# 1 项目文件结构

- `root`：
    - `README`
        - `README.md`
    - `log_plotter.py`：用于：
        - 将 *value-based* 方法中的 *Dueling DDQN* 和 *Dueling DQN* 模型在训练过程中的 `log_score.txt` 和 `log_loss.txt` 绘制成变化曲线，存储在 `./output/value_based/` 目录下；
        - 将 *policy-based* 方法中的 *TD3* 和 *DDPG* 模型在 *Humanoid-v2* 环境下训练过程中的 *avg reward in evaluation* 绘制成变化曲线，存储在 `./output/policy_based/` 目录下；
    - `output`：
        - `value_based`：存储 *log_plotter.py* 绘制的 *value-based* 相关的图像
        - `policy_based`：存储 *log_plotter.py* 绘制的 *policy-based* 相关的图像
    - `value_based`：
        - `DuelingDQN` / `DuelingDDQN`：
            - `log`：用于记录训练过程中的 *score* 和 *loss* 等信息
            - `model`：用于保存训练得到的 `Dueling_DQN_breakout.pkl` / `Dueling_DDQN_breakout.pkl` 模型
            - `atari_playground.py`：主函数
            - `RL_brain.py`：实现 `Dueling_DeepQNetwork` / `Dueling_Double_DeepQNetwork` 类
            - `atari_wrappers.py`：*Atari* 环境的封装函数
            - `utils.py`：用于模型和训练信息保存的工具函数
    - `policy_based`：
        - `TD3_results`：保存了四种 *MuJoCo* 环境下的训练信息、绘制的曲线以及模型
            - `human`：
                - `TD3` / `DDPG`：保存 *TD3 / DDPG* 算法的模型结果
                    - `models`：模型
                    - `output`：输出的曲线图片
                    - `results`：输出的用于绘制曲线的数组
            - `ant` / `halfcheetah` \ `hopper`：
                - `models`：模型
                - `output`：输出的曲线图片
                - `results`：输出的用于绘制曲线的数组
        - `TD3`：*policy-based* 模型的主文件夹
            - `main.py`：主函数
            - `TD3.py`：*TD3* 类的实现
            - `DDPG.py`：*DDPG* 类的实现
            - `utils.py`：*ReplayBuffer* 类的实现
            - `models`：模型
            - `output`：输出的曲线图片

        - `results`：输出的用于绘制曲线的数组

---

# 2 项目使用方法

## 2.1 *value-based*

- 假设现在想要运行 *Dueling DDQN* 模型，*Dueling DQN* 模型同理
- 首先进入所创建的 `conda python3.7` 虚拟环境中，进入到 `./value_based/Dueling_DDQN/` 目录下，使用如下指令开始：

```
python atari_playground.py --env_name BreakoutNoFrameskip-v4
```

- 运行过程中的 *evalueate* 会周期性输出三个值，分别是：*evaluate episode*、该 *evaluate* 周期内的平均 *reward*，以及该 *evaluate* 周期内的最大 *reward*
- 模型的状态 `dict` 被保存，因此若想测试现有模型，可以直接将 `atari_playground.py` 中的模型初始化替换为模型的加载

## 2.2 *policy-based*

- 首先进入所创建的 `conda python3.7` 虚拟环境中，进入到 `./policy_based/TD3/` 目录下，使用如下指令开始，该方法会使用 *TD3* 算法在 *Humanoid-v2* 上进行训练，并保存模型；

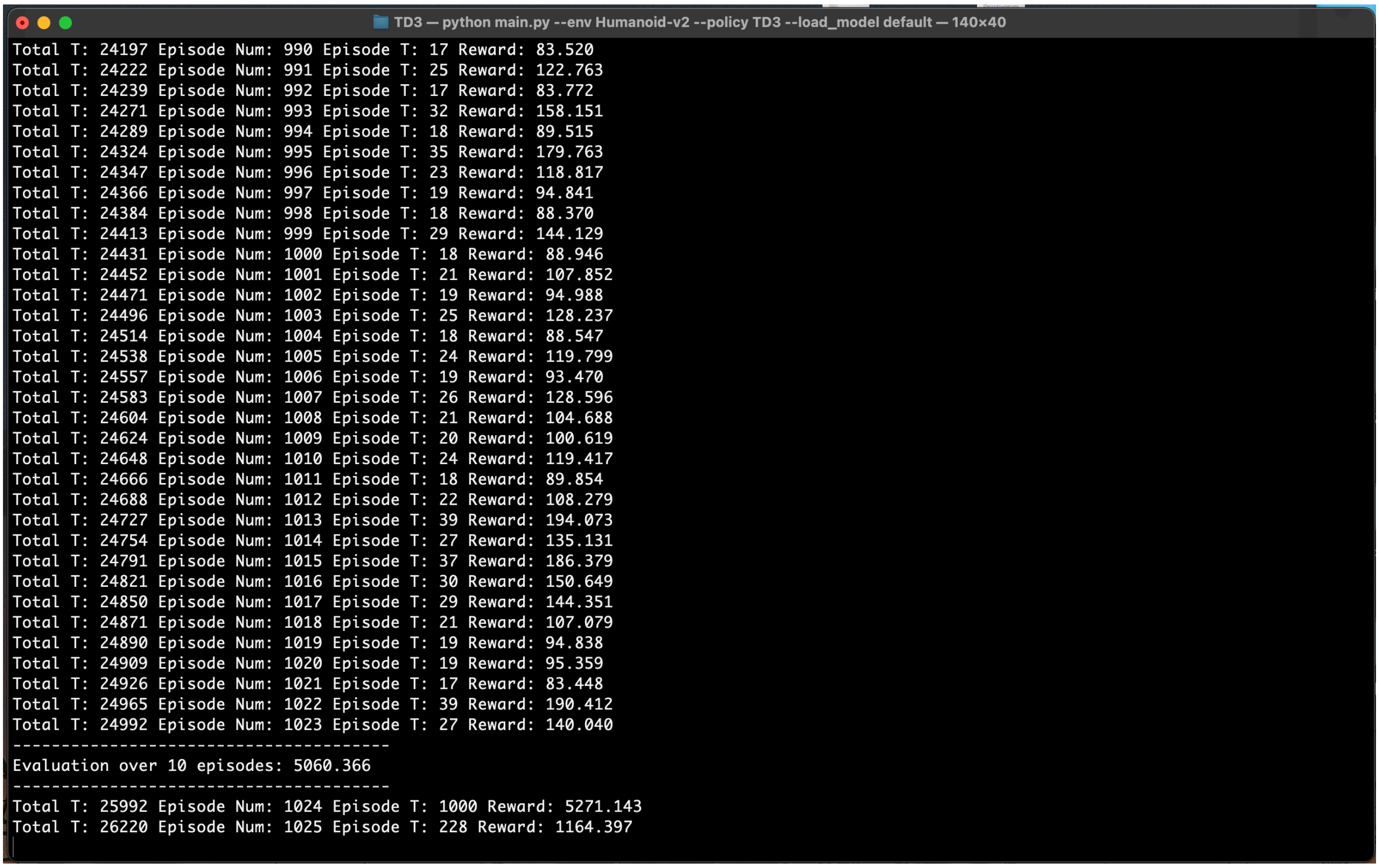
```
python main.py --env Humanoid-v2 --policy TD3 --save_model
```

- 若想要验证已有模型的效果，请添加 `--load_model default` 字段，该模型会自动去默认文件夹下寻找并加载已有模型。具体的命令行参数设置如下所示：

```
parser = argparse.ArgumentParser()
    # Policy name (TD3 or DDPG)
    parser.add_argument("--policy", default="TD3")
    # OpenAI gym environment name
    parser.add_argument("--env", default="HalfCheetah-v2")
    # Sets Gym, PyTorch and Numpy seeds
    parser.add_argument("--seed", default=0, type=int)
    # Time steps initial random policy is used
    parser.add_argument("--start_timesteps", default=2e3, type=int)
    # How often (time steps) we evaluate
    parser.add_argument("--eval_freq", default=5e3, type=int)
    # Max time steps to run environment
    parser.add_argument("--max_timesteps", default=1e6, type=int)
    # Std of Gaussian exploration noise
    parser.add_argument("--expl_noise", default=0.1)
    # Batch size for both actor and critic
    parser.add_argument("--batch_size", default=256, type=int)
```

```python
    # Discount factor
    parser.add_argument("--discount", default=0.99)
    # Target network update rate
    parser.add_argument("--tau", default=0.005)
    # Noise added to target policy during critic update
    parser.add_argument("--policy_noise", default=0.2)
    # Range to clip target policy noise
    parser.add_argument("--noise_clip", default=0.5)
    # Frequency of delayed policy updates
    parser.add_argument("--policy_freq", default=2, type=int)
    # Save model and optimizer parameters
    parser.add_argument("--save_model", action="store_true")
    # Model load file name, "" doesn't load, "default" uses file_name
    parser.add_argument("--load_model", default="")
    args = parser.parse_args()
```

注意到，当我们选择加载已有模型后，整个模型能够在 `learn_start` 之后立刻达到很高的 *reward*，如下图所示：


TD3 — python main.py --env Humanoid-v2 --policy TD3 --load_model default — 140×40

```
Total T: 24197 Episode Num: 990 Episode T: 17 Reward: 83.520
Total T: 24222 Episode Num: 991 Episode T: 25 Reward: 122.763
Total T: 24239 Episode Num: 992 Episode T: 17 Reward: 83.772
Total T: 24271 Episode Num: 993 Episode T: 32 Reward: 158.151
Total T: 24289 Episode Num: 994 Episode T: 18 Reward: 89.515
Total T: 24324 Episode Num: 995 Episode T: 35 Reward: 179.763
Total T: 24347 Episode Num: 996 Episode T: 23 Reward: 118.817
Total T: 24366 Episode Num: 997 Episode T: 19 Reward: 94.841
Total T: 24384 Episode Num: 998 Episode T: 18 Reward: 88.370
Total T: 24413 Episode Num: 999 Episode T: 29 Reward: 144.129
Total T: 24431 Episode Num: 1000 Episode T: 18 Reward: 88.946
Total T: 24452 Episode Num: 1001 Episode T: 21 Reward: 107.852
Total T: 24471 Episode Num: 1002 Episode T: 19 Reward: 94.988
Total T: 24496 Episode Num: 1003 Episode T: 25 Reward: 128.237
Total T: 24514 Episode Num: 1004 Episode T: 18 Reward: 88.547
Total T: 24538 Episode Num: 1005 Episode T: 24 Reward: 119.799
Total T: 24557 Episode Num: 1006 Episode T: 19 Reward: 93.470
Total T: 24583 Episode Num: 1007 Episode T: 26 Reward: 128.596
Total T: 24604 Episode Num: 1008 Episode T: 21 Reward: 104.688
Total T: 24624 Episode Num: 1009 Episode T: 20 Reward: 100.619
Total T: 24648 Episode Num: 1010 Episode T: 24 Reward: 119.417
Total T: 24666 Episode Num: 1011 Episode T: 18 Reward: 89.854
Total T: 24688 Episode Num: 1012 Episode T: 22 Reward: 108.279
Total T: 24727 Episode Num: 1013 Episode T: 39 Reward: 194.073
Total T: 24754 Episode Num: 1014 Episode T: 27 Reward: 135.131
Total T: 24791 Episode Num: 1015 Episode T: 37 Reward: 186.379
Total T: 24821 Episode Num: 1016 Episode T: 30 Reward: 150.649
Total T: 24850 Episode Num: 1017 Episode T: 29 Reward: 144.351
Total T: 24871 Episode Num: 1018 Episode T: 21 Reward: 107.079
Total T: 24890 Episode Num: 1019 Episode T: 19 Reward: 94.838
Total T: 24909 Episode Num: 1020 Episode T: 19 Reward: 95.359
Total T: 24926 Episode Num: 1021 Episode T: 17 Reward: 83.448
Total T: 24965 Episode Num: 1022 Episode T: 39 Reward: 190.412
Total T: 24992 Episode Num: 1023 Episode T: 27 Reward: 140.040
---------------------------------------
Evaluation over 10 episodes: 5060.366
---------------------------------------
Total T: 25992 Episode Num: 1024 Episode T: 1000 Reward: 5271.143
Total T: 26220 Episode Num: 1025 Episode T: 228 Reward: 1164.397
```