

林朋,孙成,任珂,等. 基于 PSO-XGBoost 的煤层断层智能识别方法研究[J]. 矿业科学学报, 2025, 10(1): 57-69. DOI: 10.19606/j.cnki.jmst.2024933

LIN Peng, SUN Cheng, REN Ke, et al. Research on intelligent fault identification method of coalfield based on the PSO-XGBoost algorithm[J]. Journal of Mining Science and Technology, 2025, 10(1): 57-69. DOI: 10.19606/j.cnki.jmst.2024933

# 基于 PSO-XGBoost 的煤层断层智能识别方法研究

林朋<sup>1</sup>, 孙成<sup>1</sup>, 任珂<sup>1,2</sup>, 刘育林<sup>1</sup>, 李阳<sup>1</sup>

1. 中国矿业大学(北京)地球科学与测绘工程学院, 北京 100083;
2. 空天信息大学(筹)遥感科学与技术学院, 山东济南 250299

**摘要:**为进一步提高地下断层识别准确率和解释效率,使用极限梯度提升树(XGBoost)机器学习算法对煤层断层进行智能识别,并结合粒子群算法(PSO)优化模型相关参数,构建基于 PSO-XGBoost 的断层构造识别模型。建立正演模型对 PSO-XGBoost 模型进行检验,并基于滇东矿区采集的实际数据对比分析 PSO-XGBoost 模型与 PSO-RF、PSO-SVM 模型的分类预测性能,选择准确率和对数损失值作为评价分类器预测模型的主要指标评价各模型的准确度。结果表明,基于 PSO-XGBoost 的模型在断层构造识别中展现出较高的准确率和更好的稳定性。

**关键词:**断层识别; XGBoost; PSO; 机器学习; 参数优化

中图分类号: TD 163.1

文献标志码: A

文章编号: 2096-2193(2025)01-0057-13

## Research on intelligent fault identification method of coalfield based on the PSO-XGBoost algorithm

LIN Peng<sup>1</sup>, SUN Cheng<sup>1</sup>, REN Ke<sup>1,2</sup>, LIU Yulin<sup>1</sup>, LI Yang<sup>1</sup>

1. College of Geosciences and Surveying Engineering, China University of Mining and Technology-Beijing, Beijing 100083, China;
2. School of Remote Sensing Science and Technology, Aerospace Information Technology University, Jinan Shandong 250299, China

**Abstract:** In order to further improve the accuracy and efficiency of underground fault identification, an intelligent fault recognition model based on the extreme gradient boosting tree (XGBoost) machine learning algorithm was constructed for coal seam faults, combined with the particle swarm optimization (PSO) algorithm to optimize the model's related parameters. A forward model was established to verify the PSO-XGBoost model, and the classification prediction performance of the PSO-XGBoost model was compared with that of the PSO-RF and PSO-SVM models based on actual data collected from the Dian-dong mining area. The accuracy rate and log loss value were selected as the main evaluation indicators to evaluate the accuracy of the classification prediction models for each model. The results show that the PSO-XGBoost model has a high accuracy in fault structure identification; the PSO-XGBoost model has higher accuracy and better stability in fault identification.

**Key words:** fault recognition; XGBoost; PSO; machine learning; parameters optimization

煤炭资源的高效安全开采一直备受关注。地下断层改变了煤岩层的埋藏条件<sup>[1]</sup>,增加了煤炭

的开采难度和成本<sup>[2]</sup>,断层位置处还容易发生矿井水源突水<sup>[3]</sup>、顶板透水<sup>[4]</sup>、冒顶<sup>[5]</sup>等矿山事

收稿日期: 2024-07-04 修回日期: 2024-10-20

基金项目: 国家重点研发计划(2023YFC3008902)

作者简介: 林朋(1990—),男,安徽阜阳人,博士,讲师,主要从事高分辨率成像与人工智能探测等方面的研究工作。E-mail: linpeng798@126.com

故<sup>[6]</sup>。因此,断层的准确识别和预测对于煤矿勘探与开发具有重要意义。

地震勘探是一种常用的煤矿地质构造探测方法,在煤田勘探中扮演至关重要的角色<sup>[7]</sup>。地质工作者不断尝试从地震数据中获取更多的信息用于提高断层解释的准确率<sup>[8]</sup>,解决构造解释存在的问题,但对于小断层<sup>[9]</sup>以及构造整体解释的效果仍有较大的可提升空间。当前,通过将先进勘探技术与煤田三维地震勘探进行结合,已逐步建立了针对煤矿问题的高分辨率三维地震勘探体系<sup>[10-13]</sup>。目前,断层解释人员可以通过多维度信息分析断层在地层中的分布情况,不仅降低了煤矿潜在风险、节约时间成本,同时还提升了预测准确性。

随着人工智能的兴起与发展,机器学习技术在地震断层构造识别中的应用也逐渐得到了关注<sup>[14-16]</sup>。任庆国等<sup>[17]</sup>提出 PSO-SVM-BP 预测模型对断层破碎带围岩变形进行预测,实践证明了模型的预测精度较高。杜鹏等<sup>[18]</sup>采用随机森林模型(RF)对藏东南地区的滑坡易发性进行了评价。韩成阳等<sup>[19]</sup>基于 SVM 对夏店煤矿断层进行了预测。罗家举<sup>[20]</sup>基于 UNet++ 网络结构模型实现了对断层分布的智能识别。但是上述方法都存在着不足,例如 SVM 在处理大规模数据时训练时间较长,易存在过拟合等问题;RF 的整体模型不易理解,不如单决策树直观等。杨楚龙等<sup>[21]</sup>利用 XGBoost 算法对新元煤矿东翼矿区地震地质构造进行识别,结果表明 XGBoost 算法模型的预测精确度为 0.95,虽有效控制了过拟合问题,但是在参数调优上仍有提高模型性能的空间。温后珍等<sup>[22]</sup>基于贝叶斯优化后的 XGBoost 模型实现了高维数据集的超参数快速优化,提高了石灰窑气的预测精度。然而,贝叶斯算法容易陷入局部最优解,并容易发生过拟合问题,同样有一定的提升空间。赵耀忠等<sup>[23]</sup>依赖遗传算法对 XGBoost 模型进行优化,实现了高精度煤矿安全预警,但不能避免收敛速度较慢等问题。

针对断层解释过程中效率低、主观性强的问题,以及当前主流机器学习模型存在的不足,采用极限梯度提升树(eXtreme Gradient Boosting, XGBoost)方法对断层构造进行预测识别,并使用 PSO 算法<sup>[24]</sup>对 XGBoost 模型中的超参数进行优化,提出 PSO-XGBoost 预测模型。基于矿区实际数据,对比分析 XGBoost、RF<sup>[25]</sup>和 SVM<sup>[26]</sup>3 种模型的预测结果,验证所提模型的可行性。

## 1 PSO-XGBoost 原理

### 1.1 XGBoost 算法原理

XGBoost 算法<sup>[27]</sup>是由梯度提升树(Gradient Boosting Tree, GBDT)算法发展而来,将多个弱分类器(决策树)组合成强分类器,弥补了 GBDT 在可扩展性和效率方面的不足<sup>[28]</sup>。此外, XGBoost 算法在目标函数式(1)中引入了正则化项,可以防止过拟合。

$$O_{bj} = \sum_{i=1}^m l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (1)$$

式(2)为多个基学习器的组合来预测样本的输出。

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), f_k \in F \quad (2)$$

上两式中,  $l(y_i, \hat{y}_i)$  为损失函数;  $y_i$  为真实值;  $\hat{y}_i$  为分类预测值;  $\Omega(f_k)$  为正则项;  $F$  为回归树空间;  $f_k$  为第  $k$  棵回归树;  $x_i$  为数集中第  $i$  棵回归树的数据样本;  $K$  为全部回归树个数。

通过目标函数进行误差计算。通过迭代的方式进行训练,每次迭代后添加新的回归树,  $t$  次迭代后:

$$\hat{y}_i^{(t)} = \sum_{j=1}^t f_j(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i) \quad (3)$$

式中,  $\hat{y}_i^{(t)}$  为第  $t$  次迭代后的预测值。

将式(3)代入式(1),得到第  $t$  次迭代后的目标函数:

$$O_{bj}^{(t)} = \sum_{i=1}^m l[y_i, \hat{y}_i^{(t-1)} + f_t(x_i)] + \Omega(f_k) + \sigma \quad (4)$$

将式(4)进行泰勒二阶展开,同时引入正则项  $\Omega(f_k)$  来减小过拟合情况的发生:

$$\begin{cases} O_{bj}^{(t)} \cong \sum_{i=1}^m [\partial_{y_i^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)}) f_t(x_i) + \frac{1}{2} \partial_{y_i^{(t-1)}}^2 l(y_i, \hat{y}_i^{(t-1)}) f_t^2(x_i)] + \Omega(f_k) + \sigma \\ \Omega(f_k) = \gamma T + \frac{1}{2} \lambda \|\omega\|^2 \end{cases} \quad (5)$$

式中,  $\gamma$  为叶子树惩罚系数;  $T$  为树叶子节点数目;  $\omega$  为叶子权重;  $\lambda$  为权重的惩罚系数;  $\sigma$  为与观测值  $y_i$  相关的标准差或不确定性。

XGBoost 的参数分为常规参数、增强参数(Boosting Parameters)和学习控制参数(Learning Task Parameters)。常规参数包括学习率(eta)、树的数量(n\_estimators)、最大深度(max\_depth)、最小样本分割数(min\_child\_weight)等,这些参数影响模型复杂度和学习能力的平衡;增强参数包括子

采样率 (subsample)、列采样率 (colsample\_bytree), 用于防止模型过拟合;学习控制参数包括目标函数 (objective)、评估指标 (eval\_metric), 与具体任务相关。

### 1.2 粒子群算法 (PSO)

XGBoost 超参数选取不当可能会导致模型出现过拟合或欠拟合情况,选择 PSO 算法对超参数进行优化。

PSO 算法步骤为:

(1) 设定粒子个体的位置和速度,并设置适应度函数评估其优劣。

(2) 基于粒子的历史最优位置和全局最优位置,并权衡粒子的个体和社会经验的权重,不断更新粒子的速度和位置并进行择优。

(3) 计算择优后相应粒子最终位置的适应度,并且同步更新。迭代过程中,粒子会不断重复上述步骤,直到达到终止条件(达到最大迭代次数或找到满意的解)。

### 1.3 PSO-XGBoost 模型及评价指标

#### 1.3.1 PSO-XGBoost 模型断层识别流程

PSO-XGBoost 模型断层识别<sup>[29]</sup>的流程如图 1 所示。

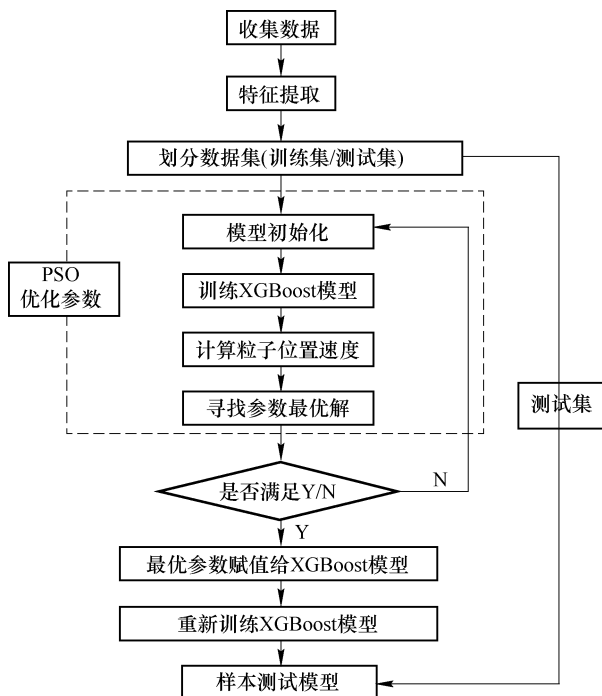


图 1 PSO-XGBoost 模型断层识别流程

Fig. 1 Flow chart of PSO-XGBoost model fault identification

(1) 建立正演模型,对正演模型采集得到的数据做基本预处理工作。

(2) 按照 7 : 3 的比例划分训练集和测试集,选择适合的适应度函数并初始化粒子个体最优值

和全局最优值。

(3) 更新优化粒子速度与位置,通过相关公式计算更新个体的适应度值,直到单体获得的最优为全局最优值时,迭代终止(或达到最大迭代次数)。

(4) 将最优参数应用到 XGBoost 分类模型进行训练集的训练。

#### 1.3.2 评价指标

对于分类预测问题,一般采用以下几项指标:

(1) 准确率 (Accuracy),是分类模型中的正确预测分类样本的比例,准确率越高越好。

$$Accuracy = (TP+TN)/(TP+TN+FP+FN) \quad (6)$$

式中,TP 为真阳性;TN 为真阴性;FP 为假阳性;FN 为假阴性。

(2) 精确率 (Precision),是指分类模型正确判断正类的概率。

$$Precision = TP/(TP+FP) \quad (7)$$

(3) 召回率 (Recall),是指分类模型正确判断正类的能力,是全部正类样本中被分类器正确判断为正类的概率。

$$Recall = TP/(TP+FN) \quad (8)$$

(4) F1 分数 (F1-score),主要是精确率和召回率的调和平均数,可以综合这两个指标的表现。

$$F1\text{-score} = 2 \times precision \times recall / (precision + recall) \quad (9)$$

采用准确率和精确率以及 F1-score 这 3 个指标进行评估,评价指标混淆矩阵见表 1。

表 1 混淆矩阵评价指标

Table 1 Chaos matrix of evaluation indicators

样本标签	预测非断层(0)	预测断层(1)
非断层(0)	真阴性(TN)	假阴性(FN)
断层(1)	假阳性(FP)	真阳性(TP)

## 2 模型检验

### 2.1 正演模型构建

为更好地分析断层位置处引发的地震响应,建立了含断层合成地质模型,如图 2 所示。模型参考了煤矿实际断层的揭露情况,总共分为 3 层,上、下层为砂岩层,波速为 3 200 m/s,密度为 2.23 g/cm<sup>3</sup>;中间煤层埋深为 300 m,波速为 3 000 m/s,密度为 2.2 g/cm<sup>3</sup>,厚度为 10 m。共设置了 5 个断层,包括 4 个正断层和 1 个逆断层,中心点位置相差 30 m,断层方向一致,断层断距分别为 3、5、5、10、20 m。观测系统中炮点与检波点重合,间距均

为 1 m,震源采用主频为 60 Hz 的雷克子波,以垂直方式进行激发。

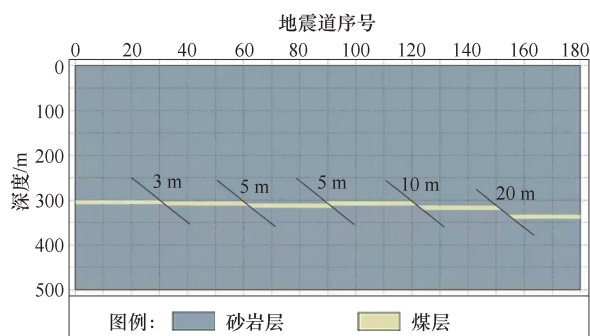


图 2 含断层合成地质模型

Fig. 2 Synthetic geologic model with faults

通过波场正演得到的零偏移距地震记录,如图 3 所示。地震记录结果与给出的正演模型结果一致,随着断层断距增加,对应的断层构造解释难度逐渐降低。

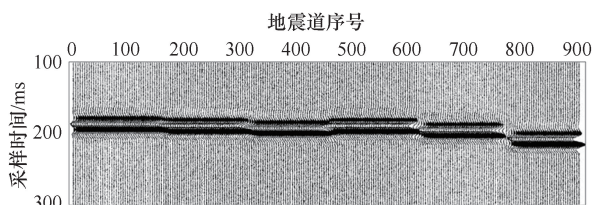


图 3 零偏移距地震记录

Fig. 3 Zero-offset seismic data

## 2.2 XGBoost 模型训练

应用传统的 XGBoost 模型进行断层分类器模型的训练,对正演剖面提取 7 种断层的地震属性:方差体、瞬时相位、瞬时频率、反射强度、混沌体、均方根振幅、最大振幅。将属性数值与断层信息进行匹配,总共得到 900 组样本点。其中,断层点 108 个,非断层点 792 个,按照 7 : 3 进行训练集和测试集的划分。模型训练所需参数根据以往训练经验进行设定,XGBoost 分类器参数取值见表 2。

表 2 XGBoost 分类器参数

Table 2 XGBoost classifier parameters

参数	参数值
学习率(learning_rate)	0.1
弱分类器(n_estimators)	100
最大树深(max_depth)	3
最小下降值(gamma)	0.01
随机采样比例(subsample)	0.5
随机采用列数占比(colsample_bytree)	0.3

将数据进行预处理后,对 XGBoost 分类预测模型进行训练,按照表 2 中的参数值进行 XGBoost 模型的学习。训练过程中,XGBoost 模型的准确率和损失函数随迭代次数的变化如图 4、图 5 所示。由图 4 可知,整个训练过程进行了 1 000 次迭代,训练准确率高达 99%。由图 5 可知,相应的损失函数下降速度较慢,模型预测效率较低。XGBoost 模型在训练集上准确率为 99%,精确率为 94%,召回率为 90%,f1-score 为 87%。

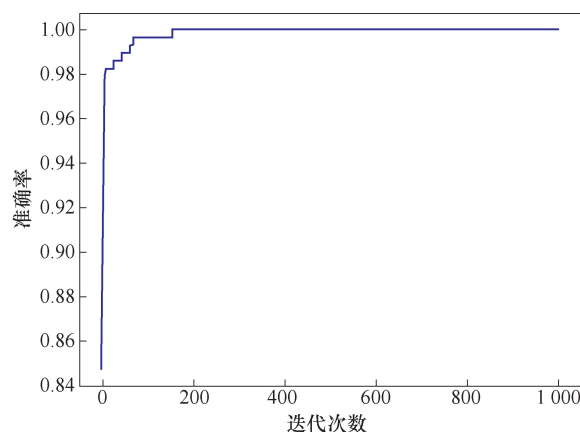


图 4 XGBoost 模型训练准确率随迭代次数变化

Fig. 4 Accuracy change of the XGBoost model with the number of iterations

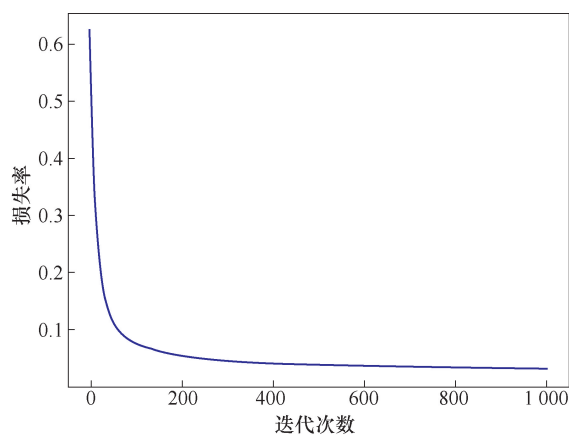


图 5 XGBoost 模型损失函数随迭代次数变化

Fig. 5 Loss function change of the XGBoost model with the number of iterations

## 2.3 PSO-XGBoost 模型训练验证

应用 PSO 算法对 XGBoost 分类模型的参数进行寻优,需要优化的参数有:learning\_rate、n\_estimators、max\_depth、gamma、subsample 和 colsample\_bytree。通过记录比较训练准确率和对数损失值来判断参数是否为最优结果,PSO 算法寻优后最佳参数值见表 3。

表 3 PSO 算法寻优后最优参数

Table 3 Optimal parameters after PSO optimization

参数	参数值
learning_rate	0.14
n_estimators	302
max_depth	2
gamma	0.03
subsample	0.732
colsample_bytree	0.332

使用 PSO 寻优后的参数,基于相同的正演数据对 PSO-XGBoost 模型进行训练,相应的准确率和损失函数随迭代次数的变化如图 6、图 7 所示。可以看出,PSO-XGBoost 模型预测准确率和损失函数值在迭代 19 次后均趋于收敛状态,相对于采

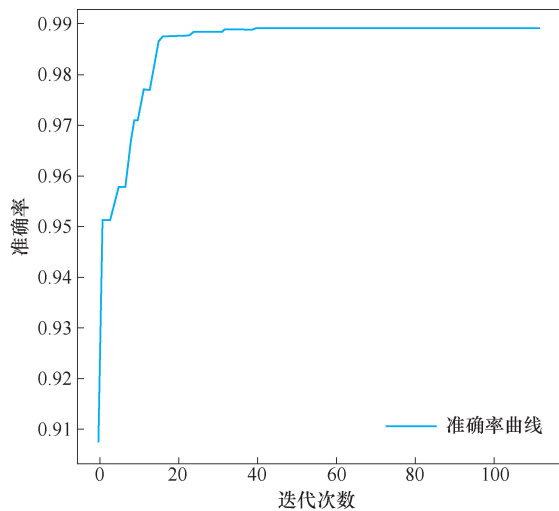


图 6 PSO-XGBoost 模型准确率随迭代次数变化

Fig. 6 Accuracy change of the PSO-XGBoost model with the number of iterations

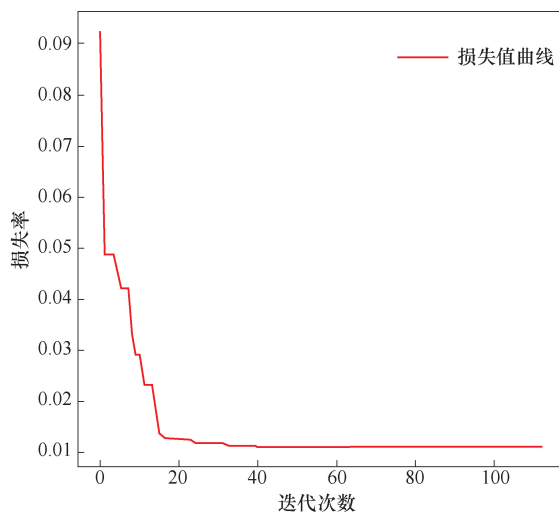


图 7 PSO-XGBoost 模型损失函数随迭代次数变化

Fig. 7 Loss function change of the PSO-XGBoost model with the number of iterations

用直接经验参数的 XGBoost 模型迭代速度明显提高;经过 100 次的迭代计算后,PSO-XGBoost 训练集结果准确率为 99.39%,精确率为 95%,召回率为 100%,f1-score 为 97.44%。应用 PSO 算法对模型参数进行优选,可以更快找到最佳的参数值,缩短模型迭代次数,提高预测效率。

采用最优参数的 PSO-XGBoost 预测模型在相应测试集上的预测结果如图 8 所示。由图 8 可知,PSO-XGBoost 模型分类性能良好,该模型在测试集上的分类准确率达到 92.59%,精确率达到 88.33%,召回率达到 93.33%,f1-score 为 92.98%。

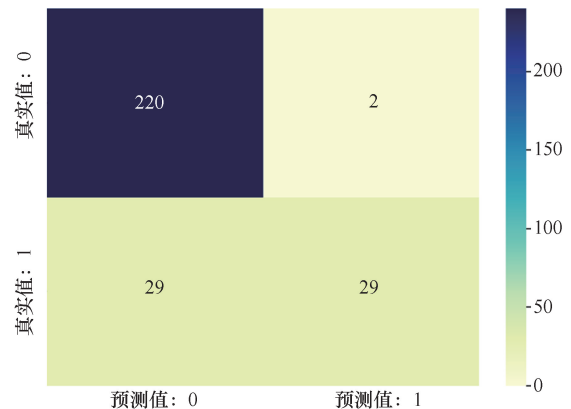


图 8 PSO-XGBoost 模型测试集预测结果

Fig. 8 Results of the test dataset using the PSO-XGBoost model

PSO-XGBoost 模型的断层预测结果如图 9 所示。PSO-XGBoost 分类模型总共预测出 6 个断层,在第一个样本处出现了预测误差,错误地预测了断层构造;正演模型中所设置的 5 个断层都被准确识别,说明该模型具有判别断层断距的能力,预测结果与正演模型所设置的情况基本一致。

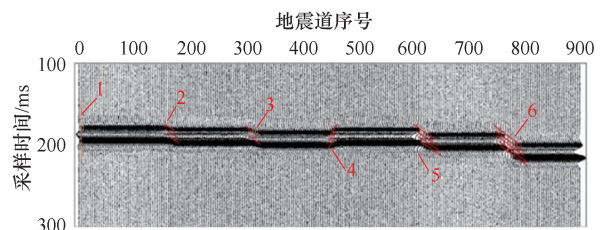


图 9 PSO-XGBoost 模型断层预测结果

Fig. 9 Fault prediction results using the PSO-XGBoost model

### 3 实际应用

#### 3.1 矿区背景介绍

滇东矿区位于中国东南部的山区地带,以该矿区玉旺井田<sup>[30]</sup>为研究背景,玉旺井田位于富源县老房煤矿区第四勘探区的西南部,受区域构造影

响,研究区内地质结构较为复杂,断层构造较为发育,对煤炭等矿产资源开采的挑战较大。

针对该研究区的复杂地质情况,围绕建设透明矿山、智慧矿井的要求<sup>[31-32]</sup>,滇东矿区开展了三维地震精细勘探。基于高分辨率三维地震数据体<sup>[33]</sup>(图10),利用精细地震解释方法提取有效地震地质信息,使煤矿的复杂地质信息透明化,促进煤矿的安全高效生产。

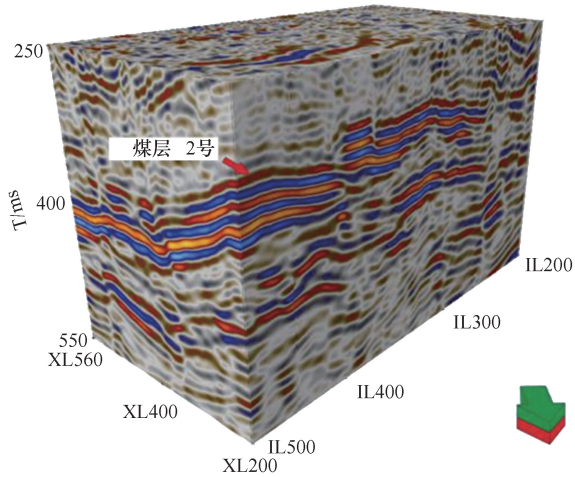


图10 地下煤层分布<sup>[33]</sup>

Fig. 10 Coal Seam Distribution Area<sup>[33]</sup>

研究区内主要可采煤层深度约200 m,煤巷走向最长超1 500 m。目标煤层中揭露的断层以正断层为主,少量逆断层,采集了11口井、4条巷道和18条揭露断层的信息(图11)。揭露断层的落差、倾角和延展长度等信息见表4。

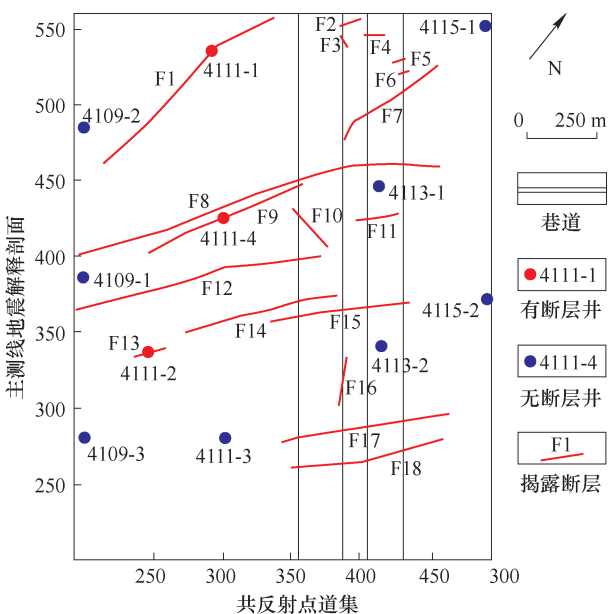


图11 研究区井、巷道及揭露断层分布情况<sup>[33]</sup>

Fig. 11 Distribution of wells, roadways and exposed faults in the study area<sup>[33]</sup>

### 3.2 数据集构建与预测

统计工区内共有106 500个数据点,由于数据量庞大,研究降低了地震属性样本数据的维数,以减少信息重叠和噪声对数据本身的影响。选用10 ms的时窗沿煤层提取地震属性信息,使用与正演模拟部分相同的7种地震属性(方差体、瞬时相位、瞬时频率、反射强度、混沌体、均方根振幅、最大振幅)作为特征,构建研究区的属性数据集。

表4 研究区揭露断层信息

Table 4 Exposed fault information in research area

断层编号	落差 /m	倾角 /( $^{\circ}$ )	长度 /m	断层编号	落差 /m	倾角 /( $^{\circ}$ )	长度 /m
1	0~10	75~85	825	10	0~5	75~80	183
2	0~3	75~85	82	11	0~5	75~85	156
3	0~5	75~80	46	12	0~40	75~85	914
4	0~3	75~85	79	13	0~3	75~80	115
5	0~5	75~85	54	14	0~15	55~65	564
6	0~5	75~85	43	15	0~12	75~85	513
7	0~5	75~85	444	16	0~5	75~80	171
8	0~60	75~80	1 375	17	0~10	50~55	619
9	0~25	75~80	611	18	0~30	50~55	562

同时,统计钻井、巷道处揭露的断层信息,“1”表示该位置存在断层,“0”表示不存在断层,将这些数据点的坐标位置与其属性值进行匹配,构建已知标签的数据集。数据经过处理后,选择断层特征明显位置的测点,筛选后共有325个数据点,其中揭露断层有200个,非断层有125个(图12)。

根据获得的实际数据,按8:2的比例划分训练集和测试集,采用PSO-XGBoost、PSO-SVM和PSO-RF 3个模型对数据标签进行预测分类,将预测结果与标记的断层进行比对,判断各模型的准确率。其中,为对比PSO-XGBoost与PSO-SVM、PSO-RF模型的性能,只对learning\_rate, n\_estimators以及max\_depth 3个主要参数进行寻优。

设置寻优参数为learning\_rate = 0.18, n\_estimators = 203, max\_depth = 2。PSO-XGBoost模型在训练集中的准确率和损失函数随迭代次数而变化如图13所示。由图13可知,PSO-XGBoost模型在训练集的准确率最高达到93.7%,在第27次迭代时收敛,损失函数值最低为0.061。

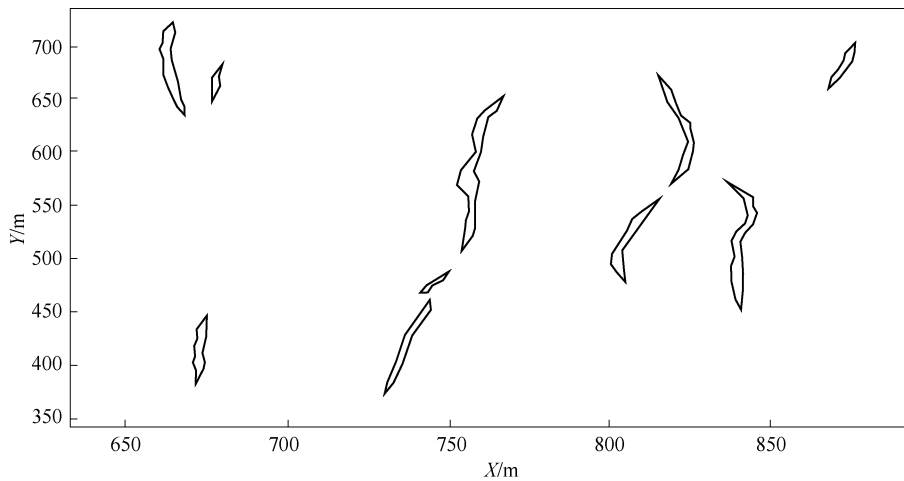


图 12 揭露的断层分布

Fig. 12 Exposed fault distribution

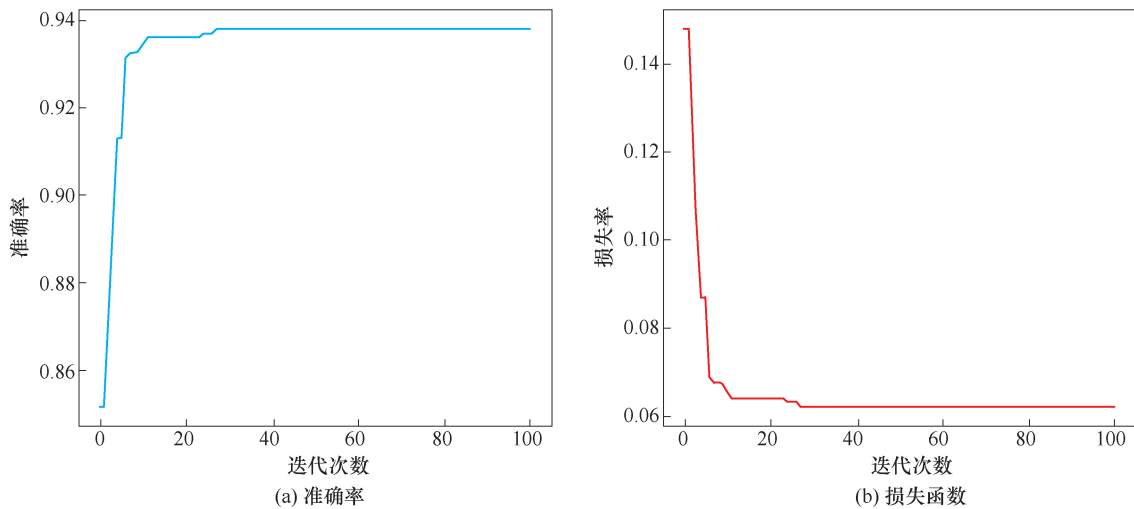


图 13 PSO-XGBoost 迭代变化

Fig. 13 Accuracy and loss change of the PSO-XGBoost model with the number of iterations

结合图 14 测试集的结果分析,可知 PSO-XG-Boost 模型的整体预测结果较好。将训练后的 PSO-XGBoost 模型用于预测地区断层,揭露断层预测结果如图 15 所示。可以看出,PSO-XGBoost 模型整体断层预测的准确率较高,在 800~850 m 之间的揭露断层预测准确率相对较低,推测可能是由于该段断层分布较密集,预测出现偏差。

### 3.3 预测结果对比

基于相同数据集,采用 PSO-SVM 和 PSO-RF 两种预测模型进行分类识别,与 PSO-XGBoost 模型进行对比。

PSO-RF 模型在第 23 次迭代时收敛,如图 16 所示,模型优化参数取值为  $n\_estimators = 21$ ,  $max\_depth = 4$ 。训练集最大准确率 87.6%。

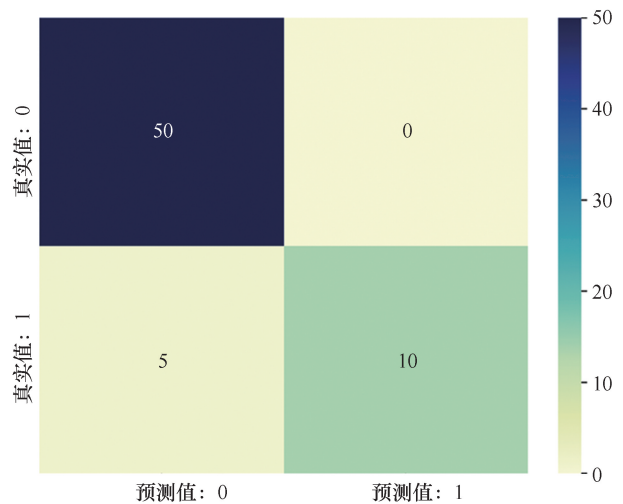


图 14 PSO-XGBoost 测试集结果

Fig. 14 Results of the PSO-XGBoost test

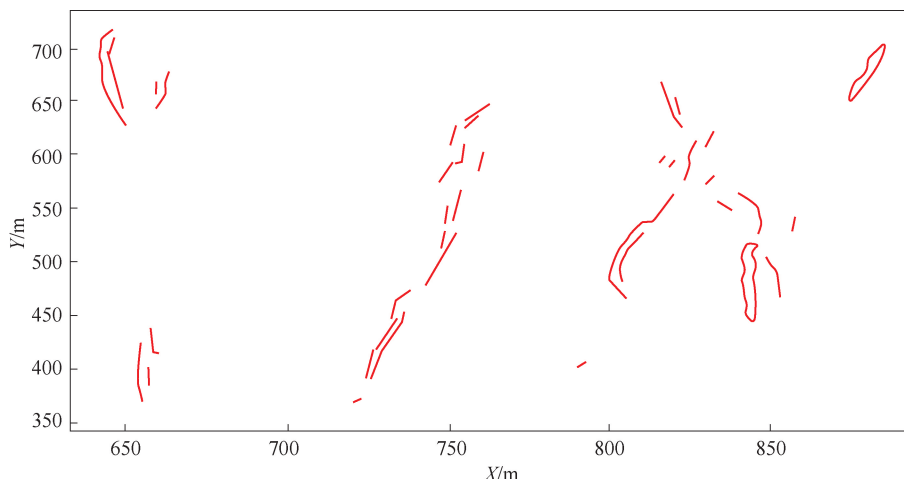


图 15 揭露断层预测 (PSO-XGBoost)

Fig. 15 Prediction of exposed fault (PSO-XGBoost)

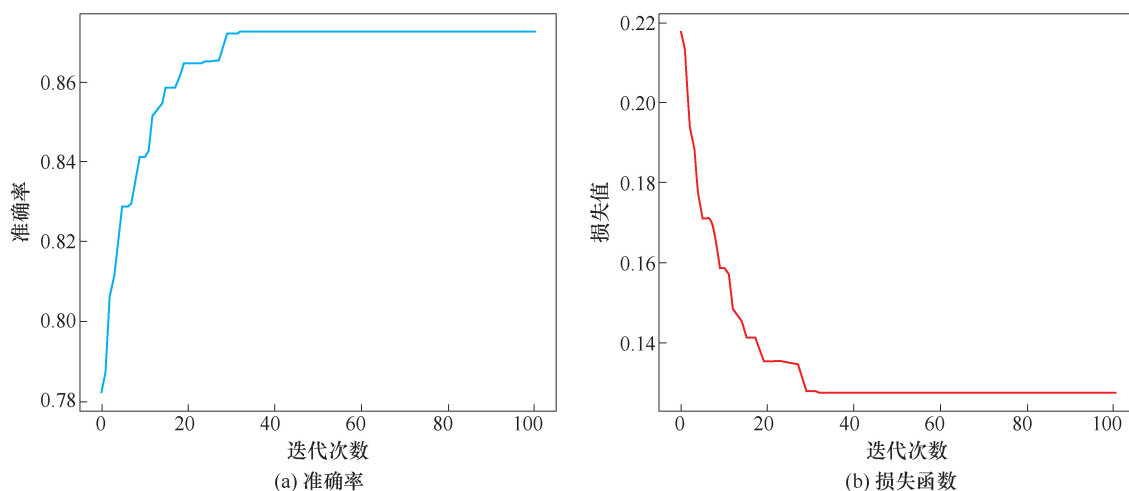


图 16 PSO-RF 迭代变化

Fig. 16 Accuracy and loss change of the PSO-RF model with the number of iterations

PSO-RF 测试集结果 (图 17) 的准确率达到 84.6%。训练后的 PSO-RF 模型预测揭露断层, 预测结果如图 18 所示, 可见预测准确率较差, 存在多处揭露断层预测错误, 预测效果一般。

PSO-SVM 在训练集上的表现如图 19 所示, 最大准确率为 74.6%, 最小对数损失值 0.255。PSO-SVM 模型通过 13 次迭代得到最优参数:  $C = 43$ ,  $\gamma = 0.0001$ ,  $\text{degree} = 19$ 。该模型在训练集的表现较差, 准确率低, 但是在测试集的表现达到了 89.6%。

图 20 中 PSO-SVM 测试集结果表现较好, 但在训练过程中准确率欠佳, 推断训练中存在欠拟合, 模型鲁棒性较差或数据与该模型不适合, 训练集预测断层效果与测试集结果不一致, 预测结果不准确。预测模型在区域整体揭露断层的预测表现

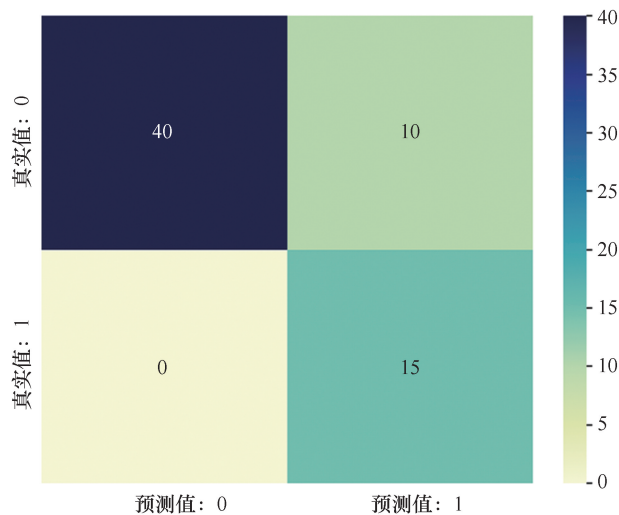


图 17 PSO-RF 测试集结果

Fig. 17 Results of the PSO-RF test

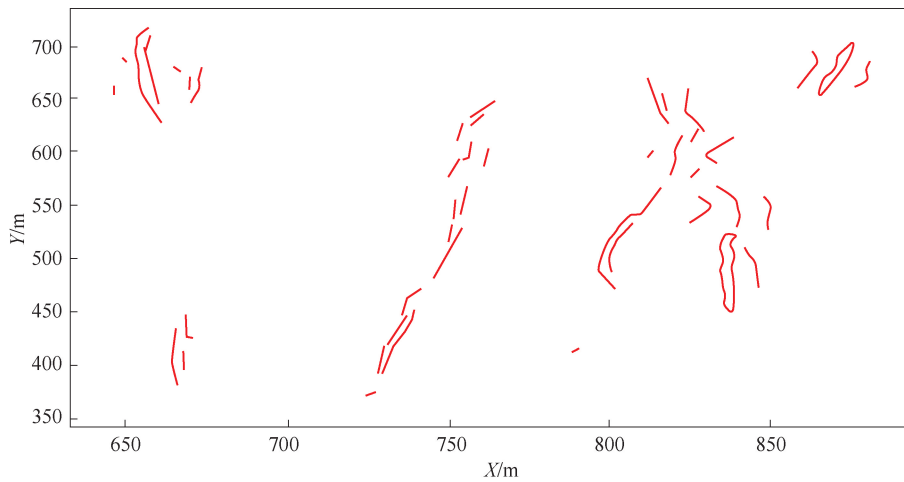


图 18 揭露断层预测 (PSO-RF)

Fig. 18 Prediction of exposed fault (PSO-RF)

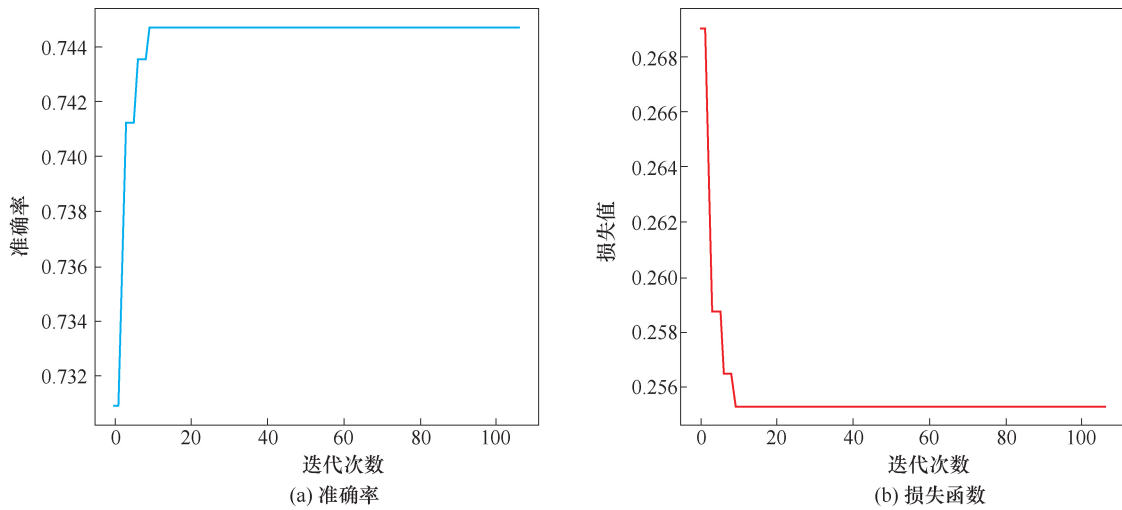


图 19 PSO-SVM 迭代变化

Fig. 19 Accuracy and loss change of the PSO-SVM model with the number of iterations

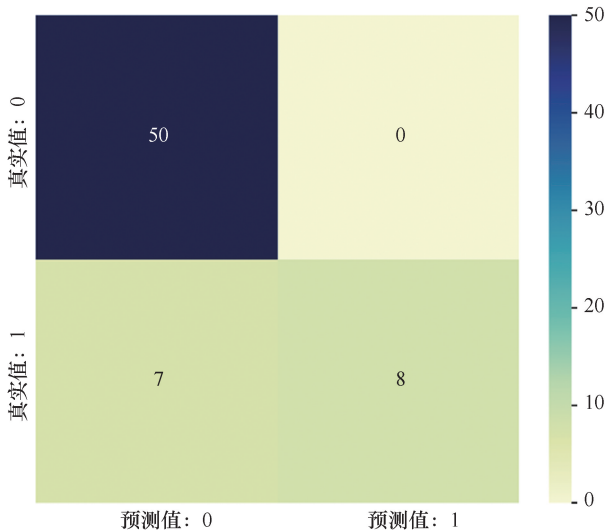


图 20 PSO-SVM 测试集结果

Fig. 20 Results of PSO-SVM test

如图 21 所示。由图 21 可知,PSO-SVM 预测模型仍然存在揭露断层预测错误情况,其效果略逊于 PSO-XGBoost 模型,且该模型测试集结果与整体揭露断层结果的预测有偏差。

表 5 不同模型结果分析

Table 5 Analysis of different model results

预测模型	训练集最大准确率/%	训练集对数损失值	测试集最大准确率/%
PSO-XGBoost	93.7	0.061	92.3
PSO-RF	87.6	0.135	84.6
PSO-SVM	74.6	0.255	89.6(训练集欠拟合)

PSO-XGBoost、PSO-SVM 和 PSO-RF3 种模型的 50 次交叉验证结果如图 22 所示,可以看出,PSO-XGBoost 模型的稳定性表现更好。

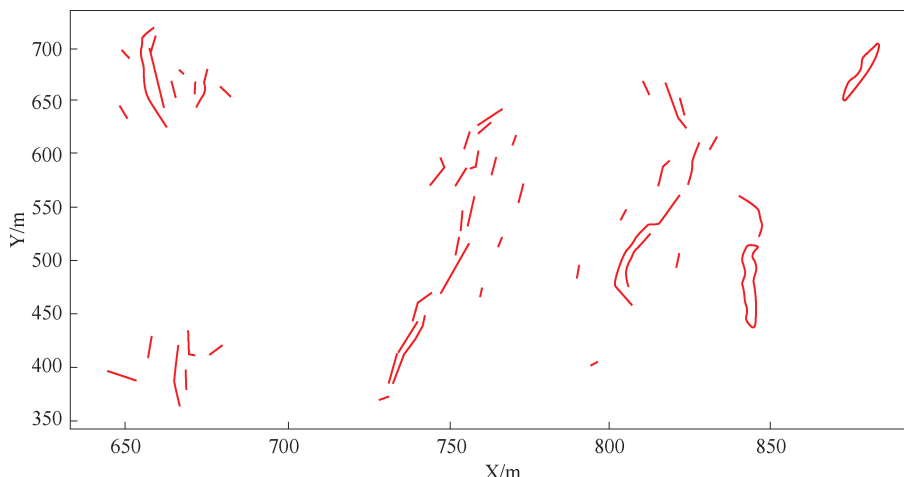


图 21 揭露断层预测 (PSO-SVM)

Fig. 21 Prediction of exposed fault ( PSO-SVM)

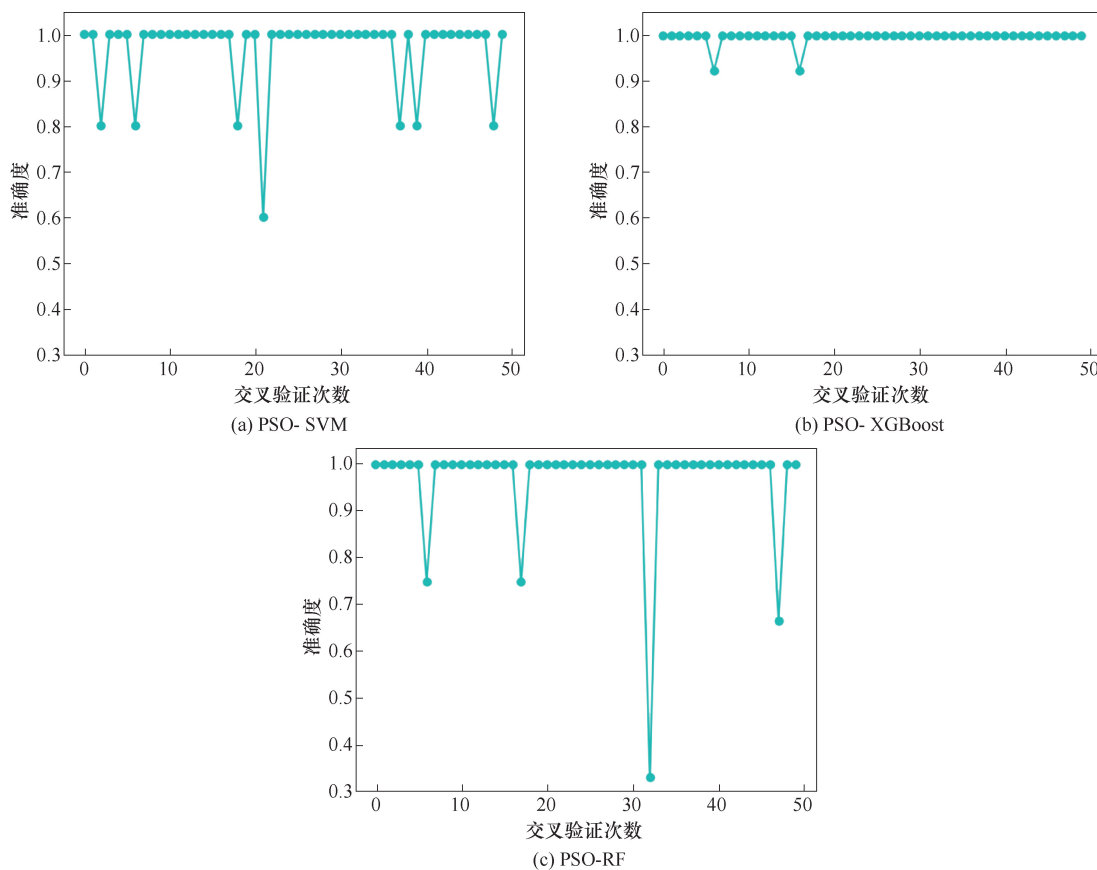


图 22 PSO-XGBoost、PSO-SVM 和 PSO-RF 三种模型的 50 次交叉验证结果

Fig. 22 Results of 50 cross validations of the PSO-XGBoost PSO-SVM and PSO-RF models

对比 3 种预测模型的预测, PSO-RF 和 PSO-SVM 的准确率均不理想, 分别出现了欠拟合与稳定性差的情况, PSO-XGBoost 在识别准确率方面表现良好。

#### 4 结论

针对地下断层识别问题, 结合 PSO 算法和

XGBoost 算法, 提出 PSO-XGBoost 断层智能识别模型, 并基于实际数据与传统的机器学习方法进行对比, 得到主要结论如下:

- (1) 采用 PSO 对 XGBoost 分类模型中的多参数进行寻优, 不仅可以显著提高 XGBoost 模型在断层识别方面的预测精度, 还能加快其收敛速度。
- (2) 基于正演模型所获得的数据对 PSO-XG-

Boost 模型进行检验,得到 PSO-XGBoost 分类预测模型准确率在 99% 以上,具有高效准确的预测效果和更强的泛化能力。

(3) 采用 PSO-XGBoost、PSO-SVM、PSO-RF 3 种模型对滇东矿区实际数据进行预测分类,得到判别准确率分别为 93.7%、87.6%、74.6%, PSO-XGBoost 预测模型在分类精度方面明显优于另外两种模型。

## 参考文献

- [1] 彭苏萍,孟召平,李玉林. 断层对顶板稳定性影响相似模拟试验研究[J]. 煤田地质与勘探, 2001, 29(3): 1-4.  
PENG Suping, MENG Zhaoping, LI Yulin. Influence of faults on coal roof stability by physical modeling study [J]. Coal Geology & Exploration, 2001, 29(3): 1-4.
- [2] 杜文凤,彭苏萍,师素珍. 深部隐伏构造特征地震解释及对煤矿安全的影响[J]. 煤炭学报, 2015, 40(3): 640-645.  
DU Wenfeng, PENG Suping, SHI Suzhen. Seismic interpretation of deep buried structure characteristics and its influence on coal mine safety [J]. Journal of China Coal Society, 2015, 40(3): 640-645.
- [3] 董东林,张陇强,张恩雨,等. 基于 PSO-XGBoost 的矿井突水水源快速判别模型[J]. 煤炭科学技术, 2023, 51(7): 72-82.  
DONG Donglin, ZHANG Longqiang, ZHANG Enyu, et al. A rapid identification model of mine water inrush based on PSO-XGBoost [J]. Coal Science and Technology, 2023, 51(7): 72-82.
- [4] 谷宇峰,张道勇,鲍志东. 测井资料 PSO-XGBoost 渗透率预测[J]. 石油地球物理勘探, 2021, 56(1): 26-37, 4-5.  
GU Yufeng, ZHANG Daoyong, BAO Zhidong. Permeability prediction using PSO-XGBoost based on logging data [J]. Oil Geophysical Prospecting, 2021, 56(1): 26-37, 4-5.
- [5] 赵惊涛,彭苏萍,陈宗南,等. 煤矿隐蔽致灾地质体地震绕射波探测方法[J]. 矿业科学学报, 2022, 7(1): 1-8.  
ZHAO Jingtiao, PENG Suping, CHEN Zongnan, et al. Seismic diffraction detection method for geological hidden disasters in coal mining [J]. Journal of Mining Science and Technology, 2022, 7(1): 1-8.
- [6] 陈阳,吕辉,毕莹,等. 国内外典型露天矿山坍塌事故案例分析与启示[J]. 中国煤炭, 2024, 50(4): 63-67.  
CHEN Yang, LV Hui, BI Ying, et al. Analysis and enlightenment of typical collapse accidents in open-pit mines at home and abroad [J]. China Coal, 2024, 50(4): 63-67.
- [7] 彭苏萍,袁亮. 淮南煤矿三维地震勘探技术应用与效果[J]. 安徽地质, 2011, 21(2): 95-99.  
PENG Suping, YUAN Liang. Research and achievements of three dimensional seismic prospecting technologies in Huainan coal mines [J]. Geology of Anhui, 2011, 21(2): 95-99.
- [8] 邵龙义,张亚星,耿苏倩,等. 煤矿固体废弃物物理化特征及生态环境影响研究[J]. 矿业科学学报, 2024, 9(5): 653-667.  
SHAO Longyi, ZHANG Yaxing, GENG Suqian, et al. Physicochemical characteristics and environmental impact of coal mine solid waste [J]. Journal of Mining Science and Technology, 2024, 9(5): 653-667.
- [9] 杨瑞召,王媛媛,王兴元,等. 煤田三维地震采区小断层解释方法及应用[J]. 矿业工程研究, 2010, 25(1): 12-15.  
YANG Ruizhao, WANG Yuanyuan, WANG Xingyuan, et al. Interpretation of small faults in coalfield 3D seismic survey: methods and application [J]. Mineral Engineering Research, 2010, 25(1): 12-15.
- [10] 蒋法文,黄晖,张振生,等. 高精度三维地震勘探技术在煤田安全生产中的应用[J]. 中国煤炭地质, 2014, 26(2): 60-64.  
JIANG Fawen, HUANG Hui, ZHANG Zhensheng, et al. Application of high precision 3D seismic prospecting technology in coalfield safety production [J]. Coal Geology of China, 2014, 26(2): 60-64.
- [11] 彭苏萍,杜文凤,赵伟,等. 煤田三维地震综合解释技术在复杂地质条件下的应用[J]. 岩石力学与工程学报, 2008, 27(S1): 2760-2765.  
PENG Suping, DU Wenfeng, ZHAO Wei, et al. Application of three-dimensional seismic comprehensive interpretation technology in coalfield in complex geological conditions [J]. Chinese Journal of Rock Mechanics and Engineering, 2008, 27(S1): 2760-2765.
- [12] 彭苏萍. 我国煤矿安全高效开采地质保障系统研究现状及展望[J]. 煤炭学报, 2020, 45(7): 2331-2345.  
PENG Suping. Current status and prospects of research on geological assurance system for coal mine safe and high efficient mining [J]. Journal of China Coal Society, 2020, 45(7): 2331-2345.
- [13] 彭苏萍,卢勇旭. 煤与瓦斯突出灾害隐患高分辨三维地震预测方法[J]. 煤矿安全, 2020, 51(10): 34-38.

- PENG Suping, LU Yongxu. High-resolution 3D seismic prediction method for hidden dangers of coal and gas outburst hazard [J]. *Safety in Coal Mines*, 2020, 51 (10): 34-38.
- [14] 杨佳润. 机器学习在断层和溶洞识别中的应用研究[D]. 北京: 中国石油大学(北京), 2021.
- YANG Jiarun. Application of machine learning in fault and Karst cave identification[D]. Beijing: China University of Petroleum (Beijing), 2021.
- [15] 王卫东, 张康辉, 吕子奇, 等. 基于深度学习的煤中异物机器视觉检测[J]. *矿业科学学报*, 2021, 6 (1): 115-123.
- WANG Weidong, ZHANG Kanghui, LÜ Ziqi, et al. Machine vision detection of foreign objects in coal using deep learning [J]. *Journal of Mining Science and Technology*, 2021, 6(1): 115-123.
- [16] 苗丙, 葛世荣, 郭一楠, 等. 煤矿数字孪生智采工作面系统构建[J]. *矿业科学学报*, 2022, 7(2): 143-153.
- MIAO Bing, GE Shirong, GUO Yinan, et al. Construction of digital twin system for intelligent mining in coal mines[J]. *Journal of Mining Science and Technology*, 2022, 7(2): 143-153.
- [17] 任庆国, 苗兰弟. 支持向量机模型在断层破碎带围岩变形预测中的应用[J]. *河北工业科技*, 2017, 34 (3): 194-201.
- REN Qingguo, MIAO Landi. Application of support vector machine model in fracture zone surrounding rock deformation prediction[J]. *Hebei Journal of Industrial Science and Technology*, 2017, 34(3): 194-201.
- [18] 杜鹏, 陈宁生, 伍康林, 等. 基于随机森林模型的藏东南地区滑坡易发性评价及主控因素分析[J]. *成都理工大学学报: 自然科学版*, 2024, 51(2): 328-344.
- DU Peng, CHEN Ningsheng, WU Kanglin, et al. Evaluation of landslide susceptibility in southeast Tibet based on a random forest model[J]. *Journal of Chengdu University of Technology: Science & Technology Edition*, 2024, 51(2): 328-344.
- [19] 韩成阳, 邹冠贵, 薛静雯. 夏店煤矿断层的支持向量机预测研究[J]. *地质论评*, 2023, 69(S1): 119-120.
- HAN Chengyang, ZOU Guangui, XUE Jingwen. Study on fault prediction of Xiadian coal mine by SVM[J]. *Geological Review*, 2023, 69(S1): 119-120.
- [20] 罗家举. 基于卷积神经网络深度学习的地质断层智能识别方法[J]. *现代矿业*, 2024, 40(8): 7-10.
- LUO Jiaju. Intelligent recognition method of geological fault based on deep learning of convolutional neural network[J]. *Modern Mining*, 2024, 40(8): 7-10.
- [21] 杨楚龙, 王怀秀, 刘最亮. 基于改进 XGBoost 的地质多属性地质构造识别方法[J]. *科学技术与工程*, 2023, 23(29): 12442-12450.
- YANG Chulong, WANG Huaixiu, LIU Zuiliang. Seismic multi-attribute geological structure identification method based on improved XGBoost[J]. *Science Technology and Engineering*, 2023, 23(29): 12442-12450.
- [22] 温后珍, 栾仪广, 孟碧霞, 等. 基于贝叶斯优化 XGBoost 的石灰窑气预测[J]. *北京化工大学学报: 自然科学版*, 2024, 51(3): 114-121.
- WEN Houzhen, LUAN Yiguang, MENG Bixia, et al. Bayesian optimization based XGBoost for lime kiln gas prediction[J]. *Journal of Beijing University of Chemical Technology: Natural Science Edition*, 2024, 51(3): 114-121.
- [23] 赵耀忠, 吴涛, 杨永军, 等. 基于遗传算法优化 XGBoost 的煤矿安全预警方法研究[J]. *电子设计工程*, 2024, 32(16): 170-173.
- ZHAO Yaozhong, WU Tao, YANG Yongjun, et al. Research on coal early warning method of mine safety based on genetic algorithm optimized XGBoost[J]. *Electronic Design Engineering*, 2024, 32(16): 170-173.
- [24] 杨艺, 赵惊涛, 付国强. 基于 PSO-LSTM 模型的地热储层温度预测研究[J]. *矿业科学学报*, 2024, 9 (4): 538-548.
- YANG Yi, ZHAO Jingtao, FU Guoqiang. Predicting geothermal reservoir temperature based on the PSO-LSTM model[J]. *Journal of Mining Science and Technology*, 2024, 9(4): 538-548.
- [25] Leo Breiman, Random Forests [J]. *Journal of Machine Learning*, 2001, 45, 5-32.
- [26] HEARST M A, DUMAIS S T, OSUNA E, et al. Support vector machines [J]. *IEEE Intelligent Systems and Their Applications*, 1998, 13(4): 18-28.
- [27] CHEN T Q, GUESTRIN C. XGBoost: a scalable tree boosting system [J]. *ArXiv e-Prints*, 2016: arXiv: 1603. 02754.
- [28] 田野, 闵锦涛. 基于 PSO-XGBoost 算法的多衰退特征锂离子电池 SOH 估计[J]. *电工材料*, 2023(1): 23-27.
- TIAN Ye, MIN Jintao. SOH prediction of lithium ion battery with multiple degradation characteristics based on PSO-XGBoost algorithm[J]. *Electrical Engineering Materials*, 2023(1): 23-27.
- [29] YIN P Y, TSAI C C, DAY R F. PSO active learning of XGBoost and spatiotemporal data for PM2.5 sensor calibration[J]. *IOP Conference Series: Earth and En-*

- vironmental Science, 2019, 227: 052048.
- [30] 车玉燕, 邹冠贵, 殷裁云, 等. 雨汪井田超压含煤地层孔隙特征及孔隙体积模量演变规律[J]. 矿业科学学报, 2022, 7(3): 275-285.
- CHE Yuyan, ZOU Guangui, YIN Caiyun, et al. The evolution law of pore characteristics and pore bulk modulus for overpressured coal-bearing formation in Yuwang Minefield [J]. Journal of Mining Science and Technology, 2022, 7(3): 275-285.
- [31] 陈柏平, 崔凡, 刘波, 等. 基于地质统计学反演的透明化矿山岩性建模参数研究及应用[J]. 矿业科学学报, 2022, 7(4): 427-436.
- CHEN Baiping, CUI Fan, LIU Bo, et al. Research and application of inversion parameters based on geological statistics inversion in transparent mines rock major modeling[J]. Journal of Mining Science and Technology, 2022, 7(4): 427-436.
- [32] 谭章禄, 王美君. 智能化煤矿数据治理概念模型及技术架构研究[J]. 矿业科学学报, 2023, 8(2): 242-255.
- TAN Zhanglu, WANG Meijun. Research on the conceptual model and technical architecture of data governance for intelligent coal mine [J]. Journal of Mining Science and Technology, 2023, 8(2): 242-255.
- [33] 任珂. 基于支持向量机的煤田断层智能化解释——以滇东矿区为例[D]. 北京: 中国矿业大学(北京), 2024.
- REN Ke. Intelligent interpretation of coalfield faults based on support vector machines - a case study of dongdong mining area [D]. Beijing: China University of Mining and Technology-Beijing, 2024.

(责任编辑: 张彩艳)