Train a Smartcab using Reinforcement Learning

Q1. *In your report, mention what you see in the agent's behavior. Does it eventually make it to the target location?*

Answer: Yes, it was able to reach destination ultimately, but after large number of steps. Also the number of steps was unpredictable.

Q2. *Justify why you picked these set of states, and how they model the agent and its environment.*

Answer: Initially I choose only other vehicles and traffic light to define the state. But after multiple trials, I figured out that this was not enough. There was a temptation of using coordinates as part of state, but than it would have increased dependence of accumulated knowledge on starting point, and would have rendered knowledge useless if this car would go to some other area.

I then decided to choose, deadline and navigation inputs as part of state.
For this I have created a 5 dimensional panel to hold State and Actions:

Actions: 4
Navigation input: 4
Deadline: 100 (maximum)
Direction of incoming traffic: 64 possibilities, hence 512 (2^9) for binary representations will be more than sufficient.
Traffic Lights: 2

So a 5DPanel of dimensions: 4 X 4 X 100 X 512 X 2
So last 4 dimensions are for State, and First Dimension is for 4 possible Action

Q3: *What changes do you notice in the agent's behavior?*

Answer: Am using a ration to control how many initial steps to use for exploration, and how many remaining for exploitation. I have observed that best results are achieved when 70%-80% of time is used for exploration, before cab starts exploiting its knowledge. I think that is because lack of enough exploration leads to lack of knowledge, and hence wrong decisions.

Q4: *Report what changes you made to your basic implementation of Q-Learning to achieve the final version of the agent. How well does it perform?*

Answer: Have parameterized exploration / exploitation ratio, and also tried with different values of alpha and gamma.

Q5: *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?*

Answer: Yes, towards the end, rewards increased and steps to reach destination decreases.

Output is of Smartcab_Output_with_Gamma_v1.txt file in same folder.