DART

Noise Injection for Robust Imitation Learning

Laskey et al. · CoRL 2017 · arXiv:1703.09327

The problem

- Behavior Cloning: errors compound
- Robot drifts to states never shown
- DAgger: robot drives, supervisor labels
- ...but tedious, slow, can be unsafe

so

The DART loop

Supervisor demos, noise in its controls

↓

Train on its clean controls

↓

Fit Σ to robot's error, scale to α

↓

Repeat K rounds on all data

Sizing the noise

- Robot's errors on noisy demo states
- Σ = their mean outer product
- Scale Σ so $T \cdot tr(\Sigma) = \alpha$
- α: guess of the final robot's error

The theory

- Shift ≤ $T \cdot \sqrt{(\frac{1}{2} \cdot KL)}$, loss in [0, 1]
- No noise: KL = ∞ once robot errs
- Full-rank noise: KL finite (Prop. 4.1)
- Assumes noise doesn't change supervisor

4 / 4

MuJoCo tasks at DAgger's level

3x

faster compute than DAgger (Humanoid)

HSR grasping in clutter



target found + path clear · 4 supervisors · 40 demos

DAgger vs DART

DAgger (on-policy)

- Robot drives (β = 0.5 mix)
- Supervisor labels its states
- Retrains after every demo
- >80% reward lost (Humanoid)

vs

DART (off-policy)

- Supervisor drives, with noise
- Labels while demonstrating
- Retrains rarely, to update Σ
- 5% reward lost (Humanoid)

Takeaway

- Noise sized to the robot's error gives corrective data, supervisor in control
- Matches DAgger in MuJoCo; 79% vs 49% on the HSR
- Limits: α is a guess; too much noise may disturb humans

bushuyeu.com/read-along-dart · paper arXiv default license · sketchnote by Pavel Bushuyeu

Active listening · DART

Write while you listen. One line per box is enough.

Laskey et al. · CoRL 2017 · arXiv:1703.09327

Date _____ Presenter _____ Role _____

Their main claim, in my words

The evidence they show

What I would push on

My question for the room

How it connects to my work

Listen for

- ☐ Where is the noise added, and what is the label?
- ☐ Who controls the robot while data is collected?
- ☐ How the noise is fit to the learner's error (α)
- ☐ What Lemma 4.1 and Prop. 4.1 assume and compare
- ☐ Which results use humans, which a TRPO supervisor?

Ask the room

1. How much noise can a human teleoperator absorb?
2. State-dependent errors vs one global Σ: how to inject?
3. Pretrained policy's error: enough to skip rounds?

Page, video and FAQ:
bushuyeu.com/read-along-dart