

摘要

本文深入探究了强化学习领域中的一个重要分支——多智能体合作问题，并专注于其面临的新挑战：可泛化的多智能体合作问题。这一问题着重研究智能体如何有效地与未见过的团队或智能体进行合作，这是多智能体系统研究中的一个关键问题。目前，在多智能体强化学习合作问题中，虽然中心化训练与分布式执行的框架占据主导地位，但它们主要面临的挑战是智能体策略在遇到未知策略或不同团队构成的环境时的协同泛化能力较弱。传统的智能体训练方法通常使得智能体在特定环境中表现良好，但这些智能体在新的、未知的环境中难以适应。此外，传统的多智能体系统侧重于在特定的环境下优化策略，以最大化在该环境中获得的奖励。这种方法的一个主要缺点是缺乏对变化环境的协同，智能体的行为策略通常是针对特定环境的最佳响应，而非面对未知情况的泛化策略。当智能体从一个训练环境转移到一个新的环境，尤其是当团队成员或任务要求发生变化时，它们的表现往往会显著下降。这是因为在原始训练过程中，智能体的学习和适应是基于特定团队和环境条件进行的，没有学习如何在不同情境下通过有效的角色适应来优化其策略。

在复杂的现实世界应用中，如自动驾驶车辆、机器人协作等场景，智能体经常需要与多种不同的智能体或人类互动。这些应用要求智能体能够理解并预测其他参与者的行为，以做出最优决策。然而，目前的多智能体强化学习算法通常只在一个固定的合作伙伴集内进行优化，缺乏适应动态变化的多智能体环境的灵活性。因此，开发能够泛化到未见过的智能体或团队的合作策略是多智能体系统研究的前沿方向。

为了应对这一挑战，本文提出了一系列基于角色多样性的强化学习算法。这些算法通过引入角色分工的概念，不仅使智能体在训练过程中学习个体任务执行的技能，还能学习如何在不同的团队配置中有效合作。在多智能体合作中引入角色分工的概念可以显著提高智能体的泛化能力。角色分工意味着每个智能体都有明确的职责和行为模式，根据其在团队中的角色定义。例如，足球队中的守门员和前锋承担不同的职责和技能需求。在多智能体学习中应用类似的原理，可以使每个智能体在面对未知环境或新队友时，快速地适应。这些方法的核心思想是通过角色的多样性分配促进智能体间的有效协同，从而在面对未曾训练的团队构成的情况下也能有效协同。

本文具体提出了两种创新的算法：第一种是最大化带约束的角色分布熵方法，该方法运用信息论和因果推断工具，通过量化角色多样性并将其转化为可优化的数值问题，以信息熵为基础构建了一个优化框架。该框架通过最大化带约束的熵原则优化角色分配，其中约束是以因果关系定义的，从而确保角色分配的有效性和策略的适应性。

第二种方法则结合了当代大语言模型的处理能力与分层强化学习策略，通过动态角色分配和策略生成，为每个智能体提供定制化的决策支持，从而提高整个系统的协同效率和灵活性。

通过实验验证，这些算法在多种多智能体合作任务中显示出优于现有基线模型的性能，特别是在处理未见过的团队配置和智能体策略时表现出色。这不仅展示了算法在实际应用中的潜力，也为深入理解多智能体合作的复杂性和泛化能力提供了实证基础。

关键词：可泛化多智能体合作；多智能体强化学习