基于 XGBoost 算法的走滑断裂内部特征带的精细识别

赵军[1*]，汪峻宇[1]，赖强[2]，文晓峰[3]，邬光辉[1]，焦世祥[1]

1. 西南石油大学地球科学与技术学院，成都 610500；
2. 中国石油西南油气田分公司勘探开发研究院，成都 610095；
3. 中国石油集团测井有限公司，西安 710000

* 通信作者，E-mail：zhaojun_70@126.com

2023-10-19 收稿；2024-03-13 修回；2024-03-15 接受

基金项目：中国石油–西南石油大学创新联合体项目(2020CX010204)

中图分类号：P631.81; P618.13　　文章编号：2096-8523(2025)02-0182-11　　doi: 10.19509/j.cnki.dzkq.tb20230583

摘　要：受制于走滑断裂内部强烈的非均质性、储集类型多样及流体分布复杂的影响，走滑断裂内部裂缝带、破碎带和溶蚀带之间的测井响应复杂多变，为有效利用成像及常规测井资料识别走滑断裂内部 3 特征带造成了困难。引入 XGboost 算法建立模型，以提高对走滑断裂内部 3 特征带的识别精度。分析了走滑断裂内部 3 特征带的测井响应特征，优选敏感测井曲线构建基于均值及方差的特征向量空间集，采用极端梯度提升算法，建立了走滑断裂溶蚀带、破碎带和裂缝带的 XGBoost 回归预测模型，并通过多分类评价指标对 XGBoost 模型的关键参数进行调优，提高了走滑断裂内部特征带的识别精度。利用构建的 XGBoost 模型对研究区走滑断裂内部特征带进行了识别，其中总样本数 234 个，识别正确样本 208 个，识别正确率达 88.89%；预测结果表明在走滑断裂内部特征带中，裂缝带分布范围最广，破碎带其次，溶蚀带最窄，这与实际走滑断裂内部特征带的分布范围相符。基于 XGBoost 算法的走滑断裂内部特征带识别模型能够有效地识别裂缝带、破碎带和溶蚀带，从而有助于对走滑断裂内部尺度更小的溶蚀孔洞及裂缝储集空间的分布进行更为有效的分析，对走滑断裂内部结构的精细刻画有一定借鉴意义。

关键词：走滑断裂；XGBoost 算法；碳酸盐岩；测井评价；特征带识别；四川盆地；高石梯–磨溪地区

四川盆地内部走滑断裂的形成是地壳构造叠合、板块运动、地质演化、构造单元交汇以及地震活动等多种地质和构造因素相互作用的结果。位于中国大陆西部的四川盆地，处于不同构造单元的交汇带，多次地壳构造运动引发了地壳内部的应力积累和释放。印度板块与欧亚板块的相对运动使得地壳扭曲、伸展和断裂，这促使了走滑断裂的形成[1-2]。四川盆地中部地区走滑断裂形成及演化主要经历了南华纪走滑断层萌芽阶段、晚震旦世–早寒武世右行张扭走滑阶段、奥陶纪–二叠纪前弱压扭走滑阶段、晚二叠世左行弱张扭走滑阶段 4 个阶段[3]。南华纪走滑断裂阶段，四川盆地西北缘经历了新元古代的火成岩发育，可能卷入罗迪尼亚超大陆的拼合和裂解[4-6]。华南板块位于活动大陆边缘环境，与环罗迪尼亚俯冲系统相符[7-10]。四川盆地在新元古代经历了来自原特提斯洋的前展–后撤俯冲作用，可能导致 NW 向板内走滑断裂形成。随后，板块俯冲后撤形成伸展背景，板内沿岩浆带裂解，形成南华纪裂谷体系[11-14]。深部地球物理资料揭示了深大基底断裂体系，为后期走滑断裂奠定了基础。晚震旦世–早寒武世右行张扭走滑阶段，走滑断裂分布与德阳–安岳“槽盆”演化相关。“槽盆”可能形成于拉张环境，不是典型裂谷。在快速发展期，“槽盆”沉积了泥页岩和张扭构造，与走滑断裂活动一致[15-17]。断裂胶结物 U-Pb 测

著录信息：赵军，汪峻宇，赖强，等. 基于 XGBoost 算法的走滑断裂内部特征带的精细识别[J]. 地质科技通报，2025，44(2)：182-192.
ZHAO Jun, WANG Junyu, LAI Qiang, et al. Fine-grained identification of internal characteristic zones within strike-slip faults via the XGBoost algorithm[J]. Bulletin of Geological Science and Technology, 2025, 44(2): 182-192.

© Editorial Office of Bulletin of Geological Science and Technology. This is an open access article under the CC BY-NC-ND license.　https://dzkjqb.cug.edu.cn

年显示了与早寒武世“槽盆”沉积匹配的活动。奥陶纪–二叠纪前弱压扭走滑阶段，四川盆地受控于 SN 向挤压和 EW 向伸展，形成槽盆–古隆起构造。走滑断裂形成于应力状态下，可能在二叠系底下复活[18-19]。奥陶纪流体活动和前期断裂活动被检测到。晚二叠世左行弱张扭走滑阶段，区域应力场转为伸展，峨眉地裂运动形成盆地构造格局[20]。走滑断裂可能复活，但变形相对较弱。在后期构造运动中，塑性层可能调节了浅层走滑断裂的形成。总体而言，走滑断裂的形成受到多重因素综合影响，包括板块俯冲、后撤和拉张。走滑断裂在油气勘探领域扮演着关键的角色，它们不仅充当了优质的油气运移通道，迅速将深部烃源岩中的油气输送至上方储层，促进了油气资源的高效传递，还对储层的地质条件产生了显著影响，包括孔隙结构和渗透性的改变[21-22]。在某些情况下，走滑断裂还能引发储层微裂缝的形成，进一步提供了储藏和运移通道，有利于油气的积聚。此外，通常在走滑断裂周围形成了有利于油气富集的聚集区，这是因为断裂改变了地下构造的均匀性，导致地下压力和温度的差异[23-24]。综上所述，走滑断裂对于油气资源的传输、富集和分布具有至关重要的影响，因此，识别和研究走滑断裂在油气勘探和开发中至关重要。

走滑断裂在纵向上可沟通多套生储盖组合且对油气的运聚和圈闭的形成具有重要的控制作用，并对深埋条件下储层的形成和改造作用显著，因此已逐渐成为油气地质学研究的热点[25]。目前，研究人员结合地质与地震等资料对走滑断裂展开了一系列研究，基本掌握了走滑断裂内部大型孔洞的分布规律，但受到地震资料分辨率的影响，未能充分掌握走滑断裂内部部分尺度较小的溶蚀孔洞和裂缝发育区域的分布规律[26]。测井资料可以提供较高的分辨率，对于尺度较小的溶蚀孔洞和裂缝发育区域的分布规律分析较为有效，因此，基于测井资料来分析溶蚀带、破碎带、裂缝带的特征，并可以以此为基础来解析走滑断裂的结构。针对走滑断裂“三带”的识别主要通过基于成像测井标定常规测井资料在单井上进一步识别出溶蚀带、破碎带与裂缝带，通过总结单井的溶蚀带、破碎带与裂缝带在常规测井数据上的差异，建立溶蚀带、破碎带、裂缝带的识别标准，进而完成整个研究区走滑断裂“三带”的识别[27]。这种方法可以在一定程度上完成对走滑断裂“三带”的识别，但受制于走滑断裂内部强烈的非均质性、储集空间多变及流体分布复杂的影响，加大了测井定性解释的多解性而造成识别精度不高[28]。

XGBoost（eXtreme gradient boosting）是一种基于决策树的机器学习算法，是一种集成学习方法，通过多次迭代训练决策树模型并将其组合起来以提高预测的准确性[29-31]。相比于结合成像测井与常规测井识别走滑断裂的方法，机器学习算法可以自动化地处理海量的数据，快速地挖掘数据中的信息，避免了传统方法中需要手动处理数据和解释的问题，同时也能够发现更为微妙和复杂的地质特征和关系[32-33]。XGBoost 作为一种性能卓越的机器学习算法，其在多个领域的应用已经得到了广泛认可，在目前众多的研究中，XGBoost 算法已被应用于岩性识别、储层划分、煤体结构识别等方面，并取得了良好的效果[34-36]。因此，利用 XGBoost 算法来识别溶蚀带、破碎带和裂缝带，将有助于提高测井识别的准确率。

1 地质背景及测井响应特征

1.1 走滑断裂“三带”的基本特征

川中高石梯–磨溪地区位于四川盆地中部的川中平缓构造带的中南部，处于近 SN 向展布的德阳–安岳克拉通内裂陷槽的中东部，总面积达到 2.7 万 km^2。在地质历史上，该地区经历了多期构造演化，包括扬子期、加里东期、海西期、印支期、燕山期和喜马拉雅期。特别是在加里东期的早期阶段，桐湾运动表现出多次活动，对于震旦纪到早寒武世的构造演化、储层发育、油气生成、运移富集以及后续的调整和改造产生了重要影响[37]。另外，由桐湾运动形成的高石梯–磨溪古隆起，在不同构造背景下形成的多期走滑断裂构造的影响下，对于安岳震旦系–寒武系的特大型气田的形成和分布具有重要控制作用[38]，进一步在川中地区形成了现今复杂的多层系多期次天然气富集格局。如图 1 所示，川中高石梯–磨溪地区发育有 $F_{Ⅰ}5$、$F_{Ⅰ}6$、$F_{Ⅰ}7$ 等共 8 条 EW 向一级走滑断裂，$F_{Ⅱ}8$、$F_{Ⅱ}9$、$F_{Ⅱ}10$ 等共 10 条 EW 向二级走滑断裂，$F_{Ⅱ}19$、$F_{Ⅱ}20$、$F_{Ⅱ}21$ 和 $F_{Ⅱ}24$ 共 4 条 NE 向二级走滑断裂，以及数条 EW 向、NE 向三级走滑断裂和四级走滑断裂。

走滑断裂内部是一个具有极其复杂地质结构的三维地质体，针对走滑断裂内部不同区域的裂缝及溶蚀程度之间的差异，将走滑断裂内部分为溶蚀带、破碎带和裂缝带 3 个特征带。走滑断裂“三带”在平面上呈现为溶蚀带靠近走滑断裂核心区域，其次为破碎带，裂缝带则处于远离走滑断裂的位置，如图 2 所示。溶蚀带：位于断裂核部靠近断裂中心的区域，

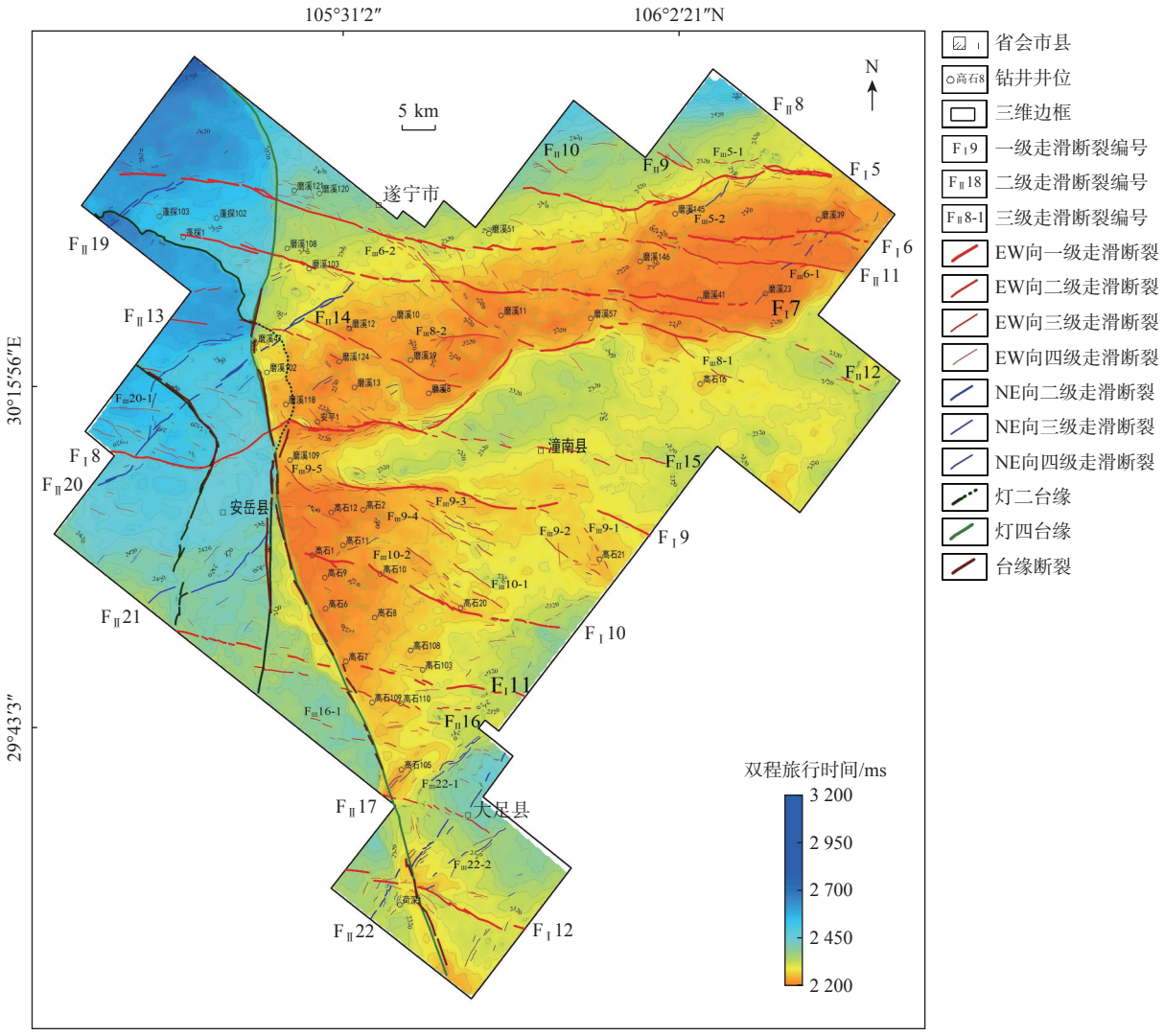


图 1　川中高石梯–磨溪地区走滑断裂展布图

Fig. 1　Distribution of strike-slip faults in the central part of Sichuan Gaoshiti-Moxi region

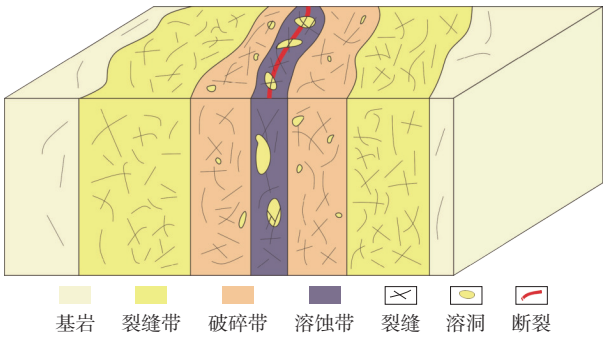


图 2　走滑断裂“三带”地质分布示意图

Fig. 2　Geological distribution of the "Three Zones" of strike-slip fault

地层大量破碎后受淡水的溶蚀作用形成大规模的溶孔和溶洞，孔渗在 3 个特征带中最好。破碎带：位于离断裂相对较近的区域，地层为断裂角砾岩和碎裂岩等断裂岩混杂带。破碎带受溶蚀作用的影响较小，内部溶蚀孔洞规模小于破碎带且发育大量裂缝，岩石往往成碎块状，孔渗较好。裂缝带：位于离断裂相对较远的区域，受溶蚀作用的影响很小，一般不发育溶蚀孔洞或发育极少量溶孔，地层内由多个不同方位且不同性质的裂缝错乱分布从而形成纵横交错的裂缝网络，孔渗较好[39-40]。

1.2　测井响应特征

走滑断裂“三带”在平面上的分布规模较大，直井内部通常仅能反映溶蚀带、破碎带和裂缝带中某一个特征带的测井响应特征，而水平井往往能够穿越多个特征带，相较于直井，水平井更能反映溶蚀带、破碎带和裂缝带 3 个特征带之间在常规测井响应之间的差距，更有利于完成对“三带”的识别及其在平面上分布的预测。水平井 gs110 井内部分布有溶蚀带、破碎带和裂缝带 3 个特征带，如图 3 所示。该井中溶蚀带的测井响应特征最为明显，该特征带存在明显的扩径伴随着自然伽马的急剧增大，声波时差

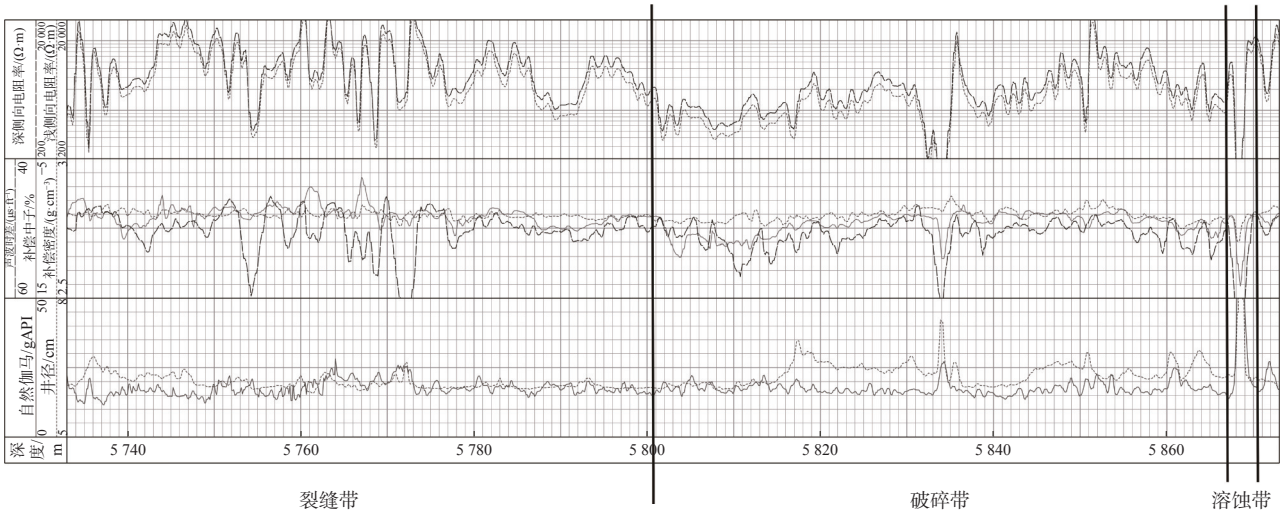


图 3　单井“三带”测井响应特征图

Fig. 3　Logging response characteristics of the "Three Zones" in a single well

和中子孔隙度呈尖刺状增大，密度降低，双侧向电阻率呈现出明显的降低趋势且其未有明显差异；破碎带测井响应特征呈现轻微扩径，自然伽马低值，声波时差及中子孔隙度相较于裂缝带略微增大，密度无明显变化，双侧向电阻率中低值且呈现明显的正差异；裂缝带中无扩径现象，自然伽马、中子孔隙度及密度均未有明显变化，声波时差总体呈现低值且偶有跳波现象，双侧向电阻率总体呈现中高值且有明显的正差异。

溶蚀带通常在常规测井中呈现出自然伽马值升高的特征，伴随着明显的扩径现象。此外，声波时差和补偿中子均呈现明显的增加，密度呈现明显的降低；在双侧向电阻率方面，呈现出明显的降低趋势，其差异通常为正差异或无差异；在成像测井图上，溶蚀带通常呈现为深色、大规模的洞穴。破碎带通常在常规测井中表现为自然伽马值较低，存在轻微扩径现象，声波时差和补偿中子增大，密度变化不大，双侧向电阻率降低且呈现中低值，差异呈现为正差异；成像测井图上显示为深色正弦条带和深色大规模不规则孔洞的组合。裂缝带通常在常规测井中呈现自然伽马值降低的特征，没有明显的扩径现象，声波时差和补偿中子呈现低值，密度变化不大，双侧向电阻率高值且表现为正差异；在成像测井图上，裂缝带通常呈现大量的深色正弦条带，如图 4 所示。

2　基于 XGBoost 的识别方法

2.1　XGBoost 算法原理

集成学习方法是指将多个弱学习器组合在一起，通过一定的策略来达到更好的预测性能、泛化能力或者解决复杂问题的目标，在实际应用中具有广泛的应用，能够提高模型的稳定性、准确性和泛化能力，适用于各种领域的分类、回归和特征选择等问题。提升算法属于集成学习方法，其基本思想是将多个弱学习算法组合起来，以获得一个更强的整体性能。这种集成学习方法被称为”Boosting”（提升）。Boosting 的基本思想是通过迭代训练一系列弱学习算法，每一轮训练都会关注之前轮次中分类错误的样本，逐步纠正错误，从而最终形成一个强大的集成模型。Boosting 的关键在于每轮训练都会聚焦于之前轮次中分类错误的样本，通过不断调整样本权重，最终使得集成模型在这些难以分类的样本上表现出色。Boosting 方法能够显著提高模型的性能，甚至在使用简单的弱学习器时也能取得很好的效果。常见的 Boosting 算法包括自适应提升算法 AdaBoost、梯度提升算法 GradientBoosting 和极端梯度提升算法（ExtremeGradientBoosting，简称 XGBoost）等[41-43]。具体流程如图 5 所示。

XGBoost 是对梯度提升树（GradientBoosting-Trees）算法的一种优化和扩展。它在梯度提升树的基础上引入了 L1 和 L2 正则化，控制树的复杂度，防止过拟合。这些正则化项被添加到损失函数中，以惩罚过于复杂的模型。使用 XGBoost 算法建立测井解释模型时，基于分类回归树（CART）作为基分类器定义目标函数，它包含损失函数与正则项：

$$obj = \sum_i l(y_i, \hat{y}_i) + \sum_k \Omega(f_k), \tag{1}$$

式中：obj 为目标函数，用于定义模型的学习目标；$l(y_i, \hat{y}_i)$ 为损失函数，用于衡量模型预测值 $\hat{y}_i$ 与真实值 y_i 之间的差异；$\Omega(f_k)$ 为正则项，用于控制模型的

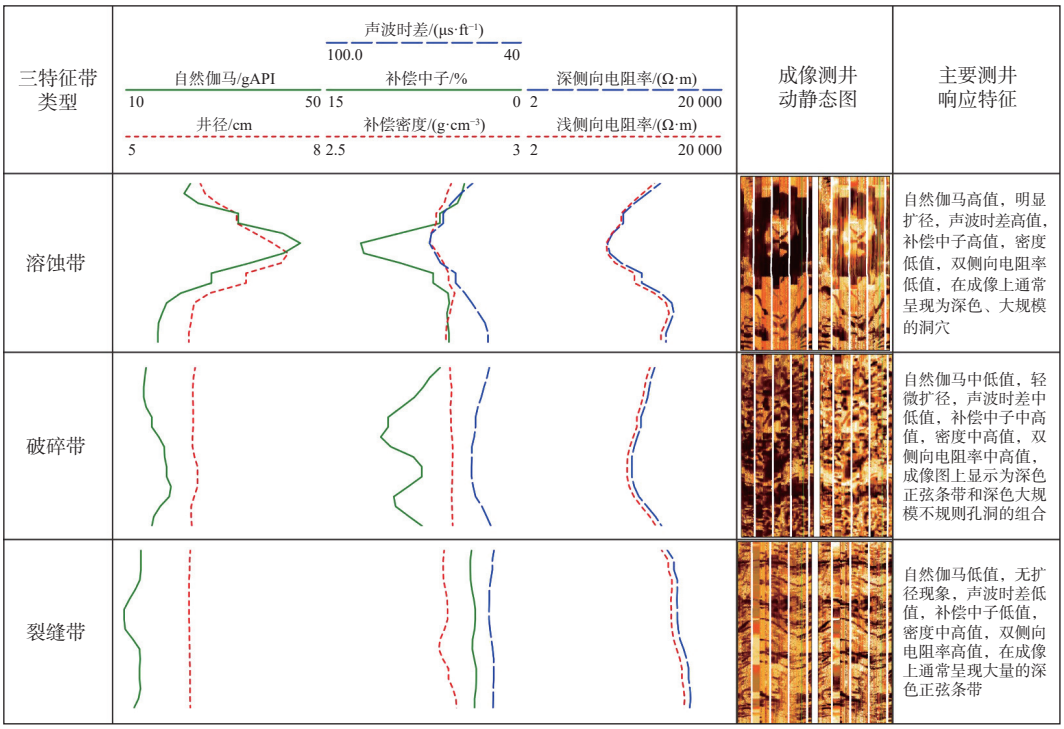


图 4　研究区溶蚀带、破碎带、裂缝带测井响应曲线特征

Fig. 4　Logging response curve characteristics of the dissolution, fault breccias and fracture (damage) zones in the study area

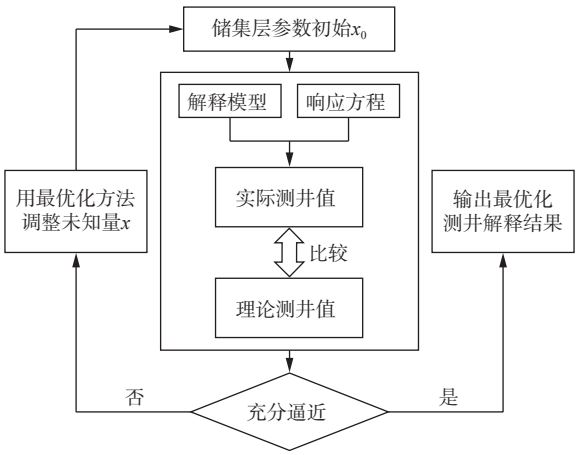


图 5　XGBoost 模型训练流程图

Fig. 5　XGBoost model training process diagram

复杂度以避免过拟合。正则项可以分为 2 部分，分别是 L1 正则化项和 L2 正则化项，用于限制叶子节点权重的大小：

$$\Omega(f_k)=\varepsilon T+\frac{1}{2}\lambda\sum_{j=1}^{T}\omega_j^2+\alpha\sum_{j=1}^{T}\left|\omega_j\right|, \tag{2}$$

式中：T 为树的叶子节点数量；ω_j 为叶子节点 j 的权重；ε，λ，α 分别为调节正则化项影响的超参数；L1 正则化项 $\alpha\sum_{j=1}^{T}\left|\omega_j\right|$ 有助于产生稀疏的解，而 L2 正则化项 $\lambda\sum_{j=1}^{T}\omega_j^2$ 则有助于控制权重的大小。

2.2　基于 XGBoost 的走滑断裂“三带”识别

2.2.1　XGBoost 模型建立

在运用 XGBoost 模型对走滑断裂内部的“三带”进行识别时，不同类型的测井曲线在量纲和数量级上存在差异，这种差异可能影响模型的识别准确度。因此，对井径 *CAL*、声波时差 *AC*、补偿密度 *DEN*、补偿中子 *CNL* 和自然伽马 *GR* 这 5 种测井曲线进行了标准化处理。由于不同井的测井曲线属性值大小各异，并且缺乏统一的上下限标准，因此采用了标准分数的方法进行标准化处理。这样可以确保各种曲线在相同的尺度范围内进行比较，从而更准确地进行走滑断裂识别，其计算式为：

$$Z=\frac{X-u}{\sigma}, \tag{3}$$

式中：Z 表示标准分数；X 是原始测井曲线属性值；u 是该属性值在所有样本上的均值；σ 是标准差。

通过对每个属性值应用上述公式，将测井曲线标准化为具有均值为 0、标准差为 1 的分布，从而确保了不同曲线之间的公平比较，进一步提高了走滑断裂识别的准确性。

将 gs110 井中 *CAL*、*AC*、*DEN*、*CNL*、*GR* 5 种测井响应参数作为特征变量，形成了 619 组样本数据。将走滑断裂溶蚀带、破碎带和裂缝带“三带”类型，分别用类型代码 1, 2, 3 表示，作为走滑断裂“三带”

识别标签数据，样本数据分区情况如表 1 所示。

通过 XGBoost 算法建立回归预测模型，需要设置的参数主要有：迭代次数（t）；控制每轮迭代中模型权重的更新步长的学习率（*eta*）；限制每棵决策树的深度，防止模型过于复杂的最大树深度（D_{max}）；限制叶子节点上的最小样本权重和，防止过拟合的叶子节点最小样本权重和（W_{min}）；每轮迭代中用于训练的样本比例，防止过拟合的子样本的比例（S）；用于控制树的生长，最小损失函数下降值（γ）。

根据前文分析，本研究以 *CAL*、*AC*、*DEN*、*CNL*、*GR* 5 类测井响应参数作为特征变量，首先确定迭代次数 t，并以此为基础对模型其他参数按照 *eta*，D_{max}，W_{min}，S 和 γ 的顺序进行调优。鉴于本研究目的为识别走滑断裂溶蚀带、破碎带和裂缝带的三分类问题，引入多分类评价指标对模型在不同参数下对走滑断裂“三带”识别的精准度进行评价从而完成对参数的调优[44-49]，其中准确率（Accuracy）表示在所有分类样本中，被正确分类的样本数量与总样本数量之比，公式为：

表 1　样本数据选取表

Table 1　Sample data selection table

“三带”类型	类型代码	训练数据	预测数据
溶蚀带	1	16	5
破碎带	2	269	67
裂缝带	3	210	52
合计/组	619	495	124
整体百分比/%		79.97	20.03

$$A=\frac{TP+TN}{N}, \tag{4}$$

式中：A 表示准确率（Accuracy）；TP 表示预测正确的类型代码为 1 的样本个数；TN 表示预测正确的类型代码为 2、3 的样本个数；N 表示样本总数。

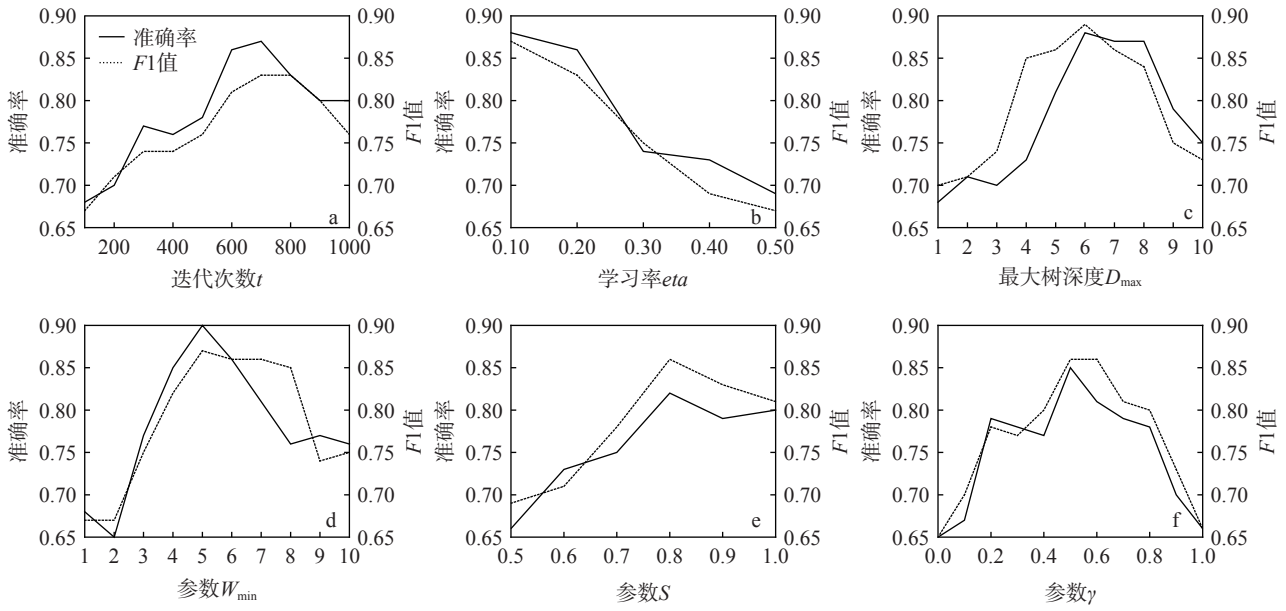
$F1$ 值（$F1$ score）：表示精确率（Precision）和召回率（Recall）的调和平均值，用于评价模型的精确度，公式为：

$$F=\frac{2\times\dfrac{TP}{TP+FP}\times\dfrac{TP}{TP+FN}}{\dfrac{TP}{TP+FP}+\dfrac{TP}{TP+FN}}, \tag{5}$$

式中：F 表示 $F1$ 值（$F1$ score）；FP 表示预测错误的类型代码为 2、3 的样本个数；FN 表示预测错误的类型代码为 1 的样本个数。

将迭代次数 t 分别设置为 100, 200, 300, 400, 500, 600, 700, 800, 900, 1 000，其他参数均选用 XGBoost 算法默认值：*eta*=0.3，D_{max}=6，W_{min}=1，S=1，γ=0。在其他参数均为默认值的条件下，改变迭代次数 t 可得到随迭代次数 t 变化 2 类评价指标 Accuracy 和 $F1$ score 的变化趋势图，如图 6 所示。从图中可以得知，当迭代次数 t 为 700 时，Accuracy 为最大值 0.87，$F1$ score 与迭代次数为 800 时同为 0.83，综合来看，选取迭代次数为 700 时该识别模型效果最好，即选取 t=700。

同理，在确定迭代次数 t=700 后，并在其他参数均为默认值的情况下，分别对 *eta*，D_{max}，W_{min}，S 和 γ 参数进行调整，最终选取的参数最优取值如表 2 所示。



W_{min}. 叶子点中最小样本权重和；S. 生成每棵树的随机样本采样的比例；γ. 最小损失函数下降值；下同

图 6　模型参数分析变化趋势图

Fig. 6　Trend chart of model parameter analysis changes

表 2　主要参数最优取值

Table 2　Optimal values of key parameters

参数名	最优取值
迭代次数t	700
学习率eta	0.1
最大树深度D_{max}	6
叶子点中最小样本权重和W_{min}	5
生成每棵树的随机样本采样的比例S	0.8
最小损失函数下降值γ	0.5

2.2.2　模型验证及结果分析

经由上述对 XGBoost 识别走滑断裂模型重要参数进行调整后，在 PyCharm 平台运行 XGBoost 代码得到走滑断裂“三带”划分训练样本及测试样本的预测情况，如图 7 所示。分析可知，训练样本共计 495 个，预测“三带”划分情况符合实际样本数 454 个，预测精度为 91.72%；测试样本共计 124 个，预测“三带”划分情况符合实际样本数 111 个，预测精度为 89.52%，表明识别结果较好。

3　应用效果分析

利用上述 XGBoost 走滑断裂识别模型分别对 gs20 井和 gs019 井进行了预测验证，在 2 口井共 234 个样本中出现了 26 个预测错误样本，主要存在将裂缝带与破碎带、溶蚀带与破碎带相互识别错误的情况，没有出现裂缝带与溶蚀带之间的错误识别，结果表明利用 XGBoost 算法对走滑断裂裂缝带、破碎带和溶蚀带“三带”的识别效果较好，识别精度为 88.89%。如图 8 所示，第 1 道为深度，第 2 道为岩性

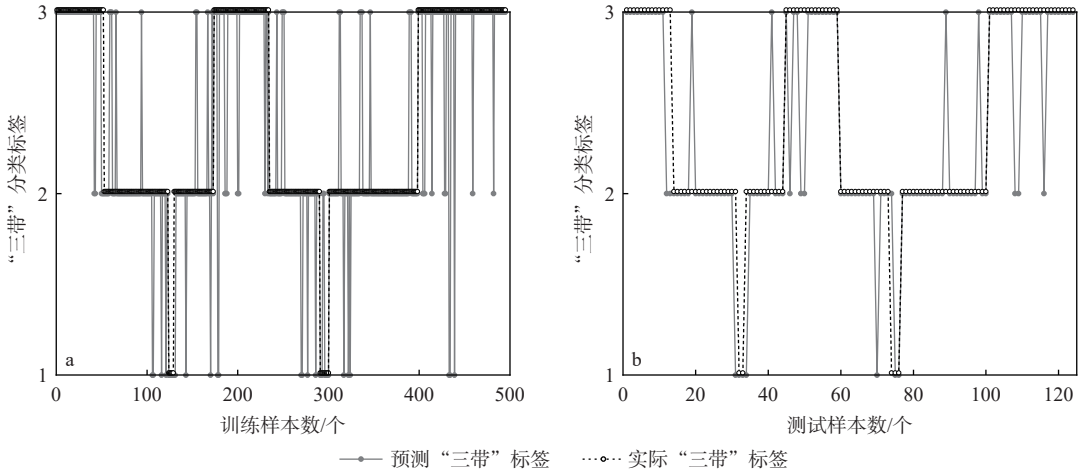


图 7　训练样本 (a) 与测试样本 (b) 的模型预测结果图

Fig. 7　Model prediction results for training (a) and testing (b) samples

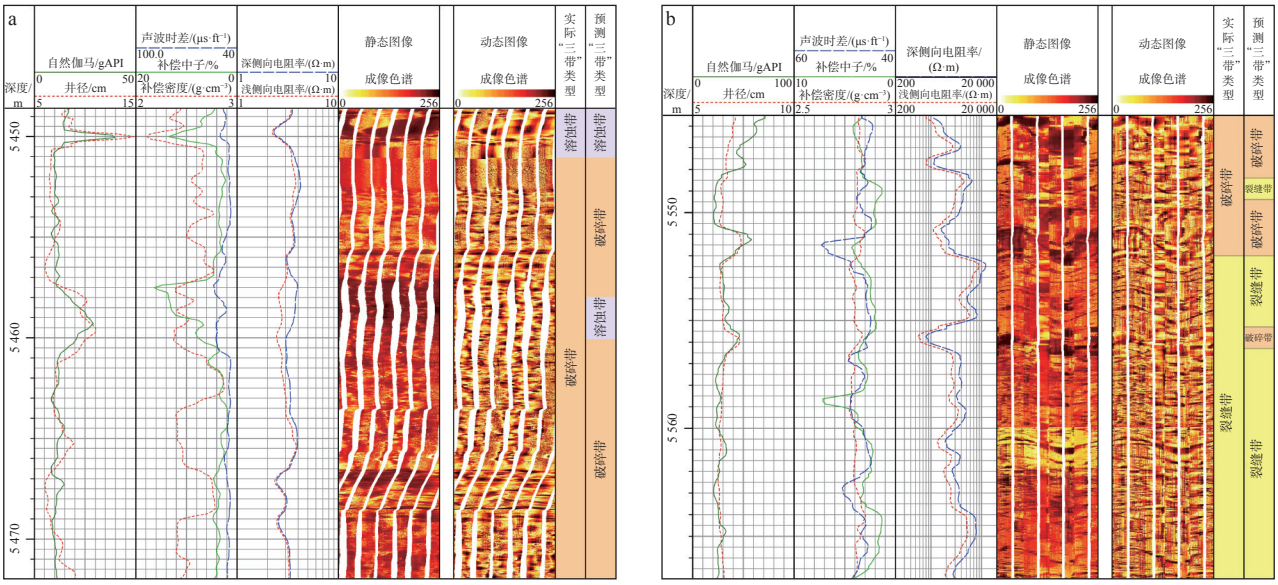


图 8　gs20(a) 和 gs019 井 (b) 走滑断裂“三带”识别效果图

Fig. 8　Identification results of the "Three Zones" of strike-slip faults in gs20 (a) and gs019 (b) wells

曲线，第 3 道为孔隙度曲线，第 4 道为电阻率曲线，第 5 道和第 6 道分别为成像测井静态图像和动态图像，第 7 道和第 8 道分别为实际“三带”类型道和预测“三带”类型道。

在高石梯–磨溪地区 EW 向一级走滑断裂 $F_{Ⅰ}11$ 和 EW 向二级走滑断裂 $F_{Ⅱ}16$ 附近区域运用 XGBoost 走滑断裂“三带”识别模型对该区域内 53 口井的井资料进行处理，完成了区域内 53 口井内部溶蚀带、破碎带及裂缝带的划分，得到走滑断裂 $F_{Ⅰ}11$ 及 $F_{Ⅱ}16$ 附近区域“三带”在平面上的分布，并利用识别结果在平面勾画出“三带”分布图，如图 9 所示。

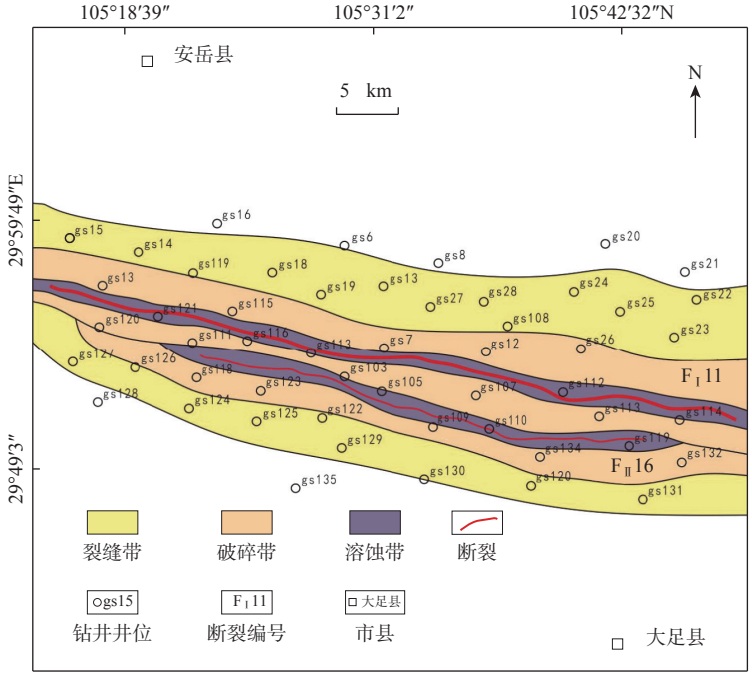


图 9　高石梯–磨溪地区走滑断裂 $F_{Ⅰ}11$ 及 $F_{Ⅱ}16$ 附近区域“三带”分布图

Fig. 9　Distribution of the "Three Zones" near strike-slip faults $F_{Ⅰ}11$ and $F_{Ⅱ}16$ in the Gaoshiti-Moxi region

4　结　论

（1）走滑断裂内部“三带”在裂缝发育及溶蚀程度上存在差异，在常规及成像测井资料上存在明显的响应特征，可利用测井资料有效识别走滑断裂内部的溶蚀带、破碎带及裂缝带 3 带。

（2）梯度提升树算法是一种集成学习方法，通过迭代训练多个弱学习算法，并逐步纠正错误，最终形成一个强大的集成模型。XGBoost 是对梯度提升树算法的优化和扩展，具有很高的准确率和稳定性，运用 XGBoost 算法可以有效地对走滑断裂内部 3 个特征带进行识别。

（3）以 *CAL*、*AC*、*DEN*、*CNL*、*GR* 5 类测井响应参数作为特征变量，并通过准确率（Accuracy）、*F*1 值（*F*1 score）、精确率（Precision）和召回率（Recall）优选迭代次数 *t*、学习率 *eta* 及最大树深度 D_{max} 等多分类评价指标优化基于 XGBoost 的识别模型，可有效提升模型的识别精度，达到较好的识别效果。

所有作者声明不存在利益冲突。

The authors declare that no competing interests exist.

参考文献：

[1] 梁瀚，唐浩，冉崎，等. 四川盆地川中地区走滑断裂的分布、类型与成因 [J]. 地质学报，2023，97(8)：2609-2620.
LIANG H，TANG H，RAN Q，et al. The distribution，type and origin of the strike-slip faults in the central Sichuan Basin[J]. Acta Geologica Sinica，2023，97(8)：2609-2620.(in Chinese with English abstract)

[2] 管树巍，姜华，鲁雪松，等. 四川盆地中部走滑断裂系统及其控油气作用 [J]. 石油学报，2022，43(11)：1542-1557.
GUAN S W，JIANG H，LU X S，et al. Strike-slip fault system and its control on oil & gas accumulation in central Sichuan Basin[J]. Acta Petrolei Sinica，2022，43(11)：1542-1557.(in Chinese with English abstract)

[3] 马兵山，梁瀚，邬光辉，等. 四川盆地中部地区多期次走滑断层的形成及演化 [J]. 石油勘探与开发，2023，50(2)：333-345.
MA B S，LIANG H，WU G H，et al. Formation and evolution of the strike-slip faults in the central Sichuan Basin，SW China[J]. Petroleum Exploration and Development，2023，50(2)：333-345.

(in Chinese with English abstract)

[4] WANG Y J, ZHANG Y Z, CAWOOD P A, et al. Early Neoproterozoic assembly and subsequent rifting in South China: Revealed from mafic and ultramafic rocks, central Jiangnan Orogen[J]. Precambrian Research, 2019, 331: 105367.

[5] YAN C L, SHU L S, FAURE M, et al. Time constraints on the closure of the Paleo-South China Ocean and the Neoproterozoic assembly of the Yangtze and Cathaysia blocks: Insight from new detrital zircon analyses[J]. Gondwana Research, 2019, 73: 175-189.

[6] LI S Z, ZHAO S J, LIU X, et al. Closure of the Proto-Tethys Ocean and Early Paleozoic amalgamation of microcontinental blocks in East Asia[J]. Earth-Science Reviews, 2018, 186: 37-75.

[7] 李献华, 李武显, 何斌. 华南陆块的形成与 Rodinia 超大陆聚合–裂解: 观察、解释与检验 [J]. 矿物岩石地球化学通报, 2012, 31(6): 543-559.
LI X H, LI W X, HE B. Building of the South China Block and its relevance to assembly and breakup of Rodinia Supercontinent: Observations, interpretations and tests[J]. Bulletin of Mineralogy, Petrology and Geochemistry, 2012, 31(6): 543-559.(in Chinese with English abstract)

[8] LIU Z, TAN S C, HE X H, et al. Petrogenesis of mid-Neoproterozoic (ca. 750 Ma) mafic and felsic intrusions in the Ailao Shan-Red River belt: Geochemical constraints on the paleogeographic position of the South China Block[J]. Lithosphere, 2019, 11(3): 348-364.

[9] CAWOOD P A, WANG Y J, XU Y J, et al. Locating South China in Rodinia and Gondwana: A fragment of greater India lithosphere?[J]. Geology, 2013, 41(8): 903-906.

[10] CAWOOD P A, STRACHAN R A, PISAREVSKY S A, et al. Linking collisional and accretionary orogens during Rodinia assembly and breakup: Implications for models of supercontinent cycles[J]. Earth and Planetary Science Letters, 2016, 449: 118-126.

[11] KONOPáSEK J, JANOUŠEK V, OYHANTçABAL P, et al. Did the circum-Rodinia subduction trigger the Neoproterozoic rifting along the Congo-Kalahari Craton margin?[J]. International Journal of Earth Sciences, 2018, 107(5): 1859-1894.

[12] 段金宝, 梅庆华, 李毕松, 等. 四川盆地震旦纪–早寒武世构造–沉积演化过程 [J]. 地球科学, 2019, 44(3): 738-755.
DUAN J B, MEI Q H, LI B S, et al. Sinian-Early Cambrian tectonic-sedimentary evolution in Sichuan Basin[J]. Earth Science, 2019, 44(3): 738-755.(in Chinese with English abstract)

[13] BIAN Q, DENG S, LIN H X, et al. Strike-slip salt tectonics in the Shuntuoguole Low Uplift, Tarim Basin, and the significance to petroleum exploration[J]. Marine and Petroleum Geology, 2022, 139: 105600.

[14] ZHU M, LIANG Z L, WANG X, et al. Mesozoic strike-slip fault system at the margin of the Junggar Basin, NW China[J]. Journal of Structural Geology, 2023, 175: 104950.

[15] LIANG S Z, LI Z, ZHANG W, et al. The characteristics of strike-slip faults and their control on hydrocarbon distribution in deep carbonate reservoirs of the central Sichuan Basin[J]. Frontiers in Earth Science, 2023, 11: 1064835.

[16] CHENG X Y, DING W L, PAN L, et al. Geometry and kinematics characteristics of strike-slip fault zone in complex structure area: A case study from the south No. 15 strike-slip fault zone in the eastern Sichuan Basin, China[J]. Frontiers in Earth Science, 2022, 10: 922664.

[17] 付小东, 张本健, 汪泽成, 等. 四川盆地中西部走滑断裂及其对油气成藏控制作用 [J]. 地球科学, 2023, 48(6): 2221-2237.
FU X D, ZHANG B J, WANG Z C, et al. Strike-slip faults in central and western Sichuan Basin and their control functions on hydrocarbon accumulation[J]. Earth Science, 2023, 48(6): 2221-2237.(in Chinese with English abstract)

[18] 马庆佑, 沙旭光, 李玉兰, 等. 塔中顺托果勒区块走滑断裂特征及控油作用 [J]. 石油实验地质, 2012, 34(2): 120-124.
MA Q Y, SHA X G, LI Y L, et al. Characteristics of strike-slip fault and its controlling on oil in Shuntuoguole region, middle Tarim Basin[J]. Petroleum Geology & Experiment, 2012, 34(2): 120-124.(in Chinese with English abstract)

[19] 吴孔友, 瞿建华, 王鹤华. 准噶尔盆地大侏罗沟断层走滑特征、形成机制及控藏作用 [J]. 中国石油大学学报(自然科学版), 2014, 38(5): 41-47.
WU K Y, QU J H, WANG H H. Strike-slip characteristics, forming mechanisms and controlling reservoirs of Dazhuluogou fault in Junggar Basin[J]. Journal of China University of Petroleum (Edition of Natural Science), 2014, 38(5): 41-47.(in Chinese with English abstract)

[20] MA D B, WANG Z C, DUAN S F, et al. Strike-slip faults and their significance for hydrocarbon accumulation in Gaoshiti-Moxi area, Sichuan Basin, SW China[J]. Petroleum Exploration and Development, 2018, 45(5): 851-861.

[21] 焦方正. 塔里木盆地顺托果勒地区北东向走滑断裂带的油气勘探意义 [J]. 石油与天然气地质, 2017, 38(5): 831-839.
JIAO F Z. Significance of oil and gas exploration in NE strike-slip fault belts in Shuntuoguole area of Tarim Basin[J]. Oil & Gas Geology, 2017, 38(5): 831-839.(in Chinese with English abstract)

[22] 许志琴, 王勤, 李忠海, 等. 印度-亚洲碰撞: 从挤压到走滑的构造转换 [J]. 地质学报, 2016, 90(1): 1-23.
XU Z Q, WANG Q, LI Z H, et al. Indo-Asian collision: Tectonic transition from compression to strike slip[J]. Acta Geologica Sinica, 2016, 90(1): 1-23.(in Chinese with English abstract)

[23] 邬光辉, 成丽芳, 刘玉魁, 等. 塔里木盆地寒武–奥陶系走滑断裂系统特征及其控油作用 [J]. 新疆石油地质, 2011, 32(3): 239-243.
WU G H, CHENG L F, LIU Y K, et al. Strike-slip fault system of the Cambrian-Ordovician and its oil-controlling effect in Tarim Basin[J]. Xinjiang Petroleum Geology, 2011, 32(3): 239-243.(in Chinese with English abstract)

[24] DENG S, LI H L, ZHANG Z P, et al. Structural characterization of intracratonic strike-slip faults in the central Tarim Basin[J]. AAPG Bulletin, 2019, 103(1): 109-137.

[25] WU G H, MA B S, HAN J F, et al. Origin and growth mechanisms of strike-slip faults in the central Tarim Cratonic Basin, NW China[J]. Petroleum Exploration and Development, 2021, 48(3): 595-607.

[26] 卢志强, 马乃拜, 尼加提, 等. 顺北油田深层断溶体内部结构物探表征研究 [J]. 内蒙古石油化工, 2022, 48(6): 115-120.
LU Z Q, MA N B, NI J T, et al. Research and application of fault-karst characterization technology in Shunbei oilfield[J]. Inner Mon-

golia Petrochemical Industry, 2022, 48(6): 115-120.(in Chinese with English abstract)

[27] 吴丰, 罗莹莹, 李昱翰, 等. 四川盆地公山庙油田大安寨段湖相灰岩-页岩裂缝特征与测井识别 [J]. 地质科技通报, 2022, 41(5): 55-67.
WU F, LUO Y Y, LI Y H, et al. Fracture characteristics and logging identification of lacustrine limestone-shale reservoirs in Da'anzhai Member, Gongshanmiao oilfield, Sichuan Basin[J]. Bulletin of Geological Science and Technology, 2022, 41(5): 55-67.(in Chinese with English abstract)

[28] 王震, 文欢, 邓光校, 等. 塔河油田碳酸盐岩断溶体刻画技术研究与应用 [J]. 石油物探, 2019, 58(1): 149-154.
WANG Z, WEN H, DENG G X, et al. Fault-Karst characterization technology in the Tahe oilfield, China[J]. Geophysical Prospecting for Petroleum, 2019, 58(1): 149-154.(in Chinese with English abstract)

[29] CHEN T Q, GUESTRIN C. XGBoost: A scalable tree boosting system[J]. ArXiv e-Prints, 2016: arXiv: 1603.02754.

[30] JOHNSON R, ZHANG T. Learning nonlinear functions using regularized greedy forest[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(5): 942-954.

[31] FRIEDMAN J H. Stochastic gradient boosting[J]. Computational Statistics & Data Analysis, 2002, 38(4): 367-378.

[32] 闫星宇, 顾汉明, 肖逸飞, 等. XGBoost 算法在致密砂岩气储层测井解释中的应用 [J]. 石油地球物理勘探, 2019, 54(2): 447-455.
YAN X Y, GU H M, XIAO Y F, et al. XGBoost algorithm applied in the interpretation of tight-sand gas reservoir on well logging data[J]. Oil Geophysical Prospecting, 2019, 54(2): 447-455.(in Chinese with English abstract)

[33] 孙予舒, 黄芸, 梁婷, 等. 基于 XGBoost 算法的复杂碳酸盐岩岩性测井识别 [J]. 岩性油气藏, 2020, 32(4): 98-106.
SUN Y S, HUANG Y, LIANG T, et al. Identification of complex carbonate lithology by logging based on XGBoost algorithm[J]. Lithologic Reservoirs, 2020, 32(4): 98-106.(in Chinese with English abstract)

[34] 丁阳阳, 赵军龙, 李兆明, 等. 基于 XGBoost 算法的煤体结构测井识别技术研究 [J]. 地球物理学进展, 2022, 37(3): 998-1006.
DING Y Y, ZHAO J L, LI Z M, et al. Research on logging recognition technology of coal structure based on XGBoost algorithm[J]. Progress in Geophysics, 2022, 37(3): 998-1006.(in Chinese with English abstract)

[35] 谷宇峰, 张道勇, 鲍志东. 利用混合模型 CRBM-PSO-XGBoost 识别致密砂岩储层岩性 [J]. 石油与天然气地质, 2021, 42(5): 1210-1222.
GU Y F, ZHANG D Y, BAO Z D. Lithology identification in tight sandstone reservoirs using CRBM-PSO-XGBoost[J]. Oil & Gas Geology, 2021, 42(5): 1210-1222.(in Chinese with English abstract)

[36] 赵军, 王伟明, 杨志冬, 等. 基于 DBN 模型的火山岩岩性识别方法: 以车排子地区为例 [J]. 地球物理学进展, 2022, 37(1): 328-337.
ZHAO J, WANG W M, YANG Z D, et al. Lithology identification method of volcanic rocks based on DBN model: An case in Chepaizi area[J]. Progress in Geophysics, 2022, 37(1): 328-337.(in Chinese with English abstract)

[37] 汪泽成, 姜华, 王铜山, 等. 四川盆地桐湾期古地貌特征及成藏意义 [J]. 石油勘探与开发, 2014, 41(3): 305-312.
WANG Z C, JIANG H, WANG T S, et al. Paleo-geomorphology formed during Tongwan tectonization in Sichuan Basin and its significance for hydrocarbon accumulation[J]. Petroleum Exploration and Development, 2014, 41(3): 305-312.(in Chinese with English abstract)

[38] 魏国齐, 杨威, 杜金虎, 等. 四川盆地高石梯—磨溪古隆起构造特征及对特大型气田形成的控制作用 [J]. 石油勘探与开发, 2015, 42(3): 257-265.
WEI G Q, YANG W, DU J H, et al. Tectonic features of Gaoshiti-Moxi paleo-uplift and its controls on the formation of a giant gas field, Sichuan Basin, SW China[J]. Petroleum Exploration and Development, 2015, 42(3): 257-265.(in Chinese with English abstract)

[39] 吴丰, 代槿, 姚聪, 等. 塔河油田奥陶系一间房组与鹰山组断溶体发育模式解剖 [J]. 断块油气田, 2022, 29(1): 33-39.
WU F, DAI J, YAO C, et al. Developmental mode analysis of the fault-karst reservoir in Yijianfang Formation and Yingshan Formation of Ordovician in Tahe oilfield[J]. Fault-Block Oil & Gas Field, 2022, 29(1): 33-39.(in Chinese with English abstract)

[40] 李梁, 孙庭斌, 陈迎楠, 等. 塔里木盆地肖塘南地区走滑断裂特征与断溶体勘探研究 [J]. 地质科技通报, 2023, 42(1): 239-245.
LI L, SUN T B, CHEN Y N, et al. Characteristics of strike-slip faults and exploration of fault-dissolution body in Xiaotangnan area of Tarim Basin[J]. Bulletin of Geological Science and Technology, 2023, 42(1): 239-245.(in Chinese with English abstract)

[41] BREIMAN L. Random forests[J]. Machine Learning, 2001, 45(1): 5-32.

[42] FRIEDMAN J, HASTIE T, TIBSHIRANI R. Additive logistic regression: A statistical view of boosting (With discussion and a rejoinder by the authors)[J]. The Annals of Statistics, 2000, 28(2): 337-374.

[43] FRIEDMAN J H. Greedy function approximation: A gradient boosting machine[J]. The Annals of Statistics, 2001, 29(5): 1189-1232.

[44] LIU C, WANG W Y, WANG M, et al. An efficient instance selection algorithm to reconstruct training set for support vector machine[J]. Knowledge-Based Systems, 2017, 116: 58-73.

[45] 尹帅, 田涛, 李俊鹿, 等. 鄂尔多斯盆地西南缘延长组断缝体特征及对油藏的调控作用 [J]. 油气地质与采收率, 2024, 31(1): 1-12.
YIN S, TIAN T, LI J L, et al. Fault-fracture body characteristics and their effect on hydrocarbon distribution of Yanchang Formation in southwestern margin of Ordos Basin[J]. Petroleum Geology and Recovery Efficiency, 2024, 31(1): 1-12.(in Chinese with English abstract)

[46] 刘粤蛟, 赖富强, 徐浩, 等. 基于测井曲线深程度耦合的页岩岩相智能识别方法 [J]. 地质科技通报, 2025, 44(1): 308-320.
LIU Y J, LAI F Q, XU H, et al. Intelligent identification methods for shale lithology based on the coupling deeply of logging curves[J]. Bulletin of Geological Science and Technology, 2025, 44(1): 308-320.(in Chinese with English abstract)

[47] 张驰, 潘懋, 胡水清, 等. 融合储层纵向信息的机器学习岩性识别方法 [J]. 地质科技通报, 2023, 42(3): 289-299.

ZHANG C, PAN M, HU S Q, et al. A machine learning lithologic identification method combined with vertical reservoir information[J]. Bulletin of Geological Science and Technology, 2023, 42(3): 289-299. (in Chinese with English abstract)

[48] 田明智, 朱超, 李森明, 等. 湖相碳酸盐岩测井岩相识别技术与应用: 以柴达木盆地英西地区为例 [J]. 中国石油勘探, 2023, 28(1): 135-143.
TIAN M Z, ZHU C, LI S M, et al. Application of logging lithofacies identification technology of lacustrine carbonate rocks: A case study of Yingxi area, Qaidam Basin[J]. China Petroleum Exploration, 2023, 28(1): 135-143. (in Chinese with English abstract)

[49] 高辉, 吴振坤, 柯雨, 等. 基于自然图像的钻探岩心识别 [J]. 煤田地质与勘探, 2023, 51(9): 64-71.
GAO H, WU Z K, KE Y, et al. Drilling core identification based on natural image[J]. Coal Geology & Exploration, 2023, 51(9): 64-71. (in Chinese with English abstract)

Summary for “基于 XGBoost 算法的走滑断裂内部特征带的精细识别”

Fine-grained identification of internal characteristic zones within strike-slip faults via the XGBoost algorithm

ZHAO Jun[1*], WANG Junyu[1], LAI Qiang[2], WEN Xiaofeng[3], WU Guanghui[1], JIAO Shixiang[1]

1. School of Earth Science and Technology, Southwest Petroleum University, Chengdu 610500, China;
2. Research Institute of Petroleum Exploration and Development, PetroChina Southwest Oil and Gas Field Company, Chengdu 610095, China;
3. China Petroleum Group Logging Co., LTD., Xi'an 710000, China

* Corresponding author, E-mail: zhaojun_70@126.com

Abstract: Because of the strong heterogeneity of fault zones, diverse reservoir types, and complex fluid distributions, the logging responses between damage, fault breccias , and dissolution zones within strike-slip faults are complex and variable, making it difficult to identify the three characteristic zones effectively inside strike-slip faults using imaging and conventional logging data. [**Objective**] The extreme gradient boosting (XGBoost) algorithm is introduced to establish a model to improve the identification accuracy of the three characteristic zones within strike-slip faults. [**Methods**] The logging response characteristics of the three characteristic zones within strike-slip faults are analyzed, and the sensitive logging curves are selected to construct a feature vector space set based on the mean and variance. The XGBoost algorithm is applied to establish XGBoost regression prediction models for the dissolution, breccias, and damage zones of strike-slip faults. The key parameters of the XGBoost model are optimized through multiclass evaluation indicators to improve the identification accuracy of the characteristic zones within strike-slip faults. [**Results**] The constructed XGBoost model was used to identify the internal characteristic zones of strike-slip faults in the study area, with a total of 234 samples; 208 samples were correctly identified, resulting in an identification accuracy of 88.89%. The prediction results reveal that, within the internal characteristic zones of strike-slip faults, the damage zone has the widest distribution, followed by the breccias zone, and the dissolution zone is the narrowest, which is consistent with the actual distribution of the internal characteristic zones of strike-slip faults. [**Conclusion**] The identification model of internal characteristic zones within strike-slip faults based on the XGBoost algorithm can be used to effectively identify the damage, breccias, and dissolution zones, thereby supporting more effective analysis of the distribution of small-scale dissolution cavities and fracture reservoir spaces inside strike-slip faults, and providing reference information for the accurate characterization of the internal structure of strike-slip faults.

Key words: strike-slip fault; XGBoost algorithm; carbonate rock; logging evaluation; characteristic zone identification; Sichuan Basin; Gaoshiti-Moxi region

Received: 2023-10-19; Revised: 2024-03-13; Accepted: 2024-03-15

doi: 10.19509/j.cnki.dzkq.tb20230583