文章编号:1671-251X(2021)08-0102-07　　DOI:10.13272/j.issn.1671-251x.2021010024

基于随机森林回归的煤矿瓦斯涌出量预测方法

吴奉亮[1,2]，　霍源[1,2]，　高佳南[1,2]

(1.西安科技大学 安全科学与工程学院，陕西 西安　710054；
2.西部矿井开采及灾害防治教育部重点实验室，陕西 西安　710054)

扫码移动阅读

摘要:为了提高煤矿瓦斯涌出量预测精度和效率，研究了基于随机森林回归的煤矿瓦斯涌出量预测方法。采用 bootstrap 自助法重采样技术采集训练样本数据，构建随机森林回归模型，取决策树输出值的均值作为煤矿瓦斯涌出量预测结果，利用袋外数据评价回归模型预测性能。通过计算袋外数据残差平方均值和拟合优度，确定随机森林回归模型的最优超参数；采用袋外数据残差平方均值的增加量来表征特征变量的重要性，用累计影响权重达到 90%的部分特征变量代替煤矿瓦斯涌出量的全部特征变量，筛选出采煤高度、煤厚、煤层瓦斯含量、采出率、埋深、日进度、开采强度、邻近层间距 8 个重要程度高的特征变量作为模型的输入变量。测试结果表明，采用全部特征变量和部分特征变量的随机森林回归模型均具有较好的预测性能，进行特征变量选择后，模型的平均绝对误差由 0.22 m^3/min 下降到 0.21 m^3/min，平均相对误差由 3.55%下降到 3.47%。基于特征变量选择的随机森林回归模型在保证较好的预测性能的前提下，降低了预测模型特征变量的维度，减少了原始数据获取工作，提高了预测效率。

关键词:瓦斯涌出量预测；随机森林回归；超参数寻优；特征变量选择；影响权重

中图分类号:TD712　　文献标志码:A

Coal mine gas emission prediction method based on random forest regression

WU Fengliang[1,2]，　HUO Yuan[1,2]，　GAO Jianan[1,2]

(1. College of Safety Science and Engineering, Xi'an University of Science and Technology, Xi'an 710054, China; 2. Key Laboratory of Western Mine and Hazard Prevention, Ministry of Education, Xi'an 710054, China)

Abstract: In order to improve the prediction accuracy and efficiency of coal mine gas emission, a coal mine gas emission prediction method based on random forest regression is proposed. The bootstrap self-service resampling technology is used to collect training sample data and construct a random forest regression model. The mean value of the decision tree output value is taken as the prediction result of coal mine gas emission and the out-of-bag data is used to evaluate the prediction performance of the regression model. The optimal hyperparameters of the random forest regression model are determined by calculating the mean of squared residuals and goodness of fit of the out-of-bag data. The increase in the mean of squared residuals of the out-of-bag data is used to characterize the importance of the characteristic variables. All the characteristic variables of coal mine gas emission are replaced by some characteristic variables with cumulative influence weight of 90%. And eight characteristic variables with high importance are selected as input variables of the model, including coal mining height, coal thickness, coal seam gas content, recovery rate, burial depth, daily progress, mining intensity and adjacent layers spacing. The

收稿日期:2021-01-10;**修回日期**:2021-07-25;**责任编辑**:胡娴。

基金项目:国家自然科学基金资助项目(51974232)。

作者简介:吴奉亮(1977－)，男，四川新都人，教授，博士，研究方向为矿井通风与安全，E-mail:15038537@qq.com。通信作者:高佳南(1991－)，男，陕西米脂人，助理工程师，硕士，研究方向为矿井通风与安全，E-mail:943157956@qq.com。

引用格式:吴奉亮，霍源，高佳南. 基于随机森林回归的煤矿瓦斯涌出量预测方法[J]. 工矿自动化，2021，47(8):102-107.
WU Fengliang, HUO Yuan, GAO Jianan. Coal mine gas emission prediction method based on random forest regression[J]. Industry and Mine Automation, 2021, 47(8): 102-107.

test results show that the random forest regression model with all characteristic variables and some characteristic variables has good prediction performance. After selecting characteristic variables, the average absolute error of the model decreases from 0.22 m^3/min to 0.21 m^3/min, and the average relative error decreases from 3.55% to 3.47%. The random forest regression model based on characteristic variable selection reduces the dimensionality of the characteristic variables of the prediction model, reduces the original data acquisition work, and improves the prediction efficiency under the premise of ensuring better prediction performance.

Key words: gas emission prediction; random forest regression; hyperparameter optimization; characteristic variable selection; influence weight

0　引言

瓦斯灾害事故是威胁煤矿安全高效开采的主要因素[1-2],为遏制瓦斯事故的发生,准确、高效的煤矿瓦斯涌出量预测方法尤为重要[3]。由于煤矿瓦斯涌出量的各影响因子间呈现出非线性关系[4],线性预测方法难以得到满意的预测结果。因此,灰色理论、神经网络和支持向量机等非线性映射方法被广泛应用于煤矿瓦斯涌出量预测。题正义等[5]基于模糊数学和灰色系统建立了瓦斯涌出量预测模型。李树刚等[6]将因子分析与 BP 神经网络相耦合,对瓦斯涌出量进行预测。付华等[7]提出了一种利用蚁群聚类算法优化 Elman 神经网络的瓦斯涌出量预测方法。孙林等[8]、董晓雷等[9]、张强等[10]利用支持向量机实现瓦斯涌出量预测;谢东海等[11]将未确知测度理论引入到瓦斯涌出量预测中。然而,煤矿瓦斯涌出量的影响因素众多且复杂,灰色理论往往不能满足预测精度要求;神经网络的预测精度依赖于样本容量,且存在训练速度慢、泛化能力差的问题;基于支持向量机的预测方法对超参数的选取有较高要求[12]。

随机森林算法具有参数少、学习速度快、适用于高维样本、可有效避免过拟合、预测精度高等优点[13-14],已被广泛用于分类和回归问题。在煤炭瓦斯防治领域,随机森林算法已在瓦斯涌出量预测、煤与瓦斯突出预测等方面有所应用,如汪明等[15]建立了回采工作面瓦斯涌出量的随机森林预测模型。温廷新等[16]提出了一种将因子分析理论与随机森林算法相结合的煤与瓦斯突出等级预测方法。郑晓亮[17]将数据挖掘多重填补算法与随机森林算法相结合进行煤与瓦斯突出预测。本文在文献[15]的基础上,分析了特征变量的影响权重,优选出影响瓦斯涌出量的关键特征变量,建立随机森林回归模型,进行煤矿瓦斯涌出量预测,提高了预测精度和效率。

1　方法原理

1.1　基于随机森林回归的煤矿瓦斯涌出量预测

随机森林算法是以 Bagging 算法思想为基础建立的集成学习算法[18],用于数据的分类和回归研究。该算法能够有效分析高维非线性数据,具有较好的泛化能力和预测性能[14]。随机森林算法用于研究连续变量的回归问题,称之为随机森林回归。

基于随机森林回归的煤矿瓦斯涌出量预测主要步骤如下[19]:

(1) 采用 bootstrap 自助法重采样技术,在原始训练集中有放回地重复随机抽取 n 个与原始训练集样本容量相等的新训练样本数据集,未被采样选中的数据称为袋外数据。

(2) 从影响煤矿瓦斯涌出量的 M 个输入特征中随机选择 m 个特征作为决策树分枝节点的备选特征变量集合,根据分枝优度准则,从集合中选取最优特征进行节点分裂,构建决策树。每棵瓦斯涌出量决策树自由完整地生长而不剪裁。

(3) 构造出 n 颗瓦斯涌出量决策树,组成随机森林回归模型。取 n 颗决策树输出值的均值作为煤矿瓦斯涌出量预测结果,利用袋外数据(如残差平方均值)评价回归模型预测性能。

1.2　超参数寻优及特征变量重要性评估

随机森林回归模型的超参数包括决策树数目 n 和随机选用的特征个数 m,其取值对回归模型的拟合和预测性能有重要影响[20]。若 n 值过小,会因回归模型训练不足而导致预测结果误差偏大;若 n 值过大,则会增加模型的计算量。若 m 值偏小,回归模型会因过拟合而导致预测精度降低;若 m 值偏大,则会降低模型运算速度。因此,构建随机森林回归模型时需对超参数 n 和 m 进行寻优。

对超参数 n 和 m 进行寻优时,首先设置超参数 m 值,回归分析中通常默认其值为输入特征变量数的 1/3[21],在 m 值不变的条件下,得到不同 n 值下袋外数据残差平方均值的变化情况。满足随机森林

回归模型稳定性(残差平方均值随 n 值变化趋于稳定值)且模型训练效率较高时对应决策树数目 n 值为最优值。设 n 取最优值不变,计算不同 m 值下袋外数据残差平方均值和回归模型的拟合优度,选取残差平方均值小、拟合优度高的 m 值作为最优值。

在采用 bootstrap 自助法重采样形成随机森林回归模型训练集的过程中,每次采样约有 36.8%的原始数据不会被取到,即袋外数据,它们将不参与回归模型训练,而是作为测试集对回归模型的泛化能力、预测性能和特征重要性进行评估[22]。

在随机森林回归模型中,通常是以特征变量的随机改变引起回归模型的变化程度来描述特征变量的重要性,具体是采用袋外数据残差平方均值的增加量(IncMSE)来表征特征变量的重要性,IncMSE 越大,说明该特征变量对因变量越重要。对 IncMSE 进行归一化处理,得到各特征变量对因变量的影响权重[23],本文用累计影响权重达到 90%的部分特征变量代替煤矿瓦斯涌出量的全部特征变量,用于构建随机森林回归模型。

2 测试分析

2.1 样本数据

根据文献[7],选用 14 个特征变量作为煤矿瓦斯涌出量预测模型的输入变量:煤层瓦斯含量 X_1,埋深 X_2,煤厚 X_3,煤层倾角 X_4,采煤高度 X_5,日进度 X_6,采煤工作面长度 X_7,采出率 X_8,邻近层瓦斯含量 X_9,邻近层厚度 X_{10},邻近层间距 X_{11},顶板管理方式 X_{12},开采强度 X_{13},层间岩性 X_{14};输出变量:瓦斯涌出量 Y。获取了 20 组原始数据,见表 1。前 16 组原始数据用于训练回归模型,其余 4 组用于测试回归模型。

表 1 随机森林回归模型训练和测试样本

Table 1 Training and testing samples of random forest regression model

序号	X_1/($m^3 \cdot t^{-1}$)	X_2/m	X_3/m	X_4/(°)	X_5/m	X_6/($m \cdot d^{-1}$)	X_7/m	X_8	X_9/($m^3 \cdot t^{-1}$)	X_{10}/m	X_{11}/m	X_{12}	X_{13}/($t \cdot d^{-1}$)	X_{14}	Y/($m^3 \cdot min^{-1}$)
1	1.92	408	2.0	10	2.0	4.42	155	0.960	2.02	1.50	20	1	1	5.03	3.34
2	2.14	421	1.8	11	1.8	4.13	145	0.950	2.64	1.62	19	1	1	4.75	3.56
3	2.58	450	2.3	10	2.3	4.67	150	0.950	2.41	1.48	18	1	2	4.91	3.67
4	2.40	456	2.2	15	2.2	4.51	160	0.940	2.55	1.75	20	1	2	4.63	4.17
5	3.22	516	2.8	13	2.8	3.45	180	0.930	2.21	1.72	12	1	2	4.78	4.60
6	2.80	527	2.5	17	2.5	3.28	180	0.940	2.81	1.81	11	1	1	4.51	4.92
7	3.23	517	2.8	13	2.8	3.46	180	0.930	2.23	1.71	12	1	2	4.76	4.61
8	3.35	531	2.9	9	2.9	3.68	165	0.930	1.88	1.42	13	1	2	1.82	4.78
9	3.61	550	2.9	12	2.9	4.02	155	0.920	2.12	1.60	14	1	2	4.83	5.23
10	3.71	573	3.2	11	3.2	2.92	175	0.910	3.11	1.46	13	1	2	4.63	5.62
11	4.21	590	5.9	8	5.9	2.85	170	0.795	3.40	1.50	18	1	3	4.77	7.24
12	4.03	604	6.2	9	6.2	2.64	180	0.812	3.15	1.80	16	1	3	4.70	7.80
13	4.80	630	6.5	9	6.1	2.77	165	0.785	3.02	1.74	17	1	3	4.62	7.68
14	4.67	640	6.3	11	6.3	2.75	175	0.802	2.56	1.75	15	1	3	4.60	7.95
15	2.43	450	2.7	11	2.7	4.32	165	0.930	2.35	1.85	16	1	2	4.58	5.06
16	3.16	544	2.7	17	2.7	3.81	165	0.930	2.81	1.79	13	1	2	4.90	4.93
17	4.62	629	6.4	13	6.4	2.80	170	0.803	3.35	1.61	19	1	3	4.63	8.04
18	4.53	635	6.2	9	6.2	2.73	160	0.716	2.94	1.73	17	1	3	4.61	7.56
19	3.87	580	3.9	11	3.9	2.85	170	0.920	3.02	1.39	14	1	2	4.72	5.82
20	3.24	509	2.5	14	2.5	4.40	160	0.930	2.79	1.72	13	1	2	4.65	4.36

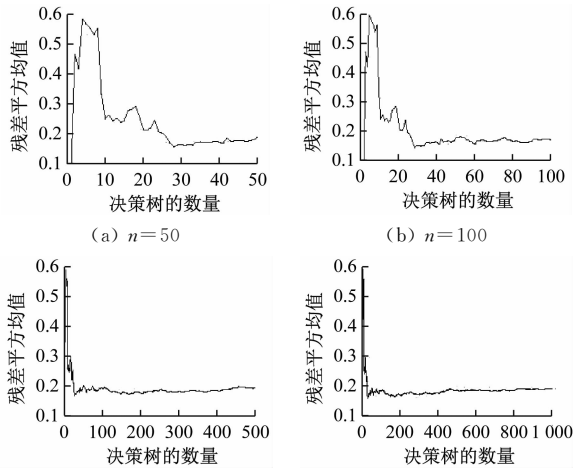
2.2 超参数确定

设参数 m 的默认值为特征变量总数的 1/3,煤矿瓦斯涌出量的影响因子有 14 个,因此,$m=5$。设决策树数目 n 为 50,100,500,1 000,利用 RStudio 软件平台进行编程,建立随机森林回归模型,得到 n 取不同值时袋外数据残差平方均值,如图 1 所示。从图 1 可知,当 n 超过 200 后,袋外数据残差平方均值变化趋于平稳。综合考虑模型运算量和预测精度,取 n 为 500。

取 $n=500$ 固定不变,m 取 1~14(步长为 1),建立随机森林回归模型,得到 m 取不同值时模型的拟合优度与残差平方均值,如图 2 所示。从图 2 可知,m 为 7 时对应拟合优度最大,残差平方均值最小,因此,确定最优值 $m=7$。

2.3 特征变量的选取

对 IncMSE 值进行归一化处理,得出各特征变量对煤矿瓦斯涌出量的影响权重及累计影响权重,如图 3 和图 4 所示。



(a) $n=50$　(b) $n=100$　(c) $n=500$　(d) $n=1\ 000$

图 1　n 取不同值时的袋外数据残差平方均值(14 个特征变量)

Fig. 1　The mean of squared residuals of out-of-bag data when n takes different values (14 characteristic variables)

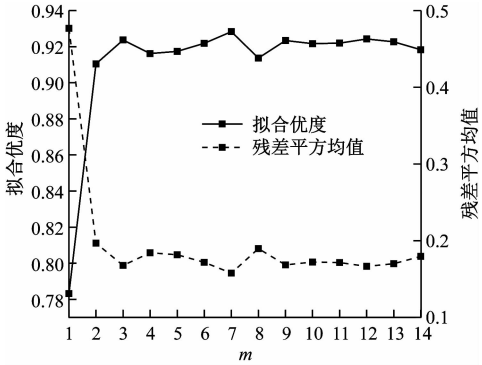


图 2　m 取不同值时的拟合优度与残差平方均值(14 个特征变量)

Fig. 2　Goodness of fit and mean of squared residuals when m takes different values (14 characteristic variables)

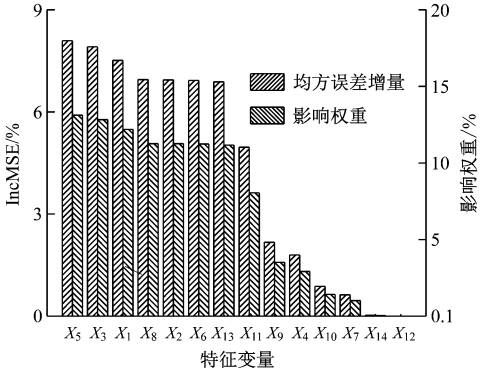


图 3　各特征变量重要性及影响权重

Fig. 3　Importance and influence weights of each characteristic variable

从图 3 可知,采煤高度、煤厚、煤层瓦斯含量、采出率、埋深、日进度、开采强度、邻近层间距等特征变量的影响权重较大,而邻近层瓦斯含量、倾角、采煤工作面长度、层间岩性、顶板管理方式、邻近层厚度等特征变量的影响权重相对较小。从图 4 可看出,采煤高度、煤厚、煤层瓦斯含量、采出率、埋深、日进度、开采强度、邻近层间距等特征变量的累计影响权重为 91.10%,按照累计影响权重达到 90%以上的关键特征变量选择规则,选取该 8 个特征变量作为影响煤矿瓦斯涌出量的主要因素。

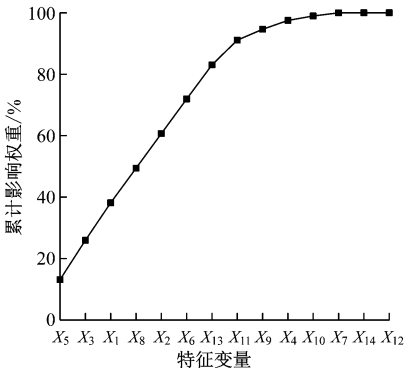
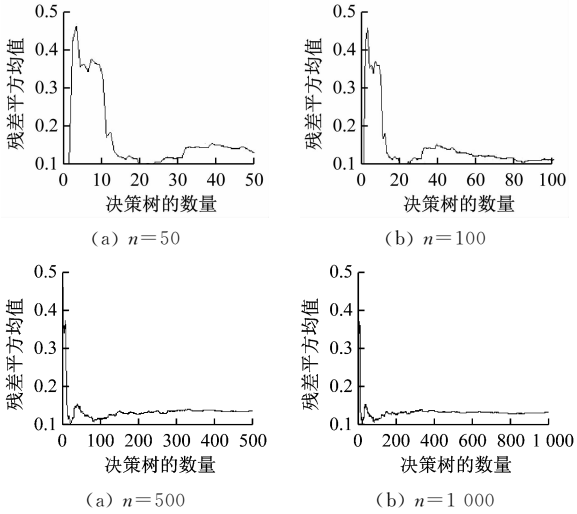


图 4　累计影响权重

Fig. 4　Cumulative influence weight

2.4　预测结果对比

选取重要性排序前 8 的特征变量作为随机森林回归模型的输入变量,在保证训练样本相同的前提下,构建随机森林回归模型。首先进行超参数寻优,结果如图 5 和图 6 所示。



(a) $n=50$　(b) $n=100$　(a) $n=500$　(b) $n=1\ 000$

图 5　n 取不同值时的袋外数据残差平方均值(8 个特征变量)

Fig. 5　The mean of squared residuals of out-of-bag data when n takes different values (8 characteristic variables)

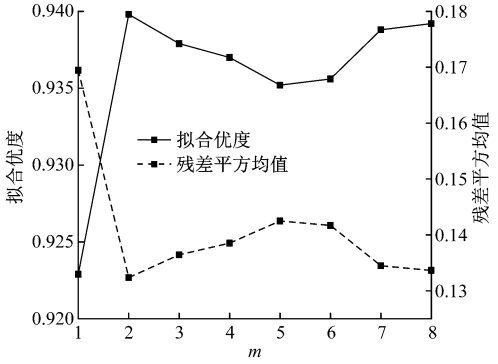


图 6　m 取不同值时的拟合优度与残差平方均值(8 个特征变量)

Fig. 6　Goodness of fit and mean of squared residuals when m takes different values (8 characteristic variables)

从图 5、图 6 可看出：n 为 500 时，模型袋外数据残差平方均值变化平稳；m 为 2 时，模型拟合优度最大，残差平方均值最小。因此，基于特征选择的随机森林回归模型的最优超参数为 $n=500$，$m=2$。

为了评价基于特征变量选择的随机森林回归模型的预测性能，在保证预测样本相同的前提下，与全部 14 个特征变量参与的随机森林回归模型的预测结果进行对比，结果见表 2。可看出，2 种情况下随机森林回归模型均具有较好的预测性能，进行特征变量选择后，平均绝对误差由 0.22 m^3/min 下降到 0.21 m^3/min，平均相对误差由 3.55%下降到 3.47%。基于特征变量选择的随机森林回归模型不仅能够保证较好的预测性能，而且降低了模型特征变量的维度，减少了原始数据获取工作，提高了预测效率。

表 2 瓦斯涌出量预测误差对比

Table 2 Comparison of gas emission prediction error

样本序号	实测值/($m^3 \cdot min^{-1}$)	全部 14 个特征变量			8 个特征变量		
		预测值/($m^3 \cdot min^{-1}$)	绝对误差/($m^3 \cdot min^{-1}$)	相对误差/%	预测值/($m^3 \cdot min^{-1}$)	绝对误差/($m^3 \cdot min^{-1}$)	相对误差/%
17	8.04	7.53	−0.51	6.32	7.65	−0.39	4.88
18	7.56	7.60	0.04	0.49	7.66	0.10	1.39
19	5.82	5.78	−0.04	0.77	5.79	−0.03	0.56
20	4.36	4.65	0.29	6.60	4.67	0.31	7.04
平均值			0.22	3.55		0.21	3.47

3 结论

(1) 研究了基于随机森林回归的煤矿瓦斯涌出量预测方法。通过计算袋外数据残差平方均值和拟合优度，确定随机森林回归模型的最优超参数。

(2) 通过特征变量重要性分析方法，从全部 14 个特征变量中优选出采煤高度、煤厚、煤层瓦斯含量、采出率、埋深、日进度、开采强度、邻近层间距 8 个关键的特征变量，建立特征选择后的随机森林预测模型。

(3) 测试结果表明，采用全部特征变量和部分特征变量的随机森林回归模型均具有较好的拟合与预测效果。进行特征变量选择后，模型的平均绝对误差由 0.22 m^3/min 下降到 0.21 m^3/min，平均相对误差由 3.55%下降到 3.47%。基于特征变量优选的随机森林回归模型保持了较高的预测精度，同时提高了模型预测效率，更加适用于煤矿瓦斯涌出量预测。

参考文献(References)：

[1] 孙继平. 煤与瓦斯突出报警方法[J]. 工矿自动化，2014，40(11)：1-5.
SUN Jiping. Alarm methods of coal and gas outburst [J]. Industry and Mine Automation，2014，40(11)：1-5.

[2] 施式亮，李润求，罗文柯. 基于 EMD-PSO-SVM 的煤矿瓦斯涌出量预测方法及应用[J]. 中国安全科学学报，2014，24(7)：43-49.
SHI Shiliang，LI Runqiu，LUO Wenke. Method for predicting coal mine gas emission based on EMD-PSO-SVM and its application [J]. China Safety Science Journal，2014，24(7)：43-49.

[3] 代巍，付华，冀常鹏，等. 回采工作面瓦斯涌出量 VMD-DE-RVM 区间预测方法[J]. 中国安全科学学报，2018，28(9)：109-115.
DAI Wei，FU Hua，JI Changpeng，et al. Interval prediction method for gas emission from coal mining face based on VMD-DE-RVM [J]. China Safety Science Journal，2018，28(9)：109-115.

[4] 周西华，孙家正. 基于主因子分析的改进 BP 神经网络瓦斯涌出量预测[J]. 矿业安全与环保，2018，45(6)：43-47.
ZHOU Xihua，SUN Jiazheng. Prediction of gas emission based on principal factor analysis and improved BP neural network[J]. Mining Safety & Environmental Protection，2018，45(6)：43-47.

[5] 题正义，杨艳国，丁涛. 瓦斯涌出量的模糊数学与灰色系统理论的预测[J]. 辽宁工程技术大学学报(自然科学版)，2000，19(2)：126-129.
TI Zhengyi，YANG Yanguo，DING Tao. The application research of fuzzy-grey system theory in the prediction of mine gas emission[J]. Journal of Liaoning Technical University (Natural Science Edition)，2000，19(2)：126-129.

[6] 李树刚，马彦阳，林海飞，等. 基于因子分析法的瓦斯涌出量预测指标选取[J]. 西安科技大学学报，2017，37(4)：461-466.
LI Shugang，MA Yanyang，LIN Haifei，et al. Selection of gas emission prediction index based on factor analysis[J]. Journal of Xi'an University of Science and Technology，2017，37(4)：461-466.

[7] 付华，谢森，徐耀松，等. 基于 ACC-ENN 算法的煤矿瓦斯涌出量动态预测模型研究[J]. 煤炭学报，2014，39(7)：1296-1301.
FU Hua，XIE Sen，XU Yaosong，et al. Gas emission

dynamic prediction model of coal mine based on ACC-ENN algorithm[J]. Journal of China Coal Society, 2014,39(7):1296-1301.

[8] 孙林,杨世元. 基于 LS-SVM 的回采工作面瓦斯涌出量预测[J]. 煤炭学报,2008,33(12):1377-1380.
SUN Lin, YANG Shiyuan. Prediction for gas emission quantity of the working face based on LS-SVM[J]. Journal of China Coal Society, 2008, 33(12): 1377-1380.

[9] 董晓雷,贾进章,白洋,等. 基于 SVM 耦合遗传算法的回采工作面瓦斯涌出量预测[J]. 安全与环境学报,2016,16(2):114-118.
DONG Xiaolei, JIA Jinzhang, BAI Yang, et al. Prediction for gas-gushing amount from the working face of stope based on the SVM coupling genetic algorithm[J]. Journal of Safety and Environment, 2016,16(2):114-118.

[10] 张强,贾宝山,董晓雷,等. PCA-GA-SVM 的回采工作面瓦斯涌出量预测[J]. 辽宁工程技术大学学报(自然科学版),2015,34(5):572-577.
ZHANG Qiang, JIA Baoshan, DONG Xiaolei, et al. Working face gas emission prediction based on PCA-GA-SVM[J]. Journal of Liaoning Technical University(Natural Science),2015,34(5):572-577.

[11] 谢东海,冯涛,朱川曲. 回采工作面瓦斯涌出量的熵权均值属性测度模型及其应用[J]. 中南大学学报(自然科学版),2013,44(6):2482-2487.
XIE Donghai, FENG Tao, ZHU Chuanqu. Attributed measurement prediction model of entropy value for coal face gas emission and its application[J]. Journal of Central South University (Science and Technology), 2013,44(6):2482-2487.

[12] 邓军,雷昌奎,曹凯,等. 采空区煤自燃预测的随机森林方法[J]. 煤炭学报,2018,43(10):2800-2808.
DENG Jun, LEI Changkui, CAO Kai, et al. Random forest method for predicting coal spontaneous combustion in gob[J]. Journal of China Coal Society, 2018,43(10):2800-2808.

[13] 张华伟,王明文,甘丽新. 基于随机森林的文本分类模型研究[J]. 山东大学学报(理学版),2006,41(3):139-143.
ZHANG Huawei, WANG Mingwen, GAN Lixin. Automatic text classification model based on random forest[J]. Journal of Shandong University(Natural Science),2006,41(3):139-143.

[14] 曹泽涛,方子东,姚瑾,等. 基于随机森林的黄土地貌分类研究[J]. 地球信息科学学报,2020,22(3):452-463.
CAO Zetao, FANG Zidong, YAO Jin, et al. Loess landform classification based on random forest[J]. Journal of Geo-information Science, 2020, 22(3): 452-463.

[15] 汪明,王建军. 基于随机森林的回采工作面瓦斯涌出量预测模型[J]. 煤矿安全,2012,43(8):182-185.
WANG Ming, WANG Jianjun. Gas emission prediction model of stope based on random forests[J]. Safety in Coal Mines,2012,43(8):182-185.

[16] 温廷新,张波,邵良杉. 煤与瓦斯突出预测的随机森林模型[J]. 计算机工程与应用,2014,50(10):233-237.
WEN Tingxin, ZHANG Bo, SHAO Liangshan. Prediction of coal and gas outburst based on random forest model[J]. Computer Engineering and Applications,2014,50(10):233-237.

[17] 郑晓亮. 基于瓦斯含量法的煤与瓦斯突出预测关键技术研究[D]. 淮南:安徽理工大学,2018.
ZHENG Xiaoliang. Research on key technology of coal and gas outburst prediction based on gas content method[D]. Huainan: Anhui University of Science & Technology,2018.

[18] BREIMAN L. Random forests[J]. Machine Learning, 2001,45(1):5-32.

[19] 王俊红,杨赛,刘富成,等. 基于随机森林算法的混凝土早期抗裂性预测研究[J]. 信阳师范学院学报(自然科学版),2021,34(1):158-165.
WANG Junhong, YANG Sai, LIU Fucheng, et al. Prediction of early crack resistance of concrete based on random forest algorithm[J]. Journal of Xinyang Normal University(Natural Science Edition), 2021, 34(1):158-165.

[20] 崔东文,金波. 基于随机森林回归算法的水生态文明综合评价[J]. 水利水电科技进展,2014,34(5):56-60.
CUI Dongwen, JIN Bo. Comprehensive evaluation of water ecological civilization based on random forests regression algorithm[J]. Advances in Science and Technology of Water Resources,2014,34(5):56-60.

[21] LIAW A, WIENER M. Classification and regression by random forest[J]. R News,2002,2/3: 18-22.

[22] 朱庆忠,胡秋嘉,杜海为,等. 基于随机森林算法的煤层气直井产气量模型[J]. 煤炭学报,2020,45(8):2846-2855.
ZHU Qingzhong, HU Qiujia, DU Haiwei, et al. A gas production model of vertical coalbed methane well based on random forest algorithm[J]. Journal of China Coal Society,2020,45(8):2846-2855.

[23] 刘继辉,许磊,马晓龙,等. 基于随机森林回归的制丝过程参数影响权重分析[J]. 烟草科技,2017,50(2):63-71.
LIU Jihui, XU Lei, MA Xiaolong, et al. Weight analysis of primary processing parameters based on random forest regression[J]. Tobacco Science & Technology, 2017,50(2):63-71.