DAgger

A Reduction of Imitation Learning and Structured Prediction to No-Regret Online Learning

Ross et al. · AISTATS 2011 · arXiv:1011.0686

The problem

- Cloning trains on the expert's states
- One mistake can lead to unseen states
- Errors can compound: up to $T^2\varepsilon$
- Earlier fixes: T policies or a random mix

so →

The DAgger loop

Run the current policy (expert in round 1) → Expert labels every state visited → Add the labels to one dataset D → Retrain on all of D, then repeat

Extra cost over T steps

Behavior cloning (Thm 2.1)

- $J \le J(\pi^*) + T^2\varepsilon$ (tight)
- ε measured on the expert's states

vs

DAgger (Thm 3.2)

- $J \le J(\pi^*) + uT\varepsilon + O(1)$
- ε: best policy's loss on all rollouts

if →

The fine print

- u small: the expert recovers in a few steps
- Bounded, strongly convex loss
- Infinite data, β decays geometrically: $N = \tilde{O}(uT)$
- Some policy in the run is good, maybe not the last

0

Kart falls per lap after 15 iterations (SMILe: about 2 after 20)

Super Mario Bros.

- Planner expert, 4 buttons
- Beats SMILe and SEARN
- Cloning: stuck at obstacles

85.5%

handwriting accuracy (supervised 83.6%, no structure 82%)

Takeaway

- Train on the states your own policy reaches
- Linear in T when $u = O(1)$: the expert recovers
- Limit: the expert must label any state reached

Active listening · DAgger

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

Ross et al. · AISTATS 2011 · arXiv:1011.0686

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 each ε measured: expert's states or learner's?
- ☐ What is u? When does the T bound turn back into T^2?
- ☐ After round 1, does the expert act or only label?
- ☐ What does β control? Which setting wins in the tests?
- ☐ Which policy do the theorems cover: last, best, random?

Ask the room

1. Labeling states the expert never visits: when is it hard?
2. Bad early policies on a real robot: how to run it safely?
3. Deep nets are not FTL on a convex loss: what survives?

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