

基于可解释机器学习的混凝土重力坝变形安全监控模型

程琳^{1,2}, 袁喜娜^{1,2}, 马春辉^{1,2}, 贾冬焱^{1,2}, 徐笑颜^{1,2}

(1. 西安理工大学旱区水工程生态环境全国重点实验室, 陕西 西安 710048;

2. 西安理工大学水利水电学院, 陕西 西安 710048)

摘要:针对目前基于机器学习的大坝安全监控模型无法给出模型预测解释的问题,引入 SHAP 值理论,并结合 LightGBM 模型,建立了一种具备可解释性的混凝土重力坝变形安全监控模型,且该模型可以量化每个影响因子的具体贡献。工程实例验证结果表明,该模型考虑了变形与环境量之间复杂的非线性关系,更接近真实情况,不仅具有良好的拟合精度和预测精度,还能对模型进行全局和局部的解释。

关键词:混凝土重力坝;变形安全监控;可解释机器学习;SHAP 值理论;LightGBM 模型

中图分类号:TV642

文献标志码:A

文章编号:1006-7647(2025)03-0077-09

Deformation safety monitoring model of concrete gravity dam based on interpretable machine learning//CHENG Lin^{1,2}, YUAN Xina^{1,2}, MA Chunhui^{1,2}, JIA Dongyan^{1,2}, XU Xiaoyan^{1,2} (1. State Key Laboratory of Water Engineering Ecology and Environment in Arid Area, Xi'an University of Technology, Xi'an 710048, China; 2. Institute of Water Resources and Hydro-electric Engineering, Xi'an University of Technology, Xi'an 710048, China)

Abstract: In order to solve the problem that current machine learning-based dam safety monitoring models cannot give the explanation of the model prediction, the Shapley additive explanations (SHAP) value theory was introduced, and combined with the light gradient boosting machine (LightGBM) model, an interpretable safety monitoring model for concrete gravity dam deformation was established. The model can quantify the specific contribution of each impact factor. The verification results of an engineering example show that the model considers the complex nonlinear relationship between deformation and the environmental quantities, which is closer to the real situation, and it not only has good fitting accuracy and prediction accuracy, but also can interpret the model globally and locally.

Key words: concrete gravity dam; deformation safety monitoring; interpretable machine learning; SHAP value theory; LightGBM model

混凝土坝会随着服役时间的推移出现不同程度的缺陷或病害隐患^[1-2]。因此,实行高效、安全、可靠的混凝土坝安全监控尤为重要^[3-4]。变形是混凝土坝真实结构性态的综合反映,所建数学监控模型中最为经典的是基于水压、温度和时效分量建立的模型^[5-6]。然而,大坝变形与其影响因素之间存在复杂的非线性关系,且受影响因子相关性的干扰,导致模型的预测精度难以进一步提高^[7-9]。

机器学习可以从大量监测数据中挖掘影响因子与变形量之间的隐含规律,具有自主学习、自我适应等新特征^[10],为解决上述问题提供了新思路。近年

来,随着数据量和计算能力的不断提升,机器学习在大坝安全监测领域得到了广泛应用。例如:Mata^[11]将人工神经网络(artificial neural network, ANN)算法与统计模型联合,提出了大坝安全监控的 ANN 模型;Kang 等^[12]采用极限学习机(extreme learning machine, ELM)算法对丰满重力坝的坝顶水平位移进行了监测,并与多种传统模型进行了对比,证明 ELM 模型具有更好的预测性能且训练速度较快;Wang 等^[13]提出了一种基于混合多种群遗传算法的神经网络算法,利用该混合模型对重力坝的位移进行了分析和预测;隋旭鹏等^[14]针对机器学习模型的

基金项目:国家自然科学基金项目(52479133);国家自然科学基金-黄河水科学研究联合基金项目(U2443230)

作者简介:程琳(1986—),男,副教授,博士,主要从事大坝安全智能监控、数字孪生和无损检测技术研究。E-mail:chenglin@xaut.edu.cn

通信作者:马春辉(1993—),男,讲师,博士,主要从事大坝安全智能监控研究。E-mail:machunhui@xaut.edu.cn

选择问题,对某拱坝径向位移利用支持向量机(support vector machine, SVM)、关联向量机、ELM 和多元线性回归法分别建立了预测模型,并进行了对比分析;丁炜等^[15]提出了一种基于 XGBoost 的水库大坝设施潜在风险评估预测方法。尽管机器学习在混凝土坝监控模型的预测方面具有独特的优势,但目前的研究大多是“黑盒子”模型,即可以根据输入给出准确的预测,却无法给出模型预测的解释^[16-17]。

可解释机器学习是近年来人工智能领域发展起来的一个重要研究方向,其兴起源于对高精度复杂机器学习模型透明化的需求^[18]。已有学者针对变量在预测模型中的贡献度解释问题进行了初步探索。例如:徐洪钟等^[19]利用 BP 神经网络建立了效应量与环境量关系的神经网络模型,进而确定了水位、温度等分量占效应量的比例;Paliwal 等^[20]提出了四分位间距法来解释神经网络中自变量的相对重要性,并与连接权重方法进行了比较,发现四分位间距法更优越;Olden 等^[21-22]提出了一种随机化方法,可以识别出对预测有显著贡献的变量,从而为神经网络提供了一种变量选择方法。目前可解释机器学习方法有部分依赖图^[23]、个体条件期望图^[24]、置换特征重要性、全局代理、局部代理和 SHAP 值理论。SHAP 值理论是 Lundberg 等^[25]基于博弈论提出的一种新的可解释机器学习方法,用 Shapley 值^[26]来量化每个特征的具体贡献,目前基于 SHAP 值理论的可解释机器学习已在抗震、抗压、环境等多个领域得到应用。例如:马先林等^[27]提出了一种基于机器学习的压裂水平井产能预测方法,并利用 SHAP 值理论对建立的模型进行了解释;Ekanayake 等^[28]利用 SHAP 值理论的可解释性对影响混凝土抗压强度的变量间的相互作用进行了解释;陈小昆等^[29]提出了一种融合 XGBoost 与 SHAP 值理论的冠心病预测及其特征分析模型,解决了模型性能无法满足临床应用要求、缺乏可解释性的问题。上述研究表明,SHAP 值理论不仅能从整体角度给出特征的重要性排序以及特征的具体影响规律,还可以从样本的角度给出特征的影响力和影响的正负性,有效地统一了机器学习方法的全局和局部可解释性,因而可以尝试将其应用于混凝土重力坝变形安全监控模型中。

本文提出了一种基于可解释机器学习的混凝土重力坝变形安全监控模型,将 SHAP 值理论本身的特性与训练好的模型结合起来,克服常规机器学习方法的“黑盒子”问题,并采用水口重力坝的变形监测数据对该模型进行了验证。

1 模型构建

1.1 混凝土重力坝变形安全监控模型

变形是最能直接反映混凝土重力坝运行工况的重要物理量之一^[30],已有的经验表明,混凝土重力坝位移主要受水压、温度和时效等因素的影响。因此,坝体位移 δ 的统计模型^[1]主要由水压分量 δ_H 、温度分量 δ_T 和时效分量 δ_θ 组成:

$$\delta = \delta_H + \delta_T + \delta_\theta \quad (1)$$

水压分量与大坝上下游水深的 1~3 次方有关^[31]:

$$\delta_H = \sum_{i=1}^3 [a_{1i}(H_u^i - H_{u0}^i)] + \sum_{i=1}^3 [a_{2i}(H_d^i - H_{d0}^i)] \quad (2)$$

式中: H_u 、 H_{u0} 分别为监测日、始测日对应的上游水头,即水位测值与坝底高程之差; H_d 、 H_{d0} 分别为监测日、始测日对应的下游水头,即水位测值与坝底高程之差; a_{1i} 、 a_{2i} 为水压因子回归系数。

一般情况下运行期的坝体温度基本呈准稳定温度场变化。因此,温度分量 δ_T 可采用周期项形式:

$$\delta_T = \sum_{i=1}^2 \left[b_{1i} \left(\sin \frac{2\pi it}{365} - \sin \frac{2\pi it_0}{365} \right) + b_{2i} \left(\cos \frac{2\pi it}{365} - \cos \frac{2\pi it_0}{365} \right) \right] \quad (3)$$

式中: t 为位移监测日到始测日的累计天数; t_0 为建模资料系列第一个监测日到始测日的累计天数; b_{1i} 和 b_{2i} 为温度因子回归系数。

大坝产生时效变形的原因极为复杂,它综合反映坝体混凝土与基岩的徐变、蠕变以及岩体地质构造的压缩变形等因素。位移变化的时效分量 δ_θ 为

$$\delta_\theta = c_1(\theta - \theta_0) + c_2(\ln \theta - \ln \theta_0) \quad (4)$$

其中 $\theta = t/100$ $\theta_0 = t_0/100$

式中 c_1 、 c_2 为时效因子回归系数。

1.2 SHAP 值理论

SHAP 值是一种衡量贡献度的指标,它是一类加性解释模型,将每个特征都视为贡献者,用 SHAP 值来计算其具体的贡献值,所有特征的贡献值之和即为模型的最终预测^[16],可以表示为

$$g(x') = \varphi_0 + \sum_{i=1}^M \varphi_i x'_i = f(x) \quad (5)$$

式中: $g(\cdot)$ 为解释模型预测值; $f(\cdot)$ 为机器学习方法预测值; x' 为简化输入,和输入 x 存在映射关系,取值为 1 或 0,1 表示该特征值与本解释案例中的值相同,0 表示该特征值缺失; φ_0 为机器学习方法在数据集上的预测均值,即若不知道任何特征,则模型的输出可以解释为数据集上的平均预测; φ_i 为第

i 个特征的贡献值(即 SHAP 值); M 为特征的总数。

1.3 LightGBM 模型

在如今的大样本和高维度的环境下,传统算法需要扫描所有的样本点,非常耗时,在效率和可扩展性上已不能满足需求。为此,Ke 等^[32]提出了一种梯度提升框架,即轻量级梯度提升机(light gradient boosting machine, LightGBM)。

对于回归问题,LightGBM 模型通过对所有回归树的预测值进行积分来生成最终预测结果。对于给定的数据集,在生成第一棵树后,基于最小化目标函数添加一棵新树:

$$\Gamma_k = \sum_{i=1}^n L[y_i, \hat{y}_i^{(k-1)} + f_k(x_i)] + \Omega(f_k) \quad (6)$$

式中: Γ_k 为目标函数; y_i 为第 i 个数据的实测值; $\hat{y}_i^{(k-1)}$ 为第 i 个数据在第 $k-1$ 棵树的预测值; $f_k(x_i)$ 为第 i 个数据在第 k 棵树的误差; $L(\cdot)$ 为损失函数; $\Omega(\cdot)$ 为正则化项; n 为样本数。

针对样本多的问题,LightGBM 模型采用基于梯度的单边采样算法(gradient-based one-side sampling, GOSS)对样本进行排序,保留梯度大的样本点,而对梯度小的样本点按比例进行随机采样。

针对特征多的问题,LightGBM 模型采用互斥特征捆绑算法(exclusive feature bundling, EFB):如果两个特征并不是完全互斥(部分情况下两个特征都是非零值),可以用一个指标对特征不互斥程度进行衡量,称之为冲突比率,当这个值较小时,可以选择把不完全互斥的两个特征捆绑。

此外,LightGBM 模型还采用直方图算法、按叶子生长算法来进行加速。与 XGBoost 模型相比,LightGBM 模型具有训练速度更快和内存消耗更少的优势。

1.4 模型构建步骤

利用 LightGBM 模型和 SHAP 值理论建立混凝土重力坝变形安全监测模型的流程见图 1。具体建模步骤如下:

- a. 数据收集。对实测数据进行归一化处理以及缺失值处理,并把 $H_u^1 - H_{u0}^1, H_u^2 - H_{u0}^2, H_u^3 - H_{u0}^3, H_d^1 - H_{d0}^1, H_d^2 - H_{d0}^2, H_d^3 - H_{d0}^3, \sin \frac{2\pi t}{365} - \sin \frac{2\pi t_0}{365}, \cos \frac{2\pi t}{365} - \cos \frac{2\pi t_0}{365}, \sin \frac{4\pi t}{365} - \sin \frac{4\pi t_0}{365}, \cos \frac{4\pi t}{365} - \cos \frac{4\pi t_0}{365}, \theta - \theta_0, \ln \theta - \ln \theta_0$ 作为输入,分别记为 $x_1, x_2, \dots, x_{12}$,把变形监测数据作为输出。
- b. 数据集划分。将数据集划分为训练集和测试集,划分比例一般分别取 70%和 30%。
- c. 建立模型。选择 LightGBM 模型,基于训练集进行模型训练;并采用网格搜索和十折交叉验证方法对模型进行优化。
- d. 模型评价。本次建模属于回归问题,模型的性能评级指标一般为数理统计指标,包括均方误差(MSE)、平均绝对误差(MAE)、均方根误差(RMSE)、平均绝对百分比误差(MAPE)和决定系数(R^2)等。
- e. 模型解释。利用 SHAP 值理论对所建立的模型进行解释。解释分为 2 个层面:全局层面上,可利用 SHAP 值分布图来描述特征的具体影响规律和相关性;局部层面上,可量化单一样本中每个特征的具体贡献。
- f. 模型预测。代入新数据,进行预测。

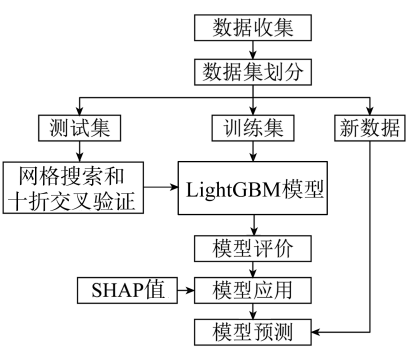


图 1 基于 LightGBM 模型和 SHAP 值理论的混凝土重力坝变形安全监控模型构建流程

- e. 模型解释。利用 SHAP 值理论对所建立的模型进行解释。解释分为 2 个层面:全局层面上,可利用 SHAP 值分布图来描述特征的具体影响规律和相关性;局部层面上,可量化单一样本中每个特征的具体贡献。
- f. 模型预测。代入新数据,进行预测。

2 工程实例验证

2.1 工程概况

水口水电站位于福建省闽江干流上,枢纽工程是以发电为主,兼有航运、过木等综合利用的大型水利工程。水口水电站枢纽由大坝、电厂、船闸、升船机等建筑物组成,设计等级为一等工程一级建筑物。水口大坝为混凝土重力坝,最大坝高 101 m,坝顶全长 783 m,坝顶高程 74 m。大坝从左往右共分为 42 个坝段。其中 1~7 号为左岸挡水坝段;8~21 号为引水坝段;23~35 号为溢流坝段;两个泄水底孔分别设在 22、36 号坝段;37、38 号为船闸及升船机上闸首坝段;39~42 号为右岸挡水坝段。

水口大坝的坝顶引张线原设计为 2 条,2002 年 2 月 9 日之后将其以 8、23、32 号坝段为界改造成 4 条,基本上每个坝段设 1 个测点,共 40 个测点,编号为 EX0~EX29、EX31~EX39、EX39-1。图 2 为大坝坝顶引张线测点布置图。

为全面分析大坝变形的情况,选取 EX06(7 号坝段)、EX10(11 号坝段)、EX17(18 号坝段)、EX21(22 号坝段)、EX27(29 号坝段)、EX31(33 号坝段)

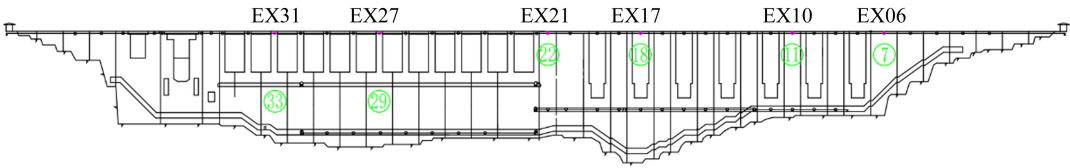


图2 水口大坝坝顶引张线测点布置

6个测点的位移进行分析。

2.2 模型拟合分析

建模选取2006年1月1日至2009年7月31日共187个数据,按照7:3的比例划分数数据集,其中训练集130个数据,测试集57个数据。以EX06测点为例,所有变量的描述性统计如表1所示。利用LightGBM模型对各测点进行拟合,并与逐步回归模型的拟合结果进行对比,结果见图3,模型拟合评价指标见表2。

由图3可知,LightGBM模型拟合值与真实值更接近,其拟合效果较逐步回归模型要好,表2的模型评价指标也印证了这一观点。

由表2计算结果可知:采用逐步回归模型拟合时, R^2 最大值为0.9399(EX27测点),最小值为0.8651(EX17测点),6个测点的 R^2 均大于0.85。采用LightGBM模型拟合时, R^2 最大值为0.9435(EX27测点),最小值为0.8839(EX17测点),6个测点的 R^2 均大于0.88,且整体上大于逐步回归模

表1 EX06测点大坝变形参数的描述性统计

变量	均值	标准差	最小值	第一四分位数	中位数	第三四分位数	最大值
x_1	0.5418	0.2288	0	0.3570	0.5867	0.7202	1
x_2	0.5297	0.2290	0	0.3429	0.5718	0.7078	1
x_3	0.5177	0.2290	0	0.3290	0.5568	0.6952	1
x_4	0.2121	0.1881	0	0.0795	0.1297	0.3315	1
x_5	0.1291	0.1502	0	0.0335	0.0587	0.1921	1
x_6	0.0775	0.1209	0	0.0125	0.0237	0.1026	1
x_7	0.4688	0.3529	0	0.1209	0.4399	0.8177	1
x_8	0.5293	0.3536	0	0.1806	0.5537	0.8777	1
x_9	0.5114	0.3578	0	0.1565	0.5172	0.8720	1
x_{10}	0.5007	0.3510	0	0.1549	0.5021	0.8482	1
x_{11}	0.5000	0.2910	0	0.2500	0.5000	0.7500	1
x_{12}	0.5287	0.2901	0	0.2844	0.5432	0.7807	1
y	2.1555	1.2291	-0.5300	1.2400	2.0900	3.0950	4.5900

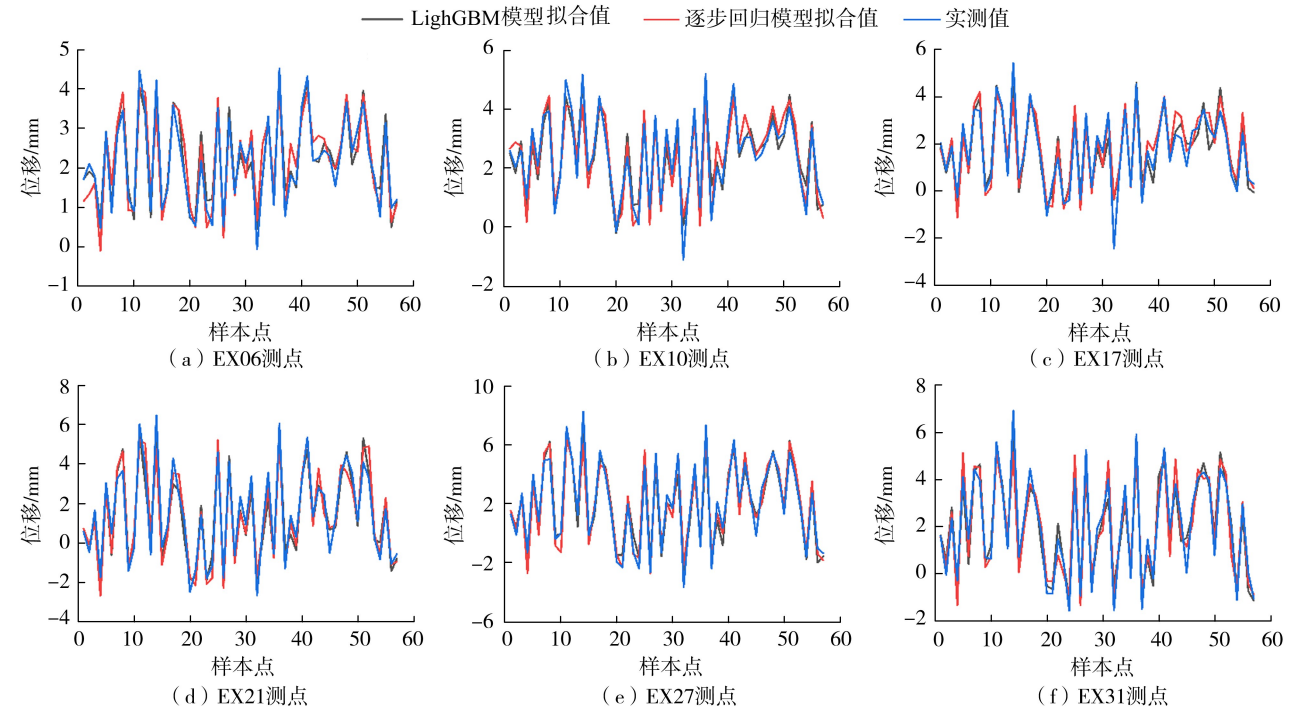


图3 各测点实测值与拟合值对比

表 2 模型拟合评价指标

测点	模型	MSE	MAE	RSME	MAPE	R ²
EX06	LightGBM	0.090 4	0.225 3	0.300 7	0.308 8	0.932 8
	逐步回归	0.172 6	0.347 5	0.415 4	0.428 0	0.872 2
EX10	LightGBM	0.179 0	0.330 4	0.423 1	0.420 5	0.913 1
	逐步回归	0.255 6	0.411 2	0.505 6	0.336 1	0.875 9
EX17	LightGBM	0.292 1	0.406 5	0.540 4	1.141 9	0.883 9
	逐步回归	0.341 2	0.458 3	0.584 1	1.644 3	0.865 1
EX21	LightGBM	0.397 3	0.519 2	0.630 3	0.461 1	0.927 0
	逐步回归	0.477 3	0.571 2	0.690 9	0.509 6	0.912 3
EX27	LightGBM	0.472 5	0.561 4	0.687 4	0.603 7	0.943 5
	逐步回归	0.502 8	0.571 9	0.709 1	1.663 8	0.939 9
EX31	LightGBM	0.340 7	0.482 2	0.583 7	1.413 6	0.926 8
	逐步回归	0.407 2	0.546 5	1.150 6	0.638 1	0.912 4

型的结果。说明 LightGBM 模型优于传统的逐步回归模型,其泛化能力更强。

2.3 模型解释

本文使用 SHAP 值理论对 LightGBM 模型进行了全局和局部解释。

2.3.1 全局解释

图 4 给出了 LightGBM 模型全局层面的解释。一个点代表一个样本,从蓝色到红色表示特征值由小到大。图 4(a)中, x_7 是代表温度分量的因子,排在第一位,其特征值越高,SHAP 值越小,表示温度是对坝顶位移影响最大的因子,其与坝顶位移呈负相关关系。对于不同的测点,因子的排列顺序也不一样,但总的来看,代表温度分量的因子 x_7 、 x_8 、 x_9 、 x_{10} 是最关键的变量,表明温度对位移的贡献最大,

其次是代表时效分量的因子 x_{11} 、 x_{12} ,而代表水压分量的因子 $x_1 \sim x_6$ 对位移的影响最小。

采用每个特征在所有样本上的 SHAP 值进行平均计算,可以量化每个影响因子的具体贡献。根据 EX06 测点的计算结果, x_7 的平均绝对 SHAP 值为 0.85,对结果的影响最大; x_{11} 、 x_1 、 x_8 、 x_4 、 x_9 、 x_{10} 的平均绝对 SHAP 值分别为 0.45、0.14、0.06、0.05、0.03、0.01; x_{12} 、 x_6 、 x_5 、 x_3 、 x_2 的平均绝对 SHAP 值为 0,对结果的影响较小,可以剔除。表 3 为采用 LightGBM 和逐步回归两种模型得到的因子选择结果,逐步回归模型基本上剔除的是 x_1 、 x_2 、 x_3 、 x_6 影响因子,而 LightGBM 模型又在上述基础上,剔除了较不敏感的 x_5 、 x_{12} 两个影响因子,可见, x_2 、 x_3 、 x_6 是对坝顶位移影响最小的因子,但不同部位的影响因子有差别。

表 4 是两种模型各分量贡献的占比。LightGBM 模型水压分量占比为 15.64%~24.25%,温度分量占比为 56.41%~70.97%,时效分量占比为 9.75%~27.95%。逐步回归模型下,水压分量占比为 19.76%~35.57%,温度分量占比为 57.22%~68.17%,时效分量占比为 0%~20.54%。二者大体上相差不大,说明采用 LightGBM 模型进行大坝位移预测是可行的。总体来看,两种模型下温度分量占比均比水压分量和时效分量占比大,说明温度变化对混凝土重力坝坝顶位移的影响比水位和时效变化对混凝土重力坝顶位移的影响大。

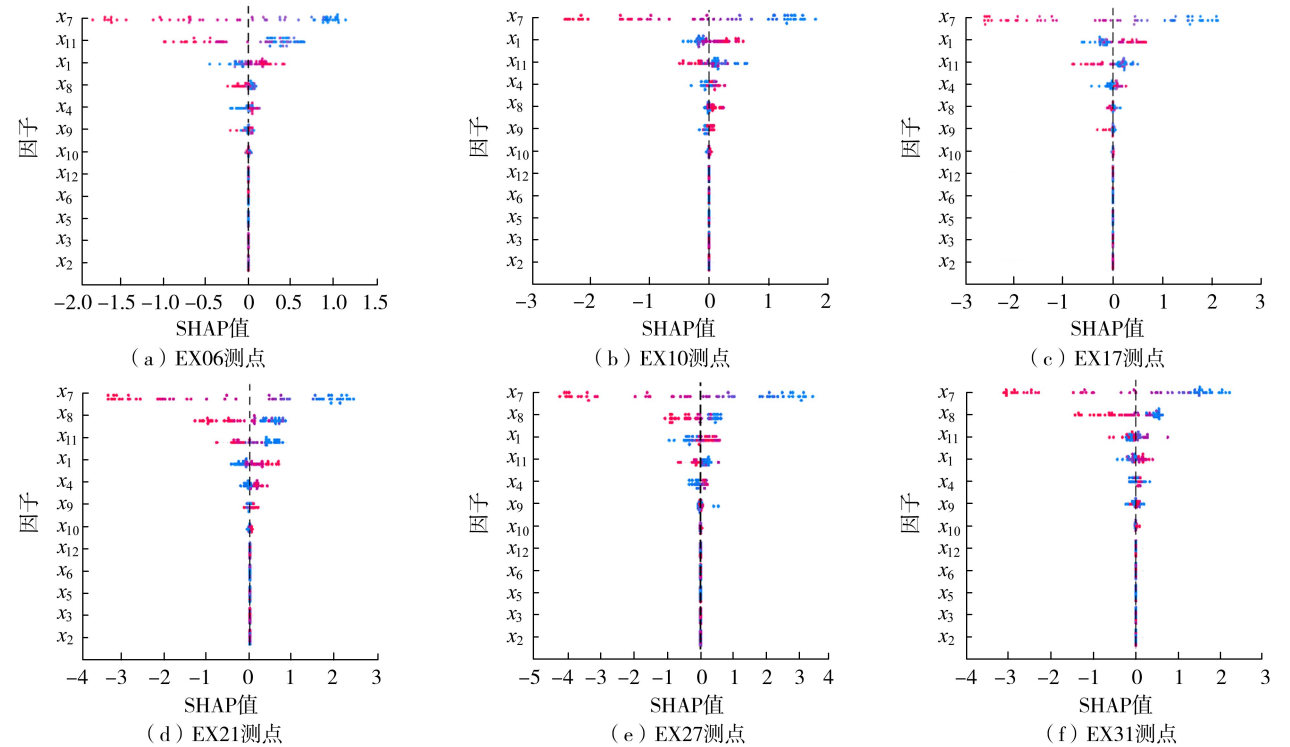


图 4 特征 SHAP 值分布

表 3 因子选择结果

测点	模型	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}
EX06	LightGBM	✓	×	×	✓	×	×	✓	✓	✓	✓	✓	×
	逐步回归	✓	×	×	✓	✓	✓	✓	✓	✓	✓	✓	✓
EX10	LightGBM	✓	×	×	✓	×	×	✓	✓	✓	✓	✓	×
	逐步回归	×	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
EX17	LightGBM	✓	×	×	✓	×	×	✓	✓	✓	✓	✓	×
	逐步回归	×	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
EX21	LightGBM	✓	×	×	✓	×	×	✓	✓	✓	✓	✓	×
	逐步回归	✓	×	✓	✓	✓	×	✓	✓	✓	✓	✓	✓
EX27	LightGBM	✓	×	×	✓	×	×	✓	✓	✓	✓	✓	×
	逐步回归	✓	×	×	✓	✓	×	✓	✓	✓	✓	✓	✓
EX31	LightGBM	✓	×	×	✓	×	×	✓	✓	✓	✓	✓	×
	逐步回归	×	×	✓	✓	✓	×	✓	✓	✓	✓	✓	✓

表 4 各分量贡献占比

测点	模型	水压分量 占比/%	温度分量 占比/%	时效分量 占比/%
EX06	LightGBM	15.64	56.41	27.95
	逐步回归	22.24	57.22	20.54
EX10	LightGBM	19.90	68.39	11.71
	逐步回归	35.57	61.21	3.22
EX17	LightGBM	24.25	65.09	10.66
	逐步回归	30.53	58.73	10.74
EX21	LightGBM	18.57	67.16	14.27
	逐步回归	21.66	68.17	10.16
EX27	LightGBM	19.28	70.97	9.75
	逐步回归	32.38	67.62	0.00
EX31	LightGBM	16.33	70.09	13.58
	逐步回归	19.76	65.86	14.38

图 5 为温度分量和水压分量 SHAP 值的分离结果统计图,其温度分量和水压分量分别为对应因子的 SHAP 值之和。从图 5 可以看出,温度分量的 SHAP 值大于水压分量的 SHAP 值;坝段中央受力复杂部位的 SHAP 值最大,两边的 SHAP 值相对较小。

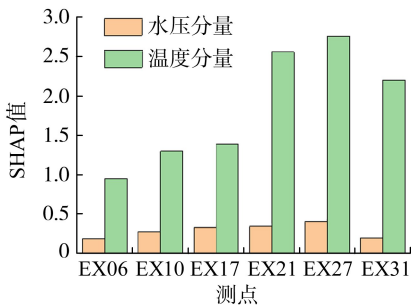


图 5 温度分量和水压分量 SHAP 值的分离结果

2.3.2 局部解释

除了对整个数据集进行全局解释之外,SHAP 值理论还为每个个体样本提供局部解释,即局部的 SHAP 值可以对某一天测值变化的成因作出解释,图 6 是 EX31 测点典型日的解释结果。在图 6(a)中,LightGBM 模型对整个数据库的平均预测值,即基准值为 2.159 mm; x_7 是最关键的影响因子,对位移有负向影响力,其 SHAP 值为-3.31, x_4 对位移有

正向影响力,其 SHAP 值为 0.17;经过各个变量的影响,最终位移结果为-0.720 mm,接近真实值-0.721 mm。

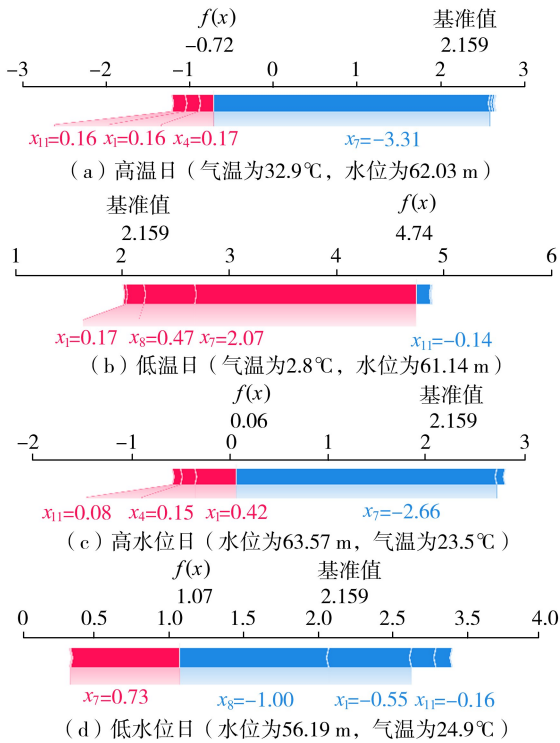


图 6 EX31 测点典型日解释结果

图 6(a)和图 6(b)是测点处于高温日和低温日的解释结果。对比分析可知, x_7 即温度因子是对位移影响最大的影响因子。不同的是,在高温日下, x_7 对位移产生负向影响力,此时,对位移影响第二大的影响因子是产生正向影响力的水压因子 x_4 ;而在低温日下, x_7 对位移产生正向影响力,对位移影响第二大的影响因子仍是温度因子,水压因子 x_1 对位移的影响排第三位。

图 6(c)和图 6(d)是测点处于高水位日和低水位日的解释结果。对比分析可知,温度因子还是对位移影响最大的影响因子。不同的是,在高水位日

下,影响最大的影响因子是产生负向影响力的 x_7 ,此时,影响第二大的影响因子是产生正向影响力的水压因子 x_1 ;而在低水位日下,影响最大的影响因子是产生负向影响力的 x_8 ,影响第二大的影响因子仍是温度因子,而水压因子 x_4 对位移的影响排第三位。

2.4 模型预测分析

以数据集的最后 10 个点作为预测点,采用 LightGBM 和逐步回归两种模型进行预测,其实测值与预测值的对比如图 7 所示,模型预测评价指标见表 5。从图 7 可以看出,LightGBM 和逐步回归模型预测的趋势与实测值的趋势大体一致,二者的预测结果都比较准确。由表 5 计算结果可知:采用逐步回归模型预测时, R^2 最大值为 0.8550 (EX27 测点),最小值为 0.8226 (EX17 测点),两个测点的 R^2 均大于 0.80。采用 LightGBM 模型预测时, R^2 最大值为 0.8714 (EX27 测点),最小值为 0.8627 (EX17 测点),两个测点的 R^2 均大于 0.85。二者虽均可以根据输入数据准确地进行预测,但整体上 LightGBM 模型的预测结果要优于逐步回归模型的预测结果。

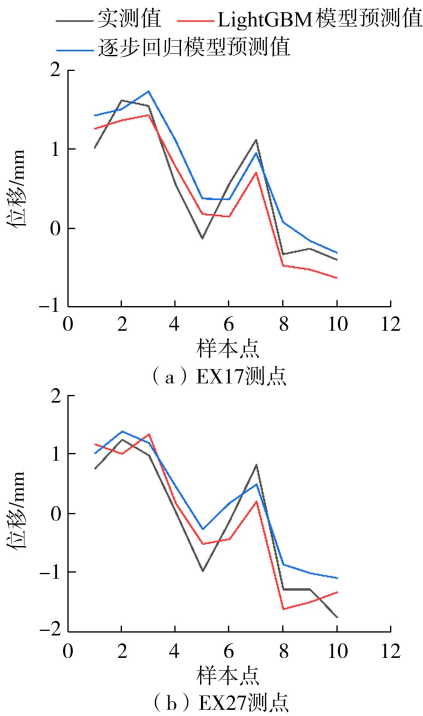


图 7 测点实测值与预测值对比

表 5 模型预测评价指标

测点	模型	MSE	MAE	RSME	MAPE	R^2
EX17	LightGBM	0.0768	0.2611	0.2772	0.6385	0.8627
	逐步回归	0.1028	0.2727	0.3207	0.7811	0.8226
EX27	LightGBM	0.1404	0.3517	0.3746	1.2805	0.8714
	逐步回归	0.1721	0.3750	0.4148	2.6596	0.8550

3 结 语

本文根据 LightGBM 模型和 SHAP 值理论,提出

了一种基于可解释机器学习的混凝土重力坝变形安全监控模型,并用水口重力坝进行了验证。验证结果表明,与逐步回归模型相比,LightGBM 模型的拟合精度更高,泛化能力更强;逐步回归和 LightGBM 模型均可以根据输入准确地给出预测结果,但整体上 LightGBM 模型的预测结果优于逐步回归模型。SHAP 值可以对模型作出全局解释。在计算各分量贡献占比和因子选择方面,LightGBM 模型和逐步回归模型具有一定的相似性,说明 LightGBM 模型结果可靠,但二者也有不同,这是因为大坝变形和环境量的映射关系本身比较复杂,传统的逐步回归模型缺乏非线性表达等问题,解释功能不完备,而 LightGBM 模型具有较强的非线性映射能力,更接近真实情况,解释的结果更准确。SHAP 值还为每个个体样本提供局部解释,即局部的 SHAP 值可以对某一天测值变化的成因作出解释。SHAP 值理论对可解释机器学习方法的全局和局部可解释性进行了统一,具有广泛的应用前景。

参考文献:

[1] 吴中如. 重大水工混凝土结构病害检测与健康诊断 [M]. 北京:高等教育出版社,2005.

[2] 赵二峰,顾冲时. 混凝土坝长效服役性态健康诊断研究述评[J]. 水力发电学报,2021,40(5):22-34. (ZHAO Erfeng,GU Chongshi. Review on health diagnosis of long-term service behaviors for concrete dams[J]. Journal of Hydroelectric Engineering, 2021, 40 (5): 22-34. (in Chinese))

[3] 吴中如,陈波. 大坝变形监控模型发展回眸[J]. 现代测绘,2016,39(5):1-3. (WU Zhongru,CHEN Bo. A review on development of dam safety monitoring models [J]. Modern Surveying and Mapping, 2016, 39 (5): 1-3. (in Chinese))

[4] 刘相杰,刘小生,张龙威. 基于二次模态分解和深度学习的大坝变形预测模型[J]. 水利水电科技进展,2024,44(3):101-106. (LIU Xiangjie, LIU Xiaosheng, ZHANG Longwei. Dam deformation prediction model based on quadratic modal decomposition and deep learning [J]. Advances in Science and Technology of Water Resources, 2024,44(3):101-106. (in Chinese))

[5] SU Huanzhi, CHEN Zhixin, WEN Zhiping. Performance improvement method of support vector machine-based model monitoring dam safety [J]. Structural Control and Health Monitoring, 2016,23(2):252-266.

[6] LI Yanlong, YIN Qiaogang, ZHANG Ye, et al. Deformation prediction model of concrete face rockfill dams based on an improved random forest model [J]. Water Science and Engineering, 2023,16(4):390-398.

[7] 李明超,任秋兵,孔锐,等. 多维复杂关联因素下的大坝

- 变形动态建模与预测分析[J]. 水利学报, 2019, 50(6):687-698. (LI Mingchao, REN Qiubing, KONG Rui, et al. Dynamic modeling and prediction analysis of dam deformation under multidimensional complex relevance[J]. Journal of Hydraulic Engineering, 2019, 50(6):687-698. (in Chinese))
- [8] 欧斌,张才溢,陈德辉,等. 基于改进 EMD-LSTM 的混凝土坝变形预测模型[J]. 水利水电科技进展, 2024, 44(6):93-99. (OU Bin, ZHANG Caiyi, CHEN Dehui, et al. Deformation prediction model for concrete dams based on improved EMD-LSTM[J]. Advances in Science and Technology of Water Resources, 2024, 44(6):93-99. (in Chinese))
- [9] 周兰庭,邓思源,柳志坤,等. 基于 SSA-LSTM-GF 的混凝土坝变形预测方法[J]. 河海大学学报(自然科学版), 2023, 51(2):73-80. (ZHOU Lanting, DENG Siyuan, LIU Zhikun, et al. Deformation prediction method of concrete dam based on SSA-LSTM-GF[J]. Journal of Hohai University (Natural Sciences), 2023, 51(2):73-80. (in Chinese))
- [10] 曾毅,刘成林,谭铁牛. 类脑智能研究的回顾与展望[J]. 计算机学报, 2016, 39(1):212-222. (ZENG Yi, LIU Chenglin, TAN Tieniu. Retrospect and outlook of Brain-inspired intelligence research[J]. Chinese Journal of Computers, 2016, 39(1):212-222. (in Chinese))
- [11] MATA J. Interpretation of concrete dam behaviour with artificial neural network and multiple linear regression models[J]. Engineering Structures, 2011, 33(3):903-910.
- [12] KANG Fei, LIU Jia, LI Junjie, et al. Concrete dam deformation prediction model for health monitoring based on extreme learning machine[J]. Structural Control and Health Monitoring, 2017, 24(10):e1997.
- [13] WANG Xiaoyu, YANG Kan, SHEN Changsong. Study on MPGA-BP of gravity dam deformation prediction[J]. Mathematical Problems in Engineering, 2017, 2017:2586107.
- [14] 隋旭鹏,王少伟,朱圣辉,等. 不同机器学习模型对混凝土坝位移预测性能的影响[J]. 水电能源科学, 2022, 40(1):107-111. (SUI Xupeng, WANG Shaowei, ZHU Shenghui, et al. Performance evaluation of different machine learning models on predicting displacement of concrete dams[J]. Water Resources and Power, 2022, 40(1):107-111. (in Chinese))
- [15] 丁炜,金有杰,张日,等. 基于 XGBoost 的水库大坝基础设施潜在风险评估预测[J]. 人民长江, 2023, 54(4):241-246. (DING Wei, JIN Youjie, ZHANG Ri, et al. Evaluation and prediction of potential risks of reservoir dam infrastructures based on XGBoost[J]. Yangtze River, 2023, 54(4):241-246. (in Chinese))
- [16] 向镇洋,包腾飞,白妍丽,等. 基于混合注意力机制和深度学习的大坝变形预测模型[J]. 水利水电科技进展, 2023, 43(2):96-101. (XIANG Zhenyang, BAO Tengfei, BAI Yanli, et al. Dam deformation prediction model based on mixed attention mechanism and deep learning[J]. Advances in Science and Technology of Water Resources, 2023, 43(2):96-101. (in Chinese))
- [17] 范向前,刘决丁,史晨雨,等. 基于人工神经网络方法的 FRP 增强混凝土断裂研究新思路[J]. 防灾减灾工程学报, 2023, 43(3):626-636. (FAN Xiangqian, LIU Jueding, SHI Chenyu, et al. Innovative idea on fracture analysis of FRP reinforced concrete using artificial neural network[J]. Journal of Disaster Prevention and Mitigation Engineering, 2023, 43(3):626-636. (in Chinese))
- [18] ANJOMSHOAE S, NAJJAR A, CALVARESI D, et al. Explainable agents and robots: results from a systematic literature review [C]//Proceedings of the 18th International Conference on Autonomous Agents and Multiagent Systems. Montréal:International Foundation for Autonomous Agents and Multiagent Systems, 2019:1078-1088.
- [19] 徐洪钟,吴中如,施斌,等. 确定大坝效应量分量比例的神经网络方法[J]. 水利学报, 2003(6):111-114. (XU Hongzhong, WU Zhongru, SHI Bin, et al. Neural network method for determining the component proportion of dam effect-variable[J]. Journal of Hydraulic Engineering, 2003(6):111-114. (in Chinese))
- [20] PALIWAL M, KUMAR U A. Assessing the contribution of variables in feed forward neural network[J]. Applied Soft Computing, 2011, 11(4):3690-3696.
- [21] OLDEN J D, JACKSON D A. Illuminating the “black box”: a randomization approach for understanding variable contributions in artificial neural networks[J]. Ecological Modelling, 2002, 154(1/2):135-150.
- [22] 郭静. 基于深度学习的异常驾驶行为检测识别[D]. 西安:长安大学, 2021.
- [23] ZHOU Jinying, PETROSIAN O L. Explainable AI: using Shapley value to explain complex anomaly detection ML-based systems[C]//Proceedings of the Machine Learning and Artificial Intelligence. Virtual Event:IOS Press, 2020:152-164.
- [24] 赵晓华,杨海益,姚莹,等. 高速公路桥梁路段交通安全风险评价及影响要素挖掘[J]. 同济大学学报(自然科学版), 2022, 50(11):1637-1646. (ZHAO Xiaohua, YANG Haiyi, YAO Ying, et al. Traffic safety evaluation and feature analysis of freeway bridge sections[J]. Journal of Tongji University (Natural Science), 2022, 50(11):1637-1646. (in Chinese))
- [25] 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.
- [26] ŠTRUMBELJ E, KONONENKO I. Explaining prediction

models and individual predictions with feature contributions [J]. Knowledge and Information Systems, 2014, 41(3): 647-665.

[27] 马先林,周德胜,蔡文斌,等. 基于可解释机器学习水平井产能预测方法[J]. 西南石油大学学报(自然科学版), 2022, 44(4): 81-90. (MA Xianlin, ZHOU Desheng, CAI Wenbin, et al. An interpretable machine learning approach to prediction horizontal well productivity [J]. Journal of Southwest Petroleum University (Science & Technology Edition), 2022, 44(4): 81-90. (in Chinese))

[28] EKANAYAKE I U, MEDDAGE D P P, RATHNAYAKE U. A novel approach to explain the black-box nature of machine learning in compressive strength predictions of concrete using Shapley additive explanations (SHAP) [J]. Case Studies in Construction Materials, 2022, 16: e01059.

[29] 陈小昆,左航旭,廖彬,等. 融合 XGBoost 与 SHAP 的冠心病预测及其特征分析模型[J]. 计算机应用研究, 2022, 39(6): 1796-1804. (CHEN Xiaokun, ZUO Hangxu, LIAO Bin, et al. Coronary artery disease prediction and feature analysis model based on XGBoost and SHAP [J]. Application Research of Computers, 2022, 39(6): 1796-1804. (in Chinese))

[30] 宋蕾,雷兆星. 基于 Inception 模块与改进 GRU 的混凝土坝变形预测模型[J]. 水利水电科技进展, 2024, 44(6): 100-105. (SONG Lei, LEI Zhaoxing. Concrete dam deformation prediction model based on Inception module and improved GRU [J]. Advances in Science and Technology of Water Resources, 2024, 44(6): 100-105. (in Chinese))

[31] 张世宝,李芳. 小湾拱坝应力监测资料分析[J]. 华北水利水电大学学报(自然科学版), 2014, 35(4): 31-34. (ZHANG Shibao, LI Fang. Analysis of stress monitoring data of Xiaowan arch dam [J]. Journal of North China University of Water Resources and Electric Power (Natural Science Edition), 2014, 35(4): 31-34. (in Chinese))

[32] KE Guolin, MENG Qi, FINLEY T, et al. LightGBM: a highly efficient gradient boosting decision tree [C]// Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017: 3149-3157.

(收稿日期: 2024 - 02 - 24 编辑: 俞云利)

(上接第 54 页)

[13] HALÁSZ, GÁBOR B. Higher order corrections for shallow-water solitary waves: elementary derivation and experiments [J]. European Journal of Physics, 2009, 30(6): 1311-1323.

[14] KOSHIZUKA S, OKA Y. Moving-particle semi-implicit method for fragmentation of incompressible fluid [J]. Nuclear Science and Engineering, 1996, 123(3): 421-434.

[15] CHIKAZAWA Y, KOSHIZUKA S, OKA Y. A particle method for elastic and visco-plastic structures and fluid-structure interactions [J]. Computational Mechanics-International Journal Then Research Journal, 2001, 27(2): 97-106.

[16] KIM K H, LEE D Y, KIM Y S. Experimental study of water impact loads on symmetric and asymmetric wedges [J]. Journal of Ocean Engineering and Technology, 2014, 28(3): 209-217.

[17] OGER G, DORING M, ALESSANDRINI B, et al. Two-dimensional SPH simulations of wedge water entries [J]. Journal of Computational Physics, 2006, 213(2): 803-822.

[18] ZHANG Youlin, TANG Zhenyuan, WAN Decheng. Simulation of water entry of a free-falling wedge by improved MPS method [C]//The 26th International Ocean and Polar Engineering Conference. Rhodes: International Society of Offshore and Polar Engineers, 2016: 220-227.

[19] TANG Zhenyuan, CHEN Xiang, WAN Decheng. Application of IMPS method in a ship bow section entering water problems [C]//The Twenty-Sixth International Ocean and Polar Engineering Conference. Rhodes: International Society of Offshore and Polar Engineers, 2016: 213-219.

[20] TANG Zhenyuan, WAN Decheng, CHEN Gang, et al. Numerical simulation of 3D violent free-surface flows by multi-resolution MPS method [J]. Journal of Ocean Engineering and Marine Energy, 2016, 2(3): 355-364.

[21] YOKOYAMA M, KUBOTA Y, KIKUCHI K, et al. Some remarks on surface conditions of solid body plunging into water with particle method [J]. Advanced Modeling and Simulation in Engineering Sciences, 2014, 1(1): 9. (DOI: 10.1186/2213-7467-1-9)

[22] SUN Zhe, DJIDJELI K, XING J T, et al. Coupled MPS-modal superposition method for 2D nonlinear fluid-structure interaction problems with free surface [J]. Journal of Fluids and Structures, 2016, 61: 29. (DOI: 10.1016/j.jfluidstructs.2015.12.002)

[23] 林鹏智,刘鑫. 光滑粒子水动力学在水利与海洋工程中的应用研究进展[J]. 水利水电科技进展, 2015, 35(5): 36-46. (LIN Pengzhi, LIU Xin. Recent advances of application of smoothed particle hydrodynamics in hydraulic and ocean engineering [J]. Advances in Science and Technology of Water Resources, 2015, 35(5): 36-46. (in Chinese))

(收稿日期: 2024 - 04 - 07 编辑: 熊水斌)