Causal Discovery from Temporal Data

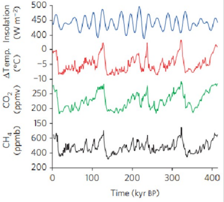
Chang Gong, Di Yao, Chuzhe Zhang, Wenbin Li, Jingping Bi, Lun Du, Jin Wang

ICT, Chinese Academy of Sciences
Microsoft Research Asia
Megagon Labs

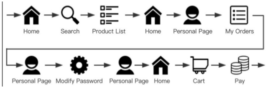
August 6, 2023

Temporal Data

- It records the status changing of complex systems.
- It consists of attribute sequences ordered by time.
- According to whether the data is calibrated, it can be categorized into two groups, *i.e.*, multivariate time series (MTS) and event sequences.



MTS



Event Sequence

Temporal Data

- It is widely collected by various application domains, such as social networks, bioinformatics, neuroscience and finance.

- Different approaches have been proposed for different tasks such as classification[1], clustering[2], prediction[3].

[1] Hassan Ismail Fawaz, Germain Forestier, Jonathan Weber, Lhassane Idoumghar, and Pierre-Alain Muller. *Deep learning for time series classification: a review*. DMKD, 2009.

[2] Saeed Aghabozorgi, Ali Seyed shirkhorshidi, and Teh Ying Wah. *Time-series clustering-a decade review*. INFORM. SYST., 2015.

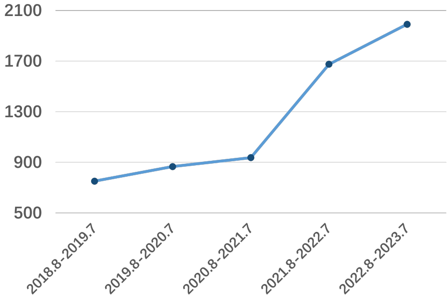
[3] Andreas S Weigend. *Time series prediction: forecasting the future and understanding the past*. Routledge, 2018

Causal Discovery for Temporal Data

- Causal discovery for temporal data recognizing the causal relations between many temporal components has become a challenging yet critical task for temporal data analysis.
- The learned causal structures could be beneficial for explaining the data generation process and guiding the design of data analysis methods.
- The wide applications in many fileds:
  - Scientific endeavours: Earth science, neuroscien, bioinformatics...
  - Industrial implementations: Anomaly detection, root cause analysis, urban data analysis, ...

Causal Discovery for Temporal Data

- The heat of causal discovery has increased rapidly in the past five years[4].

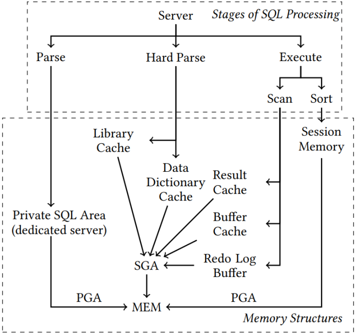


[4] https://trends.google.com

Causal Discovery for Temporal Data

Application in root cause analysis[5].

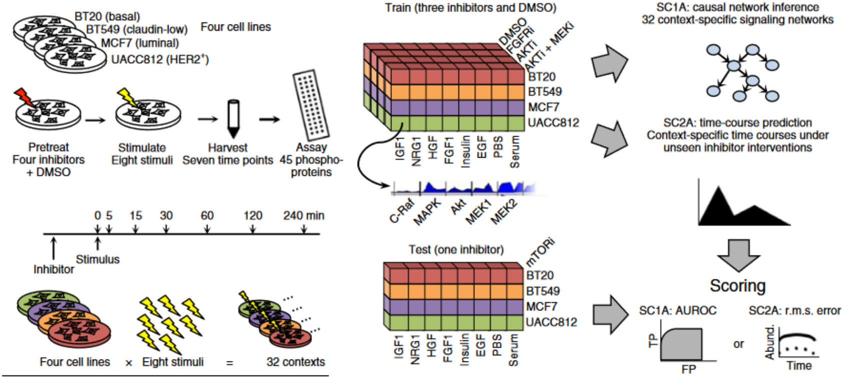
- Most applications use lots of sub-components that interact with each other in the form of a complex graph.
- Causal discovery can help to construct this graph and find the root cause of the failures in a timely fashion.



[5] Mingjie Li, Xiaohui Nie, Dan Pei. *Causal inference-based root cause analysis for online service systems with intervention recognition*. KDD, 2022

Causal Discovery for Temporal Data

- Discovering causal structures of molecular networks helps elucidate the genetic and biological mechanisms of diseases [6] [7].



[6] Jack Kelly, Carlo Berzuini, Bernard Keavney, Maciej Tomaszewski, Hui Guo. *A review of causal discovery methods for molecular network analysis.* Mol Genet Genomic Med. 2022

[7] Steven M Hill, Laura M Heiser, Thomas Cokelaer et al. *Inferring causal molecular networks: empirical assessment through a community-based effort.* Nat. Methods 2016.

Overview

This tutorial is organized as follows:

- Preliminaries (15 mins)
- Causal Discovery for Multivariate Time Series (60 mins)
- Causal Discovery for Event Sequence (60 mins)
- Discussion about Applications and New perspectives (15 mins)

1. Background
2. **Basic Concepts**
3. Assumptions
4. Problem Definitions

Basic Concepts

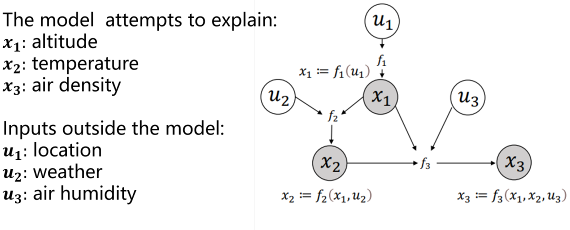
To formalize causal relationships and the task of causal discovery, these basic concepts about causality will be introduced:

- Structural Causal Model (SCM)
- Causal Markov Condition

Structural Causal Model

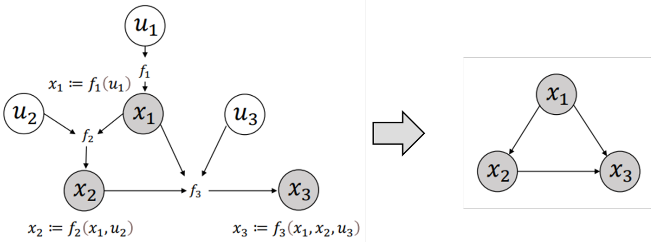
SCM is defined as a graphical representation of causal relationships, it can be represented in a 4-tuple $< V, U, F, P(U) >$

- V, U denote the set of endogenous and exogenous variables.
- $P(U)$ is the distribution of exogenous variables.
- F represents the set of the mapping function. For $f_i \in F$, the model $x_i := f_i(pa_i, u_i)$ indicates the assignment of the value x_i.



Structural Causal Model

For each SCM, we can yield a causal graph DAG G by adding one vertex for each X_i and directing edges from each parent variable in pa_i (the causes) to child X_i (the effect).



Causal Markov Condition

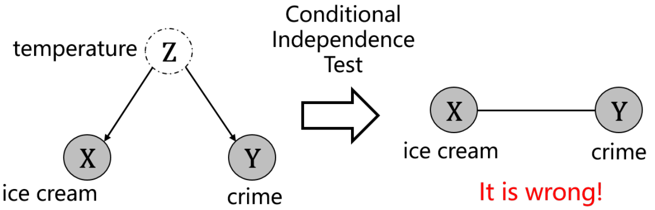
In the causal graph of SCM, each variable is conditionally independent of its non-effects (i.e., variables that do not directly cause it) given its parents. Formally, the causal Markov condition implies the joint distribution can be factorized according to the following decomposition:

$$P(\mathbf{X}) = \prod_{i}^{d} P(X_i|pa_i)$$

This condition plays an essential role in causal inference. It enables the identification of causal effects from observed data.

Causal Sufficiency

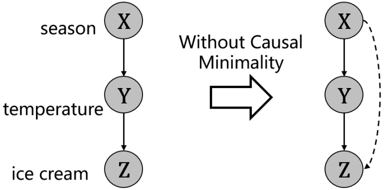
- A set of variables is causally sufficient if all common causes of all variables are observed.
- Without this assumption, the causal structure discovered by Conditional Independence Test is not credible.



- This assumption greatly simplifies the problem of causal discovery, but it does not always hold in reality.

Causal Minimality

- Consider a DAG $\mathcal{G}$ and a probability distribution P, P is said to satisfy the causal minimality with respect to $\mathcal{G}$ if P is Markovian with respect to $\mathcal{G}$ but not to any subgraph of $\mathcal{G}$.
- A distribution is not minimal with respect to the causal graph, if and only if some parents are "inactive".



- It is reasonable since a causal graph without this assumption does not provide more information than the minimal graph.

Causal Faithfulness

- Faithfulness asserts that all conditional independence relations of P that hold in the observed data are entailed by the causal model $\mathcal{G}$.
- This assumption is essential to Conditional Independence test, and it always holds except for some special cases.
- Note that faithfulness implies causal minimality. If P is faithful and Markovian with respect to $\mathcal{G}$, then the causal minimality is satisfied.

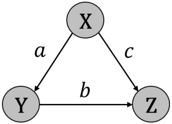
Causal Faithfulness

An example that violates this causal faithfulness assumption

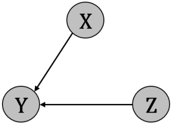
- the generation process of $\mathcal{G}_1$ as a linear Gaussian SCM:

$$\begin{aligned} X &:= N_X \\ Y &:= aX + N_Y \\ Z &:= bY + cX + N_Z \end{aligned}$$

  where the noise variables N_X, N_Y and N_Z are jointly independent and conform $\mathcal{N}(0, \sigma^2)$, and $a \cdot b + c = 0$.
- X and Z are independent, and it looks like the data was generated by $\mathcal{G}_2$



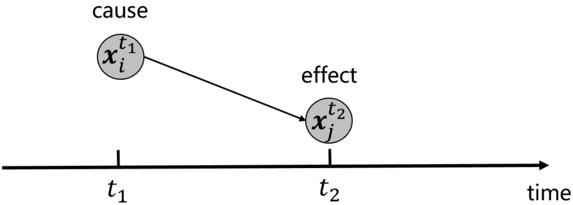
$\mathcal{G}_1$



$\mathcal{G}_2$

Temporal priority

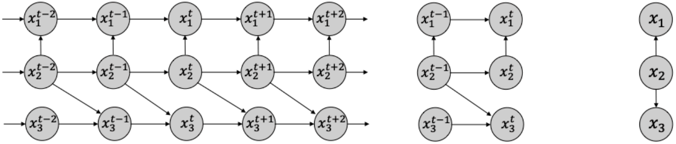
- For two variables, temporal priority means that the cause must have occurred before its effect.
- It is a foundation assumption of causal discovery from temporal data and creates an asymmetric time relationship in causal processes. It helps us to establish the direction of a causal relationship when two variables are causally linked.



Causal Structure for MTS

Given d-variate time series $\mathbf{x} = (x_1^t, \ldots, x_d^t)$:

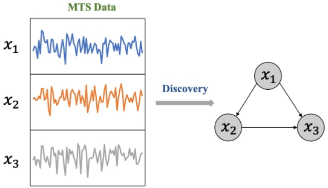
- Full-time causal graph represents a complete graph of the dynamic systems. Each variate at each time point is a vertex in the graph.
- Window causal graph is a partial full-time graph within a given time window.
- Summary causal graph ignores timestamps in the graph that each time series component is collapsed into a node.



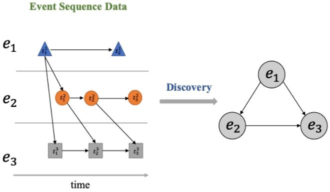
(a) Full time causal graph **(b) Window causal graph** **(c) Summary causal graph**

Causal Discovery from Temporal Data

Given a multivariate time series, causal discovery from MTS aims to find one of the causal structures, *i.e.* summary causal graph or window causal graph.



Given an event sequence: $\{(t_1, e_1), (t_2, e_2), \ldots\}$, causal discovery for event sequences aims to discover the causal relationships between different event types.



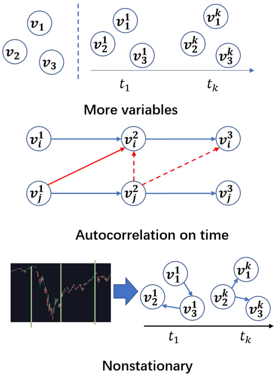
Challenges

General causal discovery is challenging due to the following reasons:

- hidden confounder, nonlinear, heterogeneity

In addition to the above challenges, there are more challenges in temporal causal discovery:

- more variables,
- autocorrelation on time
- nonstationary



What's Next?

- Causal Discovery from Multivariate Time Series (60 mins)
- Causal Discovery from Event Sequence (60 mins)
- Discussion about Applications and New perspectives (15 mins)

Thanks!