Training human-AI agents in Overcooked

Testing the potential of SyKLRBR in human-AI creation

Author: Jim Vos – j.vos-1@student.tudelft.nl – Supervisors: Robert Loftin, Frans Oliehoek

Introduction

Most cooperative problems are tackled by creating a team of agents who are optimized for each other and the problem. In this research the focus is on agents who can play in multiple unknown teams. These AI systems could be useful for human-AI interaction as different people bring a lot of variance into the system on. SyKLRBR is a new training method that showed potential in creating human-AI coordination agents but has limited tests and data [1].

Background

- **Ad-hoc play**: In this cooperative setting the players must play the game without any preknowledge of each other.
- **Overcooked:** This is a game where 2 players must cooperate to cook as my meals as possible. Researchers turned it into an environment for human-AI coordination research.
- **SyKLRBR**: This algorithm trains agents in a hierarchical order as visible in figure 1. The layers are only trained on the levels below itself where closer levels are chosen more often as the lines in figure 1 depict. The lowest agent picks its actions randomly. By training the agents differently a variance of strategies should occur. The top-level agent does encounter multiple strategies that constantly change due to synchronous training of the layers

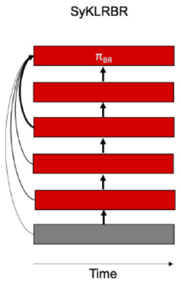

Figure 1: Visual representation of SyKLRBR [1]

Research question

The goal is to see if SyKLRBR is suitable for creating agents for human-AI coordination. In order to find out its capability this research tries to answer the question:

“Does an agent trained with SyKLRBR perform well in ad-hoc play in Overcooked”

Methodology

2 SyKLRBR agents were trained on different layouts. The cramped room visible in figure 2 is used to tests for low level coordination like collision avoidance. The layout in figure 3 is used to test for higher level strategies as it forces the players to cooperate in order to serve a meal.

The agent is played against different learning algorithms in ad-hoc play to compare its ability to generalize a strategy.

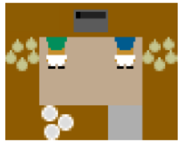

Figure 2: Cramped room

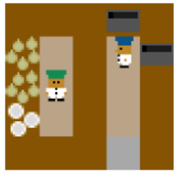

Figure 3: Forced coordination

Training method	SyKLRBR	PBT	PPO	BC	Human-proxy
SyKLRBR	123,34	120,11	90,54	111,59	106,91
PBT	120,11	155	146,79	93,29	89,23
PPO	90,54	146,79	199,19	98,43	100,67
BC	111,59	93,29	98,43	109,55	105,86
Human-proxy	106,91	89,23	100,67	105,86	103,52
Average ad-hoc	107,28	112,355	109,13	102,29	100,66

Figure 4: Average scores of cross-play in the cramped room (figure 3). All agents are trained over 5 millions steps.
*PBT = population-based, *PPO = proxy policy optimization, *BC = behavioral cloning

Results

SyKLRBR was not able to create a strategy for the forced coordination layout. The random agent made it almost impossible to learn for the lower levels, as no sparse rewards (delivering meals) could be obtained as seen in figure 5. agents could model around. Figure 6 shows how the top-level agent falls of as it starts optimizing to the failing lower strategies.

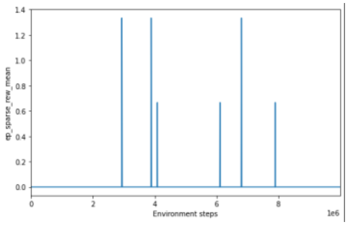

Figure 5: The points level 1 agent in SyKLRBR obtained for delivering meals in the forced coordination layout (figure 3).

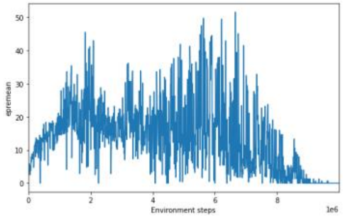

Figure 6: The total points level 5 agent in SyKLRBR obtained in the forced coordination layout (figure 3).

Cross-play shows SyKLRBR does not perform the best in ad-hoc play but is able to get consistent scores. SyKLRBR can make general for ad-hoc play strategies but lacks optimization compared to population-based training .

Conclusion

SyKLRBR does create a diverse set of training agents which makes the top-level robust against different strategies. SyKLRBR is impractical for highly cooperative settings as the random agent prohibits the learning on sparse rewards. SyKLRBR has potential to be used in human-AI with adaptation to more medium cooperative environments.

References

[1] B. Cui, H. Hu, L. Pineda, and J. Foerster, “K-level reasoning for zero-shot coordination in hanabi”, in Advances in Neural Information Processing Systems, M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan, Eds., vol. 34, Curran, Associates, Inc., 2021, pp. 8215–8228.