Transportability theory offers first-ever guarantees for **generalizing** representations to **new, unseen domains.**

Transportable Representations for Domain Generalization

Kasra Jalaldoust and Elias Bareinboim

Summary

- The assumption of identical **training** and **test** distributions is often too strong and frequently violated in practical settings.
- We study the generalizability challenge where data from multiple domains are used to train a predictor capable of excelling in a novel, unseen target domain.
- The paper's main observation is that proper generalizability depends on the **modularity and stability of the causal mechanisms**.
- This problem is studied in the literature under the rubric of **transportability theory (TR)**.
- In this paper, (1) we demonstrate how to transport queries involving representations, and (2) we develop a dual approach — graphical and data-driven — to transportability.

Notation

Observables $\mathbf{X}$ and Label Y

Source domains $\mathscr{M}^1, \mathscr{M}^2, \ldots, \mathscr{M}^T$

Source distributions: $P^1, P^2, \ldots, P^T$

Target domain $\mathscr{M}^*$ and dist. P^*

Selection diagram $\mathcal{G}^{\Delta}$

Representation $\mathbf{R} = \phi(\mathbf{X})$

Score function $\mathbb{E}_{P^*}[Y \mid \phi(\mathbf{X}) = \mathbf{r}]$

Empirical score $\mathbb{E}_{P^i}[Y \mid \phi(\mathbf{X}) = \mathbf{r}]$

Definition. The representation ϕ is said to be transportable (TR) if the score function of $\mathbf{R} = \phi(\mathbf{X})$ is unique w.r.t. the input data and assumptions, i.e.,

$$\mathbb{E}_{P\mathscr{M}_a^*}[Y \mid \mathbf{r}] = \mathbb{E}_{P\mathscr{M}_b^*}[Y \mid \mathbf{r}]$$

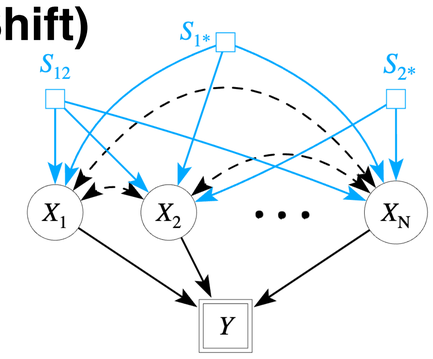
for all SCMs $\{\mathscr{M}_a^i\}, \{\mathscr{M}_b^i\}$ such that

1. $P^1, P^2, \ldots, P^T$ is entailed, and
2. $\mathcal{G}^{\Delta}$ is induced.

Example. (Covariate Shift)

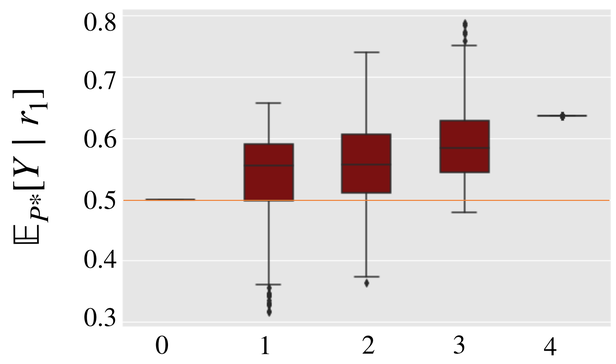
Selection diagram $\mathcal{G}^{\Delta}$:

(Binary Features X_i)



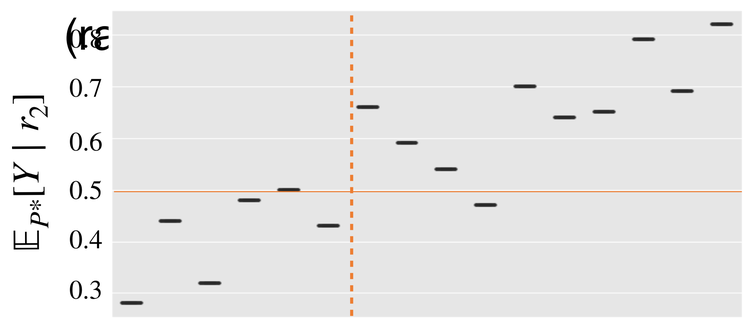
Q: Is representation $R \leftarrow \phi(\mathbf{X})$ TR?

Representation $R_1 : \|\mathbf{X}\|$



NOT TR

Representation $R_2 : \beta^\top \cdot \mathbf{X}$



TR

Theorem 1: The algorithm ϕ-TR (check QR) is sound & complete to decide whether a representation ϕ is TR across domains.

Data-Driven transportability

What if we don't have the graphical model $\mathcal{G}^{\Delta}$?

Alternative assumptions are needed that connect the data to the underlying causal structure.

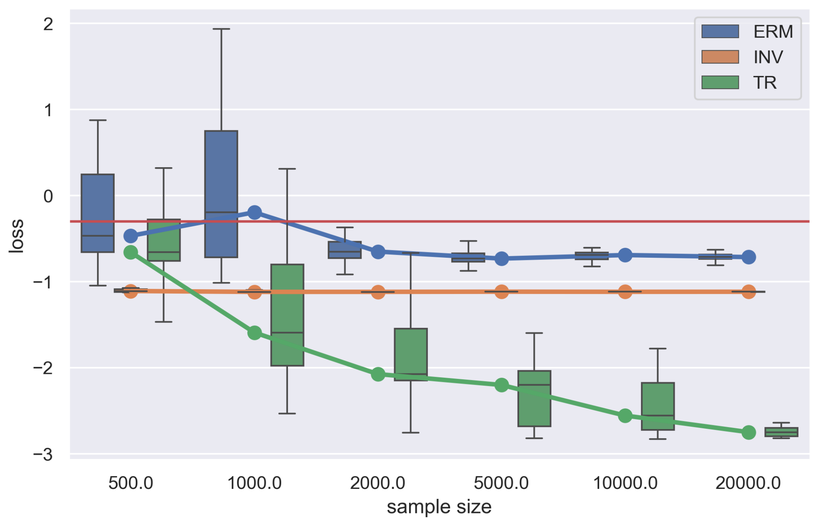
A1. R-faithfulness: The conditional independence relations implied by the data corresponds to d-separation in selection diagram.

A2. Causal Mechanistic Stability (CMS): If the mechanism of *any variable* is stable across the sources, it remains stable in the target as well.

Theorem 2: **Invariance of the empirical scores** across the sources is a sound and complete criterion for **transportability of a representation.**

This result unifies the work on invariance-based generalization under the umbrella of transportability.

Experiments



AAAI Association for the Advancement of Artificial Intelligence

COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK