Imitating Latent Policies from Observation

Ashley D. Edwards, Himanshu Sahni, Yannick Schroecker, Charles L. Isbell
Georgia Institute of Technology
Introduction

- Imitation from Observation enables learning from state sequences
- Typical approaches need extensive environment interactions
- Humans can learn policies just by watching
**Given:** Sequence of noisy expert observations

**Assumption:** Discrete actions with deterministic transitions

- $z$ is defined as a *latent* action that caused a transition to occur
- $z$ can imply a real action or some other type of transition

- A *latent* policy is the probability of taking a latent action in some state
ILPO

1. Given sequence of observations, learn \textit{latent} policy
2. Use a few environment steps to align actions
ILPO

1. Given sequence of observations, learn latent policy
2. Use a few environment steps to align actions
**ILPO**

1. Given sequence of observations, learn *latent* policy

2. Use a few environment steps to align actions

![Diagram of ILPO approach](image)
Experiments: Classic Control

- Access to expert observations only
- No reward function used in approach
- Comparison to Behavioral Cloning from Observation [1]

Experiments: CoinRun

(a) CoinRun easy
Experiments: CoinRun

(b) CoinRun hard
Thank You!

Room: Pacific Ballroom at 6:30pm (Today)!
Poster: #33