基于强化学习的智能调度系统设计与实现

张瀚驰

成都工业学院，四川省成都市，611730；

摘要：随着生产系统日益复杂化、任务调度需求动态变化，传统启发式与规则驱动的调度方法在多目标、多约束环境下表现出明显局限性。强化学习作为一种自适应、试错驱动的智能算法，为解决调度系统中的不确定性与实时优化问题提供了新的路径。本文围绕智能制造与大规模资源配置场景，构建基于强化学习的调度系统架构，通过状态空间建模、奖励函数设计与策略学习过程优化，实现对任务－资源匹配的动态决策与在线优化。系统采用模块化设计，融合神经网络与深度强化学习算法，在多个仿真与实际调度场景中进行验证，展现出较强的收敛性能与应用适应性。研究结果表明，该系统在提升资源利用效率、降低平均任务完成时间方面具有明显优势，为调度问题的智能化解决提供了可行思路与实现范式。

关键词：强化学习；调度优化；状态空间；奖励函数；系统架构

DOI：10.69979/3060-8767.25.06.083

引言

在现代制造、物流、能源与数据中心等领域，调度问题广泛存在于作业排程、资源分配、任务调配与路径选择等核心环节，其解法直接影响系统运行效率与经济效益。传统调度方法依赖人工经验、启发式规则或数理模型，其适应性与扩展性难以满足现实业务对高动态性与多约束处理的需求。近年来，强化学习因具备在线学习、自主探索与策略优化等特点，被逐渐应用于复杂环境下的调度优化任务。相较于静态建模方法，强化学习通过与环境交互学习最优行为策略，可在状态转移不确定、任务变动频繁的场景中实现动态响应。本文基于该思想，构建一套智能调度系统，围绕状态定义、动作策略、奖励函数及其在系统中的集成路径展开深入研究，并通过仿真实验与实际部署验证算法效果与系统实用性，以期为高复杂性调度问题提供一种数据驱动、自动优化的解决方案。

1 智能调度问题分析与建模基础

1.1 调度问题的复杂性与现实挑战

调度问题广泛存在于工业制造、交通运输、能源调配与云计算资源管理等实际系统中，其核心目标是通过任务与资源的匹配，使整个系统运行效率最优化。然而，真实场景下的调度问题通常表现出高度复杂性与动态变化性。系统中涉及的任务种类繁多、属性不一，资源则可能具备多种异构特征，例如加工能力差异、空闲周期不一或切换时间成本等，这使得调度优化的目标不再单一，常常需要在多个指标之间做出权衡，如缩短任务总时长、均衡资源负载、降低系统能耗等[1]。

部分任务的到达时间、加工需求甚至可能在调度决策前未知，部分资源的状态会受到外部影响而突变，例如突发故障、供应链波动或执行延迟。调度系统需应对的约束种类多样，既有硬性约束（任务顺序、时间窗、资源容量等），也有软性偏好（优先级倾向、能耗阈值、人工协作等），这导致传统启发式算法在面对大规模、高动态场景时往往收敛缓慢，难以适配或迁移。尤其在多目标优化任务中，传统调度器对冲突目标之间的协同调节缺乏智能支撑，严重制约了算法效果与系统稳定性。调度优化亟需突破单一规则主导的局限，引入自适应、可演化的算法机制，以应对复杂任务结构与环境扰动下的实际需求。

1.2 强化学习在调度问题中的理论基础

强化学习是一种以行为决策为核心，通过环境交互获取反馈，不断优化策略的学习方式。在调度系统中，强化学习可将“状态”表示为任务队列、资源占用、当前进度等系统运行信息，“动作”对应具体的调度决策（如任务－资源匹配关系），“奖励”则由调度效果计算得出，如任务提前完成奖励、设备空转惩罚等[2]。通过智能体与环境的持续交互，强化学习能不断修正策略，引导系统逐步趋向整体效率最优。

为了适应调度场景中高维状态空间与动态变化需求，深度强化学习结合神经网络进行策略函数与价值函数的逼近，大幅提升了模型的学习与泛化能力。DQN 等算法适合处理离散动作空间，可应用于资源选择与任务排序问题；而 Policy Gradient 类方法则适用于连续决

策优化，如作业起始时间与执行窗口调整。现代调度系统往往需求弹性大、反馈滞后且环境不确定性强，强化学习在这种设定下具备天然优势。它不仅能根据历史经验修正当前策略，还能在新任务模式或突发事件下自适应调整，提高整个系统的稳健性与抗干扰能力。

强化学习相较于传统方法更具可持续进化与迁移学习能力，已在多种典型调度任务中实现突破，包括云计算任务调度、无人仓储路径分配、柔性制造产线调度等。实践中，已训练好的模型还可在其他类似场景中快速迁移部署，只需少量调整即可获得较优性能。正因如此，强化学习被认为是构建下一代智能调度系统的核心算法支柱。后续章节将以此为基础，进一步探讨其在系统架构、模型设计与实际部署中的落地路径与技术实现。

2 强化学习算法设计与系统架构实现

2.1 状态空间与奖励函数设计方法

调度任务的强化学习建模过程需从调度系统自身的结构逻辑出发，构建能够准确感知系统状态与反馈调度效果的状态－动作－奖励三元组体系。在状态空间的建构方面，本系统整合了任务层级状态（任务队列长度、剩余执行时间、任务紧迫度）、资源运行状态（各设备占用情况、执行能力、冷却时间等）以及系统全局进度状态（任务完成比例、时间段内负载率波动等），以张量形式统一输入策略网络，从而增强模型对复杂时序结构与资源冲突关系的识别能力[3]。为降低状态空间维度、提高泛化能力，系统采用了基于主成分分析的状态压缩策略与离散化处理，使得在高维任务资源环境中仍能保持训练效率。

在奖励函数的构建方面，设计目标是促使调度系统自动优化调度结果，形成面向多指标目标的反馈引导机制。主奖励函数根据任务实际完成时间、平均等待时长与资源利用率等关键指标构建加权奖励模型，优先引导模型优化核心性能目标。同时，引入辅助奖励函数处理调度中出现的频繁切换、长时间空转等行为，通过负向反馈抑制低效调度行为的发生。此外，在任务密集但回报稀疏的早期训练阶段，系统设计了辅助信号机制（如完成度奖励、冲突次数奖励）以提高反馈密度，加速学习过程。整体来看，该奖励体系实现了对调度策略在局部执行效率、全局时间成本与资源综合负载三者之间的系统性调控，为策略模型在实际场景中稳定运行提供了基础保障。

2.2 系统平台的总体架构与实现机制

为实现强化学习在实际调度业务中的落地部署，本研究设计并实现了一个端到端的调度决策平台，架构整体分为三层：数据接入与预处理层、强化学习决策层、调度执行与反馈层[4]。数据接入层通过采集外部调度环境中的实时任务信息与资源配置数据，建立起基于标准协议的数据通道，将原始信息统一清洗与编码后输入至策略学习模块。同时，系统内置异常检测与数据缓冲机制，用于处理任务输入波动、设备状态突变等异常情况，保障模型输入质量。

决策层基于 Actor－Critic 结构构建策略模型，并配套目标网络用于延迟更新，提升策略更新的稳定性。模型训练时采用经验回放机制，引入多步预测与优势函数估计方法，进一步提高策略学习效率。部署阶段支持在线推理模式，模型根据实时状态输出动作指令，通过调度接口下发至底层系统。执行层通过调用调度管理模块对任务进行优先级排序与资源匹配执行，并将调度效果实时反馈至训练模块完成闭环。整个平台基于容器化部署架构，可与 ERP、MES 等系统无缝对接，同时提供图形界面展示调度轨迹与系统负载，实现从任务生成到调度执行、策略更新的全流程智能闭环。

3 算法性能验证与实际应用分析

3.1 强化学习调度算法的对比实验结果

为验证本系统强化学习调度算法的有效性，构建了包含 1000 条任务与 20 个资源节点的典型动态调度仿真平台，并引入三类对比模型：基准规则法（FCFS）、传统启发式算法（遗传算法、模拟退火）以及主流深度强化学习算法（DQN、DDPG）[5]。测试过程中设置多个动态干扰变量，包括任务突发、资源随机故障与负载飘移等，力求模拟真实生产调度系统的复杂环境。实验以平均任务完成时间、系统总响应时间、资源利用率三项指标为评价标准，全面评估各类算法的调度表现。所有算法均在相同初始条件与约束设定下进行训练与评估，确保对比数据的客观性与可复现性。

实验结果表明，本研究构建的基于 Actor－Critic 架构的策略模型在三项指标上均优于对比算法。在平均任务完成时间方面，相比 DQN 与启发式算法，分别降低约 17%与 22%；系统总响应时间也呈显著下降趋势，说明模型对系统拥堵时的应变能力较强。资源利用率方面，Actor－Critic 策略能够在高负载条件下有效分散任务执行压力，使资源保持高效调度状态。进一步分析显示，该算法策略输出稳定性较高，对任务结构变化的敏感度较低，具备较强的泛化能力。在多轮迭代训练中，其收敛速度也显著快于其他深度强化学习模型，尤其在复杂

状态组合下，依赖状态压缩与优势估计机制，策略性能表现更为鲁棒。特别是在系统初始阶段或任务分布不均时，该策略表现出更强的恢复调节能力，能够快速重构有效的调度路径，避免资源闲置与局部拥塞并存的问题。

通过增加任务规模至 3000 条，并扩大资源节点数量至 50 个，策略训练时间略有上升，但整体调度性能依旧优于传统方法，说明该算法具备良好的大规模任务适配能力。同时，在测试阶段引入部分噪声扰动与状态信息丢失模拟后，强化学习模型仍保持较高策略一致性，说明其在面对不确定性与信息缺失条件下仍具有较强的适应性与容错能力，进一步证明其在真实复杂调度环境中的应用潜力。此外，通过可视化跟踪调度过程，验证了策略输出的可解释性，管理者能够明确理解策略如何根据任务特征进行动态调整，这一能力对系统部署与后续维护具有重要意义。模型所具备的行为透明性与自我优化能力，也为调度策略从“黑盒”到“灰盒”的落地提供了现实可行的技术路径。

3.2 在智能制造与边缘算力场景中的部署成效

在工业落地验证方面，平台已在某电子制造企业 MES 系统中集成试点应用，部署周期约一周，采用灰度并行调度模式与原调度系统对比运行。现场测试数据显示，强化学习调度策略使得生产节拍提升 9.3%，设备平均空闲率下降 15%，加急任务平均响应时间缩短 23%，有效缓解了多品种小批量场景中计划频繁变更带来的调度压力。系统运维监控也显示，强化学习模块在长期运行中保持策略稳定性，未出现显著性能回退，证明其具备持续运行与自适应学习能力。该企业在推广阶段反馈良好，表示系统在复杂工单交错、高并发调度需求下依旧保持了调度精度，且支持任务优先级动态调整，满足了个性化订单频繁插单的现实需要。同时，系统具备异常状态的自动识别能力，可在设备运行波动或网络延迟出现时，实时调节调度策略，防止任务拥堵与节点空转并发。值得注意的是，策略训练过程中引入了实际设备运行数据作为初始输入，有效缩短策略冷启动周期，增强了调度策略对现场环境的适配能力与实用性。

在云边协同计算资源调度场景中，平台也完成了小规模部署试验。在 10 台边缘节点与 200 个并发任务分配测试中，系统通过调度策略自适应调整算力分配与能耗比，提升整体节点 CPU 利用率约 13%，并成功规避多次任务资源冲突。平台结合任务队列特征自动构建状态向量与动态策略映射，使调度效率得以在负载波动场景下保持稳定。相比于静态规则分配方式，强化学习策略更能快速适应资源紧张与分布不均问题，尤其在边缘计算环境中算力瓶颈频繁变化的情况下，该算法展现出显著的调控与弹性优化能力。系统还提供任务执行回放功能，可复盘调度决策逻辑并辅助系统运维人员进行故障分析与策略修正。可视化控制台同步集成任务热力图与资源占用趋势预测，提升了平台在边缘智能场景下的可视管控水平。后续部署计划还将接入多边缘异构节点集群，探索跨节点协同调度下的策略适应能力与决策时延评估，为规模化扩展奠定基础。此外，平台计划引入多目标强化学习机制，以实现对能耗、响应时间与任务成功率的动态权衡优化，进一步提升其在能源敏感型行业的推广价值。

4 结语

强化学习技术在智能调度系统中的应用，体现了人工智能方法在复杂任务分配问题中的可行性与优越性。通过构建面向实际场景的任务调度模型，并结合 Actor-Critic 结构对策略进行训练与优化，系统在调度效率、资源利用率与适应能力方面表现出明显提升。实验数据与部署反馈均证实，该调度模型不仅具备良好的稳定性与泛化能力，还能够应对动态变化与边缘算力约束，适用于制造、物流、边缘计算等多种应用环境。同时，系统的可视化模块与策略解释能力为实际运维管理提供了支持，增强了调度决策的透明性。未来研究可进一步引入多目标优化机制、异构节点策略协同与任务预判机制，持续提升调度系统的智能水平与行业适应力，为智能制造与边缘计算场景中的资源配置提供更具前瞻性与实效性的解决路径。

参考文献

[1]周华丽,邵金峰.强化学习算法在图书馆智能排架与借阅调度中的应用[J].电脑知识与技术,2025,21(13):61-63.

[2]曹小雄,廖伟文,李卓君.基于深度强化学习模型的纯电动牵引车智能充电调度系统[J].港口装卸,2025,(01):24-28.

[3]杨嘉.基于强化学习的边缘计算智能资源优化调度研究[J].信息与电脑,2025,37(01):1-3.

[4]张君,林琳,郭芮,等.基于改进深度强化学习的电网电力智能调度分析模型研究[J].自动化技术与应用,2025,44(07):139-142+177.

[5]陈宁,李法社,王霜,等.基于深度强化学习算法的分布式光伏-EV 互补系统智能调度[J].高电压技术,2025,51(03):1454-1463.