

# 基于强化学习的施工调度智能优化模型及其工程适用性评估

叶紫桢 杨青

武昌理工学院, 湖北 武汉 430000

**摘要:** 针对传统施工调度方法在应对复杂工序、动态变化及资源冲突方面的不足, 本文提出一种基于强化学习的施工调度智能优化模型。通过构建以深度 Q 网络 (DQN) 为核心的学习框架, 将施工任务状态、资源配置与调度策略有机结合, 实现调度方案的自动生成与动态优化。在某城市综合体建设项目中开展实证研究, 结果表明该模型在缩短工期、提升资源利用率和应对突发事件等方面具有显著优势。研究验证了该方法在工程实际中的适用性, 并为施工调度智能化提供了新路径。  
**关键词:** 施工调度; 强化学习; 深度 Q 网络; 资源优化; 工程应用

在建筑工程项目管理中, 施工调度作为决定工程进度、资源分配和成本控制的关键环节, 一直是项目效率优化的核心问题。然而, 传统的施工调度方法多依赖静态规则和人工经验, 如关键路径法 (CPM)、线性规划等手段, 难以应对施工现场动态变化、资源冲突频发及不确定性事件等现实挑战。随着人工智能技术的发展, 强化学习 (Reinforcement Learning, RL) 作为一种智能决策模型, 因其具备从环境中学习最优策略的能力, 逐渐成为复杂调度问题研究的新方向。本文尝试将强化学习方法, 特别是深度 Q 网络 (DQN) 引入施工调度场景, 通过构建基于状态-动作-奖励机制的模型, 自动优化任务排序和资源分配路径, 提升调度的灵活性与效率。在此基础上, 结合实际工程案例对模型进行适用性评估, 验证其在复杂工程环境下的实用效果与应用潜力。通过理论分析与实证研究结合的方式, 为施工调度智能化、自动化探索一条新路径。

## 1 强化学习在施工调度中的理论基础与建模逻辑

### 1.1 强化学习基本原理及其调度优势

强化学习作为一种基于试错和环境反馈的智能决策方法, 在处理动态优化问题时展现出独特优势。其核心由“状态-动作-奖励-策略”的循环构成, 智能体通过在特定状态下选择动作、接受环境反馈, 并更新策略, 从而逐步趋近于最优行为路径。与传统调度算法相比, 强化学习能够适应任务环境中出现的非线性、非预期变化因素, 具有较强的适应性和鲁棒性。例如, 在施工过程中资源突发性短缺或任务延误时, 传统方法需要重新计算整体计划, 而强化学习模型可以基于当前状态快速调整策略, 实现动态

响应。

在施工调度应用中, 强化学习不仅可以避免对调度规则的人工设定, 还可以通过持续学习历史执行结果, 不断提升决策能力。深度强化学习 (如 DQN) 进一步结合了神经网络的函数逼近能力, 使模型可以处理高维状态空间, 适用于具有复杂约束、多任务、多资源类型的大型工程调度场景。因此, 强化学习方法为施工调度的智能化转型提供了全新思路。

### 1.2 施工调度问题的建模逻辑

要将强化学习应用于施工调度问题, 必须首先对问题进行合理建模。施工调度可以被抽象为一个序贯决策过程, 状态空间可由多个维度构成, 如当前任务完成情况、各资源使用状态、时间窗口等; 动作空间表示当前时刻系统可执行的调度选择, 如安排某项任务的执行或调整资源分配; 奖励函数则依据目标函数进行设计, 一般以最小化总工期、降低资源闲置率和避免冲突为目标, 构建多目标复合奖励机制。

在算法实现方面, 本文选用深度 Q 网络 (DQN) 作为主干强化学习算法。Q 网络通过将状态输入转化为每个可能动作的 Q 值输出, 引导智能体做出最优选择。在训练过程中, 模型通过与施工调度仿真环境交互, 积累经验数据, 并使用经验回放机制对历史数据进行采样学习, 提升训练效率和稳定性。此外, 为使模型更贴近实际应用, 还引入了资源冲突惩罚项、任务顺序约束处理模块和工期惩罚机制, 以增强模型对真实施工现场调度特征的适应能力。

值得一提的是, 在多资源调度环境中, 各类资源 (如

人工、机械、材料)具有不同的供给能力与使用限制,如何在强化学习框架中准确刻画资源状态并合理制定动作策略,是模型建构中的关键难点。本文通过构建任务-资源关联矩阵,对各资源使用情况动态监控,并在 Q 网络中嵌入资源可行性判断模块,以保证模型输出的调度方案满足现实约束。

## 2 施工调度强化学习模型的设计与实现路径

### 2.1 模型设计与参数设定

在将强化学习算法应用于施工调度的过程中,模型设计的关键是如何准确地映射工程实际问题与算法中的状态、动作和奖励三要素。本文所构建的施工调度智能优化模型采用深度 Q 网络(Deep Q-Network, DQN)作为算法核心,通过神经网络来逼近状态-动作值函数(Q 函数),以实现最优策略的动态获取。

在状态空间设计方面,模型考虑了当前所有任务的执行进度、可用资源情况、时间窗口状态和任务之间的逻辑依赖关系。每一个状态可表示为一个多维向量,其中包括任务完成度(0~1)、资源负载率、剩余工期、资源分配矩阵等信息。动作空间则定义为“选择某一任务在当前时间执行”的一组离散动作,每次决策中,模型从当前状态出发,为下一步任务执行做出最优选择。

奖励函数的构建直接影响模型学习的目标和效率。本文设计的奖励机制不仅考虑总工期的最小化,还引入资源冲突惩罚、任务延迟惩罚以及资源闲置惩罚等负向反馈。例如,当模型输出导致两个资源调度冲突时,立即给予负奖励;当模型合理安排高优先级任务并保证工序顺序合理时,给予正向奖励。通过这种复合型奖励函数的设计,引导模型在优化总体效率的同时兼顾调度合理性和资源利用率。

在参数设置方面,考虑模型收敛效率和实际计算成本,选择了如下基本参数配置:学习率设为 0.01,以确保模型在每次梯度更新时有稳定推进;折扣因子设为 0.95,用于权衡长期收益与短期利益;经验回放池大小为 10000 条经验,结合小批量随机采样提高训练稳定性;并采用 epsilon-greedy 策略,在初期以较高概率( $\epsilon=0.9$ )探索新策略,后期逐步降低 epsilon 至 0.1,强化已学习策略的利用。

### 2.2 实施平台与系统架构

为了实现强化学习模型的训练与调度仿真,本文基于 Python 语言构建了施工调度智能优化平台,结合 TensorFlow

框架搭建神经网络结构。系统整体分为四个功能模块:数据输入与预处理模块、环境模拟与交互模块、DQN 训练模块和调度输出模块。

数据输入模块接受用户提供的施工项目任务清单、任务依赖关系图、资源配置表等基础信息,转换为模型可识别的矩阵结构。预处理过程还包括任务优先级编码、资源配额限制嵌入等操作。环境模拟模块用于模拟实际施工场景中的任务推进与资源分配情况,实时记录当前状态变化并反馈给 DQN 模型。仿真逻辑采用基于事件驱动的调度器实现(Event-based Simulator),每当任务开始、完成或资源释放时触发状态更新事件,并重新进行调度决策。DQN 训练模块是核心计算引擎,负责根据当前状态选择动作、更新 Q 值、执行训练迭代等。为提升训练效率,引入固定目标网络机制、双 DQN 结构等技术,避免 Q 值高估问题。调度输出模块将模型产生的调度决策以甘特图、资源利用率图等方式可视化输出,供用户评估模型性能。

### 2.3 训练过程与模型调优策略

在模型训练过程中,采用模拟任务图共计 40 个任务,资源种类为 5 类,每类资源均设定固定数量上限。训练初期,模型随机探索,产生的调度方案波动较大,但随着训练轮数的增加, Q 值函数逐步收敛,调度结果趋于稳定。

为保证模型泛化能力,采用 5 组不同任务组合进行交叉训练,并用一组未参与训练的施工任务作为测试集评估性能。此外,经验回放机制在每轮训练后采样一个小批量的状态-动作-奖励-下一状态四元组,用于网络权重更新,提高了样本利用率,避免模型陷入局部最优。

在调优过程中,对比了不同网络结构的表现,例如采用三层全连接网络(输入层 128 节点、隐藏层 64 节点、输出层为动作数)与两层网络在训练速度与调度质量上的差异,发现增加中间层节点数能够略微提升决策质量,但也带来一定的过拟合风险。因此最终选择中等规模网络结构,并在训练后期加入 L2 正则项以抑制过拟合。

此外,还实验了不同奖励函数配置对模型行为的引导效果,如当奖励更偏重“资源均衡使用”时,模型倾向于生成资源利用率高但总工期略延长的调度方案;而当奖励更偏向“完工时间最短”时,调度序列压缩性更强,但资源冲突频率增加。通过不断调优权重因子,最终确定一组权衡效率与稳定性的奖励参数组合。

### 3 工程案例应用与适用性评估

#### 3.1 案例项目背景与调度需求描述

为验证强化学习施工调度模型的实际适用性,本文选取“某城市综合体建设项目”为应用案例。项目任务共 45 项,涵盖结构施工、设备安装、内外装修等阶段,资源类型包括工人、塔吊、混凝土泵车等共 7 类,每类资源受限,调度周期为 120 天。项目原采用关键路径法(CPM)进行调度,面对施工任务密集、资源调配冲突频繁的复杂情境,难以及时响应计划变更,影响整体效率。

在引入强化学习模型前,将工程数据转化为状态向量输入,包括任务状态、资源使用率、时间约束等信息,构建每日为单位的调度周期。该项目具有代表性,工序依赖紧密,资源约束显著,是测试调度优化模型有效性的理想对象。

#### 3.2 模型应用流程与对比方案设定

强化学习模型按如下流程运行:首先输入当前工程状态,模型决策调度动作(即安排哪项任务执行),环境反馈更新任务与资源状态,模型据此学习策略,直至项目结束。整个训练与仿真过程在自建调度仿真平台中完成。

为评估模型表现,设置三组对比:强化学习组、传统 CPM 组、遗传算法(GA)组。在同一项目数据下分别执行 10 轮仿真实验,记录工期、资源冲突次数、资源利用率等关键指标,进行横向对比。

#### 3.3 评估结果与分析

实验数据显示,强化学习模型在效率与稳定性上表现最优。CPM 组平均工期 118.6 天,冲突 38 次,资源利用率 71%;GA 组工期 114.2 天,冲突 27 次,资源利用率 77%;强化学习组工期降至 110.4 天,冲突次数仅为 12 次,资源利用率提升至 84%。模型在避免任务堆积、提升资源协调方面明显优于传统方法。

在动态响应方面,模拟设置两种突发事件:关键资源提前调离及恶劣天气延误。CPM 和 GA 调度系统响应滞后、需要重算方案,而强化学习模型可在下轮调度中自适应调整路径,保持工期基本稳定,展现出良好的适应能力。

甘特图与资源利用图分析显示,强化学习模型生成的调度序列衔接紧密,资源空转时间减少,任务间等待时间缩短,整体调度质量显著提升。

尽管模型展现出较好性能,但也存在一定问题:模型训练时间较长,初期策略不稳定;对数据完整性要求高,

输入数据若不准确将影响最终调度结果。此外,项目规模扩大时模型复杂度和计算负担将迅速上升,需配合计算优化策略。

### 4 结论与展望

本文基于强化学习构建了一个施工调度智能优化模型,通过引入深度 Q 网络(DQN)算法,将施工任务状态、资源分配与工期控制等要素纳入统一的决策框架中,实现了对复杂调度问题的动态优化。模型在模拟与实际项目环境中的应用验证表明,相较于传统关键路径法与遗传算法,强化学习模型在调度效率、资源利用率以及应对突发变化的能力方面均表现更优,具备较强的工程适用性和推广价值。

通过对某城市综合体建设项目的实证分析,模型有效缩短了工期,减少了资源冲突次数,并提升了施工调度的智能化水平,证明强化学习方法能够在动态、约束复杂的工程环境中实现自主学习与优化决策。然而,该方法仍存在训练耗时较长、对数据依赖性强的问题,未来研究可在模型压缩、迁移学习、实时数据集成等方面进一步优化。

展望未来,强化学习在施工调度中的应用可进一步与 BIM 技术、物联网感知系统融合,实现基于实时数据驱动的智能调度平台;同时引入多智能体协同机制,有望提升跨项目、跨工序间的整体协调效率,推动施工管理模式向自动化和智能化迈进。

#### 参考文献:

- [1] 乔震霖.基于多智能体强化学习的道路施工自动化监控系统[J].电子设计工程,2024,32(23):85-88+93.
- [2] 吕娜,张然.智能制造背景下基于数字孪生的装配式建筑施工管理动态调度探究[J].绿色建造与智能建筑,2025,(01):171-174.
- [3] 张建超.智能化施工调度在建筑工程管理中的应用研究[J].房地产世界,2024,(20):101-103.
- [4] 谢宜宁.工程机械群智能配置调度优化及一体化施工系统研究[D].山东大学,2022.
- [5] 连静.装配式施工项目调度多目标优化研究[D].西安建筑科技大学,2020.
- [6] 于大成.隧道施工智能调度管理系统研究[J].运输经理世界,2021,(26):76-78.
- [7] 孙景博,王法.BIM技术在装配式建筑施工调度中的应用研究[J].新城建科技,2025,34(02):89-91.