Multi-Agent Systems

Introduction to Multi-Agent Reinforcement Learning (MARL)

Eric Pauwels (CWI & VU)

December 14, 2021

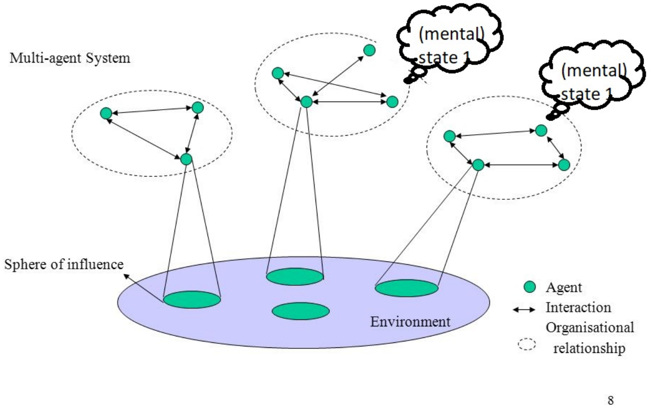
Outline

Reading

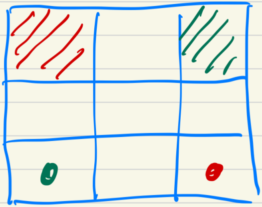
- P. Hernandez-Leal, M. Kaisers, T. Baarslag, E. Munoz de Cote: Survey of Learning in MultiAgent Environments: Dealing with Non-Stationarity. arXiv:1707.09183v2

MultiAgent Systems: Overview

Multi-agent Systems (MAS)



Simple two agent game



Goal:

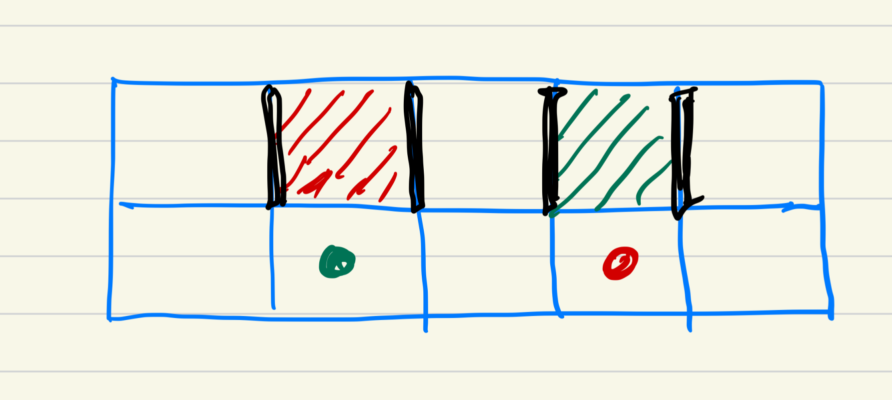
→ +10

→ +10

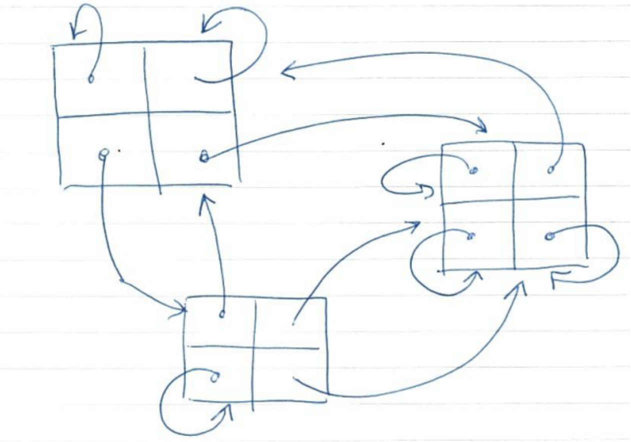
① game ends as soon as one goal is reached.

② game ends when there is a collision.

Simple two agent game

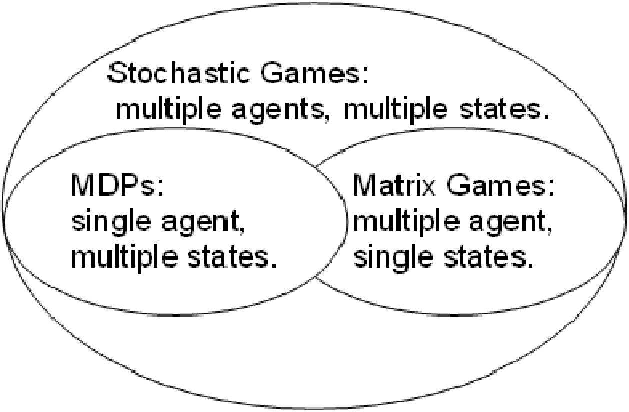


Stochastic Games (Markov Games)



Stochastic Games (Markov Games)

Stochastic games generalise **MDP** and **repeated matrix games**



MARL Problem Setting

- Several agents share environment and act on it **concurrently**:
- **Environment dynamics:** how does the environment change as a consequence of the **concurrent** actions by agents;
- **Non-stationary Problem:**
  - Environment is changing
  - **Learning:** you are **adapting your behaviour** in response to other agents;
  - **Teaching:**
    - Other agents might be learning and **adapting to your behaviour**!
    - They might be learning to take advantage of you!

Analogue for Bellman eqs. in MARL setting

- Bellman for q^* for single agent

$$
\begin{aligned}
q^*(s,a) &= \sum_{s'} p(s' \mid s,a)\left[r(s,a,s') + \gamma v^*(s')\right] \\
&= \underbrace{\sum_{s'} p(s' \mid s,a) r(s,a,s')}_{R(s,a)} + \gamma \sum_{s'} p(s' \mid s,a) v^*(s') \\
&= R(s,a) + \gamma \sum_{s'} T(s,a,s') \underbrace{\max_{a'} q^*(s',a')}_{\text{summary op.}} \\
&= R(s,a) + \gamma \sum_{s'} T(s,a,s') \underbrace{H_{a'} q^*(s',a')}_{\text{summary op.}}
\end{aligned}
$$

Analogue for Bellman eqs. in MARL setting

- **Bellman for q^* for single agent**

$$q^*(s,a) \quad = \quad R(s,a) + \gamma \sum_{s'} T(s,a,s') \underbrace{H_{a'} q^*(s',a')}_{\text{summary op.}}$$

- **Bellman for q^* for two agents**

$$s = (s_1, s_2) \quad \xrightarrow{(a',b')} \quad s' = (s_1', s_2')$$

$$q_1^*(s,(a,b)) = R_1(s,(a,b)) + \gamma \sum_{s'} T(s,(a,b),s') H_{(a',b')} q_1^*(s',(a',b'))$$

- Appropriate summarisation depends on assumptions about opponents;

Example: Maximin Q-learning for zero-sum games

- Agent (a) and Opponent (o);
- (Zero-sum game $u_a(a, o) = -u_o(a, o)$)
- Value $v(s)$ of a state (s) is safety level: using maximin strategy:

$$v(s) = \max_{\pi_a} \min_{o \in O} \sum_{a \in A} q(s, a, o)\pi_a$$

- Corresponding Q-learning rule:

$$q(s, a, o) = r(s, a, o) + \gamma \sum_{s'} p(s' \mid s, a, o) v(s')$$

Example: Nash-Q

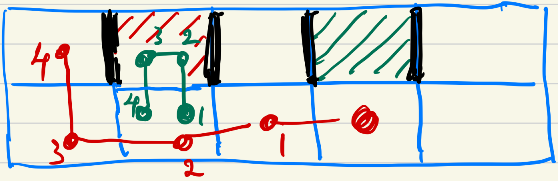
- Nash Q-value: represents an agent's expected future cumulative reward when, after choosing a specific joint action, all agents follow a joint Nash Equilibrium policy.

$$q_i(s,(a,b)) = R(s,(a,b)) + \gamma \sum_{s'} T(s,(a,b),s') v(s', \pi_1^*, \pi_2^*)$$

where $*$ indicates Nash policy.

Simple two agent game

Nash-eq. (mutual best response)



Different ways to handle non-stationarity

- **Ignore** Pretend environment is stationary;
  - Q-learing (single agent) or Fictitious Play
- **Forget** Favour more recent over older information;
- **Respond to target opponents** assume that opponent adheres to one of a class of well-defined strategies;
  - Friend-Q, Minimax-Q (Foe-Q), Nash-Q
- **Learn** model opponent and use this to plan;
- **Theory of Mind** Assume opponent is modelling you, and respond to that;

P. Hernandez-Leal, et al. Survey of Learning in MultiAgent Environments: Dealing with Non-Stationarity.

arXiv:1707.09183v2

Example of *Ignore*: Fictitious Play

- **Ignore the changes** in opponent strategies;
- **Model-based learning:** Opponent is assumed to be playing **stationary** but **unknown mixed strategy**;
- **Empirical:** Each agent uses **observed action frequencies** to estimate mixed strategy;
- Each player plays **best response to current estimate** of opponent's strategy

Fictitious Play: Convergence

Convergence in Fictious Play

If the empirical distribution of each player's strategy converge, then they converge to a Nash equilibrium.

Sufficient conditions for convergence in 2-player finite games:

- Zero sum game;
- Game is solvable using elimination of strictly dominated strategies:
- Potential game;

Overview of MARL algo's

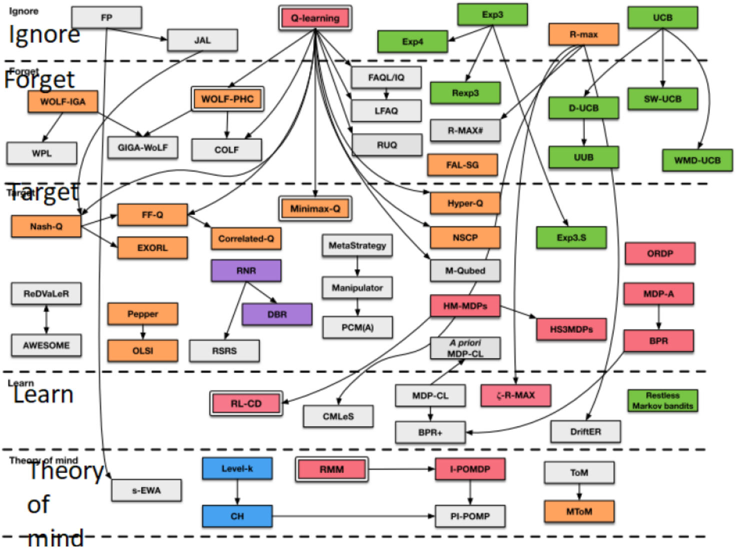


Image credit and further reading:

- P. Hernandez-Leal, et al. Survey of Learning in MultiAgent Environments: Dealing with Non-Stationarity. arXiv:1707.09183v2

L. Buşoniu, R. Babuška, and B. De Schutter, "Multi-agent reinforcement learning: An overview," Chapter 7 in *Innovations in Multi-Agent Systems and Applications – 1* (D. Srinivasan and L.C. Jain, eds.), vol. 310 of *Studies in Computational Intelligence*, Berlin, Germany: Springer, pp. 183–221, 2010.