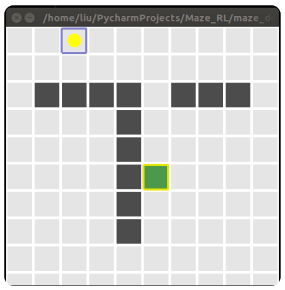

1. 环境

gym 搭迷格子环境

① reset() 结束一个 Episode 开始新训练

② step() action 对agent动作与环境互动，返回 next_state reward done

$$reward \begin{cases} goal = 10 \\ back = -1 \\ move = -0.1 \end{cases}$$

next_state agent位置

goal位置

agent和障碍物之间距离、位置信息

新物品放入状态不一致

Model 无法统一状态如何改变呢

2. QAgent

① Neural Network

relu(Dense) + relu(Dense) + Dense

② Exploitation - Exploration

$$\varepsilon_{decay} = \left(\frac{\varepsilon_{min}: 0.01}{\varepsilon: 1}\right)^{\frac{1}{MaxEpisode}}$$

$$\varepsilon \mathrel{*}= \varepsilon_{decay}$$

③ train

```
for Episode:
    reset()
    while not done:
        a0 , 按照策略选择动作.
        self.step(a0)
```

④ 经验回放

更新

$$Q_{k+1}(S_t, a_t) = Q_k(S_t, a_t) + \alpha\left(R_{k+1}(S_{t+1}, a_t) + \lambda \max_{a_{t+1}} Q_k(S_{t+1}, a_{t+1}) - Q_k(S_t, a_t)\right)$$

⑤ tensorboard 记录

loss reward 变化.