

TAAC: Temporally Abstract Actor-Critic for Continuous Control

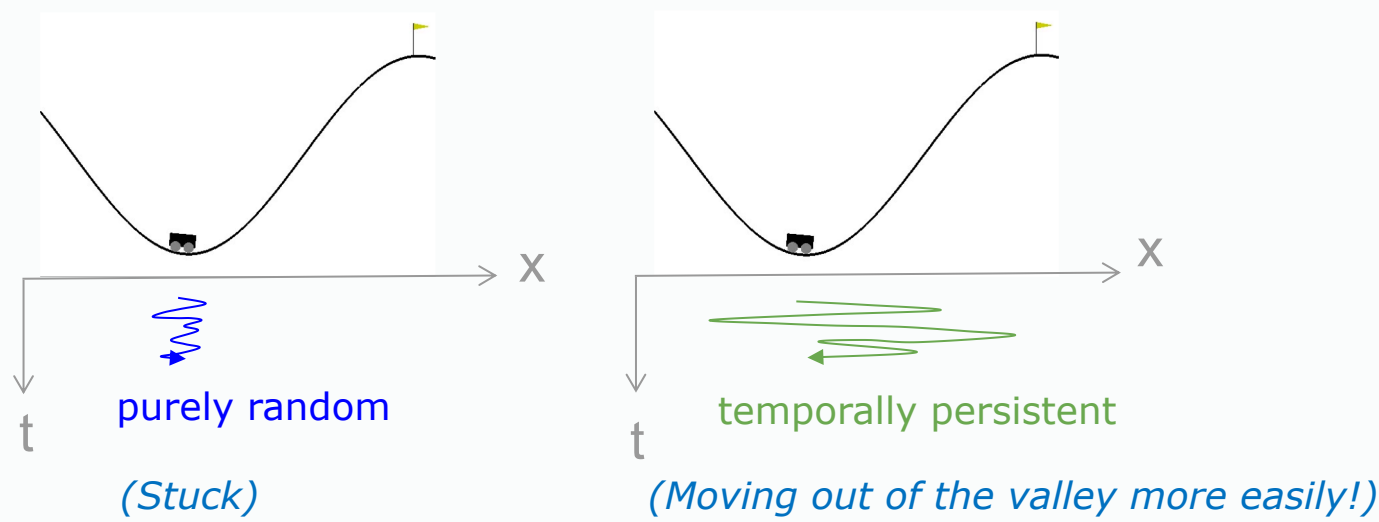
Haonan Yu, Wei Xu, Haichao Zhang

(Try code at <https://github.com/hnyu/taac>)

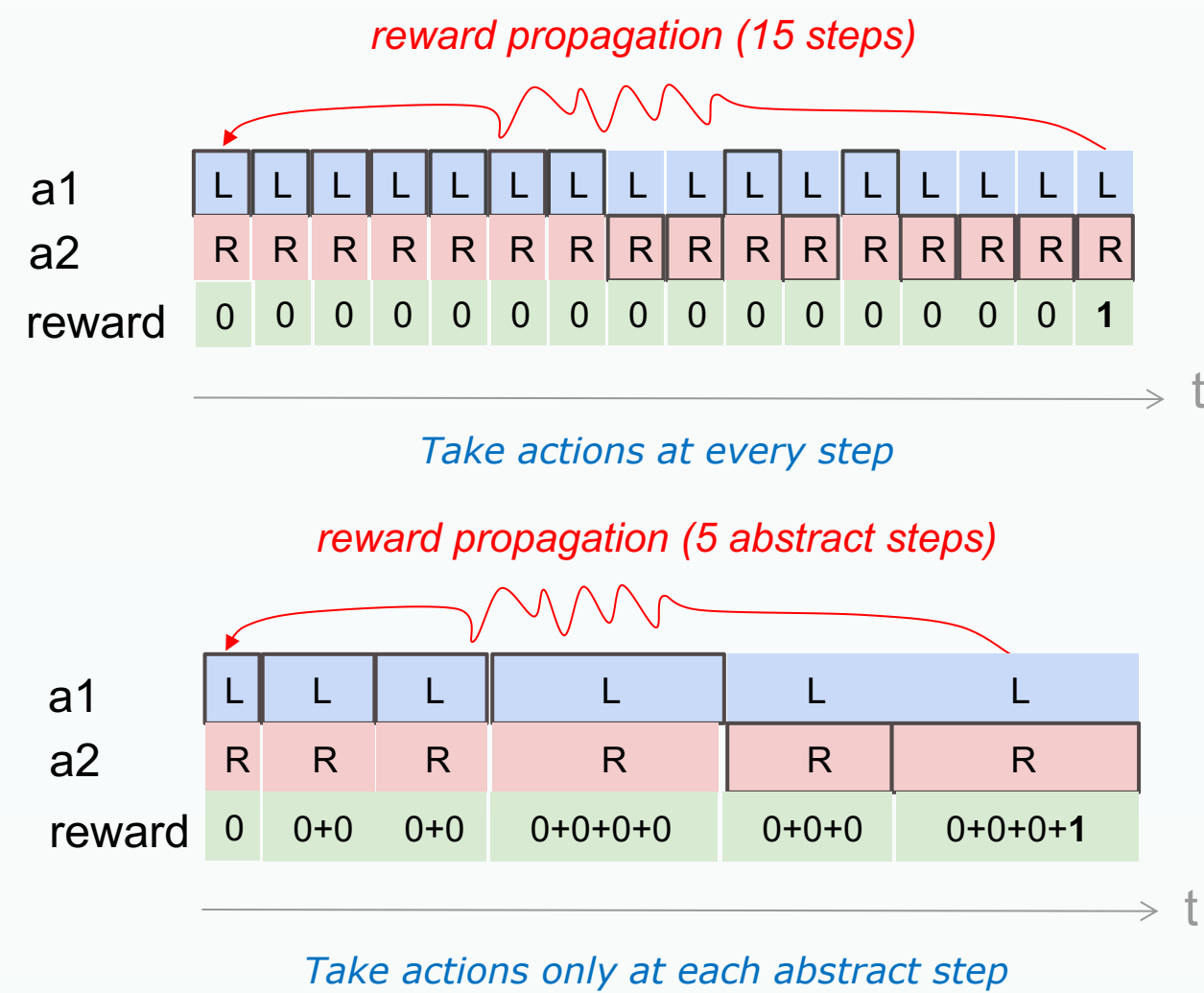


Motivation: temporal abstraction

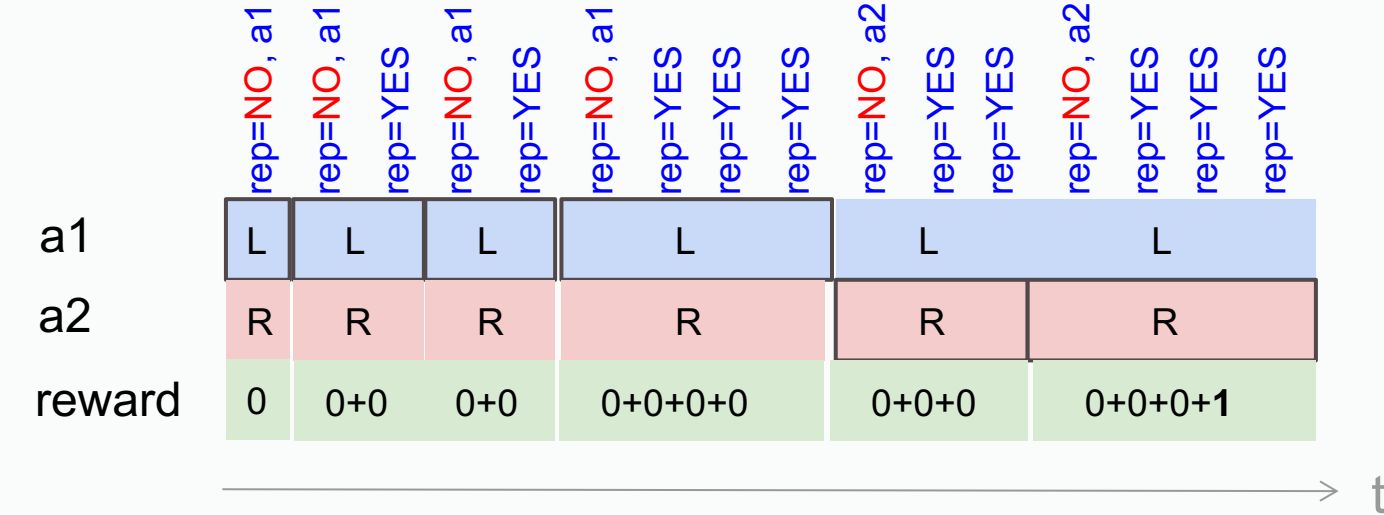
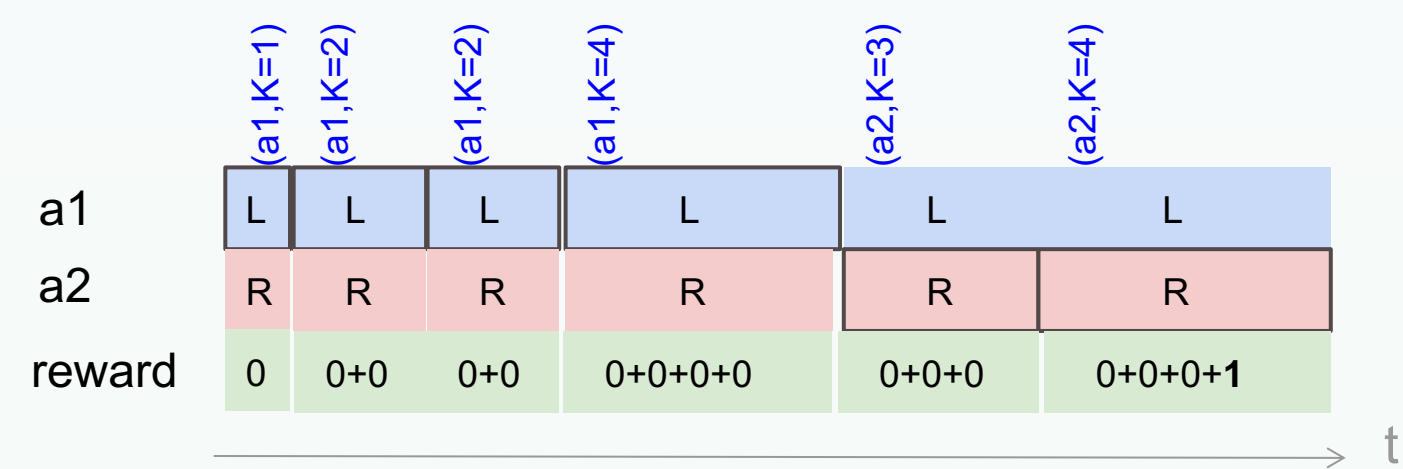
1) Persistent exploration



2) Easier credit assignment

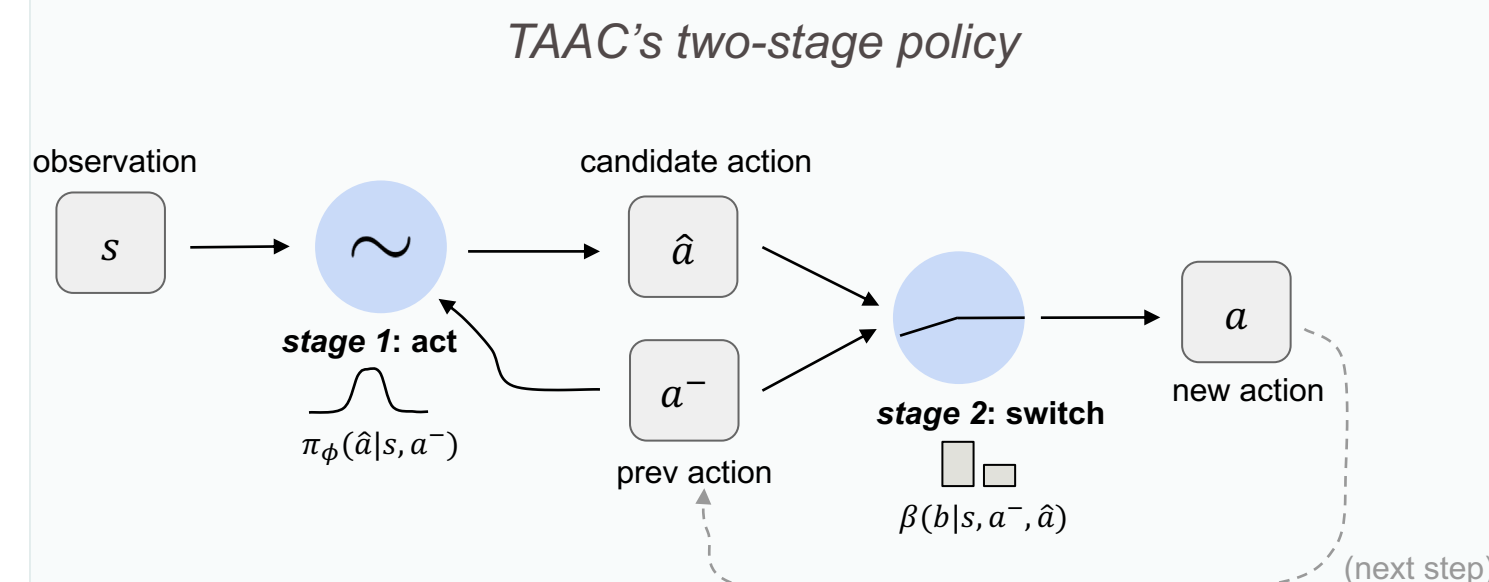


Open-loop vs. closed-loop action repetition

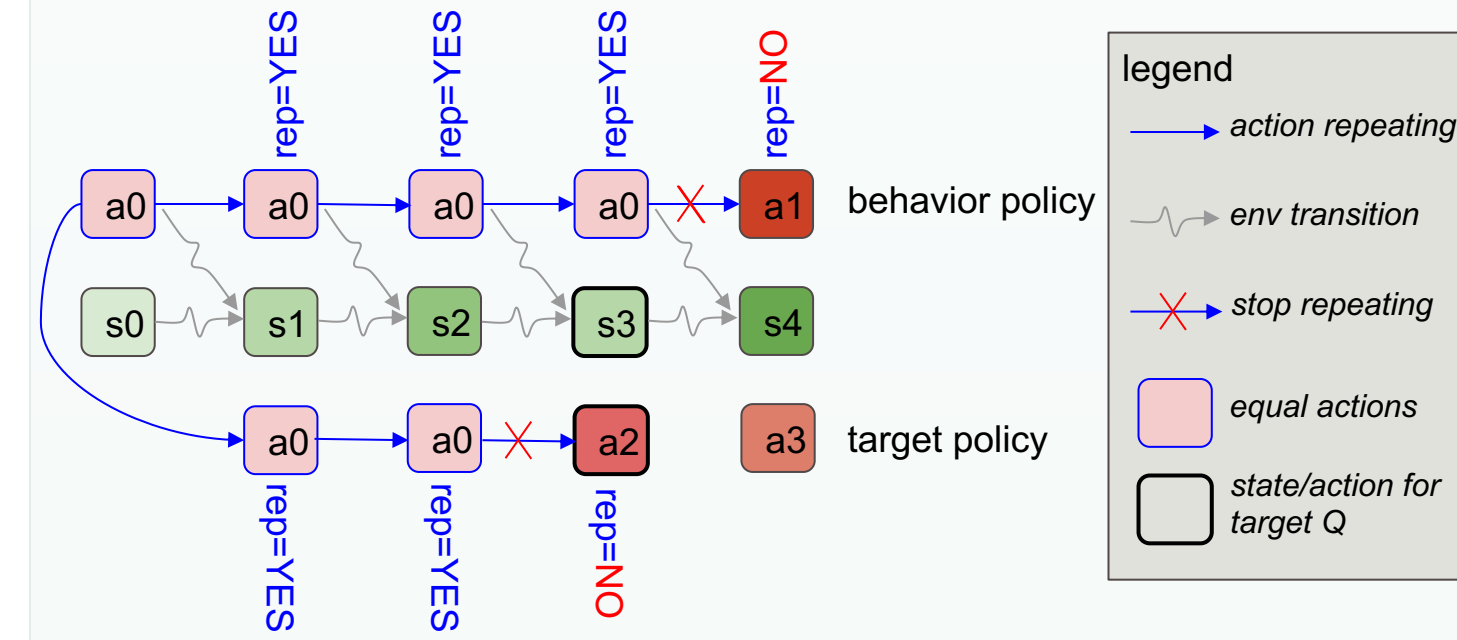


Closed-loop methods decide "act-or-repeat" at every step
[Neunert et al., 2020] [Chen et al., 2021]

TAAC incorporates closed-loop action repetition into actor-critic



Policy evaluation: a novel compare-through operator for multi-step TD backup



Example: We use $Q(s_3, a_2)$ as the target to bootstrap $Q(s_0, a_0)$ (3-step TD)

Policy improvement: similar to SAC [Haarnoja et al., 2018] with a simple approximation

$$\beta(0)^* = \exp\left(\frac{Q(s, a^-)}{\alpha}\right) / Z(s) \quad \beta(1)^* = \exp\left(\frac{Q(s, \hat{a})}{\alpha}\right) / Z(s)$$

Decide "act-or-repeat" in proportional to the exponential Q values

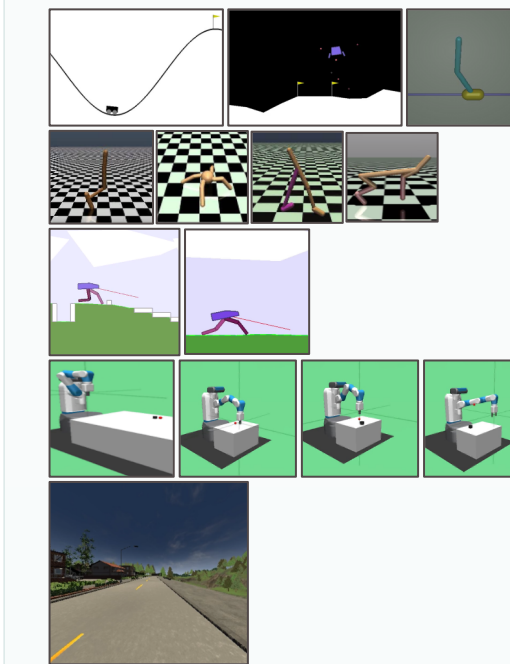
$$\pi_\phi(\hat{a}|s, a^-) \leftarrow \frac{\partial Q(s, \hat{a})}{\partial \phi} \beta^*(1)$$

Extra scaling

"dQda" as in DDPG [Lillicrap et al., 2016] and SAC [Haarnoja et al., 2018]

Experiments

Tasks



Comparison methods

SAC [Haarnoja et al., 2018]: flat RL

SAC-Ntd: SAC + Retrace [Munos et al., 2016] + N-TD

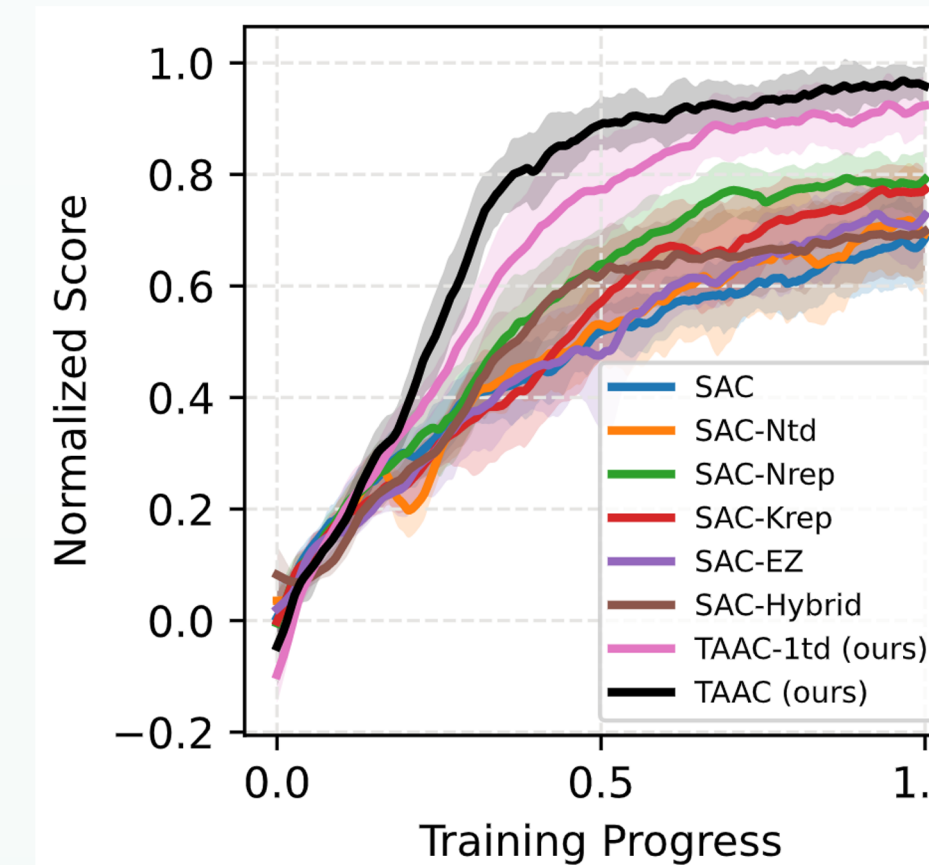
SAC-Nrep: SAC + Fixed action repetition

SAC-Krep: open-loop action repetition [Sharma et al., 2017; Biedenkapp et al., 2021]

SAC-EZ: SAC with EZ-greedy [Dabney et al., 2021]

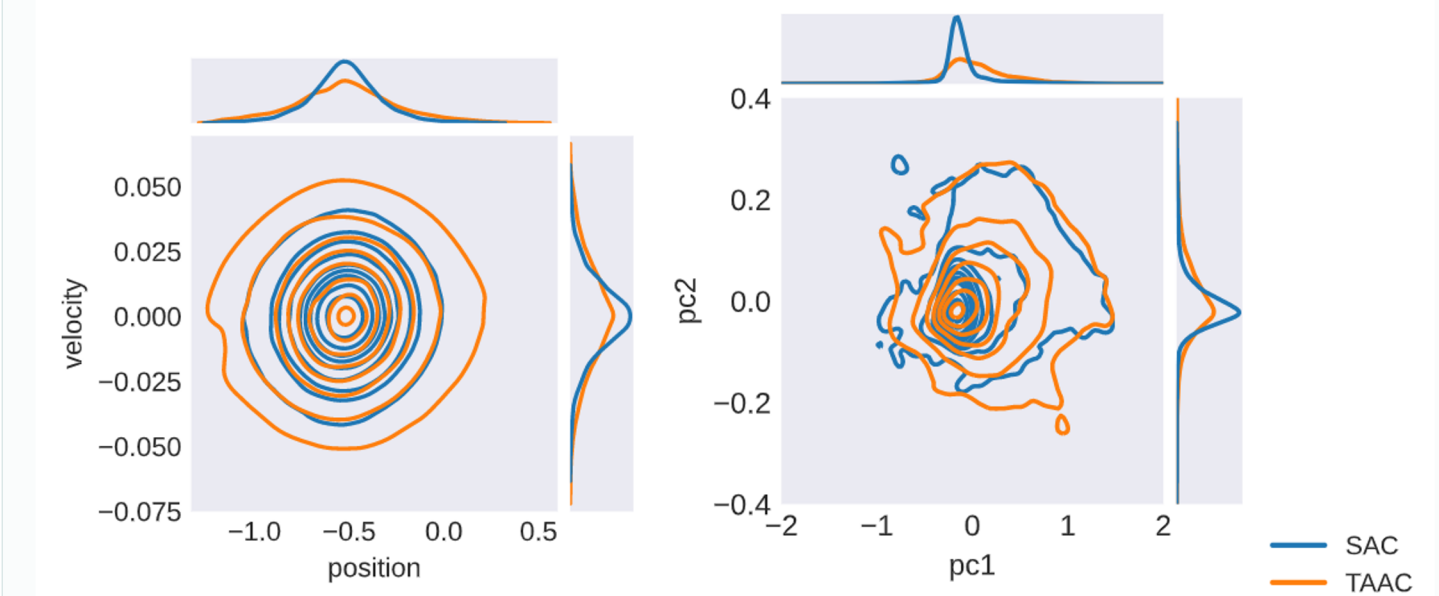
SAC-Hybrid: closed-loop action repetition formulation from H-MPO [Neunert et al., 2020] with SAC backbone

TAAC-1td: TAAC without the compare-through operator



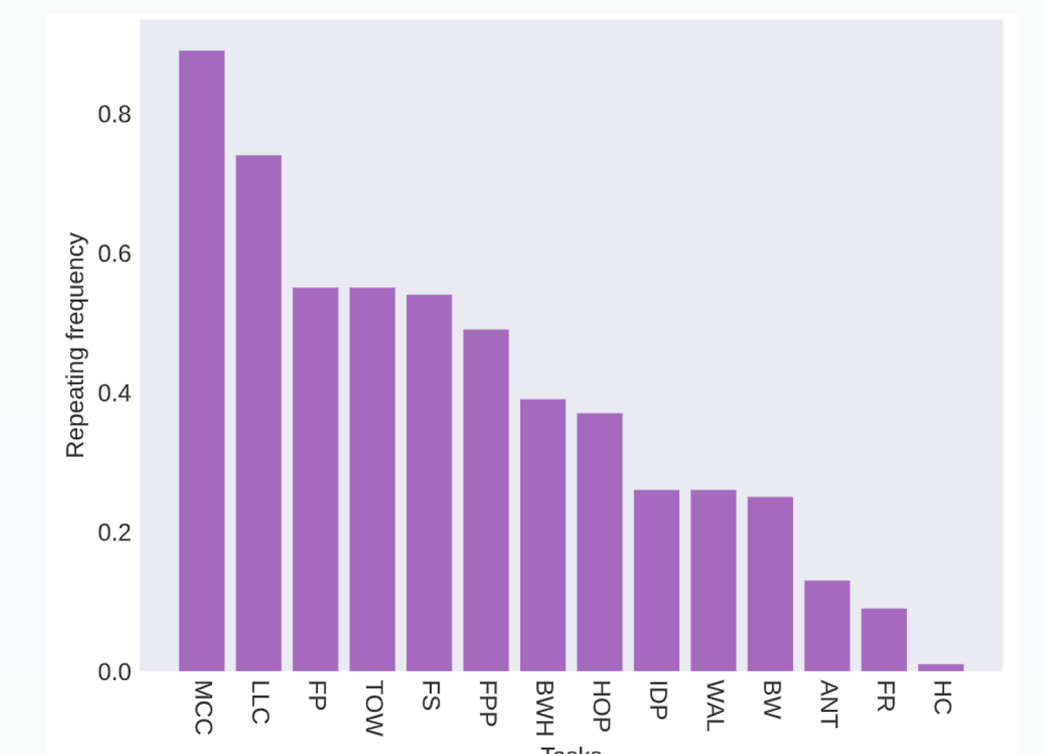
(Normalized and averaged over 14 tasks)

Exploration behavior visualization

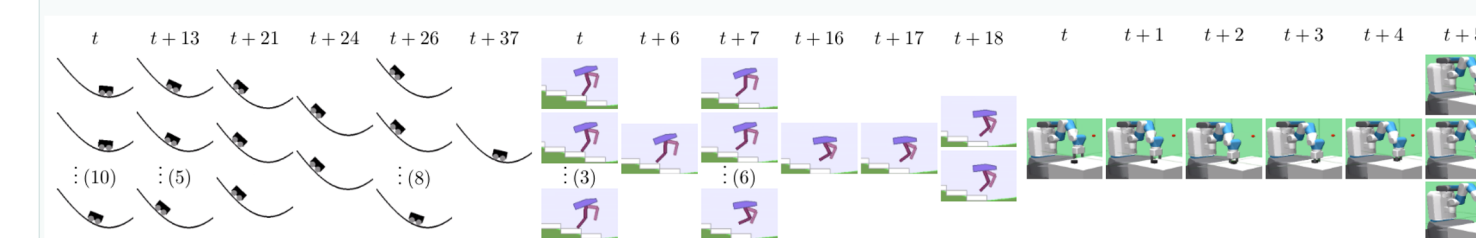


Random policies behaviors

Action repeating frequencies



Action repeating patterns



Key observation: TAAC learns to skip learning to generate new actions at non-critical states, and save the actor network's representational power for critical states!