What's done and what's to come

Maxime Toquebiau

May 10th, 2021

State-of-the-art Deep Reinforcement Learning

Value-based

- *Deep Q-Learning* (DQN; Mnih et al., 2013; 2015)
- *Deep Recurrent Q-Learning* (Hausknecht and Stone, 2015)
- *Prioritized Experience Replay* (Schaul et al., 2015)
- *Double DQN* (van Hasselt et al., 2016)
- *Dueling DQN* (Wang et al., 2016)
- *Distributional DQN* (Bellamare et al., 2017)
- *Noisy DQN* (Fortunato et al., 2017)
- *Rainbow* (Hessel et al., 2017)
- *Recurrent Replay Distributed DQN* (R2D2; Kapturowski et al., 2019)

Policy-based

- *Deep Deterministic Policy Gradient* (DDPG; Lillicrap et al., 2015)
- *Trust Region Policy Optimization* (Schulman et al., 2015)
- *Proximal Policy Optimization* (Schulman et al., 2017)
- *Twin Delayed DDPG* (Fujimoto et al., 2018)
- *Soft Actor Critic* (Haarnoja et al., 2018)

Model-based

- *AlphaGo* (Silver et al., 2016)
- *AlphaZero* (Silver et al., 2017)
- *World Models* (Ha and Schmidhuber, 2018)
- *PlaNet* (Hafner et al., 2018)
- *MuZero* (Schrittwieser et al., 2019)
- *Dreamer* (Hafner et al., 2019)
- *Plan2Explore* (Sekar et al., 2020)

Multi-agent Learning

Distribution rules

- **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)

Distribution rules

- **Shapley Value:** The average of the marginal contributions of an agent in all the possible different coalitions (Shapley, 1953)
- **Aristocrat Utility:** The difference in world utility between the agent's action and the average action (Wolpert and Tumer, 2002)
- **Wonderful Life Utility:** The change in world utility that would have arisen if the agent "had never existed" (Wolpert and Tumer, 1999; 2002)

Multi-agent Reinforcement Learning

Approaches

- **Sharing information:**
  - *Independent vs Cooperative Agents* (Tan, 1993)
- **Opponent Modelling:**
  - *Joint Action Learner* (Claus and Boutillier, 1998)
  - *LOLA* (Foerster et al., 2017)
- **Assuming the other agents' behavior:**
  - *Minimax Q-Learning* (Littman, 1994)
  - *Friend-or-foe Q-Learning* (Littman, 2001)
  - *Nash Q-Learning* (Hu and Wellman, 1998; 2003)
  - *Correlated Q-Learning* (Greenwald et al., 2003)

Approaches

- **Learning to coordinate:**
  - *Coordinated RL* (Guestrin et al., 2002)
  - *Sparse cooperative Q-Learning* (Kok and Vlassis, 2004; 2006)
- **Adaptation:**
  - *Win or Learn Fast* (WoLF; Bowling and Veloso, 2002)
  - *AWESOME* (Conitzer and Sandholm, 2005)
- **No-regret:**
  - *Generalized Infinitesimal Gradient Ascent* (GIGA; Zinkevich et al., 2003)
  - *GIGA-WoLF* (Bowling, 2005)

Limitations

- Limited environments or tasks
- Difficulties for generalizing
- Complex architectures
- Low scalability

Desirable properties

- Stability
- Adaptation
- Robustness
- Selective communication

Multi-agent Deep Reinforcement Learning

Policy-based

- **Mutli-Agent DDPG** (MADDPG; Lowe et al., 2017)
- **Counterfactual Mutlti-Agent:** (COMA; Foerster et al., 2018) credit assignment inspired by the Aristocrat Utility
- **Independent PPO** (IPPO; Schroeder de Witt et al., 2018)

Value-based

- **Stabilizing experience replay** (Foerster et al., 2017)
- **Value factorization:**
  - **Value Decomposition Networks** (VDN; Sunehag et al., 2018)
  - **QMIX** (Rashid et al., 2018)
  - **QTRAN** (Son et al., 2019)
  - **Weighted QMIX** (Rashid et al., 2020)
- **Multi-Agent Variational Exploration** (MAVEN; Mahajan et al., 2020)

Toy simulated environments

- StarCraft Multi-Agent Challenge (Samvelyan et al., 2019)
- Multi-Agent Particle Environment (Lowe et al., 2017)
- Multi-Robot Warehouse Environment
- Level-based foraging

Robotics simulated environments

- ROS
- Unity
- MuJoCo (Todorov et al., 2012)
- Robogym (OpenAI)
- Gibson Env (Xia et al., 2018)
- Multi-Agent MuJoCo (Schroeder de Witt et al., 2020)
- BOLeRo (Fabisch et al., 2020)

Definition of the subject

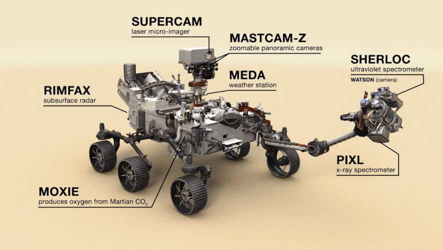
Multi-agent deep reinforcement learning in mobile robotics

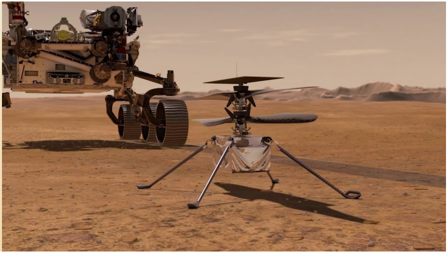
Definition of the subject

Multi-agent deep reinforcement learning in mobile robotics

$$\Downarrow$$

Space exploration





Perseverance rover and Ingenuity drone

Context

Discussions with space exploration professionals

Jorge Vago (ExoMars Project Scientist, ESA-ESTEC)

Multi-Robot Systems (MRS) applications - Space Exploration

- Usually multi-robot approaches are not worth it
- Prefer having a more capable single robot than four smaller ones
- MRS come handy when there is a *spatial dimension* that is crucial to characterise

Shreyansh Daftry (Robotics Technologist, NASA's JPL)

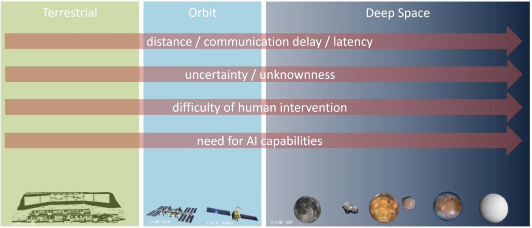
Heterogeneous MRS - Planetary Exploration - Autonomy - AI

- Humans slow missions down
- Autonomy is a major issue for NASA
- MRS are needed to scale missions
- MRS for space exploration will be heterogeneous

Context

Planetary exploration with robots

- Autonomous driving for rovers, with **very short range planning** (Olivier Toupet (NASA's JPL), IROS Workshop on Planetary Exploration (WPE) 2020)
- **Bad productivity** due to limited communication rate, bandwidth and reliance from the ground (David Wettergreen (CMU), IROS WPE 2020)
- Majority of the collected data isn't sent to the ground (David Wettergreen (CMU), IROS WPE 2020)
- Field science involves **constant wandering around** and **rescheduling** based on observations (David Wettergreen (CMU), IROS WPE 2020)



Roland Sonsalla (DFKI), IROS WPE 2020

Context

Motivations

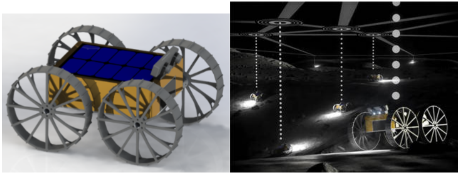
- Exploring rougher terrains (Alin Albu-Schäffer (DLR), IROS WPE 2020)
- Exploring various types of environments (canyons, gullies, caves, lava tubes...) (Alin Albu-Schäffer (DLR), IROS WPE 2020)
- Sample return (Alin Albu-Schäffer (DLR), IROS WPE 2020)
- Long range exploration (Gianfranco Visentin (ESA), IROS WPE 2020); (ADE)
- Autonomous science (PERASPERA OG-10 ADE)
- Find resources and exploit them (Gianfranco Visentin (ESA), IROS WPE 2020)

Autonomous Decision Making In Very Long Traverses (ADE), www.h2020-ade.eu

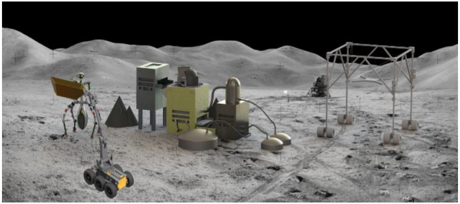
Context

Directions of space agencies

- **Long range exploration** (Gianfranco Visentin, IROS WPE 2020; ADE; Robinson et al., 2020)
- **Resource exploitation** (Govindaraj et al., 2019)
- **Planetary base assembly** (Govindaraj et al., 2019)
- **Machine learning-based autonomy** (Ono et al., 2020; Abcouwer et al.,2020)
- **Sample return** (Schuster et al., 2020)
- **Heterogeneous multi-robot systems** (Schuster et al., 2020; Ropero et al., 2019)



CADRE - NASA



PRO-ACT - ESA (Govindaraj et al., 2019)

Autonomy (David Wettergreen (CMU), IROS WPE 2020)

- Long range planning for navigation
- Scientific mission planning
- Going faster and further
- Provide more and better information

Heterogeneous MRS (Schuster et al., 2020)

- Robustness through redundancy
- Parallelization
- Complementary capabilities

Context

Existing approaches

Heterogeneous MRS

- Mathew et al., *Planning Paths for Package Delivery in Heterogeneous Multirobot Teams*, 2015
- Manjanna et al., *Heterogeneous Multi-Robot System for Exploration and Strategic Water Sampling*, 2018
- Krizmancic et al., *Cooperative Aerial-Ground Multi-Robot System for Automated Construction Tasks*, 2020

Reinforcement learning for exploration

- Viseras and Garcia, *DeepIG: Multi-Robot Information Gathering With Deep Reinforcement Learning*, 2019
- Matheron, *Integrating Motion Planning into Reinforcement Learning to solve hard exploration problems*, 2020

Reinforcement learning for heterogeneous MAS

- All papers tested on SMAC
- Xiao et al., *A Distributed Multi-Agent Dynamic Area Coverage Algorithm Based on Reinforcement Learning*, 2020

Reinforcement learning for MRS

- Chen et al., *Decentralized non-communicating multiagent collision avoidance with deep reinforcement learning*, 2017
- Long et al., *Towards Optimally Decentralized Multi-Robot Collision Avoidance via Deep Reinforcement Learning*, 2018
- Semnani et al., *Multi-Agent Motion Planning for Dense and Dynamic Environments via Deep Reinforcement Learning*, 2020

MRS in space exploration

- Schuster et al., *Towards Heterogeneous Robotic Teams for Collaborative Scientific Sampling in Lunar and Planetary Environments*, 2019
- Ropero et al., *TERRA: A path planning algorithm for cooperative UGV–UAV exploration*, 2019
- Govindaraj et al., *Multi-Robot Cooperation for Lunar Base Assembly And Construction*, 2020

Autonomous exploration for rover-drone system

Unmanned Aerial Vehicle (UAV)

System: flying drone (e.g. helicopter, quadcopter)

Tasks:

- Scout for interesting locations
- Collect data for path planning

Objective: Gather information for path planning

Unmanned Ground Vehicle (UGV)

System: planetary rover (e.g. Perseverance, ExoMars)

Tasks:

- Autonomous driving to interesting locations
- Scientific study
- Charging station for the drone

Objective: Drive fast and safely

→ Information gathering
→ Path planning
→ Interaction between heterogeneous agents

My contribution

Constraints

- Energy consumption
- Extreme temperatures
- Different conditions (gravity, atmosphere)
- Weight of payload
- Limited communication
- Robustness

Heterogeneity

- How to learn cooperative behavior with heterogeneous agents ?
- How to properly use multi-source information for path planning ?

Robustness

- How to make sure the policy we learn is reliable and robust ?

Learn complementary behavior in heterogeneous MAS with DRL

$\Downarrow$

Robust path planning from multi-source data with DRL

$\Downarrow$

Autonomous exploration of UGV-UAV system with DRL

- Collaboration with NASA
- Internship at NASA
- Collaboration with ESA or European national agencies ?

Thank you!