

基于 CTGAN 数据增强的土石坝溃坝峰值流量预测

李炎隆, 田智雯, 张野, 邱文, 王婷

(西安理工大学省部共建西北旱区生态水利国家重点实验室, 陕西 西安 710048)

摘要: 为准确预测土石坝溃坝峰值流量, 建立了包含 156 个土石坝溃坝案例的数据库, 并基于溃坝过程和相关性分析, 选取坝型、溃坝模式、溃口以上水深和溃口以上库容为预测洪峰流量的控制变量。由于数据量有限且溃坝数据难以获取, 使用条件表格生成对抗网络 (CTGAN) 增强土石坝溃坝数据, 以增加样本多样性, 丰富模型训练信息, 并提高泛化能力。基于增强数据, 采用 CatBoost 算法构建溃坝峰值流量预测模型。结果表明, 基于增强数据建立的模型预测精度更高, 拟合系数达 0.93, 展现出卓越的预测性能。结合可解释机器学习 (SHAP) 模型分析发现溃口以上库容对峰值流量的影响最为显著, 溃口以上水位次之。

关键词: 土石坝; 峰值流量; CTGAN; 数据增强; CatBoost; SHAP

中图分类号: TV122.5; TV641 **文献标志码:** A **文章编号:** 1006-7647(2025)03-0070-07

Prediction of peak discharge of earth-rock dam failure based on CTGAN data augmentation//LI Yanlong, TIAN Zhiwen, ZHANG Ye, QIU Wen, WANG Ting (State Key Laboratory of Eco-hydraulics in Northwest Arid Region, Xi'an University of Technology, Xi'an 710048, China)

Abstract: In order to accurately predict the peak discharge of earth-rock dam failure, a database containing 156 cases of earth-rock dam failure was established. Based on the failure process and correlation analysis, the dam type, failure mode, the height of water above the breach, and the volume of water stored above the breach were selected as control variables. Considering the constraints of limited available data and the difficulty in obtaining dam failure data, a conditional tabular generative adversarial network (CTGAN) was used to augment the earth-rock dam failure data, so as to increase the diversity of samples, enrich the model training information, and improve the generalization ability. Based on the augmented data, a peak discharge prediction model was constructed using the CatBoost algorithm. The results show that the model based on the augmented data achieves higher prediction accuracy, with a coefficient of determination reaching 0.93, demonstrating excellent prediction performance. Combined with the shapley additive explanation (SHAP) analysis, it is found that the volume of water stored above the breach has the most significant impact on the peak discharge, followed by the height of water above the breach.

Key words: earth-rock dam; peak discharge; CTGAN; data augmentation; CatBoost; SHAP

据统计, 水库大坝溃坝案例中 90% 以上为土石坝^[1]。溃坝峰值流量的准确预测, 对下游防洪措施制定和风险预警至关重要。然而, 由于数据总量少和数据缺失的问题, 目前峰值流量的预测值与实测值仍存在较大差距, 因此, 亟待建立准确的土石坝溃坝峰值流量预测模型。

近年来, 土石坝溃坝峰值流量的预测研究主要集中在参数模型和智能算法预测方面。参数模型依据溃坝案例数据, 通过统计分析和回归分析进行峰值流量预测, 其不用考虑物理过程, 且参数易于获取, 计算高效便捷, 并能够提供可靠的结果^[2-14]。Xu

等^[15]基于 38 个溃坝案例, 通过回归分析建立了峰值流量预测模型; Zhong 等^[16]基于 120 例溃坝案例建立了预测峰值流量的经验模型, 其拟合系数为 0.89。该模型在数据量的使用上大幅提高, 但预测精度仍有待提升。随着计算机网络的发展, 人工智能算法与峰值流量的预测逐渐融合。Pektas 等^[17]基于 35 个土石坝溃坝案例, 使用人工神经网络来预测峰值流量, 所建模型的拟合系数为 0.80, 并对选用的参数进行了敏感性分析; Elmazoghi 等^[18]收集了 93 个溃坝案例, 基于自适应神经模糊推理算法建立预测模型, 模型的拟合系数可达 0.91。以上研究

基金项目: 国家杰出青年科学基金项目 (52125904); 陕西省创新科研团队项目 (2022TD-01)

作者简介: 李炎隆 (1980—), 男, 教授, 博士, 主要从事土石坝安全评价和风险分析研究。E-mail: liyanlong@xaut.edu.cn

尽管模型评价结果较好,但溃坝案例数不足,限制了模型的泛化能力。要建立泛化能力强和预测精度高的土石坝溃坝峰值流量预测模型,需要更大规模的数据。但土石坝修建时间较早,具备详细资料的坝体较少,同时还面临数据缺失等问题,导致现有的溃坝数据有限。为了突破数据对模型训练的限制,数据增强方法逐渐得到关注。Hooshyaripor 等^[19]基于 93 个实际溃坝案例和 Copula 方法生成 1020 个合成溃坝案例,并利用神经网络预测峰值流量,模型的拟合系数达 0.87,但该模型未对不同溃坝模式和坝型进行区分;Shi 等^[20]在堰塞体稳定性评价中为解决数据缺失问题,使用自动学习方法得出最优分裂方向,从而对缺失数据进行补全。以上方法在一定程度上解决了数据总量小和数据缺失的问题。因此,本文拟引入数据增强方法,建立泛化能力更强的预测模型。

本文建立了含有 156 个土石坝溃坝案例的数据库,并依据溃坝过程和参数相关性分析,选择预测峰值流量的控制变量。采用条件表格生成对抗网络(conditional tabular generative adversarial network, CTGAN)算法对数据进行增强,对增强后的数据使用类别特征提升(categorical boosting, CatBoost)算法建立预测模型,使用均方根误差 RMSE 和拟合系数 R^2 评价模型性能,分析数据增强对所建模型预测能力的影响。最后引入可解释机器学习(shapley additive explanation, SHAP)模型对数据增强后建立的模型进行解释,以分析影响溃坝峰值流量预测的关键特征。

1 土石坝溃坝数据库的建立

基于文献资料^[9, 15-16, 20-23],建立了 156 例土石坝溃坝数据库,收集了溃坝峰值流量、溃坝模式、坝型、坝高、溃口以上水深、溃口以上库容等数据。为了确保峰值流量预测结果的准确性,需要选择与其相关性强的因素,以提高参数的有效性。同时,为了避免数据缺失导致预测结果可信度降低的问题,应避免选择存在较多数据缺失的案例。土石坝溃坝峰值流量 Q_p 与坝型 D 、溃坝模式 F 、坝高 h_d 、库容 V 、溃口以上水深 h_w 、溃口以上库容 V_w 等因素有关^[15],需依据溃坝过程和溃坝参数相关性分析选取合理的控制变量。对各变量与峰值流量的关系进行了相关性分析,结果见表 1。

土石坝溃坝可分为漫顶溃坝和渗透破坏导致的溃坝两种^[1]。在建立的数据库中,76.93%的大坝因降雨或滑坡导致漫顶而发生溃坝,21.15%因渗透破坏而溃坝,其余 1.92%未记录其溃坝模式。不同的

土石坝溃坝模式会导致不同的水流特性和溃坝峰值流量^[24-26]。在数据库中,76.93%为均质坝,17.95%为心墙坝,3.84%为混凝土面板堆石坝,1.28%的坝体未收集到坝型资料。需要注意的是,由于面板堆石坝的溃坝案例数量较少,仅有 6 例,因此分析时仅考虑均质坝和心墙坝。这两种坝型的结构和溃坝行为不同,溃坝峰值流量大小也会有所不同^[27-28],因此,需要在分析中区分不同的溃坝模式和坝型。

表 1 土石坝溃坝峰值流量预测参数统计分析

统计参数	h_d/m	V/m^3	h_w/m	V_w/m^3	$Q_p/(\text{m}^3/\text{s})$
最大值	93.00	6.50×10^8	77.40	6.6×10^8	78 100.00
最小值	1.20	1.33×10^4	1.20	7.4×10^3	9.20
中位数	11.58	6.70×10^5	11.17	7.5×10^5	495.00
平均值	16.18	2.12×10^7	14.42	2.1×10^7	3 607.00
偏 度	2.20	4.80	2.00	6.40	5.30
峰 度	5.60	24.90	5.30	44.50	31.50
相关系数	0.51	0.80	0.58	0.88	1.00

土石坝溃坝峰值流量预测时通常根据 D 和 F 对数据进行分类,并选用 h_d 、 h_w 、 V 和 V_w 作为控制变量进行回归预测^[15]。根据上述相关性分析最终选择 h_w 和 V_w 作为控制变量,进行峰值流量预测。综上所述,本文模型以 D 、 F 、 h_w 、 V_w 作为控制变量,进行土石坝溃坝峰值流量回归预测。所选参数易得,数据缺失少,可确保模型的可行性与合理性。

2 基于 CTGAN 算法的土石坝溃坝数据增强

土石坝溃坝峰值流量的精准预测受到数据规模的限制。但在实际工程中,土石坝溃坝案例有限,收集困难;另一方面,通过土石坝溃坝试验获得数据的成本较高,且与实际数据存在一定差距。因此,本文通过生成对抗网络(generative adversarial network, GAN)^[29]对实际溃坝案例数据进行增强,以提升模型的泛化能力。

GAN 是近几年在图像领域应用较为广泛的数据增强算法,然而对于表格数据,传统 GAN 表现不佳。为解决该问题,本文采用 CTGAN 算法进行数据增强。该算法引入条件生成机制,考虑了表格数据的特点,提高了生成数据的可控性,其原理见图 1,其中 L1 和 L2 表示数据中的分类变量列,其他变量含义见文献[30]。将建立的原始数据输入生成器进行学习,得到增强后的数据,并通过判别器进行真假判别,最终建立增强后的溃坝数据库。原始数据库与增强数据库各变量的范围如表 2 所示。

表 2 中的原始数据库 $D_1 = \{(x_{li}, Q_{pli})\}$,增强数据库为通过 CTGAN 算法增强的数据库 $D_2 = \{(x_{2i}, Q_{p2i})\}$ 。在原始数据库中高于 50 m 的大坝溃坝案例仅 1 例,数据增强后有 5 例;原始数据库中库容大于

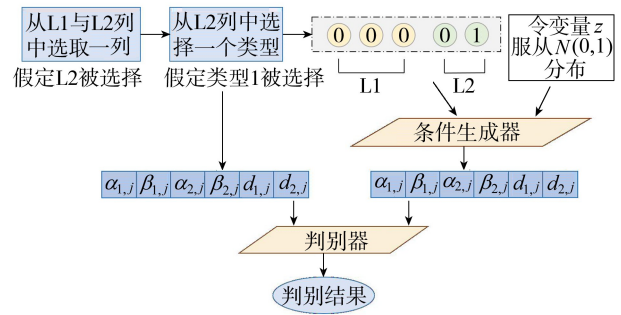


图 1 CTGAN 算法原理

表 2 原始数据库与增强数据库变量范围

统计 参数	原始数据库			增强数据库		
	h_w/m	V_w/m^3	$Q_p/(m^3/s)$	h_w/m	V_w/m^3	$Q_p/(m^3/s)$
最大值	77.40	6.60×10^8	78 100.00	55.19	3.64×10^8	120 842.00
最小值	1.20	7.40×10^3	9.20	1.52	3.65×10^3	9.13
中位数	10.50	6.85×10^5	436.50	12.06	1.26×10^6	544.74
平均值	13.79	1.84×10^7	3 406.96	14.32	1.86×10^7	4 608.65

10^8 m^3 的溃坝案例仅有 4 例,数据增强后共有 25 例;在原始数据库中峰值流量大于 $10^5 \text{ m}^3/\text{s}$ 的有 8 例,数据增强后共有 71 例。可见数据增强方法弥补了原始数据在高坝、大库容和大流量方面的数据限制。

在数据增强过程中,对比原始数据和增强数据的分布是至关重要的。原始数据反映真实样本情况,而增强数据是通过模型生成的合成样本。通过

对比两者的分布情况,可以评估数据增强的效果,了解增强数据能否准确地反映原始数据的特征。比较原始数据与增强数据中坝型、溃坝模式等分类变量的占比情况。计算得到每个类别的数量占总样本数量的比例,心墙坝、均质坝、漫顶溃坝模式和渗透破坏溃坝模式在增强数据中的占比分别为 25.2%、74.8%、55.0%、45.0%,在原始数据中的占比分别为 18.5%、81.5%、80.7%和 19.3%。可见增强数据的分类变量占比与原始数据相近。

之后分析连续型变量与溃坝参数间的数据分布,绘制原始数据与增强数据的对比。图 2(a)、图 2(b)分别展示了溃口以上水位和溃口以上库容与溃坝峰值流量的关系,可以看出,CTGAN 算法对数据进行增强后,不仅保持了原始数据的分布特征,而且丰富了数据的多样性,其生成的数据与原始数据在结构和数据分布上保持着一致性。

图 3 给出了原始数据及增强数据的各参数相关性。通过对比可以观察到,基于 CTGAN 算法的数据增强方法在模拟分类变量与其他变量之间的相关性方面表现较好。在原始数据中与 Q_p 相关性最大的变量是 V_w ,基于 CTGAN 算法的数据增强方法也较好地模拟到了这二者之间的关系。

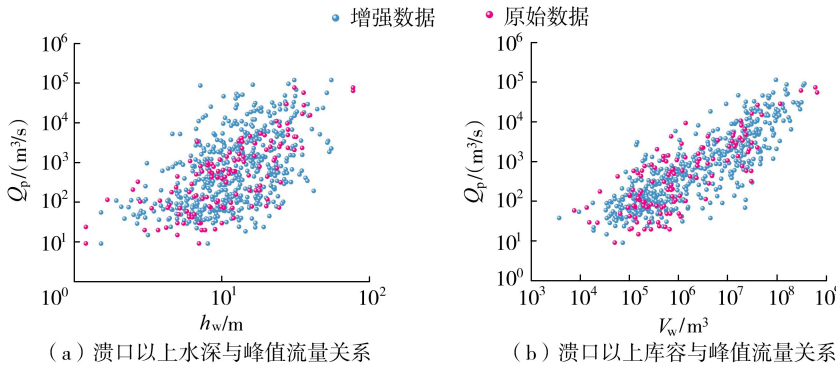


图 2 原始数据与增强数据对比

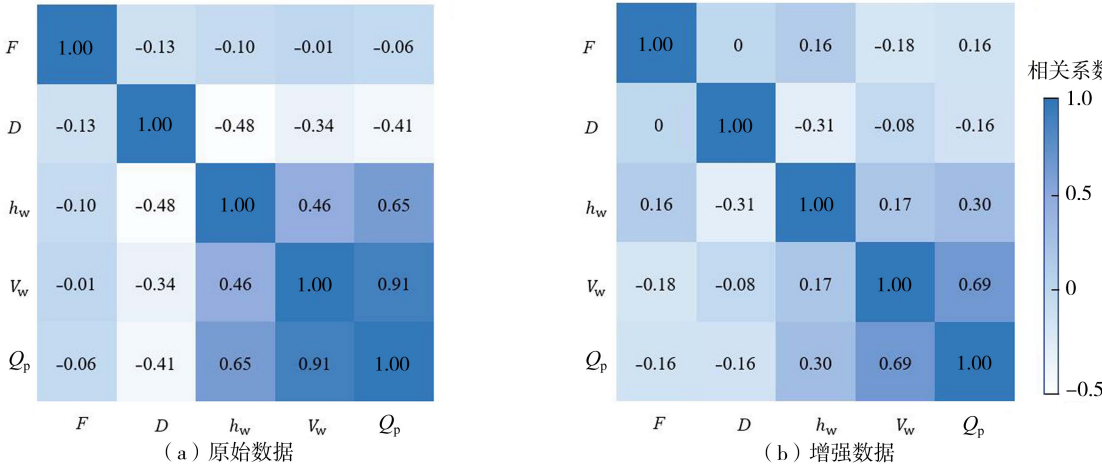


图 3 原始数据及增强数据的参数相关性对比

3 基于增强数据和 CatBoost 算法的土石坝溃坝峰值流量预测

3.1 土石坝溃坝峰值流量预测模型建立

CatBoost 算法能够直接处理原始的类型型数据,无需进行额外编码等转换,能够更好地捕捉数据中的非线性关系,有助于提高模型对复杂模式的拟合能力。CatBoost 算法是基于对称决策树为基学习器实现的,能高效合理地处理类别变量^[31],因此本文选用 CatBoost 算法进行建模,所建模型即为 CatBoost 模型。该算法在模型预测中可以解决梯度偏差和预测偏移的问题,从而减少过拟合的发生,进而提高算法的准确性和泛化能力。

对于增强后的土石坝溃坝数据库,以利用 CTGAN 算法增强的数据作为训练集,占总数据量的 80%,以原始数据作为测试集,占总数据量的 20%,基于该数据库建立增强模型。对原始数据库中数据进行随机划分,80%为训练集,20%为测试集,基于该数据库建立原始模型。图 4 为模型构建流程,先构建初始化参数模型,再将均方根误差 RMSE 作为损失函数对模型进行改进,采用网格搜索优化超参数 iterations、depth、learning_rate 和 l2_leaf_reg,超参数搜索范围和取值见表 3。

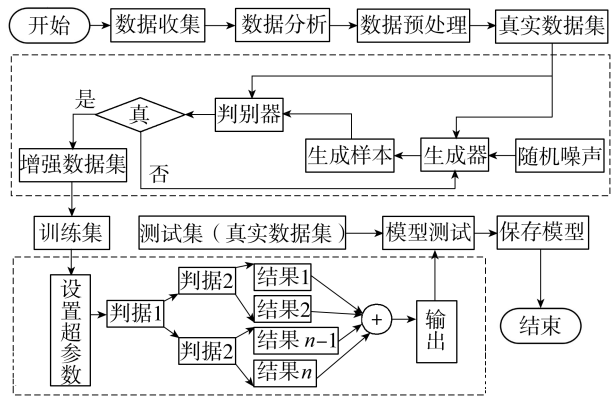


图 4 模型构建流程

表 3 模型超参数优化

超参数	搜索范围	超参数取值	
		原始模型	增强模型
iterations	10~500	130.00	340.0
depth	1~10	3.00	5.0
learning_rate	0.01~0.3	0.05	0.2
l2_leaf_reg	0.1~10	9.00	6.0

3.2 结果对比

在原始数据上,对增强模型和原始模型的预测性能进行对比。图 5(a)、图 5(b)分别为原始模型和增强模型对原始数据的预测结果。从图 5 可以看出基于增强数据库建立的增强模型预测性能更好,数据点更接近 $y_{\text{预测}} = y_{\text{实测}}$ 直线,尤其是在溃坝峰值

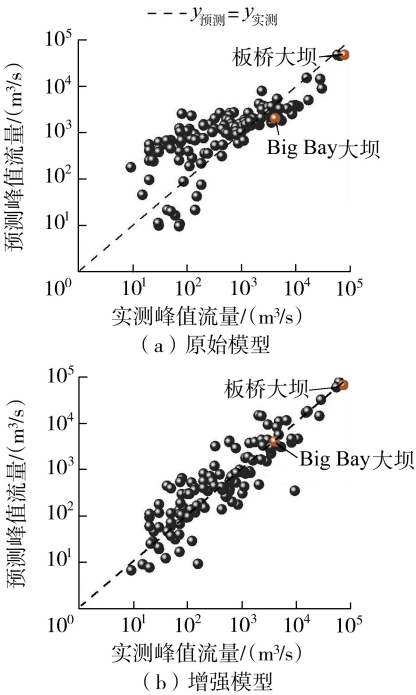


图 5 模型预测结果对比

流量处于 $10 \sim 1\,000 \text{ m}^3/\text{s}$ 的区间内的案例。从整体看,增强模型预测值更接近实测值,表明基于增强数据库建立的增强模型预测能力更优。将增强模型与现有参数模型对比,计算各模型的 R^2 和 RMSE,结果见表 4。参数模型建立时所使用的案例数较少,导致模型的泛化能力弱;随着模型的发展以及收集到的溃坝数据增加,模型的泛化能力增强。通过表 4 对比各模型的 R^2 和 RMSE 可以看出,随着模型的发展,其 R^2 呈现出增大趋势, RMSE 呈现减小趋势,模型中考虑的控制变量数量增加,其预测精度提高。通过对比表明,与现有参数模型相比,本文基于 CTGAN 算法进行数据增强、采用 CatBoost 算法建立的 CTGAN-CatBoost 模型有最大的 R^2 和最小的 RMSE,其 R^2 和 RMSE 分别为 0.93 和 $2\,936 \text{ m}^3/\text{s}$ 。

4 模型特征重要性分析及实例应用

4.1 特征重要性分析

基于增强数据库建立的溃坝峰值流量预测模型是黑箱模型,且算法复杂度高,可解释性较低,为解决模型可解释性问题,引入 SHAP 进行模型分析,见式(1)。SHAP 是在合作博弈论的启发下构建的加性解释模型,所有的特征都视为“贡献者”。对于每个预测样本,模型都会产生一个预测值,SHAP 值就是该样本中每个特征所分配到的数值。SHAP 模型表达式为

$$Q_{pi} = Q_{p\text{-base}} + f(x_{i1}) + f(x_{i2}) + \cdots + f(x_{ik}) \quad (1)$$

式中: x_{ik} 为第 i 个土石坝的第 k 个特征; Q_{pi} 为模型

表 4 增强模型与现有参数模型对比

序号	研究者或模型名称	模型计算公式	模型建立时 使用案例数	本文计算 使用案例数	R^2	RMSE/(m^3/s)
1	Kirkpatrick ^[2]	$Q_p = 1.268(h_w + 0.3)^{2.5}$	19	146	0.26	8910
2	SCS ^[3]	$Q_p = 16.6h_w^{1.85}$	13	146	0.34	8446
3	Singh 等 ^[6]	$Q_p = 13.4h_d^{1.89}$ $Q_p = 1.776V^{0.47}$	28	142	0.46	7909
4	Costa ^[10]	$Q_p = 0.981(h_d V)^{0.42}$	31	142	0.46	7897
5	Evans ^[11]	$Q_p = 0.72V_w^{0.53}$	29	142	0.59	6768
6	USBR ^[12]	$Q_p = 19.1h_w^{1.85}$	21	146	0.31	8624
7	Froehlich ^[21]	$Q_p = 0.607V_w^{0.295}h_w^{1.24}$	22	130	0.54	7251
8	Xu 等 ^[15]	$Q_p = 0.175g^{0.5}V_w^{5/6}\left(\frac{h_d}{h_r}\right)^{0.199}\left(\frac{V^{1/3}}{h_w}\right)^{1.274}e_4^B$ $Q_p = 0.1202L^{1.7856}$	75	38	0.59	6958
9	Thornton ^[23]	$Q_p = 0.863V_w^{0.335}h_d^{1.833}W_{ave}^{-0.663}$ $Q_p = 0.012V_w^{0.493}h_d^{1.205}L^{0.226}$	38	36	0.50	12989
10	Hooshyaripor 等 ^[19]	$Q_p = 0.0212V_w^{0.5429}h_w^{0.8713}$	93	130	0.60	6848
11	Froehlich ^[23]	$Q_p = 0.0175k_M k_H (gV_w h_b^2/W_{ave})^{0.5}$	41	47	0.54	11288
12	Zhong 等 ^[16]	$Q_p = \left(\frac{V_w^{1/3}}{h_w}\right)^{-1.58}\left(\frac{h_w}{h_b}\right)^{-0.76}h_d^{0.10}e^{-4.55}V_w g^{0.5}h_w^{-0.5}$ (均质坝) $Q_p = \left(\frac{V_w^{1/3}}{h_w}\right)^{-1.51}\left(\frac{h_w}{h_b}\right)^{-1.09}h_d^{-0.12}e^{-3.61}V_w g^{0.5}h_w^{-0.5}$ (心墙坝)	120	128	0.81	4531
13	CTGAN-RF		654	130	0.88	3605
14	CTGAN-GBDT		654	130	0.89	3426
15	CatBoost		130	130	0.86	4304
16	CTGAN-CatBoost		654	130	0.93	2936

注: g 为重力加速度; h_b 为溃口深; h_r 为参考坝高,设置为15 m; L 为坝轴线长; W_{ave} 为坝体平均宽度。其他物理量含义及取值见各公式相应文献。序号13~16为本文建立的算法模型,其中CTGAN-RF和CTGAN-GBDT分别为基于增强数据库建立的随机森林模型和梯度提升树模型。

对该土石坝的峰值流量预测值; Q_{p-base} 为训练集所有峰值流量的均值,为整个模型的基线; $f(x_{ik})$ 为 x_{ik} 的SHAP值。

采用SHAP模型对CTGAN-CatBoost模型进行特征重要性分析。分析结果表明,在该模型中,特征重要性排序从大到小依次为: V_w 、 h_w 、 F 、 D 。其中, V_w 和 h_w 对溃坝峰值流量的影响为正向的,随着 V_w 和 h_w 的增大,SHAP值增大,溃坝峰值流量也增大,这与实际工程相符。通常越高的土石坝具有更大的库容,对应着更高的水头,因此峰值流量也会更大。同时,从SHAP分析结果可以看出,溃坝模式和坝型对峰值流量的影响较小。

4.2 实例分析

本文选取中国板桥大坝和美国Big Bay大坝两个典型案例分析模型的适用性和泛化能力。板桥水库位于淮河支流,坝体始建于1951年,1956年扩建加固。1975年8月4日,中国内地第三号台风登陆,形成罕见的大暴雨,其暴雨中心位于河南泌阳板桥水库附近的林庄。板桥水库最大设计库容为4.92亿 m^3 ,但承受的洪水总量达7.012亿 m^3 ,最终

于1975年8月8日发生漫顶溃坝^[15]。Big Bay大坝位于密西西比州拉马尔县珀维斯以西约18 km处。坝轴线长576 m,坝高15.6 m。2004年3月由于主溢洪道附近发生渗透破坏,最终导致溃坝^[32]。计算参数和模型计算结果分别见表5和表6。

表 5 土石坝溃坝峰值流量计算参数取值

工程名称	溃坝模式	坝型	h_d/m	h_w/m	V_w/m^3	$Q_p/(\text{m}^3/\text{s})$
Big Bay 大坝	渗透破坏	均质坝	15.64	13.59	1.75×10^7	4160
板桥大坝	漫顶	心墙坝	24.50	31.00	6.08×10^8	78100

由表6可以看出,本文基于CTGAN算法进行数据增强、采用CatBoost算法建立的CTGAN-CatBoost模型相比于未采用数据增强的CatBoost模型,在Big Bay大坝和板桥大坝两个工程中,溃坝流量预测相对误差分别降低了45.77%和21.92%。同时,基于增强数据库,分别采用随机森林(random forest, RF)、梯度提升树(gradient-boosted decision trees, GBDT)和CatBoost算法进行溃坝峰值流量预测,结果表明CatBoost算法具有明显的优势。通过分析各模型对典型案例的预测结果可知,CTGAN-CatBoost模型泛化能力更强,对实际工程更具有指导意义。

表 6 模型计算结果对比

研究者或 模型名称	Big Bay 大坝		板桥大坝	
	$Q_p/(m^3/s)$	相对 误差/%	$Q_p/(m^3/s)$	相对 误差/%
Kirkpatrick ^[2]	911	-78.10	6950	-91.10
SCS ^[3]	2073	-50.17	9531	-87.79
Singh 等 ^[6]	2422	-41.78	21537	-72.42
Costa ^[10]	5368	29.10	16817	-78.47
Evans ^[11]	2385	-42.67	32569	-58.30
USBR ^[12]	2114	-49.18	10966	-85.96
Froehlich ^[21]	1762	-57.64	16739	-78.57
Hooshyaripor 等 ^[19]	2907	-30.12	20714	-73.48
Froehlich ^[23]	2020	-51.44	16212	-79.24
CTGAN-RF	3701	-11.03	91336	16.95
CTGAN-GBDT	3334	-19.86	94203	20.62
CatBoost	2046	-50.82	48431	-37.99
CTGAN-CatBoost	3950	-5.05	65549	-16.07

5 结 论

a. 为解决现有模型数据量小、预测精度低的问题,基于 156 个土石坝溃坝案例建立数据库,通过溃坝过程和参数相关性分析,选取坝型、溃坝模式、溃口以上水深和溃口以上库容作为预测洪峰流量的控制变量。利用 CTGAN 算法对数据进行增强,增强后的数据与原始数据在数据分布和参数相关性方面保持相似。

b. 分别基于原始数据库和增强数据库,采用 CatBoost 算法建立土石坝溃坝峰值流量预测模型。利用评价指标 R^2 和 RMSE 评估所建模型和现有参数模型性能,结果表明基于 CTGAN 算法进行数据增强、采用 CatBoost 算法建立的 CTGAN-CatBoost 模型预测精度更高, R^2 达到 0.93。说明数据增强可以在一定程度改善模型预测性能,且 CatBoost 算法相比 RF、GBDT 算法具有明显的优势。选取 Big Bay 和板桥水库两个典型工程案例,对比分析以上模型的预测精度,得出 CTGAN-CatBoost 模型预测的相对误差最小,验证了 CTGAN-CatBoost 模型的合理性和实用性,为土石坝溃坝峰值流量的预测提供了可靠的方法。

c. 将 CTGAN-CatBoost 预测模型与 SHAP 模型相结合,分析影响土石坝溃坝峰值流量预测结果的关键特征。分析得出其特征重要性排序从大到小依次为:溃口以上库容、溃口以上水深、溃坝模式和坝型。对土石坝溃坝峰值流量影响最大的因素是溃口以上库容,其次是溃口以上水位。

参考文献:

[1] ZHONG Qiming, WANG Lin, CHEN Shengshui, et al. Breaches of embankment and landslide dams-state of the

art review [J]. Earth-Science Reviews, 2021, 216: 103597.

[2] KIRKPATRICK G W. Evaluation guidelines for spillway adequacy [C]//Proceedings Evaluation of Foundation Conference on the Evaluation of Dam Safety. Reston: ASCE,1977:395-414.

[3] SCS. Simplified dam-breach routing procedure [R]. Washington:USDA,1981.

[4] HAGEN V K. Re-evaluation of design floods and dam safety [C]//Proceedings 14th Congress of International Commission on Large Dams. Paris: International Commission on Large Dams,1982:475-491.

[5] 沈鸿杰,徐力群,刘子茜.改进 DB-IWHR 模型及其在尾矿库溃坝影响分析中的应用 [J]. 水利水电科技进展, 2022, 42 (5) : 64-70. (SHEN Hongjie, XU Liqun, LIU Ziqian. Improved DB-IWHR model and its application in influence analysis of tailings dam break [J]. Advances in Science and Technology of Water Resources, 2022, 42 (5) :64-70. (in Chinese))

[6] SINGH K P,SNORRASON A. Sensitivity of outflow peaks and flood stages to the selection of dam breach parameters and simulation models [J]. Journal of Hydrology,1984,68 (1/2/3/4) :295-310.

[7] 李相南,陈祖煜.两种溃坝模型在唐家山堰塞湖溃决模拟中的对比 [J]. 水利水电科技进展,2017,37 (2) :20-26. (LI Xiangnan,CHEN Zuyu. Comparison of two models for back analysis of dam breach of Tangjiashan Barrier Lake [J]. Advances in Science and Technology of Water Resources,2017,37 (2) :20-26. (in Chinese))

[8] 戎贵文,袁岳,肖柏青,等.局部断面收缩的溃坝水流三维数值模拟 [J]. 水利水电科技进展,2015,35 (6) :20-25. (RONG Guiwen, YUAN Yue, XIAO Boqing, et al. Three-dimensional numerical simulation of dam-break water flow with local cross-section reduction [J]. Advances in Science and Technology of Water Resources, 2015, 35 (6) :20-25. (in Chinese))

[9] MACDONALD T C, LANGRIDGE-MONOPOLIS J. Breaching characteristics of dam failures [J]. Journal of Hydraulic Engineering,1984,110 (5) :567-586.

[10] COSTA J E. Floods from dam failures [R]. Denver:USGS, 1985.

[11] EVANS S G. The maximum discharge of outburst floods caused by the breaching of man-made and natural dams [J]. Canadian Geotechnical Journal, 1986, 23 (3) : 385-387.

[12] USBR. Downstream hazard classification guidelines [R]. Denver;U. S. Department of the Interior,1988.

[13] PIERCE M W, THORNTON C I, ABT S R. Predicting peak outflow from breached embankment dams [J]. Journal of Hydrologic Engineering,2010,15 (5) :338-349.

[14] 周建银,姚仕明,王敏,等.土石坝漫顶溃决及洪水演进

- 研究进展[J]. 水科学进展, 2020, 31(2): 287-301. (ZHOU Jianyin, YAO Shiming, WANG Min, et al. Review on overtopping failure and flood evolution of earth-rock dams[J]. Advances in Water Science, 2020, 31(2): 287-301. (in Chinese))
- [15] XU Y H, ZHANG L M. Breaching parameters for earth and rockfill dams [J]. Journal of Geotechnical and Geoenvironmental Engineering, 2009, 135(12): 1957-1970.
- [16] 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.
- [17] PEKTAS A O, ERDIK T. Peak discharge prediction due to embankment dam break by using sensitivity analysis based ANN[J]. KSCE Journal of Civil Engineering, 2014, 18(6): 1868-1876.
- [18] ELMAZOGHI H G, KARAKALE V, BENTAHER L S. Comparison of neural networks and neuro-fuzzy computing techniques for prediction of peak breach outflow [J]. Journal of Hydroinformatics, 2016, 18(4): 724-740.
- [19] 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.
- [20] SHI Ning, LI Yanlong, WEN Lifeng, et al. Rapid prediction of landslide dam stability considering the missing data using XGBoost algorithm[J]. Landslides, 2022, 19(12): 2951-2963.
- [21] FROEHLICH D C. Peak outflow from breached embankment dam [J]. Journal of Water Resources Planning and Management, 1995, 121(1): 90-97.
- [22] THORNTON C I, PIERCE M W, ABT S R. Enhanced predictions for peak outflow from breached embankment dams [J]. Journal of Hydrologic Engineering, 2011, 16(1): 81-88.
- [23] FROEHLICH D C. Predicting peak discharge from gradually breached embankment dam [J]. Journal of Hydrologic Engineering, 2016, 21(11): 04016041.
- [24] 陈生水, 钟启明, 曹伟. 黏土心墙坝漫顶溃坝过程离心模型试验与数值模拟[J]. 水科学进展, 2011, 22(5): 674-679. (CHEN Shengshui, ZHONG Qiming, CAO Wei. Centrifugal model test and numerical simulation of the breaching process of clay core dams due to overtopping [J]. Advances in Water Science, 2011, 22(5): 674-679. (in Chinese))
- [25] 钟启明, 陈生水, 邓璠. 均质土坝漫顶溃坝过程数学模型研究及应用[J]. 水利学报, 2016, 47(12): 1519-1527. (ZHONG Qiming, CHEN Shengshui, DENG Zhao. Research on mathematical model for homogeneous earthen dam breach process due to overtopping failure and its application[J]. Journal of Hydraulic Engineering, 2016, 47(12): 1519-1527. (in Chinese))
- [26] 翟巍昆, 易富, 杜常博, 等. 尾矿坝漫顶溃决溃口纵向演化与侵蚀试验[J]. 水利水电科技进展, 2023, 43(5): 45-52. (ZHAI Weikun, YI Fu, DU Changbo, et al. Longitudinal evolution of breach and erosion tests for tailings dam overtopping failure[J]. Advances in Science and Technology of Water Resources, 2023, 43(5): 45-52. (in Chinese))
- [27] 钟启明, 沈光泽, 曹伟. 土石坝溃决机理模型试验研究进展[J]. 防灾减灾工程学报, 2018, 38(4): 615-623. (ZHONG Qiming, SHEN Guangze, CAO Wei. Review on the model tests for earth-rock dam breach[J]. Journal of Disaster Prevention and Mitigation Engineering, 2018, 38(4): 615-623. (in Chinese))
- [28] 段文刚, 周赤, 杨金波. 土坝漫顶冲蚀溃决过程和峰值流量试验研究[J]. 人民长江, 2013, 44(11): 76-80. (DUAN Wengang, ZHOU Chi, YANG Jinbo. Experimental research on overtopping scouring process of earth dam and peak discharge[J]. Yangtze River, 2013, 44(11): 76-80. (in Chinese))
- [29] CRESWELL A, WHITE T, DUMOULIN V, et al. Generative adversarial networks: an overview [J]. IEEE Signal Processing Magazine, 2018, 35(1): 53-65.
- [30] XU Lei, SKOULARIDOU M, CUESTA-INFANTE A, et al. Modeling tabular data using conditional gan [C]//Proceedings of the 33rd International Conference on Neural Information Processing Systems. Red Hook: Curran Associates Inc., 2019.
- [31] 梁向荣, 葛森, 王聪霞, 等. 基于 CatBoost 模型与 SHAP 分析研究地理环境因子对健康人血尿酸参考值的影响[J]. 西安交通大学学报(医学版), 2023, 44(4): 601-607. (LIANG Xiangrong, GE Miao, WANG Congxia, et al. The influence of geographical environment factors on reference value of serum uric acid in healthy people based on CatBoost model and SHAP analysis[J]. Journal of Xi'an Jiaotong University (Medical Sciences), 2023, 44(4): 601-607. (in Chinese))
- [32] LUNDBERG S M, LEE S I. A unified approach to interpreting model predictions [C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017: 4768-4777.

(收稿日期: 2024-04-07 编辑: 骆超)