

111 | Twitter的广告点击率预估模型

2018-06-18 洪亮劼

AI技术内参

[进入课程 >](#)



讲述：初明明

时长 06:49 大小 3.12M



在上一篇文章的分享里，我们了解了 LinkedIn 这家公司是怎么来做最基本的广告预估的。LinkedIn 的广告预估模型分为三部分，这里面的核心思想是直接对“冷启动”和“热启动”进行建模，外加和 EE 策略（Exploit & Explore）结合在一起，从而提升了效果。

今天，我们就结合论文《Twitter 时间轴上的广告点击率预估》（Click-through Prediction for Advertising in Twitter Timeline）[1]，来看看 Twitter 的广告预估到底是怎么做的。

Twitter 的广告预估

我们前面提到过最基本的广告形态分类，可以分为“搜索广告”和“展示广告”。当计算广告在互联网上出现以后，这两种广告形态就迅速被很多网站和服务商所采纳。

在最近的 10 年里，随着社交媒体的发展，希望在社交媒体的**用户信息流**里投放广告的需求逐渐增强。我们之前谈到的 Facebook 的案例，其实也是往用户的信息流中插入广告。很多类似的社交媒体都争先恐后地开始进行相似的项目，这一类广告经常被称为**社交广告**。

社交广告的特点是，需要根据用户的社交圈子以及这些社交圈所产生的内容，而动态产生广告的内容。广告商和社交媒体平台都相信，不管是在投放的精准度上，还是在相关性上，社交广告都有极大的可能要强过搜索广告和展示广告。毕竟，在社交媒体上，用户有相当多的信息，例如年龄、性别，甚至在哪里工作、在哪里上学等，这些信息都有助于广告商的精准投放。而用户自己在社交媒体上追踪的各种信息，又可以让广告商清晰地知道用户的喜好。

Twitter 的工程师们在这篇论文里介绍的也是在信息流里投放的社交广告。只不过，Twitter 的工程师们认为，我们之前分享的 Facebook 的解决方案，并没有真正考虑往信息流里插入广告的难点，也就是**广告的排序**，依然把广告的排序问题当做分类问题，也就是用对数几率回归（Logistic Regression）来解决。

另外，Twitter 的工程师们认为，**社交广告比类似 Google 的搜索广告更具挑战性**。因为在社交信息流里，用户所看到的信息都是随时变化的，比如用户在 Twitter 中，可能随时有新的信息进入到信息流中，因此信息流的上下文会随时发生变化。那么，如果要投放和上下文相关的广告，这种变化无疑会带来前所未有的挑战。

利用排序学习来对广告排序

既然 Twitter 的工程师们认为，信息流广告的建模最重要的就是借鉴排序学习的办法。那么，我们就来看一看他们是怎么利用排序学习来为信息流社交广告建模的。

首先，排序学习中最基本的就是“**单点法**”（Pointwise）排序学习。回顾一下，单点法其实就是把排序学习的任务转化为分类问题。其实典型的的就是直接利用“支持向量机”（SVM）或者对数几率回归模型。

第二种比较常用的排序学习的方法就是“**配对法**”（Pairwise）排序学习。通俗地讲，配对法排序学习的核心就是学习哪些广告需要排到哪些广告之前。这种二元关系是根据一组一组的配对来体现的。学习的算法，主要是看能否正确学习这些配对的关系，从而实现整个排序正确的目的。对于配对法排序，我们依然可以使用对数几率回归。只是这个时候，我们针对的正负示例变成了某个广告比某个广告排名靠前，或者靠后。

值得一提的是，通过配对法学习排序学习，对于一般的搜索结果来说，得到最后的排序结果以后就可以了。而对于广告来说，我们还需要**对点击率进行准确的预测**。这个我们之前提到过。于是在这篇文章中专门提到了如何从配对结果到点击率的预测。

具体来说，原理其实很简单，根据配对法学习排序完成以后的广告之间顺序是绝对的，但是绝对的数值可能是不太精确的。这里进行校准的目的是根据配对法产生的预测值，再去尽可能准确地转换为实际的点击率的数值。一般来说，这里就可以再使用一次对数几率回归。也就是说，这个模型的唯一特性就是配对法产生的预测数值，然后模型的目的是去估计或者说是预测最后的实际数值。**这种使用一个回归模型来进行校准的方法，也用在比如要将支持向量机的结果转换成概率结果这一应用上。**

虽然从原理上讲，先有一个配对模型进行排序，然后再有一个校准模型对模型的绝对估计值进行重新校正，这是很自然的。但是在实际的工业级应用中，这意味着需要训练两个模型，那无疑就变成了比较繁复的步骤。

所以，在这篇文章里，作者们想到了一种结合的办法，那就是**结合单点法和配对法**。具体来说，就是直接把两者的目标函数串联在一起。这样做的好处是，可以直接用现在已经有的训练方法，而且同时解决了排序和更加准确预测点击率的问题。我们回顾一下，单点法的特性是尽可能准确地预测每一个广告的点击率，也就是刚才提到的校准的这一个步骤所需要干的事情。这种直接串联的好处是，只需要学习一个模型就可以做到既考虑了排序，又考虑了预测的绝对精准度的问题。

在机器学习应用中，串联多个目标函数是经常使用的一种技术。其目的和作用也就和这个串联的想法一样，就是希望针对多个不同的目标进行优化。一般来说，这里面的核心是，多个串联的目标函数需要**共享模型参数**才能形成有关联的总的大的目标函数；如果没有共享参数，那就仅仅是一种形式上的串联。

模型的实验

在这篇文章里，作者们也是用了 Facebook 提出的“归一化的交叉熵”，简称 NE 的概念以及业界比较常用的 AUC 来针对模型进行线下评价。

在线下实验中，配对法以及单点法和配对法结合的混合方法都在 AUC 上要超过单点法本身。这非常容易理解。只是配对法针对单点法在 NE 上的表现要差很多。这和我们刚才所说的没有对点击率进行估计有很大关系。这一点也在实验中得到了证实。

在在线实验中，单点法相对于以前的自然排序，点击率好了将近 14%，而混合法则好了 26% 左右。可以说效果非常明显。

总结

今天我为你介绍了 Twitter 广告点击率预估的核心算法。一起来回顾下要点：第一，我们讲了 Twitter 认为社交广告的难点是要解决广告的排序问题；第二，我们聊了如何利用排序学习来为点击率预估进行效果提升。

最后，给你留一个思考题，为什么 Twitter 不尝试使用树模型来对点击率进行提升呢？

参考文献

1. Cheng Li, Yue Lu, Qiaozhu Mei, Dong Wang, and Sandeep Pandey. Click-through Prediction for Advertising in Twitter Timeline. Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '15). ACM, New York, NY, USA, 1959-1968, 2015.

 极客时间

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。

精选留言 (2)

写留言



艾熊

2018-06-20



我觉得也可以tune一下树模型用树模型呀，树模型也可以融合多目标也可以适用于pair对的目标呀！



yy

2018-06-20



是不是因为树模型不太适合用户id和广告id当特征，而信息流又希望把这些算成实时的，所以不太适合