Robot Skill Learning: Real World to Sim and Back

- Motivation: RoboCup
- Sim2Real: Grounded Simulation Learning
- **Imitation Learning from Observation:**
  - Model-based approach: BCO
  - Model-free approach: GAIfO

Faraz Torabi

Garrett Warnell

Imitation Learning

Goal:

- Learn how to make decisions by trying to imitate another agent.

Imitation Learning

Goal:

- Learn how to make decisions by trying to imitate another agent.

Conventional Imitation Learning:

- Observations of other agent (demonstrations) consist of state-action pairs.[1]

[1] Niekum et al., "Learning and generalization of complex tasks from unstructured demonstrations".

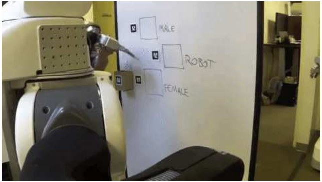
Imitation Learning

Goal:

- Learn how to make decisions by trying to imitate another agent.

Conventional Imitation Learning:

- Observations of other agent (demonstrations) consist of state-action pairs.[1]



Imitation Learning

Goal:

- Learn how to make decisions by trying to imitate another agent.

Conventional Imitation Learning:

- Observations of other agent (demonstrations) consist of state-action pairs.[1]

Challenge:

- Precludes using a large amount of demonstration data where action sequences are not given (e.g. YouTube videos).

[1] Niekum et al., "Learning and generalization of complex tasks from unstructured demonstrations".

Imitation Learning

Algorithms:

Imitation Learning

Algorithms:

- Behavioral Cloning:

Imitation Learning

Algorithms:

- Behavioral Cloning:
  - End to End Learning for Self-Driving Cars.[2]

[2] Zhang and Cho, “Query-Efficient Imitation Learning for End-to-End Simulated Driving.”

Imitation Learning

Algorithms:

- Behavioral Cloning:
  - End to End Learning for Self-Driving Cars.[2]
- Inverse Reinforcement Learning:

[2] Zhang and Cho, "Query-Efficient Imitation Learning for End-to-End Simulated Driving."

Imitation Learning

Algorithms:

- Behavioral Cloning:
  - End to End Learning for Self-Driving Cars.[2]
- Inverse Reinforcement Learning:
  - Generative Adversarial Imitation Learning.[3]
  - Guided Cost Learning.[4]

[2] Zhang and Cho, "Query-Efficient Imitation Learning for End-to-End Simulated Driving."

[3] Ho and Ermon, "Generative adversarial imitation learning".

[4] Finn, Levine, and Abbeel, "Guided cost learning: Deep inverse optimal control via policy optimization".

Imitation from Observation

Goal:

- Learn how to perform a task given state-only demonstrations.

Imitation from Observation

Goal:

- Learn how to perform a task given state-only demonstrations.

Imitation from Observation

Goal:

- Learn how to perform a task given state-only demonstrations.

Formulation:

- Given:
  - $D_{demo} = (s_0, s_1, ...)$
- Learn:
  - $\pi : \mathcal{S} \rightarrow \mathcal{A}$

Imitation from Observation

Previous work:

Imitation from Observation

Previous work:

- Time Contrastive Networks (TCN).[5]
- Imitation from observation: Learning to imitate behaviors from raw video via context translation.[6]
- Learning invariant feature spaces to transfer skills with reinforcement learning.[7]

[5] Sermanet et al., "Time-contrastive networks: Self-supervised learning from multi-view observation".

[6] Liu et al., "Imitation from observation: Learning to imitate behaviors from raw video via context translation".

[7] Gupta et al., "Learning invariant feature spaces to transfer skills with reinforcement learning".

Imitation from Observation

Previous work:

- Time Contrastive Networks (TCN).[5]
- Imitation from observation: Learning to imitate behaviors from raw video via context translation.[6]
- Learning invariant feature spaces to transfer skills with reinforcement learning.[7]

Concentrate on perception; require time-aligned demonstrations.

[5] Sermanet et al., “Time-contrastive networks: Self-supervised learning from multi-view observation”.

[6] Liu et al., “Imitation from observation: Learning to imitate behaviors from raw video via context translation”.

[7] Gupta et al., “Learning invariant feature spaces to transfer skills with reinforcement learning”.

Efficient Robot Skill Learning

- Motivation: RoboCup
- Sim2Real: Grounded Simulation Learning
- Imitation Learning from Observation:
  - **Model-based approach:** BCO
  - Model-free approach: GAIfO

Model-based Approach

- Imitation Learning: $D_{demo} = \{(s_0, a_0), (s_1, a_1), ...\}$

Model-based Approach

- Imitation Learning: $D_{demo} = \{(s_0, a_0), (s_1, a_1), ...\}$
- Imitation from Observation: $D_{demo} = \{(s_0, ?), (s_1, ?), ...\}$

Model-based Approach

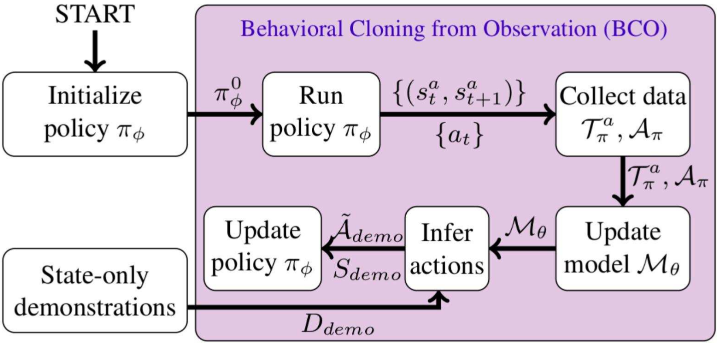
- Imitation Learning: $D_{demo} = \{(s_0, a_0), (s_1, a_1), ...\}$
- Imitation from Observation: $D_{demo} = \{(s_0, ?), (s_1, ?), ...\}$

Model-based Approach:

Learn an inverse dynamics model → Infer actions → Perform a conventional IL method

Behavioral Cloning from Observation (BCO)

Algorithm:

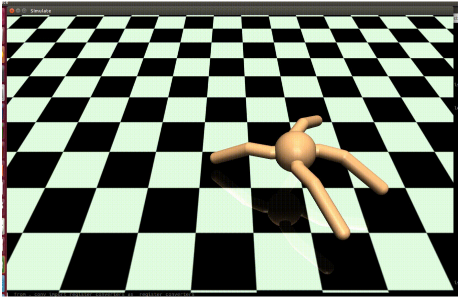


Torabi, Warnell, and Stone, IJCAI 2018

Behavioral Cloning from Observation (BCO)

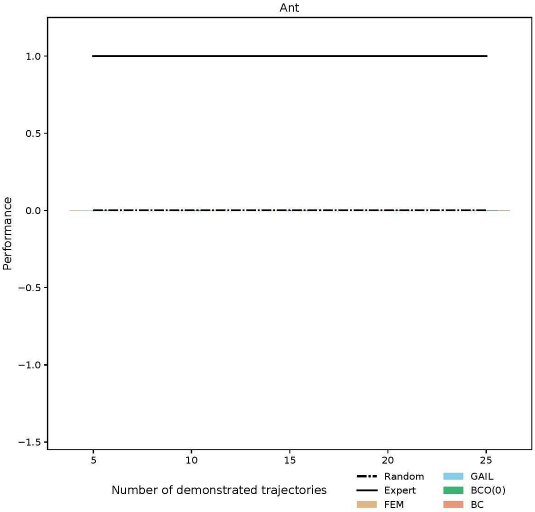
Experimental Results:

- Domain:
  - Mujoco domain "Ant" with 111 dimensional state space and 8 dimensional action space.



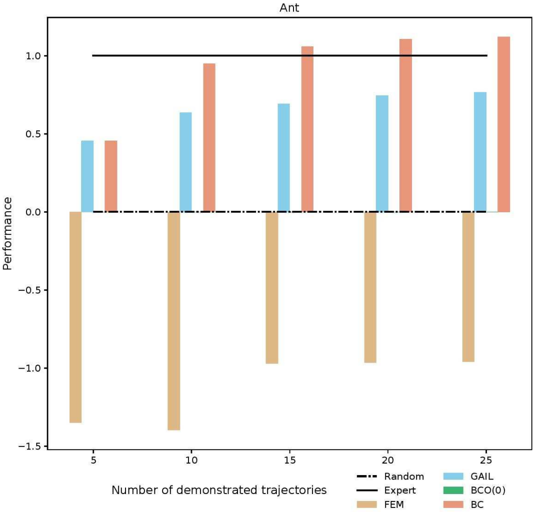
Behavioral Cloning from Observation (BCO)

Experimental Results:



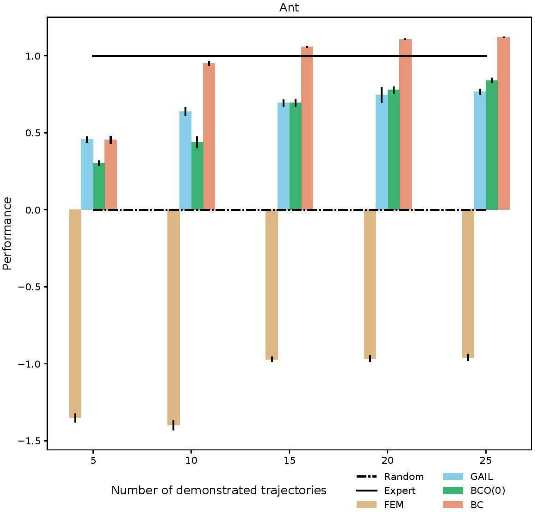
Behavioral Cloning from Observation (BCO)

Experimental Results:



Behavioral Cloning from Observation (BCO)

Experimental Results:



Behavioral Cloning from Observation (BCO(α))

Issue:

- Inverse dynamics model is learned using a random policy.

Behavioral Cloning from Observation (BCO(α))

Issue:

- Inverse dynamics model is learned using a random policy.

Solution: BCO(α)

Behavioral Cloning from Observation (BCO(α))

Issue:

- Inverse dynamics model is learned using a random policy.

Solution: BCO(α)

- Update the model with the learned policy.

Behavioral Cloning from Observation (BCO(α))

Issue:

- Inverse dynamics model is learned using a random policy.

Solution: BCO(α)

- Update the model with the learned policy.
- Parameter α controls tradeoff between performance and environment interactions

Behavioral Cloning from Observation (BCO(α))

Issue:

- Inverse dynamics model is learned using a random policy.

Solution: BCO(α)

- Update the model with the learned policy.
- Parameter α controls tradeoff between performance and environment interactions
  - $\alpha = 0$: no post-demonstration interaction.

Behavioral Cloning from Observation (BCO(α))

Issue:

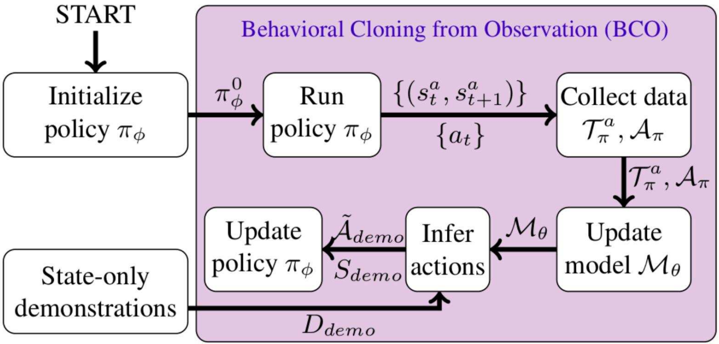
- Inverse dynamics model is learned using a random policy.

Solution: BCO(α)

- Update the model with the learned policy.
- Parameter α controls tradeoff between performance and environment interactions
  - $\alpha = 0$: no post-demonstration interaction.
  - Increasing α: increasing the number of interactions allowed at each iteration.

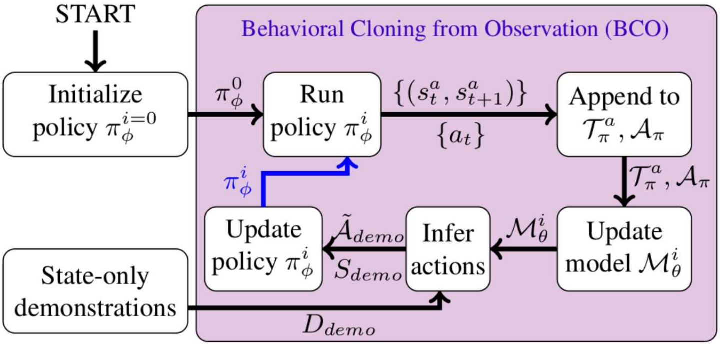
Behavioral Cloning from Observation (BCO(α))

Algorithm:



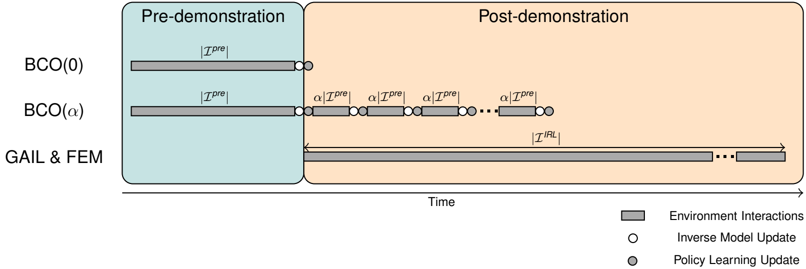
Behavioral Cloning from Observation (BCO(α))

Algorithm:



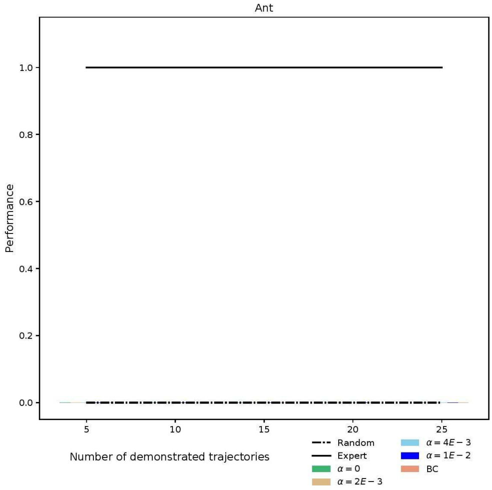
Behavioral Cloning from Observation (BCO(α))

Interaction time:



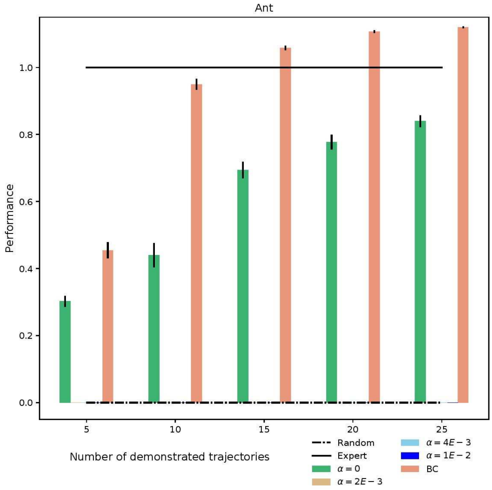
Behavioral Cloning from Observation (BCO(α))

Effect of varying α on BCO(α):



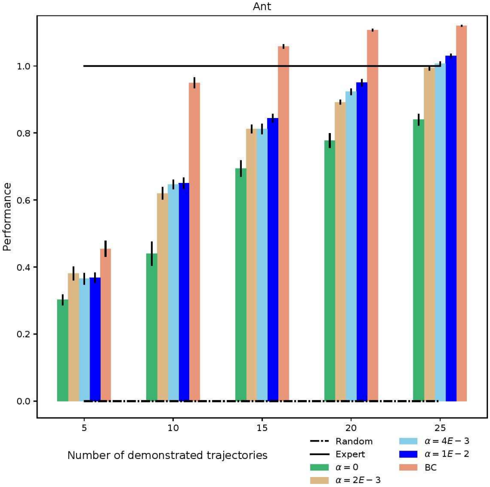
Behavioral Cloning from Observation (BCO(α))

Effect of varying α on BCO(α):



Behavioral Cloning from Observation (BCO(α))

Effect of varying α on BCO(α):

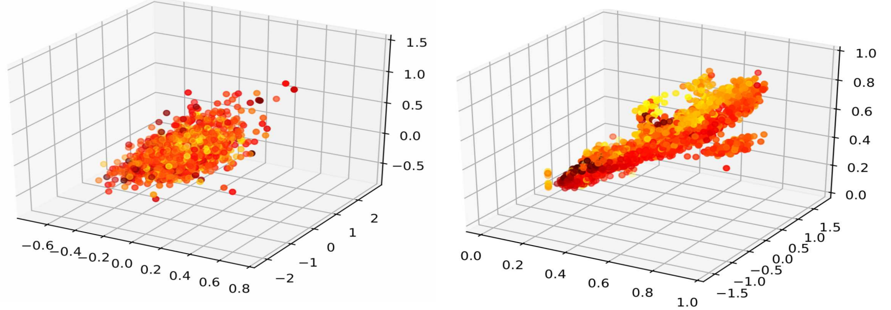


Efficient Robot Skill Learning

- Motivation: RoboCup
- Sim2Real: Grounded Simulation Learning
- Imitation Learning from Observation:
  - Model-based approach: BCO
  - **Model-free approach**: GAIfO

Gen. Adversarial Imitation from Observation (GAIfO)

Motivation:



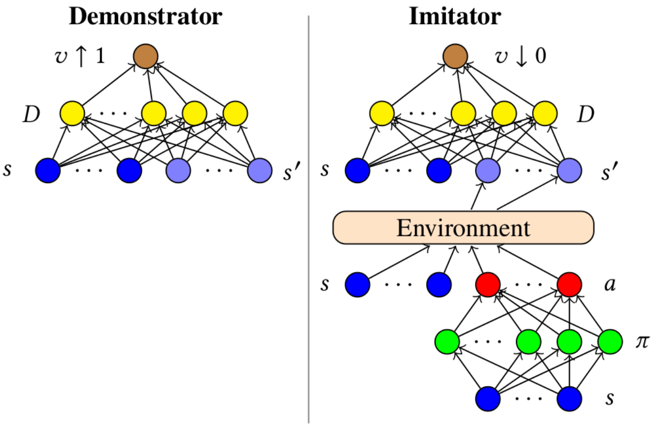
(a) Random Policy

(b) Demonstration

Figure: State transition distribution in Hopper domain.

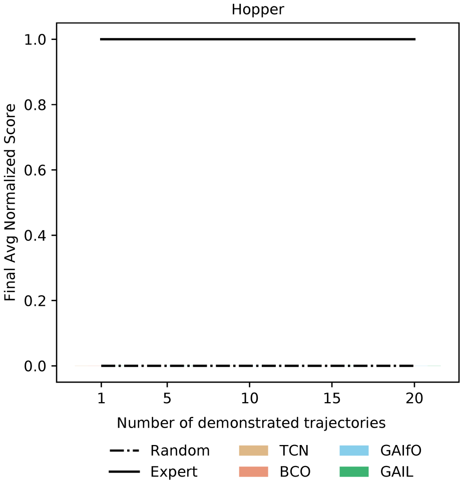
Gen. Adversarial Imitation from Observation (GAIfO)

Algorithm:



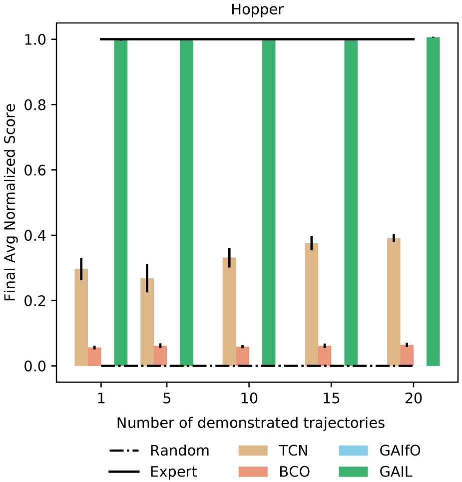
Gen. Adversarial Imitation from Observation (GAIfO)

Comparison against other IfO approaches and GAIL:



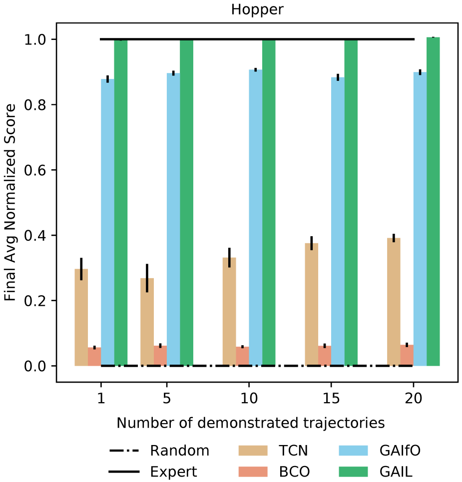
Gen. Adversarial Imitation from Observation (GAIfO)

Comparison against other IfO approaches and GAIL:



Gen. Adversarial Imitation from Observation (GAIfO)

Comparison against other IfO approaches and GAIL:



Gen. Adversarial Imitation from Observation (GAIfO)

Challenges:

Gen. Adversarial Imitation from Observation (GAIfO)

Challenges:

- States are not fully-observable.

Gen. Adversarial Imitation from Observation (GAIfO)

Challenges:

- States are not fully-observable.
- States are not Markovian.

Gen. Adversarial Imitation from Observation (GAIfO)

Challenges:

- States are not fully-observable.
- States are not Markovian.

Solution:

Gen. Adversarial Imitation from Observation (GAIfO)

Challenges:

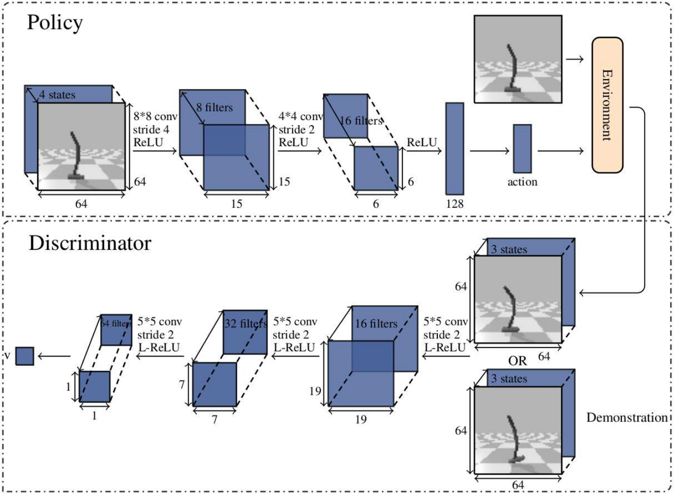
- States are not fully-observable.
- States are not Markovian.

Solution:

- Stack four frames.

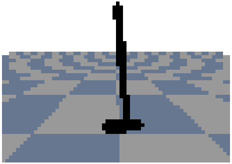
Gen. Adversarial Imitation from Observation (GAIfO)

Algorithm:



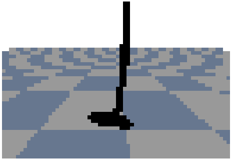
Gen. Adversarial Imitation from Observation (GAIfO)

Demonstration:



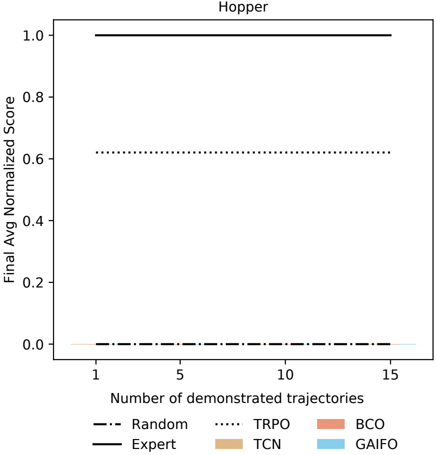
Gen. Adversarial Imitation from Observation (GAIfO)

Learned Policy:



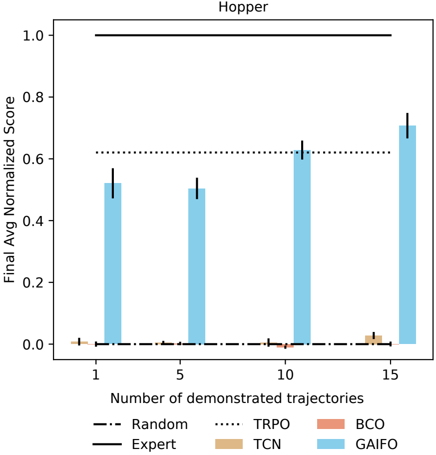
Gen. Adversarial Imitation from Observation (GAIfO)

Comparison against other IfO approaches:



Gen. Adversarial Imitation from Observation (GAIfO)

Comparison against other IfO approaches:



Ongoing Work

Ongoing Work

- Testing algorithms on more domains.

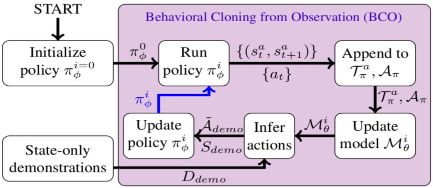
Ongoing Work

- Testing algorithms on more domains.
- Adapt algorithms for physical robots.

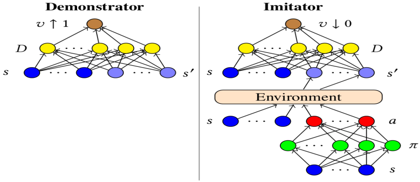
Ongoing Work

- Testing algorithms on more domains.
- Adapt algorithms for physical robots.
- Sim-to-real transfer using the algorithms.

Imitation Learning Summary



(a) BCO



(b) GAIfO

Faraz Torabi

Garrett Warnell