

# 基于大数据的水闸运行效率优化与故障预测模型研究

赵倩

河北省水利水电勘测设计研究院集团有限公司 河北 石家庄 050000

**摘要:** 随着信息技术的发展,水利工程管理越来越多地采用大数据技术。本次研究以水闸运行效率优化和故障预测为重点,目的是借助大数据分析技术来提高水闸运行维护管理水平。研究从水闸运行数据采集和处理入手,利用先进的数据预处理技术和方法保证数据准确可用。然后,构建一套水闸运行效率评价指标体系,运用数据挖掘技术对其运行效率进行优化。另外,本研究发展了一种基于机器学习的故障预测模型并对模型进行评价和改进以提高预测精度。最后通过案例分析对模型进行验证,以期水闸智能化管理提供科学依据。研究结果表明:所提模型和方法能有效地促进水闸运行效率的提高和潜在故障的精确预测,对于水利工程可持续发展至关重要。

**关键词:** 水闸; 运行效率; 故障预测; 数据挖掘; 智能化管理

## 引言

在目前水利工程管理中,水闸是重点控制节点,水闸运行效率及可靠性直接影响水资源合理配置及防洪安全问题。但是传统水闸管理方法通常依靠经验判断与人工操作相结合,很难满足水利工程管理不断增长的需要。在大数据技术崛起的背景下,如何通过大量运行数据优化水闸运行效率和对可能出现的故障进行提前预测已经成为水利工程管理领域急需解决的难题。研究目的在于探究基于大数据下水闸运行效率优化及故障预测模型构建方法,以提升水闸管理水平及运行安全性。

## 1 数据采集的重要性与挑战

### 1.1 数据采集的重要性

水闸是水利工程中的重要设施,水闸运行效率与安全直接影响着水资源合理开发利用与防洪减灾工作成效。通过收集水闸运行资料并进行分析,可找出水闸运行中存在的规律及可能存在的问题,从而为水闸运行维护及故障诊断提供科学依据。另外数据采集也是水闸运行效率优化模型与故障预测模型建设的依据,优质的数据获取工作能够提升模型精度与可靠性。<sup>[1]</sup>

水闸运行效率优化需考虑水闸运行状态,外部环境和管理措施等诸多因素。通过收集水闸运行数据并进行分析,能够充分掌握水闸运行状态,找出影响运行效率的关键因素,从而为优化措施的制定奠定基础。比如对水闸流量、水位以及闸门开度等资料进行分析,就能够对水闸运行效率进行评价,发现问题出现的位置及成因,然后采取相关优化措施来改善水闸运行效率。

水闸运行时可能发生的故障种类较多,例如闸门卡阻、启闭机失效、电气设备失效等等。这些故障如不能及时发现并处理掉,就有可能影响水闸正常工作甚至

造成安全事故。通过收集水闸运行数据并进行分析,能够对水闸运行状态进行实时监控,并及时发现异常,从而为故障诊断与预报奠定了基础。另外,通过机器学习及其他数据挖掘技术分析历史故障数据并构建故障预测模型也可用于水闸可能发生的故障预测与预警。

### 1.2 数据预处理技术与方法

大数据时代下,数据预处理作为保证数据质量的关键一步,对水闸运行效率优化和故障预测模型研究具有十分重要的意义。数据预处理包括数据清洗、转换、归一化以及降维几个步骤,其目的是增强数据可用性与准确性。第一,数据清洗就是对噪声及不一致数据进行剔除,其中包括对缺失值,异常值以及重复记录等进行处理。缺失值的处理可采用插值,删除或者预测填充等方法;异常值的识别和修正需要依赖于统计分析或者专业领域的知识。第二,数据转换就是把原始数据转换成适用于分析的形式,比如日期形式统一,文本数据编码。归一化是一种通过将数据缩小到特定尺度来减少不同量纲影响的方法,常用的实现方式包括最小-最大归一化和Z-score标准化。第三,像主成分分析(PCA)和线性判别分析(LDA)这样的降维技术能够降低数据维度,从而提升模型的计算速度和泛化性能。

### 1.3 数据存储与管理策略

高效的数据存储和管理策略,对大数据环境中水闸运行效率的优化和故障预测模型的研究具有同等重要的意义。<sup>[2]</sup>数据存储需综合考虑其大小,种类及访问频率等因素,并选用适当存储介质及格式。MySQL, PostgreSQL等关系型数据库适合结构化数据存储与管理,MongoDB, Hadoop等非关系型数据库比较适合大规模非结构化或者半结构化数据处理。数据仓库的技术,例如数据湖和

数据集市, 可以支持数据的整合、清理和转化, 为数据分析提供一个统一的视角。数据管理策略主要包括安全性、完整性、可访问性等。数据加密、访问控制、备份机制等是保证数据安全性的关键举措。数据完整性可通过事务管理, 数据校验, 审计日志等方式进行保持。数据的可访问性是基于高效的索引机制、缓存策略和查询优化技术来实现的。另外, 对数据进行生命周期管理就是要保证数据能够及时地进行更新与淘汰, 其中就涉及数据的建立、更新、存档以及删除等环节。通过采用合适的数据储存和管理方法, 我们能够在水闸的运行效率提升和故障预测模型研究提供稳固、可信赖且高效的数据基础。

## 2 水闸运行效率优化模型构建

### 2.1 运行效率评价指标体系的构建

水闸运行效率评估指标体系的构建, 是建立优化模型的第一步。评价指标体系是否科学合理, 直接影响着优化模型是否有效。构建评价指标体系时需考虑水闸运行特点及实际需要, 筛选出能综合体现水闸运行效益的关键指标。其中包括但不限于水闸开闸次数, 运行时间, 过闸流量, 能耗和故障率。同时还要为这些指标赋权, 保证评价结果客观准确。<sup>[3]</sup>

评价指标体系构建时可通过专家咨询和数据分析对各指标重要度进行评价与排序。另外, 可参考国内外水闸运行效率评价的先进体系并根据实际情况对其加以优化与完善。构建科学合理评价指标体系可为水闸运行效率优化提供强大数据支撑。

### 2.2 基于数据挖掘的运行效率优化算法

数据挖掘技术对于优化水闸运行效率有着重要应用。利用数据挖掘技术能够从大量水闸运行数据中发掘出宝贵的信息与知识, 从而为水闸运行效率优化提供决策支持。<sup>[4]</sup>首先要做好水闸运行数据预处理工作, 主要包括数据清洗、数据转换以及数据归一化工作, 从而提升数据质量以及可用性。经过预处理之后, 数据会被用作数据挖掘算法中的一个输入。其次选取适当的数据挖掘算法, 对经过预处理的数据进行了分析与挖掘。数据挖掘中经常使用的算法有: 分类方法、聚类技术、关联规则的挖掘方法以及预测技术等。水闸运行效率优化时可结合实际需要选用适当算法。最后利用数据挖掘结果对水闸运行效率进行了优化。水闸运行参数可依据挖掘出来的规律与方式进行调节与优化, 从而提升水闸运行效率。例如, 我们可以基于挖掘出的水闸开启频率与过闸流量之间的联系, 来调整水闸的启动策略; 可根据预测的故障率提前做好故障预防及处理工作等。

### 2.3 模型验证与优化策略

水闸运行效率优化模型构建过程中验证其有效性与准确性是一个关键环节。模型验证一般包括历史数据回溯分析, 检验未来数据预测能力。该研究利用交叉验证将数据集划分为训练集与测试集, 利用训练集对所建立的模型进行训练, 再在测试集上对所建模型进行性能评价。除此之外, 为了量化模型的预测准确性, 引入了多个统计指标, 包括均方误差 (MSE)、决定系数 (R<sup>2</sup>) 以及绝对百分比误差 (MAPE)。

对于模型的优化, 研究中采取了各种不同的策略。首先对该模型参数进行详细调整以便找到最佳参数组合。其次介绍L1,L2正则化等正则化技术避免模型的过拟合。除此之外, 我们还探索了如随机森林和梯度提升机 (GBM) 等多种模型整合技术, 旨在增强模型的泛化性能。综合运用上述策略, 显着提高模型预测精度。

### 2.4 案例分析: 水闸运行效率优化实践

为进一步检验该模型在实际中的应用价值, 研究选择一座大型水闸为例。在深入分析该水闸历史运行资料的基础上, 构建运行效率优化模型并用于实际运行管理。该模型应用后水闸运行效率显着提高, 能源消耗降低, 设备使用寿命增加。案例分析部分首先详细收集水闸运行数据并进行预处理, 主要包括水闸开关次数、运行时间、流量、水位等关键参数。接着在这些数据的基础上构建运行效率优化模型进行验证与优化。模型应用效果表明: 水闸运行策略优化后运行效率明显提高。比如, 通过对水闸开关时间及频率进行调节, 在确保水闸正常工作的前提下, 能够降低不必要能耗。另外, 该模型也有助于管理人员识别出可能影响水闸长期运行效率的设备老化和维修不到位等运行中存在的问题。

## 3 水闸故障预测模型研究

### 3.1 故障预测模型的构建原理

建立故障预测模型是在对水闸运行数据进行深入分析的基础上进行的, 其目的是找出故障产生的规律与方式。<sup>[5]</sup>该模型建立过程主要由数据收集、特征提取、模型选择、训练与验证组成。首先要采集水闸工作时的水位、流速、闸门开度和电机电流等多种资料。接着, 对数据进行预处理, 并利用特征工程对故障进行相关特征提取。其次选择适当的机器学习方法, 例如决策树、随机森林和支持向量机等, 来训练特征, 并据此构建一个故障预测模型。最后采用交叉验证等手段对该模型进行了评价与优化, 提高了预测准确性。

### 3.2 机器学习在故障预测中的应用

机器学习是故障预测领域中被广泛使用的有效数据

分析方法。机器学习算法可以自动地从海量历史数据中学习出故障产生的规律以达到预测故障的目的。在进行水闸故障的预测时,经常采用的机器学习方法有:决策树。通过建立决策树模型,我们可以将水闸的运行数据映射到不同的故障类别中,从而达到故障的分类和预测目的;随机森林:以决策树为基础的集成学习方法通过构造多棵决策树,对决策树进行表决或者平均来提高故障预测精度与鲁棒性;支持向量机能够通过寻找最佳的超平面来区别故障状态与正常状态,显示出良好的分类能力;神经网络:通过构造复杂网络结构和仿真人脑神经网络来进行故障非线性预测;深度学习:采用多层神经网络对数据深层次特征进行自动学习以提升故障预测精度。

### 3.3 故障预测模型的评估与改进

对故障预测模型进行评价是建立模型的一个重要步骤,旨在验证预测性能及可靠性。常见评估指标有准确率,召回率和F1分数。通过上述指标可综合评估模型预测效果。模型评估时可能发现该模型有偏差或者缺陷,有必要对其进行完善与优化。改进方法主要有以下几个方面:特征工程:通过对特征进行增加,删除或者变换来改善模型预测性能;模型融合:融合多种模型预测结果以增强模型鲁棒性与准确性;参数调优:通过调节模型中的一些超参数,例如学习率和树的深度,我们可以提高模型的整体性能;数据增强:利用数据增强手段,例如过度采样和欠采样,可以增强模型处理不平衡数据的能力;模型解释性:增强模型可解释性,有助于用户了解预测结果及决策过程。

### 3.4 故障预测模型在水闸管理中的应用案例

将故障预测模型运用于水闸管理能够显著提升水闸运行效率与安全性。某水闸管理单位为提升水闸运行安全,构建了基于机器学习的故障预测模型。首先搜集历年水闸运行资料,主要有水位、流速、闸门开度和电机电流。接着对数据进行预处理并进行特征工程以提取故

障相关特征。其次选用随机森林算法建立故障预测模型并对模型进行训练验证。模型评价阶段采用交叉验证等手段,对模型预测性能进行评价,结果表明该模型准确率与召回率都达到较高水平。模型应用阶段本单位将故障预测模型融入水闸监控系统,以实现水闸故障实时预测与预警。当该模型预测出可能出现的故障后,该系统将自动报警,给出故障诊断与维修建议。

### 结束语

本次研究通过对水闸运行数据的深入剖析,成功搭建运行效率优化模型及故障预测模型。这些模式的构建在促进水闸运行效率的同时,故障预测精度也有显著提高,对水闸智能化管理具有有力技术支撑。国内外学者一般认为将大数据技术应用于水利工程领域是水资源高效利用与可持续发展的重点。

研究结论显示:利用大数据技术获取水闸运行数据并进行处理与分析能够有效发现水闸在运行过程中存在的规律与潜在问题。对水闸进行维修与管理提供科学依据,也可为水闸优化运行及故障预防等提供新视角。通过该模型进行验证与优化,进一步验证了上述模型在实践中是有效可行的。

### 参考文献

- [1]辛苗苗. 自动化监控系统在水闸运行管理中维护与应用研究[J]. 中国水运(下半月), 2023, 23(06): 40-42.
- [2]高民望,陈剑,严凯,等. 自动化系统在水闸工程应用中的问题与解决方法[J]. 中国设备工程, 2023, (05): 205-207.
- [3]邵豫东. 水闸工程运行管理及日常维护[J]. 河南水利与南水北调, 2020, 49(11): 60-61.
- [4]刘鑫. 大运河水闸运行与管理规则[J]. 河南水利与南水北调, 2020, 49(10): 75-76.
- [5]叶铭俊. 自动化系统在广州市番禺市桥河雁洲水闸中的应用[J]. 科技风, 2020, (21): 9.