[ICLR'24] ASID: ACTIVE EXPLORATION FOR SYSTEM IDENTIFICATION IN ROBOTIC MANIPULATION

1. Link: https://weirdlabuw.github.io/asid/
2. Arthurs and institution: Marius Memmel, Andrew Wagenmaker, Chuning Zhu, Patrick Yin, Dieter Fox, Abhishek Gupta from University of Washington

NOTE:

1. Dieter Fox, Abhishek Gupta are 2 big names in robotics
2. Marius Memmel is one of the authors in URDFormer and HAMSTER
3. Patrick Yin is doing system identification. **TL;DR** A learning system that can leverage a small amount of real-world data to autonomously refine a simulation model and then plan an accurate control strategy that can be deployed in the real world. **Todos**
4. read the related works and their further works.

Thoughts and critisims

1. can we use a unified representation of \theta, which
    1. contains as much parameter of simulator
    2. contains as much information in the scene, not just an object
2. use gradient from simulator instead of the non-linear approximation

Related works

System Identification

1. theoretical (dynamic system, control theory):
    1. how to learn a model of the system dynamics efficiently
    2. how inputs to a system should be chosen to most effectively learn the system’s parameters
    3. how to deal with partial observability
    4. choose their inputs to maximize some function of the Fisher information matrix
2. real-world application
    1. active identification of physics parameter
    2. kinematic structure

3. learn the parameters of the simulator to ultimately train a downstream policy on the learned parameters, and therefore apply task-specific policies for data collection
4. exploration policies that minimize its regret

Sim2Real

1. Domain randomization
   1. change the distribution dynamically
   2. add real-world data
2. train a policy in sim and then fine-tune in the real environmen

Model-based RL

1. Definition: solve the RL problem by learning a model of the dynamics, and using this model to either plan or solve a policy optimization problem
2. most of the works focus on fully learned dynamic models

Problem formulation

1. Under MDP setting
2. the dynamics of the real environment belong to some known parametric family $\mathcal{P} = \{\mathcal{P}_\theta : \theta \in \Theta\}$, θ is unkown (mass, friction), might know the dynamics
3. The goal:
   1. 1earn as much useful information as possible about the real environment from a single episode of interaction
   2. use this information to obtain a policy that can solve the task in real as effectively as possible.

Fisher information

$$\mathcal{I}(\boldsymbol{\theta}) := \mathbb{E}_{\boldsymbol{\tau} \sim p_{\boldsymbol{\theta}}} \left[\nabla_{\boldsymbol{\theta}} \log p_{\boldsymbol{\theta}}(\boldsymbol{\tau}) \cdot \nabla_{\boldsymbol{\theta}} \log p_{\boldsymbol{\theta}}(\boldsymbol{\tau})^\top \right].$$

1. The Fisher information thus serves as a fundamental lower bound on parameter estimation error

Pázman (2013)) states that, under certain regularity conditions, the covariance of $\widehat{\boldsymbol{\theta}}(\mathfrak{D})$ satisfies:

$$\mathbb{E}_{\mathfrak{D} \sim p_{\boldsymbol{\theta}^\star}}[(\widehat{\boldsymbol{\theta}}(\mathfrak{D}) - \boldsymbol{\theta}^\star)(\widehat{\boldsymbol{\theta}}(\mathfrak{D}) - \boldsymbol{\theta}^\star)^\top] \succeq T^{-1} \cdot \mathcal{I}(\boldsymbol{\theta}^\star)^{-1}.$$

From this it follows that the Fisher information serves as a lower bound on the mean-squared error:

$$\mathbb{E}_{\mathfrak{D} \sim p_{\boldsymbol{\theta}^\star}}[\|\widehat{\boldsymbol{\theta}}(\mathfrak{D}) - \boldsymbol{\theta}^\star\|_2^2] = \mathrm{tr}(\mathbb{E}_{\mathfrak{D} \sim p_{\boldsymbol{\theta}^\star}}[(\widehat{\boldsymbol{\theta}}(\mathfrak{D}) - \boldsymbol{\theta}^\star)(\widehat{\boldsymbol{\theta}}(\mathfrak{D}) - \boldsymbol{\theta}^\star)^\top]) \geq T^{-1} \cdot \mathrm{tr}(\mathcal{I}(\boldsymbol{\theta}^\star)^{-1}). \quad (1)$$

Contributions

1. provide a real-world manipulation problem implementation of using fisher-information as learning signal for exploration policy with a non-linear approximation
2. provide details experiment on questions of interest, shows the algorithm outperforms selected baselines.

Key concepts

EXPLORATION VIA FISHER INFORMATION MAXIMIZATION

1. Goal: play an exploration policy which generates a trajectory on the real environment that provides as much information as possible

$$\arg\min_{\pi} \operatorname{tr}(\mathcal{I}(\boldsymbol{\theta}^{\star}, \pi)^{-1}).$$

2.
3. assumption to non-linear approximation

this, we make a simplifying assumption on the dynamics, that our next state, s_{h+1}, evolves as:

$$s_{h+1} = f_{\boldsymbol{\theta}}(s_h, a_h) + w_h, \tag{3}$$

where s_h and a_h are the current state and action, $w_h \sim \mathcal{N}(0, \sigma_w^2 \cdot I)$ is Gaussian process noise, and $f_{\boldsymbol{\theta}}$ are the nominal dynamics. Under these dynamics, the Fisher information matrix reduces to

4.

$$\mathcal{I}(\boldsymbol{\theta}, \pi) = \sigma_w^{-2} \cdot \mathbb{E}_{p_{\boldsymbol{\theta}}(\cdot|\pi)} \left[\sum_{h=1}^{H} \nabla_{\boldsymbol{\theta}} f_{\boldsymbol{\theta}}(s_h, a_h) \cdot \nabla_{\boldsymbol{\theta}} f_{\boldsymbol{\theta}}(s_h, a_h)^{\top} \right].$$

SYSTEM IDENTIFICATION

1. minimizing

$$\mathbb{E}_{\boldsymbol{\theta} \sim q_{\boldsymbol{\phi}}} [\mathbb{E}_{\boldsymbol{\tau}_{\mathrm{sim}} \sim p_{\boldsymbol{\theta}}(\cdot | \mathcal{A}(\boldsymbol{\tau}_{\mathrm{real}}))} [||\boldsymbol{\tau}_{\mathrm{real}} - \boldsymbol{\tau}_{\mathrm{sim}}||_2^2]]$$

SOLVING THE DOWNSTREAM TASK

run any rl mased learning policy in a simulator with identified environment.

Implementation details

Hardware

1. Franka Emika Panda robot for exploration and task performance in the real world

2. 2 D435

Software

Training

Experiments

Exams

1. sphere manipulation
2. articulation
3. rod balancing

Questions of interest

1. DOES ASID LEARN EFFECTIVE EXPLORATION BEHAVIOR

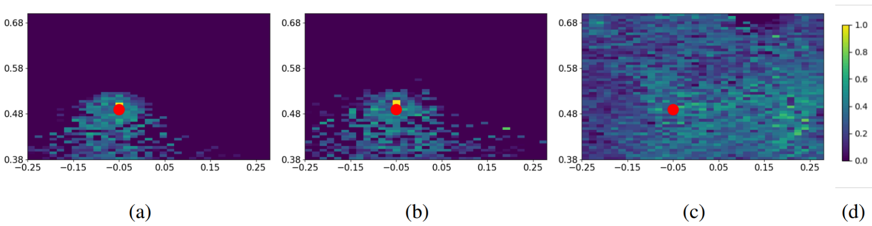

Figure 7: **Absolute sphere displacement** for different sphere starting locations. Zero means the sphere didn't get hit, higher numbers denote larger displacements. Initial endeffector position marked in red ■. a) Random Coverage, b) PPO Coverage, c) Fisher Coverage, d) Legend.

2. HOW DOES ASID PERFORM QUANTITATIVELY IN SIMULATION ON DOWNSTREAM TASKS?
3. DOES ASID ALLOW FOR REAL-WORLD CONTROLLER SYNTHESIS USING MINIMAL REAL-WORLD DATA?

Task	**Rod Balancing**			**Sphere Striking**
Metric	Tilt angle in degree° ↓			Success Rate in % ↑
Parameter	Inertia (left)	Inertia (middle)	Inertia (right)	Friction $\sim [1.1, 1.5]$
Random exploration	12.44 ± 19.6	4.20 ± 6.5	15.34 ± 15.9	10.62 ± 4.3
Kumar et al. (2019)	13.70 ± 9.3	2.82 ± 2.7	15.26 ± 9.8	9.50 ± 2.4
DR	26.69 ± 7.0	13.05 ± 7.3	1.13 ± 1.3	8.75 ± 1.5
ASID + estimator	17.73 ± 13.1	4.65 ± 5.1	9.99 ± 6.8	11.00 ± 5.2
ASID + SysID (ours)	$\mathbf{0.00 \pm 0.0}$	$\mathbf{0.72 \pm 1.0}$	$\mathbf{0.00 \pm 0.0}$	$\mathbf{28.00 \pm 9.7}$