Multi-Agent NovelD
First results & Future works

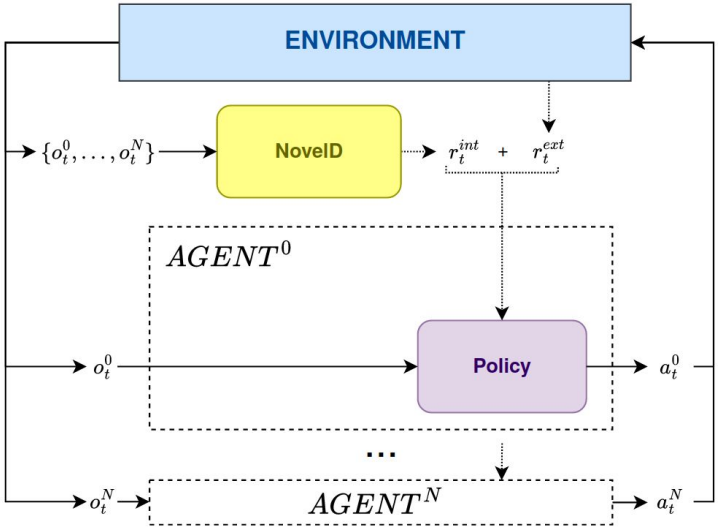
Maxime Toquebiau

November 17th, 2022

Table of Content

MA-NovelD

Architecture



MA_NOVELD

- One NovelD module for the whole multi-agent system

→ We look for novelty in the joint observations

MA-NovelD

Scenario iterations

Goal: Show that exploring the joint states is important to find a strategy that requires coordination.

New task:

- Object still needs to be pushed on the landmark
- Object can't move unless an agent stays on the button

Reward:

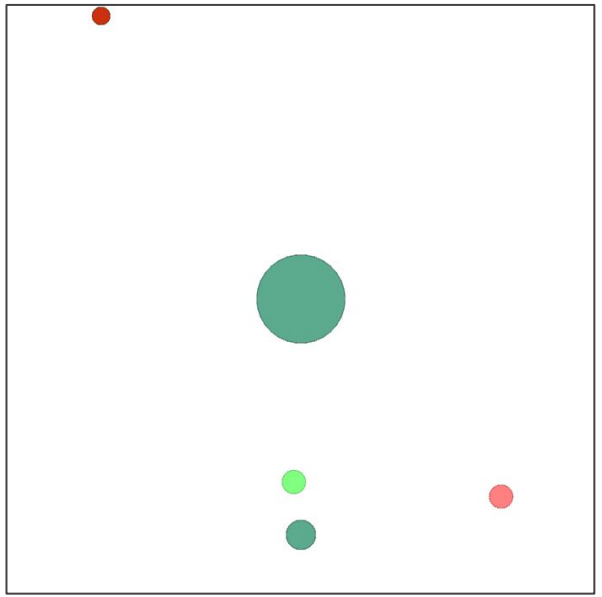
$$R_{step}(s_t) = -0,1$$

$$R_{success}(s_t) = \mathbb{1}_{success}(s_t) \times 50$$

$$R_{button}(s_t) = \mathbb{1}_{button\ pushed}(s_t) \times 0,5$$

$$R_{shaped}(s_t) = D_{obj,lm}(s_{t-1}) - D_{obj,lm}(s_t)$$

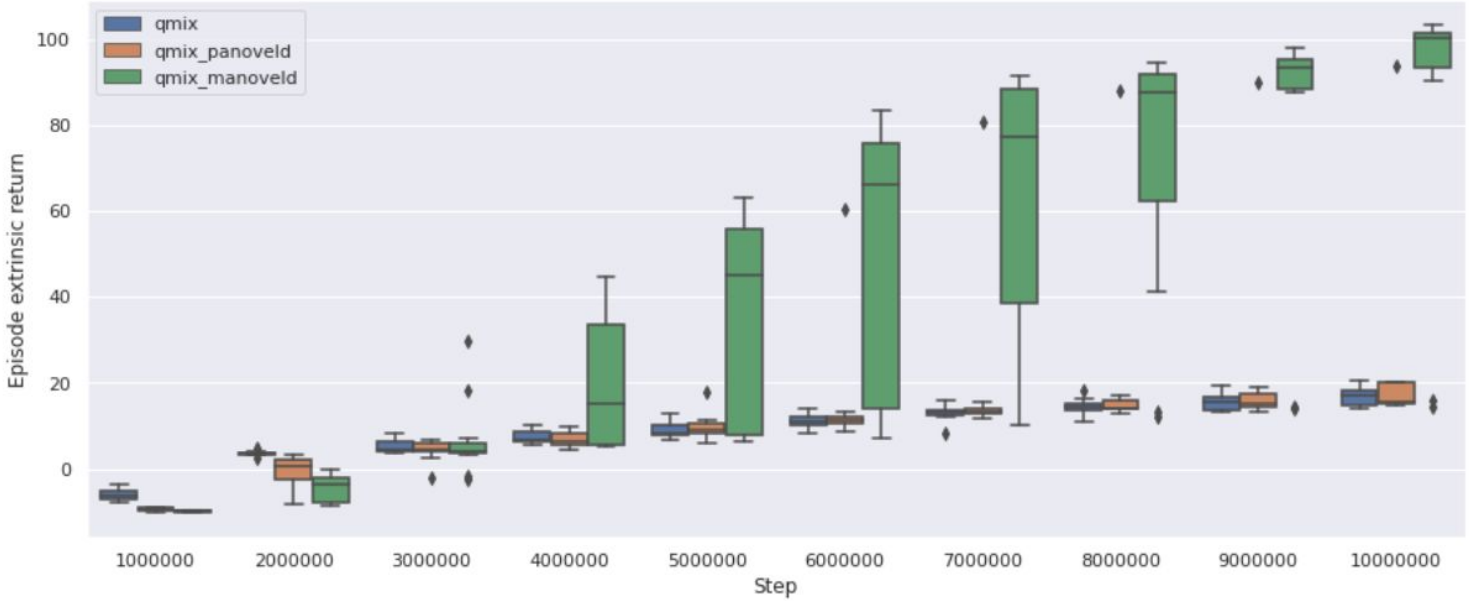
$$R_{tot} = R_{step} + R_{success} + R_{button} + 100 \times R_{shaped}$$



MA-NovelD

Results

Click-&-Push scenario, fixed initial positions of object and landmark.



MA-NovelD

Next steps

Other algorithms:

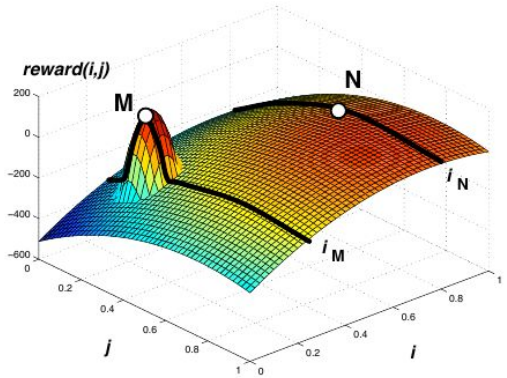
- MADDPG (policy-based, off-policy) Multi-Agent Actor-Critic for Mixed Cooperative-Competitive Environments, Lowe et al., 2017.
- MAPPO (policy-based, on-policy) Benchmarking Multi-Agent Deep Reinforcement Learning Algorithms, Yu et al., 2021.
- MASAC (policy-based, off-policy) Benchmarking Multi-Agent Deep Reinforcement Learning Algorithms, Yu et al., 2021.
- Weighted QMIX (value-based, off-policy) Weighted QMIX: Expanding monotonic value function for deep multi-agent reinforcement learning, Rashid et al., 2020.
- MAVEN (value-based, off-policy) MAVEN: Multi-Agent Variational Exploration, Mahajan et al., 2020.

MA-NovelD

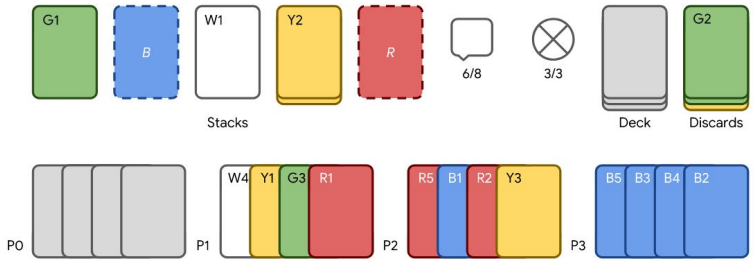
Next steps

Other tasks:

- Toy environment to showcase relative value overgeneralisation Lenient Learning in Independent-Learner Stochastic Cooperative Games, Wei and Luke, 2016.
- Predator-prey
- Hanabi The Hanabi challenge: A new frontier for AI research, Bard et al., 2020.



Relative overgeneralisation



Hanabi

MA-NovelD

Next steps: Upcoming conferences

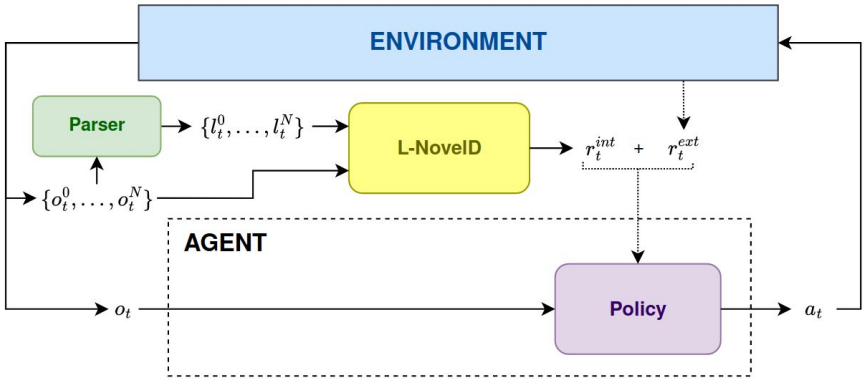
Conference	Submission deadline	Conference Date
IEEE International Conference on Distributed Computing Systems (ICDCS)	January 21, 2023	July 18-21, 2023
International Joint Conference on Artificial Intelligence (IJCAI) 2023	January 11, 2023	August 19-25, 2023
International Conference on Intelligent Robots and Systems (IROS) 2023	March 1, 2023	October 1-5, 2023
European Conference on Artificial Intelligence (ECAI) 2023	April, 2023	October 1-6, 2023

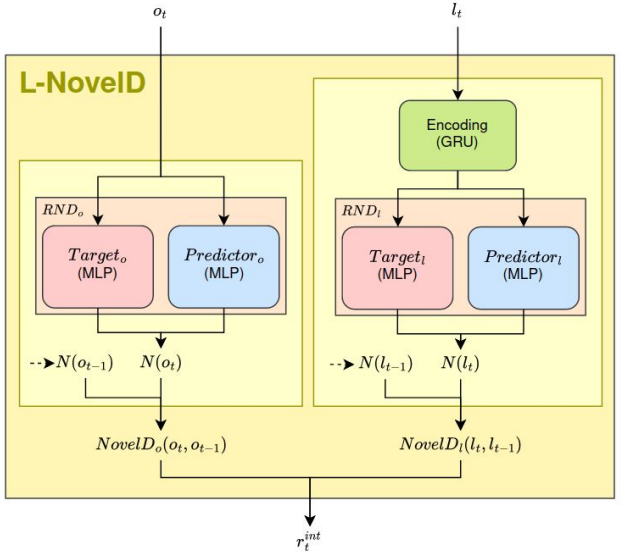
Language for intrinsic reward in MA-L-NovelD

Simplified architecture

Use language for the intrinsic reward only.

Goal: Search for novelty in language space, as language concentrate important information about the task.



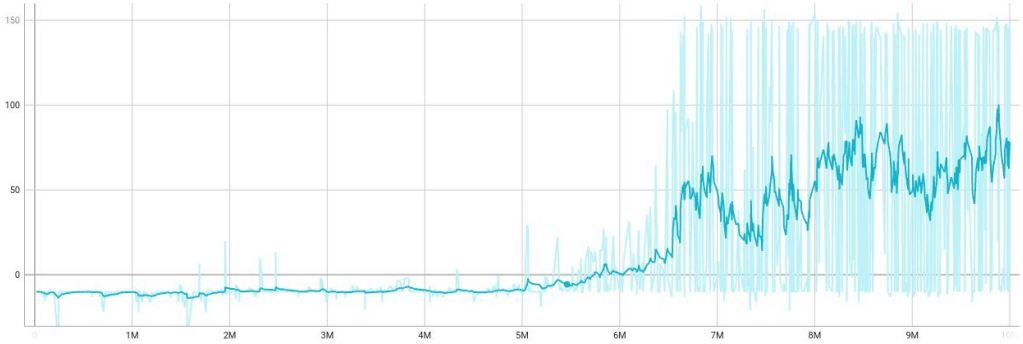


Language for intrinsic reward in MA-L-NovelD

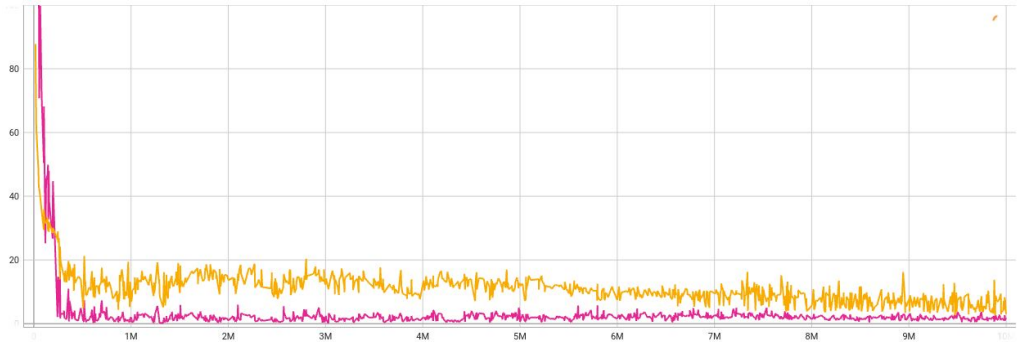
Issues

Issues with the language part of the reward:

- Language reward too low
  - ➢ Language very simple, thus very easy to learn fast
- Language reward doesn't reflect novelty of descriptions
  - ➢ New observed words don't impact the reward enough



Extrinsic reward of QMIX with MA-L-NovelD



Intrinsic rewards computed from:

- observations (yellow),
- language descriptions, scaled by 50 (pink).

Language for intrinsic reward in MA-L-NovelD

Next steps

- Play with L-NovelD hyperparameters
- Look into novelty literature for better intrinsic rewards:
  - more consistent between continuous and discrete state spaces
  - more sensitive to marginal novelty in observations
  - specific to language ?

Thank you for you attention !

Questions ?