

一、agent 从一个没有任何物品的随机起始位置开始，并负责获得钻石。只能通过浏览 Minecraft 的复杂项目层次结构来完成此任务。

AI 想要挖到钻石，并不简单：历经八个步骤，每一步都要自行探索：

第一步，收集木材。

第二步，用收来的木料造一只木镐。

第三步，拿着木镐去挖石矿，然后造一只石镐。

第四步，用新的石镐挖铁矿。

第五步，打一个炉子。

第六步，把铁熔了造个铁镐。有了铁镐，才挖得动钻石。

第七步，找钻石。并不容易，AI 要慢慢摸索，才知道钻石常常出没的地方。

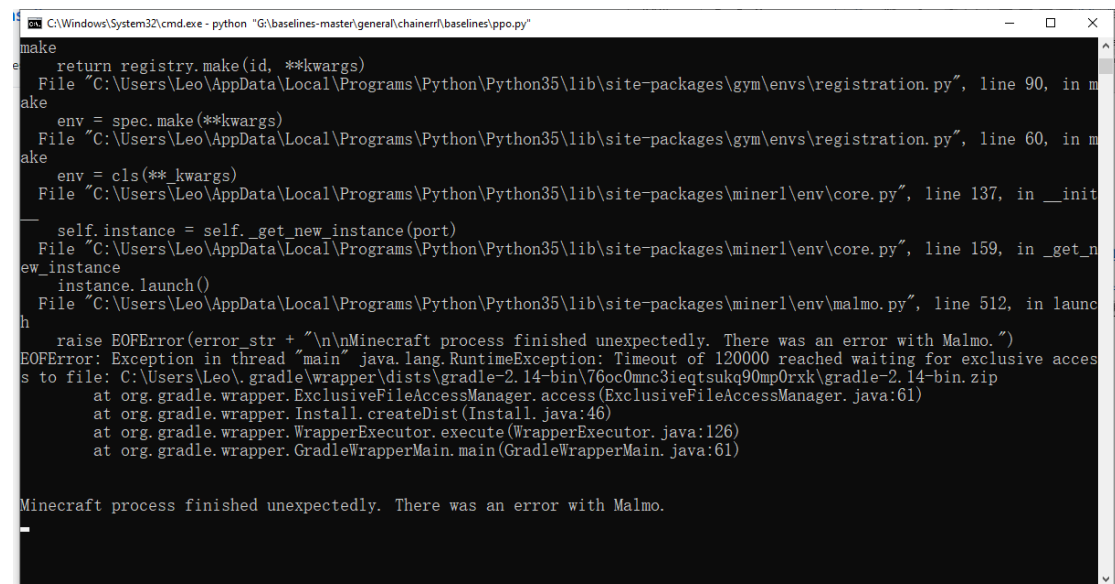
第八步，挖钻石。任务完结。

数据集 MineRL-v0。有 6,000 万帧数据，全部来自人类玩家。数据集可以大大减少 agent 的训练时间，要是用实时的数据要达到挖钻石，那得到猴年马月了。。即使有这些数据，训练时间也还是太长了一些。

## 二、调试和运行过程

### 1. 本机 windows 调试过程：

在本集中安装 [https://github.com/minerllabs/competition\\_submission\\_template](https://github.com/minerllabs/competition_submission_template) 中的环境，运行 <https://github.com/minerllabs/baselines> 中的 ppo.py 但是 windows 中一直卡在这里，看群里讨论的也有同学遇到这种问题，是 pip3 install --upgrade minerl 解决，但是，没用。。也不知道怎么搞。



```
make
return registry.make(id, **kwargs)
File "C:\Users\Leo\AppData\Local\Programs\Python\Python35\lib\site-packages\gym\envs\registration.py", line 90, in make
env = spec.make(**kwargs)
File "C:\Users\Leo\AppData\Local\Programs\Python\Python35\lib\site-packages\gym\envs\registration.py", line 60, in make
env = cls(**kwargs)
File "C:\Users\Leo\AppData\Local\Programs\Python\Python35\lib\site-packages\minerl\env\core.py", line 137, in __init__
self.instance = self._get_new_instance(port)
File "C:\Users\Leo\AppData\Local\Programs\Python\Python35\lib\site-packages\minerl\env\core.py", line 159, in _get_new_instance
instance.launch()
File "C:\Users\Leo\AppData\Local\Programs\Python\Python35\lib\site-packages\minerl\env\malmo.py", line 512, in launch
raise EOFError(error_str + "\n\nMinecraft process finished unexpectedly. There was an error with Malmo.")
EOFError: Exception in thread "main" java.lang.RuntimeException: Timeout of 120000 reached waiting for exclusive access
to file: C:\Users\Leo\gradle\wrapper\dists\gradle-2.14-bin\76oc0mmc3ieqtsukq90mp0rxk\gradle-2.14-bin.zip
at org.gradle.wrapper.ExclusiveFileAccessManager.access(ExclusiveFileAccessManager.java:61)
at org.gradle.wrapper.Install.createDist(Install.java:46)
at org.gradle.wrapper.WrapperExecutor.execute(WrapperExecutor.java:126)
at org.gradle.wrapper.GradleWrapperMain.main(GradleWrapperMain.java:61)

Minecraft process finished unexpectedly. There was an error with Malmo.
```

### 2. 云服务器中的调试过程：

安装助教给出的 docker 中配置好的环境。运行 <https://github.com/minerllabs/baselines> 中的代码： ppo.py

一开始在服务器中也提示 malmo 有错误，怎么调试也不好使，然后看群里说可以了就重新进了一下，发现确实可以了。。谜一样的环境。。

```

INFO - 2020-08-15 23:57:35,757 - [chainerrl.experiments.train_agent train_agent 59] outdir:results/20200815T234547.4730
64 step:8000 episode:0 R:0.0
INFO - 2020-08-15 23:57:35,759 - [chainerrl.experiments.train_agent train_agent 60] statistics:[('average_value', 0.023
661900237202646), ('average_entropy', 2.3937542769908906), ('average_value_loss', 9.557441258948529e-05), ('average_policy_
loss', 0.004398292129917536), ('n_updates', 672), ('explained_variance', -6.356840657404153)]
INFO - 2020-08-16 00:06:46,828 - [chainerrl.experiments.train_agent train_agent 59] outdir:results/20200815T234547.4730
64 step:16000 episode:1 R:0.0
INFO - 2020-08-16 00:06:46,829 - [chainerrl.experiments.train_agent train_agent 60] statistics:[('average_value', 0.002
4314895380375674), ('average_entropy', 2.396819194793701), ('average_value_loss', 1.517193625204527e-05), ('average_policy_
loss', 0.000557665911182994), ('n_updates', 1440), ('explained_variance', -4.98282435798739)]
INFO - 2020-08-16 00:16:25,823 - [chainerrl.experiments.train_agent train_agent 59] outdir:results/20200815T234547.4730
64 step:24000 episode:2 R:0.0
INFO - 2020-08-16 00:16:25,824 - [chainerrl.experiments.train_agent train_agent 60] statistics:[('average_value', 0.000
5441535166319227), ('average_entropy', 2.3972799215316773), ('average_value_loss', 7.472278771274432e-06), ('average_policy_
loss', -7.257209785166196e-05), ('n_updates', 2208), ('explained_variance', -10.622575452524483)]
INFO - 2020-08-16 00:26:03,121 - [chainerrl.experiments.train_agent train_agent 59] outdir:results/20200815T234547.4730
64 step:32000 episode:3 R:0.0
INFO - 2020-08-16 00:26:03,122 - [chainerrl.experiments.train_agent train_agent 60] statistics:[('average_value', 0.000
48400674165895906), ('average_entropy', 2.397660793066025), ('average_value_loss', 3.2068969039755757e-06), ('average_polic
y_loss', 0.00018989447780768388), ('n_updates', 2976), ('explained_variance', -5.654123781750957)]
INFO - 2020-08-16 00:35:00,141 - [chainerrl.experiments.train_agent train_agent 59] outdir:results/20200815T234547.4730
64 step:40000 episode:4 R:0.0
INFO - 2020-08-16 00:35:00,143 - [chainerrl.experiments.train_agent train_agent 60] statistics:[('average_value', -2.57
25508050527424e-05), ('average_entropy', 2.3976837611198425), ('average_value_loss', 1.109857161338823e-06), ('average_poli
cy_loss', 4.7872613504296165e-06), ('n_updates', 3744), ('explained_variance', -6.820766833185276)]
INFO - 2020-08-16 00:43:50,981 - [chainerrl.experiments.train_agent train_agent 59] outdir:results/20200815T234547.4730
64 step:48000 episode:5 R:0.0
INFO - 2020-08-16 00:43:50,982 - [chainerrl.experiments.train_agent train_agent 60] statistics:[('average_value', -0.00
0527678130150889), ('average_entropy', 2.397787939310074), ('average_value_loss', 7.471946399562057e-07), ('average_policy_
loss', -3.433221387240337e-05), ('n_updates', 4416), ('explained_variance', -3.353828690226095)]
INFO - 2020-08-16 00:53:07,414 - [chainerrl.experiments.train_agent train_agent 59] outdir:results/20200815T234547.4730
64 step:56000 episode:6 R:0.0
INFO - 2020-08-16 00:53:07,421 - [chainerrl.experiments.train_agent train_agent 60] statistics:[('average_value', -0.00
031374464833788807), ('average_entropy', 2.3977822515964506), ('average_value_loss', 6.056256687259065e-07), ('average_polic
y_loss', -0.00018587903728985113), ('n_updates', 5184), ('explained_variance', -2.6004463727712746)]
INFO - 2020-08-16 01:02:20,957 - [chainerrl.experiments.train_agent train_agent 59] outdir:results/20200815T234547.4730
64 step:64000 episode:7 R:0.0
INFO - 2020-08-16 01:02:20,964 - [chainerrl.experiments.train_agent train_agent 60] statistics:[('average_value', 0.000
7219635946239578), ('average_entropy', 2.3978333320617677), ('average_value_loss', 1.2894934442897465e-06), ('average_polic
y_loss', 0.00015668119040128657), ('n_updates', 5952), ('explained_variance', -4.611201002846354)]
INFO - 2020-08-16 01:11:35,726 - [chainerrl.experiments.train_agent train_agent 59] outdir:results/20200815T234547.4730
64 step:72000 episode:8 R:0.0
INFO - 2020-08-16 01:11:35,728 - [chainerrl.experiments.train_agent train_agent 60] statistics:[('average_value', -0.00
0950971498546096), ('average_entropy', 2.397846014261246), ('average_value_loss', 7.664896443415614e-07), ('average_policy_
loss', -0.00022628282677032984), ('n_updates', 6720), ('explained_variance', -12.02136550758831)]
INFO - 2020-08-16 01:21:07,597 - [chainerrl.experiments.train_agent train_agent 59] outdir:results/20200815T234547.4730
64 step:80000 episode:9 R:0.0

```

运行的 dddqn 提示 dqn\_family 有错误，看到助教在群里发的 fix 之后的新代码，试验一下，也还是有错误，调试不动。。：

```

11.2,jsonschema==3.2.0,jupyter==1.0.0,jupyter-client==6.1.6,jupyter-console==6.1.0,jupyter-core==4.6.3,jupyter-http-over-ws
==0.0.8,Keras-Preprocessing==1.1.2,keyring==10.6.0,keyrings.alt==3.0,kiwisolver==1.2.0,language-selector==0.1,lxml==4.5.2,m
acaronbakery==1.1.3,Markdown==3.2.2,MarkupSafe==1.1.1,matplotlib==3.0.3,minerl==0.3.6,mistune==0.8.4,nbconvert==5.6.1,nbfo
rmat==4.4.0,notebook==6.0.3,numpy==1.18.5,nvidia-ml-py3==7.352.0,oauthlib==3.1.0,opencv-python==4.3.0.36,opt-einsum==3.3.0,
packaging==20.4,pandocfilters==1.4.2,parso==0.7.1,pexpect==4.8.0,pickleshare==0.7.5,Pillow==7.2.0,pip==20.1.1,prometheus-cl
ient==0.8.0,prompt-toolkit==3.0.5,protobuf==3.12.2,psutil==5.7.2,ptyprocess==0.6.0,pyasn1==0.4.8,pyasn1-modules==0.2.8,pyca
iro==1.16.2,pycrypto==2.6.1,pycups==1.9.73,pydot==1.4.1,pyglet==1.5.0,Pygments==2.6.1,pygobject==3.26.1,pymacaroons==0.13.0
,PyNaCl==1.1.2,pyarsing==2.4.7,pyRFC3339==1.0,Pyro4==4.80,pyrsistent==0.16.0,python-apt==1.6.5+ubuntu0.3.python-dateutil==
2.8.1,python-debian==0.1.32,pytz==2018.3,pyxdg==0.25,PyYAML==3.12,pyzmq==19.0.1,qtconsole==4.7.5,QtPy==1.9.0,requests==2.24
.0,requests-oauthlib==1.3.0,requests-unixsocket==0.1.5,rsa==4.6,scipy==1.4.1,SecretStorage==2.3.1,Send2Trash==1.5.0,serpent
==1.30.2,setuputils==49.2.0,six==1.15.0,tensorboard==2.3.0,tensorboard-plugin-wit==1.7.0,tensorflow-estimator==2.3.0,tensor
flow-gpu==2.3.0,termcolor==1.1.0,terminado==0.8.3,testpath==0.4.4,torch==1.6.0+cu101,torchvision==0.7.0+cu101,tornado==6.0.
4,tqdm==4.48.2,traitlets==4.3.3,typing==3.7.4.3,typing-extensions==3.7.4.2,ubuntu-drivers-common==0.0.0,urllib3==1.25.10,wc
width==0.5.1,webencodings==0.5.1,Werkzeug==1.0.1,wheel==0.30.0,widgetsnbextension==3.5.1,wrapt==1.12.1,xkit==0.0.0,zip==3.
1.0
INFO - 2020-08-16 02:14:07,898 - [_main__ main 117] The first 'gym.make(MinerL*)' may take several minutes. Be patient!
INFO - 2020-08-16 02:14:07,962 - [minerl.env.malmo.instance.ff833e _launch_minecraft 671] Starting Minecraft process: [
'/tmp/tmp7r2e6u2x/Minecraft/launchClient.sh', '-port', '9184', '-env', '-runDir', '/tmp/tmp7r2e6u2x/Minecraft/run']
INFO - 2020-08-16 02:14:07,973 - [minerl.env.malmo.instance.ff833e _launch_process_watcher 694] Starting process_watcher
for process 1198 @ localhost:9184
INFO - 2020-08-16 02:15:47,807 - [minerl.env.malmo.instance.ff833e launch 533] Minecraft process ready
INFO - 2020-08-16 02:15:47,829 - [_main__ wrap_env 213] Detected 'gym.wrappers.TimeLimit'! Unwrap it and re-wrap our own
time limit.
INFO - 2020-08-16 02:15:47,830 - [minerl.env.malmo log_to_file 548] Logging output of Minecraft to results/MinerLTreech
op-v0/dddqn/20200816T021407.711957/logs/mc_184.log
INFO - 2020-08-16 02:15:47,834 - [env_wrappers _init_ 481] always pressing keys: ['attack']
INFO - 2020-08-16 02:15:47,834 - [env_wrappers _init_ 487] reversed pressing keys: ['forward']
INFO - 2020-08-16 02:15:47,835 - [env_wrappers _init_ 492] always ignored keys: ['back', 'left', 'right', 'sneak', 's
print']
INFO - 2020-08-16 02:15:47,837 - [env_wrappers _init_ 545] Dict(attack:Discrete(2), back:Discrete(2), camera:Box(low=-
180.0, high=180.0, shape=(2,)), forward:Discrete(2), jump:Discrete(2), left:Discrete(2), right:Discrete(2), sneak:Discrete
(2), sprint:Discrete(2)) is converted to Discrete(5).
INFO - 2020-08-16 02:15:47,837 - [_main__ wrap_env 213] Detected 'gym.wrappers.TimeLimit'! Unwrap it and re-wrap our own
time limit.
INFO - 2020-08-16 02:15:47,840 - [env_wrappers _init_ 481] always pressing keys: ['attack']
INFO - 2020-08-16 02:15:47,841 - [env_wrappers _init_ 487] reversed pressing keys: ['forward']
INFO - 2020-08-16 02:15:47,842 - [env_wrappers _init_ 492] always ignored keys: ['back', 'left', 'right', 'sneak', 's
print']
INFO - 2020-08-16 02:15:47,844 - [env_wrappers _init_ 545] Dict(attack:Discrete(2), back:Discrete(2), camera:Box(low=-
180.0, high=180.0, shape=(2,)), forward:Discrete(2), jump:Discrete(2), left:Discrete(2), right:Discrete(2), sneak:Discrete
(2), sprint:Discrete(2)) is converted to Discrete(5).
INFO - 2020-08-16 02:16:28,437 - [chainerrl.experiments.train_agent save_agent 266] Saved the agent to results/MinerLTree
echop-v0/dddqn/20200816T021407.711957/0 except
ERROR - 2020-08-16 02:16:28,440 - [_main__ main 112] execution failed.
Traceback (most recent call last):
  File "dqn_family.py", line 110, in main
    main(args)
  File "dqn_family.py", line 193, in main
    outdir=args.outdir, eval_env=eval_env, save_best_so_far_agent=True,
  File "/usr/local/lib/python3.6/dist-packages/chainerrl/experiments/train_agent.py", line 160, in train_agent_with_evaluat

```

### 三、小结

就跑出来了 ppo.py 就简单说一下 ppo 吧：PPO 算法本质上是一个 AC 算法，有 Actor 和 Critic 神经网络，其中，Critic 网络的更新方式和 AC 算法差不多，Actor 网络我感觉和 Q-Learning 一样有新旧神经网络，并周期性的更新旧神经网络。



在给出的 readme 中，说到 MineRLTreechop-v0 任务训练阶段算法的性能。每种算法均经过 3 次独立训练（试验），阴影区域代表三个试验的得分之间的标准偏差（而非标准误差）。通过平均观察 30 次以上的可见性来平滑曲线。

Rainbow 和 PPO 的表现优于 DDDQN。

可见 ppo 虽然是 2017 年发布的一个算法，但其稳定性和适用范围还是值得拥有的。。