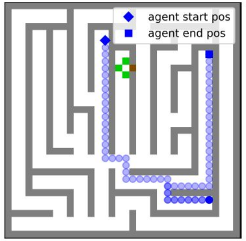
Towards Jumpy Planning

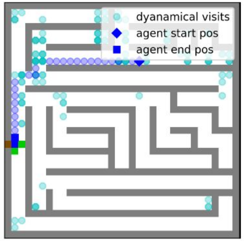
Akilesh B, **Suriya Singh**, Anirudh Goyal, Alexander Neitz and Aaron Courville.

Overview

- Model-free RL: high sample inefficiency and ignorance of the environment dynamics.
- Model-based RL at the scale of time-steps: compounding errors and high computational requirements.

- Hierarchical Reinforcement Learning framework [1,2] address limitations in classic RL through sub-tasks and abstract actions.

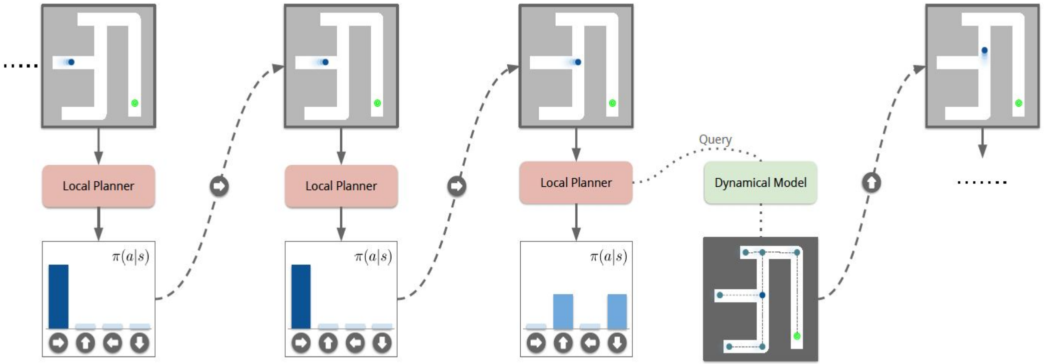




<u>This work</u>

Use a model-based planner together with a goal-conditioned policy trained with model-free learning. We use a model-based planner that operates at higher levels of abstraction i.e., *decision states* and use model-free RL between the decision states.

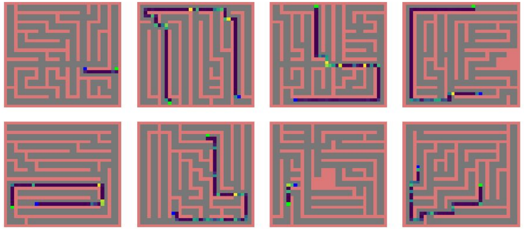
Jumpy Planning



Decision States [3] (*aka* subgoal)
states where the agent's policy has high entropy.

$$-\sum_{a\in\mathcal{A}} \pi(a|s) \ln \pi(a|s) > \tau$$

We fix τ such that a tiny fraction of states are chosen as decision states.



Dynamical Models $\mathrm{M}\text{: } (\mathrm{s},\ \mathrm{a}) \rightarrow s'$

-> argmax or sample, successively query M in BFS fashion until goal state is encountered or maximum search depth is reached.

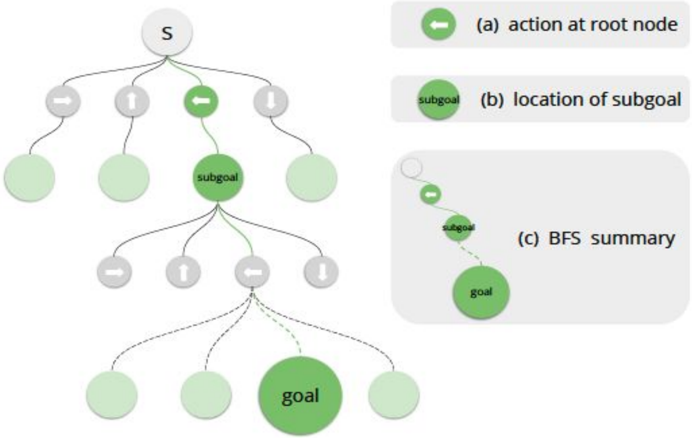
Jumpy dataset

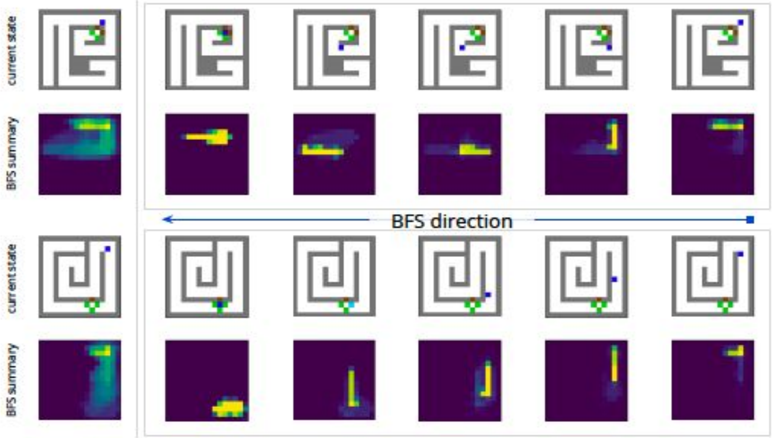
$\sum_{a\in\mathcal{A}} \pi(a|s') \ln \pi(a|s') > \tau$ or $\Delta_{min}T \leq dist(s,s') \leq \Delta_{max}T$

Funnel all intermediate states leading to the same s'

Jumpy Planning with Dynamical Model

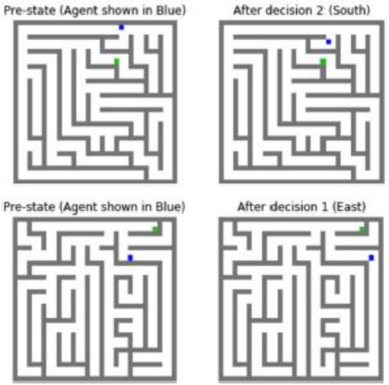
- The result of query is further passed to the agent to take action at current decision state





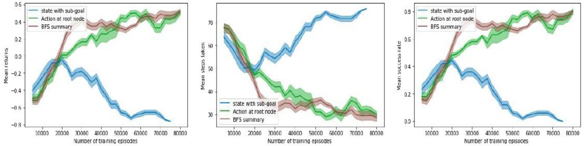
Results

Jumpy dataset

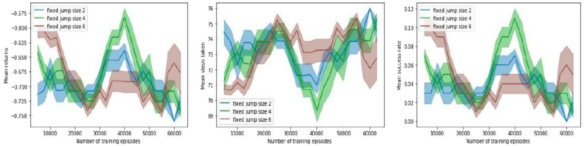


Performance Comparison

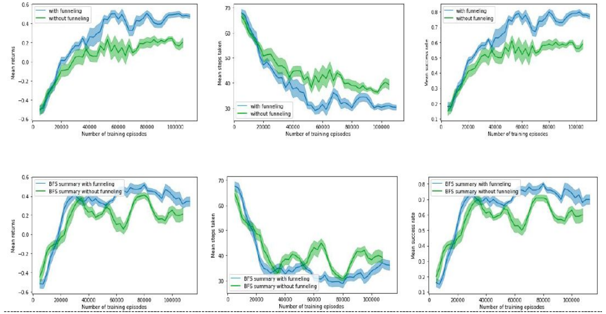
Dynamical models



Fixed time step models

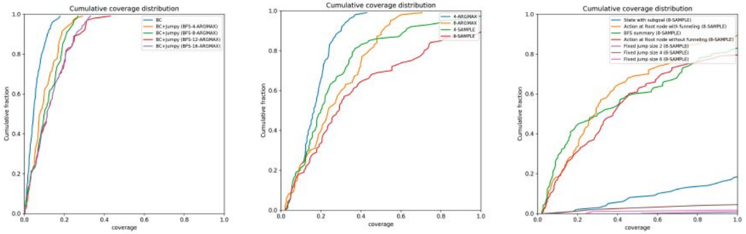


Funneling



Coverage

$$\frac{\#\text{unique states visited}}{\#\text{reachable states}}$$



Generalisation Performance