

用户推荐算法综述

14051628 孙一鸣

2016 年 12 月 28 日

用户推荐算法，能够通过数学变量的运算，推测出用户可能喜欢的东西。

在信息化社会的今天，优秀的推荐算法无疑降低了用户寻找优质内容的成本，同时也给广告定点推荐提供了良好的解决方案。

使用不同的产品，可以发现，推荐系统是给各大平台培养用户粘性的必备功能。比如，微博根据你关注的用户推荐相似你可能感兴趣的用户，虾米音乐使用“音乐基因”的方式推荐音乐，网易云音乐也提供歌曲每日推荐的服务，淘宝等电商平台则根据用户的购买记录，浏览记录实时推荐潜在的商品，知乎平台使用主题标签推荐内容，各大互联网新闻平台也提供实时准确的新闻推送，美团根据食客的评级推荐美食。

本篇综述将从推荐算法的作用，种类，效果评估，改进等，四个方面进行论述。

通过深入使用不同的平台，我发现推荐系统有如下作用：

1. 能够高效的帮助用户找到需要的商品或服务

互联网时代的商品从线下延伸到线上，使得商品的选择成指数级增长。用户往往在选择时，耗费大量的时间、精力。不过，使用推荐系统的平台，能够有效避免这方面的问题，同时，增加平台的收益。

2. 去处信息噪声

随着互联网的发展，网络每时每刻都在生产巨量的信息。个人用户要在浩如烟海的互联网寻找有价值的信息，是极其困难的。但是，推荐系统能够帮助用户过滤低价值的信息。

3. 对于互联网平台而言，提高了收益率和培养了用户粘性

观察中国的互联网环境，用户对内容的要求越来越高。低质量的内容输

出，已经满足不了用户的日常阅读需求。推荐系统则努力迎合用户的需求，秉承用户至上的原则。

4. 为广告业带来了机遇

在观看爱奇艺视频的时候，屏幕左下方经常出现男女主人公同款，电视剧商品同款等广告。贴片广告，实现了实时推荐，极大的提高了广告主的收益，同时，提高了用户的观看体验，附加体验。

在当今的互联网环境下，推荐系统的作用，被越来越多的企业所重视。各大企业也依据业务需求，开发了自己的推荐系统。无论推荐系统最终呈现出怎样的效果，他们都是基于以下几类基本的推荐算法。

1. 基于流行度的推荐算法

基于流行度的推荐算法是最简单的推荐算法。比如，微博根据新闻/事件的浏览量，评论量等大数据进行内容排序，从而将排名靠前的内容推荐给用户。

优点：适合刚刚注册的用户。比如，我今天刚申请微博，微博给我推送唐嫣和罗晋恋情。

缺点：不能提供个性化的推荐服务。

2. 协同过滤算法（Collaborative Filtering, CF）

协同过滤算法通过对未评分项进行评分，预测来实现。CF 分为 User-based CF 和 Item-based CF。

User-based CF 原理	Item-based CF 原理
1. 分析各个用户对 item 的评价（历史记录） 2. 根据用户对 Item 的评价计算得出所有用户之间的相似度 3. 选出与当前用户最相似的 N 个用户 4. 将 N 个用户评价最高并且当前用户没有浏览过的 item 推荐给当前用户	1. 分析各个用户对 Item 的浏览历史 2. 根据 history 分析出所有 item 的相似度 3. 对于当前用户评价高的 item，找出与之相似度最高的 N 个 item 4. 将 N 个 item 推荐给当前用户

优点: CF 算法描述简单, 实现容易; 推荐结果准; 推荐个性化, 自动化程度高; 能处理复杂的非结构化对象。

缺点: 对于新用户, 推荐效果差; 推荐数据取决于历史数据集; 存在矩阵稀疏问题。

3. 基于内容的推荐算法 (Content-based Recommendation)

CBR 是建立在项目的内容信息上做出推荐, 不需要依据用户对项目的评价, 更多的需要依据用机器学习的方法从关于内容的特征描述中得到用户感兴趣的资料。比如我喜欢王小波的《青铜时代》, 在当当网上买了王小波的《青铜时代》, 那么, 基于内容的推荐算法会自动给我推荐王小波的《白银时代》、《黄金时代》。

优点: 没有冷启动问题和稀疏矩阵问题; 能为具有特殊兴趣爱好的用户进行推荐; 技术相当成熟。

缺点: 对于过度专业化的问题, 采用 CBR 算法, 系统会一直推荐给用户内容密切相关的 item, 从而失去推荐内容的多样性、丰富性。

4. 基于模型的推荐算法 (Model-based)

基于模型的推荐算法与基于 Memory 的推荐算法相对。当数据量特别大时, 将用户数据, 读入内存, 进行运算, 开销较大, 如果使用基于 model 的推荐算法, 依托于机器学习的 model, 离线训练, 在线推荐, 将会减少开销。

常用的模型有:

贝叶斯分类, 决策树, 神经网络, 专家系统, 知识推理, 矩阵分界, 概率模型, 图模型, Boosting 等

优点: 推荐速度快, 准确度高, 适用于实时性较高的业务, 比如微博, 澎湃新闻等。

缺点: 属性的组合和筛选取决于模型, 如果数据集和模型之间存在较大的差异, 推荐效率将大幅度下降。

5. 组合推荐算法

在实际的应用场景，使用单一的推荐算法往往会导致推荐结果很差。所以，为了提高推荐效率，常见的做法是，综合多种推荐算法。比如，通过对不同的推荐算法计算出的结果进行加权处理，也可以在不同的计算环节运用不同的算法来混合结果。

评估推荐算法：

评价指标	具体内容
推荐准确度	通过计算预测 item 的预测评分与真实评价数据上的差别，来衡量推荐结果的准确性
覆盖率	尽量所有的商品，都可以找到各自适合的用户
多样性	考察推荐算法给用户推荐产品或者服务的多样性
信任度	在给出推荐商品或者服务的同时，给出推荐理由
健壮性	推荐系统能够自主识别恶意的点击浏览行为
停留时间	对于音乐 app，用户的停留时间
点击率和购买率	类似淘宝这样的电商平台，更多关注推荐内容的点击率，和推荐商品的购买率

改进推荐算法。

1. 各大企业间加强战略合作，分享用户数据，解决推荐系统的冷启动问题
2. 丰富用户的属性，兴趣爱好，人际关系等
3. 对于社交平台，根据用户的人际关系，推测用户的兴趣爱好等内容。
4. 推荐算法在云计算平台上进行，将解决推荐算法大数据运算问题

总结

推荐算法是极其有趣的领域。随着技术的发展，计算机能够准确的知道用户想购买的商品、性格、爱好，推荐系统降低了我们寻找内容的成本。问题是，我们可能愈加依赖推荐系统。我认为，寻找喜欢商品，音乐，朋友，是一件私人的事情，减少对计算机的依赖，培养自己的眼界、品味更加重要。