Abstracting Causal Models

Fabio Massimo Zennaro
fm.zennaro@gmail.com

Universiteit Utrecht
September 27th, 2021

Outline

1. *Introduction*: why do we care about causality and abstraction.
2. *Causality*: how do we express causal models formally.
3. *Abstraction*: how do we formalize and evaluate abstraction.
4. *State of the art*: problems and research directions.

1. Introduction

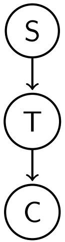
What is causality?

Some **operational** features [1]:

- ✓ *Relationship* between things/variables.
- ✓ Directed connection between *causes* and *effects*.
- ✓ *Interventional* aspect.

A driving example: *lung cancer model* [4]

- S: smoking habit
- T: tar deposits in the lungs
- C: lung cancer



What is abstraction?

Some **operational** features [1]:

- ✓ *Organization of information* on multiple levels.
- ✓ Heuristic for *efficient structuring of knowledge*.

A illustrative example: *thermodynamical systems* [5]

Microscopic description $\mathbf{p}, \dot{\mathbf{p}}$.

Macroscopic description P, T, V.

Examples abound in *computer science*, too (programming languages, OSI network stack)

Why studying causality and abstraction?

Theoretically:

- Foundational to our understanding of the world.
- Foundational to the scientific endeavour.

Practically:

- Crucial for modeling and artificial intelligence.
  - Differentiate association and causation.
  - Define interventions and policies.
  - Learn robust models in non-static settings.
  - Deal with multiple approximate models.
  - Switch between models just-in-time.

Our problem

When can causal models be considered in a relationship of abstraction?

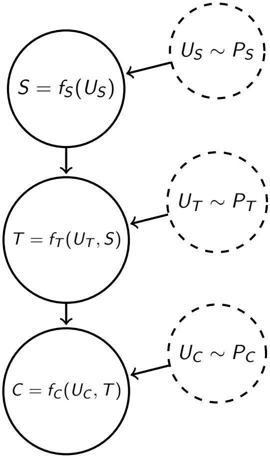
- Is a causal model an abstraction of another one?
- Is the abstraction exact or does it introduce any approximation?

2. Causality

SCMs

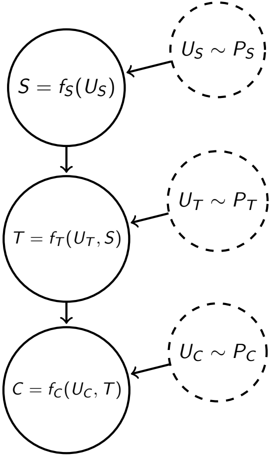
We express a causal model as a **structural causal model** $\mathcal{M}$ [1, 2]:

- $\mathcal{X}$: set of *endogenous nodes* (S, T, C) representing variables of interest
- $\mathcal{E}$: Set of *exogenous nodes* (U_S, U_T, U_C) representing stochastic factors
- $\mathcal{F}$: Set of *structural functions* (f_S, f_T, f_C) describing the dynamics of each variable
- $\mathcal{P}$: Set of *distributions* (P_S, P_T, P_C) describing the behavior of random factors



SCMs

Every SCM $\mathcal{M}$ implies a (joint) **distribution** $P_{\mathcal{M}}$:



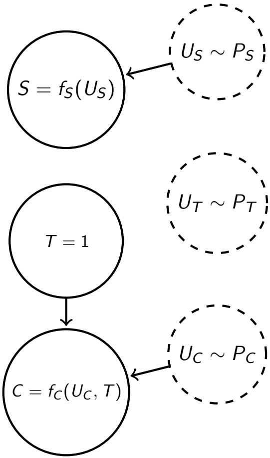
$$P_{\mathcal{M}}(S, T, C)$$

Interventions

We can perform **interventions** on a causal model:

$do(T = 1)$

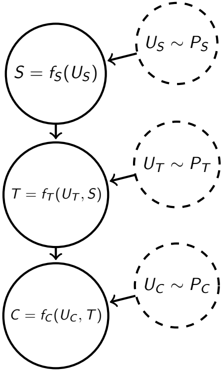
1. Remove incoming edges in the intervened node
2. Set the value of the intervened node



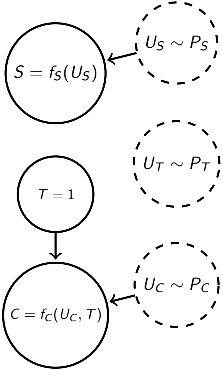
Intervened Model

An intervention ι_1 effectively defines a new **intervened model** $\mathcal{M}_{\iota_1}$.

$$\mathcal{M} = \langle \mathcal{X}, \mathcal{E}, \mathcal{F}, \mathcal{P} \rangle$$



$$\mathcal{M}_{\iota_1} = \langle \mathcal{X}, \mathcal{E}, \mathcal{F}_{\iota_1}, \mathcal{P} \rangle$$

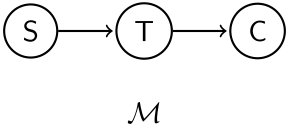


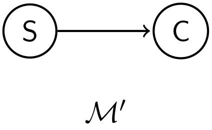
$$P_{\mathcal{M}}(S, T, C) \neq P_{\mathcal{M}_{\iota_1}}(S, T, C)$$

3. Abstraction

An example

Suppose we are given two SCMs of the lung cancer model:





What does it mean that *model $\mathcal{M}'$ is an* **abstraction** *of model $\mathcal{M}$*?

An observational meaning for abstraction

- **Observational consistency:** sampling the two models I obtain the same (observational) distributions of interest.

$$P_{\mathcal{M}}(S, C) = P_{\mathcal{M}'}(S, C)$$

An interventional meaning for abstraction

- **Interventional consistency:** under an intervention the two models produce the same (interventional) distributions of interest.

$$P_{\mathcal{M}}(C|do(S=0)) = P_{\mathcal{M}'}(C|do(S=0))$$

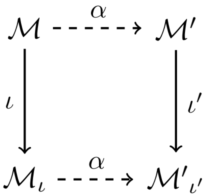
A strong meaning for abstraction

- **Abstraction-intervention commutativity:** given a model $\mathcal{M}$, the following two procedures lead to the same distribution $P_{\mathcal{M}'_{\iota'}}$:
  - Intervene on $\mathcal{M}$ and then map to the abstracted model;
  - Map $\mathcal{M}$ to the abstracted model and then intervene on it.

$$\begin{array}{ccc} \mathcal{M} & \overset{\alpha}{\dashrightarrow} & \mathcal{M}' \\ \iota \big\downarrow & & \big\downarrow \iota' \\ \mathcal{M}_{\iota} & \overset{\alpha}{\dashrightarrow} & \mathcal{M}'_{\iota'} \end{array}$$

A meaning for approximate abstraction

- **Abstraction approximation:** given a model $\mathcal{M}$, the following two procedures lead to two distributions:
  - Intervening and abstracting produces $P_{\alpha \circ \iota}$
  - Abstracting and intervening produces $P_{\iota' \circ \alpha}$



Approximation is computed using a *distance*:

$$D(P_{\alpha \circ \iota}, P_{\iota' \circ \alpha})$$

4. State of the art and challenges

Research Questions

Recent research direction with many questions.

1. *Formalizing abstractions* (more theoretical)
2. *Evaluating abstractions* (more practical)

Formalizing abstractions

How do we **express** that *model* $\mathcal{M}'$ *is an* **abstraction** *of model* $\mathcal{M}$?

$$\begin{array}{ccc} \mathcal{M} & \overset{\alpha}{\dashrightarrow} & \mathcal{M}' \\ \Big\downarrow \iota & & \Big\downarrow \iota' \\ \mathcal{M}_{\iota} & \overset{\alpha}{\dashrightarrow} & \mathcal{M}'_{\iota'} \end{array}$$

What is α?

Formalizing abstractions

Statistical formalizations:

- *Distributional*: α as a function mapping joint distributions [5]
- *Structural*: α as a collection of functions mapping variables [4]

What do we get from these approaches?

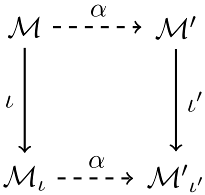
Categorical formalizations:

- *Structural*: α as a morphism between objects representing variables [4, 3]
- *Model*: α as a morphism between objects representing SCMs

What do we get from category theory?

Evaluating abstractions

How do we **measure** the abstraction **approximation** of *model* $\mathcal{M}'$ *with respect to model* $\mathcal{M}$?



Which interventions should we consider?
How do we measure? Which distances to consider?
Can we compute degree of approximation efficiently?

Evaluating abstractions

Exact abstraction:

- *Evaluation wrt a set of interventions* [5]

Approximate abstraction:

- *Jensen-Shannon distance wrt any legitimate intervention* [4, 3]
- *Composition in an enriched category* [4, 3]

Other approaches:

- *Graph-theoretical algorithms*
- *Topology-like invariance-based approaches*

Can we bound approximation with respect to time?

Further research questions

- Could abstractions be *stochastic*?
- Could abstractions express *preservation of structure*?
- Can we have different forms of *consistency*?
    - Can we evaluate *counterfactual consistency*?
- What can we learn from *physics* (renormalization theory)?

Many interesting questions and promising directions!

Thanks!

Thank you for listening!

If interested in existing approaches, feel free to check tutorials at:
`https://github.com/FMZennaro/CategoricalCausalAbstraction`

References I

[1] Judea Pearl. *Causality*. Cambridge University Press, 2009.

[2] Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. *Elements of causal inference: Foundations and learning algorithms*. MIT Press, 2017.

[3] Eigil F Rischel and Sebastian Weichwald. Compositional abstraction error and a category of causal models. *arXiv preprint arXiv:2103.15758*, 2021.

[4] Eigil Fjeldgren Rischel. The category theory of causal models. 2020.

[5] Paul K Rubenstein, Sebastian Weichwald, Stephan Bongers, Joris M Mooij, Dominik Janzing, Moritz Grosse-Wentrup, and Bernhard Schölkopf. Causal consistency of structural equation models. In *33rd Conference on Uncertainty in Artificial Intelligence (UAI 2017)*, pages 808–817. Curran Associates, Inc., 2017.