

DOI: 10.13382/j.jemi.B2407612

基于 Stacking 集成学习的孔隙度预测方法*

梁海波 马 睿

(西南石油大学机电工程学院 成都 610500)

摘要: 储层孔隙度的预测准确性决定了评估地下储层的储集空间和储层质量的可靠性。然而,现有孔隙度预测的方法存在模型算法单一、精度不高和泛化性差等问题。为了提高孔隙度预测的精度,提出了一种基于 Optuna 优化的 Stacking 集成学习方法。首先,采用灰色关联度选取声波时差、井深、岩石密度、井斜角和光电截面吸收指数作为输入参数。然后,对输入数据进行归一化处理,并通过 Optuna 优化模型参数。根据均方根误差、平均绝对误差和定系数选取随机森林(RF)、支持向量回归(SVM)和k-近邻算法(KNN)作为 Stacking 的基学习器,以及弹性网络回归(ENet)作为 Stacking 元学习器。各主流模型预测结果与 Stacking 模型比较发现:RF 在处理非线性数据时表现优异,但预测结果不稳定,Stacking 模型相较 RF 降低了约 10% 的均方根误差。SVM 具备较强的泛化能力,但参数调优复杂,Stacking 模型相较 SVM 降低了约 39% 的均方根误差。KNN 对异常值不敏感,但对高维数据效果较差,Stacking 模型相比 KNN 降低了约 21% 的误差。Xgboost 能够较好地避免过拟合,但对异常值敏感而且参数调优复杂,Stacking 模型相比 Xgboost 降低了约 30% 的误差。最终结果表明,基于 Optuna 优化的 Stacking 模型显著提高了孔隙度预测的准确性,为反应储层油气储存能力提供重要参考。

关键词: 孔隙度;预测;Optuna 优化;Stacking;集成学习;测井数据

中图分类号: TE1; TN98

文献标识码: A

国家标准学科分类代码: 440.5

Porosity prediction method based on stacking ensemble learning

Liang Haibo Ma Rui

(School of Mechanical and Electrical Engineering, Southwest Petroleum University, Chengdu 610500, China)

Abstract: The accuracy of reservoir porosity prediction is crucial for assessing the storage capacity and quality of underground reservoirs. However, existing methods for porosity prediction face challenges such as limited model algorithms, low accuracy, and poor generalization. To enhance the precision of porosity prediction, this study proposes a Stacking ensemble learning method optimized by Optuna. First, gray relational analysis is used to select input parameters, including acoustic time difference, well depth, rock density, dip angle, and photoelectric absorption index. The input data is then normalized, and Optuna is employed to optimize the model parameters. Based on metrics like root mean square error, mean absolute error, and determination coefficient, random forest, support vector regression, and K-nearest neighbors are chosen as base learners for the Stacking model, with elastic net regression serving as the meta-learner. Comparative results reveal that while RF excels in handling nonlinear data, it shows instability in predictions; the Stacking model reduces RMSE by approximately 10% compared to RF. SVM demonstrates strong generalization ability but requires complex parameter tuning, with the Stacking model achieving about a 39% reduction in RMSE compared to SVM. KNN is insensitive to outliers but performs poorly on high-dimensional data, with the Stacking model lowering the error by about 21% compared to KNN. Additionally, XGBoost effectively avoids overfitting but is sensitive to outliers and requires complex tuning, with the Stacking model reducing the error by approximately 30% compared to XGBoost. Overall, the results indicate that the Optuna-optimized Stacking model significantly improves the accuracy of porosity prediction, providing a valuable reference for evaluating the oil and gas storage capacity of reservoirs.

Keywords: porosity; prediction; Optuna optimization; Stacking; ensemble learning; logging data

收稿日期: 2024-06-21 Received Date: 2024-06-21

* 基金项目: 四川省自然科学基金创新研究群体项目(2023NSFSC1981)资助

0 引言

储层孔隙度是用于评估储层的储集能力和流体运移性,地质解释和油气储存能力的重要物性参数,指导资源开发和管理的决策。储层孔隙度的准确预测对地质解释和油气勘探开发至关重要^[1]。

在油田的勘探与开发过程中,通常采用直接测定和间接测定两种方式来确定储层参数^[2-3]。直接测定方法是在钻探过程中钻井取心并进行化验分析,从而获得各种储层参数的实际数值。虽然这种方法可以提供与真实值非常接近的储层参数,但是其成本高昂、时效性差且应用范围有限,具有一定局限性。间接测定方法是一种便捷的方法,其主要依据测井曲线,通过经验公式或者构造一个简单的地质构造模型来推断未知地层的储层参数^[4-5]。但是,这种方法存在较强的主观性强和泛化性差的问题。

储层孔隙度的预测是受到沉积环境、成岩作用和地质构造等多种因素共同作用的复杂、非线性问题。因此,测井数据与孔隙度之间的关系难以用简单的经验公式或物理模型来描述^[6-7]。监督学习^[8]和混合稀疏模型^[9]在数据分类、重建和去噪中的成功应用,表明机器学习能够有效处理复杂的储层物性参数预测问题^[10]。随着机器学习在储层孔隙度预测中的潜力逐渐显现,石油工程测井领域对提高预测精度和效率的需求也日益凸显^[11-13]。

Anietie 等^[14]采用伽马射线、电阻率、密度和深度段测井作为输入,孔隙度、渗透率和含水饱和度作为输出,建立了一个隐含层有 12 个神经元的前馈反向传播多输入多输出网络,其性能优于现有的其他模型。Wang 等^[15]通过测井资料构建了基于一维卷积神经网络和双向门控递归单元神经网络的孔隙度预测方法,相较于双向门控循环单元神经网络有显著提升。崔学慧等^[16]采用声波时差、自然电位、自然伽马、以及地层真电阻率为指标,采用交叉验证建立非线性模型,提出随机森林回归算法预测方法,发现相较传统线性回归算法随机森林回归算法精度较高。潘少伟等^[17]提出基于网格搜索法和粒子群算法寻优的 XGBoost 算法的孔隙度预测模型,相较于其他算法泛化性强。侯贤沐等^[18]采用 4 种机器学习算法(随机森林、K-近邻、支持向量机和长短期记忆网络)进行参数寻优,以实现孔隙度的预测,得到 4 种机器学习方法在储层孔隙度预测结果差异不大,需要通过调整输入参数种类,进一步寻找到最优算法。汪敏等^[19]基于少量测井和岩芯数据,定量分析和目标井的相似性,构建迁移深度神经网络实现对孔隙度进行预测,其误差小并减少对测井和岩芯数据的依赖。

上述研究表明,目前孔隙度预测大多采用单一的机

器预测模型或简单模型的优化算法,整体预测精度尚待提高。在实际应用中,每种机器学习算法都具有其独特的优缺点。因此,采用单一算法预测的孔隙度精度不高。旨在探讨 Staking 方法在储层孔隙度预测中的应用和优势,以提供一种新的技术手段用于油气勘探开发。为此,提出基于 Optuna 优化的 Stacking 集成学习的模型,通过结合不同类型的机器学习模型来提升对孔隙度预测的效果。

首先,介绍了 Stacking 集成学习算法。其次,采用灰色关联方法来构建数据集,确定模型输入参数,并利用 Optuna 方法确定各模型超参数。随后,根据评价指标来确定最佳基学习器和元学习器。最后,基于优选的基学习器和元学习器构建了 Stacking 集成学习模型,并利用某井的测井数据对模型的准确性进行了测试。结果发现,Stacking 的预测误差均小于各个单一模型,表明该算法可以有效提高孔隙度预测精度。

1 算法原理

1.1 Stacking 算法

Stacking 是一种集成学习方法,包括两层框架:第 1 层由多个基学习器组成,这些基学习器分别独立训练,以获得不同的预测值。然后,第 1 层的输出结果作为第 2 层的输入,由元模型输出最终的预测结果。Stacking 集成学习通过对多个模型输出结果的泛化,显著提升了整体评估效果^[20]。算法流程如图 1 所示。

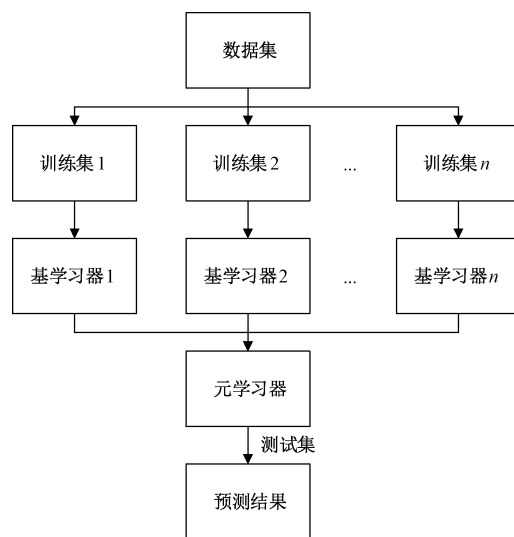


图 1 Stacking 算法流程图

Fig. 1 Algorithm flowchart

与传统的集成方法(如 Bagging 和 Boosting)相比,Stacking 可以结合多种不同类型的学习器,充分利用各自

的特点和优势,来提高模型的表达能力和泛化能力,避免过拟合或欠拟合的问题,达到提高模型的稳定性和可靠性。Stacking 还能通过调整基学习器和元学习器中参数的值来优化模型性能,增加模型的灵活性和可调节性。

1.2 基学习器

对于基学习器的选择,一般会优先考虑强模型且差异性大的模型,以实现多种模型的优势互补。选择的基学习器有随机森林 (random forest, RF)、支持向量机 (support vector machines, SVM)、k-近邻算法 (k-nearest neighbor, KNN)、XGBOOST、分类回归树算法 (classification and cegression tree, CART)。

RF 是一种集成学习方法,通过构建多棵决策树并将它们的预测结果进行平均,来提高模型的准确性和鲁棒性。适用于处理高维度和多类型的数据,能够有效降低过拟合的风险,但其训练和预测速度较慢,且对噪声数据敏感。

SVM 通过找到一个最大化边界内包含尽可能多的数据点的线性函数进行预测。能够处理线性不可分的数

据,避免维数灾难,但对参数选择敏感,对缺失数据和噪声数据的鲁棒性较差,并且训练时间较长。

KNN 算法是一种基于实例的学习方法,通过计算新样本与训练集中每个样本的距离,选择距离最近的 k 个邻居,并通过这些邻居的平均值进行预测。其简单易实现,无需训练模型,可以动态更新数据,但是计算量大,对异常值和特征尺度敏感度较高。

XGBOOST 是一种基于梯度提升的树模型。通过逐步构建树模型,每一步都在前一步的基础上进行优化,最小化损失函数。其训练速度快,支持并行计算和正则化,能够处理稀疏数据和缺失值,但其参数较多,调节困难,且容易产生过拟合。

CART 通过递归地分割数据集,使每个分割点的子集尽可能减少误差,构建二叉树。每个节点根据某个特征的值进行分割,直到满足停止条件。其模型直观易解释,适合用于处理混合类型的数据,可以进行特征选择和降维,但是容易产生过拟合或欠拟合,且对噪声和异常值敏感。各算法优缺点如表 1 所示。

表 1 各算法优缺点
Table 1 Advantages and disadvantages of each algorithm

算法	优点	缺点
RF	高鲁棒性和泛化能力,不需要特征选择,训练速度快	模型的解释性较差,存在相似决策树
SVM	适合处理高维度和非线性数据,具有较强的泛化能力	对于噪声敏感,核函数和参数选择难以确定
KNN	适合动态地更新训练数据,对异常值不敏感,鲁棒性好	预测速度较慢、参数调优较复杂
XGBOOST	具有较强的泛化能力,防止过拟合具有较强的鲁棒性	训练速度较慢,难以并行训练数据
CART	具有较强的泛化能力,处理多维度和非线性数据	模型的稳定性较差,异常值过多容易造成过拟合

1.3 Optuna-Stacking 算法

Optuna 使用了 TPE (tree-structured parzen estimator) 算法,综合了网格搜索、随机搜索和贝叶斯优化等多种搜索算法,以提高模型优化效率^[21]。Optuna 能够剪枝效果较差的试验,从而减少计算资源的浪费,在有限的迭代次数内找到最优的超参数组合。Optuna 优化参数步骤如下:

首先,确定模型的可调节参数、最大迭代次数和模型的性能评估指标;然后,循环迭代选择下一个超参数组合进行评估,并剪枝效果差的试验,直到达到预定的迭代次数或满足停止条件;最后,获取最佳超参数组合。

2 数据处理与评价指标

2.1 数据集构建

研究数据来自于我国南海某 X 油田的综合测井数据。根据测井结果,选取了 8 个与孔隙度具有较强相关性的测井参数:井深 (DEPTH)、自然伽马 (GR)、声波时差 (AC)、岩石密度 (DEN)、深电阻率 (RD)、浅电阻率

(RS)、井斜角 (DEV)、光电吸收截面指数 (PE)。以下是各参数在孔隙度预测中的重要性:

井深:描述地层深度的重要指标,能够反映地下储层的埋藏特征和层位信息,对理解地层的分布及孔隙度变化具有指导意义。

自然伽马:反映了地层中放射性矿物的含量,通常与孔隙度和水饱和度相关。通过分析自然伽马值,可以间接推测地层的孔隙特性。

声波时差:用来测量声波在地层中传播的时间,通常与孔隙度成反比。声波时差较大可能表明孔隙度较高,因此这一参数在孔隙度预测中起着重要作用。

岩石密度:提供了地层的物理特性信息,通常高孔隙度的岩石密度较低。通过测量岩石密度,可以有效评估孔隙度的变化。

深电阻率和浅电阻率:这两个电阻率参数能够反映地层的含水性 and 岩石类型,通常低电阻率值与高孔隙度及饱和水分相关。电阻率的测量对区分储层性质尤为重要。

井斜角:反映了井的倾斜程度,对孔隙度的测量和分析有一定影响,尤其是在斜井和水平井的情况下。

光电吸收截面指数:能够提供地层中流体类型和饱和度的信息,通常与孔隙度密切相关,是评估储层质量的重要指标。数据集的格式如表 2 所示。

表 2 数据集

Table 2 Data set

DEPTH/m	2 080	2 081	2 082
GR/API	77. 97	85. 72	75. 75
AC/($\mu\text{s}\cdot\text{ft}^{-1}$)	59. 85	61. 46	70. 14
DEN/($\text{g}\cdot(\text{cm}^3)^{-1}$)	2. 38	2. 29	2. 18
RD/($\Omega\cdot\text{m}$)	0. 091	0. 093	0. 094
RS/($\Omega\cdot\text{m}$)	0. 11	0. 11	0. 1
PE/($\text{b}\cdot\text{e}^{-1}$)	0. 24	0. 28	0. 34
DEV/($^{\circ}$)	11. 08	11. 54	11. 61
孔隙度/%	13. 69	12. 86	3. 45

将多种种类的测井数据作为堆叠模型的输入参数,低相关性的参数会对模型预测的准确性产生负面影响。此外,增加多参数输入会导致算法的计算耗时增加。

为择优模型输入参数,采用灰色关联度分析,其在解决多目标优化响应问题方面具有显著优势,适用于分析多因素和多变量之间的复杂关系。通过选取与孔隙度相关度高的参数,可以提高模型训练效果。因此,引入灰色关联分析法来衡量不同基学习器对目标模型的影响程度,其步骤如下:

步骤 1) 确定分析序列,包括参考序列(母序列)和比较序列(子序列);

步骤 2) 数据预处理,对原始数据采用式(1)进行无量纲化处理,采用均值化;

$$x_i(k) = \frac{x_i^*(k)}{\frac{1}{n} \sum_{k=1}^n x_i^*(k)} \quad (1)$$

其中, $x_i(k)$ 为均值化后的数据, $x_i^*(k)$ 为原始数据在第 k 个维度上的值, i 代表变量指标个数, n 代表维度数,其中 $k = 1, 2, \dots, n$, $i = 0, 1, \dots, 4$ 。

步骤 3) 计算灰色关联系数,灰色关联系数采用(2)计算,计算公式如式(2)所示。

$$\zeta_i(k) = \frac{\min_{i,k} |x_0(k) - x_i(k)|}{|x_0(k) - x_i(k)| + \rho \max_{i,k} |x_0(k) - x_i(k)|} + \frac{\rho \max_{i,k} |x_0(k) - x_i(k)|}{|x_0(k) - x_i(k)| + \rho \max_{i,k} |x_0(k) - x_i(k)|} \quad (2)$$

其中, $x_0(k)$ 是参考序列在第 k 个维度上的值, $x_i(k)$ 是第 i 个比较序列在第 k 个维度上的值, ρ 是分辨系数,一般取 0.5。

步骤 4) 计算关联度,公式如式(3)所示。

$$y(x_0, x_1) = \frac{1}{n} \sum_{k=1}^n \zeta_i(k) \quad (3)$$

其中, n 是维度数, $\zeta_i(k)$ 是第 i 个比较序列在第 k 个维度上的关联系数。

步骤 5) 关联度排序,得出与参考序列关联程度由高到低的顺序,选取关联度高的子序列作为重要因素。

根据关联度的大小排序绘制各因素对孔隙压力的关联度示意图,选取的 8 个测井参数的灰色关联度如图 2 所示。

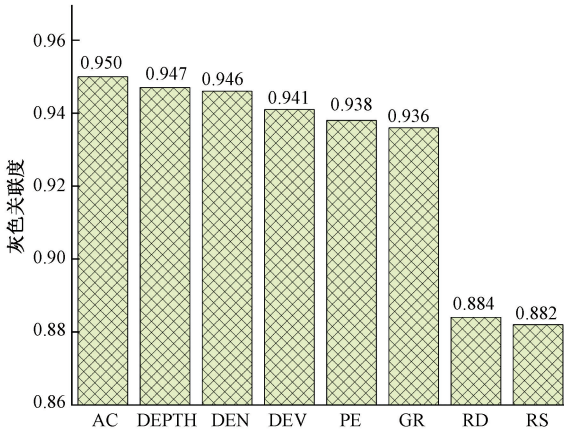


图 2 各测井参数的灰色关联度

Fig. 2 Grey correlation of each logging parameter

选取关联度前 5 的参数作为模型输入参数,依次选取声波时差、井深、岩石密度、井斜角、光电截面吸收指数。

2.2 数据预处理

为了提高孔隙度预测模型的准确率和训练效率,需要对不同量纲和数值范围的多个测井参数进行预处理。采用最大最小归一化公式对训练数据集和测试数据集进行预处理,公式如式(4)所示。

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (4)$$

其中, x' 为归一化后的值, x 为归一化前的值。经过最大最小归一化的数据将缩放值 0~1 的范围。

2.3 评价指标

为了评价堆叠模型对孔隙度的预测效果,采用均方根误差(root mean square error, RMSE)、平均绝对误差(mean absolute error, MAE)、决定系数(R^2)作为评价指标, P 作为 Stacking 模型对比其他模型均方根误差下降百分比。

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (5)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (6)$$

$$R^2 = 1 - \frac{SSR}{SST} \quad (7)$$

$$P = \frac{RMSE_o - RMSE_i}{RMSE_o}$$

(8)

其中, y_i 为真实值, $\hat{y}_i$ 为预测值, SSR 为回归平方和, SST 为总离差平方和, $RMSE_o$ 为原始模型的均方根误差, $RMSE_i$ 为改进后模型的均方根误差。

3 基于 Optuna-Stacking 集成学习孔隙度预测模型的建立

提出一种基于 Optuna-Staking 孔隙度预测模型。首先,从多个测井数据中选取了与孔隙度相关度较高的 8 个参数,并对其进行了预处理。然后,采用灰色关联分析法,选取与孔隙度相关程度最高的 5 个测井数据作为输入参数。其次,将数据集按照 7 : 3 的比例划分为训练集和测试集。在训练集上,采用 RF、SVM、KNN、XGBOOST 和 CART 训练,并用 Optuna 进行超参数优化以选取最优的基学习器。最后,将选取的基学习器输入到 Stacking 模型中,并使用 Optuna 优化超参数选取最佳的元学习器,从而得到最终的预测模型。模型预测流程图如图 3 所示。

3.1 基学习器选择

Stacking 集成学习模型要求基学习器具有较大的差异性,以实现互补优化,为了实现强的基学习器。为此,首先通过单独对每个模型进行预测,并采用结合五折交叉验证法的 Optuna 方法来使每个模型的超参数达到最优,以确保其达到最佳状态。各模型的最佳超参数如表 3 所示。基学习器的 MAE 、 $RMSE$ 和 R^2 如图 4 所示。

表 3 基学习器各模型超参数和误差
Table 3 Hyperparameters and errors of each model of the base learner

模型	最佳超参数	MAE	RMSE	R ²
RF	n_estimators = 110; max_depth = 10	0.549	0.891	0.801
SVM	kernel = 'rbf'; C = 10; gamma = 8	0.555	0.857	0.825
KNN	k = 3; weights = uniform	0.518	0.864	0.822
XGBOOST	max_depth = 3; learning_rate = 0.2; n_estimators = 100; colsample_bytree = 0.8; subsample = 0.8	0.579	0.974	0.774
CART	criterion = 'mse'; max_depth = 8	0.648	1.171	0.675

由图 4 可得,根据 Optuna 优化的各模型超参数,综

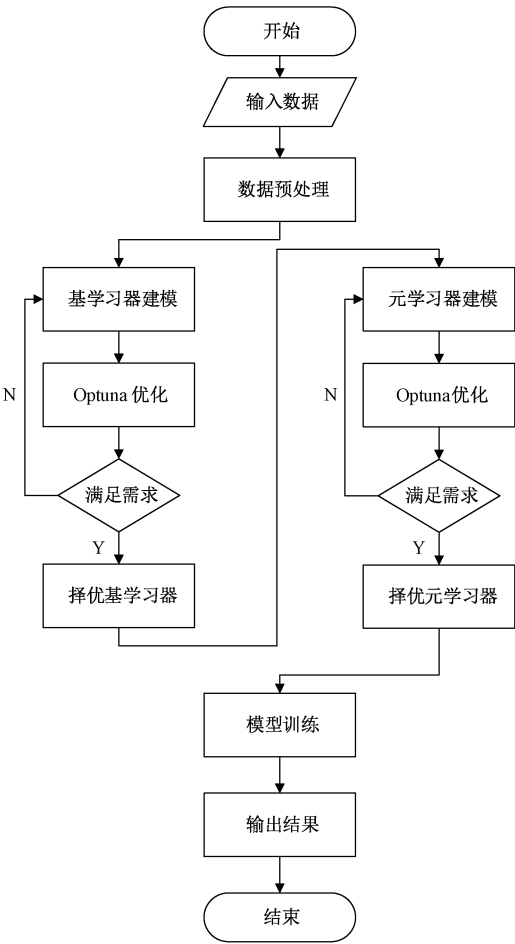


图 3 基于 Optuna-Stacking 模型的孔隙度预测流程图
Fig. 3 Flowchart of porosity prediction based on Optuna-Stakin model

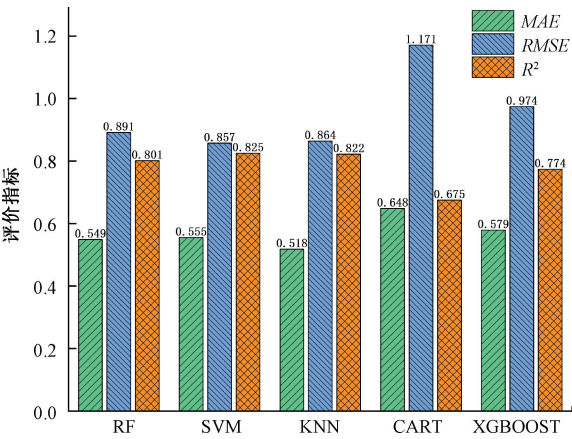


图 4 基学习器的 MAE 、 $RMSE$ 和 R^2
Fig. 4 The mean absolute error, root mean square error, and coefficient of determination of the base learner

合 $RMSE$ 、 MAE 和 R^2 选择 RF、SVM 和 KNN 作为 stacking 的基学习器。

3.2 元学习器的选择

Stacking 能够结合各基模型的优点,而忽略其缺点,但为了避免过拟合的风险,选择合适的元学习器至关重要。通常情况下,元学习器会选择泛化能力强或简单的模型。因此,元学习器选取线性回归 (linear regression, LR) 和弹性网络回归 (elastic network regression, ENet),作为元学习器,因为它们都是相对简单的模型。而梯度提升树 (gradient boosting decision tree, GBDT) 通过正则化来防止过拟合。各模型的最佳超参数如表 4 所示。元学习器的 MAE 、 $RMSE$ 和 R^2 如图 5 所示。

表 4 元学习器各超参数和误差

Table 4 Hyperparameters and errors of each model of the meta learner				
模型	超参数	MAE	$RMSE$	R^2
LR	fit_intercept=False;	0.536	0.826	0.837
	normalize=True			
GBDT	learning_rate=0.01;	0.562	0.838	0.832
	max_depth=3;			
	min_samples_split=2;			
	n_estimators=200			
ENet	alpha=0.1;	0.508	0.784	0.856
	fit_intercept=True;			
	l1_ratio=0.2;			
	normalize=False			

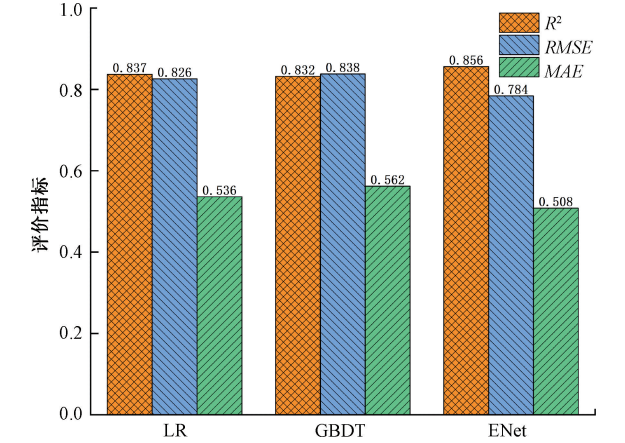


图 5 元学习器的 R^2 、 $RMSE$ 和 MAE

Fig. 5 The coefficient of determination, root mean square error and mean absolute error of the meta learner

根据图 5 可得,综合 MAE 、 $RMSE$ 和 R^2 Enet 作为元学习器相较于其他二者最低,决定系数相较其他二者最高,所以选取 Enet 作为元学习器的预测效果最好。

4 应用效果及分析

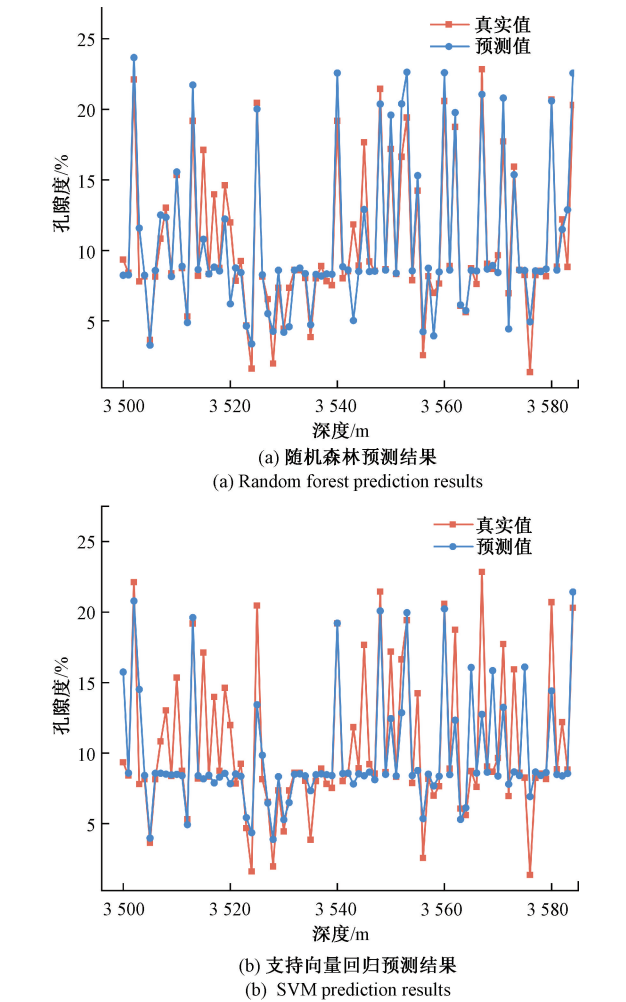
以南海 X 井的测井数据为例,进行应用效果分析。

首先对 X 井的测井数据进行数据集构建,通过灰色关联分析确定模型输入参数,然后将输入的数据进行标准化处理,形成的样本数据按照 7 : 3 的比例随机分为测试集和训练集。将训练集样本输入已经训练好的 stacking 模型进行训练。最后,将训练结果和真实值进行比较,评估现有单一回归预测结果 RF、SVM、KNN、Xgboost 和 Stacking 对应的孔隙度预测误差如表 5 所示,各模型预测结果与真实值对比如图 6 所示。

表 5 各模型预测误差

Table 5 Prediction error of each model		
模型	$RMSE$	R^2
RF	0.646	0.802
SVM	0.954	0.736
KNN	0.739	0.833
Xgboost	0.835	0.794
Stacking	0.583	0.923

从图 6 可以看出 Stacking 模型的孔隙度预测相比于单一的算法更接近真实值,直观上可以看出拟合效果优于其他模型。同样地,可以在图中观察到,SVM 模型



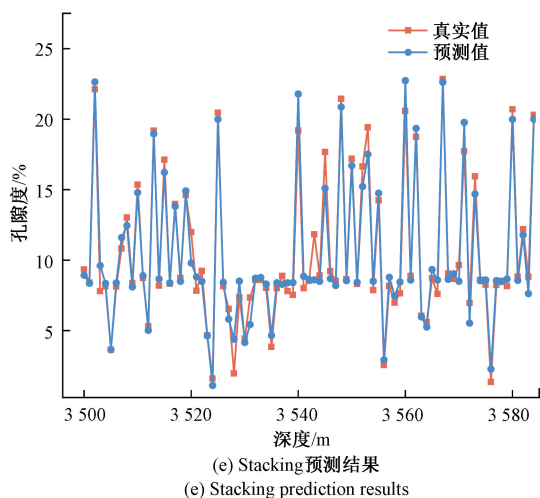
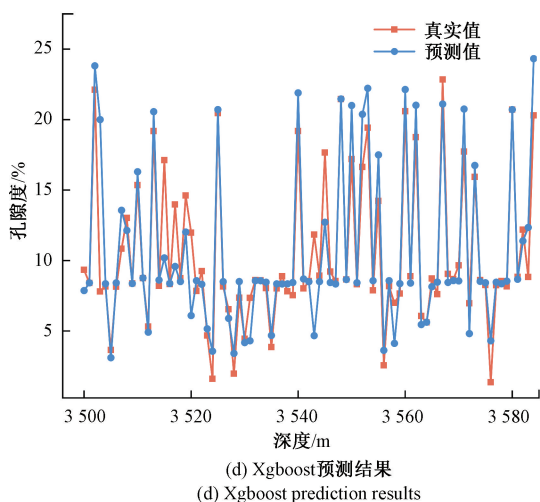
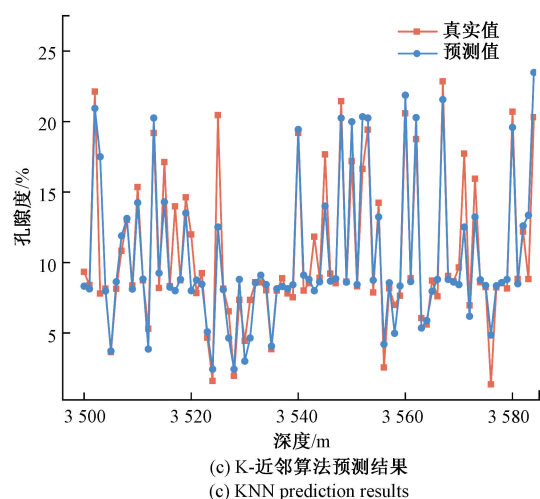


图 6 基于 RF、SVM、KNN、Xgboost 和 Stacking
各模型预测值于真实值对比

Fig. 6 Based on the comparison of the predicted values of the models of RF, SVM, KNN, Xgboost and Stacking compared with the real values

对孔隙度预测变化趋势不太敏感,KNN、RF 模型能较好的预测孔隙度变化趋势,但是容易造成过拟合的风险,Xgboost 对异常值敏感而且参数调优复杂。根据式(8)计算得到基于 Optuna 优化的 Stacking 模型在孔隙度预测相对于单个算法的均方根误差下降百分比,对于 RF、SVM、KNN 和 Xgboost 分别下降了 9.75%、38.89%、21.11% 和 30.17%,验证了该方法具有较高的准确性。与单独的算法相比,Stacking 模型对孔隙度变化规律的预测更为准确。Stacking 模型预测效果明显优于各个单一模型,表明基于 Stacking 模型的孔隙度预测误差低于单一模型,并且模型的鲁棒性更高。

5 结 论

本研究针对测井数据预测孔隙度不精确的问题,提出了一种基于 Optuna 优化的 Stacking 集成学习模型。该模型通过多模型的优势互补,克服单一模型的缺陷,有效地提高了储层孔隙度预测精度。首先,采用灰色关联度分析法,选取声波时差、井深、岩石密度、井斜角和光电截面吸收指数作为与孔隙度相关程度最高的输入参数。然后,对输入数据进行归一化处理,并使用 Optuna 优化模型参数。根据 MAE、RMSE 和 R^2 ,选取 RF、SVM 和 KNN 作为 Stacking 的基学习器,以及 ENet 作为 Stacking 元学习器,构建孔隙度预测模型。最后,对比预测孔隙度值与真实孔隙度值。通过对评价指标定量分析,验证了该方法具有较高的准确性,且在工程运用领域有较好的前景。

本研究存在的不足之处在于尚未在不同地质条件下的应用,限制了其在油气勘探开发提供更加精准和高效的技术支持。未来的研究进一步增强模型对不同地区的泛化能力,使其在多样化的地质环境中具备更强的适应性,为油气勘探提供更加精确且高效的技术支持。

参考文献

- [1] 安鹏,曹丹平,赵宝银,等. 基于 LSTM 循环神经网络的储层物性参数预测方法研究[J]. 地球物理学进展, 2019, 34(5): 1849-1858.
AN P, CAO D P, ZHAO B Y, et al. Reservoir physical parameters prediction based on LSTM recurrent neural network[J]. Progress in Geophysics, 2019, 34(5): 1849-1858.
- [2] WANG J, CAO J X. Deep learning reservoir porosity prediction using integrated neural network[J]. Arabian Journal For Science And Engineering, 2022, 47(9): 11313-11327.

- [3] BENJAMIN S, MATTHIAS G, MANFRED J. Chances and limits of single-station seismic event clustering by unsupervised pattern recognition[J]. *Geophysical Journal International*, 2015, 201(3): 1801-1813.
- [4] CHEN Y K. Automatic microseismic event picking via unsupervised machine learning[J]. *Geophysical Journal International*, 2020, 222(3): 1750-1764.
- [5] XIE J Y, CHEN W, ZHANG D, et al. Application of principal component analysis in weighted stacking of seismic data [J]. *IEEE Geoscience and Remote Sensing Letters*, 2017, 14(8): 1213-1217.
- [6] CHEN W, YANG L, ZHA B, et al. Deep learning reservoir porosity prediction based on multilayer long short-term memory network [J]. *Geophysics*, 2020, 85(4): 213-225.
- [7] 甘利灯, 张昕, 王晓钧, 等. 从勘探领域变化看地震储层预测技术现状和发展趋势[J]. *石油地球物理勘探*, 2018, 53(1): 214-225, 11.
- GAN L D, ZHANG X, WANG X J, et al. From the perspective of changes in the exploration field, the current situation and development trend of seismic reservoir prediction technology [J]. *Oil Geophysical Prospecting*, 2018, 53(1): 214-225, 11.
- [8] REYNEN A, AUDET P. Supervised machine learning on a network scale: Application to seismic event classification and detection [J]. *Geophysical Journal International*, 2017, 210(3): 1394-1409.
- [9] ZHANG D, ZHOU Y T, CHEN H M, et al. Hybrid rank-sparsity constraint model for simultaneous reconstruction and denoising of 3D seismic data [J]. *Geophysics*, 2017, 82(5): 351-367.
- [10] ZHANG Y J, ZHOU L, HU ZH J, et al. Prediction of layered thermal conductivity using artificial neural network in order to have better design of ground source heat pump system[J]. *Energies*, 2018, 11(7): 1896.
- [11] GLOVER P W J, LUO M. The porosity and permeability of binary grain mixtures [J]. *Transp. Porous Media*, 2020, 132(1): 1-37.
- [12] LUO M, GLOVER P W J, ZHAO P Q, et al. 3D digital rock modeling of the fractal properties of pore structures [J]. *Marine and Petroleum Geology*, 2020, 122: 104706.
- [13] REN Y L, GONG R B, FENG ZH, et al. Valuable data extraction for resistivity imaging logging interpretation [J]. *Tsinghua Science and Technology*, 2019, 25 (2): 281-293.
- [14] ANIETIE N O, STEVE E A, EMMANUEL M U. Artificial neural network model for reservoir petrophysical properties: porosity, permeability and water saturation prediction [J]. *Modeling Earth Systems and Environment*, 2021, 7(4): 2373-2390.
- [15] WANG J, CAO J X. Deep learning reservoir porosity prediction using integrated neural network [J]. *Arabian Journal for Science and Engineering*, 2022, 47 (9): 11313-11327.
- [16] 崔学慧, 刘长江, 薛亮. 基于随机森林回归算法预测孔隙度的研究 [J]. *西部探矿工程*, 2019, 31 (11): 99-102, 105.
- CUI X H, LIU CH J, XUE L. Research on porosity prediction based on random forest regression algorithm [J]. *West-China Exploration Engineering*, 2019, 31 (11): 99-102, 105.
- [17] 潘少伟, 郑泽晨, 雷建永, 等. 基于混合优化 XGBoost 算法的砂岩储层孔隙度预测 [J]. *计算机应用与软件*, 2023, 40(5): 103-109, 206.
- PAN SH W, ZHENG Z CH, LEI J Y, et al. Porosity prediction of sandstone reservoir based on hybrid optimization XGBoost algorithm [J]. *Computer Applications and Software*, 2023, 40(5): 103-109, 206.
- [18] 侯贤沐, 王付勇, 宰芸, 等. 基于机器学习和测井数据的碳酸盐岩孔隙度与渗透率预测 [J]. *吉林大学学报 (地球科学版)*, 2022, 52(2): 644-653.
- HOU X M, WANG F Y, ZAI Y, et al. Prediction of carbonate porosity and permeability based on machine learning and loggin data [J]. *Journal of Jilin University (Earth Science Edition)*, 2022, 52(2): 644-653.
- [19] 汪敏, 杨桃, 唐洪明, 等. 迁移深度神经网络的页岩总孔隙度预测 [J]. *西南石油大学学报 (自然科学版)*, 2023, 45(6): 69-79.
- WANG M, YANG T, TANG H M, et al. Prediction for total porosity of shale based on transfer deep neural

- network[J]. Journal of Southwest Petroleum University (Science & Technology Edition), 2023, 45(6): 69-79.
- [20] 董美娜,刘丽平,王泽忠,等. 基于 Stacking 集成学习的有源台区线损率评估方法[J]. 电测与仪表, 2023, 60(6): 134-139,173.
- DONG M N, LIU L P, WANG Z ZH, et al. A line loss rate evaluation method based on stacking ensemble learning for transformer district with DG[J]. Electrical Measurement & Instrumentation, 2023, 60(6): 134-139, 173.
- [21] AKIBA T, SANO S, YANASE T, et al. Optuna: A next-generation hyperparameter optimization framework[C]. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019: 2623-2631.

作者简介



梁海波(通信作者),2000 于西南石油大学获得学士学位,2003 于西南石油大学获得硕士学位,2008 于西南石油大学获得博士学位,现为西南石油大学教授,主要研究方向为油气井随钻测量新装备、智能测控系统、油气地质导向分析和信息交互分析。

E-mail: secondbo@126.com

Liang Haibo (Corresponding author) received his B. Sc. degree from Southwest Petroleum University in 2000, M. Sc. degree from Southwest Petroleum University in 2003, and Ph. D. degree from Southwest Petroleum University in 2008. He is currently a professor at Southwest Petroleum University. His main research interests include new equipment for measuring oil and gas wells while drilling, intelligent measurement and control system, oil and gas geological guidance analysis, and information exchange analysis.