Imitating Latent Policies from Observation

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




Introduction

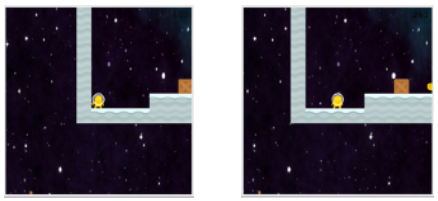
- Imitation from Observation enables learning from state sequences
- Typical approaches need extensive environment interactions
- Humans can learn policies just by watching

Approach

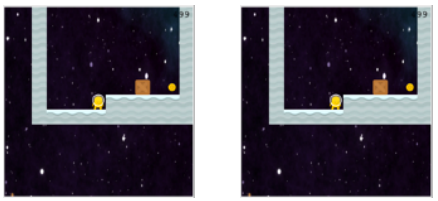
Given: Sequence of noisy expert observations

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



Action: Right
Z = 1



Action: Right
Z = 2

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

Approach

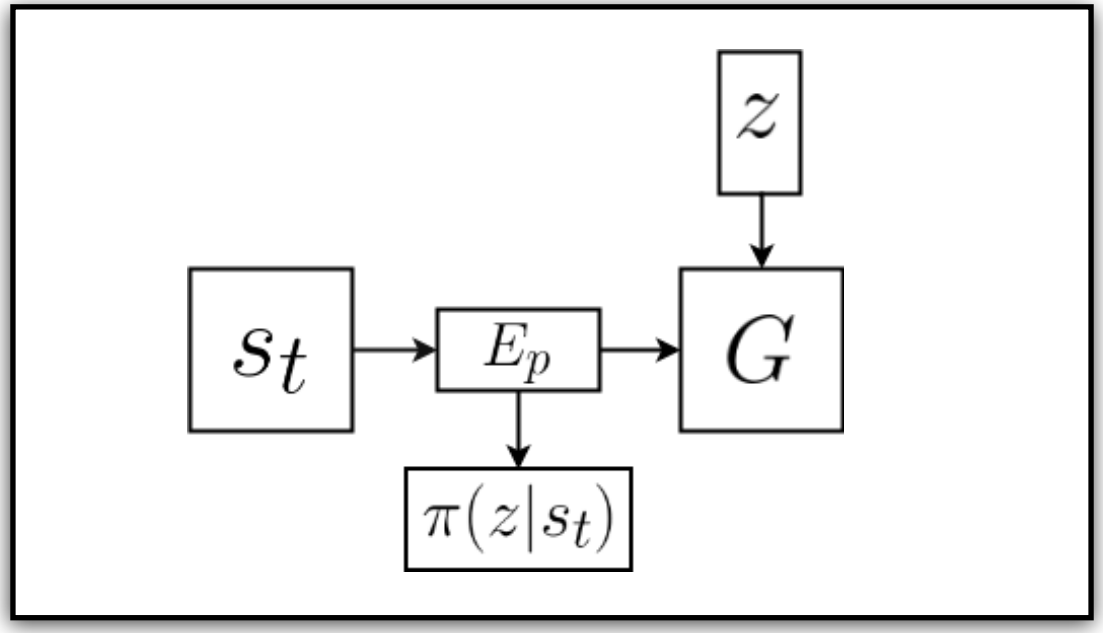
ILPO

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

Approach

ILPO

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

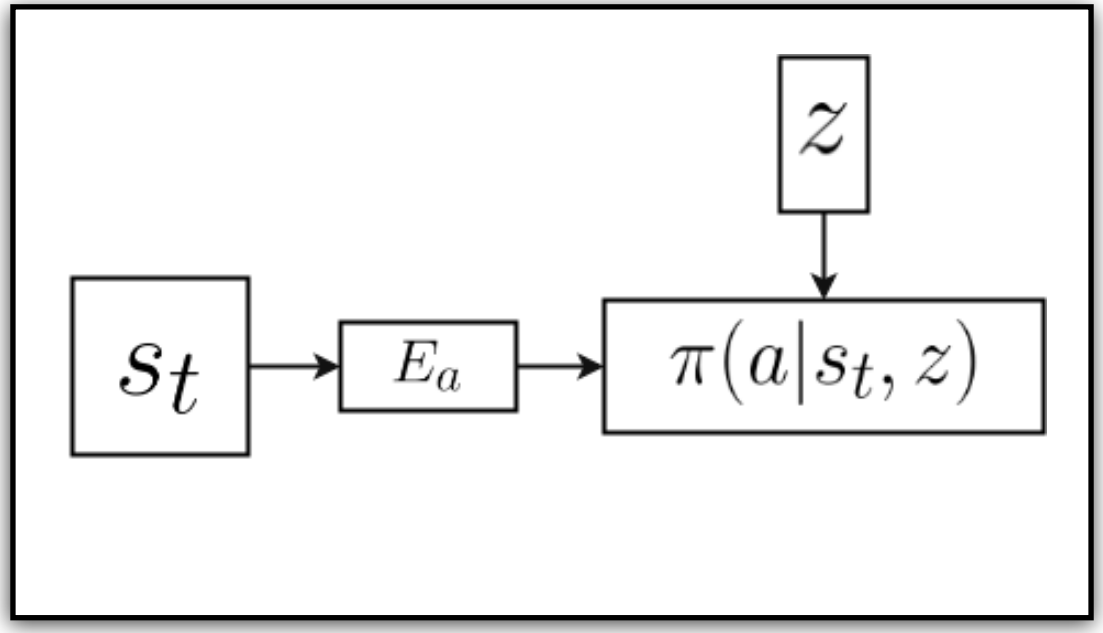


Latent policy network

Approach

ILPO

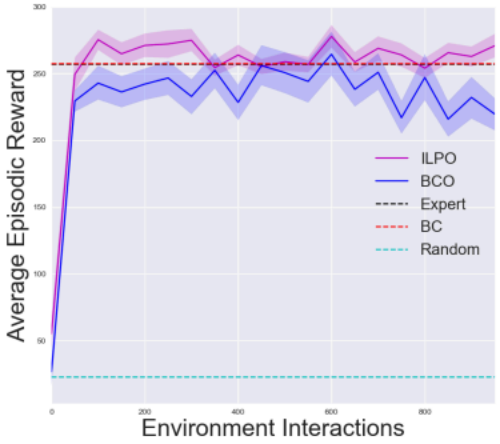
1. Given sequence of observations, learn *latent* policy
2. **Use a few environment steps to align actions**



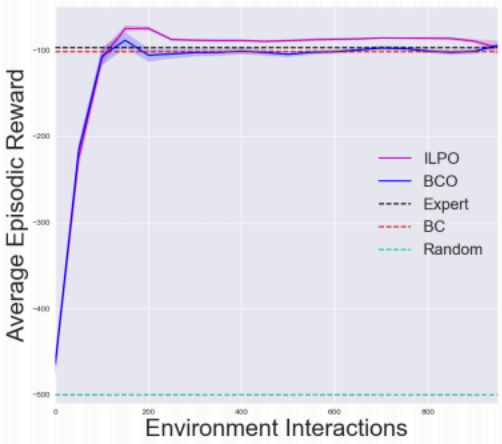
Action remapping network

Experiments: Classic Control

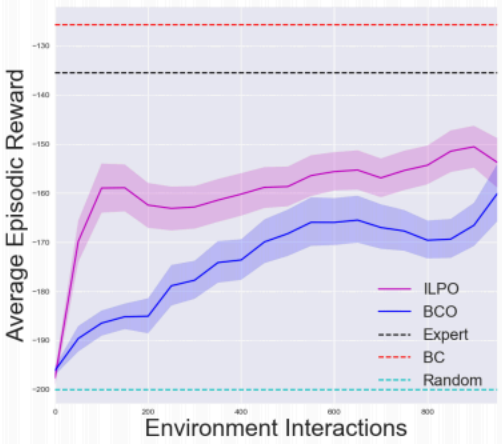
- Access to expert observations only
- No reward function used in approach
- Comparison to Behavioral Cloning from Observation [1]



(a) Cartpole



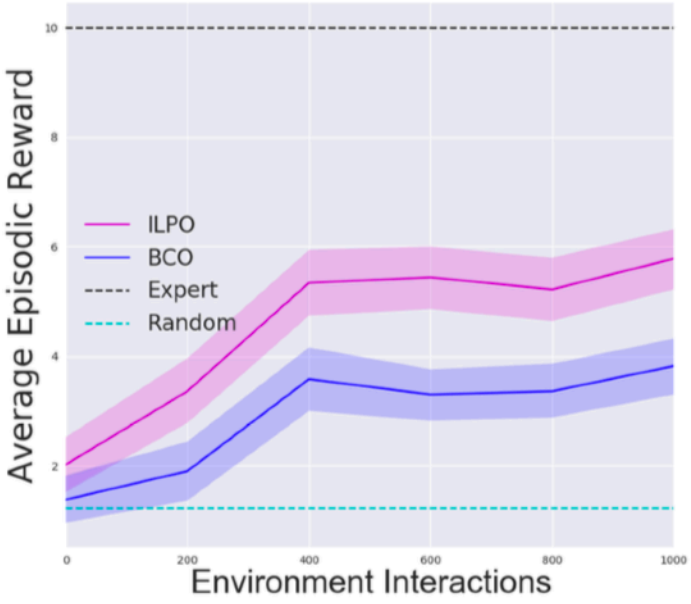
(b) Acrobot



(c) Mountain car

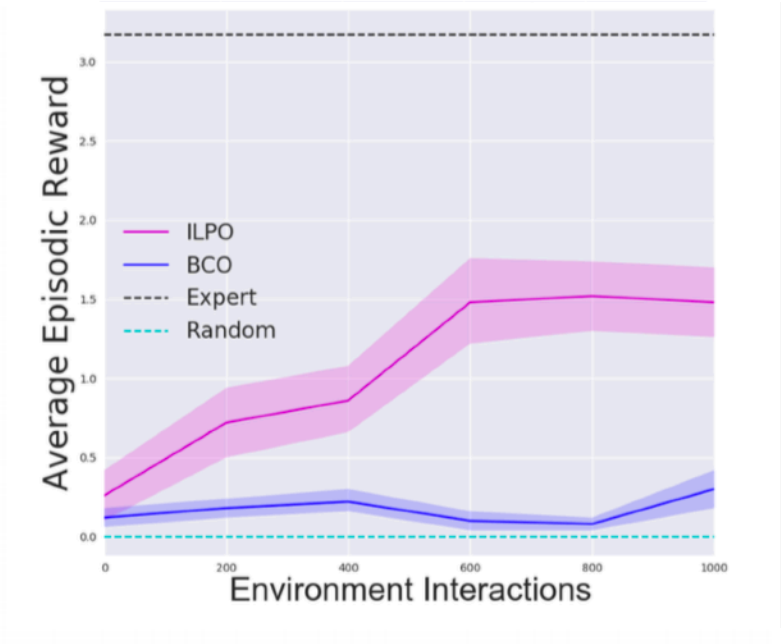
[1] Torabi, Faraz, Garrett Warnell, and Peter Stone. "Behavioral cloning from observation." *Proceedings of the 27th International Joint Conference on Artificial Intelligence*. AAAI Press, 2018.

Experiments: CoinRun



(a) CoinRun easy

Experiments: CoinRun



(b) CoinRun hard

Thank You!

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