强化学习第一次作业

提交作业时间：2024.11.8 22:00 前

作业描述：使用程序语言（建议 C++）实现以下算法

1. 使用两种方法求解贝尔曼方程：封闭式求解(closed-form solution)方法，$v_\pi = (I - \gamma P_\pi)^{-1} r_\pi$；迭代式求解(iterative solution)方法，$v_{k+1} = r_\pi + \gamma P_\pi v_k$。使用以下四种策略（如图 1 所示，设 $r_{\text{boundary}} = r_{\text{forbidden}} = -1$, $r_{\text{target}} = +1$, $r_{\text{otherstep}} = 0$, $\gamma = 0.9$），验证所实现的代码。提交 Word 文档内容包括：a. 设计思想；b.伪码描述；c.时间复杂度分析；d.源码 zip 压缩包。

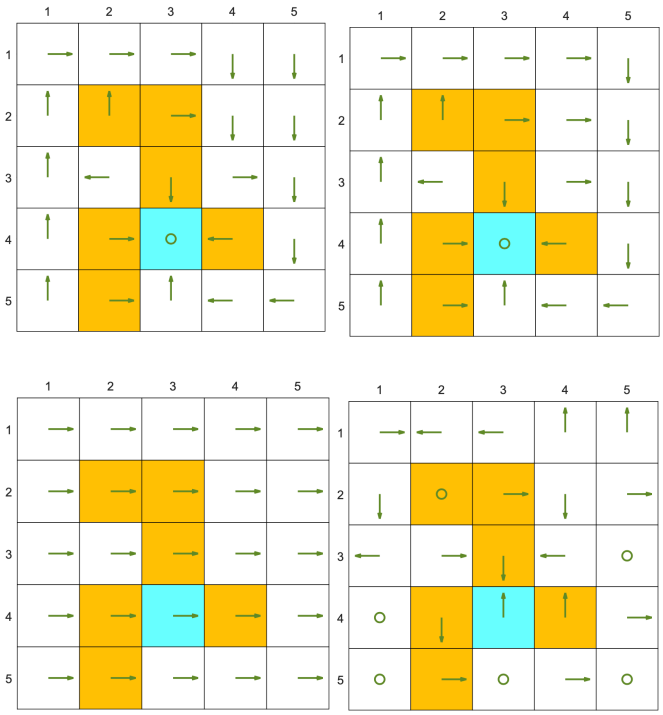

图 1. 四种策略

2. 使用迭代算法求解贝尔曼最优方程，使用以下一个简单例子进行验证。

$$v_{k+1} = f(v_k) = \max_{\pi}(r_\pi + \gamma P_\pi v_k)$$

系统（环境）：1×3 网格世界（如图 2 所示）；动作： a_l，a_0，a_r 分别代表向左走、保持不变、向右走；奖励：进入目标区域为+1； 尝试走出边界为-1；其他为零。

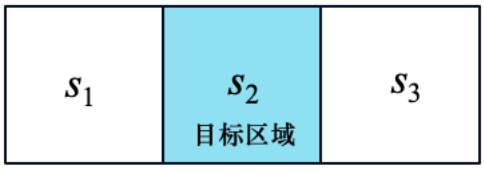

图 2. 1×3 网格世界

分析最优策略受到折扣率从 γ 以及奖励设计 r 的影响

系统（环境）：如图 1 所示的 5×5 网格世界；给出以下五种设置的最优状态值和最优策略：(a) $r_{\text{boundary}} = r_{\text{forbidden}} = -1, r_{\text{target}} = 1, r_{\text{otherstep}} = 0, \gamma = 0.9$; (b) 折扣率为 $\gamma = 0.5$, 其他与(a)相同。(c) 折扣率为 $\gamma = 0$, 其他同(a)；(d) $r_{\text{forbidden}} = -10$，其他与($a$)相同。(e) $r_{\text{boundary}} = r_{\text{forbidden}} = 0, r_{\text{target}} = 2, r_{\text{otherstep}} = 1$, 其他同($a$)。

提交 Word 文档内容包括：a. 设计思想；b.伪码描述；c.时间复杂度分析；d. 描述折扣率从 γ 以及奖励设计 r 对于最优策略的影响；e.源码 zip 压缩包。