Physical Human-Robot Adversarial Gameplay

Boling Yang*, Patrick Lancaster†, Joshua R. Smith‡

*Department of Electrical Engineering, University of Washington
Washington, Seattle 98195, Email: bolingy@uw.edu

†Paul G. Allen School of Computer Science and Engineering, University of Washington
Washington, Seattle 98195, Email: planc509@cs.washington.edu

‡Department of Electrical Engineering and Paul G. Allen School of Computer Science and Engineering
University of Washington
Washington, Seattle 98195, Email: jrs@cs.washington.edu

***Abstract*—Modern robots are no longer limited to working in structured factory settings, and the near future will bring robots into our work spaces, our public areas, and even our homes. As close proximity co-existence with robots becomes increasingly commonplace, some will be intimidated by the appearance, clumsiness and inhuman-like movements of robots. We believe that one way to reduce such anxiety around robots is to incentivize human-robot interaction. Specifically, by developing a robot that challenges the human to play games that are rich in physical interaction, the human can gain both enjoyment and familiarity with the robot. This work aims to develop a reinforcement learning Physical Human-Robot Adversary (PHRA) framework to train a robot to play physical games with a human in a two-player, zero-sum, minimax game theoretic setting.**

I. Physical Human-Robot Gameplay

Advances in artificial intelligence have enabled incredible interactions between humans and machines. For example, recent progress in natural language processing has led to the creation of numerous voice assistants that respond to verbal requests for information, to play music, and even for companionship. While there are a number of examples of studies of interaction between humans and robots, physical interaction between them has been limited to short, simple interactions, such as the robot giving the human a hug or high-five, by Fitter and Kuchenbecker [1]. We propose that viewing certain instances of physical human-robot interaction through the lens of adversarial gameplay can facilitate more complex and enjoyable behavior.

Computers (intelligent agents) have become very good at playing adversarial games with discrete state and action spaces, such as chess, poker, and go. But their capabilities are greatly diminished in the realm of games that require players to dynamically and physically respond to each other, such as in sports or games for children like patty cake, red hand, and hot potato. These types of games have rarely been examined from a human-robot gameplay perspective. In this work, we focus on developing a framework to enable a robot to play physical games with a human in a two-player, zero-sum, minimax setting. While zero-sum, minimax is just one type of adversarial game setting, other settings will be explored in the future.

A physical human-robot game can be described by a system state S_t and game score r_t. For time t, $S_t = (s_t^1, ..., s_t^k)$ is a set of k continuous values, and $r_t = (r_t^1, r_t^2)$ is a tuple of continuous values. The continuous values in S_t describe the position, velocity and/or acceleration of relevant human and robot body frames (e.g. torso, left wrist, etc...) and/or related dynamic object frame(s). r_t^1 and r_t^2 are the robot and human's instant reward at time t respectively. In the zero-sum, minimax setting, $r_t^2 = -r_t^1$. The robot will try to maximize long term reward R from start time to final time T, $R = \sum_{t=0}^{t=T-1} \gamma^t r_t^1$ (with discount factor γ), while the human will try to minimize it.

II. The PHRA Framework

Reinforcement learning tasks typically require a massive amount of data, and collecting real world human-robot gameplay data is extremely time consuming. As a result, it is nearly impossible to train a robot to robustly perform a given task with only real world data. In this framework, we propose to initialize the robot's policy in simulation. Afterward, the robot will adapt to actual human opponents through a limited amount of real-world gameplay.

A. Adversarial Training Under Simulation

To overcome the problem of generating data, we propose to use the Robust Adversarial Reinforcement Learning (RARL) method by Pinto et al. [2] to initialize the robot's policy in simulation. Although the RARL method was originally designed for training an agent in simulation that is robust to unexpected disturbances, its game theoretic setting makes it suitable for training two agents that play games against each other. A physical human-robot game can be represented by an MDP $(S, A_1, A_2, P, R, \gamma, s_0)$, where S, γ, and R are the system state, discount factor, and reward as defined above. Assuming that the first agent simulates the robot and the second agent simulates the human, A_1 and A_2 are the action spaces for the robot and human respectively. P is the state transition probability and s_0 represents the initial state distribution.

The two agents are trained in an alternating manner, where the robot (protagonist) faces off against the human (antagonist). However, in simulation the human is modeled as another robot so that there is a simple method for controlling the antagonist. The protagonist will first be trained by collecting trajectories that result from playing against an adversary with

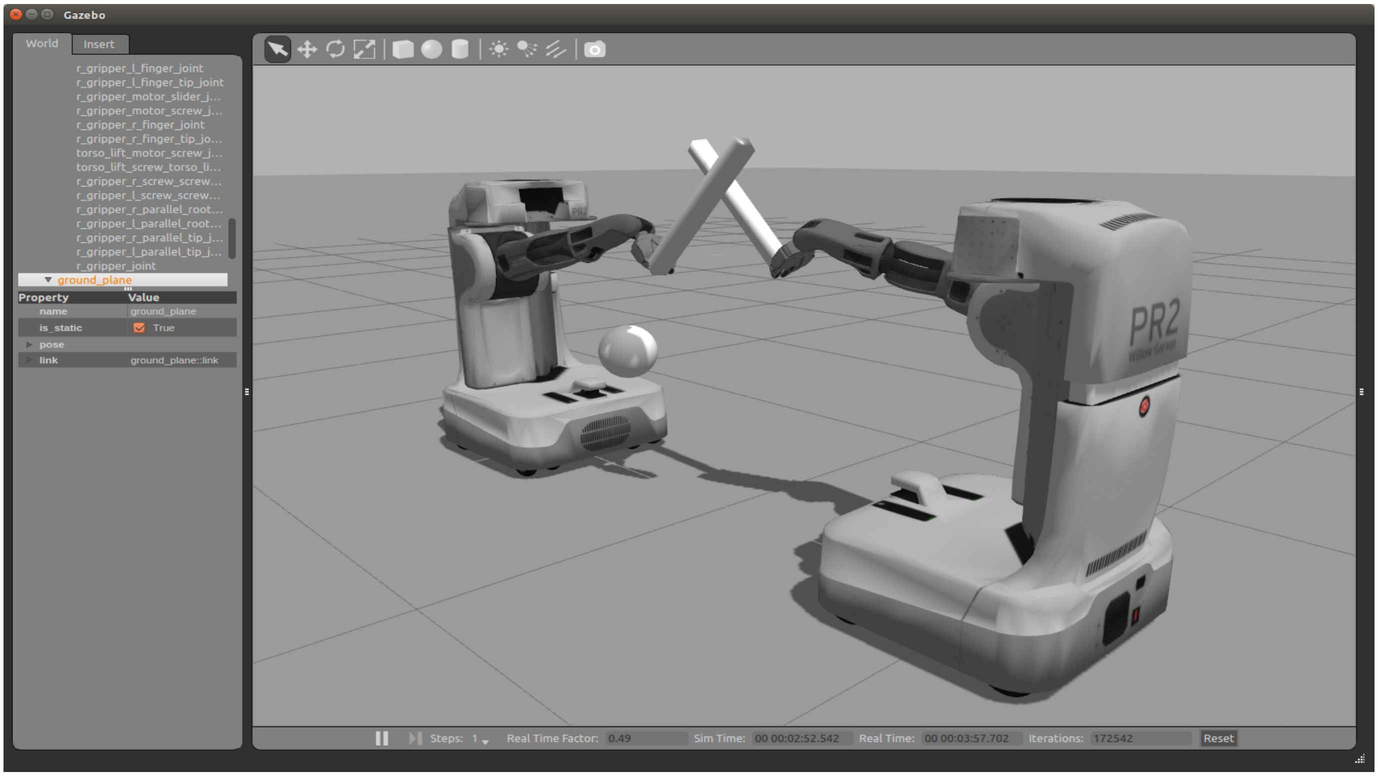

Fig. 1. The protagonist (left) attempts to use the hand-held rod to prevent the antagonist (right) from making contact with the sphere.

a static policy. This continues until the protagonist's policy achieves good performance against the adversary's current policy. The adversary will then be trained against the protagonist with a static policy in order to find a policy that the protagonist's policy is not robust to. This training sequence is repeated until the game reaches its Nash equilibrium, in which the policy of both agents have converged.

B. *Adopting Human Factors*

Once the robot's policy has been initialized by training in simulation, some amount of real world training is required in order to adapt to gameplay against a real human. Furthermore, the robot should be able to play against any human, which may require it to learn how to adapt to different 'styles' of gameplay. The proposed process to achieve such gameplay is as follows:

1) Collect trajectories from n human players.
2) Reduce the dimensionality of the trajectories through the use of time series principal component analysis.
3) Cluster trajectories based on principal components found in previous step.
4) Using the policy trained in simulation as an initialization, train dedicated policy for each cluster with model-based reinforcement learning methods.

When evaluating this process, the robot will constantly classify the human player's play style and actively select the appropriate policy. On the other hand, an on-line approach of game style adaptation will be explored in subsequent work.

III. Project Status

We will verify the proposed framework in a fencing-like scenario in which the protagonist is trying to prevent the antagonist from making contact with an object. We have implemented the learning framework and accompanying simulation, as shown in Fig. 1. Besides having the two agents train against each other in simulation, our next goal is to collect human arm movement trajectory samples to examine our proposed method for adapting the policy learned in simulation to gameplay against real humans.

References

[1] Naomi T. Fitter and Katherine J. Kuchenbecker. Analyzing human high-fives to create an effective high-fiving robot. In *ACM/IEEE International Conference on Human-Robot Interaction, HRI'14, Bielefeld, Germany, March 3-6, 2014*, pages 156–157, 2014. doi: 10.1145/2559636.2563718. URL http://doi.acm.org/10.1145/2559636.2563718.

[2] Lerrel Pinto, James Davidson, Rahul Sukthankar, and Abhinav Gupta. Robust adversarial reinforcement learning. *CoRR*, abs/1703.02702, 2017. URL http://arxiv.org/abs/1703.02702.