

基于测井数据的地层可钻性预测软件开发

数科2102班：刘洋 指导教师：刘志坤 论文类型：毕业论文

摘要：岩石可钻性是钻井工程中影响钻井效率的重要因素之一，通过预测地层的岩石可钻性可以分析当前地层的物理力学性质，进而提高钻井速度，降低钻井成本等。传统的可钻性评价方法有室内实验法，经验公式法等，存在着时效性差以及成本高等缺陷，本文以实际测井资料为基础，建立起预测岩石可钻性的单参数与多源参数的非线性回归方程，同时使用机器学习中的算法模型进一步预测岩石可钻性。本文主要开展了如下3个方面的内容：

- 一、采用室内可钻性级值试验得到的测井数据，建立出岩石可钻性级值与声波时差的单参数回归模型以及与声波时差，密度，泥质含量的多源参数回归模型。
- 二、利用多种机器学习算法，如随机森林，XG Boost进行岩石可钻性预测实验，并比较不同算法的优缺点，进而选出最优的算法模型。
- 三、根据上述建立的非线性回归模型与机器学习算法模型，搭建基于 Spring Boot + Vue + MySQL 技术栈的岩石可钻性预测系统。

通过上述开展的实验结果表明，非线性回归模型与多种机器学习算法模型均对岩石可钻性预测表现出较高的准确度以及置信度，并且机器学习算法模型较回归模型具有更高的准确性，更能适应不同的预测环境。

关键词：可钻性；回归；机器学习；测井数据

1 研究背景

近年来，随着我国油气资源勘探工程逐渐将重心转向到深层、超深层等复杂地层之中，钻井工程所面临的技术挑战同时也日益增多。其中，随着地层深度的增加，岩石的硬度，抗压强度及地层环境的复杂性也逐渐增加，这大大增加了钻井难度，严重影响了钻井效率，这时地层可钻性的准确判断在钻井参数设计、钻头选型以及降低钻井成本等方面就具有了重要意义。传统的可钻性评价方法大多依赖于钻井现场经验方法和地层录井资料数据，具有一定的主观性和滞后性，难以满足当前钻井工程中高效、安全钻井的需求。随着测井技术的快速发展和不断完善，使得钻井工程人员可以获取到较为精确的的岩石力学性质以及地层物理参数。测井数据不仅具备当前地层的实时性、连续性和物理客观性，为地层可钻性预测提供十分有效可靠的数据支撑。分析利用测井数据，预测出准确的岩石力学参数，以建立准确的地层可钻性预测模型，然后协助钻井工程人员设计决策出合适钻头，已成为当前石油工程领域的重要研究方向，同时对现实钻井工程有十分重要的指导意义。

2 发展现状

在1927年，Tillson首次提出了“可钻性”的概念，并指出它是衡量岩石在钻进过程中被破碎难易程度的重要指标。上世纪50年代的时候，我国开始引入前苏联史立涅尔提出的压入硬度法测定可钻性极值。1987年，我国在“微钻头钻进法”中取得重大突破，通过使用室内可钻性测定仪确定岩石可钻性，并且我国石油相关行业部门将可钻性极值划分为10个等级。

在2000年以后，国内学者开始加大测井资料对可钻性的研究实验，在2004年，邹德永等学者通过实验测定岩屑的声波波速，分析出岩屑波速与可钻性的关系，并定量的计算岩石可钻性极值。在2006年，邹德永等学者通过室内实验，在原有声波时差模型的基础上研究岩性，密度等参数对可钻性极值的影响。

在2023年，蒲先渤等学者通过编写LM-BP算法，合理设置预测模型参数值，以主成分分析后的数据样本作为基础，建立岩石可钻性等级预测模型。2025年，朱亮等学者提出一种基于SAE和LSTM神经网络相结合的组合模型对深部未钻地层的可钻性进行预测。

3 相关分析

3.1 单参数预测模型的建立

以实验测定的牙轮钻头可钻性级值为例，通过使用数理统计进行回归预测，选取决定系数与F检验值均高的模型，根据下表分析其中相关性最高的模型为二次函数回归模型，分析模型的拟合程度如下表所示：

表3.1 牙轮钻头可钻性级值与声波时差统计模型

数学函数	回归模型	R方	F
指数函数	$K_{druck} = 13.589 \cdot e^{-0.012 \cdot AC}$	0.841	47.763
对数函数	$K_{druck} = 23.991 - 4.3 \cdot \ln(AC)$	0.812	38.991
二次函数	$K_{druck} = 6.981 + 0.046 \cdot AC - 0.001 \cdot AC^2$	0.911	40.917

根据上表可以求解出声波时差与牙轮可钻性级值的单参数二次回归关系：

$$K_{druck} = 6.981 + 0.046 \cdot AC - 0.001 \cdot AC^2 \quad (3.1)$$

式中： K_{druck} 为牙轮钻头可钻性级值； AC 为声波时差，单位为 $\mu s / ft$ ，

根据上表可以得知，二次回归模型较比其他两个模型，表现出的拟合程度较好，同时查看相关系数检验表可知，二次预测模型的R方和F检验值远大于其对应的临界值，综合考虑认为二次回归模型较其他模型达到较高的拟合效果，且在整体数据中表现出了显著性，进而说明声波时差可以对牙轮钻头的可钻性级值进行预测。

3.2 多源参数预测模型的建立

分析模型参数值对整体模型的影响并寻求最优参数解，通过观察置信区间的范围来判断整体显著性，得到最优的参数值。求解出声波时差，密度，泥质含量与牙

轮钻头可钻性级值的多源参数回归关系：

$$K_{dr} = 9.762 - 3.125 \ln(AC) + 10.364 \ln(\rho) - 0.064 \ln(V_{sh}) \quad (3.2)$$

式中： K_{dr} 为牙轮可钻性级值； AC 为声波时差，单位为 $\mu s / ft$ ， ρ 为密度，单位为 g / cm^3 ， V_{sh} 为泥质含量，均由岩心原始实验数据得到，常数为系数项。

下表为模型概述与方差分析结果，查看F分布检验表结合下表可知，认为回归模型达到较高的拟合效果，进而说明声波时差，密度，泥质含量可以对牙轮钻头可钻性级值进行预测。

表3.2 牙轮钻头可钻性级值模型方差分析表

模型概述			
模型	R	R方	校正决定系数
多源参数模型	0.965	0.932	0.903
变量：常数项，LN(AC)，LN(DEN)，LN(Vsh)			
方差分析			
多源参数模型	平方和	均方	F
回归	10.614	3.538	32.164
残差	0.772	0.110	
总计	11.386		

在上述介绍了回归模型的建立过程，下面介绍了基于树的机器学习模型的建立过程，通过将两种不同的模型进行加权融合，通过网格搜索的方式确定最佳权重，最后整合在模型里，机器学习模型建立流程如下：选取特征，初始化模型，确定最优超参数，拟合模型。融合模型的预测性能如下表所示，可以发现，拟合程度达到了0.975，均方误差只有0.025，这说明融合模型拟合效果良好，可以满足地层可钻性预测功能要求，融合模型充分利用了两个模型在不同数据特征和模式上的优势，既保留了回归模型的解释性，又兼顾了基于树模型的非线性拟合能力，有效的提高整体预测效果。

表3.3 融合模型权重对比

融合模型	权重1	权重2	MSE	R方
随机森林与单参数公式	0.89	0.11	0.025	0.975
随机森林与多源参数公式	0.77	0.23	0.027	0.923

3.3 结果与分析

总结分析本文建立的单参数，多参数回归模型，多特征随机森林模型，以及融合模型，其中融合模型较比单独的单参数模型与多源参数模型的回归模型，以及随机森林机器学习模型表现更为良好，XG Boost模型表现最好，所有模型的均分误差

和决定系数均在良好的范围内，均可进行地层可钻性预测任务。

表3.4 机器学习模型性能分析

模型名称	MSE	R方
单参数模型		0.911
多源参数模型		0.932
多特征随机森林模型	0.038	0.962
多特征XG Boost模型	逼近于0	0.999
融合模型	0.025	0.975

4 研究结论及对策建议

4.1 研究结论

本文主要介绍了基于回归模型与基于树的机器学习模型的地层可钻性预测系统的实现，根据实验结果表明，融合模型较比单独的单参数模型与多源参数模型的回归模型，以及机器学习模型表现更为良好。同时建立起可钻性预测系统，实现完整的可钻性预测流程，通过导入测井数据，点击计算按钮，可视化可钻性级值等多通道曲线图，能够协助钻井工程人员完成可钻性预测工作。

总而言之，本次毕业设计为地层可钻性预测钻井工作提供了有益的指导意见和启示，为未来钻井工程提供宝贵的建议。

4.2 对策建议

在当前测井工程不断变化的背景下，仍然存在着诸多深入研究的内容，可以引入跨区域迁移学习、多源联合建模等方法，增强模型对多地质环境的适应能力。可考虑引入钻井工程中的钻进参数等指标，如钻压、转速、排量、机械钻速、钻头类型等，构建多维特征体系，更真实地反映钻井过程对地层可钻性的影响。