

PROJECT 3: TENNIS

P.M. Lucaci

DEEP REINFORCEMENT LEARNING NANODEGREE, Udacity

Contents

Learning Algorithm	2
Hyperparameters.....	2
NN Model Architecture.....	2
Plot of Rewards	3
Ideas for Future Work	4
Improving the report	4
Including project details.....	4
GIF of Trained Agent	4
Action Space.....	4
State Space.....	4
Rewards.....	4
Solving the Environment.....	4
Documenting Future Improvements of Learning Algorithm	4
Add Bayesian Optimization.....	4
Add Prioritised Replay Buffer.....	4
Add Progress Bar	4
Add TensorBoard	4

Learning Algorithm

DDPG stands for “Deep Deterministic Policy Gradient”, which is an algorithm learning concurrently a Q-function and a policy. This algorithm, with a shared Replay Buffer, has been implemented to train and evaluate the two agents involved in this environment.

Hyperparameters

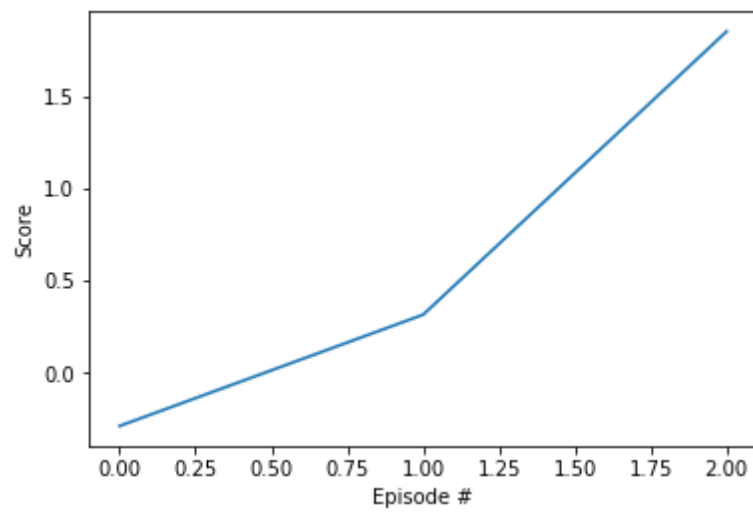
```
n_episodes = 30000 # maximum no. of episodes to train
max_t = 2000 # maximum no. of time steps
num_agents = 2 # no. of agents
random_seed = 0 # random seed
eps_start = NaN # epsilon upper limit (before any decay)
eps_end = NaN # epsilon lower limit (minimum value)
eps_decay = NaN # epsilon decay rate
BUFFER_SIZE = int(1e6) # replay buffer size
BATCH_SIZE = 512 # minibatch size
GAMMA = 0.99 # discount factor
UPDATE_EVERY = 4 # how often to update the network
TAU = 1e-3 # for soft update of target parameters
LR_ACTOR = 1e-4 # learning rate actor
LR_CRITIC = 5e-4 # learning rate actor
```

NN Model Architecture

ACTOR	BATCHNORM1D	LINEAR	ReLU	LINEAR	ReLU	LINEAR	TANH
	IN = OUT = STATE_SIZE	IN = STATE_SIZE OUT = 128		IN = OUT = 128		IN = 128 OUT = ACTION_SIZE	
STATE_SIZE = 8				ACTION_SIZE = 2			

CRITIC	BATCHNORM1D	LINEAR	ReLU	LINEAR	ReLU	LINEAR
	IN = OUT = STATE_SIZE	IN = STATE_SIZE OUT = 128		IN = 128 + ACTION_SIZE OUT = 128		IN = 128 OUT = 1
STATE_SIZE = 8				ACTION_SIZE = 2		

Plot of Rewards



Episode 3 Average Score: 0.639
Environment solved in 3 episodes

Ideas for Future Work

Improving the report

Including project details

GIF of Trained Agent

Action Space

State Space

Rewards

Solving the Environment

Documenting Future Improvements of Learning Algorithm

Add Bayesian Optimization

Add Prioritised Replay Buffer

Add Progress Bar

Add Tensor Board