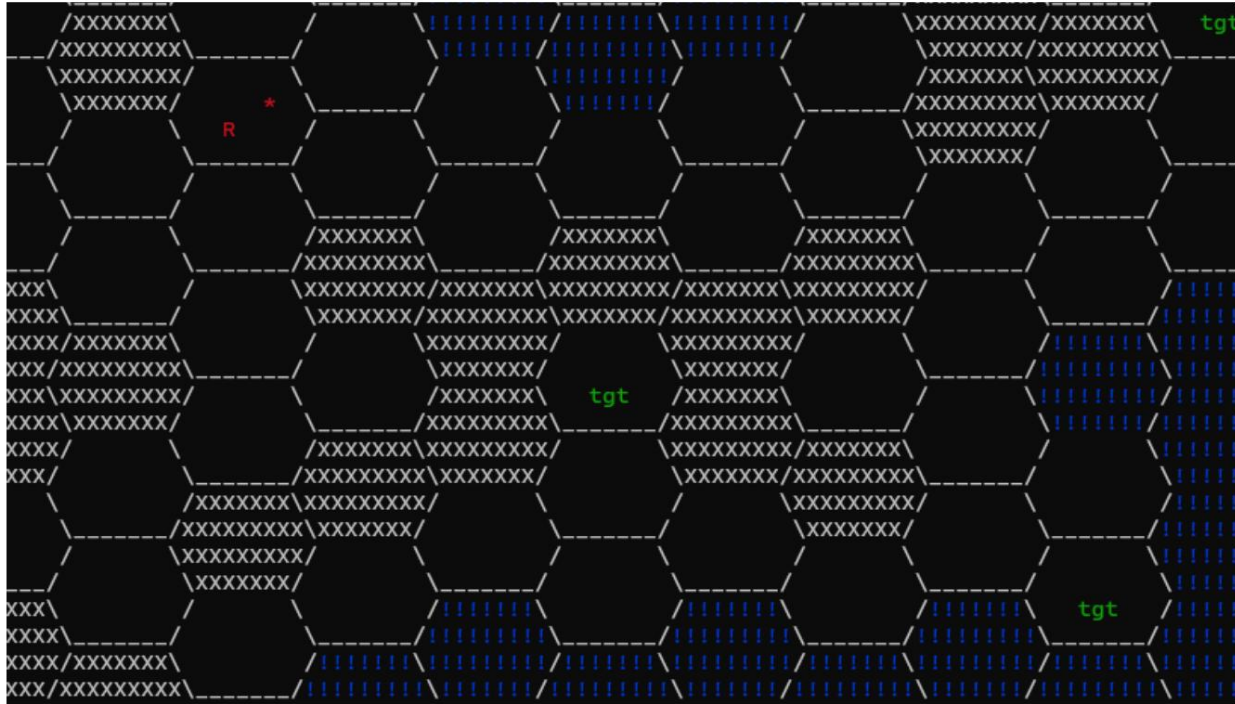


HexBot 机器人环境 (A3 更新)

COMP3702 人工智能 2022



您的任务是开发一种用于自动控制 HexBot 的**强化学习**算法,HexBot 是一种在六边形环境中运行的多功能机器人。**对于此版本的任务,HexBot 必须导航到指定目标之一,同时尽可能降低罚分。不需要考虑小部件。**为了帮助您完成这项任务,我们为 HexBot 机器人环境提供了一个模拟器和可视化,您将与之交互以开发您的解决方案。

对于 A3,HexGrid 环境具有不确定的动作结果,其概率未知,每个测试用例都是随机的。每个可用操作的成本以及与障碍物和危险碰撞的惩罚也是未知的,并且对于每个测试用例都是随机的。本文档的更新以洋红色文本显示。

六角网格

环境由六边形网格表示。十六进制网格的每个单元格都由 (行、列) 坐标索引。十六进制网格从上到下、从左到右进行索引 (即左上角的坐标为 (0, 0), 右下角的坐标为 (n_rows-1, n_cols-1))。

偶数列 (从零开始) 位于行的上半部分,奇数列位于行的下半部分。例如

第 0 行,第 0 列	第 0 行,第 1 列	第 0 行,第 2 列	第 0 行,第 3 列	...
第 1 行,第 0 列	第 1 行,第 1 列	第 1 行,第 2 列	第 1 行,第 3 列	...
...

如果六角网格中的两个单元共享一条边,则认为它们是相邻的。对于每个无边界单元,有6个相邻的单元。

机器人

HexBot 机器人占据六角网格中的一个单元。在可视化中,机器人是由标有字符 “R”的单元表示。标有 的单元一侧代表机器人的正面。机器人的状态由其 (行、列)定义坐标及其方向(即其正面指向的方向)。

机器人有 4 个可用的标称动作:

- 前进 → 向机器人前方方向移动到相邻单元 (保持相同的方向)
- 反向 → 以与机器人前方相反的方向移动到相邻单元 (保持相同的方向)
- 向左旋转 → 向左旋转 (相对于机器人正面,即逆时针)60 度 (住在同一个牢房里)
- 向右旋转 → 向右旋转 (即顺时针)60 度 (留在同一个单元中)

机器人每次选择一个动作,都有一个固定的概率 (根据每个测试用例的种子)让机器人顺时针或逆时针 “漂移”60度方向 (每个漂移方向的单独概率)在选定的标称动作之前执行。漂移发生的概率取决于选择的标称动作,其中一些动作更容易导致漂移。顺时针漂移和逆时针漂移是互斥的事件。

此外,还有一个固定的概率 (也是根据每个测试用例的种子随机设置的)让机器人 “双重移动”,即执行两次名义上的选定动作。概率发生的双重移动取决于选择了哪个动作。可能发生双重运动同时与漂移 (CW或CCW)。

每次行动后获得的奖励是奖励中最小/最负的收到的名义行动和任何额外的 (漂移/双重移动)行动。

障碍

六角网格中的一些单元格是障碍物。在可视化中,这些单元格填充有字符“X”。任何导致机器人或 Widget 的任何部分进入障碍物单元的动作都会导致碰撞,从而导致代理收到负的障碍物碰撞惩罚作为奖励。

该奖励取代了代理否则将产生的移动成本。六边形网格的外部边界与障碍物的行为方式相同。

此外,环境现在包含一种额外的障碍物类型,称为“危险”。

危险与障碍物的行为方式相同,但是当发生碰撞时,会收到不同的(更大的)惩罚作为奖励。因此,避免与危险碰撞比避免与障碍物碰撞更重要。危险用“!!!”表示在可视化中。

目标

十六进制网格包含许多“目标”单元格。在可视化中,这些单元格标有“tgt”。对于被认为已解决的 HexBot 环境,目标单元之一必须被 HexBot 占据。环境可能包含多个目标。