推荐系统评价

什么是好的推荐系统

佛陀没有定义过，人类也没有定义过。但一个系统的好坏往往需要「全链路的评定」，「贯穿于用户的整个交互过程」。之所以说「好的推荐系统」更难定义，是因为虽然算法是核心，但是个性化推荐往往不止由算法构成，这背后需要各种技术支撑。它是算法和各种技术架构以及交互设计等等「混合」在一起的产物。所以，我们「很难有一个推荐系统好不好的绝对值，只能在有参照物的情况下，选取某些常见指标然后有一个相对的评价」，「没有统一标准」，但人人心中都会有一杆秤。

这里需要提一下「客户体验价值」，它帮助企业了解每个客户的体验价值和商业价值，从而帮助决定如何对待客户体验，以及「如何调整客户体验项目的投入」。每个产品特性导致不同的评估指标，最核心的是「如何评估用户需求满足度」？不同场景、不同阶段的推荐系统各部分重要程度不同 ~~ 推荐效果影响因素：「用户需求、数据质量、算法策略、模块位置、展现样式(图文版等、契合产品)」…

「推荐目标」:实现产品目标(合适的才是最好的)

「核心问题」:如何构建一个用户对商品的评价模型

「宗旨」:服务提供方与消费方的双赢

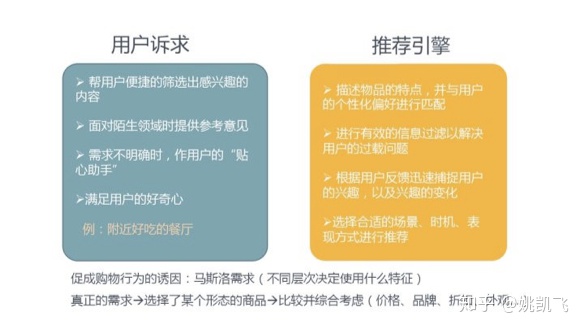
推荐系统是一个多方交互的「复杂系统」，有很多因素能够对推荐系统的效果产生影响，本文试对这些因素中的一些代表性因素进行简单的阐述分析。

▌1.1宏观

从宏观上来看，好的推荐系统就是满足用户的需求，帮助其进行「路径优化与兴趣发现」。微观上来讲，好的推荐系统一定是再为产品服务，不同的产品阶段，我们需要确立不同的推荐系统目标及评价体系。比如产品「上线初期」，该产品线上的用户为初次体验用户，我们可能更多的让用户体验到性价比(以电商推荐产品为例)，比如新客专享等福利。再比如产品「逐步发展」，我们可能会去着重考量浏览深度，点击率，转化率，客单价，gmv等指标，或是多个指标的整体值。

推荐系统评价的体系会围绕以下几点进行决策：新上线功能/算法「对产品价值多大，成本多高，收益如何」？展开的话就是可以问自己以下几个问题:目前的「推荐系统与产品和核心是否一致」，是否是在「强化现阶段的目标」(浏览深度，点击率，gmv等) ，与目前产品的「发展阶段是否相符」，这里主要考量产品自身的「用户量」，「人群大小」，「用户使用频度」等，最终就是结合以上两个问题，所使用的推荐算法策略，「是否能够强化产品核心心智」。

▏1.1.1产品域角度



1) 依赖于产品领域知识及目的

推荐产品的评价强依赖于「产品领域知识」，不同的领域对于评价也会有比较大的区别，工具类的产品重「黏性和时长」，电商重「回访和转化」。

2) 无整体评估方案存在

推荐产品的评估方案也存在「trade off」，很难做到各个方面都兼顾，只能说根据产品的各个阶段选择适合的评估方案，「不断调整及迭代」。「产品初期」可能会考虑用户的交互及浏览时长等体验指标,「产品中后期」会从商业价值实现角度考虑商业收益等指标

▏1.1.2信息检索角度

1) 「路径优化」

推荐作为信息分发的路径，定位于快速帮助用户找到合适的条目，简化和缩短用户行为路径，用完即走，当然不是永远不来。每次快速完成用户需求，并且用户持续地来。

2) 「准确的推荐」

信息分发维度，推荐就是需要准确，必然是准确的推荐是好推荐的核心要素之一。

▏1.1.3推荐系统角度

1) 「长尾挖掘」

长尾挖掘必然是推荐需要去完成的一件事，长尾作为大头的存在，分发过程中需要将把握，或者说长尾挖掘是好的推荐系统需要去完成的任务。

2) 「用户不知道什么商品存在」

好的推荐系统是既可以根据用户的反馈来推荐，也可以不断帮助用户进行探索，因为用户可能不具有某个领域内的知识，好的推荐系统还需承载帮助用户发现新事物的功能。

▏1.1.4交互角度

1) 「用户好的体验」

好的用户体验，外在表现为对产品的评价，回访、黏性、口碑等角度。

2)「教育引导」

好的推荐可以快速地教育用户，帮助用户快速地了解产品和使用产品。

3) 「说服并使用户信任」

可以让用户对产品的认知升级，并且开始信任推荐产品，对于推荐产品的结果产生依赖和信任感。

▏1.1.5商业视角

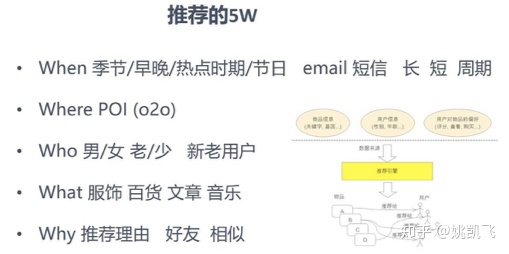
1) 可以为企业实体带来「商业价值的实现」。

2) 优化「销售边界和利润」

帮助产品「拓展边界」，通过场景构建「更多样」、「更贴心」的体验，最终提升「利润」等。

▏1.1.6推荐的5W

进一步展开，从细节来讲好的推荐系统往往会考量5个w—「when，where，who，what，why」



1) 「When」: 主要是在围绕时间维度做文章，主要是围绕「季节/早晚/热点时期/节日/周期」等因素为平台用户提供推荐服务，一般围绕时间维度的推荐会涉及用户的「购物习惯」，易耗品的「购物周期」等，举个例子来说，比如常见用户在奶粉/纸尿裤等商品上会呈现一定的购物周期规律，通过数据分析发现，用户的购物周期呈「正态分布」，如下图。再比如周末线上支付线下消费的产品销量会比工作日高等。当然泛推荐领域的信息推送等一般也会考虑时间因子进行信息「推送」(短信-email)。

2) 「Where」: 这个维度主要考虑「地理位置因素」，包涵了粗粒度的省份城市因素和细粒度的经纬度，粗细粒度一般在不同的产品场景下被使用，「细粒度」的地理位置信息一般在用在时效性/区域性要求较高的产品上，比如外卖(O2O)；「粗粒度」则更多的是的轻位置信息的电商产品重，比如衣服，百货等(B2C)，虽然轻位置信息，但是位置信息也会被使用在这类产品重，比如南方与北方姑娘喜欢的衣服风格会不一样，天气因素也会随地理位置的不同不同，所以同一时间南北方的消费需求会存在差异。

3) 「Who」:人物维度，则更多地去考虑「用户各类信息特征」，比如用户「是男是女」，「是老是少」，「是新用户还是老用户」，平时喜欢买什么品牌，喜欢什么品类的东西，风格如何等等，这些都是在描述这个人，现在这一类技术一般作为「用户画像」存在，通过用户行为和人口统计学数据构建的画像体系。

4) 「What」: 这个维度则是需要「推荐的主体」是什么，主体类型可能有以下服饰，百货，文章，音乐，美食，视频等等，「不同物料有着不同的自带属性」，并且「产品附加信息也不同」。比如文章，视频等对时效性的要求比电商类产品更高，特别是新闻类内容基本时效性要求在天级；再比如刚刚说的对位置信息的利用，美食相对于衣服这类商品对位置的要求就更高。所以在推荐的过程中我们会「根据推荐的主体不同做更多的推荐策略算法和系统的适配，最终去推动」。

5) 「Why」: 这个维度更多地会去考量推荐地「可解释性问题」，外化到产品维度则是推荐理由，比如是通过好友推荐的，比如通过浏览的商品推荐相似的商品等等 。所以上面这一切都基于数据洞察，维度基本有以下几个：「用户数，用户群体，推荐功能，推荐内容，非个性化/个性化，Top-N，列表浏览，是否实时反馈，消费需求变化情况」。

通过以上维度进行产品定位，最终可以考虑的设计要素有：「需求分析和用户调研，功能设计，界面设计，架构设计，算法设计，系统评测」。总之，「每个产品特性导致不同的评估指标，对于用户满足度不同的定义也带来了不同的评估方式，并且不同场景/不同阶段的推荐系统各个部分重要程度不同」。

用户「不只追求更优质的商品，也会期待看到商品为自己带来价值、便利性与独特性」。在积极打造更好的购物体验的同时，必须记住渠道「策略并不是一体适用」、能够「一招打天下」。电商平台上，每一名用户可以采取的购买路径就至少有10,000条。而根据内部的数据分析，一般来说，「85%的销量会来自35条高流量路径」。

1.1.7回到电商

回到电商，好的电商推挤产品，需要围绕商品更新，商品质量，商品与买家的匹配程度，好的推荐产品遵循以下四点

1).提升买家用户的体验，提高选购决策质量与效率实现优质买家的差异化服务

2).提高商品的有效曝光机会与转化率，提升卖家用户的效果与效益

3).利益均衡机制，均衡曝光机会，提升曝光商品及商家的覆盖率

4).提升买家留存率与卖家续签率，提升买卖家的忠诚度，提升商品点击机率

浏览与成交之间存在着巨大鸿沟，好的推荐将「帮助平台挖掘消费的深度」，在各个特定场景下「提升推荐转化」；「拓宽消费的广度」，「提升推荐对物料的覆盖」，「提升对用户的覆盖」，「各个场景的扩充」，「做到浅层消费到深层消费再到扩展消费」。

1.2微观

1.2.1 整体维度:

1) 「HEART模型」:Happiness 愉悦度、满意度、推荐度、感知易用性

2) 「Engagement 沉浸度」:使用频度访问频次、页面点击、停留时长, 点击/转化等比率，为产品带来点击率、转化、时长、黏性的提升

3) 「Adoption 采纳性」: 新用户率、客流量、退出率

4)「Retention ⽤用户粘度」: 老用户留存率、周回访率、周访问频率

5)「Task Success 任务完成度」

▏1.2.2 算法维度:

1) 「准确度评价指标」: 预测准确度，分类准确度，排序准确度，预测打分关联，半衰期效用指标, 准确的用户对商品的评分

2) 「准确度之外的评价指标」: 推荐列表的流行性和多样性， 召回率，覆盖率，新鲜性和意外性，用户满意度

3)「其它相关指标」:   用户对算法准确度的敏感度，算法对不同产品的普适性，广义的质量评价，个人隐私的保护，推荐系统的鲁棒性等 具体上述几个指标的计算等，我们在推荐系统的评价这一章展开。

1.2.3 线上&线下指标

1)「离线指标与评价」(mse/rmse/auc/ndcg/roc等)

2)「线上指标与评价」(abtest/ctr/cvr/gmv等),一般通过ABtest来完成

1.3开始前，先问自己几个问题

目前「即将开始的推荐与产品核心是否一致」（是否直接强化：PV/UV、GMV）；「是否与产品发展阶段相符」（用户量、感兴趣人群大小、频度）；「是否有合理的性价比/投入产出比」。「推荐功能对该产品有无价值(可以)，价值多大(值得)，成本和收益(现在)，优先级(怎样)」 。

1.4好的推荐系统迭代流程

首先需要「明确产品线当前需求」，「做好推荐系统定位」。就优化面来说：主要区分为是考虑做单产品「局部优化」还是直接考虑「全局优化」，平衡「整体收益与各模块内部收益」； 而从用户维度区分定位，则可以区分高收益用户与低收益用户。

有了定位就可以确定推荐目标了，从之前的介绍来看，我们可以选择「宏观与微观指标」，「长期与短期指标」来驱动迭代。下一步就是「方案选型」，通过系统开发周期的限定下，明确重点，选择合理的方案，并且预估一个合理的目标值。开发过程中可以逐步明确相关流程，并做调整。推荐系统的优化是一个逐步迭代的过程，这一过程中需要我们进行效果的监控，从而推动策略的迭代。当然可能在迭代过程中，我们推荐系统也会随着产品发展调整目标，也可能因为产品的相关原因，停滞或停止。

贯穿上述过程的整个流程，主要分为「需求分析和用户调研、功能设计、界面设计、架构设计、算法设计、系统评测」。

对推荐、搜索、广告和用户感兴趣的小伙伴可以阅读下面的文章汇总

[姚凯飞：推荐&amp;amp;搜索&amp;amp;广告&amp;amp;用户画像&amp;amp;深度学习整理​zhuanlan.zhihu.com](https://zhuanlan.zhihu.com/p/51015148" \t "https://zhuanlan.zhihu.com/p/_blank)