MSc in Computational Cognitive Neuroscience

Goldsmiths
UNIVERSITY OF LONDON

Episodic Deep Meta-Reinforcement Learning

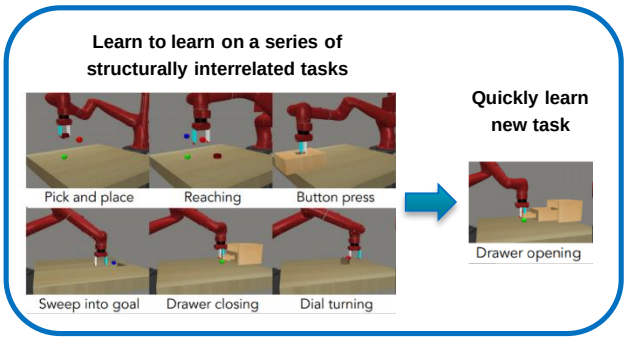
Badr AlKhamissy
Supervised by: Dr. Michael Spranger and Dr. Max Garagnani

Meta-RL and the Brain

This work addresses the sample efficiency problem observed in deep reinforcement learning (RL) systems by drawing on recent advances from both the machine learning and neuroscience literature. The goal is to develop a model that can deal with complex environments such as meta-world while still being able to reproduce neurophysiological recordings and observed behavior in humans and animals. Botvinick et al. (2019) argue that there are two main sources of slowness in deep-RL: incremental parameter adjustment and a weak inductive bias. Recent research has shown that both issues can be mitigated by augmenting an episodic memory system and using a meta-learning approach respectively.

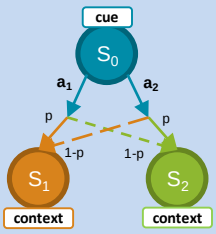
Wang et al. (2018) show that by conceptualizing the prefrontal-cortex (PFC) along with its subcortical components as its own free-standing meta-RL system one can explain a wide range of neuroscientific findings. They show that dopamine driven synaptic plasticity, that is model-free, give rise to a second more efficient model-based RL algorithm implemented in the prefrontal network's activation dynamics.

Figure 1: Meta-World Sample



Contextual Two-Step Task

This experiment is a variant of the two-step task. It is designed to disassociate multiple types of incremental and episodic learning. The emerged RL algorithm should show model-based learning in accord with observed human behavior.



Fast Learning via Slow Learning

Figure 2 shows a schematic of the model and learning procedure used in this work. The trust region policy optimization (TRPO) algorithm will be used to drive the slow learning process across a series of related tasks. The episodic memory will be implemented as a differentiable neural dictionary (DND) that is used to reinstate relevant activation dynamics encountered before to leverage prior experience and avoid redundant exploration. The working memory will be maintained in a variant of LSTM with an extra reinstatement gate. The goal is to quickly adapt to a new environment or task, that is similar to one the agent has seen before, with minimal amount of experience.

Visual Fixation Task

This experiment demonstrate the model's ability to account for Harlow's learning to learn effect. Instead of having monkeys choosing an object placed left or right. The artificial agent will either saccade to the left or right random image in order to maximize their chance of getting a reward by understanding the task's rules.

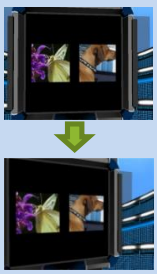
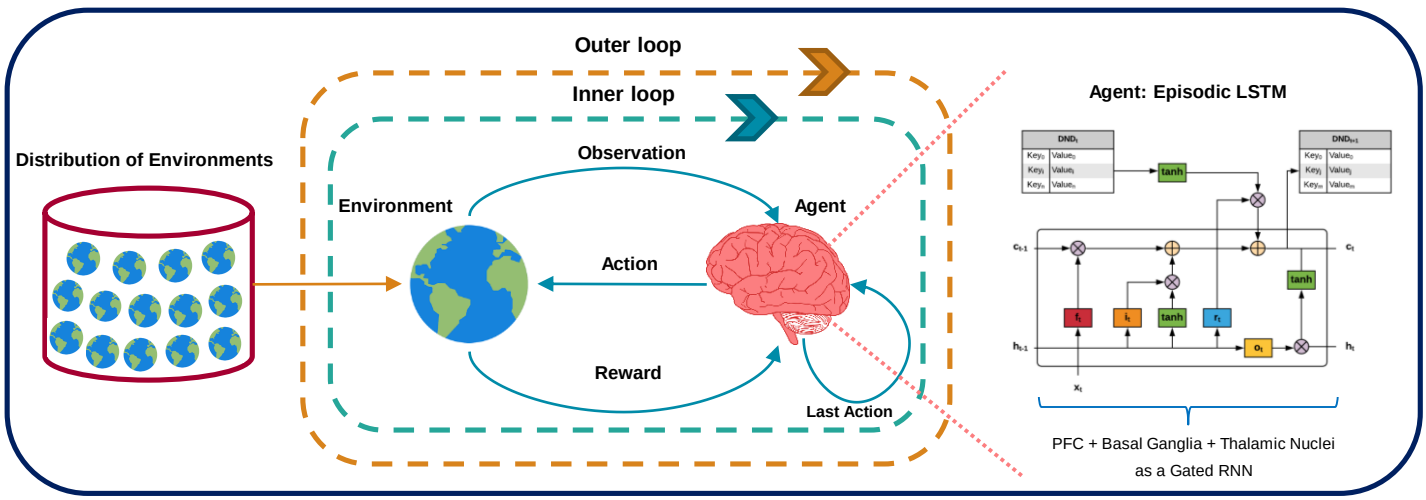


Figure 2: Model Schematic



1. Botvinick M, Ritter S, Wang JX, et al. Reinforcement Learning, Fast and Slow. Trends in Cognitive Sciences. 2019 May;23(5):408-422. DOI: 10.1016/j.tics.2019.02.006.

2. Wang, J.X., Kurth-Nelson, Z., Kumaran, D. et al. Prefrontal cortex as a meta-reinforcement learning system. Nat Neurosci 2018; 21: 860–868. https://doi.org/10.1038/s41593-018-0147-8