Assignment 2

ECE 551 - Advances in robotics and control

February 8, 2020

1 Instructions

- The goal of this assignment is to understand SARSA, Q-Learning, DQN and its training setup.
- Refer chapter-6 from Sutton and Barto [1] to review SARSA, Q-Learning. Refer [2] to review DQN. [3] provides more detailed analysis of DQN.
- We provide an partially filled in code from [4]. Do attempt to fill in the portions marked with `TODO`. You are free to use DQN codes from other resources that use other frameworks like PyTorch or Keras. Your experiments, observations and report must be original.
- Submit a zip file containing solution.ipynb, report.pdf and other additional scripts on moodle portal.
- You will be marked based on report, oral evaluation.
- Deadline Feb 28th 2020 (12 Midnight).

2 SARSA and Q-Learning (0.5 + 0.5 + 1.0)

1. Draw backup diagrams of SARSA, Q-Learning.
2. Mention 2-3 differences between MC and TD methods.

3. Suppose action selection is greedy. Is Q-learning then exactly the same algorithm as SARSA? Will they make exactly the same action selections and weight updates?. (Exercise 6.12, Sutton and Barto, 2nd Edition)

3 DQN (2.0 + 2.0 + 4.0 + *2.0)

Train a DQN to play Atari 2600 Breakout [5]. This part of assignment will be resource intensive. You need to train upto 2000-4000 episodes (12-18 hrs) to get some decent performance from your network. There will huge variation between two runs with different random seeds. So be patient and perform multiple runs. Take advantage of Google Colab [7] (with checkpoints) or Ada [8]. Your report should contain the following.

1. A plot showing the performance of your DQN. x and y axis should indicate number of time-steps and mean reward for past 30 episodes respectively. You may also use other metrics used in [2]. Add a video/gif link showing your networks final performance.

2. After the network is trained, show 3-5 screen shots of the game, corresponding input to network, Q-values for each action corresponding to that input. Briefly explain your observations.

3. Choose one hyper-parameter (loss function, learning rate, input representation, exploration policy parameter, .. etc) that you expect to affect the performance of Q-network. Run at-least two more experiments by varying this hyper parameter and comment on the performance of the network with plots. Mention your reasoning for the choice of hyper parameter.

Bonus Use the network trained on Breakout [5] to evaluate on Pong [6]. Report it's performance after 0, 100, 500 episodes of training. Show average score in the report. Show its performance plot after full training.

References

[1] http://incompleteideas.net/book/RLbook2018.pdf

[2] https://www.cs.toronto.edu/~vmnih/docs/dqn.pdf

[3] https://web.stanford.edu/class/psych209/Readings/MnihEtAlHassibis15NatureControlDeepRL.pdf

[4] https://github.com/dennybritz/reinforcement-learning

[5] https://en.wikipedia.org/wiki/Breakout_(video_game)

[6] https://en.wikipedia.org/wiki/Pong

[7] https://colab.research.google.com/

[8] http://hpc.iiit.ac.in/wiki/index.php/Ada_User_Guide