智能车项目报告

目标

智能车从一个地点开始，在有限的移动步数内到达目的地。

已知条件

1. 交通规则：绿灯亮时，仅在十字路口无直行来车时才能左转；红灯亮时，如果无直行来车左转，或左方来车直行时，则可以右转。

2. 动作：None, forward, left,right

问题分析

当智能车在每个路口时，采用随机的动作。观察到一些行为

1. 智能车大部分情况无法完成目标，即不能在有限的步数内到达目的地。但有时也可以到达目的地
2. 智能车可以穿越边界到达相反的边界
3. 智能车的每一步越靠近目的地，reward 的值越大
4. 智能车表现很怪异，有时在一个位置停留很久，整个行驶轨迹比较混乱，毫无章法。

根据观察，虽然智能车能够移动，但是并不能很好的达到目标，因此我们对智能车建模

建模分析

选取一组状态对智能车和环境建模。我们选取（next_waypoint,light，oncoming,left）作为一组状态。主要是因为一下几个原因：

1. next_waypoint：表明智能车在路口时，不考虑交通因素，选取的最快到

达目的地的动作，表明去往目的地的方向。

2. light：表明交通规则，影响步数，路口为红灯的时候，可以向右移动，
3. left：表明交通规则的影响，会一定程度影响步数，当无直行车时，可以左转。
4. oncoming：表明交通规则的影响，也会影响步数，和 left 相关，

   对于 right，影响权重比较小，不管右侧有没有车，不影响车的动作，所以不考虑这个属性

根据选择的 4 个状态，next_waypoint（4 个值），light（2 个值），left（4 个值），oncoming（4 个值），所以状态空间一共有 4*4*2*4=128 个状态，状态空间的变量比较少，每个状态都会影响目标的实现，足够做 Q-learning，使每个状态可以做出基于训练的决策。

Q-LEARNING 实现

状态更新函数：

Q(s,a)=(1-alpha)Q(s,a)+alpha(reward+gamma*max(Q(s’,a’)))

s:当前状态

a:当前动作

s’:下一个状态

a’:下一个动作

Q(s,a):当前状态，当前动作的 Q 值

Q(s’,a’):下一个状态，下一个动作的 Q 值

reward：下一个动作的奖励

alpha：学习率,下一个状态的 Q 值的权重。

gamma: 折扣因子

epsilon：随机参数，随机选取动作

实现 Q-LEARNING 之后，智能车能够很快的到达终点。和随机动作相比，智能车的行为有规律，基本可以朝着终点行驶，能够一定程度理解交通规则，红灯的情况下选择停下，尽量躲避车辆。产生这种适当的动作是因为，随着每一步的动作选择，在不同的状态下，产生奖励，随着训练的增加，状态空间慢慢收敛，在新的状态下，能够选择正确的动作。

根据观察，除了小概率的随机动作，训练 20 次之后，除了周围车的影响，智能车可以准守交通规则，路口绿灯时，开始服从导航，当训练 60 次之后，智能车完全可以准守交通规则, 除了交通规则限定的特定状态, 其他状态服从导航。当 Q-VALUE 状态空间中，某个状态 state 的对应的 4 个动作 action 中，最大的值大于等于 0 时, 可以选择正确的动作, 也就是说准守交通规则。当状态空间中，在某一个状态时，最大值的动作 action 和导航的动作 action 相同时，智能车表现和导航的动作 action 一致。随着智能车不停的训练，状态空间的值越来越多，越来越收敛，智能车的动作选择越来越准确。

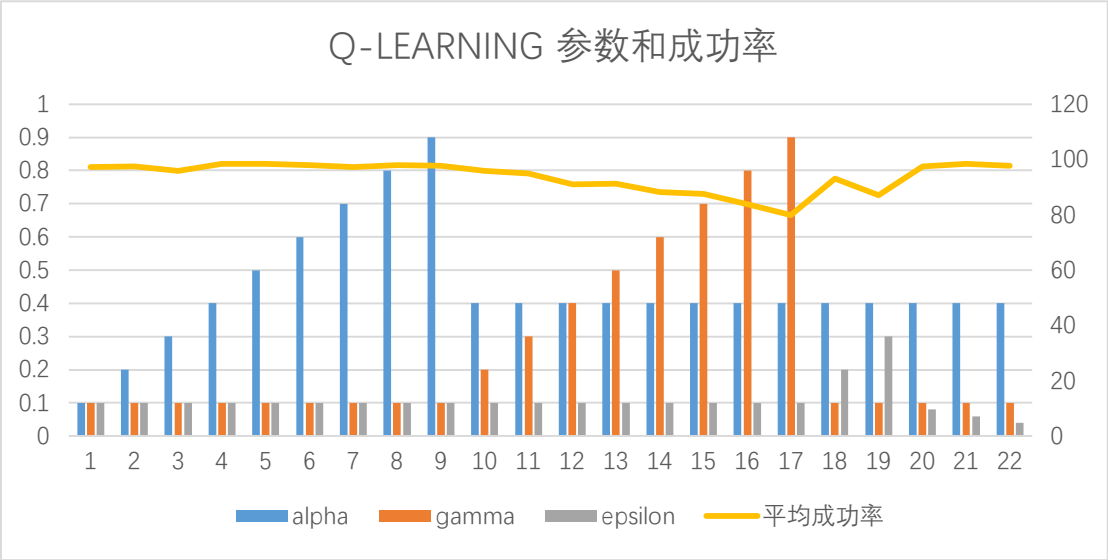
参数调优

参数是学习率 alpha, 折扣因子 gamma 和随机参数 espilon,主要是对这 3 个参数调整，达到高准确率和快速学习的能力。以下是每组 100 次学习结果的成功情况。实验结果如下：

alpha	gamma	epsilon	平均成功率(%)	第一组	第二组	第三组	第四组	第五组

0.1	0.1	0.1	97.2	97	96	97	97	99
0.2	0.1	0.1	97.6	100	94	99	96	99
0.3	0.1	0.1	95.8	97	95	95	95	97
0.4	0.1	0.1	98.4	97	99	97	100	99
0.5	0.1	0.1	98.4	98	100	98	97	99
0.6	0.1	0.1	98	98	97	97	99	99
0.7	0.1	0.1	97.2	97	100	96	97	96
0.8	0.1	0.1	98	98	99	98	96	99
0.9	0.1	0.1	97.8	99	99	98	98	95
0.4	0.2	0.1	95.8	96	98	93	95	97
0.4	0.3	0.1	95	96	97	96	94	92
0.4	0.4	0.1	91	90	87	92	92	94
0.4	0.5	0.1	91.2	90	88	93	95	90
0.4	0.6	0.1	88.2	89	92	86	86	88
0.4	0.7	0.1	87.6	87	83	86	90	92
0.4	0.8	0.1	83.8	85	89	87	80	78
0.4	0.9	0.1	80	82	68	78	86	86
0.4	0.1	0.2	93.2	95	96	92	92	91
0.4	0.1	0.3	87	87	88	86	89	85
0.4	0.1	0.08	97.6	97	97	97	98	99
0.4	0.1	0.06	98.4	100	98	98	98	98
0.4	0.1	0.04	97.8	96	99	98	98	98

对比图



从实验结果可以得出以下结论：

alpha 越大，对累积的经验依赖越小，从图上看，alpha 对成功率影响比较小，gamma 对成功率影响较大，gamma 越大，成功率越小，说明对新的状态和动作的 Q 值依赖越大。epsilon 增加，成功率明显降低，所以应该控制在较小的范围，尽量减小随机探索的情况。当 alpha=0.4，gamma=0.1，epsilon=0.1 是，智能车表现最好。正确率达到 98.4%。

最佳的策略就是智能车在不同的状态下，能够选择最佳动作，就是在准守交通规则的前提下，朝着终点移动。当性能比较好的时候，智能车大部分的动作能够达到最佳，但是也会有一部分动作，朝着远离终点的地方移动，有时在几个点转圈。朝着远离终点的地方移动，是由于 Q-VALUE 的状态空间在初始化的时候，有效的值比较少，Q-VALUE 尚未收敛，这种情况可以通过增加训练次数，让 Q-VALUE 状态空间收敛，最终让车朝着终点移动。原地转圈的情况，是 gamma 值过高，智能车贪吃小的奖励在原地转圈，不会直接朝着终点移动，可以通过参数

进行调节，降低 gamma 值。总之，通过对智能车的训练和调参，智能车能够选择最佳的策略移动，到达终点。