Partial Transportability for Domain Generalization

Kasra Jalaldoust*, Alexis Bellot*, Elias Bareinboim

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

Introduction

- We develop a framework to provide performance guarantees for predictions made in an unseen domain using knowledge of causal relations encoded in selection diagrams.
- It is based on Neural Causal Models (NCMs) and is the first general estimation technique for transportability problems [1, 2].

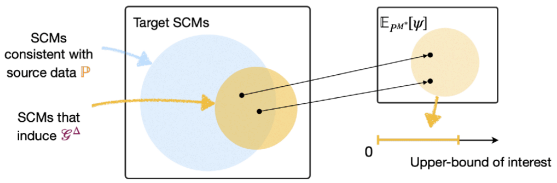

Figure 1. Partial Transportability Task.

Task: Find the tightest bound:

$$\mathbb{E}_{P^{M^*}}[\psi(\boldsymbol{V})] \leq q_{\max}, \quad (1)$$

$\forall \mathbb{M} : \{M^1, \ldots, M^k, M^*\}$ that induce $\mathcal{G}^{\Delta}$ and $\mathbb{P} : \{P^1, \ldots, P^k\}$.

Example

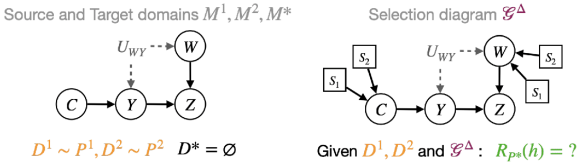

Figure 2. Selection diagrams encode structural invariances.

- **Input**: A classifier h, assumptions encoded in selection diagrams $\mathcal{G}^{\Delta}$, and data samples from source domains $D^1, D^2 \sim P^1, P^2$.
- **Goal**: Evaluate the worst case error in target M^*,

$$\max_{\text{target distributions } P^*} R_{P^*}(h), \qquad R_P(h) := \mathbb{E}_P[\mathcal{L}(y, h(\boldsymbol{x}))] \quad (2)$$

of three baseline classifiers:

$$h_1(\boldsymbol{c}, w) := w \oplus \bigoplus_{c \in \boldsymbol{c}} c, \quad h_2(\boldsymbol{c}) := \bigoplus_{c \in \boldsymbol{c}} c, \quad h_3(z) := z$$

Classifier	$R_{P^{M^1}}$	$R_{P^{M^2}}$	$R_{P^{M^*}}$
$h_1(\boldsymbol{c}, w)$	1%	4%	49%
$h_2(\boldsymbol{c})$	20%	20%	20%
$h_3(z)$	3%	5%	4%

Table 1. In-domain and worst-case out-of-domain error of classifiers.

Research Questions

Q1: Evaluate the worst-case error of a classifier h in M^*,

$$\max_{M^* \in \text{ tuple of SCMs } \mathbb{M} \text{ that entails } \mathbb{P} \text{ \& induces } \mathcal{G}^{\Delta}} R_{P^{M^*}}(h). \quad (3)$$

Q2: Optimize h for best worst-case performance in M^*,

$$\arg\min_h \max_{M^* \in \text{ tuple of SCMs } \mathbb{M} \text{ that entails } \mathbb{P} \text{ \& induces } \mathcal{G}^{\Delta}} R_{P^{M^*}}(h). \quad (4)$$

Q1: Evaluation by fitting NCMs

Theorem (Neural-TR). Given $\mathcal{G}^{\Delta}$ and distributions $\mathbb{P}$,

$$\max_{M^* \in \text{ tuple of NCMs } \mathbb{M} \text{ that entails } \mathbb{P} \text{ \& induces } \mathcal{G}^{\Delta}} R_{P^{M^*}}(h) \quad (5)$$

is a tight upper-bound on the target risk.

- NCMs are SCMs parameterized by NN.
- NCMs can be used to encode all constraints in Transportability tasks.
- Sound only in discretely-valued systems

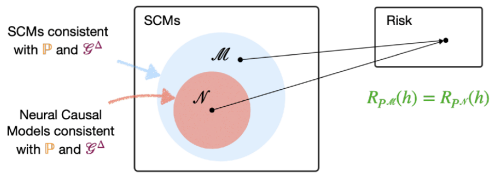

Figure 3. Expressiveness of NCMs for Transportability task.

Q2: Optimization by Adversarial Training

The CRO algorithm starts with a random classifier h_t and and empty set $\mathbb{D}$. It proceeds until the error converges, as follows:

1. Run evaluation (Neural-TR) on h_t and obtain M_t that entails the worst-case error.
2. Generate adversarial data $D_t \sim P_{M_t}, \mathbb{D} = \mathbb{D} \cup D_t$.
3. Train new classifier h_{t+1} to minimize $\max_{D \in \mathbb{D}} R_D(h)$, and iterate ...

Theorem (CRO). Given $\mathcal{G}^{\Delta}$ and distributions $\mathbb{P}$,

$$\text{CRO} \in \arg\min_h \max_{M^* \in \text{ tuple of SCMs } \mathbb{M} \text{ that entails } \mathbb{P} \text{ \& induces } \mathcal{G}^{\Delta}} R_{P^{M^*}}(h). \quad (6)$$

- In fact, with CRO we show that h_3 is **optimal** in the worst-case.

Simulations

- We consider the colored MNIST (CMNIST) dataset.
- The task is to use 2 source dataset to optimize for the worst-case performance in a test domain in which the color assignment may be arbitrary.
- After three iterations CRO converges to a classifier with an error of 25% in the worst-case, which is theoretically optimal.

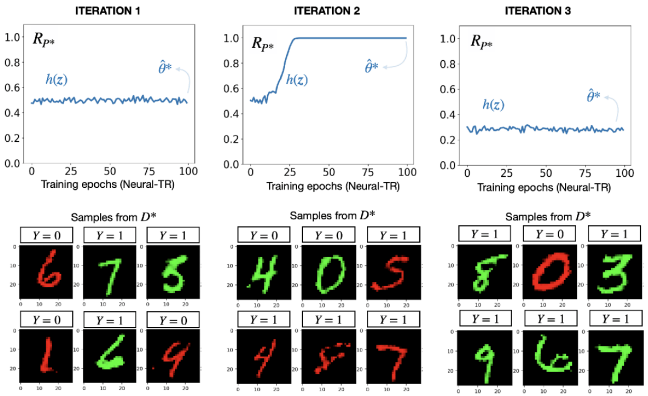

Figure 4. CMNIST.

Acknowledgements

This research was supported by the NSF, ONR, AFOSR, DARPA, DoE, Amazon, JP Morgan, and The Alfred P. Sloan Foundation.

References

[1] Elias Bareinboim and Judea Pearl. Causal inference and the data-fusion problem. PNAS, 2016.

[2] Kevin Xia, Kai-Zhan Lee, Yoshua Bengio, and Elias Bareinboim. The causal-neural connection: Expressiveness, learnability, and inference. NeurIPS, 2021.

Figure 5. Paper link.