

文章编号: 1008-2336 (2024) 01-0092-04

## 机器学习预测机械钻速及在工程上的应用

郭家<sup>1</sup>, 刘烨<sup>2</sup>, 韩雪银<sup>1</sup>, 张宝平<sup>1</sup>, 林昕<sup>1</sup>

(1. 中海油能源发展股份有限公司工程技术分公司, 天津 300452; 2. 西安石油大学计算机学院, 陕西西安 710065)

**摘要:** 机械钻速是钻井工程中的关键指标, 直接关系到钻井成本, 准确判断机械钻速是钻井工程决策的关键。针对传统机械钻速预测方法误差大、时效低的问题, 尝试通过机器学习的方法获取更准确、可靠的机械钻速预测模型。首先采用不同的机器学习算法初步建立机械钻速预测模型, 然后通过性能比选优选出梯度提升树算法, 最终通过优化参数建立机械钻速预测模型。将训练好的预测模型用于南海某盆地一口探井, 预测结果符合实际。采用模型预测的结果可以用于机械钻速评价, 识别机械钻速异常, 为工程决策提供依据。

**关键词:** 机械钻速; 预测; 机器学习; 梯度提升树

中图分类号: TE52

文献标识码: A

DOI:10.3969/j.issn.1008-2336.2024.01.092

## Prediction of ROP by Machine Learning and its Application in Engineering

GUO Jia<sup>1</sup>, LIU Ye<sup>2</sup>, HAN Xueyin<sup>1</sup>, ZHANG Baoping<sup>1</sup>, LIN Xin<sup>1</sup>

(1. CNOOC EnerTech-Drilling & Production Co., Tianjin 300450, China; 2. School of Computer Science, Xi'an Shiyou University, Xi'an Shaanxi 710065, China)

**Abstract:** Rate of penetration (ROP) is a key performance indicator in drilling engineering, which is directly related to drilling cost. The accurate judgment of ROP is the key of drilling engineering decision. To solve the problems of large error and low effective of traditional ROP prediction methods, this paper attempts to obtain a more accurate and reliable ROP prediction model through machine learning method. Different machine learning methods are used to preliminarily establish the prediction model of ROP, and then the gradient lifting tree algorithm is selected for predicting ROP through performance comparison. Finally, the prediction model of ROP is established by optimizing parameters. The trained prediction model is applied to an exploration well in a basin in South China Sea, and the prediction results are in line with the reality. The results predicted by the model can be used to evaluate the ROP, identify the abnormal ROP, and provide a basis for engineering decision-making.

**Keywords:** ROP; forecast; machine learning; gradient boosting decision tree (GBDT)

机械钻速 (ROP) 直接反映钻进时的难易程度, 对机械钻速的预测和对井场实际机械钻速正常与否的判断, 是优化钻井工艺、避免复杂情况的关键, 因此钻速预测及评价对于石油钻井和地质勘探有重要的意义。

地层可钻性分析或者机械钻速计算非常不易, 目前以经验公式、现场试验、室内试验、参数反演等方法获得。室内试验由于岩石取样数量有限, 并脱离真实的地下应力环境, 测得的数据误差较大; 现场试验则需要花费大量的人力物力, 也同样存在试验数量有限的问题<sup>[1-2]</sup>。通常利用测井 (如声波测井、密度测井) 资料对岩石可钻性进行综合预测较为有效, 但是测井资料同样存在数据有限、获取成本高等问题, 建立区

域的岩石可钻性剖面需要进行大量的岩石性质测定实验和现场数据, 这将耗费大量的人力和资金<sup>[3-5]</sup>。上述传统的方法均不能实现根据钻进参数快速获得机械钻速的目的, 对实际工程上机械钻速监测和分析的指导作用不明显, 这就需要探索建立预测机械钻速的新方法。

近年来随着人工智能和计算机技术的兴起, 采用机器学习结合领域知识的方法, 分析大量的已钻井的钻井工程参数, 实现特定区块地层的机械钻速的准确预测已成为可能<sup>[6-7]</sup>。

基于机器学习算法预测机械钻速采用监督学习的方式, 模型由分析处理实际信息而来, 因而其预测结果理论上可靠。获得可靠的机械钻速模型后, 在工程

收稿日期: 2021-12-17; 改回日期: 2022-01-29

第一作者简介: 郭家, 男, 1988年生, 硕士, 工程师, 2013年毕业于北京交通大学桥梁与隧道工程, 从事海洋石油钻井技术与研究工作。

E-mail: guojia3@cnooc.com.cn。

中可使用实际钻井参数快速做出 ROP 的预测, 以预测值为参考对比实际 ROP, 模型预测出的 ROP 可以对其是否正常做出评价。

1 机器学习基本方法

机械钻速的预测是一种无法用精确的数学表达式来描述的非线性关系, 人工智能算法能自动学习数据特征和挖掘大数据隐藏的信息, 且其预测结果具有极高的准确度和可靠性。机器学习是人工智能的核心算法, 目前已应用于多个领域, 在石油钻井工程领域的应用也已兴起。

机器学习解决的问题可分为回归、分类、聚类问题, 预测机械钻速属于回归问题。机器学习的方法虽然有许多种, 但目前具备实用价值的大部分算法都来自于监督学习, 监督学习算法主要有线性回归、决策树回归和神经网络等<sup>[8-9]</sup>。

监督学习使用标注的训练样本 (x,y), x 为输入特征向量, y 为输出目标。通过训练已有的样本获得一个最优模型 (误差函数最小或某个评价准则下最优), 即得到一个从 x 映射到 y 的函数。这样给定模型一个输入, 将产生相应的输出, 也就具有了对数据的预测能力。

机器学习的方法直接从历史数据中找到机械钻速和工程参数的模型, 采用优化算法对模型参数进行优化, 建立机械钻速与工程参数之间的映射关系, 随后

便可随时利用实际观测的工程数据来预测 ROP, 通过对比分析预测的 ROP 与实际 ROP 实现对实际机械钻速的评价。

2 建立模型和应用

实际工程中影响 ROP 的参数非常多, 只能采用机器学习技术和专业领域知识相结合的方法, 选取关键影响因素降低训练特征维度, 采用不同区块的数据训练的模型适用相应区块。

据工程实践经验, 选择出直接和间接影响机械钻速的因素, 经过数据整合和处理, 形成数据集, 通过机器学习的方法构建机械钻速预测模型, 并用真实的历史数据进行测试。

建立模型的基本步骤: 数据准备 (收集, 预处理, 删除空值错值等), 建立模型及优化, 模型检验, 实际预测应用。

2.1 数据准备

根据工程实践, 选取关键参数作为输入量, 以钻头类型、钻头尺寸、钻压、转速、扭矩、排量、钻井液密度、黏度、测深、纯钻时间等 14 个输入工程参数作为输入 x, 以机械钻速作为输出 y。

由于地质条件的不确定性和复杂性, 由某一个区块的工程数据训练的模型仅适用于一个特定区块, 因此选取南海某区块已钻探的 10 口井的钻井工程 17 000 余条数据构成数据集, 数据信息见表 1。

表 1 数据集信息示例

Table 1 Dataset example

钻头类型	钻头尺寸 /in	刀翼数量	切削齿尺寸 /mm	井深 /m	悬重 /klbs	钻压 /klbs	转速 /(r/min)	扭矩 /(klb·ft)	排量 (L/min)	立管压力 /psi	泥浆密度 /(g/cm <sup>3</sup> )	泥浆黏度/s	钻头使用时间 /h	机械钻速 /(m/h)
牙轮钻头	17.5	空	空	59	85.3	0.5	39	1.09	2 892	627	1.05	28	0.02	136.31
牙轮钻头	17.5	空	空	60	85.2	0.6	44	0.97	2 892	625	1.05	28	0.03	34.69
牙轮钻头	17.5	空	空	61	85.1	0.7	54	0.48	3 636	625	1.05	28	0.04	117.18
牙轮钻头	17.5	空	空	62	85.2	0.6	51	0.73	4 268	613	1.05	28	0.05	80.74
牙轮钻头	17.5	空	空	63	85.4	0.4	38	0.30	4 267	612	1.05	28	0.06	59.18
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
PDC钻头	8.5	5	19	2 687	196.6	14.6	121	10.52	1 817	2 202	1.35	56	55.27	30.07
PDC钻头	8.5	5	19	2 688	198.7	12.5	122	11.37	1 814	2 192	1.35	57	55.3	40.38
PDC钻头	8.5	5	19	2 689	196.8	14.4	121	10.94	1 818	2 212	1.35	57	55.33	34.13

实际数据离散性大, 对样本数据进行归一化处理, 采用正态分布标准化模型:

$$(x - x\_mean)/x\_std$$

( 1 )

式中: x\_mean 为特征的均值; x\_std 为特征的方差。

训练钻井机械钻速预测模型时, 数据集划分按照训练数据占 80%, 测试数据占 20%。采用 5 折交叉验证。

2.2 建立模型和优化

为找到最适合的模型，先使用线性回归、随机森林、K 最近邻算法、梯度提升树等主流机器学习算法学习，初步建立预测模型，各模型评分见表 2。

表 2 各模型评分		
Table 2 Rating of different models		
序号	模型	决定系数( $R^2$ )
1	线性回归	0.628
2	随机森林	0.797
3	K最近邻算法	0.791
4	梯度提升树 (GBDT)	0.806

梯度提升树是机器学习算法里性能较好的几种算法之一，在机械钻速预测问题上相关系数最高，上述计算结果符合一般认识。

以梯度提升树方法作为基础，通过网格搜获算法优化模型参数，提高训练效果，得到性能最优模型，模型基本参数见表 3。

表 3 GBDT 模型优化训练参数		
Table 3 Optimized training parameters of GBDT model		
序号	参数	设置
1	最大迭代次数	5 000
2	学习率	0.06
3	损失函数	均方差损失
4	决策树最大深度	8
5	最大特征数	8
6	叶子节点包含的最小样本数	2
7	节点包含的最小样本数	2
8	子采样比例	0.7
9	随机状态参数	10

梯度提升树模型优化后评价指标见表 4，优化后模型的  $R^2$  分数由 0.806 提高至 0.866，性能提升 10% 左右。

表 4 GBDT 优化模型评价		
Table 4 Rating of optimized GBDT model		
序号	评价指标	数值
1	决定系数( $R^2$ )	0.866
2	平均绝对误差	9.97
3	均方误差	243.96

BP 神经网络也可用于机械钻速预测<sup>[10-11]</sup>，但是神经网络模型存在训练时间长、过拟合等问题。使用相同的数据集也训练了一个 4 个隐藏层的全连接神经网络模型<sup>[12-13]</sup>。

最大训练次数为 20 000 次，在训练 7 000 次后就

达到了稳定，使用均方误差绘制学习曲线，训练轨迹如图 1 所示。

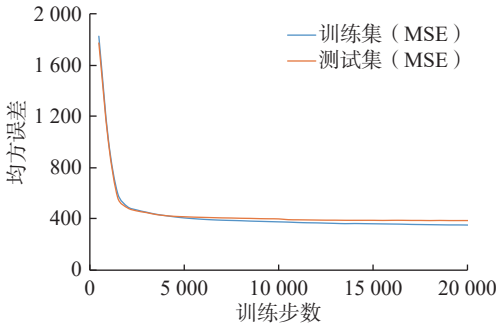
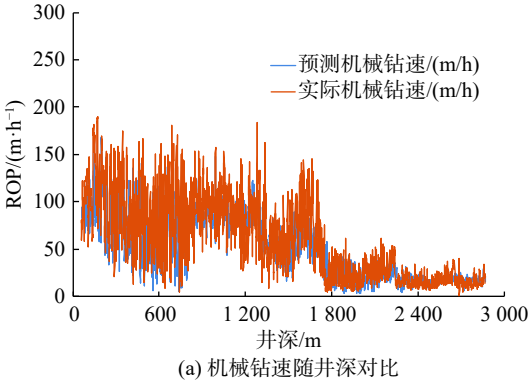


图 1 BP 模型训练轨迹图  
Fig. 1 Training error of BP network model

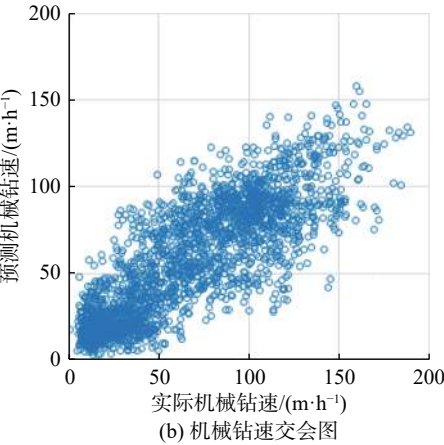
BP 神经网络模型均方误差为 383，高于 GBDT 模型，使用优化后的 GBDT 模型预测机械钻速比较适合。

2.3 模型检验

将训练好的 GBDT 模型用于测试集中的某口井的机械钻速预测，图 2 是机械钻速预测与实际值的对比，可见预测与实际值变化规律一致，且数值非常接近。



(a) 机械钻速随井深对比



(b) 机械钻速交会图

图 2 测试集中某井预测结果  
Fig. 2 Prediction results of a well in test set

该模型全井段平均误差 14.12%，除表层以外井段，平均误差在 10% 以内，该模型满足预期的要求。

2.4 应用实例

采用训练好的 GBDT 模型对南海某盆地 1 口探井

的工程数据进行实际预测分析,以期通过对比预测值和实际值,发现 ROP 是否异常。该井基本信息见表 5。

表 5 某井基本信息

Table 5 Information of a test well

钻头尺寸/in	井深/m	套管尺寸/in	井深/m
26	200	20	200
16	1 650	13-3/8	1 646
12-1/4	2 980	9-5/8	2 975
8-1/2	3 752	—	—

实际执行效果见图 3。图 3(a) 为不同井深处机械钻速预测与实际值对比,图 3(b) 为预测与实际机械钻速交会图。

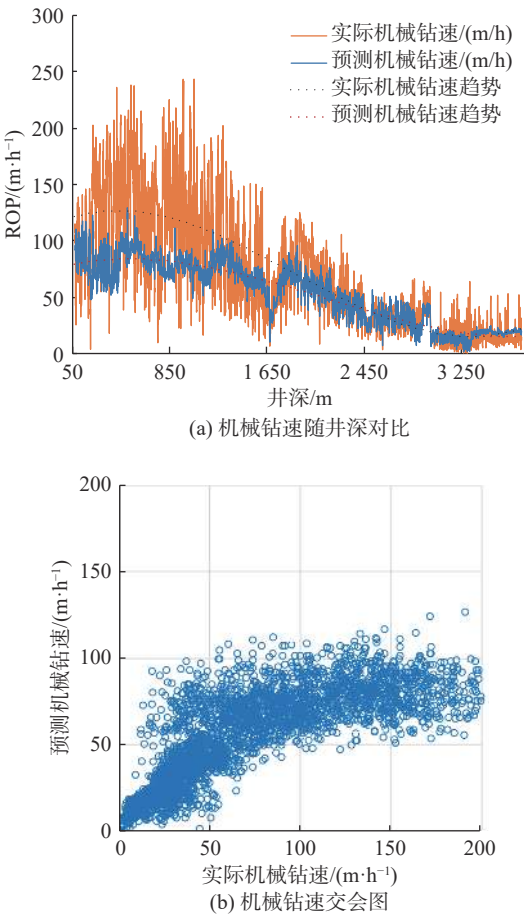


图 3 GBDT 模型实际应用效果  
Fig. 3 Application of GBDT model

结果分析: (1) GBDT 模型预测的结果整体变化规律符合实际。(2) 上部井段预测值偏差较大,原因有:地层成岩性差、地层不均质等造成机械钻速波动较大,也有钻进工艺的提高,机械学习方法本身的泛化性问题等原因,但实际值高于预测值是实际工程上所期待的,中下部井段预测较好。(3) 该井在 2500~3300 m 实际机械钻速持续低于预测值,经起钻检查钻头,发现水眼和流道被堵。

3 结论

- (1) 基于机器学习算法的机械钻速模型是一种监督学习的模型,分析处理实际信息,具有自组织、自学习、自适应的优点,因而其预测结果客观可靠,可用这种方法预测评价机械钻速。
- (2) 神经网络模型存在过拟合且训练时间长的问题,传统的机器学习 GBDT 模型训练速度快,精度较高,使用 GBDT 模型预测较好。
- (3) 基于 GBDT 方法的机械钻速预测模型,整体规律符合实际情况,在浅部地层误差较大,深部地层预测结果较好。可以使用 GBDT 模型实现对 ROP 的评价,即根据实钻 ROP 与预测 ROP 对比分析,及时发现机械钻速异常,做为实际工程上的一个决策依据。
- (4) 可继续优化模型参数,增加模型复杂程度,结合更详细的钻井信息提高预测效果。实现机器学习算法在机械钻速预测中的应用,表明人工智能在未来钻井工程中具有较大的应用前景。

参考文献:

[1] 李昌盛. 基于多元回归分析的钻速预测方法研究 [J]. 科学技术与工程, 2013, 13 (7): 1740-1744.

[2] 朱才朝, 谢永春, 伍奎, 等. 钻机系统故障智能诊断方法 [J]. 重庆大学学报, 2003, 26 (10): 97-100.

[3] 梁启明, 邹德永, 张华卫, 等. 利用测井资料综合预测岩石可钻性的试验研究 [J]. 石油钻探技术, 2006, 34 (1): 17-19.

[4] 范翔宇, 夏宏泉, 郑雷清, 等. 利用地震层速度预测地层可钻性和钻速的新方法 [J]. 钻采工艺, 2007, 30 (1): 4-6.

[5] 耿智, 樊洪海, 陈勉, 等. 区域三维空间岩石可钻性预测方法研究与应用 [J]. 石油钻探技术, 2014, 42 (5): 80-84.

[6] 谢祥俊, 张琥, 范翔宇. 基于最小二乘支持向量机的岩石可钻性研究 [J]. 西南石油大学学报 (自然科学版), 2010, 32 (1): 145-147.

[7] 吴俊杰. 钻井工程事故监测和预警方法研究 [J]. 录井工程, 2006, 17 (1): 53-55, 70.

[8] 刘胜娃, 孙俊明, 高翔, 等. 基于人工神经网络的钻井机械钻速预测模型的分析与建立 [J]. 计算机科学, 2019, 46 (6A): 605-608.

[9] Hoseinie S H, Ataei M, Osanloo M. A New Classification System for Evaluating Rock Penetrability[J]. International Journal of Rock Mechanics and Mining Sciences, 2009, 46 (8): 1329-1340.

[10] 沙林秀, 张奇志, 贺昱曜. 基于 SDCQGA 优化 BP 神经网络的岩石可钻性建模 [J]. 西安石油大学学报 (自然科学版), 2013, 28 (2): 92-97.

[11] 卢新卫, 金章东. 前馈神经网络的岩性识别方法 [J]. 石油与天然气地质, 1999, 20 (1): 84-86, 95.

[12] 邢军, 姜谔男, 邱景平, 等. 基于 DE-SVM 的岩层可钻性预测研究 [J]. 东北大学学报 (自然科学版), 2010, 31 (9): 1345-1348.

[13] Hecht-Nielsen R. Theory of the Backpropagation Neural Network[C]//Proceedings of the International 1989 Joint Conference on Neural Networks. Washington: IEEE, 1989: 593-605.