

# 大数据驱动的水利工程造价指标数据库构建与智能分析

樊彦君

黄河勘测规划设计研究院有限公司 河南 郑州 450000

**摘要:** 随着大数据技术的蓬勃发展,其在水利工程造价管理领域的应用愈发关键。本文围绕大数据驱动展开研究,先剖析理论基础与现有问题,接着构建涵盖架构、指标标准化、数据采集与质量控制的水利工程造价指标数据库。随后打造智能分析模型,应用于成本预测等场景。此外,针对数据安全、跨部门协同、模型可解释性等障碍提出突破路径,推动水利工程造价管理智能化升级。

**关键词:** 大数据驱动;水利工程;造价指标;数据库构建;智能分析

引言:在水利工程规模不断扩大、投资日益多元的当下,传统造价管理方式因数据分散、分析滞后等局限,难以满足精准管控需求。大数据技术凭借强大的数据采集、存储、分析与挖掘能力,为水利工程造价管理带来新契机。构建大数据驱动的水利工程造价指标数据库,并开展智能分析,可实现造价动态预测、风险预警与资源优化,对提升水利工程投资效益、保障工程顺利建设意义重大。

## 1 理论基础与文献综述

### 1.1 水利工程造价管理核心理论

(1) 全生命周期成本管理理论(LCC):该理论贯穿水利工程规划、设计、施工、运维至报废的全流程,核心是实现工程全周期成本最优,而非仅控制施工阶段造价。相较于传统造价管理侧重施工环节的局限,LCC理论强调统筹考虑前期设计成本、中期建设成本与后期运维成本,契合水利工程投资大、工期长、运维难度高的特点,为造价管控提供了系统性思维框架。(2) 造价指标体系构建原则:构建科学合理的水利工程造价指标体系,需严格遵循三大核心原则。系统性原则要求指标覆盖工程建设各环节、各业态,形成完整体系;科学性原则强调指标选取贴合工程实际,计算方法规范可追溯;动态性原则需兼顾市场价格波动、政策调整等因素,确保指标具备实时适配性。

### 1.2 大数据技术在工程造价领域的应用进展

(1) 数据采集与清洗技术:依托物联网设备、无人机巡检等手段,可实时监测水利工程材料消耗、施工进度等数据,同时通过数据清洗算法剔除异常值、冗余数据,保障造价数据的准确性,解决了传统数据采集效率低、误差大的问题。(2) 关联规则挖掘:通过大数据算法挖掘工程特征(如坝型、规模、地质条件)与造价指标(如单位造价、材料单价)的内在关联性,可为造价

编制提供精准参考,减少人为估算的偏差。(3) 动态预测模型:LSTM神经网络、随机森林等算法的应用,实现了水利工程造价的动态预测,能够根据施工过程中的变量的实时调整造价预测结果,提升造价管控的前瞻性<sup>[1]</sup>。

### 1.3 水利工程造价指标研究现状

(1) 现有指标体系的局限性:多数现有指标体系针对特定水利工程业态(如水库、灌区)构建,缺乏跨业态通用性,难以适配不同类型、不同规模水利工程的造价管控需求,指标的适用性和扩展性不足。(2) 数据孤岛问题与标准化缺失:不同建设单位、设计单位的造价数据分散存储,形成数据孤岛,难以实现数据共享;同时,造价数据的采集标准、统计口径不统一,标准化缺失,严重制约了造价指标的推广应用和行业整体造价管理水平的提升。

## 2 基于大数据驱动的水利工程造价指标数据库构建

### 2.1 数据库架构设计

(1) 数据层:作为数据库的核心基础,负责整合各类与水利工程造价相关的多元数据,打破数据孤岛,实现数据全面覆盖。具体整合数据包括政府造价信息(如定额标准、调价文件、招投标指导价)、企业成本数据(如施工成本、管理费用、利润水平)、市场供求数据(如建材价格波动、劳动力薪酬、设备租赁价格)、物联网实时数据(如施工过程中材料消耗、设备运行时长、工序完成进度等动态数据),通过多源数据整合,为造价指标分析与生成提供充足的数据支撑。(2) 技术层:承担数据存储、处理与分析的核心功能,结合大数据技术特性,采用适配多源异构数据的技术框架。选用Hadoop分布式存储框架,解决水利工程造价数据量大、类型复杂(结构化与非结构化并存)的存储难题,实现数据的安全存储与高效调用;搭配Spark流处理框架,提升实时数据(如物联网监测数据)的处理速度,确保数据处理的时效性,为造价动态调整提供技术保障<sup>[2]</sup>。(3) 应用层:聚

焦水利工程造价管理的实际应用场景，实现数据价值的转化，满足不同用户的使用需求。应用层核心支持三大功能，一是在线计价功能，可依托数据库中的指标数据，快速完成水利工程各分项、分部工程的计价编制；二是指标生成功能，自动提取数据层信息，按照标准化体系生成各类造价指标，减少人工计算误差；三是动态调整功能，根据市场价格波动、政策调整及工程实际变化，实时更新造价指标，确保指标的准确性与适用性，同时支持

一级（工程业态）	二级（工程专业）	三级（指标项示例）
水库工程	土建工程	钢筋含量、混凝土用量
灌区工程	机电安装工程	设备单价、安装费率

（2）指标类型：结合水利工程造价管理的核心需求，将指标分为经济指标与技术指标两大类，兼顾造价经济性与工程技术性。1）经济指标：聚焦工程成本与造价水平，主要包括单位工程规模造价（如每立方米混凝土造价、每公里渠道造价）、功能区域造价（如闸门安装每平方米造价、泵房每千瓦造价），直观反映工程的造价水平与成本分布；2）技术指标：关联工程规划、设计与施工工艺，主要包括规划指标（如灌区灌溉面积利用率、水库库容利用率）、设计指标（如钢模比、墙窗比、混凝土强度等级相关指标），为造价合理性分析提供技术支撑，实现技术与经济的协同管控<sup>[3]</sup>。（3）动态扩展机制：为适应水利工程新业态、新技术、新工艺的发展需求，避免指标体系僵化，构建“通用专业+专项专业”的动态扩展机制。通用专业涵盖各类水利工程共有的核心专业与指标，确保体系的通用性；专项专业针对新型水利工程（如智慧水利工程、生态水利工程）或特殊工艺，灵活新增专业类别与指标项，无需调整整体框架，即可实现指标体系的动态更新，提升数据库的适配性与长效性。

### 2.3 数据采集与质量控制

（1）数据来源：全面覆盖结构化与非结构化两类数据，确保数据的多样性与完整性。1）结构化数据：主要来源于企业ERP系统（如成本核算数据、人员薪酬数据）、工程量清单文件、招标投标文件、竣工结算资料等，此类数据格式规范、便于直接整理与分析，是造价指标计算的核心数据；2）非结构化数据：包括施工日志、监理报告、行业论坛讨论、政策文件、技术规范等，此类数据蕴含大量隐性造价信息，通过整理分析可补充结构化数据的不足，提升指标的全面性。（2）清洗与标准化：针对采集的数据存在的异常、格式不统一等问题，开展系统性的清洗与标准化处理。1）采用“三倍价差法”剔除异常值，通过计算同类数据的平均值与标准差，将超出

用户自定义查询、统计与导出，提升使用便捷性。

### 2.2 指标体系标准化设计

三级分类框架：按照“层层拆解、精准定位”思路，构建三级造价指标分类框架，确保覆盖全面、逻辑清晰。一级为工程业态，涵盖水库、灌区、河道整治、水电站等水利主要类型；二级为工程专业，各业态下划分土建、机电安装等核心专业；三级为指标项，各专业下细化具体指标，具体分类如下表所示：

平均值±3倍标准差的数据判定为异常值（如异常偏高或偏低的建材价格），确保数据的合理性；2）统一数据格式，针对不同计价软件生成的XML文件、Excel表格等不同格式数据，统一转换为JSON格式，实现数据的互通兼容，便于后续存储与处理；3）附加工程特征标签，为每条数据标注对应的工程特征信息，如地质条件（岩石地基、土质地基）、施工工艺（现浇施工、预制装配）、工程规模等，确保指标生成的精准性，满足不同场景下的造价管控需求<sup>[4]</sup>。

### 3 基于大数据的水利工程造价指标数据库智能分析模型构建与应用

#### 3.1 基于机器学习的动态预测模型

（1）模型输入：结合水利工程特点，筛选三类核心输入变量，确保预测的全面性与精准性。一是工程特征参数，涵盖水利工程核心指标，如坝体建筑面积、结构类型（重力坝、拱坝）、工程规模、地质条件等；二是市场价格波动，主要包括建材（钢筋、混凝土）、劳动力、设备租赁等价格的实时波动数据；三是政策变量，涵盖定额标准调整、税收政策、环保政策等影响造价的政策因素，所有输入变量均来源于前文构建的标准化数据库，确保数据一致性。（2）算法选择：结合不同变量的特性，采用“LSTM神经网络+随机森林”的组合算法，兼顾时间序列特性与高维数据处理能力。1）LSTM神经网络：重点捕捉造价指标的时间序列依赖性，如年度建材价格波动、政策调整的时序影响，解决传统算法难以处理的长期依赖问题，精准拟合价格与造价的动态关联；2）随机森林：主要处理高维非线性数据，如多标段水利项目的成本关联分析、多因素耦合下的造价波动分析，可有效规避过拟合风险，提升模型的泛化能力<sup>[5]</sup>。（3）模型优化：为提升预测精度，建立完善的模型优化与评估体系。通过网格搜索法系统调整模型超参数，如LSTM的隐藏

层节点数、随机森林的决策树数量,筛选最优参数组合;采用MAE(平均绝对误差)、 $R^2$ (决定系数)作为核心评估指标,MAE用于衡量预测值与实际值的偏差程度, $R^2$ 用于评估模型的拟合效果,确保模型预测误差控制在合理范围,满足实际应用需求。

### 3.2 智能分析场景应用

(1) 成本预测:以某地区小型水库加固项目为例,向模型输入项目建筑面积、坝体结构类型、当期钢筋混凝土价格及最新定额标准等变量,模型快速输出项目单方造价预测值,经实践验证,预测误差可控制在3%以内,相较于传统人工估算,大幅提升了造价预测的精准度与效率,为项目投资决策提供了可靠支撑。(2) 风险预警:模型实时调用数据库中的动态数据,监控项目实际成本与预算的偏差率,设定10%为预警阈值,当实际成本超出预算10%时,系统自动触发预警机制,同步推送偏差原因分析(如建材价格暴涨、施工工艺调整),提醒管理人员及时采取管控措施,规避造价失控风险。(3) 资源优化:通过模型的聚类分析功能,对水利工程各施工环节的成本数据进行分类,识别出高成本环节(如大型设备租赁、特殊建材采购),结合市场供求数据,提出针对性的资源优化建议,如调整材料采购批次、选择性价比更高的替代建材,降低工程成本,实现资源的高效合理配置。

## 4 大数据驱动下水利工程造价管理的核心障碍及突破路径

### 4.1 数据安全与隐私保护

(1) 采用区块链技术实现造价数据不可篡改存储,依托区块链去中心化、不可篡改的核心特性,对造价核心数据进行加密存储,全程记录数据录入、修改、调用轨迹,确保数据真实性与完整性,从技术层面防范数据被恶意篡改、伪造的风险。(2) 建立分级访问权限控制体系,结合用户岗位与职责,划分不同层级的访问权限,明确管理员、造价人员、监管人员等不同角色的数据访问范围,禁止越权访问敏感信息,同时留存访问日志,实现数据操作可追溯,全方位保障数据安全与隐私。

### 4.2 跨部门数据协同

(1) 推动水利、住建、财政部门数据接口标准化,联

合各相关部门制定统一的数据接口规范,明确数据采集、传输、存储的标准要求,解决不同部门数据格式不兼容的问题,为数据互通奠定基础。(2) 构建行业级数据共享平台,整合各部门核心造价相关数据,打破部门间的数据壁垒,实现定额标准、市场价格、项目审批等数据的高效共享,推动各部门协同开展造价管控、审核监管工作,提升行业整体管理效率。

### 4.3 模型可解释性

(1) 引入SHAP值分析解释机器学习模型预测结果,通过SHAP值量化各输入变量(如建材价格、工程规模)对造价预测结果的影响程度,清晰呈现预测逻辑,让管理人员直观了解不同因素如何影响造价结果,破解模型“黑箱”难题。(2) 开发可视化决策支持工具,将模型预测结果、变量影响分析等内容以图表、曲线等直观形式呈现,简化复杂数据与模型逻辑,帮助管理人员快速理解模型输出,基于可视化结果做出科学的造价管理决策,提升模型的实用性。

### 结束语

大数据驱动的工程造价指标数据库构建与智能分析,是水利工程领域顺应时代发展的创新实践。通过构建完善的数据库与智能分析模型,有效解决了传统造价管理的诸多难题,实现了造价的精准预测、风险的有效预警与资源的合理配置。未来,随着技术的持续进步,需不断优化数据库与分析模型,强化数据安全保障,推动工程造价管理向更高水平的智能化、科学化迈进。

### 参考文献

- [1] 夏蓉,张坤杰.大数据技术构建企业工程造价指数指标数据库[J].中国建设信息化,2023,(24):58-62.
- [2] 田志超.大数据时代工程造价数据库构建及其应用思考[J].工程造价管理,2023,(05):9-11.
- [3] 代端明.基于大数据技术的工程造价数据库构建[J].广西城镇建设,2021,(8):79-83.
- [4] 房光玉,刘寨民.应用大数据构建工程造价数据库[J].工程造价管理,2020,(4):82-87.
- [5] 郝婷.基于大数据的工程造价指标数据库建设与应用研究[J].建筑设计及理论,2025,(5):90-94.