中国科学技术大学

University of Science and Technology of China

人工智能原理与技术

实验 4

姓名：高茂航

学号：PB22061161

日期：2024.7.6

实验 4

1 Torch 配置

```
PS F:\Desktop\Learning\Computer\AI\AI-Homework\Lab4 - Project> nvidia-smi
Mon Jul  1 21:42:46 2024
+---------------------------------------------------------------------------------------+
| NVIDIA-SMI 531.14                 Driver Version: 531.14       CUDA Version: 12.1     |
|-----------------------------------------+----------------------+----------------------+
| GPU  Name                      TCC/WDDM | Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp  Perf            Pwr:Usage/Cap|         Memory-Usage | GPU-Util  Compute M. |
|                                         |                      |               MIG M. |
|=========================================+======================+======================|
|   0  NVIDIA GeForce RTX 3050 L...  WDDM | 00000000:01:00.0 Off |                  N/A |
| N/A   40C    P0              12W /  N/A|      0MiB /  4096MiB |      0%      Default |
|                                         |                      |                  N/A |
+-----------------------------------------+----------------------+----------------------+

+---------------------------------------------------------------------------------------+
| Processes:                                                                            |
|  GPU   GI   CI        PID   Type   Process name                            GPU Memory |
|        ID   ID                                                             Usage      |
|=======================================================================================|
|  No running processes found                                                           |
+---------------------------------------------------------------------------------------+
```

```
PS F:\Desktop\Learning\Computer\AI\AI-Homework\Lab4 - Project> python
Python 3.9.18 (main, Sep 11 2023, 14:09:26) [MSC v.1916 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license" for more information.
>>> import torch
>>> if torch.cuda.is_available():
...     print("current device: cuda.")
... else:
...     print("CUDA is not available.")
...
current device: cuda.
```

2 实验原理

2.1 二元零和马尔可夫博弈

一个涉及两个玩家在一系列状态中进行决策的过程，其中每个玩家的目标是最大化自己的累计奖励（或最小化对方的累计奖励），而这个过程的状态转移遵循马尔可夫性质。在每个状态，玩家需要基于当前的信息选择最优策略，而这个选择会影响到游戏的下一个状态和玩家的即时奖励。由于是零和博弈，一个玩家的收益等于另一个玩家的损失，使得这种博弈具有高度的竞争性。

2.2 Naive Self Play

让 AI 与其自身的副本进行对战，通过这种方式根据自身经验不断学习适应并调整，以提高其性能。该方法较为简单直接，如果要避免不收敛等问题，需要采取更复杂的算法。

2.3 Actor-Critic

这种方法旨在通过同时学习一个策略（即 Actor）和一个值函数（即 Critic）来平衡探索和利用，从而提高学习效率和性能。Actor 根据当前策略选择动作，并执行该动作。Critic 评估执行动作后的结果，即计算实际回报与预期回报之间的差异（TD 误差）。接着使用 TD 误差来更新 Critic 的值函数参数，使其预测更加准确。最后根据 Critic 提供的反馈（TD 误差），更新 Actor 的策略参数，以便在未来选择更好的动作。