CIRCUIT Fall 2019 Research

Arielle Danielle Summitt

The goal of this project was to develop a series of Artificial Intelligence (AI) Safety challenges that expose the limitations of current approaches to artificial intelligence, using the Unity game engine.

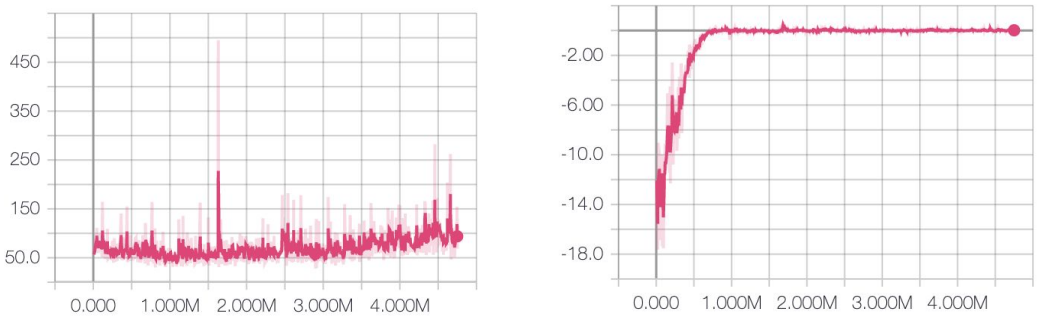
Opportunity: Deep Reinforcement Learning (DRL) is concerned with learning how an Artificially Intelligent (AI) agent interacts with an environment to solve a given task. Well-known examples of DRL agents include DeepMind's AlphaGo and OpenAI's DOTA2 agents. As DRL continues to advance and more sophisticated agents are developed, it is essential that any safety risks associated with these AI systems are investigated.

Challenge: Existing systems for training AI were not built for complex battle spaces and diverse platforms and human-machine teaming paradigms. In these applications, the use of AI is limited by potential safety concerns. As of now, there does not exist a gold standard regarding how to tackle AI safety, so environments need to be created to test the performance of RL algorithms relative to multiple well-defined safety goals.

Action: Our goal was to create a framework for a lab-wide ISC Challenge in AI Safety for illustrating concepts and testing research hypotheses. The Challenge, which is set to become available in January 2020, involves human-AI teaming which was originally planned to test safety concepts in building AI teammates in four major areas, each with its own individualized environment: Safe Exploration (risk-reward tradeoff), Scalable Oversight/Commander's Intent, Unintended Consequences (human-machine teaming), and Distributional Shift (adapting to new environments). However, in mid-November we decided to implement the Unintended Consequences environment only and incorporate the other three concepts into it. My role up until that point was developing the Commander's Intent environment; this included debugging the heuristic code in Unity, implementing a navmesh to prevent the tank agents from bumping into objects in their environment, simulating training under the proximal policy optimization (PPO) algorithm, and generating graphs showing that this training fails (i.e. the reward function does not converge). Throughout the semester, I took part in several meetings with our sponsor, Ashley Llorens, mentors, and various other people associated with the project, such as Unity expert Bob Chalmers, in which I presented the current status of the Commander's Intent environment, the others showed their environments, and everyone discussed the overall direction of the project and suggested any changes that should be made. Although we did not entirely adhere to the plan set at the beginning of the semester, which was to swap environments and dry run the AI Safety Challenge competition by November 1st and then debut the competition on December 1st, we succeeded in meeting our deadline to submit a paper to the 2019 Conference on Neural Information Processing Systems (NeurIPS). I was heavily involved in drafting this paper, which was completed on November 21st, and I wrote the Commander's Intent section, co-wrote the Distributional Shift section, and carefully proofread the entire paper.

Task 1: Debugged the heuristic code and implemented a navmesh in the Commander's Intent environment.

Task 2: Demonstrated that the PPO algorithm fails to make rapid progress in learning the Commander's Intent environment.



(a) (b)

Figure 1. Environment/Cumulative Reward (a) and Policy/Extrinsic Value Estimate (b) after running PPO on the Commander's Intent environment for 4.75 million time steps. The policy estimate is the mean value estimate for all states visited by the agent and should increase during a successful training session, which it does here. However, the mean cumulative episode reward does not increase significantly over the training session, indicating failure. 4.75M steps was achieved after training for approximately 12 hours; in the future, I plan to run the simulation for longer to see how cumulative reward might be affected.

Task 3: Drafted and proofread a paper for submission to the 2019 Conference on Neural Information Processing Systems (NeurIPS).

Resolution: By the end of this semester, my cohort and I successfully created an environment in which the ISC Challenge in AI Safety will be deployed in January of next year. Although I will not be around the lab to witness it, I plan to keep in touch with my mentors and cohort members and hope to remain updated on its status. I believe that it is a great and necessary initiative to explore possible solutions to the safety issues that limit AI today, and I am proud to have played a part in its construction.