It should include information on what you have done so far - data collection, programs you have used and programs you have written, some initial test results if you have any, and other issues you have considered.  It should also briefly describe what the remaining tasks to be accomplished for the whole project.

Midterm Progress Report

Ben Bush, Steven Hartman, Eric Tang

**Project Status**

The project is well underway and on pace with the work plan that was outlined in the Short Review Essay. Ben has undertaken the development of the A3C algorithm, Eric the DQN algorithm, and Steven the experimental data. All team members have devoted time to understand the challenge, observation space, action space, and reward function and are now fully engaged in their respective tasks.

The challenge provides a “boiler-plate” repository, with sample PPO and DDPG implementations, and instructions for how to create new agents. We have cloned this helper repository and used it as a starting point for our own work. The challenge provides the simulation environment as a Python 3 module, so all code is written in Python 3 to interface with the simulation environment. All neural network controllers are implemented using Keras and Tensorflow. Keras has a sister library, Keras-RL, that serves as an API for many common deep reinforcement learning algorithms. It allows users to program against DQN, PPO, and DDPG agents. Users can supply their standard Keras neural network architectures and any hyperparameters to the API, which then takes care of constructing the training and prediction functions that would otherwise need to be implemented. Furthermore, Keras-RL natively integrates with OpenAI Gym, which the provided simulation environment is based on. Thus, Keras-RL is a natural fit for our project and allows us to focus on some more interesting parts of the challenge, such as the neural network controller architectures and use of experimental data.

*Eric and Steven, put your stuff here. See note at the top regarding what to put in.*

*Steven: here would be a good place to talk about your data collection and supervised model from experimental data.*

Ben has nearly completed his implementation of the A3C algorithm. Unfortunately, Keras-RL, or any other deep reinforcement learning library, does not currently provide an A3C agent due to the complexity of the algorithm. As a result, we have had to develop the algorithm from scratch. Functionality of the algorithm has taken precedence over the architecture of the actor and critic networks. One major challenge presented by the A3C algorithm is that it the algorithm is asynchronous: it uses multithreading to create many rollouts in the simulation environment, and then once the rollouts have completed, each thread contributes what it has “learned” to the global gradients. Each thread is then updated with the new global gradients, and new rollouts are executed. Multithreading has many benefits: it speeds up the training process and encourages diversity in the training environment, which should make the model more robust. However, a multi-threaded agent was difficult to develop. We implemented multithreading using the threading module in Python. This module allows us each thread to execute a thread training function, which use performs the rollouts and updates the model parameters. The actor and critic networks are currently implemented as fully connected networks with all non-output layers shared, as was suggested in the original A3C paper. In the future, we intend to experiment with convolutional and recurrent neural networks for the controller. Using the simplified architecture and training for only 1000 episodes, the A3C algorithm achieves a score of around 50.

**Work Plan for Remaining Tasks**

Given that we are on pace with our original plan, we do not see the remaining tasks deviating too much from the schedule we outlined in the Short Review Essay. The next immediate task is to bootstrap the neural network controllers using the supervised model created with the experimental data. The remaining tasks involve tweaking network architecture and exploring new methodologies, such as frame skipping, to speed up training time. A time schedule for remaining tasks is outline below. Hours spend waiting for neural network controllers to finish training are not included in estimates.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Task** | **Resources Needed** | **Hours Spent** | **Hours Needed to Complete** | **Target Completion Date** |
| Midterm Progress Report | Text editor | 1 | 3 | 11/15 |
| Integrating Supervised Model | Supervised model, access to completed agents | 1 | 5 | 11/18 |
| Tweaking Controllers | GPUs | 0 | 18 | 11/28 |
| Exploring methodologies | GPUs | 0 | 6 | 12/2 |
| Presentation | PowerPoint, functioning agents | 0 | 8 | 11/29 or 12/2 |
| Final Report | Text editor, all other tasks completed | 0 | 8 | 12/6 |

**Issues**

As we noted in our Short Review Essay, it may be difficult to communicate or work on some of the hands-on tasks during the break for Thanksgiving. We have planned for this disturbance in our original schedule and plan to work on more “administrative” tasks, such as code documentation and an initial draft of the report and presentation during this time.

One major issue we foresee is availability of resources. Most of our algorithms are extremely compute-intensive. We have seen many posts on the challenge page describing some breakthrough strategies as brute force and using many GPUs to train a given agent for hundreds of hours. Although we have 3 GPUs across our team, we no longer have weeks to train our agents and may simply be constrained by compute time.