

基于实时数据流对比的钻速建模算法

李同意¹, 李 谦^{1*}, 魏思维¹, 何俊杰¹, 姜 杰²

(1. 成都理工大学环境与土木工程学院, 四川 成都 610059; 2. 成都理工大学机电工程学院, 四川 成都 610059)

摘要: 为提升钻进效率与预测精度, 本文设计了一种基于实时数据流对比的钻速建模算法。通过主井与9口历史井的数据流比对, 构建了一个面向实时更新与模型复用的预测框架。系统首先利用滑动窗口机制对主井数据动态切片, 并以 K 近邻方式从9口井中提取同深度段数据, 构建参考数据集。再结合快速傅里叶变换与频谱相似性指标, 实现主井窗口与历史数据的频域对比。当相似度高于设定阈值时, 系统复用历史模型, 否则即时触发重新建模。建模过程采用随机森林算法, 融合主井累计窗口数据与历史近邻数据, 并按“80% 训练+20% 测试”的方式进行训练验证。在最终建模效果中, 模型表现出高度的稳定性与良好的泛化能力, 平均 R^2 为0.99, 残差围绕零值分布。该系统为钻速预测提供了具备实时性、自适应性与可拓展性的建模策略, 为智能钻井决策提供了重要支撑。

关键词: 钻速预测; 实时数据流; 动态建模; 频谱特征比对

中图分类号: P634; TE242 **文献标识码:** A **文章编号:** 2096-9686(2025)S1-0097-08

The rate of penetration modeling algorithm based on real-time data stream comparison

LI Tongyi¹, LI Qian^{1*}, WEI Siwei¹, HE Junjie¹, JIANG Jie²

(1. College of Environment and Civil Engineering, Chengdu University of Technology, Chengdu Sichuan 610059, China; 2. School of Mechanical and Electrical Engineering, Chengdu University of Technology, Chengdu Sichuan 610059, China)

Abstract: In order to improve the drilling efficiency and prediction accuracy, this paper designs an ROP modeling algorithm based on real-time data stream comparison. A prediction framework oriented to real-time updating and model reuse is constructed by comparing the data streams of the main well with those of nine historical wells. The system first utilizes the sliding window mechanism to dynamically slice the data of the main well, and extracts the data of the same depth section from the nine wells in a K -Nearest Neighbor manner to construct the reference dataset. Then, it combines the Fast Fourier Transform with the spectral similarity index to realize the frequency domain comparison between the main well window and the historical data. When the similarity is higher than a set threshold, the system reuses the historical model, otherwise it instantly triggers retraining. The modeling process adopts the Random Forest algorithm, fuses the cumulative window data of the main well with the historical near-neighbor data, and carries out training and validation according to the method of “80% training+20% testing”. In the final modeling results, the model shows high stability and good generalization ability, with an average R^2 of 0.99 and its residuals distributed around zero. The system provides a real-time, adaptive, and scalable modeling strategy for ROP prediction, which provides important support for intelligent drilling decision-making.

Key words: ROP prediction; real-time data stream; dynamic modeling; spectral feature comparison

收稿日期: 2025-06-15; 修回日期: 2025-07-24 DOI: 10.12143/j.ztgc.2025.S1.015

基金项目: 四川省自然科学基金青年科学基金项目“基于数字孪生的动态时变钻进工况自适应迁移模型研究”(编号: 2024NSFSC0817)

第一作者: 李同意, 男, 汉族, 2003年生, 硕士研究生, 地质资源与地质工程专业, 研究方向为数据分析与人工智能在钻探中的应用, 四川省成都市成华区二仙桥东三路1号, 2117889532@qq.com。

通信作者: 李谦, 男, 汉族, 1987年生, 副教授, 硕士生导师, 地质资源与地质工程专业, 博士, 从事钻井智能化分析与应用研究工作, 四川省成都市成华区二仙桥东三路1号, liqian2014@cdut.edu.cn。

引用格式: 李同意, 李谦, 魏思维, 等. 基于实时数据流对比的钻速建模算法[J]. 钻探工程, 2025, 52(S1): 97-104.

LI Tongyi, LI Qian, WEI Siwei, et al. The rate of penetration modeling algorithm based on real-time data stream comparison[J]. Drilling Engineering, 2025, 52(S1): 97-104.

0 引言

全球能源需求持续增长,推动石油与天然气资源向深层、超深层等复杂地质区域拓展,钻井作业面临高温高压、岩性突变等挑战。钻速作为衡量钻进效率的重要指标,受多种因素影响,具有较强的非线性与动态性,传统基于经验或物理公式的预测方法已难以满足现代钻井对智能化与实时性的要求。随着传感器技术的发展,钻井现场可获取海量高频数据流,为实时建模提供了数据基础。基于滑动窗口与频谱分析的实时建模方法逐渐受到关注,可辅助识别钻速变化规律与岩性突变信号,实现对历史井与主井数据的动态对比。引入频域特征不仅增强模型对钻井工况变化的敏感性,也为实现高精度的钻速预测提供了新思路。

近年来,钻速预测建模研究正逐步从基于物理公式和工程经验的传统范式,迈向以数据驱动为核心的智能化建模体系。在传统机器学习方向,Hegde等^[1]强调了数据驱动方法在拟合精度上的优势,Brenjkar等^[2]则对多种主流算法(如MLPNN、RBFNN、ANFIS、SVR)在钻速预测中的应用进行了系统比较,验证了其优越性。随后的研究进一步引入集成学习与优化算法以提升模型性能,高云伟等^[3]提出结合KNN、SVM、RF的Stacking集成框架,并通过遗传算法进行参数调优;周长春等^[4]通过特征选择策略联合GBDT模型,提升了特征使用效率;张海军等^[5]基于遗传算法对随机森林结构进行优化,增强了分类能力;曾小龙等^[6]则结合复杂地层特征构建了多因子预测模型,在特殊地质条件下展现出较强的适应性。

随着深度学习技术的发展,其在非线性拟合和特征交互建模方面的优势也逐渐显现。焦通等^[7]指出深度神经网络相比传统方法在自动特征提取和参数依赖建模方面具备明显优势。Tu等^[8]设计了结合GRU与Informer结构的模型,以同时捕捉短期与长期时间依赖特征,提升了实时钻速预测效果;黄小龙等^[9]则提出将神经网络与优化算法融合应用于实时预测系统中,并结合可视化模块实现钻井现场辅助决策;Haodong等^[10]利用深度神经网络构建高精度非线性模型,任辉^[11]尝试通过RNN实现多步预测优化策略,而李博志等^[12]引入注意力机制以增强模型对关键特征的关注能力,进一步提升预测效果与泛化能力。张琦等^[13]采用粒子群优化的极

限学习机(PSO-ELM)结合非参数估计建立了钻速在线区间预测模型,并通过室内模拟实验验证了该方法能有效量化不确定性。胥知画等^[14]提出一种基于粒子群优化的融合特征选择算法,通过优化关键参数阈值筛选出最佳特征子集,实验证明该方法可有效提升多种机器学习钻速预测模型的精度。李谦等^[15]利用BP神经网络分析了钻速预测模型的数据有效性,研究确定了建立高精度模型所需的最少输入参数数量和最低取样精度

在钻井场景所涉及的实时数据流处理研究中,Mehmood等^[16]总结了该领域的主要挑战与应对路径,指出现代企业日益依赖实时流处理与分析系统获取数据优势,并提出了适用于工业数据场景的系统性框架建议。具体应用中,刘禹等^[17]开发了面向测震场景的可视化系统,验证了实时流数据在监测系统中的高效处理能力;王建胜等^[18]针对钻井现场构建了边缘计算架构,通过自适应负载均衡与分布式协同计算,有效提升了钻井数据流的处理效率与系统稳定性。此外,孙玉芬等^[19]、李岩等^[20]分别从算法模型与系统实现角度,综述了数据流挖掘技术的演化趋势,指出传统静态分析方法难以适应实时流数据的高速性、连续性与动态性,强调需发展适配性强、响应快速的实时建模与分析方法。

本文提出了一种融合实时数据流处理、频域分析与随机森林建模的钻速预测方法,旨在构建具备实时性、准确性与适应性的智能建模方案。通过引入滑动窗口机制对钻井实时数据进行动态截取,结合频谱相似性判断策略,对主井当前数据与9口历史数据进行动态对比,智能选择“主井自建”或“9口继承”建模路径,实现模型的持续更新与优化,为复杂地层条件下的钻速预测与智能化钻井作业提供了技术支撑。

1 数据采集与预处理

1.1 数据来源与采集

本文所用数据采集自10口油气井,钻井数据集包含1~10号井的46种参数数据,包括钻速、钻压、大钩载荷、3转读数等涉及多个方面的工程及地质参数。其中有15个操作参数,18个泥浆参数,6个地质参数,4个钻头参数,3个提速钻具参数。

钻井过程中每隔1 m进行一次测井,获得一组数据,10口油气井共计21912组数据。各井深度如

图1所示。10口油气井数据的采集的井深大致范围在1000~4000 m,深度相同的数据较多,具有相同深度的比较条件基础。

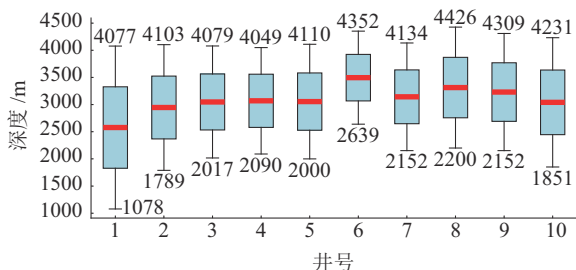
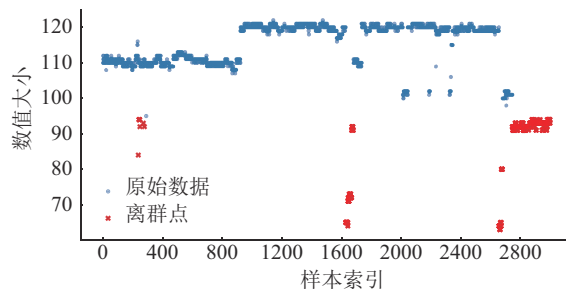


图1 各井深度箱型

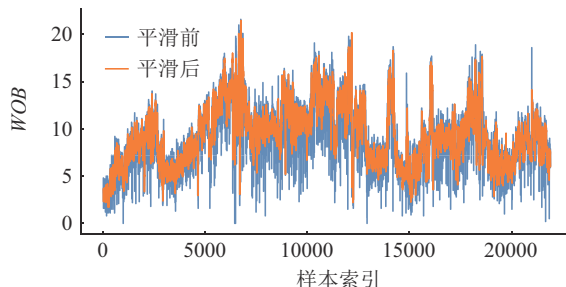
1.2 数据预处理方法及结果

针对所采集原始钻井数据,通过一系列数据预处理手段进行数据清洗,以获得能够支撑后续分析处理的数据。对原始钻井数据进行全面预处理:数值型缺失值采用线性插值,非数值型缺失值通过前向与后向填充补齐;数值型离群值利用四分位距(IQR)方法检测修正;数据经 Savitzky-Golay 滤波器(窗口长度10,二阶多项式拟合)平滑以去除噪声并保留趋势;所有数值型特征通过 Min-Max 归一化至[0,1]区间实现标准化;非数值型特征则通过多条件逻辑编码转换为数值形式,以构建完整统一的模型输入。

选取离群值处理与数据平滑方法结果作示意图,图2(a)为RPM在一号井的离群示例,红色点为离群值,蓝色点为非离群值。将筛选出来的离群值(红色标记)去除,该列的中位数来替代离群值。图2(b)为所有10口井数据的WOB特征列在使用SG滤波平滑后效果,蓝色为平滑前,橙色为平滑后。



(a) RPM一号井离群示例



(b) WOB平滑后效果

图2 预处理示意



图3 分井示意

在实时建模系统中,为模拟随时间推进的实时数据流,采用固定窗格大小的滑动窗口策略对主井进行数据切片,滑动窗口运行图解见图4。将窗口长度设置为固定值,固定窗口滑动步长为1,确保能够以最小单位获取随时间变化的实时数据特征,保证时间上的连续性,有助于捕捉短期内的数据动态变化,能够模拟实际钻井数据的采集。在单次窗口计算后,窗口向前滑动一步,重新开始计算。

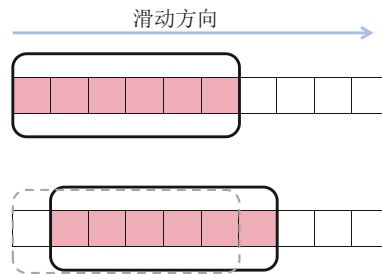


图4 滑动窗口图解

2 实时数据流处理

2.1 基本算法框架分析

为模拟实时数据流和表征实时钻进过程,将所得10口井的数据依据井号划分为主井数据和辅助井数据两部分,主井作为当前作业的目标对象,当前钻进的目标井,需要对其进行钻速预测,而辅助井是作为静态历史库,是已有的9口井历史数据,用于提供训练数据和模型构建基础。图3为数据集划分,符合石油工程中“一井一策”的作业原则并精准复现了现场“新井开钻→历史参考→实时修正”的决策链条。

在主井滑动窗口划分之后,针对不同井之间可能出现的深度偏移问题以及寻找主井窗口中心附近的历史井数据段,采用基于深度与井号以及岩性

三个维度的KNN近邻搜索机制,在9口辅助井的数据中分别找到与滑动窗口所取数据大致相同深度和岩性的数据。利用欧氏距离来衡量当前主井窗口中心深度与9口数据中每个深度点的相似性:

$$\text{Distance}(d_{\text{target}}, d_i) = \sqrt{(d_{\text{target}} - d_i)^2} \quad (1)$$

式中: d_{target} ——当前主井窗口的平均深度; d_i ——9口中第 i 个样本的深度值。

选取距离最小的前 K 个样本,除去主井外有9口井,为避免数据集的缺失,将 K 值设定为窗口大小的9倍,构成参考数据集用于后续频谱对比分析,这些深度大致相同的数据点将作为模型训练和频谱对比的基础。

2.2 频谱转化

钻速信号反映了钻头与地层相互作用的综合效应,受多种因素(如岩性变化、裂缝与钻具振动干扰等)共同影响,常在时域中表现出低频趋势与高频噪声并存的复杂形态。直接处理时域信号难以准确提取关键特征,易受噪声干扰。引入傅里叶变换,将信号转换至频域,从能量分布角度挖掘其潜在规律,其公式如下:

$$X(f) = \int_{-\infty}^{\infty} x(t) e^{-j2\pi ft} dt \quad (2)$$

式中: $X(f)$ ——对应的频域表示; $x(t)$ ——原始时域信号; t ——时间; j ——虚数单位; f ——频率变量。

傅里叶变换将信号视作不同频率正弦波的叠加,通过积分提取各频率成分的幅值与相位,所得频谱图揭示了能量在频域内的分布,有助于识别低频趋势、高频扰动及特征峰值。频谱分析不仅能揭示信号在不同频率上的能量分布,还可有效分离低频趋势与高频扰动,突出岩性变化等工况变化造成的振动特征。在钻进过程中,软硬地层交替或钻头磨损将表现为频谱图上高低频能量的结构性差异,因此通过频谱相似度进行地层工况识别具有物理意义与工程应用的合理性。

为分析钻井参数在频域里的变化规律,针对一号井全部数值参数开展了频谱分析,结果显示,钻速、钻压、转速等关键特征一般在0~5 Hz的低频段有能量集中现象,主峰频率大多集中在0.33、0.67、1 Hz等位置。这种低频占主导的谱形特征在多个参数里表现出一致性,意味着钻井系统的动态响应主要受低频波动推动。选取WOB参数的时频域图作

为代表性展示。

一号井的钻压时频域图见图5。图5(a)为时域图,其横轴表示采样点,纵轴表示归一化后的振动幅值,图5(b)为频域图,横轴为频率,表示将时域信号分解出的频率成分范围,纵轴为幅值,对应频率成分的能量强度。可以看出,在时域图里面,尽管钻压在波动,但整体趋势大致向上,与实际工程状况符合,钻进越深,钻压越大;而在频谱图里面,不难看出在低赫兹范围内出现的幅度较大,幅值在0.01~0.08间波动,无剧烈跳动,在0~2 Hz区间存在显著幅度峰,幅度约0.08,有一个主峰,多个较小的次峰,0 Hz主峰高且尖锐,说明当前钻压设置稳定,无剧烈波动。

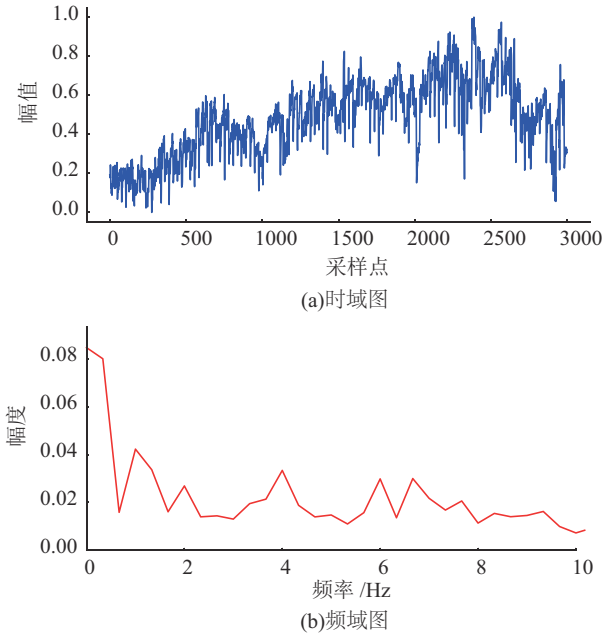


图5 一号井钻压时频域图

2.3 主井窗口对比分析

考虑到地层识别通常依赖多个钻井参数的共同作用,在频谱对比过程中同时考虑了多个特征维度。对于每一口历史井,分别计算主井窗口与9口历史井窗口在各个特征上的频谱相似度,并通过余弦相似度 c 进行评估,其公式为:

$$c = 1 - \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2 \sum_{i=1}^n B_i^2}} \quad (3)$$

式中: A_i ——主井频谱振幅值; B_i ——历史数据频谱振幅值。

本文采用余弦相似度衡量频谱图之间的相似性,因其具有幅值归一属性,能排除钻压、转速因操作波动引起的幅度变化,突出频率分布形态的匹配程度。与欧氏距离相比,余弦相似度更关注方向一致性。

WOB特征列主井原始频谱图与历史三号井的对比见图6,二者余弦相似度达到0.97,表明在此特征列中,三号井与主井具有足够的相似度。

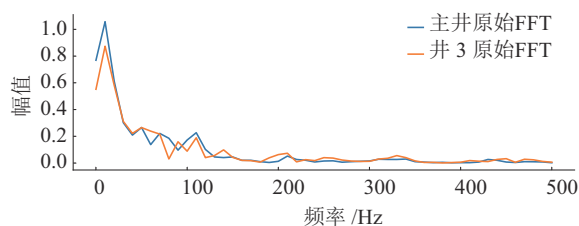


图6 频谱对比

取所有特征上的平均值作为该井与主井的整体相似度评分。最终,选取相似度最高的历史井作为主井当前地层的最佳匹配对象。若该井的最高相似度高于设定阈值,则认为主井当前地层与对应历史井高度相似,可复用其模型进行预测,否则触发重新建模流程。当前方法为对各特征频谱相似度取等权平均,然而不同参数在钻速变化中影响不同,例如钻压与转速通常与地层刚度关联性更强。后续工作中可引入特征权重机制,如基于Chatterjee相关系数确定频谱权重,进一步优化融合策略。

2.4 不同阈值下的预测精度对比

为深入探究频谱相似度阈值对建模效果的影响,设计了一组参数扫描实验。设定相似度阈值区间为0.6~0.9,每0.05一个梯度,逐一作为触发建模的判断标准,并在每一个阈值下完整运行一次预测流程,包括主井数据滑动窗口提取、与9口历史井的频谱对比、判断是否建模、建模策略选择以及最终ROP预测输出。在每轮实验完成后,统计对应阈值下所有窗口的平均 R^2 指标,用以评估整体建模性能。绘制不同相似度阈值对应的平均 R^2 变化趋势图,如图7所示,从中分析其对模型精度的影响程度,最终确定较为合理的相似度阈值。

结果表明,当相似度阈值设定为0.85时,模型能够在保证较高匹配精度的同时,最大化平均 R^2 指标,达到0.94以上。进一步提高阈值至0.9及以上时,预测性能提升幅度极小,且可用数据窗口减少。

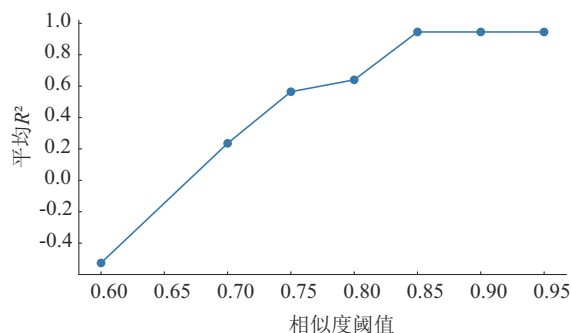


图7 不同相似度阈值下的 R^2

因此,将0.85作为最优阈值,能够在匹配准确性与样本数量之间取得良好平衡。值得注意的是,随着阈值上调,虽 R^2 略有提升,但模型重建频率也随之增加,计算时间由0.85的16 s/it(秒每迭代)上升至0.9的21 s/it,导致整体时间上升约30%。因此在应用中需权衡“精度提升”与“计算开销”,建议根据不同工况动态设定阈值以适应实时场景。

3 建模方法与结果验证

3.1 训练集测试集划分

对于训练集测试集的划分,采用了一种基于滑动窗口内部分割加上历史数据动态叠加的训练测试划分策略,最大程度兼顾预测精度与计算效率。

如前所述,在首次滑动之前,把9口井历史数据当作训练集来进行模型训练,之后滑动窗口滑动,如果频谱比对相似,就使用之前的9口井历史数据以及主井之前累计窗口所构成的数据集所训练的模型来进行预测,此时的测试集是该窗口所有数据。

如果相似度低于阈值,引入两个关键补充数据源:一个是9口参考井中与当前窗口深度最接近的数据点,保证模型在不同地质条件下有一定适应能力;另一个是之前已滑动过的主井窗口数据,这部分数据被动态累积,并逐渐融入到数据集中,有效提升模型对主井演化特征的掌握能力。将主井当前窗口与9口井中匹配深度段的数据合并,构建临时训练数据集,接着按照时间顺序把该数据集划分为训练集与测试集,此时的测试集为当前窗口的后20%数据。比如若窗口大小为100行,那么测试集就是后20行,而其余的数据为训练集,用于训练模型,如图8所示。为避免未来信息泄漏,模型仅使用当前滑动窗口及其之前的数据进行训练与预测,未

引入窗口之后的数据,确保预测流程严格符合时间顺序。

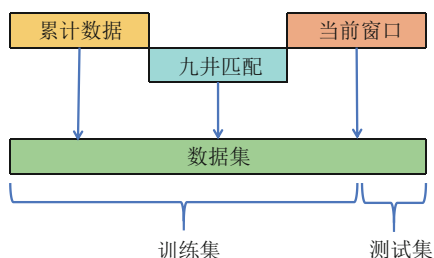


图8 不相似时数据集构成

3.2 拟用模型与评价指标

运用随机森林模型对所划定的数据进行建模训练,随机森林通过构建大量有差异且轻微过拟合的决策树,并集成其预测结果(投票/平均)来提升泛化性能。关键在于单棵树过拟合,但整体不过拟合。可对这些不同的树的结果求取平均值以降低过拟合现象,如此可以在减少过拟合状况的同时维持树的预测能力,随机森林可用于回归或者分类,借助 sklearn.ensemble 的 Random Forest Regressor 模块或 Random Forest Classifier 模块进行调用。

在本研究对模型进行评估期间,选取常用的回归类评价指标决定系数 R^2 。该指标在实际的工程建模和预测里有着广泛应用,可衡量模型的拟合效果以及预测精度。

随后设计了用于记录模型更新原因的机制,为优化工作给予了关键支持,每次进行模型更新时,系统都会注明是由于相似度低于阈值而触发更新,还是复用模型。经由对记录展开分析,可以对各参数进行调整。

3.3 模型性能评估

3.3.1 随机森林模型参数调整

为了系统评估树数量对钻速建模效果的影响,设计不同树数量下的敏感性测试。通过保持其他建模参数不变,仅调整随机森林中的树数量,从而分析该参数变化对模型 R^2 指标的影响程度。具体步骤如同阈值敏感实验一样。在实验过程中,选取了具有代表性的树数量区间,从10棵树开始,逐步递增至300棵树,记录每一次配置下的模型平均 R^2 值,如图9所示。

结果表明,当树的数量设定为100时,模型能够在保证较高精度的同时,最大化平均 R^2 分数,达到

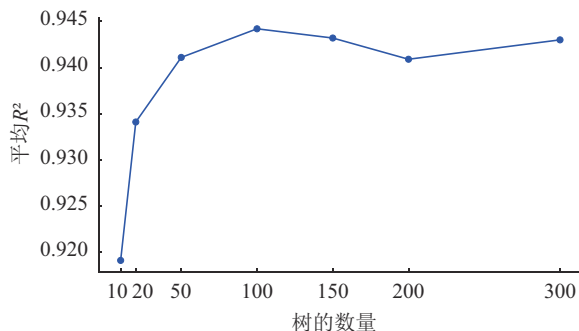


图9 不同随机森林树的 R^2

0.94以上。进一步增加树的数量对精度提升有限,但会明显增加训练时间。因此,本文最终选取窗口长度为100、树数为100作为默认参数,兼顾了预测性能与计算开销,为后续实时建模提供基础支撑。

3.3.2 窗口大小影响

如之前对于滑动窗口的论述,窗口的大小对于模型性能的影响同样重要,过大则无法捕捉短时间内的特征变化,失去精确性;过小则计算耗时大大提升,并增加重复数据。因此同上面两个实验一样,需要设置一个窗口大小的敏感度实验,具体步骤和上面实验保持一致。

从实验中得到的 R^2 变化趋势(图10)可见,窗口长度为100的数据覆盖程度适中,其平均 R^2 达到0.9651,能有效平衡建模效率与模型泛化能力,兼顾了数据完整性与模型响应速度。相比之下,窗口过小会造成特征波动过大、稳定性不足,而窗口过大则易引入历史信息冗余,影响实时性。因此,在实验中统一采用窗口长度为100的设置,作为滑动窗口建模的基础单元。

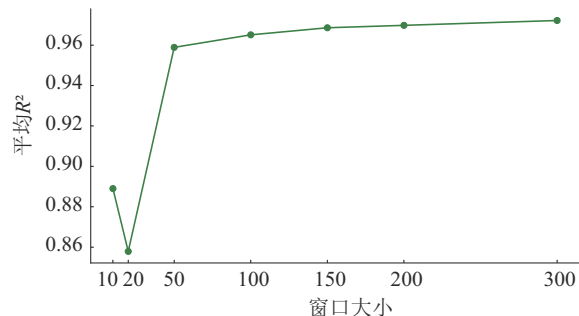


图10 不同窗口的 R^2

3.4 模型结果验证

在完成模型关键参数的设置后,有必要进一步检验所构建模型在实际钻进过程中的综合表现。

为验证本文提出的建模策略的有效性,构建了不同建模方式的对比实验:传统静态模型、本文模型,评估其在全井段范围内的预测准确度。通过真实值与预测值的差异,结合各方法下 R^2 变化趋势,全面评价本文模型在复杂工况下的适应能力与稳定性。

所设置的模型如下(见表1):(1)固定模型(Baseline):使用9口数据训练一次模型,对主井所有窗口统一预测;采用固定的随机森林模型,采用同样的数据预处理方法后,仅基于9口井的历史数据进行训练,并对主井所有窗口统一预测。(2)本文模型:采用了上述滑动窗口实时数据流处理及建模策略的处理。

表1 模型类型

类型	使用9口历史数据	使用频谱判断	实时建模
本文模型	是	是	是
固定模型	是	否	否

如图11所示,在固定模型中可以看出其 R^2 大部分聚集于0之下,均值为-4.48,负值多且波动大,证明该模型无法利用9口历史数据准确预测主井ROP,模型表现较差。

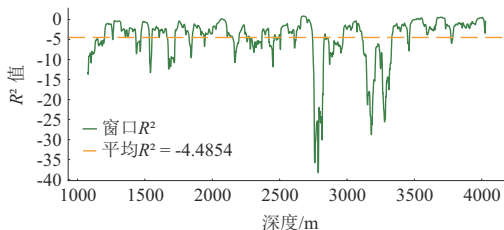


图11 Baseline模型 R^2

如图12所示,本文模型的 R^2 值在绝大多数深度区间内均保持在较高水平,稳定地接近于1,显著优于传统方法在相似场景下的预测效果,表明所构建的框架具有良好的拟合能力与实际预测能力。从图中可见, R^2 在大多数窗口中波动幅度极小,整体呈现出高位稳定的走势,其平均值达到0.99,远高于固定模型的性能基线,未出现低于零的情况,进一步证实模型在不同地层条件下均未出现预测失效或拟合崩溃的情况。

图12中还可看到某些深度段预测误差陡增,经对比钻遇地层发现,这些区域对应岩性变化与断裂带,导致钻压波动增强、ROP不稳定。结合测井岩

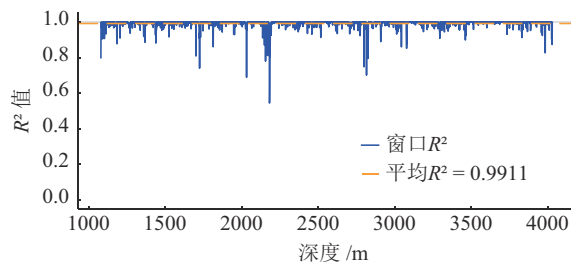


图12 本文模型 R^2

性日志分析表明,模型误差与岩性复杂度、裂缝发育密切相关,未来可进一步引入地层属性作为辅助特征,提升对突变地段的预测鲁棒性。这种现象反映出模型在应对地层突变、断层交界或参数跳变等极端复杂地质结构时,可能面临建模压力,但由于能够借助9口历史模型进行辅助回归预测,使得即便在突变区域也能保持一定的预测精度与稳定性。由此可看出,频谱对比机制在地层结构变异时仍具有良好的感知能力,而历史模型的融入有效增强了模型的泛化能力与抗扰动能力。

再进行本文模型的残差图绘制,如图13所示。

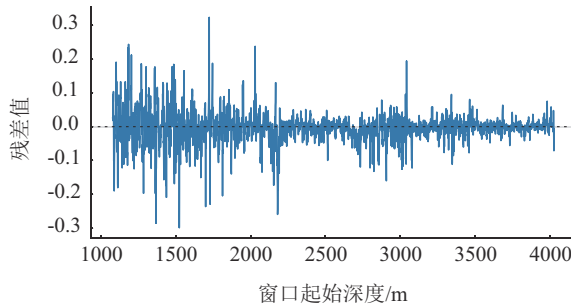


图13 本文模型残差

残差图呈现出钻速预测结果在各滑动窗口起始深度下的残差分布状况,其中残差被界定为真实ROP与预测ROP之间的差值。由图可明晰,大部分残差值围绕0展开分布,并且整体分布相对较为对称,模型于大多数窗口内的预测结果和真实值较为相近,并未出现系统性偏差,残差波动范围随着深度的增加逐渐收敛,此趋势意味着模型在初始阶段或许由于样本数量不足或者未完全掌握主井特征而表现出不稳定,随着训练数据的累积,模型预测能力逐步提高。部分极端残差对应的深度段较为集中,可能和地层突然变化或者测量误差存在关联,后续可结合地质信息加以验证,这一残差图为模型局部预测效果提供了直观的判断依据。

实验结果显示,本文模型在平均 R^2 上具有较高精度,且波动小,在地层突变区域具有更强的适应能力。表明本文所提出的机制能更有效地识别关键建模时机,在提升预测精度的同时降低不必要的计算开销。

4 结论

本文围绕钻井工程中钻速预测的需求,设计并实现了一个基于实时数据流对比的动态建模算法体系。研究从数据预处理入手,通过缺失值插补、离群值修正、平滑滤波与归一化处理,为后续建模提供干净高质量的数据源。在特征选择方面,剔除了不具备预测作用的井号与深度字段,并通过隐式使用井深信息,引导 K 近邻算法实现空间对齐,使特征与建模目标更紧密契合。

在实时数据建模策略设计中,本文引入滑动窗口机制,对主井数据进行逐点分析。每个窗口与历史9口井分别进行频谱对比,采用余弦相似度作为衡量标准,识别当前窗口与历史数据的地质相似性。当相似度高于设定阈值0.85时复用9口井模型,否则以主井历史窗口与近邻段数据为训练集,重新构建模型。参数敏感性实验进一步验证了该阈值设定的合理性,其对应的平均 R^2 值达0.94以上,展现出优良的模型选择策略。

在模型构建阶段,采用随机森林算法,调整树的数量以优化模型性能,实验表明树数量为100时可获得最优效果,平均 R^2 达到0.9651。此外,针对窗口大小对模型的影响展开对比试验,发现窗口长度为100时建模精度与实时性能达成较好平衡。

在最终结果验证中,模型整体预测性能表现优异, R^2 值稳定在0.99以上,残差围绕零值分布,且在地质突变区域也能保持较好的响应能力,展现出良好的泛化与自适应能力。

综上,本文所提出的钻速预测方法不仅能实时感知地质变化,还具备较高的建模精度和模型更新效率。相比传统静态建模方式,本系统融合实时数据流处理、频谱比对与动态建模等多个环节,有效提升了钻速预测的准确性和响应速度,具备一定的工程推广应用前景。但在极端复杂地质条件下,模型仍存在一定局限性,未来可结合更多地层参数优化。

参考文献:

- [1] Hegde C, Daigle H, Millwater H, et al. Analysis of rate of penetration (ROP) prediction in drilling using physics-based and data-driven models[J]. Journal of Petroleum Science and Engineering, 2017, 159: 295-306.
- [2] Brenjkar E, Deljani E B. Computational prediction of the drilling rate of penetration (ROP): A comparison of various machine learning approaches and traditional models[J]. Journal of Petroleum Science and Engineering, 2022, 210, 110033.
- [3] 高云伟,李博志,黄小龙,等.基于Stacking集成学习的机械钻速预测方法[J].石油学报,2022,43(5):803-811.
- [4] 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.
- [5] 张海军,张高峰,王国娜,等.基于遗传算法优化随机森林模型的机械钻速分类预测方法[J].科学技术与工程,2022,22(35):15572-15578.
- [6] 曾小龙,李谦,魏宏超,等.基于南海巨厚塑性泥岩地层特征的钻速预测模型[J].煤田地质与勘探,2023,51(11):159-168.
- [7] 焦通,张海军,李博志,等.基于深度学习方法钻速实时预测研究[J].石油机械,2021,49(6):45-50.
- [8] Tu B, Bai K, Zhan C, et al. Real-time prediction of ROP based on GRU-Informer[J]. Scientific reports, 2024, 14(1), 2133.
- [9] 黄小龙,王建胜,周长春,等.基于大数据及人工智能的钻速实时优化技术[J].石油钻采工艺,2020,42(3):307-313.
- [10] Haodong C, Yan J, Wandong Z, et al. Deep Neural Network Prediction of Mechanical Drilling Speed[J]. Energies, 2022, 15(9), 3037.
- [11] 任辉.基于循环神经网络的钻井ROP多步预测与优化研究[D].西安:西安石油大学,2023.
- [12] 李博志,黄小龙,高云伟,等.基于注意力机制的卷积神经网络机械钻速预测方法[J].石油学报,2023,44(2):297-305.
- [13] 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.
- [14] 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.
- [15] 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.
- [16] Mehmood E, Anees T. Challenges and Solutions for Processing Real-Time Big Data Stream: a systematic literature review[J]. IEEE Access, 2020, 8: 119123-19143.
- [17] 刘禹,李胜乐,李力.大数据平台的测震数据流实时可视化系统设计与实现[J].测绘地理信息,2022,47(4):77-81.
- [18] 王建胜,康芳玲.边缘计算在钻井数据流处理中的实现与优化[J].信息系统工程,2024(11):119-122.
- [19] 孙玉芬,卢炎生.流数据挖掘综述[J].计算机科学,2007,(1):1-5,11.
- [20] 李岩,王惠文,叶明.数据流分析与技术研究[J].计算机工程与应用,2008(15):8-11.

(编辑 王跃伟)