From Statistical Transportability to Estimating the Effect of Stochastic Interventions

Juan D. Correa
j.d.correa@columbia.edu

Elias Bareinboim
eliasb@columbia.edu

Columbia University
Purdue University

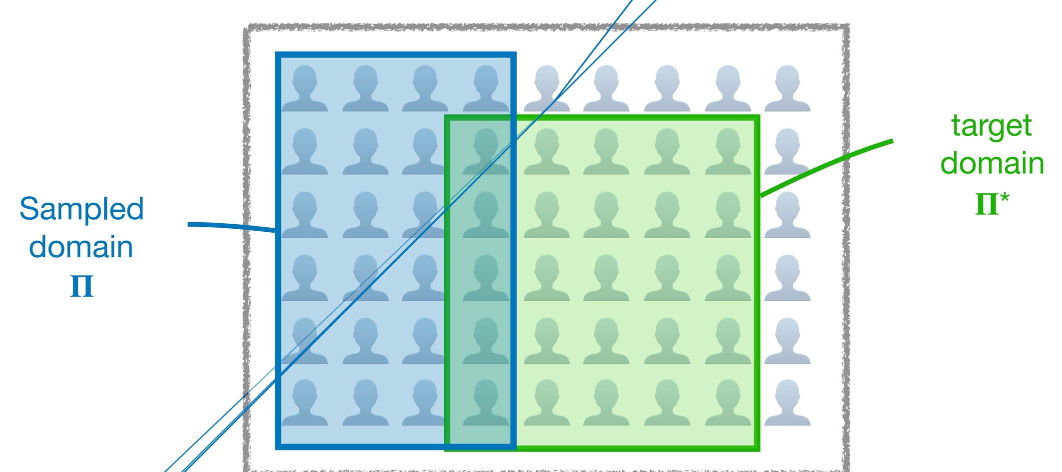
1 Introduction

One of the main achievements of modern ML technology is the efficient training of models using data collected from the corresponding, unobserved data-generating process.

In practice, however, the environment from where the data is collected rarely matches where the model is intended to be used. This leads to the problem of *transportability* (PNAS'2016).

2 The Transportability Challenge

There exists a mismatch between the sampled domain $\mathbf{\Pi}$ and the target domain $\mathbf{\Pi}^*$, due to change of mechanism or distribution, for one or more of the measured variables $\mathbf{W}$.

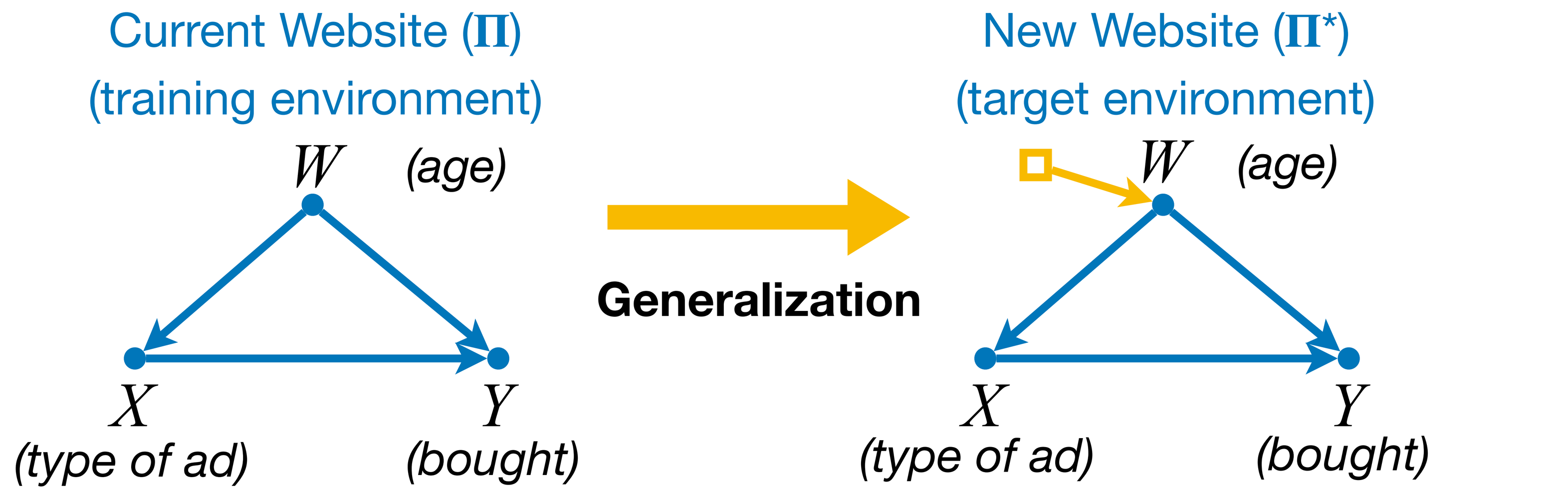


$P^*(\mathbf{w}) \neq P(\mathbf{w})$

- How to generalize the model learned in the source environment to a different, but related target environments?
- Do we need to obtain samples from $\mathbf{\Pi}^*$ and train a new model from scratch?

3 Motivating Example

- Suppose we have trained a model to predict the likelihood of people buying a product based on the type of ad used.



$P(W) \neq P^*(W)$ hence $P(y \mid x) \neq P^*(y \mid x)$

- Distributions $P(x,y,w)$ and $P^*(x,y,w)$ can be factorized as

$$P(x,y,w) = P(w)\, P(x|w)\, P(y|x,w)$$
$$P^*(x,y,w) = P^*(w)\, P^*(x|w)\, P^*(y|x,w)$$

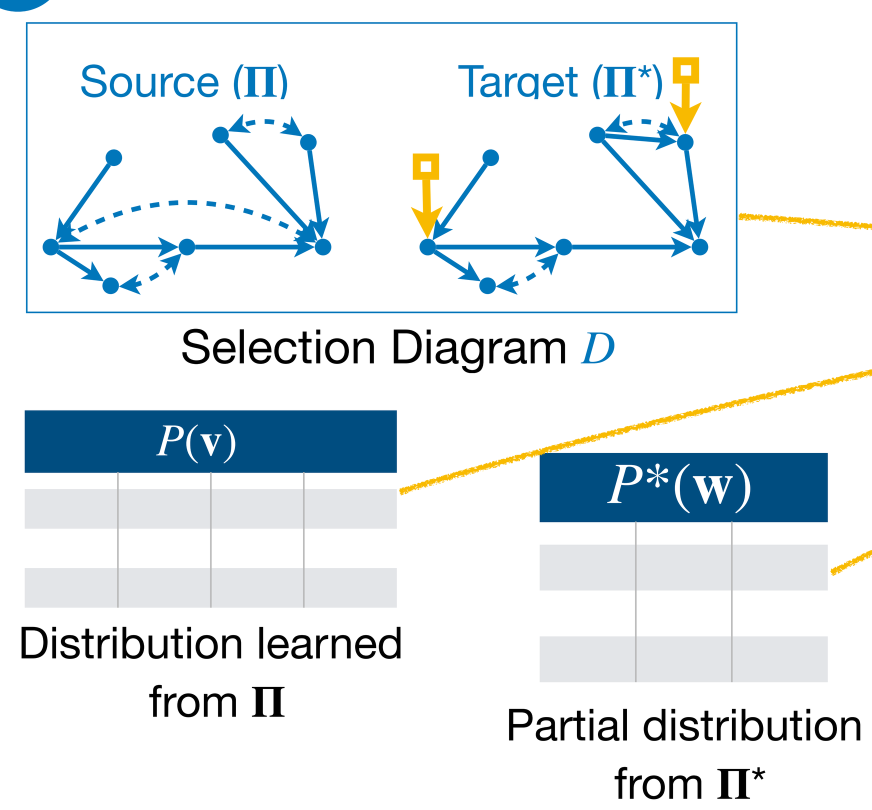
$P(x|w)$, $P(y|x,w)$ are implied to be the same in both environments

- The target distribution $P^*(y|x)$ can be expressed as:

$$P^*(y \mid x) = \frac{P^*(y,x)}{P^*(x)} = \frac{\sum_w P^*(y \mid x,w) P^*(x \mid w) P^*(w)}{\sum_w P^*(x \mid w) P^*(w)}$$
$$= \frac{\sum_w P(y \mid x,w) P(x \mid w) P^*(w)}{\sum_w P(x \mid w) P^*(w)}$$

- Under the assumptions implied by the diagram, we need only to measure $P^*(w)$ in the target environment and can reuse what was learned from the source domain.

4 Task: Deciding Statistical Transportability



Is there a function f such that

$$P^*(y \mid x) = f(P(\mathbf{v}), P^*(\mathbf{w}))\ ?$$

yes (f) / no

5 Our strategy

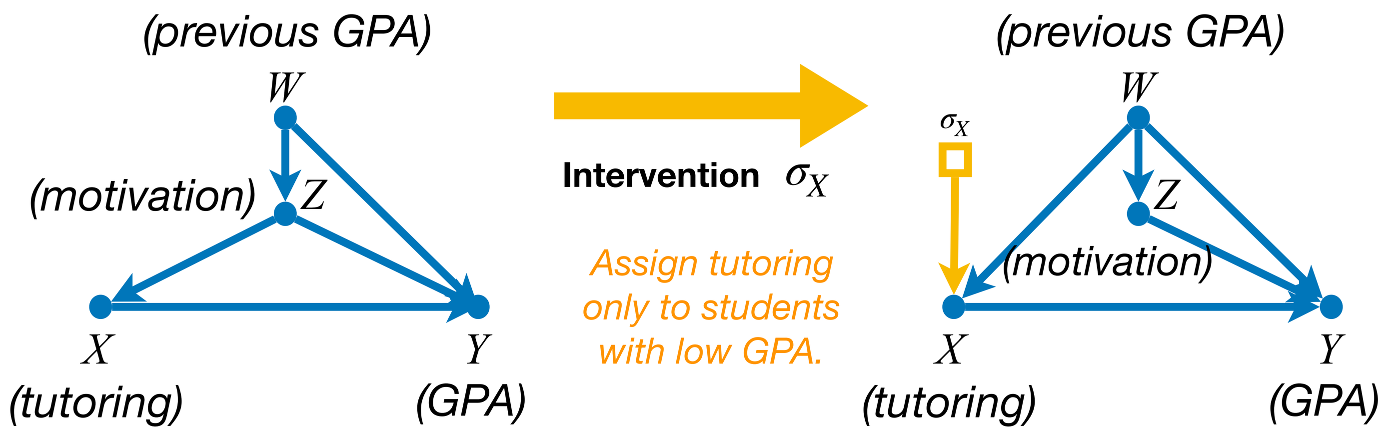
1. Formally encode the assumptions about the differences between environments. → Selection diagrams (with □)
2. Identify the mechanisms that are stable across environments.
3. Determine the variables that need to be re-measured in the target.
4. Construct an estimator using the collected data.

(2–4: Exploit Causality Theory)

6 Task: Identifying Stochastic Interventions

If the source environment corresponds to the current system, and the target environment corresponds to the source after an intervention, then transporting the distribution $P^*(\boldsymbol{y})$ is the same as identifying the effect of the intervention on an outcome $\boldsymbol{Y}$.

Students get tutoring on their own volition based on their motivation.

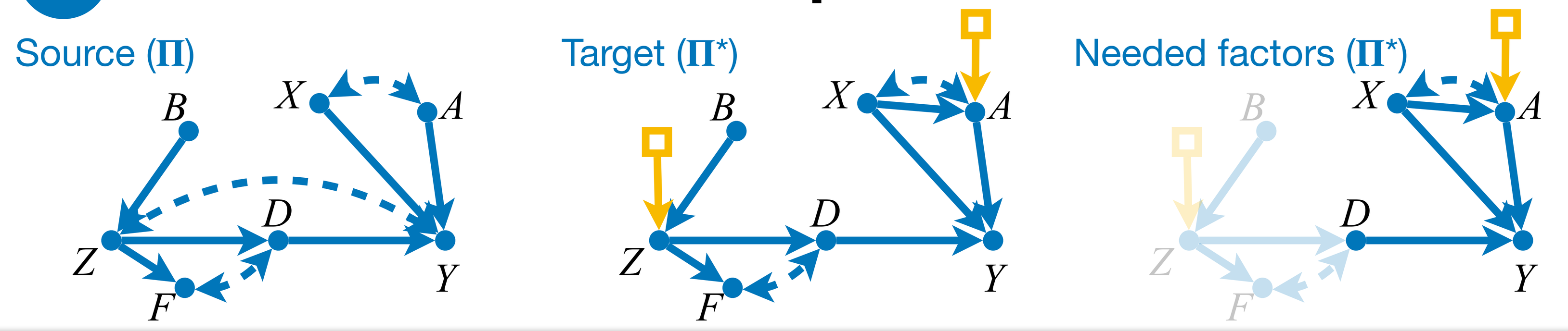


$P^*(y)$ represents the effect of σ_X on Y.

7 Results

1. We derived a novel graphical decomposition of the observed/learned distribution into factors that take into account the latent structure, suitable to reason about distributions with different sets of measured variables.
2. We developed a complete algorithm that determines if a distribution $P^*(\boldsymbol{y}|\boldsymbol{x})$ can be uniquely identified from distributions $P(\boldsymbol{v})$ and $P^*(\boldsymbol{w})$, $\boldsymbol{W} \subseteq \boldsymbol{V}$, based on the stable mechanisms shared across source and target domains.
3. Leveraging these results, we solve the problem of identifying the effect of stochastic interventions, which generalizes the corresponding do-calculus counterpart.

8 A more involved example



- Suppose the target is $P^*(y|x,z)$. After some algebra, and given $P(b,z,f,d,x,a,y)$ and $P^*(x,a)$, one can show that

$$P^*(y \mid x,z) = \sum_a P^*(a \mid x) \sum_d P(d \mid z) \sum_{z'} P(y \mid x,z',d,a) P(z')$$