Multi Agent Continuous Control Project Report

Introduction

The reader is assumed to have basic knowledge on Reinforcement Learning and Deep Learning algorithms in particular DDPG and multi-agent reinforcement learning.

This project aims at training two agents to play tennis. This observation state is a 8-dimensional space corresponding to the position and velocity of the ball and racket. The action space is a 2-dimensional continuous space corresponding to jumping and movement toward (or away) from the net. The task is considered solved when the agents obtain an average reward of +0.5 over 100 consecutives episodes.

The specificity of the **DDPG algorithm**[1] will not be described in this report, the link to the related paper is given at the end of the page.

Future Work

- Try different architecture and other hyperparameters to reach a better score in less time.
- Implement other algorithms used in multi-agent reinforcement learning.
- Stay in touch with the news in the reinforcement learning domain.

Method

I started by changing things in the architecture model itself. I used only one neural network for the actor and one neural network for the critic. Hence each agent is training the same models. Then I tried exploring various hyperparameters that led me to observe that training was unstable. It led me to add Batchnorm and to modify ReLU functions by Leaky ReLU to the models and to reduce the noise to finally resolve the task.

[1] https://arxiv.org/pdf/1509.02971.pdf

Implementation Characteristics

	MADDPG
Memory Size	1 000 000
Batch Size	128
Discount Factor	0.99
Rate of Transfer for Soft Update	1e-3
Frequency of Update of The Target Network	Every time steps
Max steps in one episode	Max as possible
Ornstein-Uhlenbeck Process Parameters	Mu=0 Theta=0.15 Sigma=0.05
Actor Architecture	Linear(24,128) ReLU Linear(128,128) BatchNorm1d(128) ReLU Linear(128,128) BatchNorm1d(128) ReLU Linear(128,2)
Critic Architecture	Linear(48,128) Leaky ReLU (Here output and action vectors are concatenated) Linear(128+4,128) BatchNorm1d(128) Leaky ReLU Linear(128,128) BatchNorm1d(128) Leaky ReLU Linear(128,1)
Learning Rate Actor	1e-3
Learning Rate Critic	1e-3

Results

MADDPG Agent average score over 100 episodes

```
Episode 100     Average Score: -0.00
Episode 200     Average Score: 0.010
Episode 300     Average Score: 0.03
Episode 400     Average Score: 0.05
Episode 500     Average Score: 0.05
Episode 600     Average Score: 0.08
Episode 700     Average Score: 0.09
Episode 793     Average Score: 0.51
```