Rethinking Causal Confusion in Imitation Learning

Han Lin
hl3199@columbia.edu
MS in Computer Science, Columbia University

Abstract

Traditional behavioral cloning methods for imitation learning maps input observations to expert actions, which might cause causal confusion due to the ignorance of causal structure of observational variables [2]. A common instance of this causal confusion occurs in the partially observed settings when expert actions are strongly correlated over time, and the imitator learns to follow the past action from expert rather than learns some valuable policies from observations [10]. In this project, we assume that the imitator could only learn policies from expert demonstrations without requiring further environmental interactions. We want to find solutions to the following questions: (1) How to effectively remove such nuisance correlations coming from past action to learn disentangled representations? (2) Is it possible (and necessary) to learn the causal structure of observational variables to solve such causal confusion problem directly?

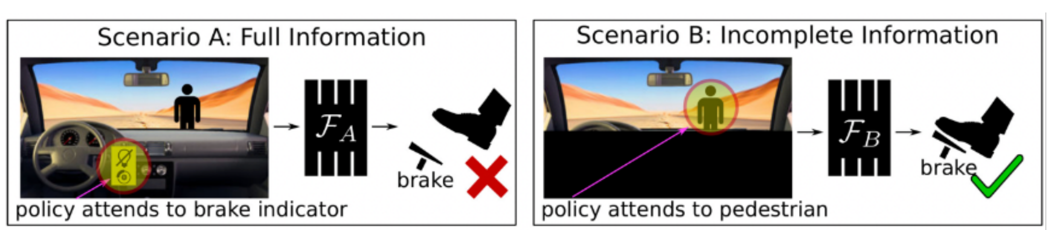
1 INTRODUCTION

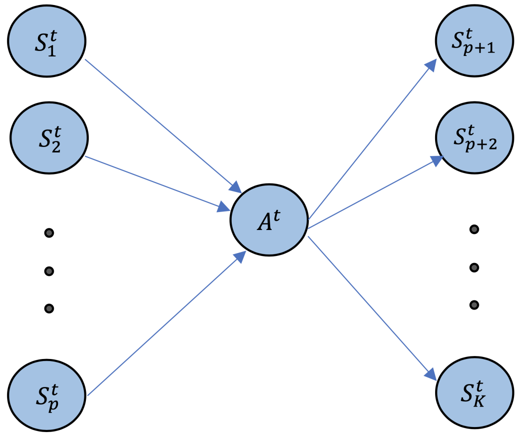
Illustrative Example: Misidentify Effect As Cause

Let's start with two examples [2, 10] to show what is *causal confusion* in imitation learning.

Consider using imitation learning to train a learner to drive a car [2]. In scenario A, the model's input is an image of the dashboard and windshield, and in scenario B, the input to the model (with identical architecture) is the same image but with the dashboard masked out. Both cloned policies achieve low training loss, but when tested on the road, model B drives well, while model A does not.

The reason: the dashboard has an yellow indicator light that comes on immediately when the brake is applied, and model A wrongly learns to apply the brake only when the brake light is on. Even though the brake light is the **effect** of braking, model A could achieve low training error by misidentifying it as the **cause** instead.





2 RELATED WORK

Causal confusion problem happens when a policy exploits the nuisance correlated in states for predicting expert actions. Distribution shift is one of the main reason that cause such confusion. Therefore, various research focus on learning disentangled representations which is robust to distribution shift. Haan [2] solves the causal confusion problem by randomly masked disentangled representations learned from β-VAE, and infers the best mask through environment interaction. Instead, [6] removes the requirement of environment interaction, and achieves better performance by learning disentangled representations through VQ-VAE. Some reviewers [1] doubt that the resolution of the Atari datasets used for experiments in [6] is not high enough, recent models including VQ-VAE2 [7] and VQ-GAN [3], which focus on large images, might be a promising way to solve their concerns. This line of work mainly focus on learning disentangled representations, and use **random masking** to sample the best subset of representation variables in the hope of removing nuisance correlates.

Considering the above methods do not learn causal graphs explicitly, we may wonder are there ways to solve causal confusion by directly learning causal graphs. However, recent progress shows a sober look that unsupervised learning of disentangled representations is fundamentally impossible without inductive biases on both the models and the data [4]. In other words, the causal confusion problem cannot be solved completely by just collecting more samples from expert, as there is no additional information for identifying the cause of expert actions, and the reward function is just learned using the given expert demonstrations where nuisance correlates exist [1]. Therefore, recent literatures for learning disentangled representations focus on using some additional labeled data to help identify the causal relationship [11, 9], but it's still unclear how can we apply these methods to imitation learning.

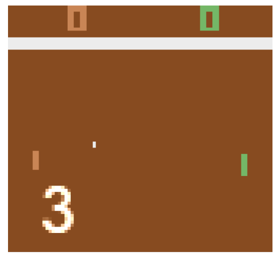
To the best of our knowledge, there's **no former research in imitation learning that explicitly solves causal confusion problem by learning the causal graphs directly** due to the obstacles stated above. Combining causal models and representation learning in imitation learning still remains an open problem [8].

3 PROBLEM TO SOLVE

In this section, we briefly illustrate the imitation learning tasks we want to solve.

Dataset: Our project is motivated by solving imitation learning problem from the Atari environment, which consists of 27 video games. For example, Pong is a two-dimensional sports game that simulates table tennis https://www.endtoend.ai/envs/gym/atari/pong/. A player can control the right green paddle to hit a ball back. And the goal for each player is to reach 11 points before the opponent. The imitator has access to discrete video images X^t as well as expert's action A^t for $t = 1, 2, ...T$ periods. And the imitator's goal is to learn a policy $\pi(X^t)$ from (X^t, A^t) tuples of expert demonstrations to achieve higher score of the game.

Confounded images: In order to address the causal confusion problem we introduced in section 1, Haan [2] added the number representing previous actions at the left bottom corner of each image (the white number 3 in the Pong example below) to make an analogy to the yellow yellow indicator light shown in Scenario A of the figure in section 1. And the problem we want to solve is to learn an effective policy that is robust to test time even with distribution shift.



4 METHODS

In this section, we will take a closer look at two directions of recent research that addresses causal structure in representation learning, and answer the following problem: **Will additional information for causal graphs (causal graph structure)**

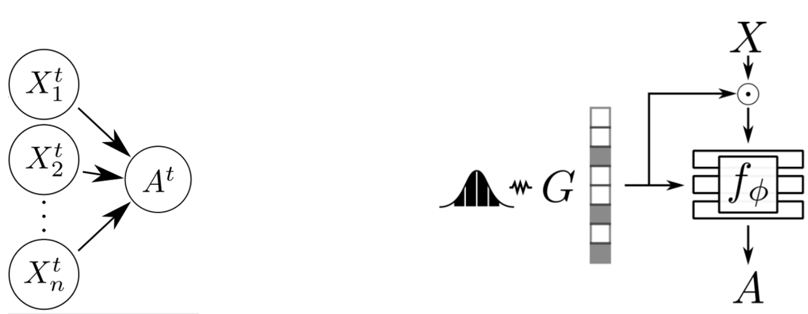
4.1 Random Masking

This is the method first proposed in Haan [2]. The author first use β-VAE to learn an disentangled representation from images. Suppose that β-VAE outputs an representation of dimension n: $\{X_i^t\}_{i=1}^n$, the author first assumes that each X_i^t are already disentangled, and some of these X_i^t's will be the cause of A^t (as shown in the left subgraph below). Since each X_i^t could either be a cause or not, there are 2^n possible graphs.

In the right subgraph below, the author parameterize the structure G of the causal graph as a vector of n binary variables, each indicating the presence of an arrow from X_i^t to A^t. In order to learn the causal graph, G is drawn uniformly at random over all 2^n possible graphs, and the paper minimizes the following loss:

$$E_G[l(f_\phi([X_i \odot G, G]), A)] \tag{1}$$

where $\odot$ is elementwise product, and $[,]$ represents concatenation. f_ϕ is the policy network mapping observations X to actions A to be trained. After training, the graph G^* that minimizes the loss l inside expectation will be the best approximation for the underlying causal graph.



Recent research [6] based on Haan [2] observes that by changing β-VAE with VQ-VAE, which is better at finding discrete representations, will achieve better performance on Atari games, but it also uses some variants of random masking to find the best approximate causal graph.

4.2 Distinguishing Cause & Effect via Neural Causation Coefficient (NCC)

For each (S_i, A) pair, $i \in [K]$, we have a bag of samples $D_i = \{(S_i^t, A^t)\}_{t=1}^T$, with S_i^t represents the i^{th} disentangled feature from some VAE model, and A_t is from expert demonstrations. Then it's natural to ask could we distinguish whether S_i is the **Cause** of A or **Effect** of A.

Following [5], let's first assume that we have a dataset $D_i' = \{\{(S_i^t, A^t)\}_{t=1}^T, l_i\}$ where l_i is the 0-1 binary label, with $l_i = 1$ represent $S_i \to A$, and 0 otherwise. Then we could just use supervised learning method to train a classifier that could distinguish cause and effect:

$$NCC(\{(S_i^t, A^t)\}_{t=1}^T) = \psi\left(\frac{1}{T}\sum_{t=1}^T \phi(S_i^t, A^t)\right) \tag{2}$$

where ϕ is a feature embedding, and $\frac{1}{T}\sum_{t=1}^T \phi(S_i^t, A^t)$ represents the mean embedding. ψ is a binary classifier that takes such mean embedding as input and outputs the estimated probability: $\widehat{P}(S_i \to A)$.

However, the problem is that we do not have such labels l_i at hand, and cannot do such supervised learning procedure directly. A possible way is to first train the model on some synthetic data generated with known cause/effect labels, and then apply the model to our (S_i, A) pairs. The procedure to generate a synthetic dataset with $X \to Y$ is as follows:

- first specify the distributions $P(f), P(X), P(E)$ that f, x, e are generated from
- for each sample $i \in [I]$, generate $x_i \sim P(X)$, $e_i \sim P(E)$, and then set $y_i \leftarrow f(x_i) + e_i$
- return the generated dataset $D = \{x_i, y_i\}_{i=1}^I$

Similarly, we could also generate the dataset with $Y \to X$ by assining x_i as a $f(y_i) + e_i$ in the above procedure.

The shortcoming of this method is two folds:

(1) We assume S_i are disentangled, which depends

4.3 Learning SCM via (semi-)supervised learning

One concern about the above method is that it assumes that each X_i^t $(i = 1, ...n)$ are disentangled, and they jointly cause action A^t, but such assumption heavily depends on the performance of disentanglement of VAE models. Moreover, the underlying factors in representations can be causally correlated. Therefore, it's natural to ask whether it is possible to learn the underlying causal relationships among directly.

However, as mentioned in the related work section, [4] proves it's impossible for pure unsupervised learning methods to learn disentangled representations. Therefore, recent works [11, 9] focus on using additional labels to learn the SCM. More specifically: **Shen [9]** assumes that (1)we know the adjacency matrix A of all latent variables, with $A_{ij} = 1$ represents variable z_i causes z_j. (2)we have a few ($\leq$ 10%) labeled data for the observed latent variables, while **Yang [11]** assumes that we have all labeled data for the observed latent variables.

Back to our Atari games, if we want to learn the causal graphs explicitly via supervised learning, we need to know at least some labeled data for observed latent variables. However, it is unclear what are the latent variables

References

[1] Object-aware regularization for addressing causal confusion in imitation learning, author's response for reviewer jmzw. `https://openreview.net/forum?id=FEhntTXAeHN`.

[2] P. de Haan, D. Jayaraman, and S. Levine. Causal confusion in imitation learning. *CoRR*, abs/1905.11979, 2019.

[3] P. Esser, R. Rombach, and B. Ommer. Taming transformers for high-resolution image synthesis. *CoRR*, abs/2012.09841, 2020.

[4] F. Locatello, S. Bauer, M. Lucic, S. Gelly, B. Schölkopf, and O. Bachem. Challenging common assumptions in the unsupervised learning of disentangled representations. *CoRR*, abs/1811.12359, 2018.

[5] D. Lopez-Paz, R. Nishihara, S. Chintala, B. Schölkopf, and L. Bottou. Discovering causal signals in images, 2016.

[6] J. Park, Y. Seo, C. Liu, L. Zhao, T. Qin, J. Shin, and T. Liu. Object-aware regularization for addressing causal confusion in imitation learning. *CoRR*, abs/2110.14118, 2021.

[7] A. Razavi, A. van den Oord, and O. Vinyals. Generating diverse high-fidelity images with VQ-VAE-2. *CoRR*, abs/1906.00446, 2019.

[8] B. Schölkopf. Causality for machine learning. *CoRR*, abs/1911.10500, 2019.

[9] X. Shen, F. Liu, H. Dong, Q. LIAN, Z. Chen, and T. Zhang. Disentangled generative causal representation learning, 2021.

[10] C. Wen, J. Lin, T. Darrell, D. Jayaraman, and Y. Gao. Fighting copycat agents in behavioral cloning from observation histories. In *NeurIPS*, 2020.

[11] M. Yang, F. Liu, Z. Chen, X. Shen, J. Hao, and J. Wang. Causalvae: Structured causal disentanglement in variational autoencoder. *CoRR*, abs/2004.08697, 2020.