Previous Knowledge Utilization In Online Anytime Belief Space Planning

IEEE Robotics and Automation Letters (RA-L), 2025

Michael Novitsky [1], Moran Barenboim [1] Vadim Indelman [2,3]

[1]Technion Autonomous System Program (TASP) [2]Stephen B. Klein Faculty of Aerospace Engineering, Technion [3]Faculty of Data and Decision Sciences, Technion

Motivation

- Reduce expensive reward calculation;
- Leverage wasted historical and expert data;
- Accelerate planning without performance loss.

Key idea

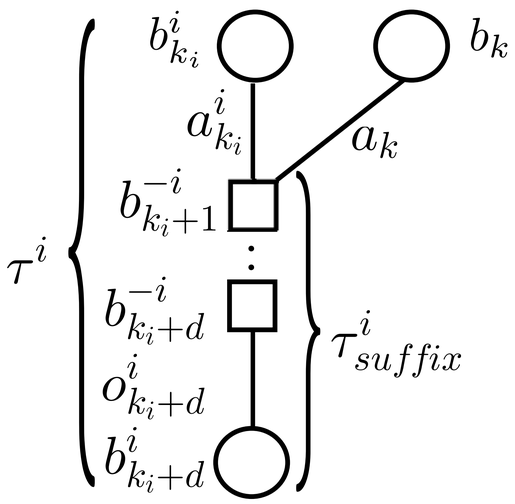
- Efficient incremental update method for the Multiple Importance Sampling estimator;
- Experience-based value estimation utilizing expert demonstrations without planning;
- An MCTS-inspired online algorithm (IR-PFT) that accelerates computations by reusing data from previous planning sessions.

Problem Formulation

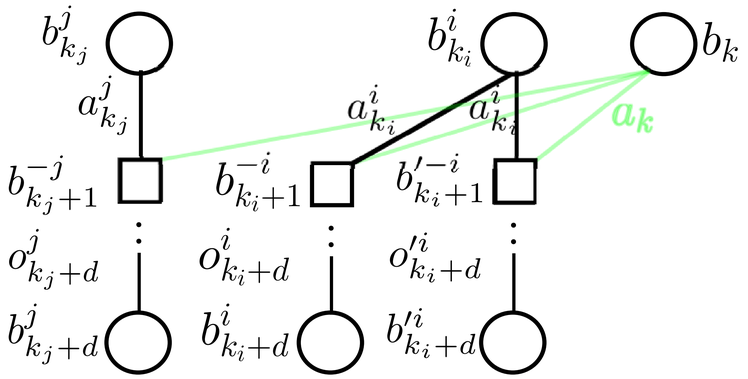
$$V^{\pi}(b_k) = {}_{b_{k+1:k+d}}[G_k \mid b_k, \pi],$$
$$\pi \triangleq \pi_{k:k+d-1},$$
$$G_k = \sum_{i=k}^{k+d-1} r(b_i, \pi_i(b_i), b_{i+1}),$$
$$Q^{\pi}(b_k, a) = {}_{b_{k+1}}[r(b_k, a, b_{k+1}) + V^{\pi}(b_{k+1})].$$

$$\hat{Q}^{\pi}_{MIS}(b_k, a_k) \triangleq \sum_{m=1}^{M} \sum_{l=1}^{n_m} \frac{P(b^{-l,m}_{k_m+1} | b_k, a_k)}{\sum_{j=1}^{M} n_j \cdot P(b^{-l,m}_{k_m+1} | b^{j}_{k_j}, a^{j}_{k_j})} \cdot \tilde{G}^{l,m}_{k}.$$

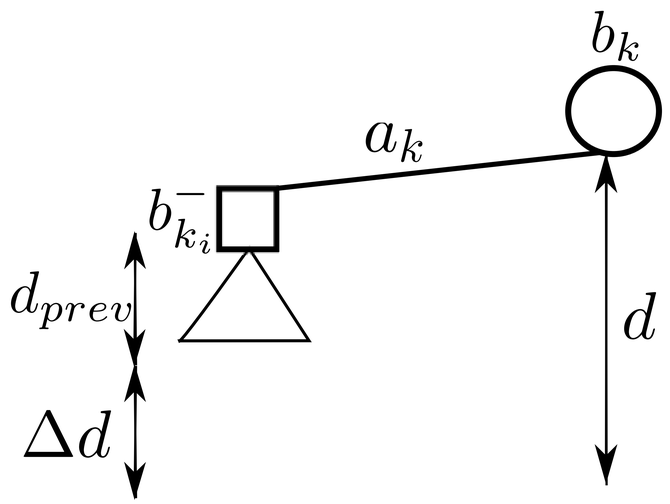
Trajectory Reuse



Multiple Trajectory Reuse



Horizon Alignment

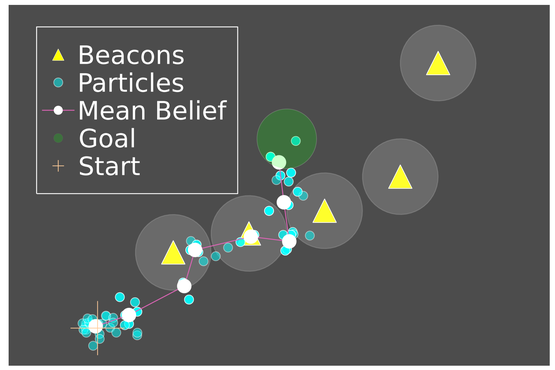


IR-PFT Decision Flow

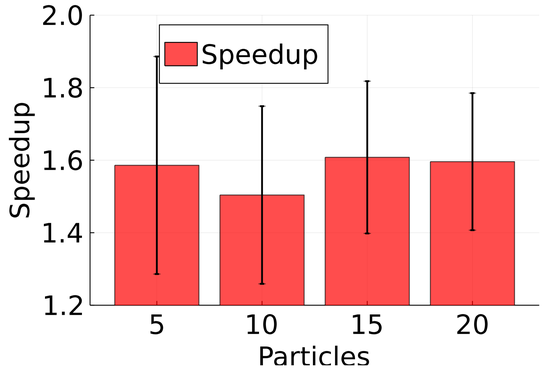
1. **Evaluate Condition:** Is current node the root and are prior candidates available?
2. If YES (Reuse Path):
   → **Select:** Retrieve best candidate b'^{-} from prior dataset D.
   → **Align:** Resolve horizon discrepancies (*Fill Horizon*).
   → **Update:** Calculate $Q(b, a)$ efficiently via Incremental MIS.
3. If NO (Simulate Path):
   → **Propagate:** Generate new belief via Particle Filter.
   → **Rollout:** Execute standard MCTS trajectory.

Results

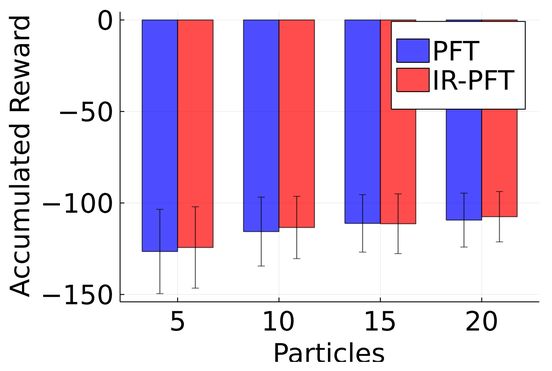
Light Dark



Speedup



Reward Comparison



miken1990@campus.technion.ac.il, moranbar@campus.technion.ac.il