**Reinforcement Learning Assignment 3**

Zirong Chen

# **MountainCarContinuous**

In the implementation of DDPG for MountainCarContinuos environment, I chose to use trajectory-wise control to conduct the training process. So, in the all the following figures from DDPG on the MountainCarContinuous environment, the x-axis stands for the number of trajectories collected.

|  |  |  |
| --- | --- | --- |
|  | REINFORCE (Vanilla PG) | DDPG |
| Average Returns  (Red lines are averaged values per 100 units) |  |  |
| To reach  convergence | ~1400 epochs(~40mins) | ~80 trajectories(~12mins) |
| Total time spent | ~6 hours | ~1.5 hours |
| Meet  termination criteria | No | Yes |
| Actor loss |  |  |
| Critic loss |  |  |
| Average length per trajectory | | Chart, histogram  Description automatically generated |

# **LunarLanderContinuous**

In the implementation of DDPG for MountainCarContinuos environment, I chose to use step-wise control to conduct the training process, which means a constraint of total timesteps within each epoch is introduced, as a result, the number of trajectories within each epoch differs from each other according to the number of interactions per trajectory.

|  |  |  |
| --- | --- | --- |
|  | REINFORCE (Vanilla PG) | DDPG |
| Average Returns  (Red lines are averaged values per 100 units) |  |  |
| To reach  convergence | ~260\* epochs(~40mins) | ~210\* epochs(~12mins) |
| Total time spent | ~2 hours | ~2 hours |
| Meet  termination criteria | No | No |
| Actor loss |  |  |
| Critic loss |  |  |
| Average length per trajectory | |  |

\*: Convergence was not guaranteed from the final log file, the average is increasing, but the performance was suffering from high variance.

# **Hyper-parameter settings**

Implementation detailed settings won’t be discussed here.

## MountainCarContinuous

|  |  |  |
| --- | --- | --- |
|  | REINFORCE (Vanilla PG) | DDPG |
| Actor learning rate | 1e-3 | 5e-4 |
| Critic learning rate | 1e-3 | 5e-4 |
| Gamma | 0.99 | 0.99 |
| Hidden size | 64 | 64 |
| Lambda for GAE | 0.97 | 0.97 |
| Max epoch length | 1000 | 1000 |
| Update frequency (actor; critic) | 1; 80 | 1; 100 |
| Polyak averaging | N/A | 0.995 |
| Noise scale | N/A | 0.15 |

## LunarLanderContinuous

|  |  |  |
| --- | --- | --- |
|  | REINFORCE (Vanilla PG) | DDPG |
| Actor learning rate | 1e-3 | 1e-3 |
| Critic learning rate | 1e-3 | 1e-3 |
| Gamma | 0.99 | 0.99 |
| Hidden size | 64 | 64 |
| Lambda for GAE | 0.97 | 0.97 |
| Max epoch length | 1000 | 500 |
| Update frequency (actor; critic) | 1; 80 | 1; 100 |
| Polyak averaging | N/A | 0.995 |
| Noise scale | N/A | 0.1 |

# **Manual book**

Run train:

python train\_[LunarLanderContinuous, MontainCarContinuous]\_chenz51.py -a [DDPG/ddpg, VPG/reinforce]

Run test:

python test\_[LunarLanderContinuous, MontainCarContinuous]\_chenz51.py -a [DDPG/ddpg, VPG/reinforce]

# **Reference**

OpenAI Implantation of RL algorithms @[spiningup](https://spinningup.openai.com/en/latest/index.html)

See code @[github](https://github.com/openai/spinningup)