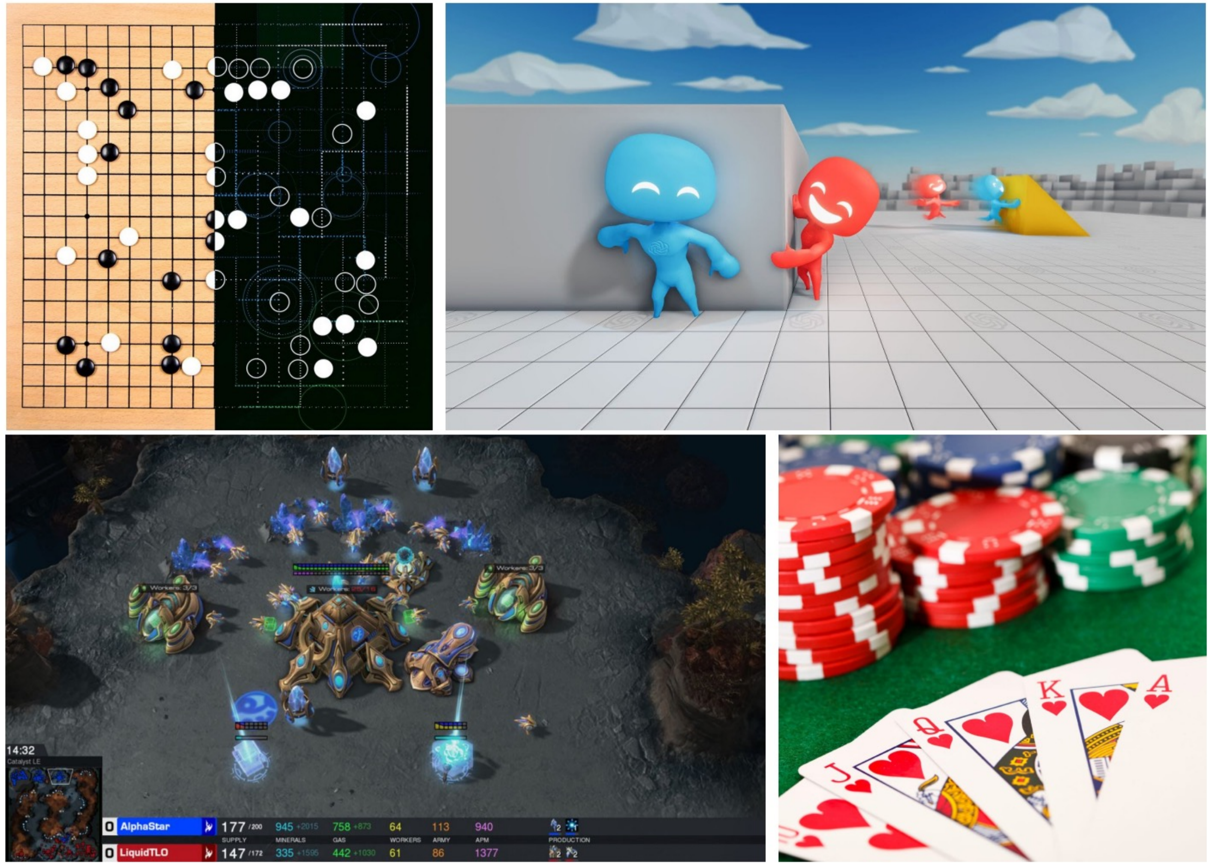
Chapter 10: Multi-agent RL

Applications of MARL

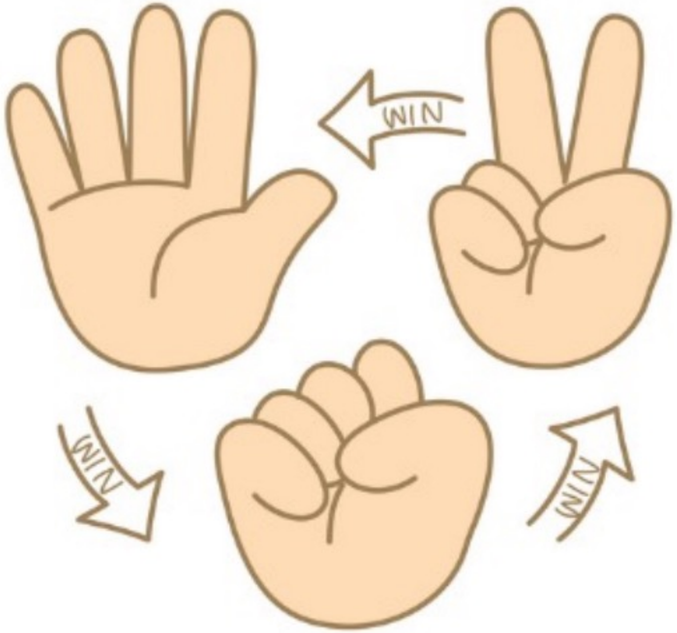


Games

Online bidding

Robot Fleet

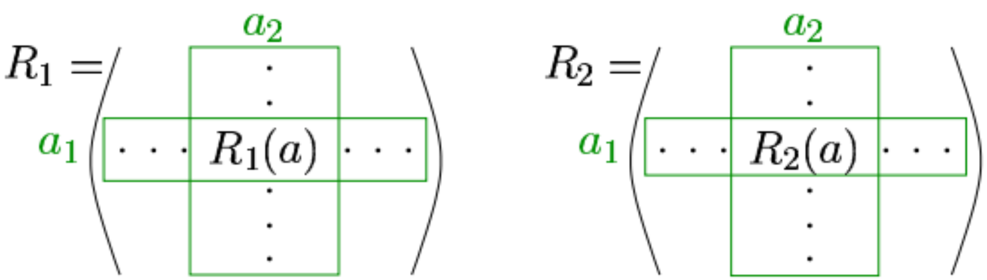
Normal-form Game – Simplest form of game



Normal-form Game – Simplest form of game

A normal-form game is a tuple $(n, \mathcal{A}_{1 \ldots n}, R_{1 \ldots n})$,

- n is the number of players,
- $\mathcal{A}_i$ is the set of actions available to player i
  - $\mathcal{A}$ is the joint action space $\mathcal{A}_1 \times \ldots \times \mathcal{A}_n$,
- R_i is player i's payoff function $\mathcal{A} \rightarrow \Re$.



Example 1: Rock-Paper-Scissors

- Two players. Each simultaneously picks an action:
  Rock, *Paper*, or *Scissors*.
- The rewards:

Rock	beats	*Scissors*
Scissors	beats	*Paper*
Paper	beats	*Rock*

- The matrices:

$$R_1 = \begin{array}{c} \\ \text{R} \\ \text{P} \\ \text{S} \end{array} \begin{array}{c} \begin{array}{ccc} \text{R} & \text{P} & \text{S} \end{array} \\ \begin{pmatrix} 0 & -1 & 1 \\ 1 & 0 & -1 \\ -1 & 1 & 0 \end{pmatrix} \end{array} \qquad R_2 = \begin{array}{c} \\ \text{R} \\ \text{P} \\ \text{S} \end{array} \begin{array}{c} \begin{array}{ccc} \text{R} & \text{P} & \text{S} \end{array} \\ \begin{pmatrix} 0 & 1 & -1 \\ -1 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix} \end{array}$$

Example 2: Prisoner's Dilemma

- Prisoner's Dilemma

$$R_1 = \begin{array}{c} \\ C \\ D \end{array} \begin{array}{c} \begin{array}{cc} C & D \end{array} \\ \begin{pmatrix} 3 & 0 \\ 4 & 1 \end{pmatrix} \end{array} \qquad R_2 = \begin{array}{c} \\ C \\ D \end{array} \begin{array}{c} \begin{array}{cc} C & D \end{array} \\ \begin{pmatrix} 3 & 4 \\ 0 & 1 \end{pmatrix} \end{array}$$

🤔What do we want to achieve? How do we define what's the desired behavior?

Example 3: Matching Pennies

- Three players. Each simultaneously picks an action: *Heads* or *Tails*.
- The rewards:

Player One	wins by matching	Player Two,
Player Two	wins by matching	Player Three,
Player Three	wins by *not* matching	Player One.

- The matrices:

$$R_1(\langle\cdot,\cdot,H\rangle) = \begin{array}{c} \\ H \\ T \end{array}\begin{array}{c} \begin{array}{cc} H & T \end{array} \\ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \end{array} \quad R_1(\langle\cdot,\cdot,T\rangle) = \begin{array}{c} \\ H \\ T \end{array}\begin{array}{c} \begin{array}{cc} H & T \end{array} \\ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \end{array}$$

$$R_2(\langle\cdot,\cdot,H\rangle) = \begin{array}{c} H \\ T \end{array}\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \quad R_2(\langle\cdot,\cdot,T\rangle) = \begin{array}{c} H \\ T \end{array}\begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$$

$$R_3(\langle\cdot,\cdot,H\rangle) = \begin{array}{c} H \\ T \end{array}\begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix} \quad R_3(\langle\cdot,\cdot,T\rangle) = \begin{array}{c} H \\ T \end{array}\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$$

"This would be a tensor"

Strategies

- What can players do?
  - Pure strategies (a_i): select an action.
  - Mixed strategies (σ_i): select an action according to some probability distribution.

Strategies

- Notation.
  - σ is a joint strategy for all players.

$$R_i(\sigma) = \sum_{a \in \mathcal{A}} \sigma(a) R_i(a)$$

  - σ_{-i} is a joint strategy for all players except i.
  - $\langle \sigma_i, \sigma_{-i} \rangle$ is the joint strategy where i uses strategy σ_i and everyone else σ_{-i}.

Types of Games

- Zero-Sum Games (a.k.a. constant-sum games)

  $$R_1 + R_2 = 0$$

  Examples: Rock-paper-scissors, matching pennies.

- Team Games

  $$\forall i, j \qquad R_i = R_j$$

  Examples: Coordination game.

- General-Sum Games (a.k.a. all games)
  Examples: Bach or Stravinsky, three-player matching pennies, prisoner's dilemma

Solution Concepts

- Prisoner's Dilemma

$$R_1 = \begin{array}{c} \\ C \\ D \end{array} \begin{array}{c} \begin{array}{cc} C & D \end{array} \\ \begin{pmatrix} 3 & 0 \\ 4 & 1 \end{pmatrix} \end{array} \qquad R_2 = \begin{array}{c} \\ C \\ D \end{array} \begin{array}{c} \begin{array}{cc} C & D \end{array} \\ \begin{pmatrix} 3 & 4 \\ 0 & 1 \end{pmatrix} \end{array}$$

🤔What kind of solution do we want?

Dominance

- An action is strictly dominated if another action is always better, i.e,

$$\exists a_i' \in \mathcal{A}_i \; \forall a_{-i} \in \mathcal{A}_{-i} \qquad R_i(\langle a_i', a_{-i} \rangle) > R_i(\langle a_i, a_{-i} \rangle).$$

- Consider prisoner's dilemma.

$$R_1 = \begin{array}{c} \\ \text{C} \\ \text{D} \end{array} \begin{array}{c} \text{C} \quad \text{D} \\ \begin{pmatrix} 3 & 0 \\ 4 & 1 \end{pmatrix} \end{array} \qquad R_2 = \begin{array}{c} \\ \text{C} \\ \text{D} \end{array} \begin{array}{c} \text{C} \quad \text{D} \\ \begin{pmatrix} 3 & 4 \\ 0 & 1 \end{pmatrix} \end{array}$$

  - For both players, D dominates C.

Minimax Optimal Solution

- Consider matching pennies.

$$R_1 = \begin{array}{c} \\ \text{H} \\ \text{T} \end{array} \begin{array}{c} \begin{array}{cc} \text{H} & \text{T} \end{array} \\ \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \end{array} \qquad R_2 = \begin{array}{c} \\ \text{H} \\ \text{T} \end{array} \begin{array}{c} \begin{array}{cc} \text{H} & \text{T} \end{array} \\ \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \end{array}$$

- Q: What do we do when the world is out to get us?
  A: Make sure it can't.

- Play strategy with the best worst-case outcome.

$$\underset{\sigma_i \in \Delta(\mathcal{A}_i)}{\operatorname{argmax}} \; \min_{a_{-i} \in \mathcal{A}_{-i}} \; R_i(\langle \sigma_i, \sigma_{-i} \rangle)$$

- Minimax optimal strategy.

Minimax Optimal Solution

- Back to matching pennies.

$$R_1 = \begin{array}{c} \\ H \\ T \end{array} \begin{array}{c} \begin{array}{cc} H & T \end{array} \\ \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \end{array} \qquad \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix} = \sigma_1^*$$

- Consider Bach or Stravinsky.

$$R_1 = \begin{array}{c} \\ B \\ S \end{array} \begin{array}{c} \begin{array}{cc} B & S \end{array} \\ \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \end{array} \qquad \begin{pmatrix} 1/3 \\ 2/3 \end{pmatrix} = \sigma_1^*$$

- Minimax optimal guarantees the saftey value.
- Minimax optimal never plays dominated strategies.

Nash Equilibria

- What action should we play if there are no dominated actions?
- Optimal action depends on actions of other players.
- A best response set is the set of all strategies that are optimal given the strategies of the other players.

$$\mathrm{BR}_i(\sigma_{-i}) = \{\sigma_i \quad | \quad \forall \sigma'_i \quad R_i(\langle \sigma_i, \sigma_{-i} \rangle) \geq R_i(\langle \sigma'_i, \sigma_{-i} \rangle)\}$$

- A Nash equilibrium is a joint strategy, where all players are playing best responses to each other.

$$\forall i \in \{1 \dots n\} \qquad \sigma_i \in \mathrm{BR}_i(\sigma_{-i})$$

Nash Equilibria

- A Nash equilibrium is a joint strategy, where all players are playing best responses to each other.

$$\forall i \in \{1 \ldots n\} \qquad \sigma_i \in \mathrm{BR}_i(\sigma_{-i})$$

- Since each player is playing a best response, no player can gain by unilaterally deviating.
- Dominance solvable games have obvious equilibria.
  - Strictly dominated actions are never best responses.
  - Prisoner's dilemma has a single Nash equilibrium.

Examples of Nash Equilibria

- Consider the coordination game.

$$R_1 = \begin{array}{c} \\ A \\ B \end{array} \begin{array}{c} \begin{array}{cc} A & B \end{array} \\ \begin{pmatrix} \boxed{2} & 0 \\ 0 & \boxed{1} \end{pmatrix} \end{array} \qquad R_2 = \begin{array}{c} \\ A \\ B \end{array} \begin{array}{c} \begin{array}{cc} A & B \end{array} \\ \begin{pmatrix} \boxed{2} & 0 \\ 0 & \boxed{1} \end{pmatrix} \end{array}$$

- Consider Bach or Stravinsky.

$$R_1 = \begin{array}{c} \\ B \\ S \end{array} \begin{array}{c} \begin{array}{cc} B & S \end{array} \\ \begin{pmatrix} \boxed{2} & 0 \\ 0 & \boxed{1} \end{pmatrix} \end{array} \qquad R_2 = \begin{array}{c} \\ B \\ S \end{array} \begin{array}{c} \begin{array}{cc} B & S \end{array} \\ \begin{pmatrix} \boxed{1} & 0 \\ 0 & \boxed{2} \end{pmatrix} \end{array}$$

Examples of Nash Equilibria

- Consider matching pennies.

$$R_1 = \begin{array}{c} \\ \text{H} \\ \text{T} \end{array} \begin{array}{c} \begin{array}{cc} \text{H} & \text{T} \end{array} \\ \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \end{array} \qquad R_2 = \begin{array}{c} \\ \text{H} \\ \text{T} \end{array} \begin{array}{c} \begin{array}{cc} \text{H} & \text{T} \end{array} \\ \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \end{array}$$

  - No pure strategy Nash equilibria. Mixed strategies?

$$\text{BR}_1\Big(\langle 1/2, 1/2 \rangle\Big) = \{\sigma_1\}$$

  - Corresponds to the minimax strategy.

Existence of Nash Equilibria

- All finite normal-form games have at least one Nash equilibrium. (Nash, 1950)

- In zero-sum games…

  - Equilibria all have the same value and are interchangeable.

    $$\langle \sigma_1, \sigma_2 \rangle, \langle \sigma'_1, \sigma'_2 \rangle \text{ are Nash } \Rightarrow \langle \sigma_1, \sigma'_2 \rangle \text{ is Nash.}$$

  - Equilibria correspond to minimax optimal strategies.

Computation of Nash Equilibria

- The exact complexity of computing a Nash equilibrium is an open problem. (Papadimitriou, 2001)
- Likely to be NP-hard. (Conitzer & Sandholm, 2003)
- For two-player games, bilinear programming solution.

Fictitious Play

(Brown, 1949; Robinson 1951)

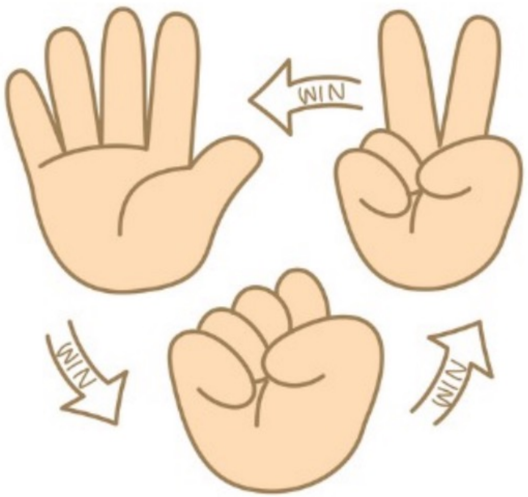
- An iterative procedure for computing an equilibrium.

  1. Initialize $C_i(a_i \in \mathcal{A}_i)$, which counts the number of times player i chooses action a_i.
  2. Repeat.
     (a) Choose $a_i \in BR(C_{-i})$.
     (b) Increment $C_i(a_i)$.

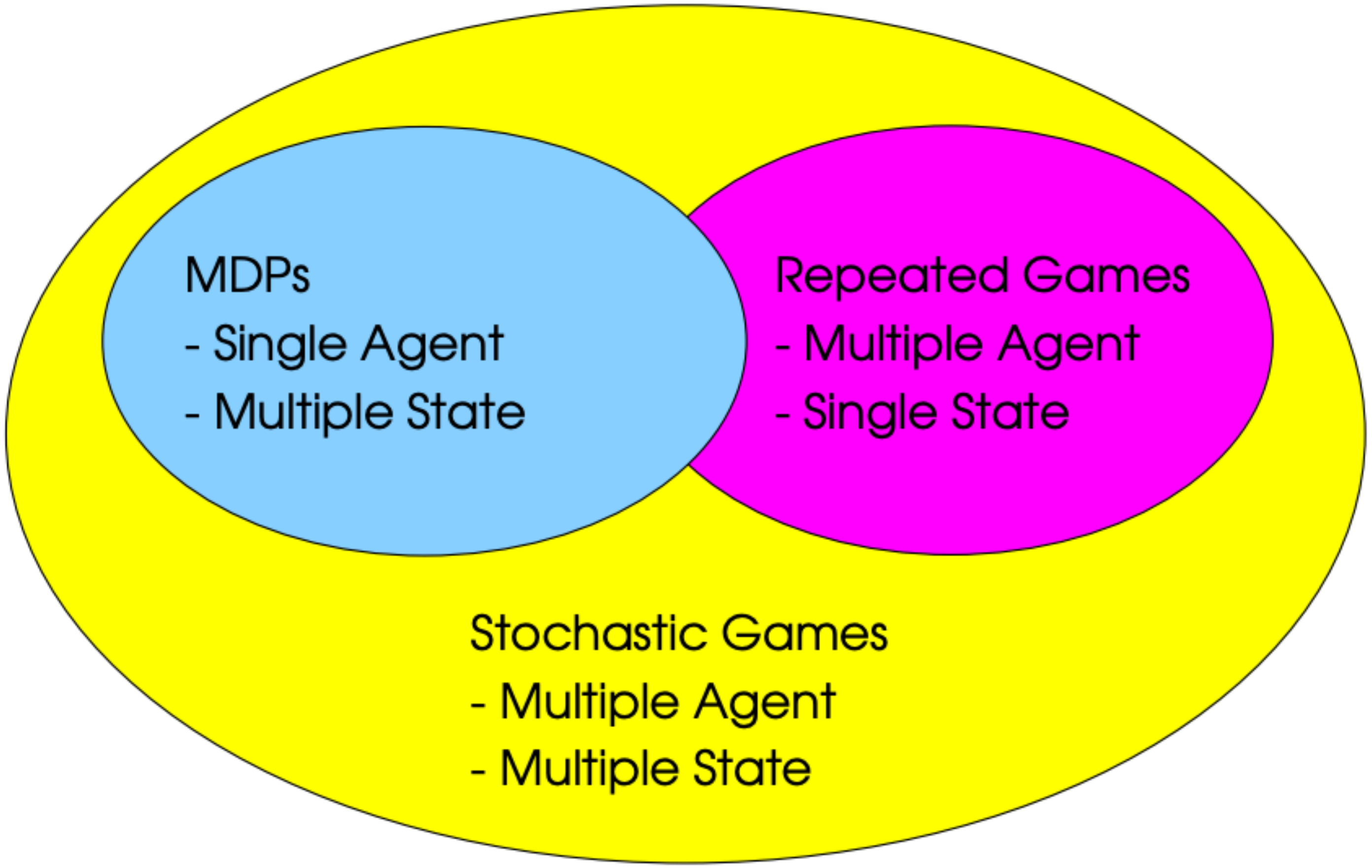
Fictitious Play

(Fudenberg & Levine, 1998)

- If C_i converges, then what it converges to is a Nash equilibrium.
- When does C_i converge?
  - Two-player, two-action games.
  - Dominance solvable games.
  - Zero-sum games.
- This could be turned into a learning rule.



Stochastic/Markov Games



Stochastic/Markov Games

A stochastic game is a tuple $(n, \mathcal{S}, \mathcal{A}_{1\ldots n}, T, R_{1\ldots n})$,

- n is the number of agents,
- $\mathcal{S}$ is the set of states,
- $\mathcal{A}_i$ is the set of actions available to agent i,
  - $\mathcal{A}$ is the joint action space $\mathcal{A}_1 \times \ldots \times \mathcal{A}_n$,
- T is the transition function $\mathcal{S} \times \mathcal{A} \times \mathcal{S} \to [0, 1]$,
- R_i is the reward function for the ith agent $\mathcal{S} \times \mathcal{A} \to \Re$.