**20250242-0: 人工智能基础**

# 第三次编程作业：火焰杯试炼问题

主讲老师：江 瑞　　负责助教：王子安

## 1 题目介绍

图 1: 《哈利 · 波特与火焰杯》剧照

背景介绍： 1994 年，霍格沃茨正举行火焰杯三强争霸赛，第四学年的哈利 · 波特正在为最后一项迷宫挑战积极备战。作为朋友兼学霸的赫敏，为了帮助哈利锻炼在未知环境中的判断力，设计了一个特殊的魔法训练场景：一个 $5 \times 5$ 的魔法迷宫。在这个迷宫中，能见度极低——受赫敏施放的“迷雾咒”（Confundus Charm，“Confundo”）影响，哈利每次只能看见自己当前所在的格子。迷宫中藏有一个火焰杯仿制品（即胜利目标），也散布着若干魔法传送陷阱，一旦哈利踩入其中，将被立刻传送回入口重新开始；如果成功找到火焰杯，就算做完成任务。更具挑战性的是，为了防止哈利靠记忆通关，赫敏还施加了“遗忘咒”（Memory Charm，“Obliviate”），每次重新开始时，哈利都会忘记上一次遇到的陷阱与火焰杯的位置。一开始，哈利凭借勇气反复尝试，但随着训练不断重复，他开始思考有没有“聪明的方法”来找到火焰杯。他回忆起两年前在麻瓜世界看到的一篇学术论文：Watkins 与 Dayan 于 1992 年提出的 Q-learning 方法——一种即使在缺乏环境模型的情况下，也能通过经验学习找到最优路径的算法。哈利意识到，他不需要记住整个迷宫，只需要学会在每一个位置上该往哪个方向走，就能避开陷阱、找到火焰杯。现在，请你模拟哈利的思考过程，设计并实现一个强化学习智能体，使其能够在这个黑暗迷宫中成功完成火焰杯试炼。

题目叙述：

1. 哈利 · 波特身处一个 $5 \times 5$ 的魔法迷宫当中，迷宫中存在位置固定的一个火焰杯仿制品（目标状态）和若干魔法传送陷阱（陷阱状态）。如图2所示，哈利波特只能进行上下左右四个方向的移动。每次移动到一个格子上时，若该格子为陷阱位置，则他会被立刻传送回起点 $(0, 0)$，并重新开始探索。每次重新开始时，他不会记得陷阱或火焰杯的位置（环境对智能体而言是无模型的）。在此情境下，请对“火焰杯试炼问题”进行强化学习建模，写出该问题的：

   - 状态空间（State space）；

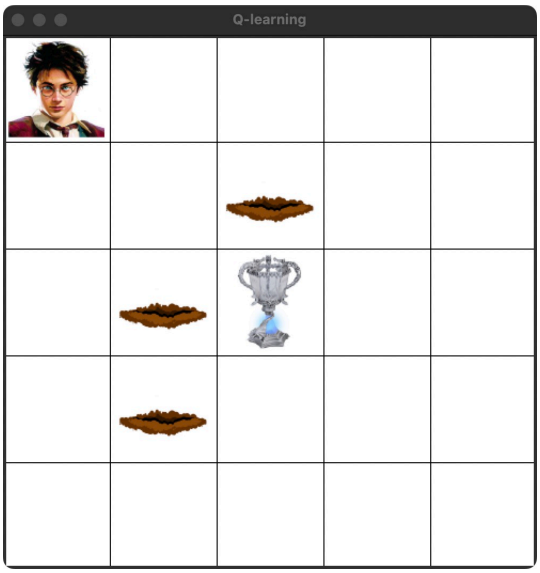


图 2: 火焰杯试炼问题

- 动作集合（Action set）;
- 状态转移概率（Transition probabilities）;
- 回报函数（Reward function）。

2. 请使用 Q-learning 算法，并利用 Python 编程语言求解“火焰杯试炼问题”。要求包括：
   - 定义 Q-table;
   - 设置适当的学习率 $\alpha$、折扣因子 $\gamma$、探索率 $\epsilon$;
   - 训练智能体，直到其能稳定找到火焰杯仿制品;
   - 可视化训练过程中的路径变化。
3. 哈利 · 波特是个非常聪明的孩子，他在阅读 Q-learning 算法的更新公式时产生了新的思考。该算法的标准更新公式为：

$$Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha \left[ R_{t+1} + \gamma \max_{a \in A} Q(S_{t+1}, a) - Q(S_t, A_t) \right]$$

   公式中，$\max_{a \in A} Q(S_{t+1}, a)$ 是一步之后的最大未来价值估计，这个算法只考虑了一步未来的回报。我们能否“看得更远”，比如估计两步的未来？为此，他提出了“2-step Q-learning”算法，它使用两步的未来回报来更新 Q 值。哈利打算利用这个改进的算法作为自己《麻瓜研究》（Muggle Studies）课程的论文主题。现在，请给出 2-step Q-learning 的更新公式，并解释每个符号的含义。请简单用语言叙述，这个公式该在什么时候更新 $t$ 时刻对应的参数 $Q(S_t, A_t)$?

## 2 作业要求

（1） 本次报告要求使用 Python 语言实现。方便起见，我们为大家提供已实现的迷宫部分代码，请你在此基础上完成强化学习的算法部分。

（2） 报告中需对所实现算法的核心代码进行解释和说明，并回答所给出的问题。

（3） 问题的分析与思考是本题的重点考核内容，建议在报告中适当展开。

## 3 提交说明

你需要写出代码，并在报告中对上述问题进行回答。提交文件格式及命名要求如下（如不按照命名规范提交会扣除少量分数）：

- 编程作业3_学号_姓名.rar（.zip）
  - hw3_学号.py（代码）
  - report3_学号_姓名.pdf（pdf版报告）
  - cup.png
  - potter.png
  - trap.png

本次作业截止日期：2025 年 6 月 4 日晚 12 点（三周后）