PEARL
Efficient Off-Policy Meta-Reinforcement Learning via Probabilistic Context Variables

Kate Rakelly*, Aurick Zhou*, Deirdre Quillen, Chelsea Finn, Sergey Levine
“Hula Beach”, “Never grow up”, “The Sled” - by artist Matt Spangler, mattspangler.com
Meta-Reinforcement Learning

**meta-testing**

Given a small amount of experience... Learn to solve the task!

**meta-training**

By learning to solve other related tasks
Meta-Reinforcement Learning

**meta-testing**
Given a small amount of experience... Learn to solve the task!

**meta-training**
By learning to solve other related tasks

requires data from each task, exacerbates sample inefficiency of RL
Meta-RL Experimental Domains

- **Half Cheetah**
- **Humanoid**
- **Ant**
- **Walker**

Variable reward function (locomotion direction, velocity, or goal)

Variable dynamics (joint parameters)

Simulated via MuJoCo (Todorov et al. 2012), tasks proposed by (Finn et al. 2017, Rothfuss et al. 2019)
ProMP (Rothfuss et al. 2019), MAML (Finn et al. 2017), RL2 (Duan et al. 2016)
ProMP (Rothfuss et al. 2019), MAML (Finn et al. 2017), RL2 (Duan et al. 2016)

20-100X more sample efficient!
Disentangle task inference from control
Off-Policy Meta-Training

Diagram illustrating the process of off-policy meta-training, including components such as a replay buffer, state representation $s_c$, function approximator $\phi$, and loss functions $L_{critic}$ and $L_{actor}$. The diagram also shows how tasks are used for training.
Efficient exploration by posterior sampling
Posterior sampling in action
Takeaways

PEARL
- First off-policy meta-RL algorithm
- 20-100X improved sample efficiency on the domains tested, often substantially better final returns
- Probabilistic belief over the task enables posterior sampling for efficient exploration

arXiv: arxiv.org/abs/1903.08254v1

GitHub: github.com/katerakelly/oyster

Come talk to us tonight at Poster 40!