Proximal Policy Optimization
A Reinforcement Learning Algorithm

Reinforcement learning algorithms in their entirety and policy gradient methods especially can be difficult to understand and therefore implement. Tutorials that explain them easily, are rare and most base on a lot of prior knowledge regarding the topic and mathematics in general. With this work we strive to provide an easy tutorial and explanation of the popular reinforcement learning algorithm Proximal Policy Optimization (PPO), how it works and what is needed to implement it.

To introduce PPO we briefly talk about the original paper of PPO and some related work, which has been done since, to address flaws in the original algorithm. We also discuss the environment Open AI Gym briefly, explaining its usage. Afterwards we briefly introduce reinforcement learning in itself, to then mention PPOs predecessors and the development from those into PPO.

After we give an more in depth view on PPOs attributes, especially concentrating on the clipping loss, actor-critic model and the probability distribution of continuous action spaces. Finally we conclude this paper with a brief explanation of the pseudocode for PPO.