ETH zürich

Impact of Reward Shaping on MicroRTS AI Agents

Timo Loher, Samuel Räber

Group 37

1 Introduction

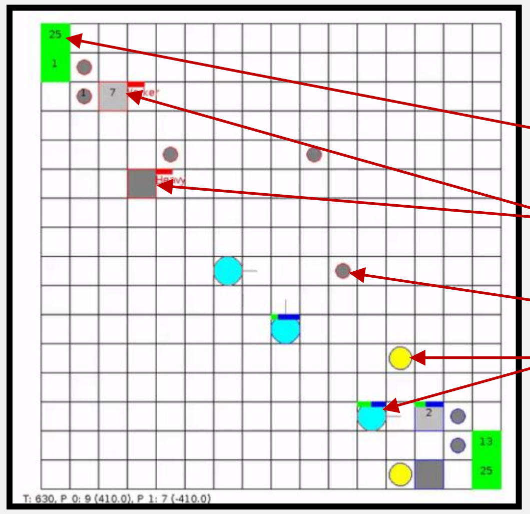
One difficulty of training AI agents on complex tasks like RTS games is the sparse reward and large action space. Reward shaping is used to give more frequent rewards on which the agent can train.

But agents learn to optimize this shaped reward instead of the true objective of the game.
Action guidance[1] is a technique which overcomes this problem. In this project we ran experiments to test the performance of action guidance, particularly with multiple auxiliary agents.

2 Method Overview

Action Guidance (AG): Training auxilary agents on shaped rewards to guide actions of main agent.
Goal: Agent reaches meaningful rewards quicker -> better sample efficiency

Environment:
We used MicroRTS-Py[2] as the training environment, a minimalistic RTS game designed for RL.



The MicroRTS environment:

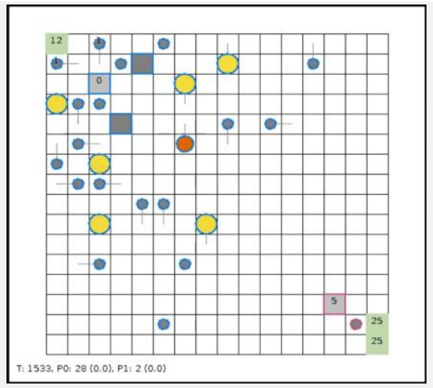
- Green squares are resources
- Grey squares are buildings
- Grey dots are workers
- Colored dots are combat units

3 Experiments

Produce Combat Units (PCU): Goal is to produce as many combat units as possible

Defeat Random Enemy (DRE): Beat an enemy which moves randomly with biased actions

Agents: Following table shows the reward weights; each color represents one auxilary agent

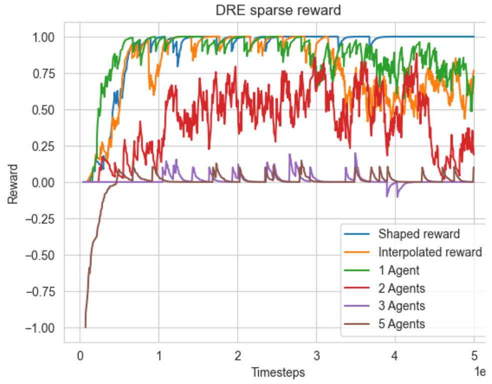


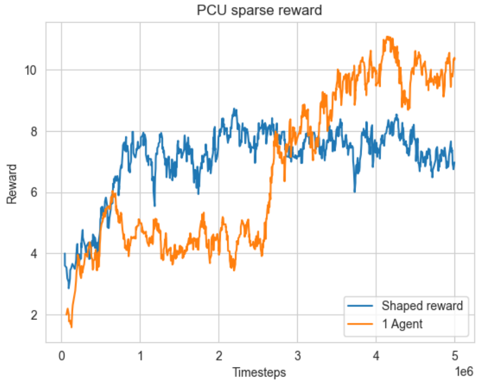
The action-guided agent performing the task Produce Combat units.

	Win	Make Worker	Harvest Resource	Make Building	Attack	Make Combat Unit
Shaped	10	1	1	0.2	1	4
1 Agent	10	1	1	0.2	1	4
2 Agents	0	1	1	0.2	1	4
3 Agents	0	1	1	0.2	1	4
5 Agents	0	1	1	1	1	1

4 Results

- On **PCU** this task we can see a clear trend that AG improves the optimization of the true reward compared to the shaped rewards
- On **DRE** all performed significantly worse than plain shaped rewards
- The more auxiliary agents, the worse the performance

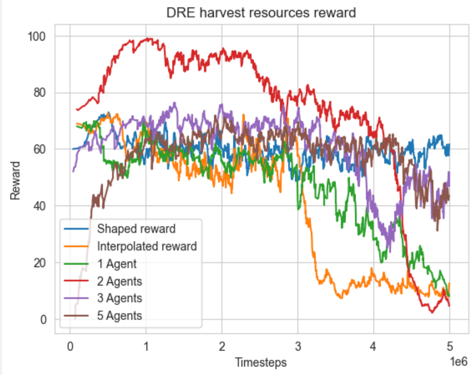


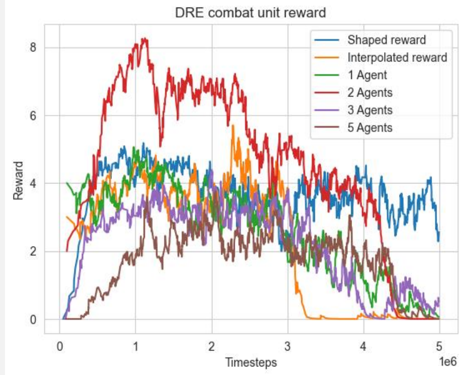


5 Discussion

- All agents learned to harvest resources and produce combat units for the DRE task
- Action guidance and interpolated agents unlearn this behavior

This leads us to assume the agents don't learn to win the game by actively attacking. They produce as much units as possible until one destroys the enemy base by chance.





We suggest to explore these options to improve the performance of AG with multiple auxiliary agents:

- Tuning the rewards: Different shaped rewards for the auxiliary agents might improve performance
- Add rewards for moving towards the enemy to counteract the erratic movement of units
- Longer training time and prolonged guidance so the main agent can better learn the more complicated DRE task

References

1. Shengyi Huang and Santiago Ontañón. Action Guidance: Getting the Best of Sparse Rewards and Shaped Rewards for Real-time Strategy Games. 2020. arXiv: 2010.03956 [cs.LG].
2. Shengyi Huang et al. Gym-µRTS: Toward Affordable Full Game Real-time Strategy Games Research with Deep Reinforcement Learning. 2021. arXiv: 2105.13807 [cs.LG].