

用户活跃度模型的搭建及应用

1：用户活跃度模型的搭建及应用；

用户活跃度模型，从建模的目标上就是要把不同活跃度的用户拆分开来。用户活跃度模型的应用场景是：区分不同活跃度的用户，针对不活跃的用户启动针对性的活跃度提升的运营策略，针对活跃用户启动针对性的加强忠诚度，引导带动不活跃用户的运营策略。它的建模搭建的思路是如何来的呢。做运营基础的 KPI 就是要提高 DAU 和 MAU，从做法上来说强调的：开源节流保活跃了。开源就是开注册的源，让更多的新用户进来；节流是节流失的流，让更少的老用户流失；保活跃就是让池子里的用户更加活跃，包括让不活跃的用户变活跃，活跃的用户保持活跃度产生忠诚度等。这三方面的手段中，开源和节流对应着 DAU 和 MAU 的增长，而保活跃对应着 DAU/MAU 的增长。

DAU/MAU 的指标就是一个测量已有用户池子活跃度的指标，其中的 DAU 取当月的每日 DAU 的平均值。如果 $DAU/MAU=1$ ，那么说明用户每天都来，所以 DAU 和 MAU 相等，而这个值的最低线就是 0.03 左右，即所有的用户一个月只来一天。所以 AU/MAU 的数值是一个介于 0.03-1 之间数字，数字越高，活跃度越高。

我曾经给一家 K12 的英语教育机构做商业化的顾问，我问他们目前商业化的困境是什么，是用户规模不够、转化率低、客单价低还是什么呢？CEO 说他们的用户量没有问题，但是转化率低，MAU 有几百万用户，但是实际的购买用户量一直上不去，所以我解决付费转化的问题。作为一个 K12 线上产品，MAU 几百万的量级确实不小，为什么没有转化呢，我决定看看他们的用户活跃度情况，我跟他要了产品的 DAU/MAU，结果这个数字大约在 0.05 左右，这个数字远低于我的预期，要知道这就意味着大量的 MAU 用户在一个月中只来了一天，这一天很可能就是注册当天，而英语学习是一个需要坚持，最好

但是数据出来后，很快就发现，这个模型并不成立。我们都知道，如果要做数据分类的时候，制定的分类标准应该让数据能够相对均匀地分到各个分类中，如果大量的数据都集中在某几个分类下，其他分类下几乎没有数据，那就说明这个分类方式是失效的。

数据分析师的数据反馈后，我非常难过的发现，DAU 的前推一周的用户活跃度基本上集中在 C、D、E 这几类，后面的数据相比前面这三类在量级上差别较大。这表示大量的人一周就来 0 到 2 天。现在回头想，这其实非常合理，除了微信这样用来取代短信电话的 APP，2016 年的 DAU/MAU 数据能够达到 0.77，而同期今日头条是 0.42，网易新闻 0.43，淘宝 0.29，腾讯新闻 0.39，百度地图 0.12。这意味着什么？意味着即便如新闻网站网易、腾讯、今日头条等，用户的使用习惯也到不了两天中来一天，这算到周活跃里面就是平均活跃度一周在 3 天左右。

3、月活跃度模型

所以周活跃度模型显然并不适合微博。所以下一步怎么办呢？当时的下一步就是想，那我必须把前推一周没有登录的用户拆出来，如果一周都不来一次，那是两周来一次？还是三周来一次？还是一个月来一次？所以索性扩展到月活跃周期来看看用户在一个个月中的活跃分布情况，表头和前面周活跃一样的形式，但是拉长了测算周期为一个月，为了方便后续应用，约定不同的自然月都以平均 30 天为记，然后拆出了一天的用户 DAU 中，在前推 30 天中的活跃天数统计表。

这个数据表分析师反馈时，我们判断分布上已经均匀了很多，但是还不能用于应用，因为加上前 30 天没有活跃的用户，一共 31 个分类的用户模型用起来太冗余，不方便通过产品形态去落地执行，我们必须对用户模型进行进一步

的精简和合并，以更好地应用到产品生态的策略上。这个时候我们想到的就是合并同类项。合并同类项能够帮助我们精简目前的 31 个分类，合并同类项的业务逻辑就是将相似的不同分类用户进行合并。那么怎么判断他们相似呢？我们又回到了属性和行为的判断上，但是用户那么多属性特征、那么多行为特征，如果都放上，势必出来的又是一个千人千面的用户画像，无法合并同类项。

所以我们回归到数据分析的一个大原则：业务导向上。所谓的业务导向就是我们的数据分析要数据源于业务，最终的分析结果又能回到业务中去指导业务。很多人觉得这个原则听起来怎么像是废话，别着急，这个原则听起来谁都能懂，用起来还特别好用，但是有些人就是不知道怎么用。在业务导向的原则下，我们考虑建立用户活跃度模型的原本目的是什么？是为了能够将不同活跃度的用户拆分出来，对低活跃度的用户，我们会采取一定的引导策略，可能是加关注、可能是引导其转发微博，来提升其用户活跃度。或者再回溯底层一些，在微博这个产品生态中，我们期待的活跃的优质用户是什么样的？我们其实不仅期待他们每天都来，我们还期望他们每天发微博、转发、评论互动、点赞、围观热门微博、加更多人的关注、拥有更多粉丝等等。

这是我们期待的一个优质的用户画像。在这样的思考下，我们再回到刚才的月活跃模型中，记录不同活跃天数用户的核心行为数据，然后把数据量相近的相邻活跃天数的用户进行合并同类项，最终得出 4-8 类用户类别，这是比较合理的分类模型。而微博也在这样的模型测算中，最终拆分了：低频用户、中低频用户、中高频用户、高频用户和全勤用户。完成了分类之后，我们就可以为用户打上活跃度的标签，每个用户登录后，根据其活跃度标签触发不同的提升活跃度的策略，这就是线上精细化运营的实现思路。总结：用户活跃度模

型是一个非常经典，而且对于达到一定用户体量的产品在进行精细化运营前进行用户分类时最常用的一个模型，也是资本市场最关注的一个模型。

DAU/MAU 的数值代表了这个模型的优劣。当然，并不是所有的产品模型都要追求这个指标越高越好，就像前面列举不同的产品的指标存在差异，并不是微信做的就比今日头条好，而是产品的使用场景决定了它基础的使用频次，像微信这样解决每日用户沟通需求的 IM 产品，使用频次自然会高于用户购物的产品，你总不能要求用户每天像和家人、同事联系沟通一样的频率，必须每天来平台买个东西吧~

所以我们在面对自己的产品的时候，要知道自己的 DAU/MAU 到底多少是健康的，首先设想一下用户对产品的需求频次常规是多少？其次不妨参考一下业内的竞争对手如果有上市公司的，看看他们的 DAU/MAU 大约在多少，做一个对标。之后再来确定对于自己的产品来说，用周活跃度模型好，还是月活跃度模型好。那这里月活跃度模型的初始统计数据表头，不妨大家当作是作业，测试一下自己的理解和掌握程度好了~ 图片是整个文章的详细目录，

目录：	
前言	
用户建模的价值和逻辑	
▼ 用户建模示例1：用户活跃度模型的搭建及应用；	
1、周活跃度模型	
2、月活跃度模型	
总结：	
▼ 用户建模示例2：用户行为模型的搭建及应用；	
1、用户关键行为指标定位	
2、用户单一行为指标定位	
总结：	
▼ 用户建模示例3：用户流失模型的搭建及应用；	
1、流失用户召回	
2、用户防流失	
总结：	
不同用户模型的交叉应用；	
最后的总结：	

4、用户行为建模（一）

用户行为模型的应用层面上有两个方向：

1、 用户关键行为指标定位：我们将不同活跃度的用户的相同行为特征进行对比，其中差异大的指标定义为关键性指标，并对不活跃用户进行该关键性为指标的重点引导，以促进其成为活跃用户。

2、 用户单一行为指标定位：我们通过用户模型拆分出不同用户的行为喜好特征，已知用户喜欢什么，就给他什么，从逻辑上来说这能够促使他更加活跃。 下面我们就这两个应用方向来分别讨论。

1、 用户关键行为指标定位 即我们将不同活跃度的用户的相同行为特征进行对比，其中差异大的指标定义为关键性指标，并对不活跃用户进行该关键性为指标的重点引导，以促进其成为活跃用户。 为第一个应用方向服务的用户行为模型的搭建思路是如何的呢？还是以微博用户为例，我们后续正好可以延续讲不通用户模型的交叉应用。从目标上，我们就是为了了解一个微博的使用用户到底在微博上喜欢做什么？初期，我们把所有微博上用户常规会有行为都列出来，包含：登录、发博、转发、评论、点赞、加关注、粉丝数、互粉数、是否上传头像、是否填写性别、教育信息等等个人信息等，这些还是微博网站的核心行为，没有拆解比如其中用微盘的、看小视频的、在 PC 登录还是移动端登录的、哪里注册来的，等等。我们先聚焦在微博主站的核心行为上。

但是即便这样，也还是行为参数有点多，我们需要有一个判断，哪个行为的差异会影响我们的核心指标？没错，就是影响我们的 DAU、MAU？所以从业务导向原则上看，用户行为的模型，需要帮助我们明确出来，哪些行为的数据差异决定了用户是否活跃。所以我们列出的初始用户行为数据表单是这样的。 见以下图：

当日DAU各频次老用户当前关系和个人资料数据及前推一个月均互动情况数据统计											
日期	用户频次	个人关系数据			个人资料填写比例			个人互动数			
		人均互粉数	人均粉丝数	人均关注数	上传头像比例	填写教育信息比例	人均上月均发博数	人均上月均转发数	人均上月均评论数	人均上月均点赞数
3.24	低频	A	B	C	D	E	F	G	H	I	J
	中频										
	高频										
.....	低频										
	中频										
	高频										
3.30	低频										
	中频										
	高频										
周均	低频										
	中频										
	高频										
数据说明： 1、老用户为注册超过30天的用户，及已经有完整的月使用频次标签记录的用户； 2、低频用户含上个月无使用的当天回流用户，高频用户含全勤用户； 3、个人关系数据取日期当天24点的即时数据，不考虑取消关注等行为影响； 4、个人资料填写数据取截止日期当天24点，用户曾经填写过即记为1，否则为0，记录填写过占该频次用户总数的比例； 4、个人互动数据取该用户从当天的前一天开始回推的30天内的相关行为总量，包含后来删除的行为； 5、数据来源：用户表、用户行为表											

在上述表单中，个人资料填写部分还可以包含性别更改、地域、年龄、职业经历等。

数据出来后，很快我们会发现不同频次的用户，某些行为的数据差异化很小，某些行为的数据差异化很大，这个时候，差异性很大的指标就可以被我们认定为影响用户活跃度的关键性指标。比如，低频用户中上传头像的比例为 20%，而高频用户中上传头像的比例为 80%，则说明上传头像是非常重要的一个关键性指标。（当然还可以进而分析上传头像的用户中，为真实头像的占比如何。）这里展开讲一下，很多产品认为上传头像不过是个装扮小功能，不涉及用户的核心功能使用。殊不知小小的头像具有非常明显的蝴蝶效应。

同样以 SNS 社区为例，上传的头像能够在虚拟社区中让对方第一眼了解这个虚拟 id 背后的喜好，如果是真实头像就有更大概率让真实的朋友认出自己，对于用户更快更好地构建用户关系有着不可磨灭的作用。在 SNS 社区中，用户希望留下更少的个人信息，却希望找到更多的朋友。结果是上来闷头找了半天朋友，却没有被朋友认出来而忽略，默默地流失。而这样的用户由于留下信息太少，产品机制上也很难帮助其找到关系、构建圈子。当然，头像跟活跃留存是相关关系，但是不能确定因果。我们很难说清楚是因为有头像导致活跃，还是活跃导致加头像，或者是其他因素同时促进了活跃和加头像。

但是引导没有上传头像的用户上传头像总不是一个错误的决策。再回到数据表单上，如果低频用户的人均互粉数是 50，而高频用户的人均互粉数是 60，中频用户的人均互粉数是 53，那么说明互粉数并不会对用户的活跃度带来关键的影响，可以从将该行为从我们此次的用户行为模型中拿掉，以后不用再关注用户的互粉数。当然，我上面示例只拿了一周的数据，我们做建模最好能取用更长时间段内的数据表现，至少两个月及以上，数据相对问题才能成为参考，如果数据波动较大，则无法成为建模的基础数据。

至此，通过前面的不同活跃度用户行为基础数据的对比分析后，我们就可以保留一部分关键性的指标，比如上面举例的是否上传头像，舍弃非关键性的指标，比如上面举例的互粉数。

在产品层面的应用上，当我们发现一个打着“低频”标签的用户登录上来，判断到他还没有上传头像，那么二话不说，马上触发引导他去上传头像，越真实越好~

5、用户行为建模（二）

用户单一行为指标定位即我们通过用户模型拆分出不同用户的行为喜好特征，已知用户最喜欢什么，就给他什么，从逻辑上来说这能够促使他更加活跃。为第二个应用方向服务的用户行为模型的搭建思路是如何的呢？这里我要举我参与的，数据相关畅销书《数源思维》作者，前新浪数据中心资深分析师；

用户成熟度模型就是还原微博上用户除活跃度外的核心使用行为，对用户进行行为特征的分类，并根据该行为特征进行专项引导的模型。在搭建的思路，就是提取用户的多种行为，进行归一分析，找到其中最聚焦稳定的行为，作为该用户的行为特征标签。

具体的核算方法非常专业，这里不做详细的展开，我记录当时的计算方法和最终的用户模型分类供大家参考。计算方法：选择用户行为：登录方式、登录、原创、转发、评论、被转发、被评论、加关注、取消关注、被关注 定义：用户成熟度，即用户行为的

稳定值，稳定值=用户 4 周行为数的平均波动程度。稳定值最高 3 分，最低 0 分，波动越小，分数越高，代表用户越成熟。 最终，得出的用户成熟度模型的分类如下图：

用户成熟度模型	
用户类型	经验标准
全面稳定型	原创、评论、转发、登录四项主动行为至少三项稳定值 ≥ 1.5
原创稳定型	原创和登录行为两项稳定值 ≥ 1.5 ，转发、评论行为稳定值 < 1.5
转发稳定型	转发和登录行为两项稳定值 ≥ 1.5 ，转发、评论行为稳定值 < 1.5
浏览稳定型	登录行为一项稳定值 ≥ 1.5 ，原创、转发、评论行为稳定值 < 1.5
不稳定型	原创、评论、转发、登录四项主动行为四项稳定值 < 1.5

有了用户的单一行为特征后，我们可以和活跃度模型一样，为每一个用户打上成熟度的标签，用户一上来我们就知道这个用户是喜欢只浏览不发言，还是喜欢发发发，或者喜欢转转转？那你还不知道要给他推荐什么内容吗？ 此外，成熟度模型不仅根据用户行为特征进行了分类，而且结合用户活跃度模型的变化，隐形反映了不同类行为特征用户的活跃度差异，在活跃度的表现上：全面稳定型>原创稳定型>转发稳定型>浏览稳定型>不稳定型。

总结： 用户行为模型是提高用户活跃度基础上非常重要的用户模型，只有知道用户的行为特征和喜好，才能知道什么样的引导不会让用户反感，什么样的引导能够促进用户活跃。 本质上来说，用户行为模型也是在帮助产品还原刻画出来在自己的产品生态中，优质用户的画像。一个优质用户，首先一定是最活跃的，其次更重要的就是他的活跃中具备了这样、那样、那那样的行为特征。 这也为我们后续搭建用户成长体系打下地基。

六、用户流失模型

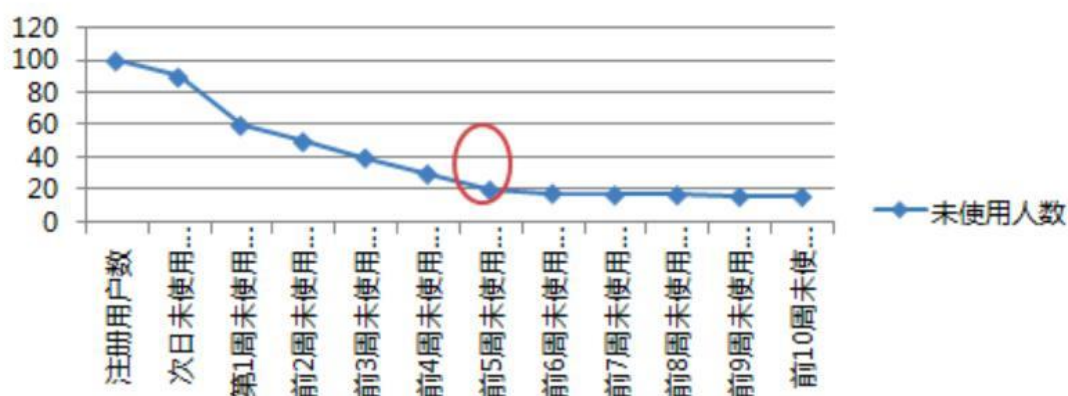
既然讲了用户活跃度和用户行为模型，那么第三个不妨来讲讲用户流失模型，这其实也是用户活跃度模型的衍生，流失用户即活跃度为 0 的用户，已经流失不再使用产品的用户。但是流失用户完全放在活跃度模型中来做，而且非常的重要。

因为用户流失模型的目标就是把流失用户根据一定的行为特征进行分类，在应用层面有两个方向： 1、 流失用户召回：应用于流失用户的召回，针对不同的用户特征启动不同的用户召回策略； 2、 用户防流失：应用于不活跃用户的防流失，我们已经知道了流失用户的特征，那么当不活跃用户出现了流失用户的特征的时候，说明他出现了流失预警，需要启动相应的防流失策略；

用户流失模型搭建的切入点是如何的呢？ 在搭建用户流失模型前，我们需要先明确流失用户的定义，也就是对于我们的产品来说，什么样的用户才被认定为流失用户。 流失用户定义 流失用户的定义可不是我们自己拍脑袋定义的，而是要看看用户的真实行为来判断，多长的时间窗内用户都没有使用，那么这个用户应该被定义为流失用户。

这里要强调一下：流失的概念，必须是长时间持续未使用，这需要和前面的用户活跃度概念相区别，我们在计算用户活跃度的时候，MAU 的统计方式就是这一个月里你只要使用过一次就算有过一次活跃，至于他在一个月中每天都来，而且一天使用好几次，还是就来一次，都属于 MAU 中的活跃用户。而流失用户的统计方式，则是在一个持续的时间窗内一直没有使用过，一次都没有，才被计算为是流失。 所以流失用户要如何定义呢，我们通过下面这张图来解释一下：

未使用人数



假设我们以周为颗粒度计算用户的流失周期，选取某一日（3月1日）注册的用户，假设为100人，这100个用户在注册后的第二天（3月2日）没有使用产品的有90人，也就意味着其中有10个人，在注册后的第二天就又来使用了产品。到了从第二天开始的一周后（3月2日-8日）仍然没有使用过产品的有60人，那也就是说，在刨去第二天的剩下的六天（3月3日-8日）内，曾经来使用过产品的用户有30人，那么加上注册第二天就来的用户，就是一共有40人在注册后的一周时间内，曾经使用过产品。这里需要强调的是，我们完全不关注这40个人在这一周时间里到底来了几次，只要他来过，就从“未使用人数”的数据表单中去除，所以可预见的，随着注册后周数的推移，“未使用人数”的数据一定是持续下降的趋势，因为只要在周数的推移中，有用户回来使用，就会被从这个表单中去掉，即便他次日来了之后就再也没有回来使用过。

根据这样的逻辑，我们模拟了下面这样一张图表。从图表分析我们可以得知：

- 1，当曲线趋于平稳时，说明我们的自然流失趋于平稳，即如果这个时候用户还留存着，基本上他就能留存很久，比较具有稳定性。换句话说，到了这个时间点，该主动回来的用户都回来了，没有回来的，就意味着已经流失了。因为后面随着周数的推移，回来的用户已经微乎其微。那么我们就可以以这个时间点

来定义我们的流失。如上图，我们就可以定义该产品用户如果连续五周没有使用，即可判断为流失；

2， 自然流失率：以上图为例，当天注册用户 100 个，其中有 20 个在此后五周完全没有使用行为，则该日的注册用户自然流失率为 $20/100=20\%$ ； 定义了流失用户后，我们回到用户流失模型的两个应用方向分别来构筑模型。

