Causal Data Science for Business Decision Making

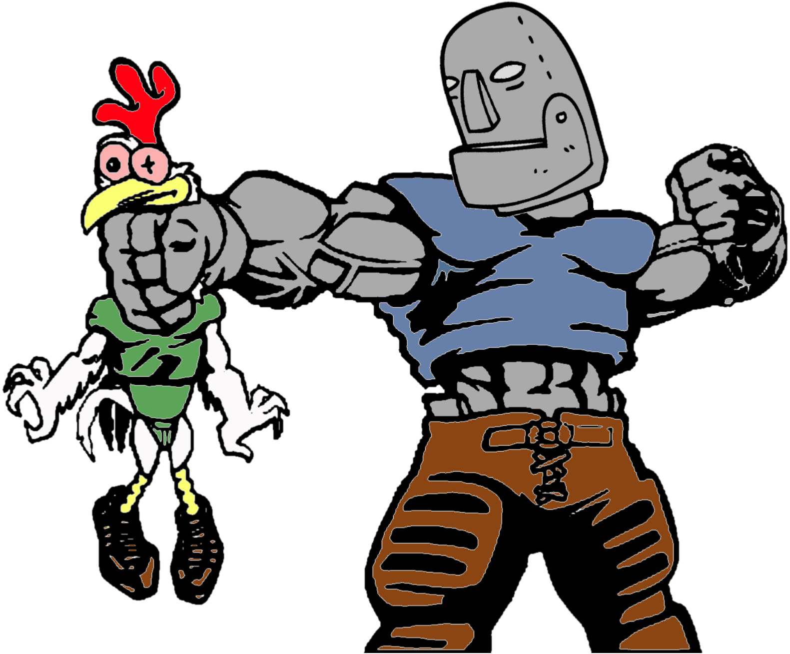
Causal Artificial Intelligence

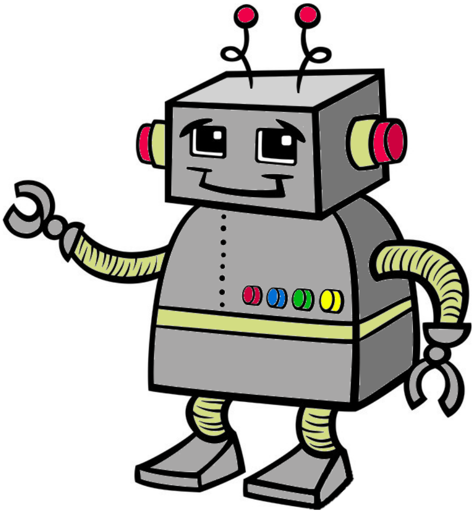
Paul Hünermund

COPENHAGEN BUSINESS SCHOOL
HANDELSHØJSKOLEN

Remember: Writing Challenge

- Work in teams of 3–5 students
- Write a 800–1500 words blog post about a topic of your choice
  - E.g., pick out a technique or topic we have discussed in class and illustrate its relevance for industry
  - Try to find practically relevant examples
  - Format similar to articles on `www.medium.com` or `www.towardsdatascience.com`
- Feedback opportunity for you!!
- The best three submissions will be published on `www.causalscience.org`
- <u>Deadline</u>: November 28, 2021



Why Causality and AI?

The way you talk about curve fitting, it sounds like you're not very impressed with machine learning.

No, I'm very impressed, because we did not expect that so many problems could be solved by pure curve fitting. It turns out they can. But I'm asking about the future — what next? Can you have a robot scientist that would plan an experiment and find new answers to pending scientific questions? That's the next step. We also want to conduct some communication with a machine that is meaningful, and meaningful means matching our intuition. If you deprive the robot of your intuition about cause and effect, you're never going to communicate meaningfully. Robots could not say "I should have done better," as you and I do. And we thus lose an important channel of communication.

Do-Calculus

- Remember: Identification task (with observational data) $\hat{=}$ we need to transform our target query $Q = P(y|do(x))$ into an expression that only contains standard probability objects
- For that we need to know the rules on how to deal with these do-objects: *do-calculus*
  - Do-calculus is a powerful symbolic machinery that provides a set of inference rules by which sentences involving interventions can be transformed into other sentences (Pearl, 2009; Pearl et al., 2016)
- Do-calculus can be used to solve the confounding problem
  - Apply the rules of *do*-calculus repeatedly until a do-expression is translated into an equivalent expression involving only standard probabilities of observed quantities

Do-Calculus (II)

Theorem: Rules of Do-Calculus (Pearl, 2009, p. 85)

Let G be the directed acyclic graph associated with a [structural] causal model [...], and let $P(\cdot)$ stand for the probability distribution induced by that model. For any disjoint subset of variables X, Y, Z, and W, we have the following rules.

Rule 1 (Insertion/deletion of observations):

$$P(y|do(x), z, w) = P(y|do(x), w) \quad \text{if } (Y \perp\!\!\!\perp Z|X, W)_{G_{\overline{X}}}.$$

Rule 2 (Action/observation exchange):

$$P(y|do(x), do(z), w) = P(y|do(x), z, w) \quad \text{if } (Y \perp\!\!\!\perp Z|X, W)_{G_{\overline{X}\underline{Z}}}.$$

Do-Calculus (III)

Rule 3 (Insertion/deletion of actions):

$$P(y|do(x), do(z), w) = P(y|do(x), w) \quad \text{if } (Y \perp\!\!\!\perp Z|X, W)_{G_{\overline{X}\,\overline{Z(W)}}},$$

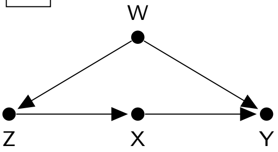
where $Z(W)$ is the set of Z-nodes that are not ancestors of any W-node in $G_{\overline{X}}$.

- $G_{\overline{X}}$ denotes the graph obtained by deleting from G all arrows pointing to nodes in X
- $G_{\underline{X}}$ denotes the graph obtained by deleting from G all arrows emerging from nodes in X

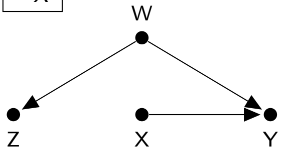
Do-Calculus Rule 1

$$P(y|do(x), z, w) = P(y|do(x), w) \quad \text{if } (Y \perp\!\!\!\perp Z|X, W)_{G_{\overline{X}}}$$

G



$G_{\overline{X}}$



The graph that results from deleting all arrows pointing into X in G is denoted by $G_{\overline{X}}$.

In $G_{\overline{X}}$, W blocks the only backdoor path between Z and Y: $Z \leftarrow W \rightarrow Y$.

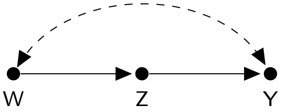
By d-separation $(Y \perp\!\!\!\perp Z|W)_{G_{\overline{X}}}$ we can therefore ignore Z in the conditional distribution of Y.

$$P(y|do(x), z, w) = P(y|do(x), w)$$

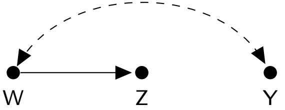
Do-Calculus Rule 2

$$P(y|do(x), do(z), w) = P(y|do(x), z, w) \qquad \text{if } (Y \perp\!\!\!\perp Z|X, W)_{G_{\overline{X}\underline{Z}}}$$

$\boxed{G}$



$\boxed{G_{\underline{Z}}}$



Assume we are interested in the query $P(y|do(z), w)$. The graph that results from deleting all arrows emitted by Z in G is denoted by $G_{\underline{Z}}$.

In $G_{\underline{Z}}$, W blocks the only backdoor path between Z and Y: $Z \leftarrow W \dashleftarrow\!\dashrightarrow Y$.

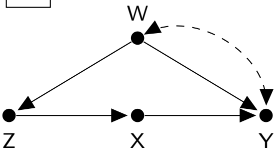
Thus, by d-separation $(Y \perp\!\!\!\perp Z|W)_{G_{\underline{Z}}}$ and therefore the second rule of do-calculus applies.

Consequently, we can get rid of the do-operator by setting $P(y|do(z), w) = P(y|z, w)$. The latter expression is estimable from observational data.

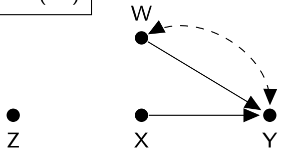
Do-Calculus Rule 3

$$P(y|do(x), do(z), w) = P(y|do(x), w) \quad \text{if } (Y \perp\!\!\!\perp Z|X, W)_{G_{\overline{X}\overline{Z(W)}}},$$

$\boxed{G}$



$\boxed{G_{\overline{X}\overline{Z(W)}}}$



Z is not an ancestor of W, therefore we delete all arrows pointing into Z and X from G to arrive at $G_{\overline{X}\overline{Z(W)}}$.

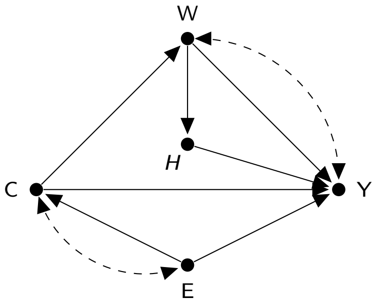
In $G_{\overline{X}\overline{Z(W)}}$, Z and Y are d-separated: $(Y \perp\!\!\!\perp Z)_{G_{\overline{X}\overline{Z(W)}}}$

Thus, we can delete the intervention on Z because it is not relevant anymore after we have already intervened on X

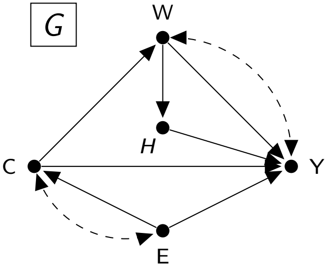
$$P(y|do(x), do(z)) = P(y|do(x))$$

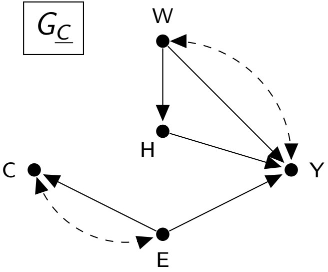
Example: Applying Do-Calculus

- Take the example of the college wage premium from one of the previous lectures
  - C: college degree
  - Y: earnings
  - W: occupation
  - H: work-related health
  - E: other socio-economic factors
- Task: Transform $P(y|do(c))$ into a do-free expression by using the rules of do-calculus



Example: Applying Do-Calculus (II)





There are two backdoor paths in G, which can both be blocked by E. Conditioning and summing over all values of E yields (law of total probability)

$$P(y|do(c)) = \sum_e P(y|do(c), e)P(e|do(c)).$$

By rule 2 of do-calculus

$$P(y|do(c), e) = P(y|c, e), \quad \text{since } (Y \perp\!\!\!\perp C|E)_{G_{\underline{C}}}.$$

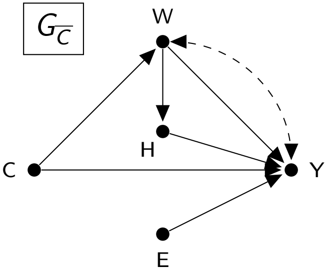
Refresher: Law of Total Probability

$$
\begin{aligned}
P(A) &= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \ldots + P(A|B_k)P(B_k) \\
&= \sum_{i=1}^{k} P(A|B_i)P(B_i)
\end{aligned}
$$

- Suppose there are 60% men in the class and 40% women
- Among women, the probability to get an A in the class is 25% while for men it is only 20%
- Then the overall probability to get an A in the class is

$$
\begin{aligned}
P(A) &= P(A|male)P(male) + P(A|female)P(female) \\
&= 0.2 \cdot 0.6 + 0.25 \cdot 0.4 \\
&= 0.22
\end{aligned}
$$

Example: Applying Do-Calculus (III)



By rule 3 of do-calculus

$$P(e|do(c)) = P(e), \quad \text{since } (E \perp\!\!\!\perp C)_{G_{\overline{C}}}.$$

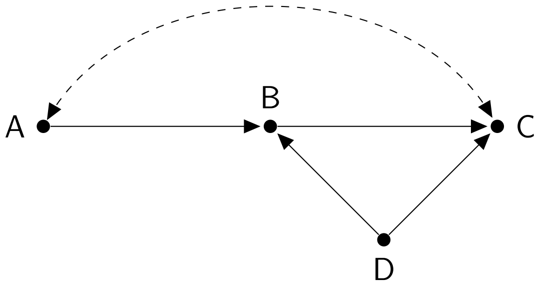
It follows that

$$P(y|do(c)) = \sum_{e} P(y|c, e)P(e).$$

The right-hand-side expression is do-free and can therefore be estimated from observational data.

Exercise: Do-Calculus

- Using the rules of do-calculus, show that $P(C|do(B))$ is identifiable via the backdoor adjustment formula in the following graph



Completeness of Do-Calculus

- Do-calculus is shown to be *complete*, meaning that if a causal effect is identifiable there exists a sequence of steps applying the rules of do-calculus that transforms the causal effect formula into an expression that includes only observable quantities (Shpitser and Pearl, 2006; Huang and Valtorta, 2006)
  - Put differently, if do-calculus fails, the causal effect is guaranteed to be unidentifiable
- Completeness proofs are notoriously difficult and showing this for the case of do-calculus was a major breakthrough in the literature (Pearl and Mackenzie, 2018)

Automatizing the Identification Task

- We know that do-calculus is complete, but the theorem is only procedural and does not tell us which series of steps leads to the desired solution
- Shpitser and Pearl (2006), building on work by Tian and Pearl (2002), propose an algorithm that automates this task
    - The algorithm takes a description of a DAG as input an returns an expression for a queried causal effect, if it exists
    - Since the algorithm is based on *do*-calculus and therefore complete, if it doesn't return a causal effect expression involving only observable quantities, no such expression exists
    - This approach is extremely user-friendly
        - Basically the computer is doing everything for you
        - But you should be aware of what's going on "under the hood"
        - To get a gist of how this algorithm works, see: `https://david-salazar.github.io/2020/07/31/causality-testing-identifiability/`

Identification Algorithms

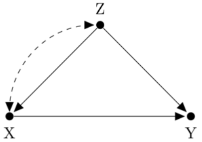
- The algorithm by Shpitser and Pearl (2006) is for the observational data case where $P(y|do(x))$ is supposed to be transformed into an expression that does not contain a do-operator (with unobservables included in the model)
- But similar algorithms have been developed for the other causal inference tasks we will discuss throughout this course
  - $\mathcal{Z}$-Identification (Bareinboim and Pearl, 2012)
  - Selection bias (Bareinboim and Tian, 2015)
  - Transportability (Bareinboim and Pearl, 2013, 2014)
- Input:
  1. A causal query Q
  2. The model in form of a diagram
  3. The type of data available
- Output: an estimable expression of Q
  - Most algorithms inherit *completeness* property from do-calculus

The Data Fusion Process

(1) Query:

$Q =$ Causal effect at target population

(2) Model:



(3) Available Data:

Observational:	$P(v)$
Experimental:	$P(v \mid do(z))$
Selection-biased:	$P(v \mid S=1)$ + $P(v \mid do(x), S=1)$
From different populations:	$P^{(source)}(v \mid do(x))$ + observational studies

Causal Inference Engine:

Three inference rules of *do-calculus*

Solution exists?

Yes

Estimable expression of Q

No

Assumptions need to be strengthened (imposing shape restrictions, distributional assumptions, etc.)

Does Your Dog Understand Cause and Effect?

Article | Open Access | Published: 15 September 2017

The effects of domestication and ontogeny on cognition in dogs and wolves

Michelle Lampe, Juliane Bräuer, Juliane Kaminski & Zsófia Virányi

Scientific Reports 7, Article number: 11690 (2017) | Cite this article

Abstract

Cognition is one of the most flexible tools enabling adaptation to environmental variation. Living close to humans is thought to influence social as well as physical cognition of animals throughout domestication and ontogeny. Here, we investigated to what extent physical cognition and two domains of social cognition of dogs have been affected by domestication and ontogeny. To address the effects of domestication, we compared captive wolves (n = 12) and dogs (n = 14) living in packs under the same conditions. To explore developmental effects, we compared these dogs to pet dogs (n = 12) living in human families. The animals were faced with a series of object-choice tasks, in which their response to communicative, behavioural and causal cues was tested. We observed that wolves outperformed dogs in their ability to follow causal cues, suggesting that domestication altered specific skills relating to this domain, whereas developmental effects had surprisingly no influence. All three groups performed similarly in the communicative and behavioural conditions, suggesting higher ontogenetic flexibility in the two social domains. These differences across cognitive domains need to be further investigated, by comparing domestic and non-domesticated animals living in varying conditions.

Thank you

Personal Website: p-hunermund.com

Twitter: @PHuenermund

Email: phu.si@cbs.dk

CBS COPENHAGEN BUSINESS SCHOOL
HANDELSHØJSKOLEN

EFMD EQUIS ACCREDITED AACSB ACCREDITED ASSOCIATION OF MBAS AMBA ACCREDITED CEMS PIM

References I

Elias Bareinboim and Judea Pearl. Causal inference by surrogate experiments: z-identifiability. In *Proceedings of the 28th Conference on Uncertainty in Artificial Intelligence*, pages 113–120, 2012.

Elias Bareinboim and Judea Pearl. A general algorithm for deciding transportability of experimental results. *Journal of Causal Inference*, 1(1):107–134, 2013.

Elias Bareinboim and Judea Pearl. Transportability from multiple environments with limited experiments: Completeness results. In Z. Ghahramani, M. Welling, C. Cortes, N.D. Lawrence, and K.Q. Weinberger, editors, *Advances of Neural Information Processing Systems*, volume 27, pages 280–288, November 2014.

Elias Bareinboim and Jin Tian. Recovering causal effects from selection bias. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015.

Yimin Huang and Marco Valtorta. Pearl's calculus of interventions is complete. In *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence (UAI2006)*, 2006. URL `https://arxiv.org/ftp/arxiv/papers/1206/1206.6831.pdf`.

Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York, United States, NY, 2nd edition, 2009.

Judea Pearl and Dana Mackenzie. *The Book of Why: The New Science of Cause and Effect*. Basic Books, New York, 2018. ISBN 9780465097609.

Judea Pearl, Madelyn Glymour, and Nicholas P. Jewell. *Causal Inference in Statistics: A Primer*. John Wiley & Sons Ltd, West Sussex, United Kingdom, 2016.

References II

Ilya Shpitser and Judea Pearl. Identification of Joint Interventional Distributions in Recursive Semi-Markovian Causal Models. In *Twenty-First National Conference on Artificial Intelligence*, 2006.

J. Tian and J. Pearl. A general identification condition for causal effects. In *Proceedings of the Eighteenth National Conference on Artificial Intelligence*, pages 567–573, Menlo Park, CA, 2002. AAAI Press/The MIT Press.