

大数据分析在建筑施工成本控制中的应用案例研究

文 / 侯田汉 安徽建工交通航务集团有限公司

摘要：建筑施工成本控制面临数据分散、预测滞后、风险难控等现实挑战，大数据分析具备支撑动态感知与预测干预的能力。通过构建数据采集、清洗、特征工程、算法建模一体化流程，可提升施工阶段成本偏差识别精度。基于SVR与LSTM等算法建立的模型能精准跟踪成本变化趋势，辅助管理者实现资源分配优化与风险预判。潭冲河家园E地块EPC总承包工程案例验证了模型在实际施工场景下的有效性，系统运行稳定，预警响应及时，预算执行率明显提升，具备推广基础与价值。

关键词：施工成本；数据建模；偏差预测；现场应用；动态控制

【DOI】10.12254/j.issn.2096-6539.2025.15.031

引言

建筑施工成本控制一直是工程项目管理的重点与难点，受工期变化、材料价格波动、施工方案调整等多重因素影响，易出现预算偏离、资源浪费与资金使用效率低下的问题。传统成本管控依赖经验判断与事后分析，响应滞后、信息断层现象普遍，难以满足复杂工程场景下对实时性与精准性的要求。随着数据驱动理念在工程建设领域逐步渗透，施工现场产生的进度、材料、设备、人力等多源数据具备可量化、可分析、可预测的特性，为构建动态成本控制模型提供了基础条件。

一、成本偏差来源与结构特征

（一）成本构成要点

建筑施工成本由多个构成单元组成，主要包括直接成本、间接成本、措施成本、规费和税金五类。直接成本是项目执行的核心支出，涵盖人工费、材料费和机械使用费，其中人工成本受工期安排与人员结构影响较大，材料成本受采购批次、损耗与市场价格波动制约，机械费用依赖设备选型与运行效率。间接成本涉及管理、办公与临设投入，虽不直接产生产值但对整体控制具有重要影响，措施成本指模板支设、安全防护、基坑支护等临时工程，具有周期短、变动性强的特点。规费如排污费、保险费、文明施工费等受项目地点政策与工程规模约束，税金主要由增值税构成，根据合同约定与结算方式确定。成本构成中直接成本占比最大，约为70%，其后为措施与间接成本，不同类型与阶段工程中成本分布略有差异，但人工、材料和机械始终是波动最敏感的控制对象^[1]。

（二）偏差形成原因

施工成本偏差表现为计划成本与实际成本间的数值差异，具有多源性和动态性。计划管理方面，预算估算粗略或定额选取不当将导致成本基准偏离实际，图纸频繁变更亦会加重调整强度。资源配置方面，材料进退场失序、人员调配结构失衡、机械调度重复闲置等问题将直接引发单位消耗上升，尤其在多专业交叉与压缩工期

下更为突出。市场环境因素如原材料价格上涨、运输周期波动易使材料成本短期超出控制范围。现场执行方面，工艺不当、返工严重、作业协同效率低等情况将引发不可预期支出。进度控制不到位还会导致资源持续占用、日耗叠加，进而放大人工与设备费用^[2]。多因素交互叠加下，偏差呈现非线性扩散趋势，导致成本控制难度上升。

（三）偏差的累积影响

成本偏差不仅表现为数值层面的预算超支，更深层影响成本结构的动态重构过程。偏差一旦发生，首先会改变原有资源配置节奏，打破人工、材料与机械的协调性，造成施工阶段成本重心偏移。例如由于材料浪费增加，材料成本在总成本中占比上升，压缩其他项的可调配空间。当偏差未及时识别与干预，预算调剂操作将频繁发生，导致间接费用分摊比例随之变化，破坏原始成本结构的均衡性。措施费用中如安全防护与基坑支护反复返工，也会形成非工程量增长性的成本堆积^[3]。多项偏差联动下，项目的总成本结构从计划态转向失衡态，形成难以追溯的“累积成本转移链”。结构性偏差如果未通过预测机制提前识别并控制，将会放大风险识别难度和干预成本，降低后续控制策略的有效性。

二、成本分析模型的构建方法

（一）数据准备流程

建筑施工项目中成本数据涉及来源广、结构异、更新频，构建分析模型前需建立完整的数据准备体系。流程首步是数据源识别，明确结构化数据与非结构化数据的覆盖范围。结构化数据包括进度计划、材料清单、人工投入台账、预算与结算表、机械使用记录、采购合同等，通常来自项目管理系统、ERP平台或BIM系统。非结构化数据如施工日报、现场图片、监控视频、现场语音记录、文档扫描件等需通过文本识别、图像识别等手段转化为可计算格式。第二步为数据清洗与标准化，对重复数据、缺失项、异常值进行筛查修复，统一字段命名与单位格式，构建数据字典与字段映射规则^[4]。

第三步为数据整合，通过时间戳、工程量编码、任务编号实现多表关联，构建覆盖人材机资源、工序进度、工地分区的多维数据矩阵。第四步为数据抽样与分层标记，划分训练集、验证集与测试集，标注关键变量如成本异常点、波动区间与预测目标值，作为后续建模的数据基础。第五步为数据归一与特征工程，采用标准差归一、最小最大值缩放等方法对数值型变量进行归约处理，构建时序特征、滑窗特征、资源组合特征等模型感知变量，增强模型对多维特征的适应能力。最终输出结构清晰、标签完整、逻辑闭环的训练数据集，为模型构建提供稳定输入源。

(二) 模型构建策略

成本分析模型构建需兼顾预测精度、计算效率与实际可用性，策略设定以“偏差趋势识别+风险预警输出”为核心目标，采用多算法融合的技术路径构建模型体系。首先确定模型功能目标，分为成本趋势预测模型与风险异常识别模型两类。趋势预测模型用于对各阶段成本累计值、单位资源耗用成本进行滚动预判，主要采用支持向量回归(SVR)、随机森林回归、长短记忆神经网络(LSTM)等算法。SVR适用于短周期、样本稳定工程，随机森林适合处理特征维度高、数据结构复杂情境，LSTM用于捕捉施工周期中长期成本波动趋势。

风险识别模型用于识别施工过程中潜在的超支节点与高波动阶段，通常采用K-means聚类、自编码器异常检测(AutoEncoder)、孤立森林等方法，通过聚类中心偏移、重构误差分析识别风险区段。模型训练中采用交叉验证与网格搜索优化参数集，提高泛化能力。模型输出包括成本预测值、置信区间、风险等级评分与偏差点定位，辅助管理人员进行成本动态干预。为保证模型稳定运行，需构建模型更新机制，定期引入最新数据进行增量学习或全量重训，构建自适应迭代体系。同时结合可视化接口，将模型输出结果接入成本控制平台，通过图表、热力图、风险雷达等方式展示，支持决策闭环执行。通过算法集成、结构设计及接口适配，构建具有预测、预警与反馈功能的施工成本分析模型体系，满足动态施工环境下的成本控制应用需求。成本分析模型架构如下图1所示：

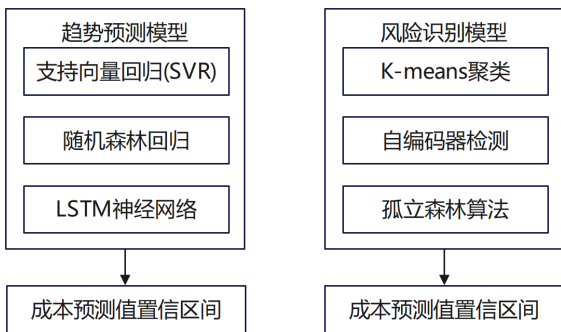


图1：成本分析模型架构

(三) 成本结构建模机制

动态施工环境中成本结构会随工序推进、资源投入节奏与外部市场环境不断变化而发生比例波动，模型若无法感知结构变化，预测结果将产生误差偏移，降低实用价值，构建具备结构适应能力的成本分析模型需在数据标签层与算法设计层进行双向改进。数据层面应在原有数值型字段基础上引入成本构成比例变量，如人工、材料、机械三类资源成本在各阶段占总成本的百分比，作为结构性参考特征进入模型输入维度。模型设计方面，通过引入组合特征构造机制，将如“人工×工期”“材料×天气”“机械×工作面”类型的交叉项作为输入增强结构关联表达能力。算法选择上采用XGBoost模型并结合SHAP可解释性工具，输出各特征对成本波动的影响贡献值，实现在不同时间节点下识别结构漂移的主导因子，构建成本结构随时间变化的敏感性路径图。

模型在训练过程中以阶段滚动窗口为基础，动态调整样本分布与特征权重分配，使模型不仅预测成本数值，还具备对成本构成变化趋势的识别能力。输出结果中除常规的预测曲线外，还形成结构偏移监测报告，标注当前周期中结构失衡趋势、波动风险等级与需关注的关键因素。该机制弥补了传统模型只识别静态偏差、忽视结构扰动的缺陷，提升了模型在多变场景下的鲁棒性与适应性，是实证验证中稳定预测能力的关键设计部分。

三、模型在实际项目中的应用分析

(一) 项目数据基础

选取的应用项目为安徽省肥西县的重点工程——潭冲河家园E地块EPC总承包工程，属大片区百姓安置重点工程。该项目施工周期长、专业工种多、工序复杂，成本控制涉及土建、装修、机电安装等多条施工主线。项目数据体系由“结构化业务数据+非结构化现场数据”构成。结构化数据来源于工程项目管理系统，涵盖预算执行表、进度计划、材料入库单、班组工日表、机械使用记录、采购合同与结算台账等，累计条目过多，按日同步更新。

非结构化数据来自施工日志、现场巡视图像、材料签收照片、质量整改单等，通过图像识别和OCR技术转化为可用于计算的标签化记录，统一编号后纳入模型数据库。项目部建立了数据服务器与中间数据库接口，支持边施工边采集，形成覆盖“人材机进度质量成本”五维的动态数据闭环。数据字段超过200项，数据更新周期最短为小时级，具备可用于模型训练与滚动预测的完整基础。

(二) 模型实施过程

模型部署阶段先由数据分析团队完成数据特征清洗与标注，通过标签分析构建时间序列特征、材料消耗指标、人工波动系数与机械稼动率等关键输入变量。采用SVR与LSTM双模型构建成本趋势预测体系，前者负责中短期施工任务预测，后者捕捉施工跨月与节点级成本偏移趋势。模型训练采用滚动滑窗与多折交叉验证策略，保证在不同工段与周期下的预测稳定性，系统部署在项目成本控制平台，嵌入BIM数据接口与计划调度平台。

现场技术人员通过移动端录入实际施工参数，系统自动调用模型接口实时输出预测值与偏差分布图，并生成任务级风险评分。当偏差评分超过设定阈值时，系统触发预警提示，反馈至施工管理人员端，实现决策闭环。模型运行过程实现全自动化推理、分区滚动更新，每24小时进行一次训练集补充与输出刷新，保持预测系统与施工进度同步演化。部署后的系统可兼容图形界面与表格双重输出模式，满足项目部现场、财务部预算与监理单位成本监管的多维需求。

(三) 成效与反馈分析

模型部署运行三个月内，项目成本控制系统监测到

的偏差率由原先均值8.4%下降至3.5%，月度预算执行率提升至95%以上，系统预警响应时间控制在10分钟以内。人工成本偏差检测准确率达91%，材料损耗预测偏移误差低于±4%，关键资源投入稳定性提升明显。管理团队反馈显示模型输出与传统经验判断相比，具备响应速度快、指标呈现直观、趋势识别提前量充足等优势，显著提高了决策前置性。特别是在室内室外同步施工阶段，通过模型分析提前识别某一区段材料超投风险，经动态调整班组配比与材料配送计划，实际成本控制优于原定目标近5%。

模型在机械调度方面反馈异常的稼动区域，引导施工单位调整施工窗口与作业重叠，月度机械空转率下降17%，模型集成后的系统运行稳定性强，支持多端接入，使用学习成本低，适配多岗位协同操作。后期反馈显示，该模型构建路径适用于具备完整信息化基础的中大型施工项目，特别是在大型基建工程中，具备可复制性与推广潜力。模型结构与算法参数也为后续优化升级提供了实际基础，形成“数据监测—模型判断—管理干预—成效反馈”的闭环控制机制。模型实施前后成本控制关键指标对比如下表1所示：

表 1：模型实施前后成本控制关键指标对比

指标类别	实施前数据表现	实施后数据表现	改善幅度
成本偏差率	8.40%	3.50%	↓ 4.9%
材料损耗率	5.80%	2.30%	↓ 3.5%
人工成本预测偏差	±12%	±5%	↓ 7%
预算执行率	85%	96%	↑ 11%
预警响应时间	48 小时	10 分钟	↓ 93%
成本调整周期	每月人工修订	每周自动预测更新	优化频次提高

结语

施工成本偏差来源多元，具有阶段性叠加与动态演化特征，传统管理依赖人工统计与事后修正，难以实现全过程精准控制，大数据分析模型结构清晰、响应及时，适配多源异构数据，能实时反馈偏差趋势，输出结果直观可解释。潭冲河家园E地块EPC总承包工程的实践证明模型在偏差识别、风险预警与辅助决策方面表现高效，显著提升资源配置效率与预算执行率。系统部署标准化程度高，接口与算法具可复制性，适用于多项目、多工种复杂场景，是推动智能化成本管控的重要路径。

参考文献

[1] 顾伟. 基于全过程动态管控视角的工程施工成本管理优化策略研究[J]. 珠江水运, 2025, (06): 32-34.

[2] 于检生. 智能化施工技术在高层建筑中的应用研究[J]. 科技与创新, 2025, (05): 211-214.

[3] 楚晶莹. 新型绿色建筑工程造价预算与成本控制研究[J]. 砖瓦, 2025, (03): 99-101.

[4] 秦艳丽. 建筑工程造价控制中施工项目成本核算对策[J]. 砖瓦, 2025, (03): 128-130.