

基于 EWMA-LightGBM-Huber 的机械钻速 在线预测方法

黄杰^{1,2}, 孔令镒^{1,2*}, 王瑜^{1,2}, 刘宝林¹, 黄新武^{1,2}, 陈一诺¹

(1. 中国地质大学(北京)深部探测与成像全国重点实验室, 北京 100083; 2. 矿山钻探救援应急管理部重点实验室, 北京 100083)

摘要:机械钻速是表征钻井效率与井下钻进状态的关键参数,其准确预测对提升钻井智能化水平具有重要意义。受地层条件、钻井参数及工况变化等因素耦合影响,传统模型难以兼顾整体趋势与局部响应,在线预测易存在未来信息泄露风险。针对上述问题,本文提出一种基于趋势-残差分解的 EWMA-LightGBM-Huber 机械钻速在线预测模型。首先,采用严格单侧处理策略对钻井数据进行异常值处理与降噪,避免未来信息泄露,再采用指数加权移动平均法提取机械钻速局部变化趋势;随后融合钻井参数与历史机械钻速特征,利用 LightGBM 学习趋势残差,并引入 Huber 损失增强模型鲁棒性;最后结合滑动窗口更新策略实现模型的随钻在线更新与连续深度预测。结果表明,所提出模型的 R^2 达 0.9333,较最优对比模型 RF 的 RMSE 降低 8.72%、MAE 降低 11.92%,表现出更优的预测精度和稳定性,消融实验验证了趋势提取、残差学习和鲁棒优化模块的有效性,滑动窗口分析进一步确定了模型在线更新的合理参数。研究结果为复杂钻进条件下机械钻速实时预测提供了一种有效的数据驱动方法。

关键词:机械钻速;在线预测;指数加权移动平均;残差学习;LightGBM;Huber 损失;滑动窗口;智能钻井
中图分类号:P634.9;TE928 **文献标识码:**A **文章编号:**2096-9686(2026)S1-0092-10

An online prediction method for rate of penetration based on EWMA-LightGBM-Huber

HUANG Jie^{1,2}, KONG Lingrong^{1,2*}, WANG Yu^{1,2}, LIU Baolin¹, HUANG Xinwu^{1,2}, CHEN Yinuo¹

(1. State Key Laboratory of Deep Earth Exploration and Imaging, China University of Geosciences, Beijing 100083, China; 2. Key Laboratory of Mine Drilling and Rescue, Ministry of Emergency Management, Beijing 100083, China)

Abstract: Rate of penetration (ROP) is a critical indicator for evaluating drilling efficiency and downhole drilling conditions. Accurate ROP prediction is essential for improving intelligent drilling. However, the coupled effects of formation properties, drilling parameters, and operational variations make it challenging for conventional models to simultaneously capture long-term trends and short-term fluctuations, and online prediction is prone to future information leakage. To address this issue, an online EWMA-LightGBM-Huber model based on trend-residual decomposition is proposed for ROP prediction. First, a strict one-sided processing strategy is adopted to perform outlier handling and noise reduction on drilling data, thereby avoiding future information leakage. The exponentially weighted moving average (EWMA) method is then employed to extract the local trend component of ROP. Subsequently, drilling parameters and historical ROP features are integrated, and LightGBM is used to learn the residual component between the actual ROP and the extracted trend. The Huber loss function is further introduced to improve model robustness against abnormal fluctuations. Finally, a sliding-window updating strategy is incorporated to

收稿日期:2026-07-31; 修回日期:2026-09-20 DOI:10.12143/j.ztgc.2026.S1.010

基金项目:国家科技重大专项专题(编号:2024ZD1003105.4);矿山钻探救援应急管理部重点实验室开放课题(编号:ZTJY202501)

第一作者:黄杰,女,土家族,2002年生,博士研究生,地质工程专业,研究方向为钻探技术,北京市海淀区学院路29号,jh@email.cugb.edu.cn。

通信作者:孔令镒,男,汉族,1993年生,副教授、博士生导师,地质工程专业,博士,从事智能定向钻探技术方面研究工作,北京市海淀区学院路29号,lrkong@email.cugb.edu.cn。

引用格式:黄杰,孔令镒,王瑜,等.基于 EWMA-LightGBM-Huber 的机械钻速在线预测方法[J].钻探工程,2026,53(S1):92-101.

HUANG Jie, KONG Lingrong, WANG Yu, et al. An online prediction method for rate of penetration based on EWMA-LightGBM-Huber[J]. Drilling Engineering, 2026,53(S1):92-101.

achieve online model updating while drilling and continuous depth-wise prediction. The results demonstrate that the proposed model achieves an R^2 of 0.9333, with the RMSE reduced by 8.72% and the MAE reduced by 11.92% compared with the best-performing benchmark model RF, exhibiting superior prediction accuracy and stability. Ablation experiments verify the effectiveness of trend extraction, residual learning, and robust optimization modules. Furthermore, sliding-window analysis determines appropriate parameters for online updating. This study provides an effective data-driven framework for real-time ROP prediction under complex drilling conditions.

Key words: rate of penetration; online prediction; exponentially weighted moving average; residual learning; LightGBM; Huber loss; sliding window; intelligent drilling

0 引言

机械钻速是衡量钻头破岩效率和钻井作业效率的关键指标,其高低直接影响钻井周期、作业成本及钻井参数优化效果。随着资源和能源领域勘探向深部发展,对钻井过程的精细化控制和智能化决策提出了更高要求。准确预测机械钻速有助于识别低效钻进井段、优化钻井参数组合,并为钻井成本控制和钻进状态评价提供依据。

传统机械钻速预测方法主要基于 Bourgoyne-Young 等经验或半经验模型^[1-3],通过建立钻压、转速、水力参数及地层因素与机械钻速之间的数学关系实现预测。该类模型具有明确的工程意义,但其参数通常依赖特定区域和地层条件,跨工况应用能力有限。随着随钻测量技术和人工智能的快速发展,数据驱动模型逐渐成为机械钻速预测研究的重要方向。依托数据挖掘与非线性建模优势,随机森林^[4-6]、支持向量机^[7-9]、人工神经网络^[10-12]、梯度提升树^[13-16]及深度学习^[17-19]等模型被广泛应用于机械钻速预测,并在复杂钻井参数与机械钻速响应关系建模方面展现出较好的应用潜力。

围绕数据驱动模型的预测性能,已有研究主要从模型结构、参数寻优和动态更新等方面展开。Wang 等^[20]将遗传算法与 LightGBM 结合用于机械钻速预测,验证了梯度提升树模型在处理多参数耦合条件下钻井过程复杂非线性关系方面的有效性。牟凯文^[21]将极端梯度提升算法与粒子群优化结合,对钻压、转速和排量等钻井参数进行了优化。Zhang 等^[22]比较多种超参数优化策略对梯度提升模型的影响,验证了参数寻优方法在提升模型预测精度和泛化能力方面的有效性。马宝东等^[23]利用多基因遗传规划算法构建显式机械钻速表达式,并采用贝叶斯优化对模型参数进行动态更新,使模型能够更好地适应钻进工况的变化。这些研究推动了机械钻速预测由单一模型拟合向模型结构与参数

协同优化发展。

模型性能不仅取决于算法结构,还受到输入特征质量和训练样本规模的制约。周长春等^[24]开展了基于融合特征选择算法的钻速预测研究,通过综合不同特征筛选结果优选钻井参数,减少单一特征选择方法可能产生的偏差,为提高预测模型的稳定性和特征解释性提供了依据。胥知画等^[25]将粒子群优化引入融合特征选择过程,利用群智能搜索获得较优特征组合,将机械钻速预测中的特征筛选转化为优化问题,有助于提高复杂钻井参数空间中的搜索效率。沙林秀等^[26]开展了基于主成分分析的 NCPSO-BP 机械钻速预测研究,通过压缩高维钻井信息并优化 BP 神经网络,减弱特征共线性和输入维度过高对模型训练的影响,为小样本条件下建立较稳定的预测模型提供了参考。除特征质量外,样本数量是否能够支撑模型训练同样关系到预测结果的可信度。李谦等^[27]分析了人工智能钻速预测模型的数据有效性下限,揭示了数据规模对模型性能的影响,为判断现有样本能否满足机械钻速建模需求提供了依据。石祥超等^[28]结合区块多井数据讨论训练样本范围和特征选取对机械钻速预测结果的影响,将模型评价由单井拟合精度进一步延伸至多井数据条件下的适用性,为预测模型的跨井验证和现场推广提供了参考。

然而,即使输入特征和模型参数得到优化,固定训练样本所建立的静态模型仍可能因地层、钻头状态及钻井参数持续变化而逐渐失效。因此,机械钻速预测开始由静态建模向能够随新增数据更新的动态预测模式发展。Rezki 等^[29]采用霍夫丁树构建机械钻速自适应预测模型,并利用 Wasserstein 距离监测数据质量和概念漂移,使模型能够根据数据分布变化进行在线调整。Chen 等^[30]构建了由 DBSCAN 异常值识别、Hermite 插值和自适应小波降噪组成的多阶段预处理流程,并融合集成学习与

深度神经网络开展机械钻速预测,说明数据质量控制是提高模型稳定性和泛化能力的重要基础。Gan等^[31]通过软测量技术获取地层可钻性参数,结合数据预处理、混合建模与增量学习策略建立机械钻速预测模型,提高了模型对实时工况变化的适应能力。在线预测还涉及预测结果修正、不确定性表达及新井初期样本不足等问题。张琦等^[32]对点预测和区间预测模型进行滑动窗口在线更新,将预测结果由单一数值扩展为预测区间,为复杂钻探条件下评价预测结果的不确定性提供了依据。甘超等^[33]通过区域多井数据优选和模型预训练引入邻井信息,缓解了新井初始阶段本井样本不足的问题,为提高深部地质钻探钻速预测的适应性提供了新的数据利用方式。与此同时,钻井参数和机械钻速沿井深连续变化,其序列依赖关系也逐渐受到关注,长短期记忆网络等循环网络能够有效挖掘钻井参数随时间变化的长程依赖特征^[34]。Pan等^[35]采用门控循环单元网络与反向传播神经网络并联结构分别处理时序和非时序特征,并结合滑动窗口策略实现模型的动态更新。彭炽等^[36]依据特征类型构建子网络,利用井深序列和钻井参数的差异化表征提高页岩气水平井机械钻速预测精度。相关研究说明,数据质量控制、特征有效性、序列信息利用和模型在线更新共同影响机械钻速动态预测的稳定性与适应能力。

尽管已有研究从数据预处理、特征选择和在线更新等方面改进机械钻速预测,但对钻速趋势、局部波动与模型动态更新的协同考虑仍显不足。因此,如何在避免未来信息泄露的前提下,有效提取钻速变化趋势并实现动态预测,仍是提高机械钻速在线预测可靠性的关键问题。针对上述问题,本文提出一种基于EWMA-LightGBM-Huber的机械钻速在线预测方法。首先,采用严格单侧的数据处理策略,避免预测过程中引入未来信息,并利用指数加权移动平均(Exponentially Weighted Moving Average, EWMA)动态提取机械钻速随井深变化的局部趋势特征;随后,利用轻量级梯度提升机(Light Gradient Boosting Machine, LightGBM)模型建立钻井参数与机械钻速趋势之间的非线性映射关系,并引入基于Huber损失函数的残差学习模型,对趋势预测过程中未充分表征的局部变化进行鲁棒修正;进一步结合滑动窗口策略,仅利用当前预测点之前

已获取的历史数据进行模型更新,实现面向随钻过程的动态预测;最后,通过与多种机器学习模型及不同模块组合的消融实验进行对比分析,验证所提出方法在连续深度机械钻速预测中的有效性和稳定性,为智能钻井过程中的实时钻速预测提供一种新的技术方案。

1 数据预处理

1.1 特征筛选

本文采用Liu等^[37]公开报道的中国新疆某页岩气井实钻数据。该数据集包含2383条数据,15种参数(含目标参数机械钻速)。由于不同钻井参数对机械钻速的影响程度存在差异,且部分输入变量可能存在冗余信息,为构建有效输入特征集,本文采用皮尔逊相关系数法定量分析各钻井参数与机械钻速的相关性^[38],如图1所示。地层压力梯度和漏斗黏度与机械钻速的皮尔逊相关系数绝对值分别为0.0097和0.0117,表明二者与机械钻速之间的线性相关性非常弱。为减少冗余特征对模型训练的干扰并降低模型复杂度,本文剔除地层压力梯度和漏斗黏度,保留其余的12个钻井参数用于后续建模。

1.2 异常值识别与处理

考虑到钻井过程具有明显的时序性和多参数耦合特性,本文从统计分布、局部波动及多变量关联关系三个方面对数据异常点进行综合识别。首先,采用Z-score^[39]、四分位距(IQR)^[40]和中位数绝对偏差(MAD)方法对数据进行统计异常检测^[41],通过分析数据点偏离整体分布中心的程度识别极端异常值。其中,Z-score适用于近似正态分布条件下的全局异常识别,IQR和MAD方法由于对异常值具有更强的鲁棒性,可进一步降低极端样本对检测结果的影响。为进一步考虑多变量耦合作用下的异常模式,引入孤立森林(Isolation Forest)和局部离群因子(LOF)算法^[42-43],从高维特征空间角度检测与整体数据分布不一致的异常样本。通过多种异常检测方法联合判别,综合确定异常值,如图2所示,其中包含117个疑似异常点和36个严重异常点。在此基础上,结合异常点对应的井段钻井状态、数据变化趋势及工程经验进行进一步复核,剔除可能由正常地层变化或钻井参数调整引起的合理波动,最终确定105个异常样本。

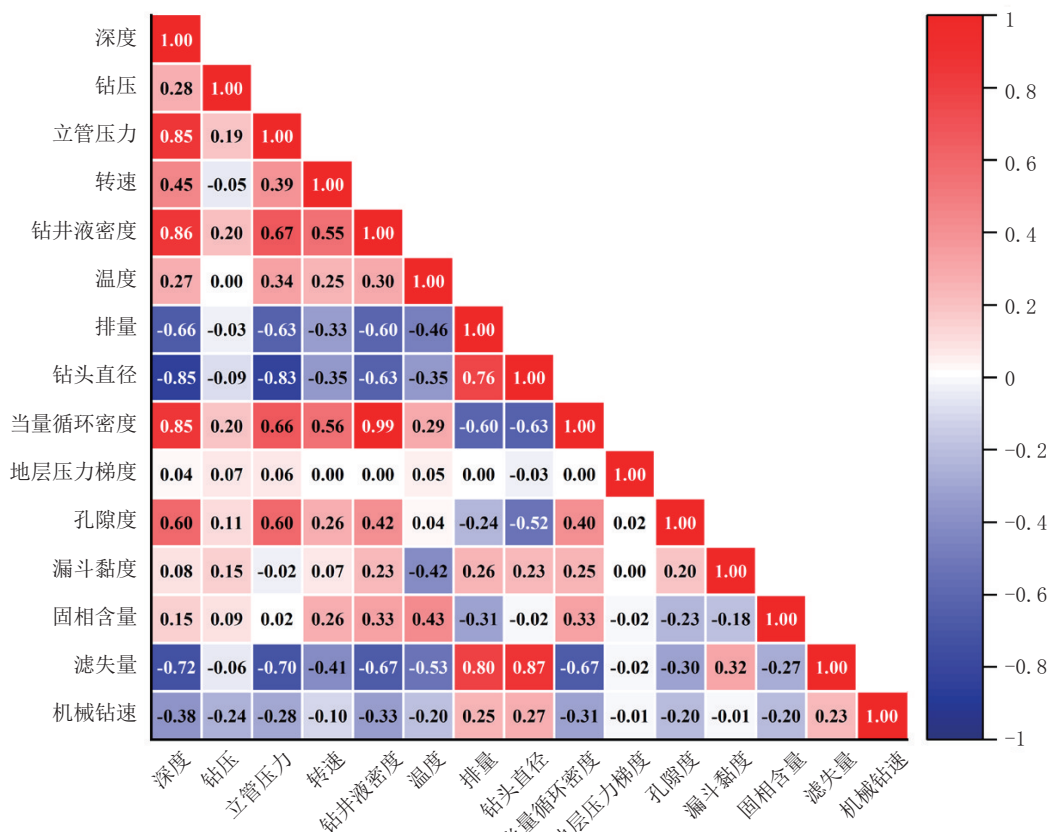


图1 钻井参数相关系数矩阵热图

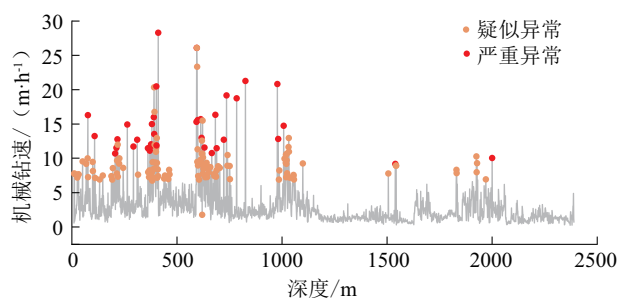


图2 机械钻速异常值识别结果

针对经人工复核确认的机械钻速异常样本,本文依据异常点在井深方向上的连续性特征,采用差异化修复策略。对于孤立分布的异常点,借鉴Hampel滤波中基于局部中位数的异常抑制思想^[44],利用5 m深度邻域范围内正常机械钻速观测值的中位数进行替换,以降低极端异常值对局部趋势估计的影响,提高异常修复的鲁棒性。对于连续出现的异常区间,将其视为局部数据失真区域,采用区间两侧最近有效观测点之间的线性插值方法进行修复,使修复后数据能够保持沿井深方向的连续变化

特征,避免连续异常点被统一替换导致局部趋势失真。对于未被确认的异常样本,则保留其原始观测值,以避免过度平滑或误修复由地层变化、钻井参数调整等真实钻进过程引起的机械钻速响应变化。处理后的数据如图3所示。

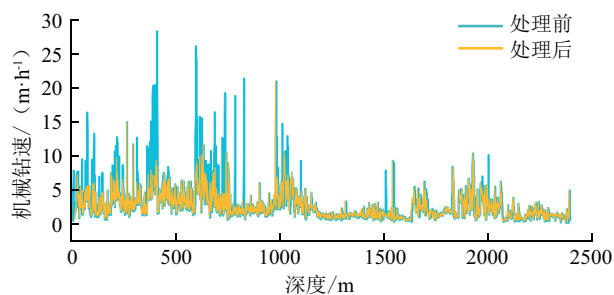


图3 机械钻速异常值处理前后对比

1.3 数据降噪

考虑到现场测量信号易受设备振动、采样误差和工况切换等因素影响,原始钻井数据中通常包含一定程度的噪声。采用SG滤波方法对机械钻速进行降噪处理^[45],以削弱瞬时尖峰和随机波动的

影响。SG滤波通过局部窗口内多项式最小二乘拟合实现信号平滑,在有效抑制高频噪声的同时,能够较好地保持信号的低阶变化趋势。传统SG滤波通常以待处理样本为中心构造对称滑动窗口,其滤波结果同时依赖当前样本之前和之后的观测值,会引入未来深度信息。为避免信息泄露,本文使用单侧SG滤波。对于第 t 个深度点,仅允许利用当前点及其历史窗口数据。

由于机械钻速数据具有明确的非负物理约束,直接在原始线性域内进行单侧SG拟合时,在边界区域或局部剧烈波动情况下可能产生非物理负值。若采用简单截断方式将负值修正为零,则容易引入大量人工近零值,进而改变原始机械钻速数据的分布特征,并削弱低速钻进区间的有效信息。为避免上述问题,本文首先对异常修复后的机械钻速序列进行自然对数变换,将其映射至对数空间后实施单侧SG滤波,并通过指数反变换恢复至原始尺度。该处理方式无需人为设置下限约束,即可保证滤波结果始终满足机械钻速非负特性,同时降低极端波动对局部平滑过程的影响,提高滤波结果的稳定

性。降噪后的数据如图4所示。

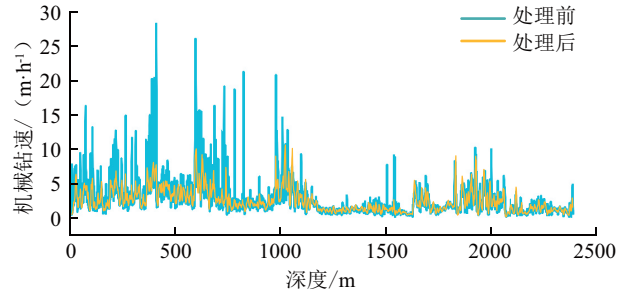


图4 机械钻速降噪前后对比

2 模型构建

本文结合时序趋势建模方法与数据驱动模型的优势,提出一种EWMA-LightGBM-Huber组合预测模型,如图5所示。该模型利用EWMA方法表征机械钻速随井深变化的连续趋势特征,并通过LightGBM-Huber残差学习模型挖掘钻井参数与局部趋势偏离之间的非线性关系,实现趋势预测与误差修正的协同融合。在此基础上,结合滑动窗口在线更新策略,建立适用于单井非平稳钻进过程的机械钻速动态预测模型。

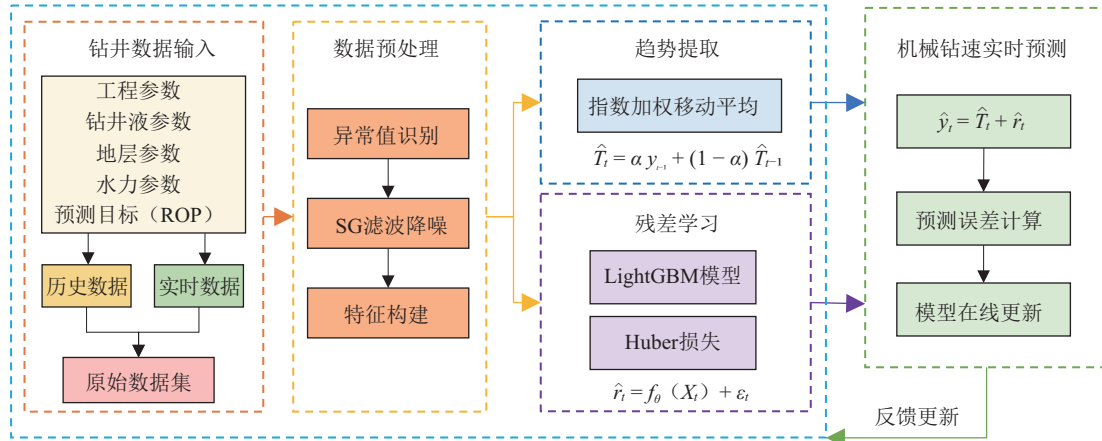


图5 机械钻速预测模型工作原理

2.1 指数加权趋势提取

指数平滑法被广泛用于一般时间序列分析和预测工作中。其中,EWMA是最常用的一种方法^[46],其通过赋予近期观测值较大的权重、赋予远期观测值逐渐衰减的权重,实现对时间序列局部变化趋势的动态跟踪。与简单移动平均相比,EWMA不需要保存全部历史数据,具有计算过程简洁、参数较少和在线更新效率高等优点,适合随钻预测场

景。对于第 t 个深度点,其趋势估计可表示为:

$$\hat{T}_t = \alpha y_{t-1} + (1 - \alpha) \hat{T}_{t-1} \quad (1)$$

式中: $\hat{T}_t$ ——第 t 个深度点的机械钻速趋势估计值; y_{t-1} ——第 $t-1$ 个深度点已经测到的机械钻速; α ——平滑系数。

该计算过程仅使用当前预测位置之前的历史信息,因而能够避免未来数据泄露。

2.2 LightGBM-Huber残差模型

EWMA本质上是一种单参数局部平滑方法,主要用于描述机械钻速随井深连续变化的总体趋势,难以充分表征钻压、转速、排量、立管压力等钻井参数变化及地层非均质性引起的复杂非线性波动。为此,本文将机械钻速 y_t 分解为EWMA趋势项 T_t 和残差项 r_t :

$$y_t = T_t + r_t \quad (2)$$

利用LightGBM模型学习钻井参数、历史机械钻速及其动态特征与残差之间的非线性关系,并引入Huber损失函数构建鲁棒残差学习模型,对预测误差进行进一步修正。最终预测结果:

$$\hat{y}_t = \hat{T}_t + \hat{r}_t \quad (3)$$

$$\hat{r}_t = f_\theta(X_t) + \epsilon_t \quad (4)$$

$$f_\theta(X_t) = \sum_{m=1}^M \eta h_m(X_t) \quad (5)$$

式中: $\hat{y}_t$ ——第 t 个深度点的机械钻速预测值; $\hat{r}_t$ ——第 t 个深度点LightGBM模型预测的机械钻速残差; θ ——模型超参数集合; M ——决策树的总数量; η ——学习率; $h_m(X_t)$ ——第 m 棵树的输出。

钻井过程中,由于地层非均质性、钻头-岩石作用突变及录井测量误差的影响,机械钻速残差序列中通常存在少量幅值较大的异常偏差。为降低此类异常残差对模型训练的干扰,本文采用Huber损失函数进行残差模型优化^[47]。当残差较小时,Huber损失采用平方形式,使模型能够充分学习趋势预测后的细微偏差特征;当残差超过设定阈值时,损失函数转变为线性形式,限制异常样本梯度贡献,避免少数极端误差点影响模型参数更新,从而提升残差修正模型对复杂钻进工况的适应能力。

$$\theta^* = \arg \min_{\theta} \left[\sum_{t \in D} L_\delta(r_t - f_\theta(X_t)) + \Omega(\theta) \right] \quad (6)$$

$$L_\delta(e_t) = \begin{cases} \frac{1}{2} e_t^2, & |e_t| \leq \delta \\ \delta \left(|e_t| - \frac{1}{2} \delta \right), & |e_t| > \delta \end{cases} \quad (7)$$

$$e_t = r_t - \hat{r}_t = r_t - f_\theta(X_t) \quad (8)$$

式中: θ^* ——经优化后得到的最优模型参数; $L_\delta(e_t)$ ——误差为 e_t 时的Huber损失值; δ ——Huber损失函数的阈值参数, $\delta > 0$; D ——训练数据集; $\Omega(\theta)$ ——模型正则化项; e_t ——第 t 个深度点机械钻速残差的预测误差,主要来源于未观测地层因素、测量

噪声及模型近似误差。

2.3 模型训练

为提高模型对钻进过程动态变化的适应能力,本文引入滑动窗口策略对训练样本进行动态更新,实现预测模型的在线迭代优化。滑动窗口方法已广泛应用于时间序列建模领域,其通过持续利用最新时间段内的有效样本更新模型参数,能够减少长期历史数据中与当前钻进状态相关性较弱的信息干扰,提高模型对时变系统的响应能力。

如图6所示,设定滑动窗口长度为 L ,随着新钻井数据的持续获取,窗口沿深度方向逐步向前移动。在每次模型更新过程中,仅选取当前窗口范围内已完成观测的历史样本作为训练数据,并基于该局部数据重新训练模型,以捕获当前钻进阶段的数据分布特征。模型训练完成后,对窗口之后的未知机械钻速值进行预测;当新的钻井数据获得后,将其纳入历史样本集,并按照上述流程更新滑动窗口及模型参数,从而实现机械钻速随钻动态预测。

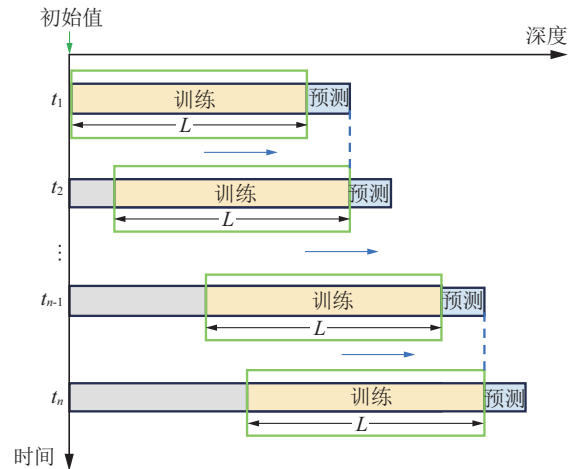


图6 滑动窗口的动态更新策略

3 结果与讨论

3.1 模型性能对比

为验证所提出EWMA-LightGBM-Huber模型的有效性,选取随机森林(RF)、极端梯度提升模型(XGBoost)、类别特征提升模型(CatBoost)、支持向量回归(SVR)和门控循环单元网络(GRU)作为对比模型。为保证不同模型比较的公平性,各模型均采用相同的滑动训练窗口策略,窗口长度设置为500 m,预测过程中按照1 m步长进行逐点输出,评价区间为510~2389 m。采用决定系数(R^2)、平均

绝对误差(MAE)、平均绝对百分比误差(MAPE)及均方根误差(RMSE)对模型性能进行评估,模型预测结果如图7所示。

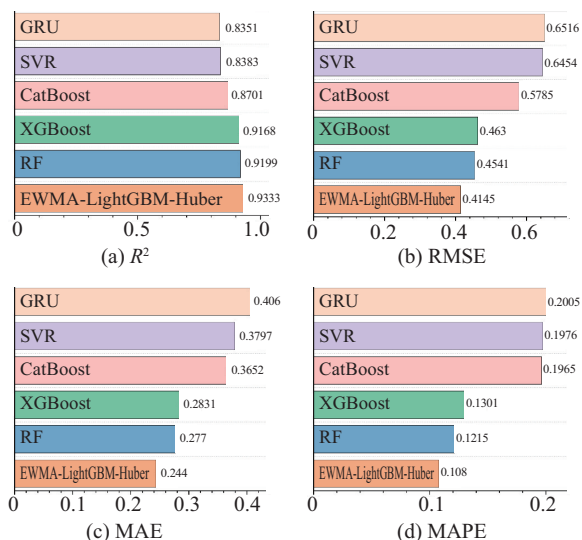


图7 不同模型评价指标对比

结果表明,EWMA-LightGBM-Huber模型在4项评价指标上均取得最优结果,表现出较好的预测精度和稳定性。与表现最好的机器学习模型RF相比,所提出模型的RMSE降低8.72%,MAE降低11.92%, R^2 提高0.0134。RF和XGBoost得到的 R^2 分别为0.9199和0.9168,明显优于CatBoost、SVR和GRU。该结果表明,在本文单井、小样本且输入变量以连续数值型钻井参数为主的数据条件下,基于树结构集成学习的方法能够有效捕获钻压、转速、排量、立管压力、当量循环密度等参数与机械钻速之间复杂的非线性映射关系。

SVR模型具有较强的小样本非线性拟合能力,但其预测性能对特征尺度、核函数类型及关键超参数较为敏感,固定核函数形式难以充分适应不同井段内钻井参数与机械钻速响应关系的动态变化。GRU模型虽然具备提取时间序列长期依赖特征的能力,但在本文研究条件下未表现出优于树模型的预测效果。这并不意味着循环神经网络不适用于机械钻速预测,而是由于单井数据规模有限、滑动窗口样本数量受限,GRU对复杂时序特征的学习优势未能充分发挥。

总体而言,EWMA-LightGBM-Huber模型性能提升主要来源于趋势提取、非线性建模和残差修

正三个环节的协同作用。因此,所提出模型能够同时利用机械钻速序列的深度连续性和钻井参数的多变量非线性特征,在连续深度在线预测场景下表现出更好的适应性和鲁棒性。

3.2 消融实验

为进一步分析EWMA趋势提取、LightGBM残差学习和Huber损失函数对模型性能的贡献,本文设计5组消融实验。所有模型采用相同的数据、500 m滑动训练窗口、1 m更新间隔和510~2389 m评价区间,结果如图8所示。

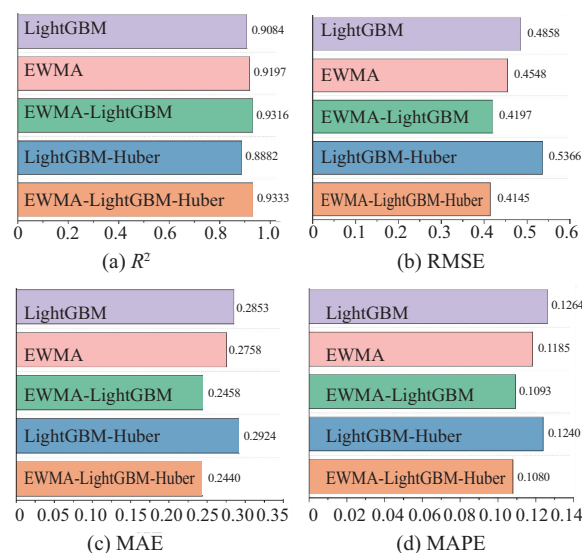


图8 模型消融实验结果

仅使用EWMA时,模型 R^2 已达到0.9197, RMSE为0.4548 m/h,说明近期历史数据对下一深度机械钻速具有较高的预测价值。然而,EWMA只利用历史数据,不能描述钻压、转速、排量、立管压力及地层参数变化对机械钻速产生的额外影响,因此在工况调整或地层突变位置仍会产生较大误差。

LightGBM模型的 R^2 为0.9084,低于EWMA模型,LightGBM模型需要同时拟合由深度变化引起的整体趋势和局部非线性扰动,在有限滑动窗口样本条件下增加了模型学习难度。在EWMA趋势基础上加入LightGBM模型后,模型 R^2 由0.9197提高至0.9316,表明LightGBM能够结合钻井参数和历史机械钻速特征,有效学习EWMA趋势之外的局部非线性变化,实现对趋势预测误差的补偿。

EWMA-LightGBM-Huber模型相较于

EWMA-LightGBM模型, R^2 由 0.9316 提高至 0.9333, 虽然提升幅度相对有限, 但 R^2 、RMSE、MAE 和 MAPE 均得到改善, 说明 Huber 损失函数能够降低大幅残差样本对模型优化过程的过度影响, 提高残差预测的稳定性。

值得注意的是, LightGBM-Huber 模型的性能低于 LightGBM 模型, 这可能是由于机械钻速中仍包含部分真实地层变化和工况响应特征, 直接采用 Huber 损失会降低大幅变化样本在模型训练中的影响, 使模型对局部突变响应能力下降。

综上, 消融实验验证了各模块的有效性: EWMA 通过提取机械钻速局部趋势降低了预测复杂度, 是模型性能提升的基础; LightGBM 进一步融合钻井参数信息, 提高了对非线性变化的刻画能力; Huber 损失则增强了残差模型对异常样本的鲁棒性。

3.3 模型优化

3.3.1 滑动训练窗口优化

滑动窗口长度决定模型每次更新时参与训练的历史样本范围, 是影响在线机械钻速预测稳定性和地层适应能力的重要因素。窗口长度过小时, 训练样本数量有限, 模型容易受到局部异常波动和短期工况变化影响; 窗口长度过大时, 虽然能够提供更多历史信息, 但可能引入与当前钻进状态相关性较低的早期地层响应特征, 降低模型对局部变化的敏感性。

由于不同窗口长度对应的首个有效预测点不同, 为避免评价区间差异对模型性能比较造成影响, 本文统一选取 710~2389 m 井段作为不同滑动窗口长度下的评价区间, 不同大小滑动窗口模型的预测性能结果如表 1 所示。

表 1 不同大小滑动窗口模型的预测性能

滑动窗口长度/m	R^2	RMSE/($\text{m}\cdot\text{h}^{-1}$)	MAE/($\text{m}\cdot\text{h}^{-1}$)	MAPE/%	模型更新次数	累计样本调用量
200	0.9254	0.3883	0.2222	11.0459	2180	436000
300	0.9236	0.3930	0.2246	11.0213	2080	624000
400	0.9244	0.3910	0.2227	10.9048	1980	792000
500	0.9271	0.3837	0.2193	10.7806	1880	940000
600	0.9274	0.3831	0.2193	10.8176	1780	1068000
700	0.9264	0.3856	0.2203	10.8685	1680	1176000

结果表明, 随着训练窗口由 200 m 增加至 600 m, 模型整体预测性能呈提升趋势, 说明适当增加历史样本范围有助于模型学习更加充分的钻井参数变化规律。其中, 600 m 窗口获得最低 RMSE 和最高 R^2 , 但相较于 500 m 窗口, 其 RMSE 仅降低 0.0006 m/h, 降幅约 0.17%, 性能提升较为有限; 相比之下, 500 m 窗口取得最低 MAPE, 同时具有更少的训练样本需求和更高的在线更新效率。

综合考虑预测精度、计算成本及模型对局部地层变化的响应能力, 本文在现有候选窗口和数据条件下, 选取 500 m 作为后续建模的滑动训练窗口长度。该取值是预测性能与实时适应性之间的折中, 并不意味着 500 m 是唯一适用或全局最优的窗口长度; 其邻近范围内的窗口也可能取得相近的预测效果, 仍需通过更小的窗口步长和多井数据进一步验证。为避免窗口长度变化干扰模型超参数优化结果, 并保证不同参数组合之间的可比性, 超参数优

化阶段将滑动训练窗口固定为 500 m。

3.3.2 LightGBM 超参数优化

LightGBM 模型的预测性能受到叶节点数量、学习率、样本采样比例、特征采样比例及正则化强度等超参数的影响。为进一步提升 EWMA-LightGBM-Huber 模型的预测精度, 本文在保持数据集、EWMA 趋势提取策略、500 m 滑动训练窗口及 1 m 更新间隔不变的条件下, 对残差预测模块中的 LightGBM 关键超参数进行优化, 优化前后的超参数如表 2 所示。采用优化后的超参数重新构建 EWMA-LightGBM-Huber 模型, 优化前后模型性能对比如表 3 所示。

结果表明, 超参数优化后模型的预测性能进一步提升, R^2 由 0.9333 提高至 0.9338, RMSE、MAE 和 MAPE 分别降低 0.44%、0.75% 和 1.45%, 最大绝对误差降低约 5.97%。尽管超参数优化带来的整体性能提升幅度相对有限, 但各项评价指标均呈改善趋

表2 LightGBM模型超参数优化结果

参 数	搜索范围	优化前	优化后
叶子节点数	15、23、31、47、63	31	63
学习率	0.02、0.03、0.05、0.08、0.10	0.05	0.02
弱学习器数量	80、120、160、200、260	120	120
叶子节点最	10、15、20、30、40、60	20	20
小样本数			
样本采样比例	0.75、0.85、0.90、1.00	0.90	0.75
特征采样比例	0.75、0.85、0.90、1.00	0.90	1.00
L1正则化系数	0、0.05、0.10、0.30、0.70	0.10	0.10
L2正则化系数	1、2、3、5、8、12	3	2
Huber阈值参数	0.80、0.85、0.90、0.95	0.90	0.95
最大深度	-1、5、7、9	-1	-1
最小分裂增益	0、0.01、0.05	0	0

表3 优化前后模型性能对比

模型	R^2	RMSE/ ($\text{m}\cdot\text{h}^{-1}$)	MAE/ ($\text{m}\cdot\text{h}^{-1}$)	MAPE/ %	最大误差/ ($\text{m}\cdot\text{h}^{-1}$)
优化前	0.9333	0.4145	0.2440	10.8031	4.1771
优化后	0.9338	0.4127	0.2421	10.6460	3.8713

势,说明优化后的参数组合能够进一步降低模型泛化误差,提高局部复杂井段的预测稳定性。

4 结论

(1)提出了EWMA-LightGBM-Huber机械钻速预测模型,该模型在连续深度实时预测任务中取得了最优预测性能,相较于传统机器学习和深度学习模型具有更高的预测精度和稳定性。

(2)通过消融实验验证了所提出模型中趋势提取、非线性学习和鲁棒优化模块的协同作用,结果表明,EWMA-LightGBM-Huber组合框架能够有效融合机械钻速序列连续性与钻井参数多变量信息,提高模型对复杂钻进条件下机械钻速动态变化的适应能力。

(3)滑动窗口长度对模型的在线适应能力与稳定性具有明显影响,参数优化能够进一步改善预测性能,但提升幅度有限,表明合理的数据处理、特征构造和模型结构比细微的超参数调整更为关键。

由于本文结果基于单井数据,后续仍需通过多井数据、岩性和测井特征及跨区块验证,进一步评价模型的泛化能力和现场适用性。

参考文献:

- [1] Maurer W C. The "perfect-cleaning" theory of rotary drilling[J]. Journal of Petroleum Technology, 1962, 14(11): 1270-1274.
- [2] Bourgoyne A T Jr, Young F S Jr. A multiple regression approach to optimal drilling and abnormal pressure detection[J]. Society of Petroleum Engineers Journal, 1974, 14 (04) : 371-384.
- [3] Warren T M. Penetration-rate performance of roller-cone bits [J]. SPE Drilling Engineering, 1987, 2(01): 9-18.
- [4] Hegde C, Wallace S, Gray K. Using trees, bagging, and random forests to predict rate of penetration during drilling[C]//SPE Middle East Intelligent Oil and Gas Symposium. SPE, 2015: D011S001R003.
- [5] Li S, Zhang J, Wu S, et al. Prediction of rate of penetration based on random forest in deep well[C]//International Conference on Information technology in Geo-Engineering. Cham: Springer International Publishing, 2019: 517-526.
- [6] Alsaihati A, Elkatatny S, Gamal H. Rate of penetration prediction while drilling vertical complex lithology using an ensemble learning model[J]. Journal of Petroleum Science and Engineering, 2022, 208: 109335.
- [7] Gan C, Cao W H, Wu M, et al. Prediction of drilling rate of penetration (ROP) using hybrid support vector regression: A case study on the Shennongjia area, Central China[J]. Journal of Petroleum Science and Engineering, 2019, 181: 106200.
- [8] Ahmed O S, Adeniran A A, Samsuri A. Computational intelligence based prediction of drilling rate of penetration: A comparative study [J]. Journal of Petroleum Science and Engineering, 2019, 172: 1-12.
- [9] Kor K, Altun G. Is Support Vector Regression method suitable for predicting rate of penetration? [J]. Journal of Petroleum Science and Engineering, 2020, 194: 107542.
- [10] Abbas A K, Rushdi S, Alsaba M, et al. Drilling rate of penetration prediction of high-angled wells using artificial neural networks [J]. Journal of Energy Resources Technology, 2019, 141(11): 112904.
- [11] Brenjkar E, Delijani E B, Karroubi K. Prediction of penetration rate in drilling operations: a comparative study of three neural network forecast methods[J]. Journal of Petroleum Exploration and Production Technology, 2021, 11(2): 805-819.
- [12] Anemangely M, Ramezanzadeh A, Tokhmechi B, et al. Drilling rate prediction from petrophysical logs and mud logging data using an optimized multilayer perceptron neural network[J]. Journal of Geophysics and Engineering, 2018, 15(4): 1146-1159.
- [13] Allawi R H, Al-Mudhafar W J, Abbas M A, et al. Leveraging boosting machine learning for drilling rate of penetration (ROP) prediction based on drilling and petrophysical parameters [J]. Artificial Intelligence in Geosciences, 2025, 6 (1) : 100121.
- [14] Liu W, Fu J, Tang C, et al. Real-time prediction of multivariate ROP (rate of penetration) based on machine learning regression algorithms: Algorithm comparison, model evaluation and parameter analysis[J]. Energy Exploration & Exploitation,

- 2023, 41(5): 1779-1801.
- [15] Han X, Zhang Z, Cheng P, et al. Application and analysis of the shapley additive explanations in predicting the rate of penetration [J]. *Petroleum Science and Technology*, 2026, 44(20): 3309-3328.
- [16] Ahmednour O, Chen D, Liu J, et al. Comparative analysis of rate of penetration prediction and optimization in deep wells using real-time continuous stacked generalization ensemble learning: A case study in Shunbei[J]. *Geoenery Science and Engineering*, 2025, 247: 213674.
- [17] Bizhani M, Kuru E. Towards drilling rate of penetration prediction: Bayesian neural networks for uncertainty quantification [J]. *Journal of Petroleum Science and Engineering*, 2022, 219: 111068.
- [18] Pacis F J, Ambrus A, Alyaev S, et al. Improving predictive models for rate of penetration in real drilling operations through transfer learning[J]. *Journal of Computational Science*, 2023, 72: 102100.
- [19] Ji H, Lou Y, Cheng S, et al. An advanced long short-term memory (LSTM) neural network method for predicting rate of penetration (ROP)[J]. *ACS Omega*, 2022, 8(1): 934-945.
- [20] Wang S S, Yan J, Geng H. Prediction of the ROP based on GA-LightGBM and drilling data[J]. *Geosystem Engineering*, 2025, 28(1): 12-30.
- [21] 牟凯文. 一种基于XGBoost回归与粒子群优化算法的大位移井钻速提升方法研究[J]. *石油地质与工程*, 2026, 40(3): 115-121.
- [22] Zhang J N, Altalbawy F M A, Vora T, et al. Application of hybrid gradient boosting machine to reliably predict rate of penetration[J]. *Physics and Chemistry of the Earth, Parts A/B/C*, 2025: 104117.
- [23] 马宝东, 祝兆鹏, 宋先知, 等. 基于BO-MGPP的机械钻速实时动态预测[J]. *石油机械*, 2025, 53(11): 18-27.
- [24] 周长春, 姜杰, 李谦, 等. 基于融合特征选择算法的钻速预测模型研究[J]. *钻探工程*, 2022, 49(4): 31-40.
- [25] 胥知画, 姜杰, 周长春, 等. 基于粒子群优化的融合特征选择钻速预测模型研究[J]. *钻探工程*, 2025, 52(2): 134-143.
- [26] 沙林秀, 胥陈卓. 基于主成分分析的NCP SO-BP机械钻速预测[J]. *石油钻采工艺*, 2022, 44(4): 515-521.
- [27] 李谦, 曹彦伟, 朱海燕. 基于人工智能的钻速预测模型数据有效性下限分析[J]. *钻探工程*, 2021, 48(3): 21-30.
- [28] 石祥超, 王宇鸣, 刘越豪, 等. 关于人工智能方法用于钻井机械钻速预测的探讨[J]. *石油钻采工艺*, 2022, 44(1): 105-111.
- [29] Rezki D, Mouss L H, Baaziz A, et al. Adaptive prediction of Rate of Penetration while oil-well drilling: A Hoeffding tree based approach[J]. *Engineering Applications of Artificial Intelligence*, 2025, 159: 111465.
- [30] Chen Y, Chen P, Dai X, et al. Prediction of rate of penetration in oil and gas well drilling: A hybrid data-driven approach with multi-stage data preprocessing[J]. *Engineering Applications of Artificial Intelligence*, 2026, 163: 113085.
- [31] Gan C, Wang Y, Cao W H, et al. Real-time formation drillability sensing-based hybrid online prediction method for the rate of penetration (ROP) and its industrial application for drilling processes [J]. *Control Engineering Practice*, 2025, 164: 106487.
- [32] 张琦, 甘超, 曹卫华. 大洋钻探过程钻速在线区间预测方法——以微型钻探船室内模拟实验为例[J]. *钻探工程*, 2024, 51(5): 45-52.
- [33] 甘超, 汪祥, 王鲁朝, 等. 基于区域多井数据优选与模型预训练的深部地质钻探过程钻速动态预测方法[J]. *钻探工程*, 2023, 50(4): 1-8.
- [34] Naderi M, Bagheri A, Khujin M N. Hybrid LSTM-Desirability concept for drilling rate of penetration (ROP) prediction and optimization[J]. *Array*, 2026: 100900.
- [35] Pan T, Song X, Ma B, et al. Predicting rate of penetration of horizontal wells based on the Di-GRU model[J]. *Rock Mechanics and Rock Engineering*, 2024: 1-16.
- [36] 彭炽, 任书江, 杨赞. 基于子网络架构的页岩气水平井机械钻速预测[J]. *钻采工艺*, 2025, 48(1): 113-120.
- [37] Liu N, Gao H, Zhao Z, et al. A stacked generalization ensemble model for optimization and prediction of the gas well rate of penetration: a case study in Xinjiang[J]. *Journal of Petroleum Exploration and Production Technology*, 2022, 12(6): 1595-1608.
- [38] Pearson K. VII. Mathematical contributions to the theory of evolution, III: Regression, heredity, and panmixia[J]. *Philosophical Transactions of the Royal Society of London. Series A, containing papers of a mathematical or physical character*, 1896 (187): 253-318.
- [39] Grubbs F E. Procedures for detecting outlying observations in samples[J]. *Technometrics*, 1969, 11(1): 1-21.
- [40] Hoaglin D C, Iglewicz B, Tukey J W. Performance of some resistant rules for outlier labeling[J]. *Journal of the American Statistical Association*, 1986, 81(396): 991-999.
- [41] Rousseeuw P J, Croux C. Alternatives to the median absolute deviation[J]. *Journal of the American Statistical Association*, 1993, 88(424): 1273-1283.
- [42] Altindal M C, Nivlet P, Tabib M, et al. Anomaly detection in multivariate time series of drilling data[J]. *Geoenery Science and Engineering*, 2024, 237: 212778.
- [43] Campos G O, Zimek A, Sander J, et al. On the evaluation of unsupervised outlier detection: measures, datasets, and an empirical study [J]. *Data Mining and Knowledge Discovery*, 2016, 30(4): 891-927.
- [44] Zhang B, Xiong J, Zhang N, et al. Improved method of processing downhole pressure data on smart wells[J]. *Journal of Natural Gas Science and Engineering*, 2016, 34: 1115-1126.
- [45] Savitzky A, Golay M J E. Smoothing and differentiation of data by simplified least squares procedures[J]. *Analytical Chemistry*, 1964, 36(8): 1627-1639.
- [46] Hunter J S. The exponentially weighted moving average [J]. *Journal of Quality Technology*, 1986, 18(4): 203-210.
- [47] Friedman J H. Greedy function approximation: a gradient boosting machine[J]. *Annals of Statistics*, 2001: 1189-1232.

(编辑 王跃伟)