

# 面向非稳态霍克斯过程的格兰杰因果发现算法研究

陈济斌, 蔡瑞初

广东工业大学计算机学院, 广东 广州  
Email: 18824152582@163.com

收稿日期: 2021年3月12日; 录用日期: 2021年4月7日; 发布日期: 2021年4月14日

## 摘要

发现非稳态霍克斯过程中潜在的格兰杰因果关系具有重要意义。现有的因果发现算法主要基于稳态性假设, 无法适用于非稳态的情况。为此, 文中提出了一种面向非稳态霍克斯过程的格兰杰因果发现算法: 首先, 建立非稳态霍克斯过程因果网络结构学习模型; 然后, 利用贪婪算法来发现某段霍克斯过程存在的模式; 最后, 利用基于极大似然估计的稀疏组套索(MLE-SGL)的方法发现模式对应的因果关系。在模拟数据上的实验效果验证了算法的正确性和有效性, 并在交互式网络电视(IPTV)数据集上发现了一些不同模式及有趣的因果关系。

## 关键词

霍克斯过程, 非稳态, 格兰杰因果关系, 交互式电视网络, 贪婪算法

# Research on Algorithms of Granger Causality Discovery for Non-Stationary Hawkes Process

Jibin Chen, Ruichu Cai

Faculty of Computer Science, Guangdong University of Technology, Guangzhou Guangdong  
Email: 18824152582@163.com

Received: Mar. 12<sup>th</sup>, 2021; accepted: Apr. 7<sup>th</sup>, 2021; published: Apr. 14<sup>th</sup>, 2021

## Abstract

The discovery of potential Granger Causality in the non-stationary Hawkes process is of great

significance. Existing causal discovery algorithms are mainly based on the assumption of stationary situation and thus cannot be applied to non-stationary situations. For this reason, this paper proposes an algorithm named Granger Causality for Non-stationary Hawkes Process (GC-NOHP): First, the work establishes the non-stationary Hawkes process causal network structure learning model; Then, the greedy algorithm is implemented to discover the patterns that are hidden in a certain Hawkes process and an MLE-SGL-based method is utilized to discover the corresponding Granger causality of the patterns. The experimental results on synthetic data verified the correctness and effectiveness of the algorithm, and the algorithm also can find different patterns and some interesting causal relationships on the Interactive Internet TV (IPTV) data set.

## Keywords

Hawkes Processes, Non-Stationary, Granger Causality, Interactive Internet TV, Greedy Algorithm

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

从观测数据中学习潜在的格兰杰因果关系(Granger Causality)对现实应用具有重要意义[1] [2] [3] [4], 如 IPTV 节目推荐[5]、用户商品购买推荐[6]、地震预警[7]、金融分析[8], 社交网络分析[9] [10]和生物信息学分析[11] [12]。这些场景中获取到的数据大多都是连续时间内的动态离散数据, 这样的数据被称为点过程数据。由于数据动态变化, 呈现出非稳态的特性, 如果想利用这些数据挖掘出数据背后蕴含的事物发生机制, 为人们提供有效的信息, 关键在于找出这些点过程数据下潜藏的模式, 即发现非稳态的格兰杰因果关系。

现有的大多数时序离散数据上的格兰杰因果关系发现的工作[5]-[11]主要关注一类特殊的点过程, 称为霍克斯过程, 其目的是模拟历史事件对未来事件有影响的复杂事件序列, 并试图发现这些过程的格兰杰因果关系。格兰杰因果关系可以通过挖掘出数据背后蕴含的事物发生机制实现, 同时能为人们提供有效的信息。比较经典的方法是 Xu 等人的工作[13]提出的一种有效的学习算法(MLE-SGL), 该算法将最大似然估计器(MLE)与稀疏组套索(SGL)正则化器相结合。此外, 他的模型的灵活性允许将聚类结构事件类型合并到学习框架中; Zhou 等人的工作[14]提出了一种凸优化方法(ODE), 通过将不同个体的重复事件建模为多维霍克斯过程来发现社会影响力的隐藏网络。此外, 此外他的工作中还设计了一种算法 ADM4, 该算法结合了乘法器的交替方向方法和主化最小化技术; Eichler 等人的工作[15]通过使用无穷阶自回归, 引入了一种基于点过程的时间离散版本的链接函数的新非参数估计器的方法(LS), 从而得出新估计量的一致性。上述的这些算法大都假设格兰杰因果关系的霍克斯过程是稳态的, 故在一个霍克斯过程稳态的情况下, 其可以稳健和合理性地恢复了格兰杰因果关系。

在现实中, 多元霍克斯过程中存在多个甚至重复的模式, 这意味着一个霍克斯过程通常是非稳态的。例如, 对于一个有孩子的家庭, 一个月内的电视观看记录数据可能存在两种不同的模式。因为观看电视的观众不同, 如成年人或儿童, 它的电视观看偏好将是不同的。以前的方法[13] [14] [15]的基本假设是数据是稳态的, 缺乏考虑到霍克斯过程中存在的非稳态情况, 无法知道某一段数据到底是归属于哪一个模式, 这将导致他们无法学习出不同模式对应正确的格兰杰因果关系, 甚至会得到虚假的因果关系。以图

1 为例, 具体来说, 在模式 A (工作日模式) 中, “法律”、“军事”和“新闻”节目之间存在着格兰杰因果关系, 即此时的观看群体在观看完“法律”或“军事”节目后, 很可能会接着收看“新闻”节目; 在模式 B (周末模式) 中, “儿童”、“音乐”和“娱乐”节目之间存在着格兰杰因果关系, 即此时的观看群体在观看完“儿童”节目后, 很可能会接着收看“音乐”节目, 最后也很可能在观看完“音乐”节目后, 接着收看“娱乐”节目。

因此, 为了解决现有方法无法发现非稳态的格兰杰因果关系问题, 本文提出了一种面向非稳态霍克斯过程的格兰杰因果发现算法框架。该框架首先检测非稳态模式出现的时刻, 然后基于检测到的时刻确定稳态情况下的区间, 最后在稳态的区间中学习格兰杰因果关系。基于此框架, 本文提出了一种面向非稳态霍克斯过程的格兰杰因果发现算法首先使用贪婪算法(Granger Causality for Non-stationary Hawkes Process, GC-NOHP)。GC-NOHP 算法首先利用贪婪算法寻找出现非稳态模式之间的分割点, 得到多个分段的时间序列。然后, 在每个分段上基于霍克斯过程构建格兰杰因果关系模型, 利用 MLE-SGL 算法[13]求解模型, 得到不同事件间的格兰杰因果关系。上述两个步骤迭代进行, 直到所有模型参数收敛为止。本文的贡献可总结为以下三个方面: 1. 提出了一个框架来学习非稳态霍克斯过程的格兰杰因果关系; 2. 提供了一种基于贪婪算法的方法来发现非稳态霍克斯过程背后的不同模式; 3. 提出的方法在模拟数据和真实数据中得到了验证, 且从真实数据实验中发现了一些有趣的结果。

## 2. 基础概念

### 2.1. 点过程

本点过程是描述随机点分布的随机过程。很多随机现象发生的时刻、地点、状态等往往可以用某一空间上的点来表示。例如, 服务台前顾客的到来时刻, 真空管阴极电子的发射时刻。又如, 天空中某一区域内星体的分布, 核医疗中放射性示踪物质在人体器官的各处出现, 不同能级地震的发生, 都可用二维以上空间的点表示, 点过程就是描述这类现象的理想化的数学模型。点过程可以在数轴上的非负实线上来表示, 其中非负线被用来表示时间, 点过程实际上是在  $t_i \in [0, T], i \in 1, 2, \dots, n$  时间段发生在数轴上的一系列离散事件[16]。点过程也可以被描述为一个计数过程  $N(t) = \{N(t) | t \in [0, T]\}$ , 其中  $N(t)$  记录了发生在  $t$  时刻之前发生的事件次数。一个多维具有  $c$  个类型事件点过程可以表示为  $\{N_c\}_{c=1}^C$ 。因此  $N_c(t) = \{N_c(t) | t \in [0, T]\}$  是某类型为  $c$  的时间在  $t$  时刻之前发生的次数。一种表示点过程的方法是使用条件强度函数来捕获自触发或自校正模式。定义为给定历史信息的发生类型为  $c$  的事件的瞬时发生率:

$$\lambda_c(t) dt = \lambda_c(t | \mathcal{h}_t^c) dt \quad (1)$$

其中  $\mathcal{h}_t^c = \{(t_i, c_i) | t_i < t, c_i \in \mathbb{C}\}$  包含所有类型的所有事件, 其时间戳小于时间  $t$ ,  $\mathbb{C} = \{1, \dots, C\}$  是一组事件类型。

前面提到的点过程中, 每个事件之间的发生事件戳都是独立地, 以恒定的速率  $\lambda$  (对于泊松过程) 或由强度函数  $\lambda(t)$  (对于非均匀泊松) 控制。然而, 对于一些应用场景, 已知事件的到达增加了在不久的将来观察事件的可能性。其中事件发生率显式地取决于或依赖于过去的事件发生: 自我激励过程, 这就是霍克斯过程。多维或多元霍克斯过程是一个计数过程也是一种特殊的点过程, 由一种特定称为强度函数的形式来表示:

$$\lambda_c(t) = \mu_c + \sum_{j=1}^{j=U} \int_0^t b_{cj} k(s) dN_j(t-s) \quad (2)$$

其中  $\mu_c$  是一个与历史事件无关的确定性基(base), 是一个常数, 而方程的第二部分捕获变量之间的互相影响影响,  $k(s)$  是核函数。本文根据现实场景, 限制核函数是单调递减的, 即事件的影响程度随着时间

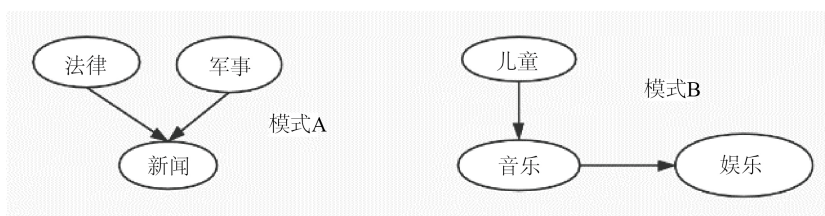
的推移而衰减或减少。该核函数基于历史类型  $c'$  事件对接下来发生的类型  $c$  事件的影响来度量衰减效应。

## 2.2. 核函数

虽然核函数  $k(t)$  不必单调递减, 但本文将形式设置为函数的递减族, 因为在现实世界中, 随着时间的推移, 自然会注意到事件衰减的影响。本文更多地关注变量之间的格兰杰因果关系, 而不是选择最优核函数。因此, 本文选择了广泛使用的指数函数[17]作为核函数, 定义为  $k(t) = \tau e^{-\tau t}$  (在这里  $\tau$  是常数)。

## 2.3. 格兰杰因果关系

格兰杰因果关系其实只是统计意义上的因果关系, 而不是哲学意义上探讨的因果关系。格兰杰因果关系是这样的一种因果关系: 即在包括  $X$  和  $Y$  的所有过去信息的条件下, 对于  $Y$  的预测结果要好于只包括  $Y$  的过去信息的条件下对于  $Y$  的预测结果, 那么便认为他们之间存在格兰杰因果关系。



**Figure 1.** For the viewing records of IPTV users with a time span of one week, it is likely to be a non-steady Hawkes process. This is because the structure of TV viewing audiences on weekdays and weekends is different, so there are two different Granger causalities (or two patterns)

**图 1.** 对于时间跨度为一周的 IPTV 用户观看记录, 它很可能是一个非稳态霍克斯过程, 这是因为工作日与周末电视观看观众的结构的不同, 因此存在着两种不同的格兰杰因果关系(或两种模式)

## 3. 算法模型

### 3.1. 问题定义

考虑长度为  $T$  的非稳态观测时间序列  $X_{original} = [x_1, x_2, \dots, x_T]$ , 其中  $x_i \in R_n$  是第  $i$  次观测;  $x_i = [t, c]$ , 其中  $t$  表示时间戳,  $c$  表示  $x_i$  属于的事件类型。在这里, 对于某一个数据点  $x_i$ , 本文根据这个数据点发生的时间戳之前发生的事件或数据作为背景来综合考虑与看待这个数据。因此, 本文转而考虑一个长度为  $l < T$  的短子序列, 而不是单单只考虑某个数据点  $x_i$ 。本文将观测值  $x_{t-l+1}, \dots, x_t$  链接起来然后把它们合成的新子序列称为  $X_{t-l+1}$ 。本文将这个从  $X_1$  到  $X_{t-l+1}$  的新子序列表示为  $X_{new}$ 。将孤立点组合成新的列是有帮助的, 因为它们为每个  $x_i$  观测数据点提供了上下文背景, 如果两个子序列属于同一模式, 它们应该有着的上下文信息或变量之间的关系。本文的目的是发现隐藏在非稳态序列数据背后的不同模式, 并捕获每个模式的格兰杰因果关系。

在这里本文定义了不同模式有不同的霍克斯过程模型参数, 类似于 Zhou 等人工作[14]中的做法。对于霍克斯过程, 本文将方程(2)中的第二项参数化为  $b_{cj}k(s)$ , 在二值化的因果矩阵(infectivity matrix)  $B = [\text{sign}(b_{cj})]$  是对应格兰杰因果图的邻接矩阵。如果  $b_{cj} \geq 0$ , 则表示  $j$  类型事件对  $c$  类型事件有影响, 否则, 这两种类型的事件是相互独立的, 本文提的霍克斯模型参数就是  $\theta = \{b_{cj}, \mu_c\}$ 。

### 3.2. 非稳态霍克斯过程的格兰杰因果关系模型

如果在时间序列中考虑某一时间点的数据时, 用类似于 Hallac 的思想[18]把上下文一起纳入考虑范



围, 那么这将提供更多有用的信息, 而不仅仅是将一个单独的数据点纳入考虑。因此, 对于具有一定长度的序列观测, 本文将其看成具有多个时间长度  $w$  的短子序列集合, 而不是看成孤立的数据点集合。本文假设变量在时间长度为  $w$  的短子序列中的局部格兰杰因果关系属于相同的模式(标记为  $\psi$ )是相同的, 而其时间不变属性也与现实相一致。也就是说, 假设相同的局部格兰杰因果关系意味着这些时间序列共享相同的霍克斯过程模型。由于具有相同霍克斯过程模型参数的子序列数据将具有相同的数据似然, 因此本文可以利用该数据似然作为来区分两个子序列是否来自相同的模式的参考信息, 从而发现霍克斯进程中存在的非稳态或不同的模式。为了控制相邻子序列通常属于同一模式的实际事实, 本文还添加了  $\beta * 1\{x_{t-1} \notin \psi_q\}$  这个惩罚项来调节时间一致性: 随着惩罚值  $\beta$  的增大, 相邻子序列属于同一模式的可能性更大。

对于霍克斯过程, 本文将方程(2)中的第二项参数化为  $b_{cj}k(s)$ , 在二值化的因果矩阵(infectivity matrix)  $B = [\text{sign}(b_{cj})]$  是对应格兰杰因果图的邻接矩阵。如果  $b_{cj} \geq 0$ , 则表示  $j$  类型事件对  $c$  类型事件有影响, 否则, 这两种类型的事件是相互独立的。故霍克斯模型参数为  $\theta = \{b_{cj}, \mu_c\}$ 。

类似于 Zhou 等人的工作[14], 由于具有相同霍克斯过程模型参数的子序列数据将具有相同的数据似然度, 因此可以利用该数据似然作为来区分两个子序列是否来自相同的模式的参考信息, 从而发现霍克斯进程中存在的非稳态或不同的模式。针对同一模式, 假设有  $C$  组时间序列数据  $\mathbb{C} = \{e_c\}_c^C$ ,  $e_c$  表示第  $c$  组的数据集, 参照 Xu 等人的工作[13], 观测数据的似然函数  $-ll(x_t, \theta_q)$  的完整形式为,

$$-ll(x_t, \theta_q) = \sum_{c=1}^C \left\{ \sum_{i=1}^{N_c} \log \lambda_{u_i^c}(t_i^c, \theta_q) - \sum_{u=1}^U \int_0^{T_c} \lambda_u(s, \theta_q) ds \right\} \quad (3)$$

其中  $u \in \{1, \dots, U\}$  表示事件的类型, 在式子(3)中, 等式左侧的  $x_t$  在右侧转化为该数据发生的时间戳  $t$  和时间类型  $u$ ;  $N$  为数据集,  $c$  为数据集编号。

假设每个模式  $q$  都有其霍克斯过程模型参数  $\theta_q$ , 为了控制相邻子序列通常属于同一模式的实际事实, 本文引入  $\beta * 1\{x_{t-1} \notin \psi_q\}$  这个惩罚项来调节时间一致性: 随着惩罚值  $\beta$  的增大, 相邻子序列属于同一模式的可能性更大。因此, 为了找到不同的模式, 并获得每个模式的最好模型参数, 有以下目标函数,

$$\arg \min_{\psi} \sum_{q=1}^K \left[ \sum_{x_t \in \psi_q} \left( -ll(x_t, \theta_q) + \beta * 1\{x_{t-1} \notin \psi_q\} \right) \right] \quad (4)$$

其中  $-ll(x_t, \theta_q)$  是  $x_t$  属于模式  $q$  的对数似然,  $\beta * 1\{x_{t-1} \notin \psi_q\}$  表示  $x_{t-1}$  和  $x_t$  不属于相同模式的惩罚代价。

### 3.3. GC-NOHP 算法

为了可以同时做到最小化负对数似然以及通过正则化参数  $\beta$  控制时间一致性, 本文通过将每个  $T$  子序列分配给  $K$  模式之一来解决优化问题。当  $\beta = 0$  时, 子序列  $X_1, \dots, X_T$  倾向于独立分配到不同的  $T$  个模式, 因为没有惩罚来促进相邻子序列被分配到相同模式的情况。通过将每个点分配给最大化其可能性的模式来解决这个问题, 随着  $\beta$  的值越来越大, 相邻子序列被分配到同一模式的可能性越来越大。

为了求解方程(3), 从随机初始化模式参数开始, 迭代步骤 1 和步骤 2, 直到每个数据点在在每个模式  $\psi$  的分布情况(决定哪些数据属于哪个模式)和模型参数  $\theta_q$  的更新是趋于稳定的(即参数已经收敛)。在步骤 2 中, 本文利用了 Xu 等人[13]的 MLE-SGL 方法, 在数据已经模式分段划分完成的情况下, 对每个分段的格兰杰因果关系(学习出每个模式对应的模型参数)进行学习。算法 1 对本文算法(GC-NOHP)框架进行了概述。

算法 1 基于 MLE-SGL 框架的 GC-NOHP 算法

输入 观测  $X_{ori} = [x_1, x_2, \dots, x_T]$ , 模式数  $K$ , 惩罚项值  $\beta$ 。

输出 模型参数  $\Theta = \{\theta_1, \theta_2, \dots, \theta_q, \dots, \theta_K\}$ , 模式参数  $\Psi = \{\psi_1, \psi_2, \dots, \psi_q, \dots, \psi_K\}$ 。

1. 将数据点分配到对应的模式  $\psi^n \rightarrow \psi^{n+1}$
2. 使用 MLE-SGL 更新每个模式对应的模型参数  $\Theta^n \rightarrow \Theta^{n+1}$
3. 重复执行步骤 1 和步骤 2 直到  $\Theta$  和  $\Psi$  收敛。

### 3.4. 非稳态模式发现方法

下面是 3.2 节中算法 1 的步骤 1 的实现细节: 假定有  $K$  个不同的模式, 给定每个模式的各自的模型参数  $\theta_q = \{b_{ij}^q, \mu_i^q\}$ ,  $q \in \{1, 2, \dots, K\}$ , 本文通过将  $T$  个子序列  $X_1, \dots, X_T$  分配给这些  $K$  个模式来求解方程(3), 其方法是最小化数据的负对数似然, 同时尽量让相邻的  $X_{i+1}$  与  $X_i$  被分配到同一个模式(即减小 2.2 节中提到的惩罚项)。给定  $T$  个子序列, 它们可分配到  $K$  个模式, 这个组合优化问题, 总共有  $K * T$  种可能。根据 3.2 节的算法 1 中可以看出, 算法在进行数据点(本文这里将一个子序列看成一个数据点)分配到对应模式时, 模型参数是固定的, 那么在这种情况下对于每个数据点  $x_t$ , 它的似然  $-ll(x_t, \theta_q)$  与其他数据点  $x_{i, i \in t}$  的似然  $-ll(x_{i, i \in t}, \theta_q)$  是相互独立的, 也就是说每个数据点被分配到哪一个模式是可以独立完成的, 那么本文可以通过基于贪婪算法来让某一个数据点选择最优似然的分配决策来选择它应该属于的最佳模式, 从而完成最佳的数据点的分配方案, 具体策略, 见算法 2。

**算法 2** 基于贪婪算法的数据点分配方法

输入 观测  $X_{ori} = [x_1, x_2, \dots, x_T]$ , 模式  $K$  数, 惩罚项值  $\beta$ 。

输出 finalList[] (表示对应数据点编号被分配到的对应模式的情况)。

//初始化

1. finalList[1...T] ← 1, tmpList[1...T] ← 1

//对数据进行循环操作, 选择似然度最大的情况

2. for i in [1, T] do //对 T 个数据进行循环操作

MinValue ← min value of  $\{-ll(x_i, \theta_1), \dots, -ll(x_i, \theta_K)\}$ ;

for q in [1, K] do

if q != tmpList[i]

if  $\beta + (-ll(x_i, \theta_q)) < -ll(x_i, \theta_{finalist[i]})$

finalist[i] ← q.

算法 2 首先初始化所有数据点, 所有数据点初始默认属于模式 1, 然后对  $T$  个数据点进行循环操作, 在循环中如果当前数据属于另一个模式的跳转损失小于当前数据没有发生模式跳转的损失, 则将这个数据属于的模式进行改变, 否则让这个数据属于的模式不改变。

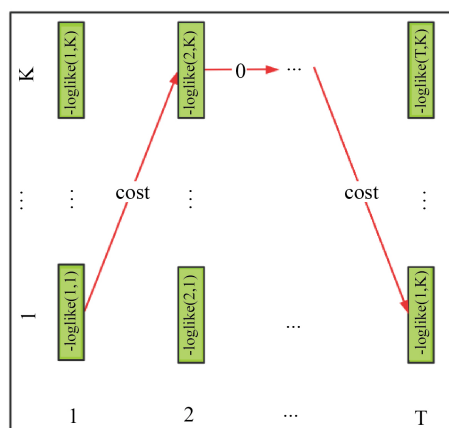
## 4. 算法验证

为了验证本文的方法的正确性和有效性, 我们进行了仿真数据和真实数据实验。我们通过对模拟数据实验的评估度量本文提出的方法是否能优于目前的一些对比方法, 以及我们的方法是否能从真实数据集中发现非稳态模式, 并且学习每个模式的格兰杰因果关系。

### 4.1. 仿真数据实验

我们生成了随机生成了 3 个不同维度的观测变量  $D \in \{3, 4, 5\}$  的合成数据集, 每个数据集都包含时间长度  $T = 200$  的事件序列  $N = 1000$ 。所有数据集都基于不同的二值化的因果矩阵(用于表示不同的格兰杰

因果关系或模式), 每个数据集包含 3 种不同的模式。3 个模拟数据集根据以下规则生成: 1) 随机生成因果矩阵。图 2 中的每个子图是归一化的因果矩阵, 对应元素  $(i, j)$  亮度越大, 说明  $j$  事件的发生有更高的概率导致接下来  $i$  事件发生; 2) 根据生成的因果矩阵, 参考 Jesper 等人提出的霍克斯过程的近似仿真方法[19]来生成模拟数据, 也就是说通过此方式生成的模拟数据是符合图 3 中的格兰杰因果关系。此模拟数据实验是基于三个不同的因果矩阵生成长度为  $3T$  的数据集对于时间戳大于 0 小于  $T$  的数据点, 基于第一个因果矩阵生成; 对于时间戳大于  $T$  且小于等于  $2T$  的数据点, 基于第二因果矩阵生成; 对于时间戳大于  $2T$  小于等于  $3T$  的数据点, 同理。



**Figure 2.** A small rectangular box in the figure represents the likelihood of a certain data, the vertical axis represents the number of patterns, and the horizontal axis represents time. Algorithm 2 can be equivalent to finding a red path from the data with a time length of  $T$  based on the principle of greed. The small rectangular boxes connected on the path represent the pattern distribution of data from time stamps 1 to  $T$

图 2. 图中的一个矩形小框代表某一个数据似然, 纵轴代表模式数量, 横轴代表时间, 算法 2 可以等价于从时间长为  $T$  的数据基于贪婪原则找到一条如上图的红色路径, 这路径上连接的矩形小框, 就代表着时间戳 1 到  $T$  的数据的模式分配情况

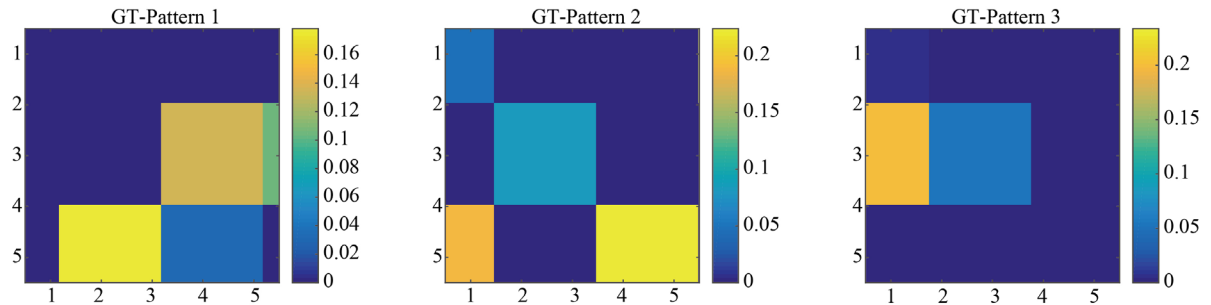
我们对每个数据集进行了参数组  $(w, \beta)$  不同设置情况下的实验。我们用  $w \in \{50, 70, 100\}$  改变子序列的长度, 一致性因子的值  $\beta \in \{50, 100, 150\}$ , 以验证这两个参数如何影响我们的方法的性能。我们还将我们的方法与现有的方法进行了比较, 例如基于 MLE-SGL 算法[13]和 ODE 的算法[14]和最小二乘(LS)算法[15], 我们根据一下两个评价指标进行比较:

- 比较学习出来的因果矩阵的可视化效果;
- 给定估计(根据学习而得出来的)因果矩阵  $B_{est}$ , 计算的其与真实的因果矩阵  $B_{real}$  的相对误差。

$D = 5$  (即考虑五个不同变量)的情况的可视化结果如图 3~5 所示, 以验证我们的方法和对比方法(通过评估度量 1)之间的性能差别, 如图 3~5 所示。在图 6 显示了在  $D \in \{3, 4, 5\}$  的情况下, 不同方法的学习出来的因果矩阵的相对误差 RelErr。

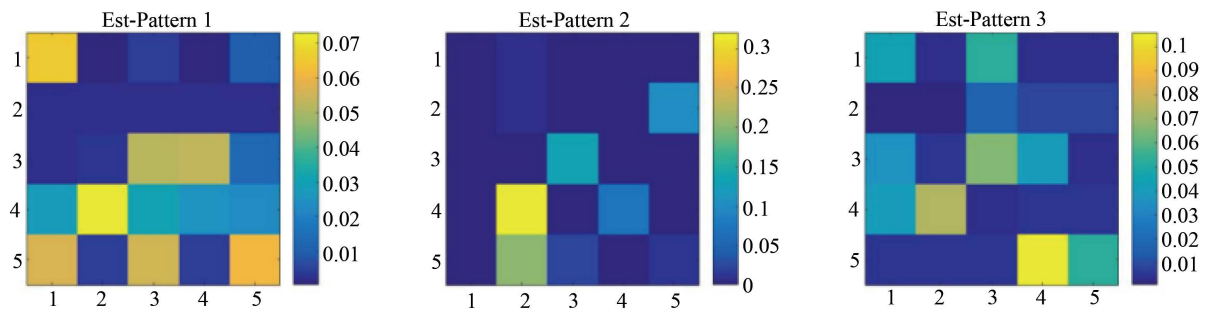
从仿真实验的结果可以看出: a) 根据图 3~5, 我们的方法可以发现存在于非稳态霍克斯过程数据集的模式, 学习每个模式的格兰杰因果关系(尽管存在着些许偏差)因此用本论文的方法可以学习出三个不同的因果矩阵, 但对比方法由于不能发现非稳态的模式, 它们每个方法只能学习出对应的一个因果矩阵, 显然由于它们不能发现非稳态模式, 它们得出的因果矩阵是于图 2 的 Ground Truth 真实因果矩阵是不符合的。b) 较低的相对误差 RelErr 意味着在恢复某一模式的 Ground Truth 真实格兰杰因果关系方面有更好

表现, 从图 6 中, 我们的方法相对于其他对比方法得到了较低的相对误差 RelErr, 这表明了我们的可靠性和有效性。c) 造成 a)和 b)的本质上的原因对比方法(MLE-SGL、ODE、LS)都假设格兰杰因果关系的霍克斯过程是稳态的, 无法适用于非稳态的情况, 而模拟数据是非稳态的, 故在一个霍克斯过程是非稳态的情况下, 对比方法并不可以稳健和合理地恢复了格兰杰因果关系。



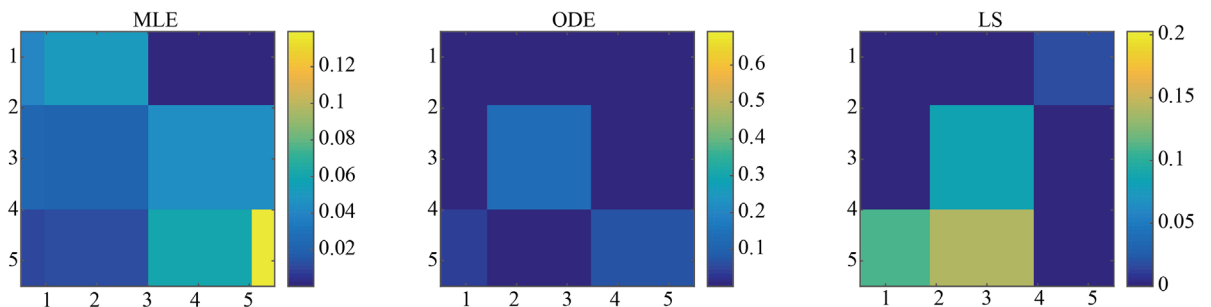
**Figure 3.** The Ground Truth visualization result of the causal matrix used to generate the pair of  $D = 5$  data sets. Each small image is a normalized (probability value) causal matrix. The greater the brightness of the corresponding element, the higher the occurrence of the event Probability leads to the next event

图 3. 用于生成  $D = 5$  数据集的对用的因果矩阵的 Ground Truth 可视化结果, 每个小图是归一化的(概率值)因果矩阵, 对应元素  $(i, j)$  亮度越大, 说明  $j$  事件的发生有更高的概率导致接下来  $i$  事件发生



**Figure 4.** In the case of  $D = 5$  data set, the visual result of the causal matrix learned by the algorithm of this paper, compared with Figure 3 shows that the method of this paper can basically restore Ground Truth with non-steady-state and non-steady-state characteristics

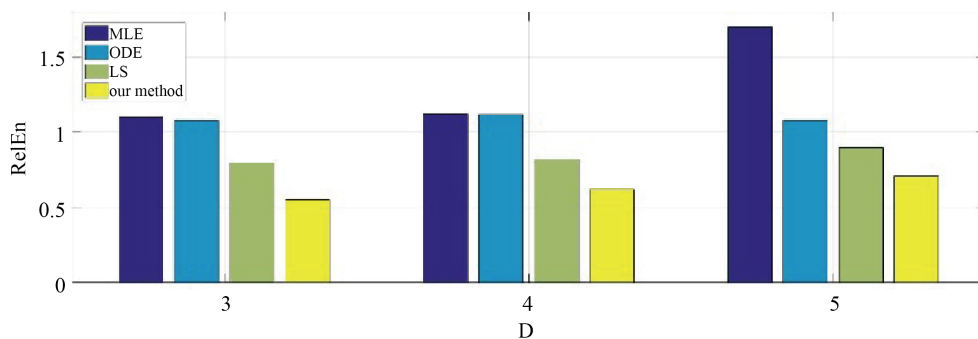
图 4. 在  $D = 5$  数据集的情况下, 用本文的算法学习出来的因果矩阵的可视化结果, 与图 3 对比可知本文的方法可以基本恢复出具有非稳态非稳态特性 Ground Truth



**Figure 5.** In the case of the  $D = 5$  data set, the visualization results of the causal matrix learned by the comparison algorithm (MLE, ODE, LS), it can be seen that the learning effect of the three algorithms on Ground Truth with non-steady state characteristics is not ideal

图 5. 在  $D = 5$  数据集的情况下, 用对比算法(MLE, ODE, LS)学习出来的因果矩阵的可视化结果, 可见这三个算法对具有非稳态特性 Ground Truth 的学习效果不理想



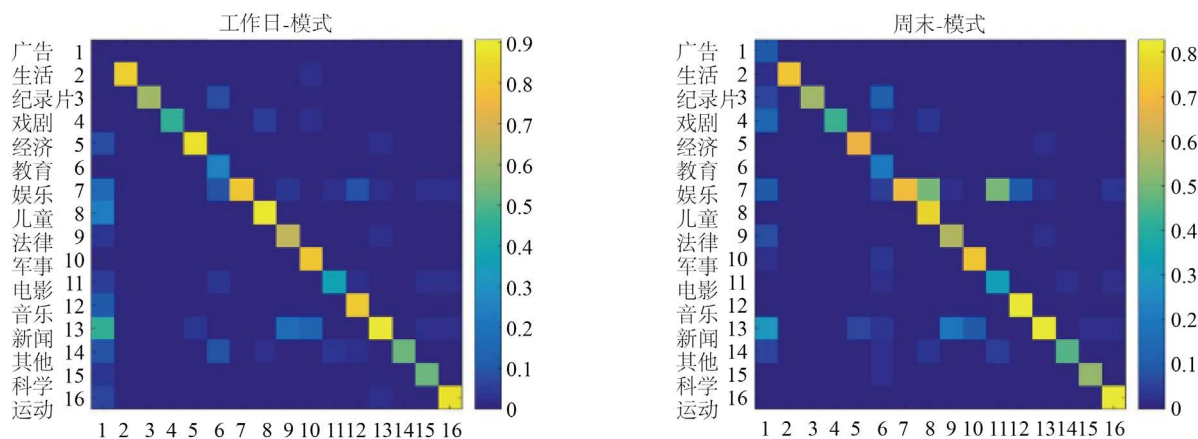


**Figure 6.** In the case of  $D = 3, 4, 5$  data sets, the comparison of the relative error RelErr of the causal matrix learned by different methods (MLE, ODE, LS, our method-GC-NOHP) can be seen in the case of different dimensions, The relative error RelErr obtained by the GC-NOHP algorithm is the lowest among these four algorithms

**图 6.** 在  $D = 3, 4, 5$  数据集的情况下, 不同方法(MLE, ODE, LS, our method-GC-NOHP)的学习出来的因果矩阵的相对误差 RelErr 的对比, 可见在不同维度的情况下, GC-NOHP 算法得到的相对误差 RelErr 在这四种算法中是最低的

## 4.2. 真实数据实验

我们使用 IPTV 电视观看记录数据集[20]并且通过学习这非稳态霍克斯过程中潜在的格兰杰因果关系来验证我们的方法。该数据集具有  $N = 302$  个序列, 该序列记录在很长一段时间内用户的电视观看行为, 时间跨度从 2012 年 1 月至当年 11 月。此数据集包含 16 种类型的电视节目。我们通过一个类似于 Xu 等人[13]的霍克斯过程来模拟用户的电视节目查看行为。我们尝试用本文的方法去发现这份数据集中用户观看电视节目的观看偏好序列(也是尝试发现其中潜在的非稳态观看偏好序列), 即电视节目的类别存在自我和相互触发的模式。例如, 观看一部电影的一集将导致观看以下几集(自我触发)和相关的新闻(可以在新闻类别中看到)演员(相互触发)。因此, 类别之间的因果关系不仅取决于播放时间表, 而且也取决于用户的观看选择。学习到这份这种触摸模式将有利于电视节目的排挡, 从而让电视台从中获得更高的商业利润。



**Figure 7.** The method proposed in this paper is applied to the IPTV data set to obtain two steady-state causal matrices (workday mode and weekend mode)

**图 7.** 本文提出的方法应用于 IPTV 数据集上得到两个稳态情况下的因果矩阵(工作日模式和周末模式)

在这里我们将长达一年的数据集, 按一周的时间跨度进行分割预处理, 然后尝试发现在长度为一周

的时间观测窗口, 本文算法能找到一些潜在于其中的非稳态格兰杰因果关系。可视化结果如图 7, 在我们的学习结果中观察到了一些有趣的现象: a) 我们的方法发现了 IPTV 查看记录数据集背后的两种不同模式: 例如, 工作日模式更多地关注新闻, 这种情况可能更经常发生在工作日, 而周末模式中有更多的娱乐节目可能发生在周末, 因此结果重新考虑了我们的方法考虑了非稳态的情况。b) 我们可以仔细看看每种模式的局部部分。从图 7 中的周末模式中, “娱乐消遣” 类别很可能是由“儿童”和“电影”触发的, 换句话说, 如果观众看“儿童”和“电影类”节目, 他们更有可能喜欢看一些关于“娱乐消遣”的电视节目。因此, 周末模式更像是有小孩子在操控着遥控器, 看着更多的“儿童”, “电影”, “娱乐消遣”类的节目, 这种观看偏好很明显区别于工作日模式。

## 5. 结束语

本文提出了一种有效的算法来学习非稳态霍克斯过程中潜在的格兰杰因果关系。该方法可以发现存在于数据中的非稳态的模式, 并学习每个模式对应的格兰杰因果关系。本文通过实验证明了我们在模拟数据和真实数据方面的工作的稳健性和合理性。通过真实的 IPTV 电视节目观看记录数据, 我们发现了工作日和非工作日用户观看的电视节目间的不同因果关系。这对电视广告营销、电视节目内容推送等具有一定的帮助。但是本文的方法的算法复杂度较高, 未来将通过优化数据点分配的算法, 选择其他可能的优化算法去替代本文使用的贪婪算法。未来也将尝试将该方法运用到别的真实数据集上, 尝试发现其他数据集中潜藏的有趣现象。

## 参考文献

- [1] Wei, C., Cai, R. and Hao, Z. (2019) Mining Hidden Non-Redundant Causal Relationships in Online Social Networks. *Neural Computing and Applications*, **32**, 6913-6923. <https://doi.org/10.1007/s00521-019-04161-5>
- [2] Cai, R., Zhang, Z. and Hao, Z. (2016) Understanding Social Causalities behind Human Action Sequences. *IEEE Transactions on Neural Networks and Learning Systems*, **28**, 1801-1813. <https://doi.org/10.1109/TNNLS.2016.2556724>
- [3] 蔡瑞初, 陈薇, 张坤, 郝志峰. 基于非时序观察数据的因果关系发现综述[J]. 计算机学报, 2017, 40(6): 1470-1490.
- [4] 陈薇, 蔡瑞初, 伍运金, 谢峰, 郝志峰. 基于多组典型相关变量的因果关系发现算法[J]. 计算机应用研究, 2021, 38(1): 53-56.
- [5] Gurleen, K. and Sukhmani, A. (2011) Study of TV Viewership Patterns among Youngsters in Northern India. *Zenith International Journal of Multidisciplinary Research*, **1**, 141-160.
- [6] Levin, A.M., Levin, I.P. and Weller, J.A. (2005) A Multi-Attribute Analysis of Preferences for Online and Offline Shopping: Differences Across Products, Consumers, and Shopping Stages. *Journal of Electronic Commerce Research*, **6**, 281-290.
- [7] Daley, D.J. and Vere-Jones, D. (2007) An Introduction to the Theory of Point Processes: Volume II: General Theory and Structure. Springer-Verlag, New York.
- [8] Bacry, E., Delattre, S., Hoffmann, M. and Muzy, J.F. (2013) Some Limit Theorems for Hawkes Processes and Application to Financial Statistics. *Stochastic Processes and their Applications*, **123**, 2475-2499. <https://doi.org/10.1016/j.spa.2013.04.007>
- [9] Farajtabar, M. Gomez-Rodriguez, M., Wang, Y., Li, S., Zha, H. and Song, L. (2018) COEVOLVE: A Joint Point Process Model for Information Diffusion and Network Co-Evolution. *Companion Proceedings of the Web Conference 2018*, Lyon, April 2018, 473-477. <https://doi.org/10.1145/3184558.3186236>
- [10] 蔡瑞初, 谢泳, 陈薇, 曾艳, 郝志峰, 杜文俊. 面向社交媒体的直接因果网络发现算法[J]. 计算机应用研究, 2020, 37(9): 2689-2693.
- [11] Reynaud-Bouret, P. and Schbath, S. (2010) Adaptive Estimation for Hawkes Processes: Application to Genome Analysis. *The Annals of Statistics*, **38**, 2781-2822. <https://doi.org/10.1214/10-AOS806>
- [12] Carstensen, L., Sandelin, A., Winther, O. and Hansen, N.R. (2010) Multivariate Hawkes Process Models of the Occurrence of Regulatory Elements. *BMC Bioinformatics*, **11**, Article No. 456. <https://doi.org/10.1186/1471-2105-11-456>

- 
- [13] Xu, H., Farajtabar, M. and Zha, H. (2016) Learning Granger Causality for Hawkes Processes. *33th International Conference on Machine Learning*, New York, 20-22 June 2016, 1717-1726.
- [14] Zhou, K., Zha, H. and Song, L. (2013) Learning Social Infectivity in Sparse Low-Rank Networks Using Multi-Dimensional Hawkes Processes. *16th International Conference on Artificial Intelligence and Statistics*, Scottsdale, 29 April-1 May 2013, 641-649.
- [15] Eichler, M., Dahlhaus, R. and Dueck, J. (2017) Graphical Modeling for Multivariate Hawkes Processes with Nonparametric Link Functions. *Journal of Time Series Analysis*, **38**, 225-242. <https://doi.org/10.1111/jtsa.12213>
- [16] Rizoiu, M.-A., Lee, Y., Mishra, S. and Xie, L. (2017) A Tutorial on Hawkes Processes for Events in Social Media. arXiv preprint arXiv:1708.06401.
- [17] Hawkes, A.G. (1971) Spectra of Some Self-Exciting and Mutually Exciting Point Processes. *Biometrika*, **58**, 83-90. <https://doi.org/10.1093/biomet/58.1.83>
- [18] Hallac, D., Vare, S. and Boyd, S. (2018) Toeplitz Inverse Covariance-Based Clustering of Multivariate Time Series Data. *27th International Joint Conference on Artificial Intelligence*, Stockholm, 13-19 July 2018, 5254-5258. <https://doi.org/10.24963/ijcai.2018/732>
- [19] Møller, J. and Rasmussen, J.G. (2006) Approximate Simulation of Hawkes Processes. *Methodology and Computing in Applied Probability*, **8**, 53-64. <https://doi.org/10.1007/s11009-006-7288-z>
- [20] Luo, D., Xu, H., Zha, H., Du, J., Xie, R., Yang, X. and Zhang, W. (2014) You Are What You Watch and When You Watch: Inferring Household Structures from IPTV Viewing Data. *IEEE Transactions on Broadcasting*, **60**, 61-72. <https://doi.org/10.1109/TBC.2013.2295894>