CIS 520, Machine Learning, Fall 2021
Homework 10
Due: Tuesday, November 30th, 11:59pm
Submit to Gradescope

Yixuan Meng, Zhouyang Fang

December 1, 2021

Problem 1

1. The state space $\mathcal{S}$ is the state 0,1,2,3,4.
   The action space $\mathcal{A}$ consists of the forward (f) and backward (b) actions.
   The γ is the discount factor, $\gamma = 0.9$ for this question.
   The specification of p and r are shown in Table 1

s	a	s'	p(s'—s,a)	r(s,a,s')
0	f	0	0.2	3
0	b	0	0.8	3
0	f	1	0.8	0
0	b	1	0.2	0
1	f	0	0.2	3
1	b	0	0.8	3
1	f	2	0.8	0
1	b	2	0.2	0
2	f	0	0.2	3
2	b	0	0.8	3
2	f	3	0.8	0
2	b	3	0.2	0
3	f	0	0.2	3
3	b	0	0.8	3
3	f	4	0.8	0
3	b	4	0.2	0
4	f	0	0.2	3
4	b	0	0.8	3
4	f	4	0.8	6
4	b	4	0.2	6

Table 1: component p and r for all possible s,a,s'

2. The optimal state-value function V for the five states are: [24.61, 24.96, 26.85, 30.31, 35.11]

3. The optimal deterministic policy for the five states are: ['b', 'b', 'f', 'f', 'f']

Problem 2

2.1

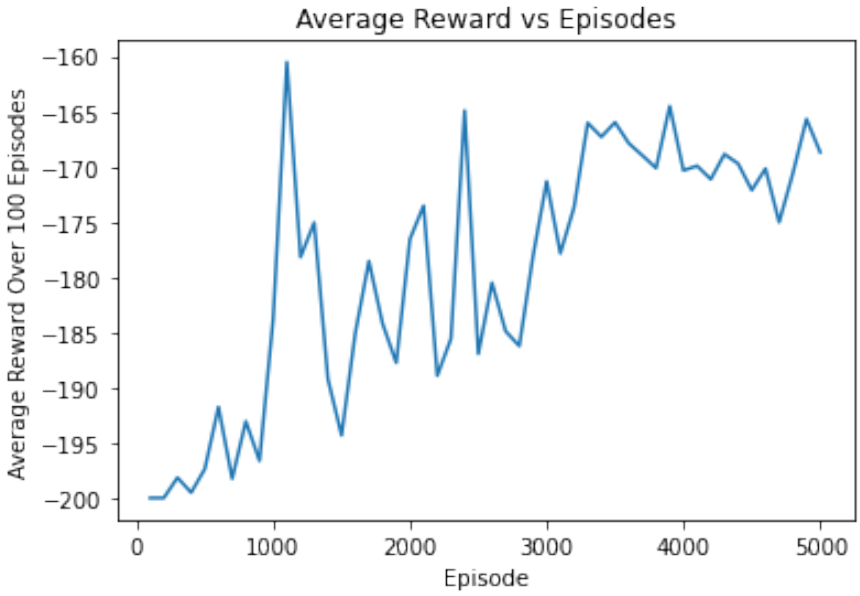

Figure 1: Average Reward

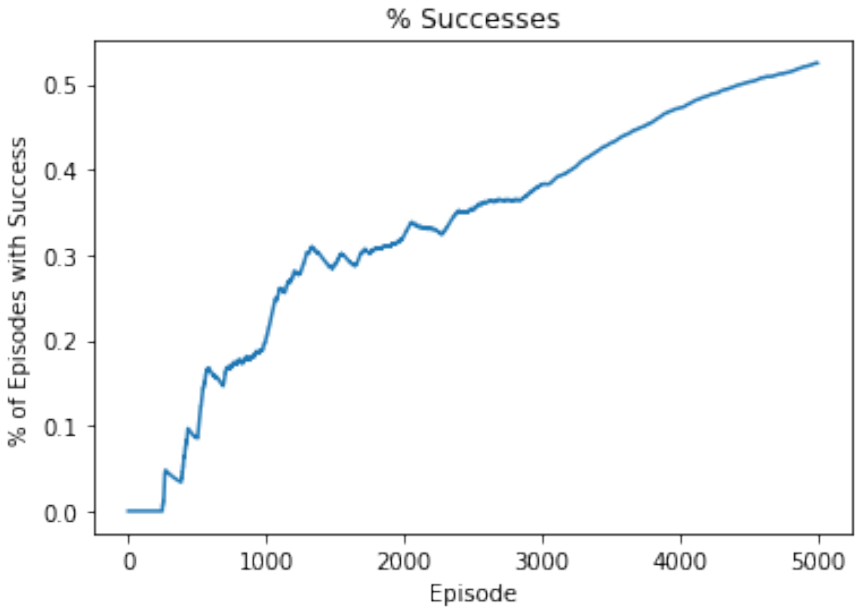

Figure 2: Success Rate

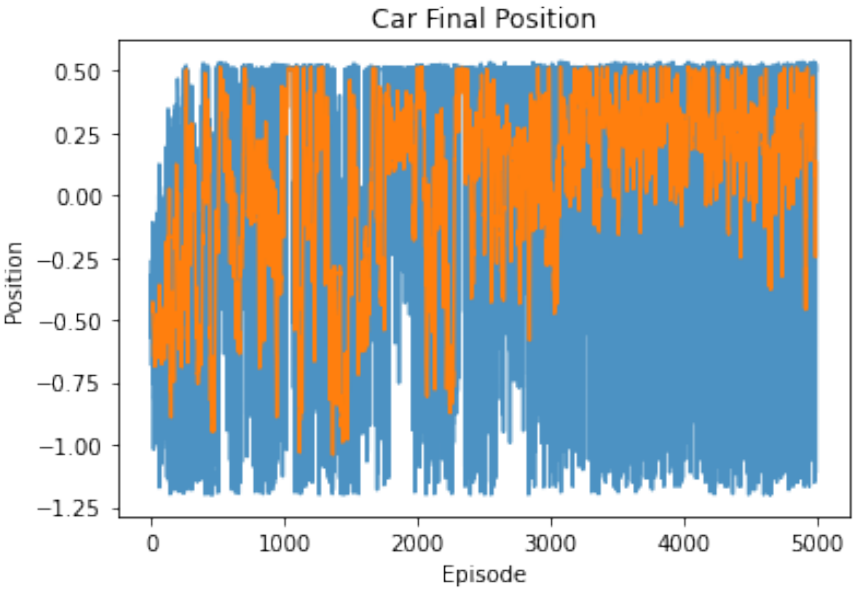

Figure 3: Car Final Position vs Episodes

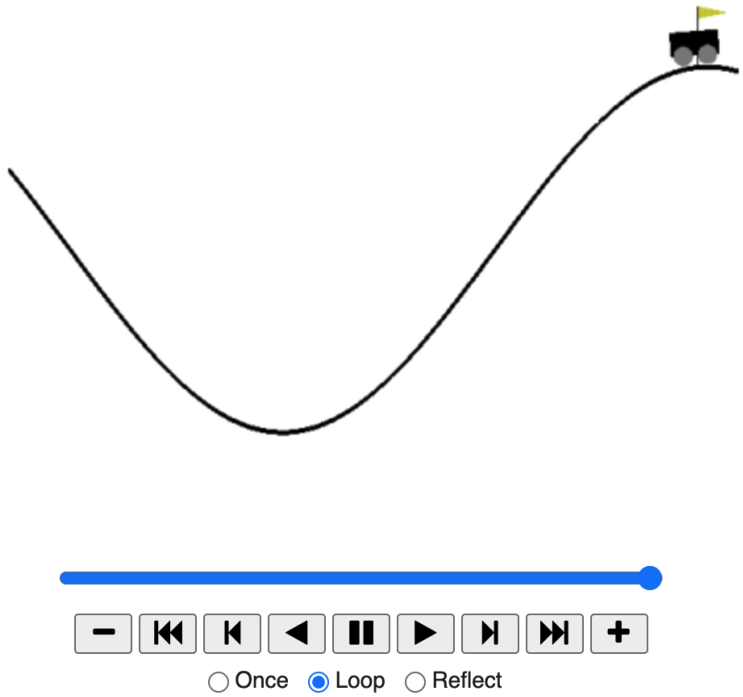

Figure 4: Car reaching the flag in the final episode

2.2

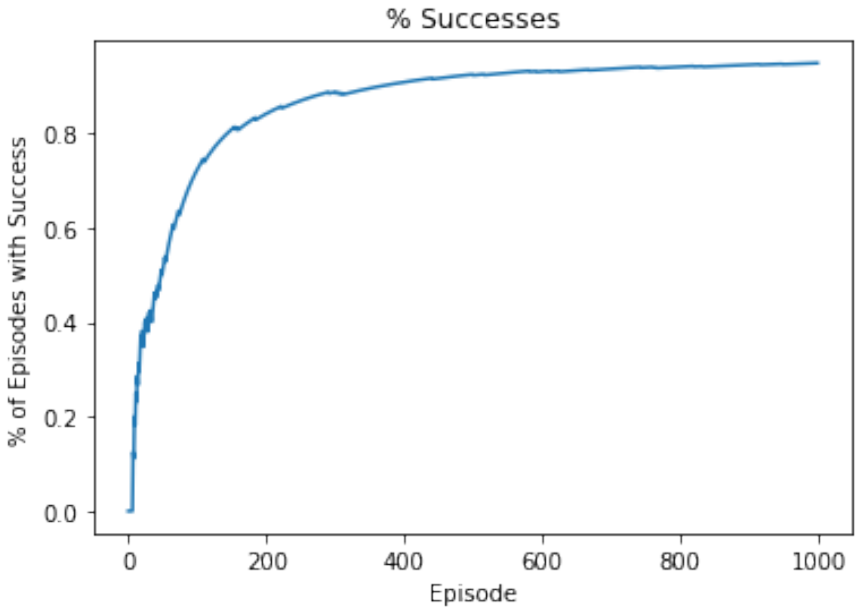

Figure 5: Success Rate

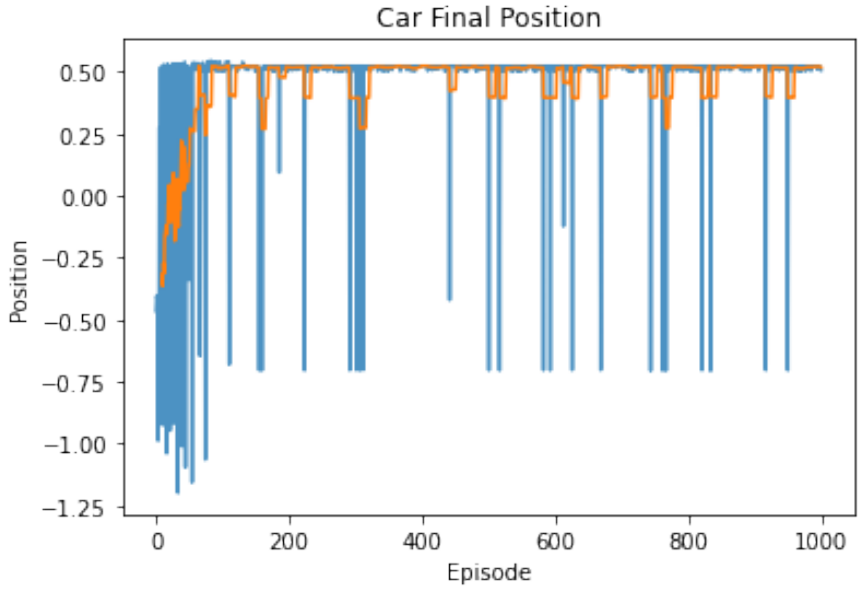

Figure 6: Car Final Position vs Episodes

2.3

1. The new rewarding function is more helpful in two perspectives. Firstly, the reward is a direct reflection of our goal. We want to climb to the hill, so the closer we get, the more reward we get. However, the original reward structure gives penalty to the time we spend, which somewhat deviates from our initial goal.

   Secondly, the new reward-shaping function provides a straightforward and frequently updated guide to the car, avoiding the trouble of passing the reward state by state slowly, making it much faster to train and perform.

2. Deep Q-learning learns much more quickly due to two main reasons. One is the improved reward-shaping function provides a more straightforward guide and is closer to our original goal, making learning speed extremely improved.

   The other reason is deep network provides a better model for the hidden states, which is closer to the complicated structure of optimal policy while not overfitting.