

引用格式:刘杰,肖聪,张士诚,等.基于改进Stacking融合模型的煤层气含气量预测方法[J].油气藏评价与开发,2026,16(5):983-991.

Liu Jie, Xiao Cong, Zhang Shicheng, et al. A coalbed methane content prediction method based on an improved Stacking ensemble model[J].

Petroleum Reservoir Evaluation and Development, 2026, 16(5): 983-991.

DOI: 10.13809/j.cnki.cn32-1825/te.2025291

基于改进Stacking融合模型的煤层气含气量预测方法

刘杰¹, 肖聪¹, 张士诚¹, 贺甲元²

(1. 中国石油大学(北京)石油工程学院, 北京 102249; 2. 中国石化石油勘探开发研究院, 北京 102206)

摘要:传统解吸法测定煤层气含气量存在成本高、周期长、区域覆盖难等问题,难以满足经济可采性快速评价与降本增效的需求。为此,提出一种基于改进Stacking融合模型的煤层气含气量预测方法。该方法通过网格搜索优化支持向量机、随机森林、极端梯度提升、轻量梯度提升机和多层感知机5种异构基模型的超参数,创新地将原始训练集特征与各基模型预测结果进行加权平均,构建新的元训练集,输入线性回归元模型进行二次学习,实现原始特征与基模型预测信息的双重融合。将模型应用于沁水盆地南部煤层气区块,结果表明:改进Stacking融合模型的测试集决定系数达0.94,较最优单一基模型和标准Stacking模型分别提升10.64%和6.38%;平均绝对误差、均方根误差和均方误差较标准Stacking模型分别降低14.3%、25.2%和43.7%。SHAP(是一种用于解释机器学习模型预测的方法)可解释分析表明,煤层斜深、自然电位和自然伽马是影响含气量预测的主要特征。现场应用显示,8口井预测结果与实测值吻合良好,平均绝对误差小于0.84 m³/t,平均相对误差控制在6%以内。该模型显著提升了煤层气含气量预测的精度与可靠性,为煤层气资源选区评价和高效开发提供了一种新的技术手段。

关键词:煤层气;含气量;机器学习;Stacking融合模型;可解释分析

中图分类号:TE35

文献标识码:A

A coalbed methane content prediction method based on an improved Stacking ensemble model

Liu Jie¹, Xiao Cong¹, Zhang Shicheng¹, He Jiayuan²

(1. College of Petroleum Engineering, China University of Petroleum (Beijing), Beijing 102249, China; 2. Sinopec Petroleum Exploration and Production Research Institute, Beijing 102206, China)

Abstract: The traditional desorption method for determining coalbed methane (CBM) gas content suffers from high costs, long cycles, and limited regional coverage, making it unsuitable for rapid economic recoverability assessments, cost reduction, and efficiency improvement. To address this, a CBM gas content prediction method based on an improved Stacking ensemble algorithm was proposed. This method optimizes the hyperparameters of five heterogeneous base models—support vector machine (SVM), random forest (RF), XGBoost, LightGBM, and multi-layer perceptron (MLP)—through grid search. It constructed a new meta-training set by innovatively weighting and averaging the raw training-set features and prediction results from each base model, which was then input into a linear regression meta-model for secondary learning. This achieved dual integration of raw features and base-model prediction information. The model was applied to CBM blocks in the southern Qinshui Basin. The results demonstrated that the improved Stacking ensemble model achieved a coefficient of determination of 0.94 on the test set, representing improvements of 10.64% and 6.38% over the optimal single base model and standard Stacking model, respectively. Compared with the standard Stacking model, the mean absolute error, root mean square error, and mean squared error are reduced by 14.3%, 25.2%, and 43.7%, respectively. SHAP (SHapley Additive exPlanations, a method for interpreting machine learning model predictions) interpretability analysis indicated that coal seam slant depth, spontaneous potential, and natural gamma ray are the primary features influencing gas content prediction. Field application demonstrated good agreement between predicted and measured values for eight wells, with a mean absolute error below 0.84 m³/t and a mean relative error controlled within 6%. This model significantly enhances the accuracy and reliability of CBM gas content prediction, providing a new technical approach for CBM resource area selection and efficient development.

Keywords: coalbed methane (CBM); gas content; machine learning; Stacking ensemble model; interpretability analysis

收稿日期:2025-10-17。

第一作者简介:刘杰(2000—),男,在读博士研究生,从事人工智能在压裂增产改造领域的研究。地址:北京市昌平区府学路18号,邮政编码:102249。E-mail:18323672600@163.com

通信作者简介:肖聪(1989—),男,博士,副教授,从事深度学习智能反演优化理论与方法、非常规油气藏智能压裂理论与方法研究。地址:北京市昌平区府学路18号,邮政编码:102249。E-mail:18810907235@163.com

基金项目:国家自然科学基金项目“页岩油平台井闷井压力干扰响应机理与智能诊断方法研究”(52304055);安徽省自然资源厅地质工作项目“两淮矿区重点区域煤层气储层可改造性调查评价”(ZX20240114)。

随着全球能源需求持续攀升以及对清洁能源日益增长的需求,煤层气作为一类关键的非常规天然气资源,其勘探与开发受到广泛关注。煤层含气量是评估和开发此类资源的核心参数,不仅对于资源总量估算至关重要,同时也是识别优质开采区域的基础^[1-4]。目前,煤层含气量参数的获取主要通过钻孔取心测试、物理模拟、数值模拟及定量预测等方法,其中物理模拟主要采用校正系数、气体相态、带压逐级降压解吸等方法,定量预测包括含气梯度法、煤级-压力-等温吸附曲线法、地球物理解释技术^[5-7]。然而,传统解吸法测定煤层气含量成本高昂且难以实现全面覆盖,地球物理测井技术作为一种有效的替代方案,能够提供反映煤层岩石性质、物理特性和导电性能等多方面信息的数据集,逐渐成为实验室分析之外的重要选择^[8]。尽管如此,由于煤层内部结构的复杂性和气体状态的不确定性,现有预测方法仍需进一步优化与完善^[9]。

近年来,机器学习技术在油气开发及地质资源预测领域的广泛应用取得了较大成效。连承波等^[10]构建了BP(反向传播)神经网络模型对煤层气含量进行预测,但在大规模搜索空间中易陷入局部最小值,限制了预测精度^[11]。胡锦宏^[12]采用遗传算法优化神经网络,提高了网络训练的准确性,构建了基于GA-BP(遗传算法-反向传播)神经网络的煤层气含气量测井评价智能模型。同时,向旻等^[13]的研究显示,神经网络在煤层含气量预测中的表现优于多元回归及Langmuir方程,进一步证实了机器学习在煤层气含气量计算中的巨大潜力。集成学习方法因其出色的预测性能和良好的泛化能力在机器学习领域广受关注,梁海波等^[14]基于Stacking的集成学习方法在孔隙度预测中实现了较单一模型更优的准确性;刘磊等^[15]提出的分频地震属性Stacking融合方法,其预测结果与砂体厚度的相关系数高达0.92,且融合属性特征的空间展布形态更为清晰,充分证明了该方法在复杂地质

参数预测中的可靠性。然而,传统的Stacking方法在处理高维、非线性数据时仍存在一定的局限性。

以沁水盆地南部煤层气藏为研究对象,该区相较于盆地内已开发的樊庄、潘庄等区块^[16-17],具有埋深更大、渗透率更低及非均质性更强等特点,对勘探技术与预测模型提出了更高要求。基于区内系统的煤层气探井资料,结合钻压取心测试与测井解释数据,构建了含气量预测样本库,并提出一种改进的Stacking融合预测方法。该方法通过对基模型预测结果进行误差加权,并将优化后的预测值与原始训练集重构为新的训练样本,输入元模型进行二次学习,从而有效捕捉原始数据与新生成特征间的潜在关联。改进融合模型整合了多算法优势,通过结构优化与参数调校显著提升了预测精度与适应性,不仅为煤层气含气量预测提供了新技术途径,也为机器学习在该领域的应用提供了重要参考。

1 机器学习方法

1.1 基模型及元模型

在Stacking融合模型^[18]中,基模型和元模型的选择很关键。基模型的选择遵循“好且多样性”的原则,即考虑单个基模型性能好坏的同时,也要考虑单个基模型结构不相似性,能够捕捉到不同方面的数据特征。研究从K近邻、装袋法、AdaBoost、XGBoost、随机森林、轻量梯度提升机、多层感知机、岭回归、决策树回归及支持向量机10个候选模型中,通过初步实验筛选出支持向量机、随机森林、极端梯度提升、轻量梯度提升机与多层感知机5个表现最优的模型作为基模型。这些模型覆盖了线性、非线性、树结构以及神经网络等多样性,确保融合效果,各模型核心机理与特性详见表1,而元模型则可以选择简单且效果良好的线性回归模型,用来融合基模型的预测结果。

表1 单一机器学习模型优缺点对比分析
Table 1 Comparative analysis of advantages and disadvantages of single machine learning models

算法	算法机理	优点	缺点
支持向量机(SVM)	通过核函数映射高维空间,构建最优超平面,最小化预测误差并保留边界	有多种核函数,灵活处理非线性关系;不易过拟合	不适用大量样本;可解释度较低;参数调整困难
随机森林(RF)	构建多棵独立决策树,随机选择特征,通过投票或平均输出结果	可处理高维数据;并行计算,计算速度较快	对噪声和异常值敏感;低维数据预测性能较差
极端梯度提升(XGBoost)	基于梯度提升树(GBDT)改进,引入正则化项,采用二阶泰勒展开优化损失函数	能处理缺失值;引入正则化防止过拟合	对超参数选择敏感
轻量梯度提升机(LightGBM)	基于GBDT,采用直方图优化与层次学习策略,通过梯度提升优化损失函数	良好的预测性能和泛化能力;对噪声和异常值较稳定	需要调整多个超参数;训练时间较长
多层感知机(MLP)	由输入层、隐藏层、输出层构成的前馈神经网络,通过全连接结构捕捉高维非线性信息	适用于处理数据非线性关系及复杂高维问题	对初始占比和特征选择较敏感,容易陷入局部最优解

1.2 标准融合模型

Stacking方法最早由 Wolpert^[19]于1992年提出,是一种基于堆叠结构的集成学习算法。其核心流程采用 k 折交叉验证构建多层预测结构:首先,将训练集划分为 k 折,依次以 $k-1$ 折作为训练数据,剩余1折用于生成各基模型的预测结果;随后,将各基模型在所有折上的预测结果组合为第二层(元模型)的训练特征。同时,对原始测试集进行预测并取均值,构成元模型的测试集。最终,基于基模型输出所构建的二级特征矩阵,训练元模型以得到集成预测结果。Stacking融合模型的核心优势在于能够综合利用多个异构基模型对原始数据的不同解析角度,通过集成其优势提升整体预测性能。该模型结构见图1。

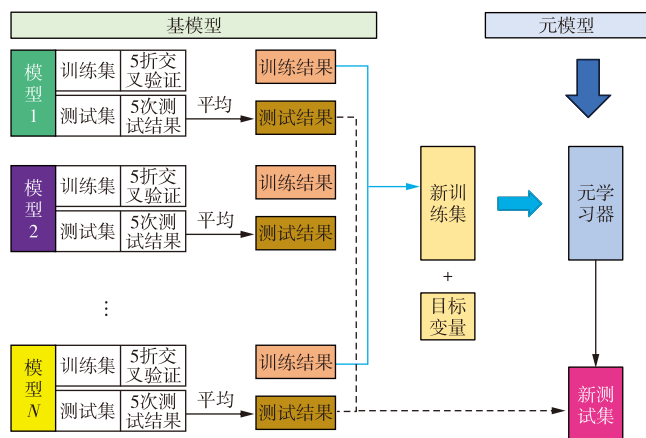


图1 标准融合模型结构示意图

Fig. 1 Schematic diagram of standard ensemble model structure

1.3 改进融合模型

标准的Stacking框架中,第二层级的元模型训练所采用的数据集,是由第一层级各基模型的预测输出构建而成。这样,每个基模型的数量就对应着输入特征的数量。这种方法虽然在一定程度上能够有效防止模型的过拟合,但也可能导致原始训练集中部分信息的丢失。此外,各基模型的预测性能存在差异,为了更好地提取基模型中的有效信息,根据第一层中基模型的表现,对其输出结果进行加权平均。因此,为了进一步提升Stacking模型的预测精度,提出了一种改进的方法,具体流程如下:

1) 数据集划分:将样本数据集按照80:20的比例划分为训练集 X_{train} 和测试集 X_{test} 。

2) 基模型训练与预测:将数据集投入到 N 个基模型中,采用5折交叉验证进行训练和预测,并计算各基模型的均方误差(MSE)值作为性能评价指标。

3) 权重计算与归一化:根据各基模型的均方误差值计算权重,误差越小,权重越大。权重计算公式为:

$$w_i = \frac{1}{\sigma_{\text{MSE}_i}} \quad (1)$$

式中: w_i 为各模型权重; σ_{MSE_i} 为第 i 个基模型的均方误差值,单位 m^3/t 。

为避免权重绝对值过大或过小,对权重进行归一化,确保所有权重之和为1。

$$w'_i = \frac{w_i}{\sum_{i=1}^N w_i} \quad (2)$$

式中: w'_i 为各模型归一化后的权重; N 为基模型数量。

归一化后,权重直接反映各基模型在融合中的相对贡献比例,形成新的训练集 X_{Ntrain} 和测试集 X_{Ntest} 。

$$X_{\text{Ntrain}} = \sum_{i=1}^N w'_i y_{\text{train}_i} \quad (3)$$

$$X_{\text{Ntest}} = \sum_{i=1}^N w'_i y_{\text{test}_i} \quad (4)$$

式中: y_{train_i} 为第 i 基模型在训练集的预测值,单位 m^3/t ; y_{test_i} 为第 i 基模型在测试集的预测值,单位 m^3/t 。

4) 重构元模型训练集:把新的训练集 X_{Ntrain} 和测试集 X_{Ntest} 与原始数据集 X_{train} 和 X_{test} 组合为最终的训练集和测试集。

5) 元模型训练与预测:将最终形式的训练集作为输入重新训练学习器的权重。使用训练好的元模型对测试集进行预测,得到最终任务的预测结果。

这种方式使得次元学习器能够捕捉到原始训练集与新生成训练集之间的潜在关联,突破传统Stacking仅依赖基模型输出的局限,加强了元模型对原始特征参数间数据分布及关联的适应性,有助于提升模型的预测性能。改进的Stacking融合模型结构如图2所示。

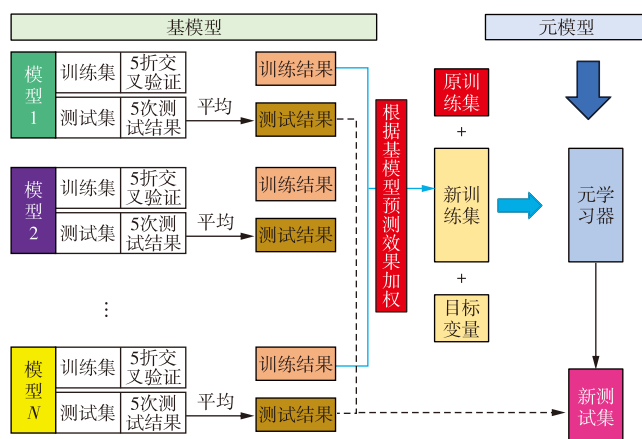


图2 改进Stacking融合模型结构

Fig. 2 Improved Stacking ensemble model structure

2 基于改进Stacking集成学习的煤层气含气量预测

2.1 数据集构建

通过对沁水盆地南部煤层气区块的钻压取心测试数据及常规测井数据进行系统整理,初步构建了含气量预测的原始数据集。采用箱线图法识别并剔除异常值。同时考虑沁水盆地煤层气赋存机制在埋深超过1 200~1 400 m后可能由吸附气主导转变为吸附气与游离气共存。为减少不同赋存机制对模型学习的干扰,使模型更专注于吸附气主导的含气量预测,剔除了含气量异常高样本。最终筛选出277条有效样本构成建模数据集。清洗后,各特征参

数的统计分布如图3所示,煤层气含气量介于0.93~27.16 m³/t,更符合中浅层吸附气为主的分布特征。

输入特征参数的选择基于煤储层气体吸附赋存机理与区块低成本常规测井数据的实际情况。优选了对煤岩属性与储层条件响应敏感的参数:声波时差直接响应煤体压缩性与孔隙度;密度间接反映灰分含量及吸附能力;煤层斜深直接关联控制吸附平衡的地层压力与温度;自然伽马可有效指示降低吸附能力的黏土矿物含量;自然电位在区块内,其负异常幅度的增大与煤层纯度呈正相关,而纯净煤通常具备更强的甲烷吸附能力,故自然电位可作为间接指示含气量潜力的有效参数。

研究最终以煤层斜深、自然伽马、自然电位、声波时差和密度5个参数作为模型输入的特征变量,以实测含气量作为预测目标变量^[20-21]。

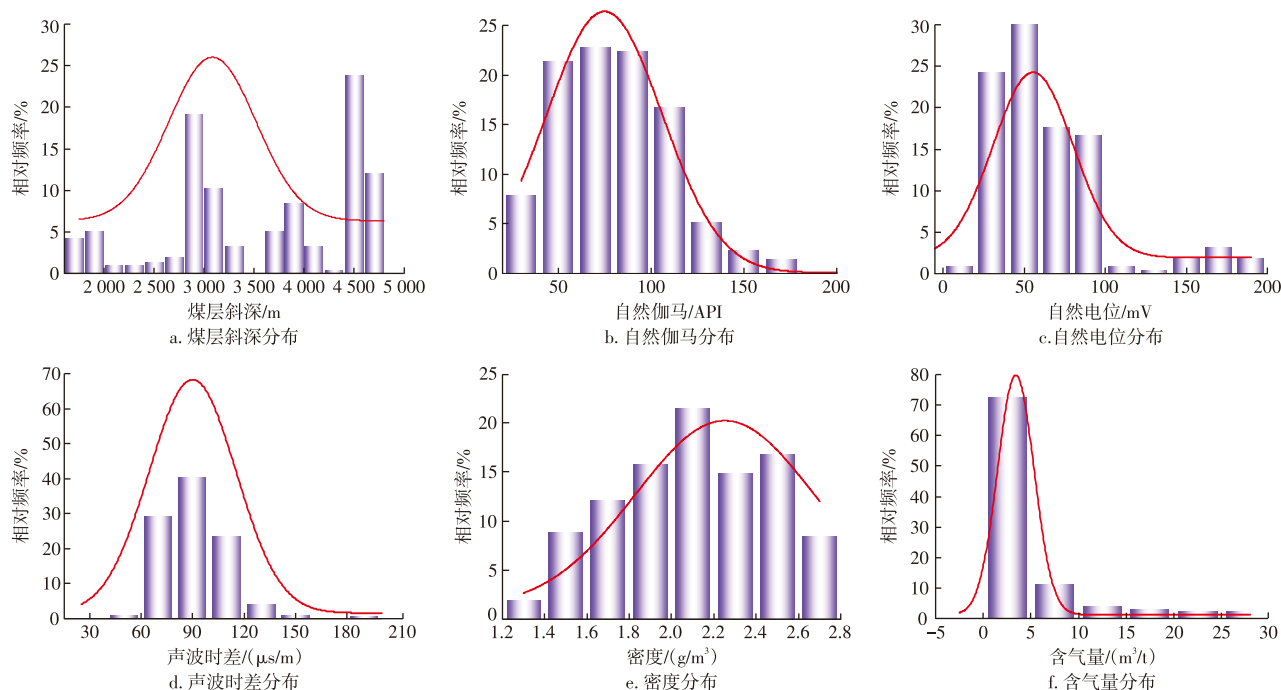


图3 特征参数数据统计

Fig. 3 Characteristic parameter statistics

2.2 数据预处理

特征参数分布不均匀,在量级上存在较大的差异。因此,为了消除量纲影响,加快算法的收敛速度和提升模型的精准度,在进行模型训练前对数据进行归一化处理,公式如下:

$$x = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (5)$$

式中: x 为归一化后的特征参数数据; x_i 为特征参数第 i 个样本数据; $x_{\max}$ 为特征参数数据最大值; $x_{\min}$ 为特征参数数据最小值。

为规避多重共线性导致的过拟合风险,采用皮尔逊相关系数法分析特征间相关性。

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (6)$$

式中: r 为相关系数; y_i 为第 i 个样本的含气量,单位m³/t; $\bar{x}$ 为特征参数平均值; $\bar{y}$ 为含气量平均值,单位m³/t; n 为样本数量。

测井参数与含气量之间以及各测井参数彼此间均存在一定关联。为直观揭示测井参数与含气量之间的相关

性,绘制了相关系数热图(图4)。结果显示,最大相关系数为0.32,最小相关系数为-0.26,表明所选特征参数间冗余性较低。因此,可将密度、声波时差、自然电位、自然伽马及煤层斜深作为预测含气量的关键特征参数。单因素分析表明,声波时差与含气量相关性最高,其次为煤层斜深。尽管密度与含气量相关系数为0,但其与声波时差、煤层深度及自然电位仍存在弱相关。因此,在构建含气量预测的机器学习模型时,密度仍可作为输入特征之一。

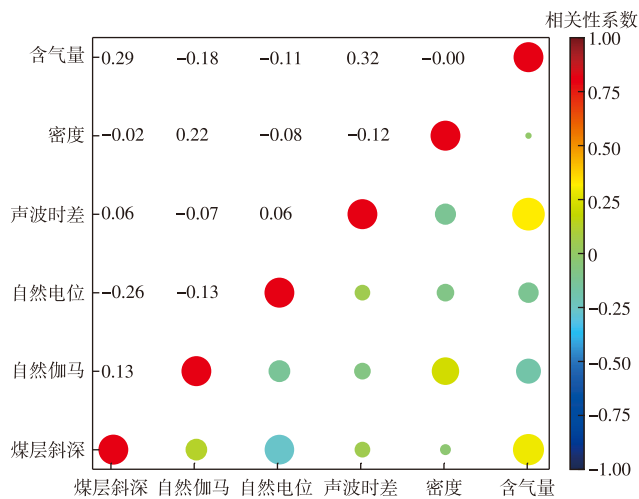


图4 特征参数相关性矩阵
Fig. 4 Correlation matrix of feature parameters

2.3 评价指标及超参数优化

为评估含气量预测模型泛化性能,实验将数据集按80:20划分为训练集和测试集,其结果可视为泛化能力的保守估计。选取均方误差(MSE)值、均方根误差(RMSE)值、平均绝对误差(MAE)值、决定系数(R^2)4个评价指标,综合评估模型对煤层气含气量的预测能力,其中MSE值、RMSE值、MAE值用来检验模型预测精度,其值越小越好, R^2 用来检验模型的拟合优度,值越接近于1效果越好,具体计算公式如下:

$$\sigma_{\text{MSE}} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \tag{7}$$

$$\sigma_{\text{RMSE}} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \tag{8}$$

$$\sigma_{\text{MAE}} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \tag{9}$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \tag{10}$$

式中: R^2 为决定系数; σ_{MSE} 为均方误差值,单位 m^3/t ; σ_{RMSE} 为均方根误差值,单位 m^3/t ; σ_{MAE} 为平均绝对误差值,单位 m^3/t ; $\hat{y}_i$ 为第*i*个样本的含气量预测值,单位 m^3/t 。

模型性能亦受超参数影响,采用网格搜索法^[22]进行优化。超参数的初始范围基于算法原理与预实验设定,通过多轮迭代逐步收敛,最终确定表2所示的优化组合。

表2 基模型网格搜索最优超参数组合
Table 2 Optimal hyperparameter combinations for base model grid search

算法	算法超参数	优化组合	最优组合
支持向量机(SVM)	惩罚系数 <i>C</i> 值	[100, 200, 300, 400]	200
	核函数 gamma(核函数宽度)值	高斯核函数,线性函数 [scale, auto]	高斯核函数 auto
随机森林(RF)	决策树个数	[200, 300, 400, 500]	400
	最大深度	[4, 5, 6, 7, 8]	7
	最小分离样本数	[2, 4, 4, 5]	4
极端梯度提升(XGB)	决策树个数	[300, 500, 600, 700]	600
	学习率	[0.01, 0.10]	0.10
	最大深度	[3, 5, 6, 7]	5
轻量梯度提升机(LGBM)	决策树个数	[100, 200, 300, 500]	200
	学习率	[0.01, 0.10]	0.10
	最大深度	[3, 6, 7, 8]	7
多层感知机(MLP)	层数	[1, 2, 3]	2
	神经元数量	[32, 64, 128, 160]	128
	激活函数	[relu, tanh]	relu

2.4 模型性能对比与结果分析

研究基于SVM、RF、XGB、LGBM和MLP这5种机器学习模型预测煤层气含气量,采用网格搜索方法优化超参数,并基于所述4种评价指标评估模型性能,结果见表3。在单一基模型中,LGBM模型表现最优,测试集上 R^2 为0.84,MAE值、RMSE值和MSE值分别为1.33、1.84和3.39;XGB模型次之,相应指标为0.82、1.52、2.25和5.06;MLP、RF与SVM模型性能相对较弱, R^2 分别为0.76、0.75和0.60。LGBM作为一种基于梯度提升决策树的算法,凭借其高效的直方图优化与层次学习策略,在任务中展现出较高的训练与预测效率及精度。

表3 不同煤层气含气量预测模型的测试集评价指标
Table 3 Test set valuation criteria for different coalbed methane gas content prediction models

预测模型	$\sigma_{\text{MAE}}/(\text{m}^3/\text{t})$	$\sigma_{\text{RMSE}}/(\text{m}^3/\text{t})$	$\sigma_{\text{MSE}}/(\text{m}^3/\text{t})$	R^2
SVM	2.86	3.61	13.05	0.60
RF	2.31	2.90	8.42	0.75
XGB	1.52	2.25	5.06	0.82
LGBM	1.33	1.84	3.39	0.84
MLP	1.84	2.35	5.52	0.76
标准Stacking	0.91	1.47	2.15	0.88
改进Stacking	0.78	1.10	1.21	0.94

进一步构建基于上述5种模型的Stacking融合模型,其预测性能显著优于任一单一模型。该融合模型在验证集上 R^2 达0.88,较各基模型提升4.55%~31.80%;MAE值、RMSE值和MSE值分别为0.91、1.47和2.15,均较基模型有不同程度下降,表明Stacking能有效整合各模型优势,提升预测精度。

图5展示了各模型在测试集上的预测效果,提出的

改进Stacking融合模型取得最佳性能, R^2 为0.94,MAE值、RMSE值和MSE值分别为0.78、1.10和1.21。相较于标准Stacking模型, R^2 提升6.38%,其余指标分别下降14.3%、25.2%和43.7%。改进Stacking模型通过元模型捕捉原始训练集与新生成训练集间的潜在关联,有效提升了预测精度,验证了模型在煤层气含气量预测中的有效性与应用潜力。

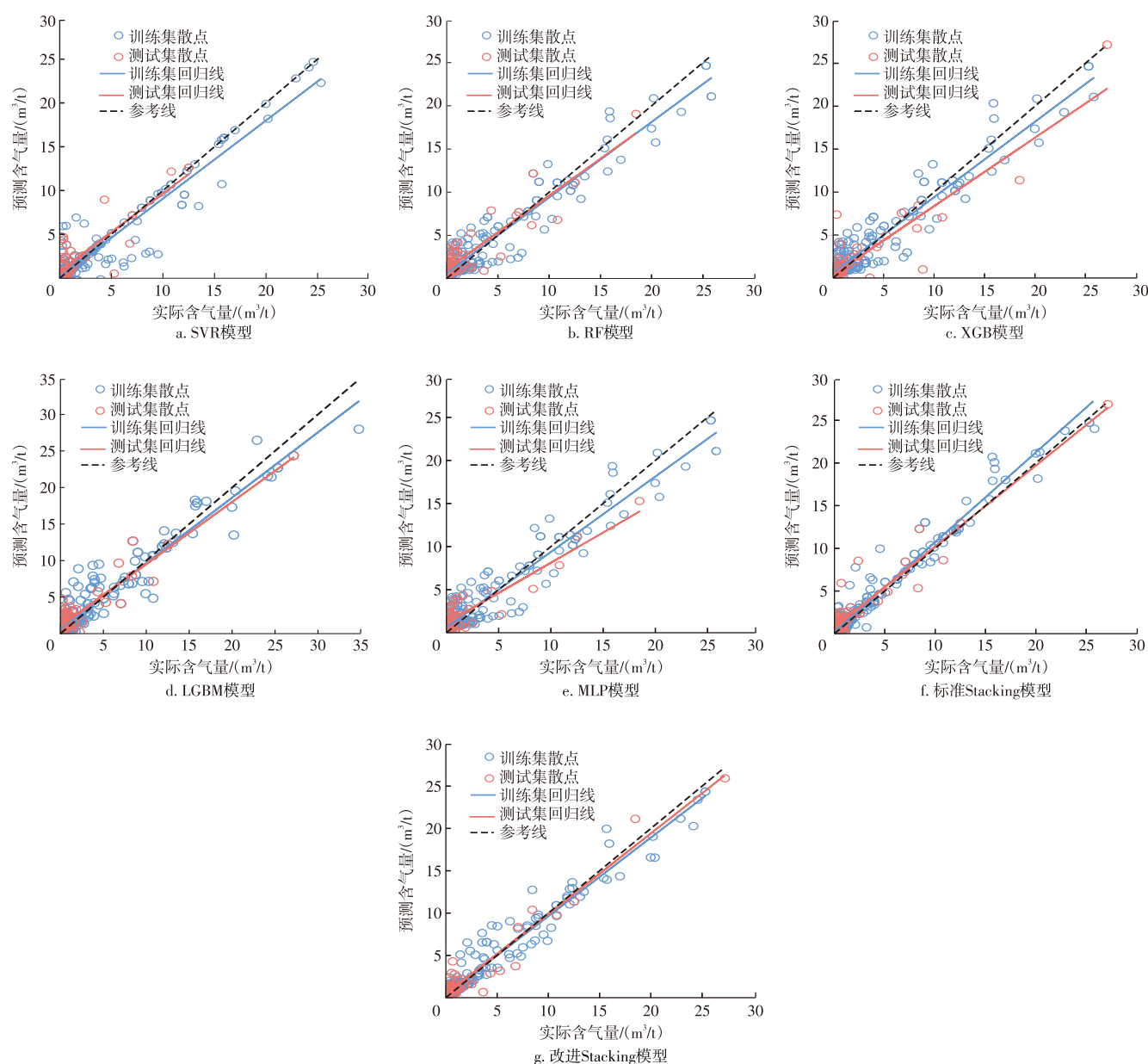


图5 不同煤层气含气量预测模型的预测值与真实值的关系

Fig. 5 Relationship between predicted values and actual values of different coalbed methane gas content prediction models

2.5 可解释分析

为探究输入特征参数对含气量的影响,采用SHAP可解释性分析框架^[23]对预测模型进行特征重要性评估。图6a展示了基于SHAP的基模型融合分析结果,表明煤

层斜深、自然电位和自然伽马对含气量的影响最为显著。图6b中每个样本点对应的SHAP值分布以颜色表示特征数值大小(蓝至红对应低到高),横坐标SHAP值的正负反映特征对预测结果的正负贡献。分析显示:自然伽马值与含气量呈负相关,其值减小会导致SHAP值增大,进

而提升预测含气量。自然伽马通常与灰分含量正相关,而高灰分煤层吸附能力一般较低,从而导致含气量减少;声波时差则与含气量呈正相关,其值增大会带动SHAP值上升,推高预测含气量;煤层斜深、自然电位和密度参数的SHAP分布呈现非线性特征,反映其与含气量关系的复杂性。整体趋势表明:自然电位负异常幅度增大(测井值减小,对应图中蓝色点)倾向于产生正SHAP值,从而提高含气量预测,这与前文论述一致,即自然电位负异常指示岩性纯度提高,间接反映更高的含气潜力。模型成功捕捉了自然电位与含气量之间间接且非单一的关系。

煤层斜深则直接关联储层温压条件,通过影响吸附-解吸平衡调控含气量。SHAP分析揭示各参数通过不同物理机制共同作用于含气量,其分布规律与地质解释机理一致,验证了模型预测的科学性。图7展示了SHAP对元模型的解释结果。在元模型中,各基模型预测值的加权平均对最终预测贡献最大;而在引入原始训练集后,煤层斜深、密度、自然电位、自然伽马和声波时差特征参数亦对元模型预测产生一定影响。结果表明,改进的Stacking融合模型能够有效捕捉原始训练集与新生成训练集之间的潜在关联,从而提升预测精度。

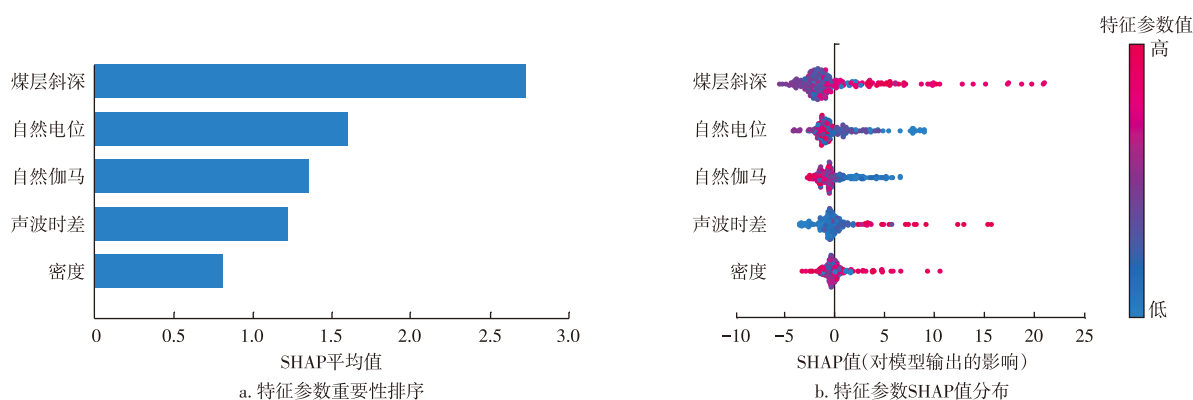


图6 改进Stacking模型的基模型特征参数重要性与SHAP值分布

Fig. 6 Feature parameter importance and SHAP value distribution of base models in improved Stacking model

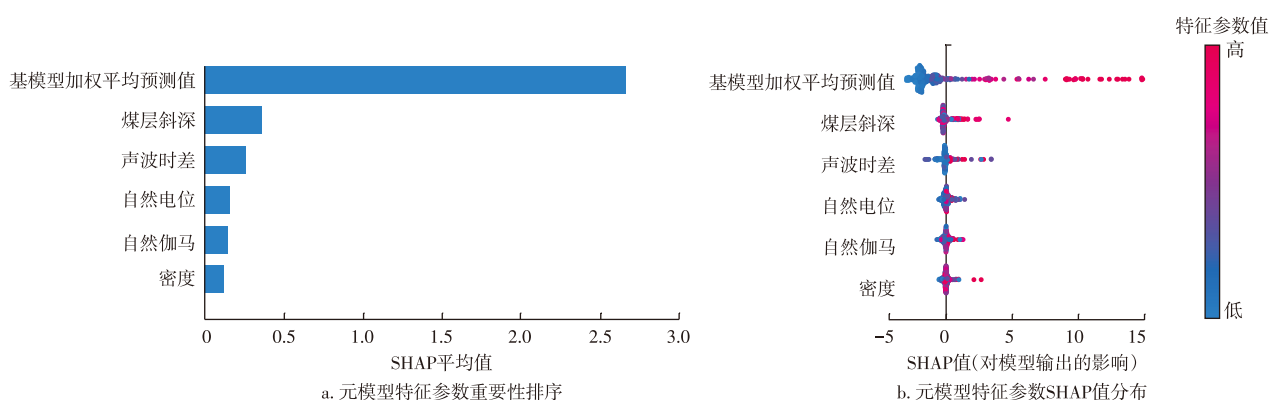


图7 改进Stacking模型的元模型特征参数重要性与SHAP值分布

Fig. 7 Feature parameter importance and SHAP value distribution of meta models in improved Stacking model

3 现场实例应用

利用训练好的改进Stacking预测模型,对沁水盆地煤层气已有测井数据的单井进行煤层气含气量预测。选取区块内含气量较高的4口井以及含气量较小的4口井进行预测。与前文训练模型一致以煤层斜深、自然伽马、自然电位、声波时差和密度作为输入。预测结果如表4所示,改进Stacking预测模型对沁水盆地的煤层气含气量预测值与钻压取心测试结果进行比较,高含气量井平

均绝对误差为 $1.56 \text{ m}^3/\text{t}$;低含气量井平均绝对误差为 $0.11 \text{ m}^3/\text{t}$,因低含气量数据稳定性以及训练样本数据更多则预测精度更高。改进模型通过加权融合基模型优势,在该区块场景中保持平均绝对误差值在 $0.84 \text{ m}^3/\text{t}$ 以内,平均相对误差在6%以内,表明改进Stacking预测模型具有较好的外推性,可用于区块现场实际应用。通过与前人构建的煤层气含气量最优预测模型在实际应用中的表现进行对比分析(表5),结果表明,改进后的Stacking融合模型具有相对一致的预测效果,进一步验证了其良好的工程适用性。

表4 现场8口井煤层气含气量预测结果对比

Table 4 Comparison of coalbed methane gas content prediction results for eight wells at site

井号	煤层斜深/m	自然伽马/API	自然电位/mV	声波时差/(μ s/m)	密度/(g/cm^3)	含气量测量值/(m^3/t)	含气量预测值/(m^3/t)	绝对误差/(m^3/t)	相对误差/%
1	3 008.00	37.88	88.57	80.62	2.32	1.94	2.05	0.11	5.93
2	3 357.50	74.17	81.27	398.39	1.82	2.06	1.94	0.12	6.24
3	3 967.00	71.57	33.87	97.87	1.70	2.53	2.45	0.08	3.28
4	3 837.50	73.17	61.57	387.39	1.85	3.06	2.93	0.13	6.24
5	4 632.25	43.38	77.46	315.10	2.15	28.06	29.85	1.79	4.71
6	1 713.87	47.24	73.29	82.08	2.41	15.41	15.09	0.32	2.04
7	4 470.00	48.03	57.09	295.48	2.20	24.58	21.05	3.53	14.35
8	3 218.50	80.45	136.07	131.96	1.64	16.99	17.59	0.60	3.51

表5 不同预测模型在煤层气含气量预测中的性能对比

Table 5 Performance comparison of different prediction models in coalbed methane gas content prediction

模型	含气量测量均值/(m^3/t)	含气量预测均值/(m^3/t)	平均绝对误差/(m^3/t)	平均相对误差/%
改进Stacking模型	11.703	11.495	0.835	5.875
细菌觅食优化广义回归神经网络 ^[24]	11.938	11.994	0.566	5.791
灰色系统模型 ^[20]	15.274	15.410	0.831	4.933
高斯过程回归模型 ^[25]	9.621	9.623	—	4.533

4 结论

1) 单一的煤层气含气量预测模型中,LGBM 预测效果最好,验证集上的 R^2 为0.84,更加适用于该数据集场景。标准的Stacking融合模型能够有效地集合各基模型优点,相比于SVM、RF、XGB、LGBM和MLP五种基模型, R^2 均不同程度提升,MAE值、RMSE值、MSE值均不同程度降低,全面提升煤层气含气量预测效果。

2) 提出的一种基于改进Stacking融合模型的煤含气量预测方法,将原始数据集也作为元模型训练的一部分,有利于元学习器能够捕捉到原始训练集与新生成训练集之间的潜在关联,相比于标准Stacking融合模型 R^2 提升了6.38%,MAE值、RMSE值、MSE值分别降低了14.3%、25.2%、43.7%,验证了该模型应用于煤层气含气量预测的可行性和有效性,可为此类问题的解决提供方法参考。

3) 基于改进Stacking融合模型,从全局和元模型两个层面进行了可解释性分析。结果表明,煤层斜深、自然电位、自然伽马值输入对模型预测含气量影响较大。各基模型预测值的加权平均值对元模型最终预测值影响最大,但是原始训练集的特征参数对元模型预测煤层含气量依旧有一定的影响,验证了模型内在的预测逻辑与工程实践一致性,模型具有一定的合理性和可靠性。

参考文献

[1] 邵林海,刘池阳,丁清香,等.韩城煤层气三维地震勘探区11#煤层含气量预测[J].地质科技情报,2016,35(1):147-151.

Shao Linhai, Liu Chiyang, Ding Qingxiang, et al. Prediction of 11#

coalbed gas content in Hancheng coalbed methane 3D seismic survey [J]. Geological Science and Technology Information, 2016, 35(1): 147-151.

[2] 李金龙.地质信息技术在云南煤层气领域的应用情况评析[J].中国煤层气,2013,10(3):33-36.

Li Jinlong. Application analysis of geological information technology in the coalbed methane sector of Yunnan Province[J]. China Coalbed Methane, 2013, 10(3): 33-36.

[3] 孙四清,杨帆,郑玉岐,等.煤矿区煤层气开发技术应用现状及展望[J].油气藏评价与开发,2025,15(6):972-982.

Sun Siqing, Yang Fan, Zheng Yuqi, et al. Current applications and prospects of coalbed methane development technologies in coal mining areas[J]. Petroleum Reservoir Evaluation and Development, 2025, 15(6): 972-982.

[4] 陈峥嵘,时贤,刘伟,等.考虑支撑剂嵌入的煤层气井压裂优化方法[J].断块油气田,2025,32(1):159-164.

Chen Zhengrong, Shi Xian, Liu Wei, et al. Optimization method of coalbed methane well fracturing considering proppant embedding[J]. Fault-Block Oil & Gas Field, 2025, 32(1): 159-164.

[5] 傅雪海,张小东,韦重韬.煤层含气量的测试、模拟与预测研究进展[J].中国矿业大学学报,2021,50(1):13-31.

Fu Xuehai, Zhang Xiaodong, Wei Chongtao. Review of research on testing, simulation and prediction of coalbed methane content[J]. Journal of China University of Mining & Technology, 2021, 50(1): 13-31.

[6] 唐颖,张金川,刘珠江,等.解吸法测量页岩含气量及其方法的改进[J].天然气工业,2011,31(10):108-112.

Tang Ying, Zhang Jinchuan, Liu Zhujiang, et al. Use and improvement of the desorption method in shale gas content tests[J]. Natural Gas Industry, 2011, 31(10): 108-112.

[7] 周宝艳,傅雪海,周荣福.沁水煤田煤层含气量校正系数研究[J].中国煤层气,2010,7(2):24-27.

Zhou Baoyan, Fu Xuehai, Zhou Rongfu. Research on correction

- coefficient of gas content in coal seams of Qinshui Coalfield[J]. China Coalbed Methane, 2010, 7(2): 24–27.
- [8] 张景越, 肖小玲, 王鹏飞, 等. 基于多信息融合的层次聚类测井曲线自动分层方法[J]. 断块油气田, 2024, 31(1): 42–49.
- Zhang Jingyue, Xiao Xiaoling, Wang Pengfei, et al. Automatic stratification method of hierarchical clustering logging curve based on multi-information fusion[J]. Fault-Block Oil & Gas Field, 2024, 31(1): 42–49.
- [9] 李丹丹, 王振国, 降文萍, 等. 基于测井参数的煤层含气量定量评价方法研究: 以寿阳区块15煤为例[J]. 中国煤炭地质, 2022, 34(8): 34–40.
- Li Dandan, Wang Zhenguo, Jiang Wenping, et al. Approach study on coal gas content quantitative assessment based on well logging parameters: a case study of coal No. 15 in Shouyang Block[J]. Coal Geology of China, 2022, 34(8): 34–40.
- [10] 连承波, 赵永军, 李汉林, 等. 煤层含气量的主控因素及定量预测[J]. 煤炭学报, 2005, 30(6): 726–729.
- Lian Chengbo, Zhao Yongjun, Li Hanlin, et al. Main controlling factors analysis and prediction of coalbed gas content[J]. Journal of China Coal Society, 2005, 30(6): 726–729.
- [11] 李伟超, 宋大猛, 陈斌. 基于遗传算法的人工神经网络[J]. 计算机工程与设计, 2006, 27(2): 316–318.
- Li Weichao, Song Dameng, Chen Bin. Artificial neural network based on genetic algorithm[J]. Computer Engineering and Design, 2006, 27(2): 316–318.
- [12] 胡锦宏. 基于遗传神经网络的煤层气含量预测[J]. 绿色科技, 2015, 17(2): 260–263.
- Hu Jinhong. Prediction of coalbed gas content based on genetic neural network[J]. Journal of Green Science and Technology, 2015, 17(2): 260–263.
- [13] 向旻, 齐兴华, 张峰玮, 等. 深度学习在煤层气测井解释中的应用研究[J]. 地质与勘探, 2020, 56(6): 1305–1312.
- Xiang Min, Qi Xinghua, Zhang Fengwei, et al. Application of deep learning to coalbed methane logging interpretation[J]. Geology and Exploration, 2020, 56(6): 1305–1312.
- [14] 梁海波, 马睿. 基于Stacking集成学习的孔隙度预测方法[J]. 电子测量与仪器学报, 2024, 38(12): 202–210.
- Liang Haibo, Ma Rui. Porosity prediction method based on Stacking ensemble learning[J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(12): 202–210.
- [15] 刘磊, 李伟, 杜玉山, 等. 基于Stacking集成学习的分频地震属性融合储层预测方法[J]. 石油地球物理勘探, 2024, 59(1): 12–22.
- Liu Lei, Li Wei, Du Yushan, et al. Reservoir prediction method of fusing frequency-decomposed seismic attributes using Stacking ensemble learning[J]. Oil Geophysical Prospecting, 2024, 59(1): 12–22.
- [16] 孔祥伟, 谢昕, 王存武, 等. 基于灰色关联方法的深层煤层气井压后产能影响地质工程因素评价[J]. 油气藏评价与开发, 2023, 13(4): 433–440.
- Kong Xiangwei, Xie Xin, Wang Cunwu, et al. Evaluation of geological engineering factors for productivity of deep CBM well after fracturing based on grey correlation method[J]. Petroleum Reservoir Evaluation and Development, 2023, 13(4): 433–440.
- [17] 蒋新立, 李明忠, 宋文宇, 等. 鄂北大牛地气田深层煤层气钻完井关键技术及应用[J]. 断块油气田, 2025, 32(5): 863–868.
- Jiang Xinli, Li Mingzhong, Song Wenyu, et al. Key drilling and completion techniques and applications in deep coalbed methane of Daniudi gas field in northern Ordos Basin[J]. Fault-Block Oil & Gas Field, 2025, 32(5): 863–868.
- [18] 张宝东, 杜支文, 闫昭, 等. 基于集成学习算法的尾气处理装置SO₂排放预测模型[J]. 石油与天然气化工, 2025, 54(1): 9–17.
- Zhang Baodong, Du Zhiwen, Yan Zhao, et al. Sulfur dioxide emissions predictive model of tail gas treatment unit based on ensemble learning algorithm[J]. Chemical Engineering of Oil & Gas, 2025, 54(1): 9–17.
- [19] Wolpert D H. Stacked generalization[J]. Neural Networks, 1992, 5(2): 241–259.
- [20] 郭建宏, 张占松, 张超谟, 等. 基于灰色系统与测井方法的煤层气含量预测及应用[J]. 物探与化探, 2020, 44(5): 1190–1200.
- Guo Jianhong, Zhang Zhansong, Zhang Chaomo, et al. Prediction and application of coalbed methane content based on gray system and logging method[J]. Geophysical and Geochemical Exploration, 2020, 44(5): 1190–1200.
- [21] 陈涛, 丁海琨, 张占松, 等. 基于测井资料与优化通用向量机的煤层气含量预测模型[J]. 煤矿安全, 2022, 53(11): 157–166.
- Chen Tao, Ding Haikun, Zhang Zhansong, et al. Prediction model of CBM content based on logging data and optimized general vector machine[J]. Safety in Coal Mines, 2022, 53(11): 157–166.
- [22] Lerman P M. Fitting segmented regression models by grid search[J]. Applied Statistics, 1980, 29(1): 77.
- [23] Kannangara K K P M, Zhou W, Ding Z, et al. Investigation of feature contribution to shield tunneling-induced settlement using Shapley additive explanations method[J]. Journal of Rock Mechanics and Geotechnical Engineering, 2022, 14(4): 1052–1063.
- [24] 张瑞, 陈刚, 潘保芝, 等. 基于细菌觅食优化广义回归神经网络的煤层气含量预测[J]. 物探与化探, 2016, 40(2): 327–332.
- Zhang Rui, Chen Gang, Pan Baozhi, et al. The prediction of the coalbed methane content based on bacteria foraging optimizing generalized regression neural network[J]. Geophysical and Geochemical Exploration, 2016, 40(2): 327–332.
- [25] 马晓荷, 谭成仟, 郭泽坤. 博文盆地M煤层气田含气量主控因素分析及预测[J]. 科学技术与工程, 2023, 23(11): 4586–4595.
- Ma Xiaohu, Tan Chengqian, Guo Zekun. Analysis and prediction of main controlling factors of gas content in coalbed of M coalfield in Bowen basin[J]. Science Technology and Engineering, 2023, 23(11): 4586–4595.

(编辑 李青)