

陈佳煜

北京市海淀区双清路 30 号清华大学

(+86) 1733 1878 868 ◊ jiayu-ch19@mails.tsinghua.edu.cn

教育背景

(本科) 清华大学电子工程系	2015.8 - 2019.7
(博士) 清华大学电子工程系	2019.8 - 2024.12 (预计毕业)

科研经历

清华大学电子工程系线路与系统研究所 2019.9-至今

指导教师: 杨华中, 清华大学电子工程系长聘教授; 汪玉, 清华大学电子工程系长聘教授; 吴翼, 清华大学交叉信息学院助理教授

研究领域: 多智能体强化学习; 课程学习

研究高数据效率多智能体强化学习方法与系统, 从算法和系统应用 2 个层面展开研究。主要包括

- 设计针对稀疏奖励任务的高效自动化课程学习训练算法。这项工作中, 我们介绍了一种变分自动课程学习算法 (VACL), 从变分角度, 将强化学习目标分解成两部分: 当前分布上的任务学习和对新任务分布的课程更新。我们解决了具有大量智能体的稀疏奖励问题。使用一台计算机, VACL 在二维粒子世界环境中使用 100 个智能体实现了 98% 的覆盖率, 并重现了最初在 OpenAI 的捉迷藏项目中的使用梯子的行为。
- 在复杂的零和博弈中用多智能体强化学习 (MARL) 来学习纳什均衡 (NE), 计算成本极高。课程学习是加速学习的有效方法, 但如何生成特定难度的子博弈 (从特定状态开始诱发的博弈) 来设计课程是一个未被充分探索的领域。在这项工作中, 我们提出了一个新的零和博弈的子博弈课程学习框架。它采用了一种自适应的初始状态分布, 将智能体重置到一些以前访问过的状态, 从而使得他们可以快速学习以提高性能。在这个框架的基础上, 我们推导出一个子博弈的选择度量, 近似于 NE 值的平方距离, 并进一步采用基于粒子的状态采样器来生成子博弈。综合这些技术, 我们提出了新的算法: Subgame Automatic Curriculum Learning (SACL)。SACL 可以与任何 MARL 算法相结合, 如 MAPPO。在二维粒子世界和谷歌足球环境中的实验表明, SACL 产生的策略比基线强得多。在具有挑战性的捉迷藏环境中, SACL 产生了所有四个突发阶段, 并且只使用了 MAPPO 一半的样本。
- 本文研究的是多智能体追捕-逃跑问题, 即在有障碍物的狭窄环境中, 慢速追逐者合作捕捉快速逃跑者。现有方法主要集中于二维追逐-逃跑任务, 任务初始条件固定, 包括场景大小、逃跑者和追逐者的速度比、障碍物的数量和速度等。在这项工作中, 我们介绍了一种灵活的多智能体追捕任务求解器 (TaskFlex Solver, TFS), 它能够求解二维和三维场景中具有动态任务条件下的追捕-逃避任务。TFS 利用课程学习算法, 根据训练进度构建任务分布, 从而提高训练效率和最终性能。我们的算法由两部分组成: 任务评估器 (Task Evaluator) 用于评估任务成功率, 并选择难度适中的任务; 任务采样器 (Task Sampler) 用于构建训练分布, 从而最大限度地改善当前策略。我们在二维多智能体粒子环境和基于 NVIDIA Isaac Sim 的三维环境中对 TFS 进行了评估, 其中既有固定任务配置, 也有动态任务配置。我们的方法优于非学习方法和强化学习方法, 在所有场景中都取得了接近 100% 的成功率。
- 设计异步多智能体强化学习训练算法, 提高强化学习策略从仿真器到真实世界的泛化能力。我们将经典的多智能体强化学习算法 MAPPO 扩展到异步设置, 并额外应用动作延迟随机化来强制策略泛化到现实世界中的不同动作延迟。在简单的网格场景中, 我们的方法与经典方法相比, 减少了超过 10% 的实际探索时间。在基于视觉的高保真 habitat 环境, 实现了 28% 的性能提升。

- 设计了多智能体神经拓扑映射算法 (Multi-Agent Neural Topological Mapping, MANTM), 提高多智能体探索任务的探索效率和泛化能力。MANTM 主要包括拓扑映射器和基于 RL 的新型层次拓扑规划器 (HTP)。拓扑映射器采用视觉编码器和基于距离的启发式方法来构建包含主节点及其相应幽灵节点的图。HTP 利用图神经网络以从粗到细的方式捕捉代理与图节点之间的相关性, 从而实现有效的全局目标选择。在物理仿真模拟器 Habitat 中进行的实验表明, 与基于规划的基线相比, MANTM 至少减少了 26.40% 的探索步数, 与基于 RL 的竞争对手相比, 在未见场景中至少减少了 7.63% 的探索步数。

发表文章

- [1] **Jiayu Chen**, Yuanxin Zhang, Yuanfan Xu, Huimin Ma, Huazhong Yang, Jiaming Song, Yu Wang, Yi Wu, “Variational automatic curriculum learning for sparse-reward cooperative multi-agent problems”, Advances in Neural Information Processing Systems (NeurIPS), 2021. **机器学习顶会 CCF-A, 引用量 28**
- [2] **Jiayu Chen**, Zelai Xu, Yunfei Li, Chao Yu, Jiaming Song, Huazhong Yang, Fei Fang, Yu Wang, Yi Wu, “Accelerate Multi-Agent Reinforcement Learning in Zero-Sum Games with Subgame Curriculum Learning”, Association for the Advancement of Artificial Intelligence (AAAI), 2024. **机器学习顶会 CCF-A**
- [3] **Jiayu Chen**, Guosheng Li, Chao Yu, Xinyi Yang, Botian Xu, Huazhong Yang and Yu Wang, “TaskFlex Solver for Multi-Agent Pursuit via Automatic Curriculum Learning”, IEEE Robotics and Automation Letters (RAL). **机器人期刊, SCI 2 区**
- [4] Chao Yu, Xinyi Yang, Jiaxuan Gao, **Jiayu Chen**, Yunfei Li, Jijia Liu, Yunfei Xiang, Ruixin Huang, Huazhong Yang, Yi Wu, Yu Wang, “Asynchronous Multi-Agent Reinforcement Learning for Efficient Real-Time Multi-Robot Cooperative Exploration”, International Joint Conference on Autonomous Agents and Multi-agent Systems (AA-MAS), 2022. **多智能体顶会 CCF-B, 引用 12**
- [5] Xinyi Yang, Yuxiang Yang, Chao Yu, **Jiayu Chen**, Jingchen Yu, Haibing Ren, Huazhong Yang and Yu Wang, “Active Neural Topological Mapping for Multi-Agent Exploration”, IEEE Robotics and Automation Letters (RAL). **机器人期刊, SCI 2 区**

获得荣誉

全国高中生物理竞赛二等奖	2014
清华大学“清华之友-郑格如奖学金”	2016
清华大学校设奖学金	2016
全国大学生物理竞赛一等奖	2016
清华大学“清华之友-泉州英才奖学金”	2023

实习经历

曾在商汤自动驾驶实验室实习, 参与 OpenDILab 自动驾驶场景的代码工作	2021.7-2021.9
---	---------------

技术优势

熟练	Python, MATLAB, PyTorch, Tensorflow
中等	LaTeX, Linux, Git, JavaScript
基本	C, C++