UDACITY

PROJECT

Train a Smartcab to Drive

A part of the Machine Learning Engineer Nanodegree Program

PROJECT REVIEW

CODE REVIEW

NOTES

SHARE YOUR ACCOMPLISHMENT!

Meets Specifications

You have a good understanding of Q-Learning already.

Please take some time to think through the major steps of this project and digest (with NO RUSH). 😄

- Random action (no learning)
- State selection
- Basic Q-Learning (with Exploration and Exploitation)
- Parameter tuning
- Optimal policy (and possible limitations and concerns)

Hope you have a better idea of Q-Learning and Reinforcement Learning...

实施基础驾驶智能体

✓

学生能够实施智能体接受指定输入所需的接口。

✓

驾驶智能体产生有效输出（one of None、'forward'、'left'、'right'），以响应输入。

✓

驾驶智能体在模拟器中正常运行，未产生任何错误。奖励和惩罚没有必要——我们允许智能体出错。

确认并更新状态

✓

学生确认了模拟驾驶智能体和环境的状态，并给出了合理的理由。

Great job to exclude "deadline" and "location" in the agent states.
"Next_waypoint" is used to give the agent "direction", like the GPS, to the destination. Therefore, "location" will not help agent to make decision; "Deadline" has no direct help either. If we include any of them in the agent states, the Q table will grow too big, and it will take too much resources to retrieve, calculate, and update/store Q-table.

✓

驾驶智能体运行时会根据当前输入更新自己的状态。确切的状态并不重要，也无需与输入有关系，但应该在运行时进行相应更改。

实现 Q 学习

✓

驾驶智能体会正确更新 Q 值的表/映射，从而实施 Q 学习算法。

✓

鉴于某个状态当前的 Q 值集，智能体会选择可用的最佳动作。

✓

学生报告了观察到的行为变化，并提供了合理解释。

增强驾驶智能体

✓

驾驶智能体不断按规定时间到达目的地，并且净奖励保持为正数。

✓

已报告了学生在 Q 学习基础实现期间所做的改进。

You got the idea that parameter tuning is very important.
You certainly can do more and present it more professionally...

✓

从最终驾驶智能体与学习最佳策略的关系角度讨论了该智能体的性能。

Please notice that, the agent is trained in a simplified simulation environment. All the conditions, e.g. next_waypoint, traffic lights, step rewards, and deadline, are all designed to help the agent to learn. Any changes of those conditions shall affect the agent performance.

However, the purpose of training the smartcab is to make it reach the destination safely and promptly (could be measured differently for shortest time, shortest turns, or shortest distance). The optimal policy is not environment specific.

As you mention, "Even the agent can have 100% success rate, there are still in-time situation occurs at driving. This model still need to be done for improvement.". It is really helpful to understand the effects of the training environment and the limitation. It will be more challenging to have the agent run in the real world situation even it can have 100% performance.

DOWNLOAD PROJECT

Have a question about your review? Email us at review-support@udacity.com and include the link to this review.

RETURN TO PATH

Student FAQ