


UCB算法升职记——LinUCB算法

 刑无刀 · 1 年前

UCB再回顾

上回书说到，UCB这个小伙子在做EE(Exploit-Explore)的时候表现不错，只可惜啊，是一个不关心组织的上下文无关(context free)bandit算法，它只管埋头干活，根本不观察一下面对的都是些什么样的arm。

进一步送UCB去深造之前，我们再把UCB算法要解决的问题描述一下：

面对固定的K个item（广告或推荐物品），我们没有任何先验知识，每一个item的回报情况完全不知道，每一次试验要选择其中一个，如何在这个选择过程中最大化我们的回报？

UCB解决这个Multi-armed bandit问题的思路是：用置信区间。置信区间可以简单地理解为不确定性的程度，区间越宽，越不确定，反之亦反之。

每个item的回报均值都有个置信区间，随着试验次数增加，置信区间会变窄（逐渐确定了到底回报丰厚还是可怜）。

每次选择前，都根据已经试验的结果重新估计每个item的均值及置信区间。

选择置信区间上限最大的那个item。

“选择置信区间上界最大的那个item”这句话反映了几个意思：

1. 如果item置信区间很宽（被选次数很少，还不确定），那么它会倾向于被多次选择，这个是算法冒风险的部分；
2. 如果item置信区间很窄（备选次数很多，比较确定其好坏了），那么均值大的倾向于被多次选择，这个是算法保守稳妥的部分；
3. UCB是一种乐观的算法，选择置信区间上界排序，如果时悲观保守的做法，是选择置信区间下界排序。

给UCB插上特征的翅膀

UCB还是很有前途的，所以算法大神们还是有心提携它一把。

这不，Yahoo!的科学家们在2010年发表了一篇论文[1]，给UCB指了一条明路，同时还把改造后的UCB算法用在了Yahoo!的新闻推荐中，深造后的UCB算法现在title叫LinUCB。

