



引文格式: 兰孟平, 唐旭, 蒿多民, 等. 基于自适应机器学习的油田注采方法优化及其应用[J]. 西北地质, 2026, 59(4): 122–132. DOI: 10.12401/j.nwg.2026042

Citation: LAN Mengping, TANG Xu, HAO Duomin, et al. Optimization of Oilfield Injection-Production Methods Based on Adaptive Machine Learning and Its Application[J]. Northwestern Geology, 2026, 59(4): 122–132. DOI: 10.12401/j.nwg.2026042

基于自适应机器学习的油田注采方法优化及其应用

兰孟平^{1,2}, 唐旭^{1,*}, 蒿多民³, 胡铭源², 牛云鹏², 吴清燕⁴

(1. 中国石油大学(北京)经济管理学院, 北京 100100; 2. 西安洛科电子科技股份有限公司, 陕西 西安 710076; 3. 中国石油集团西部钻探工程有限公司试油公司, 新疆 克拉玛依 834000; 4. 海南热带海洋学院商学院, 海南 三亚 572022)

摘要: 随着油田开发进入中后期, 地层能量亏空严重, 注采矛盾日益突出, 传统的经验配注及数值模拟方法在应对非均质性强、动态变化快的复杂油藏时, 常面临计算耗时长、历史拟合困难及预测精度不足等挑战。基于上述问题, 笔者提出一种融合生产动态数据、物理约束与自适应机器学习的注采优化方法。该方法以 LSTM 和 XGBoost 构建并行预测框架, 通过多指标竞争选择预测模型; 以卡尔曼滤波和非线性扩散滤波重构注采响应信号, 结合物理约束 ANN 和梯度敏感性分析反演井间连通系数; 在此基础上, 建立兼顾产油量提高与含水率控制的双目标配注优化模型, 并采用差分进化算法求解。B 采油厂实例表明, 物理约束 LSTM 的 R^2 达到 0.952 1, MSE 降至 0.651 3, 预测精度和物理一致性均优于标准 LSTM 和 XGBoost; 井间连通性反演结果能够识别高含水优势通道并与地质认识相吻合; 优化后区块产油量由 $175.00 \text{ m}^3/\text{d}$ 提高至 $176.10 \text{ m}^3/\text{d}$, 日增油为 $1.10 \text{ m}^3/\text{d}$ 。研究表明, 数据驱动模型与渗流物理约束耦合, 可减少反复历史拟合的依赖, 提高注采方案评价效率, 为高含水后期油田精细配注和稳油控水提供技术支撑。

关键词: 机器学习; 产量预测; 井间连通性; 卡尔曼滤波; 差分进化算法; 注采优化

中图分类号: P618.13; TE33+1

文献标志码: A

文章编号: 1009-6248(2026)04-0122-11

Optimization of Oilfield Injection-Production Methods Based on Adaptive Machine Learning and Its Application

LAN Mengping^{1,2}, TANG Xu^{1,*}, HAO Duomin³, HU Mingyuan², NIU Yunpeng², WU Qingyan⁴

(1. China University of Petroleum (Beijing), Beijing 100100, China; 2. Xi'an Luoke Electronic Technology Co., Ltd., Xi'an 710076, Shaanxi, China; 3. Testing Company, CNPC Western Drilling Engineering Co., Ltd., Karamay 834000, Xinjiang, China; 4. Hainan Tropical Ocean University, Sanya 572022, Hainan, China)

Abstract: As oilfield development enters the middle and late stages, formation energy deficit becomes increasingly severe and injection-production conflicts become more prominent. Conventional empirical injection alloca-

收稿日期: 2026-06-15; 修回日期: 2026-06-25; 责任编辑: 吕鹏瑞

基金项目: 国家自然科学基金项目(72174206)资助。

作者简介: 兰孟平(1985–), 男, 博士, 高级工程师, 从事油气田数智化注采联动系统研究。E-mail: 273402908@qq.com。

* 通讯作者: 唐旭(1985–), 男, 教授, 博士生导师, 从事资源环境政策与管理研究。E-mail: tangxu2001@163.com。

© Editorial Office of *Northwestern Geology*. OA under CC BY-NC-ND

tion and numerical simulation methods often face challenges such as high computational cost, difficult history matching, and insufficient prediction accuracy when applied to complex reservoirs with strong heterogeneity and rapid dynamic changes. To address these problems, an injection-production optimization method integrating production dynamic data, physical constraints, and adaptive machine learning is proposed. In this method, LSTM and XGBoost are used to construct a parallel prediction framework, and the prediction model is selected through multi-index competition. Kalman filtering and nonlinear diffusion filtering are adopted to reconstruct injection-production response signals, and a physics-constrained ANN combined with gradient sensitivity analysis is used to invert inter-well connectivity coefficients. On this basis, a dual-objective injection allocation optimization model considering both oil production enhancement and water-cut control is established and solved using the differential evolution algorithm. A case study from B Oil Production Plant shows that the physics-constrained LSTM achieves an R^2 of 0.952 1 and reduces the MSE to 0.651 3, outperforming the standard LSTM and XGBoost in both prediction accuracy and physical consistency. The inter-well connectivity inversion results identify dominant high-water-cut flow channels and agree well with geological understanding. After optimization, block oil production increases from 175.00 m³/d to 176.10 m³/d, with an incremental oil production of 1.10 m³/d. The results indicate that coupling data-driven models with seepage-physics constraints can reduce reliance on repeated history matching, improve the evaluation efficiency of injection-production schemes, and provide technical support for refined injection allocation and stable oil production with water control in high-water-cut oil fields.

Keywords: machine learning; production prediction; inter-well connectivity; Kalman filter; differential evolution algorithm; injection-production optimization

机器学习作为人工智能的重要分支,已在图像识别、自然语言处理等领域取得突破性进展。近年来,随着“数字油田”建设的推进,海量的生产动态数据为数据驱动的研究范式提供了契机,机器学习技术逐步应用于油气勘探开发领域,为解决复杂油藏工程问题提供了新的技术方法(张凯等, 2020; 匡立春等, 2021; 闵超等, 2024)。基于机器学习的剩余油分布预测新方法给出 12 个维度参数的具体计算方法(谷建伟等, 2020)。人工智能方法对于测井数据分析具有较好的适用性,开发结合高端测井资料(如成像测井资料)与常规测井曲线联合解释的智能解释流程,可更好地提升岩相分类及参数反演的效果(王华等, 2021; 程希等, 2023; 徐碧兰等, 2025)。相对于 BP (Back Propagation)神经网络等传统机器学习算法模型,深度森林模型可以准确预测油井产量,且具有算法参数少、调参及应用简单等优势(薛永超等, 2022)。人工智能方法与油藏工程交叉融合辅助注水开发方案调整存在一些核心问题,未来注水开发方案优化的研究重点将聚焦于“动态数据+物理约束+人工智能算法”的深度融合(刘合等, 2023)。然而,现有机器学习在油田注采管理中的应用仍存在明显不足。①预

测模型多忽略井间时间延迟效应与多层位非均质性,导致产量预测偏差较大;②优化算法在注水分配中难以动态响应井间连通性的实时变化;③系统层面缺乏自动化的定时调度机制,无法灵活适应数据更新与生产需求波动。

在油田开发中,注水采油是常用的增产技术,注水量分配和油井产量的动态变化受地层特性、井间连通性及时间延迟效应等多种复杂因素影响。传统的注采优化方法多依赖经验公式或简单的数学模型,难以准确预测产量趋势或优化注水策略。在时间序列产量预测方面,长短期记忆网络(Long Short-Term Memory, LSTM)凭借其独特的门控机制(任杉等, 2025),能够有效捕捉产油序列的长期依赖关系。LSTM 应用于油田产量预测,验证了其在处理非线性时序数据方面的优越性(侯春华, 2019)。相对于反向传播(Back Propagation, BP)神经网络等传统机器学习算法(李培月等, 2025),深度森林模型在油井产量预测中具有算法参数少、调参简单等优势,且预测精度更高(薛永超等, 2022)。此外, XGBoost(Xtreme Gradient Boosting)等集成学习方法因具备出色的非线性拟合能力和特征重要性分析功能,也被广泛应用于产油

量预测研究。在注采参数优化算法层面,差分进化(Differential Evolution, DE)算法、粒子群优化(Particle Swarm Optimization, PSO)算法等智能优化方法已被应用于注水量分配优化问题。然而,现有算法在适应性方面仍存在一些不足。①物理一致性不足,现有机器学习预测模型多将注采问题视为纯数据拟合任务,缺乏对渗流力学方程的显式约束,预测结果可能出现违反物质守恒定律的“伪解”,降低了模型的工程可信度;②优化目标单一,现有注采优化算法多以产油量最大化为单一目标,难以兼顾含水率控制、注水成本约束等实际生产需求,未能实现“稳油控水”的多目标平衡。

针对上述问题,笔者围绕“产量预测-连通性反演-配注优化”的技术主线开展相关研究。①构建 LSTM 与 XGBoost 并行的产量预测模型,并引入物理约束损失项提高预测结果的工程合理性;②采用卡尔曼滤波与非线性扩散滤波对注采动态信号进行分阶段重构,在此基础上利用自适应 ANN 和梯度敏感性分析定量表征井间连通关系;③将连通系数、含水率和配注约束耦合到双目标优化模型中,通过差分进化算法形成分层配注调整方案。该研究重点在于将数据驱动预测、渗流物理约束和优化决策进行闭环耦合,为高含水后期油田注采结构精细调整提供可解释、可计算和可应用的技术流程。

1 研究方法

1.1 产量预测模型

油水井生产过程中会输出海量的动态参数,这些数据表现出典型的非线性特征。为捕捉注水采油过程中复杂的时序依赖与非线性特征,此处构建了两种并行的预测模型,并通过竞争机制自动选择最优者作为最终预测引擎。

(1)LSTM 网络模型:作为循环神经网络的改进变体, LSTM 由遗忘门、输入门和输出门组成的门控机制,有效解决了传统 RNN 在处理长序列数据时的梯度消失问题,能够跨越数天至数月的时滞,捕捉注水脉冲与油井产量响应之间的动态关系。基于 PyTorch 框架,搭建 2 层堆叠式 LSTM,每层配置 64 个隐藏神经元,引入 Dropout 层(比率为 0.2)防止过拟合,采用 Adam 优化器与 MSE 损失函数,超参数通过麻雀搜索算法(SSA)自动寻优。各门控信号的计算过程表

达为公式(1)~公式(3)。

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (3)$$

式中: f_t 、 i_t 和 o_t 分别为 t 时刻遗忘门、输入门和输出门的激活向量; σ 为 Sigmoid 激活函数; W_f 、 W_i 和 W_o 分别为各门控对应的权重矩阵; h_{t-1} 为上一时刻状态; x_t 为 t 时刻输入特征向量; b_f 、 b_i 和 b_o 分别为各门控偏置项。

(2)XGBoost 集成学习模型: XGBoost 是基于梯度提升决策树的高效实现(林琴等, 2024), 通过串行迭代拟合前一棵树的残差, 利用二阶泰勒展开加速收敛, 其内置的正则化项(L1/L2)可有效控制模型复杂度(图 1)。目标函数近似为公式(4)。

$$\mathcal{L}^{(t)} \approx \sum_{i=1}^n \left[g_i f_i(x_i) + \frac{1}{2} h_i f_i^2(x_i) \right] + \Omega(f_i) \quad (4)$$

式中: g_i 和 h_i 分别为损失函数的一阶和二阶导数; $\Omega(f_i)$ 为正则化项, 控制叶子节点个数和权重。XGBoost 超参数采用贝叶斯优化自动寻优。

(3)多模型竞争机制: 两种模型经超参数优化后, 在测试集上基于 MAPE、RMSE 和 R^2 进行综合评估, 竞争评分式为公式(5)。

$$S_{total} = w_1(1 - MAPE) + w_2 \left(1 - \frac{RMSE}{RMSE_{max}} \right) + w_3 R^2 \quad (5)$$

式中: w_1 、 w_2 、 w_3 为各指标权重, 满足和为 1; 根据综合评分自动择优部署。

(4)物理约束嵌入策略: 为避免纯数据驱动模型在稀疏区域产生违背物理规律的预测结果(如产油量为负), 将物质平衡方程作为软约束嵌入损失函数, 总损失函数定义为公式(6)。

$$L_{total} = L_{data} + \lambda \cdot L_{physics} \quad (6)$$

式中: L_{data} 为数据拟合项(MSE); $L_{physics}$ 为物理约束项。基于油相物质平衡方程构建表达为公式(7)和公式(8)。

$$L_{data} = \frac{1}{N} \sum_{i=1}^N (y_{pred,i} - y_{true,i})^2 \quad (7)$$

$$L_{physics} = \frac{1}{N} \sum_{i=1}^N \left[\frac{\partial(\phi \rho_o S_o)}{\partial t} + \nabla \cdot (\rho_o v_o) - q_o \right]^2 \quad (8)$$

式中: ϕ 为储层孔隙度(%); ρ_o 为地层原油密度(kg/m^3); S_o 为含油饱和度(%); v_o 为油相渗流速度(m/d); q_o 为质量源汇项密度($\text{kg/m}^3 \cdot \text{d}$); λ 为约束权重,

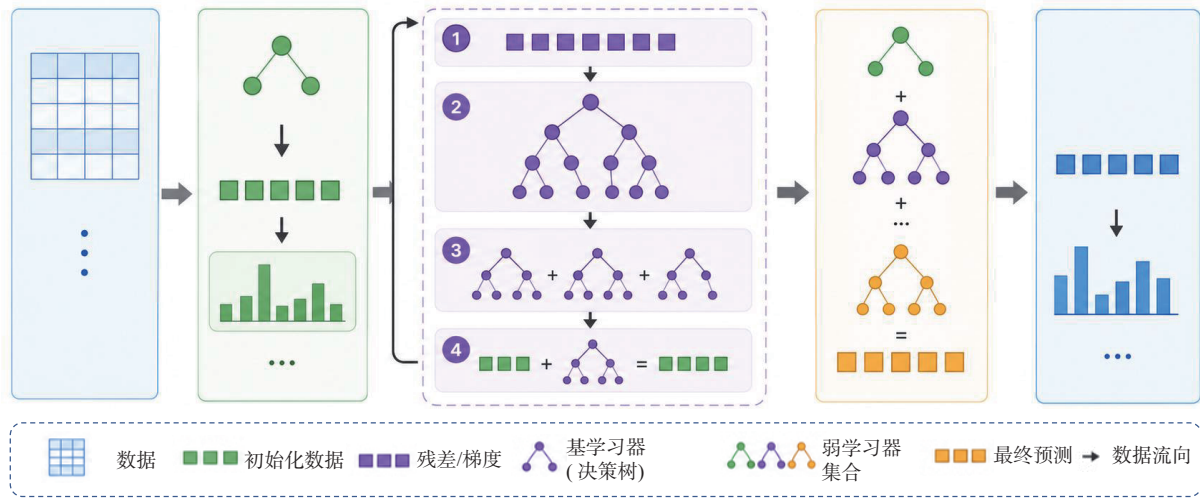


图1 XGBoost 模型结构图

Fig. 1 XGBoost model architecture diagram

无因次, $\lambda \in [0.001, 0.1]$ 。该范围依据数据拟合项与物理约束项量级匹配原则确定: 当 $\lambda < 0.001$ 时, 物理约束对损失函数贡献过小, 难以抑制非物理预测; 当 $\lambda > 0.1$ 时, 约束项可能主导优化过程, 导致模型对观测数据拟合不足。因此, 在该范围内采用验证集误差和物理残差共同约束的自适应调整策略, 以兼顾预测精度与物理一致性。

1.2 井间连通性反演方法

准确量化注水井与生产井之间的动态连通关系, 是优化注水分配的前提。采用数据驱动与物理约束相结合的方式反演。

(1) 卡尔曼滤波信号重构: 考虑到注入水在地层中的运移存在时间延迟与幅度衰减, 引入卡尔曼滤波对原始注水序列进行最优估计与平滑重构(图2)。状态方程与观测方程为公式(9)和公式(10)。

$$x_k = Fx_{k-1} + w_k \quad (9)$$

$$z_k = Hx_k + V_k \quad (10)$$

式中: x_k 为状态向量; F 为状态转移矩阵; H 为观测矩阵; w_k 和 V_k 分别为过程和观测噪声。通过“预测-更新”递推循环实现最优状态估计, 有效平滑注水和产液曲线, 去除测量误差。

(2) 非线性扩散滤波器信号重构: 针对注采过程中由于地层渗流时滞及压力传播衰减所导致的非线性动态特征, 采用非线性扩散滤波器对原始序列进行物理一致性重构(图3)。该方法基于渗流扩散思想, 将数据平滑过程类比为非均匀介质中的扩散过程, 通过引入梯度依赖的扩散系数, 实现对高频噪声抑制与

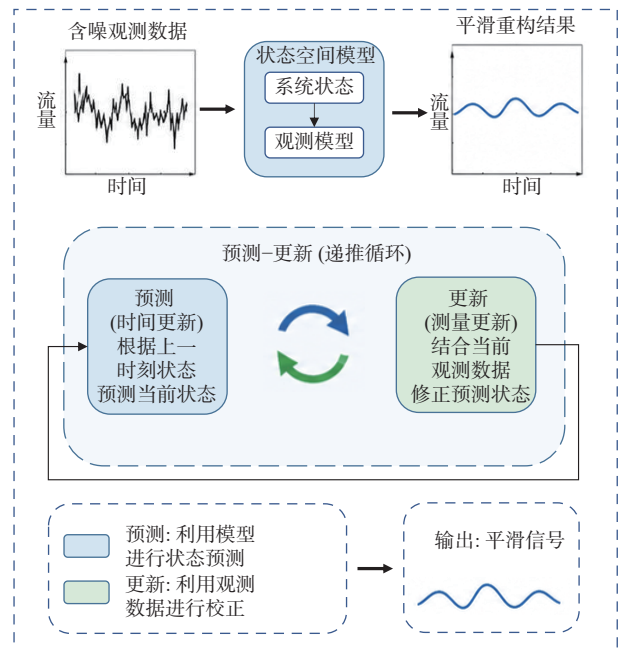


图2 卡尔曼滤波原理图

Fig. 2 Kalman filter schematic diagram

结构信息保留的统一, 其连续形式可表示为扩散偏微分方程(公式11)。

$$\frac{\partial u(x,t)}{\partial t} = \nabla \cdot (c(x,t) \nabla u(x,t)) \quad (11)$$

式中: $u(x,t)$ 表示待重构的注采信号; ∇ 为梯度算子; $c(x,t)$ 为扩散系数函数, 用于控制不同区域的平滑强度化(Perona et al., 1990)。

(3) 物理约束 ANN 反演模型: 构建以重构后注水量为输入、产油量为输出的 ANN(图4), 训练中嵌入物质平衡等物理约束, 通过 Optuna-TPE 算法动态寻优

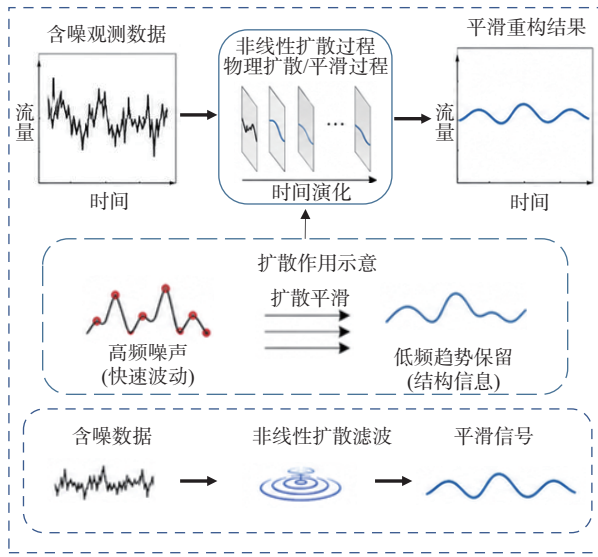


图3 非线性扩散滤波器原理图

Fig. 3 Nonlinear diffusion filter schematic diagram

网络架构(Raissi et al., 2019)。

(4)梯度敏感性分析与连通系数。网络训练完成后,利用自动微分技术计算输出对输入的偏导数作为敏感性指标,表达为公式(12)。

$$S_{ij} = \frac{\partial Q_{o,i}}{\partial Q_{inj,j}} \quad (12)$$

式中: S_{ij} 为第 j 口注水井对第 i 口生产井的产量敏感性。经 Min-Max 归一化后得到连通系数 $f_{ij} \in [0, 1]$, 划分为极强 (>0.8)、强 ($0.5 \sim 0.8$)、中等 ($0.2 \sim 0.5$)、弱 ($0.05 \sim 0.2$)、极弱 (<0.05) 5 级。该方法无需踪剂测试或历史拟合,实现连通性的可解释量化。

1.3 注采参数双层协同优化模型

在获得井间连通系数的基础上,建立多目标协同优化模型,综合考虑产油量最大化与产水量最小化(含水率控制)的双重诉求,完成注采参数协同优化。

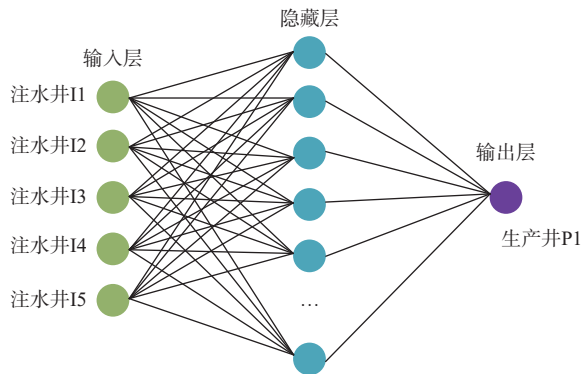


图4 人工神经网络(ANN)结构图

Fig. 4 Artificial neural network (ANN) architecture diagram

(1)双目标函数:传统单目标优化仅追求产油量最大化,可能导致含水率急剧攀升。构建如下双目标优化模型为公式(13)和公式(14)。

$$\max f_1 = \sum_{i=1}^M Q_{o,i} \quad (13)$$

$$\min f_2 = \sum_{i=1}^M Q_{w,i} \quad (14)$$

式中: f_1 为区块总产油量; f_2 为区块总产水量; $Q_{o,i}$ 和 $Q_{w,i}$ 分别为第 i 口生产井的产油量和产水量。约束条件:各注水井配注比例之和为 1,单井配注比例 $0.1 \leq a_j \leq 0.6$,总注水量为常数。

(2)差分进化算法求解:采用 DE 算法进行全局寻优,不依赖梯度信息,主要包含种群初始化、变异操作、交叉操作、选择操作等过程,并重复执行直至达到最大迭代次数。其中变异操作和交叉操作分别可表示为公式(15)和公式(16)。

$$v_i = X_{r1} + F \cdot (X_{r2} - X_{r3}) \quad (15)$$

$$u_{i,j} = \begin{cases} v_{i,j}, & \text{if } \text{rand}(0,1) \leq CR \text{ or } j = j_{\text{rand}} \\ x_{i,j}, & \text{otherwise} \end{cases} \quad (16)$$

式中: x_i 第 i 个个体; v_i 第 i 个个体对应的变异向量; u_i 第 i 个个体对应的交叉后的试验向量; $r1, r2, r3$ 为随机选择且互不相同的索引,满足 $r1, r2, r3 \neq i$ 。 $F \in [0.5, 0.9]$ 为缩放因子; $CR \in [0.7, 0.95]$ 为交叉概率。选择操作基于非支配排序与拥挤度距离,迭代直至收敛(Storn et al., 1997)。

(3)Pareto 前沿与方案优选。DE 收敛后生成 Pareto 最优解集,采用 TOPSIS 方法计算各方案相对贴近度,其计算方程表达为公式(17)。

$$C_i = D_i^- / (D_i^+ + D_i^-) \quad (17)$$

式中: D_i^+ 和 D_i^- 分别为到正负理想解的欧氏距离。选择 C_i 最大的方案作为推荐方案,在产油量最大化和含水率控制两个目标间取得最佳平衡。

2 实验设计

2.1 实验数据描述

本研究以 B 采油厂实际生产数据为基础开展实验验证。B 采油厂位于某盆地,构造类型为背斜构造岩性圈闭,储层埋深约为 2000 m,平均渗透率为 150 mD,平均孔隙度为 18%,原油黏度为 25 mPa·s,属常规稠油油藏。选取 5 口生产井和 4 口注水井作为研究对象,

原始数据集包含采集时间、流量、含水率、层段温度、油压、套压、油嘴开度、层段名称等物理参数。数据时间跨度为 2010 年至 2024 年,采集频率为每日一次,经清洗后有效样本总量约 5 000 条。

2.2 数据清洗与标准化

由于现场采集设备故障、信号传输中断等原因,原始数据中不可避免地存在缺失值和异常值。对于短时间的缺失,采用线性插值法利用前后时刻的数据进行填充;对于长时间缺失,结合邻井趋势进行趋势项补全。对于异常数据,利用 3σ 原则(拉依达准则)

识别离群点,当瞬时流量突变为负值或超出物理极限时,将其判定为异常值并剔除。此外,为消除不同物理量纲之间的数量级差异,采用 Min-Max 标准化方法(公式 18)将所有特征映射至 $[0, 1]$ 区间。

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (18)$$

2.3 影响因素相关性分析

特征选择是机器学习建模的关键步骤。结合 Pearson 相关系数对各因素进行量化分析(图 5),该系数取值范围为 $[-1, +1]$,计算公式为公式(19)。

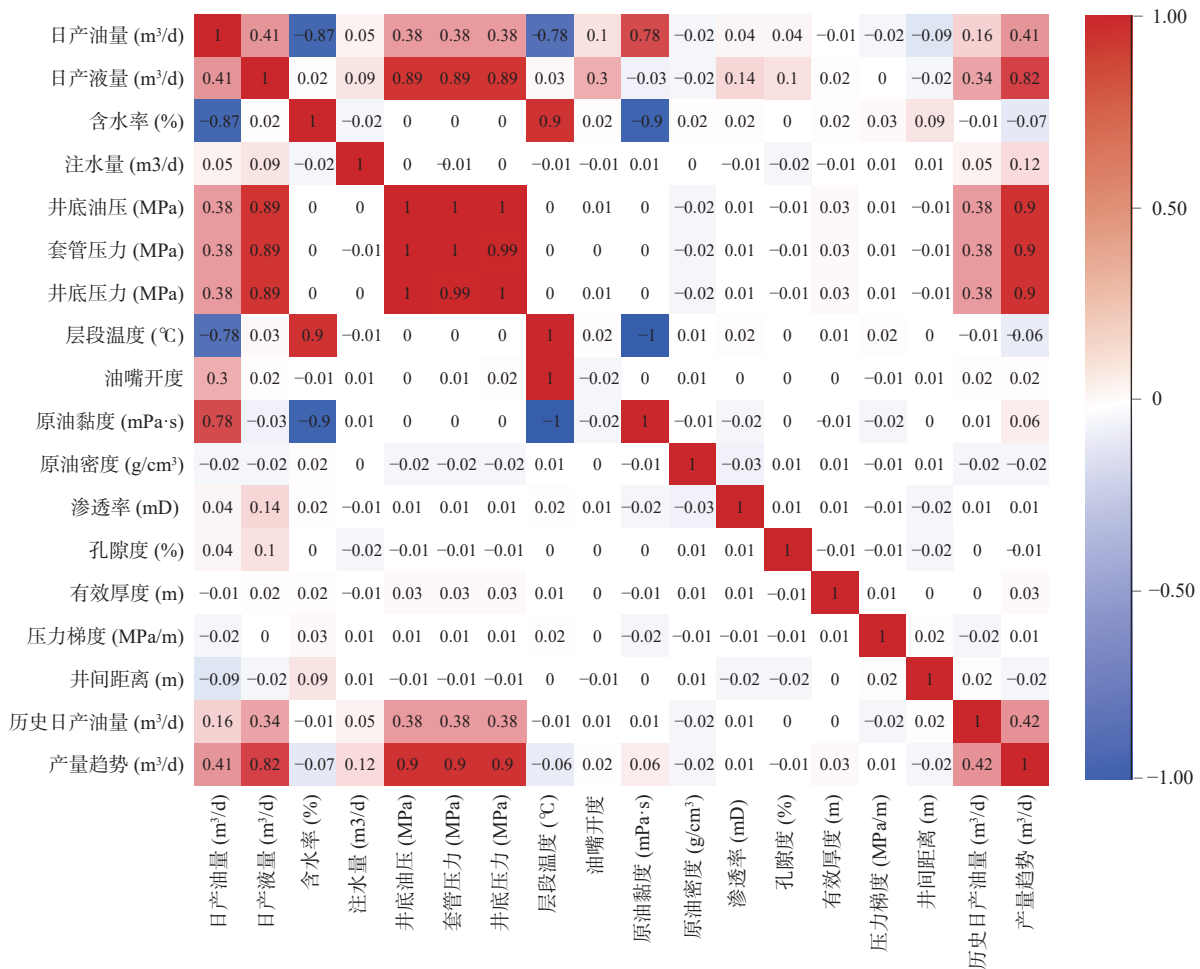


图5 基于 Pearson 的相关性分析

Fig. 5 Pearson correlation analysis

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (19)$$

式中: $\bar{x}$ 、 $\bar{y}$ 分别为 x 、 y 中元素的平均值。明显地, Pearson 相关系数 r_{xy} 是 $[-1, 1]$ 中的实数, 当 $r_{xy} > 0$ 时, 两变量正相关; 反之, 则负相关。 $|r_{xy}|$ 越大, x 和 y 的相关

程度越高。分析结果(图 5)表明: 油压与套压作为驱动流体流动的直接动力参数, 与产量呈现显著相关性; 层段温度反映了流体粘度的变化, 与产量存在较强的非线性关系; 油嘴开度直接控制井口流出阻力, 是产量的主控因素之一。

笔者从原始生产动态参数中初步识别与产量响

应相关的潜在影响因素,重点反映各变量与产量之间的相关方向和相关强度(图 5)。在各变量与产量相关性分析(图 5)基础上,结合模型训练过程中的特征重要性评价,对进入预测模型的候选变量进行进一步排序和筛选(图 6)。通过相关性分析,笔者确定以含水率、层段温度、原油粘度、井底油压、套管压力、油嘴开度作为输入特征向量(图 6),构建高维特征空间。

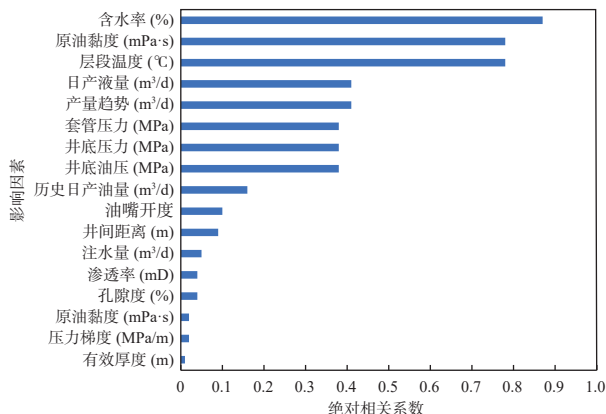


图6 候选输入特征重要性排序

Fig. 6 Importance ranking of candidate input features

2.4 时序特征工程构建

为充分发挥 LSTM 的性能,构建了丰富的高维时序特征空间:①滞后特征($t-1$ 、 $t-12$ 、 $t-24$),捕捉产量的瞬时连续性和周期性波动规律;②滚动统计特征(窗口 24 d、72 d 和 168 d 的移动均值与标准差),平滑高频噪声并揭示长期趋势;③滑动窗口机制(时间步长为 7 d),将数据重构为 $N \times 7 \times F$ 的三维张量作为模型输入,其中 F 为特征总数。

2.5 产量预测模型构建

2.5.1 LSTM 模型训练及优化

基于 PyTorch 深度学习框架,搭建输入层、2 层堆叠 LSTM 隐层(每层 64 神经元)、全连接输出层。采用 Adam 优化器与 MSE 损失函数,超参数通过麻雀搜索算法(SSA)寻优。训练过程中记录损失曲线(图 7)。

2.5.2 XGBoost 模型训练及优化

构建 XGBoost 模型作为 LSTM 的互补,通过串行构建多棵决策树拟合残差,引入 L1/L2 正则化防止过拟合,自动输出特征重要性排序。超参数采用贝叶斯优化寻优,记录损失曲线(图 8)。

2.6 基于信号处理的数据降噪

当产量预测模型显示注采效果不理想时,需深入剖析地下的井间连通状态。注采数据中往往夹杂高

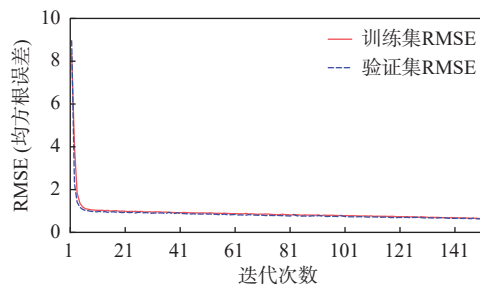


图7 LSTM 损失曲线

Fig. 7 LSTM loss curve

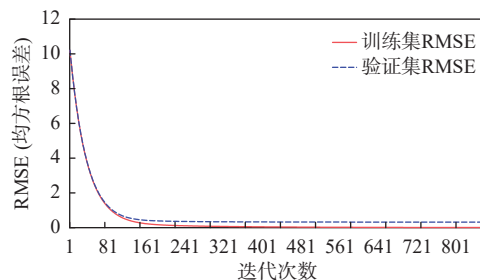


图8 XGBoost 损失曲线

Fig. 8 XGBoost loss curve

频噪声,且存在流体在地层中传播的滞后效应。采用“卡尔曼滤波-非线性扩散滤波”分阶段处理流程:卡尔曼滤波建立状态空间模型,利用“预测-更新”循环从含噪观测数据中估计系统真实状态,主要用于平滑注水和产液曲线并削弱测量误差;非线性扩散滤波器基于渗流扩散原理,体现压力传播时滞和幅度衰减,在保留有效动态结构的同时抑制非物理振荡。两者并非替代选择关系,而是前后衔接的联合预处理步骤,其输出作为 ANN 连通性反演模型的输入(图 9)。

3 结果与讨论

3.1 模型预测结果分析

XGBoost 模型和 LSTM 模型在整个测试集上的产量预测值与实际值的对比结果(图 10, 表 1)显示, XGBoost 的决定系数 R^2 为 0.940 3, 均方误差 MSE 为 0.719 1, 可解释方差值 EVS 为 0.940 3, 平均绝对误差 MAE 为 0.615 8; LSTM 的 R^2 为 0.926 5, MSE 为 0.885 6, EVS 为 0.926 8, MAE 为 0.647 1。物理约束 LSTM 在引入渗流力学先验约束后,进一步提升了模型性能,其 R^2 达到 0.952 1, MSE 降至 0.651 3, EVS 为 0.952 4, MAE 为 0.589 2, 各项指标均优于 XGBoost 和标准 LSTM。

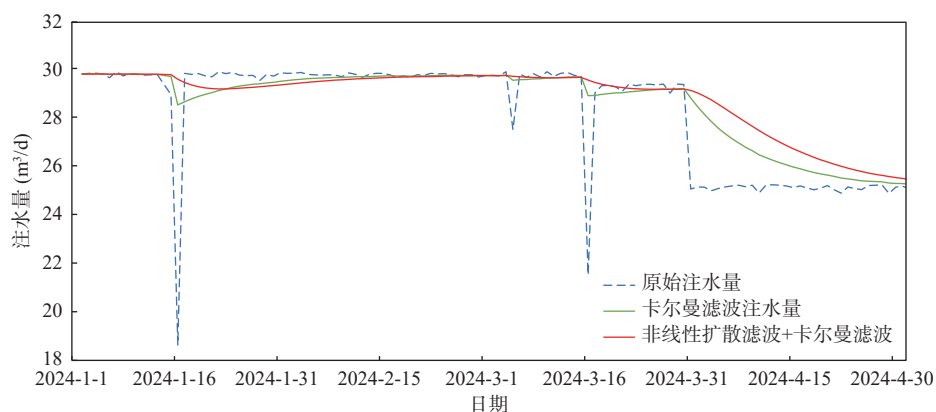
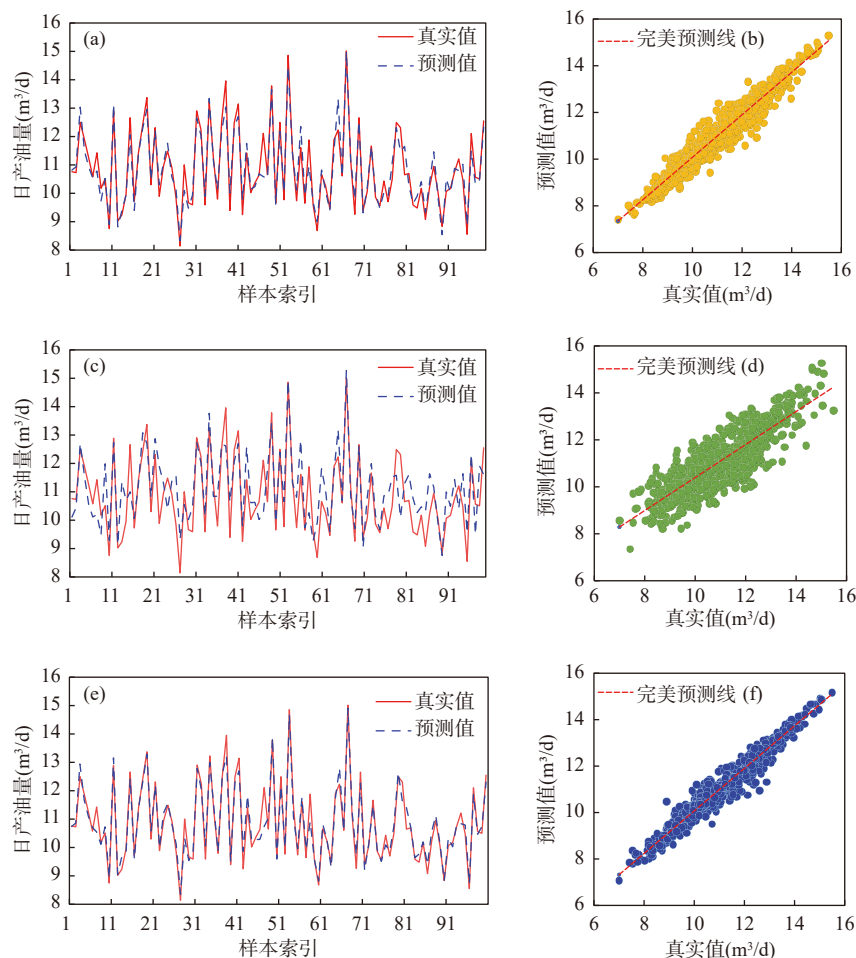


图9 卡尔曼滤波及非线性扩散滤波器

Fig. 9 Kalman filter and nonlinear diffusion filter

分析可知, XGBoost 基于梯度提升决策树的集成学习框架, 在处理结构化特征数据时展现出强大的非线性拟合能力, 能够高效捕捉注采参数与产量之间的

复杂映射关系。LSTM 模型通过门控机制对时序依赖进行建模, 在捕捉产量动态趋势方面具有一定优势, R^2 更高(达到 0.926 5), 但在局部极值点的拟合精度上



a. XGBoost 模型真实值与预测值对比图; b. XGBoost 模型真实值与预测值散点图; c. LSTM 模型真实值与预测值对比图; d. LSTM 模型真实值与预测值散点图; e. 物理约束 LSTM 真实值与预测值对比图; f. 物理约束 LSTM 真实值与预测值散点图

图10 LSTM、XGBoost 及物理约束的 LSTM 模型真实值预测值对比图

Fig. 10 Comparison of true and predicted values for LSTM, XGBoost, and Physics-constrained LSTM models

表 1 XGBoost 和 LSTM 产量预测效果对比

Tab. 1 Comparison of Production Prediction Performance between XGBoost and LSTM

预测方法	决定系数R ²	均方误差MSE	可解释方差值EVS	平均绝对误差MAE
XGBoost	0.940 3	0.719 1	0.940 3	0.615 8
LSTM	0.926 5	0.885 6	0.926 8	0.647 1
物理约束LSTM	0.952 1	0.651 3	0.952 4	0.589 2

略逊于 XGBoost。物理约束 LSTM 通过引入渗流力学先验约束损失项,在预测精度和物理一致性方面表现更优,R²提升至 0.952 1,显著改善了模型的外推能力和泛化性能,体现了物理信息嵌入对提升模型可靠性的重要作用。XGBoost 预测曲线与实际值的整体贴合度最高,尤其在产量平稳期拟合效果优异;LSTM 在产量突变阶段的响应存在一定的相位滞后;物理约束 LSTM 的预测结果在物理合理性方面显著改善,主要表现为产量变化趋势更符合达西渗流规律,避免了标准 LSTM 中出现的非物理振荡和负产量异常,同时在注采参数突变时的响应更加平滑合理。建议优先选用物理约束 LSTM 模型进行实际油藏产量预测,以兼顾高精度和物理可解释性。

3.2 井间连通性反演结果

当产量预测模型显示注采效果不理想时,需深入剖析地下井间连通状态。注采数据实际上包含了注水信号在地层中的响应,但数据中往往夹杂高频噪声,且存在流体在地层中传播的滞后效应。卡尔曼滤波建立状态空间模型,利用“预测-更新”循环从含噪观测数据中估计系统真实状态,有效平滑注水和产液曲线并去除测量误差;非线性扩散滤波器基于渗流力学的压降叠加原理,考虑注入水引发的压力波

传播至生产井所需时间。经信号处理后,通过注采响应分析反演得到井组分层连通系数。为直观展示井间连通系数在多层优化中的作用,设定一个简化的 2 注 1 采(3 层)井组作为仿真对象。该井组包含注水井 W-A、W-B 和生产井 O-1,生产井 O-1 分为 3 个生产层位。区块总注水量 I=500 m³/d,配注比约束设置为 0.1≤α_i≤0.6。

通过生产测井数据反演得到各层含水率,并以 1-含水率作为含油潜力表征值(表 2),用于近似反映不同层位剩余油动用潜力。该指标为分层配注优化的辅助评价参数,并不等同于严格意义上的剩余油饱和度。井间连通系数矩阵值(表 3)代表注入水从注水井 W-i 到生产井 O-1 的层位 k 的传输效率。

连通系数反演结果(图 11)与地质认识具有较好的一致性。层位 1 含水率高达 90%,连通系数显示 W-B 以 0.612 0 的强连通性构成主要水窜通道,这与该层处于高含水后期、注入水沿优势通道快速突进的地质特征相符。层位 2 和层位 3 的连通性分布则揭示了不同注水井对各层位的能量供给差异,为后续分层配注优化提供了定量依据。

3.3 注采优化效果分析

使用差分进化算法(DE)以最大化总产油量为目

表 2 井组分层含水率分析

Tab. 2 Layered water cut analysis of well groups

层位	O-1含水率	含油潜力表征值	连通性分析
层位 1	90%	10%	高含水后期,有效驱替潜力较低
层位 2	75%	25%	中含水阶段,具一定剩余油动用潜力
层位 3	65%	35%	中低含水阶段,剩余油动用潜力相对较高

表 3 井组分层连通系数矩阵

Tab. 3 Interlayer connectivity coefficient matrix of well groups

连通性系数	注水井 W-A	注水井 W-B	连通性评价
层位 1	0.388 0	0.612 0	W-B连通性极强,是当前主要的水窜通道
层位 2	0.621 8	0.378 2	W-A连通性极强,且该层潜力中等,是重点增注目标
层位 3	0.476 5	0.523 5	连通性相对均衡,但 W-B略占优势,该层潜力最大

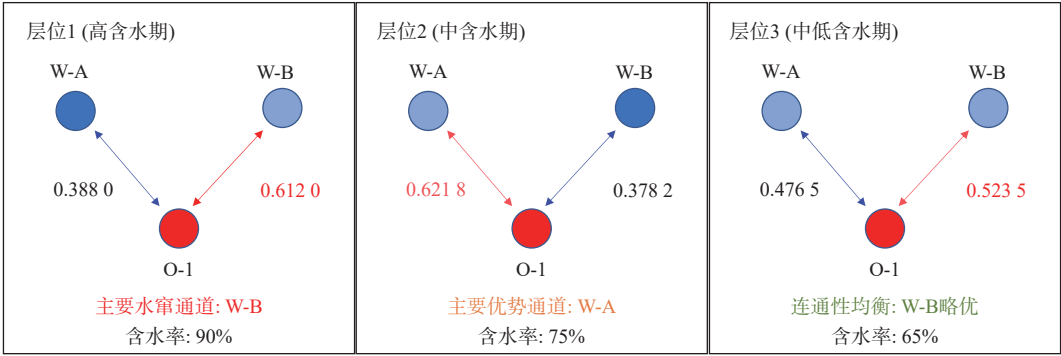


图11 井间连通性可视化对比图

Fig. 11 Well-to-Well Connectivity Visualization Comparison Diagram

标进行优化,得到的最优配注比(表 4)。优化后的总产油量从 175.00m³/d 提高到 176.10m³/d,日增油 1.10m³/d,相对提升 0.63%。优化后的总产油量从 175.00 m³/d 提高到 176.10 m³/d,日增油 1.10 m³/d,相对提升 0.63%。虽然绝对提升百分比较小,但考虑到该油井处于高含水后期,微小的增效也具有重要经济意义。目标函数成功指导算法执行了“抑强扶弱、集中驱替潜力层”

的配注策略:W-B 是高含水层 1 的主要连通源,其配注比被降低至 40%,减少了无效注水循环;W-A 是高潜力层 2 的主要连通源,其配注比被提高至 60%,实现了注水能量向最具产油效益层位的集中。该优化基于实时分层连通系数和含水率数据,展示了多源数据融合与 DE 优化在油田多层配注精细化调整中的高效性和实用性。

表 4 井组分层配注比及注水量分析

Tab. 4 Layered injection allocation ratio and water injection volume analysis of well groups

参数	基础方案	优化方案	策略分析
总注水量 (m ³ /d)	500	500	保持总注水规模不变,优化井间分配结构
W-A配注比	0.5	0.6	提高W-A配注比例,增强对含油潜力较高层位的有效驱替
W-B配注比	0.5	0.4	降低W-B配注比例,抑制高含水优势通道的低效循环
W-A单井注水量 (m ³ /d)	250	300	单井注水量增加 50 m ³ /d
W-B单井注水量 (m ³ /d)	250	200	单井注水量减少 50 m ³ /d
总产油量 (m ³ /d)	175.00	176.10	日增油 1.10 m ³ /d,相对提升 0.63%

4 结论

- (1)构建了融合物理约束的 LSTM 与 XGBoost 的多模型预测框架,通过物理约束嵌入损失函数提升了预测物理一致性,B 采油厂区块验证表明物理约束 LSTM 取得最优性能 R² 达 0.952 1,优于标准 LSTM 和 XGBoost 模型,验证了物理信息嵌入对提升模型外推能力和泛化性能的重要作用。
- (2)提出了卡尔曼滤波增强的自适应 ANN 连通性反演方法,通过梯度敏感性分析定量表征井间连通系数,无需示踪剂测试或历史拟合即可实现连通性的可解释量化。
- (3)建立了产油量最大化与含水率控制的双目标

协同优化模型,采用 DE 算法求解 Pareto 最优解集,仿真验证表明优化后区块产油量相对提升 0.63%。该方法计算效率高,可为油田日常配注调整提供近实时决策支持,对推动高含水后期油田智能化开发具有工程实践意义。

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):

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.
- 谷建伟, 任燕龙, 王依科, 等. 基于机器学习的平面剩余油分布预测方法[J]. 中国石油大学学报(自然科学版), 2020, 44(4): 39–46.
- GU Jianwei, REN Yanlong, WANG Yike, et al. Prediction methods of remaining oil plane distribution based on machine learning[J]. Journal of China University of Petroleum, 2020, 44(4): 39–46.
- 侯春华. 基于长短期记忆神经网络的油田新井产油量预测方法[J]. 油气地质与采收率, 2019, 26(3): 105–110.
- HOU Chunhua. New well oil production forecast method based on long short-term memory neural network[J]. Petroleum Geology and Recovery Efficiency, 2019, 26(3): 105–110.
- 匡立春, 刘合, 任义丽, 等. 人工智能在石油勘探开发领域的应用现状与发展趋势[J]. 石油勘探与开发, 2021, 48(1): 1–11.
- KUANG Lichun, LIU He, REN Yili, et al. Application status and development trend of artificial intelligence in petroleum exploration and development[J]. Petroleum Exploration and Development, 2021, 48(1): 1–11.
- 李培月, 梁豪, 杨俊岩, 等. 基于 GM(1, 1) 与 BP 神经网络模型的西安市地下水位动态特征及趋势预测研究[J]. 西北地质, 2025, 58(3): 236–245.
- LI Peiyue, LIANG Hao, YANG Junyan, et al. Dynamic Characteristics and Trend Prediction of Groundwater Level in Xi'an City, China Based on GM (1, 1) and BP Neural Network Models[J]. Northwestern Geology, 2025, 58(3): 236–245.
- 林琴, 郭永刚, 吴升杰, 等. 基于梯度提升的优化集成机器学习算法对滑坡易发性评价: 以雅鲁藏布江与尼洋河两岸为例[J]. 西北地质, 2024, 57(1): 12–22.
- LIN Qin, GUO Yonggang, WU Shengjie, et al. Evaluation of Landslide Susceptibility by Optimization Integrated Machine Learning Algorithm Based on Gradient Boosting: Take Both Banks of Yarlung Zangbo River and Niyang River as Examples[J]. Northwestern Geology, 2024, 57(1): 12–22.
- 刘合, 张凯. 人工智能与油藏工程交叉融合辅助注水开发方案调整[J]. 石油勘探与开发, 2023, 50(4): 784–796.
- LIU He, ZHANG Kai. Integration of artificial intelligence and reservoir engineering for assisted adjustment of waterflooding development schemes[J]. Petroleum Exploration and Development, 2023, 50(4): 784–796.
- 闵超, 文国权, 李小刚, 等. 可解释机器学习在油气领域人工智能中的研究进展与应用展望[J]. 天然气工业, 2024, 44(9): 114–126.
- MIN Chao, WEN Guoquan, LI Xiaogang, et al. Research progress and application prospects of explainable machine learning in artificial intelligence for oil and gas[J]. Natural Gas Industry, 2024, 44(9): 114–126.
- 任杉, 宋金民, 李柯然, 等. 基于机器学习的川中地区雷口坡组三段二亚段泥质灰岩储层分布预测[J]. 成都理工大学学报(自然科学版), 2025, 52(5): 914–930.
- REN Shan, SONG Jinmin, LI Keran, et al. Prediction of the distribution of marly limestone reservoirs in the second sub-member of the third member of the Leikoupo Formation, central Sichuan Basin, based on machine learning[J]. Journal of Chengdu University of Technology (Science & Technology Edition), 2025, 52(5): 914–930.
- 王华, 李明, 张强. 人工智能方法在测井数据分析中的适用性研究[J]. 石油勘探与开发, 2021, 48(3): 521–530.
- WANG Hua, LI Ming, ZHANG Qiang. Applicability of artificial intelligence methods in well logging data analysis[J]. Petroleum Exploration and Development, 2021, 48(3): 521–530.
- 薛永超, 赵辉. 基于深度森林模型的油井产量预测方法[J]. 油气地质与采收率, 2022, 29(4): 78–85.
- XUE Yongchao, ZHAO Hui. Oil well production prediction method based on deep forest model[J]. Petroleum Geology and Recovery Efficiency, 2022, 29(4): 78–85.
- 徐碧兰, 何建华, 吴炎峰, 等. 多源数据与地质逻辑融合的深层页岩储层构造裂缝测井智能识别[J]. 成都理工大学学报(自然科学版), 2025, 52(5): 931–950.
- XU Bilan, HE Jianhua, WU Yanfeng, et al. Intelligent logging identification of structural fractures in deep shale reservoirs based on the fusion of multisource data and geological logic[J]. Journal of Chengdu University of Technology (Science & Technology Edition), 2025, 52(5): 931–950.
- 张凯, 赵兴刚, 张黎明, 等. 智能油田开发中的大数据及智能化理论和方法研究现状及展望[J]. 中国石油大学学报(自然科学版), 2020, 44(4): 28–38.
- ZHANG Kai, ZHAO Xinggang, ZHANG Liming, et al. Research status and prospect of big data and intelligent optimization theory and methods in intelligent oilfield development[J]. Journal of China University of Petroleum (Edition of Natural Science), 2020, 44(4): 28–38.
- Perona P, Malik J. Scale-Space and Edge Detection Using Anisotropic Diffusion[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1990, 12(7): 629–639.
- Raissi M, Perdikaris P, Karniadaki G E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations[J]. Journal of Computational Physics, 2019, 378: 686–707.
- Storn R, Price K. Differential Evolution: A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces[J]. Journal of Global Optimization, 1997, 11(4): 341–359.