DRL Homework 03

Leon Schmid

Deadline: July 3rd 23.59

Abstract

Homework 3 is your first actual Deep Reinforcement Learning Homework: You are asked to implement DQN.

Contents

1 Task 01 - Experience Replay Buffer

Implement the experience Replay buffer. Remember: The experience replay buffer is a 'fifo' buffer - meaning first in first out. It should have a maximum size (e.g. 100.000, in many actual implementations 1.000.000). When an element is added to the buffer and this this maximum size is reached, the buffer is supposed to forget the oldest element (which is the reason for the name - the first element written into it is first forgotten, etc.). Each element should be a tuple (s, a, r, s'). I recommend (and for eligibility for bonus points expect) wrapping the experience replay buffer into a TF dataset for training on it.

2 Task 02 DQ Network

You are tasked to implement DQN for the Lunar Lander Environment:
https://www.gymlibrary.ml/environments/box2d/lunar$_l ander$/

Use Tensorflow to implement an appropriate Neural Network for the task. Hint: In DRL Neural Networks, especially for such toy tasks, can be very small! Your Network Implementation should be parameterized with the appropriate input (check the observation space of the environment) size and shape and the

appropriate output size (check the action space). For eligibility for Bonus Points this parameterization should simply take the environment as an argument (and automatically check the action and observation spaces to create the appropriate network), additionally your network should use a functional forward step (wrapping in @tf.function or writing as a functional Model).

3 Task 03 DQN Training

Implement the necessary training algorithm around your Network and ERB. You will need to implement some sort of delay (delayed target network or Polyak-Averaging for the target network). Notice you will need to run this algorithm for quite some timesteps to have it work - generally DRL algorithms run much longer as compared to typical supervised Learning tasks. Finally Implement some evaluation (at least a graph of the average return per run) of your algorithm. For eligibility for Bonus points, also implement evaluation via inspection of behaviour by saving a video of the agents behaviour in regular intervals during training.

4 Task 04 Review

Again, find 3 other groups for which you can review homework 02. Include the written review in the folder you submit.