

125 | SIGIR 2018论文精读：偏差和“流行度”之间的关系

2018-07-20 洪亮劼

AI技术内参

[进入课程 >](#)



讲述：初明明

时长 08:34 大小 3.92M



2018 年的 SIGIR（国际信息检索研究与发展大会）于 7 月 8 日~12 日在美国密歇根州的安娜堡举行。从今天开始，我将精选几篇大会上最有价值的论文，和你一起来读。

我先简单介绍一下这个大会。SIGIR 从 1978 年开始举办，有 40 年的历史，是信息检索和搜索领域的顶级会议。SIGIR 2018，全称是 The 41st International ACM SIGIR Conference on Research and Development in Information Retrieval。

从最初举办开始，这个会议就成为了信息检索领域，特别是搜索技术和推荐技术方面的权威学术会议。会议的内容往往包含了搜索、推荐、广告、信息提取、互联网数据挖掘等诸多领域的优秀论文，每年都吸引着来自世界各地的学者和工程师参会，来分享他们最新的研究成果。

今天，我们首先来看一看今年的最佳论文，标题是《推荐系统中流行度有效性的概率分析》
([Should I Follow the Crowd? A Probabilistic Analysis of the Effectiveness of Popularity in Recommender Systems](#))。

这篇论文一共有两位作者，均来自马德里自治大学 (Universidad Autónoma de Madrid)。第一作者罗西奥·卡纳马雷斯 (Rocio Cañamares) 已经发表了好几篇相关主题的论文，第二作者帕布罗·卡斯蒂罗斯 (Pablo Castells) 是马德里自治大学、甚至是整个欧洲的信息检索学术权威。论文有超过 5 千次的引用。

论文的主要贡献

想要理解清楚这篇论文的主要贡献，我们首先要从推荐系统，或者是从更大的方向上，来看所有信息检索系统都存在的一个核心问题，那就是“**偏差**” (Bias)。偏差会带来一系列问题。这对推荐系统甚至信息检索系统的建模和评价都带来了巨大的挑战。

那么，为什么信息检索系统会有偏差呢？

我这里举一个简单的例子来说明。假设我们有两个物品和很多用户。对于每一个用户来说，系统都按照随机的顺序，分别给用户展示这两个物品，并且询问用户是否喜欢。

在这样的假设里，顺序是随机的，因此对于同一个用户来说，用户是否喜欢某一个商品，就完全是取决于这个物品本身的属性。对于所有用户来说，在整体上呈现的用户对这两个物品的喜好，则完全来自于大家对这两个物品本身的一种评价。那么，我们可以看到这里面没有任何的偏差。

然而，只要这个场景稍微有一些改变，就很容易引入各种偏差。比如，我们有超过一万件物品。尽管我们还是随机地展示给用户，但用户可能在看过一定数量的物品之后就慢慢厌倦了，那么，用户对于物品的喜好判断或许就会受到厌倦的影响，甚至，用户还很有可能直接放弃查看后面的物品。

还有很多相似的情况，比如我们不是把每个商品逐一展示给用户看，而是提供一个列表。那么，用户很有可能会以为这个列表有一定的顺序，比如在列表排名上方的物品可能是比较重要的。有研究表明，在有列表的情况下，用户很可能会按照列表的顺序提供某种喜好判断。很明显，在这样的情况下，用户的喜好判断就受到了这个列表顺序的干扰。

上面我们提到的都是“**表现偏差**”（Presentation Bias）。除此以外，一个信息系统其实还有很多类型的偏差，比如**系统性偏差**：一个新闻系统，只给用户推荐娱乐新闻，而不给用户看时政新闻，在这样的情况下，用户表现出来的喜好性就是有偏差的，因为系统没有给用户表达对时政新闻喜好的可能性。

信息检索和推荐系统的学者其实很早就意识到了偏差对于建模的影响。不管是我们这里提到的表现偏差还是系统性偏差，如果我们直接利用用户和系统交互产生的数据，那么训练出来的模型以及我们采用的衡量模型的办法也会有偏差，那我们得出的结论有可能就是不精准的。

这篇论文就是希望能够**系统性地讨论偏差在推荐系统所带来的问题**。具体来说，这篇论文主要是探讨偏差和“**流行度**”（Popularity）之间的关系。

这里描述的是这样一种情况：有一些物品很有可能曾经给很多人推荐过，或者同时还被很多人喜欢过或者评价过，那么，这种流行度高的物品会不会对推荐结果的评价带来意想不到的偏差呢？

在过去的研究中，大家只是对这种流行度高的物品有一种直观上的怀疑，认为如果一个推荐系统仅仅能够推荐流行的物品，那肯定是有偏差的。但之前的很多工作并没有定量地去解释这里面偏差和评价之间的关系。**这篇论文就提供了一个理论框架，指导我们去理解偏差以及偏差带来的评测指标的一些变化。**

论文的核心方法

今天我们不去讲这篇论文的理论框架细节，我会重点提供一个大体的思路，帮助你理解这篇论文希望达到的目的。

简单来说，为了表达偏差和流行度之间的关系，作者们用了这么几个**随机变量**：用户是否对某个物品打分，用户是否对某个物品有喜好，以及用户是否观看某个物品。这里面的一个细节，或者说技巧，就是如何用概率的语言把这三者之间的关系给表达清楚。

作者其实采用了一些简化的假设，比如假设在测试集上的物品是训练集上没有出现过的等等。这样，就能够写出在测试集上用户对物品评价的一个**期望关系**，这个期望关系包含用户对所有测试物品是否有喜好。有了这层期望关系以后，就开始推导出，在测试集上**理想状态下的最佳排序是一个什么样子**。在这里的理论讨论其实并没有很大的现实意义，但是这是第一次研究人员用数学模型去详细表征一个最优的在测试集上的按照流行度排序的结论。

紧接着，作者们还讨论了这个最优排序在两种极端情况下的变化。一种情况是用户过往的行为都是仅依赖于物品本身的属性，而没有任何其他偏差。另外一种情况是用户过往的行为和物品本身的属性无关，意思就是仅依赖于其他的偏差。

在第一种极端情况下，最优的排序其实也就是我们所能观测到的最优排序，那就是按照物品的流行度。在第二种极端情况下，最优的排序其实是**按照平均打分**。

当然，你可能会说讨论这两种极端情况并没有现实意义呀。但这两种极端情况的讨论其实就证明了，**只有在没有偏差的情况下，按照物品的流行度排序才是平均情况下最优的**。而很明显，现实存在偏差，因此依靠流行度的排序，即便是平均情况下，也不是最优的选择。

然后，论文讨论了用户是否观看某一个物品对用户行为的影响。关于这一部分的讨论，其实之前已经有很多工作都做了一些类似的探索。不过这篇论文得出了一个有意思的结论。在考虑了用户观看物品的偏差以后，通过模拟的方法，我们会发现：**随机结果的效果其实要比之前的观测值要好很多，而按照流行度的排序虽然不错，但是比随机的效果并没有好很多，而基于平均打分的结果其实要优于按照流行度的排序**。可以说，这是一个有别于之前很多工作的新发现。

延申讨论

虽然这篇论文获得了 SIGIR 2018 的最佳论文奖，但是如果我们站在更大的角度上来分析这篇论文，其实就会发现，作实际上作者们是开发了一套特有的理论框架来描述推荐系统中的某一种偏差。**更加普适化的对偏差的建模其实需要有随机化的数据，以及利用因果推断的办法来对任意情况下的偏差进行分析**。文章提出的概率模型仅仅在这篇文章讨论的假设情况下才能成立。

当然，瑕不掩瑜，这篇文章不管是从结论上，还是从实际的分析过程中，都为我们提供了很多有意义的内容，帮我们去思考偏差对于建模所带来的挑战以及我们应该如何应对。

总结

今天我为你们讲了今年 SIGIR 2018 的最佳论文。

一起来回顾下要点：第一，我们详细介绍了这篇文章要解决的问题和贡献，探讨偏差和流行度之间的关系，系统性地来讨论偏差在推荐系统所带来的问题；第二，我们简要地介绍了

文章提出方法的核心内容，包括设定随机变量、期望关系以及推导理想状态下的最佳排序；第三，针对论文我们简单进行了讨论。

最后，给你留一个思考题，在不考虑偏差的情况下，为什么一般的推荐系统会偏好于推荐流行物品的算法呢？

欢迎你给我留言，和我一起讨论。

 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。

上一篇 124 | 数据科学家必备套路之三：广告套路

下一篇 126 | SIGIR 2018论文精读：如何利用对抗学习来增强排序模型的普适性？

精选留言 (1)

 写留言



晏

2018-07-20



推荐流行物品算法，只要针对所有用户的点击率和购买率等进行建模。针对某一类人员的特别推荐，识别其他的偏差点，需要重新建模，难度更大。

拼课微信：171614366!