

基于机器学习算法的井漏预测与诊断理论模型研究

张欣, 田英英*, 韩泽龙, 赵明, 蒋亚峰

(中国地质科学院勘探技术研究所, 河北廊坊 065000)

摘要:在钻井作业中发生的井漏事故具有突发性、难治理等特点,导致现场堵漏工程难度大、成功率低、成本高等。为了“快、准、狠”地解决该类难题,必须准确预测漏失并提出堵漏措施。本文利用梯度提升决策树算法(GBDT)进行了井漏预测模型研究,通过分析若干机器学习算法,论证了GBDT作为井漏钻前预测模型的可行性;利用案例推理技术建立井漏案例诊断库,使用欧几里得算法和LB_Keogh算法进行井漏案例的检索,提出相应的井漏处理措施,通过实例验证了相似性计算的可靠性。

关键词:井漏;井漏预测;梯度提升决策树算法(GBDT);案例推理;井漏诊断

中图分类号:P634 **文献标识码:**A **文章编号:**2096-9686(2022)02-0058-09

Research on lost circulation prediction and diagnosis theoretical model based on machine learning algorithm

ZHANG Xin, TIAN Yingying*, HAN Zelong, ZHAO Ming, JIANG Yafeng

(Institute of Exploration Techniques, CAGS, Langfang Hebei 065000, China)

Abstract: The lost circulation incident in drilling operation has the characteristics of sudden occurrence and difficulty in treatment, which leads to low success rate of plugging and high cost. In order to solve this kind of problem quickly, accurately and efficiently, we must accurately predict lost circulation and put forward plugging measures. In this paper, the gradient boosted decision tree algorithm (GBDT) is used to study the loss circulation prediction model with the feasibility of GBDT as a lost circulation prediction model verified by analyzing some machine learning algorithms. The lost circulation case decision base is established by using case-based reasoning technology, and the Euclidean algorithm and LB_Keogh algorithm are used to search lost circulation cases to put forward corresponding lost circulation treatment measures. The reliability of similarity calculation has been verified through field cases.

Key words: lost circulation; lost circulation prediction; gradient boosted decision tree (GBDT); case-based reasoning (CBR); lost circulation diagnosis

0 引言

井漏是指相当一部分或者全部钻井液从井筒损失进入到具有高渗透率的地层中,比如洞穴、裂缝或者松散地层。国外较为成熟的井漏预测与诊

断的技术有精细控压钻井技术、ALS-K井涌井漏快速探测系统、DrillEdge软件^[1-3],分别通过微量溢流测量、钻井液进出口流量测量和案例推理技术来实现。国内相关的井漏预测与诊断技术有综合录井

收稿日期:2021-11-12; 修回日期:2022-01-18 DOI:10.12143/j.ztgc.2022.02.008

基金项目:中国地质调查局地质调查项目“×××先导试验区试采实施(中国地质科学院勘探技术研究所)”(编号:DD20190227)、“×××先导试验区资源评价与试采(中国地质科学院勘探技术研究所)”(编号:DD20211349);南方海洋科学与工程广东省实验室(广州)人才团队引进重大专项“天然气水合物钻采船单筒双井及大洋钻探技术研究”(编号:GML2019ZD0504)

第一作者:张欣,男,汉族,1993年生,助理工程师,石油与天然气工程专业,硕士,从事海洋钻探新技术、新设备、新工艺的研发工作,河北省廊坊市广阳区金光道77号,zxincompany@163.com。

通信作者:田英英,女,汉族,1993年生,工程师,油气田开发专业,硕士,从事海洋钻探新技术、新设备、新工艺的研发工作,河北省廊坊市广阳区金光道77号,tianyingying0704@163.com。

引用格式:张欣,田英英,韩泽龙,等.基于机器学习算法的井漏预测与诊断理论模型研究[J].钻探工程,2022,49(2):58-66.

ZHANG Xin, TIAN Yingying, HAN Zelong, et al. Research on lost circulation prediction and diagnosis theoretical model based on machine learning algorithm[J]. Drilling Engineering, 2022, 49(2): 58-66.

技术方法、分离器液面检测方法和井下微流量检测方法等^[4-7],这些方法对井漏的处理主要是专家结合井漏的现场情况提出解决方案,过于依靠现场专家的经验,导致无法实现钻前诊断,忽略了其他堵漏决策方案^[8]。而应用人工智能方法预测与诊断井漏等复杂井下情况成为了钻井工程领域的发展趋势,学者多采用神经网络^[9-12]、贝叶斯网络^[13-14]和支持向量机(SVM)^[15-16]等机器学习算法建立井漏预测与诊断模型。每种机器学习算法都有其优缺点和适用范围,其中神经网络适用于复杂非线性数据建模^[10],但是该算法容易陷入局部最优解问题^[17],不充分的数据难以维持算法计算;贝叶斯网络在基于钻井参数的不确定性预测中发挥了重要作用,但还需要对时间尺度、趋势判断界限和阈值优化进行深入研究,提高预测准确性^[14];支持向量机适用于小样本、非线性及高纬度的数据处理^[15],不适用于数据量庞大的井漏诊断与识别问题。

梯度提升决策树算法(GBDT, Gradient Boosted Decision Tree),是一种集成学习算法,该算法通过建立一个决策树,以适应它前面的树的残差,就像其他提升算法一样,GBDT算法是一种迭代过程,可以将几个弱的“学习器”组合成一个强的“学习器”,提高预测的准确性。同时,其对原始数据处理要求低,能够有效应对复杂的工程问题^[18-20]。目前,该算法已应用于医学^[21-22]、城市交通^[23-25]、电力^[26-27]、石油^[17,28]等领域。

综上所述,本文旨在研究GBDT在井漏预测与诊断中的应用,通过井漏机理分析,选取地层参数和钻井工程参数,分析GBDT作为井漏钻前诊断模型的可行性;利用案例推理技术建立井漏案例诊断库,使用欧几里得算法和LB_Keogh算法进行井漏案例的检索,提出相应的井漏处理措施,通过实例计算出相似性计算的可靠性。

1 基于GBDT的井漏预测模型分析

1.1 井漏特征参数分析

井漏的发生受到地质、工程等多种原因影响,也会通过地面监测仪器参数表现出来,因此,在使用GBDT算法对井漏进行预测时,应将这些原因考虑进去,提高预测精度。

1.1.1 井深

在同一区块的油气井发生的井漏一般会集中在一个深度范围。不同的井深对应的漏失工况、钻井液性能和所钻遇的地层岩性等因素亦不相同。

1.1.2 地层压力

当地层中存在天然漏失通道或钻井液密度过大时,地层孔隙压力 $<$ 井内压力,钻井液会从井内流入地层,诱发井漏事故^[14]。

1.1.3 破裂压力

在钻井液漏失机理中,当漏失压力 \geq 破裂压力时,地层中的裂缝会被压开或者由于压力过大地层出现新裂缝,形成漏失通道,造成钻井液的漏失事故。

1.1.4 坍塌压力

在钻井过程中,有一些地层会出现“宁漏勿塌”的现象,这时的漏失压力和坍塌压力近乎相等,井底压力稍微大一些地层就会出现漏失事故,井底压力稍微小一些地层就会出现坍塌事故。

1.1.5 地层岩性、孔隙度和渗透率

在钻井设计时,考虑地层分层及其特征并进行故障提示是非常有必要的,比如碳酸盐岩地层裂缝、孔洞发育;粗砂、砾石地层渗透率高;页岩具有弱平面的特点,在钻井作业中容易被压裂形成漏失通道。所以地层岩性、孔隙度和渗透率需要参与到井漏诊断的计算中。

1.1.6 钻井液密度、粘度、排量

钻井液密度对井漏的影响主要是由于密度升高,增大了液柱压力,从而增大了井内与地层之间的压差。粘度较大时,对防止井漏或减轻井漏是有益的。但是在实际钻井过程中,过大的粘度将会使得循环压力过高,从而使薄弱地层产生人为裂缝,引起更大的漏失^[29]。而过大的排量也会对井眼稳定性产生影响,因此在易漏地层,应适当调整钻井液的密度、粘度和排量值^[30]。

1.1.7 钻井液固相颗粒

漏失通道的尺寸应大于钻井液固相颗粒粒径,否则固相颗粒过大会在井壁形成泥饼,而固相颗粒过小,则会导致井筒内钻井液的漏失^[31],较差的固相颗粒级配也会造成泥饼质量差,导致漏失。

1.1.8 泵压

泵压升高或降低一定程度上反映了井下是否发生了复杂事故,亦可作为漏失是否发生的判定条件之一。

1.1.9 钻井液池体积、进出口流量差、立压、大钩负载和钻速

在发生井漏时,钻井液池体积、进出口流量差、立压、大钩负载和钻速也会发生变化,根据地下情况的不同,变化规律也不尽相同^[32]。单一的地面参数变化不能完全表示井漏的发生,需要多参数协同表示,才能够提高预测和诊断的精度。

1.2 数据归一化处理

GBDT使用回归树模型作为弱学习器,通过多次迭代,每次迭代产生一个弱学习器,每个学习器在上一轮学习器的梯度基础上进行训练,即让损失函数沿着梯度方向下降,使当前迭代轮次的损失函数最小。当迭代完成后,就可以得到最终的强学习器,实现预测的目的。在构建井漏预测模型的过程中,需要对模型进行数据训练来提高模型的精度。

在进行数据训练之前首先对数据进行预处理,剔除由人为因素或外界因素干扰造成的异常数据,对含有缺失值的数据进行忽略,并将数值进行归一化,从而避免特征量纲差异对预测结果造成的影响。本研究选用最大最小化方法将特征归一到 $[0, 1]$ 之间:

$$X_{\text{new}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (1)$$

式中: X_{new} ——归一化后特征; X_{\max} 、 X_{\min} ——分别为特征的最大值和最小值。

需要进行处理的数据分为2类,一类是数值型数据,一类是文字型数据。数值型数据需要统一格式再进行处理,采用能够代表井漏特征的参数进行表示,其中使用马氏漏斗粘度来表示钻井液粘度,其中马氏漏斗粘度通过流出一定体积的钻井液所经历的时间来衡量粘度大小;用颗粒级配中的中值粒径 D_{50} 来表示固相颗粒的大小程度,其中 D_{50} 表示颗粒累积分布百分数达到50%时所对应的粒径值;文字型数据需要进行量化,使文字数据转化为数值数据。在井漏特征参数中只有地层岩性是文字型数据,将其进行量化的处理方法是按照0~100的顺序

逻辑对特征属性的所有取值赋予一个对应的正整数,方便后续计算。

1.3 预测模型建立

本文将井漏参数分别进行归一化处理。将训练数据代入到GBDT模型中进行训练,建立井漏预测模型。

(1)输入:

数据集 $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$,

损失函数 $L[y, f(x)]$ 。

输出:梯度提升模型 $F(x)$ 。

初始化:

$$f_0(x) = \arg \min_c \sum_{i=0}^n L(y_i, c) \quad (2)$$

式中: c ——使损失函数达到极小值的常数。

(2)假设有 $m = 1, 2, \dots, M$ 次迭代,对样本 $i = 1, 2, \dots, n$, 计算损失函数的负梯度——残差的估计值为:

$$r_{mi} = - \left[\frac{\partial L[y_i, f(x_i)]}{\partial f(x_i)} \right] f(x) = f_{m-1}(x) \quad (3)$$

(3)利用数据集 (x_i, r_{mi}) 拟合下一轮基础模型,得到对应的叶子结点 $R_{mj}, j = 1, 2, \dots, J$, 计算每个叶子结点 R_{mj} 的最佳拟合值,使损失函数最小:

$$c_{mj} = \arg \min_c L[y_i, f_{m-1}(x_i) + c] \quad (4)$$

(4)计算所有叶子节点对应的 c_{mj} 之和得到第 m 轮的基础模型:

$$f_m(x) = \sum_{j=1}^J c_{mj} I(x_i \in R_{mj}) \quad (5)$$

结合前 $m - 1$ 轮的基础模型得到最终强学习器:

$$f_M(x) = f_{M-1}(x) + f_m(x) = \sum_{m=1}^M \sum_{j=1}^J c_{mj} I(x_i \in R_{mj}) \quad (6)$$

2 基于案例推理的井漏诊断模型建立

案例推理既是一种人工智能技术,又是利用相似案例的相关信息解决当前问题的类比推理方法^[33]。该方法的工作流程为案例表示、案例检索和案例学习。案例的表示需要将所有的案例使用统一的方式描述出来,建立案例数据库。数据库中的案例越多,案例越具有代表性,数据库越成熟。本研究

将收集到的数据采用三元组的形式进行描述,分别是:(1)井漏案例基本信息,此部分包括案例编号、油田名称、区块名称、井口坐标、井号、事故类型、漏失层位、漏失岩性、漏失位置、钻井工况、漏速和堵漏次数等;(2)井漏特征属性信息,此部分内容包括案例编号和特征参数;(3)井漏事故结论信息,此部分内容包括案例发生经过、处理方案和处理效果。

案例检索的相似度计算是诊断模型的核心^[34]。本研究针对井漏特征参数中的非时间序列数据和时间序列数据分别采用欧几里得数学模型和 LB_Keogh 数学模型进行相似度计算。当输入新案例井的信息时,诊断模型将搜索与新案例的信息匹配度最高的案例,将其调用出来,技术人员可参考调用出来的案例中防漏堵漏的信息,对新案例井进行处理。

诊断模型需要不断加入新案例进行自主学习^[35],也需要删减案例,提高案例质量。若诊断模型未检索出与新案例井相似的案例,则将新案例井的信息存储进去;若某一案例被检索出来的频率很高,则将此案例作为经典案例;若部分案例经常一起被检索出来,则删除其余案例,保留一个代表性案例。

2.1 井漏参数权重计算

在井漏诊断中,本研究利用层析分析法中的标度法,对井漏参数构建判断矩阵,其中各标度值见表 1。层次分析法通过对影响因素进行赋权,每一层影响因素的权重都会影响到结果,可以将复杂问题变成层次分明的阶梯结构,是一种解决多目标复杂问题的决策分析方法。

表 1 特征参数标度值

Table 1 Value scale of characteristic parameters

参 数	破裂 压力	地层 压力	地层 岩性	钻井 液密 度	钻井 液粘 度	排量	泵压	钻井 液池 体积	出入 口流 量差	大钩 负荷	立压	钻速	井 深	孔隙 度	渗透 率	D_{50}	坍塌 压力
破裂压力	1	1	1	2	2	2	2	5	5	5	5	5	7	7	7	7	7
地层压力	1	1	1	2	2	2	2	5	5	5	5	5	7	7	7	7	7
地层岩性	1	1	1	2	2	2	2	5	5	5	5	5	7	7	7	7	7
钻井液密度	1/2	1/2	1/2	1	1	1	1	2	2	2	2	2	4	4	4	4	4
钻井液粘度	1/2	1/2	1/2	1	1	1	1	2	2	2	2	2	4	4	4	4	4
排量	1/2	1/2	1/2	1	1	1	1	2	2	2	2	2	4	4	4	4	4
泵压	1/2	1/2	1/2	1	1	1	1	2	2	2	2	2	4	4	4	4	4
钻井液池体积	1/5	1/5	1/5	1/2	1/2	1/2	1/2	1	1	1	1	1	2	2	2	2	2
出入口流量差	1/5	1/5	1/5	1/2	1/2	1/2	1/2	1	1	1	1	1	2	2	2	2	2
大钩负荷	1/5	1/5	1/5	1/2	1/2	1/2	1/2	1	1	1	1	1	2	2	2	2	2
立压	1/5	1/5	1/5	1/2	1/2	1/2	1/2	1	1	1	1	1	2	2	2	2	2
钻速	1/5	1/5	1/5	1/2	1/2	1/2	1/2	1	1	1	1	1	2	2	2	2	2
井深	1/7	1/7	1/7	1/4	1/4	1/4	1/4	1/2	1/2	1/2	1/2	1/2	1	1	1	1	1
孔隙度	1/7	1/7	1/7	1/4	1/4	1/4	1/4	1/2	1/2	1/2	1/2	1/2	1	1	1	1	1
渗透率	1/7	1/7	1/7	1/4	1/4	1/4	1/4	1/2	1/2	1/2	1/2	1/2	1	1	1	1	1
D_{50}	1/7	1/7	1/7	1/4	1/4	1/4	1/4	1/2	1/2	1/2	1/2	1/2	1	1	1	1	1
坍塌压力	1/7	1/7	1/7	1/4	1/4	1/4	1/4	1/2	1/2	1/2	1/2	1/2	1	1	1	1	1

建立判断矩阵:

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \quad (7)$$

式中: a_{mn} ——判断矩阵中第 m 个特征参数相对于第 n 个特征参数的重要判断标度值。

由表 1 的特征参数判断标度值得判断矩阵为:

$$A = \begin{bmatrix} 1 & 1 & 1 & 2 & 2 & 2 & 2 & 5 & 5 & 5 & 5 & 5 & 7 & 7 & 7 & 7 & 7 \\ 1 & 1 & 1 & 2 & 2 & 2 & 2 & 5 & 5 & 5 & 5 & 5 & 7 & 7 & 7 & 7 & 7 \\ 1 & 1 & 1 & 2 & 2 & 2 & 2 & 5 & 5 & 5 & 5 & 5 & 7 & 7 & 7 & 7 & 7 \\ 1/2 & 1/2 & 1/2 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 2 & 2 & 4 & 4 & 4 & 4 & 4 \\ 1/2 & 1/2 & 1/2 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 2 & 2 & 4 & 4 & 4 & 4 & 4 \\ 1/2 & 1/2 & 1/2 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 2 & 2 & 4 & 4 & 4 & 4 & 4 \\ 1/2 & 1/2 & 1/2 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 2 & 2 & 4 & 4 & 4 & 4 & 4 \\ 1/5 & 1/5 & 1/5 & 1/2 & 1/2 & 1/2 & 1/2 & 1 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 2 & 2 \\ 1/2 & 1/5 & 1/5 & 1/5 & 1/2 & 1/2 & 1/2 & 1/2 & 1 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 2 & 2 \\ 1/5 & 1/5 & 1/5 & 1/2 & 1/2 & 1/2 & 1/2 & 1 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 2 & 2 & 2 \\ 1/5 & 1/5 & 1/5 & 1/2 & 1/2 & 1/2 & 1/2 & 1 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 2 & 2 & 2 \\ 1/5 & 1/5 & 1/5 & 1/2 & 1/2 & 1/2 & 1/2 & 1 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 2 & 2 & 2 \\ 1/7 & 1/7 & 1/7 & 1/4 & 1/4 & 1/4 & 1/4 & 1/2 & 1/2 & 1/2 & 1/2 & 1/2 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1/7 & 1/7 & 1/7 & 1/4 & 1/4 & 1/4 & 1/4 & 1/2 & 1/2 & 1/2 & 1/2 & 1/2 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1/7 & 1/7 & 1/7 & 1/4 & 1/4 & 1/4 & 1/4 & 1/2 & 1/2 & 1/2 & 1/2 & 1/2 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1/7 & 1/7 & 1/7 & 1/4 & 1/4 & 1/4 & 1/4 & 1/2 & 1/2 & 1/2 & 1/2 & 1/2 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1/7 & 1/7 & 1/7 & 1/4 & 1/4 & 1/4 & 1/4 & 1/2 & 1/2 & 1/2 & 1/2 & 1/2 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \quad (8)$$

为了确保最终计算结果的准确性,需要进一步构建最优传递矩阵,如果判断矩阵满足下式,则判断矩阵为一致性矩阵:

$$b_{ij} = \frac{1}{n} \sum_{k=1}^n (a_{ik} + a_{kj}) \quad (9)$$

构建一致性矩阵:

$$c_{ij} = e^{b_{ij}} \quad (10)$$

通过计算一致性矩阵 C 每行元素的乘积:

$$C_i = \prod_{j=1}^n c_{ij} \quad (11)$$

最终通过计算每个乘积的 n 次方根,并对结果做归一化处理得到每个特征参数的权重值,最后计算结果见图1。

$$u_i = \sqrt[n]{C_i} \quad (12)$$

$$\omega_i = \frac{u_i}{\sum_{i=1}^n u_i} \quad (13)$$

式中: n ——特征参数的个数。

为确保各特征参数之间重要程度之间的协调性,需要对举证做一致性检验,即可靠性论证,以免出现参数 A 比参数 B 重要,参数 B 比参数 C 重要,但参数 C 比参数 A 重要的矛盾。一致性指标见式(14)。

$$C_R = \frac{C_I}{R_I} \quad (14)$$

式中: C_R ——致性指标; C_I ——致性系数; R_I ——平均随机一致性指标。

其中平均随机一致性指标与矩阵阶数 n 有关,具体数值见表2。

矩阵中最大特征根 $\lambda_{\max} = 18.6$, $C_I = 0.097$, $R_I = 1.61$, $C_R = 0.06 < 0.1$, 满足矩阵一致性要求。

2.2 非时间序列的相似度计算

本研究选用欧几里得距离法来计算属于非时间序列的井漏特征参数。下面结合2个案例具体研究如何使用欧几里得法计算井漏案例在特征参数上的相似度^[36]。

设 A 井漏案例中的所有非时间序列参数的取值组成的集合为:

$$A = \{A_1, A_2, A_3, \dots, A_n\} \quad (15)$$

设 B 井漏案例中的所有非时间序列参数的取值组成的集合为:

$$B = \{B_1, B_2, B_3, \dots, B_n\} \quad (16)$$

式中: A_j, B_j ——2个井漏案例的第 j 个特征属性的取值; n ——井漏案例的非时间序列参数的总个数。

计算相似度之前,需要将钻井作业中采集到的数据进行预处理,消除数据量纲或数据数值过大对

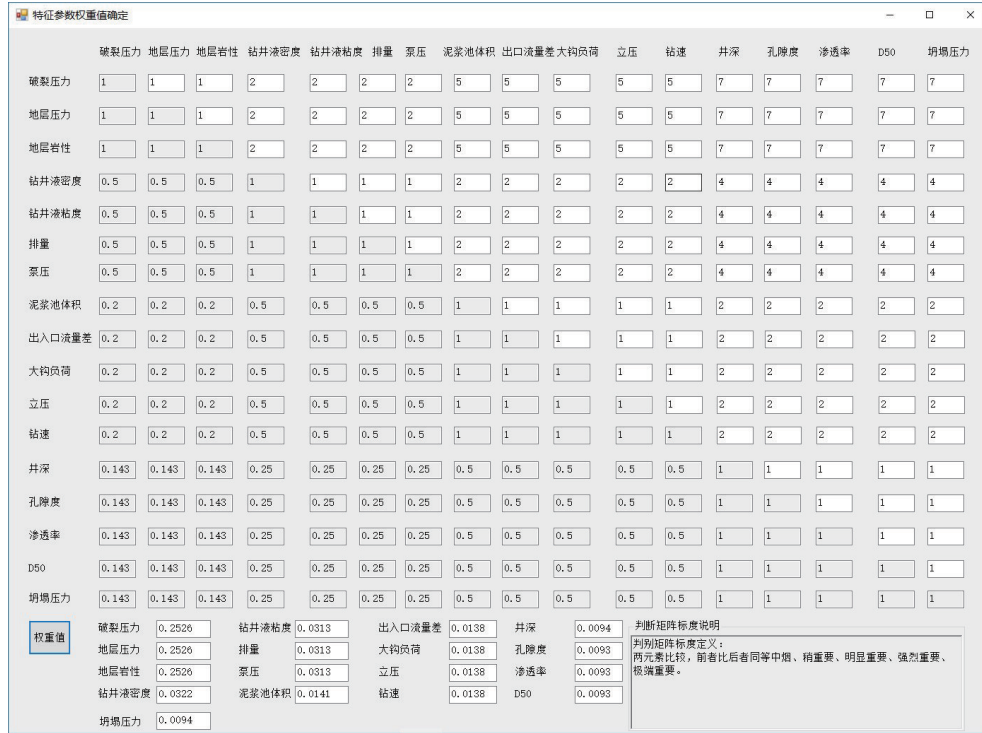


图 1 层次分析法权重求取结果

Fig.1 AHP weighting results

表 2 平均随机一致性指标 R_1 数值

Table 2 Values of mean random consistency index R_1

阶数	1	2	3	4	5	6	7	...	17
R_1	0	0	0.58	0.89	1.12	1.24	1.35	...	1.61

计算造成的不良影响。以地层压力为例,本文利用归一化算法和欧几里得法计算 2 个井漏案例在钻井液密度上的相似度,见式(17):

$$sim(A_2, B_2) = 1 - \left| \frac{A_2 - B_2}{max - min} \right| \quad (17)$$

式中: max ——地层压力特征属性的最大取值; min ——地层压力特征属性的最小取值

在计算 2 个案例某一特征的相似度之后,要计算 2 个案例的所有特征的相似度,并且根据层次分析法求取的每个特征参数的权重进行求和,计算出 2 个案例的整体相似度,见式(18):

$$S(a) = \omega_1 sim(A_1, B_1) + \omega_2 sim(A_2, B_2) + \dots + \omega_n sim(A_n, B_n) \quad (18)$$

式中: $S(a)$ ——井漏案例在所有特征参数上的相似度; ω_i ——第 i 个特征参数所具有的权重; $sim(A_i, B_i)$ ——井漏案例在第 i 个特征参数上的相似度。

2.3 时间序列的相似度计算

本研究采用动态弯曲算法中的 LB_Keogh 算法^[37]进行时间序列的相似度计算,在相似度计算之前给时间序列 Q 设施加一个边界约束,分别为 U 和 L 。使得时间序列 Q 被 U 和 L 包裹在其中,由此设置时间序列 U 和 L 满足下式:

$$U_i = \max(q_{i-r}, q_{i+r}) \quad (19)$$

$$L_i = \min(q_{i-r}, q_{i+r}) \quad (20)$$

式中: $r \geq 1$ 。

对于任意的 U, L 和 Q 都满足下式:

$$\forall U_i \geq q_i \geq L_i \quad (21)$$

定义了 U 和 L 之后,使用它们来定义一个下界度量:

$$LB_{Keogh}(Q, C) = \sqrt{\sum_{i=1}^n \begin{cases} (c_i - U_i)^2 & \text{if } c_i > U_i \\ (c_i - L_i)^2 & \text{if } c_i < L_i \\ 0 & \text{otherwise} \end{cases}} \quad (22)$$

则时间序列 Q 和 C 之间的相似度为:

$$SIM(Q, C) = 1 - \frac{1}{L} LB_{Keogh}(Q, C) \quad (23)$$

2 个案例之间的相似度为所有井漏特征参数乘以权重再求和,见式(24):

$$S(b) = \sum_{i=1}^n \omega_i SIM(Q, C) \quad (24)$$

以同在珍珠冲层位发生小漏的川6B1和川6B2井为例,以钻井液池总体积特征参数为例说明此时间序列检索模型的应用。川6B1和川6B2井发生井漏时钻井液池总体积参数数据及其归一化计算见表3。

表3 川6B1与川6B2井井漏特征参数
Table 3 Characteristic parameters of lost circulation in Chuan-6B1 and Chuan-6B2 wells

井号	井深/m	钻井液池总体积/m ³	归一化数据
川6B1井	2845	63.7	0.5967
	2849	63.2	0.5911
	2851	62.9	0.5878
	2854	61.4	0.5711
	2856	61.3	0.5700
	2865	61.1	0.5678
	2870	60.0	0.5556
	2871	58.3	0.5367
	2874	56.4	0.5156
川6B2井	2953	74.6	0.7178
	2955	68.2	0.6467
	2956	67.8	0.6422
	2959	61.0	0.5667
	2962	57.6	0.5289
	2965	54.6	0.4956
	2966	53.8	0.4867
	2968	52.2	0.4689
2971	48.8	0.4311	

定义川6B1井的归一化数据为时间序列 Q ,川6B2井的归一化数据为时间序列 C 。则根据上述公式得:

$$U_i = (0.5911, 0.5967, 0.5967, 0.5967, 0.5967, 0.5967, 0.5967, 0.5967, 0.5967) \quad (25)$$

$$L_i = (0.5911, 0.5711, 0.5678, 0.5367, 0.5156, 0.5156, 0.5156, 0.5156, 0.5156) \quad (26)$$

$$LB_{Keogh(Q,C)} = 0.21 \quad (27)$$

$$SIM(Q,C) = (1 - 0.21/9) \times 100\% = 97.7\% \quad (28)$$

由于发生井漏的2口井漏案例为同一区块且位于同一层位的2个案例,因此在钻井液池总体积参数上的相似度很大,其他参数的相似度计算与上文计算步骤类似,再此不一一赘述。通过此实例说明

LB_Keogh算法在对时间序列相似度求取上的可靠性。

2.4 整体相似度的计算

对非时间序列和时间序列分别进行相似度计算后,设置非时间序列的相似度在案例整体相似度中所占权重为0.4,时间序列的相似度在案例整体相似度中所占权重为0.6,得到整体权重的相似度计算公式:

$$S(a,b) = \omega_a S(a) + \omega_b S(b) = 0.4S(a) + 0.6S(b) \quad (29)$$

式中: $S(a)$ ——非时间序列的相似度; $S(b)$ ——时间序列的相似度。

通过案例相似度的计算,可以根据相似度进行井间的匹配,对相似度设定某一阈值,只有相似度大于这一阈值的案例井才能被调出,其防漏堵漏措施能够应用于新案例井。

3 结论

(1)本文通过分析,论证了梯度提升决策树算法在预测模型上的可行性,采用案例推理方法建立了井漏诊断模型,在预测出新井某一深度的漏失情况后,可使用诊断模型计算相似度,快速检索与新井最相似的案例井,现场工作人员可参考案例井的防漏堵漏措施,提前准备处理方案。

(2)在井漏诊断模型中,使用层次分析法确定井漏特征参数权重,使用欧几里得距离法和LB_Keogh动态弯曲算法对案例相似度进行计算,并验证了相似度算法的可靠性。

参考文献(References):

- [1] Kulikov S, Veliev G, Bakhtin A, et al. "Secure Drilling" services for safe and effective drilling[C]//SPE Arctic and Extreme Environments Technical Conference and Exhibition, 2013.
- [2] Raja H, Sørmo F, Vinther M L. Case-based reasoning: Predicting real-time drilling problems and improving drilling performance[C]//SPE Middle East Oil and Gas Show and Conference, 2011.
- [3] 贾爱贵. ALS-K井涌井漏快速探测系统[J]. 录井技术, 1998(2): 48-52.
JIA Aigui. ALS-K kick and lost circulation fast detection system [J]. Mud Logging Engineering, 1998(2): 48-52.
- [4] 姜永, 白生平, 赵书滨, 等. 综合录井技术在钻井工程上的应用[J]. 石油钻采工艺, 2009, 31(S2): 79-82.
JIANG Yong, BAI Shengping, ZHAO Shubin, et al. Applica-

- tion of compound logging technology in drilling engineering[J]. Oil Drilling & Production Technology, 2009,31(S2):79-82.
- [5] 李治伟.塔里木井控装备配套技术研究[D].成都:西南石油大学,2011.
LI Zhiwei. Study on matching technology of well control equipment in Tarim[D]. Chengdu: Southwest Petroleum University, 2011.
- [6] 苟开海,胡泽,葛亮.基于DSP的井底微流量测量系统设计[J].计量与测试技术,2011,38(9):17-19.
GOU Kaihai, HU Ze, GE Liang. Design of detective instrument for tiny flow of well based on DSP[J]. Metrology & Measurement Technique, 2011,38(9):17-19.
- [7] 屈俊波,陈平,马天寿,等.精确监测井底溢流的井下微流量装置设计与试验[J].石油钻探技术,2012,40(5):106-110.
QU Junbo, CHEN Ping, MA Tianshou, et al. Design and test of down-hole micro-flow device for monitoring overflow[J]. Petroleum Drilling Techniques, 2012,40(5):106-110.
- [8] 王海彪.井漏智能识别及处理决策研究[D].成都:西南石油大学,2017.
WANG Haibiao. Study of intelligent recognition and management decisions of lost circulation[D]. Chengdu: Southwest Petroleum University, 2017.
- [9] 李雪松,张骁,管震,等.基于图像识别技术的钻井井漏溢流智能报警系统开发[J].世界石油工业,2021,28(1):48-54.
LI Xuesong, ZHANG Xiao, GUAN Zhen, et al. Development of the drilling mud loss and overflow intelligent alarm system based on the image recognition technology[J]. World Petroleum Industry, 2021,28(1):48-54.
- [10] 和鹏飞,刘晓宾,陈真,等.基于深度神经网络模型的钻井井漏预测研究[J].天津科技,2019,46(S1):21-23.
HE Pengfei, LIU Xiaobin, CHEN Zhen, et al. Research on prediction of lost circulation based on deep neural network model[J]. Tianjin Science & Technology, 2019,46(S1):21-23.
- [11] 谢平,蒋丽雯,赵尧,等.基于神经网络的井涌井漏实时预测方法研究[J].现代计算机(专业版),2018(11):23-28.
XIE Ping, JIANG Liwen, ZHAO Yao, et al. Research on real-time prediction and analysis based on neural network for well kick and lost circulation [J]. Modern Computer, 2018(11):23-28.
- [12] 李克智,袁本福.红河油田井漏风险实时识别研究与应用[J].钻采工艺,2013,36(4):20-22,134.
LI Kezhi, YUAN Benfu. Research and application on circulation loss risk real-time identification in Honghe Oilfield [J]. Drilling & Production Technology, 2013,36(4):20-22,134.
- [13] 徐哲,李建,王兵,等.基于贝叶斯网络的钻井井漏问题研究[J].石油天然气学报,2013,35(12):125-129,8-9.
XU Zhe, LI Jian, WANG Bing, et al. Research on well leakage in drilling based on Bayesian network[J]. Journal of Oil and Gas Technology, 2013,35(12):125-129,8-9.
- [14] 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):114-121,144.
- [15] 李晓桐,徐英卓,何坤鹏,等.钻井异常智能预警系统研究[J].福建电脑,2013,29(4):1-3.
LI Xiaotong, XU Yingzhuo, HE Kunpeng, et al. Research on intelligent early warning system of drilling anomaly[J]. Fujian Computer, 2013,29(4):1-3.
- [16] 刘彪,李宸晓,李双贵,等.基于支持向量回归的井漏预测[J].钻采工艺,2019,42(6):17-20,1-2.
LIU Biao, LI Chenxiao, LI Shuanggui, et al. Lost circulation prediction based on support vector regression [J]. Drilling & Production Technology, 2019,42(6):17-20,1-2.
- [17] 谷宇峰,张道勇,鲍志东,等.利用梯度提升决策树(GBDT)预测渗透率——以姬塬油田西部长4+5段致密砂岩储层为例[J].地球物理学进展,2021,36(2):585-594.
GU Yufeng, ZHANG Daoyong, BAO ZhiDong, et al. Permeability prediction using gradient boosting decision tree (GBDT): A case study of tight sandstone reservoirs of member of Chang 4+5 in western Jiyuan Oilfield[J]. Progress in Geophysics, 2021,36(2):585-594.
- [18] 韩启迪,张小桐,申维.基于梯度提升决策树(GBDT)算法的岩性识别技术[J].矿物岩石地球化学通报,2018,37(6):1173-1180.
HAN Qidi, ZHANG Xiaotong, SHEN Wei. Lithology identification technology based on gradient boosting decision tree (GBDT) algorithm [J]. Bulletin of Mineralogy Petrology and Geochemistry, 2018,37(6):1173-1180.
- [19] 石运良,罗宇,陈正科.基于GBDT算法的焊缝背面熔宽预测[J].热加工工艺,2021,50(17):110-114.
SHI Yunliang, LUO Yu, CHEN Zhengke. Prediction of weld backside molten pool width based on GBDT algorithm[J]. Hot Working Technology, 2021,50(17):110-114.
- [20] 贾南,何昌原,段海鹏.基于集成算法的森林火灾风险预警模型研究[J].武警学院学报,2021,37(12):5-9.
JIA Nan, HE Changyuan, DUAN Haipeng. Research on fire risk early warning model based on integrated algorithm [J]. Journal of Chinese People's Armed Police Force Academy, 2021,37(12):5-9.
- [21] 徐安.基于机器学习的慢性疾病预测关键技术研究[D].成都:电子科技大学,2019.
XU An. Research of chronic disease prediction technology based on machine learning[D]. Chengdu: School of Computer Science and Engineering, 2019.
- [22] 龚谊承,都承华,张艳娜,等.基于主成分和GBDT对血糖值的

- 预测[J]. 数学的实践与认识, 2019, 49(14): 116-122.
- GONG Yicheng, DU Chenghua, ZHANG Yanna, et al. Prediction of blood glucose based on principal component and GB-DT[J]. Mathematics in Practice and Theory, 2019, 49(14): 116-122.
- [23] 黄沼洋, 薛旺星, 蔡铭, 等. 基于梯度提升算法的道路交通噪声预测模型研究[J]. 环境科学与技术, 2020, 43(4): 46-53.
- HUANG Zhaofeng, XUE Wangxing, CAI Ming, et al. A traffic noise prediction model based on gradient boosting algorithm[J]. Environmental Science & Technology, 2020, 43(4): 46-53.
- [24] 翁剑成, 付宇, 林鹏飞, 等. 基于梯度推进决策树的日维度交通指数预测模型[J]. 交通运输系统工程与信息, 2019, 19(2): 80-85, 93.
- WENG Jiancheng, FU Yu, LIN Pengfei, et al. GBDT method based on prediction model of daily dimension traffic index[J]. Journal of Transportation Systems Engineering and Information Technology, 2019, 19(2): 80-85, 93.
- [25] 廖璐, 张亚东, 葛晓程, 等. 基于GBDT的列车晚点时长预测模型研究[J]. 铁道标准设计, 2021, 65(8): 149-154, 176.
- LIAO Lu, ZHANG Yadong, GE Xiaocheng, et al. Research on prediction model of train delay time based on GBDT[J]. Railway Standard Design, 2021, 65(8): 149-154, 176.
- [26] 陈宏, 邓芳明, 吴翔, 等. 基于梯度提升决策树的电力电子电路故障诊断[J]. 测控技术, 2017, 36(5): 9-12, 20.
- CHEN Hong, DENG Fangming, WU Xiang, et al. Fault diagnosis of power electronics circuit based on gradient boosting decision tree[J]. Measurement & Control Technology, 2017, 36(5): 9-12, 20.
- [27] 刘金硕, 刘必为, 张密, 等. 基于GBDT的电力计量设备故障预测[J]. 计算机科学, 2019, 46(S1): 392-396.
- LIU Jinshuo, LIU Biwei, ZHANG Mi, et al. Fault prediction of power metering equipment based on GBDT[J]. Computer Science, 2019, 46(S1): 392-396.
- [28] 苏兴华, 孙俊明, 高翔, 等. 基于GBDT算法的钻井机械钻速预测方法研究[J]. 计算机应用与软件, 2019, 36(12): 87-92.
- SU Xinghua, SUN Junming, GAO Xiang, et al. Prediction method of drilling rate of penetration based on GBDT algorithm[J]. Computer Applications and Software, 2019, 36(12): 87-92.
- [29] 杜青才. 准噶尔南缘复杂构造地质力学分析与井下复杂机理研究[D]. 成都: 西南石油学院, 2004.
- DU Qingcai. Geomechanical analysis of complex structures in the southern margin of Zhunger basin and study on underground complex mechanism[D]. Chengdu: Southwest Petroleum University, 2004.
- [30] 金衍, 卢运虎, 李再均. 一种井漏层位钻前风险预测新方法[J]. 石油钻采工艺, 2008(3): 24-28.
- JIN Yan, LU Yunhu, LI Zaijun. A new method predicting pre-drilling risk in lost circulation formation[J]. Oil Drilling & Production, 2008(3): 24-28.
- [31] 胡莎莎. 压力衰竭地层井漏预测技术研究[D]. 东营: 中国石油大学(华东), 2013.
- HU Shasha. The prediction of lost circulation technology research in pressure depleted formation[D]. Dongying: China University of Petroleum (East China), 2013.
- [32] 刘寿军. 钻井液液面监测与自动灌浆装置的研制[J]. 石油机械, 2006(2): 29-30, 78.
- LIU Shoujun. Development of device for drilling fluid level detection and automatic grout system[J]. China Petroleum Machinery, 2006(2): 29-30, 78.
- [33] Denhiere G. Dynamic memory, a theory of reminding and learning in computers and people-schank, RC [J]. Annee Psychologique, 1985, 85(4): 607-608.
- [34] 张学洪, 李黔. 基于案例推理的井漏风险预警方法[J]. 断块油气田, 2017, 24(2): 255-258, 263.
- ZHANG Xuehong, LI Qian. Risk pre-warning method of well leakage based on case reasoning[J]. Fault-Block Oil & Gas Field, 2017, 24(2): 255-258, 263.
- [35] 谷淑娟, 高学东, 孙冉. 一种改进的CBR案例检索相似性度量模型[J]. 中国管理信息化, 2011, 14(9): 50-55.
- GU Shujuan, GAO Xuedong, SUN Ran. An improved case similarity measure for case retrieval in CBR based on emergency decision[J]. China Management Informationization, 2011, 14(9): 50-55.
- [36] 郭双双. 基于模型算法的网络艺术考级研究分析[D]. 杭州: 浙江大学, 2018.
- GUO Shuangshuang. Research and analysis of network art grading based on model and algorithm[D]. Hangzhou: Zhejiang University, 2018.
- [37] Bassily H. A comparative fault diagnosis methodology based on time series analysis of system's signals[D]. Clemson: Clemson University, 2007.

(编辑 李艺)