

# 归因模型

## Attribution Model

郭谦 2016年11月14日

- 背景知识
  - 什么是Attribution?
  - 归因模型使用的基本概念
  - 归因模型和传统机器学习应用场景的对比
  - 与Marketing Mix Model的对比
- 启发式模型
- Data-driven(数据驱动)模型
- 应用与思考

## 背景知识--什么是Attribution?

### ► Attribution

在社会心理学中，归因主要有两重意思

#### 1、对于行为的解释

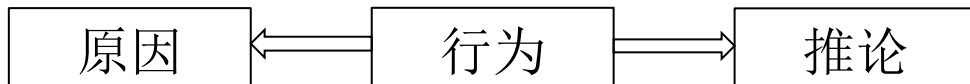
例如：某种事件是由哪些原因导致

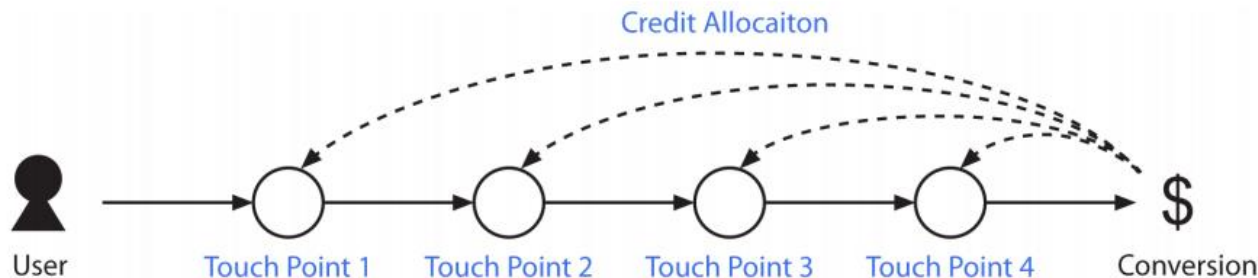
#### 2、对于某种观察到的行为的推论

例如：通过一个人的行为去推断他是否犯罪

这两种意义均包含一个重要的分配(**Assigning**)过程

在接下来的分享中，归因指的是第一种意义，即对于行为的解释





- 广告事件(ad events)用户与广告之间的交互行为称之为广告事件(曝光或点击)
- Touch points: 广告事件发生的媒介
- Channels: 相同或相似的Touch points被聚合为Channels, channels和touch points有时会相互替代, 总体来说, Channel粒度更粗  
例如: 付费搜索渠道可以认为是一个Channel  
用户通过搜索关键词后曝光了一次广告, 可以认为是一个touch point
- 转化(conversion) 通常意义指购买或下单
- 转化路径(conversion path)  
用户最终的转化行为通常是由一系列的touch points综合总用的结果  
这些touch points构成了一条转化路径

## 背景知识—归因模型和传统机器学习应用场景的对比

### ➤ 相同点：

归因模型可以认为是一种分类或排序  
使用相同的数据表示，features and labels

### ➤ 不同点：

利用归因模型的目的是，分析用户转化路径对最终用户决定的影响(是否转化)  
希望给予这些影响一个稳定的，准确的解释。

对于用户的转化行为，可能由一个或多个channel综合作用，归因模型可以给每个channel归纳出合理的credit。可以对在不同channel中投入的预算进行重新分配，使其更加合理，最终达到提高转化的作用

传统机器学习的目的是利用各种信息，去准确的对用户进行分类或排序

### Marketing Mix Model

- 起源于上个世纪80年代
- **自顶向下的策略**
  - 首先确定需要的营销策略和渠道
  - 准确设计所需的对照组
  - 利用对照组的实验结果得出结论
- 先有策略 → 利用实验验证
- 可以同时包含线上和线下的渠道
- 粗粒度
  - 数据在对照组中往往会被聚合，难以达到user-level
- 往往适用于长期的策略
  - 季度/半年/年

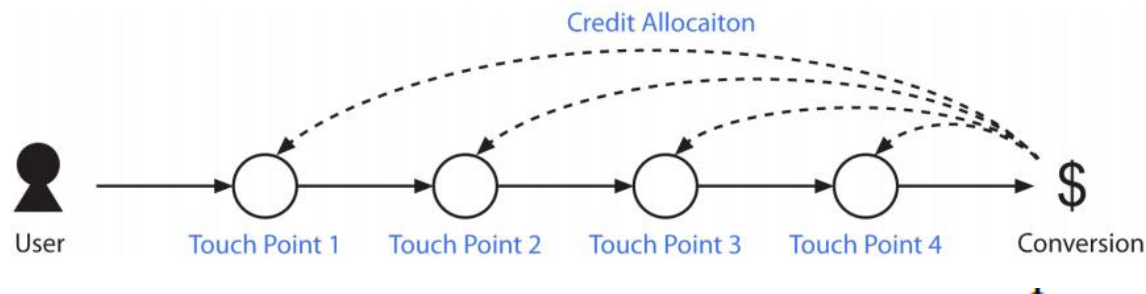
### Multi Touch Attribution

- 最近7-8年
  - 原因是随着互联网和数字营销的发展，可以非常方便的获取大量用户数据
- **自底向上的策略**
  - 首先从大量的用户数据中构建user-level的转化路径
  - 使用归因模型来分析用户数据得出结论
- 先有数据 → 利用模型归因
- 只适用于可以方便的获取用户数据的渠道，主要为线上渠道(付费搜索，展示广告，email等)
- 细粒度，支持用户级别的分析
- 支持短期的策略

## 内容概要

---

- 背景知识
- 启发式模型
  - 常见的启发式模型
  - 启发式模型的局限
- Data-driven (数据驱动)模型
- 应用与思考



Model	Attribution			
Touch point	1	2	3	4
Last touch	0%	0%	0%	100%
First touch	100%	0%	0%	0%
Linear touch	25%	25%	25%	25%
Time decay	10%	20%	30%	40%
Position based	40%	10%	10%	40%
Customised	5%	25%	15%	55%



## 启发式模型—常用的启发式模型



- Last touch: 将全部credit分配给最后一个touch point  
更加强调转化，使用最多



- First touch: 将全部credit分配给第一个touch point  
扩大品牌的影响力，增加品牌的认知度



- Linear touch: 将credit平均分配给所有的touch points  
对所有渠道一视同仁



- Time decay: 各touch points 的分配比例随着时间衰减  
离转化越近，该渠道越重要



- Position based: 将大部分credit分配给首尾的touch points  
同时关注拉新和转化
- Customized: 用户自主定义各touch points比例

## 启发式模型—启发式模型的局限

- 使用启发式模型可能会得出错误的结论

	转化前最后一次点击的渠道
1	搜索推广
2	搜索推广
3	搜索推广
4	搜索推广
5	展示广告
6	展示广告
7	搜索推广
8	搜索推广
9	搜索推广
10	搜索推广

示例：共有10次转化，最后一次的channel如左图，有8次为搜索广告，2次为展示广告。

如果盲目使用Last touch归因规则，将全部的转化价值分配给转化前的最后一个channel

广告主可能会得出结论认为，展示广告的转化成效低，因此有可能会降低在展示广告的预算投入，增加搜索广告的预算投入。

## 启发式模型—启发式模型的局限

	转化前最后一次点击的渠道	转化路径
1	搜索推广	展示广告 → 搜索推广
2	搜索推广	展示广告 → 搜索推广
3	搜索推广	展示广告 → 搜索推广
4	搜索推广	搜索推广
5	展示广告	展示广告
6	展示广告	展示广告
7	搜索推广	展示广告 → 搜索推广
8	搜索推广	展示广告 → 搜索推广
9	搜索推广	展示广告 → 搜索推广
10	搜索推广	搜索推广

当我们观察，最后两次channel时,发现在8次搜索广告的转化中，展示广告协助了6次，因此降低展示广告的预算，可能会抑制搜索广告的转化效果。因此，使用启发式模型，需要对数据有丰富的经验，细致的分析，精确的建模，才能达到所需的效果。

## 内容概要

---

- 背景知识
- 启发式模型
- Data-driven 模型
  - 基于机器学习
  - 基于概率
    - 简单的条件概率
    - 沙普利值(博弈论)
    - 统一的概率表示
- 应用与思考

样本表示为

$(\mathbf{x}, y)$ , 其中  $\mathbf{x} = [x_1, x_2, \dots, x_n]$ ,  $x_i$  为二元变量表示该样本是否接触过  $channel_i$

$y$  表示该样本是否形成一次转化

- 1、利用随机抽样的方法，随机抽取一定的channel和样本训练，进行多轮训练
- 2、每一轮迭代中利用LR来进行训练
- 3、将各个channel的权重作为该channel的重要程度

问题：简单的机器学习模型无法表示channel之间的相互作用。

复杂的机器学习模型能够拟合高维的交叉特征，但其解释性不强。

## Data-driven模型—基于概率模型—简单的条件概率

利用各channel对转化条件概率，计算出各channel对于转化的贡献  
样本表示为

$(\mathbf{x}, y)$ , 其中 $\mathbf{x} = [x_1, x_2, \dots, x_n]$ ,  $x_i$ 为二元变量表示该样本是否接触过 $channel_i$   
 $y$ 表示该样本是否形成一次转化

我们可以计算所有的channel及二阶channel组合对于转化的条件概率  

$$P(y = 1 | x_i), \quad P(y = 1 | x_i, x_j)$$

最终各channel对于转化的贡献如下:

$$V(x_i) = P(y | x_i) + \frac{1}{2N_{j \neq i}} \sum_{j \neq i} (P(y | x_i, x_j) - P(y | x_i) - P(y | x_j))$$

$$V(x_i) = \frac{1}{2} P(y | x_i) + \frac{1}{2N_{j \neq i}} \sum_{j \neq i} (P(y | x_i, x_j) - P(y | x_j))$$

问题：直接利用一阶和二阶条件概率估计贡献值不够精确，忽略了二阶条件概率中，对两个channel的影响可能不一致

更加精确的贡献比例分配方法—沙普利值

沙普利值是可以用在合作任务中，它可以公平的衡量每个合作者对于任务的促进效用。

例子(来自知乎)

当前任务：3个程序员需要合作写500行代码

各程序员及其组合的代码能力：

1号独立写：100行 2号独立写：125行 3号独立写：50

1,2合作写：270行 2,3合作写：350行 1,3合作写：375行

1,2,3合作写：能写500行

那么各个程序员的奖金分配比例是多少？

各程序员的奖金分配比例取决于计算每个程序员的边际贡献

$$\delta_i(S) = v(S \cup \{i\}) - v(S)$$

3人，共有 $3! = 6$ 种合作情况，各情况的边际贡献入下

概率	顺序	1号边际贡献	2号边际贡献	3号边际贡献
1/6	1,2,3	$v(\{1\}) = 100$	$v(\{1,2\}) - v(\{1\}) = 270 - 100 = 170$	$v(\{1,2,3\}) - v(\{1,2\}) = 500 - 270 = 230$
1/6	1,3,2	$v(\{1\}) = 100$	$v(\{1,2,3\}) - v(\{1,3\}) = 500 - 375 = 125$	$v(\{1,3\}) - v(\{1\}) = 375 - 100 = 275$
1/6	2,1,3	$v(\{2,1\}) - v(\{2\}) = 270 - 125 = 145$	$v(\{2\}) = 125$	$v(\{1,2,3\}) - v(\{1,2\}) = 500 - 270 = 230$
1/6	2,3,1	$v(\{1,2,3\}) - v(\{2,3\}) = 500 - 350 = 150$	$v(\{2\}) = 125$	$v(\{2,3\}) - v(\{2\}) = 350 - 125 = 225$
1/6	3,1,2	$v(\{1,3\}) - v(\{3\}) = 375 - 50 = 325$	$v(\{1,2,3\}) - v(\{1,3\}) = 500 - 375 = 125$	$v(\{1\}) = 50$
1/6	3,2,1	$v(\{1,2,3\}) - v(\{2,3\}) = 500 - 350 = 150$	$v(\{2,3\}) - v(\{3\}) = 350 - 50 = 300$	$v(\{1\}) = 50$

边际贡献  $\delta_i(S) = v(S \cup \{i\}) - v(S)$



各程序员沙普利值为

$$1号: \frac{1}{6}(100 + 100 + 145 + 150 + 325 + 150) = \frac{970}{6}$$

$$2号: \frac{1}{6}(170 + 125 + 125 + 125 + 125 + 300) = \frac{970}{6}$$

$$3号: \frac{1}{6}(230 + 275 + 230 + 225 + 50 + 50) = \frac{1060}{6}$$

转化为百分比为各占 32.3%, 32.3%, **35.3%**,

该比例即代表了3个程序员对任务的贡献

在线广告的场景中，最终的任务目标是转化。

不同的channel可以认为是参与任务的合作者，那么最终各channel的沙普利值可以认为是各个channel对于最终转化的贡献。

Shapley value(沙普利值)的形式化定义

$$V_k = \sum_{S \subseteq C \setminus k} \overbrace{w_{S,k}}^{\text{概率}} \cdot \overbrace{(\mathbb{E}[y|S \cup k] - \mathbb{E}[y|S])}^{\text{边际期望}}$$

$$w_{S,k} = \frac{|S|!(|C| - |S| - 1)!}{|C|!};$$

C为channel全集，S为(C与k的差集的所有真子集，包括空集)  
权重 $w_{S,k}$ 表示为以S,k开头的所有转化路径出现的概率

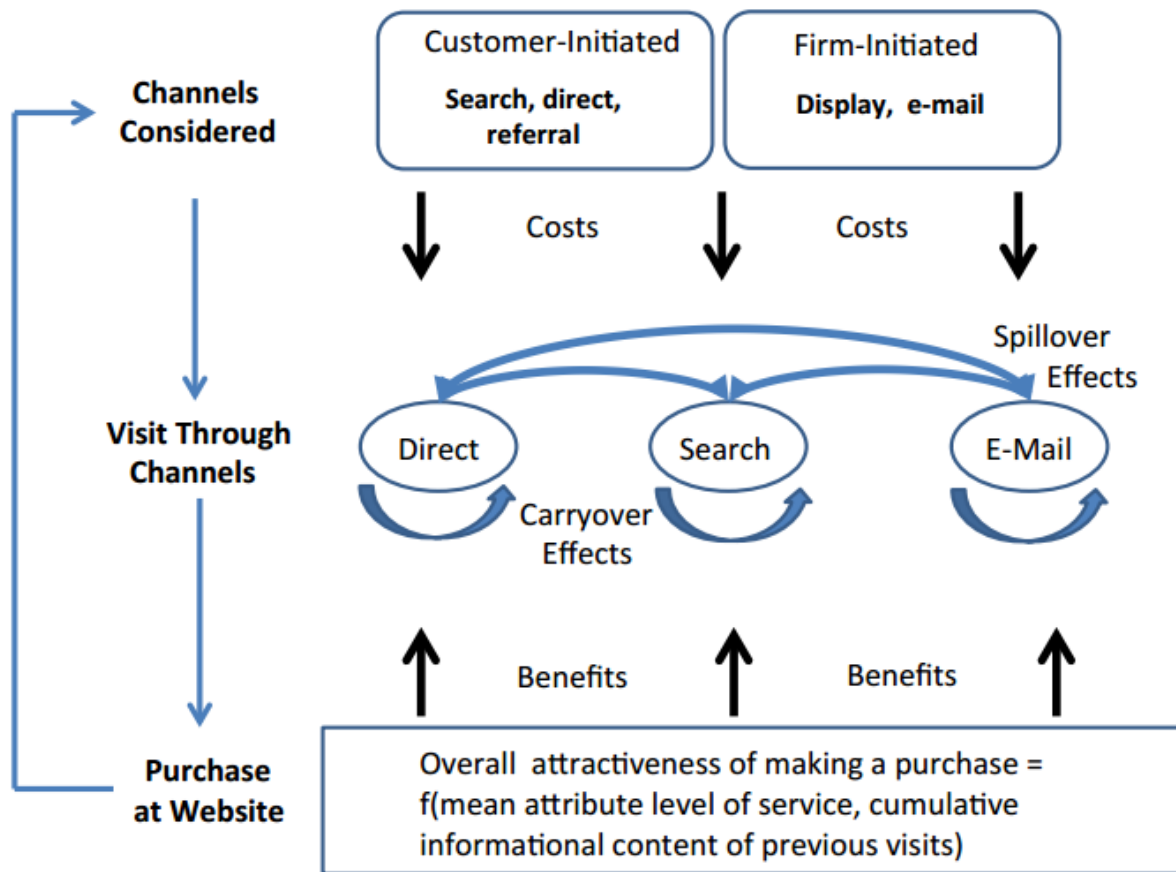
Channel k 的沙普利值表示为channel k 对于转化的边际贡献期望的加权均值

问题：没有区分用户在整体的转化路径中的不同阶段的情况

Spillover effects: 用户被某一Channel接触到，会增加它在其他channel形成转化的可能

Carryover effects: 用户被某一Channel接触到，会增加它在同一channel形成转化的可能

利用 consideration, visit, and purchase stages构建联合似然函数，使用MCMC方法对参数进行估计。



统一的概率表示

$$L(B|\theta) = \prod_{n=1}^{N_i} \prod_{i=1}^I \prod_{j=1}^J \sum_{k=1}^{2^J-1} \underbrace{P(C_i = H_k | \alpha, \Sigma)}_{\text{统一的概率表示}} \times \left[ b_{ijn1}^{B_{ijn}} b_{ijn0}^{(1-B_{ijn})} \right]$$

统一的概率表示

$$b_{ijn1} = P(V_{in} = j | C_i; \beta, \mu, \tau) P(B_{ijn} = 1 | C_i, V_{in} = j; \gamma, \lambda)$$

$$b_{ijn0} = \underbrace{P(V_{in} = j | C_i; \beta, \mu, \tau)}_{\text{visit stage}} \underbrace{[1 - P(B_{ijn} = 1 | C_i, V_{in} = j; \gamma, \lambda)]}_{\text{purchase stage}}$$

- 背景知识
- 启发式模型
- Data-driven 模型

- 应用与思考
  - 宏观上

Budget allocation: 用来调整在不同channel上的预算分配

- 微观上

Bidding strategies: 用来调整在RTB出价中的策略

- 思考

## Budget allocation(贪心求解)

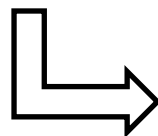
n个channel  $X_i$ , 每个channel最大预算 $S_i$ , ROI为 $R_i$ (利用归因模型来估计),  
如何分配总预算 $B$ 到各个channel, 得到 $B_i$   
使得回报最大, 即

$$\text{maximize } U = \sum_{i=1}^n R_i B_i$$

其中约束为

$$\forall j \in [1, n] \quad B_j \leq S_j$$

$$\sum_{i=1}^n B_i \leq B$$



$$R_i = \frac{\sum_a P(x_i|a) r_a}{\text{Cost in } x_i}$$

### Bidding strategies

目前常用的方法是通过估计曝光的价值(基于VALUE)来出价

$r$ : 转化的价值,  $\theta$ : 转化率(cvr)

$$b\_value = r \times \theta$$

基于MTA的出价策略估计曝光带来的提升价值

$r$ : 转化的价值,  $\theta$ : 提升转化率(通过归因模型计算, lift-cvr, 看过广告后的转化率与不看广告转化率的差)

$$b\_lift = r \times \Delta\theta$$

优点: 基于MTA的出价策略能带来更多的转化。

原因: 基于VALUE的出价策略往往关注高转化率的用戶, 而基于MTA的出价策略更关注能促进转化率的用戶

## 应用与思考-- Bidding strategies—局限

使用lift-based的出价策略往往会有利于广告主，而有可能损害DSP的利益

例子：a, b两个用户，他们看广告后的转化率分别为0.04,0.02，不看广告，转化概率分别是0.03,0.001(该广告对于a的转化率提升为0.01,b为0.019)

固定CPA = 100，竞争对手固定出价为3.5

Value-based: 对于2个用户的出价为 4,2，其中对于a出价胜出，b失败  
总共带来的转化量为  $0.04 + 0.001 = 0.041$ ，dsp的花费为3.5，收入为4

Lift-based: 对于2个用户的出价为 2,3.8，其中对于a出价失败，b成功  
总共带来的转化量为  $0.03 + 0.02 = 0.05$ ，dsp花费为3.5，收入为2

对于广告主来说，他们花费更少，获得了更高的转化量。

但对于DSP来说，并不一定是好事。

原因是目前的行业基本采用Last-touch

因此，只有MTA被所有广告主的广告计划应用，甚至RTB全行业推广MTA时候，才能更广泛的使用lift-based bidding，使得营销更有效果



- 归因模型并不能解决所有网络营销的疑问，广告主需要结合其他深入的数据分析及不断作优化尝试，方可持续地将推广计划的效益提升。
- 认清推广目标。建立清楚及明确的推广目的。不同的目标会影响所应选择的归因模型。
- 归因模型有助广告主了解其自身产品的转化漏斗(Conversion Funnel)。
- 效率和质量平衡(全量计算+增量更新)

### 与京东目前业务的结合

- 不同产品线之间组合投放的营销工具

大盘概览：

提供基本的投放组合效果分析

Top N 的投放组合

品牌用户：

当前投放组合分析

未来投放组合的优化建议

- 不同维度流量渠道的质量监控

- RTB:提供基于MTA思想的出价指导，优化平台出价策略

Wang J, Zhang W, Yuan S. Display Advertising with Real-Time Bidding (RTB) and Behavioural Targeting[J]. arXiv preprint arXiv:1610.03013, 2016.

Williams R. Data-driven multi-touch attribution models.[C]// ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, Ca, Usa, August. 2011:258-264.

Li H, Kannan P K. Attributing Conversions in a Multichannel Online Marketing Environment: An Empirical Model and a Field Experiment[J]. Journal of Marketing Research, 2014, 51(1):40-56.

Dalessandro B, Perlich C, Stitelman O, et al. Causally motivated attribution for online advertising[C]//Proceedings of the Sixth International Workshop on Data Mining for Online Advertising and Internet Economy. ACM, 2012: 7.

Geyik S C, Saxena A, Dasdan A. Multi-touch attribution based budget allocation in online advertising[C]//Proceedings of the Eighth International Workshop on Data Mining for Online Advertising. ACM, 2014: 1-9.

Xu J, Shao X, Ma J, et al. Lift-Based Bidding in Ad Selection[J]. arXiv preprint arXiv:1507.04811, 2015.

Shapley L S. A value for n-person games[J]. The Shapley value, 1988: 31-40.

# Thank You