Autonomous Intelligence Systems

Labortory

Exercise Nr. 4

Reinforcement Learning

Group members
Basharat Basharat
Sindhu Singh
date : January 8, 2019

Introduction:

The goal of this exercise is to experiment with reinforcement learning. It is to implement a DQN agent and evaluate its performance on CartPole and CarRacing game environments from OpenAI gym.

1. CartPole:

The CartPole environment has the four states (cart position, cart velocity, pole angle, pole velocity at tip) and two actions (push cart to the left, push cart to the right). In order to implement the learning and optimizing actions a basic neural network is designed with 20 hidden units. It is considered that optimizing with Adam optimizer together with mean square error gives slightly a better performance. The game is considered solved with average score is greater than or equal to 195 over 100 consecutive frames. The online_training was run with exploration method namely epsilon-greedy. The following parameters' values were used while training the agent.

1.1 Training with E-Greedy:

batch_size = 64, learning_rate = 3e-4, epsilon = 0.5, gamma = 0.95, exploration = greedy, episodes = 1000

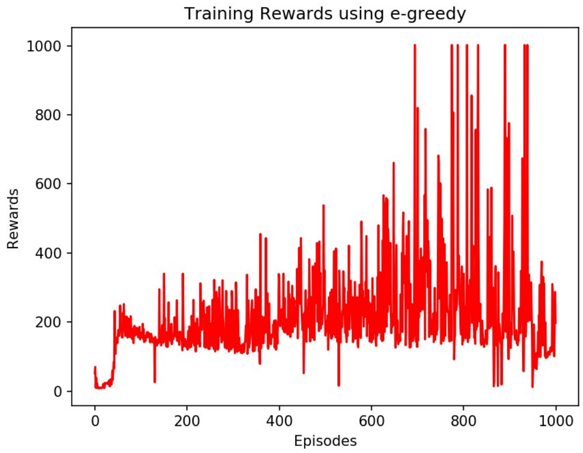

Figure : 1 Training reward using e-greedy

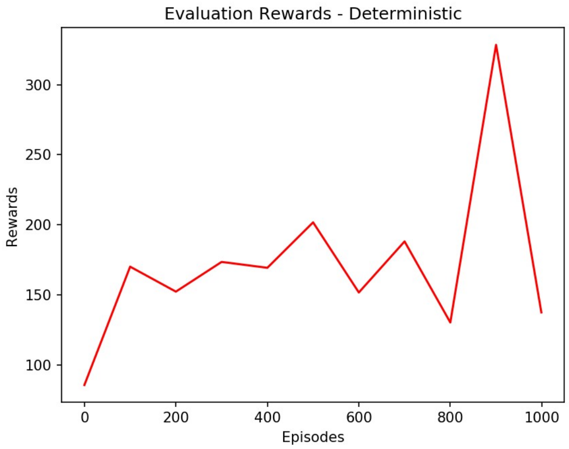

Figure : 2 Evaluation Rewards - Deterministic

Results:

Mean : 208.4
Std. : 29.57

2. CarRacing:

The convolutional neural network was used for the CarRacing environment. The network was designed with three layers with 256 hidden units and one fully connected layer. The layers' parameters are filters (32, 64, 64), kernel_size (8, 4, 3), and strides (4, 2, 1) respectively with padding ('valid') and the Adam optimizer. The state dimension and nr. of actions are (96, 96) and (7) respectively. The online_training was run with two different exploration methods namely e-greedy and boltzmann as follow:

2.1 Training with E-Greedy:

batch_size = 64, learning_rate = 0.001, epsilon = 1.0, gamma = 0.95, exploration = greedy , epsilon_min = 0.01, epsilon_decay = 0.95, episodes = 500 , history_length= 2, skip_frames = 3

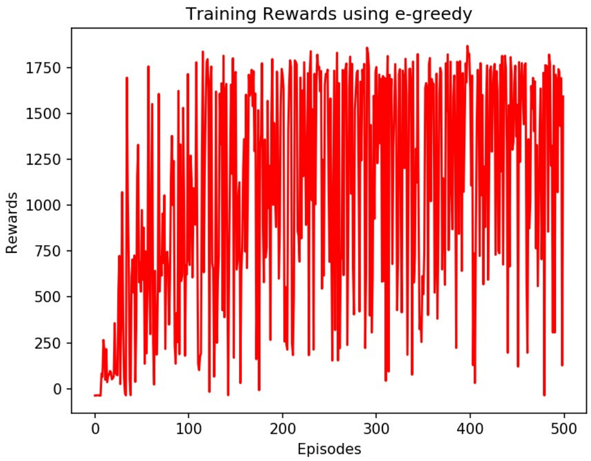

Figure : 3 Training reward using e-greedy

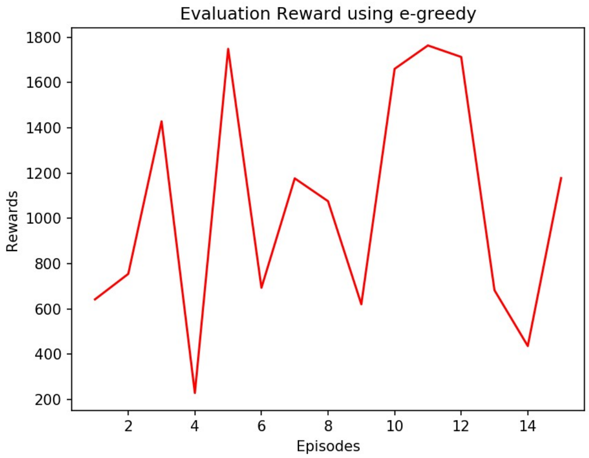

Figure : 4 Evaluation Rewards using e-greedy

Results:

Mean : 1053.62
Std. : 499.20

2.1 Training with Boltzmann:

batch_size = 64, learning_rate = 0.0003, epsilon = 0.95, gamma = 0.95, exploration = Boltzmann , epsilon_min = 0.05, episodes = 1000 , history_length= 2, skip_frames = 2

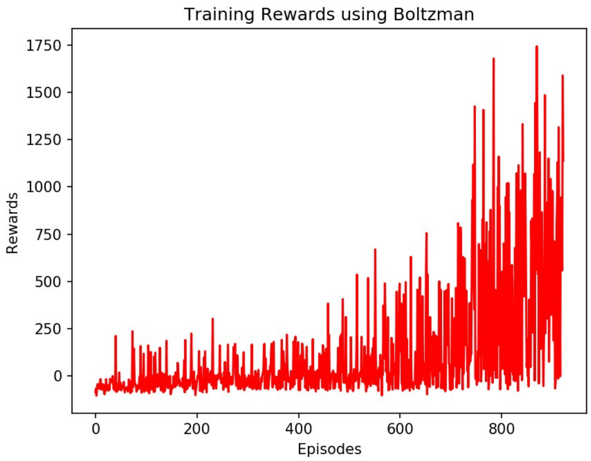

Figure : 5 Training reward using Boltzmann

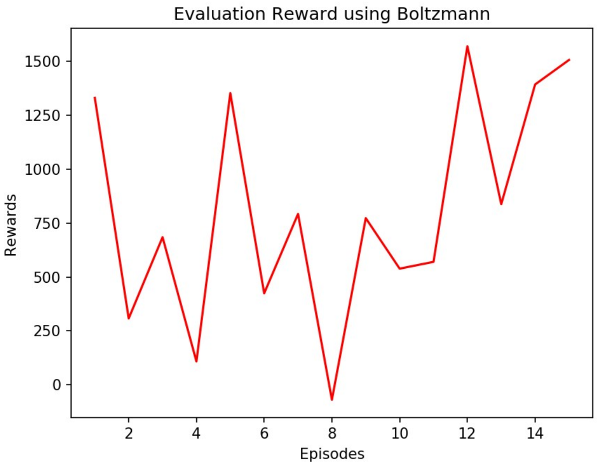

Figure:6 Evaluation Rewards using Boltzmann

Results:

Mean : 807.53
Std. : 501.78