Adaptive Learning Under Uncertainty With Variational Belief Deep Reinforcement Learning

Po-Chen Kuo (pckuo@uw.edu)
Neuroscience Graduate Program, University of Washington
1705 NE Pacific St, Seattle, WA 98195, USA

Han Hou (han.hou@alleninstitute.org)
Allen Institute for Neural Dynamics
615 Westlake Ave North, Seattle, WA 98109, USA

Edgar Y. Walker (eywalker@uw.edu)
Department of Physiology and Biophysics, University of Washington
1705 NE Pacific St, Seattle, WA 98195, USA

Abstract

Animals live in environments that are inherently uncertain and constantly changing. To thrive, they must learn to mitigate uncertainty and achieve their goals. For instance, when foraging in stochastic and dynamic environments, animals learn to adapt their strategies based on experience and trade off exploration and exploitation. Adaptive learning under uncertainty involves not only acquiring action-outcome contingencies but also discovering environmental regularities. Past computational modeling has largely studied the two aspects separately: contingency learning through reinforcement learning (RL) or structure learning through Bayesian inference. However, recent studies show animals may combine different strategies, which asks for an integrated approach to understand the computational basis of adaptive learning. Leveraging advances in deep RL and variational inference, we develop a flexible computational framework – variational belief deep RL – to incorporate Bayesian inference with RL. Focusing on a series of dynamic foraging tasks with various reward and temporal structures, we show how variational belief deep RL can provide effective modeling tools for both structural inference and fast adaptation to understand the computational and neural mechanisms of adaptive learning under uncertainty.

Keywords: deep reinforcement learning; Bayesian inference; decision-making under uncertainty; foraging

Introduction

To survive, animals need to adapt behavior and mitigate uncertainty to achieve their goals despite imperfect information from the environment. For instance, foraging is a canonical problem of adaptive learning under uncertainty in nature in which animals decide where and when to look for food among multiple options. Animals and humans are found to learn associations between actions and outcome in spite of the stochasticity and volatility of reward sources (Behrens, Woolrich, Walton, & Rushworth, 2007; Bari et al., 2019). Further, they make foraging decisions to balance the trade-off between exploiting familiar food sources versus exploring unknown alternatives (Hogeveen et al., 2022). Finally, organisms can use knowledge of the environmental structure and dynamics to guide efficient foraging decisions and flexibly adapt to new environments (Vertechi et al., 2020; Harhen & Bornstein, 2023). Understanding the computational and neural mechanisms of adaptive learning under uncertainty is a core challenge for computational and systems neuroscience.

Two families of computational models are used to explain how animals learn and adapt under uncertainty. Reinforcement learning (RL), by learning action-outcome association based on environmental feedback, learns to maximize cumulative reward (Bari et al., 2019). However, RL models often fall short of explaining animals' ability to reason about and adapt to hidden environmental states (Bromberg-Martin, Matsumoto, Hong, & Hikosaka, 2010). By contrast, Bayesian inference assumes animals use world models to perform inference to solve the task (Vertechi et al., 2020). However, it requires many assumptions and is computationally expensive. Recent studies indicate that animals may combine different strategies (Le et al., 2023), which asks for an integrated approach to understand the computational basis for learning and adaptation in uncertain environments.

Combining RL and Bayesian inference leads to the theory of Bayesian RL, providing elegant solutions to optimally trade off exploration and exploitation (Ghavamzadeh, Mannor, Pineau, Tamar, et al., 2015). However, its use is limited due to typically intractable computation. Building upon advances in deep RL and variational inference, we develop a flexible computational framework integrating RL and Bayesian inference to understand the computation basis of adaptive learning under uncertainty and demonstrate its application in neuroscience relevant tasks for further investigating neural mechanisms.

Variational Belief Deep Reinforcement Learning

The critical computation in Bayesian RL is first performing Bayesian inference to derive posterior distribution – or *belief* – of the true underlying environmental states and then learning a policy using RL algorithms. To circumvent the intractable computation, Zintgraf et al. (2021) proposed an ap-

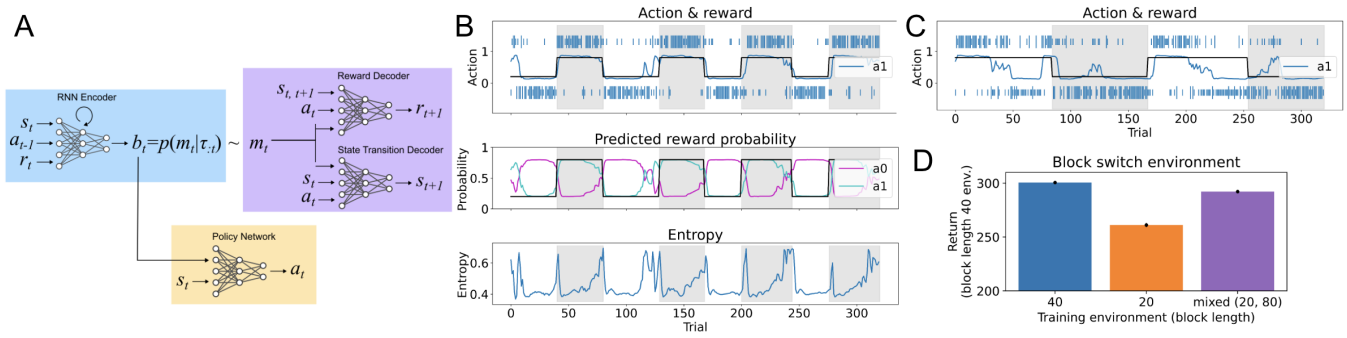

Figure 1: A) Model architecture of the proposed variational belief deep reinforcement learning framework, consisting of an recurrent neural network (RNN) encoder to approximate Bayesian update (blue), a set of reward and state transition decoder to promote representation learning (purple), and a deep RL policy network for decision-making (yellow). B) Example rollout in block switch environments. Top: action and reward trajectory, the long (short) ticks indicate (un)rewarded chosen actions for each trial, the blue curve shows policy (probability of choosing a_1), and the black curve is the underlying reward probability for a_1, white and gray shading demarcates blocks. Middle: predicted reward probability by the learned model, with magenta and cyan for a_0 and a_1, respectively. Bottom: entropy of the policy matches underlying environmental volatility. C) Same model as in B) but tested in environments with longer blocks. Note how incorporation of structural prior and ongoing experience shapes the behavior. D) Return (cumulative reward) of models trained in environments with average block length of 40 (blue), 20 (orange), and mix of 20 and 80 trials (purple), and evaluated in environments with 40-trial average block length.

proximation approach in stationary environments leveraging variational autoencoder, deep RL, and meta-learning. Extending this approach to dynamic environments, we develop variational belief deep RL, a tractable computational framework integrating for Bayesian RL. As shown in Figure 1A, the framework consists of three components: a recurrent neural network (RNN) encoder which learns to approximate Bayesian update by taking as input trial-to-trial observations s_t, actions a_{t-1}, and rewards r_t, and updating current belief of the true environmental states $b_t = p(m_t|\tau_{:t})$, where m_t is the environmental hidden states, and $\tau_{:t}$ is the trajectory till time t; secondly, a set of decoders predicting future reward and state transition to make the learned latent states relevant for the task; and finally a policy network that learns to maximize cumulative rewards. The entire network is trained using meta-learning to optimize performance across the task distribution.

To examine how the proposed model captures the computation of adaptive learning under uncertainty, we focus on the *dynamic foraging* task which is widely used in neuroscience to study how animals learn and adapt (Bari et al., 2019; Hattori & Komiyama, 2022). A two-armed bandit task where reward probabilities change over time, dynamic foraging can be designed with different reward structures (coupled reward, independent reward) and temporal structures (block switch, random walk), and hence are ideal for examining how agents learn both action-outcome contingencies and environmental regularities for decision-making under uncertainty.

Adaptive Learning of Dynamic Foraging

We first examine how variational belief deep RL models can effectively learn different task structures and use the learned structural prior to mitigate uncertainty. We applied variational belief deep RL to various dynamic foraging tasks with distinct reward (coupled and independent) and temporal (block switch or random walk) structures. Across different task structures, the trained models learn strategies that utilize corresponding structural knowledge and efficiently adapt to solve the tasks. For instance, Figure 1B shows one example rollout in coupled reward and block switch environments. The model trained on environments with average block length of 40 trials demonstrates effective choice allocation to match the underlying reward probabilities and efficient choice switch when the underlying environmental states are changed (Figure 1B, top). Further, the learned reward prediction model acquires not only accurate action values but also block structure (Figure 1B, middle). Finally, the policy network learns to adjust entropy of its choice according to the underlying volatility, effectively mitigate higher uncertainty around block transitions with a more stochastic policy and higher learning rate (Figure 1B, bottom).

Impact of Structural Priors on Generalization

To further investigate how the learned structural prior shapes strategy, we test the same model above in an unseen environment with average block length of 80 trials. As shown in Figure 1C, the model's behavior is the result of competition and combination of inference based on structural prior and adaptation to ongoing experience. The consequence of such incorporation leads to sometimes early switching due to the mismatched prior on block length acquired during training and other times successful adjustment to avoid early switching.

Finally, we examine how different structural priors learned by variational belief deep RL across different training task distributions affect its ability to generalize to new environments. Specifically, in Figure 1D, when the models are trained in en-

vironments with average block length of 20 trials but tested in environments with that of 40 trials (orange), there is a decrease in return due to mismatch in structural prior as compared to baseline models trained on the matching environments with 40-trial average block length. However, when the model is trained in environments with average block length of a mixture of 20 and 80 trials (purple), even though the model still has not seen environments with 40-trial average block length, the model performance is comparable to the baseline models. This demonstrates how variational belief deep RL can extract relevant task structures during training to facilitate efficient adaptation in unseen but related environments.

Conclusion and Discussion

To model how contingency learning through RL and structure learning through Bayesian inference jointly contribute to adaptive learning under uncertainty, we develop variational belief deep RL, a flexible computational framework to approximate Bayesian RL. Focusing on a series of dynamic foraging tasks, we demonstrate how variational belief deep RL effectively learns distinct task structures, mitigates uncertainty by combining structural priors and fast adaptation, and leverages learned priors for generalization to new environments.

Further analysis is needed to examine the learned representations of belief states and how they relate to task structures. In addition, analysis on the trained RNN encoders will help understand how approximate Bayesian inference and RL are achieved via network connectivity and dynamics. Finally, variational belief deep RL offers a computational model to investigate the neural mechanisms of how animals continuously updating their beliefs about the environment and adapting their decisions to mitigate uncertainty accordingly.

Acknowledgments

We thank Matthew Bull, Patrick Zhang, Jeremiah Cohen, and Karel Svoboda for helpful discussions.

References

Bari, B. A., Grossman, C. D., Lubin, E. E., Rajagopalan, A. E., Cressy, J. I., & Cohen, J. Y. (2019). Stable representations of decision variables for flexible behavior. *Neuron*, *103*(5), 922–933.

Behrens, T. E., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. (2007). Learning the value of information in an uncertain world. *Nature neuroscience*, *10*(9), 1214–1221.

Bromberg-Martin, E. S., Matsumoto, M., Hong, S., & Hikosaka, O. (2010). A pallidus-habenula-dopamine pathway signals inferred stimulus values. *Journal of neurophysiology*, *104*(2), 1068–1076.

Ghavamzadeh, M., Mannor, S., Pineau, J., Tamar, A., et al. (2015). Bayesian reinforcement learning: A survey. *Foundations and Trends in Machine Learning*, *8*(5-6), 359–483.

Harhen, N. C., & Bornstein, A. M. (2023). Overharvesting in human patch foraging reflects rational structure learning and adaptive planning. *Proceedings of the National Academy of Sciences*, *120*(13), e2216524120.

Hattori, R., & Komiyama, T. (2022). Context-dependent persistency as a coding mechanism for robust and widely distributed value coding. *Neuron*, *110*(3), 502–515.

Hogeveen, J., Mullins, T. S., Romero, J. D., Eversole, E., Rogge-Obando, K., Mayer, A. R., & Costa, V. D. (2022). The neurocomputational bases of explore-exploit decision-making. *Neuron*, *110*(11), 1869–1879.

Le, N. M., Yildirim, M., Wang, Y., Sugihara, H., Jazayeri, M., & Sur, M. (2023). Mixtures of strategies underlie rodent behavior during reversal learning. *PLOS Computational Biology*, *19*(9), e1011430.

Vertechi, P., Lottem, E., Sarra, D., Godinho, B., Treves, I., Quendera, T., ... Mainen, Z. F. (2020). Inference-based decisions in a hidden state foraging task: differential contributions of prefrontal cortical areas. *Neuron*, *106*(1), 166–176.

Zintgraf, L., Schulze, S., Lu, C., Feng, L., Igl, M., Shiarlis, K., ... Whiteson, S. (2021). Varibad: Variational bayes-adaptive deep rl via meta-learning. *Journal of Machine Learning Research*, *22*(289), 1–39.