

# 强化学习与深度学习结合的多智能体系统研究

姚昊林

合肥大学 安徽 合肥 230000

**摘要:** 本论文围绕强化学习与深度学习融合的多智能体系统展开深入研究。针对动态环境下算法稳定性欠佳、样本利用效率低下以及算法可扩展性不足等关键难题,提出系统性优化策略。从动态环境适应性、样本高效利用、算法灵活扩展三个维度,对多智能体深度强化学习算法进行针对性改进,并创新性地结合自监督学习与迁移学习设计全新算法。深入探究多智能体协作竞争机制、通信交互技术及环境感知建模等关键技术,完善系统运行框架。同时,从理论层面严格证明算法的收敛性,细致分析算法复杂度,全面验证算法稳定性与鲁棒性,构建科学的性能评价体系。研究成果显著提升了多智能体系统算法性能,为其在复杂多变场景中的广泛应用提供了坚实的理论基础与技术保障,对推动人工智能领域发展具有重要意义。

**关键词:** 强化学习; 深度学习; 多智能体系统; 算法优化; 算法设计

## 引言

随着人工智能技术发展,多智能体系统在多领域应用广泛,但复杂环境下的算法难题制约其发展。强化学习与深度学习融合虽为多智能体系统带来新方向,却面临动态环境适应、样本利用及算法扩展等挑战<sup>[1]</sup>。本研究旨在突破这些瓶颈,探索算法优化与设计方法,提升系统性能,对推动多智能体系统理论发展与实际应用具有重要意义。

## 1 强化学习与深度学习结合的多智能体系统算法优化策略

### 1.1 动态环境下的算法稳定性增强

多智能体系统运行时,强化学习与深度学习融合的算法常因环境动态变化而面临稳定性难题。环境状态的频繁更迭与不可预测性,极易干扰算法的收敛进程,引发训练结果的剧烈波动甚至崩溃。要提升算法在动态场景中的稳定性,需从模型构建、训练机制及策略优化三个层面进行综合改进。构建深度学习模型时,采用能动态捕捉环境变化特征的架构,可增强模型对复杂动态环境的适应性;在强化学习训练环节,借助正则化手段和经验回放策略,能有效降低训练数据的关联性,防止模型陷入过拟合困境,进而提升算法的抗干扰能力。而设计具备自适应能力的学习率调整方案,配合灵活的探索-利用平衡策略,可使算法根据环境变化实时调整学习步调,确保算法在动态环境中始终保持稳定运行状态。

### 1.2 多智能体系统中的样本效率提升

多智能体系统运行过程中,强化学习与深度学习结合的算法在样本利用方面存在显著挑战。单个智能体收集的样本往往带有局限性,多个智能体积累的海量数据

又容易产生重复信息,这导致算法训练需耗费大量时间与计算资源<sup>[2]</sup>。要提升样本效率,可通过设计高效的数据共享机制打破智能体间信息壁垒,让各智能体采集的样本得以互补复用,避免重复探索。构建跨智能体的知识迁移框架也是有效途径,将已训练智能体的经验与策略,合理迁移到其他智能体的学习过程中,减少从零开始学习的样本消耗。此外,运用先进的数据筛选与处理技术,对原始样本进行精细化分析和筛选,保留最具价值的信息用于模型训练,能大幅提升样本利用效率,推动算法在有限样本条件下实现高效学习与优化。

### 1.3 算法可扩展性优化

随着多智能体系统规模不断扩大、应用场景日益复杂,强化学习与深度学习结合的算法面临可扩展性不足的瓶颈。当智能体数量增多、环境维度增加时,原有算法架构容易出现计算资源过载、训练效率下降等问题。优化算法可扩展性,需从架构设计、功能模块解耦和分布式部署等方面进行突破。采用分层模块化的算法架构,将复杂任务拆解为多个相对独立的子任务模块,每个模块可根据实际需求进行独立优化与扩展。在功能模块设计上,确保各模块之间低耦合,便于新增智能体或扩展任务时快速集成与调整。同时,引入分布式计算框架,将算法训练和决策过程分散到多个计算节点,利用并行计算能力提升处理大规模数据和复杂任务的效率,从而实现算法在不同规模多智能体系统中的灵活适配与稳定运行。

## 2 多智能体系统中的关键技术研究

### 2.1 多智能体协作与竞争机制

在多智能体系统运行过程中,智能体间的协作与竞争关系深刻影响着系统整体效能。多个智能体为达成共

同目标时,需要建立有效的协作机制,通过信息共享与行动协同,发挥各自优势实现资源整合,提升任务完成效率。然而,在有限资源分配或目标存在冲突时,智能体间又不可避免地产生竞争行为。若竞争过度,可能导致智能体陷入局部最优解,阻碍系统整体目标达成;协作不足,则难以形成合力应对复杂任务。优化多智能体协作与竞争机制,需构建动态平衡体系,依据任务需求和环境变化,灵活调整智能体间的关系。通过设计合理的奖励机制引导智能体在竞争中寻求协作机会,同时利用约束规则避免恶性竞争,从而在复杂场景下保障多智能体系统稳定高效运行<sup>[3]</sup>。

## 2.2 通信与信息交互技术

多智能体系统的高效运行依赖智能体间稳定的通信链路与精准的信息交互。现实应用场景中,通信环境的复杂性给信息传递带来诸多挑战,无线信号干扰导致数据丢包、有限的网络带宽限制传输速率,这些问题直接削弱智能体间的协作效能。优化通信与信息交互技术需从多方面着手,首先要设计抗干扰能力强的专用通信协议,通过编码优化和差错控制机制,保障复杂环境下信息传输的完整性与可靠性。针对不同任务特性构建定制化信息交互策略也至关重要,明确各智能体信息传输优先级与交互时序,既能避免关键信息延迟又能减少冗余数据传输。面对多源异构的海量交互数据,采用轻量化数据压缩算法与实时处理技术,降低传输负载的同时提升信息解析效率。只有系统性地改进通信与信息交互技术,才能构建稳定高效的信息流通网络,为多智能体系统协同作业筑牢基础。

## 2.3 环境感知与建模技术

多智能体系统运行时,环境的动态性和复杂性给智能体决策带来巨大挑战,准确的环境感知与高效的建模技术成为系统运行的关键。实际场景中,环境信息往往呈现多源性、不确定性特点,传统感知方式难以全面捕捉环境变化。为解决这一问题,需采用多模态感知技术融合视觉、听觉、传感器等多源信息,拓宽智能体对环境的认知维度。在建模方面,构建动态化环境模型尤为重要,利用深度学习强大的特征提取能力,将环境状态转化为可计算的表征形式,使智能体能够快速理解环境变化规律。同时,结合强化学习的反馈机制,持续优化环境模型参数,确保模型随环境动态演变而更新。通过先进的环境感知与建模技术,智能体得以更精准地把握环境态势,为制定合理决策、实现系统目标奠定坚实基础。

## 3 基于强化学习与深度学习的多智能体系统算法设计

### 3.1 多智能体深度强化学习算法改进

多智能体深度强化学习算法应用于实际场景时,智能体交互关系的复杂性与环境的动态特性,常使算法陷入局部最优困境,训练效率也因此受到严重制约。突破这一现状需对算法架构与训练机制进行系统性革新<sup>[4]</sup>。在架构层面,引入分层注意力机制能够帮助智能体在海量环境信息中精准筛选关键数据,增强对全局状态的感知与理解能力,从而有效应对复杂环境带来的挑战。针对大规模系统中传统集中式训练引发的通信拥堵问题,采用分布式训练架构将计算任务合理分散至多个节点,大幅缩短训练周期。在训练机制优化上,设计动态奖励分配策略可依据智能体在任务执行过程中的实际贡献调整奖励额度,这种差异化激励方式能充分调动智能体协作积极性,避免因奖励分配不均导致的学习方向偏差,最终实现多智能体深度强化学习算法性能的全面提升。

### 3.2 结合自监督学习与迁移学习的算法设计

多智能体系统在训练过程中常遭遇数据匮乏、任务场景差异大等难题,限制算法泛化能力与学习效率。将自监督学习与迁移学习融入算法设计,可有效攻克这些瓶颈。自监督学习能挖掘数据自身蕴含的监督信号,让智能体从无标注数据中自主学习特征表示,减少对大量人工标注样本的依赖,尤其适用于多智能体系统中复杂环境下的海量未标注数据处理。迁移学习则能将源任务中学习到的知识、经验迁移到目标任务,使智能体在新任务场景下无需从头训练,大幅缩短学习周期。在实际应用中,先通过自监督学习构建通用特征提取器,再利用迁移学习技术将其适配到不同任务,既能提升算法对多场景的适应能力,又能降低训练成本,实现多智能体系统在复杂环境下的高效学习与决策。

### 3.3 算法性能分析与比较

评判强化学习与深度学习融合的多智能体系统算法,性能分析和比较是不可或缺的重要环节。构建全面的算法性能评价体系,需要从多个关键维度展开。任务完成效率作为核心考量因素,通过统计智能体在限定时间内达成目标的成功率,能够直观展现算法决策的速度与精准度。而算法收敛速度同样影响着学习效率,通过监测训练过程中累计奖励值的增长趋势或损失函数的下降曲线,可判断算法能否快速且稳定地获取最优策略。面对复杂多变的实际应用场景,算法的泛化能力决定其适用范围,将训练后的算法置于不同环境条件下运行,对比性能变化情况,便能检验其对多样化场景的适应能力。在进行算法比较时,选择业内经典算法作为参照基准,在统一的实验环境中运行各算法,借助量化指标对比与可视化图表展示,能够清晰揭示不同算法在各项性

能指标上的优劣差异,为算法的改进方向和实际应用选型提供可靠的数据支撑。

#### 4 算法优化与设计的理论分析

##### 4.1 算法收敛性证明

评估强化学习与深度学习融合的多智能体系统算法质量,收敛性证明是关键理论环节。若算法训练过程无法收敛,智能体难以形成稳定决策策略,导致系统目标无法达成。开展算法收敛性证明工作,需紧密围绕算法结构与运行机制展开深入研究。先精准界定算法目标函数,透彻剖析参数更新规则,运用数学推导手段,跟踪每次迭代过程中目标函数数值演变轨迹。多智能体系统内,各智能体行为相互影响,群体交互产生的复杂动态特性,显著作用于算法收敛进程,此影响因素在证明过程中必须充分考量。可引入李雅普诺夫函数等数学分析工具,搭建稳定性理论模型,从数学逻辑层面严格论证算法在特定条件下,能够逐步逼近最优解或近似最优解。考虑到实际应用中环境的动态变化,还需进一步验证算法在环境扰动下的收敛保持能力,以此确保算法在复杂场景中具备可靠的应用价值<sup>[5]</sup>。

##### 4.2 算法复杂度分析

研究强化学习与深度学习融合的多智能体系统算法,复杂度分析是评估其实际应用可行性的重要理论依据。算法复杂度直接关系到运行效率与资源消耗,若复杂度超出硬件承载能力,即便算法理论上有效,也难以落地实施。对这类算法进行复杂度分析,需从时间和空间两个维度展开深入研究。时间复杂度层面,要细致分析算法完成一次训练迭代或决策所需的计算步骤数量,考量智能体数量增加、环境状态维度扩展时计算量的增长趋势。空间复杂度方面,则需评估算法运行过程中存储模型参数、中间计算结果及经验数据所需的内存资源,尤其在多智能体产生海量交互数据的情况下,合理控制空间占用至关重要。通过全面的复杂度分析,能够清晰把握算法在不同规模场景下的资源需求,为算法优化和实际部署提供关键参考。

##### 4.3 算法稳定性与鲁棒性理论验证

在强化学习与深度学习融合的多智能体系统中,算

法稳定性与鲁棒性是决定其实际应用价值的重要属性。算法稳定性确保智能体在常规运行状态下维持可靠决策,而鲁棒性则体现算法在面对噪声干扰、环境突变或系统参数波动时的抗干扰能力。对算法稳定性进行理论验证,需构建稳定性分析模型,通过分析算法参数更新过程中状态变量的变化趋势,判断算法是否能收敛到稳定状态且保持长时间稳定运行。针对鲁棒性验证,需模拟多种干扰场景,如人为添加数据噪声、改变环境关键参数、设置通信故障等,观察算法性能指标的波动情况。利用数学推导和仿真实验相结合的方式,从理论层面证明算法在不同干扰强度下仍能保持可接受的性能水平,确保算法在复杂多变的实际应用场景中具备稳定、可靠的运行能力。

#### 结论

本研究在多智能体系统算法优化与设计方面取得进展,提出的策略与算法改进方案经理论分析验证有效。然而,实际应用中仍面临复杂环境适配、多场景通用性等问题。未来研究可进一步探索跨领域知识迁移,优化算法在极端场景下的性能,推动强化学习与深度学习结合的多智能体系统技术创新与应用拓展。

#### 参考文献

- [1]张浩洋,彭世国,张添乐.编码-解码方案下具有网络不确定的离散多智能体系统的均方一致性[J].陕西科技大学学报,2025,43(03):202-209.
- [2]张保俊,马旭峰,邢阳阳.基于有向图和DoS攻击的多智能体系统平均一致性研究[J].网络安全技术与应用,2025,(05):33-35.
- [3]吴燕,班曼露,邓红勇.具有动态量化器和数据丢包的随机多智能体系统的事件触发控制[J].遵义师范学院学报,2025,27(02):94-97.
- [4]耿燕,常杜辉,贺兴时.多智能体系统的一致性数据驱动最优迭代学习控制[J].西安工程大学学报,2025,39(02):118-126.
- [5]甘凤金,陈展衡.事件触发机制下时滞二阶多智能体系统一致性研究[J].科技创新与应用,2025,15(11):23-27.