推荐系统：持续的刺激用户内容的生产和消费

小流量实验技巧：

1. 小流量长，全量跌：
   1. 小流量更偏利用，所以当大流量帮你探索了哪些好哪些坏以后，你就有了这些信息，这些特征，然后小流量会利用这些特征和信息，直接利用相对较好的结果，所以小流量会长。
   2. 小流量做头部的东西，比如低俗的，擦边球的，猎奇的等各种高ctr的东西，这些肯定就比总体要好。
2. 波动很大：

2.1，实验的流量小，导致探索出的东西其实置信度很低，所以可能在今天表现很好，明天就查了。

2.1，供货有问题，供货质量每天的高低也会影响指标的涨跌

3，ctr长，时长长，dau跌

3.1，这其实是最值得警惕的，也是最容易麻痹人的。如果供货内容质量不高，都是时长相对较长的，就会出现这种情况。因为质量不高，可能是一些猎奇，高ctr 的东西，这些到时点击率高，时长边长。但是dau一定会跌。

所以，其实ctr一方面是推荐系统带来的，另一方面，内容本身是带ctr的，一些猎奇，低俗的东西本来ctr就是高。但是这些东西可能不是用户的兴趣，或者说不是一些人的兴趣，总是推这个东西，总体表现肯定很好，但是一定会流失用户。因为这些其实并不是个性化推荐带来的红利，报纸上都有。这些不是用户的差别化兴趣，用户的diversity并没有体现。

4，小流量不长，全量大涨：

4.1，小流量的探索流量是不够的，即使小流量的策略或者算法比全量或者基线好，但是由于流量不足以探索，所以导致全量后会大涨

5，全量长，后面更长：

内容质量

反向实验桶，back bucket

关于ctr涨的一些理解：

1，推荐是用户想消耗时间还是想让用户快速找到一些东西。推荐并不是一个用户的指标，而是一个系统，构建一个网络的指标

2，

头条：图文的ctr稳定在15%-20%左右。

视频：稳定在25%

3，Ctr涨，长在什么上很重要：

1. 内容侧：ctr长在长尾的高质量内容上。那说明算法在内容的理解和挖掘，还有用户兴趣的探索是有效的。如果长在头部，那说明对用户兴趣的探索是有效的。如果长在头部，低俗，猎奇的东西，那说明算法其实是无效的。因为你既没有探索出用户的兴趣，又没有挖掘出高质量内容。
2. 用户侧：长在新用户上，长在老用户上。长在新用户上，说明对新用户的最大的一个兴趣点的探索是有效的，或者说是兴趣探索的时间更快。如果长在老用户上，说明兴趣的范围是更广的。

最后点击到底能反映多少信息，其实很少。因为点击只有正反馈和负反馈，但是正多少，负多少很难衡量。比如一些极度引起用户反感的东西和用户看看标题就明白没有点进去的东西其实差别相当大。

2,

增长怎么做：

一个人群一个人群的打，先很好的满足一个人群的需求，然后在满足另一个人群的需求。

1，从用户侧：

新用户如果更加快速的长到第一个兴趣点，做的是兴趣探索。老用户怎样探索更多的兴趣点，做的是兴趣转移。新用户其实关注的相关性，老用户关注的其实是惊喜性。

怎样探索用户的兴趣：这里就有一个很重要的机制：流转机制：老用户给新用户输血。用老用户验证过的内容给新用户。老内容给新内容输血，用验证过的老内容去验证新内容的标签。

这里其实有一个非常重要的技术点：知识图谱。其实不用知识图谱也能探索兴趣，但是用知识图最大的优点就是把时间复杂度从O（n）降到了log（n）

2，从供货侧：

时效性内容的供给，爬取。

用户评论，点赞，分享，评论等交互式产品的设计以及用户体验的提升。创建一个和谐，积极的用户内容生产的一个环境。因为这些这些才是用户真正喜欢一个东西的表现。

用户短视频，自媒体等一些激烈机制，包括曝光分发的一些公平性，这点快手做的是非常好的。这个最重要的还是一个好的激励方法以及运营模型，但是算法其实做的就是如何让优质的内容曝光在需要他的人的身上。

所以公平性和多样性一定很重要。不能扼杀小的创作者，同时也要让已经优质的创作者的收入得到保证。

用户满意度靠什么：

1. 内容质量
2. 相关性
3. 时效性

探索怎么做：

探索利用的实验其实很难做，需要很长的时间才能做好。探索的实验一定要明确实验目标，比如用户对于内容的兴趣，这个可以通过用户点没点，赞美赞，完成度等各种指标进行建模。但是内容的diversity探索，这个其实不是内容，而是每一刷的指标情况。他探索的是一个版面的pvctr或者其他指标。这个东西就和排版有很大的关系了。

所以相关性其实不是简单拿去做召回，pctr模型特征，而是影响整个排序的参数，例如diverstiy，准入，rank function，混排等

探索的一些技巧：

几个概念：

1，用户和Item都是有置信度的，什么时候用用户的置信度，什么时候关注Item的置信度

1，用一个好的算法在好的用户上做基线探索。假如我发明了一种算法叫A，IDST的算法叫B，IDST的算法比我的算法好。，那我可以让优质的用户在优质的算法上做探索，因为像阿里的IDST算法，是基于Tompson抽样的，那么这部分优质的用户运行在优质的算法上做探索肯定指标是不错的。用置信的用户做实验，因为老用户兴趣相对固定。 然后我让新用户或者不置信的用户在我的A算法上做分发，对A分发的东西很简单我根据点击来分发，那么指标肯定也会涨。

用户画像怎么做：

用户画像就是用户的需求，用户兴趣的分类其实就是内容的分类，但是除了这个还有用户兴趣的转移程度，用户接受新鲜内容的程度，用户空闲时间等多种维度的刻画。

标签怎么做：

运营是配置一些标签，然后算法利用这些标签绘制用户画像，不同的用户画像在不同的APP是不适用的，适用于这个app的用户画像是产品，运营和算法共同努力的结果。标签不仅仅是内容的分类，也有内容的时效性，质量等多个维度的评估。

标签和内容的关系：

内容分类标签和用户兴趣兴趣标签其实是一套东西，做的时候也是利用流转机制去不断的推断内容和用户兴趣的标签，但是单靠流转机制其实效率是很低的，而且推得不好也会影响用户体验，所以一个合理的looklike模型是很重要的。内容测就要涉及到一些自然语言处理的东西，用户侧可能更多的是数据挖掘，利用用户和内容的特征把looklike做好。

Diversity怎么做：

假如说男女对素材有很大的区分，那么我创造了一个新的素材分发了以后发现，男女并没有区分，那可能说明素材是有问题的，因为素材本来就没有区分度。用一个素材的基尼系数或者熵去衡量一个素菜的在某些人群和标签的聚集程度，然后同一个页面排列多个熵小的素材，那么这个页面的diversity自然就大。

关于产品和推荐的理解

推荐系统是建立在用户对产品认知的情况下做的内容分发。比如，用户来优酷，就是为了看片的，看视频的，你给用户总是推治疗癌症，痔疮这些一定会引起永不反感，因为他不是来治病的，这不符合用户来平台当时的心智和需求。在进一步思考，优酷在中国人心中是什么样的产品定位，产品形态其实比较定型的，我就是到上面看电视剧和电影，综艺，再不济我也看个和他们相关的短视频，但是用户不会来上面看直播，看西瓜，火山这样的短视频和小视频的，所以优酷的电流就没做起来。其实一个产品在用户心中的认知就决定了这个产品的定位和天花板，想改变这个真的太难了。比如医药养生类APP，用户去看就是为了看健康的，你推明星的花边新闻不是很合适。