FINDING INVARIANT PREDICTORS EFFICIENTLY VIA CAUSAL STRUCTURE

Kenneth Lee, Md Musfiqur Rahman, Murat Kocaoglu

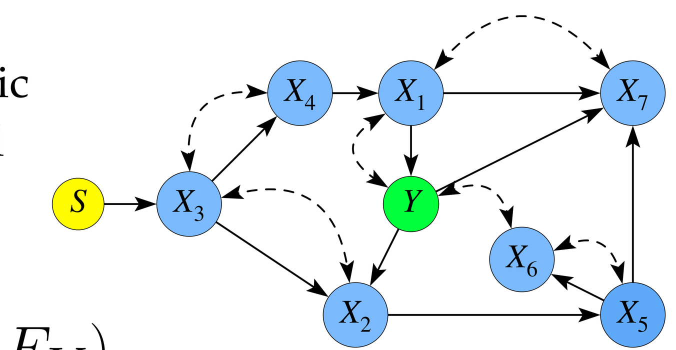
Elmore Family School of Electrical and Computer Engineering, Purdue University

INTRODUCTION

- **Objectives:** Find a set of features that are invariant to the distribution shifts for predicting the target variable *using distributional invariance via Pearl's do-calculus.*
- **Motivation:** An existing state-of-the-art approach called Graph Surgery Estimator (GSE) *takes exponential time to search for this set of features* [Adarsh et al. '19].
- **Contributions:** Develop a sound and polynomial-time algorithm that searches for surgery estimators, if any.

BACKGROUND

- *Pearlian framework* [Pearl' 09]:
  Acyclic Directed mixed graphs (ADMGs) encode causal relation between variables.
- Arrows: Non-deterministic functional relations called *structural equations.*

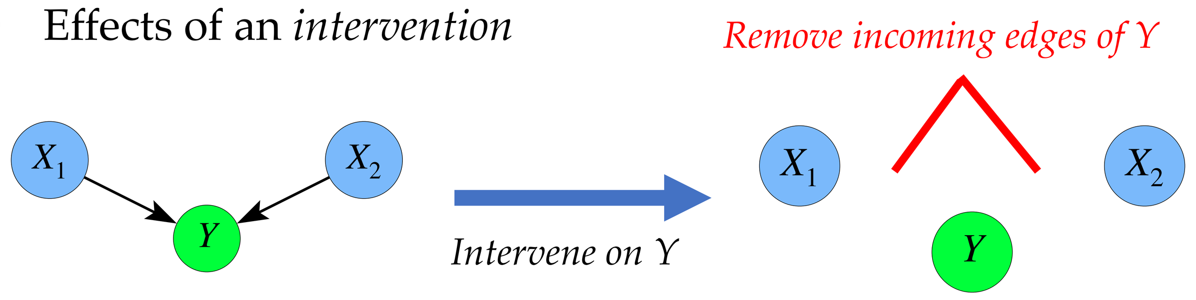


$$Y = f(X_1, U_Y^1, U_Y^2, E_Y)$$

- In general $X_i = f(Pa_{X_i}, U_{X_i}, E_{X_i})$
- *A path* p between X and Y is *active relative to* Z if
  - all non-colliders on p are not in **Z** and
  - all colliders on p are ancestors of some Z in **Z.**
- Assumptions:
  - An ADMG is given a priori with the known *selection variable* (yellow), a set of predictors (blue) and the target (green).

EFFECTS OF INTERVENTIONS

- Effects of an *intervention*



$Y = f(X_1, X_2, E_Y)$ $\qquad$ $Y = y$

CONTACT INFORMATION

Web: muratkocaoglu.com/CausalML

Email: lee4094@purdue.edu

Code: github.com/kenneth-lee-ch/id4ip

Scan for paper

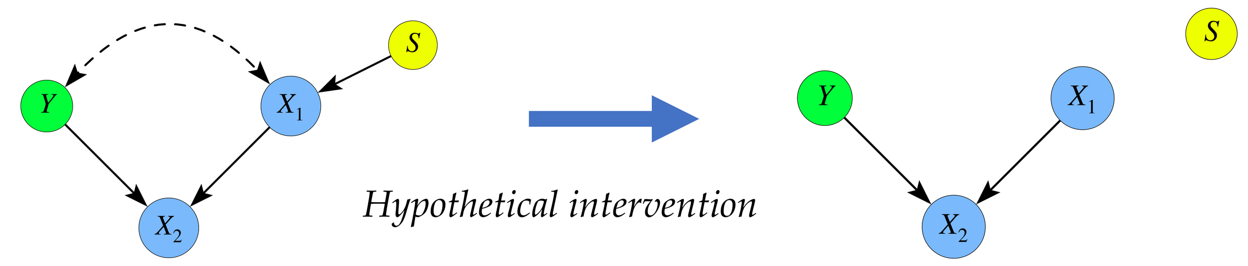
GRAPH SURGERY ESTIMATOR

- A sound and complete algorithm named **ID** that uses do-calculus rules to *identify a predictive model from interventional distribution* **[Tian et al. '08] .**
- **Rule 2:** $P(Y|do(X,Z),W) = P(Y|do(X),Z,W)$ if $(Y \perp\!\!\!\perp Z|X,W)_{G_{\overline{X}\underline{Z}}}$
- Two others

$P(Y|do(\mathbf{Q}),\mathbf{W})$ is called *a graph surgery estimator* if

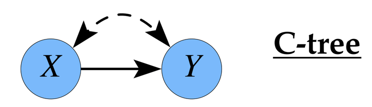
1. $P(Y|do(\mathbf{Q}),\mathbf{W})$ is identifiable from observational distribution and
2. $(Y \perp\!\!\!\perp S|\mathbf{W})_{G_{\overline{\mathbf{Q}}}}$ and
3. $P(Y|do(\mathbf{Q}),\mathbf{W})) \neq P(Y)$

Example: $P(Y|do(X_1),X_2) = \frac{P(X_2|X_1,Y)P(Y)}{\sum_Y P(X_2|X_1,Y)P(Y)}$



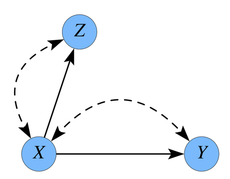
- Interventional distribution is *invariant* to changes in how X1 is generated.
- **How to search for different surgery estimators?**
  - Search through all possible Q and W [Adarsh et. al '19].

IDENTIFIABILITY AND HEDGE



C-tree

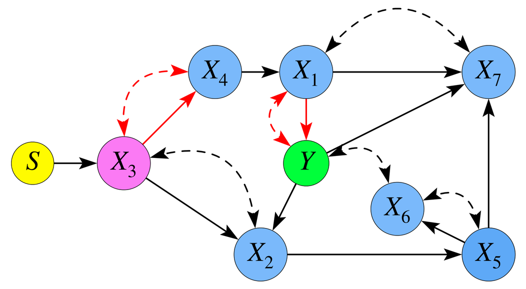
- Each node has at most one child
- Only one vertex has no children



C-forest

- Each node has at most one child
- More than one vertex has no children

- Causal identifiability relates to a specific graphical structure called *hedge* [Shiptser et al. '08]
- **Generalized hedge condition for G =(V, E)**
  - Let $\mathbf{Z} \subseteq \mathbf{V}$ be the maximal subset such that $P(\mathbf{Y}|do(\mathbf{Q}),\mathbf{W}) = P(\mathbf{Y}|do(\mathbf{Q},\mathbf{Z}),\mathbf{W} \setminus \mathbf{Z})$. There is a hedge for $P(\mathbf{Y}|do(\mathbf{Q}),\mathbf{W})$ if
    - Two **R**-rooted C-forests F, F' exist where **R** is in $An(\mathbf{Y} \cup (\mathbf{W} \setminus \mathbf{Z}))_{G_{\overline{\mathbf{X} \cup \mathbf{Z}}}}$ and **Q** is in F but not in F' in G.
- Example: $P(Y|do(X_3))$
  - $F = \{X_3, X_4\}$
  - $F' = \{X_4\}$



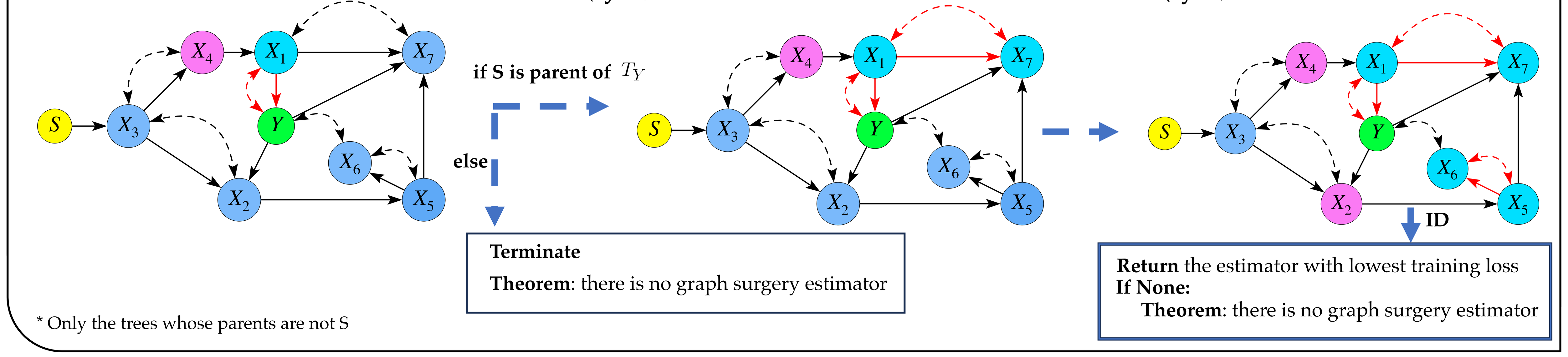
- **Theorem:** Generalized hedge ⇔ Unidentifiability
  - Helps avoiding searching for unidentifiable queries.

PROPOSED ALGORITHM: ID4IP

- There exists a complete algorithm that finds conditional independence query between two sets of variables [Shiptser et al. '08] .
- Greedy feature selection on the selected conditioning sets.
- **Theorem:** Each predictor found is guaranteed to be identifiable and ID4IP finds at least one predictor if exists.

ID4IP algorithm:

1. Find Y-rooted C-tree T_Y **(red)** and intervene on its parents **(purple)** with conditioning on members of the tree **(cyan).**
2. Find each Ch(Y)-rooted C-trees* union with Y-rooted C-tree **(red)** and intervene on its parents **(purple)** with conditioning on members of the tree **(cyan).**
3. Find each bidirected-nbr(Y)-rooted C-trees* union with previous C-trees **(red)** and intervene on its parents **(purple)** with conditioning on members of the tree **(cyan).**

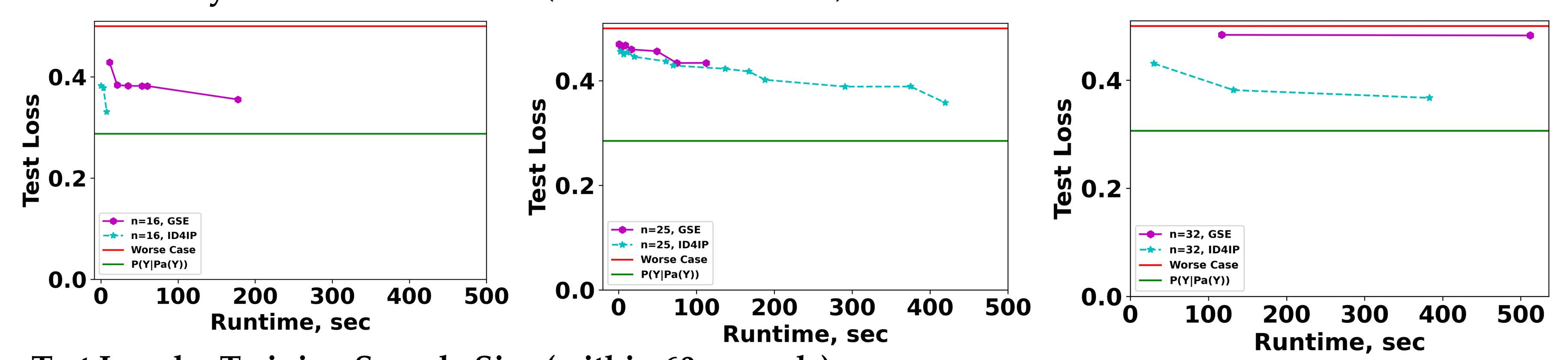


* Only the trees whose parents are not S

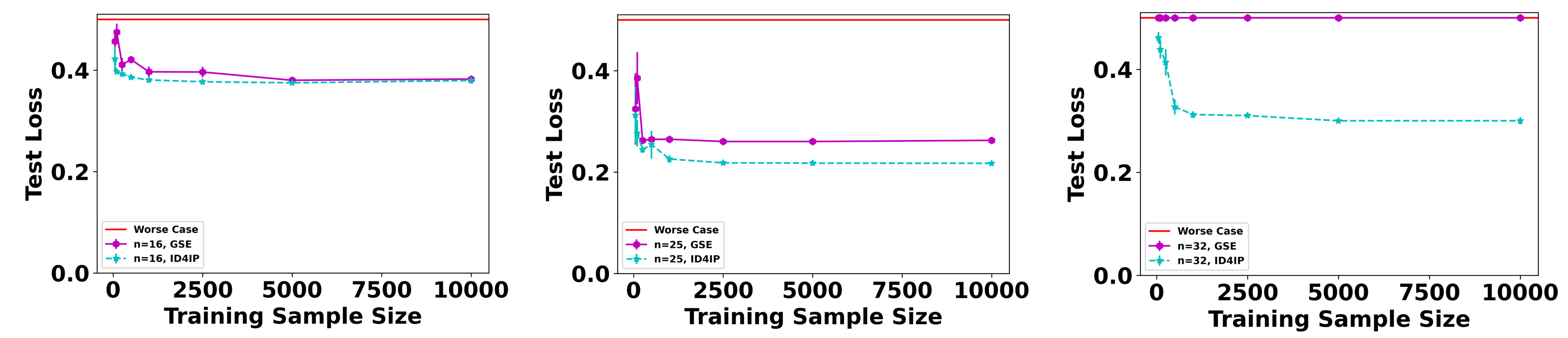
EXPERIMENTAL RESULTS

(Simulated) (Semi-synthetic)

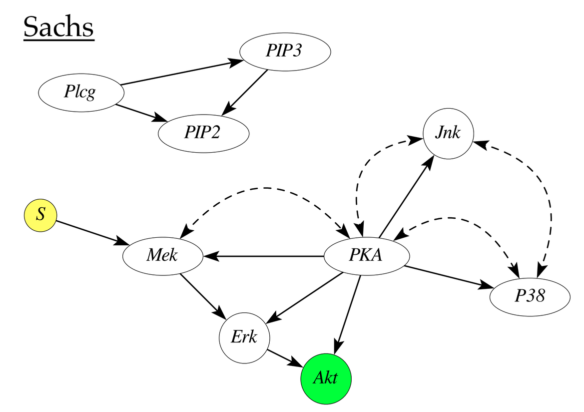
- **Test Loss by Runtime Restriction (within 600 seconds):**



- **Test Loss by Training Sample Size (within 60 seconds):**



- Left to right: **n= 16, 25, 32 with n/2 latent confounders** and **3n directed edges** randomly generated.
- Each variable is **binary.**
- Selection variable is randomly assigned and the marginal of the child will be changed in the test distribution.
- Set the test loss to be 0.5 if the model fails to find a predictor.
- For the top row, we use **population distribution** to train both models.
- For the bottom row, turned all bidirected edges to directed edges if their children has no directed edge to learn the training distribution using greedy hill algorithm.



Algorithm	Sachs (11 nodes)	Alarm (37 nodes)
GSE	0.80	0.57
ID4IP	0.80	**0.83**
Logistic regression	0.53	0.52

F1 score **within 120 seconds** for two real-world datasets: Sachs and Alarm.

- Original graphs were altered to introduce latents and selection variable.
- **Practical scenarios:**
  - Medical record transfer [Agniel et.al 2018]

CONCLUSION & FUTURE WORK

- We utilize a graphical characterization of the identifiability of conditional causal queries with greedy search to increase the efficiency of finding invariant predictors.
- Our algorithm is sound that runs in polynomial time in contrast to the existing work that requires exponential time.
- For future work, several directions are worth pursuing: improve the algorithm for completeness, approximation guarantees for greedy-search methods for invariant causal prediction, and combining with causal discovery algorithms.

REFERENCES

- J. Pearl, Causality: Models, Reasoning and Inference. Cambridge University Press, 2009.
- Shpitser, Ilya, and Judea Pearl. "Complete identification methods for the causal hierarchy." Journal of Machine Learning Research 9 (2008): 1941-1979.
- Shpitser, Ilya, and Judea Pearl. Dormant independence. AAAI Press, 2008.
- Tian, Jin, and Judea Pearl. "A general identification condition for causal effects." AAAI. 2002.
- Subbaswamy, Adarsh, Peter Schulam, and Suchi Saria. "Preventing failures due to dataset shift: Learning predictive models that transport." The 22nd International Conference on Artificial Intelligence and Statistics. PMLR, 2019.
- Denis Agniel, Isaac S Kohane, and Griffin M Weber. Biases in electronic health record data due to processes within the healthcare system: retrospective observa- tional study. Bmj, 361, 2018.