《强化学习》课程大作业说明

背景

深度强化学习是一种通过环境交互进行学习的方法。它类似于人类的学习过程，根据外界反馈进行学习。因此，它在物理世界得到广泛应用，例如机器人、互联网、自动驾驶等。

本课程以引导学生将强化学习的相关理论应用到科研中为目标，该目标要求学生同时掌握强化学习的算法理论以及相关算法的实现与应用。本学期的课程已经介绍了许多经典的深度强化学习理论，包括深度神经网络、多臂老虎机、马尔科夫决策过程、动态规划法、蒙特卡洛法、时分求解法等。但是，仅仅通过课堂上的理论学习和课下作业是无法满足本课程的目标的。因此本课程设计了课程大作业，意图通过给定的应用场景锻炼学生的应用与实践能力，从而为后续研究工作打下坚实的基础。

考虑到课程中的学生的研究方向不同，若针对某特定领域可能会增加学生理解应用场景的时间成本，这与本课程的目标是背道而驰的。因此本课程设计一种几乎所有人都耳熟能详的无禁手五子棋(Gomoku)作为应用场景。无禁手五子棋是一种两人对弈的纯策略型棋类游戏，通常双方分别使用黑白两色的棋子，下在棋盘直线与横线的交叉点上，先形成5子连线者获胜。通过已有的Botzone平台，学生可以在平台上运行自己的模型，并通过和其他同学的模型进行排位赛从而量化模型的性能。Botzone平台可以通过如下链接进入：https://www.botzone.org.cn/game/Gomoku

结合本次大作业的应用实例，学习强化学习相关方法的使用，从而使学生具备使用强化学习工具开展研究或应用的良好知识基础。在实现大作业的过程中，学生往往需要阅读大量的前沿论文，这将加深学生对强化学习基础理论的理解。

无禁手五子棋(Gomoku)在Botzone平台的运行规则

五子棋运行规则

- 棋盘大小为15*15。
- 黑白双方轮流落子。黑方为先手。
- 首先在横、竖、斜方向上成五（连续五个己方棋子）者为胜。超过五子也算。

具体详情请见https://wiki.botzone.org.cn/index.php?title=Gomoku#.E6.B8.B8.E6.88.8F.E8.A7.84.E5.88.99

Bot运行规则

详情请见https://wiki.botzone.org.cn/index.php?title=Bot#.E4.BA.A4.E4.BA.92

- 请注意Botzone对时间的限制以及不同库的版本支持问题。

代码样例

BotZone交互接口使用方式Demo

```
# BotZone
# 基于随机走子的Gomoku的交互Demo
```

```
3
4  import json
5  import numpy
6  import random
7
8  SIZE = 15
9
10 # 放置棋子
11 def place(board, x, y):
12     if x >= 0 and y >= 0:
13         board[x][y] = True
14
15 # 随机产生决策
16 def randplace(board):
17     empty_grid = []
18     for x in range(SIZE):
19         for y in range(SIZE):
20             if not board[x][y]:
21                 empty_grid.append((x, y))
22     return random.choice(empty_grid)
23
24 # 处理输入，还原棋盘
25 def restoreBoard():
26     fullInput = json.loads(input())
27     requests = fullInput["requests"]
28     responses = fullInput["responses"]
29     board = numpy.zeros((SIZE, SIZE), dtype=numpy.bool)
30     turn = len(responses)
31     for i in range(turn):
32         place(board, requests[i]["x"], requests[i]["y"])
33         place(board, responses[i]["x"], responses[i]["y"])
34     place(board, requests[turn]["x"], requests[turn]["y"])
35     return board
36
37
38 board = restoreBoard()
39 x, y = randplace(board)
40 print(json.dumps({"response": {"x": x, "y": y}}))
```

Baseline样例

1. 该Baseline采用MCTS，仅作为算法实现参考。
2. 该实现同时也作为在Botzone平台上多文件上传的样例。
3. 下载地址为https://disk.pku.edu.cn:443/link/3DDDE27222AFFD5A30C5DFF75B0F0E11

大作业要求

- 原则上不超过4人一组，在Botzone上注册账号。考虑到隐私因素，昵称为RL2022+组名。
- 在Botzone平台上提交至少一次比赛，作为模型性能参考。该参考作为大作业得分的重要参考。若不在Botzone提交，大作业记为0分。若因Botzone平台资源限制，算法无法达到最优性能，请在大作业报告中体现。
- 针对本组做的相关内容做ppt展示，具体时间定为12.22、12.29两周。若时间有变化，则另行通知。
- 该作业最终提交代码、书面课程报告、Botzone的比赛记录、展示的ppt。提交日期另行通知。
- 书面报告需要写清成员分工、工作量占比、算法具体实现与性能。

- 若有调整将在群里另行通知。