

从零开始，构建数据化运营体系

数据化运营是一个近年来兴起的概念，它在运营的基础上，提出了以数据驱动决策的口号。

在了解数据化运营前，运营们有没有过如下的问题：

1. 不同渠道，效果究竟是好是坏？
2. 活跃数下降了，到底是因为什么原因？
3. 这次活动推广成效如何？
4. 发布了版本，用户喜不喜欢？
5. 我们总是说传播，传播到底有多大？

这是产品和运营每天每时每刻都会遇到的问题。数据化运营，实际以解决这些问题为根本。它从来不是 BAT 的专属，也不是大数据的独宠，每一家互联网公司，都有适合的数据运营土壤。

数据运营体系，是数据分析的集合与应用，也是数据先行的战略，它不仅是运营人员的工作，也是产品、市场和研发的共同愿景。从管理角度，是自上而下的推动，如果领导不重视，那么执行者数据用得再好，也是半只腿走路。

如何构建数据化运营体系呢？以下是我的总结思考。

我将数据化运营体系划分成四层架构，每一层架构都逐步演进互相依赖，每一层又不可缺少。这四层分别是数据收集层，数据产品层，数据运营层，用户触达层。它是以运营人员为视角的框架。

数据收集层

数据化运营体系的底层是数据收集，数据是整个体系中的石油。



数据收集的核心是尽可能收集一切的数据，它有两个原则：宜早不宜晚，宜全不宜少。

- **宜早不宜晚**，意思是产品从创立阶段，就需要有意识的收集数据，而不是等到公司发展到 B 轮、C 轮才去收集。数据化运营贯彻产品全阶段，不同阶段有不同的运营方法。

高价值资源分享社群——清晰洞察世界

存有**20万份+**报告与资料

为您的创业、投资、职场、学习研究… **超强赋能**

【资源属性】

运营、产品、增长黑客、营销、数据分析、案例拆解、行业报告、趋势研究、白皮书、电子书…

【覆盖行业】

教育、电商、直播及短视频、内容产业、泛娱乐产业、保险及金融、投资、人工智能、智能硬件、AR/VR、零售、汽车、社交、社区、社群、母婴儿童、区块链、物联网及产业互联网、B端及企业服务、跨境海外、地产、婚恋美妆、运动体育、餐饮美食、工具类…



扫码关注公众号（或添加微信：zifumeng001）

加入微信交流群 获取最新资料



万千资源合集的知识店铺

- **宜全不宜少**，指的是只有不合适的数据，而没有烂数据。像历史数据、变更记录或者细节处的数据，都存在价值。

举一个例子，有一家金融产品，它的征信系统会详细记录用户的行为，用户在借贷时上传担保资料，会记录用户在这些页面的操作步骤和时间。这里有一个假设，上传担保资料普通人一定是谨慎小心的，如果这步骤完成的非常顺畅快速，很可能是会违约和欠款的人群：你操作那么溜，是不是想捞一笔？属于熟练工作案。征信系统会把这些数据作为特征判断风险。

需要收集的数据能划分成四个主要类型：行为数据、流量数据、业务数据、外部数据。

行为数据

它是记录用户在产品上一系列操作行为的集合，按时间顺序记录。用户打开 APP，点击菜单，浏览页面是行为；用户收藏歌曲、循环播放歌曲，快进跳过歌曲是行为。

行为数据的核心是描述哪个用户在哪个时间点、哪个地方，以哪种方式完成了哪类操作。

我们可以利用其分析用户的偏好，页面停留时间的长短，浏览的频繁程度，点赞与否，都可以成为依据。另外一方面，用户行为也是用户运营体系的基础，按不同行为，如购买、评论、回复、添加好友等，

划分出不同梯度，定义核心用户、重要用户、普通用户、潜在用户的分层。

行为数据通过埋点技术收集。埋点有不同种的实现方式，采集到的数据内容倒是没有差别，主要以用户 ID，用户行为，行为时间戳为最主要的字段。用表格画一个简化的模型：

userId	active	timestamp
1	浏览首页	2017/1/22 14:0
1	浏览A商品页	2017/1/22 14:
1	点击收藏	2017/1/22 14:
2	浏览首页	2017/1/23 16:0
2	浏览B商品页	2017/1/23 18:2
2	浏览首页	2017/1/25 19:0

userId 用来标示用户唯一身份，通过它来确定具体是谁，理解成身份证号就行。active 就是具体操作的行为，需要在技术层面设置和定义，timestamp 就是发生行为的时间点，我这里只精确到分，一般会精确到毫秒。用户的行为记录应该详细，比如浏览了什么页面，此时页面有

哪些元素（因为元素是动态的，比如价格），它是半结构化的 NoSQL 形式，我这里简化了。

有时候为了技术方便，行为数据只会采集用户在产品浏览的页面，像点击、滑动这类操作不记录。属于折衷的方法。

除此以外，行为数据还会记录用户设备、IP、地理位置等更详细的信息。不同设备的屏幕宽度不一样，用户交互和设计体验是否会有差异和影响，怎么拿来分析？这也是数据化运营的应用之一，是宜全不宜少的体现。

流量数据

流量数据是行为数据的前辈，是 Web1.0 就兴起的概念。它一般用于网页端的记录，行为数据在产品端。

流量数据和行为数据最大的差异在于，流量数据能够知道用户从哪里来，是通过搜索引擎、外链还是直接访问。这也是 SEO、SEM 以及各渠道营销的基础。

虽然现在移动时代，Web 时代的流量数据并不过时。比如微信朋友圈的内容都是 HTML 页面，活动运营需要基于此统计效果，我们可以把它看作一类流量数据。另外，不少产品是原生+Web 的复合框架，内置的活动页大多通过前端实现，此时即算行为，也算流量数据，当我们将

活动页发送到朋友圈时，相应的统计只能依赖基于前端的流量数据来采集了。

流量数据是基于用户访问的网页端产生。主要字段为用户 ID、用户浏览页面、页面参数、时间戳四类，简化模型如下：

userId	url	param	timestamp
1	首页	source=baidu&city=上海	2017/1
1	一级目录A	city=上海	2017/1
1	二级目录B	city=上海	2017/1
2	首页	source=SEM&city=北京	2017/1
2	一级目录C	city=北京	2017/1
2	首页	city=北京	2017/1

url 是我们访问的页面，以 `***.com/***` 形式记录，param 是描述这个页面的参数，我们在页面上的搜索、属性信息会以参数的形式记录。和行为数据一样，如果流量数据需要更详细的统计，也是以半结构化为佳，囊括操作记录。

它是活动及内容运营的好基友，活动的转化率，文章被发到朋友圈的阅读量等，都是作为流量数据被记录。主要通过 JS 采集。

流量数据的统计已经比较成熟，Google Analytics 和百度统计都是知名的第三方工具，最为常用。不过它们不支持私有化的部署，只能提供统计，我知道这个页面有 100 人访问，但这一百人是谁不能定位，数据也无法记录在数据库中，这对数据化运营是一种麻烦。一些新式的工具则能支持这种更精细的需求，不过要收费。

如果有可靠和先进的技术手段，我们是能做到将行为数据和流量数据统一到一起，这是未来的趋势。

业务数据

业务数据在产品运营过程中伴随业务产生。比如电商产品，我进行了促销，多少用户领取了优惠券，多少优惠券被使用，优惠券用在哪个商品上，这些数据和运营息息相关又无法通过行为和流量解释，那么就归类到业务数据的范畴。

库存、用户快递地址、商品信息、商品评价、促销、好友关系链、运营活动、产品功能等都是业务数据，不同行业的业务数据是不一样的，业务数据没有固定结构。

业务数据需要后端研发进行配置，因为结构不能通用化，最好提前和研发们打声招呼提下需求。

行为数据、流量数据、业务数据构成了数据来源的三驾马车。统称为原始数据，指没有经过任何加工。

外部数据

外部数据是一类特殊的数据，不在内部产生，而是通过第三方来源获取。比如微信公众号，用户关注后我们就能获取他们的地区、性别等数据。比如支付宝的芝麻信用，很多金融产品会调用。还有公开数据，像天气、人口、国民经济的相关指标。

另外一种外部数据的获取方式是爬虫，我们可以爬取豆瓣电影评分、微博内容、知乎回答、房地产信息为我们所用。第三方不可能支持你获取，很多时候会有防爬虫机制。它需要一定的技术支持，不属于稳定轻松的来源。

外部数据因为质量难以保证，更多是一种参考的作用，不像内部数据能产生巨大的作用。

这四类数据构成了数据化运营的基石。随着互联网公司数据化水平的提高，能够利用的数据越来越多。数据结构逐步从 SQL 到 NoSQL；信息来源更加丰富，图形和声音数据越来越多；技术由单服务器演变成分布式；响应从离线批处理到实时流式，都是数据收集的挑战。

当我们有了数据以后，进入下面一层，数据产品层。

数据产品层

数据产品是对数据的加工和利用，它属于技术和自动化的范畴，由计算机对原始数据进行处理。它不是传统意义上的数据产品（如广告系统），而是以发挥数据价值和生产力为目的，理解成进行数据加工的产品也可。



原始数据并不能直接为运营所用，通常脏乱差，我们需要按照一定的标准整合、加工。

比如行为数据和流量数据，用户在微信朋友圈看到一则活动觉得不错，于是下载 APP，注册后参与了活动。这里的行为数据和流量数据是完全独立的。微信朋友圈的浏览，记录的是用户 weixinOpenId 和 cookie，下载后则是产品内部使用的的 userId，两者无法对应，这就需要数据整合，将 cookie、手机号、userId 等信息映射(mapping)到同一个人。

这是技术层面的数据清洗。整个过程叫做 ETL。

数据发挥价值的方式有很多种。即能通过 BI，将原始数据以维度和数量的方式聚合，进行各类可视化的决策分析，也能数据挖掘。根据业务和场景决定数据的不同使用。这里最重要的是先有指标。

数据指标

我强调过尽可能的收集数据，然而原始数据那么多，怎么才能指导我们的业务呢？这要求我们从庞大的数据中找出方向。这时我们就要建立指标，指标就是我们的方向，它是业务和原始数据的连接器。



可以这样说，指标在数据化运营体系中是承上启下的润滑油，它由原始数据加工而来，反过来又驱动其他产品。

需要有 BI？BI 肯定是围绕指标建立仪表盘；要用机器学习算法？算法的目的就是提升指标效果的；你要运营？内容、用户、活动模块的 KPI 也是围绕指标的。

指标不是一个通常意义的产品，我更喜欢的解释，是数据届的产品经理，是驱动、规划其他数据产品以及配合运营迭代业务的。这样一说，大家就明白了。

指标如何设立，是根据运营业务所决定，也是运营的第一驱动力。

我们简单看一下指标如何由原始数据加工而来，下图是原始数据中记录的用户打开 APP 的情况。

userId	timestamp
1	2017/2/1 10:03
1	2017/2/2 14:10
1	2017/2/2 15:52
2	2017/2/2 16:00
2	2017/2/2 18:27
2	2017/2/4 19:00

每一个时间戳意味着对应的用户打开过 APP 一次，通过该表我们能计算每天有多少用户打开过 APP，这是打开量，将用户数去重，就是运营中的重要指标：活跃用户数。通过对该表的进一步复杂运算，譬如用 SQL 的 Left Join，能获得留存率。

文章阅读量、日销售额、活动参与人数，这些几乎都是由原始数据汇总加工而出。指标汇总以后，就是运营人员产品人员每日的报表 Dashboard。

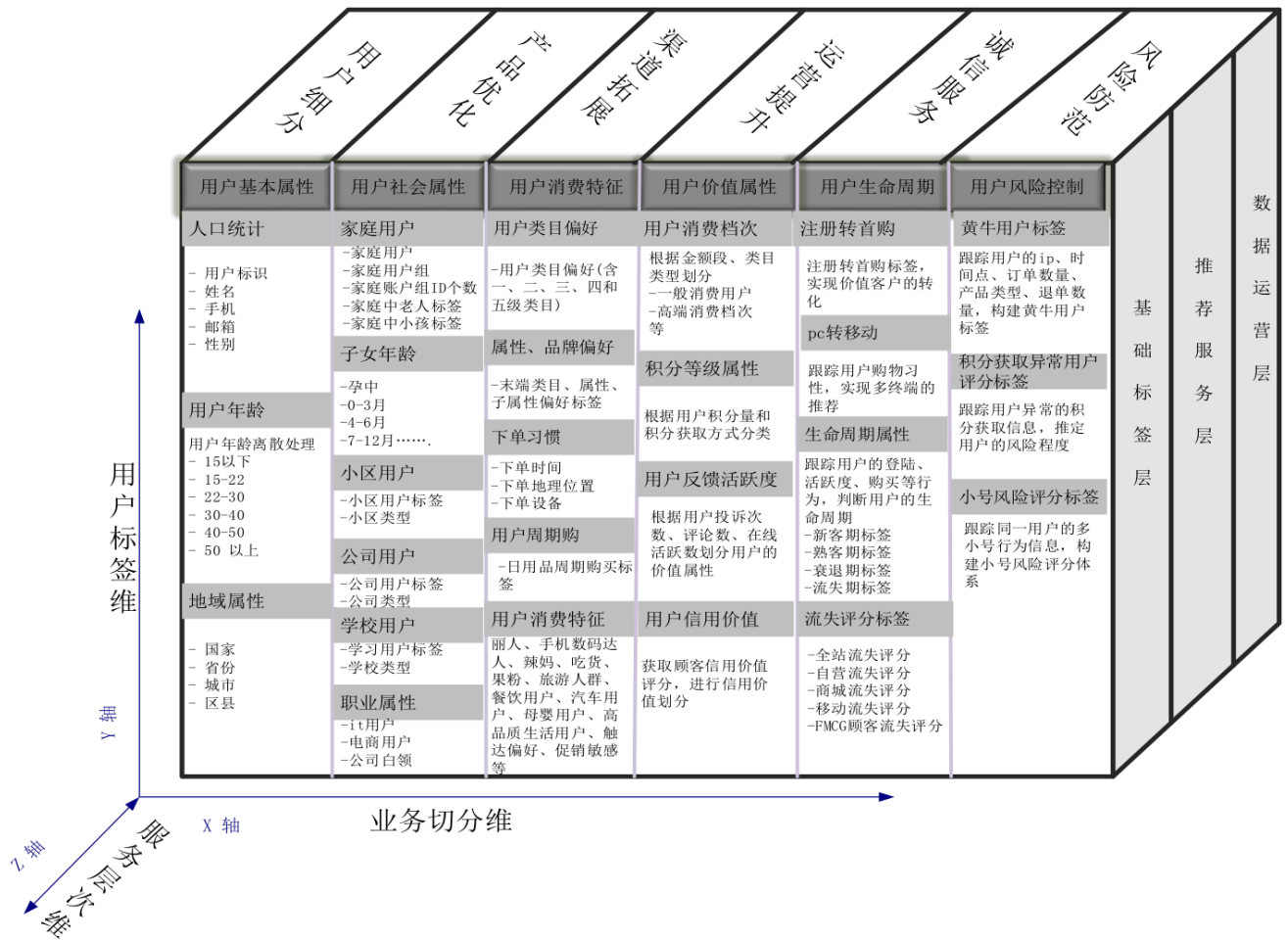
有了指标，我们再看其他的数据产品，因为篇幅有限，我着重介绍一下用户画像。

用户画像

用户画像是常用的数据产品，对产品和运营人员往往带有神秘色彩。它有两种解释，也是很多新手歧义根源，一种用户画像属于市场营销和用户调研领域，叫做 Persona，更准确的翻译是用户角色，描绘的是一个自然人的社会属性，用于用户需求和场景的确定。

而数据领域的用户画像，叫做 Profile，是将一系列数据加工出来描述人物属性的数据标签。最知名的例子就是淘宝的千人千面：用户去购买孕期的孕妇产品，很大可能被打上孕妇标签；浏览了汽车相关商品，会被打上汽车兴趣的标签。

用户画像是一个依赖大数据和机器学习的复杂体系。准确丰富的用户画像能呈指数级的提高运营效果。



用户画像也有简单的用法，没有数据挖掘不要紧。用户的性别、年龄、地区这些信息不难拿到吧？用户行为简单做一个喜爱偏好区分也不难吧。那么我们就有用户画像 V1.0 了。

推荐系统，精准营销、广告投放都是常见的基于用户画像的应用。你要推送化妆品促销活动，选择女性标签的用户肯定有更高的成功率，更进一步，如果运营知道女性用户偏好哪个品类的化妆品，效果会更好。

用户画像可以通过已有数据提炼获得，比如拥有用户的身份证信息，就能准确获得性别、籍贯、出生年月这三个标签。也能通过算法计算

获得，比如在淘宝购物遗留的收件人姓名，通过机器学习，以概率的形式获得买家是男是女，建国很大可能是男性，翠兰很大可能是女性。

用户画像是基于原始数据的加工，原始数据越全，用户画像就越丰富。

数据产品层中，我们将数据加工为指标，以其为核心，构建和规划数据产品。如何展现指标（BI），如何提高指标（算法），如何计算出指标（ETL），如何与指标组合（用户画像）。

我们现在获得了这些「产品」，接下来就是使用，运营和产品人员就是它们的用户。

数据运营

数据运营层，是运营人员将数据转化成运营策略。以人为主要生产力和数据产品的计算机自动化对应。



在我们谈及具体的方法前，强调一下人的作用。不论我们前面打造了多好的数据产品，员工的数据化运营意识提高不上去，一切等于零。

对人的要求有三点：

(1) 以数据做决策，既要知道数据能够做什么，也要知道数据做不了什么

前者很容易理解，我工作中遇到很多次，在有数据可以提供决策的情况下，依旧相信个人经验。这是应该规避的思维，不是一个人，而是团队要做到。

数据化运营也不是企业运营的灵丹妙药，得客观承认，公司体量越大，数据化运营所能发挥的效果也越好。在创业公司或者小公司，会受到一定的限制，比如没有技术支持，提升效果不够，数据体量缺乏等原因，造成优先级的延后。这是没办法的取舍问题，只能以解决问题为首先依据。

(2) 本身数据分析和运营水平不过关

虽然有意识地利用，可员工仅限于求平均数的水平，那么也别期待太高了。

这一点，得通过不断地系统培训，人员招聘解决。自上而下的倡导和发起是最好的结果，高层有数据化运营的战略和意识、管理层有数据

化运营的指导经验，执行层能将数据化运营的落地，那么整个体系也推行成功了。

(3) 产品工具的使用

这是对员工的技能要求，诸如 MySQL 查询数据、BI 多维度分析、精准营销、AB 测试、转化率分析，都是必须的。将数据相关的工具玩得顺溜，员工才能在发挥够大的价值。

运营和产品如何进行数据运营，具体的技巧和方法论太多了，我以核心思想为引子。大家着重了解思维。

1、不是全量，而是精细；不止精细，更是精益

全量运营是一种集中运营的策略，活动、内容推送、营销、用户关系维护，这些方式如果针对所有的用户，这是运营资源的浪费，你不可能通过一种方式满足所有的用户，也不可能用一种方式做到最好。

用户间是有差异的，这种差异需要用精细化运营弥补。

精细是是将目标拆分成更细的粒度，全国销量变成上海销量北京销量、全年销量变成第一季度销量第二季度销量，用户变成新用户老用户。电商卖口罩，是卖给北京的用户好，还是海南的？促销化妆品，目标人群选择男人女人也是显而易见的。精细(拆分)是一种数据分析的思路，也是一种运营手段。

精益比精细更进一步，精细是手段，精益是目标。什么是精益？精益就是二八法则，找出最关键的用户。我们都知道要将化妆品卖给女人，但一定会有部分女人支付更多，20%的女人占了80%的销量，精益就是找准这20%。

对最适合的用户在最恰当的时机采取最合适的手段以产生最大的价值。

前面三个「最」说的是精细，后面一个「最」指的是精益：价值 / 目标最大化。我有CRM，那么就从CRM中找出最有价值的客户去维护；我有风险管理，就找出最可能违约的投资；要做活动，欢迎的是产出最大而不是薅羊毛的用户；积分中心，效果最好的只会是最优质的那批客户。

2、未来比现在重要，现在比过去重要

这个第二个核心，数据化运营能够预测未来，把握当下。传统的运营方式，是知晓过去已经发生的事，销量是多少，活跃数是多少，这在日益严酷的竞争环境中还不够。

把握当下，是能获得数据的立即反馈。你要推广一个活动，可以提前挑选5%的用户做一个测试，及时获知用户的反馈，转化率高不高，响不响应，然后按照数据决定后续运营是继续还是改进。这是技术带来的进步优势。

预测未来，是机器学习的领域，通过数据建模，获得概率性的预测，用户可不可能流失，会不会喜欢和购买这个商品，新上线的电影会否偏好…运营则利用这些概率针对性的运营。

如果限于技术无法使用机器学习，则需要根据现有数据趋势去估计，这取决于运营人员的经验和数据敏感性。

3、系统化与自动化

数据化运营体系的搭建过程中，运营人员会用到很多的工具。

用户积累到一定数量，我们考虑引入积分中心增加用户粘性；产品涉及到地推和销售人员，则要加入 CRM（客户关系管理）以维系客群；O2O 和电商，基本配置肯定有优惠券的发送；反馈越来越多，我们也需要客服中心解决各类疑问。这些与运营息息相关的工具，在数据运营体系中占据中重要的比例。

为了更好的达成目标，会将其独立成运营模块 / 运营后台。好的运营后台和用户端的产品同等重要，也需要后台产品经理规划。

以我们经常接触的优惠券为例，它肯定要设置一套规则，核心目标是财务数据，是优惠券成本和收入之间的平衡：你不能滥发，那肯定亏钱，也不能少发，用户连这东西都不知道。有哪些券、怎么发、发了多少用了多少、未来准备发多少、发了有多少没用掉，都是一套大框架的东西，于是做成了发券系统。

优惠券能和 CRM 结合，CRM 通过几个指标将用户划分成了不同的价值和人群。这个用户特别喜欢花钱，那么优惠券给他满 1000 减 100，肯定比满 200 减 20 过瘾。那个用户还没有消费过，要用首单优惠刺激他。还有用户有段时间不消费了，运营们得加把劲营销。上面东西从更高的视野看，是一连串效果、ROI、盈利的评估。这就是用数据做运营策略。

CRM 又能和客服中心结合，电话号码肯定和用户的数据绑定，VIP 用户电话进来了，我们选客户主管去接待，宾至如归。普通用户呢，也不能粗心，客服至少需要通过后台的用户画像知道这个用户是什么情况，这也有针对性的服务。数据运营体系不止服务于运营和产品的。

系统化，要求的是我们把运营的整个过程和策略流程也当作一款产品去缔造：哪些方法好用，哪些手段效果好，哪种活动能持续做，把这些都固定下来，打造出一个运营用的产品后台，作为日常和招数。这种系统化思维也叫「复用」，之后则是把系统做得越来越自动，功能越来越强大，也是另外一种精益了。

以上种种，是将数据、产品运营、系统和人员四者结合起来。系统之所以是系统，就是脱离了粗放阶段，一切皆是有序、规则和充满策略。数据就是系统的润滑剂，你没有数据，怎么能有选择性的发券、做活动、推送，维护用户呢？

数据产品层加工出来的各类标签、用户画像、模型…就是要在数据运营层最大化的被员工使用。数据本身没有价值，变成策略才有价值。

这三条要点总结一下：我们系统化的使用各种加工后的数据，以精细和精细为手段目标，以把握未来为方向，指定运营策略。这是数据运营层的核心。

用户触达

我们整个体系进行到最后的环节，它需要面向用户。数据收集得再多、加工得再好，运营得再努力，如果不将它们传递给用户，体系就是失败的。

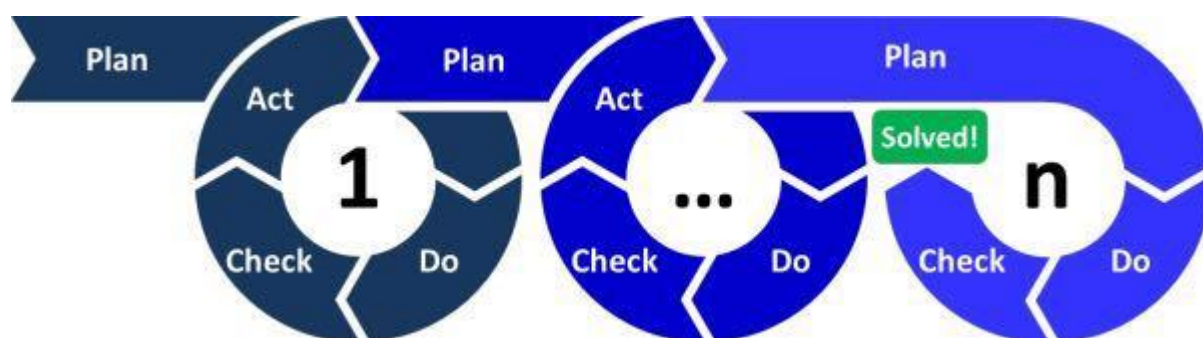


整个体系的前三层用户都感知不到。用户直接感知到的是产品的推送通知、Banner、广告位、活动、文案、商品的展示顺序等。在与产品交互的过程中，用户会以直接的反馈表达自己喜恶。

感兴趣的会点击，喜爱的会够买，讨厌的会退出…这些构成了新一轮的行为数据，也构成了反馈指标：点击率、转化率、跳出率、购买率等。这些指标就是用户触达层的结果体现，也是数据化运营的结果体现。

好与不好，都需要验证。

结果不是终点。管理学有个概念叫 PDCA，翻译成中文是计划-执行-检查-改进，以此为循环。用户触达层不是数据化运营体系的结束，它是另外一种开始。通过反馈获得的数据去优化去改进。



我的点击率 5%，那么我能不能通过运营优化，达到 10%？用户接受推送后选择了卸载，我们有什么方法挽回？留存率被提高，这种策略能不能应用到其他用户上面。

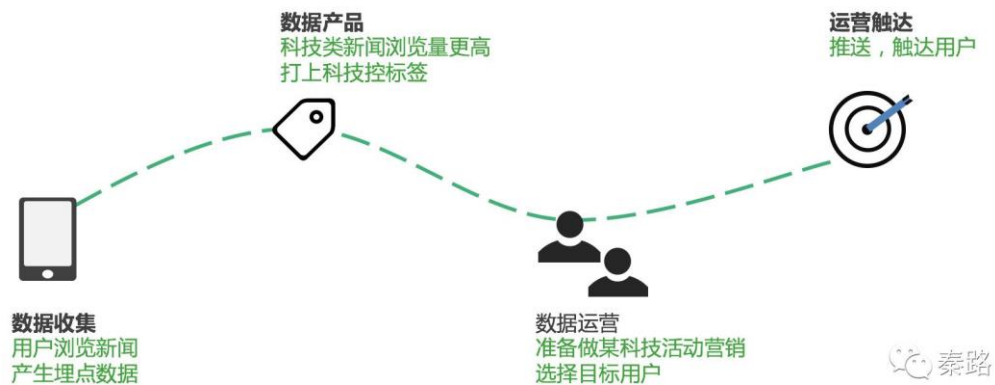
也许我们数据化运营后，不会获得一个满意的结果，但如果我们连优化改进都不去做，那么连好的机会都不会有。

你看，优秀的员工，不会以数据化运营的结果沾沾自喜，而是进行新一轮的开始。

是终点，又是起点，此过程就是迭代，是体系的核心。

总结

我们将四层串联起来看待，下图是一款产品简化的数据化运营闭环。



- **数据收集层：**当用户打开 APP 时，浏览新闻，通过埋点记录用户的行为数据：何时何地是谁看了哪些新闻。
- **数据产品层：**计算机将收集上来的行为数据进行加工，统计用户对军事、科技、经济等不同类型新闻的阅读数。用卡方检验得到用户的阅读偏好在科技新闻，将其写入到用户画像 / 标签系统。
- **数据运营层：**近期有一个科技类的活动，需要一定用户量参与。运营不能选择全部的用户推送吧，那么就从用户池中筛选中对科技感兴趣的用户。
- **用户触达层：**选择用户进行精准推送，用户在手机端接收到消息。后台则会记录用户是否打开推送通知，是否浏览页

面，是否参与了活动。转化率作为反馈会被记录下来，用以下次迭代改进。

该例就是一次合格的闭环。数据化运营体系既能简单到用 Excel 完成，也能引入机器学习数据挖掘分布式系统等高端技术，看的是思维和应用。我们将体系中的四层简化成四个模型，帮助大家理解：

- **数据收集：**以用户和产品的交互为输入，原始数据（行为、业务、流量、外部）为输出。
- **数据产品：**以原始数据为输入，以加工数据（标签、画像、维度、指标、算法结果）为输出。
- **数据运营：**以加工数据为输入，以运营策略（用户、内容、活动、电商）为输出。
- **用户触达：**以运营策略为输入，以反馈行为（转化率、点击率、响应率）为输出。

用户产生的反馈行为作为新的交互输入，迭代和优化，数据化运营体系就良好地运作起来。好的数据化运营体系也是高度自动化的运作，像个性化推荐，可以略过数据运营层，服务器实时计算后直接将推荐结果给用户，人就不用参与其中了。

高价值资源分享社群——清晰洞察世界

存有**20万份+**报告与资料

为您的创业、投资、职场、学习研究… **超强赋能**

【资源属性】

运营、产品、增长黑客、营销、数据分析、案例拆解、行业报告、趋势研究、白皮书、电子书…

【覆盖行业】

教育、电商、直播及短视频、内容产业、泛娱乐产业、保险及金融、投资、人工智能、智能硬件、AR/VR、零售、汽车、社交、社区、社群、母婴儿童、区块链、物联网及产业互联网、B端及企业服务、跨境海外、地产、婚恋美妆、运动体育、餐饮美食、工具类…



扫码关注公众号（或添加微信：zifumeng001）

加入微信交流群 获取最新资料



万千资源合集的知识店铺

这是四个互相联系有先后顺序的系统，以此构成数据化运营体系。因为技术手段差异，实现方式会有不同，哪怕是 Excel，也能发出数据化运营的光芒。

以上就是产品和运营视角的数据化运营体系，没有过多的牵涉研发技术，实际复杂程度还要再高一点。当然，万千用法，存乎一心，希望大家学到的是理念和思维，实际工作中，还是有很多玩法留待大家挖掘。

作者：秦路，微信公众号 ID: tracykanc

本文由 @秦路 原创发布于人人都是产品经理。未经许可，禁止转载。