GETTING AI TO COOPERATE: SHARING A CRITIC IN A VIDEO GAME

RESEARCH QUESTION

How does MAPPO compare to PPO when it comes to efficiency and collaborating with human-like agents in the Overcooked-AI environment?

AUTHOR

Jelle Groenendijk

AFFILIATIONS

TU Delft

Delft University of Technology

INTRODUCTION

Artificial Intelligence has proven to be able to beat humans in multiple games, but something that is more of a challenge is cooperating with humans. Overcooked-AI[1] is an environment where AIs have to cooperate with their partners to succeed.

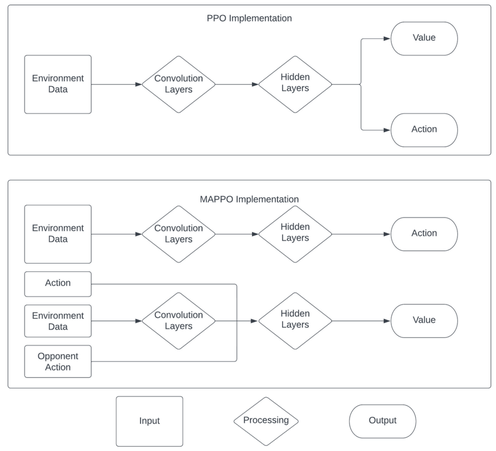
OBJECTIVE

The goal of my research is to let the trainer of the AI know what its partner is doing during training in the Overcooked-AI environment, hopefully allowing it to learn better by reacting to the strategy of its partner.

METHODOLOGY

Currently the Proximal Policy Optimisation[2] (PPO) algorithm is implemented in the Overcooked-AI environment. This algorithm trains a neural network by letting it take actions and then making it smarter by evaluating its actions. During the evaluation step, I will give it access to the actions its partner takes, turning the algorithm into a variant called Multi-Agent Proximal Policy Optimisation[3] (MAPPO).

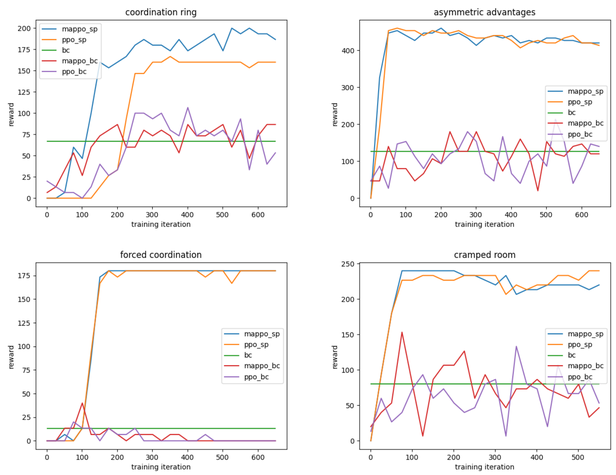
The Proximal Policy Optimisation implementation currently found in the Overcooked-AI environment uses just the output of the hidden layers to determine how well the agent has done.



The Multi-Agent Proximal Policy Optimisation implementation will also use knowledge of the opponents action to determine how well the agent has done.

Here we can see the results of the agent playing with itself in Self Play (SP), and the agent playing with a Behavioural Cloning agent (BC)

The results are quite similar for both the PPO and the MAPPO implementation. Both in how fast the agent learns, and the final results the agent achieves.



The only notable exception here being the MAPPO agent learning faster (in both SP and BC) and achieving higher scores (in SP) in the Coordination Ring map.

CONCLUSION

I believe these results are so similar to the PPO results because the agent evaluator already has a full view of the environment in the PPO algorithm. The MAPPO agent only adds the current action of the opponent, which seems to only really have an effect in the coordination ring map. Due to a lack of time, I also have not done any hyperparameter tuning to the MAPPO algorithm. This might also have changed the outcome.

FUTURE RESEARCH

Future research could tune the hyperparameters of the MAPPO model, as well as getting the Population Based Training from the original paper implemented in the current environment.

REFERENCES

[1] Micah Carroll et al. On the Utility of Learning about Humans for Human-AI Coordination. 2020. arXiv: 1910.05789 [cs.LG].
[2] John Schulman et al. Proximal Policy Optimization Algorithms. 2017. arXiv: 1707.06347 [cs.LG].
[3] Chao Yu et al. The Surprising Effectiveness of PPO in Cooperative, Multi-Agent Games. 2022. arXiv: 2103.01955 [cs.LG]