

108 | Facebook的广告点击率预估模型

2018-06-11 洪亮劼

AI技术内参

[进入课程 >](#)



讲述：初明明

时长 06:33 大小 3.00M



上一篇文章我们讲了整个计算广告领域最核心的一个问题：广告回馈预估。广告回馈预估，就是预测“用户与广告的交互以及达成交易这种行为”的概率，也就是点击率预估和转化率预估。广告回馈预估存在着数据稀疏等难点和挑战，目前在这个领域比较流行的模型有对数几率回归和数模型等。

今天，我们就来看一个广告回馈预估的实例：**Facebook 的广告点击率预估**。我们会结合2014年发表的一篇论文《Facebook 的广告点击率预估实践经验》（Practical Lessons from Predicting Clicks on Ads at Facebook）来进行分析 [1]。

Facebook 广告预估

Facebook 的广告不是我们之前介绍过的搜索广告或者展示广告的简单应用，而是**社交广告**。可以说，社交广告是最近 10 年慢慢崛起的一种新的广告类型。在论文发表的时候，也就是 2014 年，Facebook 有 7.5 亿“日活跃用户”（Daily Active Users）和超过 1 百万的广告商，这个数字在当时是相当惊人的。而今天，在 Facebook 上活跃的大约有 14.5 亿用户和 5 百万广告商。因此，广告系统所需要应对的规模是成倍增加的。

我们说 Facebook 的广告是社交广告，也就是说，这些广告不依赖于用户输入的搜索关键词。从 Facebook 的角度来说，广告商在其平台上投放广告的巨大优势，在于能够精准地根据用户的地理位置、年龄、性别等重要信息进行有针对性的投放，因此这些信息能够帮助平台选择什么样的广告适合什么样的人群。那这里的难点就是，对于某一个人群来说，可能符合的广告数量是巨大的，这对广告的回馈预估以及整个系统都是一个不小的挑战。

广告点击率的评估

在我们详细解释 Facebook 点击率系统的一些核心组件之前，我们首先来看一看 Facebook 的研究人员是怎么评测他们的系统的。

我们之前提到过，广告系统中的一个巨大挑战就是数据的不均衡。负例，也就是用户没有点击过的广告非常多；而正例，也就是点击过的广告相对比较少。这个比例，根据不同的广告系统会不太一样，但是大体说来，负例与正例的比大概是 10:1、100:1 甚至 1000:1。

在这样的情况下，如果把点击率预估当做是一个分类问题，按照一般分类问题的评价标准，例如准确率，我们只要预测绝大多数，甚至是全部的实例为负例，那么就可以取得很高的准确率。因此，单独看准确率并不是一个很好的评测标准。

这个时候，一个比较通行的评测不均衡数据分类问题的指标是“**曲线下面积**”，或者简称为 **AUC**，这个评测办法可以算是一种替代方法。简单来说，AUC 就是看我们是不是能够把正例给排序到负例上面。也就是说，如果每一个正例和负例都有一个预测数值，那么我们按照这个数值排序，去数每一个正例下面有多少负例，然后对所有正例所对应的数取平均。

AUC 的数值高，则代表我们可以把绝大多数正例排序到负例前面。

当然，AUC 也不是万能的。AUC 的一个最大问题就是它并不在乎所有实例的绝对预测数值，而只在乎它们的相对位置。这在广告系统中可以说是一个非常大的缺陷。我们之前也提过，有很多广告系统组件依赖于对于广告点击率的精确预估，比如收费系统，流量预测等。因此，仅有一个相对位置的正确是不够的。

在这篇论文中，Facebook 团队提到了一个概念叫“**归一化的交叉熵**”，简称 **NE**，用于衡量广告系统的好坏。NE 实际上是一个比值，比值的分母是数据中观测到的实际的点击率的数值，也可以叫作数据的“背景估计”（Background Estimation）；而分子是某一个模型对点击率的估计。这样做的归一化，目的就是来看，在去除了背景估计的情况下，对点击率的估计是否依然好或者坏。

层次化的点击率预估模型

Facebook 的研究人员在这篇论文中提出的点击率预估模型分为两个层次。也就是说，从最初的模型特性输入，需要经过两个不同的模型才对点击率做出最终的预测。这个两层架构对后来的很多点击率预估模型有巨大的影响。

我们首先来看第一层模型，这里的输入是最初的特性，其中连续数值的特性已经被转换成了**离散的数值**。然后，这些离散的数值经过了一个**GBDT 树**来进行特性转换。这里为什么会用 GBDT 呢？主要有两层意义。

第一，GBDT 可以对特性进行非线性组合。也就是说，GBDT 的输出一定是之前特性的非线性的转换，这是由树模型本身的性质所带来的，这个性质对于线性模型来说会有巨大的优势。

第二，经过 GBDT 转换之后，树模型其实选择出了对目标有用的特性，因此这里还起到一个“特性筛选”（Feature Selection）的作用。也就是说，经过 GBDT 的模型，最后剩下的特性肯定是要远小于最初的输入特性的，毕竟有作用的特性是少数的。

在经过了 GBDT 之后，Facebook 的研究者用树模型最后的叶节点当做新的特性，然后再学习了一个**线性的分类模型**。这里的思想其实和后来流行的深度学习的想法很类似，也就是先对输入特性进行非线性转换，然后再经过一个线性分类器来进行最后的预测。这个第二层的线性分类器可以用类似 SGD 的方法进行“在线学习”（Online Learning）。因此，学习到这样一个模型就相对比较容易。

在论文的实验中，作者们不仅展示了两层模型的优势，并且还讨论了很多选取特性方面的经验以及训练模型的经验，比如广告过去的历史信息非常重要，而绝大多数重要的特性都和历史信息有关。

总结

今天我为你介绍了 Facebook 的广告点击率预估的核心算法。一起来回顾下要点：第一，Facebook 的广告是社交广告，有其自身的特点和难点；第二，Facebook 对广告进行评测的指标主要有 AUC 和 NE；第三，Facebook 提出了两层模型的架构，其主要思想是先经过 GBDT 来进行特性转化，再经过一个线性分类器进行最后的预测。

最后，给你留一个思考题，对于两层架构来说，除了模型性能上的优势以外，在训练的方便程度上，这样的架构还没有什么优势或者劣势呢？

欢迎你给我留言，和我一起讨论。

参考文献

1. Xinran He, Junfeng Pan, Ou Jin, Tianbing Xu, Bo Liu, Tao Xu, Yanxin Shi, Antoine Atallah, Ralf Herbrich, Stuart Bowers, and Joaquin Quiñero Candela. Practical Lessons from Predicting Clicks on Ads at Facebook. Proceedings of the Eighth International Workshop on Data Mining for Online Advertising (ADKDD'14). ACM, New York, NY, USA, , Article 5 , 9 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。

精选留言 (7)

写留言



艾熊

2018-06-20

1

hi, 因为这篇论文是2014年发的, 想问一下现在Facebook还是基于这个模型来做广告ctr预估吗? 这个模型适用和不适用的规模和场景可以分享一下吗?



星星之火

2018-06-15

1

老师能不能讲下具体的实现, 架构中的难点

展开



帅帅

2019-01-16

1

反正加个GBDT就是比单纯的LR好就对了

展开



夏洛克的救...

2018-09-27

1

如何了解谷歌的广告模型

展开



arfa

2018-09-13

1

老师, id怎么加进模型?

展开



夏洛克的救...

2018-09-10

1

老师怎么看待微信广告现状和前景？

展开 ∨



旭

2018-06-11



无需人工特征工程也是优势，缺点应是gbdt的离线更新以及预测时间消耗

展开 ∨