

骆永乐

联系方式: 15609682400 电子邮箱: lyl1994@mail.ustc.edu.cn
作品链接: [Github 10500+stars](#) 博客链接: [知乎: 强化学徒-5K 关注](#)
求职意向: RLHF, 基于 LLM 的具身智能或者 LLM+论文辅助工具

教育经历

郑州大学	自动化	本科	2013-2017
● 社团经历: 担任郑州大学第二十五届马列主义研究会会长			
中国科学技术大学	模式识别与智能系统	研二转博; 博士四年级在读	2017-
● 导师: 双丰研究员 (百人计划, 硕士导师); 宋博研究员 (青年千人计划, 博士导师)			

研究经历

深度强化学习代码库 DRLib:

- 基于 Spinning UP 封装的深度强化学习算法: DQN、DDPG、TD3、SAC、PPO、**PER**、**HER** 等。
- 深度强化学习算法链接: [DRLib](#) (369 star)

稀疏奖励矫正密集奖励的强化学习

- 论文综合稀疏奖励全局收敛但效率低下以及密集奖励收敛快但容易局部最优的特点, 提出 dense2sparse 解决方案, 兼顾二者的优势, 同时提高探索效率和最终性能。
- 《Balance Between Efficient and Effective Learning: Dense2Sparse Reward Shaping for Robot Manipulation with Environment Uncertainty》(共一, 机器人会议 2022 AIM, Oral Presentation)
- 23 年改进版《D2SR: Transferring Dense Reward Function to Sparse by Network Resetting》, 彻底解决稳定性的问题, 性能大幅提升 (一作, 机器人 EI 会议 RCAR, 本月底, 出结果)

乒乓球仿真搭建和真机验证—深度强化学习的单步决策高效学习

- 基于 Mujoco 物理引擎的乒乓球击球平台 (唯一一个基于 Mujoco 的乒乓球环境), 实现与真机类似的击球效果。
- 《SIRL: Self-Imitation Reinforcement Learning for Single-step Hitting Tasks》(一作, 机器人会议 ARM, 同上)

自我引导持续强化学习—彻底解决深度强化学习, 在稀疏奖励下复杂序列任务中效率低下的问题

- 首次提出自我引导探索的强化学习框架。面对奖励反馈稀疏的复杂任务, 该算法可以让智能体从失败中提取有效信息, 积极探索, 不断积累优势, 最终实现高效学习。在一到三物体的各类操作任务中都取得极高探索效率, 真机实验从零开始训练仅需 250 回合即可达 100%成功率, 是本人**博士期间最有学术价值的工作**。
- 在此基础上的另外一个侧重于策略优化的工作正在撰写, 可以使得样本效率再次提高 60%以上。
- 代码已开源: [RHIER](#); 论文已在 ArXiv 公布: [Relay Hindsight Experience Replay](#) (一作, NeuroComputing 刚拿到大修结果, 审稿人对写作, 创新点以及性能提升非常认可)

证书及项目经历

- 证书: 英语四级、英语六级、计算机二级、心理咨询师三级
- 项目经历:
 - 基于强化学习的竞技型乒乓球机器人运动控制系统研发 (横向, 148w, 本人负责 仿真系统搭建和强化算法)
 - 开源 [ChatPaper](#), 获得 10K star, GitHub 连续三天热榜第五, 正在商业化, 并在推动 ChatOpenReview

自我评价

- 熟练掌握经典深度强化学习算法, 拥有丰富的机器人仿真和真实系统搭建经验, 7 成几率秋季毕业, 3 成几率明年毕业。
- 品行良好, 为人坦率靠谱。工程能力强, 编程基础扎实。擅长将人类学习经验应用于人工智能领域, 科研能力优秀, 拥有丰富的团队合作经验, 热爱开源、技术分享和教学。
- 对于目前的强化发展预估是, 简单基于 MLP 的强化学习已经结束, 下一步必须融入 LLM, 而 LLM/LMM 的世界模型, 是机器人落地的基础, 才能满足实际生产需求。
- 个人对 ChatPaper 系列的规划, 是希望能做一个垂直领域, 为科研工作者加速, 并且整个项目适合商业化运营。
- 希望能有一个机会, 加入这次 LLM 的发展浪潮中。