Bayesian Nonparametrics for Offline Skill Discovery

Valentin Villecroze, Harry Braviner, Panteha Naderian,
Chris J. Maddison, Gabriel Loaiza-Ganem

ICML 2022

Key concepts

- Hierarchical reinforcement learning.

Key concepts

- Hierarchical reinforcement learning.
- Offline skill discovery.

Key concepts

- Hierarchical reinforcement learning.
- Offline skill discovery.
- Bayesian nonparametric models.

Our contribution

Our contribution

- We use a variational approach to offline option learning.

Our contribution

- We use a variational approach to offline option learning.

- We introduce a scheme to make the method nonparametric, which can be applied to other skill learning frameworks.

Our contribution

- We use a variational approach to offline option learning.
- We introduce a scheme to make the method nonparametric, which can be applied to other skill learning frameworks.
- We propose a practical implementation of a nonparametric (thus infinite) prior over skills.

Our model

- We assume the expert trajectories are generated by a hierarchical policy using options.

Our model

- We assume the expert trajectories are generated by a hierarchical policy using options.
- The high-level policy and the hidden trajectories of options and termination variables are considered as latent variables to be inferred through variational inference.

Our model

- We assume the expert trajectories are generated by a hierarchical policy using options.
- The high-level policy and the hidden trajectories of options and termination variables are considered as latent variables to be inferred through variational inference.
- We introduce an approximate posterior respecting the conditional independence inherent to the trajectories, which allows us to optimize the ELBO in a tractable way.

Nonparametric version

- The nonparametric version replaces the K-dimensional Dirichlet prior over the high-level policy to a GEM prior which results in assuming a countably infinite number of options.

Nonparametric version

- The nonparametric version replaces the K-dimensional Dirichlet prior over the high-level policy to a GEM prior which results in assuming a countably infinite number of options.

- In practice we still consider only a finite number of options K in the posterior but allow it to increase during training.

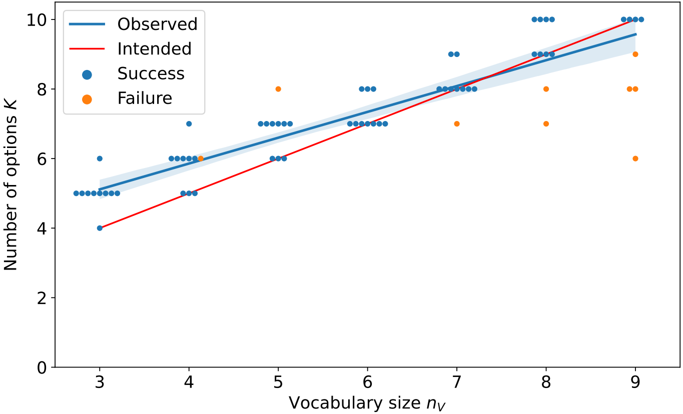
Nonparametric version

- The nonparametric version replaces the K-dimensional Dirichlet prior over the high-level policy to a GEM prior which results in assuming a countably infinite number of options.

- In practice we still consider only a finite number of options K in the posterior but allow it to increase during training.

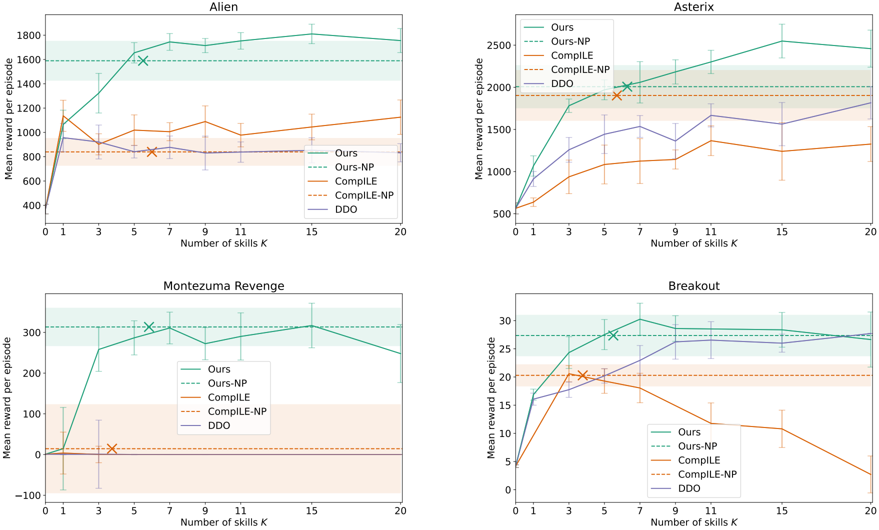
- Every n_K epochs, we check how much each option is used by the encoder. If every option is used more than a certain threshold, we add a new one (i.e. we increase K).

Proof-of-concept experiment



In this environment, the agent receives a message $m \in \{0, \ldots, n_V - 1\}$ at $t = 0$ and has to emit the same message as $t = 4$, i.e. take the action $a = m$.

Atari experiment



Conclusion

We introduced a novel approach for offline option discovery, and highlighted an unexplored connection between skill discovery and Bayesian nonparametrics.

Conclusion

We introduced a novel approach for offline option discovery, and highlighted an unexplored connection between skill discovery and Bayesian nonparametrics.

Thank you for listening!