Causal Discovery from Temporal Data

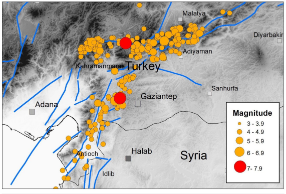
Event Sequence

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

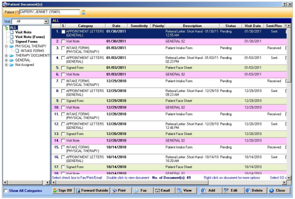
Event Sequence Example



Geometry: earthquake records …

Crime: crime activity records …

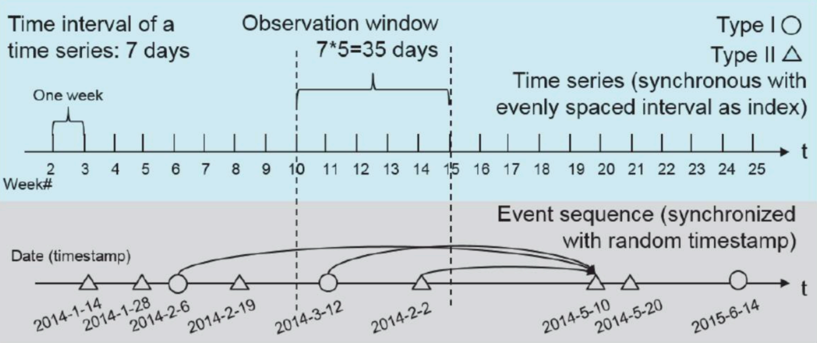
Finance: trading records …



Medicine: Electronic health records …

Event Sequence: $\{(t_1, e_1), (t_2, e_2), \dots\}$ times & types

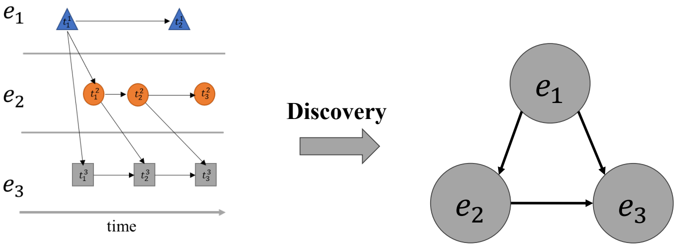
Comparison Between Time Series Data and Event Sequence Data



Event Sequences:
Record the occurrence $\rightarrow$ random timestamp, random interval

Problem Definition

Aim: find the causal relationships between different event types in event sequences

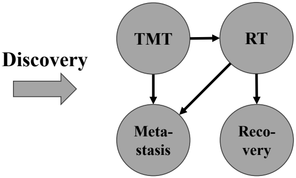


$$\{(t_1^1, e_1), (t_1^3, e_3), (t_1^2, e_2), \dots\}$$

e_1: event type 1, e_2: event type 2, e_3: event type 3

Example 1

2021/11/02 07:01:35 Tumor Marker Test (TMT)
2021/11/10 10:24:42 Radiation Therapy (RT)
2022/09/21 15:40:32 Metastasis
2022/10/19 10:24:42 Radiation Therapy (RT)
2023/06/20 12:09:56 Recovery
……

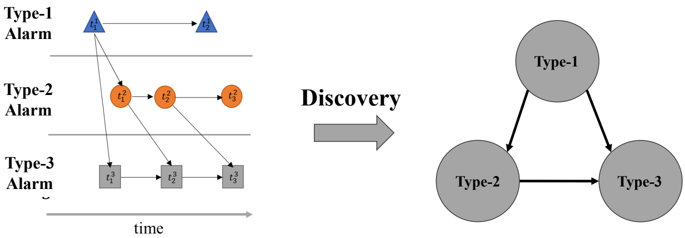


1

[1] Zhuochen Jin et al. "Visual causality analysis of event sequence data". In: *IEEE transactions on visualization and computer graphics* 27.2 (2020), pp. 1343–1352.

Example 2

a cellular network alarm sequence



2022/11/2 07:01:35   Type-1 Alarm
2022/11/2 10:24:42   Type-2 Alarm
2022/11/3 15:40:32   Type-3 Alarm
……

Event Sequences and Multivariate Point Processes

Multivariate Point Processes (MPP)

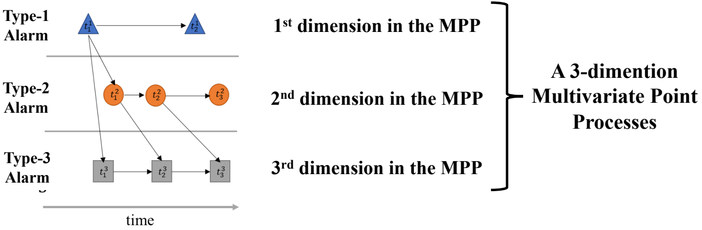
$$\{(t_1, e_1), (t_2, e_2), \ldots\}$$

- Use Multivariate Point Processes to characterize Event sequences

Multivariate Point Processes

An example

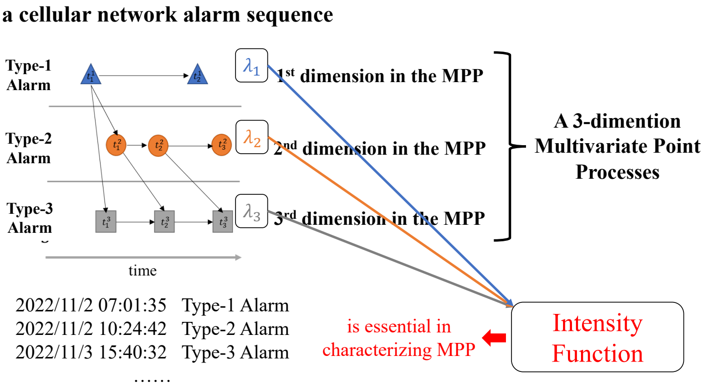
a cellular network alarm sequence



2022/11/2 07:01:35 Type-1 Alarm
2022/11/2 10:24:42 Type-2 Alarm
2022/11/3 15:40:32 Type-3 Alarm
……

Multivariate Point Processes

Intensity Functions



- $\lambda_e(t)$ denotes the expected instantaneous rate of type-e event's occurrence at time t given the history

Multivariate Point Processes

Intensity Functions

Intensity Functions of Multivariate Point Processes:

Occurrence Count

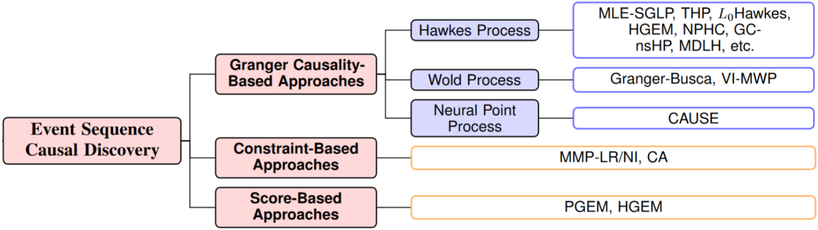
Historical Information

$$\lambda_{\boxed{e}}(t) = \frac{\mathbb{E}[d\boxed{N_e(t)}|\boxed{\mathcal{H}_{last}}]}{dt}, \text{ where } \mathcal{H}_{last} = \{(t_i, e_i)|t_i < t, e_i \in \mathcal{E}\}$$

Event type

- $\lambda_e(t)$ denotes the expected instantaneous rate of type-e event's occurrence at time t given the history

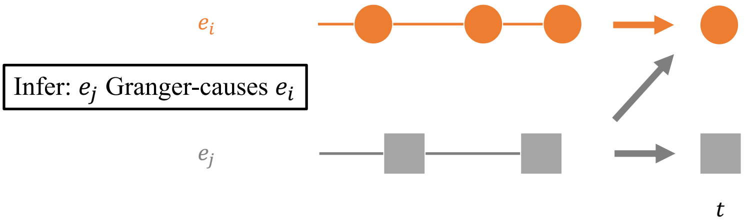
Framework


Event Sequence Causal Discovery
Granger Causality-Based Approaches
Constraint-Based Approaches
Score-Based Approaches
Hawkes Process
Wold Process
Neural Point Process
MLE-SGLP, THP, L_0Hawkes, HGEM, NPHC, GC-nsHP, MDLH, etc.
Granger-Busca, VI-MWP
CAUSE
MMP-LR/NI, CA
PGEM, HGEM

Granger Causality in Event Sequence

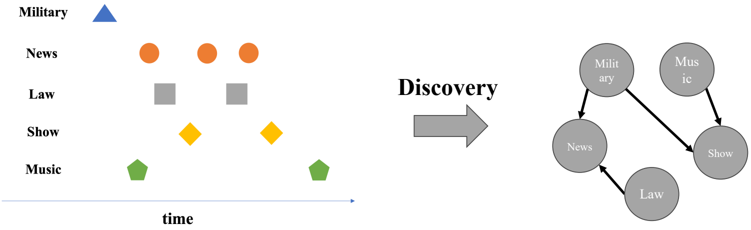
Granger Causality in Event Sequence

e_j-type events Granger cause e_i if $\{e_j(t)|t < t_0\}$ is useful in forecasting $e_i(t)$.



- A weaker, prediction-based causality

An Example for Discovering Granger Causality



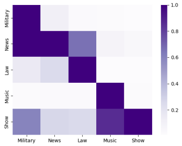
Discover the Granger Causality between different TV categories (in this example, Military → News, Law → News, Military → Show, Music → Show)[1]

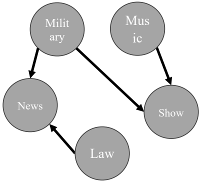
[1] Dixin Luo et al. "You are what you watch and when you watch: Inferring household structures from IPTV viewing data". In: *IEEE Transactions on Broadcasting* 60.1 (2014), pp. 61–72.

An Example for Discovering Granger Causality

Input:
2012/01/01 13:08:30, Law
2012/01/01 13:20:22, News
2012/01/01 14:01:56, Show
......
Output: Granger Causality Heat Map and Graph:

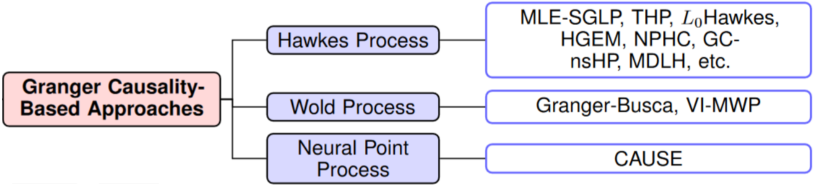
$$\begin{pmatrix} 1 & 0.12 & 0.03 & 0.02 & 0.01 \\ 1 & 1 & 0.66 & 0.08 & 0.05 \\ 0.16 & 0.25 & 1 & 0.01 & 0.01 \\ 0.02 & 0.03 & 0.03 & 1 & 0.02 \\ 0.6 & 0.27 & 0.26 & 0.86 & 1 \end{pmatrix}$$





(Granger Causality from B to A: $(rowA, columnB)$ element in the heat map)

Granger Causality - Based Approaches Framework



Methods for Hawkes Processes

- Show great performance in modeling event sequences
- Show a close relationship to Granger causality

Multivariate Hawkes Processes

The Multivariate Hawkes Processes is a type of Multivariate point processes. Each dimension of it has a fixed form of intensity function:

$$\lambda_{e_i}(t) = \mu_{e_i} + \sum_{e_j=1}^{E} \int_0^t \phi_{e_i e_j}(s) dN_{e_j}(t-s)$$

Self excitatory, decaying

Used to model earthquakes, financial transactions, crimes

Methods for Hawkes Processes

the Intensity Function

Multivariate Hawkes Processes

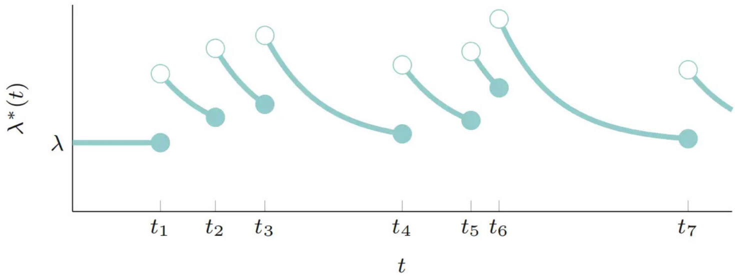
The Multivariate Hawkes Processes is a type of Multivariate point processes. Each dimension of it has a fixed form of intensity function:

$$\lambda_{e_i}(t) = \underbrace{\mu_{e_i}}_{\text{baseline intensity}} + \sum_{e_j=1}^{E} \int_0^t \underbrace{\phi_{e_i e_j}(s)}_{\text{impact function: our focus}} dN_{e_j}(t-s)$$

- μ_{e_i} - the baseline intensity - can only be affected by exogenous events, hence, is a constant over time
- $\phi_{e_i e_j}(s)$ - the impact function - measures the decay of the excitement on future type-e_i events triggered by historical type-e_j events

Methods for Hawkes Processes

the Intensity Function



Methods for Hawkes Processes

Granger Causality in Hawkes Processes

Multivariate Hawkes Processes

$$\lambda_{e_i}(t) = \mu_{e_i} + \sum_{e_j=1}^{E} \int_0^t \phi_{e_i e_j}(s) dN_{e_j}(t-s)$$

Granger Causality in Hawkes Processes

$$e_j \text{ does not Granger-cause } e_i \iff \phi_{e_i e_j}(s) = 0, \forall s \in R$$

$$\downarrow$$

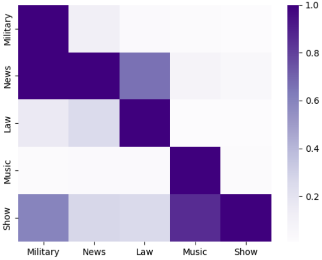
Our goal is to model $\phi_{e_i e_j}(t)$ for each event and all $t \in R$

Methods for Hawkes Processes

Inferring Granger Causality Matrix

The Granger Causality matrix G is built by

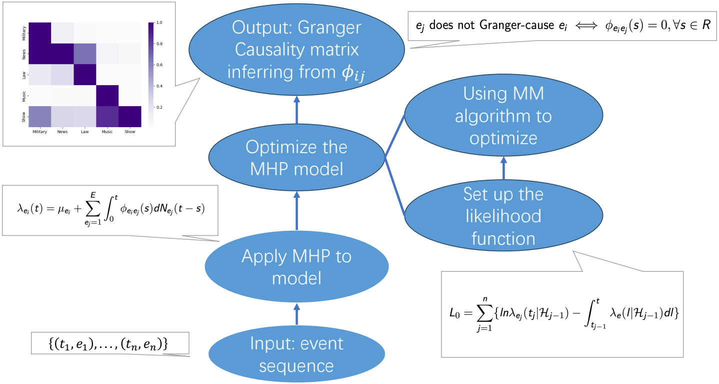
$$(g_{ij}) = \int_0^{\infty} \phi_{ij}(s)ds$$



(Granger Causality from B to A: $(rowA, columnB)$ element in the heat map)

Methods for Hawkes Processes

The Flowchart of the Vanilla Hawkes Processes Methods (MLE-SGL)



1

[1] Hongteng Xu, Mehrdad Farajtabar, and Hongyuan Zha. "Learning granger causality for hawkes processes". In: *International conference on machine learning*. PMLR. 2016, pp. 1717–1726.

Methods for Hawkes Processes

Optimize the MHP Model

$$\lambda_{e_i}(t) = \boxed{\mu_{e_i}} + \sum_{e_j=1}^{E} \int_0^t \boxed{\phi_{e_i e_j}(s)} dN_{e_j}(t-s)$$

Parameter set: $\{\mu_{e_i}, \phi_{e_i e_j}(s) | e_i, e_j = 1, \dots, E\}$

Optimize the parameter s.t. $\{\lambda_{e_i}\}$ become more consistent with the ground truth intensity of the input event sequence

Optimization method: e.g., Maximum Likelihood Estimation (MLE)

Learn from the event sequence data to optimize the MHP model

Methods for Hawkes Processes

Maximum Likelihood Estimation (MLE) in MHP Model

The log-likelihood function: $\widetilde{L_0} \triangleq \ln \boxed{f((t_1, e_1), ..., (t_n, e_n)|(t_0, e_0))}$ Probability Density Function of the MPP

by the chain rule $\boxed{=} \ln(\prod_{j=1}^{n} f((t_j, e_j)|\mathcal{H}_{lastj}))$

$$= \sum_{j=1}^{n} \ln f(t_j|e_j, \mathcal{H}_{lastj}) + \sum_{j=1}^{n} \ln f(e_j|\mathcal{H}_{lastj})$$

$$\lambda_e(t|\mathcal{H}_{n-1}) = \frac{\mathbb{E}[N_e(t+dt) - N_e(t)|\mathcal{H}_{n-1}]}{dt} = -\frac{d}{dt}\ln(1 - \int_{t_{n-1}}^{t} f(l|d, \mathcal{H}_{n-1})dl)$$

by definition

$$L_0 = \sum_{j=1}^{n} \ln f(t_j|e_j, \mathcal{H}_{lastj})$$

$$= \boxed{\sum_{j=1}^{n}\{\ln \lambda_{e_j}(t_j|\mathcal{H}_{j-1}) - \int_{t_{j-1}}^{t} \lambda_e(l|\mathcal{H}_{j-1})dl\}}$$

important!!

Objective: $\min_{\mu_{e_i}, \phi_{e_i e_j}(s)} -L_0$

Methods for Hawkes Processes

Parameterize The Impact Function

$$\boxed{\text{MHP:} \quad \lambda_{e_i}(t) = \mu_{e_i} + \sum_{e_j=1}^{E} \int_0^t \phi_{e_i e_j}(s)\, dN_{e_j}(t-s)}$$

To simplify the model:

Parameterize $\phi_{e_i e_j}(s)$: $\quad \phi_{e_i e_j}(s) = \underbrace{a_{e_i e_j}}_{\text{Event Interaction part}} \underbrace{g(s)}_{\text{Time-decaying part}}$

Let $A = \left(a_{e_i e_j}\right), \mu = (\mu_{e_i})$

$$L_0 \xrightarrow{\phi_{e_i e_j}(s) = a_{e_i e_j} g(s)} L(A, \mu)$$

Objective: $\quad \min\limits_{A \geq 0, \mu \geq 0} -L(A, \mu) + \underbrace{\boxed{\|A\|}}_{\text{Regularization term}}$

Methods for Hawkes Processes

How to Optimize the Objective Function?

Depending on the regularization term, this optimizing problem could be non-convex

- Majorization-Minimization (MM) or other non-convex optimizing algorithm
- Simulation method

Methods for Hawkes Processes

Shortcomings of the Vanilla MLE-Hawkes Model

- high computational complexity
- bad modeling capacity

Methods for Hawkes Processes

Several Improvements to the Vanilla MLE-Hawkes Model

To improve modeling capacity:

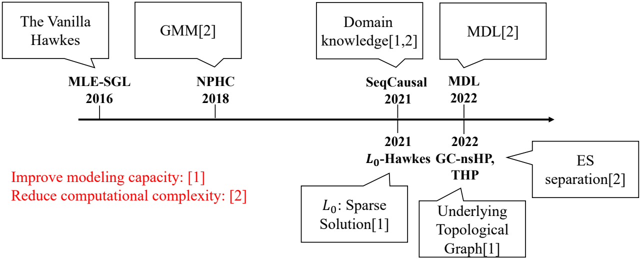
- For intensity functions: Parameterization strategies
- For likelihood functions: Regularization methods

To reduce computational complexity:

- Generalized method of moments
- Event sequence separation

Methods for Hawkes Processes

Hawkes Granger Causality Timeline



Methods for Hawkes Processes

Hawkes Granger Models

MLE-SGL[1] (The Vanilla Hawkes)

- Optimization approach: MLE+MM
- Parameterization: $\phi_{e_i e_j}(s) = \sum_{m=1}^{M} a_{e_i e_j}^{m} \kappa_m(s)$ - a linear combination of basic functions $\kappa_m(s)$
- Regularization: group-lasso, L_1, pairwise similarity: $\sum_{e_i=1}^{E} \sum_{e_j \in C_e} \|a_{e_i \cdot} - a_{e_j \cdot}\|_F^2 + \|a_{\cdot e_i} - a_{\cdot e_j}\|_F^2$

[1] Hongteng Xu, Mehrdad Farajtabar, and Hongyuan Zha. "Learning granger causality for hawkes processes". In: *International conference on machine learning*. PMLR. 2016, pp. 1717–1726.

Methods for Hawkes Processes

Hawkes Granger Models

NPHC[1]

- Reduce computational complexity
- Optimization approach:
  Infer from $G_0 = (g_{ee'}) = (\int_0^{+\infty} \phi_{ee'}(s)ds)$+GMM
- The output Granger Causality matrix is the optimized r.v. $\hat{G}_0$: $G = \hat{G}_0$
- GMM (Generalized Method of Moment): model G_0 using its specific moments (in this case, centered second and third-order moments)

[1] Massil Achab et al. "Uncovering causality from multivariate Hawkes integrated cumulants". In: *International Conference on Machine Learning*. PMLR. 2017, pp. 1–10.

Methods for Hawkes Processes

Hawkes Granger Models

SeqCasual[1]

- Improve modeling capacity, Reduce computational complexity
- Optimization approach: MLE+MM
- Parameterization: $\phi_{e_i e_j}(s) = \sum_{m=1}^{M} a_{e_i e_j}^{m} \kappa_m(s)$
- Regularization: Manual Updated:

$$\underset{\mu,\alpha}{argmin} \quad -L + \alpha_v \sum_{e_i, e_j} \|a_{e_i e_j}(\hat{G})\|_F$$

$$\text{s.t.} \quad a_{e_i e_j} = 0 \quad \text{for} \quad (e_j \to e_i) \notin \hat{G}$$

$$a_{e_i e_j}(\hat{G}) = \begin{cases} 0; & \text{if} \quad (e_j \to e_i) \quad \text{is confirmed} \\ a_{e_i e_j}; & \text{otherwise} \end{cases}$$

[1] Zhuochen Jin et al. "Visual causality analysis of event sequence data". In: *IEEE transactions on visualization and computer graphics* 27.2 (2020), pp. 1343–1352.

Methods for Hawkes Processes

Hawkes Granger Models

L_0-**Hawkes**[1]

- Improve modeling capacity
- Optimization approach: MLE+MM
- Parameterization:

$$\lambda_{e_i}(t) = \mu_{e_i} + \sum_{e_j=1}^{E} A_{d_i d_j} \phi_{e_i}(t - t_j)$$

- Regularization: L_0 - the number of matrix's non-zero elements

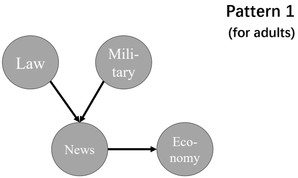
[1] Tsuyoshi Idé et al. "Cardinality-regularized hawkes-granger model". In: *Advances in Neural Information Processing Systems* 34 (2021), pp. 2682–2694.

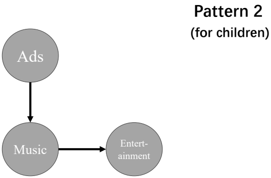
Methods for Hawkes Processes

Hawkes Granger Models

GC-nsHP[1]

- Reduce computational complexity
- Optimization and Parameterization approach is the same as MLE-SGL's
- Event sequence separation:

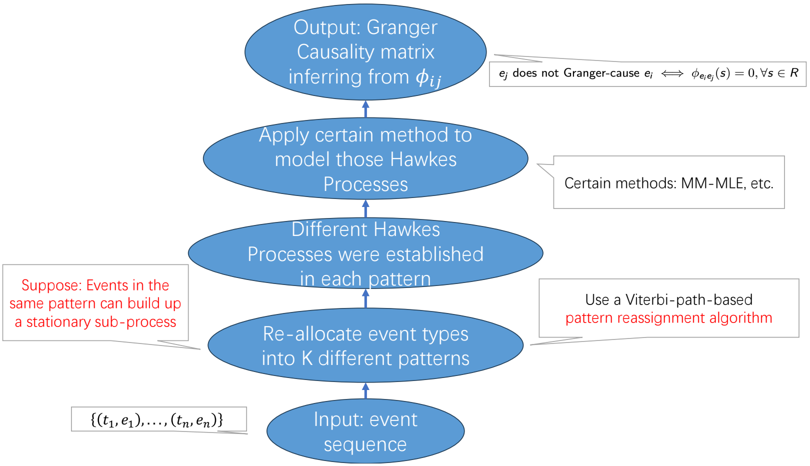




[1] Wei Chen et al. "Learning granger causality for non-stationary Hawkes processes". In: *Neurocomputing* 468 (2022), pp. 22–32.

Methods for Hawkes Processes

Framework of GC-nsHP

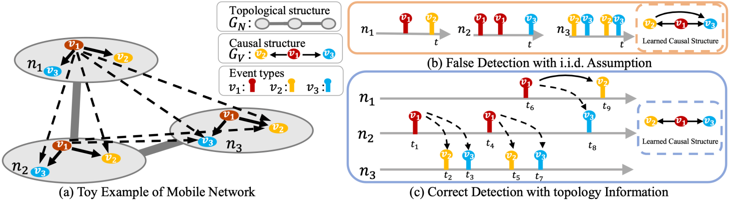


Methods for Hawkes Processes

Hawkes Granger Models

THP[1]

- Improve modeling capacity
- Optimization approach: MLE+MM
- Parameterization: $\lambda_{e_i}(n,t) = \mu_{e_i} + \sum_{e_j \in E} (g_{e_i e_j} * s_{e_i,e_j,t})_{G_N}(n)$
- Regularization: L_0
- Consider the underlying topological relationships:



[1] Ruichu Cai et al. "THPs: Topological hawkes processes for learning causal structure on event sequences". In: *IEEE Transactions on Neural Networks and Learning Systems* (2022).

Methods for Wold Processes

Basic Ideas

Advantage: Less complexity:
(computation complexity per iteration)
Vanilla-Hawkes: $O(MK^2N^3)$
GMM-Hawkes: $O(K^3)$
GRANGER-BUSCA[1] (Wold Processes): $O(N(lnN + lnK))$
[N: the number of observations in all processes, K: the number of processes, M: the number of basis functions]

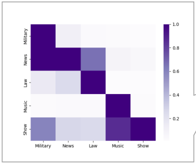
Suppose: $\delta_i = t_i - t_{i-1}$ only relate to δ_{i-1}:
$\{\delta_i, i \in \mathrm{N}\}$ forms a Markov chain

The inherent Markov property within the Wold processes makes them suitable for modeling the dynamics of certain complex systems

[1] Flavio Figueiredo et al. "Fast estimation of causal interactions using wold processes". In: *Advances in Neural Information Processing Systems* 31 (2018).

Methods for Wold Processes

Framework



Output: Granger Causality matrix inferring from $\alpha_{e_i e_j}$

e_j does not Granger-cause e_i $\iff$ $\alpha_{e_i e_j} = 0$

Optimize the MWP model

$$\lambda_{e_i}(t) = \mu_{e_i} + \sum_{e_j \in E} \frac{\alpha_{e_i e_j}}{\beta_{e_j} + \Delta_{e_i^p e_j^q}(t)}$$

Apply MWP to model

Input: event sequence

$\{(t_1, e_1), \ldots, (t_n, e_n)\}$

Methods for Wold Processes

Intensity Functions of Wold Processes

GRANGER-BUSCA[1] (Multivariate Wold Processes)

$$\lambda_{e_i}(t) = \underbrace{\mu_{e_i}}_{\text{baseline intensity}} + \sum_{e_j \in E} \frac{\alpha_{e_i e_j}}{\underbrace{\beta_{e_j}}_{\text{flat value}} + \underbrace{\Delta_{e_i^p e_j^q}(t)}_{\text{time gap}}}$$

Here, $\Delta_{e_i^p e_j^q}(t) = t_{e_i^p} - t_{e_j^q}$, $t_{e_i^p} : p = argmax\{p | t_{e_i^p}\}_{<t}$, $t_{e_j^q} : q = argmax\{q | t_{e_j^q} < t_{e_i^p}\}$.

Intuition: larger $\Delta_{e_i^p e_j^q}(t) \rightarrow$ weaker evidence for the influence of e_j on e_i

Recall Multivariate Hawkes Processes:

$$\lambda_{e_i}(t) = \mu_{e_i} + \sum_{e_j=1}^{E} \int_0^t \phi_{e_i e_j}(s) dN_{e_j}(t-s)$$

[1] Flavio Figueiredo et al. "Fast estimation of causal interactions using wold processes". In: *Advances in Neural Information Processing Systems* 31 (2018).

Methods for Wold Processes

Granger Causality in Wold Processes

GRANGER-BUSCA (Multivariate Wold Processes)

$$\lambda_{e_i}(t) = \underbrace{\mu_{e_i}}_{\text{baseline intensity}} + \sum_{e_j \in E} \frac{\alpha_{e_i e_j}}{\underbrace{\beta_{e_j}}_{\text{flat value}} + \underbrace{\Delta_{e_i^p e_j^q}(t)}_{\text{time gap}}}$$

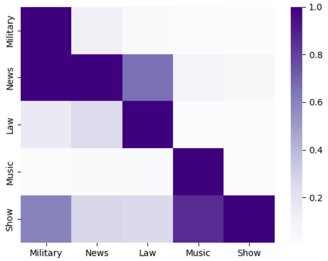
Granger Causality in Wold Processes

$$e_j \text{ does not Granger-cause } e_i \iff \alpha_{e_i e_j} = 0$$

Methods for Wold Processes

Inferring Granger Causality Matrix

The Granger Causality matrix G is built by $(g_{ij}) = \alpha_{ij}$



(Granger Causality from B to A: $(rowA, columnB)$ element in the heat map)

Methods for Neural Point Processes

Main Idea and Pros & Cons

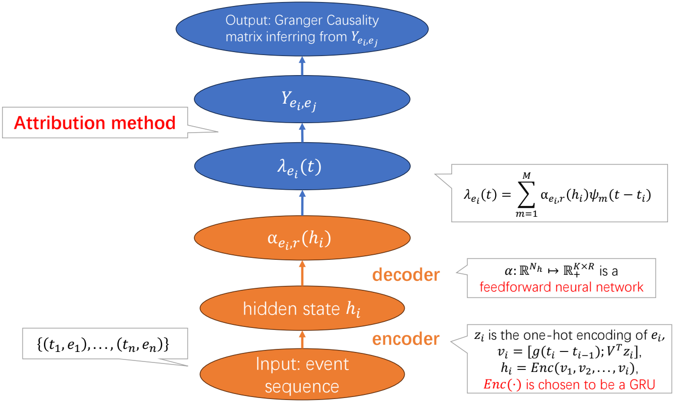
Main idea:
use neural networks to infer the intensity function $\lambda_{e_i}(t)$.

Pros: introducing more freedom in event sequences modeling $\rightarrow$ better modeling capacity

Cons: No direct relationship with Granger Causality, need extra attribution method

Methods for Neural Point Processes

Framework: Network and Parameterization Settings



1

[1] Wei Zhang et al. "Cause: Learning granger causality from event sequences using attribution methods". In: *International Conference on Machine Learning*. PMLR. 2020, pp. 11235–11245.

Methods for Neural Point Processes

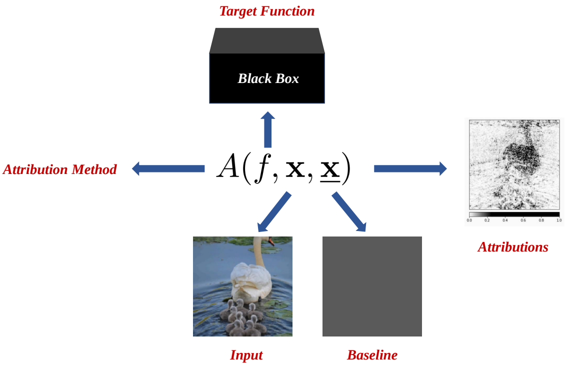
Parameterization Settings

Comparison:

- Multivariate Hawkes Processes:
  $\lambda_{e_i}(t) = \mu_{e_i} + \sum_{e_j=1}^{E} \int_0^t \phi_{e_i e_j}(s) dN_{e_j}(t-s)$
- GRANGER-BUSCA (Multivariate Wold Processes):
  $\lambda_{e_i}(t) = \mu_{e_i} + \sum_{e_j \in E} \frac{\alpha_{e_i e_j}}{\beta_{e_j} + \Delta_{e_i^p e_j^q}(t)}$
- CAUSE (Neural Point Processes):
  $\lambda_{e_i}(t) = \sum_{m=1}^{M} a_{e_i,m}(h_i)\kappa_m(t-t_i)$, in which, $\kappa_m(t)$ can be pre-selected, decoder outputs the $a_{e_i,m}(h_i)$.

Methods for Neural Point Processes

Attribution Methods



1

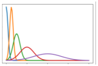
[1] Wei Zhang et al. "Cause: Learning granger causality from event sequences using attribution methods". In: *International Conference on Machine Learning*. PMLR. 2020, pp. 11235–11245.

Methods for Neural Point Processes

Attribution Methods

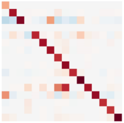
Target Function

$$\int_{t_i}^{t_{i+1}} \lambda_k^*(t')dt'$$



Integrated Gradient

$$A_j(f_k, \mathbf{x}_i, \underline{\mathbf{x}}_i)$$



Event Contributions

$$\mathbf{x}_i \triangleq [t_1, \mathbf{z}_1, \ldots, t_i, \mathbf{z}_i, t_{i+1}] \qquad \underline{\mathbf{x}}_i \triangleq [t_1, \mathbf{0}, \ldots, t_{i-1}, \mathbf{0}, t_{i+1}]$$

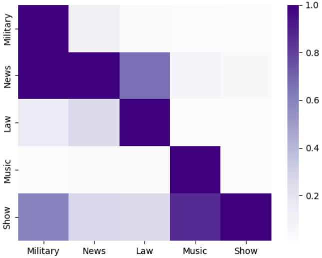
Input ***Baseline***

$$A_j(f_e, x_i^s, \tilde{x}_i^s) \xrightarrow{Normalization} Y_{e,e'}$$

Methods for Neural Point Processes

Inferring Granger Causality Matrix

The Granger Causality matrix G is built by $(g_{ij}) = Y_{i,j}$



(Granger Causality from B to A: $(rowA, columnB)$ element in the heat map)

Comparison of Granger Causal Discovery Methods for Event Sequences

Model	Idea	Model Capacity	Complexity	Data Volume
Vanilla-Hawkes	Hawkes+MLE+MM	Low	High	Low
NPHC	Hawkes+GMM	Low	Low	High
SeqCausal	Hawkes+Domain knowledge	Medium	Medium	Medium
L_0-Hawkes	Hawkes+L_0 penalty	High	High	Low
GC-nsHP	Hawkes+ES separation	Low	Low	High
THP	Hawkes+Topological graph	High	High	Low
Vanilla-Wold	Wold+Simulation	Low	Low	High
Vanilla-Neural	Encoder-decoder+Attribution method	Medium	XXX	XXX

How to select?

- Big data v.s. Small data
- High computation resource v.s. Low computation resource
- Have different event patterns?
- Have underlying topological graph?
- Have expert's domain knowledge?

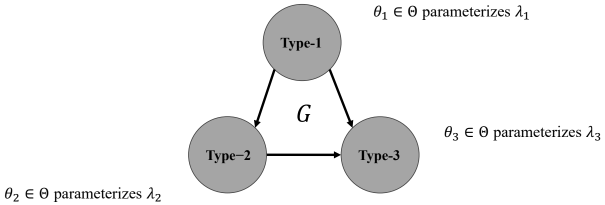
1. Background
2. Granger Causality Based Approaches
3. Constraint Based & Score Based Approaches

Other Causality Inferring Approaches

- Based on Graph Event Models (GEM)
- Directly model the relationships between different processes in the Graph Event Model
- On GEM, there are constraint-based and score-based methods

Graphical Event Model (GEM)

- Graphical Event Models (GEMs) have been proposed as a graphical representation for Multivariate Point Processes
- A GEM is a pair $< G, \Theta >$



- Vertices are for each event type $E = \{e_1, ..., e_n\}$
- Edges represent potential dependencies
- $\theta_{e_i} \in \Theta$ parameterizes e_i's intensity function

Process Independence (PI)

(Conditional) Process Independence[1]

For processes X, Y, Z, s.t. $Y \cap Z = \varnothing$, X is a process independent of Y given Z if all events in X have conditional intensities such that if historical information of events in Z is known, then those events in process Y do not provide any further information.[2]

[1] Christopher Meek. "Toward Learning Graphical and Causal Process Models.". In: *CI@ UAI*. 2014, pp. 43–48.

[2] Debarun Bhattacharjya et al. "Process independence testing in proximal graphical event models". In: *Conference on Causal Learning and Reasoning*. PMLR. 2022, pp. 144–161.

Constraint-Based Approaches

d-separation & d^*-separation

d-separation:
Two sets of variables are said to be *d*-separated if every path in the (Probabilistic) Graphical Model between them is blocked.

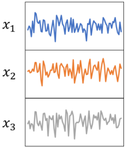
$\downarrow$

In a Graph Event Model (GEM), A path d^*-connects[1] nodes X and Y given the set of vertices Z in graph G if every collider on the path is an ancestor of Z and every non-collider is not in Z . For processes $X, Y, Z\ s.t. Y \cap Z = \varnothing$, X is d^*-separated from Y by Z in G if and only if there does not exist a non-trivial path that d^*-connects any node in X to any node in Y given Z in G

[1] Debarun Bhattacharjya et al. "Process independence testing in proximal graphical event models". In: *Conference on Causal Learning and Reasoning*. PMLR. 2022, pp. 144–161.

Constraint-Based Approaches

Event Sequences & Multivariate Time Series



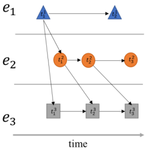
Multivariate Time Series

- *Structural Causal Model*
- *Conditional Independence*
- *d-separation*
- *Causal Faithfulness Assumption*

Analogous to ←

Event Sequence

- *Graph Event Model*
- *Conditional Process Independence*
- *d*-separation*
- *Causal Dependence Assumption*



- *Apply PC algorithm*
- ***Difficulty: Process Independence Tests*** ★

1

[1] Debarun Bhattacharjya et al. "Process independence testing in proximal graphical event models". In: *Conference on Causal Learning and Reasoning*. PMLR. 2022, pp. 144–161.

Constraint-Based Approaches

PC Algorithm on GEMs

Algorithm 1 PC Algorithm for Structure Discovery in GEMs

Data: Event label $X \in \mathcal{L}$, event dataset D (over $\mathcal{L}$), threshold parameter for tester α

Result: Parents $\mathbf{U}$ for X

```
U = L
for all Y in L do
    flag = False, n = 0, Z* = U \ Y
    while n ≤ |Z*| and flag = False do
        for all Z that are subsets of size n in Z* do
            Obtain score from a process independence test
            if score ≤ τ = g(α) (indicating process independence) then
                flag = True, U = U \ Y
                Break from loop
            end
        end
        n = n + 1
    end
end
```

1

[1] Debarun Bhattacharjya et al. "Process independence testing in proximal graphical event models". In: *Conference on Causal Learning and Reasoning*. PMLR. 2022, pp. 144–161.

Constraint-Based Approaches

Process Independence Testers

Process Independence testers[1]:

- It can test a dependence of an event type on set of event histories
- PI testers: NI score, LR score[2], etc.
- When the score is less than τ - a predefined threshold , the processes are independent conditional on Z, i.e., $Y \not\rightarrow X|Z$

[1] Christopher Meek. "Toward Learning Graphical and Causal Process Models.". In: *CI@ UAI*. 2014, pp. 43–48.

[2] Debarun Bhattacharjya et al. "Process independence testing in proximal graphical event models". In: *Conference on Causal Learning and Reasoning*. PMLR. 2022, pp. 144–161.

Score-Based Approaches

Analogous to cases in SCMs:

- In GEM, We can choose a score function (like BIC or Bayesian Gamma score[1])
- Then, we can apply a search algorithm (e.g., a forward-backward search (FBS) algorithm[2]) to find the graph $\widetilde{G}$ with the smallest score that comply with certain criterions $\rightarrow$ learn the structure of GEM

[1] Debarun Bhattacharjya et al. "Score-based learning of graphical event models with background knowledge augmentation". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 37. 10. 2023, pp. 12189–12197.

[2] Debarun Bhattacharjya, Dharmashankar Subramanian, and Tian Gao. "Proximal graphical event models". In: *Advances in Neural Information Processing Systems* 31 (2018).

Thanks!