Temporal Abstraction in Animal Exploration in a Complex Environment

Umesh K Singla (usingla@princeton.edu)
Princeton Neuroscience Institute, Princeton University
Department of Computer Science, UC San Diego

Marcelo G Mattar (marcelo.mattar@nyu.edu)
Department of Psychology and Center for Neural Science, New York University
Department of Cognitive Science, UC San Diego

Abstract

Exploration in sequential decision problems is a computationally challenging problem. Yet, animals exhibit effective exploration strategies, discovering shortcuts and efficient routes toward rewarding sites. Characterizing this efficiency in animal exploration is an important goal in many areas of research, from ecology to psychology and neuroscience to machine learning. In this study, we aim to understand the exploration behavior of animals freely navigating a complex maze with many decision points. We propose an algorithm based on a few simple principles of animal movement from foraging studies in ecology and formalized using reinforcement learning. Our approach not only captures the search efficiency and turning biases of real animals but also uncovers longer spatial and temporal dependencies in the decisions of animals during their exploration of the maze. Through this work, we aspire to unveil a novel approach in cognitive science of drawing interdisciplinary inspiration to advancing the field's understanding of complex decision-making.

Keywords: Exploration; Reinforcement Learning; Temporal Abstraction; Animal Behavior; Ecology; Foraging

Introduction

Understanding the exploratory and search behavior of humans and animals is a key focus in many scientific fields. The dynamics of exploration in neuroscience have been studied across mostly shorter temporal scales: from characterizing choice behavior in bandit tasks or studying head turns on encountering a novel object, but few studies have tried to model animals' exploratory behavior in larger or more complex environments. This study aims to fill that gap. The natural world is full of complex environments that require animals to navigate through intricate paths. Neuroscience experiments often fall short of replicating that complexity, limiting what we can learn about true animal behavior. This is in contrast to the field of spatial ecology, which has focused extensively on studying animal movement in naturalistic settings, from prey hunting in plain fields to bird migrations across oceans. However, there is a gap in exchange of ideas between ecology and neuroscience, partly because of the differences in the scale of investigation of the two fields.

With advanced animal tracking, there is now an increase in the use of complex environments with many choice points to study animal behavior. One such experiment conducted by Rosenberg, Zhang, Perona, and Meister (2021) involves ten mice, each exploring a complex labyrinth for close to 7 hours without any human interference. Animals had access to sufficient food and water in the home cage but curiously, even though the maze offered no explicit reward, animals continued to enter and exit the maze throughout the night to explore the maze. While this behavior supports the role of intrinsic motivation in driving animals to explore, the structure and remarkable efficiency exhibited in their exploration strategies constitute a perfect example of the complex and naturalistic behavior that remains poorly understood in the behavioral sciences. In their original paper, Rosenberg et al. (2021) characterized the animals' exploratory behavior using a computational model composed of four parameters that governed the probabilities of actions at each junction. However, this model was tailored to the specific dataset and maze layout. As such, it remains unclear if there are general computational principles capable of explaining the efficiency of animal exploration in this and other complex environments. Such principles should, ideally, also relate to known tenets of animal movement in spatial ecology.

In this study, using the maze exploration data from Rosenberg et al. (2021) as a case study, we built an exploration agent based on a few simple principles of animal movement from foraging studies and formalized using the framework of reinforcement learning (RL). Our main hypothesis is that, during exploration, animals rely on temporal abstraction to circumvent the complexity of sequential decision-making, giving rise to stereotyped action sequences. Computationally, we express this hypothesis in terms of a temporally-extended ε-greedy algorithm, recently proposed as a general exploration framework in RL by Dabney, Ostrovski, and Barreto (2020). Temporally-extended ε-greedy uses temporal abstraction to yield efficient exploration in a range of RL settings. However, Dabney et al. (2020) only compared this algorithm against perfect memory agents or neural networks, here we test its ability to explain real animal behavior. Our work makes a novel contribution to the field of cognitive science by providing a parsimonious characterization of the exploratory behavior of animals in a complex maze.

Model

We formalize the problem of exploration in the current maze as a Markov Decision Process (MDP). The set of states constitutes all the 63 nodes at T-junctions, 64 end nodes and the home node. At each of the 63 T-junctions in the maze, there are 3 actions available to go left, right or back. The transition probability matrix is deterministic and assumed to be known. There is 0 reward throughout the maze.

Temporally-persistent ε-greedy exploration A common strategy used in RL to promote exploration in sequential environments is ε-greedy. However, in a reward-free environment, relying on an ε-greedy strategy can be very inefficient (Dabney et al., 2020). In ε-greedy, the probability of being able to move away from one part of the environment to another reduces exponentially with the number of steps required. To tackle this, Dabney et al. (2020) proposed a temporally-extended version of ε-greedy. Rather than sampling an action at every time step, this algorithm instead samples a sequence of actions and executes this "composite" action. These composite actions, also known as options in the semi-MDP literature, abstract away the intermediate steps and allow flexible behavioral policies (Sutton, Precup, & Singh, 1999). The temporally-extended ε-greedy strategy requires choosing an exploration probability ε and an appropriate set of options O.

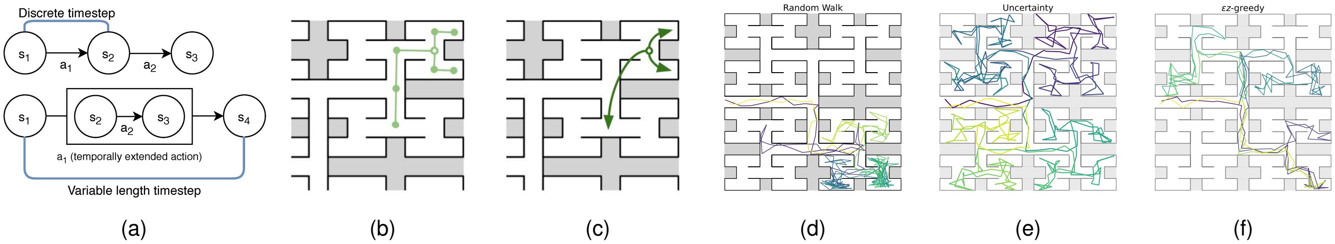

Figure 1: (a) The semi-MDP framework (bottom) allows the length between timesteps to be variable in comparison to standard MDP (top) and as such can support temporally extended actions. (b) If modeled as MDP, the agent has to choose an action at every intermediate turn in the maze. (c) In εz-greedy model, based on a semi-MDP framework, the agent first chooses a direction at random and samples an entire action-sequence and executes it in one go, abstracting away the decisions at intermediate timesteps. (d) A random walk agent, simulated for 100 time steps, gets stuck in a small region of the maze. (e) An exploration-bonus-based agent exhibits a very systematic behavior and performs very efficiently, however, at the expense of requiring intensive computations. (f) An εz-greedy agent, in the same amount of time, covers a much larger portion of the maze with minimal resources and generates efficient trajectories similar in structure to animals. Panel (a) adapted from Hutsebaut-Buysse et al. (2022).

Then, as with vanilla ε-greedy, it samples an option $w \in O$ with probability ε or follows a baseline policy with probability $1-\varepsilon$. For pure exploration, we set $\varepsilon = 1$, so that the agent always samples an option, eliminating the need to specify a baseline policy and a learning algorithm.

We adopt the spatial version of temporally-extended ε-greedy for our problem, called εz-greedy (Dabney et al., 2020). εz-greedy constructs an option w_{an} that takes the same action a for n time steps and terminates. The complete set of options O is made up of all such “action-repeats”, for all combinations of valid actions and durations where the duration n is sampled from some distribution z. These “action-repeats” allow an agent to persist in one direction and not get stuck in a local region, in contrast to a vanilla ε-greedy agent. To test εz-greedy on our data, we construct an appropriate set of options that encode a similar sense of directional persistence in our maze environment. For the duration distribution, we use the heavy-tailed Lévy distribution $z(n) \sim n^{-\mu}$ with $\mu = 2$. Being heavy-tailed, it samples a lot of short steps and spends time in one region but also has a non-zero probability to switch to a different region when a large step size is sampled. Such heavy-tailed distributions have been observed in many animal foraging studies in ecology (Viswanathan et al., 1999).

Results

Exploration efficiency The εz-greedy model exhibits efficient movement in the maze. We use the definition of exploration efficiency from the original study as the total number of nodes visited N_{half} required to survey half the end nodes, and define $E = 32/N_{half}$. An optimal agent with perfect memory visits the end nodes systematically without any repeats, resulting in an efficiency of $E = 1.0$. A random agent with no memory repeats a node before having visited all of them results in an efficiency of $E = 0.23$ when simulated. The exploration efficiencies observed for animals lie in the middle of the two, with an average of $E = 0.39 \pm 0.03$. The εz-greedy model gives an efficiency of $E = 0.35$ and accounts for 91% of the variance observed in the animals' efficiencies.

Turning biases The εz-greedy model also recovers the turning biases of animals. Data analysis showed animals exhibited a strong preference, consistent across all animals, to go forward at T-junctions and alternate at turns left and right. The εz-greedy model recovered all the turning biases within $\sim 90\%$ of animals' values. Rosenberg et al. (2021) had speculated on the presence of these consistent biases in animals in their paper, questioning if such rules are genetic. However, we show that just adhering to the general principle of directional persistence in an environment is sufficient to replicate these biases.

Our results show that a temporally-persistent agent captures the efficiency, the turning biases and many other aspects of mice behavior not included in this abstract. However, its main strength lies in its interpretability. The exploration movement patterns of humans and animals in open environments are known to be superdiffusive in nature and resemble Lévy walks (Viswanathan, Da Luz, Raposo, & Stanley, 2011). The success of εz-greedy model implies that mice exhibit superdiffusive movement within the maze and are optimizing for search efficiency. That is, once the animals have chosen to go in a certain direction, they do not make decisions at intermediate turns but continue to persist in that direction. By segregating the learning process and the innate mechanical aspects of a behavior, models like εz-greedy serve additional purpose by aiding in the selection of the appropriate formulation of the action space and the behavior policies. We now know that an RL policy that considers a spatiotemporally flexible state-action space is going to be more effective than the one trying to learn the local turning rules. Finally, being able to execute a temporally-extended option in this maze indicates that mice can sample a “jump” in arbitrary directions, even when those directions appear to be obstructed by the presence of maze walls. This implies a high degree of flexibility in their spatial decision-making and planning.

Acknowledgments

We are grateful to Rosenberg et al. (2021) for sharing their data in public. We thank Homero Esmeraldo, Kokila Perera, Matthieu Le Cauchois and Kris Jensen for many discussions on the analysis and modeling of the behavior as well as feedback on the manuscript. We thank departments of Cognitive Science and Computer Science at UC San Diego for providing resources and supporting interdisciplinary research.

References

Dabney, W., Ostrovski, G., & Barreto, A. (2020, June). Temporally-Extended {\epsilon}-Greedy Exploration. *arXiv:2006.01782 [cs, stat]*. (arXiv: 2006.01782)

Hutsebaut-Buysse, M., Mets, K., & Latré, S. (2022). Hierarchical reinforcement learning: A survey and open research challenges. *Machine Learning and Knowledge Extraction*, *4*(1), 172–221.

Rosenberg, M., Zhang, T., Perona, P., & Meister, M. (2021). Mice in a labyrinth show rapid learning, sudden insight, and efficient exploration. *Elife*, *10*, e66175.

Sutton, R. S., Precup, D., & Singh, S. (1999). Between mdps and semi-mdps: A framework for temporal abstraction in reinforcement learning. *Artificial intelligence*, *112*(1-2), 181–211.

Viswanathan, G. M., Buldyrev, S. V., Havlin, S., da Luz, M. G. E., Raposo, E. P., & Stanley, H. E. (1999, October). Optimizing the success of random searches. *Nature*, *401*(6756), 911–914. doi: 10.1038/44831

Viswanathan, G. M., Da Luz, M. G., Raposo, E. P., & Stanley, H. E. (2011). *The physics of foraging: an introduction to random searches and biological encounters*. Cambridge University Press.