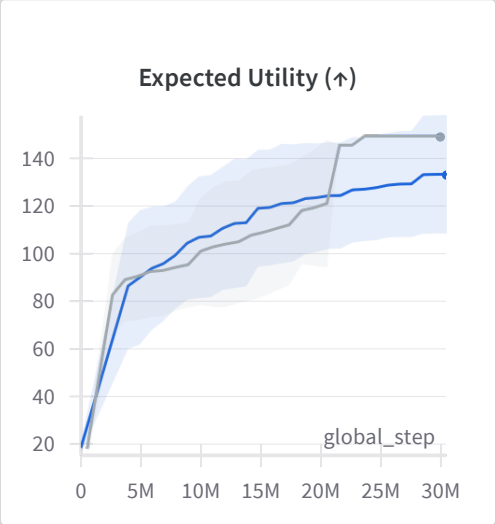
Share Comment Star

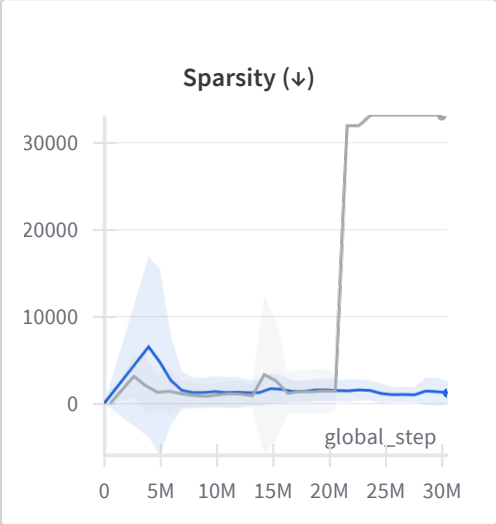
MORL using EAs for Diverse Policy Selection

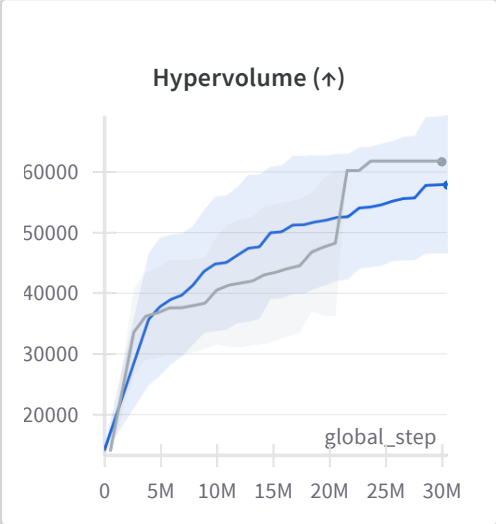
This report presents the results of multi-objective reinforcement learning (MORL) experiments integrated with evolutionary algorithms (EAs) and a novel policy selection mechanism. The work explores policy optimization for environments such as mo-halfcheetah-v4 and mo-hopper-2d-v4 while focusing on balancing multiple competing objectives.

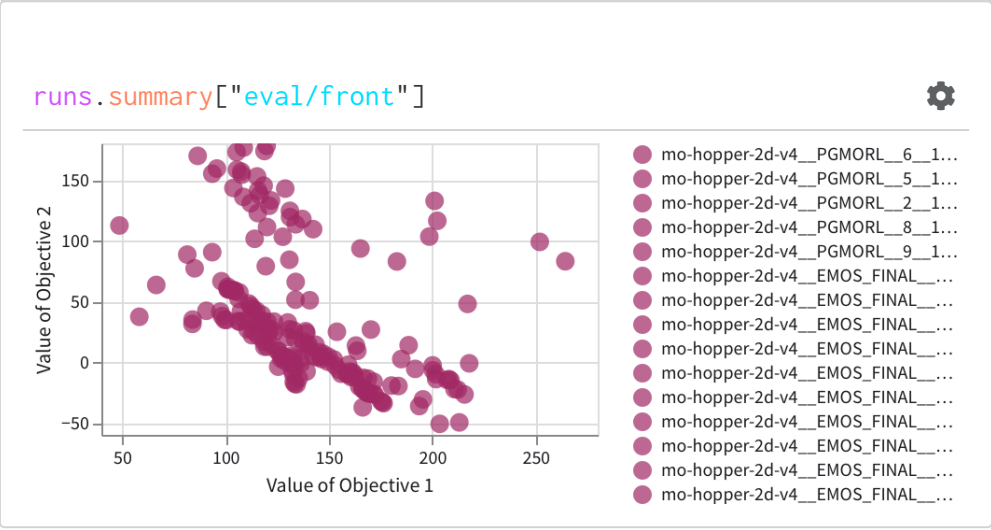
Clarissa Kümhof

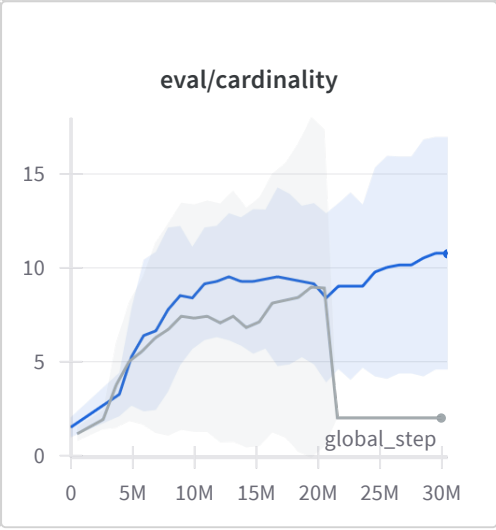
Created on February 11 | Last edited on February 12


Expected Utility (↑)
140
120
100
80
60
40
20
global_step
0
5M
10M
15M
20M
25M
30M

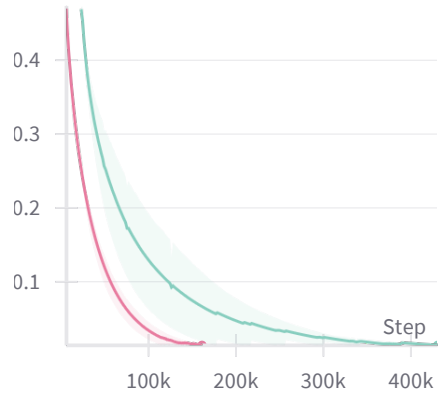

Sparsity (↓)
30000
20000
10000
0
global_step
0
5M
10M
15M
20M
25M
30M


Hypervolume (↑)
60000
50000
40000
30000
20000
global_step
0
5M
10M
15M
20M
25M
30M

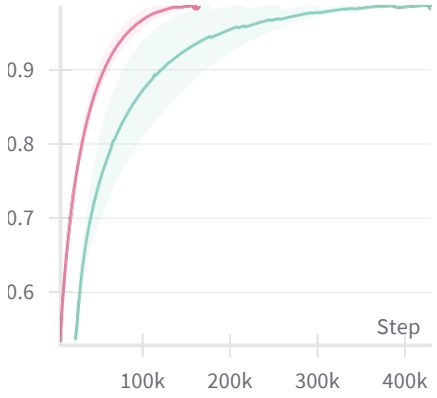

runs.summary["eval/front"]
Value of Objective 2
150
100
50
0
−50
50
100
150
200
250
Value of Objective 1
mo-hopper-2d-v4__PGMORL__6__1...
mo-hopper-2d-v4__PGMORL__5__1...
mo-hopper-2d-v4__PGMORL__2__1...
mo-hopper-2d-v4__PGMORL__8__1...
mo-hopper-2d-v4__PGMORL__9__1...
mo-hopper-2d-v4__EMOS_FINAL__...
mo-hopper-2d-v4__EMOS_FINAL__...
mo-hopper-2d-v4__EMOS_FINAL__...
mo-hopper-2d-v4__EMOS_FINAL__...
mo-hopper-2d-v4__EMOS_FINAL__...
mo-hopper-2d-v4__EMOS_FINAL__...
mo-hopper-2d-v4__EMOS_FINAL__...
mo-hopper-2d-v4__EMOS_FINAL__...
mo-hopper-2d-v4__EMOS_FINAL__...


eval/cardinality
15
10
5
0
global_step
0
5M
10M
15M
20M
25M
30M

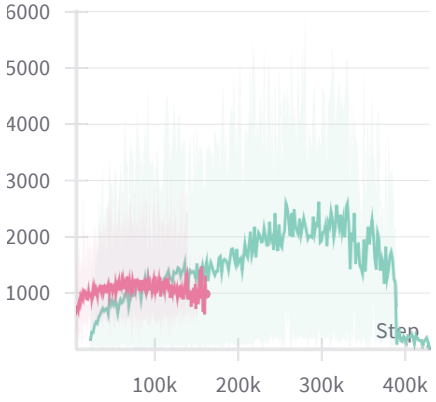
dynamic exploration weight



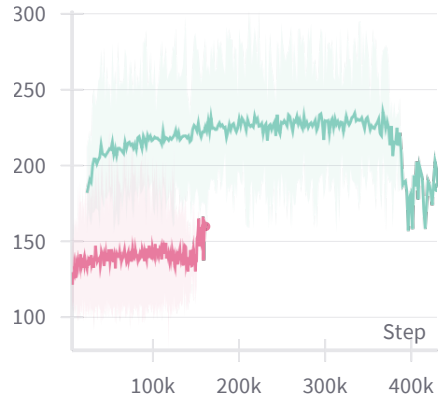
dynamic exploitation weight



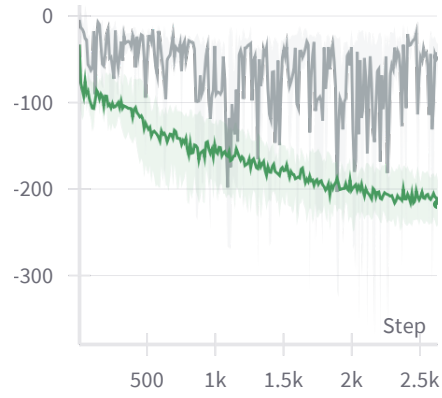
new/average_uncertainty



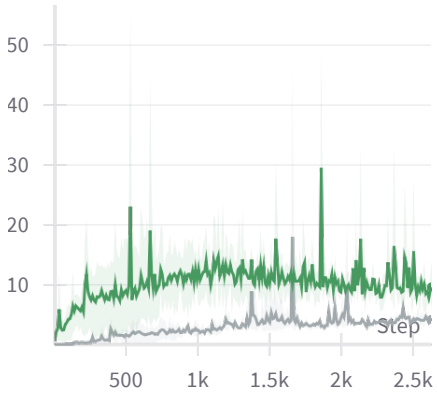
new/average_regret

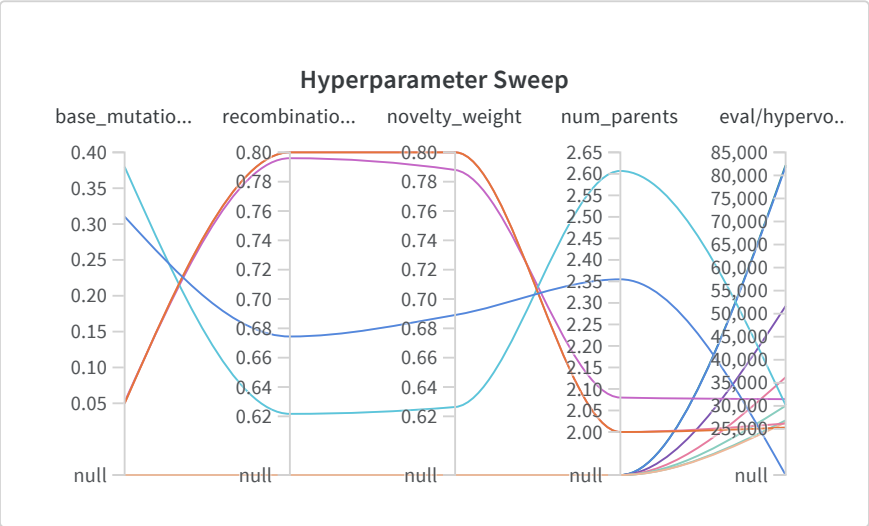


losses/policy_loss

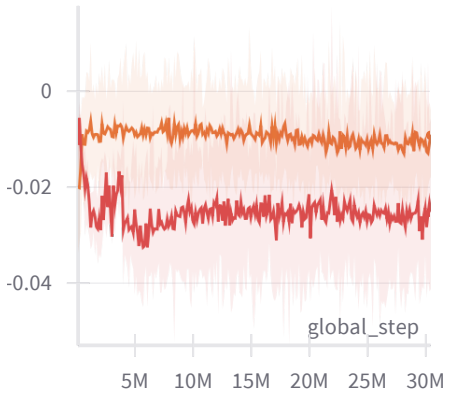


losses/critic_loss

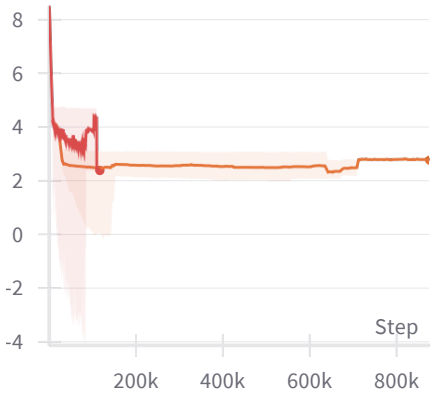



Hyperparameter Sweep
base_mutatio...
recombinatio...
novelty_weight
num_parents
eval/hypervo..
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
null
0.80
0.78
0.76
0.74
0.72
0.70
0.68
0.66
0.64
0.62
null
0.80
0.78
0.76
0.74
0.72
0.70
0.68
0.66
0.64
0.62
null
2.65
2.60
2.55
2.50
2.45
2.40
2.35
2.30
2.25
2.20
2.15
2.10
2.05
2.00
null
85,000
80,000
75,000
70,000
65,000
60,000
55,000
50,000
45,000
40,000
35,000
30,000
25,000
null

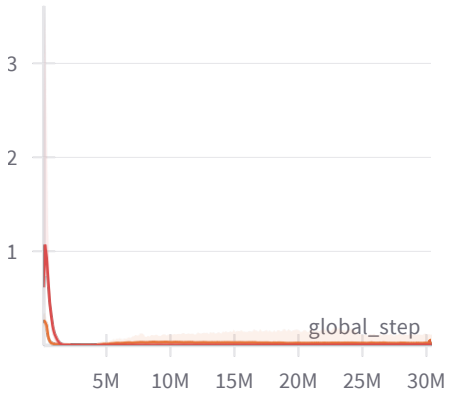
losses_5/policy_loss



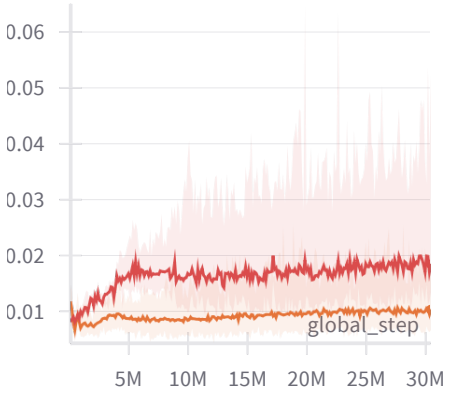
losses_5/entropy



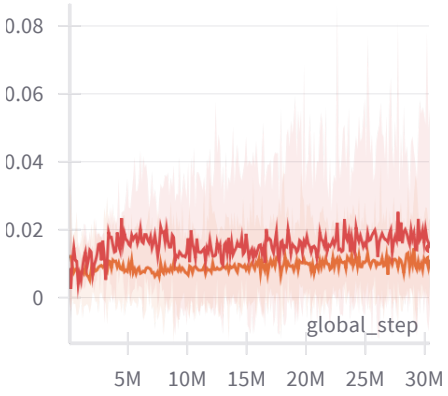
losses_5/value_loss



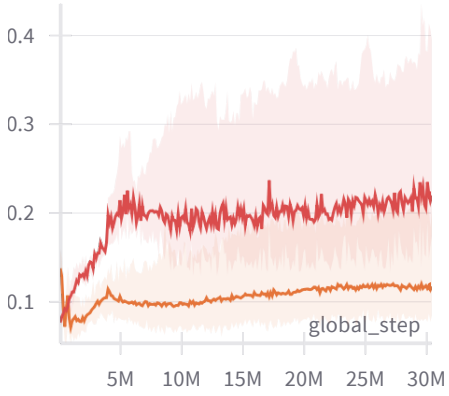
losses_5/approx_kl



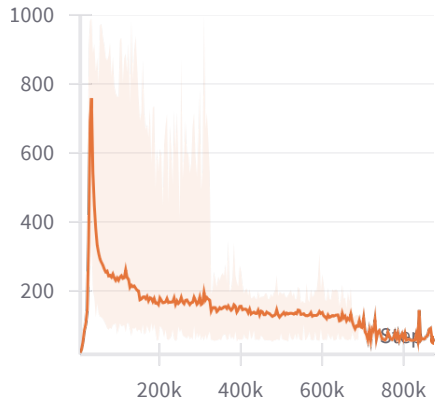
losses_5/old_approx_kl



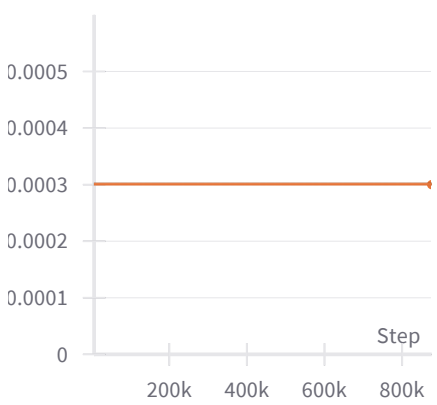
losses_5/clipfrac



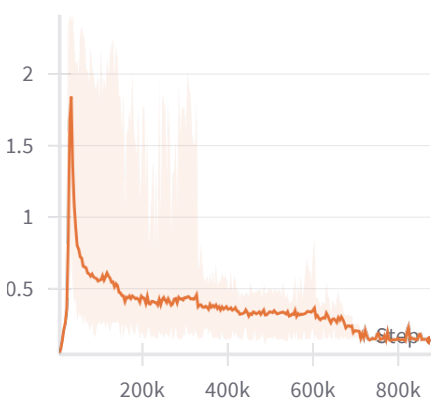
charts_5/timesteps_per_episode



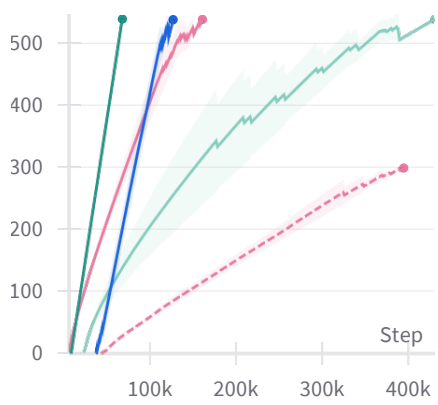
charts_5/learning_rate



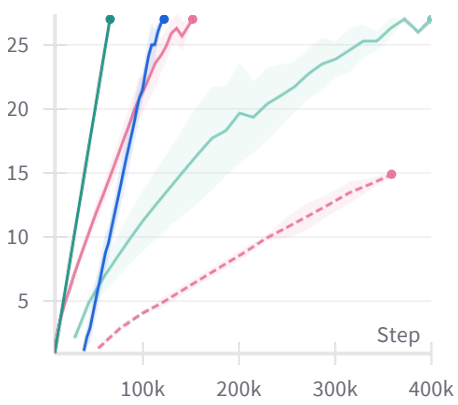
charts_5/episode_time



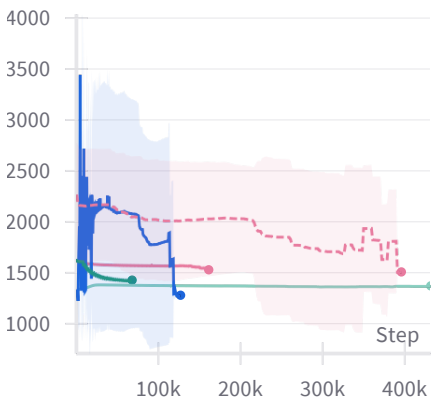
charts/evolutionary_iterations



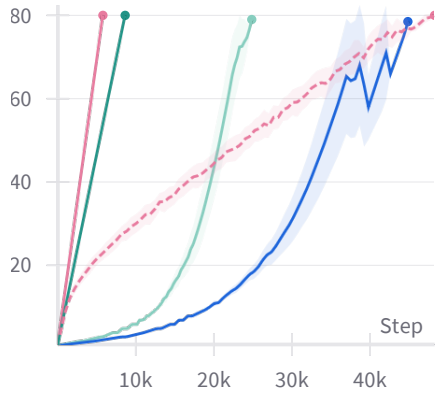
charts/evolutionary_generation

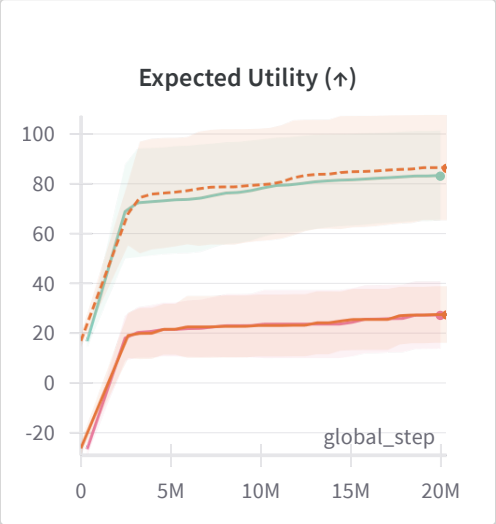


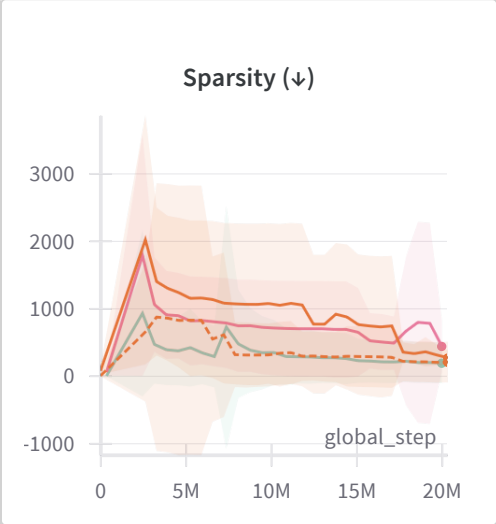
charts/SPS

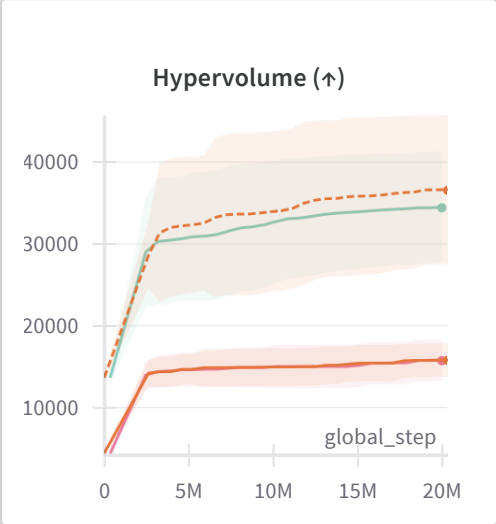


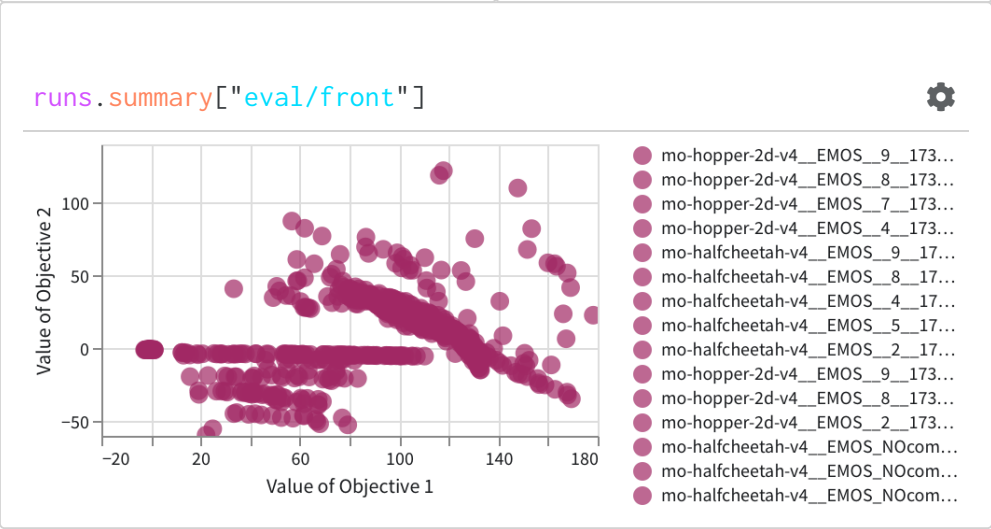
charts/warmup_iterations

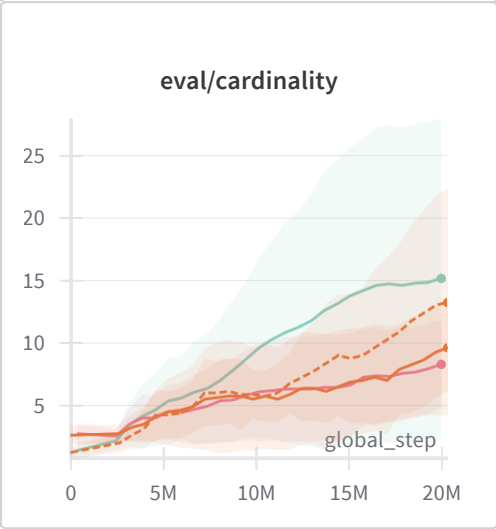


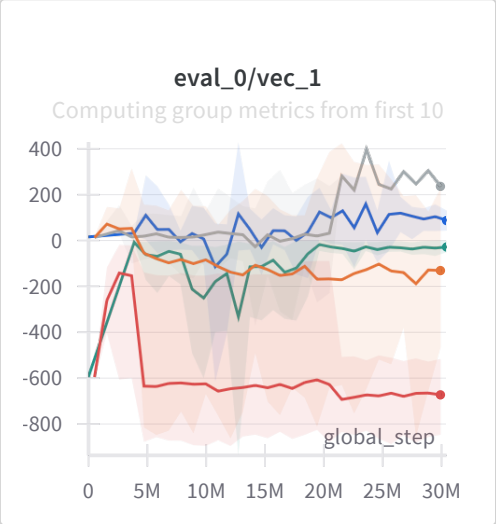

Expected Utility (↑)
100
80
60
40
20
0
-20
global_step
0
5M
10M
15M
20M

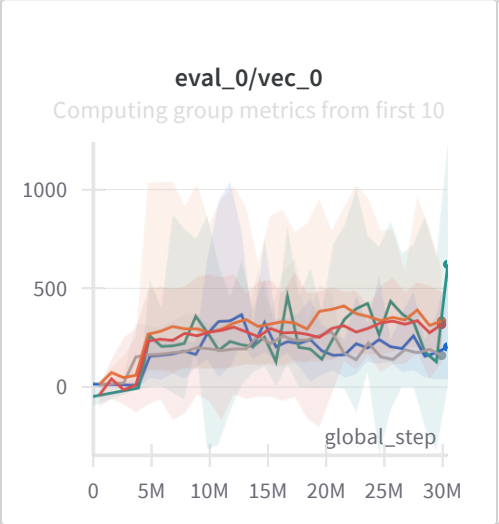

Sparsity (↓)
3000
2000
1000
0
-1000
global_step
0
5M
10M
15M
20M

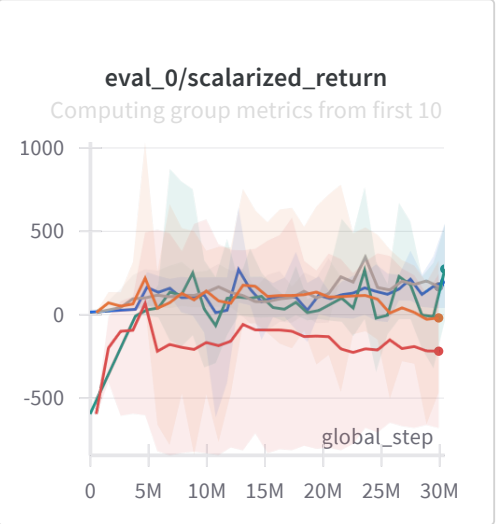

Hypervolume (↑)
40000
30000
20000
10000
global_step
0
5M
10M
15M
20M

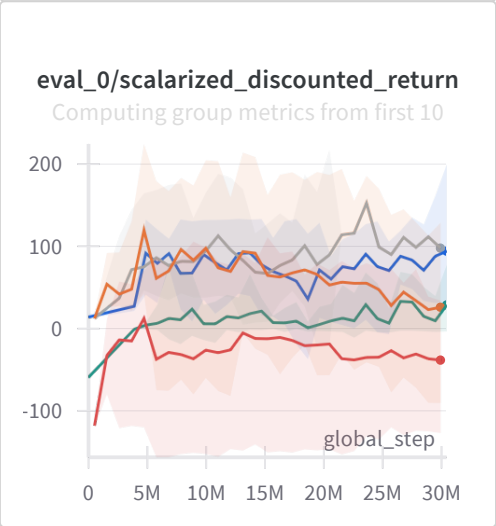

runs.summary["eval/front"]
Value of Objective 2
100
50
0
−50
−20
20
60
100
140
180
Value of Objective 1
mo-hopper-2d-v4__EMOS__9__173…
mo-hopper-2d-v4__EMOS__8__173…
mo-hopper-2d-v4__EMOS__7__173…
mo-hopper-2d-v4__EMOS__4__173…
mo-halfcheetah-v4__EMOS__9__17…
mo-halfcheetah-v4__EMOS__8__17…
mo-halfcheetah-v4__EMOS__4__17…
mo-halfcheetah-v4__EMOS__5__17…
mo-halfcheetah-v4__EMOS__2__17…
mo-hopper-2d-v4__EMOS__9__173…
mo-hopper-2d-v4__EMOS__8__173…
mo-hopper-2d-v4__EMOS__2__173…
mo-halfcheetah-v4__EMOS_NOcom…
mo-halfcheetah-v4__EMOS_NOcom…
mo-halfcheetah-v4__EMOS_NOcom…

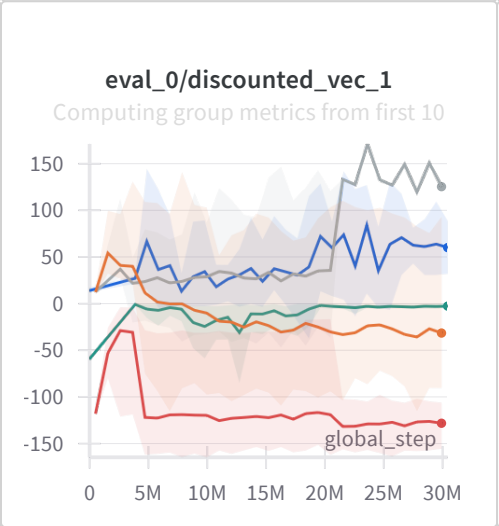

eval/cardinality
25
20
15
10
5
global_step
0
5M
10M
15M
20M

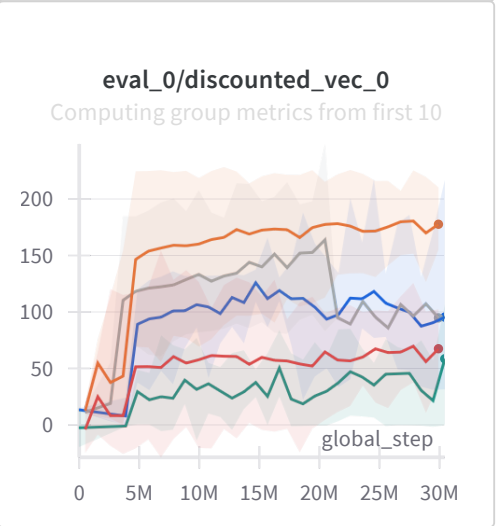

eval_0/vec_1
Computing group metrics from first 10
400
200
0
-200
-400
-600
-800
global_step
0
5M
10M
15M
20M
25M
30M


eval_0/vec_0
Computing group metrics from first 10
1000
500
0
global_step
0
5M
10M
15M
20M
25M
30M

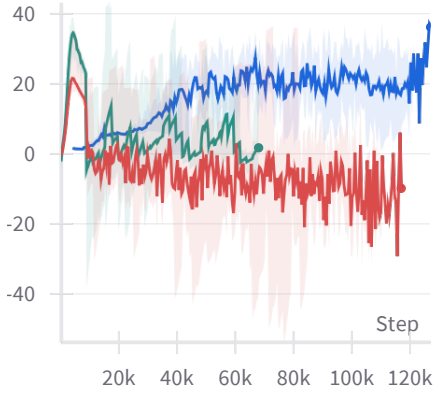

eval_0/scalarized_return
Computing group metrics from first 10
1000
500
0
-500
global_step
0
5M
10M
15M
20M
25M
30M


eval_0/scalarized_discounted_return
Computing group metrics from first 10
200
100
0
-100
global_step
0
5M
10M
15M
20M
25M
30M

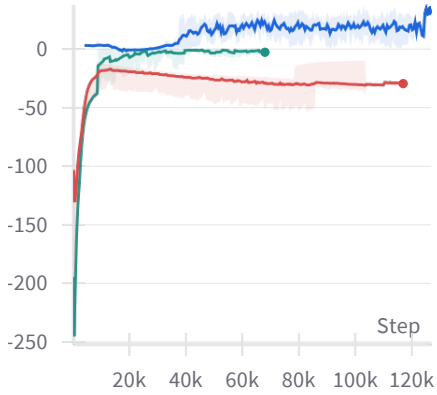

eval_0/discounted_vec_1
Computing group metrics from first 10
150
100
50
0
-50
-100
-150
global_step
0
5M
10M
15M
20M
25M
30M


eval_0/discounted_vec_0
Computing group metrics from first 10
200
150
100
50
0
global_step
0
5M
10M
15M
20M
25M
30M

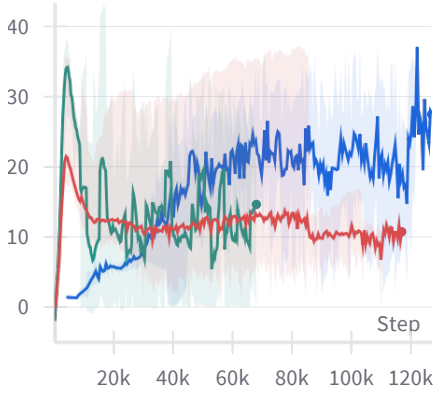
metrics_5/scalarized_episode_return



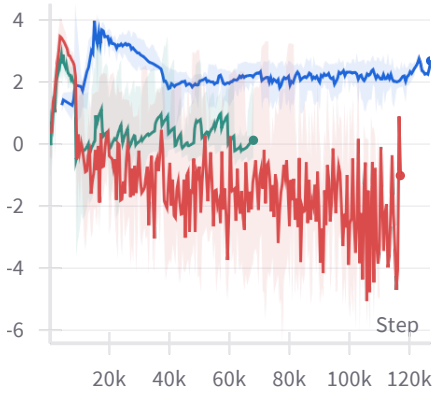
metrics_5/episode_return_obj_1



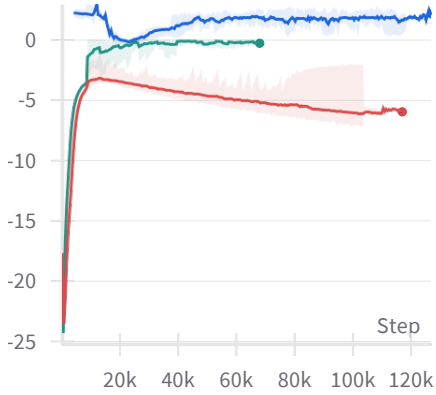
metrics_5/episode_return_obj_0



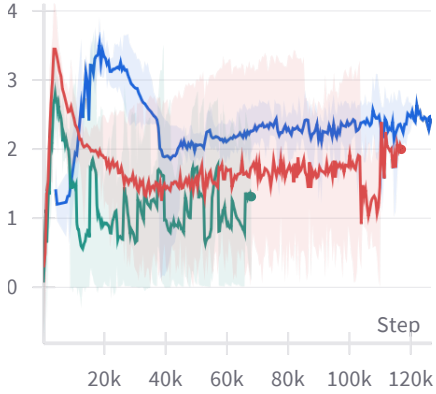
...

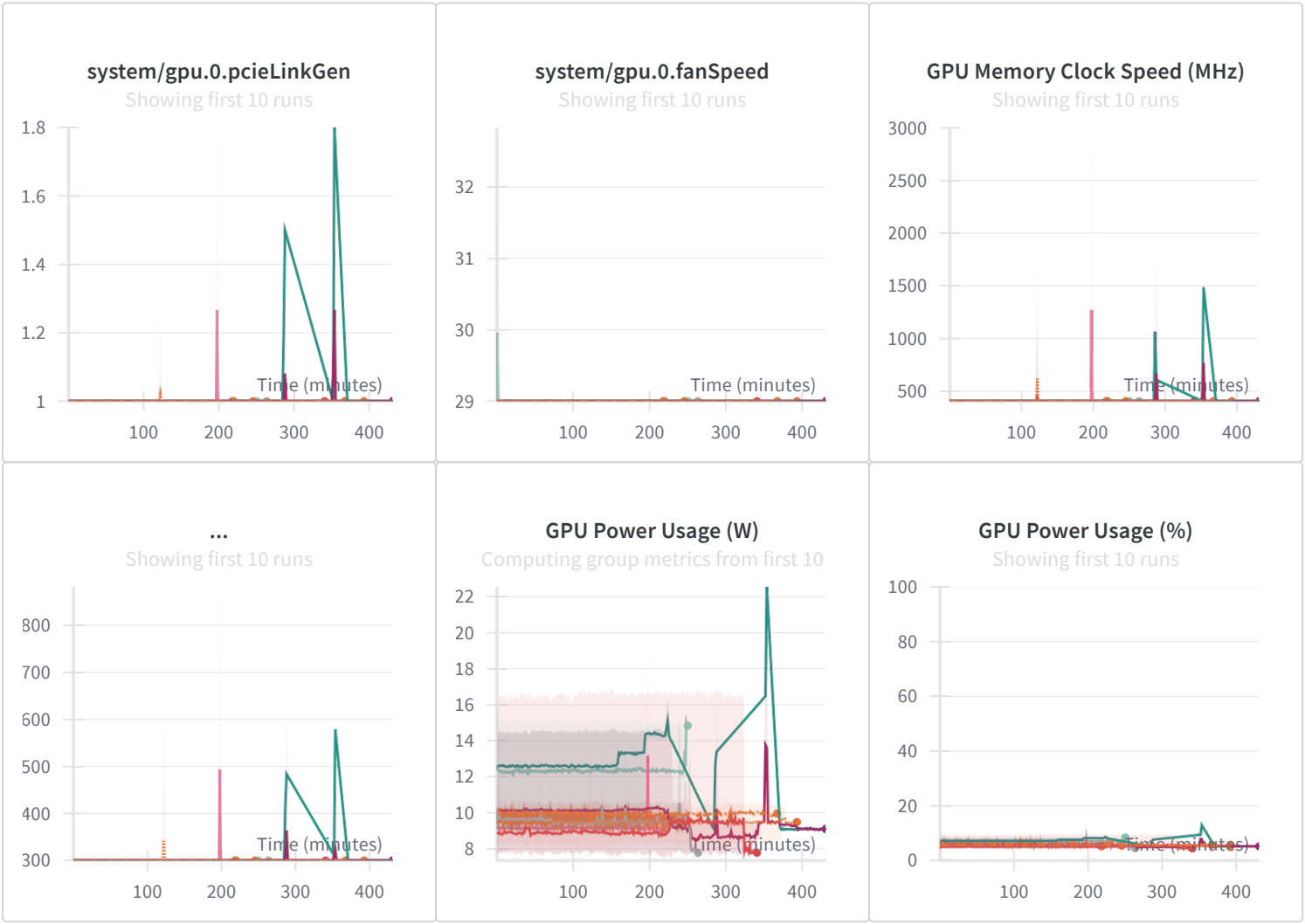


metrics_5/disc_episode_return_obj_1



metrics_5/disc_episode_return_obj_0





Created with ❤️ on Weights & Biases.

https://wandb.ai/clarissa-kuemhof-ludwig-maximilianuniversity-of-munich/MORL-baselines/reports/MORL-using-EAs-for-Diverse-Policy-Selection--VmlldzoxMTI5OTc2MA?accessToken=uff3bzjgz1r0d69z423q33lkj29z7709kll4muu9sewf24fmgqfaitnzub81jv48