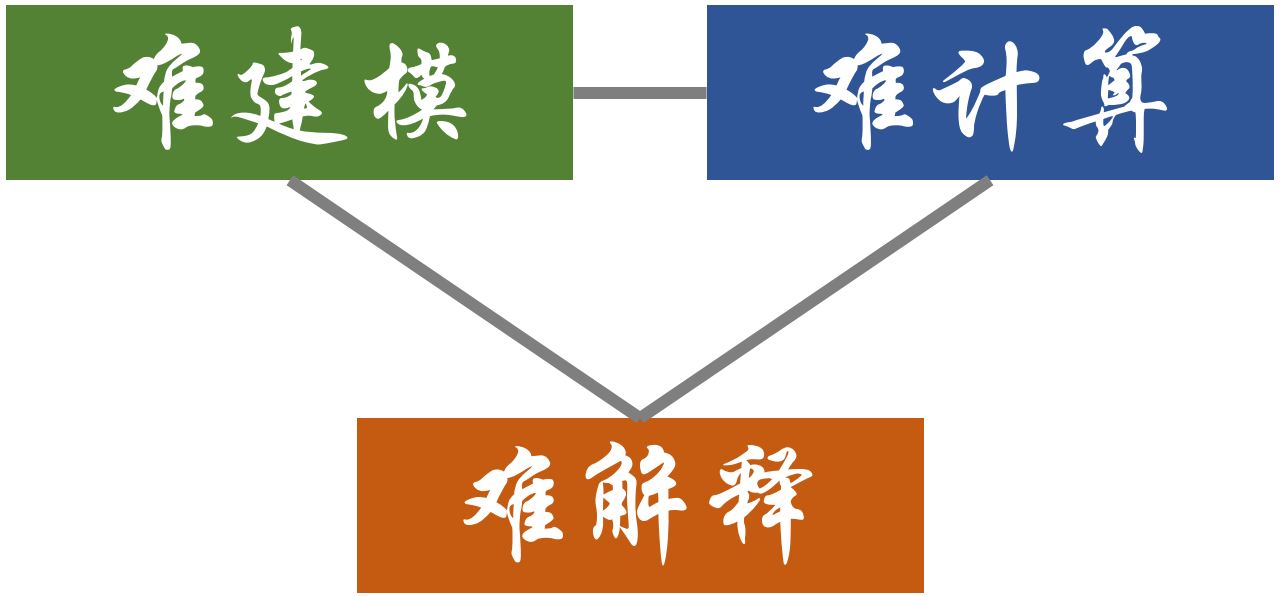
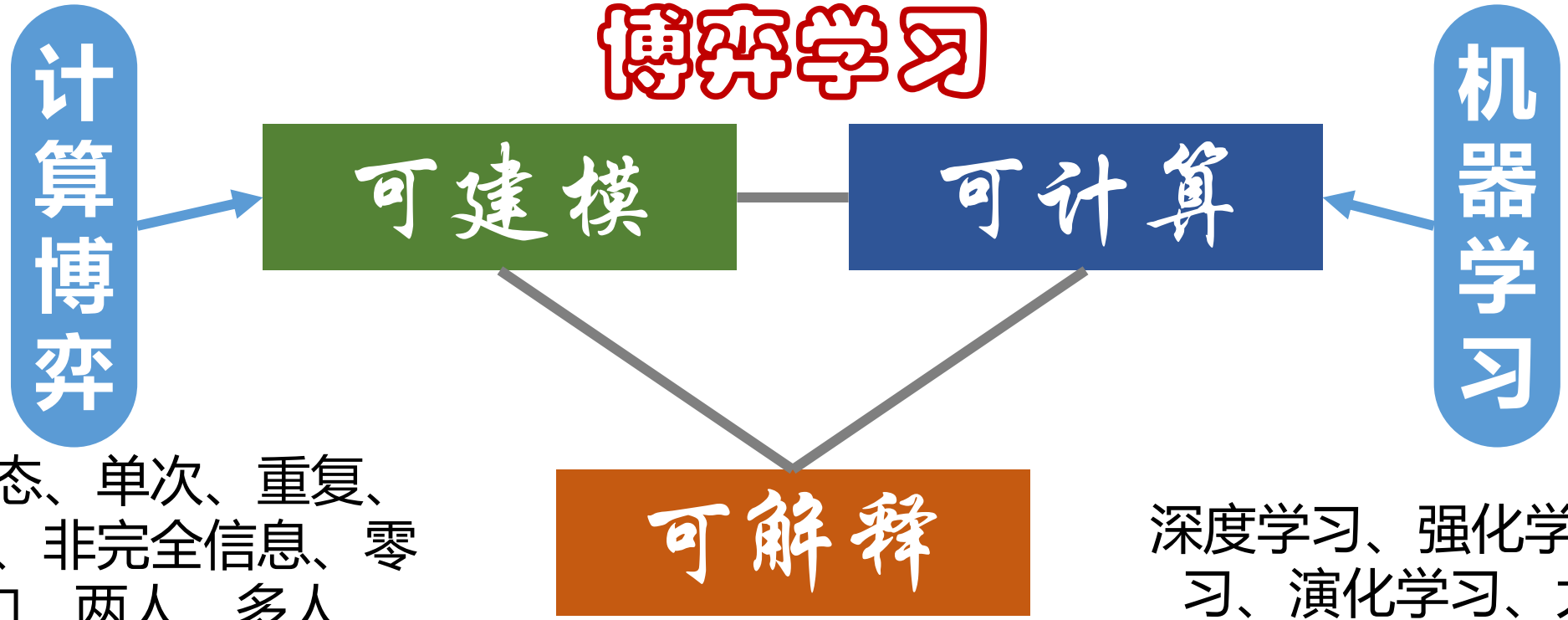
研究方向：博弈学习

- 现实世界复杂博弈决策问题求解面临诸多挑战

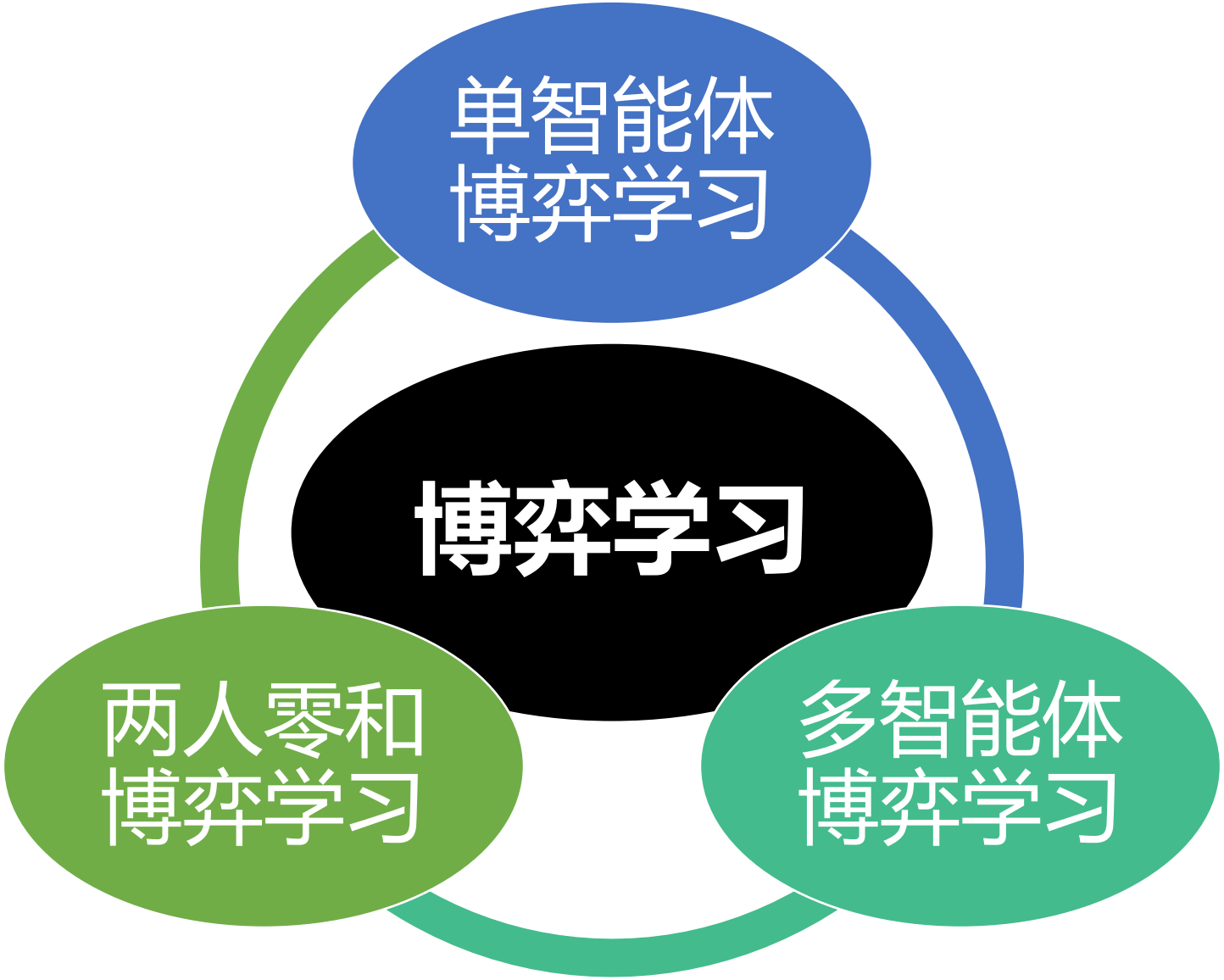


研究方向：博弈学习

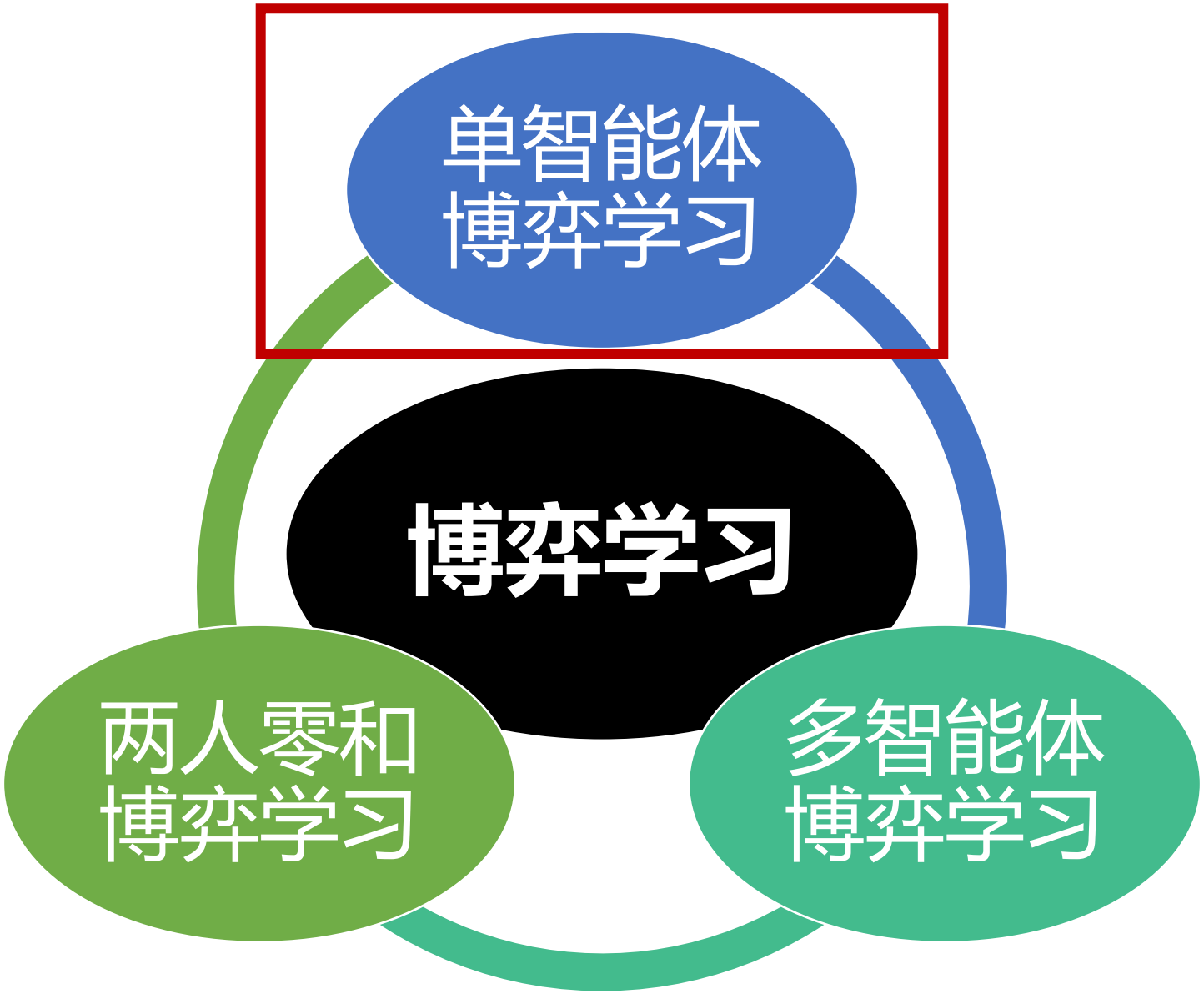
- 目标是实现复杂博弈决策问题的可建模、可计算和可解释的结合



具体研究工作介绍



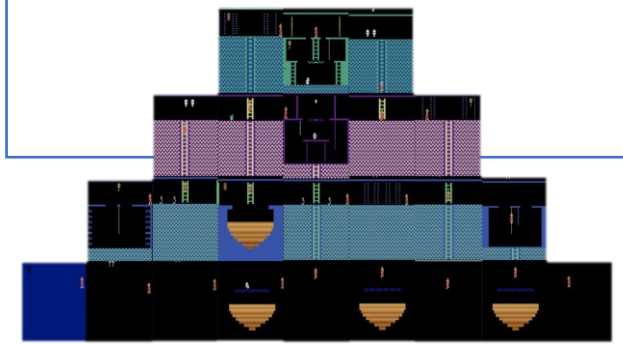
单智能体博弈学习相关工作介绍



单智能体博弈学习：学习目标与关键问题

学习目标

智能体与外部环境的博弈，极大化期望累积奖励



关键问题

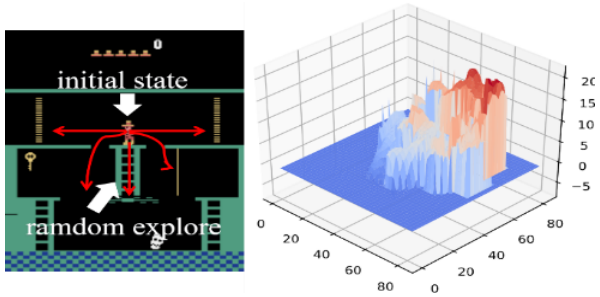
- 如何对环境进行有效探索
- 如何高效完成多个任务

博弈/学习视角

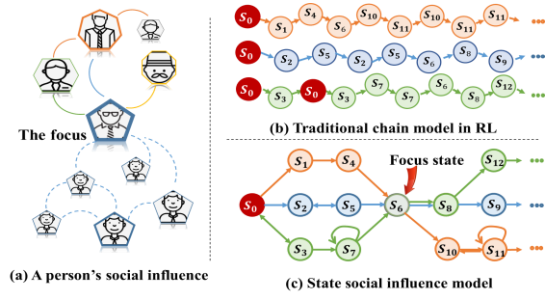
- 效用最大化/深度强化学习
- 帕累托最优/多任务学习

单智能体博弈学习：高效探索学习系列工作

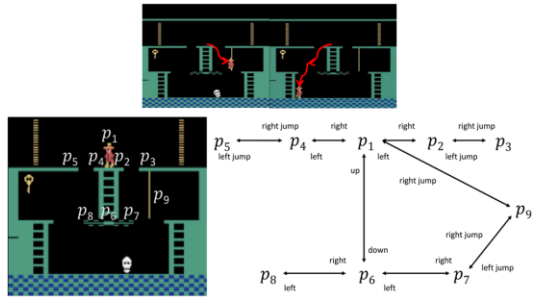
- 从不同角度出发提出了一系列单智能体高效探索学习算法，提高了智能体在复杂环境中的学习效率



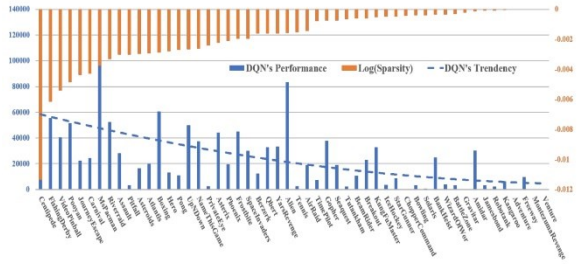
基于势能化经验回放的智能体探索算法，IJCAI 2021



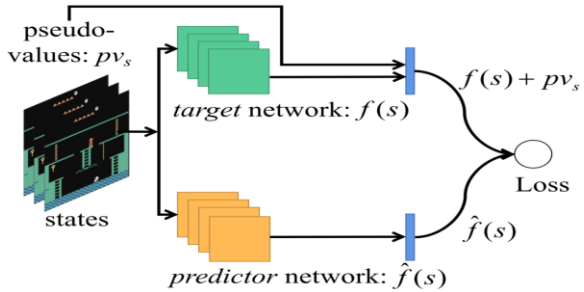
基于社交影响力的智能体探索算法，AAAI 2021



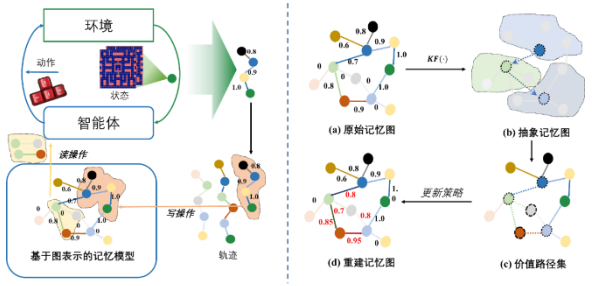
基于图表示学习的智能体探索算法，IJCNN 2021



基于稀疏度度量的智能体探索算法，ICONIP 2022



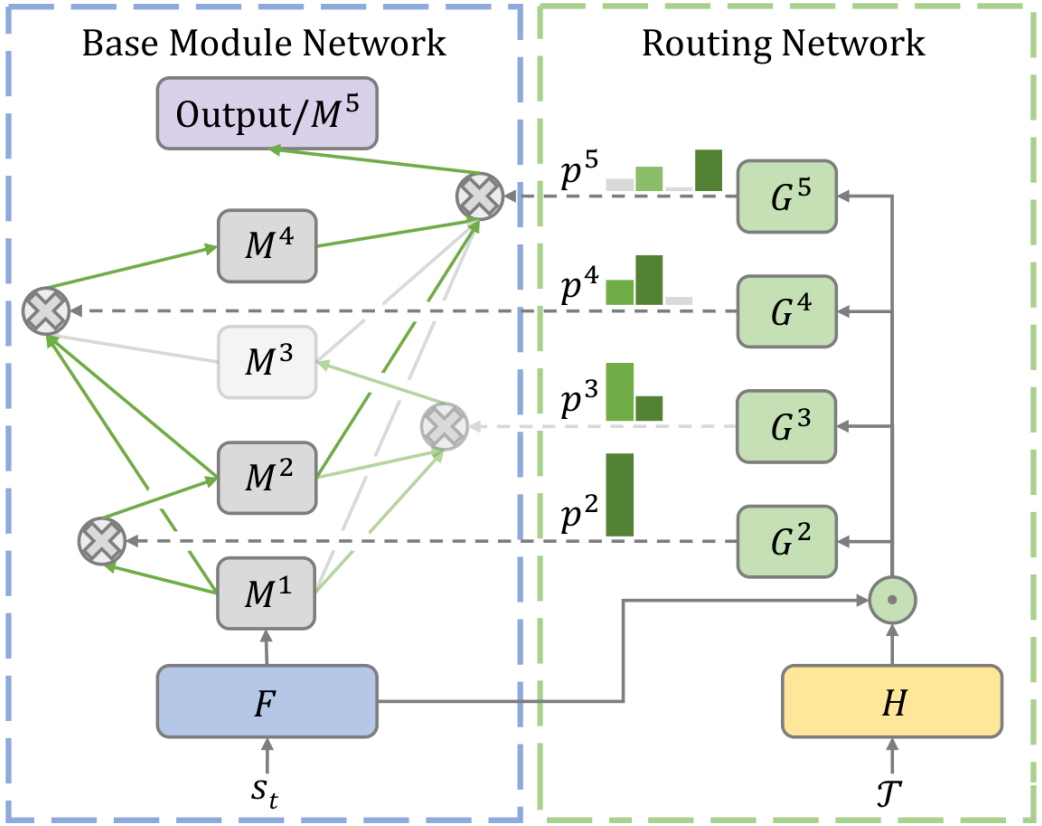
基于伪价值网络蒸馏的智能体探索算法，IJCNN 2022

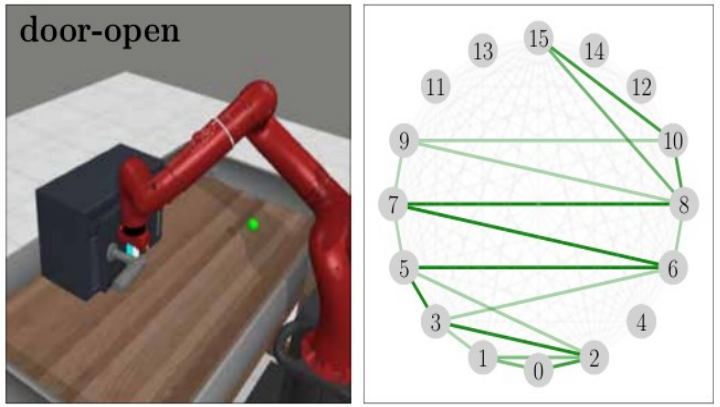


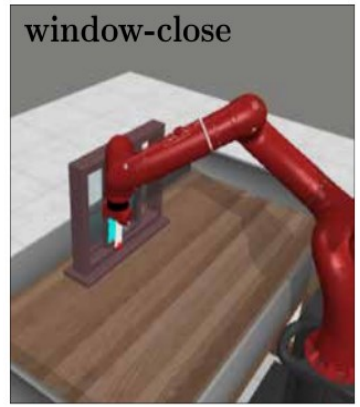
基于图记忆重构的智能体探索算法， TAI 2023

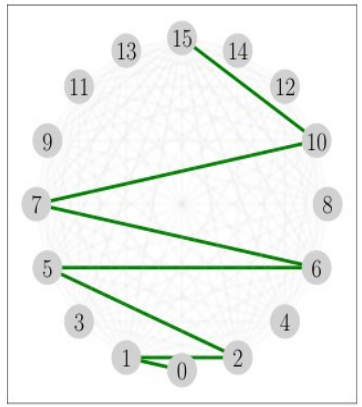
单智能体博弈学习：多任务强化学习（工作一）

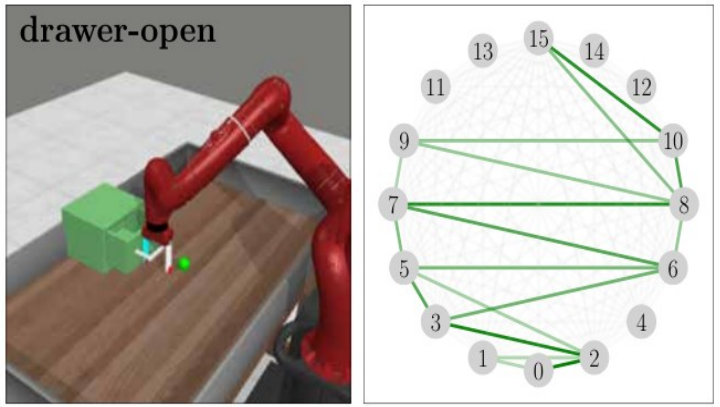
- **基于动态深度路由的多任务强化学习算法**
  - 不同任务难度不同，所需策略复杂度不同，动态深度路由可以为任务选择合适的网络结构，大大提高各任务的执行效率，AAAI 2024

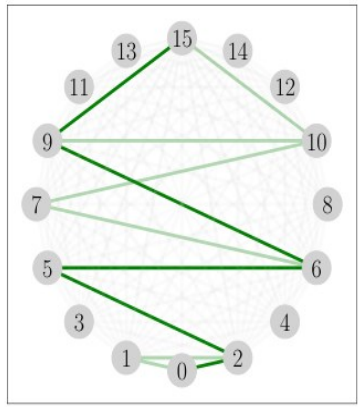






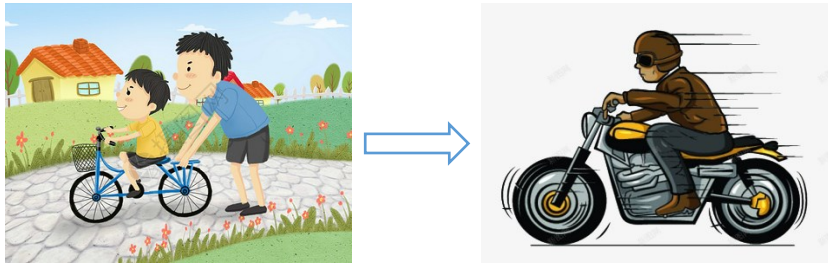




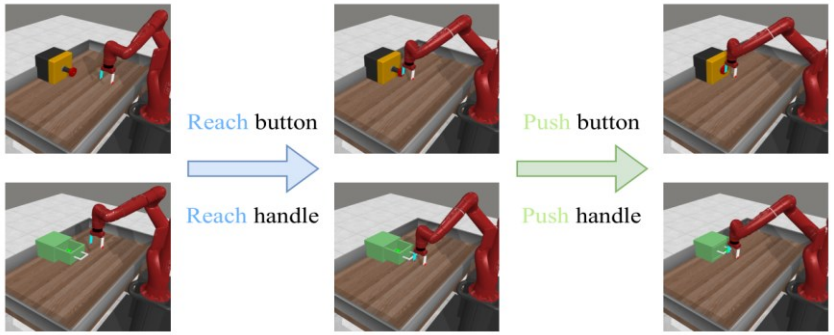


单智能体博弈学习：多任务强化学习（工作二）

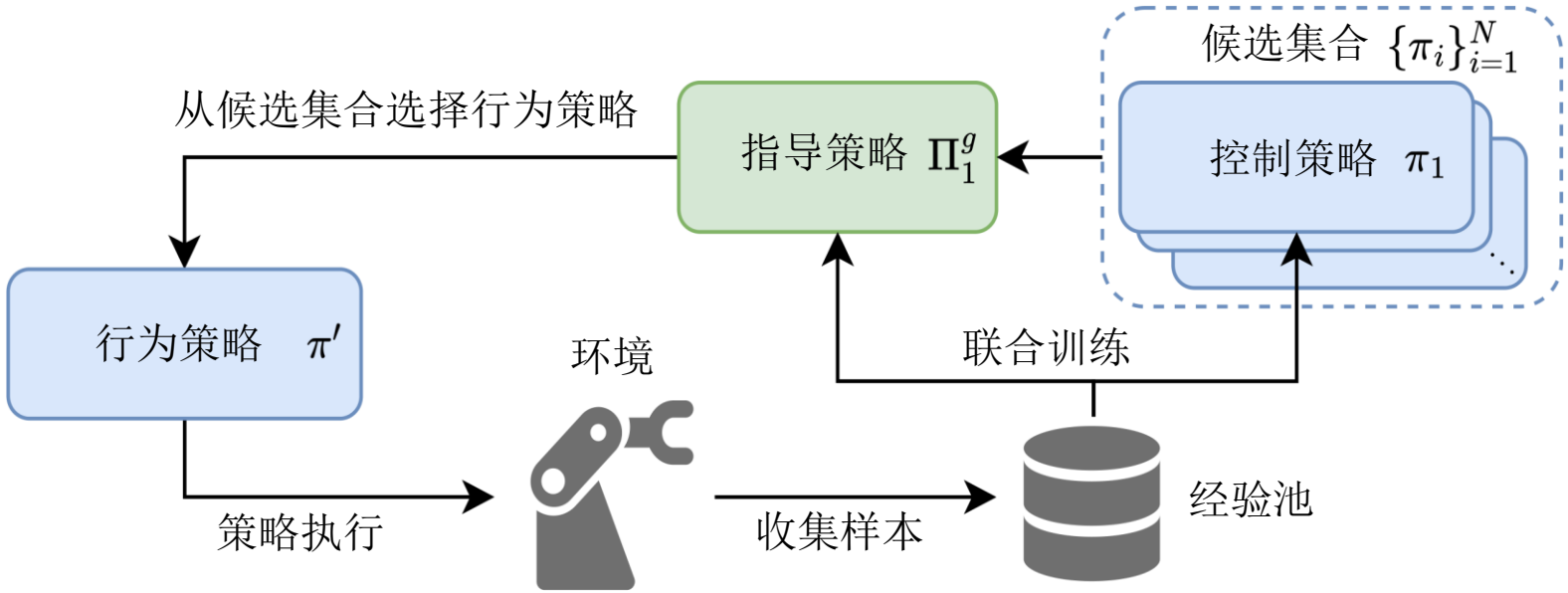
- **跨任务经验指导的多任务强化学习框架**
  - 通过学习一个指导策略来自适应地决定何时采用哪个任务的经验来辅助当前任务的学习，利用已掌握任务的经验知识，引导困难任务的探索学习，从而达到多任务高效联合学习的目的



人类的经验共享机制

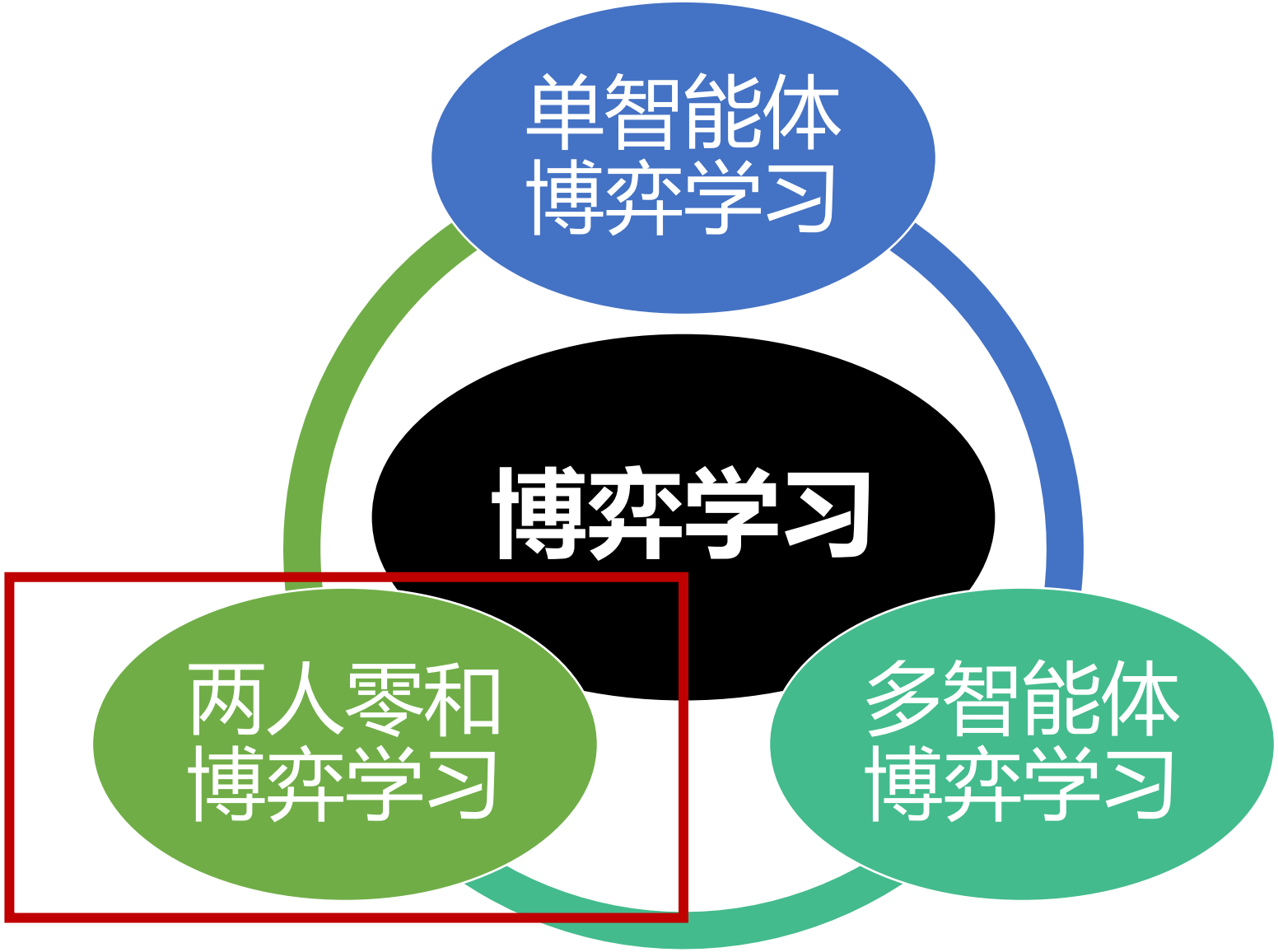


通用智能体也存在经验共享的可能



跨任务经验指导的多任务策略学习架构

两人零和博弈相关工作介绍



两人零和博弈学习：学习目标与关键问题

目标1

确保不输

- **博弈均衡求解**
  - **关键问题：**大规模复杂博弈如何高效求解均衡
  - **博弈/学习视角：**纳什均衡/AutoML、演化学习

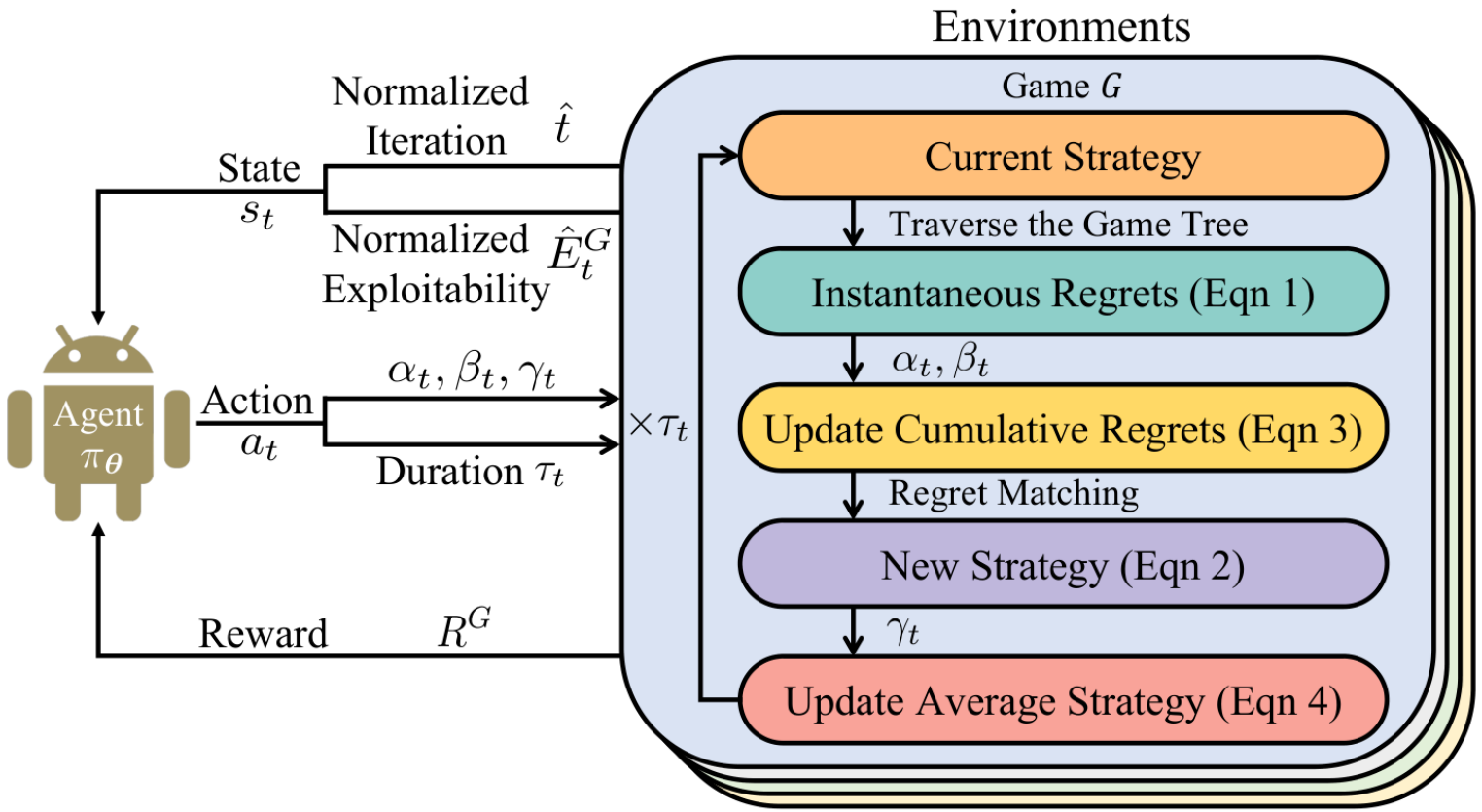
目标2

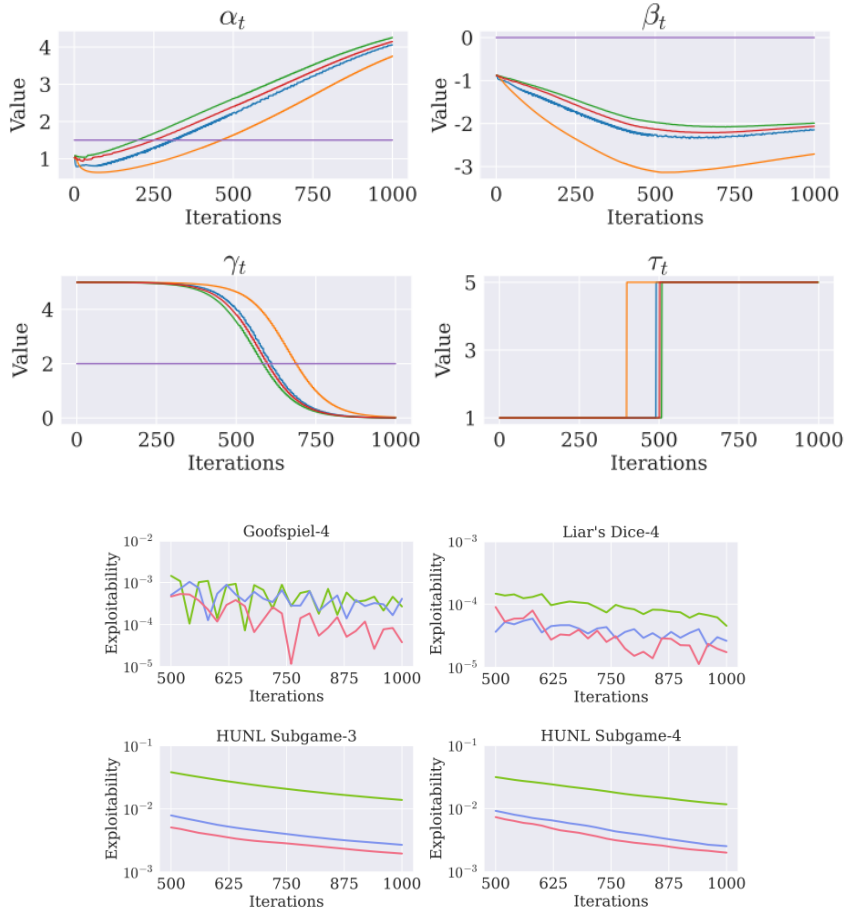
争取多胜

- **对手行为建模**
  - **关键问题：**如何实现快速、高效、精准的对手建模
  - **博弈/学习视角：**最优反应/元学习、In-Context学习

两人零和博弈学习：高效纳什均衡求解（工作一）

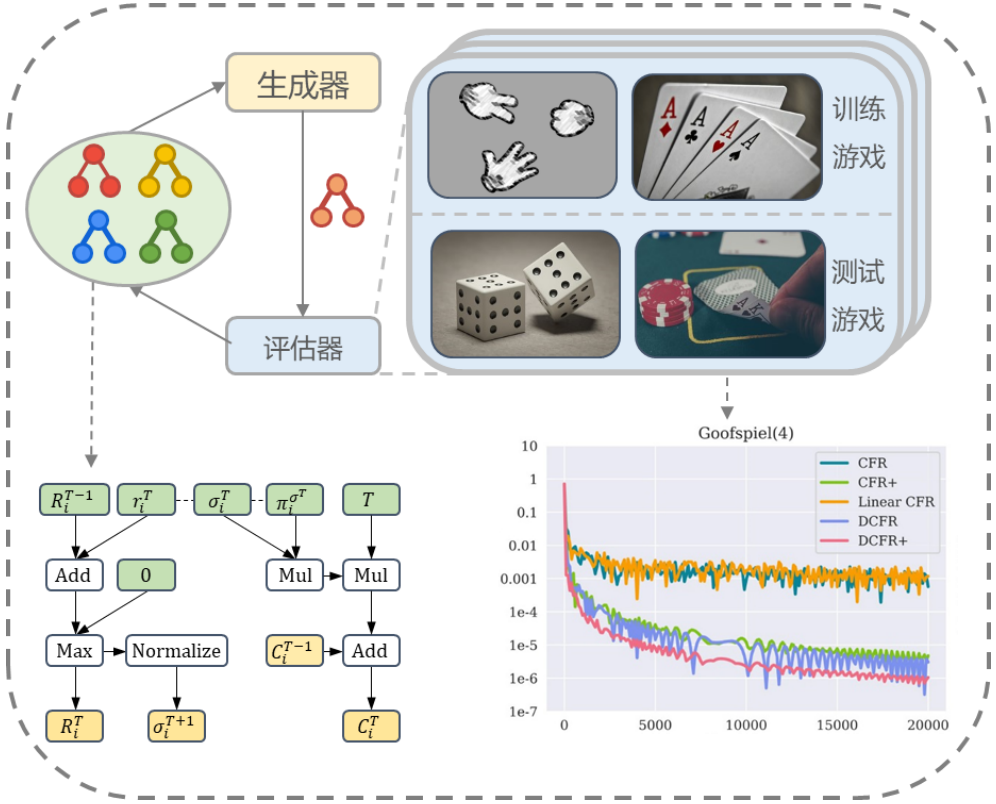
- **基于动态超参调整的纳什均衡求解算法**
  - 将超参数调整的过程转化为智能体的决策过程，利用强化学习训练自动调参策略
  - ICLR 2024，Spotlight, Top 5%



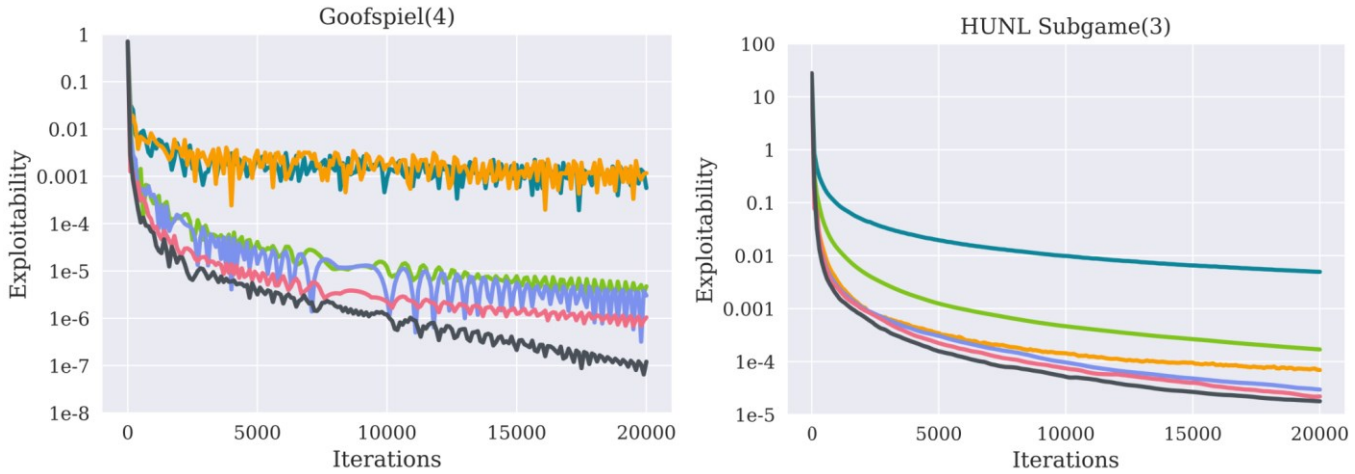


两人零和博弈学习：高效纳什均衡求解（工作二）

- **自动机器学习驱动的纳什均衡求解算法搜索框架**
  - 利用元学习和演化搜索技术来寻找高性能可泛化的纳什均衡求解算法新变体
  - AAAI 2022

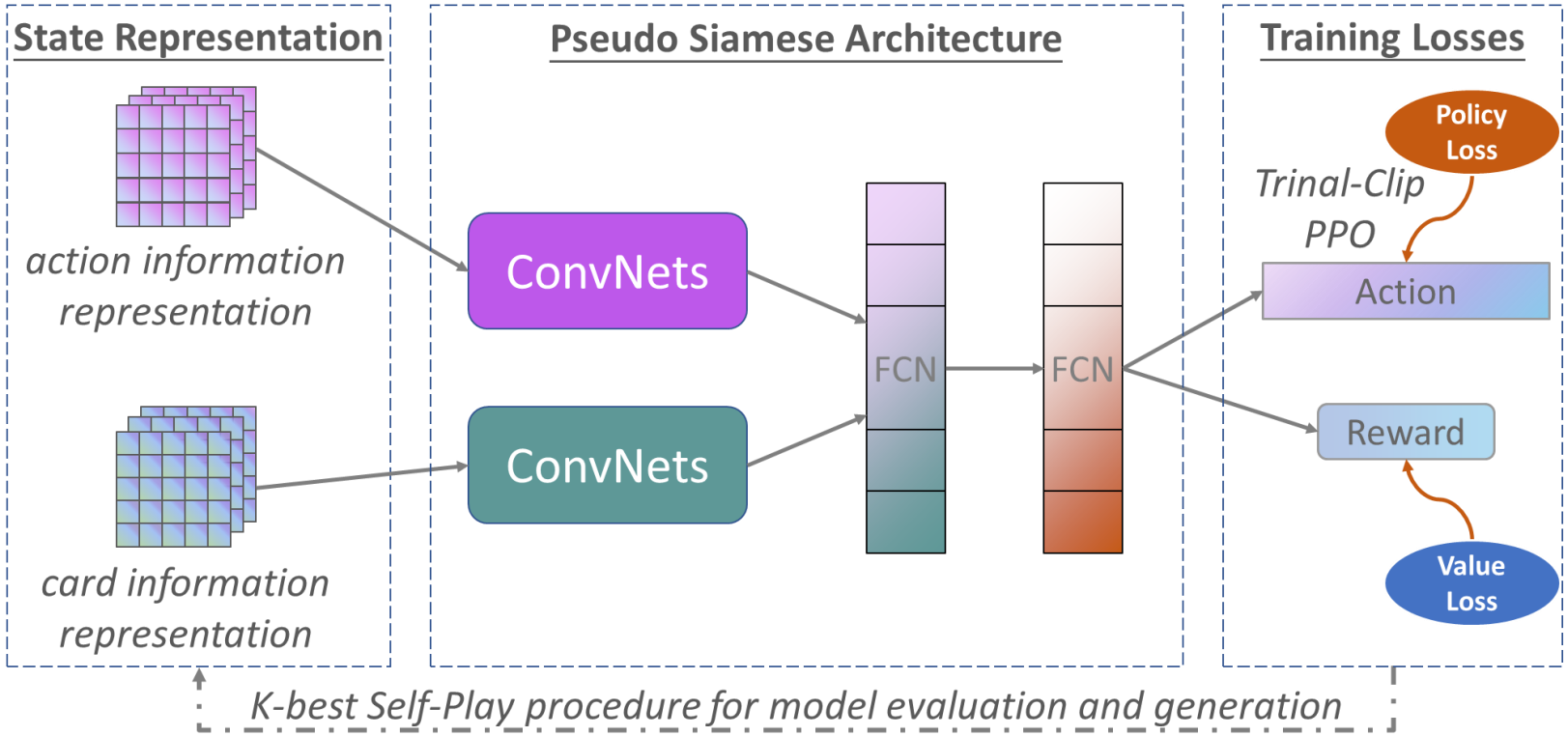


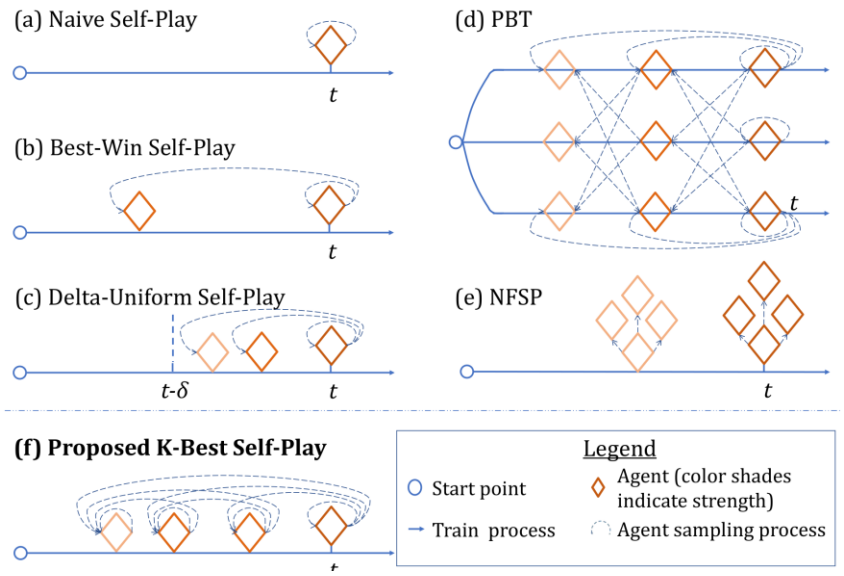
Theorem 1. *Assume that T iterations of DCFR+ are conducted in a two-player zero-sum game. Then the weighted average strategy profile is a $5|\mathcal{I}|\Delta((\sqrt{|A|}+1.5)\sqrt{T}+1.5\sqrt{|A|}(\ln T+1))/T$-Nash equilibrium.*



两人零和博弈学习：高效纳什均衡求解（工作三）

- **自博弈强化学习驱动的大规模非完备信息博弈求解框架**
  - 不依赖领域知识，端到端从零开始自博弈学习；三截断PPO损失函数；K-Best自博弈算法；**AAAI 2022卓越论文奖**





两人零和博弈学习：高效纳什均衡求解（工作四）

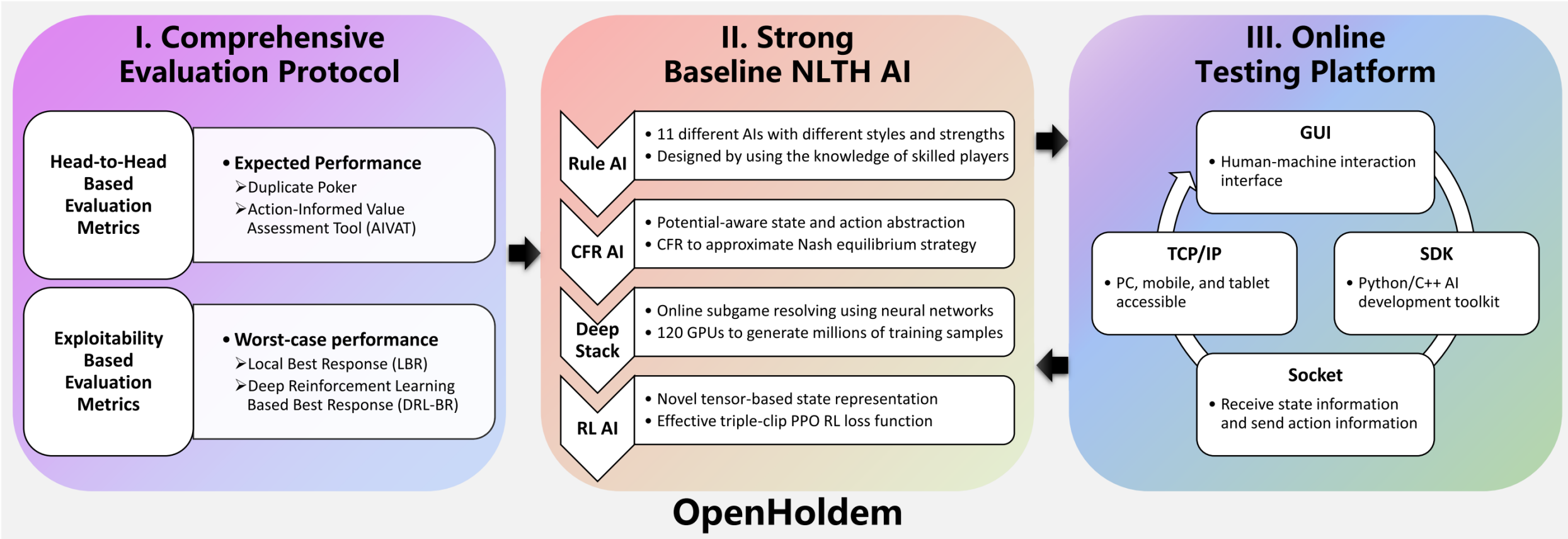
- **具有理论收敛保证的深度强化学习驱动的纳什均衡求解算法**
  - 传统计算博弈算法：具有理论收敛保证、可扩展性差
  - 深度强化学习算法：可扩展性强、缺乏理论收敛保证
  - 有机结合两者的优点，形成一种收敛性质有保证、计算规模可扩展的新算法
  - ICLR 2022

Theorem 1. *NW-CFR is equivalent to a type of weighted CFR with Hedge when* $w_t(s) = f_p^{\mu_t}(s) > 0$, *given that enough trajectories are sampled and* $y(a|s;\theta_t)$ *is sufficiently close to* $R_t^a(s,a)$. *Further, if* $\eta(s) = \sqrt{8\ln|\mathcal{A}(s)|/\{[w_h(s)]^2\Delta^2(s)T\}}$ *and* $w_t(s) = f_p^{\mu_t}(s) \in [w_l(s), w_h(s)] \subset (0,1], t = 1,\ldots,T$, *the average policy*[5] $\bar{\pi}$ *of the corresponding weighted CFR with Hedge and equivalently NW-CFR with* $\bar{\pi}_p(a|s) = \sum_{t=1}^{T}\left[f_p^{\pi_t}(s)\pi_t(a|s)\right]/\sum_{t=1}^{T} f_p^{\pi_t}(s), \forall p \in \mathcal{P}$, *has* ϵ *exploitability after* T *iterations, where*

$$\epsilon \leq |\mathcal{S}|\Delta\sqrt{\frac{1}{2T}\ln|\mathcal{A}|} + \Delta\sum_{s\in\mathcal{S}}\frac{w_h(s)-w_l(s)}{w_h(s)}.$$

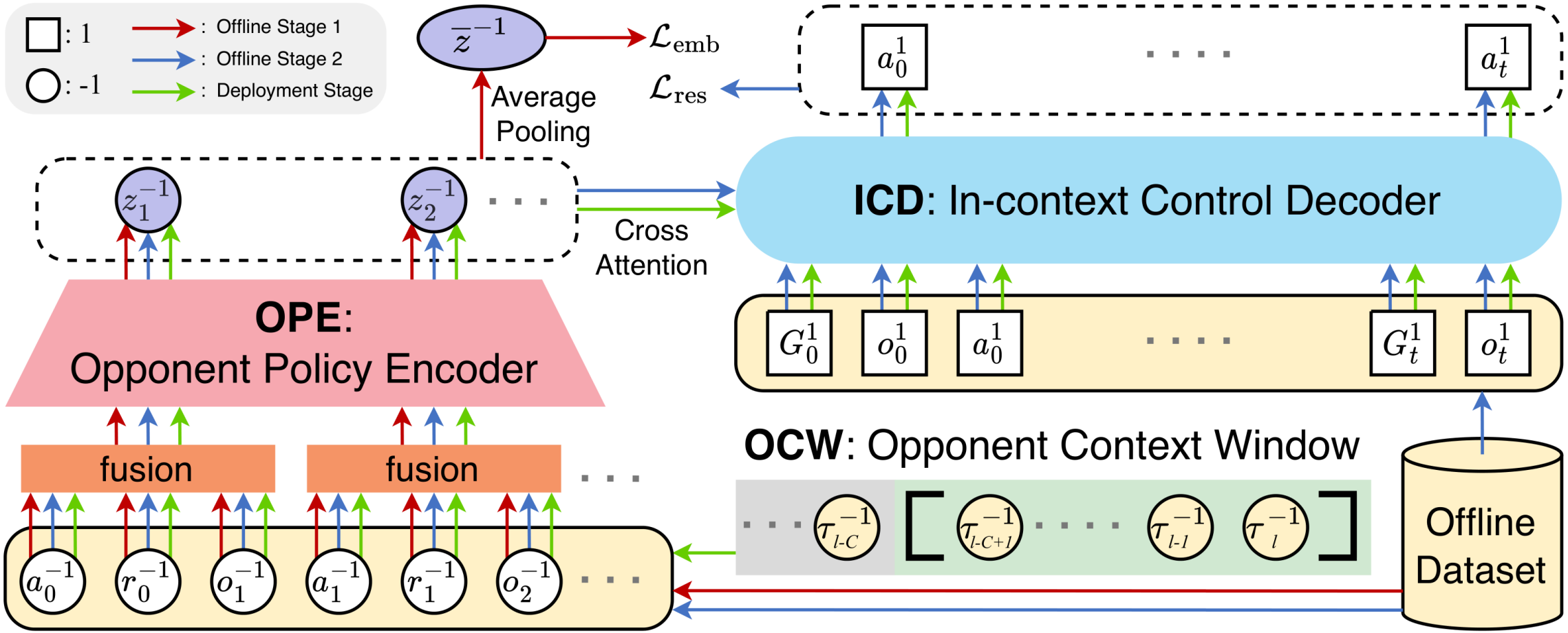
两人零和博弈学习：高效纳什均衡求解（工作五）

- **学界首个大规模非完备信息博弈研究平台OpenHoldem**
  - 标准的评估体系、丰富的基准算法、在线人人、人机、机机对抗接口
  - TNNLS, 2023



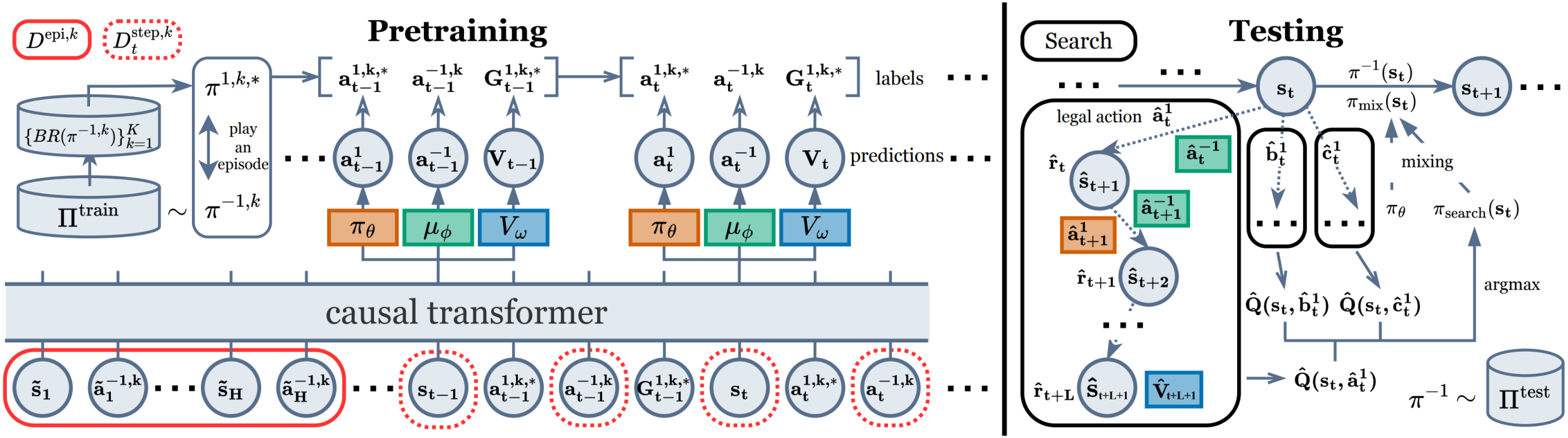
两人零和博弈学习：对手建模（工作一）

- **基于Transformer架构的对手建模算法**
  - 利用Transformer的in-context能力来有效学习对手特点并加以应对，ICLR 2024



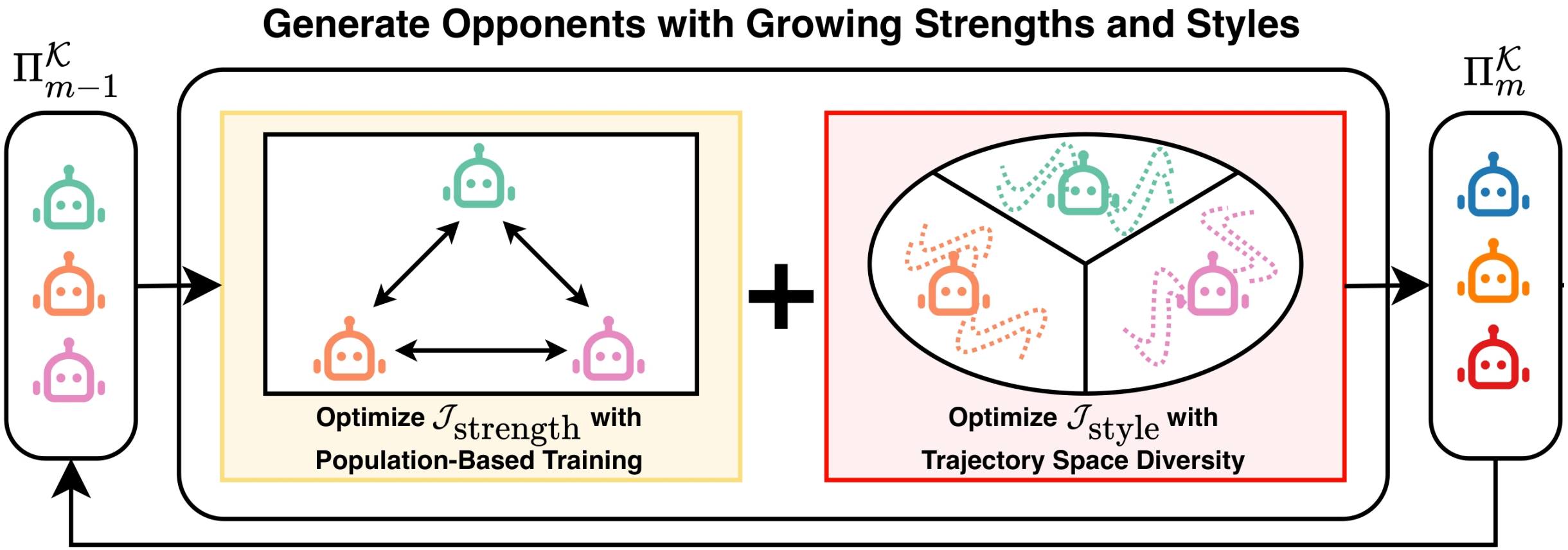
两人零和博弈学习：对手建模（工作二）

- **基于在线搜索的高效对手建模算法**
  - 在上一个工作基础上加入在线搜索过程大大提高性能
  - Richard S. Sutton's "The Bitter Lesson":
    - ***The two methods that seem to scale arbitrarily are learning and search***



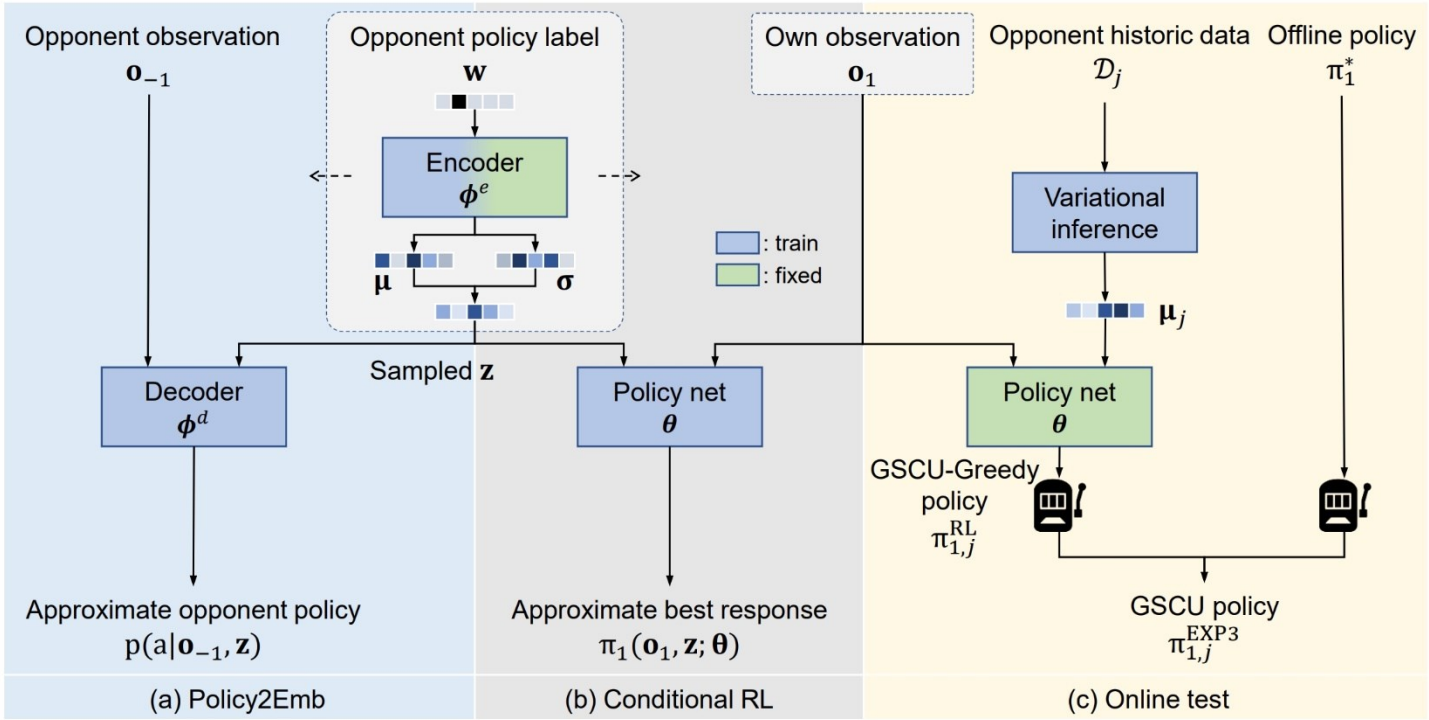
两人零和博弈学习：对手建模（工作三）

- **基于多样化策略生成的开放式对手建模算法**
  - 自动生成包含各种强度的多样化对手策略，形成数据飞轮



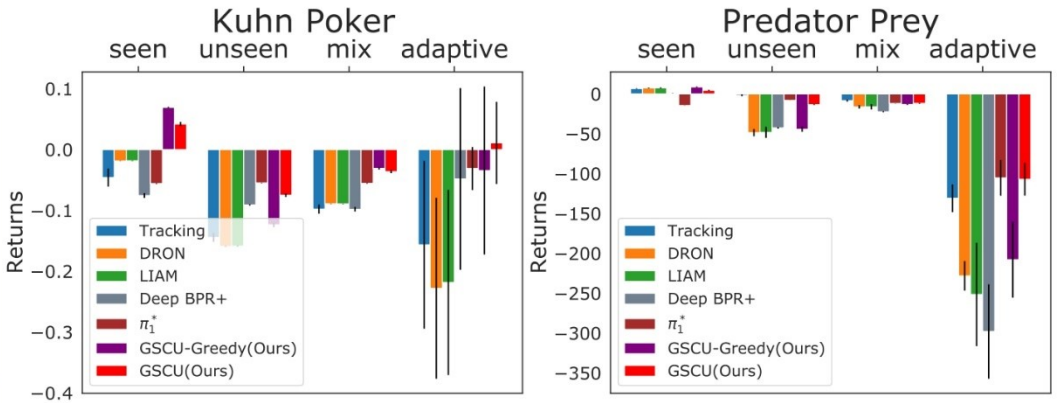
两人零和博弈学习：对手建模（工作四）

- **基于在线变分推断的安全对手建模算法**
  - 对手类型非常确定时，采取最优反应策略，Greedy when Sure
  - 对手不确定时，采取较为保守的均衡策略，Conservative when Uncertain
  - ICML 2022, Spotlight

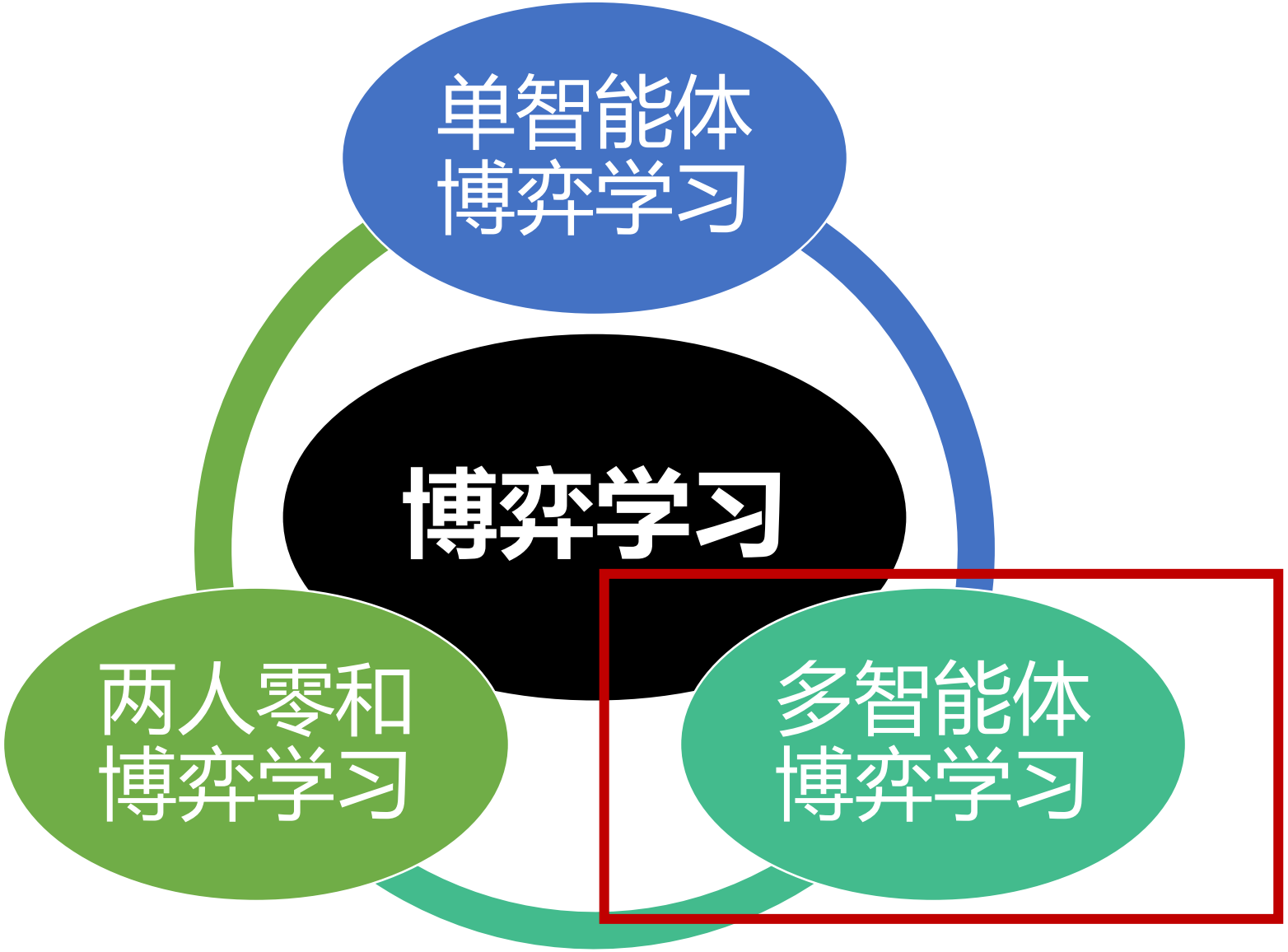


Theorem 3.1. *When* $\eta = \min\left\{1, \sqrt{\frac{2\ln 2}{(e-1)\Delta T}}\right\}$, *the regret of playing* $\pi_{1,j}^{EXP3}$ *for* T *episodes is upper bounded:*

$$R_T(\pi_{1,j}^{EXP3}) \leq 3.1\sqrt{\Delta T} + \min\left\{R_T(\pi_1^*), R_T(\pi_{1,j}^{RL})\right\}.$$



多智能体博弈学习相关工作介绍



多智能体博弈学习：学习目标与关键问题

学习目标

多个智能体协作完成指定任务

关键问题

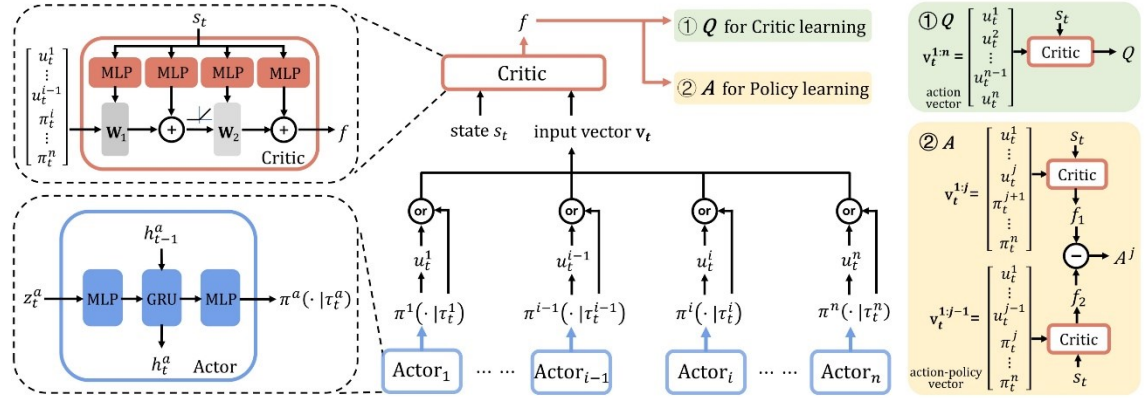
如何有效**信用分配**促进**分工合作**

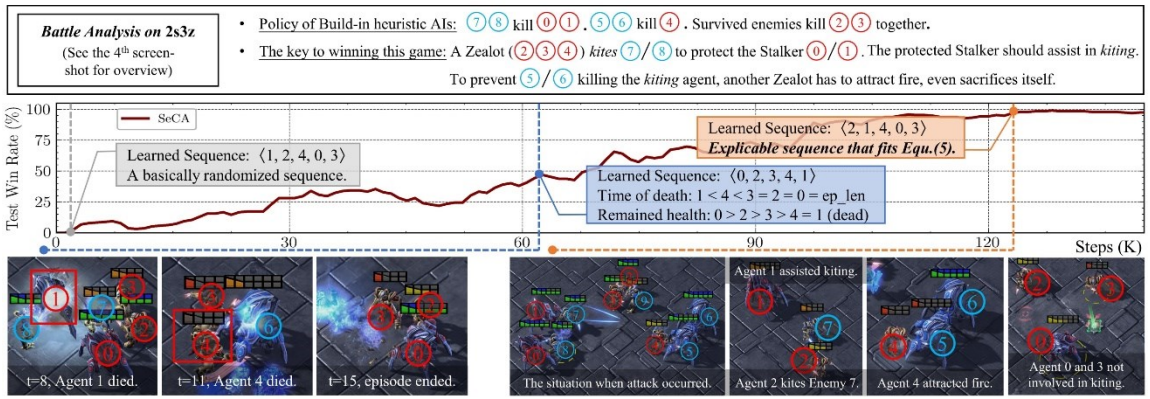
博弈/学习视角

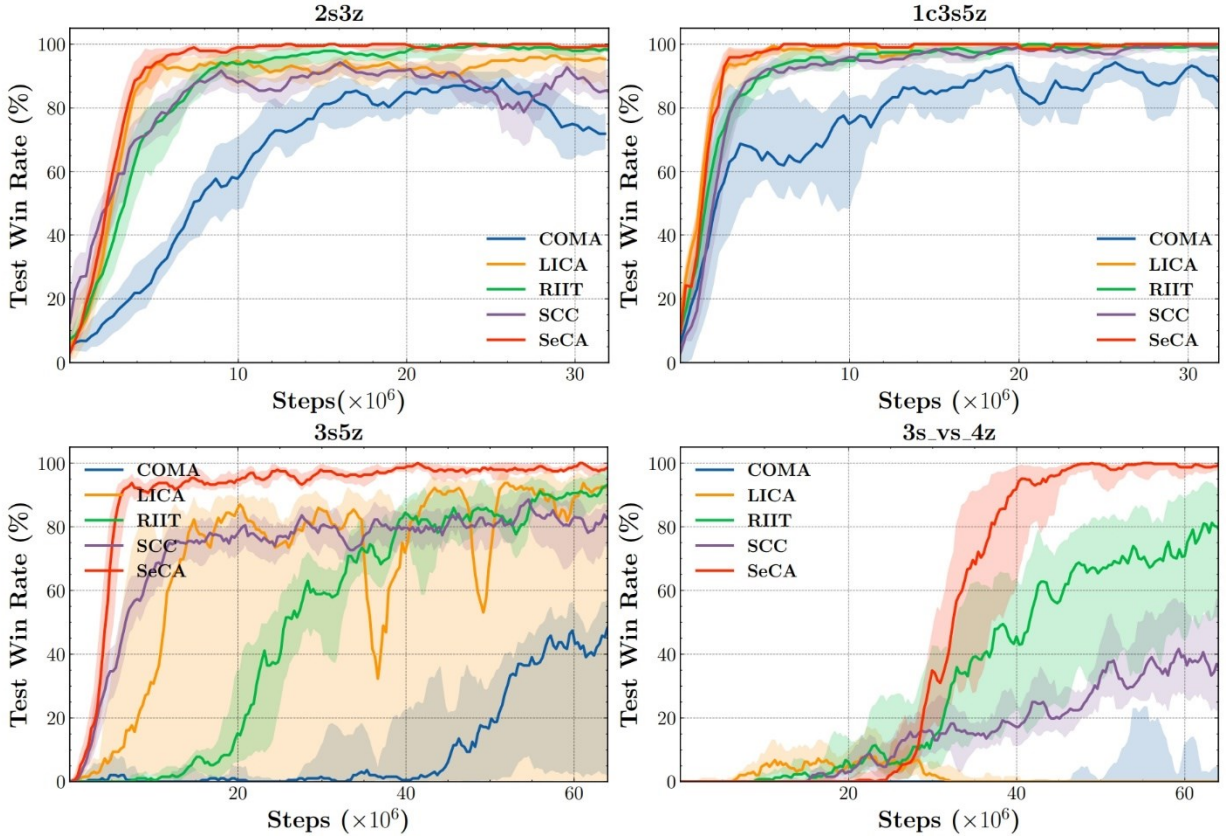
合作博弈/多智能体强化学习

多智能体博弈学习：多智能体强化学习（工作一）

- **基于序列化信用分配的多智能体高效学习算法**
  - 简化多智能体系统分析的复杂度，按照自动学习得到的顺序依次分析单个智能体的贡献，提高系统整体性能，AAMAS 2023

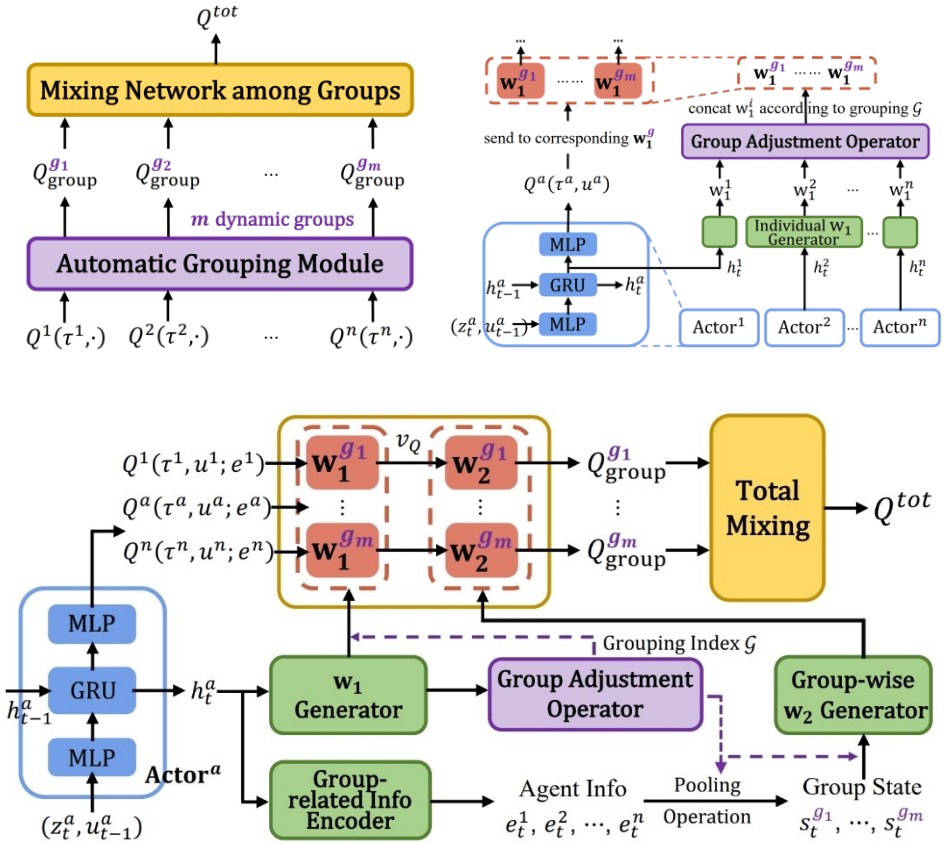


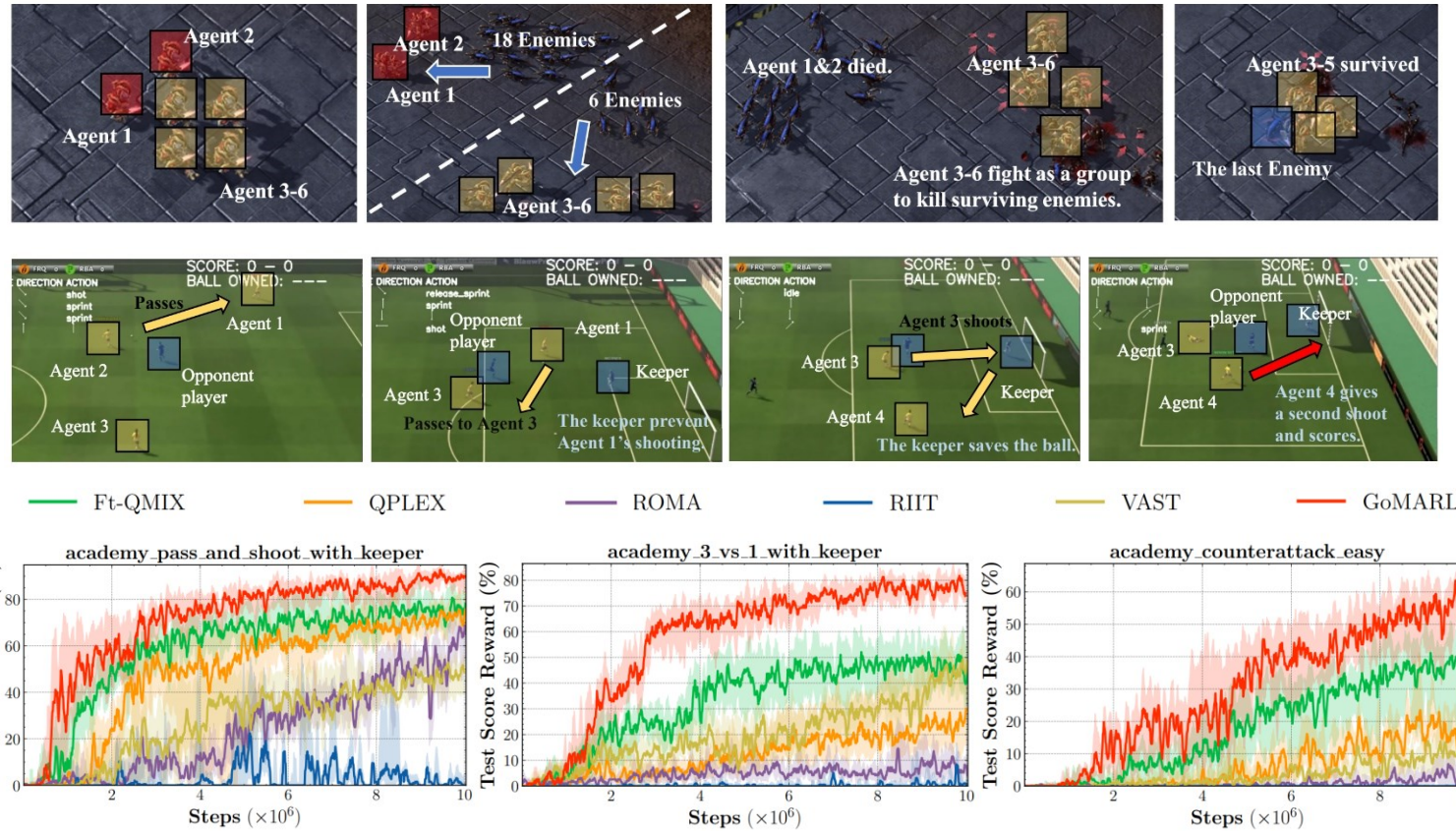




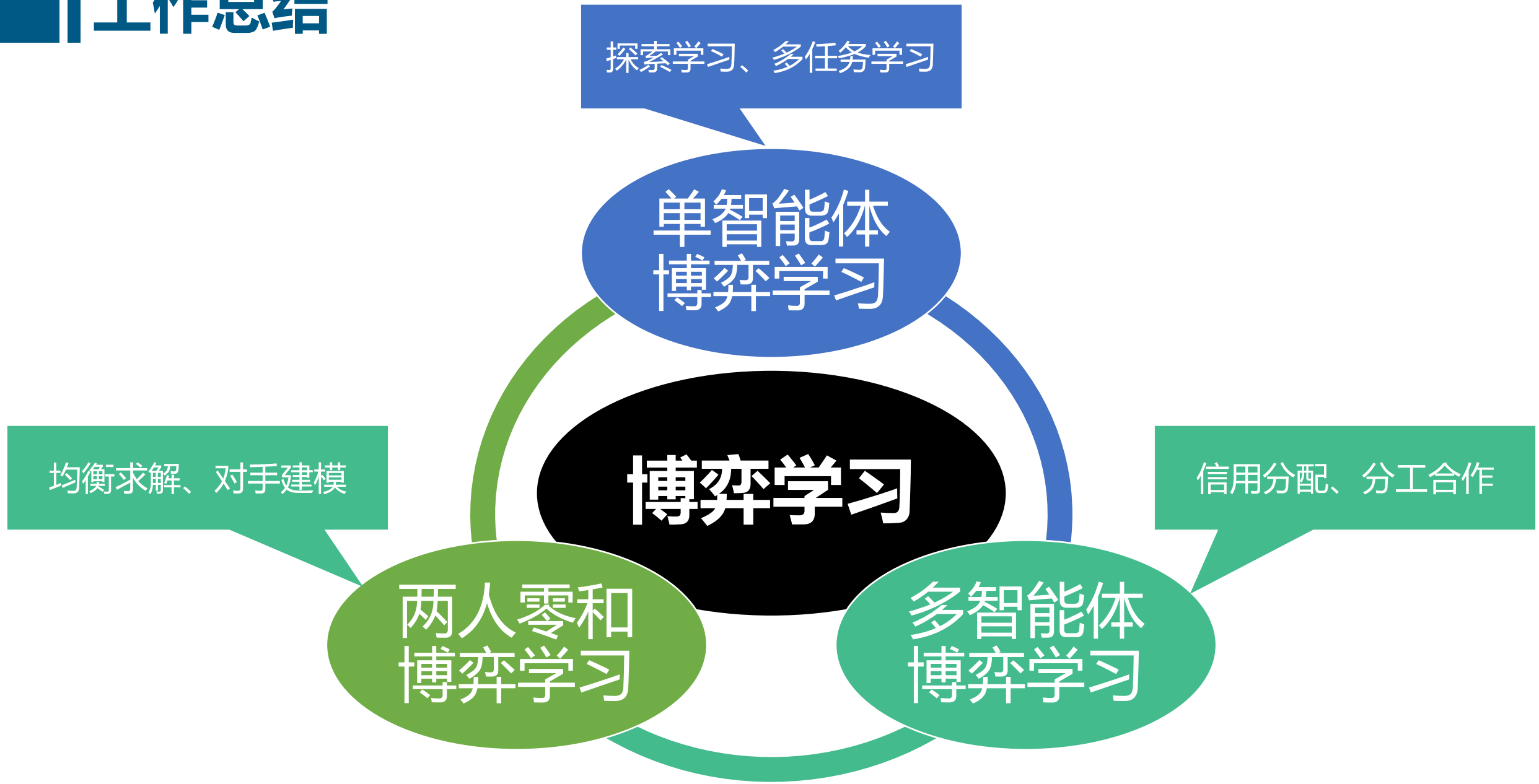
多智能体博弈学习：多智能体强化学习（工作二）

- **基于自动分组学习的多智能体高效学习算法**
  - 可实现多智能体按照角色自动分组，组别个数以及组内成员均可动态变化，以此促进良好配合实现高效合作，NeurIPS 2023





工作总结


探索学习、多任务学习
单智能体
博弈学习
博弈学习
均衡求解、对手建模
信用分配、分工合作
两人零和
博弈学习
多智能体
博弈学习

后续研究计划

考虑人的因素

- 如何建模人
- 人机协作配合

与大模型结合

- 大模型如何帮助强化学习
- 强化学习如何帮助大模型

更多落地应用

- 具身智能
- 自主无人系统