

# 基于大数据的建筑工程造价预测模型研究

马 磊

河南省三门峡市渑池县国有渑池林场 河南 三门峡 472400

**摘要:** 在数字化时代背景下,大数据技术的崛起对建筑工程造价管理领域产生了深远的影响。通过大数据的深度挖掘和分析,能够揭示传统方法难以触及的造价规律和趋势,为建筑工程的决策提供更为精准和前瞻性的支持。本文深入探讨了基于大数据的建筑工程造价预测模型的构建过程、应用前景以及面临的挑战,旨在为行业内的理论研究和实践应用提供有价值的参考。

**关键词:** 大数据; 建筑工程; 造价预测; 模型研究; 数据挖掘

## 引言

随着全球经济一体化和市场竞争的加剧,建筑工程造价管理的复杂性和挑战性日益凸显。传统的造价管理方法受限于数据处理能力和分析手段,难以适应快速变化的市场需求和项目特点。而大数据技术的兴起,以其强大的数据处理和分析能力,为建筑工程造价管理提供了新的解决思路。本文旨在系统研究基于大数据的建筑工程造价预测模型的构建与应用,以期提升工程造价管理水平提供科学依据和实践指导。

### 1 大数据在建筑工程造价管理中的应用潜力

一是数据挖掘与知识发现:通过深度挖掘和分析海量的建筑工程数据,可以发现隐藏在数据背后的知识、规律和趋势,为造价预测提供更为准确和全面的信息支持。二是实时动态监控:大数据技术能够实现对建筑工程造价的实时动态监控,及时发现造价偏差和异常,为管理者提供及时的预警和调整建议。三是风险预测与防范:基于大数据的预测模型可以对建筑工程造价中的潜在风险进行预测和评估,帮助管理者提前制定风险防范措施,降低风险损失。

### 2 基于大数据的建筑工程造价预测模型构建过程

#### 2.1 数据收集与整合

在构建基于大数据的建筑工程造价预测模型时,数据收集与整合是至关重要的第一步。首先,需要全面收集与建筑工程造价相关的各类数据。这包括历史项目数据,如过往项目的造价、工期、规模等,这些数据能够反映建筑工程在不同条件下的成本情况。同时,市场行情数据也是不可或缺的,包括材料价格、人工费用、机械设备租赁费用等,这些数据直接影响着工程造价的波动。此外,政策法规数据也是重要的参考因素,如税收政策、环保要求等,都会对工程造价产生直接或间接的影响<sup>[1]</sup>。在收集到原始数据后,还需要进行清洗、整合和

格式化处理。数据清洗的目的是去除重复、错误或无关的信息,以确保数据的准确性和一致性。数据整合则是将不同来源、不同格式的数据进行统一处理,形成一个完整、规范的数据集。格式化处理则是将数据按照特定的结构和格式进行排列,以便于后续的分析和挖掘。这一过程需要借助专业的数据处理工具和技术,如数据仓库、数据挖掘算法等,以实现高效、准确的数据处理。同时,还需要建立完善的数据管理制度和流程,确保数据的及时性、完整性和安全性。

#### 2.2 特征提取与选择

特征提取与选择是基于大数据的建筑工程造价预测模型构建中的关键环节。在这一步骤中,需要运用先进的数据挖掘技术,对收集到的多维数据进行深入地分析和挖掘。通过算法的运用,我们可以提取出与造价预测密切相关的特征,这些特征包括但不限于项目类型、地理位置、建筑面积、材料价格等。这些特征在很大程度上决定了建筑工程的造价水平,是预测模型中不可或缺的重要因素。项目类型是影响造价的重要因素之一,不同类型的项目往往具有不同的成本结构和造价特点。地理位置则直接关系到工程所在地的市场环境、交通运输等因素,进而对工程造价产生影响。建筑面积是衡量工程规模的重要指标,通常与造价成正比关系。而材料价格作为建筑工程成本的重要组成部分,其波动会直接导致造价的变化。在提取出关键特征后,还需要根据特征的重要性进行筛选和降维处理。这是因为并非所有特征都对造价预测具有同等重要的影响,一些冗余或相关性较弱的特征可能会增加模型的复杂性,降低预测的准确性。因此,需要通过特征选择算法,如主成分分析、决策树等,对特征进行筛选和降维,保留最具代表性的特征集合,以提高预测模型的性能和效率。

#### 2.3 预测模型构建与训练

在完成特征提取与选择后，接下来的关键步骤是预测模型的构建与训练。这一步是基于已提取的特征，构建一个能够有效预测建筑工程造价的模型，并通过历史数据对模型进行训练，使其能够掌握数据间的内在规律和联系。首先，需要根据问题的特性和数据的性质选择合适的预测模型。在建筑工程造价预测中，常用的模型包括回归模型、神经网络模型等。回归模型适用于分析变量间的数量关系，能够揭示自变量与因变量之间的依赖关系。而神经网络模型则具有强大的非线性拟合能力，能够处理复杂的非线性关系。在选择好模型后，需要利用历史数据对模型进行训练和学习。这一过程是通过将历史数据输入到模型中，不断调整模型的参数和结构，使模型能够逐渐拟合数据的变化规律。在训练过程中，还需要采用合适的优化算法和学习策略，以提高模型的训练效率和准确性。此外，为了防止模型出现过拟合或欠拟合等问题，还需要对模型进行验证和测试。通过将部分历史数据作为验证集和测试集，对模型的预测性能进行评估和比较，可以及时发现模型的不足并进行相应的调整和优化<sup>[2]</sup>。通过反复地训练和验证，我们可以得到一个能够较为准确地预测建筑工程造价的模型。这个模型不仅能够帮助我们理解数据间的内在联系和规律，还可以为实际的建筑工程造价预测提供有力的支持。同时，模型的构建和训练过程也需要不断地迭代和优化，以适应不断变化的数据环境和预测需求。

#### 2.4 模型验证与优化

在构建完预测模型并进行初步训练后，模型验证与优化环节变得至关重要。这一步骤的核心在于确保模型在实际应用中的准确性和可靠性，以及提升其泛化能力，即模型对未知数据的预测能力。为了全面评估模型的预测性能，我们采用多种方法，其中交叉验证是一种常用的技术。通过将数据集分成多个部分，并轮流使用其中的一部分作为测试集，其余部分作为训练集，我们可以有效地评估模型在不同数据子集上的表现。这种方法有助于减少因数据划分不当而导致的模型性能评估偏差。误差分析是另一种重要的评估手段。通过对模型预测结果与实际值之间的差异进行详细分析，我们可以了解模型在哪些方面的预测存在偏差，进而针对性地优化模型。常见的误差分析指标包括均方误差、平均绝对误差等，它们能够量化模型的预测准确性，为模型优化提供明确的方向。在模型验证过程中，如果发现模型的预测性能不佳或存在过拟合等问题，就需要对模型进行参数调整和优化。这可能涉及修改模型的网络结构、调整学习率等超参数，或者引入正则化等技巧来增强模型的

泛化能力。优化过程需要反复迭代，直到模型的预测性能够达到满意的水平。

#### 2.5 预测结果解释与应用

将训练好的模型应用于实际建筑工程造价预测中，我们首先需要确保输入数据的准确性和完整性。这是因为模型的预测结果直接依赖于输入数据的质量。通过输入项目类型、地理位置、建筑面积、材料价格等关键特征，模型能够输出一个预测的造价范围或具体数值。在得到预测结果后，我们需要对其进行解释和评估。这包括分析预测结果的合理性和可信度。合理性可以通过比较预测值与历史数据或行业标准来判断，而可信度则可以通过模型的预测性能指标来评估。如果预测结果既合理又可信，那么我们就可以放心地将其应用于实际的造价管理中<sup>[3]</sup>。根据预测结果，我们可以制定相应的造价管理策略和措施。例如，如果预测结果显示某个项目的造价可能会超出预算，那么我们可以提前调整项目计划或寻找成本更低的替代方案。这种基于预测结果的决策方式可以帮助我们更好地控制项目成本，降低造价风险。此外，预测模型还可以用于项目投标阶段的成本估算、项目执行过程中的成本控制以及项目结束后的成本审计等各个环节。通过持续应用和优化预测模型，我们可以实现建筑工程造价的精细化管理，提高项目的经济效益和社会效益。

### 3 面临的挑战与应对策略

#### 3.1 数据质量挑战

大数据环境下，由于数据来源的多样性和数据采集、存储、传输过程中的各种干扰因素，数据中往往夹杂着噪声、存在缺失值和异常值等问题。噪声数据可能导致模型学习到错误的模式，从而降低预测的准确性；缺失值则可能使模型在训练过程中丢失重要信息，影响模型的完整性；而异常值则可能导致模型对数据的分布产生误判，进而影响模型的稳定性。为了应对这些挑战，我们需要采取有效的数据清洗和预处理技术。首先，对于噪声数据，我们可以采用滤波、平滑等方法进行去噪处理，以减少其对模型的影响。其次，对于缺失值，我们可以根据数据的分布特征采用填充、插值等方法进行补充，以保证数据的完整性。最后，对于异常值，我们可以利用统计学方法或机器学习算法进行检测和处理，以确保模型能够正确反映数据的真实分布。此外，为了提高数据质量，我们还需要建立完善的数据质量管理体系和数据监控机制。通过定期的数据质量检查和评估，我们可以及时发现并解决数据中存在的问题，确保数据的准确性和可靠性。同时，通过数据监控机

制,我们可以实时掌握数据的动态变化,及时调整数据处理策略,以适应不断变化的数据环境。

### 3.2 模型复杂性与可解释性挑战

在构建建筑工程造价预测模型时,我们经常会面临一个关键的权衡:模型的复杂性与可解释性。随着模型复杂度的增加,例如引入更多的特征、使用更深的神经网络结构等,其预测性能往往会有所提高。这是因为更复杂的模型能够捕捉到数据之间更细微、更复杂的联系,从而更准确地预测未来的造价。然而,模型复杂度的增加往往伴随着可解释性的降低。当模型变得过于复杂时,我们可能很难直观地理解其内部的运作机制,以及各个特征是如何影响预测结果的。这不仅使得模型的应用变得困难,也增加了模型在未知数据上表现不稳定的风险。因此,在构建预测模型时,我们需要在保证预测性能的前提下,尽量简化模型结构,提高模型的可解释性和易用性。这可以通过选择合适的模型类型、控制模型的复杂度、使用可解释性强的特征等方法来实现。例如,我们可以选择线性回归、决策树等相对简单的模型,这些模型能够直接展示特征与目标变量之间的关系;或者我们可以使用特征选择技术来减少特征的数量,从而降低模型的复杂度<sup>[4]</sup>。此外,为了提高模型的可解释性,我们还可以采用一些专门的解释性工具或方法,如部分依赖图、排列特征重要性等。这些工具和方法能够帮助我们理解模型在不同特征值下的预测行为,以及各个特征对预测结果的重要性。

### 3.3 技术与人才短缺挑战

技术型人才的短缺主要体现在对既懂大数据技术,又深谙建筑工程造价管理知识的复合型人才的需求上。随着大数据技术的飞速发展,其在各行各业的应用越来越广泛,但在建筑工程造价管理领域,能够将大数据技术与实际业务相结合的专业人才却相对匮乏。这主要是因为大数据技术涉及复杂的算法、数据分析和处理技能,而建筑工程造价管理则需要深厚的行业知识和实践

经验。因此,要找到同时具备这两方面能力的人才并不容易。为了应对这一挑战,我们需要加强跨学科的人才培养和团队建设。首先,高校和企业应该合作,共同开设跨学科课程,培养既懂技术又懂业务的复合型人才。这些课程应该结合大数据技术和建筑工程造价管理的最新发展,注重理论与实践的结合,提高学生的综合素质和能力水平。其次,企业也应该加大对现有从业人员的培训力度,提高他们的大数据技术和建筑工程造价管理知识。通过定期的培训、研讨会和交流活动,从业人员可以及时了解最新的技术动态和业务知识,不断提升自己的专业能力。最后,我们还需要加强团队建设,培养一支既懂技术又懂业务的跨学科团队。这样的团队能够充分发挥大数据技术在建筑工程造价管理中的作用。

### 结语

本文系统研究了基于大数据的建筑工程造价预测模型的构建与应用过程,并深入探讨了面临的挑战和应对策略。通过大数据技术的深度挖掘和分析,可以揭示传统方法难以触及的造价规律和趋势,为建筑工程的决策提供更为精准和前瞻性的支持。未来,随着技术的不断发展和应用的深入拓展,基于大数据的建筑工程造价预测模型将在提高项目管理水平、降低项目风险等方面发挥越来越重要的作用。同时,也需要关注数据安全与隐私保护等新兴问题,推动大数据技术在建筑工程造价管理领域的健康可持续发展。

### 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,(05):6-11.
- [2]门金瑞.大数据背景下工程造价系统的设计[J].黑龙江科学,2023,14(17):133-134+137.
- [3]闫真真.大数据时代BIM技术在工程造价管理中的应用[J].中国招标,2023,(09):164-165.
- [4]陈辉阳.大数据应用下建筑工程造价预结算审核路径[J].中国建筑金属结构,2023,22(07):153-155.