Toward Human Cognition-inspired High-Level Decision Making For Hierarchical Reinforcement Learning Agents

Rousslan Fernand Julien Dossa† and Takashi Matsubara‡

†Graduate School of System Informatics, Kobe University, Hyogo, 657-8501 Japan
‡Graduate School of Engineering Science, Osaka University, Osaka, 560-8531 Japan
Email: doss@ai.cs.kobe-u.ac.jp, matsubara@sys.es.osaka-u.ac.jp

Abstract—Hierarchical Reinforcement Learning methods aim to leverage the concept of temporal abstraction to efficiently solve long-horizon, sequential decision-making problems with sparse and delayed rewards. However, most HRL methods often base the decision-making process of the agent directly on low-level observations, while also using fixed temporal abstraction. We propose the Hierarchical World Model which can capture more flexible high-level, temporally abstract dynamics, as well as low-level dynamics of the system, which we posit is a natural extension to the HRL framework.

1. Background

Conventional Deep Reinforcement Learning (DRL) methods can be very sample inefficient when applied to long-horizon, sequential decision-making tasks, which usually overlap with sparse and delayed rewards problems. The Hierarchical Reinforcement Learning (HRL) framework aims to improve the efficiency of conventional (flat) RL by structuring the agent as a hierarchy of policies, where each level acts at a coarser time scale than the level below. In practice, most HRL methods rely on *fixed length temporal abstraction*. Moreover, the decision-making occurring at higher levels in the agent's hierarchy is still often based *on the observations at the lowest level*.

Concurrently, a growing body of studies of human behavior, cognitive science, and computational biology suggests that human behavior is hierarchically organized. Not only is the representation of knowledge structured in different levels of abstraction, hence forming a hierarchy, but so is the planning process and the execution of resulting plans [1]. It would thus be desirable to have HRL agents endowed with the ability to explicitly plan and act at a different level of abstraction.

2. Proposed method

In this work, we combine World Modeling methods [2] with the framework of Variational Temporal Abstraction [3] to form the **Hierarchical World Model** (HWM). Owing to its hierarchical structure, the proposed model inherently provides (1) a *temporally abstract state representation* summarizing an arbitrary number of lower-level states, and (2) an adaptive temporal abstraction mechanism to divide long-horizon, sequential decision tasks into smaller tasks of variable lengths.

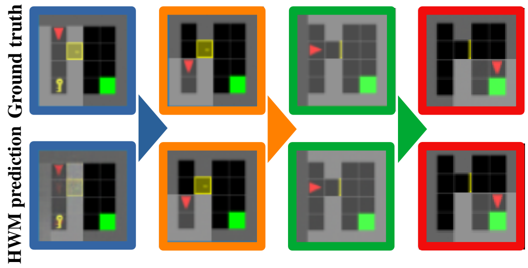

Figure 1: Low-level, and high-level temporally abstract dynamics captured by the proposed HWM

3. Results

We applied the proposed model to trajectories sampled from a partially observable, sequential decision-making task we refer to as *MiniGrid-DoorKey*. In this task, the RL agent starts in a maze where it needs to consecutively pick the key, open the door, and exit the maze. Preliminary results documented in Fig. 1 demonstrate that the HWM can predict low-level dynamics and perform reconstruction of partial observations (bottom row in Fig. 1), as in standard Model-based RL methods. Additionally, the model divides the trajectory of the agent into segments of variable lengths. Each segment corresponds to a temporally abstract state, represented by a different border color in Fig. 1, and often coincides with semantically meaningful states such as *picked up the key*, *opened the door*, and *exiting the maze*.

We posit that such high-level states learned can be used to augment existing HRL agents, allowing them to plan and explore more efficiently in temporally coarser, abstract state spaces while learning and executing variable-length low-level policies.

References

[1] M. Botvinick et al., "Hierarchically organized behavior and its neural foundations: A reinforcement learning perspective", *Cognition*, 2009.

[2] D. Hafner et al., "Dream to Control: Learning Behaviors by Latent Imagination", *International Conference on Learning Representations*, 2020.

[3] T. Kim et al., "Variational Temporal Abstraction", *Neural Information Processing Systems*, 2019.