Causal Discovery from Temporal Data

Multi-variant Time-series

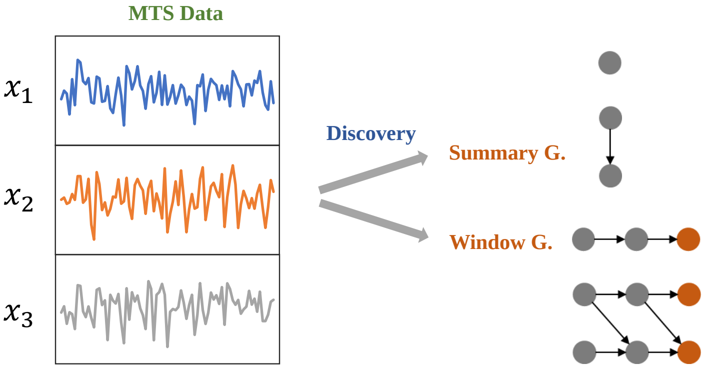
Chang Gong, Di Yao, Chuzhe Zhang, Wenbin Li, Jingping Bi, Lun Du, Jin Wang

ICT, Chinese Academy of Sciences
Microsoft Research Asia
Megagon Labs

August 6, 2023

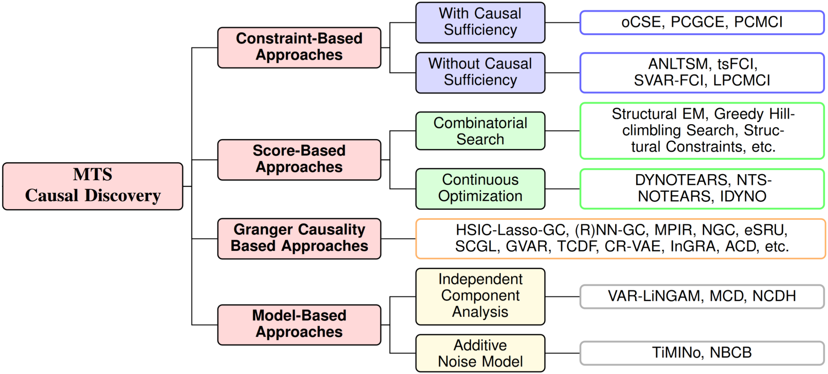
Recap: Problem Definition

Causal discovery from MTS aims to find one of the causal structures, *i.e.*, summary causal graph or window causal graph.



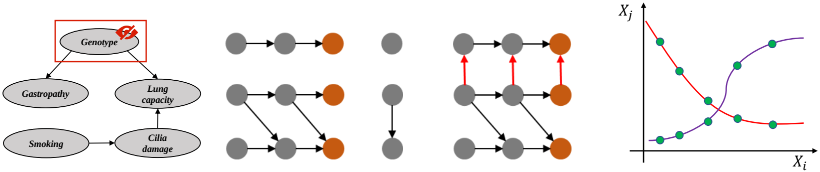
Overview

Categorize existing methods into four groups:



Characteristics

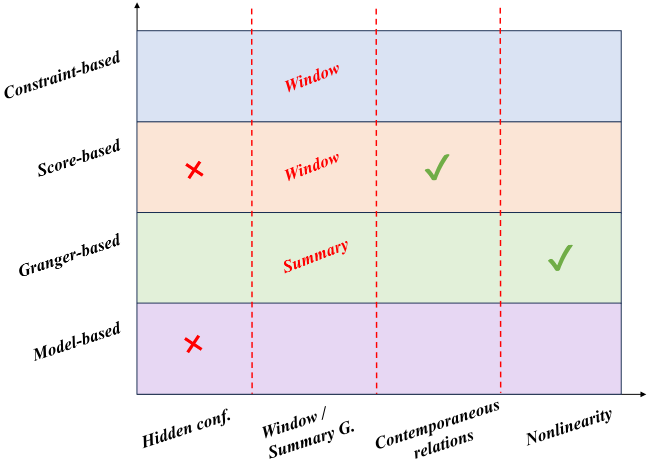
Four characteristics to qualify the existing methods



i) Hidden confounders *ii) Window / Summary graph* *iii) Contemporaneous relations* *iv) Nonlinearity*

Characteristics

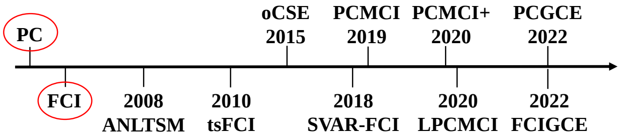
Four characteristics to qualify the existing methods



Overview

Basic Idea: Learn Causal Relations via Conditional Independence Test. $X_i \perp\!\!\!\perp X_j | \{X_k\}$

According to whether assume no **hidden confounders**:



- **PC**'s extension (Assume No Hidden Confounders)
  - PCMCI, PCMCI+, ...
- **FCI**'s extension (Under Hidden Confounder)
  - tsFCI, SVAR-FCI, LPCMCI, ...

PC Algorithm

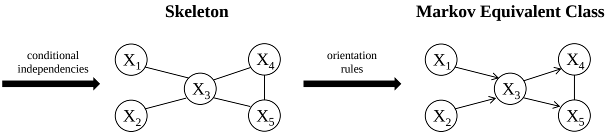
Non-temporal Origins

Peter-Clark Algorithm

- Non-temporal origins of the following methods
- Assume **no hidden confounder** / causal sufficiency
- Steps:
  - Skeleton identification: ***d*-separation and conditional independence test**
  - Orientation rules

Data

X_1	X_2	X_3	X_4	X_5
0.1	0.5	1.2	0.9	0.9
0.2	0.3	1.5	1.1	1.0
…	…	…	…	…

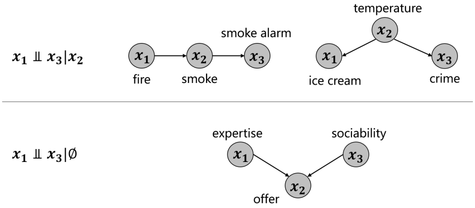


Sprites, Glymour, and Scheines. *Causation, Prediction, and Search, Second Edition.* MIT Press, 2000.

PC Algorithm

d-separation and conditional independence test

- A set of variables S d-separates two variables if S blocks all paths between them.
- If S d-separates two variables X and Y, then X and Y are independent conditioned on S. For instance, x_1 and x_3 are independent conditioned on x_2 in following figures:



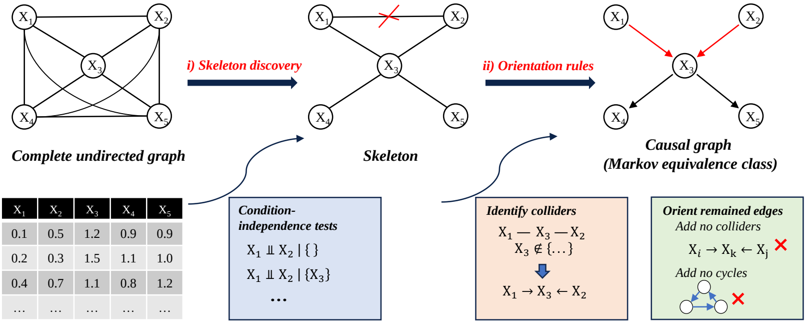
- **Helpful in skeleton identification and orientations!**

PC Algorithm

Procedures

Peter-Clark Algorithm

- Steps in detail

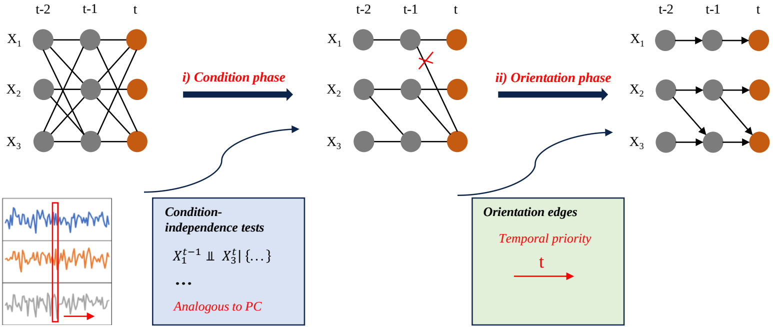


X_1	X_2	X_3	X_4	X_5
0.1	0.5	1.2	0.9	0.9
0.2	0.3	1.5	1.1	1.0
0.4	0.7	1.1	0.8	1.2
…	…	…	…	…

PC's Extension

PC's Extension to Time Series Data

- Direct extension: **time window**
- **Spurious correlation** due to **autocorrelation**

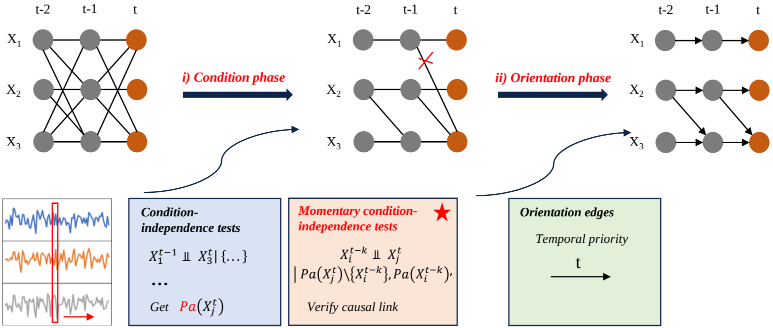


Runge, Nowack, Kretschmer, Flaxman, and Sejdinovic. *Detecting and quantifying causal associations in large nonlinear time series datasets*. Science advances, 2019.

PC's Extension

PC's Extension to Time Series Data

- To tackle autocorrelation : momentary conditional independence test (**PCMCI**)



Runge, Nowack, Kretschmer, Flaxman, and Sejdinovic. *Detecting and quantifying causal associations in large nonlinear time series datasets*. Science advances, 2019.

PC's Extension

PCMCI[1]

- Motivation: tackle autocorrelation in time series
- Idea: PC + **MCI test**
- Advantage: eliminate spurious correlation $\rightarrow$ false positive rate is reduced
- Feature: hidden conf. $\times$, window G., contemporaneous $\times$, nonlinear $\surd$,

PCMCI+[2]

- Motivation: + model **contemporaneous** relations
- Idea: + contemporaneous MCI test
- Feature: hidden conf. $\times$, window G., contemporaneous $\surd$, nonlinear $\surd$

[1] Runge, Nowack, Kretschmer, Flaxman, and Sejdinovic. *Detecting and quantifying causal associations in large nonlinear time series datasets*. Science advances, 2019.

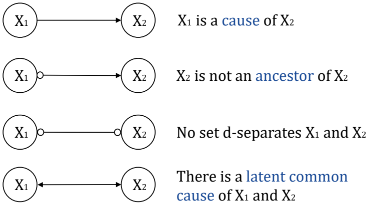
[2] Runge. *Discovering contemporaneous and lagged causal relations in autocorrelated nonlinear time series datasets*. UAI, 2020.

FCI Algorithm

Non-temporal Origins

Fast Causal Inference Algorithm

- Tackle **hidden confounders**
- Analogous to **PC** algorithm:
  - Skeleton identification
  - Orientation rules
- Difference:
  - **New form of causal graph**
  - New orientation rules



https://causal-learn.readthedocs.io/

FCI Algorithm

Non-temporal Origins

Fast Causal Inference Algorithm

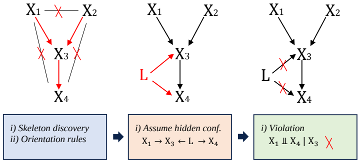
- Difference:
  - **New orientation rules**

Example 1:

$X_1 \perp\!\!\!\perp X_2$;

$X_1 \perp\!\!\!\perp X_4 | X_3$;

$X_2 \perp\!\!\!\perp X_4 | X_3$.

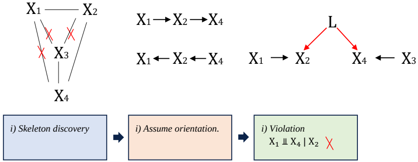


Example 2:

$X_1 \perp\!\!\!\perp X_3$;

$X_1 \perp\!\!\!\perp X_4$;

$X_2 \perp\!\!\!\perp X_3$.

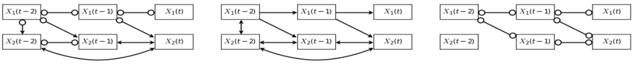


Jiji Zhang. *Causal reasoning with ancestral graph*. JMLR, 2008.

FCI's Extension

tsFCI[1]

- Motivation: extend FCI to MTS
- Idea: **time window (MTS $\rightarrow$ non-temporal), temporal priority and stationary assumption for orientation**
- Feature: hidden conf. $\surd$, window G., contemporaneous $\times$, nonlinear $\surd$



(a) original FCI with oracle (b) tsFCI with oracle (c) original FCI on data

SVAR-FCI[2]

- Motivation: Enable contemporaneous causal relationships
- Idea: $+$ structural VAR assumption
- Feature: hidden conf. $\surd$, window G., contemporaneous $\surd$, nonlinear $\times$

[1] Entner and Hoyer. *On causal discovery from time series data using FCI*. PGM, 2010.

[2] Malinsky and Sprites. *Causal structure learning from multivariate time series in settings with unmeasured confounders*. KDD Workshop on Causal Discovery, 2018.

Basic Idea

Score-Based Approaches

$$\min_{\mathbf{A}\in\mathbb{R}^{d\times d}} S(\mathbf{A})$$
$$\text{subject to } G(\mathbf{A}) \in \text{DAGs}$$

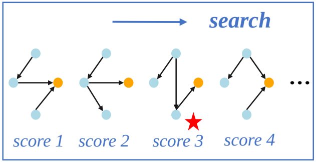
i) score function (e.g., BIC)

$$S(\mathbf{A}) = \boxed{-log[\hat{p}(D|\mathbf{A})]} + \boxed{\Delta(D,\mathbf{A})}$$

goodness of fit *penalize graph complexity*

ii) search strategy

X_1	X_2	X_3	X_4
0.1	0.5	1.2	0.9
0.2	0.3	1.5	1.1
0.4	0.7	1.1	0.8
…	…	…	…



Basic Idea

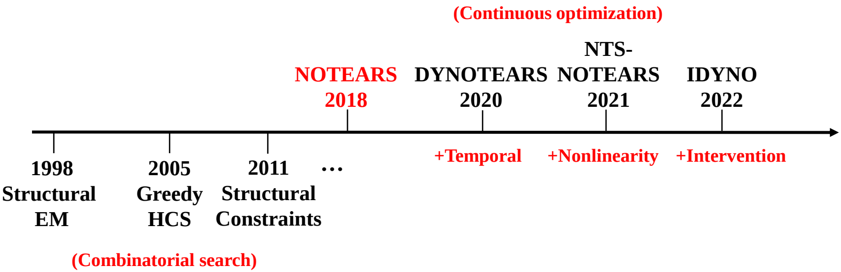
Score-Based Approaches

- **Too many DAGs[1], Combinatorial Search**

d	Number of DAGs with d nodes
1	1
2	3
3	25
4	543
5	29281
6	3781503
7	1138779265
8	783702329343
9	1213442454842881
10	4175098976430598143
11	31603459396418917607425
12	521939651343829405020504063
13	18676600744432035186664816926721
14	1439428141044398334941790719839535103

[1]Peters, Janzing, and Schölkopf. *Elements of causal inference: foundations and learning algorithms*. The MIT Press, 2017.

Overview



- Continuous Optimization Approaches
  - DYNOTEARS, NTS-NOTEARS, IDYNO
- Combinatorial Search Approaches
  - Greedy Hill-climb Search ...

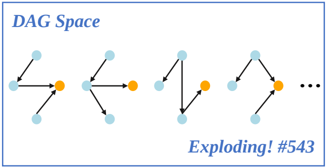
Continuous Optimization Approaches

Non-temporal Origins

NOTEARS[1]

- Non-temporal origins of the following methods

Combinatorial Search	***Continuous Optimization***
$\min_{\mathbf{A}\in\mathbb{R}^{d\times d}} S(\mathbf{A})$ subject to $G(\mathbf{A}) \in \mathrm{DAGs}$	$\min_{\mathbf{A}\in\mathbb{R}^{d\times d}} S(\mathbf{A})$ subject to $h(\mathbf{A}) = 0$



$$h(\mathbf{A}) = \mathrm{tr}(e^{\mathbf{A}\odot\mathbf{A}}) - d.$$

How can h(·) constrain DAGness?

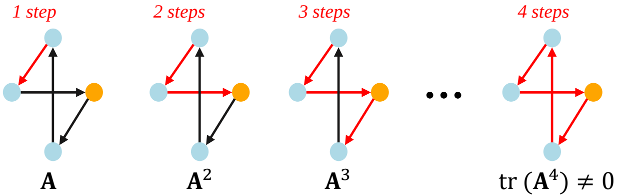
[1] Zheng, Aragam, Ravikumar, and Xing. *Dags with NO TEARS: continuous optimization for structure learning.* NeurIPS, 2018.

Continuous Optimization Approaches

Non-temporal Origins

NOTEARS

- Acyclicity constraint function



Acyclicity:

$$\mathrm{tr}\,(\mathbf{A}) + \mathrm{tr}\,(\mathbf{A}^2) + \mathrm{tr}\,(\mathbf{A}^3) + \cdots = 0$$

Approximation

Taylor expansion:

$$\mathrm{tr}\,(e^{\mathbf{A}}) - \mathrm{tr}\,(\mathbf{A}^0) = \frac{1}{1!}\mathrm{tr}\,(\mathbf{A}) + \frac{1}{2!}\mathrm{tr}\,(\mathbf{A}^2) + \frac{1}{3!}\mathrm{tr}\,(\mathbf{A}^3) + \cdots = 0$$

Non-negative:

$$\mathrm{tr}\,(e^{\mathbf{A}\odot\mathbf{A}}) - \mathrm{tr}\,((\mathbf{A}\odot\mathbf{A})^0) = \mathrm{tr}\,(e^{\mathbf{A}\odot\mathbf{A}}) - d$$

Continuous Optimization Approaches

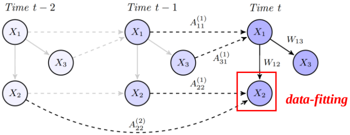
For Time Series

DYNOTEARS[1]

- NOTEARS' extension to time series data

i) Model specification:

$$\mathbf{x}_{m,t}^{\top} = \mathbf{x}_{m,t}^{\top}\mathbf{W} + \mathbf{x}_{m,t-1}^{\top}\mathbf{A}_1 + \ldots + \mathbf{x}_{m,t-p}^{\top}\mathbf{A}_p + \mathbf{z}_{m,t}^{\top}$$



ii) Training objective:

on contemporaneous edges, analogous to NOTEARS

$$\min_{\mathbf{W},\mathbf{A}} \; f(\mathbf{W},\mathbf{A}) \;\; \text{s.t.} \;\; \boxed{\mathbf{W} \text{ is acyclic}}$$

$$\text{with } f(\mathbf{W},\mathbf{A}) = \underbrace{\boxed{\ell(\mathbf{W},\mathbf{A})}}_{\textit{goodness of fit}} + \underbrace{\boxed{\lambda_{\mathbf{W}}\|\mathbf{W}\|_1 + \lambda_{\mathbf{A}}\|\mathbf{A}\|_1}}_{\textit{penalize graph complexity}}$$

$$\ell(\mathbf{W},\mathbf{A}) = \frac{1}{2n}\|\mathbf{X} - \mathbf{X}\mathbf{W} - \mathbf{Y}\mathbf{A}\|_F^2.$$

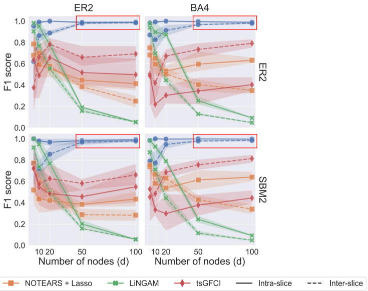
[1] Pamfil, Sriwattanaworachai, Desai, Pilgerstorfer, Georgatzis, Beaumont, and Aragam. *DYNOTEARS: structure learning from time-series data*. AISTATS, 2020.

Continuous Optimization Approaches

For Time Series

DYNOTEARS[1]

- Motivation: extend NOTEARS to MTS, model both contemporaneous and lagged effects
- Feature: hidden conf. $\times$, window G., contemporaneous $\surd$, nonlinear $\times$
- Finding: good performance under large graph size



[1] Pamfil, Sriwattanaworachai, Desai, Pilgerstorfer, Georgatzis, Beaumont, and Aragam. *DYNOTEARS: structure learning from time-series data*. AISTATS, 2020.

Continuous Optimization Approaches

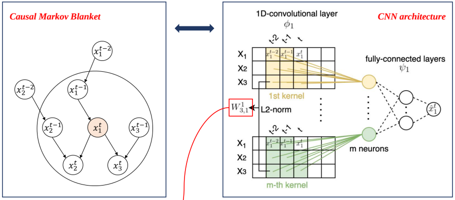
For Time Series

NTS-NOTEARS[1]

- 1D-CNN for capturing nonlinear relations

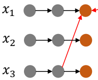
i) Temporal CNN model: bridge between causal meaning and NN architecture

$$\mathbb{E}[X_j^t | PA(X_j^t)] = CNN_j(\{\boldsymbol{X}^{t-k} : 1 \leq k \leq K\}, \boldsymbol{X}_{-j}^t)$$



ii) From local CNN to causal weight:

$$W_{ij}^k = ||\phi_{i,j}^k||_{L^2} \text{ for } k = 1, \ldots, K+1$$



iii) Training objective: $\min_\theta F(\theta)$

subject to $\boxed{h(W^{K+1}) = 0}$ ***on contemporaneous edges, analogous to NOTEARS***

$$F(\theta) = \frac{1}{T}\left(\sum_{j=1}^{d} \mathcal{L}(X_j^t, \boxed{CNN_{\theta_j}(\{\boldsymbol{X}^{t-k} : 1 \leq k \leq K\}, \boldsymbol{X}_{-j}^t)}\right)$$

model nonlinearity

$$+\boxed{\sum_{k=1}^{K+1} \lambda_1^k \cdot ||\phi_j^k||_{L^1} + \frac{1}{2}\lambda_2 \cdot ||\theta_j||_{L^2})}$$

penalize graph complexity

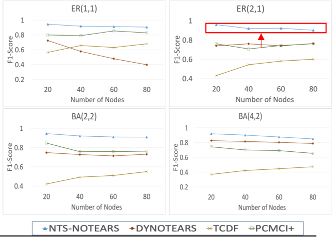
[1] Sun, Liu, Poupart, and Schulte. *NTS-NOTEARS: learning nonparametric temporal dags with time-series data and prior knowledge*. AISTATS, 2023 (ArXiv 2021).

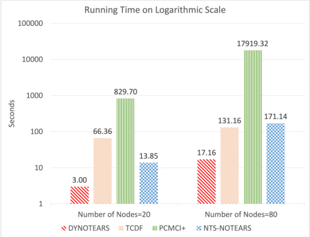
Continuous Optimization Approaches

For Time Series

NTS-NOTEARS[1]

- Motivation: $+$ model nonlinearity
- Idea: 1D CNN for relationship capturing
- Feature: hidden conf. $\times$, window G., contemporaneous $\surd$, nonlinear $\surd$
- Finding: good performance under nonlinearity, comparatively **faster** than Constraint-based methods





[1] Sun, Liu, Poupart, and Schulte. *NTS-NOTEARS: learning nonparametric temporal dags with time-series data and prior knowledge*. AISTATS, 2023 (ArXiv 2021).

Continuous Optimization Approaches

For Time Series

IDYNO[1]

- Motivation:
  - $+$ model nonlinearity
  - $+$ use both observational and **interventional** data
- Idea: objective calibrated for nonlinear and interventional condition
- Feature: hidden conf. $\times$, window G., contemporaneous $\surd$, nonlinear $\surd$

[1] Gao, Bhattacharjya, Nelson, Liu, and Yu. *IDYNO: learning nonparametric dags from interventional dynamic data*. ICML, 2022.

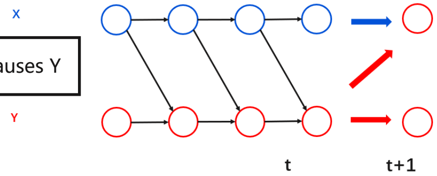
Basic Idea

Granger Causality in MTS

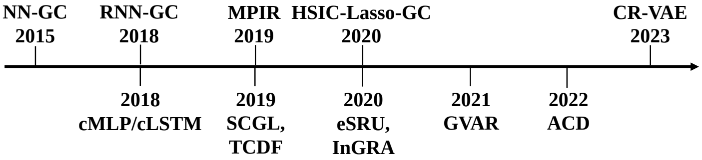
A time series X Granger-causes Y if past values of X provide unique, statistically significant information about future values of Y. Formally, if $X \to Y$,

$$\mathrm{var}[Y^t - \mathcal{P}(Y^t|\mathcal{H}^{<t})] < \mathrm{var}[Y^t - \mathcal{P}(Y^t|\mathcal{H}^{<t}\backslash X^{<t})].$$

Infer: X Granger-causes Y

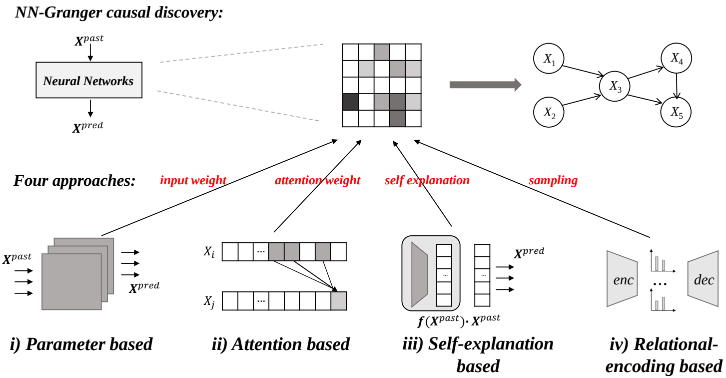


Overview



- Classic Methods
  - Lasso-GC, CGCI, ...
- NN-Based Methods
  - **Parameter based:** cMLP/cLSTM, eSRU, ...
  - **Attention based:** TCDF, InGRA, ...
  - **Self-explanation based:** GVAR
  - **Relational-encoding based:** ACD

NN-Based Methods



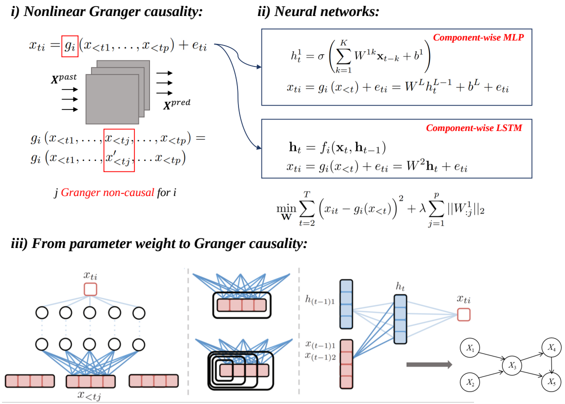
Gong, Yao, Zhang, Li and Bi. *Causal Discovery from Temporal Data: An Overview and New Perspectives*. Arxiv, 2023.

NN-Based Methods

Parameter based

cMLP/cLSTM

- component-wise MLP/LSTM for modeling nonlinearity



Tank, Covert, Foti, Shojaie, and Fox. *Neural Granger causality*. TPAMI, 2022. (ArXiv 2018)

NN-Based Methods

Parameter based

cMLP/cLSTM[1]

- Motivation: model nonlinear and complex relationships
- Idea:
  - NN for prediction
  - input matrix as causal weights
- Feature: hidden conf. $\times$, summary G., contemporaneous $\times$, nonlinear $\surd$

eSRU[2] LSTM $\rightarrow$ SRU

- Motivation: for better modeling MTS, data-scarcity & over-fitting
- Idea: eSRU (a special type of RNN designed for modeling MTS), fewer parameters
- Feature: hidden conf. $\times$, summary G., contemporaneous $\times$, nonlinear $\surd$

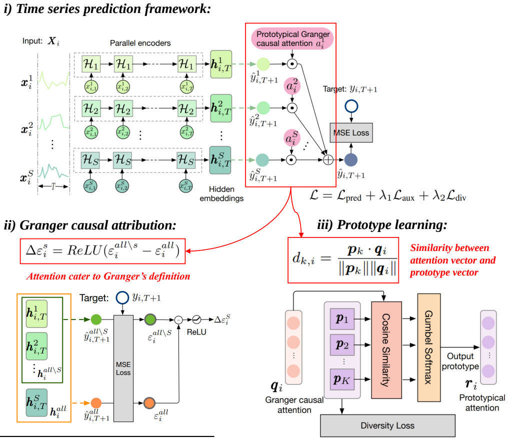
[1] Tank, Covert, Foti, Shojaie, and Fox. *Neural Granger causality*. TPAMI, 2022. (ArXiv 2018)

[2] Khanna and Tan. *Economy statistical recurrent units for inferring nonlinear Granger causality*. ICLR, 2020.

NN-Based Methods

Attention based

InGRA



Chu, Wang, Ma, Jia, Zhou, and Yang. *Inductive Granger causal modeling for multivariate time series*. ICDM, 2020.

NN-Based Methods

Attention based

InGRA

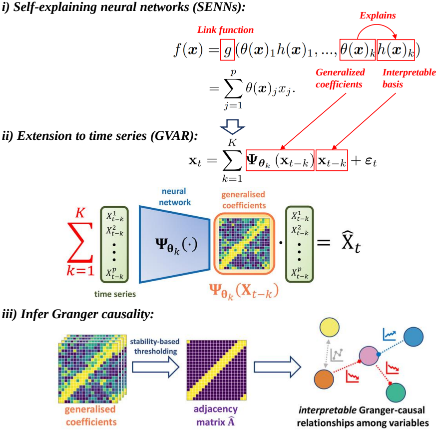
- Motivation:
  - the interpretability of **attention** mechanism
  - **heterogeneous** samples with common structures
- Idea:
  - Granger causal attention mechanism
  - prototype learning
- Feature: hidden conf. $\times$, summary G., contemporaneous $\times$, nonlinear $\surd$

Chu, Wang, Ma, Jia, Zhou, and Yang. *Inductive Granger causal modeling for multivariate time series*. ICDM, 2020.

NN-Based Methods

Self-explanation based

GVAR



Marcinkevics and Vogt. *Interpretable models for Granger causality using self-explaining neural networks*. ICLR, 2021.

NN-Based Methods

Self-explanation based

GVAR

- Motivation: more interpretable (detect signs of Granger causality)
- Idea: Self-explaining Networks: $f(\mathbf{x}) = g(\theta(\mathbf{x})_1 h(\mathbf{x})_1, \ldots, \theta(\mathbf{x})_k h(\mathbf{x})_k)$
- Feature: hidden conf. $\times$, summary G., contemporaneous $\times$, nonlinear $\surd$

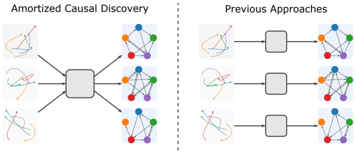
Marcinkevics and Vogt. *Interpretable models for Granger causality using self-explaining neural networks*. ICLR, 2021.

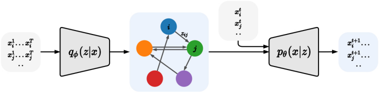
NN-Based Methods

Relational-encoding based

ACD (Amortized Causal Discovery)

- Motivation: **heterogeneous** samples sharing relevant information
- Idea: GNN + VAE
- Feature: hidden conf. $\surd$, summary G., contemporaneous $\times$, nonlinear $\surd$, **heterogeneous**



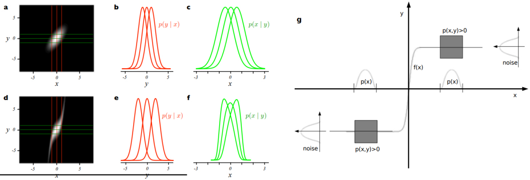


Löwe, Madras, Shilling, and Welling. *Amortized causal discovery: Learning to infer causal graphs from time-series data*. CLeaR, 2022.

Basic Idea

Structural Causal Model + Additional Assumptions → Causal Identification

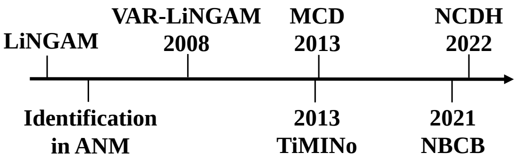
- **Linear Non-Gaussian Acyclic Model (LiNGAM)** [1]
  - Assume **linear** data generation, non-Gaussian disturbances
  - linear Independent Component Analysis (ICA): $\mathbf{x} = \mathbf{A} \cdot \mathbf{e}$
- **Nonlinear Additive Noise Model (ANM)** [2]
  - Identifiability in almost any nonlinear model with additive noise
  - Test $y = f(x) + n \rightarrow$ independence test between 'cause' variable x and residual: $\hat{n} = y - \tilde{f}(x)$



[1] Shimizu, Hoyer, Hyvärinen, Kerminen, and Jordan. *A linear non-Gaussian acyclic model for causal discovery*. JMLR, 2006.

[2] Hoyer, Janzing, Mooij, Peters, and Schölkopf. *Nonlinear causal discovery with additive noise models*. NeurIPS, 2008.

Overview



- LiNGAM's extension
  - VAR-LiNGAM, MCD, NCDH
- ANM's extension
  - TiMINo, NBCB

LiNGAM

Linear Non-Gaussian Acyclic Model (LiNGAM)

i) Model assumption:

$$x_i = \sum_{k(j)<k(i)} b_{ij}\, x_j + \boxed{e_i}$$

Non-Gaussian

ii) Transformation:

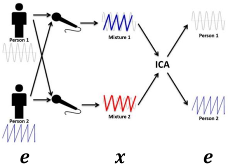
$$\boldsymbol{x} = \boldsymbol{B}\boldsymbol{x} + \boldsymbol{e}$$

$$\boxed{\boldsymbol{x} = (\boldsymbol{I} - \boldsymbol{B})^{-1}\boldsymbol{e} = \boldsymbol{A}\boldsymbol{e}}$$

X_1	X_2	X_3	X_4	X_5
0.1	0.5	1.2	0.9	0.9
0.2	0.3	1.5	1.1	1.0
…	…	…	…	…

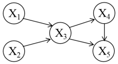
iii) Independent Component Analysis:

$$\boldsymbol{x} = \boldsymbol{A} \cdot \boldsymbol{e}$$



iv) Infer Causal Graph:

$$\boldsymbol{A} \rightarrow \boldsymbol{B}$$



Shimizu, Hoyer, Hyvärinen, Kerminen, and Jordan. *A linear non-Gaussian acyclic model for causal discovery*. JMLR, 2006.

Lubo-Robles. *Development of independent component analysis for reservoir geomorphology and unsupervised seismic facies classification in the Taranaki Basin, New Zealand*. 2018.

LiNGAM's Extension

For Time Series

VAR-LiNGAM[1,2]

i) Model assumption:

$$\boldsymbol{x}(t) = \sum_{\tau=0}^{k} \boldsymbol{B}_\tau \boldsymbol{x}(t-\tau) + \boldsymbol{e}(t)$$

ii) Transformation:

$$\boldsymbol{n}(t) = \boldsymbol{x}(t) - \sum_{\tau=1}^{k} (\boldsymbol{I} - \boldsymbol{B}_0)^{-1} \boldsymbol{B}_\tau \boldsymbol{x}(t-\tau)$$

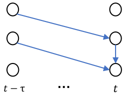
Residual as causal variable

$$\boldsymbol{n}(t) = \boldsymbol{B}_0 \boldsymbol{n}(t) + \boldsymbol{e}(t)$$

$$\boxed{\boldsymbol{n}(t) = (\boldsymbol{I} - \boldsymbol{B}_0)^{-1} \boldsymbol{e}(t)}$$

v) Infer Causal Graph:

$\boldsymbol{M}_\tau, \boldsymbol{B}_0 \rightarrow \boldsymbol{B}_\tau$ $\quad \boldsymbol{B}_0$

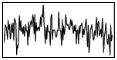


iv) Independent Component Analysis:

*Analogous to **LiNGAM*** $\quad \boldsymbol{B}_0$

iii) VAR estimation:

$\boldsymbol{M}_\tau = (\boldsymbol{I} - \boldsymbol{B}_0)^{-1} \boldsymbol{B}_\tau$



$\boldsymbol{n}(t)$

[1] Hyvärinen, Shimizu, and Hoyer. *Causal modelling combining instantaneous and lagged effects: an identifiable model based on non-Gaussianity*. ICML, 2008.

[2] Hyvärinen, Zhang, Shimizu, and Hoyer. *Estimation of a structural vector autoregression model using non-gaussianity*. JMLR, 2010.

LiNGAM's Extension

VAR-LiNGAM

- Motivation: extend to MTS, model lagged and contemporaneous effects
- Feature: hidden conf. $\times$, window G., contemporaneous $\surd$, nonlinear $\times$

NCDH[1]

- Motivation: model nonlinear and nonstationary relations
- Idea:
  - nonlinear ICA
  - HMM for nonstationary scenarios
- Feature: hidden conf. $\times$, summary G., contemporaneous $\surd$, nonlinear $\surd$

[1] Wu, Wu, Wang, Liu, and Chen. *Nonlinear causal discovery in time series.* CIKM, 2022.

ANM's Extension

ANM for modeling nonlinear causal relations

TiMINo[1]

- Motivation: nonlinear
- Idea: SCM for MTS, ANM
- Feature: hidden conf. $\times$, summary G., contemporaneous $\surd$, nonlinear $\surd$

NBCB[2]

- Motivation: high dimensional (computational complexity)
- Idea: Extension of TiMINo, leverage CB to prune unnecessary edges
- Feature: hidden conf. $\times$, summary G., contemporaneous $\surd$, nonlinear $\surd$

[1] Peters, Janzing, and Schölkopf. *Causal inference on time series using restricted structural equation models.* NeurIPS, 2013.

[2] Assaad, Devijver, Gaussier, and Aït-Bachir. *A mixed noise and constraint-based approach to causal inference in time series.* ECML-PKDD, 2021.

1. Constraint-Based Approaches
2. Score-Based Approaches
3. Granger Causality Based Approaches
4. Model-Based Approaches
5. Summary

Characteristics of reviewed methods

Section	Method	Causal Graph	Nonlinear	Instantaneous effects	Hidden confounders	Sufficiency Asm.	Markov Asm.	Faithfulness Asm.	Minimality Asm.
Constraint-based	oCSE (2015)	Summary	Yes	No	No	Yes	Yes	Yes	
	PCGCE (2022)	Extended	Yes	Yes	No	Yes	Yes	Yes	
	PCMCI (2019)	Window	Yes	No	No	Yes	Yes	Yes	
	PCMCI$^+$ (2020)	Window	Yes	Yes	No	Yes	Yes	Yes	
	ANLTSM (2008)	Window	Yes	Yes	Yes	No	Yes	Yes	
	tsFCI (2010)	Window	Yes	No	Yes	No	Yes	Yes	
	SVAR-FCI (2018)	Window	No	Yes	Yes	No	Yes	Yes	
	FCIGCE (2022)	Extended	Yes	Yes	Yes	No	Yes	Yes	
	LPCMCI (2020)	Window	Yes	Yes	Yes	No	Yes	Yes	
Score-based	DYNOTEARS (2020)	Window	No	Yes	No	Yes	Yes	No	No
	NTS-NOTEARS (2021)	Window	Yes	Yes	No	Yes	Yes	No	No
	IDYNO (2022)	Window	Yes	Yes	No	Yes	Yes	No	No
FCM-Based	VAR-LiNGAM (2008)	Window	No	Yes	No	Yes	Yes	No	Yes
	NCDH (2022)	Summary	Yes	No	No	Yes	Yes	No	Yes
	TiMINo (2013)	Summary	Yes	Yes	No	Yes	Yes	No	Yes
	NBCB (2021)	Summary	Yes	Yes	No	Yes	Yes	Yes	Yes
Granger Causality	HSIC-Lasso-GC (2020)	Summary	Yes	No	No	No	No	No	No
	(R)NN-GC (2015,2018)	Summary	Yes	Yes	No	No	No	No	No
	MPIR (2019)	Summary	Yes	No	No	No	No	No	No
	NGC (2022)	Summary	Yes	No	No	No	No	No	No
	eSRU (2020)	Summary	Yes	No	No	No	No	No	No
	SCGL (2019)	Summary	Yes	No	No	No	No	No	No
	GVAR (2021)	Summary	Yes	Yes	No	No	No	No	No
	TCDF (2019)	Window	Yes	Yes	Yes	No	No	No	No
	CR-VAE (2023)	Summary	Yes	Yes	No	No	No	No	No
	InGRA (2020)	Summary	Yes	No	No	No	No	No	No
	ACD (2022)	Summary	Yes	No	Yes	No	No	No	No
Others	DBCL (2010)	Summary	Yes	Yes	Yes	No	Yes	Yes	
	NGM (2022)	Summary	Yes	Yes	No	No	No	No	No
	CCM (2012)	Summary	Yes	No	No	No	No	No	No
	PCTL(c) (2009,2011)	Summary	Yes	No	No	No	No	No	No

Gong, Yao, Zhang, Li and Bi. *Causal Discovery from Temporal Data: An Overview and New Perspectives*. Arxiv, 2023.

Comparison of Causal Discovery Methods for MTS

Methods	Idea	Causal Graph	Complexity	# of Variables	Data Volume
Constrain-based	Independent Test	Window	High	5	Low
Score-based	Optimization	Window	Middle	100	High
Model-based	Assumption	Window	High	5	Low
Granger-based	Prediction	Summary	Low	10	High

How to select?

- Window causal graph v.s. Summary causal graph
  - Constrain-based, Score-based v.s. Granger-based
- Big data v.s. Small data
  - Score-based, Granger-based v.s. Constrain-based, Model-based
- Large scale variables v.s. few variables
  - Score-based, Granger-based v.s. Constrain-based, Model-based

Thanks!