基于 TPE 优化集成学习的岩石弹性模量预测模型

孟祥龙[a,b*]，王胜建[a,b]，朱迪斯[a,b]，马彦彦[a,b]，李大勇[a,b]，迟焕鹏[a,b]，张家政[a,b]，岳伟民[a,b]

中国地质调查局 a. 油气资源调查中心；b. 非常规油气地质重点实验室，北京 100083

* 通信作者，E-mail：mengxianglong001@mail.cgs.gov.cn

2024-06-13 收稿；2024-09-23 修回；2024-09-24 接受

基金项目：中国地质调查局油气资源调查中心科技创新基金项目（油科创 [2023]-QN02）；中国地质调查局项目（DD20242400）

中图分类号：TE311；TP181　　文章编号：2096-8523（2026）01-0342-09　　doi：10.19509/j.cnki.dzkq.tb20240325

摘　要：油气工程中常利用地球物理资料获取地层弹性模量并结合小样本的岩心实验数据进行校正，但这种方法在复杂地质条件下往往表现不佳。为提高岩石弹性模量的预测精度和泛化能力，提出了一种利用基本岩石物性参数的弹性模量智能预测模型。分别采用 3 种集成学习算法（RandomForest, XGBoost, LightGBM）构建了岩石弹性模量智能预测模型，并采用 TPE 方法对模型进行超参数优化，最后利用 SHAP 归因分析探讨了各输入变量对模型的贡献。结果表明：①提出的智能预测模型明显优于传统模型，能够实现弹性模量的精确预测并具有较强的泛化能力，其中 XGBoost 模型表现最佳（决定系数 R^2=0.87，均方根误差 *RMSE*=6.94，平均绝对误差 *MAE*=4.96）；②横波速度对模型贡献最大，纵波速度次之，密度最小，精确横波波速对弹性模量预测有重要意义。该方法无需对工区及地层进行预先识别即可实现弹性模量的精准预测，研究成果对油气工程设计及实施有重要参考意义。

关键词：弹性模量；TPE；集成学习；SHAP；横波

岩石弹性参数的获取方式分为静态法和动态法。静态法对完整岩心进行三轴压缩实验测试其变形，是获得岩石弹性参数最可靠的手段[1]。深部地层岩心获取成本高，动态法能够通过地球物理资料获得地层的弹性参数[2-3]，成本低廉且能够在地层条件下直接测试，广泛应用于油气工程等深部工程。但岩石并非理想的线弹性材料，动态法获得的弹性模量（动态弹性模量）和三轴压缩实验获得的弹性模量（静态弹性模量）并不一致，无法直接用于表征静力荷载作用下的岩土体的变形行为，非弹性变形、排水效应、频散效应以及岩石的非均质性和各向异性都被认为是造成动、静态弹性模量之间差异的原因[4-8]，且目前没有任何岩石物理模型能够囊括所有原因。工程中通常采用少量岩心实验回归经验方程的方式来对动态弹性模量进行校正[9-10]，但由于岩心往往仅从油气储层采集，工区之间以及工区内部不同年代的地层之间存在较大差异，由小样本实验数据构建的经验方程在跨工区及复杂地质条件下往往预测效果不佳。

近年来，一些人工智能算法被引入弹性模量等工程参数的预测。王文文等[11]利用 32 块砂岩样品的纵波速度、密度和孔隙度数据构建了基于遗传规划算法的弹性模量预测公式；ELKATATNY 等[12]利用伽马、中子孔隙度、密度、纵波时差、横波时差等参数基于多元回归实现弹性模量预测；NASIRI 等[13]提出了利用孔隙度、点载荷指数、P 波速度和 Schmidt 锤回弹数基于 ANN 预测弹性模量的方法。这些方法的研究区域大多针对单一工区或地层，新引入的一些参数无法直接观测或计算过程需要人为给定控制参数，给这些方法的应用带来限制。另一方面，这些模型中新引入的孔隙度等参数和波速密切相关，这也说明波速等基本岩石物性参数和弹性模量之间存在更为复杂的关系。

相对于单一的人工智能算法，集成学习算法通

著录信息：孟祥龙，王胜建，朱迪斯，等. 基于 TPE 优化集成学习的岩石弹性模量预测模型［J］. 地质科技通报，2026，45（1）：342-350.
MENG Xianglong, WANG Shengjian, ZHU Disi, et al. Prediction model for rock elastic modulus based on TPE-optimized ensemble learning［J］. Bulletin of Geological Science and Technology, 2026, 45（1）: 342-350.

© Editorial Office of Bulletin of Geological Science and Technology. This is an open access article under the CC BY-NC-ND license.　https://dzkjtb.cug.edu.cn

过构建大量基学习器获得更高的预测精度和泛化能力[14-16]，建模简单且在无需考虑确切的机理条件下挖掘数据中的隐藏关系以解决复杂的参数预测问题，这将有助于解决岩土工程参数预测中对大量经验公式的依赖问题，如李超等[17]利用 384 组数据构建的土固结压力智能预测模型效果显著优于经验模型，SUN 等[18]利用 XGBoost 算法实现了岩石单轴抗压强度的智能预测，为提高弹性模量的预测精度和泛化能力提供了一种新思路。

本研究选用纵波速度、横波速度和密度 3 种能同时通过岩心和地球物理方法直接测量的基本岩石物理参数作为输入特征参数，分别利用 3 种集成学习算法（RandomForest, XGBoost, LightGBM）构建了岩石弹性模量智能预测模型，结合统计回归模型对智能预测模型的性能进行了评估，并探讨了各特征参数对模型的贡献。

1 技术方法

1.1 集成学习原理

（1）RandomForest 算法

随机森林（RandomForest）算法是一种以决策树为基学习器的算法，对于回归问题，其采用分类与回归树（classification and regression tree，简称 CART）决策树算法进行回归运算，采用袋装法（Bagging）将训练结果进行聚合[19]。即在训练过程中随机抽取 t 个独立的子集进行独立的训练，构建 t 个 CART 模型，将所有的模型聚合取平均值作为最终的预测结果，其表达式为式（1）。RandomForest 算法能够集成多个回归模型，改善单一回归模型的不足，使预测结果更具有抗噪性和稳定性。

$$\hat{y}_i = \frac{1}{t}\sum_{k=1}^{t} f_k(x_i), \tag{1}$$

式中：$\hat{y}_i$ 为第 i 个样本的预测值；t 为决策树（基学习器）数量；x_i 为第 i 个样本的输入数据；f_k 为第 k 个决策树回归模型。

（2）XGBoost 算法

极限梯度提升树（extreme gradient boosting，简称 XGBoost）算法[20]是在梯度提升决策树（gradient boosting decision tree，简称 GBDT）算法基础上的一种改进算法，其基于 Boosting 算法对基学习器进行训练，与随机森林构建大量独立的基学习器（CART）不同，Boosting 算法在每一个基学习器的训练中都会考虑上一个基学习器和实测值之间的差异，通过不断迭代达到逼近实际值的效果，最后累加所有基学习器输出结果，表达式为：

$$\hat{y}_i = \sum_{k=1}^{t} f_k(x_i) = \sum_{k=1}^{t-1} f_k(x_i) + f_k^{(t)}(x_i), f_k \in F, \tag{2}$$

式中：F 为回归树的样本空间；$f_k^{(t)}$ 为当前需要优化的决策树。

XGBoost 在目标函数中引入正则化项以控制运行效率和防止过拟合，通过二阶泰勒展开来简化损失函数的运算实现最优解的求取，目标函数 Obj 的表达式为：

$$Obj = \sum_{i=1}^{N} L(y_i, \hat{y}_i) + \sum_{k=1}^{t} \Omega(f_k), \tag{3}$$

$$\Omega(f_k) = \gamma_\Omega T + 1/2\lambda_\Omega \sum_{k=1}^{T} w_k^2, \tag{4}$$

式中：$\sum_{i=1}^{N} L(y_i, \hat{y}_i)$ 为损失函数；N 为样本数量；y_i 为第 i 个样本的实际值；$\sum_{k=1}^{t} \Omega(f_k)$ 为惩罚函数（正则项）；T 为叶子节点数量；w_k^2 为叶子权重节点的 L2 范数；γ_Ω 和 λ_Ω 为惩罚系数。

（3）LightGBM 算法

轻量梯度提升机（light gradient boosting machine，简称 LightGBM）算法[21]和 XGBoost 类似，基于 Boosting 算法对基学习器进行训练，其采用直方图决策树算法寻找特征值最优的分割点，并利用基于梯度的单边采样（gradient-based one-side sampling，简称 GOSS）算法进行样本采样同时采用带有深度限制的叶子生长（leaf-wise）算法，其主要特点是能够有效减少计算时间和计算规模，在降低模型复杂度的同时提高抗过拟合能力。

1.2 超参数优化

超参数的选取直接关系到模型的性能，但超参数和预测结果之间的关系无法用函数表达，是典型的“黑盒问题”，为了实现模型的最佳预测效果以及便于应用，自动化的超参数优化至关重要[22]。

树结构 Parzen 估计器（tree-structured parzen estimator，简称 TPE）优化算法是一种基于树结构的贝叶斯优化的算法[23]。其通过构建代理函数代替实际模型，采用序贯模型对全局进行优化，即利用前序采样的结果建立概率分布，利用采集函数搜索下一个测试点，在不断迭代训练过程中，实现预测效果增强，这种优化算法能够降低运算量，大大减少运算时间。相对遍历所有点的网格搜索算法和随机搜索若干个点的随机搜索算法更具优势。

具体来说，TPE 方法对条件分布概率 $P(x|y)$ 和边缘概率 $P(y)$ 建模，定义 $P(x|y)$ 的表达式为：

$$P(x|y)=\begin{cases}l(x), & y\leqslant y^*\\ g(x), & y>y^*\end{cases}, \tag{5}$$

式中：x 为观测值；y 为输出值；y^* 为设置的阈值；$l(x)$ 和 $g(x)$ 分别为损失函数≤ y^* 和＞ y^* 的密度组成。

期望函数（*EI*）被用作采集函数，利用贝叶斯原理计算后验概率，采集函数（*EI*）的表达式为：

$$EI=\int_{-\infty}^{y^*}(y^*-y)P(y|x)\mathrm{d}y=\int_{-\infty}^{y^*}\frac{P(x|y)P(y)}{P(x)}\mathrm{d}y。 \tag{6}$$

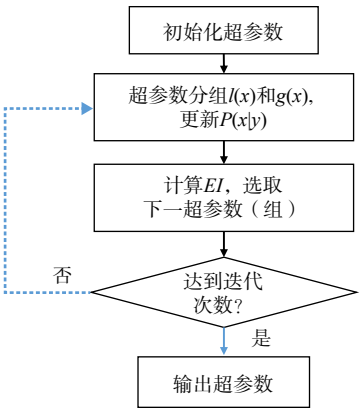
定义 $\gamma=P(y\leqslant y^*)$，则可以变换得到

$$P(x)=\int_{\mathrm{R}}P(x|y)P(y)\mathrm{d}y=\gamma l(x)+(1-\gamma)g(x)。 \tag{7}$$

因此，

$$EI=\frac{\gamma y^* l(x)-l(x)\int_{-\infty}^{y^*}P(y)\mathrm{d}y}{\gamma l(x)+(1-\gamma)g(x)}\propto\left(\gamma+\frac{g(x)}{l(x)}(1-\gamma)\right)^{-1}。 \tag{8}$$

根据式（8），由最大概率的 $l(x)$ 和最小概率的 $g(x)$ 获得最大的 *EI* 来实现下一组超参数的选择，进而不断更新 $P(x|y)$ 和 $P(y)$，直到迭代停止，获得最优化的超参数，优化算法流程见图 1。



$l(x)$，$g(x)$ 分别为损失函数≤ y^* 和＞ y^* 的密度组成；$P(x|y)$ 为条件分布概率；*EI* 为期望函数

图 1　TPE 超参数优化算法流程示意图

Fig. 1　Schematic workflow of TPE hyperparameter optimization algorithm

1.3　模型的可解释性

SHAP（SHapley Additive exPlanation）[17] 是一种基于博弈论原理的事后解释方法，利用线性模型对机器学习模型进行逼近，对模型中各个特征及其交互项的边际贡献进行计算来衡量它们的影响大小，从而实现对“黑盒模型“的解释。*SHAP* 值（*Shapley* 值）用来表征输入贡献度，*SHAP* 值越大代表贡献越高[24-26]。SHAP 模型的表达式为：

$$f(x_i)=\phi_0+\sum_{i=1}^{M}\phi_i, \tag{9}$$

式中：M 为模型中的特征数量；ϕ_0 为所有样本数量的平均值；ϕ_i 为第 i 个特征的 *SHAP* 值。

ϕ_i 的理论表达式为：

$$\phi_i=\sum_{S\subseteq\{M\backslash x_i\}}\frac{|S|!(M-|S|-1)!}{M!}\{f(x_{S\cup\{i\}})-f(x_S)\}, \tag{10}$$

式中：S 为 $\{M\backslash x_i\}$ 的特征子集；$f(x_{S\cup\{i\}})$ 和 $f(x_S)$ 分别为各种特征组合下 x_i 加入模型及不加入模型时输出结果；$|S|!(M-|S|-1)!$ 为各种特征组合对应的概率；| |为集合的元素个数；!为阶乘。

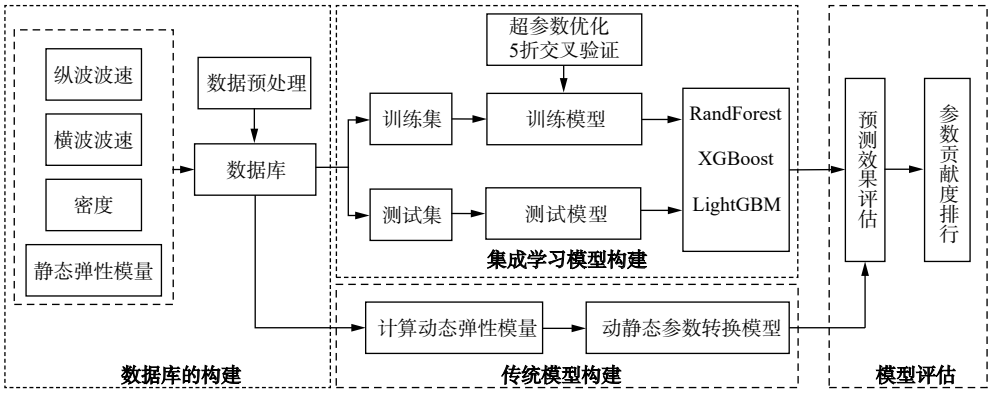
1.4　技术路线

本研究通过建立岩心实验数据库对 RandomForest, XGBoost, LightGBM 3 种集成学习模型训练和评估，具体步骤为：①搜集公开发表文献中的岩石力学实验数据以构建数据库，每组数据包括岩心的密度、纵波速度、横波速度、三轴压缩实验获得的弹性模量以及岩石样本的基本信息，基于数据库计算岩心的动态弹性模量，构建岩石的动静态弹性模量转换统计模型；②将数据集划分为训练集和测试集，利用 Python 语言的 Scikit-learn 机器学习库联合 Numpy、Pandas 等数据处理库对 RandomForest, XGBoost, LightGBM 进行训练，利用 K 折交叉验证和 TPE 进行优化，得到最优化模型；③基于优化后的超参数训练模型，采用测试集对比几种智能模型的预测效果，并结合动静态弹性模量转换统计模型评估模型性能；④利用 SHAP 归因分析，探讨不同变量对模型的贡献（图 2）。

2　实例研究

2.1　数据基本情况

本研究采用的数据全部来源于公开发表文献，主要为砂岩、泥岩、石灰岩等沉积岩，具体包括岩心采集的基本信息以及密度、纵波速度、横波速度和三轴压缩实验获得的弹性模量剔除有明显偏差及不规范的数据，共计收集到 397 组有效数据，来源于多个国家的十余个工区，具体见表 1[11, 27-35]。虽然岩心实验数据来源范围广，但其采用的实验方法原理基本一致，考虑到不同年代测试精度的不同，这里对数据采用统一单位的预处理方式，密度单位采用 g/cm^3，



RandomForest. 随机森林; XGBoost. 极限梯度提升树; LightGBM. 轻量梯度提升机; 下同

图 2　技术路线

Fig. 2　Technical roadmap

表 1　数据来源

Table 1　Data sources

数据来源	岩性	采样地区
文献[11]	砂岩	—
文献[27]	泥岩	中国泸州
文献[28]	砂岩	中国鄂尔多斯
文献[29]	砂岩、砂砾岩等	中国准噶尔盆地
文献[30]	各种岩性	—
文献[31]	泥灰岩、石灰岩、砂岩	伊朗
文献[32]	石灰岩	中东伊朗
文献[33]	粉砂岩、石灰岩和白云岩	美国
文献[34]	白云岩	埃及
文献[35]	灰岩和白云岩	伊朗扎格罗斯山

波速单位采用 km/s, 弹性模量单位采用 GPa。

表 2 和图 3 展示了各参数的数据分布情况和参数间的相关性。本研究所采用的 3 种智能算法都是以对变量特征缩放不明显的决策树为基学习器, 虽然变量的数据分布没有严格服从正态分布, 也无需进行标准化或归一化处理。图 3 显示各变量之间普遍具有较强相关性, 由于变量之间的共线性不会影响集成学习的效果, 所有变量都将被用于建模。

2.2　动静态弹性模量转换统计模型

动态法计算弹性模量基于线弹性假设, 机械波在固体中传播引起完全弹性形变, 由纵波速度(V_p)、横波速度(V_s)和密度(ρ)可获得弹性参数[36-38]。动态弹性模量(E_{dyn})的计算公式为:

$$E_{dyn}=\frac{\rho V_s^2\left(3V_p^2-4V_s^2\right)}{V_p^2-V_s^2}。\tag{11}$$

图 4 显示了动态弹性模量和静态弹性模量之间的关系, 可以看出, 无论是单一数据源内部还是数据整体, 动态和静态弹性模量之间都存在明显的正相关关系。基于统计回归构建动静态弹性模量转换模型, 线性拟合效果较幂函数、指数函数和对数函数好, 拟合方程为:

$$E_{sta}=0.87E_{dyn}-7,\tag{12}$$

式中: E_{sta} 为三轴压缩实验得到的弹性模量, GPa。

表 2　数据分布情况

Table 2　Data distribution

参数	密度/($g\cdot cm^{-3}$)	静态弹性模量/GPa	纵波速度/($km\cdot s^{-1}$)	横波速度/($km\cdot s^{-1}$)
平均值	2.49	26.64	4.13	2.41
标准差	0.25	18.99	1.10	0.58
中位数	2.53	12.00	3.51	2.13
最大值	3.16	93.9	6.57	3.73
最小值	1.18	0.77	1.78	0.90

2.3　数据集的划分与交叉验证

本研究将数据集按照 8∶2 的比例划分为训练集和测试集, 训练集的数据将被用于训练模型, 测试集的数据用于对模型预测性能进行评估。考虑到数据规模较小, 为提升模型的泛化能力, 采用 K-fold 交叉验证的方式引入验证集, 模型在训练中被划分成 K 个互斥子集, 每次选取一个子集作为验证集, 其他子集作为训练集, 能够进行 K 次训练, 采用 K 次训练指标的平均值作为模型的评估指标来减少预测偏差和避免过拟合, 本研究采用 5 折交叉验证。

2.4　基于 TPE 的超参数优化

本研究利用 hyperopt 库实现 TPE 方法对集成学习模型的超参数优化, 在 5 折交叉验证的条件下, 以决定系数(R^2)最大为目标, 设置 100 次迭代, 优化后的参数如表 3 所示, 其余参数则采用默认设置。

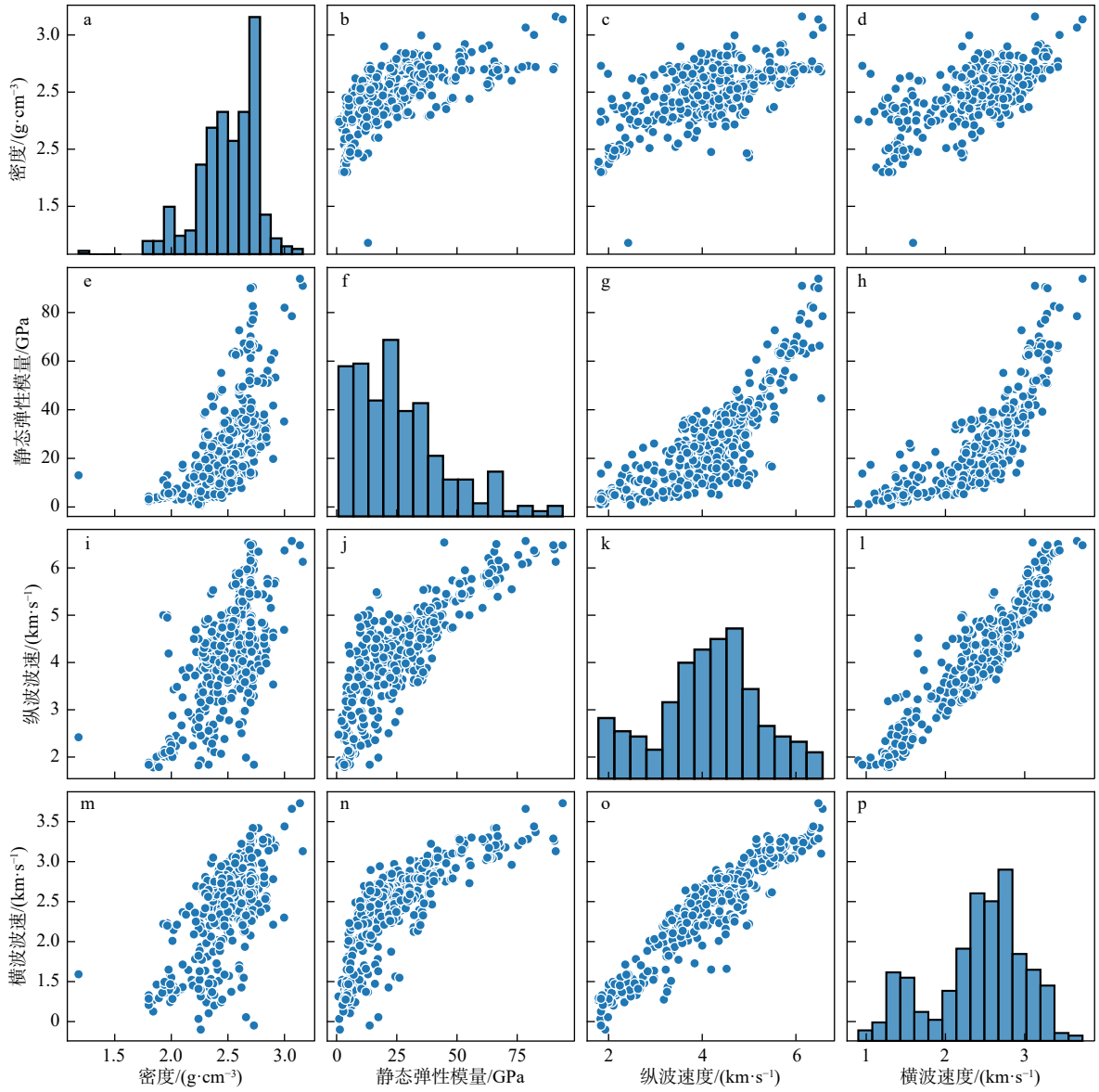


图 3　变量数据分布特征和相关性

Fig. 3　Distribution characteristics and correlation of variable data

2.5　模型评价指标

以下指标用于评估模型的预测效果：R^2 在 0 和 1 之间，值越大表明拟合程度越高；平均绝对误差（*MAE*）是模型预测误差的平均数，值越小则反映模型预测效果更好；均方根误差（*RMSE*）值越小则反映模型预测效果更好，其对异常偏离数据更加敏感。

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y}_i)^2}, \tag{13}$$

$$MAE = \frac{\sum |y_i - \hat{y}_i|}{N}, \tag{14}$$

$$RMSE = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{N}}, \tag{15}$$

式中：$\bar{y}_i$ 为样本平均值。

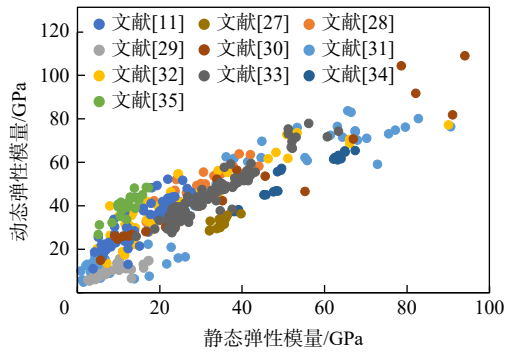


图 4　动态和静态弹性模量关系

Fig. 4　Relationship between Dynamic and static elastic moduli

表 4 列出了各个模型的评价指标数据，可以看出 3 种集成学习模型均表现出了良好的效果。在训练集上，3 种模型的 R^2 均接近 1，且 *RMSE* 和 *MAE*

表 3　优化后的超参数

Table 3　Optimized hyperparameters

LightGBM 超参数	取值	XGBoost 超参数	取值	RandomForest 超参数	取值
alpha	11.05	*alpha*	9.23	*max_depth*	25
learning_rate	0.19	*learning_rate*	0.09	*n_estimators*	1011
max_depth	36	*max_depth*	35		
min_data_in_leaf	16	*subsample*	0.46		
n_estimators	310	*n_estimators*	163		
num_leaves	43	*num_leaves*	93		
reg_lambda	1.71	*reg_lambda*	2.86		
		eta	0.02		

表 4　预测模型的性能对比

Table 4　Performance comparison of prediction models

预测模型		RandomForest	XGBoost	LightGBM	统计回归模型
R^2	训练集	0.98	0.98	0.97	0.78
	测试集	0.86	0.87	0.86	
	交叉验证集	0.86	0.86	0.85	
RMSE	训练集	2.54	2.70	3.41	9.11
	测试集	7.16	6.87	7.14	
	交叉验证集	6.90	6.94	7.04	
MAE	训练集	1.78	1.88	2.27	6.80
	测试集	4.93	4.94	5.00	
	交叉验证集	4.87	4.96	5.19	

注：R^2. 决定系数；*RMSE*. 均方根误差；*MAE*. 平均绝对误差

都较小，说明 3 种模型对该组数据都具有较强的学习能力，其中 XGBoost 和 RandomForest 表现效果相似，略优于 LightGBM。在测试集上，3 种模型同样表现效果较好，表现出了良好的预测能力，其中 XGBoost 模型预测效果最佳，优于 RandomForest 和 LightGBM 模型。交叉验证集的测试效果能反映模型的泛化能力，3 个模型基本相似，说明 3 个模型都具有较强的泛化能力。图 5 展示了模型预测值和实际值的关系，可见 3 种集成学习模型预测结果和实际值构成的散点均沿着对角线分布。综上，3 种集成学习模型都具有良好的学习和预测能力和较强的泛化能力，XGBoost 模型在本研究的数据集中表现效果最佳。

动静态弹性模量统计回归模型的 R^2 仅为 0.78，其他各项指标均差于集成学习模型。图 4 中数据点呈现为较宽的条带状分布，每一个静态弹性模量对

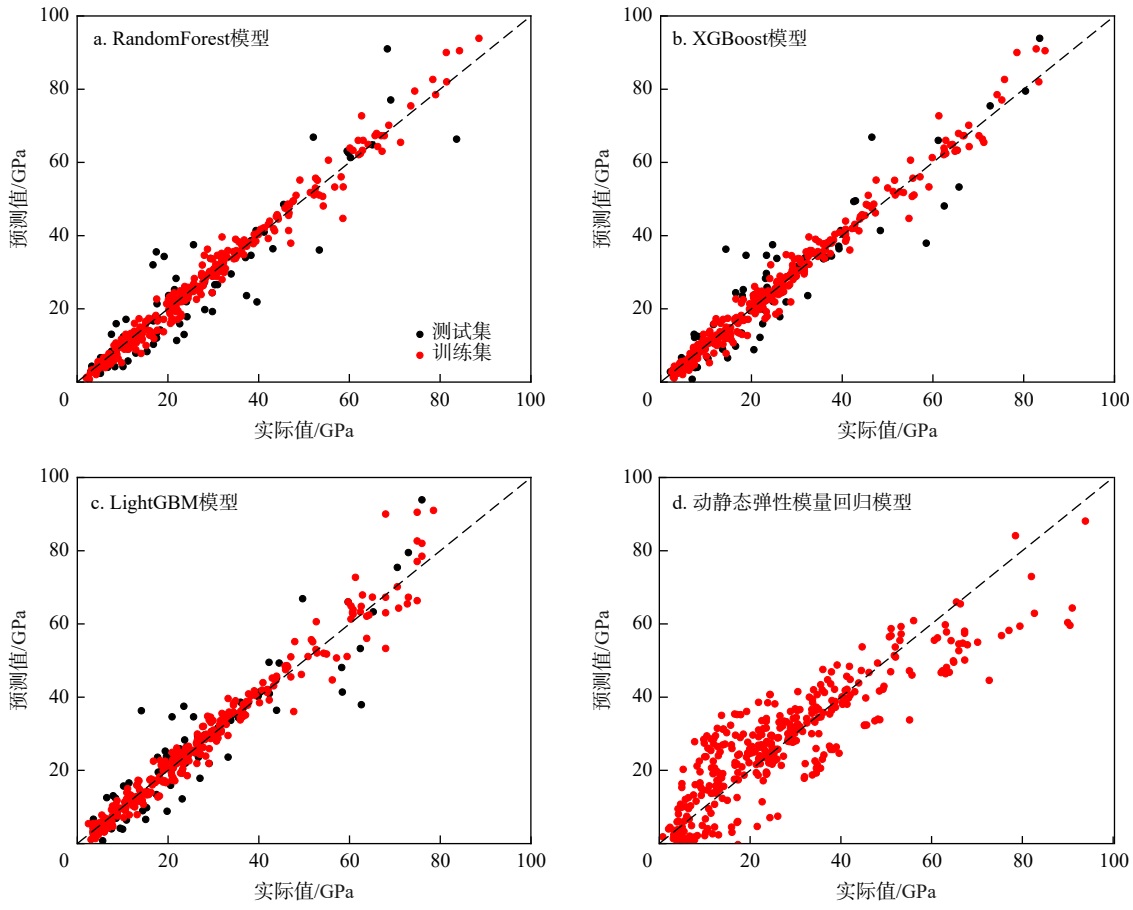


图 5　模型预测值和实际值对比

Fig. 5　Comparison between model-predicted values and actual values

应多个动态弹性模量，这表明单一的函数无法实现动静态弹性模量的有效转换，引入更多的特征的集成学习算法在弹性模量预测上更具有优势。

2.6　特征重要性分析

为探索各特征参数对模型的影响，本研究对表现最佳的 XGBoost 模型进行 SHAP 归因分析以评估各输入特征参数对模型的贡献度。图 6 显示了 3 个变量的 *SHAP* 值：横波速度>纵波速度>密度，横波波速对模型的贡献度最大。岩石纵横波速度之间的关系与岩石物性及孔隙微观结构密切相关，而岩石的孔隙结构也和力学行为具有直接关系，相对于纵波，横波速度较慢，采集精度更高，且频散效应更弱，不同频率下速度的变化更为稳定，这可能是横波贡献更大的原因之一。另一方面，伍顺伟等[39] 和赵旭阳等[40] 基于实际工程数据的研究也指出了精确横波数据对岩石力学参数预测的重要性。横波速度包含的特征将有助于提高弹性模量的预测精度和泛化能力，精确横波数据对弹性模量的预测有重要意义。

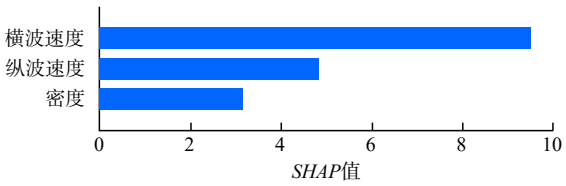


图 6　XGBoost 模型各输入变量的 *SHAP* 值

Fig. 6　*SHAP* values of input variables in XGBoost model

3　结论和展望

（1）利用 TPE 优化 3 种集成学习算法（RandomForest, XGBoost, LightGBM）分别构建了岩石弹性模量智能预测模型，能够在无需识别工区和地层情况下实现对弹性模量的精准预测，其中 XGBoost 模型表现效果最佳，明显优于统计回归模型，具有较好的预测精度和泛化能力，该方法能够为油气工程的设计和实施提供重要参考。

（2）基于 SHAP 归因分析，得到了各参数对于 XGBoost 模型贡献大小：横波速度>纵波速度>密度，精确横波速度对弹性模量预测具有重要意义，但其中具体的机制仍有待进一步的研究。

（3）本研究采集的数据量较少且样本多为沉积岩，深层油气及新能源和矿产的开发涉及更多地区和地层，进一步的研究可引入更多的岩性和更大量的数据，同时将标准化的岩心实验归位到测井曲线能够引入更多特征参数以构建更大和更详细范围的数据库，这将有助于进一步认知岩石物理参数之间的内在联系，构建泛化能力更强的模型，推广地球物理方法在弹性参数预测中的应用。

所有作者声明不存在利益冲突。

The authors declare that no competing interests exist.

参考文献：

[1] 蔡美峰. 岩石力学与工程 [M]. 北京：科学出版社，2002.
CAI M F. Rock mechanics and engineering[M]. Beijing: Science Press, 2002.（in Chinese）

[2] 李红斌，王贵文，庞小娇，等. 苏北盆地古近系阜宁组页岩工程品质测井评价 [J]. 地质科技通报，2023，42(3)：311-322.
LI H B, WANG G W, PANG X J, et al. Logging evaluation of the engineering quality of the Paleogene Funing Formation oil shales in the Subei Basin[J]. Bulletin of Geological Science and Technology, 2023, 42(3): 311-322.（in Chinese with English abstract）

[3] 舒红林，仇凯斌，李庆飞，等. 页岩气地质力学特征评价方法：中国南方海相强改造区山地页岩地质力学特征 [J]. 天然气工业，2021，41(增刊 1)：1-13.
SHU H L, QIU K B, LI Q F, et al. A method for evaluating the geomechanical characteristics of shale gas: The geomechanical characteristics of the mountain shale in the intensively reworked marine area of South China[J]. Natural Gas Industry, 2021, 41(S1): 1-13.（in Chinese with English abstract）

[4] 葛洪魁，陈颙，林英松. 岩石动态与静态弹性参数差别的微观机理 [J]. 石油大学学报（自然科学版），2001，25(4)：34-36.
GE H K, CHEN Y, LIN Y S. Microscopic mechanism of difference between static and dynamic elastic parameters of rock[J]. Journal of the University of Petroleum, China, 2001, 25(4): 34-36.（in Chinese with English abstract）

[5] 尤明庆，苏承东，申江. 岩石材料的非均质性与动态参数 [J]. 辽宁工程技术大学学报（自然科学版），2001，20(4)：492-494.
YOU M Q, SU C D, SHEN J. Effect of heterogeneity on the dynamic parameters of rock[J]. Journal of Liaoning Technical University（Natural Science Edition），2001，20(4)：492-494.（in Chinese with English abstract）

[6] FJÆR E. Relations between static and dynamic moduli of sedimentary rocks[J]. Geophysical Prospecting, 2019, 67(1): 128-139.

[7] 陈勉，金衍，张广清. 石油工程岩石力学 [M]. 北京：科学出版社，2008.
CHEN M, JIN Y, ZHANG G Q. Rock mechanics of petroleum engineering[M]. Beijing: Science Press, 2008.（in Chinese）

[8] LOZOVYI S, BAUER A. From static to dynamic stiffness of shales: Frequency and stress dependence[J]. Rock Mechanics and Rock Engineering, 2019, 52(12): 5085-5098.

[9] 蔡文军，邓金根，冯永存，等. 泥页岩地层区域三维地质力学建模与应用 [J]. 钻采工艺，2023，46(1)：8-14.
CAI W J, DENG J G, FENG Y C, et al. Three-dimensional geomechanical modeling and application in shale formation[J]. Drilling & Production Technology, 2023, 46(1): 8-14.（in Chinese with English abstract）

[10] 杨宝刚，潘仁芳，赵丹，等. 四川盆地长宁示范区龙马溪组页岩岩石力学特性及脆性评价 [J]. 地质科技情报，2015，34(4)：183-188.
YANG B G, PAN R F, ZHAO D, et al. Mechanical properties and

brittleness evaluation of Longmaxi shale rock of Changning demonstration area in Sichuan Basin[J]. Geological Science and Technology Information, 2015, 34(4): 183-188. (in Chinese with English abstract)

[11] 王文文, 李诺, 韩建强, 等. 基于多基因遗传规划的储层岩石静态模量预测 [J]. 应用声学, 2020, 39(2): 300-305.
WANG W W, LI N, HAN J Q, et al. Static modulus prediction for reservoir rocks based on multi-gene genetic programming[J]. Journal of Applied Acoustics, 2020, 39(2): 300-305. (in Chinese with English abstract)

[12] ELKATATNY S, MAHMOUD M, MOHAMED I, et al. Development of a new correlation to determine the static Young's modulus[J]. Journal of Petroleum Exploration and Production Technology, 2018, 8(1): 17-30.

[13] NASIRI H, HOMAFAR A, CHELGANI S C. Prediction of uniaxial compressive strength and modulus of elasticity for Travertine samples using an explainable artificial intelligence[J]. Results in Geophysical Sciences, 2021, 8: 100034.

[14] 杨剑锋, 乔佩蕊, 李永梅, 等. 机器学习分类问题及算法研究综述 [J]. 统计与决策, 2019, 35(6): 36-40.
YANG J F, QIAO P R, LI Y M, et al. A review of machine-learning classification and algorithms[J]. Statistics & Decision, 2019, 35(6): 36-40. (in Chinese with English abstract)

[15] 徐继伟, 杨云. 集成学习方法: 研究综述 [J]. 云南大学学报(自然科学版), 2018, 40(6): 1082-1092.
XU J W, YANG Y. A survey of ensemble learning approaches[J]. Journal of Yunnan University (Natural Sciences Edition), 2018, 40(6): 1082-1092. (in Chinese with English abstract)

[16] 肖立志. 机器学习数据驱动与机理模型融合及可解释性问题 [J]. 石油物探, 2022, 61(2): 205-212.
XIAO L Z. The fusion of data-driven machine learning with mechanism models and interpretability issues[J]. Geophysical Prospecting for Petroleum, 2022, 61(2): 205-212. (in Chinese with English abstract)

[17] 李超, 汪磊, 陈洋, 等. 基于贝叶斯集成学习算法的土体先期固结压力预测模型 [J]. 地球科学, 2023, 48(5): 1780-1792.
LI C, WANG L, CHEN Y, et al. Prediction model of soils' preconsolidation pressure based on Bayesian ensemble learning algorithm[J]. Earth Science, 2023, 48(5): 1780-1792. (in Chinese with English abstract)

[18] SUN J L, ZHANG R, ZHANG A L, et al. Rock strength prediction based on machine learning: A study from prediction model to mechanism explanation[J]. Measurement, 2024, 238: 115373.

[19] BREIMAN L. Random forests[J]. Machine Learning, 2001, 45: 5-32.

[20] CHEN T Q, GUESTRIN C. XGBoost: A scalable tree boosting system[C]//Anon. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco California USA: ACM, 2016: 785-794.

[21] KE G, MENG Q, FINLEY T, et al. LightGBM: A highly efficient gradient boosting decision tree[C]//Anon. Neural Information Processing Systems. [S. 1.]: [s. n.], 2017: 3146-3154.

[22] 李阳, 代宗仰, 张洁伟, 等. 基于无监督学习的多参数储层评价: 以蒲包山地区下三叠统飞仙关组礁滩储层为例 [J]. 地质科技通报, 2023, 42(5): 285-292.
LI Y, DAI Z Y, ZHANG J W, et al. Multiparameter reservoir evaluation method based on unsupervised learning: A case study of the reef beach reservoir of the Lower Triassic Feixianguan Formation in the Pubaoshan area[J]. Bulletin of Geological Science and Technology, 2023, 42(5): 285-292. (in Chinese with English abstract)

[23] BERGSTRA J S, BARDENET R, BENGIO Y, et al. Algorithms for hyper-parameter optimization[C]//Anon. Neural Information Processing Systems. [S. 1.]: [s. n.], 2011: 2546-2554.

[24] LUNDBERG S M and LEE S I. A unified approach to interpreting model predictions[C]//Anon. Proceedings of the 31st International Conference on Neural Information Processing Systems. [S. 1.]: [s. n.], 2017: 4765-4774.

[25] LUNDBERG S M, ERION G, CHEN H, et al. From local explanations to global understanding with explainable AI for trees[J]. Nature Machine Intelligence, 2020, 2(1): 56-67.

[26] 陈曦泽, 贾俊峰, 白玉磊, 等. 基于 XGBoost-SHAP 的钢管混凝土柱轴向承载力预测模型 [J]. 浙江大学学报(工学版), 2023, 57(6): 1061-1070.
CHEN X Z, JIA J F, BAI Y L, et al. Prediction model of axial bearing capacity of concrete-filled steel tube columns based on XGBoost-SHAP[J]. Journal of Zhejiang University (Engineering Science), 2023, 57(6): 1061-1070. (in Chinese with English abstract)

[27] 李新林. 川南深层页岩气储层地应力测井评价方法研究 [D]. 武汉: 长江大学, 2022.
LI X L. Study on logging evaluation method of in-situ stress for deep shale gas reservoir in southern Sichuan[D]. Wuhan: Yangtze University, 2022. (in Chinese with English abstract)

[28] 江昊焱. X 地区致密砂岩储层岩石力学特性及体积压裂甜点研究 [D]. 西安: 西安石油大学, 2020.
JIANG H Y. Study on rock mechanical characteristics and volume fracturing "sweet spot"of tight sandstone reservoir in area X[D]. Xi'an: Xi'an Shiyou University, 2020. (in Chinese with English abstract)

[29] 王艳梅. 岩石力学动、静态参数关系研究 [D]. 成都: 西南石油大学, 2012.
WANG Y M. Study on the relationship between dynamic and static parameters of rock mechanics[D]. Chengdu: Southwest Petroleum University, 2012. (in Chinese with English abstract)

[30] LAMA R D, VUTUKURI V S. Handbook on mechanical properties of rocks: Testing techniques and results[M]. Clausthal, Bay Village, Ohio : Trans. Tech. Publications, 1978.

[31] MORADIAN Z A, BEHNIA M. Predicting the uniaxial compressive strength and static Young's modulus of intact sedimentary rocks using the ultrasonic test[J]. International Journal of Geomechanics, 2009, 9(1): 14-19.

[32] NAJIBI A R, GHAFOORI M, LASHKARIPOUR G R, et al. Empirical relations between strength and static and dynamic elastic properties of Asmari and Sarvak limestones, two main oil reservoirs in Iran[J]. Journal of Petroleum Science and Engineering, 2015, 126: 78-82.

[33] YALE D P, JAMIESON W H. Static and dynamic rock mechanical properties in the Hugoton and Panoma fields, Kansas[C]//Anon. SPE Mid-Continent Gas Symposium. Richardson: Society of Petroleum Engineers, 1994: SPE 27939-MS.

[34] ABD EL-AAL A K, SALAH M K, KHALIFA M A. Acoustic and strength characterization of Upper Cretaceous dolostones from the Bahariya Oasis, western Desert, Egypt: The impact of porosity and diagenesis[J]. Journal of Petroleum Science and Engineering, 2020, 187: 106798.

[35] SHARIFI J, NOORAIEPOUR M, AMIRI M, et al. Developing a

relationship between static Young's modulus and seismic parameters[J]. Journal of Petroleum Exploration and Production Technology, 2023, 13(1): 203-218.

[36] NAJIBI A R, GHAFOORI M, LASHKARIPOUR G R, et al. Reservoir geomechanical modeling: In situ stress, pore pressure, and mud design[J]. Journal of Petroleum Science and Engineering, 2017, 151: 31-39.

[37] HARRYPERSAD-DANIEL A M, BLAKE O O, RAMSOOK R. Determining the static Young's modulus and Poisson's ratio, and compressive strength of the friable Erin Formation rocks using P-wave velocity[J]. Journal of Applied Geophysics, 2022, 198: 104557.

[38] QIN X, ZHANG Y H, WANG Y, et al. Experimental study on the static and dynamic elastic stiffness of rock samples from a shale oil reservoir[J]. International Journal of Rock Mechanics and Mining Sciences, 2023, 170: 105446.

[39] 伍顺伟, 高阳, 胡俊, 等. 玛湖地区横波预测方法优选及其对岩石力学参数计算的影响 [J]. 特种油气藏, 2021, 28(3): 9-14.
WU S W, GAO Y, HU J, et al. Prediction method optimization for S-wave in Mahu area and its influence on calculation of rock mechanics parameters[J]. Special Oil & Gas Reservoirs, 2021, 28(3): 9-14. (in Chinese with English abstract)

[40] 赵旭阳, 郭海敏, 李紫璇, 等. 基于测井横波预测的地应力场及岩石力学参数建模 [J]. 断块油气田, 2021, 28(2): 235-240.
ZHAO X Y, GUO H M, LI Z X, et al. Modeling of in situ stress field and rock mechanics parameters based on logging shear wave prediction[J]. Fault-Block Oil & Gas Field, 2021, 28(2): 235-240. (in Chinese with English abstract)

Summary for “基于 TPE 优化集成学习的岩石弹性模量预测模型”

Prediction model for rock elastic modulus based on TPE-optimized ensemble learning

MENG Xianglong[a,b*], WANG Shengjian[a,b], ZHU Disi[a,b], MA Yanyan[a,b], LI Dayong[a,b], CHI Huanpeng[a,b], ZHANG Jiazheng[a,b], YUE Weimin[a,b]

a. Oil & Gas Survey; b. Key Laboratory of Unconventional Oil and Gas Geology, China Geological Survey, Beijing 100083, China

* Corresponding author, E-mail: mengxianglong001@mail.cgs.gov.cn

Abstract: [**Objective**] Geophysical data is often used to determine the elastic modulus of formations in oil and gas engineering, with experimental data from small core samples used for calibration. However, acquiring core samples from every stratum is impractical and often leads to inadequate performance under complex geological settings. To improve the predictive accuracy and generalizability of the rock elastic modulus, an intelligent prediction model based on fundamental rock physical properties is proposed. [**Methods**] Using 397 sets of core experimental data from diverse sources, with compressional and shear wave velocities and density as input variables, intelligent prediction models for rock elastic modulus were developed based on three ensemble learning algorithms (Random Forest, XGBoost, LightGBM). The TPE method was employed to optimize the models. The dynamic and static elastic modulus regression models were constructed based on current methods used in petroleum engineering to provide a comprehensive assessment of the performance of the intelligent predictive model using statistical indicators. Additionally, the SHAP attribution analysis was utilized to assess the contribution of each input variable to the model. [**Results**] The research findings indicated that: ① The proposed intelligent prediction model using TPE was significantly better than traditional statistical regression models, achieving accurate predictions of the elastic modulus without distinguishing geological layers, with strong generalization ability. Among the three models, the XGBoost model performed the best (R^2=0.87, *RMSE*=6.94, *MAE*=4.96). ②Shear wave velocity made the greatest contribution to the model, followed by compressional wave velocity, with density having the least impact. Accurate shear wave velocity was crucial for predicting the elastic modulus. [**Conclusion**] This method allows for the precise prediction of elastic modulus without the need for prior identification of the work area and strata, providing valuable insights for the design and implementation of oil and gas engineering projects.

Key words: elastic modulus; TPE; ensemble learning; SHAP; shear wave

Received: 2024-06-13; Revised: 2024-09-23; Accepted: 2024-09-24

doi: 10.19509/j.cnki.dzkq.tb20240325