

基于GBDT算法的页岩储层脆性地震预测方法

——以渝西足201井区龙马溪组页岩储层为例

李若霖, 冯加阳, 王仲旭, 李金龙

重庆科技大学石油与天然气工程学院, 重庆

收稿日期: 2025年9月30日; 录用日期: 2025年11月5日; 发布日期: 2025年11月17日

摘要

随着页岩气勘探开发的深入, 页岩气储层的脆性评价显得越来越重要, 脆性指数评估直接影响水平井钻探的效率和安全性, 可靠的页岩脆性预测方法对提高页岩气开发效益具有着重要的意义。本研究以渝西足201井区龙马溪组页岩为例, 首先基于Zoeppritz方程进行叠前地震反演获得纵、横波速度以及密度从而得到所需要的弹性参数地震反演体; 利用岩石弹性参数实验数据、阵列声波测井数据及全岩X-射线衍射实验数据, 经过校正及解释获得连续的矿物脆性指数数据及对应的岩石弹性参数数据作为机器学习算法的样本集; 训练高效准确的学习方法和GBDT梯度决策树预测模型; 采用叠前地震反演岩石弹性参数体及GBDT梯度决策树学习算法, 实现研究区页岩储层脆性的智能化预测。

关键词

页岩气, 脆性, 梯度决策树, 地震预测, 叠前反演

Brittle Seismic Prediction Method of Shale Reservoir Based on GBDT Algorithm

—Taking the Shale Reservoir of Longmaxi Formation in the 201 Well Area of Western Chongqing as an Example

Ruolin Li, Jiayang Feng, Zhongxu Wang, Jinlong Li

School of Petroleum Engineering, Chongqing University of Science and Technology, Chongqing

Received: September 30, 2025; accepted: November 5, 2025; published: November 17, 2025

Abstract

With the deepening of shale gas exploration and development, the brittleness evaluation of shale

文章引用: 李若霖, 冯加阳, 王仲旭, 李金龙. 基于 GBDT 算法的页岩储层脆性地震预测方法[J]. 矿山工程, 2025, 13(6): 1368-1377. DOI: 10.12677/me.2025.136152

gas reservoirs is becoming more and more important. The brittleness index evaluation directly affects the efficiency and safety of horizontal well drilling. Reliable shale brittleness prediction methods are of great significance to improve the efficiency of shale gas development. In this study, the shale of the Longmaxi Formation in the 201 well area of western Chongqing is taken as an example. Firstly, pre-stack seismic inversion is performed based on the Zoeppritz equation to obtain the P-wave, S-wave velocity and density to obtain the required elastic parameter seismic inversion body. Using the experimental data of rock elastic parameters, array acoustic logging data and whole rock X-ray diffraction experimental data, continuous mineral brittleness index data and corresponding rock elastic parameter data are obtained after correction and interpretation as the sample set of machine learning algorithm. Train efficient and accurate learning methods and GBDT gradient decision tree prediction models; the pre-stack seismic inversion of rock elastic parameters and GBDT gradient decision tree learning algorithm are used to realize the intelligent prediction of shale reservoir brittleness in the study area.

Keywords

Shale Gas, Brittleness, Gradient Decision Tree, Earthquake Prediction, Prestack Inversion

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

通过压裂造缝以连通天然微裂缝、水平层理缝或基质孔隙是页岩气高效开发最有效的措施，而影响压裂的重要因素是，页岩层的脆弱性是否能承受人工引发的多数裂缝[1]-[4]。研究表明，形成的脆性高的页岩地层是人造网状缝发育不可或缺的组成部分[5][6]。脆性越高的页岩，造缝能力就越强，越易形成网缝。而且，页岩内游离气体的产生，大多受成岩作用和构造运动产生裂缝的影响。因此，评估页岩的脆性对于确定页岩气的优选区域非常重要，北美 Barnett、Haynesville 页岩气田和国内涪陵页岩气田等的高产与页岩脆性均有十分显著的关系[7]。多年来，国内外学者对页岩脆性地震预测问题进行了很多理论研究和方法分析。但是，脆性指数预测方法主要通过矿物组分法来实现对页岩脆性的评价，但是很难实现矿物组分法地震数据的脆性预测。传统的脆性指数地震预测方法在叠前地震反演得到岩石弹性参数后，一般采用较简单的瑞克曼或者线性数学关系预测脆性指数，预测结果精度有限[8]-[10]。

目前，有的研究者针对单纯考虑脆性矿物或岩石弹性参数进行脆性预测中存在的不足，提出将脆性矿物组分和岩石弹性参数相结合进行页岩脆性预测，但主要采用传统的地质统计分析预测方法，即利用有代表性的样本数据进行回归分析，确定出页岩脆性预测模型，并以此来预测研究区的储层脆性分布情况。但是，由于回归分析受到模型选择、校验水平确定、方程组的病态、原始数据异常值等因素影响，使回归模型常常不切实际，预测误差偏大，达不到预期效果。

人工智能(AI)已经成为现代科技的重要组成部分，推动了许多领域的创新与进步。在人工智能的诸多子领域中，机器学习(ML)无疑是最关键和最具影响力的一个。机器学习通过自动分析和学习数据中的模式，赋予计算机以改进性能和做出预测的能力。机器学习在分析样本数据时具有高效率的决策能力，与传统建模技术相比，机器学习模型具有较高的灵活性和精度。其中 GBDT 梯度提升决策树算法通过多轮迭代，每轮迭代产生一个弱分类器，每个分类器在上一轮分类器的残差基础上进行训练，从而逐步降低模型的误差，预测的准确性较高。GBDT 算法在各种实际应用中都得到较好的应用效果，能够提供高精度的预测结果。

本项目在前人研究的基础上,从岩石矿物脆性指数和弹性参数入手,采用梯度决策树预测方法,建立页岩脆性预测的梯度决策树模型。通过叠前地震反演获得岩石弹性参数预测数据体,利用 GBDT 梯度决策树模型获得基于地震反演的页岩储层脆性的预测结果,实现基于机器学习法的全区储层脆性的地震预测,提供一种页岩储层脆性预测的新方法。

2. 原理

本研究以地震叠前同时反演预测的杨氏模量、泊松比及剪切模量等岩石弹性参数预测结果为基础,采用 GBDT 神经网络算法建立的岩石弹性参数预测岩石脆性的网络模型,实现页岩储层脆性指数的预测方法,其中主要涉及叠前同时反演及 GBDT 神经网络两大核心理论。

2.1. 叠前同时反演预测岩石弹性参数基本原理

叠前同时反演是一种基于地震叠前数据(如角道集或偏移距道集)同步预测多种岩石弹性参数(如纵波速度 V_p 、横波速度 V_s 、密度 ρ 等)的高分辨率反演方法。叠前反演的核心依据是地震反射振幅随入射角(或偏移距)的变化规律(AVO/Zoeppritz 方程)。不同弹性参数组合会导致不同的 AVO 响应特征,通过拟合实际观测的 AVO 曲线,可反推地下介质的弹性参数。由于 Zoeppritz 方程复杂,实际应用中常采用其线性或高阶近似式,例如公式(1)所示的 Aki-Richards 近似表达式。方法流程包括以下几个部分:地震数据优化处理、测井曲线的标准化、地震子波提取及合成地震记录制作;以地震解释层位为约束结合测井曲线建立低频模型;将地震数据体、测井曲线以及层位作为数据输入,最终得到反演数据体。

$$R_{pp}(\theta) \approx a \frac{\Delta V_p}{\bar{V}_p} + b \frac{\Delta V_s}{\bar{V}_s} + c \frac{\Delta \rho}{\bar{\rho}} \quad (1)$$

2.2. GBDT 算法基本原理

GBDT (Gradient Boosting Decision Tree, 梯度提升决策树)是一种基于 Boosting 集成学习的机器学习算法,通过迭代训练多个弱学习器(决策树),逐步优化模型的预测能力。其核心思想是利用梯度下降法最小化损失函数,每一轮迭代中,新训练的树会拟合当前模型的残差(负梯度方向),最终将所有树的结果加权求和得到预测值。GBDT 的核心原理为: (1) Boosting (提升): 通过串行训练多个弱模型(如浅层决策树),每个模型专注于修正前一个模型的错误。具有模型之间强依赖(前一棵树的输出影响后一棵树的训练)、最终预测是所有弱模型的加权累加的特点。(2) 梯度下降(Gradient Descent): 通过每一轮计算损失函数的负梯度(即伪残差),用新的决策树拟合这些梯度,逐步逼近最优解得到最小化损失函数(如均方误差、交叉熵)。(3) 决策树(弱学习器): GBDT 通常使用 CART 回归树(即使分类问题也是回归树拟合概率)、树的深度较浅(如 3~6 层),避免过拟合(图 1)。

3. 材料与实验方法

以地震资料、测井资料以及地质资料为基础,在分析研究区沉积构造环境的基础上,采用全岩 X-射线衍射实验数据得到样本点页岩矿物组分计算矿物脆性指数;利用交叉偶极阵列声波测井资料,结合三轴岩石力学试验确定样本点的静态弹性参数;基于页岩矿物脆性指数和弹性参数作为梯度决策树模型的基本特征量分析储层脆性与这些因素之间的相关关系,建立梯度决策树预测模型并进行模型训练效果评估;针对三维地震资料通过叠前地震反演方法获取所需的弹性参数,利用梯度决策树模型对弹性参数进行换算处理,对研究区储层进行脆性预测(图 2)。

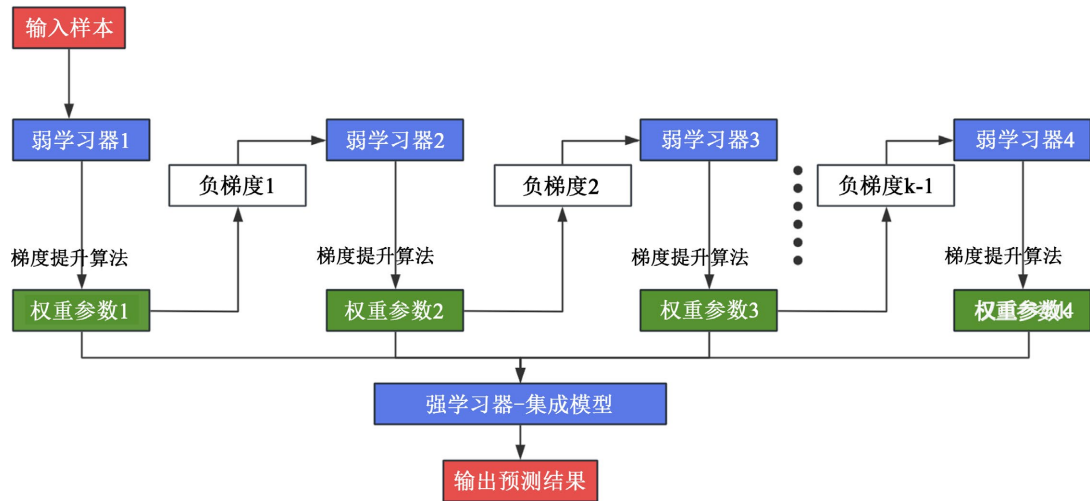


Figure 1. Flow chart of GBDT algorithm

图 1. GBDT 算法流程图

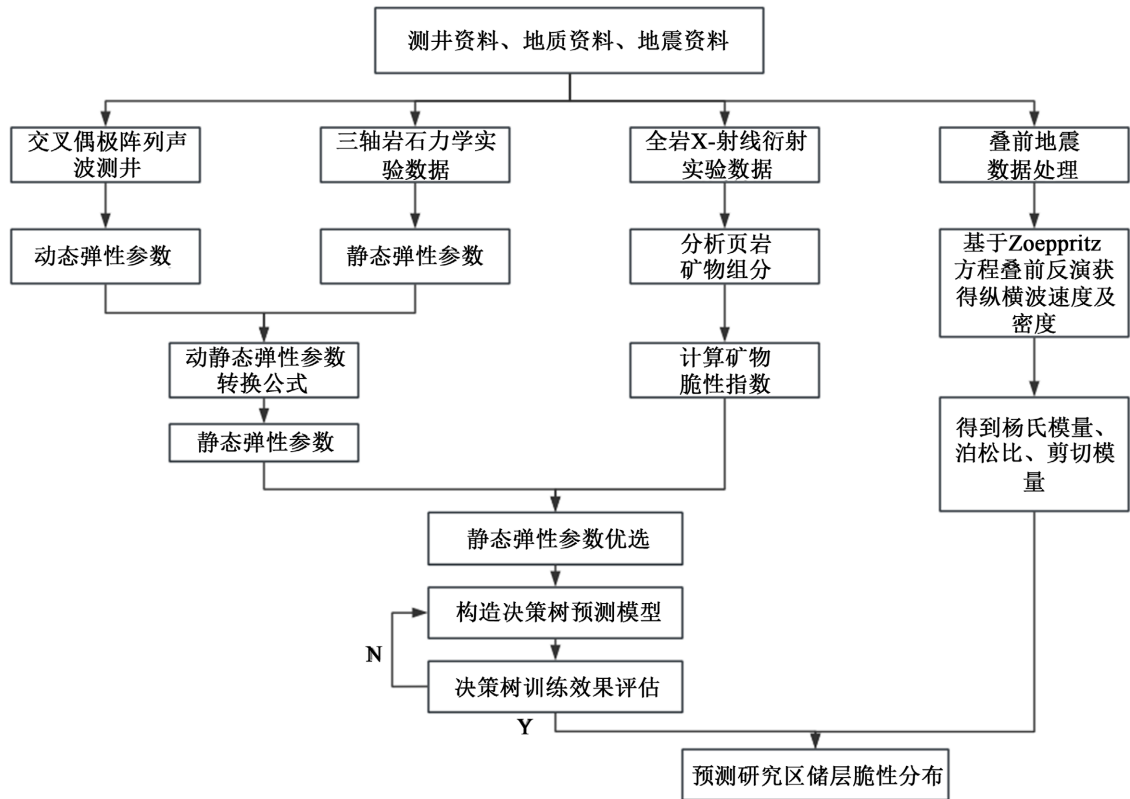


Figure 2. This paper studies the flow chart

图 2. 本文研究流程图

3.1. GBDT 神经网络模型构建

3.1.1. 数据预处理

(1) 岩石弹性参数数据

通过研究区全岩 X-射线衍射实验分析脆性矿物组分, 计算出研究区矿物脆性指数。根据研究区三轴

岩石力学试验得到的静态弹性参数数值范围,利用交叉偶极声波测井资料提供的纵波时差(DTP)、横波时差(DTS)和体积密度(DEN)计算动态杨氏模量(E_d _dynamic)和动态泊松比(ν_d _dynamic),

$$E_d = \frac{\rho V_s^2 (3V_p^2 - 4V_s^2)}{V_p^2 - V_s^2} \times K \quad (K \text{ 为单位转换系数}) \quad (2)$$

$$\nu_d = \frac{(V_p/V_s)^2 - 2}{2((V_p/V_s)^2 - 1)} \quad (3)$$

利用动静弹性参数转换公式,将大量、连续的动态弹性参数转换为实验室静态弹性参数数值。进而实现动态参数的静态化校正,为GBDT模型训练提供充足的数据点。

(2) 脆性指数数据

基于常规测井数据与X衍射实验矿物组分数据建立测井解释模版可获得连续的矿物曲线,再通过公式(3)对矿物曲线进行计算,即可得到连续的矿物脆性指数数据点。

$$B_l = \frac{w_{QFM} + w_{CARM}}{w_{QFM} + w_{CARM} + w_{CM} + w_{PY}}, \quad (4)$$

其中, B_l 为矿物脆性指数, w_{QFM} 为页岩中石英和长石含量,%; w_{CARM} 为页岩中方解石和白云石含量,%; w_{CM} 为页岩中黏土矿物含量,%; w_{PY} 为页岩中黄铁矿含量,%。

3.1.2. 模型结构

GBDT的结构核心可以概括为:用决策树作为基学习器的梯度提升集成模型。虽然它常被称为“模型”,但将其理解为一个构建模型的框架或算法更为准确。其整体结构可以看作是一个线性叠加的模型,其预测结果是所有基学习器(树)预测结果的加权和:由一系列弱学习器(决策树)按顺序组合而成。

$$F_M(x) = F_0(x) + \eta \sum_{m=1}^M f_m(x) \quad (5)$$

其中, $F_M(x)$ 为经过 M 轮迭代后的最终强学习器(最终模型), $F_0(x)$ 为初始化的模型,通常是一个常数(例如,对于回归问题,是所有样本标签的均值), M 为基学习器的数量(树的棵数), $f_m(x)$ 为第 m 棵决策树(基学习器)。

3.1.3. 模型训练与评估标准

通过研究区中实际测井资料以及地震、地质资料得到矿物脆性指数、及对应的岩石弹性参数数据。随机选择4口井中的连续数据剔除异常值,其中70%数据进行模型的训练。假设损失函数为 $L(y, F(x))$, 则构建一个由 M 棵树组成的加法模型 $F_M(x)$ 来预测 y , 首先初始化模型,用一个常数值初始化模型,通常是使损失函数最小的常数值:

$$F_0(x) = \arg \min_r \sum_{i=1}^n L(y_i, r) \quad (6)$$

其次进行迭代训练(对于 $m = 1$ 到 M): a 计算伪残差; b 拟合一颗新树; c 为每个叶子节点计算最佳值; d 更新模型,并调整树结构、损失函数、子采样以在偏差和方差之间取得最佳平衡后得到最终模型。

评估是衡量模型性能的关键,本研究为预测连续值的回归任务类型,则选取最常用、可导但对异常值敏感的均方误差(MSE)及表示模型可解释的方差比例决定系数(R^2)作为评估指标评估模型较为合适。

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

$$R^2 = \frac{\left[\sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})(y_i - \bar{y}) \right]^2}{\sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2 \sum_{i=1}^n (y_i - \bar{y})^2}, \quad (8)$$

其中, n 为样本数, $\hat{y}_i$ 为第 i 个样本的预测值, y_i 为第 i 个样本的真实值, $\bar{\hat{y}}$ 为样本预测值中的平均值, $\bar{y}$ 为样本真实值的平均值。

通过标准的训练与评估流程, 可以系统地构建和评估高性能、高泛化能力的 GBDT 模型。

3.1.4. 效果验证

使用剩下的 30% 数据的进行模型效果的评估与验证, 测试不同学习率和树数量组合的性能, 找到最佳的学习率与树数量的平衡点并分析该井预测值与真实值之间的相关性, 采用均方误差(MSE)与决定系数(R^2)检验模型预测效果。

3.2. 叠前地震弹性参数反演

3.2.1. 优化叠前地震数据并进行井震精细标定

首先将地震数据进行预处理、保幅去噪和精细道集拉平, 其次对处理后的角道集应用超道集处理优化数据, 生成一个高信噪比、保留全角度信息的超道集数据体。最后进行部分角度叠加减少数据量, 提高反演计算效率。为叠前反演提供高信噪比、高保真度、振幅相对可靠且道集拉平的高质量角道集数据。

对测井数据中的曲线进行环境矫正、编辑和平滑, 去除异常值, 并提取井旁地震道、建立深度域的测井数据与时间域的地震数据之间可靠的时深关系, 提取合适的地震子波最大化合成记录与地震道的相关系数, 继而进行精细化标定。建立井与地震之间的准确对应关系。利用标定结果指导进一步的地震数据处理, 再次精细标定确保匹配度最高。最后将得到的子波和低频模型用于叠前同时反演。

3.2.2. 弹性参数体计算

基于已有的井位、层位及地震子波数据, 构建包含纵波速度、横波速度和密度信息的初始模型, 并结合实际地震资料实施叠前反演, 最终获取对应的纵波速度、横波速度及密度反演数据体。在此基础上, 应用岩石力学关系计算得到敏感性优良的弹性参数数据体。

3.3. 基于 GBDT 训练模型的页岩储层脆性指数地震预测

采用地震数据反演得到的储层脆性指数敏感弹性参数数据体作为神经网络输入数据, 利用训练成功并验证合格的神经网络系统预测页岩储层脆性指数数据体, 进而得到研究区基于 GBDT 神经网络的脆性指数地震预测结果。

将储层脆性指数敏感弹性参数数据体作为训练模型输入并验证, 利用模型预测页岩储层脆性指数三维数据, 最终形成研究区基于 GBDT 训练模型的脆性指数地震预测成果。

4. 结果

4.1. GBDT 训练模型构建

4.1.1. 数据预处理结果

本次采用研究区同时可具备岩石力学参数及脆性指数数据的 201、203H2-1、203 及 206 四口井龙马溪组一亚段数据点作为模型基础数据, 采用静态化校正公式校正连续的动态岩石参数, 以及由矿物法获得的 4 口井目的层脆性指数数据, 具备 1592 个连续的脆性指数数据点。

4.1.2. 模型结构的确定

首先明确模型结构调优的关键参数及其候选范围, 包括学习率(learning_rate)、树最大深度(max_depth)、叶节点最少样本数(min_samples_leaf)、内部节点分裂所需最少样本数(min_samples_split)以及弱学习器数量(n_estimators), 选取训练数据与测试数据输入模型, 通过调整不同参数对模型结构的影响进行确定模型结构。

通过对比使用不同参数时模型性能的表现, 训练数据与测试数据的决定系数随着树数量的增多而提升(图 3); 当学习率选取值为 0.05~0.1 时, 树数量对负均方误差的影响程度很小并且模型负均方误差绝对值较小(图 4)。最终选取最佳参数: Best parameters found: {'learning_rate': 0.1, 'max_depth': 5, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 200}。

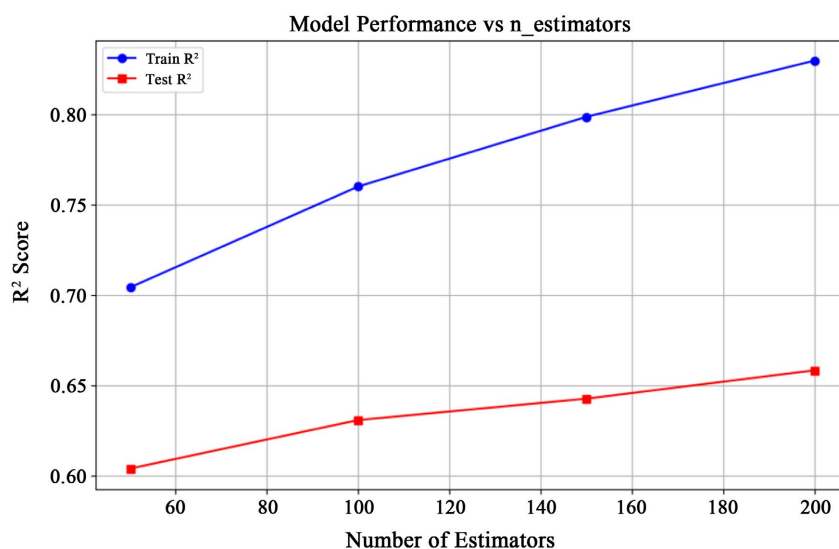


Figure 3. Relationship between model performance and tree number parameters (n_estimators)

图 3. 模型表现与树数量参数(n_estimators)的关系

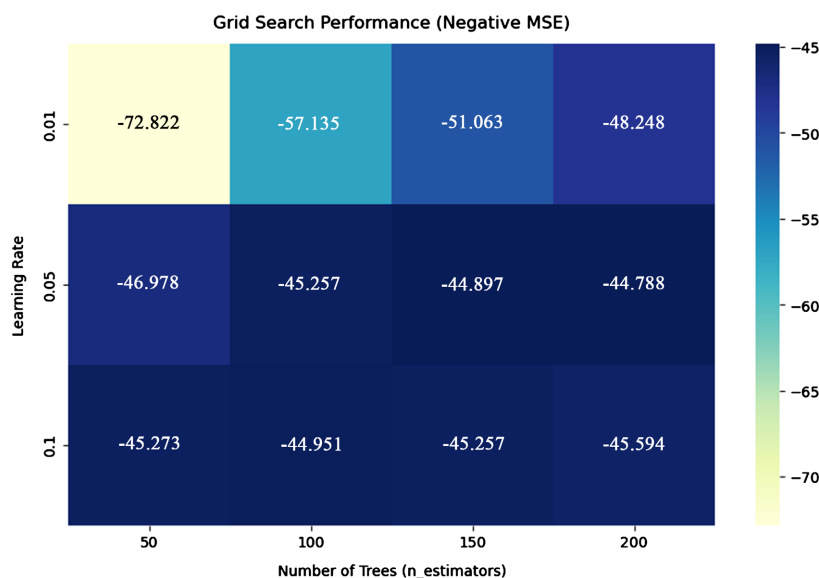


Figure 4. Combination performance of different learning rates and tree numbers

图 4. 不同学习率和树数量组合性能图

4.1.3. 模型训练与测试

选取研究区中 203H2-1 井、203 井、206 及 201 井 4 口井的杨氏模量(E)、泊松比(σ)数据作为模型输入特征变量,以对应脆性指数数据作为输出特征变量。4 口井共具备 1592 组有效数据点,选取 203H2-1、203、206 三口井用于训练与测试,保留 201 井用于验证模型效果。系统训练数据预测结果与实际脆性数据相关性分析表明(图 5),二者相关决定系数为 0.895,训练效果理想;系统测试数据预测结果与实际数据相关性分析表明,二者相关决定系数为 0.793,测试效果较好。

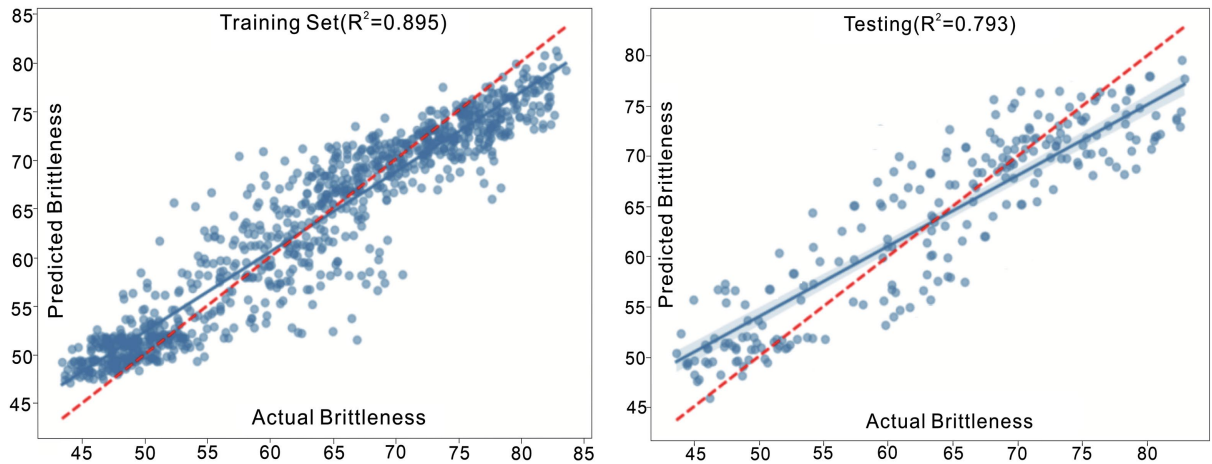


Figure 5. Correlation analysis diagram between model training and test data. (a) The correlation analysis diagram of training data prediction results and actual data; (b) The correlation analysis diagram between the test data prediction results and the actual data

图 5. 模型训练与测试数据相关性分析图。(a) 训练数据预测结果与实际数据相关性分析图; (b) 测试数据预测结果与实际数据相关性分析图

4.1.4. 模型效果验证

采用未参与模型训练与测试的 201 井 291 组数据点中的杨氏模量、泊松比参数作为输入,对比分析预测模型输出的脆性指数与对应实际脆性指数。通过残差(图 6)分析评估和诊断模型适用性,检测数据中的异常值以及了解模型的预测能力,可以从图中看出数据没有异常点,模型的拟合充分,几乎不存在系统系偏差。

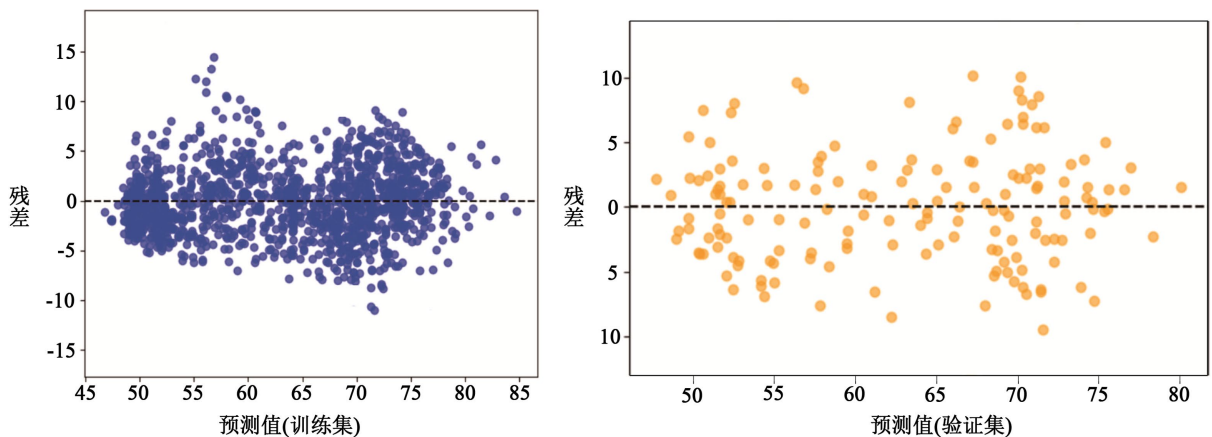


Figure 6. Residual analysis diagram of training set and validation set

图 6. 训练集、验证集残差分析图

4.2. 弹性参数反演体计算

针对原始地震道集超道集处理,提高原始数据信噪比(图 7),再对超道集数据进行角度叠加处理为角道集数据;道集质量得到明显改善,信噪比明显提高,获得高品质的叠前地震数据。在层位数据、井数据及地质模式约束下完成纵波速度、横波速度和密度的联合反演,进而得到杨氏模量、泊松比弹性参数数据体。

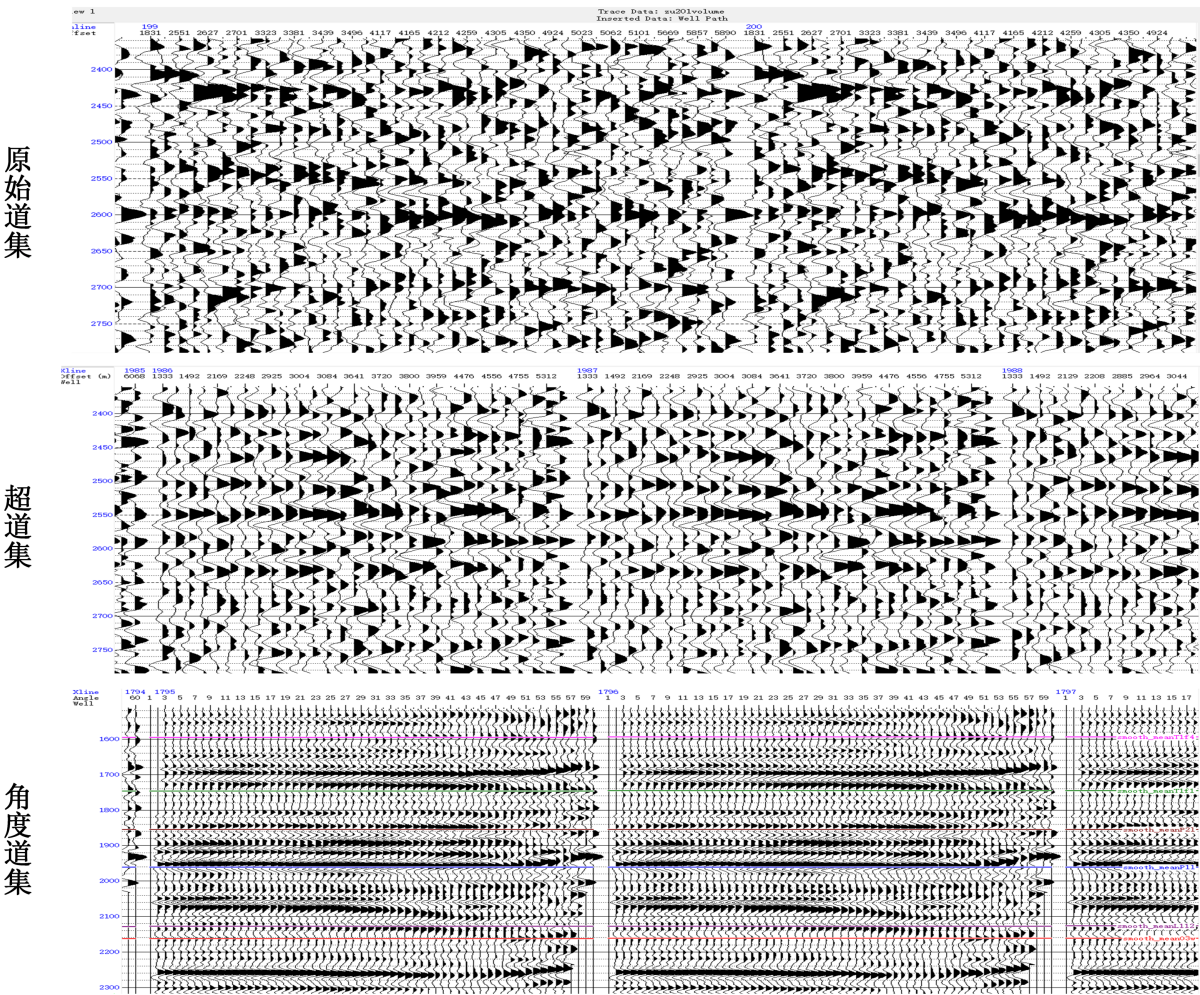


Figure 7. Signal-to-noise ratio of seismic data
图 7. 地震数据信噪比

4.3. 脆性指数地震预测

将叠前反演的杨氏模量、泊松比预测数据体,输入到构建好的 GBDT 模型中,即可得到脆性指数预测结果,预测结果平面分布图如图 8 所示。分析 GBDT 神经网络法脆性指数预测平面结果图可以看出平均脆性指数在 60%~70%之间,脆性指数平面规律呈现由西北至东南逐渐变小。

5. 结论

本次应用于页岩脆性指数地震预测方法的 GBDT 预测模型,优选了杨氏模量、泊松比弹性参数作为输入样本数据。采用 3 口井的 1592 组数据对模型进行训练,1 口井的 291 组数据进行模型效果验证。

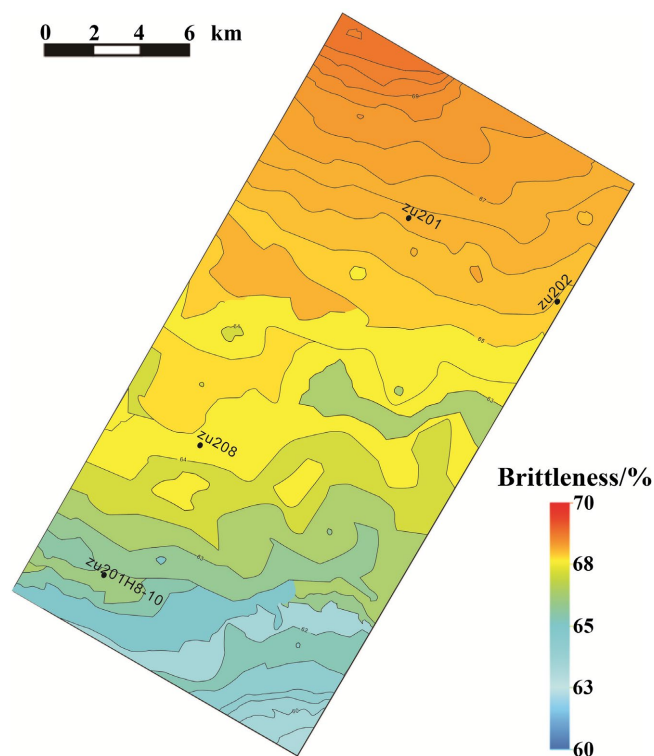


Figure 8. The plane distribution of brittleness index predicted by GBDT model
图 8. GBDT 模型预测脆性指数结果平面分布图

预测模型训练及验证结果表明,本次建立的 GBDT 神经网络预测模型具有较高的可靠性,脆性指数预测值符合率为 91.79%,该方法显著提高了页岩储层脆性地震预测方法的预测精度。

基金项目

重庆科技大学科技创新基金项目“基于 GBDT 算法的页岩储层脆性地震预测方法——以渝西足 201 井区龙马溪组页岩储层为例”(立项编号: YKJCX2420112)资助。

参考文献

- [1] 涂良程. 丁山地区龙马溪组页岩储层脆性分析[D]: [硕士学位论文]. 青岛: 中国石油大学(华东), 2020.
- [2] 刘震, 张军华, 于正军, 等. 非常规储层脆性研究进展及展望[J]. 石油地球物理勘探, 2023, 58(6): 1499-1507.
- [3] 孙彪, 刘小平, 舒红林, 等. 湖相泥页岩储层脆性评价及影响因素分析——以苏北盆地海安凹陷曲塘次凹泥页岩为例[J]. 石油实验地质, 2021, 43(6): 1006-1014.
- [4] 杜睿山, 李宏杰, 孟令东, 等. 基于 PSO-BiLSTM 的储层岩石脆性指数预测[J]. 海南大学学报(自然科学版), 2023, 41(3): 260-267.
- [5] 于庭, 巴晶, 钱卫, 等. 非常规油气储层脆性评价方法研究进展[J]. 地球物理学进展, 2019, 34(1): 236-243.
- [6] 曹东升, 曾联波, 吕文雅, 等. 非常规油气储层脆性评价与预测方法研究进展[J]. 石油科学通报, 2021, 6(1): 31-45.
- [7] 陈国庆, 吴家尘, 蒋万增, 等. 基于弹性能演化全过程的岩石脆性评价方法[J]. 岩石力学与工程学报, 2020, 39(5): 901-911.
- [8] 温韬, 张馨, 孙金山, 等. 基于峰前和峰后能量演化特征的岩石脆性评价[J]. 地球科学, 2021, 46(9): 3385-3396.
- [9] 张超, 白允, 安永林, 等. 基于全应力应变曲线的岩石脆性特征评价新方法[J]. 公路交通科技, 2021, 38(3): 63-72.
- [10] 姜宇. 基于深度学习的致密砂岩储层预测方法研究及应用[D]: [硕士学位论文]. 北京: 中国石油大学(北京), 2023.