

DOI: 10.11779/CJGE2023S20025

# 基于集成学习模型的正常固结土抗剪强度指标预测方法

王钰轲<sup>1,2,3</sup>, 冯爽<sup>1,2,3</sup>, 钟燕辉<sup>1,2,3</sup>, 张蓓<sup>1,2,3</sup>

(1. 郑州大学水利与交通学院, 河南 郑州 450001; 2. 重大基础设施检测修复技术国家地方联合工程实验室, 河南 郑州 450001;  
3. 地下工程灾变防控省部共建协同创新中心, 河南 郑州 450001)

**摘要:** 抗剪强度指标的准确获取对工程设计具有决定性作用, 目前试验方法确定的强度指标过于依靠工程经验, 导致最终的取值具有不确定性。集成学习模型是机器学习的一个子类, 在处理复杂的数据和任务时表现出强大的性能。为了能更准确地获得抗剪强度指标, 以正常固结土为研究对象, 利用不同的集成学习算法建立其抗剪强度指标的预测模型。通过均方根误差 (RMSE)、可决系数 ( $R^2$ )、绝对值误差 (MAE) 评估不同模型的泛化能力, 并采用 Adaboost 算法进行输入参数的敏感性分析。结果表明抗剪强度采用 Adaboost、内摩擦角采用 RF、黏聚力采用 Adaboost 算法具有最佳的泛化能力, 其测试集  $R^2$  可分别达到 0.925, 0.965, 0.942。敏感性分析结果显示, 对抗剪强度影响最大的参数为干密度、含水率和法向应力; 对内摩擦角影响最大的参数为曲率系数、黏粒含量和含水率; 对黏聚力影响最大的参数为含水率、干密度和液限。本文所建立的预测模型可为工程中抗剪强度指标的选取及运用机器学习方法研究土体强度参数提供参考。

**关键词:** 正常固结土; 抗剪强度指标; 集成学习; 机器学习; 敏感性分析

中图分类号: TU43 文献标识码: A 文章编号: 1000-4548(2023)S2-0183-06

作者简介: 王钰轲(1989—), 男, 博士, 教授, 主要从事土的基本特性与本构理论相关的研究工作。E-mail: ykewang@163.com。

## A data-driven model for predicting shear strength indexes of normally consolidated soils

WANG Yuke<sup>1,2,3</sup>, FENG Shuang<sup>1,2,3</sup>, ZHONG Yanhui<sup>1,2,3</sup>, ZHANG Bei<sup>1,2,3</sup>

(1. School of Water Conservancy and Transportation, Zhengzhou University, Zhengzhou 450001, China; 2. National Local Joint Engineering Laboratory of Major Infrastructure Testing and Rehabilitation Technology, Zhengzhou 450001, China; 3. Provincial and Ministerial Collaborative Innovation Center for Underground Engineering Disaster Prevention and Control, Zhengzhou 450001, China)

**Abstract:** The accurate determination of shear strength indexes is crucial for engineering soils. Currently, the strength indexes determined by the test method depend too much on engineering experience, which leads to the uncertainty of the final value. The ensemble learning is a subclass of machine learning that exhibits strong performance when dealing with complex data and tasks. To enhance the precision of the shear strength indexes, based on the normally consolidated soils, a model is established using the diverse ensemble learning algorithms for predicting the shear strength indexes of soils. Various models are assessed for their generalization capability using the root mean square error (RMSE), coefficient of determination ( $R^2$ ) and absolute value error (MAE). The Adaboost algorithm is employed for the sensitivity analysis of input parameters. The findings indicate that the Adaboost algorithm yields the best generalization for the shear strength the RF for the internal friction angle, and the Adaboost algorithm for the cohesion, achieving respective test-set  $R^2$  values of 0.925, 0.965 and 0.942. The sensitivity analyses reveal that the dry density, moisture content and normal stress exert the most significant influence on the shear strength, while the key factors for the internal friction angle are the coefficient of curvature, viscous grain content and water content. The water content, dry density and liquid limit are identified as the primary influencers on the cohesion. The data-driven model established herein offers guidance for selecting the shear strength indexes in engineering and investigating strength parameters of soils through the machine learning methods.

**Key words:** normally consolidated soil; shear strength index;

ensemble learning; machine learning; sensitivity analysis

基金项目: 国家自然科学基金项目(52109140); 河南省优秀青年基金项目(232300421069); 中原科技创新领军人才资助项目(234200510014); 河南省交通运输厅科技项目(2022-5-5)

收稿日期: 2023-11-30

## 0 引言

在岩土工程中, 基础、挡土墙、边坡等土工结构的设计以及路堤稳定性、土压力和地基承载力的计算等均需用到土的抗剪强度指标。抗剪强度指标的选取关乎土工结构的设计与安全, 不合理的取值会造成工程灾害发生、工程造价提高等不良影响<sup>[1]</sup>, 中国某些软土地基基坑坍塌就是由于强度指标选择不当而造成的。但土的抗剪强度指标受到多种因素的影响, 首先取决于土本身的物理性质, 即土的组成、土的状态和土的结构, 其次还取决于它所受的应力状态<sup>[2]</sup>, 各因素之间相互作用及其影响机理较为复杂<sup>[3]</sup>, 难以解释, 因此目前主要还是依靠室内试验和现场原位试验确定其取值。

胡海军等<sup>[4]</sup>比较了不同试验测试方法及其适用性, 指出试验方法的选择对确定抗剪强度指标有很大的影响, 且不同施工期需采用不同的抗剪强度指标; 闫澎旺等<sup>[5]</sup>指出十字板剪切试验不能准确反映实际滑动面的强度, 并通过大量的试验结果回归分析得到软黏土的抗剪强度指标, 但其回归方程的某些参数也是依靠经验取值。由于试验手段获取的抗剪强度指标会受到试验条件的限制, 具有精度难以保证以及过于依靠“经验”而导致的不确定性等缺点, 越来越多的学者选择采用数值分析或机器学习等非试验方法获取抗剪强度指标<sup>[6-7]</sup>。其中, 机器学习 (machine learning) 是人工智能领域的 1 个子领域<sup>[8]</sup>, 它可以通过让计算机系统自动识别数据中的模式、规律和趋势, 从而使其能够对目标参数进行准确的预测和决策。随着近些年人工智能领域的发展, 机器学习在工程领域中已经得到大量的应用并表现出惊人的能力<sup>[9]</sup>。

对于土体抗剪强度指标的预测, 也有不少学者采用不同的机器学习算法展开了大量研究。Das 等<sup>[10]</sup>采用人工神经网络 ANN 预测黏土的内摩擦角, 其测试集的可决系数  $R^2$  达到 0.868; Pham 等<sup>[11]</sup>采用自适应模糊推理系统 ANFIS 预测软土的抗剪强度, 并用 SVM, ANN 两种算法做对比, 得出结论粒子群优化的 ANFIS 的预测效果最好, 其拟合度为 0.601; 刘晓燕等<sup>[12]</sup>采用回归树和模型树对黏土的不排水抗剪强度展开预测, 其用模型树法预测的测试集可决系数  $R^2$  达到 0.95; Rabbani 等<sup>[13]</sup>采用灰狼算法等 6 种启发式算法对 ANN 进行优化, 训练土体抗剪强度预测模型, 最终得到增强灰狼算法优化 ANN 的预测效果最好, 其测试集的可决系数  $R^2$  达到 0.819; 顾春生等<sup>[14]</sup>采用 PAC 优化 BP 神经网络的方法预测黏性土的抗剪强度, 其抗剪强度拟合度达到 0.85, 内摩擦角拟合度达到 0.72。

虽然学者们已经使用各种算法对抗剪强度指标的预测模型进行了研究, 但是这些模型的预测精度并未达到理想水平。且以上研究尚未针对抗剪强度  $\tau_f$ 、内摩擦角  $\varphi$  和黏聚力  $c$  这 3 个指标进行统一的建模研究。集成学习模型是一种将多个算法模型集成在一起以提高模型性能的方法。它有几种不同的方法, 包括 Bagging, Boosting 和堆叠法, 在处理复杂数据和任务时表现出强大的能力, 具有泛化能力高、鲁棒性好等优点。针对上述问题, 采用两种集成学习算法 (RF 算法和 Adaboost 算法) 对  $\tau_f$ ,  $\varphi$ ,  $c$  3 个指标的预测进行统一的建模研究。采用 RMSE3,  $R^2$ , MAE 评估不同预测模型的泛化能力, 使该预测模型具备泛化误差小、稳定性好、不易过拟合等特点。最后采用 Adaboost 算法对输入参数进行敏感性分析, 找到对抗剪强度、内摩擦角、黏聚力影响最大的输入参数。

## 1 数据集准备及预处理

### 1.1 特征筛选

从影响土体抗剪强度的因素出发, 考虑土体的抗剪强度首先取决于其基本物理性质: 即土的组成、土的状态和土的结构, 其次还取决于它当前所受的应力状态, 因此选择用土体的基本物理参数以及所受的应力状态来预测其抗剪强度及其强度指标。故将模型的输入参数确定为相对质量密度  $G_s$ 、塑限  $w_p$ 、液限  $w_l$ 、黏粒含量  $d_{0.005}$ 、塑性指数  $I_p$ 、不均匀系数  $C_u$ 、曲率系数  $C_c$ 、含水率  $w$ 、干密度  $\rho_d$  以及所受的法向应力  $\sigma$ 。输出参数为抗剪强度  $\tau_f$  和抗剪强度指标为内摩擦角  $\varphi$ 、黏聚力  $c$ 。

### 1.2 数据集准备及统计值描述

建立一个精确的抗剪强度指标预测模型, 必须收集大量的试验数据, 并通过机器学习算法对这些试验数据展开研究, 找到输入变量和输出变量之间复杂的映射关系, 数据集的质量对预测模型的精度起到决定性的作用。从文献[15~19]中共收集到 695 条有效的抗剪强度指标的试验数据, 构成训练模型的数据集。绘制了  $\tau_f$ ,  $\varphi$ ,  $c$  这 3 个输出参数的直方分布图, 如图 1 所示, 以更清晰地观察其数据分布情况。并绘制了 3 个参数的箱型图进行异常值检测, 如图 2 所示。据图 2 显示, 抗剪强度和黏聚力均存在异常值的情况, 且大部分异常值是超越了箱型图上限, 因此在训练模型之前需将这些异常值剔除, 内摩擦角则几乎不存在异常值的情况。将在拟合曲线上方的数据视为异常值, 经筛选, 抗剪强度存在 11 个异常值, 黏聚力存在 3 个异常值, 将异常值所在的数据列删除, 最终用于训练模型的数据集为 681 条有效数据。

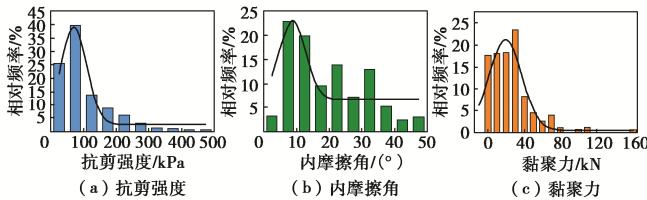


图 1 输出参数的数据分布图

Fig. 1 Data distribution of output parameters

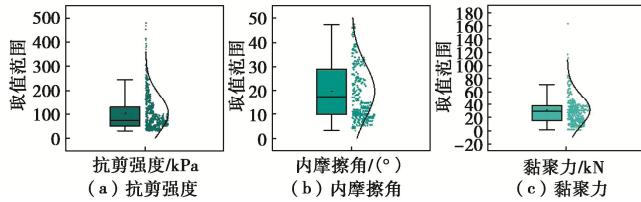


图 2 输出参数异常值检测图

Fig. 2 Outlier detection of output parameters

### 1.3 相关性分析

为了防止所建立的模型出现多重共线性, 需要在训练模型之前进行输入变量之间的相关性分析。

Spearman 相关分析方法可用于衡量两个变量之间的非线性关系, 且 Spearman 相关系数不要求数据满足正态分布或其他特定分布, 对异常值相对较不敏感, 具有较好的鲁棒性<sup>[20]</sup>。因此, 采用 Spearman 相关系数法计算所有输入参数之间的相关系数。Spearman 相关系数的计算公式为

$$r_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (1)$$

式中:  $d_i$  为  $x_i$  与  $y_i$  的等级差;  $n$  为样本组数。

图 3 为各输入参数的 Spearman 等级相关系数矩阵热图。 $|r_s|$  越接近于 1, 说明两个参数之间的相关性越强, 正值代表两个参数之间具有正相关性, 负值代表两个参数之间具有负相关性。可以看出, 对角线上的相关系数为 1, 而其他大部分位置的相关系数都小于 0.6, 说明不同输入参数之间的相关性较低。因此, 预测精度不受输入变量自相关性的影响。

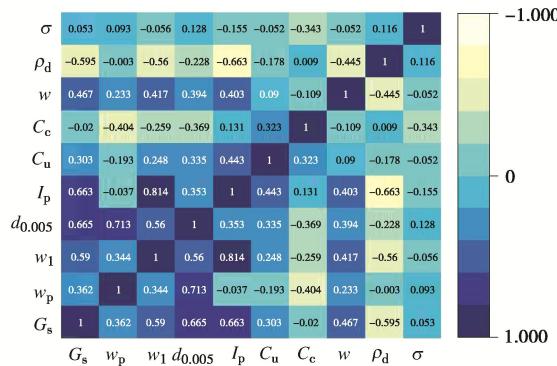


图 3 输入变量相关系数热图

Fig. 3 Correlation coefficients of input variables

## 2 集成学习算法及训练策略

### 2.1 集成学习模型

集成学习是一种通过结合多个学习器来提高整体性能的机器学习方法。根据个体学习器的生存方式, 集成学习大致可分为: ①个体学习器间存在强依赖关系的序列化方法, 如 Adaboost 算法; ②个体学习器间不存在依赖关系的并行化方法, 如随机森林算法。

其中, Adaboost 算法属于序列化集成学习算法, 其核心思想是通过结合多个弱学习器(通常是简单且性能较差的学习器)来构建一个强学习器<sup>[21]</sup>。将该算法用于回归任务中的优点包括该算法对噪声数据具有鲁棒性、在处理复杂问题时的高准确性等, 并且不需要特别复杂的弱学习器。然而, 需要注意的是数据的离散化可能会导致信息损失, 并且 Adaboost 在处理高维数据时效果不如其他算法, Adaboost 算法的原理示意图见图 4。

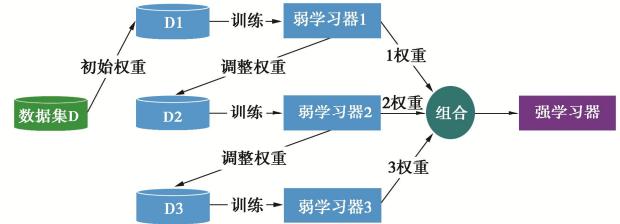


图 4 Adaboost 算法原理示意图

Fig. 4 Schematic diagram of Adaboost algorithm

随机森林 (RF) 是 Bagging 的一个扩展变体, 它是以决策树为基学习器构建 Bagging 集成的基础上, 在决策树的训练过程中引入随机属性的选择。随机森林的训练效率常优于 Bagging<sup>[22]</sup>, 因为在个体决策树的构建过程中, Bagging 使用的是“确定型”决策树, 在选择划分属性时要对节点的所有属性进行考察; 而随机森林使用的是“随机型”决策树, 只需考察一个属性子集。随机森林简单、容易实现、计算代价小, 但在很多现实任务中展现出强大的性能, RF 算法的原理示意图见图 5。

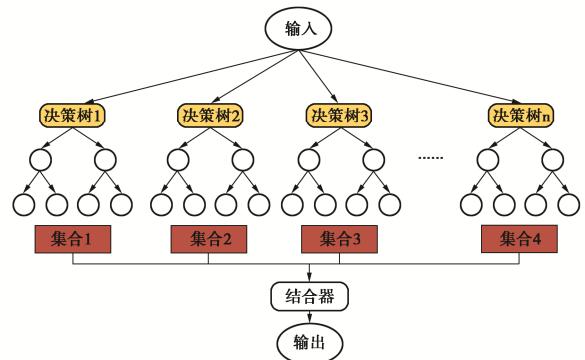


图 5 随机森林算法结构示意图

Fig. 5 Structure of random forest algorithm

## 2.2 训练策略

将数据集按照 8:2 的比例划分为训练集和测试集, 训练集用于模型的训练, 测试集用于评估模型的泛化能力。由于输入变量和输出变量的原始数据跨度范围相当大, 为了避免奇异样本数据的出现, 减少训练时间, 使模型更快收敛, 对原始数据基于以下线性映射函数进行归一化处理:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad . \quad (2)$$

式中:  $x'$  为各参数的归一化数据;  $x_{\max}$  为样本数据的最大值;  $x_{\min}$  为数据样本中的最小值。

为了比较不同模型的泛化能力, 使用评估回归模型最常用的 RMSE,  $R^2$ , MAE 3 个指标来评估各模型的预测性能:

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \bar{y}_i)^2} \quad , \quad (3)$$

$$R^2 = 1 - \frac{\sum_{i=1}^m (y_i - \bar{y}_i)^2}{\sum_{i=1}^m (y_i - \bar{y}_i)^2} \quad , \quad (4)$$

$$MAE = \frac{1}{m} \sum_{i=1}^m |y_i - \bar{y}_i| \quad . \quad (5)$$

式中:  $y_i$  为真实的观测值,  $\bar{y}_i$  为预测值,  $\bar{y}_i$  为真实观测值的平均值。

## 3 结果与讨论

### 3.1 泛化能力比较

在预测抗剪强度的任务中, RF 算法和 Adaboost 算法在训练集和测试集上的 3 个评价指标统计于表 1。可决系数  $R^2$  最接近于 1 的算法是 Adaboost 算法, 训练集和测试集分别达到 0.998, 0.925。绝对值误差 MAE 最小的算法是 Adaboost 算法, 训练集和测试集分别为 0.942, 13.272。均方根误差 RMSE 最小的算法是 Adaboost 算法, 训练集和测试集的 RMSE 分别为 3.809, 22.244。综合 3 个评价指标, Adaboost 算法预

测土体抗剪强度的能力优于 RF 算法。

表 1  $\tau_f$ ,  $\varphi$ ,  $c$  预测模型各指标统计表

Table 1 Indexes of prediction model for  $\tau_f$ ,  $\varphi$  and  $c$

模型	$R^2$		MAE		RMSE	
	训练	测试	训练	测试	训练	测试
$\tau_f$	RF	0.967	0.932	10.717	15.528	13.947
	Adaboost	0.998	0.925	0.941	13.274	3.809
$\varphi$	RF	0.997	0.965	0.901	1.469	1.167
	Adaboost	0.991	0.964	0.192	0.579	0.413
$c$	RF	0.952	0.910	2.860	3.978	4.087
	Adaboost	0.978	0.942	0.734	1.251	2.777

在预测内摩擦角的任务中, RF 算法和 Adaboost 算法在训练集和测试集上的 3 个评价指标统计于表 1。可决系数  $R^2$  最接近于 1 的算法是 RF 算法, 训练集和测试集分别达到 0.997, 0.965。绝对值误差 MAE 最小的算法是 Adaboost, 训练集和测试集分别为 0.192, 0.579。均方根误差 RMSE 最小的算法是 RF, 训练集和测试集的 RMSE 分别为 1.167, 1.536。但是 Adaboost 算法训练集和测试集的 RMSE 分别为 0.413, 2.189, 测试集的误差远大于训练集的误差, 可能出现了过拟合的现象。综合 3 个评价指标, RF 算法预测内摩擦角的能力优于 Adaboost 算法, 且 Adaboost 算法预测内摩擦角可能会出现过拟合。

在预测黏聚力的任务中, RF 算法和 Adaboost 算法在训练集和测试集上的 3 个评价指标统计于表 1。 $R^2$  最接近于 1 的算法是 Adaboost 算法, 训练集和测试集分别达到 0.978, 0.942。MAE 最小的算法是 Adaboost, 训练集和测试集分别为 0.734, 1.251。RMSE 最小的算法是 Adaboost, 训练集和测试集的 RMSE 分别为 2.777, 4.377。综合 3 个评价指标, Adaboost 算法预测黏聚力的能力优于 RF 算法。

### 3.2 不同模型综合能力比较结果

根据以上分析, 综合考虑模型的泛化能力, 预测抗剪强度、内摩擦角、黏聚力表现能力最好的算法分别是 Adaboost, RF, Adaboost。3 个算法在测试集中真实值和预测值的拟合参考图见图 6。

其中使用 Adaboost 算法预测内摩擦角时, 训练结

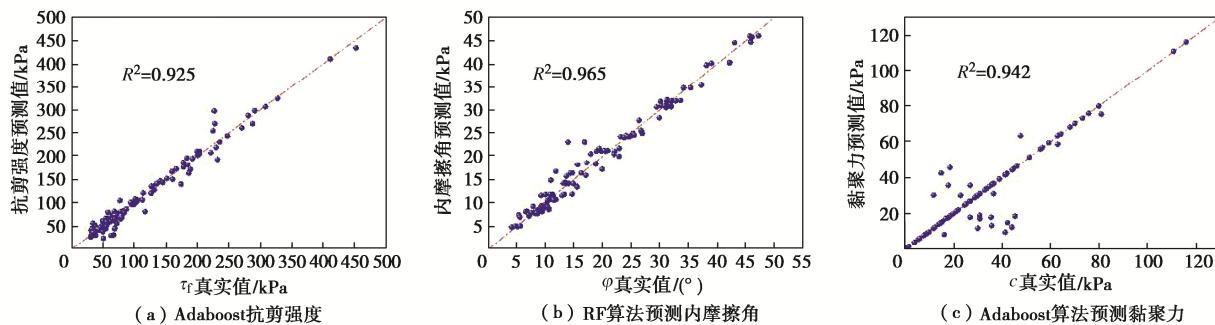


图 6 3 个模型测试集中真实值、预测值拟合参考图

Fig. 6 Reference plots of true and predicted values fitted in three model test sets

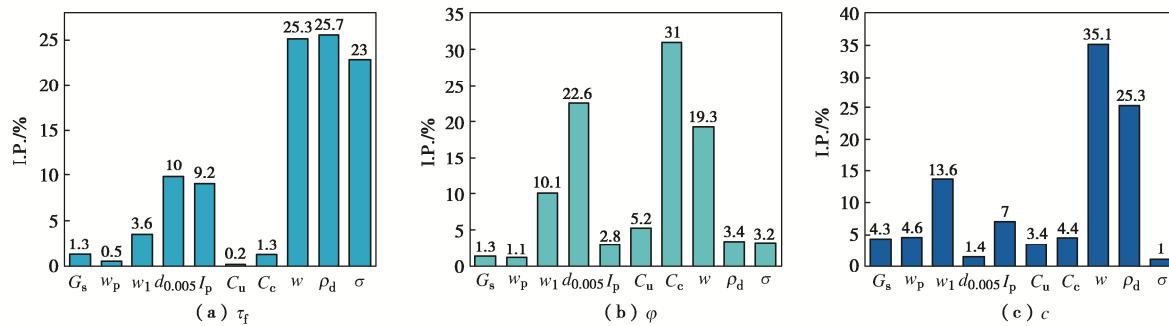


图 7 敏感性分析结果

Fig. 7 Results of sensitivity analysis

果显示测试集的误差远大于训练集的误差, 可能出现了过拟合的现象。究其原因, 可能为: ①AdaBoost 算法对异常值比较敏感, 因为这些异常值可能在每一轮的训练过程中得到强调, 从而影响模型的泛化能力。②Adaboost 算法采用的弱学习器数量过多, 虽然增加弱学习器的数量可以提高模型的复杂度, 但也容易导致过拟合。

### 3.3 输入参数敏感性分析

由于 AdaBoost 是一种迭代的学习方法, 每一轮都在调整样本的权重以关注先前被错误分类的样本, 因此对参数的敏感性较高。在 Adaboost 算法训练完成后会得到每个弱分类器的权重, 每个弱分类器对应一个参数, 这样就可以识别不同参数对模型性能的影响程度<sup>[23]</sup>。因此, 为了评估模型的输入参数对抗剪强度指标的影响程度, 采用 Adaboost 算法对参数做敏感性分析, 分析结果见图 7。根据训练结果可以得到, 对抗剪强度  $\tau_f$  影响程度最大的 3 个参数为干密度、含水率和法向应力, 分别占比 25.7%, 25.3%, 23%; 对内摩擦角影响程度最大的 3 个参数为曲率系数、黏粒含量和含水率, 分别占比 31%, 22.6%, 19.3%; 对黏聚力影响程度最大的 3 个参数是含水率、干密度和液限, 分别占比 35.1%, 25.3%, 13.6%。在使用数据驱动模型预测相应指标时, 若数据集缺失某些输入参数, 可优先考虑将以上参数作为模型的输入参数。

## 4 结 论

提出了一种基于集成学习的正常固结土抗剪强度指标预测模型, 利用土体常用的物理参数来预测土体的抗剪强度、内摩擦角和黏聚力, 为工程中抗剪强度指标的选取提供参考。首先通过数据集的扩建及预处理, 使用集成学习算法 (RF, Adaboost) 进行模型的构建, 然后利用 RMSE,  $R^2$ , MAE 3 个评价指标评估模型的泛化能力, 最后采用 Adaboost 算法进行输入参数的敏感性分析, 得到 2 点结论。

(1) 所建立的模型对  $\tau_f$ ,  $\phi$ ,  $c$  3 个参数的预测均取得了不错的效果。比较各模型的泛化能力, 预测

抗剪强度效果最好的算法是 Adaboost 算法, 其测试集的  $R^2$  可达 0.925; 预测内摩擦角效果最好的算法是 RF, 其测试集的  $R^2$  可达 0.965; 预测黏聚力效果最好的算法是 Adaboost, 其测试集的  $R^2$  可达 0.942。

(2) 敏感性分析结果表明: 对抗剪强度影响最大的参数是干密度 (25.7%)、含水率 (25.3%)、法向应力 (23%); 对内摩擦角影响最大的 3 个参数是曲率系数 (31%)、黏粒含量 (22.6%)、含水率 (19.3%); 对黏聚力影响最大的 3 个参数是含水率 (35.1%)、干密度 (25.3%)、液限 (13.6%)。

### 参 考 文 献:

- [1] VANAPALLI S K, FREDLUND D G. Comparison of different procedures to predict unsaturated soil shear strength[C]// Advances in Unsaturated Geotechnics. Denver, 2000.
- [2] 雷洁, 张国明, 刘连友, 等. 土壤抗剪强度测定与影响因素研究进展[J]. 北京师范大学学报(自然科学版), 2016, 52(4): 486-492. (LEI Jie, ZHANG Guoming, LIU Lianyou, et al. Measuring soil shear strength and impact factors[J]. Journal of Beijing Normal University (Natural Science), 2016, 52(4): 486-492. (in Chinese))
- [3] 雷国辉, 高翔, 徐可, 等. 饱和软黏土固结过程中的不排水抗剪强度特性[J]. 岩土工程学报, 2019, 41(1): 41-49. (LEI Guohui, GAO Xiang, XU Ke, et al. Behavior of undrained shear strength of saturated soft clay under consolidation[J]. Chinese Journal of Geotechnical Engineering, 2019, 41(1): 41-49. (in Chinese))
- [4] 胡海军, 魏丽敏, 冷伍明. 黏土强度指标的测定和选取[J]. 岩土力学, 2007, 28(12): 2698-2704. (HU Haijun, WEI Limin, LENG Wuming. Determination and choice of strength index of clay[J]. Rock and Soil Mechanics, 2007, 28(12): 2698-2704. (in Chinese))
- [5] 闫澍旺, 封晓伟, 侯晋芳, 等. 用十字板强度推算软黏土抗剪强度指标的方法及应用[J]. 岩土工程学报, 2009, 31(12): 1805-1810. (YAN Shuwang, FENG Xiaowei, HOU Jinfang, et al. Deduction and application of strength parameters of soft

clay by use of vane strength[J]. Chinese Journal of Geotechnical Engineering, 2009, **31**(12): 1805-1810. (in Chinese))

[6] ZHANG P, YIN Z Y, JIN Y F. Machine learning-based modelling of soil properties for geotechnical design: review, tool development and comparison[J]. Archives of Computational Methods in Engineering, 2022, **29**(2): 1229-1245.

[7] ZHANG W G, LI H R, LI Y Q, et al. Application of deep learning algorithms in geotechnical engineering: a short critical review[J]. Artificial Intelligence Review, 2021, **54**(8): 5633-5673.

[8] LI Y Y, RAHARDJO H, SATYANAGA A, et al. Soil database development with the application of machine learning methods in soil properties prediction[J]. Engineering Geology, 2022, **306**: 106769.

[9] ZHANG Q, BARRI K, JIAO P, et al. Genetic programming in civil engineering: advent, applications and future trends[J]. The Artificial intelligence review, 2021, **54**(3): 1863-1885.

[10] DAS S K, BASUDHAR P K. Prediction of residual friction angle of clays using artificial neural network[J]. Engineering Geology, 2008, **100**(3/4): 142-145.

[11] PHAM B T, SON L H, HOANG T A, et al. Prediction of shear strength of soft soil using machine learning methods[J]. Catena, 2018, **166**: 181-191.

[12] 刘晓燕, 蔡国军, 邹海峰, 等. 基于 CPTU 数据融合技术的黏性土应力历史与强度特性评价研究[J]. 岩土工程学报, 2019, **41**(7): 1270-1278. (LIU Xiaoyan, CAI Guojun, ZOU Haifeng, et al. Prediction of stress history and strength of cohesive soils based on CPTU and data fusion techniques[J]. Chinese Journal of Geotechnical Engineering, 2019, **41**(7): 1270-1278. (in Chinese))

[13] RABBANI A, SAMUI P, KUMARI S. A novel hybrid model of augmented grey wolf optimizer and artificial neural network for predicting shear strength of soil[J]. Modeling Earth Systems and Environment, 2023, **9**(2): 2327-2347.

[14] 顾春生, 唐 鑫, 朱常坤, 等. 基于主成分分析法优化神经网络的漏湖组黏性土抗剪强度预测[J]. 科学技术与工程, 2023, **23**(28): 11980-11989. (GU Chunsheng, TANG Xin, ZHU Changkun, et al. Prediction of shear strength of cohesive soil in Gehu formation based on back propagation neural network optimized by principal component analysis[J]. Science Technology and Engineering, 2023, **23**(28): 11980-11989. (in Chinese))

[15] 黄鑫中. 不同击实功下非饱和黄土强度特性研究[D]. 西安: 长安大学, 2021. (HUANG Xinzong. Study on Strength Characteristics of Unsaturated Loess Under Different Compaction Work[D]. Xi'an: Changan University, 2021. (in Chinese))

[16] 雷昊楠. 黄土剪切面特征及抗剪强度影响因素试验研究[D]. 西安: 长安大学, 2021. (LEI Haonan. Research for Loess Shear Surface Characteristics and Influencing Factors of Shear Strength[D]. Xi'an: Changan University, 2021. (in Chinese))

[17] 朱志坤. 干密度和含水率对太原重塑黄土强度及强度参数影响的研究[D]. 太原: 太原理工大学, 2018. (ZHU Zhikun. Effect of Dry Density and Water Content on the Strength and Strength Parameters of Remolded Loess in Taiyuan[D]. Taiyuan: Taiyuan University of Technology, 2018. (in Chinese))

[18] 陈梦芸. 击实黏土抗拉强度研究[D]. 南京: 南京大学, 2014. (CHEN Mengyun. Study on Tensile Strength of Compacted Clay[D]. Nanjing: Nanjing University, 2014. (in Chinese))

[19] 杨永亮. 不同击实因素对黄土状粉土击实特性及抗剪强度影响的试验研究[D]. 太原: 太原理工大学, 2019. (YANG Yongliang. Effect of Different Compaction Factors on Compaction Characteristics and Shear Strength of Loess Silt[D]. Taiyuan: Taiyuan University of Technology, 2019. (in Chinese))

[20] DE WINTER J C F, GOSLING S D, POTTER J. Comparing the Pearson and Spearman correlation coefficients across distributions and sample sizes: a tutorial using simulations and empirical data[J]. Psychological Methods, 2016, **21**(3): 273-290.

[21] SHRESTHA D L, SOLOMATINE D P. AdaBoost RT: a boosting algorithm for regression problems[J]. Proceedings of the IEEE International Joint Conference on Neural Networks, 2004, **2**: 1163-1168.

[22] GONG H R, SUN Y R, SHU X, et al. Use of random forests regression for predicting IRI of asphalt pavements[J]. Construction and Building Materials, 2018, **189**: 890-897.

[23] WEI P F, LU Z Z, SONG J W. Variable importance analysis: a comprehensive review[J]. Reliability Engineering & System Safety, 2015, **142**: 399-432.