

plain_text (0.93)

处不在,但即使在最先进的多智能体系统中也很难实现。

于是,人们尝试了很多关于智能体自主意图的研究工作^[1]。Qi 等^[49]提出了一种意图感知的多智能体规划框架以及学习算法,在此框架下,智能体计划在目标空间中最大化预期效用,并且在规划过程中考虑了其他智能体的意图;他们提出了一个简单但有效的效用函数的近似函数,而不是将学习问题公式化为部分可观察的马尔可夫决策过程。Raileanu 等^[50]提出了一种新的学习方法 SOM(Self Other-Modeling),其中智能体使用自己的策略来预测另一个智能体的策略,并在线更新其隐藏的状态。Rabinowitz 等^[51]提出了一种心智理论的神经网络(Theory of Mind neural network, ToMnet),该网络使用元学习,通过观察其行为来构建智能体所遇到的模型。研究者将 ToMnet 应用到简单的格子环境中的智能体上,研究结果表明它可以学习和模拟来自不同群体的行为,并且它通过了经典的 ToM 任务测试。可以通过训练 ToMnet 来预测其他智能体的意图,并且其明确地揭示了其他智能体的错误意图。

title (0.96)

3.2.3 奖励机制 2

多智能体系统中的奖励机制比单智能体中的奖励机制^[3]加复杂。目前,在 MARL 中应用的一些奖励信号存在问题。比如,全局奖励信号向所有智能体分配相同的奖励而不区分它们的贡献,这可能会鼓励懒惰的智能体;而本地奖励信号仅基于个体行为向每个智能体提供不同的本地奖励,这又可能会产生自私的智能体。如何设计一个合适的奖励信号来加速学习和稳定收敛,是一个关键问题。有研究人员设计了一些混合的奖励信号,智能体通过混合信号能学习到更好的策略,实验结果也得到提高。除了混合信号外,Omidshafiei 等^[52]的研究表明在智能体之间添加通信机制能显著提高团队的最终绩效,他们第一次在多智能体强化学习中引入了教学机制。其主要贡献是将元学习的思想引入合作的多智能体强化学习中,将智能体的学习进度作为奖励反馈机制。

title (0.93)

4 MARL 的应用和前 4

plain_text (0.98)

近几年,MARL 在许多领域都得到了实际应用,目前^[5]机器人系统、人机博弈和自动驾驶等领域开展的研究较多。在机器人系统领域,Gu 等^[53]提出了一种基于离线策略的深度强化学习算法,其可以有效地训练真实的物理机器人。机器人可以在没有任何演示或手动设计的情况下,学习各种仿真与复杂的操作技巧。Foerster 等^[10]将多智能体强化学习方法应用到机器人交流领域,他们采用集中学习分散执行的方式,第一次实现了机器人之间的深层次交流。Duan 等^[54]在机器人控制上的实践工作对多智能体强化学习的应用也具有启发意义。人机博弈一直都是人工智能领域最具挑战性也最令人兴奋的工作,2018 年 Open AI 和 DeepMind 相继取得重大进展。Open AI 在实时 5v5 策略游戏 dota2 中战胜人类顶级玩家;DeepMind 在复杂的第一人称多人游戏 QuakeIII 中达到人类水平,还能与人类玩家合作^[11-13]。在自动驾驶领域,Shalev-Shwartz 等^[14]针对自动驾驶的安全性和环境的不可预测性问题进行了改进和优化,并展示了如何在没有 MDP 假设的情况下使用策略梯度迭代,以及用随机梯度上升来最小化梯度估计的方差。

plain_text (0.98)

可以预见的是,未来 MARL 还将被更广泛地应用到各^[6]各业,如互联网、资源管理、交通系统、医疗和金融等领域。目前,在互联网领域,Ji 等^[15]将聚类的思想与 MARL 方法相结合,面对大量的广告商,对即时在线投标的性能进行了优化;为了平衡广告商之间的竞争与合作,提出并实现了一种实用的分布式协调多智能体竞标算法。在资源管理方面,Xi 等^[16]提出了一种 MARL 新算法,新算法不基于马尔可夫假设,具有更快的收敛速度和更强的鲁棒性,使得电网系统能够在更复杂的条件下提高新能源的利用率。Perolat 等^[17]运用部分马尔可夫观测模型对公共资源的占用主体进行了建模,揭示了排他性、可持续性和不平等性之间的关系,并提出了解决方案,提高了资源管理能力。Kofinas 等^[55]提出的模糊 Q 学习方法有效地提高了分散式微电网的能量管理能力。在交通控制领域,Chen 等^[56]提出了一种基于 MARL 的协同控制框架,并用其来实时缓解公交车道上的公交拥挤。Vidhate 等^[57]提出了一种基于协同多智能体强化学习的交通流模型用于控制优化交通系统,模型能够很好地处理未知的复杂状态。目前,多智能体强化学习应用在医疗和金融等领域的研究还较少,需要人们进行更进一步的探索。

plain_text (0.98)

MARL 虽然已经在诸多领域中实现了应用,但依然存^[7]很多问题。在 MARL 未来的研究工作和实际应用中,有一些方向需要进行进一步的关注和探索。首先,可扩展性依然是最核心的问题。目前,MARL 在电脑游戏、人机博弈(包括小型机器人体系)中已经取得了不错的成果,其中 DeepMind 公司和 Open AI 公司做出了重要的贡献。但是,MARL 在自动驾驶、船舶制造、商业交易、资源配置等大型机器人体系中的应用还不够成熟,因此将 MARL 应用到海量智能体以及更复杂的环境中是未来研究的重要方向。其次,现在普遍的多智能体强化学习都是基于马尔可夫随机过程环境的,它给许多模型提供了一个简单明确的数学框架,但是现实环境中很多场景和问题是非马尔可夫的,在大部分多智能体场景中,其他智能体的行为是不可预测的。上文提到的 Shalev 等^[14]和 Xi 等^[16]提出的 MARL 新算法都不基于马尔可夫假设。因此,在没有马尔可夫假设的情况下建模实现 MARL,需要进行进一步的研究与探索。最后,无监督的生成模型、计算图模型、注意力机制等其他机器学习方法与 MARL 的结合可以解决许多实际应用的问题。迁移学习也将被更多地应用到 MARL 方法中,以缓解真实任务场景中训练数据缺乏的问题。借助云服务器端的分布式的多智能体协同学习也是未来^[8]一个重要的方向。

plain_text (0.98)

结束语 本文对多智能体强化学习进行了简要分析,^[8]包括 MARL 的基本背景、经典算法、研究进展以及实际应用。在许多现实问题和领域中,多智能体强化学习都展现出了巨大的潜力,研究成果层出不穷,各种算法不断涌现。本文首先介绍了多智能体强化学习的理论背景和经典算法,包括马尔可夫框架和随机博弈模型等;然后主要从算法的可扩展性方面详细综述了近期多智能体强化学习算法的创新和改进;最后对多智能体强化学习的实际应用和未来前景展开讨论。多智能体强化学习是一个新兴的、活跃的、快速扩展的研究领域,还有很多问题值得进一步探索和研究。随着 MARL 理论和方法研究的不断深入,可以预见,“解决智能,并用智能解决