

## 110 | LinkedIn的广告点击率预估模型

2018-06-15 洪亮劼

AI技术内参

[进入课程 >](#)



讲述：初明明

时长 05:18 大小 4.85M



上一篇文章我们讲了雅虎的广告预估模型。雅虎早期的模型主要集中在如何利用两轮架构来对点击率进行精确建模，后期的模型回归到了比较传统的利用线性模型外加特性哈希来进行大规模点击率预估的模式。

今天，我们继续来做公司的案例分析，结合论文《LASER：在线广告的可扩展响应预测平台》（LASER: a scalable response prediction platform for online advertising）[1]，来了解 LinkedIn 这家公司是怎么来做最基本的广告预估的。

### LinkedIn 广告预估模型

我们首先来看一看 LinkedIn 的广告预估模型。这个模型的一大“卖点”就是直接充分考虑了“冷启动”和“热启动”两种模式。

那么，什么是“冷启动”，什么又是“热启动”呢？

从我们之前的分享中可以看出，很多点击率预估的模型，都强烈依赖于对用户过去信息以及对广告过去表现的建模。比如刚刚讲过的雅虎预估模型，在它早期的模式中就已经见到了这种信息的作用。

然而，当我们出现新用户或者新广告时，就会有“冷启动”的问题。也就是说，“冷启动”主要是针对新用户或者新广告而言的。这时候基于历史信息特性都无法起作用了，一般来说需要有专门的处理。

相对于“冷启动”，“热启动”指的是我们已经掌握了用户或者广告的一定信息，然后利用这些历史信息来对点击率进行预测。

这么说来，我们一般需要有两套对策，一套针对“冷启动”，一套针对“热启动”。LinkedIn 的方法就是**希望通过一个模型来同时解决这两个问题**。

具体来说，LinkedIn 把对点击率的预估拆成了三个部分。

**第一部分，是利用用户、广告和上下文所建立的全局性预测。**什么意思呢？就是我们利用用户特性、广告特性以及上下文特性来对点击率进行预测。这部分的核心思路就是**这些特性所对应的系数是全局性的**。也就是说，对于不同的用户、不同的广告以及不同的上下文所对应的系数是相同的。因为是全局性的系数，因此这部分其实提供了一种“冷启动”的需求，也就是不管是任何新的用户或是广告，只要有一定的特性，我们总能通过这部分得到一种粗略的估计。

**第二部分，是利用第一部分的用户、广告和上下文信息组成交叉特性，从而学习这些特性之间的关系。**如果说第一部分直接就是线性的预测，那么第二部分其实就是“交叉项”形成的非线性的部分。我们之前在讲推荐系统的时候提到过“分解机”（Factorization Machines）这个模型，讲到过这种“交叉项”所带来的非线性预测的好处。虽然这里和分解机的构成不完全一样，但是整体上表达了相似的意思。

**第三部分，是 LinkedIn 模型提出来的独特之处（和其他公司模型不太一样的地方）。**那就是同样是利用用户、广告和上下文特性，但是 LinkedIn 所提模型的系数则是每个用户、广告和上下文都不同。作者们认为这可以实现“热启动”效果。也就是说，当某个用户、某个广告或者某个上下文已经有比较多的数据以后，就可以依靠这些用户、广告或者上下文自己

的系数了，而不仅仅依靠第一部分的全局系数。这个第三部分只有当数据比较多时才能够起作用。

## 模型的其他特性

这个模型在增加了这些系数的先验概率信息之后变得相对比较复杂。这篇论文介绍了一系列的模型训练思路，都是不错的可以借鉴的工业界经验。

首先，作者们认为，刚才模型中所说的**三个部分所需要的模型更新频率是不一样的**。比如第一部分和第二部分都可以认为是全局模型，也就是说系数是全局性的。因此这些模型的变化会比较慢，作者们建议一个星期对模型进行一次更新。而第三部分则是在已经积累了历史信息后慢慢呈现出的效果，因此对于数据会非常敏感，而且每个用户和每个广告都是不同的系数，因此需要在短时间内，比如半个小时甚至几分钟内，就重新训练模型，以达到个性化的目的。

其次，作者们还把**提出的模型和 EE (Exploit & Explore) 策略结合了起来**。我们在讲推荐系统时介绍过 EE 的思路，简单回顾一下 EE 的目的，主要就是探索那些并没有太多机会被展示的物品，在这里也就是广告。我们刚才说了，所有的系数都加上了先验概率，因此其实可以很容易结合数据计算**后验概率分布**。有了后验概率分布，作者们提出了**以汤普森采样为主的 EE 模式**。这也可以算是论文提出模型的一大亮点。

最后我们提一下这个模型的训练算法。因为要在大规模的数据上对模型进行训练，这篇文章采用了一种**ADMM 算法**。在文章提出来的时候，作者们还是希望能够利用单个服务器对所有的模型参数进行训练。和其他的算法相比，一般认为 ADMM 这种算法的收敛速度更快，但是，利用这种算法的其他公司并不太多。

## 总结

今天我为你介绍了 LinkedIn 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 LinkedIn 把点击率预估分为三个部分，从而分别解决“冷启动”和“热启动”的思路；第二，我们聊了如何更加有效地对这个提出的模型进行训练学习。

最后，给你留一个思考题，回顾我们讲过的推荐系统模块，我们还介绍过什么方法可以结合“冷启动”和“热启动”呢？

欢迎你给我留言，和我一起讨论。

## 参考文献

1. Deepak Agarwal, Bo Long, Jonathan Traupman, Doris Xin, and Liang Zhang. LASER: a scalable response prediction platform for online advertising. Proceedings of the 7th ACM international conference on Web search and data mining (WSDM '14). ACM, New York, NY, USA, 173-182, 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 109 | 雅虎的广告点击率预估模型

下一篇 111 | Twitter的广告点击率预估模型

## 精选留言 (2)

💬 写留言



艾熊

2018-06-20

👍 5

洪博士，通过一篇论文的简单导读来探讨业界应用的话，希望能够听到更深层次更多角度的东西。毕竟，这篇论文所是pr还是内部真的在应用了，应用效果是什么，现在是否还在用，真正应用中卡主的可能是一些没有在论文里写出来的东西...这些都不得而知。论文表层

的东西确实给了题目大家去一看也就是这样，那么希望洪博士多提供一些发散的和批判性的分析，甚至遗留的问题也是比较好的点。感觉每一次留下来思考的点几乎没有反馈和...  
展开 ▾

---



仲毅

2018-06-18

👍 1

請問如何達成個人化的係數，在實作上有點無法想像

展开 ▾