Amy Chou <amyachou>
Rory Lipkis <rlipkis>
Victoria Tsai <vitsai>

Run. Reward. Repeat.

Training Musculoskeletal Models with Deep Deterministic Policy Gradients

Abstract

We apply deep reinforcement learning to musculoskeletal models with the objective of learning to run. Using the actor-critic deep deterministic policy gradient method, we learn on successive observations drawn from the simulated openAI gym, and generate a policy that maps high-dimensional observation states to muscle excitation vectors.

Methodology

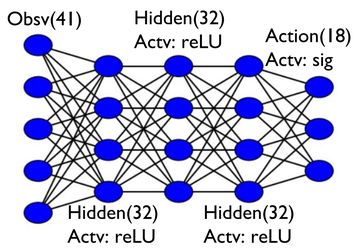
Deep Deterministic Policy Gradient (DDPG)

Off-Policy; Model-Free
Deterministic Policy
Continuous Action Space

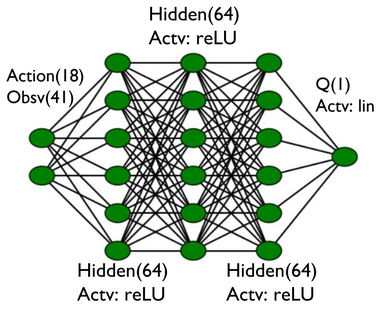
Observations:
41 dimensions
OpenAI Gym

Actions:
18 dimensions
Muscle Excitations

Actor: $\pi(o_t)$



Critic: $Q^{\pi}(o_t, a_t)$

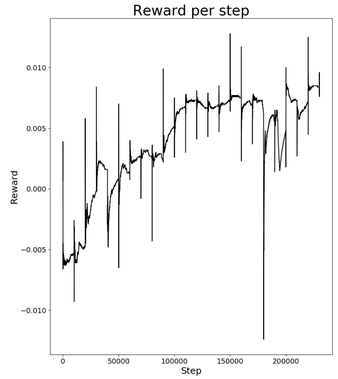


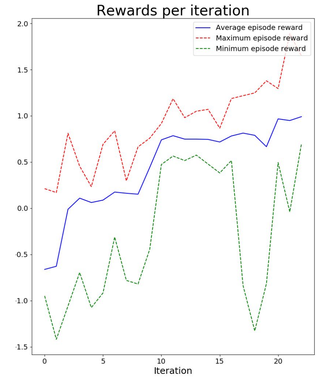
Deterministic Policy Gradient Theorem:

$$\nabla_\theta J(\pi_\theta) = \mathbb{E}_{s\sim\rho^\pi, a\sim\pi_\theta}\left[\nabla_\theta \log \pi_\theta(a|s) Q^w(s,a)\right]$$

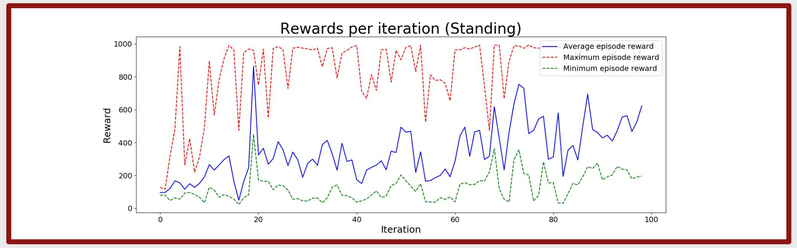
Results

We trained the agent with a series of trials ranging from 100,000 steps to 500,000 steps, computing statistics on reward per episode every iteration of 10,000 training steps. We see that the rewards per episode increased over the number of iterations trained and resulted in the agent's ability to take a step toward its goal after 200,000 training steps.





Furthermore, in order to demonstrate the applicability of DDPG to high-dimensional musculoskeletal models, we also trained the agent to stand using a reward function based on the amount time it stayed upright. After one million iterations, the agent was able to stay balanced for at least 3,000 simulation steps while it stood. In the training of the standing agent, the reward was capped at 1,000 per episode.



Discussion

The agent successfully learned to take stable steps towards its goal, although it did not achieve long-term rapid locomotion in the allocated training time.

It also demonstrated the ability to hold and balance in a position for an arbitrarily long period of time (at least 3000 simulation steps).

The primary obstacle to progress was the heavy computation required to simulate the precise physics of the agent, independent of optimizations made to the learning algorithm.

The project would be best extended by gaining access to a high-performance cluster for a long period of time.

References

1. Silver, David, Guy Lever, Nicolas Heess, Thomas Degris, Daan Wierstra, and Martin Riedmiller. "Deterministic policy gradient algorithms." In *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, pp. 387-395. 2014.

2. Reinbolt, J, Seth, A, Habib, A, Hamner, S. osim-rl (2017). GitHub repository. https://github.com/stanfordnmbl/osim-rl