

基于不同相关性的钻速预测模型对比

魏思维¹, 李 谦^{1*}, 何俊杰¹, 李同意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: 针对钻进过程中的钻速预测与优化问题, 提出了一种基于相关性分析的建模算法。对来自10口油气井的21912组数据进行分析。首先对数据进行了预处理, 包括去除无关参数、填补缺失值、处理离群值、数据平滑和归一化等步骤, 确保了数据质量的可靠性。随后, 基于4种相关性计算方法(Pearson/Spearman/Kendall/Chatterjee), 分析了多个钻井参数与钻速之间的相关性, 发现钻井液密度、固相含量、泵量等参数对钻速影响显著。在建模方法上, 采用了多层感知机(MLP)和随机森林(RF)两种回归模型, 结合不同数量的相关性参数进行训练和验证。对比不同相关性算法与特征数量对模型性能的影响, Pearson算法在多层感知机模型中表现最佳, 而在随机森林模型中, 所有算法的表现均较为稳定, R^2 值高达0.96, 同时, 随机森林较多层感知机模型表现更为优秀。结果表明, 相关性分析不仅能够有效提取影响钻速的关键参数, 还能优化模型的建模过程。结合机器学习和特征选择技术, 本研究为钻速预测与优化提供了一种智能化的新途径, 并为实际钻井工程提供了理论支持。

关键词: 钻速预测; 相关性分析; 机器学习; 特征选择; 数据预处理

中图分类号: P634; TE242 **文献标识码:** A **文章编号:** 2096-9686(2025)S1-0112-07

Comparison of ROP prediction models based on different correlation measures

WEI Siwei¹, LI Qian^{1*}, HE Junjie¹, LI Tongyi¹, 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: A correlation-based modeling algorithm is proposed to address the prediction and optimization of the rate of penetration (ROP) in the drilling process. By analyzing 21912 data points from ten oil and gas wells, the data were preprocessed, including the removal of irrelevant parameters, filling missing values, handling outliers, data smoothing, and normalization to ensure data quality and reliability. Subsequently, four correlation calculation methods (Pearson/Spearman/Kendall/Chatterjee) were used to analyze the relationships between various drilling parameters and ROP, revealing that parameters such as drilling fluid density, solid-phase content, and pump rate significantly impact ROP. In the modeling approach, two regression models, Multi-Layer Perceptron (MLP) and Random Forest (RF), were employed, and different quantities of correlation parameters were used for training and validation. Pearson's method performs best in the MLP model, while the performance of all algorithms in the RF model remains stable with an R^2 value reaching 0.96. Additionally, the RF model outperforms the MLP model. The findings indicate that correlation analysis not only effectively extracts key parameters affecting ROP but also optimizes the modeling process. By combining machine learning and feature selection techniques, this study provides an intelligent new approach for ROP prediction and optimization and offers theoretical support for practical drilling engineering.

Key words: rate of penetration prediction; correlation analysis; machine learning; feature selection; data preprocessing

收稿日期: 2025-06-15; 修回日期: 2025-07-31 DOI: 10.12143/j.ztgc.2025.S1.017

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: 魏思维, 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, 3210408869@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: 魏思维, 李谦, 何俊杰, 等. 基于不同相关性的钻速预测模型对比[J]. 钻探工程, 2025, 52(S1): 112-118.

WEI Siwei, LI Qian, HE Junjie, et al. Comparison of ROP prediction models based on different correlation measures [J]. Drilling Engineering, 2025, 52(S1): 112-118.

0 引言

随着油气资源开发向深层及复杂地质结构的持续推进,钻井过程中实时预测与智能控制的需求变得日益强烈,钻速作为衡量钻井效率以及地层响应状态的关键参数,建模及其预测精度对于整个钻井工程的经济性和安全性有着决定性作用。

钻速预测与优化是提升钻井效率的关键。在传统方法中,Bourgoyne等^[1]提出的多元回归模型奠定了钻速预测的基础。近年来人工智能与数据科学技术发展迅速,研究者逐步引入机器学习与智能优化算法以提升模型精度与适应性,如甘超^[2]借助蝙蝠算法优化BYM参数,误差减少12%;Ehsan等^[3]则证实了PSO优化的MLPNN模型优于传统Bingham模型;Hegde等^[4]进一步指出数据驱动模型在复杂工况中具有更高精度与适应能力;王若禹^[5]通过结合卡尔曼滤波与遗传算法优化随机森林,显著提升了预测效果。在钻井参数优化方面,张海军等^[6]利用遗传算法与随机森林融合实现了82.1%的分类准确率;黄小龙等^[7]和曹祺^[8]分别提出了适用于实时调整的优化模型,提升了钻速与模型的环境适应性;王正^[9]从智能化角度出发,引入本体建模和范例推理实现参数优化。神经网络已成为主流的预测工具,严顾鑫^[10]通过粒子群算法优化网络结构;任辉^[11]采用RNN实现多步预测;屈峰涛^[12]融合大数据与智能算法提升钻速;焦通^[13]则提出引入注意力机制的CNN模型在预测表现上最为优越。与此同时,特征选择作为影响预测精度的核心步骤也被广泛关注,周长春等^[14]通过优化特征选择减少冗余提高准确率;曾小龙等^[15]在南海区域实测中利用KNN与随机森林显著提升了预测效果。

在数据科学推动下相关性分析和维度约简成为高维数据处理与建模的基础手段,殷辛瑜^[16]提出的基因表达相关性建模提升了预测精度;宋金玉^[17]基于最大信息系数优化了特征筛选效果;Chatterjee^[18]提出新型相关系数以克服传统Pearson系数的不足,为高相关特征选择提供更稳健的理论基础。

本文主要从相关性分析的角度出发,利用4种不同的相关性计算方法(Pearson/Spearman/Kendall/Chatterjee),分析多个钻井参数与钻速之间的相关性,结合机器学习和特征选择,对比不同相关性算法与特征数量对模型性能的影响。

1 分析数据综述

1.1 原始数据集

本文采用的数据均来自于10口油气井,包含超过40余种工程参数,将其分为操作参数、泥浆参数、地质条件、钻头参数、提速钻具5个大类,井深1000~4000 m。操作参数包含井径(Diameter)、钻速(ROP)、钻压(WOB)、扭矩(TORQUE)等14个参数,泥浆参数包含泥浆密度(MW)、固相含量、滤失量、稠度系数等18个参数,地质条件包含孔隙压力、破裂压力、上覆压力、岩性等6个参数,钻头参数由喷嘴数量、等效直径、内排磨损和外排磨损组成,提速钻具则由液力提速马达、水力脉冲和钻头齿形组成。钻井过程中每隔1 m进行一次测试,10口井获得的所有参数数据,组成了一个包含21912组数据的数据集。

1.2 数据预处理

(1)删除指定列:移除位置、井号两列,这两列与目标变量钻速无直接关联,移除岩性一列,其所表达的信息能用数据集中的岩性编号来表示。

(2)初步数据清洗:删除全部都是NaN的行,同时删除非空值少于70%的列,确保数据的质量,保留有效信息,避免无效数据影响后续分析,提高分析的可靠性和准确性。

(3)缺失值填充:本文采用的缺失值填充方法为中位数填充,中位数在统计学中表示将数据从小到大排列后处于中间位置的数值。

(4)离群值处理:其核心是基于数据的分布特征识别并修正那些明显偏离主体数据分布的极端数值,本文使用IQR(四分位距)法。

(5)数据平滑处理:采用Savitzky-Golay滤波器进行数据平滑处理,该方法基于局部多项式最小二乘拟合原理,不仅能有效滤除高频噪声,同时又可保留关键的工程特征。针对数据的边界效应,采用了向前填充和向后填充,确保了窗口的数据在数据边界也能正常计算。

(6)归一化处理:使用Min-Max标准化处理,将每一个特征列的数值线性映射到[0,1]区间上,有效解决了钻井参数量纲不统一导致的模型偏差问题。标准化之后,所有的特征被压缩到相同的尺度空间,数据分布不变。

以离群值处理和数据平滑处理的结果作为数据预处理的效果展示,见图1、图2。

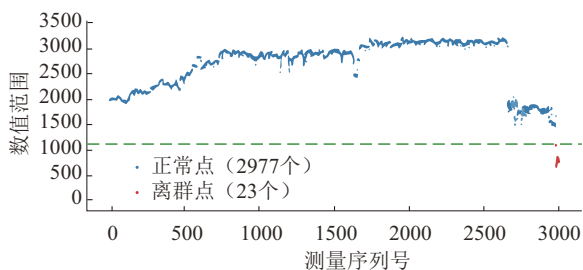


图1 离群值处理

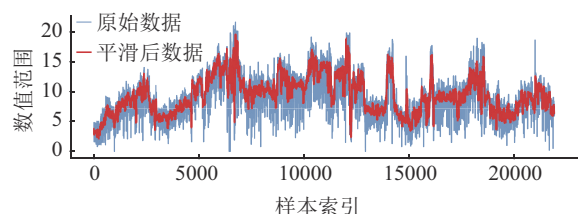


图2 数据平滑处理

图1是以一号井的立管压力(SPP)列的部分数据为例,其中绿线为下分界线,红点即为删除的离群值,蓝点为保留的数据。图2是以钻压(WOB)列为例,其中蓝线为数据平滑处理之前的情况,红线为数据平滑处理之后的情况,观察到数据平滑处理有效滤除高频噪声,同时又可以保留关键的工程特征(如钻压突变、扭矩峰值)。

由此可见,数据预处理确保了数据质量可靠,数据更加干净、一致,为后续相关性分析和机器学习建模奠定基础。

2 不同相关性的基本理论与钻井应用

2.1 相关性基本分析理论

(1) Pearson 理论:由 Karl Pearson 于 1895 年提出,作为线性相关的经典度量方法,其核心是通过标准化协方差衡量两个连续变量之间的线性关联强度。Pearson 相关系数 r 计算公式:

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

式中: n ——样本量; x_i, y_i ——第 i 个观测值; $\bar{x}, \bar{y}$ ——变量 x 和 y 的样本均值。

该系数的值域为 $[-1, 1]$, 其中 1 表示完全正线性相关, -1 表示完全负线性相关, 系数绝对值在 $[0.7, 1]$ 属于高相关性, 在 $[0.3, 0.7]$ 属于中相关性, 在 $(0, 0.3)$ 属于低相关。

(2) Spearman 理论:由 Charles Spearman 于

1904 年提出,是一种非参数方法,其核心价值在于摆脱分布假设和捕捉单调趋势,通过将原始数据转换为秩次后,计算 Pearson 相关系数来评估变量间的单调关系。Spearman 相关系数 ρ 计算公式:

$$\rho = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} \quad (2)$$

式中: d_i ——变量对应数据的秩次差,其本质是评估两变量单调关系的强度。

该系数的值域为 $[-1, 1]$, 系数绝对值在 $[0.7, 1]$ 属于高相关性, 在 $[0.4, 0.7]$ 属于中相关性, 在 $(0, 0.4)$ 属于低相关。

(3) Kendall 理论:由 Maurice Kendall 于 1938 年提出,通过评估数据对的协同性(Concordance)来量化变量间的单调关联。Kendall 相关系数 τ 为协同对数 C 与不协同对数 D 的标准化差值:

$$\tau = \frac{C - D}{n(n-1)/2} \quad (3)$$

式中: C ——协同对数(两变量排序一致); D ——不协同对数(排序相反)。

该系数的值域为 $[-1, 1]$, 系数绝对值在 $[0.5, 1]$ 属于高相关性, 在 $[0.2, 0.5]$ 属于中相关性, 在 $(0, 0.2)$ 属于低相关。

(4) Chatterjee 理论:由 Sourav Chatterjee 于 2021 年提出,是一种突破传统单调性限制的非参数相关性度量,能够检测任意形式的依赖关系(包括非函数型和非对称关联)。通过按 X 排序后观察 Y 秩次的局部波动性来量化依赖强度,Chatterjee 系数 ϵ 的计算公式:

$$\epsilon(X, Y) = \frac{\sum_{i=1}^{n-1} |R(y_{(i+1)}) - R(y_{(i)})|}{n^2 - 1} \quad (4)$$

式中: $R(y_{(i)})$ ——按 X 排序后 Y 的秩。

该系数的值域为 $[0, 1]$, 系数值在 $[0.5, 1]$ 属于高相关性, 在 $[0.3, 0.5]$ 属于中相关性, 在 $(0, 0.3)$ 属于低相关。

2.2 不同相关性计算结果对比

利用 Pearson 算法进行全局相关性计算,结果见图 3。Pearson 相关系数的绝对值在 $(0, 0.72]$, 相关系数中位数为 0.53, 由饼状图可知,各种工程参数中,高相关性占 14.3%, 中相关性占 47.6%, 低相关性占 38.1%, 相关系数绝对值最大的几个参数为滤失量、扭矩、深度、钻井液密度、泵量。

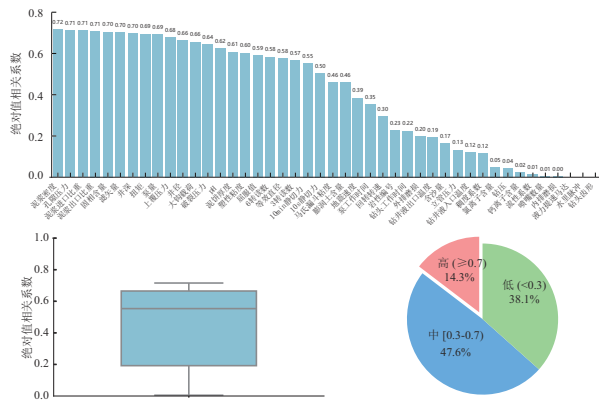


图3 Pearson相关系数

利用Spearman算法进行全局相关性的计算,结果见图4。Spearman相关系数的绝对值在(0, 0.79],相关系数中位数为0.54,由饼状图可知,各种工程参数中,高相关性占21.4%,中相关性占35.7%,低相关性占42.9%,其中,相关系数绝对值最大的几个参数为钻井液密度、固相含量、出口泥浆密度、入口泥浆密度、泵量、深度。

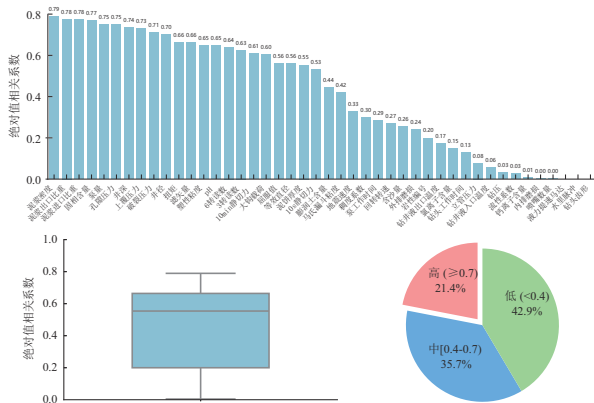


图4 Spearman相关系数

利用Kendall算法进行全局相关性的计算,结果见图5。Kendall相关系数的绝对值在(0, 0.58],相关系数中位数为0.38,由饼状图可知,各种工程参数中,高相关性占23.8%,中相关性占40.5%,低相关性占35.7%,其中,相关系数绝对值最大的几个为钻井液密度、井径、泵量、固相含量、出口泥浆密度。

利用Chatterjee算法进行全局相关性的计算,结果见图6。Chatterjee相关系数值在[0, 0.59],相关系数中位数为0.3,由饼状图可知,各种工程参数中,高相关性占8.9%,中相关性占42.2%,低相关性占48.9%,其中,相关系数值最大的几个参数为固相

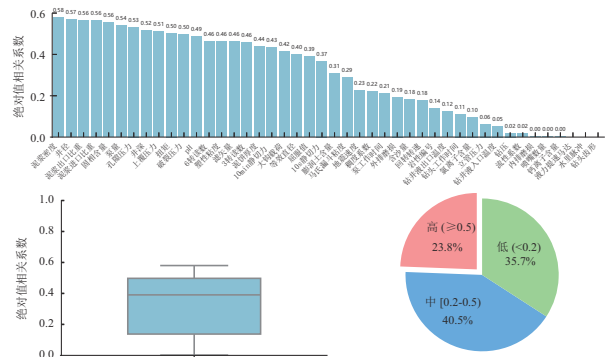


图5 Kendall相关系数

含量、钻井液密度、入口泥浆密度、出口泥浆密度、泵量、稠度系数。

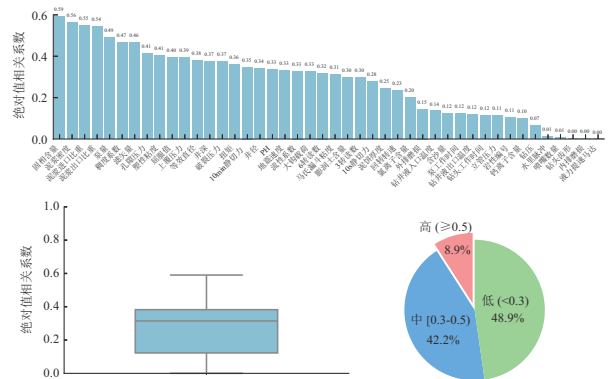


图6 Chatterjee相关系数

在进行全局相关性的分析之后,结合4种相关性计算方法各自高、中、低的阈值,绘制出4种相关性确定的高中低相关性参数数量对比图,见图7。Kendall算法的强相关性参数多达10个,而Chatterjee算法的强相关性参数仅有4个,差别较大,Pearson算法和Chatterjee算法的中相关性参数较多,分别为20个和19个,对于低相关性的参数数量而言,Kendall算法最少,仅有15个,而Chatterjee算法最多,多达22个。

综上所述,使用其他所有参数与钻速进行4种相关系数的计算之后分析发现,4种计算方法的结

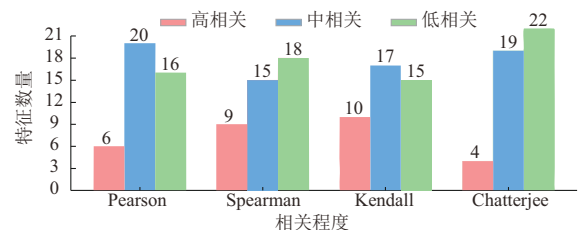


图7 高中低相关性参数数量对比

果在系数范围和参数的排序上存在着一定的差异,但是钻井液密度、固相含量、入口泥浆密度、出口泥浆密度、泵量、深度等几个参数在4种计算方法中都显著突出,表明这些参数对钻速具有系统性的影响。在这之中,钻井液密度和固相含量在4类相关系数中都排在前5名,很大的可能是二者都直接参与了对井壁压力的平衡控制:钻井液密度通过静液柱压力来维持井壁的稳定,固相含量过高可能会导致滤失量的增加,流变性能恶化,从而影响当前泥浆的循环密度,井壁压力的平衡一旦受到影响,钻速也会发生很大的变化。入口泥浆密度和出口泥浆密度同时具有显著性,表明钻井液性能在整个钻井循环的过程中的动态变化对钻井系统产生了累积性影响。泵量与深度都具有强相关性,反映出二者可能共同地影响着环空压耗和钻头的水力参数,进而影响钻速,在某些深井条件下,为了与井深相匹配,需要合理调整泵量,以避免压力的波动。地层的力学参数(上覆压力和孔隙压力)也具有普遍的显著性,因其与井壁的力学环境直接相关,这些参数的数值异常可能诱发井漏或者井涌。最后注意到,在Chatterjee相关系数的计算中,许多流变参数如稠度系数、塑性黏度的相关系数值较高,这说明了非线性的作用机制:钻井液的流变特性通过影响剪切速率的分布和携带岩屑的能力,以一种复杂的方式作用于钻速,不同的计算方法之间的差异性可能是因为各类参数与钻速之间关系的本质特征(线性关系、单调非线性关系、复杂非参数关联),也可能是因为钻速受到了参数间共同作用的影响。

3 钻速建模方法综述及整体建模效果对比

3.1 拟采用的回归算法及评价指标

本文拟采用的回归算法主要是多层感知机和随机森林。

(1)多层感知机:本文所用多层感知机模型采用双隐藏层结构,并且利用ReLU作为激活函数,多层感知机不仅可用于分类和聚类任务,也可用于回归任务。利用其在回归任务中的优点,强大的非线性建模的能力,可以拟合复杂的非线性关系。相比于传统的线性回归等方法,多层感知机可以有效捕捉钻速与各类钻井参数之间可能存在的非线性相互作用,其架构图见图8。

如图8所示,其架构为双隐藏层设计,第一隐藏

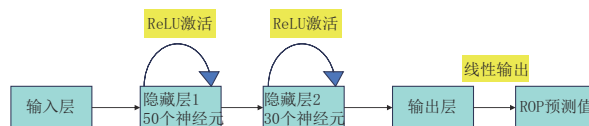


图8 多层感知机架构

层包含50个神经元,第二隐藏层包含30个神经元,通过这种深层结构增强了模型对钻井参数非线性关系的捕捉能力。输入层接收预处理后的特征数据(如钻井液密度、泵量等关键参数),通过全连接方式传递至隐藏层,隐藏层采用ReLU激活函数,该函数能有效缓解梯度消失问题并提升训练效率,最终,输出层生成连续的钻速预测值(ROP),实现端到端的回归任务。

(2)随机森林:随机森林算法适用于分类问题和回归问题的研究,对于钻速预测这种包含了复杂特征之间的相互关联的地质工程回归问题,随机森林展示出了卓越的适用性。其核心原理是通过构建多棵决策树的集成学习框架来实现高准确度的预测,其步骤如下:①首先通过Bootstrap采样从原始的数据集中随机抽取很多个子数据集(有放回的抽取);②从开始选择的每一个子数据集中随机选择 K 个特征,依据这 K 个特征建立决策树;③应用其中的每一棵决策树来预测结果;④最后根据所有决策树的预测结果,如果执行的是回归问题的研究,则对所有树的预测结果进行平均值的输出作为最后结果,如果执行的是分类问题的研究,则进行投票决策,即选择票数最高的模型作为最终决策。

图9展示了随机森林回归模型架构,该架构基于集成学习框架,通过构建多棵决策树实现高精度钻速预测。输入层接收预处理后的钻井参数特征(如钻井液密度、泵量等关键变量),通过Bootstrap采样随机抽取多个子数据集(有放回抽样),确保模型多样性和鲁棒性。在每棵决策树构建阶段,随机选择特征子集(如 K 个特征),避免过拟合并捕捉特征间的非线性交互关系。决策树通过递归分裂节点生成预测规则,最终通过集成机制输出结果:对于回归任务(如钻速预测),采用平均值聚合所有树的预测值;对于分类任务,则采用投票机制。

拟采用的评价指标:本文使用的两种回归模型的评价指标均为 R^2 ,该指标是衡量这两种回归模型的钻速预测能力的重要指标之一,通过比较每个回归模型的预测值与简单的预测计算平均值之间的

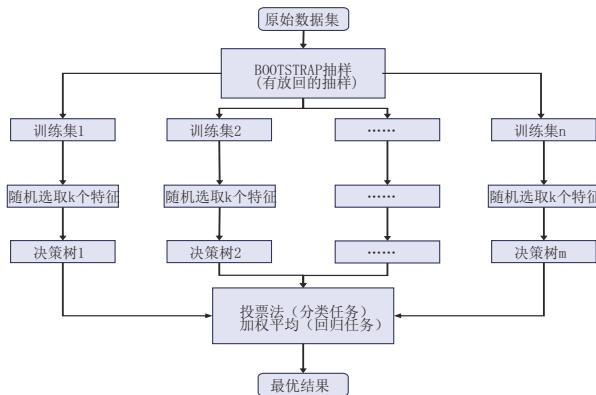


图9 随机森林架构

误差来衡量模型的拟合程度。 R^2 的计算公式:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (5)$$

式中: y_i ——实际观测值; $\hat{y}_i$ ——预测值; $\bar{y}$ ——平均值。

该评价指标对异常值较为敏感(因为是基于平方误差),与本文中数据预处理步骤中的Savitzky-Golay平滑滤波、箱线图法离群值处理之间形成了闭环,确保了评估结果的地质意义。

3.2 基于不同相关性的整体建模效果对比

进一步验证不同的相关性算法在钻速预测中的建模性能的差异,探索显著特征的数量对于模型预测精度的影响,找到较为合适的特征参数数量用以建模,步骤如下:(1)特征选择策略:针对4种相关性算法,根据高中低参数数量对比可知,4种算法的高相关性参数数量均小于10,为了让最终结果更可靠且不增加多余的计算时间和引入噪声,按相关系数绝对值大小排序后,依次取TOP5、TOP6、...、TOP15的特征参数,用以考察不同的特征参数数量对于模型性能的影响;(2)回归模型的选择:采用多层感知机和随机森林两种模型进行对比;(3)数据划分:数据按80%作为训练集、20%作为测试集随机划分;(4)评价指标: R^2 。

在上述操作步骤之后,得到了关于不同特征参数数量的建模效果图,为了更加直观地对比不同相关系数算法以及不同模型最后在训练集和测试集的结果上面的差异,更加直观地显示出随着特征数量的增加,模型的评分 R^2 的变化趋势,将训练集和测试集分开绘制出折线图,每一种算法用一种特定的折线表示,见图10~13。

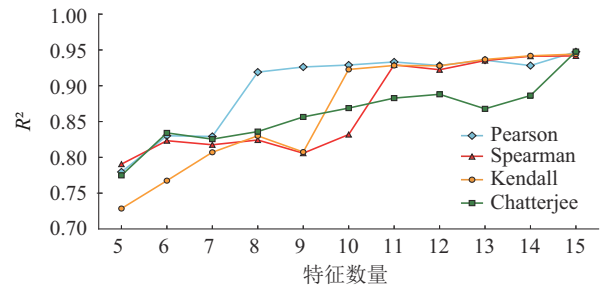


图10 多层感知机训练集

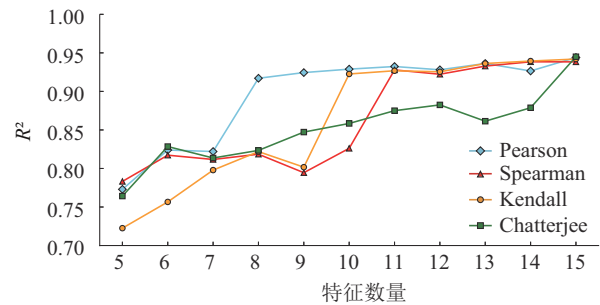


图11 多层感知机测试集

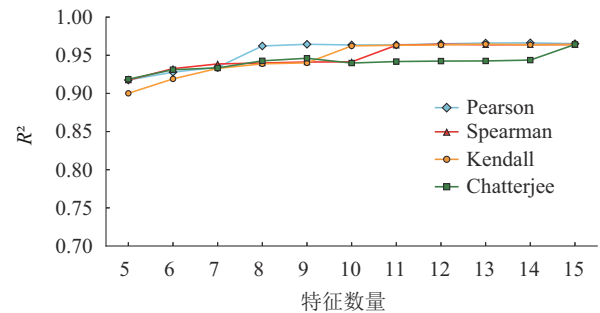


图12 随机森林训练集

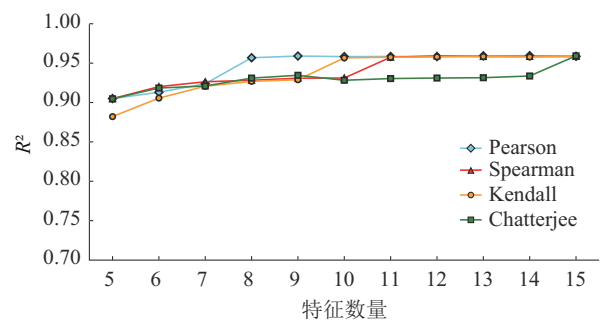


图13 随机森林测试集

分析图10~13可知:在整体建模中,对于多层感知机而言,Pearson算法整体的建模表现更好,特征参数数量选取为8个时,训练集和测试集的 R^2 均能达到0.90以上,特征参数数量达到10时,训练集

和测试集 R^2 均能达到 0.93。在特征参数较少时(数量 < 10), Kendall 算法表现最差, 特征参数数量为 5 和 6 的时候, R^2 均低于 0.80, 特征参数数量大于 10 时, 模型表现逐渐良好。Chatterjee 算法表现较为平均, 但均不高, 大多数情况下, R^2 在 0.80 至 0.90。由此可见, 针对多层感知机而言, 利用 Pearson 算法进行相关系数的计算之后再选取合适参数建模的方法是比较可靠的。对比 4 种算法可以发现, 特征参数数量在 11 之后均无明显变化。对于随机森林而言, 4 种算法的模型表现都比较优秀, 每种算法无论选取多少个特征参数进行建模, 模型表现都比较稳定, R^2 大多数在 0.90 至 0.96。综合比较, Pearson 算法的表现最为优秀, 大多数情况下 $R^2 > 0.96$ 。

最后, 进行两种回归模型的综合比较, 在整体建模中, 随机森林比多层感知机综合表现更加优秀, 与此同时, 在全局分析中, 利用 Pearson 算法进行特征参数的筛选更加合适, 同时, 特征参数的数量在 10 之后无特别明显的变化, 因此特征参数为排名中的前 10 个比较合适。

4 结论

本研究通过相关性分析和机器学习, 针对钻井过程中的钻速预测与优化问题进行了深入探讨。通过对 10 口井 21912 组数据进行分析, 发现钻井液密度、固相含量、泵量等参数与钻速存在显著相关性。采用 Pearson、Spearman、Kendall 和 Chatterjee 4 种相关性计算方法, 成功识别了影响钻速的关键参数, 并揭示了钻速受多种因素共同作用的复杂性。在回归建模中, Pearson 算法在多层感知机模型中表现最佳, 而随机森林模型在所有算法中表现更为稳定, R^2 值可达到 0.96, 表明随机森林模型更适用于钻速预测。此外还发现, 选择前 10 个重要特征参数能够有效提升模型性能, 特征数量增多后, 模型表现趋于稳定。结合机器学习与相关性分析, 本研究不仅提高了钻速预测的准确性, 还为实际钻井工程中的智能化决策提供了新的思路和方法。综上所述, 本研究为钻速预测与优化提供了有效工具, 具有重

要的应用价值, 能够在提升钻进效率和降低成本方面发挥积极作用。

参考文献:

- [1] Bourgoyne A T, et al. A Multiple Regression Approach to Optimal Drilling. SPE J 1974.
- [2] 甘超. 复杂地层可钻性场智能建模与钻速优化[D]. 武汉: 中国地质大学(武汉), 2019.
- [3] Ehsan B, Ebrahim D 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.
- [4] 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.
- [5] 王若禹. 基于机器学习的钻井机械钻速预测方法研究[D]. 西安: 西安石油大学, 2023.
- [6] 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.
- [7] 黄小龙, 刘东涛, 宋吉明, 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, 2021, 43(4): 442-448.
- [8] 曹祺. 钻井过程钻速实时预测和优化算法研究与应用[D]. 成都: 电子科技大学, 2023.
- [9] 王正. 钻井参数优化及其知识建模的研究[D]. 西安: 西安石油大学, 2013.
- [10] 严顾鑫. 基于优化神经网络的钻井机械钻速预测模型及应用研究[D]. 成都: 西南石油大学, 2022.
- [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] 屈峰涛. 基于大数据和智能算法的钻井参数优选模型与应用研究[D]. 西安: 西安石油大学, 2021.
- [13] 焦通. 基于深度学习方法的钻速实时预测研究[D]. 西安: 西安石油大学, 2023.
- [14] 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.
- [15] 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.
- [16] 殷辛瑜. RNA-Seq 数据的建模及相关性估计[D]. 合肥: 中国科学技术大学, 2023.
- [17] 宋金玉. 基于相关性分析的区间型数据特征选择方法研究[D]. 太原: 太原师范学院, 2024.
- [18] Chatterjee S. A new coefficient of correlation[J]. Journal of the American Statistical Association, 2020, 116(536): 1-21.

(编辑 王跃伟)