

Adaptive Multi-Source Causal Inference from Observational Data

Thanh Vinh Vo ¹

Pengfei Wei ²

Trong Nghia Hoang ³

Tze-Yun Leong ¹

¹ National University of Singapore

² AI Lab Speech & Audio Bytedance, Singapore

³ Washington State University

CIKM 2022, Atlanta, Georgia, USA



- 1 Motivation
- 2 Causal quantities of interest
- 3 AdaTRANS: the proposed method
- 4 Experimental evaluation
- 5 Summary: limitations and future work

Why do we need causal inference?

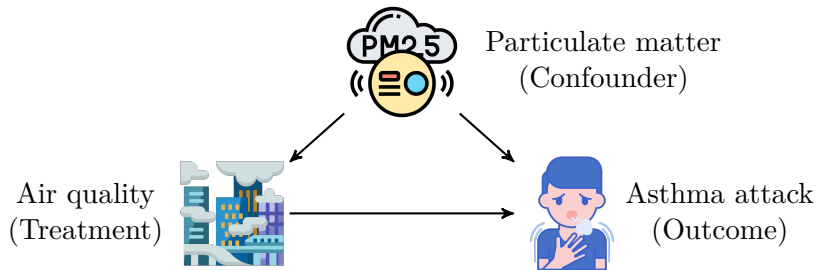
- Effect of a ‘new medicine’ on ‘blood pressure’ of patients.

Why do we need causal inference?

- Effect of a ‘new medicine’ on ‘blood pressure’ of patients.
- Effect of ‘air quality’ on ‘asthma attack’.
- Effect of ‘smoking’ on ‘cancer’.
- Effect of ‘coronary heart disease’ on ‘mortality’.
- Effect of ‘fertilizer’ on ‘crop yield’.

Typical regression would give a biased estimand because of confounders.

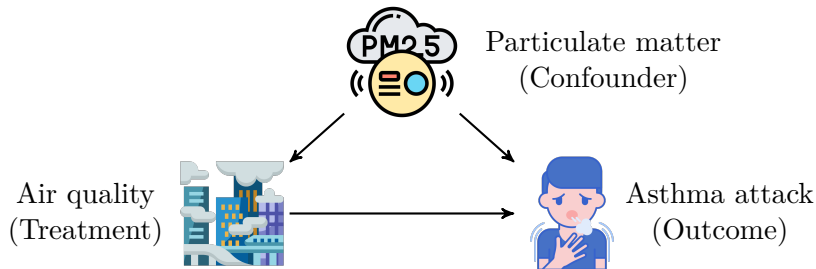
Motivation



Two approaches to estimate causal effects:

- Randomized control trial
- Inference from observational data
 - Potential outcomes framework (PO) ([Rubin, 1974, 1975, 1976, 1977, 1978](#); [Rosenbaum & Rubin, 1983](#))
 - Structural causal model (SCM) ([Pearl, 1995, 2000](#))

Motivation

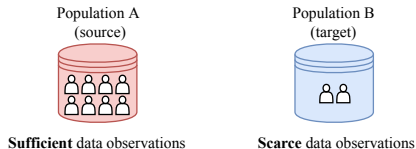


Problem:

- Observational data in a specific population might be **scarce**.
- For example:
 - Vaccination data of the elder in a country might be scarce and much less than the younger.

Motivation

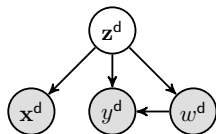
- Estimate causal effects in a target population with scarce data observation



Combining data might lead to poor causal estimands.

		Observed confounder	Latent confounder	With transfer	Randomized data
Without transfer	Louizos et al. (2017)		✓		
	Madras et al. (2019)		✓		
	Hill (2011)	✓			
	Shalit et al. (2017)	✓			
	Künzel et al. (2019) (to name a few)	✓			
With transfer	Bareinboim & Pearl (2014)			✓	
	Bareinboim & Pearl (2016)			✓	
	Aglietti et al. (2020)			✓	✓
	AdaTRANS (proposed method)		✓	✓	

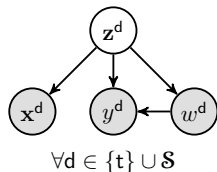
Causal quantities of interest



$\forall d \in \{t\} \cup \mathcal{S}$

- t : target population
- $\mathcal{S} = \{s_1, s_2, \dots, s_m\}$: collection of source populations
- y^d : the outcome
- w^d : the treatment
- z^d : the latent confounder
- x^d : the covariate

Causal quantities of interest



- t : target population
- $\mathcal{S} = \{s_1, s_2, \dots, s_m\}$: collection of source populations
- y^d : the outcome
- w^d : the treatment
- z^d : the latent confounder
- x^d : the covariate

We estimate

$$\text{ite}(x) = E[y^t | \text{do}(w^t = 1), \mathbf{x}^t = x] - E[y^t | \text{do}(w^t = 0), \mathbf{x}^t = x]$$

$$\text{ate} = E_X[\text{ite}(X)]$$

The proposed method

Expectation of the outcome given intervention on w_i^t : $E[y_i^t | \text{do}(w_i^t), \mathbf{x}_i^t]$

The proposed method

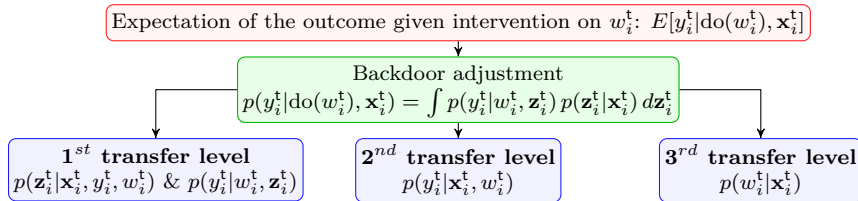
Expectation of the outcome given intervention on w_i^t : $E[y_i^t | \text{do}(w_i^t), \mathbf{x}_i^t]$



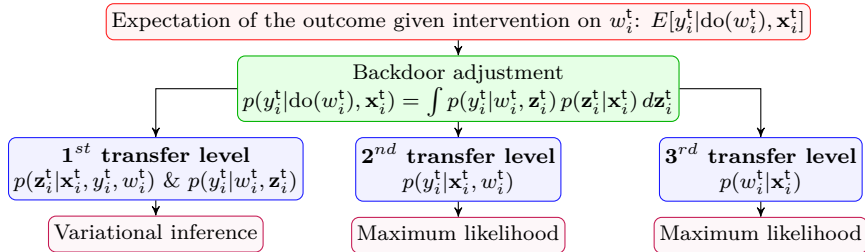
Backdoor adjustment

$$p(y_i^t | \text{do}(w_i^t), \mathbf{x}_i^t) = \int p(y_i^t | w_i^t, \mathbf{z}_i^t) p(\mathbf{z}_i^t | \mathbf{x}_i^t) d\mathbf{z}_i^t$$

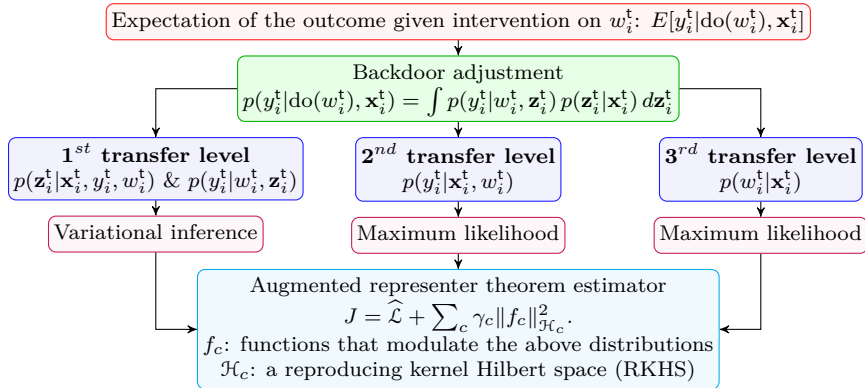
The proposed method



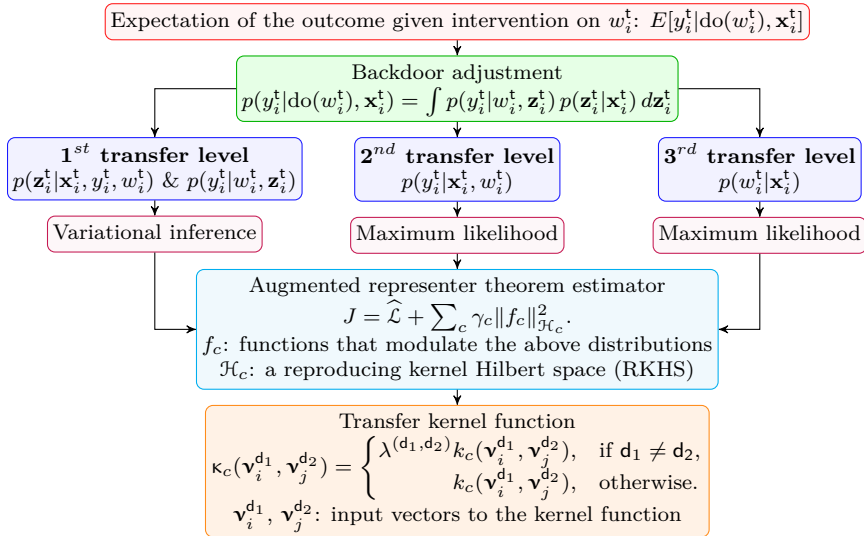
The proposed method



The proposed method



The proposed method



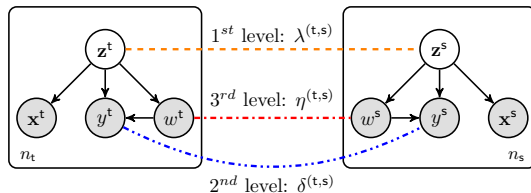
The proposed method

The transfer kernel function

$$\kappa_c(\mathbf{v}_i^{\mathbf{d}_1}, \mathbf{v}_j^{\mathbf{d}_2}) = \begin{cases} \lambda^{(\mathbf{d}_1, \mathbf{d}_2)} k_c(\mathbf{v}_i^{\mathbf{d}_1}, \mathbf{v}_j^{\mathbf{d}_2}), & \text{if } \mathbf{d}_1 \neq \mathbf{d}_2, \\ k_c(\mathbf{v}_i^{\mathbf{d}_1}, \mathbf{v}_j^{\mathbf{d}_2}), & \text{otherwise.} \end{cases}$$

The transfer factor $\lambda^{(\mathbf{d}_1, \mathbf{d}_2)}$ is learned.

- $\lambda^{(\mathbf{d}_1, \mathbf{d}_2)} = 0$: no transfer
- $\lambda^{(\mathbf{d}_1, \mathbf{d}_2)} = 1$: fully transfer
- $0 < \lambda^{(\mathbf{d}_1, \mathbf{d}_2)} < 1$: partially transfer



The aims of experiments

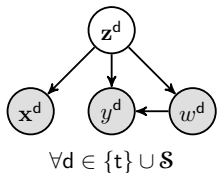
- We study the performance when the sources' distributions and the target's distributions are similar/different.
- We illustrate the performance when the similarity of sources and target slowly changes.

Experiments

Datasets

■ Synthetic data

- Simulate based on a ground truth causal graph.



$$\mathbf{z}_i^d \sim \mathcal{N}(\mathbf{0}, \sigma_z^2 \mathbf{I}_2),$$

$$x_{ij}^d \sim \text{Bern}(\varphi(a_{0j} + (\mathbf{z}_i^d)^\top \mathbf{a}_{1j})),$$

$$w_i^d \sim \text{Bern}(\varphi(b_0 + (\mathbf{z}_i^d)^\top \mathbf{b}_1^d)),$$

$$y_i^d(0) \sim \mathcal{N}(\zeta(c_0 + (\mathbf{z}_i^d)^\top \mathbf{c}_1^d), \sigma_y^2),$$

$$y_i^d(1) \sim \mathcal{N}(\zeta(d_0 + (\mathbf{z}_i^d)^\top \mathbf{d}_1^d), \sigma_y^2).$$

- We keep \mathbf{x}_i^d , w_i^d and either $y_i^d(0)$ or $y_i^d(1)$ as observed data.
- \mathbf{b}_1^d , \mathbf{c}_1^d , \mathbf{d}_1^d are set differently on each population d .

Datasets

- Twins data
 - Study the impact of twins' weight (treatment) on their mortality (outcome).
 - Source data: 1594 entries, Target data: 457 entries

Evaluation metrics

- Precision in estimation of heterogeneous effects (PEHE) ([Hill, 2011](#))

$$\varepsilon_{\text{PEHE}} = E \left[\left(\underbrace{(y_1 - y_0)}_{\text{True ITE}} - \underbrace{(\hat{y}_1 - \hat{y}_0)}_{\text{Estimated ITE}} \right)^2 \right]$$

- Absolute error

$$\varepsilon_{\text{ATE}} = \left| \underbrace{E(y_1 - y_0)}_{\text{True ATE}} - \underbrace{E(\hat{y}_1 - \hat{y}_0)}_{\text{Estimated ATE}} \right|$$

Experiments

The importance of adaptively causal transfer

$$\mathbf{b}_1^t = [1.1, 1.7]^\top,$$

$$\mathbf{b}_1^s = \mathbf{b}_1^t + \Delta^s [1, 1]^\top,$$

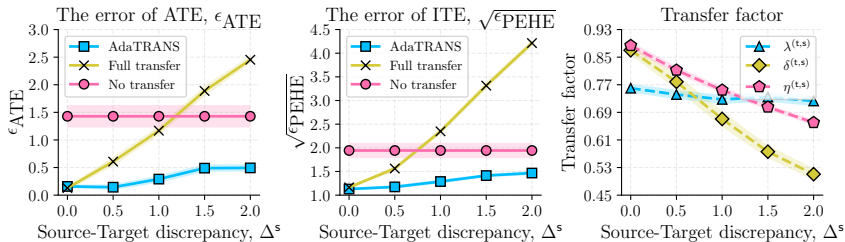
$$\mathbf{c}_1^t = [1.5, 1.8]^\top,$$

$$\mathbf{c}_1^s = \mathbf{c}_1^t + \Delta^s [1, 1]^\top,$$

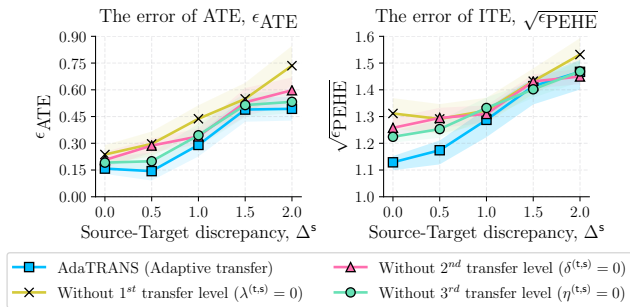
$$\mathbf{d}_1^t = [1.5, 2.8]^\top,$$

$$\mathbf{d}_1^s = \mathbf{d}_1^t + \Delta^s [1, 1]^\top,$$

We vary $\Delta^s \in \{0.0, 0.5, 1.0, 1.5, 2.0\}$ to obtain different instances of the source data.



Which transfer level is the most important?



Performance analysis: compare with the baselines (Lower is better)

Method	The error of ITE ($\sqrt{\varepsilon_{\text{PEHE}}}$)			The error of ATE (ε_{ATE})		
	0-source	2-sources	4-sources	0-source	2-sources	4-sources
CEVAE _{stack}	3.1±.30	4.6±.39	4.8±.40	1.7±.29	2.8±.30	2.5±.26
CFRNet _{stack}	4.6±.51	8.9±.50	6.0±.19	1.6±.41	6.1±.48	4.0±.17
SITE _{stack}	6.0±.98	8.9±.61	7.5±.60	3.3±.67	6.4±.79	5.0±.76
BART _{stack}	2.5±.06	2.3±.03	2.2±.06	1.2±.13	0.7±.08	0.6±.09
R-learner _{stack}	3.0±.27	2.2±.11	1.8±.09	1.4±.35	1.2±.17	1.0±.10
X-learner _{stack}	2.0±.13	2.2±.12	1.9±.13	1.0±.17	1.0±.11	1.1±.13
OrthoRF _{stack}	6.2±.40	2.4±.03	2.2±.03	1.2±.37	0.5±.08	0.6±.06
CEVAE _{1-hot}	—	5.0±.43	3.3±.12	—	3.1±.42	1.9±.23
CFRNet _{1-hot}	—	4.4±.26	3.3±.21	—	3.3±.26	2.1±.17
SITE _{1-hot}	—	5.8±.99	3.2±.25	—	3.4±.67	2.1±.21
BART _{1-hot}	—	2.3±.03	2.2±.04	—	0.7±.10	0.4±.10
R-learner _{1-hot}	—	2.0±.07	1.7±.15	—	0.8±.15	0.8±.20
X-learner _{1-hot}	—	1.9±.12	1.8±.10	—	0.7±.13	0.6±.12
OrthoRF _{1-hot}	—	5.5±.30	4.1±.16	—	3.9±.22	2.6±.17
AdaTRANS	1.6±.09	1.3±.03	1.3±.02	1.1±.13	0.2±.05	0.1±.03

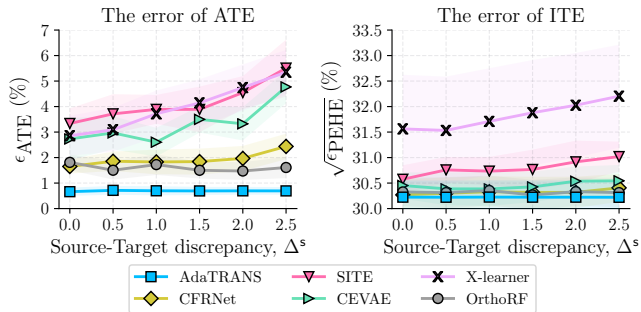
■ The range of true ATE: (1.51, 1.87).

■ The range of true ITE: (-5.69, 13.14).

Experiments

■ Twins data

- Study the impact of twins' weight (treatment) on their mortality (outcome).
- Source data: 1594 entries, Target data: 457 entries



- We developed an adaptive method that estimates causal effects for a target population whose data is scarce.
- The proposed method required no prior information about the discrepancy about the source and target population.
- We assume that the source and the target share the same causal graph, but different structural equations.
- Limitations:
 - Causal effects in each population are identifiable.
 - The populations share similar causal graph & data features.
 - The confounders are independent and identically distributed.
- Future direction: allowing in the target population to be unidentifiable and have different set of data features.

Q & A

This research/project is supported by the National Research Foundation Singapore and DSO National Laboratories under the AI Singapore Programme (AISG Award No: AISG2-RP-2020-016).