

<https://doi.org/10.3799/dqkx.2022.409>



准噶尔盆地玛湖凹陷风城组陆相页岩油储层测井裂缝智能识别

陆国青^{1,2},董少群^{1,3},黄立良⁴,曾联波^{1,2,5*},刘国平⁵,何文军⁴,
杜晓宇⁵,杨森⁴,高文颖⁴

1. 中国石油大学油气资源与探测国家重点实验室,北京 102249
2. 中国石油大学地球科学学院,北京 102249
3. 中国石油大学理学院,北京 102249
4. 中国石油新疆油田分公司,新疆克拉玛依 834000
5. 北京大学能源研究院,北京 100871

摘要:准噶尔盆地玛湖凹陷北部斜坡区风城组陆相页岩油储层为多物源混合沉积,多种岩性频繁互层,岩石力学层厚度小,导致其裂缝尺度小,裂缝常规测井响应弱,识别难度大。针对页岩裂缝测井识别的难题,应用集成学习中的极端梯度提升树方法,通过深度挖掘裂缝信息与测井数据之间的非线性关系,将多个弱分类器集成强分类器,降低裂缝识别的不确定性,以提高裂缝的识别能力。该方法将岩心裂缝描述和井壁成像测井裂缝解释结果作为标签,常规测井信息作为模型训练的输入数据,在异常点筛查、SMOTE过采样处理和特征优选的基础上,通过网格搜索方法获得裂缝智能识别模型的最优超参数。通过与目前常用的支持向量机和逻辑回归等机器学习方法对比,极端梯度提升树具有比其他两种非线性机器学习方法更好的裂缝识别效果,测试集识别准确率可达到90%。A1井风3段识别结果反映了该段裂缝较为发育,且模型对于裂缝段与非裂缝段都具有较好的识别效果,与岩心观察结果符合率较高。表明极端梯度提升树具有较好的裂缝识别能力,能够为玛湖凹陷陆相页岩油储层的裂缝智能识别提供有效手段。

关键词:极端梯度提升树;裂缝智能识别;陆相页岩油;玛湖凹陷;准噶尔盆地;石油地质学。

中图分类号: P618.13

文章编号: 1000-2383(2023)07-001-13

收稿日期: 2022-05-12

Fracture Intelligent Identification Using Well Logs of Continental Shale Oil Reservoir of Fengcheng Formation in Mahu Sag, Junggar Basin

Lu Guoqing^{1,2}, Dong Shaoqun^{1,3}, Huang Liliang⁴, Zeng Lianbo^{1,2,5*}, Liu Guoping⁵, He Wenjun⁴,
Du Xiaoyu⁵, Yang Sen⁴, Gao Wenyi⁴

1. State Key Laboratory of Petroleum Resources and Prospecting, China University of Petroleum, Beijing 102249, China
2. College of Geosciences, China University of Petroleum, Beijing 102249, China
3. College of Science, China University of Petroleum, Beijing 102249, China

基金项目:国家自然科学基金项目(Nos.42090020, U1663203)。

作者简介:陆国青(1995—),男,博士研究生,从事裂缝识别与预测、人工智能等相关研究。ORCID: 0000-0003-4570-0345. E-mail: zsdhdl-gq@126.com

*通讯作者:曾联波,ORCID: 0000-0002-6470-8206. E-mail:lbzeng@sina.com

引用格式:陆国青,董少群,黄立良,曾联波,刘国平,何文军,杜晓宇,杨森,高文颖,2023.准噶尔盆地玛湖凹陷风城组陆相页岩油储层测井裂缝智能识别.地球科学,48(7):1—13.

Citation: Lu Guoqing, Dong Shaoqun, Huang Liliang, Zeng Lianbo, Liu Guoping, He Wenjun, Du Xiaoyu, Yang Sen, Gao Wenyi, 2023. Fracture Intelligent Identification Using Well Logs of Continental Shale Oil Reservoir of Fengcheng Formation in Mahu Sag, Junggar Basin. *Earth Science*, 48(7):1—13.

4. PetroChina Xinjiang Oilfield Company, Karamay 834000, China

5. Institute of Energy, Peking University, Beijing 100871, China

Abstract: The continental shale oil reservoir of Fengcheng Formation in the northern slope area of Mahu Sag, Junggar Basin is a mixed deposition of multiple provenances, with frequent interbedding of various lithologies and small thickness of rock mechanics layer, resulting in small fracture scale, weak conventional logging response of fractures and great difficulty in identification. Aiming at the logging identification of shale fractures, this paper applies the extreme gradient boosting (XGBoost) method in ensemble learning to deeply mine the nonlinear relationship between fracture information and logging data, integrates multiple weak classifiers into strong classifiers, reduces the uncertainty of fracture identification, and improves the accuracy of fracture identification. In this method, core fracture description and fracture interpretation results of imagelogging are used as labels, and conventional logging information is used as input data for model training. On the basis of outlier screening, SMOTE oversampling and feature optimization, the optimal hyperparameters of fracture intelligent identification model are obtained through grid search method. Compared with the commonly used machine learning methods such as support vector machine and logical regression, the XGBoost has better fracture identification ability than the other two nonlinear machine learning methods, and the accuracy of test set identification reaches 90%. The identification results of the P_{f_3} of well A1 reflect that the fractures in this section are relatively developed, and the model has a good identification ability for both fractured and non-fractured sections, with a high coincidence rate with the core observation results. It can provide effective means for intelligent identification of fractures in continental shale oil reservoirs in Mahu Sag.

Key words: XGBoost; fracture intelligent identification; continental shale oil; Mahu Sag; Junggar Basin; petroleum geology.

0 引言

随着油气勘探开发从常规资源向非常规资源转移,页岩油气资源正成为研究热点(Ghosh *et al.*, 2018; Gong *et al.*, 2021).页岩油作为一种非常规油气资源,已成为全球能源格局多元化的重要组成部分.迄今为止,大多数页岩油勘探目标都是海相地层,而我国主要为陆相页岩油,具有油质重、黏度高、流动能力差的特征(金之钧等,2021).准噶尔盆地西北缘玛湖凹陷二叠系风城组发育中-高成熟度碱湖优质烃源岩,是一种新型的碱性湖盆沉积页岩油系统(曹剑等,2015;支东明等,2021).由于风城组页岩油储层物性较差,仅靠岩石基质孔隙难以形成有效渗流体系,而天然裂缝可以有效改善储层物性,提高油气产能(Zeng, 2010; Zeng *et al.*, 2016; Liu *et al.*, 2020a),因此,天然裂缝的研究对于风城组页岩油的有效勘探开发极为必要.此外,风城组岩相变化快、非均质性强,对页岩油储层天然裂缝单井识别技术提出了更高的要求(Liu *et al.*, 2020b; Zeng *et al.*, 2021).

常规测井资料识别裂缝实际是对微弱信号的提取与放大的过程.然而,研究区风城组裂缝测井曲线响应特征复杂,没有单一的测井曲线可以明确指示裂缝的存在,此外,流体类型、成岩作用和粘土含量等外部因素也可以产生与裂缝相似的测井响

应特征,导致裂缝段与非裂缝段在测井曲线交会图中存在相当大的重叠区域,使常规测井裂缝识别成为了复杂的非线性问题(Dong *et al.*, 2020a).

机器学习通过建立测井曲线与裂缝之间的非线性识别模型,对未知样本进行了预测(罗刚等,2022;宋随宏等,2022;宋先知等,2022).前人针对裂缝的智能识别展开过大量研究,石广仁(2008)利用支持向量机对南襄盆地泌阳凹陷安棚油田安1井和安2井进行裂缝识别,取得了较好的分类效果.Li *et al.*(2018)从声波测井数据中提取多种特征,利用GA-SVM方法预测了鄂尔多斯盆地碳酸盐岩储层的裂缝密度,通过对比岩心观察和电阻率测井解释结果,验证了模型预测效果良好.陈钢花等(2015)从缝洞充填物测井响应特征入手,优选测井敏感性参数,利用BP神经网络建立缝洞充填物解释模型,提高了缝洞充填物的识别精度.Xue *et al.*(2014)利用GA-BP神经网络,通过深、浅电阻率测井的振幅及其差异,对松辽盆地新立油田进行裂缝识别,取得了较好的裂缝识别效果.

由于机器学习模型在页岩裂缝识别中的性能在很大程度上取决于特征选择,特征选择用于从原始特征中选择特征子集,以实现更高的分类精度.与BP、SVM等机器学习算法相比,极端梯度提升树(XGBoost)算法在底层采用CART分类回归树,并建立了一种有效的特征选择机制.此外,

极端梯度提升树可将多个弱学习器组合成强学习器(Gu et al., 2021),模型通过阶段学习策略调整权重,避免模型分类能力有限带来的裂缝识别效果不佳的问题,有效提高了识别精度。同时,考虑到玛湖陆相页岩油储层岩性复杂且纵向变化快,不同岩性脆性矿物含量和地应力的差异导致构造裂缝发育程度不同,因此,将岩石脆性曲线也加入极端梯度提升树的模型训练中,并通过后验井段检测该方法的识别效果。

1 研究区概况

玛湖凹陷位于准噶尔盆地西北缘,为一个二级构造单元(图1a),凹陷西侧自北向南为乌夏断裂带、克百断裂带和中拐凸起,凹陷东部自北向南为石英滩凸起、英西凹陷、三个泉凸起、夏盐凸起和达巴松凸起(支东明等,2021)(图1b)。受到西准噶尔洋向哈萨克斯坦板块的强烈碰撞挤压运动影响,尤其是中晚石炭世—早二叠世,准噶尔地块与哈萨克斯坦板块的碰撞作用加剧,在盆地西北缘形成大型推覆体构造,是玛湖凹陷最重要的形成时期(雷德文等,2017)。下二叠统风城组沉积时期是西部前陆盆地系统剧烈发育期,形成了盆地最重要的一套烃源岩(支东明等,2021),目前已发现与之有关的克-乌、玛湖西斜坡两大百里油区,累计探明储量 17.9×10^8 t,是全球著名的大油区(唐勇等,2019)。为了推进玛湖凹陷页岩油勘探,MY1井于2008年在玛湖凹陷斜坡区相对较浅的位置进行了钻探,2020年试油峰值日产油 50.6 m³,展现了良好的勘探开发前景。

玛湖凹陷风城组主体为深-半深碱湖背景下的多源混合细粒沉积建造(支东明等,2019),存在内源化学沉积、火山物质沉积,以及陆源碎屑沉积(支东明等,2021)。风城组自下而上划分为风一段(P_{lf_1})、风二段(P_{lf_2})、风三段(P_{lf_3}),不同深度段沉积物差异明显,形成了砂岩、白云质岩、泥岩、凝灰岩等多种岩石类型(图1c);其中,白云质泥岩以及泥质粉细砂岩中脆性矿物含量高,在多期构造变形作用下,天然裂缝普遍发育。根据裂缝走向可划分为四组,分别为近东西向、近南北向、北西-南东向和北东南西向。经岩心观察描

2.2 极端梯度提升树的裂缝智能识别流程

由于研究区风城组陆相页岩油储层岩性复杂

且频繁互层,控制裂缝形成与分布的岩石力学层厚度变薄,裂缝尺度小,导致裂缝测井响应弱,识别难

2 原理与方法

2.1 极端梯度提升树原理

集成学习是机器学习中的一种重要类型,对于解决非线性分类问题具有较好的效果(Fei et al., 2021; 鲍明阳等,2023)。集成学习可分为套袋算法(Bagging)和推进算法(Boosting),套袋算法为随机有放回的采样,在采样过程中保持数据集规模不变,在训练过程中,每个弱学习器具有相同权重,彼此独立无关联,通过取平均值或者投票机制得到最终分类结果。推进算法是通过序列的方式生成弱学习器,后面生成的与前一个相关,首先使用初始权重从训练集中训练出一个弱学习器,根据弱学习器的学习误差率来更新样本的权重,误差率高的样本会获得更高的权重在后面的弱学习器中得到更多的重视,每添加一个弱学习器权重会重新调整,如此循环,直到得到指定数量的学习器;再通过结合策略进行整合,得到最终的强学习器(鲍明阳等,2023)。

极端梯度提升树属于典型的推进算法,具有以下优点:(1)使用CART树作为基学习器,不需要关心特征之间是否存在相关性,识别速度快,能够自动组合多个特征,快速处理特征之间的交互关系,使得训练过程更为高效(Torlay et al., 2017).(2)每棵决策子树并不是独立的,而是彼此之间能够相互学习,减小联合估计量的偏差;对于输出结果而言,该方法不是采取多数投票产生识别结果,而是对一系列决策子树结果进行加权求和,是基于权值的弱学习器集成(图2).(3)极端梯度提升树在损失函数中加入正则化项,对叶子节点个数和叶子输出分别做L1、L2正则化,可以缓解过拟合的情况,在计算损失时使用了二阶泰勒展开,更好地拟合了损失下降曲线;同时,该方法也加入了收缩参数 ρ ,在每次迭代完决策树后将叶子节点权重乘以该系数,削弱每棵树的影响,进而控制学习速率。

基于上述优点,极端梯度提升树在构建裂缝智能识别模型时,决策子树中被错误分类的样本将有更大的机会被选择或分配更高的权重,提高了裂缝识别的准确率。

且频繁互层,控制裂缝形成与分布的岩石力学层厚度变薄,裂缝尺度小,导致裂缝测井响应弱,识别难

两行排版错位,改样时会调整

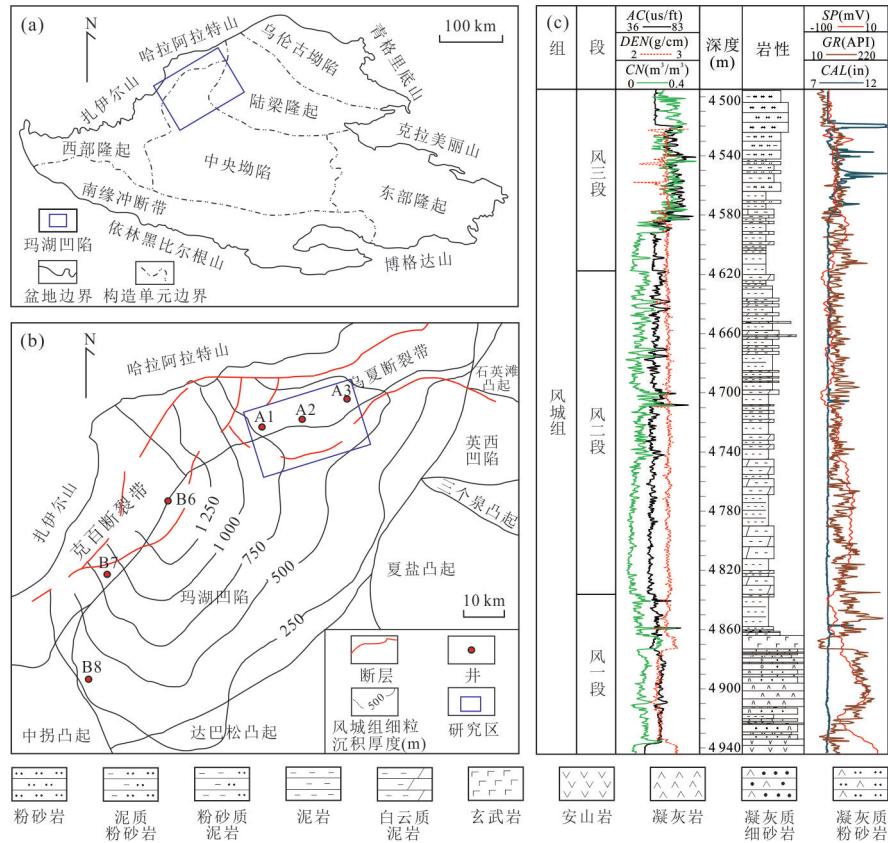


图 1 准噶尔盆地玛湖凹陷构造区划图

Fig. 1 Tectonic map of Mahu Sag in Junggar Basin

据支东明等(2021)修改。a. 准噶尔盆地构造区划分图;b. 玛湖凹陷构造区划分图;c. 风城组综合柱状图

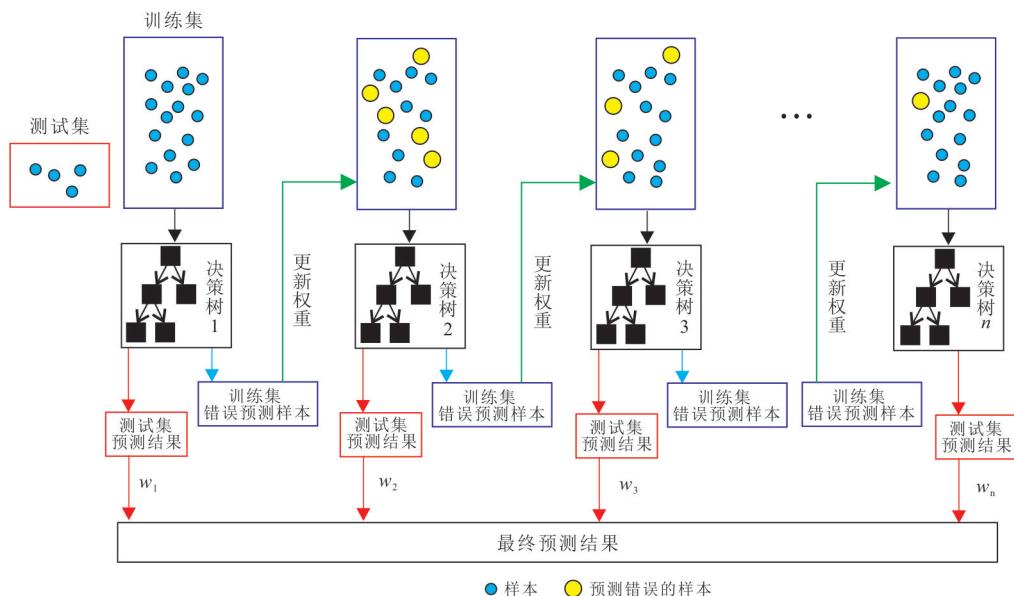


图 2 极端梯度提升树原理图

Fig.2 Schematic diagram of XGBoost

度大。因此,本文利用集成学习中极端梯度提升树深度挖掘裂缝与常规测井曲线之间的非线性关系,以提高天然裂缝识别精度。根据极端梯度提升

树的基本原理,总结出极端梯度提升树识别裂缝的工作流程(图3)。其中,输入数据包括岩心、成像测井标记的常规测井曲线,由于常规测井

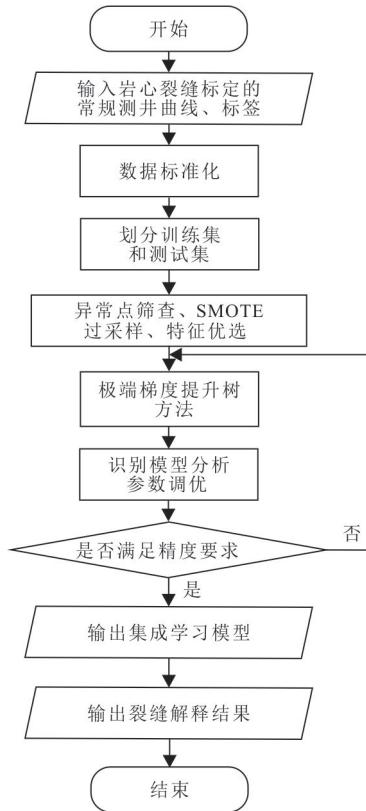


图3 基于极端梯度提升树的裂缝智能识别流程图

Fig.3 Flow chart of intelligent fracture identification based on XGBoost

曲线取值范围变化很大,可能会导致目标函数识别错误,因此所有的数据首先将进行标准化,并按照7:3的比例划分训练集和测试集;在异常点筛查和去除、SMOTE过采样处理以及特征优选后输入极端梯度提升树,并随机划分训练子集参与模型训练,经多次迭代得到最优识别模型,并用于后验井段以检测模型识别效果。

3 玛湖凹陷风城组陆相页岩油储层裂缝智能识别

3.1 裂缝测井响应分析

由于页岩油储层裂缝段与非裂缝段测井响应特征因其特定的物理性质会存在差异(Zeng *et al.*, 2010; Dong *et al.*, 2020b),因此,理论上讲测井曲线幅度变化可以反映裂缝段是否发育;但在实际应用中,测井曲线的响应特征不仅仅受裂缝段的影响,还受粘土矿物、孔隙流体、泥浆压力等地质、钻井因素的影响,使得裂缝测井响应更为复杂(董少群等,2023),如井径曲线(CAL)测量井眼的大小和形状

(Dong *et al.*, 2020a).裂缝发育段可能引起泥浆漏失或泥饼增厚,进而导致井径曲线读数变化(Tokhmechi *et al.*, 2009),但松散地层也会影响井眼延伸,导致井径曲线读数变化。伽马曲线(GR)是由于存在放射性元素(铀、钍和钾)形成的放射性记录(Dong *et al.*, 2020a)。放射性元素溶解于水中,可在裂缝中重新沉积,导致伽马曲线读数偏高(Tokhmechi *et al.*, 2009);但粘土矿物也具有天然的放射性,引起伽马曲线读数增大。声波时差曲线(AC)记录声波穿过固定长度地层所需的时间(Dong *et al.*, 2020a)。裂缝段会影响声波速度,引起声波时差曲线增大,然而高角度裂缝和直立裂缝对波速下降影响很小,可能并不会导致曲线幅度剧烈变化。密度测井(DEN)根据伽马射线散射和吸收的物理现象进而反映地层体密度(Dong *et al.*, 2020a)。裂缝段会使密度测井曲线读数减小,特别是当开启裂缝中充满流体时,然而泥浆压力高于地层压力时,钻井泥浆可能会穿透开敞裂缝,导致地层密度增加(Aghli *et al.*, 2016)。补偿中子测井(CNL)反映氢含量,在裂缝段,如果裂缝处于张开并充满流体,流体中的氢会导致补偿中子读数增大(Ja'Fari *et al.*, 2012),但孔隙流体中的氢元素也会引起曲线读数增大。在裂缝段中,浅、深电阻率测井经常会出现差异,特别是当裂缝处于张开状态和钻井泥浆侵入时。

研究区取心井段以高角度剪切缝为主(图4a、4b、4d、4e),在成像测井中表现为幅度相对较大的正弦形态(图4g),多被局部充填、半充填或全充填(图4a~4d),少数裂缝为无充填(图4e),充填物以钙质和硅质充填物为主,部分井段受多条裂缝交切的影响形成破碎带(图4f)。为研究裂缝段与非裂缝段测井曲线响应特征,根据取心段8条常规测井曲线建立交会图(图5),曲线峰的重叠和分离反映了测井曲线识别裂缝的敏感性。裂缝段CNL值分布在0~0.4 m³/m³之间,GR值分布在50~150 API之间;非裂缝段CNL值分布在0.04~0.35 m³/m³之间,GR值分布在45~220 API之间(图5a)。裂缝段AC值分布在50~83 μs/ft之间,DEN值分布在2.3~2.8 g/cm³之间;非裂缝段AC值分布在45~83 μs/ft之间,DEN值分布在2.3~2.9 g/cm³之间(图5b)。裂缝段CAL值分布在21.3~24.1 cm之间,lg(RXO)值分布在-2~5之间;非裂缝段CAL值分布在21.1~25.4 cm之间,lg(RXO)值分布在-2~6之间(图5c)。裂缝段lg(RI)值分布在2~7.3之间,lg(RT)值

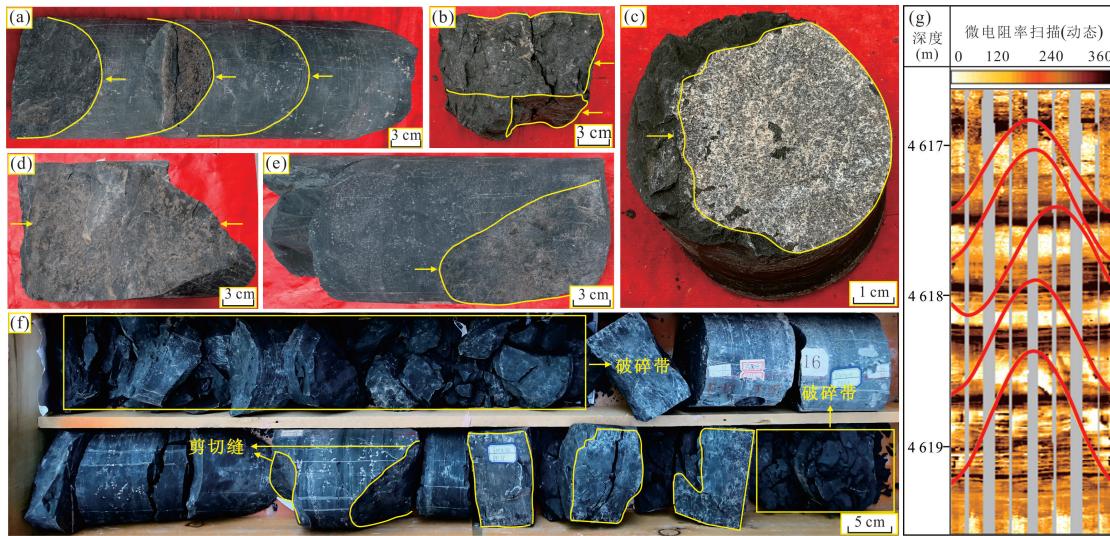


图 4 研究区风城组取心段裂缝照片

Fig.4 Photos of fractures in the core section of Fengcheng Formation in the study area

a. A1井, 4 037.01 m, 一组3条高角度剪切缝, 被方解石半充填; b. A1井, 4 026.28 m, 两组高角度剪切缝, 被方解石局部充填; c. A1井, 4 027.73 m, 低角度剪切缝被方解石全充填; d. A1井, 4 031.04 m, 高角度剪切缝被方解石半充填; e. A1井, 4 036.88 m, 高角度剪切缝无充填; f. A1井, 4 033.6~4 033.35 m, 破碎带与剪切缝发育段; g. A2井, 4 616.6~4 619.6 m, 成像测井中的高角度剪切裂缝

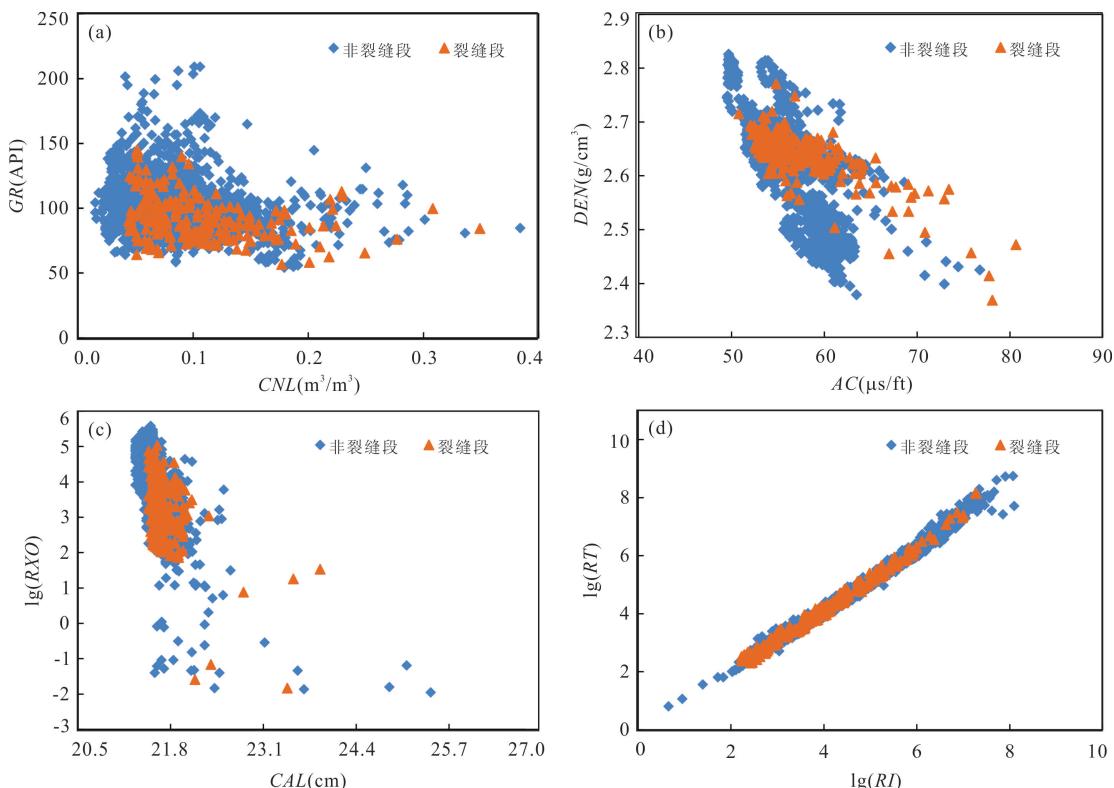


图 5 裂缝段与非裂缝段测井曲线交会图

Fig.5 Intersection diagram of logging curves of fractured section and non-fractured section

分布在2~8.5之间;非裂缝段 $\lg(RI)$ 值分布在0.5~8.2之间, $\lg(RT)$ 值分布在-0.5~9.1之间(图5d)。以上结果表明,研究区页岩油储层裂缝段相比非裂缝段 GR 值偏小、 AC 值偏大,电阻率值偏小,但裂缝

段与非裂缝段在测井曲线交会图中依然存在相当大的重叠区域,仅通过测井曲线交会图难以区分裂缝段与非裂缝段,需利用极端梯度提升树建立裂缝智能识别模型以达到预期效果。

3.2 裂缝智能识别模型特征优选

选取研究区 A1 井、A2 井、A3 井风城组共 4 162 个样本,标签取自岩心描述、成像测井解释结果。选取声波时差(AC)、密度测井(DEN)、中子孔隙度(CNL)、自然伽马(GR)、井径(CAL)、原状地层电阻率(RT)、侵入带电阻率(RI)、冲洗带电阻率(RXO)共 8 条测井曲线作为模型输入数据。由于裂缝发育程度受岩石脆性控制,因此,基于已有测井资料,使用 Techlog 软件解释得到岩石脆性曲线,与其余常规测井曲线作为训练数据,并采用孤立森林方法(Isolation Forest)检测去除数据中的异常点;经标准化处理后,随机选取 70% 共 2 913 个样本点作为训练集,并利用 SMOTE 算法进行过采样处理以平衡不同标签样本数量之间的差距;其余 30% 共 1 249 个样本点作为测试集,用于检验模型分类性能。

由于不同曲线对于裂缝的敏感性不同,需要通过特征优选方法,选择出对于裂缝识别贡献度大,对裂缝响应更敏感的特征,以提高模型的可靠性。特征优选的过程分为两步:首先,明确特征贡献度排序;其次,明确优选特征的个数。

极端梯度提升树这类集成学习方法的特点是可以通过对其中每种特征的权重进行分析,进而获得每个特征在决策树上的贡献度,并进行大小对比从而获得特征贡献排序。由于极端梯度提升树中单个决策树通过每个特征分裂点改进性能度量的量来计算特征重要性,也就说一个特征对分裂点改进性能度量越大,即越靠近根节点,权值越大,被越多提升树所选择,特征越重要。

将训练集 9 条曲线,即 9 个特征,作为极端梯度提升树的输入数据进行特征优选。特征贡献度排序结果表明,岩石脆性对裂缝识别模型的贡献度最大,其次为 RXO、GR 和 AC,贡献度最小的为 RI 和 RT(图 6)。在此基础上,将训练集再次按照 7:3 进行随机划分,根据贡献度排序,依次选取 70% 数据的 1~9 个特征进行模型训练,其余数据以测试模型训练效果。结果表明,当去除 1 个贡献度最低的特征时,分类准确率最高,而随着特征的减少,信息流失的越多,分类准确率反而降低(图 7)。根据以上测试结果,最终选择将数据集中的 RT 去除并进行裂缝识别模型的构建。

3.3 裂缝智能识别模型超参数调优

由于极端梯度提升树属于有监督学习,存在偏差和方差,交叉验证可有效降低方差和偏

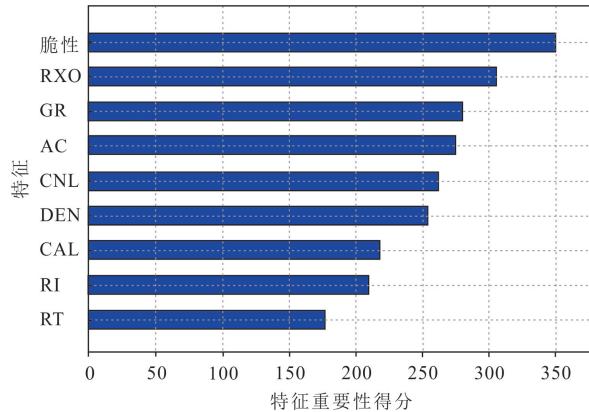


图 6 特征贡献度排序

Fig. 6 Ranking of feature contribution

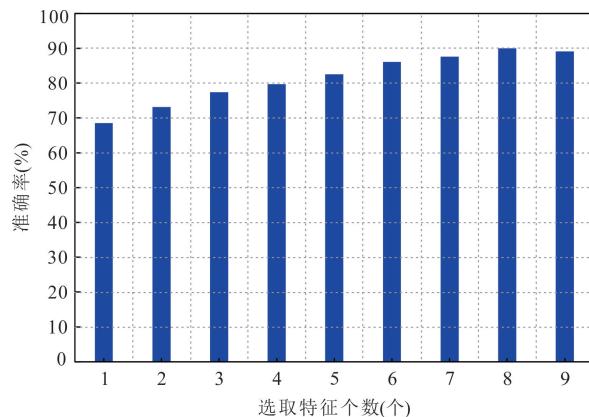


图 7 特征选择个数与分类准确率的关系

Fig. 7 Relationship between the number of feature selections and classification accuracy

差,并消除当前数据划分方式对模型输出结果造成的影响。本文采用 5 折交叉验证,即将原始训练样本随机分成 5 个大小相同的样本数据集,交叉验证过程重复了 5 次,每次选取 1 个样本数据集作为验证数据,其余为训练数据;然后,对整个过程中的 5 个结果计算平均。

在模型训练过程中,对机器学习分类器进行超参数调优可使模型获得更好的分类能力(杨笑等,2019),超参数调优有经验法和自动调参法。由于极端梯度提升树有多个超参数,分别对每个超参数进行调优不一定使模型达到最好分类效果,而对多个超参数同时进行调优才能找出所有超参数的全局最优解(Xie et al., 2018)。经验法不仅对操作者要求较高,同时仅通过人为经验寻找全局最优解难度过大,因此,本文采用自动调参法中的网格搜索方法以寻找超参数的最优取值;基本思想是将待优化的超参数在一定的空间范围中划

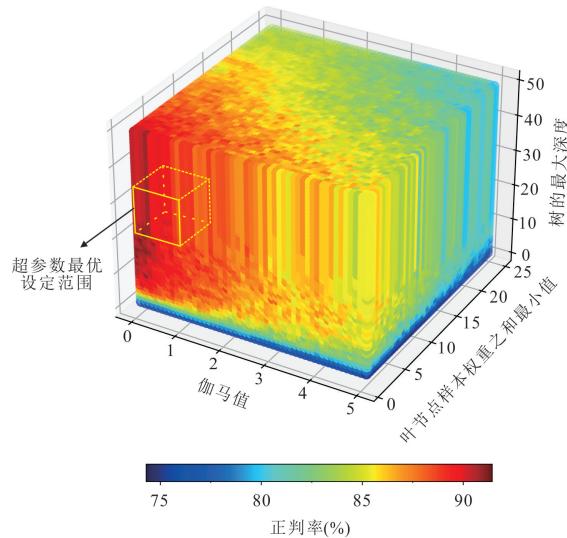


图 8 极端梯度提升树超参数网格搜索过程

Fig.8 XGBoost hyperparameter grid search process

分成网格,通过遍历网格中所有的交点寻找超参数的最优解(Dong *et al.*, 2016).

网格搜索过程选取步距较小的网格对伽马值、树的最大深度、叶节点样本权重之和最小值3个可决定模型中叶子节点分裂条件和模型复杂度的超参数进行调优.其中,伽马值代表叶子节点分裂时所需要的最小的损失下降值,这个值越大,叶子节点越难分裂;树的最大深度代表决策树的最大深度,值越大模型复杂度越高;叶节点样本权重之和最小值代表最小的叶子结点权重,如果在一次分裂中,叶子节点上所有样本的权重小于该值则停止分裂.由图8可知,伽马值、叶节点样本权重之和最小值的增大会导致模型分类性能的下降;而树的最大深度过小则会导致模型欠拟合.模型的最优超参数集如表1所示,伽马值、树的最大深度和叶节点样本权重之和最小值全局最优解分别是0.2、27.1.

3.4 裂缝智能识别模型评价

在模型调优的基础上,需建立统一的裂缝识别

表 1 极端梯度提升树超参数选取及最优设定

Table 1 Hyperparameter selection and optimal setting table of XGBoost

分类器	超参数	范围	最优设定
极端梯度提升树	伽马值	0~5	0.2
	树的最大深度	1~50	27.0
	叶节点样本权重之和最小值	1~25	1.0

评价指标,以便评价模型的裂缝识别能力.由于裂缝识别为二分类问题,即将输入数据分为裂缝和非裂缝两类,利用二分类数据混淆矩阵对模型进行评价(表2).表2中TP、FP、FN、TN分别代表将实际为裂缝的样本识别为裂缝、实际为非裂缝的样本识别为裂缝、实际为裂缝的样本识别为非裂缝、实际

表 2 二分类数据混淆矩阵

Table 2 Confusion matrix for binary data

		识别结果	
		正类	负类
正类	TP	FN	
	FP	TN	

为非裂缝的样本识别为非裂缝.

评价裂缝识别模型可以用准确率(A_c)、精确率(P_r)、召回率(R_e)和 f_1 得分(f_1)等参数进行定量评价.其中,准确率(A_c)为正确分类的样本数与总样本数之比:

$$A_c = (TP + TN) / (TP + TN + FP + FN), \quad (1)$$

准确率(A_c)反映该模型所有样本的综合分类性能.对于裂缝识别模型,准确率并不能完全代表分类器识别裂缝的能力,裂缝识别相比岩性分类更注重于将实际裂缝样本识别正确的能力,而非全局准确率,若准确率的贡献主要来源于较高的非裂缝样本识别率,而实际为裂缝的样本识别率很低,则不能说明分类器裂缝识别效果好.

精确率(P_r)用于定义识别正确的裂缝样本数与识别为裂缝的样本总数的比率:

$$P_r = TP / (TP + FP), \quad (2)$$

精确率越高,说明将非裂缝样本误识别为裂缝的样本越少.

召回率(R_e)用于定义识别正确的裂缝样本数与实际裂缝样本数的比率:

$$R_e = TP / (TP + FN), \quad (3)$$

召回率越高,说明将裂缝样本误识别为非裂缝的样本越少.

f_1 得分(f_1)同时兼顾了分类模型的精确率和召回率,是模型精确率和召回率的一种调和平均:

$$f_1 = 2TP / (2TP + FP + FN), \quad (4)$$

f_1 得分越高,说明将裂缝、非裂缝样本错误识别的总数越少,裂缝的识别准确率就越高.

利用准确率(A_c)、精确率(P_r)、召回率(R_e)和 f_1

表3 极端梯度提升树、支持向量机、逻辑回归测试集准确率、精确率、召回率和 f_1 得分统计
Table 3 XGBoost, SVM, Logistic Regression test set accuracy, Precision, Recall and f_1 score statistics

分类器	准确率(A_c)	精确率(P_r)	召回率(R_e)	f_1 得分(f_1)	种类
极端梯度提升树	90%	93%	93%	93%	0(非裂缝)
		84%	84%	84%	1(裂缝)
支持向量机	87%	88%	94%	91%	0(非裂缝)
		83%	69%	76%	1(裂缝)
逻辑回归	68%	82%	71%	76%	0(非裂缝)
		48%	63%	54%	1(裂缝)

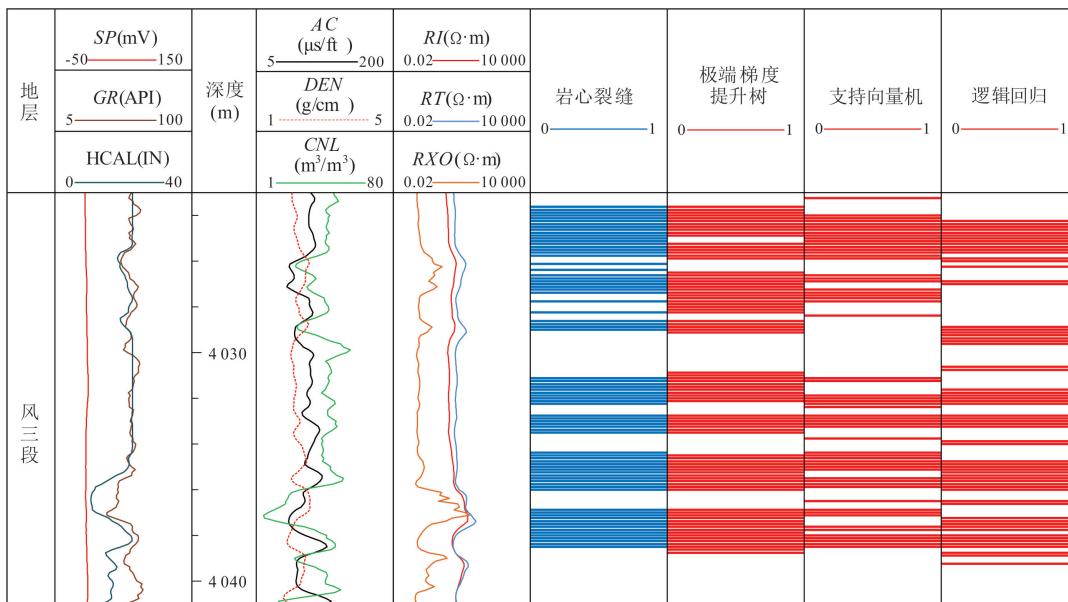


图9 A1井极端梯度提升树裂缝智能识别结果与其他两种方法对比

Fig.9 Comparison of the intelligent fracture identification results of XGBoost in Well A1 and the other two methods

得分(f_1)可以定量评估极度梯度分类器在测试集中的裂缝识别能力,反映了所建立的裂缝识别模型对研究区陆相页岩油储层裂缝识别的可靠性.

3.5 裂缝智能识别效果检验

利用超参数调优后的模型对测试集数据进行裂缝识别,测试集为A1、A2和A3井中除训练集外剩余的1 249个样本. 极端梯度提升树测试集准确率为90%,非裂缝样本识别精确率、召回率和 f_1 得分均为93%,裂缝样本识别精确率、召回率和 f_1 得分均为84%(表3). 其中,准确率较高,说明极度梯度提升树总体分类效果较好,非裂缝样本识别的精确度、召回率、 f_1 得分达90%以上,说明该方法对于非裂缝段的识别精度较高,将裂缝识别为非裂缝、非裂缝识别为裂缝的错误识别情况较少. 裂缝样本识别的精确度、召回率、 f_1 得分均低于非裂缝段,这是由于裂缝段测井响应弱且复杂、以及较少的样本数

量造成的,虽然识别率低于非裂缝段,但3个评价指标均超过80%,识别效果相对理想(表3).

3.6 极端梯度提升树与其他方法识别效果对比

支持向量机和逻辑回归等方法是目前测井裂缝识别常用的机器学习方法(石广仁,2008; Dong et al., 2020a). 利用相同的数据集训练支持向量机和逻辑回归,支持向量机识别准确率为87%,非裂缝样本识别精确率、召回率、 f_1 得分为88%、94%、91%,裂缝样本识别精确率、召回率、 f_1 得分为83%、69%、76%(表3);逻辑回归识别准确率为68%,非裂缝样本识别精确率、召回率、 f_1 得分为82%、71%、76%,裂缝样本识别精确率、召回率、 f_1 得分为48%、63%、54%(表3).

从准确率来看,极度梯度提升树略高于支持向量机,但二者差异较小,逻辑回归则与前两个分类器具有较大差距. 而裂缝识别更应关注裂缝样本的

识别精度,极端梯度提升树裂缝样本识别精确度、召回率和 f_1 得分为 3 个分类器中最高(表 3),说明该分类器裂缝识别精度最高;支持向量机裂缝样本精确率比极端梯度提升树略低,二者差距不大;召回率和 f_1 得分却远低于极端梯度提升树,说明支持向量机相比极端梯度提升树更容易将大量的裂缝误识别为非裂缝(表 3);逻辑回归裂缝样本精确率、召回率、 f_1 得分均远低于其他两个分类器,说明该分类器对于裂缝和非裂缝识别效果都不好(表 3)。

根据 A1 井裂缝识别结果可知,极端梯度提升树裂缝识别结果与岩心裂缝标定结果吻合程度较高,但在 4 024.94~4 025.24 m、4 032.14~4 032.28 m 等井段出现有裂缝未能识别,在 4 027.8~4 028.2 m、4 029.0~4 029.2 m 等井段出现无裂缝但错误识别(图 9)。支持向量机可识别出的裂缝与岩心裂缝符合率较高,但出现了大量裂缝不可识别的情况,这与表 3 中所得到的裂缝样本的高精确率、低召回率和低 f_1 得分的结果相符合。逻辑回归在后验井段出现了大量有裂缝未能识别和无裂缝却错误识别的现象,与表 3 中得到的裂缝样本识别的低精确率、低召回率和低 f_1 得分特征相符合。

4 讨论

4.1 噪声干扰问题

由于极端梯度提升树的原理基于对错误点的纠正,而噪声即异常点的存在,很容易让模型误判为错误分类的点,逐轮提升时会放大误差的干扰,让模型分类性能变差。针对该问题,本文提出的解决方案可分为两步:首先,对数据集进行异常点检测,并去除异常点,这样可以减少噪声干扰;其次,对特征进行重要性排序和特征优选,特征越多,所携带的噪音可能越多,而删除部分贡献度较低的特征,也可有效消除噪音干扰。

本文基于孤立森林(Isolation Forest)方法进行异常点检测和去除,该方法利用隔离的思想使用垂直于坐标轴的随机超平面对数据空间进行切割从而“孤立”离群点,与基于密度和聚类的方法先相比其避免了计算距离、密度,所以算法在时间开销上更有优势,同时也减少了异常的掩盖和淹没效应(周杭和蒋瑜,2023);通过孤立森林方法,从原始数据中共筛查出 209 个异常点,其中分数小于 0 代表样本点异常(图 10)。在此基础上,本文也进行了特征贡献度排序(图 6,

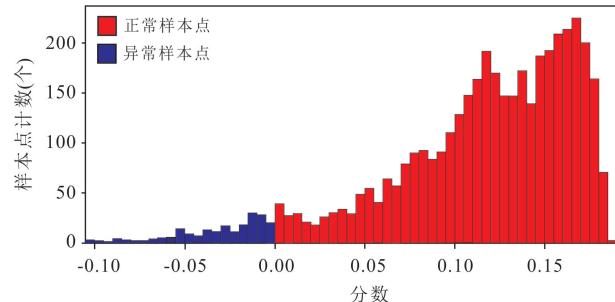


图 10 孤立森林异常点检测结果直方图
Fig. 10 Histogram of detection results of isolated forest for abnormal data points

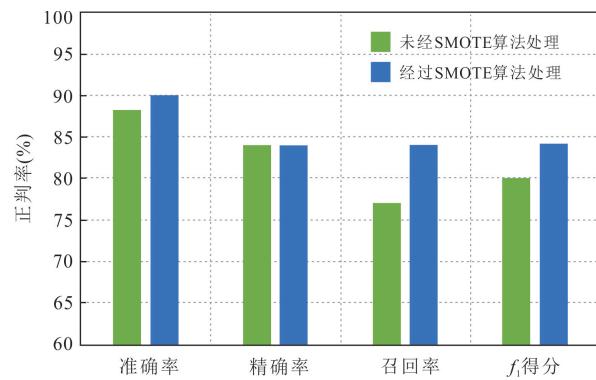


图 11 极端梯度提升树原始数据与过采样数据识别结果对比
Fig. 11 Comparison of recognition results between original data and oversampled data using XGBoost

图 7),并去除了不重要特征,进一步减少了噪音干扰,提高了模型的分类性能和可靠性。

4.2 数据不均衡对裂缝智能识别效果的影响

研究区陆相页岩油储层裂缝较为发育,但岩心数据和测井数据依旧存在非裂缝标签比裂缝标签多的问题,比值大约为 5:2,因此裂缝识别为典型的数据不均衡分类问题。由于非裂缝标记占主导地位,使用基于简单的精度评估(如最小化均方误差)的传统拟合方法,往往过度预测主要特征。即使评估准确率很高,其贡献主要来源于非裂缝样本,次要特征(裂缝标记)往往会被低估。因此,本文在数据预处理阶段利用 SMOTE 算法,对裂缝数据进行过采样,使裂缝标签数量达到非裂缝标签数量的 0.8 倍,以此来缓解由数据不均衡导致的次要特征被低估的问题。经对比,基于原始数据训练的模型与经过 SMOTE 处理数据训练的模型测试集准确率为 88% 和 90%,精确率都为 84%,召回率为 77% 和 84%, f_1 得分为 80% 和 84%(图 11, 图 12)。经 SMOTE 算法处理过的数据,测试集准确率提

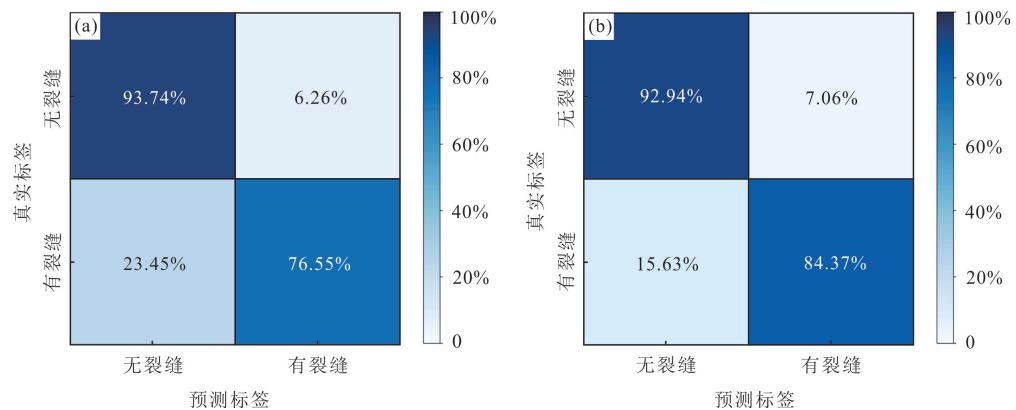


图12 极端梯度提升树原始数据与过采样数据识别结果混淆矩阵对比

Fig.12 Confusion matrix comparison between original data and oversampled data recognition results of XGBoost

a.未经SMOTE算法处理;b.经过SMOTE算法处理

高2%,召回率和 f_1 得分分别提高7%和4%,模型对裂缝的识别精度有所提高,表明裂缝识别中利用SMOTE算法处理训练数据的必要性。

5 结论

(1)针对陆相页岩油储层裂缝尺度小、测井响应弱、识别难度大等难题,本文将岩心观察和成像测井解释裂缝作为标签,应用集成学习中的极端梯度提升树,通过网格搜索获得裂缝智能识别模型的最优超参数,深度挖掘了裂缝信息与常规测井数据之间的非线性关系,有效提高了陆相页岩油储层测井裂缝的识别精度。

(2)极端梯度提升树在准噶尔盆地玛湖凹陷风城组陆相页岩油储层应用结果表明,该方法对于测试集的分类准确率可达到90%,对于A1井风3段的裂缝段与非裂缝段都具有较好的识别效果,与岩心观察结果符合率较高,且智能识别效果优于目前常用的支持向量机和逻辑回归等方法,反映了集成学习方法在测井裂缝智能识别中的优越性。

(3)基于机器学习的测井裂缝智能识别是一个非均衡数据集的二分类问题,非裂缝标签的数量远远大于裂缝标签,经SMOTE算法处理后,有效提高了训练集中裂缝标签的数量,结果表明经过采样数据集训练模型的准确率、召回率和 f_1 得分都明显高于原始数据集。

(4)极端梯度提升树可对准噶尔盆地玛湖凹陷风城组陆相页岩油储层提供一种有效的裂缝识别手段,但该方法也存在对噪声较为敏感,逐轮提升时会放大误差的干扰的缺

点,需配合数据清洗和特征提取等手段以减小噪声对模型所产生的负面影响。

References

- Aghli, G., Soleimani, B., Moussavi-Harami, R., et al., 2016. Fractured Zones Detection Using Conventional Petrophysical Logs by Differentiation Method and Its Correlation with Image Logs. *Journal of Petroleum Science and Engineering*, 142: 152–162. <https://doi.org/10.1016/j.petrol.2016.02.002>
- Bao, M.Y., Dong, S.Q., Zeng, L.B., et al., 2023. Artificial Intelligence Prediction Method for Tight Carbonate Reservoir Fracture Distribution Based on Seismic Attributes. *Earth Science* (in Chinese with English abstract).
- Cao, J., Lei, D. W., Li, Y. W., et al., 2015. Ancient High-Quality Alkaline Lacustrine Source Rocks Discovered in the Lower Permian Fengcheng Formation, Junggar Basin. *Acta Petrolei Sinica*, 36(7): 781–790 (in Chinese with English abstract).
- Chen, G. H., Hu, C., Zeng, Y. L., et al., 2015. Logging Identification Method of Fillings in Fractures and Caverns in Carbonate Reservoir Based on BP Neural Network. *Geophysical Prospecting for Petroleum*, 54(1): 99–104 (in Chinese with English abstract).
- Dong, S. Q., Wang, Z. Z., Zeng, L. B., 2016. Lithology Identification Using Kernel Fisher Discriminant Analysis with Well Logs. *Journal of Petroleum Science and Engineering*, 143: 95–102. <https://doi.org/10.1016/j.petrol.2016.02.017>
- Dong, S. Q., Zeng, L. B., Lyu, W. Y., et al., 2020a. Fracture Identification by Semi-Supervised Learning Using Conventional Logs in Tight Sandstones of Ordos Basin, China. *Journal of Natural Gas Science and Engineering*

- ing, 76: 103131. <https://doi.org/10.1016/j.jngse.2019.103131>
- Dong, S. Q., Zeng, L. B., Lyu, W. Y., et al., 2020b. Fracture Identification and Evaluation Using Conventional Logs in Tight Sandstones: A Case Study in the Ordos Basin, China. *Energy Geoscience*, 1(3–4): 115–123. <https://doi.org/10.1016/j.engeos.2020.06.003>
- Dong, S.Q., Zeng, L.B., Che, X.H., et al., 2023. Application of Artificial Intelligence in Fracture Identification Using Well Logs in Tight Reservoirs. *Earth Science* (in Chinese with English abstract).
- Fei, S. P., Hassan, M. A., He, Z. H., et al., 2021. Assessment of Ensemble Learning to Predict Wheat Grain Yield Based on UAV-Multispectral Reflectance. *Remote Sensing*, 13(12): 2338. <https://doi.org/10.3390/rs13122338>
- Ghosh, S., Galvis-Portilla, H. A., Klockow, C. M., et al., 2018. An Application of Outcrop Analogues to Understanding the Origin and Abundance of Natural Fractures in the Woodford Shale. *Journal of Petroleum Science and Engineering*, 164: 623–639. <https://doi.org/10.1016/j.petrol.2017.11.073>
- Gong, L., Wang, J., Gao, S., et al., 2021. Characterization, Controlling Factors and Evolution of Fracture Effectiveness in Shale Oil Reservoirs. *Journal of Petroleum Science and Engineering*, 203: 108655. <https://doi.org/10.1016/j.petrol.2021.108655>
- Gu, Y. F., Zhang, D. Y., Lin, Y. B., et al., 2021. Data-Driven Lithology Prediction for Tight Sandstone Reservoirs Based on New Ensemble Learning of Conventional Logs: A Demonstration of a Yanchang Member, Ordos Basin. *Journal of Petroleum Science and Engineering*, 207: 109292. <https://doi.org/10.1016/j.petrol.2021.109292>
- Ja'Fari, A., Kadkhodaie-Iikhchi, A., Sharghi, Y., et al., 2012. Fracture Density Estimation from Petrophysical Log Data Using the Adaptive Neuro-Fuzzy Inference System. *Journal of Geophysics and Engineering*, 9(1): 105–114. <https://doi.org/10.1088/1742-2132/9/1/013>
- Jin, Z. J., Zhu, R. K., Liang, X. P., et al., 2021. Several Issues Worthy of Attention in Current Lacustrine Shale Oil Exploration and Development. *Petroleum Exploration and Development*, 48(6): 1276–1287 (in Chinese with English abstract).
- Lei, D. W., Chen, G. Q., Liu, H. L., et al., 2017. Study on the Forming Conditions and Exploration Fields of the Mu Hu Giant Oil (Gas) Province, Junggar Basin. *Acta Geologica Sinica*, 91(7): 1604–1619 (in Chinese with English abstract).
- Li, T. Y., Wang, R. H., Wang, Z. Z., et al., 2018. Prediction of Fracture Density Using Genetic Algorithm Support Vector Machine Based on Acoustic Logging Data. *Geophysics*, 83(2): D49–D60. <https://doi.org/10.1190/geo2017-0229.1>
- Liu, G. P., Zeng, L. B., Sun, G. Q., et al., 2020a. Natural Fractures in Tight Gas Volcanic Reservoirs and Their Influences on Production in the Xujiaweizi Depression, Songliao Basin, China. *AAPG Bulletin*, 104(10): 2099–2123. <https://doi.org/10.1306/05122017169>
- Liu, G. P., Zeng, L. B., Wang, X. J., et al., 2020b. Natural Fractures in Deep Tight Gas Sandstone Reservoirs in the Thrust Belt of the Southern Junggar Basin, Northwestern China. *Interpretation*, 8(4): SP81–SP93. <https://doi.org/10.1190/int-2020-0051.1>
- Luo, G., Xiao, L. Z., Shi, Y. Q., et al., 2022. Machine Learning for Reservoir Fluid Identification with Logs. *Petroleum Science Bulletin*, 7(1): 24–33 (in Chinese with English abstract).
- Shi, G.R., 2008. Superiorities of Support Vector Machine in Fracture Prediction and Gassiness Evaluation. *Petroleum Exploration and Development*, 35(5): 588–594 (in Chinese with English abstract).
- Song, S.H., Shi, Y.Q., Hou, J.G., 2022. Review of a Generative Adversarial Networks (GANs)-Based Geomodeling Method. *Petroleum Science Bulletin*, 7(1): 34–49
- Song, X. Z., Yao, X. Z., Li, G. S., et al., 2022. A Novel Method to Calculate Formation Pressure Based on the LSTM-BP Neural Network. *Petroleum Science Bulletin*, 7(1): 12–23 (in Chinese with English abstract).
- Tang, Y., Guo, W.J., Wang, X.T., et al., 2019. A New Breakthrough in Exploration of Large Conglomerate Oil Province in Mu Hu Sag and Its Implications. *Xinjiang Petroleum Geology*, 40(2): 127–137 (in Chinese with English abstract).
- Tokhmechi, B., Memarian, H., Noubari, H. A., et al., 2009. A Novel Approach Proposed for Fractured Zone Detection Using Petrophysical Logs. *Journal of Geophysics and Engineering*, 6(4): 365–373. <https://doi.org/10.1088/1742-2132/6/4/004>
- Torlay, L., Perrone-Bertolotti, M., Thomas, E., et al., 2017. Machine Learning-XGBoost Analysis of Language Networks to Classify Patients with Epilepsy. *Brain Informatics*, 4(3): 159–169. <https://doi.org/10.1007/s40708-017-0065-7>
- Xie, Y. X., Zhu, C. Y., Zhou, W., et al., 2018. Evaluation of Machine Learning Methods for Formation Lithology

- Identification: A Comparison of Tuning Processes and Model Performances. *Journal of Petroleum Science and Engineering*, 160: 182—193. <https://doi.org/10.1016/j.petrol.2017.10.028>
- Xue, Y. C., Cheng, L. S., Mou, J. Y., et al., 2014. A New Fracture Prediction Method by Combining Genetic Algorithm with Neural Network in Low-Permeability Reservoirs. *Journal of Petroleum Science and Engineering*, 121: 159—166. <https://doi.org/10.1016/j.petrol.2014.06.033>
- Yang, X., Wang, Z. Z., Zhou, Z. Y., et al., 2019. Lithology Classification of Acidic Volcanic Rocks Based on Parameter-Optimized AdaBoost Algorithm. *Acta Petrolei Sinica*, 40(4): 457—467 (in Chinese with English abstract).
- Zeng, L. B., 2010. Microfracturing in the Upper Triassic Sichuan Basin Tight-Gas Sandstones: Tectonic, Overpressure, and Diagenetic Origins. *AAPG Bulletin*, 94(12): 1811—1825. <https://doi.org/10.1306/06301009191>
- Zeng, L. B., Jiang, J. W., Yang, Y. L., 2010. Fractures in the Low Porosity and Ultra-Low Permeability Glutinite Reservoirs: A Case Study of the Late Eocene Hetaoyuan Formation in the Anpeng Oilfield, Nanxiang Basin, China. *Marine and Petroleum Geology*, 27(7): 1642—1650. <https://doi.org/10.1016/j.marpetgeo.2010.03.009>
- Zeng, L. B., Lyu, W. Y., Li, J. A., et al., 2016. Natural Fractures and Their Influence on Shale Gas Enrichment in Sichuan Basin, China. *Journal of Natural Gas Science and Engineering*, 30: 1—9. <https://doi.org/10.1016/j.jngse.2015.11.048>
- Zeng, L. B., Lyu, W. Y., Zhang, Y. Z., et al., 2021. The Effect of Multi-Scale Faults and Fractures on Oil Enrichment and Production in Tight Sandstone Reservoirs: A Case Study in the Southwestern Ordos Basin, China. *Frontiers in Earth Science*, 9: 1—12. <https://doi.org/10.3389/feart.2021.664629>
- Zhi, D. M., Song, Y., He, W. J., et al., 2019. Geological Characteristics, Resource Potential and Exploration Direction of Shale Oil in Middle-Lower Permian, Junggar Basin. *Xinjiang Petroleum Geology*, 40(4): 389—401 (in Chinese with English abstract).
- Zhi, D. M., Tang, Y., He, W. J., et al., 2021. Orderly Co-existence and Accumulation Models of Conventional and Unconventional Hydrocarbons in Lower Permian Fengcheng Formation, Mahu Sag, Junggar Basin. *Petroleum Exploration and Development*, 48(1): 38—51 (in Chinese with English abstract).
- Zhou, H., Jiang, Y., 2023. Improved Isolation Forest Method Based on High Contrast Subspace. *Application Research of Computers*, 40(2): 388—393 (in Chinese with English abstract).
- ### 附中文参考文献
- 鲍明阳, 董少群, 曾联波, 等, 2023. 基于地震属性的致密碳酸盐岩储层裂缝分布的人工智能预测方法. 地球科学.
- 曹剑, 雷德文, 李玉文, 等, 2015. 古老碱湖优质烃源岩: 准噶尔盆地下二叠统风城组. 石油学报, 36(7): 781—790.
- 陈钢花, 胡琮, 曾亚丽, 等, 2015. 基于BP神经网络的碳酸盐岩储层缝洞充填物测井识别方法. 石油物探, 54(1): 99—104
- 董少群, 曾联波, 车小花, 等, 2023. 人工智能在致密储层裂缝测井识别中的应用. 地球科学.
- 金之钩, 朱如凯, 梁新平, 等, 2021. 当前陆相页岩油勘探开发值得关注的几个问题. 石油勘探与开发, 48(6): 1276—1287.
- 雷德文, 陈刚强, 刘海磊, 等, 2017. 准噶尔盆地玛湖凹陷大油(气)区形成条件与勘探方向研究. 地质学报, 91(7): 1604—1619.
- 罗刚, 肖立志, 史燕青, 等, 2022. 基于机器学习的致密储层流体识别方法研究. 石油科学通报, 7(1): 24—33.
- 石广仁, 2008. 支持向量机在裂缝预测及含气性评价应用中的优越性. 石油勘探与开发, 35(5): 588—594
- 宋随宏, 史燕青, 侯加根, 2022. 基于生成对抗网络的储层地质建模方法研究进展. 石油科学通报, 7(1): 34—49
- 宋先知, 姚学喆, 李根生, 等, 2022. 基于LSTM-BP神经网络的地层孔隙压力计算方法. 石油科学通报, 7(1): 12—23.
- 唐勇, 郭文建, 王霞田, 等, 2019. 玛湖凹陷砾岩大油区勘探新突破及启示. 新疆石油地质, 40(2): 127—137
- 杨笑, 王志章, 周子勇, 等, 2019. 基于参数优化 AdaBoost 算法的酸性火山岩岩性分类. 石油学报, 40(4): 457—467.
- 支东明, 宋永, 何文军, 等, 2019. 准噶尔盆地中-下二叠统页岩油地质特征、资源潜力及勘探方向. 新疆石油地质, 40(4): 389—401
- 支东明, 唐勇, 何文军, 等, 2021. 准噶尔盆地玛湖凹陷风城组常规-非常规油气有序共生与全油气系统成藏模式. 石油勘探与开发, 48(1): 38—51
- 周杭, 蒋瑜, 2023. 基于高对比度子空间的改进孤立森林方法. 计算机应用研究, 40(2): 388—393.