2050 追逐早上七八点钟的太阳 云栖

MediaGamma UCL

YunQi 2050 - DRL Session

Communication in Multi-agent Reinforcement Learning

Ying Wen
Department of Computer Science, University College London
MediaGamma Ltd.
ying.wen@cs.ucl.ac.uk

30 May, 2018

Multi-agent in Real-World

Human Teams

Transportation Networks

Economies Markets

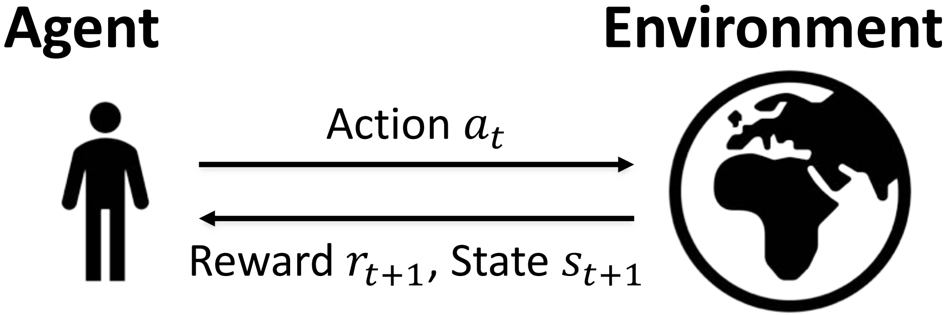
Games

Communication Networks

Agenda

- Generalizing Reinforcement Learning
  - Single Agent Reinforcement Learning
  - Multi-agent Reinforcement Learning (MARL)
- Challenges in MARL
  - Nonstationary Environment
  - Model Free Learning
  - Increasing Agent Number even Millions
- Communication and Learning
- Implicit Communication
- Dynamic Interaction

Reinforcement Learning



Agent

Environment

Action a_t

Reward r_{t+1}, State s_{t+1}

Optimal Policy $a = \pi^*(s)$ ← Maximise Long Term Reward $\sum r_t$

Multi-Agent System

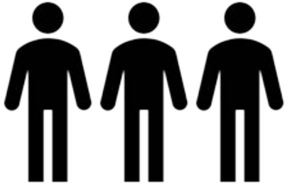
- **Multiagent system** is a collection of multiple autonomous (intelligent) **agents**, each acting towards its **objectives** while all **interacting** in a **shared environment**, being able to **communicate** and possibly **coordinating** their actions.

Types of Agent Systems

Single-Agent

Multi-Agent

Cooperative

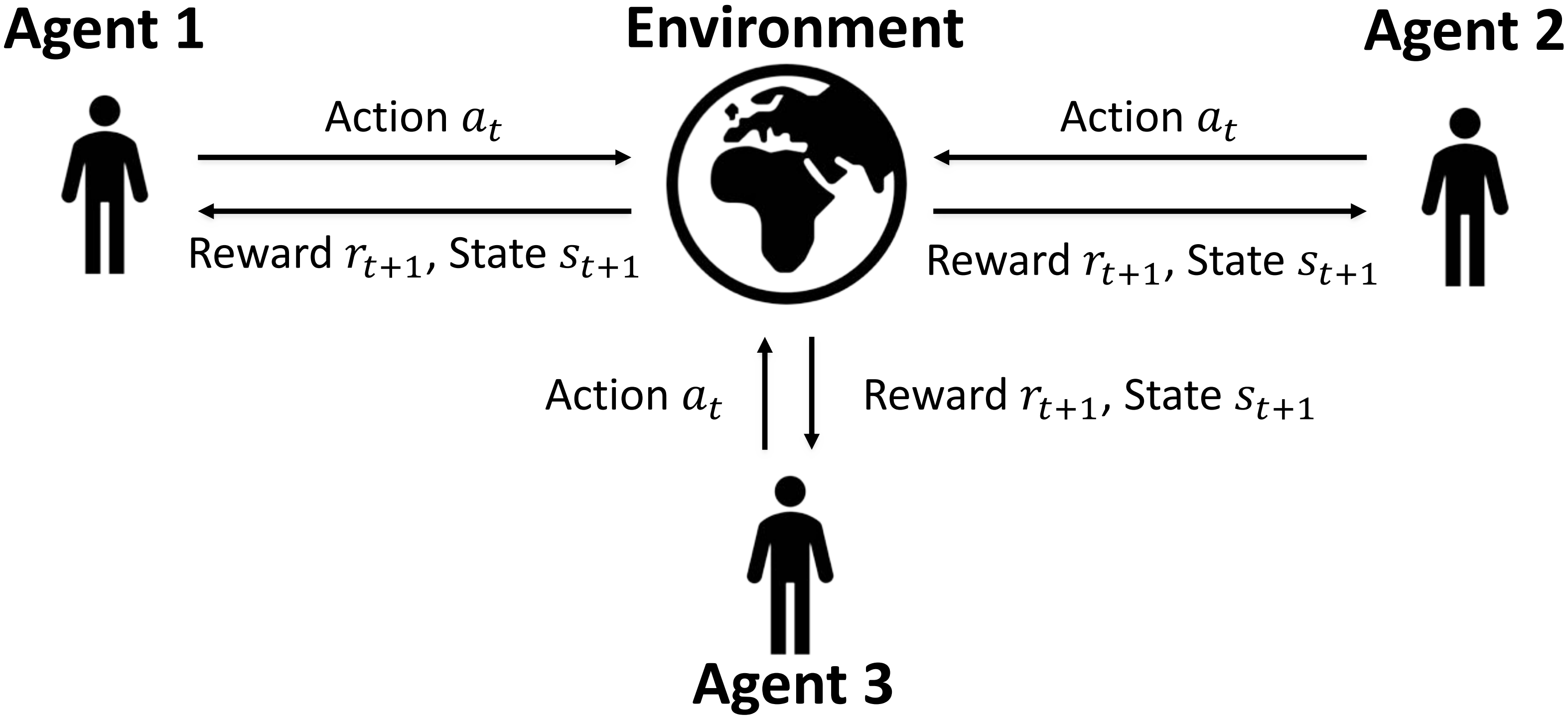


single
shared utility

Competitive

multiple
different utilities

Multi-agent Reinforcement Learning



Agent 1

Environment

Agent 2

Action a_t

Reward r_{t+1}, State s_{t+1}

Action a_t

Reward r_{t+1}, State s_{t+1}

Action a_t

Reward r_{t+1}, State s_{t+1}

Agent 3

Challenges in MARL

1. Non-stationary Environment
   - Needs for communication

2. Model Free - Agent Awareness
   - Intent / Opponent Modelling

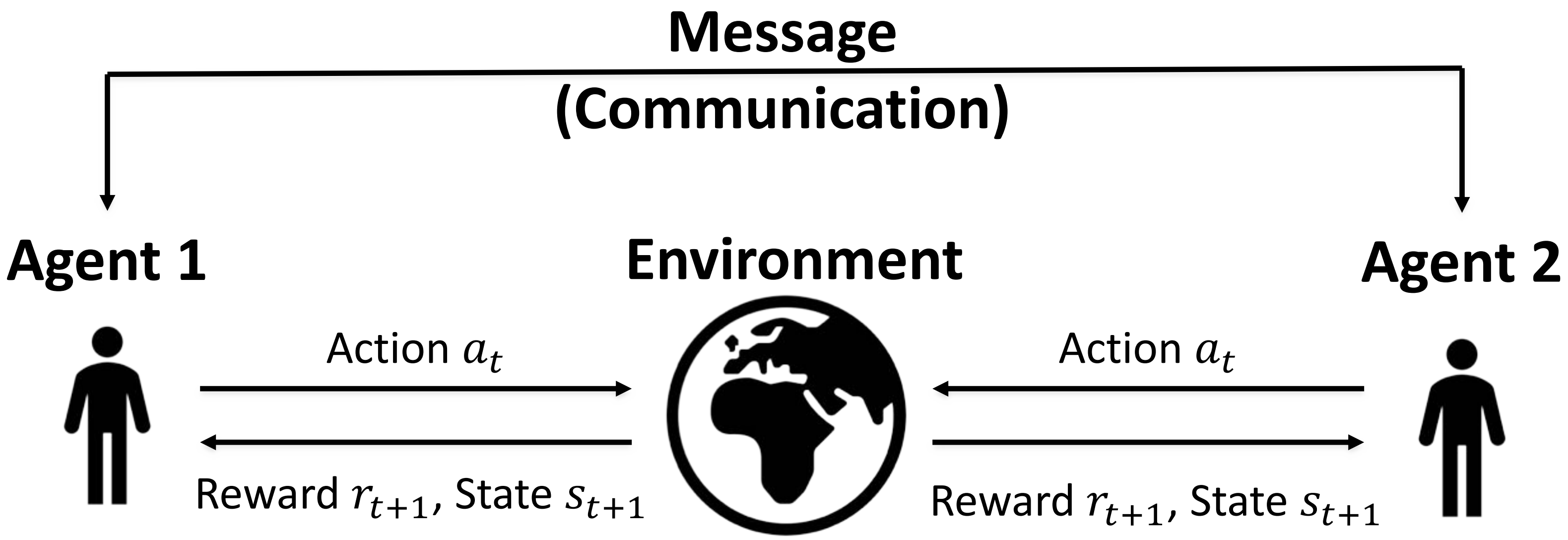
3. Increasing Number of Agents
   - Approximation of other agents
   - Dynamics of agents

Multi-Agent Perspective

1. **Micro Perspective**, The agent design problem:
   - How should agents act to carry out their tasks? Optimal Policy.

2. **Macro Perspective**, The society design problem:
   - How should agents interact to carry out their tasks? Dynamic Interaction.

MARL with Communication



Message (Communication)

Agent 1

Environment

Agent 2

Action a_t

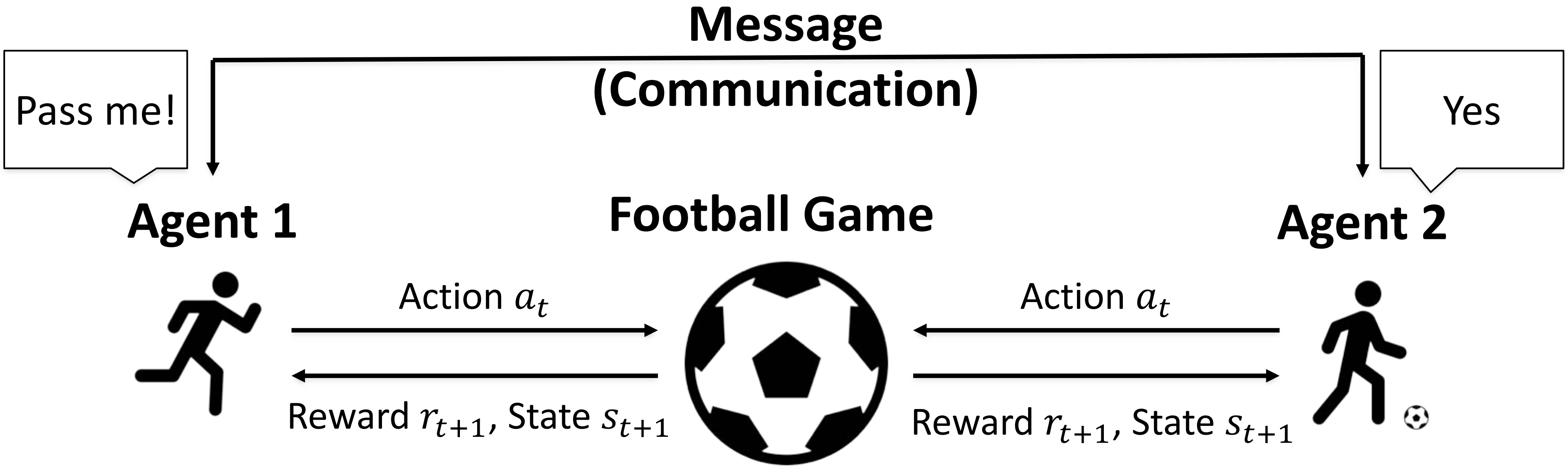
Reward r_{t+1}, State s_{t+1}

Action a_t

Reward r_{t+1}, State s_{t+1}

How to cooperate? -> with Communication

MARL with Communication - Example



Message (Communication)

Pass me!

Yes

Agent 1

Football Game

Agent 2

Action a_t

Reward r_{t+1}, State s_{t+1}

Action a_t

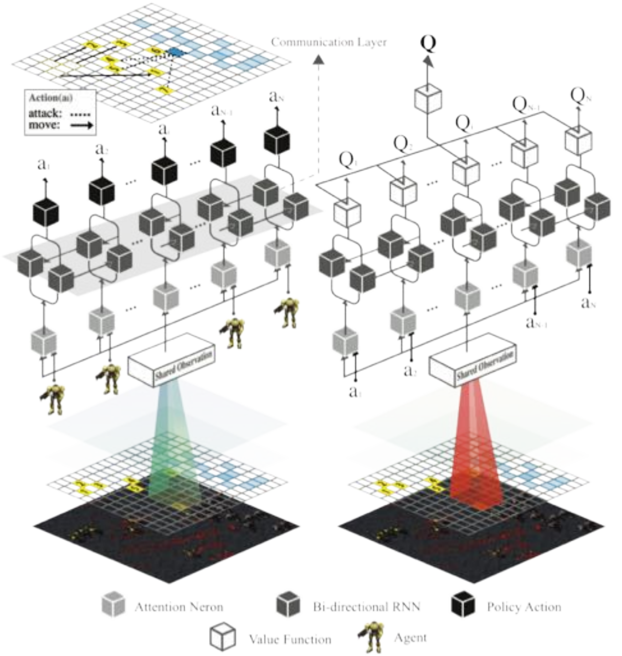
Reward r_{t+1}, State s_{t+1}

How to cooperate? -> with Communication

Bi-directionally Coordinated Network

- Bi-directional recurrent networks
  - Means of communication
  - Connect each individual agent’s policy and and Q networks

- Multi-agent deterministic actor-critic



(a) Multiagent policy networks (b) Multiagent Q networks

How It Works

- High Q-value steps are aggregated in the same area.

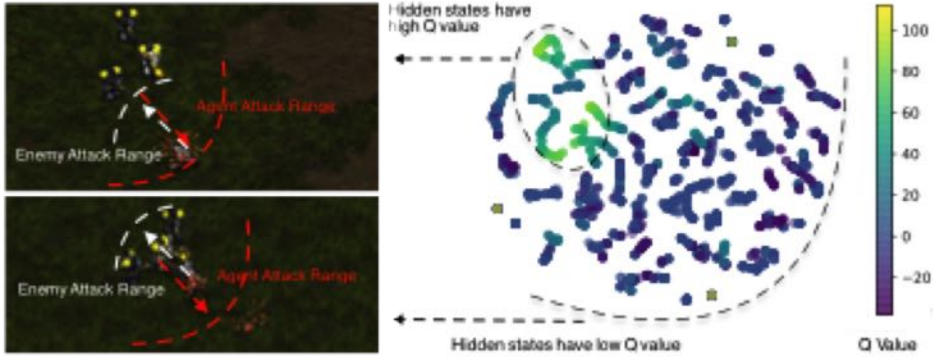


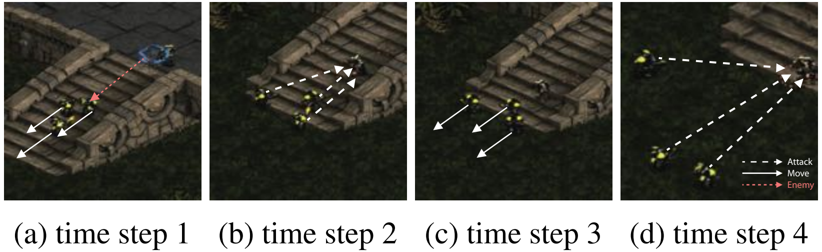
Figure 4: Visualisation for 3 Marines vs. 1 Super Zergling combat. **Upper Left**: State with high Q value; **Lower Left**: State with low Q value; **Right**: Visualisation of hidden layer outputs for each step using TSNE, coloured by Q values.

Emerged Human-level Coordination

- Hit and Run tactics

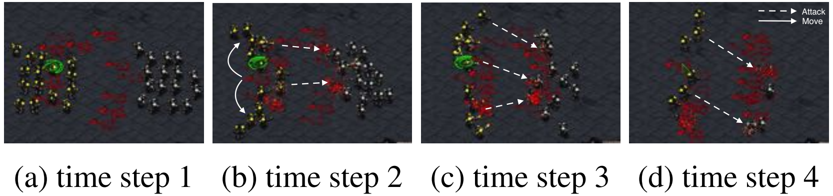
- Focus fire without overkill

- ……



(a) time step 1 (b) time step 2 (c) time step 3 (d) time step 4

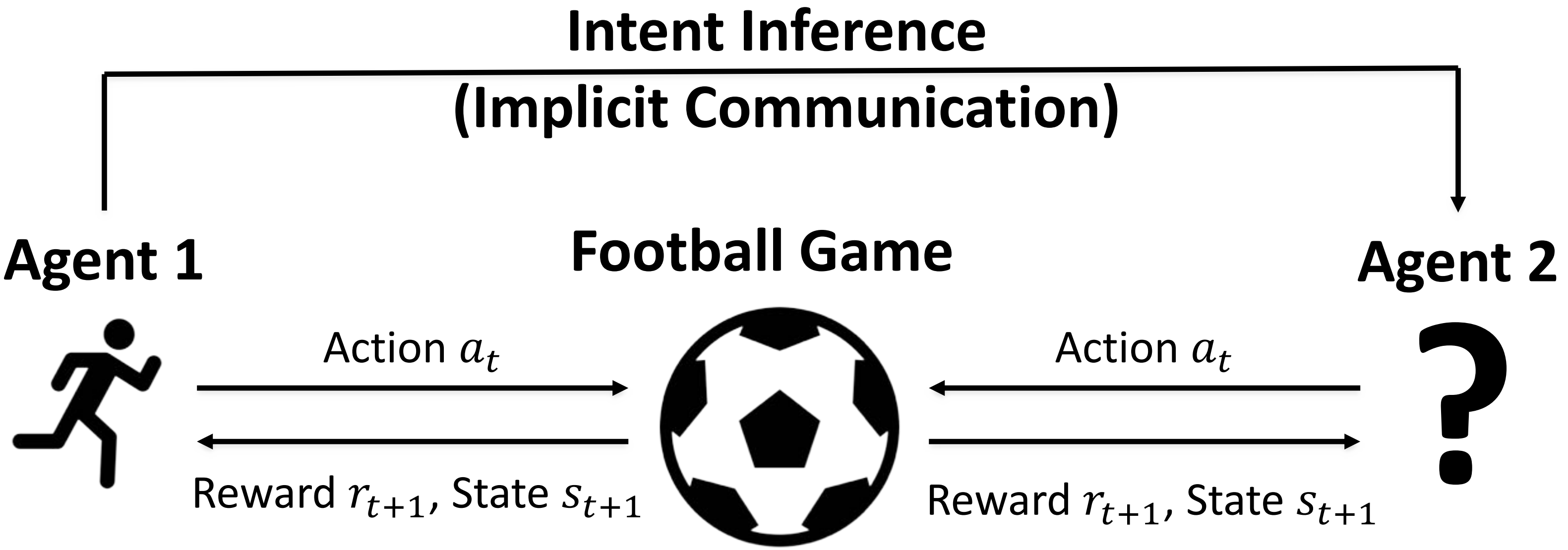
Figure 7: *Hit and Run* tactics in combat *3 Marines (ours) vs. 1 Zealot (enemy).*



(a) time step 1 (b) time step 2 (c) time step 3 (d) time step 4

Figure 9: ”focus fire” in combat *15 Marines (ours) vs. 16 Marines (enemy).*

MARL with Implicit Communication



Intent Inference

(Implicit Communication)

Agent 1

Football Game

Agent 2

Action a_t

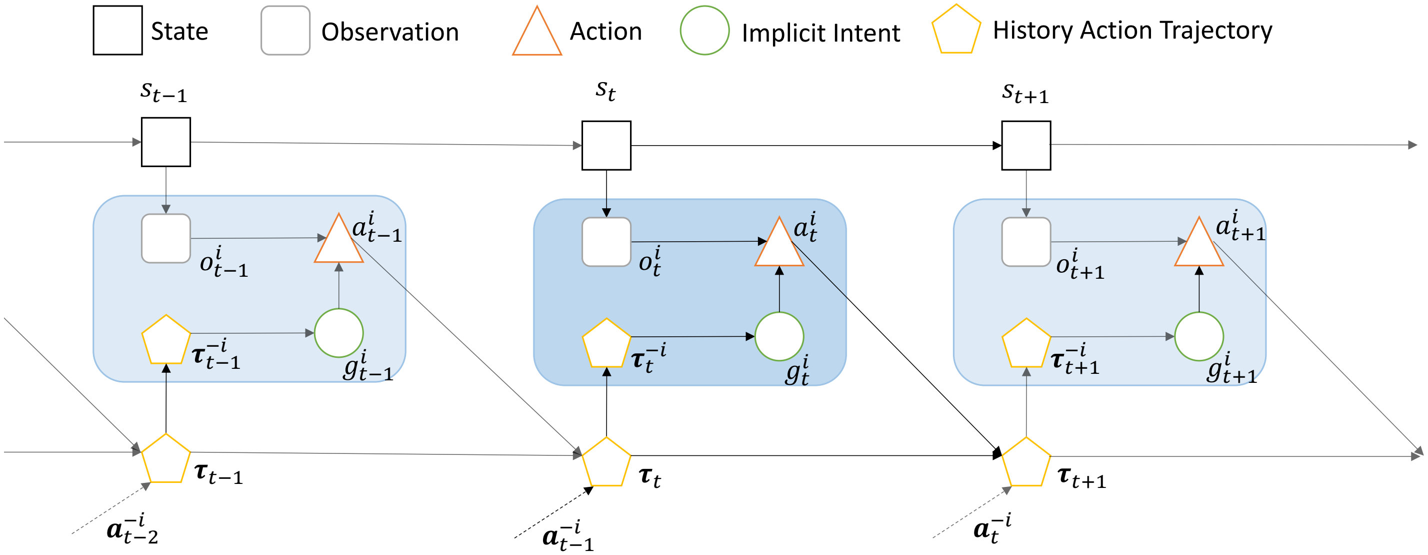
Reward r_{t+1}, State s_{t+1}

Action a_t

Reward r_{t+1}, State s_{t+1}

How to know learn with unknown agents? -> Agent Awareness

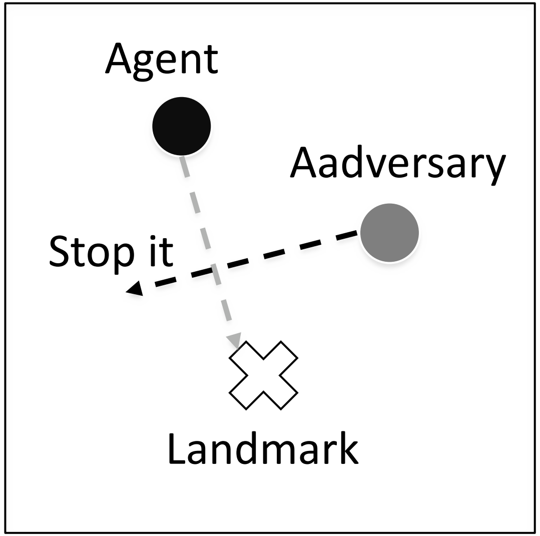
Implicit Intent Inference in MARL



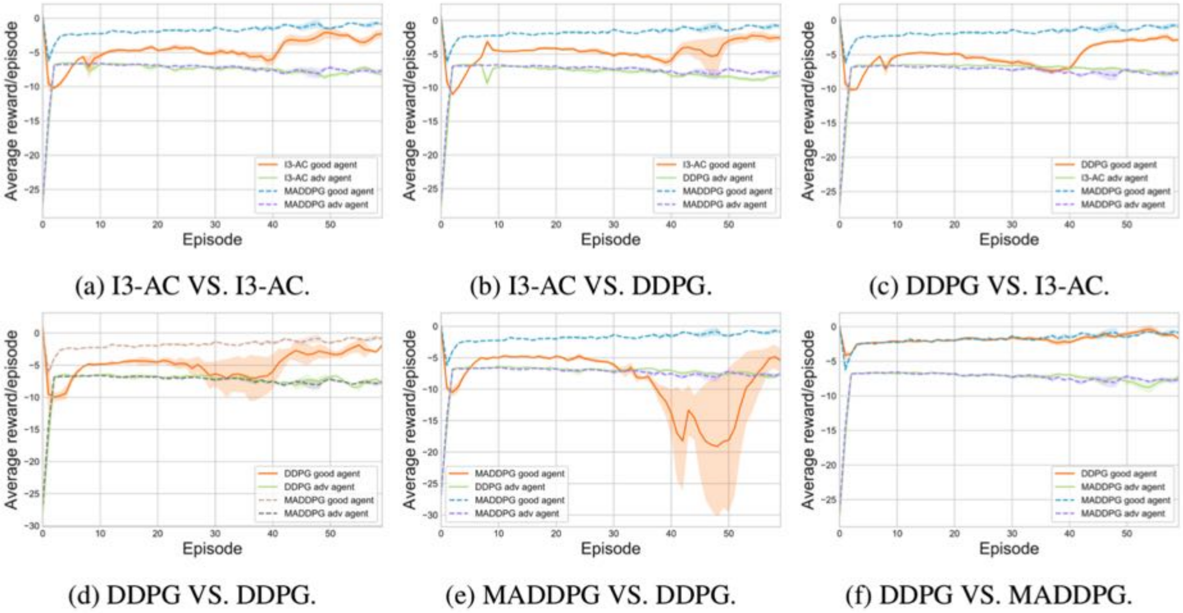
□ State
□ Observation
△ Action
○ Implicit Intent
⬠ History Action Trajectory
s_{t-1}
s_t
s_{t+1}
o^i_{t-1}
a^i_{t-1}
$\boldsymbol{\tau}^{-i}_{t-1}$
g^i_{t-1}
o^i_t
a^i_t
$\boldsymbol{\tau}^{-i}_t$
g^i_t
o^i_{t+1}
a^i_{t+1}
$\boldsymbol{\tau}^{-i}_{t+1}$
g^i_{t+1}
$\boldsymbol{\tau}_{t-1}$
$\boldsymbol{\tau}_t$
$\boldsymbol{\tau}_{t+1}$
$\boldsymbol{a}^{-i}_{t-2}$
$\boldsymbol{a}^{-i}_{t-1}$
$\boldsymbol{a}^{-i}_t$

Implicit Intent Inference Network to Learn the Intent Embedding

Implicit Intent Inference in MARL



Keep Away Game



(a) I3-AC VS. I3-AC.

(b) I3-AC VS. DDPG.

(c) DDPG VS. I3-AC.

(d) DDPG VS. DDPG.

(e) MADDPG VS. DDPG.

(f) DDPG VS. MADDPG.

Mean Field MARL

- When the number of agents becomes thousands even millions

- Mean action approximation

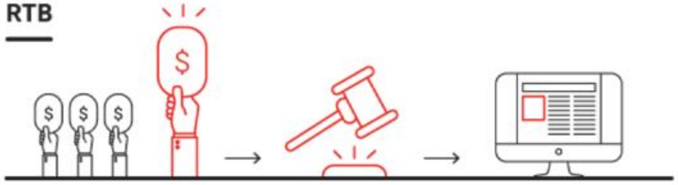
Agent 1
Agent 2
……
Agent N

Mean Field MARL – Real-time Bidding

- **Mean Field Equilibrium**
  learning in real-time bidding

- **High Volume** and **High Liquid**

- **Second Price** Auction only
  pay the second highest price



Multi-Agent Perspective

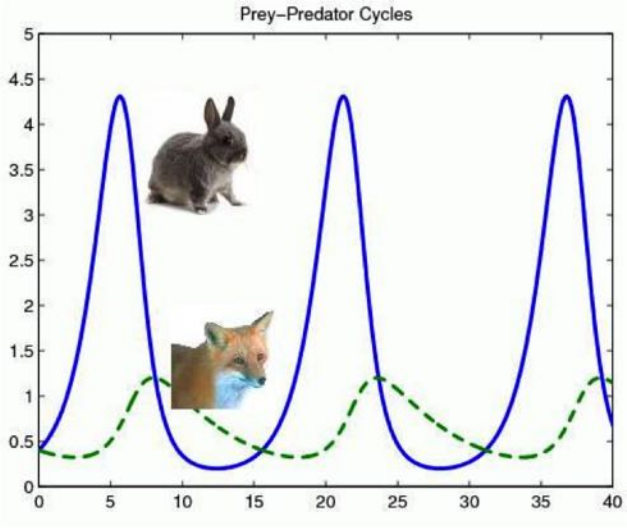
1. **Micro Perspective**, The agent design problem:
   - How should agents act to carry out their tasks? Optimal Policy.

2. **Macro Perspective**, The society design problem:
   - How should agents interact to carry out their tasks? Dynamic Interaction.

Population Dynamics in Million-agent RL

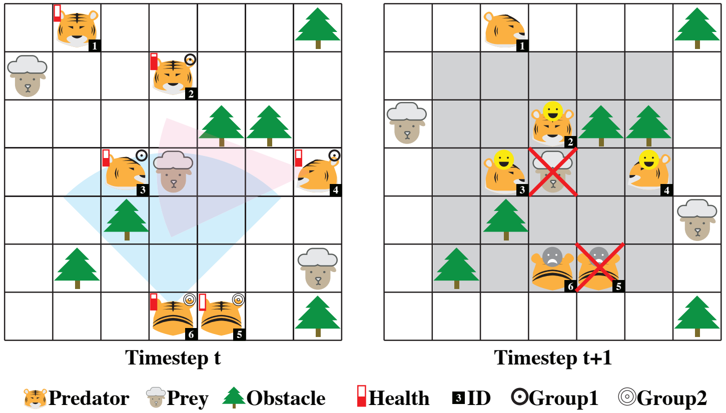
- A major topic of population dynamics is the cycling of predator and prey populations

- The **Lotka-Volterra** model is used to model this.



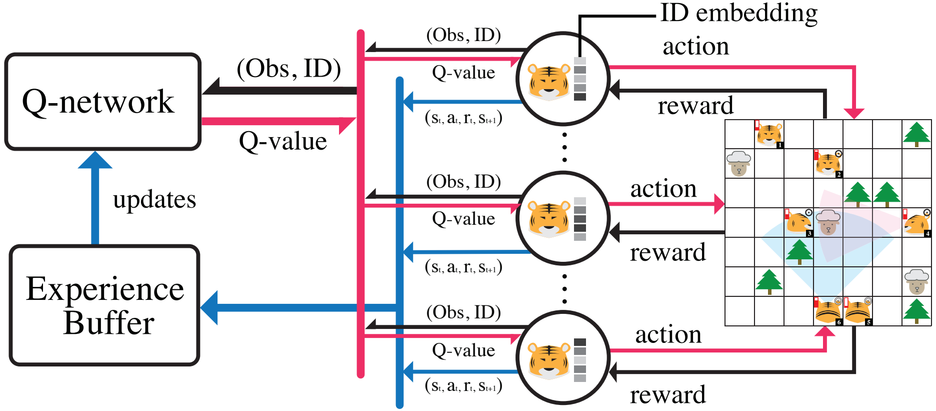
Population Dynamics in Million-agent RL

- **Predators** hunt the **prey** so as to survive from starvation
- Each **predator** has its own health bar and eyesight view
- **Predators** can form a group to hunt, and are scaled to 1 million

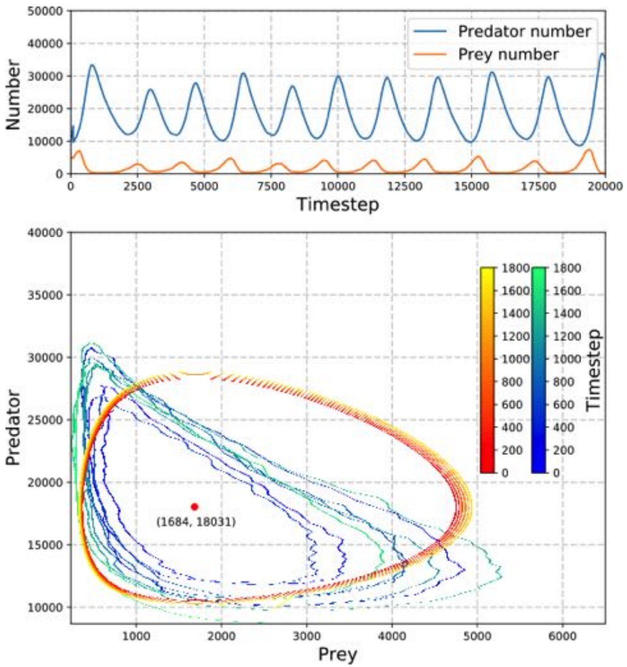


Population Dynamics in Million-agent RL

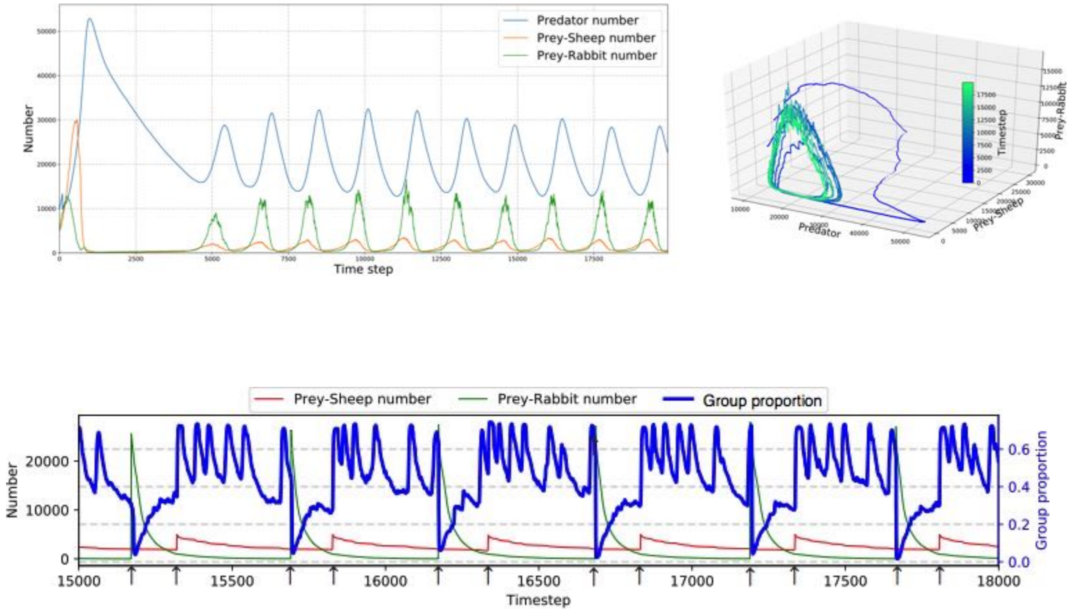
- The action space: {move forward, backward, left, right, rotate left, rotate right, stand still, join a group, and leave a group}.



Population Dynamics in Million-agent RL



The Dynamics of the Artificial Population



Tiger-sheep-rabbit: Grouping

Reference

[1] Peng, Peng*, Ying Wen*, Yaodong Yang, Quan Yuan, Zhenkun Tang, Haitao Long, and Jun Wang. "Multiagent Bidirectionally-Coordinated nets for learning to play StarCraft combat games.”

[2] Wen, Ying, Hui Chen and Jun Wang. " Implicit Intent Inference with Action Trajectories in Multi-agent Reinforcement Learning."

[3] Yang, Yaodong, Rui Luo, Minne Li, Ming Zhou, Weinan Zhang, and Jun Wang. "Mean Field Multi-Agent Reinforcement Learning."

[4] Wen, Ying and Jun Wang. “A Mean Field Approximation for Real Time Bidding with Budget Constraints.”

[5] Yang, Yaodong, Lantao Yu, Yiwei Bai, Ying Wen, Jun Wang, Weinan Zhang, and Yong Yu. "A Study of AI Population Dynamics with Million-agent Reinforcement Learning."

Thank You!

Ying Wen
ying.wen@cs.ucl.ac.uk