 DOI: https://doi.org/10.12349/edc.v8i5.10290

Construction and Application of Engineering Cost Forecasting Model in the Context of Big Data

Wenjian Gong

Beijing Jianzhida Engineering Management Co., Ltd. Anhui Branch, Hefei, Anhui, 230001, China

Abstract

With the rapid advancement of big data technology, traditional engineering cost forecasting methods have shown limitations in accuracy and adaptability, making them inadequate for complex and dynamic market environments. This study proposes an engineering cost forecasting model based on data mining and machine learning, integrating big data technologies. By analyzing key factors influencing project costs, the model combines multiple forecasting techniques to enhance prediction accuracy and efficiency. Research demonstrates that this big data-driven approach significantly improves forecasting precision, enabling construction entities, design firms, and contractors to better budget and control project costs, thereby providing robust support for successful project implementation.

Keywords

big data; project cost; prediction model; data mining; machine learning

大数据背景下工程造价预测模型构建与应用

龚文剑

北京建智达工程管理股份有限公司安徽分公司，中国·安徽 合肥 230001

摘　要

随着大数据技术的迅速发展，传统的工程造价预测方法存在精度不足、灵活性差等问题，难以应对复杂多变的市场环境。本文结合大数据技术，提出了一种基于数据挖掘和机器学习的工程造价预测模型。通过分析影响工程造价的关键因素，建立了一个集成多种预测方法的模型，以提高预测的准确性和效率。研究表明，基于大数据的工程造价预测模型能显著提升预测精度，帮助建设单位、设计单位以及承包商合理预算和控制工程造价，为工程项目的顺利实施提供有力支持。

关键词

大数据；工程造价；预测模型；数据挖掘；机器学习

1 引言

随着社会经济的发展，建设工程行业逐渐成为国家经济的重要组成部分。尤其在我国，建筑行业的持续增长推动了工程造价的逐步攀升。工程造价是建设项目的关键指标之一，它直接影响项目的盈利性与可行性。传统的工程造价预测方法往往依赖经验或基于简单的数学模型，精度有限，难以应对多变的工程环境。随着大数据技术的飞速发展，传统的造价预测方法逐渐无法满足日益复杂的工程需求。大数据背景下，工程造价预测模型能够通过采集和分析大量历史数据，发现潜在的影响因素及其关系，从而提高预测的准确性。数据挖掘技术和机器学习算法在工程造价预测中的应用为解决传统方法中的不足提供了新的思路和方向。本文将从大数据技术出发，探讨如何构建一个适应现代工程需求的造价预测模型，并分析其实际应用效果。

2 工程造价预测的理论基础

2.1 工程造价的定义与构成

工程造价是工程项目从设计、采购到施工全过程中所发生费用的综合反映，其构成包括人工费、材料费、机械设备使用费、管理费、利润及税金等多种费用要素。造价水平不仅受到项目规模、工期安排、地理条件和施工难度的影响，也与市场环境、政策调控及技术条件密切相关。精准的造价预测能够为预算编制、成本控制和施工管理提供重要依据，是工程管理体系中的关键环节。随着工程项目规模不断扩大、结构日趋复杂，人工经验判断已难以满足精细化管理的需求。传统估算方式在应对复杂工程时易出现偏差，难以支撑现代工程建设的高效决策。因此，引入大数据技术和机器学习方法以提高造价预测的科学性与精度，已成为工程造价研究的重要发展趋势。

【作者简介】龚文剑（1981–），男，中国江苏金湖人，本科，工程师，从事工程造价研究。

2.2 传统造价预测方法的局限性

传统工程造价预测方法主要包括类比法、经验法与计量法，这些方法在早期工程管理中发挥了重要作用，但其局限性日益突出。类比法依赖历史项目的相似性，预测易受主观判断影响；经验法依托专家经验，在复杂情境下受个人认知差异影响明显；计量法虽然较为精细，但需要大量准确的基础数据，且计算过程繁琐，难以适应工程项目快速变化的特性。更为关键的是，这些方法大多采用线性推断或少量变量分析，难以处理工程造价中存在的多维度、非线性关系。同时，传统方法不具备及时更新能力，无法快速响应市场材料价格波动、政策变化等外部因素。从现代工程管理需求来看，传统方法难以满足高精度、高效率的造价预测要求，亟须引入更加智能的数据驱动方法。

2.3 大数据技术在造价预测中的应用前景

大数据技术在工程造价预测中的应用为解决传统方法的不足提供了新的思路。通过对海量历史项目数据、市场行情信息、施工过程数据等多源数据进行整合分析，可以全面揭示造价变化的内在规律。机器学习和数据挖掘技术能够在高维数据中自动识别关键影响因素，捕捉隐藏在复杂关系中的造价波动模式，提高预测模型的准确性和稳定性。在大数据支持下，造价预测不仅可以基于历史数据进行静态估算，还能结合实时市场动态、政策变化与项目实际进展进行动态预测，实现预测结果的及时更新与持续优化。这种基于数据驱动的预测方式具有更高的适应性和科学性，为现代工程造价管理提供了有效的技术支撑。未来，随着数据规模进一步扩大和算法不断优化，大数据驱动的造价预测模型将在工程决策与成本控制中发挥更加重要的作用。

3 工程造价预测模型的构建

3.1 数据采集与预处理

数据采集与预处理是工程造价预测模型构建的基础环节，其质量直接影响模型的训练效果与预测精度。工程造价相关数据来源复杂，包括财务投入、材料与设备价格、施工进度、技术方案及项目管理信息等多维度内容，且数据结构呈现非结构化或半结构化特点。为了提高数据的可用性，需要对采集到的原始数据进行系统化预处理。数据清洗是其中至关重要的步骤，通过剔除噪声数据、填补缺失值与识别异常点，使数据更加真实可靠。在此基础上，还需对数据进行标准化和归一化处理，以消除不同量纲带来的影响，增强模型的稳定性。此外，对时间序列数据及类别数据进行合理编码，可有效提升后续特征提取与建模的效率。完善的数据预处理流程不仅确保输入数据的质量，也为模型预测性能的提升奠定坚实基础。

3.2 特征工程

特征工程在工程造价预测中起到决定模型表现的重要作用，其核心任务是从大量原始数据中提取能够反映造价变化规律的关键因素。工程造价受到项目规模、结构复杂度、地质条件、所用材料、工期安排、施工工艺及管理水平等多维因素影响，因此需要通过特征分析识别造价波动的主要驱动变量。常见的方法包括统计相关分析、主成分分析等，可用于判断各特征与造价之间的关系强度。在此基础上，借助决策树、随机森林等机器学习特征选择算法，可进一步筛选出对预测结果贡献度最高的特征，使模型在保证精度的同时减少冗余信息，提高计算效率。特征工程不仅优化模型结构，也使其具有更强的解释性，为后续预测提供更可靠的数据支撑。

3.3 预测模型的选择与构建

工程造价预测模型的构建需要综合考虑数据特点、任务需求与模型性能。常见方法包括回归分析、支持向量机、神经网络、决策树与随机森林等。回归分析模型结构简单、具有较强解释性，适用于线性关系明显的造价数据；支持向量机适合处理高维与非线性问题，但计算复杂度较高；神经网络可自动学习复杂模式，在大规模数据背景下表现优异；决策树具有良好的可解释性，而随机森林通过集成策略显著提高预测稳定性与精度。针对工程造价数据变量多、非线性特征突出的特点，本文构建了随机森林与神经网络相结合的混合模型，以兼具随机森林的稳定性与神经网络的表达能力。实验结果显示，混合模型在预测精度、泛化能力与计算效率方面均优于单一模型，能够更准确地反映工程造价的变化趋势，为工程管理决策提供有效支持。

4 工程造价预测模型的应用

4.1 模型验证与实验设计

为检验工程造价预测模型的有效性，本文选取多个具有代表性的工程项目数据集开展实验研究。实验采用训练集与测试集分离的设计方式，通过对历史造价数据进行模型训练，使模型能够学习造价变动规律；随后在测试集上进行独立验证，以评估其泛化能力与预测精度。实验中设置多个对照模型，包括传统统计方法和常规机器学习模型，以便从不同维度评价所构建混合模型的性能。结果显示，该模型在均方误差、预测偏差率和趋势拟合度等指标上均表现更优，尤其在结构复杂、成本构成多样的工程项目中，能够有效捕捉造价的变化趋势，显著提升预测的稳定性与可靠性。实验结果表明，基于大数据驱动的混合模型在工程造价预测场景中具有明显的优势与应用潜力。

4.2 实际应用案例分析

为进一步验证模型在真实情境中的适用性，本文选取某大型建筑工程项目进行案例分析，将构建的预测模型应用于其造价预测任务，并与项目竣工后的实际造价数据进行对比。案例研究结果表明，预测值与实际造价之间的偏差处于合理可控范围内，说明模型不仅在实验环境中表现出较高精度，在工程实践中同样具有较强的可靠性。此外，模型可引

入实时市场数据，如材料价格指数、劳务成本变化及政策调整信息，通过动态更新预测结果，使造价预测具备时效性和自适应特征。案例分析证明，模型能够在工程决策、预算制定和风险管理等环节提供有效支持，提高造价测算工作的科学性与精细化水平。

4.3 模型的应用前景

随着大数据技术和智能算法的持续发展，工程造价预测模型的应用前景愈发广阔。在未来工程建设领域，项目数据的种类与规模将不断扩大，使模型能够从更丰富的数据维度中提取规律，从而进一步提升预测准确性。同时，深度学习、知识图谱、自适应学习等先进技术的融入，将为造价预测模型带来更强的表达能力，使其能够处理更复杂的数据结构与成本影响因素。模型还可扩展至全过程造价管理，实现从方案设计、施工组织到决算审核的全周期动态监测与预测。此外，与 BIM、智慧工地系统的融合将推动模型在数字化建设环境中的深度应用，为工程成本控制提供更加智能化与自动化的技术支撑。总体来看，该类模型将在工程管理领域发挥越来越重要的作用，为提升造价控制水平和优化资源配置提供长期价值。

5 工程造价预测模型的挑战与未来发展

5.1 数据质量与隐私保护问题

在工程造价预测模型的构建过程中，数据质量直接决定模型训练的有效性与预测结果的可靠性。大数据环境中数据来源复杂，存在缺失、噪声、冗余等问题，若未经过严格筛选与清洗，可能导致模型偏差增大、预测稳定性下降。同时，工程造价相关数据涉及企业经营信息、项目合同内容与人员配置等敏感内容，隐私泄露风险不容忽视。如何在保证数据可用性的同时实现隐私保护，是未来应用中必须解决的关键问题。可采用数据脱敏、匿名化处理、差分隐私等技术降低敏感数据暴露概率，并通过建立数据质量评估体系提升数据的完整性与准确性。此外，制定严格的数据管理制度与权限控制措施，有助于在数据共享与安全保护之间取得平衡，为高质量预测模型提供可靠的数据基础。

5.2 模型的适应性与可解释性

尽管机器学习模型在工程造价预测中表现出较高的精度与泛化能力，但其内部结构复杂、决策逻辑难以直接呈现，使模型在实际应用中常被视为“黑箱”，限制了结果的可信度与推广应用。工程造价决策高度依赖透明的逻辑推断，因此提升模型的可解释性成为未来研究的重要方向。可通过引入可解释人工智能技术（如特征贡献分析、规则提取、可视化决策路径等）增强模型行为的透明度，使决策者能够理解预测结果的依据。此外，工程项目类型多样、成本构成复杂，模型需要具备较强的适应性，以便在不同地域、不同规模与不同政策环境下保持稳定性能。通过动态更新模型参数、构建迁移学习框架与集成多模型结构，可提高模型对变化环境的适应能力，为工程造价管理提供更可靠的技术支撑。

5.3 跨领域数据融合

工程造价预测不仅依赖建筑工程本体数据，还受到宏观经济环境、材料市场波动、政策调控力度等多领域因素影响，因此跨领域数据的融合能力将决定预测模型的全面性与深度。单一来源的数据往往无法反映造价变化的复杂机制，而跨领域数据的差异性、结构差别与格式不统一，使融合过程面临较大挑战。未来研究应重点发展多源异构数据融合技术，通过构建统一的数据标准、建立多维数据关联机制与采用图模型、知识图谱等结构化工具，实现经济、市场、政策与工程数据之间的深层关联。此外，随着实时数据获取能力的增强，动态数据流的融合将进一步提升造价预测的时效性与敏感度。多领域数据的有效整合不仅有助于提高模型的准确性，也能为工程造价管理提供更加全面的辅助决策依据。

6 结语

随着大数据技术在工程建设领域的深入应用，工程造价预测正逐步迈向数据驱动的智能化阶段。基于大数据构建的预测模型能够从海量工程信息中提取造价变化规律，显著提升预测的精度与时效性，为项目决策、预算编制和成本控制提供可靠依据。数据挖掘与机器学习技术的引入，使模型能够处理复杂的非线性关系，突破了传统造价预测方法在维度单一、更新滞后和人工依赖方面的局限。未来，工程行业将积累更多可用于分析的高质量数据，预测模型也将不断优化与迭代，形成更具泛化能力和稳定性的造价预测体系。可以预见，大数据驱动的预测方法将成为工程造价管理的核心工具，为推动行业数字化与精细化发展提供重要支撑。

参考文献

[1] 胥筱娅.大数据背景下工程造价数据挖掘与成本预测方法探析[J].城市建设,2025,(28):23-25.

[2] 陈龙.基于大数据的工程造价预测模型构建[J].中华建设,2025,(11):64-66.

[3] 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.

[4] 张柯锐.基于大数据分析的工程造价预测模型与应用[J].中国建筑金属结构,2025,24(11):13-15.

[5] 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,(07):34-36.