

# 基于特征选择与机器学习的煤与瓦斯突出危险等级协同预测方法

林海飞<sup>1,2</sup>, 周捷<sup>1</sup>, 金洪伟<sup>1,2</sup>, 李树刚<sup>1,2</sup>, 赵鹏翔<sup>1,2</sup>, 刘时豪<sup>1</sup>

(1. 西安科技大学安全科学与工程学院, 陕西 西安 710054;

2. 西安科技大学煤炭行业西部矿井瓦斯智能抽采工程研究中心, 陕西 西安 710054)

**摘要** 煤与瓦斯突出危险性预测可有效防止煤矿井下突出灾害事故。为进一步提高煤与瓦斯突出危险等级预测的科学性及准确性, 构建了基于多算法和多元分析的煤与瓦斯突出动态预测模型。选择 51 组煤与瓦斯突出工程案例数据作为样本集, 对样本数据进行空值填补、数据标准化等预处理, 通过引入 6 种特征选择方法及 6 种有监督机器学习算法构建了 42 种煤与瓦斯突出危险等级预测模型。采用准确率、混淆矩阵、Kappa 系数及  $F1$  值等指标对预测模型的性能进行验证与评估, 筛选出精度及稳定度高的 4 种机器学习算法和 3 种特征参数组合, 确定了 8 种最优分类模型, 并对 8 组典型的煤与瓦斯突出事故案例进行等级预测。结果表明: 8 种最优分类预测模型准确率为 0.667~0.961, Kappa 系数为 0.625~0.920,  $F1$  值为 0.615~1; 实际案例煤与瓦斯突出预测准确率为 100%, 突出等级预测准确率为 87.5%。所构建的多参数、多算法、多组合、多判定指标的煤与瓦斯突出等级协同预测模型精度较高, 且具有一定的普适性, 可为煤与瓦斯突出危险等级预测提供一种新思路。

**关键词** 煤与瓦斯突出; 机器学习; 特征选择; 等级划分

中图分类号 TD 353

文献标志码 A

DOI 10.13545/j.cnki.jmse.2022.0010

## Cooperative prediction method of coal and gas outburst risk grade based on feature selection and machine learning algorithm

LIN Haifei<sup>1,2</sup>, ZHOU Jie<sup>1</sup>, JIN Hongwei<sup>1,2</sup>, LI Shugang<sup>1,2</sup>, ZHAO Pengxiang<sup>1,2</sup>, LIU Shihao<sup>1</sup>

(1. College of Safety Science and Engineering, Xi'an University of Science and Technology, Xi'an, Shaanxi 710054, China;

2. Coal Industry Engineering Research Center for Western Mine Gas Intelligent Extraction, Xi'an University of Science and Technology, Xi'an, Shaanxi 710054, China)

**Abstract** Coal and gas outburst risk prediction is essential to effectively prevent underground coal and gas outburst disasters. In order to improve the scientificity and accuracy of coal and gas outburst risk level prediction, proposed a dynamic prediction model of coal and gas outburst based on multi-algorithm and multivariate analysis. The system selected 51 sets of coal and gas outburst engineering case data as the sample set. Null filling and data standardization were used for preprocessing of the sample data. The 42 prediction models of the coal and gas outburst risk level were built by introducing 6 feature selection methods and 6 supervised machine learning algorithms. The accuracy, confusion matrix, Kappa coefficient

收稿日期: 2022-01-11

责任编辑: 苗青

基金项目: 国家自然科学基金重点项目 (51734007); 陕西省杰出青年项目 (2020JC-48); 新疆维吾尔自治区创新环境建设专项项目 (PT201)

作者简介: 林海飞 (1979—) 男, 山西省天镇县人, 博士, 教授, 从事矿山安全及瓦斯灾害防治方面的研究。

E-mail: lhaifei@163.com

Tel: 029-85587450

cient and  $F1$  value were used to verify and evaluate the performance of the prediction mode. 8 optimal classification models were determined. Finally, the classification model is used for the prediction of 8 typical coal and gas outburst cases. The results show that the accuracy rate of the 8 optimal classification prediction models is 0.667–0.961, the Kappa coefficient is 0.625–0.920, and the  $F1$  value is 0.615–1. The actual case of coal and gas outburst prediction accuracy rate is 100%, and the outburst grade prediction accuracy rate is 87.5%. The constructed multi-parameter, multi-algorithm, multi-combination, and multi-identification index collaborative prediction system of coal and gas outburst level has high accuracy and a certain degree of universality, which can provide a new way for the prediction of the coal and gas outburst risk level.

**Key words** coal and gas outburst; machine learning; feature selection; classification

煤与瓦斯突出是一种井下特殊的瓦斯涌出现象,严重影响煤矿的安全生产<sup>[1-2]</sup>。我国 90% 以上的煤矿是井工开采,随着煤炭开采深度的增加,高瓦斯压力、高地应力等作用更加明显,煤与瓦斯突出发生概率及严重程度加大。如何精准预测煤层瓦斯突出危险性及其突出等级,采取相应的对策措施成为保障煤矿安全生产的主要任务之一<sup>[3]</sup>。

由于影响煤与瓦斯突出的各参数具有动态变化及各因素间存在非线性关系等特点,运用简单的数学方法很难对突出危险程度进行准确判别<sup>[4]</sup>。近年来,基于案例分析的突出预测方法得到了较快发展,该方法主要借助机器学习算法、数据挖掘技术等数学工具进行预测研究。张建国等<sup>[5]</sup>在分析矿区自然地质动力条件的前提下,对突出发生的潜在危险性进行了预测。Li 等<sup>[6]</sup>、王佳信等<sup>[7]</sup>改进 BP 神经网络模型预测煤与瓦斯突出危险性。He 等<sup>[8]</sup>采用向后神经网络(BA-ANN)及指数评估方法(EEM)对深井中煤与瓦斯突出危险类别进行分类。张友谊等<sup>[9]</sup>改进灰色关联模型并预测多参数下煤与瓦斯突出危险性。Zhang 等<sup>[10]</sup>、李冬等<sup>[11]</sup>将人工神经网络(ANN)、贝叶斯判别等方法运用到煤与瓦斯突出危险分级中,并定义了不同的危险警报级别。Li 等<sup>[12]</sup>认为逻辑回归(LR)方法能够对突出因变量和自变量之间的关系进行回归建模,从而实现分级预警。

目前,煤与瓦斯突出等级大部分是通过某种单一数学模型进行预测,没有全面考虑影响突出的因素与瓦斯突出之间复杂的非线性关系,预测精准度需进一步提高。针对该问题,本文引入信息增益、卡方检验、基尼系数、随机森林、Spearman 系数及递归特征消除等特征选择算法筛选预测指标,采用支持向量机、决策树、随机森林、朴素贝叶斯、K 最近邻及 BP 神经网络等经典机器学习算法预测突出等级,运用准确率、混淆矩阵、Kappa 系数及  $F1$  值等指

标综合评估模型的性能,选择分类器协同预测组合,并进行实际事故案例验证,以期为准预测煤与瓦斯突出危险等级、采取合理有效防治措施提供一定依据。

## 1 煤与瓦斯突出危险等级预测数据样本

### 1.1 煤与瓦斯突出影响因素及强度划分

#### 1.1.1 煤与瓦斯突出指标选取

目前在煤与瓦斯突出发生机理方面受到广泛认可的是综合作用假说,该假说认为煤与瓦斯突出主要受地应力、瓦斯和煤体物理力学性质 3 方面因素共同影响<sup>[13]</sup>。

1) 地应力。地应力是造成突出的首要原因,包括岩层应力、集中应力及构造应力<sup>[14]</sup>。埋深是影响岩层应力及集中应力的主要因素,煤层埋深增加便于瓦斯赋存,进而导致煤与瓦斯突出发生的可能性加大。构造应力是由于地质构造作用引起的应力,其与构造复杂度等因素有关,按断层、褶皱、节理等特征可将构造复杂程度划分为 4 类,其中 1 类表示简单构造,2 类表示一般构造,3 类表示较复杂构造,4 类表示极其复杂构造<sup>[15]</sup>。

2) 瓦斯因素。瓦斯作为煤与瓦斯突出的必要条件,在突出发生时主要起抛出和搬运煤体的作用<sup>[16]</sup>。当井下瓦斯压力和瓦斯含量达到某一极值时,突出才有可能发生及进一步发展。瓦斯放散初速度越大表示煤体解吸释放瓦斯越快,突出的可能性越大。

3) 煤体物理力学性质。该因素决定了突出发生、发展的难易程度,与煤的坚固性系数、煤的破坏类型、煤层厚度、煤层倾角等因素有关<sup>[17]</sup>。煤的坚固性系数越小,煤体越软,则突出危险性越大。根据《煤与瓦斯突出矿井鉴定规范》(AQ 1024—2006),煤的破坏类型划分为 5 类。其中,1 类为非破坏煤,2 类为破坏煤,3 类为强烈破坏煤(片状煤),4 类为粉碎煤(粒状煤),5 类为全粉状煤(土状

煤)。煤层厚度越大,煤层的稳定性越差,越容易导致突出。煤层倾角越大,煤层自重影响越大,突出危险性越大。

综上,可建立综合反映煤与瓦斯突出危险性的指标体系,如图 1 所示。其中地应力包含煤层埋深、构造复杂度;瓦斯因素包括煤层瓦斯压力、煤层瓦斯含量以及煤的瓦斯放散初速度;煤的物理力学性质包括煤的坚固性系数、煤的破坏类型、煤层厚度及煤层倾角。

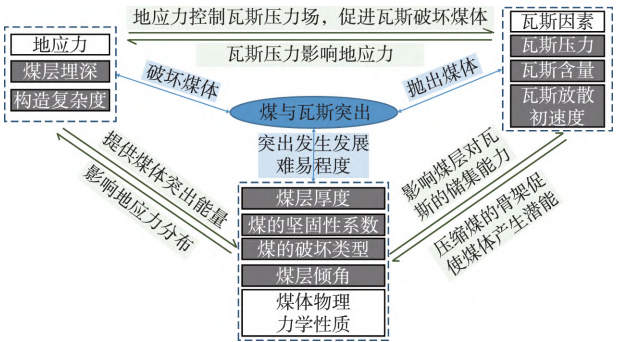


图 1 煤与瓦斯突出指标

Fig. 1 Coal and gas outburst indicators

1.1.2 煤与瓦斯突出强度划分

本文将煤与瓦斯突出危险等级划分为 4 类<sup>[18]</sup>:无突出、小型突出(Ⅰ类,突出煤量<100 t)、中型突出(Ⅱ类,突出煤量 100~500 t)、大型突出(Ⅲ类,突出煤量>500 t)。

1.2 煤与瓦斯突出案例获取

本文选取 51 组突出案例构成煤与瓦斯突出预测样本集,其数据均来自文献[19-22]。51 组突出预测样本主要来自贵州省盘江矿区、六盘水矿区及河北省开平矿区。根据前文煤与瓦斯突出强度划分规则,样本集分为 4 个等级:无突出(17 例)、小型突出(Ⅰ类,12 例)、中型突出(Ⅱ类,8 例)及大型突出(Ⅲ类,14 例)。这些案例包含了上述突出评价指标,如煤层埋深( $D$ )、构造复杂程度( $SC$ )、煤层瓦斯压力( $P$ )、煤层瓦斯含量( $W$ )、瓦斯放散初速度( $\Delta p$ )、煤的坚固性系数( $f$ )、煤的破坏类型( $DTC$ )、煤层厚度( $H$ )以及煤层倾角( $DAC$ )。样本集中有 10 组煤层倾角数据缺失,其余指标均完整。表 1 为样本集不同突出评价指标的统计参数。

表 1 煤与瓦斯突出样本集不同指标统计参数

Table 1 Statistical parameters of different indexes of coal and gas outburst sample set

突出等级	取值类型	$D/\text{m}$	$SC$	$P/\text{MPa}$	$W/(\text{m}^3 \cdot \text{t}^{-1})$	$\Delta p/(\text{mL} \cdot \text{s}^{-1})$	$f$	$DTC$	$H/\text{m}$	$DAC/(\text{°})$
无	最小值	178.00	1.00	0.58	9.68	7.83	0.40	2.00	0.58	15.00
	最大值	212.00	2.00	0.72	16.37	20.07	0.60	3.00	1.91	21.00
	均值	195.07	1.13	0.67	14.12	10.30	0.51	2.27	1.37	17.73
	标准差	10.80	0.34	0.05	1.94	2.93	0.06	0.44	0.51	1.84
Ⅰ	最小值	239.00	1.00	0.43	6.18	9.35	0.25	2.00	1.83	6.00
	最大值	411.00	3.00	1.40	19.28	23.08	0.46	3.00	2.67	25.00
	均值	285.86	1.71	0.80	15.41	16.73	0.40	2.71	2.23	19.43
	标准差	56.68	0.88	0.27	4.48	4.95	0.08	0.45	0.30	6.86
Ⅱ	最小值	262.00	1.00	0.49	5.08	8.15	0.22	2.00	2.00	8.00
	最大值	440.00	3.00	0.76	21.12	23.08	0.56	3.00	2.90	24.00
	均值	333.83	2.50	0.68	13.36	16.84	0.37	2.83	2.33	17.50
	标准差	74.97	0.76	0.11	6.75	6.07	0.11	0.37	0.28	6.40
Ⅲ	最小值	380.00	1.00	0.48	3.78	13.11	0.11	3.00	2.20	3.00
	最大值	1 130.00	4.00	3.60	17.56	20.19	0.53	5.00	4.71	17.00
	均值	615.92	2.38	1.60	11.09	16.64	0.28	3.77	3.46	8.54
	标准差	218.51	1.00	0.99	4.46	2.25	0.13	0.80	0.77	4.89

1.3 数据清洗

1.3.1 空值填补

以往在进行数据处理时会直接删除缺失部分参数,然而这种处理方法会导致样本数据大量减少,引起分类模型训练不充分从而导致分类算法辨识度低等问题。为尽可能多地保留样本,本文对 10

组煤层倾角缺失数据进行空值填充,用均值、中位数、众数、回归插补、多重插补(多重插补采用 4 次插补)等方法进行填补,比较各方法计算结果,最终选取模型准确率( $R^2$ )最高时的空值填充方法,即多重插补法(四重插补)填补空值。各方法空值插补准确性结果如表 2 所示。

表 2 煤层倾角空值插补准确性统计

Table 2 Accuracy statistics of null interpolation of dip angle of coal seam

序列	实际值	均值	中位数	众数	回归插补	一重插补	二重插补	三重插补	四重插补
1	6	8.75	7	3	10.38	7	7	5	7
2	23	20.17	19	18	21.00	17	25	25	20
3	23	20.17	19	18	21.00	19	15	13	24
4	25	19.00	19	18	21.00	20	25	24	24
5	22	18.17	16	18	21.00	16	23	25	21
6	24	19.00	17	17	15.00	19	24	25	19
7	23	19.00	17	17	15.00	17	24	24	21
8	15	14.33	18	18	19.26	17	24	19	14
9	18	18.17	18	18	19.26	19	15	19	21
10	15	14.33	18	18	19.26	24	19	17	14
准确率		0.64	0.35	0.29	0.29	0.18	0.44	0.57	0.83

### 1.3.2 数据标准化

煤与瓦斯突出案例中选用的 9 个突出评价指标均是数值型数据,不同特征取值范围不同,甚至可能数量级不同。为获得准确的分类结果,确保每一特征作用的发挥,对特征进行零-均值规范化(z-score)处理,以减少特征规模、量纲对模型的影响。

对序列  $x_1, x_2, \dots, x_n$  进行变换:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (1)$$

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (2)$$

$$h_i = \frac{x_i - \bar{x}}{s} \quad (3)$$

式中:  $x_i$  为原始序列  $i \in [1, n]$ ;  $\bar{x}$  为序列平均值;  $s$  为方差;  $h_i$  为变换后新序列  $i \in [1, n]$ 。

## 2 煤与瓦斯突出危险等级预测模型

### 2.1 煤与瓦斯突出危险等级预测流程

煤与瓦斯突出危险等级预测流程如图 2 所示。

#### 1) 样本数据处理

采用多重插补法对样本集数据中的 10 组随机煤层倾角指标进行插补,选择准确率  $R^2$  最高的插补方法补全缺失值。运用 z-score 数据标准化方法处理样本集数据。

#### 2) 模型确定

① 参数组合确定。采用信息增益、卡方检验、基尼系数、随机森林、Spearman 相关系数及递归特征消除等方法获取 7 种不同特征参数组合。

② 分类模型确定。预测模型训练过程采用五折网格搜索交叉验证进行模型的参数寻优。用支

持向量机、决策树、随机森林、朴素贝叶斯、K 最近邻及 BP 神经网络等方法预测各参数组合下的突出分类;根据总体精确率及混淆矩阵反映模型整体分类性能并确定最优分类器;用 Kappa 系数及  $F1$  值来衡量分类器对不同突出等级的预测性能。

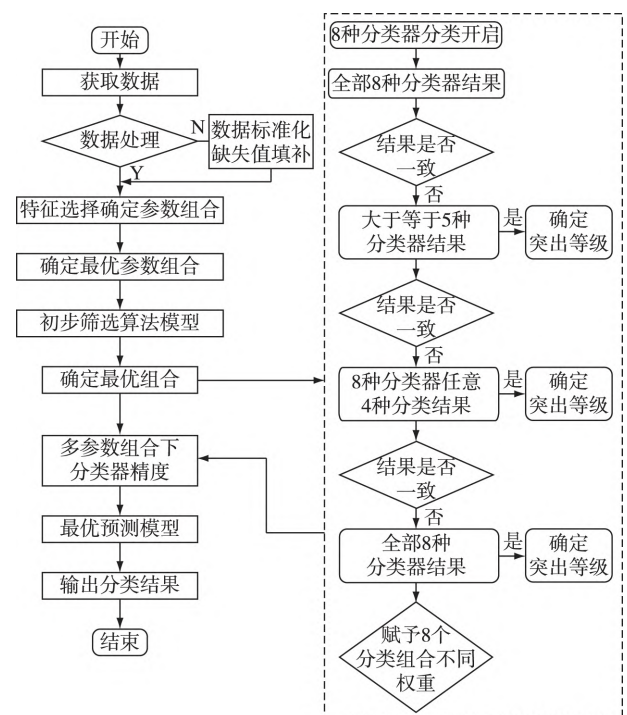


图 2 基于机器学习的煤与瓦斯突出等级预测流程图

Fig. 2 Flow chart of coal and gas outburst grade prediction based on machine learning

#### 3) 模型验证

采用分类器对样本集进行分类并得到最终结果,与实际等级对比得到模型准确率,并进行实例验证。

2.2 煤与瓦斯突出危险等级预测模型建立方法

2.2.1 特征选择方法

分类模型中能否选择最佳参数子集是影响分类准确率的重要一环,不同参数组合对各类机器学习算法的敏感性存在差异。因此,特征选择应着重选择足以代表输入数据的变量,减少不相关和冗余数据的影响。本文以增加(删除)任何特征不会产生比预测精度更高的特征子集为原则,确定在不同特征选择方法下的特征数目。

通过计算信息增益<sup>[23]</sup>可确定特征消除随机变量不确定性能力的大小。信息增益越大能力越强,则该变量越重要。本文取信息增益大于 0.5 的变量。

卡方检验<sup>[24]</sup>的基本思想是比较理论频数  $T$  与实际频数  $A$  的一致性,构造卡方统计量  $\chi^2$ ,越小表示理论频数  $T$  与实际频数  $A$  越接近。本文取  $\chi^2$  小于 50 的变量。

基尼系数<sup>[25]</sup>表示样本在集合中被错误分类的

可能性大小,其值越小表示样本越不容易被错误分类。本文取基尼系数小于 40 的变量。

随机森林由多棵决策树组成,结合每棵决策树得到的变量重要性,得到最终变量重要性排序。Mean Decrease Accuracy(MDA) 越小则该指标的重要性越大。本文取 MDA 值小于 10 的变量。

Spearman 相关系数表明变量  $X$  和  $Y$  的相关方向及大小。当  $X$  与  $Y$  同时增加,则相关系数为正;如果  $X$  增加时  $Y$  趋向于减少,则相关系数为负。本文取 Spearman 系数大于 0.4 的变量。

递归特征消除<sup>[26]</sup>是给每一个特征指定一个权重,在获取到特征的权重值后删除权重绝对值最小的特征。依此方法不断循环递归,直至剩余的特征数量其误差最小,最终取均方根误差最小的特征组合。

图 3 为各方法特征选择的方法指标值分布,表 3 为各方法特征选择后的特征组合。

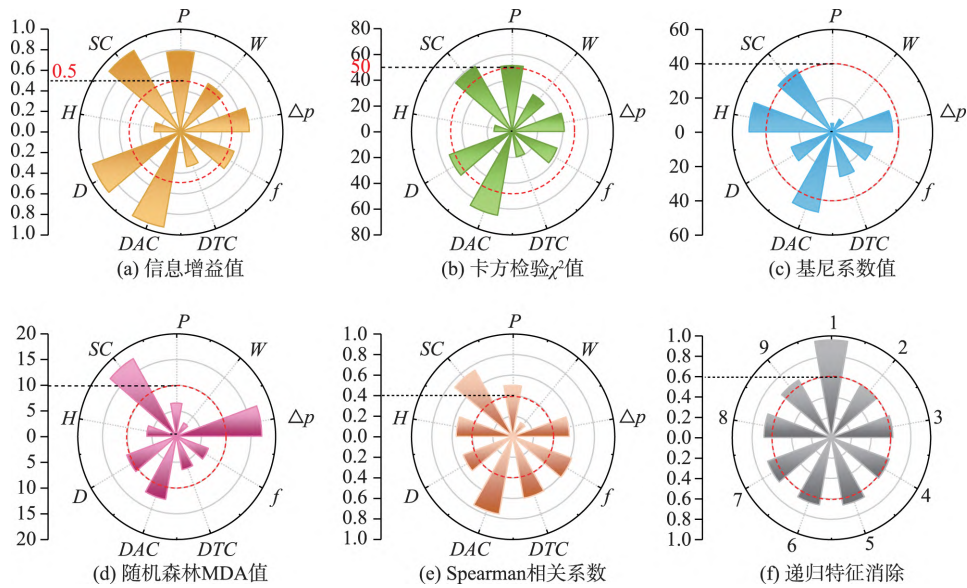


图 3 特征参数选择指标数据分布

Fig. 3 Distribution of characteristic parameter selection index data

表 3 7 种不同的特征组合

Table 3 7 different feature combinations

特征选择方法	组合名称	煤与瓦斯突出相关参数								
		$D$	$SC$	$P$	$W$	$\Delta p$	$f$	$DTC$	$H$	$DAC$
信息增益	F-1	✓	✓	✓	✓	✓	✓			✓
卡方检验	F-2				✓	✓	✓	✓	✓	
基尼系数	F-3	✓		✓	✓	✓	✓	✓		
随机森林	F-4			✓	✓		✓	✓	✓	
Spearman 系数	F-5	✓	✓	✓		✓	✓	✓	✓	✓
递归特征消除	F-6	✓				✓				
全部特征	F-7	✓	✓	✓	✓	✓	✓	✓	✓	✓



2.2.2 机器学习算法

基于 R 语言机器学习基础算法包,本文采用 6 种有监督学习算法对突出数据集进行学习,包括支持向量机(Support Vector Machine,SVM)、决策树(Decision Tree,DT)、随机森林(Random Forest,RF)、朴素贝叶斯(Naive Bayesian,NB)、K 最近邻(K-Nearest Neighbor,KNN)及 BP 神经网络(Back Propagation Neural Network,BPNN)。各算法关键参数如表 4 所示。

表 4 各算法关键参数

Table 4 Key parameters of each algorithm

算法	关键参数
SVM	C = 10 ,kernal = RBF ,gamma = 0.1
DT	splitter = best ,min_samples_leaf = 1 ,min_samples_split = 2
RF	trees = 500 ,max_features = sqrt ,estimators = 100
NB	$\lambda = 1$
KNN	kernel = triangular ,k = 4
BPNN	Hidden node = 10 , $\alpha = 0.01$ ,epsilon = 0.001

2.3 煤与瓦斯突出危险等级预测模型确定

2.3.1 模型初选

1) 参数组合初选

表 5 不同算法采用不同参数组合时的分类准确率

Table 5 Classification accuracy of different algorithms using different parameter combinations

特征组合	SVM	DT	RF	NB	KNN	BPNN	平均值
F-1	0.882	0.902	0.863	0.725	0.922	0.863	0.860
F-2	0.843	0.843	0.765	0.784	0.843	0.922	0.833
F-3	0.863	0.922	0.824	0.725	0.863	0.824	0.837
F-4	0.902	0.784	0.804	0.824	0.863	0.961	0.856
F-5	0.922	0.863	0.882	0.804	0.902	0.902	0.879
F-6	0.686	0.784	0.745	0.667	0.824	0.725	0.739
F-7	0.941	0.882	0.863	0.725	0.922	0.863	0.866
平均值	0.863	0.854	0.821	0.751	0.877	0.866	—

2) 分类器初选

由前述可知,采用特征参数组合 F-1、F-5、F-7 进行预测的效果较好,为进一步明确煤与瓦斯突出危险等级预测模型的性能,采用测试集混淆矩阵来表征真实等级与预测等级之间的关系。由参数组合 F-1、F-5、F-7 与各机器学习算法混淆矩阵图(如图 4~6 所示)可知,对参数组合 F-1 进行煤与瓦斯突出危险性分类预测时,各算法对无突出的预测效果最好,准确率为 0.94~1,Ⅲ类及Ⅱ类突出的分类准确率次之,Ⅰ类突出分类结果较不理想,最低值为采用 NB 分类器时的预测结果,其预测精度为

采用 6 种有监督学习算法及 7 种参数组合构建的模型准确率如表 5 所示。由表 5 可知,不同算法采用不同参数组合时的模型准确率在 0.667~0.941。在各分类器中 K 最近邻(KNN)算法预测精度最高,对特征变化反应最小,其准确率为 0.824~0.922,平均值为 0.877;其次是 BP 神经网络(BPNN),准确率为 0.725~0.961,平均值为 0.866。

除 KNN 算法、BPNN 算法外,SVM、DT、NB、RF 四种算法对特征参数组合的选择较敏感,SVM 算法准确率波动最大,采用参数组合 F-6 时最小为 0.686,采用参数组合 F-7 时达最大值 0.941。对不同特征参数组合进行预测,采用参数组合 F-5 时,不同算法的准确率为 0.804~0.922,平均值为 0.879;其次为采用参数组合 F-7 时准确率为 0.725~0.941,平均值为 0.866;采用参数组合 F-1 时准确率为 0.725~0.922,平均值为 0.860。值得注意的是,参数组合 F-7 采用了全部特征参数,但与其与 NB 算法组合的预测精度为 0.725。参数组合 F-2、F-3、F-4 及 F-6 的预测精准较差,尤其是参数组合 F-6,其预测精度为 0.667~0.824,平均值仅为 0.739。

0.50;在采用参数组合 F-5 进行煤与瓦斯突出危险性预测时,各算法对无突出及Ⅲ类突出的预测精度较高,Ⅰ类次之,Ⅱ类精度最低。采用参数组合 F-7 作为煤与瓦斯突出危险性预测时,无突出与Ⅲ类突出的预测效果较好,Ⅰ类、Ⅱ类预测精度较低,整体预测精度分布在 0.5~1。

综上可知,参数组合 F-1、F-5、F-7 中 SVM、DT、KNN 及 BPNN 精度分布在 0.863~0.941,均高于平均水平。因此,本文选取 SVM、KNN、DT 及 BPNN 为最优分类器。

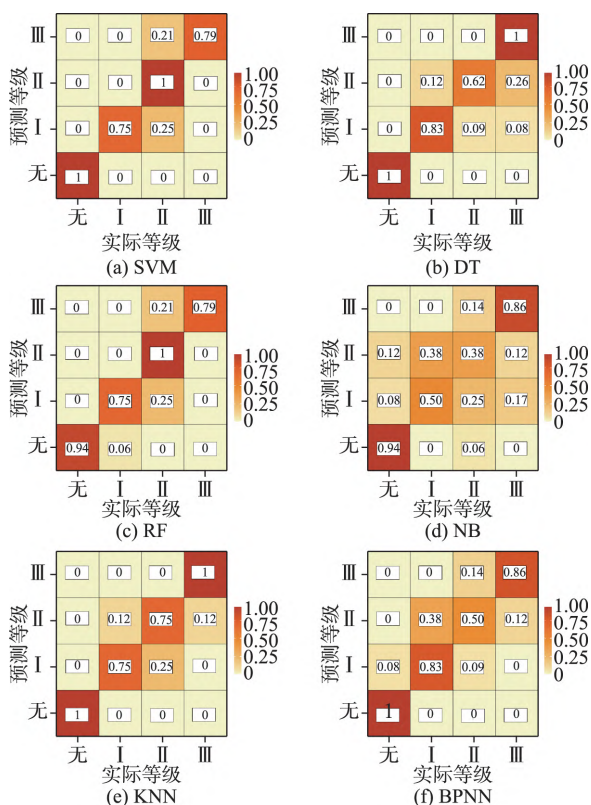


图4 F-1 各机器学习算法混淆矩阵图

Fig. 4 F-1 confusion matrix diagram of each machine learning algorithm

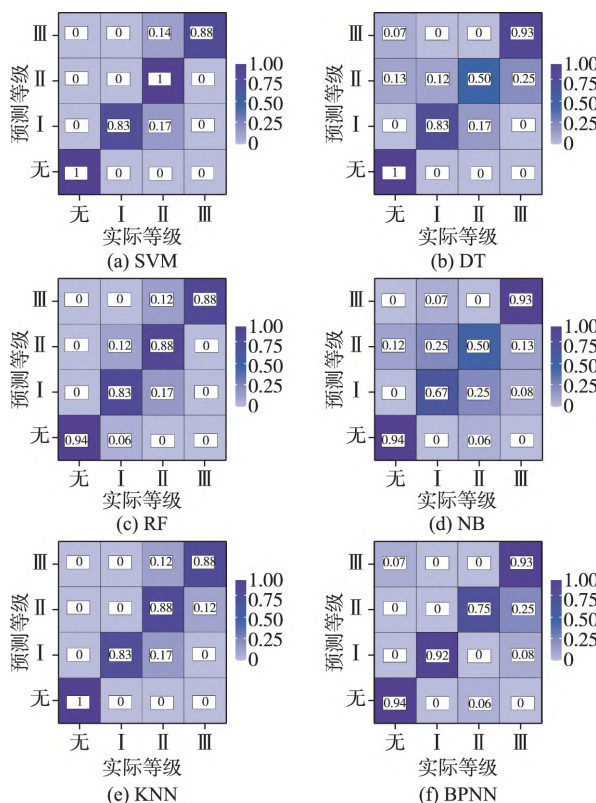


图5 F-5 各机器学习算法混淆矩阵图

Fig. 5 F-5 confusion matrix diagram of each machine learning algorithm

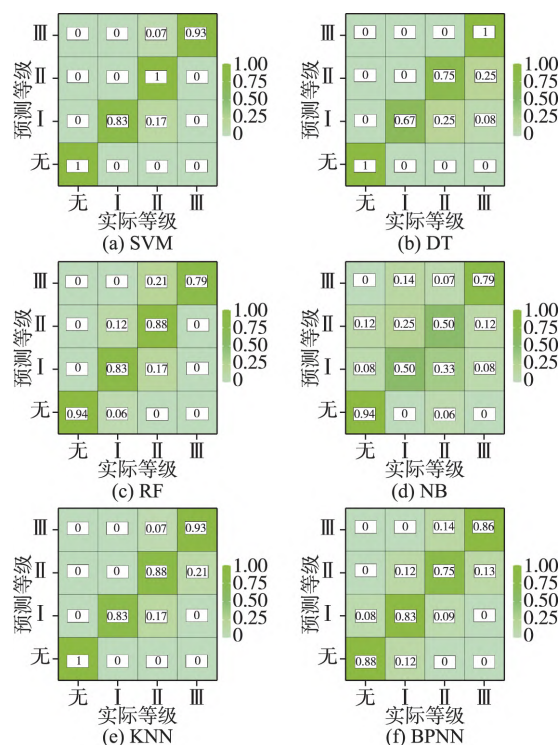


图6 F-7 各机器学习算法混淆矩阵图

Fig. 6 F-7 confusion matrix diagram of each machine learning algorithm

### 2.3.2 模型优选

F-1、F-5 及 F-7 三种参数组合的 Kappa 系数平均数分别为 0.800、0.835 及 0.814。由图 7 可知, 3 种参数组合 Kappa 系数平均数均大于 0.75, SVM、DT、KNN 及 RF 的 Kappa 系数均高于 0.800。初步选择 F-1、F-5、F-7 与 SVM、DT、KNN、RF 组合的 12 种模型。

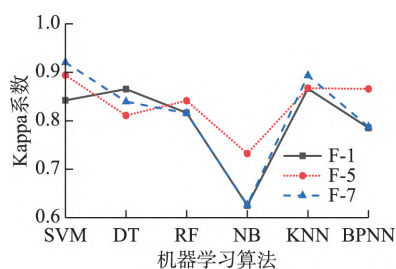


图7 F-1、F-5、F-7 参数组合 Kappa 系数分布

Fig. 7 Kappa coefficient distribution of F-1, F-5, F-7 parameter combination

### 2.3.3 确定预测模型

根据上述结果, 对采用参数组合 F-1、F-5、F-7 不同算法的  $F1$  值进行计算, 结果如图 8 所示。从图 8 可看出, 无论采用何种分类器, 无突出与 III 类突出的  $F1$  值均大于 I 类突出和 II 类突出。F-5、F-7 对于 I、II 类突出分类的  $F1$  值普遍较低, 故剔除 F-5、

F-7 与 DT、RF 组合的结果。

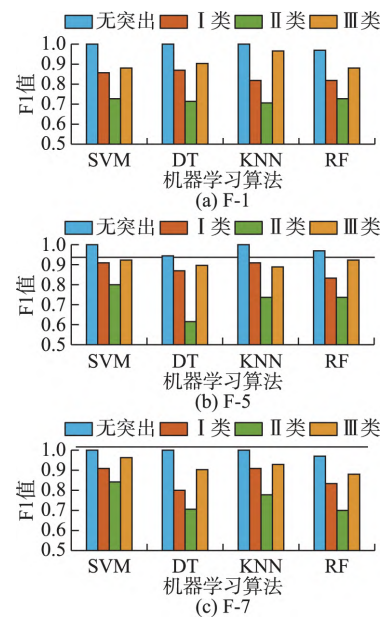


图 8 各参数组合中各分类器 F1 值  
Fig. 8 F1 value of each classifier in each parameter combination

最终预测组合为 F-5、F-7 与 SVM、KNN 结合

的 4 种预测组合; F-1 分别与 SVM、DT、KNN、RF 结合的 4 种组合, 共 8 种。

3 突出等级预测模型验证

3.1 煤与瓦斯突出案例数据

用 8 起煤与瓦斯突出案例对突出等级预测模型进行验证, 各案例参数如表 6 所示。煤与瓦斯突出事故发生的时间地点分别为 2004 年 10 月 20 日河南大平煤矿 21 轨道下山岩石掘进工作面(案例 1); 2007 年 11 月 12 日平煤集团十矿己<sub>15-16</sub>-24110 综采工作面(案例 2)<sup>[27]</sup>; 2012 年 11 月 24 日贵州响水煤矿河西采区 1135 运输巷掘进工作面(案例 3); 2014 年 3 月 21 日河南长虹矿东翼下延采区二<sub>1</sub>-21010 机巷掘进工作面(案例 4)<sup>[28]</sup>; 2016 年 3 月 6 日吉林松树镇煤矿+100 m 东一采区 4112 运输巷掘进工作面(案例 5); 2016 年 9 月 6 日河南平顶山十三矿己<sub>15-17</sub>-11111 综采工作面(案例 6); 2017 年 5 月 15 日河南薛湖煤矿 23 采区 2306 风巷掘进工作面(案例 7); 2018 年 11 月 27 日黑龙江新立煤矿一采区四阶段三片左部瓦斯巷掘进工作面(案例 8)。

表 6 煤与瓦斯突出案例各指标数据  
Table 6 Index data of coal and gas outburst cases

案例	<i>D</i>	<i>SC</i>	<i>P</i> / MPa	<i>W</i> / ( $\text{m}^3 \cdot \text{s}^{-1}$ )	<i>p</i> / ( $\text{L} \cdot \text{s}^{-1}$ )	<i>f</i>	<i>DTC</i>	<i>H</i> / m	<i>DAC</i> / ( $^\circ$ )	突出煤量/ t	突出 等级
1	612	4	2.00	12.59	31.00	0.12	5	7.00	19.0	1 894	Ⅲ
2	960	3	1.03	6.49	11.75	0.26	4	2.00	15.5	2 000	Ⅲ
3	203	4	1.65	13.91	18.24	0.33	3	3.50	19.5	490	Ⅱ
4	273	2	0.28	5.63	27.00	0.19	5	3.80	19.0	970	Ⅲ
5	467	3	3.98	11.31	2.60	0.35	5	3.00	10.0	455	Ⅱ
6	720	3	3.60	16.98	9.07	0.32	4	5.80	17.0	301	Ⅱ
7	886	1	0.92	8.71	4.00	0.22	4	2.50	7.0	116	Ⅱ
8	200	3	1.60	10.30	10.34	0.32	4	2.60	25.0	562	Ⅲ

3.2 煤与瓦斯突出预测结果

结果如表 7 所示。

用前述 8 种分类器对突出案例进行分析, 预测

表 7 煤与瓦斯突出案例等级预测结果  
Table 7 Grade prediction results of coal and gas outburst cases

案例	突出 等级	F-1				F-5		F-7		预测 等级
		SVM	DT	KNN	RF	SVM	KNN	SVM	KNN	
1	Ⅲ	Ⅲ	Ⅲ	Ⅲ	Ⅱ	Ⅱ	Ⅲ	Ⅱ	Ⅲ	Ⅲ
2	Ⅲ	Ⅱ	Ⅲ	Ⅲ	Ⅱ	Ⅲ	Ⅲ	Ⅲ	Ⅲ	Ⅲ
3	Ⅱ	Ⅱ	Ⅲ	Ⅱ	Ⅱ	Ⅱ	Ⅲ	Ⅱ	Ⅱ	Ⅱ
4	Ⅲ	Ⅲ	Ⅲ	Ⅲ	Ⅱ	Ⅱ	Ⅱ	Ⅱ	Ⅲ	Ⅲ
5	Ⅱ	Ⅱ	Ⅱ	Ⅱ	Ⅱ	Ⅱ	I	Ⅱ	Ⅱ	Ⅱ
6	Ⅱ	Ⅱ	Ⅱ	Ⅲ	Ⅱ	Ⅱ	Ⅲ	Ⅱ	Ⅲ	Ⅱ
7	Ⅱ	I	Ⅱ	Ⅱ	Ⅱ	Ⅱ	Ⅱ	Ⅱ	Ⅱ	Ⅱ
8	Ⅲ	Ⅱ	Ⅲ	Ⅱ	Ⅲ	Ⅱ	Ⅱ	Ⅱ	I	Ⅱ



当8种分类结果不同时,以多数类( $\geq 5$ )为准。当8个分类组合不能确定多数类时,赋予8个分类组合不同权重(归一化的准确率)重新投票确定,具体流程如图2所示。由表7可看出前7组案例全部预测正确,最后一组案例将Ⅲ类突出预测为Ⅱ类突出。综上所述,预测突出率为100%,预测突出等级准确率为87.5%。

## 4 结 论

1) 建立了考虑地应力、瓦斯因素和煤体物理力学性质3个二级指标、9个三级指标的煤与瓦斯突出危险等级预测指标体系。

2) 提出了集样本数据处理、特征参数组合选择、预测模型建立、模型初选、优化及预测为一体的煤与瓦斯突出危险等级预测模型。

3) 采用准确率、混淆矩阵、Kappa系数及F1值等指标,确定了F-5、F-7与SVM、KNN,F-1与SVM、DT、KNN、RF结合的8种最优分类组合模型。用所构建的8种最优分类模型对实际案例进行煤与瓦斯突出危险等级预测,突出预测准确率为100%,突出等级预测准确率为87.5%。

## 参考文献:

- [1] 唐巨鹏,郝娜,潘一山,等. 基于声发射能量分析的煤与瓦斯突出前兆特征试验研究[J]. 岩石力学与工程学报,2021,40(1): 31-42.  
TANG Jupeng, HAO Na, PAN Yishan, et al. Experimental study on precursor characteristics of coal and gas outbursts based on acoustic emission energy analysis[J]. Chinese Journal of Rock Mechanics and Engineering, 2021, 40(1): 31-42.
- [2] 李云波,张腾飞,宋党育,等. 基于多重分形奇异性理论的瓦斯地质规律预测方法——以淮南矿区潘一矿为例[J]. 煤炭学报,2018,43(12): 3436-3446.  
LI Yunbo, ZHANG Tengfei, SONG Danguyu, et al. Prediction method of methane geological law based on the multifractal singularity theory: An example from the Panyi Coalmine, Huainan coalfield, China[J]. Journal of China Coal Society, 2018, 43(12): 3436-3446.
- [3] 袁亮. 煤矿典型动力灾害风险辨识及监控预警技术研究进展[J]. 煤炭学报,2020,45(5): 1557-1566.  
YUAN Liang. Research progress on risk identification, assessment, monitoring and early warning technologies of typical dynamic hazards in coal mines[J]. Journal of China Coal Society, 2020, 45(5): 1557-1566.
- [4] ZHOU B, XU J, PENG S J, et al. Test system for the visualization of dynamic disasters and its application to coal and gas outburst[J]. International Journal of Rock Mechanics and Mining Sciences, 2019, 122: 104083.
- [5] 张建国,兰天伟,王满,等. 平顶山矿区深部矿井动力灾害预测方法与应用[J]. 煤炭学报,2019,44(6): 1698-1706.  
ZHANG Jianguo, LAN Tianwei, WANG Man, et al. Prediction method of deep mining dynamic disasters and its application in Pingdingshan mining area[J]. Journal of China Coal Society, 2019, 44(6): 1698-1706.
- [6] LI B, WANG E Y, SHANG Z, et al. Deep learning approach to coal and gas outburst recognition employing modified AE and EMR signal from empirical mode decomposition and time-frequency analysis[J]. Journal of Natural Gas Science and Engineering, 2021, 90: 103942.
- [7] 王佳信,周宗红,张继华,等. 煤与瓦斯突出危险性预测的S $\alpha$ S-PNN模型及应用[J]. 传感技术学报,2017,30(7): 1112-1118.  
WANG Jiaxin, ZHOU Zonghong, ZHANG Jihua, et al. S $\alpha$ S-PNN model for forecast of coal and gas outburst risk and its application[J]. Chinese Journal of Sensors and Actuators, 2017, 30(7): 1112-1118.
- [8] HE X Q, CHEN W X, NIE B S, et al. Classification technique for danger classes of coal and gas outburst in deep coal mines[J]. Safety science, 2010, 48(2): 173-178.
- [9] 张友谊,崔金雷,焦向东. 煤与瓦斯突出多指标耦合预测模型研究及应用[J]. 工程科学学报,2018,40(11): 1309-1316.  
ZHANG Youyi, CUI Jinlei, JIAO Xiangdong. Study of the multi-index coupling forecasting model of coal and gas outburst and its application[J]. Chinese Journal of Engineering, 2018, 40(11): 1309-1316.
- [10] ZHANG R L, LOWNDES I S. The application of a coupled artificial neural network and fault tree analysis model to predict coal and gas outbursts[J]. International Journal of Coal Geology, 2010, 84(2): 141-152.
- [11] 李冬,彭苏萍,杜文凤,等. 煤层瓦斯突出危险区综合预测方法[J]. 煤炭学报,2018,43(2): 466-472.  
LI Dong, PENG Suping, DU Wenfeng, et al. Comprehensive prediction method of coal seam gas outburst danger zone[J]. Journal of China Coal Society, 2018, 43(2): 466-472.
- [12] LI Z H, WANG E Y, OU J C, et al. Hazard evaluation of coal and gas outbursts in a coal-mine roadway based on logistic regression model[J]. International Journal of Rock Mechanics and Mining Sciences, 2015, 80: 185-195.
- [13] 袁亮,王伟,王汉鹏,等. 巷道掘进揭煤诱导煤与瓦斯突出模拟试验系统[J]. 中国矿业大学学报,2020,49(2): 205-214.  
YUAN Liang, WANG Wei, WANG Hanpeng, et al. A simulation system for coal and gas outburst induced by coal uncovering in roadway excavation[J]. Journal of

- China University of Mining & Technology, 2020, 49 (2): 205-214.
- [14] 高魁, 乔国栋, 刘泽功, 等. 煤与瓦斯突出机理分类研究构想及其应用探讨[J]. 采矿与安全工程学报, 2019, 36(5): 1043-1051.
- GAO Kui, QIAO Guodong, LIU Zegong, et al. On classification conception of coal and gas outburst mechanism and its application[J]. Journal of Mining & Safety Engineering, 2019, 36(5): 1043-1051.
- [15] 于不凡. 煤矿瓦斯灾害防治及利用技术手册[M]. 北京: 煤炭工业出版社, 2005.
- [16] 许江, 刘东, 彭守建, 等. 煤样粒径对煤与瓦斯突出影响的试验研究[J]. 岩石力学与工程学报, 2010, 29(6): 1231-1237.
- XU Jiang, LIU Dong, PENG Shoujian, et al. Experimental research on influence of particle diameter on coal and gas outburst[J]. Chinese Journal of Rock Mechanics and Engineering, 2010, 29(6): 1231-1237.
- [17] 王刚, 武猛猛, 王海洋, 等. 基于能量平衡模型的煤与瓦斯突出影响因素的灵敏度分析[J]. 岩石力学与工程学报, 2015, 34(2): 238-248.
- WANG Gang, WU Mengmeng, WANG Haiyang, et al. Sensitivity analysis of factors affecting coal and gas outburst based on a energy equilibrium model[J]. Chinese Journal of Rock Mechanics and Engineering, 2015, 34(2): 238-248.
- [18] 唐巨鹏, 杨森林, 王亚林, 等. 地应力和瓦斯压力作用下深部煤与瓦斯突出试验[J]. 岩土力学, 2014, 35(10): 2769-2774.
- TANG Jupeng, YANG Senlin, WANG Yalin, et al. Experiment of coal and gas outbursts under ground stress and gas pressure in deep mines[J]. Rock and Soil Mechanics, 2014, 35(10): 2769-2774.
- [19] 陈恋, 袁梅, 高强, 等. 主成分-费歇尔判别模型在煤与瓦斯突出等级预测中的应用[J]. 工矿自动化, 2020, 46(3): 55-62.
- CHEN Lian, YUAN Mei, GAO Qiang, et al. Application of principal component-Fisher discrimination model in grade prediction of coal and gas outburst[J]. Industry and Mine Automation, 2020, 46(3): 55-62.
- [20] 隆能增, 袁梅, 敖选俊, 等. 基于 LLE-FOA-BP 模型的煤与瓦斯突出强度预测[J]. 工矿自动化, 2019, 45(10): 68-73.
- LONG Nengzeng, YUAN Mei, AO Xuanjun, et al. Prediction of coal and gas outburst intensity based on LLE-FOA-BP model [J]. Industry and Mine Automation, 2019, 45(10): 68-73.
- [21] 付华, 李海霞, 卢万杰, 等. 一种改进的极限学习机煤与瓦斯突出预测模型[J]. 传感技术学报, 2016, 29(1): 69-74.
- FU Hua, LI Haixia, LU Wanjie, et al. A prediction model of coal and gas outburst based on improved extreme learning machine[J]. Chinese Journal of Sensors and Actuators, 2016, 29(1): 69-74.
- [22] 孙福成. 基于优化量子门节点神经网络的煤与瓦斯突出预测模型研究[D]. 阜新: 辽宁工程技术大学, 2020.
- [23] 尹涛, 王祥宇, 周越. 基于 Bayesian 证据推断与信息增益的参数化有限元修正模型选择[J]. 振动与冲击, 2018, 37(12): 159-166.
- YIN Tao, WANG Xiangyu, ZHOU Yue. Model selection in updating of parametric FE model based on Bayesian evidence inference and information divergence [J]. Journal of Vibration and Shock, 2018, 37(12): 159-166.
- [24] 张文, 陈剑平, 刘存合, 等. 卡方检验在裂隙岩体统计均质区划分中的应用研究[J]. 岩土工程学报, 2011, 33(9): 1440-1446.
- ZHANG Wen, CHEN Jianping, LIU Cunhe, et al. Application of chi-square test in structural homogeneity dividing of fractured rock mass [J]. Chinese Journal of Geotechnical Engineering, 2011, 33(9): 1440-1446.
- [25] GE J H, YUSA N, FAN M B. Frequency component mixing of pulsed or multi-frequency eddy current testing for nonferromagnetic plate thickness measurement using a multi-gene genetic programming algorithm [J]. NDT & E International, 2021, 120: 102423.
- [26] YOU W J, YANG Z J, JI G L. Feature selection for high-dimensional multi-category data using PLS-based local recursive feature elimination [J]. Expert Systems with Applications, 2014, 41(4): 1463-1475.
- [27] 司亚超. 回采工作面煤与瓦斯突出发生原因及对策研究[D]. 焦作: 河南理工大学, 2015.
- [28] 王武超. 豫西松软煤层掘进工作面低指标突出机理及应用研究[D]. 西安: 西安科技大学, 2017.