Robin Stringer

P3: Competition and Collaboration

Introduction

In this project, two agents control tennis rackets to hit a ball over the net. The agents receive a reward of +0.1 for doing so, and a reward of -0.1 should the ball hit the ground, net or go out of bounds. The primary goal of the agents is to keep the ball in play.

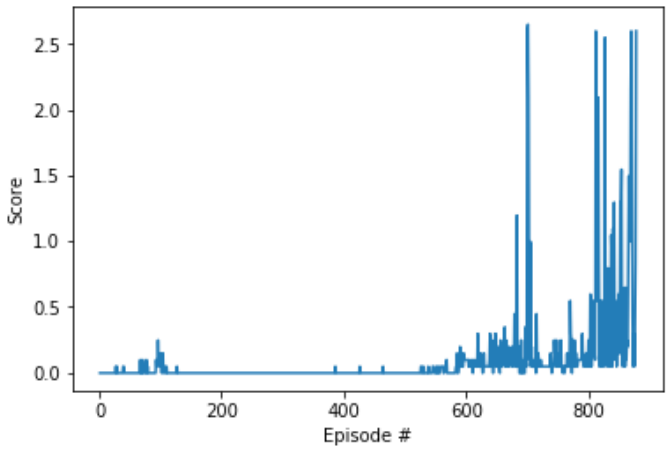
This episodic task is considered solved after the agents achieve an average score of +0.5 over 100 consecutive episodes.

Algorithm

While Deep Q Networks can solve problems with high-dimensional observation spaces, they struggle beyond discrete and low-dimensional action spaces. The deep deterministic policy gradient, DDPG, is a model-free, off-policy actor-critic algorithm using deep function approximators able to learn policies in high-dimensional, continuous action spaces. This approach was first outlined by Google Deepmind authors in the 2016 paper '*Continuous Control With Deep Reinforcement Learning*' (https://arxiv.org/pdf/1509.02971.pdf).

The Actor network comprises a hidden layer of 400 nodes, a RELU activation function, followed by a batch normalization function before a second hidden layer of 300 nodes with a further RELU activation. The Critic features the same design, with a gradient clip added. The agents use the same actor network to select actions and use a shared replay buffer to learn from the experience.

Training DDPG and multi-agent DDPG networks can bring unstable and unpredictable results, and this project was a case in point. Improvements were flat until around episode 600, four-fold gains around episode 625-675. The learning then flattened again before rising around episode 800.



Further work

Multi-agent collaboration is one of the most exciting areas of deep reinforcement learning and I am keen to explore the subject further. I would like to compare similar approaches such as

PPO to DDPG in a multli-agent environment. I also hope to try an approach similar to the one taken for this project on a more complex task with more agents, such as soccer.