

大数据时代下工程造价预测模型的构建与实证研究

文 / 肖琦 中铁房地产集团（贵州）有限公司 贵州贵阳 550000

【摘要】随着大数据时代的到来，工程造价预测模型的构建和实证研究成为工程管理领域的热点问题。本文以大数据时代下工程造价预测模型的构建与实证研究为研究对象，探讨了数据收集与处理、模型选择与优化的关键步骤，并通过实证数据与样本选择、模型验证与评估、模型应用与案例分析展开了深入研究。研究结果表明，在大数据时代背景下，构建的工程造价预测模型具有良好的准确性和适用性，为工程管理决策提供了可靠的支持。

【关键词】工程造价预测模型；大数据时代；数据处理；模型优化；实证研究

【DOI】10.12334/j.issn.1002-8536.2024.12.036

引言：

在大数据时代，工程造价预测模型的构建对于工程管理具有重要意义。大数据的涌现为工程造价预测提供了更多的信息和可能性，然而如何有效地利用这些数据并构建可靠的预测模型仍然是一个挑战。

1. 大数据时代下工程造价预测模型的重要性

在当今的大数据时代，工程造价预测模型成为了工程管理领域中不可或缺的一部分。随着数据量的爆炸式增长，工程造价预测模型的重要性愈发凸显，它能够利用大数据技术，对海量的工程数据进行分析和挖掘，为工程项目的管理和决策提供更为科学、精准的依据^[1]。首先，大数据的涌现为工程造价预测提供了前所未有的数据来源。传统的工程造价预测往往依赖于有限的历史数据和经验判断，难以全面反映工程项目的实际情况和市场变化。而现在，

借助大数据技术，可以获取到更为丰富和多样的数据，包括工程项目的成本构成、材料价格、人工费用、施工进度等各个方面，从而更准确地捕捉工程造价的变化规律和趋势。其次，通过对大规模数据的分析和挖掘，工程造价预测模型能够更好地应对复杂多变的市场环境和工程项目特征。在工程项目的实施过程中，市场环境的变化、工程设计的调整、施工进度的延误等因素都可能对工程造价产生影响。而利用大数据分析技术，能够对各种因素进行全面的分析和预测，更好地把握市场动态和工程项目的实际情况，为工程决策提供更为科学、精准的参考依据。此外，大数据技术的运用也为工程造价预测模型的构建和优化提供了更为强大的工具和方法。传统的工程造价预测模型往往依赖于人工建模和经验判断，难以应对复杂多变的市场环境和工程项目特征。而现在，随着机器学习、深度学习等人工智能技术的发展，可以构建更为智能、自适应的预

测模型,使其能够更好地理解和预测工程造价的变化趋势。

2. 大数据时代下工程造价预测模型的构建

2.1 数据收集与处理

2.1.1 数据来源的多样性与获取方式

在大数据时代,工程造价预测模型的构建离不开多样化的数据来源和获取方式。数据可以从多个渠道获取,包括历史工程项目数据、政府公开数据、行业报告、专业机构数据、以及企业内部数据等。这些数据来源的多样性为构建工程造价预测模型提供了丰富的信息基础,有助于提高模型的全面性和准确性。数据的获取方式也多种多样,包括数据采集、数据购买、数据共享、以及数据挖掘等手段。通过多样的获取方式,可以获取到更加全面和真实的数据,为工程造价预测模型的构建提供更可靠的数据支持^[2]。因此,数据来源的多样性与获取方式的灵活运用是工程造价预测模型构建中不可或缺的重要环节,也是确保模型数据质量和可靠性的关键一步。

2.1.2 数据质量的评估与筛选标准

在构建工程造价预测模型的过程中,数据质量评估包括数据完整性、准确性、一致性和时效性等方面。针对数据完整性,需要确保所收集的数据涵盖了工程项目的各个关键方面,包括成本、资源投入、工程规模等,以及相关影响因素的全面性。数据的准确性则要求对数据进行验证和核实,避免因数据错误导致的预测偏差。此外,数据的一致性也是评估的重要指标,确保数据在不同来源和不同时间点上的一致性,以避免数据冲突和混乱。同时,数据的时效性也需要被考虑,特别是对于长期工程项目的数据,需要确保数据的时效性以反映当前的市场和环境情况。在筛选数据时,需要根据以上评估标准对数据进行逐步筛选,剔除低质量和不一致的数据,以确保最终用于模型构建和分析的数据集是高质量、可靠的,从而提高工程造价预测模型的准确性和可靠性。

2.1.3 数据清洗与异常值处理

在大数据时代下构建工程造价预测模型时,数据清洗与异常值处理是至关重要的步骤。数据清洗包括缺失数据处理、重复数据处理和数据格式统一等,旨在确保数据的完整性和一致性。通过对缺失数据进行填充或删除,可以减少对模型建立和预测的干扰。同时,处理重复数据可以避免对模型分析和训练过程的影响。异常值处理则旨在识别和处理数据中的异常数值,以避免异常值对模型的扭曲和误导。通过采用统计方法、可视化分析和领域知识相结合的方式,可以有效识别和处理异常值,保证模型的稳健

性和准确性。

2.1.4 数据转换与特征工程的实施

在大数据时代下构建工程造价预测模型时,数据转换包括对数据进行标准化、归一化、离散化等处理,以确保数据的可比性和一致性,为模型的训练和预测提供更好的数据基础。同时,特征工程则涉及对原始数据进行特征提取、特征选择和特征构建等操作,旨在挖掘数据中的潜在信息和规律,为模型提供更具代表性和预测能力的特征。通过特征工程的实施,可以提高模型的泛化能力和预测准确性,为工程造价预测模型的构建奠定坚实基础。

2.2 模型选择与优化

2.2.1 机器学习模型的比较与选择

针对工程造价预测,可以考虑多种机器学习模型,如线性回归、决策树、随机森林、支持向量机、神经网络等。在比较与选择模型时,需要考虑模型的预测性能、可解释性、计算效率以及对大规模数据的适用性。不同模型具有各自的特点和适用场景,因此需要综合考虑模型的优缺点,并结合实际需求 and 数据特征进行选择。此外,还需考虑模型的泛化能力、过拟合和欠拟合问题,以及模型的参数调优和性能评估等方面,以确保所选模型能够在实际工程造价预测中取得良好的效果。机器学习模型的比较与选择是构建工程造价预测模型中至关重要的环节,直接影响模型的预测准确性和实际应用效果。

2.2.2 深度学习模型的可行性分析

在大数据时代下构建工程造价预测模型时,进行深度学习模型的可行性分析至关重要。深度学习模型以其对大规模数据的学习能力和对复杂特征的提取能力而闻名,对于工程造价预测也具有潜在的应用前景。然而,深度学习模型也面临着数据需求量大、计算资源消耗大、模型解释性较弱等挑战。在进行可行性分析时,需要充分考虑工程造价预测的数据特点、数据规模以及业务需求。同时,还需要评估深度学习模型在当前场景下是否能够带来明显的性能提升,并且是否值得投入大量的数据和计算资源。此外,还需要考虑深度学习模型的训练时间、模型调参的复杂性,以及模型结果的可解释性对工程预测实际应用的影响。综合考量之后,才能够准确评估深度学习模型在工程造价预测中的可行性,并决定是否采用深度学习模型作为最终的预测工具。

2.2.3 参数调整与模型性能优化策略

在进行模型选择后,还需要进行参数调整以优化模型的性能。这不仅是为了提高模型的泛化能力,更是为了提高模型的预测准确性。为了达到这个目的,可以采用交叉

验证等技术来选择最佳的参数组合。具体来说，对于神经网络模型，可以调整学习率、采用正则化、批量归一化等手段来提升模型性能。这些优化策略可以更好地训练模型，使其在训练和测试数据上都能取得更好的表现。另外，在模型性能优化过程中，还需要特别注意过拟合和欠拟合问题。过拟合是指模型在训练数据上表现很好，但在测试数据上表现较差；而欠拟合则是指模型在训练数据和测试数据上都表现较差。为了解决这些问题，可以采用早停法等技术来防止过拟合，或者采用增加模型复杂度、增加特征等方法来处理欠拟合问题。还需要考虑模型在实际应用中的稳定性和可靠性。一个好的模型不仅应该在训练数据上表现良好，还应该在不同的场景下都能保持稳定的性能。因此，在模型训练过程中，需要对模型进行充分的测试和验证，以确保其在实际情况下的稳定性和可靠性。

2.2.4 模型稳定性与可解释性的平衡探讨

模型的稳定性指的是模型在面对不同数据分布或输入变化时能够保持良好的预测性能，而可解释性则是指模型能够清晰地解释预测结果的原因和逻辑。在实际应用中，稳定性和可解释性往往是相互制约的。为了在模型稳定性与可解释性之间取得平衡，可以采取多种策略。例如，在特征工程阶段可以选择那些对模型预测具有较强解释性的特征，并且对数据进行合理的转换和处理，以提高模型的可解释性；同时，在模型选择阶段可以优先考虑一些对数据分布变化不敏感的模型结构，以提高模型的稳定性。此外，还可以采用集成学习的方法，综合多个模型的预测结果，以平衡稳定性和可解释性的需求。最终的平衡取决于实际应用场景和需求，需要在模型构建过程中综合考量。在构建工程造价预测模型时，平衡模型的稳定性与可解释性是一个需要认真思考和权衡的问题，关系到模型在实际工程预测中的可靠性和可应用性。

3. 大数据时代下工程造价预测模型的实证研究

3.1 实证数据与样本选择

3.1.1 实证数据的来源与多样性

在构建工程造价预测模型时，实证数据的来源和多样性对模型的准确性和适用性起着至关重要的作用。实证数据的来源应当尽可能全面，可以包括历史工程项目的造价数据、相关行业报告数据以及政府部门发布的工程造价统计数据等。同时，数据的多样性也至关重要，应涵盖不同地区、不同规模和不同类型的工程项目数据，以确保模型具有较强的泛化能力和适应性。通过综合利用多来源、多样性的实证数据，可以更准确地反映工程造价的实际情况，

为模型的构建和应用提供坚实的数据基础^[3]。

3.1.2 样本选择的合理性与代表性

在构建预测模型时，样本选择的合理性及其代表性直接关系到模型的可靠性和泛化能力。一个合理的样本选择不仅需要考虑到整个数据集的特征和分布情况，还需要考虑到各种工程类型、地域和规模等因素。只有这样，才能确保模型具有更好的适用性和泛化能力。一个具有代表性的样本能够充分反映整个工程造价数据的特征和分布情况，从而使得模型能够更好地学习和预测未来的数据。不同的工程类型、地域和规模可能会对工程造价产生不同的影响，因此在选择样本时需要综合考虑这些因素。这样可以确保模型在面对不同类型、地域和规模的工程时都能够表现出较好的预测性能。充足且高质量的样本能够提高模型的稳定性和准确性。在选择样本时，需要根据实际情况来确定样本的数量，并尽可能选择质量较高的样本。例如，如果数据集中存在一些异常值或缺失值，需要对这些数据进行处理或剔除，以确保样本的质量。

3.2 模型验证与评估

3.2.1 模型预测结果与实际数据的对比分析

模型预测结果与实际数据的对比分析是验证模型准确性和泛化能力的关键步骤。通过将模型的预测结果与实际工程项目数据进行对比，可以评估模型的预测精度和可靠性。这种对比分析有助于发现模型预测中的偏差和误差，从而为模型的改进和优化提供重要的指导。通过对比分析，研究人员可以深入了解模型在不同工程项目中的预测表现，发现模型在特定条件下的优势和局限性，为模型的实际应用提供可靠的支持。这种对比分析也为工程管理决策提供了可靠的依据，帮助项目管理者更准确地评估工程项目的成本，优化资源配置，降低风险，提高项目的执行效率和成功率。因此，模型预测结果与实际数据的对比分析不仅是工程造价预测模型研究的重要环节，也是推动工程管理实践向数据驱动决策转变的关键一步。

3.2.2 模型预测误差的评估与分析

在大数据时代，工程造价预测模型的验证与评估至关重要。模型预测误差的评估与分析是确保模型准确性和可靠性的关键环节。通过对模型预测结果与实际数据之间的误差进行评估与分析，可以全面了解模型的预测偏差和稳定性。常用的误差评估指标包括均方误差（MSE）、平均绝对误差（MAE）、均方根误差（RMSE）等。这些指标的评估与分析有助于揭示模型的预测表现，帮助识别模型在不同情况下的表现差异，为模型改进提供指导。此外，对模型预测误差进行深入分析还能帮助发现造成误差

的潜在原因，例如特定的工程类型、地理位置、材料成本波动等因素，从而为模型的进一步优化和定制提供重要参考。模型预测误差的评估与分析不仅有助于提高模型的预测精度和稳定性，也为工程管理决策提供了可靠的依据，帮助项目管理者更准确地评估工程项目的成本，优化资源配置，降低风险，提高项目的执行效率和成功率。

3.3 模型应用与案例分析

3.3.1 模型在实际工程项目中的应用与效果展示

将模型应用于实际工程项目中，并展示其效果可以验证模型的实用性和可靠性。通过对多个实际工程项目的案例展示，可以清晰展现模型在不同工程类型和规模下的适用性和预测效果。这种案例分析有助于工程管理者及决策者深入了解模型在实际项目中的应用情况，从而增强对模型的信任度。同时，通过案例分析还可以揭示模型在实际工程项目中取得的成效，例如在成本控制、资源优化、风险管理等方面的效果，为工程管理实践提供可靠的依据^[4]。

3.3.2 模型在不同工程类型中的适用性分析

在大数据时代，针对不同工程类型，包括住宅建筑、商业建筑、基础设施等，对模型的适用性进行分析可以揭示模型在不同领域的预测能力和局限性。通过对不同工程类型的数据进行模型验证与评估，可以发现模型在不同领域中的表现差异，进而为工程项目的成本预测提供更精准的依据。适用性分析还有助于识别特定工程类型下的影响因素和特征，为模型的定制和优化提供指导。此外，对模型在不同工程类型中的适用性进行分析还能帮助工程管理者及决策者更好地理解模型在实际项目中的应用范围，为不同工程类型的成本管理和决策提供支持。

4. 大数据时代下工程造价预测模型评价与改进

4.1 模型准确性评价

在评价工程造价预测模型的准确性时，需要综合考虑模型的预测精度、偏差和稳定性。通过与实际数据的对比分析，可以评估模型的预测准确性，包括平均绝对误差、均方误差和相关系数等指标^[5]。此外，对于不同工程类型和规模的预测结果进行分析，可以更全面地评价模型的准确性。在评价准确性时，还需关注模型的预测范围和置信区间，以确保模型在不同情况下均具有可靠的预测能力。

4.2 模型适用性评价

评价工程造价预测模型的适用性需要考虑模型在不同工程类型、不同地域和不同规模下的适用情况。模型的适用性评价还需要关注模型的灵活性和可扩展性，即

模型是否能够适应新的数据和新的情境，并具有一定的泛化能力。此外，模型的预测结果是否符合工程实际需求，是否能够为工程管理决策提供有益的参考，也是评价适用性的重要考量因素。综合考虑这些方面，可以全面评价工程造价预测模型的适用性，并为模型的改进和推广提供指导。

4.3 模型改进措施

为进一步提高工程造价预测模型的准确性和适用性，可以考虑以下改进措施：针对数据质量和多样性，可以采用更加全面和精细的数据清洗和特征工程方法，以提高数据的质量和信量；可以结合领域知识和专家经验，进一步优化模型的选择和参数调整，以提高模型的预测能力和稳定性；可以通过引入新的数据源和应用新的建模技术，不断完善和更新模型，以适应工程管理实践中的新需求和挑战^[6]。

结语：

综上所述，本研究在大数据时代下构建了适应工程造价预测需求的模型，并通过实证研究验证了其在工程管理决策中的有效性和实用性。未来的研究将继续关注工程造价预测模型的改进与优化，进一步提高模型的准确性和适用性，为工程管理决策提供更可靠的支持。

参考文献：

- [1] 董娜, 卢泗化, 熊峰. 大数据背景下基于 ABC-SVM 的建筑工程造价预测 [J]. 技术经济, 2021, 40(08): 25-32.
- [2] 刘哲, 何子东, 靳健欣, 等. 大数据技术在输变电工程造价指标预测中的应用研究 [J]. 微型电脑应用, 2021, 37(05): 154-157.
- [3] 俞敏, 王愿翔, 闫园, 等. 架空线路改造工程造价的组合预测方法 [J]. 电力科学与技术学报, 2020, 35(01): 24-30.
- [4] 荣鹏, 张寒, 代勇, 等. 电网基建工程造价预测模型构建方法 [J]. 电子技术与软件工程, 2019(24): 137-138.
- [5] 妙旭娟, 刘锦明, 高亮, 等. 基于主成分分析法和神经网络的技改工程造价预测模型 [J]. 内蒙古科技与经济, 2019(19): 37-40+42.
- [6] 王轩, 杨文波. 建筑工程造价预测模型的设计与实现 [J]. 微型电脑应用, 2019, 35(08): 95-97+104.