Learning Unknown Intervention Targets in Structural Causal Models from Heterogeneous Data

Yuqin Yang[1] **Saber Salehkaleybar**[2] **Negar Kiyavash**[3]

[1]Georgia Tech [2]Leiden University [3]EPFL

Motivation

Causal structure learning from **interventional data**

- We may not fully control the intervention target
- Intervention is done by an unknown source

Task: Learn intervention targets from multi-domain data

Existing method: (possibly a byproduct)

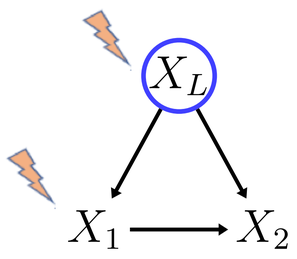
- Limited to linear systems
- Requiring exponential CI / invariance tests
- Unable to handle latent confounders

Model Description

- SCM: $X_i = f_i(PA_i, N_i),\ X_i \in \mathbf{X}$, not given as input
  - Partitioned into $[\mathbf{O}; \mathbf{L}]$ under latent confounding
- Soft intervention: $X_i = f_i(PA_i, N_i')$
- We collect data from D domains

$$\mathbf{T} := \{X_i | \exists d, d' \in [D], p_d(N_i) \neq p_{d'}(N_i)\}$$

- Goal: Recover intervention target set $\mathbf{T}$ ($\mathbf{T} \cap \mathbf{O}$)



- Two obs. variables, one latent
- Two noises change across environments

We propose **Locating Intervention Target (LIT)** algorithm which includes Recovery phase and Matching phase.

Recovery Phase

Recover the noises $\mathbf{N_T} = \{N_i | X_i \in \mathbf{T}\}$ up to permutation and component-wise invertible transformations using contrastive learning approach.

- Mixing function: $\mathbf{X} = \mathbf{g}(\mathbf{N})$
- Auxiliary / domain variable U

Proposition 1: Assume $\min(D-1, |\mathbf{O}|) \geq |\mathbf{T}|$. Under certain conditions on $\mathbf{N}$, the recovery is possible when:

a) $\mathbf{L} = \emptyset$ and $\mathbf{g}$ is invertible;

b) $\mathbf{L} \neq \emptyset$ and $\exists$ invertible $\tilde{\mathbf{g}} : \mathbb{R}^{|\mathbf{O}|} \to \mathbb{R}^{|\mathbf{O}|}$ such that $\tilde{\mathbf{g}}(\mathbf{O}) = (\mathbf{N_T}; \mathbf{V})$, where $\mathbf{V} \perp\!\!\!\perp U$ and $\mathbf{V} \perp\!\!\!\perp \mathbf{N_T} | U$.

Invertibility holds when the model is a linear SCM, nonlinear ANM, or $\{f_i\}$ are MLPs with ReLU activation function and positive coefficients.

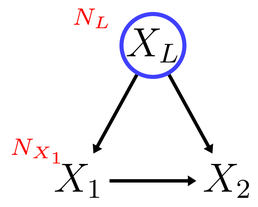
Matching Phase

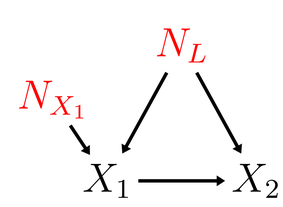
Match each recovered noise in $\tilde{\mathbf{N}}_\mathbf{T}$ to its corresponding observed variable (if such a variable exists)

- A noise may correspond to a latent confounder.
- An observed variable does not correspond to any recovered noise if it is not in $\mathbf{T}$.

Matching Phase (Cont'd)

Causal Diagram





Augmented Graph

T-faithfulness assumption

d-separation between noise and observed variable on the augmented graph is equivalent to independency.

- **Indicator set** $\mathcal{I}$: $\mathbf{I}_i = \{\tilde{N}_j | \tilde{N}_j \not\perp\!\!\!\perp X_i\}$
  - Includes all noises in $An(X_i) \cap \mathbf{N_T}$

Matching under causal sufficiency

Theorem 1: The intervention targets can be uniquely identified based on $\tilde{\mathbf{N}}_\mathbf{T}$, $\mathbf{X}$ and $\mathcal{I}$ using three conditions.

- LIT: Checking Cond. (I) - (III) for all variables
- Requires quadratic CI tests: Bounded by $|\mathbf{T}| \cdot |\mathbf{X}|^2$

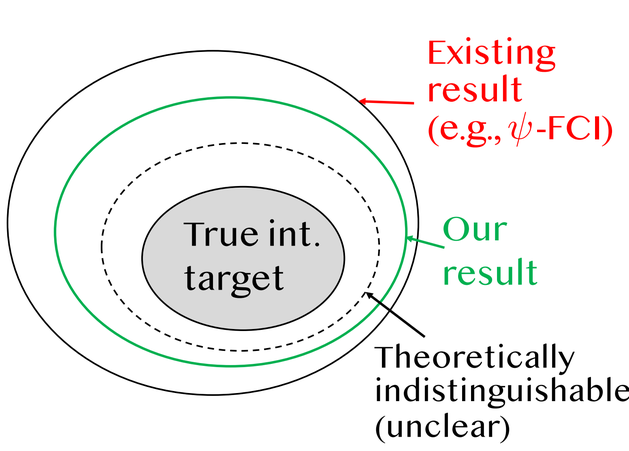
Algorithm 1: LIT algorithm

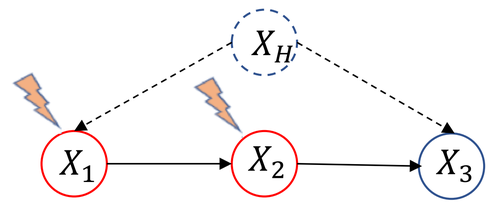
1 Obtain $\tilde{\mathbf{N}}_\mathbf{T}$ and $\mathcal{I}$; $\mathbf{U} \leftarrow \mathbf{X}$; $\mathbf{K} \leftarrow \emptyset$;
2 Check Cond. (I), (IV), (II) for every X_i. Add X_i to $\mathbf{K}$ if it satisfies (II). Otherwise, exclude it from $\mathbf{K}$;
3 Partition $\mathbf{U}$ into disjoint subsets $\mathbf{U}_1, \cdots, \mathbf{U}_r$ according to the indicator sets;
4 For every $\mathbf{U}_i$, add $X_{i_k} \in \mathbf{U}_i$ satisfying Cond. (III) to $\mathbf{K}$ (resp. exclude variables satisfying (III-L));
5 **return** $\mathbf{K}$

Matching under latent confounding

Theorem 2: By adding Cond. (IV) and changing (III), LIT can return a superset of the true intervention targets.

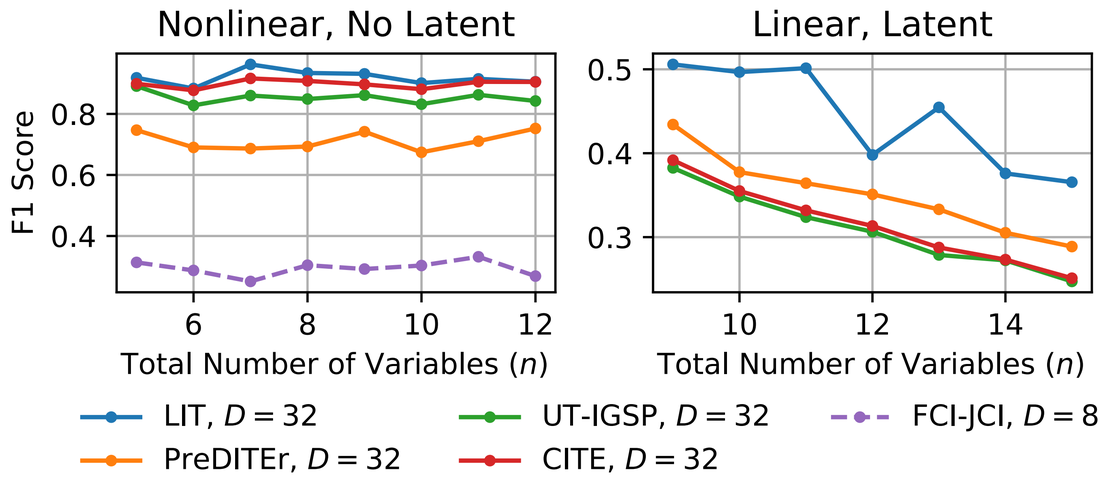
- Graphical characterization: Auxiliary graph
- Can handle latent intervention targets, i.e., $\mathbf{T} \cap \mathbf{L} \neq \emptyset$
- More informative than baselines when $\mathbf{T} \cap \mathbf{L} = \emptyset$





- LIT: Recovers the targets
- $\mathbf{I}_1 = \{\tilde{N}_1\}, \mathbf{I}_2 = \mathbf{I}_3 = \{\tilde{N}_1, \tilde{N}_2\}$
- $\tilde{N}_2 \perp\!\!\!\perp X_3 | X_1, X_2$
- Existing work (ψ-FCI): Returns $\{X_1, X_2, X_3\}$

Simulations



- Compare the recovery of the intervention targets
- PreDITEr: linear Gaussian; UT-IGSP: causal suff.; CITE: both
- # of CI tests: LIT ~80; PreDITEr ~30000; CITE ~20000