

Switching Successor Measures for Hierarchical Zero-Shot Reinforcement Learning

Stefan Stojanovic, Alexandre Proutiere
KTH Royal Institute of Technology, Sweden
stesto@kth.se

MOTIVATION



Most hierarchical RL methods only work for goal-reaching tasks.



Hierarchy is imposed through fixed temporal abstractions. (ex. fixed subgoal horizons)



Can successor representations induce hierarchical policies?

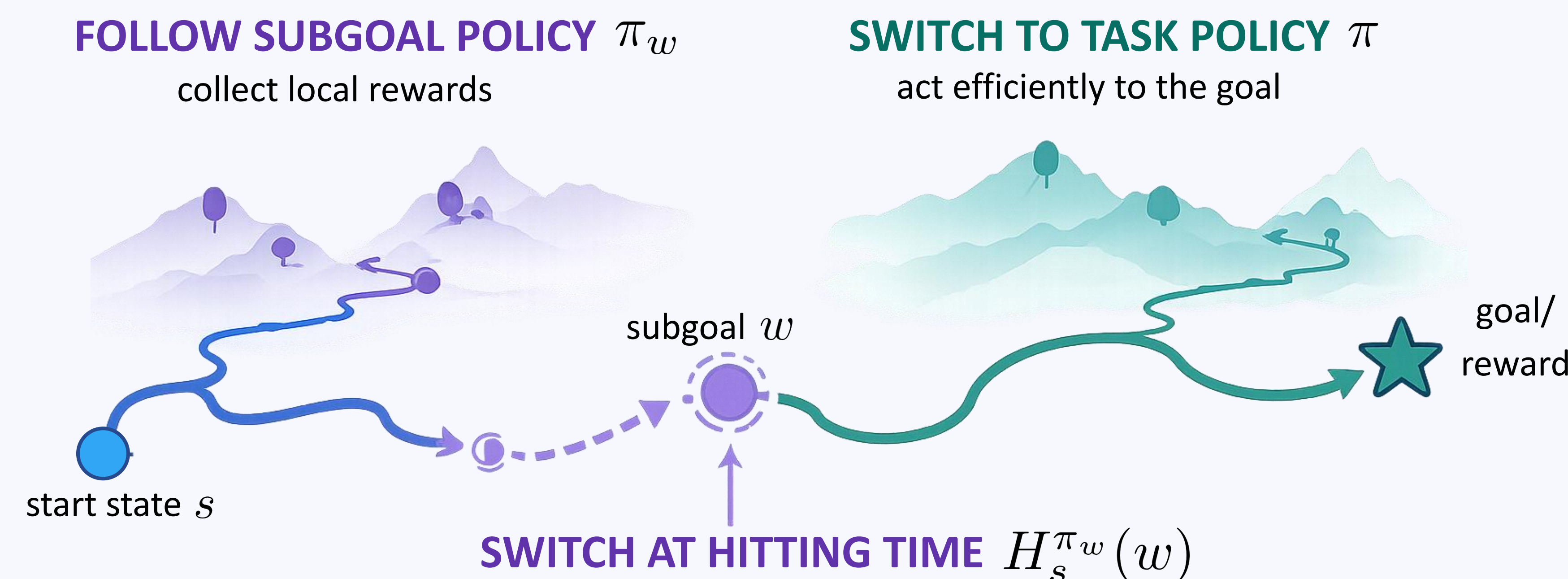
Our answer: YES.

CONTRIBUTIONS

- ✓ First hierarchical framework built directly from successor measures.
- ✓ Works for general reward functions, not only goal-reaching.
- ✓ No fixed planning horizon: switching happens at hitting time.
- ✓ Hierarchical policies emerge from a single learned representation.
- ✓ Competitive with HIQL and stronger than non-hierarchical baselines.

THE KEY IDEA

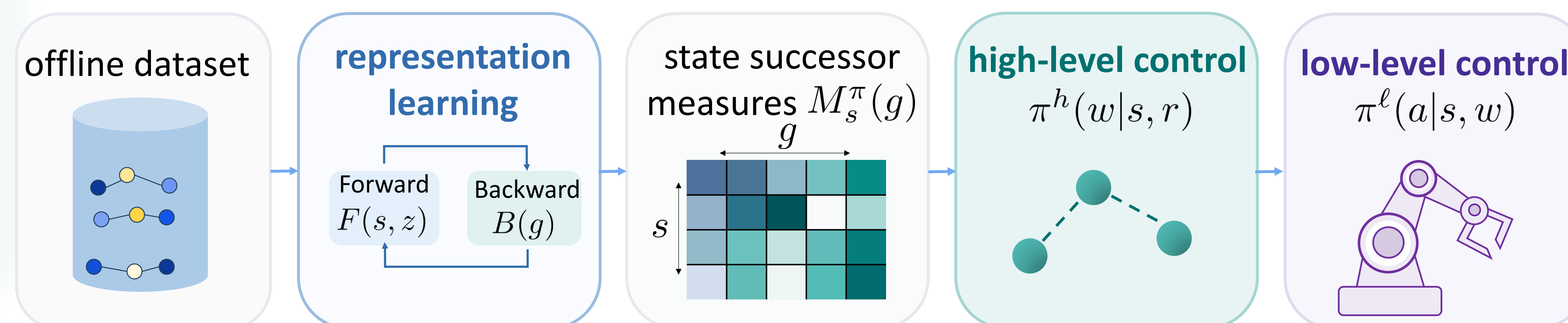
We introduce **Switching Successor Measures** for agents that follow a subgoal policy until the subgoal is reached, then switch back to the task policy.



$$M_s^{\pi_w \rightarrow \pi}(s') = \underbrace{M_s^{\pi_w}(s')}_{\text{subgoal policy term}} + \underbrace{\frac{M_s^{\pi_w}(w)}{M_w^{\pi_w}(w)}}_{\text{switches at hitting time}} \left(\underbrace{M_w^{\pi}(s')}_{\text{post-switching state occupancy}} - \underbrace{M_w^{\pi_w}(s')}_{\text{pre-switch correction term}} \right)$$

- ✓ captures local rewards
- ✓ switches at hitting time
- ✓ supports all rewards

FB π-SWITCH: ALGORITHM PIPELINE



RESULTS

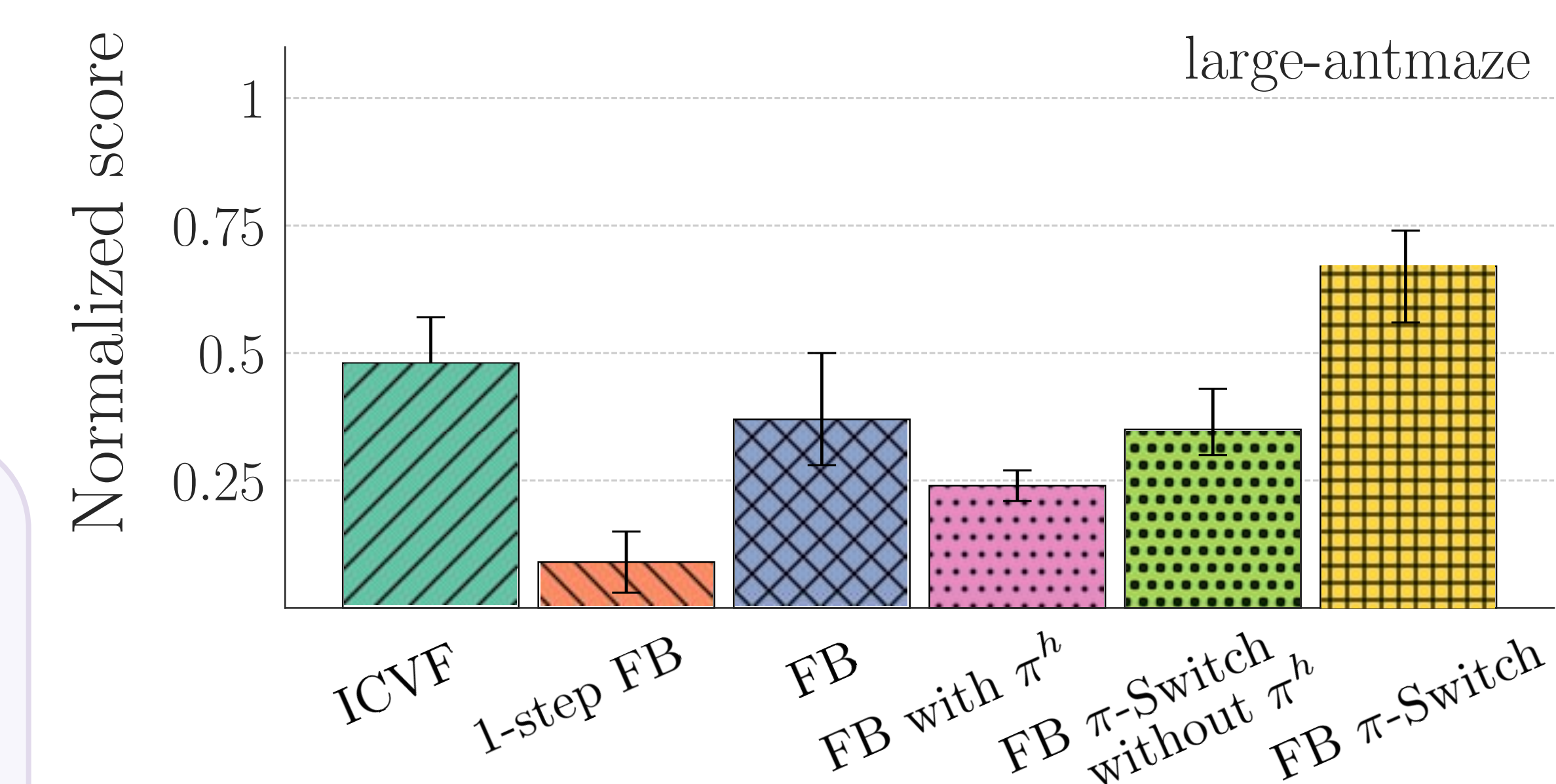
GOAL-CONDITIONED RL

Antmaze	HIQL*	ICVF	FB	FB π-Switch
medium	93 ± 1	46 ± 37	47 ± 1	87 ± 6
large	73 ± 6	28 ± 3	21 ± 5	66 ± 9
teleport	42 ± 4	23 ± 6	25 ± 7	40 ± 6

- ✓ comparable to the state-of-the-art (HIQL)
- ✓ consistently better than non-hierarchical FB baseline

GENERAL REWARD FUNCTIONS

Evaluated on novel tasks requiring a combination of sparse goal-reaching and dense region rewards.



- ✓ FB π -Switch achieves the highest IQM return across diverse reward landscapes



TAKE-AWAY

Hierarchies are already encoded within successor representations.

Switching successor measures reveal this structure, enabling hierarchical zero-shot RL without manually designed temporal abstractions.

Project
page

