Universiteit Leiden
The Netherlands

Enhancing autonomy and efficiency in goal-conditioned reinforcement learning

Yang, Z.

Citation

Yang, Z. (2025, February 26). *Enhancing autonomy and efficiency in goal-conditioned reinforcement learning*. *SIKS Dissertation Series*. Retrieved from https://hdl.handle.net/1887/4196074

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/4196074

Note: To cite this publication please use the final published version (if applicable).

Summary

Reinforcement learning is a framework that enables agents to learn in a manner similar to humans, i.e. through trial and error. Ideally, we would like to train a generalist agent capable of performing multiple tasks and achieving various goals. Goal-conditioned reinforcement learning is a step toward training such an agent. The goal-conditioned reinforcement learning framework comprises four steps: 1) defining the goal space; 2) selecting an interesting goal for the agent; 3) the agent learning to reach the goal; 4) the agent post-exploring.

In the thesis, we proposed four research questions and focused on three parts of the goal-conditioned reinforcement learning framework. First, in Chapter 3, we studied a more autonomous goal-conditioned reinforcement learning setting where there is no access to reset, which poses challenges on exploration. By using world models and selecting various goals to command the agent for data collection, our proposed agent can autonomously operate in reset-free tasks and outperform state-of-the-art baselines.

Once a goal is selected, a goal-conditioned policy is typically trained to reach the goal. Previous research on a non-parametric method, called model-free episodic control (Blundell, Uria, Pritzel, Li, Ruderman, Leibo, Rae, et al., 2016), shows that episodic control can quickly latch onto previously discovered solutions and learn faster than deep RL methods which are known to be slow. Instead of training a goal-conditioned policy, episodic control can also be employed to train the agent to reach the selected goal. However, model-free episodic control was designed for tasks with a discrete action space. In Chapter 4, we extend the episodic control to continuous episodic control that can now tackle tasks with a continuous action space, and demonstrate its strong performance on various continuous robotic control tasks. Moreover, limitations of episodic control methods are identified in Chapter 5, namely, they will be non-optimal in stochastic tasks. Then we proposed to combine episodic control with deep reinforcement learning methods to gain from both approaches. Experimental results

show that by combining these two methods, the unified agent achieves both the fast learning attributed to episodic control and the optimality attributed to reinforcement learning.

In Chapter 6, we demonstrated that additional exploration after reaching frontier goals—referred to as post-exploration—enhances the efficiency of GCRL agents. Without post-exploration, the agent resets after reaching the frontier goals, preventing it from expanding its knowledge boundary. By engaging in post-exploration, agents effectively step into new, unseen areas, acquiring more diverse data. This leads to improved performance compared to agents that do not perform post-exploration.

The methods proposed in the thesis improved components of the goal-conditioned reinforcement learning framework, including goal selection, policy learning and exploration, consequently enhancing the performance and increasing the autonomy of the entire goal-conditioned reinforcement learning framework. In future work, we hope that the goal-conditioned reinforcement learning framework will serve as a recipe for training a generalist agent, and ultimately, an embodied artificial general intelligent agent.