Active Reward Learning with Comparative Language Feedback

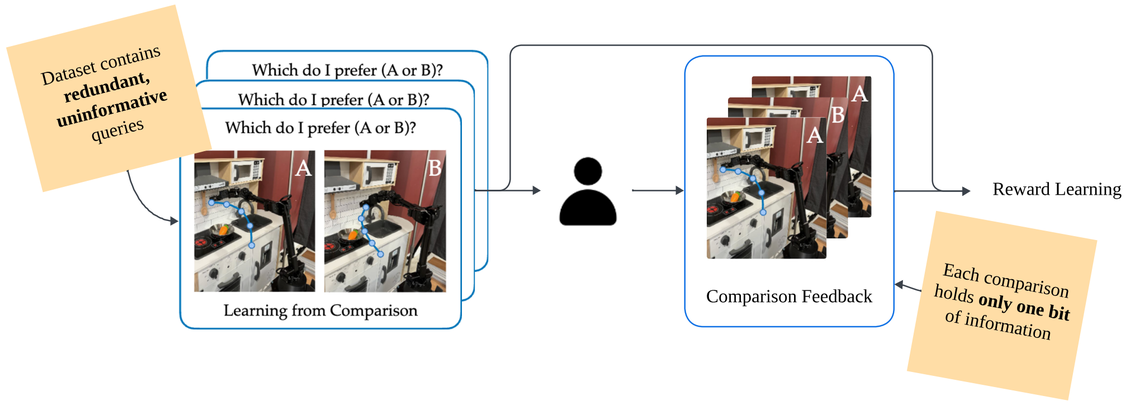
USC University of Southern California

Eisuke Hirota, Erdem Bıyık

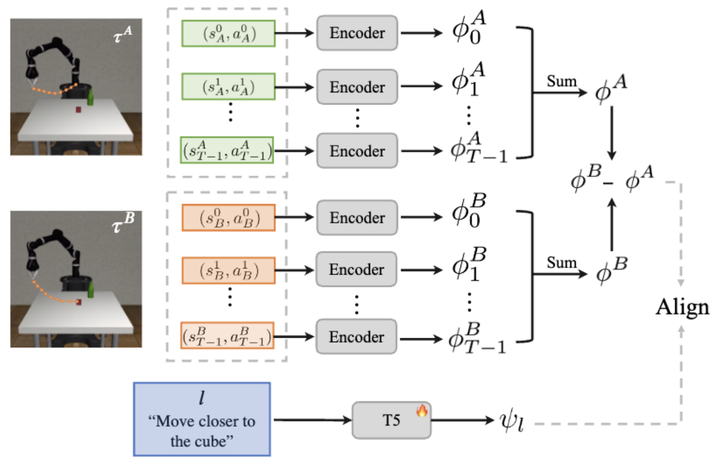
LiraLab

Our framework enables integrating Language Feedback for Active Reward Learning

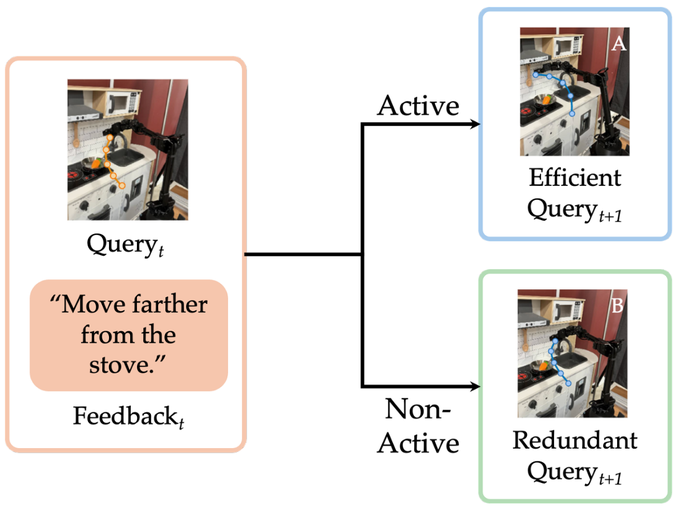
Naïve Reward Learning



Language Reward Learning



Language Active Reward Learning



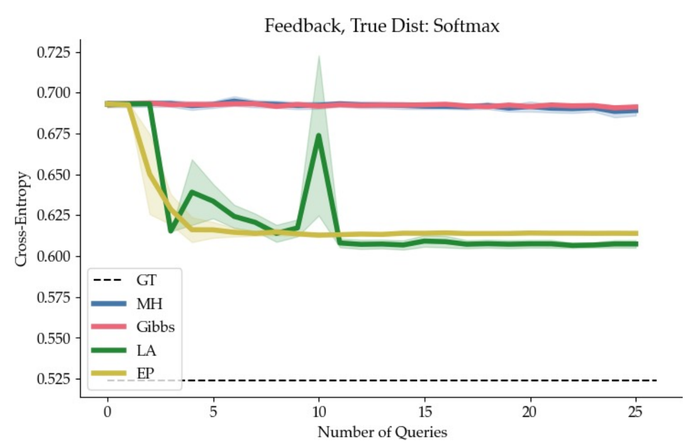
Algorithm 1 Active Learning Pseudocode

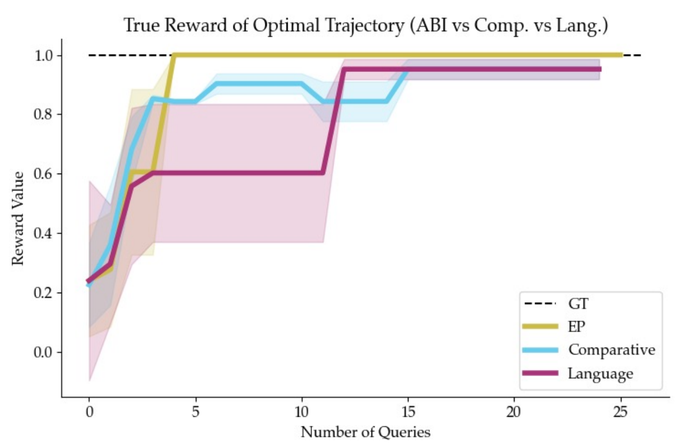
Require: $\mathcal{L}, \mathcal{T}, \text{fUpdater}(), \text{gUpdater}(), \text{infoGain}(), \epsilon$
$f = \mathcal{N}(\mathbf{0}_{512}, \mathbf{I}_{512})$ and $g = \mathcal{N}(\mathbf{0}_{512}, \mathbf{I}_{512})$
$\{\omega\}_{i=1}^m \sim f$ and $\mathbf{Q} = \{\}$
Score $= \infty$
while Score $> \epsilon$ **do**
  $f \leftarrow \text{fUpdater}(f, Q)$ and $\{\omega\}_{i=1}^m \sim f$
  $g \leftarrow \text{gUpdater}(g, Q, \{\omega\}_{i=1}^m)$ and $\{l\}_{i=1}^n \sim g$
  τ, Score $\leftarrow \text{infoGain}(\{\omega\}_{i=1}^m, \{l\}_{i=1}^n, \mathcal{L}, \mathcal{T})$
  Query the human with τ and receive l
  Append $\{\tau, l\}$ to $\mathbf{Q}$
end while

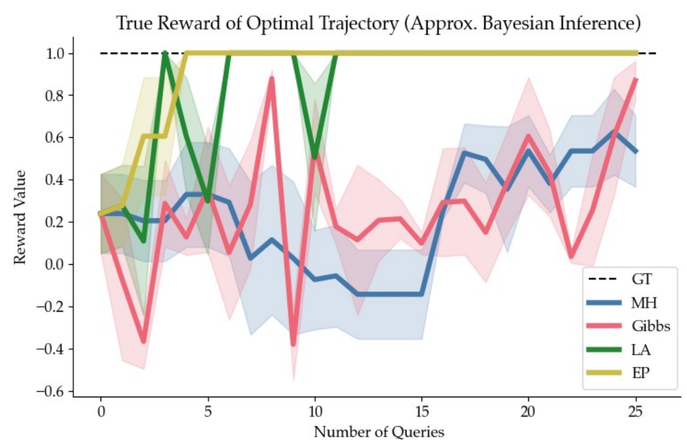
We experiment with four different Approx. Bayesian Inference methods:

- Metropolis-Hastings
- Stochastic Gibbs
- Laplace Approximation
- Expectation Propagation

Results and Key Insights







- We can learn a shared latent space between trajectories and language
- Using language feedback enables faster learning of the optimal reward function vs. comparison feedback
- Active Learning with language feedback enables extra speed up in terms of convergence

References

Tien, J., Yang, Z., Jun, M., Russell, S. J., Dragan, A., & Bıyık, E. (2024). Optimizing Robot Behavior via Comparative Language Feedback.

Acknowledgement

This work has taken place at Learning and Interactive Robot Autonomy Lab (LiraLab), University of Southern California. This research is supported by grants from the National Science Foundation.