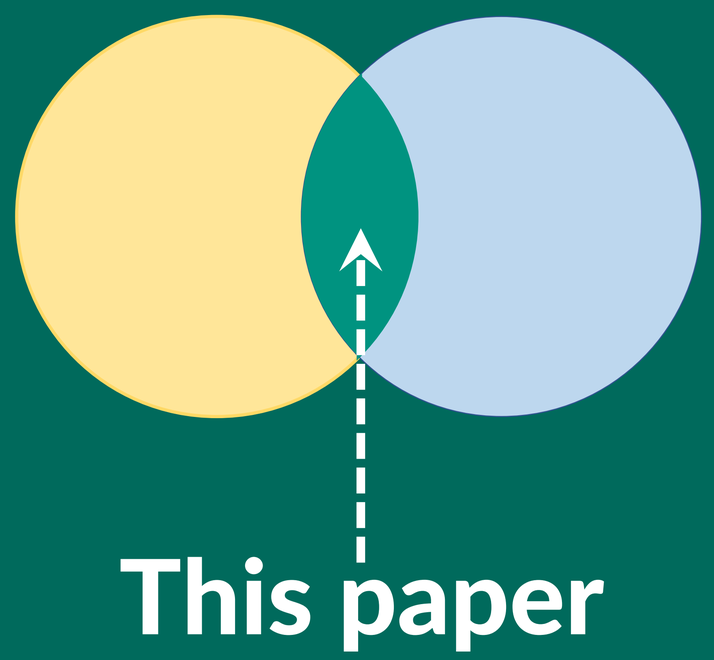


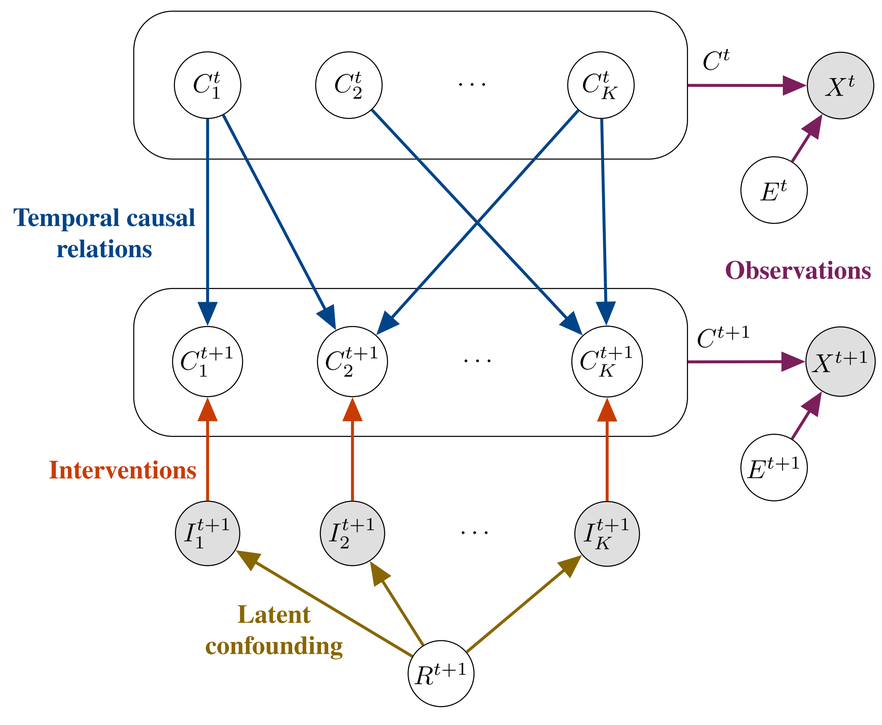
Intervention Design for Causal Representation Learning

Phillip Lippe[1] (p.lippe@uva.nl), Sara Magliacane[1,2], Sindy Löwe[1], Yuki M. Asano[1], Taco Cohen[3], Efstratios Gavves[1]

[1]University of Amsterdam, Informatics Institute, [2]MIT-IBM Watson AI Lab, [3]Qualcomm AI Research, The Netherlands

PROBLEM SETTING

- *Experiment:* Set of variables that are simultaneously intervened
- *Intervention Design for Causal Discovery*: $\lfloor \log_2 K \rfloor + 1$ experiments for K known variables identify the causal graph (worst case) [1]
- *CITRIS [2]*: Causal variables are identifiable from videos with interventions if intervention targets are not deterministic functions of each other ⇒ How many experiments are necessary?



METHOD

- Four conditions: Each variable must be observed (1) passively and (2) intervened at least once, and for each pair of variables, the targets cannot be (3) equal or (4) different at *all* time steps
- Derivation by considering targets as binary codes, and results:

$\lfloor \log_2 K \rfloor + 2$ experiments identify the minimal causal variables

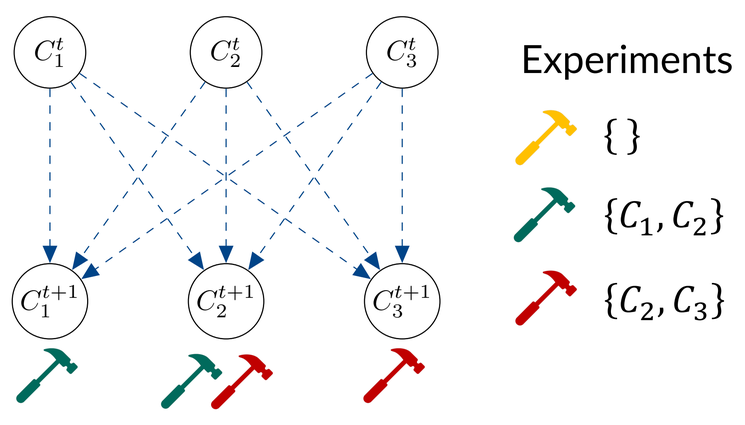
ALGORITHM

Require: Number of variables K

1: Create all possible binary codes of length $L = \lfloor \log_2(K) \rfloor + 1$ as set $\mathcal{B} = \{0, 1\}^L$
2: Remove the code of observing a variable passively in all experiments, $\{0\}^L$, from $\mathcal{B}$
3: Extend all codes in $\mathcal{B}$ by appending $\{0\}$, *i.e.* an experiment where all variables are passively observed
4: From the remaining codes in $\mathcal{B}$, (arbitrarily) pick K unique codes $b^1, ..., b^K$, one for each causal variable C_i
5: Create experiments by using the codes as binary intervention targets: $E_l = \{C_i | i \in [\![1..K]\!], b_l^i = 1\}$

Example

$\lfloor \log_2 3 \rfloor + 2 = 3$ experiments needed



EXPERIMENTS

- Temporal Causal3DIdent with 6 variables ⇒ $\lfloor \log_2 6 \rfloor + 2 = 4$ experiments necessary
- CITRIS with minimal experiments is close to full set of experiments in disentanglement

Experimental setting	Triplets ↓	R^2 diag ↑	R^2 sep ↓
iVAE - Full experiments	0.15 (±0.01)	0.78 (±0.04)	0.21 (±0.10)
CITRIS - Full experiments	0.04 (±0.00)	0.98 (±0.00)	0.01 (±0.00)
CITRIS - Minimal experiments	0.12 (±0.02)	0.94 (±0.05)	0.08 (±0.05)

References
[1] Hyttinen, Antti, Frederick Eberhardt, and Patrik O. Hoyer. "Experiment selection for causal discovery." Journal of Machine Learning Research 14 (2013): 3041-3071.
[2] Lippe, Phillip, et al. "CITRIS: Causal Identifiability from Temporal Intervened Sequences." *International Conference on Machine Learning*. PMLR, 2022.

Check out the full paper!