Report

Neural Network Graph Definition:

Standard 3 layers network (input, hidden, output) with 128 hidden unit was defined to engage with environment observation. Weights were initialized by random normal initializer with mean=0 standard deviation=0.3. Bias were initialized with 0.

By experiment, AdamOptimizer with default learning rate was decided to use.

Replay Experience and batching:

Experience replay buffer was designed by a list with size 2000. After the size of buffer is over the batch size (32), old experience data will be used as fixed target Q values for training, in order to improve the learning speed.

By experiment, classify and distribute the high and low reward actions can improve the learning performance.

Low reward will be distributed at the front-part of experience replay buffer, and high reward will be set at the end-part.

When the experience buffer size is full, replace the oldest experience with the latest one.

Other Parameter Settings:

GAMMA (discount rate) = 0.9

EPSILON_DECAY_STEPS = 100

INITIAL_EPSILON = 0.5

FINAL_EPSILON = 0.01