

砂泥岩地层岩性智能识别方法与应用

蔡明 周庆文 杨聪 陈枫 伍东 林旺 章成广 张远君 苗雨欣

引用本文:

蔡明, 周庆文, 杨聪, 等. 砂泥岩地层岩性智能识别方法与应用[J]. 煤田地质与勘探, 2025, 53(1): 235–244.

CAI Ming, ZHOU Qingwen, YANG Cong, et al. An intelligent lithology identification method for sandstone and mudstone strata and its applications: A case study of the Jurassic strata in the Lunnan area, Xinjiang, China[J]. *Coal Geology & Exploration*, 2025, 53(1): 235–244.

在线阅读 View online: <https://dx.doi.org/10.12363/issn.1001-1986.24.07.0503>

您可能感兴趣的其他文章

Articles you may be interested in

人工智能测井：基础、原理、技术及应用

Artificial intelligence logging: Fundamental, principle, technique, and application

煤田地质与勘探. 2024, 52(8): 145–164 <https://doi.org/10.12363/issn.1001-1986.23.12.0813>

基于智能岩性识别的爆破岩体三维实体模型建立

Establishment of a 3D solid model blasting rock mass based on intelligent lithology identification

煤田地质与勘探. 2022, 50(4): 92–97 <https://doi.org/10.12363/issn.1001-1986.21.06.0340>

煤矿采空区覆岩“三带”智能识别方法

Intelligent identification method for overburden three zones of a goaf

煤田地质与勘探. 2024, 52(4): 164–171 <https://doi.org/10.12363/issn.1001-1986.23.08.0460>

基于机器学习的钻井液流变参数智能识别方法

Intelligent identification method of drilling fluid rheological parameters based on machine learning

煤田地质与勘探. 2024, 52(5): 183–192 <https://doi.org/10.12363/issn.1001-1986.24.01.0055>

基于钻孔多参数测井的煤层顶底板岩性及厚度识别

Identifying the lithologies and thicknesses of coal seam roofs and floors based on multiparameter logging of boreholes

煤田地质与勘探. 2024, 52(12): 190–200 <https://doi.org/10.12363/issn.1001-1986.24.03.0177>

锚固钻孔软弱夹层结构特征智能识别方法

A method for intelligent identification of weak interlayer characteristics during the drilling of anchor boreholes

煤田地质与勘探. 2024, 52(11): 200–210 <https://doi.org/10.12363/issn.1001-1986.24.04.0245>



移动阅读

蔡明, 周庆文, 杨聪, 等. 砂泥岩地层岩性智能识别方法与应用——以新疆轮南侏罗系岩层为例[J]. 煤田地质与勘探, 2025, 53(1): 235–244. doi: 10.12363/issn.1001-1986.24.07.0503

CAI Ming, ZHOU Qingwen, YANG Cong, et al. An intelligent lithology identification method for sandstone and mudstone strata and its applications: A case study of the Jurassic strata in the Lunnan area, Xinjiang, China[J]. Coal Geology & Exploration, 2025, 53(1): 235–244. doi: 10.12363/issn.1001-1986.24.07.0503

砂泥岩地层岩性智能识别方法与应用

——以新疆轮南侏罗系岩层为例

蔡 明^{1,2}, 周庆文^{1,2}, 杨 聪³, 陈 枫^{2,5}, 伍 东⁴, 林 旺⁴,
章成广^{1,2,*}, 张远君^{1,2}, 苗雨欣^{1,2}

(1. 油气资源与勘探技术教育部重点实验室(长江大学), 湖北 武汉 430100; 2. 长江大学 地球物理与石油资源学院, 湖北 武汉 430100; 3. 中钢集团武汉安全环保研究院有限公司, 湖北 武汉 430081; 4. 中国石油集团工程技术研究院有限公司信息中心, 北京 102206; 5. 安徽省煤田地质局第三勘探队, 安徽 宿州 234000)

摘要: 【目的】岩性识别是储层精细评价的基础, 传统方法一般仅用 2~3 种测井参数的交互关系进行岩性识别, 测井信息利用率低, 对于岩性测井响应差异小的地层岩性识别精度低, 严重制约了老井复查效果。高效的智能分类算法 CatBoost 可充分挖掘多源测井信息与岩性的关联。【方法】以新疆轮南地区侏罗系砂泥岩储层为研究对象, 通过敏感性分析选取自然伽马、自然电位、深浅电阻率比值、声波时差和密度 5 个测井参数, 构建基于 CatBoost 算法的岩性智能识别模型。利用优化的模型处理实际井资料以进行地层岩性识别, 通过准确率、精确率和召回率综合评估模型的岩性识别效果, 并对比分析了其与随机森林和 KNN 算法模型的识别效果。【结果和结论】结果表明: 轮南侏罗系大类岩性包括泥岩、砂岩和砾岩, 细分岩性复杂; 根据岩性敏感测井参数利用 CatBoost 算法建立的岩性智能预测模型对目标储层细分岩性的识别准确率达 92.64%, 显著高于随机森林模型的 82.95% 和 KNN 模型的 70.16%, 证明该方法能有效解决研究区的岩性识别问题。研究成果不仅为轮南地区老井复查和进一步勘探开发提供了科学依据, 还为复杂岩性精细识别方法研究提供重要参考。

关键词: 测井; 岩性识别; 人工智能; CatBoost; 梯度提升算法

中图分类号: TE257; P631 **文献标志码:** A **文章编号:** 1001-1986(2025)01-0235-10

An intelligent lithology identification method for sandstone and mudstone strata and its applications: A case study of the Jurassic strata in the Lunnan area, Xinjiang, China

CAI Ming^{1,2}, ZHOU Qingwen^{1,2}, YANG Cong³, CHEN Feng^{2,5}, WU Dong⁴, LIN Wang⁴,
ZHANG Chengguang^{1,2,*}, ZHANG Yuanjun^{1,2}, MIAO Yuxin^{1,2}

(1. Key Laboratory of Exploration Technologies for Oil and Gas Resources, Ministry of Education, Yangtze University, Wuhan 430100, China; 2. College of Geophysics and Petroleum Resources, Yangtze University, Wuhan 430100, China; 3. Wuhan Safety & Environmental Protection Research Institute Co., Ltd., Sinosteel, Wuhan 430081, China; 4. Information Center, CNPC Engineering Technology R&D Company Limited, Beijing 102206, China; 5. Anhui Provincial Coal Geological Exploration Bureau Third Team, Suzhou 234000, China)

Abstract: [Objective] Lithology identification lays the foundation for fine-scale reservoir evaluation. However, traditional identification methods generally utilize the interactive relationships between only 2–3 logging parameters, suffer-

收稿日期: 2024-07-31; 接收日期: 2024-12-19

基金项目: 国家自然科学基金项目(42104126, 42474177); 中国石油科技创新基金项目(2022DQ02-0301); 湖北省教育厅科学技术研究项目(Q20211309)

第一作者: 蔡明, 1986 年生, 男, 湖北随州人, 博士, 副教授, 硕士生导师。E-mail: caiming@yangtzeu.edu.cn

*通信作者: 章成广, 1963 年生, 男, 浙江宁海人, 博士, 教授, 博士生导师。E-mail: zhangcglog@163.com

© Editorial Office of Coal Geology & Exploration. OA under CC BY-NC-ND

ing from low utilization rates of logging information and low identification accuracy for strata with small differences in logging responses. This seriously restricts the effects of old well reexamination. The efficient, intelligent CatBoost classification algorithm can fully mine the correlations between multi-source logging information and lithology. [Methods] This study investigated the Jurassic sandstone and mudstone reservoirs in the Lunnan area, Xinjiang, China. Using five logging parameters determined through sensitivity analysis, i.e., natural gamma-ray value, spontaneous potential, deep and shallow resistivity ratio, sonic interval transit time, and density, this study developed an intelligent lithology identification model based on the CatBoost algorithm. The optimized model was employed to deal with actual logging data for lithology identification, and its performance was evaluated using accuracy, precision, and recall and was then compared with the lithology identification results of the random forest (RF) and k-nearest neighbors (KNN) algorithms. [Results and Conclusions] The results indicate that the large-scale lithologies of the Jurassic strata in the Lunnan area include mudstones, sandstones, and conglomerates, with complex fine-scale lithologies. In the identification of the fine-scale lithologies of the target reservoir, the intelligent lithology identification model, established using the CatBoost algorithm and lithology-sensitive logging parameters, yielded an accuracy of 92.64%, significantly higher than that of the random forest model (82.95%) and the KNN model (70.16%). This result demonstrates that the CatBoost model can effectively address of the challenges of lithology identification in the study area. The results of this study will provide a scientific basis for the review and further exploration and development of old wells in the Lunnan area. Besides, these results can serve as a valuable reference for research on methods for fine-scale identification of complex lithologies.

Keywords: logging; lithology identification; artificial intelligence (AI); CatBoost; gradient boosting algorithm

随着油气勘探开发的不断发展,具有良好物性的油气藏逐渐减少且油气产量下降^[1-2],许多老油田将目光转移到已经开发过的老井上,期望寻找出新的潜力油气层。潜力层的寻找依赖于测井综合解释,而岩性的准确识别是测井综合解释中不可缺少的一环^[3-5]。因此,如何从测井数据中准确地获取岩性数据逐渐成为了测井领域的研究热点。传统的岩性识别方法主要包括测井交会图版法和成像测井方法^[6-12]。交会图版法是通过分析不同岩性之间的测井响应特征差异从而建立交会图来进行岩性划分,但该方法难以充分利用测井曲线包含的岩性信息,对于非常规油气藏,不同岩性成分地层对应的测井响应相似,很难准确实现岩性分类^[13]。例如,新疆轮南地区侏罗系层位地层主要岩性包括泥岩、粉砂质泥岩、粉砂岩、泥质粉砂岩、细砂岩等,这些岩性地层的自然伽马、声波时差和电性测井曲线具有一定的相似性,用交会图方法无法准确地确定各种岩性的测井响应参数上下限。通过成像测井资料识别岩性虽然比较直观且分辨率高,但成像测井成本昂贵,且大部分老井只有常规测井资料,所以无法依靠成像资料完成研究区的岩性岩相划分工作^[14]。

近些年机器学习和人工智能技术得到了飞速发展,研究人员开始将这些方法引入地质领域,期望充分挖掘常规测井资料中隐藏的信息以改善岩性识别的精度和效率^[15-19]。在机器学习方法中,常用的随机森林算法训练速度快、准确率高、抗过拟合能力较强且在训练结束后可以比较各输入特征的重要性,但对于低维数据(测井数据)分类准确率并不理想,容易出现很多相似的决策树,掩盖真实结果,且对于模型参数也需要通过多次尝试才能找到最优解,无法控制模型的内部运行,可解释性较差^[20-21]。K 最邻近值(K-nearest neighbor, KNN)

算法理论成熟,思想简单,当训练样本量大时准确率非常高,但模型的建立需要大量的内存,且在岩性识别训练过程中对于数据量较少的样本预测准确率^[22-23]。基于梯度提升的机器学习算法可以很好地解决数据分布不均衡的问题,即使数据集中存在缺失值和异常值对训练模型的影响也不大,因此,该类算法成为了岩性识别研究的焦点^[24-27];其中 CatBoost 算法是一种基于对称决策树的梯度提升算法,相比于传统机器学习算法和其他梯度提升算法而言,其具有如下明显的优势:有更高的准确性和泛化能力;能够高效合理地处理类别型特征、梯度偏差和预测偏移等问题;其决策树结构使模型能够可视化,可根据输入层数据特征的重要性增强模型的可解释性并显著降低过拟合发生的概率^[28-30]。

目前 CatBoost 算法通常用于重大疾病预测和市场营销分析,这 2 个领域内的数据通常异常值较多且数据分布不均衡,而 CatBoost 模型可以自动处理类别型特征,无需独热编码,且对于异常值和噪声数据具有鲁棒性,超参数的调节也较为方便^[31-32]。轮南地区的岩性分布也十分不均衡,且存在扩径等导致的异常值。因此,以该区常规测井曲线数据和取心资料为基础,对测井参数进行敏感性分析,并利用过采样方法增加样本使数据分布均衡,进而建立了基于 CatBoost 算法的岩性识别模型,通过准确率、精确率和召回率 3 个指标来评估模型的识别效果,以提升识别精度,为地层岩性准确高效识别提供新的思路。

1 CatBoost 算法

CatBoost 算法属于 Boosting 算法系列,是一种基于梯度提升决策树(gradient boosting decision trees, GB-

DT) 的新型机器学习算法框架^[33]。GBDT 算法是一种用于回归和分类的算法, 能够避免由于内部集成多个决策树并累积多个决策树而导致单个决策树产生过度拟合的问题。GBDT 算法在每个迭代步构建沿梯度最陡方向降低损失的学习器, 以弥补当前已构建模型的不足, 即利用梯度下降法进行优化^[34]。但由于 GBDT 算法中弱学习器之间存在依赖关系, 难以并行训练数据; 同时, 算法缺乏平滑性, 学习得到的模型仍可能产生过拟合问题。CatBoost 算法由 L. Prokhorenkova 等于 2017 年首次提出^[28], 使用对称树作为基学习器解决了传统 GBDT 算法的过拟合问题, 其采用 Ordered Boosting 方法获得梯度的无偏估计, 以减轻梯度估计偏差的影响, 从而提高模型的泛化能力。使用伪码描述排序提升算法, 见表 1, CatBoost 算法对每一个样本 x_i 都会训练生成一个单独的模型 M_i , 模型 M_i 是由不包含样本 x_i 的训练集训练得到的, 使用 M_i 来得到关于样本的梯度估计。

表 1 排序提升算法^[28]
Table 1 Sort boosting algorithm^[28]

```

输入:  $\{(x_k, y_k)\}_{k=1}^n, I$ ;
 $\sigma \leftarrow$  random permutation of  $[1, n]$ ;
 $M_i \leftarrow 0$  for  $i = 1, 2, \dots, n$ ;
for  $t \leftarrow 1$  to  $I$  do
    for  $i \leftarrow 1$  to  $n$  do
         $r_i \leftarrow y_i - M_{\sigma(i)-1}(x_i)$ ;
        for  $j \leftarrow 1$  to  $n$  do
             $\Delta M \leftarrow \text{LearnModel}((x_j, r_j) : \sigma(j) \leq i)$ ;
             $M_i \leftarrow M_i + \Delta M$ ;
return  $M_n$ 

```

同时 CatBoost 还具有自适应学习率, 自适应学习率可以帮助算法更好地控制每一轮迭代中弱学习器的贡献, 从而提高整个模型的精度。自适应学习率的计算方法如下:

$$\eta_t = \frac{1}{\sqrt{t+1}} \quad (1)$$

$$\alpha_t = \frac{\sum_{i=1}^t \eta_i}{t} \quad (2)$$

上述分析表明, CatBoost 算法不需要过多训练, 即可获得较高的模型精度, 并且不需要对超参数进行过多的调优, 降低了过拟合发生的概率。

2 岩性测井响应特征分析

2.1 测井曲线标准化与岩心深度归位

在油气勘探中, 不同的测井仪器、不同的测量条件以及不同的数据处理方法都可能导致数据之间存在明显差异。这些差异可能导致地层识别、储层评价等工作的准确性下降。通过标准化处理, 可以消除这些差异,

使数据更加准确可靠, 从而提高后续处理的精度。常用的测井曲线标准化一般有频率直方图法、均值校正法和趋势面分析法。本文采用趋势面法对研究区的测井曲线进行标准化处理。将多口井的标准层测井数据与井的空间位置相结合, 利用最小二乘法拟合出趋势面。趋势面能在数据没有局部变化的情况下, 反映测井数据的变化规律。

本文统计了研究区内 58 口关键井测井曲线数据, 并选取侏罗系四油组地层 JIV 砂岩中最为稳定的一段约 10 m 厚地层为标准层。

以声波时差标准化为例, 读取每口井相应标准层的声波时差中值来进行趋势面的拟合, 图 1 为声波时差拟合出的趋势面, 东北方向声波时差为低值, 东南方向声波时差为高值。声波时差趋势面拟合公式为:

$$\Delta DT = 0.0022X + 0.7659Y + 8.29 \times 10^{-9}X^2 - 5.56 \times 10^{-8}XY + 8.95 \times 10^{-9}Y^2 - 1.775 \times 10^6 \quad (3)$$

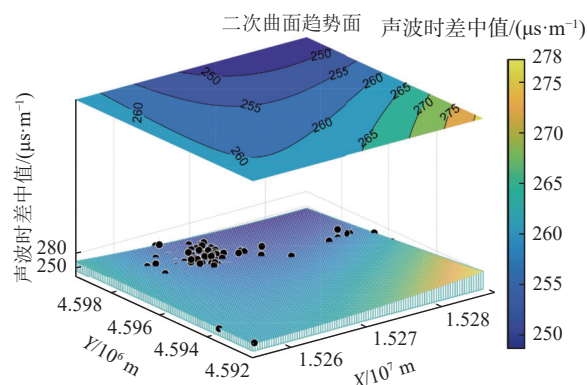


图 1 DT 值趋势面法拟合结果

Fig. 1 The fitting result of DT by trend surface method

通过岩样密度与密度测井曲线相关对比法进行岩心深度归位, 目标地区 Y1 井岩心深度归位结果如图 2 所示。图 2 中第 5 道为岩心原深度范围: 4 554.80~4 567.70 m, 第 6 道为校正后深度范围: 4 557.11~4 570.01 m, 校正量: +2.311 m。

2.2 研究区岩性类型分析

轮南地区位于塔克拉玛干沙漠北部边缘, 构造上位于塔北隆起中部偏东, 是塔北隆起的一个二级构造单元, 称为轮南低凸起。该区侏罗系为河—湖沉积体系, 该层位自下而上分为 JIV、JIII、JII、JI 共 4 个油层组, 下统对应于 JIV 油层组, 为辫状河三角洲沉积; 中统对应于 JIII 和 JII 油层组; 上统对应于 JI 油层组, 为滨浅湖沉积^[35]。该区砂岩储层虽处于有利沉积相带, 但成岩作用类型较多, 对储层物性的影响较为复杂。JI、JII、JIII 油层组主要以泥岩、粉砂岩、细砂岩为主, 而 JIV 油层组岩性分布较为复杂, 粉、细、中、粗、不等粒砂岩、砾岩和砂砾岩等均有出现, 给测井岩性分类带来了困难。

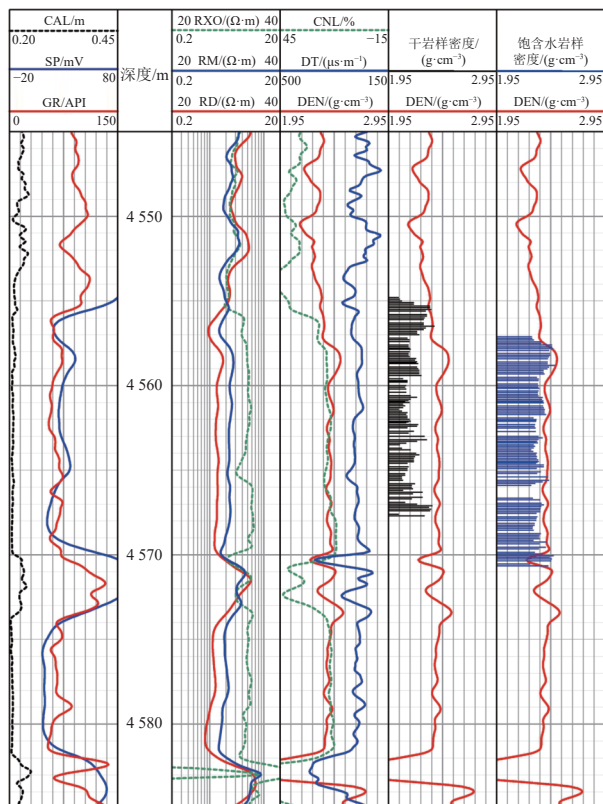


图 2 Y1 井岩心深度归位成果

Fig.2 Core depth calibration result of well Y1

本次研究共统计了 4 口井的取心资料,将其导入测井资料处理专业软件 CIFLog 中并设置采样间隔为 0.100 m,统计每种岩性的样本数量,结果见表 2;其中,粉砂岩数量最多,其次是泥岩,中砂岩数据量较少,将其和粗砂岩归为一类称为中粗砂岩,砂砾岩、角砾岩和细砾岩归为一类称为砾岩。

表 2 目的层位岩性类型统计结果
Table 2 Statistics of lithology types at the target horizon

岩石分类	样本数量	岩性编号
泥岩	泥岩	548
	粉砂质泥岩	176
砂岩	粉砂岩	1 921
	泥质粉砂岩	207
	细砂岩	224
	中砂岩	44
砾岩	粗砂岩	171
	砂砾岩	80
	角砾岩、细砾岩	82

2.3 岩性测井响应特征分析

通过对取心和测井资料的分析,选取了对岩性较为敏感的自然伽马(GR)、自然电位(SP)、深浅电阻率比值(RD/RXO)、声波时差(DT)、密度(DEN)共 5 种测井响应参数。在多井资料分析中为消除不同测量工具的

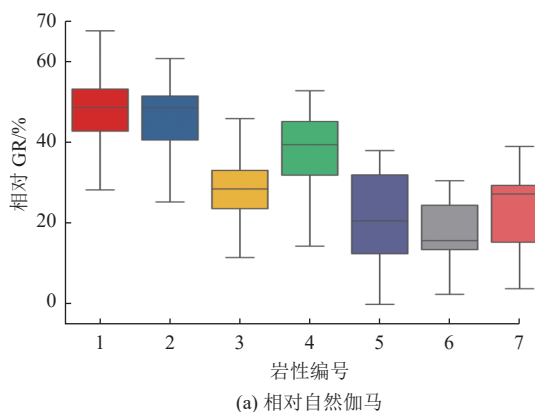
系统误差,本文对 GR 和 SP 曲线采取了单井归一化使其不同井的 GR 和 SP 曲线具有统一的标准。标准化公式如下:

$$\Delta SP = \frac{SP - SP_{\min}}{SP_{\max} - SP_{\min}} \times 100\% \quad (4)$$

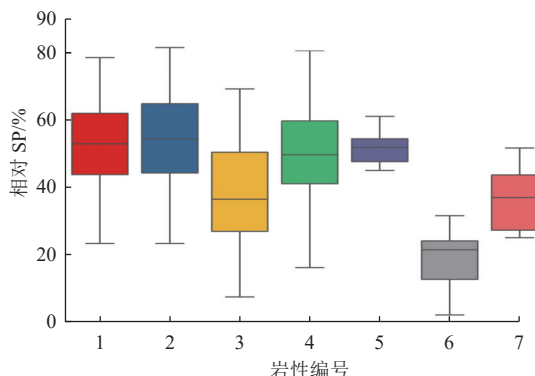
$$\Delta GR = \frac{GR - GR_{\min}}{GR_{\max} - GR_{\min}} \times 100\% \quad (5)$$

将 4 口取心井的岩性数据合并,对泥岩、粉砂质泥岩、粉砂岩、泥质粉砂岩、细砂岩、中粗砂岩和砾岩 7 种岩性(编号依次为 1~7)进行箱型图分析,结果如图 3 所示。由图 3a 中可知,含有泥质的岩性相对自然伽马值都较高,且泥质含量越高的岩性相对自然伽马值越高,而非泥质岩性的相对自然伽马较低且有一定的重合。而细砂岩的相对 SP 和密度(图 3b 和图 3d)明显高于中粗砂岩和砾岩,且砾岩的值大于中粗砂岩。在图 3c 中砾岩的深浅电阻率比值(RD/RXO)最低,而泥岩、粉砂质泥岩和泥质粉砂岩的值有很大部分重叠且高于其他岩性,粉砂岩、细砂岩和中粗砂岩的值也大致相同。总体来说,不同岩性的各测井响应特征虽然有差异,但存在一定程度上的重叠,尤其是泥岩、粉砂质泥岩、粉砂岩和泥质粉砂岩很难直接通过箱型图来直接区分。

由图 3 可知,由于数据中存在异常值导致个别岩性的测井参数响应区间较大,甚至超出了正常值,这对于岩性识别模型的准确率明显不利。为了进一步统计有



(a) 相对自然伽马



(b) 相对自然电位

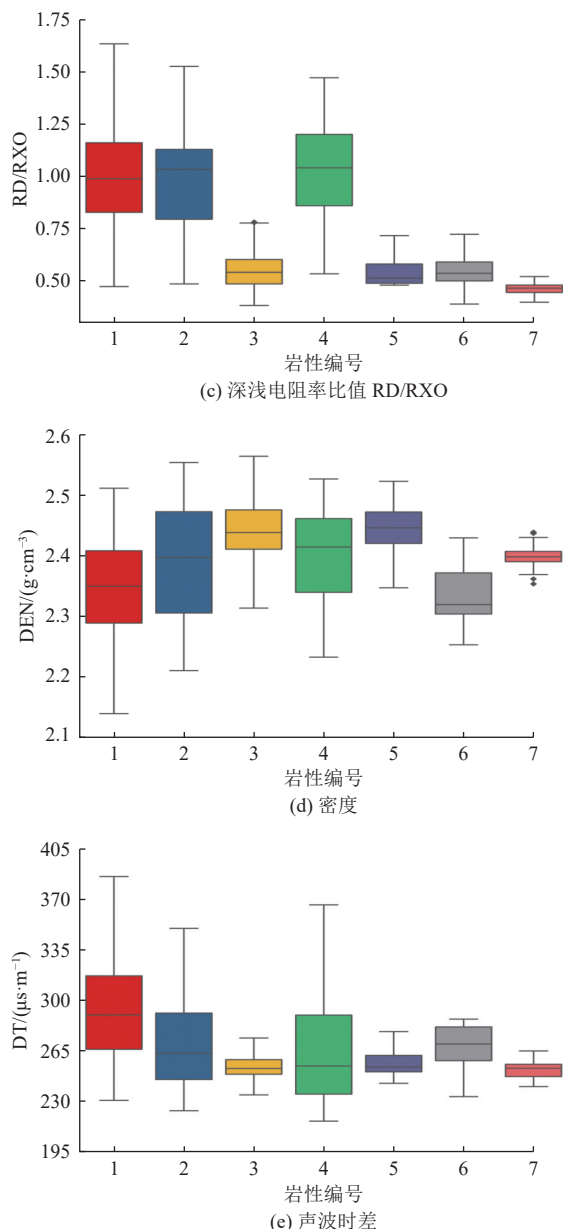


图 3 轮南地区侏罗系取心岩性及测井参数箱型图

Fig.3 Box diagrams showing the lithologies and logging parameters of the cores from Jurassic strata in the Lunnan area

价值的测井响应参数,将箱型图第 85 百分位和第 15 百分位作为实际岩性响应范围的上下限,见表 3。

表 3 轮南地区侏罗系取心井不同岩性地层的测井响应特征(15%~85%)

Table 3 Logging responses of varying lithologies of Jurassic strata in cored wells in the Lunnan area (15%~85%)

岩性	相对GR/%	相对SP/%	RD/RXO	DEN/(g·cm ⁻³)	DT/(μs·m ⁻¹)
泥岩	38.9~54.8	39.7~66.4	0.719~1.286	2.268~2.436	256.138~323.829
粉砂质泥岩	37.1~53.0	36.0~76.3	0.730~1.206	2.283~2.508	238.757~300.568
粉砂岩	20.7~35.1	19.1~55.6	0.458~0.637	2.392~2.494	245.597~260.692
泥质粉砂岩	27.8~47.7	34.5~76.1	0.744~1.223	2.313~2.471	230.233~294.032
细砂岩	6.9~35.0	46.7~57.3	0.483~0.608	2.387~2.484	246.010~265.669
中粗砂岩	12.4~26.3	11.3~25.3	0.483~0.639	2.298~2.380	249.429~280.194
砾岩	7.2~30.7	26.0~44.3	0.435~0.487	2.372~2.415	244.268~255.932

3 岩性预测模型建立与实例分析

3.1 数据预处理

本文研究的数据来自轮南地区 4 口取心井的测井曲线和取心数据,在建立岩性识别模型前需要对测井参数和取心数据进行预处理^[20]。

(1) 异常值处理。根据表 2 所划定的不同岩性测井响应特征对数据进行筛选。

(2) 数据集划分。将其中 3 口井按 8 : 2 随机划分为训练集和测试集,留一口井作为验证集。

(3) 数据标准化。由于不同测井特征值量纲和单位都不同,需要对数据进行标准化。本文使用 Z-Score 标准化可以将不同量级的数据转换为同一量级,先处理训练集数据并保留其均值和标准差,再用训练集的均值和标准差对测试集进行标准化。

$$x' = \frac{x - \bar{x}}{\sigma} \quad (6)$$

(4) 数据过采样。由于泥岩和粉砂岩的数据量比其他 2 种岩性的数据量大很多,岩性识别模型可能会对数据量较少的类识别效果不佳。针对这一问题,本文采用 SMOTE (synthetic minority over-sampling technique) 算法对训练集数据进行过采样使数据分布均衡^[20]。SMOTE 算法公式如下:

$$x_{\text{new}} = x_i + \text{rand}(0, 1) \times (\hat{x}_i - x_i) \quad (7)$$

3.2 基于 CatBoost 算法的岩性智能预测模型建立

基于 CatBoost 算法的岩性智能识别模型建立流程如图 4 所示。利用 SMOTE 算法对训练样本进行过采样处理使少数类岩性样本与多数类岩性样本数量相同;将新得到的训练样本集输入建立的 CatBoost 模型中进行训练,并利用网格搜索和交叉验证(Grid SearchCV)方法进行模型调优,优选对模型影响最大的 3 个参数进行搜索:树深范围(1, 10),步长为 1;学习率(0.01、0.03、0.06、0.10、0.15);正则系数(1、4、9),通过网格遍历后确定 CatBoost 算法的最优参数组合:树的深度、学习率、正则系数分别取 4、0.15 和 1,此时基于 CatBoost 算法

的岩性预测模型最优,对应的交叉验证准确率为 97.87%。同时该模型还给出了各特征参数的权重,见表 4。

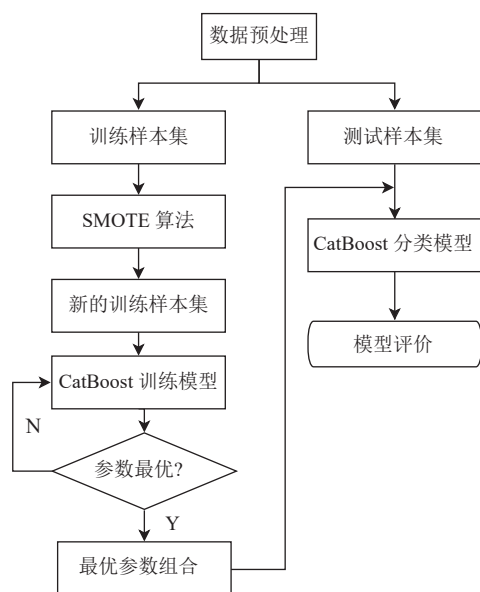


图 4 岩性智能识别流程

Fig.4 Flow chart of intelligent lithology identification

表 4 CatBoost 模型的特征参数权重
Table 4 Weights of the characteristic parameter of the CatBoost model

特征参数	权重/%
相对GR	13.25
相对SP	21.52
RD/RXO	23.19
DEN	20.02
DT	22.02

3.3 模型测试结果与对比分析

利用上述优选的参数和训练集对 CatBoost 算法进行训练学习,并且与随机森林算法和 KNN 算法在测试集上进行分类预测对比,从而评估模型的岩性识别效果。图 5 为轮南地区不同模型岩性识别混淆矩阵。表 5 为 CatBoost、随机森林、KNN 这 3 种机器学习算法在测试集上的岩性识别效果,随机森林算法和 KNN 算法对岩性识别的准确率均低于 90%,而 CatBoost 算法的岩性识别准确率为 92.12%,岩性识别效果最好。另外, CatBoost 模型对于全部岩性的精确率和召回率都很高,说明该模型对于轮南地区侏罗系层位的岩性识别效果很好;但粉砂质泥岩的召回率较低,表明该模型还有进一步改善的空间。

由表 5 可以看出,随机森林模型中泥质粉砂岩和粉砂质泥岩的精确率和召回率都低于 80%,由随机森林岩性识别混淆矩阵(图 5b)可以看出这 2 种岩性主要是

被误判为泥岩和粉砂岩;该模型对于细砂岩的召回率很低,有 36.36% 的细砂岩被误判为粉砂岩,中粗砂岩中也有 17.24% 被识别为粉砂岩,说明随机森林岩性识别模型只对于粉砂岩和砾岩有很好的区分效果。

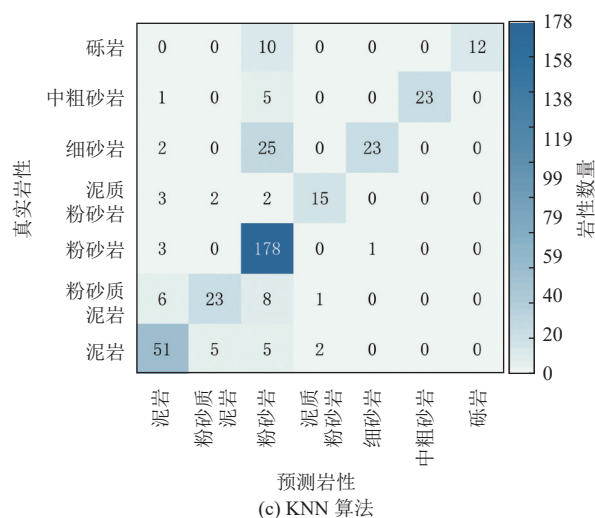
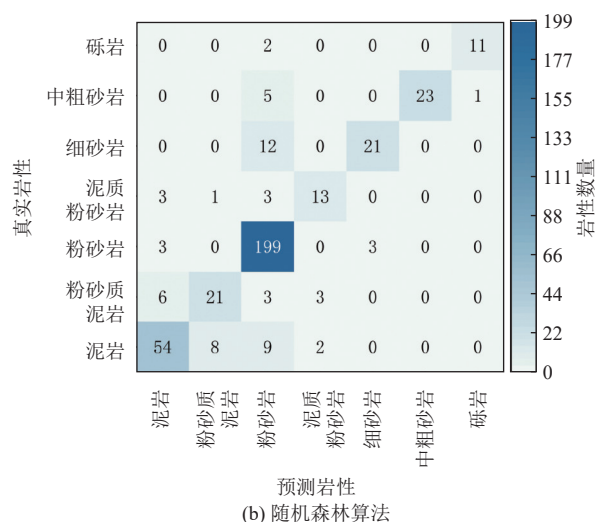
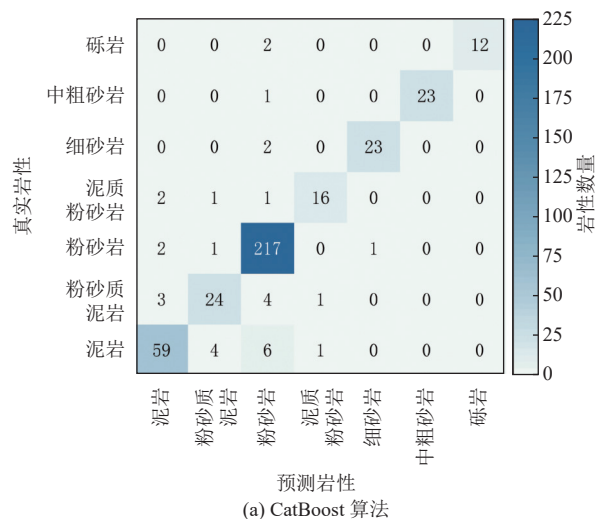


图 5 轮南地区不同模型岩性识别混淆矩阵

Fig.5 Confusion matrices of varying models for lithology identification in the Lunnan area

表 5 轮南地区取心井不同方法的岩性识别准确率、召回率和精确率
Table 5 Accuracy, recall, and precision of different methods for lithology identification of cored wells in the Lunnan area %

岩性	CatBoost		随机森林		KNN	
	精确率	召回率	精确率	召回率	精确率	召回率
泥岩	89.39	84.29	81.82	73.97	77.27	80.95
粉砂质泥岩	80.00	75.00	70.00	63.64	76.67	60.53
粉砂岩	93.13	98.19	85.41	97.07	76.39	97.80
泥质粉砂岩	88.89	80.00	72.22	65.00	83.33	68.18
细砂岩	95.83	92.00	87.50	63.64	95.83	46.00
中粗砂岩	100.00	95.83	100.00	79.31	100.00	79.31
砾岩	100.00	85.71	91.67	84.62	100.00	54.55
准确率	92.12		84.09		80.05	

KNN 岩性识别模型对于粉砂质泥岩和泥质粉砂岩的识别准确率同样不高(表 5), 且 KNN 模型对于粉砂岩和泥岩精确率都低于 80%, 说明该模型不能很好地区分泥岩、粉砂质泥岩、粉砂岩、泥质粉砂岩 4 种岩性。另外, KNN 岩性识别模型对细砂岩、中粗砂岩、砾岩的召回率都很低, 结合图 5c(KNN 岩性识别模型混淆矩阵)可以看出, 这 3 种岩性主要被误判为粉砂岩。

3.4 应用实例分析

利用上述建立的 3 种模型分别对轮南地区 X1 井 4 455~4 495 m 井段进行岩性识别, 结果如图 6 所示, 其中第 1 道为 GR 和 SP 曲线, 第 3 道为深、中、浅电阻率曲线, 第 4 道为 DT 和 DEN 曲线, 第 5—第 8 道分别为取心岩性、CatBoost 模型预测岩性、随机森林模型预测岩性和 KNN 模型预测岩性。由图 6 可知: X1 井取心层段由上往下第一段 4 456.45~4 457.66 m 为泥岩, CatBoost、随机森林、KNN 都识别为泥岩; 4 457.66~4 458.09 m 为粉砂质泥岩, 随机森林和 KNN 识别为泥岩, 只有 CatBoost 识别为粉砂质泥岩; 4 458.09~4 459.07 m 是粉砂岩, CatBoost 识别为粉砂岩而随机森林和 KNN 则识别为顶端小段泥岩, 底端为泥质粉砂岩; 4 459.08~4 459.49 m 为粉砂质泥岩, 3 种岩性识别模型均判断正确; 4 460.20~4 460.77 m 为粉砂质泥岩, CatBoost 和 KNN 识别正确, 只有随机森林误判为粉砂岩; 4 460.79~4 464.61 m 为粉砂岩, CatBoost 模型识别最为准确, KNN 和随机森林模型识别该段时都夹杂了其他岩性效果不理想; 4 469.96~4 472.73 m 为细砂岩, 3 种算法都准确识别, 但从测井曲线来看 4 468~4 472.73 m 都为细砂岩, 而随机森林和 KNN 算法将该段顶部都识别为了粉砂岩; 在 4 472.73~4 477.73 m 井段, CatBoost 和随机森林的预测岩性与取心岩性相同, 而 KNN 算法则是将粉砂岩预测成了泥岩和粉砂质泥岩; 4 485.31~4 485.93 m 井段为粉砂岩, KNN 和随机森林模型识别为上部泥岩, 底端为粉砂质泥岩, CatBoost 算法识别为

粉砂岩; 4 486.85~4 491.02 m 井段为粉砂岩, CatBoost 岩性识别模型预测准确, 而随机森林岩性识别模型和 KNN 岩性识别模型则将上部识别为泥质粉砂岩。整体而言, CatBoost、随机森林、KNN 3 种模型岩性识别准确率分别为 92.64%、82.95%、70.16%。

上述分析表明, 在岩性识别结果中, 随机森林和 KNN 岩性识别模型对粉砂岩的识别效果不好, 容易将其识别为粉砂质泥岩和泥质粉砂岩。CatBoost 模型预测的岩性与取心岩性的符合率最高, 但在岩性的过渡带岩性识别精度相对较低。

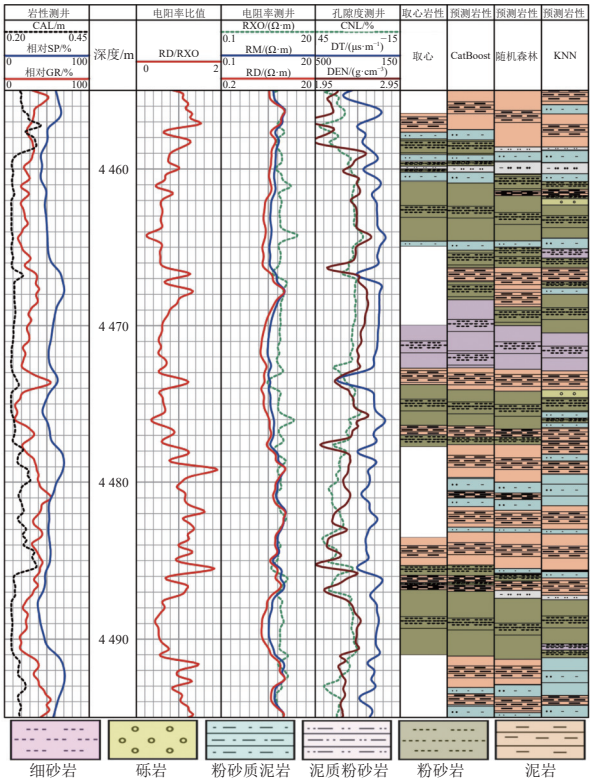


图 6 轮南地区 X1 井(4 455~4 495 m)不同算法岩性识别结果

Fig.6 Lithology identification results of well X1 (4 455~4 495 m) in the Lunnan area derived using by different algorithms

4 结论

(1) 岩心分析表明, 轮南地区侏罗系细分岩性主要包括泥岩、粉砂质泥岩、粉砂岩、泥质粉砂岩、细砂岩、中粗砂岩和砾岩。这些岩性测井响应差异小, 难以通过传统的方法来精确地划分。

(2) 通过敏感性分析提炼了岩性敏感测井参数, 构建了基于岩性敏感参数和高效分类算法 CatBoost 的岩性智能识别模型, 并应用于新井目标储层岩性识别。结果表明: 新建的 CatBoost 模型细分岩性识别准确率达到 92.64%, 显著高于随机森林模型的 82.95% 和 KNN 模型的 70.16%, 证实了该岩性智能识别模型的适用性与智能模型构建方案的可行性。

(3) 构建的 CatBoost 岩性智能识别模型对于粉砂岩、细砂岩、中粗砂岩和砾岩的识别效果相对最好, 准确率高达 93% 以上; 粉砂岩和细砂岩储层是该地区的主要含油储层; 因此, 可以在分区域分层构建孔渗饱和模型的基础上再根据岩性精细识别结果进一步细化模型, 进而提高储层有效性和含油性评价结果的精度, 为优化老井复查开发方案及提高采收率奠定基础。

(4) 3 种算法模型对于泥质粉砂岩和粉砂质泥岩的预测精度均较低, 常与粉砂岩混淆。后续可考虑先通过 GR、SP 和 CAL 曲线进行泥岩和砂岩的大类划分, 再通过 DT、DEN 和 RD/RXO 等曲线进行细分类别划分, 以进一步提高岩性划分的精度。

致谢: 感谢中国石油集团有限公司给予测试数据支持以及匿名审稿人提供建设性修改意见。

符号注释:

CAL 为井径, m; CNL 为补偿中子测井, %; DEN 为密度测井, g/cm^3 ; DT 为声波时差, $\mu\text{s}/\text{m}$; ΔDT 为经过校正后的值, $\mu\text{s}/\text{m}$; GR_{\min} 为纯砂岩段的 GR 值, API; GR_{\max} 为纯泥岩段的 GR 值, API; ΔGR 为相对 GR 值, %; RD、RM、RXO 分别为深、中、浅电阻率测井, $\Omega \cdot \text{m}$; I 为树的数量; j 为样本序号; k 为样本序号; 下标 max、min 分别为相应测井数值的最大值、最小值; M_0 为初始化模型; M_i 为样本对应的单个模型; M_n 为最终输出模型; $M_{\sigma(i)-1}$ 为第 $\sigma(i)-1$ 个模型; ΔM 为模型的增量更新值; n 为样本总数; SP_{\min} 为纯泥岩段的 SP 值, mV; SP_{\max} 为纯砂岩段的 SP 值, mV; ΔSP 为相对 SP 值, %; t 为迭代次数; x 为原始数据; x_i 为原始样本; x_{new} 为新样本点; x' 为标准化后的数据; \bar{x} 为均值; \hat{x}_i 为 x_i 的随机邻近点; X 为井位横坐标, m; X_i 为样本特征值; Y 为井位纵坐标, m; y_i 为 $M_{\sigma(i)-1}(X_i)$ 得出的值, 即残差值, 用 r_i 表示; $\{(x_k, y_k)\}_{k=1}^n$ 为训练样本, 按照顺序排列为 $(x_{\sigma 1}, y_{\sigma 1}), (x_{\sigma 2}, y_{\sigma 2}), \dots$,

$(x_{\sigma n}, y_{\sigma n})$; α_t 为前 t 轮迭代的平均学习率; η_t 为第 t 轮迭代的学习率; σ 为标准差。

利益冲突声明/Conflict of Interests

所有作者声明不存在利益冲突。

All authors disclose no relevant conflict of interests.

参考文献(References)

- [1] 蔡明, 章成广, 唐军, 等. 参数估计法声波远探测反射波提取效果影响因素研究[J]. 西安石油大学学报(自然科学版), 2020, 35(1): 42–48.
CAI Ming, ZHANG Chengguang, TANG Jun, et al. Study on factors of influencing extraction effect of reflection wave in acoustic remote detection using parameter estimation method[J]. Journal of Xi'an Shiyou University (Natural Science Edition), 2020, 35(1): 42–48.
- [2] 蔡明, 章成广, 韩闯, 等. 微裂缝对横波衰减影响的实验研究及其在致密砂岩裂缝评价中的应用[J]. 中国石油大学学报(自然科学版), 2020, 44(1): 45–52.
CAI Ming, ZHANG Chengguang, HAN Chuang, et al. Experimental research of effect of microfracture on shear wave attenuation and its application on fracture evaluation in tight sand formation[J]. Journal of China University of Petroleum (Edition of Natural Science), 2020, 44(1): 45–52.
- [3] 雷明, 陈涛, 韩乾凤, 等. 一种基于常规测井资料计算碳酸盐岩储层裂缝孔隙度新方法[J]. 地球科学, 2023, 48(7): 2678–2689.
LEI Ming, CHEN Tao, HAN Qianfeng, et al. A new method for calculating fracture porosity based on conventional logging data[J]. Earth Science, 2023, 48(7): 2678–2689.
- [4] 胡玉双, 王志勇, 李金宝. 塔南凹陷南二段低渗透砂岩有利储层预测研究[J]. 能源与环保, 2023, 45(3): 7–12.
HU Yushuang, WANG Zhiyong, LI Jinbao. Study on prediction of favorable reservoirs of low permeability sandstone in South 2nd Section of Tanan Sag[J]. China Energy and Environmental Protection, 2023, 45(3): 7–12.
- [5] 李晓峰, 王黎, 李剑平, 等. 基于岩性识别的渗透率计算模型[J]. 测井技术, 2022, 46(1): 84–88.
LI Xiaofeng, WANG Li, LI Jianping, et al. Permeability calculation model based on lithology identification[J]. Well Logging Technology, 2022, 46(1): 84–88.
- [6] 薛桂玉, 菅红军, 赵永忠, 等. 吉木萨尔凹陷芦草沟组页岩油储层岩性识别方法[J]. 测井技术, 2021, 45(6): 636–641.
XUE Guiyu, JIAN Hongjun, ZHAO Yongzhong, et al. Lithology identification method for the shale oil reservoir of Lucaogou Formation in Jimusar Depression[J]. Well Logging Technology, 2021, 45(6): 636–641.
- [7] 罗兴平, 庞旭, 苏东旭, 等. 电成像测井在复杂砂砾岩储集层岩性识别中的应用: 以准噶尔盆地玛湖凹陷西斜坡百口泉组为例[J]. 新疆石油地质, 2018, 39(3): 345–351.
LUO Xingping, PANG Xu, SU Dongxu, et al. Recognition of complicated sandy conglomerate reservoir based on micro-resistivity imaging logging: A case study of Baikouquan Formation in

- western slope of Mahu Sag, Junggar Basin[J]. *Xinjiang Petroleum Geology*, 2018, 39(3): 345–351.
- [8] 周海超, 付广, 王艳, 等. 测井资料交会图法在碎屑岩岩性识别中的应用: 以十屋断陷为例[J]. *大庆石油地质与开发*, 2009, 28(1): 136–138.
- ZHOU Haichao, FU Guang, WANG Yan, et al. Application of the crossplot method of well logging data in the lithology recognition of clastic rock: Taking Shiwu Fault Depression as example[J]. *Petroleum Geology & Oilfield Development in Daqing*, 2009, 28(1): 136–138.
- [9] 陈龙, 魏五峰. 松辽盆地滨北地区变质岩岩性测井识别方法[J]. *测井技术*, 2019, 43(4): 423–426.
- CHEN Long, WEI Wufeng. Log identification methods of metamorphic lithology in Binbei area, Songliao Basin[J]. *Well Logging Technology*, 2019, 43(4): 423–426.
- [10] 王楠. 乍得 Bongor 盆地潜山岩性识别及储层测井综合评价[J]. *科学技术与工程*, 2019, 19(27): 66–73.
- WANG Nan. Lithology identification and well-logging comprehensive evaluation of basement reservoir in Bongor Basin, Chad[J]. *Science Technology and Engineering*, 2019, 19(27): 66–73.
- [11] 王霄宇, 谢然红, 毛治国, 等. 主成分分析法在致密砂岩岩性识别的应用研究[J]. *天然气与石油*, 2021, 39(1): 88–93.
- WANG Xiaoyu, XIE Ranhong, MAO Zhiguo, et al. Application research of principal component analysis method in lithology identification of tight sandstone[J]. *Natural Gas and Oil*, 2021, 39(1): 88–93.
- [12] 俞初安, 司马献章, 李建国, 等. 鄂尔多斯盆地直罗组地层岩性测井响应特征[J]. *煤田地质与勘探*, 2018, 46(6): 33–39.
- YU Reng'an, SIMA Xianzhang, LI Jianguo, et al. Response characteristics of lithologic logging of Zhiluo Formation in Ordos Basin[J]. *Coal Geology & Exploration*, 2018, 46(6): 33–39.
- [13] QI Tingting, ZHAO Hui, FU Jia. Research on the reservoir characteristics and lithology recognition based on asphalt sandstone[J]. *IOP Conference Series Earth and Environmental Science*, 2019, 237(3): 032044.
- [14] 李国欣, 赵太平, 石玉江, 等. 鄂尔多斯盆地马家沟组碳酸盐岩储层成岩相测井识别评价[J]. *石油学报*, 2018, 39(10): 1141–1154.
- LI Guoxin, ZHAO Taiping, SHI Yujiang, et al. Diagenetic facies logging recognition and evaluation of carbonate reservoirs in Majiagou Formation, Ordos Basin[J]. *Acta Petrolei Sinica*, 2018, 39(10): 1141–1154.
- [15] 程希, 周军, 傅海成, 等. 机器学习算法在地球物理测井中的适用性及应用[J]. *西北地质*, 2023, 56(4): 336–348.
- CHENG Xi, ZHOU Jun, FU Haicheng, et al. Applicability and application of machine learning algorithm in logging interpretation[J]. *Northwestern Geology*, 2023, 56(4): 336–348.
- [16] 许振浩, 马文, 李术才, 等. 岩性识别: 方法、现状及智能化发展趋势[J]. *地质论评*, 2022, 68(6): 2290–2304.
- XU Zhenhao, MA Wen, LI Shucai, et al. Lithology identification: Method, research status and intelligent development trend[J]. *Geological Review*, 2022, 68(6): 2290–2304.
- [17] GONG Renbin, ZHOU Jingrun. Research on application of deep learning in lithology recognition of oil and gas reservoir[C]//2021 IEEE International Conference on Power, Intelligent Computing and Systems (ICPICS). Shenyang, China. IEEE, 2021: 110–115.
- [18] 刘明军, 李恒堂, 姜在炳. GA-BP 神经网络模型在彬长矿区测井岩性识别中的应用[J]. *煤田地质与勘探*, 2011, 39(4): 8–12.
- LIU Mingjun, LI Hengtang, JIANG Zaibing. Application of genetic-BP neural network model in lithology identification by logging data in Binchang mining area[J]. *Coal Geology & Exploration*, 2011, 39(4): 8–12.
- [19] 陈刚, 汪凯斌, 蒋必辞, 等. 随钻测井中岩性识别方法的对比及应用[J]. *煤田地质与勘探*, 2018, 46(1): 165–169.
- CHEN Gang, WANG Kaibin, JIANG Bici, et al. Comparison and application of LWD lithology identification method[J]. *Coal Geology & Exploration*, 2018, 46(1): 165–169.
- [20] 黄安, 蔡文渊, 魏新路, 等. 基于改进随机森林的火山岩测井岩性识别[J]. *科学技术与工程*, 2023, 23(9): 3696–3704.
- HUANG An, CAI Wenyuan, WEI Xinlu, et al. Lithology identification of volcanic logging based on improved random forest[J]. *Science Technology and Engineering*, 2023, 23(9): 3696–3704.
- [21] 王敏, 杨添微, 刘永震, 等. 基于随机森林算法的复杂碳酸盐岩岩性识别[J]. *工程地球物理学报*, 2020, 17(5): 550–558.
- WANG Qi, YANG Tianwei, LIU Yongzhen, et al. Identification of complex carbonate lithology based on random forest algorithm[J]. *Chinese Journal of Engineering Geophysics*, 2020, 17(5): 550–558.
- [22] 孙岗. 基于改进 KNN 算法的潜山复杂岩性测井识别方法[J]. *特种油气藏*, 2022, 29(3): 18–27.
- SUN Kui. Logging identification method of complex lithology in buried hill based on the improved KNN algorithm[J]. *Special Oil & Gas Reservoirs*, 2022, 29(3): 18–27.
- [23] 赵彤彤, 张春雷, 张春雨, 等. 基于模糊熵的 KNN 分类模型在岩性识别中的应用[J]. *计算机工程与应用*, 2018, 54(24): 260–265.
- ZHAO Tongtong, ZHANG Chunlei, ZHANG Chunyu, et al. Application of KNN classification model based on fuzzy entropy in lithology recognition[J]. *Computer Engineering and Applications*, 2018, 54(24): 260–265.
- [24] KUMAR T, SEELAM N K, RAO G S. Lithology prediction from well log data using machine learning techniques: A case study from Talcher coalfield, eastern India[J]. *Journal of Applied Geophysics*, 2022, 199: 104605.
- [25] YU Zhichao, WANG Zhizhang, ZENG Fancheng, et al. Volcanic lithology identification based on parameter-optimized GBDT algorithm: A case study in the Jilin Oilfield, Songliao Basin, NE China[J]. *Journal of Applied Geophysics*, 2021, 194: 104443.
- [26] 马陇飞, 萧汉敏, 陶敬伟, 等. 基于梯度提升决策树算法的岩性智能分类方法[J]. *油气地质与采收率*, 2022, 29(1): 21–29.
- MA Longfei, XIAO Hanmin, TAO Jingwei, et al. Intelligent lithology classification method based on GBDT algorithm[J]. *Petroleum Geology and Recovery Efficiency*, 2022, 29(1): 21–29.
- [27] 彭英, 李克文, 朱应科, 等. 基于 FL-XGBoost 算法的砂泥岩识别方法: 以胜利油田牛庄地区为例[J]. *油气地质与采收率*, 2023, 30(1): 76–85.
- PENG Ying, LI Kewen, ZHU Yingke, et al. FL-XGBoost algorithm-based method for identifying sandstone and mudstone: A

- case study of Niuzhuang area in Shengli Oilfield[J]. Petroleum Geology and Recovery Efficiency, 2023, 30(1): 76–85.
- [28] PROKHORENKOVA L, GUSEV G, VOROBIEV A, et al. CatBoost: unbiased boosting with categorical features[C]//Proceedings of the 32nd International Conference on Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates Inc. , 2018: 6639–6649.
- [29] 李华洋, 谭强, 朱施杰, 等. 基于 CatBoost 算法的孔隙压力预测方法及其在井壁稳定分析中的应用[J]. 中国安全生产科学技术, 2023, 19(2): 136–142.
- LI Huayang, TAN Qiang, ZHU Shijie, et al. Prediction method of pore pressure based on CatBoost algorithm and its application in wellbore stability analysis[J]. Journal of Safety Science and Technology, 2023, 19(2): 136–142.
- [30] LU Chengang, ZHANG Sui'an, XUE Dan, et al. Improved estimation of coalbed methane content using the revised estimate of depth and CatBoost algorithm: A case study from southern Sichuan Basin, China[J]. Computers & Geosciences, 2022, 158: 104973.
- [31] 马晓君, 宋嫣琦, 常百舒, 等. 基于 CatBoost 算法的 P2P 违约预测模型应用研究[J]. 统计与信息论坛, 2020, 35(7): 9–17.
- MA Xiaojun, SONG Yanqi, CHANG Baishu, et al. Application research of P2P default prediction model based on CatBoost algorithm[J]. Statistics & Information Forum, 2020, 35(7): 9–17.
- [32] 罗仁泽, 庾娟娟, 倪华玲, 等. 基于改进集成学习的测井岩性识别方法研究[J]. 石油物探, 2023, 62(2): 212–224.
- LUO Renze, TUO Juanjuan, NI Hualing, et al. Logging lithology identification method based on improved ensemble learning[J]. Geophysical Prospecting for Petroleum, 2023, 62(2): 212–224.
- [33] 顾崇寅, 徐潇源, 王梦圆, 等. 基于 CatBoost 算法的光伏阵列故障诊断方法[J]. 电力系统自动化, 2023, 47(2): 105–114.
- GU Chongyin, XU Xiaoyuan, WANG Mengyuan, et al. CatBoost algorithm based fault diagnosis method for photovoltaic arrays[J]. Automation of Electric Power Systems, 2023, 47(2): 105–114.
- [34] FRIEDMAN J H. Greedy function approximation: A gradient boosting machine[J]. The Annals of Statistics, 2001, 29(5): 1189–1232.
- [35] 徐雄飞, 王嘉伟, 杨波, 等. 塔里木盆地轮南地区侏罗系砂岩储层成岩作用研究[J]. 岩性油气藏, 2011, 23(5): 21–27.
- XU Xiongfei, WANG Jiawei, YANG Bo, et al. Study on diagenesis of Jurassic sandstone reservoirs in Lunnan area, Tarim Basin[J]. Lithologic Reservoirs, 2011, 23(5): 21–27.

(责任编辑 聂爱兰)