

项目审阅	代码审阅 4	注释
------	---------------------	----

Requires Changes

SHARE YOUR ACCOMPLISHMENT

11 SPECIFICATIONS REQUIRE 变化



- 新年快乐～
- 总体来说做的不错～
- 但是你的一些细节没有回答到位，我都给你以【问题】或者【优化】的标记指出了。通过针对这些标记进行修改，你的回答能够更严谨。
- 同时，你对 state 中 feature 的挑选还有改进的空间。一个 state 挑选不完善的小车是非常危险的——你一定不希望未来你坐的自动驾驶小车不能够感知十字路口右侧的情况吧！你也许了解过 tesla 自动驾驶的车祸，那就是极端情况下某个状态没有被充分考虑到（训练到）导致的车祸。训练到A(A+) 并不是我们的最终目的（这个指标也有一定的局限性）——我们希望的一定是一个有着很高安全性、能够对周围充分感知的小车。
- 如有参考引用其他作业或资料，请额外给出你的参考来源与链接。

（可选）开始

 学生给出了对与智能车有互动关系的环境的认识。

【优化】

- 绿灯的时候，是否总是负的奖励？你可以考虑增大 `update_delay`，多观察几轮，再得出新的结论。


 学生正确的回答了关于*训练智能出租车*代码的有关问题。


【优化】


- 想在 Markdown 里插入代码引用，请在代码首尾使用 ``` 符号，即 `Tab` 上方的那个键，如

``code``

实施基础驾驶智能体

 当需要做动作时，小车能做一个有效动作。在模拟器中能够产生于智能车行动匹配的奖励或惩罚。

 学生做出了一个基本自动驾驶智能体结果的可视化。

 学生总结了对基本智能体行为的观察。如果学生制作了相应的可视化图表（可选），也根据图表做了相应的分析。

通知智能体

 学生论证了哪些特征最能对智能体在环境中的驾驶状态来建模。不必要的特征没有包括在状态里，并且也给出了理由。

- 根据你的表述，我认为你对 `inputs` 这个 feature 的理解不是很透彻，请看如下的解释：

【解释】


- `oncoming`、`left`、`right` 表示交叉路口对面路、左路、右路的车辆情况，每个特征都有4种可能的方向：
 - 它们表示我们训练小车当前所在路口的对面、左侧、右侧路上车辆行驶的方向。如果是 `None` 的情况下可能是小车停止或者没有小车。
 - 如 `oncoming` 上的值是 `left`，那么代表路对面车下次行动的转向是左转，即这个下次车会走到我们小车在当前路口的右侧方向上的路上。
- 那么你回答中类似「多数的情况是确定右边没有车的时候才会转右」这样的表述其实是不正确的，请你修改之。

【问题】

- 同时 `right` 从理论上来说也是应该使用的
- 首先对于安全性，我们认为给小车足够多、足够完整的状态还是很重要的。
- 从交通规则上说，右转的确不太受其他车的干扰。但是在真实的情况下，难免会出现其他车违背交通规则的情况。因此我们需要对环境进行完善建模～
- 因而，尽管根据交通规则分析能够带来不错的结果，但是一个完善的建模还是能够在某些特殊情况下帮助小车避免事故。

【优化】

- 对 `deadline` 的直觉是正确的，但还可以更充分。
- 你可以从这个方面考虑（不选用的）原因：
 - 针对这个题目的特征（例如起点、终点不停改变），这个 `deadline` 是否具有泛化能力与实际含义？
 - 增加它对状态空间的大小影响是怎么样的？具体来说，增加它前后状态空间扩大了多少？

 学生正确的计算了状态总共有多少种可能。并且讨论了一个合理的测试轮数下，智能车是否能够学会一个可行的策略。

【问题1】

- 你提到 `waypoint` 有3个状态值，但实际上在 `planner.py` 中 `waypoint` 有四种取值 `('left', 'right', 'forward', None)`，请具体指出此处你说它是3个状态值的原因～


【问题2】

- “这个空间的值不是很大，应该在可接受的训练次数里完成。”，这个结论看起来不是这么地显然，请你补充一些具体的分析：
 - 例如，我们是否可以在一定的训练次数训练的到可靠的策略？
 - 这个训练次数是怎么得到的呢？

 根据状态的定义和给定的输入，智能车成功地更新了它的状态。

请参考上述审阅优化你的代码。

实现智能车的 Q-Learning

 智能体在给定状态下的Q值下，能够在可选动作中选出最佳的那个。此外，智能车能够依照学习率和收到的奖励或惩罚，正确地更新映射到特定状态中的Q值。

【思考】

- 为什么我们这么强调对相同最大值的 action 的选择要随机呢？
 - 你可以从小车刚开始模拟、Qtable 还是空的情况还是进行分析。
 - 如果在那个时候，不随机选择 action，而用类似“总是挑列表的第一个最大的”这样的方法会造成什么问题？
- 请在报告中“问题5”下方加上你对这个问题的回答。

 学生给出了一个正确捕捉Q-Learning智能体初始 / 默认情况下的结果。

请在调整了 state 之后再重新生成此处的图像。

 学生总结了观察到的在初始 / 默认状态下 Q-Learning 智能体的行为，并把它与基本智能体做了比较。如果含有可视化内容，学生也做了相应的分析。

请在调整了 state 之后再根据重新生成此处的图像、微调此处的分析。

提高驾驶智能体

 智能车做了除初始 / 默认设定之外的其它可选参数的尝试。

请在调整了 state 之后，尝试重新调整参数（注意，训练次数可能会有所增大）。

【解释1】

- 首先给你补充一下对于 epsilon greedy 算法的解释：
- 对于 epsilon-greedy 算法，你可以参考论坛中的 [这个帖子](#)：


Q: 如何理解 greed-epsilon 方法 / 如何设置 epsilon / 如何理解 exploration & exploitation 权衡？
A: (1) 我们的小车一开始接触到的 state 很少，并且如果小车按照已经学到的 qtable 执行，那么小车很有可能出错或者绕圈。同时我们希望小车一开始能随机的走一走，接触到更多的 state。(2) 基于上述原因，我们希望小车在一开始的时候不完全按照 Q learning 的结果运行，即以一定的概率 epsilon，随机选择 action，而不是根据 maxQ 来选择 action。然后随着不断的学习，那么我会降低这个随机的概率，使用一个衰减函数来降低 epsilon。(3) 这个就解决了所谓的 exploration and exploitation 的问题，在“探索”和“执行”之间寻找一个权衡。

【解释2】

- 再给你补充一下对 alpha 的解释。alpha 是一个权衡上一次学到结果和这一次学习结果的量，如：`Q = (1-alpha)*Q_old + alpha*Q_current`。
- alpha 设置过低会导致小车不在乎之前的知识，而不能积累新的 reward。一般取 0.5 来均衡以前知识及新的 reward。
- 希望你根据这些理解，优化一下你的 epsilon 和 alpha 设置～

 用可视化的方式捕捉了经过提高的Q-Learning智能体的行驶结果。


请相应地重新生成此处的图像。

 学生总结了优化过的Q-Learning智能体和它的行为，近一步比较了观察到的与初始 / 默认情况下的不同。如果含有可视化内容，学生也做了相应的分析。

请相应地修改此处的分析。

【问题】

- 「训练了约240个示例，并对50个示例进行了测试，相当于对一个数据集的80-20分割。」这个表述不正确。强化学习和之前监督学习、非监督学习的理论框架是不同的。（非）监督学习他们是基于样本来训练模型，而强化学习是通过和环境的交互来训练我们的学习体（agent）。那么训练240次，相当于让 agent 在环境中运行240回合，通过和环境交互来学习到对应的 Qtable，接着再用50回合，测试 agent 在环境中的表现。建议你把这个过程用数据集来进行类比。
- 你也可以参考 [这篇文章](#)，加深对 Q learning 的理解。

 智能体能够安全可靠地引导智能出租车在规定时间内到达目的地。学生存在安全性和可靠性上获得至少都为A的评分。

【问题】

- 在「问题8」中，你对于具体情况的策略做了很好的总结。
- 但是，针对所有的 state 的所有策略，它们都有一定的共性，你需要对这个共性——也就是所谓的宏观最优策略——是进行归纳总结，进行泛化的评价。
 - 就如这个例子：训练决策树，具体的策略可以是“遇到XX特征就把它分到YY类”，但是宏观的策略就是按照某种内在的规则或者目标对数据进行分类。
- 同时，这个宏观的策略也是你训练小车的目的。
 - 如你驾驶的时候你在各个情况下，无论作出的具体行动如何，最后的目的无非就是不出事故、及时到达目的地等。
- 相信这样你应该能够 get 这个题目的含义。

 学生正确地陈述了该项目的两个特点，使得Q-Learning 中未来的奖励在这里没有意义。

- 很有思考的回答，但是你这边的分析有没有击中要害～


【优化】

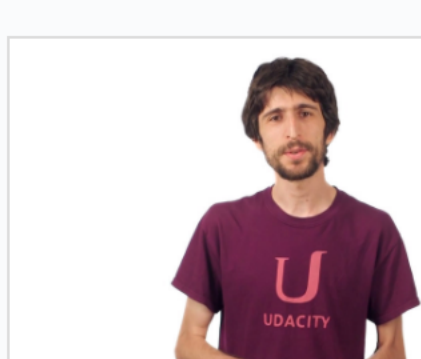
- 这边你可从数学公式出发，讨论增加的未来奖励项的含义以及它对更新公式的影响是怎样的，然后对整个 Q-table 中的数值的影响又是怎样的。[Q-learning](#)
- 再次，你还可以从项目的特点出发，这边增加未来的奖励在这个项目情景下是否合理、有没有什么问题？

 重新提交项目

 下载项目

4 代码审阅评论





重新提交项目的最佳做法

Ben 与你分享修改和重新提交的 5 个有益的小贴士。

 [观看视频](#) (3:01)

给...