

强化学习-2021秋-课程作业二

作业内容

实现Q-learning算法

作业描述

环境描述

本次作业的环境为网格世界(gridworld)，玩家可以通过选择动作来移动人物，走到出口。如下图，人形网格是由玩家控制的单位，白色砖状网格为不可通行的墙壁，右下角的褐色网格是出口，其余黑色部分表示可以通行的通道。



玩家得到的观测：一个二维数组(x,y)，表示玩家所处的位置坐标。

可执行的动作：{0, 1, 2, 3}分别表示上下左右四个方向移动。

奖励：玩家在游戏中每移动一步会得到-1的奖励，走到出口时将额外得到100的奖励。

游戏目标：尽可能达到高的累计奖励/移动小人以尽可能少的步数到达出口。

任务描述

请完成：

1. 与环境交互，在环境中采样并记录轨迹。
2. 依据Q-learning算法，学习一个移动小人走到出口的策略。
3. 绘制你实现的Q-learning算法的性能图（训练所用的样本与得到的累计奖励的关系图）

代码描述

代码文件夹code由'main.py', 'arguments.py', 'algo.py', 'env.py' 组成。

'main.py': 包含了代码的主要结构，包括环境初始化、如何与环境交互的样例等等。**你需要在其中实现Q-learning算法的相关部分**，并用你的策略(agent.select_action)来玩游戏并进行性能测试。

'arguments.py': 包含了默认的参数，可以修改。

'algo.py': 包含了待填充的Q-learning算法QAgent，**请继承其中的QAgent来实现你自己的算法。**

'env.py': 包含了内置的环境，请勿修改。

运行代码前请安装：

numpy、argparse、pickle、gym、matplotlib、PIL

提交方式

完成的作业请通过sftp上传提交。上传的格式为一份压缩文件，命名为'学号+姓名'的格式，例如'MG21370001张三.zip'。文件中需包含 'main.py', 'arguments.py', 'algo.py', 'env.py', 'performance.png' 和'Document.pdf'（一份pdf格式的说明文档），文档内容至少需要包含：

1. 实验效果说明。
2. 如何复现实验效果。
3. Q-learning算法的实现说明。
4. 如果有相关的改进，也请在其中说明。

文档模板参见'Document2.tex'和'Document2.pdf'。（也可以使用自己的模板。）