Connect the Goal and the State Using Language For a RL Agent

Jianhai Su and Qi Zhang

University of South Carolina

University of South Carolina

Background & Motivation

- To solve a new task, reinforcement learning agent generally needs to learn from the scratch. Several existing works [1-2] aim to improve the learning efficiency by reusing previously learned skills to solve subtasks in a hierarchical reinforcement learning setting. But none of them can support an extensible library of skills and active subtask suggestion together.
- Instruction-following agent [3] is designed to understand human's instruction to support human-robot collaboration scenarios. Besides, language description of visual observation can speed the learning and support policy transfer [4] by abstracting and clustering states.
- Considering the capability of pretrained large models (e.g., Flamingo [5], it will be possible and useful to enable the agent to actively reason in language to support mission completion.

Problem

- How to build a reasoning module for the agent to improve learning efficiency by reusing available skills and provide explanations of its policy decision along the mission?

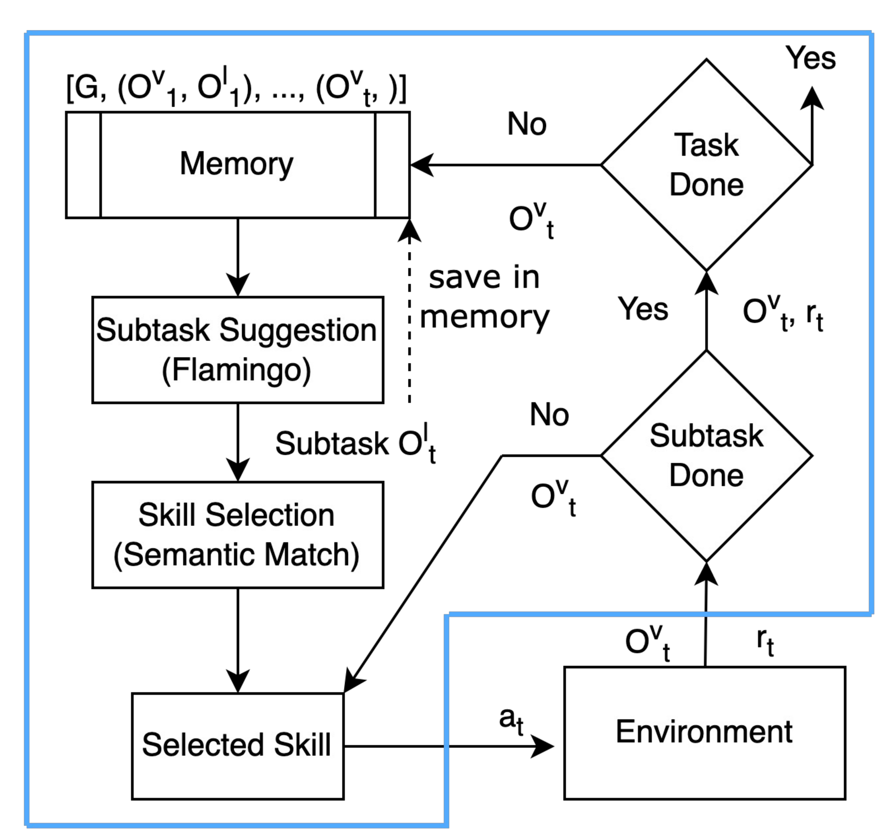
Application

- Human collaborators without expert knowledge in agent learning can understand the agent's behavior through the provided language explanation and provide guidance and instruction.
- In some scenarios, e.g. after earthquake, rescue robots are sent to search for victims. Instead of sending back videos and images that requires large transmission band, cost a big amount of time and energy, it could be more efficient to have the agent send back language explanation of its decision and summary of the environment for human collaborators.

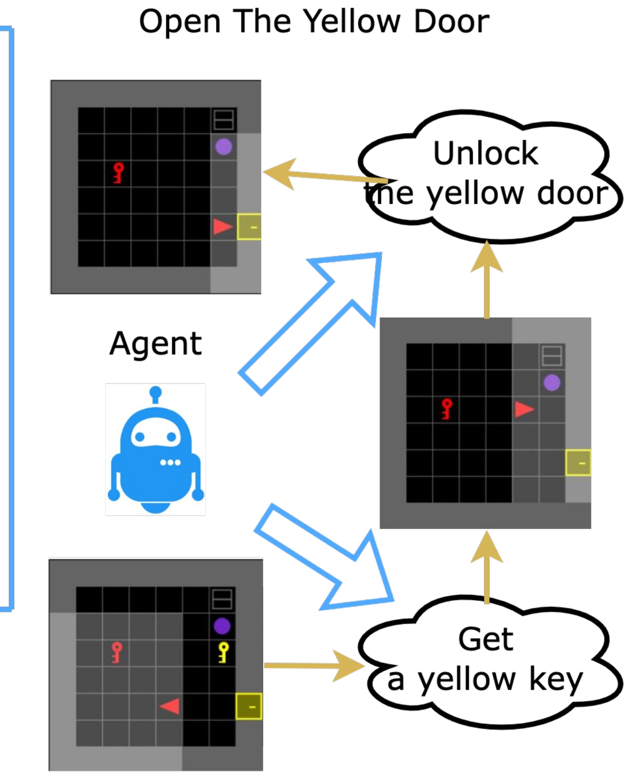
Approach

- In order to solve a task, the agent is trained in a hierarchical reinforcement learning (HRL) setting to iteratively prompt a meaningful subtask and select a promising skill for it.
- The agent proposes a subtask based on an accumulated semantic representation of the mission status and environment. The semantic representation is provided by the vision-language model, Flamingo [5], by cross-attending mission description, visual observations and subtask descriptions.
- A promising skill is selected using a pretrained large language model to measure the semantic distance between the subtask description and each skill's description.

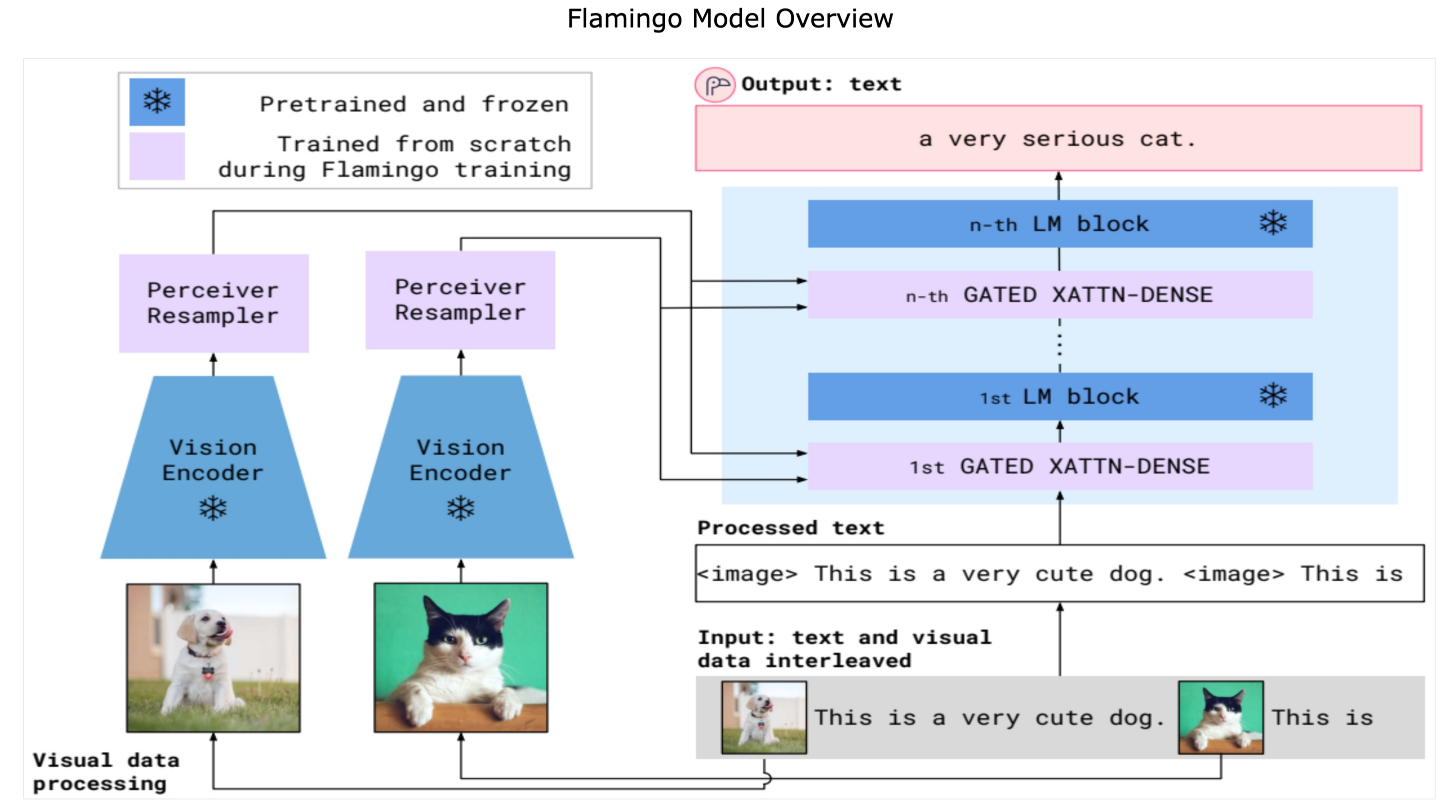
Agent Design



Gridworld Environment Example: Open The Yellow Door



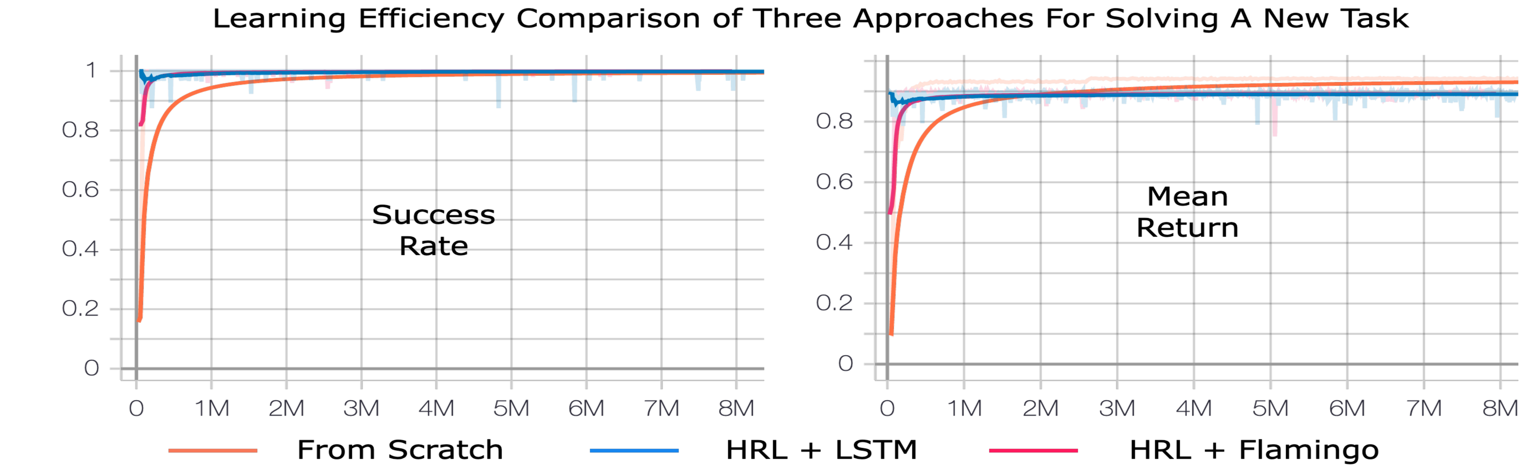
Flamingo Model Overview



Current Progress

- Implemented an episode-based PPO algorithm to train a HRL agent to apply Flamingo model to determine the next optimal skill. The module, "Skill Selection", is currently replaced by a policy header which decide which of the available skill should be used.
- The experiment is conduct in BabyAI platform [6] that provides partial-observation environment with a mission description supported.
- As show in the figure, the designed HRL agent can quickly learn to solve the task, unlock the door. This tells that the accumulated semantics by the agent is useful.

Learning Efficiency Comparison of Three Approaches For Solving A New Task



Future Work

- Design and implement the "Skill Selection" module
- When there is no promising skill, teaching agent to learn a new skill for achieving the subgoal
- When selected promising skill does not achieve expected performance, teaching the agent to polish the skill

Reference

1. Andreas, J., Klein, D. and Levine, S., 2017, July. Modular multitask reinforcement learning with policy sketches. In *International Conference on Machine Learning* (pp. 166-175). PMLR.
2. Jiang, Y., Gu, S.S., Murphy, K.P. and Finn, C., 2019. Language as an abstraction for hierarchical deep reinforcement learning. *Advances in Neural Information Processing Systems*, *32*.
3. Luketina, J., Nardelli, N., Farquhar, G., Foerster, J., Andreas, J., Grefenstette, E., Whiteson, S. and Rocktäschel, T., 2019. A survey of reinforcement learning informed by natural language. *arXiv preprint arXiv:1906.03926*.
4. Narasimhan, K., Barzilay, R. and Jaakkola, T., 2018. Grounding language for transfer in deep reinforcement learning. *Journal of Artificial Intelligence Research*, *63*, pp.849-874.
5. Alayrac, J.B., Donahue, J., Luc, P., Miech, A., Barr, I., Hasson, Y., Lenc, K., Mensch, A., Millican, K., Reynolds, M. and Ring, R., 2022. Flamingo: a visual language model for few-shot learning. *arXiv preprint arXiv:2204.14198*.
6. Chevalier-Boisvert, M., Bahdanau, D., Lahlou, S., Willems, L., Saharia, C., Nguyen, T.H. and Bengio, Y., 2018. Babyai: A platform to study the sample efficiency of grounded language learning. *arXiv preprint arXiv:1810.08272*.