CMP9137M: Advanced Machine Learning Lecture 7: Deep Reinforcement Learning 2

Attendance: 437288

Heriberto Cuayahuitl

20th April 2022

Reinforcement Learning

An agent must find its behaviour via its interaction with an environment, this can be p or virtual environment, including humans, depending on if this is a learning task or carrying out an objective. The task being to find the best behaviour possible for accomplishing the objectives. This requires us to equip our machine with a brain, this brain is called a policy. The policy is defined in a variety of ways and often includes a “reward”. In reinforcement learning we are trying to maximise the rewards the machine will get through interacting with its environment.

For assignment:

Try 3 different methods each of them with a different exploration method.

Episode = initial state to final state (iterations)

Double Deep Q-Networks

Green- estimate, purple – true values

Need to change the equation to use the online and target weights

DQN w/Prioritized Experience Replay (DQN-PER) ref. 3

We need a way to show that one learning experience is better than another because it can lead to better rewards

DQN – Only applies to problems with discrete actions

PPO can apply to any problems, discrete or continuous actions - it generally performs better than many other algorithms. This is possible because PPO works with probability distributions.