Report Navigation

Author: Dorzhi M.

Approach description

We apply Deep Q-Learning to solve banana collection problem. The approach we applying is based on '[Human-level control through deep reinforcement learning]' (Mnih et al.)

The environment's state space has dimension of 37, actions space has dimension of 4.

The following neural net is used for parametrizing Q function:

- 37-dimension input layer
- 2 hidden layers with dimension of 64
- 4-dimension output layer

We set following hyperparameters:

- maximum number of training episodes (by default 2000)
- maximum number of timesteps per episode (by default 2000)
- starting value of epsilon, for epsilon-greedy action selection (by default 2000)
- minimum value of epsilon (by default 2000)
- multiplicative factor (per episode) for decreasing epsilon (by default 2000)

In order to solve the environment, the agent must get an average score of +13 over 100 consecutive episodes.

Agent Performance

We were able to solve the problem in 372 episodes.