

















Requires Changes

**还需满足 6 个要求 变化**

* 报告如果有参考其他同学的作业或他人的分析，请在报告中对应的部分附上参考链接／引用文献。
* 总体来说做的不错，你的问题我也在审阅中给了你答复。
* 但是你的许多细节地方都没有回答到位，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。
* 此外，我还给了你不少思考的内容，以【思考】标注。思考这些思考题能够帮助你更深入理解项目的机理，并对数学、概率大的知识有所回顾与提升。

**（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。**

**【问题】**

* 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。**

很细致的回答，非常好！

**【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。**

* 这边想强调一下，可靠性是衡量小车能不能在有限时间内到达目的地，安全性是衡量小车出事故的情况。

**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。**

**【问题】**

* 你的说法“由于deadline最大为100，可能的取值太多”是不太完整、不太正确的。请在 environment.py 中查找定义 deadline的代码，说明这个变量的取值究竟是如何取的～之后再说明一下这个变量的选用带来的影响。

**【优化】**

* 此外，你对 deadline 的分析还可以更充分。
  + 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？

**学生正确的计算了状态总共有多少种可能。并且讨论了在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。**

**【优化】**

* 你这个问题回答的不够好～
* 这实际上在组合问题中是一个乘法问题，请你指出 state 中每一个特征有几种取值情况，并具体列举是哪几个取值，然后再将它们相乘。

**【思考】**

* 从严谨的角度上说，有 384 个 state，那么究竟多少次，才能保证多少百分率的 state 被训练充分呢？可以从这几个方面出发：
  + 考虑一个简单的独立事件模型，如果每一次行为都是纯随机（与上次行动无关），那么多少次能遍历多少百分比的状态？
  + 考虑一个简单的马尔科夫模型，如果每一次行为都是只与上次行动有关，那么多少次能遍历多少百分比的状态？
  + 一个状态要被遍历多少次才能算充分训练了？
  + 多少个状态被充分训练了才能保证小车的驾驶符合要求？
* 尽管你要求了直接给你提示，但我认为你应当先独立思考，尝试写出一些答案，我再给你进一步的分析。

**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。**

**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。**

**【问题】**

* 在选择 action 的时候，我们要考虑多个 action 同时出现相同最大 Q 值的时候，要怎么处理。在这个项目中，我们选择随机挑选一个 action 的做法。
* 请修改对应的函数~

**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。**

**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。**

**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。**

* 你的训练次数较多。可以结合 alpha、epsilon 下降函数来优化你的训练次数。

**【问题】**

* 有尝试过稍微加快一下 epsilon 的率减速度，减少训练次数吗？我们知道，一个随机性很大的小车肯定效果不好，分析左上方和右上方的图，发现事故曲线和 epsilon 下降曲线基本重合，可以发现是 epsilon 下降的速度过慢导致了学习次数过多，因而可以考虑加快 epsilon 的下降速度。建议你能够再多尝试一下，将训练次数下降到1000次以下。
* 这边是要你能够在 “训练次数-训练效果” 这对矛盾的关系之间寻求一种平衡，并不一定是“训练时间越长、训练结果越好”。

**【优化】**

* 我们希望，在训练初期的时候多探索，在训练末期的时候少探索，那么我建议你选择类似 cos 这样的上凸函数。你可以查看这个链接 [函数的凹凸性](https://baike.baidu.com/item/%E5%87%BD%E6%95%B0%E7%9A%84%E5%87%B9%E5%87%B8%E6%80%A7/4583322) （wiki 怕打不开）。这个函数的特点是，一开始导数负得小，逐渐地导数负得越来越大。你可以和指数函数对比一下。

**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。**

**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。**

你这边回答的很好，不过你可以更深入地思考一些问题。

**【思考】**

* 我们知道，学习率 α 的目的是为了在更新 Q 值的同时也保留过去学到的结果，那么对于不同的 state，实际上学习的进度是不一样的。那么此处对所有的 state 统一设置 α，似乎并不是最优的做法。你可以考虑对每个 state 设置不同的学习率，该 state 学习完毕后其对应的 α 衰减，而其他 state 对应的 α 不变。可以参考 周志华 的 《机器学习》（西瓜书）中相关的内容。

**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。学生在在安全性和可靠性上获得至少都为A的评分。**

**【问题】**

* 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，这个共性是要你能够进行归纳总结，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。**

建议你思考一下这个问题～