Paper Critique

Shuvrajeet Das, DA24D402

Course: DA7400, Fall 2024, IITM
Paper: [Mastering Atari Games with Limited Data]
Date: [18-10-2024]

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 challenge of mastering Atari games with limited data, focusing on reinforcement learning agents that require substantial training data to perform well. The problem is to improve the efficiency of data utilization while maintaining or improving performance in the task of playing Atari games. The goal is to find a policy π_θ that maximizes the expected cumulative reward while minimizing the amount of training data required.

2 Key contributions of the paper

The paper introduces several methods to enhance data efficiency in reinforcement learning for mastering Atari games, summarized as follows:

- **Efficient Data Augmentation:** Proposes novel augmentation techniques $\mathcal{A}$ to transform states $s \in \mathcal{S}$, improving generalization without requiring additional gameplay data.
- **Improved Policy Optimization:** Utilizes advanced optimization methods to find a policy π_θ that is more robust to limited data scenarios, improving the convergence of $J(\theta)$.
- **Data-Efficient Training Framework:** Implements a training pipeline $\mathcal{T}$ that reduces the total number of environment interactions N required to achieve a high-performing policy π_θ.
- **Evaluation on Limited Data:** Provides empirical results showing that the proposed approach outperforms baselines with fewer samples, proving the method's efficacy in environments where N is restricted.

3 Proposed algorithm/framework

Algorithm 1 Data-Efficient Reinforcement Learning Framework

```
Input: Initial policy parameters $\theta_0$, initial value function parameters $\phi_0$, learning rate $\alpha_\theta$, $\alpha_\phi$, environment $\mathcal{E}$, augmentation function $\mathcal{A}$, number of iterations $N$
for $i = 0, 1, \ldots, N - 1$ do
    Sample initial state $s_0$ from environment $\mathcal{E}$
    for $t = 0, 1, \ldots, T - 1$ do
        Augment state: $s'_t = \mathcal{A}(s_t)$
        Choose action: $a_t \sim \pi_{\theta_i}(s'_t)$
        Execute action $a_t$ in environment $\mathcal{E}$, observe reward $r_t$ and next state $s_{t+1}$
        Store transition $(s_t, a_t, r_t, s_{t+1})$ in replay buffer $\mathcal{D}$
    end for
    Sample a mini-batch of transitions from $\mathcal{D}$
    Compute loss $\mathcal{L}_\pi$ for policy and update parameters:
        $\theta_{i+1} \leftarrow \theta_i - \alpha_\theta \nabla_\theta \mathcal{L}_\pi$
    Compute loss $\mathcal{L}_V$ for value function and update parameters:
        $\phi_{i+1} \leftarrow \phi_i - \alpha_\phi \nabla_\phi \mathcal{L}_V$
end for
Output: Optimized policy $\pi_{\theta_N}$
```

4 How the proposed algorithm addressed the problem

- **Data Efficiency:** The use of data augmentation $\mathcal{A}(s_t)$ enables the agent to generalize better by learning from a transformed state space, thus reducing the number of environment interactions N required to reach a robust policy π_θ.

- **Efficient Policy Updates:** By optimizing the policy loss $\mathcal{L}_\pi$ and value function loss $\mathcal{L}_V$ with mini-batches from a replay buffer $\mathcal{D}$, the algorithm avoids overfitting to limited data and accelerates convergence to an optimal policy.

- **Replay Buffer:** Storing transitions (s_t, a_t, r_t, s_{t+1}) in $\mathcal{D}$ and sampling mini-batches allow the agent to reuse past experiences, further reducing the demand for new data and improving sample efficiency.

- **Augmentation for Generalization:** The augmentation function $\mathcal{A}$ generates diverse state variations, enhancing the model's ability to generalize across unseen states, thereby addressing the problem of data scarcity while maintaining high performance.