

“大数据与智能钻探技术装备专题”编者按:通过融合大数据与人工智能技术,形成面向深部和复杂地质条件的高端钻探技术与装备,是突破高风险钻探工程核心技术的关键,也是钻探工程科技发展的方向。为此,中国地质大学(武汉)与编辑部联合组织了“大数据与智能钻探技术装备专题”,由甘超副教授担任客座编辑。本专题遴选7篇论文发表,内容涵盖钻探过程中大数据的分析应用、智能钻探装备、智能钻探钻井技术等,以期促进钻探技术与装备的发展创新。

深部地质钻探钻进过程流式大数据分析 与动态预处理

——以辽宁丹东3000 m科学钻探工程为例

甘超^{1,2,3}, 曹卫华^{*1,2,3}, 王鲁朝⁴, 吴敏^{1,2,3}

(1. 中国地质大学(武汉)自动化学院,湖北武汉430074;

2. 复杂系统先进控制与智能自动化湖北省重点实验室,湖北武汉430074;

3. 地球探测智能化技术教育部工程研究中心,湖北武汉430074;

4. 山东省第三地质矿产勘查院,山东烟台264000)

摘要:深部地质钻探钻进过程数据价值密度低,传统方法难以在钻中流式大数据条件下有效去除尖峰、毛刺等各类钻进过程数据噪声。本文提出一种深部地质钻探钻进过程流式大数据分析与动态预处理方法,并成功应用于辽宁丹东3000 m科学钻探工程。首先,深入分析过程工艺和数据处理需求,建立深部地质钻探钻进过程流式大数据分析与动态预处理框架结构;然后,运用限幅滤波结合过程数据分布特征、司钻/机长人工操作经验去除过程数据中的离群值;接着,引入滑动窗口策略对流式钻进大数据进行动态处理,在每个窗口中运用Savitzky Golay滤波进一步提升数据质量。仿真实验和工程应用结果验证了本文方法具有很好的工程适用性和有效性。

关键词:深部地质钻探;钻进过程;流式大数据分析;流式大数据预处理;科学钻探

中图分类号:P634 **文献标识码:**A **文章编号:**2096-9686(2022)04-0001-07

收稿日期:2022-05-05; 修回日期:2022-06-10 DOI:10.12143/j.ztgc.2022.04.001

基金项目:国家自然科学基金青年项目“基于多源井震信息融合的地质钻进过程钻速智能优化”(编号:62003318);国家自然科学基金面上项目“复杂地质钻进过程效率动态优化与安全智能预警”(编号:62173313);国家自然科学基金重点项目“复杂地质钻进过程智能控制”(编号:61733016);湖北省自然科学基金创新群体项目“地质钻探智能化技术及应用”(编号:2020CFA031);中央高校基本科研业务费专项资金科研项目“考虑复杂地质环境的钻进过程钻速优化”(编号:CUG2106350)

第一作者:甘超,男,汉族,1990年生,中国地质大学(武汉)副教授,硕士生导师,主要从事复杂地质钻进过程建模与优化控制的研究工作,湖北省武汉市鲁磨路388号,ganchao@cug.edu.cn。

通信作者:曹卫华,男,汉族,1972年生,中国地质大学(武汉)自动化学学院院长,教授,博士生导师,主要从事过程控制、智能系统和机器人技术的研究工作,湖北省武汉市鲁磨路388号,weihuacao@cug.edu.cn。

引用格式:甘超,曹卫华,王鲁朝,等.深部地质钻探钻进过程流式大数据分析与动态预处理——以辽宁丹东3000 m科学钻探工程为例[J].钻探工程,2022,49(4):1-7.

GAN Chao, CAO Weihua, WANG Luzhao, et al. Streaming big data analysis and dynamic pre-processing in deep geological drilling process: A case study on the 3000m scientific drilling project in Dandong, Liaoning province[J]. Drilling Engineering, 2022, 49(4): 1-7.

Streaming big data analysis and dynamic pre-processing in deep geological drilling process: A case study on the 3000m scientific drilling project in Dandong, Liaoning province

GAN Chao^{1,2,3}, CAO Weihua^{1,2,3}, WANG Luzhao⁴, WU Min^{1,2,3}

(1.School of Automation, China University of Geosciences, Wuhan Hubei 430074, China;

2.Hubei Key Laboratory of Advanced Control and Intelligent Automation for Complex Systems, Wuhan Hubei 430074, China;

3.Engineering Research Center of Intelligent Technology for Geo-Exploration, Ministry of Education, Wuhan Hubei 430074, China;

4.No.3 Exploration Institute of Geology and Mineral Resources of Shandong Province, Yantai Shandong 264000, China)

Abstract: The data quality in deep geological drilling process is poor, and traditional methods are hard to effectively remove all kinds of data noise such as spikes and burrs. A streaming big data analysis and dynamic pre-processing method for deep geological drilling process was proposed and successfully applied to the 3000m scientific drilling project in Dandong, Liaoning province. Firstly, the process mechanism and requirements of data processing are deeply analyzed, and the framework of streaming big data analysis and dynamic pre-processing in deep geological drilling process is established. After that, the outliers in the process data are removed by limiting filtering combined with the distribution characteristics of the process data and the driller's manual operation experience. Then, the moving window strategy is introduced to dynamically process the big data of convective drilling, and savitzky Golay filter is used in each window to further improve the data quality. Finally, results of simulations and engineering application show that the proposed method has good engineering applicability and effectiveness.

Key words: deep geological drilling; drilling process; streaming big data analysis; streaming big data pre-processing; scientific drilling

0 引言

深部地质钻探钻进过程中遇到的地层层位多、压力体系复杂,具有高地温、高地应力、高陡构造和钻进扰动的复杂地质力学环境^[1-5]。受地层环境、钻进工艺、传感元件等多种因素综合影响,在深部地质钻探钻进过程中采集的各类数据常具有尖峰和毛刺等噪声^[6-9]、数据价值密度低,给进一步分析和利用数据提升深部地质钻探的安全性和效率带来较大困难。因此,深入地分析与预处理深部地质钻探钻进过程的流式大数据对提升数据质量、实现安全高效深部钻探具有重要的科学与工程意义。

通常,分析与预处理深部地质钻探钻进过程数据的方法可以分为2类。一类是将过程数据收集齐后再统一分析与预处理的静态法;另一类则是在钻进过程中“边钻边分析与预处理”的动态法。本文主要基于后一类方法开展研究。

许多学者运用静态法对钻进过程数据展开了大量研究^[10-13],在分析数据分布特征、去除数据噪声等方面取得一定成效。作者^[8]整理了湖北省神农架林

区某口地热井的钻进过程数据,首先通过频谱分析和互信息分析方法,确定了数据存在尖峰与毛刺、以及高低频噪声并存等特点;然后使用小波滤波方法有效提升了钻进过程数据质量。伊朗阿米尔卡比尔理工大学 Ashrafi 等^[10]收集了伊朗西南部马伦油田一口直井的1000组钻进过程数据,并采用 Savitzky Golay(SG)滤波方法有效降低数据噪声。此外,成都理工大学李谦等^[13]对中国南海10口井的钻进过程数据进行了分析和处理,通过数据清洗、缺失值补充、错误值删除等方法,确定了20000多组有效的钻进过程数据。这类方法是对钻进过程数据开展的钻后分析与处理,虽然能够为分析数据分布和去除数据噪声提供参考,但是无法直接运用于钻中对动态流式大数据进行分析与处理。

为克服静态法在处理动态流式钻进过程数据时的不足,部分学者开展了动态法的研究,并应用于实际钻进过程。范海鹏等^[2]综合运用限幅滤波和滑动窗口策略,有效去除了钻进过程数据中的离群值,为后续开展钻进过程的建模与优化控制研究奠定了良

好的数据质量基础,工程应用结果证明了所提方法的有效性。韩国科学技术联合大学院大学 DIAZ 等^[14]采集了一口4000 m地热井钻进过程数据并模拟在线钻进过程,先后运用快速傅里叶变换和滑动窗口策略实现了钻速等钻进过程数据的在线预处理。动态法常将各类滤波方法与滑动窗口策略等流式大数据处理方法相结合,从而实现钻进过程流式大数据的分析与预处理。但是深部地质钻探钻进过程地层情况复杂,过程数据噪声大,如何有效地分析与预处理深部地质钻探钻进过程流式大数据仍是亟待解决的难题。

针对上述难点,本文提出了一种深部地质钻探钻进过程流式大数据分析与动态预处理方法,优势之处在于:

(1)针对深部地质钻探钻进过程数据突变明显、高低频噪声并存等特点,综合运用限幅滤波、滑动窗口策略、SG滤波方法,有效去除了钻进过程流式大数据中存在的尖峰与毛刺;

(2)相比其他钻进过程数据分析与静态处理方法,本文方法运算速度快、能在线运行,满足实际工程需要;

(3)本文方法极大地提高了深部地质钻探钻进过程数据质量,为进一步利用数据实现深部地质钻探钻进过程建模与优化控制奠定了重要基础。

1 深部地质钻探钻进过程流式大数据分析与预处理框架

深部地质钻探钻进过程流式大数据分析与预处理框架如图1所示。框架包括2层,分别是基础自动化层和深部地质钻探钻进过程流式大数据分析与预处理层。2层结构可以各自完成相应功能,并实现双向通讯。

基础自动化层主要完成回转控制系统、进给控制系统和循环控制系统等的过程参数采集和变频控制,从而实现旋转钻进、起下钻等基本功能。深部地质钻探钻进过程流式大数据分析与预处理层是在已有基础自动化层的基础之上开发的一套软硬件系统(图2),它是由中国地质大学(武汉)自动化学院研究团队开发的复杂地质钻进过程智能控制系统的一套子系统。系统软件指深部地质钻探钻进过程流式大数据分析与预处理软件,首先通过限幅滤波去除掉钻进过程数据中的离群值,接着采用滑动窗口策

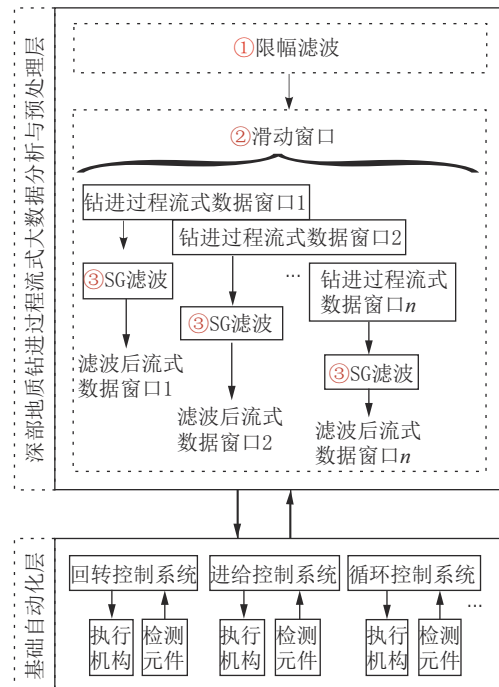


图1 深部地质钻探钻进过程流式大数据分析与预处理框架

Fig.1 Framework of streaming big data analysis and pre-processing in deep geological drilling process

略将过程数据装载入动态滑动的窗口,然后对窗口中的过程数据进行SG滤波处理,最终实现了深部地质钻探钻进过程流式大数据的分析与动态预处理。系统硬件主要包括:工业控制计算机(工控机)、触摸屏、专用通讯接口和线缆;工控机及其搭载的深部地质钻探钻进过程流式大数据分析与预处理软件放置在电控房内,它通过用于过程控制的对象链接和嵌入OPC(OLE for Process Control)协议与基础自动化层的可编程逻辑控制器(Programmable Logic Controller, PLC)进行通讯,获取钻进过程数据并将经过分析与预处理后,再进一步将计算得到的最优钻进操作参数下发给基础自动化层进行跟踪控制。司钻房中的触摸屏与电控房中的工控机相连,主要用于显示处理前后的钻进过程参数和最优钻进操作参数供司钻人员参考,且具有最优钻进操作参数设定的功能。

2 深部地质钻探钻进过程流式大数据分析与预处理方法

基于图1的框架结构,深部地质钻探钻进过程流式大数据分析与预处理方法包括3部分内容。



图2 深部地质钻探钻进过程流式大数据分析与预处理系统与工程现场情况
Fig.2 System of streaming big data analysis and pre-processing in deep geological drilling process, and drilling site situation

2.1 限幅滤波

深部地质钻探钻进过程采集的数据受地层、工艺、传感器等多因素影响,存在大量离群值。如果不能有效处理这些离群值,则不利于后续开展钻进过程建模与优化控制研究。通常,在实际工程中一般会结合过程数据分布特征、司钻/机长人工操作经验来确定限幅滤波阈值。

2.2 滑动窗口

在深部地质钻探钻进过程中会采集钻压、钩载、钻速、转速、扭矩等核心参数,这些参数的采样频率一般是1 Hz。以一口3000 m的深部地质钻探工程来看,工程周期往往长达数月甚至超过1年,在这过程中会采集上千万组的钻进过程流式大数据。如果采用传统静态法或动态累积式方法进行分析与预处理,会使得计算时间随着数据的积累越来越长,不适用于对实时采集的流式大数据进行分析与预处理。采用滑动窗口策略则可以将采集的过程数据放入数据窗口,在窗口中对有限的过程数据进行分析与处理,加快处理速度适应实时计算需求,并通过窗口滑动的形式实现动态流式大数据的迭代计算。滑动窗口策略中滑动窗口宽度和滑动距离对处理效果有一定影响,需要具体分析后确定。

2.3 SG滤波

SG滤波^[15]是1964年由Savitzky和Golay提出的时域内基于多项式最小二乘拟合的滤波方法,该方法可以提高数据平滑性、降低数据噪声的干扰。SG滤波中有2个核心参数,影响着滤波的效果,分别是多项式的阶数与滑动窗口的宽度。多项式阶数越大时,拟合精度越高,但是泛化性能越低,反之亦

然;滑动窗口宽度越小时,滤波前后数据相差的数值越小,反之亦然。

上述方法中限幅滤波阈值、滑动窗口宽度、滑动距离、多项式阶数等核心参数的合理选择,直接关系到深部地质钻探钻进过程流式大数据分析与预处理方法的表现。因此,需要开展仿真实验,对实验结果进行分析,并确定上述核心参数,然后在此基础上再开展工程应用实践。

3 仿真实验与工程应用情况

首先采集来自辽宁丹东3000 m科学钻探工程的钻进过程数据,开展仿真实验分析,确定第2章中相关方法的核心参数;然后再将本文方法应用于辽宁丹东3000 m科学钻探工程(图2),具体情况如下。

3.1 仿真实验

本节选取从辽宁丹东3000 m科学钻探工程采集得到的80万组多维(钻压、钩载、钻速、转速、扭矩)钻进过程原始数据进行仿真实验。部分原始数据的最大值、最小值、平均值和标准差统计如表1所示。

表1 辽宁丹东3000 m科学钻探工程部分原始数据统计

Table 1 Some raw data of 3000m scientific drilling project in Dandong, Liaoning province

名称	钻压/ kN	钩载/ kN	钻速/(cm· min ⁻¹)	转速/(r· min ⁻¹)	扭矩/ (N·m)
最大值	99.930	201.600	999.4	1023	9987
最小值	0	0	0	0	0
平均值	7.349	57.777	4.0	118.7	514.3
标准值	9.850	46.971	38.7	127.6	582.1

根据表 1 数据分析可知,深部地质钻探钻进过程数据存在大量离群值,运用核密度估计法对上述数据的分布情况进行计算,得到如图 3 所示的辽宁

丹东 3000 m 科学钻探工程部分原始数据概率密度估计。可以看出数据呈非高斯分布,数据分布离散,对应的数据概率密度估计值低。

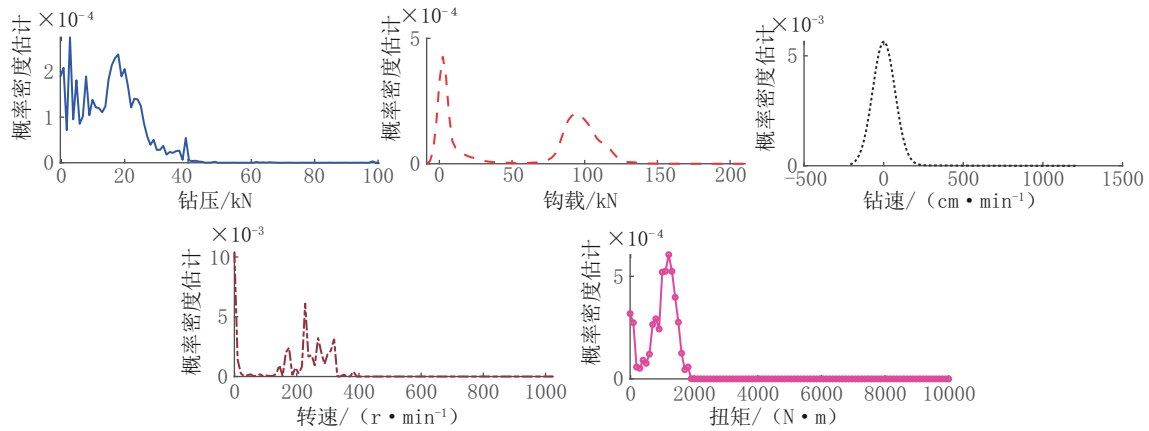


图 3 辽宁丹东 3000 m 科学钻探工程部分原始数据概率密度估计

Fig.3 Density estimation of some raw data of 3000m scientific drilling project in Dandong, Liaoning province

运用本文方法对图 3 中的原始数据进行处理,结合过程数据分布特征、司钻/机长人工操作经验以及试凑法得到本文方法中的核心参数取值,获得处理后的辽宁丹东 3000 m 科学钻探工程部分数据概率密度统计如图 4 所示。可以明显看出,经过本文

方法分析与预处理后的过程数据分布更加集中,去除了数据中存在的尖峰和毛刺,处理后数据的概率密度估计值相比原始数据要大许多,且近似符合高斯分布,提升了深部地质钻探钻进过程数据质量,为本文方法的工程应用奠定了重要基础。

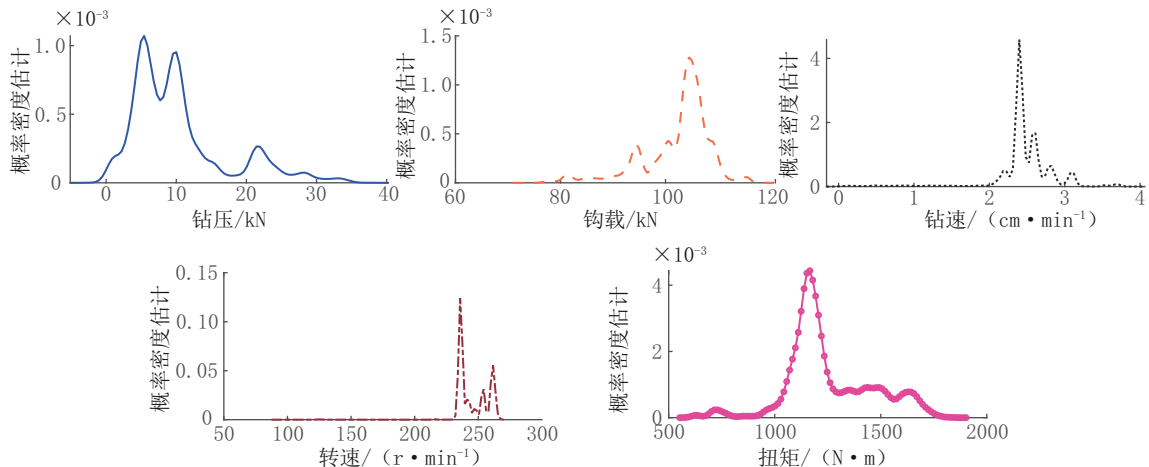


图 4 辽宁丹东 3000 m 科学钻探工程部分处理后数据概率密度估计

Fig.4 Density estimation of pre-processed data of 3000m scientific drilling project in Dandong, Liaoning province

3.2 工程应用

当深部地质钻探钻进过程流式大数据分析与动态预处理系统在现场稳定运行一段时间,并征得领导专家和现场工程技术人员的同意后,我们自 2021 年 10 月 9 日起对系统进行了工程应用(图 2),并已

在辽宁丹东 3000 m 科学钻探工程现场稳定工作超半年时间。

将本文方法与限幅滤波结合滑动窗口的方法对比,得到如图 5 所示的部分工程应用结果。图 5 是某一时刻数据窗口中滤波效果的对比,可以分析得知

本文方法进一步去除了深部地质钻探钻进过程数据的噪声,提升了过程数据质量。特别是,本文方法

在对相对变化幅度较大的钻速和扭矩的处理方面,表现出明显优势。

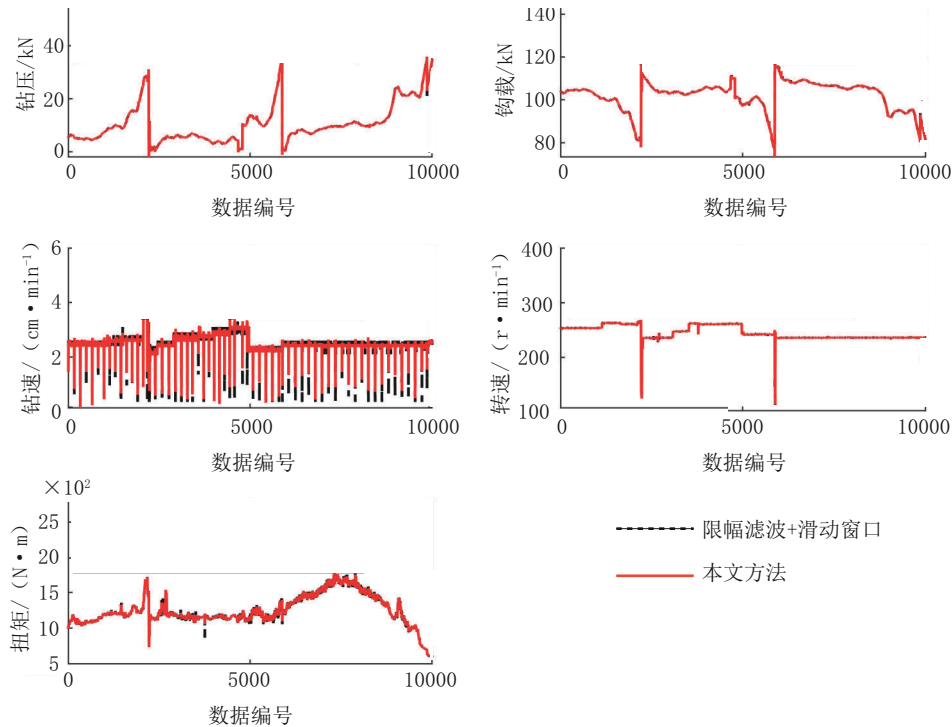


图5 部分工程应用结果

Fig.5 Some results of industrial application

钻速作为表征钻进过程效率的关键指标,由于受到地层条件、钻进工艺等不同因素影响,它在工程应用井段具体表现为在 $0\sim 4\text{ cm/min}$ 之间跳变,如果不能很好地对钻速数据进行分析与预处理,将不利于高精度预测钻速进而实现钻速优化^[7-8]。在图5中的钻速子图里,红色曲线是运用本文所提方法的工程应用结果,可以看出所提方法有效地去除了钻速数据中的尖峰与毛刺。此外,扭矩是钻进过程的另一个核心参数,主要反映过程的安全性。它在工程应用井段具体表现为在 $500\sim 2000\text{ N}\cdot\text{m}$ 之间变化。通过图5扭矩子图的红色曲线,可以看出本文所提方法能够很好地抑制扭矩数据的异常突变。工程应用结果表明,本文所提方法有效提升了深部地质钻探钻进过程数据的价值密度。

4 结论

针对深部地质钻探钻进过程地层条件复杂、数据价值密度低,传统数据分析与预处理方法难以在

保证噪声滤波效果的同时兼顾实时计算速度的问题。本文提出了一种深部地质钻探钻进过程流式大数据分析与动态预处理方法,并成功应用于辽宁丹东3000 m科学钻探工程。本文方法能够在线处理深部地质钻探钻进过程流式大数据,去除数据中的尖峰与毛刺。方法的工程适用性和有效性在仿真实验和工程应用中都得到了验证。

将来,深部地质钻探钻进过程流式大数据分析与动态预处理方法会在工程实践中不断更新与完善。

参考文献(References):

- [1] 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.
GAN Chao. Intelligent modeling of formation drillability field and drilling rate of penetration optimization in complex conditions [D]. Wuhan: China University of Geosciences, 2019.
- [2] 范海鹏, 吴敏, 曹卫华, 等. 基于钻进状态监测的智能工况识别

- [J].探矿工程(岩土钻掘工程),2020,47(4):106-113.
FAN Haipeng, WU Min, CAO Weihua, et al. Intelligent drilling mode identification based on drilling state monitoring while drilling[J]. Exploration Engineering (Rock & Soil Drilling and Tunneling), 2020,47(4):106-113.
- [3] 张正,赖旭芝,陆承达,等.基于贝叶斯网络的钻进过程井漏井涌事故预警[J].探矿工程(岩土钻掘工程),2020,47(4):114-121,144.
ZHANG Zheng, LAI Xuzhi, LU Chengda, et al. Lost circulation and kick accidents warning based on Bayesian network for the drilling process[J]. Exploration Engineering (Rock & Soil Drilling and Tunneling), 2020,47(4):114-121.
- [4] 张正,朱恒银.深部钻探关键设备选择原则及配置优化[J].探矿工程(岩土钻掘工程),2017,44(9):17-20.
ZHANG Zheng, ZHU Hengyin. Selection principles and configuration optimization of the key equipments in deep drilling[J]. Exploration Engineering (Rock & Soil Drilling and Tunneling), 2017,44(9):17-20.
- [5] 陈师逊,翟育峰,王鲁朝,等.西藏罗布莎科学钻探施工对深部钻探技术的启示[J].探矿工程(岩土钻掘工程),2012,39(11):1-3,9.
CHEN Shixun, ZHAI Yufeng, WANG Luzhao, et al. Enlightenment to deep drilling technology from scientific drilling in Luobusha of Tibet[J]. Exploration Engineering (Rock & Soil Drilling and Tunneling), 2012,39(11):1-3,9.
- [6] GAN Chao, CAO Weihua, LIU Kangzhi, et al. A novel dynamic model for the online prediction of rate of penetration and its industrial application to a drilling process[J]. Journal of Process Control, 2022,109:83-92.
- [7] GAN Chao, CAO Weihua, LIU Kangzhi, et al. A new hybrid bat algorithm and its application to the ROP optimization in drilling processes[J]. IEEE Transactions on Industrial Informatics, 2020,16(12):7338-7348.
- [8] GAN Chao, CAO Weihua, WU Min, 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.
- [9] GAN Chao, CAO Weihua, WU Min, et al. Two-level intelligent modeling method for the rate of penetration in complex geological drilling process[J]. Applied Soft Computing, 2019, 80: 592-602.
- [10] ASHRAFI Seyed Babak, ANEMANGELY Mohammad, SABAH Mohammad, et al. Application of hybrid artificial neural networks for predicting rate of penetration (ROP): A case study from Marun Oil Field[J]. Journal of Petroleum Science and Engineering, 2019,175:604-623.
- [11] DIAZ Melvin B, KIM Kwang Yeom, SHIN Hyu Soung, et al. Predicting rate of penetration during drilling of deep geothermal well in Korea using artificial neural networks and real-time data collection[J]. Journal of Natural Gas Science and Engineering, 2019,67:225-232.
- [12] 熊虎林,李谦.基于地层成分和钻进参数的钻速预测模型[J].探矿工程(岩土钻掘工程),2018,45(10):195-201.
XIONG Hulin, LI Qian. ROP prediction model based on formation composition and drilling parameters[J]. Exploration Engineering (Rock & Soil Drilling and Tunneling), 2018, 45(10):195-201.
- [13] 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.
LI Qian, CAO Yanwei, ZHU Haiyan. Discussion on the lower limit of data validity for ROP prediction based on artificial intelligence[J]. Drilling Engineering, 2021,48(3):21-30.
- [14] DIAZ Melvin B, KIM Kwang Yeom, SHIN Hyu Soung. On-line prediction model for rate of penetration (ROP) with cumulating filed data in real time [C]//YSRM & NDRMGE, Jeyu, Korea, May 10-13, 2017.
- [15] SAVITZKY Abraham, Golay M J E. Smoothing and differentiation of data by simplified least squares procedures[J]. Analytical Chemistry, 1964,36(8):1627-1639.

(编辑 李艺)