

## 39 | 如何预测用户的喜好?

2019-01-26 李智慧

从0开始学大数据

[进入课程 >](#)



讲述：李智慧

时长 08:38 大小 7.91M



在用户对自己需求相对明确的时候，可以用搜索引擎通过关键字搜索很方便地找到自己需要的信息。但有些时候，搜索引擎并不能完全满足用户对信息发现的需求。一方面，用户有时候其实对自己的需求并不明确，期望系统能主动推荐一些自己感兴趣的内容或商品；另一方面，企业也希望能够通过更多渠道向用户推荐信息和商品，在改善用户体验的同时，提高成交转化率，获得更多营收。而这中间发现用户兴趣和喜好的就是推荐引擎。

在豆瓣中打开电影《肖申克的救赎》的页面，你会发现这个页面还会推荐一些其他电影。如果你喜欢《肖申克的救赎》，那么有很大概率你也会喜欢下面这些电影，这就是推荐引擎发挥的作用。

喜欢这部电影的人也喜欢 ·····



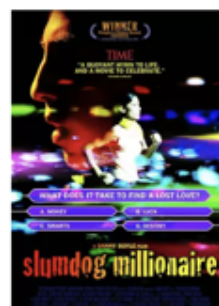
阿甘正传



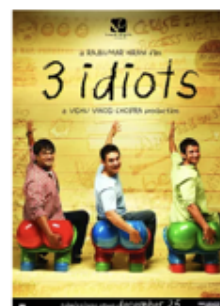
当幸福来敲门



飞越疯人院



贫民窟的百万富翁



三傻大闹宝莱坞



楚门的世界



荒岛余生



心灵捕手



国王的演讲



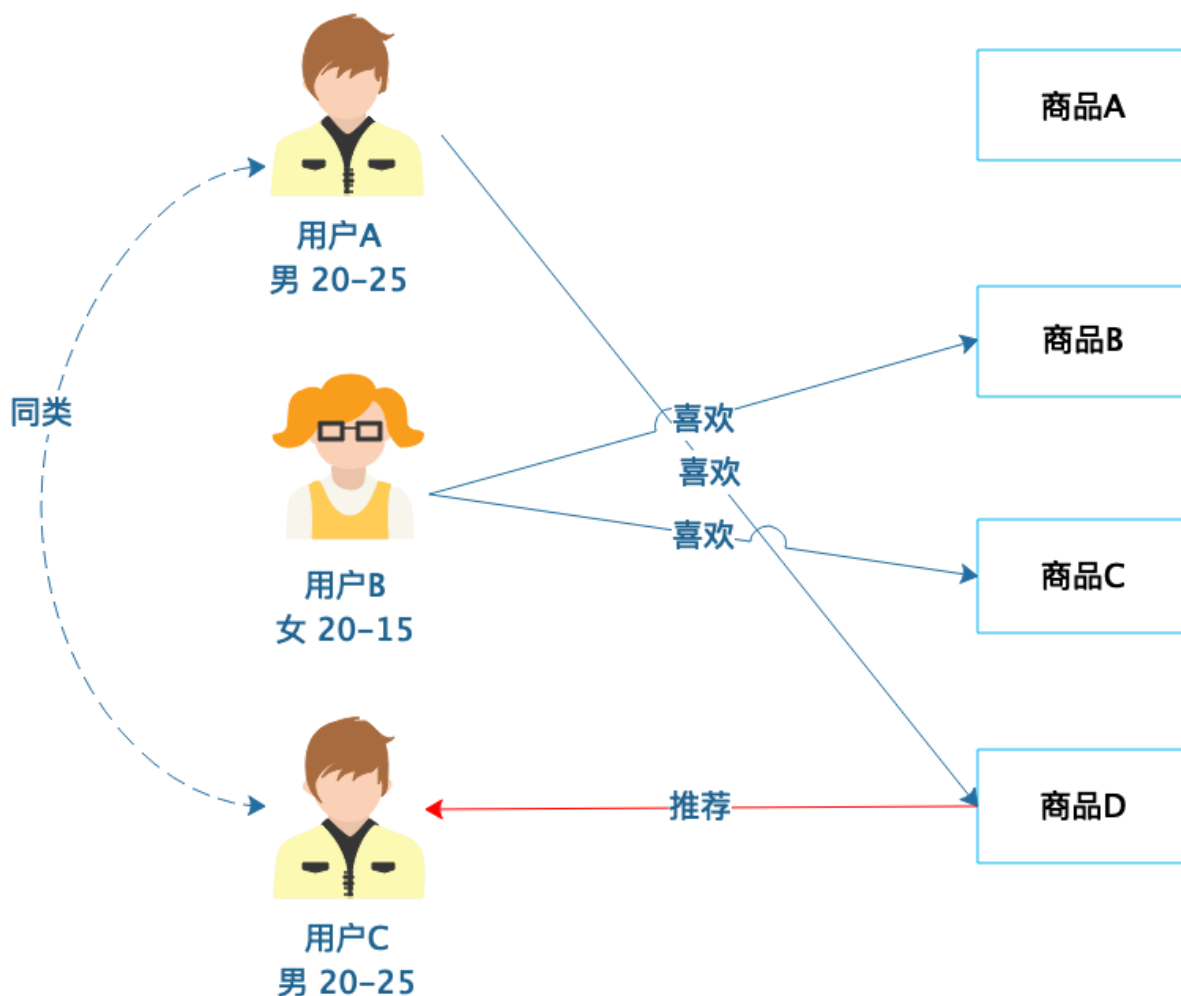
闻香识女人

推荐引擎的思想其实很早就存在了，后来随着大数据技术的发展，推荐引擎的普及程度和重要性也越来越高，淘宝曾经就主推“千人千面”，要让每个用户打开的淘宝都不一样，背后的核心技术就是推荐引擎。现在稍有规模的互联网应用几乎都有推荐功能，而一些新兴崛起的互联网产品，推荐功能甚至是其核心产品特点与竞争优势，比如今日头条，就是靠智能推荐颠覆了互联网新闻资讯领域。

那么推荐引擎如何预测用户的喜好，进行正确的推荐呢？主要就是依靠各种推荐算法，常用的推荐算法有：基于人口统计的推荐、基于商品属性的推荐、基于用户的协同过滤推荐、基于商品的协同过滤推荐。

## 基于人口统计的推荐

基于人口统计的推荐是相对比较简单的一种推荐算法，根据用户的基本信息进行分类，然后将商品推荐给同类用户。



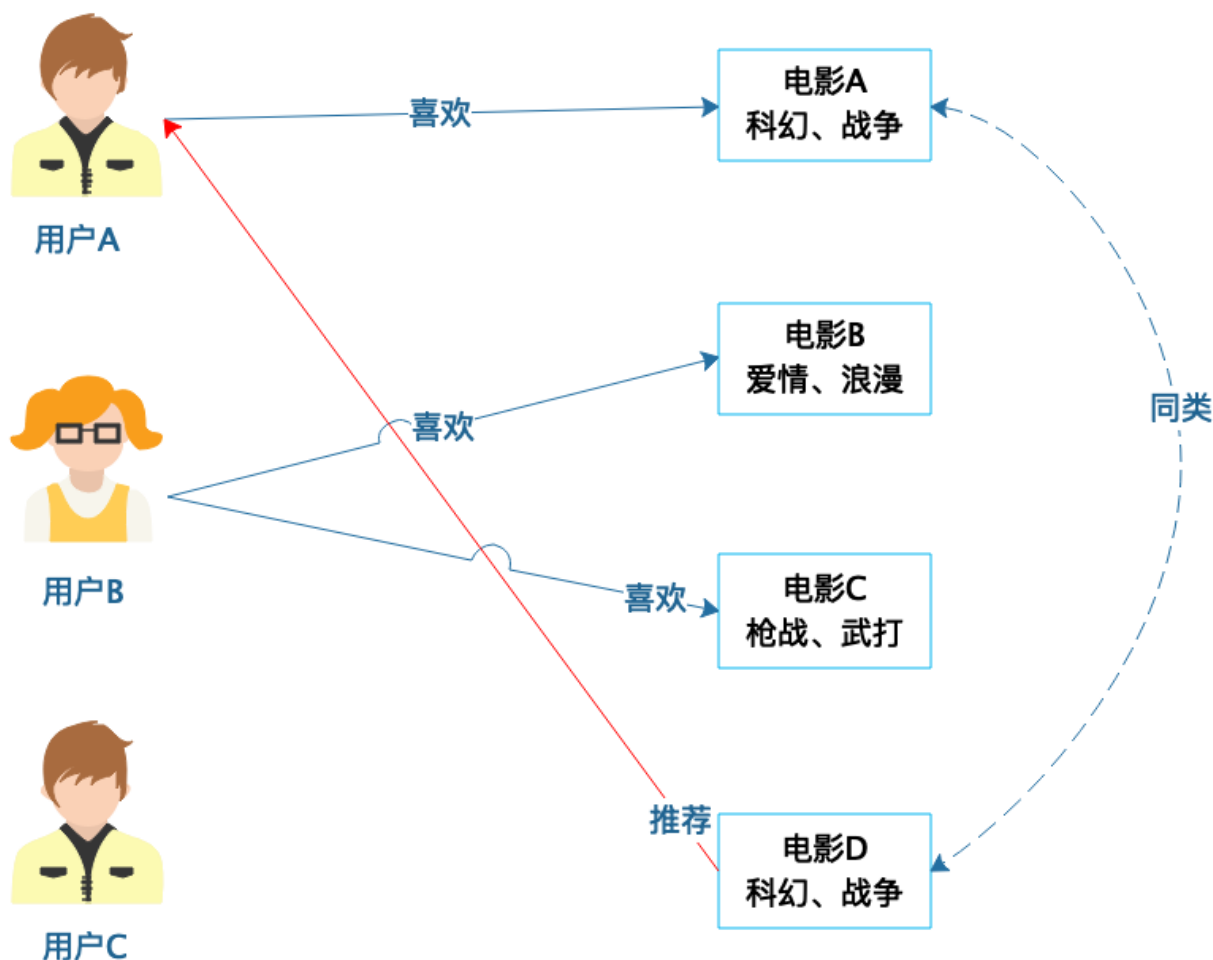
从图中可以看到，用户 A 和用户 C 有相近的人口统计信息，划分为同类，那么用户 A 喜欢（购买过）的商品 D 就可以推荐给用户 C。基于人口统计的推荐比较简单，只要有用户的基本信息就可以进行分类，新注册的用户总可以分类到某一类别，那么立即就可以对他进行推荐，没有所谓的“冷启动”问题，也就是不会因为不知道用户的历史行为数据而不知道该如何向用户推荐。

而且这种推荐算法也不依赖商品的数据，和要推荐的领域无关，不管是服装还是美食，不管是电影还是旅游目的地，都可以进行推荐，甚至可以混杂在一起进行推荐。

当然也正因为这种推荐算法比较简单，对于稍微精细一点的场景，推荐效果就比较差了。因此，在人口统计信息的基础上，根据用户浏览、购买信息和其他相关信息，进一步细化用户的分类信息，给用户贴上更多的标签，比如家庭成员、婚姻状况、居住地、学历、专业、工作等，即所谓的用户画像，根据用户画像进行更精细的推荐，并进一步把用户喜好当做标签完善用户画像，再利用更完善的用户画像进行推荐，如此不断迭代优化用户画像和推荐质量。

## 基于商品属性的推荐

前面一个算法是基于用户的属性进行分类，然后根据同类用户的行为进行推荐。而基于商品属性的推荐则是将商品的属性进行分类，然后根据用户的历史行为进行推荐。

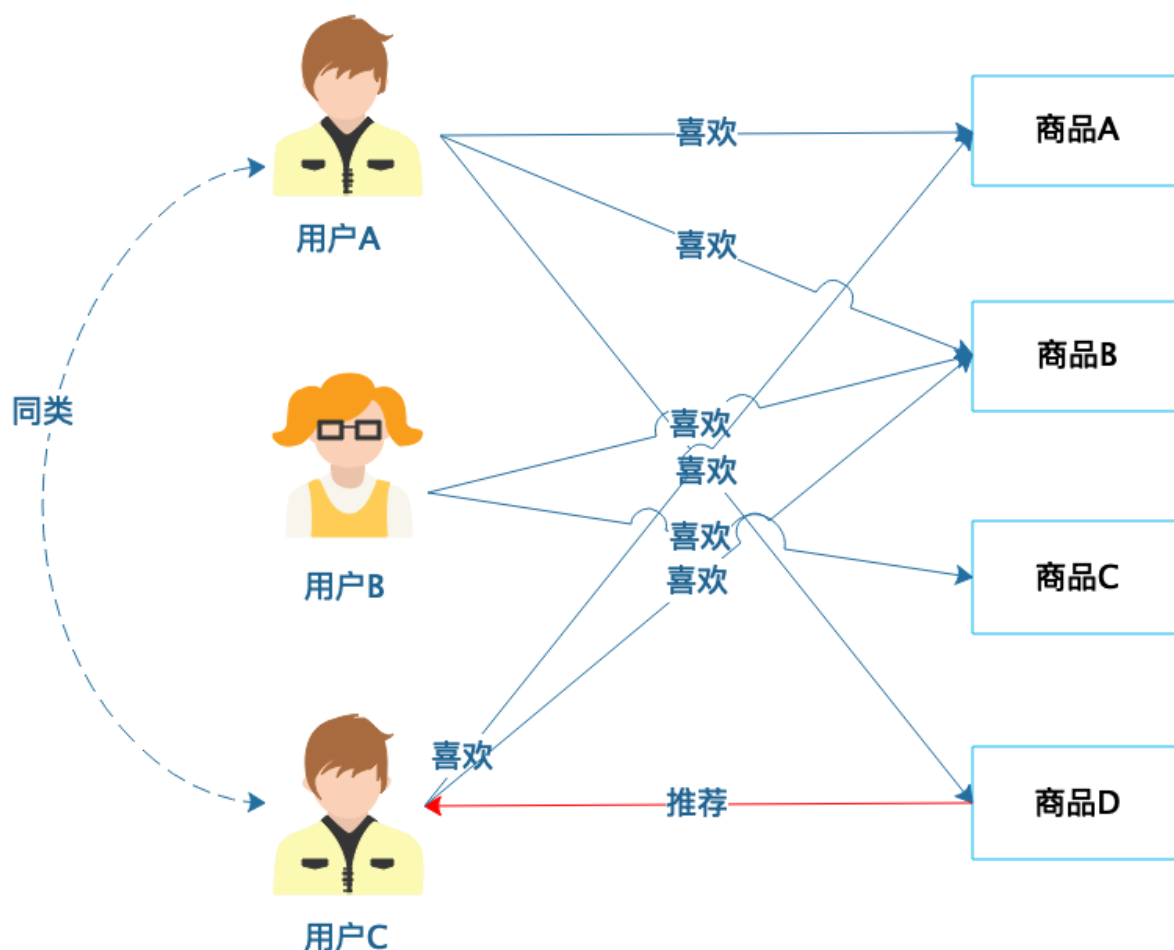


从图中可以看到，电影 A 和电影 D 有相似的属性，被划分为同类商品，如果用户 A 喜欢电影 A，那么就可以向用户 A 推荐电影 D，比如给喜欢《星球大战》的用户推荐《星际迷航》。一般来说，相对于基于人口统计的推荐，基于商品属性的推荐会更符合用户的口味，推荐效果相对更好一点。

但是基于商品属性的推荐需要对商品属性进行全面的分析和建模，难度相对也更大一点，在实践中，一种简单的做法是提取商品描述的关键词和商品的标签作为商品的属性。此外，基于商品属性的推荐依赖用户的历史行为数据，如果是新用户进来，没有历史数据，就没有办法进行推荐了，即存在“冷启动”问题。

## 基于用户的协同过滤推荐

基于用户的协同过滤推荐是根据用户的喜好进行用户分类，常用的就是我前面讲过的 KNN 算法，寻找和当前用户喜好最相近的 K 个用户，然后根据这些用户的喜好为当前用户进行推荐。



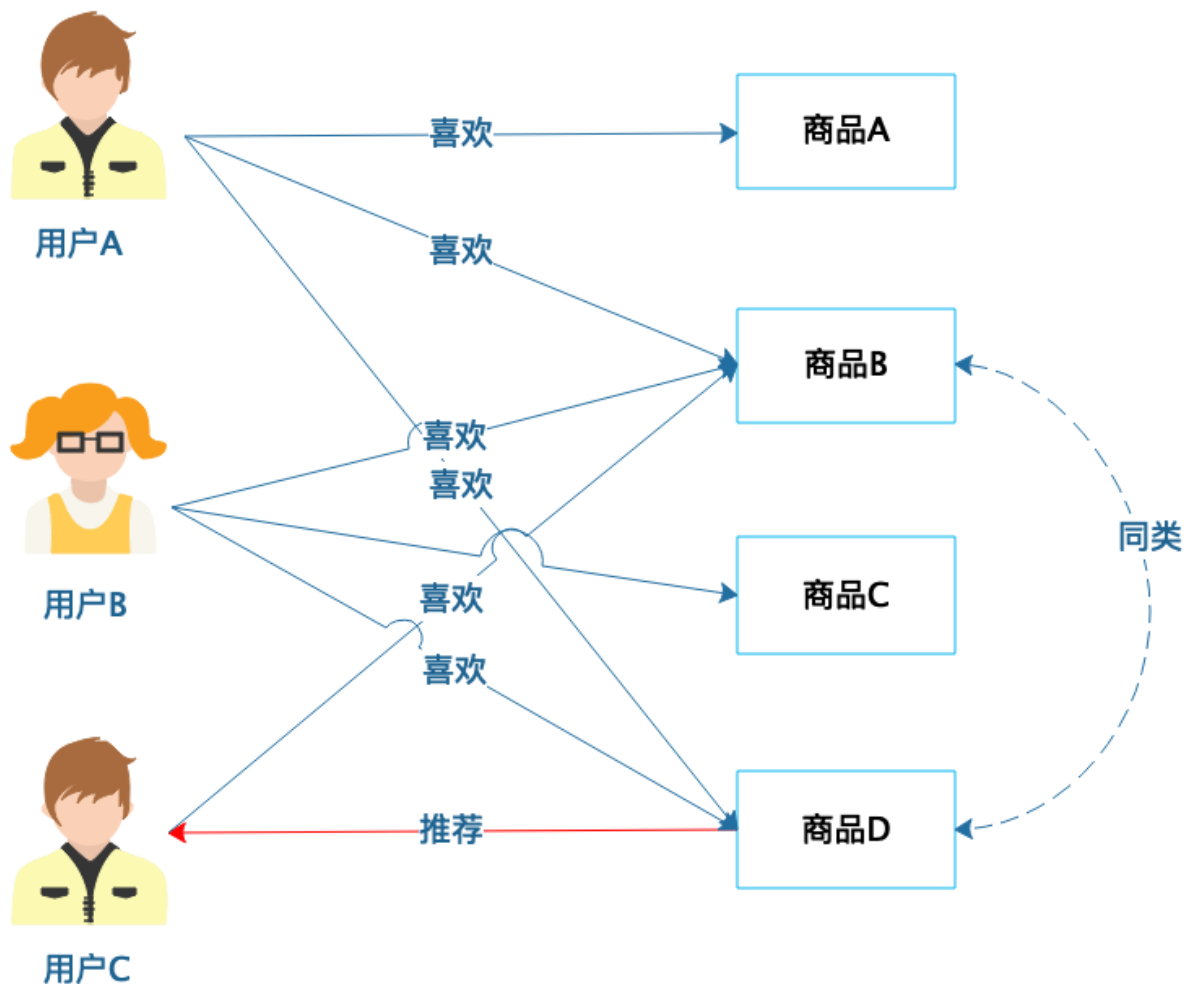
从图中可以看到，用户 A 喜欢商品 A、商品 B 和商品 D，用户 C 喜欢商品 A 和商品 B，那么用户 A 和用户 C 就有相似的喜好，可以归为一类，然后将用户 A 喜欢的商品 D 推荐给用户 C。

基于用户的协同过滤推荐和基于人口统计的推荐都是将用户分类后，根据同类用户的喜好为当前用户进行推荐。不同的是，基于人口统计的推荐仅仅根据用户的个人信息进行分类，分类的粒度比较大，准确性也较差；而基于用户的协同过滤推荐则根据用户历史喜好进行分类，能够更准确地反映用户的喜好类别，推荐效果也更好一点。今天文章开头举的推荐电影的例子，就是基于用户的协同过滤进行推荐。

## 基于商品的协同过滤推荐

基于商品的协同过滤推荐是根据用户的喜好对商品进行分类，如果两个商品，喜欢它们的用户具有较高的重叠性，就认为它们的距离相近，划分为同类商品，然后进行推荐。





从图中可以看到，用户 A 喜欢商品 A、商品 B 和商品 D，用户 B 喜欢商品 B、商品 C 和商品 D，那么商品 B 和商品 D 的距离最近，划分为同类商品；而用户 C 喜欢商品 B，那么就可以为其推荐商品 D。商品的分类相对用户的分类更为稳定，通常情况下，商品的数目也少于用户的数目，因此使用基于商品的协同过滤推荐，计算量和复杂度小于基于用户的协同过滤推荐。

## 小结

除了上面这些推荐算法，还有基于模型的推荐，根据用户和商品数据，训练数学模型，然后进行推荐。前面我们讨论过的关联分析，也可以进行推荐。在实践中，通常会混合应用多种算法进行推荐，特别是大型电商网站，推荐效果每进步一点，都可能会带来巨大的营收转化，如果你经常在网上购物，肯定也能感受电商网站这些年在推荐方面的巨大进步。

互联网发展到现在，历史上那种用户主动搜索，然后选择信息的产品模式几乎已经走到尽头。用户无需做任何操作，打开产品就能立即看到自己想看的消息，正成为新的产品模式，而最近几年快速崛起的互联网产品，也都有这样的特性。今日头条根据你点击的新闻，预测

你的关注偏好，不断推荐你感兴趣的内容，让你一打开 App，看到的几乎都是你感兴趣的内容。抖音、快手这些产品也是如此，通过不断计算、分析用户喜好，从而优化推荐的结果。某些社交产品甚至将人作为商品进行推荐，不断推荐你可能喜欢的人，遇到喜欢的人向左滑动，不喜欢的向右滑动，然后根据你的喜好进一步优化推荐。

在未来，随着用户对推荐功能接受程度的不断提高，推荐算法的不断改进，以及包括自然语言处理在内的各种 AI 技术的进步，系统主动推荐会逐渐成为主要的用户交互方式。也许在不久的将来，你不需要再打开各种 App 看新闻、点外卖、刷微博、逛淘宝，你的手机就已经非常了解你，它会主动推荐你该看点什么、吃点什么、买点什么、玩点什么。未来的生活可能不是你唤醒你的手机，而是你的手机唤醒你。

## 思考题

你在使用互联网产品过程中，发现使用推荐算法的功能有哪些？可能使用了哪些推荐算法？

欢迎你点击“请朋友读”，把今天的文章分享给好友。也欢迎你写下自己的思考或疑问，与我和其他同学一起讨论。

 极客时间

# 从 0 开始学大数据

## 智能时代你的大数据第一课

**李智慧**  
同程艺龙交通首席架构师  
前 Intel 大数据架构师



新版升级：点击「 请朋友读」，10位好友免费读，邀请订阅更有**现金**奖励。

© 版权归极客邦科技所有，未经许可不得传播售卖。页面已增加防盗追踪，如有侵权极客邦将依法追究其法律责任。

## 精选留言 (5)

写留言



强哥

2019-01-26

3

介绍了几种常用的推荐算法的解读，但是缺乏对应的场景应用。结合使用场景来介绍每种场景下应该用哪个推荐算法更合理会觉得更好点？



kobepeng

2019-01-26

1

看完这篇之后，让我感觉到大数据在生活中无处不在，除了那些高深莫测的数学模型，简单的分类分组其实也算是大数据的一部分，让自己在学习大数据的过程中有一种亲近感、带入感。感谢老师~

展开



小老鼠

2019-02-02

1

一般大型电商是不是结合这四类方案。

展开



shuifa

2019-01-27

1

mark

展开



杰之7

2019-01-26

1

通过这一节的阅读学习，毫无疑问，智能推荐已经成为今天产品功能的重点。它将用户和产品黏贴的足够紧，有一天很可能不是我们需要手机里的产品做什么，而是产品要求我们做什么，因为产品更了解我们自身。

在今天的学习中，老师介绍了常用的4种推荐，其中第三种和第四种推荐也是第一二种推...



展开 ✓