Spatial State Representations for Deep Reinforcement Learning, Milestone 2
15-400, Spring 2019

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

January 30, 2019

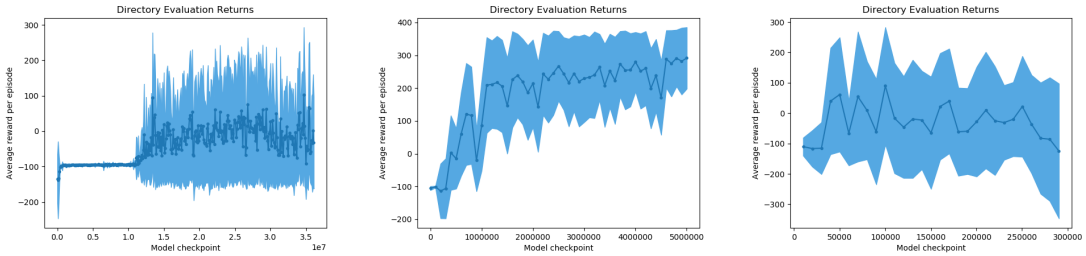

Figure 1: **Left:** best baseline MLP model, trained on 60 body dataset and evaluated on 60 body dataset. **Center:** best body-space model, trained on 60 body dataset and evaluated on 60 body dataset. **Right:** baseline MLP model, trained on single canonical body and evaluated on 60 body dataset.

1 Major Changes

We are planning to push to submit an extended abstract to the International Conference on Learning Representations (ICLR) 2019 Structure and Priors in Reinforcement Learning (SPiRL) workshop, which has a submission deadline of March 7, 2019. We decided to push for this deadline after evaluations completed showing that our models were significantly better than the baselines, both in learning speed and generalizability (see Figure 1).

2 What You Have Accomplished Since Your Last Meeting

As mentioned before, the long-running full evaluations have completed. Additionally, over break, I thought about the contributions of this work and its place in the literature and fleshed out quite a bit of the paper, writing down the intuitions behind our representations as well as our model architectures and training strategies. Since the semester started, I also spent some time building out tooling for visualizing learned policies and datasets, as well as generating plots from saved evaluation data and thought about how to best show the results we want.

3 Surprises

Early this week, we decided to quantify the difficulty of the 60 body training dataset (made up of variations on the original canonical walker body) we were using to train the bulk of our models—we would randomly sample a body from this dataset, run a training episode on it, and repeat—by training models on a single, canonical body, and evaluating those models' performance on the 60 body training dataset. Unfortunately, we found that the baseline model with the exact same architecture is able to learn to walk fairly well (average 80 reward per episode over all bodies, 300 reward per episode is considered "solved") on the 60 body training dataset in 100-200x fewer gradient steps. Of course, it has no chance of achieving more reward than that, but the fact that it was able to do so well on the 60 bodies despite never having seen them while the baseline model that was trained on those bodies took so much longer to show signs of improvement is a very strong sign that the training method of sampling a random body every training episode is very bad.

These results are certainly notable and will be written up, but we will have to think about how best to train models moving forward and how to best structure tasks to show off the advantages of our novel representations.

4 Looking Ahead

Over the next two weeks, I plan to immediately implement a way to add a similar amount of spatial information to the baseline models so that the baselines and our models both have the same amount of information. I will also begin to train new baseline models in this fashion, and continue to think about a better training strategy. Finally, I will evaluate the difficulty of several smaller, handcrafted (10-12 bodies) training datasets by running the canonical body baseline on them.

5 Revisions to Your Future Milestones

We have approxmiately achieved the 75% project milestone—as we are aiming for the earlier March 7 deadline (in about a month), I intend to focus more on fleshing out results on the bipedal walker bodies rather than splitting between those and the raptor walker body. If time allows, of course, I wish to explore a bit towards the raptor body, but for the SPiRL extended abstract, I would prefer to have a smaller but more solid body of results.

In this light, I will revise the next project milestone to be completing evaluations of the baseline models with the added positional information, as well as having figured out a better way to train models. I also would revise future milestones by focusing less on different bodies, as once we complete results for the bipedal walker body getting results for the raptor body (and centipede body, if I decide to explore that) will be very similar.

6 Resources Needed

I have all of the resources I need.