

基于不同机器学习算法的地浸 测井数据岩性识别对比

邹玉涵¹, 李沁慈^{1,2}, 阙为民^{1,2}, 黄亮³, 吴童盼⁴, 杜志明^{1,2}, 姜振蛟³, 陈炫沂³

(1. 核工业北京化工冶金研究院, 北京 101149;

2. 中核矿业科技集团有限公司, 北京 101149;

3. 吉林大学 地下水资源与环境教育部重点实验室, 吉林 长春 130015;

4. 南华大学 资源环境与安全工程学院, 湖南 衡阳 421001)

摘要: 机器学习算法可从大量地质数据中自动学习和提取特征, 进而实现对岩性的快速准确识别。以内蒙古某砂岩型铀矿区为研究对象, 将数口井的测井数据随机划分为训练集和验证集, 通过调整模型结构和优化超参数进行训练, 利用 BC1401、BC2802、BC4603 和 BC7206 等 4 口井进行测试, 实现对随机森林、XGBoost、K 值邻近算法、BP 网络和 SMOTE-LSTM 算法等模型应用效果的对比。分析结果显示, SMOTE-LSTM 模型应用的稳定性和准确度最优, 准确率为 84.6%。

关键词: 地浸采铀; 机器学习; 砂岩型铀矿; 岩性识别; 测井数据; SMOTE-LSTM 模型; XGBoost 模型

中图分类号: TD853; TQ028; TL212 **文献标志码:** A **文章编号:** 1000-8063(2025)02-0009-09

DOI: 10.13426/j.cnki.yky.2024.10.09

天然铀是重要的矿产能源, 是核工业发展的物质基础和国家紧缺的战略资源。目前, 国内天然铀主要来源于砂岩型铀矿^[1]。砂岩型铀矿储层中的岩性空间结构控制着铀矿富集位置, 影响天然状态下和地浸采铀过程中的渗流和溶质运移过程^[2-4]。因此, 高精度识别铀矿的岩性结构对于提高砂岩型铀矿开采的靶向性和开采效率具有重要意义。

传统的岩性识别方法主要有岩芯观察和薄片鉴定, 存在勘探成本高、工程周期长问题; 在矿产开发过程中, 若要实现全井段取芯的岩性识别工作, 耗费的人力与时间成本较高^[5]。此外, 测井数据的处理与分析也是岩性识别的一个重要途径。传统的基于测井数据的岩性识别方法主要有交会图法^[6]、概率统计法^[7]和聚类分析法^[8-9]等; 但面对复杂的地下地质环境, 通过建立测井数据与地层岩性之间的数学关系来映射实际的地下地质情况仍存在较大困难。

机器学习算法能从大量地质数据中自动学习和提取特征, 实现对岩性的快速准确识别, 其已被广泛利用于测井数据识别岩性领域。安鹏等利用深度卷积神经网络构建了一个石油地球物理勘探领域的岩性识别神经网络模型, 提高了岩性识别的效率和准确性^[10]。段友祥等通过结合不同的机器学习模型, 创建了一个基于油藏地质领域异构多层集成学习的模型, 显著提升了岩性识别的效率和准确度^[11]。岳中文等利用机器学习算法分析随钻数据, 比较了各类算法在岩性识别应用中的优缺点, 为岩性识别领域下一步的发展方向提出了建议^[12]。刘颖等利用监督学习理论与机器学习算法进行不整合面型铀矿找矿预测研究, 并根据最优模型进一步划分成矿远景区^[13]。叶玉金等利用机器学习算法对铀矿 γ 能谱测井的能谱放射性核素的定量解释进行研究, 为铀矿 γ 能谱测井定量解释方法的选择提供了新思路^[14]。Wang 等探讨了如何利用机器学习算法优化地浸

收稿日期: 2024-10-18

基金项目: 中核集团青年英才项目(A101-8, 基于地浸钻孔数据的采区多源信息智能化提取研究)。

第一作者简介: 邹玉涵(1998—), 女, 四川乐山人, 在读硕士, 主要研究方向为地浸采铀工艺技术。

通信作者简介: 阙为民(1965—), 男, 湖南衡阳人, 研究员级高级工程师, 主要从事地浸采铀科研和管理工作。

采铀过程中的关键参数,以提高铀的回收效率^[15]。Kuchin 等利用机器学习的方法,对哈萨克斯坦 Inkai 铀矿床的勘探数据进行研究,提高对铀矿床宿主岩石过滤特性评估的准确性^[16]。综上所述,机器学习算法已在地学领域有了一定的研究基础,但在铀矿开采的岩性识别方面仍有待深入研究。因此,建立一套准确性高、应用性强的砂岩型铀矿岩性智能识别方法极为重要。

笔者利用随机森林算法(Random Forest, RF)、极端梯度提升算法(Extreme Gradient Boosting, XGBoost)、K 值邻近算法(K-Nearest Neighbor, KNN)、前馈神经网络算法(Back Propagation Neural Network, BP)及 SMOTE-LSTM 算法等 5 种模型对内蒙古某砂岩型铀矿区测井数据开展岩性识别计算,对比分析它们在岩性识别方面的优劣,为后续进行砂岩型铀矿区岩性识别方法的选择提供参考。

1 机器学习算法的基本原理

1.1 随机森林算法

随机森林算法是一种基于集成学习的监督式机器学习算法^[17]。基本单元是决策树,通过集成多个决策树而形成整体模型(图 1)。算法步骤如下^[18]:1)从训练集中随机抽取样本(有放回),形成 n 个子训练集,作为决策树根节点处的样本;2)决策树通过样本的属性进行分裂,划分训练集,每次节点分裂时仅考虑部分特征,而非全部特征,以减少特定特征对模型的影响;3)按照以上步骤分裂出大量决策树,直至不再分裂(子树与父树属性相同时),形成 k 个独立的决策树;4)集成随机森林网络,最终结果由多棵树投票决定。随机森林通过集成多个决策树来降低过拟合风险,并利用随机性增强模型的泛化能力。

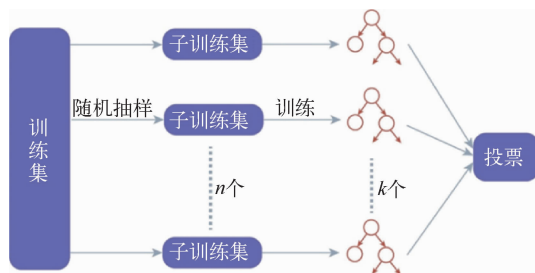


图 1 随机森林算法分类示意图^{[19]35}

Fig. 1 Random forest algorithm classification diagram

1.2 极端梯度提升算法

极端梯度提升算法是一种基于决策树的集成机器学习算法^[20],是由梯度提升决策树算法(Gradient Boosting Decision Tree, GBDT)发展而来,其优化了性能和效率,具备高运行速度和稳定的预测精度^[21]。与 GBDT 相比, XGBoost 对损失函数进行二阶泰勒展开,还支持自定义损失函数,可解决更复杂的问题。同时, XGBoost 将正则化项加入到损失函数中,可有效防止过拟合。在模型训练过程中,目标函数包括梯度提升算法损失项和正则化项,见式(1)。正则化项可以表示模型的复杂程度,见式(2)。

$$L(\Phi) = \sum_{i=1}^n l(y'_i, y_i) + \sum_k f(k), \quad (1)$$

式中: n 表示训练样本总数; y'_i 表示训练样本的模型预测标签值; y_i 表示训练样本的真实标签值; l 为凸函数,表示单个样本的损失值; k 为决策树中正则化项的数量。

$$f(k) = \alpha T + \frac{1}{2} \beta \sum_{j=1}^T \omega_j^2, \quad (2)$$

式中: α 和 β 是正则项的参数,可以手动设置; ω 是由决策树中所有叶子节点的值形成的向量; T 是叶子节点的个数。

1.3 K 值邻近算法

K 值邻近算法是一种较为简单且实用的监督学习分类算法^[22],主要思想是求出预测数据集与训练数据集的每个点之间的距离,取前 K 个数据作为结果集,选择结果集中出现最多的类别标记作为预测结果(图 2)。

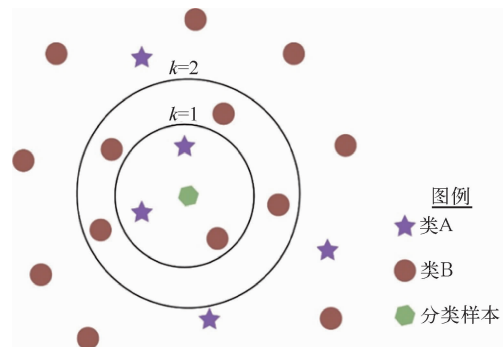


图 2 KNN 算法示意图^{[19]37}

Fig. 2 K-Nearest neighbors algorithm diagram

1.4 前馈神经网络算法

前馈神经网络算法是在人工神经网络的基础

上增加了误差反向传播和梯度下降的方法^[23]。算法步骤如下:在信息从输入层通过隐藏层到达输出层后,将输出的数据与真实数据进行对比;随后误差进行反向传播,从输出层到隐藏层,最后到输入层;通过梯度下降法依次调节隐藏层到输出层和输入层到隐藏层的权重和偏置系数,直到输

出结果和期望误差达到一定的标准(图 3)。BP 神经网络算法模型,通过反向传播和梯度下降法进行神经网络模型中的权重更新,这个过程也被称为“训练”,经神经网络参数的不断更新和调整,最终得到一个训练好的神经网络模型。

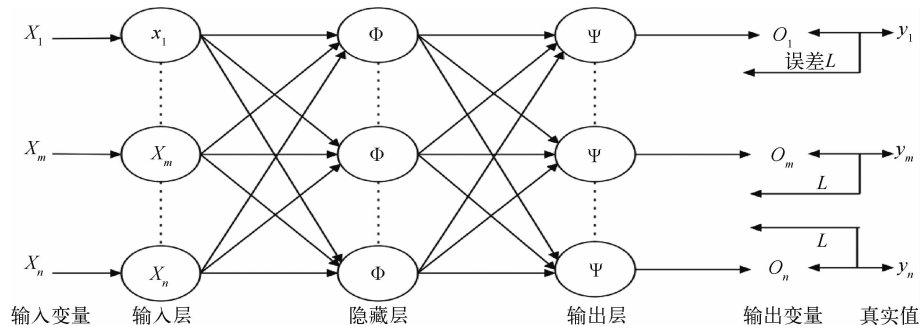


图 3 BP 神经网络结构示意图^{[17]39}

Fig. 3 Back propagation neural network structure diagram

1.5 SMOTE-LSTM 算法

1.5.1 SMOTE 算法

SMOTE 是一种用于处理不平衡数据集的过采样技术,适用于在分类问题中增强少数类别的样本数量。通过在少数类别的样本之间生成合成样本,改善模型训练过程中的类别不平衡问题,从而提高模型对少数类别的识别能力^[24]。该算法采用了 K 值邻近技术,算法流程如下(图 4):1)在少数类样本中随机选择一个样本;2)在该样本附近找到 K 个邻近样本(以 K=3 为例);3)随机在 K 个邻近样本中选出一个样本;4)在该样本点和随机选出的样本点之间连线,并在该线上随机找一个点,将其定义为人工合成的新样本点。

1.5.2 LSTM 算法

LSTM 是 RNN 算法的改进,通过增加单元状态实现信息的遗忘和保留,能很好地处理和预测较长时间序列^[25]。LSTM 记忆单元由遗忘门、输入门和输出门 3 种门结构组成,3 种门结构通过抛弃无用信息、增添有用信息来保证信息选择性通过,使记忆单元状态不断更新,达到控制和保护记忆单元状态的目的^[26]。LSTM 神经网络的记忆单元结构见图 5。具体的计算流程见式(3)~式(8)。

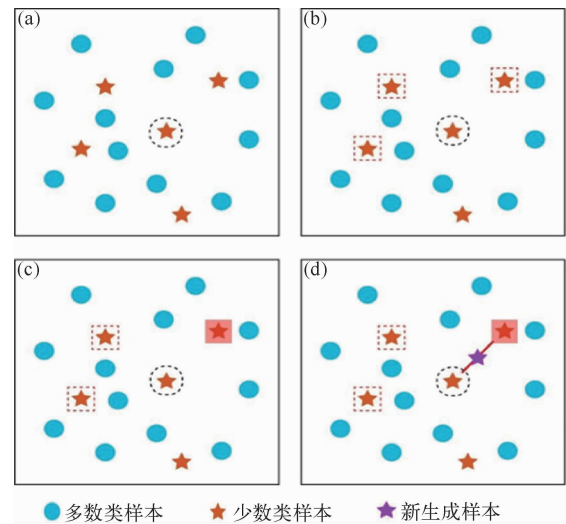


图 4 SMOTE 算法流程示意图^{[19]30}

Fig. 4 SMOTE algorithm process diagram

利用 $t-1$ 时刻单元的输出计算得到遗忘门的结果

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f); \quad (3)$$

利用 $t-1$ 时刻单元的输出计算得到输入门的结果

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i); \quad (4)$$

利用 $t-1$ 时刻单元的输出计算得到候选值的结果

$$\tilde{C}_t = \tanh(W_c[h_{t-1}, x_t] + b_c); \quad (5)$$

利用 $t-1$ 时刻的单元状态, 遗忘门、输入门和候选值更新单元状态

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t; \quad (6)$$

利用 $t-1$ 时刻单元输出计算得到输出门的结果

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o); \quad (7)$$

利用输出门和单元状态计算此时刻单元的输出

$$h_t = o_t * \tanh(C_t); \quad (8)$$

式中: f_t 表示遗忘门; i_t 表示输入门; o_t 表示输出门; h_{t-1} 表示 $t-1$ 时刻单元的输出; h_t 表示 t 时刻单元的输出; x_t 表示 t 时刻的输入; C_{t-1} 表示 $t-1$ 时刻的单元状态; C_t 表示 t 时刻的单元状态; \tilde{C}_t 表示候选值; σ 表示 sigmoid 函数; W_f 、 W_i 、 W_c 、 W_o 分别表示遗忘门、输入门、候选值、输出门的权重系数; b_f 、 b_i 、 b_c 、 b_o 分别表示遗忘门、输入门、候选值、输出门的偏移系数。

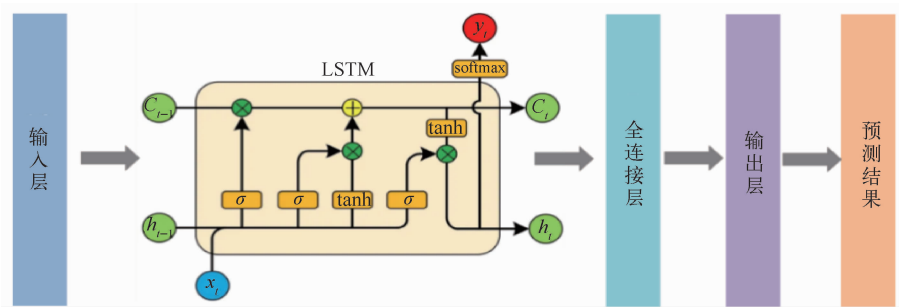


图5 LSTM神经网络结构示意图^{[17]32}

Fig. 5 LSTM Neural network structure diagram

SMOTE-LSTM 算法输入数据以深度为索引项, 以不同深度测井参数为特征数据, 以岩性编录数据为目标数据。由于利用了多个测井数据进行岩性预测, 输入测井数据为多维度序列数据, 在归一化基础上, 先对数据进行 SMOTE 处理, 之后再依据 LSTM 神经网络对测试集测井数据进行预测。

2 数据预处理及评估指标

2.1 数据来源

本研究所使用的数据源自内蒙古某砂岩型铀矿区生产井的测井数据和对应的岩性编录数据, 选取其中 9 口井的数据, 共计 6 900 个样本。这些数据涵盖了多种岩性, 包括泥岩、粉砂质泥岩、含砾细砂岩、含砾中砂岩、砂砾岩等(表 1)。为提高模型的可训练能力, 对研究区的岩性分类进行简化, 并赋予相应的数字符号: 泥岩、粉砂质泥岩归为一类, 符号为 1; 含砾细砂岩为一类, 符号为 2; 砂砾岩和含砾中砂岩归为一类, 符号为 3。按照 7:2 的比例将数据整体划分为训练集和验证

集, 其中数据选取的方式为随机取样, 以保证算法模型识别的广泛性和适用性。然后, 依托选取的数据样本依次建立随机森林、XGBoost 算法、K 值邻近算法、BP 神经网络及 SMOTE-LSTM 算法等 5 种算法模型。

2.2 数据处理方法

选取的生产井测井曲线类别包括自然伽马 (GR)、井径 (CAL)、自然电位 (SP) 和三侧向电阻率 (Rt)。由于不同的测井数据之间存在量纲不同和数值差异的问题, 若将不加以处理的测井数据直接输入模型, 会导致数值较小的测井数据被算法模型忽略, 而数值较高的测井数据会出现过拟合的现象。这会对研究区的岩性识别和预测产生较大影响, 因此需要对原始数据进行异常值和归一化预处理。

1) 异常值处理

对原始数据两端存在的异常、未定义值, 按照经验可直接对其进行剔除; 对测井数据中的极端值进行剔除, 采用插值进行补充, 避免其影响模型对数据的敏感度而降低预测结果的准确度。

表 1 部分测井岩性数据类别统计

Table 1 Classification statistics of some logging lithology data

岩性类别	BC7004		BC7005		BC7006	
	样本数	占比/%	样本数	占比/%	样本数	占比/%
泥岩	269	23.1	264	24.5	298	27.4
粉砂质泥岩	135	11.5	20	1.8	172	15.8
含砾细砂岩	254	21.8	208	19.3	428	39.4
含砾中砂岩	307	26.3	303	28.2	76	7.0
砂砾岩	151	17.3	282	26.2	113	10.4
总计	1 166	100.0	1 076	100.0	1 087	100.0
岩性类别	BC7007		BC7104		BC7105	
	样本数	占比/%	样本数	占比/%	样本数	占比/%
泥岩	375	34.5	174	15.8	229	21.2
粉砂质泥岩	105	9.6	262	23.7	117	10.8
含砾细砂岩	338	31.2	162	14.7	250	23.1
含砾中砂岩	133	12.2	324	29.4	320	29.5
砂砾岩	137	12.5	181	16.4	167	15.4
总计	1 088	100.0	1 103	100.0	1 083	100.0
岩性类别	BC7106		BC7108		BC7204	
	样本数	占比/%	样本数	占比/%	样本数	占比/%
泥岩	291	26.8	300	27.8	105	9.4
粉砂质泥岩	45	4.3	81	7.5	290	26.0
含砾细砂岩	258	23.9	213	19.8	248	22.2
含砾中砂岩	289	26.7	313	29.0	293	26.2
砂砾岩	198	18.3	172	15.9	181	16.2
总计	1 081	100.0	1 079	100.0	1 117	100.0

2) 归一化处理

将数据按比例缩放,便于不同单位或量级的指标能够进行比较和加权,其目的是将数据映射到指定的范围,通常是 $[0,1]$,见式(9)。

$$\hat{x}_i = \frac{x_i - \min_{1 \leq j \leq n} \{x_j\}}{\max_{1 \leq j \leq n} \{x_j\} - \min_{1 \leq j \leq n} \{x_j\}}, \quad (9)$$

式中: \hat{x}_i 是归一化之后的数值, x_i 是归一化之前的数值, $\max_{1 \leq j \leq n} \{x_j\}$ 是整个数据集的最大值, $\min_{1 \leq j \leq n} \{x_j\}$ 是整个数据集的最小值, n 是数据集的长度。

测井数据通过 Min-Max 标准化处理,数值最终分布于 $[0,1]$ 区间内,归一化处理不会改变数据的相对大小,归一化之后的数值可作为岩性预测

模型的输入数据。由模型对输入的多维度序列数据进行分析处理,从而得到测井岩性编录的预测结果,最终基于评价指标量化模型准确度。

模型的分类预测性能和准确率与模型的结构和超参数的选择均息息相关,在对数据进行预处理的基础上,以准确率和精确率作为量化依据,对比 5 种模型在不同超参数组合条件下的岩性预测结果,选择最优的超参数组合,优化模型结构。其中,RF 模型选择的超参数包括树的数量值(12)、最大深度值(7)、随机数生成器种子值(11)、叶节点最小样本数值(2);XGBoost 模型选择的超参数包括子节点最小权重参数值(6)、最大深度值(15)、分类数值(4);KNN 模型选择的超参数包括最近邻居数量值(10);BP 神经网络模型选择的超

参数包括学习率值(0.001)、迭代次数值(25)、单次训练样本数量值(28);SMOTE-LSTM模型选择的超参数包括学习率值(0.001)、时间步长值(6)、迭代次数值(100)、单次训练样本数量值(28)。

2.3 评价指标

评价指标采用准确率(Accuracy)和精确率(Precision),量化模型的准确性和稳定性。准确率是正确分类的样本占样本总数的比例,精确率是预测的正样本中实际为正的样本的百分比,以二分类为例(类1和类2),准确率(A)的计算式为^{[19]39}

$$A = \frac{N_1 + F_1}{N_1 + N_2 + F_1 + F_2}, \quad (10)$$

精确率(P)的计算式为^{[19]40}

$$P = \frac{N_1}{N_1 + F_2}, \quad (11)$$

式中: N_1 是原本是类1的样本被预测为类1的样本数量; N_2 是原本是类1的样本被预测为类2的样本数量; F_1 是原本是类2的样本被预测为类2的样本数量; F_2 是原本是类2的样本被预测为类1的样本数量。

3 岩性识别算法的应用效果对比

为对比不同机器学习算法在砂岩型铀矿岩性识别的应用效果,用该研究区内BC1401、BC2802、BC4603和BC7206等4口井的测井数据对已训练好的RF模型、XGBoost模型、KNN模型、BP神经网络模型及SMOTE-LSTM模型的岩性分类效果进行对比研究。

3.1 算法预测准确率、精确率对比

对4口井预测的准确率和精确率见表2。可以看出,在BC1401井中,SMOTE-LSTM的预测准确率略低于BP神经网络模型;但在BC2802、BC4603、BC7206测试井中,SMOTE-LSTM算法的准确率均高于BP神经网络模型,并大幅超过其他三类模型;SMOTE-LSTM模型的平均准确率最高(达84.6%),且最为稳定。

XGBoost模型与RF模型、KNN模型、BP神经网络模型相比,在BC1401和BC4603井中,XGBoost模型的准确率低于BP神经网络模型,但远高于RF模型和KNN模型;在BC2802和BC7206中,XGBoost模型的准确率均高于其他三类模型,平均准确率为81.4%。

测试结果表明,SMOTE-LSTM模型预测的应用效果最好,这5种模型的准确率由高到低依次为:SMOTE-LSTM模型、XGBoost模型、BP神经网络模型、RF模型、KNN模型。

根据4口验证井的精确率对比结果可知,XGBoost模型和BP神经网络模型的精确率均高于RF模型和KNN模型。在BC2802和BC7406中,XGBoost模型的精确率高于BP神经网络模型,但在其他两口井中,BP神经网络模型的精确率高于XGBoost模型;而SMOTE-LSTM模型的精确率始终处于最优。测试结果表明,针对4口验证井的平均精确率由高到低依次为:SMOTE-LSTM模型(83.6%)、BP神经网络模型(80.5%)、XGBoost模型(79.4%)、RF模型(72.0%)、KNN模型(65.0%)。

表2 5种岩性识别模型的准确率、精确率

Table 2 Accuracy and precision of 5 lithology identification models

算法模型	BC1401		BC2802		BC4603		BC7206	
	准确率/%	精确率/%	准确率/%	精确率/%	准确率/%	精确率/%	准确率/%	精确率/%
SMOTE-LSTM	84.8	83.8	83.6	82.5	85.1	83.1	85.0	84.8
KNN	67.8	62.8	63.4	63.3	64.6	65.2	61.8	68.8
BP	86.1	82.7	71.7	78.2	84.7	82.4	78.9	78.5
RF	70.1	70.2	71.6	71.1	71.7	71.8	75.2	74.7
XGBoost	78.5	71.2	80.0	81.6	84.5	82.1	82.5	82.5

3.2 岩性预测效果对比

对5种不同模型预测的岩性结果和真实测井

岩性进行对比,结果见图6。SMOTE-LSTM、XGBoost和BP模型的预测结果与真实岩性基本

一致,RF 和 KNN 模型预测结果与真实岩性的偏差较大,预测结果误差集中在将泥岩预测为含砾细砂岩。此外,5 种模型对于薄层岩性的预测均具有一定准确性,但 RF 和 KNN 预测模型中出

现较多预测错误的薄岩层,XGBoost 和 BP 网络模型对一些薄层岩性也未能做出准确的预测。相对而言,SMOTE-LSTM 模型对于薄层岩性的预测更为全面与准确。

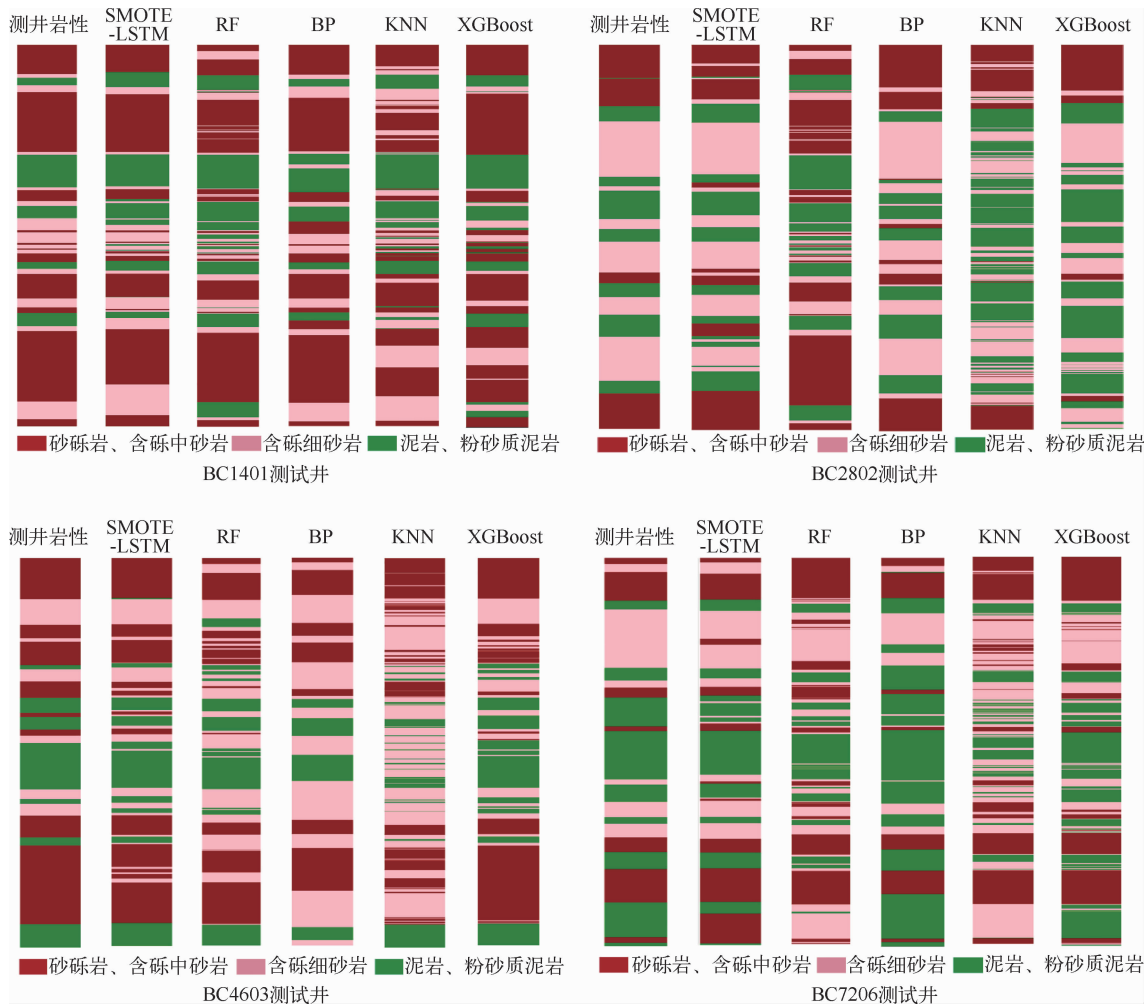


图 6 测试井预测岩性与真实岩性对比

Fig. 6 Comparison between predicted lithology and true lithology in test wells

总体来说,与其他 4 种模型相比,SMOTE-LSTM 模型在 4 口测试井的岩性识别上的准确率和精确率最高,预测性能最稳定,表现出一定的优越性。这主要是由于 SMOTE-LSTM 模型可以利用 SMOTE 算法来平衡样本,使得模型不会偏向于多数类样本,同时可利用 LSTM 算法学习数据的复杂时序与非线性关系,从而通过两种算法的结合提高模型的预测准确性和泛化能力,并实现对一维钻孔岩性的精细刻画。

4 结论与建议

4.1 结论

利用随机森林、XGBoost、K 值邻近、BP 神经网络及 SMOTE-LSTM 等机器学习算法,依托研究区内 9 口具有典型代表的测井数据,调整模型结构和优化超参数进行训练,建立了内蒙古某砂岩型铀矿区泥岩、粉砂质泥岩、含砾细砂岩、含砾中砂岩、砂砾岩等 5 种岩性和自然伽马、井径、自

然电位、三侧向电阻率等4种测井特征曲线之间的非线性映射关系;并对研究区内的BC1401、BC2802、BC4603和BC7206等4口井进行岩性预测,比较了5种机器学习算法识别的效果。结果显示:SMOTE-LSTM的预测准确率最高,为84.6%。SMOTE-LSTM算法模型与其他4种模型相比,性能最优,研究结果为砂岩型铀矿的岩性识别提供了一个新的思路。

4.2 建议

本次研究区域仅局限于目标砂岩型铀矿区,未能对其他区块的砂岩型铀矿进行试验,模型的可复制性有待进一步检验。此外,在超参数优化过程中,由于不同模型需要调节的参数不同,未来可将寻找更适合不同机器学习算法进行超参数选择的方法,作为研究方向。

参考文献:

- [1] 阙为民,王海峰,田时丰,等.我国地浸采铀研究现状与发展[J].铀矿冶,2005,24(3):113-117.
QUE Weimin, WANG Haifeng, TIAN Shifeng, et al. Research status and development of in-situ leaching uranium techniques in China[J]. Uranium Mining and Metallurgy, 2005, 24(3): 113-117 (in Chinese).
- [2] 焦养泉,吴立群,荣辉,等.沉积、成岩与铀成矿:中国砂岩型铀矿研究的创新发现与认知挑战[J].地球科学,2022,47(10):3580-3602.
JIAO Yangquan, WU Liqun, RONG Hui, et al. Sedimentation, diagenesis and uranium mineralization: Innovative discoveries and cognitive challenges in study of sandstone-type uranium deposits in China [J]. Earth Science, 2022, 47(10): 3580-3602 (in Chinese).
- [3] 张友澎,赵利信,王亚奴,等.高能气体致裂技术在低渗砂岩型铀矿地浸开采中的应用[J].铀矿冶,2024,43(3):1-8.
ZHANG Youpeng, ZHAO Lixin, WANG Yanu, et al. Application of high energy gas fracturing in in-situ leaching of low-permeable sandstone uranium deposit[J]. Uranium Mining and Metallurgy, 2024, 43(3): 1-8 (in Chinese).
- [4] 禹宝利,张振强.松辽盆地南部钱家店凹陷姚家组含水层水成铀矿作用分析[J].铀矿冶,2022,41(4):346-351.
YU Baoli, ZHANG Zhenqiang. Analysis on hydrogenic uranium mineralization of Yaojia Formation aquifer in Qianjiadian Depression of South Songliao Basin[J]. Uranium Mining and Metallurgy, 2022, 41(4): 346-351 (in Chinese).
- [5] 付光明,严加永,张昆,等.岩性识别技术现状与进展[J].地球物理学进展,2017,32(1):26-40.
FU Guangming, YAN Jiayong, ZHANG Kun, et al. Current status and progress of lithology identification technology[J]. Progress in Geophysics, 2017, 32(1): 26-40 (in Chinese).
- [6] 徐德龙,李涛,黄宝华,等.利用交会图法识别国外M油田岩性与流体类型的研究[J].地球物理学进展,2012,27(3):1123-1132.
XU Delong, LI Tao, HUANG Baohua, et al. Research on the identification of the lithology and fluid type of foreign M oilfield by using the crossplot method[J]. Progress in Geophysics, 2012, 27(3): 1123-1132 (in Chinese).
- [7] 刘子云,王向公.利用概率统计方法判断岩性[J].江汉石油学院学报,1989(2):35-40.
LIU Ziyun, WANG Xiangong. Determination of lithology through probability statistics[J]. Journal of Oil and Gas Technology, 1989(2): 35-40 (in Chinese).
- [8] 寻知锋,余继峰.聚类和判别分析在测井岩性识别中的应用[J].山东科技大学学报(自然科学版),2008(5):10-13.
XUN Zhifeng, YU Jifeng. The application of cluster and discriminant analyses in logging lithology recognition[J]. Journal of Shandong University(Natural Science), 2008(5): 10-13 (in Chinese).
- [9] 关涛.基于交会图和贝叶斯聚类分析法的岩性识别方法[J].科学技术与工程,2013,13(4):976-979.
GUAN Tao. Method of lithologic identification based on crossplot and bayesian cluster analysis algorithm[J]. Science Technology and Engineering, 2013, 13(4): 976-979 (in Chinese).
- [10] 安鹏,曹丹平.基于深度学习的测井岩性识别方法研究与应用[J].地球物理学进展,2018,33(3):1029-1034.
AN Peng, CAO Danping. Research and application of logging lithology identification based on deep learning[J]. Progress in Geophysics, 2018, 33(3): 1029-1034 (in Chinese).
- [11] 段友祥,赵云山,马存飞,等.基于多层集成学习的岩性识别方法[J].数据采集与处理,2020,35(3):572-581.

- DUAN Youxiang, ZHAO Yunshan, MA Cunfei. Lithology identification method based on multi-layer ensemble learning[J]. *Journal of Data Acquisition and Processing*, 2020, 35(3): 572-581 (in Chinese).
- [12] 岳中文, 闫逸飞, 王煦, 等. 基于随钻数据的岩性识别机器学习算法研究进展[J]. *科学技术与工程*, 2023, 23(10): 4044-4057.
- YUE Zhongwen, YAN Yifei, WANG Xu, et al. Research progress of machine learning algorithm for lithology identification based on data while drilling[J]. *Science Technology and Engineering*, 2023, 23(10): 4044-4057 (in Chinese).
- [13] 刘颖. 基于机器学习的加拿大阿萨斯卡盆地东南部铀成矿预测与评价[D]. 抚州: 东华理工大学, 2022.
- [14] 叶玉金. 基于机器学习算法的铀矿 γ 能谱测井定量解释研究[D]. 抚州: 东华理工大学, 2023.
- [15] WANG B, LUO Y, QIAN J Z, et al. Machine learning-based optimal design of the in-situ leaching process parameter (ISLPP) for the acid in-situ leaching of uranium[J]. *Journal of Hydrology*, 2023, 626(Pt. A).
- [16] KUCHIN Y, MUKHAMEDIEV R, YUNICHEVA N, et al. Application of machine learning methods to assess filtration properties of host rocks of uranium deposits in Kazakhstan[J]. *Applied Sciences-Basel*, 2023, 13(19): 12.
- [17] LU X, WANG P, NIYATO D, et al. Wireless networks with RF energy harvesting: A contemporary survey[J]. *IEEE Communications Surveys and Tutorials*, 2015, 17(2): 757-789.
- [18] 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.
- JIANG Li, ZHANG Zhimo, WANG Qiwei, et al. Comparative study on lithology classification of oil logging data based on different machine learning models[J]. *Geophysical and Geochemical Exploration*, 2024, 48(2): 489-497 (in Chinese).
- [19] 陈炫沂. 砂岩型铀矿潜力区岩性特征智能识别方法研究[D]. 长春: 吉林大学, 2023.
- [20] SAGI O, ROKACH L. Approximating XGBoost with an interpretable decision tree[J]. *Information sciences*, 2021, 572: 522-542.
- [21] ASSELMAN A, KHALDI M, Aammou S. Enhancing the prediction of student performance based on the machine learning XGBoost algorithm[J]. *Interactive Learning Environments*, 2023, 31(6): 3360-3379.
- [22] CHEN, Q F, LI D, TANG C K, et al. KNN matting[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 35(9): 2175-2188.
- [23] MONICA B, GIOVANNI Z, AUREL P, et al. The 4.2 ka BP Event in the Mediterranean region: An overview[J]. *Climate of the Past*, 2019, 15(2): 555-577.
- [24] CHAWLA N V, BOWYER K W, HALL L O, et al. SMOTE: Synthetic minority over-sampling technique[J]. *Journal of Artificial Intelligence Research*, 2002, 16: 321-357.
- [25] YU Y, SI X S, HU C H, et al. A review of recurrent neural networks: LSTM cells and network architectures[J]. *Neural computation*, 2019, 31(7): 1235-1270.
- [26] 潘少伟, 王朝阳, 张允, 等. 基于长短期记忆神经网络补全测井曲线和混合优化 XGBoost 的岩性识别[J]. *中国石油大学学报(自然科学版)*, 2022, 46(3): 62-71.
- PAN Shaowei, WANG Chaoyang, ZHANG Yun, et al. Lithology identification based on LSTM neural networks completing log and hybrid optimized XGBoost[J]. *Journal of China University of Petroleum (Edition of Natural Sciences)*, 2022, 46(3): 62-71 (in Chinese).

(下转第 37 页)

