Spatial State Representations for Deep Reinforcement Learning, Milestone 1
15-300, Fall 2018

Joshua Zhanson
http://jzhanson.com/15400-s19

December 16, 2018

1 Major Changes

No major changes. We are targeting CoRL 2019, which has a submission deadline in June 2019 and a conference date of September or October 2019.

2 What have you accomplished so far

So far, we have created an OpenAI Gym environment based on the BipedalWalker-v2 environment that allows for arbitrary JSON agent body configurations to be input, and created both baseline MLP models and new convolutional models that integrate with the environment to get a projection of the agent body into its input grid. We have run these models for some time on several of our created datasets, which include a randomized body dimensions (within 25% tolerance) and randomized number of body segments BipedalWalker body dataset, and a BipedalWalker dataset with randomized dimensions, number of segments, and off-center leg attachment point, with promising results—we have at least one convolutional model that learns to solve the environment faster than the baseline MLP with the same amount of information.

3 Meeting your milestone

I have not yet met the milestone as described in my project proposal. Despite promising preliminary test episode reward curves showing that some convolutional models are able to solve the task faster than the baseline MLP models, full, more exhaustive evaluations of all the model checkpoints of these promising models are still in progress. The delay lies in the fact that evaluations take a great deal of time to complete, especially since it has become clear that we can only run 10-20 evaluation at a time for the several dozen validation bodies on the possibly hundreds of model checkpoints, per model. About as much time is spent on training models as I expected, however.

4 Surprises

I expected most of the time to be spent waiting for models to train, and while that is a significant portion of the project, I discovered that model evaluations take as much if not more time, especially since our evaluation framework, which involves the parallel execution command

`sem`, does not play very well with a high number of worker threads (sometimes across cluster machines), likely due to the finite number of semaphores of the directory where we are saving the evaluation results. I plan to spend a great deal of time in the next month optimizing and streamlining our evaluation process.

5 Revisions to your 15-400 milestones

I expect once I have streamlined our evaluation procedure and commands and we are able to run more than two dozen evaluations commands at a time that most of the blockers should dissolve, so I still project being able to meet the rest of my 15-400 milestones. Also, I intend to spend some time to clean up the code already written and reduce some of the technical debt I have incurred thus far to make the environment and the models more easily extensible in the future and get some tools for monitoring and interacting with the cluster up and running to make the training and evaluation process easier.

6 Resources Needed

So far, yes. I will do some careful study into the machinery used for running multiple commands in parallel as well as the scripts I wrote to run these commands and see if there are any good alternatives that do not suffer from the aforementioned issues. I am confident that there is a good solution to this problem and, if not, there is a dirty hack I can come up with to work around this problem, such as writing evaluations to different directories. If I am unable to find either a good solution or a kludge, then it might be time to consider using AWS for evaluations—AWS is also a worthy alternative for selective, judicious use if the MLD cluster will be crowded in the future, since I have some leftover credits from previous courses.