Project Report Smart Cab

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?

When the agent is making random actions the smartcab sometimes makes it to the destination by chance within the deadline, but given sufficient time the agent will reach the destination as brownian motion in 2d will always return to the starting point given that there is infinite time. Source:
http://www.alexchinco.com/recurrence-in-1d-2d-and-3d-brownian-motion/

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

Answer:
The smart cab travels through intersections towards a destination. I think the states appropriate to this problem is the current intersection information and the destination. The intersection information is the color of the traffic light: green or red. Front oncoming traffic: Oncoming or None. Left oncoming traffic: Oncoming or None. We don't have to worry about oncoming from the right, as they either have red light when we have green. If they turn on red we have the right of way.

OPTIONAL: How many states in total exist for the smartcab in this environment? Does this number seem reasonable given that the goal of Q-Learning is to learn and make informed decisions about each state? Why or why not?

Answer:
Possible inputs are: Color:2, Front:2, Left:2, Waypoint:(1 to 8, 1 to 6) = 48. In total there are 2*2*2*48=384 states in this environment.

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:
Now that the q function is implemented the action taken is no longer random. The agent is able to “learn” from previous actions and make informed actions based on what was previously most rewarding from that state. As the reward is higher for the correct decisions the agent will more likely reach the final destination. This is exactly what I am observing as the agent is successful in reaching the destination before the deadline after a some tries.

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?

In order to tune the parameters one can do grid search across parameters. Consider the parameters gamma and alpha. Alpha is a measure of the learning rate. Gamma is the discount factor that determines the importance of future rewards. In this project the grid space of alpha from 0.1 to .5 with 5 values equally spaced between the interval, and similarly for gamma from values 0.1 to 1.

By doing this I found that there are several parameters that at first run succeeded 98 out of 100 times.

Gamma	Alpha
1	0.1
0.1	0.4
0.325	0.3
0.1	0.1

Running several times yields Gamma=0.1 and Alpha=0.1 as the most occurring parameter for highest success rate with consistent rate above 98%.

By further studying the statistics of these parameters I found that the mean time used to travel to the destination in steps, median and standard deviation. Mean: 13.73 steps, median: 13 steps, standard deviation: 6.8 steps) Where the run for not reaching the destination could occur after several runs.

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?

An optimal policy would be the one that can reach the destination with the least steps required without penalties. As the Board is 8x6 squares and the mean step length is 13.8 I

would say that the policy is not optimal considering that it uses more steps than required. This is most likely due to the start position not being a part of the state. By using a direction to destination instead of the waypoint as a state it could improve the accuracy.