P4 - Smart Cab Report

-Prafull Kumar Soni

QUESTION: *Observe what you see with the agent's behavior as it takes random actions. Does the smartcab eventually make it to the destination? Are there any other interesting observations to note?*

ANSWER: When agent is taking random actions it's just exploring randomly around the environment and getting feedback as rewards when it luckily makes a correct decision it is rewarded with positive reward and negative otherwise. The agent isn't concerned about the actions and rewards and just explores. Smartcab did reached destination a few times when it got lucky not by learning.

QUESTION: *What states have you identified that are appropriate for modeling the smartcab and environment? Why do you believe each of these states to be appropriate for this problem?*

ANSWER: The inputs to the agent by sensing the environment and directions from route planner are appropriate for creating states for modeling. The inputs from environment gives all the details regarding the intersection like which light is on (Red/Green), is there any traffic at the intersection, and the direction of on going traffic is any. While these information is enough to create all the states at the intersection but, we need one more features which is direction in which we have to move forward. This is provided by the route planner, since router planner gives us directions where to go at every intersection, the agent will be able to learn, with this collective information. However, we have deadline feature available to us which might be useful if our agent can access complete route. Since, our agent suppose that it is only one intersection away form goal state, we don't need to care about deadline as irrespective of deadline it will always believe that goal state is only I step away.

QUESTION: *What changes do you notice in the agent's behavior when compared to the basic driving agent when random actions were always taken? Why is this behavior occurring?*

ANSWER: After implementing Q-Learning agent was able to train itself with the feedback given in form of rewards. Agent was able to se use the knowledge it collected and improve itself. After first few rounds Agent showed great improvement and was quickly able to learn rules of intersection such as it should stop at red light, should proceed only if the light is green and taking a right turn on red if no ongoing traffic. This behavior is because agents likes to be rewarded and becomes greedy to get rewards my making right decision and stops making actions which earns him a negative reward or punishment. Agent in the process of maximizing his rewards learns how to drive around the environment.

QUESTION: *Report the different values for the parameters tuned in your basic implementation of Q-Learning. For which set of parameters does the agent perform best? How well does the final driving agent perform?*

ANSWER: For Tuning Q-Learning I tried different values of alpha and gamma and my observations are as follows:

Alpha	Gamma	Total Rewards	Total Actions
0.25	1.0	2182.0	1269
0.5	1.0	2284.5	1427
0.8	1.0	2230	1293
1.0	1.0	2319	1515
0.5	0.5	2137.5	1236
0.5	0.8	2280.0	1331

After trying various combinations of alpha and gamma the results were nearly the same for all cases in terms of rewards / action. However I observed model to performed best when alpha = 0.8 and gamma = 1.0 since, by setting gamma to 1.0 makes the algorithm to be prefer long term rewards over maximizing locale i.e. current state rewards.

QUESTION: *Does your agent get close to finding an optimal policy, i.e. reach the destination in the minimum possible time, and not incur any penalties? How would you describe an optimal policy for this problem?*

ANSWER: Yes, I would argue that the agent was able to get very close to finding optimal policy, and was able to reach to destination in near to minimum possible time with very occasional and very minimal penalties. According to me an optimal policy for the problem is when an agent is able to make optimal decision for taking actions which doesn't led to penalties i.e. it doesn't break any traffic rules and makes to the destination in time.