

<https://doi.org/10.3799/dqkx.2024.051>



基于机器学习的黄土关键力学 参数概率预测统一框架体系

宋超¹, 赵腾远¹, 高重阳^{1,2*}

1. 西安交通大学人居环境与建筑工程学院, 陕西西安 710049

2. 东北电力大学建筑工程学院, 吉林吉林 132012

摘要: 为实现黄土关键力学参数的准确预测, 并合理刻画预测结果的不确定性, 提出了基于机器学习方法的黄土关键力学参数概率预测统一框架体系, 通过对训练集的预测偏差进行概率分布拟合, 进而构建预测结果的 95% 置信区间, 置信区间的大小反映了预测结果的合理与否。基于随机森林、决策树、极限梯度提升和自适应提升 4 种方法预测黄土黏聚力, 对应的决定系数 R^2 分别达到了 0.84、0.75、0.81 和 0.79, 4 种方法所构建的 95% 置信区间包含真正的试验结果的比例均在 95% 左右。表明通过训练集的预测偏差得到的 95% 置信区间是相对可靠的, 可对预测结果的不确定性进行合理量化。此外, 基于上述 4 种方法可实现黄土黏聚力的相对准确的预测。

关键词: 黄土; 黏聚力; 数据驱动; 概率预测; 偏差; 不确定性。

中图分类号: P642.3

文章编号: 1000-2383(2026)02-386-12

收稿日期: 2024-01-29

Unified Framework for Probabilistic Prediction of Critical Mechanical Parameters of Loess by Machine Learning Methods

Song Chao¹, Zhao Tengyuan¹, Gao Chongyang^{1,2*}

1. School of Human Settlements and Civil Engineering, Xi'an Jiaotong University, Xi'an 710049, China

2. School of Civil Engineering and Architecture, Northeast Electric Power University, Jilin 132012, China

Abstract: In order to predict the critical mechanical parameters of loess accurately and quantify the uncertainty corresponding to the prediction results reasonably, a unified framework for probabilistic prediction of critical mechanical parameters of loess by machine learning methods is proposed. By fitting probability density function to the bias of the training dataset, a 95% confidence interval for the prediction results is constructed, and the size of the confidence interval reflects the rationality of the prediction results. (Result) Predicting cohesion of loess based on four machine learning methods, namely, random forest, decision tree, extreme gradient boosting and adaptive boosting, the corresponding coefficients of determination R^2 reached 0.84, 0.75, 0.81 and 0.79, respectively. The proportion of measurement data included in the 95% confidence interval constructed by the four methods is around 95%. It is shown that the 95% confidence interval obtained from the bias based on the training dataset is relatively reliable and can quantify the uncertainty of the prediction results reasonably. In addition, the cohesion of loess can be predicted accurately

基金项目: 吉林省教育厅科学技术研究基金项目(No.JJKH20230141KJ); 国家自然科学基金青年基金项目(No.42107204)。

作者简介: 宋超(1993-), 女, 博士研究生, 研究方向为岩土工程数据分析与机器学习建模。ORCID: 0000-0003-0519-0870. E-mail: song_chao@stu.xjtu.edu.cn

* **通讯作者:** 高重阳, ORCID: 0000-0002-3776-7796. E-mail: gaochongyang99@stu.xjtu.edu.cn

引用格式: 宋超, 赵腾远, 高重阳, 2026. 基于机器学习的黄土关键力学参数概率预测统一框架体系. 地球科学, 51(2): 386-397.

Citation: Song Chao, Zhao Tengyuan, Gao Chongyang, 2026. Unified Framework for Probabilistic Prediction of Critical Mechanical Parameters of Loess by Machine Learning Methods. *Earth Science*, 51(2): 386-397.

using the four machine learning methods.

Key words: loess; cohesion; data-driven; probabilistic prediction; bias; uncertainty.

0 引言

我国黄土分布广泛,尤其在黄土高原地区,其力学性质与地质灾害防治、基础设施建设、交通运输规划等息息相关(Yan *et al.*, 2023; 张俊然等, 2023).黄土的黏聚力是黄土重要的力学参数之一,反映了土体颗粒间的相互关系以及吸附作用,黄土黏聚力的准确测定具有十分重要的理论价值与工程意义(党进谦和李靖, 1997).例如,黄土的黏聚力随其埋深、地质成因、赋存环境等的不同存在显著差异(Xu *et al.*, 2023; Zuo *et al.*, 2024),与黄土的变形、沉降等密切相关,因此在黄土地区重大线性工程(如高速铁路、高速公路)施工过程中,需充分考虑黄土黏聚力的影响;黄土浸水后,其黏聚力降低导致土体强度衰减,在地震等外部动荷载作用下易发生边坡失稳等灾害,对周边地区人民的生命和财产安全产生了极大威胁.因此,对研究区或施工现场的黄土黏聚力进行合理估计十分必要.

传统土力学试验中,黏聚力的测试主要通过直接剪切试验或三轴压缩试验,然而,由于土体的取样、包装、运输等均需耗费较多的人力、物力、时间,且受限于试验条件、施工进度、施工成本等因素的制约,经常难以满足要求.因此,可利用经验公式或机器学习方法,建立易获得的黄土物理力学参数与其黏聚力之间的关系,间接估计黄土的黏聚力.经验公式法可明确显示易获得的黄土物理力学参数与其黏聚力之间的定量转换关系,但其固有的方程形式导致其经常难以准确估计目标值(Zhao *et al.*, 2022).机器学习方法通过建立易获得的黄土物理力学参数与其黏聚力之间的复杂非线性关系,可显著提高模型的预测性能(宋超等, 2023).

然而,传统机器学习方法往往仅能对目标值进行预测,预测结果的合理性通常难以定量刻画(Zhang *et al.*, 2022b).因此,本文提出了基于机器学习方法的黄土关键力学参数概率预测统一框架体系,通过对训练集的预测偏差进行概率分布拟合,进而构建预测结果的95%置信区间,从而实现预测结果不确定性的合理刻画.相比于基于贝叶斯理论的不确定性量化方法,本文所提方法适用范围更广,且无需进行马尔科夫链蒙特卡洛模拟,计算

效率更高.

本文研究内容总体框架如下:首先,介绍了4种常见的机器学习方法的基本原理,本文以随机森林、决策树、极限梯度提升和自适应提升4种常见的机器学习模型为例,验证所提方法不确定性量化的合理性与可靠性,对于其他机器学习模型,本文所提方法同样适用;其次,提出了基于机器学习方法的黄土关键力学参数概率预测统一框架体系,并对其不确定性量化的原理与步骤进行阐述;然后,通过一个粉黄土黏聚力概率预测工程实例验证了本文所提的不确定性量化方法的合理性与可靠性,并对4种机器学习方法的预测结果进行了对比分析;最后,探讨了基于训练集的预测误差进行不确定性量化的合理性,并与本文所提的黄土关键力学参数概率预测统一框架体系进行了对比.

1 黄土黏聚力预测的机器学习方法

1.1 决策树方法

决策树(DT)方法通过一系列决策节点和分支实现目标值的预测.决策树的组成主要包括根节点、内部节点、叶节点和树枝.其中非叶节点用于表征在某个输入特征属性上的划分,根据测量数据在该输入特征属性上的取值将其划分为若干子集,分支表示预测目标值,叶节点用于存放预测目标值(董学超等, 2023).决策树方法不仅可以处理离散和连续型数据,且在数据量较多时,具有较高的计算效率.其主要实施步骤如下(井彦林等, 2011; Kardani *et al.*, 2022):首先,通过信息增益、信息增益比、基尼指数等方法从所有输入特征中确定用于数据集划分的输入特征;其次,在每个节点上分别确定用于数据集划分的输入特征,对数据集进行逐级划分,直至达到预定条件或数据集已不能被继续划分;最后,为避免出现过拟合,需对决策树进行剪枝,主要包括在决策树构建过程中通过限制其深度、节点数等实现剪枝,和决策树构建完成后通过对子树进行剪枝,从而降低模型的复杂度.

1.2 随机森林方法

随机森林(RF)方法是由Breiman(2001)提出的一种集成学习算法.通过构建多个决策树,取所

有决策树输出目标的平均值作为代理模型的预测目标值(Ewusi-Wilson *et al.*, 2023). 随机森林方法不仅可以处理高维数据, 还可根据数据集特征对其中的缺失值进行补充, 并对输入特征的重要性进行评估(Chen *et al.*, 2022). 此外, 在数据量较多时, 随机森林方法仍具有较高的计算效率. 随机森林方法的具体实施步骤如下(Liu *et al.*, 2021; 吴禄源等, 2023):

首先, 根据预置的比例, 从训练集中有放回地进行 N 次随机抽样, 构成 N 个用于代理模型训练的测量数据子集; 其次, 基于每一个测量数据子集进行模型训练, 建立包含 N 棵决策树的随机森林模型; 最后, 综合 N 棵决策树的预测结果, 计算其算术平均值, 即可得到预测目标值.

1.3 极限梯度提升方法

极限梯度提升(XGBoost)方法是一种基于决策树的集成学习算法, 由加法模型(即一系列弱学习器累加形成强学习器)和前向分布算法(即新一轮迭代生成的新学习器基于上一轮计算结果)两部分组成. 极限梯度提升方法通过在损失函数中增加正则化项, 有效降低了模型过拟合发生的概率, 此外, 其通过对损失函数进行泰勒二阶展开, 显著提升了模型的预测性能(Zhang *et al.*, 2021). 基于极限梯度提升方法预测目标值可表示为(Nguyen *et al.*, 2023):

$$y_{p,i} = \sum_{k=1}^K f_k(x_i), f_k \in F, \quad (1)$$

式中: $y_{p,i}$ 表示与第 i 组易获得的黄土物理力学参数 X_i 对应的黄土黏聚力预测值; $f_k(\cdot)$ 表示第 k 棵决策树; K 表示决策树的总数; F 表示所有决策树的集合; x_i 表示第 i 组易获得的黄土物理力学参数 X_i 的测量值; $f_k(x_i)$ 表示第 k 棵决策树中第 i 个黄土黏聚力预测值的得分. 极限梯度提升方法通过在每一次迭代中拟合一个新的弱学习器, 达到了降低上一轮迭代的残差的目的, 进而实现了模型预测性能的不断提升(Abdi *et al.*, 2023; 李山有等, 2024).

1.4 自适应提升方法

自适应提升(AdaBoost)方法是一种增强学习算法, 通过训练多个弱学习器, 并利用权重调整弱学习器的组合方式, 形成强学习器, 从而提升模型的预测性能(Liu *et al.*, 2020). AdaBoost方法的自适应体现在会根据前一个弱学习器的预测性能自动调整其权重系数, 并用来训练下一个弱学习器. 同时, 在每一轮迭代中, 加入一个新的弱学习器, 直

到达到某个设定的足够小的残差或达到预置的最大迭代次数, 进而确定最终的强学习器. 自适应提升方法仅需调整迭代次数即可自动调整弱学习器的权重, 不需对弱学习器进行单独调参, 此外自适应提升方法在处理大规模数据集时具有较高的计算效率(Wen *et al.*, 2023; 杨玲和魏静, 2023).

2 黄土黏聚力概率预测统一框架体系

2.1 框架体系一

在基于黄土物理力学参数与机器学习方法预测其黏聚力过程中, 预测结果的不确定性主要来源于以下几个方面: 其一为黄土自身的不确定性, 主要由黄土自身的性质导致, 如土体的埋深、形成原因、赋存环境等; 其二为外界因素导致其黏聚力存在一定的不确定性, 如取样过程中土体扰动而产生的不确定性、试验人员的操作水平、试验仪器的精度导致的不确定性等(Zhang *et al.*, 2022a). 已有机器学习方法主要聚焦于目标值的准确预测, 而通常忽略了对预测结果合理性/不确定性的合理刻画. 尽管基于贝叶斯学派相关理论可量化预测结果的不确定性, 但通常难以得到解析解, 需通过马尔科夫链蒙特卡洛模拟得到代理模型参数的后验概率分布, 计算效率相对较低, 且对多数机器学习方法并不适用(Bao and Burghardt, 2022). 基于此, 本节提出了一种基于机器学习方法的黄土关键力学参数概率预测统一框架体系. 接下来, 将详细阐述本节所提不确定性量化方法的具体实施步骤:

(1) 将包含黄土黏聚力 y 与易获得的黄土物理力学参数 $X = \{X_1, X_2, \dots, X_9\}$, 其中 X_1 = 埋深, X_2 = 含水率, X_3 = 孔隙比, X_4 = 液限, X_5 = 塑限, X_6 = 液性指数, X_7 = 塑性指数, X_8 = 饱和度和 X_9 = 干密度的数据集划分为两部分, 分别为训练集和测试集;

(2) 基于训练集中的测量数据, 以易获得的黄土物理力学参数作为输入特征, 黄土黏聚力作为输出目标值, 建立机器学习模型, 将得到的黄土黏聚力预测值记作 y_p^{train} , 已有的训练集中黄土黏聚力测量值记作 y_m^{train} ;

(3) 定义机器学习模型的预测偏差 λ_d 为训练集中黄土黏聚力测量值 y_m^{train} 与预测值 y_p^{train} 的比值, 即 $\lambda_d = y_m^{\text{train}}/y_p^{\text{train}}$;

(4) 根据黄土黏聚力自身属性特征, 其值为大于 0 的数, 因此上述步骤中得到的机器学习模型的预测偏差 λ_d 大于 0. 基于此, 通过对数正态分布对预

测偏差 λ_d 进行概率分布拟合,可得到预测偏差 λ_d 对应的概率密度函数(PDF),进一步地,可获得PDF函数的95%置信区间(CI)以及置信区间的上、下边界.95%置信区间体现了预测结果的不确定性;

(5)定义覆盖比例 CP_{95} ,即黄土黏聚力测量值落入95%置信区间的数量与黄土黏聚力测量值的总数的比值,覆盖比例 CP_{95} 体现了不确定性量化的可靠性与合理性(Song *et al.*, 2024).因此,根据步骤4中的结果可得到训练集的覆盖比例 CP_{95}^{train} ;

(6)将测试集中易获得的黄土物理力学参数带入步骤2中建立的机器学习模型,可得到测试集的黄土黏聚力预测值,记作 y_{test}^{train} .将 y_p^{train} 与步骤4中得到的95%置信区间的上、下边界分别相乘,得到测试集中黄土黏聚力预测值对应的95%置信区间;

(7)根据步骤5中覆盖比例的定义,计算测试集的覆盖比例 CP_{95}^{train} .

2.2 框架体系二

本文2.1节详细阐述了一种黄土黏聚力概率预测的统一框架体系,即通过对预测偏差进行概率分布拟合,从而达到量化预测结果不确定性的目的.然而,在2.1节步骤4中拟合概率分布时,直接通过训练集的预测偏差对应的概率密度函数得到95%置信区间,由于测试集中的测量数据为全新的数据,代入2.1节步骤2中建立的机器学习模型,容易导致测试集的预测偏差相比于训练集偏大.此时,通过2.1节得到的95%置信区间用于测试集不确定性量化表征时,覆盖比例 CP_{95} 可能偏小,进而导致不确定性量化结果的合理性与可靠性降低.鉴于此,本节对黄土黏聚力概率预测的统一框架体系一进行了进一步的改进,即框架体系二,具体如下:

首先,将包含黄土黏聚力 y 与易获得的黄土物理力学参数 $X = \{X_1, X_2, \dots, X_9\}$ 的数据集划分为两部分,分别为训练集和测试集;其次,进一步划分训练集,采用五折交叉验证方法,分别用于模型训练和模型验证;最后,基于一次随机试验结果,通过对训练集中模型验证部分的预测偏差进行对数正态分布拟合得到预测偏差对应的概率密度函数.该不确定性量化方法的其他步骤与2.1节完全一致,不再详细阐述.

本节所提不确定性量化方法通过对训练集中模型验证部分的预测偏差进行概率分布拟合,由于训练集中模型验证部分也未参与机器学习模型的训练,因此其可有效改善测试集中不确定性量化结

表1 黄土物理力学参数的基本统计特征

Table 1 Statistical characteristics of the physical and mechanical parameters of loess

| 黄土参数 | 单位 | 最大值 | 最小值 | 平均值 | 标准差 | 偏度 | 峰度 |
|------|-------------------|--------|-------|-------|-------|-------|------|
| 黏聚力 | kPa | 54.60 | 11.70 | 28.20 | 8.03 | 0.37 | 2.99 |
| 埋深 | m | 44.00 | 1.00 | 17.13 | 9.89 | 0.34 | 2.41 |
| 含水率 | % | 25.30 | 8.20 | 16.63 | 3.72 | 0.00 | 2.30 |
| 孔隙比 | 1 | 1.15 | 0.55 | 0.74 | 0.11 | 0.80 | 3.39 |
| 液限 | % | 33.70 | 21.90 | 24.86 | 2.02 | 1.44 | 5.32 |
| 塑限 | % | 20.30 | 15.20 | 16.82 | 0.79 | 1.21 | 4.86 |
| 液性指数 | 1 | 1.11 | -1.17 | -0.06 | 0.42 | -0.30 | 2.56 |
| 塑性指数 | 1 | 13.40 | 6.20 | 8.04 | 1.29 | 1.44 | 5.25 |
| 饱和度 | % | 100.00 | 24.00 | 62.55 | 17.48 | -0.17 | 1.89 |
| 干密度 | g/cm ³ | 1.75 | 1.26 | 1.56 | 0.09 | -0.51 | 2.80 |

果偏小的情况.本节所提不确定性量化方法的可靠性与合理性将在第3章通过一个粉黄土黏聚力概率预测工程实例进行详细说明.

3 黄土黏聚力概率预测实例分析

3.1 数据集概述

本章通过一个工程实例验证本文所提不确定性量化方法的可靠性与合理性.采用的黄土类型为粉黄土,数据集中包含350组测量数据,由1个输出目标与9个输入特征构成,即 y =黄土黏聚力, X_1 =埋深, X_2 =含水率, X_3 =孔隙比, X_4 =液限, X_5 =塑限, X_6 =液性指数, X_7 =塑性指数, X_8 =饱和度和 X_9 =干密度.测量数据的基本统计特征如表1所示,包括最大值、最小值、平均值、标准差、偏度以及峰度.图1给出了测量数据分布的直方图和数据之间的相关关系图.

3.2 基于不同机器学习方法的黄土黏聚力概率预测

本节将详细探讨基于第2章所提的通过黄土黏聚力概率预测统一框架体系——训练集中测量值与预测值的比值进行不确定性量化结果的合理性,以框架体系二为例,将本文所提不确定性量化方法与不同机器学习模型进行耦合,并对比其预测性能.训练集和测试集分别包含280组和70组测量数据.首先,探讨预测偏差与通过随机森林(RF)、决策树(DT)、极限梯度提升(XGBoost)和自适应提升(AdaBoost)4种机器学习方法得到的黄土黏聚力预测值之间的关系,分别计算了4种机器学习模型对应的预测偏差与黄土黏聚力预测值之间的斯皮尔曼相关系数 r_s ,并进行了 P 值检验,统计结果如表

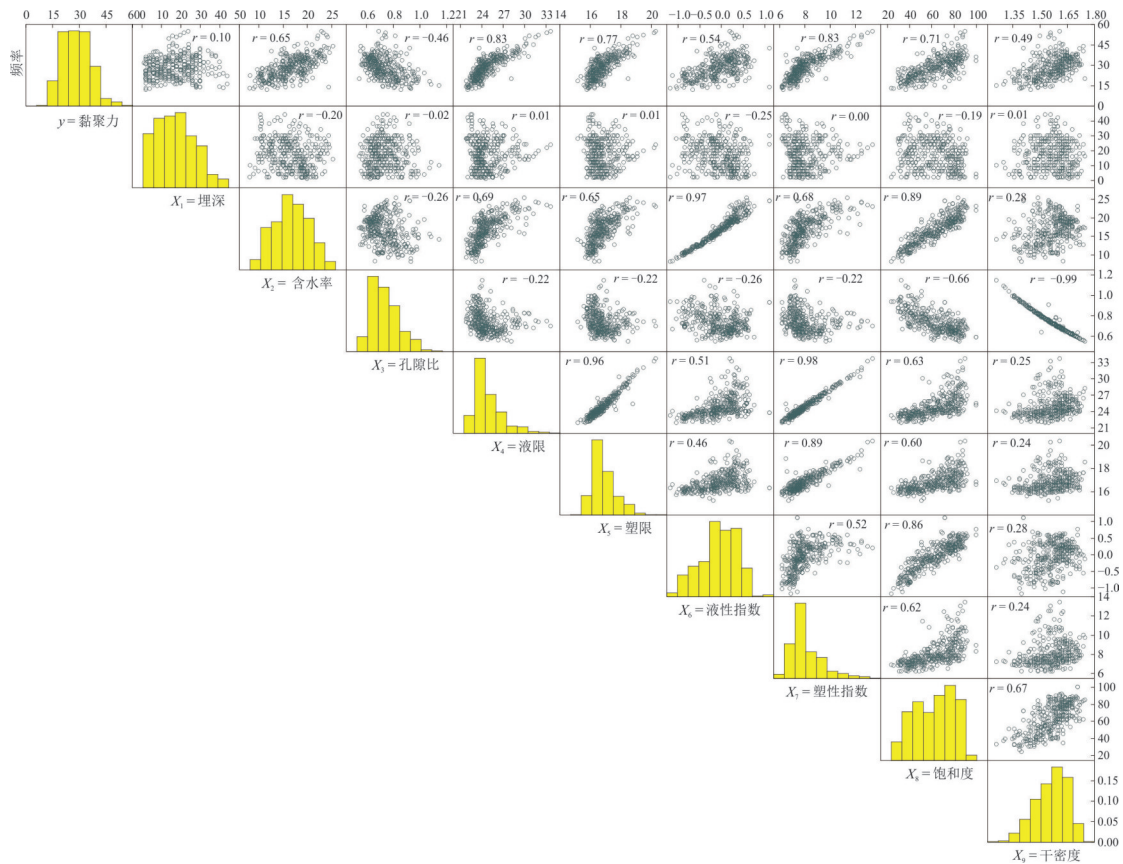


图 1 黄土物理力学参数分布直方图及相关关系图

Fig. 1 The histograms of the physical and mechanical parameters of loess and correlations among them

2所示.在假设检验中,当 P 值 <0.05 时,通常认为观察到的结果在原假设为真时发生的概率极低,从而拒绝原假设(Ching *et al.*, 2019; Liu *et al.*, 2021).由表2可以看出,RF、DT、XGBoost和AdaBoost 4种机器学习方法对应的斯皮尔曼相关系数 r_s 分别为0.089、-0.156、-0.076和0.163, P 值分别为0.138、0.009、0.203和0.006,说明基于RF、DT、XGBoost和AdaBoost 4种机器学习方法得到的黄土黏聚力预测值与预测偏差无关.图2进一步给出了RF、DT、XGBoost和AdaBoost 4种机器学习方法对应的黄土黏聚力预测值与预测偏差之间的散点图,可以看出散点随着黄土黏聚力预测值的增加无明显变化趋势,再次说明了基于RF、DT、XGBoost和AdaBoost方法得到的黄土黏聚力预测值与预测偏差之间不相关,与表2所得结论一致.此外,表2还给出了RF、DT、XGBoost和AdaBoost 4种机器学习方法对应的预测偏差与各输入特征之间的相关关系,可以看出,基于框架体系二得到的预测偏差与各输入特征无关,说明基于不同机器学习方法均可实现黄土黏聚力的无偏估计,且预测结果为

多因素共同作用得到,这与实际工程是一致的.黄土黏聚力与其埋深、含水率、孔隙比、液限、塑限、液性指数、塑性指数、饱和度和干密度均存在一定的相关性.此外,由表2还可以看出,预测偏差与 X_3 =孔隙比和 X_7 =塑性指数之间的斯皮尔曼相关系数已经达到了0.24以上,但其对应的 P 值为0,意味着原假设不成立,即预测偏差与 X_3 =孔隙比和 X_7 =塑性指数之间不相关.

进一步地,为对不同机器学习方法预测黄土黏聚力的准确性进行定量评价,采用决定系数(R^2) (Chen *et al.*, 2024)和平均绝对百分误差(MAPE) (Zhang *et al.*, 2022c)两个评价指标,计算公式分别如下:

$$R^2 = 1 - \frac{\sum_{i=1}^{n_D} (y_{m,i} - y_{p,i})^2}{\sum_{i=1}^{n_D} (y_{m,i} - \mu_y)^2}, \quad (2)$$

$$\text{MAPE} = \frac{1}{n_D} \sum_{i=1}^{n_D} \left| \frac{y_{p,i} - y_{m,i}}{y_{m,i}} \right| \times 100\%, \quad (3)$$

式中: n_D 表示数据集中包含的易获得的黄土物理力

表 2 基于框架体系二的预测偏差与黄土黏聚力预测值及输入特征的斯皮尔曼相关系数及 P 值检验结果

Table 2 Spearman's correlation coefficient between bias obtained from unified framework 2 and predicted cohesion/ input variables and corresponding results of P-value test

| 黄土参数 | RF | | DT | | XGBoost | | AdaBoost | |
|----------------------|--------|-------|--------|-------|---------|-------|----------|-------|
| | r_s | P 值 | r_s | P 值 | r_s | P 值 | r_s | P 值 |
| y=黏聚力 | 0.089 | 0.138 | -0.156 | 0.009 | -0.076 | 0.203 | 0.163 | 0.006 |
| X ₁ =埋深 | 0.140 | 0.019 | 0.213 | 0.000 | 0.052 | 0.389 | 0.122 | 0.041 |
| X ₂ =含水率 | 0.068 | 0.256 | 0.049 | 0.413 | 0.027 | 0.652 | 0.105 | 0.078 |
| X ₃ =孔隙比 | -0.154 | 0.010 | -0.091 | 0.131 | -0.062 | 0.298 | -0.244 | 0.000 |
| X ₄ =液限 | 0.122 | 0.041 | 0.089 | 0.137 | 0.064 | 0.287 | 0.225 | 0.000 |
| X ₅ =塑限 | 0.085 | 0.154 | 0.060 | 0.315 | 0.052 | 0.391 | 0.175 | 0.003 |
| X ₆ =液性指数 | 0.060 | 0.319 | 0.044 | 0.463 | 0.021 | 0.732 | 0.074 | 0.218 |
| X ₇ =塑性指数 | 0.129 | 0.031 | 0.097 | 0.104 | 0.062 | 0.300 | 0.240 | 0.000 |
| X ₈ =饱和度 | 0.131 | 0.028 | 0.083 | 0.168 | 0.048 | 0.423 | 0.204 | 0.001 |
| X ₉ =干密度 | 0.161 | 0.007 | 0.095 | 0.115 | 0.068 | 0.257 | 0.253 | 0.000 |

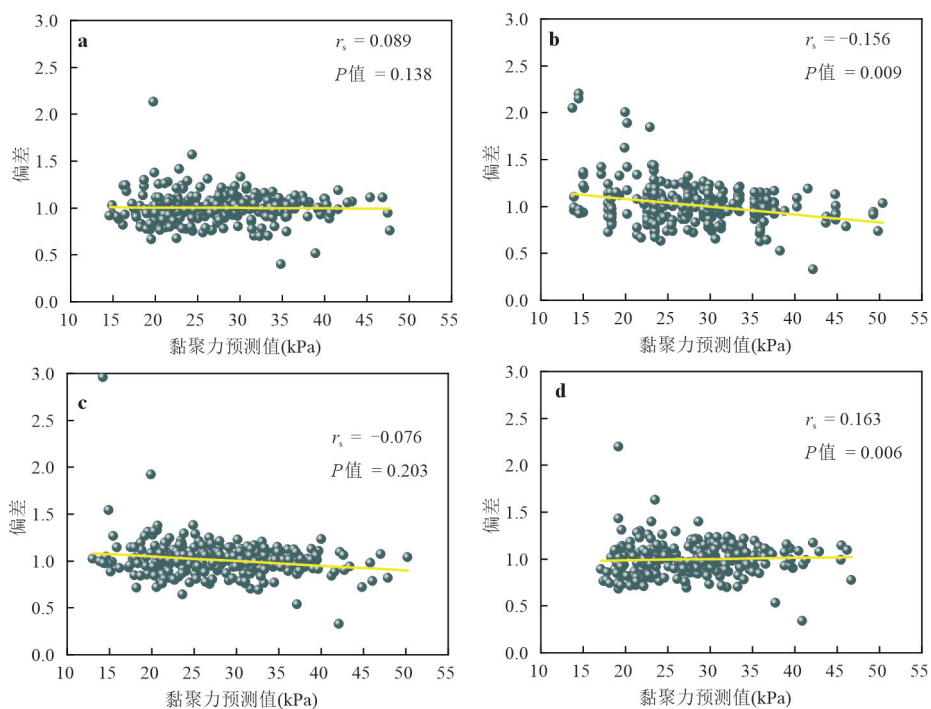


图 2 基于框架体系二的预测偏差与黄土黏聚力预测值分布散点图

Fig.2 Scatter plots between bias obtained from unified framework 2 and predicted cohesion of loess

a. RF; b. DT; c. XGBoost; d. AdaBoost

学参数和黄土黏聚力测量数据样本量; $y_{m,i}$ 表示与第 i 组易获得的黄土物理力学参数 X_i 对应的黄土黏聚力测量值; μ_y 为黄土黏聚力测量值的平均值. 其中, 决定系数 R^2 的范围为 $[0, 1]$, R^2 的值越大, 代表模型的预测性能越好; 平均绝对百分误差 MAPE 的范围为 $[0, +\infty]$, MAPE 的值越小, 代表模型的预测性能越好. 图 3 给出了随机森林 (RF)、决策树 (DT)、极限梯度提升 (XGBoost) 和自适应提升 (Ada-

Boost) 4 种机器学习方法测试集所对应的预测性能以及量化的不确定性, 可以看出, RF 和 XGBoost 方法对应的 R^2 分别达到了 0.84 和 0.81, MAPE 分别为 9.99% 和 10.10%, DT 和 AdaBoost 方法的 R^2 分别为 0.75 和 0.79, MAPE 分别为 12.19% 和 11.86%, RF 和 XGBoost 方法对应的 R^2 略大于 DT 和 AdaBoost 方法, 对应的 MAPE 略小于 DT 和 AdaBoost 方法, 说明基于 RF 和 XGBoost 方法可实现黄土黏

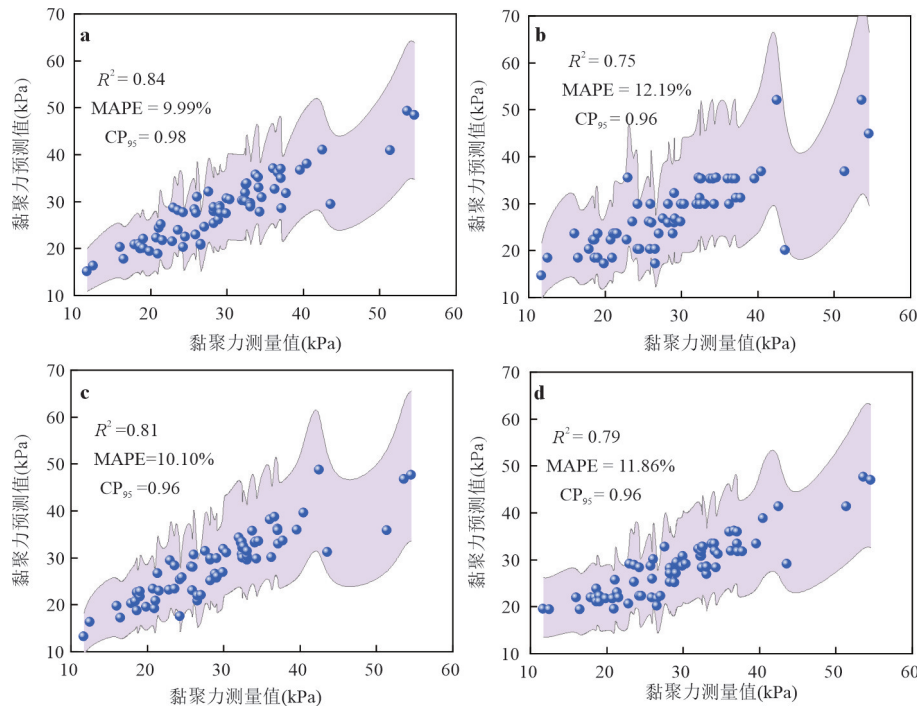


图3 基于框架体系二的不同机器学习方法预测结果及其不确定性

Fig.3 Results of cohesion prediction and uncertainty quantification for different machine learning methods based on unified framework 2

a. RF; b. DT; c. XGBoost; d. AdaBoost

聚力更为准确的预测.图3还给出了RF、DT、XGBoost和AdaBoost 4种机器学习方法对应的覆盖比例 CP_{95} ,分别为0.98、0.96、0.96和0.96,均在0.95附近波动.此外,由图3还可以看出,在数据密集的区域量化的不确定性较小,在数据稀疏的区域,不确定性较大,这是由于在数据稀疏的位置,可获得的信息有限,可能导致预测结果波动较大.此外,DT方法得到的不确定性量化结果明显大于其他3种机器学习方法,尤其在数据稀疏的区域,这是由于DT方法通过一系列决策节点和分支实现目标值的预测,而RF和XGBoost是对多个决策树模型预测结果进行集成,实现目标值的估计;AdaBoost方法通过训练并调整多个弱学习器实现目标值的预测.集成/增强学习算法相比于单一模型,预测性能提升,尤其对于数据稀疏的区域,单一模型的预测结果可能与实际情况偏差较大,而集成/增强学习算法由于耦合了多个简单学习器,可显著提升数据稀疏区域模型预测的准确性.因此可以得出结论,基于本文所提方法可对预测结果的不确定性进行合理量化.

3.3 统计特征分析

上述为一次随机试验的统计结果,按照与3.2节相同的训练集、测试集比例,将数据集随机划分

100次,并重复上述过程,可得到100组对应的RF、DT、XGBoost和AdaBoost 4种机器学习方法的 R^2 、MAPE和 CP_{95} ,统计分布如图4所示.可以看出RF、DT、XGBoost和AdaBoost方法对应的测试集的 R^2 的均值分别达到了0.83、0.71、0.78和0.78,MAPE的100次统计结果的均值分别为10.85%、13.15%、11.31%和12.54%,说明基于RF、XGBoost和AdaBoost方法预测黄土黏聚力结果更为合理,基于DT方法的预测准确性略低,分析认为是由于RF、XGBoost和AdaBoost方法是在一系列弱学习器的基础上进行了整合或权重的重新分配,提升了模型的预测性能.此外,由图4还可以看出,RF、DT、XGBoost和AdaBoost方法对应的 CP_{95} 的训练集100次统计结果的均值分别为0.96、0.95、0.95和0.96,测试集分别为0.96、0.95、0.95和0.95,说明基于本文所提方法可对不同机器学习模型预测结果的不确定性进行合理量化,即使预测结果与实际情况偏差较大(如基于DT方法的预测结果),也可合理刻画其不确定性,本文所提方法具有较高的鲁棒性与可靠性.

3.4 不同框架体系不确定性量化结果对比

上述详细探讨了基于黄土黏聚力概率预测统一框架体系二进行不确定性量化,RF、DT、XG-

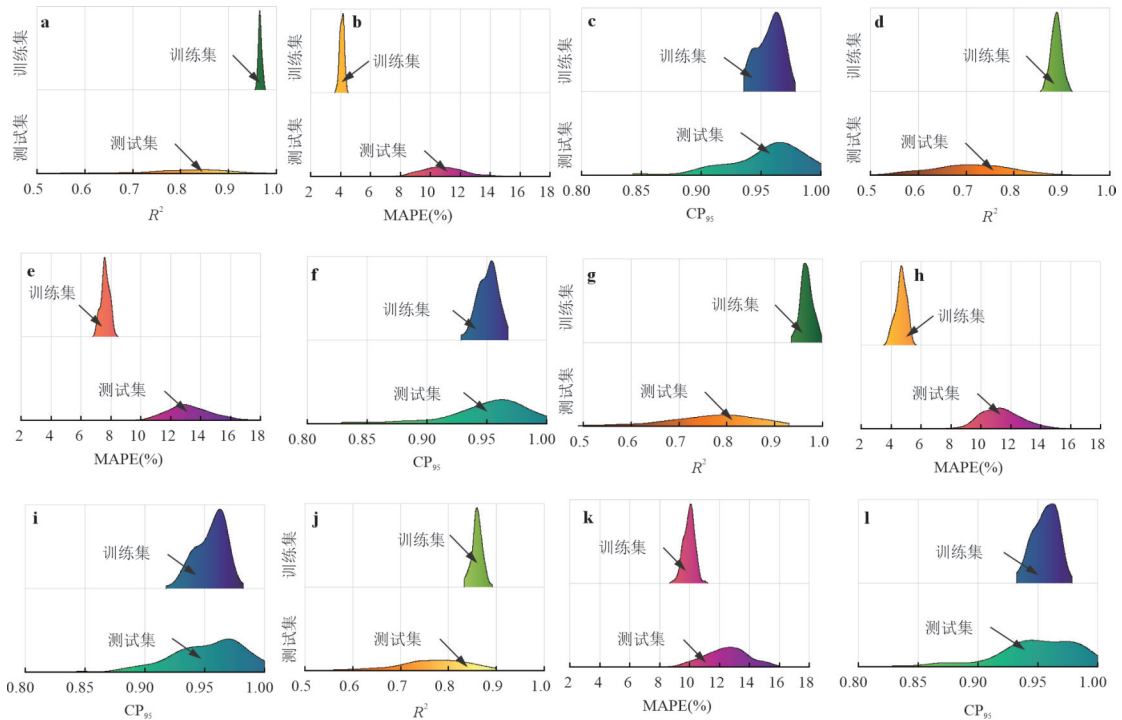


图 4 基于框架体系二的 100 次随机试验黄土黏聚力预测性能与不确定性量化结果统计特征分析

Fig. 4 Statistical characteristic results of predicted cohesion of loess and uncertainty quantification for 100 experiments based on unified framework 2

a. RF 方法: R^2 ; b. RF 方法: MAPE; c. RF 方法: CP_{95} ; d. DT 方法: R^2 ; e. DT 方法: MAPE; f. DT 方法: CP_{95} ; g. XGBoost 方法: R^2 ; h. XGBoost 方法: MAPE; i. XGBoost 方法: CP_{95} ; j. AdaBoost 方法: R^2 ; k. AdaBoost 方法: MAPE; l. AdaBoost 方法: CP_{95}

Boost 和 AdaBoost 4 种机器学习模型的预测性能及不确定性的量化结果,本节将对框架体系一的不确定性量化结果进行阐述,并与框架体系二进行对比.基于框架体系一,RF、DT、XGBoost 和 AdaBoost 4 种机器学习方法对应的训练集的 CP_{95} 分别为 0.97、0.95、0.93 和 0.97,测试集的 CP_{95} 分别为 0.67、0.90、0.90 和 0.89,相比于框架体系二的计算结果,4 种机器学习方法的测试集的 CP_{95} 分别下降 31.63%、6.25%、6.25% 和 7.29%.也就是说,通过框架体系一量化得到的 95% 置信区间是偏小的,这是由于测试集中的数据并未参与模型训练,为全新的测量数据,代入已有机器学习模型后预测值与测量值之间的误差相比于训练集偏大,因此导致通过框架体系一得到的不确定性应用于测试集时偏小.表 4 给出了 RF、DT、XGBoost 和 AdaBoost 4 种机器学习方法对应的预测偏差与各输入/输出特征之间的相关关系.可以看出,基于框架体系一得到的预测偏差与各输入/输出特征无明显相关性.

上述为一次随机试验的统计结果,为进一步验证基于框架体系一得到的预测偏差进行不确定性量化结果的鲁棒性与可靠性,按照与上述相同的训

练集、测试集比例进行 100 次随机试验,可得到 100 组对应的测试集的 CP_{95} ,如图 5 所示.图中 IQR 表示四分位距.由图 5 可以看出,RF、DT、XGBoost 和 AdaBoost 4 种机器学习方法对应的测试集的 CP_{95} 小提琴图的中值分别为 0.64、0.83、0.89 和 0.89,均小于 0.95,再次说明基于框架体系一进行黄土黏聚力概率预测,得到的 95% 置信区间(即量化的不确定性)是偏小的,与前述结论一致.此外,通过对比框架体系一和框架体系二的小提琴图可以发现,基于框架体系一进行不确定性量化,结果的合理性与可靠性相比于框架体系二略差.

4 讨论

以上探讨了基于不同机器学习方法的黄土关键力学参数概率预测统一框架体系进行不确定性量化,模型的概率预测结果与统计特征,核心为预测偏差的合理估计.类似地,本节基于预测残差对黄土黏聚力预测结果的不确定性进行刻画.预测残差定义为训练集中模型验证部分黄土黏聚力测量值 y_m^{train} 与预测值 y_p^{train} 的差值,记作 $\lambda_{mi} = y_m^{train}/y_p^{train}$.此

表 3 基于框架体系一的预测偏差与黄土黏聚力预测值及输入特征的斯皮尔曼相关系数及 P 值检验结果Table 3 Spearman's correlation coefficient between bias obtained from unified framework 1 and predicted cohesion/ input variables and corresponding results of P -value test

| 黄土参数 | RF | | DT | | XGBoost | | AdaBoost | |
|-------------|--------|-------|--------|-------|---------|-------|----------|-------|
| | r_s | P 值 | r_s | P 值 | r_s | P 值 | r_s | P 值 |
| y =黏聚力 | 0.418 | 0.000 | 0.024 | 0.693 | 0.196 | 0.001 | 0.305 | 0.000 |
| X_1 =埋深 | 0.116 | 0.053 | 0.134 | 0.025 | 0.026 | 0.667 | 0.175 | 0.003 |
| X_2 =含水率 | 0.083 | 0.167 | 0.019 | 0.749 | 0.001 | 0.990 | 0.091 | 0.131 |
| X_3 =孔隙比 | -0.144 | 0.016 | -0.119 | 0.047 | 0.017 | 0.781 | -0.279 | 0.000 |
| X_4 =液限 | 0.148 | 0.013 | 0.062 | 0.304 | 0.012 | 0.847 | 0.222 | 0.000 |
| X_5 =塑限 | 0.111 | 0.063 | 0.027 | 0.659 | 0.027 | 0.649 | 0.186 | 0.002 |
| X_6 =液性指数 | 0.074 | 0.216 | 0.005 | 0.930 | -0.008 | 0.888 | 0.054 | 0.365 |
| X_7 =塑性指数 | 0.154 | 0.010 | 0.077 | 0.201 | -0.002 | 0.971 | 0.224 | 0.000 |
| X_8 =饱和度 | 0.134 | 0.025 | 0.073 | 0.224 | -0.017 | 0.775 | 0.204 | 0.001 |
| X_9 =干密度 | 0.152 | 0.011 | 0.123 | 0.040 | 0.008 | 0.894 | 0.286 | 0.000 |

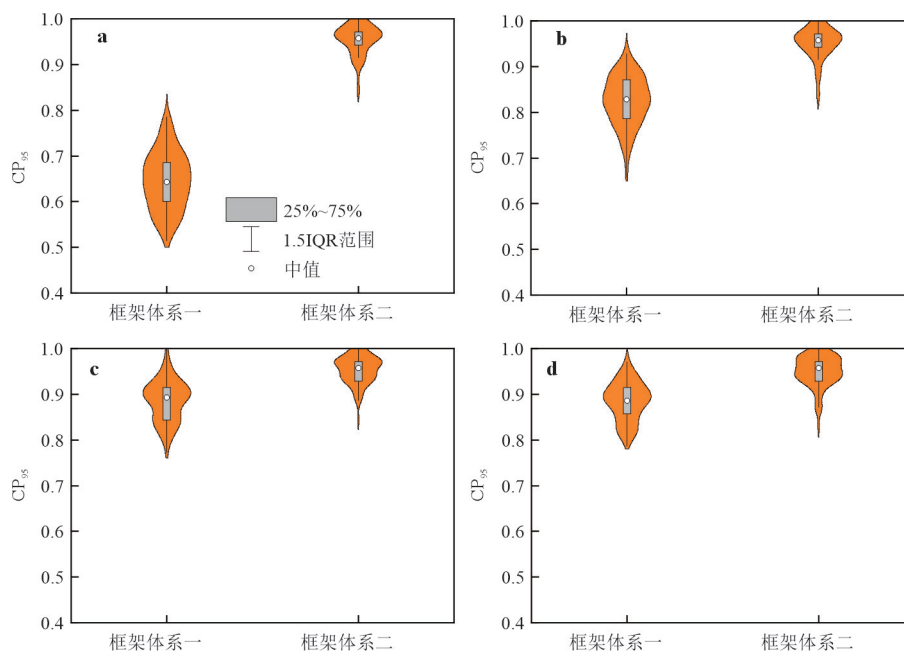


图 5 基于框架体系一和二的 100 次随机试验黄土黏聚力概率预测结果统计特征分析

Fig. 5 Statistical characteristic results of probabilistic prediction of cohesion of loess for 100 experiments based on unified framework 1 and 2

a. RF; b. DT; c. XGBoost; d. AdaBoost

时,预测残差 λ_{mi} 的范围为 $[-\infty, +\infty]$, 因此,可通过正态分布对预测残差 λ_{mi} 进行概率分布拟合,测试集的 95% 置信区间通过 y_p^{test} 与已有的训练集 95% 置信区间的上、下边界分别相加得到,其他步骤与第 2 章完全一致,由于文章篇幅有限,不再赘述.基于预测残差进行不确定性量化,RF、DT、XGBoost 和 AdaBoost 4 种机器学习方法对应的训练集的 CP_{95} 分别为 0.96、0.95、0.96 和 0.96,测试集分别为 0.97、0.96、0.97 和 0.97,均在 0.95 附近波动,说明基

于预测残差也可实现黄土黏聚力预测结果不确定性的合理量化.此外,表 4 给出了 RF、DT、XGBoost 和 AdaBoost 4 种机器学习方法对应的预测残差与各输入/输出特征之间的相关关系.可以看出,预测残差与各输入/输出特征无明显相关性.

上述为一次随机试验的计算结果,为进一步验证基于预测残差进行不确定性量化结果的鲁棒性与可靠性,重复上述过程 100 次,得到 100 组测试集的 CP_{95} ,如图 6 所示.可以看出,RF、DT、XGBoost

表 4 预测残差与黄土黏聚力预测值及输入特征的斯皮尔曼相关系数及 P 值检验结果

Table 4 Spearman's correlation coefficient between residual and predicted cohesion/ input variables and corresponding results of P-value test

| 黄土参数 | RF | | DT | | XGBoost | | AdaBoost | |
|-------------|--------|-------|--------|-------|---------|-------|----------|-------|
| | r_s | P 值 | r_s | P 值 | r_s | P 值 | r_s | P 值 |
| y =黏聚力 | 0.098 | 0.101 | -0.169 | 0.004 | -0.075 | 0.211 | 0.164 | 0.006 |
| X_1 =埋深 | 0.151 | 0.011 | 0.224 | 0.000 | 0.057 | 0.341 | 0.129 | 0.032 |
| X_2 =含水率 | 0.078 | 0.193 | 0.036 | 0.547 | 0.022 | 0.718 | 0.109 | 0.068 |
| X_3 =孔隙比 | -0.160 | 0.007 | -0.079 | 0.185 | -0.060 | 0.320 | -0.247 | 0.000 |
| X_4 =液限 | 0.132 | 0.027 | 0.079 | 0.186 | 0.067 | 0.266 | 0.231 | 0.000 |
| X_5 =塑限 | 0.094 | 0.115 | 0.052 | 0.382 | 0.056 | 0.352 | 0.183 | 0.002 |
| X_6 =液性指数 | 0.067 | 0.262 | 0.032 | 0.597 | 0.012 | 0.842 | 0.074 | 0.216 |
| X_7 =塑性指数 | 0.139 | 0.020 | 0.087 | 0.149 | 0.065 | 0.279 | 0.245 | 0.000 |
| X_8 =饱和度 | 0.143 | 0.016 | 0.070 | 0.242 | 0.044 | 0.461 | 0.212 | 0.000 |
| X_9 =干密度 | 0.168 | 0.005 | 0.082 | 0.169 | 0.066 | 0.275 | 0.256 | 0.000 |

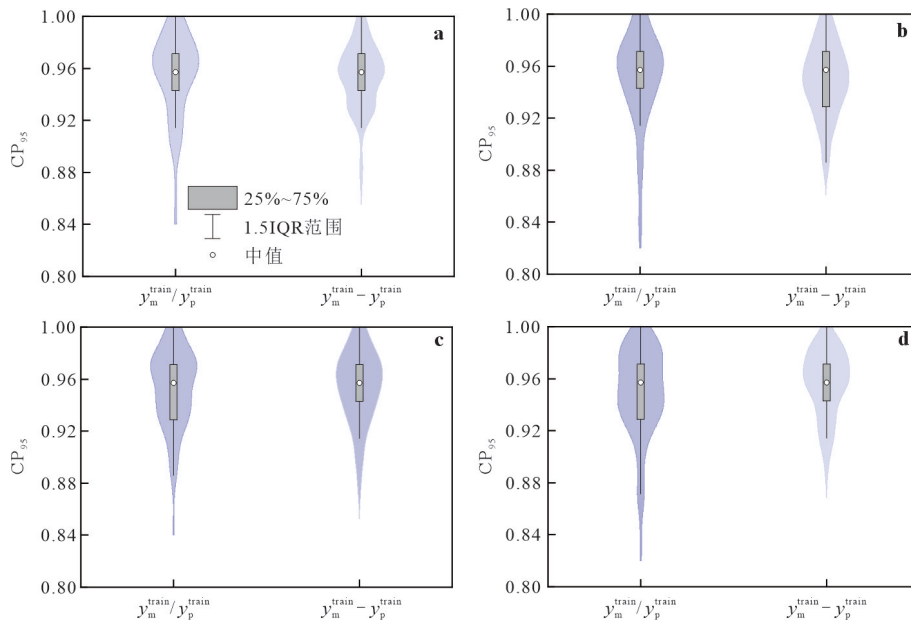


图 6 基于预测残差的 100 次随机试验黄土黏聚力概率预测结果统计特征分析

Fig. 6 Statistical characteristic results of probabilistic prediction of cohesion of loess for 100 experiments based on residual
a. RF; b. DT; c. XGBoost; d. AdaBoost

和 AdaBoost 4 种机器学习方法对应的测试集的 CP_{95} 小提琴图的中值均在 0.95 附近,与基于预测偏差的统计特征十分相似,再次证明了基于预测残差进行不确定性量化结果的合理性.然而,由于基于预测残差进行不确定性量化时,预测残差可能出现负值,与实际工程不符,因此,建议采用第 2 章所提的基于机器学习方法的黄土关键力学参数概率预测统一框架体系对黄土黏聚力进行概率预测.

5 结论

本文提出了基于机器学习方法的黄土关键力学参数概率预测统一框架体系,通过对训练集的预测偏差进行概率分布拟合,进而确定 95% 置信区间,实现了机器学习模型预测结果不确定性的合理刻画.通过一个粉黄土黏聚力预测工程实例验证了本文所提方法不确定性量化的可靠性、合理性和鲁棒性.主要结论如下:

(1) 通过训练集的测量值与预测值的比值定义

的偏差与机器学习模型的输入/输出特征之间无明显相关性。

(2) 基于框架体系一得到的 95% 置信区间(即预测结果的不确定性)偏小, 基于框架体系二得到的 95% 置信区间是更加可靠, 可对黄土黏聚力预测结果的不确定性进行合理量化。

(3) 基于本文所提方法可对不同机器学习模型预测结果的不确定性进行合理量化, 即使预测结果与实际情况偏差较大, 也可合理刻画其不确定性, 说明本文所提方法具有较高的鲁棒性与可靠性。

(4) 基于随机森林(RF)、决策树(DT)、极限梯度提升(XGBoost)和自适应提升(AdaBoost)4 种方法预测黄土黏聚力, 对应的 R^2 分别达到了 0.84、0.75、0.81 和 0.79, 说明基于上述 4 种方法与 9 个易获得的黄土物理力学参数可实现黄土黏聚力的相对准确的概率预测。

但本文所提方法仍存在一定的局限性: 当测量数据过少, 无法充分捕捉其统计特征时, 基于预测偏差拟合的对数正态分布可能与实际情况偏离, 进而可能导致不确定性量化的合理性降低。此时, 可采用 SMOTE 等数据增广算法首先基于已有测量数据生成部分模拟数据, 进而采用模拟数据和测量数据的组合建立黄土黏聚力概率预测模型, 再基于预测偏差进行概率分布拟合, 具体可参考 Song *et al.* (2024)。

References

- Abdi, Y., Momeni, E., Armaghani, D. J., 2023. Elastic Modulus Estimation of Weak Rock Samples Using Random Forest Technique. *Bulletin of Engineering Geology and the Environment*, 82(5): 176. <https://doi.org/10.1007/s10064-023-03154-y>
- Bao, T., Burghardt, J., 2022. A Bayesian Approach for In-Situ Stress Prediction and Uncertainty Quantification for Sub-surface Engineering. *Rock Mechanics and Rock Engineering*, 55(8): 4531–4548. <https://doi.org/10.1007/s00603-022-02857-0>
- Breiman, L., 2001. Random Forests. *Machine Learning*, 45(1): 5–32. <https://doi.org/10.1023/A:1010933404324>
- Chen, J. F., Zhao, Z. H., Zhang, J. T., 2024. Predicting Peak Shear Strength of Rock Fractures Using Tree-Based Models and Convolutional Neural Network. *Computers and Geotechnics*, 166: 105965. <https://doi.org/10.1016/j.compgeo.2023.105965>
- Chen, Y., Xu, Y. F., Jamhiri, B., et al., 2022. Predicting Uniaxial Tensile Strength of Expansive Soil with Ensemble Learning Methods. *Computers and Geotechnics*, 150: 104904. <https://doi.org/10.1016/j.compgeo.2022.104904>
- Ching, J., Phoon, K. K., Li, K.H., et al., 2019. Multivariate Probability Distribution for Some Intact Rock Properties. *Canadian Geotechnical Journal*, 56(8): 1080–1097. <https://doi.org/10.1139/cgj-2018-0175>
- Dang, J. Q., Li, J., 1997. Strength Characteristics of Unsaturated Loess. *Chinese Journal of Geotechnical Engineering*, (2): 59–64(in Chinese with English abstract).
- Dong, X.C., Guo, M.W., Wang, S.L., et al., 2023. Inclination Prediction of a Super-Sized Open Caisson Foundation During Sinking Process Based on Ensemble Learning. *Chinese Journal of Rock Mechanics and Engineering*, 42(S1): 3812–3822(in Chinese with English abstract).
- Ewusi-Wilson, R., Lee, C., Park, J., 2023. Artificial Intelligence - Optimized Design for Dynamic Compaction in Granular Soils. *Acta Geotechnica*, 19(6): 3487–3503. <https://doi.org/10.1007/s11440-023-02081-2>
- Jing, Y.L., Wu, Y.Q., Lin, D.J., et al., 2011. Study of Relationship Between Loess Collapsibility and Index of Compaction Test. *Rock and Soil Mechanics*, 32(2): 393–397(in Chinese with English abstract).
- Kardani, N., Aminpour, M., Nouman Amjad Raja, M., et al., 2022. Prediction of the Resilient Modulus of Compacted Subgrade Soils Using Ensemble Machine Learning Methods. *Transportation Geotechnics*, 36: 100827. <https://doi.org/10.1016/j.trgeo.2022.100827>
- Li, S.Y., Chen, X., Lu, J.Q., et al., 2024. Real-Time Discrimination Model for Local Earthquake Intensity Threshold Based on XGBoost. *Earth Science*, 49(2): 379–390(in Chinese with English abstract).
- Liu, D., Lin, P. Y., Zhao, C. Y., et al., 2021. Mapping Horizontal Displacement of Soil Nail Walls Using Machine Learning Approaches. *Acta Geotechnica*, 16(12): 4027–4044. <https://doi.org/10.1007/s11440-021-01345-z>
- Liu, Q. S., Wang, X. Y., Huang, X., et al., 2020. Prediction Model of Rock Mass Class Using Classification and Regression Tree Integrated AdaBoost Algorithm Based on TBM Driving Data. *Tunnelling and Underground Space Technology*, 106: 103595. <https://doi.org/10.1016/j.tust.2020.103595>
- Nguyen, T., Ly, D. K., Huynh, T. Q., et al., 2023. Soft Computing for Determining Base Resistance of Super-Long Piles in Soft soil: A Coupled SPBO-XGBoost Approach. *Computers and Geotechnics*, 162: 105707. <https://doi.org/10.1016/j.compgeo.2023.105707>
- Song, C., Zhao, T. Y., Xu, L., et al., 2024. Probabilistic

- Prediction of Uniaxial Compressive Strength for Rocks from Sparse Data Using Bayesian Gaussian Process Regression with Synthetic Minority Oversampling Technique (SMOTE). *Computers and Geotechnics*, 165: 105850. <https://doi.org/10.1016/j.compgeo.2023.105850>
- Song, C., Zhao, T. Y., Xu, L., 2023. Estimation of Uniaxial Compressive Strength Based on Fully Bayesian Gaussian Process Regression and Model Selection. *Chinese Journal of Geotechnical Engineering*, 45(8): 1664–1673(in Chinese with English abstract).
- Wen, L. F., Li, Y. L., Zhao, W. B., et al., 2023. Predicting the Deformation Behaviour of Concrete Face Rockfill Dams by Combining Support Vector Machine and AdaBoost Ensemble Algorithm. *Computers and Geotechnics*, 161: 105611. <https://doi.org/10.1016/j.compgeo.2023.105611>
- Wu, L. Y., Li, J. H., Ma, D., et al., 2023. Prediction for Rock Compressive Strength Based on Ensemble Learning and Bayesian Optimization. *Earth Science*, 48(5): 1686–1695 (in Chinese with English abstract).
- Xu, L., Zhou, G. P., Zhao, T. Y., et al., 2023. Characterization of Inherent Spatial Variability of Loess Deposit Properties in Shaanxi Province, China. *Journal of Soils and Sediments*, 23(7): 2862–2877. <https://doi.org/10.1007/s11368-023-03517-8>
- Yan, D. D., Zhao, T. Y., Xu, L., et al., 2023. Statistical Modeling of Multivariate Loess Properties in Taiyuan Using Regular Vine Copula with Optimized Tree Structure. *Transportation Geotechnics*, 41: 101025. <https://doi.org/10.1016/j.trgeo.2023.101025>
- Yang, L., Wei, J., 2023. Prediction of Rockburst Intensity Grade Based on SVM and Adaptive Boosting Algorithm. *Earth Science*, 48(5): 2011–2023(in Chinese with English abstract).
- Zhang, J. R., Song, C. Y., Jiang, T., et al., 2023. Hydromechanical Characteristics and Microstructure of Unsaturated Loess Under High Suction. *Rock and Soil Mechanics*, 44(8): 2229–2237(in Chinese with English abstract).
- Zhang, L., Wang, M., Zhao, H. B., et al., 2022a. Uncertainty Quantification for the Mechanical Behavior of Fully Grouted Rockbolts Subjected to Pull-out Tests. *Computers and Geotechnics*, 145: 104665. <https://doi.org/10.1016/j.compgeo.2022.104665>
- Zhang, P., Yin, Z. Y., Jin, Y. F., 2022b. Bayesian Neural Network-Based Uncertainty Modelling: application to Soil Compressibility and Undrained Shear Strength Prediction. *Canadian Geotechnical Journal*, 59(4): 546–557. <https://doi.org/10.1139/cgj-2020-0751>
- Zhang, W. G., Li, H. R., Tang, L. B., et al., 2022c. Displacement Prediction of Jiuxianping Landslide Using Gated Recurrent Unit (GRU) Networks. *Acta Geotechnica*, 17(4): 1367–1382. <https://doi.org/10.1007/s11440-022-01495-8>
- Zhang, W. G., Wu, C. Z., Zhong, H. Y., et al., 2021. Prediction of Undrained Shear Strength Using Extreme Gradient Boosting and Random Forest Based on Bayesian Optimization. *Geoscience Frontiers*, 12(1): 469–477. <https://doi.org/10.1016/j.gsf.2020.03.007>
- Zhao, T. Y., Song, C., Lu, S. F., et al., 2022. Prediction of Uniaxial Compressive Strength Using Fully Bayesian Gaussian Process Regression (fB - GPR) with Model Class Selection. *Rock Mechanics and Rock Engineering*, 55(10): 6301–6319. <https://doi.org/10.1007/s00603-022-02964-y>
- Zuo, L., Xu, L., Baudet, B. A., et al., 2024. Small-Strain Shear Stiffness Anisotropy of a Saturated Clayey Loess. *Géotechnique*, 74(4): 325–336. <https://doi.org/10.1680/jgeot.21.00179>

中文参考文献

- 党进谦, 李靖, 1997. 非饱和黄土的强度特征. *岩土工程学报*, 19(2): 56–61.
- 董学超, 郭明伟, 王水林, 等, 2023. 基于集成学习的超大型沉井基础下沉倾斜程度预测. *岩石力学与工程学报*, 42(S1): 3812–3822.
- 井彦林, 仵彦卿, 林杜军, 等, 2011. 黄土的湿陷性与击实试验指标关系研究. *岩土力学*, 32(2): 393–397.
- 李山有, 陈欣, 卢建旗, 等, 2024. 基于XGBoost的现地地震烈度阈值实时判别模型. *地球科学*, 49(2): 379–390.
- 宋超, 赵腾远, 许领, 等, 2023. 基于贝叶斯高斯过程回归与模型选择的岩石单轴抗压强度估计方法. *岩土工程学报*, 45(8): 1664–1673.
- 吴禄源, 李建会, 马丹, 等, 2023. 基于集成学习与贝叶斯优化的岩石抗压强度预测. *地球科学*, 48(5): 1686–1695.
- 杨玲, 魏静, 2023. 基于支持向量机和增强学习算法的岩爆烈度等级预测. *地球科学*, 48(5): 2011–2023.
- 张俊然, 宋陈雨, 姜彤, 等, 2023. 非饱和黄土高吸力下的水力学特性及微观结构分析. *岩土力学*, 44(8): 2229–2237.