Project Proposal: MuJoCo Basketball

Authors. Ijay Narang (in5787@princeton.edu), Windsor Nguyen (mn4560@princeton.edu), Caleb Shim (calebs@princeton.edu)

Do you need GPUs? Yes.

Project type. Applying RL to a new problem

Description Our task is to use reinforcement learning to teach a figure in a MuJoCo environment to shoot a basketball through a hoop. We would like to see how the human shoots the basketball and what form it achieves to eventually get the ball through the basket. Reinforcement learning is an appropriate solution because the humanoid must take a series of sequential decisions (leg movement, arm movement etc) to learn how to best shoot the ball at the hoop. This is indeed a continuous state/action MDP (where the states are based on position of the ball/limbs/humanoid) and the actions are based on the joints (torque on hinges) that you move within the humanoid. We may have to do some reward shaping to get good shooting behavior (maybe something based on the trajectory of the ball after a movement), but this is unknown as of now. This type of problem clearly can't be solved by bandits because (1) the complexity of the above parameters, (2) the fact that we need sequential decision making that are dependent on each other (the way a shooter shoots the ball is based on the decisions first taken with regards to legs, arms, hands etc), and (3) we likely need to be clever about the reward function as a simple binary function based on make or miss will not work, and we need to somehow take arc, final destination and trajectory of ball into consideration.

Resources

1. MuJoCo: https://mujoco.org/
2. Humanoid Environment: https://www.gymlibrary.dev/environments/mujoco/humanoid/

Challenges

1. Modifying the environment/humanoid
   - From a conceptual standpoint: In reality, finger placement on the ball is incredibly important for a consistent shooting form (from personal experience). However, to do implement hands with fingers may or may not be a difficult task. We would need to figure out the physics for hands/fingers.
   - From a practical standpoint: Defining all of these properly and modifying the XML file accordingly
2. Reward Shaping with regards to what a better trial is. We can't do binary (0/1) rewards as that would not encourage improvement. Furthermore, a scheme such as L2-norm wouldn't work as we need to encourage arc as well.
3. The RL algorithm we need to use is also a critical design challenge. Currently, we think that we need a powerful continuous RL model which is also efficient. One thought was just copying the DDPG implementation from a previous homework. We also thought that applying PPO or TRPO black box may work, but need to think about these further (and would probably discuss with course staff)

Team We will work together for every portion of the task from problem formulation, production, and analysis. We plan to meet once every week.