

【概念篇】你真的需要个性化推荐系统吗？

2018-03-05 刑无刀



【概念篇】你真的需要个性化推荐系统吗？

朗读人：黄洲君 08'42" | 3.99M

个性化推荐的历史和我们国家的改革开放历史差不多，已经有些年头了。它已经从一个非常小众的方向，演变成了今天互联网产品的常见“配件”。

再加上一些以“个性化”为品牌卖点的互联网产品的成功，也给相关从业者打了一剂“强心针”，并让更多人跃跃欲试，想给自己的产品加上个性化元素。

但是别急，在问“该怎么做”之前，先来回答一下“要不要做”，毕竟谁都不想去实现一个无脑PM拍脑袋给出的需求不是？我先帮你更加透彻地理解“推荐系统”这个概念，只有彻底理解了事物的本质，才能轻松做出“适不适合”的判断，毕竟知根知底再相爱更容易到白头。

什么是推荐系统？

到底什么是推荐系统？按照维基百科的定义：它是一种信息过滤系统，手段是预测用户（User）对物品（Item）的评分和偏好。这个定义不是很好理解，也不恰当。它用“怎么做”来定义了“是什么”，这相当于变相规定了推荐系统的实现路径。

让我们来换一个角度回答三个问题，从而重新定义什么是推荐系统：

1. 它能做什么；
2. 它需要什么；
3. 它怎么做。

对于第一个问题“它能做什么”，我的回答是：推荐系统可以把那些最终会在用户（User）和物品（Item）之间产生的连接提前找出来。

这里简单说一下“连接”这个词，这个词含义非常广泛，凡是能够产生关系的都是连接，比如用户对物品做出了一个行为，或者用户的某些属性和物品的属性一样等等，有关系就是连接。

为什么这么说呢？这是基于这么一个事实：万事万物有相互连接的大趋势，比如人和人倾向于有更多社会连接，于是有了各种社交产品；比如人和商品有越来越多的消费连接，于是有了各种电商产品；人和资讯有越来越多的阅读连接，于是有了信息流产品。

这还只是纯数字世界，随着各种物理实体智能化趋势越来越明显，万物互联还会进一步强化。世界是一个数字化的大网，但里面只有两类节点：人和其他。

人是互联的终极意义，“其他”统称为物品，物品可能是人、资讯、消费品、服务等。推荐系统就是要在这张巨大的网中，不断去发现那些很可能会和人发生连接的另一类物品节点，让它们和用户真的建立连接。

提炼一下上述逻辑：

1. 世界的发展趋势是万物倾向于建立越来越多的连接；
2. 人是这一切趋势的意义所在，为人建立连接是要义；
3. 根据已有的连接预测和人有关的连接，就是推荐系统。

为了更形象，我再举几个例子。

1 一个社交产品，比如脸书（Facebook），如果它的 20 亿活跃用户之间已经都有社交关系了，那么它的“感兴趣的人”这一推荐系统就该寿终正寝了。

从已经建立社交关系的用户身上去推测你还可能对哪些人感兴趣，本质上就是提前把那些可能的用户连接找出来，然后再按照用户分别呈现在每一个人面前。

2 一个信息流资讯阅读产品，比如今日头条，只有当用户不断点进源源不断的内容物品中，每一次点击，就是一个连接，每一次阅读也是一个连接，不同层次不同重要性的连接在推荐系统的帮助下不断建立，所主要依据的就是那些已经存在的连接，即：用户过去都点击阅读了哪些内容。

3 一个电商平台，用户刚买过什么，常买什么，你正在浏览什么，这些都是用户和物品之间已经存在的连接，用这些连接去预测还会买什么，还会看什么也是推荐系统。

按照上面的分析，我也同时回答了第二个问题“它需要什么”：推荐系统需要已经存在的连接，从已有的连接去预测未来的连接。

第三个问题：怎么做？

维基百科的定义提供了一个说法：预测用户评分和偏好。这是推荐系统背后相关算法和技术的两大分类，在后面的专栏内容中我会讲到；但比这个定义更抽象的实现方式分类是：机器推荐和人工推荐，也就是通常说的“个性化推荐”和“编辑推荐”。

两者之间还存在现在最常见的领域专家推荐，也就是网红推荐，如何为用户找到适合他的网红也属于推荐系统范畴，编辑推荐偏玄学了，我在这个专栏里不会重点讨论。

总结一下推荐系统就是：用已有的连接去预测未来用户和物品之间会出现的连接。

你需要推荐系统吗？

我已经根据“能做什么”“需要什么”“怎么做”三个方面，讨论了什么是推荐系统。那么只要前两个条件成熟，你就需要一个推荐系统，至于“怎么做”的问题则简单得多，否则的话就是暂时不需要。那么，如何判断条件是否成熟了呢？

我们可以考虑两点。

第一，看看产品的目的。如果一款产品的目的是建立越多连接越好，那么它最终需要一个推荐系统。有哪些产品的目的不是建立连接呢？一种典型的产品就是工具类，如果是单纯提高人类某些工作的效率而存在的产品，比如一个视频编辑器，则不需要。虽然如今很多产品都从工具切入最后做成社区了，至少在工具属性很强时不需要推荐系统。

第二，看看产品现有的连接。如果你的产品中物品很少，少到用人工就可以应付过来，那么用户产生的连接肯定不多，因为连接数量的瓶颈在于物品的数量，这时候不适合搭建推荐系统。

或者用户和物品数量在某些手段下也变得很多，但是用户和物品之间的连接很少，表现就是用户的留存回访很低，这时候也不是很需要一个推荐系统。

你应该是想办法找到用户流失的原因，直到他们能贡献第一批连接才行。当然，用户很少时，人工完全可以应付一对一服务时，也是不需要推荐系统的。

关于第二点，“长尾理论”可以帮助我们理解，如何把用户和物品各种可能的连接汇总，包括用户属性、物品属性等，应该要有长尾效应才可能让推荐系统发挥效果。

这里我介绍一个简单指标，用于判断是不是需要推荐系统：

$$\frac{\Delta connection}{\Delta user \times \Delta item}$$

分子是增加的连接数，分母是增加的活跃用户数和增加的有效物品数。

这个简单的指标我解释一下：

如果增加的连接数主要靠增加的活跃用户数和增加的物品数贡献，则该值会较小，不适合加入推荐系统；如果增加的连接数和新增活跃用户和物品关系不大，那说明连接数已经有自发生长的趋势了，适合加入推荐系统加速这一过程。

不过，具体并没有判断标准，因产品而异。

总结

到底要不要上推荐系统，如果仅仅从战术上来看，是一个关乎投入产出比的问题，搭建一个推荐系统的前期投入不小，你需要：组建团队、购置计算资源、积累数据、花费时间优化。

这些成本在早期不必须或者不成熟的情形下投入，显然投入产出比不是很优；但如果是战略问题，那就不在本文的讨论范围内了，至于哪些算是战略问题，我列举几个例子：

1. 产品要有这个属性，方便产品融资；
2. 团队要有相关人才，因为不好招聘，先提前屯着；
3. 要培养这样的思维，从而形成数据导向的产品文化。

类似等等，都不属于战术问题，具体问题具体分析。

现在我们来回顾一下，今天我先从三个角度定义了什么是推荐系统，然后从定性和定量两个角度分析了到底是否需要给自己的产品加上推荐系统。

那么，你可以试着从这两个角度分析一下，你自己的产品是不是需要推荐系统呢？欢迎留言，我们一起讨论。



版权归极客邦科技所有，未经许可不得转载

精选留言



墨明棋妙

16

后面课程，会有代码示例么？

2018-03-05

作者回复

尽量少。因为音频不好读。

2018-03-06



叶晓锋

9

老师你好，我有一个疑问。您讲到如果用户留存回访率低，不太需要推荐系统，而应该去找到用户流失的原因。问题是，很多消费并不是高频消费，产品属性决定了用户高流失率，例如airbnb的业务，他的业务会存在大量的潜在链接，又是低频消费，请问类似这种情况是否需要推荐系统？

2018-03-05

作者回复

帮助每个用户在低频场景做好每次决策，也是价值所在，不是都要留住用户贡献注意力的。我说的留存，有个前提，就是你以多长时间为统计周期，即使低频应用，比如几个月用一次，只要每次用都想起你，也是高留存。

2018-03-06



Classtag

8

其实从我个人的经验来看，很多产品其实不需要太多个性化的元素，需要的是的业务和用户的理解。亚马逊的推荐就是帮用户做决策，而不是帮亚马逊买更多的商品。

2018-03-05

作者回复

你得到了！

2018-03-06



Outsider

👍 6

要是作者本人朗读的就更好了，不一定要字正腔圆。但只有作者才能朗读出原文的轻重缓急。另外的朗读者来朗读始终觉得稍显平淡。

2018-03-09

作者回复

我有空试试。

2018-03-10



jt120

👍 6

希望有实战，有代码，最好有github上的配套

2018-03-06

作者回复

会有，专栏还是着重用浅显直觉方式呈现“有什么”，后续会有图书计划，书中会有更多代码和公式等“深内容”，专供自我要求较高的人享用。

2018-03-10



softpower2018

👍 4

个性化的推荐是不重要还是工程中难以实现，我感觉老师您不太强调个性化的精准推荐

2018-03-07

作者回复

精准是个坑，不要跳，不要跳，不要跳。我从不敢说精准，怕被老板听见，因为那是一个望山跑死马的终极目标，在实际中，不如多关注量化指标的增长。

2018-03-08



yxj

👍 3

老师您好，想问下是否可以单纯理解为只要所储备数据量到一定程度，然后推荐系统能帮用户提供更好的服务就可以指引团队去着手推荐系统了呢。因为数据量攒到一定量其实也说明项目已经到了中期运营过程，而且也保证了团队制作的推进系统有一定质量和可靠度。我的工作主要在于机器学习，所以在我看来搭建这样的推荐系统其实只要数据到一定程度根据现有的机器学习算法就可以在成本明晰且并不是很高的情况下实现了。

2018-03-05

作者回复

数据量还是个很低维度的指标，不能这样说，多大数据量是个大呢？

2018-03-06



AlexYoung

👍 3

非常浅显易懂且专业的表述。想请教邢老师一个问题，是否需要推荐系统的评判方法，涉及到用户与物料的规模和公司增长阶段。但对于某些特殊行业，例如房地产，物料可以理解为

在售的房源，这类物料的存量通常不会很大，增长速度也随市场行情和政策等因素不稳定波动。针对这类虽然2C但被行业因素限制的企业，是否还能适用文中公式来做评判，谢谢。

2018-03-05

作者回复

首先，这是个经验公式，衡量的是用户物品关系网络自增长的情况，你可以根据实际情况再定义其他公式。其次，和行业无关，假如房产数量仅仅十万量级，且静态不变，但用户贡献了很多连接，如收藏等，也是可以给他推荐相关房源的。

2018-03-06



小诸葛

👍 2

看了老师的分析，对于我目前负责的产品，从用户数和物品数这两个维度作简要分析：

- 1、目前的拉新更侧重是量&活动，这两者之间主要是激发活跃用户和增加物品数的贡献。
- 2、产品现有连接数较少，暂不能拥有自发生产的趋势。

因此还不太适宜做推荐系统。知其然而知其所以然很重要。

2018-03-07

作者回复

推荐系统锦上添花，先把产品本身打磨好。

2018-03-08



fifi兔

👍 2

判断是否需要推荐系统的那个指标，是否也可以用于后续评估推荐系统的推荐效果

2018-03-05

作者回复

这个没法说，这里只是衡量网络自增长情况的。万一有缘，和你的业务效果一致，谁说不可以呢？

2018-03-06



宁静致远

👍 1

刚接触推荐系统，想着以后从事这方面的工作，老师以后会讲如何去实现一个自己的推荐系统么？

2018-05-24



13683815260

👍 1

老师可以附加一些资料和链接什么的。毕竟跨学科的，里面也有很多公式，和得到的纯文字还是不一样的。

2018-05-14



王王王

👍 1

老师好，我是做电商内容推荐的，想要了解类似淘宝，京东这些大厂的内容推荐逻辑，可以通过哪些渠道呀？求回复

2018-04-05



sprinty

👍 1



老师您好，最近买了个打印机，已经入手。但是某些电商或网站还是老是给我推荐打印机，从我个人角度怎么阻止向我推荐打印机？推荐系统有什么好的方式规避这种问题，或者说如何断开用户和物品的链接？

2018-03-07

作者回复

一般是分品类去解决。多巴胺类是买了还想买，所以会推类似的，工具类物品则应强调推荐组合。阻止推荐打印机，抱歉这个我恐怕无法帮你了。

2018-03-08



rushui

1

说的真的非常浅显易懂，一看文风就是有很多工业界实战经验的，期待后面更精彩的分享！

2018-03-05

作者回复

欢迎多多分享传阅。

2018-03-06



逗逼师父

0

公司人力资源有限，我是这方面的小白，准备操刀推荐系统，跟老师目前算是同行业了，搞房产中介，希望老师以后能多多指教

2018-08-01



ayesd

0

可不可以认为，数据量少的，就通过咨询公司或者中介就解决了？

2018-06-12

作者回复

可以。

2018-07-23



会飞的书2008

0

写得通俗易懂，本人也在做推荐系统，处于刚入门，这篇文章把我自己一直想回答的问题表达得很清楚。

2018-05-12



daisy

0

判断公式不是很理解。可以举产品实例说明吗？

2018-05-01



三竹先生

0

社交产品的推荐系统我认为很难的

2018-03-22

作者回复

Facebook 做得很好。

2018-03-23

