Proximal Policy Optimization with Dynamic Clipping

Student: Rishikesh Vaishnav
Mentor: Sicun Gao

August 13, 2018

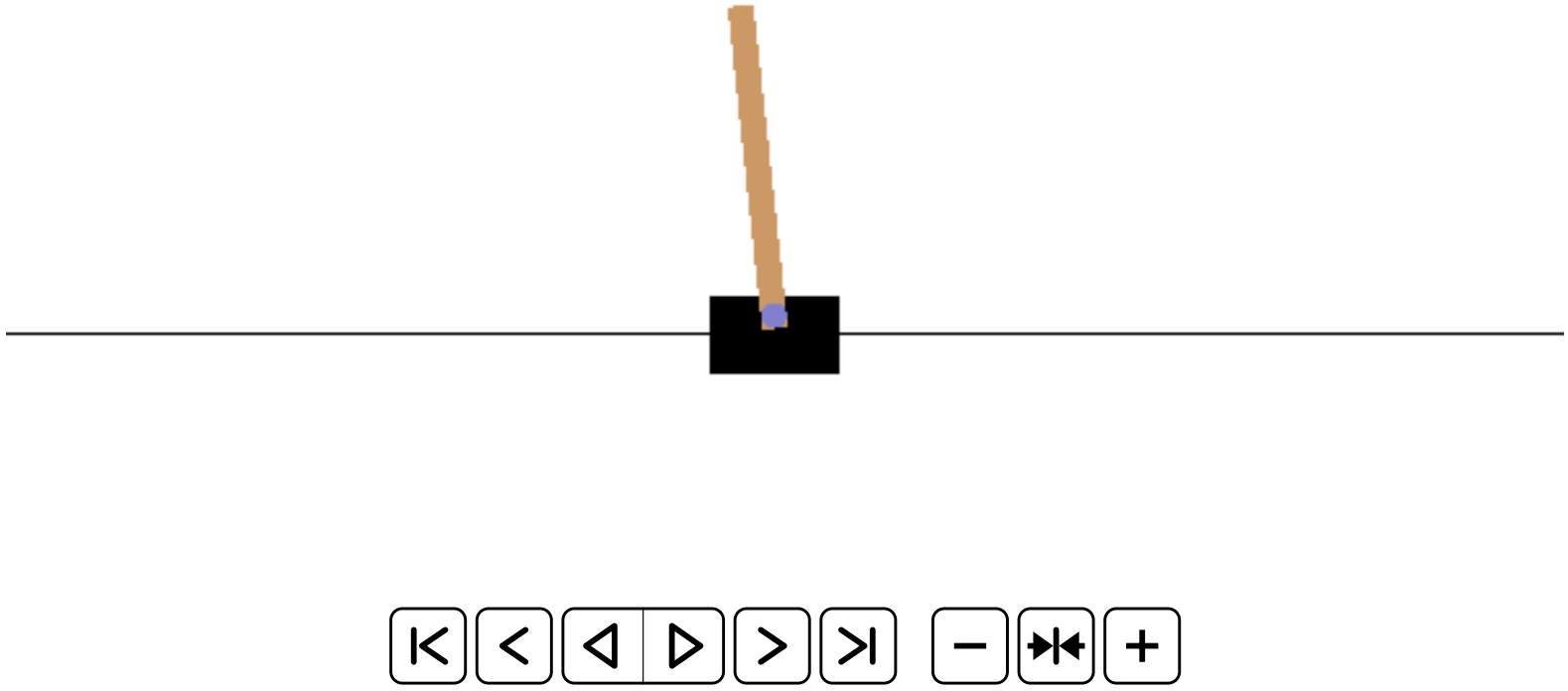
Introduction

Reinforcement Learning

- A general algorithmic technique that seeks to replicate behavioral learning.
- Basic vocabulary:
  - **Environment**: a general setting with changeable parameters in which actions can be performed that affect these parameters
  - **State**: a specific configuration (i.e. "snapshot") of an environment
  - **Agent**: an entity that learns to accomplish a task in a specific evironment
  - **Action**: a decision made by the agent that is intended to affect subsequent states
  - **Episode**: a sequence of states and actions in an environment
  - **Reward**: a number associated with a state-action pair
- Overall goal: train an agent that picks actions such that the sum of the rewards over an episode is maximimized.

Introduction (contd.)

– Example: cart-pole demo



Introduction (contd.)

Trust Region Policy Optimization Proximal Policy Optimization

Potential Shortcomings of PPO

Idea

Results

Future Directions