Causal models are generative models
What is special about them ?

Michèle Sebag

TAU, CNRS – INRIA – LISN, U. Paris-Saclay

Collège de France: Feb 14th, 2024

Coll: I. Guyon (Google & U. PSaclay); D. Lopez-Paz (FAIR); O. Goudet (U. Angers); P. Caillou, G. Bied, S. Dong, D. Kalainathan, A. Lacombe (U. PSaclay), O. Allais (INRAE), J. Mink (U. Bonn); B. Crépon (ENSAE)

PEPR-IA-Causalite: M. Clausel (U. Nancy)
E. Gaussier, E. Devijver (U. Grenoble); E. Chzhen, M. Sebag (UPSaclay)

Summary

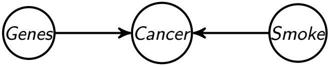
- A causal model is a generative model
- Why — Positioning wrt Machine Learning
- What — Formal background
- How — Observational causal discovery
- An example: Causality and Human Resources analysis

Example

Pearl, 2009; Pearl & Mackenzie 2018

Variables

- Covariates $X_1 \dots X_d$ — genes G; smoke S
- Outcome Y — cancer C



Intervention: do($X = v$)

Setting the value of X to v

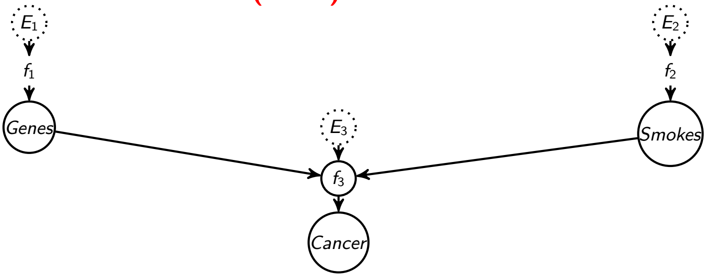
Definition: X is a cause of Y, noted $X \rightarrow Y$ iff

$$P(Y|do(X = v)) \neq P(Y|do(X = v'))$$

Beware: **Intervening is *not* conditioning**

- Conditioning: observing what happens for smokers
- Intervening: making everyone smoke; and observing what happens

Functional causal models (FCM)



Variables

- Endogenous variables: $X_1, \ldots X_d$
- Exogenous noise variables: independent, one for each variable

  Noise variables E_1, E_2, E_3 **model the known unknowns**

Causal graph

- Edge (X, Y) $(X \rightarrow Y)$ iff X is a direct cause of Y

Distribution

- Causal mechanism f_i

$$X_i \sim f_i(PA(X_i), E_i)$$

with $PA(X_i)$ the immediate causes of X_i

$$P(X_1, \ldots X_d) = \prod P(X_i | PA(X_i), E_i)$$

Why Causal Modeling ? Position w.r.t. Machine Learning

Formal background
- Ideally, Randomized Controlled Trials
- Alternative: Causal Discovery from Observational Data

Observational causal discovery
- Dependency tests-based
- Score based
- Based on supervised learning

Structural Agnostic Modelling

An application to Human Resources

The AI/ML/big data promise

Auguste Comte 

Knowledge → Prediction → Control

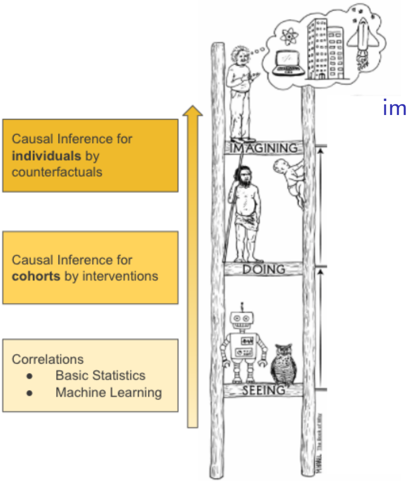
Savoir pour prévoir afin de pouvoir

ML models will expectedly enable control: **= Interventions**

- health and nutrition
- education
- economics/management
- climate

Learning causal models

Pearl 2003, 2009, 2018; Schölkopf et al., 2021, Barenboim et al., 2022



The causal hierarchy

imagining: would a given patient have suffered heart failure if they had started exercising a year earlier ?

doing: how does the probability of heart failure change if we convince a patient to exercise regularly ?

seeing: what is the probability of heart failure given certain diagnostic measurements ?

Model accurate at level $i \not\rightarrow$ model accurate at level $i+1$

Mainstream Machine Learning: discriminative or generative modelling

Given a training set iid samples $\sim P(X, Y)$

$$\mathcal{E} = \{(\mathbf{x}_i, y_i), \mathbf{x}_i \in \mathbb{R}^d, i \in [[1, n]]\}$$

Find

- Supervised learning: $\widehat{h} : X \mapsto Y$ or $\widehat{P(Y|X)}$
- Generative model $\widehat{P(X, Y)}$

Predictive modelling might be based on correlations

If umbrellas in the street, Then it rains

Using Machine Learning models out of their scope

If you can predict...

... can you make things happen ?

Causal modeling for more trustworthy AI

Wanted: An AI with common decency

- Fair — no biases
- Accountable — model can be explained
- Transparent — decisions can be explained
- Robust

The dark side of AI:

Zeynep Tufekci	We're building a dystopia just to make people click on ads
C. O'Neill	Weapons of Math Destruction, 2016
Timnit Gebru	www.technologyreview.com/2020/12/04/1013294/google-ai-ethics-research-paper-forced-out-timnit-gebru

Causal modeling for more trustworthy AI, 2

- Decreased sensitivity wrt data distribution
- Support interventions
- Hopes of explanations
- Hopes of bias detection

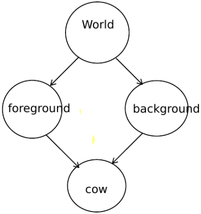
The robustness issue

Perona, 2022

When it works

When it does not work

Tentative interpretation



out-of-distribution generalization.

Caveat

Causal models $\neq$ efficient predictive models

$$
\begin{array}{lll}
X & \sim U[0,1] & \\
 & & \\
Y & \leftarrow 0.5X + E_Y & E_Y \sim \mathcal{N}(0,1) \\
Z & \leftarrow Y + E_Z & E_Z \sim \mathcal{N}(0,1)
\end{array}
$$

Predicting Y

$$\widehat{Y} = 0.25X + 0.5Z$$

If interpreted as a causal model, suggests that Y depends on Z.

Why Causal Modeling ? Position w.r.t. Machine Learning

Formal background

- Ideally, Randomized Controlled Trials
- Alternative: Causal Discovery from Observational Data

Observational causal discovery

- Dependency tests-based
- Score based
- Based on supervised learning

Structural Agnostic Modelling

An application to Human Resources

Causal Discovery: The royal road

Gold standard: Randomized controlled trials, RCT

- Draw iid samples, form two subsets:
  - T=1: treatment group
  - T=0: control group
- Compute Average Treatment Effect (ATE);
  Conditional Average Treatment (CATE)

Notations

- Y: outcome (survival)
- X: covariates (age, gender,...)
- $Y_i(0)$: outcome of the i-th sample if it does not get the treatment
- $Y_i(1)$: outcome of the i-th sample if it does get the treatment

Challenges:

- For i-th individual, either $Y_i(0)$ or $Y_i(1)$ is known
- In many cases, no RCTs (infeasible; unethical; too expensive)

Goal

Estimate the Average Treatment Effect

$$ATE = \mathbb{E}[Y(1) - Y(0)]$$

Under assumptions, it works $(X \perp\!\!\!\perp T)$

$$\begin{aligned}
ATE &= \mathbb{E}[Y(1) - Y(0)] \\
&= \mathbb{E}[Y(1)] - \mathbb{E}[Y(0)] \\
&\qquad \textit{linearity of expectation} \\
&= \mathbb{E}_X[\mathbb{E}[Y(1)|X]] - \mathbb{E}_X[\mathbb{E}[Y(0)|X]] \\
&\qquad \textit{expectation over covariates} \\
&= \mathbb{E}_X[\mathbb{E}[Y(1)|T = 1, X]] - \mathbb{E}_X[\mathbb{E}[Y(0)|T = 0, X]] \\
&\qquad \textit{no hidden confounder (no unbserved common causes)} \\
&\qquad \textit{overlap assumption, T=1 and T=0 are observed in the data} \\
&= \mathbb{E}_X[\mathbb{E}[Y|T = 1, X]] - \mathbb{E}_X[\mathbb{E}[Y|T = 0, X]] \\
&\qquad \textit{consistency: } Y_i(1) \sim Y|T = 1, X = X_i
\end{aligned} \tag{1}$$

Estimating ATE: when it does not work ($X \not\perp\!\!\!\perp T$)

The Simpson paradox: comparing treatments A and B of kidney stones

Reporting the survival rate:

Stone size	Treatment A	Treatment B
Small stones	93% (81/87)	87% (234/270)
Large stones	73% (192/263)	69% (55/80)
Total	78% (273/350)	83% (289/350)

Bottom line: treatment B dominates treatment A

Not really: treatment A applied more often on severe cases

(because it's more efficient)

because applied on more severe cases, A looks less efficient...

Observational Causal Discovery

WHAT

- Given data,
- Infer causal models
- Challenges: data quality; data quantity; learning criterion...

Some applications

- For human resources — Coll. ENSAE, Pôle Emploi; Bied et al. 23
  Quality of life at work / economic profitability
  Coll. SECAFI, La fabrique de l'industrie; Kalainathan et al. 18
- For health — Coll. INRAE
  - Diet / Diabetes type 2. — Gasnikova et al. 21
  - Impact of pesticides on newborn diseases

Why Causal Modeling ? Position w.r.t. Machine Learning

Formal background

- Ideally, Randomized Controlled Trials
- Alternative: Causal Discovery from Observational Data

Observational causal discovery

- Dependency tests-based
- Score based
- Based on supervised learning

Structural Agnostic Modelling

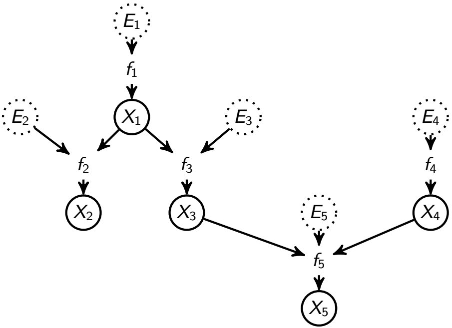
An application to Human Resources

Observational causal discovery

Goal

Given $\mathcal{D} \sim P(X_1, \ldots X_d)$

Find Functional Causal Model



$$\begin{cases} X_1 = f_1(E_1) \\ X_2 = f_2(X_1, E_2) \\ X_3 = f_3(X_1, E_3) \\ X_4 = f_4(E_4) \\ X_5 = f_5(X_3, X_4, E_5) \end{cases}$$

Requirements on solutions

- Identifiability unicity of the FCM solution
- The causal graph is directed acyclic

Observational causal discovery, 2

Reichenbach principle

If $X \not\perp\!\!\!\perp Y$, then exists Z s.t. $Z \rightarrow X$ and $Z \rightarrow Y$.
3 cases: $X \rightarrow Y$ or $Y \rightarrow X$ or $X \leftarrow Z \rightarrow Y$ (Z confounder)

Usual assumptions

- Causal sufficiency: no unobserved common confounders
- Causal Markov, Causal Faithfulness:

$$P(X_1, \ldots X_d) == \prod P(X_i | PA(X_i), E_i)$$

Observational Causal Discovery

Main approaches

- Dependence/independence tests based
- Score-based
- Learning-based

Main drawbacks

- Find a graphical model: doubly exponential
- Data hungry: Tests in $\mathcal{O}(d^3)$ + multiple hypothesis testing
- Representativity and biases in data

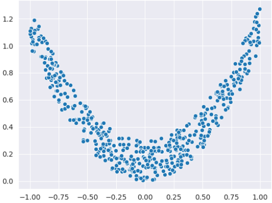
Observational Causal Discovery based on tests, 1

Independent variables X and Y ($X \perp\!\!\!\perp Y$)

$$X \perp\!\!\!\perp Y \text{ iff } P(X,Y) = P(X).P(Y)$$

Dependency tests

- Correlation — limited to linear dependencies



$$Y = X^2 + \varepsilon$$
$$\mathit{Correlation}(X,Y) \approx 0$$

- HSIC, Hilbert-Schmitt Independence Criterion

Gretton et al. 05

Observational Causal Discovery based on tests, 2

Conditional dependence a.k.a. V-structure

$$X = \text{Complex machine} \rightarrow Z = \text{Accident} \leftarrow Y = \text{Inexperienced Worker}$$

X and Y are independent; but given $Z = \textit{true}$ they are not independent (either the machine is complex or the worker is inexperienced...)

Observational Causal Discovery based on tests, 3

Algorithms

- Find dependencies and conditional independencies
- Constraint propagation

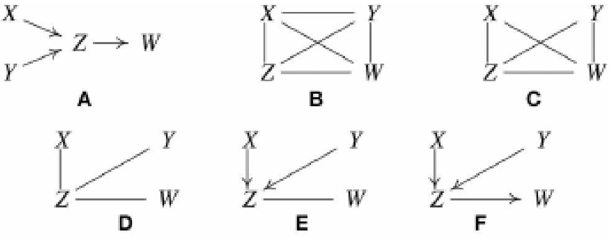


FIGURE 1 | Illustration of how the PC algorithm works. **(A)** Original true causal graph. **(B)** PC starts with a fully-connected undirected graph. **(C)** The $X - Y$ edge is removed because $X \perp Y$. **(D)** The $X - W$ and $Y - W$ edges are removed because $X \perp W \mid Z$ and $Y \perp W \mid Z$. **(E)** After finding v-structures. **(F)** After orientation propagation.

PC Algorithm (PC) Spirtes et al. 00; Glymour et al., 20

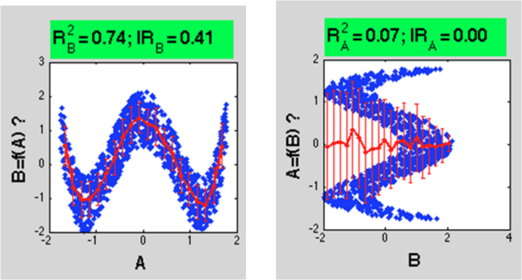
Non-linear extensions (CI tests): PC-HSIC (KCI-test), PC-RCIT

Zhang 12, Strobl 17

Observational Causal Discovery based on scores, 1

Leveraging Occam's razor principle Janzig 19

$\rightarrow$ the causal model as the one being the simplest model that fits the data.



Intuition

- If $X \rightarrow Y$, then $p(X, Y) = p(Y|X).p(X)$

Observational Causal Discovery based on scores, 2

Bayesian Information Criterion

$$BIC(\mathcal{G}) = -2 \ln L + k * \ln n$$

with L: Likelihood of the model, k: number of parameters, n: Number of samples

The graph is optimized with the operators:

- add edge
- remove edge
- revert edge

Alg: Greedy Equivalence Search (GES) Chickering 02, 20

Observational Causal Discovery based on scores, 2

Leveraging the data inverse covariance matrix Θ

Loh Buhlman 2014; Shimizu et al. 06, Dong et al. 23

If

$$\mathbf{X} = B^t\mathbf{X} + \mathbf{E}$$

Then

$$(I - B)\, D^{-1}\, (I - B)^t = \Theta \tag{2}$$

with $D = Diag(var(E_1), \ldots var(E_d))$

Issues

- Estimating Θ
- Enforcing DAG-ness — Zheng et al. 18
- Solution of Eq. 2 not unique

Observational Causal Discovery based on Machine Learning

Guyon et al, 2014-2015; 2019

Pair Cause-Effect Challenges

- Gather data: a sample is a pair of variables (A_i, B_i)
- Its label ℓ_i is the "true" causal relation (e.g., age "causes" salary)

Input

$$\mathcal{E} = \{(A_i, B_i, \ell_i), \ell_i \; in \; \{\rightarrow, \leftarrow, \perp\!\!\!\perp\}\}$$

Example A_i, B_i	Label ℓ_i
A_i causes B_i	$\rightarrow$
B_i causes A_i	$\leftarrow$
A_i and B_i are independent	$\perp\!\!\!\perp$

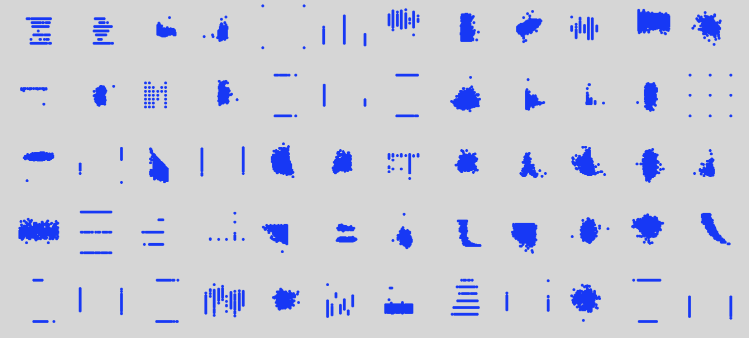
Output

using supervised Machine Learning

$$\text{Hypothesis} \; : (A, B) \mapsto \; \text{Label}$$

Guyon et al, 2014-2015

Pair (A_i, B_i) can be represented by a distribution = an image:



The Cause-Effect Pair Challenge

Learn a causality classifier (causation estimation)

- Like for any supervised ML problem from images   ImageNet 2012

More

- Guyon et al., eds, *Cause Effect Pairs in Machine Learning*, 2019.

The Cause-Effect Pair Challenge, Limitation

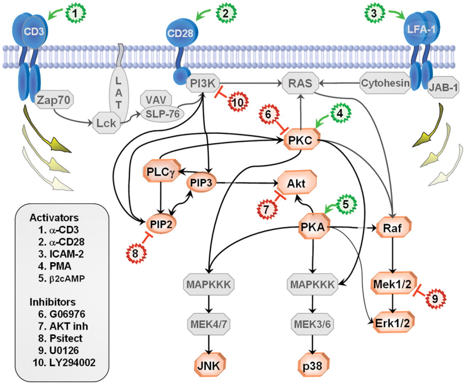
Using predictive ML to estimate the causation sense: where is the problem ?

The Cause-Effect Pair Challenge, Limitation

Using predictive ML to estimate the causation sense: where is the problem ?
Predictive learning requires **examples**: many, representative,...

Where do the CEP examples come from ?

- Common sense (age causes salary...)
- Variations on regulatory genes



Structure Agnostic Modeling

Kalainathan et al. 22

Principle

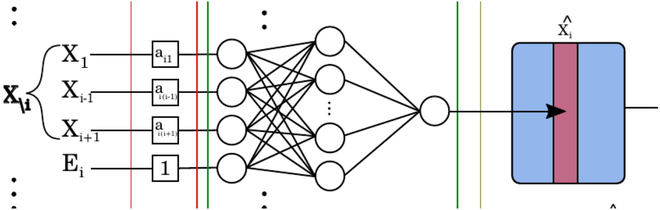
- For each X_i, learn a Markov kernel

$$X_i \sim f_i(\text{Subset of } \mathbf{X}, E_i)$$

  Rk: Sufficient subset of $\mathbf{X}$: Markov blanket (X_i) (causes, effects, spouses)
  Yu et al., 2018

- Learn simultaneously all kernels
  Rk: Avoids combinatorial search for structure
- Add regularization
  To enforce parsimony (MB(X_i))
  To enforce DAGness ($PA(X_i)$)

Structure Agnostic Modeling, 2



The *i*-th neural net

- Learns conditional distribution $P(X_i|X_{\backslash i})$ as $\hat{f}_i(X_{\backslash i}, E_i)$
- Gate variables $a_{i,j}$ in $\{0,1\}$

$$f_i(X_{\backslash i}, E_i) = \sum_k \beta_{i,k}\, \phi_{i,k}(a_{i,1}X_1, \ldots, a_{i,d}X_d, E_i)$$

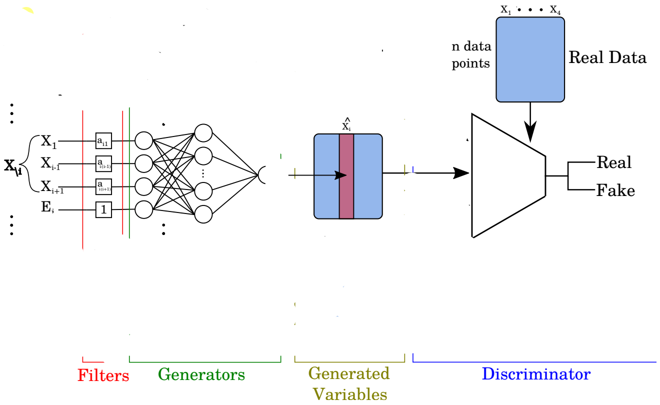
Structure Agnostic Modeling, 2

Adversarial training

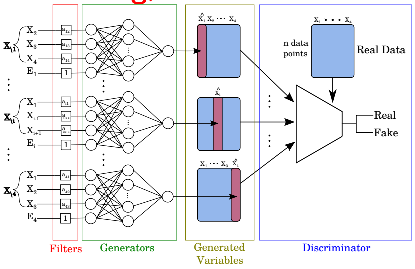
Goodfellow 14; 17; Arjovsky Bottou 2017

- Generate $\{\tilde{x}_i^{(j)}\}$ with j-th component of $\tilde{x}_i^{(j)}$ set to $\hat{f}_i(x_i, \epsilon)$, $\epsilon \sim \mathcal{N}(0,1)$
- Discriminate true from generated data

$$\min_G \max_D \mathbb{E}_x[log(D(x))] + \mathbb{E}_{\hat{x}}[log(1 - D(x)]$$



Structure Agnostic Modeling, 3



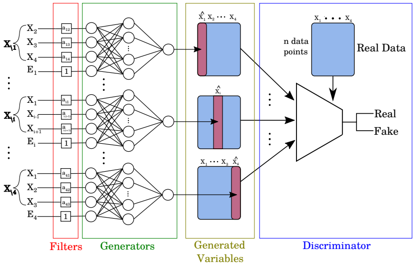
Given observational data $\{x_1, \ldots, x_n\} \sim P(X_1, \ldots, X_d)$ $\quad x_i$ in $\mathbb{R}^d$

Adversarial learning

- Generate $\{\tilde{x}_i^{(j)}\}$ with j-th component of $\tilde{x}_i^{(j)}$ set to $\hat{f}_i(x_i, \epsilon)$, $\epsilon \sim \mathcal{N}(0,1)$
- Discriminator D among observational data $\{x_i\}$ and generated data $\{\tilde{x}_i^{(j)}, i = [[1,n]], j = [[1,d]]\}$
- Learning criterion (adversarial + sparsity)

$$\min_G \max_D \left(\mathbb{E}_x[log(D(x))] + \mathbb{E}_{\hat{x}}[log(1 - D(x)] + \lambda \sum_{i,j} |a_{i,j}| \right)$$

Structure Agnostic Modeling, 4



Structure of causal graph: $A = (a_{i,j})$

Competition between discriminator and sparsity term $||A||_1$

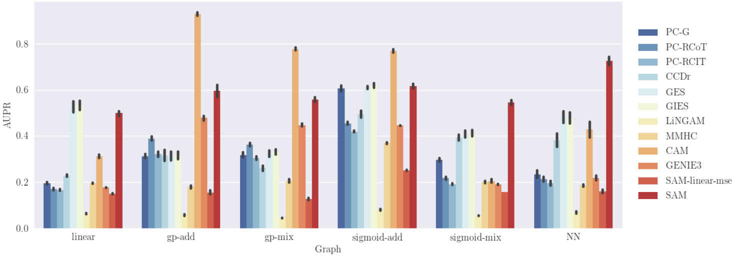
- Avoids combinatorial search for structure
- Cycles are possible
- DAGness achieved by enforcing constraints on trace of $exp(A)$

Zheng et al., 18

Validation on synthetic graphs

Experimental setting on synthetic graphs
causal mechanisms: linear; sigmoid additive or mixed; Gaussian Process additive or mixed; NN

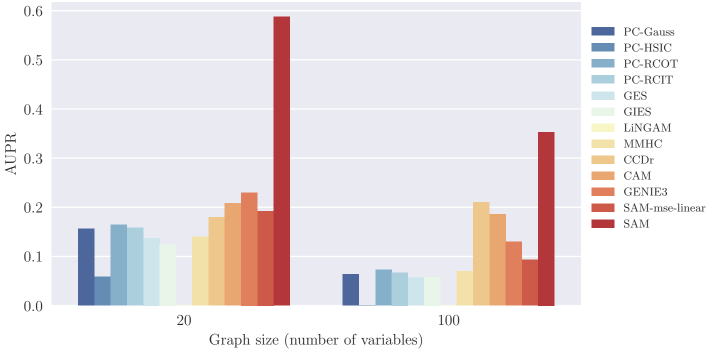
Comparative performance, Area Under Precision Recall Curve ($d = 100$)



On simulated biological datasets

On SynTREN graphs

Comparative performance, Area Under Precision Recall Curve

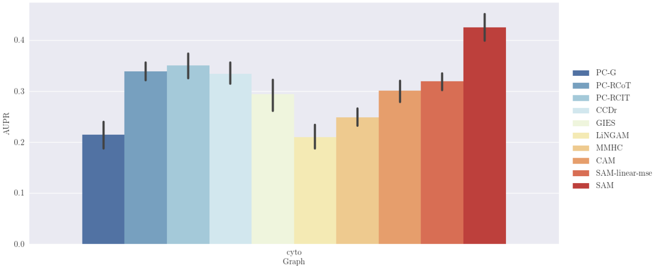


On Sachs, 2005

On protein network problem

Sachs et al., 05

Comparative performance, Area Under Precision Recall Curve



Quantitative results

Computational time (sec., per graph)

CPU: 48-core Intel(R) Xeon(R) CPU E5-2650 CPU.
GPU: Nvidia RTX 2080Ti GPU

AUPR	Linear	GP AM	GP Mix	Sigmoid AM	Sigmoid Mix	NN	CPU Time	GPU Time
PC-Gauss	0.19 (0.01)	0.31 (0.02)	0.32 (0.02)	0.61 (0.02)	0.30 (0.01)	0.23 (0.03)	13	
PC-HSIC	-	-	-	-	-	-	-	
PC-RCOT	0.18 (0.01)	0.39 (0.02)	0.36 (0.01)	0.45 (0.01)	0.22 (0.01)	0.21 (0.02)	31,320	
PC-RCIT	0.17 (0.01)	0.32 (0.02)	0.31 (0.01)	0.52 (0.01)	0.19 (0.01)	0.19 (0.02)	46,440	
GES	**0.53** (0.04)	0.32 (0.03)	0.32 (0.02)	0.61 (0.01)	0.41 (0.03)	0.48 (0.04)	1	
GIES	**0.53** (0.03)	0.31 (0.03)	0.33 (0.02)	0.62 (0.02)	0.41 (0.02)	0.48 (0.04)	5	
MMHC	0.20 (0.01)	0.18 (0.01)	0.21 (0.01)	0.37 (0.01)	0.20 (0.01)	0.19 (0.01)	5	
LiNGAM	0.06 (0.01)	0.06 (0.01)	0.04 (0.01)	0.08 (0.01)	0.05 (0.01)	0.07 (0.01)	5	
CAM	0.31 (0.01)	**0.93** (0.01)	**0.78** (0.01)	**0.77** (0.01)	0.20 (0.01)	0.43 (0.05)	45,899	
CCDr	0.23 (0.01)	0.31 (0.04)	0.26 (0.02)	0.49 (0.02)	0.39 (0.02)	0.38 (0.05)	3	
GENIE3	0.18 (0.01)	0.48 (0.02)	0.45 (0.01)	0.45 (0.01)	0.19 (0.01)	0.22 (0.02)	511	
SAM-lin-mse	0.15 (0.003)	0.16 (0.02)	0.13 (0.01)	0.25 (0.004)	0.16 (0.002)	0.16 (0.01)	3,076	74
SAM-mse	0.21 (0.01)	0.30 (0.03)	0.20 (0.01)	0.33 (0.005)	0.20 (0.01)	0.26 (0.03)	12,896	118
SAM-lin	0.41 (0.01)	0.29 (0.02)	0.22 (0.01)	0.51 (0.01)	0.46 (0.02)	0.47 (0.04)	8,746	516
SAM	0.50 (0.02)	0.60 (0.04)	0.56 (0.02)	0.62 (0.02)	**0.55** (0.02)	**0.72** (0.03)	15,361	519

Why Causal Modeling ? Position w.r.t. Machine Learning

Formal background

- Ideally, Randomized Controlled Trials
- Alternative: Causal Discovery from Observational Data

Observational causal discovery

- Dependency tests-based
- Score based
- Based on supervised learning

Structural Agnostic Modelling

An application to Human Resources

Causal Modeling and Human Resources

Position of the problem

A Quality of life at work — employee's perspective

B Economic performance — firm's perspective

- ... are correlated

Question: Are there causal relationships ?
$A \to B$; or $B \to A$; or $\exists C$ / $C \to A$ and $C \to B$

- Answering the question is key to evolve management strategies.

Data

- Gathered by Group Alpha Secafi (trade union advisor)
- Tax files + social audits for 408 firms

Data

Firms

Category 1	255	Chocolatier, Lesieur, Dassault, Compagnie des fromages,
Category 2	312	ArcelorMittal, St Gobain, Lafarge, Vallourec, Michelin,
Category 3	197	Air Liquide, Thales, Mersen, Filtrauto, Fenwick,...
Category 4	105	Hispano-Suiza, TurboMéca, Sanofi, Snecma,...

Variables

- Total number of employees, av. salary, productivity, profits, ..
- age, Average seniority, Physical effort, Permanent contract rate, Manager rate, Fixed-term contract rate, Temporary job rate, Shift and night work, Turn-over
- Frequency & Gravity of work injuries, Safety expenses, Safety training expenses, Absenteeism (diseases), Occupational-related diseases
- Percentage of women (employees, managers); wage gap

Phase 1: General Causal Relations

General Causal Relations

Access to training $\nearrow$

- $\searrow$ Gravity of work injuries
- $\searrow$ Occupational-related diseases

Termination rate $\nearrow$

- $\nearrow$ Absenteism (diseases)

Percentage of managers $\nearrow$

- $\nearrow$ Access to training
- $\searrow$ Shift or night working hours

Age $\nearrow$

- $\searrow$ Fixed-term contract rate
- $\searrow$ Productivity (weak impact)

?

- Productivity $\nearrow$ $\rightarrow$ Participation rate $\nearrow$

Global relations between QLW and performance ?

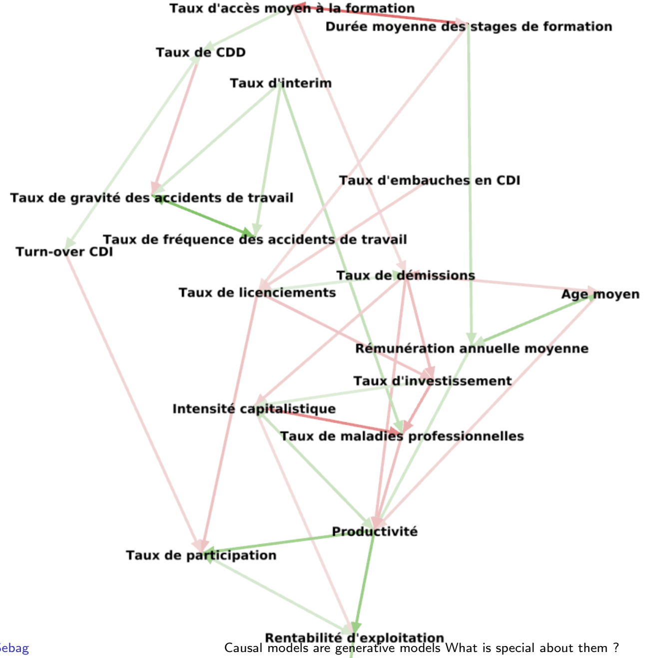
Failure

- Nothing conclusive

Interpretation

- Exist confounders (controlling QLW and performance) $C \rightarrow A$ and $C \rightarrow B$
- One such confounder is the activity sector
- In different activity sectors, causal relations are different (hampering their identification)
- $\Rightarrow$ Condition on confounders (independently handle the activity sectors)

Conditioning on confounder ! Focus on Low-tech industry



Category 1 (low-tech industry)

- Resignation rate $\nearrow$, Productivity $\searrow$
- Average salary $\nearrow$, Productivity $\nearrow$ very significant
- Occupational-related diseases $\nearrow$, Productivity $\searrow$
- Temporary job rate $\nearrow$, Gravity of work injuries $\nearrow$
- Permanent contract rate $\nearrow$, Safety training $\searrow$
- Duration training stints $\nearrow$, Termination rate $\searrow$

Conclusion

SAM: Feasibility of causal generative modelling

+ Avoid combinatorial optimization + parallelization
+ Covers large families of causal mechanisms
− Better than competitors; still limited

Applications

- Always: Confirmatory studies needed
- Always: Confounders.

Open source: SAM + Causal Toolbox

- Causal Discovery Tolbox
- SAM

Perspectives

Relaxing the requirements

- From identifiability to stability

Change of representation

- Dimensionality reduction — Wang Blei 2021
- Handling confounders — Karlsson Krijthe 2023
- Causal structures in latent space — Roy et al, 2023

Divide and Conquer

- Subsets of variables (e.g., $MB(X_i)$)
- Partial solutions $\mathcal{G}_i$
- Reconcile $\mathcal{G}_i$

We are hiring ! https://sites.google.com/view/causali-t-ai/home

References

- Arjovsky, M. and Bottou, L. Towards principled methods for training generative adversarial networks. ICLR 2017
- Arjovsky, M. et al. Wasserstein generative adversarial networks. ICML 2017
- Elias Bareinboim, Juan D. Correa, Duligur Ibeling, and Thomas Icard. On Pearl's Hierarchy and the Foundations of Causal Inference, pages 507–556. Association for Computing Machinery, 2022
- Guillaume Bied, Solal Nathan, Elia Perennes, Morgane Hoffmann, Philippe Caillou, Bruno Crépon, Christophe Gaillac, Michèle Sebag: Toward Job Recommendation for All. IJCAI 2023
- Statistically Efficient Greedy Equivalence Search, David Maxwell Chickering, UAI 2020
- Diviyan Kalainathan, Olivier Goudet, Isabelle Guyon, David Lopez-Paz, Michèle Sebag: Structural Agnostic Modeling: Adversarial Learning of Causal Graphs. J. Mach. Learn. Res. 23: 219:1-219:62 (2022)
- Diviyan Kalainathan, Olivier Goudet, Ritik Dutta:Causal Discovery Toolbox: Uncovering causal relationships in Python. J. Mach. Learn. Res. 21: 37:1-37:5 (2020)
- Ksenia Gasnikova, Olivier Allais, Michèle Sebag: Towards causal modeling of nutritional outcomes. CAWS 2021: 5-19
- Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. Advances in Neural Information Processing Systems, 2014
- Arthur Gretton, Olivier Bousquet, Alexander J. Smola, Bernhard Schölkopf: Measuring Statistical Dependence with Hilbert-Schmidt Norms. ALT 2005: 63-77
- Isabelle Guyon, Alexander Statnikov, Berna Bakir Batu (eds). Cause Effect Pairs in Machine Learning Springer-Verlag, 2019.
- Rickard Karlsson, Jesse Krijthe. Detecting hidden confounding in observational data using multiple environments, NeurIPS 2023

References, 2

- Cathy O'Neill, Weapons of Math Destruction, 2016.
- Judea Pearl. Causality: models, reasoning, and inference. Econometric Theory, 19(675-685):46, 2003.
- Judea Pearl. Causality. Cambridge university press, 2009
- Pietro Perona: Generalization and Abstraction, Out-of-Distribution Generalization, ECCV 2022.
- Saptarshi Roy, Raymond K. W. Wong, Yang Ni: Directed Cyclic Graph for Causal Discovery from Multivariate Functional Data, NeurIPS 2023
- Shohei Shimizu, Patrik O Hoyer, Aapo Hyvärinen, and Antti Kerminen. A linear non-gaussian acyclic model for causal discovery. Journal of Machine Learning Research, 2006
- Bernhard Schölkopf, Francesco Locatello, Stefan Bauer, Nan Rosemary Ke, Nal Kalchbrenner, Anirudh Goyal, Yoshua Bengio: Toward Causal Representation Learning. Proc. IEEE 109(5): 612-634, 2021
- Yixin Wang, David M. Blei: A Proxy Variable View of Shared Confounding. ICML 2021
- Kui Yu, Lin Liu, and Jiuyong Li. A unified view of causal and non-causal feature selection. arXiv preprint arXiv:1802.05844, 2018
- Xun Zheng, Bryon Aragam, Pradeep K Ravikumar, and Eric P Xing. DAGs with NO TEARS: Continuous optimization for structure learning. In Advances in Neural Information Processing Systems, volume 31, 2018