Paper Critique

Shuvrajeet Das, DA24D402

Course: DA7400, Fall 2024, IITM
Paper: [Unsupervised Curricula for Visual Meta-Reinforcement Learning]
Date: [25-10-2024]

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 problem of *unsupervised curriculum generation for meta-reinforcement learning (meta-RL)*. Current meta-RL approaches rely on manually defined training task distributions, which can be challenging and time-consuming to design. The authors propose a method to automatically generate these task distributions in an unsupervised manner, allowing for pre-training in visual environments without explicitly specified tasks.

They frame unsupervised meta-RL as an **information maximization** problem between a latent task variable and the meta-learner's data distribution. By alternating between updating the task distribution and meta-learning, the approach adapts the curriculum based on the shifting data distribution.

2 Key contributions of the paper

- The development of an unsupervised algorithm for inducing an adaptive meta-training task distribution, which serves as an automatic curriculum for meta-reinforcement learning (meta-RL).
- Formulation of unsupervised meta-RL as an information maximization problem, maximizing the mutual information between a latent task variable z and the meta-learner's trajectory data τ:
$$\max I(\tau; z) = H(\tau) - H(\tau|z)$$
- Introduction of a practical method that alternates between reorganizing the task distribution and meta-learning, leading to a curriculum that adapts as the meta-learner's data distribution shifts.
- Demonstration of how discriminative clustering can support trajectory-level task acquisition and exploration, especially in domains with high-dimensional, pixel-based observations.
- Empirical evaluation of the proposed method in vision-based navigation and manipulation domains, showing successful unsupervised meta-learning that transfers to downstream tasks defined by hand-crafted reward functions.

3 Proposed algorithm/framework

Algorithm 1 CARML Algorithm

Require: C, an MDP without a reward function

Initialize f_θ, an RL algorithm parameterized by θ
Initialize D, a reservoir of state trajectories, via a randomly initialized policy
while not done **do**
  E-step: Fit a task scaffold q_ϕ to D (using Algorithm 2)
  for desired mixture model-fitting period **do**
    Sample a latent task variable $z \sim q_\phi(z)$
    Define the reward function $r_z(s)$ and a task $T = C \cup r_z(s)$
    Apply f_θ on task T to obtain a policy $\pi_\theta(a|s, D_T)$ and trajectories $\{\tau_i\}$
    M-step: Update f_θ via meta-RL (e.g. RL2 algorithm)
  end for
  Add new trajectories to D: $D \leftarrow D \cup \{\tau_i\}$
end while
Return a meta-learned RL algorithm f_θ tailored to C

4 How the proposed algorithm addressed the problem

- **Unsupervised Task Discovery:** CARML automatically generates task distributions without requiring human-specified rewards. This solves the problem of manually designing training task distributions, which is time-consuming and impractical for complex environments.
- **Information Maximization:** The algorithm maximizes the mutual information between a latent task variable z and the meta-learner's data distribution τ:

$$\max I(\tau; z) = H(\tau) - H(\tau|z)$$

  This ensures that the tasks discovered are diverse yet structured, providing the meta-learner with a variety of tasks that are both distinguishable and learnable.

- **Adaptive Curriculum:** By alternating between task acquisition (E-step) and meta-learning (M-step), CARML continually updates the task distribution based on the meta-learner's evolving data distribution. This addresses the need for a curriculum that adapts to the learner's capabilities over time.
- **Scalability to Visual Domains:** CARML uses discriminative clustering and deep generative models to enable unsupervised task acquisition from high-dimensional pixel-based observations, addressing the challenge of task discovery in visually rich environments.
- **Improved Transferability:** The discovered tasks allow for unsupervised meta-learning that transfers to downstream tasks, meaning that the learned strategies are reusable and provide better pre-training for more efficient supervised meta-learning in new task distributions.