Udacity Deep Reinforcement Learning
Project Navigation

Rohil Pal (rohilpal9763@gmail.com)

Abstract

In this project, we have to train an agent to navigate (and collect bananas!) in a large, square world. This environment is provided by Unity Machine Learning agents (ML-agents). We are free to use any value-based methods that were taught in the course.

1 Environment

The **state space** has 37 dimensions each of which is a continuous variable. It includes the agent's velocity, along with ray-based perception of objects around the agent's forward direction.
The **action space** contains the following 4 legal actions:

- move forward (0)
- move backward (1)
- turn left (2)
- turn right (3)

A reward of +1 is provided for collecting a yellow banana, and a reward of -1 is provided for collecting a blue banana. Thus, the goal of your agent is to collect as many yellow bananas as possible while avoiding blue bananas.

The task is **episodic**, and in order to solve the environment, your agent must get an average score of +13 over 100 consecutive episodes.

2 Method

Two variants of DQNs are used to solve this problem.

1. *Vanilla DQN* with *Experience Replay* and *Fixed Q-targets*.
2. *Double DQN*

3 DQN Architecture

It has 3 fully connected layers each 150 neurons. Each of them is followed by a $ReLU$ activation function. To avoid overfitting, *dropout* with a probability of 0.5 is used after every fully connected layer except the last layer.

The network accepts a tensor of dimension 37 which is the dimension of each state and outputs a tensor of 4 dimension which is the number of actions an agent can perform at each state.

4 Hyperparameters

4.1 Vanilla DQN

- Replay buffer size : e^5
- Batch Size : 64
- Discount factor ($gamma$) : 0.99
- Soft Update parameter (TAU) : e^{-3}
- Learning rate ($alpha$) : e^{-3}
- Frequency of network update : 50

4.2 Double DQN

- Replay buffer size : e^5
- Batch Size : 32
- Discount factor ($gamma$) : 0.99
- Soft Update parameter (TAU) : e^{-3}
- Learning rate ($alpha$) : e^{-4}
- Frequency of network update : 4
- Frequency of target network update : 100

5 Future work

Use of advanced DQN variants like *Prioritised Experience Replay* and *Dueling DQN*.