

井级约束下长尾测井岩性分类方法

王焜龙¹ 张亚壕¹ 何 胜¹ 陈琪璟¹ 李 攀^{1,2}

WANG Kunlong ZHANG Yahao HE Sheng CHEN Qijing LI Pan

摘 要

针对测井岩性分类中少数类识别困难及同井样本相关性可能造成评价偏差的问题,文章基于 FORCE 2020 数据集构建井级划分下的 11 类岩性分类任务,并提出井级约束的 WGR-BSMOTE 方法。该方法仅在训练井中增强边界少数类样本,并结合验证集错分风险与测井响应重叠程度自适应调整采样强度。实验结果表明, XGBoost 的 Macro- F_1 由 0.486 1 提高至 0.497 9, Long-tail Avg- F_1 由 0.449 2 提高至 0.476 4。进一步分析表明,所提方法对长尾类别均衡识别具有一定改善作用,可作为井级约束条件下风险可控的边界增强策略。

关键词

测井岩性分类; 长尾分布; WGR-BSMOTE; 井级划分; 少数类识别; 测井响应

doi: 10.3969/j.issn.1672-9528.2026.07.038

0 引言

岩性识别是储层评价、地质建模和油气开发中的基础工作。测井资料连续性强、纵向分辨率高,能够反映不同岩性在自然伽马、密度、声波和电阻率等曲线上的响应差异,是开展岩性智能识别的重要数据来源。近年来,随机森林、XGBoost、LightGBM 等树模型以及 CNN、GRU 等深度模型被广泛用于测井岩性分类^[1]。

实际测井岩性数据通常呈明显长尾分布。页岩、砂岩等多数类样本较多,而白云岩、岩盐、凝灰岩等特殊岩性或薄层类别样本较少^[2]。若仅依赖总体 Accuracy,模型容易被多数类主导,从而掩盖少数类识别不足^[3]。

此外,测井样本沿深度方向具有强相关性。若同井相邻样本同时进入训练集和测试集,可能造成评价偏高。本文在井级划分条件下构建长尾测井岩性分类实验,并提出井级约束的地质响应感知 Borderline-SMOTE 方法,结合验证集错分风险与测井响应重叠程度调整少数类采样强度,重点分析长尾类别识别的收益、代价和能力边界。

1 数据与方法

1.1 数据来源与预处理

数据来源于 FORCE 2020 Well log and lithofacies dataset for machine learning competition 公开测井岩性数据集。该数

据集由 Bormann 等^[4]发布于 Zenodo,来源于 Norwegian Sea 海域 118 口井的测井与岩性解释资料,曾用于 FORCE 2020 机器学习岩性预测竞赛。数据集中包含井名、测深、井位坐标、地层分组、岩性标签以及 CALI、RDEP、RHOB、GR、NPHI、PEF、DTC 等多种常规测井曲线,能够反映不同井之间的测井响应差异和岩性类别分布不均衡特征^[5-6]。考虑到 Basement 类样本数量极少,且在井级划分条件下容易造成评价指标不稳定,本文将其剔除,并构建 11 类测井岩性分类任务^[7]。输入变量选取 GR、RDEP、RHOB、NPHI、DTC 等 16 条有效测井曲线,并进一步用于窗口统计特征构建。为降低同井样本相关性和预处理信息泄露风险,本文以井为最小单位划分训练集、验证集和测试集;缺失值填补、异常值裁剪和标准化参数均仅在训练集上拟合,验证集和测试集仅使用训练集统计量进行转换^[8-10],如图 1 所示。

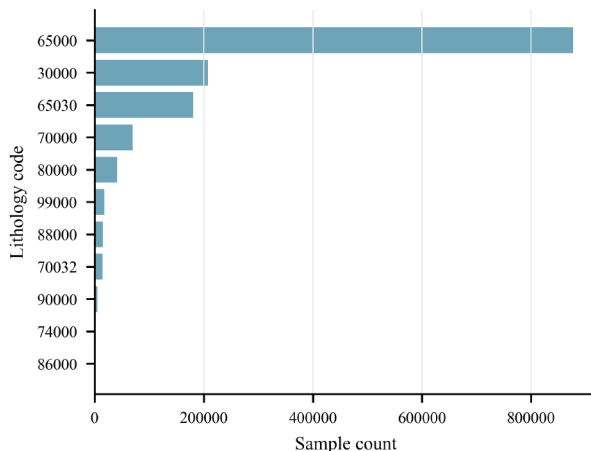


图 1 FORCE 2020 数据集岩性类别长尾分布

1. 应急管理大学计算机科学与工程学院 河北廊坊 065201
2. 河北省高校智慧应急应用技术研发中心 河北廊坊 065201
[基金项目] 中央高校基本科研业务费研究生科技创新基金资助 (ZY20260323)

1.2 模型与评价指标

本文选取 CNN-GRU、LightGBM 和 XGBoost 作为对比模型，并从总体性能和少数类识别能力两个方面进行评价。评价指标包括 Accuracy、Macro- F_1 、Weighted- F_1 、Long-tail Avg- F_1 、Target Minority Avg- F_1 和每类 F_1 。其中 Long-tail Avg- F_1 面向 70032、74000、86000、88000、90000 和 99000 等整体长尾类别，Target Minority Avg- F_1 面向 74000、88000 和 99000 等重点少数类^[11-12]。

1.3 实验设置与复现细节

实验采用固定井级划分，训练集、验证集和测试集分别包含 70、26 和 22 口井，同一口井样本不同时出现在训练集、验证集和测试集中。为降低数据泄漏风险，缺失值填补、异常值裁剪和标准化参数仅在训练集上拟合，验证集和测试集仅使用训练集统计量进行转换。所有主要实验重复 5 次，随机种子为 2021、2022、2023、2024 和 2025。深度模型输入为 21×16 测井窗口，树模型输入为 64 维 C/M/S/K 窗口统计特征^[13-14]。模型评价采用 Accuracy、Macro- F_1 、Weighted- F_1 、Long-tail Avg- F_1 、Target Minority Avg- F_1 以及“关键少数”类 Recall 和 F_1 ^[15]。

1.4 井级约束的地质响应感知 WGR-BSMOTE 方法

普通 SMOTE 通过在少数类近邻之间合成样本缓解类别不平衡^[16]，Borderline-SMOTE 进一步关注少数类边界危险样本^[17]。但在测井岩性分类中，样本具有井内连续性和井间分布差异，若不加约束地合成样本，可能放大噪声或引入评价偏差。因此，本文提出井级约束的地质响应感知 Borderline-SMOTE 方法（WGR-BSMOTE）。

WGR-BSMOTE 只在训练井内部执行边界少数类增强，验证集和测试集保持原始类别分布。该方法不直接作用于 21×16 原始序列窗口，而是作用于树模型使用的 64 维 C/M/S/K 窗口统计特征，以降低合成样本破坏深度方向测井结构的

风险。对长尾类别 i ，首先利用训练集基线模型在验证集上的 Recall 计算错分风险，并根据验证集混淆矩阵确定该类主要混淆类别；随后在 GR_C 、 GR_M 、 DTC_M 、 $RHOB_M$ 、 $NPHI_M$ 、 $RDEP_M$ 、 GR_S 和 $RHOB_S$ 等关键统计特征上计算测井响应重叠分数。错分风险和最终采样比例分别定义为：

$$E_i = 1 - \text{Recall}_i \quad (1)$$

$$r_i = \min(r_{\max}, r_0 \times (1 + \alpha O_i) \times (1 + \beta E_i)) \quad (2)$$

式中： E_i 表示第 i 类在验证集上的错分风险， Recall_i 表示第 i 类在验证集上的召回率， O_i 表示第 i 类与主要混淆类别之间的测井响应重叠分数， r_i 表示相对于该少数类原训练样本数的扩增比例， r_0 表示基础采样比例， α 和 β 分别表示响应重叠权重和错分风险权重， $r_{\max}$ 表示最大采样比例上限，用于

避免过度过采样。本文默认 $r_0=0.5$ ， $\alpha=0.5$ ， $\beta=0.5$ ，并在参数敏感性分析中进一步考察其影响。

测井响应重叠分数 O_i 基于验证集中的主要混淆类别计算。具体而言，先根据验证集混淆矩阵确定第 i 类最容易预测为哪一类，再在 GR_C 、 GR_M 、 DTC_M 、 $RHOB_M$ 、 $NPHI_M$ 、 $RDEP_M$ 、 GR_S 和 $RHOB_S$ 等特征上计算两类样本四分位区间重叠度，并对多项特征取平均。上述特征结合了测井响应敏感性和树模型统计特征重要性，用于刻画边界样本的潜在混淆风险，但不作为绝对地质判据。

对于某一特征 f ，设少数类 i 的四分位区间为 $[Q_{1i}^f, Q_{3i}^f]$ ，主要混淆类 j 的四分位区间为 $[Q_{1j}^f, Q_{3j}^f]$ 。定义两类四分位区间的交集长度和并集长度分别为：

$$A_i^f = \min(Q_{3i}^f, Q_{3j}^f) - \max(Q_{1i}^f, Q_{1j}^f) \quad (3)$$

$$B_i^f = \max(Q_{3i}^f, Q_{3j}^f) - \min(Q_{1i}^f, Q_{1j}^f) \quad (4)$$

$$O_i^f = \max(0, A_i^f / B_i^f) \quad (5)$$

当两类四分位区间完全分离时，交集长度 A_i^f ，对应的重叠分数 $O_i^f=0$ ；当两类四分位区间重叠程度较高时， O_i^f 逐渐接近 1。最终，类别 i 的测井响应重叠分数 O_i 取 GR_C 、 GR_M 、 DTC_M 、 $RHOB_M$ 、 $NPHI_M$ 、 $RDEP_M$ 、 GR_S 和 $RHOB_S$ 等关键特征重叠分数的平均值。

Borderline-SMOTE 的近邻参数设置为 $k_{\text{neighbors}}=5$ ， $m_{\text{neighbors}}=10$ ， $\text{random}_{\text{state}}$ 与对应实验随机种子一致。WGR-BSMOTE 仅在训练井样本集合内执行，验证井和测试井不参与过采样；当前实现允许不同训练井中的同类少数样本共同构成近邻空间，但所有合成样本均来源于训练井集合，不使用验证集或测试集样本。

值得一提的是，WGR-BSMOTE 对不同少数类的作用存在差异，尤其对测井响应高度重叠、测试样本规模有限或跨井分布差异明显的类别，仍可能效果有限。WGR-BSMOTE 的整体流程如图 2 所示，包括井级划分、训练集预处理、统计特征构建、错分风险与测井响应重叠计算、自适应采样、模型训练以及测试评估。

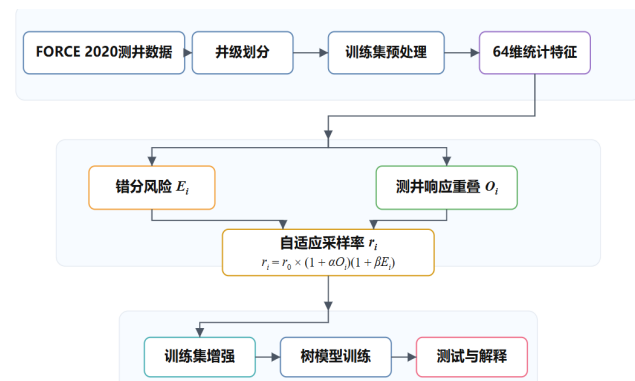


图 2 WGR-BSMOTE 方法流程图

2 结果与分析

2.1 主模型与少数类识别结果

主实验结果表明，不同模型在总体精度和类别均衡指标上存在差异。LightGBM 的 Accuracy 和 Weighted- F_1 较高，XGBoost 的 Macro- F_1 和 Target Minority Avg- F_1 较高，CNN-GRU 在总体 Accuracy 上接近树模型，但少数类识别能力相对不足。

表 1 主模型性能对比

模型	Accuracy/%	Macro- F_1	Weighted- F_1
XGBoost	68.29±0.34	0.486 1±0.0054	0.697 8±0.003 2
LightGBM	69.61±0.16	0.482 3±0.0049	0.705 0±0.001 5
CNN-GRU	68.36±2.86	0.385 1±0.0482	0.680 7±0.022 3

如表 1 所示，各类别 F_1 结果表明，少数类识别效果不仅受样本数量影响，也与类别之间的测井响应可分性有关。表 2 仅选取代表性多数类、重点少数类及综合少数类指标进行展示，以反映不同类别识别效果的差异。

表 2 重点岩性类别 F_1 对比

类别	测试 样本数	CNN- GRU- F_1	LightGBM- F_1	XGBoost- F_1
30000 砂岩	27 884	0.637 8	0.675 4	0.679 2
65000 页岩	124 839	0.828 5	0.825 1	0.812 3
74000 白云岩	378	0.013 6	0.020 3	0.014 0
88000 岩盐	3 899	0.630 3	0.950 2	0.970 3
99000 凝灰岩	2 369	0.059 1	0.228 8	0.298 6
Target Minority Avg- F_1	—	0.234 3	0.399 8	0.427 6

2.2 WGR-BSMOTE 长尾优化效果

由表 3 和图 3 可见，在 5 个随机种子的重复实验中，WGR-BSMOTE 与其他长尾优化方法表现出不同的性能特征。结果显示，WGR-BSMOTE 相比普通 Borderline-SMOTE，在

Macro- F_1 和 Long-tail Avg- F_1 上更稳定，但 Accuracy 并非最高，说明长尾优化存在总体精度与少数类指标之间的权衡。

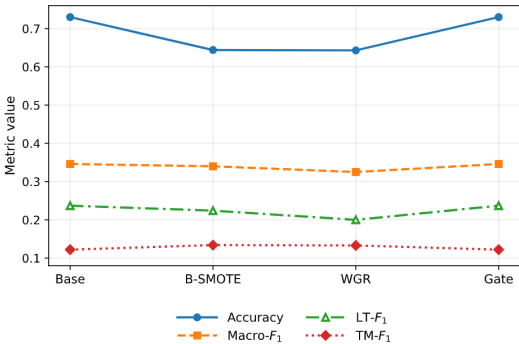


图 3 不同长尾优化策略的多指标权衡结果

从 5 个随机种子的平均结果看，LightGBM+WGR-BSMOTE 的 Macro- F_1 为 0.490 0，Long-tail Avg- F_1 约为 0.462 9；XGBoost+WGR-BSMOTE 的 Macro- F_1 为 0.497 9，Long-tail Avg- F_1 约为 0.476 4。但两者 Accuracy 均低于对应基线或类别权重方法，说明总体精度和长尾类别均衡识别之间存在权衡。

2.3 白云岩识别困难与测井响应解释

WGR-BSMOTE 能改变部分样本的预测结果分布，但未能稳定改善 74000 白云岩识别，说明极少数类瓶颈并非单纯由类别数量不足造成。该结果提示，74000 白云岩识别还受到测井响应重叠、测试样本规模有限和跨井分布差异的共同影响。

该现象提示，74000 白云岩与页岩、石灰岩在关键测井响应空间中存在较强重叠，同时受测试样本规模有限和跨井分布差异影响。相比之下，88000 岩盐虽然样本较少，但测井响应可分性更强，因此在树模型中保持较高 F_1 。

窗口统计特征分析表明，中心值、均值、标准差和斜率共同刻画了测井曲线的局部水平、波动和变化趋势。GR、DTC、RHOB 等曲线的统计特征具有明确测井物理意义，可

表 3 WGR-BSMOTE 与不同长尾方法的 5 个随机种子对比结果

模型 / 方法	Accuracy	Macro- F_1	Long-tail Avg- F_1	Target Minority Avg- F_1	R74000	F99000
LGBM/Base	0.696±.002	0.482±.005	0.445±.008	0.400±.009	0.011±.009	0.229±.023
LGBM/CW	0.689±.004	0.494±.007	0.475±.009	0.414±.008	0.000±.000	0.291±.024
LGBM/BSM	0.675±.007	0.478±.009	0.446±.014	0.408±.008	0.001±.001	0.271±.025
LGBM/WGR	0.676±.004	0.490±.008	0.463±.015	0.418±.012	0.005±.007	0.297±.032
XGB/Base	0.683±.003	0.486±.005	0.449±.007	0.428±.007	0.015±.014	0.299±.012
XGB/CW	0.677±.001	0.484±.002	0.453±.002	0.440±.006	0.018±.008	0.333±.015
XGB/BSM	0.665±.007	0.465±.008	0.420±.015	0.398±.004	0.008±.004	0.249±.014
XGB/WGR	0.668±.008	0.498±.005	0.476±.007	0.425±.006	0.001±.001	0.324±.018

帮助解释树模型在长尾类别上的差异表现。

由图 4 可见，74000 白云岩在 GR、DTC、RHOB 等关键统计特征上与页岩、砂岩等类别存在重叠，这为其低 Recall 提供了测井响应层面的解释。

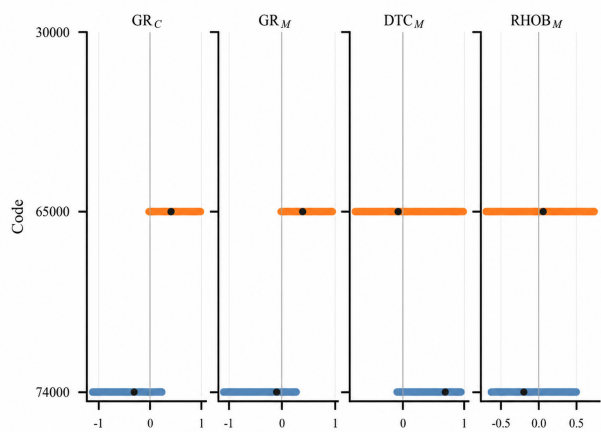


图 4 74000 白云岩与主要混淆类别的关键测井响应重叠

2.4 消融实验与参数敏感性分析

为检验响应重叠模块和错分风险模块的贡献，本文设置 Baseline、普通 Borderline-SMOTE、WGR-no-overlap、WGR-no-risk 和完整 WGR-BSMOTE 进行消融，具体消融结果见表 4。结果表明，不同模块对不同指标的影响存在差异，完整 WGR-BSMOTE 在 Macro- F_1 和 Long-tail Avg- F_1 上具有较稳定改善趋势，但 74000 Recall 仍未稳定提高。

表 4 WGR-BSMOTE 消融实验结果

模型	方法	Accuracy	Macro- F_1	Long-tail	R74000
				Avg- F_1	
LGBM	Baseline	0.696 1	0.482 3	0.445 3	0.011 1
LGBM	B-SMOTE	0.674 6	0.478 0	0.446 3	0.000 5
LGBM	WGR-BSMOTE	0.675 5	0.490 0	0.462 9	0.004 8
LGBM	no-overlap	0.675 7	0.486 2	0.458 8	0.004 2
LGBM	no-risk	0.677 3	0.486 6	0.459 8	0.005 3
XGB	Baseline	0.682 9	0.486 1	0.449 2	0.015 3
XGB	B-SMOTE	0.664 8	0.465 4	0.420 2	0.007 9
XGB	WGR-BSMOTE	0.667 9	0.497 9	0.476 4	0.000 5
XGB	no-overlap	0.670 3	0.494 6	0.469 7	0.000 0
XGB	no-risk	0.667 8	0.492 5	0.467 7	0.000 0

为进一步检验 WGR-BSMOTE 参数设置的合理性，本文围绕基础采样比例 r_0 、响应重叠权重 α 和错分风险权重 β 开展参数敏感性分析。在固定 $r_0=0.5$ 的条件下，设置 α 和 β 分别在 $\{0, 0.25, 0.5, 0.75, 1.0\}$ 范围内变化；在固定 $\alpha=0.5$ 、 $\beta=0.5$ 的条件下，设置 r_0 在 $\{0.25, 0.5, 0.75, 1.0, 1.25\}$ 范围内

变化。

图 5~6 展示了不同参数组合下模型性能的变化情况。结果表明，WGR-BSMOTE 的长尾均衡指标对 α 和 β 存在一定敏感性。其中，Long-tail Avg- F_1 的较优结果出现在 $\alpha=1.0$ 、 $\beta=0.75$ 附近，达到 0.491 2；默认参数组合 $\alpha=0.5$ 、 $\beta=0.5$ 下 Long-tail Avg- F_1 为 0.486 0，与较优结果差距较小，说明默认参数位于相对稳定区域。 r_0 扫描结果进一步表明，采样强度并非越高越好，过强的少数类增强可能改变模型决策边界，从而带来 Accuracy 与长尾均衡指标之间的权衡。

需要注意的是，74000 白云岩 Recall 在参数扫描中仍接近 0，说明该类别的识别困难并不能单纯依靠调整采样比例解决。这一结果进一步支持前文关于白云岩测井响应重叠、测试样本规模有限和跨井泛化困难的分析。

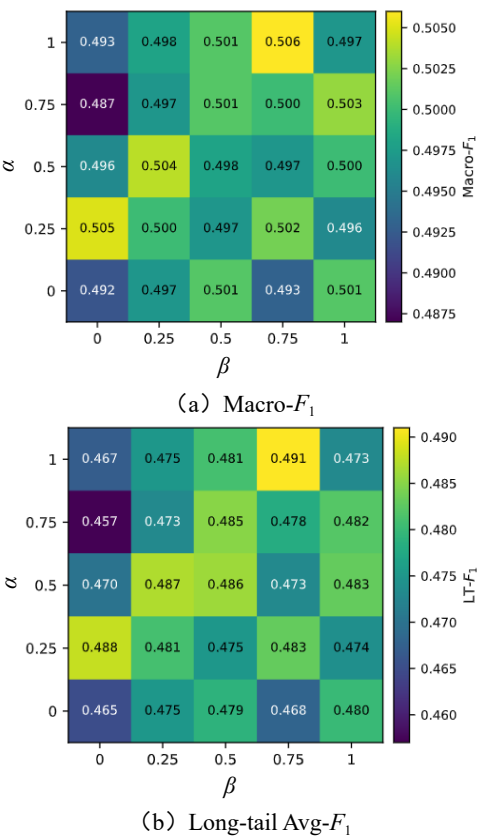


图 5 WGR-BSMOTE α - β 参数敏感性热力图

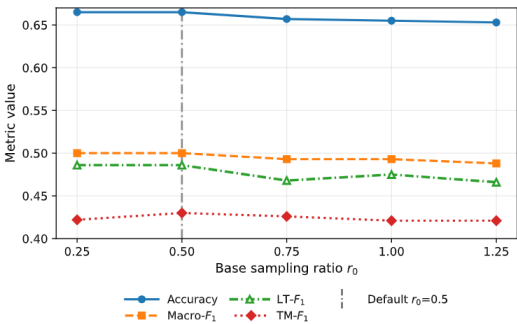


图 6 r_0 采样比例下的多指标权衡

2.5 跨井划分与门控风险控制

为进一步检验结果是否依赖于单一固定井级划分，本文设置 5 次重复井级随机划分实验。每次实验均以井为最小划分单元，重新划分训练集、验证集和测试集，并仅使用训练集统计量完成缺失值填补、异常值裁剪和标准化处理。在每次划分中，WGR-BSMOTE 仅作用于训练井样本，验证集和测试集保持原始类别分布。

表 5 为跨井划分条件下不同方法的平均结果。结果显示，当测试井组合发生变化时，不同长尾增强策略的指标表现会出现一定波动，说明井间地质差异和少数类空间分布会影响模型推广效果。对于实际工区应用，建议结合验证集表现对过采样策略进行筛选，避免在收益不足的井组组合中盲目增强。

表 5 跨井划分条件下不同方法的应用敏感性结果

模型	方法	Accuracy	Macro- F_1	Long-tail Avg- F_1	Target Minority Avg- F_1	R74000
XGB	Base	0.730±.032	0.346±.068	0.237±.125	0.122±.115	0.000±.000
XGB	BSM	0.644±.037	0.340±.055	0.224±.117	0.134±.125	0.020±.024
XGB	WGR	0.643±.036	0.325±.045	0.200±.090	0.133±.125	0.003±.004

因此，WGR-BSMOTE 更适合在固定工区或地质条件相近的井组内应用，并可配合门控机制进行风险筛选。该结果说明，长尾测井岩性分类方法在实际应用中不仅要关注固定划分下的指标提升，也需要考虑跨井组合变化带来的应用条件。

为降低跨井划分条件下盲目过采样带来的风险，本文进一步引入验证集门控机制 WGR-Gate。该策略仅依据验证集表现决定是否启用 WGR-BSMOTE，不使用测试集调参或选择模型，从而使过采样策略更符合实际工区应用中的风险控制需求。

具体而言，记 LT 为验证集 Long-tail Avg- F_1 中 WGR-BSMOTE 相对 Baseline 的变化，ACC 为验证集 Accuracy 的变化。当 $LT > 0$ 且 $ACC > -0.02$ 时启用 WGR-BSMOTE，否则保留 Baseline。测试集仅用于评价门控策略选定后的模型。

实验结果表明，门控机制能够识别部分划分条件下过采样收益不足或精度代价较高的情况，从而为实际工区应用提供风险筛选依据。该机制并非追求在所有井组组合中强制启用过采样，而是在验证集支持增强有效时才使用 WGR-BSMOTE，以提高方法应用的安全性。

WGR-BSMOTE 在固定井级划分下对 Macro- F_1 和 Long-tail Avg- F_1 表现出一定改善，主要原因是其采样强度不仅考虑少数类样本数量，还结合了验证集错分风险和测井响应重叠程度，从而能够更有针对性地增强边界少数类样本。然而，重复井级划分和 WGR-Gate 结果表明，该方法的收益受测试

井组合影响较大，不能简单外推为稳定的跨井泛化优势。因此，WGR-BSMOTE 更适合作为一种风险可控的长尾边界增强策略，而非适用于所有井组组合的通用方法。

74000 白云岩仍然难以识别，说明极少类分类瓶颈不能仅依靠边界过采样解决。该类别可能同时受到测试样本规模有限、测井响应与页岩和石灰岩重叠、跨井分布差异以及多数类决策区域挤压等因素影响。相比之下，99000 凝灰岩和 88000 岩盐更容易取得较好识别效果，可能与其统计响应边界相对清晰或测井响应特征较独立有关。本文的地质解释仍主要基于测井响应和统计特征，后续可结合岩心、层位或沉积相资料进一步验证。

3 结论

(1) FORCE 2020 测井岩性数据存在明显长尾分布。采用井级划分，并仅使用训练集统计量进行预处理，有助于降低同井样本相关性和预处理信息泄露带来的评价偏差。

(2) WGR-BSMOTE 在 5 个随机种子的重复实验中，对 Macro- F_1 和 Long-tail Avg- F_1 表现出一定改善，但 Accuracy 并非最高，说明长尾优化需要在总体精度和少数类识别能力之间进行权衡。

(3) 74000 白云岩 Recall 未得到稳定提升，说明该类识别困难不仅受样本数量影响，还与测井响应重叠、测试样本规模有限和跨井分布差异有关。

(4) 总体来看，WGR-BSMOTE 可作为井级约束条件下长尾测井岩性分类的一种边界增强策略。结合验证集门控机制，可进一步识别跨井划分条件下的过采样风险，提高方法应用的稳定性。后续可结合更多独立井资料、层位约束或沉积相先验，进一步提升极少类岩性的识别能力。

参考文献：

[1] 江丽,张智谟,王琦玮,等.基于不同机器学习模型的石油测井数据岩性分类对比研究[J].物探与化探,2024,48(2):489-497.

[2] 王恒,姜亚楠,张欣,等.基于梯度提升算法的岩性识别方法[J].吉林大学学报(地球科学版),2021,51(3):940-950.

[3] 程希,周军,傅海成,等.机器学习算法在地球物理测井中的适用性及应用[J].西北地质,2023,56(4):336-348.

[4] Peter B, Peder A, Fahad D, et al. FORCE 2020 well log and lithofacies dataset for machine learning competition[EB/OL].(2020-12-18)[2026-07-10].<https://zenodo.org/records/4351156>.

[5] Xie Yunxin, Zhu Chenyang, Zhou Wen, et al. Evaluation of

machine learning methods for formation lithology identification: a comparison of tuning processes and model performances[J]. Journal of Petroleum Science and Engineering, 2018, 160: 182-193.

- [6]Hall B. Facies classification using machine learning[J]. The Leading Edge, 2016, 35(10): 906-909.
- [7]Liu Jingjing, Liu Jianchao. Integrating deep learning and logging data analytics for lithofacies classification and 3D modeling of tight sandstone reservoirs[J]. Geoscience Frontiers, 2022, 13(1): 101311.
- [8]Breiman L. Random forests[J]. Machine Learning, 2001, 45: 5-32.
- [9]Chen Tianqi, Guestrin C. XGBoost: a scalable tree boosting system[PP/OL].(2016-06-10)[2026-07-13].<https://doi.org/10.48550/arXiv.1603.02754>.
- [10]Ke Guolin, Meng Qi, Finley T, et al. LightGBM: a highly efficient gradient boosting decision tree[C]//Advances in Neural Information Processing Systems. NewYork: ACM, 2017: 3146-3154.
- [11]LeCun Y, Bengio Y, Hinton G. Deep learning[J]. Nature, 2015, 521: 436-444.
- [12]Cho K, Van Merriënboer B, Gulcehre C, et al. Learning phrase representations using RNN encoder-decoder for statis-

tical machine translation[PP/OL].(2014-09-03)[2026-07-13].
<https://doi.org/10.48550/arXiv.1406.1078>.

- [13]He Haibo, Garcia E A. Learning from imbalanced data[J]. IEEE Transactions on Knowledge and Data Engineering, 2009, 21(9): 1263-1284.
- [14]Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//2017 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2017: 2980-2988.
- [15]Cui Yin, Jia Menglin, Lin T Y, et al. Class-balanced loss based on effective number of samples[PP/OL].(2019-01-16)[2026-07-13].<https://doi.org/10.48550/arXiv.1901.05555>.
- [16]Chawla N V, Bowyer K W, Hall L O, et al. SMOTE: synthetic minority over-sampling technique[PP/OL].(2011-06-09)[2026-07-13].<https://doi.org/10.48550/arXiv.1106.1813>.
- [17]Han Hui, Wang Wenyuan, Mao Binghuan. Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning[C]//Advances in Intelligent Computing. Berlin: Springer, 2005: 878-887.

【作者简介】

王焜龙（1999—），男，河南平顶山人，硕士研究生，研究方向：机器学习、图像识别、灾害信息处理。

（收稿日期：2026-06-18 修回日期：2026-07-17）

（上接第 157 页）

参考文献：

- [1] 贺英,王天一,李莹莹.线性光学克隆机改进的离散极化调制连续变量量子密钥分发可组合安全性分析[J].物理学报, 2024, 73(23): 49-58.
- [2] 王坚,黄普善,金东方,等.基于无线信道密钥生成的群组通信密钥共享方法[J].通信技术, 2024, 57(12):1291-1296.
- [3] 姜震,张勇,王莉.基于经验模态分解的企业等值连接共享数据加密方法[J].信息技术,2023(11):120-125.
- [4] 陈金木,陈少英,黄朝阳.基于云服务器应用的密码管理方案设计[J].河南工程学院学报(自然科学版), 2023, 35(2): 54-57.
- [5] 何志成.基于 Shamir 算法的多 QKD 网络密钥共享策略[J].信息与电脑(理论版), 2023, 35(2): 231-234.
- [6] 解建国,刘晶,吴涵,等.基于超晶格 PUF 的轻量级信息论安全密钥达成协议[J].密码学报(中英文), 2024, 11(2): 387-402.
- [7] 唐长虹,赵艳琦,杨晓艺,等.加权门限 SM2 签名方案[J].软件学报, 2025, 36(8): 3883-3895.

[8] 臧洪睿,王天琰,姜姝宇,等.物联网环境下去中心化的联邦学习加密仿真[J].计算机仿真, 2025, 42(1): 400-404.

[9] 张泽辉,李庆丹,富瑶,等.面向非独立同分布数据的自适应联邦深度学习算法[J].自动化学报, 2023, 49(12): 2493-2506.

【作者简介】

李文科（1985—），男，贵州榕江人，本科，高级工程师，研究方向：人工智能、数据应用及治理。

王益彰（1998—），男，贵州安顺人，本科，助理工程师，研究方向：大数据。

田钺（1996—），男，贵州施秉人，本科，工程师，研究方向：电力行业数字化技术、数据运营。

董若烟（1995—），女，贵州贵阳人，硕士，工程师，研究方向：大数据、人工智能。

姚舜（1994—），男，贵州安顺人，硕士，研究方向：大数据、人工智能。

（收稿日期：2025-11-19 修回日期：2026-07-08）