15-888 Computational Game Solving (Fall 2025)

Course Project Proposal

Student: Boxiang Fu (boxiangf@andrew.cmu.edu) **Date: September 29, 2025**

1 Description

The overall aim of the project is to use CFR to train a robust model of human-robot interaction when the human is non-cooperative or adversarial. More concretely, a Franka robotic manipulator interacts in the environment in the presence of a human. The robot wishes to perform a particular task (e.g. picking up an item) to maximize its reward. The human attempts to act adversarially and minimize the reward of the robot. This would be modeled as a two-player zero-sum game and CFR self-play would be used to train a model for the robot manipulator that is robust in adversarial human-robot interaction settings. Such a model could be useful in human-robot interactions (such as factories or production lines) where the human is not cooperative or might not know the presence of the manipulator. We model the human as adversarial to prepare for the worst.

2 Game Description (Tentative)

2.1 Agents

- Player 1: Robot manipulator (most likely the Franka Emika Panda arms available for use in NSH).
- Player 2: Human.

2.2 Actions

- Player 1: Choose different paths, choose different poses, wait/advance, re-plan, etc.
- Player 2: Block region, nudge object, occlude camera, do nothing, etc.

2.3 Information Sets

The human would have perfect recall of its path actions. The human would also have perfect information of the robot's actions. The robot could potentially have partial information of human's actions (e.g. might have occluded camera and also blocked region, but robot may not know if region is blocked).

2.4 Payoff

The payoff would be zero-sum. Positive rewards would be given to the robot if it achieves its desired goal (e.g. picking up an item). Negative rewards would occur if it hits the human. Contrastingly, negative rewards would be given to the human it the robot achieves its goal. Positive rewards would occur if the manipulator hits the human. If neither happens, each agent gets a payoff of zero.

2.5 Time

The game will be played sequentially in alternating fashion between the robot and the human. If the game has not ended after T time-steps, it is considered a draw and each player gets 0 payoff. If it ends before then, then payoff is determined as above.

3 Implementation

3.1 High Level Planning

The CFR would dictate the high level planning for the robot (and human). The action space and game states could be abstracted using game abstraction into discrete information sets that CFR can run through.

3.2 Low Level Control

For low level control, position or force control could be used to move the robot manipulator. A good choice could be the `Frankapy` package developed here at CMU.

4 Deliverables

For the deliverables of the final project, it would be in the form of an algorithm that trains via self-play to output an epsilon-Nash equilibrium in the robot's average strategies. This would give a robust high-level planning algorithm that is more safe around humans to feed to the low-level controllers. It will also output what strategies the human could perform to be most adversarial to the robot. The deliverables for the baseline, standard, and stretch goal implementations are as follows:

- **Baseline**: CFR implementation in a perfect information setting. The robot arm would fully observe the actions of the human.
- **Standard**: CFR implementation in an imperfect information setting. The human is able to occlude the robot so that the manipulator is only able to partially observe the human's actions.
- **Stretch**: CFR implementation in a perfect information setting. However, CFR would be run at every game tree node so that the resulting solution is a sub-game perfect Nash equilibrium and the human-robot interaction is robust/safe in all possible actions the player can take (including in off-equilibrium settings).

References

Kevin Zhang, Mohit Sharma, Jacky Liang, and Oliver Kroemer. A modular robotic arm control stack for research: Franka-interface and frankapy. *arXiv preprint arXiv:2011.02398*, 2020.

Noam Brown, Adam Lerer, Sam Gross, and Tuomas Sandholm. Deep counterfactual regret minimization, 2019. URL https://arxiv.org/abs/1811.00164.

Jean Tarbouriech, Evrard Garcelon, Michal Valko, Matteo Pirotta, and Alessandro Lazaric. No-regret exploration in goal-oriented reinforcement learning, 2020. URL https://arxiv.org/abs/1912.03517.

Yongyuan Liang, Yanchao Sun, Ruijie Zheng, Xiangyu Liu, Benjamin Eysenbach, Tuomas Sandholm, Furong Huang, and Stephen McAleer. Game-theoretic robust reinforcement learning handles temporally-coupled perturbations, 2024. URL https://arxiv.org/abs/2307.12062.

Catherine Glossop, William Chen, Arjun Bhorkar, Dhruv Shah, and Sergey Levine. Cast: Counterfactual labels improve instruction following in vision-language-action models, 2025. URL https://arxiv.org/abs/2508.13446.

Stefanos Nikolaidis, Swaprava Nath, Ariel D. Procaccia, and Siddhartha Srinivasa. Game-theoretic modeling of human adaptation in human-robot collaboration. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, HRI '17, page 323–331. ACM, March 2017. doi: 10.1145/2909824.3020253. URL http://dx.doi.org/10.1145/2909824.3020253.

Yong Wang, Pengchao Sun, Liguo Shuai, and Daifeng Zhang. Double counterfactual regret minimization for generating safety-critical scenario of autonomous driving. *Electronics*, 13(21), 2024. ISSN 2079-9292. doi: 10.3390/electronics13214303. URL https://www.mdpi.com/2079-9292/13/21/4303.