

文章编号:1009-9603(2022)01-0190-07

DOI:10.13673/j.cnki.cn37-1359/te.2022.01.024

基于核主成分分析-半监督极限学习机的钻井溢流诊断方法

李仙琳,左 信,高小永,岳元龙
(中国石油大学(北京)信息科学与工程学院,北京 102249)

摘要:溢流是钻井作业中最常见的事故之一,如果对溢流监测与诊断不及时,可能导致严重的井控风险,甚至井喷。钻井现场监测数据较多,直接采用这些数据作为溢流诊断模型的输入会增加模型的复杂度,影响模型的准确率,并且在诊断模型建立过程中存在溢流样本数据标记代价较高的问题。为此建立了基于核主成分分析-半监督极限学习机(KPCA-SSELM)的钻井溢流诊断方法。首先利用核主成分分析(KPCA)算法对钻井各参数进行信息整合,提取其主成分以反映原数据的核心信息,然后选用半监督极限学习机(SSELM)算法进行模型训练,最后利用现场钻井数据与SSELM和KPCA-ELM等模型进行对比实验,验证模型的有效性。结果表明,基于KPCA-SSELM的模型较其他模型具有较高的溢流诊断率及模型泛化能力,采用半监督极限学习机算法能够在钻井数据标记样本比较少的情况下充分挖掘无标签样本所包含的信息训练网络,进一步提高模型的性能,具有很好的应用前景。

关键词:钻井溢流;核主成分分析;半监督学习;极限学习机

中图分类号:TE319

文献标识码:A

Intelligent diagnosis method for kick based on KPCA-SSELM

LI Xianlin, ZUO Xin, GAO Xiaoyong, YUE Yuanlong

(College of Information Science and Engineering, China University of Petroleum(Beijing), Beijing City, 102249, China)

Abstract: Kick is one of the most common accidents in drilling operations. If the kick monitoring and diagnosis are not timely, serious well control risks such as blowouts can be caused. Due to the massive monitoring data in drilling fields, directly using these data as the input of a kick diagnosis model will increase the complexity of the model and thus affect the accuracy of the model. Moreover, the labeling cost of kick samples is high in the process of constructing the diagnosis model. To solve these problems, this paper develops an intelligent diagnosis method for kick diagnosis, which is based on the kernel principal component analysis-semi-supervised extreme learning machine(KPCA-SSELM). Firstly, KPCA is adopted to integrate the drilling parameters, and the principal components are extracted to reflect the core information of the original data. Then, the SSELM algorithm is employed for model training. Finally, the field drilling data is compared with the results of SSELM, KPCA-ELM, and other models to verify the effectiveness of the model. The results show that the proposed model based on KPCA-SSELM has a higher kick diagnosis rate and model generalization ability than other models. The semi-supervised learning method can make full use of the information contained in the unlabeled data to train the network when the number of labeled samples is relatively small, which can further improve the model performance and has a good application prospect.

Key words: kick; kernel principal component analysis; semi-supervised learning; extreme learning machine

钻井过程中,由于井下情况复杂多变,存在极大的随机性与不确定性,钻井事故发生频繁,极大

收稿日期:2021-04-16。

作者简介:李仙琳(1994—),女,安徽池州人,在读硕士研究生,从事钻井事故实时诊断研究。E-mail:lixl0324@163.com。

基金项目:中国石油天然气集团有限公司-中国石油大学(北京)战略合作科技专项“物探、测井、钻完井人工智能理论与应用场景关键技术研究”(ZLZX2020-03),国家重点研发计划资助项目“水下生产系统智能控制关键技术研究”(2016YFC0303703)。

程度上增加了钻井非生产时间,影响整个钻井过程的进度、建井质量和勘探效益^[1]。钻井事故种类繁多,溢流事故当属其中对于安全钻井作业威胁最大的事故之一。若侵入井筒的油气未被及时发现并得到有效控制,不仅会严重减缓钻井开发速度,甚至可能引发卡钻、井塌、井喷等其他井下一系列安全事故,将严重威胁到人身、设备、环境及地下油气资源的安全,同时也会给企业带来恶劣的社会影响和巨大的经济损失^[2]。因此,实时有效的钻井溢流诊断至关重要。

近年来,人工智能、大数据、物联网等新技术的快速发展,以及学科交叉、跨界融合,引发了油气行业的巨大变革^[3-9],钻井过程的数字化与信息化极大提升了钻井效率^[10-12]。在溢流诊断方面,司孟菡对钻井溢流发生前后的表征参数提出了基于改进的BP神经网络的溢流事故监测方法与溢流发展趋势模型^[13];张禾等将钻井专家的知识经验与溢流发生后的参数变化相结合构建了基于专家系统的溢流智能预警技术^[14];涂曦予等将钻井过程看作时间序列分析,建立了基于大规模时间序列分析的钻井事故预警方法^[15];肖肖燕等提出了基于随机森林的溢漏实时判断方法,利用随机森林对初选的特征进行重要性分析^[16];王茜等根据溢流发生前后的表征规律,定义2个分别与钻井工程参数和钻井液参数相关的监测指标,通过时间序列分析算法实现实时监测溢流发生时的异常趋势从而识别溢流^[17]。

尽管在溢流事故诊断方面已有大量成果,但大多忽视了钻井数据中大量未标记数据的价值。在现有的基于机器学习的溢流诊断方法研究中,均建立在假设足够多的训练样本且数据标签已知的情况下。而模型的建立通常需要由具有丰富经验的专家根据现场日志或钻井数据来完成数据标记工作。获取数据标签往往需耗费大量额外的人力、物力。而半监督学习可以实现未标记的数据和有标记的数据同时提供关于数据分布的更有价值的信息。如果在钻井过程中可以利用少量的有标签样本和大量的无标签本来提高溢流诊断性能,可以使机器学习方法在钻井溢流事故诊断领域得到更好的应用。另外,模型的输入均聚焦于总池体积、钻井液进出口流量等溢流发生后的显著特征,其余数据均被抛弃,造成很大的数据浪费。挖掘钻井数据隐含的信息、利用少量的标记样本和大量的未标记本来提升溢流诊断性能是本文的主要出发点。

笔者以真实的钻井溢流数据为基础,首先通过核主成分分析方法对钻井各参数进行核心信息提

取,然后利用半监督极限学习机算法在标记数据较少的情况下,充分挖掘无标记数据所包含的信息训练模型,建立基于核主成分分析-半监督极限学习机(KPCA-SSELM)的钻井溢流诊断方法。实例应用结果表明所提算法利用大量的无标签样本进行辅助学习,能够在故障标记样本比较少情况下达到较高的故障诊断率。

1 方法原理

1.1 核主成分分析算法

核主成分分析(KPCA)算法是在主成分分析算法基础上提出的一种非线性特征提取理论^[18-19],其主要优点就是引入了核方法思想,使算法对非线性数据的处理能力得到了很大程度的提高。钻井各大系统的各特征彼此耦合,所产生的各种数据之间具有极强的相关性及复杂的非线性,因此可以利用KPCA方法来进行钻井数据的信息整合,提取主成分反映核心信息。

1.2 极限学习机算法

极限学习机(ELM)算法是应用于训练单隐层神经网络的算法^[20-21]。假设训练样本有 N 个任意的样本 (x_i, y_i) ,对于一个有 \tilde{N} 个隐层节点的单隐层前馈神经网络可以表示为:

$$\sum_{i=1}^{\tilde{N}} \beta_i g(x_j) = \sum_{i=1}^{\tilde{N}} \beta_i g(w_i \cdot x_j + b_i) = f(x_j) \quad (1)$$

ELM算法旨在学习一个基于训练数据的近似函数或决策规则来估计 y_j 的值,也就是意味着存在 β_i, w_i, b_i ,使得 $\sum_{i=1}^{\tilde{N}} \beta_i g(w_i \cdot x_j + b_i) = y_j$,对应的矩阵形式表示为:

$$H\beta = Y \quad (2)$$

针对模型求解,是期望找到一组特定的解 $\hat{w}_i, \hat{\beta}_i, \hat{b}_i$,使得损失函数 $E = \sum_{j=1}^N \left(\sum_{i=1}^{\tilde{N}} \beta_i g(w_i \cdot x_j + b_i) - y_j \right)^2$ 最小化。

区别于传统的函数逼近,要求求解出隐层参数,极限学习机的网络结构中,在训练过程中仅需设置合适的隐层节点个数,隐层参数可以通过某一概率分布来随机分配,并且在学习过程中,一旦隐层参数被随机确定,隐层输出矩阵 H 就可以保持不变,只需要优化隐层神经元个数。那么训练单隐层神经网络就转化成求解一个线性系统 $\|H\beta - Y\|$ 的最小二乘解 $\hat{\beta}$ 。

1.3 半监督极限学习机算法

半监督学习方法综合了有监督的学习方法和无监督的学习方法,旨在考虑利用少量的标记样本和大量的未标记样本共同参与训练分类器,在减少标记代价的同时提供有效的分类器,弥补有监督学习和无监督学习的不足^[15],学习方法示例如图1所示。

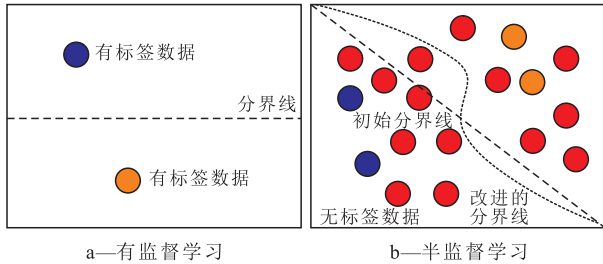


图1 半监督学习方法示例

Fig.1 An example of semi-supervised learning

半监督学习有多种实现方法,本文选择基于图的半监督分类方法^[22-24],该方法基于流形假设和光滑性假设,通过对输入的样本建立一个图,样本代表图上的每个顶点。如果图上的样本之间的权值越大那么它们的预测标记应该相似的可能性就越大。流形正则化是属于基于图的半监督方法中常见的方法之一。半监督极限学习机(SSELM)算法将流形正则化和极限学习机相结合,充分利用大量无标签数据,减少标记数据的工作量,同时继承了极限学习机无需迭代、模型执行高效的优点。因此基于流形正则化的SSELM被描述为:

$$\begin{aligned} \min \frac{1}{2} \|\beta\|^2 + \frac{1}{2} \sum_{i=1}^l C_i \|e_i\|^2 + \frac{\lambda}{2} \text{Tr}(\mathbf{F}^T \mathbf{L} \mathbf{F}) \\ \text{s.t. } \mathbf{H}(\mathbf{x}_i) \beta = \mathbf{y}_i^T - \mathbf{e}_i^T \quad i = 1, 2, \dots, l \\ f_i = \mathbf{H}(\mathbf{x}_i) \beta \quad i = 1, 2, \dots, l + u \end{aligned} \quad (3)$$

通过对不平衡样本中的不同类赋予不同的惩罚参数,缓解过拟合等问题。比如说样本 \mathbf{x}_i 标签为类 \mathbf{y}_i , 同时类 \mathbf{y}_i 有 N_{y_i} 个样本数,取 $C_i = C_0/N_{y_i}$ 。

(3)式可以写成:

$$\min \frac{1}{2} \|\beta\|^2 + \frac{1}{2} \left\| \mathbf{C}^{\frac{1}{2}} (\mathbf{Y} - \mathbf{H}\beta) \right\|^2 + \frac{\lambda}{2} \text{Tr}(\beta^T \mathbf{H}^T \mathbf{L} \mathbf{H} \beta) \quad (4)$$

通过求解(4)式,当梯度为0时,可得:

$$\beta^* = (\mathbf{I} + \mathbf{H}^T \mathbf{C} \mathbf{H} + \lambda \mathbf{H}^T \mathbf{L} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C} \mathbf{Y} \quad (5)$$

2 诊断方法建立

提出了一种基于KPCA-SSELM的钻井溢流诊断方法,其流程如图2所示。在模型训练过程中,主要包括以下步骤:①钻井数据预处理。受传感器测

量等因素影响在现场会收集到部分无效数据,同时存在部分测量噪声,这些数据无法为溢流的诊断提供有效的信息,因此需要对数据进行清洗和质量优化。针对部分缺失数据利用相邻的数据平均值代替,将无效的数据剔除。针对存在噪声与离群点数据采用滑动窗平滑滤波以及 3σ 准则进行处理。特征缩放对于将每个特征权重推广到机器学习算法中是至关重要的。钻井原始数据具有不同的尺度,在训练时会导致过长的训练时间及过度拟合,因此需要对输入数据进行标准化处理,消除各个钻井参数之间的量纲影响。②核主成分分析。利用KPCA法对归一化处理后的数据进行特征提取,首先选择高斯径向基核函数计算核矩阵,然后计算核矩阵对应的特征值和特征向量,最后计算得分主元,选择累积贡献率大于90%的各主成分建立新的模型样本。③模型输入。将经过核主成分分析后的新集合划分为训练集、验证集和测试集。其中训练集包含有标签样本与无标签样本两部分,验证集用于SSELM中的超参数选取,测试集用于模型评估。④模型训练。利用SSELM算法进行模型训练,分为2个阶段。第一阶段生成随机隐层;在第二阶段,取相似性度量函数 $w_{ij} = e^{-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / (2\sigma^2)}$ 建立图拉普拉斯矩阵,训练模型。其中在SSELM算法中寻找最优超参数 C_0 和 λ 过程中是基于验证集性能评估从指数序列 $\{10^{-6}, 10^{-5}, \dots, 10^6\}$ 进行优选,并通过(5)式计算 β , 模型输出为 $f_i = \mathbf{h}(\mathbf{x}_i) \beta$ 。⑤模型评估。利用测试集对训练的最优SSELM模型进行性能评估。

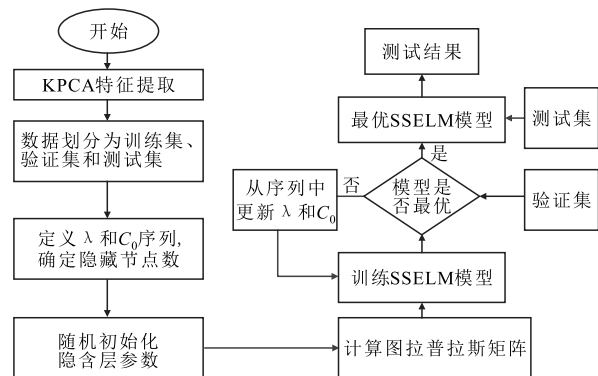


图2 基于KPCA-SSELM的钻井溢流诊断模型建立
Fig.2 Kick diagnosis process based on KPCA-SSELM

3 模型训练

3.1 数据获取

选用中国某油田同一区块的钻井现场数据,其中收集到的数据涉及到正常钻进、起下钻、接单根

等操作变化。在事故报告中发现所获溢流事故数据大都发生于钻进过程中,因此通过比较井深与钻头深度、钻压和转速大于零等条件分析钻进状态和停钻状态。仅与钻进状态对应的样本数据记录在数据集中,供本研究进行进一步分析。通过整理,最终将同区块的4口井共计6次发生在钻进过程的溢流事故的当日报井数据作为原始数据集,4次事故数据用于训练模型,2次事故数据用于验证模型。

实时监测的钻井过程数据主要来自传感器或测量工具,反映钻井作业的安全状态。所获数据包括:井深、钻头深度、钻压、转速、钻速、扭矩、立管压力、大钩载荷、大钩高度、泵冲、钻井液进出口流量、钻井液进出口温度、钻井液进出口电导率、钻井液进出口密度、总池体积等共19维。

由于钻井正常样本远远多于溢流样本,容易造成模型过拟合问题,取溢流发生前后半小时之内的数据作为数据集,整个数据集由4部分组成:有标签与无标签样本的训练集、验证集和测试集。有标签和无标签数据集用于训练模型,验证集用于模型参数的选择,在训练出最优模型之后,测试集用于最终的验证。最终训练集样本数为6 000,验证集样本数为1 800,测试集样本数为3 000。

3.2 实验设计与模型评估

将溢流的诊断问题建模为钻井测量数据为自变量、是否发生溢流为结果的分类模型,数据的分类结果为离散值,0表示无溢流发生,1表示溢流发生。综合录井数据不包含溢流标签信息,但是现场事故记录报告记录了溢流发生时间和结束时间。考虑到人工坐岗监测的滞后性,以事故记录报告作为参考,通过观察相关数据在该时间段前后的变化来进行重新标注,其中主要关注钻井液进出口流量差、总池体积、钻速、钻压等参数的变化。

为了验证KPCA-SSELM方法在溢流事故诊断上的有效性,设置了2组实验,分别与SSELM和KPCA-ELM等进行对比。其中SSELM中输入数据选用溢流表征参数,包括总池体积、钻井液进出口流量、立管压力、钻压、钻速等6个参数,用来对比验证KPCA特征提取的有效性。同时为了验证针对钻井过程中少量有标签样本,半监督学习方法用于溢流诊断的有效性,在相同的有标签样本下与有监督学习方法KPCA-ELM进行对比。

模型性能评估一般通过分类准确率来评价,但针对钻井样本中存在正常数据与溢流数据不平衡的问题,分类准确率不具有太大说服力。因此另外选择精确率、召回率、 F_1 值来评估模型性能优劣,各

评价指标计算公式为:

$$A = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

$$P = \frac{TP}{TP + FP} \quad (7)$$

$$R = \frac{TP}{TP + FN} \quad (8)$$

$$F_1 = \frac{2PR}{P + R} \quad (9)$$

3.3 模型训练与结果分析

首先对选取的样本数据进行标准化处理,然后采用KPCA对处理后的数据进行主元分析,按照经验确定累积贡献率为90%。核函数选取高斯径向基核函数,其中参数 σ 设置为100,分析处理后得到的钻井各参数特征值及对应的贡献率(表1)。经过KPCA处理后,前5个核主成分的累积贡献率达到了90.78%,也就是说,所获的钻井数据集中的19个特征指标所携带的信息量经KPCA处理后大部分集中在前面的5个核主成分上;并且获得的主成分的贡献率比较集中,较好地保存了原始数据的绝大部分信息,也很好达到了降维的目的。

表1 钻井各参数特征值及对应的贡献率

Table1 Eigenvalue and contribution rate of drilling parameters

主成分	特征值	方差贡献率/%	累积贡献率/%
1	8.268 4	32.92	32.92
2	4.642 5	18.48	51.4
3	3.799 1	15.12	66.52
4	3.171 6	12.63	79.15
5	2.921 9	11.63	90.78

SSELM与KPCA-SSELM模型对比 将KPCA方法处理后的数据与利用溢流表征参数做特征选择的数据作为输入分别送入SSELM中进行训练。训练中网络隐含节点数为120,激活函数选择Sigmoid函数, $C_0 = 0.1, \lambda = 0.01$ 。有标签样本数与无标签样本数各自取训练集总数的50%。由对比结果(表2)可知,KPCA-SSELM性能好于基于特征选择的SSELM,证明了KPCA针对钻井数据在特征提取上有很大的优势。面对未来钻井过程中可监测参数更多的情况,基于核主成分分析的特征提取方法用于钻井溢流诊断可以获得较好的结果。

表2 不同模型性能对比结果

Table2 Performance comparison results of models %

模型	准确率	精确率	召回率	F_1 值
SSELM	91.08	88.83	86.89	87.84
KPCA-SSELM	92.16	90.51	89.45	89.97

KPCA-ELM与KPCA-SSELM模型对比 为了

验证针对钻井过程中少量有标签样本,半监督学习用于溢流诊断的有效性,首先在训练集中,有标签样本在正常和溢流的样本中随机选取10%。KPCA-ELM方法仅使用这少量的有标签样本进行训练,KPCA-SSELM方法同时使用有标签样本和剩下的无标签样本进行训练。此外针对不同数量的有标签样本条件下2种算法性能的比较,分别从有标签样本数占训练样本总数的10%开始递增进行训练,性能比较结果(图3,表3,图4)显示,引入半监督学习的KPCA-SSELM方法始终优于监督学习KPCA-ELM方法,这表明KPCA-SSELM方法能够有效地探索未标记数据,以实现比纯监督学习方法更好的性能。可以发现,当有少量的标记数据时,对于KPCA-ELM方法来说,训练集样本数很少,模型性能并不好,而增加了半监督学习的KPCA-SSELM方法,由于增加了大量的无标签样本信息,模型性能

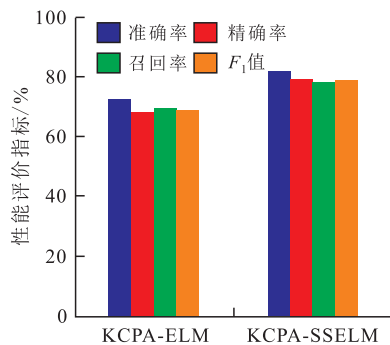


图3 10%有标签样本数的算法性能对比
Fig.3 Comparison of model performance with 10% labeled samples

表3 不同数量有标签样本条件下2种算法性能对比结果
Table3 Performance comparison results of two algorithms with different proportion of labeled samples %

模型	有标签样本数占训练样本总数的比例	准确率	精确率	召回率	F ₁ 值
KPCA-ELM	10	72.67	68.34	69.66	68.99
	20	76.12	72.59	71.45	72.01
	30	80.27	78.11	76.37	77.23
	40	86.45	84.69	83.14	83.90
	50	89.03	88.24	88.45	88.34
	60	91.28	90.07	89.93	89.99
KPCA-SSELM	10	82.14	79.45	78.57	79.00
	20	84.21	82.31	80.81	81.55
	30	85.63	83.43	82.83	83.12
	40	88.92	87.11	86.65	86.87
	50	93.16	90.51	89.45	89.97
	60	93.68	91.73	91.41	91.56

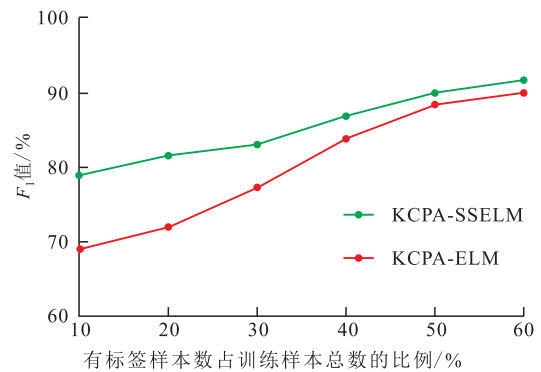


图4 不同有标签样本数下的模型性能对比

Fig.4 Performance comparison results with different proportion of labeled samples

有了一定的提升。随着有标签样本数量的增加,2种模型性能均不断提高。但在每一种情况下,加入半监督学习的KPCA-SSELM方法均高于有监督学习KPCA-ELM方法,这是因为无标签样本提供了更多的信息,表明KPCA-SSELM方法可以充分挖掘无标签样本的信息来进一步提高模型的性能。

但从每个测试样本来看,整体模型性能略低。因为在模型训练中,虽然为考虑数据不平衡,选择了溢流前后半小时之内的数据训练,但正常样本仍多于溢流样本,较少的溢流样本无法使得模型学习到全部的信息。

3.4 实际测试

在实际钻井过程中,除了对每个测试样本进行模型的相应评估外,还需要对模型的应用进行分析与评估。图5给出了测试样本中1次溢流事故发生前后20分钟的数据变化以及模型诊断结果,其中该测试结果建立在有标签样本数为训练样本总数的50%上。混淆矩阵结果见表4。

从图5可以看出,总池体积在9:05开始增加,超过1 m³记为溢流发生,但是在此时间之前,钻压、钻速、出口流量有了一定的变化,通过模型输出可以看到在9:03左右模型开始预警,虽然存在一定的错误报警,但比传统的总池体积监测溢流提前了2分钟。

4 结论

提出了一种基于KPCA-SSELM的钻井溢流智能诊断方法。首先利用KPCA方法分析得出对钻井原始数据关联超过90%的5个主成分作为模型的输入向量,可以很好地提升模型诊断率;然后将半监督学习方法引入极限学习机的模型中,该方法可以有效地利用大量无标签样本来辅助有标签样本进

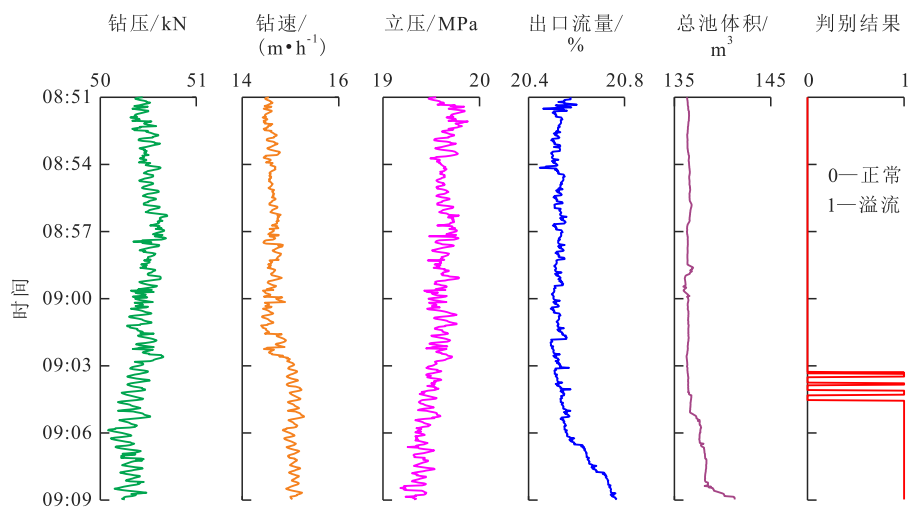


图5 溢流事故诊断图

Fig.5 Kick diagnosis diagram

行训练,很好地解决了钻井数据标记费时费力的问题。通过利用无标签样本进行辅助训练,获得比传统极限学习机更高的溢流诊断正确率,具有很好的应用前景。

表4 混淆矩阵
Table4 Confusion matrix

实际	预测	
	溢流	正常
溢流	322	35
正常	33	810

符号解释

- A——准确率;
- b_i ——偏置向量;
- C_i ——类间惩罚参数;
- C_0 ——自定义正则化参数;
- e_i ——误差;
- F ——经模型运算后的输出矩阵,其第*i*行的值为 f_i ;
- FN ——假阴性,代表溢流被错误认定为非溢流的个数;
- FP ——假阳性,代表非溢流被错误识别为溢流的个数;
- g ——激活函数;
- H ——隐层节点的输出;
- I ——单位矩阵;
- l, u ——有标签样本和无标签样本的个数;
- L ——由有标签样本和无标签样本构建的图拉普拉斯矩阵, $L \in \mathbb{R}^{(l+u) \times (l+u)}$;
- N ——训练样本数量;
- \tilde{N} ——隐层节点数量;
- P ——精确率;
- R ——召回率;
- TP ——真阳性,代表正确识别溢流的个数;

- TN ——真阴性,代表非溢流被正确识别为非溢流的个数;
- $Tr(\cdot)$ ——矩阵的迹;
- w_i ——输入和输出之间的权重向量;
- x_j ——输入向量, $j = 1, 2, \dots, N$;
- Y ——期望输出;
- β ——输出权重;
- σ ——宽度函数;
- β_i ——连接模型中第*i*个隐层和输出之间的权重向量, $i = 1, 2, \dots, \tilde{N}$;
- λ ——流形正则化的正则化因子。

参考文献

- [1] 王江萍, 孟祥芹, 鲍泽富. 钻进参数实时监测与故障诊断技术[J]. 钻采工艺, 2008, 31(1): 49-52.
WANG Jiangping, MENG Xiangqin, BAO Zefu. Technology of drilling parameters real-time monitoring and fault diagnosis [J]. Drilling and Production Technology, 2008, 31(1): 49-52.
- [2] 王海涛. 关于钻井井控设备管理的研究与应用[J]. 石化技术, 2020, 27(1): 290-291.
WANG Haitao. Research and application of drilling well control equipment management [J]. Petrochemical Industry Technology, 2020, 27(1): 290-291.
- [3] 吴君达, 李治平, 孙妍, 等. 基于神经网络的剩余油分布预测及注采参数优化[J]. 油气地质与采收率, 2020, 27(4): 85-93.
WU Junda, LI Zhiping, SUN Yan, et al. Neural network-based prediction of remaining oil distribution and optimization of injection-production parameters [J]. Petroleum Geology and Recovery Efficiency, 2020, 27(4): 85-93.
- [4] 刘巍, 刘威, 谷建伟, 等. 利用卡尔曼滤波和人工神经网络相结合的油藏井间连通性研究[J]. 油气地质与采收率, 2020, 27(2): 118-124.
LIU Wei, LIU Wei, GU Jianwei, et al. Research on interwell con-

- nectivity of oil reservoirs based on Kalman filter and artificial neural network [J]. *Petroleum Geology and Recovery Efficiency*, 2020, 27(2): 118-124.
- [5] 王泽龙, 刘先贵, 唐海发, 等. 基于多次数据吸收集合平滑算法的自动油藏历史拟合研究[J]. *特种油气藏*, 2021, 28(3): 99-105.
WANG Zelong, LIU Xiangui, TANG Haifa, et al. Study on automatic reservoir history matching based on ES-MDA algorithm [J]. *Special Oil & Gas Reservoirs*, 2021, 28(3): 99-105.
- [6] 程翊珊, 李治平, 许龙飞, 等. 预测油层无机积垢的BP神经网络方法[J]. *大庆石油地质与开发*, 2021, 40(3): 84-93.
CHENG Yishan, LI Zhiping, XU Longfei, et al. BP neural network method for predicting the inorganic scaling in the reservoir [J]. *Petroleum Geology & Oilfield Development in Daqing*, 2021, 40(3): 84-93.
- [7] 石玉江, 刘国强, 钟吉彬, 等. 基于大数据的测井智能解释系统开发与应用[J]. *中国石油勘探*, 2021, 26(2): 113-126.
SHI Yujiang, LIU Guoqiang, ZHONG Jibin, et al. Development and application of intelligent logging interpretation system based on big data [J]. *China Petroleum Exploration*, 2021, 26(2): 113-126.
- [8] 余晓露, 叶恺, 杜崇娇, 等. 基于卷积神经网络的碳酸盐岩生物化石显微图像识别[J]. *石油实验地质*, 2021, 43(5): 880-885, 895.
YU Xiaolu, YE Kai, DU Chongjiao, et al. Microscopic recognition of micro fossils in carbonate rocks based on convolutional neural network [J]. *Petroleum Geology and Experiment*, 2021, 43(5): 880-885, 895.
- [9] 黄家宸, 张金川. 机器学习预测油气产量现状[J]. *油气藏评价与开发*, 2021, 11(4): 613-620.
HUANG Jiachen, ZHANG Jinchuan. Overview of oil and gas production forecasting by machine learning [J]. *Reservoir Evaluation and Development*, 2021, 11(4): 613-620.
- [10] 李凤霞, 崔茂荣, 王丽华, 等. 综合录井技术在实时监测钻井事故中的应用[J]. *断块油气田*, 2007, 14(3): 66-68.
LI Fengxia, CUI Maorong, WANG Lihua, et al. Application of comprehensive logging technology in real time monitoring drilling accidents [J]. *Fault-Block Oil and Gas Field*, 2007, 14(3): 66-68.
- [11] 葛景凯. 综合录井参数在钻井工程异常预警中的应用探讨[J]. *信息系统工程*, 2020, (3): 97-98.
GE Jingkai. Discussion on application of comprehensive logging parameters in early warning of drilling engineering anomalies [J]. *China CIO News*, 2020, (3): 97-98.
- [12] 李根生, 宋先知, 田守增. 智能钻井技术研究现状及发展趋势[J]. *石油钻探技术*, 2020, 48(1): 1-8.
LI Gensheng, SONG Xianzhi, TIAN Shouceng. Intelligent drilling technology research status and development trends [J]. *Petroleum Drilling Techniques*, 2020, 48(1): 1-8.
- [13] 司孟菡. 钻井溢流实时预警方法研究[D]. 成都: 西南石油大学, 2016.
SI Menghan. Research on real-time warning method of drilling kick [D]. Chengdu: Southwest Petroleum University, 2016.
- [14] 张禾, 李祁颖, 张露之, 等. 基于模糊专家系统的钻井溢流智能预警技术[J]. *西南石油大学学报: 自然科学版*, 2016, 38(2): 169-175.
ZHANG He, LI Qiyang, ZHANG Luzhi, et al. Early intelligent kick warning in well drilling based on fuzzy expert system [J]. *Journal of Southwest Petroleum University: Science & Technology Edition*, 2016, 38(2): 169-175.
- [15] 涂曦予, 于露, 耿子辰, 等. 基于大规模时间序列的井漏事故预警方法[J]. *信息技术*, 2018, (12): 1-4.
TU Xiyu, YU Lu, GENG Zichen, et al. Research on intelligent warning of loss circulation based on large-scale time series data [J]. *Information Technology*, 2018, (12): 1-4.
- [16] 史肖燕, 周英操, 赵莉萍, 等. 基于随机森林的溢漏实时判断方法研究[J]. *钻采工艺*, 2020, 43(1): 9-12.
SHI Xiaoyan, ZHOU Yingcao, ZHAO Liping, et al. Study on how to determine kicks and losses in real time on basis of random forest method [J]. *Drilling and Production Technology*, 2020, 43(1): 9-12.
- [17] 王茜, 张菲菲, 李紫璇, 等. 基于钻井模型与人工智能相耦合的实时智能钻井监测技术[J]. *石油钻采工艺*, 2020, 42(1): 6-15.
WANG Xi, ZHANG Feifei, LI Zixuan, et al. Real-time intelligent drilling monitoring technique based on the coupling of drilling model and artificial intelligence [J]. *Oil Drilling & Production Technology*, 2020, 42(1): 6-15.
- [18] CHEN Z G, REN H D, DU X J, et al. Minimax probability machine classifier with feature extraction by kernel PCA for intrusion detection [C]. Dalian: 4th International Conference on Wireless Communications, Networking and Mobile Computing, 2008.
- [19] 范玉刚, 李平, 宋执环. 基于特征样本的KPCA在故障诊断中的应用[J]. *控制与决策*, 2005, 20(12): 1 415-1 418, 1 422.
FAN Yugang, Li Ping, SONG Zhihuan. KPCA based on feature samples for fault detection [J]. *Control and Decision*, 2005, 20(12): 1 415-1 418, 1 422.
- [20] HUANG G B, ZHU Q Y, SIEW C K. Extreme learning machine: a new learning scheme of feedforward neural networks [C]. Budapest: 2004 IEEE International Joint Conference on Neural Networks.
- [21] HUANG G B, ZHU Q Y, SIEW C K. Extreme learning machine: theory and applications [J]. *Neurocomputing*, 2006, 70(1/3): 489-501.
- [22] 梁吉业, 高嘉伟, 常瑜. 半监督学习研究进展[J]. *山西大学学报: 自然科学版*, 2009, 32(4): 528-534.
LIANG Jiye, GAO Jiawei, CHANG Yu. The research and advances on semi-supervised learning [J]. *Journal of Shanxi University: Natural Science Edition*, 2009, 32(4): 528-534.
- [23] BELKIN M, MATVEEVA I, NIYOGI P. Regularization and semi-supervised learning on large graphs [C]. Montreal: 17th Annual Conference on Learning Theory, 2004: 624-638.
- [24] BELKIN M, NIYOGI P, SINDHWANI V. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples [J]. *The Journal of Machine Learning Research*, 2006, 7(1): 2 399-2 434.