

在 gym 中可以获得讯息

- 一辆车在一个二维的轨道上，位于两座山之间。
- 目标是冲上右边的山头。
- 动作集：一维的离散空间：motor=(left,neutral,right)
- 状态变量：二维的连续状态空间：
- Velocity=(-0.07,0.07) , Position=(-1.2,0.6)
- 奖励：每过去一个时间，reward=-1
- 状态更新：

Velocity=Velocity+(Action)*0.001+cos(3*Position)*(-0.0025)Velocity=Velocity+(Action)*0.001+cos(3*Position)*(-0.0025),

Position=Position+VelocityPosition=Position+Velocity

- 初始状态：Position=-0.5 Position=-0.5, Velocity=0.0 Velocity=0.0
 - 终止状态：Position≥0.6
1. 为什么选择 DQN,问题本身是 MDP 的知道状态和连续动作，可以用物理学推导出下一个状态和连续动作。但是问题是，状态是连续的不能用单一的一个大 Q 表表示。所以选择使用 DQN

Policy gradient:

选择的理由：更好的收敛性质

能有效处理连续空间，高维度资讯

能学习复杂的策略

2. Q-learning is an off-policy algorithm, since it updates the Q values without

making any assumptions about the actual policy being followed.

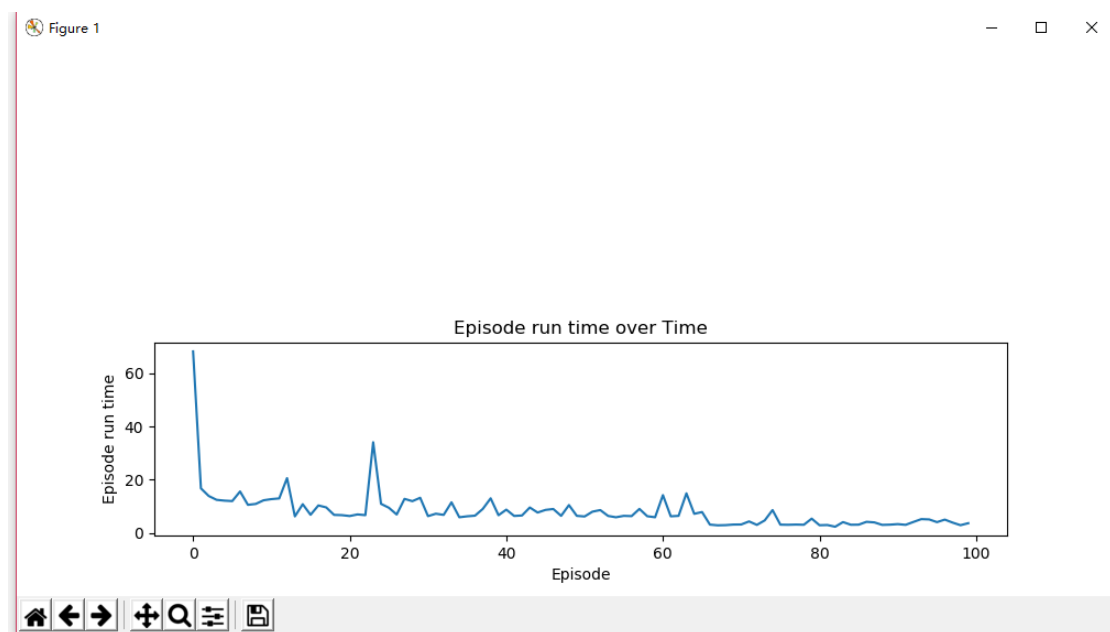
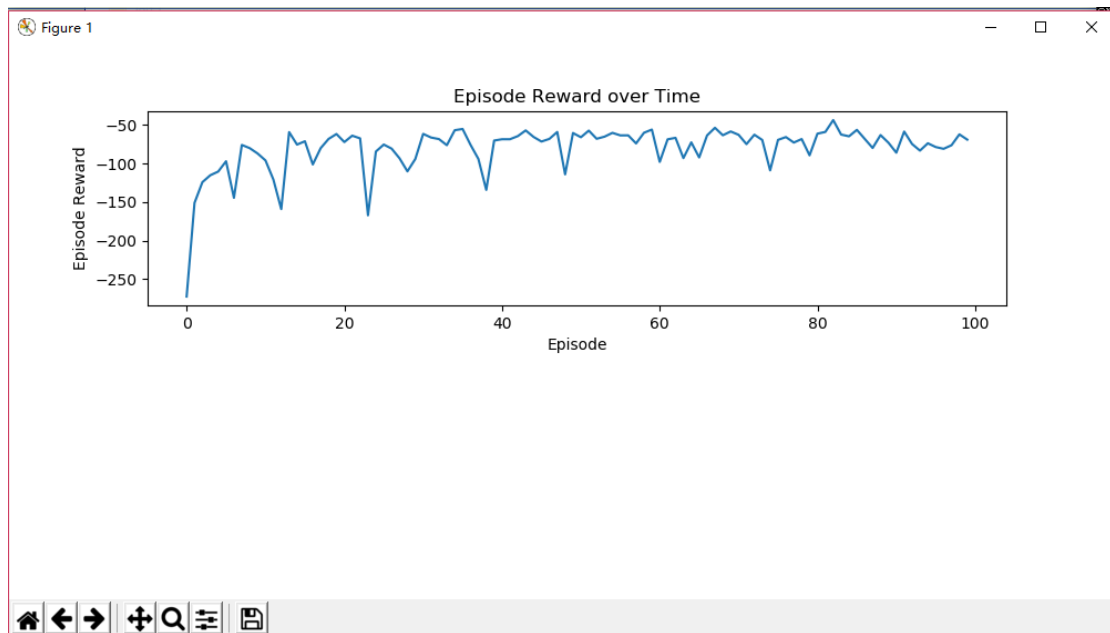
Policy gradient is an on-policy algorithm, which will directly learn the policy on each state

3. It uses a replay buffer to store the **experiences** of the agent during training, and then randomly sample experiences to use for learning in order to break up the temporal correlations within different training episodes. This technique is known as **experience replay**.

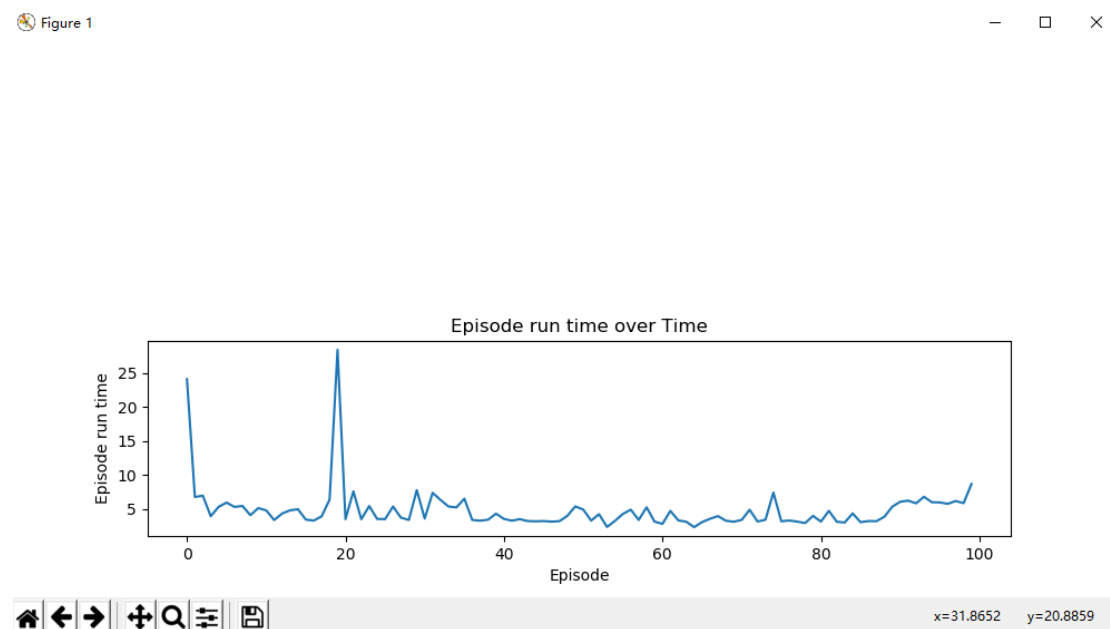
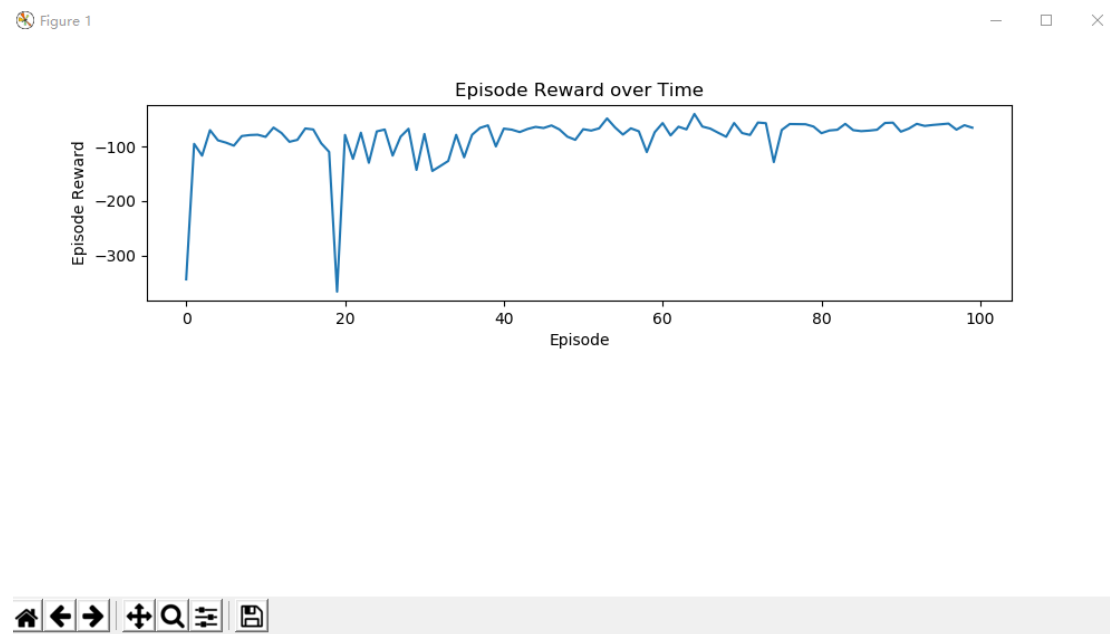
实验结果分析：

不同 batch_size 对 train 影响：

Batch_size 32:

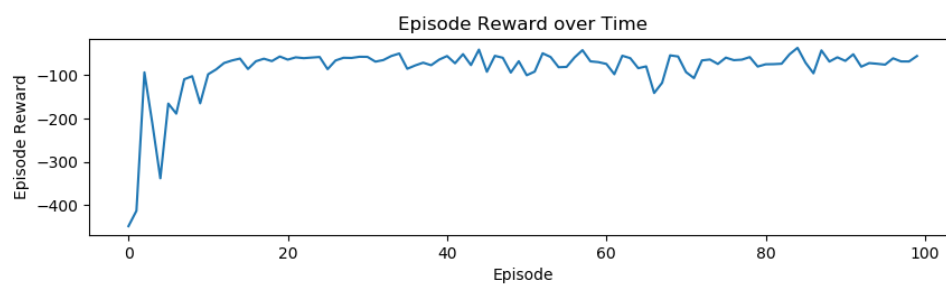


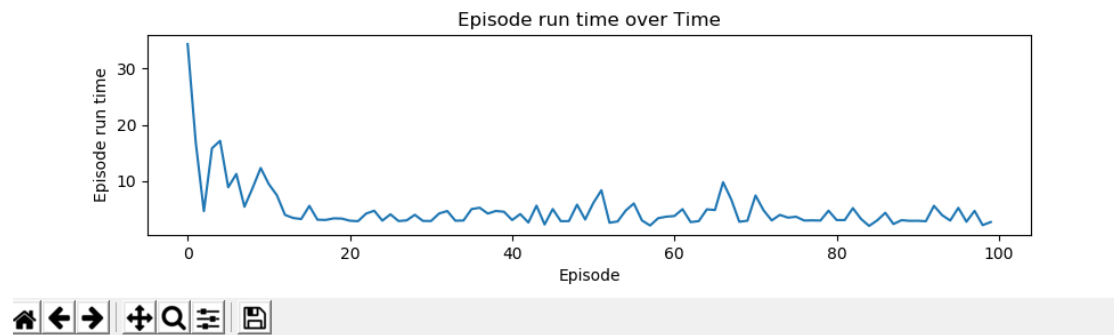
Batch_size:128



Batch_size 16:

Figure 1





实验结果显示：batch_size 越大收敛的越快，但在训练过程中会出现比较大的震荡，小的 batch_size 会比较平滑的收敛。

Learning_rate:0.1

Figure 1

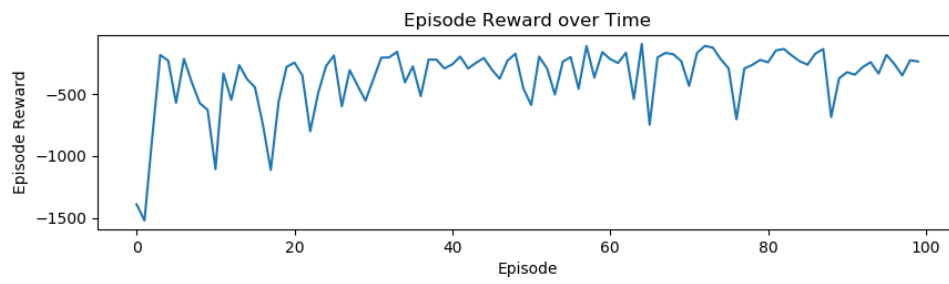
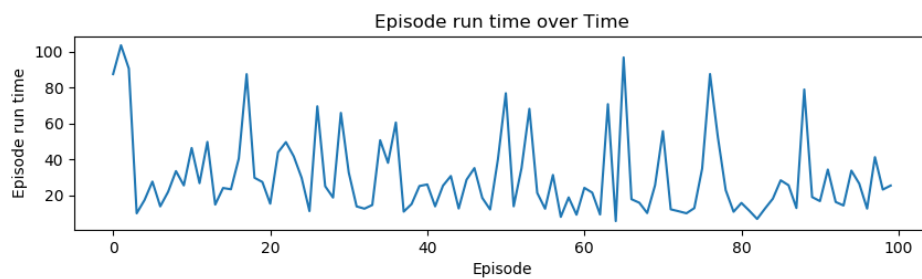


Figure 1



x=32.0057 y=45.006

Learning rate:0.001

Figure 1

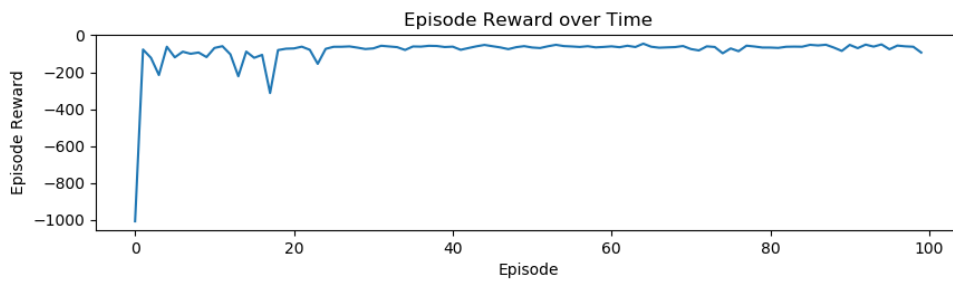
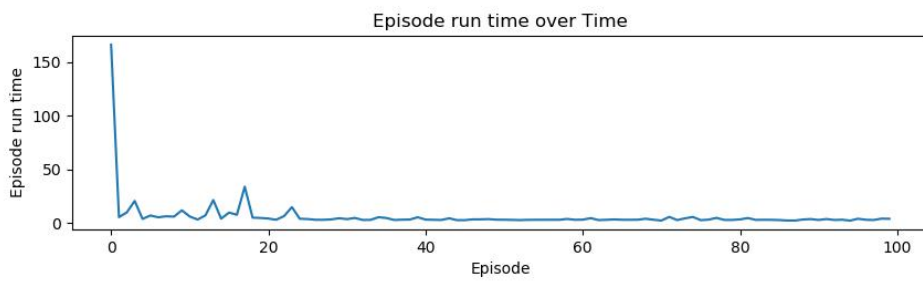
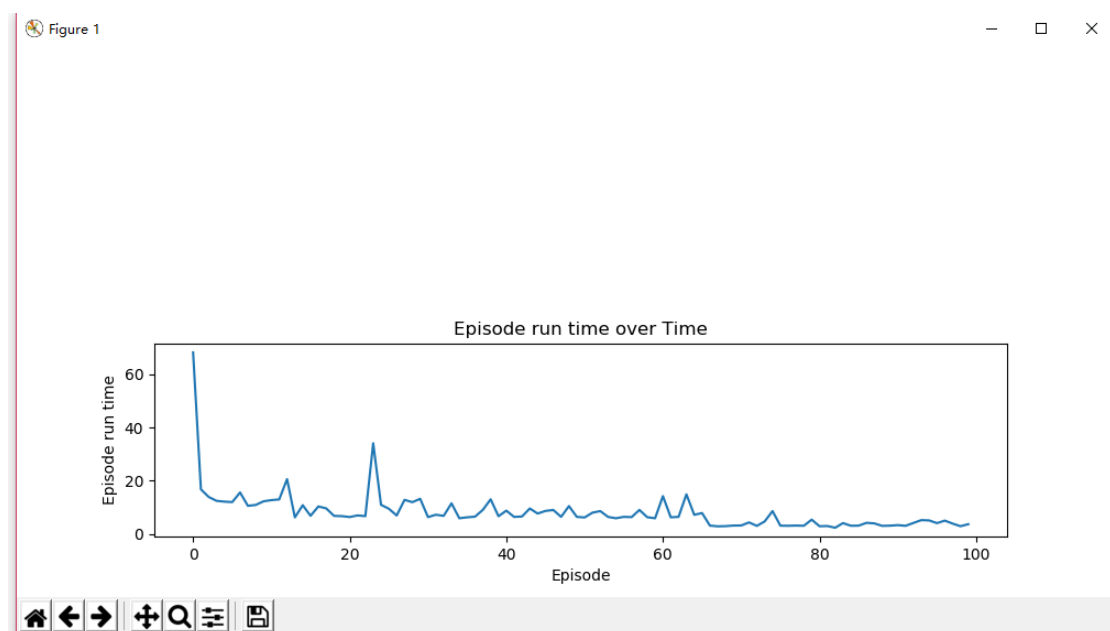
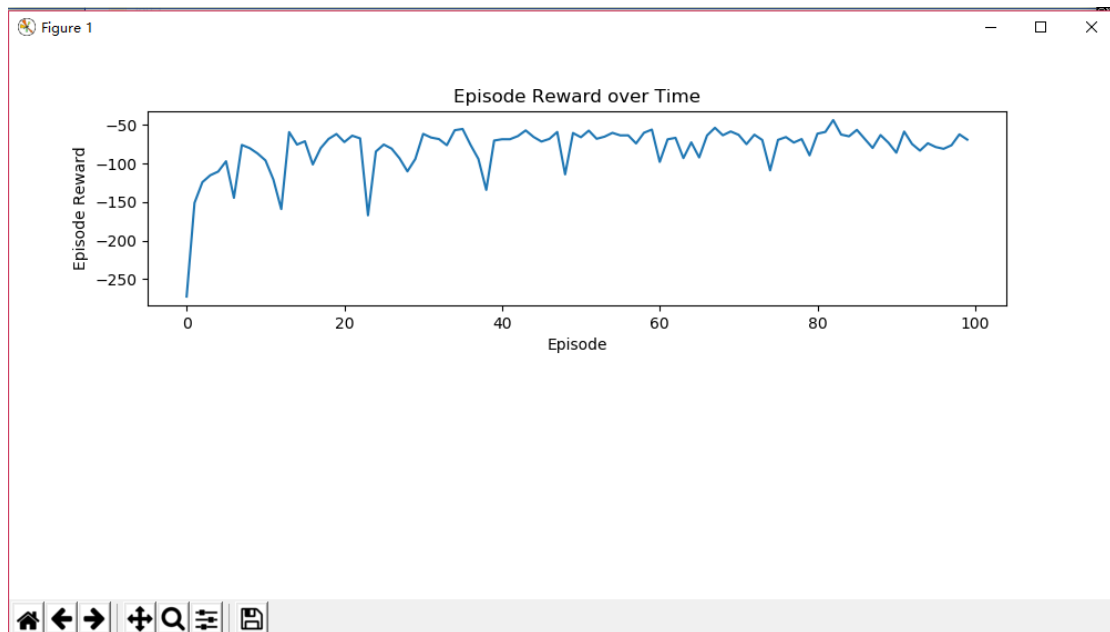


Figure 1



Learning rate 0.01 :



Learning rate 0.1 可以比较快速的收敛，但是很难学习到 global minimal，

Learning rate 0.01，和 0.001 都可以收敛到比较好的结果，并且 learning rate 0.01 有更快的收敛速度，但 learning rate 0.001 有更少的震荡 故选择 learning rate 0.001

Test every 100 epoch in policy gradient：

100：

```
episode: 100 episode_reward: 2819
6.999279975891113
episode: 0 episode_reward 0.00
1.8969252109527588
episode: 1 episode_reward 0.00
2.594059705734253
episode: 2 episode_reward 0.00
0.8487305641174316
episode: 3 episode_reward 0.00
5.1671788692474365
episode: 4 episode_reward 0.00
2.1771762371063232
episode: 5 episode_reward 0.00
2.129305124282837
episode: 6 episode_reward 0.00
3.068793296813965
episode: 7 episode_reward 0.00
0.860694169998169
episode: 8 episode_reward 0.00
7.122951030731201
```

200：

```
episode: 200 episode_reward: 1001  
3.3949201107025146  
episode: 0 episode_reward 0.00  
2.622981548309326  
episode: 1 episode_reward 0.00  
5.055476427078247  
episode: 2 episode_reward 0.00  
4.509939670562744  
episode: 3 episode_reward 0.00  
9.58536148071289  
episode: 4 episode_reward 0.00  
6.712046146392822  
episode: 5 episode_reward 0.00  
3.116661310195923  
episode: 6 episode_reward 0.00  
2.2639455795288086  
episode: 7 episode_reward 0.00  
3.6302881240844727  
episode: 8 episode_reward 0.00  
6.542503833770752
```

300:

```
1.3813037872314453  
episode: 0 episode_reward 0.00  
0.5295841693878174  
episode: 1 episode_reward 0.00  
1.5628182888031006  
episode: 2 episode_reward 0.00  
0.7480006217956543  
episode: 3 episode_reward 0.00  
1.4989893436431885  
episode: 4 episode_reward 0.00  
7.878925800323486  
episode: 5 episode_reward 0.00  
7.952727794647217  
episode: 6 episode_reward 0.00  
3.114668130874634  
episode: 7 episode_reward 0.00  
4.71538782119751  
episode: 8 episode_reward 0.00  
1.8051707744598389  
episode: 9 episode_reward 0.00
```

400:

```
2.383891733838438
episode: 0 episode_reward 0.00
0.7699398994445801
episode: 1 episode_reward 0.00
3.0677952766418457
episode: 2 episode_reward 0.00
1.278580665588379
episode: 3 episode_reward 0.00
2.2649407386779785
episode: 4 episode_reward 0.00
5.070439100265503
episode: 5 episode_reward 0.00
1.0731298923492432
episode: 6 episode_reward 0.00
1.225722312927246
episode: 7 episode_reward 0.00
1.0821056365966797
episode: 8 episode_reward 0.00
1.5408785343170166
episode: 9 episode_reward 0.00
```

500:

```
episode: 500 episode_reward 0.00
1.8660078048706055
episode: 0 episode_reward 0.00
7.102005243301392
episode: 1 episode_reward 0.00
1.7622864246368408
episode: 2 episode_reward 0.00
1.6356244087219238
episode: 3 episode_reward 0.00
2.978034496307373
episode: 4 episode_reward 0.00
5.061460971832275
episode: 5 episode_reward 0.00
5.669813394546509
episode: 6 episode_reward 0.00
1.0581672191619873
episode: 7 episode_reward 0.00
2.080434560775757
episode: 8 episode_reward 0.00
1.5707998275756836
episode: 9 episode_reward 0.00
```

600:

```
1.7562737464904785
episode: 0 episode_reward 0.00
11.016526937484741
episode: 1 episode_reward 0.00
2.3816323280334473
episode: 2 episode_reward 0.00
4.597683668136597
episode: 3 episode_reward 0.00
10.205702066421509
episode: 4 episode_reward 0.00
0.7759242057800293
episode: 5 episode_reward 0.00
1.1309764385223389
episode: 6 episode_reward 0.00
2.0983877182006836
episode: 7 episode_reward 0.00
6.635253190994263
episode: 8 episode_reward 0.00
2.318798065185547
episode: 9 episode_reward 0.00
```

700:

```
episode: 0 episode_reward 0.00
3.1814913749694824
episode: 1 episode_reward 0.00
1.648590087890625
episode: 2 episode_reward 0.00
2.848381757736206
episode: 3 episode_reward 0.00
4.639591455459595
episode: 4 episode_reward 0.00
5.696763277053833
episode: 5 episode_reward 0.00
1.2785804271697998
episode: 6 episode_reward 0.00
1.3603615760803223
episode: 7 episode_reward 0.00
1.2686069011688232
episode: 8 episode_reward 0.00
3.2163984775543213
episode: 9 episode_reward 0.00
```

800:

```
1.0342347621917725
episode: 0 episode_reward 0.00
5.200090646743774
episode: 1 episode_reward 0.00
2.8792998790740967
episode: 2 episode_reward 0.00
3.6412606239318848
episode: 3 episode_reward 0.00
2.3467226028442383
episode: 4 episode_reward 0.00
1.4650804996490479
episode: 5 episode_reward 0.00
1.7932043075561523
episode: 6 episode_reward 0.00
3.2832188606262207
episode: 7 episode_reward 0.00
2.744659185409546
episode: 8 episode_reward 0.00
1.9328303337097168
episode: 9 episode_reward 0.00
```

900:

```
4.2935168743133545
episode: 0 episode_reward 0.00
5.543172597885132
episode: 1 episode_reward 0.00
0.7978391647338867
episode: 2 episode_reward 0.00
2.6658685207366943
episode: 3 episode_reward 0.00
3.364629030227661
episode: 4 episode_reward 0.00
2.460397958755493
episode: 5 episode_reward 0.00
6.779865503311157
episode: 6 episode_reward 0.00
1.035231590270996
episode: 7 episode_reward 0.00
3.103700637817383
episode: 8 episode_reward 0.00
1.4381535053253174
episode: 9 episode_reward 0.00
```

Policy gradient: reward time

