

互联网 | PM能看懂的个性化推荐（一）：五大适用场景

原创 卿一 掌仙人儿 2019-03-08

仙人儿不是算法产品经理，
最近要在自己的业务中应用算法推荐，
把一些学习所得做一点分享~~

VR、区块链非常流行的时候，很多人都希望能乘技术和资本的风，破产品的浪。而事实上很多时候，技术是好技术，用在你的产品上也是可以，但没必要。

个性化算法推荐这个概念已经不陌生，技术也相对成熟，而且不像其他昙花一现的“技术革命”，它已经为很多公司带来了长久而切实的利益，成为了这些产品的制胜核心。例如连续六个季度保持 > 30% 营收增长的netflix，把音像制品租赁这种“夕阳产业”转型为流媒体，还利用算法推荐做得风生水起，甚至扭转了劣币驱逐良币的行业诅咒。再例如我们身边，在新闻领域异军突起的宇宙条，拿得起放不下的抖音，都因为算法推荐带来了不断增长的日活和留存。

算法推荐毕竟挺贵的（人力&算力），到底在什么产品中使用算是用在了刀刃上？

个性化算法推荐为什么越来越普及

最近20年，随着互联网的蓬勃发展，我们经历了信息匮乏到信息过载的过渡。在从前，我们为用户提供商品、内容、服务，基本上全靠用户主动的搜索和检索，用户有明确的需求或者需求范围才会来到我们的产品。

现在我们有海量的条目，如果还等着用户主动获取，就会效率非常低下。

一方面会有大量的长尾无法获取曝光机会。用户如果不知道它的存在是不会主动去寻找它的，即使用户很需要。

另一方面产品过于被动。用户的大量时间在等着消耗，我们如果能让用户在没有明确需求的时候也想要来我们产品，就要比用户更了解他自己，他不知道自己想要什么的时候，我们要知道他想要什么，主动喂给他（feed名词来源？）。这样用户每次使用我们的产品，都有的吃，自然会经常光顾。

可以举例一个大家熟悉的场景：

如果我身边只有小卖部，我想买辣条的时候会去，想买三文鱼的时候不会去，因为我知道它没有。（搜索场景）

后来附近开了大超市，我家冰箱空了的时候会去超市食品区看看有什么可以买，但是没有明确需求的时候不会去，因为不知道要买什么。（分类检索词场景）

再后来超市给我配了私人导购（个性化推荐），她会告诉我：上个月你买了10连包的卫生纸，估计快用完了，这是你常用的品牌，补个货吧（基于用户行为）；还有你爱吃的草莓新上市了一个品种很受欢迎，可以尝尝哦（基于用户标签）；对了，经常和你一起来的那个朋友前几天买了一瓶混合花生酱，你要不要也试试（基于社交关系或协同推荐）？那我可能有空的时候都会去逛逛，很可能有收获。

从上面可以看出，个性化推荐对于内容的分发，特别适合内容信息过载的产品。具体有以下几种：

视频

视频内容大致分两类：PGC和UGC。

PGC主要指专业产生的内容，比如需要购买版权或自制的电影综艺电视剧，和影视公司合作宣发的网大，和专业生产团队合作的影视解说等。PGC内容由于制作门槛较高，所以内容数量有限（与用户量对比）；内容质量好，所以有沉淀价值，长尾内容也同样有受众，值得挖掘，尤其是电影品类。

比如优质的文艺电影或者是七十年代的经典电影，如果给它们分配高流量的展示渠道属于浪费资源，因为数据肯定不如新热内容好；但是如果流量都分配给新热内容，又会造成马太效应，浪费了片库，用户也不容易找到自己喜欢的内容。

这里举个大佬的例子：netflix。早在音像制品租赁的时代，netflix就已经意识到个性化推荐的重要性。2009年它发起的Netflix Prize百万美金竞赛，首次把这个概念从学术界引入商业界。直到现在，netflix在算法推荐方面也是视频内容分发的行业标杆，国内的视频产品在算法方面都难以望其项背。

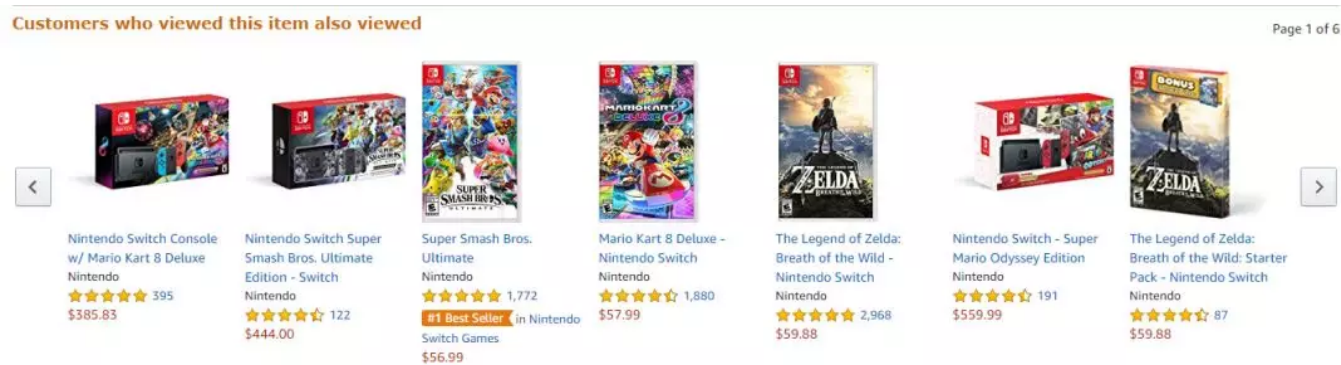


UGC主要指用户产生的内容，形式五花八门，质量良莠不齐。UGC视频制作门槛很低，所以数量巨大，而且时效性强。这类内容的分发重点不再是长尾内容的高价值利用和挖掘，更多的是更快扩散热点，以及通过feed流等低操作成本、低时间成本的产品方式，吸引用户并让用户产生沉浸感。

电子商务

用户在购物网站上是否能迅速找到所需商品，直接影响到订单成交的漏斗转化。

在全球电商巨头亚马逊，有20%~30%的销售来源于算法推荐。在用户购物的每个环节，算法推荐都有渗透。比如在首页，根据用户的历史消费行为给予推荐；在商品页面，推荐与该商品有关的物品，例如“浏览过该商品的用户最后还买了……”等。



淘宝最近几年的个性化推荐也逐渐提高了应用比例。在去年（好像是，不太记得了），淘宝首页简化掉了之前的分类版块，从第2屏开始就进入了弱化分类的个性化推荐feed流。用户点击item右下角的更多按钮或长按图片，可以呼出负反馈蒙层，校准个人特征。

广告

比起简单粗暴的无差别广告投放，利用个性化可以提高广告的CPM、CPC等指标。

例如谷歌的Adsense，通过分析用户当前正在浏览的页面内容，投放和这个内容相关的广告。谷歌Adsense可以说是全球最出名的广告联盟，开发者可以在自己的网站嵌入一段google提供的javascript脚本，网站通过审核后，用户来这个网站时就会看到和页面内容相关的广告，一方面更容易命中用户，另一方面丰富网站内容。如果用户点击了广告，网站开发者就可以获得一点收入。



日常中经常有朋友惊呼，XXX卖我的信息！我在XXX搜了罗技鼠标，去YYY的时候立刻给我推了罗技鼠标的广告！这种情况大概率就是XXX和YYY加入了同一家广告联盟，而不是他俩之间直接交易用户信息。

讲到这里，就涉及到了个性化算法对于用户信息的获取和利用，该坚持什么道德底线。这是另一个话题以后再讨论。

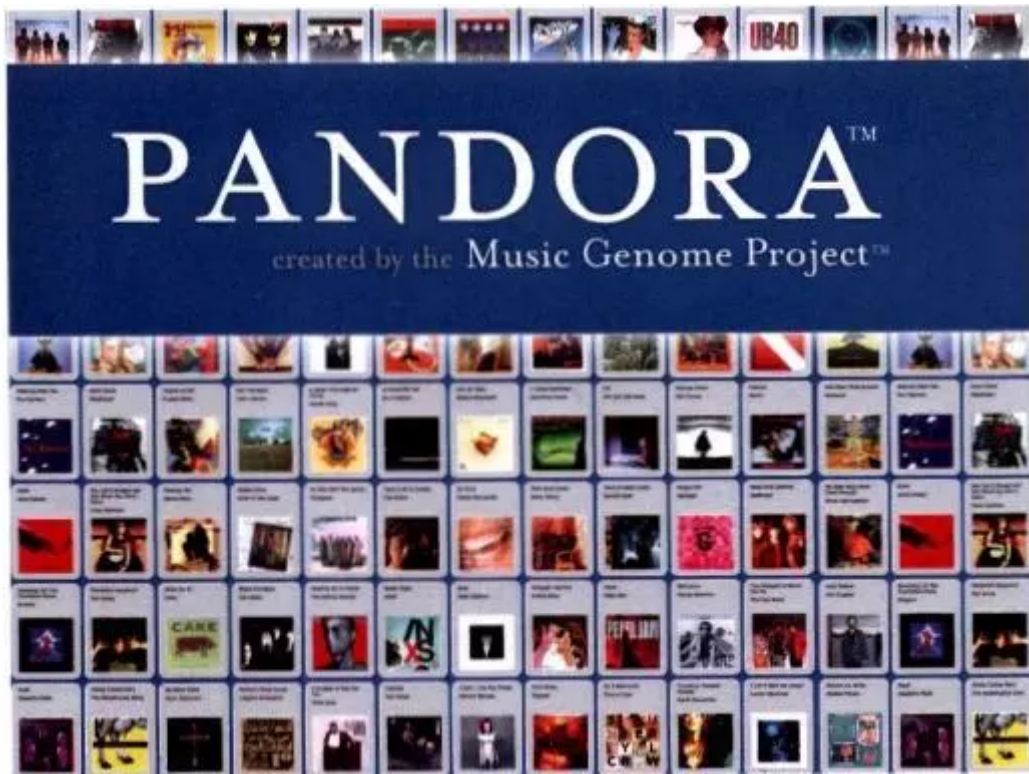
音乐&阅读.....

其实音乐、阅读等领域的个性化推荐，与视频很相似，都属于可消费内容的分发。

这块单说一下音乐吧。

对内容的算法推荐，召回通道中，标签是最不能忽视的一类。尤其像音乐一般单条只有几分钟，而且可以作为背景音来播放，消费成本很低。所以用户是否对某条音乐进行播放、播放的时长，正反馈的权重不像视频的播放那样高。这时候音乐本身的特征就变得非常重要。

最近被收购的美国音乐流媒体服务公司Pandora，从创立到现在十多年，做得最多的事就是推荐！推荐！推荐！早在2000年初，潘多拉就开启了音乐基因组计划，让专业的音乐家和对音乐有追求的工程师对音乐进行分类、打tag、提炼各维度特征。最终通过越来越了解用户，获取了美国境内超过50%的网络电台收听量。虽然最终因为巨亏被收购，那都是盈利模式的问题了，不可否认在推荐这件事上，潘多拉是认真的。



社交媒体

社交媒体的两大个性化场景：用户推荐和信息流展示。基于用户在社交网络的社交关系和大量行为，这些都比较容易做到。

用户推荐也就是好友推荐，或者关注人推荐，利用最多的是三元闭包原理，简单来说朋友的朋友就是朋友，一般二度好友作为推荐候选集准确率和基数量都有优势。其次参考社交兴趣，一般基于社交和兴趣的好友推荐的准确率会比较高。

信息流这个就比较有趣了。社交平台的信息流和普通的内容发布平台不尽相同，对于一般内容平台，信息过载的情况下，只有个性化分发才能信息效率最大化。但是社交平台的信息，大部分都是已关注的用户信息，还要不要做个性化呢？

如果不做个性化，用户可能很快就刷到底，然后因为没有新内容而离开，低活跃用户也因为偶尔来一趟页面，不容易找到重要的信息转而流失。如果做了个性化，可能会惹毛用户，因为面对支离破碎的时间轴，用户完全不清楚黑盒中的推荐逻辑是什么，会没有安全感。

Twitter 从16年2月更改了首页排序方式，企图加强对信息流的控制力，然鹅这次玩火最终在18年底认怂，把排序方式的选择入口拿到了首页，允许用户改回干净的时间线排

序。



总结

从“我就坐等用户找到他想要的”，到“我不管用户想要什么反正我就推荐我的”，到“我大概了解大部分用户想要什么，我尽量精细运营投其所好”，到现在的“我不用知道你想要什么，但我靠机器深度学习了你的全貌和内心”，可以说达到了无我相，无众生相的推荐境界。

后续会继续总结最近的学习和实践成果，包括个性化算法推荐的几大类召回方式、评测方法、道德讨论。