

Learning Effective Value Function Factorization via Attentional Communication

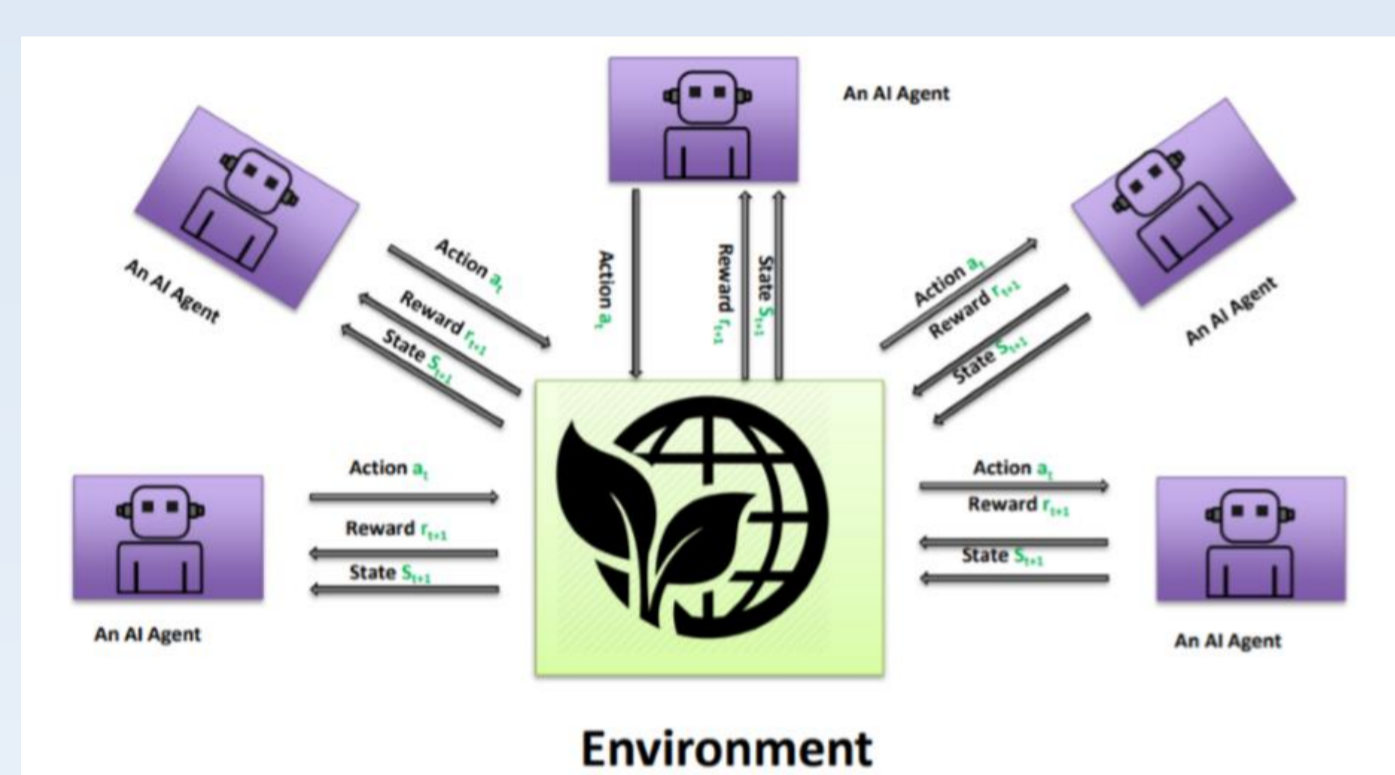
基于注意力通信学习有效的值函数分解

吴波* 杨晓亚 孙楚雄 王瑞 胡晓惠

*通讯方式: 15927642981 | wubo2018@iscas.ac.cn

背景

多智能体强化学习方法(Multi-Agent Reinforcement Learning, MARL)在人工智能和自动控制等领域得到了广泛的关注和研究。在MARL中, 多个智能体与环境进行交互来尽可能获得更多的累计奖励, 通常可建模成去中心化部分可观测马尔科夫决策过程(Dec-POMDP)。

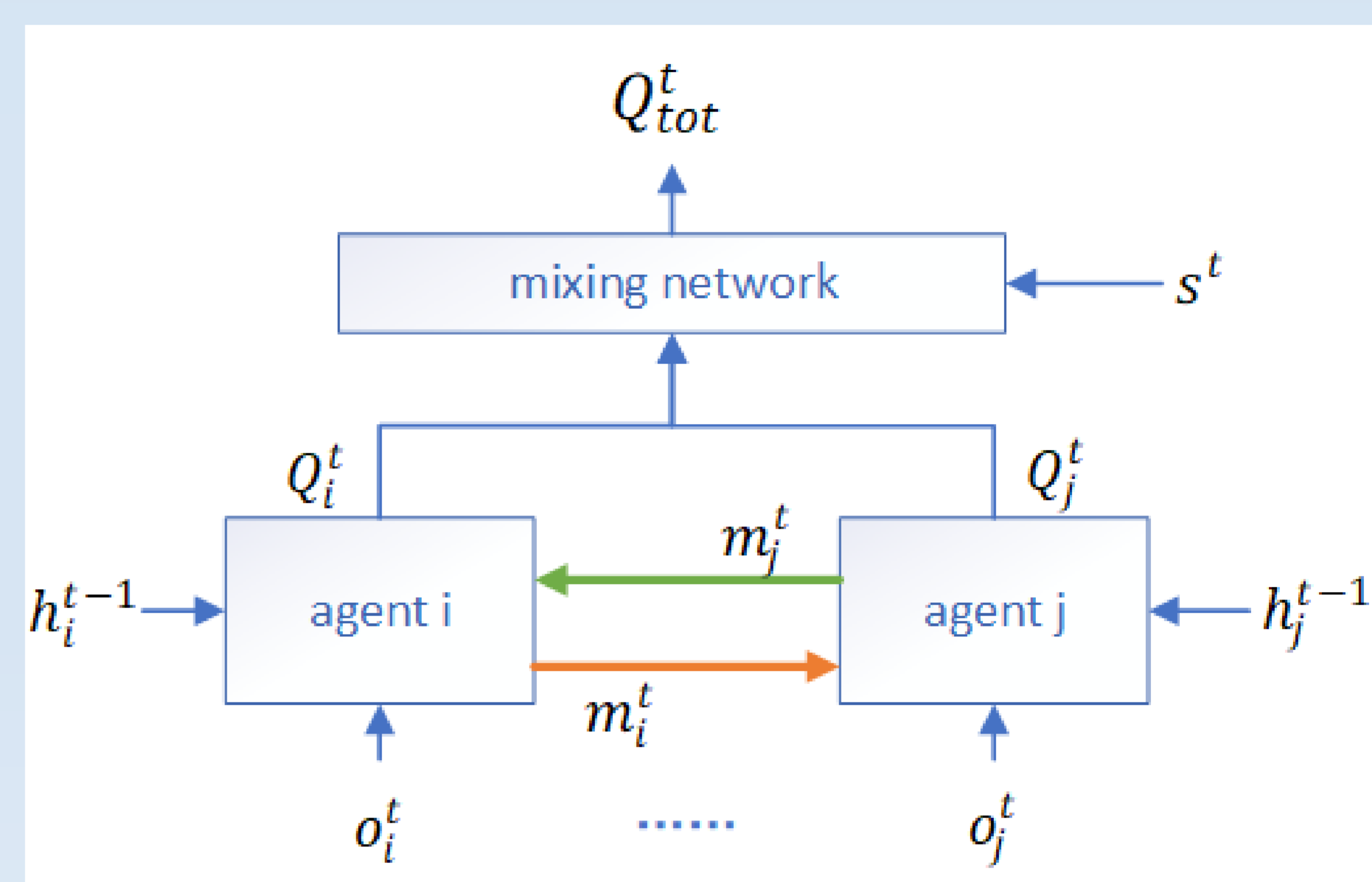


- 连续决策过程
- 无监督, 只有奖励信号反馈
- 反馈通常是延迟的
- 智能体行为互相影响

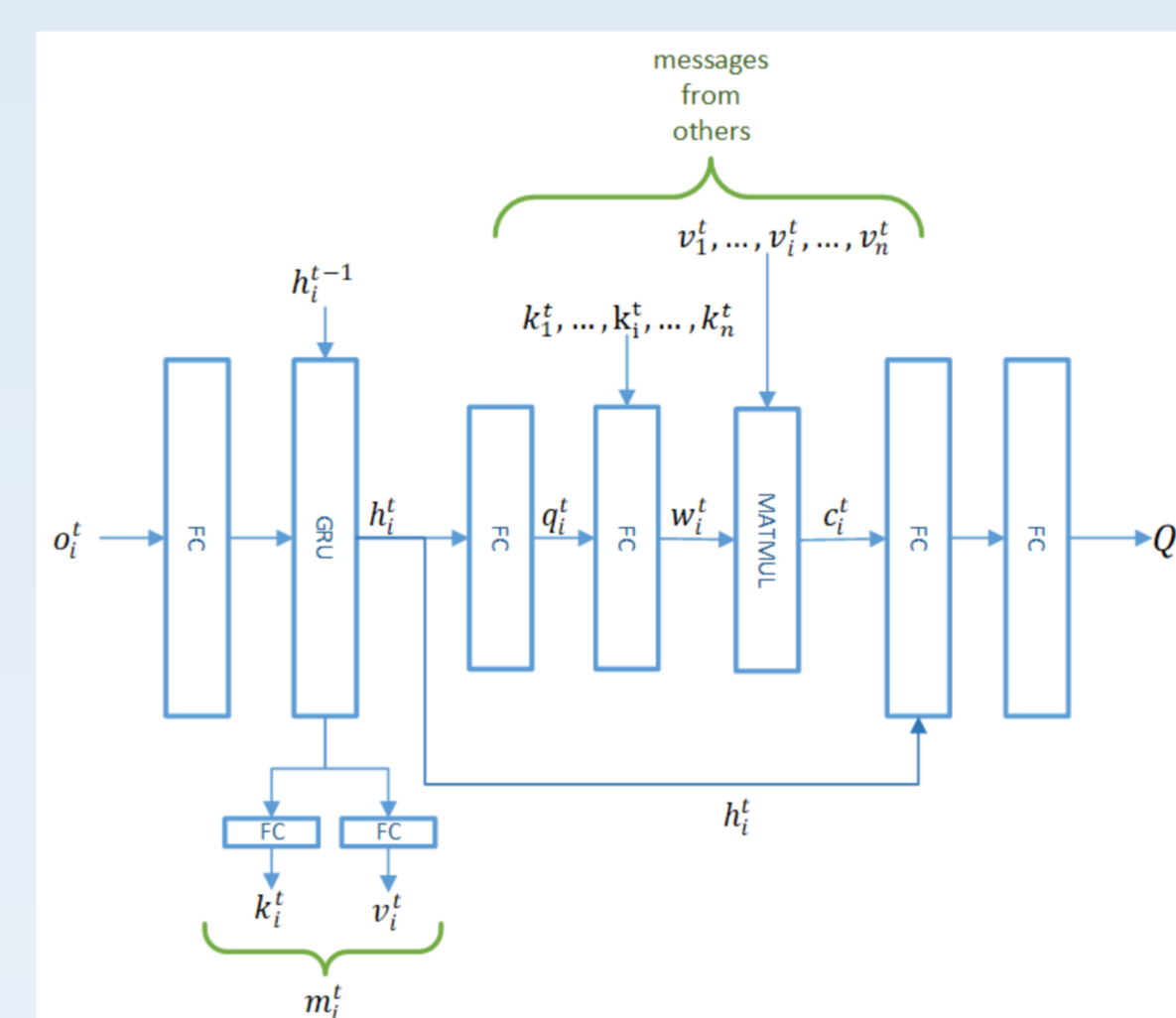
创新点

① 通信辅助值函数分解学习

引入通信来降低混合网络的“负担”, 学习更通用的值函数分解形式, 进而表达更复杂的协作模式。



② 基于软注意力的信息整合



- [key, value]形式的消息拆分设计
- 依据相关性高效整合信息

星际争霸小场景

StarCraft II micromanagement tasks

对战难点:

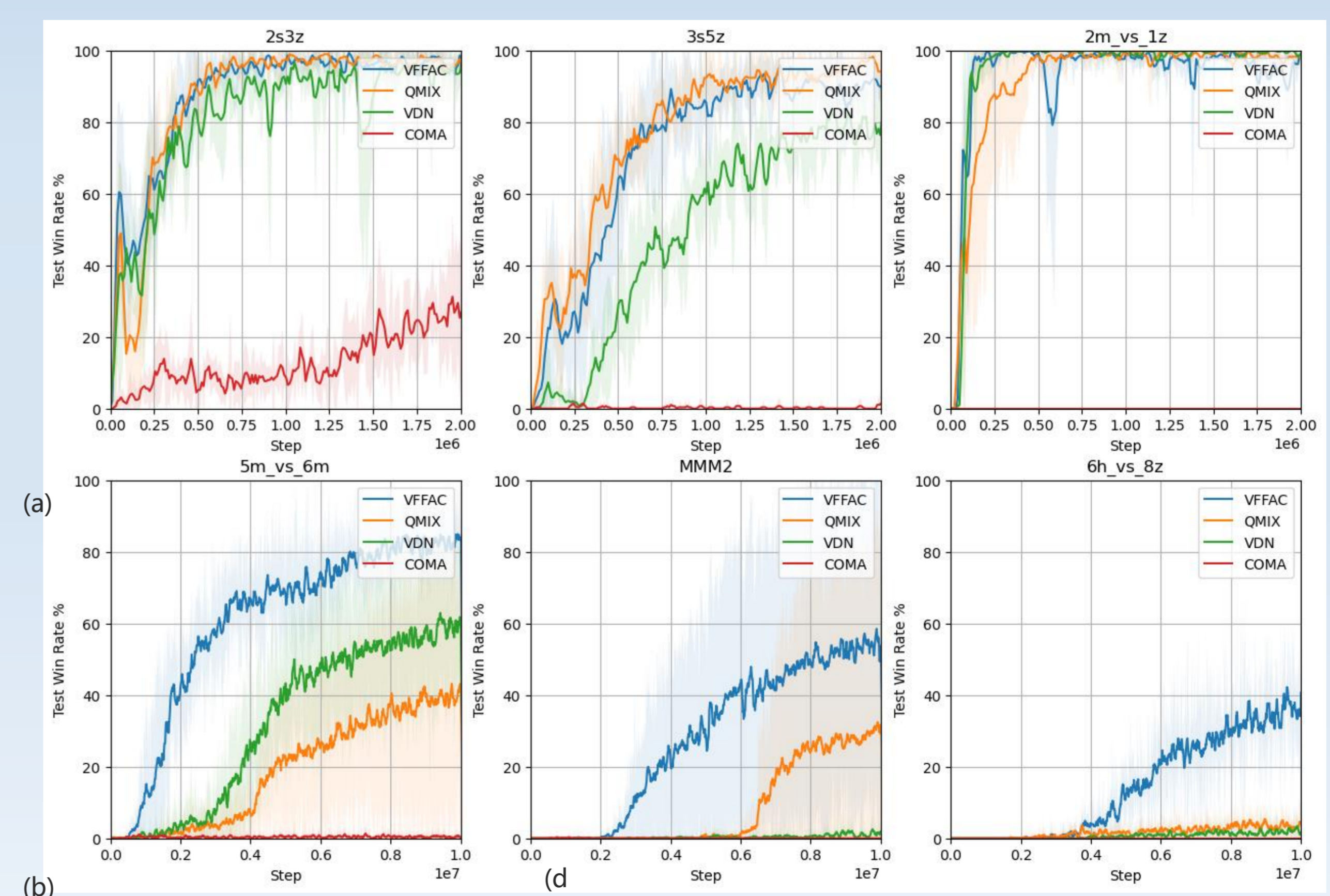
- 视野受限
- 兵力不对等

微操技能:

- 集火
- 风筝(走a)



实验结果



总结与展望

- 通信行为本身就表达了某种意义上的“协作”。
- 引入注意力通信后, 值函数分解算法获得了显著的性能提升。
- 研究展望: 有限通信带宽下的算法优化, 通信组群的自动学习。