

109 | 雅虎的广告点击率预估模型

2018-06-13 洪亮劫

AI技术内参

[进入课程 >](#)



讲述：初明明

时长 05:46 大小 2.64M



在上一篇文章中，我和你分享了广告回馈预估的一个实例：Facebook 的广告点击率预估。我们重点讲了 Facebook 利用“归一化的交叉熵”这个概念，来衡量广告模型的精准度。另外，我们还分享了 Facebook 提出的两层模型架构，也就是一开始先利用树模型 GBDT 来对特性进行非线性转换和选择，然后再利用一个在线学习的线性模型进行最后的预测，这个架构后来得到了很多研究者和实践者的追随。

今天，我们来分析另外一个经典的公司实例：**雅虎的广告点击率预估模型**。

雅虎早期的广告预估模型

对于曾经的互联网巨头雅虎来说，广告系统很早就成为其重要的经济支柱。早在 2007 年甚至更早的时候，雅虎的工程师和研究人员就投入到了对广告点击率预估的研究中。这方面的

几个代表作我列在了文末的参考文献（[1]、[2] 和 [3]）里，如感兴趣，你可以进一步学习。

我们在这里集中讲讲这些工作的核心思路。

首先，雅虎的研发人员很早就注意到了广告预估模型的一个难点，那就是**数据的稀疏性**。这些具有稀疏性的数据带来了一个后果，那就是**模型对很多“罕见事件”（Rare Events）的估计会出现很大的不稳定性**。

我们来举个例子，如果广告 1 在纽约地区展示了 1 万次，点击 200 次，点击率是 0.02；而广告 2 在旧金山地区展示了 1 百次，点击 3 次，点击率是 0.03，在这样的情况下，我们能判断广告 2 就一定比广告 1 有更高的点击率吗？这里面至少有两个因素需要考虑。

第一，**上下文特性信息非常重要**。这两个广告可能是类型不同，可能展示的地区不同，因此并不能完全直接来对这两个广告进行比较。第二，广告 2 在旧金山地区的展示次数还比较少，因此 0.03 这个**预估值**可能是非常不准确的，或者说至少是**不稳定的**，它的误差要大于第一个广告。

这个时候，研发人员就会思考，如何对这些广告的预估进行更加精确的处理呢？

在这样的背景下，雅虎的研发人员提出了一个点击率估计方法，其实也是一种**两层模型**。第一层模型就是最原始的对点击率的估计，也就是类似我们上面所说的直接按照数据进行估计。当然，这里的问题我们刚才也已经提到了，就是估计的不稳定性。第二层模型是对第一层模型的修正。所谓修正，就是**利用层次化信息来对原始的估计值进行“平滑”（Smoothing）**。

什么是层次化信息呢？我们举例来说明。比如，两个广告来自于同一个广告商，因此它们应该有一定的类似的点击率；两个广告被展示到同一个地区，它们也应该有一定的类似的点击率。这些层次信息给了我们一些启示，来对原始估计值进行修正。当然，根据我们这两个例子你就可以看出，一个广告可以受到多个层次信息的影响，比如广告商的层次信息，地理位置的层次信息，以及类别的层次信息等。所以，要想设计一套完善的基于层次信息的平滑方案也并非易事。

这个时期，雅虎在这方面的工作都围绕着一个主题，那就是**如何对平滑方案进行创新**。一种方法是利用“产生式模型”（Generative Model）的概念，把层次信息的叶子节点的数据

产生过程，定义为基于其父节点数据的一个概率分布产生过程，从而把整个平滑方案的问题转换成为了一个有向无环图上的每个节点的后验概率参数的估计问题（参考文献 [1] 和 [2]）。另外一种方法则采取了一个不太一样的思路，那就是在做平滑的时候，在这种产生式建模之后，还追加了一个过程，利用树模型来对平滑的结果进行再次修正，使得最后的结果能够达到更高的精度（参考文献 [3]）。

这一系列工作虽然在概念上有很高的学术和实践价值，特别是如何利用层次性信息来对预测进行平滑这个方面，但是从整体来说，预估方案变得非常复杂而且环节太多。

雅虎后期的广告预估模型

雅虎后期的广告预估模型又从比较复杂的两层模式转换为了一**层模式**。这个转换主要是考虑到了整个流水线（Pipeline）的复杂度以及需要处理的数据规模逐渐变大，那么利用更加稳定和简单的方法就势在必行了。

对于雅虎后期的广告预估模型，我参考论文《简单和可扩展的展示广告响应预测》（Simple and Scalable Response Prediction for Display Advertising）[4]，在这里为你简单做一个总结。

总体来说，整个模型回到了相对简单的“**对数几率回归**”（Logistic Regression），并且直接对所有的特性（Feature）进行建模。这里面唯一可能和之前的很多工作不太一样的地方，是大量使用了“**特性哈希**”（Feature Hashing）的方法。简单来说，特性哈希就是把原来大规模的有可能是极其稀疏的特性给压缩到了一个固定维度的特性空间里。当然，这肯定会对精度等性能有一定影响，因此这是一个需要有一定取舍的决策。

在这篇论文中，作者们还介绍了如何对大量的数据进行采样，以及如何利用配对的特性（也就是把两种不同的特性，比如广告商和地理位置进行配对）来自动产生更多的非线性因素的方法。

那么这个一层模式的方法所达到的效果怎样呢？论文中论述，相比于之前的两层结构，这个方法所达到的效果有很大程度的提升。

总结

今天我为你介绍了雅虎的广告点击率预估模型。一起来回顾下要点：第一，雅虎早期的广告预估算法，其重点放在了一种两层模型架构上；第二，雅虎广告预估后期的一些思路，重点

则放在了回归到一种更加简单的架构上。

最后，给你留一个思考题，如何在对数几率回归这样的线性模型中引入层次化的平滑思路呢？

欢迎你给我留言，和我一起讨论。

参考文献

1. Deepak Agarwal, Andrei Zary Broder, Deepayan Chakrabarti, Dejan Diklic, Vanja Josifovski, and Mayssam Sayyadian. Estimating rates of rare events at multiple resolutions. Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '07). ACM, New York, NY, USA, 16-25, 2007.
 2. Deepak Agarwal, Rahul Agrawal, Rajiv Khanna, and Nagaraj Kota. Estimating rates of rare events with multiple hierarchies through scalable log-linear models. Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '10). ACM, New York, NY, USA, 213-222, 2010.
 3. Nagaraj Kota and Deepak Agarwal. Temporal multi-hierarchy smoothing for estimating rates of rare events. Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '11). ACM, New York, NY, USA, 1361-1369, 2011.
 4. Olivier Chapelle, Eren Manavoglu, and Romer Rosales. Simple and Scalable Response Prediction for Display Advertising. ACM Trans. Intell. Syst. Technol. 5, 4, Article 61 (December 2014), 34 pages, 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 108 | Facebook的广告点击率预估模型

下一篇 110 | LinkedIn的广告点击率预估模型

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。