

Report

实现基本的驾驶智能体

问题: 在你的报告中, 说明观察到的智能体的行为。它最终到达目标位置了吗? 有什么其他值得注意的地方吗?

Answer:

由于智能体的行为是随机的, 运气好的话, 它是可以到达目标位置的, 但是到达目标位置时一般都超过了deadline。如果deadline = -100时智能体还没有到达目标位置的话, 就会尝试失败。

Inform the Driving Agent

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:

1. next_waypoint (下一路径点的位置)
2. light (信号灯的状态)
3. oncoming (前面车的行驶方向)
4. left (左边车的行驶方向)

next_waypoint是根据目标位置得到的, 用于指示前进方向。

light, oncoming, left是为了不违反交通规则, 不发生撞车事故。

没有选择right是因为右边的车不会对智能体产生影响, 不需要考虑。

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:

在这个环境中一共有 $3*2*4*4=96$ 个状态，对于Q-Learning来说是合理的。这些状态包含了所有可能会使智能体违反交通规则和发生撞车事故的情况。

Implement a Q-Learning Driving Agent

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:

智能体不再是随机的移动。经过学习以后，智能体能够不违反交通规则，不发生撞车，快速的到达目标位置。即使有deadline，对智能体来说也能轻松应对。这是因为我们设置的Qtable，并且根据智能体每一次的行为都对Qtable进行了更新。当智能体在某个位置出现了错误的行为或者去了错误的方向时都会被惩罚，经过几轮学习之后，智能体就能够自己规范行为并去向目标位置的方向。

Improve the Q-Learning Driving Agent

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:

alpha

只要不把alpha调到大于1，或者减小到0，学习效果基本没有变化。因此设置为0.3。

gamma

情况和alpha一样，只要在(0,1)就是合理的。因此设置为0.1。

epsilon

在训练次数小于60次的时

候，`self.epsilon = (80 - self.ntrain) / 60`，在最开始的时候epsilon的值应该大一些，智能体应该多尝试不同的情况以适应不同的环境。同

时epsilon的值不断减小，直到60次以后， $\epsilon = 0$.

*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: 在程序中的智能体依据优化，已经可以顺利到达目的地，并且能够保证不会得到任何处罚。对于这个智能体来说，最优算法应该是在不违反交通规则，不发生事故的前提下快速到达目标位置。