



西安理工大学学报
Journal of Xi'an University of Technology
ISSN 1006-4710, CN 61-1294/N

《西安理工大学学报》网络首发论文

题目： 基于 XGBoost 的堆场软土渗透系数反演研究
作者： 林玉祥，林浩东，莫品强，褚锋，庄培芝
收稿日期： 2022-03-17
网络首发日期： 2022-11-02
引用格式： 林玉祥，林浩东，莫品强，褚锋，庄培芝. 基于 XGBoost 的堆场软土渗透系数反演研究[J/OL]. 西安理工大学学报.
<https://kns.cnki.net/kcms/detail/61.1294.n.20221102.0853.002.html>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

基于 XGBoost 的堆场软土渗透系数反演研究

林玉祥¹, 林浩东², 莫品强², 褚 锋³, 庄培芝⁴

(1. 中交第三航务工程勘察设计院有限公司, 上海 200032; 2. 中国矿业大学 深部岩土力学与地下工程国家重点实验室 力学与土木工程学院, 江苏 徐州 221116; 3. 山东高速建设管理集团有限公司, 山东 济南 250098; 4. 山东大学 齐鲁交通学院, 山东 济南 250002)

摘要: 软土渗透系数的确定一直是其渗流研究的热点和难点, 针对渗透系数缺乏合理确定手段的现状, 以连云港堆场软土勘探项目为例, 通过对静力触探数据的预处理分析选取特征参数和标签值, 根据渗透系数在水平向和垂直向的差异分别调整超参数, 构建了基于 XGBoost 机器学习算法的渗透系数反演模型。最后利用工程实测资料将 XGB 模型与传统的 BP 神经网络模型和前人经验公式进行对比分析, 结果显示, XGB 模型对渗透系数的预测准确度高于其它方法。

关键词: 静力触探; 渗透系数; 反演模型; XGBoost; BP 神经网络

中图分类号: TV223.6

文献标识码: A

Inversion of permeability coefficient of soft soil in storage yard based on XGBoost

LIN Yuxiang¹, LIN Haodong², MO Pinqiang², CHU Feng³, ZHUANG Peizhi⁴

(1. CCCC Third Navigation Engineering Survey and Design Institute Co. Ltd, Shanghai 200032, China; 2. State Key Laboratory for GeoMechanics and Deep Underground Engineering, School of Mechanics and Civil Engineering, China University of Mining and Technology, Xuzhou 221116, China; 3. Shandong Hi-Speed Group Co., Ltd., Jinan 250098, China; 4. School of Qilu Transportation, Shandong University, Jinan 250002, China)

Abstract: The determination of the coefficient of permeability of soft soil has always been its seepage research hot spot and the difficulty. In view of the present situation of lack of reasonable methods to determine the permeability coefficient, taking Lianyungang storage yard soft soil exploration project as an example. Characteristic parameters and label values are selected by pre-processing analysis of data of cone penetration test, hyperparameters are adjusted according to the difference between horizontal and vertical permeability coefficients, a permeability coefficient inversion model based on XGBoost machine learning algorithm is constructed. Finally, the XGB model is compared with the traditional BP neural network model and the previous empirical formulas by using the engineering measured data. The results show that the XGB model is more accurate than other methods in predicting the permeability coefficient.

Key words: cone penetration test; permeability coefficient; inversion model; XGBoost; BP neural network

随着连云港沿海开发规划政策的落实, 连云港海积软土工程特性的研究逐渐引起业内人士的关注, 其中土体渗透特性的研究为当前热点之一。渗透系数 k 是能定量反映土体渗透性强弱的重要指标, 其数值的正确确定对渗透特性研究至关重要。渗透系数可由土工试验或现场井孔抽注水试验获得, 但

试验过程均繁琐且耗时耗力。因此, 业界学者长期致力于提出一种简单可靠的渗透系数确定方法。

国内外学者对粗粒土的渗透特性研究较多^[1-2], 提出了大量基于天然土的物理性指标、粒径级配曲线等计算渗透系数的经验公式。然而相关公式通常不适用于细粒土, 限制了其在实际工程中的应用。

收稿日期: 2022-03-17

基金项目: 国家自然科学基金青年基金资助项目(51908546); 国家自然科学基金面上基金资助项目(52178374)

第一作者: 林玉祥, 男, 硕士, 工程师, 研究方向为岩土工程勘察。E-mail: 541009687@qq.com

通信作者: 林浩东, 男, 硕士生, 研究方向为静力触探数据解译。E-mail: TS21030056A31LD@cumt.edu.cn

为了得到适用于广泛土类的渗透系数计算方法, 部分学者利用静力触探可以连续反映土层性质的特点, 基于静力触探数据对渗透系数进行预测计算。Robertson^[3]提出渗透系数与土类指数 I_c 的平均关系, 并给出通过静探数据计算 I_c 的公式。Elsworth等^[4-5]和Chai等^[6]分别结合球面流和半球面流模型, 采用位错理论进行推导, 得到了静探测试数据与渗透系数的理论关系式。李镜培等^[7]基于前人研究提出了圆柱面径向渗流模型, 推导出了修正后的经验公式。此外, 有学者在渗透特性研究中引入机器学习算法, 取得了一定成果。许增光等^[8]基于实测电阻率, 采用BP神经网络对土体渗透系数进行预测, 通过计算实例证明了相比经验公式, 神经网络法的平均误差更小。徐丽等^[9]借助极限学习机建立渗透系数与水头之间的映射关系, 并结合遗传算法构建了渗透系数反演分析模型。Zhao等^[10]使用测井数据作为输入, 训练了7种机器学习模型以预测珠江口盆地低渗透砂岩的渗透率, 其中, XGBoost模型表现出了最优的性能。Tian等^[11]通过对3种特征选择方法和6种机器学习模型的分析, 提出了一种将数字岩石物理模型与机器学习模型相结合的方法来改进地下多孔介质的渗透率预测。Pham等^[12]基于越南岷港项目勘察数据构建土体渗透系数预测模型, 并对采用的3种机器学习算法进行了比较分析。前人多致力于对已有经验公式的修正或只简单结合机器学习算法, 而缺少对原始数据的合理利用和预处理研究。且已有经验公式或反演模型鲜有考虑渗透系数在水平向和垂直向的差异, 而方向问题在实际工程应用中往往不可忽略。

本文根据土体静力触探数据与其渗透特性间的关联性, 基于XGBoost机器学习算法构建一种渗透系数反演模型。该模型通过对原始数据的预分析选取特征参数和标签值, 根据工程数据集的特点调整超参数, 且考虑渗透系数在水平向和垂直向的差异。最后根据连云港开展的大荷载矿石堆场深厚软土地基加固研究项目的工程勘察数据, 将本文模型与BPNN模型和前人经验公式进行对比分析。

1 XGBoost 反演模型

1.1 XGBoost 算法

XGBoost即极限梯度提升算法(以下简称XGB)在普通树模型基础上改进算法以提高精度, 充分调用CPU的多线程并行以使提升树达到自身的计算极限, 是近年来性能最强的机器学习算法之一^[13]。

基于梯度增强决策树算法, XGB最小化目标函数至期望范围, 预测值计算公式如下:

$$\hat{y}_i = \varphi(X_i) = \sum_{k=1}^K f_k(X_i), f_k \in F, i \in n \quad (1)$$

式中, \hat{y}_i 为预测值, X_i 为输入变量, i 为均值样本, f_k 代表第 k 个弱评估函数, n 为样本个数, F 为所有分类回归树的集合。

目标函数计算公式如下:

$$L(\varphi) = \sum_{i=1}^n l(\hat{y}_i, y_i) + \sum_{k=1}^K \Omega(f_k) \quad (2)$$

式中, $\sum_{i=1}^n l(\hat{y}_i, y_i)$ 为预测值与实测值的偏差;

$\sum_{k=1}^K \Omega(f_k)$ 为限制算法复杂度的正则化项。

最小化目标函数, 经过正则化项对算法学习权重的平滑, 最终得到目标函数的最优解如下所示:

$$\tilde{L}^{(t)}(q) = -\frac{1}{2} \sum_{j=1}^T \frac{(\sum_{i \in q_j} g_i)^2}{\sum_{i \in q_j} h_i + \lambda} + \gamma T \quad (3)$$

式中, T 为剩余节点数, λ 和 γ 为正则化系数, g_i 和 h_i 分别是损失函数 $l(\hat{y}_i^{(t-1)}, y_i^t)$ 对 $\hat{y}_i^{(t-1)}$ 的一阶导数和二阶导数, t 为迭代次数。

XGB允许使用者根据实际使用场景调整损失函数, 对于渗透系数预测这样的回归问题而言, 采用最小二乘损失函数, 公式如下:

$$Loss = \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (4)$$

XGB算法本质是基于梯度提升树实现的集成算法, 涉及参数极多。考虑到使用的数据集较为简单, 故仅对属于集成算法和弱评估器的部分重要参数进行调参(部分重要参数如表1所示), 将其它与提升模型性能无关的复杂参数设为默认值。

表1 XGBoost算法重要参数表(部分)

Table 1 Important parameters of XGBoost algorithm(part)

参数	含义	默认值
num_round	集成中弱评估器的数量	10
eta	集成中的学习率	0.3
subsample	随机抽样抽取的样本比例	1
max_depth	树的最大深度	6
colsample_bytree	生成树时随机抽样特征的比例	1
colsample_bylevel	生成层时随机抽样特征的比例	1
colsample_bynode	生成节点随机抽样特征的比例	1

1.2 考虑 k_v 和 k_h 的差异

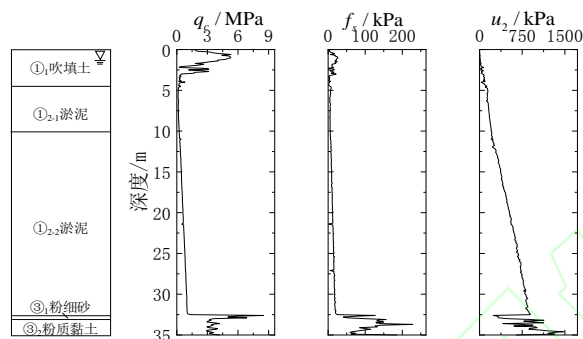
在对渗透系数水平向和垂直向的差异研究上, Jamiolkowski等^[14]并利用径向渗透仪装置将两种标

准的垂直测量值(k_v)与水平值(k_h)进行了比较, 提出了软粘土 k_h/k_v 的可能值范围: $k_h/k_v = 1\sim 1.5$ (均质沉积物), $2\sim 4$ (发达宏观结构), $3\sim 15$ (明显分层)。以往的渗透系数经验公式往往不考虑渗透系数在方向上的差异, 本模型将在前人研究基础上, 分别对 k_v 和 k_h 进行预测, 使预测结果更符合实际工程需要。

2 工程应用

2.1 工程概况

场地位于连云港港旗台作业区南部, 工程范围内已由 30 万吨航道二期疏浚土吹填成陆, 土体地下深度约 4m 内均为吹填土。孔压静力触探试验采用荷兰 GEOMIL 静力触探试验设备, 静力触探试验曲线图如图 1 所示(以 JT33 静探孔为例)。



(a) 土层划分图 (b) 静力触探曲线图

图 1 Geomil 静力触探试验成果图(锥尖、侧壁、孔压)

Fig.1 Geomil cone penetration test results (cone tip, side wall, pore pressure)

工程布置有静力触探试验孔 65 个, 取土试样钻孔 15 个, 且每个钻孔都与一静探孔相邻。根据取土钻孔的位置及取土土样的深度匹配其对应的静探孔的静探数据, 以建立 CPTU 数据与室内试验结果的严格对应关系。室内土工试验共设置了 11 组垂直向渗透系数(k_v)和 8 组水平向渗透系数(k_h)测定试验, 以“QT 钻孔号-土样编号”的形式命名土样, 试验结果即原始数据集如表 2 所示。

表 2 土工试验渗透系数数据集

Tab.2 Permeability coefficient data set of geotechnical test

钻孔及土样 编号	取土深度 /m	土类名称	土工试验渗透系数	
			k_v /cm/s	k_h /cm/s
QT12-08	12.5-13.0	淤泥	1.2E-7	1.7E-7
QT12-10	19.5-20.0	淤泥	6.7E-8	1.2E-7
QT12-12	26.5-27.0	淤泥	4.8E-8	7.5E-8
QT15-08	11.0-11.5	淤泥	1.5E-7	1.6E-7
QT15-10	18.0-18.5	淤泥	7.0E-8	1.0E-7
QT15-12	23.0-23.5	淤泥	7.0E-8	1.2E-7

QT15-14	34.1-34.3	粉质粘土	4.8E-8	1.3E-7
QT06-101	1.6-1.8	细砂	1.2E-3	1.2E-3
QT06-102	3.1-3.3	粉砂	1.1E-3	
QT14-01	1.0-1.5	粉砂	1.2E-3	
QT14-02	1.5-2.0	粉砂	9.2E-4	

本工程钻孔取得土试样等级为 I~II 级, 即扰动程度为不扰动~轻微扰动。土样的采取、蜡封、储存、运输及室内试验严格按照有关规定执行, 因此认为室内土工试验结果较为可靠。

2.2 XGBoost 模型的训练与预测

2.2.1 输入变量和标签值的选取

由于静探孔的探测精度为 0.1m, 即随深度每 0.1m 可得一组 CPTU 数据, 为将其与渗透系数测定值对应, 假定土样的取土深度范围内每 0.1m 的土层渗透系数都等于该段土样的渗透系数值。对 CPTU 数据的异常值进行剔除后, 最终总数据集共有 47 组 CPTU- k_v 数据和 41 组 CPTU- k_h 数据。模型的输入变量从 CPTU 测试指标中选取, 根据经验, 考虑使用静探数据转换后得到的归一化锥尖阻力 Q_{tn} 和归一化摩擦比 F_r 的组合, 或直接使用土类指数 I_c 作为输入变量对标签值进行预测, 相关计算公式^[15]如下:

$$I_c = [(3.47 - \lg Q_{tn})^2 + (\lg F_r + 1.22)^2]^{0.5} \quad (5)$$

$$Q_{tn} = ((q_t - \sigma_{v0})/p_a) \times (p_a/(\sigma'_{v0}))^n \quad (6)$$

$$F_r = f_s/(q_t - \sigma_{v0}) \times 100\% \quad (7)$$

式中, $n = 0.381 \times I_{c(RW)} + 0.05 \times (\sigma'_{v0}/p_a) - 0.15$; σ_{v0} 、 σ'_{v0} 分别为总上覆土压力和上覆土有效应力; q_t 为孔压修正锥尖阻力, $q_t = q_c + (1 - a)u_2$, a 为探头面积比, 取 0.8; p_a 为参考压力, $p_a = 100\text{kPa} = 0.1\text{MPa}$ 。

对于标签值的选取, 考虑到表 2 中渗透系数的数量级均为 10^{-2} 以下, 故对渗透系数取对数后作为标签值, 并以原数值设置对照组进行对比。

本文算法均通过 Python 的 xgboost 模块实现。采用默认超参数值建立预分析模型, 对两类输入变量和两类标签值(以 k_v 为例)形成的 4 种预测组合分别进行训练预测。以 R^2 为评价指标, 取模型运行 5 次结果的均值, 结果如表 3 所示。

表 3 预分析结果表

Tab.3 Results of pre-analysis

		标签值	
		k_v	$\lg(k_v)$
输入变量	R^2		
	$Q_{tn} + F_r$	0.768	0.914
	I_c	-0.159	0.002

从表3中可见,以 $Q_{tn} + F_r$ 作为输入变量, $\lg(k_v)$ 作为标签值时有着最高的 R^2 。可见 $Q_{tn} + F_r$ 比 I_c 单值能更全面地反应土体的性质,更适于模型学习;渗透系数原数值过小且接近,作为标签值不便于模型分辨,在取对数后模型效果有较大提升。所以最后选取 $Q_{tn} + F_r$ 作为输入变量,渗透系数的对数值作为标签值建立预测模型。

2.2.2 模型的训练

1) k_v 预测模型

本模型采用先通过学习曲线调整弱评估器数量,再通过网格搜索调整其余超参数的调参策略。将总数据集按7:3的比例随机划分为训练集和测试集,为保证严谨性,只对训练集而非全集使用K折交叉验证来估计模型的泛化误差。考虑到样本数据量小,取 $K = 5$ 。作出模型表现随弱评估器数量增加而变化的学习曲线图(以 R^2 为评价指标),如图2所示。

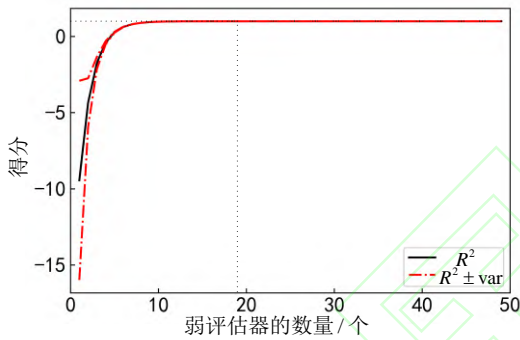


图2 训练集学习曲线图

Fig.2 Learning curve of training set

图2中, R^2 、方差、总泛化误差分别在弱评估器数量为29、14、19时达到最佳取值,故最后取 $\text{num_round} = 19$ 。在经过网格搜索对其余参数进行调整后,确定了一组最佳的超参数组合如表4所示。

表4 k_v 预测模型最佳超参数

Tab.4 Optimal hyperparameters of k_v prediction model

超参数	最佳值
num_round	19
eta	0.1
subsample	1
max_depth	7
colsample_bytree	0.8
colsample_bylevel	0.9
colsample_bynode	0.9

设置了该组超参数、经过训练集训练的模型对测试集的预测表现如图3所示:

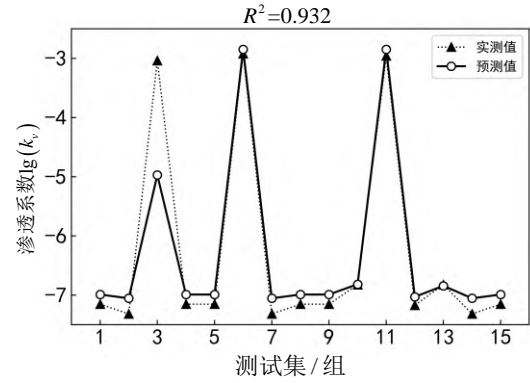


图3 k_v 预测模型结果可视化对比

Fig.3 Visual comparison of results of k_v prediction model

图3中,该模型对测试集的预测值与实测值的决定系数 R^2 达到0.932,意味着模型的输入变量即归一化锥尖阻力 Q_{tn} 和归一化摩阻比 F_r 对渗透系数的对数标签值有较强的解释性。注意到模型对 $k = 10^{-3}\text{cm/s}$ 左右的砂土和 $k = 10^{-7}\text{cm/s}$ 左右的软土均有不错的预测表现,说明模型对砂土和软土都能适用,泛化性较强。

2) k_h 预测模型

同上节,在经过学习曲线和网格调参后,确定了一组适用于 k_h 预测的最佳超参数如表5所示。

表5 k_h 预测模型最佳超参数

Tab.5 Optimal hyperparameters of k_h prediction model

超参数	最佳值
num_round	21
eta	0.2
subsample	1
max_depth	5
colsample_bytree	0.9
colsample_bylevel	0.7
colsample_bynode	1

该模型在测试集表现如图4所示:

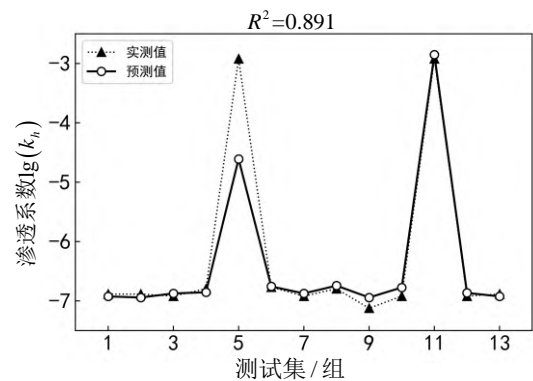


图4 k_h 预测模型结果可视化对比

Fig.4 Visual comparison of results of k_h prediction model

图4中,由于砂土水平向渗透系数数据较少,

故模型对砂土的渗透系数预测效果略有下降, 对软土预测效果仍较好。总体实测值与预测值的 R^2 为0.891, 仍为较高水平。

2.2.3 模型预测效果

将上述 k_v 和 k_h 预测模型应用于项目 CPTU 数据集, 将土工试验结果的实测值与模型预测值直接进行对比, 结果可视化对比如图 5 所示。

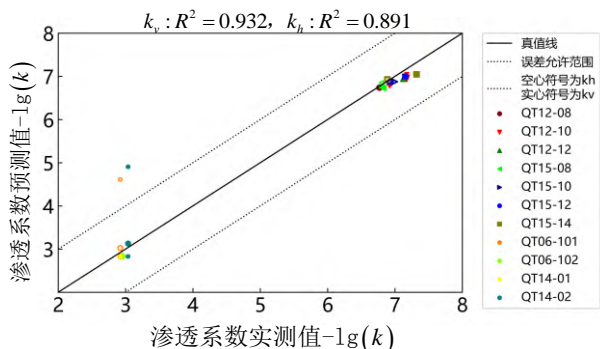


图 5 XGB 模型预测值-实测值对比图

Fig.5 Comparison of predicted and measured values of XGB model

图 5 中, 考虑到工程实践里渗透系数的计算可接受一个数量级范围内的误差^[16], 故加入距实测值一个数量级的两误差线, 该范围内的预测值认为都达到精度要求。为每个取土孔及土样编号指定特定符号, 将该土样深度范围内每 0.1m 对应的渗透系数预测值显示在对比图中, 该深度范围内渗透系数的平均值以较大的符号显示。从图 5 可见, 除了在 $-\lg(k) = 3$ 附近的两个数据在误差范围外, 其余数据的预测值均与实测值较为接近。两个不准数据来源于 QT06-101、QT14-02 土样, 均为砂土, 这可能与土工试验结果中砂土数据较少有关, 但该土样深度范围内渗透系数预测值的平均值也与实测值相当接近。在测试集上, k_v 和 k_h 的实测值与预测值取对数后的 R^2 分别为 0.932 和 0.891, 均达较高水平。

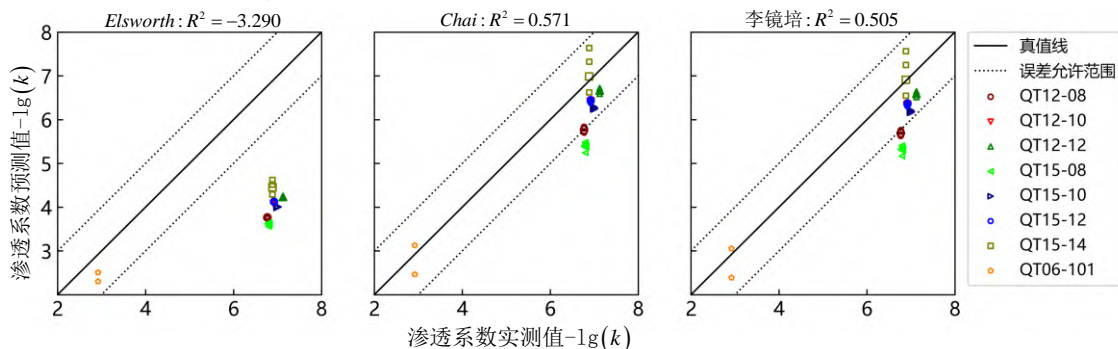


图 7 预测值-实测值对比图

Fig.7 Comparison of predicted and measured values

图 7 中, 三经验公式的计算值与实测值取对数后的 R^2 分别为-3.290、0.571、0.505, 可见是较为经

3 各预测方法结果对比

3.1 Robertson 经验公式

Robertson^[3]提出了一种土壤渗透系数 k 与土类指数 I_c 的平均关系式, 实测值在该经验公式图中分布情况如图 6 所示。

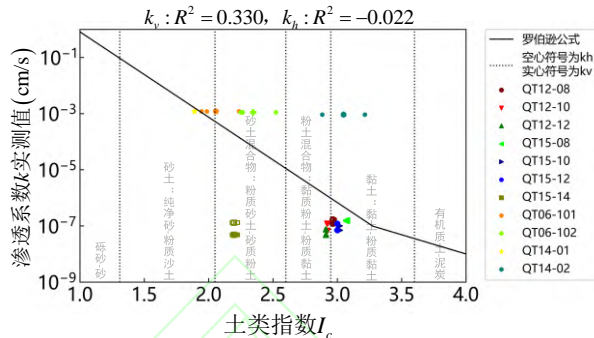


图 6 实测值 $k - I_c$ 分布图

Fig.6 $k - I_c$ distribution of measured values

图 6 中, 由于数据量较小, 实测值与公式曲线离散度较大。把 Robertson 公式计算值分别与实测值的垂直向和水平向渗透系数取对数后比较, 垂直向的 R^2 为0.330, 水平向的 R^2 为-0.022 (R^2 为负意味着预测值的预测效果还不如实测值的均值), 总体 R^2 为 0.282。注意到垂直向的 R^2 远高于水平向的, 说明罗伯逊公式可能更适合预测 k_v 而非 k_h 。

3.2 (半) 球面流模型经验公式

Elsworth 等^[4-5]、Chai 等^[6]、李镜培等^[7]分别结合球面流、半球面流、圆柱面径向渗流模型推导出了 CPTU 测试指标与水平向渗透系数之间的理论关系式, 其适用于正常或轻微超固结的黏性土和松散的无黏性土。将 CPTU 测试指标和探头参数代入上述各经验公式, 将计算值与土工试验结果的水平向渗透系数实测值对比, 结果可视化如图 7 所示。

典的 Chai 公式拥有最好的表现。

3.3 BP 神经网络模型

BP 神经网络虽然性能上比 XGBoost 和随机森林等树模型稍逊，但优势在于结构简单，能够提取出较为简洁的公式以供使用。训练过程不再赘述。

将 k_v 和 k_h 预测模型的权值和阈值参数在三层 BP 神经网络中反算出来，代入提取公式可得：

$$\lg(k_v) = 0.34 \cdot \text{tansig}(2.46 \cdot F_r - 1.16 \cdot Q_{tn} + 2.08) - 1.31 \cdot \text{tansig}(2.28 \cdot F_r - 0.24 \cdot Q_{tn} + 0.17) - 0.28 \cdot \text{tansig}(-2.24 \cdot F_r - 0.44 \cdot Q_{tn} - 2.62) - 0.07 \tag{8}$$

$$\lg(k_h) = -0.79 \cdot \text{tansig}(-0.29 \cdot F_r + 2.67 \cdot Q_{tn} + 1.99) - 1.66 \cdot \text{tansig}(2.47 \cdot F_r - 1.62 \cdot Q_{tn} + 0.98) - 0.08 \cdot \text{tansig}(-0.65 \cdot F_r + 2.16 \cdot Q_{tn} - 2.67) + 0.15 \tag{9}$$

将 CPTU 数据代入上述公式计算，并将结果反归一化后可得渗透系数的 BPNN 公式预测值。将实

测值与预测值直接进行对比，结果如图 8 所示。

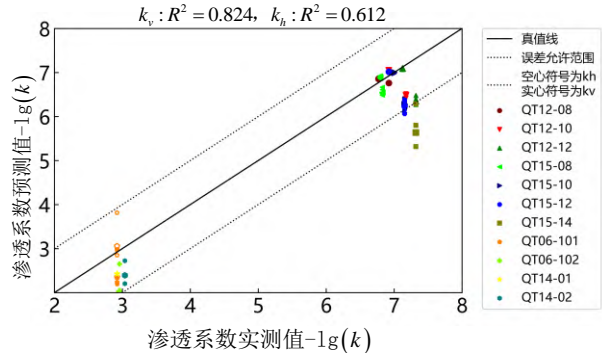


图 8 BPNN 模型预测值-实测值对比图

Fig.8 Comparison of predicted and measured values of BPNN

model

总体来说，BPNN 预测模型对 k_v 预测效果较好， R^2 为 0.824，对 k_h 预测效果较差， R^2 为 0.612。

3.4 各预测方法结果对比

将基于 CPTU 数据的各预测方法所得预测值与土工试验实测值进行对比，如表 6 所示。

表 6 各预测方法预测值与实测值对比表

Tab.6 Comparison of predicted and measured values of each prediction method

实测值		Robertson	Elsworth	Chai	李镜培	BPNN 预测模型		XGB 预测模型	
k_v	k_h	k	k_h	k_h	k_h	k_v	k_h	k_v	k_h
/cm/s	/cm/s	/cm/s	/cm/s	/cm/s	/cm/s	/cm/s	/cm/s	/cm/s	/cm/s
1.2E-7	1.7E-7	8.6E-07	1.7E-04	1.7E-06	2.0E-06	1.7E-07	1.4E-07	1.5E-07	1.8E-07
6.7E-8	1.2E-7	1.2E-06	8.4E-05	3.9E-07	4.7E-07	3.3E-07	8.9E-08	1.1E-07	1.5E-07
4.8E-8	7.5E-8	1.3E-06	5.8E-05	2.2E-07	2.6E-07	4.5E-07	8.1E-08	9.0E-08	1.1E-07
1.5E-7	1.6E-7	4.1E-07	2.4E-04	4.0E-06	4.8E-06	2.9E-07	1.3E-07	1.9E-07	1.7E-07
7.0E-8	1.0E-7	6.0E-07	9.9E-05	5.5E-07	6.6E-07	5.3E-07	1.0E-07	1.0E-07	1.3E-07
7.0E-8	1.2E-7	6.8E-07	7.5E-05	3.7E-07	4.4E-07	5.9E-07	9.7E-08	1.0E-07	1.3E-07
4.8E-8	1.3E-7	1.9E-04	3.5E-05	1.0E-07	1.2E-07	2.3E-06	6.2E-10	9.0E-08	1.2E-07
1.2E-3	1.2E-3	6.8E-04	3.1E-02	2.6E-02	3.1E-02	4.4E-03	8.8E-04	1.5E-03	9.5E-04
1.1E-3		8.8E-05				9.4E-03		1.5E-03	
1.2E-3		1.6E-03				3.7E-03		1.5E-03	
9.2E-4		8.4E-07				4.0E-03		7.4E-04	

表 6 中，取土孔取得土样的淤泥、粉质粘土、细砂、粉砂作为均质沉积物，渗透系数实测值都基本符合 $k_h/k_v = 1 \sim 1.5$ 。XGB 模型的预测值也符合该相关关系，但 BPNN 模型的 k_h 预测值普遍低于 k_v 预测值，所以选择 XGB 模型用于预测渗透系数更合理。

将各预测方法的计算值与实测值取对数值后的距离使用 R^2 、均方误差 MSE、均方根误差 RMSE、平均绝对误差 MAE、平均绝对百分比误差 MAPE 等 5 个常用评估指标进行评价，以 R^2 为主要评价标准降序排序如表 7 所示。其中对于机器学习预测模型，则根据测试集而非全集的表现进行评估。

表 7 各预测方法计算值与实测值距离评估指标

Tab.7 Evaluation indexes of distance between calculated and measured values of each prediction method

预测方法	R^2	MSE	RMSE	MAE	MAPE
XGB (k_v)	0.932	0.269	0.519	0.296	7.165
XGB (k_h)	0.891	0.225	0.475	0.192	5.442
BPNN (k_v)	0.824	0.711	0.843	0.746	15.098
BPNN (k_h)	0.612	0.681	0.825	0.361	5.572
Chai (k_h)	0.571	0.754	0.868	0.756	14.268
李镜培 (k_h)	0.505	0.870	0.933	0.812	15.273
Robertson (k_v)	0.327	2.714	1.647	1.273	25.609
Robertson (k_h)	-0.028	1.806	1.344	1.037	15.582
Elsworth (k_h)	-3.290	7.539	2.746	2.695	42.482

从表 7 中可见, 基于 XGB 的 k_v 预测模型表现最好, 基于 XGB 的 k_h 预测模型其次。基于 BP 神经网络的提取公式的预测表现虽然不如 XGB 模型, 但也高于其它经验公式。前人经验公式中, Chai 公式有着最好的预测表现。由于本文所用数据集的数据量有限, 两个基于机器学习方法的预测模型不一定能在其他地区的土壤渗透系数预测中也能达到如此表现, 但对于连云港地区类似的吹填土区域的渗透特性研究应当有一定的参考价值。

4 结论

1) 本文根据土体的 CPTU 数据与其渗透系数之间存在的关联性, 提出了一种基于 XGBoost 机器学习算法的渗透系数反演模型, 该模型适用于广泛土类, 且考虑了水平向和垂直向渗透系数的差异。

2) 通过对输入变量和标签值的合理选取, 以及对最优超参数组合的确定, 模型最终达到了较高的预测精度。该 XGBoost 预测模型的 k_v 预测值与实测值 R^2 可达 0.932, k_h 预测值与实测值 R^2 可达 0.891, 远高于传统的 BPNN 模型和前人经验公式。

3) 通过本文的工程实例表明, 基于机器学习算法构建的预测模型合理可行。XGBoost 模型学习速度快、预测精度高、需调参数少, 适合复杂原理相关的土体参数的反演, 对于渗透特性研究及类似研究有一定参考价值。

参考文献:

[1] 包孟碟, 朱俊高, 吴二鲁, 等. 基于级配方程的粗粒土渗透系数经验公式及其验证[J]. 岩土工程学报, 2020, 42(8): 1572-1576.
BAO Mengdie, ZHU Jungao, WU Erlu, et al. Empirical

formula of permeability coefficient of coarse-grained soil based on gradation equation and its verification[J]. Chinese Journal of Geotechnical Engineering, 2020, 42(8): 1572-1576.

- [2] 赵茜, 苏立君, 刘华, 等. 黄土粒度组成对其渗透系数各向异性影响的试验研究[J]. 中南大学学报(自然科学版), 2020, 51(6): 1615-1626.
ZHAO Qian, SU Lijun, LIU Hua, et al. Experimental study on the effect of grain size composition on anisotropy of loess permeability coefficient[J]. Journal of Central South University(Science and Technology), 2020, 51(6): 1615-1626.
- [3] ROBERTSON P K. Estimating in-situ soil permeability from CPT & CPTU[C]// Proceedings of the 2nd International Symposium on Cone Penetration Testing (CPT'10). Huntington Beach, California: California State Polytechnic University Pomona, 2010: 535-542.
- [4] ELSWORTH D, LEE D S. Permeability determination from on-the fly piezocone sounding[J]. Journal of Geotechnical and Geoenvironmental Engineering, 2005, 131(5): 643-653.
- [5] ELSWORTH D, LEE D S. Limits in determining permeability from on-the-fly CPTu sounding[J]. Géotechnique, 2007, 57(8): 679-686.
- [6] CHAI J C, AGUNG P M A, HINO T, et al. Estimating hydraulic conductivity from piezocone soundings[J]. Géotechnique, 2011, 61(8): 1-10.
- [7] 李镜培, 李险峰, 张亚国. CPTU 确定饱和土体水平渗透系数的改进方法[J]. 哈尔滨工业大学学报, 2016, 48(8): 185-188.
LI Jingpei, LI Xianfeng, ZHANG Yaguo. An improved method for determining horizontal permeability coefficient of saturated soil by CPTU[J]. Journal of Harbin Institute of Technology, 2016, 48(8): 185-188.
- [8] 许增光, 王雪, 柴军瑞, 等. 基于电阻率求解渗透系数不同方法的对比研究[J]. 西安理工大学学报, 2018, 34(1): 43-47.
XU Zengguang, WANG Xue, CHAI Junrui, et al. Comparison of different methods for calculating permeability coefficient based on resistivity[J]. Journal of Xi'an University of Technology, 2018, 34(1): 43-47.
- [9] 徐丽, 沈振中. 基于 ELM-GA 的复杂土石坝渗透系数反演模型及其应用[J]. 水电能源科学, 2021, 39(9): 86-90.

- XU Li, SHEN Zhenzhong. Inversion model of permeability coefficient of complex Earth-rock dam based on ELM-GA and its application[J]. *Water and Power Energy Science*, 2021, 39(9): 86-90.
- [10] ZHAO Xiaobo, CHEN Xiaojun, HUANG Qiao, et al. Logging-data-driven permeability prediction in low-permeable sandstones based on machine learning with pattern visualization: A case study in Wenchang A Sag, Pearl River Mouth Basin[J]. *Journal of Petroleum Science and Engineering*, 2022, 214: 110517.
- [11] TIAN J W, QI Chongchong, PENG Kang, et al. Improved Permeability Prediction of Porous Media by Feature Selection and Machine Learning Methods Comparison[J]. *Journal of Computing in Civil Engineering*, 2022, 36(2): 04021040.
- [12] PHAM B T, NGUYEN M D, AL-ANSARI N, et al. A Comparative Study of Soft Computing Models for Prediction of Permeability Coefficient of Soil[J]. *Mathematical Problems in Engineering*, 2021, 2021: 7631493.
- [13] CHEN T, GUESTRIN C. XGBoost: A Scalable Tree Boosting System[A]. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*[C]. San Francisco, 2016: 785-794.
- [14] BEAR J. *Dynamics of fluid in porous media*[M]. New York: Elsevier, 1972.
- [15] ROBERTSON P K, WRIDE C E. Evaluating cyclic liquefaction potential using the cone penetration test[J]. *Canadian Geotechnical Journal*, 1998, 35(3): 442-459.
- [16] 邹海峰, 蔡国军, 刘松玉. 基于位错理论的饱和土渗流特性 CPTU 评价研究[J]. *岩土工程学报*, 2014, 36(3): 519-528.
- ZOU Haifeng, CAI Guojun, LIU Songyu. Evaluation of coefficient of permeability of saturated soils based on CPTU dislocation theory[J]. *Chinese Journal of Geotechnical Engineering*, 2014, 36(3): 519-528.