

引用格式:李裕涛,李潮流,魏兴云,等.水平井井地模型人工智能预测方法研究与应用[J].油气藏评价与开发,2025,15(5):858-871.

LI Yutao, LI Chaoliu, WEI Xingyun, et al. Research and application of artificial intelligence-based prediction method for horizontal well-formation modelling[J]. Petroleum Reservoir Evaluation and Development, 2025, 15(5): 858-871.

DOI: 10.13809/j.cnki.cn32-1825/te.2025.05.015

水平井井地模型人工智能预测方法研究与应用

李裕涛,李潮流,魏兴云,王浩

(中国石油勘探开发研究院,中国北京 100089)

摘要:水平井钻探已成为油公司提高致密与非常规油气单井产量的重要手段,但由于水平井井眼轨迹与地层的空间关系复杂,传统直井分析思路无法有效应用,准确描述水平井井眼轨迹与目的层及围岩的空间组合关系是水平井测井解释的首要任务。基于导眼井构建地层初始模型,利用测井资料正演约束逐段调整模型是目前主流做法,但该方法时效性低,针对同一地区不同井都需要大量重复的正演计算。因此,在水平井测井数据的处理和解释中,建立合理的井地模型是关键。井地模型能准确描述井眼与地层界面之间的空间关系,包括井眼位置与地层界面距离,井轴位置与地层法线方向的夹角等。同时,基于机器学习和人工智能技术的测井数据分析方法,通过智能模型的训练,已经应用于测井数据解释的各个方面,借助人工智能技术有望突破传统方法的瓶颈。因此,提出一种基于多模型集成与深度神经网络的自动化水平井测井解释方法:首先构建包含不同井眼轨迹与地层组合关系的理论模型,生成测井响应样本库;然后整合极端梯度提升树(eXtreme Gradient Boosting, XGBoost)、轻量级梯度提升机(Light Gradient Boosting Machine, LightGBM)、分类提升机(Categorical Boosting, CatBoost)等机器学习模型,通过多层感知器(Multi-Layer Perceptron, MLP)进一步融合各模型的预测结果;最后对实际测井资料开展井眼轨迹与围岩几何关系的智能自动判别。实例分析显示,该方法在准确捕捉水平井复杂测井响应特征的同时,显著提高了解释速度和精度,能够适应相似地质环境下多口井的快速分析需求,为水平井测井解释提供了一种高效的智能化手段。

关键词:水平井;测井解释;人工智能;深度学习;地层建模

中图分类号:TE243.1

文献标识码:A

Research and application of artificial intelligence-based prediction method for horizontal well-formation modelling

LI Yutao, LI Chaoliu, WEI Xingyun, WANG Hao

(PetroChina Research Institute of Petroleum Exploration & Development, Beijing 100089, China)

Abstract: Horizontal well drilling has become an important method for oil companies to enhance single-well production in tight and unconventional oil and gas reservoirs. However, due to the complex spatial relationship between the wellbore trajectory of horizontal wells and the formation layers, traditional vertical well analysis methods cannot be effectively applied. Accurately describing the spatial combination relationship between the wellbore trajectory, the target layer, and the surrounding rock is a primary task in horizontal well logging interpretation. To address this issue, the mainstream approach is to construct an initial stratigraphic model based on a pilot well and then adjust the model segment by segment using forward modeling constraints from logging data. However, this process is time-consuming and requires numerous repetitive forward modeling calculations for different wells in the same area. Therefore, in the processing and interpretation of horizontal well logging data, developing a reasonable well-formation model is essential. This model enables an accurate description of the spatial relationship between the wellbore and the formation interfaces, including the distance between the wellbore position and formation interfaces and the angle between the wellbore axis and the formation normal direction. At the same time, logging data analysis methods based on machine learning and artificial intelligence (AI) technologies have been applied to various aspects of logging data interpretation by training intelligent models. With the support of AI technologies, it is expected to overcome the bottlenecks of traditional methods. To this end, the study proposed an automated horizontal well logging interpretation method based on multi-model integration and deep neural networks. First, a theoretical model was constructed incorporating different wellbore trajectories and formation combination relationships, and a logging response sample library was generated. Then, machine learning models such as eXtreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), and Categorical Boosting (CatBoost) were integrated, and their prediction results

收稿日期:2024-11-18。

第一作者简介:李裕涛(2000—),男,硕士,助理工程师,从事水平井人工智能解释研究。地址:北京市海淀区学院路20号院,邮政编码:100089。

E-mail: 1902176217@qq.com

基金项目:国家自然科学基金资助项目“深层页岩气藏岩石物理多尺度融合与储层品质井震智能评价方法”(42430810)。

were further fused using a multi-layer perceptron (MLP). Finally, intelligent automatic recognition of the geometric relationship between the well trajectory and the surrounding rock was carried out using actual logging data. Case analysis showed that this method accurately captured the complex logging response characteristics of horizontal wells while significantly improving interpretation speed and accuracy. The proposed method meets the demand for rapid analysis of multiple wells in similar geological environments and provides an efficient, intelligent approach to horizontal well logging interpretation.

Keywords: horizontal wells; well logging interpretation; artificial intelligence; deep learning; formation modelling

随着中国油气勘探开发力度的加大,水平井钻探工作量显著上升。相比于直井,水平井通过增加与油气层的接触面积以提高单井产量和采收率^[1],但水平井的测井响应特征更多表现为非线性,主要受控于井轨迹与目的层或围岩的组合关系控制,在复杂地质背景下对这种组合关系的准确描述是一项重要挑战^[2]。

由于沉积岩固有的各向异性特征,水平井中声波和电阻率系列的测井响应与直井存在显著差异,井轴或仪器轴距围岩距离、相对夹角(简称为井地模型)以及非对称泥浆侵入(仅对电缆测井而言)成为其主要影响因素^[3-5]。水平井测井处理解释的首要任务是要给出准确合理的井地模型,这与传统的直井测井处理解释思路完全不同。因此,有必要研究能够精确处理复杂地质条件的水平井测井解释新方法^[6-8]。

当前,用于描述水平段井地模型的主要方法包括以下2种:井震结合地质建模方法和数值模拟方法。井震结合地质建模方法通过邻井或导眼井构建层状模型,结合构造特征和地震信息约束将层状模型推广到水平井。该方法能够反映一定的地质特征,但对水平井测井资料反映的模型变化缺乏分析,尤其在复杂多变的沉积环境下其局限性更加明显^[9]。数值模拟方法通过正演计算不同模型的测井响应,逐段调整模型直至模拟结果与实际测井资料吻合。该方法能够充分利用水平井测井资料蕴含的井眼-地层关系信息,但模拟计算量大,在测井数据噪声较大的情况下算法稳定性差^[10],并且对同一井区相近地质背景下不同井眼都需要重复相关流程,时效性低。

近年来,随着大数据和人工智能技术的迅速发展,基于机器学习的地质建模方法展现出巨大潜力。这些方法通过学习大量样本,能够在不依赖专家经验的情况下自动构建地质模型。LIU等^[11]提出了一种基于深度学习的岩相分类和三维地质建模方法,提高致密砂岩储层的建模精度。LIAO等^[12]提出了一种基于时间卷积网络(Temporal Convolutional Network, TCN)、双向门控循环单元(Bidirectional Recurrent Unit, BiGRU)和多头自注意力机制的密度测井曲线重构方法,该方法结合物理约束,显著提高了重构曲线的准确性和稳定性。与此同时,范翔宇等^[13]研究了基于自动机器学习(Automated Machine Learnin, AutoML)的测井曲线重构技术,通过集成多模型选择和超参数调优,优化了测井曲线的重构流程。刘军

等^[14]在准噶尔盆地夏子街油田的研究中,通过引入BP神经网络模型,对低含油饱和度砂岩储层的孔隙度、渗透率和含油饱和度进行了预测。此外,武娟等^[15]提出了一种基于大语言模型(Large Language Models, LLMs)的新方法,通过引入利用变压器生成真实的关系与表格数据(Realistic Relational and Tabular Transformer, REaLTabFormer)增强的对齐框架模型(Rational Transfer Function-Large Language Model Meta AI, RTF-LLaMA),实现了致密砂岩储层含水饱和度的精准预测。杨春生等^[16]通过井震结合技术成功提高了曲流型河道砂体的识别精度,并通过动态 Q (品质因子)补偿的偏移处理技术、基于 S 变换(时频分析的一种方法)的三原色(Red Green Blue, RGB)融合技术,精准预测了点坝砂体。综上所述,前人利用人工智能算法评价复杂储层已开展大量研究,但机器学习算法在处理复杂地层时仍然面临诸多挑战,特别是在水平井中利用智能算法开展地层模型自动识别研究尚未见报道。

针对上述问题,提出一种基于多模型集成与深度神经网络架构的新方法,在大斜度井、水平井中不考虑泥浆侵入的前提下,构建大量井地模型模拟计算测井响应形成样本库,然后利用多种机器学习模型进行特征学习,最后对实际井的测井资料进行智能预测,并引入神经网络对预测结果进行融合,输出合理井地模型,为储层定量评价提供基础。

1 研究内容

1.1 测井理论响应模型

利用机器学习做预测,其前提是要确保解释变量(预测变量)和响应变量(被预测变量)之间的内在逻辑关系与供训练学习的样本库基本一致。对1口水平井利用测井资料预测其可能的井地模型,首先要提供大量的与该井实际钻遇地层模型相一致的理论测井响应样本库,然后选择合适的算法进行训练和参数优化。以常见的层状砂泥岩地层为例,选择苏里格气田二叠系下石盒子组致密砂岩作为对象。由于水平井段的钻井液侵入特征极其复杂,因此,研究暂时不考虑侵入过程对测井响应的影响。

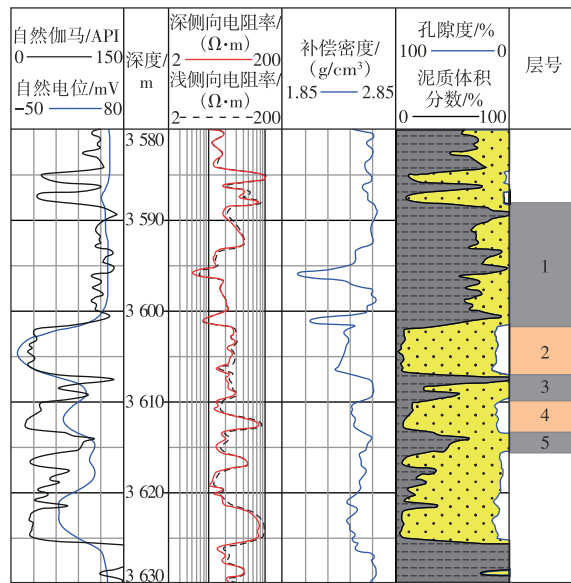
苏里格气田位于鄂尔多斯盆地中北部的中央坳陷带,地质构造整体较为平缓,主要目的层二叠系下石盒子组以三角洲平原沉积为主,发育致密砂岩,储层物性差,

孔隙度介于6%~12%,渗透率普遍低于 $1 \times 10^{-3} \mu\text{m}^2$,目的层厚度一般介于2~10 m^[17-18]。以苏47井区为例,水平井苏47-A-H1井的导眼井为苏47-A井,其测井曲线及小层分布见图1a。从图1a可见,苏47-A井目的层段1号、3号和5号层为泥岩,2号和4号层为砂岩气层,各小层的详细测井参数响应值如表1所示。

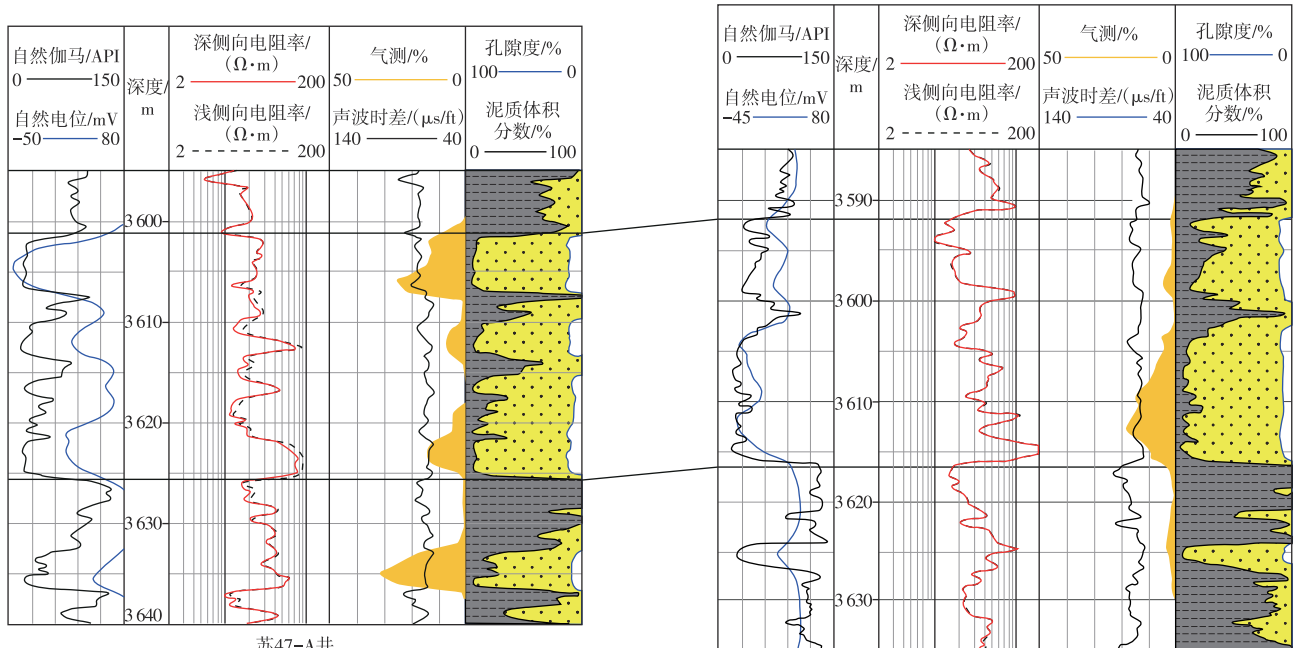
图1b为苏47-A井与邻井苏47-B井的连井剖面图,根据地质设计,2口直井在平面上相距1 610 m,进行海拔高度校正后2口井1号层底部相差8.5 m,换算地层倾角

约0.3°,选择4号层为目标靶体实施水平井钻探。由于陆相沉积的非均质性,图2进一步揭示在平面上4号层厚度、局部倾角甚至岩性都可能发生变化,钻井过程也不可能做到准确控制井轨迹保持在4号层中部。因此,苏47-A-H1井的水平段井轨迹主体应该保持在4号层内部,但很有可能在局部钻遇1~5号5个小层,井轴与4号层界面的夹角也会不断变化,所有这些未知且不可控的影响因素都会在水平段的测井响应中得到体现和表征。

基于苏47-A-H1井的水平段测井资料预测其井地



a. 苏47井区苏47-A导眼井测井曲线



b. 苏47-A与邻井苏47-B井连井剖面

■ 泥岩 ■ 砂岩 ■ 砂岩层 ■ 泥岩层 ■ 全烃

注: 1 ft=30.48 cm。

图1 鄂尔多斯盆地苏里格气田苏47-A-H1井的导眼井苏47-A井及其与邻井苏47-B井连井剖面

Fig. 1 IConnecting well profiles of pilot hole Su 47-A of well Su 47-A-H1 and its adjacent well Su 47-B in Sulige gas field, Ordos Basin

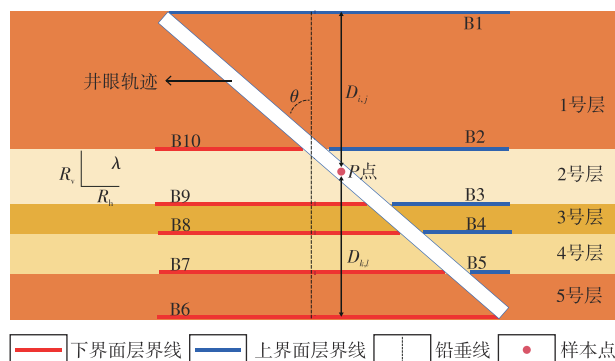
表1 鄂尔多斯盆地苏里格气田苏47-A-H1井的导眼井苏47-A井目的层及围岩5个小层测井响应值
Table 1 Logging response values of target layer and five sub-layers of surrounding rocks in pilot hole Su 47-A of well Su 47-A-H1, Sulige gas field, Ordos Basin

层号	厚度/m	岩性	自然伽马/API	密度/(g/cm ³)	深侧向电阻率/(Ω·m)	浅侧向电阻率/(Ω·m)
1	13.5	泥岩	122.3	2.52	17.5	22.1
2	5.5	砂岩	24.8	2.41	22.9	26.1
3	2.8	粉砂质泥岩	87.1	2.64	22.5	26.2
4	3.5	砂岩	31.6	2.48	25.6	28.3
5	2.2	泥岩	113.4	2.61	16.4	20.3

模型,首先必须考虑井轴与1~5号5个小层呈任意不同组合关系时的理论响应。为此,以图1a所示的5个小层作为层状地层模型,模拟仪器处于不同位置、不同角度时的理论测井响应,形成样本库。

1.2 样本库构建

在图1基础上构建如图2所示的五层水平井地模型,井轨迹就是测井仪器轴。为了确保样本库能够适应复杂地质背景,在样本构建中充分考虑岩性、电各向异性、井眼轨迹与地层界面关系等因素。



注: θ 为井轨迹与地层界面法线法向的夹角,单位($^{\circ}$); λ 为目的层的电各向异性系数; R_v 为垂直电阻率,单位($\Omega \cdot m$); R_h 为水平电阻率,单位($\Omega \cdot m$); $D_{i,j}$ 为P点到某一上界面的铅垂距离,单位m, i 为界面序数($1, \dots, 5$), j 为P点沿井眼轨迹的采样点序数($1, \dots, n$), n 为任意正整数; $D_{k,l}$ 为P点到某一下界面的铅垂距离,单位m, k 为界面序数($1, \dots, 5$), l 为P点沿井眼轨迹的采样点序数($1, \dots, n$)。

图2 井地模型测井响应模拟示意图

Fig. 2 Schematic diagram of logging response simulation for well-formation model

就地层的电各向异性而言,在样本库中考虑垂直电阻率(R_v)和水平电阻率(R_h)等关键参数针对不同岩性(砂岩和泥岩)对测井响应的差异,在样本库中区分砂岩与泥岩的电各向异性特征:①砂岩地层的电各向异性指数(λ)取值范围设置为1.0~1.2,步长0.05,参数范围基于目标地层的岩石物理特性统计结果及导眼井测井曲线的电阻率特征值。②泥岩层具有更显著的电各向异性, λ 的取值范围设置为1.2~1.5,步长0.1,这一范围结合了页岩气

藏中典型泥岩测井响应曲线的特征值以及实际地质资料。

通过正演模拟,即根据已知的地层模型和物理定律,模拟计算出测井曲线等数据,分析不同 λ 条件下测井曲线的响应特征,结果表明,各向异性对双侧向电阻率具有显著的控制作用。

在选择井眼轨迹参数,如倾角(θ)、井眼到层界面的距离(D)等时,以目标区域的实际地质条件为基础,同时适当拓展了参数范围,以确保样本库具有更好的泛化能力:① θ 的范围设为 $70^{\circ} \sim 90^{\circ}$,步长为 0.1° ,比地层的可能倾角范围增加约10%~20%,保证了样本库在地质条件参数上的多样性覆盖;② D 的范围设为0~30 m,这一范围基于苏里格气田目标层厚度分布统计结果。

以此为基础建立3个参数(表2):倾角(θ)、砂岩的电各向异性系数(λ_s)和泥岩的电各向异性系数(λ_M)。任意组合的井地模型1 100个,使用美国Maxwell Dynamics公司的商业软件LogXD正演模拟计算每个模型的自然伽马、密度、双侧向电阻率等测井响应理论值。

研究样本库构建中,假设测井仪器位于井眼中心位置,因此,未考虑偏心影响。这一假设简化了测井响应的计算过程,同时也符合研究重点,即在理想条件下模拟井眼与地层组合关系对测井响应的影响。然而,在实际地层中,偏心效应(测井工具位置偏离井眼中心时,导致测量结果发生偏差的现象)可能对电阻率、密度测井结果产生非对称性影响,需要进一步研究其对模型预测的潜在影响。

以上构建的每一个样本点P的正演计算结果,都可以看成该点距层界面一定距离、该点的切线与层界面呈一定角度组合关系时5个小层的综合反映。如上所述,考虑每一个小层界面并非完全水平、厚度在横向上可能有变化这2个重要因素,需给定井轨迹上任一点P相对于每个小层的距离。

以图2为例,以井眼为界,规定井轨迹上方的层界面为上界面(即B1~B5),下方的层界面为下界面(即B6~B10),P点到某一上界面的铅垂距离为 $D_{i,j}$,其中, i 为界面序数($1, \dots, 5$), j 为P点沿井眼轨迹的采样点序数($1, \dots, n$), n 为任意正整数。例如, $D_{1,1}$ 表示沿井眼轨迹的第1个采样点到B1的铅垂距离, $D_{1,1} \sim D_{1,n}$ 的集合构成

表2 井地模型数量分类

Table 2 Classification of well-formation model numbers

倾角 $\theta/(^\circ)$	λ_s	λ_M	模型数量/个
70~75	1.00	1.2	50
75~80	1.00	1.3	50
80~85	1.00	1.4	50
85~90	1.00	1.5	70
70~75	1.05	1.2	50
75~80	1.05	1.3	50
80~85	1.05	1.4	50
85~90	1.05	1.5	70
70~75	1.10	1.2	50
75~80	1.10	1.3	50
80~85	1.10	1.4	50
85~90	1.10	1.5	70
70~75	1.15	1.2	50
75~80	1.15	1.3	50
80~85	1.15	1.4	50
85~90	1.15	1.5	70
70~75	1.20	1.2	50
75~80	1.20	1.3	50
80~85	1.20	1.4	50
85~90	1.20	1.5	70

数组 $Y_1, D_{5,1} \sim D_{5,n}$ 的集合构成数组 Y_5 。

同理, P 点到某一下界面的铅垂距离为 $D_{k,l}$, 其中 k 为界面序数 ($6, \dots, 10$), l 为 P 点沿井眼轨迹的采样点序数 ($1, \dots, n$)。例如, $D_{6,1}$ 表示沿井轨迹第 1 个采样点到 B6 的铅垂距离。 $D_{6,1} \sim D_{6,n}$ 的集合构成了数组 Z_1 , $D_{10,1} \sim D_{10,n}$ 的集合构成了数组 Z_5 。最终的样本点数据结构如图 3 所示, 对于实际水平段井地模型解释的反演问题 (通过实际的测井数据来推断地下地层的特性),

$Y_1 \sim Y_5, Z_1 \sim Z_5$ 就是需要预测的目标变量。如果水平井轨迹钻穿更多个小层, 则图 3 所示的样本数据结构还需要增加更多的目标变量。

1.3 多模型集成智能算法

1.3.1 数据预处理

在训练前, 首先对前面计算的样本库中测井数据进行预处理, 包括 2 个步骤: 数据清洗和归一化。在数据预处理中除了归一化外, 也尝试了采用主成分分析 (Principal Component Analysis, PCA) 以减少特征冗余并提升训练效率。然而, 尽管特征降维技术可有效降低输入特征维度, 但降维后的数据在训练中并未显著提升模型预测精度, 部分实验甚至出现精度略微下降的情况。这一现象可能与以下原因相关: ①测井数据的高维特征中包含重要的非线性信息, 而特征降维技术本质上是一种线性降维方法, 未能捕捉这些复杂特征; ②样本库中各测井参数与目标变量的相关性较强, 直接应用特征降维技术可能弱化了这些重要特征的权重, 从而影响了模型性能。

数据清洗是为了填补个别样本可能存在的缺失值, 确保数据的完整性和一致性^[19]。由于表 1 所示的测井响应值分布区间不统一, 在归一化过程中, 将所有数值缩放到 0~1, 以减少不同量纲之间的差异对模型训练的影响^[20]。

1.3.2 机器学习算法设计

智能算法种类很多, 根据常用算法的优势比较, 优选 6 种算法: 极端梯度提升树 (eXtreme Gradient Boosting, XGBoost)、随机森林 (Random Forest, RF)、支持向量机

自然伽马/ API	密度/ (g/cm ³)	浅侧向电阻率/ ($\Omega \cdot m$)	深侧向电阻率/ ($\Omega \cdot m$)	倾角/($^\circ$)	Y_1	Y_2	Y_3	Y_4	Y_5	Z_5	Z_4	Z_3	Z_2	Z_1
31.669 2	2.485 1	24.599 1	24.461 4	74.000 0	24.638 3					-0.213 0	-0.319 5	-0.461 5	-0.710 0	-2.911 2
31.671 3	2.485 1	24.512 6	24.459 8	74.000 0	24.673 8					-0.177 5	-0.234 0	-0.426 0	-0.674 5	-2.875 7
31.675 6	2.485 1	24.430 0	24.452 8	74.000 0	24.709 3					-0.142 0	-0.248 5	-0.390 5	-0.639 0	-2.840 2
31.681 3	2.485 1	24.342 3	24.440 2	74.000 0	24.744 3					-0.105 5	-0.213 0	-0.355 0	-0.603 5	-2.804 7
31.689 3	2.485 1	24.248 7	24.421 1	74.000 0	24.744 3					-0.071 0	-0.177 5	-0.319 5	-0.563 0	-2.769 1
31.702 9	2.485 1	24.148 5	24.395 3	74.000 0	24.815 8					-0.035 5	-0.142 0	-0.284 0	-0.532 5	-2.733 6
31.722 7	2.485 1	24.043 1	24.364 1	74.000 0	24.851 3					0.000 0	-0.106 5	-0.248 5	-0.497 0	-2.698 1
31.753 1	2.485 1	23.931 8	24.326 9	74.000 0	24.886 3	11.218 6					-0.071 0	-0.213 0	-0.461 5	-2.662 5
31.799 9	2.485 1	23.816 1	24.284 5	74.000 0	24.922 3	11.254 1					-0.035 5	-0.177 5	-0.426 0	-2.627 1
31.371 9	2.485 2	23.699 9	24.238 1	74.000 0	24.957 8	11.289 6					0.000 0	-0.142 0	-0.390 5	-2.591 6
31.984 3	2.485 2	23.583 5	24.187 7	74.000 0	24.993 4	11.325 1	5.928 8				-0.106 5	-0.355 0	-0.603 5	-2.556 1
32.161 5	2.485 3	23.463 9	24.132 0	74.000 0	25.028 9	11.350 6	5.964 3				-0.071 0	-0.319 5	-0.563 0	-2.520 5
32.445 1	2.485 4	23.325 6	24.060 9	74.000 0	25.064 4	11.396 1	5.999 8				-0.035 5	-0.284 0	-0.523 5	-2.485 1
32.901 2	2.485 7	23.149 2	23.963 6	74.000 0	25.099 9	11.431 6	6.035 3					0.000 0	-0.248 5	-2.449 6
33.654 7	2.486 2	22.904 9	23.809 9	74.000 0	25.135 4	11.467 1	6.070 8	3.301 7				-0.213 0	-0.603 5	-2.414 1
34.313 2	2.487 3	22.508 9	23.493 6	74.000 0	25.170 9	11.502 6	6.106 3	3.337 2				-0.177 5	-0.563 0	-2.378 5
37.313 2	2.489 9	21.953 0	22.999 2	74.000 0	25.206 4	11.538 1	6.141 8	3.372 7				-0.142 0	-0.523 5	-2.343 1
42.240 7	2.495 7	21.389 7	22.465 7	74.000 0	25.241 9	11.573 6	6.177 3	3.408 2				-0.106 5	-0.484 0	-2.307 6
53.283 8	2.515 3	20.810 9	21.922 5	74.000 0	25.277 4	11.609 1	6.212 8	3.443 7				-0.071 0	-0.444 5	-2.272 1
63.456 2	2.531 5	20.262 9	21.413 8	74.000 0	25.312 9	11.644 6	6.248 3	3.479 2				-0.035 5	-0.405 0	-2.236 5
73.180 6	2.546 5	19.821 1	21.023 3	74.000 0	25.348 4	11.680 1	6.283 8	3.514 7	0.000 0			0.000 0	-0.365 5	-2.201 1
82.954 4	2.561 7	19.470 3	20.728 4	74.000 0	25.383 9	11.715 6	6.319 3	3.550 2	0.035 5				-0.326 0	-2.165 6
93.282 4	2.578 2	19.117 9	20.420 3	74.000 0	25.419 4	11.751 1	6.354 8	3.585 7	0.071 0				-0.286 5	-2.130 1
103.926 2	2.596 2	18.786 4	20.124 9	74.000 0	25.454 9	11.786 6	6.390 3	3.621 2	0.106 5				-0.247 0	-2.094 6
103.235 0	2.601 9	18.491 7	19.862 8	74.000 0	25.490 4	11.822 1	6.425 8	3.656 7	0.142 0				-0.207 5	-2.059 1
110.370 3	2.604 1	18.244 5	19.646 4	74.000 0	25.525 9	11.857 6	6.461 3	3.692 2	0.177 5				-0.168 0	-2.023 6

注: 空白处为空值, 表明该点铅垂线未在该深度穿过该层界面。

图3 样本点数据结构示意图

Fig. 3 Schematic diagram of sample point data structure

(Support Vector Machine, SVM)、轻量级梯度提升机(Light Gradient Boosting Machine, LightGBM)、分类提升机(Categorical Boosting, CatBoost)和极限学习机(Extreme Learning Machine, ELM),并分析其基本原理及优势(表3)^[21-22]。对于XGBoost算法使用网格搜索进行超参数优化,其他模型则采用经验参数^[23-25]。

针对需要预测的10个目标变量 $Y_1 \sim Y_5, Z_1 \sim Z_5$,依据不同算法的均方误差(Mean Square Error, MSE)、决定系数

(R^2)和平均绝对误差(Mean Absolute Error, MAE),筛选每一个算法适用的目标变量(表3)^[26-27]。MSE是反映预测值与真实值之间误差的平方平均值,MSE值越小表示模型预测越精确,但对异常值敏感; R^2 反映模型对目标变量的解释能力,取值范围为[0, 1],值越接近1表示模型拟合效果越好;MAE反映预测值与真实值之间误差的绝对值平均大小,MAE值越小表示模型预测误差越小,且对异常值的敏感性低于均方误差。

表3 目标变量预测的适用算法组合及算法特性分析

Table 3 Analysis of applicable algorithm combinations and algorithm characteristics for target variable prediction

算法名称	基本原理	优势	适用的目标变量
极端梯度提升树 (eXtreme Gradient Boosting, XGBoost)	基于梯度提升决策树通过迭代添加决策树优化性能,目标函数包括损失函数和正则化项	快速的训练速度和较高的预测精度,良好的泛化能力	$Y_2, Y_4, Y_5, Z_3, Z_4, Z_5$
随机森林 (Random Forest, RF)	集成多棵决策树提高模型的鲁棒性和泛化能力,每棵树在随机选择的样本和特征上训练,减少过拟合风险	有效减少过拟合,提高预测稳定性	$Y_1, Y_2, Y_3, Y_4, Y_5, Z_1, Z_2, Z_3, Z_4, Z_5$
支持向量机 (Support Vector Machine, SVM)	在高维空间中构建最优超平面,通过优化数据点到超平面最小距离提高模型的鲁棒性	能处理高维数据,具有良好的泛化能力	不适用
轻量级梯度提升机 (Light Gradient Boosting Machine, LightGBM)	通过对特征值进行直方图离散化提高训练速度和效率,同时保持高预测精度	训练速度快,能处理大规模数据集,预测精度高	Y_1, Y_3, Z_1, Z_2
分类提升机 (Categorical Boosting, CatBoost)	优化的梯度提升算法,通过目标编码和其他优化技术提高对类别数据的处理能力	有效处理类别特征,提高预测精度和效率	$Y_1, Y_2, Y_3, Y_4, Y_5, Z_1, Z_2, Z_3, Z_4, Z_5$
极限学习机 (Extreme Learning Machine, ELM)	通过随机生成的隐藏层提高训练速度和预测性能,无需迭代训练,具有很高的计算效率	训练速度快,大多数情况下预测精度高	不适用

MSE、 R^2 和MAE的定义分别如式(1)一式(3):

1) MSE:

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

式中: σ_{MSE} 为MSE值; N 为样本数量, y_i 为第*i*个样本的实际值, \hat{y}_i 为第*i*个样本的预测值^[28]。

2) R^2 :

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (2)$$

式中: R^2 为决定系数; \bar{y} 为实际值的平均值^[29-30]。

3) MAE^[31]:

$$\sigma_{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (3)$$

式中: σ_{MAE} 为MAE值。

为综合评估模型性能与选择最优模型,采用多指标综合评价:优先选择MSE值和MAE值最低,且 R^2 接近1的机器学习模型。同时,在测井数据中,特征维度较高且可能存在多重共线性,这会增加模型复杂度并导致过拟合风险。针对高维数据的特性,采用正则化方法进行优化:在模型训练中引入L1正则化和L2正则化,两者均用于约束模型参数,避免权重过大导致的过拟合,L1正则化用于最小绝对收缩和选择算子(Lasso, Least Absolute Shrinkage and Selection Operator)回归模型,L2用于岭

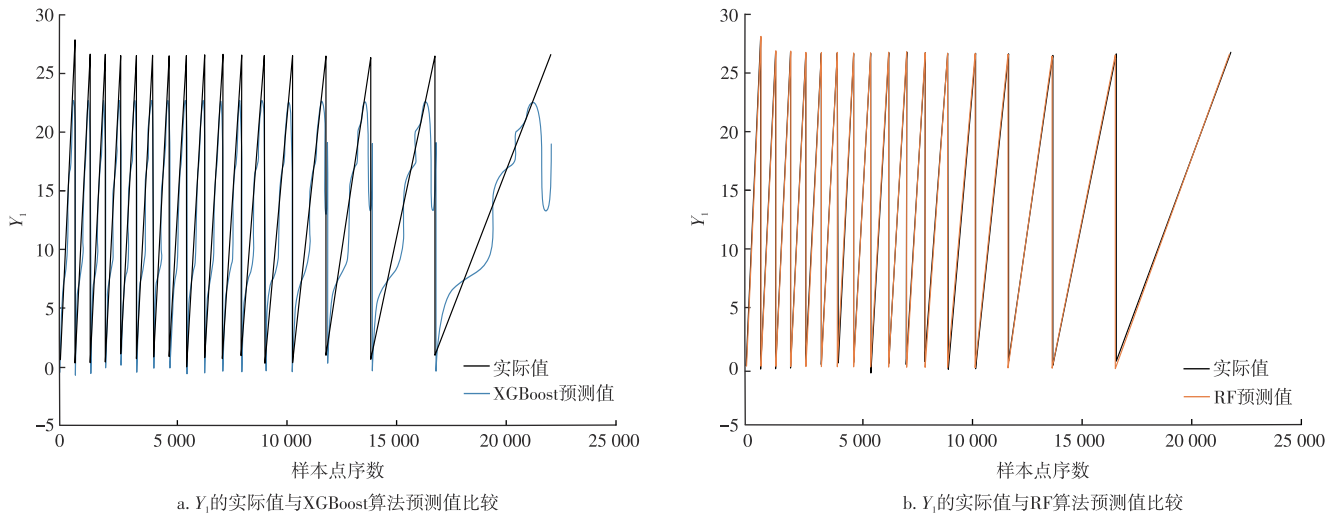
(Ridge)回归模型。表4列出了部分预测算法对10个目标变量的MSE值、MAE值和 R^2 的评估结果。

以 Y_1 为例,由不同算法预测效果的对比(图4)可以看出,RF算法的预测曲线在大多数情况下与实际值高度吻合,MSE值仅为0.008 6, R^2 高达0.998 3,表明该算法能够准确拟合 Y_1 的实际变化趋势,尤其在中等幅度的变化区域,RF算法能够较为精确地跟踪实际值的波动,具有较强的稳定性。LightGBM算法和CatBoost算法的预测值与实际值也基本重合,而XGBoost算法、SVM算法和ELM算法预测结果误差较大。基于上述分析,最终选择RF算

表4 各目标变量的机器学习算法性能评估结果

Table 4 Performance evaluation results of machine learning algorithms for each target variable

目标变量	模型	MSE值	决定系数(R^2)	MAE值
Y_1	RF	0.008 6	0.998 3	0.007 8
Y_2	RF	0.021 6	0.990 1	0.014 2
Y_3	RF	0.010 3	0.998 1	0.013 1
Y_4	RF	0.003 9	0.998 4	0.007 8
Y_5	XGBoost	0.000 7	0.997 4	0.005 3
Z_1	RF	0.024 3	0.998 6	0.097 0
Z_2	RF	0.000 7	0.999 8	0.097 0
Z_3	RF	0.000 5	0.998 8	0.009 9
Z_4	XGBoost	0.000 4	0.998 1	0.014 6
Z_5	RF	0.000 1	0.999 1	0.009 4

图4 目标变量 Y_1 实际值与不同算法预测值对比Fig. 4 Comparison between actual values of target variable Y_1 and predicted values by different algorithms

法、LightGBM 算法和 CatBoost 算法对 Y_1 变量进行预测。对其他 9 个目标变量的预测,采用类似筛选择优过程。

1.3.3 多层感知机(MLP)设计与多模型网络结构设计

为充分利用不同算法的优势,虽然对某一个目标变量可能会采取多个机器学习算法进行预测,但最终结果只能保留其中之一(表3)。因此,采用多模型集成策略以确保能够适应陆相沉积横向复杂多变的地质特征。图3所示的样本点具有典型的高维度、非线性以及数据连续等特性,非常适合使用多层感知机(Multi-Layer Perceptron, MLP)进行多模型集成^[32-34]。

在设计 MLP 模型时,针对隐藏层节点数、激活函数、学习率等超参数进行了优化,以确保模型能够适应不同地质背景下的测井数据:①隐藏层节点数:第一隐藏层的节点数直接影响模型的表达能力与计算复杂度,研究分别测试了每层 16、32、64、128 个节点的结构;②激活函数:选择激活函数对模型非线性映射能力至关重要,研究比较了整流线性单位(Rectified Linear Unit, ReLU)函数、Sigmoid 函数和双曲正切(Tanh)函数这 3 种常用激活函数的表现;③学习率:学习率控制模型的训练步长,测试中分别设定为 10^{-3} 、 10^{-4} 、 10^{-5} 。

为验证所选超参数对模型性能的影响,以 Y_1 为例针对隐藏层节点数、激活函数、学习率等参数进行灵敏度实验,实验结果如表5所示。

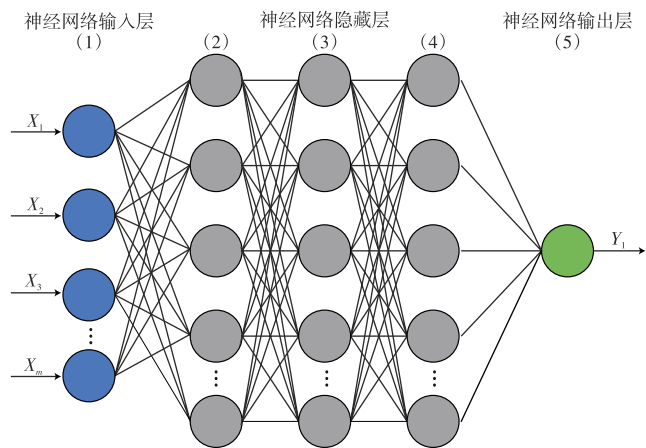
从表5中可以看出,随着隐藏层节点数的增加,模型的验证集 MSE 值和 MAE 值均有所下降,但当节点数超过 64 时,性能提升趋于平缓,而训练时间显著增加。因此,选择隐藏层节点数为 64 是一个性能和效率的折中点。并且在同一节点数,相同学习率下 ReLU 激活函数的表现优于 Sigmoid 和 Tanh,验证集的 MSE 值和 MAE 值更低,

表5 目标变量 Y_1 隐藏层节点数、激活函数、学习率等参数灵敏度实验Table 5 Sensitivity experiments on parameters of target variable Y_1 (number of hidden-layer nodes, activation function, learning rate, etc.)

隐藏层节点数	激活函数	学习率	训练时间/s	验证集 MSE 值	验证集 MAE 值	验证集 R^2
16	ReLU	10^{-4}	115	0.020 5	0.120 1	0.995 0
32	ReLU	10^{-4}	132	0.016 0	0.105 0	0.997 8
64	ReLU	10^{-4}	204	0.008 2	0.007 1	0.999 1
128	ReLU	10^{-4}	310	0.007 7	0.006 8	0.999 2
64	Sigmoid	10^{-4}	220	0.010 3	0.011 4	0.996 5
64	Tanh	10^{-4}	215	0.016 5	0.010 8	0.997 2
64	ReLU	10^{-5}	250	0.009 1	0.008 3	0.999 5
64	ReLU	10^{-3}	150	0.013 4	0.115 0	0.995 2

同时训练时间相对更短。这表明 ReLU 更适合处理测井数据的非线性特征。在学习率为 10^{-4} 时,模型在验证集上的表现最佳(MSE 值接近理想值 0, R^2 接近理想值 1)。较大的学习率(如 10^{-3})可能导致收敛不稳定,而较小的学习率(如 10^{-5})则导致训练速度显著下降。

MLP 是一种基础的深度学习模型,采用输入层、多个隐藏层和输出层的结构(图5),能够快速有效提取这些高维度特征,捕捉层界面的复杂变化,根据灵敏度分析结果最终确定采用非线性激活函数(ReLU)。研究中所采用的 MLP 结构的主要特点包括:①输入层的节点数量等于参与集成的算法个数,例如,对于目标变量 Y_1 , MLP 的输入层包含 3 个节点,分别接收 RF、CatBoost 和 LightGBM 的预测结果;②包含 3 层隐藏层,分别有 64、32 和 16 个节点。每个隐藏层之后均增加批量归一化层以提升训练的稳定性,并使用 ReLU 激活函数增加网络的非线性表达能力;③输出层为一个单节点的线性层,通过一个神经元



注: m 为任意正整数。

图5 MLP结构

Fig. 5 MLP structure

和线性激活函数生成最终的预测值;④每个神经元接收上一层所有神经元的输出,然后进行加权求和。

多模型网络结构(图6)包括输入层、多模型层、聚合层、稠密层、输出层和成图层。输入层的作用是接收原始测井数据并进行预处理,每一个算法独立地接收输入并生成一组预测结果;多模型层由多种智能算法构成,这些模型并行运行,提取输入数据中的模式信息;聚合层和稠密层对提取的特征进行整合和优选,利用非线性转换增强特征的表达能力。

在研究的代码实现中,稠密层即是MLP,由多个“全连接层+ReLU激活函数+批量归一化”组合模块层叠构成。为优化稠密层的训练效果,引入学习率调度器,通过分阶段的学习率调整机制来改善模型的收敛表现^[35-38]。

调度器的设计逻辑是:在前20个Epoch(训练轮数,即模型在训练过程中遍历整个训练数据集一次的过程)中进行学习率预热,从初始学习率逐渐增加到目标学习率,以确保初期训练的平稳;在第20个Epoch之后,则采用阶梯式下降机制,每经过10个训练轮次,学习率按比例衰减,以实现模型的稳步收敛^[39-42]。

输出层将所有处理结果汇总,生成最终的预测输出,导出变量 Y 和 Z 的最终预测结果。成图层(或称可视化层)负责将结果进行可视化展示。研究中,输出层生成的一维数组 U 、 V 代表地层边界信息,可直接作为成图层的输入数据,用于绘制地层轮廓和水平井井眼轨迹的穿越关系。具体步骤如下:①提取地层边界坐标:从输出层获取 U 、 V 数组,提取每一层的边界坐标,这些边界坐标定义了地层的形状轮廓;②叠加井眼轨迹:在绘制水平段目标层轮廓的基础上,将实测的井眼轨迹逐点叠加,从而展现井眼穿越地层的过程,为模型预测提供空间位置背景;③动态更新:如果输出层的 U 、 V 数组更新,成图层也可以实时更新地层边界,以动态展示水平井轨迹在不同地层中的位置变化。

以目标变量 Y_1 为例,采用MLP架构的训练损失与验证损失随迭代轮次变化情况如图7所示。从定量分析来看,从训练开始到第25个轮次,训练损失的MSE值从初始阶段的250迅速下降,而验证损失的MSE值也从初始阶段的300快速下降,验证损失曲线逐步与训练损失曲线重叠;随着训练的深入,损失曲线在25个轮次后趋于稳定,训练和验证损失均保持在接近0的较低水平。表明MLP结构在前期阶段就有效地捕捉到了数据中的重要特征。对其他目标变量的分析也有类似结论。因此,

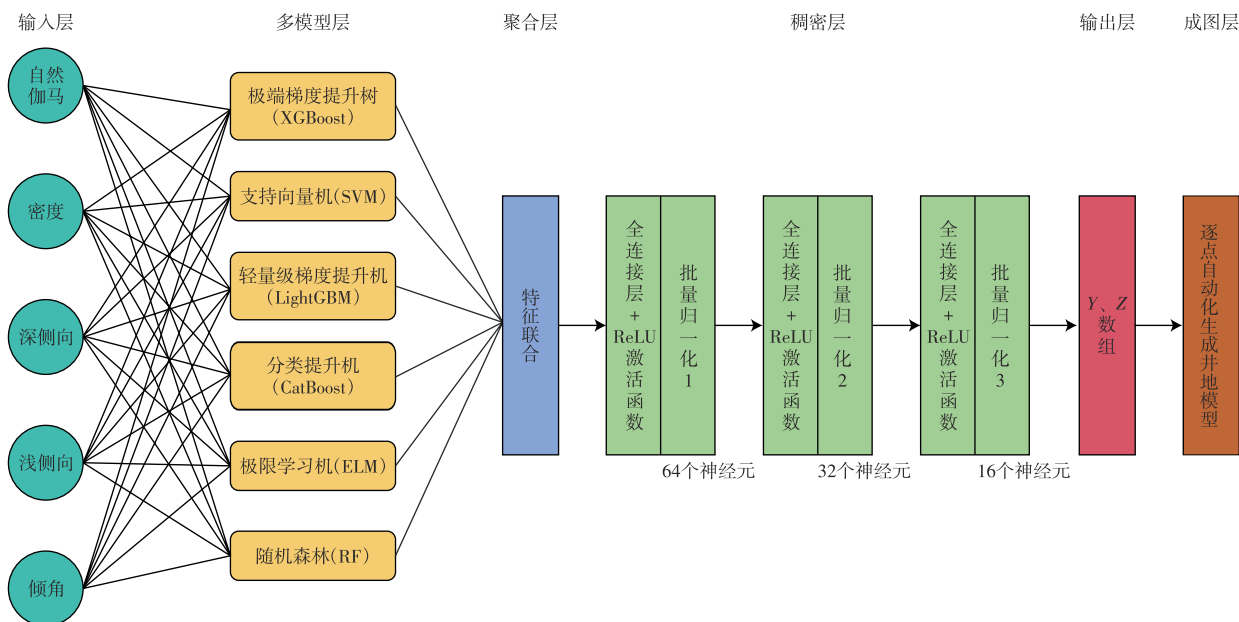
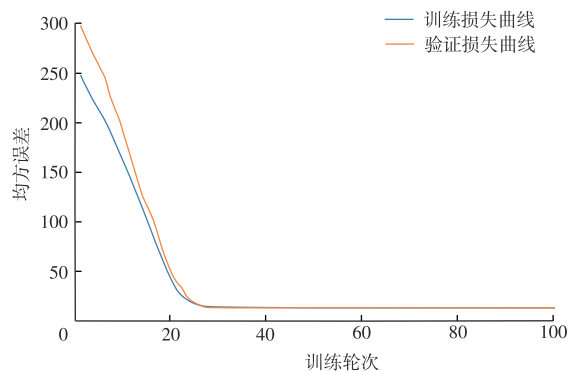


图6 多模型网络结构

Fig. 6 Multi-model network structure

图7 目标变量 Y_1 的训练与验证损失曲线Fig. 7 Training and validation loss curves of target variable Y_1

基于多层感知机的预测方法在地层界面预测任务中具备良好的快速收敛性。

2 典型实例分析

2.1 苏里格井区

采集苏47井区苏47-A-H1水平井水平段的自然伽马、密度、双侧向电阻率、井斜等测井曲线数据。将这些数据进行预处理后,输入到基于图1所示5层地层模型理论样本库训练的多模型网络中,并经多模型集成,最终输出的井地模型界面如图8所示。

从图8可以看出,在利用新建立的智能算法预测的苏47-A-H1水平井段井地模型中,1~5号层沿水平段呈一定角度抬升,估算倾角约为 0.38° ,与图2所示的地层真

倾角相差 0.08° ,说明这种智能算法能够预测出大致的地层构造变化趋势。对井轨迹进行具体分析可以看出:水平井在3 689 m进入1号层,随后在3 737 m和3 770 m依次穿越2号层和3号层,从3 785 m起井斜角逐渐增大,至3 806 m进入目标4号层后井斜角达到 90° ,之后主要沿4号层顶部穿行。通过进一步观察细节可以发现:对应的自然伽马曲线从3 689 m至3 806 m处呈现“升—降—升—降”的变化趋势,与井轨迹穿越不同岩性层(泥岩—砂岩—泥岩—砂岩)的响应特征一致,验证了该算法在定性描述井地模型上的可靠性。此外,水平井在4号层顶部穿行时的局部自然伽马曲线高值表明井轨迹靠近上覆3号泥岩层,反映了该算法对地层倾角细微变化的敏感性,进而支持其在定量刻画井轨迹与目的层空间关系中的应用潜力。

研究中的模型假设测井仪器始终位于井眼中心,然而在实际操作中,井眼形态、重力及地层环境可能导致仪器发生偏心效应,尤其在造斜段和水平段等复杂井段中,这种效应可能使电阻率曲线呈现非对称性响应。为解决这一局限性,未来研究可引入偏心参数 E (即测井仪器中心与井眼中心的径向偏移距离,单位m)及相关物理校正模型,模拟偏心条件下的测井响应特征,并结合实际测井数据进行定量分析,以提升模型的适应性和预测精度。此外,在实例分析中尝试将主成分分析(Principal Component Analysis, PCA)降维方法应用于数据预处理,结果表明,虽然PCA降维能够减少训练时间,但对地层组合关系的预测精度提升有限。以苏47-A-H1水平井

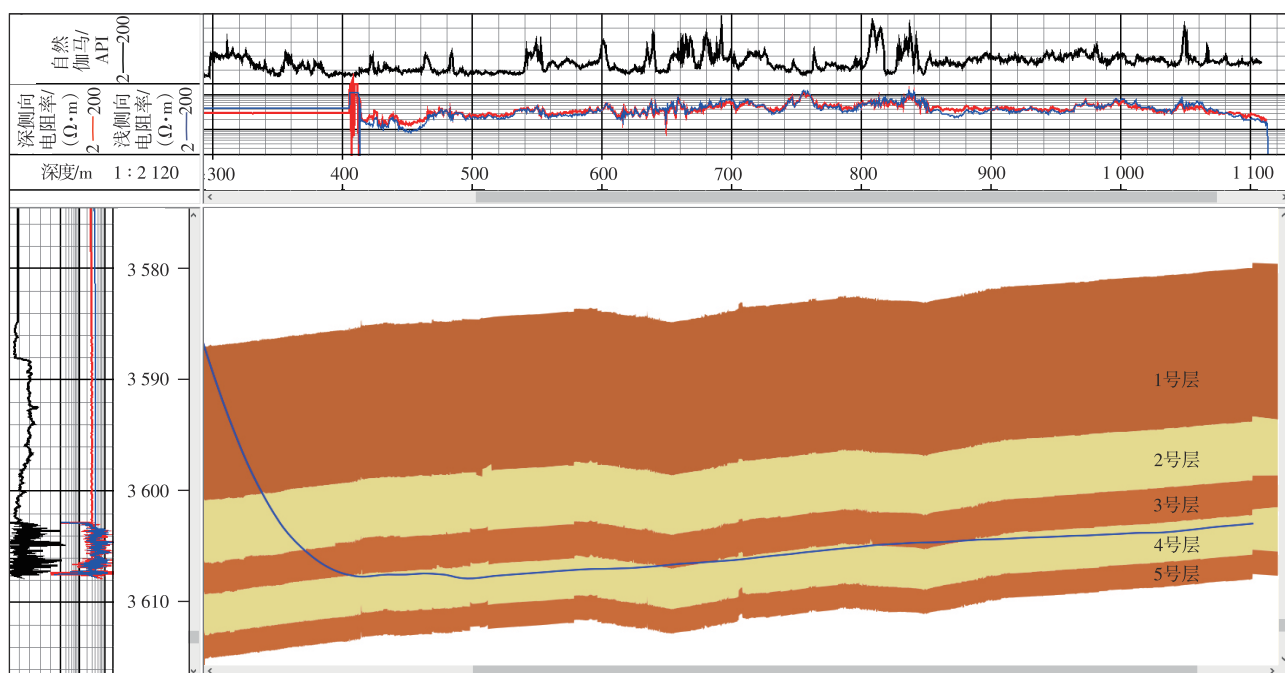


图8 机器学习预测的水平井苏47-A-H1井地模型界面

Fig. 8 Interface of machine learning-predicted well-formation model for horizontal well Su 47-A-H1

段为例,降维后的模型与未降维模型相比,在关键地层界面(如2号与3号层交界处)的预测误差略有增加。因此,最终选择保留原始归一化数据作为模型输入,以在训练效率和预测精度之间取得平衡。

同时在在苏47井区相似地质背景下,传统方法依赖人工逐段对比测井曲线,或针对相近地质条件下的多口井进行重复建模和调剖;而新建的方法只需直接输入测井曲线信息即可对多口井展开快速自动智能化的井地模型分析,无需重复建模,显著提升了解释速度和精度。以苏47-A-H1井地模型(图8)为例,在获得该预测结果后,可进一步开展水平段电阻率曲线的围岩校正和储层岩石物理参数计算,为后续完井方案的设计提供可靠指导。

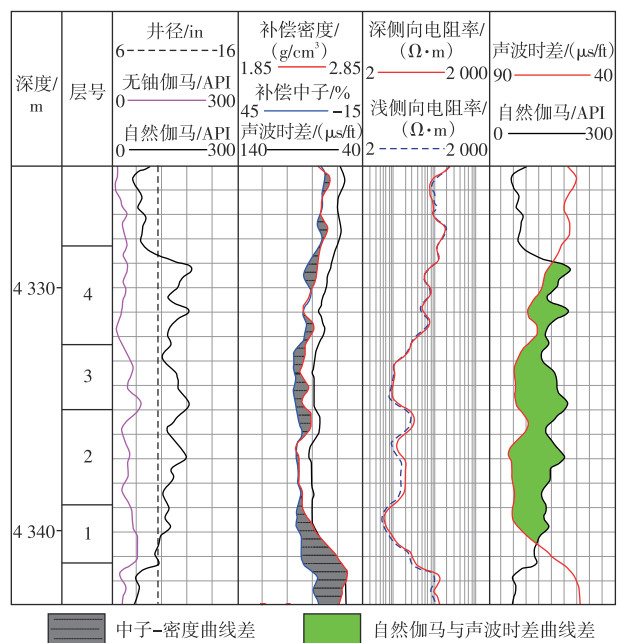
2.2 大页井区实例

为验证新建的方法在不同地质背景下水平井测井解释的适用性,以四川盆地川中地区深层页岩气井DY1H井为例。该井导眼井测井曲线(图9a)显示其目标层为一套高伽马页岩,结合岩心分析分为4个小层,其中1号层电阻率最低,3号层电阻率略高于1号层,4号层电阻率最高,2号层电阻率居中。因2号层有机质含量高且脆性好,被选为DY1H水平井的目标靶体。

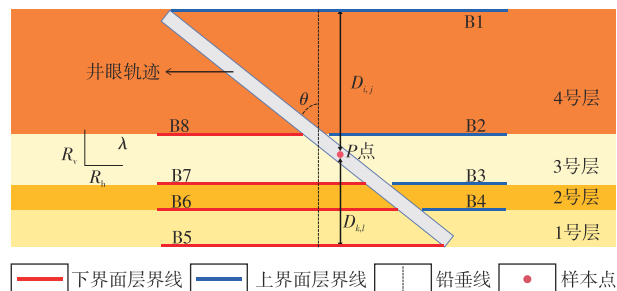
基于导眼井揭示的4个小层,建立初始层状四层水平井地层模型(图9b)。通过将 D 设置为0~15 m(步长0.125 m), θ 设置为 $80^\circ\sim 95^\circ$ (步长 0.2°), λ 设置为1~1.5(步长0.1),生成1 050个井地模型组合,构建样本库。利用该样本库训练多模型网络并优化参数,得到适用于DY井区的预测模型。最终,基于DY1H水平段测井数据生成的井地模型界面如图10a所示。

从图10b可以看出,DY1H井轨迹在4 500 m处进入4号小层,入靶后继续向下穿行,穿过3号层后达到2号层,之后主要沿2号目的层内部穿行,并且多次穿出和穿回2号层。从细节上分析,水平井分别在4 500、4 524.7、4 544.8 m依次进入4号、3号、2号层,对应电阻率测井曲线从4 500 m至4 544.8 m先由高降低,再由低到高的变化趋势,与导眼井揭示的4个小层电阻率相对高低规律一致。进入2号目的层后,在4 544.8~4 986 m处,井轨迹先后3次从2号小层顶部穿出进入3号小层又返回2号层,对应电阻率曲线的3次起伏,验证了人工智能预测方法确定的该段井地模型的合理性。

图10b为地质专家手工调剖得到的DY1H井地模型界面,井轨迹从4号层依次穿越至2号层,并在2号层与3号层间多次穿插,与图10a预测结果在趋势和界面位置上基本一致,且已获生产数据验证。相较于传统逐段调剖,新建的方法基于导眼井揭示的层状初始模型构建



a. DY1H井导眼井测井曲线



b. 针对DY1H井初始地层模型

注:1 in=2.54 cm;1 ft=30.48 cm; θ 为井轨迹与地层界面法线方向的夹角,单位($^\circ$); λ 为目的层的电各向异性系数; R_v 为垂直电阻率,单位($\Omega\cdot m$); R_h 为水平电阻率,单位($\Omega\cdot m$); $D_{i,j}$ 为P点到某一上界面的铅垂距离,单位m, i 为界面序号(1, ..., 5), j 为P点沿井眼轨迹的采样点序号(1, ..., n); $D_{k,l}$ 为P点到某一下界面的铅垂距离,单位m, k 为界面序号(1, ..., 5), l 为P点沿井眼轨迹的采样点序号(1, ..., n)。

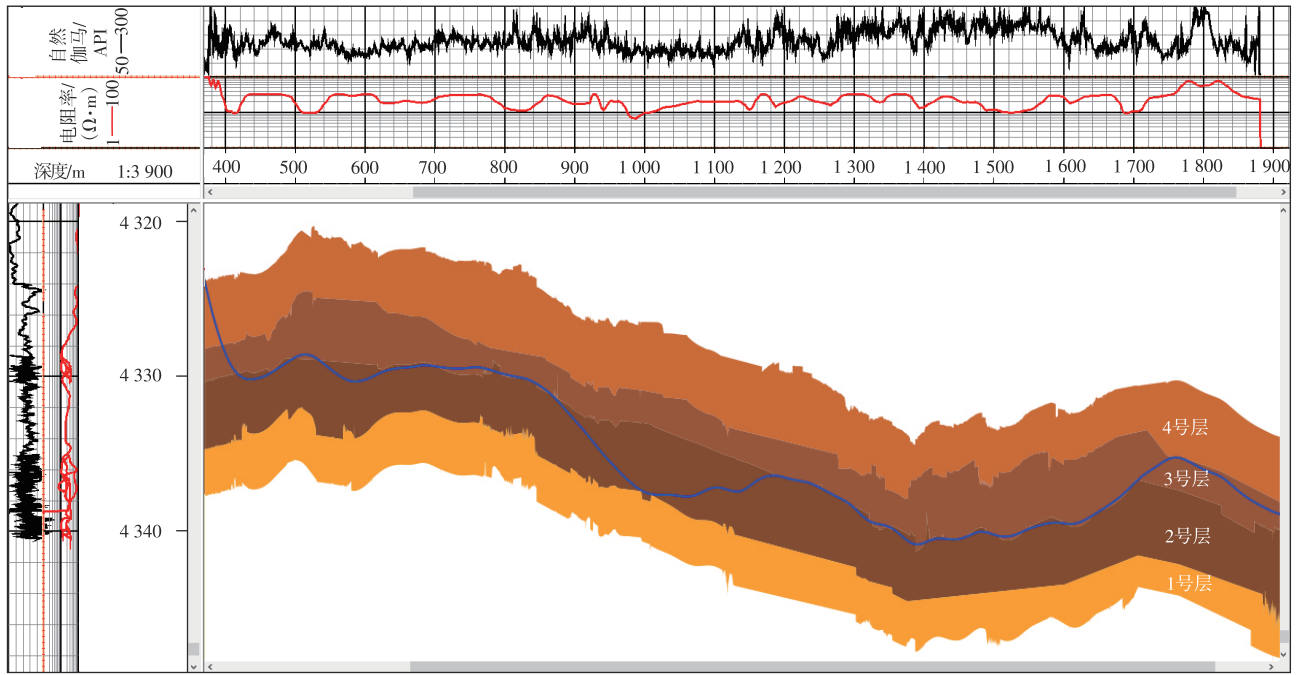
图9 四川盆地川中地区深层页岩气井DY1H井的导眼井曲线及初始地层模型

Fig. 9 Pilot well curves and initial stratigraphic model of deep shale gas well DY1H in central Sichuan Basin

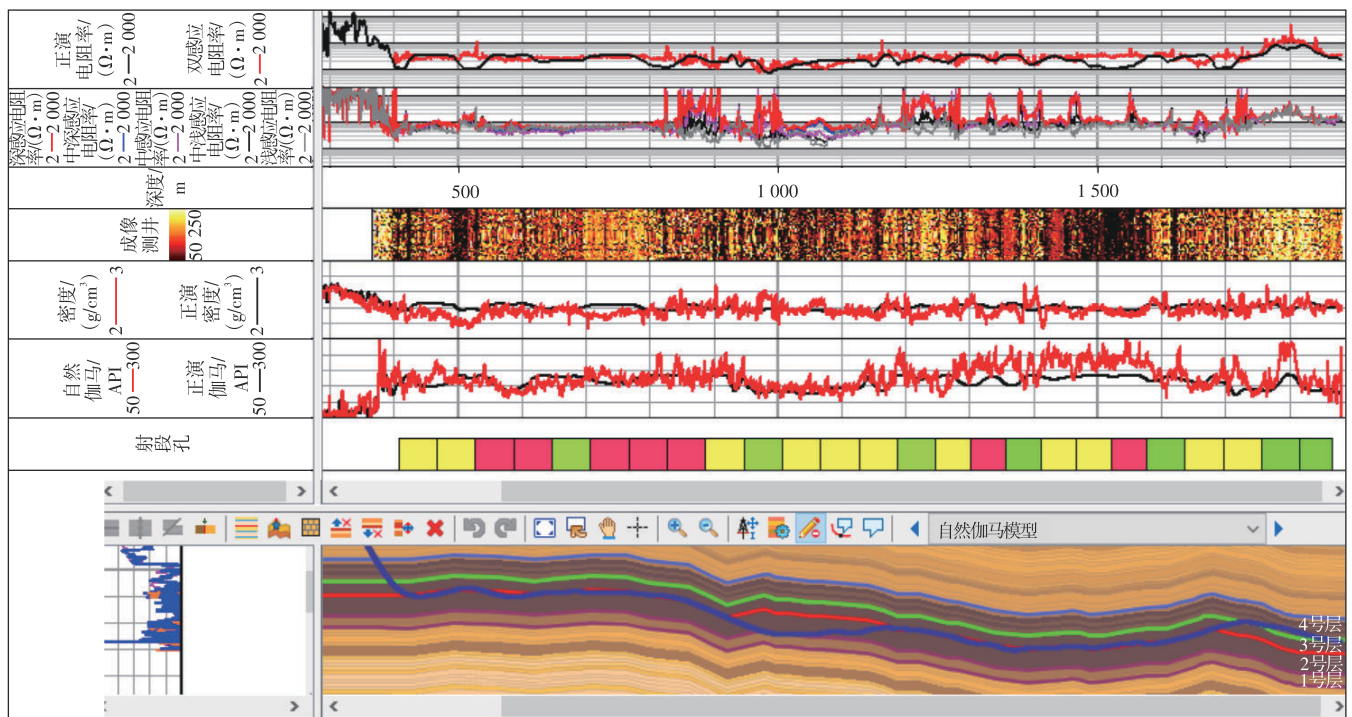
样本库,通过机器学习算法的训练和优化,仅需测井数据输入即可快速预测井地模型,展现了较高的自动化水平和稳定性。这表明该方法不仅适用于苏47井区,在不同地质背景下(如大页井区)也具有较强的泛化能力和可靠性。

2.3 实验结果讨论与方法适用性分析

通过对苏47井区平缓层序地层,以及DY井区复杂地层的实例研究,系统验证了基于机器学习与深度学习方法MLP的集成预测模型在不调整初始核心参数的情况下,对多区域地质背景的适应性、鲁棒性和预测精度。



a. 机器学习预测的水平井DY1H井地模型界面



b. 地质专家解释的水平井DY1H井地模型界面

图10 四川盆地川中地区深层页岩气井DY1H井机器预测井地模型界面与专家井地模型界面对比

Fig. 10 Comparison between machine-predicted and expert's well-formation model interfaces of deep shale gas well DY1H in central Sichuan Basin

不同地质背景的实验结果表明,该模型能够在较高的计算效率下,兼顾总体趋势捕捉和局部层界刻画,但其局限性和优化空间在处理复杂局部特征时仍需深度探讨。

在实验中,针对除 XGBoost 外的其他算法,如 RF、CatBoost 等机器学习方法,统一使用初始默认的参数设定,并未对特定区域或目标变量进行超参数的单独优化。以 RF 算法为例,针对 2 个井区未调整初始设置的树的深

度或分割策略,但依然能够展现良好的鲁棒性,特别是在地层情况复杂性较强的 DY 井区中,RF 算法通过树结构的集成投票机制,有效降低了因单一决策路径可能导致的误差放大。同时,CatBoost 算法凭借对类别特征的处理优势,能够在包括目标变量 $Y_1 \sim Y_3$ 和 $Z_1 \sim Z_5$ 在内的广泛任务中表现出较强的预测稳定性。

实验结果表明,各算法由于各自的特点在不同地质

背景下的表现各具优势。相比之下,SVM算法因其对高维特征的依赖性和对样本规模的敏感性,在本实验的适用性中表现有限,不适用于苏47井区和DY井区的预测任务。而LightGBM算法通过特征离散化优化模型效率的方法,则在目标层序变化相对平稳的苏47井区表现出较高的精度,能够通过优化层序边界,准确捕捉测井数据的局部特征波动。

RF算法和CatBoost算法的综合优势在层厚变化显著的DY井区得以进一步体现。实验中模型通过集成学习的方式处理层序变化显著的目标变量,不仅表现出准确的趋势预测能力,还显著提升了对主层界的稳健性。此外,多算法组合的进一步集成使RF算法和CatBoost算法能够携手有效捕捉岩性分布中的局部异常。

进一步分析苏47井区与DY井区实验的对比可以发现,尽管同一模型在无需修改初始设定参数的情况下适应了多种地质背景,但不同的地质复杂程度对模型的响应引发了性能差异。在数据特性较为均匀的苏47井区,集成模型与MLP结合后的表现几乎能够实现岩性界面的完整拟合;而在数据非线性依赖性更强的DY井区,虽然模型成功捕捉了主要趋势,但仍存在局部优化的潜力。特别是对于测井数据中异常特征的响应能力(如层交界处细小的厚度波动),当前的统一参数设置可能无法完全满足细节预测需求。

3 结论

针对水平井测井解释中井轨迹与地层空间几何关系的问题,提出了一种基于多模型集成和深度学习的智能测井解释方法。通过集成多种机器学习模型并结合多层感知机(MLP)神经网络,新建立的方法在实际地质应用中表现出较高的预测精度和鲁棒性。

实验结果和模型性能分析验证了该方法在多区域复杂地层下的可行性,得出以下主要研究结论和关键发现与改进方向。

1) 主要研究结论

① 研究采用多种机器学习模型的集成方法,有效综合了不同模型在特征处理上的互补优势。RF算法通过集成多棵决策树展现了对复杂非线性特征的强捕捉能力,而XGBoost算法和CatBoost算法则在提升模型效率和处理类别变量数据方面表现突出。在实验中,该集成方法能够较好地适应苏47井区的平缓层序和DY井区的裂缝性复杂层序,尤其在目标层的层序预测任务中,整体预测误差显著降低,展现了极高的泛化能力和稳定性。

② 深度学习有效提取非线性特征:通过MLP神经网络对多模型输出进行融合,进一步提升了对测井响应的

非线性特征捕捉能力,从而实现了复杂地质条件下更高的预测精度。其3层隐藏层设计(64、32、16节点递减结构)在泛化性能和收敛稳定性上表现良好,并通过ReLU激活函数与批量归一化的结合,增强了异质性强区域的特征提取能力。

③ 智能解释效率显著提升:在苏里格气田(苏47井区)和平缓层序,以及四川盆地(DY井区)裂缝性层序的实际应用中,该方法通过自动化和多任务处理能力,相比于传统手工逐段调剖方式,提高了预测效率,显著简化了解释过程。其快速部署能力能满足相似地质背景下多井批量解释的实际需求,对非常规油气开发中的自动化测井处理具有重要意义。

2) 关键发现与改进方向

① 机器学习算法在统一参数下展现适配性,但可进一步优化:机器学习模型参数的统一设置保证了算法在多区域的通用性,但DY井区裂缝层等地质特征分布差异较大的场景表明,动态优化模型参数(如RF算法的树深)或结合自动化调参技术(如自动搜索森林深度、动态调整学习率策略)可以进一步提升模型适用性。

② MLP的网络结构和激活函数优化潜力尚待挖掘:当前MLP已通过多模型输出的深度融合有效提升了目标特征的刻画能力,但未来可探索网络结构和激活函数改进方向,以进一步增强复杂区域的预测性能。多任务学习架构的引入也可能扩展其对多目标变量的建模能力。

③ 局地训练样本覆盖与损失权重优化:未来将增加对关键地质特征(如裂缝性区域、层厚快速变化地带)的代表性样本覆盖,以更好地覆盖复杂特征空间。同时,在损失函数优化过程中,可将裂缝区域、层界急变点与关键特征波动区域赋予更高的权重,从而提升模型对这些区域的敏感性。

④ 探索多数据源融合与自动化优化:引入AutoML框架,结合多区域背景的数据特征分析,有望实现适用于更复杂应用场景的自动化调参模型重训练。未来还将探索多源数据融合技术(如结合地震数据、测井数据、岩心数据共同分析),进一步提升模型的复杂地质背景适用性。

参考文献

- [1] 潘继平. 中国油气勘探开发新进展与前景展望[J]. 石油科技论坛, 2023, 42(1): 23-31.
PAN Jiping. New progress and outlook of China's oil and gas exploration and development[J]. Petroleum Science and Technology Forum, 2023, 42(1): 23-31.
- [2] 庞驰斌, 邵昕, 寇宇静. 水平井测井评价方法及应用探究[J]. 石化技术, 2022, 29(8): 88-90.
PANG Chiyu, SHAO Xin, KOU Yujing. Exploration on logging

- evaluation method and application of horizontal wells[J]. *Petrochemical Industry Technology*, 2022, 29(8): 88-90.
- [3] 陈猛, 谢韦峰, 张煜, 等. 水平井油水两相流阵列电磁波持水率计算方法及应用[J]. *油气藏评价与开发*, 2023, 13(4): 505-512.
CHEN Meng, XIE Weifeng, ZHANG Yu, et al. Methods and application for water holdup calculation and flowing image based on array electromagnetic wave instrument in horizontal water-oil wells [J]. *Petroleum Reservoir Evaluation and Development*, 2023, 13(4): 505-512.
- [4] 卢比, 胡春锋, 马军. 南川页岩气田压裂水平井间干扰影响因素及对策研究[J]. *油气藏评价与开发*, 2023, 13(3): 330-339.
LU Bi, HU Chunfeng, MA Jun. Influencing factors and countermeasures of inter-well interference of fracturing horizontal wells in Nanchuan shale gas field[J]. *Petroleum Reservoir Evaluation and Development*, 2023, 13(3): 330-339.
- [5] 邱小雪, 钟光海, 李贤胜, 等. 不同井斜页岩气水平井流动特征的CFD模拟研究[J]. *油气藏评价与开发*, 2023, 13(3): 340-347.
QIU Xiaoxue, ZHONG Guanghai, LI Xiansheng, et al. CFD simulation of flow characteristics of shale gas horizontal wells with different inclination[J]. *Petroleum Reservoir Evaluation and Development*, 2023, 13(3): 340-347.
- [6] 刘伟男, 张超谟, 朱林奇, 等. 页岩气水平井TOC测井评价新方法[J]. *物探与化探*, 2021, 45(2): 423-431.
LIU Weinan, ZHANG Chaomo, ZHU Linqi, et al. A new method for TOC logging evaluation in shale gas for horizontal well[J]. *Geophysical and Geochemical Exploration*, 2021, 45(2): 423-431.
- [7] 汪中浩, 易觉非, 赵乾富, 等. 水平井测井资料地质解释应用[J]. *江汉石油学院学报*, 2004, 26(3): 70-72.
WANG Zhonghao, YI Juefei, ZHAO Qianfu, et al. Application of geologic interpretation of horizontal well logging data[J]. *Journal of Jiangnan Petroleum Institute*, 2004, 26(3): 70-72.
- [8] 吕萍, 张永敏. 水平井咨询系统的原理与应用[J]. *测井技术*, 2004, 28(5): 455-457.
LU Ping, ZHANG Yongmin. Theory and application of horizontal well advisement system[J]. *Well Logging Technology*, 2004, 28(5): 455-457.
- [9] 胡松, 周灿灿, 王昌学, 等. 泥页岩油气藏水平井评价对策与实践[J]. *地球物理学进展*, 2013, 28(4): 1877-1885.
HU Song, ZHOU Cancan, WANG Changxue, et al. The strategies and practice of horizontal well evaluation in shale hydrocarbon reservoir [J]. *Progress in Geophysics*, 2013, 28(4): 1877-1885.
- [10] 苏林帅, 蔡明, 郑占树, 等. 井眼扩径对水平井声波测井响应影响的数值模拟[J]. *物探与化探*, 2022, 46(2): 467-473.
SU Linshuai, CAI Ming, ZHENG Zhanshu, et al. Numerical simulation of the effects of borehole enlargement on sonic logging response of horizontal wells[J]. *Geophysical and Geochemical Exploration*, 2022, 46(2): 467-473.
- [11] LIU J J, LIU J C. Integrating deep learning and logging data analytics for lithofacies classification and 3D modeling of tight sandstone reservoirs[J]. *Geoscience Frontiers*, 2022, 13(1): 101311.
- [12] LIAO W, GAO C, FANG J, et al. A TCN-BiGRU density logging curve reconstruction method based on multi-head self-attention mechanism[J]. *Processes*, 2024, 12(8): 1589.
- [13] 范翔宇, 孟凡, 赵鹏斐, 等. 一种基于自动机器学习的声波测井曲线重构方法: CN118534544A[P]. 2024-08-23.
- [14] 刘军, 钟洁, 倪振, 等. 基于机器学习的低含油饱和度砂岩储层参数预测: 以准噶尔盆地夏子街油田夏77井区下克拉玛依组为例[J]. *石油实验地质*, 2024, 46(5): 1123-1134.
LIU Jun, ZHONG Jie, NI Zhen, et al. Machine learning - based prediction of low oil saturation sandstone reservoir parameters: A case study of Lower Karamay Formation in Xia 77 well block of Xiazijie Oilfield, Junggar Basin[J]. *Petroleum Geology & Experiment*, 2024, 46(5): 1123-1134.
- [15] 武娟, 罗仁泽, 雷璨如, 等. 基于大语言模型的致密砂岩储层测井含水饱和度预测[J]. *天然气工业*, 2024, 44(9): 77-87.
WU Juan, LUO Renze, LEI Canru, et al. Prediction of water saturation in tight sandstone reservoirs from well log data based on the large language models(LLMs)[J]. *Natural Gas Industry*, 2024, 44(9): 77-87.
- [16] 杨春生, 焦艳丽, 蔡东梅, 等. 井-震结合曲流型储层构型研究及应用: 以大庆长垣西部斜坡区为例[J]. *石油地球物理勘探*, 2024, 59(6): 1330-1339.
YANG Chunsheng, JIAO Yanli, CAI Dongmei, et al. Meandering reservoir architecture based on well-seismic analysis and its application: Taking the western slope area of Daqing Oilfield as an example[J]. *Oil Geophysical Prospecting*, 2024, 59(6): 1330-1339.
- [17] 程敏华, 孟德伟, 王丽娟, 等. 致密砂岩气藏水平井差异化开发效果评价: 以鄂尔多斯盆地苏里格气田为例[J]. *中国矿业大学学报*, 2023, 52(2): 354-363.
CHENG Minhua, MENG Dewei, WANG Lijuan, et al. Evaluation of horizontal well differential development effect in tight sandstone gas reservoir: A case study of Sulige gas field, Ordos Basin[J]. *Journal of China University of Mining & Technology*, 2023, 52(2): 354-363.
- [18] 冀光, 贾爱林, 孟德伟, 等. 大型致密砂岩气田有效开发与提高采收率技术对策: 以鄂尔多斯盆地苏里格气田为例[J]. *石油勘探与开发*, 2019, 46(3): 602-612.
JI Guang, JIA Ailin, MENG Dewei, et al. Technical strategies for effective development and gas recovery enhancement of a large tight gas field: A case study of Sulige gas field, Ordos Basin, NW China[J]. *Petroleum Exploration and Development*, 2019, 46(3): 602-612.
- [19] 韩扬. 数据清洗和支持向量机技术研究[D]. 天津: 河北工业大学, 2015.
HAN Yang. Research on technologies of data cleaning and support vector machine[D]. Tianjin: Hebei University of Technology, 2015.
- [20] 张豹. 基于机器学习算法的数据分类与标准化方法研究[J]. *信息与电脑(理论版)*, 2024, 36(6): 53-55.
ZHANG Bao. Research on data classification and standardization methods based on machine learning algorithms[J]. *Information & Computer*, 2024, 36(6): 53-55.
- [21] 江丽, 张智谟, 王琦玮, 等. 基于不同机器学习模型的石油测井数据岩性分类对比研究[J]. *物探与化探*, 2024, 48(2): 489-497.
JIANG Li, ZHANG Zhimo, WANG Qiwei, et al. Comparative study on lithology classification of oil logging data based on different machine learning models[J]. *Geophysical and Geochemical Exploration*, 2024, 48(2): 489-497.
- [22] HUANG G B, CHEN L, SIEW C K. Universal approximation using incremental constructive feedforward networks with random hidden nodes[J]. *IEEE Transactions on Neural Networks*, 2006, 17(4):

- 879-892.
- [23] SHI S Z, SHI G F, PEI J B, et al. Porosity prediction in tight sandstone reservoirs based on a one - dimensional convolutional neural network - gated recurrent unit model[J]. Applied Geophysics, 2023.
- [24] ROSTAMI A, BAGHBAN A, MOHAMMADI A H, et al. Rigorous prognostication of permeability of heterogeneous carbonate oil reservoirs: Smart modeling and correlation development[J]. Fuel, 2019, 236: 110-123.
- [25] 李宁, 徐彬森, 武宏亮, 等. 人工智能在测井地层评价中的应用现状及前景[J]. 石油学报, 2021, 42(4): 508-522.
- LI Ning, XU Binsen, WU Hongliang, et al. Application status and prospects of artificial intelligence in well logging and formation evaluation[J]. Acta Petrolei Sinica, 2021, 42(4): 508-522.
- [26] WU J, YIN X, XIAO H. Seeing permeability from images: Fast prediction with convolutional neural networks[J]. Science Bulletin, 2018, 63(18): 1215-1222.
- [27] 黎子豪, 蒋恕. 基于机器学习和SHAP算法的声波测井曲线重构及可解释性分析[J]. 地质科技通报, 2025, 44(1): 321-331.
- LI Zihao, JIANG Shu. Reconstructing and interpreting analysis of sonic logging curves based on machine learning and SHAP algorithm [J]. Bulletin of Geological Science and Technology, 2025, 44(1): 321-331.
- [28] 田仁飞, 李山, 刘涛, 等. 基于XGBoost算法的 v_p/v_s 预测及其在储层检测中的应用[J]. 石油地球物理勘探, 2024, 59(4): 653-663.
- TIAN Renfei, LI Shan, LIU Tao, et al. v_p/v_s prediction based on XGBoost algorithm and its application in reservoir detection[J]. Oil Geophysical Prospecting, 2024, 59(4): 653-663.
- [29] 张家臣, 邓金根, 谭强, 等. 基于XGBoost的测井曲线重构方法[J]. 石油地球物理勘探, 2022, 57(3): 697-705, 496.
- ZHANG Jiachen, DENG Jingen, TAN Qiang, et al. Reconstruction of well logs based on XGBoost[J]. Oil Geophysical Prospecting, 2022, 57(3): 697-705, 496.
- [30] FENG R, GRANA D, BALLING N. Imputation of missing well log data by random forest and its uncertainty analysis[J]. Computers & Geosciences, 2021, 152: 104763.
- [31] XIA J, ZHANG S, CAI G, et al. Adjusted weight voting algorithm for random forests in handling missing values[J]. Pattern Recognition, 2017, 69: 52-60.
- [32] 韩如冰, 高严, 张元峰. 基于BP神经网络的分层相控碳酸盐岩储层渗透率预测方法[J]. 中国海上油气, 2024, 36(1): 100-108.
- HAN Rubing, GAO Yan, ZHANG Yuanfeng. Intelligent prediction method for permeability of layered phase controlled carbonate reservoirs based on BP neural network[J]. China Offshore Oil and Gas, 2024, 36(1): 100-108.
- [33] XIANG M, QIN P, ZHANG F. Research and application of logging lithology identification for igneous reservoirs based on deep learning [J]. Journal of Applied Geophysics, 2020, 173: 103929.
- [34] 陈芊澍, 文晓涛, 何健, 等. 基于极限学习机的裂缝带预测[J]. 石油物探, 2021, 60(1): 149-156.
- CHEN Qianshu, WEN Xiaotao, HE Jian, et al. Prediction of a fracture zone using an extreme learning machine[J]. Geophysical Prospecting for Petroleum, 2021, 60(1): 149-156.
- [35] 邓少贵, 张凤姣, 陈前, 等. 基于混合机器学习算法的页岩薄互层识别方法[J]. 石油学报, 2023, 44(7): 1097-1104.
- DENG Shaogui, ZHANG Fengjiao, CHEN Qian, et al. Identification of shale thin interbeds based on hybrid machine learning algorithm [J]. Acta Petrolei Sinica, 2023, 44(7): 1097-1104.
- [36] HUANG G B, ZHOU H, DING X, et al. Extreme learning machine for regression and multiclass classification[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 2012, 42(2): 513-529.
- [37] 高飞, 曲志鹏, 魏震, 等. 基于机器学习方法的测井岩相分类研究 [J]. 地球物理学进展, 2024: 39(3): 1173-1192.
- GAO Fei, QU Zhipeng, WEI Zhen, et al. Study on well-log lithofacies classification based on machine learning methods[J]. Progress in Geophysics, 2024: 39(3): 1173-1192.
- [38] 李盼池, 李文杰. 基于一维卷积神经网络的岩石物理相识别[J]. 吉林大学学报(信息科学版), 2022, 40(1): 51-63.
- LI Panchi, LI Wenjie. Identification of petrophysical facies based on one-dimensional convolutional neural networks[J]. Journal of Jilin University(Information Science Edition), 2022, 40(1): 51-63.
- [39] 刘学锋, 张晓伟, 曾鑫, 等. 采用机器学习分割算法和扫描电镜分析页岩微观孔隙结构[J]. 中国石油大学学报(自然科学版), 2022, 46(1): 23-33.
- LIU Xuefeng, ZHANG Xiaowei, ZENG Xin, et al. Pore structure characterization of shales using SEM and machine learning-based segmentation method[J]. Journal of China University of Petroleum (Edition of Natural Science), 2022, 46(1): 23-33.
- [40] AO Y, ZHU L, GUO S, et al. Probabilistic logging lithology characterization with random forest probability estimation[J]. Computers & Geosciences, 2020, 144: 104556.
- [41] DURGA KANNAIAH P V, MAURYA N K. Machine learning approaches for formation matrix volume prediction from well logs: Insights and lessons learned[J]. Geoenergy Science and Engineering, 2023, 229: 212086.
- [42] 王嵩然, 李潮流, 刘英明, 等. 高阻地层HDIL阵列感应影响因素分析与校正方法[J]. 石油科学通报, 2024, 9(1): 50-61.
- WANG Songran, LI Chaoliu, LIU Yingming, et al. Analysis of influence factors and method of correction in high resistivity formation by a high definition induction log[J]. Petroleum Science Bulletin, 2024, 9(1): 50-61.

(编辑 顾丽媛)