Combining Reinforcement Learning and Large Language Models for Transfer Learning

Mohammad Saadati

Supervisor: Prof. Majid Nili Ahmadabadi

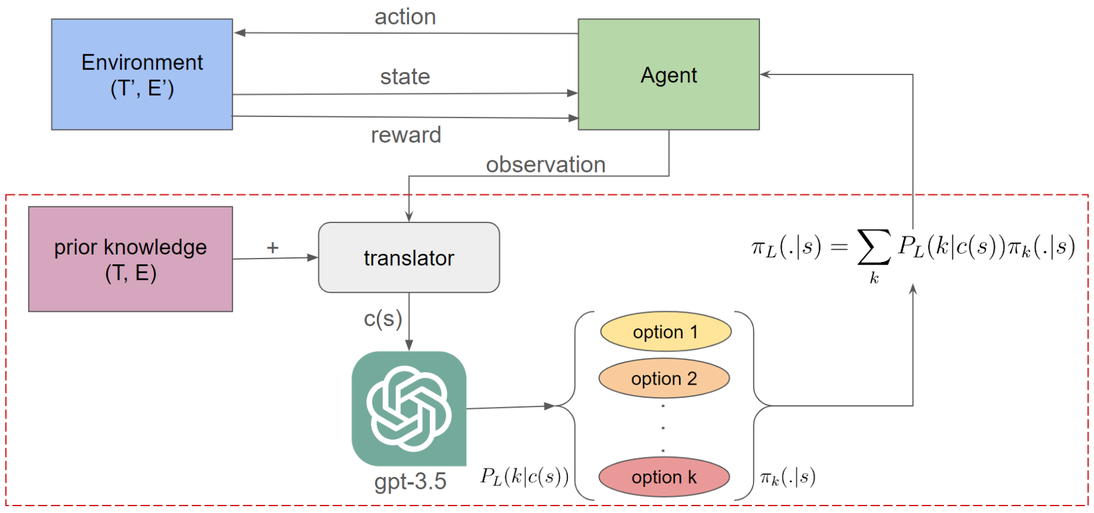
School of Electrical and Computer Engineering, University of Tehran

Introduction

Reinforcement learning, a powerful paradigm in machine learning, often faces challenges such as high exploration costs and slow learning speeds when applied to diverse and evolving environments. These challenges can lead to failures in developing effective models or require prolonged learning processes. Transfer learning leverages past experiences and knowledge to address new problems, thereby mitigating some of the obstacles encountered in reinforcement learning when dealing with modern, diverse issues. Consequently, employing transfer learning methods to enhance the efficiency of reinforcement learning has become a prominent research focus. In this project, we utilize large language models to identify similarities and differences between prior knowledge and new challenges, facilitating improved knowledge transfer from previous tasks to current ones. Our experiments demonstrate that large language models significantly boost learning and decision-making efficiency, thereby improving the effectiveness of reinforcement learning in solving complex problems.

Method

In this project, we propose RL3M4TL (Reinforcement Learning with Large Language Models for Transfer Learning), a novel framework that integrates large language models (LLMs) with reinforcement learning to enhance transfer learning. Our approach utilizes a pre-trained GPT-3.5 model to provide high-level guidance to RL agents. The goal of RL3M4TL is to mitigate exploration challenges in new environments and enhance the RL agent's ability to generalize learned policies across various tasks and scenarios. This ultimately improves the agent's performance in transferring knowledge from source domains (T, E) to target domains (T', E').



The RL3M4TL framework is illustrated in the figure above. The expert agent, based on an LLM, responds to state observations from the environment along with past knowledge by providing soft instructions, which are a distribution over a set of recommended actions. The learning agent balances maximizing expected rewards with following the LLM's guidance, gradually becoming more independent as it gains expertise.

When the agent sends a textual explanation $c(s)$ to the expert agent, the expert responds by providing a soft decision $\pi_L(\cdot|s)$, which is a distribution over existing policies. This is computed as follows:

$$\pi_L(\cdot|s) = \sum_k P_L(k|c(s))\pi_k(\cdot|s)$$

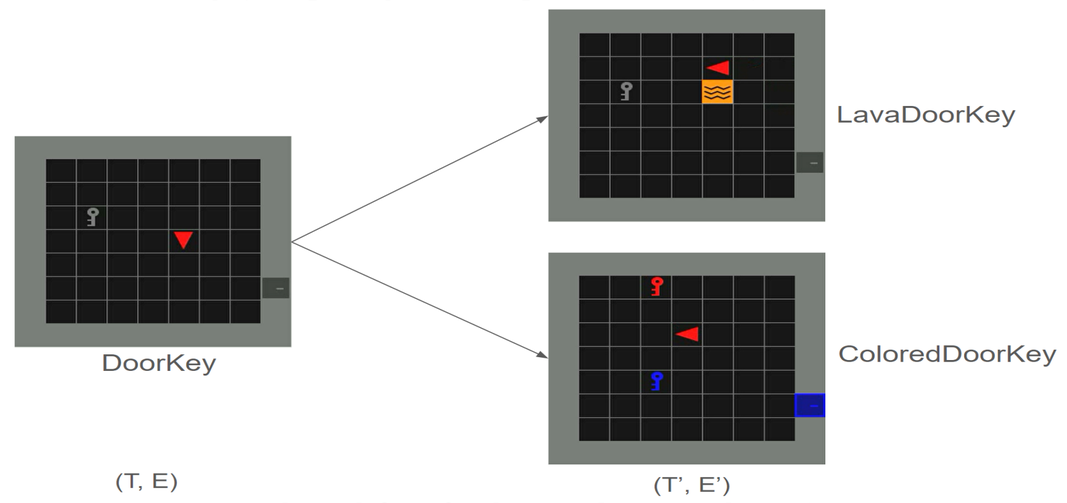
The RL agent's policy $\pi_\theta(\cdot|s)$ is learned by minimizing a combined loss function that balances the traditional reinforcement learning objective with the expert's guidance. The loss function is defined as follows:

$$\mathcal{L}(\theta) = \mathcal{L}_{RL}(\theta) + \lambda E_{s\sim\pi_\theta}\mathcal{H}(\pi_L(\cdot|s)||\pi_\theta(\cdot|s))$$

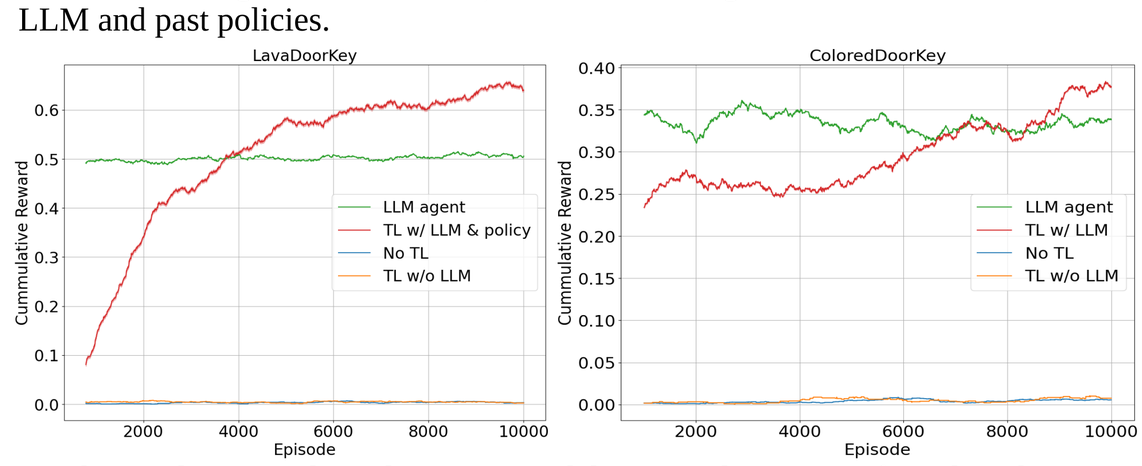
Here, $\mathcal{L}_{RL}(\theta)$ represents the standard reinforcement learning loss function aimed at maximizing expected rewards, while the additional term indicates the divergence between the expert guidance $\pi_L(\cdot|s)$ and the agent's policy $\pi_\theta(\cdot|s)$. The parameter λ controls the influence of the expert's guidance, where a higher value indicates a stronger reliance on the LLM's instructions.

Results

For our experiments, we used the MiniGrid environment, which provides configurable grid-based tasks with varying levels of complexity. These tasks pose challenges for reinforcement learning due to sparse rewards, requiring effective exploration. We utilized three MiniGrid tasks: DoorKey (finding a key to unlock the exit door), LavaDoorKey (avoiding hazards), and ColoredDoorKey (using a key matching the door's color).



Our experiments evaluated five fundamental approaches: 1) Training an agent without any prior transfer learning. 2) Learning based solely on a large language model (LLM) agent. 3) Transfer learning without using an LLM. 4) Transfer learning with an LLM. 5) Transfer learning with a combination of an LLM and past policies.



As observed, using a large language model as a guide can significantly enhance the learning process of an RL agent.

Conclusion

In this project, we introduced a framework that integrates large language models with reinforcement learning to enhance transfer learning. By leveraging large language models, this framework improves the speed of learning, decision-making, and policy generalization across tasks. Our experiments demonstrated that our framework significantly outperforms traditional reinforcement learning approaches in terms of sample efficiency and task completion success rates. The agent trained under this framework showed superior performance in both source and target tasks, indicating effective transfer learning. The inclusion of LLM guidance not only enhanced the agent's learning process but also reduced the computational resources required during online experiments.

The proposed approach has significant industrial applications in robotics, healthcare, finance, autonomous vehicles, and supply chain management, where it can enhance learning and adaptability to new environments.

References

1. **Taylor ME, Stone P. Transfer learning for reinforcement learning domains: A survey. Journal of Machine Learning Research. 2009 Jul 1;10(7).**
2. **Beck N, Rajasekharan A, Tran H. Transfer Reinforcement Learning for Differing Action Spaces via Q-Network Representations. arXiv preprint arXiv:2202.02442. 2022 Feb 5.**
3. **Schmitt S, Hudson JJ, Zidek A, Osindero S, Doersch C, Czarnecki WM, Leibo JZ, Kuttler H, Zisserman A, Simonyan K, Eslami SM. Kickstarting deep reinforcement learning. arXiv preprint arXiv:1803.03835. 2018 Mar 10.**