

基于协同权重的堤防堵口案例推荐系统研究

徐斌^{1,2}, 李旭¹, 刘俊国³, 关靖⁴, 黄伟⁵, 庞锐^{1,2}

(1. 大连理工大学建设工程学院, 辽宁 大连 116024; 2. 海岸和近海工程国家重点实验室, 辽宁 大连 116024; 3. 华北水利水电大学, 450046, 郑州; 4. 黄河勘测规划设计研究院有限公司, 450046, 郑州; 5. 华北水利水电大学管理与经济学院, 450046, 郑州;)

摘要: 堤防溃决后, 现有堵口方案制定方法多依赖专业技术人员临时决策, 缺乏系统化的历史案例支持, 导致方案制定的效率和准确性受限。本文构建了一种基于协同权重的堤防堵口案例推荐系统, 通过整合历史案例数据, 为抢险决策提供科学依据。通过收集近年来国内部分堤防堵口案例, 总结工程实践中影响堵口方案选择的关键特征, 引入共现水平修正的相关性分析对溃口特征进行缺失值补齐, 建立堤防堵口案例数据库。进一步引入标签编码统一语义和数值特征, 采用随机森林评估各特征对堵口方案选择的客观权重, 并建立了预案到主观相似性权重的计算流程。结果表明, 道路条件在堵口方案选择中起主导作用, 以其为侧重的随机森林和协同权重组合在相似性匹配模型中的平均准确率最大值高 10~25%, 检索出的方案在抢险救援现场有一定的参考价值。

关键词: 堤防溃口; 堵口方案; 缺失值插补; 协同权重; 案例推理

中图分类号: TV871.3

文献标志码: J

Case Recommendation System for Levee Breach Closure Based on Collaborative Weights

Xu Bin^{1,2}, Li Xu¹, Liu Junguo³, Guan Jing⁴, Huang Wei⁵, Pang Rui^{1,2}

(1. School of Infrastructure Engineering, Dalian University of Technology, Dalian, 116024, China;

2. State Key Laboratory of Coastal and Offshore Engineering, Dalian University of Technology, Dalian, 116024, China;

3. North China University of Water Resources and Electric Power, Zhengzhou, 450046, China;

4. Yello River Engineering Consulting Co., Ltd.;

5. School of Management and Economics, North China University of Water Resources and Electric Power, Zhengzhou, 450046, China)

Abstract: Existing methods for levee breach closure planning predominantly rely on on-site decisions by technical personnel, lacking systematic support from historical cases, which compromises both efficiency and accuracy in emergency response. To address this limitation, this study proposes a collaborative weight-based case recommendation system for levee breach closure, integrating historical case data to provide a scientific foundation for decision-making in disaster mitigation. This paper collects domestic cases of levee breach closures from recent years and summarizes the key features in the selection of closure plans in engineering practice. By employing correlation analysis modified by co-occurrence levels, missing values in breach characteristics are completed, and a database of levee breach closure cases is established. Furthermore, label encoding is introduced to unify semantic and numerical features, and random forest algorithm is used to assess their objective weights in the selection of breach closure plans. A computational process for determining the subjective similarity weights of contingency plans is also established. The results indicate that the feature of road conditions plays a dominant role in the selection of breach closure plans. The random forest model, when focused on this feature and combined with

收稿日期: 2025-04-06

接受日期: 2025-05-20

基金项目: 国家重点研发计划项目 (2023YFC3011400); 国家自然科学基金 (52479094)

作者简介: 徐斌 (1981—), 男, 教授. E-mail: xubin@dlut.edu.cn

通信作者: 庞锐 (1989—), 男, 副教授. E-mail: pangrui@dlut.edu.cn

collaborative weights, achieves an average accuracy improvement of 10-25% in the similarity matching model. The retrieved plans provide valuable reference points during emergency rescue operations on-site.

Keywords: levee breach; breach repairing schemes; missing values imputation; collaborative weights; case-based reasoning

0 引言

我国地域辽阔,地形复杂,年降雨量集中,洪水灾害历来都是中华民族的心腹之患。洪涝灾害导致的直接经济损失占整个气象灾害损失的 39.5%,死亡人数超过总数的一半,是对我国社会经济影响最严重的自然灾害之一^[1,2]。堤防作为保护人民群众工农业生产免遭洪涝灾害影响的重要设施,截至 2020 年年底,全国已建成 5 级及以上江河堤防 32.8 万公里^[3],保护其免遭洪水冲毁和堤防溃决后迅速组织抢险工作,是守住安全度汛底线的关键,符合国计民生的重大需求。

堤防溃决后,通常由地方政府和防汛部门启动应急响应机制,组织抢险救援工作,其中迅速勘测现场条件,合理规划抢险方案是堵口工作的必要支撑。现有抢险技术通常采取在溃口现场建立一线指挥部,专业抢险技术人员在了解溃口口门情况后,临时组织制定方案^[4,5]。近年来,随着知识工程技术的发展,在堤防溃决抢险领域开展专业知识文献、应急预案资料和历史堵复案例的数字化应用,是提升防汛抢险工作效率的可行手段之一。

案例推理 (Case-based Reasoning) 的框架由 Roger Schank 在 1982 年首次提出^[6],并由 Kolodner 构建并完善了名为 CTRUS 的 CBR 系统^[7],其开创了问题事例表示、事例检索和事例重用等程序,为解决实际问题提供了可靠的方法。CBR 技术在故障诊断、预算决策、设计规划等领域已有了初具规模的应用,例如 Chang^[8]建立了零件设计案例库,使用最邻近检索相似案例,为设计者提供有益参考。Mohammed 等^[9]将遗传算法和 CBR 系统结合提出了一种设备故障的诊断方法,检测准确率达到 98.7%。冯为民等^[10]基于案例模糊推理提出了一种土木工程造价估算方法,为项目预算决策提供了有

价值的参考依据。沈扬^[11]等基于知识图谱构建了面向混凝土坝施工管理的案例库,实现了施工知识的快速提取和利用。但在堤防溃决抢险领域,利用以往案例建立规范案例库的研究较少,难以从历史经验中为新发险情抢险提供有益指导。

本文搜集历年来堤防溃决及堵复的案例文本,抽取关键特征,通过相关性分析对缺失值进行拟合补齐,建立了堤防堵口案例数据库。通过引入随机森林和专家权重相结合的协同权重,采用最邻近匹配从案例库中筛选相似案例,为堤防堵口现场方案制定提供历史案例参考。

1 分析计算方法

1.1 系统框架

本文从科研文献、学术书籍、网络报道等渠道搜集了自 1980 年以来我国主要的堤防堵口案例,结合部分溃决试验成果数据,建立堤防堵口案例数据库。其中的实发溃决案例往往由于情形紧急、设备条件、整理欠缺等因素,存在普遍的特征值缺失问题。因此采用相关性分析方法,引入参数共现水平以修正案例各维度特征间的相关系数,确定了由完整性较好的特征向较差特征间的补齐流程并进行拟合插值,从而得到完整案例数据库的流程。以堵口方案为目标,采用随机森林方法计算溃口各维度特征值对于堵口方案选择的客观权重,采用多种方法对比评估专家主观评价影响下的主观权重,使用二者结合得到的协同权重作为实发溃决案例与案例库最邻近匹配时的权重,通过相似度计算得到最优的历史案例进行推荐。本文案例推荐系统流程如图 1 所示:

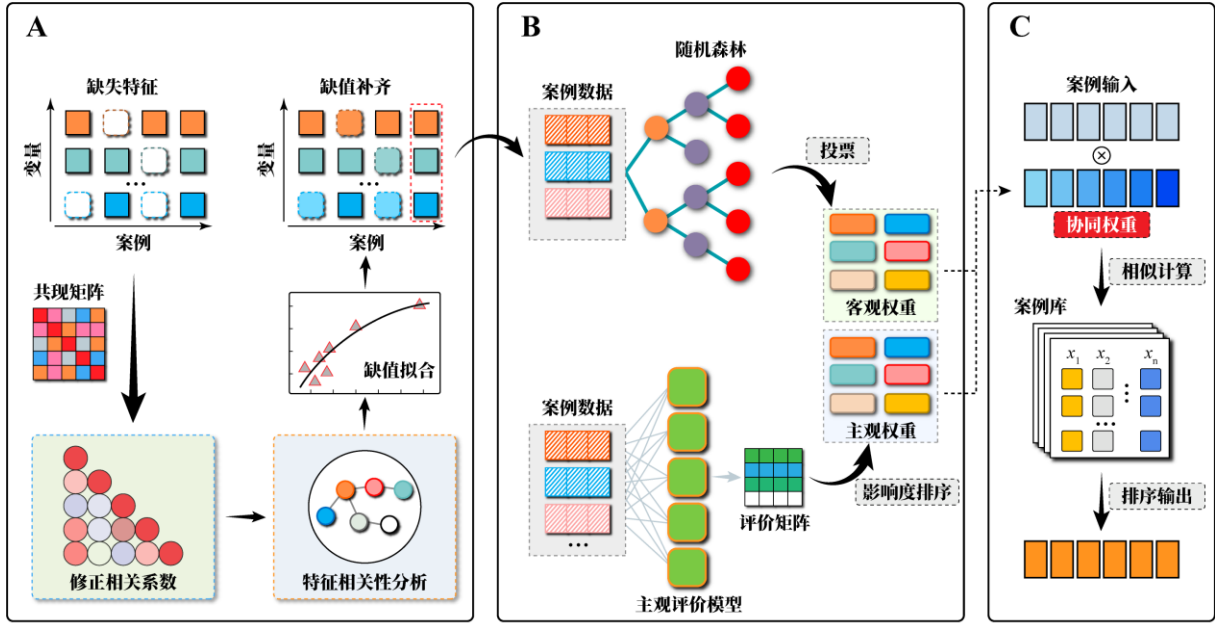


图 1 堤防堵口案例推荐系统流程示意图

Fig.1 Example process flowchart of the case recommendation system for levee breach closure

1.2 算法原理

1.2.1 基于相关性分析的案例缺失特征值插补方法

总结堤防堵口案例特点，主要涉及到的参数和信息有溃口宽度、溃口深度、口门流速、口门流量、堵口物料总量、复堤时间和道路情况等。由于现场救援工作的紧急性，其中部分参数通常无法及时准确地测量记录，属于完全随机缺失（Missing Completely at Random, MCAR）类型，因此需通过其他特征分布估计缺失值的可能取值。通常采用的插补方法包括单一插补、多重插补和 KNN 插补等^[12-14]，考虑到堤防堵口案例数据库容量偏小且不同维度特征间存在的相关关系，本文主要采用单一插补中的回归插补处理缺失值，同时以归一化均方根误差（Normalized Root Mean Squared Error, NRMSE）和归一化平均绝对误差（Normalized Mean Absolute Error, NMAE）为评估指标，对比分析了 KNN 插补和基于链式方程的多重插补的结果以评估补值性能^[15]。

（1）回归插补方法

以皮尔逊相关系数^[16]作为特征间相关关系度量标准，计算公式为：

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (1)$$

式中： n 为样本总数， X_i 和 Y_i 分别指代将要计算的两维特征的变量值。

但由于缺失值的存在，如果一对特征序列 (X_{pi}, X_{qi}) ($i=1, 2, \dots, n$) 所包含的非空数据过少，也会使特征间表现出极强的线性相关性^[17, 18]，因此提出一种基于共现水平对相关系数进行修正的方法：

对于数据集 D_{ij} ($i=1, 2, \dots, n; j=1, 2, \dots, m$)，其中 n 和 m 分别为数据集的大小和维度，非空数据标识矩阵 D' 定义为：

$$D'_{ij} = \begin{cases} 0, & \text{if } D_{ij} \text{ is NaN} \\ 1, & \text{otherwise} \end{cases} \quad (2)$$

则共现水平矩阵为：

$$C = D'^T g D' \quad (3)$$

对于特征 p, q ，如果共现水平 C_{pq} 低于一定阈值 s_{\min} ，则对相关系数 r_{pq} 施加修正系数，降低相关系数对此类数据点的敏感程度以增强鲁棒性。修正后的皮尔逊相关系数为：

$$r'_{pq} = r_{pq} \operatorname{gmin} \left\{ 1, \frac{C_{pq}}{s_{\min}} \right\} \quad (4)$$

当有效数据仅占原始数据 30% 时, 插值效果出现下降^[19], 考虑到堤防溃口案例样本规模较小, 阈值 s_{\min} 设置为样本总数的 20%。根据修正相关系数的大小排序特征相关性, 由缺失水平较低的特征开始向其他维度特征逐渐进行数据拟合补齐缺失值。

回归分析中拟合模型的确定不仅包括变量选择, 还涉及函数形式的选择。评价所选函数在数据集上拟合优度的一个指标是残差, 定义为 $e_i = y_i - \hat{y}_i$, 其中 y_i 和 \hat{y}_i 分别是第 i 个数据的观测值和预测值。基于最小二乘法, 通过最小化残差平方和来获得回归系数 b_i 的估计值:

$$\begin{aligned} \min_{b_0, b_1, \dots, b_m} \sum_{i=1}^n e_i^2 &= \sum_{i=1}^n (y_i - \hat{y}_i)^2 \\ &= \sum_{i=1}^n [y_i - f(x_i | b_0, b_1, \dots, b_m)]^2 \end{aligned} \quad (5)$$

特大规模溃口案例的部分特征值显著偏离主流溃口分布, 但在防汛抢险领域具有特殊意义^[20]。为避免其对数据库常规拟合关系的影响, 对该部分非常规溃口数据单独建模, 引入 Huber 损失函数对插值拟合模型进行修正^[21]:

$$L_\delta(e) = \begin{cases} \frac{1}{2}e^2, & \text{if } |e| \leq \delta \\ \delta \left(|e| - \frac{1}{2}\delta \right), & \text{if } |e| > \delta \end{cases} \quad (6)$$

式中: e 为残差; δ 为阈值参数。当预测误差小于 δ 时, 仍采用残差平方和估计模型参数, 否则采用绝对误差修正模型, 增强模型对于此类案例的泛化能力。

考虑到堤防溃决堵复案例特征间关系可能是线性或非线性, 拟合时用时采用一次和二次两种模型进行拟合回归, 引入三种评价标准对不同模型的拟合效果进行评估: 调整 R^2 、AICc 信息准则和区间均方根误差, 从而选取更合适的模型进行拟合补值。

① 调整 R^2

对于线性回归, 通常采取决定系数 R^2 来表征因变量 Y 的变异中能够被自变量 X 所解释的程度^[22]:

$$R^2 = 1 - \frac{\sum_i (\hat{y}_i - y_i)^2}{\sum_i (\bar{y}_i - y_i)^2} \quad (7)$$

对于二次曲线回归, 相当于向模型中增加了一个自变量, 为了考察新增二次项是否会增加模型的拟合度, 因此采用调整 R^2 来评价模型的拟合优度^[23]:

$$R_{adj}^2 = 1 - \left[\frac{(1 - R^2)(n - 1)}{n - k - 1} \right] \quad (7)$$

式中: n 为数据集大小, k 为拟合模型使用的参数个数。如果额外的参数不能显著改善模型解释数据变化的能力, 则 k 个参数的 R_{adj}^2 值小于 $k-1$ 个变量的。

因此, 通过比较 R_{adj}^2 值, 可以选择一个只涉及少量参数的简化预测模型。

② AICc 信息准则

为了权衡模型估计的复杂程度和拟合优度, 赤池弘次^[24]提出一种信息熵计算方法, 赤池信息量准则 (Akaike Information Criterion, AIC), 用似然函数代表拟合效果, 同时考虑参数数量对模型简洁性的负面影响, 其计算公式为:

$$AIC = 2k - 2\ln(L) \quad (8)$$

式中: k 为模型自由度, L 为模型的对数似然函数。在样本容量较小时, 模型选择应该更加侧重与对变量个数的惩罚程度, 趋向于选择参数更少的模型。修正赤池信息准则 (Corrected Akaike Information Criterion, AICc) 基于参数估计的期望值, 适用于数据量偏少的样本^[25]:

$$AICc = AIC + \frac{2k(k+1)}{n-k-1} \quad (9)$$

AICc 值越小, 说明模型拟合更能兼顾拟合效果和简洁性。考虑到堤防溃决案例的样本容量偏小, 因此本文采用 AICc 准则作为模型选择的评价标准之一。

③ 区间均方根误差

堤防堵口案例特征值受溃口规模影响, 口门宽深、流速、流量等参数取值在一定区间内分布, 计算样本总体均方根误差不能很好地反映拟合模型对于各尺度溃口的拟合效果。因此将横轴特征划分

为若干个区间，计算每个区间内的均方根误差，以对模型的拟合效果进行比选。

(2) KNN 及其补值算法

K 最近邻算法 (K-nearest neighbor, KNN) 通过计算输入样本各维度特征与所有已知样本的距离进行相似度排序^[26]，采用加权欧氏距离 L_E 进行相似度计算：

$$L_E = \sqrt{\sum_{i=1}^m w_i (x_i - X_i)^2} \quad (10)$$

式中： m 为特征数量； w_i 为特征对应的权重； x_i 和 X_i 分别为输入样本和已知样本的特征参数。加权欧氏距离通过调整各特征的权重来反映特征对相似性的贡献程度差异，展现出良好的有效性和灵活性^[27, 28]。

KNN 补值算法通过筛选出距离最小的 K 个已知样本构成相邻集 $N_K(X) = \{Y_1, Y_2, \dots, Y_K\}$ ，以其均值作为插补值。

(3) 基于链式方程的多重插补算法

基于多重插补链式方程 (Multiple Imputation by Chained Equations, MICE)^[29] 的补值算法，首先由简易插补方法如均值或中位数初步填补缺失值，得到初始完整数据集 M_0 。对于其中有缺失值的特征 X_j ，将其作为目标预测变量，其余特征 X_j 作为线性回归模型的自变量进行插补，对所有特征重复上述过程，完成一次插补，重复 p 次，模拟多种填补结果的采样分布，生成预测数据集 $\{M_1, M_2, \dots, M_p\}$ ，使用其均值作为缺失值估计值：

$$\hat{X}_j^{miss} = \frac{1}{p} \sum_{i=1}^p X_j^{miss(i)} \quad (11)$$

1.2.2 基于协同权重的案例匹配方法

堤防堵口案例数据库中，最终采用的堵口方案是由溃口各特征参数共同决定的。分析案例库的历史数据，可以得到各特征对于堵口方案选择的客观权重。专家根据现场救援要求，给出溃口各特征对于抢险工作的重要程度，进而计算得到主观权重。

(1) 客观权重计算

随机森林 (Random Forest, RF) 算法通过构建多颗决策树，计算使不纯度降低最明显的节点特征进行数据集分割，对所有树的输出进行投票来给出预测结果，同时得到各维特征对于数据分类的重要

度，具有良好的抗噪能力和较高的泛化性能^[30, 31]。本文采用基尼指数 (Gini Impurity) 计算根据某维特征分类后数据集的不纯度减少来确定各特征的权重，基尼指数表示样本集合 D 中一个选中的随机样本被分类错误的概率，计算公式^[32]为：

$$Gini(D) = \sum_{k=1}^K p_k (1 - p_k) \quad (12)$$

式中： K 为数据标签类别数， p_k 为类别 k 在数据集 D 中的比例。基尼指数取值在 $[0, 1]$ 之间，值越小表示数据集纯度越高。

按照特征 A 对数据集进行二分类后得到子集 D_1 、 D_2 ，此时数据集的基尼指数为：

$$Gini(D, A) = \frac{|D_1|}{|D|} Gini(D_1) + \frac{|D_2|}{|D|} Gini(D_2) \quad (13)$$

进而得到以 A 为分割特征的基尼不纯度减少值，作为特征在第 k 颗决策树上的重要度 $I_A^{(k)}$ ：

$$I_A^{(k)} = |Gini(D) - Gini(D, A)| \quad (14)$$

将所有特征的重要度进行归一化得到客观权重 w_A ：

$$w_A = \frac{\sum_{k=1}^T I_A^{(k)}}{\sum_{j=1}^m \sum_{k=1}^T I_A^{(k)}} \quad (15)$$

式中， m 为特征总数， T 为参与构建随机森林的决策树数量。

(2) 主观权重计算

各特征值的主观权重可以采用层次分析法 (Analytic Hierarchy Process, AHP)^[33]、DEMATEL 法 (Decision-making Trial and Evaluation Laboratory)^[34] 和 Shapley 法^[35] 进行计算。针对堤防堵口案例，通过对各特征要素进行两两比较，确定各要素在堵口方案选择中的主观权重。

① 层次分析法

首先构建准则层评价矩阵 $K = (k_{ij})_{n \times n}$ ，根据经验给出重要性标度；计算评价矩阵的最大特征值 λ_{max} 对应的特征向量 ξ_0 ，并计算对应的一致性指标

$CI = (\lambda_{max} - n)(n-1)$ ；根据随机一致性指标 RI 计算一致性比率 $CR = CI / RI$ ；若 $CR < 0.1$ ，表明评价矩阵 K 的一致性程度在容许范围内，采用特征向量 ξ_0 作为主观权重 w_B 参与相似度计算。其中随机一致性指标的取值如表 1 所示：

表 1 1~6 阶随机一致性指标取值表
Table 1 Random consistency index table for 1st to 6th order

矩阵阶数	1	2	3	4	5	6
RI	0	0	0.58	0.90	1.12	1.24

② DEMATEL 法

收集专家评分，构造直接影响矩阵 $E = (e_{ij})_{n \times n}$ ，归一化后计算总影响矩阵 $T = (t_{ij})_{n \times n} = E(1 - E)^{-1}$ ，并基于 T 计算单一特征的综合影响度，归一化后得到某特征的主观权重^[34]：

$$w_{B,m} = \frac{\sum_{j=1}^n t_{ij} + \sum_{i=1}^n t_{ij}}{\sum_{m=1}^n (\sum_{j=1}^n t_{ij} + \sum_{i=1}^n t_{ij})} \quad (16)$$

③ Shapley 法

假设因素集合 N 构成的指标评价矩阵 K ，函数 $v(S)$ 表示因素子集 $S \subseteq N$ 对于方案选择的总贡献，本文定义为 K 中对应因素的行和。每个因素的 Shapley 值为^[35]：

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|! g(|N| - |S| - 1)!}{|N|!} g[v(S \cup \{i\}) - v(S)] \quad (17)$$

式中： S 为不含因素 i 的所有子集； $|S|$ 是子集 S 中的元素数量； $|N|$ 为总因素数量， $[v(S \cup \{i\}) - v(S)]$ 是因素 i 对子集 S 的边际贡献。

对每个因素的 Shapley 值进行归一化得到主观权重 w_B 。

(3) 协同权重计算

在案例推理时采用随机森林方法与主观经验结合以应对复杂工程案例，相比较于单一评价指标体系在相似度计算时的准确性更高^[36]。本文采用协同权重 $W = (w_1, w_2, \dots, w_n)^T$ 融合特征客观贡献与主观判断，由主客观权重加权计算得到^[37]：

$$W = \eta w_A + (1 - \eta) w_B \quad (18)$$

式中： η 表示实际工程中考虑抢险现场情形的侧重因子，在 $[0, 1]$ 范围取值。随机森林通过特征重要性计算量化溃口特征对方案选择的贡献，同时主观权重反映现场对方案制定的实际需求。两种权重的耦合以数据库案例的客观性为基础，叠加主观性判断，实现多维度信息融合，提高权重结构的解释性和可靠性^[38]。若现场情况紧急，需要匹配物资调度和道路水平等关键因素，则更加侧重于主观判断；若情况总体可控，则更加侧重于匹配到溃口参数更接近的历史案例。协同权重结合了溃口特征对于抢险方案选择的客观影响和主观考量，相比较于单一评价体系更接近实际情况^[39]。

对于实发险情，根据式(10)计算协同权重参与下与数据库中案例的相似程度，进行排序推荐。

2 堤防堵口案例数据库构建

2.1 数据收集及特征相关性分析

收集 1980 年来我国 60 组堤防堵口实际案例，总体情况如图 2 所示。从实发险情地域分布看出，河南、湖北、湖南、江西四省受灾最频繁，且溃口规模较大。经统计，溃口平均宽度 90.85m，宽度 100m 内的溃口数量为 41 处，险情主要集中在长江黄河流域。

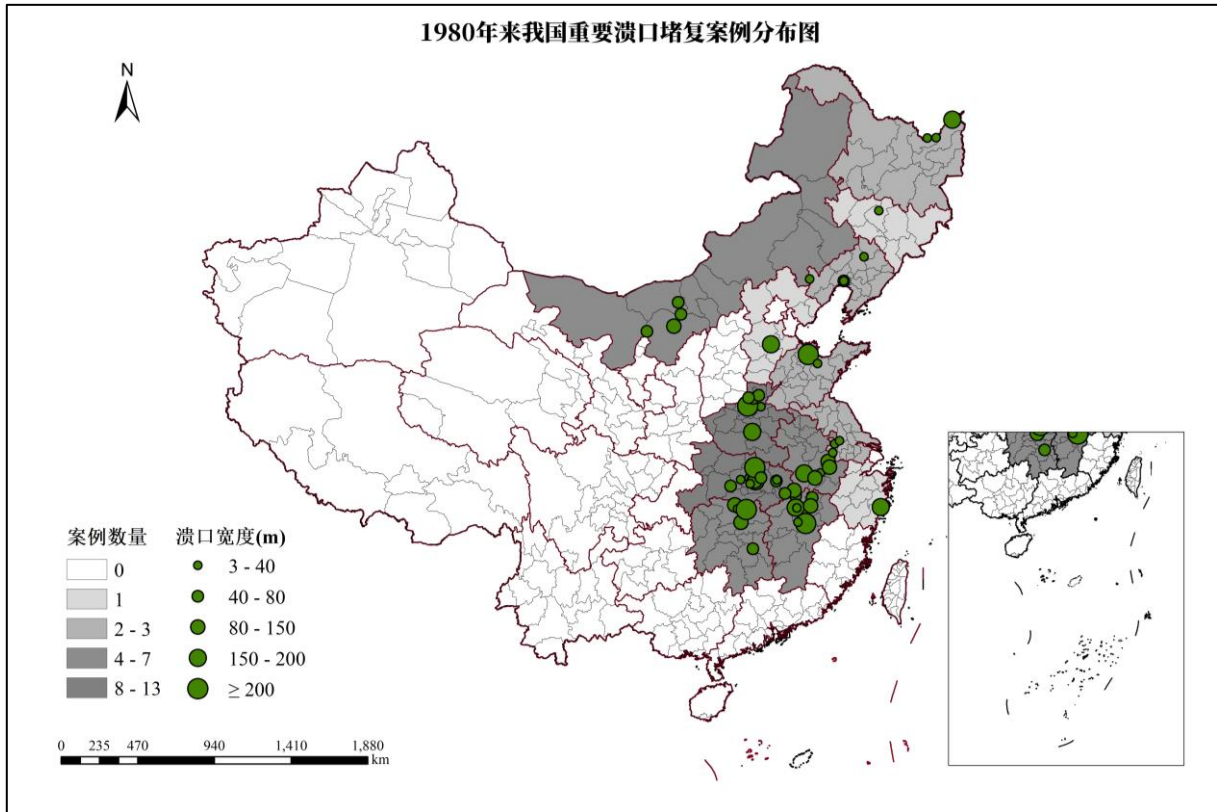


图2 1980年来我国重要溃口堵复案例分布图

Fig.2 Map of significant embankment breach closure cases in China since 1980

总结堤防溃口记录的特点,通常以溃口宽度 B 、溃口口门深度 H 、流速 v 、流量 Q 作为衡量溃口规模大小的典型特征^[40-42]。由于受灾地方各方部门协调调度能力和险情响应缓急要求不同,所需要的堵口物料总量 E_v 和最小复堤时间 T 也都呈现出明显的案例特点。此外,现场道路情况 R_c 决定能够采用的施工机械,同样影响方案决策。因此在案例匹配时,特别关注以上特征。本文采取的堤防堵口案例特征如图3所示。

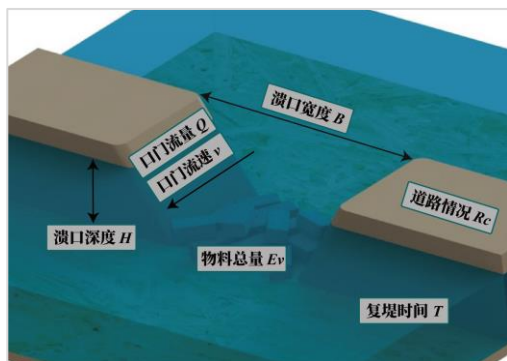


图3 堤防堵口案例特征示意图

Fig.3 Schematic diagram of characteristics of levee breach closure cases

考虑到实发溃口案例可能因缺少设备、测量条件差、整理遗漏等原因造成特征数据部分缺失甚至大部分缺失。为了提高数据的完整性,基于已有数据对缺失数值型数据进行拟合插值。通过分析特征维度之间的相关性,以记录质量较高或数据完整性较好的特征为基准进行插补,引入拟合效果评估准则,筛选适用的插值模型。同时,增加一部分模型试验的数据作为溃口特征值相关性分析及拟合插值样本集的补充^[43-45],以提高模型泛用性。

针对一个特定的案例数据 X_i ,若某两维特征值 X_{ip} 和 X_{iq} 均未缺失,则构成一对有效数据对参与拟合,反之则将其作为补值的对象。在整个案例数据库层级上,若某两维特征 X_q 和 X_p 存在的有效数据对过少,则小样本可能导致线性关系估计的假阳性。因此基于共现水平,对有效数据对数量小于样本总数 20% 的特征相关系数进行修正。对搜集到 119 组的堤防溃口特征数据进行分析,共现矩阵如表 2 所示。

表 2 堤防溃口案例数据特征共现水平矩阵
Table 2 Co-occurrence matrix of levee breach case data characteristics

案例特征		B	H	v	Q	Ev	T
共现水平	B	110	64	51	45	22	43
	H	64	73	29	29	17	28
	v	51	29	51	35	6	10
	Q	45	29	35	52	8	7
	Ev	22	17	6	8	24	18
	T	43	28	10	7	18	44

其中溃口宽度 *B* 数据记录情况最好, 缺失值仅占案例总数的 4.5%, 这是因为抢险现场通常可以通过目视测距、直接测量、无人机观测等手段快速获取溃口宽度数据; 而溃口深度 *H*、口门流速 *v*、口门流量 *Q* 的缺失可能由于没有直接测量设备或不具备测量条件而疏于记录; 堵口物料总量 *Ev* 和复堤时间 *T* 往往因为救灾工作结束后总结工作不到位而存在更多缺失值, 分别占案例总数的 79% 和 62%。但以上特征直接关系到堵口方法的选择, 因此进行缺失值插补工作具有积极意义。

基于式(1)~(4)计算修正后的特征相关系数, 如图 4 所示:

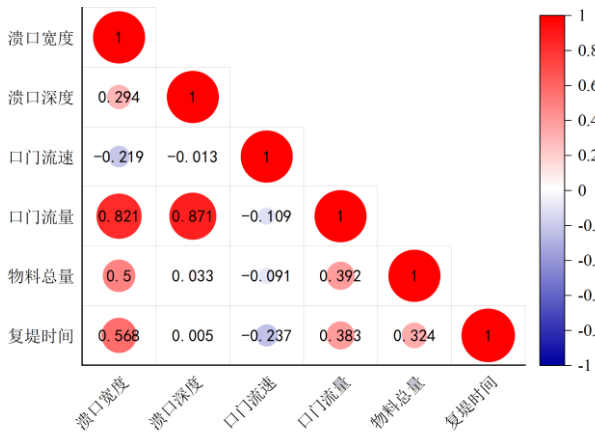


图 4 堤防溃口案例数据特征修正相关系数

Fig.4 Adjusted correlation coefficient of levee breach case data characteristics

分析特征间相关系数, 口门流量 *Q* 与溃口深度 *H*、溃口宽度 *B* 间的线性相关系数最显著, 相关系数分别为 0.871 和 0.821; 其次溃口宽度 *H* 与复堤时间 *T*、物料总量 *Ev* 间分别存在 0.568 和 0.50 的相关系数。因此首先使用溃口宽度数据拟合口门流量、物料总量和复堤时间, 进而使用口门流量拟合口门溃口深度数据。

口门流速 *v* 与溃口其他单一特征参数间不存在明显的相关关系, 国内外学者通常使用非线性或多参数模型, 使用其他特征值计算得到^[46-48]。对于沿堤防轴线纵截面为梯形的一般溃口, 其横截面形状和尺寸可由溃口深度 *H_b*、溃口顶宽 *B_t*、溃口均宽 *B_{ave}*、溃口底宽 *B_b* 和边坡系数 *z* 中的任意三个确定^[49], 则口门流速可由口门断面流量和面积由 $v = Q / A$ 计算得到, 其中 $A = \left(B_t - \frac{1}{2} H_b z \right) H_b$ 。

2.2 拟合分析及模型选取

按照 2.1 节确定的拟合补值顺序, 对溃口宽度-口门流量、溃口宽度-复堤时间、溃口宽度-物料总量和口门流量-溃口深度四对表现出显著或一定相关性的特征对进行拟合回归分析, 根据参数特点使用线性和二次两种模型对有效数据集进行拟合插值。

(1) 溃口宽度 *B*-口门流量 *Q* 拟合

由于部分案例的口门流量数值偏离常规规律, 为避免其对数据库常规拟合关系的影响, 引入 Huber 损失函数^[21]对插值拟合模型进行修正。其中损失阈值 δ 对于两种模型拟合效果的影响如图 5 和图 6 所示:

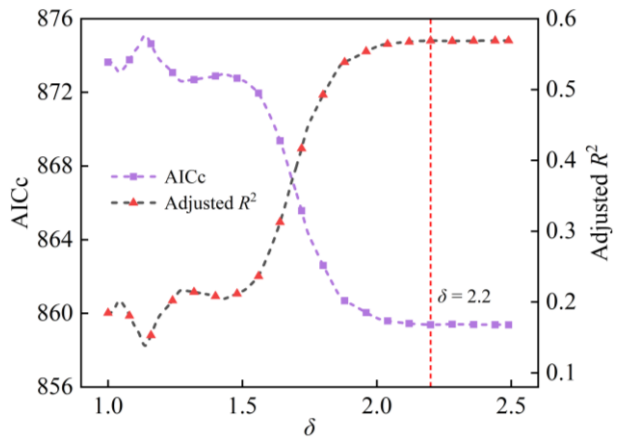


图 5 基于 Huber 损失函数的溃口宽度 *B*-口门流量 *Q* 线性拟合 δ 参数优化

Fig.5 Optimization of δ Parameters for Linear Fitting of *B*-*Q* Based on Huber Loss Function

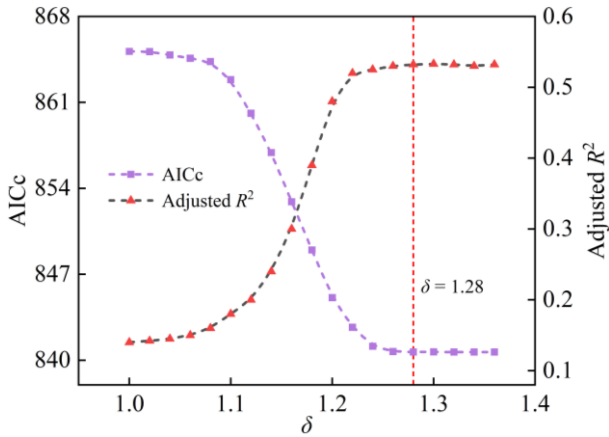


图 6 基于 Huber 损失函数的溃口宽度 B -口门流量 Q 二次拟合 δ 参数优化

Fig.6 Optimization of δ Parameters for Quadratic Fitting of B - Q Based on Huber Loss Function

确定线性和二次模型的 Huber 损失阈值 δ 分别为 2.2 和 1.28。拟合结果和区间均方根误差如图 7 所示。

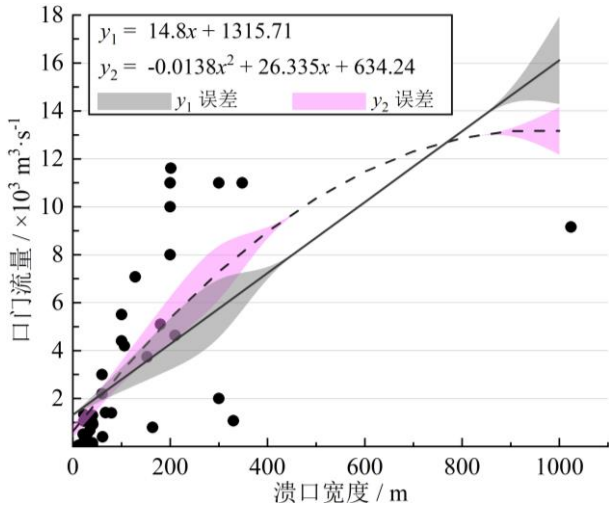


图 7 溃口宽度 B -口门流量 Q 拟合及区间误差

Fig.7 Fitting and interval error of B - Q

(2) 其余特征拟合

对 B - T 、 B - Ev 、 Q - H 三对特征进行拟合插值，拟合结果和区间均方根误差如图 8~10 所示。

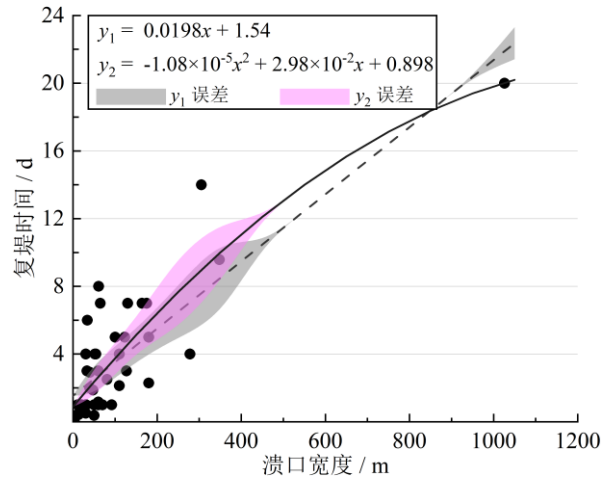


图 8 溃口宽度 B -复堤时间 T 拟合及区间误差

Fig.8 Fitting and interval error of B - T

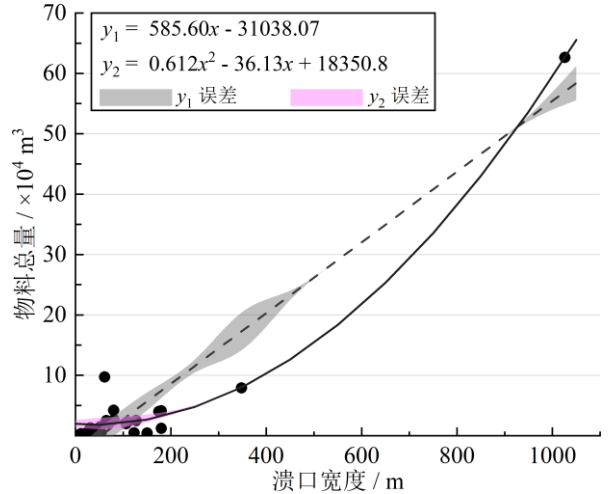


图 9 溃口宽度 B -物料总量 Ev 拟合及区间误差

Fig.9 Fitting and interval error of B - Ev

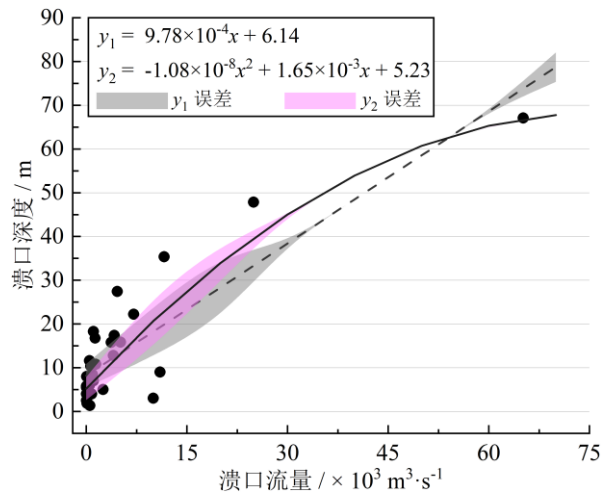


图 10 溃口流量 Q -溃口深度 H 拟合及区间误差

Fig.10 Fitting and interval error of Q - H

(3) 拟合模型选取

比较两种拟合模型对于四对特征各自的 R_{adj}^2 和 AICc 值, 结果如表 3 所示:

表 3 模型拟合优度比较

Table 3 Comparison of model goodness-of-fit

	模型阶数	R_{adj}^2	AICc
$B-Q$	1	0.57	859.38
	2	0.532	840.67
$B-T$	1	0.69	191.35
	2	0.71	190.11
$B-Ev$	1	0.90	533.38
	2	0.97	506.33
$Q-H$	1	0.55	474.39
	2	0.55	476.42

考察不同拟合模型的拟合信息, R_{adj}^2 值均符合拟合优度检验要求。 $B-Q$ 、 $B-T$ 、 $B-Ev$ 三对特征下, 二次曲线的调整-AICc 值相较于线性模型表现出更好的拟合优度, 说明模型参数增加对于改善拟合效果有积极意义。

对于 $B-T$ 拟合, 两种模型误差表现接近, 选择 R_{adj}^2 和 AICc 值更优的二次模型作为拟合补值模型。对于 $B-Ev$ 拟合, 二次模型在拟合优度和区间误差上均有更好表现; 对于 $B-Q$ 和 $Q-H$ 拟合, 进一步采用肖克利奇经验公式考察拟合函数形态^[50]:

$$Q_{\max} = \frac{8}{27} \sqrt{g} \left(\frac{L}{b} \right)^{\frac{1}{4}} b H^{\frac{3}{2}} \quad (19)$$

式中: Q_{\max} 为溃口最大流量; g 为重力加速度; L 为主坝长度; b 为溃口宽度; H 为溃口上游水深。两种模型在拟合优度接近情况下, 二次模型函数形态更符合对应特征值量纲关系, 选择作为拟合补值模型。

2.3 补值结果分析

根据选定的补值模型和口门流速计算方法, 对堤防堵口数据库中的缺失值进行处理, 拟合结果相对误差如图 11 所示, 90% 案例的特征值拟合相对误差在 5% 以内, 表明模型具有良好的预测效果。

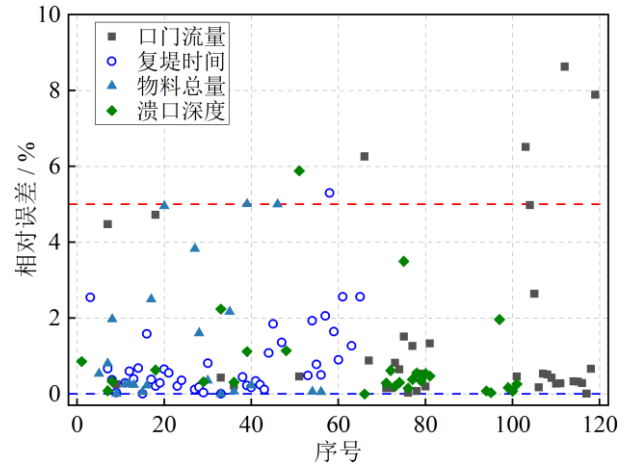


图 11 特征值拟合相对误差分布

Fig.11 Relative Error Distribution of Eigenvalue Fitting

同时使用 KNN 和 MICE 插补算法对目标数据集进行缺失值处理, 取通过随机掩模机制模拟 30% 的随机缺失率对原始真实数据进行遮蔽^[51], 以 NRMSE 和 NMAE 作为评价指标, 对比其与回归插补模型的性能差异, 结果如图 12 所示:

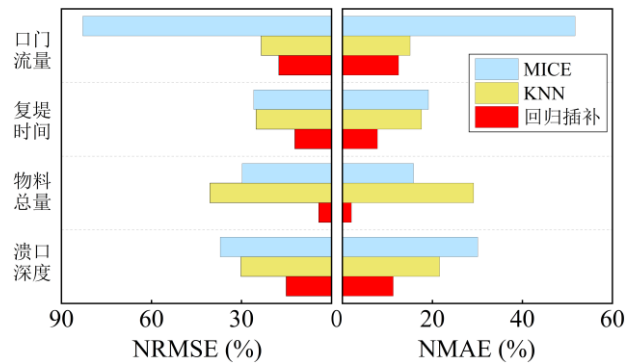


图 12 三种补值方法性能对比

Fig.12 Performance Comparison of Three Imputation Methods

在四组特征上, 回归插补方法的误差仅为 KNN 和 MICE 的 10%~80%, 在目标数据集上呈现出良好的缺失值填补性能。

针对实发溃口案例特征值分布进行分析, 其中湖北省荆州市监利县上车湾溃口宽度达到 1024m, 对于一般溃口堵复工作已不具备普遍指导意义, 排除后其余数据补值结果及典型案例如图 13 所示:

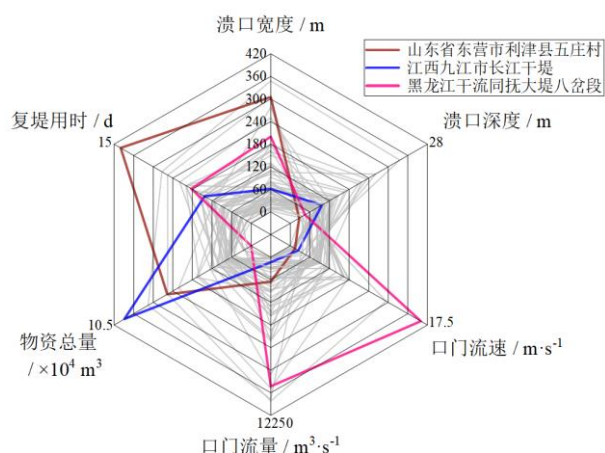


图 13 堤防堵口数据补值结果及典型案例

Fig.13 Imputation results of levee breach closure data and typical cases

分析补值结果,对于溃口宽度 $B < 240\text{m}$ 的大部分堤防溃口,其各特征值分布比较集中,对于新发溃口抢险方案制定有一定参考价值,但同时也观察到部分案例的某些特征值不符合一般规律,例如:

① 山东省东营市利津县五庄村,溃口复堤时间相较于同等规模溃口多出近一倍。该溃决发生于1955年,堵口时以人力修筑搂厢埽,筑垛截流,进而修正边坝协同进占^[52],机械化设备尚未投入使用,因此复堤时间偏长。

② 江西九江市长江干堤,溃口宽度 60m ,但投入堵口物料总量近十万方^[53]。与该溃口位于九江市城区,相关部门高度重视,短时间内调集大量物资参与抢险救援有关。

③ 黑龙江干流同抚大堤八岔段,其口门流速根据观测数据计算得到,存在明显异常。所记录的溃口宽度 200m ,深度 3m ,但口门流量为 $10000\text{m}^3/\text{s}$,可能是因为现场记录数据不准确导致。

因此考虑到险情实际,进行历史相似案例推荐时不仅需分析溃口参数尺寸、口门流场特征等客观数据,还应结合现场物资调度、灾情缓急等因素对参评特征进行专家评估和重要程度排序,因此引入的主观权重对于筛选更合适的推荐案例,进而合理调整救灾规划,具有积极意义。

3 基于协同权重的堤防堵口相似案例推荐方法

3.1 特征协同权重计算

3.1.1 随机森林确定客观权重

堤防堵口案例数据集中,每一条数据由8个维度的数据组成,除参与拟合补值的6个特征外,还包括道路情况 Rc 和堵口方案 Rm ,均二者为文本形式表述的语义信息。以堵口方案为案例标签,进行相似案例推荐时,以溃口宽度、深度、口门流速、流量、能够调度的物资总量和道路信息为输入特征,根据推荐案例指导堵口方案的选择,同时提供复堤时间的参考,特征属性如表4所示。

表 4 案例特征属性及说明

Table 4 Case characteristics attributes and description

特征	符号	说明	特征	符号	说明
溃口宽度	B	输入特征	物料总量	Ev	输入特征
溃口深度	H	输入特征	道路情况	Rc	输入特征
口门流速	V	输入特征	堵口方案	Rm	数据标签
口门流量	Q	输入特征	复堤时间	T	决策参考

总结堵口案例中出现的各种道路情况和采用的堵口方案,采用标签编码将文本转化为数值便于参与计算,如表5所示。

表 5 语义特征标签编码

Table 5 Semantic feature label encoding

道路情况	编码	堵口方案	编码
两侧正常通车	0	双向机械化立堵	0
可单侧正常通车	1	单向机械化立堵	1
仅可上人	2	平堵抛投	2
		钢木土石组合坝	3
		人工打桩筑埽	4
		人工抛投进占	5

为避免不同量纲或量级的数据在进行加权计算时数值较高大的指标对模型判断计算的影响,采用 Z-score 方法对数据进行标准化处理^[54]:

$$y_i = \frac{x_i - \bar{x}}{s} \quad (20)$$

式中: y_i 为标准化后的变量, s 为变量 x 的标准差。由式(12)~式(15),对数据集进行五折交叉验证,以

均方根误差为评价指标, 取最优的计算结果。得到堤防堵口案例数据集上, 各输入特征对应的随机森林客观权重如表 6 所示:

表 6 案例特征随机森林权重
Table 6 Case feature random forest weight

特征	B	H	Q	v	Ev	Rc
权重	0.136	0.134	0.105	0.093	0.083	0.450

分析客观权重结果, 道路情况 Rc 权值最高, 表明堵口方案的选择较大程度依赖于抢险现场是否具备大型机械化设备进入施工的条件, 否则应该考虑结合水面设备进行平堵法施工或人力抛投堵口物料进占进行抢险。

3.1.2 主观权重评估

一般情况下, 溃口特征是堵口方案选择的首要参考, 但在实际抢险中, 应对策略也因道路情况、物料调度的不同而有所差异。例如, 2020 年吉林德惠饮马河决口抢险中, 决口南侧因受淹无法通车, 优先考虑了道路情况对堵口方案选择的影响^[55]; 1998 年广东南海市丹灶镇荷村决口, 堵口物料短时间筹集难度大, 方案制定受制于现场物料的调度能力^[56]。本研究基于类似案例进行归纳分析, 考虑特征数量, 使用五标度法^[57]针对不同特征侧重构建了三种候选预案的相对标度, 如表 7 所示:

表 7 三种预案的特征相对标度
Table 7 Relative scale of features for three plans

	B	H	Q	v	Ev	Rc
①	1	1	1	1	3	5
②	1	1	1	1	5	3
③	5	5	5	5	1	3

采用 AHP 法, 计算得到三种预案对应的最大特征值和特征向量, 计算一致性检验指标 CR 均小于 0.1, 满足检验条件。结果如表 8 所示:

表 8 三种预案的主观权重和一致性检验指标
Table 8 Subjective weights and consistency test indices for three plans

	λ_{max}	主观权重 w_B	CR
①	6.0	(0.1, 0.1, 0.1, 0.1, 0.2, 0.4)	0
②	6.0	(0.1, 0.1, 0.1, 0.1, 0.4, 0.2)	0
③	6.0	(0.211, 0.211, 0.211, 0.211, 0.053, 0.105)	0

DEMATEL 法和 Shapley 法对应的主观权重计算结果如表 9 和表 10 所示:

表 9 三种预案的主观权重 (DEMATEL)

Table 9 Subjective weights for three plans (DEMATEL)

主观权重 w_B	
④	(0.15, 0.15, 0.15, 0.15, 0.16, 0.24)
⑤	(0.15, 0.15, 0.15, 0.15, 0.23, 0.17)
⑥	(0.15, 0.15, 0.15, 0.15, 0.26, 0.14)

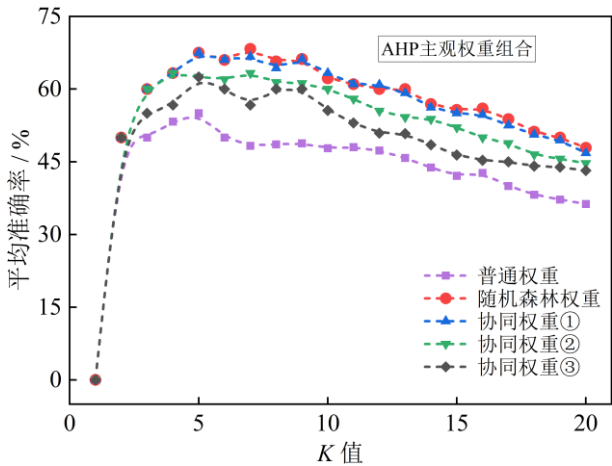
表 10 三种预案的主观权重 (Shapley)

Table 10 Subjective weights for three plans (Shapley)

主观权重 w_B	
⑦	(0.083, 0.083, 0.083, 0.083, 0.258, 0.41)
⑧	(0.083, 0.083, 0.083, 0.083, 0.42, 0.248)
⑨	(0.21, 0.21, 0.21, 0.21, 0.04, 0.12)

3.2 相似度匹配模型优化

由式(18), 取 $\eta = 0.5$, 组合随机森林客观权重与主观权重, 得到三种预案的协同权重, 同时研究 KNN 算法中 K 取值对相似度匹配模型判断准确率的影响, 从而改善模型能力。将数据库按 5: 1 的比例划分为训练集和测试集, 以返回的 K 个案例中与测试数据输入案例堵口方案标签相同的比例作为准确率, 将测试集上的平均准确率作为模型匹配能力的评价标准, 得到平均准确率关于 K 值与权重组合的关系, 如图 14 所示。



(a)

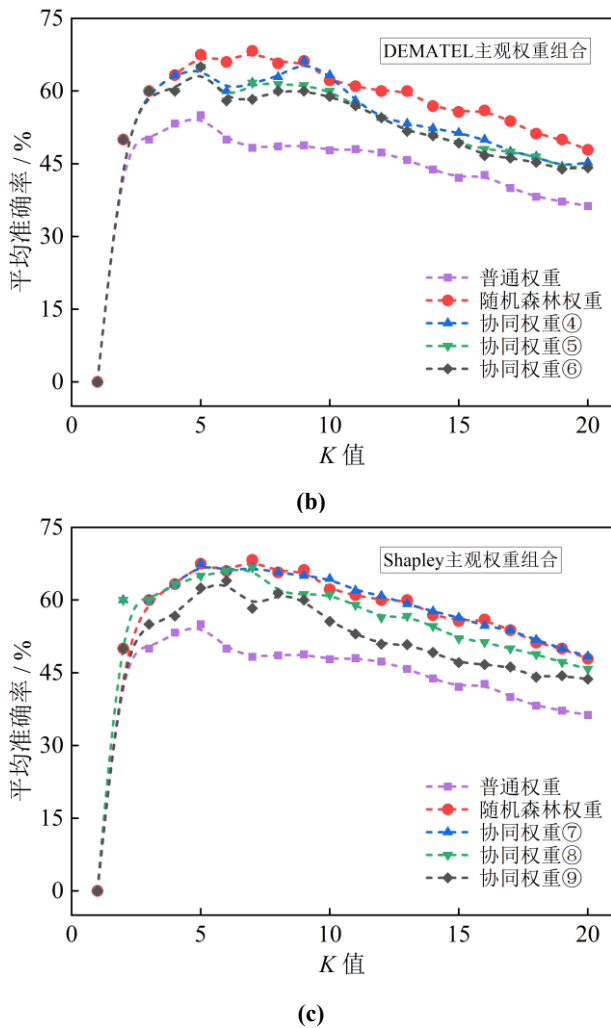


图 14 不同权重组合下 K 值与平均准确率关系

(a) AHP (b) DEMATEL (c) Shapley

Fig.14 Relationship between K value and average accuracy under different weight combinations
(a) AHP (b) DEMATEL (c) Shapley

图中随机森林权重组合和将道路情况作为最重要特征的两种协同权重组合①、⑦在案例相似度匹配上的表现最好,在 $K=5$ 时均达到最高平均准确率,分别为 68.3%、67.5% 和 67.5%,相较于其他权重组合平均准确率最大值高 10~25%;基于 DEMATEL 法的协同权重中,组合④在 $K=9$ 时达到最高准确率 66.2%,应用性较差;普通权重匹配案例时为所有特征赋予相等权值,准确率表现最差,再次说明了道路情况这一特征在堵口方案选择时占据主导地位。

3.3 相似案例推荐实例应用

2016 年 8 月 6 日,湖北省汉川市老观湖外堤发生溃决险情,最大溃口宽度达到 175m,导致 8 万亩农田、2.54 万人受灾^[58]。以此为例,通过输入案例各特征值,采用协同权重①作为相似度匹配权值,检索案例库中相似度排序前 5 的案例作为输出。输入特征值如表 11 所示:

表 11 实发案例输入特征值

Table 11 Actual case feature value input

特征	单位	输入值	特征	单位	输入值
溃口宽度	m	175	口门流速	m/s	0.4
溃口深度	m	6	物料总量	m ³	\
口门流量	m ³ /s	\	道路情况	\	两侧通车

为防止缺值输入后对相似度计算造成影响,根据本文确定的拟合补值模型插补后参与匹配,最终排序输出相似度最高的 5 条案例如表 12 所示。

表 12 相似度匹配结果

Table 12 Similarity match result

序号	案例名	溃口宽度	堵口方案	复堤时间
1	安徽省南陵县西七圩决口	106m	双向机械化进占	3.93d
2	江西省鄱阳县问桂道圩决口	127m	单向机械化进占	3d
3	内蒙古乌海市乌兰木头河民堤	40m	双向机械化进占	2d
4	安徽省桐城大沙河溃口	150m	钢木土石组合坝	5.12d
5	安徽省池州市万子圩一号溃口	110m	钢木土石组合坝	4.04d

匹配的前五条案例中,双向机械化进占方案所属溃决案例的相似度相较于钢木土石组合坝所属案例的更高,且复堤时间更短。因此建议采用双向机械化进占作为抢险方案,预计复堤时间为 3~5 天。

4 总结

本文通过总结历年来国内重要堤防堵口案例资料,采用基于共现水平修正的相关性系数对案例特征间关系进行分析,并对其中的缺失值进行拟合补齐以建立案例数据库。在此基础上引入随机森林确定的客观权重和层次分析确定的主观权重相结

合的协同权重, 提出历史相似案例的推荐方法。具体总结如下:

(1) 本文分析堤防堵口案例数据, 对溃口宽度、溃口深度、口门流量、口门流速、物料总量和复堤时间等这 6 个关键特征, 通过共现水平矩阵修正特征间相关系数, 确定了案例数据缺失值的补齐流程。进一步引入调整 R^2 、区间均方根误差和 AICc 信息准则对不同拟合模型进行评价, 从而建立了堤防堵口案例缺失值的补值流程, 为后续相似案例推荐工作奠定了基础。

(2) 引入标签编码将语义型特征转化为数值参与相似度匹配计算, 通过随机森林方法计算案例库中各因素对于堵口方案选择的客观权重, 并采用层次分析法计算三种侧重观点下方案选择的主观权重。结果表明, 道路情况在堵口方案选择时占据主导地位。

(3) 以平均准确率为评价指标, 进行了不同权重组合下相似案例推荐的算法模型评估, 在 $K=5$ 条件下随机森林权重和协同权重的匹配准确率分别达到 68.3% 和 67.5%, 相较于其他权重组合平均准确率最大值高 10~25%。证明了协同权重能够综合考虑溃口特征和主观考量对于堵口方案选择的影响, 在抢险救援现场有一定参考价值。

参考文献 (References)

- [1] 赵珊珊, 高歌, 黄大鹏, 等. 2004—2013 年中国气象灾害损失特征分析[J]. 气象与环境学报, 2017, 33(01): 101-107.
ZHAO Shanshan, GAO Ge, HUANG Dapeng, et al. Characteristics of meteorological disaster losses in China from 2004 to 2013[J]. Journal of Meteorology and Environment, 2017, 33(01): 101-107.(in Chinese)
- [2] 麦合木提·图达吉, 童瑞, 徐宝宁, 等. 北京“23·7”特大洪水复盘分析[J]. 水力发电学报, 2024, 43(04): 12-22.
Tudaji Mahmut, TONG Rui, XU Baoning, et al. Hindcasting on "July 2023" flood event in Beijing[J]. Journal of Hydroelectric Engineering, 2024, 43(04): 12-22.(in Chinese)
- [3] 中华人民共和国水利部. 2020 年全国水利发展统计公报[R], 2020
Ministry of Water Resources of the People's Republic of China. 2020 Statistic Bulletin on China Water Activities[R], 2020(in Chinese)
- [4] 王茂, 柳欢, 张文超. 辽宁盘锦绕阳河溃口封堵处置经验[J]. 人民黄河, 2023, 45(S1): 19-20.
WANG Mao, LIU Huan, ZHANG Wenchao. Experiences in the Breach Closure and Management of the Raoyang River in Panjin, Liaoning[J]. Yellow River, 2023, 45(S1): 19-20.(in Chinese)
- [5] 杨广西, 张轩庄, 由淑明, 等. 水陆联合快速封堵技术在团洲垸特大决口抢险中的应用[J]. 人民黄河, 2024, 46(S2): 38-39.
YANG Guangxi, ZHANG Xuanzhuang, YOU Shuming, et al. Application of Rapid Land-Water Joint Sealing Technology in the Emergency Repair of the Massive Breach at Tuanzhou Polder[J]. Yellow River, 2024, 46(S2): 38-39.(in Chinese)
- [6] Schank Roger C. Dynamic Memory: A Theory of Reminding and Learning in Computers and People[M]. 1982.
- [7] Kolodner J. L. Improving Human Decision Making through Case-Based Decision Aiding[J]. Ai Magazine, 1991, 12(2): 52-68.
- [8] Guangshu Chang. Applying Case-Based Reasoning Approach to Assembly Part Design[J]. Applied Mechanics and Materials, 2013, 392: 237-241.
- [9] Mohammed M. A., Ghani M. K. A., Arunkumar N., et al. Genetic case-based reasoning for improved mobile phone faults diagnosis[J]. Computers & Electrical Engineering, 2018, 71: 212-222.
- [10] 冯为民, 曹跃进, 任宏. 基于案例模糊推理的土木工程估价估算方法研究[J]. 土木工程学报, 2003, (03): 51-56.
FENG Weimin, CAO Yuejin, REN Hong. The Study on the Case-Based Reasoning Method of the Cost-Estimation in Civil Engineering[J]. China Civil Engineering Journal, 2003, (03): 51-56.(in Chinese)
- [11] 沈扬, 田丹, 刘昊, 等. 面向混凝土坝施工管理的知识图谱智能构建[J]. 水力发电学报, 2022, 41(09): 118-128.
SHEN Yang, TIAN Dan, LIU Hao, et al. Knowledge graph intelligent establishment for concrete dam construction management [J]. Journal of Hydroelectric Engineering 2022, 41(09): 118-128.(in Chinese)
- [12] Allison P. D. Multiple imputation for missing data - A cautionary tale[J]. Sociological Methods & Research, 2000, 28(3): 301-309.
- [13] Emmanuel Tlamele, Maupong Thabiso, Mpoeleng Dimane, et al. A survey on missing data in machine learning[J]. Journal of Big Data, 2021, 8(1): 140.
- [14] Memon Shaheen M. Z., Wamala Robert, Kabano Ignace H. A comparison of imputation methods for categorical data[J]. Informatics in Medicine Unlocked, 2023, 42: 101382.

- [15] Li Jiang, Yan Xiaowei S., Chaudhary Durgesh, et al. Imputation of missing values for electronic health record laboratory data[J]. NPJ Digital Medicine, 2021, 4(1): 147.
- [16] Pearson Karl. Mathematical Contributions to the Theory of Evolution. III. Regression, Heredity, and Panmixia[J]. Philosophical Transactions of the Royal Society A, 1895, 187: 253-318.
- [17] Lin Huang, Eggesbø Merete, Peddada Shyamal Das. Linear and nonlinear correlation estimators unveil undescribed taxa interactions in microbiome data[J]. Nature Communications, 2022, 13(1): 4946.
- [18] Shantal Mohammed, Othman Zalinda, Bakar Azuraliza Abu. Impact of Missing Data on Correlation Coefficient Values: Deletion and Imputation Methods for Data Preparation[J]. Malaysian Journal of Fundamental and Applied Sciences, 2023, 19: 1052-1067.
- [19] Almadani Murad, Bin Waheed Umair, Masood Mudassir, et al. Dictionary learning with convolutional structure for seismic data denoising and interpolation[J]. Geophysics, 2021, 86(5): V361-V374.
- [20] Szopos Noémi Mária, Holb Imre J., Dávid Abriha, et al. Flood risk assessment of a small river with limited available data[J]. Spatial Information Research, 2024, 32(6): 787-800.
- [21] Huber Peter J. Robust Estimation of a Location Parameter[J]. The Annals of Mathematical Statistics, 1964, 35(1): 73-101.
- [22] Steel R. G. D., Torrie J. H. Principles and Procedures of Statistics with Special Reference to the Biological Sciences[J]. Journal of the Royal Statistical Society Series D: The Statistician, 1960, 11(3): 170-171.
- [23] Karch J. Improving on Adjusted R-Squared[J]. Collabra-Psychology, 2020, 6(1).
- [24] Akaike H. New look at statistical-model identification[J]. Ieee Transactions on Automatic Control, 1974, AC19(6): 716-723.
- [25] Burnham Kenneth P., Anderson David R. Model selection and multimodel inference: a practical information-theoretic approach[J]. Journal of Wildlife Management, 2003, 67: 655.
- [26] Cover T. M., Hart P. E. Nearest neighbor pattern classification[J]. Ieee Transactions on Information Theory, 1967, 13(1).
- [27] Abulaish M., Jahiruddin, Assoc Comp Machinery. A Novel Weighted Distance Measure for Multi-Attributed Graph; proceedings of the 10th Annual ACM India Compute Conference, Assoc Comp Machinery Bhopal, Bhopal, INDIA, F Nov 16-18, 2017 [C]. 2017.
- [28] 管清云, 陈雪龙, 王延章. 基于距离熵的应急决策层信息融合方法[J]. 系统工程理论与实践, 2015, 35(01): 216-227.
GUAN Qingyun, CHEN Xuelong, WANG Yanzhang. Distance entropy based decision-making information fusion method[J]. Systems Engineering - Theory & Practice, 2015, 35(01): 216-227.(in Chinese)
- [29] Azur Melissa J., Stuart Elizabeth A., Frangakis Constantine, et al. Multiple imputation by chained equations: what is it and how does it work?[J]. Int J Methods Psychiatr Res, 2011, 20(1): 40-49.
- [30] Breiman L. Random forests[J]. Machine Learning, 2001, 45(1): 5-32.
- [31] Ghosh Supriya, Saha Soumik, Bera Biswajit. Flood susceptibility zonation using advanced ensemble machine learning models within Himalayan foreland basin[J]. Natural Hazards Research, 2022, 2(4): 363-374.
- [32] Breiman L., Friedman J., Olshen R.A., et al. Classification and Regression Trees[M]. 1984.
- [33] Saaty T. L. How to make a decision - the analytic hierarchy process[J]. European Journal of Operational Research, 1990, 48(1): 9-26.
- [34] Tzeng Gwo-Hshiung, Chiang Cheng-Hsin, Li Chung-Wei. Evaluating intertwined effects in e-learning programs: A novel hybrid MCDM model based on factor analysis and DEMATEL[J]. Expert Systems with Applications, 2007, 32(4): 1028-1044.
- [35] Shapley L. S., Tucker Albert William, Kuhn Harold William. A VALUE FOR n-PERSON GAMES[M]. Princeton: Princeton: Princeton University Press, 2016: 307-318.
- [36] Beden Sadeer, Lakshmanan Kayal, Giannetti Cinzia, et al. Steelmaking Predictive Analytics Based on Random Forest and Semantic Reasoning[J]. Applied sciences, 2023, 13(23): 12778.
- [37] 冯婧瑜. 基于案例推理的危化品突发环境事件应急方案选择和调整方法[D]. 长沙: 中南大学, 2022.
FENG Jingyu. Selection and Adaption Method of Emergency Environment Accidents Caused by Hazardous Chemicals Based on Case-based Reasoning[D]. Changsha: Central South University, 2022.(in Chinese)
- [38] 王金虎, 张东晖, 王钰尧. 城市洪涝风险的组合赋重方法研究与应用: 以淮安市为例[J]. 水利水电技术(中英文), 2025: 1-19.
WANG Jinhu, ZHANG Donghui, WANG Yuyao. Research and application of combined empowerment method for urban flood risk: Take Huai'an City as an example[J]. Water Resource and Hydropower Engineering, 2025: 1-19.(in Chinese)

- [39] 唐清竹, 徐宗学, 王京晶, 等. 深圳河流域城市洪涝风险分析[J]. 水力发电学报, 2023, 42(06): 13-22.
- TANG Qingzhu, XU Zongxue, WANG Jingjing, et al. Urban flood risk assessment for Shenzhen River basin[J]. Journal of Hydroelectric Engineering, 2023, 42(06): 13-22.(in Chinese)
- [40] Tadesse Y. B., Fröhle P. Modelling of Flood Inundation due to Levee Breaches: Sensitivity of Flood Inundation against Breach Process Parameters[J]. Water, 2020, 12(12).
- [41] Kakinuma T., Shimizu Y. Large-Scale Experiment and Numerical Modeling of a Riverine Levee Breach[J]. Journal of Hydraulic Engineering, 2014, 140(9).
- [42] 陈启刚, 张大伟, 王忠祥, 等. 堤防溃口水流特性与封堵技术研究进展[J]. 中国防汛抗旱, 2021, 31(08): 1-6.
- CHEN Qigang, ZHANG Dawei, WANG Zhongxiang, et al. Progress inflow characteristics and closure technology of dike breaches[J]. China Flood & Drought Management, 2021, 31(08): 1-6.(in Chinese)
- [43] 李火坤, 曾智超, 邓冰梅, 等. 堤防决口封堵的水力学特性[J]. 水利水运工程学报, 2017, (03): 8-15.
- LI Huokun, ZENG Zhichao, DENG Bingmei, et al. Hydraulic characteristics of levee breach closure[J]. Hydro-Science and Engineering, 2017, (03): 8-15.(in Chinese)
- [44] 孙芦忠, 赵建均, 严建国, 等. 堤防决口的水力学试验研究[J]. 人民长江, 2003, (11): 41-42.
- SUN Luzhong, ZHAO Jianjun, YAN Jianguo, et al. Hydraulic test studies of dike breach[J]. Yangtze River, 2003, (11): 41-42.(in Chinese)
- [45] 田治宗, 梁跃平, 解吉祥, 等. 堤防溃口口门区水力及冲淤特性模型试验研究[J]. 人民黄河, 2003, (03): 32-33.
- TIAN Zhizong, LIANG Yueping, XIE Jixiang, et al. Model test studies on hydraulic and eroding and depositing characteristics in the gate areas of embankment breach[J]. Yellow River, 2003, (03): 32-33.(in Chinese)
- [46] Peng M., Zhang L. M. Breaching parameters of landslide dams[J]. Landslides, 2012, 9(1): 13-31.
- [47] Danka Jozsef, Zhang L. M. Dike Failure Mechanisms and Breaching Parameters[J]. Journal of Geotechnical and Geoenvironmental Engineering, 2015, 141(9): 04015039.
- [48] Wu Weiming. Simplified Physically Based Model of Earthen Embankment Breaching[J]. Journal of Hydraulic Engineering, 2013, 139(8): 837-851.
- [49] 马洪福. 阵列锚技术在堤防溃口堵复中的应用研究[D]. 郑州: 华北水利水电大学, 2019.
- MA Hongfu. Research on the application of array anchor technique in the levee breach plugging[D]. Zhengzhou: North China University of Water Resources and Electric Power, 2019.(in Chinese)
- [50] 可友国, 雷宏军, 王永高, 等. 尖岗水库溃坝洪水计算与风险评估[J]. 人民黄河, 2008, (07): 36-37.
- KE Youguo, LEI Hongjun, WANG Yougao, et al. Jiangang Reservoir Dam Break Flood Calculation and Risk Assessment[J]. Yello River, 2008, (07): 36-37.(in Chinese)
- [51] Pereira Ricardo Cardoso, Abreu Pedro Henriques, Rodrigues Pedro Pereira, et al. Imputation of data Missing Not at Random: Artificial generation and benchmark analysis[J]. Expert systems with applications, 2024, 249: 123654.
- [52] 黄河水利委员会. 中国江河防洪丛书: 黄河卷[M]. 北京: 中国水利水电出版社, 1996.
- Resources Yellow River Conservancy Commission of the Ministry of Water. Series on flood control of rivers in China: volume of the Yellow River[M]. Beijing: China Water&Power Press, 1996.(in Chinese)
- [53] 杨光煦. 九江长江江堤堵口实录及经验[J]. 人民长江, 1998, (11): 4-7+49.
- YANG Guangxu. The record and experience of river dike closure in Jiujiang City of the Yangtze River[J]. Yangtze River, 1998, (11): 4-7+49.(in Chinese)
- [54] Al Shalabi L., Shaaban Z., Kasasbeh B. Data mining: a preprocessing engine[J]. Journal of Computer Sciences, 2006, 2(9): 735-739.
- [55] 林金良, 宁占金, 崔堃鹏, 等. 2020 年吉林德惠饮马河决口抢险案例分析[J]. 东北水利水电, 2022, 40(03): 58-60.
- LIN Jinliang, NING Zhanjin, CUI Kunpeng, et al. Emergency case analysis of Yinmahe River breach in Dehui City of Jilin Province in 2020[J]. Water Resources & Hydropower of Northeast China, 2022, 40(03): 58-60.(in Chinese)
- [56] 李雁翀, 陈晖. 堵口决战——南海市丹灶镇荷村决堤堵口战斗纪实[J]. 珠江水运, 1998, (08): 6-7.
- LI Yanchong, CHEN Hui. The Decisive Battle of Dike Closure: A Chronicle of the Fight Against Levee Breach in He Village, Danzao Town, Nanhai City[J]. Pearl River Water Transport, 1998, (08): 6-7.(in Chinese)
- [57] 骆正清, 杨善林. 层次分析法中几种标度的比较[J]. 系统工程理论与实践, 2004, 24(9): 51-60.

LUO Zhengqing, YANG Shanlin. Comparative Study on Several Scales in AHP[J]. Systems Engineering — Theory & Practice, 2004, 24(9): 51-60.(in Chinese)

[58] 查显能. 汉北河中下游流域洪水风险分析[D]. 武汉: 华中科技大学, 2020.

ZHA Xianneng. Flood Risk Analysis of the Middle and Lower Reaches of Hanbei River[D]. Wuhan: Huazhong University of Science & Technology, 2020.(in Chinese)