**TRAINING PHASE (source2source)**

LR using PPO = 3e-4, LR using TRPO = 1e-3 (IF NOT SPECIFIED)

|  |  |  |
| --- | --- | --- |
| AGENT | TIMESTEPS | RESULT (mean +/- std dev on the reward) |
| PPO | 1K | 75.59 +/- 1.54 |
| PPO | 10K | 74.05 +/- 1.41 |
| PPO | 20K | 221.78 +/- 1.13 |
| PPO | 50K | 366.84 +/- 1.48 |
| PPO | 100K | 575.52 +/- 33.32 |
| TRPO | 1K | 67.73 +/- 1.52 |
| TRPO | 10K | 253.03 +/- 1.57 |
| TRPO | 20K | 200.66 +/- 0.75 |
| TRPO | 50K | 388.70 +/- 218.39 |
| TRPO | 100K | 827.65 +/- 260.86 |
| TRPO (lr=3e-4) | 50K | 289.30 +/- 1.68 |
| TRPO (lr=3e-4) | 100K | 404.18 +/- 3.46 |
| TRPO (lr = 3e-4) | 10K | 290.37 +/- 1.97 |
| PPO (lr = 3e-5) | 100K | 181.55 +/- 2.18 |
| PPO (lr = 1e-3) | 100K | 1369.29 +/- 123.25 |
| PPO | 150K | 1421.12 +/- 31.52 |
| PPO | 200K | 761.03 +/- 144.89 |
| PPO | 175K | 1387.40 +/- 5.54 |
| PPO (lr=3e-4) | 150K | 520.35 +/- 92.73 |
| PPO (lr=1e-3) | 150K | 997.20 +/- 350.13 |
| PPO (lr=1e-4) | 150K | 494.84 +/- 51.98 |
| TRPO | 150K | 1380.45 +/- 36.29 |
| TRPO | 200K | 1519.37 +/- 11.66 |
| TRPO | 300K | 1667.76 +/- 7.55 |
| TRPO | 500K | 1660.24 +/- 2.87 |
|  |  |  |
|  |  |  |

The previous results can be used to fulfill the next table (they are the source-source average return)!

**TABLE FOR STEP 2.3**

LR using PPO = 3e-4, LR using TRPO = 1e-3 (IF NOT SPECIFIED)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ALGORITHM | HYPERPARAMETERS | SOURCE-SOURCE AVERAGE RETURN | SOURCE-TARGET AVERAGE RETURN | TARGET-TARGET AVERAGE RETURN |
| TRPO | TS=300K | 1667.76 +/- 7.55 | 977.76 +/- 16.55 | 1545.72 +/- 139.68 |
| TRPO | TS=500K | 1660.24 +/- 2.87 | 1002.65 +/- 5.69 | 1652.47 +/- 5.33 |
| TRPO | TS=200K | 1519.37 +/- 11.66 | 1440.56 +/- 185.18 | 1465.65 +/- 7.18 |
| PPO | TS=150K | 1421.12 +/- 31.52 | 844.19 +/- 261.51 | 1071.64 +/- 26.01 |
| PPO | TS=175K | 1387.40 +/- 5.54 | 1316.47 +/- 10.34 | 715.40 +/- 37.81 |
| TRPO | TS=200K, LR=3E-4 | 1276.75 +/- 153.49 | 889.60 +/- 165.36 | 873.66 +/- 92.97 |
| TRPO | TS=300K, LR=3E-4 | 1160.93 +/- 243.49 | 1463.48 +/- 38.29 | 769.00 +/- 210.86 |
| TRPO | TS=500K, LR=3E-4 | 1626.47 +/- 9.49 | 1095.82 +/- 105.24 | 1541.97 +/- 5.37 |

**TABLE FOR STEP 2.3**

LR using PPO = 3e-4, LR using TRPO = 1e-3 (IF NOT SPECIFIED)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ALGORITHM | HYPERPARAMETERS | UDR distribution | SOURCE-SOURCE AVERAGE RETURN | SOURCE-TARGET AVERAGE RETURN |
| TRPO | TS=200K | Uniform  [min(original\_masses), max(original\_masses)] | 1468.80 +/- 2.62 | 1610.92 +/- 15.71 |
| TRPO | TS=300K, LR=3E-4 | Uniform  [min(original\_masses), max(original\_masses)] | 1299.02 +/- 242.96 | 795.21 +/- 63.22 |
| PPO | TS=175K | Uniform  [min(original\_masses), max(original\_masses)] | 1361.61 +/- 44.49 | 990.87 +/- 101.86 |
|  |  |  |  |  |
|  |  |  |  |  |