Coupled and Model-based cooperative planning in Overcooked AI

Author: Nils van Veen (N.vanVeen-3@student.tudelft.nl)
Supervisors: Frans Oliehoek, Robert Loftin

TU Delft

1. Background

- **Overcooked AI**: Simplified version of Overcooked [1] to simulate collaborative tasks.
- **Coupled planning (CP) with replanning**: Compute (near-)optimal joint & re-plan route.
- **Model based planning**: Compute (near-)optimal decision based on learned human model.

2. Research Question & Goals

What are the strengths and weaknesses of coupled planning with replanning as a solution to the ad-hoc teamwork problem?

- Reproduce results for CP and Model-based planning from [2].
- Improve upon the obtained results for CP with a specific focus on adapting to human behaviour.

3. Methodology

- Run existing planning experiments in Overcooked AI on a subset (⅔: Figure 2) of the experiment layouts:
  1. **Cramped room**: Tests the ability of how an agent can optimize the result, while colliding easily.
  2. **Asymmetric Advantages**: Tests whether players can choose high-level strategies to play to their strength.
- Compare results (Figure 1) with original paper [2].
- Change how Coupled failures are handled in CP to create and evaluate CPx.

CP: Coupled planning with replanning.
CPx: Improved Coupled planning with replanning.
$P_{H_{Proxy}}$: Model-based planning with respect to a true human model.
P_{BC}: Model-based planning with BC (Behavioural Cloning).
H_{Proxy}: Human proxy = "simulated" human model [2].

4. Results

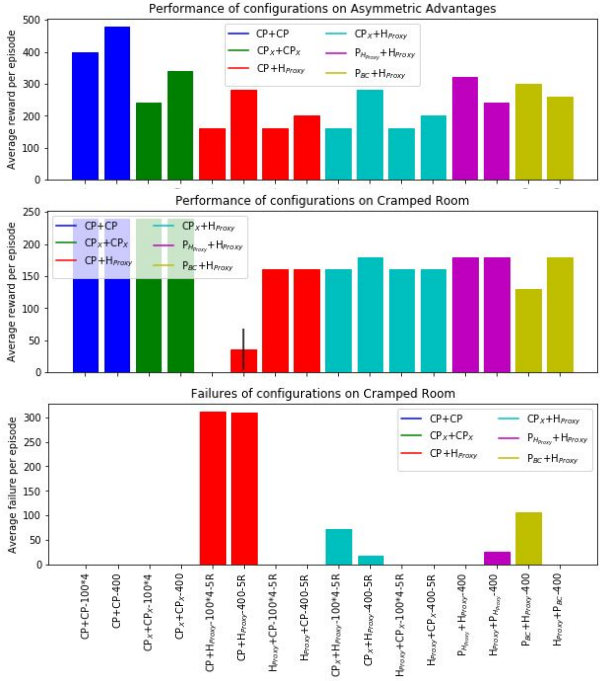

Figure 1: (a) Performances obtained on Asymmetric Advantages. (b) Performances obtained on Cramped Room (c) Collision failures on Cramped Room. *x*-axis: Planning configuration - steps * multiplication - runs R. Higher *Average reward per episode* and lower *Average failure per episode* is better.

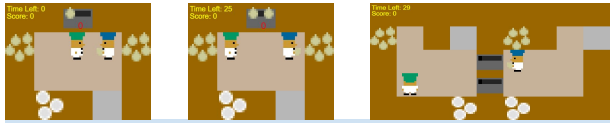

Figure 2: Cramped Room (failure cases 1 & 2) and Asymmetric Advantages, flTR .

5. Conclusions

- **CP+CP**: Better self-play performance than reinforcement learning. Matches original [2] results.
- **$P_{H_{Proxy}}$+ H_{Proxy} & P_{BC}+H_{Proxy}**: Matches original [2] results.
- **CP + H_{Proxy}**: Inconsistent results, due to collision failures. Matches original [2] results.
- **Extrapolation suboptimality**: Evaluation on 100 step horizon multiplied by 4 is less than evaluation on 400 step horizon.
- **Collision failures**: Blocking agents, impossible moves. See Figure 2, failure cases.
- **Reduce Collision failures**: Deviate from optimal play by walking into the opposite direction, let the human solve problem.

6. Future Work

- Add position states and orientations to CP.
- Combine ATPO [3] with Coupled Failures.
- Use statistics from [4] to predict whether human understands collision and switch agent roles accordingly.

7. References

[1] G.T Games, *Overcooked*, https://ghosttowngames.com/overcooked/, Published: 2016.
[2] M. Carroll, R. Shah, M. K. Ho, et al., "On the utility of learning about humans for human-ai coordination," CoRR, vol. abs/1910.05789, 2019. arXiv:1910 . 05789. [Online]. Available: http://arxiv.org/abs/1910.05789.
[3]J. G. Ribeiro, C. Martinho, A. Sardinha, and F. S. Melo, "Assisting unknown team- mates in unknown tasks: Ad hoc teamwork under partial observability," CoRR, vol. abs/2201.03538, 2022. arXiv: 2201.03538 [Online]. Available: https://arxiv.org/abs/2201.03538.
[4] S. van Waveren, C. Pek, J. Tumova, and I. Leite, "Correct me if i'm wrong: Using non-experts to repair reinforcement learning policies," Mar. 2022, p. 2022.