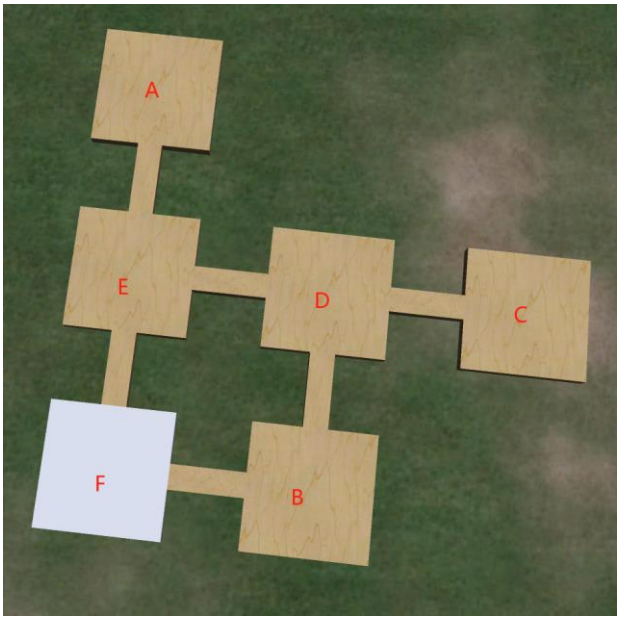
机器智能实验 6

强化学习实验报告

实验步骤：

1. 搭建房间场景，如下图所示



2. 强化学习建模

对于一个 Agent，定义它的三种状态：ready、learn、start

- ready 状态

  此状态为初始状态，Agent 的行为是初始化 Q、R 矩阵和学习系数等参数。

  在 ready 状态下用 llListen 监听指令，根据指令跳转到 learn 或者 start 状态。

- learn 状态

  此状态进行 Q 学习，当接收到学习次数 epoch 和开始学习的指令后，按照如下伪代码行动

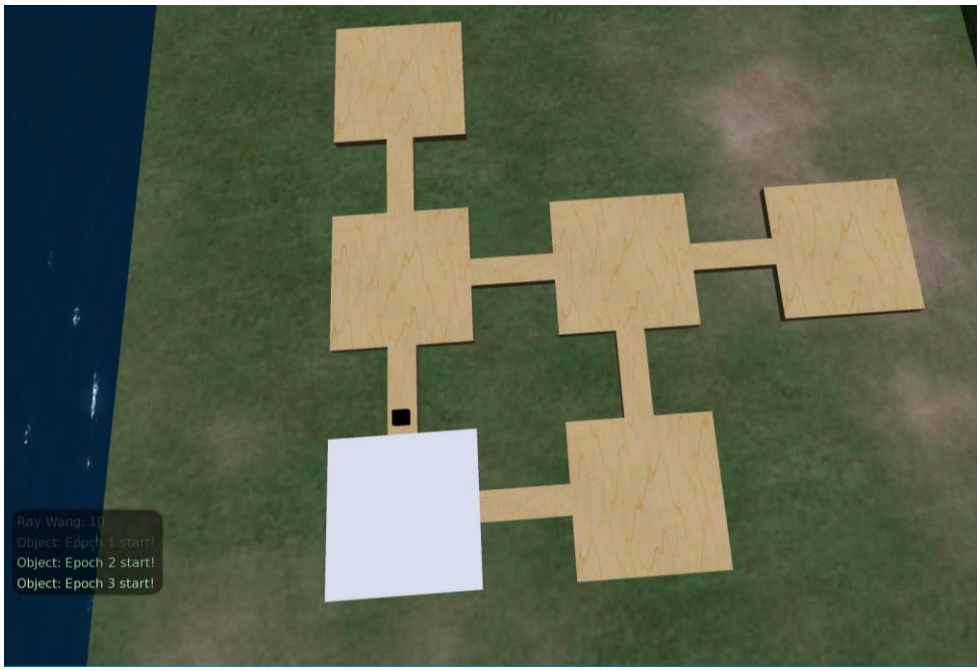
```
For each episode:
    Select random initial room
    Do while (not reach room F)
        All_next_rooms ← All rooms current room connects
        Next_room ← Random_Select(All_next_rooms)
        Compute Q[current room, next room] =
            R[current room, next room] +  γ Max[Q[next room, all actions]]
        Current_room ← Next_room
    End Do
End For
```

- start 状态

  此状态利用 Q 学习学到的知识，给定初始状态，按照最优路径前往目标 F 房间。

3. 在 Second Life 中实现建模

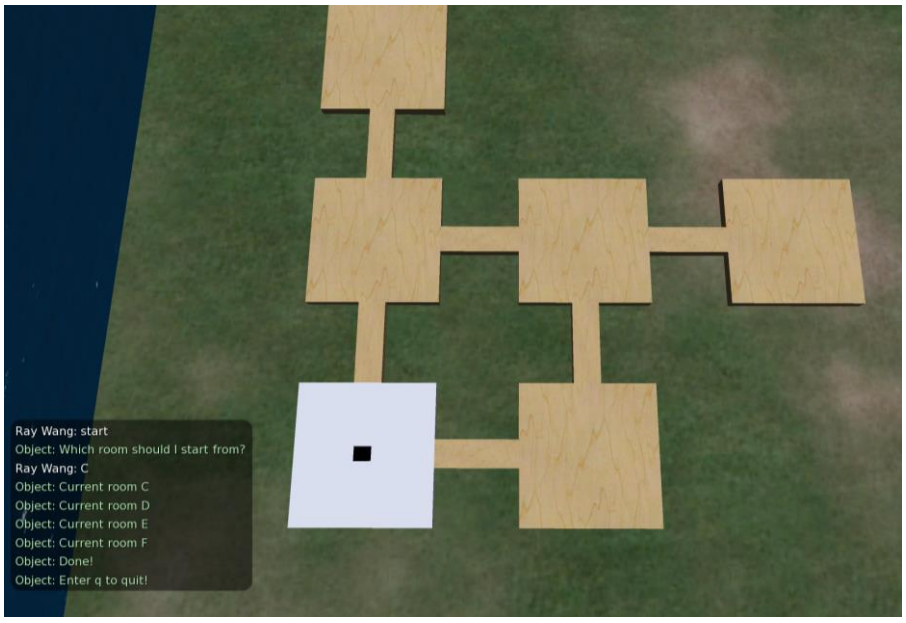
新建一个正方体 Prim，将 Qlearning.txt 的代码添加为其内容，在对话框中输入 learn，并输入 epoch 数量 50，然后 Prim 开始进行学习，如下图所示



50 轮学习完成后，Prim 输出它学习到的 Q 矩阵如下

```
Object: Learning Done!
Object: Here's the Q matrix
Object: 0.000000 0.000000 0.000000 0.000000 382.679367 0.000000
Object: 0.000000 0.000000 0.000000 306.143494 0.000000 482.679367
Object: 0.000000 0.000000 0.000000 306.143494 0.000000 0.000000
Object: 0.000000 372.936511 244.914795 0.000000 382.679367 0.000000
Object: 306.143494 0.000000 0.000000 306.143494 0.000000 482.679367
Object: 0.000000 382.679367 0.000000 0.000000 378.349209 482.679367
```

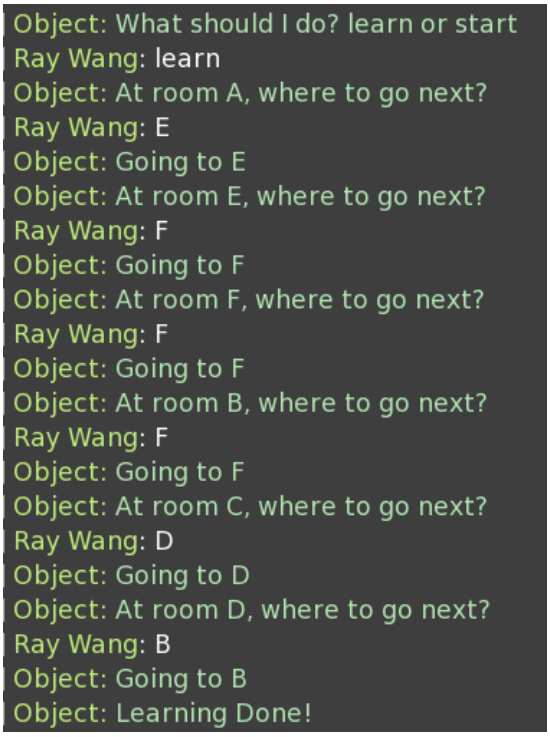
然后输入 start，并给定一个初始房间 C，Prim 每次选取 Q 矩阵中对应最大值的那个房间进入，并最终到达目标房间 F，如下图所示



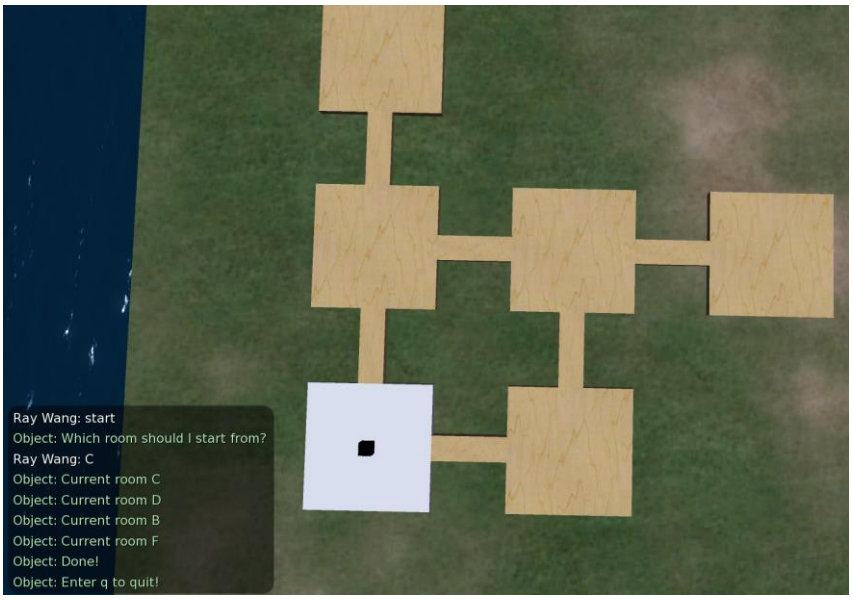
4. 监督学习的实现

Prim 每出现在一个房间里，就由 Avatar 主人作为老师给出最优路径值，将所有的状态学习完，即可开始应用规则。监督学习的代码在 SupervisedLearning.txt 中，与 Qlearning.txt 的使用方法一致。

学习的过程如下图所示



学习完成后，即可按照学习到的知识选择最优路线，如下图所示



5. 监督学习和强化学习的区别分析

5.1. 强化学习的样本通过不断与环境进行交互产生，即试错学习，而监督学习的样本由人工提供

5.2 强化学习的反馈信息只有奖励，并且是可能延迟的，而监督学习需要在每一步都做明确的指导