# Multi-agent systems: basic concepts

ECE PARIS ÉCOLE D'INGÉNIEURS
ISIR INSTITUT DES SYSTÈMES INTELLIGENTS ET DE ROBOTIQUE

Maxime Toquebiau

11th of March 2021

**Agent**
A computational system able to perform actions in his environment based on information received from the environment. (Panait and Luke, 2005)

**Multi-agent environment**
One in which there are more than one agent. Agents are able to interact with one another and influence different parts of the environment. This may lead to dependency relationships between agents.

**Multi-agent learning**

The application of machine learning to problems involving multiple agents. These problem are by definition more complex:

- Multiple agents $\Rightarrow$ larger search space
- Multiple learners $\Rightarrow$ non-stationarity
- Small changes in learned behaviors $\Rightarrow$ big changes in the properties of the MAS

**Nash equilibrium**
In game theory, it is a set of strategies (one for each agent) such that no agent can gain by changing their strategy, as long as all the other agents keep theirs unchanged.
Thus it forms a stable state of the system.

**Pareto efficiency**
An outcome is Pareto efficient if there is no other outcome that makes at least one person better off without leaving anyone worse off.

| **Game theory** | **Reinforcement learning** |
| --- | --- |
| Player | Agent |
| Strategy | Policy |
| Best response | Greedy policy |
| Utility | Reward |
| State | State |
| Move | Action |
| Coalition | Team |
| Collective | System |

**Team learning** VS **Concurrent learning**

**Centralized** VS **Decentralized**

**Cooperative** VS **Competitive**

# Credit Assignment

Def: How to distribute the reward obtained at a team level to the individual learners.

Distribution rules: (Panait and Luke, 2005)

- **Global reward:** Split the team reward equally to each of the learners
- **Local reward:** Each agent gets its own individual reward based on his own individual behavior
- **Observational reinforcement:** Reward obtained by observing other agents and imitating their behavior (Mataric, 1994)
- **Vicarious reinforcement:** Small reward received whenever other agents are rewarded (Mataric, 1994)

## Aristocrat Utility (AU)

Def: The difference in world utility between the agent's action and the average action (Wolpert and Tumer, 2002).
The AU of an agent $i$ is defined as:

$$u_i(s, \overrightarrow{a}) = U(s, \overrightarrow{a}) - \sum_{\overrightarrow{a}' \in \overrightarrow{A}} Pr[\overrightarrow{a}']U(s, \overrightarrow{a}_{-i}, \overrightarrow{a}'_i),$$

with $\boldsymbol{u_i(s, \overrightarrow{a})}$: agent $i$'s reward in state $s$,
**U**: the world utility (i.e. global reward),
$\boldsymbol{\overrightarrow{a}}$: joint action of all agents in the MAS,
$\boldsymbol{\overrightarrow{a}_{-i}}$: joint action of all agents except $i$,
$\boldsymbol{Pr[\overrightarrow{a}']}$: probability that $\overrightarrow{a}'$ happens.

# Wonderful Life Utility (WLU)

Def: The change in world utility that would have arisen if the agent "had never existed" (Wolpert and Tumer, 1999; 2002).
The WLU of an agent $i$ is given by:

$$u_i(s, \overrightarrow{a}) = U(s, \overrightarrow{a}) - U(s, \overrightarrow{a}_{-i}, CL(\overrightarrow{a}_i)),$$

with $\boldsymbol{u_i(s, \overrightarrow{a})}$: agent $i$'s reward in state $s$,
**U**: the world utility (i.e. global reward),
$\boldsymbol{\overrightarrow{a}}$: joint action of all agents in the MAS,
$\boldsymbol{\overrightarrow{a}_{-i}}$: joint action of all agents except $i$,
$\boldsymbol{CL(\overrightarrow{a}_i)}$: agent $i$'s action clamped (e.g. replaced by $\overrightarrow{0}$, or averaged).

# Shapley Value

Def: The average of the marginal contributions of an agent $i$ in all the possible different coalitions (Shapley, 1953).
In a coalition game defined by a set $N$ of $n$ agents and $v(S)$ giving the worth of any coalition S, the Shapley value of agent $i$ is:

$$\varphi_i(v) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(n - |S| - 1)!}{n!} (v(S \cup \{i\}) - v(S))$$

$$\Leftrightarrow \varphi_i(v) = \frac{1}{n} \sum_{S \subseteq N \setminus \{i\}} \binom{n-1}{|S|}^{-1} (v(S \cup \{i\}) - v(S))$$

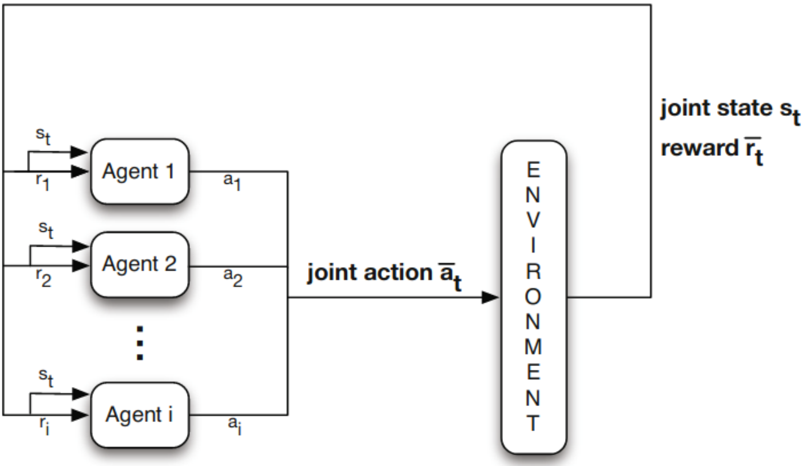
$$\Leftrightarrow \varphi_i(v) = \frac{1}{\text{number of players}} \sum_{\text{coalitions excluding i}} \frac{\text{marginal contribution of } i \text{ to coalition}}{\text{number of coalitions excluding } i \text{ of this size}}$$

Advantages:
Desirable properties:

- **Efficiency:** $\sum_{i \in N} \varphi_i(v) = v(N)$
- **Symmetry:** if $v(S \cup \{i\}) = v(S \cup \{j\})$ then $\varphi_i(v) = \varphi_j(v)$
- **Linearity:** $\varphi_i(v + w) = \varphi_i(v) + \varphi_i(w)$
- **Fairness:** if a player $i$ is *null* $(v(S \cup \{i\}) = v(S))$ then $\varphi_i(v) = 0$

Disadvantage:
Exponential computation time

Application examples:
Network design (Anshelevich et al., 2008; Michalak et al., 2014), MADRL (Shapley Q-value DDPG by Wang et al., 2020)




Nowé, Vrancx and De Hauwere, 2012

Thank you!