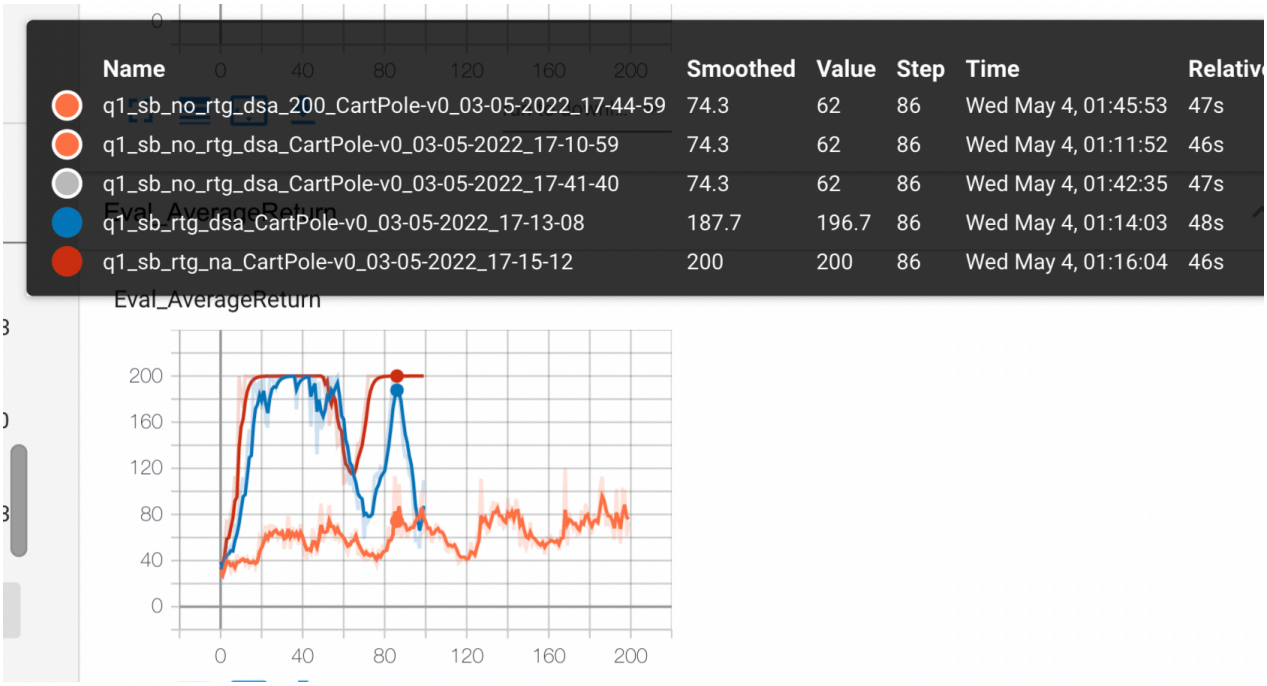
hw2

sec1

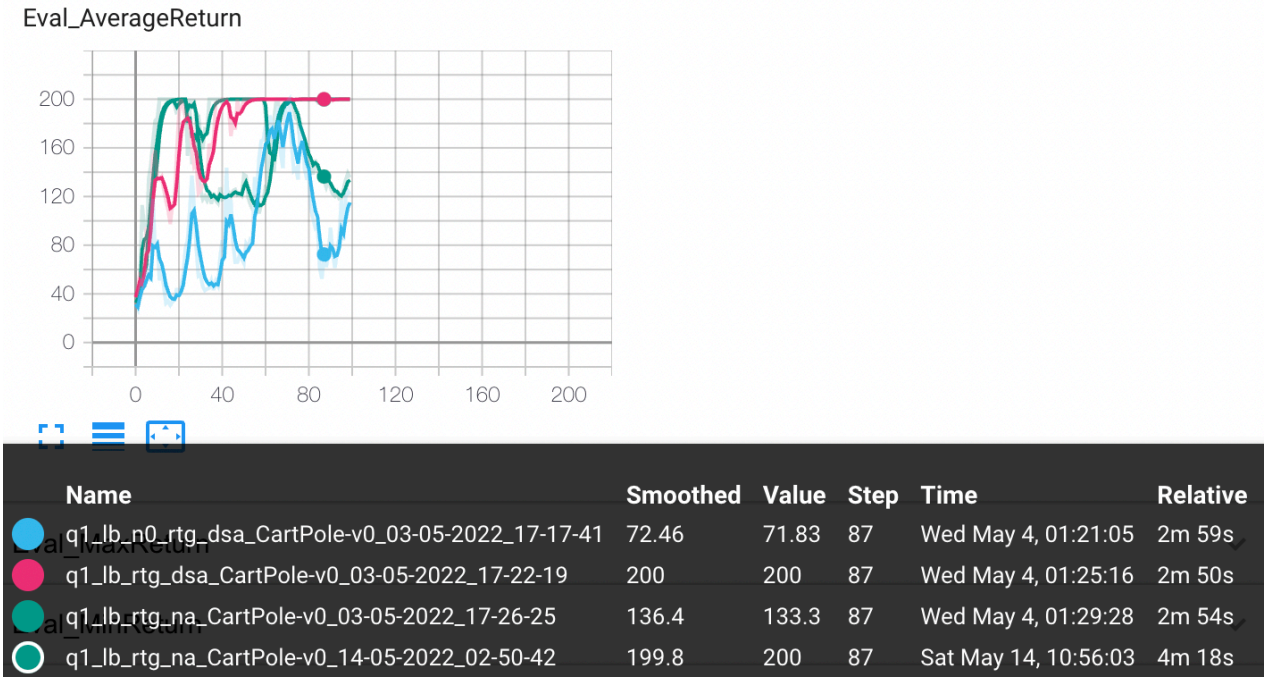
1. Deliverables for report:

   • Create two graphs:

   - – In the first graph, compare the learning curves (average return at each iteration) for the experiments prefixed with q1_sb_. (The small batch experiments.)



   - – In the second graph, compare the learning curves for the experiments prefixed with q1_lb_. (The large batch experiments.)

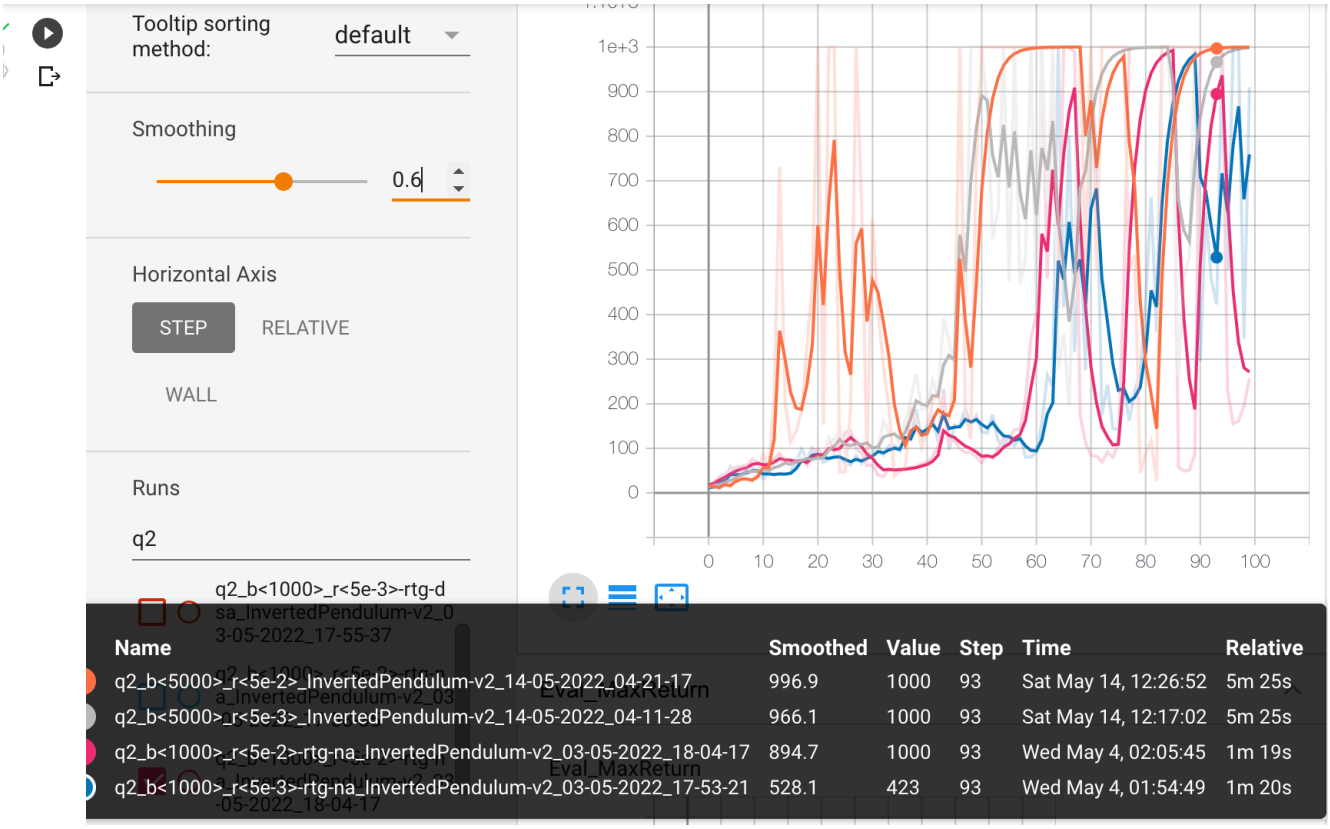


   • Answer the following questions briefly:

- – Which value estimator has better performance without advantage-standardization: the trajectory- centric one, or the one using reward-to-go?
- – Did advantage standardization help? Yes,but indistinctive.
- – Did the batch size make an impact?

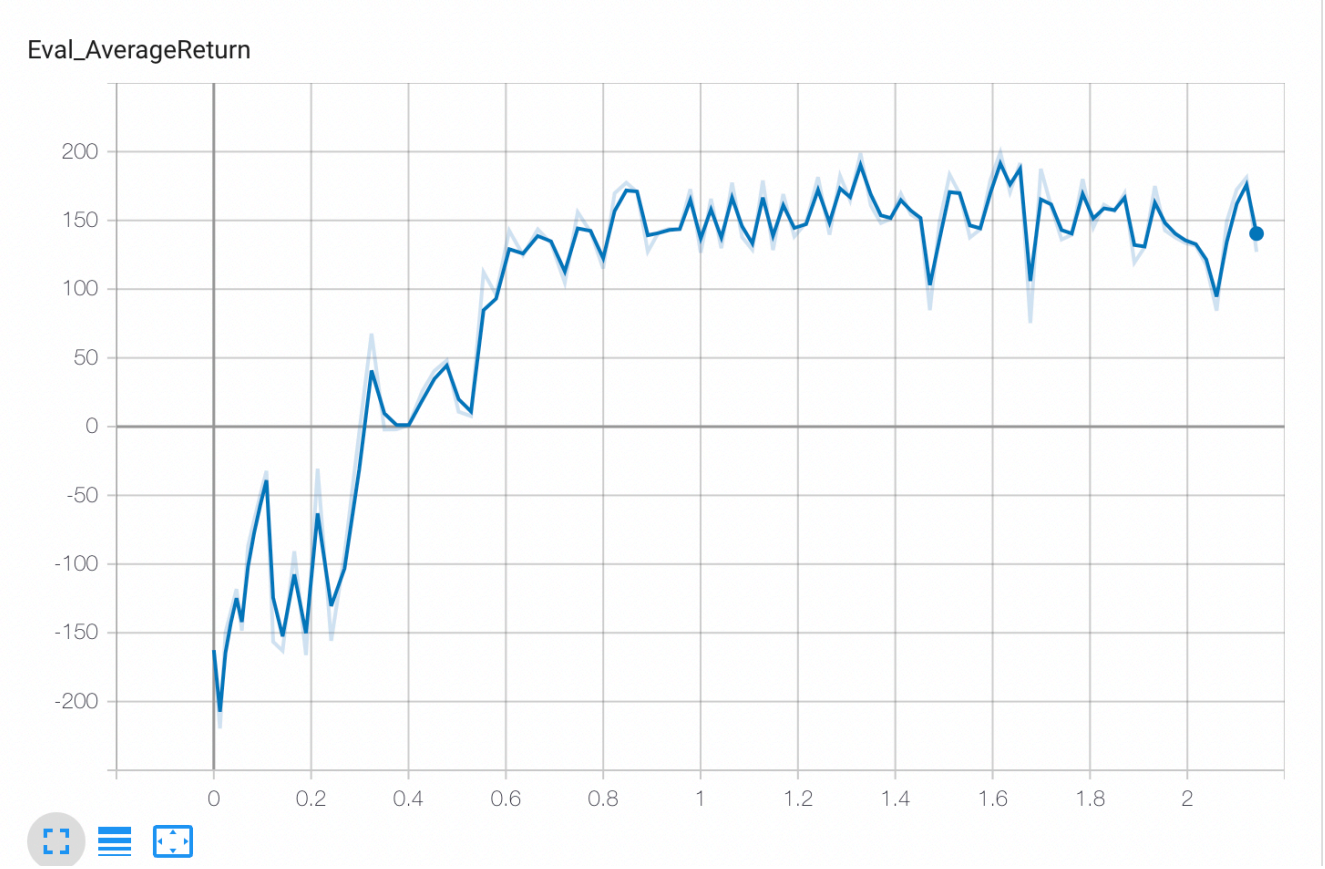
sec2

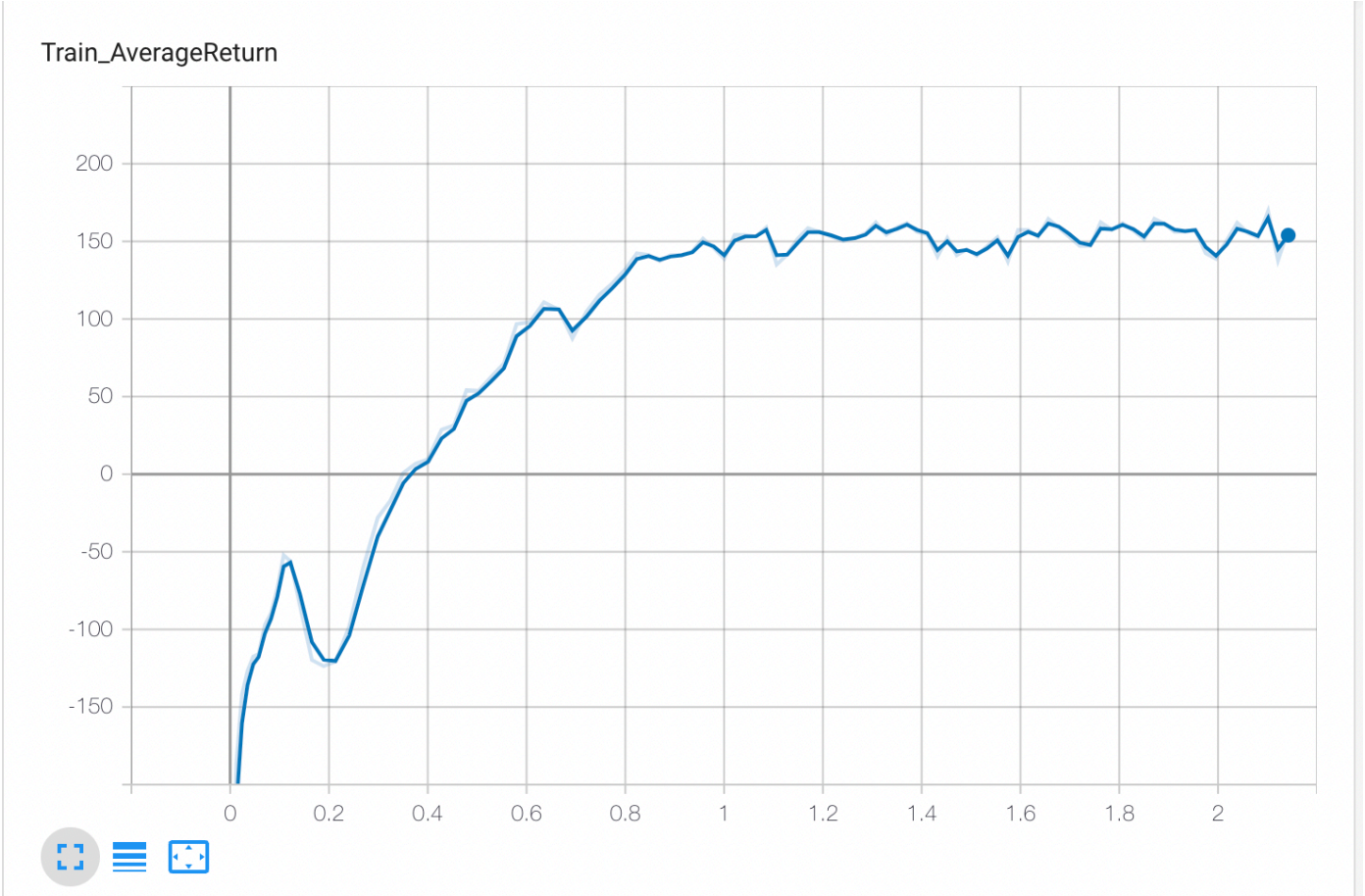
- Large batch size:stable, low variance.
- Small learning rate: converges quicker but unstable



sec3

- maxreturn is over 200

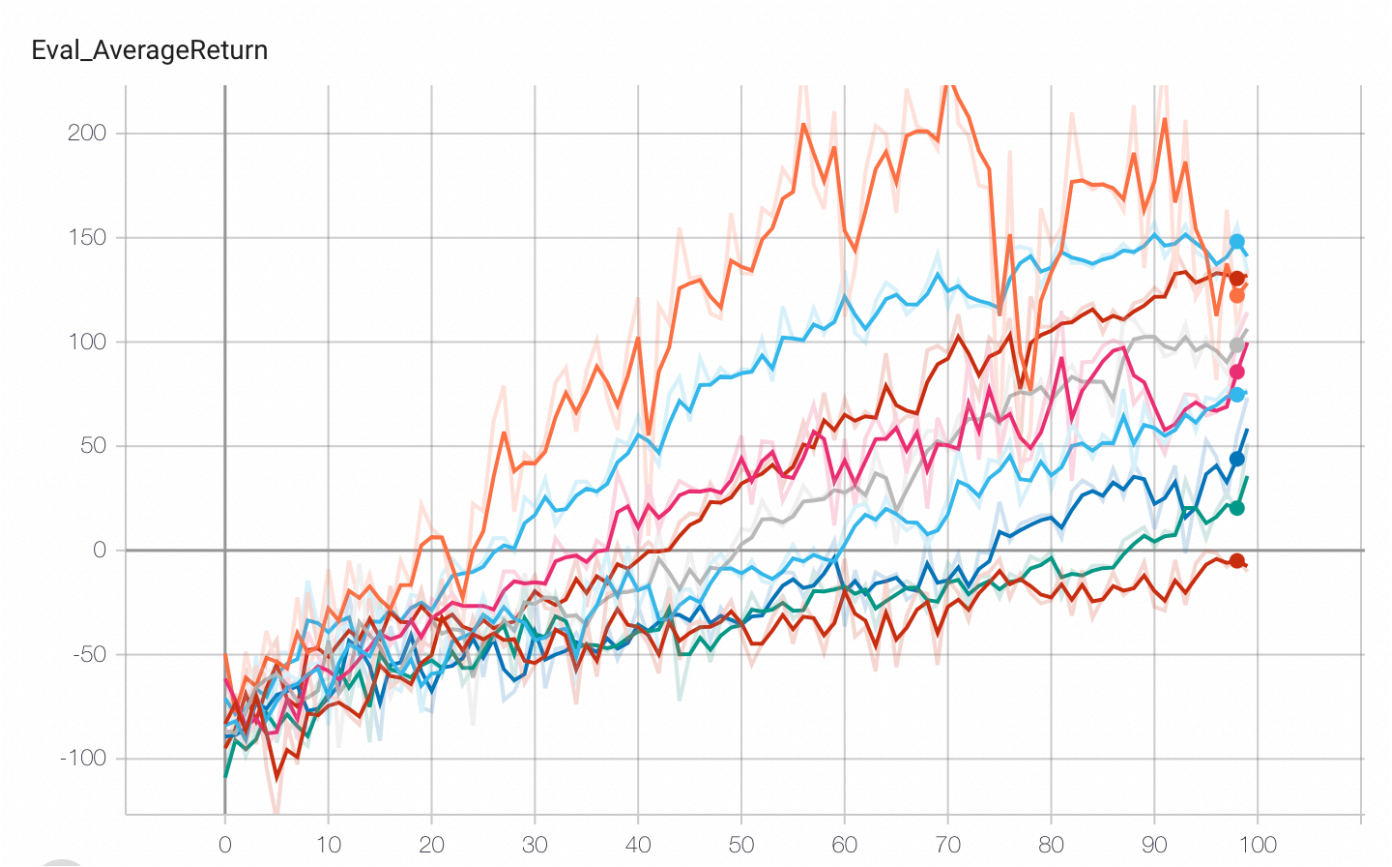




(lec7)要不知道mdp，要不在一个状态用同个策略做好多次（来模拟出状态转移概率）。

而利用q-function，不依赖于策略，只依赖于mdp,在任何策略下采样都可以(模拟mdp)，不知道mdp也无妨

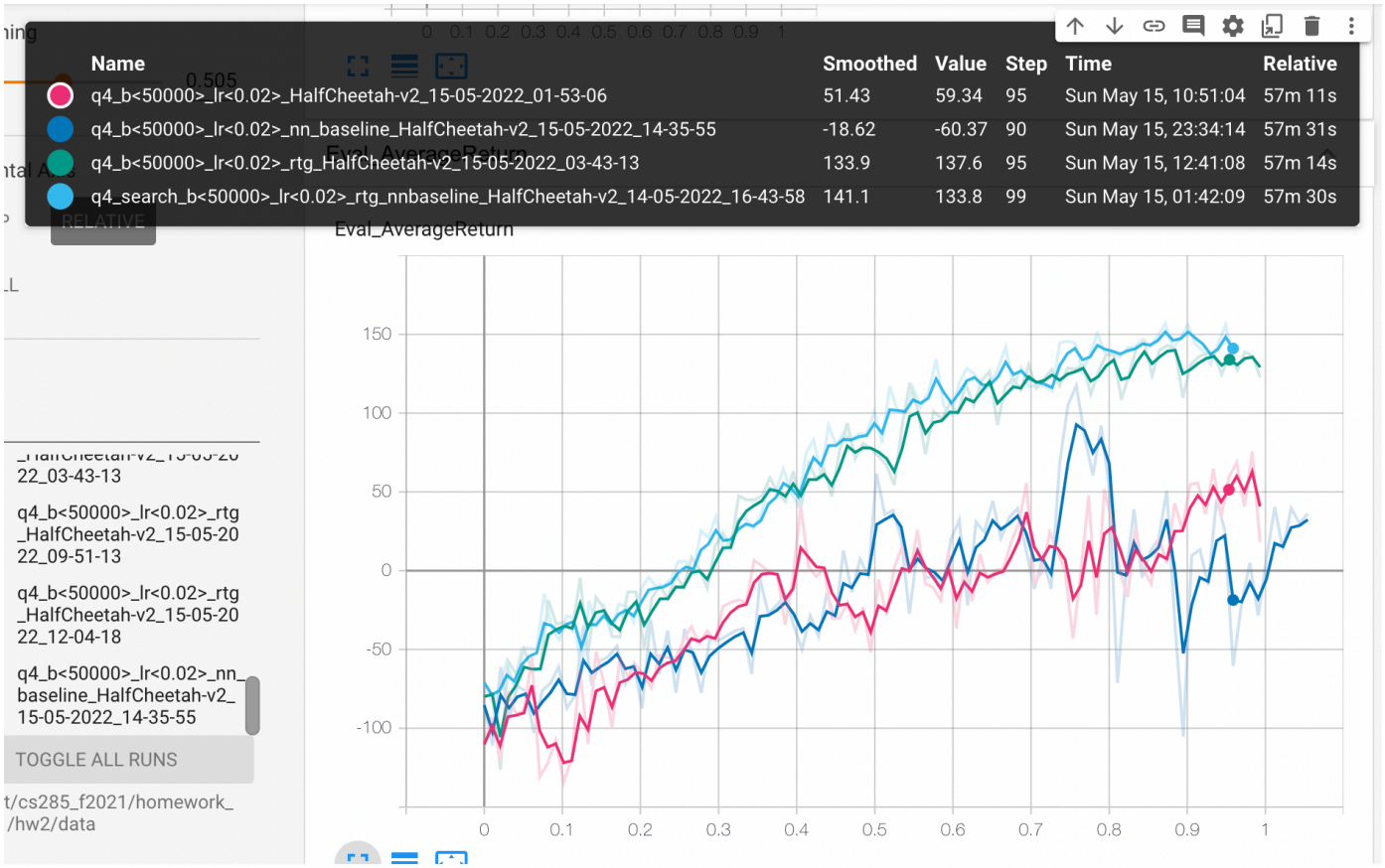
sec4



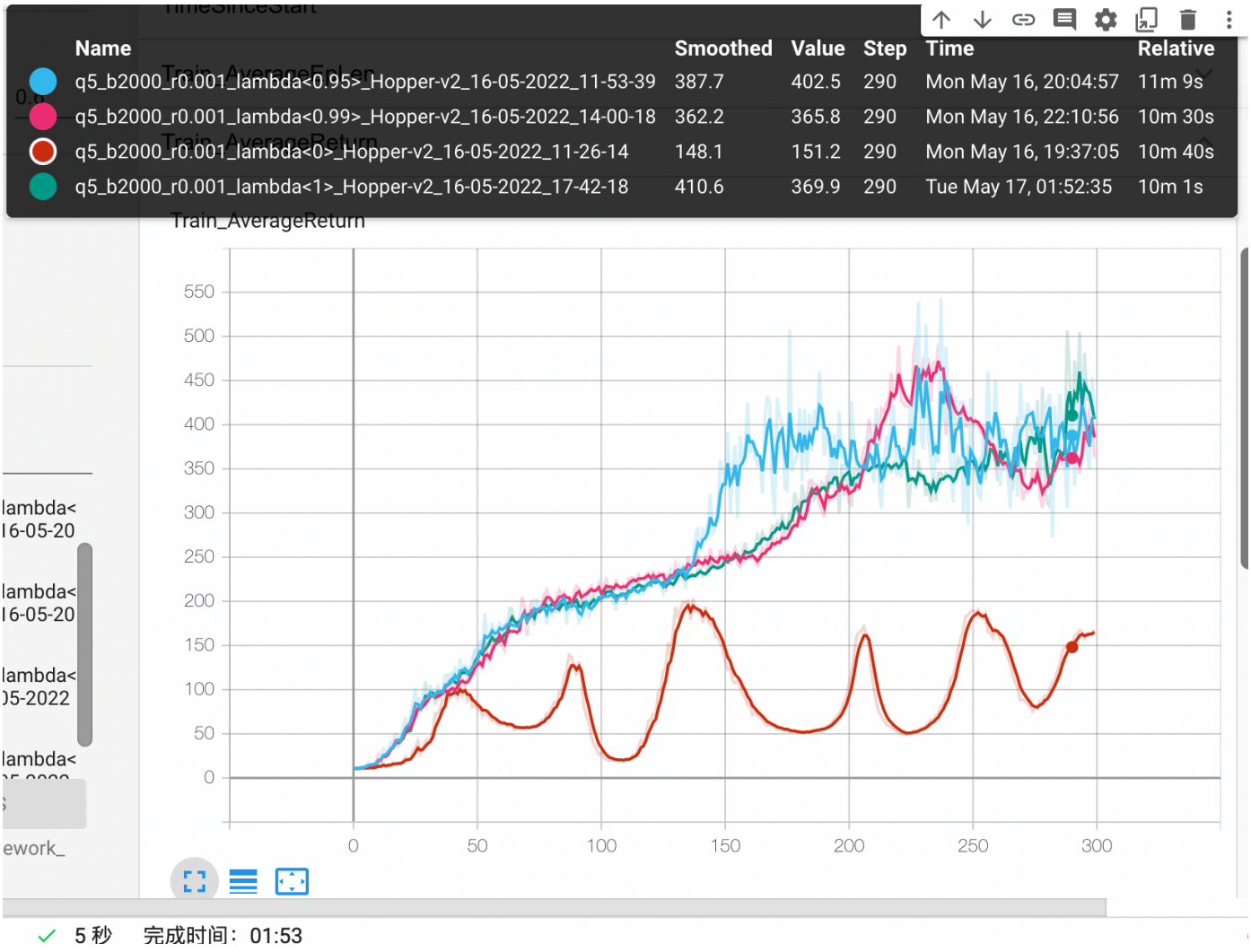
Name	Smoothed	Value	Step	Time	Relative
q4_search_b<10000>_lr<0.005>_rtg_nnbaseline_HalfCheetah-v2_14-05-2022_11-01-38	-29.28	-24.19	49	Sat May 14, 19:07:37	5m 44s
q4_search_b<10000>_lr<0.01>_rtg_nnbaseline_HalfCheetah-v2_14-05-2022_11-20-55	-8.62	-8.496	49	Sat May 14, 19:26:54	5m 46s
q4_search_b<10000>_lr<0.02>_rtg_nnbaseline_HalfCheetah-v2_14-05-2022_11-38-20	33.62	39.67	49	Sat May 14, 19:44:31	5m 47s
q4_search_b<30000>_lr<0.005>_rtg_nnbaseline_HalfCheetah-v2_14-05-2022_11-54-53	-36.98	-33.02	49	Sat May 14, 20:12:03	16m 44s
q4_search_b<30000>_lr<0.01>_rtg_nnbaseline_HalfCheetah-v2_14-05-2022_12-29-30	-4.106	0.877	49	Sat May 14, 20:47:20	17m 23s
q4_search_b<30000>_lr<0.02>_rtg_nnbaseline_HalfCheetah-v2_14-05-2022_13-09-41	139	161.8	49	Sat May 14, 21:31:05	20m 49s
q4_search_b<50000>_lr<0.005>_rtg_nnbaseline_HalfCheetah-v2_14-05-2022_14-31-09	-33.34	-35.37	49	Sat May 14, 23:00:21	28m 24s
q4_search_b<50000>_lr<0.01>_rtg_nnbaseline_HalfCheetah-v2_14-05-2022_15-31-34	25.41	27.94	49	Sun May 15, 00:00:26	28m 10s
q4_search_b<50000>_lr<0.02>_rtg_nnbaseline_HalfCheetah-v2_14-05-2022_16-43-58	83.09	82.7	49	Sun May 15, 01:12:50	28m 11s

b<30000> lr<0.02>提升很快，方差极大。最终最好的是b<50000>lr<0.02>

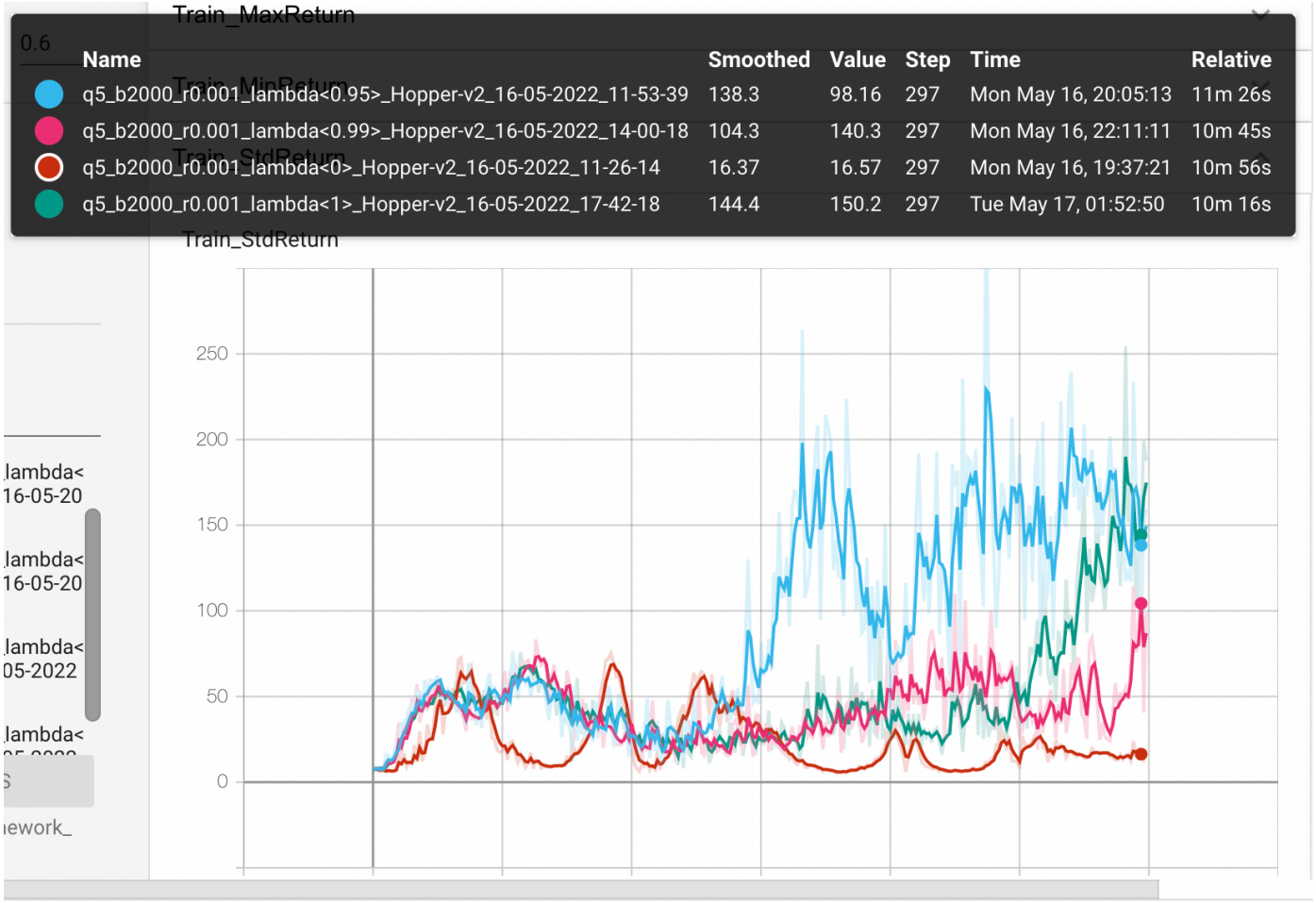
rtg nn-baseline对照


Name
Smoothed
Value
Step
Time
Relative
q4_b<50000>_lr<0.02>_HalfCheetah-v2_15-05-2022_01-53-06
51.43
59.34
95
Sun May 15, 10:51:04
57m 11s
q4_b<50000>_lr<0.02>_nn_baseline_HalfCheetah-v2_15-05-2022_14-35-55
-18.62
-60.37
90
Sun May 15, 23:34:14
57m 31s
q4_b<50000>_lr<0.02>_rtg_HalfCheetah-v2_15-05-2022_03-43-13
133.9
137.6
95
Sun May 15, 12:41:08
57m 14s
q4_search_b<50000>_lr<0.02>_rtg_nnbaseline_HalfCheetah-v2_14-05-2022_16-43-58
141.1
133.8
99
Sun May 15, 01:42:09
57m 30s
Eval_AverageReturn
150
100
50
0
-50
-100
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
22_03-43-13
q4_b<50000>_lr<0.02>_rtg
_HalfCheetah-v2_15-05-20
22_09-51-13
q4_b<50000>_lr<0.02>_rtg
_HalfCheetah-v2_15-05-20
22_12-04-18
q4_b<50000>_lr<0.02>_nn_
baseline_HalfCheetah-v2_
15-05-2022_14-35-55
TOGGLE ALL RUNS
t/cs285_f2021/homework_
/hw2/data

sec5


Name
Smoothed
Value
Step
Time
Relative
q5_b2000_r0.001_lambda<0.95>_Hopper-v2_16-05-2022_11-53-39 387.7 402.5 290 Mon May 16, 20:04:57 11m 9s
q5_b2000_r0.001_lambda<0.99>_Hopper-v2_16-05-2022_14-00-18 362.2 365.8 290 Mon May 16, 22:10:56 10m 30s
q5_b2000_r0.001_lambda<0>_Hopper-v2_16-05-2022_11-26-14 148.1 151.2 290 Mon May 16, 19:37:05 10m 40s
q5_b2000_r0.001_lambda<1>_Hopper-v2_16-05-2022_17-42-18 410.6 369.9 290 Tue May 17, 01:52:35 10m 1s
Train_AverageReturn
550
500
450
400
350
300
250
200
150
100
50
0
0
50
100
150
200
250
300

✓ 5 秒 完成时间：01:53


Train_MaxReturn
Train_MinReturn
Train_StdReturn
Train_StdReturn
0.6
Name
Smoothed
Value
Step
Time
Relative
q5_b2000_r0.001_lambda<0.95>_Hopper-v2_16-05-2022_11-53-39
138.3
98.16
297
Mon May 16, 20:05:13
11m 26s
q5_b2000_r0.001_lambda<0.99>_Hopper-v2_16-05-2022_14-00-18
104.3
140.3
297
Mon May 16, 22:11:11
10m 45s
q5_b2000_r0.001_lambda<0>_Hopper-v2_16-05-2022_11-26-14
16.37
16.57
297
Mon May 16, 19:37:21
10m 56s
q5_b2000_r0.001_lambda<1>_Hopper-v2_16-05-2022_17-42-18
144.4
150.2
297
Tue May 17, 01:52:50
10m 16s
250
200
150
100
50
0
lambda<
16-05-20
lambda<
16-05-20
lambda<
05-2022
lambda<
ework_