UMassAmherst

Manning College of Information & Computer Sciences

Efficient Control in Real Time Reinforcement Learning Using Temporally Layered Architecture

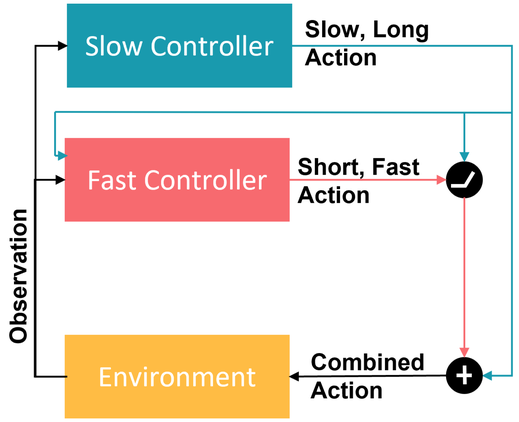
Devdhar Patel[1], Francesca Walsh[1], Joshua Russell[1], Zhongyang Zhang[1], Tauhidur Rahman[1], Terrence Sejnowski[2], Hava Siegelmann[1]
[1]University of Massachusetts Amherst, Amherst MA
[2] Salk Institute for Biological Studies, La Jolla, CA

Introduction

- A fast constant response time for an RL agent is energy intensive and makes the learning difficult.
- We developed a temporally layered architecture where each layer has a different response time allowing the RL agent to adapt its response time.
- The fast action is gated based on its effect on the final action, reducing the action frequency and improving efficiency.
- The fast layer is trained after the slow layer.
- To measure efficiency, we introduce a new metric: Return Per Action (RPA).
- We demonstrate performance on the MuJoCo continuous control environments.

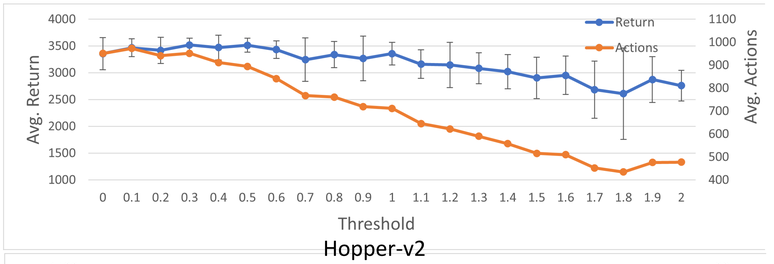
Results Overview

- TLA outperforms the constant frequency networks on all environments tested while demonstrating efficient performance.
- Learning curves suggest more stable learning with a lower standard deviation.
- Gating the fast actions improves the efficiency and performance.
- State-action values suggest that the fast controller discriminates between states that seem similar to the slow controller.

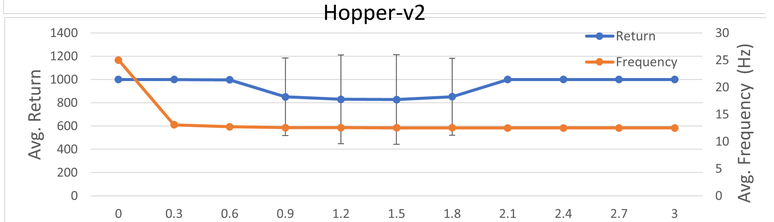


Temporally Layered Architecture (TLA)

Environment	TLA			TD3		
	Response Time	Avg. Return	RPA	Response Time	Avg. Return	RPA
InvertedPendulum-v2	0.04s, 0.08s	**1000 ± 0**	**1.9**	0.04s	1000 ± 0	1.00
InvertedDoublePendulum-v2	0.05s, 0.1s	**9358.94 ± 0.82**	**18.49**	0.05s	9358.48 ± 2.5	9.36
Hopper-v2	0.008s, 0.016s	**3443.21 ± 131.6**	**3.74**	0.008s	3032.25 ± 262.8	3.20
Walker2d-v2	0.008s, 0.016s	**3694.04 ± 128.58**	**3.96**	0.008s	3233.77 ± 895.3	3.23

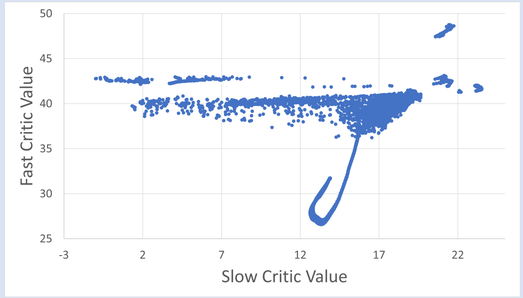


Hopper-v2

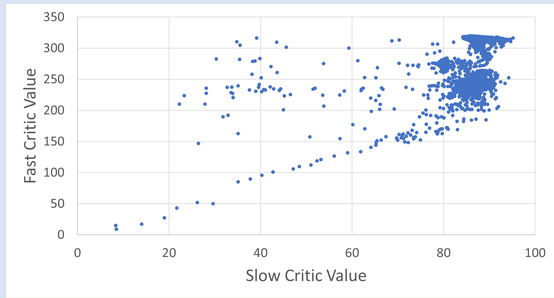


InvertedPendulum-v2

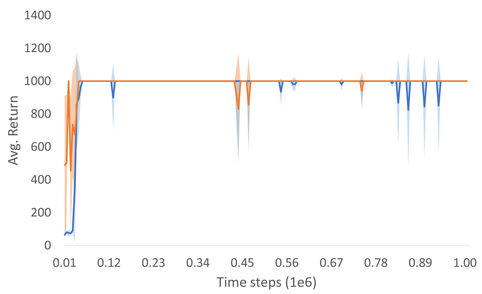
Critic Value Analysis



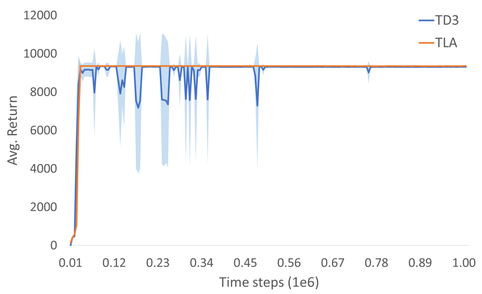
InvertedPendulum-v2



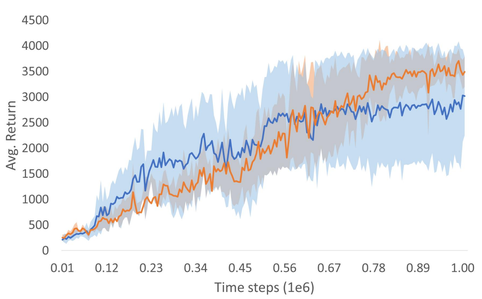
InvertedDoublePendulum-v2



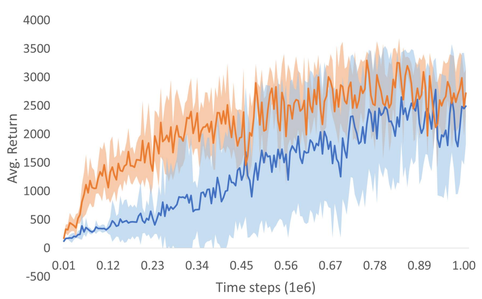
InvertedPendulum-v2



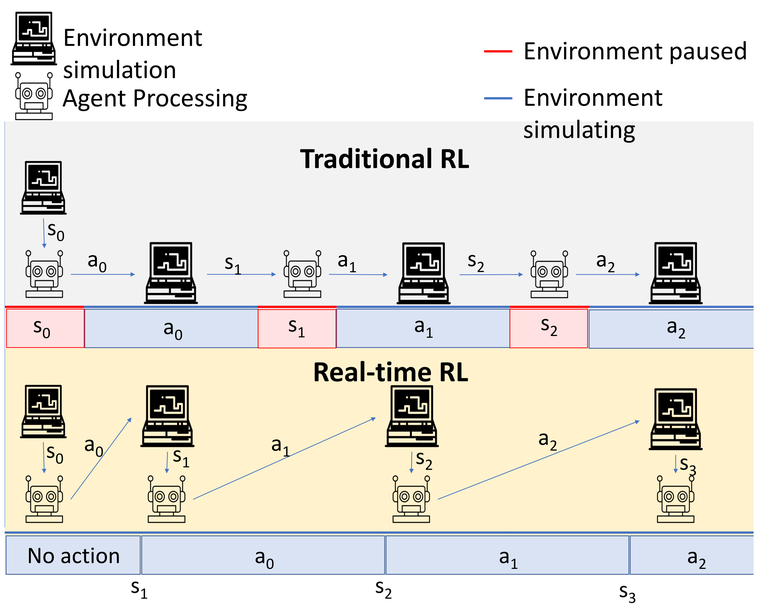
InvertedDoublePendulum-v2



Walker2d-v2

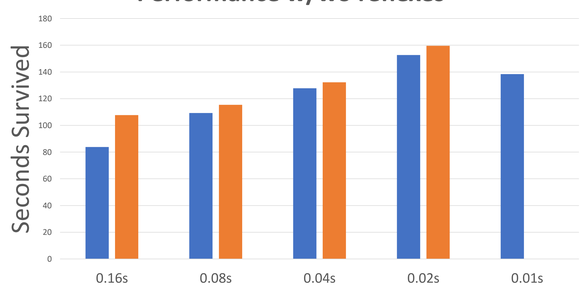


Hopper-v2



Manually designed Reflexes

Performance w/wo reflexes



	Seconds survived (Decisions made)				
Res. speed	0.16s	0.08s	0.04	0.02	0.01
w/o Reflex	83.81 (534)	109.21 (1366)	127.8 (3195)	152.68 (7634)	138.4 (13840)
Reflex	107.72 (674)	129.14 (1618)	132.23 (3306)	159.63 (7982)	