Paper Critique

Shuvrajeet Das, DA24D402

Course: DA7400, Fall 2024, IITM
Paper: [Hierarchical Imitation and Reinforcement Learning]
Date: [27-08-2024]

Make sure your critique Address 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 paper addresses the challenge of learning effective sequential decision-making policies, especially in environments with sparse rewards and long time horizons, which are typically difficult for reinforcement learning (RL) algorithms.

Sparse Rewards: Situations where feedback (rewards) from the environment is infrequent, making it hard for RL algorithms to learn which actions are beneficial.

Long Time Horizons: Scenarios where the outcome of actions only becomes apparent after many steps, adding to the difficulty of learning effective policies.

The authors propose a framework called hierarchical guidance, which leverages the hierarchical structure of problems to combine different types of expert feedback. The framework integrates Imitation Learning (IL) and Reinforcement Learning (RL) at varying levels of the hierarchy.

2 Key contributions of the paper

2.1 Two-Level Hierarchy

- **High-Level (HI) Tasks**: The agent selects and sequences subtasks or subgoals (e.g., "go to the elevator," "take the elevator down," "walk out of the building").
- **Low-Level (LO) Tasks**: The agent executes these subtasks by performing a sequence of primitive actions (e.g., navigating to the elevator, pressing buttons).

2.2 Learning Process

- **HI-Level**: The agent learns a policy (meta-controller) that selects subgoals based on the observed state.
- **LO-Level**: The agent learns subpolicies that execute the chosen subgoals through sequences of primitive actions until the subgoal is achieved or a new one needs to be chosen.

2.3 Hierarchical Trajectories

- **HI-Level Trajectory**: A sequence of subgoals chosen by the meta-controller.
- **LO-Level Trajectory**: A sequence of actions that accomplish each subgoal.
- **Full Trajectory**: The concatenation of all LO-level trajectories, forming the overall sequence of actions from start to finish.

2.4 Expert Supervision

The agent can receive various forms of expert feedback:

- **Hierarchical Demonstration (HierDemo)**: The expert provides a full hierarchical trajectory, demonstrating the correct sequence of subgoals and actions.
- **HI-Level Labeling (LabelHI)**: The expert labels the next subgoal for each state in a given HI-level trajectory.
- **LO-Level Labeling (LabelLO)**: The expert provides the next primitive action towards a subgoal in a given LO-level trajectory.
- **LO-Level Inspection (InspectLO)**: The expert verifies whether a subgoal was accomplished, without providing action labels.
- **Full Trajectory Labeling (LabelFULL)**: The expert labels the entire sequence of actions without considering the hierarchy.
- **Full Trajectory Inspection (InspectFULL)**: The expert verifies if the overall goal was achieved.

3 Proposed algorithm/framework

Algorithm 1 Hierarchical Agent Behavior

```
for $h_{HI} = 1$ to $H_{HI}$ do
    observe state $s$ and choose subgoal $g \leftarrow \mu(s)$
    for $h_{LO} = 1$ to $H_{LO}$ do
        observe state $s$
        if $g(s)$ then
            break
        end if
        choose action $a \leftarrow \pi_g(s)$
    end for
end for
```

4 How the proposed algorithm addressed the described problem

The hierarchical guidance framework is presented as a method to accelerate learning and minimize the cost of expert feedback in both hierarchical imitation learning and hybrid imitation–reinforcement learning. The approach could be extended by considering weaker forms of feedback, such as preference-based or gradient-style feedback, or by using bandit-style imitation learning, where feedback only indicates if an action is correct or incorrect.