

# 基于大数据的工程造价预测模型研究

郭 云

云南华电龙盘水电开发有限公司 云南 昆明 650233

**摘 要：**本文聚焦基于大数据的工程造价预测模型研究。阐述大数据技术“5V”特征及在造价领域的应用生态，分析传统预测局限与现代理论进展。探讨造价影响因素分类、关联性及不同工程类型关键因素。构建分层模块化预测模型框架，对比单一模型并设计融合模型。提出数据质量、模型性能及全生命周期动态预测优化策略。研究表明大数据技术推动造价预测向智能、精准、动态方向演进。

**关键词：**工程造价；大数据分析；预测模型优化

引言：在建筑行业快速发展背景下，工程造价预测的准确性对项目成本控制和效益提升至关重要。传统预测方法受数据维度单一、模型适应性差、时效性不足等局限，难以满足复杂多变的工程需求。大数据技术凭借其海量数据处理、实时分析等优势，为工程造价预测带来新契机。通过整合多源数据、革新分析方法、转变决策模式，大数据技术有望突破传统边界，构建更科学精准的预测模型，推动建筑行业高质量发展。

## 1 大数据与工程造价预测的基础理论

### 1.1 大数据技术概述

大数据技术是以分布式计算框架为核心，涵盖数据采集、存储、处理、分析及可视化全流程的技术体系。其核心特征体现在“5V”维度：Volume（体量）要求存储系统具备PB级扩展能力，如Hadoop分布式文件系统（HDFS）通过多节点并行存储实现数据的高效管理；Velocity（速度）强调实时处理能力，ApacheKafka流处理框架可实现每秒百万级消息的实时捕获；Variety（多样性）涉及结构化与非结构化数据的融合处理，MongoDB等NoSQL数据库支持JSON、XML等多格式数据存储；Value（价值）通过机器学习算法挖掘数据价值，如TensorFlow框架在工程造价预测中可识别材料价格波动规律；Veracity（真实性）依赖数据清洗与校验机制，确保分析结果的可靠性。在工程造价领域，大数据技术已形成完整的应用生态<sup>[1]</sup>。广联达造价云平台整合了全国30个省份的工程数据，通过区块链技术实现数据不可篡改；北京城建集团开发的国际项目造价管理系统，利用Spark内存计算框架将投标报价分析时间从72小时压缩至8小时。这些实践表明，大数据技术正从单一工具向行业基础设施演进。

### 1.2 工程造价预测的基本原理

传统工程造价预测遵循“数据-模型-决策”的线性逻辑，

但存在三大局限：其一，数据维度单一，仅依赖历史造价指标，忽略市场波动、政策变化等动态因素；其二，模型适应性差，线性回归等统计方法难以处理非线性关系；其三，时效性不足，季度更新的造价指数无法反映实时市场变化。现代预测理论引入复杂系统思维，将工程造价视为多要素交互作用的动态系统。基于控制论的反馈机制，通过实时数据流修正预测偏差；运用灰色系统理论处理信息不完全性，在缺乏完整数据时仍可进行趋势推断；结合博弈论分析多方利益主体的决策互动，如业主、承包商、供应商在材料定价中的策略博弈。某商业综合体项目应用动态博弈模型后，将造价偏差率从12%降至4.3%。

### 1.3 大数据与工程造价预测的结合点

二者的融合体现在三个层面：数据维度扩展，整合BIM模型、物联网传感器、市场行情等多源数据，形成包含200+特征变量的数据集；分析方法革新，采用LSTM神经网络处理时间序列数据，在某地铁项目中准确预测出钢材价格波动对盾构机租赁成本的影响；决策模式转变，构建“预测-预警-干预”闭环系统，当混凝土用量偏离基准值15%时自动触发成本核查流程。具体技术实现包括：利用自然语言处理（NLP）解析政策文本，提取环保标准提升对造价的影响系数；通过图计算技术分析供应链网络，识别关键材料供应商的定价权；运用强化学习优化资源调配，在某超高层项目中减少塔吊闲置时间32%。这些创新使造价预测从经验驱动转向数据驱动。

## 2 基于大数据的工程造价影响因素分析

### 2.1 造价影响因素的分类与识别

影响因素可划分为四个维度：项目特征（建筑面积、建筑高度、结构类型等）、市场环境（材料价格、人工成本、机械租赁费）、政策法规（环保标准、税收政策、招投标规定）、管理效能（供应链协同、施工组

织、风险管控)。某跨海大桥项目数据显示,政策因素对造价的影响权重达27%,远高于传统认知<sup>[2]</sup>。识别方法包括:专家调查法,组织20+行业专家进行德尔菲法评分;文本挖掘,对10万份工程合同进行关键词提取,发现“赶工费”出现频率与工期压缩率呈正相关;因果推断,利用双重差分模型评估营改增政策对造价的影响,结果显示税率下调使建筑企业税负降低1.8个百分点。

## 2.2 基于大数据的因素关联性分析

关联性分析采用三种技术路线:统计关联,计算皮尔逊相关系数,发现混凝土强度等级与单方造价的相关性达0.89;机器学习关联,运用随机森林算法计算特征重要性,在某医院项目中识别出“手术室净化等级”是影响机电造价的关键因素;复杂网络关联,构建包含500+节点的造价影响网络,发现“钢材价格”通过“塔吊选型”间接影响“垂直运输费”。典型案例中,某住宅项目通过分析历史数据发现:当铝合金门窗K值从2.5降至1.8时,虽然材料成本增加8%,但因保温性能提升使空调能耗降低15%,综合造价反而下降3.2%。这种非线性关系揭示了传统单因素分析的局限性。

## 2.3 典型工程类型的关键影响因素提炼

不同工程类型的影响因素存在显著差异:住宅工程,户型标准化程度、装配式建筑应用率、精装修交付标准是核心因素;工业厂房,跨度、吊车吨位、工艺管道复杂度起决定性作用;基础设施工程,地质条件、征地拆迁难度、生态保护要求影响突出。以超高层建筑为例,其造价影响因素呈现“三高”特征:技术复杂度高,核心筒施工工艺选择影响结构造价20%以上;资源集中度高,塔吊、施工电梯等垂直运输设备的配置方案决定工期成本;风险敏感度高,台风、地震等极端天气对幕墙、机电系统造成额外防护成本。某600米级超高层项目通过大数据分析,优化了动臂塔吊的附着点布局,减少钢平台改造费用1200万元。

## 3 基于大数据的工程造价预测模型构建

### 3.1 模型构建的总体框架

模型架构采用“分层-模块化”设计:数据层,构建包含结构化数据(工程量清单)、半结构化数据(BIM模型)、非结构化数据(施工日志)的混合数据库;特征层,通过PCA降维将原始特征从200+压缩至30个关键变量;算法层,集成线性回归、支持向量机、神经网络等多种模型;应用层,开发造价预测API接口,与项目管理软件无缝对接。某省级造价站建立的预测系统包含四大模块:数据治理模块,实现GB级数据的秒级查询;特征工程模块,自动生成组合特征如“钢筋含量×混凝土强

度”;模型训练模块,支持TensorFlow、PyTorch等框架的并行计算;结果可视化模块,通过热力图展示各因素对造价的贡献度。

### 3.2 数据预处理流程

预处理流程包含六个关键步骤:数据清洗,采用孤立森林算法检测异常值,在某市政道路项目中识别出32处错误的路基压实度数据;缺失值处理,运用多重插补法填补缺失的混凝土配合比参数;数据转换,将分类变量(如结构类型)转换为独热编码;特征缩放,对连续变量(如建筑面积)进行Min-Max标准化;数据集成,通过ETL工具整合ERP、MES、CRM等系统数据;数据增强,利用SMOTE算法生成少数类样本,解决造价数据分布不均衡问题<sup>[3]</sup>。某商业综合体项目的数据预处理实践显示:经过清洗后,数据准确率从78%提升至99.2%;通过特征缩放,神经网络训练时间缩短60%;数据增强使模型对罕见工程类型的预测误差降低42%。

### 3.3 单一预测模型的构建与对比

常见单一模型包括:线性回归,适用于变量间呈强线性关系的场景,在某标准厂房项目中 $R^2$ 达0.92;支持向量机,通过核函数处理非线性关系,在某桥梁工程中MAE为3.8%;随机森林,可解释性强,在医院项目中识别出“净化空调系统”为首要影响因素;LSTM神经网络,擅长处理时间序列数据,在某地铁项目中准确预测出钢材价格波动对盾构机租赁成本的影响。模型对比实验表明:在数据量<1万条时,线性回归表现最优;数据量在1万-10万条时,随机森林综合性能最佳;数据量>10万条时,深度学习模型优势显著。

### 3.4 融合预测模型的设计

融合模型通过组合不同算法的优势提升预测精度。常见融合策略包括:加权融合,根据模型历史表现动态调整权重,在上海某住宅项目中,通过合理分配不同预测模型的权重,使RMSE(均方根误差)降低18%,有效提高了造价预测的准确性;栈式融合,将初级模型的输出作为次级模型的输入,在苏州某工业厂房项目中,先由基础模型进行初步预测,再将结果输入更复杂的模型进一步优化,成功将 $R^2$ (决定系数)提升至0.95,极大增强了模型的拟合优度;模型蒸馏,用大型模型指导小型模型训练,在移动端部署桥梁造价预测模型时,借助大型高性能模型的丰富知识,指导小型模型学习,使预测速度提升10倍。在港珠澳跨海大桥项目开发的融合模型包含三个层级:底层采用XGBoost处理结构化数据,中层运用CNN提取BIM模型特征,顶层通过注意力机制融合多源信息。该模型在台风季的预测实验中,将造价波动

预测准确率从67%提升至89%。

#### 4 基于大数据的工程造价预测模型优化策略

##### 4.1 数据质量优化策略

数据质量提升需构建“采集-清洗-监控”闭环体系：源头控制，制定《工程造价数据标准》，统一200+数据项的定义与格式；过程清洗，开发自动化清洗工具，在一个大型EPC项目中识别并修正1200余处数据错误；动态监控，建立数据质量评分卡，对供应商的数据准确率、完整性、及时性进行量化考核。省级某建设数据平台的数据治理实践显示：通过实施数据质量提升工程，历史项目数据的可用率从65%提升至92%，模型预测偏差率从8.3%降至3.1%。特别在材料价格数据方面，引入爬虫技术实时抓取电商平台价格，使钢材价格预测的MAE从120元/吨降至45元/吨。

##### 4.2 模型性能优化策略

性能优化从三个维度展开：算法优化，采用贝叶斯优化调整超参数，在秦山核电站项目中将神经网络训练时间从72小时压缩至18小时；架构优化，运用分布式计算框架，在集群环境中实现模型并行训练；部署优化，开发轻量化模型，通过模型剪枝将参数量从1000万降至200万，满足移动端实时预测需求。国内某知名造价咨询公司建立的模型优化体系包含四项机制：自动调参机制，基于Hyperopt库实现超参数自动搜索；模型版本管理，通过MLflow记录120+版本模型的性能指标；A/B测试机制，对新旧模型进行并行测试，确保升级决策的科学性；反馈迭代机制，将预测偏差反向输入模型训练，形成持续改进闭环<sup>[4]</sup>。

##### 4.3 全生命周期动态预测优化

全生命周期优化需构建“预测-预警-干预”三位一体系统：设计阶段，运用BIM+大数据技术进行多方案比

选，在上海浦东机场项目中优化航站楼结构形式，节省钢材用量12%；施工阶段，通过物联网传感器实时采集进度、成本数据，当混凝土用量偏离基准值15%时自动触发预警；运维阶段，建立设施管理数据库，为深圳某写字楼改造项目提供成本基准，使预算编制时间缩短60%。杭州某城市综合管廊项目开发了全生命周期造价管理系统：在设计阶段，利用遗传算法优化管廊断面布局，减少土方开挖量18%；在施工阶段，通过数字孪生技术模拟施工过程，提前识别出3处管线碰撞风险；在运维阶段，建立设备健康档案，预测电梯等关键设备的更换周期，使全生命周期成本降低22%。该系统应用后，项目获评国家级绿色施工示范工程。

#### 结束语

基于大数据的工程造价预测模型研究，为建筑行业成本管理提供了创新思路与方法。从基础理论到模型构建，再到优化策略，全方位展现了大数据在造价预测中的应用潜力。随着数字孪生、元宇宙等新兴技术涌现，未来造价预测将更智能精准、动态实时。建筑企业应积极拥抱大数据技术，加强数据治理与模型应用，提升成本管控水平，以适应行业变革，实现可持续发展。

#### 参考文献

- [1]姜华.基于大数据的工程造价预测模型研究[J].建筑与预算,2024(7):34-36.
- [2]王继龙.基于大数据分析的工程造价预测模型研究[J].门窗, 2025(14): 166-168.
- [3]程方圆.基于大数据分析的工程造价智能预测模型方法分析[J].办公自动化,2024,29(21):72-74.
- [4]马瑜琴.基于大数据的建筑工程造价精准预测与成本管控路径研究[J].城镇建设,2025(8):221-223.