Causal Inference Team
Shohei Shimizu
因果推論チーム　清水 昌平

What is causal discovery?

- Methodology for inferring causal graphs using data

Assumptions

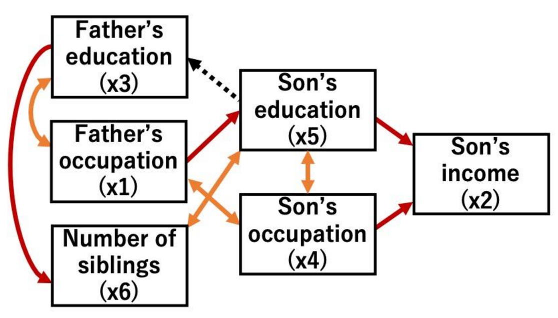
- Functional form?
- Distribution?
- Hidden common cause present?
- Acyclic? etc.

+

Data

x1	x2	x3	x4	x5	x6
87.9	45433	17	76.3	17	1
87.9	55071	16	86	18	2
62.1	113159	16	87.9	16	0
78.5	30289	16	30.1	14	4
32.3	113159	20	63.5	20	7
60.6	55071	17	83.7	17	1
76.4	55071	16	78	14	2
63.5	37173	12	63.2	16	3
63.2	113159	14	86.5	17	1
36.5	37173	12	83.7	12	4

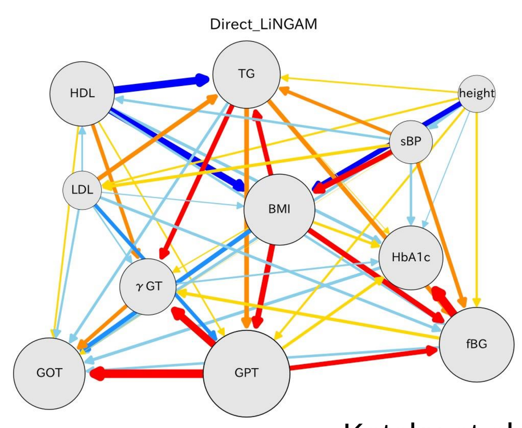
Causal graph



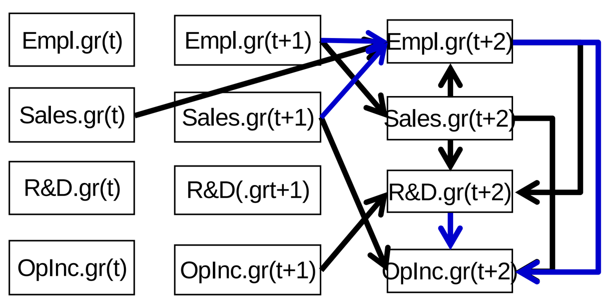
Maeda and Shimizu (2020)

Application examples

- Biology (Maathuis et al., 2010)
- Prevention medicine (Kotoku et al., 2020)
- Chemistry (Campomanes et al., 2014)
- Material Science (Nelson et al., 2021)
- Climatology (Liu et al., 2020)
- Economics (Moneta et al., 2013)
- Psychology (von Eye et al., 2012)
- Policymaking (高山ら, 2021)
- Network log data (Jarry et al., 2021)



Kotoku et al. (2020)



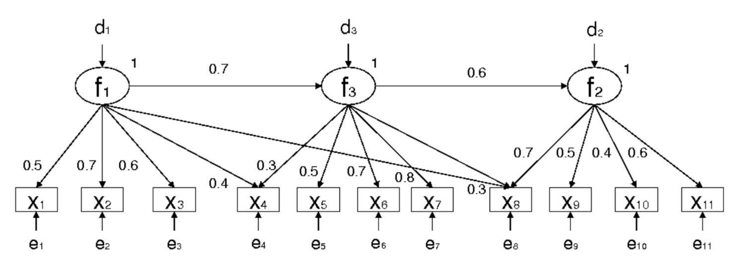
Moneta et al. (2013)

https://www.shimizulab.org/lingam/lingampapers/applications-and-tailor-made-methods

LiNGAM for latent factors (Shimizu et al., 2009)

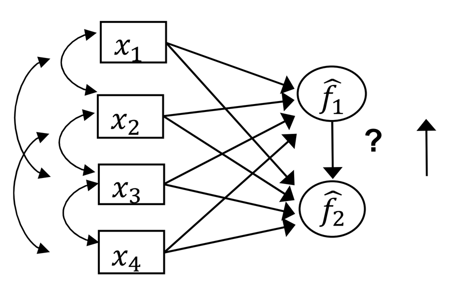
- Model:

$$\boldsymbol{f} = B\boldsymbol{f} + \boldsymbol{\epsilon}$$
$$\boldsymbol{x} = G\boldsymbol{f} + \boldsymbol{e}$$



  – 2 pure measurement variables per latent needed to identify the measurement model (Silva et al., 2006; Xie et al., 2020)

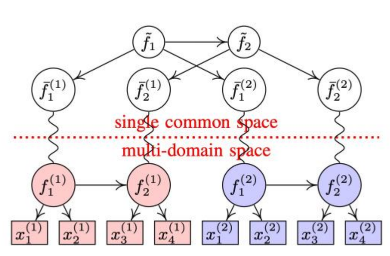
- Estimate the latent factors and then their causal graph



Find common and unique factors across multiple datasets (Zeng et al., 2021)

- Model

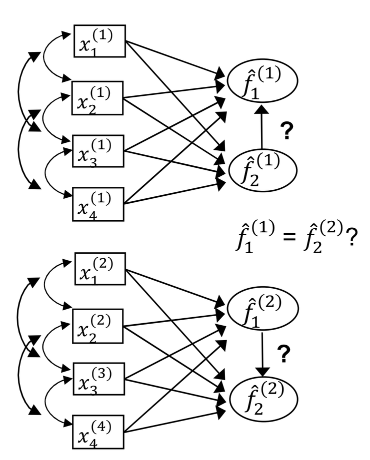
$$\boldsymbol{f}^{(m)} = B^{(m)}\boldsymbol{f}^{(m)} + \boldsymbol{\epsilon}^{(m)}$$
$$\boldsymbol{x}^{(m)} = G^{(m)}\boldsymbol{f}^{(m)} + \boldsymbol{e}^{(m)}$$
$$m = 1, \ldots, M$$



- Score function: likelihood + DAGness (Zheng et al., 2018)

DAGness (Zheng et al., NeurIPS2018): $h(B) = \mathrm{tr}(e^{B \circ B}) - \mathrm{N.\,factors}$

- Feature extraction across multiple datasets + causal discovery of latent factors



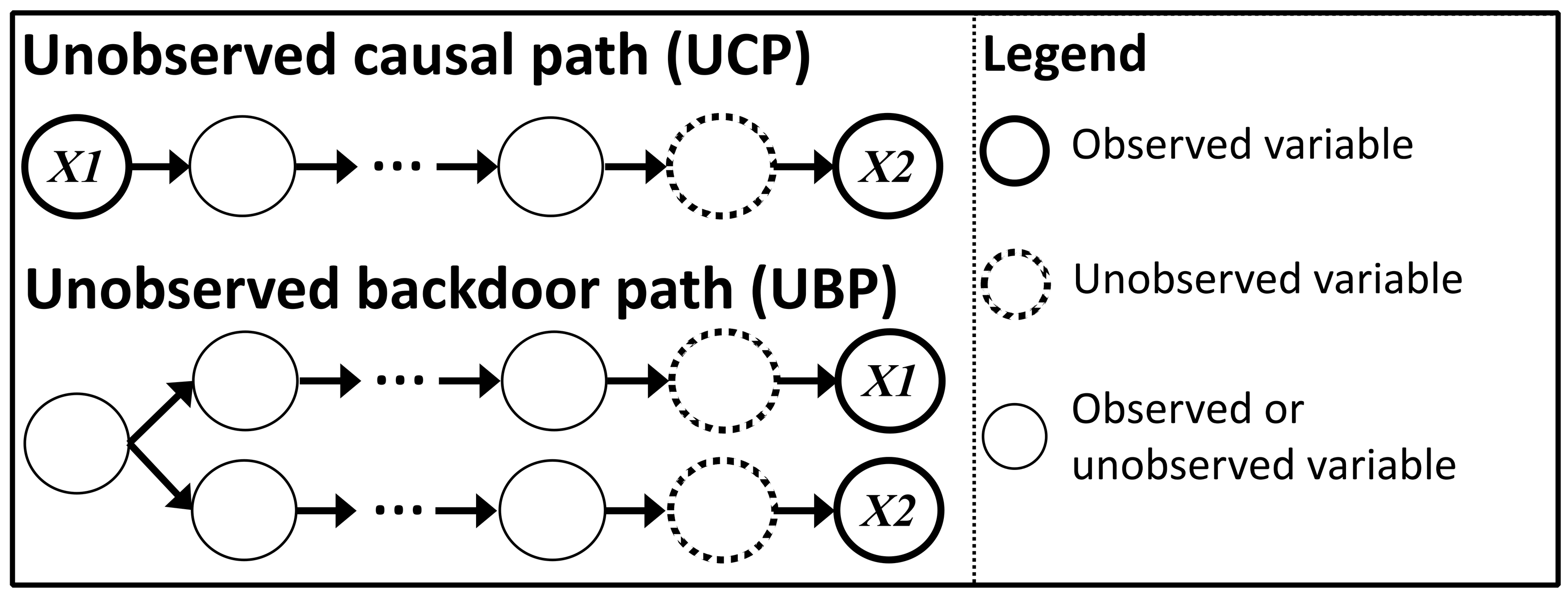
Y. Zeng, **S. Shimizu**, R. Cai, F. Xie, M. Yamamoto, Z. Hao. **Causal discovery with multi-domain LiNGAM for latent factors**. In *Proc. the 30th International Joint Conference on Artificial Intelligence (IJCAI2021), 2021.*

Causal additive models with unobserved variables

Identifiability of indirect causal relationships in linear and nonlinear cases

- **Assume that X1 has an indirect causal effect on X3 mediated through unobserved variable Y2.**
- **In linear cases:**
  - The residual obtained by the regression of X3 on X1 is independent of X1.
- **In nonlinear cases (Cascade additive noise models):**
  - The residual obtained by the regression of X3 on X1 cannot be independent of X1.

Cases in which causal relationships between variables cannot be identified



Causal additive models with unobserved variables (CAM-UV)

$$v_i = \sum_{x_j \in P_i} f_j^{(i)}(x_j) + \sum_{y_k \in Q_i} f_k^{(i)}(y_k) + n_i$$

The effects of unobserved variables

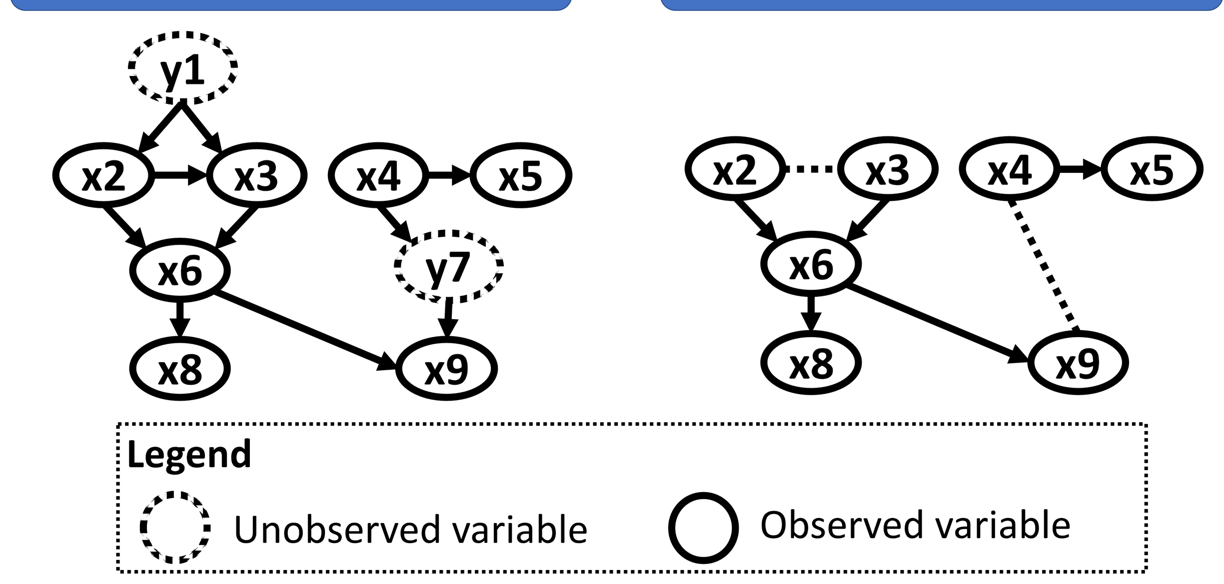
- v_i : Observed or unobserved variable
- x_i : Observed variable
- y_i : Unobserved variable
- n_i : External effects
- f : Nonlinear function
- P_i : Direct observed cause of x_i
- Q_i : Direct unobserved cause of x_i.

The objective of CAM-UV

- **Infer the causal directions of all the variable pairs that do not have a UCP or UBP.**

Data generation process

Causal graph



T. N. Maeda, S. Shimizu. **Causal additive models with unobserved variables**. In Proc. 37th Conf. on Uncertainty in Artificial Intelligence (UAI2021), 2021