Paper Critique

Shuvrajeet Das, DA24D402

Course: DA7400, Fall 2024, IITM
Paper: [Language as an Abstraction for Hierarchical Deep Reinforcement Learning]
Date: [06-09-2024]

Make sure your critique Addresses the following points:
1. The problem the paper is trying to address
2. Key contributions of the paper
3. Proposed algorithm/framework
4. How the proposed algorithm addressed the described problem
Note: Be concise with your explanations. Unnecessary verbosity will be penalized. Please don't exceed 2 pages.

---

1 The problem the paper is trying to address

The key problem is enabling agents to generalize and efficiently learn compositional, temporally extended tasks. Traditional RL methods struggle with such tasks due to their inherent complexity and the need for both long-term planning and systematic generalization.The authors propose that using language as an abstraction can significantly improve the process. Language provides a compositional structure that allows for the representation and reuse of concepts and sub-skills across different tasks. This hierarchical approach, where a high-level policy generates language-based instructions and a low-level policy follows these instructions, aims to make RL agents more flexible and able to generalize across tasks more effectively.

2 Key contributions of the paper

The contribution given in the paper are:

1. A hierarchical RL framework (HAL) that uses language as an abstraction between policies:

$$\pi_h(g_t \mid s_t), \quad \pi_l(a_t \mid s_t, g_t)$$

2. Compositional language improves generalization by leveraging structured instructions.

3. Hindsight Instruction Relabeling (HIR) to tackle sparse rewards:

$$R(s_t, a_t, s_{t+1}, g) = \Psi(s_{t+1}, g) \oplus \Psi(s_t, g)$$

4. An environment based on the CLEVR engine for evaluating long-horizon tasks.

5. Empirical results show better sample efficiency and generalization compared to existing approaches.

3 Proposed algorithm/framework

Algorithm 1 Overall Hierarchical Training

1: **Inputs:** Low-level RL algorithm A_l; High-level RL algorithm A_h; Environment $\mathcal{E}$;
2: $\pi_l(a \mid s, g) \leftarrow$ low-level policy trained with A_l and other appropriate inputs
3: $\pi_h(g \mid s) \leftarrow$ high-level policy trained with $A_h, \pi_l(a \mid s, g)$ and other appropriate inputs
4: **return** $\pi_l(a \mid s, g)$ and $\pi_h(g \mid s)$

4 How the proposed algorithm addressed the problem

1. Hierarchical Structure

The HAL framework introduces a two-layer hierarchical policy:

- **High-level policy** $\pi_h(g_t \mid s_t)$ generates language-based goals or instructions g_t, providing abstraction for long-horizon tasks.
- **Low-level policy** $\pi_l(a_t \mid s_t, g_t)$ follows the generated instructions g_t, completing specific subtasks associated with the goals.

2. Compositionality of Language

Language is inherently compositional, meaning it can represent complex tasks through combinations of simpler instructions. The algorithm leverages this property:

- **Language as a flexible abstraction**: The high-level policy operates in the space of language instructions g_t, which can represent abstract goals (e.g., spatial relationships between objects).
- **Reusable sub-skills**: The low-level policy π_l can be reused across different tasks as it operates based on language instructions, allowing generalization to unseen combinations of sub-tasks.

3. Hindsight Instruction Relabeling (HIR)

To deal with sparse rewards in long-horizon tasks, the algorithm employs **Hindsight Instruction Relabeling (HIR)**, inspired by Hindsight Experience Replay (HER). HIR relabels failed attempts by changing the original instruction g to a new instruction g' based on the outcome of the agent's actions.