**Q-Learning代码说明**

**一、代码部分说明**

1. 环境说明：这里15个状态，0为起点，13为终点，14表示的是障碍物，16是环境边界，机器人可以在0-13中移动。

2. 动作说明：机器人可以从起始状态0运动到终止状态13，对于每一个状态，机器人都有4中动作。0-前，1-左，2-后，3-右。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 16 | 16 | 16 | 16 | 16 | 16 |
| 16 | 0 | 1 | 2 | 3 | 16 |
| 16 | 4 | 5 | 14 | 6 | 16 |
| 16 | 7 | 14 | 13 | 8 | 16 |
| 16 | 9 | 10 | 11 | 12 | 16 |
| 16 | 16 | 16 | 16 | 16 | 16 |

环境说明图

3. Q-Learning更新Q值公式:

Q(St,At) = Q(St,At) + a[Rt-1+ymaxQ(St+1,a)-Q(St,At)] = (1-a)Q(St,At) + aRt-1 + aymaxQ(St-1,a)

**二、伪代码**

定义变量：奖励表R，动作表A，位置表S，折扣因子g，学习率a，贪心方法e，训练幕数Episode ……

For大循环:

Choose\_Action(position) //通过位置，选择和返回最大Q值的动作（此处可加入贪心策略）

Run(action) //基于该动作，返回位置，奖励，判断自己有无撞到物/到达目的地

Q\_Learn（position,action,reward,next\_position） //用公式更新Q值表