arXiv:2302.06075v1 [stat.ME] 13 Feb 2023

用于理解多触点归因中移除效应的图形点过程框架

Jun Tao[1,2], Qian Chen[1], James W.Snyder Jr.[2], Arava SaiKumar[2],Amirhossein Meisami[2],和 Lingzhou Xue[1]

[1]宾夕法尼亚州立大学； [2]Adobe公司

初稿：2022年5月；本版本：2023年2月

摘要

营销人员利用各种在线广告渠道触达客户，并尤其关注归因——衡量单个触点对最终转化的贡献程度。个体客户级购买路径数据的可用性，以及在线营销渠道和触点类型的不断增加，为这一基础问题带来了新的挑战。我们旨在通过在路径层级开展归因，以更精细的粒度解决归因问题。为此，我们开发了一种新颖的图形点过程框架，用于同时研究直接转化效应以及众多类型触点之间完整的关系结构。利用转化的时间点过程和图形结构，我们进一步提出图形归因方法，为每位客户的购买路径中的各个触点或相应渠道分配适当的路径级转化贡献，称为归因得分。我们提出的归因方法将归因得分视为移除效应，并采用严格的概率定义推导出两种移除效应。我们通过大量模拟研究考察所提方法的表现，并将其与常用归因模型进行比较。我们还通过一项现实归因应用展示了所提方法的表现。

关键词：格兰杰因果关系、图模型、高维统计学、多触点归因、点过程。

1 引言

归因是一个经典问题。数据的粒度在研究这一问题时发挥着重要作用。早期研究利用汇总数据考察渠道间的协同效应，以获取客户定向和营销预算分配方面的洞见（例如，Naik & Raman 2003年）。随着在线数据收集技术的发展，个体客户级购买路径数据变得可用，这些数据描述了个体客户在购买漏斗中与各种渠道互动的时间和方式。此类数据的出现使归因建模进入了一个新阶段。不同客户的购买路径各不相同，触点之间的时间距离也有所不同。在图1中，我们给出了两个客户级购买路径数据的示例。如路径1所示，一位客户在时间 t_1收到了关于焦点产品的电子邮件广告，随后在时间 t_2在社交媒体上看到了该产品的广告。之后，这位客户在搜索引擎中搜索该产品，并在时间 t_3看到该产品的付费广告出现在搜索结果页顶部。这位客户在时间 t_4 点击了该产品广告，访问产品网站，并在时间 t_5 购买了该产品。如路径2所示，另一位客户在时间 t_1' 看到了关于该产品的搜索引擎广告，随后在时间 t_2' 点击了该广告，访问产品网站，并在时间 t_3'完成转化。

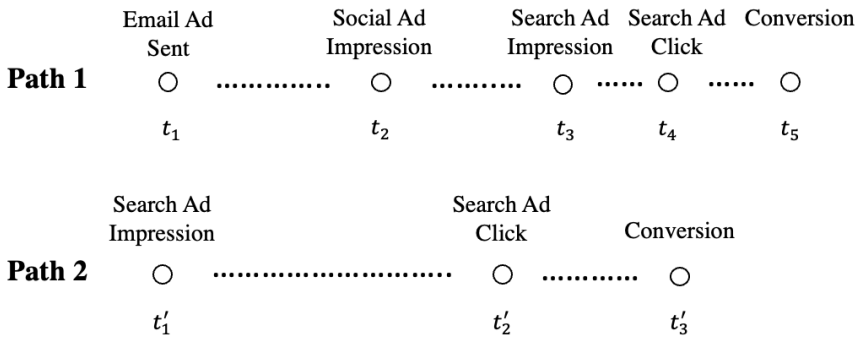

图1：客户级购买路径数据示意图。

研究人员已开始利用客户级购买路径数据进行归因建模（例如，Xu 等，2014、Li 和 Kannan，2014 年、 Anderl 等人，2016 年）。近年来，在线营销渠道和触点类型日益增多。这些渠道包括各种搜索引擎（例如谷歌、必应）、社交媒体平台（例如领英、Twitter、Instagram、TikTok、Snapchat、WhatsApp）、展示广告、电子邮件、网页横幅广告、应用横幅广告、客户支持、桌面通知等。同时，每个搜索引擎、平台或媒体中还包含许多不同类型的触点。

例如，广告曝光和广告点击是不同的触点，可能产生不同的转化效应。为了实现精细化营销，营销人员希望评估每个搜索引擎、平台或媒体的营销效果，甚至评估某一特定类型触点的效果。但在现有的大多数归因建模研究中，路径上的触点通常会被汇总到其所属的渠道（例如搜索、展示广告、电子邮件）；同一渠道内的不同搜索引擎、平台或媒体不会被区分，同一搜索引擎、平台或媒体内的不同触点类型也不会被区分。例如，搜索广告曝光和点击都被视为客户与搜索渠道之间的互动，未被单独研究；通过 Instagram、Twitter 及其他社交媒体平台产生的触点也不会被区分，而是作为一个整体社交媒体渠道进行研究。此外，在现有的大多数研究中，同时研究的渠道数量也很少。

与近期归因建模研究类似，我们也聚焦于购买路径数据。但我们旨在通过在路径层级进行归因并提供触点级评分，实现更精细的分析。具体而言，我们希望针对每位客户的购买路径，将适当的转化功劳分配给每个触点或一组触点。这样，我们便能为单个触点和渠道提供归因分数，帮助营销人员做出更好的营销决策。此外，我们还希望同时研究来自更多渠道的触点，以更好地满足企业当前的需求。

为每个触点分配适当的功劳极具挑战性，因为触点之间既存在渠道内互动，也存在跨渠道互动。例如，看到电子邮件和社交广告，可能会提高客户今后在搜索引擎中搜索该产品，继而点击搜索广告访问企业网站并完成在线购买的概率。此外，一个触点（例如搜索广告曝光）可能会引发同一渠道内未来触点的出现（例如搜索广告点击），进而带来转化。关于较早出现的触点可能提高未来触点出现概率并促成转化的现象，Li & Kannan (2014年将其称为延续效应（通过相同渠道）和溢出效应（通过其他渠道），并对此进行了深入讨论。

为了恰当捕捉各种触点的直接和间接转化效应，我们必须在归因建模中考虑以下特征。首先，不同类型触点的发生概率可能存在显著差异，它们会激发其他类型的触点并影响购买转化。因此，有必要对不同类型触点的多变量特性进行建模。其次，所有营销效应，包括转化效应

以及触点之间的交互效应，都会随时间衰减。在归因建模中，必须考虑两个触点之间的时间间隔。仅利用触点的顺序信息会忽略这种衰减效应，从而导致估计偏差。第三，触点和购买转化很可能会在时间线上聚集成簇。换言之，购买路径数据具有聚集性（Zhang 等，2013年）。第四，归因建模及其估计方法必须具备可扩展性，以便分析大量营销渠道和类型各异的触点。

1.1 相关文献

归因问题在过去二十多年里一直受到学术界和业界研究人员的广泛关注。本小节简要概述现有方法，更多详情可参阅综述论文（Kannan 等，2016、Gaur 与 Bharti，2020）。现有归因模型可分为两大类：基于规则的启发式方法和数据驱动方法。

1.1.1 基于规则的启发式方法

简单的基于规则的启发式方法在实践中广泛用于多触点归因。例如，末次触点归因方法将全部功劳分配给转化前的最后一个触点；首次触点归因将功劳分配给客户旅程中的第一个触点；U 型方法则将同等份额的功劳分配给第一个和最后一个触点，并将剩余功劳平均分配给其他触点。这些基于规则的启发式方法仍可轻松应用于路径转化数据，但它们要么忽略其他触点的影响，要么未考虑触点之间的交互效应。此外，这些方法也因存在偏差，且缺乏充分理由来证明其作为归因衡量方法的合理性而受到批评（Singal 等人，2022年）。

1.1.2 数据驱动方法

数据驱动方法主要包括增量价值（或移除效应）方法和夏普利值方法。该领域以往研究的主要创新体现在以下两个方面。大多数研究侧重于提出描述用户行为的新模型（例如，Shao 与 Li，2011，Breuer 等人，2011年、Danaher 与 van Heerde，2018、Xu 等，2014、Zhao 等人，2019年），而其他研究则侧重于探讨归因评分方法，例如夏普利值的理论依据和公平性（例如，

Dalessandro 等人，2012年，Singal 等人，2022年）。

- 增量价值（或移除效应）方法通过从客户路径中移除一个或一组触点，计算转化概率的变化。因此，转化概率的变化也称为移除效应。过去十年间，研究人员开发了各种描述消费者行为的模型，例如回归模型（例如，Shao 与 Li，2011、Breuer 等人，2011年、Danaher 与 van Heerde，2018、Zhao 等人，2019年）、马尔可夫模型（Yang 和 Ghose，2010年、Anderl 等人，2016年、Berman，2018年、Kakalejˇcík 等人，2018年）、贝叶斯模型（例如，Li 和 Kannan，2014年）、时间序列模型（Kireyev 等人，2016年、De Haan 等人，2016年）、基于生存理论的模型（Zhang 等人，2014年、Ji 等人，2016年）、深度学习模型（Li 等人，2018年、Kumar 等人，2020年）等。这一系列以往研究的主要创新在于对用户行为进行建模。大多数研究将触点视为确定性事件而非随机事件，并忽略了这些营销传播与干预措施之间的动态交互作用（Xu 等人，2014）。然而，考虑这些触点的激发效应十分重要。此外，正如Singal 等人(2022）所指出的，基于增量价值的归因几乎没有（如果确实存在的话）理论依据。

- 夏普利值方法运用基于博弈论的夏普利值概念（Shapley，1953年），为合作博弈中的各个参与者分配功劳。受夏普利值本身特性的影响，它通常只能提供渠道级归因，而无法提供路径级归因或触点级归因得分。此外，现有基于夏普利值的方法未考虑购买路径数据中触点之间的时间间隔，包括（Dalessandro 等人，2012年，De Haan 等人，2016年，Kireyev 等人，2016年，Berman，2018年，Singal 等人，2022年）。例如，Singal 等人最新的研究（Singal 等人（2022年）使用离散马尔可夫链模型描述客户在转化漏斗中的客户旅程里状态转换，但该模型未纳入客户在一次转换中从一个状态转移到另一个状态时的时间间隔。

- 归因也从其他角度进行了研究。例如，Xu 等人(2014年)提出了一种采用多元点过程的贝叶斯方法，并通过模拟计算归因分数。然而，这种基于模拟的归因方法计算开销很大，且无法为每条包含转化的观测路径提供路径级得分。

1.2 我们的方法

据我们所知，现有模型都无法研究跨多个渠道的多种触点之间的完整关系结构，而这对于理解每种触点及其对应渠道的直接和间接转化效应至关重要。鉴于研究中涉及的触点类型可能数量众多，因此需要一种能够同时研究这些触点类型之间的交互效应及其转化效应的模型。为弥补这些研究空白，我们在本文中作出以下努力：

首先，我们提出一种新颖的图形点过程归因模型，利用客户级购买路径数据描述多渠道场景中的客户行为。该图模型不仅能学习多种类型触点的直接转化效应，还能同时估计不同类型触点之间的激发效应。

其次，我们提出图形归因方法，为每位客户的购买路径中的各个触点及其对应渠道分配适当的转化功劳，称为归因得分。我们的方法将归因得分视为移除效应。我们利用转化的时间点过程和图形结构，推导出两种移除效应。

第三，我们提出一种正则化方法，用于同时进行边选择和参数估计。我们设计了一种定制的交替方向乘子法（ADMM），以高效且可扩展的方式求解这一优化问题。此外，我们通过建立参数估计值的渐近理论，提供了理论保证。

下面我们简要介绍所提出的图形点过程模型的基本思想。首先，我们将购买路径数据建模为多元时间点过程。更具体地说，所提出的方法将各条触点路径视为相互独立的事件流。每个事件流都由在共同时间线上以不规则且异步方式发生的各种类型触点（事件）组成。为了捕捉大量相互独立的事件流中触点之间的动态相互依赖关系（例如激发模式），所提出的方法将事件流建模为多元时间点过程。多元时间点过程通常使用条件强度函数（Gunawardana 等人 2011年）进行刻画，这些函数描述了在给定先前触点历史的情况下，未来触点发生的瞬时速率。所提出的模型能够考虑触点之间的时间距离及其簇集特性。Xu 等人 (2014年）是第一篇使用多元时间点过程处理归因建模的论文。他们将广告点击和购买视为相互依赖的随机

事件，并在连续时间框架下将模型构建为贝叶斯层级框架。

所提出的模型进一步引入了格兰杰因果图，这是一种用于表示各种事件类型之间依赖关系的有向图。图中的节点表示事件类型，有向边描述某一类型事件对其他事件的历史影响，这些影响被称为格兰杰因果关系（Granger 1969）。所提出的图形模型支持大量在线营销渠道和多种触点类型。通过充分捕捉各种事件类型之间的格兰杰因果关系，该模型能够同时衡量多种触点对转化的影响，以及不同类型触点在渠道内部和渠道之间产生的激发效应。

基于所学习的图，我们提出了图形归因方法，为每种触点或相应渠道的转化分配适当的功劳。转化功劳也称为归因分数，从细粒度视角出发，在路径层级进行计算。必要时，这些分数可以汇总到渠道层级。第一种归因方法衡量感兴趣事件对转化的直接效应，即仅从路径中移除感兴趣的事件，并假设路径上的其他触点不受影响时，转化强度的相对变化。第二种归因方法充分利用图形因果结构，衡量感兴趣事件的总移除效应。相应的归因得分，是在不仅考虑移除感兴趣事件，还考虑后续客户发起事件可能损失的情况下，预期转化强度的边际提升。

我们通过模拟研究考察所提出方法的表现，并使用两组模拟数据将其与常用的归因模型进行比较。一组数据由多元霍克斯过程（Hawkes 1971）模拟生成。另一组数据则由谷歌公司开发的数字广告系统模拟（DASS）的一个修改版本模拟生成。模拟数据包含在线客户浏览行为，以及注入的、会影响客户行为的广告事件。在衡量各渠道对转化的贡献时，我们提出的方法优于基准模型。此外，我们还展示了所提出方法在现实归因应用中的表现。

1.3 我们的贡献

我们为从业者提供了一种新的归因建模工具，帮助他们在如今可能存在大量不同类型触点的线上多渠道场景中，以更精细的粒度了解不同营销活动如何促成转化。

我们的工具通过在个体客户路径层级进行归因建模，为各个触点及其对应渠道合理分配功劳。该工具有助于企业开展精细化营销运营、进行预算分配、实现利润最大化等。

除实质性贡献外，我们还在方法论方面为现有文献作出了以下贡献。

首先，我们提出一种图形点过程模型，利用客户个体层面的购买路径数据描述客户行为，从而为归因建模文献作出贡献。在这一框架中，我们对多种类型触点的激发效应进行严谨建模，并开发了一种高效的惩罚算法来进行模型估计。我们将格兰杰因果关系应用于营销情境，以研究营销活动之间的时间关系。此外，我们还建立了参数估计值的渐近理论。

其次，我们提出的图形归因方法为增量价值（或移除效应）方法相关文献作出了贡献。我们对归因得分给出了严谨的概率定义，并推导出两类移除效应，即直接移除效应和总移除效应。我们分别开发了一种新颖高效的基于稀疏化的模拟方法和一种反向传播算法，用于计算这两类移除效应。

本文其余部分安排如下。第2节介绍所提出的图形点过程模型，第3节介绍所提出的图形归因方法。第4节阐述模型估计、计算细节和渐近性质。随后，我们在第5节利用模拟数据展示所提方法和算法的性能，并在第6节给出一项实证应用。最后，第7节总结本文。

2 图形点过程模型

为解决归因问题，我们首先提出一种图形点过程模型。该模型利用客户级购买路径数据，学习不同类型触点之间的完整关系结构。本节将介绍所提出的图形点过程模型。

我们的图形点过程模型将观测到的、通向购买的个体路径视为相互独立的事件流，其中每条事件流都由不同类型的事件（即触点和转化）组成，这些事件在同一时间线上以不规则且不同步的方式发生。假设共有 p 种独特的事件类型，可标记为 1、...、p。客户的路径 D 表示为 $\{(t_i, e_i)\}_{i=1}^{m}$，其中 $0 \leq t_1 < t_2 < \cdots < t_m \leq T$，

m 是已发生事件的总数，T 是观测长度。对于第 i 个事件（t_i，e_i），t_i 是其时间戳，$e_i \in \mathcal{E} = \{1, 2, \ldots, p\}$ 是其事件标签，用于指明触点类型（例如，社交广告曝光、社交广告点击、邮件发送、邮件广告点击、搜索广告曝光、搜索广告点击、展示广告曝光、展示广告点击）或转化。一般性不失，令 $e = 1$ 表示转化的标签。给定路径 D，如果存在 $i \leq m$ 使得 $e_i = 1$，这意味着发生了转化事件，那么这样的路径 D 称为正路径。否则，D 称为负路径。假设我们观测到n条路径，D_1，...，D_n，其中 $D_j = \{(t_i^j, e_i^j)\}_{i=1}^{m_j}$ 是第 j 条路径，其长度为 T_j。

为捕捉大量相互独立的事件流中各触点之间的动态相互依赖关系，所提出的框架将事件流建模为多元时间点过程。该框架引入有向格兰杰因果图，以表示事件流中各类触点之间的依赖关系。本节首先简要介绍时间点过程和格兰杰因果图，然后详细介绍我们提出的模型。

2.1 时间点过程

对于标记为 e 的事件类型，我们可以将其在时间线上的发生情况描述为时间点过程 $N_e(t)$。函数 $N_e(t)$表示截至时间 t 发生的类型为 e 的事件数量，它是一个右连续、非递减的分段函数 $\mathbb{R}_{\geq 0} \to \mathbb{N}$:

$$N_e(t) = \sum_{i:e_i=e} \mathbb{1}_{\{t_i \leq t\}}.$$

我们假设 $N_e(t)$ 几乎必然以 1 为步长，且不会同时跳跃。若 $t' < t$，则 $N_e(t) - N_e(t')$ 表示在区间 $(t', t]$ 内发生的类型为 e 的事件数量，也可记为 $N_e((t', t])$。将 p 种事件类型放在一起，我们令 $\mathbf{N}(t)$ 表示计数函数向量 $(N_1(t) \ldots, N_p(t))^\top$。坐标 $N_e(t)$，$e = 1, \ldots, p$，由其条件强度刻画

$$\lambda_e(t \mid \mathcal{H}_t) := \lim_{\Delta t \to 0^+} \frac{1}{\Delta t} \mathbb{P}(N_e(t + \Delta t) - N_e(t) > 0 \mid \mathcal{H}_t),$$

它描述未来类型为 e 事件的瞬时发生速率。滤子$\mathcal{H}_t := \sigma\{\mathbf{N}(u) : u < t\}$ 是所有截至但不包括 t 时刻的事件所构成的 σ-代数，表示时间 t 之前的历史信息。

一个例子是多元霍克斯过程（Hawkes 1971），其条件强度函数为

$$\lambda_e(t \mid \mathcal{H}_t) = \mu_e + \sum_{e'=1}^{p} \int_0^t h_{e'e}(t - u) dN_{e'}(u).$$

这里， $\mu_e > 0$ 是基线强度，作为不受历史影响的背景发生率； $h_{e'e}(\cdot) \geq 0$ 称为影响函数。直观地说，每个过去的e'类事件（其 $h_{e'e}(\cdot) > 0$ ）都会通过提高 e类事件的条件强度，对当前的e类事件的发生产生正向贡献，而且这种影响可能会随时间推移而减弱。这种正向影响称为从 e' 类事件到 e类事件的激发效应。

2.2 格兰杰因果图

令 $\mathcal{E} = \{1, 2, \ldots, p\}$ 表示各种类型的标记事件（即触点和转化）构成的集合，这些事件彼此之间的历史影响备受关注。我们使用格兰杰因果关系（Granger 1969）来描述所研究的触点类型之间的时间依赖关系。如果事件类型 e' 的历史记录有助于预测事件类型e ，且其预测效果超出仅依据事件类型 e 历史记录所能达到的水平，则称事件类型 e' “格兰杰导致”事件类型 e。格兰杰因果关系最初由Granger (1969)在原始论文中提出并加以讨论，后续论文（Granger1980、1988）也对此进行了探讨，而Sims(1972)则给出了格兰杰因果关系的另一种定义。

对于任意事件标签子集 $W \subseteq \mathcal{E}$，令 $\mathbf{N}_W(t)$ 表示子过程（$N_e(t)$）$_{e \in W}$。例如，$\mathbf{N}^{\mathcal{E}\setminus\{e,e'\}}(t)$ 表示除 e 和 e' 之外所有事件类型构成的子过程。对于时间点过程，格兰杰因果关系定义如下。

定义 *1.*（局部独立性（*Didelez 2008*）） 给定 $\boldsymbol{N}^{\mathcal{E}\setminus\{e,e'\}}(t)$，时间点过程 $N_e(t)$ 关于 $N_{e'}(t)$ 局部独立，记作 $N_{e'} \nrightarrow N_e \,|\boldsymbol{N}^{\mathcal{E}\setminus\{e,e'\}}$，当且仅当条件强度 $\lambda_e(t)$ 关于 $\sigma\{\boldsymbol{N}^{\mathcal{E}\setminus\{e,e'\}}(u)$ 可测：$u < t\}$ 对所有 $t < T$ 均成立。否则，称 $N_e(t)$ 在给定 $\boldsymbol{N}^{\mathcal{E}\setminus\{e,e'\}}(t)$ 时关于 $N_{e'}(t)$ 局部依赖，记作 $N_{e'} \to N_e \,|\boldsymbol{N}^{\mathcal{E}\setminus\{e,e'\}}$。

上述定义由 Didelez(2008) 针对区间内的标记点过程提出 $[0, T]$。 Eichler (2012研究了 $\mathbb{R}$ 中类似于多元霍克斯过程的平稳多元点过程，并采用了“ 格兰杰非因果性 ”这一概念。上述定义等价于说，时间点过程 $N_{e'}(t)$ 不会格兰杰导致$N_e(t)$，就历史信息 $\mathcal{H}_t$ 而言。否则， $N_{e'}(t)$ 会格兰杰导致 $N_e(t)$，就 $\mathcal{H}_t$ 而言。为便于理解，我们也会明确地说事件类型 e' 格兰杰导致事件类型 e 。

我们使用有向图 $\mathcal{G} = (\mathcal{E}, \mathcal{A})$ 如图 2 所示，用以表示转化路径中各种触点类型之间的时间依赖关系。该图的节点集

表示事件类型的集合 $\mathcal{E}$。边集 $\mathcal{A} \subseteq \mathcal{E} \times \mathcal{E}$表示各种事件类型之间的格兰杰因果关系。如果就 $\mathcal{H}_t$而言， $N_{e'}(t)$ 格兰杰导致 $N_e(t)$，则从节点 e' 到节点 e的有向边，记作 $(e' \to e)$，被认为属于 $\mathcal{A}$。这样的图 $\mathcal{G} = (\mathcal{E}, \mathcal{A})$ 称为格兰杰因果图。对于多元时间点过程，格兰杰因果图定义如下。

定义 2. *(*格兰杰因果图（*Didelez 2008*， *Eichler 2012*）） 多元时间点过程 $\mathbf{N}(t)$ 称为遵循格兰杰因果图 $\mathcal{G} = (\mathcal{E}, \mathcal{A})$，如果对于任意一对节点 e， $e' \in \mathcal{E}$，

$$(e' \to e) \notin \mathcal{A} \iff N_{e'} \nrightarrow N_e \mid \mathbf{N}_{\mathcal{E}\setminus\{e,e'\}}.$$

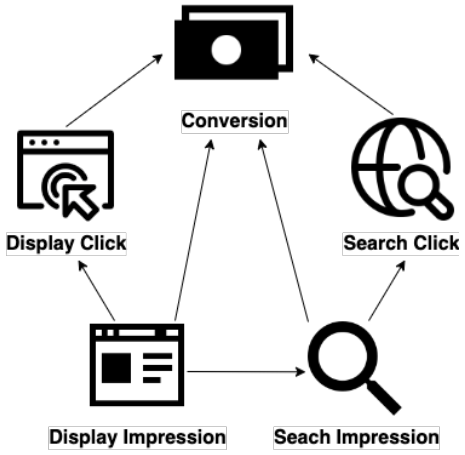

图2：归因格兰杰因果图的一个简单示例。四种类型的触点都会对转化产生格兰杰因果关系。在展示渠道或搜索渠道中，曝光都会对点击产生格兰杰因果关系。在两个渠道之间，展示广告曝光会对搜索曝光产生格兰杰因果关系。

2.3 所提出的图形点过程模型

我们将每条购买路径视为一个 p维点过程$\mathbf{N}(t)$。在归因问题中，事件类型分为两类：企业发起的事件类型和客户发起的事件类型。企业发起的事件类型（Wiesel 等人，2011）由企业发起，例如电子邮件发送、社交广告曝光和展示广告曝光。客户发起的事件类型（Bowman 与 Narayandas，2001，Wiesel 等人，2011）由客户或潜在客户发起，包括转化、搜索广告曝光和点击、邮件点击等。

等等。令 $\mathcal{E}_{\mathrm{f}}$ 和 $\mathcal{E}_{\mathrm{c}}$ 分别表示企业发起的事件类型集合和客户发起的事件类型集合，因此 $\mathcal{E}=\mathcal{E}_{\mathrm{f}}\cup\mathcal{E}_{\mathrm{c}}$ 和 $\mathcal{E}_{\mathrm{f}}\cap\mathcal{E}_{\mathrm{c}}=\emptyset$。假设客户发起的事件类型有 q 种，其中 $1\leq q<p$。不失一般性，令 $\mathcal{E}_{\mathrm{c}}=\{1,\ldots,q\}$，其中转化标记为 $e=1$ 和 $\mathcal{E}_{\mathrm{f}}=\{q+1,\ldots,p\}$。我们假设企业发起的事件类型的点过程$\mathbf{N}_{\mathcal{E}_{\mathrm{f}}}(t)$由企业或广告商策略性地控制。换言之，$\mathbf{N}_{\mathcal{E}_{\mathrm{f}}}(t)$仅作为观测输入，其条件强度函数无需学习。对于客户发起的事件类型的点过程$\mathbf{N}_{\mathcal{E}_{\mathrm{c}}}(t)$，我们通过以下条件强度对其进行建模：

$$\lambda_e(t\mid\mathcal{H}_t)=\mu_e+\sum_{e'=1}^{p}\alpha_{e'e}\int_0^t\psi_{e'e}(t-u)dN_{e'}(u),\quad\text{for } e\in\mathcal{E}_{\mathrm{c}},\tag{1}$$

- $\mu_e\geq 0$ 是基线强度，它对应于除自身及其他类型触点的发生历史之外，导致类型-e事件发生的来源，即事件发生的内在倾向。

- $\psi_{e'e}(\cdot)\geq 0$，$e'=1$，……，p是一个有界、左连续的核函数，$\psi_{e'e}(t)=0$当 $t\leq 0$ 且 $\int_0^\infty\psi_{e'e}(t)dt=1$. 核函数 $\psi_{e'e}(\cdot)$ 描述了触点影响的形状。例如，$\psi_{e'e}(t)=\frac{1}{T_0}\cdot\mathbb{1}_{\{0<t\leq T_0\}}$ 表示某个先前类型-e' 事件在其发生后的 T_0 内产生恒定影响；$\psi_{e'e}(t)=\frac{1}{T_0}\exp(-\frac{t}{T_0})\cdot\mathbb{1}_{\{t>0\}}$适用于指数衰减影响；当衰减影响更快时，可以使用 $\psi_{e'e}(t)=\sqrt{\frac{2}{\pi T_0^2}}\exp(-\frac{t^2}{2T_0^2})\cdot\mathbb{1}_{\{t>0\}}$。

- $\alpha_{e'e}\geq 0$，$e'=1$，……，p 是格兰杰因果关系系数。$\alpha_{e'e}$ 的值描述了时间依赖的程度。

为了诠释系数 $\alpha_{e'e}$，我们引入以下定理。

定理 1. 假设点过程 $\mathbf{N}(t)$ 的条件强度函数定义如(1)，且遵循格兰杰因果图 $\mathcal{G}=(\mathcal{E},\mathcal{A})$。对于任意事件标签 $e'\in\mathcal{E}$和 $e\in\mathcal{E}_{\mathrm{c}}$，如果条件 $N_{e'}(T)>0$ 成立，那么

$$(e'\to e)\notin\mathcal{A}\Longleftrightarrow\alpha_{e'e}=0.$$

该结果是从多元霍克斯过程（Eichler2012，Xu 等，2016）的情形推广而来。与多元霍克斯过程类似，$\alpha_{e'e}$的含义可分为两种情况：$\alpha_{e'e}>0$ 表示事件类型 e' 通过提高其条件强度 e，对事件类型 λ_e产生激发效应（$t\mid\mathcal{H}_t$）；而 $\alpha_{e'e}=0$ 则表示事件类型 e' 与事件类型 e 之间不存在格兰杰因果关系，因为 λ_e（$t\mid\mathcal{H}_t$）不受任何事件发生的影响

的任何类型-e' 事件。通过这种构造，我们可以用矩阵 $A = (\alpha_{e'e})_{e' \in \mathcal{E},\ e \in \mathcal{E}_c} \in \mathbb{R}_{\geq 0}^{p \times q}$ 表示客户发起的事件类型的图形化格兰杰因果结构。

由于企业发起的事件类型完全由营销人员或广告商控制，因此在我们提出的模型中，假设它们不依赖于其他类型的事件。也就是说，对于任意 $e' \in \mathcal{E}$， $\mathcal{A}$ 的真实值对于 $e \in \mathcal{E}_{\mathrm{f}}$ 而言是已知的，且为 $N_{e'} \nrightarrow N_e \mid \mathbf{N}^{\mathcal{E} \setminus \{e, e'\}}$ 。

3 图形化归因方法

归因的目标是将转化的功劳合理分配给每种触点或相应的渠道。我们采用路径级评分方法，以获得更细致的视角。该方法不依赖于企业发起的事件类型$\mathbf{N}_{\mathcal{E}_{\mathrm{f}}}(t)$的分布。因此，它适用于客户采用不同广告策略的情形。

路径级功劳分配，也称为归因得分，表示在缺少某些事件的情况下，某条路径上的转化可能损失的比例。我们将这一数值称为这些事件的移除效应。本节首先通过分析每个触点的增量贡献，利用转化的时间点过程推导一种移除效应，即直接移除效应。随后，我们充分利用图形结构，提出另一种移除效应，即总移除效应，用以解释转化概率的边际增量。

3.1 直接移除效应

在本小节中，我们从图形化点过程的角度，使用直接移除效应计算归因得分。我们将直接移除效应 （DRE）定义为：仅从路径中移除感兴趣的事件，同时假设路径上的其他触点不受影响时，转化强度的相对变化。也就是说，直接移除效应仅考虑所研究事件对转化发生的影响，而忽略其对其他事件的影响。从图形角度来说，直接移除效应关注转化的直接格兰杰因果父节点，而不依赖于这些节点之外的层级结构。

对于正路径 D，假设在 $t = t_{i^\star}$ 发生了一次转化，其中 $1 < i^\star \leq m$，即 $e_{i^\star} = 1$。令 $F_t^W(D) := \{(t_i,\ e_i) \in D\colon\ t_i < t,\quad e_i \in W\}$ 表示在 t 之前发生且事件标签属于 W 的事件集合，其中 $W \subseteq \mathcal{E}$ 是任意事件标签子集。特别地，令 $F_t(D)$ 表示 $F_t^{\mathcal{E}}(D)$，即 t 之前的截断路径。我们关注触点子集 $R \subseteq F_{t^\star}(D)$ 对此次转化的影响，并将其称为

R移除集。令 $\mathbf{N}^D(t) = (N_1^D(t) \ldots, N_p^D(t))^\top$ 表示关于路径 D、具有 $N_e^D(t) = \sum_{(t_i,\ e_i) \in D:\ e_i = e} \mathbb{1}_{\{t_i \le t\}}$ for $e = 1, \ldots, p$ 的点过程。令 $\mathcal{H}_t^D$为对应的滤子 $\sigma\{\mathbf{N}^D(u) : u < t\}$。我们可以将 R 关于路径 D的归因得分计算为其直接移除效应：

$$\mathrm{att}_{t_{i^\star}}^{(\mathrm{direct})}(R \mid D) := \frac{\lambda_1(t_{i^\star} \mid \mathcal{H}_{t_{i^\star}}^D) - \lambda_1(t_{i^\star} \mid \mathcal{H}_{t_{i^\star}}^{D \setminus R})}{\lambda_1(t_{i^\star} \mid \mathcal{H}_{t_{i^\star}}^D)}. \tag{2}$$

上述定义是一个一般性陈述，不依赖于任何点过程建模假设。

假设转化事件满足模型（1），这意味着该转化的条件强度函数具有如下形式

$$\lambda_1(t_{i^\star} \mid \mathcal{H}_{t_{i^\star}}) = \mu_1 + \sum_{i < i^\star} \alpha_{e_i 1} \psi_{e_i 1}(t_{i^\star} - t_i). \tag{3}$$

我们可以根据方程（2）和（3）推导出更具体的归因分数。例如，触点（t_i，e_i）的直接移除效应可计算为

$$\mathrm{att}_{t_{i^\star}}^{(\mathrm{direct})}(\{(t_i, e_i)\} \mid D) = \frac{\alpha_{e_i 1} \psi_{e_i 1}(t_{i^\star} - t_i)}{\lambda_1(t_{i^\star} \mid \mathcal{H}_{t_{i^\star}}^D)}.$$

该表达式展示了格兰杰因果图与归因得分之间的关系。只有当触点（t_i，e_i）的触点类型是图中转化节点的父节点，或者（$e_i \to 1$） $\in \mathcal{A}$时，才能为其分配得分。一般而言，对于 $F_{t^\star}$（D）的一个子集 R，其直接移除效应为

$$\mathrm{att}_{t_{i^\star}}^{(\mathrm{direct})}(R \mid D) = \sum_{(t_i, e_i) \in R} \frac{\alpha_{e_i 1} \psi_{e_i 1}(t_{i^\star} - t_i)}{\lambda_1(t_{i^\star} \mid \mathcal{H}_{t_{i^\star}}^D)}. \tag{4}$$

此外，在模型（1）下，基线效应可定义为$\lambda_1 \overset{\mu_1}{(t_{i^\star} | \mathcal{H}_{t_{i^\star}}^D)}$。例如，图3 展示了由三个触点和一次转化组成的路径中各触点的直接移除效应。

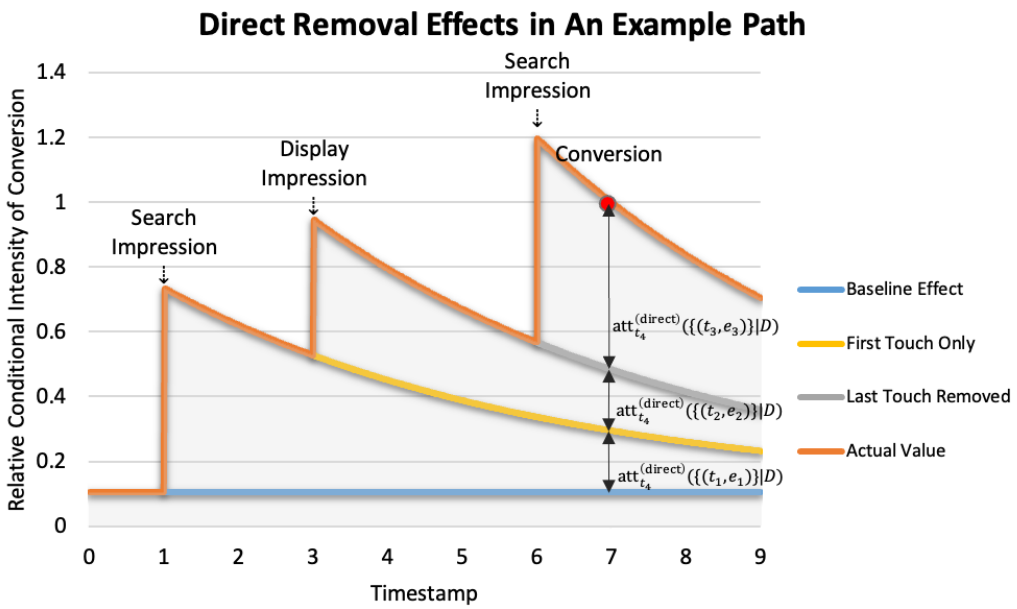

图3：示例路径中各触点的直接移除效应，以及转化相对条件强度的分解。该路径包含三个触点，其中 $t_1 = 1$、$t_2 = 3$、 $t_3 = 6$ 和 e_1、 e_2、和 e_3 分别表示搜索曝光、展示广告曝光和搜索曝光。在 $t_4 = 7$发生了一次转化。

在以下内容中，我们将用概率语言解释归因得分的一般形式（2）。考虑一个关于客户 A 和 B 的理想实验，图 4 展示了一个简化示例。假设 A 和 B 以相同的方式对触点作出反应。他们的客户发起的事件类型的条件强度函数遵循相同的模型。我们观察到D_{A}，即 A 的路径，并且在 $t = t_{i^\star} > 0$ 处发生了转化。令 D_{B} 表示 B 的路径。假设$F_{t_{i^\star}}^{\mathcal{E}\setminus\{1\}}(D_{\mathrm{B}}) \subsetneqq F_{t_{i^\star}}^{\mathcal{E}\setminus\{1\}}(D_{\mathrm{A}})$ 中， D_{B} 上的触点是 D_{A}上触点的子集。假设没有关于 $N_1^{D_{\mathrm{B}}}(t)$ 的信息。也就是说，我们无法观察 B 是否发生转化。我们可以对过程 $N_1^{D_{\mathrm{B}}}(t)$ 进行采样，以获得完整路径 D_{B}。我们利用时间点过程的稀疏化操作，该操作依据某一确定规则删除基本点过程中的点，从而生成一个新的点过程。

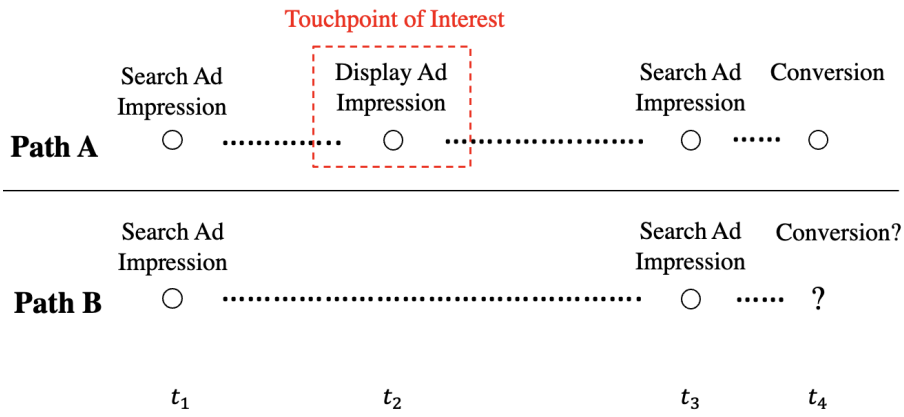

图 4：通过展示广告曝光的直接移除效应解释的问题，涉及一条包含三个触点的路径。移除展示广告曝光后，直接移除效应描述了在相同时间戳看到转化的可能性。

定理2。 *(Lewis 和 Shedler 1979年，Ogata 1981年)* 假设一个单变量时间点过程 $N(t)$ 定义在 $[0,T]$ 上，其强度函数为 $\lambda(t)$。令 t_1，……，$t_{N(T)}$ 为 $N(t)$ 的时间戳。存在一个函数 $\lambda'(t)$ 几乎处处满足

$$\lambda'(t) \leq \lambda(t), \quad for \;\; 0 < t < T.$$

对于 $i = 1, \ldots, N(T)$，以概率 $1 - \lambda'(t_i)/\lambda(t_i)$ 删除位于 $t = t_i$ 的点。然后，剩余的点构成区间 $[0,T]$ 上的点过程 $N'(t)$，其强度函数为 $\lambda'(t)$。

我们将点过程 $N(t)$ 称为目标点过程 $N'(t)$ 的背景过程。一般而言，Ogata (1981年) 建议在点过程模拟中使用泊松过程作为背景过程。要对 $N_1^{D_{\mathrm{B}}}(t)$ 进行采样，我们可以寻找一个满足 λ $\lambda \geq \lambda_1(t \mid \mathcal{H}_t^{D_{\mathrm{B}}})$ 的常数 $t \in [0,T]$，并使用速率为 λ 的泊松过程作为背景过程。

仔细考察这个问题后，我们会发现，无需将泊松过程作为背景过程进行采样。A 的转化点过程 $N_1^{D_{\mathrm{A}}}(t)$ 也可以作为背景过程。根据 (3)，在 B 发生任何转化之前，我们有 $\lambda_1(t \mid \mathcal{H}_t^{D_{\mathrm{B}}}) \leq \lambda_1(t \mid \mathcal{H}_t^{D_{\mathrm{A}}})$。定理 2 指出，如果以概率 $t = t_{i^\star}$ 在 $N_1^{D_{\mathrm{A}}}(t)$ 中删除时刻 $1 - \lambda_1(t_{i^\star} \mid \mathcal{H}_{t_{i^\star}}^{D_{\mathrm{B}}})/\lambda_1(t_{i^\star} \mid \mathcal{H}_{t_{i^\star}}^{D_{\mathrm{A}}})$，那么我们就得到一个服从式 (3) 所述分布的 B 的过程。与式 (2) 相比，这个概率正是 $D_{\mathrm{A}} \setminus D_{\mathrm{B}}$ 相对于路径 D_{A} 的直接移除效应。

3.2 总移除效应

给定事件子集的直接移除效应，描述的是将不包含该子集的路径与原始路径进行比较时，预期转化损失的大小。在本小节中，我们将研究另一种归因得分，它衡量沿路径移除该事件子集所产生的整体影响。相应的归因得分是预期转化强度的边际提升，其计算不仅考虑移除感兴趣的事件，还考虑其他相关客户发起事件可能造成的损失。

例如，可以通过研究第一个触点的直接移除效应（t_1、e_1）$\in D$ 来分析 $D \setminus \{$（t_1、e_1）$\}$。但是，移除路径早期阶段的某个触点，其影响不仅限于移除该触点本身。一些较晚出现的触点可能会受到影响，从而趋于不再发生。因此，$D \setminus \{$（t_1、e_1）$\}$ 可能无法很好地表示模型（1）所描述的某条可能的剩余路径，因为$\mathbf{N}_{\mathcal{E}_c}^{D\setminus\{(t_1,e_1\}}$）$(t)$无法保证遵循与$\mathbf{N}_{\mathcal{E}_c}^{D}(t)$相同的模型。换言之，直接移除效应可能无法充分反映移除该触点所产生的整体影响。

根据模型（1），对于任意 $e \in \mathcal{E}_c$（包括转化），

$$\lambda_e(t \mid \mathcal{H}_t^{D\setminus R}) \le \lambda_e(t \mid \mathcal{H}_t^{D}), \quad t \in (0,T).$$

从 R 中移除 D 可能会导致某些后续触点以及付费转化缺失。实际的剩余路径可能包含更少的事件发生次数，从而形成 $D \setminus R$的子集。这启发我们根据定理**2**（t）对N 执行稀疏化操作，而不是仅对单变量过程 N_1（t）执行稀疏化操作。

令 $i_{\min}(R) = \min\{1 \le i \le m : (t_i, e_i) \in R\}$ be 表示R中第一个事件的事件索引。回顾客户 A 和 B 的理想实验。现在假设，A 的 $D_A = D$ 和 B 的客户发起的事件类型过程 $\mathbf{N}_{\mathcal{E}_c}^{D_B}(t)$ 未知。首先令 $D_B = D \setminus R$。根据定理 2，如果我们从 D_B 中依次以概率 $1 - \lambda_{e_i}($ $t_i \mid \mathcal{H}_{t_i}^{D_B})/\lambda_{e_i}(t_i \mid \mathcal{H}_{t_i}^{D_A})$ 删除 $F_{t_{i^\star}}^{\mathcal{E}_c}(D_B)$ 中的触点 (t_i, e_i)，其中 $i > i_{\min}(R)$，那么得到的过程 $\mathbf{N}_{\mathcal{E}_c}^{D_B}(t)$ 将遵循模型 (1)。假设 $D \setminus R^\diamond$ 是对 D_B 进行稀疏化后得到的版本，其中 $R^\diamond \supseteq R$ 为实际移除集。图5展示了这一想法的一个简单示例。因此，给定路径 D 的一个子集 R，其总移除效应（TRE）定义为

$$\mathrm{att}_{t_{i^\star}}^{(\mathrm{total})}(R \mid D) := \mathbb{E}[\mathrm{att}_{t_{i^\star}}^{(\mathrm{direct})}(R^\diamond \mid D) \mid D] = \frac{\lambda_1(t_{i^\star} \mid \mathcal{H}_{t_{i^\star}}^{D}) - \mathbb{E}[\lambda_1(t_{i^\star} \mid \mathcal{H}_{t_{i^\star}}^{D\setminus R^\diamond}) \mid D]}{\lambda_1(t_{i^\star} \mid \mathcal{H}_{t_{i^\star}}^{D})},$$

其中，条件期望是相对于实际移除集而言的 $R^\diamond$，该移除集来自随机稀疏化操作。采用稀疏化操作的算法总结于算法1。

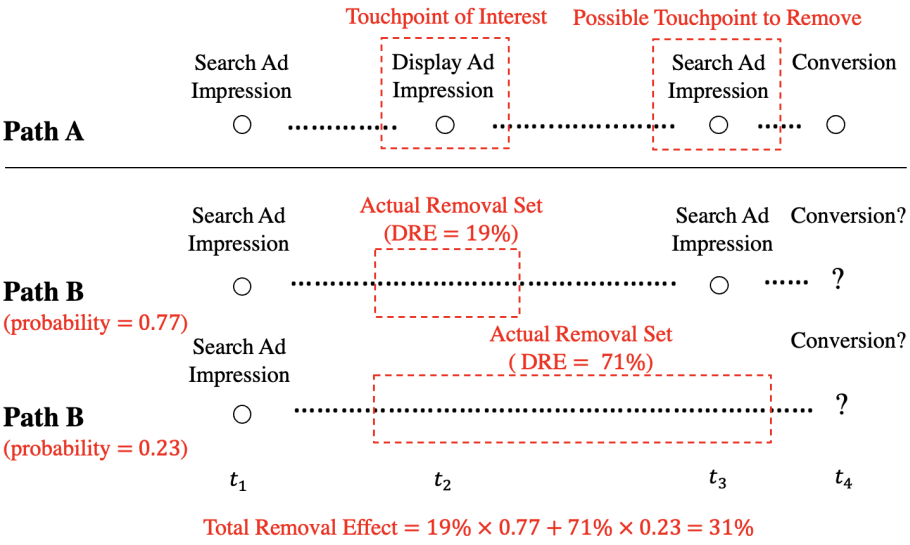

图5：对于包含三个触点的示例路径，展示广告曝光的总移除效应可视为直接移除效应的期望，其不确定性在于实际移除集。移除展示广告曝光可能导致后续搜索曝光也被移除。

算法1 总移除效应：稀疏化

输入：路径 D、指定的转化时间戳 $t_{i^\star}$ $(i^\star > 1)$ 和非空移除集 $R \subseteq F_{t^\star}(D)$。模型参数 $\boldsymbol{\mu}$、$\boldsymbol{\alpha}_1$、……、$\boldsymbol{\alpha}_q$ 以及一个较大的整数 L。对于 $\ell = 1$、……、L，执行$R^\diamond = R$。对于 $\{i > i_{\min}(R)$ 中按 t_i、e_i：$\in F^{\mathcal{E}_c}_{t_{i^\star}}(D \setminus R)\}$（升序）排列的 i：以概率 $1 - {}^{\lambda_{e_i}}(t_i|\mathcal{H}^{D\setminus R^\diamond}_{t_i})\,\lambda_{e_i}$ $(t_i|\mathcal{H}^{D}_{t_i})$ 更新 $R^\diamond = R^\diamond \cup \{(t_i$、$e_i)\}$。结束循环$x_\ell =$ 于 $^{(\text{direct})}_{t_{i^\star}}$ $(R^\diamond \mid D)$。结束循环 返回：$\text{att}^{(\text{total})}_{t_{i^\star}}(R \mid D) = \frac{1}{L}\sum_{\ell=1}^{L} x_\ell$。

$R^\diamond$ 的最大值为 $R \cup \{(t_i,\ e_i) \in F^{\mathcal{E}_c}_{t_{i^\star}}(D \setminus R) :\ i > i_{\min}(R)\}$，为简化记号，记为 Ω。利用稀疏化算法，我们可以推导出总移除效应的显式形式：

$$\text{att}^{(\text{total})}_{t_{i^\star}}(R \mid D) = \sum_{R \subseteq R' \subseteq \Omega} \text{att}^{(\text{direct})}_{t_{i^\star}}(R' \mid D)\mathbb{P}(R^\diamond = R' \mid D).$$

在稀疏化操作过程中，$R^{\diamond}$ 的值取决于对一列伯努利随机变量进行采样。令 $U_i \sim \text{Bernoulli}(1 - {}^{\lambda_{e_i}}\!/_{(t_i|\mathcal{H}_{t_i}^{D\setminus R'})}\,\lambda_{e_i}(t_i|\mathcal{H}_{t_i}^{D}))$ 表示用于稀疏化事件 $(t_i,\ e_i) \in \Omega$ 的伯努利随机变量。那么，$R^{\diamond}$ 的条件概率质量函数为

$$\begin{aligned}\mathbb{P}(R^{\diamond} = R' \mid D) &= \prod_{(t_i,e_i)\in\Omega\setminus R} \mathbb{P}(U_i = \mathbb{1}_{\{(t_i,e_i)\in R'\}}) \\ &= \prod_{(t_i,e_i)\in\Omega\setminus R} \left(1 - \frac{\lambda_{e_i}(t_i \mid \mathcal{H}_{t_i}^{D\setminus R'})}{\lambda_{e_i}(t_i \mid \mathcal{H}_{t_i}^{D})}\right)^{\mathbb{1}_{\{(t_i,e_i)\in R'\}}} \left(\frac{\lambda_{e_i}(t_i \mid \mathcal{H}_{t_i}^{D\setminus R'})}{\lambda_{e_i}(t_i \mid \mathcal{H}_{t_i}^{D})}\right)^{1-\mathbb{1}_{\{(t_i,e_i)\in R'\}}}.\end{aligned}$$

由于涉及组合子集计算，上述概率质量函数在实际应用中可能难以实现。在模型（1）下，我们可以改变求和顺序，根据（4）中直接移除效应的线性分解，推导出总移除效应。

$$\begin{aligned}\text{att}_{t_{i^\star}}^{(\text{total})}(R \mid D) &= \sum_{R\subseteq R'\subseteq\Omega} \text{att}_{t_{i^\star}}^{(\text{direct})}(R' \mid D)\mathbb{P}(R^{\diamond} = R' \mid D) \\ &= \sum_{R\subseteq R'\subseteq\Omega} \sum_{(t_i,e_i)\in R'} \text{att}_{t_{i^\star}}^{(\text{direct})}(\{(t_i,e_i)\} \mid D)\mathbb{P}(R^{\diamond} = R' \mid D) \\ &= \sum_{(t_i,e_i)\in\Omega} \text{att}_{t_{i^\star}}^{(\text{direct})}(\{(t_i,e_i)\} \mid D)\mathbb{P}((t_i,e_i) \in R^{\diamond} \mid D).\end{aligned}$$

这一结果意味着还可以采用另一种无需模拟且效率更高的迭代算法。其基本思路是重新分配直接移除效应所得到的分数。该算法采用直观的反向传播评分方法，如算法2所示。由 Rumelhart 等人 (1986年)提出的反向传播，近年来在人工神经网络中日益普及。

算法 2 总移除效应：反向传播

输入：路径 D、指定的转化时间戳 $t_{i^\star}$ $(i^\star > 1)$ 和非空移除集 $R \subseteq F_{t^\star}(D)$。模型参数 $\boldsymbol{\mu}, \boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_q$。初始化：$y_i = \text{att}_{t_{i^\star}}^{(\text{direct})}(\{(t_i, e_i)\} \mid D)$，对于 $i_{\min}(R) \le i \le i^\star - 1$。对于 i 中的 $\{i > i_{\min}(R)$：按 t_i, e_i 的降序排列 $\in F_{t_{i^\star}}^{\mathcal{E}_c}(D \setminus R)\}$，然后对于 i' 中的 $\{1 \le i' \le i-1 :\ (e_{i'} \to e_i) \in \mathcal{A}\}$ 执行 $y_{i'} = y_{i'} + y_i \cdot \left(1 - {}^{\lambda_{e_i}(t_i|\mathcal{H}_{t_i}^{D\setminus\{(t_{i'},e_{i'})\}})}\!/_{\lambda_{e_i}(t_i|\mathcal{H}_{t_i}^{D})}\right)$。结束循环。结束循环。返回：$\text{att}_{t_{i^\star}}^{(\text{total})}(R \mid D) = \sum_{(t_i,e_i)\in R} y_i$。

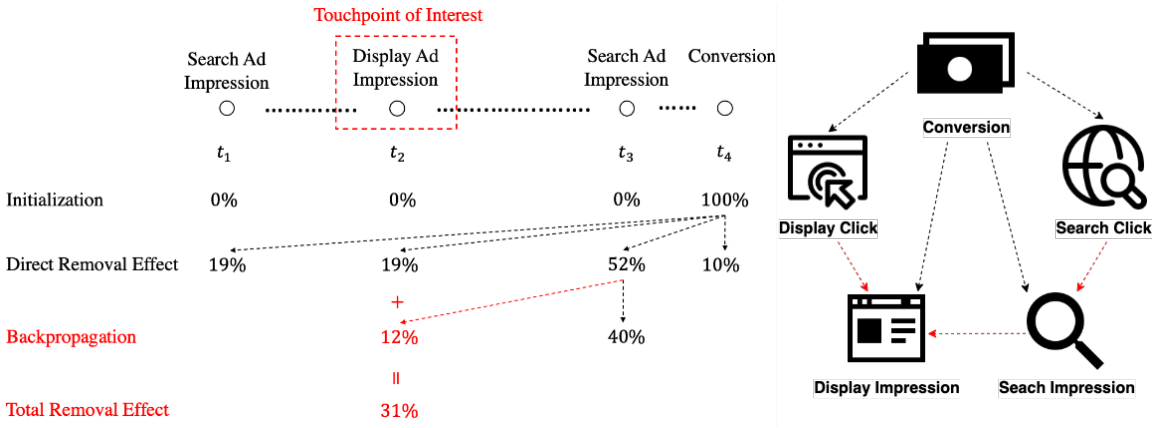

图 6：对于包含三个触点的示例路径，可以通过分数反向传播计算展示广告曝光的总移除效应。它由两部分组成：其自身的直接移除效应，以及它在后续搜索广告曝光的直接移除效应中所占的份额。右侧的图描述了分数在事件类型之间的流动情况，其中黑色和红色箭头分别表示直接移除效应归因和分数反向传播。

如图 6中的示例路径所示，分数并不是沿路径向前流动，而是从最后一个触点向之前的触点反向流动。这种流动遵循格兰杰因果图的逆图，其箭头从某一事件类型指向其格兰杰因果父节点。

3.3 关于两种归因评分方法的说明

在本小节中，我们讨论所提出的两种归因评分方法及其在一种极端情况下的实现之间的差异。

直接移除效应将每个触点的贡献视为独立成分。而指定触点子集的总移除效应，则是其沿路径移除所产生的累积影响。直接移除效应将转化事件视为唯一响应，而客户发起的事件类型的完整向量/子过程则是总移除效应的响应。直观地说，在图 6所示的分数流图中，直接移除效应使分数直接从转化节点流出，而总移除效应则基于分数的持续流动，直到其到达指定节点或没有父节点的节点。

这两个分数在可加性方面有所不同。假设 R_1 和 R_2 是截断路径 $F_{t^\star}(D)$ 的两个互斥子集。模型 (1) 下的直接移除效应

具有定义所蕴含的可加性：

$$\mathrm{att}_{t_{i^\star}}^{(\mathrm{direct})}(R_1 \cup R_2 \mid D) = \mathrm{att}_{t_{i^\star}}^{(\mathrm{direct})}(R_1 \mid D) + \mathrm{att}_{t_{i^\star}}^{(\mathrm{direct})}(R_2 \mid D).$$

然而，对于总移除效应，在反向传播过程中，只要两个集合的事件类型之间存在格兰杰因果关系，我们就允许分数在这两个集合之间流动。因此，总移除效应具有次可加性，因而不是增量性的

$$\mathrm{att}_{t_{i^\star}}^{(\mathrm{total})}(R_1 \cup R_2 \mid D) \leq \mathrm{att}_{t_{i^\star}}^{(\mathrm{total})}(R_1 \mid D) + \mathrm{att}_{t_{i^\star}}^{(\mathrm{total})}(R_2 \mid D).$$

因此，要获得某个渠道的总移除效应，我们应将移除集设为所有相应的单个触点，而不是对触点级评分求和。

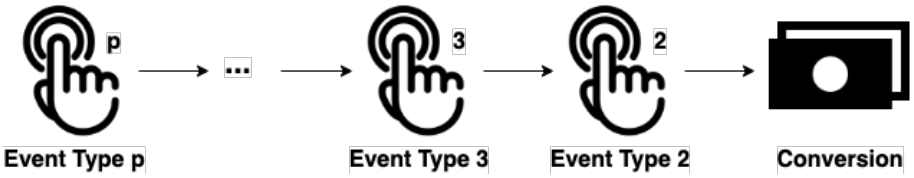

图 7：格兰杰因果图呈线形的极端情形。

我们考虑一个极端情形，如图 7中的格兰杰因果图所示。在实践中，我们认为每个渠道中的触点分别对转化具有格兰杰因果关系，因此，转化节点在格兰杰因果图中有多个父节点，如图 2所示。但在这里，我们用这个例子来论证所提出的两种归因分数。在图中，转化是标记为 1 的事件类型，其他所有事件类型 2、.....、p都是触点，其中标记为 p 的触点类型是唯一一种企业发起的事件类型。假设任何客户发起的事件类型都没有基线强度，即 $\mu_e = 0$对于 $e = 1$、.....、$p - 1$。那么，类型为（$e + 1$）的事件将是类型为 e 的事件发生的唯一原因。假设我们观察到路径 D，其模式与图 7相同。也就是说，

$$D = \{(t_1, p), \ldots, (t_{p-2}, 3), (t_{p-1}, 2), (t_p, 1)\}.$$

如果我们考察每个触点的直接移除效应，就会得到

$$\mathrm{att}_{t_p}^{(\mathrm{direct})}(\{(t_i, e_i)\} \mid D) = \begin{cases} 100\%, & i = p - 1; \\ 0\%, & i = 1, \ldots, p - 2, \end{cases}$$

其中 $e_i = p - i + 1$ 对于 $i = 1$，.....，$p - 1$。直接移除效应忽略了早期阶段触点的重要性，因为它们无法立即触发转化。另一方面，

总移除效应给出

$$\mathrm{att}_{t_p}^{(\mathrm{total})}(\{(t_i, e_i)\} \mid D) = 100\%, \quad i = 1, \ldots, p-1.$$

总移除效应似乎会过度分配，因为路径上只有 1 次转化 D。这一问题也曾由Singal 等人（2022）指出，认为这是基于移除效应的归因方法的一个缺点。然而，这并不意味着这些得分是错误的。这种极端的过度分配问题表明，为每个触点计算此类归因得分可能并不合适。在图7中，事件类型 2 至 p 很可能都属于同一个渠道。一个详细的例子是：电子邮件发送、电子邮件打开和邮件点击是触点，而转化只能由邮件点击触发（通过电子邮件链接进行的仅限受邀购买）。在这种情况下，我们可以使用移除集 $R = \{（t_1，\ p）\ldots\ldots、\ （t_{p-2}，\ 3），\ （t_{p-1}，\ 2）\}$将事件类型 2 至视为一个整体。于是我们得到

$$\mathrm{att}_{t_p}^{(\mathrm{direct})}(R \mid D) = \mathrm{att}_{t_p}^{(\mathrm{total})}(R \mid D) = 100\%.$$

如果事件类型 2 至 p 实际上来自不同的渠道，这意味着该图不太可能真实，或者由于这些渠道高度依赖，因此没有必要将它们彼此区分。

总而言之，直接移除效应是一种解释性得分，它根据各触点的增量影响，将转化事件的功劳分配给每个触点。这两种所提出的方法都能给出渠道级分数，但总移除效应为转化提供了边际视角。换言之，当所关注的是转化的留一渠道损失时，这种方法更为适用。

4 估计

本节中，我们为所提出的模型构建一个正则化估计量。我们设计了一种计算高效的交替方向乘子法（ADMM）算法，用于求解相应的优化问题。我们还推导了渐近情形下的图模型选择一致性，以及模型估计值和归因分数的收敛速率。

为便于阐述，我们引入一些必要的记号。对于任意一个 $K \geq 1$ 为任意整数的 K 维向量 $\boldsymbol{x} = (x_1，\ \ldots\ldots，\ x_K)^\top \in \mathbb{R}^K$，令 $\|\boldsymbol{x}\|_1 := \sum_{k=1}^K |x_k|$、$\|\boldsymbol{x}\|_2 := \sqrt{\sum_{k=1}^K x_k^2}$ 和 $\|\boldsymbol{x}\|_\infty := \max_{1 \leq k \leq K} |x_k|$ 分别表示其 L_1-范数、L_2-范数和 L_∞-范数。令 $\boldsymbol{x}_+$ 表示 $\boldsymbol{x}$ 的非负部分，其定义为 $\boldsymbol{x}_+ :=$

$(\max(0,\ x_1),\ \ldots\ldots,\ \max(0,\ x_K))^\top$。对于矩阵 $M=(M_{ij})_{i,\ j}$，矩阵 L_1-范数定义为 $\|M\|_{1,\ \infty}:=\max_i(\sum_j|M_{ij}|)$。

4.1 模型学习

假设总共有 n 条路径，$D_1,\ldots,D_n$，其中 D_j 是时间间隔 $[0,T_j]$ 上的第 j 条路径。假设核函数 $\psi_{e'e}(\cdot)$ $e'=1,\ldots,p,\ e=1,\ldots,q$, 已知。令 $\boldsymbol{\mu}=(\mu_1,\ldots,\mu_q)^\top\in\mathbb{R}^q_{\geq 0}$ 为 $\mathbf{N}\mathcal{E}_c(t)$ 的基准强度向量，$\boldsymbol{\alpha}_e=(\alpha_{1e},\ldots,\alpha_{pe})^\top\in\mathbb{R}^p_{\geq 0}$ 为 A 中 $e=1,\ldots,q$ 的第 e 列。模型学习转化为一个优化问题：

$$\min_{\boldsymbol{\mu},\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_q\in\mathbb{R}^p_{\geq 0}}\frac{1}{n}\sum_{j=1}^{n}\Phi(D_j;\boldsymbol{\mu},\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_q),$$

其中，$\Phi(D_j;\ \boldsymbol{\mu},\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_q)$ 是给定第 j 条路径 D_j 时的损失函数。考虑一个最小二乘泛函：

$$\Phi(D_j;\boldsymbol{\mu},\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_q)=\sum_{e=1}^{q}\left\{\frac{1}{2}\int_0^{T_j}[\lambda_e(t\mid\mathcal{H}_t^{D_j})]^2dt-\int_0^{T_j}\lambda_e(t\mid\mathcal{H}_t^{D_j})dN_e^{D_j}(t)\right\}.$$

与负对数似然相比，最小二乘泛函具有更高的计算效率。Lin & Ying 提出了其等价形式，用于估计加性风险模型（Lin & Ying（1994年）。Hansen 等人也（2015）以及Bacry 等人（2020年）将其用于点过程模型。

假设边集是稀疏的，我们对系数施加如下稀疏性约束：

$$\min_{\boldsymbol{\mu},\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_q\in\mathbb{R}^p_{\geq 0}}\frac{1}{n}\sum_{j=1}^{n}\Phi(D_j;\boldsymbol{\mu},\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_q)+\sum_{e=1}^{q}\gamma_e\|\boldsymbol{\alpha}_e\|_1,\tag{5}$$

其中，γ_1、……、$\gamma_q\geq 0$ 是用于控制系数向量 $\boldsymbol{\alpha}_1$、……、$\boldsymbol{\alpha}_q$各自稀疏性的正则化参数。值得指出的是，$\{\mu_e,\boldsymbol{\alpha}_e\}_{e=1}^q$ 在目标函数（5）中是可分离的。因此，该优化问题可以分解为逐节点模型学习。对于每个节点 $e\in\mathcal{E}_c$，学习其父节点可得

$$\min_{\mu_e\geq 0,\boldsymbol{\alpha}_e\in\mathbb{R}^p_{\geq 0}}\frac{1}{n}\sum_{j=1}^{n}\phi_e(D_j;\mu_e,\boldsymbol{\alpha}_e)+\gamma_e\|\boldsymbol{\alpha}_e\|_1,\tag{6}$$

其中

$$\phi_e(D_j;\mu_e,\boldsymbol{\alpha}_e)=\frac{1}{2}\int_0^{T_j}[\lambda_e(t\mid\mathcal{H}_t^{D_j})]^2dt-\int_0^{T_j}\lambda_e(t\mid\mathcal{H}_t^{D_j})dN_e^{D_j}(t).$$

在下文中，我们记 $\boldsymbol{\theta}=(\theta_0,\ \theta_1,\ \ldots,\ \theta_p)^\top=(\mu_e,\ \boldsymbol{\alpha}_e^\top)^\top\in\mathbb{R}^{p+1}_{\geq 0}$。令 $\boldsymbol{X}_j(t)=(X_{j,\,0}(t)\ X_{j,\,1}(t)\ \ldots,\ X_{j,\,p}(t))^\top$，其中

$$X_{j,0}(t)=1,\quad X_{j,k}(t)=\int_0^t \psi_{ke}(t-u)dN_k^{D_j}(u),\ k=1,\ldots,p.$$

现在，条件强度可以写成 $\lambda_e(t\mid\mathcal{H}_t^{D_j})=\boldsymbol{\theta}^\top\boldsymbol{X}_j(t)$。令 $V=(V_{kk'})_{k,\,k'=0}^{p}\in\mathbb{R}^{(p+1)\times(p+1)}$ 以及 $\boldsymbol{b}=(b_0,\ \cdots,\ b_p)^\top\in\mathbb{R}^{p+1}$，其中对于 $k=0,\ \ldots\ldots,\ p$,

$$V_{kk'}=\frac{1}{n}\sum_{j=1}^{n}\int_0^{T_j}X_{j,k}(t)X_{j,k'}(t)dt,\quad b_k=\frac{1}{n}\sum_{j=1}^{n}\int_0^{T_j}X_{j,k}(t)dN_e^{D_j}(t).$$

那么，正则化解满足

$$\hat{\boldsymbol{\theta}}=\underset{\boldsymbol{\theta}\in\mathbb{R}^{p+1}_{\geq 0}}{\arg\min}\ \frac{1}{2}\boldsymbol{\theta}^\top V\boldsymbol{\theta}-\boldsymbol{b}^\top\boldsymbol{\theta}+\gamma_e\|\boldsymbol{\alpha}_e\|_1. \tag{7}$$

上述问题等价于一个线性约束问题

$$\min_{\boldsymbol{\theta}\in\mathbb{R}^{p+1}_{\geq 0},\boldsymbol{\alpha}_e=\boldsymbol{\alpha}_e'}\ \frac{1}{2}\boldsymbol{\theta}^\top V\boldsymbol{\theta}-\boldsymbol{b}^\top\boldsymbol{\theta}+\gamma_e\|\boldsymbol{\alpha}_e'\|_1.$$

应用交替方向乘子法（ADMM），其对应的增广拉格朗日函数为

$$\mathcal{L}_\eta(\boldsymbol{\theta},\boldsymbol{\alpha}_e',\boldsymbol{\omega})=\frac{1}{2}\boldsymbol{\theta}^\top V\boldsymbol{\theta}-\boldsymbol{b}^\top\boldsymbol{\theta}+\gamma_e\|\boldsymbol{\alpha}_e'\|_1+\boldsymbol{\omega}^\top(\boldsymbol{\alpha}_e-\boldsymbol{\alpha}_e')+\frac{1}{2}\eta\|\boldsymbol{\alpha}_e-\boldsymbol{\alpha}_e'\|_2^2.$$

学习算法如算法3所示。

算法3 基于交替方向乘子法的图点过程学习

输入：路径 D_1、……、D_n 以及正则化参数 γ_e。

预计算：V，$\boldsymbol{b}$。

初始化：设置 $\eta>0$，并为 μ_e、$\boldsymbol{\alpha}_e$、$\boldsymbol{\alpha}_e'$ 和 $\boldsymbol{\omega}$ 设置适当的初始值。

当未收敛时**执行**

$$\begin{pmatrix}\mu_e\\ \boldsymbol{\alpha}_e\end{pmatrix}=\left[\left(V+\begin{pmatrix}0 & \\ & \eta I_p\end{pmatrix}\right)^{-1}\left(\boldsymbol{b}+\begin{pmatrix}0\\ \eta\boldsymbol{\alpha}_e'-\boldsymbol{\omega}\end{pmatrix}\right)\right]_+.$$

$\boldsymbol{\alpha}_e'=(\boldsymbol{\alpha}_e+\eta^{-1}\boldsymbol{\omega}-\eta^{-1}\gamma_e\mathbb{1}_p)_+$.

$\boldsymbol{\omega}=\boldsymbol{\omega}+\eta(\boldsymbol{\alpha}_e-\boldsymbol{\alpha}_e')$.

结束循环

返回：$\boldsymbol{\theta}=(\mu_e,\ \boldsymbol{\alpha}_e^\top)^\top$。

4.2 渐近性质

在本小节中，我们推导所提出估计量的收敛速率。模型中由客户发起的部分（1）与Hawkes过程（Hawkes 1971）相似。但与典型的多元霍克斯过程不同，模型（1）并不保证是平稳点过程。因此，我们在样本量 $n \to \infty$趋于无穷的假设下推导渐近结果，而非令观测长度 $T \to \infty$ 趋于无穷（Guo et al. 2018、Yu et al.2020）。在生存分析领域，Lin& Ying （1994年）研究了加性风险模型，而 Lin &Lv （2013年）证明了相应的 L_1-正则化估计量的一致性。结合这些研究，我们将证明，在包括不可表示性在内的若干条件下，所提出的估计量在经典固定 p 设定和稀疏高维设定下均具有一致性；在稀疏高维设定中，log(p)与样本量 n相当。

假设1。 （独立同分布）客户发起的事件类型过程 $\mathbf{N}_{\mathcal{E}_c}^{D_j}$ （t）、 j =1、.....、 n 相互独立，并服从模型 （1）。

估计量的收敛性质依赖于观测值的同分布性。我们不需要假设外部（企业发起的）事件在不同路径之间 独立同分布。 它们可以因路径而异，但模型（1）应对每个个体均成立。

假设2。 （有界输入）存在常数 T 和 X ，使得 $T_j < T$ 和$\sup_{t\in(0,T_j)} \left\|\boldsymbol{X}_j\ (t)\right\|_\infty < X$ *a.s.* 对于 $j = 1$、.....、n均成立。

定义活跃集 $S = \{0\} \cup \{e' \in \mathcal{E}:\ \alpha_{e'e} > 0\}$ 及其补集 $S^c = \{e' \in \mathcal{E}:\ \alpha_{e'e} = 0\}$。我们使用 $s = \mathrm{Card}(S)$ 表示活跃集 S的基数。令 $G = (G_{kk'})_{k,\ k'=0}^{p} = \mathbb{E}V$为 V的总体版本。假设子矩阵 $G_{SS} = (G_{kk'})_{k,\ k'\in S}$是非奇异的，并定义 $\kappa := \left\|G_{SS}^{-1}\right\|_{1,\ \infty}$。

假设3。 （不可表示性）存在常数 $\xi \in (0,\ 1)$ ，使得

$$\left\|G_{S^cS}G_{SS}^{-1}\right\|_{1,\infty} < 1 - \xi.$$

该条件改编自Lin 和 Lv(2013年) 中的条件3，它是Wainwright(2009) 中线性回归 LASSO 条件（15）的一般化。

现在我们建立正则化估计量的收敛速率和模型选择一致性。为简化记号，令 $\boldsymbol{\theta}$ 表示真实值。我们考虑两种情形：事件类型的数量 p 固定，或 p 趋于无穷，同时活跃集 S 是稀疏的，即 s 有界于 p。

定理 3. 在假设 *1-3* 下，存在 $C_1 > 0$ 和 $C_2 > 0$，使得正则化ˆ估计量 $\boldsymbol{\theta}$ 在 (7)中满足以下性质：

(i) 若 p 固定，则对于任意常数 $0 < \nu < 1$，当适当选择 γ_e，且 n 足够大时，以至少 $1 - 2(p+1)(p+2)\exp(-n^{\nu})$ 的概率，

- *(Edge selection)* $\hat{\boldsymbol{\theta}}_{S^c} = 0$
- *(L_∞-error)* $\left\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right\|_\infty \leq 10\xi^{-1}\kappa C_1^{-\frac{1}{2}} C_2 n^{-\frac{1-\nu}{2}}$;

(ii) 若 p 发散，则对于任意常数 $\zeta > 2$，当适当选择 γ_e，且 n 足够大时，以至少 $1 - 3/(p+1)^{\zeta-2}$ 的概率，

- *(Edge selection)* $\hat{\boldsymbol{\theta}}_{S^c} = 0$
- *(L_∞-error)* $\left\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right\|_\infty \leq 10\xi^{-1}\kappa C_1^{-\frac{1}{2}} C_2 \sqrt{\frac{\zeta \log(p+1)}{n}}$.

在上述定理中，为便于表述，我们未具体说明 γ_e 和 n 的取值。详细取值见补充材料。

接下来，我们讨论归因分数的收敛速率。对于在 $t = t_{i^\star}$ 处发生转化的正路径 D，我们希望分析子集 $R \subseteq F_{t^\star}(D)$ 的直接移除效应。在实践中，对于方程（2），我们只能得到条件强度的估计值。因此，对于模型（1），我们使用如下给出的估计直接移除效应

$$\widehat{\mathrm{att}}^{(\mathrm{direct})}_{t_{i^\star}}(R \mid D) := \frac{\sum_{(t_i, e_i) \in R} \hat{\alpha}_{e_i 1} \psi_{e_i 1}(t_{i^\star} - t_i)}{\hat{\mu}_1 + \sum_{i < i^\star} \hat{\alpha}_{e_i 1} \psi_{e_i 1}(t_{i^\star} - t_i)}.$$

令 $r = \mathsf{Card}(R)$ 表示移除集 R 的基数。对于核函数，假设存在 $\psi_1 > 0$，使得 $\max_{1 \leq e \leq p} \sup_{t>0} |\psi_{e1}(t)| < \psi_1$。

定理 4。 给定一条在 $t = t_{i^\star}$ 处发生转化的路径 D，$i^\star > 1$，在上述关于 $e = 1$ 的假设下，存在一个依赖于 ψ_1 和 D 的 $C_3 > 0$，使得对移除集 $R \subseteq F_{t^\star}(D)$ 估计得到的直接移除效应满足以下任一条件：

(i) 若 p 固定且 n 充分大，则对于任意常数 $0 < \nu < 1$，以至少 $1 - 2(p+1)(p+2)\exp(-n^{\nu})$ 的概率，

$$\left|\widehat{\mathrm{att}}^{(\mathrm{direct})}_{t_{i^\star}}(R \mid D) - \mathrm{att}^{(\mathrm{direct})}_{t_{i^\star}}(R \mid D)\right| \leq 10\xi^{-1}\kappa C_1^{-\frac{1}{2}} C_2 C_3 r n^{-\frac{1-\nu}{2}};$$

(ii) 若 p 发散且 n 足够大，则对于任意常数 $\zeta > 2$，以至少 $1 - 3/(p+1)^{\zeta-2}$ 的概率，

$$\left|\widehat{\mathrm{att}}^{(\mathrm{direct})}_{t_{i^\star}}(R \mid D) - \mathrm{att}^{(\mathrm{direct})}_{t_{i^\star}}(R \mid D)\right| \leq 10\xi^{-1}\kappa C_1^{-\frac{1}{2}} C_2 C_3 r \sqrt{\frac{\zeta \log(p+1)}{n}}.$$

常数 ξ、 κ、 C_1和 C_2 来自定理 3 ，对应于 $e = 1$。基于对估计的直接移除效应的分析，我们进一步给出了总移除效应估计的误差界。我们将估计的总移除效应称为

$$\widehat{\mathrm{att}}^{(\mathrm{total})}_{t_{i^\star}}(R \mid D) := \mathbb{E}[\widehat{\mathrm{att}}^{(\mathrm{direct})}_{t_{i^\star}}(\widehat{R^\diamond} \mid D) \mid D],$$

其中 $\widehat{\ }$ $R^\diamond \supseteq R$ 是使用估计的稀疏化概率，通过算法 1中的稀疏化操作得到的实际移除集。此外，对于 ψ_1，对于每个 $e = 2, \ldots, \overline{q}$，假设存在$\psi_e > 0$ 使得 $\max_{1 \le e' \le p} \sup_{t>0} |\psi_{e'e}(t)| < \psi_e$。

定理 5. 给定一条在 $t = t_{i^\star}$处发生转化的路径 D 、 $i^\star > 1$ 以及移除集$R \subseteq F_{t^\star}(D)$，在上述关于 $e = 1$、…、 q 的假设下，存在一个依赖于ψ_1、…、 ψ_q、 D和 R 的 $C_4 > 0$ ，使得 R 的估计总移除效应满足以下条件之一：

(i) 若 p 固定且 n 足够大，则对于任意常数 $0 < \nu < 1$，以至少 $1 - 2p(p+1)(p+2)\exp(-n^\nu)$的概率，

$$\left|\widehat{\mathrm{att}}^{(\mathrm{total})}_{t_{i^\star}}(R \mid D) - \mathrm{att}^{(\mathrm{total})}_{t_{i^\star}}(R \mid D)\right| \le C_4 n^{-\frac{1-\nu}{2}};$$

(ii) 如果 p 发散且 n 足够大，那么对于任意常数 $\zeta > 3$，其概率至少为 $1 - 3/(p+1)^{\zeta-3}$，

$$\left|\widehat{\mathrm{att}}^{(\mathrm{total})}_{t_{i^\star}}(R \mid D) - \mathrm{att}^{(\mathrm{total})}_{t_{i^\star}}(R \mid D)\right| \le C_4 \sqrt{\frac{\zeta \log(p+1)}{n}}.$$

5 模拟研究

本节中，我们进行两项模拟实验，以考察所提出的图形归因方法的性能。第一部分中，我们使用从多元霍克斯过程中模拟生成的数据，验证所提出的图形归因方法。第二部分中，我们使用根据谷歌公司开发的数字广告系统模拟（DASS）的改进版本模拟生成的数据，将所提出的方法与常用的归因模型进行比较。模拟数据包括在线客户浏览行为，以及人为注入的、会影响该客户行为的广告事件。

我们首先介绍归因方法中的一些渠道级指标。一般来说，假设有 Z 个渠道，分别标记为 $z = 1$、……、 Z。受此前文献的启发 Anderl

等(2016年)、Li & Kannan (2014年，我们计算每个渠道的渠道级转化计数占比（CCC 占比） z，即

$$\mathsf{p}_z := \frac{\mathrm{CCC}_z}{\sum_{z=1}^{Z} \mathrm{CCC}_z} \quad \text{with } \mathrm{CCC}_z = \sum_{j=1}^{n} N_1^{D_j}((0, T_j]) - \sum_{j=1}^{n} N_1^{D_j^{(z\text{-off})}}((0, T_j]),$$

其中， $D_j^{(z\text{-off})}$ 是一个路径，其分布与 D_j 相同，但不包含渠道 z的任何触点。CCC_z 被解释为由渠道 z带来的转化次数。CCC_z 的第一项是 n 路径中的转化次数，第二项则是在关闭渠道 z 时 n 路径中的转化次数。要在合成数据中获得这一数值，我们可以禁用所有相关的触点类型，并使用相同的随机种子再次运行模拟器。然后，总转化次数的减少量就是所需的值。

令 $\mathcal{C}_z \subseteq \mathcal{E}$ 表示属于渠道 z的触点类型集合，其中 $z = 1, \ldots, Z$。那么，对于路径 D 中位置 t 处的转化，相应的移除集为$F_t^{\mathcal{C}_z}(D) = \{(t_i, e_i) \in D \colon t_i < t,\ e_i \in \mathcal{C}_z\}$。我们使用渠道级聚合得分占比（CAS 占比）来估计渠道 z的 CCC 占比，即

$$\hat{\mathsf{p}}_z := \frac{\mathrm{CAS}_z}{\sum_{z=1}^{Z} \mathrm{CAS}_z} \quad \text{with } \mathrm{CAS}_z = \sum_{j=1}^{n} \sum_{i \le m_j \colon e_i^j = 1} \widehat{\mathrm{att}}_{t_i^j}(F_{t_i^j}^{\mathcal{C}_z}(D_j) \mid D_j).$$

回顾一下，归因得分是相对于转化而言的增量成分或边际损失。其聚合版本 CAS 表示各渠道的转化次数相对于转化总次数的整体下降，因此可用于估计 CCC。令 $\mathbf{p} = (\mathsf{p}_1, \ldots, \mathsf{p}_Z)^\top$表示 CCC 比例向量，$\hat{\mathbf{p}} = (\hat{\mathsf{p}}_1, \ldots, \hat{\mathsf{p}}_Z)^\top$表示 CAS 比例向量。我们采用 KL 散度 $D_{\mathrm{KL}}(\cdot \,\|\, \cdot)$ 和海灵格距离 $H(\cdot, \cdot)$ 来衡量估计精度，其中

$$D_{\mathrm{KL}}(\mathbf{p} \,\|\, \hat{\mathbf{p}}) := \sum_{z=1}^{Z} \mathsf{p}_z \log\left(\frac{\mathsf{p}_z}{\hat{\mathsf{p}}_z}\right)$$

$$H(\mathbf{p}, \hat{\mathbf{p}}) := \sqrt{\frac{1}{2} \sum_{z=1}^{Z} \left(\sqrt{\hat{\mathsf{p}}_z} - \sqrt{\mathsf{p}_z}\right)^2}.$$

5.1 基于 HawkesProcess 的模拟

在本小节中，我们使用从多元霍克斯过程（Hawkes 1971）模拟生成的数据集验证我们的归因方法。

From \ To	展示广告点击	搜索曝光	搜索点击	转化
展示广告曝光	0.08	0.08	0	0.01
展示广告点击	0	0	0	0.08
搜索曝光	0	0	0.08	0.02
搜索点击	0	0	0	0.1
转化	0	0	0	0

表1：模拟数据的格兰杰因果系数。

根据方程（1），如果对于每个 N_e，t）都是泊松过程，那么我们的模型就会简化为多元霍克斯过程。换言之，多元霍克斯过程是我们模型的一个特例。因此，我们按照图2，使用多元霍克斯过程模拟一个数据集，其中包含 $Z = 2$ 个渠道（展示广告和搜索）以及 4 种触点类型：展示广告曝光、展示广告点击、搜索曝光和搜索点击。展示广告曝光被视为企业发起的事件类型，其遵循速率为 0.02 的泊松过程分布。共模拟 $n = 10,000$ 条路径，每条路径包含 $T_j = 365$ 天，适用于 $j = 1$、.....、n。我们对每一对可能相连的节点取 $\psi_{e'e} = \frac{1}{10}\exp(-\frac{t}{10}) \cdot \mathbb{1}_{\{t>0\}}$。搜索曝光和转化的基准强度分别设为 0.02 和 1×10^{-4}，两种点击触点类型的基准强度均设为零。

渠道：p_z	TRE	DRE
展示广告：0.3799	0.3782 (0.0104)	0.3491 (0.0112)
搜索：0.6201	0.6218 (0.0104)	0.6509 (0.0112)
KL散度	0.0002 (0.0002)	0.0023 (0.0015)
海灵格距离	0.0064 (0.0042)	0.0226 (0.0083)

表2：总移除效应与直接移除效应中 CAS 所占比例的比较。报告的数值为 100 次独立运行的平均值，括号内为标准误。第一列列出渠道级转化计数（CCC）占真实值的比例。

格兰杰因果系数见表 1。模拟结果汇总于表 2，这些结果是基于 100 次独立运行计算得出的。如表 2所示，我们的图形化TRE归因方法考虑了不同类型触点之间的格兰杰因果关系，能够准确估计搜索渠道和展示渠道的真实移除效应。TRE计算得出的两个渠道的CAS占比都非常

接近真实值。此外，TRE的KL散度和海灵格距离都非常小。这并不意外，因为定理 2 保证 $D_j \setminus (F^{\mathcal{C}_z}_{T_j}(D_j))^{\diamond}$ 和 $D_j^{(z\text{-off})}$ 服从相同分布。相比之下，图形化DRE方法往往会低估展示渠道的贡献，因为它忽略了展示广告曝光对搜索曝光的激发效应。

5.2 基于 DASS 的模拟

接下来，我们将图形归因方法与现有方法进行比较，包括 DNAMTA（Li 等人，2018年）、逻辑回归、马尔可夫模型（Anderl 等人，2016年），以及基于规则的方法，包括末次触点、首次触点、线性、时间衰减和 U 形。表 3列出了所比较模型的说明。其中，马尔可夫模型直接提供渠道级分数，其他模型则提供可汇总至渠道层级的路径级得分。

方法	Type	评分说明
TRE	数据驱动	图形点过程模型的总移除效应。
DRE	数据驱动	图形点过程模型的直接移除效应。
DNAMTA	数据驱动	由深度神经网络的转化概率推导出的增量评分 Li 等人提出的注意力多触点归因模型于 (2018年)。
逻辑回归	数据驱动	根据逻辑回归的转化概率得出的增量评分。
马尔可夫	数据驱动	Anderl 等人提出的马尔可夫模型的移除效应Anderl 等人(2016年)。
Last	基于规则	末次触点归因，将所有功劳分配给最接近转化的触点。
第一	基于规则	首次触点归因，将所有功劳分配给路径上的初始触点。
线性	基于规则	线性归因，即为转化前的每个触点分配相同的功劳。
衰减	基于规则	时间衰减归因，即距离转化越近的触点获得的 功劳越多，距离转化时间越远的触点获得的功劳则越少。
U形	基于规则	U形归因，将40%的功劳分别分配给第一个触点和 最后一个触点，其余触点均分剩余的20%。

表3：对比中的归因方法说明。

为了比较模型，我们使用数字广告系统模拟器（DASS）的一个修改版本来模拟数据。DASS（Sapp 等，2016）由谷歌公司开发，是业内广受欢迎的归因模拟器，其有效性也得到了从业者的普遍认可。“它生成可供观察性模型使用的数据，也支持针对模拟客户开展虚拟实验，以衡量营销的实际增量价值，从而进行直接比较”（Sapp 等，2016）。我们对 DASS 进行了修改，使其具备两项所需功能，并将其称为 DASS+。DASS 模拟每位

客户浏览状态之间的转移，不会明确考虑时间戳；而 DASS+使用一个转移矩阵，反映每分钟的这些浏览状态。另一方面，DASS 并未明确限制可投放的广告数量。使用 DASS+时，广告数量会限制在固定数目，并可在整个模拟时间范围内按预先确定的分布进行投放。

该合成数据涉及 $Z = 4$ 渠道：电子邮件、展示广告、搜索和社交，包含 9 种触点：邮件发送、电子邮件打开、邮件点击、展示广告曝光、展示广告点击、搜索曝光、搜索点击、社交曝光和社交点击。开启所有渠道后，我们从 100，000 名客户中获得了 $n = 98,986$ 条有效路径。其中，62，287 条为正向路径，36，699 条为负向路径。每条路径的模拟周期为 90 天。进行模型学习时，我们将第一个事件的时间戳用作起始时间（$t_1 = 0$），并将路径上最后一个事件的时间戳用作终止时间。

图 8 展示了我们提出的模型针对模拟数据学习得到的格兰杰因果图。大多数类型的触点都对转化产生激发效应。与第一项模拟研究类似，即使存在更多渠道，我们仍确认了每个渠道的渠道内延续效应。此外，我们还确认了其他渠道对搜索曝光的渠道间溢出效应。

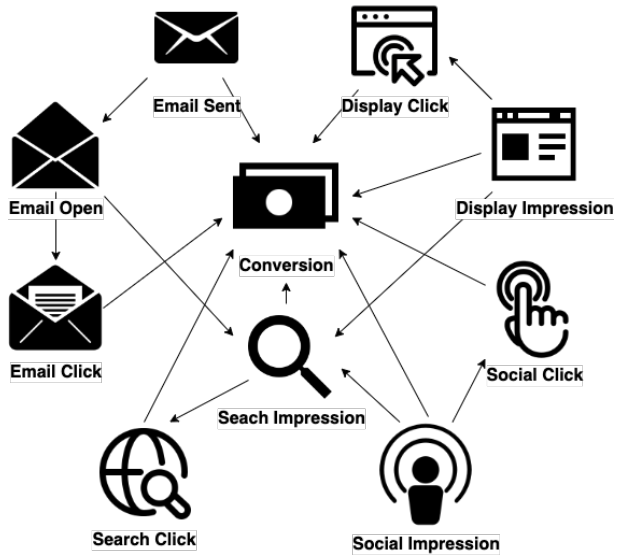

图8：模拟数据的格兰杰因果图。

渠道：p_z	TRE	DRE	DNAMTA	逻辑回归	马尔可夫	Last	第一	线性	衰减	U形
展示广告:0.207	0.185 (0.006)	0.137 (0.006)	0.172 (0.027)	0.275 (0.004)	0.221 (0.000)	0.371 (0.002)	0.397 (0.003)	0.380 (0.001)	0.372 (0.001)	0.383 (0.001)
电子邮件：0.171	0.139 (0.008)	0.101 (0.007)	0.130 (0.019)	0.217 (0.004)	0.356 (0.000)	0.277 (0.001)	0.293 (0.002)	0.288 (0.001)	0.282 (0.001)	0.286 (0.001)
搜索：0.485	0.574 (0.006)	0.680 (0.007)	0.579 (0.034)	0.324 (0.003)	0.206 (0.000)	0.135 (0.001)	0.078 (0.001)	0.108 (0.000)	0.127 (0.000)	0.107 (0.001)
社交： 0.137	0.102 (0.009)	0.082 (0.008)	0.119 (0.028)	0.183 (0.004)	0.217 (0.000)	0.217 (0.002)	0.232 (0.002)	0.224 (0.001)	0.219 (0.001)	0.224 (0.001)
KL散度	0.008 (0.002)	0.011 (0.001)	0.012 (0.007)	0.024 (0.002)	0.093 (0.002)	0.154 (0.004)	0.255 (0.005)	0.194 (0.004)	0.164 (0。004)	0.196 (0.004)
海灵格距离	0.067 (0.007)	0.141 (0。007)	0.078 (0.023)	0.117 (0.005)	0.226 (0.003)	0.277 (0.004)	0.342 (0.003)	0.306 (0.003)	0.285 (0.003)	0.307 (0。003)

表4：不同方法的渠道级聚合得分（CAS）比例比较。报告数值为10次独立运行的平均值，括号内为标准误。第一列列出了作为真实值的渠道级转化次数（CCC）比例。

表 4 总结了 CAS 的比例，以及不同方法的 KL散度和海灵格距离，这些指标是基于 10 次独立运行计算得出的。我们的两种图形归因方法取得了最准确的结果。基于规则的方法（即末次触点、首次触点、线性、时间衰减和 U 型方法）低估了搜索渠道的贡献，并高估了其他渠道的贡献，尤其是展示渠道。它们无法将基线效应纳入考量，因此表现不如其他方法。在所有方法中，我们的图形归因方法估计误差最小，这证明了我们提出的图形归因方法在衡量各渠道对转化的贡献方面具有优势。

6 实际应用

本节将提出的方法应用于一个真实场景案例。数据涉及一家财富 500 强公司的在线订阅产品在连续 4 个月内的付费转化。总计有 2,887, 657 条路径和 74, 440 次转化。

这些触点属于 $Z = 4$渠道，并且包含更具体的细节。对于搜索渠道，品牌搜索是指广告宣传特定的公司或产品，而非品牌搜索则是通用搜索结果，不针对特定公司或产品。对于社交渠道，自有社交点击由公司控制且无需付费（例如：

公司在领英上发布的帖子）；自然社交点击则不受公司控制，但同样无需付费（例如：第三方分享公司在领英上发布的帖子）。对于电子邮件渠道，认知意味着广告只是试图让客户了解产品。促销邮件意味着正在提供折扣产品。行动号召意味着客户已经熟悉该产品，且广告中包含明确的行动号召（例如：立即购买）。

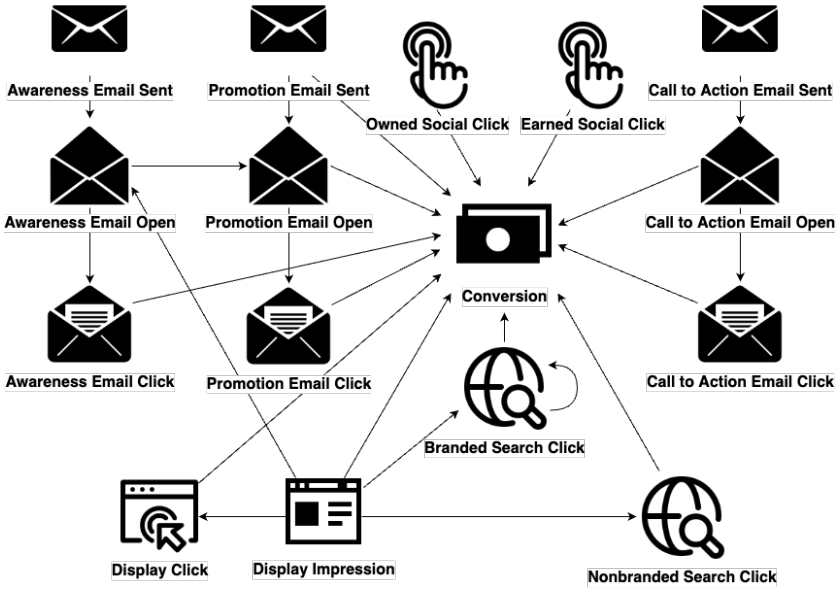

图9：真实数据的格兰杰因果图。

学习得到的格兰杰因果图如图所示9。除了触点对潜在转化的激发作用外，图形点过程模型还发现了渠道内部及不同渠道间触点之间的相互作用。例如，认知邮件打开触点可能会提高打开促销邮件的概率。展示广告曝光可能会触发搜索点击和认知邮件打开。品牌搜索点击还存在一个自环，这意味着此类点击往往会成簇出现。图10 展示了不同方法所得 CAS 的占比。两种图形归因方法和 DNAMTA 的结果相近，均将最高比例的功劳归于搜索渠道（$\geq$ 70%）。对于展示渠道，与直接移除效应（10.7%）相比，总移除效应分配了更多功劳（14.2%），因为展示广告曝光可能会触发搜索点击。逻辑回归比其他算法方法更强调展示渠道的重要性，其赋予展示渠道的分数（33.6%）接近搜索渠道（44.8%）。基于规则的方法与

马尔可夫模型倾向于给电子邮件渠道最高分（$\geq 38\%$）。

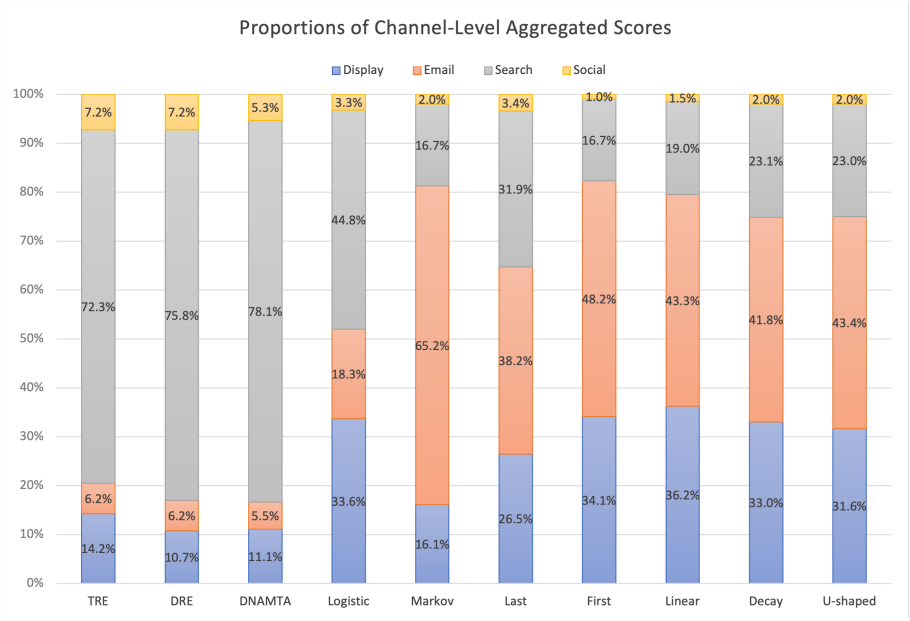

图10：图形归因方法与其他方法的 CAS 占比比较。每个柱形表示一种归因方法的 CAS 占比，四个部分分别以不同颜色表示四个渠道。

根据格兰杰因果图，事件类型确实存在层级结构，因此有必要构建一个响应变量不局限于转化的模型。通过我们的图形方法得到的渠道级聚合得分表明，搜索渠道是效果最显著的渠道。总移除效应则凸显了展示渠道的重要性，它可能在正路径的早期阶段发挥关键作用。

7 结论

本文提出了一种用于多触点归因的新型图形点过程框架。首先，我们构建了一个图形点过程模型，用于分析客户层级的购买路径数据。该图模型利用格兰杰因果关系揭示触点之间的激发效应，以及多种类型触点的直接转化效应。随后，在点过程框架下，我们进一步提出图形归因方法，为每位客户的购买路径中的各个触点及相应渠道分配适当的转化功劳。我们提出的归因方法将

归因得分视为移除效应，并研究了两种类型的移除效应。我们给出了移除效应的概率定义和数学形式。我们开发了一种新颖、高效的基于稀疏化的模拟方法和一种用于计算的反向传播算法。我们采用正则化方法，同时进行边的选择和参数估计。我们开发了交替方向乘子法，以期望的计算效率和可扩展性求解这一优化问题。此外，我们通过建立参数估计值的渐近理论，提供了理论保证。

参考文献

Anderl, E., Becker, I., Von Wangenheim, F. & Schumann, J. H.（2016年），‘绘制客户旅程：从基于图的在线归因建模中汲取的经验’，《国际营销研究杂志》**33**(3)，457–474。

Bacry, E., Bompaire, M., Gaïffas, S.& Muzy, J.-F.（2020年），‘稀疏低秩多元霍克斯过程’，《机器学习研究杂志》**21**(50)，1–32。

Berman, R.（2018年），‘超越末次触点：在线广告中的归因’，《营销科学》**37**(5)，771–792。

Bowman, D. & Narayandas, D.（2001），‘管理客户发起的与制造商的联系：对品类需求份额和口碑行为的影响’，《市场营销研究杂志》**38**(3)，281–297。

Breuer, R., Brettel, M.& Engelen, A.(2011年), ‘在确定不同类型在线广告的效果时纳入长期影响’,《营销快报》**22**(4), 327–340。

Dalessandro, B., Perlich, C., Stitelman, O. & Provost, F. (2012), 在线广告的因果动机归因，载于‘第六届在线广告与互联网经济数据挖掘国际研讨会论文集’，第 1–9 页。

Danaher, P. J. & van Heerde, H. J. (2018年), ‘归因错觉：将归因用于多媒体预算分配时的注意事项’,《市场营销研究杂志》**55**(5), 667–685。

De Haan, E.、Wiesel, T. 与 Pauwels, K.（2016年），‘多渠道归因框架下不同形式在线广告对购买转化的有效性’，《国际营销研究杂志》**33**(3)，491–507。

Didelez, V.（2008年），‘基于局部独立性的标记点过程图模型’，英国皇家统计学会会刊：*B*辑70(1)，245–264。Eichler, M.（2012年），‘多元时间序列的图建模’，概率论及相关领域153(1)，233–268。Gaur, J. 与 Bharti, K.（2020年），‘营销中的归因建模：文献综述与研究议程’，营销研究学会期刊24(4)，1–21。Granger, C. W.（1969年），‘通过计量经济模型和交叉谱方法研究因果关系’，《计量经济学》37(3)，424–438。Granger, C. W.（1980年），‘因果关系检验：个人观点’，《经济动态与控制杂志》2，329–352。Granger, C. W.（1988年），‘因果关系概念的一些近期发展’，《计量经济学杂志》39(1-2)，199–211。Gunawardana, A.、Meek, C. 与 Xu, P.（2011年），‘事件流中的时间依赖关系模型’，神经信息处理系统进展24，1962–1970。Guo, X.、Hu, A.、Xu, R. 与 Zhang, J.（2018年），‘多元霍克斯过程正则化极大似然估计的一致性与计算’，*arXiv*预印本 *arXiv:1810.02955*。Hansen, N. R.、Reynaud-Bouret, P. 与 Rivoirard, V.（2015年），‘多元点过程的LASSO与概率不等式’，《伯努利》21(1)，83–143。Hawkes, A. G.（1971年），‘一些自激励与相互激励点过程的谱’，《生物计量学》58(1)，83–90。Ji, W.、Wang, X. 与 Zhang, D.（2016年），《在线广告的概率多触点归因模型》，载于‘第25届ACM信息与知识管理国际会议论文集’，第1373–1382页。Kakalejˇc´ık, L.、Bucko, J.、Resende, P. A. 与 Ferencova, M.（2018年），‘使用马尔可夫链进行多渠道营销归因’，《应用管理与投资杂志》7(1)，49–60。Kannan, P.、Reinartz, W. 与 Verhoef, P. C.（2016年），‘购买路径与归因建模：专题部分导论’，《国际营销研究杂志》33(3)，449–456。

Kireyev, P., Pauwels, K. & Gupta, S.（2016年）, 《展示广告会影响搜索吗？在线广告中的归因与动态》, 《国际营销研究杂志》**33**(3), 475–490。Kumar, S., Gupta, G., Prasad, R., Chatterjee, A., Vig, L. & Shroff, G.（2020年）, Camta：用于多触点归因的因果注意力模型, 载于《2020年数据挖掘研讨会国际会议（ICDMW）》, IEEE, 第79–86页。Lewis, P. W. & Shedler, G. S.（1979年）, 《通过稀疏化模拟非齐次泊松过程》, 《海军研究后勤季刊》**26**(3), 403–413。Li, H. & Kannan, P.（2014年）, 《多渠道在线营销环境中的转化归因：实证模型与现场实验》, 《市场营销研究杂志》**51**(1), 40–56。Li, N., Arava, S. K., Dong, C., Yan, Z. & Pani, A.（2018年）, 《用于多渠道多触点归因的注意力深度神经网络》, *arXiv*预印本 *arXiv:1809.02230* 。Lin, D. Y. & Ying, Z.（1994年）, 《加性风险模型的半参数分析》, 《生物计量学》**81**(1), 61–71。Lin, W. & Lv, J.（2013年）, 《高维稀疏加性风险回归》, 《美国统计学会杂志》**108**(501), 247–264。Naik, P. A. & Raman, K.（2003年）, 《理解协同效应对多媒体传播的影响》, 《市场营销研究杂志》**40**(4), 375–388。Ogata, Y.（1981年）, 《关于点过程的Lewis模拟方法》, 《*IEEE*信息论汇刊》**27**(1), 23–31。Rumelhart, D. E., Hinton, G. E. & Williams, R. J.（1986年）, 《通过误差反向传播学习表征》, 《自然》**323**(6088), 533–536。Sapp, S., Vaver, J., Shi, M. & Bathia, N.（2016年）, DASS：数字广告系统模拟, 技术报告, 谷歌公司。Shao, X. & Li, L.（2011年）, 数据驱动的多触点归因模型, 载于《第17届ACM SIGKDD知识发现与数据挖掘国际会议论文集》, 第258–264页。

Shapley, L. S. (1953), A value f或n人博弈，载于 H. W. Kuhn & A. W. Tuck[等, e] d 等，《博弈论贡献 II》，普林斯顿大学出版社，第 307–317 页。Sims, C. A. (1972), ‘货币、收入与因果关系’，《美国经济评论》62(4), 540– 552。Singal, R., Besbes, O., Desir, A., Goyal, V. & Iyengar, G. (2022), ‘Shapley 与均匀分布相遇：在线广告归因的公理化框架’，《管理科学》68(10), 7457–7479。Wainwright, M. J. (2009), ‘使用 ℓ_1 约束二次规划（lasso）实现高维噪声稀疏恢复的尖锐阈值’，《*IEEE*信息论汇刊》55(5), 2183–2202。Wiesel, T., Pauwels, K. & Arts, J. (2011), ‘实践奖论文——营销的利润影响：量化在线与线下漏斗进程’，《营销科学》30(4), 604–611。Xu, H., Farajtabar, M. & Zha, H. (2016), 学习 Hawkes 过程的格兰杰因果关系，见 ‘国际机器学习会议’，PMLR，第 1717–1726 页。Xu, L., Duan, J. A. & Whinston, A. (2014), ‘购买路径：用于在线广告和转化的相互激发点过程模型’，《管理科学》60(6), 1392– 1412。Yang, S. & Ghose, A. (2010), ‘分析自然搜索广告与赞助搜索广告之间的关系：正向、负向还是零相互依赖？’，《营销科学》29(4), 602–623。Yu, X., Shanmugam, K., Bhattacharjya, D., Gao, T., Subramanian, D. & Xue, L. (2020), Hawkes图事件模型，见 ‘国际概率图模型会议’，PMLR，第 569–580 页。Zhang, Y., Bradlow, E. T. & S mall, D. S. (2013), ‘发生数据聚集性的新度量’，《应用统计学杂志》40(11), 2533–2548。Zhang, Y., Wei, Y. & Ren, J. (2014), 基于生存理论的在线广告多触点归因，发表于‘2014 年 IEEE国际数据挖掘会议’，IEEE，第 687– 696 页。Zhao, K., Mahboobi, S. H. & Bagheri, S. R. (2019), ‘在线广告的基于收入的归因建模’，《国际市场研究杂志》61(2), 195–209。