Formal and Natural Language assisted Curriculum Generation for Reinforcement Learning Agents

Doctoral Consortium

Yash Shukla
Tufts University
Medford, USA
yash.shukla@tufts.edu

ABSTRACT

Reinforcement Learning (RL) has proven successful in learning behaviors for artificial agents and robots when the transition dynamics of the environment are unknown. Despite this progress, many sequential decision making tasks are prohibitively expensive to learn. For my research, I intend to utilize and synthesize existing symbolic knowledge available to supplement the RL techniques for improved efficiency and faster learning progress. This symbolic information be in the form of formal language specifications (such as LTL) or in the form of natural language derived using Large Language Models (LLMs). I have developed various methods and frameworks that propose novel techniques in the curriculum learning domain to improve the learning efficiency of RL agents. I further want to implement these techniques on physical manipulator robot and show its efficacy for solving problems in the real world.

KEYWORDS

Large Language Models; RL; Curriculum Learning

ACM Reference Format:
Yash Shukla. 2024. Formal and Natural Language assisted Curriculum Generation for Reinforcement Learning Agents: Doctoral Consortium. In *Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), Auckland, New Zealand, May 6 – 10, 2024*, IFAAMAS, 3 pages.

1 INTRODUCTION

Despite the progress made in Reinforcement Learning, numerous sequential decision-making tasks are still excessively costly and challenging to learn in a practical manner. In response to this challenge, various research avenues have investigated the sequencing of tasks or data samples themselves in a curriculum, aiming to facilitate the learning progress of of agents in complex problems [3]. However, generating and optimizing a curriculum in a realistic scenario still requires extensive interactions with the environment. Other lines of work have investigated incorporation of symbolic knowledge (either in the form of PDDL, LTL or LLM) to aid the learning progress of the agent [1, 2]. Little work has examined the interplay between Symbolic knowledge and Curriculum learning. I will examine how Symbolic knowledge can be abstracted and utilized so that the RL policy shows improved performance with a faster convergence. Such techniques would enable us to learn several complex tasks on physical robots using RL techniques, which will make the RL policies and behaviors robust, and will also reduce the work-load on the humans engineering the task. I will begin by studying representations, and how symbolic knowledge can be abstracted either from LTLs or LLMs for efficient RL. With theoretical contributions using these two techniques, I will perform empirical evaluation on physical robots for robotic manipulation tasks. This will enable robots to learn a wide range of tasks, without needing to have an explicit policy for the individual task they are attempting.

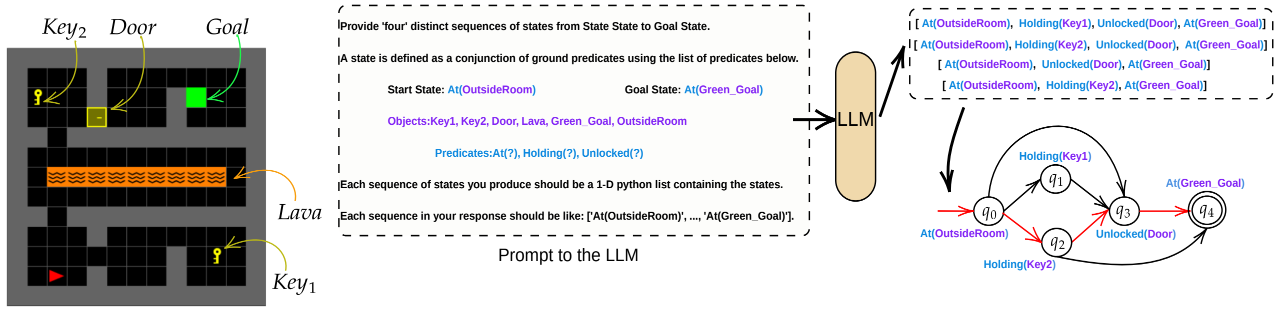
Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), N. Alechina, V. Dignum, M. Dastani, J.S. Sichman (eds.), May 6 – 10, 2024, Auckland, New Zealand.

2 BACKGROUND

I research Curriculum Learning techniques for Reinforcement Learning (RL) agents. ***All*** RL algorithms suffer from the sample-complexity problem, requiring millions of data points to train a simple and effective policy. In the realm of robotic research, this translates to extensive man-hours invested in setting up environments for expensive data collection and training behavior policies, consuming tens of graphics processing unit (GPU) days and contributing to the significant carbon emissions generated by AI systems. Curriculum, as applied in RL, explores how the arrangement of data samples can reduce the overall task sample complexity. However, a notable drawback of several curriculum learning algorithms is that the time required to generate a sample-efficient curriculum often surpasses the time needed to learn the target task from scratch. This undermines the purpose of having an automatically generated curriculum. Moreover, certain curriculum techniques necessitate task-specific engineering, relying on a curated reward function that proves ineffective when the task undergoes changes.

To investigate sample efficient algorithms for learning robust policies on a robot, I focus on two key areas - *Symbolic Knowledge abstraction for RL* and *Curriculum Learning*. Both of the above mentioned techniques are targeted toward efficient reinforcement learning for time-intensive robotic applications, however, little has been studied how the interplay of these two techniques can aid efficient and quicker learning.

Previously, I worked on Automaton-Guided Curriculum Generation for Reinforcement Learning Agents [5], in which the task goal is represented using a high-level specification language, such as LTL_f, and the goal is to come up with a curriculum that returns an ordered list or a graphical representation of tasks (Fig. 2). Long-horizon tasks have been traditionally challenging to solve, given the problem of catastrophic forgetting of RL systems. Hence, I used the high-level specification language LTL_f that allows representing task goals using temporal specifications, and the DFA form provides



(a) Gridworld domain

(b) Prompt to the LLM (left) and the LLM output and corresponding DAG (right)

Figure 1: (a) Gridworld domain and descriptors. The agent needs to collect one of the keys and open the door to reach the goal; (b) The prompt to the LLM that contains information about the number of paths n expected from the LLM and the symbolic information such as the entities, predicates and the high-level initial and goal states of the of the environment. The LLM prompt is converted to a DAG. The path chosen by *LgTS* is highlighted in red in the DAG in Fig. 1b

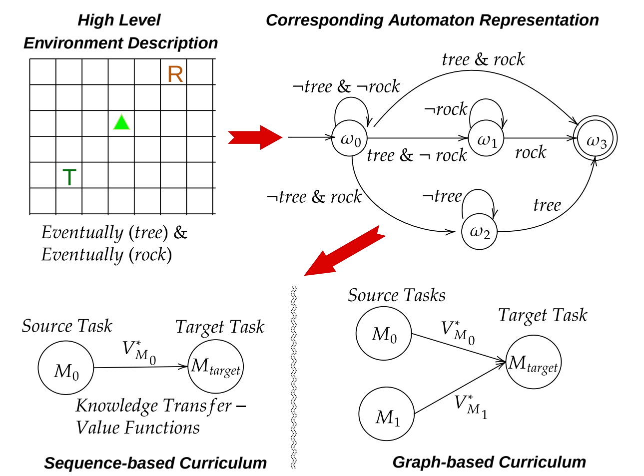


Figure 2: Overview of the automaton-guided curriculum generation approach. We use the automaton to come up with a sequence and graph-based curriculum.

a graphical representation of the sequence in which sub-goals must be achieved. However, it does not specify the individual sub-tasks of the curriculum. Thus, generating a curriculum is non-trivial as it requires the agent to reason over multiple potential curriculum environment configurations for the same sub-goal objective.

To come up with tasks in increasing order of difficulty, I used the Object-Oriented MDP representation to propose tasks that are nor too easy nor too difficult given the current learning capacity of the agent. Thus, the agent has a curriculum, in the form of a sequence or a graph based ordering of tasks, which help the agent learn the complex target task in fewer number of interactions. We demonstrated improved learning performance on a minecraft-like gridworld domain, as well as on two robotic domains - one robotic manipulation domain, and one robotic navigation domain. Our approach, *AGCL*, reduces the number of interactions with the target environment by orders-of-magnitude when compared to state-of-the-art automaton-guided RL baselines.

Later, I worked on LgTS [4]. Instead of having a human specify the high-level specification in the form of LTL, the graphical structure for the sub-goals of the task can be obtained by prompting an off-the-shelf Large Language Model (LLM) to provide us with a graphical representation of sub-goal sequences. This provides the agent a graphical structure, similar to the one obtained from human-specified LTL formulas as in [5]. The agent later proceeds to learn RL policies for solving the final task objective.

3 FUTURE WORK

In the near future, I want to use human-specified high-level language to dynamically choose tasks that can lead to quicker learning. In this work, instead of generating a manual curriculum, I will employ an interplay between two agents - A Teacher agent that proposes new tasks based on the LTL specification to the Student agent that attempts to learn a low-level RL policy for the task proposed by the Teacher agent. Here, the Teacher and the Student agents are RL policies, and the aim for the Teacher agent is to find the most suitable task based on the LTL specification for the student, and the aim for the Student is to learn the proposed task using RL policy. This interplay between the two agents promotes a curriculum strategy that improves the learning progress on the overall task.

In future, I want to move to learning robust policies for physical robots. The robot has access to sensory information, which can be abstracted to generate symbolic information. This symbolic information can be further enriched to be provided to an RL agent using off-the-shelf LLMs along with a Visual Language Model that can guide the low-level control of the robot by informing the robot regarding important spatial and environmental awareness that the LLM cannot provide. By combining all the tools that I have mentioned above, namely, symbolic knowledge abstracted using LLMs, situated awareness using VLMs, and curriculum learning techniques, I want to improve real world learning capacity for robots using reinforcement learning. While this is a challenging goal, reinforcement learning for robotics has been limited to high-level policies that are not robust due to limitations of the low-level control policies. End-to-end trained RL policies will mitigate this issue and promote robust and dynamic behaviors for task solving.

REFERENCES

[1] Rodrigo Toro Icarte, Toryn Klassen, Richard Valenzano, and Sheila McIlraith. 2018. Using reward machines for high-level task specification and decomposition in reinforcement learning. In *Intl. Conf. on Machine Learning*. 2107–2116.

[2] Kishor Jothimurugan, Suguman Bansal, Osbert Bastani, and Rajeev Alur. 2021. Compositional reinforcement learning from logical specifications. *Neural Information Processing Systems* (2021).

[3] Sanmit Narvekar, Bei Peng, Matteo Leonetti, Jivko Sinapov, Matthew E Taylor, and Peter Stone. 2020. Curriculum Learning for Reinforcement Learning Domains: A Framework and Survey. *JMLR* 21 (2020), 1–50.

[4] Yash Shukla, Wenchang Gao, Vasanth Sarathy, Alvaro Velasquez, Robert Wright, and Jivko Sinapov. 2024. LgTS: Dynamic Task Sampling using LLM-generated subgoals for Reinforcement Learning Agents. *International Conference on Autonomous Agents and Multiagent Systems.* (2024).

[5] Yash Shukla, Abhishek Kulkarni, Robert Wright, Alvaro Velasquez, and Jivko Sinapov. 2023. Automaton-Guided Curriculum Generation for Reinforcement Learning Agents. In *Proceedings of the 33rd International Conference on Automated Planning and Scheduling*.