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

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

November 2, 2018

1 Project Description

I will be working with Professor Ruslan Salakhutdinov and his graduate student Emilio Parisotto in the Machine Learning Department on a new spatial state representation for deep reinforcement learning applied to robot walkers in simulation.

The problem we aim to solve is that existing state representations for robot control are often just a vector of joint velocities or positions or angles, which does not capture the position of limbs in relation to each other. A vector of scalars is a fairly basic state representation and not particularly expressive. It is our hope that adding a spatial inductive bias in the state representation should aid policy learning and policy robustness. Intuitively, it makes sense that a richer state representation incorporating a spatial factor would improve learning, since animals and humans depend on more than just their knowledge of the positions and rotations of their joints to learn motor skills. In particular, athletes depend on their knowledge of the relative positions of their limbs to execute complex maneuvers.

We introduce and experiment with several different ways of encoding relative position of body parts, most notable of which is drawing a grid centered on the agent and projecting joint and body part information into the appropriate grid cell to form a new state representation. This grid-state representation allows us to replace the traditional multi-layer perceptron (MLP) neural network architecture with convolutional neural networks (CNNs), which have been traditionally used in image processing and computer vision and are capable of capturing local and global dependencies in two- and three-dimensional spaces. We hope to demonstrate the viability of this grid state representation and to show improved model learning and performance over the existing vector state representation on the same environments and new custom environments tailored to our new representation.

If we are able to obtain favorable results, we will have introduced a new state representation to the field of deep reinforcement learning for robotic control that would be shown to work better than the existing vector state representation, at least on our targeted tasks. This state representation could be used in appliactions and tasks beyond simple locomotion—certainly, our grid representation could also be applied to robotic manipulation and other complex tasks.

Most of our work will be done using the Box2D physics engine, provided as a reinforcement learning environment through OpenAI Gym, on custom variants of the BipedalWalker-v2 environment and agent body, as well as custom bodies including a raptor body. We also hope to move into more complex three-dimensional physics simulation environments, such as MuJoCo or Bullet.

Major challenges of this project will include adapting existing reinforcement learning code to our particular uses, building new environments and robot bodies, and the common difficulties in training deep models where it is not clear whether the models are incapable of learning the task or if we have just not discovered the right hyperparameters. It will also be challenging comparing model performance besides learning speed, since there is very little middle ground between models that are able to solve the walking task, and models that are completely unable to solve the walking task.

2 Project Goals

2.1 75% Project Goal

- Determine whether grid state representation increases training speed on simple/standard bodies, and policy robustness to different body shapes or variations on a canonical body.
- Successfuly train a single model for control on many variations on a canonical body shape, with same number of body parts.
- Achieve zero- or few-shot transfer to bodies with same number of body parts but different body part dimensions.
- Targeted bodies: BipedalWalker

2.2 100% Project Goal

- Determine whether grid state representation increases training speed on simple/standard bodies, and policy robustness to different body shapes or variations on a canonical body.
- Successfuly train a single model for control on many variations on a canonical body shape, with same number of body parts.
- Achieve zero- or few-shot transfer to bodies with same number of body parts but different body part dimensions.
- Successfully train a single model for control on many variations on a canonical body shape with different number of body parts.
- Achieve zero- or few-shot transfer to bodies with unseen numbers of body parts.
- Targeted bodies: BipedalWalker, RaptorWalker, CentipedeWalker
- Show above results or similar in 3D environments.

2.3 125% Project Goal

- Determine whether grid state representation increases training speed on simple/standard bodies, and policy robustness to different body shapes or variations on a canonical body.
- Successfuly train a single model for control on many variations on a canonical body shape, with same number of body parts.
- Achieve zero- or few-shot transfer to bodies with same number of body parts but different body part dimensions.
- Successfully train a single model for control on many variations on a canonical body shape with different number of body parts.

- Achieve zero- or few-shot transfer to bodies with unseen numbers of body parts.
- Targeted bodies: BipedalWalker, RaptorWalker, CentipedeWalker, DogWalker, HumanoidWalker
- Show above results or similar in 3D environments.
- Experiment with more complex model architectures and transfer techniques such as policy similarity in a shared feature space or learned walk-cycle embeddings. This may be reaching into the scope of a second paper.

3 Project Milestones

3.1 First Technical Milestone

As this research is already in progress and some preliminary trial evaluations have been completed, the first technical milestone will be completing rock-solid evaluations and have publishable results on the BipedalWalker environment showing our state representation is better or at least comparable to the state-of-the-art, i.e. the 75% project goal. By this time, we should have shown some good transfer results on BipedalWalker.

3.2 First Biweekly Milestone: February 1st

By the first biweekly milestone, I hope to have preliminary results for RaptorWalker and CentipedeWalker suggesting that our state representation can be applied to RaptorWalker and CentipedeWalker for comparable performance as the state-of-the-art.

3.3 Second Biweekly Milestone: February 15th

By the second biweekly milestone, I hope to have rock-solid, publishable results for RaptorWalker and CentipedeWalker that models trained with our state representation are comparable to or better than the state-of-the-art. By this time, we should have shown good transfer results on RaptorWalker and CentipedeWalker, and have obtained some preliminary results on DogWalker and HumanoidWalker or begun to construct bodies and environments for three-dimensional robot control. Here, I also want to begin revisions on the paper and moving towards a final submittable draft with Emilio and Professor Salakhutdinov.

3.4 Third Biweekly Milestone: March 1st

By the third biweekly milestone, I hope to have completed evaluations and have publishable results for DogWalker and HumanoidWalker, or have some preliminary results on simple three-dimensional environments.

3.5 Fourth Biweekly Milestone: March 22nd

By the fourth biweekly milestone, I anticipate to be still working on the three-dimensional environments but should have a more-or-less submittable final paper draft with the 100% project results.

3.6 Fifth Biweekly Milestone: April 5th

By the fifth biweekly milestone, I hope to have publishable results on 3D and a more-or-less satisfactorily completed research project. Depending on whether we decide to pursue this project further or begin work on a new research project, I hope to begin building models and systems for more complex transfer techniques in this setting and/or begin finalizing my next paper idea with Professor Salakhutdinov.

3.7 Sixth Biweekly Milestone: April 19th

By the sixth biweekly milestone, I hope to have begun training models with the additional transfer techniques and started a literature survey for my next research project which will likely become my undergraduate senior thesis.

3.8 Seventh Biweekly Milestone: May 3nd

By the seventh and final biweekly milestone, I hope to have completed the first round literature survey for my next research project and begun writing code and training preliminary models. Also, if we chose to pursue this paper further, I hope to have preliminary results for the additional transfer techniques.

Note: since I intend to be working on this new paper idea over the summer, it is possible that if we choose to pursue this paper or a follow-up paper, I will likely be able to commit to these two research projects full-time.

4 Literature Search

As of this writing, I have completed the first draft of the "Introduction" and "Related Work" sections of the paper after a preliminary literature search. There are a couple notable papers from which we draw inspiration, which are cited at the end of this paper. [1] [2] [3] Some other papers with interesting or thought-provoking results that I have encountered are also cited below. [4] [5] [6] [7] [8]

5 Resources Needed

We will be using Python and the deep learning framework Pytorch and the reinforcement learning environment OpenAI Gym, all of which we have successfully set up in an Anaconda environment to be used on remote cluster machines. Also, we intend to use open-source project TensorboardX to use Tensorboard logging with Pytorch, which I have also recently set up.

To complete this research project, nontrivial computing power will be required, preferably on GPU machines. To this end, Emilio has given me a ssh login to his MLD `lithium` cluster account for training models and running evaluations on the GPU machines there. I anticipate this should be more than sufficient.

References

[1] Tingwu Wang, Renjie Liao, Jimmy Ba, and Sanja Fidler. Nervenet: Learning structured policy with graph neural networks. In *International Conference on Learning Representations*, 2018.

[2] Tao Chen, Adithyavairavan Murali, and Abhinav Gupta. Hardware conditioned policies for multi-robot transfer learning. In *Neural Information Processing Systems*, 2018.

[3] Hexiang Hu, Liyu Chen, Boqing Gong, and Fei Sha. Synthesize policies for transfer and adaptation across tasks and environments. In *Neural Information Processing Systems*, 2018.

[4] Coline Devin, Abhishek Gupta, Trevor Darrell, Pieter Abbeel, and Sergey Levine. Learning modular neural network policies for multi-task and multi-robot transfer. In *International Conference on Robotics and Automation*.

[5] Abhishek Gupta, Coline Devin, Yuxuan Liu, Pieter Abbeel, and Sergey Levine. Learning invariant feature spaces to transfer skills with reinforcement learning. In *International Conference on Learning Representations*, 2017.

[6] Lerrel Pinto and Abhinav Gupta. Learning to push by grasping: Using multiple tasks for effective learning. *CoRR*, abs/1609.09025, 2016.

[7] Maruan Al-Shedivat, Trapit Bansal, Yuri Burda, Ilya Sutskever, Igor Mordatch, and Pieter Abbeel. Continuous adaptation via meta-learning in nonstationary and competitive environments. *CoRR*, abs/1710.03641, 2017.

[8] Joshua Tobin, Rachel Fong, Alex Ray, Jonas Schneider, Wojciech Zaremba, and Pieter Abbeel. Domain randomization for transferring deep neural networks from simulation to the real world. *CoRR*, abs/1703.06907, 2017.