

DQN implementation for Banana collection

Jonas Stiernström

I wanted to get a baseline for the task so went for a basic DQN with Experience learning. After playing with the hyperparameters I found a sweet spot allowing me to succeed at the task at around 594-694 episodes. With this baseline ready, I could go on and test more advanced methods and compare them to this first test.

Algorithm

The algorithm as mentioned was a basic DQN with experience learning. The agent and model used was taken from the Reinforcement Learning DQN course at Udacity and in the end not modified too much (as the results suffered). The model consists of 3 layers, the first taking as input the size of the state. The second layer takes as input the output of the first layer, and the third layer takes as input the second output. The output of the last layer is the action_size.

DQN for Banana Collection

Meaning: The input will be our states, and the output will be our action values. The model does not use dropout.

The agent then interacts with the environment, feeding the available states through the network and using the received action values to make decisions on how to act (using the epsilon values to prevent an entirely greedy outcome).

The agent will basically follow this scenario, repeatedly, to learn and improve: act on current state -> receive updates -> save experience in experience buffer -> repeat

Hyperparameters were chosen with trial and error. For example these experiments were made:

Test 1: LR $5e-4$ as in the QNetwork course - success on the task after 594-694 episodes

Test 2: LR $5e-2$ - learning was an issue and an average of even 2 was never reached.

Test 3: LR $5e-5$ - Learning was made, but took much longer than the first test. Approx. 1300 episodes

Test 4: LR $6e-4$ - no learning at all. It barely reached an average of 1.

Test 5: LR $6e-4$ this worked, but learning was even slower than Test 3.

Multiple tests like these were done for the hyperparams and in the end the ones chosen seemed to give me the best results.

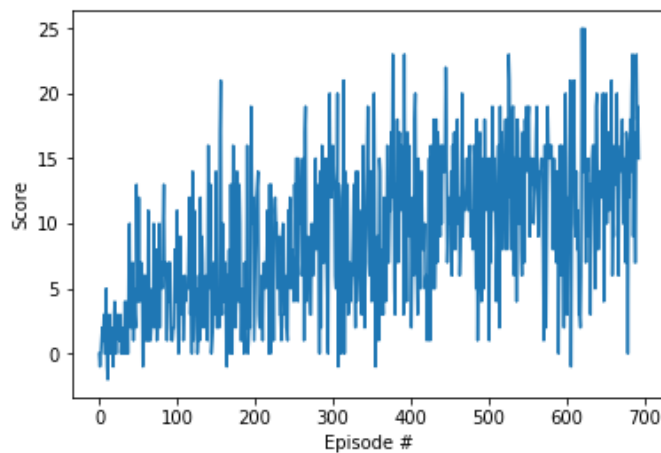
DQN for Banana Collection

Result

The result was in the end good (594-694 episodes). And here's some data:

Episode 100	Average Score: 3.45
Episode 200	Average Score: 5.80
Episode 300	Average Score: 8.03
Episode 400	Average Score: 9.73
Episode 500	Average Score: 11.12
Episode 600	Average Score: 12.71
Episode 694	Average Score: 13.05

Environment solved in 594 episodes! Average Score: 13.05



Improvements

Future improvements could definitely be made. And since a basic DQN was used, it could be worth trying a Double or Dueling DQN to see how the avoidance of over optimistic values in the Double DQN would change the results. Or the Aggregation of a dueling version. They both come with their own set of improvements and should be worth investigating for this particular challenge.

DQN for Banana Collection