yl_cartpole_Ex

October 10, 2023

1 Reinforcement Learning for Cartpole Balancing

Name: *Yunlong Pan*

Email:*yunlong.pan@stonybrook.edu*

- Implemented a reinforcement learning agent to control the Cartpole environment using the OpenAI Gym toolkit. Successfully trained the agent to balance the pole on the cart.
- Employed the Proximal Policy Optimization (PPO) algorithm to train the agent and fine-tuned hyperparameters for optimal performance. Utilized deep neural networks as function approximators for policy and value functions.
- Visualized training progress and analyzed learning curves to monitor agent's training process.

[ ]:
```
import gymnasium as gym
import numpy as np
import ray
from ray import tune
from ray.rllib.algorithms.ppo.ppo import PPO
from gym.wrappers import RecordVideo

ray.init()
# tune.run("PPO",
#          config={"env": "CartPole-v1",
#                  "evaluation_interval": 2,    # num of training iter between␣
 ↪evaluations
#                  "evaluation_num_episodes": 20
#                  },
#          local_dir='/Users/ylpan/Desktop/yl_AMS691_RL_Project/yl_gymnasium',
#          checkpoint_freq= 5,
#          )

agent = PPO(config={"env": "CartPole-v1",
                    "evaluation_interval": 2,
                    "evaluation_num_episodes": 20
                    }
            )
```

```
agent.restore("/Users/ylpan/Desktop/yl_AMS691_RL_Project/yl_gymnasium/
 ↪PPO_2023-10-09_13-15-49/PPO_CartPole-v1_7e4a4_00000_0_2023-10-09_13-15-49/
 ↪checkpoint_000180")

env = RecordVideo(gym.make('CartPole-v1',render_mode="rgb_array"),␣
 ↪'yl_cartpole_result')
# obs = env.reset()
# print(obs[0])
# print(agent.compute_single_action(obs[0]))
obs = env.reset()
action = agent.compute_single_action(obs[0])
# for _ in range(30):
#     # print(f"Pole angle at step start: {np.degrees(obs[0])}", end=" ")
#
#     # print(agent.compute_single_action(obs[0]))
#     obs, rewards, done, _c, _d = env.step(action)
#     action = agent.compute_single_action(obs)
#     # print(f"Reward in this step: {rewards}")
#     env.render()

for ep in range(1):
    print(f"Episode number is {ep+1}")
    obs = env.reset()
    for _ in range(100):
        print(f"Pole angle at step start: {np.degrees(obs[0])}", end=" ")
        obs, reward, done, _, _ = env.step(action)
        action = agent.compute_single_action(obs)
        print(f"Pole angle at step end: {np.degrees(obs[0])}", end=" ")
        print(f"Reward in step: {reward}, done: {done}")
        if done:
            break
ray.shutdown()
# env.close()
```

```
2023-10-10 12:19:27,048 WARNING deprecation.py:50 -- DeprecationWarning:
`DirectStepOptimizer` has been deprecated. This will raise an error in the
future!
2023-10-10 12:19:29,213 INFO worker.py:1633 -- Started a local Ray instance.
View the dashboard at 127.0.0.1:8265
2023-10-10 12:19:29,861 WARNING algorithm_config.py:2578 -- Setting
`exploration_config={}` because you set `_enable_rl_module_api=True`. When
RLModule API are enabled, exploration_config can not be set. If you want to
implement custom exploration behaviour, please modify the `forward_exploration`
method of the RLModule at hand. On configs that have a default exploration
config, this must be done with `config.exploration_config={}`.
2023-10-10 12:19:29,861 WARNING deprecation.py:50 -- DeprecationWarning:
`AlgorithmConfig.evaluation(evaluation_num_episodes=..)` has been deprecated.
```

```
Use `AlgorithmConfig.evaluation(evaluation_duration=..,
evaluation_duration_unit='episodes')` instead. This will raise an error in the
future!
/Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-
packages/ray/rllib/algorithms/algorithm.py:484: RayDeprecationWarning: This API
is deprecated and may be removed in future Ray releases. You could suppress this
warning by setting env variable PYTHONWARNINGS="ignore::DeprecationWarning"
`UnifiedLogger` will be removed in Ray 2.7.
  return UnifiedLogger(config, logdir, loggers=None)
/Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-
packages/ray/tune/logger/unified.py:53: RayDeprecationWarning: This API is
deprecated and may be removed in future Ray releases. You could suppress this
warning by setting env variable PYTHONWARNINGS="ignore::DeprecationWarning"
The `JsonLogger interface is deprecated in favor of the
`ray.tune.json.JsonLoggerCallback` interface and will be removed in Ray 2.7.
  self._loggers.append(cls(self.config, self.logdir, self.trial))
/Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-
packages/ray/tune/logger/unified.py:53: RayDeprecationWarning: This API is
deprecated and may be removed in future Ray releases. You could suppress this
warning by setting env variable PYTHONWARNINGS="ignore::DeprecationWarning"
The `CSVLogger interface is deprecated in favor of the
`ray.tune.csv.CSVLoggerCallback` interface and will be removed in Ray 2.7.
  self._loggers.append(cls(self.config, self.logdir, self.trial))
/Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-
packages/ray/tune/logger/unified.py:53: RayDeprecationWarning: This API is
deprecated and may be removed in future Ray releases. You could suppress this
warning by setting env variable PYTHONWARNINGS="ignore::DeprecationWarning"
The `TBXLogger interface is deprecated in favor of the
`ray.tune.tensorboardx.TBXLoggerCallback` interface and will be removed in Ray
2.7.
  self._loggers.append(cls(self.config, self.logdir, self.trial))
(pid=17616) DeprecationWarning: `DirectStepOptimizer` has been
deprecated. This will raise an error in the future!
2023-10-10 12:19:34,287 WARNING algorithm_config.py:2578 -- Setting
`exploration_config={}` because you set `_enable_rl_module_api=True`. When
RLModule API are enabled, exploration_config can not be set. If you want to
implement custom exploration behaviour, please modify the `forward_exploration`
method of the RLModule at hand. On configs that have a default exploration
config, this must be done with `config.exploration_config={}`.
2023-10-10 12:19:34,293 WARNING deprecation.py:50 -- DeprecationWarning:
`ValueNetworkMixin` has been deprecated. This will raise an error in the future!
2023-10-10 12:19:34,293 WARNING deprecation.py:50 -- DeprecationWarning:
`LearningRateSchedule` has been deprecated. This will raise an error in the
future!
2023-10-10 12:19:34,294 WARNING deprecation.py:50 -- DeprecationWarning:
`EntropyCoeffSchedule` has been deprecated. This will raise an error in the
future!
2023-10-10 12:19:34,294 WARNING deprecation.py:50 -- DeprecationWarning:
```

```
`KLCoeffMixin` has been deprecated. This will raise an error in the future!
2023-10-10 12:19:34,308 WARNING algorithm_config.py:2578 -- Setting
`exploration_config={}` because you set `_enable_rl_module_api=True`. When
RLModule API are enabled, exploration_config can not be set. If you want to
implement custom exploration behaviour, please modify the `forward_exploration`
method of the RLModule at hand. On configs that have a default exploration
config, this must be done with `config.exploration_config={}`.
(RolloutWorker pid=17616) 2023-10-10 12:19:34,270  WARNING
algorithm_config.py:2578 -- Setting `exploration_config={}` because you set
`_enable_rl_module_api=True`. When RLModule API are enabled, exploration_config
can not be set. If you want to implement custom exploration behaviour, please
modify the `forward_exploration` method of the RLModule at hand. On configs that
have a default exploration config, this must be done with
`config.exploration_config={}`.
(RolloutWorker pid=17616) 2023-10-10 12:19:34,275  WARNING
deprecation.py:50 -- DeprecationWarning: `ValueNetworkMixin` has been
deprecated. This will raise an error in the future!
(RolloutWorker pid=17616) 2023-10-10 12:19:34,275  WARNING
deprecation.py:50 -- DeprecationWarning: `LearningRateSchedule` has been
deprecated. This will raise an error in the future!
(RolloutWorker pid=17616) 2023-10-10 12:19:34,275  WARNING
deprecation.py:50 -- DeprecationWarning: `EntropyCoeffSchedule` has been
deprecated. This will raise an error in the future!
(RolloutWorker pid=17616) 2023-10-10 12:19:34,275  WARNING
deprecation.py:50 -- DeprecationWarning: `KLCoeffMixin` has been deprecated.
This will raise an error in the future!
2023-10-10 12:19:34,321 WARNING util.py:68 -- Install gputil for GPU system
monitoring.
2023-10-10 12:19:34,325 WARNING algorithm_config.py:2578 -- Setting
`exploration_config={}` because you set `_enable_rl_module_api=True`. When
RLModule API are enabled, exploration_config can not be set. If you want to
implement custom exploration behaviour, please modify the `forward_exploration`
method of the RLModule at hand. On configs that have a default exploration
config, this must be done with `config.exploration_config={}`.
2023-10-10 12:19:34,326 WARNING algorithm_config.py:672 -- Cannot create
PPOConfig from given `config_dict`! Property __stdout_file__ not supported.
2023-10-10 12:19:34,357 INFO trainable.py:984 -- Restored on 127.0.0.1 from
checkpoint: Checkpoint(filesystem=local, path=/Users/ylpan/Desktop/yl_AMS691_RL_
Project/yl_gymnasium/PPO_2023-10-09_13-15-49/PPO_CartPole-v1_7e4a4_00000_0_2023-
10-09_13-15-49/checkpoint_000180)
/Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-
packages/gym/wrappers/record_video.py:75: UserWarning: WARN: Overwriting
existing videos at /Users/ylpan/Desktop/yl_AMS691_RL_Project/yl_cartpole_result
folder (try specifying a different `video_folder` for the `RecordVideo` wrapper
if this is not desired)
  logger.warn(
```

```
Episode number is 1
Pole angle at step start: [-2.3076787  -1.1653538   1.6054516   0.46877775] Pole
angle at step end: -2.3309860229492188 Reward in step: 1.0, done: False
Pole angle at step start: -2.3309860229492188 Pole angle at step end:
-2.1311728954315186 Reward in step: 1.0, done: False
Pole angle at step start: -2.1311728954315186 Pole angle at step end:
-2.155400276184082 Reward in step: 1.0, done: False
Pole angle at step start: -2.155400276184082 Pole angle at step end:
-1.9564155340194702 Reward in step: 1.0, done: False
Pole angle at step start: -1.9564155340194702 Pole angle at step end:
-1.981394648551941 Reward in step: 1.0, done: False
Pole angle at step start: -1.981394648551941 Pole angle at step end:
-1.7830814123153687 Reward in step: 1.0, done: False
Pole angle at step start: -1.7830814123153687 Pole angle at step end:
-1.808663249015808 Reward in step: 1.0, done: False
Pole angle at step start: -1.808663249015808 Pole angle at step end:
-1.610883116722107 Reward in step: 1.0, done: False
Pole angle at step start: -1.610883116722107 Pole angle at step end:
-1.6369366645812988 Reward in step: 1.0, done: False
Pole angle at step start: -1.6369366645812988 Pole angle at step end:
-1.4395661354064941 Reward in step: 1.0, done: False
Pole angle at step start: -1.4395661354064941 Pole angle at step end:
-1.4659736156463623 Reward in step: 1.0, done: False
Pole angle at step start: -1.4659736156463623 Pole angle at step end:
-1.2689008712768555 Reward in step: 1.0, done: False
Pole angle at step start: -1.2689008712768555 Pole angle at step end:
-1.2955546379089355 Reward in step: 1.0, done: False
Pole angle at step start: -1.2955546379089355 Pole angle at step end:
-1.0986770391464233 Reward in step: 1.0, done: False
Pole angle at step start: -1.0986770391464233 Pole angle at step end:
-1.1254773139953613 Reward in step: 1.0, done: False
Pole angle at step start: -1.1254773139953613 Pole angle at step end:
-0.9286980032920837 Reward in step: 1.0, done: False
Pole angle at step start: -0.9286980032920837 Pole angle at step end:
-0.9555498957633972 Reward in step: 1.0, done: False
Pole angle at step start: -0.9555498957633972 Pole angle at step end:
-0.7587756514549255 Reward in step: 1.0, done: False
Pole angle at step start: -0.7587756514549255 Pole angle at step end:
-0.7855867743492126 Reward in step: 1.0, done: False
Pole angle at step start: -0.7855867743492126 Pole angle at step end:
-0.5887258052825928 Reward in step: 1.0, done: False
Pole angle at step start: -0.5887258052825928 Pole angle at step end:
-0.6154037714004517 Reward in step: 1.0, done: False
Pole angle at step start: -0.6154037714004517 Pole angle at step end:
-0.4183633327484131 Reward in step: 1.0, done: False
Pole angle at step start: -0.4183633327484131 Pole angle at step end:
-0.44481372833251953 Reward in step: 1.0, done: False
Pole angle at step start: -0.44481372833251953 Pole angle at step end:
```

```
-0.24749763309955597 Reward in step: 1.0, done: False
Pole angle at step start: -0.24749763309955597 Pole angle at step end:
-0.2736213803291321 Reward in step: 1.0, done: False
Pole angle at step start: -0.2736213803291321 Pole angle at step end:
-0.07592782378196716 Reward in step: 1.0, done: False
Pole angle at step start: -0.07592782378196716 Pole angle at step end:
-0.10161897540092468 Reward in step: 1.0, done: False
Pole angle at step start: -0.10161897540092468 Pole angle at step end:
0.09656202048063278 Reward in step: 1.0, done: False
Pole angle at step start: 0.09656202048063278 Pole angle at step end:
0.07141899317502975 Reward in step: 1.0, done: False
Pole angle at step start: 0.07141899317502975 Pole angle at step end:
0.27020812034606934 Reward in step: 1.0, done: False
Pole angle at step start: 0.27020812034606934 Pole angle at step end:
0.245741069316864 Reward in step: 1.0, done: False
Pole angle at step start: 0.245741069316864 Pole angle at step end:
-0.0018929778598248959 Reward in step: 1.0, done: False
Pole angle at step start: -0.0018929778598248959 Pole angle at step end:
-0.02554064244031906 Reward in step: 1.0, done: False
Pole angle at step start: -0.02554064244031906 Pole angle at step end:
-0.27247369289398193 Reward in step: 1.0, done: False
Pole angle at step start: -0.27247369289398193 Pole angle at step end:
-0.29552313685417175 Reward in step: 1.0, done: False
Pole angle at step start: -0.29552313685417175 Pole angle at step end:
-0.0947832390666008 Reward in step: 1.0, done: False
Pole angle at step start: -0.0947832390666008 Pole angle at step end:
-0.11743801087141037 Reward in step: 1.0, done: False
Pole angle at step start: -0.11743801087141037 Pole angle at step end:
-0.3633871078491211 Reward in step: 1.0, done: False
Pole angle at step start: -0.3633871078491211 Pole angle at step end:
-0.38545462489128113 Reward in step: 1.0, done: False
Pole angle at step start: -0.38545462489128113 Pole angle at step end:
-0.6309106349945068 Reward in step: 1.0, done: False
Pole angle at step start: -0.6309106349945068 Pole angle at step end:
-0.6525694727897644 Reward in step: 1.0, done: False
Pole angle at step start: -0.6525694727897644 Pole angle at step end:
-0.8977030515670776 Reward in step: 1.0, done: False
Pole angle at step start: -0.8977030515670776 Pole angle at step end:
-0.9191194176673889 Reward in step: 1.0, done: False
Pole angle at step start: -0.9191194176673889 Pole angle at step end:
-1.1640912294387817 Reward in step: 1.0, done: False
Pole angle at step start: -1.1640912294387817 Pole angle at step end:
-1.1854230165481567 Reward in step: 1.0, done: False
Pole angle at step start: -1.1854230165481567 Pole angle at step end:
-0.9832013249397278 Reward in step: 1.0, done: False
Pole angle at step start: -0.9832013249397278 Pole angle at step end:
-1.0046024322509766 Reward in step: 1.0, done: False
Pole angle at step start: -1.0046024322509766 Pole angle at step end:
```

```
-1.2495205402374268 Reward in step: 1.0, done: False
Pole angle at step start: -1.2495205402374268 Pole angle at step end:
-1.2707608938217163 Reward in step: 1.0, done: False
Pole angle at step start: -1.2707608938217163 Pole angle at step end:
-1.0684089660644531 Reward in step: 1.0, done: False
Pole angle at step start: -1.0684089660644531 Pole angle at step end:
-1.0896410942077637 Reward in step: 1.0, done: False
Pole angle at step start: -1.0896410942077637 Pole angle at step end:
-1.3343497514724731 Reward in step: 1.0, done: False
Pole angle at step start: -1.3343497514724731 Pole angle at step end:
-1.3553402423858643 Reward in step: 1.0, done: False
Pole angle at step start: -1.3553402423858643 Pole angle at step end:
-1.5998826026916504 Reward in step: 1.0, done: False
Pole angle at step start: -1.5998826026916504 Pole angle at step end:
-1.6207787990570068 Reward in step: 1.0, done: False
Pole angle at step start: -1.6207787990570068 Pole angle at step end:
-1.4181123971939087 Reward in step: 1.0, done: False
Pole angle at step start: -1.4181123971939087 Pole angle at step end:
-1.4390572309494019 Reward in step: 1.0, done: False
Pole angle at step start: -1.4390572309494019 Pole angle at step end:
-1.683504343032837 Reward in step: 1.0, done: False
Pole angle at step start: -1.683504343032837 Pole angle at step end:
-1.704256534576416 Reward in step: 1.0, done: False
Pole angle at step start: -1.704256534576416 Pole angle at step end:
-1.948582649230957 Reward in step: 1.0, done: False
Pole angle at step start: -1.948582649230957 Pole angle at step end:
-1.969282865524292 Reward in step: 1.0, done: False
Pole angle at step start: -1.969282865524292 Pole angle at step end:
-1.7664395570755005 Reward in step: 1.0, done: False
Pole angle at step start: -1.7664395570755005 Pole angle at step end:
-1.7872252464294434 Reward in step: 1.0, done: False
Pole angle at step start: -1.7872252464294434 Pole angle at step end:
-2.0315299034118652 Reward in step: 1.0, done: False
Pole angle at step start: -2.0315299034118652 Pole angle at step end:
-2.0521554946899414 Reward in step: 1.0, done: False
Pole angle at step start: -2.0521554946899414 Pole angle at step end:
-2.2963695526123047 Reward in step: 1.0, done: False
Pole angle at step start: -2.2963695526123047 Pole angle at step end:
-2.31697154045105 Reward in step: 1.0, done: False
Pole angle at step start: -2.31697154045105 Pole angle at step end:
-2.561229705810547 Reward in step: 1.0, done: False
Pole angle at step start: -2.561229705810547 Pole angle at step end:
-2.5819430351257324 Reward in step: 1.0, done: False
Pole angle at step start: -2.5819430351257324 Pole angle at step end:
-2.379195213317871 Reward in step: 1.0, done: False
Pole angle at step start: -2.379195213317871 Pole angle at step end:
-2.400155544281006 Reward in step: 1.0, done: False
Pole angle at step start: -2.400155544281006 Pole angle at step end:
```

```
-2.19753098487854 Reward in step: 1.0, done: False
Pole angle at step start: -2.19753098487854 Pole angle at step end:
-2.2184953689575195 Reward in step: 1.0, done: False
Pole angle at step start: -2.2184953689575195 Pole angle at step end:
-2.462939977645874 Reward in step: 1.0, done: False
Pole angle at step start: -2.462939977645874 Pole angle at step end:
-2.4836678504943848 Reward in step: 1.0, done: False
Pole angle at step start: -2.4836678504943848 Pole angle at step end:
-2.727947950363159 Reward in step: 1.0, done: False
Pole angle at step start: -2.727947950363159 Pole angle at step end:
-2.74858021736145 Reward in step: 1.0, done: False
Pole angle at step start: -2.74858021736145 Pole angle at step end:
-2.992833137512207 Reward in step: 1.0, done: False
Pole angle at step start: -2.992833137512207 Pole angle at step end:
-3.0135059356689453 Reward in step: 1.0, done: False
Pole angle at step start: -3.0135059356689453 Pole angle at step end:
-2.810681104660034 Reward in step: 1.0, done: False
Pole angle at step start: -2.810681104660034 Pole angle at step end:
-2.8315300941467285 Reward in step: 1.0, done: False
Pole angle at step start: -2.8315300941467285 Pole angle at step end:
-3.075943946838379 Reward in step: 1.0, done: False
Pole angle at step start: -3.075943946838379 Pole angle at step end:
-3.096723794937134 Reward in step: 1.0, done: False
Pole angle at step start: -3.096723794937134 Pole angle at step end:
-3.3411388397216797 Reward in step: 1.0, done: False
Pole angle at step start: -3.3411388397216797 Pole angle at step end:
-3.3619890213012695 Reward in step: 1.0, done: False
Pole angle at step start: -3.3619890213012695 Pole angle at step end:
-3.1593589782714844 Reward in step: 1.0, done: False
Pole angle at step start: -3.1593589782714844 Pole angle at step end:
-3.180419683456421 Reward in step: 1.0, done: False
Pole angle at step start: -3.180419683456421 Pole angle at step end:
-3.4250643253326416 Reward in step: 1.0, done: False
Pole angle at step start: -3.4250643253326416 Pole angle at step end:
-3.4460952281951904 Reward in step: 1.0, done: False
Pole angle at step start: -3.4460952281951904 Pole angle at step end:
-3.2435965538024902 Reward in step: 1.0, done: False
Pole angle at step start: -3.2435965538024902 Pole angle at step end:
-3.264742851257324 Reward in step: 1.0, done: False
Pole angle at step start: -3.264742851257324 Pole angle at step end:
-3.5094268321990967 Reward in step: 1.0, done: False
Pole angle at step start: -3.5094268321990967 Pole angle at step end:
-3.530451536178589 Reward in step: 1.0, done: False
Pole angle at step start: -3.530451536178589 Pole angle at step end:
-3.77508807182312 Reward in step: 1.0, done: False
Pole angle at step start: -3.77508807182312 Pole angle at step end:
-3.796137809753418 Reward in step: 1.0, done: False
Pole angle at step start: -3.796137809753418 Pole angle at step end:
```

```
-3.59368634223938 Reward in step: 1.0, done: False
Pole angle at step start: -3.59368634223938 Pole angle at step end:
-3.6149067878723145 Reward in step: 1.0, done: False
Pole angle at step start: -3.6149067878723145 Pole angle at step end:
-3.8596928119659424 Reward in step: 1.0, done: False
Pole angle at step start: -3.8596928119659424 Pole angle at step end:
-3.880847454071045 Reward in step: 1.0, done: False
Pole angle at step start: -3.880847454071045 Pole angle at step end:
-3.678457021713257 Reward in step: 1.0, done: False

(pid=17617) DeprecationWarning: `DirectStepOptimizer` has been
deprecated. This will raise an error in the future!
(RolloutWorker pid=17617) 2023-10-10 12:19:34,270  WARNING
algorithm_config.py:2578 -- Setting `exploration_config={}` because you set
`_enable_rl_module_api=True`. When RLModule API are enabled, exploration_config
can not be set. If you want to implement custom exploration behaviour, please
modify the `forward_exploration` method of the RLModule at hand. On configs that
have a default exploration config, this must be done with
`config.exploration_config={}`.
```

[ ]: