CSE190 Final Report

Zeyu Chen: A98005611
Jiapeng Li: A10691672

Topic: Reinforcement Learning Without The Knowledge of Robot Motion Pattern

Introduction to the project:

In this project, we explored the Q-learning method without the knowledge of robot's action knowledge in a variety of situations. Since Q-learning has it uncertainty once the ground truth CPT is unknown, the performance of this algorithm depends on both the samples we get from each state and the learning rate we have set. In order to evaluate the different parameters, we conducted experiments on under controlled conditions and observed how learning rate and number of samples could significantly influence the convergence of the algorithm.

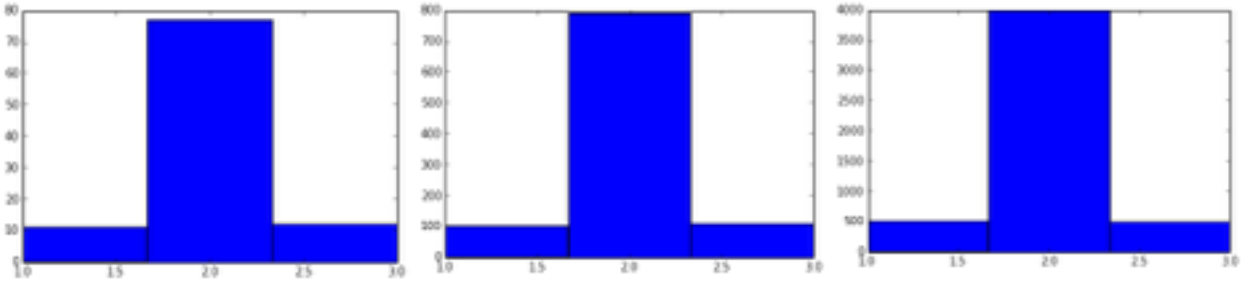
Github repository:
https://github.com/NeroNL/Model_Free_Learning.git

YouTube link: https://youtu.be/K37woRmLcWM

Total Analysis:

The Law of Large Numbers:

In this project , we approximated the unknown action CPT by repeating simulating the robot's action in a large number. By the Law of Large Number, we should be able to get the true distribution of the robot's actions if we have enough number of samples. In the figure below we can see that as our number of samples increase the distribution of the robot's action also comes closer to the 'true' distribution of the robot.



Problem Formulation:

The target formula we want to improve is:

$$V^{\pi}(S) = \sum_{S'} P(S'|S,\pi(S))\{R(S,\pi(S),S') + \gamma V^{\pi}(S')\}$$

$$V^{\pi}_{new}(S) \leftarrow (1-\alpha)V^{\pi}_{old}(S) + \alpha U^{\pi}_{i}$$

Where the value could be obtained by:

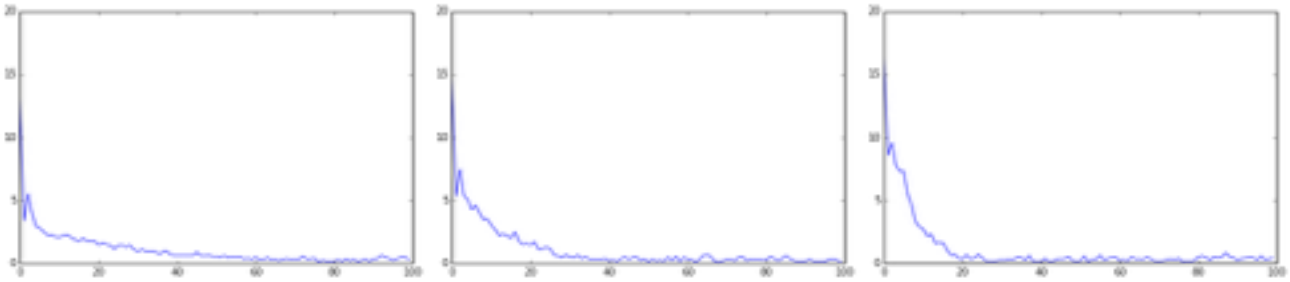
$$\begin{aligned} U_1 &= R(S,Up,A) + \gamma V^{\pi}(A) \\ U_2 &= R(S,Up,S) + \gamma V^{\pi}(S) \\ U_3 &= R(S,Up,A) + \gamma V^{\pi}(A) \\ &\cdots \\ U_N &= R(S,Up,A) + \gamma V^{\pi}(A) \end{aligned}$$

$$V^{\pi}(S) = \frac{1}{N}\sum_{i=1}^{N} U_i$$

The Importance of Learning Rate:

Since we also need to update the value using learning rate, we conducted our experiments on the 3x4 map with a fixed 1000 samples drawn from each state. We experimented on the convergence of the algorithm based on 0.4, 0.6, and 0.8 learning rate. Interestingly, if we conduct the experiment on learning rate 0.2 and 0.1, the algorithm diverges rather than converges. If the learning rate is small, the convergence of the algorithm is slower and more stable, while if we have a larger learning rate, the convergence will be steeper and closer to the output of value iteration. Learning rate could be useful if the number of samples is small such

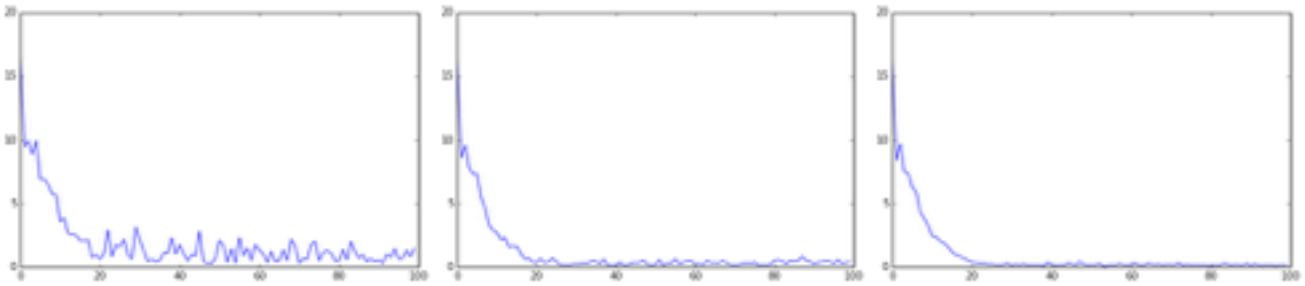
that there is a smaller probability to go to the wrong way; however, this could also lead to slower convergence. The figure below describes the influence of learning rate. We experimented this experiment with 0.4, 0.6, 0.8 from left to right.



The Importance of Number of Samples:

In our scenario, we assume that we do not know the robots action pattern, so in order to simulate the action pattern, we need to simulate the actions for a number of samples and get use the expected reward we get from our samples to compute the best policy.

If the number of samples is small the convergence of the algorithm fluctuates much more in the end than we have larger number of samples. We can conclude from the Law of Large Numbers and the convergence based on large number of samples that it is better to have a larger number of samples to compute the best policy than smaller number of samples.



From left to right the number of samples are 100, 1000, 5000 with a fixed learning rate 0.8 and the map size of 3x4.