

北京大学第十五届“江泽涵杯”数学建模竞赛题目

请先阅读竞赛规则

<http://portal.math.pku.edu.cn:8000/news/read.php?newsid=13231>

A 题：如何对排行榜类产品的排序结果好坏进行衡量？

「如果一件事不能被测量，就不能被改进。」

搜索引擎主要关注用户是否在一次搜索中获得了想找的信息、用户想找的信息是否排在搜索结果页靠前的位置，因此会使用用户在搜索结果页「有点击行为发生的比例」「点击条目在结果中的排序」「点击一个条目后是否在短时间内点击了其他条目或修改关键词重新搜索」等可统计量化的数据来衡量产品的表现。

信息流内容类产品主要关注展现的信息是否是用户感兴趣的、内容是否能吸引用户在未来频繁回访，因此会使用「展示的内容有多大比例被点击/有效观看」「包含观看每个单条内容的时间在内用户停留的总时长」「在某个时间周期内（如一周）用户重复使用的频次」等可统计量化的数据来衡量产品的表现。

以上例子中给出的指标与产品的具体实现方式无关。在设计衡量产品整体表现的指标时，应当把产品的实现方式视为黑箱。如果进行 A/B 测试，无需关注试验与对照方案分别采用的具体交互设计、技术实现、推荐算法，只依据这些核心指标的差异，就可以对产品方案的优劣做出评价。使用这些指标，甚至可以对不同公司制作的同类产品进行比较。

请选择若干款你常用的排行榜类互联网产品，为其设计一套量化数据指标，来衡量这些产品的表现。需要对指标给出清晰的描述和定义。

比如大众点评会对餐厅按照人气、服务、口味等纬度进行排序¹，豆瓣会对电影按照评分、热度等维度进行排序²。这类排序没有绝对客观的标准，使用不同时间窗口的数据、赋予不同用户不同的权重，就会得到不同的结果。另一方面，当某个商户或影片上榜后，得到更大的曝光，其热度和评价也会变化，对原本的排序依据产生正/负反馈。我们想知道，让哪些条目上榜，让谁排在前面，能够带来更大的用户价值。如果可以设计一套量化评价指标，衡量排行榜的好坏，就可以使用 A/B 测试等常用方法辅助改进了。

¹ 参考网址: http://www.dianping.com/shoplist/search/2_10_0_popscore

² 参考网址: <https://movie.douban.com/explore#!type=movie&sort=recommend>

在指标设计的过程中可以使用用户的浏览和点击行为、产品元素在界面上的展示和交互，以及其他任何可以通过技术手段记录的数据。也可以为了使结果可衡量，在不严重影响使用体验的情况下，对产品本身做出必要的调整。