



RLChina 2021

习题课1

课程概况

林舒

中国科学院自动化研究所

2021年8月16日

课程简介

RLChina 2021 强化学习暑期课习题课

- 今年暑期课新增的重要部分
- 以“演示+动手+答疑”的形式，帮助大家更近距离接触强化学习落地实践
- 提供了详细的步骤演示，引导了解各类经典的强化学习算法，并亲自动手在特定环境下进行训练和评测
- 时间：2021年8月16-20日，每晚19:00-20:00
- 主讲人：林舒，杨梦月

课程安排

序号	课程内容	作业环境	作业算法	作业要求
1	强化学习基础及平台介绍	推箱子	随机	提交并验证通过
2	基于表的强化学习算法	悬崖漫步	Q-learning SARSA	成绩优于随机
3	基于神经网络的强化学习算法	车杆	DQN	成绩优于随机
4	多智能体合作	贪吃蛇5P	IQL	成绩优于随机
5	多智能体对抗	贪吃蛇1V1	任意	参与贪吃蛇1V1擂台 成绩优于随机

课程形式

每节课内容安排如下：

1. 介绍本次课使用的算法理论，以及环境规则
2. 现场演示如何在该环境下训练该算法
3. 学生自己动手完成算法的本地/线上训练
4. 答疑环节（关注RLCN微信公众号后，发消息进行提问）

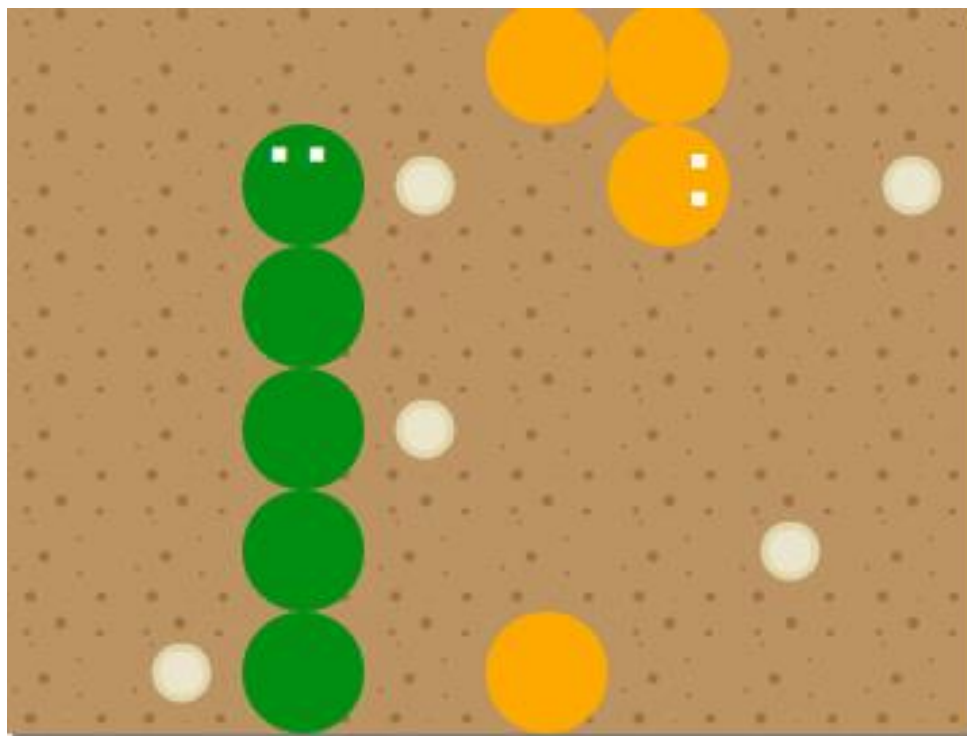
结课要求

- 在规定时间内（**8月27日23:00**）前完成所有五次作业
- 提交的算法必须符合相应作业的要求



习题课大擂台

- 习题课最后一次作业是贪吃蛇1v1擂台
- 名列前茅的学员可得到认证徽章



课程平台

- 及第Jidi: 用于作业（包括擂台）的发布、提交及评测
 - <http://www.jidiai.cn/homepage>
- 动手学强化学习: 提供了强化学习的在线教程
 - <http://hrl.boyuai.com/>
- 和鲸平台: 为线上训练提供了工作台和算力
 - <https://www.heywhale.com/>
- RLChina微信公众号: 用于提问答疑和问题反馈



RLChina 2021

习题课1

强化学习基础

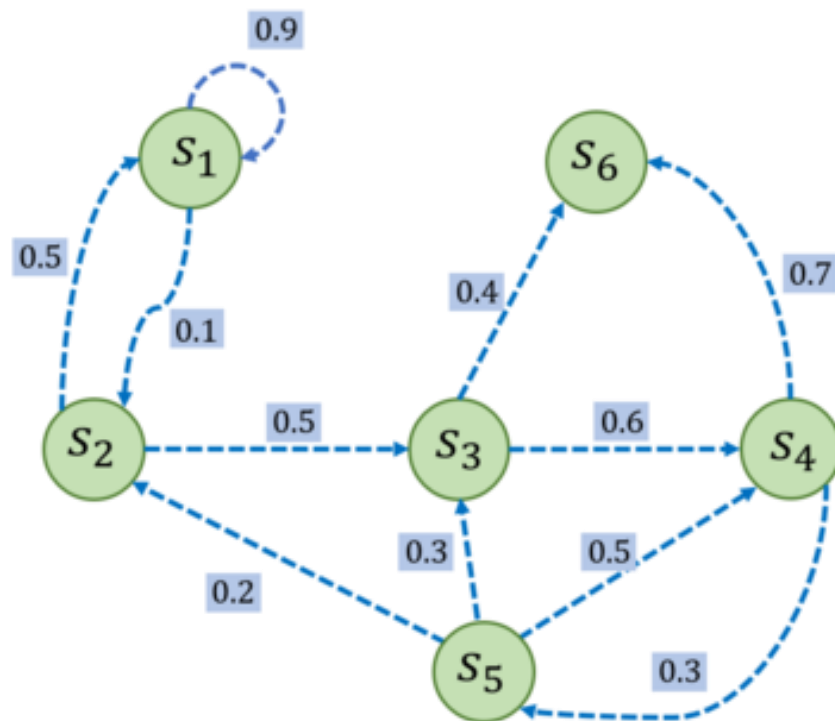
林舒

中国科学院自动化研究所

2021年8月16日

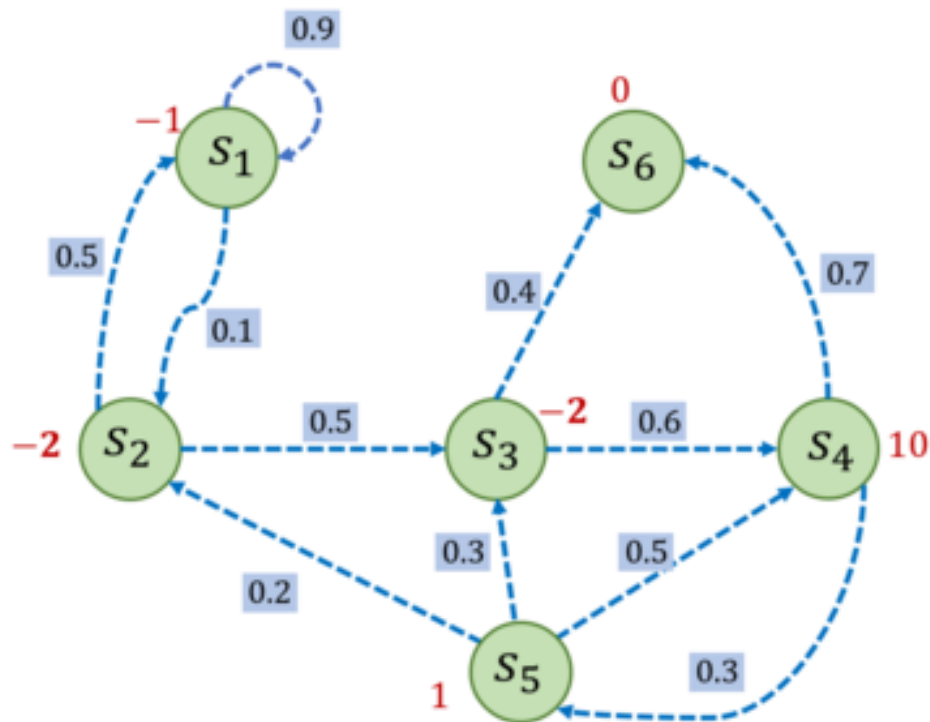
* 课程内容参考《动手学强化学习》<http://hrl.boyuai.com/>

马尔可夫过程 Markov Process



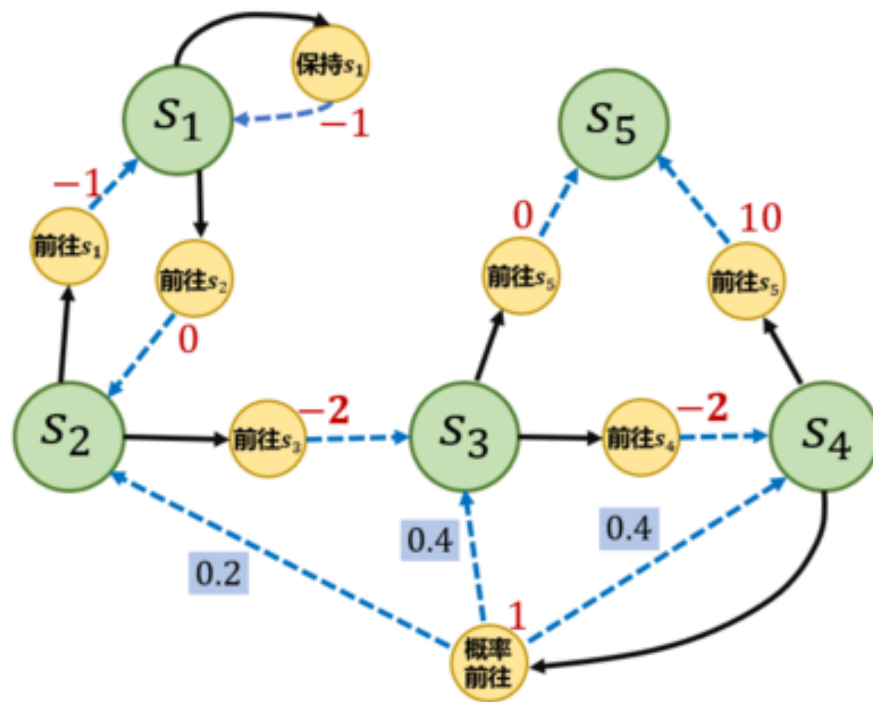
- $MP = \langle S, p \rangle$
 - 状态集 $S = \{s_1, s_2, \dots, s_n\}$
 - 状态转移函数 $p(s' | s) = \Pr(S_{t+1} = s' | S_t = s)$

马尔可夫奖励过程 Markov Reward Process



- $MRP = \langle S, p, r, \gamma \rangle$
 - 状态集 $S = \{s_1, s_2, \dots, s_n\}$
 - 状态转移函数 $p(s' | s) = \Pr(S_{t+1} = s' | S_t = s)$
 - 奖励函数 $r(s)$
 - 折扣因子 $\gamma \in [0, 1]$

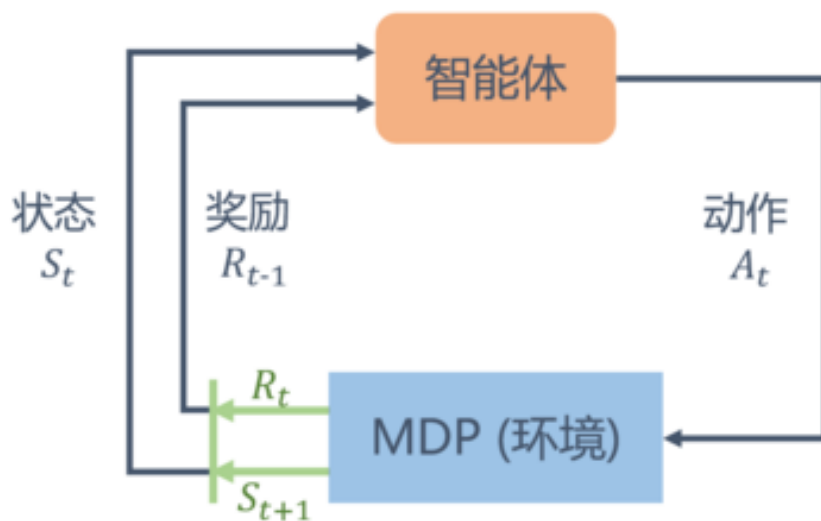
马尔可夫决策过程 Markov Decision Process



- $MDP = \langle S, A, p, r, \gamma \rangle$
 - 状态集 $S = \{s_1, s_2, \dots, s_n\}$
 - 动作集 $A = \{a_1, a_2, \dots, a_m\}$
 - 状态转移函数 $p(s' | s, a) = \Pr(S_{t+1} = s' | S_t = s, A_t = a)$
 - 奖励函数 $r(s, a)$
 - 折扣因子 $\gamma \in [0, 1)$

智能体与环境交互

智能体策略 $\pi(a | s) = \Pr(A_t = a | S_t = s)$



示例：随机策略 $\bar{\pi}(a | s) = \begin{cases} 1/k_s, & a \text{ 是 } s \text{ 下的合法动作} \\ 0, & a \text{ 不是 } s \text{ 下的合法动作} \end{cases}$

注： k_s 是 s 下的合法动作数



RLChina 2021

习题课1 及第 平台介绍

林舒

中国科学院自动化研究所

2021年8月16日

丰富的
环境

优质的
竞赛

实时的
讨论

公正的
排行

前沿的
算法

及第 JDI
在线算法对抗平台

及第——开始您的AI晋级之路

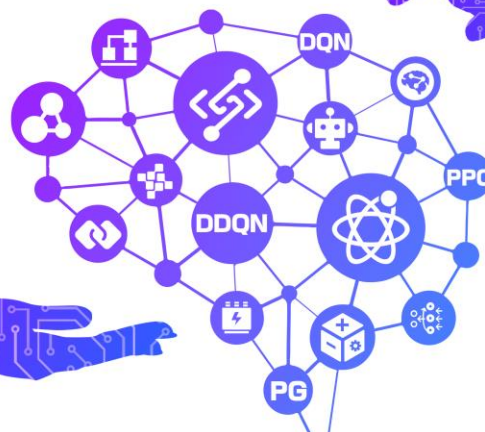
丰富的博弈环境

Rich Game Environments



前沿的开源算法

Advanced Open Source Algorithms



公正的智能体排行

Fair AI Rankings



开放的智能体竞赛

Open AI Competitions



及第主页



及第功能

- 金榜：各游戏环境的排行榜以及总排行榜
 - 科目：不同游戏环境的环境介绍、提交接口等
 - 秘籍：各种经典算法的原理介绍、复现代码等
 - 擂台：各类竞赛的竞赛说明、赛程安排等
 - 论道：实时讨论区
 - 赶考：新手引导、常见问题等
-
- 所有环境和算法代码及详细说明均已发布到Github仓库：
https://github.com/jidiai/ai_lib

第一次作业：完成新手任务1

1. 注册、登录

2. 填写真实姓名

① 右上角头像→“个人中心”

② 左边栏“个人信息”右侧的编辑按钮

③ 填写真实姓名，点击“修改”

3. 进入“推箱子(1p)”科目

导航栏“科目”→“推箱子(1p)”

4. 下载样例Random算法

“提交说明”→“提交样例”→复制“submission.py”到本地

5. 提交样例Random算法

重新进入“推箱子(1p)”科目→右侧“提交-推箱子(1p)”

输入“算法名称”→上传“submission.py”→点击“提交”

6. 查看结果

经过一段时间后，可在“个人中心”中查看验证结果

若验证通过，可以在“金榜”中看到自己的排名