

北京大学2024秋季强化学习课程大作业：双升AI

作业要求

利用提供的双升强化学习环境和代码框架，训练双升游戏AI。

准备工作

- 各队伍注册Botzone账号并加入小组“北京大学2024秋-强化学习双升大作业”（需填写表单）
(Botzone使用方法请见课程群内大作业说明ppt)
- 了解双升游戏规则和Botzone双升交互方式 <https://wiki.botzone.org.cn/index.php?title=%E5%8F%8C%E5%8D%87>

环境介绍

- 提供一个双升强化学习的交互环境和训练框架，支持出牌阶段策略的训练。环境内已实现一个PPO算法和一个包含三层卷积的策略网络。
- 交互环境的详细信息请见代码包内的环境使用说明。
- 本次作业使用Botzone双升(Tractor)作为双升AI的评测环境。

作业内容

本次作业中，需要大家完成的工作包括：

- 环境建模。双升是一个多人多轮多阶段游戏，提供的交互环境仅包含单轮出牌阶段的决策和反馈，同学们可以使用给定的环境进行训练，也可以自行对问题进行建模，修改已有交互环境或构造新的环境。
- 特征处理。给定的训练框架中采用最基本的游戏信息来训练策略网络，在训练过程中引入新的特征（例如，其他玩家的手牌等实际对局中可能不可见的特征），或利用人类经验显式地发掘一些潜在特征（例如结合对手决策和规则判断出其他玩家局部的手牌信息）可能对提升智能体决策水平有帮助。同学们可以使用给定的特征输入，也可以自行探索和设计一些特征及特征结构来提高AI的表现。除了输入输出特征外，奖励函数也是一个可优化的重点。
- 算法设计。提供的代码中实现了一个PPO算法，同学们可以使用给定的算法，也可以自主尝试其他的强化学习算法。在时间有余裕的情况下，建议大家多尝试几种算法，比较不同算法在训练和决策时的性能，并写进报告中。
- 模型探索。代码包中的模型为三层卷积，结构较为简单。同学们可以使用原有的模型，也可以尝试不同的模型结构并比较不同模型的表现，选择最优的模型进行决策或在决策时参考多个模型输出的行动。由于Botzone用户空间有限，请注意控制模型的参数量。
- 参数调优。在训练过程中，可以对各种参数进行调整来最大化训练效率，节省训练时间和提升模型性能。我们鼓励同学们积极探索合适的训练参数，并把探索的过程和结果写入报告中。
- 策略创新。同学们可以在强化学习的基础上融合其他的基于规则、搜索或学习的方法来提升bot的表现，也可以提出自己的新想法。由于双升是一个2v2游戏，同学们也可以灵活地利用双升游戏规则的特点来构建配合更默契的AI。

提交内容

- 每支队伍提交一个bot参加Botzone的对战评测
- 每支队伍在教学网上提交一份报告（队长提交即可，队员无需重复提交）

评测方式与评分标准

- Botzone评测（70%）

所有队伍的bot参与Botzone评测小组的积分赛，按照最终排名给分。积分赛每一局调用两支队伍提交的bot，每个bot控制同一方的两个玩家（两个玩家在信息上是独立的）。积分赛的对局为单轮模式，不考虑升级。

最终的积分赛前，会组织多次练习赛供大家了解bot的水平，练习赛成绩不计入总成绩。积分赛时间暂定于12月29日晚23:55。

- 实验报告（30%）

在实验报告中，我们要求各队伍必须说明：

- 使用的环境。请注明使用的环境是原始环境，或介绍对环境所作改动的思路。
- 特征处理。请说明算法对特征作了何种处理，模型的输入/输出结构、奖励函数等。
- 使用的算法。请注明使用的算法和实现过程，若对算法有创新，请介绍算法设计的思路。
- 使用的模型。（如使用了神经网络）请注明使用的模型结构，以及为什么选择了这种模型结构。
- 其他创新点。如果使用了其他方法（例如基于规则或搜索的方法），请额外介绍实现的过程和思路。如果对算法、环境、模型等有特殊的创新，请特别说明。
- 运行参数。请说明在训练模型的过程中使用的机器配置、训练时间及模型的参数大小。

在实验报告中，我们建议各队伍可以额外介绍：

- 探索过程和经验。在算法/模型/参数/特征的设计上，经历了哪些探索，尝试了哪些不同的设计/组合？这些探索提供了什么样的经验？
- 遇到的问题和解决方法。在构建环境/训练模型的过程中，遇到了哪些困难，最终是如何解决的，进行了哪些尝试，有什么样的发现/获得了什么样的经验？
- 分析和思考。对实验过程中出现的各种现象，或者对自己和其他队伍Bot的表现有没有观察或留意？有没有有价值的分析和发现？

实验报告不限语言，不限格式，篇幅尽量不要超过五页。有条件的队伍可以用图表等方式辅助说明。

两人队伍请在实验报告末尾注明两人的分工，这一部分不包含在篇幅限制内。