Neuro-Nav: A Library for Neurally-Plausible Reinforcement Learning

Arthur Juliani
Araya Inc.
Minato-ku, Tokyo Japan
arthur_juliani@araya.org

Samuel A. Barnett
Princeton University
Princeton, NJ USA
samuelab@princeton.edu

Brandon Davis
Massachusetts Institute of Technology
Cambridge, MA USA
davisb@mit.edu

Margaret Sereno
University of Oregon
Eugene, OR USA
msereno@uoregon.edu

Ida Momennejad
Microsoft Research
New York City, NY USA
idamo@microsoft.com

Abstract

Here we propose Neuro-Nav, an open-source library for neurally plausible reinforcement learning (RL), with a focus on navigating diverse state spaces. RL is among the most common modeling frameworks for studying decision making, learning, and navigation in biological organisms. In utilizing RL, cognitive scientists often handcraft environments and agents to meet the needs of their particular studies. On the other hand, artificial intelligence researchers often struggle to find benchmarks for neurally and biologically plausible representation and behavior (e.g., in decision making or navigation). In order to streamline this process across both fields with transparency and reproducibility, Neuro-Nav offers a set of standardized environments and RL algorithms drawn from canonical behavioral and neural studies in rodents and humans. We demonstrate that the toolkit replicates relevant findings from a number of studies across both cognitive science and RL literatures. We furthermore describe ways in which the library can be extended with novel algorithms (including deep RL) and environments to address future research needs of the field.

Keywords: reinforcement learning, open source, navigation, neuroscience

Acknowledgements

1 Introduction

Understanding how humans and other animals make non-reactive decisions in their environments remains a longstanding interest of psychologists and neuroscientists alike, especially following the proposal of cognitive maps a century ago [1]. Applying the formalism of reinforcement learning to decision making has accelerated the progress in this field within recent decades [2]. Indeed, there is neuroscientific evidence that various forms of learning in humans and other animals can be modeled using the formalism of RL, and thus be described by classical algorithms from the RL literature [3, 4].

Applying the RL paradigm has led to a deeper understanding of learning from describing simple forms of conditioning in animals [5], to providing a basis for complex navigational strategies in humans [6]. This work is often made possible by abstracting the underlying complexity of the real environment-agent system into an idealized model. This process involves the real environment of the animal being abstracted into a computationally tractable Markov Decision Process (MDP) [7], consisting of an underlying graph of state nodes, which are connected to one another via edges corresponding to actions taken by the agent. Likewise, the biological organism making the decisions is abstracted into an artificial agent modeled by a policy and value function, which can interact with and potentially learn the relevant dynamics of the MDP.

This process of abstraction is largely non-standardized, and has seen various ad-hoc implementations over the decades, often driven by the particular experimental needs of the researchers at the time. Furthermore, many of these concrete implementations are not available in open-source forms, change when translated from one coding language to another, or the code has been lost altogether. As a result, the ability to replicate or build on previous work has been limited, confusing trainees, slowing progress in the field, and preventing the straightforward comparison of various models or algorithms to animal data. On the other hand, most AI benchmarks remain limited to maximizing scores and superhuman performance, which misses the opportunity for neurally and biologically plausible representation and behavior.

In this work, we present Neuro-Nav, an open-source library providing a standardized set of benchmark environments and RL algorithms which can be used to both replicate previous findings, test the biological plausibility of new models, and provide scaffolding for future experimental work in the field. The benchmarks focus on domains that are the basis of many experimental studies: spatial navigation in mazes (Figure 1 Top), associative learning, and graph navigation tasks (Figure 1 Bottom) [4]. In addition, we provide standardized implementations of classical RL algorithms such as Q-Learning, Successor Representation, and Model-based RL with Value Iteration, which can be evaluated using the benchmark environments. We provide these all within the context of a documented and tested open-source repository, which can be freely used and developed further by the broader research community.

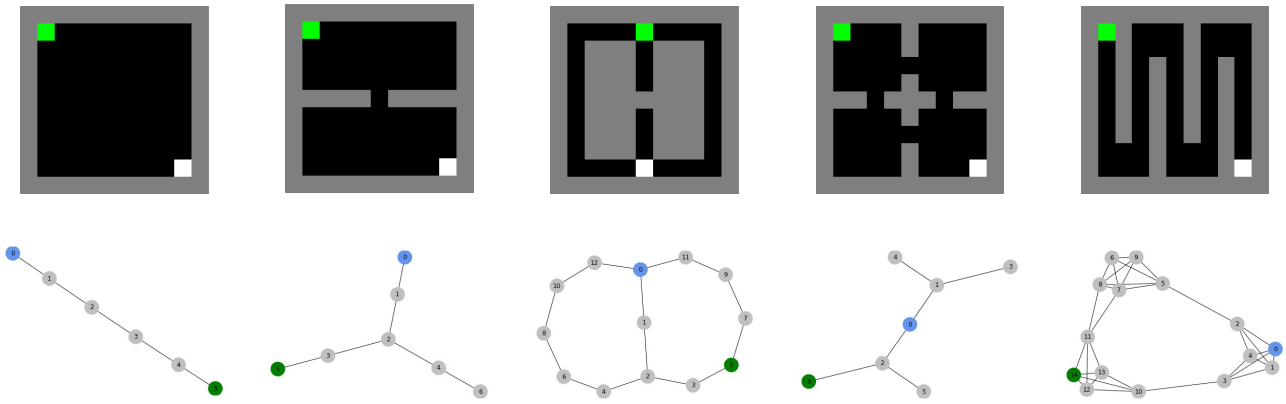

Figure 1: **Top**: Examples of maze environments. White square corresponds to agent location. Green square corresponds to goal location. Grey squares correspond to walls. **Bottom**: Examples of graph environments. Blue circle corresponds to agent location. Green circle corresponds to goal location. Environments drawn from studies in neuroscience [8, 9], classical RL [2], and cognitive science [10, 11].

2 Neuro-Nav Library

Neuro-Nav is an open source library[1] consisting of three components: a set of benchmark environments, a toolkit containing artificial reinforcement learning agents and algorithms, and a set of interactive notebooks for replicating findings

[1] https://github.com/awjuliani/neuro-nav

in the literature. All of the code in Neuro-Nav is written in the python programming language, and utilizes the numpy computational programming library to accelerate the relevant tensor mathematics.

2.1 Benchmark Environments

The environments included in Neuro-Nav consist of a set of graph and maze navigation tasks. In both cases, the underlying environment dynamics consists of a graph of 'state' nodes connected by 'action' edges. The agent occupies one of the nodes in the graph. The graph may also contain nodes that provide reward upon the agent occupying them, as well as a goal node, which both provides a reward as well as terminates the episode. Maze environments may be seen as a specific class of graph environments where the structure is mapped to a 2D grid, and the action space always consists of four actions: movement in each of the cardinal directions. Both types of environment are implemented using the OpenAI Gym interface, thus allowing for integration with a wide number of pre-existing open source RL codebases [12].

For each class of environments, we provide both an underlying library for defining and interacting with the environment as well as a set of pre-defined environments, which can be utilized for evaluation. These environments are often taken from existing literature, and serve to aid in the partial or complete replication of various findings from those original papers. In particular, we draw from environments defined in the context of neuroscience [8, 9], classical RL [2], and cognitive science [10, 11]. See Figure 1 for examples of a representative subset of the included maze and graph environments.

We also provide an abstraction for defining the observation space of the environment, which is completely orthogonal to that of the structure or topography of the environment. By default, we utilize a one-hot encoding for states, but also provide a set of alternative observational spaces such as euclidean distance from walls in the maze tasks, or CIFAR images in the graph tasks.

2.2 Algorithms Toolkit

Algorithm	Function(s)	TD	Replay	Value Iteration
TD-Q	$Q(s,a)$	Yes	No	No
TD-AC	$V(s)$, $\pi(a\|s)$	Yes	No	No
TD-SR	$\psi(s,a)$, $\omega(s)$	Yes	No	No
Dyna-Q	$Q(s,a)$	Yes	Yes	No
Dyna-SR	$\psi(s,a)$, $\omega(s)$	Yes	Yes	No
Dyna-AC	$V(s)$, $\pi(a\|s)$	Yes	Yes	No
MBV	$Q(s,a)$, $T(s'\|s,a)$	No	No	Yes
MBSR	$\psi(s,a)$, $\omega(s)$, $Q(s,a)$, $T(s'\|s,a)$	Yes	No	Yes

Table 1: Reinforcement learning algorithms included in the Neuro-Nav toolkit. 'TD' column: the algorithm utilizes an online temporal difference learning rule. 'Replay' column: the algorithm utilizes offline replay algorithm. 'Value Iteration' column: the algorithm utilizes an offline model-based learning procedure.

In addition to graph and maze environments, we provide a set of artificial agents, which implement canonical algorithms from the RL literature. We focus on tabular rather than deep learning algorithms, as this has likewise been the focus of much of the literature at the intersection of neuroscience and RL [11]. We include algorithms implemented either as model-free or model-based learning algorithms, as well as hybrid algorithms such as Dyna. See Table 1 for a description of the provided agent types. Each agent implements one of multiple possible algorithms for learning a policy and value function. Decisions are taken by the agent with either a softmax policy or epsilon greedy strategy, both of which are available to use in any agent type.

3 Experimental Results

We validate Neuro-Nav by replicating well-known results from human and rodent navigation and decision-making experiments. In particular, we replicate two classes of results: behavioral, which compares the observable behavior of biological and artificial agents, and representational, comparing learned representations of biological agents (using neuroscience) and artificial agents. Here we present Neuro-Nav results using benchmark from both categories.

In order to validate Neuro-Nav for replication of behavioral results, we focused on two recent studies in the literature, [10] and [11]. In the former, agent performance is compared to that of humans in a set of navigation tasks, where aspects of the environments are changed partway during learning. In the latter, agent performance is compared to that of rodents in a set of maze navigation tasks. We present replications of the findings from Experiment 1 of [10] in Figure 2. We present results consistent with the results of [11] in Figure 3.

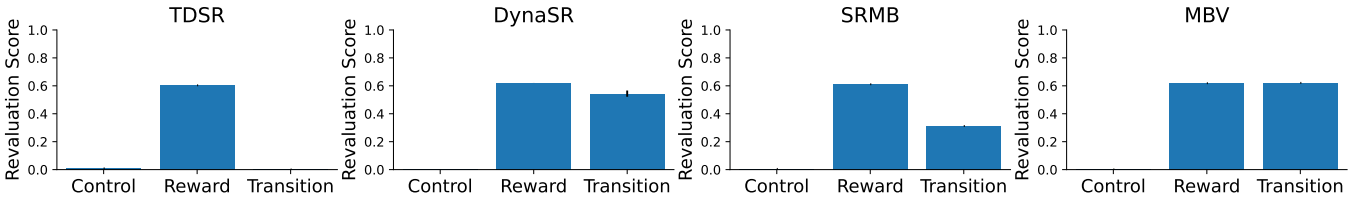

Figure 2: Replication of results from revaluation task described in Experiment 1 from [10]. Higher revaluation score corresponds to greater change in behavior after the re-learning phase. Both DynaSR and SRMB algorithms show most human-like revaluation behavior. Results averaged over ten experiments.

Beyond simply analyzing behavior, the internal content and structure of information utilized for decision making is of interest to many researchers. Here we focus on three relevant internal representations: value estimations, synthetic place fields, and grid fields [9]. We present Figure 4, which contains examples of each of these within an open field maze environment taken from an agent utilizing a successor representation.

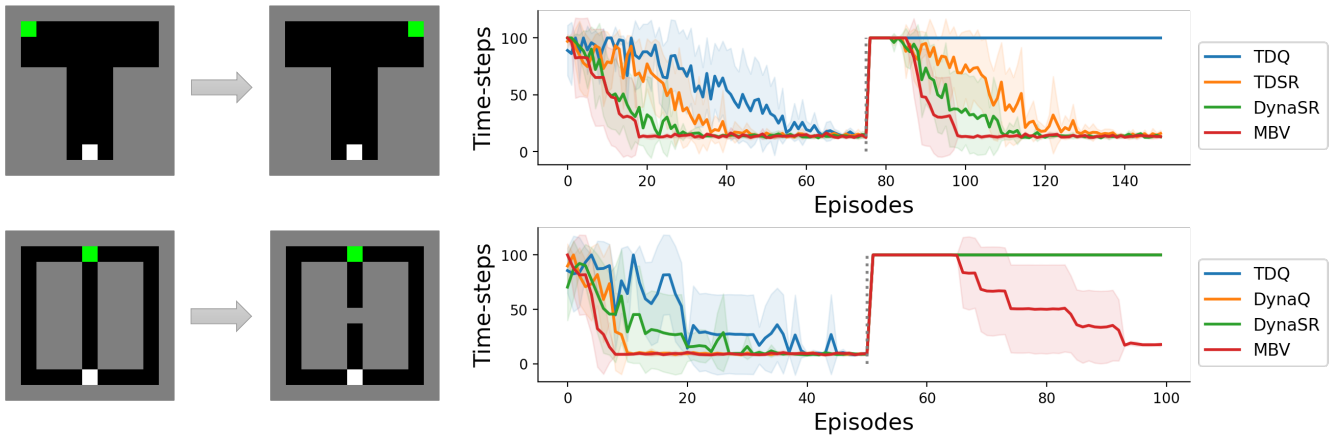

Figure 3: **Top**: Performance comparison between agent types on a reward transfer problem. Goal location changes at episode 75. SR and model-based algorithms successfully adapt to change. **Bottom**: Performance comparison between agents on a structure transfer problem. Environment structure changes at episode 50. Only model-based algorithm adapts to structural change. Graphs display average performance over five separate experiments. Tasks adapted from [11].

4 Discussion

Here we presented an open-source library of benchmark environments and an agent toolkit for performing reinforcement learning experiments on decision making and navigation. Neuro-Nav aims to empower reproducibility and standardization of evaluation within research in neurally plausible RL models of navigation and decision making. This project was developed with future extensions in mind, and as such, Neuro-Nav empowers users to easily develop new environments and algorithms beyond what is demonstrated in this work.

Neuro-Nav users can extend the benchmark environments in two ways. The first is by creating novel environments, i.e., MDPs, graphs, and maze topographies, to evaluate specific navigation, associative learning, or decision making experiments. The second is by defining novel observation spaces for the agents, such as using fractal or face/scene images as nodes or states, or perceiving euclidean distance from maze walls. These examples, among others, capture relevant paradigms in the neuroscience literature [8, 13]. Future work using Neuro-Nav can add novel environments and novel algorithms to the open-source library, compare the performance of RL algorithms on all environments, and thus replicate a broader class of studies within the field.

Currently, the Neuro-Nav library only supports tabular learning agents. As such, the agents provided here are not capable of learning from the more varied class of possible observation spaces. In contrast, it is clear that humans and other animals learn from complex multi-dimensional sensory signals. While much unrelated work in the field of deep RL has focused on learning from high-dimensional observation spaces [14], we believe there is an opportunity for a middle

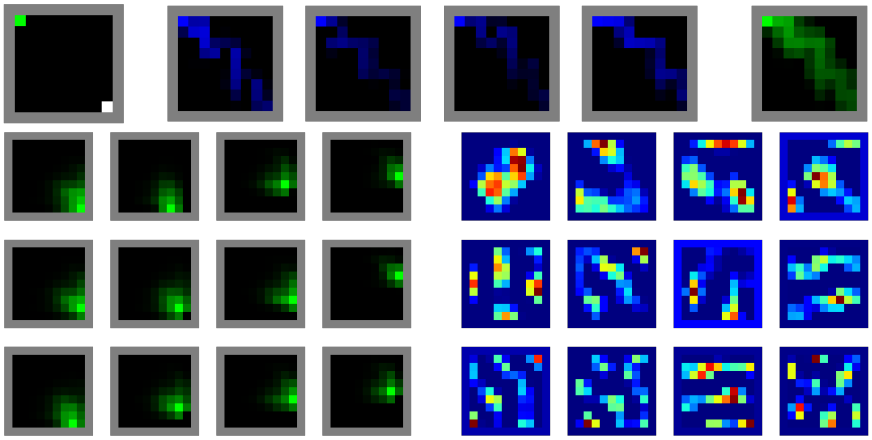

Figure 4: **Top Left**: Maze environment. **Top Middle**: Values of $Q(a|s)$ function in a maze environment after learning for each of the four actions. **Top Right**: Averaged value function over all actions, i.e. $V(s)$. **Bottom Left**: Values of a selection of units in the $\psi(s)$ function in a maze environment after learning. Units show place-like spatially selective fields. **Bottom Right**: PCA of $\phi(s)$ function in a maze environment after learning. Units show grid-like spatial selectivity. Procedure for generating place and grid fields adapted from [9].

path of lower-dimensional, but biologically grounded observation spaces, which enables linear models or simple neural networks to learn more expressive and generalizable behavioral policies. While we plan to include deep RL algorithms in future versions, we believe the current version of Neuro-Nav offers a promising step toward biologically plausible benchmarks, and a toolkit with potentially significant contributions to the field.

References

[1] Edward C Tolman. Cognitive maps in rats and men. *Psychological review*, 55(4):189, 1948.

[2] Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018.

[3] Yael Niv. Reinforcement learning in the brain. *Journal of Mathematical Psychology*, 53(3):139–154, 2009.

[4] Ida Momennejad. Learning structures: predictive representations, replay, and generalization. *Current Opinion in Behavioral Sciences*, 32:155–166, 2020.

[5] Richard S Sutton and Andrew G Barto. Time-derivative models of pavlovian reinforcement. 1990.

[6] Dylan Alexander Simon and Nathaniel D Daw. Neural correlates of forward planning in a spatial decision task in humans. *Journal of Neuroscience*, 31(14):5526–5539, 2011.

[7] Richard Bellman. A markovian decision process. *Journal of mathematics and mechanics*, pages 679–684, 1957.

[8] Anna C Schapiro, Timothy T Rogers, Natalia I Cordova, Nicholas B Turk-Browne, and Matthew M Botvinick. Neural representations of events arise from temporal community structure. *Nature neuroscience*, 16(4):486–492, 2013.

[9] Kimberly L Stachenfeld, Matthew M Botvinick, and Samuel J Gershman. The hippocampus as a predictive map. *Nature neuroscience*, 20(11):1643–1653, 2017.

[10] Ida Momennejad, Evan M Russek, Jin H Cheong, Matthew M Botvinick, Nathaniel Douglass Daw, and Samuel J Gershman. The successor representation in human reinforcement learning. *Nature human behaviour*, 1(9):680–692, 2017.

[11] Evan M Russek, Ida Momennejad, Matthew M Botvinick, Samuel J Gershman, and Nathaniel D Daw. Predictive representations can link model-based reinforcement learning to model-free mechanisms. *PLoS computational biology*, 13(9):e1005768, 2017.

[12] Greg Brockman, Vicki Cheung, Ludwig Pettersson, Jonas Schneider, John Schulman, Jie Tang, and Wojciech Zaremba. Openai gym. *arXiv preprint arXiv:1606.01540*, 2016.

[13] William de Cothi and Caswell Barry. Neurobiological successor features for spatial navigation. *Hippocampus*, 30(12):1347–1355, 2020.

[14] Kai Arulkumaran, Marc Peter Deisenroth, Miles Brundage, and Anil Anthony Bharath. Deep reinforcement learning: A brief survey. *IEEE Signal Processing Magazine*, 34(6):26–38, 2017.