

119 | 归因模型：如何来衡量广告的有效性

2018-07-06 洪亮劼

AI技术内参

[进入课程 >](#)



讲述：初明明

时长 06:24 大小 2.94M



在互联网广告生态系统的环境中，我们已经分享了不少关于点击率优化和竞价排名以及如何优化出价的内容。接下来我们开始讨论一些计算广告相关的高级话题。之所以说这些是高级话题，是因为作为机器学习在计算广告的应用，这些话题往往都比较偏冷，但在现实中又特别有实用价值。

今天我们先来聊一聊[归因模型](#)，这种技术在计算广告业中被广泛使用。

什么是归因模型

归因模型 (Attribution Model) 是一种计算广告中分配“贡献”的机制。

在现代网站或者应用中，每一个用户都有可能在每一次会话中看到多个不同的广告，或者在多个不同的会话中看到相同广告的不同展示。那么，当用户点击了某个广告，或者是当用户转化以后，比如购买了某个商品或是订阅了某种服务，广告商通常希望知道究竟是哪一个广告起了更大的作用。也就是说，广告商想知道用户接收到的不同广告对这个最后的转化事件都起了什么作用，这个问题就是归因模型研究的核心。

归因模型之所以重要，是因为这里面牵涉到了**广告有效性**这个话题。那么，如何来衡量广告的有效性呢？

衡量广告的有效性，就需要利用归因模型，针对每一个转化来分配“贡献”。这样，对于广告商来说，就可以通过**贡献值的叠加**来看某一个渠道或者某一个内容发布平台的转化效果。

然而，归因模型的难点在于，这里面并没有完全的“基本事实”（Ground Truth），全部都基于一定的假设。同时，归因模型直接关系到广告是否有效的计算，也就关系到我们能否推行一个“公平”的市场，以及能否防止其他的广告商在整个平台上进行博弈。

那么，现在各个平台普遍都在使用的是什么样的归因模型呢？下面我给你介绍几个最基本的归因模型。当然了，说这些方法是模型其实也是不够准确的，因为这些方法大多没有理论支撑，主要是基于经验或者基于传统的方法。

第一种经验方法叫“**最后触碰**”（Last Touch）。

顾名思义，最后触碰指的就是**在转化前的最后一个广告拿走 100% 的贡献值**。这是目前使用最广泛的归因方法，主要是因为它的简单直观。

我们之前讨论过的所有点击率或者转化率的计算都是基于这个归因方法的。一个可以去博弈“最后触碰”的方法就是让 DSP（需求侧平台）把广告投放给那些已经对品牌或者服务产生兴趣的人，从而能够以较大的概率获得用户的转化。在这个过程中，广告的投放其实并没有起作用，而 DSP 也并没有试图去转化新用户。

举一个例子，如果我们已知一个用户喜欢可口可乐，并且很可能在过去购买过可口可乐，那么给这个用户展示最新的可口可乐促销广告，就很有可能让这个用户点击广告并购买了一箱促销的可乐。但是，在这个情况下，我们还可以认为，这个用户很有可能不需要看这个广告也会购买可乐，所以这个广告其实是浪费了资源。**“最后触碰”其实是鼓励了 DSP 采用更加保守的投放方式。**

既然有“最后触碰”，那肯定就有“**第一次触碰**”（First Touch）的经验方法。这种方式
和“最后触碰”截然相反，那就是只要一个用户最后转化了，那么这个用户第一次看到的广告就获得了 100% 的贡献值。尽管用户可能在第一次看到这个广告后还看了其他的广告，但是这些其他广告都不算数了。“**第一次触碰**”其实鼓励了 DSP 尽可能广地投放广告，把广告的投放当做品牌宣传。

除了这两种比较极端但是被广泛使用的归因方法以外，还有一系列的经验方法，都算是这两种方法的某种平衡状态。比如一种叫“**线性碰触**”（Linear Touch）的方法，是给用户在转化的道路上每一个广告都赋予一样的贡献值。还有一种叫“**位置碰触**”（Position Based）的方法，其实就是“最后触碰”和“第一次触碰”的结合。另外一种经验方法“**时间递减**”（Time Decay），则是按照由远到近，对所有的广告位都给一定的贡献值。离转化事件时间越近的广告获得的贡献值越多。

总之，你可以看到，这些林林总总的经验方法虽然都比较直观，但是在实践中，都有可能给一些广告商利用系统进行不公平投放的机会。

基于模型的归因方法

下面我们来看一些具备一定理论基础的归因方法，介绍一个在这方面比较早的探索研究 [1]。在这个研究里，作者们首先介绍了一种叫 **Bagged Logistic Regression** 的方法，这个方法根据当前广告的“触碰”信息，也就是用户看了什么广告，来预测用户是否将会转化。在这个模型里，所有的特征就是二元的用户是否观看了某个广告的信息，然后标签就是用户是否转化。通过这些二元的特征学习到的系数就表达了这个广告在这个预测模型下的贡献度。当然，作者们利用了 Bagged 的方法学习到所有的系数都是正的，确保能够解释这个模型的含义。

同时，作者们还提出了一个对归因问题的概率解法，我来介绍下这个概率解法的直观思路。某一个广告对用户转化的最后作用都来自两个部分：第一部分是这个广告对用户转化的直接作用；第二个部分是当前这个广告和另外一个广告一起作用使用户转化的概率。当然，这个第二部分的联合作用需要减去这两个广告分别单独作用于用户的情况。那么，一个广告对于用户的影响，就是这两个部分概率的加和，这其实就是考虑了一阶和二阶的关系下的归因模型。

知道了归因信息之后，我们还可以把这个信息利用到广告的竞价中。直白来说，就是针对有价值的渠道进行有效的出价，而对没有效果的渠道进行控制 [2]。除此以外，归因信息还可以帮助广告商来分配自己的预算，把大部分的预算用在优质的渠道中来投放广告 [3]。

总结

今天我为你介绍了在线计算广告的一个高级话题：归因模型。

一起来回顾下要点：第一，归因模型是一种计算广告中分配贡献的机制，广泛使用的方法有最后触碰和第一次触碰等；第二，有一些有一定理论基础的归因方法，我们其实可以拓展归因信息的应用场景。

最后，给你留一个思考题，如何来衡量一个归因方法是否有效呢？

欢迎你给我留言，和我一起讨论。

参考文献

1. Shao, X. and Li, L. **Data-driven multi-touch attribution models**. Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 258–264. ACM, 2011.
 2. Xu, J., Shao, X., Ma, J., Lee, K.-c., Qi, H., and Lu, Q. **Lift-based bidding in ad selection**. Proceedings of the 30th AAAI Conference on Artificial Intelligence, 2016.
 3. Geyik, S. C., Saxena, A., and Dasdan, A. **Multitouch attribution based budget allocation in online advertising**. Proceedings of 20th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pages 1–9. ACM, 2014.
-

AI 技术内参

你的360度人工智能信息助理

洪亮劼

Etsy 数据科学主管
前雅虎研究院资深科学家



新版升级：点击「👤 请朋友读」，10位好友免费读，邀请订阅更有**现金**奖励。

© 版权归极客邦科技所有，未经许可不得传播售卖。页面已增加防盗追踪，如有侵权极客邦将依法追究其法律责任。

上一篇 118 | 聊一聊“程序化直接购买”和“广告期货”

下一篇 120 | 广告投放如何选择受众？如何扩展受众群？

精选留言

写留言

由作者筛选后的优质留言将会公开显示，欢迎踊跃留言。