Paper Critique

Anirud N, CE21B014

Course: DA7400, Fall 2024, IITM
Paper: LEARNING MULTI-LEVEL HIERARCHIES WITH HINDSIGHT
Date: 30 August 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 Problem Statement

To realize the true potential of HRL and maximize its benefits, multi-level hierarchical levels must be introduced. The existing algorithms do not work efficiently learn multiple levels in parallel. Reasons for issues with learning multiple include:

- Non-stationary state function
- For the agents to effectively value actions, the distribution of states to which those actions lead should be stable
- Changes in the state function of one level can change the transition function for the other level.

2 Key Contributions

- Framework consists of two main components: (i) a specific hierarchical architecture and (ii) a method for concurrently learning multiple levels of policies despite sparse rewards.
- Introduces the concept of hindsight transitions - two types - action hindsight and goal hindsight transitions to enable learning multiple levels in parallel
- HER to accelerate learning in sparse reward tasks.

3 Proposed Algorithm/Framework

Multiple-level strategy - each level gives output as a subgoal that becomes the goal for the next lower level. HAc learns a combination or a nesting of different levels. This starts from the higher level - and the top level gives the policy to reach the goal state, which is a set of subgoals. It stops either when we run out of the allowed number of actions or when we reach the goal state. These subgoals are the goal state for the next level, and they output primitive actions. In other words - the action space at each level i is the goal space for the next lower level. $i = 0$ is the lowest level, and it uses primitive actions. Overall, the training is done by initializing the policies at each level and then using a Universal FUnction approximator to estimate them.

3.1 Hindsight action transition

This is used to get the transition function for the current level using the lower level optimal policy, i.e., π_{i-1}. The hindsight action transition uses the state reached (instead of the proposed subgoal) as the action. The reward function is -1 if the goal is not achieved and 0 if it is achieved. This gives more information about the sub-goals that the lower-level policies would otherwise not reach.

3.2 Hindsight Goal transition

Even if the rewards are sparse, the agent must learn effectively. The current lower-level policy is used to see the likeliness of achieving a subgoal. If the agent cannot reach the subgoal within H number of actions, we will pay the reward by H.

4 Conclusions and Results

On the whole, this algorithm introduces a way to implement a multi-level hierarchical framework. But there are a few issues in this algorithm :

- Restricted subgoal Learning: The level i can only learn from the transitions at level i-1, which is limited by most H transitions. This limit essentially means that there might be issues of not properly exploring state space - leading to missing out a few states.
- There is no mechanism to give positive rewards or incentivize whenever the relevant subgoal levels. This can result in subgoal proposals that the lower level cannot execute effectively with its current policy hierarchy, leading to inefficiencies.