

考虑特征选择的土石坝溃口峰值流量预测模型

张美满^{1,2}, 李晶³, 张友明⁴, 杨旭^{1,2}, 雷天宇⁴, 陈瀚^{1,2}, 徐津^{1,2}, 王玲玲^{1,2}

(1. 河海大学流域水循环与水安全全国重点实验室; 2. 河海大学水利部水循环与水动力系统重点实验室;
3. 淮河水利委员会治淮工程建设管理局; 4. 江苏省防汛防旱抢险中心(江苏省防汛抢险训练中心))

摘要: 针对溃口特征维度高且特征间高度相关性导致预测模型性能下降的问题, 提出了一种融合 Lasso 算法和 XGBoost 模型的土石坝溃口峰值流量预测模型。该模型采用斯皮尔曼相关系数法分析溃口特征间的相关性, 使用 Lasso 算法进行进一步的特征选择, 并通过剔除冗余特征得到最优特征子集, 再将该特征子集输入 XGBoost 模型进行溃口峰值流量预测。与支持向量回归和岭回归机器学习模型对比结果表明, 该模型具有良好的非线性信息挖掘能力, 可对高维特征进行有效降维, 在减少模型复杂性的同时提高了模型预测精度。

关键词: 土石坝溃口; 峰值流量预测; 特征选择; 多重共线性; Lasso 算法; XGBoost 模型

Prediction model for peak discharge of earth-rock dam breaches based on feature selection//Zhang Meiman^{1,2}, Li Jing³, Zhang Youming⁴, Yang Xu^{1,2}, Lei Tianyu⁴, Chen Han^{1,2}, Xu Jin^{1,2}, Wang Lingling^{1,2} (1. State Key Laboratory of Water Cycle and Water Security, Hohai University; 2. Key Laboratory of Hydrologic-Cycle and Hydrodynamic-System of Ministry of Water Resources, Hohai University; 3. Project Construction Management Bureau of the Huaihe River Water Resources Commission; 4. Jiangsu Provincial Flood Control and Drought Relief Center (Jiangsu Provincial Flood Control and Emergency Rescue Training Center))

Abstract: To address the issue of degraded predictive model performance caused by high-dimensional breach features and strong correlations among features, a peak discharge prediction model for earth-rock dam breaches was developed by integrating the Lasso algorithm with the XGBoost model. This model uses the Spearman correlation coefficient method to analyze the correlations between features, utilizes the Lasso algorithm for further feature selection, and obtains an optimal feature subset by eliminating redundant features. The feature subset is input into the XGBoost model to predict the peak breach discharge. Comparative results with support vector regression and ridge regression machine learning models show that the proposed model exhibits strong nonlinear information mining capability, effectively reduces the dimensionality of high-dimensional features, and improves the prediction accuracy while reducing model complexity.

Key words: earth-rock dam breach; peak discharge prediction; feature selection; multicollinearity; Lasso algorithm; XGBoost model

水库大坝和圩垸堤防作为我国国民经济的重要基础设施,在防洪减灾、城乡供水、水力发电、农田灌溉等方面发挥了巨大作用。然而,在长期服役过程中,土石类堤坝易发生开裂、滑坡、浸散、管涌、流土和冲刷等险情,一旦溃决将导致洪水泛滥,严重威胁人民生命财产安全,影响工农业生产和社会生活^[1]。堤坝溃决突发性强,溃口发展态势难以预料,极大地增加了抢险救灾的难度。溃口峰值流量的快速预测对于抢险救灾和下游洪灾评估具有重要意义,也一直是水利工程领域的重要研究热点。

国内外众多学者针对溃口峰值流量预测模型开展了大量研究。Wahl 等^[2-4]基于收集的溃坝案例,利用多元非线性回归建立溃口峰值流量预测公式; Sammen 等^[5]采用 140 余组案例分析了现有预测公式的不确定性,进而提出采用机器学习方法取代传统的回归分析方法。人工神经网络(ANN)是目前水利工程领域应用最广泛的机器学习方法之一^[6]。Nourani 等^[7]构建了 2 个 ANN 模型模拟土坝溃口流量过程线,但由于缺乏足够的实测资料训练模型,模型精度不高。支持向量机(SVM)是另一种机器学

基金项目: 国家重点研发计划项目(2022YFC3202600);江苏省水利科技项目(2023013,2024009);国家自然科学基金项目(52479062,52309086);亚洲开发银行贷款项目(3704-PRC)

作者简介: 张美满(1996—),男,博士研究生,主要从事水力学及河流动力学研究。E-mail: zhangmeiman@hhu.edu.cn

通信作者: 王玲玲(1966—),女,教授,博士,主要从事计算水力学研究。E-mail: wanglingling@hhu.edu.cn

习算法,适用于小样本训练数据集。Nobarinia 等^[8]的研究表明 SVM 的预测性能优于传统回归模型。虽然 SVM 已成功应用于多个领域,但其对超参数选择较为敏感^[9-10]。近年来,为弥补上述算法的不足,一系列新型机器学习算法被相继提出。其中,极限梯度提升(XGBoost)算法作为一种高效的集成学习算法,凭借其独特的优势获得了广泛认可^[11]。该算法在损失函数的计算中增加了二阶导数与正则项,并在一定程度上实现了并行计算,可以更精确地拟合非线性关系^[12-14]。

在机器学习领域,“特征”指描述样本的属性或变量,特征选择是训练和评估机器学习模型的基础,合理的特征选择能够显著提升模型的性能和准确度。本文将机器学习中的“特征”对应于水利工程中的“溃坝关键影响参数”。已有研究^[15-18]表明,溃口峰值流量受坝高、溃口底部以上水深、水库库容及溃口宽度等多项参数(即特征)共同影响。当这些特征之间存在线性相关性时,会引发多重共线性问题,导致模型精度下降甚至参数估计失真;同时过多特征的输入也会加剧模型的复杂度,增加预测难度^[19]。改善多重共线性的常用方法有主成分分析(PCA)、线性判定分析(LDA)等。尽管这些方法具有较好的降维效果,但在降维过程中会丢失部分细节信息,从而可能影响后续分析的准确性^[19]。鉴于此,本文提出一种融合 Lasso 特征选择与 XGBoost 模型的溃口峰值流量预测模型(Lasso-XGBoost 模型),该模型先利用 Lasso 算法的 L1 正则化特性剔除冗余特征,筛选出最优特征子集,有效降低数据维度;随后将所选特征输入 XGBoost 模型进行溃口峰值流量预测。通过构建 Lasso-XGBoost 耦合模型,旨在解

决高维特征及多重共线性导致的模型复杂和精度下降问题,以期为抗洪抢险、减灾防灾提供技术支持。

1 数据库及特征描述

1.1 数据库建立

已有研究^[16]表明,溃坝水流首先在下游坝坡形成初始冲坑,并逐步向上游溯源冲刷;当冲刷发展到坝顶上游边缘时,溃决过程变得十分迅速剧烈,伴随着溃口边坡的间歇性失稳坍塌,最终形成相对稳定的溃口。典型的溃口横断面、纵断面及相关参数如图 1 所示。

本文收集整理了国内外土石类堤坝溃决数据和水利部大坝安全管理中心的相关资料^[20],构建了一个包含 192 个溃坝案例的基础数据库。为了充分考虑上游水库影响和溃口形状特征,选取 7 个关键溃坝特征作为研究对象,其物理含义、详细描述及在数据库中的完整资料占比见表 1。

1.2 数据预处理

在应用机器学习算法前,必须对溃坝数据库的原始数据进行预处理,包括数据清洗、转换和处理,以提高模型的准确性和泛化能力。由表 1 可见,7 个特征中溃口宽度缺失最为严重,缺失率高达 46.35%,其次是水库库容,缺失率为 23.40%。大量数据缺失会影响模型训练的准确性,因此需要对原始数据库进行数据填充。数据填充方法主要包括固定值填充、平均值填充、众数填充、相邻值填充及插值法填充等方法^[21]。相较于直接剔除缺失数据或其他复杂的插补方法,平均值填充法能够在最大程度保留样本完整性的同时,保持特征的统计均值不变,避免了因样本量减少而导致的信息损失,满足了

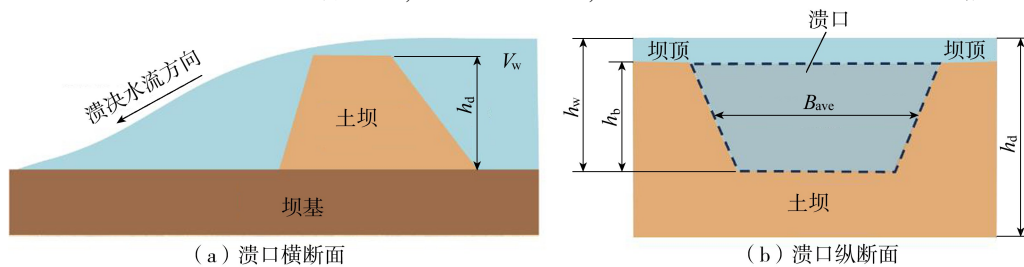


图 1 土坝溃口形状及其表征参数

表 1 溃坝数据库主要特征描述

特征名称	符号	特征描述	单位	数据完整率/%
坝高	h_d	坝顶的最低点与原始河床中的最低点之间的垂直高差	m	87.50
溃口底部以上水深	h_w	水位与最终溃口最低点的高差	m	100.00
溃口最终深度	h_b	溃口处坝顶和溃口最低点的高差	m	99.95
水库库容	V	正常蓄水位对应的水量	m^3	76.60
溃口底部以上库容	V_w	最终溃口底部高程以上的水量	m^3	99.95
峰值流量	Q_p	溃口处流量达到的最大值	m^3/s	92.71
溃口宽度	B_{ave}	最终溃口顶部和底部宽度的平均值	m	53.65

机器学习模型训练对数据完整性的要求。因此,本文采用平均值填充法补充缺失值,建立一个包含192个溃坝案例的完整数据库,用于训练机器学习模型以预测溃口峰值流量。

2 模型构建

首先使用斯皮尔曼相关系数法和 Lasso 算法依次对特征进行筛选,得到最优特征子集,然后将其输入 XGBoost 模型进行溃口峰值流量预测。

2.1 基于斯皮尔曼相关性分析的特征选择

斯皮尔曼相关系数法是一种衡量两个特征之间单调关系的非参数统计方法,适合于非线性单调关系的数据分析,尤其当数据存在异常值时,采用斯皮尔曼相关系数法可以降低异常值的影响。因此,本文采用斯皮尔曼相关系数法对除峰值流量外的6个特征与峰值流量之间的相关性进行评估,并据此进行初步筛选。斯皮尔曼相关系数的计算步骤如下:

步骤1 确定秩次。对两个特征 X 和 Y 的所有数据分别按大小进行排序,确定每个数据在各自序列中的秩次,得到两个特征的秩序列 $R(X)$ 和 $R(Y)$ 。

步骤2 计算秩次差 d_i 。对于样本点 i ,计算两个特征之间的秩次差 $d_i = R(X_i) - R(Y_i)$,其中 $R(X_i)$ 、 $R(Y_i)$ 分别为特征 X 、 Y 中第 i 个数据的秩次。

步骤3 计算相关系数。采用以下公式计算斯皮尔曼相关系数 r_s :

$$r_s = 1 - \frac{6 \sum_{i=1}^N d_i^2}{N(N^2 - 1)} \quad (1)$$

式中: N 为样本数据点的数量。当 r_s 接近 1 时,表示特征之间呈强正相关关系;当 r_s 接近 -1 时,表示呈强负相关关系;当 r_s 接近 0 时,表示相关性极弱。

2.2 基于 Lasso 算法的进一步特征选择

由于斯皮尔曼相关系数法忽略了特征之间的相互作用,因此在实际应用中,需要结合其他方法以获得更准确的特征选择结果^[22]。Lasso 算法的基本原理是在最小二乘法的基础上,引入一个带权重 λ 的惩罚项来约束机器学习模型中各特征的回归系数,以达到特征选择的目的^[23]。通过增大 λ 值,与目标变量不显著相关的特征的回归系数会逐渐被压缩为 0 并被剔除,从而最终获得一个特征数量较少且效率更高的模型。回归系数计算公式为:

$$J(\mathbf{w}) = \sum_{i=1}^N \left(y_i - \sum_{j=1}^n w_j x_{ij} \right)^2 + \lambda \sum_{j=1}^n |w_j| \quad (2)$$

式中: J 为目标函数; n 为特征数量; \mathbf{w} 为回归系数向

量; w_j 为第 j 个特征的回归系数; x_{ij} 为第 i 个样本的第 j 个特征值; y_i 为第 i 个样本的真实值; λ 为惩罚函数权重,即回归系数的压缩程度。

采用 K 折交叉验证方法获得 Lasso 模型的性能最优估计。 K 折交叉验证是一种常用的模型评估技术,用于评估机器学习模型的泛化性能。其具体流程为:将原始数据集分为 K 个子集(称为折);随后对模型进行 K 次训练与验证,在每次训练中,选取其中 1 个折作为验证集,剩下的 $K-1$ 个折作为训练集;最后对 K 次验证得到的评估指标计算平均值,作为最终的模型性能评估结果。

2.3 模型构建流程

将 Lasso 算法筛选出的特征作为 XGboost 模型的输入变量,建立溃口峰值流量预测模型,即 Lasso-XGBoost 模型。XGBoost 作为一种新兴的机器学习算法,采用迭代方式将多个弱学习器集成提升为强学习器。该算法通过最小化包含正则化项的目标函数来优化模型,在每次迭代中利用损失函数的二阶泰勒展开信息来构建新的决策树,从而逐步降低预测误差。XGboost 模型的具体步骤如下:①初始化每个样本的预测值;②计算损失函数对于每个样本预测值的导数;③根据导数信息建立一棵新决策树;④将新决策树的预测结果累加到原来的预测值上;⑤重复步骤②~④,直到满足停止条件;⑥输出预测结果。溃口峰值流量 Lasso-XGBoost 模型构建流程见图 2。

3 模型验证

为了构建预测精度较高的溃口峰值流量预测模型,将 Lasso 算法的特征选择结果作为 XGBoost 模型的输入参数,并对模型参数进行调试,选择最优参数对溃口峰值流量进行预测。

3.1 特征确定

本文采用斯皮尔曼相关性热图直观展示各特征之间的相关性关系,结果如图 3 所示。由图 3 可见,峰值流量 Q_p 与 h_d 、 h_w 、 h_b 、 V 、 V_w 、 B_{ave} 的斯皮尔曼相关系数分别为 0.51、0.42、0.46、0.97、0.97、0.42。各特征与 Q_p 的相关系数均大于 0.3; V 、 V_w 与 Q_p 的相关性较高,而 h_d 、 h_w 、 h_b 、 B_{ave} 与 Q_p 的相关性较低,但其最低相关系数为 0.42,表明它们仍可能包含有价值的信息。因此,根据斯皮尔曼相关系数的分析结果,可初步将影响峰值流量的特征分为强相关与弱相关两类。值得注意的是,4 个弱相关变量的系数值接近,不能简单地将其剔除。

根据本文数据库的样本数据,采用十折交叉验证计算不同惩罚权重 λ 下的均方误差 (MSE),确定

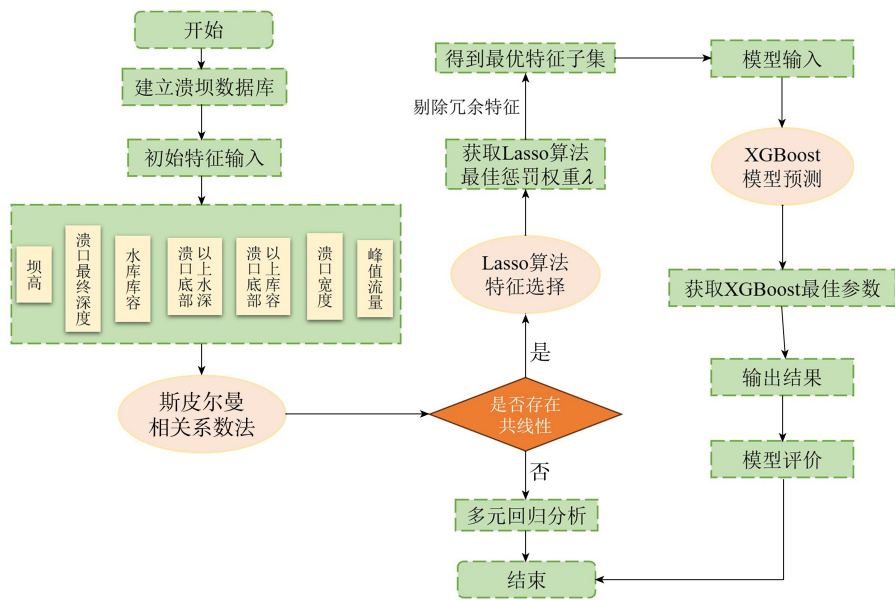


图2 Lasso-XGBoost 模型构建流程

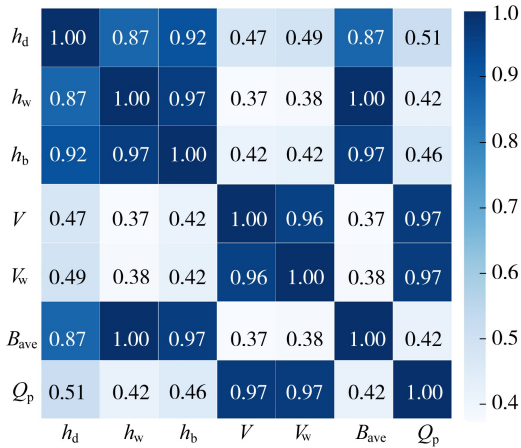


图3 斯皮尔曼相关性热图

了最优惩罚权重,并绘制了除峰值流量之外其余6个特征的Lasso路径(图4)。由图4可见,随着 λ 不断增大,各特征的回归系数逐渐趋向于零,其中,回归系数被压缩至0的速度越缓慢,表明该特征的影响越重要。当 $\ln\lambda$ 取值为0.1,即最优惩罚权重 λ 为-1时, h_w 、 h_d 的回归系数被压缩至0,表明其对 Q_p 的影响较弱,可作为冗余特征剔除;而其他4个特征

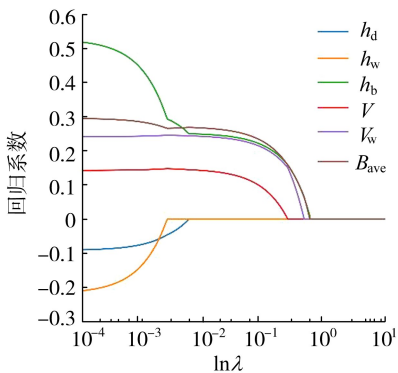


图4 Lasso 路径

h_b 、 V 、 V_w 、 B_{ave} 的回归系数均为正值,分别为 0.254、0.102、0.251、0.265,表明这4个特征与峰值流量的相关性较高。与主成分分析等重构型降维方法不同,Lasso 算法通过参数压缩而非空间变换实现降维,因此保留了原始参数的物理意义,这对水利工程的机理分析尤为重要。最终,通过 Lasso 算法从初始6个特征中挖掘出4个高相关性特征(h_b 、 V 、 V_w 、 B_{ave}),从而实现了特征选择。

3.2 XGBoost 参数设置与模型评价指标

XGBoost 算法的参数较多,根据多次训练结果,本文选取对预测结果影响较大的3个参数(学习率、树的最大深度、最佳迭代次数)使用K折交叉检验进行调参,调参范围如表2所示。为了全面、可靠地评价预测模型的预测性能,采用均方根误差(RMSE)、平均绝对误差(MAE)、均方误差(MSE)和决定系数(R^2)作为评价指标^[19]。

表2 XGBoost 模型参数调优范围

参数	调参范围	步长
学习率	[0.05,0.4]	0.05
树的最大深度	[3,9]	1
最佳迭代次数	[100,400]	100

3.3 预测精度分析

为了验证 Lasso-XGBoost 模型的优越性及特征选择方法的有效性,本文设计了多组对比实验。首先,选取支持向量回归(SVR)和岭回归(RR)作为对比基准模型;其次,分别对 XGBoost、SVR 和 RR 3种模型在特征选择前后的预测性能进行对比分析。将溃坝数据库按8:2比例划分为训练集和测试集,基于训练集数据对各模型进行训练,利用训练好的模型对测试集数据进行预测,获得对应的预测结果。

特征选择前后各模型的预测结果如图 5 所示,图中带状阴影表示 95% 置信区间。各模型测试集预测性能结果如表 3 所示。

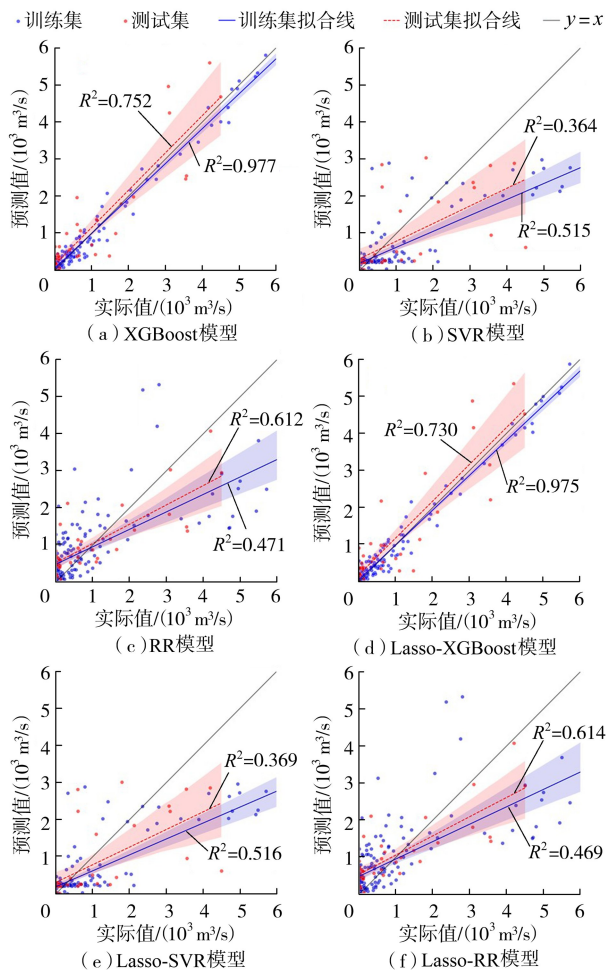


图 5 各模型预测结果对比

表 3 各模型测试集预测性能评价指标对比

模型	RMSE/(m³/s)		MAE/(m³/s)	
	训练集	测试集	训练集	测试集
XGBoost	190	686	140	484
Lasso-XGBoost	277	627	197	483
SVR	1016	1052	562	639
Lasso-SVR	1015	1047	562	637
RR	1061	819	676	607
Lasso-RR	1063	821	671	604

模型	MSE/(m³/s)		R²	
	训练集	测试集	训练集	测试集
XGBoost	36410	470995	0.729	0.864
Lasso-XGBoost	76834	393557	0.774	0.983
SVR	1032834	1106724	0.515	0.364
Lasso-SVR	1030457	1098066	0.516	0.369
RR	1126548	671781	0.471	0.612
Lasso-RR	1130599	671781	0.469	0.614

由图 5 和表 3 可见,在 XGBoost、SVR 和 RR 3 个模型中,XGBoost 模型的训练集拟合线和测试集拟合线最接近 $y=x$ 线,且测试集的 RMSE、MAE 和 MSE 均为最小, R^2 最大,表明 XGBoost 模型性能的优越性。经过特征选择后,3 个模型的预测精度均

有不同程度的提升,其中 XGBoost 模型的提升效果最显著,表明使用 Lasso 算法进行特征选择可以有效改善模型性能。综合来看,Lasso-XGBoost 模型最适用于溃口峰值流量预测,能够在保证模型精度的同时,提高特征选择的效率。在计算效率方面,由于 Lasso 特征选择将特征维度从 6 维降至 4 维,XGBoost 决策树的平均深度减少 2 层,使得 Lasso-XGBoost 模型的平均训练耗时 (8.7 s) 较全参数 XGBoost 模型 (15.3 s) 缩短了 43.1%,计算效率显著提升。

4 结论

a. 斯皮尔曼相关系数法和 Lasso 算法相结合的特征选择方法是有效的,Lasso 算法具有良好的降维效果。该特征选择方法能够有效识别并剔除由多重共线性产生的冗余特征,可筛选出对峰值流量影响较大的特征,经过特征选择后的溃口峰值流量预测模型性能更加优越。

b. 对比 Lasso 算法在 SVR、RR 及 XGBoost 3 个机器学习模型中的应用效果,发现 XGBoost 模型的溃口峰值流量预测精度最高,表明 XGBoost 模型与 Lasso 算法的适配性最佳。该组合具有良好的非线性信息挖掘能力,可以有效捕捉溃口特征间复杂的非线性函数关系,在溃坝发展机理研究和溃口规模预测中具有较好的应用价值。

参考文献:

[1] 向衍,盛金保,刘成栋,等.土石坝长效服役与风险管理研究进展[J].水利水电科技进展,2018,38(5):86-94. (Xiang Yan, Sheng Jinbao, Liu Chengdong, et al. Research progress in long-term service and risk assessment of earth-rockfill dams[J]. Advances in Science and Technology of Water Resources, 2018, 38(5): 86-94. (in Chinese))

[2] Wahl T L. Uncertainty of predictions of embankment dam breach parameters[J]. Journal of Hydraulic Engineering, 2004, 130(5): 389-397.

[3] Xu Yuanhua, Zhang L M. Breaching parameters for earth and rockfill dams [J]. Journal of Geotechnical and Geoenvironmental Engineering, 2009, 135(12): 1957-1970.

[4] Froehlich D C. Peak outflow from breached embankment dam [J]. Journal of Water Resources Planning and Management, 1995, 121(1): 90-97.

[5] Sammen S S, Mohamed T A, Ghazali A H, et al. An evaluation of existent methods for estimation of embankment dam breach parameters[J]. Natural Hazards, 2017, 87(1): 545-566.

[6] Hooshyaripor F, Tahershamsi A, Golian S. Application of

- copula method and neural networks for predicting peak outflow from breached embankments[J]. *Journal of Hydro-Environment Research*, 2014, 8(3):292-303.
- [7] Nourani V, Hakimzadeh H, Amini A B. Implementation of artificial neural network technique in the simulation of dam breach Hydrograph[J]. *Journal of Hydroinformatics*, 2012, 14(2):478-496.
- [8] Nobarinia M, Kalateh F, Nourani V, et al. Dam failure peak outflow prediction through GEP-SVM meta models and uncertainty analysis[J]. *Water Supply*, 2021, 21(7):3387-3401.
- [9] Ahmad M W, Reynolds J, Rezugui Y. Predictive modelling for solar thermal energy systems: a comparison of support vector regression, random forest, extra trees and regression trees[J]. *Journal of Cleaner Production*, 2018, 203:810-821.
- [10] Zhou Lanting, Long Guanlin, Hu Cancan, et al. Reservoir water level prediction using combined CEEMDAN-FE and RUN-SVM-RBFNN machine learning algorithms [J]. *Water Science and Engineering*, 2025, 18(2):177-186.
- [11] Chen Tianqi, Guestrin C. XGBoost: a scalable tree boosting system [C]//*Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco: ACM, 2016:785-794.
- [12] 佟大威, 杨传会, 余佳, 等. 基于 XGBoost-PSO 的混凝土重力坝体型多目标优化设计[J]. *河海大学学报(自然科学版)*, 2023, 51(3):91-98. (Tong Dawei, Yang Chuanhui, Yu Jia, et al. Multi-objective shape optimization of concrete gravity dam based on XGBoost-PSO [J]. *Journal of Hohai University (Natural Sciences)*, 2023, 51(3):91-98. (in Chinese))
- [13] 马春辉, 侯媛媛, 杨杰, 等. 基于 RUN-XGBoost 算法的土石坝渗流预测模型[J]. *水利水电科技进展*, 2024, 44(2):72-78. (Ma Chunhui, Hou Yuanyuan, Yang Jie, et al. Seepage prediction model of earth-rockfill dams based on RUN-XGBoost algorithm[J]. *Advances in Science and Technology of Water Resources*, 2024, 44(2):72-78. (in Chinese))
- [14] 许秋鸿, 刘晓青. 基于 VMD-XGBoost-GRU 模型的危岩体变形预测[J]. *水利水电科技进展*, 2024, 44(2):92-98. (Xu Qiuhong, Liu Xiaoqing. Deformation prediction of dangerous rock mass based on VMD-XGBoost-GRU model [J]. *Advances in Science and Technology of Water Resources*, 2024, 44(2):92-98. (in Chinese))
- [15] Morris M W, Hassan M A A M, Vaskinn K A. Breach formation: field test and laboratory experiments [J]. *Journal of Hydraulic Research*, 2007, 45(S1):9-17.
- [16] Zhang Jianyun, Li Yun, Xuan Guoxiang, et al. Overtopping breaching of cohesive homogeneous earth dam with different cohesive strength[J]. *Science in China Series E: Technological Sciences*, 2009, 52(10):3024-3029.
- [17] 陈生水, 陈祖煜, 钟启明. 土石坝和堰塞坝溃决机理与溃坝数学模型研究进展[J]. *水利水电技术*, 2019, 50(8):27-36. (Chen Shengshui, Chen Zuyun, Zhong Qiming. Progresses of studies on failure mechanism and numerical dam failure model of earth-rockfill dam and landslide dam [J]. *Water Resources and Hydropower Engineering*, 2019, 50(8):27-36. (in Chinese))
- [18] 李炎隆, 田智雯, 张野, 等. 基于 CTGAN 数据增强的土石坝溃坝峰值流量预测[J]. *水利水电科技进展*, 2025, 45(3):70-76. (LI Yanlong, TIAN Zhiwen, ZHANG Ye, et al. Prediction of peak discharge of earth-rock dam failure based on CTGAN data augmentation [J]. *Advances in Science and Technology of Water Resources*, 2025, 45(3):70-76. (in Chinese))
- [19] 梅世昂, 陈生水, 钟启明, 等. 土石坝溃坝参数模型研究[J]. *工程科学与技术*, 2018, 50(2):60-66. (Mei Shi'ang, Chen Shengshui, Zhong Qiming, et al. Parametric model for breaching analysis of earth-rock dam [J]. *Advanced Engineering Sciences*, 2018, 50(2):60-66. (in Chinese))
- [20] Zhong Qiming, Chen Shengshui, Fu Zhongzhi, et al. New empirical model for breaching of earth-rock dams [J]. *Natural Hazards Review*, 2020, 21(2):06020002.
- [21] Dzulkalnine M F, Sallehuddin R. Missing data imputation with fuzzy feature selection for diabetes dataset [J]. *SN Applied Sciences*, 2019, 1(4):362.
- [22] 崔鸿雁, 徐帅, 张利锋, 等. 机器学习中的特征选择方法研究及展望[J]. *北京邮电大学学报*, 2018, 41(1):1-12. (Cui Hongyan, Xu Shuai, Zhang Lifeng, et al. The key techniques and future vision of feature selection in machine learning [J]. *Journal of Beijing University of Posts and Telecommunications*, 2018, 41(1):1-12. (in Chinese))
- [23] 陈茜, 黄连兵. 基于 LASSO-LARS 的回采工作面瓦斯涌出量预测研究[J]. *煤炭科学技术*, 2022, 50(7):171-176. (Chen Qian, Huang Lianbing. Gas emission prediction from coalface based on Least Absolute Shrinkage and Selection Operator and Least Angle Regression [J]. *Coal Science and Technology*, 2022, 50(7):171-176. (in Chinese))

(收稿日期:2024-12-26 编辑:雷燕)