Paper Critique

Shuvrajeet Das, DA24D402

Course: DA7400, Fall 2024, IITM
Paper: [Automatic Curriculum Learning through Value Disagreement]
Date: [30-10-2024]

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 inefficiency in sample usage in multi-goal reinforcement learning (RL). The primary goal is to develop an *automatic curriculum* that selects and prioritizes goals for training, specifically sampling goals at the *frontier* of the agent's abilities—those that are neither too easy nor too difficult. To achieve this, the authors introduce a *Value Disagreement Sampling (VDS)* technique. VDS measures the epistemic uncertainty in the Q-function, selecting goals with high uncertainty to maximize learning efficiency.

2 Key contributions of the paper

- Proposed a novel *Value Disagreement Sampling (VDS)* method for goal selection in multi-goal reinforcement learning (RL), which leverages epistemic uncertainty in the Q-function to create an automatic curriculum.
- Demonstrated that VDS improves sample efficiency by focusing on goals at the frontier of the agent's learning capability—those with high uncertainty that are neither too easy nor too hard.
- Validated VDS through experiments on 18 tasks, showing significant gains over existing curriculum-based methods in terms of sample efficiency and success rate.
- Showcased that VDS is compatible with existing techniques like Hindsight Experience Replay (HER), enhancing its performance on challenging tasks.

3 Proposed algorithm/framework

Algorithm 1 Curriculum Learning with Value Disagreement Sampling (VDS)

Input: Policy learning algorithm $\mathcal{A}$, goal set G, replay buffer R
Initialize: Learnable parameters θ for policy π_θ, and ensemble parameters $\phi_1, \ldots, \phi_k$ for Q-functions $Q_\pi^{\phi_1}, \ldots, Q_\pi^{\phi_k}$
for $n = 1, 2, \ldots, N_{\text{iter}}$ **do**
    Sample a set of goals $G = \{g_i\}$ from goal space
    Compute approximate goal distribution $\hat{C}_{\pi_\theta}$ using:

$$\hat{C}_{\pi_\theta}(g) = \frac{1}{\hat{Z}} f(\delta^\pi(g))$$

    where $\delta^\pi(g)$ is the standard deviation across the ensemble for goal g
    Sample goal $g \sim \hat{C}_{\pi_\theta}(\cdot)$
    Collect a goal-conditioned trajectory $\tau_n(\pi_\theta \mid g)$ and add transitions to R
    for each Q-function Q_π^ϕ in ensemble **do**
        Update ϕ via Bellman update:

$$Q_\pi^\phi(s, a, g) \leftarrow r + \gamma \mathbb{E}_{a' \sim \pi(\cdot|s', g)} \left[Q_\pi^\phi(s', a', g) \right]$$

    end for
    Update policy parameter θ using algorithm $\mathcal{A}$
end for
Return: Learned policy parameters θ

4 How the proposed algorithm addressed the described problem

The proposed *Value Disagreement Sampling (VDS)* algorithm addresses the inefficiency in multi-goal reinforcement learning (RL) by:

- **Targeted Goal Sampling:** Rather than sampling goals randomly, VDS uses the epistemic uncertainty of an ensemble of Q-functions to identify and prioritize goals that lie on the *frontier* of the agent's current capability, ensuring that these goals are neither too easy nor too hard.
- **Dynamic Curriculum Learning:** By focusing on high-uncertainty goals, VDS generates an automatic curriculum where the goals evolve with the agent's proficiency. This allows the agent to continually tackle progressively challenging tasks, optimizing sample efficiency.
- **Enhanced Learning Signal:** The algorithm's focus on uncertain, boundary-level goals provides a strong learning signal, maximizing the information gained per sample. This approach avoids repeatedly sampling solved (easy) goals and unsolvable (hard) goals, which improves convergence rates.