README

文件列表

```
args.py
ddpg_v2.py
main.py
rainbow_cv_v1.py
rainbow_v5.py
```

依赖

以下为部分关键依赖，其中`mujoco`必须为`2.3.*`版本，否则会出现与`gymnasium`不匹配的问题，为解决该问题需要先安装`gymnasium[mujoco]`，再安装`mujoco==2.3.0`。

```
pytorch
gymnasium==0.29.1
gymnasium[atari]
gymnasium[accept-rom-license]
gymnasium[mujoco]
mujoco==2.3.0
```

运行

程序主入口为`main.py`，运行时需要指定`--env_name`参数。

例如：

```
python .\main.py --env_name VideoPinball-ramNoFrameskip-v4
python .\main.py --env_name Pong-ramNoFrameskip-v4
python .\main.py --env_name Breakout-ramNoFrameskip-v4
python .\main.py --env_name HalfCheetah-v4
python .\main.py --env_name Ant-v4
python .\main.py --env_name Hopper-v4
python .\main.py --env_name Humanoid-v4
```

默认情况下不会进行渲染，如果需要渲染，需要指定`--render`参数，如：

```
python .\main.py --env_name VideoPinball-ramNoFrameskip-v4 --render
python .\main.py --env_name Humanoid-v4 --render
```

通过`--episode_limit`参数可以指定最大训练轮数，如：

```
python .\main.py --env_name VideoPinball-ramNoFrameskip-v4 --episode_limit 100
```

仅在以下环境进行过测试：

- VideoPinball-ramNoFrameskip-v4
- Pong-ramNoFrameskip-v4
- Breakout-ramNoFrameskip-v4

- BreakoutNoFrameskip-v4
- HalfCheetah-v4
- Ant-v4
- Hopper-v4
- Humanoid-v4

但是，理论上也支持其他环境，如果想运行其他环境，可以通过`--force_run`参数指定运行方法，方法包括：

- value-base：value-based方法，环境state需要为一维向量
- value-base-cv：value-based方法，使用CNN进行图像处理，环境state为图像
- policy-base：policy-based方法，环境state需要为一维向量

以下为运行示例：

```
python .\main.py --env_name VideoPinball-v4 --render --force_run value-base-cv
```

通过禁用改进方法，可以运行原始DQN，如：

```
python .\main.py --env_name VideoPinball-ramNoFrameskip-v4 --disable_dueling --
disable_noisy --disable_double_dqn --disable_priority --multi_step 1
```

还可以通过命令行控制其他参数，部分参数仅对Value-based或Policy-based其中一种生效，完整参数如下：

```
python .\main.py -h
usage: main.py [-h] --env_name ENV_NAME [--replay_buffer_capacity
REPLAY_BUFFER_CAPACITY]
               [--train_batch_size TRAIN_BATCH_SIZE] [--episode_limit
EPISODE_LIMIT] [--render]
               [--gamma GAMMA] [--force_run {value-base,value-base-cv,policy-
base}] [--target_dqn]
               [--disable_noisy] [--disable_double_dqn] [--disable_priority] [--
disable_dueling]
               [--multi_step MULTI_STEP] [--target_update_delay
TARGET_UPDATE_DELAY]
               [--learning_rate LEARNING_RATE] [--test_delay TEST_DELAY]
               [--init_epsilon INIT_EPSILON] [--min_epsilon MIN_EPSILON]
               [--epsilon_decay EPSILON_DECAY] [--tau TAU] [--sigma SIGMA] [--
actor_lr ACTOR_LR]
               [--critic_lr CRITIC_LR]

optional arguments:
  -h, --help            show this help message and exit
  --env_name ENV_NAME
  --replay_buffer_capacity REPLAY_BUFFER_CAPACITY
  --train_batch_size TRAIN_BATCH_SIZE
  --episode_limit EPISODE_LIMIT
                        maximum episode for training
  --render              render environment
  --gamma GAMMA         discount factor
  --force_run {value-base,value-base-cv,policy-base}
```

```
                        force run method, ignore check of environment, choose
from value-base,
                        value-base-cv, policy-base

Value-base:
  For value-based method (Rainbow)

  --target_dqn          use target DQN
  --disable_noisy       disable noisy net
  --disable_double_dqn  disable double DQN
  --disable_priority    disable priority replay buffer
  --disable_dueling     disable dueling DQN
  --multi_step MULTI_STEP
                        multi-step for n-step DQN, set to 1 for vanilla DQN
  --target_update_delay TARGET_UPDATE_DELAY
                        delay for updating target network or exchanging double
network
  --learning_rate LEARNING_RATE
                        learning rate
  --test_delay TEST_DELAY
                        delay for testing
  --init_epsilon INIT_EPSILON
                        initial epsilon value for epsilon-greedy exploration
  --min_epsilon MIN_EPSILON
                        minimum epsilon value for epsilon-greedy exploration
  --epsilon_decay EPSILON_DECAY
                        epsilon decay rate for epsilon-greedy exploration

Policy-base:
  For policy-based method (DDPG)

  --tau TAU             soft update parameter
  --sigma SIGMA         noise parameter
  --actor_lr ACTOR_LR   actor learning rate
  --critic_lr CRITIC_LR
                        critic learning rate
```