Causal Confusion in Imitation Learning

Pim de Haan, Dinesh Jayaraman, and Sergey Levine (NeurIPS 2019)

Dongmin Lee

Biointelligence Laboratory

Seoul National University

dmlee@bi.snu.ac.kr

April, 2020

Outline

Introduction

What is imitation learning?

- Learning a policy from examples of expert behavior
- The goal of imitation learning is to reproduce the expert's demonstrations with generalized intention

Introduction

Methods of imitation learning

1. Behavioral cloning
2. Inverse reinforcement learning

Introduction

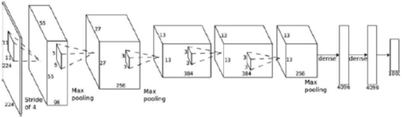
Methods of imitation learning

1. Behavioral cloning - Selected!
2. Inverse reinforcement learning

Introduction

What is behavioral cloning?

$\mathbf{o}_t$



$\pi_\theta(\mathbf{a}_t|\mathbf{o}_t)$

$\mathbf{a}_t$

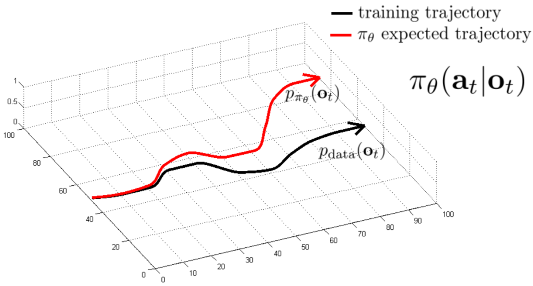
$\mathbf{o}_t$

$\mathbf{a}_t$

$\pi_\theta(\mathbf{a}_t|\mathbf{o}_t)$

Introduction

Fundamental problem of behavioral cloning: **distributional shift**

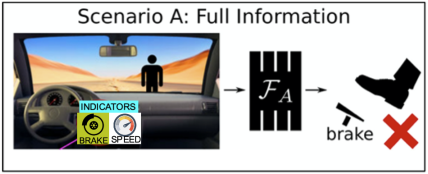


- Training and testing state distributions are different, induced respectively by the expert and learned policies

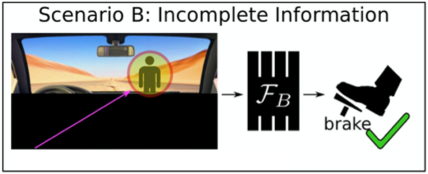
Introduction

What happens under distributional shift? **causal misidentification**

- Behavioral cloning to learn to drive a car in two scenarios (tasks):



policy attends to indicators

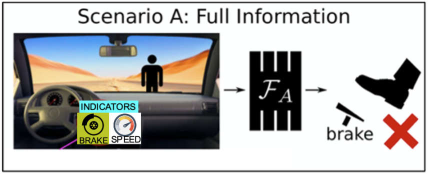


policy attends to pedestrian

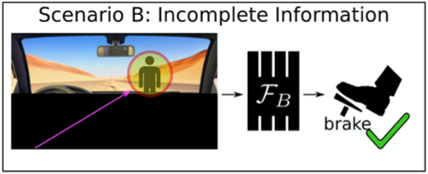
Introduction

What happens under distributional shift? **causal misidentification**

- Behavioral cloning to learn to drive a car in two scenarios (tasks):



policy attends to indicators



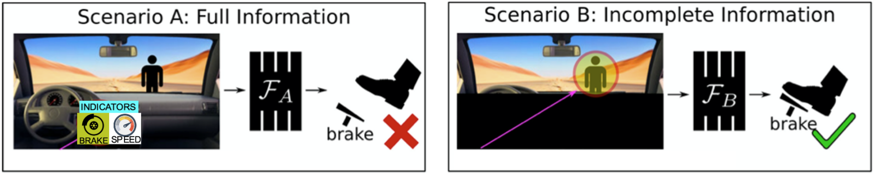
policy attends to pedestrian

- More information yields worse performance

Introduction

What happens under distributional shift? **causal misidentification**

- Behavioral cloning to learn to drive a car in two scenarios (tasks):



policy attends to indicators — policy attends to pedestrian

- More information yields worse performance
- Distinguishing correlates of expert actions in demonstration set from true causes is usually very difficult...

Causality and Causal Inference

What is causality and causal inference?

Causality and Causal Inference

Who am I?

Causality and Causal Inference

Who am I?

Why do you think this picture is a cat?

Causality and Causal Inference

What is causality?

- Causality (also referred to as causation, or cause and effect) is efficacy by which one event **(a cause)** contributes to the production of another event **(an effect)**
- The cause is partly responsible for the effect, and the effect is partly dependent on the cause

Causality and Causal Inference

What is causality?

- Causality (also referred to as causation, or cause and effect) is efficacy by which one event **(a cause)** contributes to the production of another event **(an effect)**
- The cause is partly responsible for the effect, and the effect is partly dependent on the cause

What is causal inference?

- Causal inference is general problem of deducing cause-effect relationships among variables
- **Causal discovery** approaches allow causal inference from pre-recorded observations under constraints

Causality and Causal Inference

Yoshua Bengio: presentations at NeurIPS, 2019 and AAAI, 2020

MISSING TO EXTEND DEEP LEARNING TO REACH
HUMAN-LEVEL AI

- **Out-of-distribution generalization & transfer**
- **Higher-level cognition: system 1 → system 2**
  - *High-level semantic representations*
  - *Compositionality*
  - *Causality*
- **Agent perspective:**
  - *Better world models*
  - ***Causality***
  - *Knowledge-seeking*
- **Connections between all 3 above!**

Causality and Causal Inference

Yann LeCun: ”Lots of people in ML/DL know that causal inference is an important way to improve generalization. The question is how to do it.”

Causality and Causal Inference

Yann LeCun: ”Lots of people in ML/DL know that causal inference is an important way to improve generalization. The question is how to do it.”

Recent papers:

- 25 papers at NeurIPS 2018 workshop on causal learning
- 18 papers at NeurIPS 2019 conference
- 13 papers at AAAI 2020 conference

$\Longrightarrow$ **Currently hot topic of research in machine learning!**

Causality and Causal Inference

Concepts of causality:

- Association vs Causation
- Intervention and do-calculus

Causality and Causal Inference

Association: $P(Y|X)$

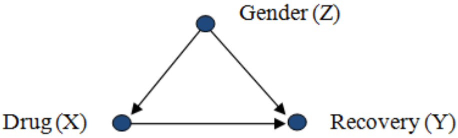
- Is there correlation between two variables?
- Is mutual information zero?
- Statistical inference for **observations** of two variables

Causation: $P(Y|do(X))$

- Is there causality between two variables?
- Does one variable respond to intervention?
- Randomized **intervention** for one variable

Causality and Causal Inference

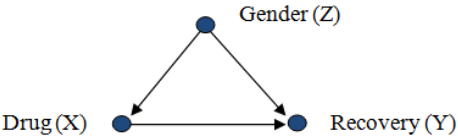
Confounder (also confounding variable, confounding factor) in causality:



- The cause of Y is X, and when there is **confounder** Z that affects X and Y at the same time, **confounding effect** occurs

Causality and Causal Inference

Confounder (also confounding variable, confounding factor) in causality:



- The cause of Y is X, and when there is **confounder** Z that affects X and Y at the same time, **confounding effect** occurs
- We define that

$$P(Y|do(X)) \neq P(Y|X)$$

  because observational quantity contains information about correlation between X and Z, while interventional quantity does not (since X is not correlated with Z in randomized intervention)

Causality in Imitation Learning

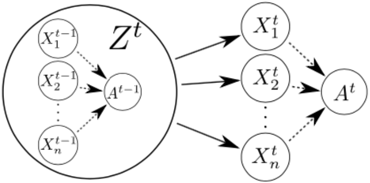
Imitation learning (i.e., behavioral cloning) setting:

- $X^t = [X_1^t, X_2^t, ..., X_n^t]$: expert's state observations at each time t
- A^t: expert's actions

$\Longrightarrow$ The goal is to learn a mapping π from X^t to A^t using all (X^t, A^t) from the expert's demonstrations

Causality in Imitation Learning

Causal structures:



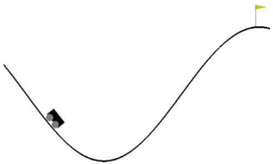
- Expert actions A^t are influenced by some information in state observations X^t
- A **confounder** Z^t influences each state variable in X^t
- Some unknown subset of disentangled factors of X^t **(causes)** affect expert actions, and the rest **(nuisance variables)** do not
- **Interventional query**:

$$P(A^t|do(X^t))$$

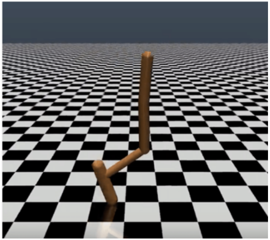
Experiments Setting

Benchmarks:

- We add information about **previous action**, which tends to correlate with **current action** in the expert data



Mountain Car
2 state dims
past action integer added



MuJoCo Hopper
11 state dims
3 past action dims added

Experiments Setting

Benchmarks:

- We add information about **previous action**, which tends to correlate with **current action** in the expert data

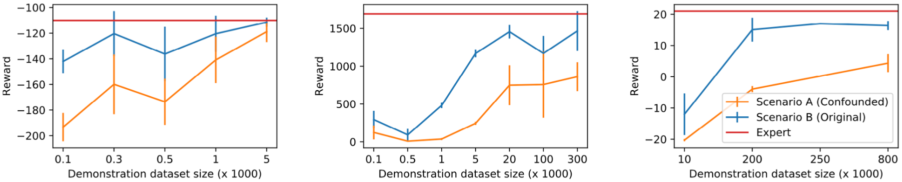
(a) Pong (b) Enduro (c) UpNDown

VAE encoding

30 state variables **(disentangled observations)**
setting to a disentangled representation by training β-VAE

Experiments Setting

Demonstrating **causal misidentification**



(a) MountainCar (b) Hopper (c) Pong

Resolving Causal Misidentification

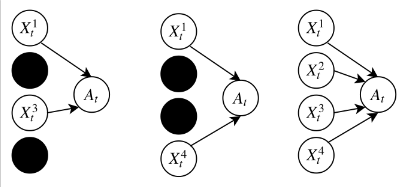
Proposed solutions:

1. Learn **a policy** that map from states to expert actions using randomized interventional queries

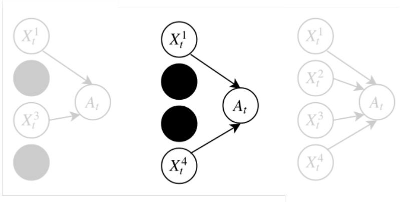
2. Find **the true causal intervention** using the trained policy

Resolving Causal Misidentification

1. Causal graph-parametrized policy learning

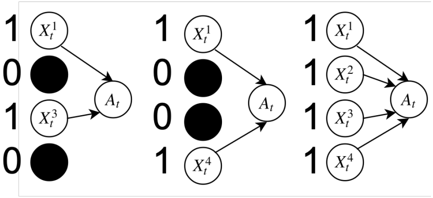


2. Targeted intervention: find true graph
   - Expert queries
   - Rewards



Causal Graph-Parameterized Policy Learning

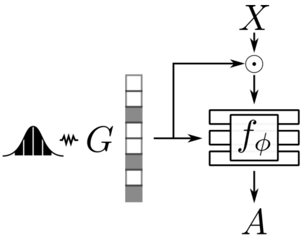
Causal graphs as masks



- Each state variable X_i in X may either be a cause or not, so there are 2^n possible graphs
- We parameterize **causal graph** G as a vector of n binary variables

Causal Graph-Parameterized Policy Learning

Graph-parameterized policy



- A single graph-parameterized policy

$$\pi_G(X) = f_\phi([X \odot G, G])$$

where $\odot$ is element-wise multiplication, and $[\cdot, \cdot]$ denotes concatenation

Causal Graph-Parameterized Policy Learning

Parameters ϕ of policy network f_ϕ are trained through gradient descent to minimize:

$$\mathbb{E}_G[\ell(f_\phi([X_i \odot G, G]), A_i)]$$

- G is drawn uniformly as a bernoulli random vector over all 2^n graphs
- ℓ: mean squared error loss or cross-entropy loss
- X_i, A_i: observations and actions in batches

Targeted Intervention

Expert query mode and policy execution mode

Algorithm 1 Expert query intervention

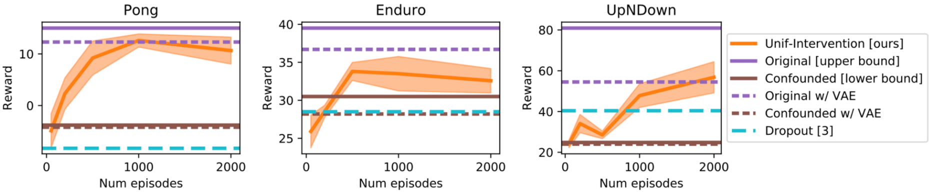
Input: policy network f_ϕ s.t. $\pi_G(X) = f_\phi([X \odot G, G])$
Initialize $w = 0, \mathcal{D} = \emptyset$.
Collect states $\mathcal{S}$ by executing π_{mix}, the mixture of policies π_G for uniform samples G.
For each X in S, compute disagreement score:
$$D(X) = \mathbb{E}_G[D_{KL}(\pi_G(X), \pi_{mix}(X))]$$
Select $\mathcal{S}' \subset \mathcal{S}$ with maximal $D(X)$.
Collect state-action pairs $\mathcal{T}$ by querying expert on $\mathcal{S}'$.
for $i = 1 \ldots N$ **do**
  Sample $G \sim p(G) \propto \exp\langle w, G\rangle$.
  $\mathcal{L} \leftarrow \mathbb{E}_{s,a\sim\mathcal{T}}[\ell(\pi_G(s), a)]$
  $\mathcal{D} \leftarrow \mathcal{D} \cup \{(G, \mathcal{L})\}$
  Fit w on $\mathcal{D}$ with linear regression.
end for
Return: $\arg\max_G p(G)$

Algorithm 2 Policy execution intervention

Input: policy network f_ϕ s.t. $\pi_G(X) = f_\phi([X \odot G, G])$
Initialize $w = 0, \mathcal{D} = \emptyset$.
for $i = 1 \ldots N$ **do**
  Sample $G \sim p(G) \propto \exp\langle w, G\rangle$.
  Collect episode return R_G by executing π_G.
  $\mathcal{D} \leftarrow \mathcal{D} \cup \{(G, R_G)\}$
  Fit w on $\mathcal{D}$ with linear regression.
end for
Return: $\arg\max_G p(G)$

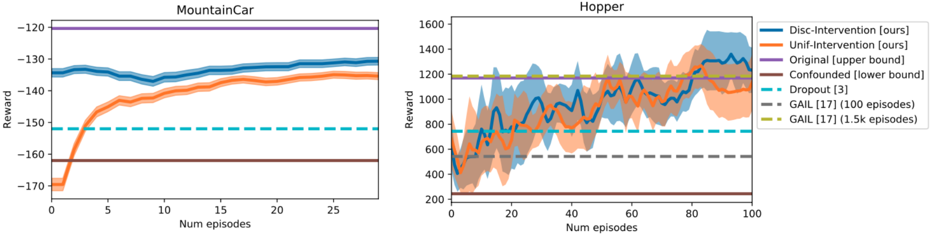
Experiments

Intervention by policy execution on Atari games



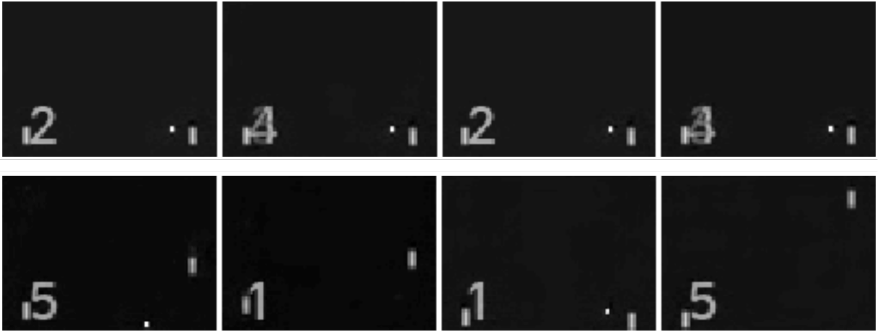
Experiments

Intervention by policy execution on MountainCar and Hopper



Experiments

Interpreting the learned causal graph



samples from (top row) learned causal graph
and (bottom row) random causal graph on Pong

Experiments

Necessity of disentanglement

Mode	Representation	Reward
Policy execution	Disentangled	**-137**
	Entangled	-145
Expert queries	Disentangled	**-140**
	Entangled	-165

intervention on (dis)entangled MountainCar

Thank You!

Any Questions?