A Framework for Predictable Actor-Critic Control

Affiliation: Texas A&M University
Authors: Josiah Coad, James Ault, Jeff Hykin, Guni Sharon

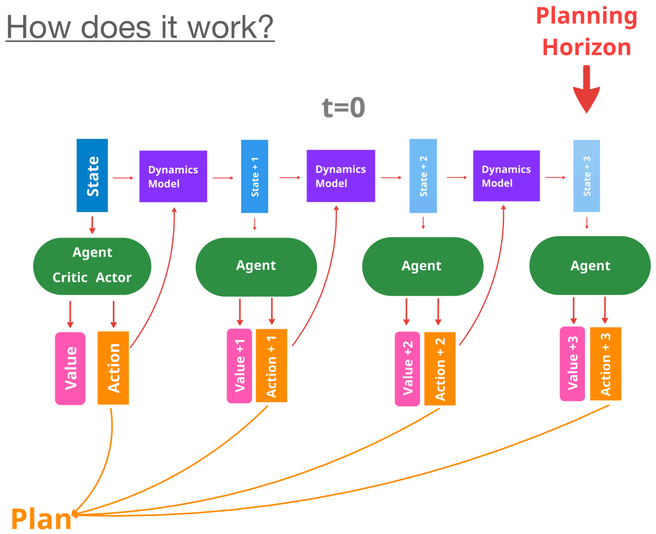
What problem is being addressed?

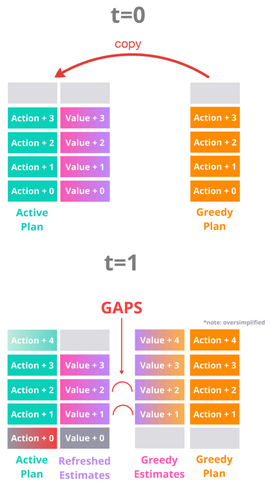
Agents that rigidly adhere to a plan usually encounter significant decrease in total reward. Despite this, following a plan is often desirable for both theoretical and practical applications. Can we do better than strict adherence?

How might this work be extended?

- The dynamics model is designed to be replaced with a more advanced model. We expect this would significantly increase plan length.
- The utility of non-rigid plan commitment could have applications in multi-agent environments. Performing a multi-agent baseline could make for future work.

How does it work?





What does this paper contribute?

Non-rigid plan commitment with theoretical bounds and empirical support

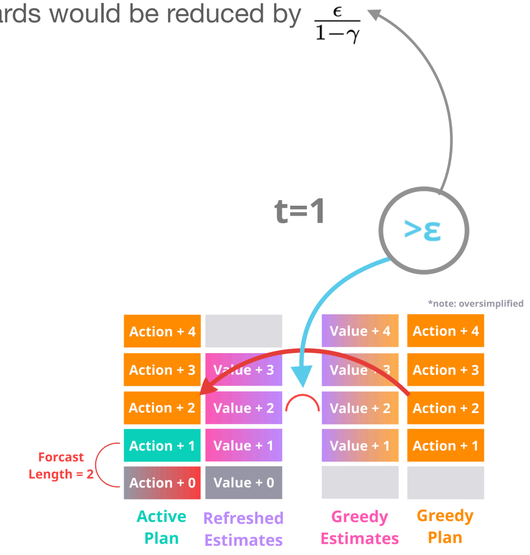
Guarantees

Assumptions

- Accurate Q values for a policy
- A discount factor $0 < \gamma < 1$

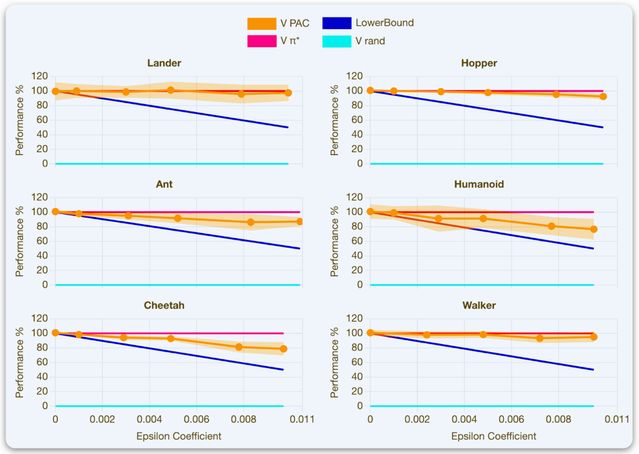
Bound

In the worst case, the sum of discounted rewards would be reduced by $\frac{\epsilon}{1-\gamma}$



the future-most parts of the active plan are most likely to be revised. This happens any time the gap is more than ε

What are the empirical results?



Performance stays well above the theoretical bound

Forecast length increases as epsilon is relaxed