

梁嘉骏,周小淇,史培新.面向钻孔策略优化的机器学习三维地层建模研究[J].辽宁工程技术大学学报(自然科学版),2026,45(4): 426-435. DOI:10.11956/j.issn.1008-0562.20260046

Liang Jiajun, Zhou Xiaoqi, Shi Peixin. Three-dimensional stratigraphic modeling for borehole scheme optimization using machine learning method[J]. Journal of Liaoning Technical University (Natural Science), 2026, 45(4): 426-435. DOI: 10.11956/j.issn.1008-0562.20260046

面向钻孔策略优化的机器学习三维地层建模研究

梁嘉骏,周小淇*,史培新

(苏州大学 轨道交通学院,江苏 苏州 215000)

摘要: 为降低勘探数据稀疏对地层模型准确性的影响,提出一种基于多信度地勘数据的集成学习三维地层隐式建模方法。该方法采用遗传算法(GA)构建优化的集成学习(Stacking)模型,学习基于地质勘探数据的地层分布特征,生成与虚拟钻孔对应的地层分类数据集;采用基于多二次核的径向基函数隐式建模方法构建地层模型。在此基础上,通过信息熵量化建模过程的不确定性,优化钻孔布设方案,并以某城市轨道交通地勘数据为依托开展实例验证。结果表明:GA-Stacking模型在测试集上的F1分数与准确率分别达到90%与89%;相比单一机器学习模型,该模型在地层分类预测中具有更高的分类准确性;模型生成的地层分布与地层基准剖面的平均绝对误差为0.23 m,剖面吻合度平均为85%,证实该模型的可靠性;新增5个钻孔后地层整体不确定性降低24.1%,验证了所提方案的有效性。研究结论为三维地层精细建模提供新的途径,并为地质勘探布孔策略提供理论支撑。

关键词: 三维地层建模;集成学习;多信度数据融合;不确定性量化;布孔策略优化

中图分类号:TU195

文献标志码:A

文章编号:1008-0562(2026)04-0426-10

Three-dimensional stratigraphic modeling for borehole scheme optimization using machine learning method

Liang Jiajun, Zhou Xiaoqi*, Shi Peixin

(School of Rail Transportation, Soochow University, Suzhou 215000, China)

Abstract: To reduce the adverse impact of sparse geological exploration data on the precision of stratigraphic models, this study proposes an implicit three-dimensional (3D) stratigraphic modeling method integrating ensemble learning and multi-fidelity geological survey data. A genetic algorithm (GA) is utilized to construct an optimized Stacking-based ensemble learning framework, which learns stratigraphic distribution patterns from raw geological exploration data and generates stratigraphic classification datasets corresponding to virtual boreholes. A radial basis function with a multiquadric kernel is adopted for implicit stratigraphic modeling. On this basis, information entropy is introduced to quantify the modeling uncertainty and further optimize the borehole layout scheme. A case study is conducted using geological survey data from an urban rail transit project. The results indicate that the GA-Stacking model achieves an F1-score of 90% and an accuracy of 89% on the test dataset. Compared with standalone machine learning models, the proposed method exhibits superior performance in stratigraphic classification and prediction. The mean absolute error between the modeled stratigraphic distribution and benchmark stratigraphic profiles is 0.23 m, with an average profile consistency of 85%, demonstrating the reliability of the developed model. Moreover, the overall stratigraphic uncertainty is reduced by 24.1% after supplementing five additional boreholes, which validates the effectiveness of the proposed optimization strategy. This research provides a novel insight into refined 3D stratigraphic modeling and offers theoretical guidance for borehole layout optimization in geological exploration.

收稿日期:2026-01-25 修回日期:2026-03-18 接受日期:2026-04-10 责任编辑:张雯

基金项目:国家自然科学基金面上项目(52278405)

作者简介:梁嘉骏(2001-),男,甘肃天水人,硕士研究生,主要从事地层建模方面的研究。E-mail: 2407353978@qq.com

通信作者:周小淇(1998-),女,江苏徐州人,博士,讲师,主要从事人工智能与地下工程方面的研究。E-mail: xqzhou@suda.edu.cn

Keywords: three-dimensional stratigraphic modeling; ensemble learning; multi-fidelity data fusion; uncertainty quantification; borehole layout optimization

0 引言

城市土地资源的紧缺与地下建造技术的进步推动地下空间开发利用的高速发展。相较于地面工程,地下工程修建于岩土体中,具有造价高、建设环境复杂、质量要求高的特点^[1]。岩土体埋于地下,无法直接观测其分布与性质^[2]。开展地质勘探,确定每层岩土体的组分与工程性质,建立准确的地层模型,是地下工程规划、设计的基础,也是施工、运营安全的保障^[3]。

传统地层建模方法主要依托工程地质勘探获取的实测数据,由工程技术人员手动连接相邻地层分界点,形成用于工程设计的地层剖面,并赋予每个地层相应的性质参数^[4]。该方法需要大量人机交互,且受勘探点稀疏、间距较大的限制,单一的线性连接难以准确反映地层在相邻区域的实际分布。针对这些问题,学者提出隐式建模方法^[5],基于地质统计学与空间插值理论,自动、快速地构建地层曲面模型。然而,隐式建模的性能易受原始数据分布影响,当勘探间距过大时,传统插值方法难以准确捕捉地层的局部细节变化,导致建模局部精度不高^[6]。

近年来,机器学习方法在岩土工程领域的应用日益广泛。该算法具备强大的非线性拟合能力,能够深入挖掘数据的潜在模式,整合并分析工程参数间的复杂关联,有效改善传统地层划分中信息处理不足、分析能力有限的问题,实现快速、准确的地层划分^[7]。多项研究已证实机器学习方法在辅助地层建模中的有效性,Kumar等^[8]采用模糊多层感知机(fuzzy multilayer perceptron, F-MLP)模型,结合钻孔数据与人工合成数据,实现钻孔间地层类别的“数据驱动式”插值。Shi等^[9]基于有限的钻孔数据以及训练图像,提出数据驱动的极限梯度提升(extreme gradient boosting, XGB)模型,用于插值地层剖面。Abedi等^[10]将支持向量机(support vector machine, SVM)应用于矿床勘探建模。郭甲腾等^[11]使用SVM与反向传播神经网络(back propagation neural network, BP)实现隐式三维地质建模。王昊等^[12]将BP神经网络用于矿床外围的矿物分布预测。朱峻生等^[13]通过改进的K近邻(K-nearest neighbor, KNN)算法进行地层薄

层预测,提升预测的准确性。机器学习虽然能够有效辅助地层建模,但如何有效提高模型的泛化能力与分类精度,仍是地层建模领域亟待解决的问题。

集成学习通过组合多个学习器提升整体预测能力与泛化性能^[14],该方法在地质数据建模中得到广泛的应用。Jia等^[15]结合地球物理数据,使用Stacking模型学习构建三维地质模型。扶金铭等^[16]利用集成学习模型学习矿体轮廓线的点云数据分布特征,并借助径向基函数构建隐式场,减少曲面边界的自拟合现象,提升模型的连续性。邓怡徽等^[17]使用人工蜂群算法优化基学习器的选择,构建组合模型以提高虚拟钻孔的预测精度。

本研究针对地勘数据稀疏性与不均匀性对地层建模精度的影响,提出一种混合集成学习的三维地层隐式建模方法。该方法采用遗传算法构建优化集成学习模型,学习地层分布特征,生成虚拟钻孔作为原始数据补充。采用径向基函数隐式建模方法构建地层模型,并开展实例验证。同时,基于信息熵指标量化建模过程的不确定性,进行钻孔布设策略优化。

1 多信度地勘数据融合

岩土工程地质勘探的手段主要为现场钻孔和测试,测试包括标准贯入试验(standard penetration testing, SPT)、静力触探试验(cone penetration testing, CPT)与室内试验。SPT核心优势在于土样可见性与室内试验可靠性高,属于高信度数据。但其测点稀疏且数据离散性显著,通常每1~2 m获取一个土样,取样与试验过程的扰动亦可能影响最终结果(见图1)。仅依靠SPT数据与室内试验数据难以实现精确的地层划分。CPT的优势在于快速、经济且数据采集密集(采样间隔0.1 m,见图2),其局限性在于无法获取土样,仅靠经验公式进行土的分类缺乏直观验证,在实际工程中,CPT数据常被视为低信度数据。将CPT数据、SPT数据与室内试验数据相结合,可以有效整合多信度地质勘探数据,从而实现更准确的地层分类,在数据输入层面降低误差,建立更精细的地层模型。

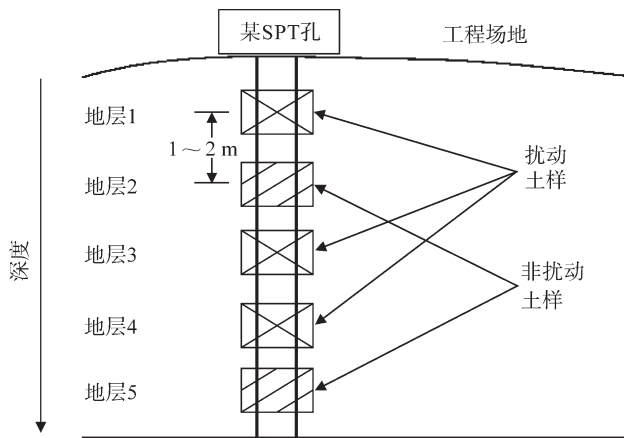


图1 标准贯入试验示意

Fig.1 Schematic of standard penetration test

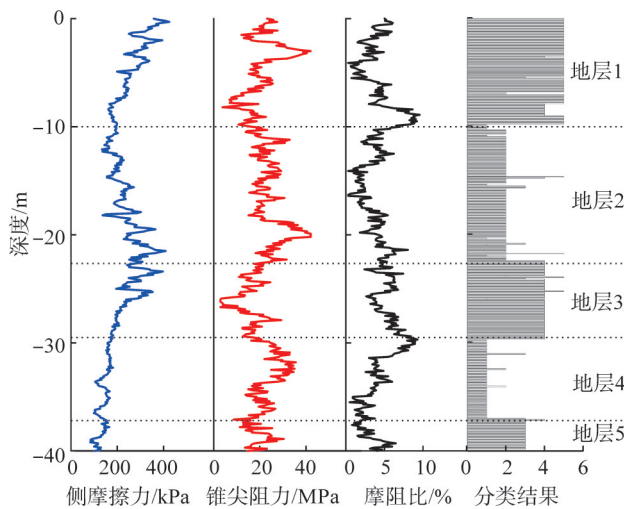


图2 静力触探试验示意

Fig.2 Schematic of cone penetration testing

采用机器学习分类算法,以CPT数据与SPT数据作为输入数据,通过训练模型预测SPT未取样位置的地层信息,见图3。在得到未取样处的预测数据后,将离散的SPT数据、连续的CPT数据、预测的未取样数据、室内试验数据与既有地层资料整合为多信度融合数据,见图4。

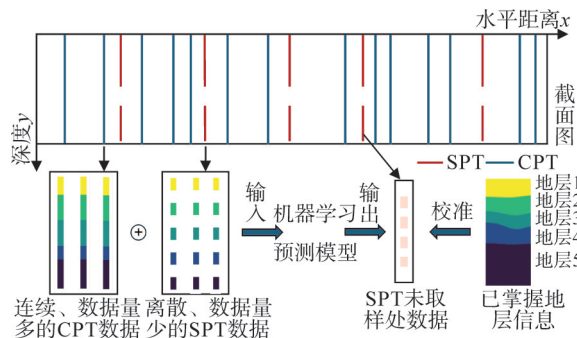


图3 未取样数据预测方法

Fig.3 Prediction method for unsampled data

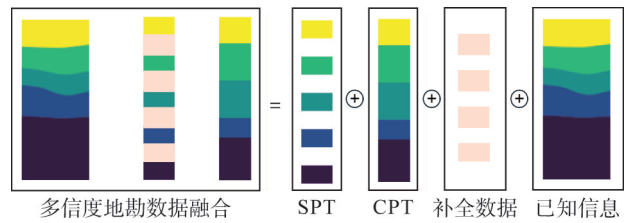


图4 多信度地勘数据融合

Fig.4 Multi-reliability fusion of geological exploration data

2 GA-Stacking集成学习模型框架

集成学习通过组合多个学习器提升整体性能^[18]。Stacking模型采用分层结构,基学习器对原始数据进行学习,元学习器整合其预测结果,进一步提升模型性能。鉴于基学习器与元学习器选择对Stacking模型性能的关键影响,提出一种基于遗传算法与两层Stacking模型的混集成方法。该方法利用遗传算法模拟自然进化中的选择、交叉与变异操作,在候选学习器池中迭代寻优,自动识别最优学习器组合,在保持基学习器多样性的同时,提升模型预测精度与泛化能力,为复杂分类任务提供解决方案。

2.1 Stacking集成学习算法

假设训练数据集为

$$D = \{(a_i, b_i)\}_{i=1}^n, \quad (1)$$

式中: $a_i \in \mathbf{R}^d$ 为第 i 个勘探点的特征向量, d 为特征维度; b_i 为对应的地层类别标签。

Stacking的学习结构通常分为两层。第一层中,基学习器在原始稀疏勘探数据集 D 上训练多个基学习器 $h_1, h_2, \dots, h_k$, 各学习器分别输出各地层的隶属度概率,形成新的特征表示

$$c_i = (h_1(a_i), h_2(a_i), \dots, h_k(a_i)). \quad (2)$$

第二层中,元学习器将基学习器的预测结果 c_i 作为输入特征,训练一个元学习器 H , 得到最终预测结果

$$\hat{b}_i = H(c_i) = H(h_1(a_i), h_2(a_i), \dots, h_k(a_i)). \quad (3)$$

Stacking模型的整体形式为

$$\hat{b} = H(h_1(a), h_2(a), \dots, h_k(a)). \quad (4)$$

该分层结构使元学习器能够自适应学习不同基模型在复杂三维地层空间分布中的决策权重,而非简单投票或加权平均,从而更精准地捕捉地层空间渐变与非线性分界规律。

元学习器的优化目标为最小化整体损失函数,表达式为

$$\hat{L} = \sum_{i=1}^n L(b_i, H(c_i)), \quad (5)$$

式中, L 为交叉熵损失函数,适用于地层多分类任

务,通过最小化 L ,元学习器可实现基模型在三维空间地层隶属度预测中的最优组合,提升稀疏条件下地层建模的稳定性与精度。

基于统计学习、核方法、树模型与线性模型等理论框架,构建包含7种算法的基学习器池,分别为SVM、KNN、轻量梯度提升机(light gradient boosting, LGB)、决策树(decision tree, DT)、XGB、随机森林(random forest, RF)与逻辑回归(logistic regression, LR),多维度捕捉地层空间分布复杂规律。通过集成策略对各基学习器的优势进行互补融合,增强整体模型的泛化能力与分类精度,从而用于复杂地质条件下的地层预测任务。

2.2 遗传算法优化策略

遗传算法(GA)是基于自然选择与遗传学机理的优化算法^[19],通过模拟生物进化中的选择、交叉、变异等过程,在搜索空间中不断进化,逐步逼近全局最优解。为提升Stacking集成模型的性能,将各基学习器的选用状态编码为二进制向量,形成个体染色体,并以模型在验证集上的加权F1值作为适应度函数。在进化过程中,GA通过选择、交叉与变异操作不断迭代更新个体,使得种群逐步向性能更优的学习器组合收敛,得到最优基学习器组合,提高模型整体准确率。

3 建模方法及不确定性量化

3.1 基于径向基函数的三维地层建模方法

GA-Stacking集成算法生成虚拟钻孔后,采用径向基函数算法实现地层自动化构建。对每个勘探孔位置,按照深度对地层数据进行排序,识别连续地层单元的上界面与下界面深度,构建完整的地层界面控制点数据集 S_k 。基于研究区域数据的空间分布特征,确定合理计算边界,并设置缓冲区以消除潜在的边界效应。网格生成采用等间距划分方法。

$$x_{\text{grid}} = W_1(x_{\min} - \delta, x_{\max} + \delta, N), \quad (6)$$

$$y_{\text{grid}} = W_1(y_{\min} - \delta, y_{\max} + \delta, N), \quad (7)$$

式(6)、式(7)中: x_{grid} 、 y_{grid} 为研究区域平面坐标系内 x 、 y 方向上的等间距网格坐标; $x_{\min}$ 、 $x_{\max}$ 、 $y_{\min}$ 、 $y_{\max}$ 依次为原始数据在 x 、 y 方向上的坐标值下限和坐标值上限; $W_1(\cdot)$ 为网格生成函数; δ 为缓冲区长度; N 为网格分辨率。

高分辨率网格划分策略为曲面插值提供密集的计算支撑点,确保界面重建的精度。径向基函数插值算法构建地层界面曲面,基于空间控制点的距离加权求和。对于给定的控制点集 S_k ,插值函数为

$$f(x, y) = \sum_{i=1}^n \lambda_i \phi(\|(x, y) - (x_i, y_i)\|), \quad (8)$$

式中: $\phi(\cdot)$ 为核函数; λ_i 为待求系数, $\|(x, y) - (x_i, y_i)\|$ 为欧几里得距离范数。

引入正则化约束构建求解方程:

$$(\Phi + \sigma I)z = \mathbf{z}, \quad (9)$$

式中:矩阵 Φ 内元素满足 $\Phi_{ij} = \phi(\|(x_i, y_i) - (x_j, y_j)\|)$; σ 为平滑参数; I 为单位矩阵; $\mathbf{z}$ 为地层界面控制点的观测值向量。

此方法能够处理稀疏的地质勘探数据,自动生成连续地层界面曲面,为后续的三维地层建模和不确定性量化提供技术支撑。

3.2 不确定性量化

为进一步量化建模过程中的不确定性,引入信息熵来衡量模型的不确定性。信息熵通过定量描述地质模型局部地层单元隶属度,对模型的不确定模拟结果进行定量评价^[20-21]。信息熵计算式为

$$H = - \sum_{i=1}^n p_i \log_2 p_i, \quad (10)$$

式中: p_i 为当前单元隶属于地层 i 的概率; n 为所有地层类型的数目。

GA-Stacking模型基于多信度融合数据进行预测,得到带有概率分布的虚拟钻孔数据。通过计算虚拟钻孔的信息熵,插值得到三维不确定性分布,用于衡量模型的不确定性。

4 工程实例

4.1 数据来源与预处理

以苏州某轨道交通为背景,基于稀疏数据构建地层模型。实验收集钻孔日志、CPT数据及室内试验数据。研究区域共计12个CPT孔、14个钻孔,工程基准地层剖面见图5,该剖面基于现场勘探数据与室内试验数据构建,并在施工过程中根据观察到的实际地层进行校准,可视为基准剖面。自地表向下依次为①杂填土、②黏土、③粉质黏土、④粉土夹粉砂、⑤粉砂、⑥黏土、⑦粉质黏土、⑧粉砂。

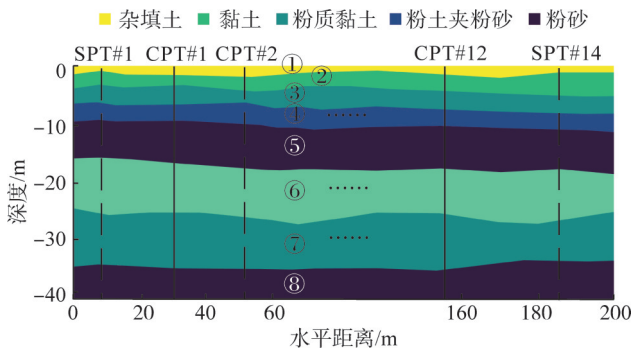


图5 工程基准地层剖面

Fig.5 Engineering benchmark stratigraphic profile

4.2 GA-Stacking模型训练及结果

算法中基学习器的训练采用Python 3.8, 结合Scikit-Learn 机器学习库实现。以F1分数作为优化目标, 通过随机搜索方法确定各基学习器的最优参数配置, 见表1。

表1 各基学习器超参数设置

Tab.1 Hyperparameter settings of base classifiers

基学习器	最优参数
XGB	正则系数取0.2, 分裂最小损失阈值取0.1, 学习率取0.1
LGB	样本采样比例取0.8, 叶子节点数取31, 基树数量取100, 学习率取0.2
KNN	权重模式采用距离加权, 近邻数量取3, 距离度量方式采用欧氏距离
LR	优化求解器采用lbfgs, 正则化类型采用 l_2 , 最大迭代次数取: 1 000, 正则强度系数C取1
RF	基树数量取100, 树最大深度取15, 内部节点最小分裂样本数取2, 叶子节点最小样本数取1, 特征采样规则采用 $\log_2$
DT	内部节点最小分裂样本数取2, 叶子节点最小样本数取1, 分裂准则采用熵
SVM	核函数类型采用径向基核, 核函数系数采用auto, 正则强度系数C取100

在得到基学习器最优参数后, 随机生成多个基学习器组合作为初始种群, 每个组合包括3~6个不同的基学习器, 个体即为被选中的基学习器组合, 通过随机选择不同数量与类型的基学习器保证初始种群的多样性。对种群中的每个个体构建Stacking模型, 采用交叉验证计算加权F1分数作为适应度值, 反映该组合泛化性能。遗传算法性能主要受种群规模、迭代次数、交叉概率、变异概率与精英保留比例影响。以模型准确率为优化目标, 对遗传算法开展参数敏感性分析, 通过

控制变量验证不同参数组合对模型性能与精度的影响, 确定最优参数组合: 种群规模为14, 迭代代数为20, 交叉概率为0.2, 变异概率为0.15, 精英保留率为10%。

遗传算法迭代过程见图6, 适应度呈现稳定上升趋势并最终收敛。GA-Stacking模型框架见图7, 经过优化后, 算法最终从KNN、XGB、LGB、SVM、LR、DT与RF这7个基学习器中, 选择LGB、XGB与DT作为最优基学习器组合, LGB作为元学习器。采用F1分数与准确率作为评估依据, 见表2。对比发现GA-Stacking模型在测试集上领先于其他学习器, 其F1分数达到0.90, 比DT高出3个百分点, 比LR高出41个百分点。与其他基学习器相比, GA-Stacking模型准确率最高。

基学习器与GA-Stacking模型混淆矩阵见图8。LGB在地层6(⑥黏土)产生较大误差, 把部分黏土误判为粉质黏土, XGB在地层3(③粉质黏土)产生较大误差, 把部分粉质黏土分类为粉土夹粉砂, 而GA-Stacking模型在每个土类标签处都具有较高的准确率。

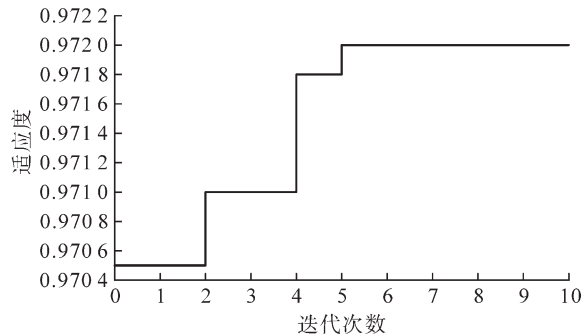


图6 遗传算法迭代过程

Fig.6 Genetic algorithm iteration process

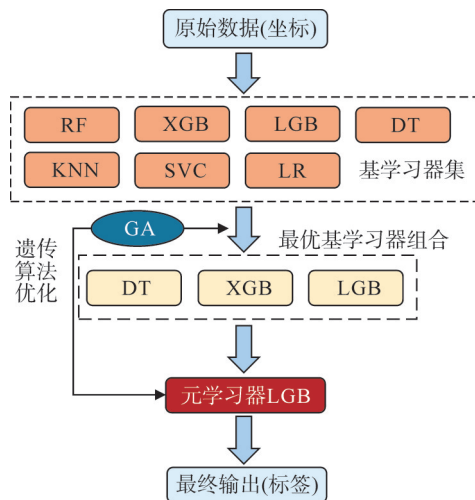


图7 GA-Stacking模型框架

Fig.7 Architecture of the GA-Stacking model

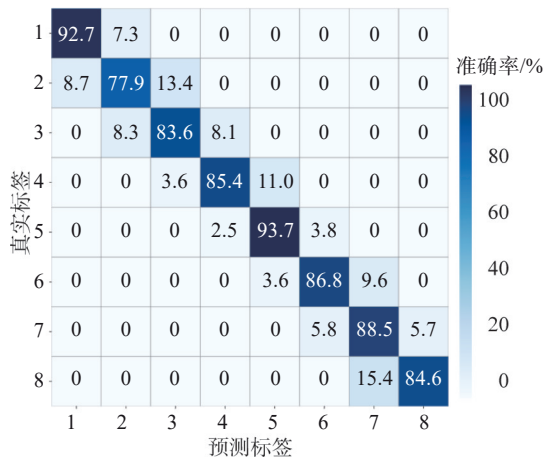
表2 各学习器F1分数与准确率对比
Tab.2 F1-score vs accuracy across learners

基学习器	F1分数		准确率	
	训练集	测试集	训练集	测试集
XGB	0.98	0.87	0.98	0.86
LGB	0.99	0.87	0.99	0.87
KNN	0.99	0.83	0.99	0.83
LR	0.51	0.49	0.51	0.49
RF	0.97	0.80	0.97	0.80
DT	0.99	0.87	0.99	0.87
SVM	0.93	0.83	0.93	0.83
GA-Stacking	0.99	0.90	0.99	0.89

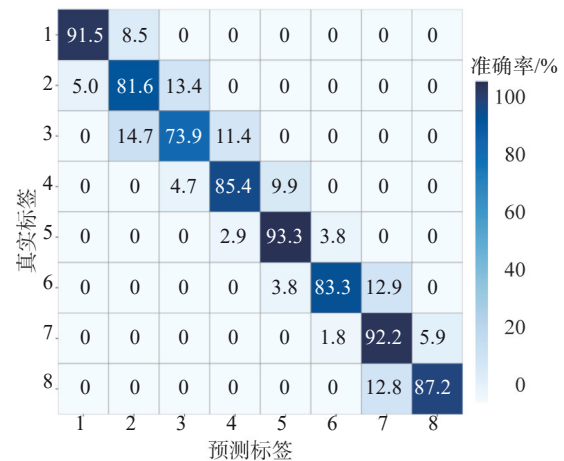
在原有26个实测CPT孔与钻孔数据基础上,综合地质勘探规范中钻孔布置区间与工程实际布孔经验,选取30 m、50 m、100 m间距分别生成虚拟钻孔。对比分析建模精度与计算效率后,选择50 m间距生成虚拟钻孔,共生成25个虚拟钻孔。通过GA-Stacking模型基于已有地勘数据构建包含

虚拟钻孔的地层分类数据集。基于这些数据,考虑研究区范围与各地层厚度分布特征,将研究区域按10 m×10 m×0.1 m的剖分尺度进行网格剖分。

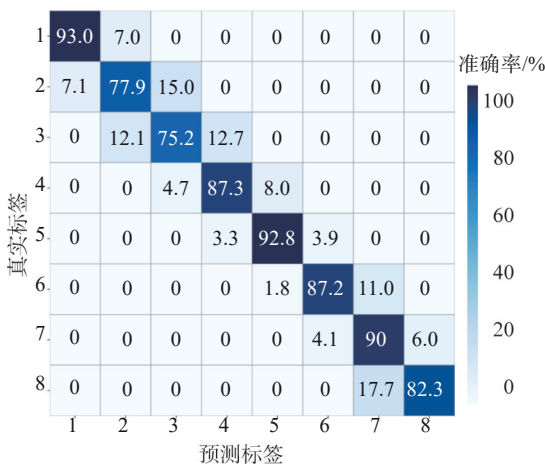
基于径向基函数隐式建模方法,选取线性、三次样条、五次样条、薄板样条、多二次、逆多二次及高斯7种核函数开展对比实验,以RMSE为精度评价指标,结果表明多二次核函数在本研究地质三维空间数据中具备最高插值精度与数值稳定性,可更好地适配地层空间连续分布特性,因此选用该核函数开展三维隐式地层建模,分别生成地层界面图与地层实体图(见图9)。建模总耗时294 s,包括GA-Stacking训练时间(264 s)、预测时间(8 s)与建模时间(22 s),相较于传统建模方法,建模过程中所需的人力、算力与时间显著减少,地层建模效率提升。



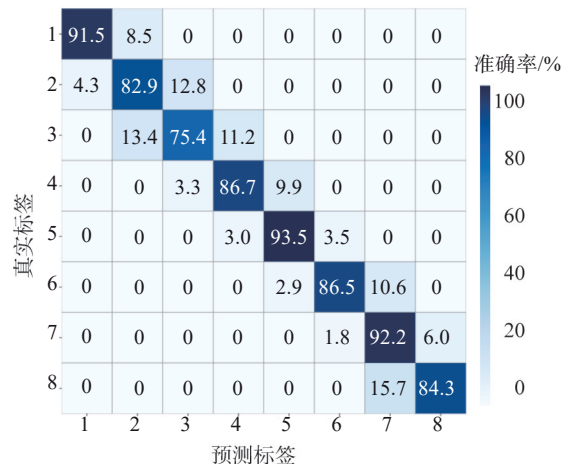
(a) DT



(b) LGB



(c) XGB



(d) GA-Stacking

图8 基学习器与GA-Stacking模型混淆矩阵

Fig.8 Base learners & GA-Stacking confusion matrices

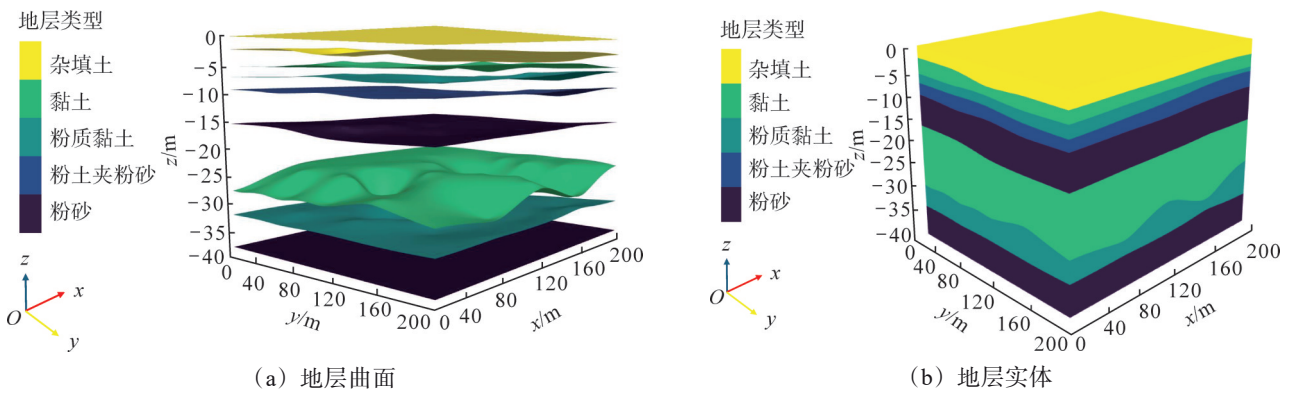


图9 建模结果示意

Fig.9 Schematic of modeling results

为对比验证本文构建的三维地层模型在局部建模精度方面的提升效果,设置3种方案构建地层模型。方法1:根据本文方法,基于多信度地勘数据与虚拟钻孔数据构建模型;方法2:根据本文方法,基于原始地勘数据构建模型;方法3:根据克

里金方法,基于原始地勘数据构建模型。采用主成分分析法选择钻孔最多的方向作为校验主方向,对比各组模型剖面与工程基准地层剖面,对比结果见图10和图11,其中红色标记区域为3种建模方法差异较大的区域。

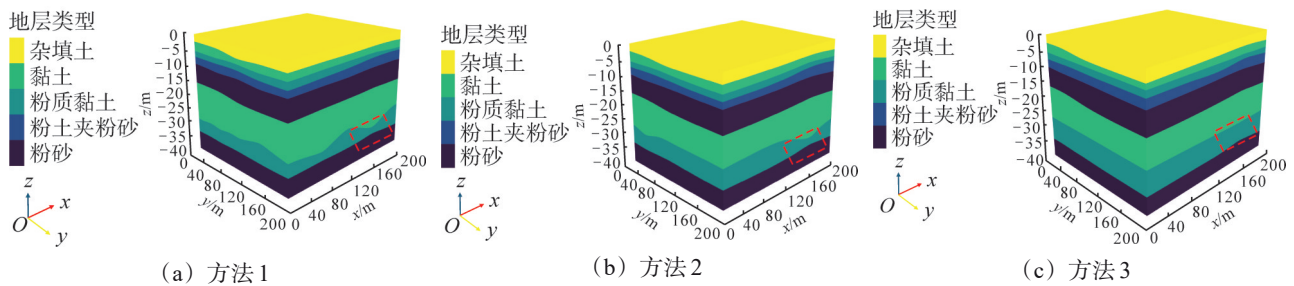


图10 建模结果对比

Fig.10 Comparison of modeling results using different methods

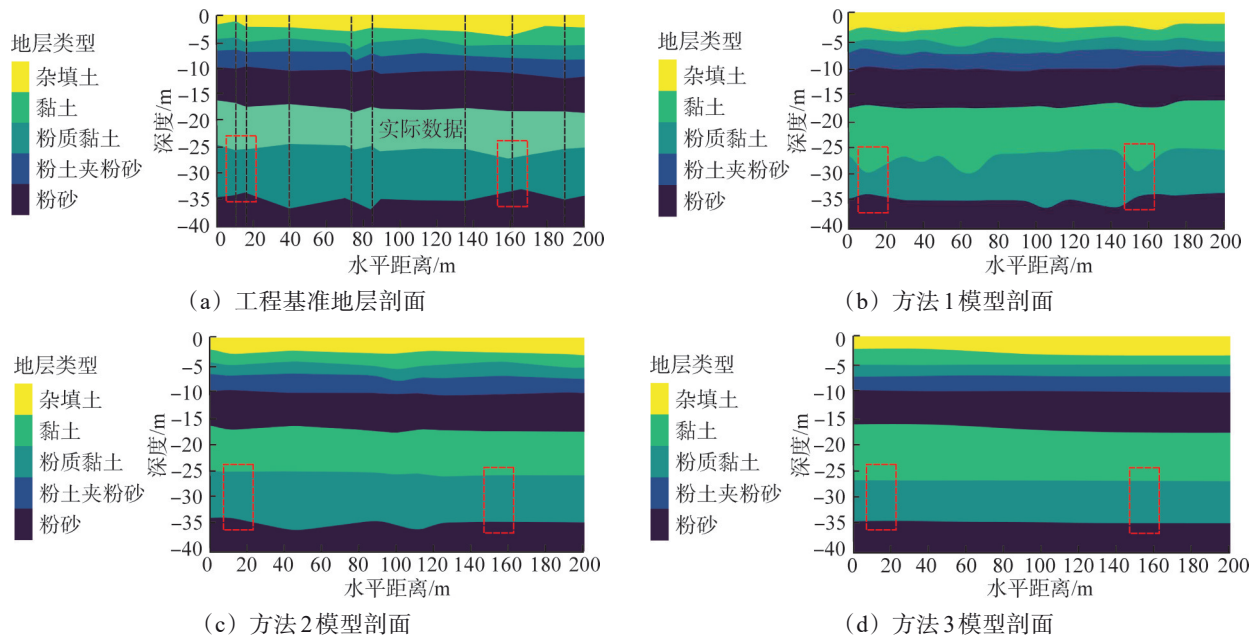


图11 基于剖面的建模对比

Fig.11 comparison of stratigraphic modeling results based on cross-section

由图 10 和图 11 可以看出, 仅基于实际钻孔勘探数据的方法 2 能一定程度反映该区域的地层变化趋势, 但在数据稀疏区域难以捕捉可能存在的地层变化特征; 克里金插值建模的方法 3 由于数据稀疏, 难以反映该区域的地层变化趋势; 相比之下, 方法 1 将 GA-Stacking 集成学习策略生成的虚拟钻孔数据作为补充约束, 所构建的地层模型能够更好地刻画地层的局部变化, 通过引入更多的数据点, 较好地保持地层的空间连续性与合理性。

为定量评估各模型的准确性, 采用抽稀检验法剔除部分原始数据构建模型, 通过对比剔除的原始数据与地层模型中生成的数据, 量化评价模型的准确性, 结果见表 3。

表 3 不同建模方法下的平均误差

Tab.3 Average error comparison of modeling methods

地层类型	平均误差/m		
	方法 1	方法 2	方法 3
①杂填土	0.15	0.18	0.24
②黏土	0.14	0.43	0.38
③粉质黏土	0.12	0.63	0.69
④粉土夹粉砂	0.24	0.46	0.39
⑤粉砂	0.05	0.11	0.21
⑥黏土	0.60	0.64	2.09
⑦粉质黏土	0.32	0.91	0.80
平均值	0.23	0.47	0.63

相较于方法 2, 方法 1 的建模平均误差减小 0.24 m; 相较于方法 3 平均误差减小 0.40 m, 具有更高的精度。将模型剖面 and 实际剖面的重叠面积与实际剖面总面积之比定义为剖面吻合度, 用以表征模型对实际地层的模拟能力, 各地层的剖面吻合度见表 4。

表 4 不同建模方法下各地层的剖面吻合度

Tab.4 Profile consistency of each stratum under different modeling methods

地层类型	吻合度/%		
	方法 1	方法 2	方法 3
①杂填土	86	83	75
②黏土	87	81	72
③粉质黏土	87	77	66
④粉土夹粉砂	85	80	77
⑤粉砂	88	83	78
⑥黏土	77	74	68
⑦粉质黏土	83	80	73
平均值	85	80	73

方法 1 各地层剖面吻合度平均值为 85%, 方法 2、方法 3 的剖面吻合度平均值分别为 80%、73%。

方法 1 各地层吻合度均为最优, 其中, ⑤粉砂吻合度高达 88%, ③粉质黏土吻合度为 87%, 高于其他两种方法 (77% 与 66%)。同时, 在数据稀疏的条件下没有发生如方法 3 在⑦黏土的不吻合现象。上述定量评估结果表明, 方法 1 在提高三维地质模型的精度与可靠性方面具有一定优势, 尤其是在数据稀疏区域能够更准确地刻画地层形态与变化趋势。

4.3 不确定性分析及钻孔方案优化

建模完成后, 对全部虚拟钻孔逐一求解对应的信息熵, 插值得到信息熵三维分布, 见图 12。模型的不确定性主要集中在纵向的两端与中心部分区域, 由于勘探数据的缺失, 建模生成的地层模型在此处有很高的不确定性。将各区域平均熵值最大处作为新增钻孔位置, 在熵值分布中进行更新, 重新计算各区域平均熵值, 确定后续新增钻孔位置, 迭代结果见表 5。

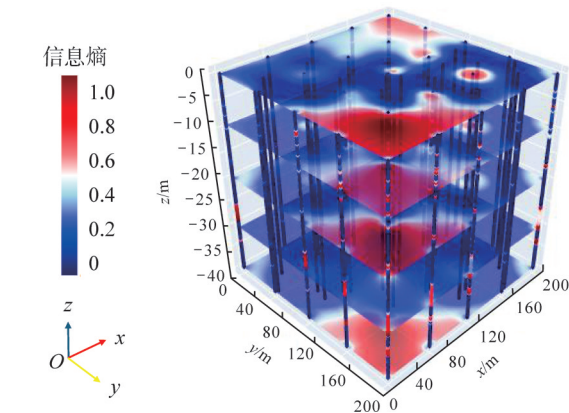


图 12 不确定性分布示意

Fig.12 Schematic of uncertainty distribution

表 5 加入新增钻孔后不确定性变化结果

Tab.5 Change in uncertainty after adding new boreholes

新增钻孔编号	初始平均熵值	新增后平均熵值	最小熵值	最大熵值	不确定性降低/%
1	0.110	0.102	0.08	0.87	7.7
2	0.102	0.094	0.04	0.65	8.2
3	0.094	0.087	0.05	0.55	7.7
4	0.087	0.079	0.07	0.63	8.2
5	0.079	0.073	0.03	0.53	8.6

新增 5 个钻孔后, 原空间平均熵值由 0.110 下降至 0.073, 不确定性降低 34.5%。新增钻孔位置示意图 13, 新增钻孔周围均无原始钻孔数据, 且均为该区域熵值较集中点, 符合工程现场实际情况。

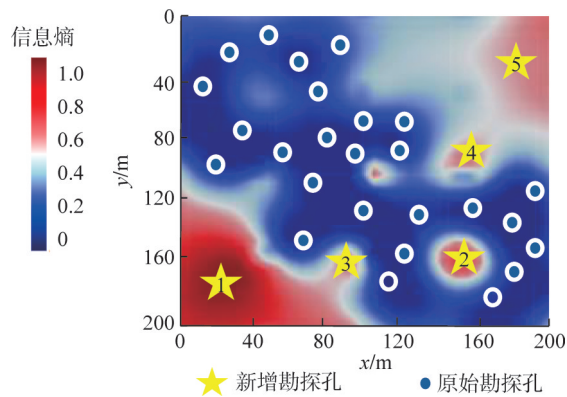


图 13 新增钻孔位置示意

Fig.13 Schematic of new borehole locations

5 结论

(1) 通过多信度地勘数据融合, 将离散、不均匀、不同信度的原始地勘数据处理为连续、均匀的融合数据, 从数据层面提升建模精度, 减小误差。

(2) GA-Stacking 模型在地层分类预测中表现出较好的准确性与稳定性。经过实例验证, 模型在测试集上的 F1 分数与准确率分别达到 90% 与 89%, 相比单一机器学习模型具有更高的准确性。

(3) 基于 GA-Stacking 模型生成的虚拟钻孔数据能够作为地层三维隐式建模的有效约束, 所构建的地层模型能更准确地刻画地层的局部变化特征, 通过实例生成的地层分布与实际勘探地层分布误差平均值为 0.23 m, 剖面吻合度平均达 85%, 验证了该方法的有效性。

(4) 以不确定性大小作为新增钻孔选择的依据, 在实例车站高不确定性区域新增 5 个钻孔, 使总不确定性降低 34.5%, 为工程后期钻孔布设方案提供参考。

参考文献(References):

- [1] 陈开明, 陈元芳, 吴亚东, 等. 地下空间安全开发利用三维评价——以珠海市为例[J/OL]. 地下空间与工程学报, 2025-10-15. <https://link.cnki.net/urlid/50.1169.TU.20251014.1618.004>.
Chen Kaiming, Chen Yuanfang, WU Yadong, et al. Three-dimensional evaluation of safe exploitation and utilization of underground space: a case study of Zhuhai city[J/OL]. Chinese Journal of Underground Space and Engineering, 2025-10-15. <https://link.cnki.net/urlid/50.1169.TU.20251014.1618.004>.
- [2] 钟茂华. 发展地下空间安全科学, 构筑重大工程安全屏障[J]. 清华大学学报(自然科学版), 2025, 65(3): 413.
- [3] De Rienzo F, Oreste P, Pelizza S. Subsurface geological-geotechnical modelling to sustain underground civil planning[J]. Engineering Geology, 2008, 96(3/4): 187-204.
- [4] Caumon G, Gray G, Antoine C, et al. Three-dimensional implicit stratigraphic model building from remote sensing data on tetrahedral meshes: theory and application to a regional model of La popa basin, NE Mexico[J]. IEEE Transactions on Geoscience and Remote Sensing, 2013, 51(3): 1613-1621.
- [5] Belbèze S, Rohmer J, Guyonnet D, et al. Improving spatial interpolation for anomaly analysis in presence of sparse, clustered or imprecise data sets[J]. Journal of Geochemical Exploration, 2025, 279: 107868.
- [6] 杭振泉, 薛涛, 史玉金, 等. 城市地下空间三维地质建模效率和精度提升方法——以上海为例[J]. 工程地质学报, 2024, 32(3): 1-12.
Hang Zhenquan, Xue Tao, Shi Yujin, et al. Methods for improving 3D geological modeling efficiency and enhancing property interpolation accuracy of urban underground space: A case study of Shanghai[J]. Journal of Engineering Geology, 2024, 32(3): 1-12.
- [7] Zhou X Q, Shi P X. Multi-scale generative adversarial network for 2D subsurface reconstruction using multi-fidelity geological exploration data[J]. Advanced Engineering Informatics, 2025, 66: 103482.
- [8] Kumar J K, Konno M, Yasuda N. Subsurface soil-geology interpolation using fuzzy neural network[J]. Journal of Geotechnical and Geoenvironmental Engineering, 2000, 126(7): 632-639.
- [9] Shi C, Wang Y. Development of subsurface geological cross-section from limited site-specific boreholes and prior geological knowledge using iterative convolution XGBoost[J]. Journal of Geotechnical and Geoenvironmental Engineering, 2021, 147(9): 04021082.
- [10] Abedi M, Norouzi G H, Bahrudi A. Support vector machine for multi-classification of mineral prospectivity areas[J]. Computers & Geosciences, 2012, 46: 272-283.
- [11] 郭甲腾, 刘寅贺, 韩英夫, 等. 基于机器学习的钻孔数据隐式三维地质建模方法[J]. 东北大学学报(自然科学版), 2019, 40(9): 1337-1342.
Guo Jiateng, Liu Yinhe, Han Yingfu, et al. Implicit 3D geological modeling method for borehole data based on machine learning[J]. Journal of Northeastern University (Natural Science), 2019, 40(9): 1337-1342.
- [12] 王昊, 严加永, 祁光, 等. 基于重磁三维建模与机器学习的成矿预测方法——以朱溪外围为例[J]. 地球物理学进展, 2023, 38(2): 734-747.

- Wang Hao, Yan Jiayong, Qi Guang, et al. Metallogenic prediction method based on gravity and magnetic three-dimensional modeling and machine learning: a case study of Zhuxi[J]. Progress in Geophysics, 2023, 38(2): 734-747.
- [13] 朱峻生, 王胜, 柏君, 等. 基于改进KNN算法的有限钻孔预测全域地质特征的方法[J]. 隧道建设(中英文), 2023, 43(S2): 348-358.
- Zhu Junsheng, Wang Sheng, Bai Jun, et al. An improved K-nearest neighbor algorithm method using finite boreholes for predicting full-area geological features[J]. Tunnel Construction, 2023, 43 (S2): 348-358.
- [14] 王牧帆, 罗周全, 于琦. 基于Stacking模型的采空区稳定性预测[J]. 黄金科学技术, 2020, 28(6): 894-901.
- Wang Mufan, Luo Zhouquan, Yu Qi. Stability prediction of goaf based on stacking model[J]. Gold Science and Technology, 2020, 28(6): 894-901.
- [15] Jia R, Lv Y K, Wang G W, et al. A stacking methodology of machine learning for 3D geological modeling with geological-geophysical datasets, Laochang Sn camp, Gejiu (China) [J]. Computers & Geosciences, 2021, 151: 104754.
- [16] 扶金铭, 胡茂胜, 方芳, 等. Stacking集成策略下的径向基函数曲面复杂矿体三维建模方法[J]. 地球科学, 2024, 49(3): 1165-1176.
- Fu Jinming, Hu Maosheng, Fang Fang, et al. Complex orebody 3D modeling using radial basis function surface incorporating stacking integration strategy[J]. Earth Science, 2024, 49(3): 1165-1176.
- [17] 邓怡徽, 邹艳红, 李延申. 混合ABC-Stacking机器学习的钻孔数据地层三维隐式建模方法[J]. 成都理工大学学报(自然科学版), 2025, 52(5): 1020-1034.
- Deng Yihui, Zou Yanhong, Li Yanshen. A three-dimensional implicit formation modeling method for drilling data with a mixed artificial bee colony stacking machine learning strategy[J]. Journal of Chengdu University of Technology (Science & Technology Edition), 2025, 52(5): 1020-1034.
- [18] 徐继伟, 杨云. 集成学习方法: 研究综述[J]. 云南大学学报(自然科学版), 2018, 40(6): 1082-1092.
- Xu Jiwei, Yang Yun. A survey of ensemble learning approaches[J]. Journal of Yunnan University (Natural Sciences Edition), 2018, 40 (6): 1082-1092.
- [19] Katoch S, Chauhan S S, Kumar V. A review on genetic algorithm: past, present, and future[J]. Multimedia Tools and Applications, 2021, 80(5): 8091-8126.
- [20] 王丽芳, 刘肖莉, 徐坤, 等. 数据和知识融合的Bayesian-MCMC三维地质建模[J]. 地球科学, 2024, 49(8): 3056-3070.
- Wang Lifang, Liu Xiaoli, Xu Kun, et al. Bayesian-MCMC (Markov chain Monte Carlo)based three-dimensional geological model optimization by data and knowledge fusion[J]. Earth Science, 2024, 49(8): 3056-3070.
- [21] 代月晨, 王英, 束学道, 等. 基于GA-BP的弧形端坯料辊切成形工艺优化[J]. 塑性工程学报, 2026, 33(2): 157-165.
- Dai Yuechen, Wang Ying, Shu Xuedao, et al. Process optimization of roll cutting forming for arc-end blanks based on GA-BP[J]. Journal of Plasticity Engineering, 2026, 33(2): 157-165.