CEIP: Combining Explicit and Implicit Priors for Reinforcement Learning with Demonstrations

Kai Yan, Alexander G. Schwing, Yu-Xiong Wang

https://289371298.github.io/jekyll/update/2022/10/25/CEIP

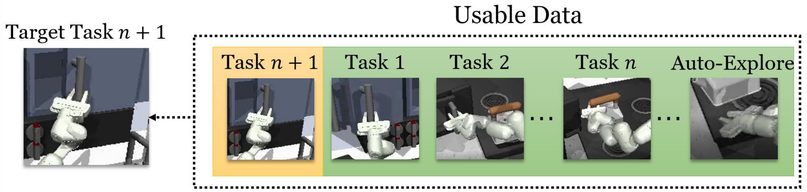
Motivation

Goal: Improve RL sample efficiency with demonstrations

- Two types of demonstrations available in real life

A) Task-specific: identical to target, expensive, few, need to be collected for every new task

B) Task-agnostic: less related, cheap, many, accumulated from existing tasks / auto exploration



- Two ways to use demonstrations

A) Implicit Prior: deep net encodes action; expressive

B) Explicit Prior: database retrieves action; structured

s: states
a: action

Examples | Implicit Prior: Normalizing Flow | Explicit Prior: Database

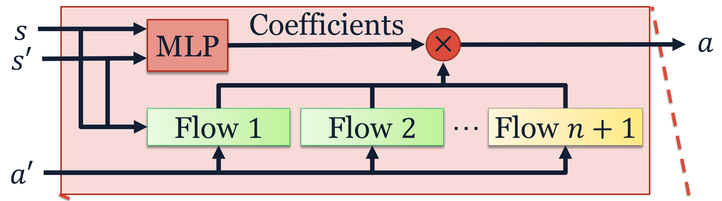
- We use **both types** of demonstrations in **both ways**

Contribution

- A novel architecture that introduces explicit prior into RL community, combines explicit and implicit prior, and reaches state-of-the-art
- A new type of normalizing flow mixture, and the first to use flow mixture in RL / pioneering work for flow mixture application

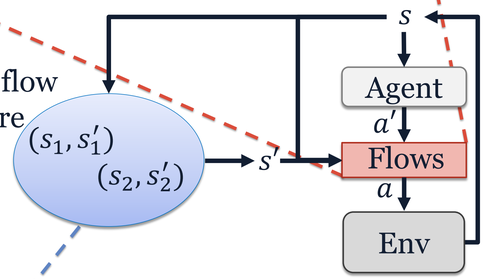
Implicit Prior

- Linear mixture of $n+1$ single-layer normalizing flows, each trained on one task in task-agnostic data (Task $1 \sim n$) and one trained on task-specific data (Task $n+1$)



Combination

- Step 1: Train single flow
- Step 2: Train mixture
- Step 3: RL finetune



Explicit Prior

- $s' = s'_i, i = \text{argmin}_x \|s_x - s\|_2^2 + C(s'_x)$, where $C(s'_x)$ is a penalty to push the agent forward along the path
- $C(s'_x) > 0$ if this (s, a) pair or another (s, a) pair later in the trajectory has been retrieved in this episode
- Enhances the input of each flow by predicting future state following expert trajectory

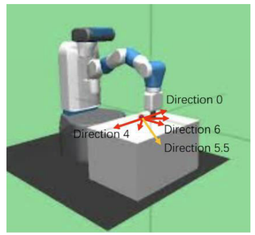
Results

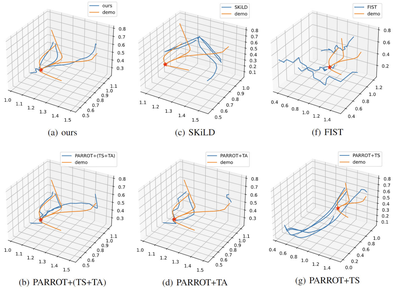
- We test our CEIP on 3 robot arm manipulation testbeds: fetchreach, kitchen and office and compare to 3 state-of-the-art methods: FIST, PARROT and SKiLD

Reward (higher is better; see paper for details)

Env Name	CEIP (ours)	PARROT	FIST	SKiLD
Fetchreach	**-10.03**$\pm$**0.64**	-20.30$\pm$10.62	-34.80$\pm$8.33	-39.91$\pm$0.14
Kitchen	**3.94**$\pm$**0.07**	2.27$\pm$0.24	0.53$\pm$0.50	1.67$\pm$0.58
Office	**6.33**$\pm$**0.30**	1.97$\pm$0.22	5.50$\pm$1.12	0.50$\pm$0.50

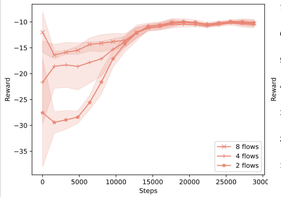
- CEIP on fetchreach yields the best trajectories





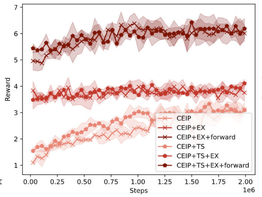
Left: Fetchreach Env.; **right:** trajectories (red: goal, orange: expert, blue: agent)

- More flows, better performance



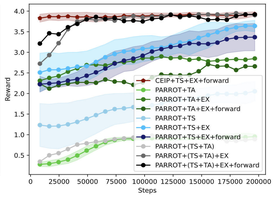
CEIP on Fetchreach with different #flows

- Explicit prior is very important for CEIP



Ablation of CEIP on Office
EX: explicit prior
Forward: push-forward technique (see paper)
TS: flow $n+1$

- Explicit prior improves PARROT



Ablation of PARROT on kitchen
EX: explicit prior
TA/TS: use task-agnostic/task-specific dataset to train