

## 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  $4 \times 4 \times 2 \times 4 = 128$  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：

$$Q(s,a)=(1-\alpha)Q(s,a)+\alpha(\text{reward}+\gamma\max_{a'}(Q(s',a')))$$

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

参数是学习率 alpha, 折扣因子 gamma 和随机参数 epsilon,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。

实验结论

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	45	25	46	32	55	67
0.2	0.1	0.1	56.4	56	68	53	50	55
0.3	0.1	0.1	39.4	42	36	52	33	34
0.4	0.1	0.1	36.8	53	31	23	33	44
0.5	0.1	0.1	36.6	39	28	46	31	39
0.6	0.1	0.1	30.6	39	30	21	39	24
0.7	0.1	0.1	24.8	22	20	26	34	22
0.8	0.1	0.1	29.8	33	23	24	28	41
0.9	0.1	0.1	25.4	31	18	32	17	29
0.2	0.2	0.1	45.8	29	37	62	51	50
0.2	0.3	0.1	45.6	61	46	39	46	36
0.2	0.4	0.1	40	52	25	48	43	32
0.2	0.5	0.1	47.6	58	52	29	42	57
0.2	0.6	0.1	56	62	56	55	53	54

0.2	0.7	0.1	60.8	49	47	61	92	55
0.2	0.8	0.1	51.4	65	58	43	53	38
0.2	0.9	0.1	31.2	43	25	24	34	30
0.2	0.7	0.2	39.6	54	19	51	28	46
0.2	0.7	0.3	28.6	32	35	19	19	38
0.2	0.7	0.08	54.2	46	41	60	63	61
0.2	0.7	0.06	58.8	62	60	64	37	71
0.2	0.7	0.04	41.6	25	31	63	66	23

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，gamma 越大，对新的状态和动作的 Q 值依赖越大，epsilon 是随机参数，应该控制在较小的范围，尽量减小探索的情况。当 alpha=0.2, gamma=0.8, epsilon=0.1 是，智能车表现最好。正确率达到 60.8%。

智能车在部分状态下，能够给出最佳的状态，不遇到任何惩罚。但是 Q 值没有学习到的状态或这初始化时选取的随机状态，或探索率，都会影响最佳策略的选择。增加一些训练数据，使所有的状态的 Q 值收敛，智能车在不同的状态下就会选择最佳策略。

