

大数据背景下工程造价精准估算模型构建与应用

侯广冲

徐州方正会计师事务所有限公司 江苏 徐州 221000

摘要: 大数据背景下,工程造价精准估算对项目决策与管理意义重大。本文阐述大数据技术框架,分析造价数据特征,构建涵盖数据预处理、模型算法选择与优化、评估验证的精准估算模型。探讨模型在全生命周期造价管理、风险评估、协同管理等方面的应用,提出模型优化与持续改进策略,通过自适应性增强、数据质量提升、算法迭代创新,提升模型性能,为工程造价精准估算提供有效方法。

关键词: 大数据;工程造价;精准估算模型;模型应用;模型优化

引言:在工程建设领域,工程造价精准估算对项目投资决策、成本控制及资源合理配置至关重要。传统估算方法依赖人工经验,存在效率低、精度差等问题。大数据技术发展,为解决这些问题提供新途径。通过整合多源数据,运用机器学习等算法构建精准估算模型,可提升估算准确性与效率。

1 大数据技术框架与工程造价数据特征

1.1 大数据技术核心架构

数据采集层作为整个架构的基础,承担着多源数据整合的关键任务。在工程造价领域,采集的数据来源广泛,涵盖BIM模型蕴含的详细工程信息、物联网传感器实时反馈的施工动态数据、市场价格平台提供的各类材料价格波动数据,以及历史项目数据库积累的丰富经验数据。这些数据来源不同、格式各异,却共同为精准估算提供了全面支撑^[1]。数据处理层紧随其后,对采集到的原始数据进行深度加工。清洗环节去除数据中的噪声与错误信息,确保数据质量;标准化处理统一数据格式,消除因来源不同导致的差异;特征提取与融合则进一步挖掘数据价值,通过时间序列匹配将不同时间点的数据关联起来,语义映射使不同表述的数据含义一致,权重分配突出关键数据对估算结果的影响。模型构建层是核心所在,运用机器学习、深度学习与智能优化算法构建估算模型。回归分析用于探索变量间的线性关系,神经网络处理复杂的非线性问题,集成学习通过组合多个模型提升预测精度。应用展示层将模型结果直观呈现,为决策提供有力支持。动态仪表盘实时展示造价变化趋势,风险热力图清晰标识高风险成本项,预测曲线则对未来造价走向进行合理预判。

1.2 工程造价数据特征分析

工程造价数据类型丰富,既有结构化数据,如工程量、材料价格等,便于直接分析与计算;也有非结构化

数据,像设计图纸、合同文本等,蕴含着大量潜在信息。数据维度多元,时间维度反映历史造价变化规律,空间维度体现区域间造价差异,工程特征维度则关联建筑类型、施工工艺等因素对造价的影响。这些数据共同构成了一个动态演化的知识体系,其价值密度不均,核心的价格与工程量数据直接影响估算结果,而大量的过程文本与日志则间接反映成本动因。然而,这些数据也面临诸多挑战,噪声多影响数据准确性,缺失值频繁增加数据处理难度,格式不统一阻碍数据整合与共享。此外,数据的动态性与实时性特征日益显著,市场价格、政策法规的更新变化要求数据平台必须具备高效的同步与融合机制,以确保估算基础的时效性。

2 精准估算模型构建方法

2.1 数据预处理与特征工程

数据预处理是构建精准估算模型的基础环节。在数据清洗阶段,异常值检测通过统计方法或机器学习算法识别偏离正常范围的数据点,避免这些异常值对模型训练产生误导^[2]。缺失值填充则根据数据分布特征,采用均值填充、中位数填充或基于相似样本的插值方法,确保数据的完整性。噪声过滤环节运用滤波算法或数据平滑技术,消除数据中的随机波动,提升数据质量。特征提取阶段,统计分析方法通过计算均值、方差等统计量,捕捉数据的集中趋势与离散程度,为模型提供基础特征。聚类分析根据项目特征将相似项目归为一组,挖掘数据内在结构,辅助后续建模。文本挖掘技术针对合同条款等非结构化文本,提取关键信息并转化为结构化数据,扩展特征维度。特征选择环节聚焦于筛选对造价预测影响显著的关键因素。信息增益通过衡量特征与目标变量之间的关联程度,优先保留信息量大的特征。卡方检验分析特征与目标变量之间的独立性,剔除无关特征。相关性分析计算特征间的相关系数,避免高度相关

的特征导致模型过拟合,确保模型简洁高效。

2.2 模型算法选择与优化

模型算法的选择需结合数据特性与业务需求。监督学习算法中,线性回归通过拟合线性关系实现简单快速的造价预测,支持向量机则擅长处理非线性问题,提升预测精度。无监督学习算法中,主成分分析通过线性变换将高维数据投影到低维空间,实现数据降维,减少计算复杂度。聚类分析将项目划分为不同类别,为差异化建模提供依据。深度学习算法中,卷积神经网络凭借局部连接与权重共享特性,有效识别BIM模型中的结构特征,捕捉空间信息。循环神经网络通过记忆单元处理时间序列数据,适应造价随时间变化的动态特性。混合神经网络融合文本、图像、数值等多模态数据,综合利用不同类型数据的优势,提升模型鲁棒性与解释性。智能优化算法为模型参数调整提供高效搜索策略。遗传算法模拟自然选择过程,通过交叉、变异等操作实现参数全局搜索。粒子群算法借鉴群体行为规律,通过个体与群体的信息交互快速定位最优参数,提升模型性能。

2.3 模型评估与验证

模型评估指标是衡量模型性能的关键依据。均方根误差与平均绝对误差通过计算预测值与真实值之间的偏差,量化模型预测精度。决定系数反映模型对目标变量变异的解释程度,值越接近1表明模型拟合效果越好。例如,在某住宅项目群的成本预测中,优化后的模型在测试集上取得了RMSE为85.3元/平方米、MAE为62.7元/平方米的精度,其 R^2 值达到0.94,表明模型能解释绝大部分成本变异。为进一步确保模型决策的可靠性,还需引入更贴近业务的评估维度,如针对高风险成本项预测的精确率与召回率,以及在不同工程类型或造价区间上的稳定性表现。验证方法确保模型具备泛化能力。交叉验证将数据划分为训练集与测试集,通过多次训练验证评估模型稳定性。留一法验证在数据量较小的情况下,逐一留出样本进行验证,充分利用数据信息。稳定性分析测试模型在不同数据集与环境下的表现一致性,确保模型在实际应用中可靠运行。最终,模型需通过历史项目回溯测试与前瞻性试点应用的双重检验,才能正式投入工程实践,形成从实验室评估到现场验证的完整闭环。

3 模型应用场景与功能实现

3.1 全生命周期造价管理

在投资决策阶段,模型深度挖掘历史项目数据中的成本规律,结合当前市场趋势与行业动态,生成科学合理的预算方案。这一过程不仅考虑直接成本,还纳入间接费用与潜在风险成本,为项目可行性研究提供全面支

撑^[3]。设计优化阶段,模型实现BIM模型与造价数据的深度联动。通过自动解析BIM模型中的构件信息,精准计算工程量并关联对应造价数据,避免人工计算的误差与遗漏。成本敏感性分析功能可模拟不同设计参数变化对总造价的影响,辅助设计师在满足功能需求的前提下优化方案,实现成本与质量的平衡。施工控制阶段,模型接入物联网传感器与供应链系统,实时获取材料价格波动、施工进度偏差等动态数据。当材料价格超出预设阈值时,模型自动触发成本预警,并基于历史数据推荐替代材料或调整采购计划。进度偏差修正功能通过对比实际进度与计划进度,动态调整后续工序的资源分配,确保项目按期完成且成本可控。后期结算阶段,模型对实际造价与预测造价进行对比分析,验证模型准确性并提炼经验参数。这些参数反馈至模型数据库,持续优化后续项目的估算精度,形成“预测-实施-反馈-优化”的闭环管理体系。

3.2 风险评估与应对策略

风险因素识别环节,模型整合市场监测数据、政策文件与供应链信息,全面梳理可能影响造价的市场波动、政策调整、供应链中断等风险源。通过自然语言处理技术解析政策文本,提取与造价相关的关键条款,提前预判政策影响。风险预测模型构建阶段,模型联合外部经济指标、气候数据等多维度信息,构建多因素耦合的预测体系。例如,结合原材料价格指数与通货膨胀率预测材料成本走势,利用气候数据评估极端天气对施工进度影响,提升风险预测的前瞻性。动态调整机制根据风险等级自动触发预警与应对方案。低风险事件通过移动端推送提醒,中高风险事件则启动应急预案,如调整采购渠道、增加备用资源等,将风险影响降至最低。

3.3 协同管理与决策支持

跨部门数据互通通过统一数据标准与接口规范实现设计、施工、监理单位的数据共享。模型搭建云端协作平台,各参与方实时上传与获取数据,避免信息孤岛与重复工作,提升协同效率。平台内置的版本管理与权限控制机制,确保了数据流过程的可追溯性与安全性,为跨阶段、多主体的协同作业奠定了信任基础。公众参与式管理借助移动端反馈系统整合社会监督数据。公众可通过移动应用上报施工噪音、环境污染等问题,模型自动分类并推送至责任单位,增强项目透明度与责任感。这种开放式的数据收集渠道,不仅辅助了现场管理,也为评估项目的社会综合成本提供了新的数据视角。可视化决策界面以多维交互仪表盘为核心,集成造价趋势分析、影响因素敏感性分析等功能。决策者可通

过拖拽操作调整参数，实时查看不同情景下的造价变化，为投资决策提供直观依据。

4 模型优化与持续改进

4.1 自适应增强

在线学习机制赋予模型实时进化能力。当新数据持续流入时，模型并非简单叠加信息，而是通过增量学习算法动态调整内部参数与结构。这种在线迭代方式摒弃了传统“离线训练-部署应用”的静态模式，能够在不中断服务的前提下完成参数更新，有效解决了数据分布漂移带来的模型性能衰减问题^[4]。这种调整既保留历史知识，又快速适应数据分布变化，尤其在材料价格波动频繁或施工工艺更新的场景下，能及时捕捉关键特征并优化预测逻辑。跨项目知识迁移方法进一步突破单一项目的数据局限，实现经验复用。通过构建工程特征与造价指标的映射知识库，模型可提取不同项目间的共性规律，弱化项目个性化差异对估算精度的干扰。面对小样本新项目时，模型从历史项目中提取通用模式，构建跨项目知识库。通过迁移学习技术，将相似项目的成功经验快速适配到新场景，避免从零训练导致的过拟合风险。这种策略在区域差异显著或工程类型特殊的项目中表现尤为突出，显著提升模型在新环境中的泛化性能。

4.2 数据质量提升

数据增强技术破解稀缺数据困境。生成对抗网络通过生成器与判别器的博弈，合成与真实数据分布高度一致的虚拟样本。生成器基于真实造价数据的特征分布规律，模拟出涵盖极端价格波动、特殊工程结构等边缘场景的样本，判别器则通过迭代优化提升合成样本的真实性与有效性。这些样本不仅扩充数据规模，还覆盖原始数据中未出现的边缘场景，增强模型对极端情况的应对能力。例如在罕见材料价格预测中，合成数据可模拟市场突发波动时的价格走势，提升模型鲁棒性。噪声过滤算法精准识别异常数据。基于密度聚类的分析方法通过计算数据点在特征空间的局部密度，自动划分正常数据与离群点。该算法引入时间权重系数，对不同采集时段的数据赋予差异化权重，优先保留近期数据的有效性，

同时过滤因传感器故障、人工录入失误产生的孤立噪声点。该方法无需预设阈值，能适应不同数据分布特征，尤其擅长处理工程造价中因录入错误或设备故障产生的异常值，确保模型训练基于高质量数据。

4.3 算法迭代与创新

新型神经网络架构解锁工程关系建模。图神经网络将工程构件、施工工序等要素抽象为图结构节点，通过消息传递机制捕捉节点间的依赖关系。这种架构天然适配工程造价中复杂的空间与逻辑关联，例如在BIM模型解析中，可同时处理构件几何信息与施工顺序约束，提升工程量计算的准确性。混合智能模型融合物理规律与数据驱动优势。针对工程领域强物理约束的特点，模型将经典造价理论公式嵌入神经网络结构，形成“白箱+黑箱”的混合预测框架。这种设计既利用数据驱动方法挖掘隐藏规律，又通过物理约束避免违背工程常识的预测结果，在保证解释性的同时提升预测精度，为复杂工程场景提供可靠决策依据。

结束语

大数据背景下工程造价精准估算模型的构建与应用，有效提升了造价估算的精度与效率。通过全生命周期管理、风险评估及协同管理等多场景应用，模型为工程建设各阶段提供了有力支持。自适应增强、数据质量提升及算法迭代创新等优化策略，进一步保障了模型的可靠性与实用性。该模型在实际工程中的成功应用，为工程造价管理提供了新思路与方法，推动行业向智能化、精细化方向发展。

参考文献

- [1]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(5): 6-11.
- [2]张艳梅.大数据时代工程造价精准估算模型构建与应用[J].居舍,2025,(11):169-172.
- [3]吴兵.基于主成分回归法的建设工程造价估算模型研究[J].中国建筑金属结构,2024,23(12):19-21.
- [4]胡永波.大数据驱动的工程造价预测模型构建与优化研究[J].新城建科技,2025,34(01):180-182.