Modulated Policy Hierarchy

Alexander Pashevich*, Danijar Hafner,
James Davidson, Rahul Sukthankar, Cordelia Schmid

1 Overview

Hierarchical reinforcement learning (HRL) is an intuitive approach to address long-horizon problems with sparse rewards.

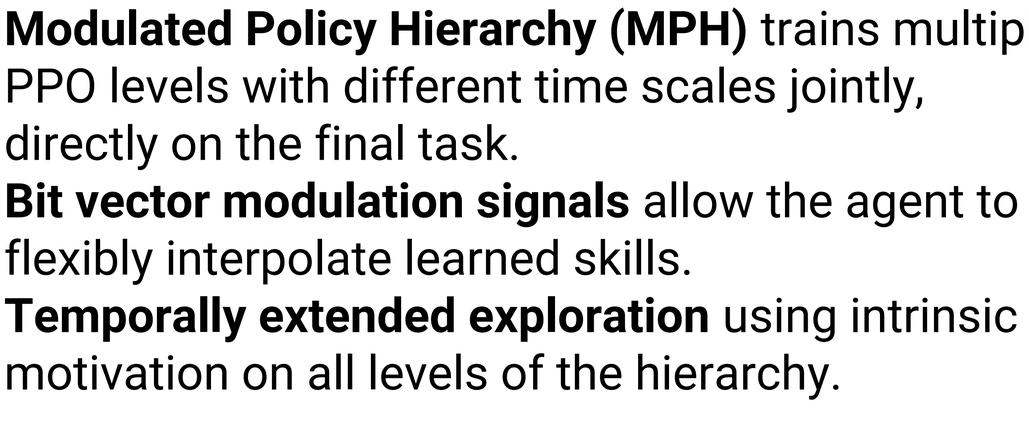
Previous HRL methods often require manual reward shaping (Riedmiller 2018), alternating training phases (Frans et al., 2018), or manually defined subtasks (Florensa et al., 2017).

Our goal is to solve tasks with sparse rewards using a hierarchy end-to-end.

Solution:

- **Modulated Policy Hierarchy (MPH)** trains multiple PPO levels with different time scales jointly, directly on the final task.
- **Bit vector modulation signals** allow the agent to flexibly interpolate learned skills.
- **Temporally extended exploration** using intrinsic motivation on all levels of the hierarchy.

2 Modulation Signals



a) Options b) One-hot modulation c) MPH (ours)

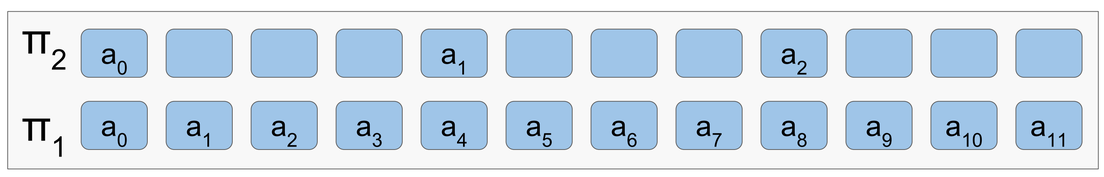
(a) The options agent selects between separate skill networks using a categorical master policy.

(b) The one-hot agent combines the skills into a single network modulated by a 1-hot signal.

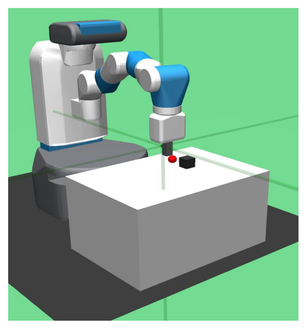
(c) MPH sends a binary vector, allowing for richer communication and mixing of skills.

3 Temporal Abstraction

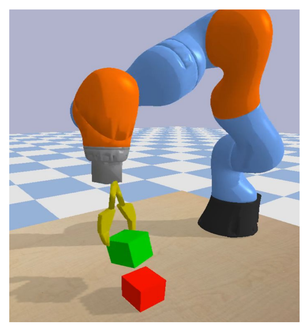
Each level in the hierarchy attends to different time-scales, higher levels activate less frequently. When a level doesn't tick it ignores the input and its modulation output stays constant.



4 Sparse Reward Tasks



FetchPush-v0



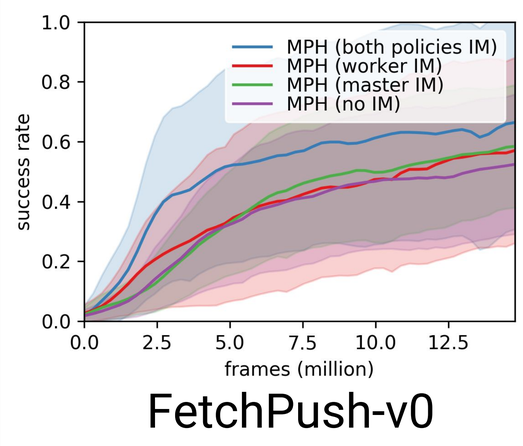
Block Stacking

5 Hierarchical Intrinsic Motivation

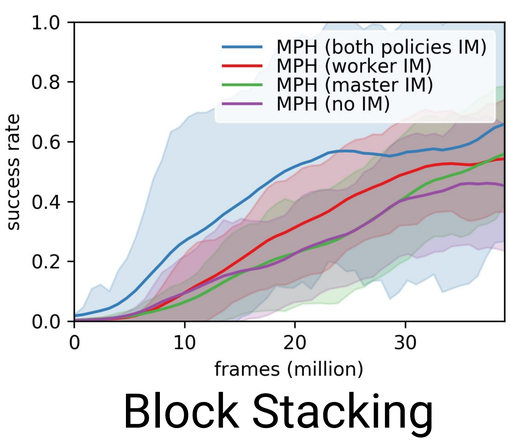
Use forward prediction error as curiosity bonus. Independently for each level of the hierarchy and on the corresponding time-scale:

$$R_t^k = R_t^{\text{env}} + \left\| \hat{\phi}_k^F(s_{t+1}^k) - \phi_k(s_{t+1}^k) \right\|_2$$

Intrinsic motivation on both levels improves exploration both for long and short term effects:

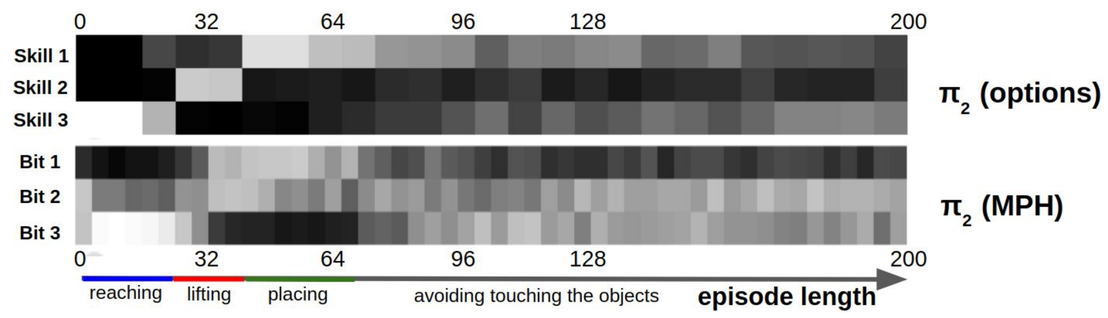


FetchPush-v0



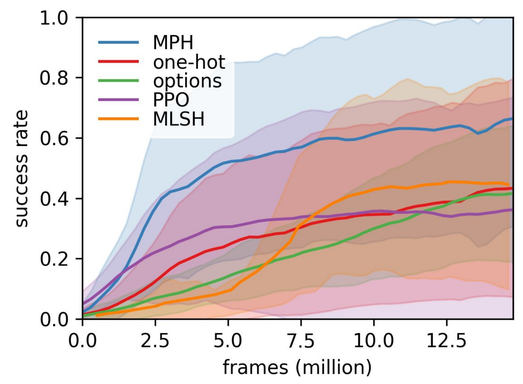
Block Stacking

6 Learned Modulation

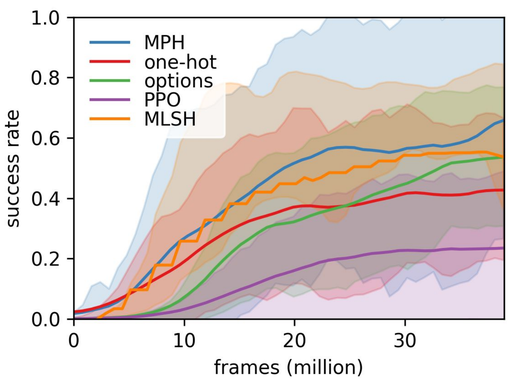


7 Learned Modulation

MPH outperforms options, one-hot, PPO, and MLSH:



FetchPush-v0



Block Stacking

* Could not attend conference because the Canadian government did not process the visa in time.
Contact: alexander.pashevich@inria.fr