Chain of Thought Imitation with Procedure Cloning

Mengjiao (Sherry) Yang, Dale Schuurmans, Pieter Abbeel, Ofir Nachum
Advances in Neural Information Processing Systems 35 (NeurIPS 2022)

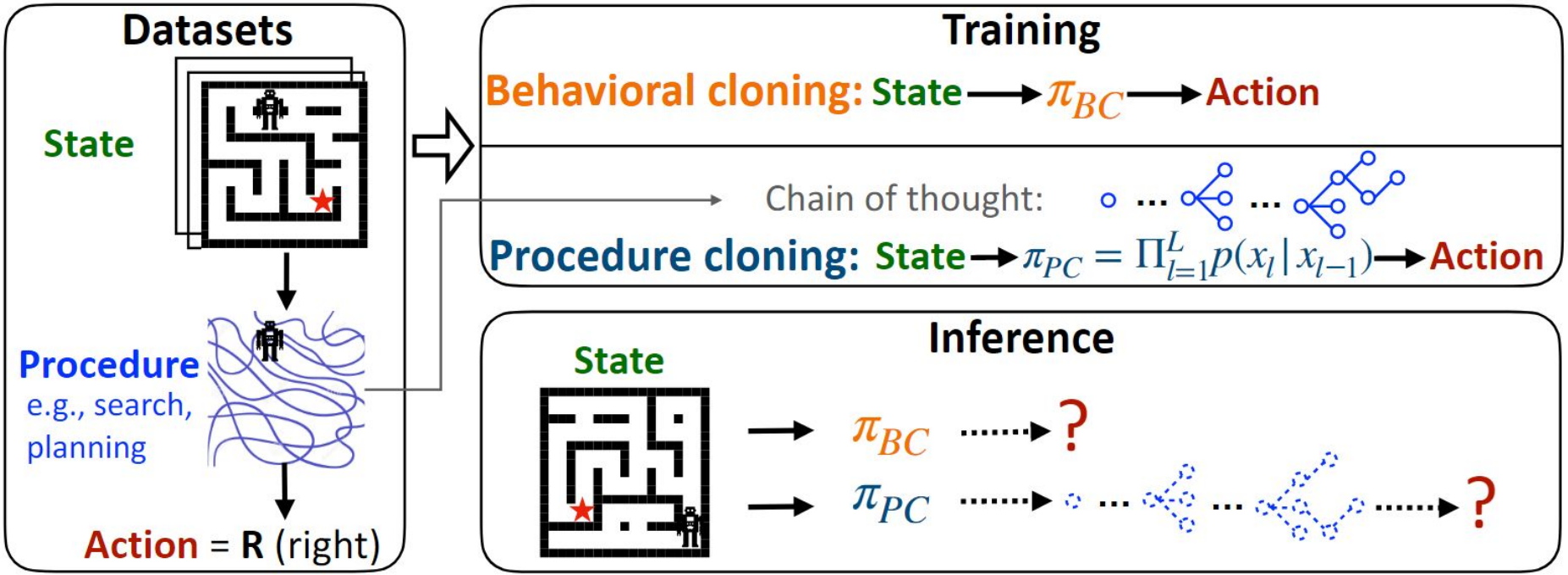
Purpose

- The aim of this paper is improvement of behavioral cloning methods.
  - Agents trained with BC objective can quickly fail to generalize to unseen states

- Chain of thought imitation: imitation learning using procedure cloning.

Procedure cloning

- Learn not only the expert's actions, but also the process for taking those actions.
- Predicts optimal behavior by autoregressively generating an expert's thought process


Datasets
State
Procedure
e.g., search,
planning
Action = R (right)
Training
Behavioral cloning: State → π_{BC} → Action
Chain of thought:
Procedure cloning: State → $\pi_{PC} = \Pi_{l=1}^{L} p(x_l \mid x_{l-1})$ → Action
Inference
State
π_{BC} ?
π_{PC} ?

Auxiliary Behavioral Cloning (Aux BC)

Add procedural information to Behavioral cloning as supporting role for more effective learning

Augmented Behavioral Cloning (Aug BC)

Improve agent performance by extending expert demonstrations and generating more diverse training data

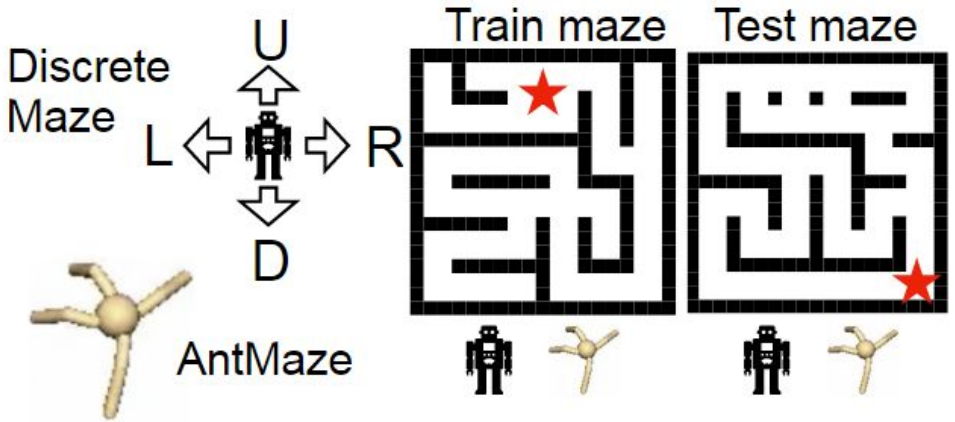
robot navigation in 2D Maze

discrete maze

agents use four discrete actions (up, down, left, right) to reach a target point. BFS

AntMaze

Four-legged agent "Ant" aims for the goal in a 2D maze.



Navigation - result

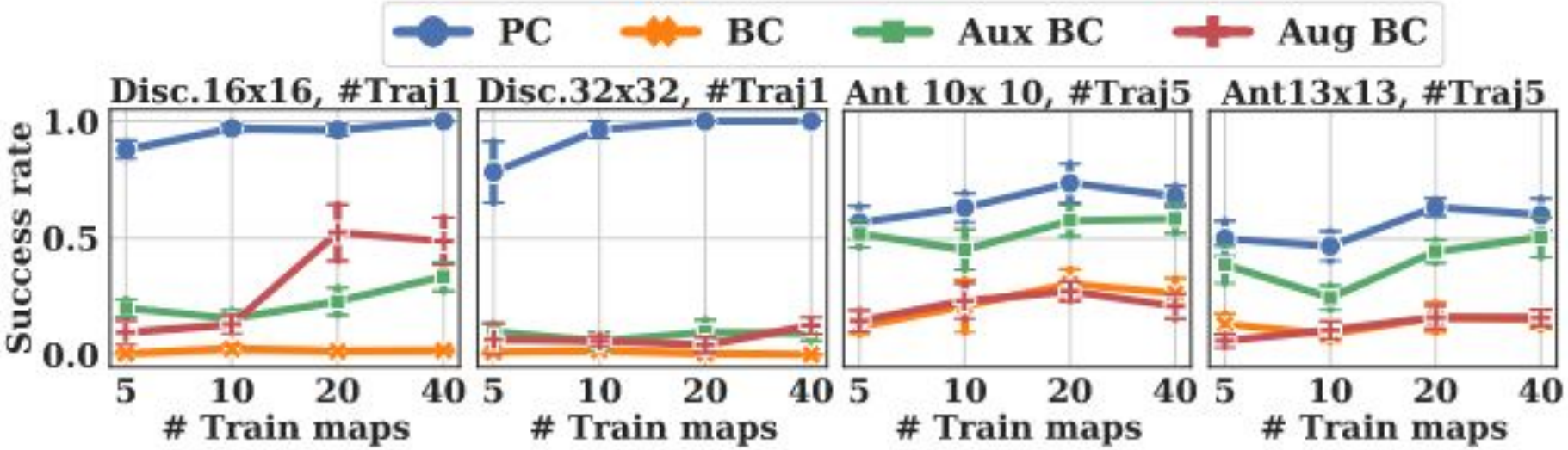
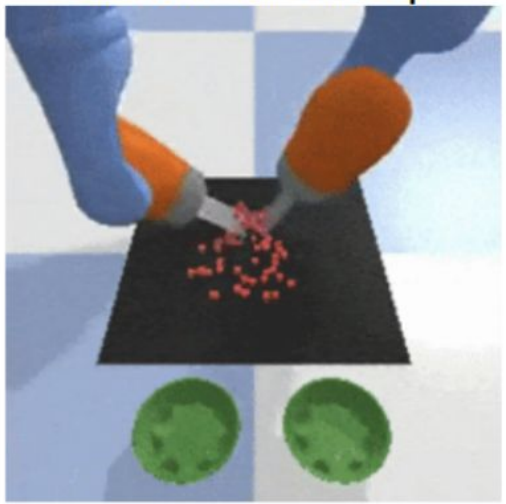


Image-based robot manipulation

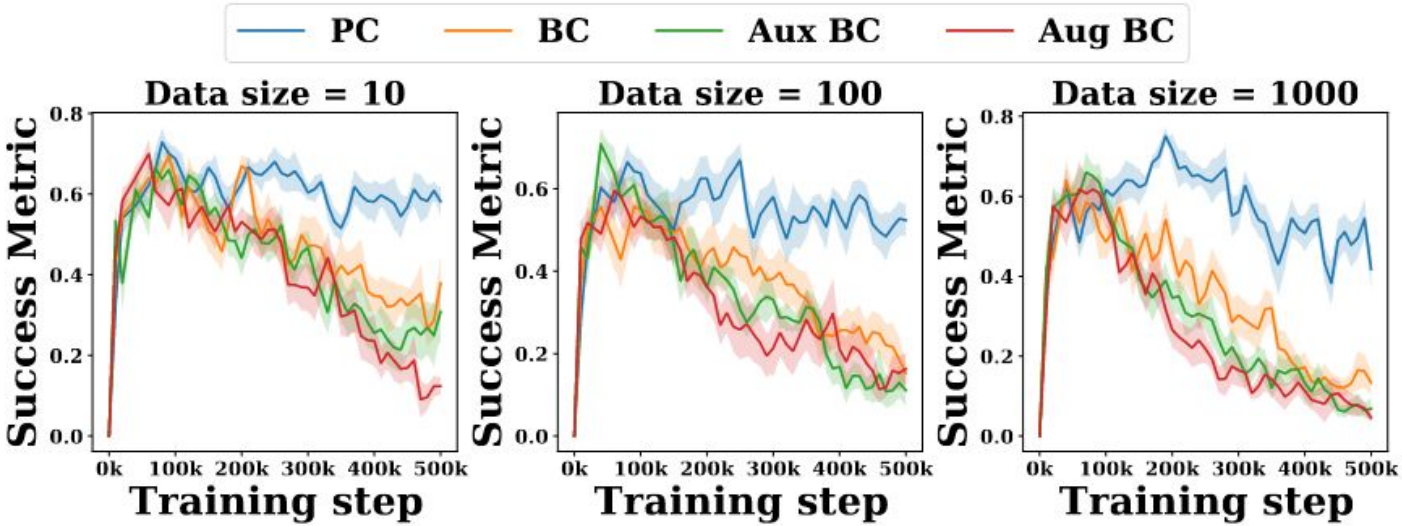
The two-handed sweeping task

sweep a pile of particles evenly into two bowls while avoiding dropping particles

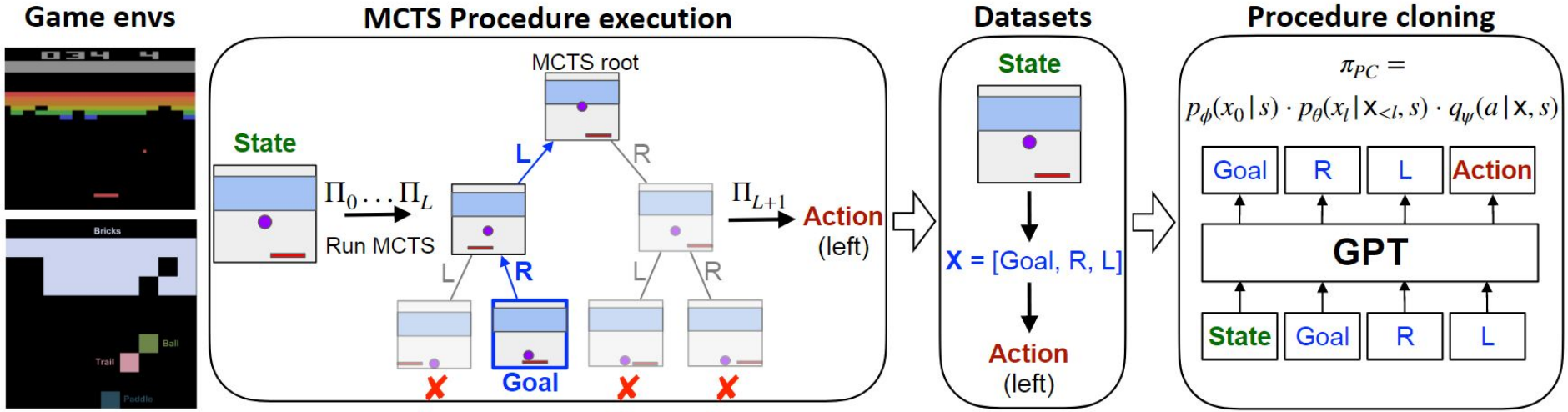
Bimanual sweep

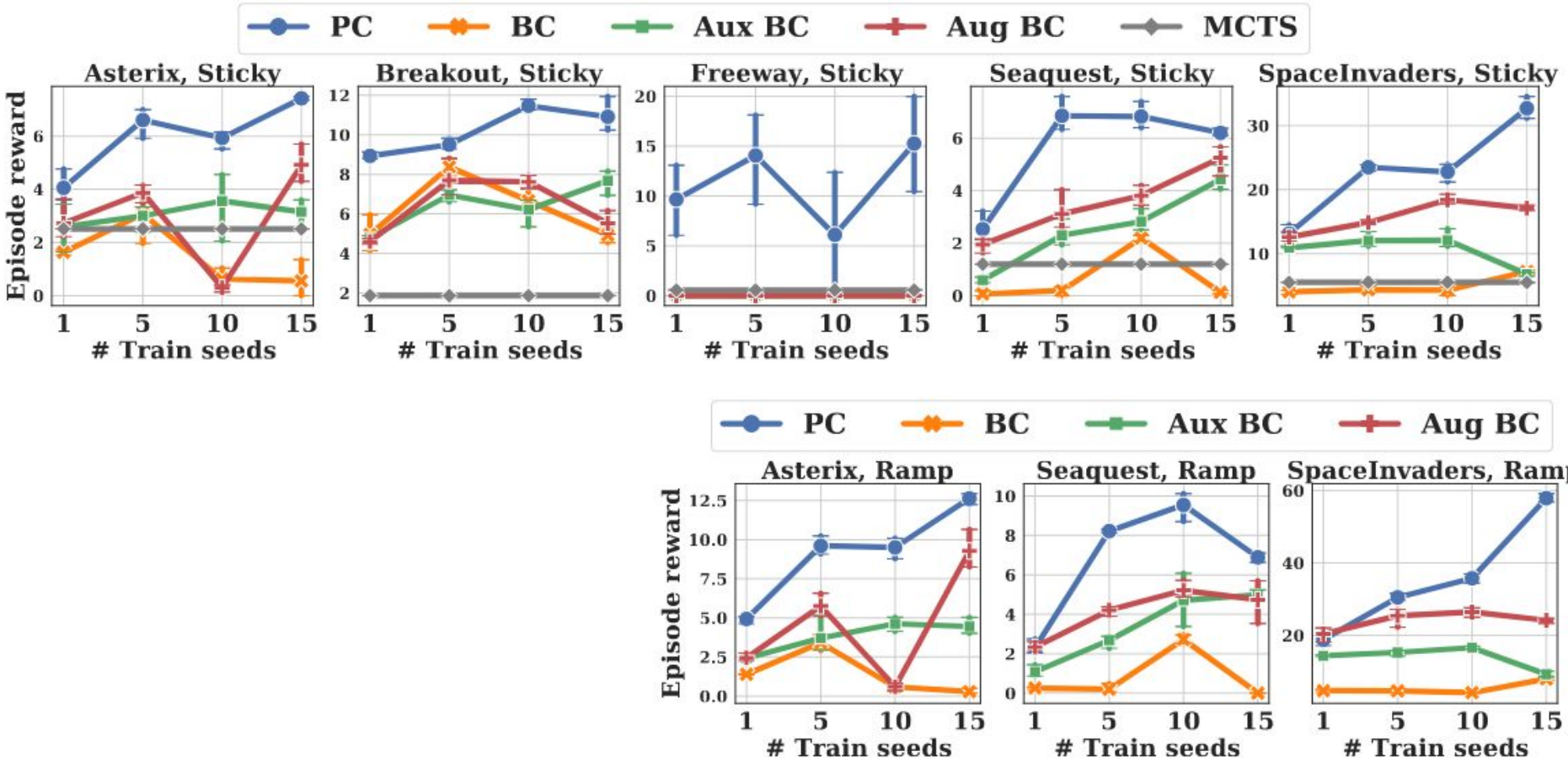


Robotic Manipulation - result



Strategy games in MinAtar




PC
BC
Aux BC
Aug BC
MCTS
Asterix, Sticky
Breakout, Sticky
Freeway, Sticky
Seaquest, Sticky
SpaceInvaders, Sticky
Episode reward
Train seeds
PC
BC
Aux BC
Aug BC
Asterix, Ramp
Seaquest, Ramp
SpaceInvaders, Ramp
Episode reward
Train seeds

Conclusion

Procedure cloning algorithms perform better than other alternative methods and can improve the ability of agents to perform goal-achieving tasks in unknown environments

The procedure cloning algorithm is useful in robot control and reinforcement learning

Thank you for listening