Deep Q Networks

Sungsoo Lim

November 22, 2020

1 Breakout-v0

1.1 (a)

The Mean max Q values were saved after each game, as the text file saving the results were becoming too big for each iteration. The algorithm was run for 10 million iterations. The dip in the plot is due to changing of the skip_game parameter in between sessions.

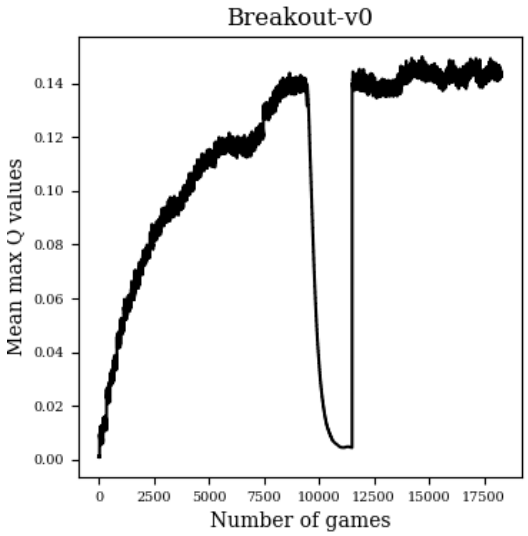

Figure 1: Mean max Q values vs. number of games for Breakout-v0.

1.2 (b)

Based on the Mean max Q values, the algorithm is determined to be learning as seen by the increasing and plateauing of the mean max Q-values. However, when the trained online network was run for 200 games, the score was worse than the random action results (data not shown).

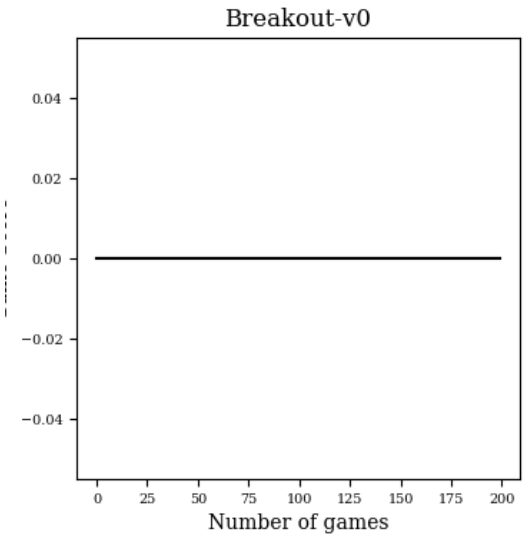

Figure 2: Game scores vs. number of games for Breakout-v0.

2 MsPacman-v0

2.1 (a)

The Mean max Q values were saved after each game, as the text file saving the results were becoming too big for each iteration. The algorithm was run for 10 million iterations.

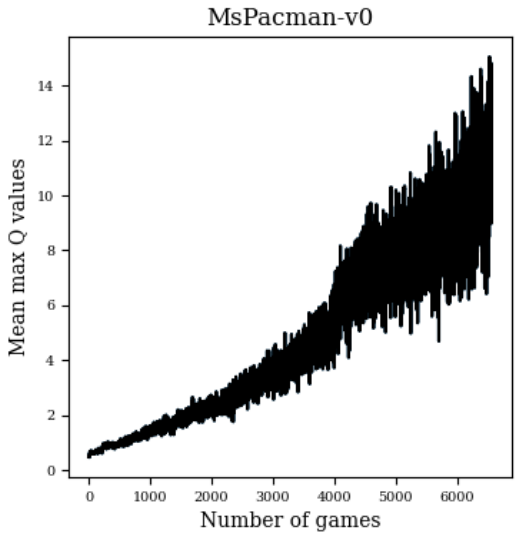

Figure 3: Mean max Q values vs. number of games for MsPacman-v0.

2.2 (b)

Based on the Mean max Q values, the algorithm is determined to be learning as seen by the increasing and plateauing of the mean max Q-values. When the trained online network was run for 200 games (please excuse me, I was only able to run the game 200 times, as the saved score file was lost after the game was run for 1000 times), the score is better than the random action results (data not shown).

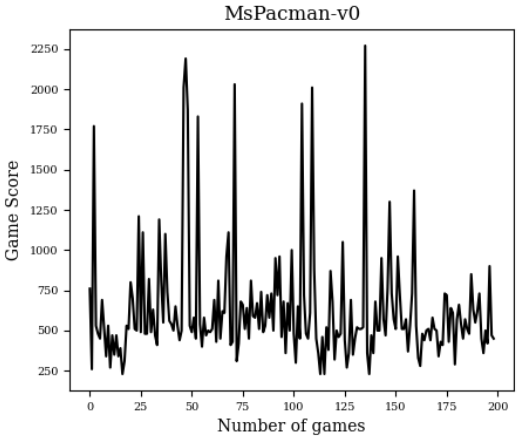

Figure 4: Game scores vs. number of games for MsPacman-v0.