Multi-agent deep reinforcement learning in mobile robotics

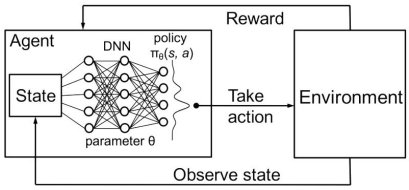
Maxime Toquebiau[1,2], Nicolas Bredeche[2], Faïz Benamar[2], Jae-Yun Jun[1]

[1]ECE Paris,
[2]Sorbonne Universités, CNRS, ISIR

Phd Thesis started in January 2021.

Three dimensions

Deep reinforcement learning (DRL)



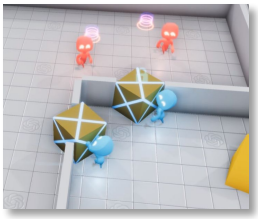
Goal

Maximise long-term cumulative reward

Tools

- **Deep learning**: extract valuable information from observations
- **Reinforcement learning**: learn a policy (sequence of actions) that maximises the reward

Multi-agent systems (MAS)



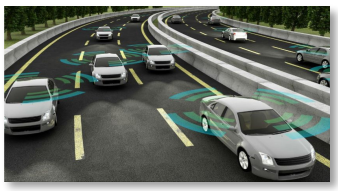
OpenAI (2019) - Hide and Seek

Multiple agents interacting with one another and with their environments.

Issues

- **Complexity**: More agents means that the joint spaces of states and actions are bigger and harder to process.
- **Non-stationarity**: With multiple agents constantly learning new behaviours, the environment is endlessly evolving.
- **Partial Observability**: Each agent's reward depends on events it can't observe.
- **Credit assignment**: All agents in a team have a different impact on the reward received at a team-level. We need ways to divide the reward based individual performance.

Mobile robotics



Constraints

- Complexity of real-world environments
- Imperfections of sensors and actuators
- Safety
- Limited communication
- Dynamic environments

Problematics

- How to design more relevant simulated environments ?
- How to use deep reinforcement learning in real-world environments ?
- How to use and interpret the actions of artificial agents ?
- How to interact with robots ?
- How to use a policy learnt in simulation on a real-world robot ?

Four directions

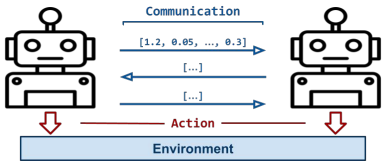
Emergent communication

Context

Despite its clear potential, communication isn't used enough in the multi-agent DRL literature.

Contribution

Study emergent communication in MAS and their impact on performance.



Learning an existing language

Context

- Robots will need to **interact** with human beings, and thus communicate with an **interpretable language**.
- Using a pre-existing and discrete language can not only help communicate with humans, but also drive the understanding of the environment.

Contribution

Design a language that can be learnt by artificial agents by interaction and that improve their performance in a chosen task.

Multi-agent credit assignment

Context

- In a cooperative multi-agent setting, the global reward isn't suitable for agents to learn their optimal policy, as they are not able to deduce the quality of their actions from the reward obtained at a team level.
- Communication takes part in finding the best solution to a given task, thus it must be taken into account when assigning credit to agents.

Contribution

- Study existing solutions and find the best suited for our context.
- Link credit assignment to communication.

Macro-actions as a sim-to-real approach

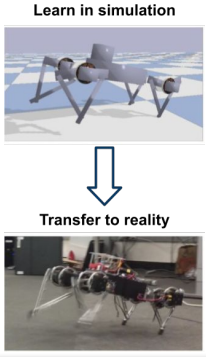
Context

DRL methods are notoriously hard to apply in real world environments due to:

- very low sample efficiency,
- need to experience very bad outcomes to find the optimal policy.

Contribution

Use macro-actions to learn high-level policies in simulation and transfer them easily in the real world.



Contact: maxime.toquebiau@ece.fr