SMART CAB

Reinforcement Learning

Author:
Vayne Lover

Supervisor:
Udacity

August 17, 2016

1 Background

1.1 Tools

- $\LaTeX$
- Excel
- Python

1.2 Q Learning

In this part i want to say my comprehension about Q Learning.Firstly Q Learning is a reinforcement learning algorithm.This algorithm help machine to explore without supervising. The key of this algorithm is the Q(s,a),which store the values of Q.We can learn from the formula:

$$Q(s,a) = (1-\alpha)Q(s,a) + \alpha(R(s,a) + \gamma \cdot max\{Q(\tilde{s},\tilde{a})\})$$

This formula means that we will choose the action which can lead to the max Q in Q(s,a).The parameter α and γ is the learning rate.If α is high,it means that current is more important because Q(s,a) depends less on past.If γ is high,it means that now action has a good influence on Q(s,a) because Q(s,a) depends more on current.

2 Implement a Basic Driving Agent

2.1 Question 1

Q: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?

A:Firstly when i open the agent.py and run it,i find that it doesn't move.Only when i change the code by adding:

action=random.choice(Environment.valid_actions)

Then i find the cab start to move.However,it just move randomly,and if we close the deadline we can see that the cab finally will arrive at the destination.And usually it costs a lot of time.

3 Inform the Driving Agent

3.1 Question 2

Q: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?

A:I add the states oncoming,left and right.Firstly we know that we must keep the passengers safe.Secondly i think only consider these states will we decrease the time we get negative rewards.

3.2 Question 3

Q: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?

A:In general,there are next_waypoint,light,oncoming,left and right states.Yes,i think only when we combine these states will we can find the best way to arrive at destnation with safety.

4 Implement a Q-Learning Driving Agent

4.1 Question 4

Q: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?

A:We can observe that as first it always gets negative rewards.And it's very ridiculous that the cab circles to avoid other car.However,after some time trying,it usually gets positive rewards and especially last several times it does the right actions and reaches the destination.

5 Improve the Q-Learning Driving Agent

5.1 Question 5

Q: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?

A:I choose different α,γ and epsilon,and finally find that when α=0.5,γ=0.3 and epsilon=0.9 it has the best result.We can know that the accuracy is 96%.

α	γ	epsilon	Steps	Rewards	Success	Total
0.3	0.3	0.7	1889	2181.0	81	100
0.3	0.5	0.7	1927	2462.0	78	100
0.3	0.7	0.7	2075	2392.0	76	100
0.5	0.3	0.7	1737	1971.5	82	100
0.5	0.5	0.7	1922	2184.0	72	100
0.5	0.7	0.7	2024	2230.0	75	100
0.7	0.3	0.7	1927	2145.5	82	100
0.7	0.5	0.7	1895	2130.5	81	100
0.7	0.7	0.7	2123	2182.5	76	100
0.3	0.3	0.9	1552	2251.5	93	100
0.3	0.5	0.9	1537	2401.5	90	100
0.3	0.7	0.9	1835	2732.0	87	100
0.5	0.3	0.9	1559	2286.0	96	100
0.5	0.5	0.9	1653	2344.5	93	100
0.5	0.7	0.9	1899	2979.5	82	100
0.7	0.3	0.9	1516	2268.5	93	100
0.7	0.5	0.9	1655	2493.0	93	100
0.7	0.7	0.9	1738	2545.0	84	100

5.2 Question 6

Q: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?

A:In my opinion,when i choose the parameters α=0.5,γ=0.3 and epsilon=0.9 the cab usually do the optimal policy,it can usually reach the destination using minimum time and without penalties but sometimes it doesn't follow the best policy. As for the optimal policy,i think the cab should satisify:

- Arrive at destination with minimum time cost.
- Keep the passengers safe.In another way is obeying the rule of traffic.

6 Reference

1. https://classroom.udacity.com/nanodegrees

2. http://mnemstudio.org/path-finding-q-learning-tutorial.htm

3. https://discussions.youdaxue.com/c/nd009-p4-train-a-smartcab-to-drive