的决定系数 R^2 达到 0.999 9, 平均相对误差 MRE 和均方根误差 RMSE 皆为 0.01%。结果表明: 经 SG 预处理过后的光谱数据, 比 MSC 预处理效果更精确, 改进的粒子群优化算法更适用于该实验的矿井水源分类。

关键词 激光诱导荧光技术; 预处理; 改进的粒子群优化算法

中图分类号: O657.3 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2021)07-2288-06

引言

煤矿在生产建设中会遇到各种安全问题^[1-2], 根据国家煤矿安全监察局事故查询系统公布的煤矿安全数据, 仅 2020 年上半年全国煤矿企业共发生事故 48 起, 死亡 74 人。减少矿井事故的发生对人民的生命安全和国家的生产建设具有极其重要的意义^[3]。矿井事故主要有瓦斯事故、顶板事故、机电事故、放炮事故、水灾事故、火灾事故、运输事故以及其他事故^[4-6], 其中尤以矿井水害事故最为严重。矿井水害事故一般发生在采煤工作面, 当采煤工作面遇到地下暗河或者富水性强的含水层等, 就会出现水源的大规模涌出, 进入采

煤巷道^[7-8], 而一旦发生矿井突水事故, 轻则淹没采煤工作面, 摧毁煤矿开采工具, 重则导致矿井损毁, 造成人员财产的重大损失。

传统对于煤矿水源种类的判别主要是利用水化学的方法^[9-11], 利用水源中各种离子浓度的不同对水源进行分类, 比如代表离子法就是利用水源中 7 种元素的离子浓度, 通过对比已分类的含水层水质来辨识水源类型^[12-14]。但是由于水化学方法实验的特殊性, 操作复杂且数据不稳定, 因而提出一种利用激光诱导荧光的方法获取水源的荧光光谱, 再通过 IPSO 算法(改进的粒子群优化算法)优化 BP 神经网络模型来预测分类煤矿水源类型。

收稿日期: 2020-12-30, 修订日期: 2021-03-26

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 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: 13329078709@163.com

1 实验部分

实验以对矿井水源分类为目的, 首先收集样本并获取光谱数据, 对繁杂的数据进行 PCA 降维等预处理, 提取处理后数据的光谱特征并进行神经网络的建模分类。具体过程如图 1 所示。

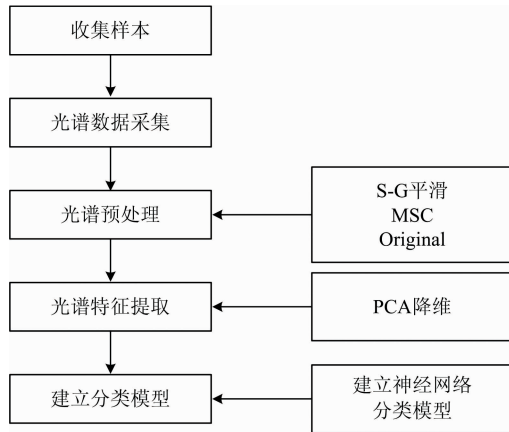


图 1 矿井水源预测分类模型流程图

Fig. 1 Flow chart of mine water source prediction and classification model

1.1 材料准备

由于煤矿水害水源类型多为老空水, 实验选择新庄孜煤矿采集到的岩溶水与老空水以及按两者按不同比例混合作为实验样本, 混合比例分别为 10 : 3, 10 : 6, 10 : 10, 6 : 10, 3 : 10, 每种样本取 30 份, 共计 210 个样本, 皆避光存储。

1.2 光谱采集

待测样本经 405 nm 入射光激发光谱, 设定激光功率 100 mW, 积分时间设定为 1/1 000 nm, 采用浸入式荧光探头(广州标旗光电 FPB-405-V3), 直接将其放入待测水体获取待测光谱信息, 光谱采集与接收是由 OceanView 软件记录, 整个实验过程均在暗室中进行。

1.3 光谱预处理

OceanView 软件收集的样本原始光谱图如图 2 所示。光波长检测范围是 400~800 nm, 光谱的变化范围主要集中在光波长为 400~700 和 700~800 nm 之间光谱强度趋于稳定。

由于原始荧光光谱受到仪器噪声以及功率波动等因素的影响, 采集的干扰信息会影响实验结果的准确度, 因此采用预处理对光谱数据进行定量定性分析。采用多元散射矫正(MSC)和 SG 平滑对原始光谱数据进行预处理, 可以有效减少光谱噪声, 提高光谱的分辨能力, 增强模型的性能。

(1) SG 平滑算法

由于是利用激光激发待测水源的荧光光谱信息, 而光谱的噪声属于白噪声, 可以利用 SG 平滑算法经过多次测量光谱信息数据求平均值来降低其自身所携带的随机误差、提高光谱的信噪比, 经过平滑之后可以有效地减少光谱噪声对实验数据的影响。

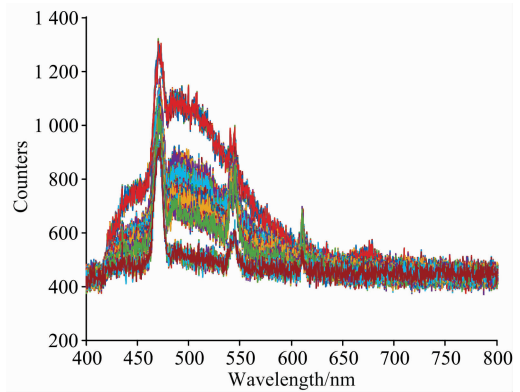


图 2 原始光谱图

Fig. 2 Original spectrum

(2) 多元散射矫正 MSC

多元散射矫正的目的是消除由于光程变化或者待测物质分布不均匀而对红外光漫反射光谱的影响。MSC 算法主要是通过一组实验样本的光谱信息, 基于统计的方法来修正光谱因为散射所带来的线性变化, 消除样本间的基线平移和漂移现象, 增强光谱特异性。

对实验采集的原始光谱作 MSC 预处理: 首先选取平均光谱作为标准光谱; 用原始光谱和标准光谱进行一元线性回归, 得出截距 b_i 和系数 k_i ; 然后对每个样本进行校正见式(1)~式(3)。

$$\bar{A} = \frac{1}{n} \sum_{i=1}^n A_i \quad (1)$$

$$A_i = k_i \bar{A} + b_i \quad (2)$$

$$A_{i(MSC)} = \frac{A_i - b_i}{k_i} \quad (3)$$

各预处理图像如图 3(a, b, c)所示。

1.4 PCA 降维

对所选取的 210 个样本进行 PCA 降维, 主成分数取 7, 其主要成分贡献率如图 4 所示。仅 7 种主成分的累计贡献率就达到了 95%。

在第一主成分与第二主成分作对比时, 水样的聚类效果明显, 并且模型的决定系数达到 0.9, 如图 5 所示。

1.5 PSO-BP 神经网络预测模型

1.5.1 BP 神经网络参数选取

BP 神经网络隐含层节点数 m 的计算公式如下

$$h < n - 1 \quad (4)$$

$$h < \sqrt{(m+n)} + a \quad (5)$$

$$h = \log_2 n \quad (6)$$

式(4)~式(6)中, h 为隐含层节点数, m 为输出层节点数, n 为输入层节点数, $a=1, 2, \dots, 10$ 是隐含层之间的调节常数。

a 取值的不同会影响神经网络的训练误差, 在隐含层神经元个数可以随意调整的前提下, 将隐含层神经元个数分别设为 5~15, 创建神经网络模型并进行网络训练, 得出不同隐含层节点的训练误差, 如表 1 所示。在经过训练后, 隐含层节点数为 13 时, 训练误差最小, 因此在本次矿井水源分类 BP 神经网络预测实验中隐含层神经元个数为 13。

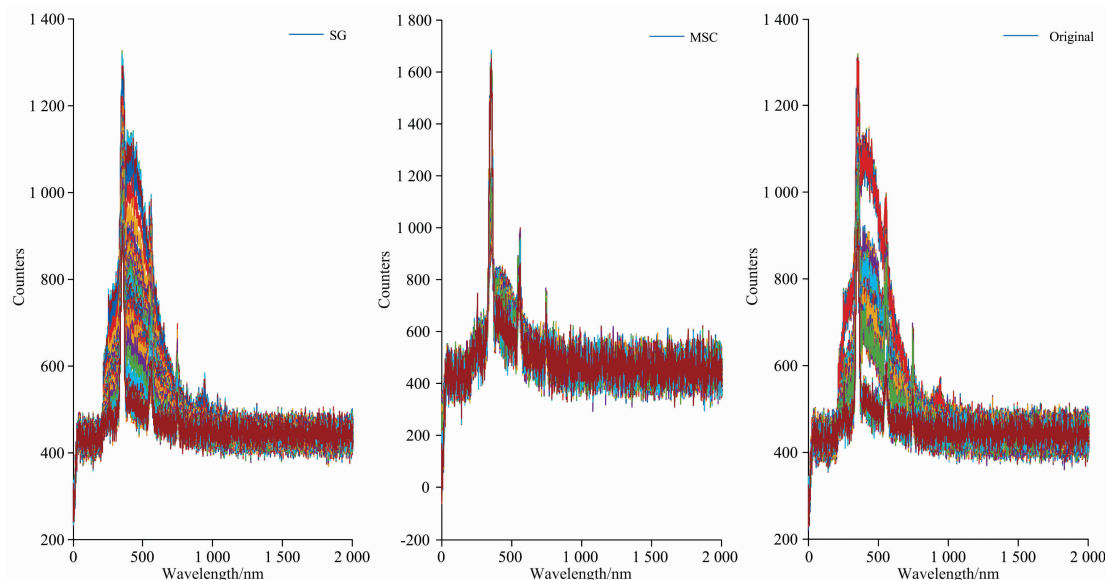


图 3 原始光谱图及预处理后的图像

(a): SG; (b): MSC; (c): Original 预处理

Fig. 3 Original spectrogram and preprocessed image

(a): SG; (b): MSC; (c): Original preprocessed

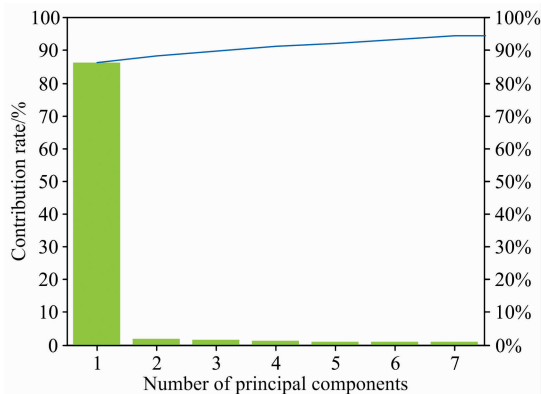


图 4 主成分累积贡献率

Fig. 4 Cumulative contribution rate of principal components

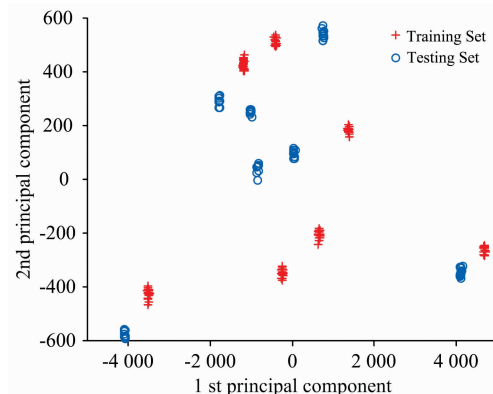


图 5 得分分布图

Fig. 5 Score distribution

表 1 不同隐含层神经元个数的训练误差

Table 1 Training error of neuron number in different hidden layers

隐节点数	5	6	7	8	9	10	11	12	13	14	15
训练误差	1.914	1.338	4.066	5.230	1.036	0.039	0.038	0.039	0.037	0.039	0.039

1.5.2 PSO 算法优化模型

PSO 算法即粒子群优化算法，是一种源于对鸟类捕食的研究并用于求解优化问题的优化算法。PSO 算法用位置、速度和适应度值表示该粒子特征，通过追踪个体极值 Pbest 和群体极值 Gbest 来确定最优位置。

假设有一个 D 维的目标搜索空间，有 n 个粒子组成的种群 $Z=(X_1, X_2, \dots, X_n)$ ，种群 Z 中有某一 i 个粒子可以表示为一个 D 维向量 $X_i=[x_{i1}, x_{i2}, \dots, x_{iD}]^T$ ，代表某一问题

在 D 维空间的一个潜在的解向量，根据目标函数确定每个 X_i 粒子位置对应的适应度值。第 i 个粒子的速度为 $V_i=[V_{i1}, V_{i2}, \dots, V_{iD}]^T$ ，个体极值为 $P_i=[P_{i1}, P_{i2}, \dots, P_{iD}]^T$ ，种群的群体极值为 $P_g=[P_{g1}, P_{g2}, \dots, P_{gD}]^T$ 。

在迭代过程中，粒子的速度和位置的更新通过式(7)和式(8)得到

$$V_{id}^{k+1} = \omega V_{id}^k + c_1 r_1 (P_{id}^k - X_{id}^k) + c_2 r_2 (P_{gd}^k - X_{id}^k) \quad (7)$$

$$X_{id}^{k+1} = X_{id}^k + V_{id}^{k+1} \quad (8)$$

式中 ω 为惯性权重; $d=1, 2, \dots, D$; $i=1, 2, \dots, n$; k 为当前迭代次数; V_{id} 是粒子速度; c_1 和 c_2 是加速常数; r_1 和 r_2 是 $[0, 1]$ 之间的随机数。

1.5.3 建立 PSO-BP 预测模型

PSO-BP 预测模型的基本步骤主要包括:

(1) BP 神经网络选择单隐层网络拓扑结构, 矿井水源种类的影响因素作为网络输入, 水源种类为网络输出, 根据调节常数 a 确定最优隐层节点数。

(2) 对 PSO 算法粒子和速度进行初始化处理, 确定粒子种群 n , 粒子速度 v 以及种群的维度 D 。

(3) 将降维以及预处理后的数据集划分为训练集和测试集, 导入 PSO-BP 预测模型。

(4) 运行 PSO-BP 模型, 根据适应度函数计算粒子适应度值, 通过追踪个体 Pbest 和群体极值 Gbest 更新个体位置, 设置收敛精度, 使 Gbest 不断逼近此精度。

(5) 将在收敛精度范围内的 Gbest 赋值给 BP 神经网络作为网络的权值和阈值。

(6) 根据训练组训练的神经网络, 并以此作为标准对测试集进行预测处理并检验结果。

1.5.4 IPSO-BP 模型的建立

PSO-BP 算法模型的优点是具有快速的收敛性和很强的通用性, 但正是由于这一优点容易存在早熟收敛、搜索精度偏低、迭代效率不高等缺点, 为了使算法跳出局部最优, 本实验改进了 PSO-BP 算法模型, 联想到遗传算法的变异思想, 改进的 PSO-BP 算法在种群更新和优化选择的基础上引入了自适应变异操作, 在普通 PSO-BP 算法中粒子每次更新之后添加变异算子, 对某些变量以一定概率进行初始化处理, 使粒子可以跳出先前的最优值位置, 在更大空间内开展搜索, 提高算法寻找到更优值的可能性。

2 结果与讨论

实验中 7 种比例的水样样本, 分别编号 1—7, 每种样本包含 30 组实验数据, 将每种样本的前 20 组数据作为训练集进行模型演算, 后 10 组数据作为测试集, 共计 140 组训练数据, 70 组测试数据。在经过 MSC 与 SG 预处理算法后, 经 PCA 进行降维, 将降维之后的数据代入各预测模型中并进行预测值和真实值作比较, 如图 6 所示, 决定系数最小的是采用 MSC 预处理的 PSO-BP 预测模型, 最大的是采用 SG 预处理的 IPSO-BP 预测模型, 其决定系数无限接近 1。

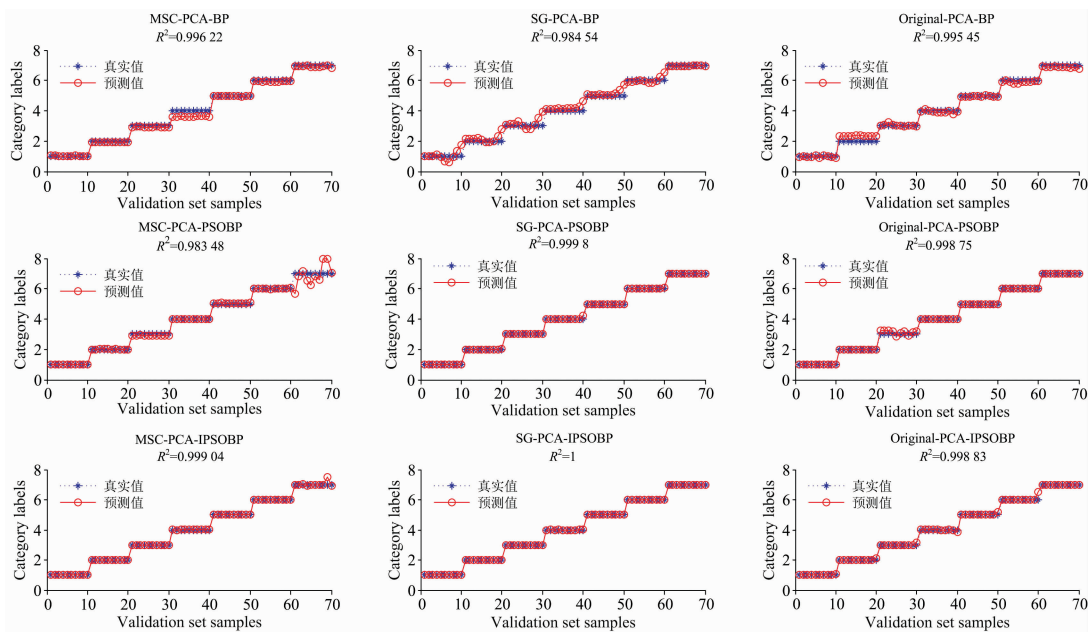


图 6 九种不同预处理方法下的各预测模型真实值与预测值对比

Fig. 6 Comparison between the real value and the predicted value of each prediction model under nine different pretreatment methods

根据表 2 的各分类模型的评价指标对比可知, 在 BP 神经网络中, 三种预处理均表现出大体上进行分类, 但是具体分类出现较大偏差; PSO-BP 分类模型中, 三种预处理预测性能较好, MSC 预处理在 60—70 号样本中发生错判, 而 SG 预处理在对每组样本的最后一个样本分类时会发生较小差值, Original 预处理在 20—30 号样本之中会发生明显错判; IPSO-BP 预测模型中, MSC 预处理在第 69 号样本预测分类时发生较大错判; SG 预处理分类结果与真实值完全相符,

表现良好; Original 预处理方式下, 在 20—60 的每组的最后一个样本发生错判现象。因此在三种预处理方式下的 BP, PSO-BP 以及 IPSO-BP 预测分类模型中, SG-IPSO-BP 分类性能良好。

预测效果最佳的是 SG-IPSO-BP 算法预测分类模型, 其平均相对误差 MRE 和均方根误差 RMSE 皆是 0.01%, 预测错误率无限接近于 0。实验中 9 种预测模型效果最差的是 SG-BP 模型, 其决定系数为 0.9845, 平均相对误差 MRE 为

7.39%，与同种预处理方式下的 IPSO-BP 相差 7.38%，均方根误差 RMSE 为 7.25%，二者相差 7.24%。BP 神经网络模型经过 SG, MSC 和 Original 三种预处理，其中经过 Original 处理过后效果最好；在 PSO-BP 预测模型种经过 SG 预处理算法预测效果最佳。

表 2 不同预处理方法下的各预测模型评价指标对比

Table 2 Comparison of evaluation indexes of each prediction model under different pretreatment methods

预处理方式	评价指标	BP 模型	PSO-BP 模型	IPSO-BP 模型
SG 预处理	R^2	0.984 5	0.999 8	0.999 9
	MRE	0.073 9	0.001 7	0.000 1
	RMSE	0.072 5	0.000 8	0.000 1
MSC 预处理	R^2	0.996 2	0.983 5	0.999 0
	MRE	0.028 8	0.019 7	0.002 3
	RMSE	0.023 5	0.066 9	0.003 9
Original	R^2	0.995 5	0.998 8	0.998 8
	MRE	0.044 1	0.009 1	0.004 2
	RMSE	0.024 9	0.005 4	0.004 9

如图 7 所示为 BP, PSO-BP, IPSO-BP 分别对经 MSC, SG 和 Original 预处理之后数据做出的预测值和真实值的绝对误差图，误差幅度的剧烈变化主要集中在 0—20 之间的样本，也就是编号为 1 和 2 的样本预测误差变化幅度较大，其

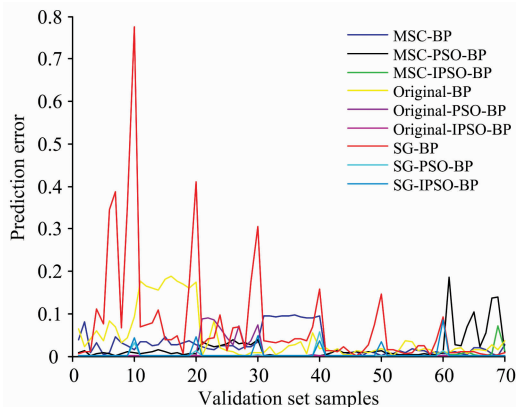


图 7 各模型预测值与真实值的绝对误差图

Fig. 7 Absolute error diagram of predicted value and true value of each model

中 SG-BP 的误差幅度远远大于其他几种预测分类模型，该模型的绝对误差最大值为 0.775 1，最小值为 0.001；绝对误差最小的模型为 SG-IPSO-BP 预测模型，预测的绝对误差最大值为 0.084 7，最小值是 0 误差。

三种分类模型的适应度曲线如图 8 所示，最终得到的最优个体适应度值为 IPSO-BP 的 0.047 0 要优于 PSO-BP 以及 BP 预测分类模型。

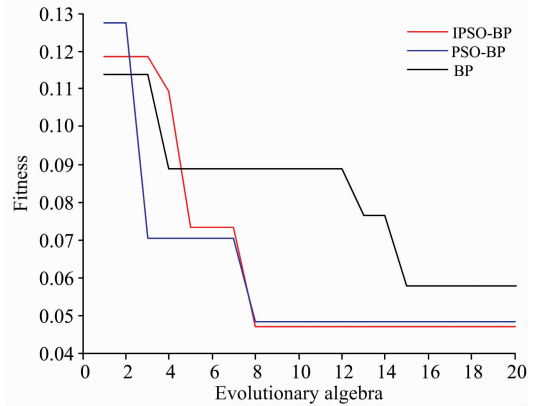


图 8 适应度曲线

Fig. 8 Fitness curves

综上所述，从预测值和实际值的对比图以及各模型误差结果分析的可以看出 SG-BP 神经网络的预测分类能力最差，SG-IPSO-BP 的预测值与真实值完全相符，分类性能较好，对水源种类的预测分类模型中，IPSO-BP 要优于 PSO-BP 与 BP 神经网络，而 PSO-BP 的神经网络预测模型要优于 BP 神经网络。

3 结 论

对于 BP, PSO-BP 和 IPSO-BP 算法预测分类模型在 SG, MSC 以及 Original 三种预处理方式中，SG 被选为最适宜对 7 种水样进行预处理的算法，结合 PCA 降维，对比同种预处理方式下的不同分类预测模型发现 SG-IPSO-BP 分类性能最佳。在本次实验中，SG 预处理可以很好的提高光谱的平滑性，并降低噪声的干扰；经过 PCA 降维，去除冗余信息，减少计算量；在 PSO 算法模型的基础上提出一种变异因子的改进算法，增加了模型寻找最优解的可能性，对水源在激光诱导荧光的实验环境下提供了一种新的分类方法。

References

- [1] ZHU Yun-fei, WANG De-ming, QI Xu-yao, et al(朱云飞, 王德明, 戚绪尧, 等). Safety in Coal Mines (煤矿安全), 2018, 49(10): 241.
- [2] PENG Su-ping(彭苏萍). Journal of China Coal Society(煤炭学报), 2020, 45(7): 2331.
- [3] JIN De-wu(靳德武). Coal Science and Technology(煤炭科学技术), 2017, 45(5): 141.
- [4] FENG Chang-gen, WANG Ya-jun(冯长根, 王亚军). Journal of Safety and Environment(安全与环境学报), 2007, (5): 146.
- [5] JIANG Xing-xing, LI Chun-xiang(蒋星星, 李春香). Coal Engineering(煤炭工程), 2019, 51(1): 101.
- [6] XU Lei, LI Xi-jian(徐磊, 李希建). Safety in Coal Mines(煤矿安全), 2018, 49(3): 98.
- [7] SHENG Chao(盛超). Coal Engineering(煤炭工程), 2018, 50(11): 6.
- [8] YIN Shang-xian, LIAN Hui-qing, LIU De-min, et al(尹尚先, 连会青, 刘德民, 等). Coal Science and Technology(煤炭科学技术),

- 2019, 47(11): 1.
- [9] LI Yong-jun, LI Chen, SUN Hao, et al(李永军, 李琛, 孙浩, 等). Coal Technology(煤炭技术), 2016, 35(7): 88.
- [10] ZHANG Shu-ying, HU You-biao, JU Qi-ding(张淑莹, 胡友彪, 嵇棋定). Mining Safety & Environmental Protection(矿业安全与环保), 2018, 45(6): 53.
- [11] CAO Hai-dong(曹海东). Safety in Coal Mines(煤矿安全), 2015, 46(S1): 16.
- [12] JI Zhuo-chen, DING Xiang, HOU En-ke, et al(纪卓辰, 丁湘, 侯恩科, 等). Coal Geology & Exploration(煤田地质与勘探), 2020, 48(5): 97.
- [13] JIANG Zi-hao, HU You-biao, JU Qi-ding, et al(姜子豪, 胡友彪, 嵇棋定, 等). Industry and Mine Automation(工矿自动化), 2020, 46(4): 28.
- [14] QU Xing-yue, QIU Mei, SHI Long-qing(曲兴玥, 邱梅, 施龙青). Coal Technology(煤炭技术), 2018, 37(12): 137.

Classification of Coal Mine Water Sources by Improved BP Neural Network Algorithm

YAN Peng-cheng^{1,2}, SHANG Song-hang^{2*}, ZHANG Chao-yin², ZHANG Xiao-fei²

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 Coal mine safety is very important to the healthy and sustainable development of the coal industry, and the coal mine flood is a major hidden danger of coal mine accidents. Therefore, coal mine water source data processing is of great significance to prevent mine water inrush accidents. In this experiment, the laser-induced fluorescence technology was used to obtain the data information of 7 water sources. The laser power was set at 100 mW, 405 nm laser was emitted to the measured water, and 210 groups of fluorescence spectrum data of experimental water samples were obtained. In order to eliminate the influence of fluorescence background, detector noise and power fluctuation, SG smoothing and multiplicative scatter correction (MSC) preprocessing is used to reduce the noise and improve the spectral specificity of the data. Due to a large amount of initial data operation, data compression, redundancy elimination and data noise elimination, principal components analysis (PCA) is used to analyze the seven water samples. Row modeling and dimensionality reduction are used to obtain small data and keep the original data characteristics. In order to identify the water inrush type of coal mine water source, particle swarm optimization (PSO) is used to optimize BP neural network for dimension reduced data. PSO algorithm updates the position of individual extremum and population extremum by comparing the fitness value of new particle with that of individual extremum and population extremum, PSO algorithm updates the position of individual extremum and population extremum by comparing the fitness value of new particle with that of individual extremum and population extremum, and endows the optimal initial weight and threshold value to BP neural network, so as to predict and analyze the types of water samples to be measured. The common PSO optimized BP neural network is prone to premature convergence, so mutation factor is introduced into the improved PSO algorithm to improve the possibility of finding a better solution. Experimental results show that the SG algorithm performs well among SG, MSC, and original preprocessing methods and improves the correlation of models. On the premise of SG pretreatment, the determination coefficient R^2 of BP is 0.984 5, the mean relative error MRE is 7.39%, and the root mean square error is 7.25%; the determination coefficient R^2 of PSO-BP is 0.999 8, the mean relative error MRE is 0.17%, the root mean square error is 0.08%; the determination coefficient R^2 of IPSO-BP is 0.999 9, the MRE and RMSE are 0.01%. The results show that the spectral data preprocessed by SG is more accurate than that by MSC, and the improved particle swarm optimization algorithm is more suitable for mine water source classification in this experiment.

Keywords Laser-induced fluorescence technology; Pretreatment; Improved particle swarm optimization algorithm

(Received Dec. 30, 2020; accepted Mar. 26, 2021)

* Corresponding author