Learning Causal Abstractions

Fabio Massimo Zennaro

University of Bergen

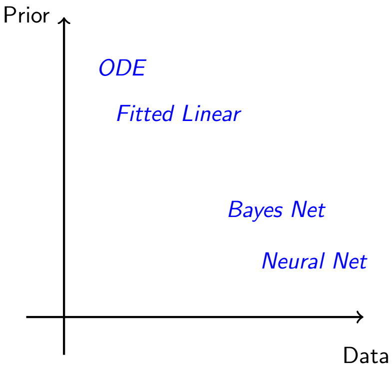
November 22, 2023

1. Structural Causal Modelling

Modelling

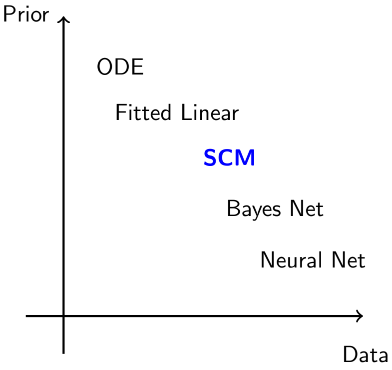
Assume we want to model a *system*.

Different types of model will negotiate a trade-off between priors and data:



Structural Causal Modeling

Structural causal models rely on a strong prior given by *causality* [6, 7].

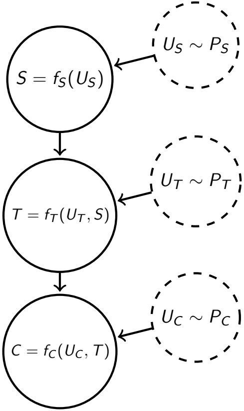


- It discriminates *correlations* and *causes*.
- It allows for reasoning about *interventions*.
- It allows for reasoning about *counterfactuals*.
- It implies a *causality ladder* of reasoning.
- It requires more than data.

SCMs

We express a causal model as a **structural causal model** $\mathcal{M} = \langle \mathcal{X}, \mathcal{U}, \mathcal{F}, \mathcal{P} \rangle$ [6, 7]:

- $\mathcal{X}$: set of *endogenous nodes* (S, T, C) representing variables of interest
- $\mathcal{U}$: 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 random factors



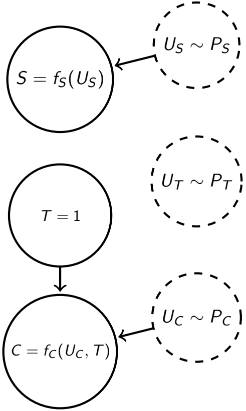
Every SCM $\mathcal{M}$ implies a (joint) **distribution** $P_{\mathcal{M}}$: $P_{\mathcal{M}}(S, T, C)$

Interventions

We can perform **interventions** on a causal model [6, 7]:

$do(T = 1)$

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



An intervention ι_1 effectively defines a new **intervened model** $\mathcal{M}_{\iota_1}$ such that $P_{\mathcal{M}}(S, T, C) \neq P_{\mathcal{M}_{\iota_1}}(S, T, C)$

2. Abstraction

Levels of Abstraction

Systems may be represented at different **levels of abstraction** (LoA) [3].

Thermodynamics example:

Low-level / Base model:
Microscopic description $\mathbf{x}, \dot{\mathbf{x}}$.

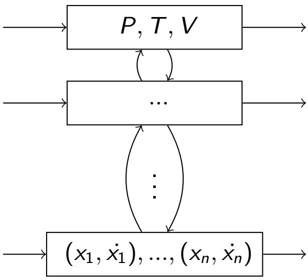
High-level / Abstracted model:
Macroscopic description P, T, V.

LoA may be inaccessible, so we may want to *shift* among LoAs.

1. We need a *mapping* between LoAs.
2. We want the mapping to be *consistent*.

Abstraction

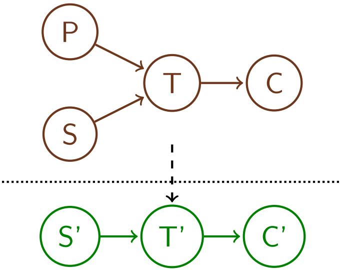
Abstraction (aka, *multi-level modelling* or *multi-resolution modelling*) aims at relating these levels.



- It combines models from *different sources*.
- It aggregates information from *different resolutions*.
- It allows for *computation with minimal effort*.

A Motivating Example

Lung cancer scenario example:



$$\mathcal{M}[S] = \mathcal{M}[P] = \mathcal{M}[T] = \mathcal{M}[C] = \{0, 1\}$$

$$\mathcal{M}'[S'] = \mathcal{M}'[C'] = \{0, 1\}$$

- The *transformation* approach [10, 1]
- The α-**abstraction** approach [9, 8]
- The Φ-*abstraction* approach [4, 5]

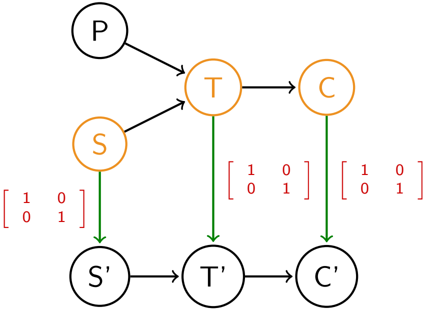
α-Abstraction [9]

An **abstraction** $\boldsymbol{\alpha}$ is a tuple

$$\langle R, a, \alpha_i \rangle$$

where:

- $R \subseteq \mathcal{X}$ are relevant variables;
- $a : R \to \mathcal{X}'$ is a surjective function between *variables*;
- $\alpha_i : \mathcal{M}[a^{-1}(X_i')] \to \mathcal{M}'[X_i']$ is a collection of surjective functions between *outcomes*.



$$\boldsymbol{\alpha} : \begin{cases} R = \{S, C, T\} \\ a(S) \mapsto S', a(T) \mapsto T', a(C) \mapsto C' \\ \alpha_{S'}(s) \mapsto s, \alpha_{T'}(t) \mapsto t, \alpha_{C'}(c) \mapsto c \end{cases}$$

Abstraction Error [9]

Given two (disjoint set of) variables in $\mathcal{X}'$, we evaluate **abstraction error** in terms of *interventional consistency* $E_{\boldsymbol{\alpha}}(X', Y')$ as the maximum *distance between interventional distributions*.

$$
\begin{tikzcd}
\mathcal{M}[S] \arrow[r, "\mu", "P_{\mathcal{M}_\iota}(T|do(S))"'] \arrow[d, "\alpha_{S'}"'] & \mathcal{M}[T] \arrow[d, "\alpha_{C'}"] \\
\mathcal{M}'[S'] \arrow[r, "\nu", "P_{\mathcal{M}'_{\iota'}}(T'|do(S'))"'] & \mathcal{M}'[T']
\end{tikzcd}
$$

$$E_{\boldsymbol{\alpha}}(S', T') = \max_{\iota} D_{JSD}(\alpha_{T'} \cdot \mu, \nu \cdot \alpha_{S'})$$

Abstraction Errors [9]

An abstraction implies multiple *abstraction errors*.

(Global) abstraction error $e(\boldsymbol{\alpha})$ is the maximum abstraction error over all disjoint sets of variable.

$$\begin{tikzcd} \mathcal{M}[S] \arrow[r, "\mu"] \arrow[d, "\alpha_{S'}"'] & \mathcal{M}[T] \arrow[d, "\alpha_{T'}"] \\ \mathcal{M}'[S'] \arrow[r, "\nu"] & \mathcal{M}'[T'] \end{tikzcd} \qquad \begin{tikzcd} \mathcal{M}[T] \arrow[r, "\mu'"] \arrow[d, "\alpha_{T'}"'] & \mathcal{M}[C] \arrow[d, "\alpha_{C'}"] \\ \mathcal{M}'[T'] \arrow[r, "\nu'"] & \mathcal{M}'[C'] \end{tikzcd}$$

$$\begin{tikzcd} \mathcal{M}[S] \arrow[r, "\mu' \circ \mu"] \arrow[d, "\alpha_{S'}"'] & \mathcal{M}[C] \arrow[d, "\alpha_{C'}"] \\ \mathcal{M}'[S'] \arrow[r, "\nu' \circ \nu"] & \mathcal{M}'[C'] \end{tikzcd}$$

$$e(\boldsymbol{\alpha}) = \sup_{\mathbf{X}', \mathbf{Y}' \subseteq \mathcal{X}'} E_{\boldsymbol{\alpha}}(\mathbf{X}', \mathbf{Y}')$$

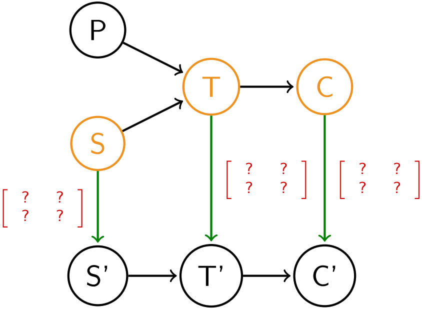
3. Abstraction Learning

Joint work of FMZ, M. Drávucz, G. Apachitei, W.D. Widanage and T. Damoulas

Problem statement [11]

Given a partially define *abstraction* $\boldsymbol{\alpha}$ in terms of $\langle R, a \rangle$ can I learn α_i as:

$$\min_{\boldsymbol{\alpha}} \; e(\boldsymbol{\alpha})$$



Challenges [11]

(i) *Multiple related problems*

(ii) *Combinatorial optimization*

(iii) *Surjectivity constraints*

Baselines: parallel or sequential approaches.

$$\alpha_{S'} = \begin{bmatrix} ? & ? \\ ? & ? \end{bmatrix}, \alpha_{T'} = \begin{bmatrix} ? & ? \\ ? & ? \end{bmatrix}, \alpha_{C'} = \begin{bmatrix} ? & ? \\ ? & ? \end{bmatrix}$$

$$\begin{tikzcd} \mathcal{M}[S] \arrow[r, "\mu"] \arrow[d, "\alpha_{S'}"'] & \mathcal{M}[T] \arrow[d, "\alpha_{T'}"] \\ \mathcal{M}'[S'] \arrow[r, "\nu"] & \mathcal{M}'[T'] \end{tikzcd} \qquad \begin{tikzcd} \mathcal{M}[T] \arrow[r, "\mu'"] \arrow[d, "\alpha_{T'}"'] & \mathcal{M}[C] \arrow[d, "\alpha_{C'}"] \\ \mathcal{M}'[T'] \arrow[r, "\nu'"] & \mathcal{M}'[C'] \end{tikzcd}$$

$$\begin{tikzcd} \mathcal{M}[S] \arrow[r, "\mu' \circ \mu"] \arrow[d, "\alpha_{S'}"'] & \mathcal{M}[C] \arrow[d, "\alpha_{C'}"] \\ \mathcal{M}'[S'] \arrow[r, "\nu' \circ \nu"] & \mathcal{M}'[C'] \end{tikzcd}$$

Relaxation and parametrization [11]

We address *(ii) combinatorial optimization* by *relaxing* and *parametrizing* all α_i.

$$\min_{\boldsymbol{\alpha}(\mathbf{W})} \; e(\boldsymbol{\alpha}(\mathbf{W}))$$

$$\alpha_{S'}, \alpha_{T'}, \alpha_{C'} \in \mathbb{R}^{2\times 2}$$

$$\begin{bmatrix} 0.7 & 1.2 \\ -0.2 & 3.3 \end{bmatrix}$$

We add *tempering* $t(W) = \frac{e^{\frac{W_i j}{T}}}{\sum_i e^{\frac{W_i j}{T}}}$ along the matrix columns to binarize them.

$$\mathcal{L}_1 : \min_{\boldsymbol{\alpha}(\mathbf{W})} \; e(\boldsymbol{\alpha}(t(\mathbf{W})))$$

$$\alpha_{S'}, \alpha_{T'}, \alpha_{C'} \in [0,1]^{2\times 2}$$

$$t\left(\begin{bmatrix} 0.7 & 1.2 \\ -0.2 & 3.3 \end{bmatrix}\right) = \begin{bmatrix} 0.99 & 0.02 \\ 0.01 & 0.98 \end{bmatrix}$$

Enforcing surjectivity [11]

We address *(iii) surjective constraints* through a *penalty function*:

$$\mathcal{L}_2 : \min_{\mathbf{W}} \sum_{\mathbf{W}} \sum_{i} \left(1 - \max_{j} t(W)_{ij}\right)$$

$$\alpha_{S'}, \alpha_{T'}, \alpha_{C'} \in [0,1]^{2\times 2}$$

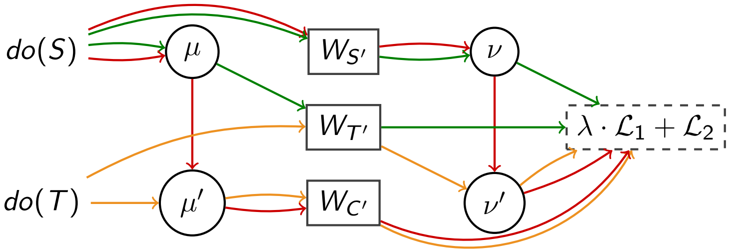
$$\begin{bmatrix} 0.99 & 0.02 \\ 0.01 & 0.98 \end{bmatrix} \overset{\mathcal{L}_2}{\rightsquigarrow}$$

$$(1-0.99)+(1-0.98)$$

Solution by gradient descent [11]

We address *(i) multiple related problems* by *jointly* solving all the problems via *gradient descent*:

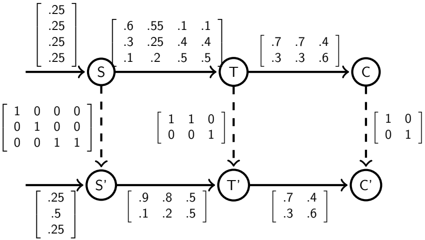
$$\begin{tikzcd} \mathcal{M}[S] \arrow[r, "\mu"] \arrow[d, "\alpha_{S'}"'] & \mathcal{M}[T] \arrow[d, "\alpha_{T'}"] \\ \mathcal{M}'[S'] \arrow[r, "\nu"] & \mathcal{M}'[T'] \end{tikzcd} \qquad \begin{tikzcd} \mathcal{M}[T] \arrow[r, "\mu'"] \arrow[d, "\alpha_{T'}"'] & \mathcal{M}[C] \arrow[d, "\alpha_{C'}"] \\ \mathcal{M}'[T'] \arrow[r, "\nu'"] & \mathcal{M}'[C'] \end{tikzcd} \qquad \begin{tikzcd} \mathcal{M}[S] \arrow[r, "\mu' \circ \mu"] \arrow[d, "\alpha_{S'}"'] & \mathcal{M}[C] \arrow[d, "\alpha_{C'}"] \\ \mathcal{M}'[S'] \arrow[r, "\nu' \circ \nu"] & \mathcal{M}'[C'] \end{tikzcd}$$

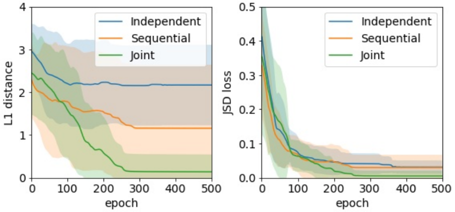


Synthetic Experiments [11]

We evaluated our learning method:

- On multiple synthetic models;
- Against *independent* and *sequential* approach;
- Monitoring *loss functions*, *L1-dist from ground truth*, *wall-clock time*.





Real-World Experiments [11]

We want to model the stage of **coating** in lithium-ion battery manufacturing:

$$\text{Mass Loading} = \text{f(input)}$$

Experiments are costly, so we want to integrate data[1] collected by two groups running similar (but not identical) experiments:

LRCS (France)	**WMG** (UK)
Collection of few statistics in each a few stages of battery manufacturing [2].	Collection of detailed space- and time-dependent measurements during coating.

[1] https://chemistry-europe.onlinelibrary.wiley.com/doi/full/10.1002/batt.201900135
https://github.com/mattdravucz/jointly-learning-causal-abstraction/

Real-World Experiments [11]

We evaluated our learning method:

- Performing abstraction of data from base to abstracted (WMG $\rightarrow$ LRCS);
- Evaluating change in performance using aggregated data when predicting *out-of-sample* (k).

	Training set	**Test Set**	**MSE**
(a)	LRCS[$CG \neq k$]	LRCS[$CG = k$]	1.86 ± 1.75
(b)	LRCS[$CG \neq k$] + WMG	LRCS[$CG = k$]	0.22 ± 0.26
(c)	LRCS[$CG \neq k$] + WMG[$CG \neq k$]	LRCS[$CG = k$] + WMG[$CG = k$]	1.22 ± 0.95

Conclusion

- *Causality* and *abstraction* may both play important role in modelling.
- A first proposal for *learning abstraction*.
- Preliminary results show promise for *transporting* data.

Large space for conceptual and practical development of **causal abstraction frameworks**:

- *Foundations* of the framemorks
- *Characterization* of these frameworks
- *Algorithmic and empirical* development

More about causal abstraction:
https://github.com/FMZennaro/CausalAbstraction/

Thanks!

Thank you for listening!

References I

[1] Sander Beckers and Joseph Y Halpern. Abstracting causal models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 2678–2685, 2019.

[2] Ricardo Pinto Cunha, Teo Lombardo, Emiliano N Primo, and Alejandro A Franco. Artificial intelligence investigation of nmc cathode manufacturing parameters interdependencies. *Batteries & Supercaps*, 3(1):60–67, 2020.

[3] Luciano Floridi. The method of levels of abstraction. *Minds and machines*, 18(3):303–329, 2008.

[4] Jun Otsuka and Hayato Saigo. On the equivalence of causal models: A category-theoretic approach. *arXiv preprint arXiv:2201.06981*, 2022.

[5] Jun Otsuka and Hayato Saigo. The process theory of causality: an overview. 2022.

References II

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

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

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

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

[10] 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.

References III

[11] Fabio Massimo Zennaro, Máté Drávucz, Geanina Apachitei, W. Dhammika Widanage, and Theodoros Damoulas. Jointly learning consistent causal abstractions over multiple interventional distributions. In *2nd Conference on Causal Learning and Reasoning*, 2023.