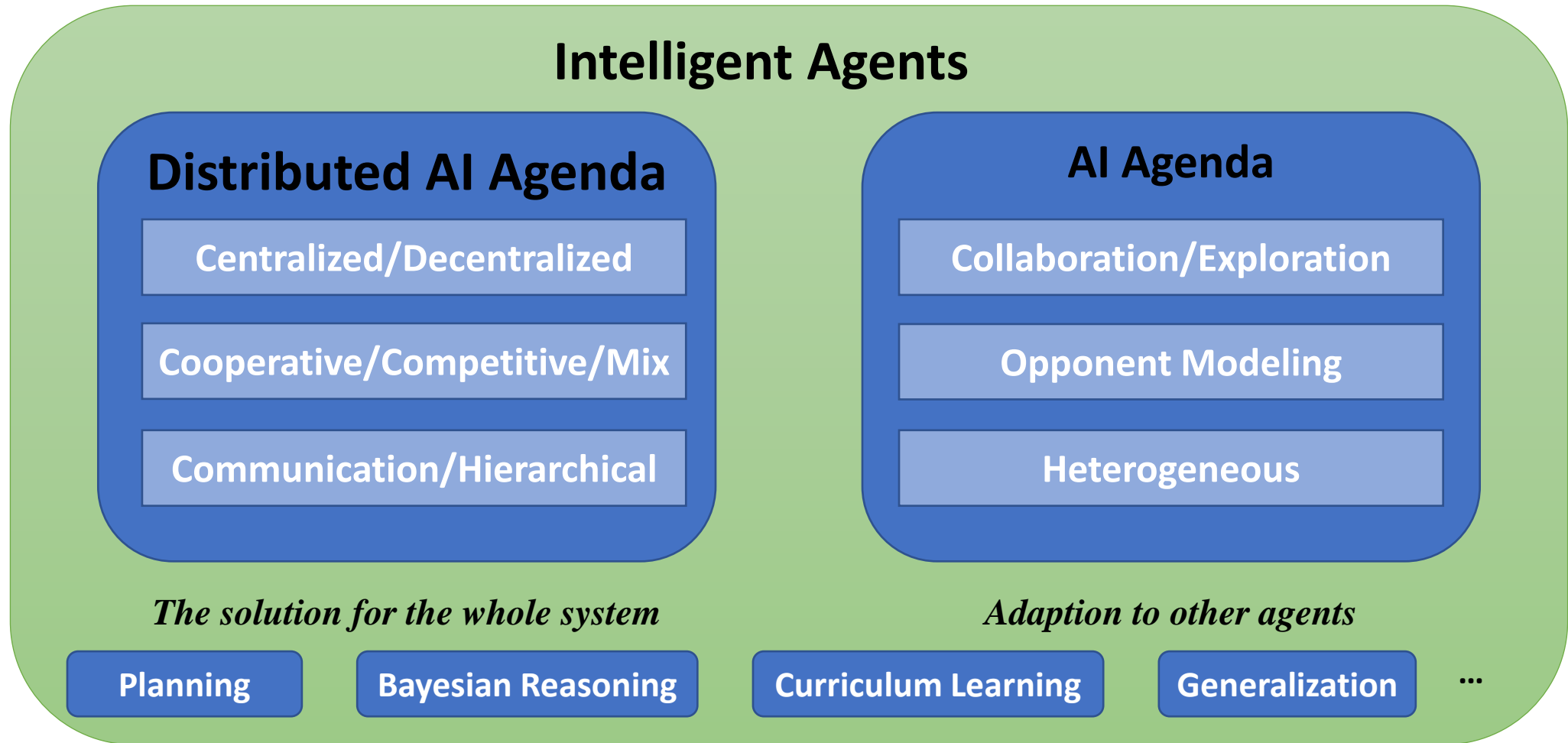


MARL研究现状

Linghui Meng

Taxonomy



Distributed AI Agenda

	Fully Centralized	Fully Decentralized	Uncertainty	Large Scale Agent
优点	Stationarity	Scalability	Completeness	Complex Tasks
缺点	Non-scalability	Instability and Bias	Convergence	Exponential Action Space
Solutions	Centralized Training Decentralized Execution		Communication	Hierarchical RL
Framework	VDN, QMIX, COMA, QTRAN, QPLEX		RIAL, CommNet, IC3Net, NDQ	ROMA, RODE, GR2

AI Agenda

	Independent-Training	Opponent Modelling	Large Scale Agent
优点	Scalability	Stationary	Complex Tasks
缺点	Ignore other Agents	Complexity	Exponential Action Space
Solutions	Collaboration	Constraints	Heterogeneous Agent
Framework	Lenient, FMQ	DPIQN, LOLA	ROMA, RODE

New Possible Methods

Planning

在复杂任务下需要有人介入，构建AI Planner实现层次化RL

Curriculum Learning

学习task的生成和顺序，以达到完成复杂任务的目的，用到迁移学习

Bayesian Reasoning

针对Agent的决策可解释性，可引入概率方法做推理

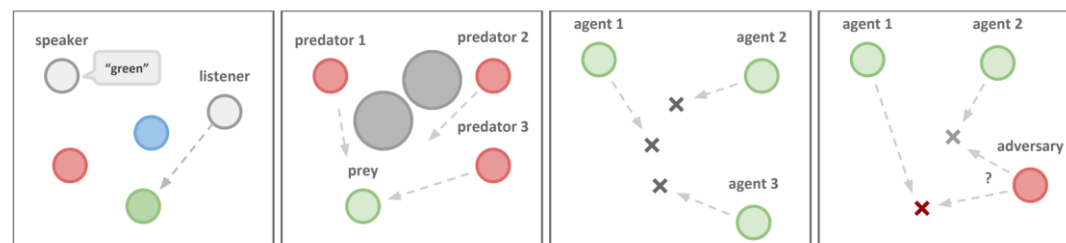
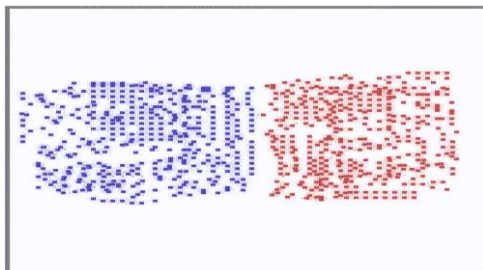
Generalization

offline训练时对online的数据泛化能力不足，可考虑用变分方法改善

Open Resource Env & Implementation

Environment

- [SMAC](#) - StarCraft Multi-Agent Challenge
- [Flatland](#) 简易的交通场景
- [MAgent](#) 内置battle、gather、pursuit等场景
- [Particle world](#) predator-prey, Cooperative navigation等场景



Implementation

1. [IMPALA的Pytorch实现](<https://github.com/Sheepsody/Batched-Impala-PyTorch>)
2. DRL框架GARAGE, RLlib, Catalyst, rlpyt, pymarl, rlkit

Benchmark on MARL

- The Apprentice Firemen Game
- Pommerman
- StarCraft Multiagent Challenge
- Particle world (predator prey, cooperative navigation)
- MuJoCo Multiagent Soccer
- Neural MMO

State of the art

RODE: Winrate outperforms 10 of 14 scenarios on SC sota.

Challenge & Future Direction

1. **Scalability**: 可扩展性CTDE (Centralized Training and Decentralized Execution)
2. **Credit Assignment**: each agent's contribution to the team
3. **Uncertainty** (non-stationary) : partial and noisy observation通过communication解决来自环境的不确定性, 多智能体会相互影响
4. **Heterogeneity**: 异构性 requiring diverse behaviors of agents /role-based
5. **Hierarchical**: 层次化Agent, 多级agent面对模型
6. **Coordination**: 协调
7. **Generality**: RL的泛化性, off-policy和在推理时数据不同

Thanks