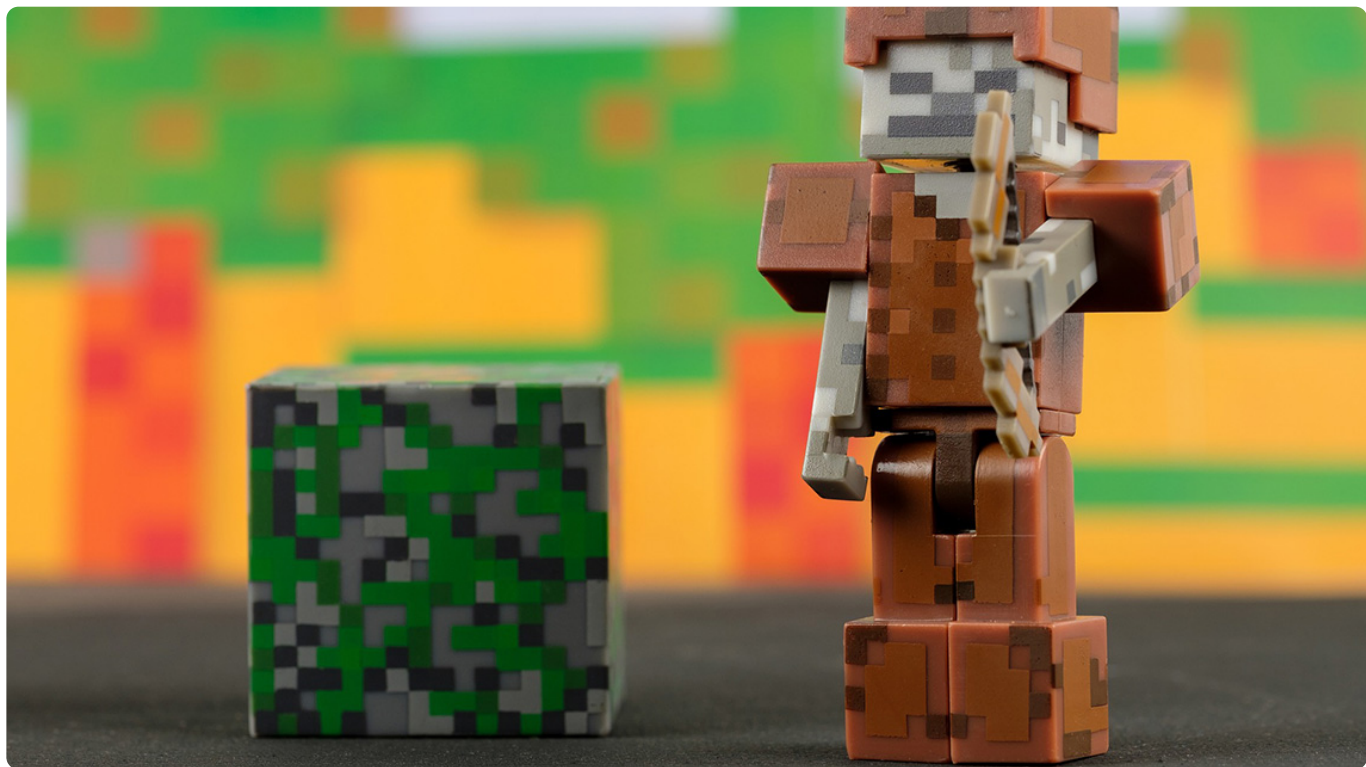


073 | 推荐的Exploit和Explore算法之二：UCB算法

2018-03-21 洪亮劼

AI技术内参

[进入课程 >](#)



讲述：初明明

时长 06:32 大小 2.99M



这周，我们来讨论 EE 策略，周一介绍了 EE 的综合情况。今天来看一种最基本的思路，叫作 **UCB (Upper Confidence Bound) 算法**。

EG 算法

在介绍 UCB 算法之前，我们先来看一种更加简单的 EE 算法，叫 **EG (Epsilon-Greedy) 算法**。

我们先来回顾一下 EE 的主要目的。EE 的核心思想是说，我们对当前物品的估计往往是有限的、不准确的，需要不断尝试来增强对整个环境的了解，进而能够更加准确地对每个物品进行估计。

可以说，**EG 算法是最简单也是最基本的 EE 算法**。EG 算法的基本思路是这样的：既然我们当前对所有物品的估计是不完整的，那就可以随机地显示所有物品来获取数据。假设我们现在有一千个物品，我们对每个物品都需要估计一个数值，比如点击率。很显然，这个点击率的估计受以下两个因素的影响：已经显示了什么样的物品和显示的次数。那么，要想进一步提高这个估计值的准确度，EG 算法认为我们必须对所有物品进行“探索”（Explore）。

具体来说，**EG 算法的流程是这样的：对于所有的物品，在概率 P 的情况下，按照当前的估计值来显示物品**。回到刚才点击率的例子，那就是在概率 P 的情况下，谁的点击率高就显示谁。然后在概率 $1-P$ 的情况下，随机地在所有物品中选择显示对象。如果我们从所有用户的角度来看，也就是说， $P\%$ 的用户看到的是根据点击率排序的物品，而 $(1-P)\%$ 的用户看到的是随机的物品。

EG 的想法是，虽然在最开始的时候，这种随机性可能会带来用户体验的下降，也就是那 $(1-P)\%$ 的用户会持续看到随机的内容，但是在牺牲这部分用户体验的情况下，随着时间的推移，慢慢地从整体上来看，对所有物品的估计会更加准确， $P\%$ 的那部分用户的体验会增加。这也就是一种牺牲小部分用户体验来换取大部分用户体验的思路。

UCB 算法的核心思路

我们刚才讲了 EG 算法的基本思路。很显然，**EG 有一个很大的问题，那就是有一个固定百分比的用户持续看到的都是随机的内容，这就太过于局限**。

那么，我们能不能根据对物品的估计，来动态地调整显示物品的可能性呢？

回到我们刚才说的物品点击率预测的例子。一般来说，我们可以根据每个物品的显示记录来预测点击率。这个数值，其实是一个估计的“均值”。然而，这个估计可能是很不准确的，或者说，估计的置信度不高。

那么，**如何来衡量一个物品的置信度呢**？在统计中，一个比较好的方法，就是利用“标准差”（Standard Deviation）。从感性上来理解，标准差描述了数据的离散程度，也就是说，标准差其实是量化了数据在均值周围的分布情况。标准差小说明我们对这个数值的估计比较有信心，而标准差大则说明了不确定性大。

有了标准差的思路之后，我们再回到最初的问题，怎样才能动态地调整显示物品的可能性呢？我们沿着这个思路再稍微展开一下。很显然，我们需要考虑物品的当前估计，但同时也需要考虑这个估计的置信度。这个置信度告诉我们是不是需要更多地去“探索”这个物品。

那么，很自然地，**我们就可以同时用均值和标准差来表达对一个物品的整体估计**，然后根据这个估计来排序显示物品。因为标准差已经表达了这种不确定因素，因此，我们的结果里面，不确定性特别大的物品，会被显示到前面来。

具体来说，**UCB 采用的数值是均值加上两倍的标准差来作为最终排序的实用数值**。当然，不同类型的 UCB 算法在最终的数值上会有所偏差，但是大体思路基本相同。在这样的思路下，每一轮计算，我们都根据当前的数据计算出物品点击率的均值和当前的标准差，然后根据 UCB 的计算，我们再基于物品的数值，也就是刚才提到的均值加上两倍的标准差来排序。

在这样的一个机制下，经过多轮显示，当某个物品的数据越来越多的时候，标准差也会慢慢减小，最终 UCB 的数值会收敛到均值上。因此，**这个算法本身其实是同时考虑了物品现在的情况以及在这种情况下的置信度，并且寄希望通过多次迭代来达到减小标准差，提高置信度的目的**。

UCB 算法的讨论

UCB 的方法一经提出，因为它的简单，并且有数学基础，马上备受学术界的关注。另外从概念上来说，UCB 的确要比 EG 要好。EG 有一个固定的群体需要忍受“探索”的不确定性结果；而 UCB，这部分“牺牲”消失了。不仅如此，我们之前提到 EG 中有一个概率 P ，这个参数在 UCB 中也完全消失了。这个概率 P 是一个需要调整的参数，而且没人知道这个参数该怎么设置才是最优的。而在 UCB 中，每一个物品的“随机度”是不一样的，并没有一个全局的类似 P 这样的参数。

那是不是 UCB 就没有问题了呢？

UCB 的最大问题在于其对物品打分的机制，也就是均值加上两倍的标准差。这个机制听上去很合理，但在实际中，比如一些大型网站，会有上百上千甚至几百万的物品，那么，在没有任何特殊处理的情况下，对绝大多数物品的打分数值是相同的。什么意思？比如，很多物品从来没有被显示过，估计的均值就可能是 0，或者是一个默认的初始值，在这样的情况下，物品的标准差自然也是一样的。那对于所有这些一样的物品，UCB 本身并没有设计任何机制来加以区分。

这其实说明了一个问题，UCB 算法本质上还是“确定性”（Deterministic）算法，也就是说并没有随机性。表面上通过标准差引入的不确定性其实是一种假象，这个算法从根本上还并不能真正做到随机探索。

小结

今天我们继续讨论推荐系统的一个重要问题，EE 策略。我们介绍了一个很重要的算法，UCB 算法。

一起来回顾下要点：第一，我们首先介绍了比 UCB 算法更加简单的 EG 算法；第二，我们介绍了 UCB 的核心思想；第三，我们讨论了 UCB 存在的一些问题。

最后，给你留一个思考题，如果有一大堆物品的 UCB 打分值是一样的，我们该如何解决这个问题呢？

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 072 | 推荐的Exploit和Explore算法之一：EE算法综述

下一篇 074 | 推荐的Exploit和Explore算法之三：汤普森采样算法

精选留言 (2)

 写留言



不玩

2018-11-04



请问标准差是怎么计算的？我理解经验的点击率就是总的click除以总的show

展开 ∨



林彦

2018-03-23



如果有一大堆物品的 UCB 打分值是一样的，是不是可以考虑引进一个随机数来打破平衡？

搜索资料过程中提到LinUCB可以根据物品的特征，涉及的用户特征，所在页面的信息得出一些特征值，用这些特征值的组合计算来预估物品的期望值和UCB值作为打分，再根据实际结果更新特征值。我的理解分数一样时也是先随机选择一个。...

展开 ∨