

零基礎科普：4種簡單推薦算法背後的原理

機器學習算法與Python實戰 今天

以下文章來源於大數據DT，作者李智慧



大數據DT

提供大數據、AI等領域乾貨學習資源的「寶藏號」，跟50萬技術人共同成長，一起玩轉大數據、Python、數據分析、數據科學、人工智能...



導讀：大數據平台只是提供了數據獲取、存儲、計算、應用的技術方案，真正挖掘這些數據之間的關係讓數據發揮價值的是各種機器學習算法。在這些算法中，最常見的當屬智能推薦算法了。下面通過幾種簡單的推薦算法來了解一下推薦算法背後的原理。

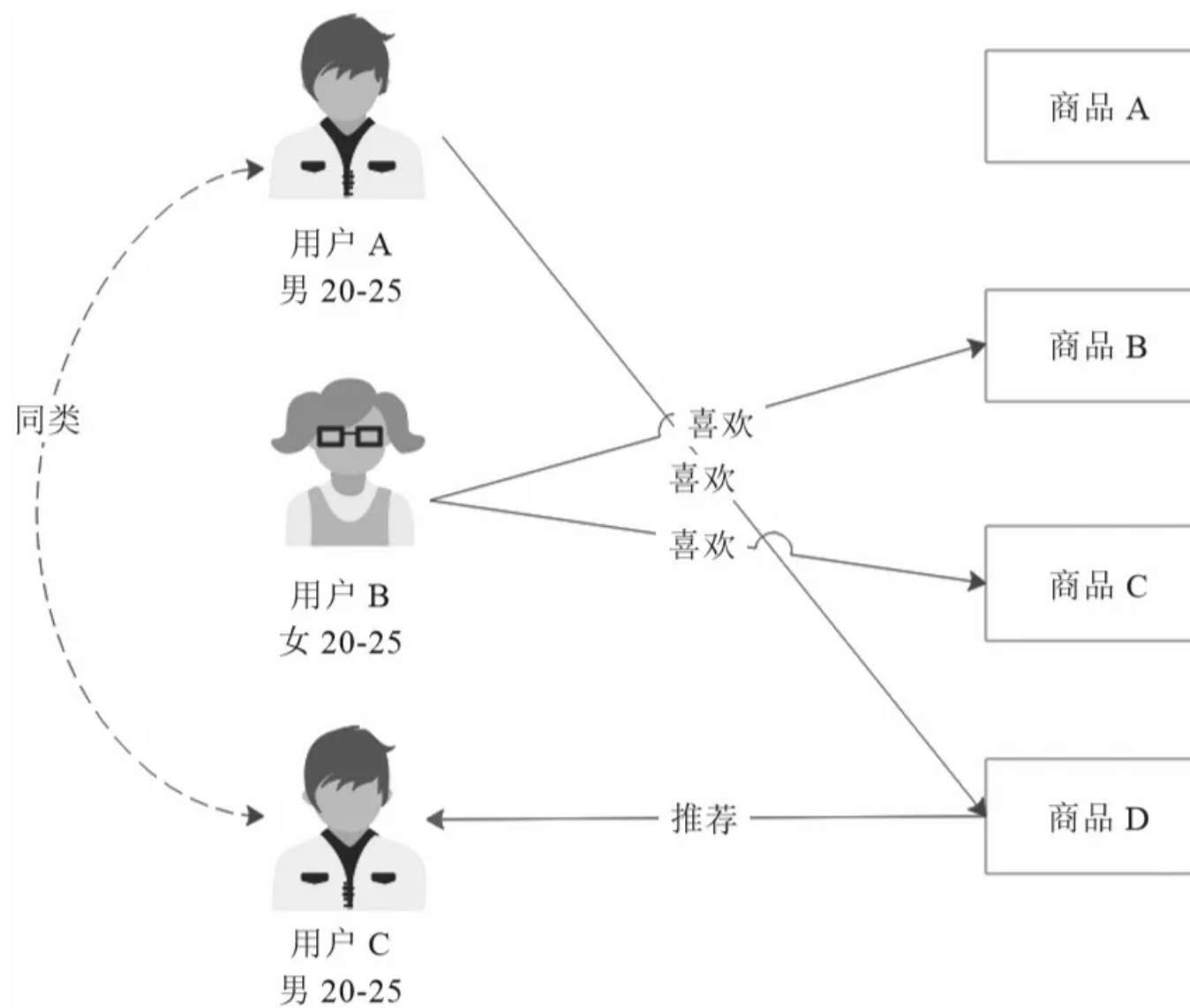
作者：李智慧 來源：大數據DT (ID: hzdashuju)



我們在淘寶購物，在頭條閱讀新聞，在抖音刷短視頻，背後其實都有智能推薦算法。這些算法不斷分析、計算我們的購物偏好、瀏覽習慣，然後為我們推薦可能喜歡的商品、文章、視頻。這些產品的推薦算法如此智能、高效，以至於我們常常**一打開淘寶就買個不停，一打開抖音就停不下來。**

01 基於人口統計的推薦

基於人口統計的推薦是相對簡單的一種推薦算法，它會根據用戶的基本信息進行分類，然後將商品推薦給同類用戶，如圖1所示。



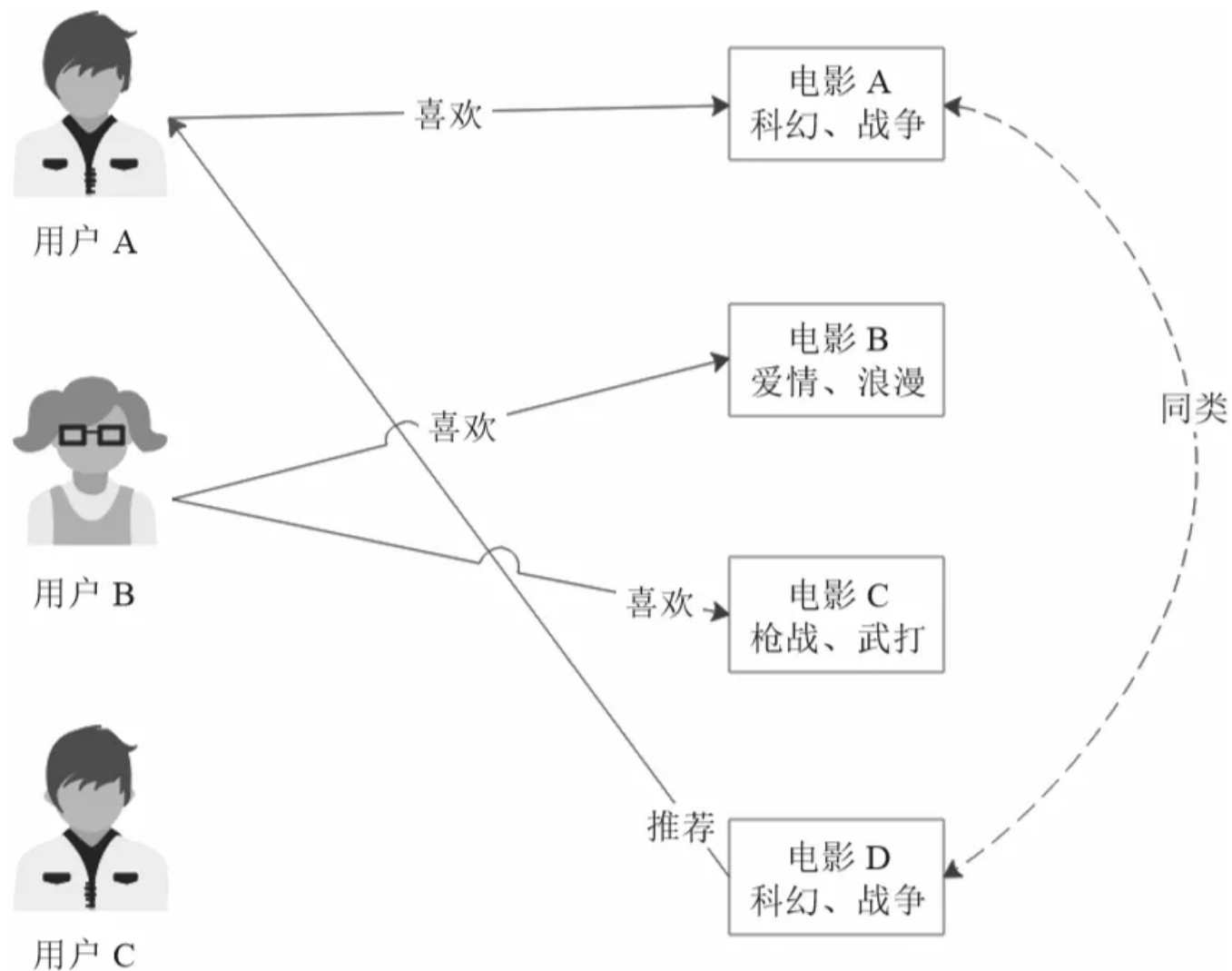
▲圖1 基於人口統計的推薦算法

用户A和用户C的年龄相近、性别相同，可以将他们划分为同类。用户A喜欢商品D，因此推测用户C可能也喜欢这个商品，系统就可以将这个商品推荐给用户C。

图1中的示例比较简单，在实践中，**还应该根据用户收入、居住地区、学历、职业等各种因素对用户进行分类，以使推荐的商品更加准确。**

02 基于商品属性的推荐

基于商品属性的推荐和基于人口统计的推荐相似，只是它是根据商品的属性进行分类，然后根据商品分类进行推荐的，如图2所示。



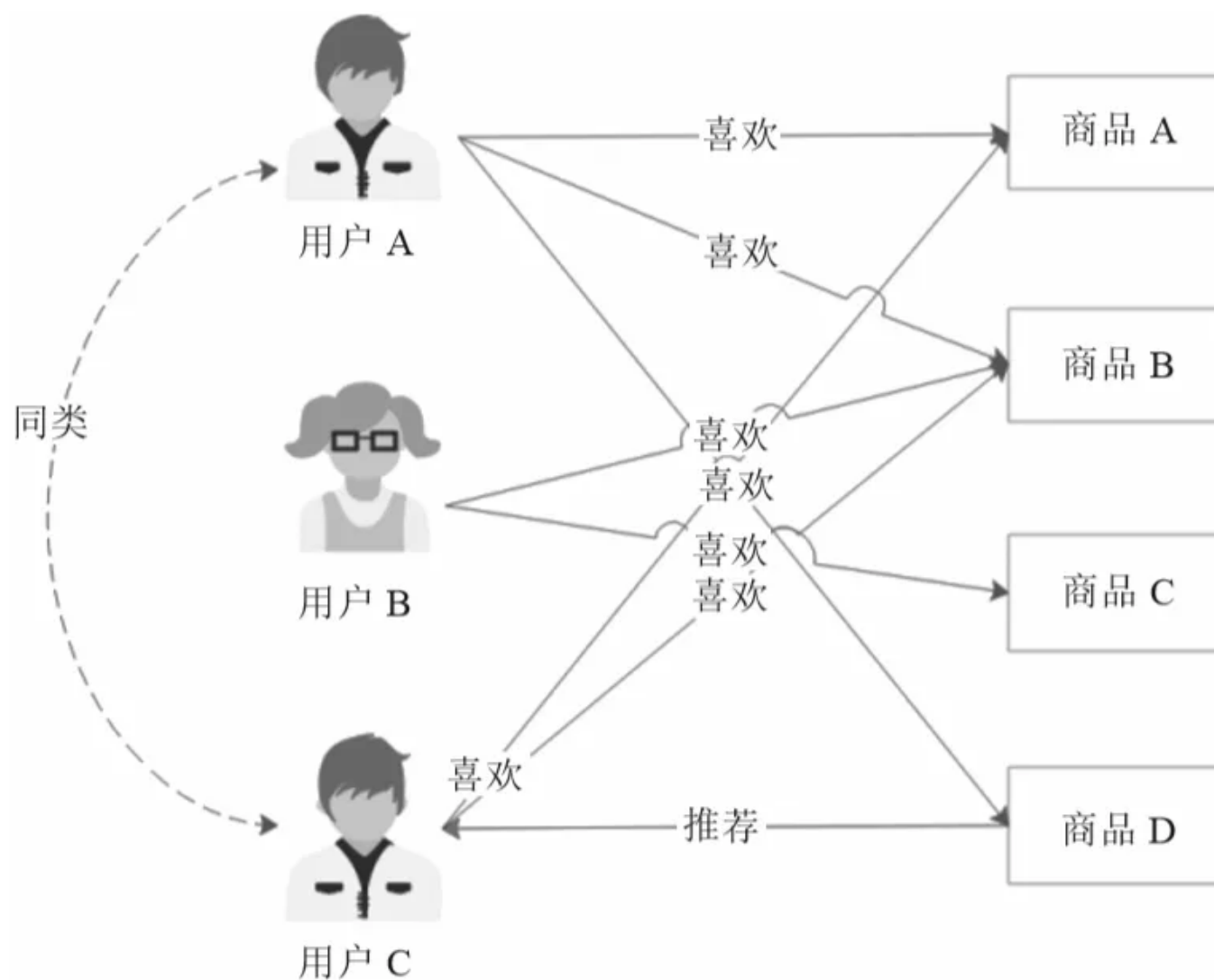
▲图2 基于商品属性的推荐

电影A和电影D都是科幻、战争类型的电影，如果用户A喜欢电影A，很有可能他也会喜欢电影D，因此就可以给用户A推荐电影D。

这我们的生活常识也是相符合的。如果一个人连续看了几篇关于篮球的新闻，那么再给他推荐一篇篮球的新闻，他很大可能会有兴趣看。

03 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，然后根据用户分类进行推荐，如图3所示。



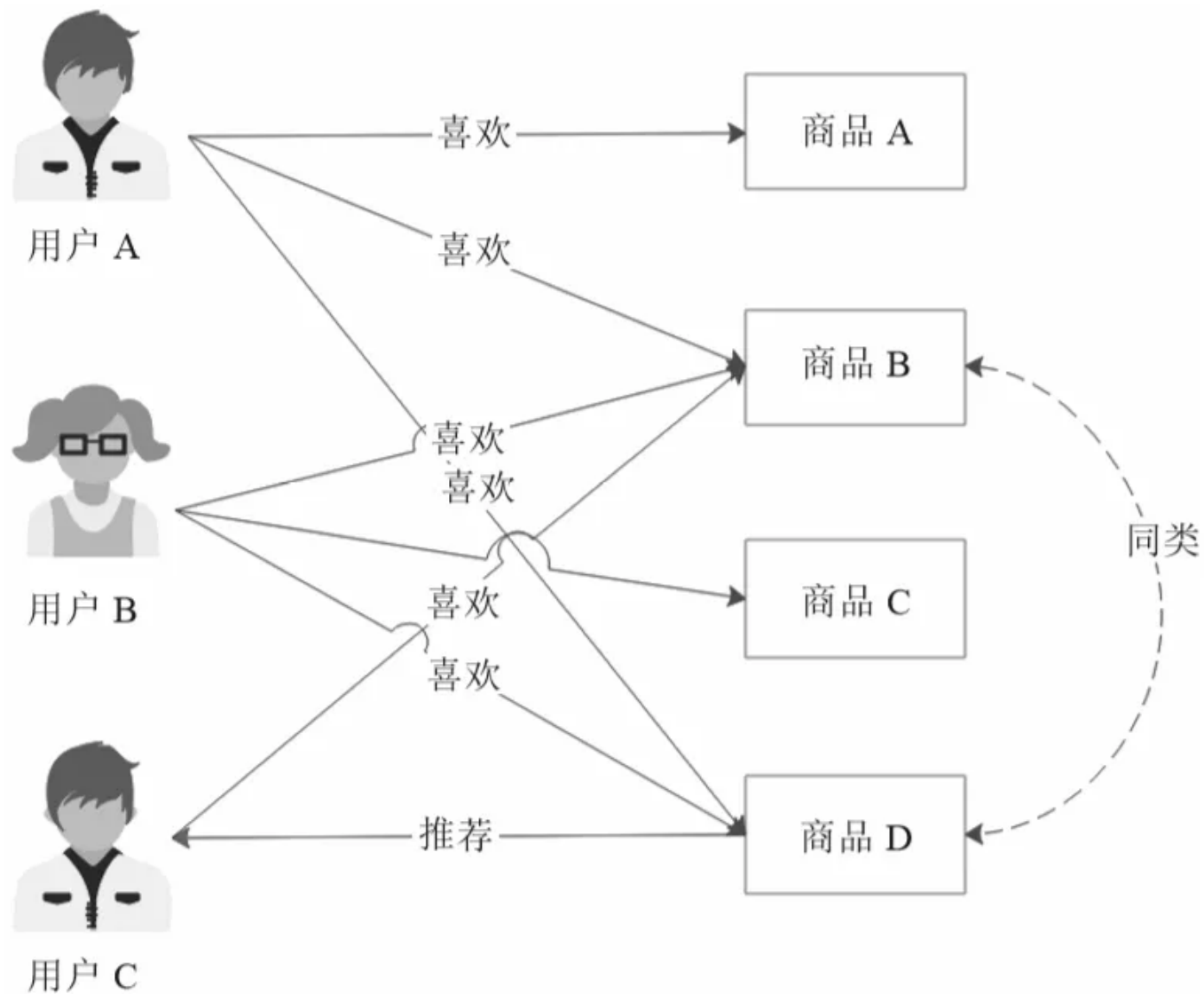
▲图3 基于用户的协同过滤推荐

这个示例中，用户A和用户C都喜欢商品A和商品B，根据他们的喜好可以分为同类。用户A还喜欢商品D，那么将商品D推荐给用户C，他可能也会喜欢。

现实中，**跟我们有相似喜好、品味的人也常常被我们当作同类，我们也愿意去尝试他们喜欢的其他东西。**

04 基于商品的协同过滤推荐

基于商品的协同过滤推荐则是根据用户的喜好对商品进行分类，然后根据商品分类进行推荐，如图4所示。



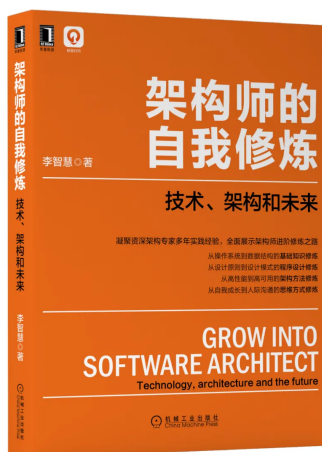
▲图4 基于商品的协同过滤推荐

这个示例中，喜欢商品B的用户A和用户B都喜欢商品D，那么商品B和商品D就可以分为同类。对于同样喜欢商品B的用户C，很有可能也喜欢商品D，就可以将商品D推荐给用户C。

这里描述的推荐算法比较简单。事实上，要想做好推荐其实是非常难的，用户不要你觉得他喜欢，而要自己觉得喜欢。现实中，**有很多智能推荐的效果并不好，被用户吐槽是“人工智障”**。推荐算法的优化需要不断地收集用户的反馈，不断地迭代算法和升级数据。

关于作者：李智慧，资深架构专家，同程旅行交通首席架构师，曾在NEC、阿里巴巴、Intel等知名企业担任架构师，也曾在WiFi万能钥匙等企业担任CTO。长期从事大数据、大型网站的架构和研发工作，领导设计过多个日活用户在千万级以上的互联网系统架构，实战经验丰富。曾设计、开发过 Web 服务器防火墙、分布式 NoSQL 系统、大数据仓库引擎、反应式编程框架等各种类型的软件系统。

本文摘编自《架构师的自我修炼：技术、架构和未来》，经出版方授权发布。



架构师的自我修炼：技术 架构和未来

作者：李智慧

 当当

 购买

延伸阅读《架构师的自我修炼：技术、架构和未来》

点击上图了解及购买

转载请联系微信：DoctorData

推荐语：大型网站技术架构作者李智慧新作，通过架构师的4项自我修炼，构建你的架构师知识体系，完整展示架构师修炼之道。



也可以加一下老胡的微信
围观朋友圈~~~

推荐阅读 （点击标题可跳转阅读）

10部顶级数学纪录片

2021年，机器学习研究风向要变了？

【机器学习】随机森林是我最喜欢的模型

Python之父：Python 4.0可能不会来了

【2021版】机器学习、深度学习调参手册

亚马逊首席科学家李沐博士：工作五年反思

【下载】80页笔记看遍机器学习基本概念、算法、模型

老鐵，三連支持一下，好嗎？↓↓↓

[閱讀原文](#)

喜歡此內容的人還喜歡

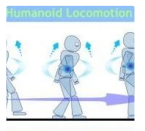
11 種數據降維算法，代碼已開源！

圖靈人工智能



17個機器學習的常用算法！

圖靈人工智能



基於Python 的11 種經典數據降維算法

Python那些事

