EE488F - Project Proposal

Mukul Chodhary (20236375), Kevin Octavian (20233989)

1 Motivation

Current best-performing reinforcement learning algorithms require a vast number of samples to train the agent in solving specific problems [1, 2]. The inefficiency arises due to the usage of semantic memory, which stores only the statistical summaries of event trajectories. However, solely relying on semantic memory is inefficient as some events happen infrequently, causing inferences on those states to be inaccurate. Another type of memory, called episodic memory, saves episodic traces along with their returns [3]. These episodic traces, then, will be utilized for future references when the agent is facing a similar state with the saved experiences. Combining semantic memory and episodic memory might allow RL agents to learn more efficiently and improve the data efficiency of RL algorithms.

2 Problem Statement

Given that current RL algorithms suffer from data inefficiency, we are planning to increase the efficiency and utility of a training dataset by using bio-inspired episodic memory for the agent. In this project, we limit the RL problems to those with discrete action space. Atari games will be used as the main environment to test our agent as they are widely used problems in the discrete action space domain. The specificity of the type of game is irrelevant, as different environments will be used to evaluate the agent's performance.

3 Problem History

Several attempts have been made to improve the sample efficiency of RL algorithms. In some research areas, researchers have tried to incorporate problem-specific priors to the agent by pre-training them with physical simulation data [4, 5], which results in up to 10-fold lower training epochs. Although the results are promising, this method can't be generalized to other applications and needs some expertise in determining which physical information should be used.

A more general idea for making RL more efficient is to configure how the agent samples the memory [6] or how the agent stores the trajectories into a memory/replay buffer [7]. In the former work, each trajectory transition is weighted differently according to the temporal-difference (TD) error. That way, an agent can put more focus on learning the high-error transitions. Unlike the former work, the latter work emphasizes the architecture of the memory itself. It implements episodic memory by borrowing the concept of a dictionary to the memory buffer and a neighbor-based algorithm to retrieve the samples. By rapidly integrating recent experiences to estimate the value function – as opposed to relying on gradient descent over the whole trajectory – this method outperforms most of the modern RL algorithms. Although the performance of this method excels in the earlier training periods, the prioritized replay buffer still manages to outperform the episodic memory in the long-term case.

4 Methods

We propose an agent with two modules, 1. the RL module and 2. an interface module. The RL module will consist of the reinforcement learning agent with episodic memory and the interface module will be responsible for converting the environment state into correct inputs to the RL module. Thus, allowing us to train the model on some known games and test it on an unknown game.

Furthermore, since most of the research direction in episodic memory is about how to store and retrieve experiences efficiently, we would like to explore the possibility of adding pattern-learning functionality inside the replay buffer such that the agents can do more exploration and exploitation more effectively. If time permits, we would also like to embed a layer into the model with environment-specific knowledge or features which may help the agent reach the goal state much quicker than a random exploration approach.

5 Project Timeline

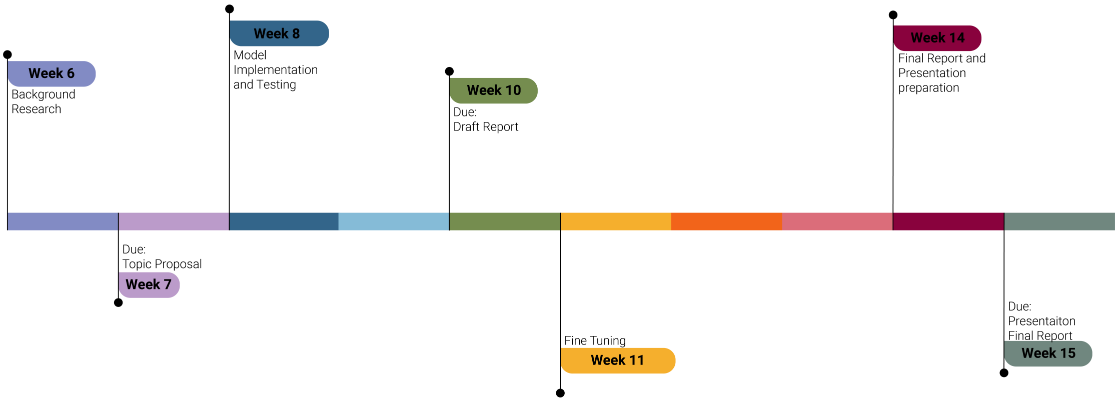

Figure 1: Project Timeline

References

[1] Volodymyr Mnih et al. "Human-level control through deep reinforcement learning". In: *Nature* 518.7540 (Feb. 2015), pp. 529–533. ISSN: 1476-4687. DOI: 10.1038/nature14236. URL: https://doi.org/10.1038/nature14236.

[2] Tuomas Haarnoja et al. *Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor*. arXiv:1801.01290 [cs, stat]. Aug. 2018. DOI: 10.48550/arXiv.1801.01290. URL: http://arxiv.org/abs/1801.01290 (visited on 10/11/2023).

[3] Samuel J. Gershman and Nathaniel D. Daw. "Reinforcement Learning and Episodic Memory in Humans and Animals: An Integrative Framework". In: *Annual Review of Psychology* 68.1 (2017). _eprint: https://doi.org/10.1146/annurev-psych-122414-033625, pp. 101–128. DOI: 10.1146/annurev-psych-122414-033625. URL: https://doi.org/10.1146/annurev-psych-122414-033625 (visited on 10/11/2023).

[4] Chaejin Park et al. *Physics-informed reinforcement learning for sample-efficient optimization of freeform nanophotonic devices.* arXiv:2306.04108 [physics]. June 2023. DOI: 10.48550/arXiv.2306.04108. URL: http://arxiv.org/abs/2306.04108 (visited on 10/11/2023).

[5] Colin Rodwell and Phanindra Tallapragada. "Physics-informed reinforcement learning for motion control of a fish-like swimming robot". en. In: *Scientific Reports* 13.1 (July 2023). Number: 1 Publisher: Nature Publishing Group, p. 10754. ISSN: 2045-2322. DOI: 10.1038/s41598-023-36399-4. URL: https://www.nature.com/articles/s41598-023-36399-4 (visited on 10/11/2023).

[6] Tom Schaul et al. *Prioritized Experience Replay.* arXiv:1511.05952 [cs]. Feb. 2016. DOI: 10.48550/arXiv.1511.05952. URL: http://arxiv.org/abs/1511.05952 (visited on 10/11/2023).

[7] Alexander Pritzel et al. "Neural Episodic Control". en. In: *Proceedings of the 34th International Conference on Machine Learning.* ISSN: 2640-3498. PMLR, July 2017, pp. 2827–2836. URL: https://proceedings.mlr.press/v70/pritzel17a.html (visited on 10/11/2023).