

[论文分享]华为PAL论文：解决推荐、广告中的position-bias问题

广告算法 2019-10-11

论文地址：

https://www.researchgate.net/publication/335771749_PAL_a_position-bias_aware_learning_framework_for_CTR_prediction_in_live_recommender_systems

华为2019文章，解决推荐、广告中的position-bias问题。

一段话总结

解决推荐、广告中的position-bias问题（训练的时候知道位置信息，预估的时候不知道）。文章将广告被点击的概率分为两个因素：广告被用户看到的概率和用户看到广告后点击广告的概率。进一步假设用户看到广告后点击广告的概率与广告的位置无关。因此建模成两个模块：

广告被看到的概率预估模型和用户看到广告后，点击的概率预估模型。同时在线上servering的时候，只需要预估第二个模型（不需要位置信息）。

个人觉得主要问题在于假设太强：

第一个假设，广告是否被用户看到只跟广告位置有关，这个假设在广告场景是不成立的。跟广告、以及用户的属性都有关系（广告大图、小图等）。不过可以对第一个模型更精细建模解决（看论文中这个模型只用了position信息）。

第二个假设，用户看到广告后，点击广告的位置，其实跟广告是否被点击很可能还是有关的。比如都在一个页面，用户同时看到了位置1的广告和位置3的广告，但用户点击位置1的广告的概率更大。

背景

用户是否点击广告有两个影响因素：

1. 广告的位置

2. 用户的兴趣

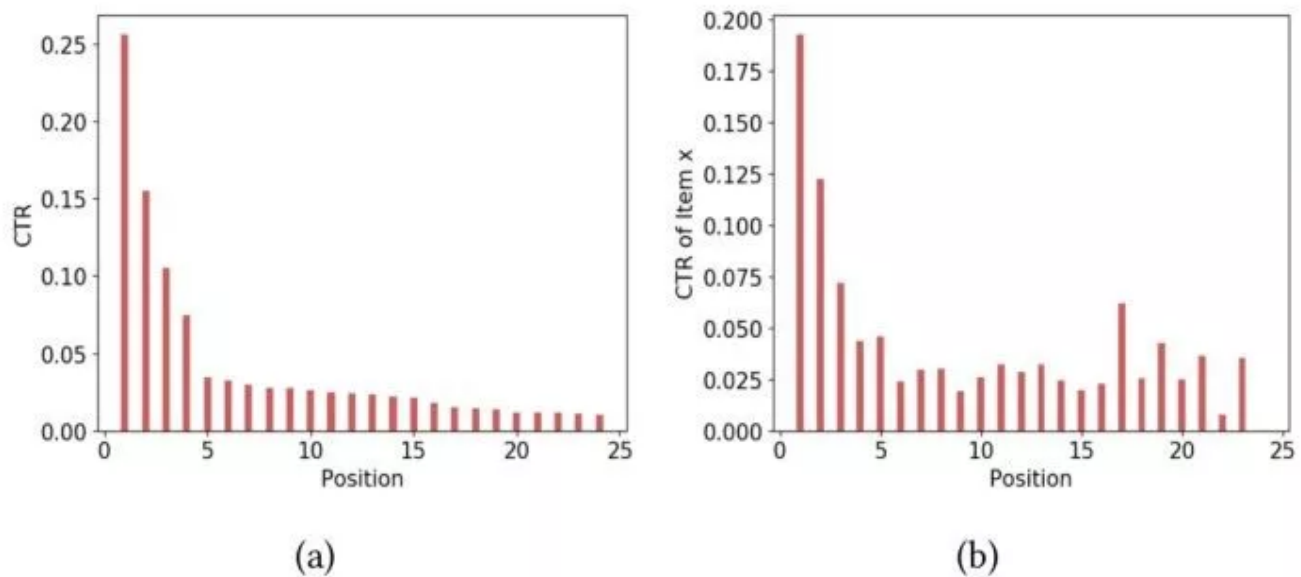


Figure 1: CTRs at different positions

广告算法

推荐广告都会存在position-bias的问题：

训练的样本里面有广告位置信息，但是inference的时候无法知道广告位置（预估之后才决定广告的位置，因此预估时是无法知道的）。

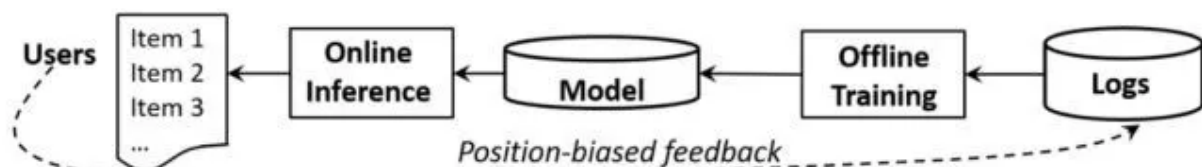


Figure 2: Workflow of Recommendation

广告算法

工业界常见的解决方法：

训练的时候，把广告位置信息当做特征进行训练，inference的时候，用一个默认值来预估。不同的默认值会影响到最终线上的效果，因此一般只能获取到一个次优解。

解决方案

用户点击广告的概率由两部分组成：

1. 广告被用户看到的概率
2. 用户看到广告后，点击广告的概率

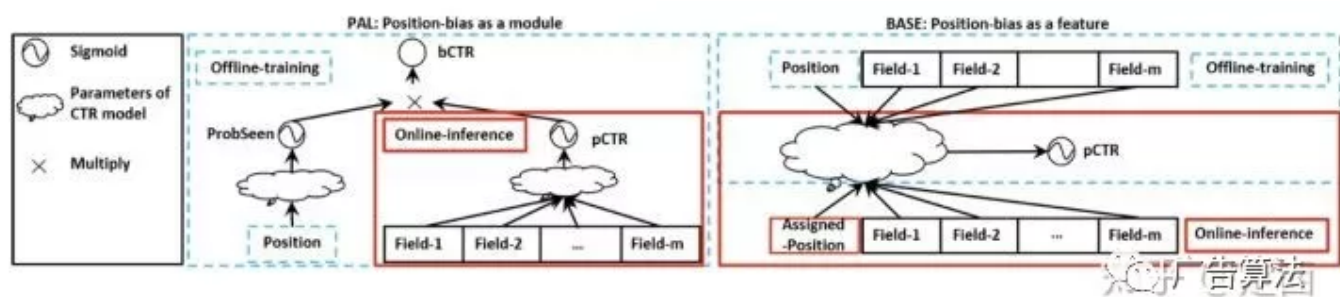
$$p(y = 1|x, pos) = p(seen|x, pos)p(y = 1|x, pos, seen)$$
广告算法

进一步假设：

1. 用户是否看到广告只跟广告的位置有关系
2. 用户看到广告后，是否点击广告与广告的位置无关

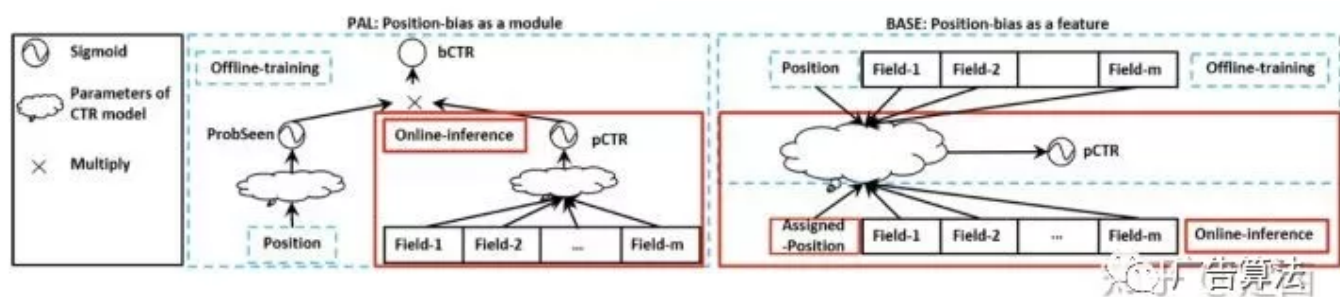
$$p(y = 1|x, pos) = p(seen|pos)p(y = 1|x, seen)$$
广告算法

基于该假设，就可以分开建模：



模型框架：

基于该假设就可以分开建模（左边是分开建模，右边是工业界的常用方法，也是该文章实验的baseline）



有两个模块组成：

1. ProbSeen：预估广告被用户看到的概率
2. pCTR：用户看到广告后，点击广告的概率

训练方法：

同时训练两个模型（单独训练会得到次优解）

Loss Function:

$$L(\theta_{ps}, \theta_{pCTR}) = \frac{1}{N} \sum_{i=1}^N l(y_i, bCTR_i) = \frac{1}{N} \sum_{i=1}^N l(y_i, ProbSeen_i \cdot pCTR_i),$$

参数更新:

$$\theta_{ps} = \theta_{ps} - \eta \cdot \frac{1}{N} \sum_{i=1}^N (bCTR_i - y_i) \cdot pCTR_i \cdot \frac{\partial ProbSeen_i}{\partial \theta_{ps}}. \quad (4)$$

$$\theta_{pCTR} = \theta_{pCTR} - \eta \cdot \frac{1}{N} \sum_{i=1}^N (bCTR_i - y_i) \cdot ProbSeen_i \cdot \frac{\partial pCTR_i}{\partial \theta_{pCTR}}$$

线上servering的时候，直接预估pCTR，不需要知道位置信息（这个点需要注意下，线上servering的时候，就预估假设这个广告被曝光后，用户点击的概率，只需要预估第二个模型）。

实验

不同位置默认值对于离线AUC、LogLoss的影响

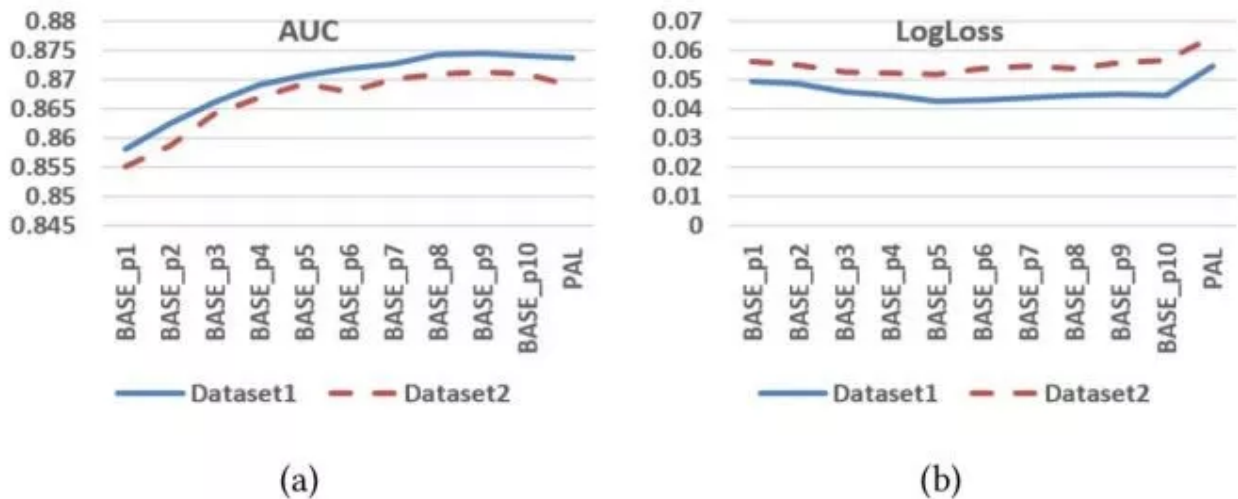


Figure 5: Offline Experimental Results

线上效果

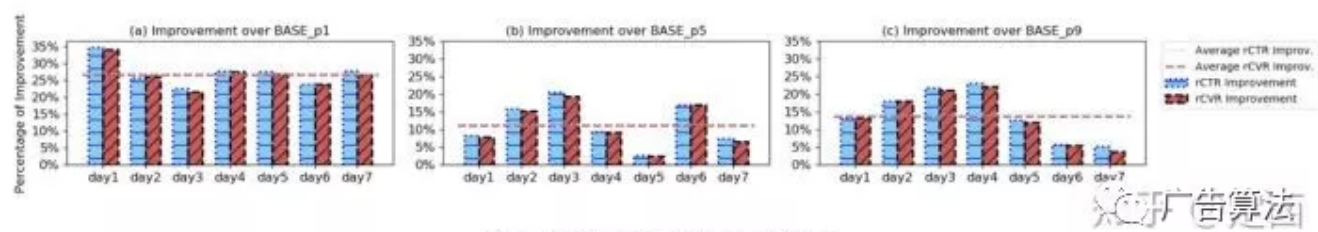


Figure 4: Results of Online AB Test.

广告算法

REFERENCES

其他两篇也是关于position-bias的文章，google和微软的，也很值得看一遍。

- Google: Learning to Rank with Selection Bias in Personal Search
- Bing: Model Ensemble for Click Prediction in Bing Search Ads

阅读原文

喜欢此内容的人还喜欢

高智商刽子手！ 兰州大学教师虐杀流浪小动物！

没有买卖没有杀害9

大触专访 | 这三渲二场景太帅了！ 风格多样、技法独特的网易宝藏原画师

原画人