DRLND - Continuous Control Project Report

Fourat Thamri

March 2021

1 DDPG Actor and Critic Models Specifications

The deep RL method used to train the agent is a Deep Deterministic Policy Gradient with experience replay and added noise like the one learnt from the Udacity 3rd course module. Experience replay is a method that saves agent experience in batches and then learn from this experience in a random way rather than learn from ordered episode sequences. This technique is proven to improve the learning process a lot.

We used the second version of the environment that includes 20 agents acting running and aggregating experience in the same time.

The Actor and Critic networks used are fully connected ANNs consisting of 2 hidden layers with 128 nodes each. The table below lists all the hyperparameters used to achieve the results presented in the next section. The model parameters can also be found in the file agent.py and the associated jupyter notebook.

Parameters	Value	Note
Actor and Critic Hidden Layers nb	2	-
Nodes nb	128	-
NN Batch size	512	-
Learning Rate	0.001	-
Replay Buffer size	100000	-
Update Frequency	12	How often to update the network
Tau	0.001	For soft update of target parameters
Gamma	0.99	Discount factor

Table 1: Hyperparameters list

2 Results

The model learned its goal well and efficiently by solving the task in 104 episodes. The task is considered solved when the agent achieves a mean score of >30 over

100 consecutive episodes. The plot below shows the reward history of the agent.
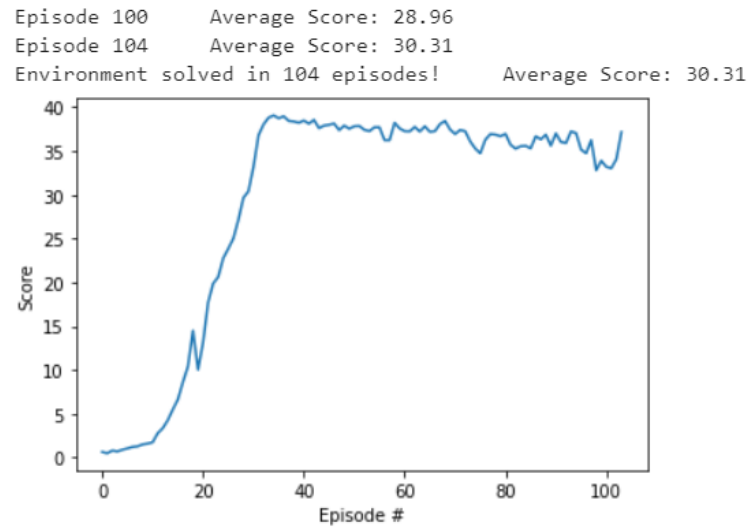

Figure 1: Training scores plot

3   Future work

To further improve the model, a longer training is needed along with a proper optimal hyperparameters search. A next step would be to improve sample efficiency by using the generalized advantage estimation approach.