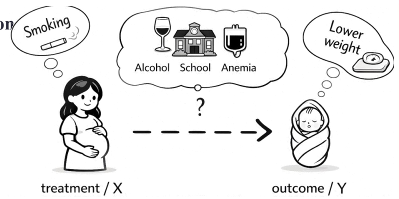
Local Covariate Selection for Average Causal Effect Estimation without Pretreatment and Causal Sufficiency Assumptions

Zeyu Liu, Zheng Li, Feng Xie, Yan Zeng, Hao Zhang, Kun Zhang

ICML International Conference On Machine Learning

Motivation

- Existing local methods rely on:
  - Pretreatment assumption
  - Causal sufficiency assumption
  - Global structure learning
- Problems:
  - Unrealistic in practice
  - Miss valid adjustment sets
  - Computationally expensive



Goal:
Develop a sound & complete local method for causal effect estimation with latent confounders.

Method (local)	Soundness	Completeness	Without Pretreatment
CEELS	✓	✗	✗
LSAS	✓	✓	✗
LDP	✓	✗	✓
LCS (Ours)	✓	✓	✓

Core concept

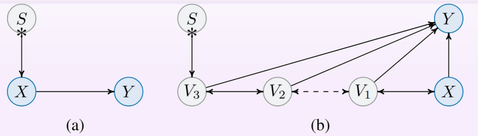
Generalized Adjustment Criterion

A set Z is a valid adjustment set for estimating the causal effect from X to Y if :

1. G is adjustment amenable relative to (X, Y), which means that the first edge of every possible directed path from Xto Y is visible,

2. Z ∩ Forb(X, Y) = ∅ , means that Z cannot contain the descendants of any node on any possible directed path from X to Y , and

3. all definite status non-causal paths from X to Y are blocked by Z.



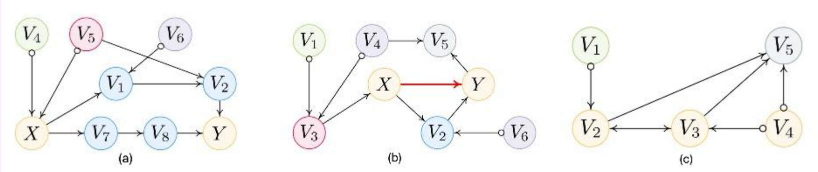
Visible Edge

A directed edge X → Y is visible if there exists a node S not adjacent to Y such that: S has an edge into X, or there exists a collider path from S to X into X, where all intermediate nodes are parents of Y. Otherwise, the edge is called invisible.

Theory

With the background knowledge above, we first give Theorem 1, which narrows the selection range of the adjustment set to MB(X).

Theorem 1 Let O be the set of observed variables, and let (X, Y) be a pair of target variables in O. Then, there exists a subset of O that is a valid adjustment set for estimating the average causal effect of X on Y if and only if there exists a subset of MB(X) that is a valid adjustment set for X and Y . (e.g. Fig(a))



(a) demonstrates that MB(Y) is insufficient, while R1 works. (b) presents a case where R1 fails but R2 succeeds. (c) demonstrates Rule 3.

identifiable causal effect

RULE ONE : Let P be the PAG over O, (X, Y)be a pair of ordered target variables. A subset Z ⊆ MB(X) \ PossDe(X, P) is a valid adjustment set w.r.t. (X, Y) if there exists a variable S ∈ MB(X) \ ({Y } ∪ Ch(X, P)) such that

(i) S /⊥ Y | Z and (ii) S ⊥ Y | Z ∪ {X}.

✓ **CASE 1-1: Two CI tests can determine the presence of a causal effect**

✗ **CASE 1-2: A special case**

In Fig. (b), Rule 1 fails to identify a valid adjustment set, while Rule 2 succeeds. Before introducing Rule 2, we first present a necessary definition.

Definition 1. Let $P_{MB^+}(X)$ be the induced subgraph of MB+(X). The definitely non-collider possible parents of X are the nodes that are adjacent to X via an edge of the form V0 ∘→ X, where V0 does not act as a collider on any path within this subgraph. This set is denoted as NCPa(X, $P_{MB^+}(X)$).

RULE TWO : Let P be the PAG over O, (X, Y) be a pair of ordered target variables. Suppose that in the local adjacency structure around X, the mark at the X-endpoint is always determined. If the causal effect from X to Y is identifiable, then **the set Z = Pa(X) ∪ NCPa(X, P_{MB^+}(X)) is a valid adjustment set w.r.t. (X, Y)** for identifying the total effect of X on Y , and it satisfies Z ⊆ MB(X) \ PossDe(X, P)

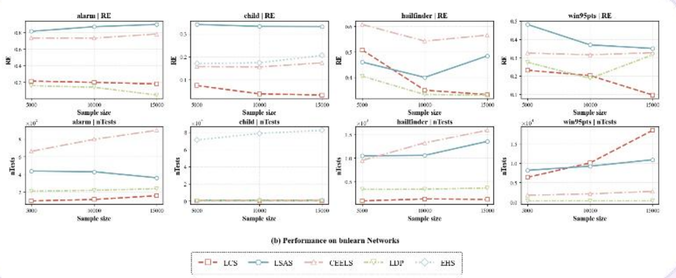
identifiable zero effect

RULE THREE : Let P be the PAG over O, (X, Y) be a pair of ordered target variables. Then, X has no causal effect on Y if there exists a subset Z ⊆ MB(X) with Z ∩ PossDe(X, P) = ∅ and a variable S ∈ MB(X) \ ({Y } ∪ Ch(X, P)) such that at least one of the following conditions holds: **(i) X ⊥ Y | Z, or (ii) S /⊥ X | Z and S ⊥ Y | Z.** (e.g. Fig(c) : (V_2, V_4); (V_2, V_3))

non-identifiable

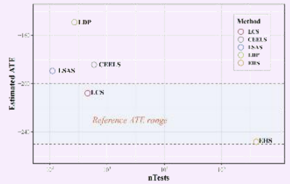
Theorem 5. Under causal Markov and Faithfulness assumptions, the causal effect of X on Y must fall into one of the three identifiable cases characterized by Theorem 2, Theorem 3, and Theorem 4. If none of these conditions hold, then the effect is not identifiable from observable conditional independence and dependence relations.

Simulation



We evaluate LCS on synthetic data using random graphs and bnlearn benchmarks. LCS outperforms existing methods across almost all metrics and sample sizes.

Application to Real-World Data:



Method	ATT	Bias (%)	nTests
EHS	930.606	5.04	1732
LSAS	989.346	11.67	176
CEELS	683.406	22.87	183
LDP	-13597.588	1635.98	61
LCS (Ours)	**854.528**	**3.55**	303

On the Cattaneo2 dataset, LCS remains within the reference range.
On the Jobs dataset, LCS achieves the lowest ATT estimation bias (3.55%).

✓ Conclusions and Future work

Conclusions

- Characterize a local boundary containing a valid adjustment sets and develop local identification rules.
- Propose a sound and complete data-driven algorithm for identifying adjustment sets.
- Experiments on synthetic and real-world datasets demonstrate the effectiveness and efficiency of LCS.

Future work

- incorporate expert knowledge and address selection bias for greater robustness.