第 27 卷 第 3 期
2025 年 6 月

古 地 理 学 报
JOURNAL OF PALAEOGEOGRAPHY (Chinese Edition)

Vol. 27 No. 3
Jun. 2025

文章编号：1671-1505(2025)03-0763-14 DOI:10.7605/gdlxb.2025.00.026

基于 XGBoost 算法的页岩岩相测井预测方法*

闫佳飞[1] 李胜利[1] 魏泽德[1] 吴忠宝[2] 陈建阳[2]

1 中国地质大学（北京）能源学院，海相储层演化与油气富集机理教育部重点实验室，北京 100083
2 中国石油勘探开发研究院，北京 100083

摘 要 页岩岩相的识别与预测对于分析确定页岩油气甜点层段非常重要。在缺乏岩心信息进行单井岩相研究时，测井数据扮演着十分重要的角色，而基于 XGBoost 算法可以充分挖掘多维测井数据所揭示的页岩岩相信息，从而达到预测单井页岩岩相的目的。本研究应用具有监督学习算法的 XGBoost 机器学习方法，利用常规测井数据作为变量数据集，建立了可预测页岩岩相类型的计算模型。首先建立适合具体研究区的页岩岩相划分标准，该标准应能体现研究区页岩岩相的辨识差异性，再统计不同矿物含量，确定不同岩相的具体矿物含量和 TOC 含量界限。在建立计算模型时，相关变量可能会提供相似的信息，导致模型过于依赖这些特征，需注意去除相似信息。XGBoost 算法在参数优选方面，其网格搜索具有全面性，在网格搜索过程中应该进行多次优选，不断缩小搜索范围以求取最优值。以松辽盆地松南地区赞字井区块为例，采用矿物组分含量、沉积构造及 TOC 含量建立页岩岩相划分标准，青山口组可划分出 5 类主要页岩；在应用 XGBoost 算法进行变量优选时，对于具有较高相关性的深侧向电阻率（LLD）和浅侧向电阻率（LLS）曲线，保留一条即可，结果表明模型准确率可提高 4%左右；经过变量选择及参数调优后，最终模型预测岩相的准确率可达 90.03%。

关键词 页岩岩相预测 XGBoost 算法 变量选择 参数调优 测井信息 青山口组 松辽盆地

第一作者简介 闫佳飞，男，1999 年生，中国地质大学（北京）在读硕士研究生，地质资源与地质工程专业。E-mail：1275585596@qq.com。

通讯作者简介 李胜利，男，1971 年生，博士，中国地质大学（北京）能源学院教授，博士生导师，研究方向为沉积储层和开发地质。E-mail：slli@cugb.edu.cn。

中图分类号：P512.2 **文献标志码**：A

Shale lithofacies prediction method with well-logging data based on XGBoost algorithm

YAN Jiafei[1] LI Shengli[1] WEI Zede[1] WU Zhongbao[2] CHEN Jianyang[2]

1 *Key Laboratory of Marine Reservoir Evolution and Hydrocarbon Enrichment Mechanism (Ministry of Education), School of Energy Resources, China University of Geosciences (Beijing), Beijing* 100083, *China*
2 *Research Institute of Petroleum Exploration and Development, PetroChina, Beijing* 100083, *China*

Abstract The identification and prediction of shale lithofacies are crucial for identifying favorable intervals ("sweet spots") in shale oil and gas reservoirs. In the absence of core data, logging data plays a key role in lithofacies analysis at the single-well level. By applying the XGBoost algorithm, useful lithofacies information can be extracted from multidimensional logging data, enabling effective prediction of

* 国家自然科学基金项目(编号：42172112)资助。[Financially supported by the National Natural Science Foundation of China(No. 42172112)]
收稿日期：2024-05-04 改回日期：2024-11-27

shale lithofacies in individual wells. In this study, the XGBoost machine learning method, a supervised learning algorithm, is used to build a predictive model based on conventional logging datasets. First, a lithofacies classification scheme tailored to the specific study area is established, which captures the variability in shale lithofacies identification. The boundaries of mineral compositions and TOC content for different lithofacies types are then determined using statistical proportion analysis. During model construction, care must be taken to eliminate redundant variables, as highly correlated features may provide overlapping information and cause overfitting. XGBoost's grid search approach allows comprehensive parameter tuning. Multiple rounds of optimization should be conducted, with the search range gradually narrowed to determine the optimal parameter set. Using the Zanzijing block in the Songnan area as a case study, five major shale lithofacies types are defined based on mineral composition, sedimentary structures, and TOC content. During variable selection, for instance, only one of the highly correlated LLD and LLS logs is retained, which results in a model accuracy improvement of approximately 4%. After feature selection and parameter tuning, the final model achieves a lithofacies prediction accuracy of up to 90.03%.

Key words shale lithofacies prediction, XGboost algorithm, variable selection, parameter tuning, well-logging data, Qingshankou Formation, Songliao Basin

About the first author YAN Jiafei, born in 1999, is a graduate student of China University of Geosciences (Beijing), and he specializes in geological resources and geological engineering. E-mail: 1275585596 @qq.com.

About the corresponding author LI Shengli, born in 1971, Ph.D., is a professor and doctoral supervisor at School of Energy, China University of Geosciences (Beijing), and he is mainly engaged in sedimentary reservoir and development geology. E-mail: slli@cugb.edu.cn.

0 引言

岩相可以指示沉积岩的形成环境及其特征，可以反映沉积物的来源、沉积方式、沉积环境、结构、构造、颜色等信息（陈登辉等，2019）。在页岩的岩相划分方面国内外已进行了大量研究，提出了多种划分方案，比如基于“岩石组分—沉积构造—有机质”的特征进行页岩岩相的划分（彭丽等，2019）；基于“矿物组分—纹层发育—岩性—电性”的特征进行页岩岩相的划分（赵贤正等，2019）；基于“岩性—矿物—纹层密度—沉积构造—有机质丰度”5 种参数的岩相划分方案（付秀丽等，2022）。其中最关键的就是按照构成岩石组分的黏土矿物、硅质矿物（石英+长石）以及碳酸盐矿物三者相对含量的差异来进行岩相划分（毛玉丹，2023）。通过以上方法的确能有效地识别出岩相类型，但是对于缺少矿物含量、岩石薄片、成像测井及 TOC 含量数据的钻井，则需要建立岩相与测井信息之间的联系，从而实现各个单井的页岩岩相划分。薛纯琦等（2021）在利用测井曲线识别岩相方面开展了大量工作，其中机器学习法尤为突出。由于页岩具有较强的非均质性（张益粼等，2023），通常需要大量的样品进行各项实验分析，在划分页岩岩相类型时往往出现岩石样品数量较少的情况，岩相类型的偶然性会较大，因此将机器学习算法当作回归器是一种可行的页岩岩相划分方法。谌丽等（2023）通过机器学习中集成学习 Bagging 算法组合多个基分类器，优化各类岩相的分类性能；也有学者利用集成算法，如随机森林算法（王民等，2023）和 XGBoost 算法（薛纯琦等，2021）可以很好地识别泥页岩岩相，都取得较高的模型准确率。

XGBoost 算法是在对梯度提升算法的大量研究基础上提出的一种基于提升树的机器学习系统（Chen and Guestrin，2016），是一种已经被证实的表现力很好的有监督机器学习算法。该算法一般不适合过于稀疏和单一维度的数据，因其可能导致模型难以学习到有效信息（Zhu *et al.*，2021）；但对于多维数据，该算法具有准确性高、灵活性强、可扩展性好、计算速度快等特点而被广泛应用（Zheng *et al.*，2022）。XGBoost 算法适合处理复杂

的非线性问题，尤其是适合处理具有多个特征的数据。同时，对含有噪声的数据具有一定的鲁棒性（李占山和刘兆赓，2019）。所以，多种（条）测井曲线可以充分发挥 XGBoost 算法的优点，充分捕捉测井曲线有效信息并与页岩岩相建立联系，从而达到较为准确识别与预测页岩岩相的目的。其中在建立模型过程中，特征选择及参数调优是提高模型准确性的有效方法。

本研究首先结合松辽盆地松南地区赞字井区页岩中的矿物含量，建立页岩岩相划分的标准；然后利用测井曲线作为模型特征，建立与岩相对应的机器学习算法模型，从而实现研究区单井岩相预测。松辽盆地青一段沉积时期发生了大规模的水进，形成了湖盆广布、水深可达半深湖—深湖的大型坳陷湖盆，沉积了厚层富有机质的暗色泥岩和页岩。青一段既是该盆地的主力烃源岩层系，也是页岩油主要发育层段，松南地区赞字井区则位于前三角洲相带（柳波等，2021），其中的泥页岩夹杂着粉砂岩。由于分布面积较小，侧向变化快，成分多样，结构各异，具有一定的成因多解性。并且该区域单井的 TOC 含量、成像测井、岩心等资料不全，对赞字井区的页岩岩相的单井划分带来了困难。作者利用机器学习模型来建立常规测井与真实岩相的联系来进行该区域的岩相预测，为后续非常规油气地质“甜点”评价（赖锦等，2023）提供参考。在建立岩相划分方案时，不是采用传统的黏土矿物、长英质矿物及碳酸盐矿物 50% 为界来划分，而是根据研究区不同矿物含量所反映的岩性差异性，确定相对合理的矿物含量界限，从而建立相应的岩相划分方案。文中的岩相划分方案采用矿物含量与沉积构造为主、有机质丰度为辅的页岩岩相划分方案。

1　页岩岩相的划分方案

页岩的非均质性较强，时空变化较大、有机质分布的非均质性较强、纹层结构及其组合类型差异明显，因此，盆地内富有机质页岩的分布预测以及页岩油气“甜点”段/区的评价需要明晰页岩岩相的变化特征（张晋言，2013）。利用矿物组分含量和 TOC 含量数据，通过成像测井拾取纹层等沉积构造，可实现岩相的测井判别（赖锦等，2023）。矿物含量可以直观地反映出岩石矿物组成成分及比例，在页岩岩相命名方面具有重要作用（沈骋等，2021），因此文中的页岩岩相命名采用将沉积构造和矿物类型相结合的方法。为了便于使用，既要解决名称过长的问题，对命名进行简化处理，也要保留传统名称及意义（Peng and Guo，2023）。

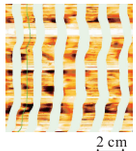
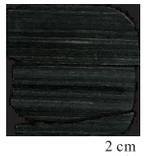
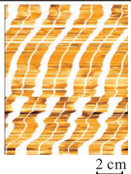
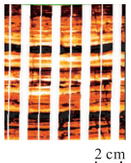
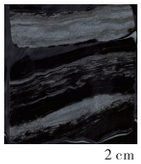
首先利用岩心薄片观察、成像测井、TOC 含量及矿物含量识别出真实岩相，再通过建立机器学习模型的方式与常规测井建立联系，用于识别该区域无岩心的其他单井中的岩相。文中的岩相划分在沉积构造方面采用层状与纹层状来进行区分（表 1）。其中，层状构造主要表现为夹杂于泥页岩段中的粉（细）砂岩，单层厚度大于 1 cm。各层之间有清晰的界限，层与层之间大致平行（表 1）。而页岩的纹层反映岩层内部的沉积构造特征（田瀚等，2023）。纹层也称为页岩层理，是指在沉积过程中，由于沉积环境的变化而形成的一种层状构造，其厚度通常小于 1 cm（Pang *et al.*，2024）。纹层在页岩中常呈现出薄而平行的层状结构（车世琦，2018；庞小娇等，2023）。纹层作为页岩独具特色的组构特征，它的发育直接导致页岩非均质性变化明显，进而对页岩生烃、储集性以及含气性等产生影响（徐传正等，2021）。在页岩储集层评价中，纹层的特征和发育程度也是重要的指标之一（Sun *et al.*，2022），纹层构造不仅可以反映页岩储集层的微观结构和储集性能，还会直接影响水平井体积压裂裂缝的扩展规律和压裂效果（何伟等，2021），所以该沉积构造在划分页岩岩相类型时尤为重要。

岩相类型划分的基本原则强调分类方案具有较强的可操作性和实用性，同时又能反映岩石中最重要的成因特征（彭军等，2022；李宁等，2023）。文中岩相划分基于“矿物组分—沉积构造—岩性—有机质丰度”的页岩岩相划分方案，操作步骤如下：在样品 X 射线衍射全岩矿物分析基础上，以碳酸盐矿物、长英质矿物、黏土矿物为三端元进行岩石类型划分，然后结合页岩中的纹层发育情况，以及 TOC 信息，进行岩相综合划分。其中纹层发育情况可以通过岩心与成像测井信息识别，而 TOC 数据可以通过测试分析获得。同时，在缺乏足够信息的情况下，过于复杂的分类不利于实际应用，因此文中的岩相划分方案不采用纹层密度作为划分依据。

首先，利用矿物组分含量划分出 3 大类页岩岩

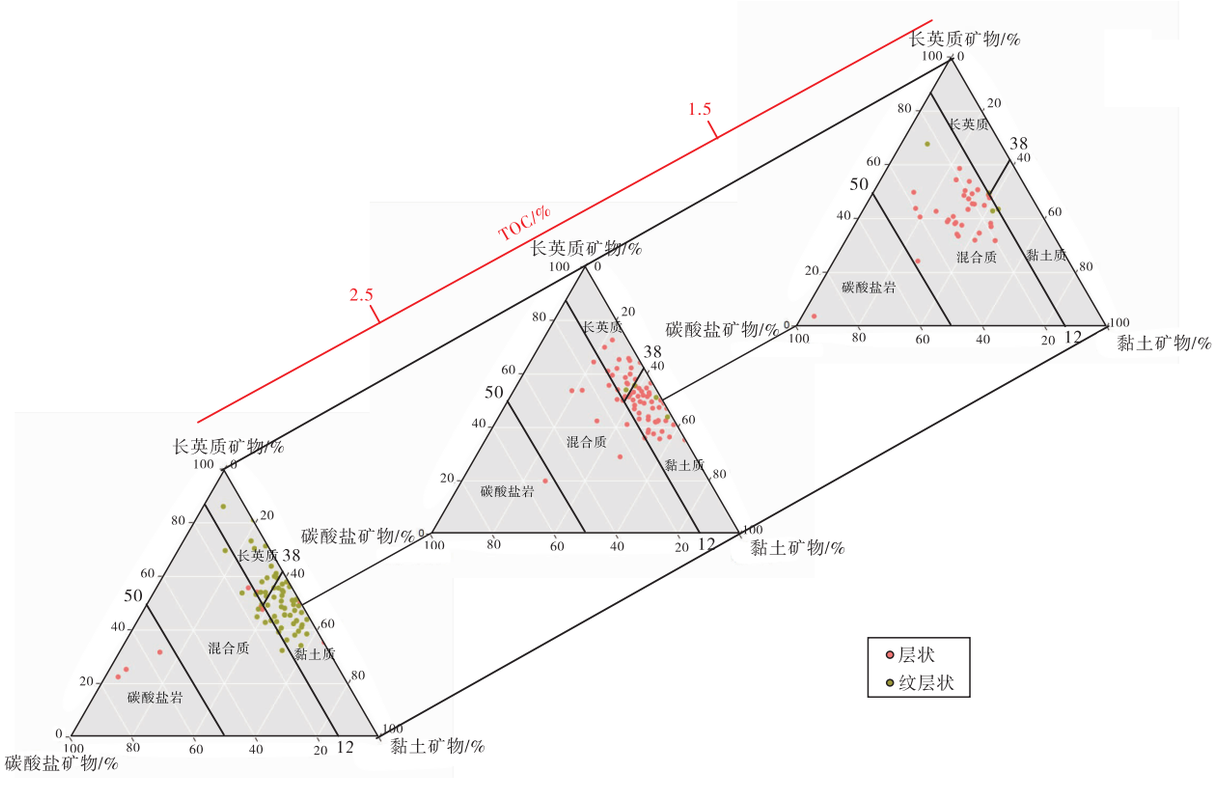
表 1 松辽盆地赞字井区青山口组页岩岩相划分标准

Table 1 Shale lithofacies division standard of the Qingshankou Formation in Zanzijing area in Songliao Basin

序号	页岩岩相类型	成像测井/(°) 0 120 240 360	岩心观察	TOC/%	沉积构造	矿物含量/%
1	富有机质纹层状长英质页岩	 2 cm	 2 cm	≥2.5	长英质纹层发育	黏土 < 38 长英质≥50 碳酸盐 < 12
2	中等有机质层状长英质页岩	 2 cm	 2 cm	1.5~2.5	长英质层状构造发育	黏土 < 38 长英质≥50 碳酸盐 < 12
3	富有机质纹层状黏土质页岩	 2 cm	 2 cm	≥2.5	黏土质纹层发育	黏土≥38 长英质 < 50 碳酸盐 < 12
4	中等有机质层状黏土质页岩	 2 cm	 2 cm	1.5~2.5	黏土质层状构造发育	黏土≥38 长英质 < 50 碳酸盐 < 12
5	混合质页岩	 2 cm	 2 cm	<1.5	黏土质、长英质、碳酸盐交互层	黏土 < 38 长英质 < 50 碳酸盐≥12

相：长英质页岩、黏土质页岩和混合质页岩，传统的岩石学命名方法以矿物含量 50% 为界确定岩石主名，有学者认为页岩中矿物组分丰富，通常将黏土矿物 、碳酸盐矿物和长英质矿物（长石+石英）作为三端元并将矿物含量 50% 作为界限，长英质矿物含量大于 50% 时为长英质页岩，黏土矿物含量大于 50% 时为黏土质页岩，碳酸盐矿物含量大于 50% 时为混合质页岩（王民等，2023）。但不同地区有不同的划分方案及适用范围，应该根据统计占比的方法，以最能反映岩相差异性的矿物含量数值确定具体界限。以松南地区赞字井区块为例，当碳酸盐矿物含量大于 12% 而小于 50% 时，该范围内数据占比 30% 左右，则定义为混合质页岩；黏土矿物含量大于 38% 时该范围内数据占比 40% 左右，则定义为黏土质页岩；当碳酸盐矿物含量大于 50% 时，可认为是碳酸盐岩，不过该范围数据点极少；剩余部分黏土矿物含量小于 38% 且数据约占 30%，则可命名为长英质页岩(图 1)。

然后，结合岩心与成像测井识别出相应的沉积构造(表 1)，其中纹层状长英质页岩的成像测井表现为明暗相间，有明显的纹层层理，岩心颜色以深灰色、灰黑色为主，沉积构造以平直的含长英质纹层为主，纹层厚度 1~2 mm；而层状长英质页岩的成像测井颜色条带表现出典型的层状特征，岩心颜色以深灰色、灰色为主，沉积构造以连续—断续的含长英质层状为主，层厚度 1~2 cm。纹层状黏土质页岩的成像测井表现为明显的纹层层理，岩心颜色以灰黑色、黑色为主，沉积构造以平直且连续的含黏土质纹层为主，纹层厚度小于 1 mm；层状黏土质页岩的成像测井层状构造特征明显，岩心颜色



松辽盆地青山口组赞字井区矿物含量井位包括 H2、C34-7、D86 和 H197 井

图 1　松辽盆地赞字井区青山口组页岩岩相划分方案四端元图

Fig. 1　Four-end diagram of shale lithofacies division scheme of the Qingshankou Formation in Zanzijing area in Songliao Basin

以灰黑色为主，沉积构造以连续—断续的含黏土质层状构造为主，层状厚度 1～2 cm。利用成像测井与岩心将长英质页岩和黏土质页岩细分为纹层状长英质页岩、层状长英质页岩、纹层状黏土质页岩和层状黏土质页岩(表 1)。

最终，通过统计得出研究区各类岩相 TOC 含量介于 0.5%～6%之间，其中 TOC 含量大于等于 2.5%的数据点占比 40%左右，TOC 含量介于 1.5%～2.5%之间的数据点占比 30%，TOC 含量小于 1.5%的数据点占比 30%左右，以最能反映岩相差异性的 TOC 含量数据确定具体界限。以松南地区赞字井区为例，纹层状长英质页岩和纹层状黏土质页岩 TOC 含量大多数大于等于 2.5%，可定义为富有机质，层状长英质页岩和层状黏土质页岩 TOC 含量大多数介于 1.5%～2.5%之间，可定义为中等有机质，混合质页岩 TOC 含量则大多数小于 1.5%，可定义为低有机质。因此可根据有机质丰度和占比区分出富有机质的 TOC 含量大于等于 2.5%、中等有机质 TOC 含量介于 2.5%～1.5%之间与低有机质 TOC 含量小于 1.5%的岩相类型(图 1)，确立富有机质纹层状黏土质页岩、中等有机质层状黏土质页岩、富有机质纹层状长英质页岩、中等有机质层状长英质页岩以及混合质页岩 5 种页岩岩相类型(表 1)。研究区目的层位松辽盆地青山口组，真实岩相的划分方案适用于松辽盆地青山口组赞字井区，是否适用于其他区块，还是要根据其他区块的具体区域背景及资料来判断，四端元图(图 1) 的划分方法（矿物含量三角图和 TOC 含量相结合，其中矿物含量三角图中又区分出不同的构造类型）具有一定的推广意义，但是该页岩岩相划分方案中的具体岩相类型划分界限需根据不同研究区的具体情况而定。

2　XGBoost 算法

XGBoost 是一种有监督的集成学习算法，它在梯度提升算法的基础上进行了一系列的改进和优化。XGBoost 算法的核心思想是通过不断地拟合残差，构建一系列弱学习器，并将它们组合起来形成

一个强学习器。与传统的梯度提升算法相比，XGBoost 具有一些显著的优势，因为它引入了二阶导数信息(Chen and Guestrin, 2016)，加快了算法的收敛速度，并能够处理稀疏数据。XGBoost 算法在许多数据挖掘和机器学习任务中都取得了很好的效果(李红斌等，2022)，例如分类、回归、排序等(罗钰涵等，2022；史长林等，2022)。构建分类模型的具体步骤(Chen and Guestrin, 2016) 如下：(1)数据准备：收集和整理数据，包括特征和目标变量，可能需要进行数据清洗、特征工程等预处理步骤；(2)加载 XGBoost 库：使用适当的方式导入 XGBoost 库；(3)创建 XGBoost 分类器：通过调用 XGBClassifier 类创建一个分类器对象；(4)设置模型参数：例如调整决策树数量、学习率等参数，可以根据经验或通过网格搜索等方法找到最优参数；(5)数据划分：将数据集划分为训练集和测试集；(6)拟合训练数据：使用训练集对模型进行训练；(7)模型评估：在测试集上评估模型的性能；(8)进行预测：使用训练好的模型对新数据进行预测。

2.1 Boosting 算法的基本原理与算法流程

Boosting 算法是一种集成学习框架，而 XGBoost 算法是基于 Boosting 框架的一种高效实现(刘忠宝等，2019)。在 Boosting 算法体系中，通常采用迭代串行的方式生成一系列模型，然后将这些模型进行线性加权相加，得到最终的集成学习器。假设已经迭代了 m−1 次，得到的集成模型为 F_{m-1}。那么在接下来的一次迭代中，需要训练的是 F_m，它应该是能使新生成的集成模型在训练集上损失信息最小的模型 (He *et al.*, 2016)(图 2)。

2.2 XGBoost 变量重要性度量

根据变量重要性的度量结果，可以进行特征选择，去除不重要或冗余的变量，从而减少模型计算的复杂度，提高模型的泛化能力(Xue *et al.*, 2024)。

在 XGBoost 中衡量变量重要性有 3 个选项，分别为：(1) Weight (权重)：某个特征被用于在所有树中拆分数据的次数，这个参数反映了该特征的重要性，也即权重；(2) Cover (样本数量)：首先得到某个特征被用于在所有树中拆分数据的次数(数量)，然后要利用经过这些拆分点的训练数据数量赋予权重；(3) Gain (平均增益值)：使用某个变量进行拆分时，获得的平均训练损失减少量，也即表示一个变量在所有计算树中对预测结果的平均增益值。这个参数反映了该变量在每个节点上的分裂能力。对于这 3 种重要性度量指标，有以下公式 (李占山和刘兆赓，2019)：

$$FScore = |X| \tag{1}$$

$$AverageGain = \frac{\sum Gainx}{FScore} \tag{2}$$

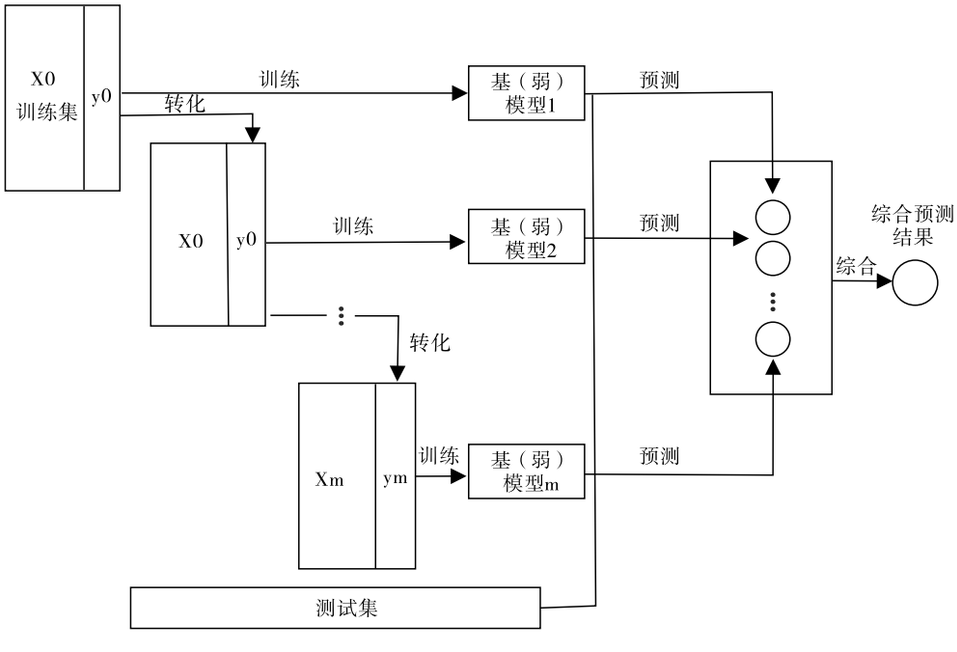


图 2 Boosting 算法流程图

Fig. 2 Flow chart of boosting algorithm

$$AverageGain = \frac{\sum Coverx}{FScore} \tag{3}$$

X 表示将所求变量分类到叶子节点的集合；Gainx 是 X 中每个叶子节点在分割时的节点增益值；Coverx 则是 X 中落在每个节点的样本数量。

2.3　XGBoost 参数调优

参数调优的意义在于找到模型的最佳参数配置，以提高模型的性能和准确性。通过调整参数，可以优化模型性能，不同的参数值可能对模型的预测能力（Su *et al.*, 2023）产生显著影响。

其中有 3 种较为常见的参数调优方法：（1）随机搜索（Random Search）：通过随机选择参数值尝试不同的组合；（2）网格搜索（Grid Search）：网格搜索系统地遍历参数空间中定义的网格；（3）基于绩效的调优（Performance-based Tuning）：这种方法是根据模型在验证集或其他评估指标上的表现来调整参数。

XGBoost 算法主要调整的模型参数有决策树的数量、树的最大深度、最小叶子节点样本数、正则化参数、学习率等。

3　XGBoost 算法预测页岩岩相

3.1　数据准备

本次使用有监督学习算法的 XGBoost 作为回归器，利用常规 7 条（种）测井数据作为特征变量数据集，即：自然伽马测井（GR）、自然电位（SP）、深侧向电阻率（LLD）、浅侧向电阻率（LLS）、密度（DEN）、声波时差（AC）和井径测井（CAL）7 种曲线（图 3），建立可预测页岩岩相类型的模型。首先，根据测井数据特点，可以选择删除包含过多缺失值的样本，由于各类测井曲线的量纲不同，如果直接将测井数据作为输入训练模型，可能会影响模型的预测结果，而对测井数据进行归一化处理是一个很好的解决量纲差异问题的方法（王民等，2023）。可通过公式（4）最大最小归一化函数将输入曲线值映射到［0，1］这个区间：

$$X^* = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{4}$$

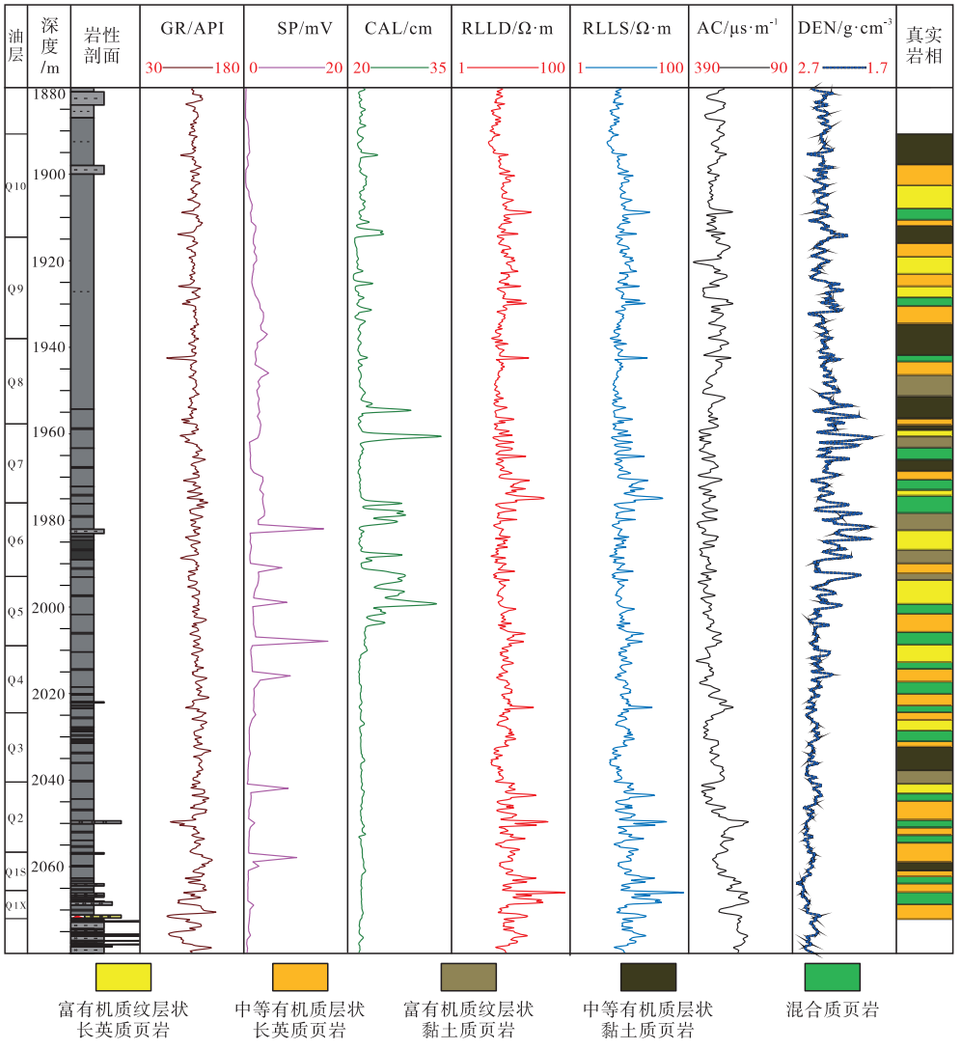
X^* 代表归一化后数据，$X_{\min}$ 为样本数据最小值，$X_{\max}$ 为样本数据最大值。

研究区不同类型的页岩岩相的测井响应特征区分较为明显，基本可以看出富有机质纹层状长英质页岩常规测井特征表现为中低伽马、中低声波、中高电阻、中高密度、中低自然电位；中等有机质层状长英质页岩则表现为低伽马、中低声波、中高电阻、中高密度、中低自然电位；富有机质纹层状黏土质页岩表现为高伽马、高声波、低电阻、低密度、中高自然电位；中等有机质层状长英质页岩表现为中高伽马、高声波、低电阻、中低密度、中高自然电位；混合质页岩则表现为中低伽马、低声波、高电阻、高密度、中低自然电位（图 3）。因此可以由这 7 种测井数据与岩相建立联系。页岩岩相样品数据分布比例如表 2 所示，其中此次模型训练按照一般常用的训练集与测试集 8∶2 的比例进行。

3.2　特征分析

在 XGBoost 模型的建立中，通过选择合适的特征变量，可以降低空间的维度，减少计算时间和过拟合的风险。一些变量可能包含噪声或对模型的预测能力没有贡献，通过变量选择可以去除这些干扰因素（王民等，2023），并且重要的变量（张家臣等，2022）可以帮助模型更好地学习数据的模式和规律，从而提高预测准确性。本研究将 7 条测井曲线作为模型特征变量，并分析其中各个变量之间相关性，做出有效的取舍。去除冗余变量，较多的相关特征可能会增加模型的复杂度，导致过拟合（Wang and Zhang, 2024）。保留具有较高信息量的变量，如果一个变量与目标变量有较强的相关性，并且能够提供独特的信息，那么它可能比其他相关变量更有价值。变量散点矩阵即可反映出变量间的相关性：通过观察散点矩阵中的点分布，可以初步判断变量之间的相关性。如果 2 个变量的数据点在矩阵中呈现出明显的线性趋势，可能表示它们之间存在较强的相关性。散点矩阵中的点的密集程度和分布情况可以反映每个变量的取值范围和数据分布情况。5 种岩相分别用 0-4 的编号表示（图 4），可以看出对岩相具有区分能力的测井组合主要有 SP-GR、AC-DEN、AC-GR、SP-AC、CAL-DEN 组合。其他组合分类界线并不明显。

在 XGBoost 算法中，变量热力图可以反映很多有用信息（图 5），通过观察变量热力图中相邻区域的颜色相似性，可以初步了解变量之间的相关性。



不同测井参数的取值可分为高、中、低 3 类，其中：高伽马：110~130 API，中伽马：90~110 API，低伽马：70~90 API；高声波：320~360 μs/m，中声波：260~320 μs/m，低声波：220~260 μs/m；高电阻：11~15 Ω·m，中电阻：8~11 Ω·m，低电阻：5~8 Ω·m；高密度：2.3~2.5 g/cm^3，中密度：2.1~2.3 g/cm^3，低密度：1.9~2.1 g/cm^3；高自然电位：10~15 mV，中自然电位：5~10 mV，低自然电位：0~5 mV

图 3 松辽盆地赞字井区 H2 井青山口组特征测井曲线柱状图

Fig. 3 Histogram of logging curve characteristics of the Qingshankou Formation of Well H2 in Zanzijing area in Songliao Basin

表 2 松辽盆地赞字井区 H2 井青山口组页岩岩相样品统计

Table 2 Statistics of shale lithofacies samples of the Qingshankou Formation from Well H2 in Zanzijing area in Songliao Basin

岩相编号	岩相名称	样品点/个	占比/%
0	富有机质纹层状长英质页岩	270	19.7
1	中等有机质层状长英质页岩	447	30.9
2	富有机质纹层状黏土质页岩	155	11.8
3	中等有机质层状黏土质页岩	277	16.7
4	混合质页岩	303	20.9
	总计	1452	100

本次实验中深侧向电阻率（LLD）与浅侧向电阻率（LLS）这 2 个变量的相关性较大，高达 99%，这可能对模型影响较大，保留 1 个深侧向电阻率（LLD）即可。

最后通过公式（2）和公式（3）计算出 XGBoost 变量重要性度量，可以帮助了解模型在进行预测时对不同变量的依赖程度。这有助于解释模型的决策过程，并确定哪些变量对模型的预测结果具有较大的影响，从而做出有利取舍，提高模型准确率。可以直观发现研究区 SP 的重要性最高，CAL、AC、DEN、LLD、GR 较高（图 6）。LLS 重要性得

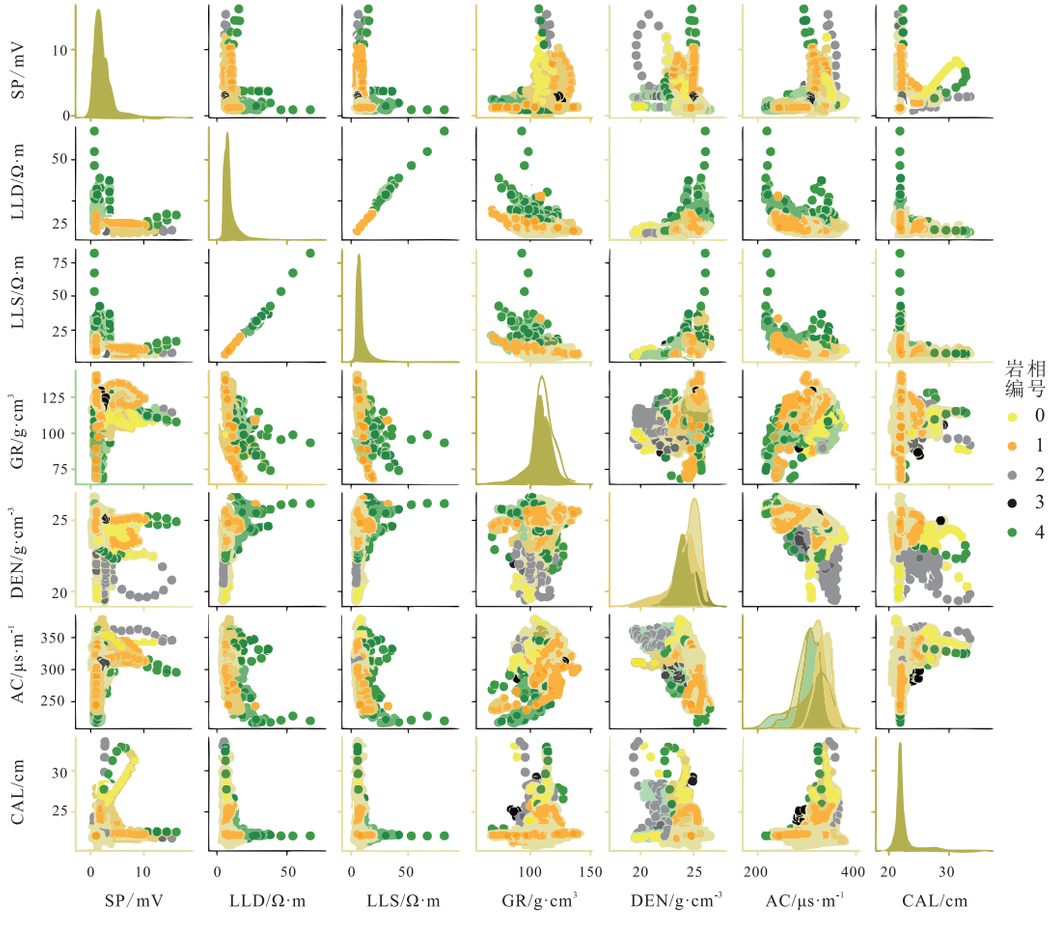


图 4　松辽盆地赞字井区 H2 井青山口组 7 种测井曲线变量的散点矩阵图(岩相编号含义见表 2)

Fig. 4　Scatter matrix plots diagram of seven logging curve variables of the Qingshankou Formation of Well H2 in Zanzijing area in Songliao Basin(The meanings of lithofacies code are shown in Table 2)

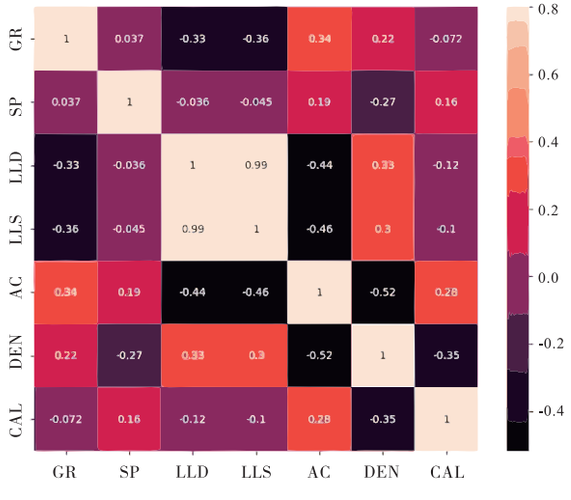


图 5　松辽盆地赞字井区 H2 井青山口组 7 种测井曲线变量热力图

Fig. 5　Feature heat map of seven logging curve variables of the Qingshankou Formation of Well H2 in Zanzijing area in Songliao Basin

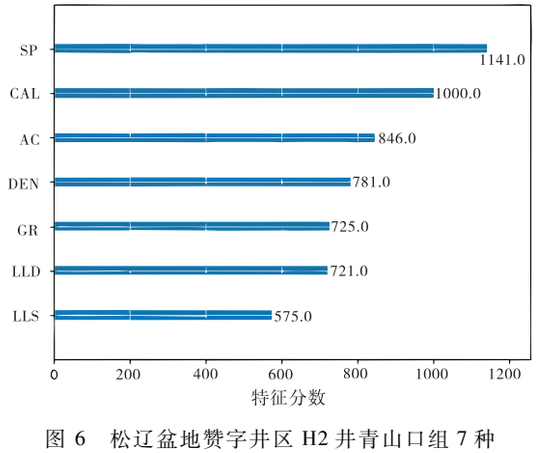


图 6　松辽盆地赞字井区 H2 井青山口组 7 种测井曲线变量重要性分析图

Fig. 6　Importance analysis diagram of seven logging curve variables of the Qingshankou Formation of Well H2 in Zanzijing area in Songliao Basin

分最低，且 LLS 与 LLD 曲线的相关性较高，删除 LLS 变量进行训练后并通过公式（5）计算得到的测试集准确率可以提高大概 4%（表 3 中的模型 2）。准确率通常是通过在测试集或验证集上进行预测并与真实标签进行比较来计算的。准确度 ACC：正例和负例中预测正确数量占总数量的比例（李占山等，2019），用公式表示：

$$ACC = \frac{TP + TN}{TP + FP + FN + TN} \quad (5)$$

TP（True Positives）：真正例，实际为正例且被预测为正例；

FP（False Positives）：假正例，实际为负例却被预测为正例；

FN（false Negatives）：假负例，实际为正例却被预测为负例；

TN（True Negatives）：真负例，实际为负例且被预测为负例。

井径（CAL）曲线作为特征值重要性很高，并且井径变量参与模型训练可以使得模型通过公式（5）计算出的准确度提高大概 6%（表 3 中的模型 3），所以 CAL 对于此次岩相的划分具有重要作用。并且从 CAL 曲线（图 7）可以看出黏土质页岩岩相

表 3 松辽盆地赞字井区 H2 井青山口组 XGBoost 模型实验结果统计

Table 3 Statistics of experimental results of XGBoost model of the Qingshankou Formation of Well H2 in Zanzijing area in Songliao Basin

模型	模型特征	模型准确率/%
模型1	含 LLS 曲线，无 CAL 曲线；未进行网格搜索及参数优选	78.73
模型2	无 LLS 曲线，无 CAL 曲线；未进行网格搜索及参数优选	82.07
模型3	无 LLS 曲线，含 CAL 曲线；未进行网格搜索及参数优选	88.32
模型4	无 LLS 曲线，含 CAL 曲线；进行网格搜索及参数优选	90.03
模型5	无 LLS 曲线，含 CAL 曲线；进行网格搜索及参数优选；去除 SP 曲线	78.35

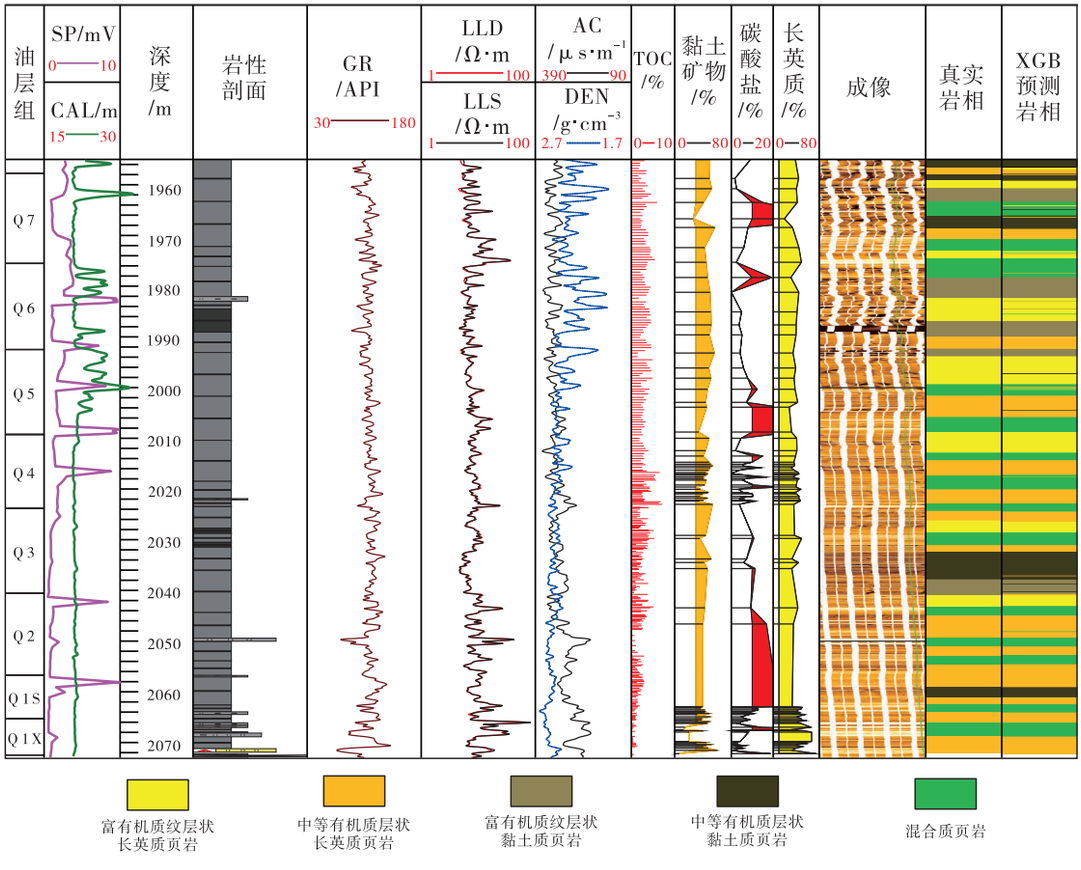


图 7 松辽盆地赞字井区 H2 井青山口组综合柱状图

Fig. 7 Comprehensive histogram of the Qingshankou Formation of Well H2 in Zanzijing area in Songliao Basin

可能会发生遇水膨胀，且该目的层青一段泥页岩为含粉砂质泥页岩（柳波等，2021），会出现扩径问题，所以 CAL 曲线对该目的层的页岩岩相识别具有一定指示作用。SP 曲线重要性较高(图 6)，去掉 SP 曲线进行模型训练会使得模型准确率下降(表 3 中的模型 5)，其中 SP 测井曲线反应地层渗透率（车世琦，2018），从 SP 曲线(图 7）趋势上可以看出黏土质页岩岩相和含粉砂质泥页岩层 SP 曲线有回返趋势。通过以上的特征曲线分析可以有效地提高机器学习模型的准确率。

3.3　参数调优

网格搜索（Grid Search）是一种常用的参数调优方法，它在给定的参数空间中系统地尝试不同的参数组合，以找到最优的参数配置。本研究利用网格搜索交叉验证的方式进行参数优选，主要有以下步骤：(1）定义参数空间：确定要调节的参数及其可能的取值范围；(2）创建网格：根据参数空间生成一个网格，包含所有可能的参数组合；(3）遍历网格：对于每个参数组合，进行实验；(4）评估结果：根据特定的评估指标（如准确性、误差等），比较不同参数组合的性能；(5）选择最优参数：根据评估结果，选择表现最好的参数组合。

通过此方法先筛选出最佳模型数量及决策树的数量，再对树的最大深度、最小叶子节点样本数、随机采样比例、特征随机采样比例、正则化参数、学习率进行网格搜索，各参数搜索范围如表 4 所示，并通过公式（5）最终确定测试集准确率最高值为 90.03%(表 3)，由此确定出相应的最优参数(表 4)。这些参数在进行网格搜索的过程中需要进行多次搜索，不断缩小搜索范围以求取最优值。

其中可以通过损失曲线(Loss Curve）来进行评估模型性能，它展示了模型在训练过程中损失函数（通常是预测误差的度量）随迭代次数或训练数据量的变化情况。损失曲线有助于了解模型在训练过程中的收敛情况。如果曲线逐渐下降并趋于稳定，说明模型正在学习和改进，可能已经接近最优状态。如果损失曲线在训练集上持续下降，但在验证集或测试集上开始上升，可能意味着模型出现了过拟合。相反，如果损失曲线在所有数据集上都没有明显下降，可能表示模型欠拟合。比较不同模型的损失曲线有助于选择性能更好的模型或超参数设置。根据损失曲线的趋势，可以确定何时停止训练，以避免过拟合或浪费计算资源。本次模型的损失曲线逐渐下降并趋于稳定，说明模型正在学习和改进，可能已经接近最优状态(图 8)。参照 loss 曲线判别标准(图 9）可以说明此次模型的学习率最优，模型的参数筛选较优。

表 4　松辽盆地赞字井区 H2 井青山口组 XGBoost 模型最优参数统计表

Table 4　Statistics of optimal parameters of XGBoost model of the Qingshankou Formation of Well H2 in Zanzijing area in Songliao Basin

参数	参数网格搜索范围	最优参数值
决策树数量	10~300	70
树的最大深度	3~10	5
最小叶子节点样本数	1~6	1
随机采样比例	0.3~1	0.8
特征随机采样比例	0.3~1	0.8
Ll 正则化权重项	[0,0.01,0.1,1,100]	0
Gamma	0~0.5	0
学习率	0.01~0.1	0.1

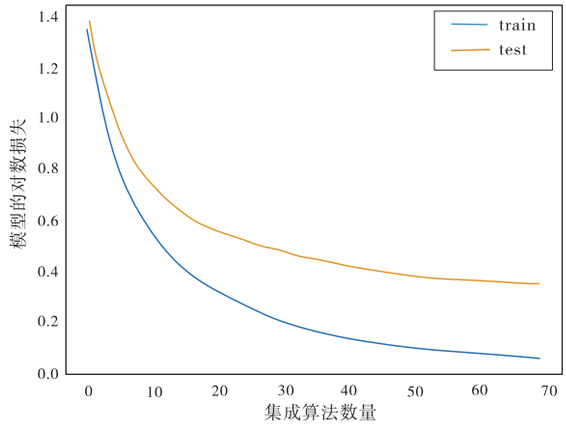


图 8　松辽盆地赞字井区 H2 井青山口组 XGboost 模型 loss 曲线

Fig. 8　Loss curves of XGboost model of the Qingshankou Formation of Well H2 in Zanzijing area in Songliao Basin

模型训练的好坏重点还要关注最终的准确率，准确率是指模型预测正确的结果所占的比例，经过特征筛选以及参数调优，通过公式（5）计算出本次最终的模型测试集准确率提高 12%左右，高达 90.03%(图 7)。经过模型特征分析及参数调优后，通过公式（5）计算出测试集准确率从 78.73%上升到 90.03%，有明显提升。

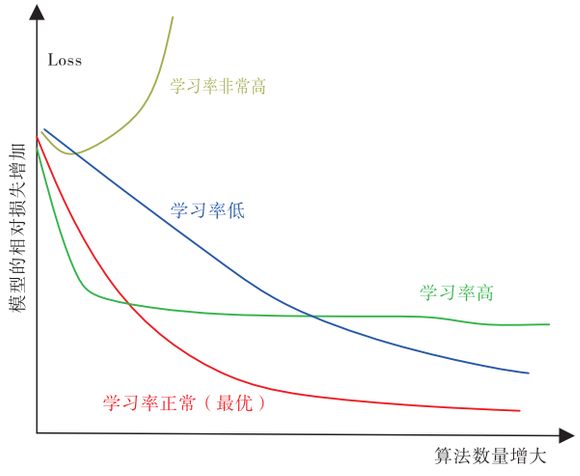


图 9 loss 曲线判别标准

Fig. 9 Discrimination criteria of loss curves

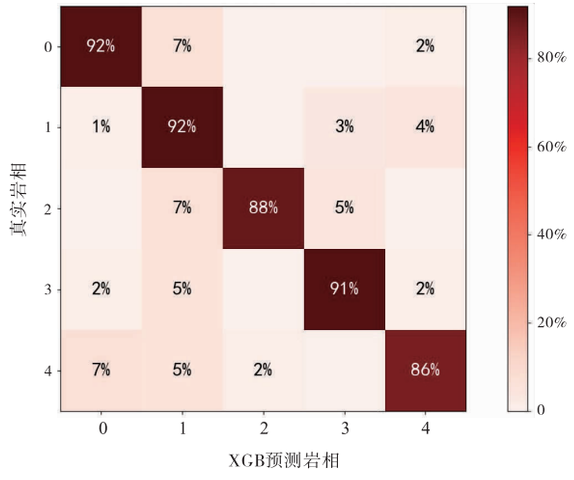


图 10 松辽盆地赞字井区 H2 井青山口组不同岩相识别模型混淆矩阵图

Fig. 10 Confusion matrix diagram of different lithofacies identification models of the Qingshankou Formation of Well H2 in Zanzijing area in Songliao Basin

3.4 结果分析

最终将训练好的模型 4（表 3）保存后，与前述通过岩心观察、成像测井、矿物含量及 TOC 含量确定的页岩岩相类型的划分方案划分出的真实岩相进行比对（图 7），预测出的岩相与实际划分的岩相拟合度较高，通过公式（5）计算出测试集准确率达 90.03%。通过对不同岩相的预测值与真实值相比较（图 10），可以看出拟合度最高的为编号为“0”的富有机质纹层状长英质页岩和编号为“1”的中等有机质层状长英质页岩，2 种岩相的真实值和预测值拟合度高达 92%，整体上不同的岩相拟合度都在 90%左右，说明模型准确率较高。虽然本次算例中有部分曲线异常，但最终对结果影响不大，说明 XGBoost 算法在处理数值变化幅度较大的测井曲线时，仍然具有一定稳定性。事实表明此次岩相预测模型在研究区可信性较高。

4 结论

XGBoost 算法能够充分捕捉测井曲线有效信息并与页岩岩相建立联系，是一种已经被证实的表现力很好的有监督的机器学习算法，在页岩岩相的识别中拥有很好的应用前景。

1）利用有监督的机器学习算法来预测页岩岩相时，建立的岩相划分标准要能充分反映出研究区页岩岩相的辨识差异性，不同地区有不同的划分方案及适用范围，应该根据统计占比的方法，以最能反映岩相差异性的数值来确定具体矿物含量和 TOC 含量的界限。

2）在变量选择时，XGBoost 算法提供了变量重要性的评估指标，相关变量可能会提供相似的信息，导致模型过于依赖这些变量，因此应该进行变量分析并筛选，从而使得数据更具有分类能力。说明合理选择变量对岩相划分过程具有重要作用。

3）在参数优选上网格搜索具有全面性，XGBoost 算法可以系统地遍历参数空间中的所有可能组合，确保不会遗漏任何潜在的最优解。相对其他复杂的调优方法，网格搜索的实现较为简单，适用于多种问题并且结果相对较稳定。在网格搜索过程中应该注意进行多次优选，不断缩小搜索范围以求取最优值。

参考文献（References）

车世琦. 2018. 测井资料用于页岩岩相划分及识别：以涪陵气田五峰组—龙马溪组为例. 岩性油气藏，30(1)：121-132.［Che S Q. 2018. Shale lithofacies identification and classification by using logging data：a case of Wufeng-Longmaxi Formation in Fuling Gas Field，Sichuan Basin. Lithologic Reservoirs，30(1)：121-132］

陈登辉，隋清霖，赵晓健，荆德龙，滕家欣，高永宝. 2019. 西昆仑穆呼锰矿晚石炭世含锰碳酸盐岩地质地球化学特征及其沉积环境. 沉积学报，37(3)：477-490.［Chen D H，Sui Q L，Zhao X J，Jing D L，Teng J X，Gao Y B. 2019. Geology，geochemical characteristics，and sedimentary environment of Mn-bearing carbonate from the Late Carboniferous Muhu manganese deposit in West Kunlun. Acta

Sedimentologica Sinica,37(3): 477-490]

付秀丽,蒙启安,郑强,王忠杰,金明玉,白月,崔坤宁. 2022. 松辽盆地古龙页岩有机质丰度旋回性与岩相古地理. 大庆石油地质与开发,41(3): 38-52. [Fu X L, Meng Q A, Zheng Q, Wang Z J, Jin M Y, Bai Y, Cui K N. 2022. Cyclicity of organic matter abundance and lithofacies paleogeography of Gulong shale in Songliao Basin. Petroleum Geology & Oilfield Development in Daqing,41(3): 38-52]

何伟,陈杨,雷玉雪,钱程,陈科,林拓,宋腾. 2021. 鄂西地区五峰组—龙马溪组岩石相与页岩气富集关系分析: 以鄂红地 1 井为例. 煤炭学报,46(3): 1014-1023. [He W, Chen Y, Lei Y X, Qian C, Chen K, Lin T, Song T. 2021. Analyses of the relationship between lithology and shale gas accumulation for the Wufeng Formation to Longmaxi Formation in the west of Hubei Province: a case study of the Erhongdi 1 well. Journal of China Coal Society,46(3): 1014-1023]

赖锦,李红斌,张梅,白梅梅,赵仪迪,范旗轩,庞小娇,王贵文. 2023. 非常规油气时代测井地质学研究进展. 古地理学报,25(5): 1118-1138. [Lai J, Li H B, Zhang M, Bai M M, Zhao Y D, Fan Q X, Pang X J, Wang G W. 2023. Advances in well logging geology in the era of unconventional hydrocarbon resources. Journal of Palaeogeography(Chinese Edition),25(5): 1118-1138]

李红斌,王贵文,王松,庞小娇,刘士琛,包萌,彭寿昌,赖锦. 2022. 基于 Kohonen 神经网络的页岩油岩相测井识别方法: 以吉木萨尔凹陷二叠系芦草沟组为例. 沉积学报,40(3): 626-640. [Li H B, Wang G W, Wang S, Pang X J, Liu S C, Bao M, Peng S C, Lai J. 2022. Shale oil lithofacies identification by Kohonen neural network method: the case of the Permian Lucaogou Formation in Jimusaer Sag. Acta Sedimentologica Sinica,40(3): 626-640]

李宁,冯周,武宏亮,田瀚,刘鹏,刘英明,刘忠华,王克文,徐彬森. 2023. 中国陆相页岩油测井评价技术方法新进展. 石油学报,44(1): 28-44. [Li N, Feng Z, Wu H L, Tian H, Liu P, Liu Y M, Liu Z H, Wang K W, Xu B S. 2023. New advances in methods and technologies for well logging evaluation of continental shale oil in China. Acta Petrolei Sinica,44(1): 28-44]

李占山,刘兆赓. 2019. 基于 XGBoost 的特征选择算法. 通信学报,40(10): 101-108. [Li Z S, Liu Z G. 2019. Feature selection algorithm based on XGBoost. Journal on Communications,40(10): 101-108]

刘忠宝,刘光祥,胡宗全,冯动军,朱彤,边瑞康,姜涛,金治光. 2019. 陆相页岩层系岩相类型、组合特征及其油气勘探意义: 以四川盆地中下侏罗统为例. 天然气工业,39(12): 10-21. [Liu Z B, Liu G X, Hu Z Q, Feng D J, Zhu T, Bian R K, Jiang T, Jin Z G. 2019. Lithofacies types and assemblage features of continental shale strata and their significance for shale gas exploration: a case study of the Middle and Lower Jurassic strata in the Sichuan Basin. Natural Gas Industry,39(12): 10-21]

柳波,孙嘉慧,张永清,贺君玲,付晓飞,杨亮,邢济麟,赵小青. 2021. 松辽盆地长岭凹陷白垩系青山口组一段页岩油储集空间类型与富集模式. 石油勘探与开发,48(3): 521-535. [Liu B, Sun J H, Zhang Y Q, He J L, Fu X F, Yang L, Xing J L, Zhao X Q. 2021. Reservoir space and enrichment model of shale oil in the first member of Cretaceous Qingshankou Formation in the Changling Sag, southern Songliao Basin, NE China. Petroleum Exploration and Development,48(3): 521-535]

罗钰涵,葛政俊,谌廷姗,洪亚飞,林波,刘宗堡. 2022. 基于卷积神经网络的陆相页岩油岩相类型识别方法及系统. 中国专利: CN114881171A.2024-11-29. [Luo Y H, Ge Z J, Shen T S, Hong Y F, Lin B, Liu Z B. 2022. The identification method and system of continental shale facies based on convolutional neural network are introduced in this paper. Chinese Patent: CN114881171A. 2024-11-29]

毛玉丹. 2023. 页岩岩相测井识别方法. 石油知识,(3): 54-55. [Mao Y D. 2023. Identification method of shale lithofacies by logging. Petroleum Knowledge,(3): 54-55]

庞小娇,王贵文,匡立春,赵飞,李红斌,韩宗晏,白天宇,赖锦. 2023. 沉积环境控制下的页岩岩相组合类型及测井表征: 以松辽盆地古龙凹陷青山口组为例. 古地理学报,25(5): 1156-1175. [Pang X J, Wang G W, Kuang L C, Zhao F, Li H B, Han Z Y, Bai T Y, Lai J. 2023. Logging evaluation of lithofacies and their assemblage under control of sedimentary environment: a case study of the Qingshankou Formation in Gulong sag, Songliao Basin. Journal of Palaeogeography(Chinese Edition), 25(5): 1156-1175]

彭军,曾垚,杨一茗,于乐丹,许天宇. 2022. 细粒沉积岩岩石分类及命名方案探讨. 石油勘探与开发,49(1): 106-115. [Peng J, Zeng Y, Yang Y M, Yu L D, Xu T Y. 2022. Discussion on classification and naming scheme of fine-grained sedimentary rocks. Petroleum Exploration and Development,49(1): 106-115]

彭丽,伍轶鸣,练章贵,彭鹏,王剑,苏洲,易珍丽. 2019. 陆相断陷湖盆高频层序特征及其沉积演化: 以渤海湾盆地济阳坳陷沙三下亚段为例. 石油与天然气地质,40(4): 789-798. [Peng L, Wu Y M, Lian Z G, Peng P, Wang J, Su Z, Yi Z L. 2019. Features and sedimentary evolution of high-frequency sequence in continental lacustrine rift basin: example of the lower Shahejie member 3 in Jiyang Depression, Bohai Bay Basin. Oil & Gas Geology,40(4): 789-798]

沈骋,任岚,赵金洲,陈铭培. 2021. 页岩岩相组合划分标准及其对缝网形成的影响: 以四川盆地志留系龙马溪组页岩为例. 石油与天然气地质,42(1): 98-106, 123. [Shen C, Ren L, Zhao J Z, Chen M P. 2021. Division of shale lithofacies associations and their impact on fracture network formation in the Silurian Longmaxi Formation, Sichuan Basin. Oil & Gas Geology,42(1): 98-106,123]

史长林,魏莉,张剑,杨丽娜. 2022. 基于机器学习的储层预测方法. 油气地质与采收率,29(1): 90-97. [Shi C L, Wei L, Zhang J, Yang L N. 2022. Reservoir prediction method based on machine learning. Petroleum Geology and Recovery Efficiency,29(1): 90-97]

谌丽,王才志,宁从前,刘英明,王浩. 2023. 基于机器学习的鄂尔多斯盆地陇东地区长 7 段岩相测井识别方法. 油气藏评价与开发,13(04): 525-536. [Shen L, Wang C Z, Ning C Q, Liu Y M,

Wang H. 2023. Well-log lithofacies classification based on machine learning for Chang－7 member in Longdong area of Ordos Basin. Petroleum Reservoir Evaluation and Development,13(4): 525－536]

田瀚,闫伟林,武宏亮,闫学洪,李潮流,郑建东,冯周. 2023. 一种陆相页岩油岩相测井定量识别方法. 地球物理学进展,38(5): 2122－2134. [Tian H,Yan W L,Wu H L,Yan X H,Li C L,Zheng J D,Feng Z. 2023. Logging quantitative identification method for lithofacies of continental shale oil. Progress in Geophysics,38(5): 2122－2134]

王民,杨金路,王鑫,李进步,徐亮,言语. 2023. 基于随机森林算法的泥页岩岩相测井识别. 地球科学,48(1): 130－142. [Wang M,Yang J L,Wang X,Li J B,Xu L,Yan Y. 2023. Identification of shale lithofacies by well logs based on random forest algorithm. Earth Science,48(1): 130－142]

徐传正,李鑫,田继军,吝文,蒋立伟,张治恒. 2021. 四川盆地南缘龙马溪组混合岩相页岩及其沉积环境. 煤炭科学技术,49(5): 208－217. [Xu C Z,Li X,Tian J J,Lin W,Jiang L W,Zhang Z H. 2021. Mixed lithofacies shale and depositional environment of Longmaxi Formation in southern margin of Sichuan Basin. Coal Science and Technology,49(5): 208－217]

薛纯琦,吴建光,张健,张守仁,吴翔,程璐,钟建华. 2021. 机器学习在页岩岩相识别中的应用: 以鄂尔多斯临兴地区山西太原组页岩为例. 2021 年煤层气学术研讨会,2021－10－10. [Xue C Q,Wu J G,Zhang J,Zhang S R,Wu X,Cheng L,Zhong J H. 2021. The application of machine learning in shale lithofacies identification is taken as an example of Taiyuan Formation shale in Linxing area of Ordos. Annual CBM Academic Symposium in 2021,2021－10－10]

张家臣,邓金根,谭强,石林. 2022. 基于 XGBoost 的测井曲线重构方法. 石油地球物理勘探,57(3): 697－705,496. [Zhang J C,Deng J G,Tan Q,Shi L. 2022. Reconstruction of well logs based on XGBoost. Oil Geophysical Prospecting,57(3): 697－705,496]

张晋言. 2013. 泥页岩岩相测井识别及评价方法. 石油天然气学报,35(4): 96－103,167－168. [Zhang J Y. 2013. Shale lithofacies logging identification and evaluation. Journal of Oil and Gas Technology,35(4): 96－103,167－168]

张益粼,王贵文,宋连腾,包萌,黄玉越,赖锦,王松,黄立良. 2023. 页岩岩相测井表征方法: 以准噶尔盆地玛湖凹陷风城组为例. 地球物理学进展,38(1): 393－408. [Zhang Y L,Wang G W,Song L T,Bao M,Huang Y Y,Lai J,Wang S,Huang L L. 2023. Logging identification method of shale lithofacies: a study of Fengcheng Formation in Mahu Sag,Junggar Basin. Progress in Geophysics,38(1): 393－408]

赵贤正,周立宏,蒲秀刚,金凤鸣,时战楠,肖敦清,韩文中,姜文亚,张伟,汪虎. 2019. 断陷湖盆湖相页岩油形成有利条件及富集特征: 以渤海湾盆地沧东凹陷孔店组二段为例. 石油学报,40(9): 1013－1029. [Zhao X Z,Zhou L H,Pu X G,Jin F M,Shi Z N,Xiao D Q,Han W Z,Jiang W Y,Zhang W,Wang H. 2019. Favorable formation conditions and enrichment characteristics of lacustrine facies shale oil in faulted lake basin: a case study of Member 2 of Kongdian Formation in Cangdong sag,Bohai Bay Basin. Acta Petrolei Sinica,40(9): 1013－1029]

Chen T Q,Guestrin C. 2016. XGBoost: a scalable tree boosting system. The ACM SIGKDD International Conference. DOI: 10.1145/2939672.2939785.

He J H,Ding W L,Jiang Z X,Li A,Wang R Y,Sun Y X. 2016. Logging identification and characteristic analysis of the lacustrine organic-rich shale lithofacies: a case study from the Es3L shale in the Jiyang Depression,Bohai Bay Basin,Eastern China. Journal of Petroleum Science & Engineering,145(1): 238－255.

Pang Q,Hu G,Hu C W,Meng F S,Wang B Z,Zhang J Y. 2024. The lithofacies of sandstones interbedded with shales: implication for organic matter accumulation of Triassic deep lacustrine setting,Southern Ordos Basin. ACS Omega,9(22): 23266－23282.

Peng Y X,Guo S B. 2023. Lithofacies analysis and paleosedimentary evolution of Taiyuan Formation in Southern North China Basin. Journal of Petroleum Science & Engineering,220: 111127.

Su K,Yuan X,Huang Y K,Yuan Q,Yang M H,Sun J W,Li S Y,Long X Y,Liu L,Li T W,Yuan Z Q. 2023. Improved prediction of knee osteoarthritis by the machine learning model XGBoost. Indian Journal of Orthopaedics,57(10): 1667－1677.

Sun B,Liu X P,Liu J,Wang G C,Shu H L,Luo Y F,Liu T,Hua Z X. 2022. The heterogeneity of lithofacies types,combination modes,and sedimentary model of lacustrine shale restricted by high-frequency sequence. Geological Journal,57(10): 1.

Wang D,Zhang Y N. 2024. Coupling of SME innovation and innovation in regional economic prosperity with machine learning and IoT technologies using XGBoost algorithm. Soft Computing,28(4): 2919－2939.

Xue C Q,McBeck J A,Lu H J,Yan C H,Zhong J H,Wu J G,Renard F. 2024. Classification of shale lithofacies with minimal data: application to the early Permian shales in the Ordos Basin,China. Journal of Asian Earth Sciences,259: 105901.

Zheng D Y,Hou M C,Chen A Q,Zhong H T,Qi Z,Ren Q,You J C,Wang H Y,Ma C. 2022. Application of machine learning in the identification of fluvial-lacustrine lithofacies from well logs: a case study from Sichuan Basin,China. Journal of Petroleum Science and Engineering,215: 110610.

Zhu X,Chu J,Wang K D,Wu S F,Yan W,Chiam K. 2021. Prediction of rockhead using a hybrid N-XGBoost machine learning framework. Journal of Rock Mechanics and Geotechnical Engineering,13(6): 1231－1245.

(责任编辑 李新坡; 英文审校 徐 杰)