

## 010 | 数据科学家高阶能力之评估产品

2017-10-25 洪亮劼

AI技术内参

[进入课程 >](#)



讲述：初明明

时长 11:15 大小 5.16M



**“如果你无法衡量它，你就无法改进它。”**（If you can't measure it you can't improve it.）这是一句你可能会经常听到的话，这句话也被应用到很多不同的场景中。那么，对于人工智能工程师和数据科学家来说，这句话其实是他们工作核心的核心。不管是模型和算法，还是产品迭代，都离不开“指标”和“评估”这两个方面。

评估一个产品的好坏，是一项说起来最容易但做起来最困难的工作。任何人，从用户到产品经理，对某一个产品都可能有自己的主观意见。然而对一个产品，特别是要面对成千上万用户的产品来说，依靠主观感觉是很难有一个完整、全面的评价的。同时，**有一个成熟的产品评价体系可以成为产品不断迭代的领航标。**

今天，我就来聊一聊如何评估一个数据驱动型产品，又如何从评估产品的角度来推动产品的迭代。我们需要建立层次化的评估体系，需要一个衡量产品好坏的框架。这个框架要从宏观

到微观，能够对你的产品进行全方位的检测，并且这种检测能够帮助你更容易地进行决策。

## 产品的经济收益

你可能要问，是的，我们需要评估一个产品，但是如何找到衡量产品的这些指标呢？

比方说你要做一个社交网络的网站，怎么来制定检测指标呢？首先，你要问自己，我做这个社交网络的最终目的是什么？很明显，一个商业网站的终极目标是赚钱，也就是说，你最终的指标是你网站的经济收益。知道了这一点远远不够，你至少还需要思考两个问题。**第一，如何衡量你的经济收益；第二，你能否用经济收益来直接指导你的产品构建。**

我们先谈谈第一个问题。衡量经济收益看似简单其实不易。从比较大的维度上来说，你可以衡量总收入，你也可以衡量利润，你可以衡量收入的年增长率，还可以衡量季度增长率。从比较具体的维度来说，很多社交网站依靠广告收入，对广告收入的衡量本身就是一个非常复杂的问题。

总体来看，衡量经济收益，有两点值得你思考。其一，如何衡量你收入的现状。其二，如何衡量你收入的增长。今天，关于收入的指标我就不展开讨论了。

刚才讲的第二个问题就更加复杂。衡量经济收入的指标往往太过宏观，而且衡量起来有难度，因此用经济指标来指导产品的发展是很困难的。我刚才说了一些经济收益指标，比如年收入、年增长率、季度增长率，这些指标的衡量需要至少等待一个季度以上，甚至一年的时间。这些有时间间隔的指标，无法给产品的快速迭代带来很大的指导意义。

另一方面，很多产品并不直接产生经济结果。也就是说，经济收益是一个“副产品”。这个时候，如果我们只看经济收益，就无法真正指导我们构建更好的产品。比如，我刚才提到，对于一个社交网站来说，广告收入是一个“副产品”，绝大多数用户来到这个网站的主要目的不是点击广告。因此，仅仅衡量广告有可能让社交网络产品的迭代误入歧途。

## 层次化的评估体系

如果单从经济指标无法对产品有全面的指导作用，那怎么才能更加有效地建立评估体系呢？这里就引出下一个话题，那就是多层次的评估体系。

接下来，我就由低到高依次从五个层面来说明一下，这个层次化的评估到底是什么意思。

**最低层次的评估主要围绕着产品的最小组成单元。**比如我们刚才用的社交网络的例子，社交网络的各个页面上的模块就可以是最小的被评估的单元。

为什么要用这个概念呢？原因是这样的，每一个模块往往是产品的一个逻辑单元，一个最小的承载产品理念的单元。不管是工程团队还是产品团队的运作，基本上都是为这些模块而工作。因此，观察最小单元的效果对产品和工程团队都有直接的指导意义。如果团队目前对这个模块做了一些更改，那么最直接的效果就是这个模块的一些指标会发生变化。这是产品迭代的一个重要组成部分。

在这个层次，衡量模块的指标主要是模块的直接效果指标。比如，模块本身的点击率，模块本身的驻留时间，模块上一些其他的用户活跃指标等。这些都是最低层次的模块级别的指标，和产品、工程团队的运作有密切联系。

**第二个层次的指标是从单个模块上升到一个页面。**这个时候，就不仅需要理解单个模块的情况，还需要对整个页面上所有模块产生的功能群进行深入研究。在这个层次，产品功能群的思考可能会涉及到多个产品团队，也可能会出现模块间冲突的情况。

比如不少现代搜索引擎的搜索页面往往都有广告模块。长期的经验告诉我们，广告模块的效果和普通搜索模块的效果往往有相反关系的耦合。也就是说，普通搜索模块的效果提高了，广告模块的某些指标反而可能下降。反过来，广告模块的效果提高了，也很有可能是因为普通搜索模块的质量突然变差。因此，在有经验的产品团队面前，广告效果有意想不到的提高，可能并不意味着是件好事情。

第二个层次的指标比第一个层次变得复杂起来。不过这个层次的指标依然是可以直接测量的。比如页面的点击率，页面的驻留时间，页面上其他的用户指标等等。这些指标虽然可以直接测量，但是分析时需要对面上的所有模块有全面了解。

前两个层次的指标主要是测量用户在某一个模块或者页面上的表现，核心是看产品的更改对用户的直接影响。而且，第一层次和第二层次的指标非常易于检测。通常情况下，如果页面和模块发生了什么问题，这些指标就能很快地反映出页面的情况。然后通过排查，我们就能快速发现问题，这也就是通常所说的，这些指标都比较“敏感”。

“敏感”指标的第一个好处是，这些指标具有非常强的指导意义，能够帮助产品团队快速认识问题并提出解决方案。“敏感”指标的第二个好处无疑就是，产品团队的绝大多数改动都能够比较容易地反应到这些指标上。因此，这是一个容易建立的、良性循环的指标体系。当

然，仅有这两个层次的指标还是远远不够的，我们可以看到，这两个层次的指标和一个产品最终目标的衡量还有一定距离。

**第三个层次的指标，就从某一个模块、某一个页面上升到了用户这个层级，主要是检测用户在一个会话（Session）中的表现。**这个时候，用户往往在一个会话中，和多个模块、多个页面进行非常复杂的互动。在这个层次上，我们已经很难仅凭观测就能琢磨出用户在这个会话中是否真正感觉满意。这个时候，我们往往就需要建立用户模型（User Model），以及通过一些统计的方法建模，从而实现真正理解用户行为的目的。

举一个例子，如果我们构建一个电子商务网站，在一个用户会话中检测用户是否购买了一些商品，这些商品的总价值又是多少。这个监测指标有时候被称作 GMV（Gross Merchandise Value），也就是通常所说的网站成交金额。GMV 还是比较容易计算的，就是计算每个会话之后用户购买的商品价值，然后对所有会话的结果求和。但是要真正理解用户会话行为对 GMV 的影响，就是一个比较困难的任务了。

我们可以想象，即便是同一个用户，是否在一个会话中购买商品，这是一个非常复杂的决策过程。在一个会话中，用户可能会接触到搜索页面，可能会接触到各种推荐的模块，也可能会跳转某个商品的页面，还可能会跳转首页。并且，每个用户的用户轨迹不同，接触各个页面和模块的流程也是不一样的。可以肯定地说，任何一个流程中的每一个环节，都有可能对用户是否购买商品、以及购买多少价钱的商品有或多或少的影响。而如何来测量和建模这样的影响，就是第三层次指标的核心挑战。

**第四个层次的指标是从一个用户会话上升到多个用户会话。**这个时候，我们关心的是用户较长时间的体验问题。对于一些复杂的任务，用户需要多个会话才能完成。套用我们刚才举的电商 GMV 的例子，很多用户购买比较贵重或者是一些有特定需求的商品（比如婚纱）时，往往无法在一个用户会话中完成决策。

那么这种情况下，检测指标的复杂性又进一步提高。比如说，用户可能先在电商网站上搜索了关于婚纱的信息，但在这一次会话中并没有完成交易。用户之后可能又从其他途径了解了一些更多的信息，然后又重新到电商网站开始一个新的会话。在这个会话中，用户也许重点比较了好几个婚纱，然后决定购买其中一件。这个例子还是一个比较简单的情况了。

第三和第四层次的指标有两个特点。第一，相对于第一、第二层次的指标而言，这些指标已经不那么“敏感”了。也就是说，仅改变某一个模块甚至某一个页面，是很难在短时间内改变第三，特别是第四层次的指标的。从上面的例子可以看出，用户的购买行为是非常复杂

的，仅仅因为提高了某个推荐模块，是不是就能让用户多买贵的东西，答案是不确定的。第二个特点就是，第三和第四层次的指标依然可以用传统的 A/B 测试来进行观测，只不过需要很仔细地设计实验。

**第五个层次的指标就是用户和产品的长期指标。**我们最开始提到的经济指标其实就是第五层次的指标。类似的指标还包括月活跃用户、年度活跃用户等等。这些指标有两个特点。第一，这些指标往往是产品的终极目标，一般极其难以撼动，特别是对于成熟的产品而言。第二个特点是，这些指标往往无法通过 A/B 测试进行衡量。也就是说，我们往往无法通过常规的实验就能够观测到这些长期指标的变化，这也是为什么这些指标被称为“长期”的原因。

## 小结

今天我为你讲了数据科学家和人工智能工程师如何评估产品的能力，这属于比较高阶的分析问题的能力。一起来回顾下要点：第一，我们如何来认识衡量产品经济收益这件事情。第二，我们很详细地阐述了什么是五个层次的评估体系，以及这个评估体系每个层次的特点。

最后，给你留一个思考题，如果第五个层次无法直接通过 A/B 测试进行观测，那我们如何在平时进行 A/B 测试的时候，就能确保是在优化第五个层次的指标，也就是我们产品的终极目标呢？

欢迎你给我留言，和我一起讨论。

---



# 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 009 | 数据科学家高阶能力之分析产品

下一篇 011 | 数据科学家高阶能力之如何系统提升产品性能

## 精选留言 (3)

写留言



劳永超 置顶

2017-10-25

👍 5

洪老师你好，我觉得你的这个话题很好，你的内容我一般都是开车的时候听，朗读质量也不错。这次最后的思考题，如何在ab test的时候确保对产品终极指标的正向影响，我认为还是要基于先验经验来做，ab test可以看出对短期敏感指标的影响，例如次留、点击率、某页面转化率、崩溃率等，这些指标对长期指标的影响有一些是显而易见的，有一些可通过历史数据来看关联性。

展开 ▾



沉

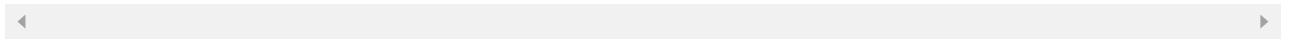
2017-10-27

👍 2

这篇非常有用，我们的产品就是一个付费订阅的模式，之前对付费用户有做过一些简单的

分析和统计学习，但没有一个全面的产品分析意识，现在看来还有很多东西可以做  
展开 ▾

作者回复: 谢谢。



晏

2018-07-16



最好有一个产品评估体系的5层次模型图。

展开 ▾