

# 算法向善与个性化推荐 发展研究报告

# 目录

第一章 推荐系统.....	
1.1. 什么是推荐系统.....	
1.2. 推荐系统在人类生活中的体现方式.....	
1.2.1. 资讯推荐.....	
1.2.2. 视频、音乐推荐.....	
1.2.3. 电商平台推荐.....	
1.2.4. 电子政务推荐.....	
第二章 推荐系统的研究背景及国内外研究现状.....	
2.1. 推荐系统的发展历史.....	
2.2. 推荐系统的价值.....	
2.2.1. 整合市场需求.....	
2.2.2. 技术发展规律.....	
2.2.3. 产业发展需求.....	
2.3. 推荐系统的研究现状.....	
2.3.1. 学术界.....	
2.3.2. 产业界.....	
第三章 推荐领域面临的问题.....	
3.1. 信息茧房.....	
3.2. 算法黑盒.....	
3.3. 价值观及伦理问题.....	
3.3.1. 虚假新闻与低俗内容泛滥.....	
3.3.2. 算法偏见和歧视.....	
3.3.3. 算法公平性.....	
3.4. 用户隐私与安全问题.....	
第四章 推荐领域研究趋势.....	
4.1. 信息茧房.....	
4.2. 算法黑盒.....	
4.3. 价值观及伦理问题.....	
4.4. 用户隐私与安全问题.....	
第五章 推荐算法面临问题总结与思考.....	

# 第一章 推荐系统

## 1.1. 什么是推荐系统

推荐系统作为一个专业术语，对于人们来说可能比较陌生，但实际上，我们已经逐渐习惯了它的存在。无论你是用抖音刷着小视频，或是在今日头条 APP 中浏览着资讯信息，还是打开淘宝找寻想要购买的商品，你无时无刻都在和推荐系统进行交互。设想一下，当你打开今日头条 APP 想要看一会儿最新资讯时，主页面会给你自动推荐你感兴趣的领域的信息，在你选择其中一则新闻并看完之后，你发现在该资讯的页面下方出现了许多可能与你所看资讯有关的信息，在这个场景中决定给您推荐哪些信息的决策过程就是推荐系统。



图 1 现实生活中常见的推荐系统。

随着信息技术的发展和电子设备的广泛使用，我们现在已经进入了数据大爆炸的时代。据互联网数据中心发布的《数据时代 2025》报告显示，全球每年产生的数据将从 2018 年的 33ZB（十万亿亿字节）增长到 175ZB，相当于每天产生 491EB（百亿亿字节）的数据。虽然相比于过去信息匮乏的时代，如今信息量的增加值得欣喜，但是一旦信息量过大，我们就需要合适的方法来获取需要或者感兴趣的信息，否则杂乱无章的数据对于我们来说是无益的。除此之外，信息量过大也会导致数据信息的质量无法得到保证，海量数据通常会掺杂许多没有意义甚至虚假的信息，形成噪声，对信息的使用和挖掘产生负面影响。面对包含噪声的海量数据，对于用户来说，用户无法或者很难在其中找到满足自己真正需要的内容或者对自己真正有益的信息；对于企业来说，企业无法有效管理和分析这些数据，也无法通过简单的技术手段挖掘其中的有效信息来为用户提供他们可能感兴趣的内容。面对这种窘境，推荐系统应运而生，旨在从海量数据中寻找满足用户意图或者兴趣偏好的内容，从而减少用户浏览大量无效数据而造成的时间和精力上的浪费。本质上，推荐系统是一个信息过滤系统，是用户历史数据、推荐模型和用户交互行为循环交互的过程。具体而言，其工作原理及其流程

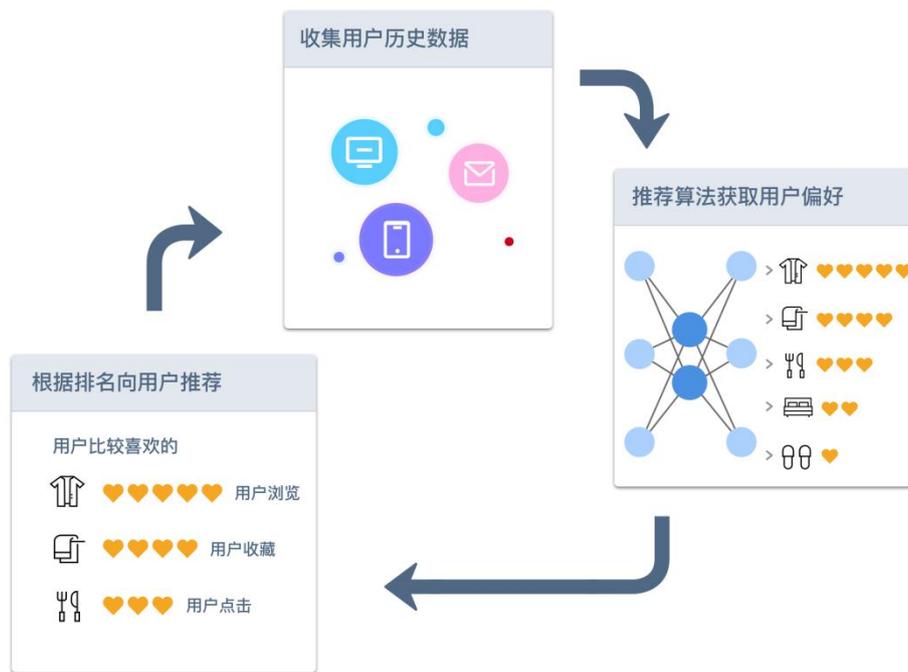


图2 推荐系统工作原理及其流程

主要包括 3 个步骤：（1）从移动端、网页、社交网络等各种途径获取用户大量的行为历史数据，通过清洗和筛选产生有效的用户日志数据；（2）推荐算法从用户日志数据中挖掘用户特征，对用户的兴趣偏好进行建模，预测用户对候选商品、服务或者内容的喜爱程度；（3）根据预测的用户对候选商品、服务或者内容的喜爱程度对其排序，将排名较高的商品、服务或者内容推荐给用户。用户对推荐的服务或者内容进行浏览、点击或者购买等行为，形成新的用户交互记录，用于更新用户兴趣偏好。

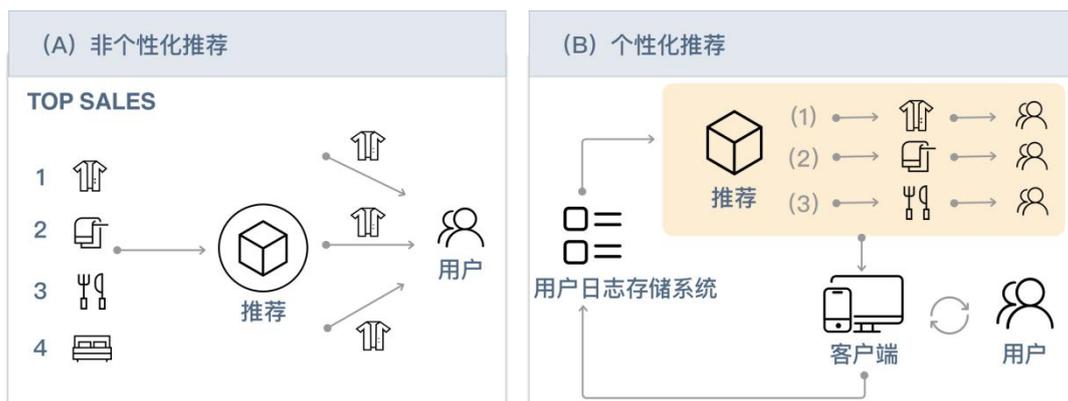


图3 推荐系统分类

纵观推荐系统的发展历史，可以将推荐系统分为非个性化推荐系统和个性化推荐系统。顾名思义，非个性化推荐就是不考虑每个用户的特性和其历史行为，忽略对用户兴趣偏好的个性化建模，使得推荐结果形式化的展示为‘千人一面’。该类推荐系统直接根据点击量、销量、热度、流行度等属性特征将物品、服务或者内容排序给用户做推荐。这样的推荐模式

使得即使是有不同兴趣偏好的用户也会拥有几乎相同的推荐结果；而个性化推荐系统就是为每一个特定用户提供特定的服务，实现‘千人千面’。个性化推荐系统的核心功能是根据用户的历史行为、社交关系、兴趣点、所处上下文环境等信息来识别用户的兴趣偏好，借此预测用户是否对某种物品、服务或者内容感兴趣，并根据可能感兴趣的程度高低来给用户进行推荐。总的来说，个性化的推荐算法是现今推荐产品的主流技术，也是推动推荐系统不断优化核心。同时，个性化程度的高低是决定推荐系统性能好坏的关键，也是决定用户留存和产品是否能长远发展的关键因素。

## 1.2. 推荐系统在人类生活中的体现方式

进入人工智能时代，推荐系统以各种各样的方式影响着我们的生活。日常生活中经常使用的各种软件以及主流的 App 都基本应用到了推荐系统。例如，今日头条、腾讯新闻等会给你推送你感兴趣的资讯；京东、淘宝、亚马逊、拼多多等会给你推荐“可能喜欢”的物品；抖音、快手、西瓜等短视频 App 会给你推送感兴趣的短视频；饿了么、美团、大众点评、小红书等会给你推荐可能喜欢的食物或者饭店；携程、去哪儿等都会给你推荐机票、酒店等等。基于推荐系统潜在的应用价值，本小节简单介绍几种常见的推荐系统应用场景。

### 1.2.1. 资讯推荐

随着交互式通信技术的发展，互联网已经成为人们获取信息的主要渠道。互联网在每一秒都会有大量新的资讯出现，为人们了解和认识世界提供源源不断的资源。随着时代更替，科学技术得到了一定的发展，与传统报纸、电视等媒体相比，在互联网时代下产生的网络信息平台在一定程度上缓解了海量资讯生产和分发的瓶颈问题，但是对于新闻等资讯的价值界定并未作出明显区分。造成这种局面的可能原因之一就是缺乏规定的程序来及时提供各种各样的资讯，以及系统无法以更好的方式模拟用户兴趣偏好。因此，网络应用逐渐运用推荐系统技术，根据读者和社会的需求及时更新资讯，进一步对用户偏好建模，为其呈现个性化推荐结果。在资讯推荐场景中，推荐系统需要关注的点是如何理解文章的内容和用户的偏好，并基于此为用户提供满足其兴趣的推荐资讯列表。与传统媒体相比，基于个性化的资讯推荐相当于为每一个用户提供一份自己喜欢看或者感兴趣的‘报纸’。与人们日常生活息息相关的资讯推荐系统，如今日头条、腾讯新闻、百度首页新闻推荐模块等根据用户的观看、浏览记录、用户的评论以及用户的搜索意图等显式和隐式的反馈来建模用户的兴趣偏好，为用户呈现其感兴趣的推荐结果。以今日头条的推荐系统为例，如图 4 所示，当用户打开今日头条 App 时，系统的推荐模块根据用户历史观看过的内容、用户关注的其他用户和账号等向用户推荐其感兴趣的新闻内容。比如用户点击或搜索过关于新冠疫苗的内容，推荐模块会给用户推荐对接种新冠疫苗的注意事项以及该不该打疫苗的内容；热榜模块会根据话题的热度、点击量、国情要事等向用户呈现当下较为受关注的资讯，比如对于国外品牌抵制新疆棉

花以及国内相关部门如何应对此事件的内容；充分体现国情的抗疫模块向用户及时推荐国内以及全球的疫情新闻和动态，比如关于本土的新增病例数量及分布的情况，接种新冠疫苗的情况等；今日头条综合模块中的头条寻人功能综合人脸识别和推荐技术向社区推荐寻人消息，集大众力量给走失家人的家庭提供重新团聚的可能。该功能充分体现了以用户价值服务大众，取之于民用之于民的理念。

此外，与人类生活密切相关的资讯类推荐 App 还有腾讯新闻、北京时间、凤凰新闻、360 快资讯、新浪看点等，通过推荐算法为用户找到相匹配的资讯内容，以提供不同领域、不同形式的资讯来丰富用户的认知世界。在个性化推荐的时代，为了满足受众的个性化需求，这类以推荐算法为主的资讯平台将资讯主体进行重构，使得资讯分发由原先的以传播者为核心逐渐向以受众为核心进行转变，逐步提升了受众在传播过程的地位。同时，个性化资讯推荐力求精准的满足用户差异性的兴趣需求，为用户提供其感兴趣的新闻，有助于实现私人订制。总的来说，个性化资讯推荐在人们日常生活中扮演着重要角色，是人们了解世界、接触世界的有效途径。它的出现不仅是科学技术的进步的体现，也是提升人民主体地位的重要手段。



图 4 资讯类推荐在人类生活中的体现举例

### 1.2.2. 视频、音乐推荐

随着移动设备的广泛使用，以视频、音乐为主的娱乐领域得到了快速发展。许多相关平台已经打破了专业人士和一般用户之间的界限，使得在视频和音乐等资源上的创作增长尤为迅速。聚焦到视频产业中，一类是专业视频内容为主的网站，例如 HBO、Netflix、腾讯视频、优酷视频、爱奇艺等，而另一类以用户生产内容为主的网站，即 UGC 网站，例如 YouTube、抖音、西瓜视频、哔哩哔哩。这些视频网站或者 App 为人类提供了多姿多彩的智能生活。第一类视频推荐网站主要以长视频为主，根据用户观看视频的类型、题材、相关的主体人物（导演、演员等）为用户推荐其感兴趣的长视频。比如某用户观看了具有战争、奋斗、红色主题的视频《觉醒年代》，推荐系统推测用户可能喜欢具有爱国题材的视频，可将与《觉醒年代》具有相似属性的视频《理想照耀中国》推荐给用户。现实生活中，典型的



图 5 视频、音乐类推荐在人类生活中的体现举例

以长视频推荐为主 App 腾讯视频的“猜你会追”模块、爱奇艺的“猜你喜欢”模块、优酷的“猜你在追”模块，根据用户曾经观看过的视频属性向用户推荐其可能感兴趣的视频（包括但不限于电视剧、电影和综艺等）。相应的，以用户生产内容为主的网站侧重短视频的推荐，根据用户观看记录、所处地域以及用户关注的短视频博主和账号等信息为用户推荐其感兴趣的短视频。相比如长视频推荐，短视频可以充分利用用户的碎片化时间，为用户呈现无缝的观看体验，以抖音短视频为例，在主界面提供“推荐”、“同城”和“关注”模块，系统会向用户实时推送其感兴趣的短视频。在“推荐”模块，推荐系统根据用户历史点击或观看的小视频的点赞量、评论量、转发量和完播率判断用户偏好，向用户推荐其可能感兴趣的短视频。在“关注”模块，推荐系统根据用户关注的抖音账号、用户标签和内容标签，向用户智能分发其可能感兴趣的内容。比如用户关注了“中央新闻”抖音账号，推荐系统捕捉到用户比较关注国家大事，就向用户推荐国情要事；在“同城”模块，根据用户的地理位置，将与用户地理位置相近的发生故事、视频等推荐给用户。在对用户进行个性化智能推荐的同时，为加强对青少年的保护以及防止青少年沉迷网络，抖音积极响应国家号召，纷纷开启青少年模式，设置时间锁，规定青少年模式下无法充值，以加大对未成年人良性上网的引导力度。

在音乐推荐场景下，为了防止用户在多元化产品中无法找到自己感兴趣的点，许多企业

已经开发了音乐推荐系统支持用户自己上传音源，利用用户的历史兴趣（曾经听过的音乐的风格、发行者、作词者等信息）以及其好友的相关信息为用户提供个性化的推荐结果，以提高用户的体验感。以 QQ 音乐为例，在“今日为你打造”模块，推荐系统根据用户历史点击或所听的音乐记录，向用户推荐满足其偏好的个性化电台、每日 30 首等音乐列表；在“智能分类”模块，推荐系统根据用户的历史信息以及相互关注的好友收听的音乐信息向用户推荐其可能感兴趣的音乐。同时，QQ 音乐还提供直播模块，满足用户对音乐多元化的需求，提高用户满意度。此外，QQ 音乐还智能化的设置用户的个性化权限，非常友好得让用户自由切换推荐模式。

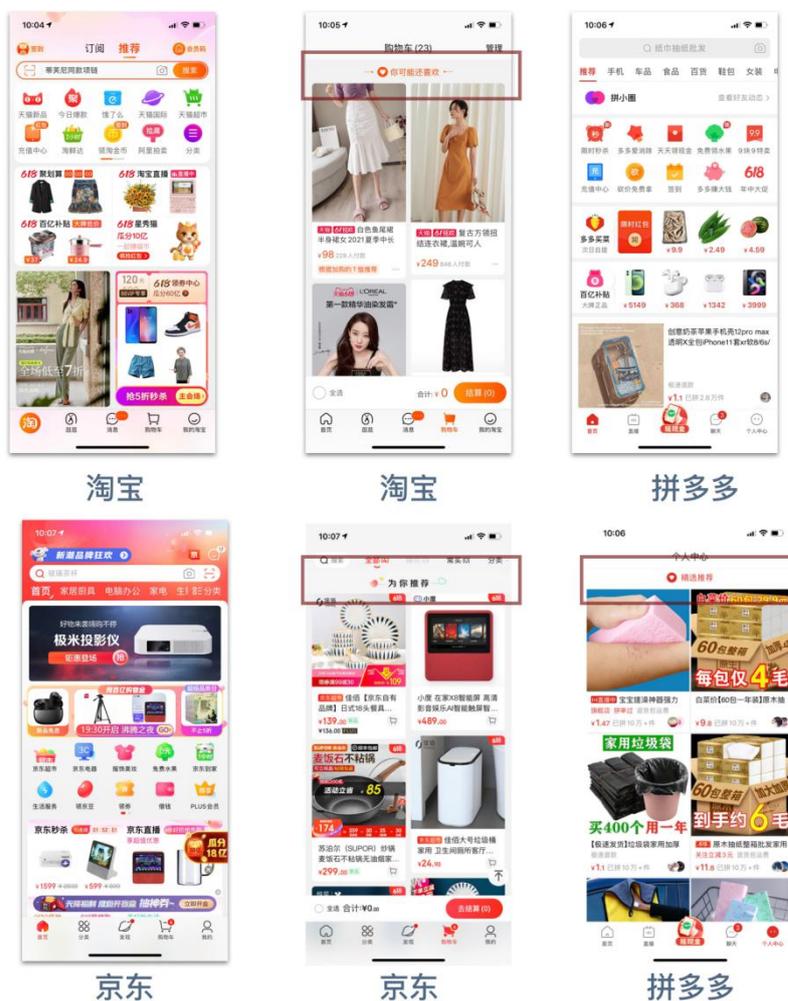


图 6 电商推荐在人类生活中的体现举例

### 1.2.3. 电商平台推荐

随着互联网的普及，电商产业发展迅速，为人们提供了更加便捷的购物渠道，使得人们的生活更加便利。但随着流量红利的消退，提高用户满意度、用户黏度以及用户转化率变得越来越重要，这些因素直接决定了电商平台能否实现更多的利润和长期发展。推荐系统通过

推荐算法为用户提供满足其兴趣偏好的信息或者服务,在很大程度上能够提高用户满意度和用户黏性,从而提高用户的转化率。一般的,电商平台将推荐算法挑选出来的推荐商品放置于用户主页,当用户能够发现自己感兴趣的商品时,他们会继续点击商品详情浏览产品,而如果用户浏览片刻后无法找到满足自己需要的产品,则会离开页面,间接造成用户流失以及点击量的减少。因此,个性化的精准推荐逐渐成为各电商平台非常看中的技术手段,也成为实现电商平台长期稳定发展的根本要素。到目前为止,国内外著名的电商平台,像亚马逊、淘宝、京东、拼多多等已经相继开发自己独有的推荐系统用于提高用户黏度和转化率。以阿里巴巴的淘宝网举例,在“推荐”和“你可能还喜欢”模块,推荐系统根据用户的历史点击行为,向用户推荐可能感兴趣的产品,比如,当用户浏览过“半身裙”,推荐系统捕获到用户的购买意图,向用户推荐各种款式和价位的“半身裙”以及与“半身裙”相关的产品“连衣裙”等。这样的推荐系统能够很灵活地捕捉到用户的兴趣偏好,及时将用户可能购买的商品推荐给用户。

基于电商领域的特点,推荐系统在商品推荐中不仅仅只关注商品本身,还需要考虑到不同商品有不同的复购周期,例如用户可能会在短期内购买衣物、食品等商品,但是手机、家具一般是不可能在短时间内再次被用户购买的。除此之外,物品之间购买的相关性、时序性也是需要考虑的。如此看来,推荐系统在电商中是一种多场景推荐算法的融合,同时由于目前还有许多更为复杂的逻辑关系在推荐系统中尚未考虑,推荐算法在电商场景上还有很大的优化空间。

#### 1.2.4. 电子政务推荐

在互联网时代,面对互联网的高速发展带来的信息过载,企业和公民无法从所面对的信息范围中做出有效的选择,而且用户难以找到正确的信息,最终影响用户的忠诚度。为顺应时代的发展,政府也在逐步实现线上政务,从而为公民和企业提供更好的信息和服务。电子政务推荐算法则是希望通过综合用户数据优先向用户推送利益关联最大的办事项,虽然有研究者提出该方面的理论,但是近几年相关研究较少,可以看到推荐系统目前在电子政务上应用不广,我国许多电子政务网站依旧是按照时间排列,无法满足公民和企业的个性化需求,因此推荐系统与电子政务是未来值得研究的一个方向。

为方便读者理解不同推荐场景下推荐算法的不同侧重点,本小节将推荐系统按照使用的特征以及关注的内容、物品或服务的类型等,形式化地总结了各类主流应用中推荐算法相关的信息。

推荐场景	推荐产品	推荐项	推荐使用的特征
资讯推荐	今日头条、百度新闻、腾讯新闻、知乎、北京时间、凤凰新闻、360	资讯	新闻标题、内容、主题、用户行为、用户画像、年龄、性别、教育背景、用户社交网络、设备、时间、位置

	快资讯、新浪看点		
视频、音乐 推荐	抖音、快手、西瓜视频、YouTube、哔哩哔哩	短视频	视频内容、音频、类型、热度、标题和描述、用户行为、用户社交网、用户人口统计学信息、设备、时间、位置
	腾讯视频、爱奇艺、优酷、HBO、Netflix、芒果TV	长视频	视频内容、音频、类型、热度、标题和描述、用户行为、用户人口统计学信息、设备、时间、位置
	QQ音乐、网易云音乐	音乐	音频类型、演唱者、作曲者等、用户行为、用户人口统计学信息
电商推荐	淘宝、京东、亚马逊、拼多多、考拉海购	商品	商品标题、描述、图片、类别、流行度；用户行为、反馈、用户人口统计学信息

表 1 不同推荐场景下的推荐应用总结

总结发现，推荐系统的主要原理是从用户的历史交互日志中挖掘用户特征，继而将满足用户偏好的物品或者服务推荐给用户。其接受信息的方式是被动且需求是模糊而不明确的。以电商场景为例，用户在购物需求不明确的情况下进入商场，需要推荐系统来告诉用户有哪些优质的商品、哪些合适的内容等。但如果用户购物需求非常明确，知道当下需要购买哪个品牌、什么型号的商品时，直接搜索对应的店铺购买就行，这就是搜索引擎的功能。因此，无论是在互联网还是线下场景中，为满足用户的这两种需求，推荐系统和搜索引擎都是联系十分紧密且大量并存的。尽管两者都是用户获取信息的途径，但两者存在本质的区别：在获取信息方式上，推荐系统是被动且需求模糊，而搜索引擎是主动且需求明确；在对结果的关注上，推荐系统重视个性化的推荐结果，期望对不同的用户有不同的推荐结果，而搜索引擎重视结果的排序是否满足用户的搜索意图，需要将好的结果尽可能排在前面。推荐系统和搜索引擎虽然有很多差异，但两者都是大数据技术的应用分支，存在着大量的交叠。近年来，逐步融合推荐系统和搜索引擎已经成为人工智能领域的研究热点。

## 第二章 推荐系统的研究背景及国内外研究现状

### 2.1. 推荐系统的发展历史

随着信息技术和互联网的快速发展，人类从信息匮乏时代走向了**信息过载 (Information Overload)** 时代。信息过载是当今信息化时代经常被研究者提起的一个词，它形容的是信息量远远超过了个人或系统处理信息的能力，从而导致信息处理效率下降的现象。信息过载是信息化带来的负面影响之一。面对信息过载现象，对于信息消费者（用户）来说，从大量信息中找到自己感兴趣的信息变得越来越困难；对于信息生产者，让自己生产的信息在众多信息中脱颖而出也变得越来越困难。信息过载虽然是近期才被人们广泛提出，但是在上个世纪 80、90 年代就已经有了这个概念。当时由于电子邮箱逐渐普及，人们收到的消息繁多而杂乱，用户通常没有足够的时间去阅读所有的邮件，这也就导致了人们难以在这么多信息中找到真正对自己来说重要的信息。早期的研究提出了通过信息检索和过滤 (**information retrieval and filtering**) 的方式来解决这个问题，到了 90 年代中期，许多研究者开始关注用户评级数据，通过预测用户对推荐的物品、内容或服务的评分来解决信息过载问题。因此，推荐系统作为独立的研究领域出现了。

推荐系统的兴起与互联网的发展息息相关，早期的推荐算法研究延续了信息检索和过滤的相关思想，产生了大量经典的协同过滤推荐算法。最早的自动化协同过滤系统可以追溯到 1994 年明尼苏达大学双城分校计算机系的研究组设计的 **GroupLens** 新闻推荐系统，该系统收集了读者对于文章的评分，并使用这些评分来预测其他未阅读文章的读者希望阅读某篇文章的程度。在之后十几年的时间内，其他一些著名的协同过滤算法也相继被提出，以不断丰富、完善和突破现有的推荐技术。在推荐系统的发展历程中，将推荐系统的研究推向高潮的标志性事件是 **Netflix** 在 2006 至 2009 年举办了 **Netflix Prize** 比赛，该比赛宣布提供一百万美元奖金给第一个能将现有推荐算法的准确度提升 10% 以上的参赛者，吸引了众多专业人士开始投身于推荐系统领域的研究工作，也开启了推荐领域的机器学习时代。早期的以机器学习为主的推荐系统主要关注如何提高针对单一问题的算法方案的准确度，而忽视了人机交互的时间性和系统性，对于变化莫测的用户行为和外部环境无法完整建模。其中最典型的问题就是推荐结果千篇一律，例如若一个用户喜欢汽车，那么系统将会反复给他推荐汽车相关的信息。在推荐算法经典框架下，基于监督学习的模型大多无法彻底解决这一问题。面对这一问题，雅虎研究院最早开始了在这一方面的探索，并尝试将人工智能中另一方向——强化学习相关的算法应用到推荐系统中来，雅虎新闻推荐和 **Linked In** 的首页信息流推荐都应用了强化学习，在数据量较少时，他们证明了强化学习对系统和用户的长期交互流程优化有显著效果。然而，这类算法在大规模的推荐系统实践下并不理想。伴随着深度学习的浪潮对整个人工智能领域的影响，许多研究者开始研究如何将深度学习与传统算法结合来优化推荐效果。其中一个典型案例就是 **YouTube** 采用深度学习来提取视频特征，通过视频特征和

用户兴趣偏好的匹配度，决定是否给用户进行推荐。同时，也有许多学者尝试用深度学习技术解决以往较难的推荐问题，例如预测用户下一个想要播放的歌曲、购买的商品以及想看的电影等。深度学习技术对于推荐系统的影响是巨大的，从开始的非智能时代一步步踏入深度学习时代，推荐系统已经成为了许多应用的基石，对于基于大数据的人工智能时代具有非凡的影响。但从现实来看，推荐系统要真正成为智能系统，还有很长的路要走，这也是未来学者研究的方向。

推荐系统发展至今，其核心技术可以划分为基于协同过滤的推荐方法，基于内容的推荐方法和混合推荐方法。协同过滤推荐方法的本质是“人以群分，物以类聚”，根据相似的用户具有相似的喜好，推荐给他们喜欢的物品、内容或者服务（后文统称为“项”），或者将与用户曾经浏览、收藏、购买过的相似项推荐给用户。基于内容的推荐方法一般只依赖于用户自身的行为为用户提供推荐，不涉及到其他用户的行为。该类方法根据项的相关信息、用户相关信息及用户对项的操作行为来构建推荐算法模型，为用户提供推荐服务。一般的，项的相关信息可以是项的描述信息、标签、用户评论、人工标注信息等；用户相关信息可以是用户的职业、年龄、性别、偏好、地域、收入等；用户对项的行为信息可以是评论、收藏、点赞、观看、浏览、点击、加购物车、购买等。而混合推荐方法目的是衡量各推荐方法的利弊，扬长避短，通过加权、切换、混杂、特征组合等方式避免或弥补各推荐技术的弱点，进一步提升推荐方法的性能。

## 2.2. 推荐系统的价值

### 2.2.1. 整合市场需求

进入人工智能时代，由于数据量的激增，用户有着希望能从海量数据中得到对自己来说有用信息的需求，而信息生产者也有期望能让自己生产的信息被看到的需求，推荐系统通过算法计算用户与信息生产者生产的内容之间的匹配程度，将用户和信息生产者联系起来，实现对这两种潜在需求的整合，从而减少了用户和信息生产者在市场上寻找双方的时间，在一定程度上提高了市场运行的效率，使得用户和信息生产者的需求都能得到满足。



图 7 推荐系统改变了用户和信息生产者的联系方式

### 2.2.2. 技术发展规律

当推荐算法并未成为推荐系统的核心时，推荐大部分是由人工完成的，根据一些专业人士的建议和想法，为用户提供他们认为值得推荐的物品或服务。本质上，传统报纸就可以看成是这样的一个依赖人工的推荐系统，用户在报纸上所能看到的内容全都是专业编辑认为应该关注的文章。人工推荐系统是一种低效的方式，它对于人工的数量和专业素养要求都很高，很难处理大量数据，并且人工推荐总是会带有主观性，无法满足每一个用户的个性化需求。

从技术发展规律可以看到，技术总是在不断提高社会运行效率的，不断使社会多元化。过去人们听歌的方式是通过广播或者自己购买碟片和磁带，这些歌曲往往是热门歌曲，小众歌曲和歌手往往没有对口市场或者很难找到对口市场，而现在的音乐软件中各种类型的音乐都有，不管是流行还是一些小众音乐都有其喜好群体。因此，推荐算法的发展也是合乎技术发展的规律的，它让每个人都能拥有属于自己的个性化服务。

### 2.2.3. 产业发展需求

在传统营销方式中，存在一个大家普遍承认的定律——二八定律，这是一位意大利的经济学家在 1897 年归纳出了一个统计结论，即 20% 的人口享有 80% 的财富。之后也广泛指代一种不平衡关系，即少数的人或事情可以造成主要的、重大的影响。根据该理论，传统商家主要关注那 20% 能创造 80% 收益的顾客，而忽视其他 80% 的顾客。在 2004 年，一位杂志主编在形容亚马逊和 Netflix 的商业模式时，首次运用了“长尾”这一词，表示需求和销量不高的产品所占据的共同市场份额可以和主流产品的市场份额持平或更高。因此，当传统商业中的二八定律已经无法进一步创造更多利润时，一些商家们转而关注“长尾”的潜在价值，即关注小人群的差异化偏好，而推荐算法能有效的实现企业从二八定律转向“长尾效应”的过渡，从而挖掘那些具有不同特性的 80% 用户能够带来的商业价值。在这个过程中，企业能够在以较低成本了解这 80% 人群的需求同时为他们提供满足其需求的物品或服务，从而实现盈利的增加。与此同时，以往市场只关注大众需求，而如今 80% 人群的个性化需求也可以得到满足，提高了用户的满意程度和消费质量，实现了双赢。

## 2.3. 推荐系统的研究现状

推荐系统是学术界和工业界研究的热门话题。学术界侧重理论层面的分析和模型性能的提升，而工业界更侧重实践层面的发展以及用户体验的提升。以下将从学术界和工业界两个方面介绍推荐系统的研究现状。

### 2.3.1. 学术界

上文提到推荐系统的主要原理是利用推荐算法从用户的历史日志中挖掘用户偏好，将满

足用户兴趣偏好的物品或者服务推荐给用户。根据用户日志数据的输入形式和推荐算法的设计机制，可将现有的推荐算法划分为 3 类，分别是基于协同过滤的推荐、基于内容的推荐以及混合推荐。

**协同过滤推荐算法**是诞生最早，并且较为著名的推荐算法，在整个推荐系统发展史上有举足轻重的作用。基于协同过滤的推荐算法又分为基于用户的协同过滤算法和基于项的协同过滤算法。其中，基于用户的协同过滤算法主要考虑用户与用户之间的相似度，在用户群中找到与指定用户具有相似兴趣偏好的相似用户，将相似用户喜欢的物品或者服务推荐给指定用户。举个简单的例子，如图 8 (A) 所示，对于用户 Up 来说，根据用户 Up 的观看记录知道用户 Up 喜欢《钢铁侠》和《复仇者联盟》这类奇幻、冒险类电影，而用户 Jark 也喜欢看奇幻、冒险类电影，如《钢铁侠》、《复仇者联盟》和《美国队长》等。因此，可以通过某种方法了解到 Jark 和用户 Up 有相似的好，所以基于用户的协同过滤算法会把 Jark 喜欢而 Up 还未观看过的电影《美国队长》推荐给 Up。与基于用户的协同过滤推荐算法不同，基于项的协同过滤推荐算法从项（物品或者服务）的角度出发，重点考虑项与项之间的相似度。该类方法认为用户更倾向于喜欢曾经购买、收藏、点击、浏览过物品、内容或者服务，将与用户历史交互过的物品、内容或者服务相似的项推荐给用户。同样的，该类方法通过分析指定用户的历史交互记录，挖掘用户对项的历史偏好，找到与其相似的项推荐给用户。举例来说，如图 8 (B) 所示，基于项的推荐算法观测到用户 Up 观看过的《钢铁侠》和《复仇者联盟》等具有动作和科幻类型属性的电影，认为用户更喜欢该种类型的电影。因此，与具有爱情属性的电影《美女与野兽》相比，具有相同动作和科幻类型属性的电影《美国队长》更适合给用户做推荐。

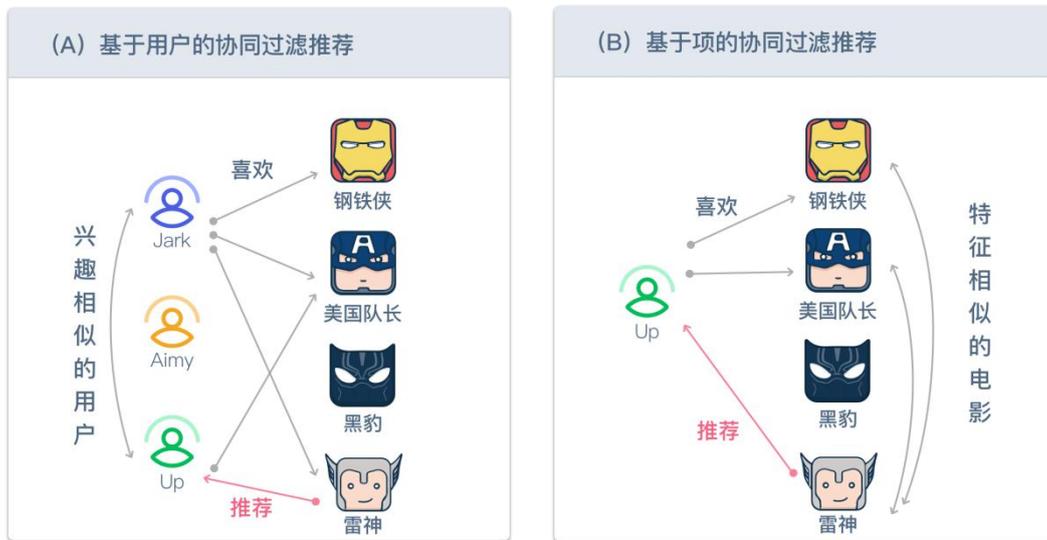


图 8 协同过滤推荐举例

在学术界，目前已有大量协同过滤的推荐算法被提出，并取得了较好的推荐性能。如图9所示，早期的协同推荐方法<sup>[1][2][3][4]</sup>一般采用矩阵分解的方式从用户对项目的评分矩阵中学习用户和项的潜在因子，利用学习的用户和项的潜在因子之间的相似性做推荐预测。但此类方法面临着冷启动问题和数据稀疏问题的困扰。比如，某物品、内容或者服务项没有得到用户的评分时，矩阵分解的推荐算法不会将该候选项推荐给用户；当有新的用户时，矩阵分解的推荐算法无法根据已有的评分矩阵学习新用户的潜在因子；当评分矩阵较为稀疏时，矩阵分解的推荐算法无法分解得到有效的用户和项的潜在因子。为解决上述问题，一些学者利用深度学习技术在捕获向量交互特征和复杂语义信息方面的优势，将深度学习技术和矩阵分解技术融合在一起学习用户和项的潜在因子。比如，较为经典的方法 Autoencoders<sup>[5]</sup>设计了协同过滤式的深度学习模型，利用边缘化去噪自动编码器和矩阵分解技术，将评分矩阵的分解和特征学习结合在一起，以学习有效的用户和项的潜在因子。CDAE<sup>[6]</sup>以更灵活的结构将几种先进的矩阵分解模型进行概括，提出协同去噪自动编码器框架形成用户和项目的反馈数据来学习用户和项的分布式表示，以解决 Top-N 推荐问题。DMF<sup>[7]</sup>构造了一个具有显式评分和非偏好隐式反馈的用户项矩阵。以该矩阵为输入，提出了一种深层结构的学习体系结构，用于学习用户和项分布式表示的低维向量空间。协同过滤作为一种经典的推荐算法种类，在工业界应用广泛，它的优点很多，模型通用性强，不需要太多对应数据领域的专业知识，工程实现简单，效果也不错。这些都是它流行的原因。当然，协同过滤也有些难以避免的难题，比如令人头疼的“冷启动”问题，我们没有新用户任何数据的时候，无法较好的为新用户推荐物品。同时也没有考虑情景的差异，比如根据用户所在的场景和用户当前的情绪做个性化的满足用户需求的推荐。

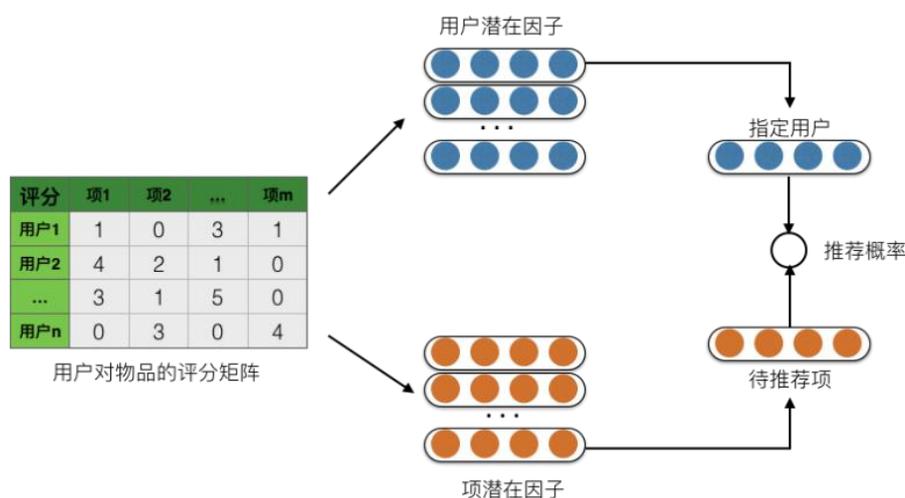


图9 协同过滤推荐算法图例

**基于内容的推荐算法**不同于协同过滤推荐方法，比较擅长根据用户的交互记录为用户提供个性化的推荐结果。基于内容的推荐算法的基本原理是根据用户和项的属性特征以及用户

的历史行为，获得用户的兴趣偏好，为用户推荐跟他的兴趣偏好相匹配的项。图 10 形象地给出了基于内容的推荐算法的例子：某一舞者用户曾经购买过舞鞋和裙子，推荐系统从用户购买过的舞鞋和裙子的文本内容（标题、文本描述、用户反馈等）中挖掘项的特征，将具有相似特征的舞鞋和裙子推荐给用户。基于内容的个性化推荐，一般需要三个步骤，分别是：

（1）通过深度学习技术挖掘用户信息及用户历史行为以构建用户特征表示；（2）根据项的文本内容信息（标题、文本描述、用户反馈等）构建项的特征表示；（3）使用用户和项特征表示之间的相似度匹配做为衡量用户接受被推荐项的概率。在基于内容的推荐算法中，用户的历史交互行为通常是一系列用户点击或浏览过的项序列。一般地，该类推荐方法在构建用户特征表示的步骤中会设计相应的项表示机制，通过集成用户一系列的历史交互项特征来挖掘用户特征，并基于此用户特征做个性化推荐。

早期的基于内容的推荐方法<sup>[8]</sup>一般采用统计策略从用户的历史行为中挖掘用户特征，比如 TF-IDF 词频-逆文档频率评估词对于用户历史交互文档的重要程度；余弦相似度计算内容间关于词的相关性。但此类方法在两个不相同的物品可能拥有同样的内容的情况下对内容分析不合理。后续一些方法利用机器学习和统计学习技术，如决策树，最近邻居、聚类神经网络等，来有效地解决了现有方法存在的问题。其中，较为典型的方法 DeepWide<sup>[9]</sup>、DeepFM<sup>[10]</sup>使用神经网络学习用户和项之间的高层交互特征来做推荐概率的预测。聚焦到目前主流的基于深度神经网络的推荐算法，利用深度神经网络对用户历史交互序列建模以挖掘用户的偏好特征，是获得个性化推荐的有效方法。该类方法一般采用卷积神经网络 CNN、循环神经网络 RNN、注意力网络或者记忆网络从用户历史序列中抽取局部特征或者序列特征来表示用户的分布式表示，并基于此用户特征表示计算与候选物品项的匹配概率。比如 DeepJoNN<sup>[11]</sup>、ATEM<sup>[12]</sup>、MANN<sup>[13]</sup>等。但这些方法由于假设任意相邻的项都紧密相关，容易捕获一些错误和冗余的依赖关系。事实上，在现实世界中，一些用户与项的交互序列是严格排序，而有些用户与项的排序可能不是严格排序的。比如在用户的购物序列{牛奶、黄油、面包}，先买牛奶还是先买黄油不重要，但是同时购买这两种商品后要买面粉的可能性更高。也就是说，牛奶和黄油两者之间没有严格的命令，但面粉顺序取决于牛奶和黄油的结合。为克服这些问题，一些方法利用神经网络从用户的历史交互序列中挖掘用户的长期兴趣和短期兴趣，融合用户的长期和短期兴趣做个性化推荐。这类方法称为以会话为主的推荐方法。通常情况下，用户的历史交互记录包含多个会话，不同的会话反应了用户不同时期的兴趣偏好。以会话为主的推荐方法将用户当前会话作为用户短期兴趣的体现，当前会话之前的历史会话作为用户长期兴趣的体现，通过从用户的短期兴趣和长期兴趣两个层面挖掘用户的特征偏好，基于用户的长短期特征计算与候选物品的匹配概率，以此概率作为向用户推荐候选项的可能性，即用户对该候选物品的满意度。

**混合推荐方法**侧重衡量各推荐方法的利弊，扬长避短，通过加权、切换、混杂、特征组合等方式避免或弥补各推荐技术的弱点，进一步提升推荐方法的性能。近年来，知识图谱由

于其结构化、可解释、可推理等特性被广泛应用到推荐系统、自然语言处理、问答系统、文本分类等领域。相应地，融合知识图谱的混合推荐方法受到了学术界和工业界的广泛关注，

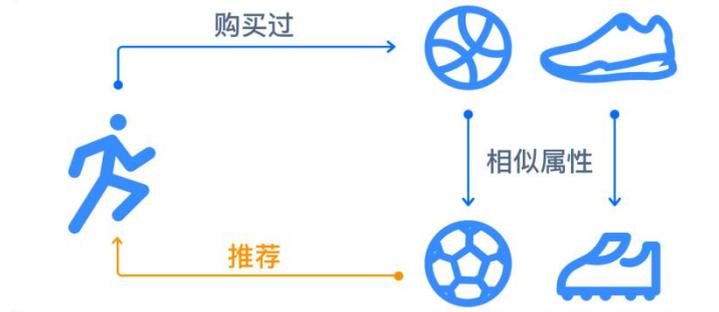


图 10 基于内容的推荐方法示意图

已然成为当前的研究热点。融合知识图谱的推荐方法利用知识表示学习技术将知识图谱蕴含的丰富语义信息融入推荐过程，使推荐结果更加精准地满足用户的需求。现有的基于知识图谱的推荐方法一般首先会根据推荐场景中的用户与项的交互记录和项的属性信息建立用户-项知识图谱。然后利用知识图谱的三元组表示学习技术、路径推理技术、强化学习技术以及图神经网络技术将用户与项之间购买关系的推荐任务建模为用户-项知识图谱的补全任务，从用户-项知识图谱中学习用户和项的分布式表示，预测三元组（用户，购买关系，项）的匹配概率，即向用户推荐该项的概率。比如较为经典的 KPRN 方法使用循环神经网络挖掘用户与物品之间的图谱路径，利用知识图谱的路径推理技术建模用户与候选推荐项的可解释性推荐。融合知识的推荐方法是当前学术界较为前沿的技术，是解决推荐算法面临黑盒问题的有效途径，对提高推荐结果的可解释性具有显著的提升。因此，利用知识图谱提升推荐结果的可解释性也是未来推荐系统较为前沿的研究方向。

总而言之，协同过滤推荐算法倾向于利用相似用户的偏好和相似的项为用户做推荐，基于文本内容的推荐算法从用户或项的文本内容中挖掘特征为用户做推荐，而基于知识的推荐算法利用知识图谱的可推理特性，在提高推荐性能的同时增加推荐结果的可解释性。这三类推荐算法从不同角度建模推荐过程，可以将这三类方法的优势相互组合，构建混合推荐方法以提高推荐性能。表 2 列举了上述三类方法的经典模型、各模型使用的核心技术以及评测指标等。

推荐场景	方法举例	核心技术	评测指标
协同过滤推荐方法	Autoencoders[5]、CADE[6]、DMF[7]、CKE[14]	矩阵分解、张量分解、神经网络	accuracy、MRR、ndcg@k、hit@k
基于文本内容推荐	TransRec[20]、DeepCoNN[21]、JRL[22]、DKN[23]、DAN[24]	卷积神经网络、循环神经网络、注意力机制	accuracy、MRR、ndcg@k、hit@k

混合推荐方法	KPRN[15]、RippleNet[16]、KGCN[17]、VGAE[18]、PGPR[19]	强化学习、图卷积神经网络、深度神经网络、注意力神经网络	accuracy、MRR、ndcg@k、hit@k
--------	---	-----------------------------	---------------------------

表 2 三类推荐算法举例

### 2.3.2. 产业界

基于协同过滤、内容的推荐算法是最早应用于工程实践的推荐算法。在工业届的不同领域已有大量的应用案例。根据应用领域的不同，本文将按照资讯、视频、电商和社交等场景下介绍产业界目前使用的推荐技术。

**推荐系统在资讯场景下的应用。**目前推荐算法已经成为各资讯类应用中的“标配”。当今较为流行的资讯类推荐应用包括今日头条、腾讯新闻、百度首页新闻推荐以及新浪看点等。各推荐应用具有类似的推荐思想，但在具体落地实施中会根据特有场景设计不同的推荐算法。以国内头部资讯平台今日头条的推荐算法为例，详细讲述其如何根据用户的历史行为信息为用户做个性化推荐。宏观上，今日头条的推荐技术的算法原理是通过资讯内容、用户特征以及环境特征三个维度拟合用户对推荐内容的满意程度。对于资讯内容，今日头条已建立了一个包括图文、问答、视频、UGC 小视频等多元化内容的综合性平台。根据多元化内容的相应特征，提取不同类型的内容特征做推荐；对于用户特征，包括年龄、性别、职业等显式特征和从大量用户数据中挖掘得到的隐式特征，如用户感兴趣的类别、主题、关键词及兴趣特征（国际新闻、旅游、汽车等）、身份特征等；对于环境特征，强调用户在不同场景中的信息偏好是有所差异的，例如在工作和旅行时，用户的兴趣偏好是完全不同的。今日头条通过多种方式，比如协同过滤推荐模型、深度学习模型、逻辑回归、分解模型和监督学习模型设计推荐算法以融合上述三个维度的特征，进而有效地预测在该场景下某用户是否喜欢相应的推荐内容。以上只是一个简单的推荐算法框架，实际的推荐系统需要灵活的适应各种情况，比如如何去除噪声的影响、优化无法直接衡量的目标、组合多种算法以调整模型结构等。今日头条及字节跳动旗下其他几款产品都在沿用同一套强大的算法推荐系统，但模型架构会根据业务场景的不同有所调整。为便于公众为推荐系统工作机制的充分理解，本文从输入特征、模型在实际场景中的训练与评价机制以及模型推荐内容质量等方面介绍今日头条推荐系统：

(1) **输入特征。**为了进一步细化推荐算法，为公众提供满足其兴趣的推荐内容，今日头条除了设计高质量推荐算法，还重点关注哪些推荐特征对推荐性能有影响。在今日头条的特征体系中，重点关注相关性特征、环境特征、内容热度特征和协同特征。其中，相关性特征就是评估内容的属性与用户兴趣是否匹配，包括关键词匹配、分类匹配、来源匹配、主题匹配等显式匹配，以及用户向量与内容向量的距离相似性的隐式匹配；环境特征包括地理位置、时间等因素，用于构建一些匹配特征；内容热度特征包括全局热度、分类热度，主题热

度，以及关键词热度等，可以有效缓解用户冷启动问题；协同特征是通过分析不同用户间的相似性，比如点击相似、兴趣词相似，用户标签相似甚至用户间的向量相似来扩展模型的探

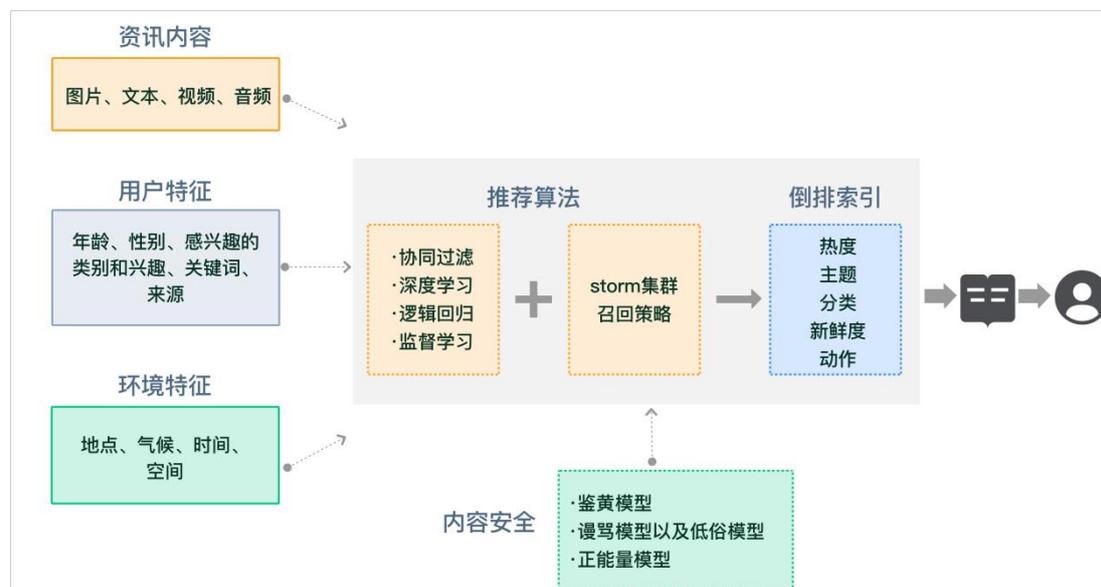


图 11 今日头条资讯推荐系统示意图

索能力。在某种程度上，协同特征在部分程度上可以帮助解决所谓的算法越推越窄的问题。在上述四种特征中，相关性特征和协同特征是今日头条推荐系统的两大基石，以下简要介绍两种特征的相关知识：

(a) 对于相关性特征，文本内容属性是计算用户对推荐资讯满意程度的重要因素。推荐系统中的文本分析就是对用户兴趣的建模，没有内容及文本标签，无法得到用户兴趣标签。比如，只有知道文章标签是互联网，用户看了互联网标签的文章，才能知道用户有互联网标签。今日头条主要抽取的文本特征包括由人打上的显式语义标签以及基于主题和关键词的隐式语义特征。显示的标签体系是预定义的且每个标签有明确的意义，隐式的文本语义特征是通过算法计算关键词概率分布，主题特征等获得的。今日头条推荐系统的线上标签体系分类采用典型的层次化文本分类算法。最上面根结点，下面第一层的分类是像科技、体育、财经、娱乐，体育这样的大类，再下面细分足球、篮球、乒乓球、网球、田径、游泳等，足球再细分国际足球、中国足球，中国足球又细分中甲、中超、国家队等。相比单独的分类器，利用层次化文本分类算法能更好地解决数据倾斜的问题。而隐式的语义文本特征是衡量文本相似性的重要因素，也是解决不重复推荐相同或类似内容的关键。在头条，曾经用户反馈最大的问题之一就是为什么总推荐重复的内容。这个问题的难点在于，每个人对重复的定义不一样。举个例子，有人觉得这篇讲皇马和巴萨的文章，昨天已经看过类似内容，今天还说这两个队那就是重复。但对于一个重度球迷而言，尤其是巴萨的球迷，恨不得所有报道都看一遍。解决这一问题需要根据判断相似文章的主题、行文、主体等内容，根据这些特征做线上策略。同样，还有时空特征，分析内容的发生地点以及时效性。比如武汉限行的事情推给北京用户

可能就没有意义。最后还要考虑质量相关特征，判断内容是否低俗，色情，是否是软文，鸡汤？目前，隐式语义特征已经可以很好的帮助推荐，而语义标签需要持续标注，新名词新概念不断出现，标注也要不断迭代。显式语义标签做好的难度和资源投入要远大于隐式语义特征，但语义标签是必要的，原因是有一些产品上的需要，比如频道需要有明确定义的分类内容和容易理解的文本标签体系。同时，语义标签的效果也是检查一个公司自然语言技术水平高低的试金石。

(b) 对于协同特征，主要通过分析不同用户间的相似性，比如点击相似、兴趣词相似，用户标签相似甚至用户间的向量相似来扩展模型的探索能力。其中，用户标签工程挑战较大。今日头条常用的用户标签包括用户感兴趣的类别和主题、关键词、来源、基于兴趣的用户聚类以及各种垂直兴趣特征（车型，体育球队，股票等），还有性别、年龄、地点等信息。这些用户标签非常有助于推荐。当然最简单的用户标签是浏览过的内容标签。但这里涉及到一些数据处理策略。主要包括：一、过滤噪声。通过停留时间短的点击，过滤标题党。二、热点惩罚。对用户在一些热门文章上的动作做降权处理。理论上，传播范围较大的内容，置信度会下降。三、时间衰减。用户兴趣会发生偏移，因此策略更偏向新的用户行为。因此，随着用户动作的增加，老的特征权重会随时间衰减，新动作贡献的特征权重会更大。四、惩罚展现。如果一篇推荐给用户的文章没有被点击，相关特征（类别，关键词，来源）权重会被惩罚。

(2) **模型在实际场景中的训练与评价机制**。在实际应用场景中，今日头条面临的资讯内容容量非常大，推荐系统不可能对所有内容进行预估。为给用户提供实时的资讯，一方面，今日头条采用 storm（开源的分布式实时大数据处理框架）集群实时处理样本数据，不仅节省资源，还拥有反馈快的优势，可以快速捕捉用户行为信息并反馈至下一刷的推荐效果。另一方面，今日头条设计了一些召回策略，每次推荐时从海量内容中筛选出千级别的内容库作为给用户推荐的候选池。实际应用中的召回策略要求极致的性能且召回时间不能超过 50 毫秒。今日头条主要采用倒排索引维护一个倒排，在后续的线上召回过程中可以迅速从倒排中根据用户兴趣标签对内容做截断，高效的从很大的内容库中筛选比较靠谱的一小部分内容推荐给用户。这个倒排的关键标签可以是分类、主题、实体、来源、热度、新鲜度、动作等。

为有效评估推荐效果的好坏，今日头条尝试综合尽可能多的指标合成唯一的评估指标以全面的评估推荐系统。事实上，很多因素都会影响推荐效果，比如候选集合变化，召回模块的改进或增加，推荐特征的增加，模型架构的改进，算法参数的优化等等。全面的评估推荐系统，需要完备的评估体系、强大的实验平台以及易用的经验分析工具。所谓完备的体系指不能完全由点击率、阅读时间、点赞、评论、转发等可以量化的数据指标衡量，需要引入数据指标之外的其他指标，比如对广告和特频内容的频控，对低俗内容、虚假新闻、标题党的治理，以及对重要内容的置顶、强插和加权等。过去几年，今日头条通过人工和模型的结合过滤低俗、负能量内容、建立了反低俗模型、正能量模型以建立综合的评估体系，旨在做到

既兼顾短期指标与长期指标，又兼顾用户指标、广告主利益和生态指标，

(3) **模型推荐内容质量**。在人工智能技术快速发展的今天，内容安全一直是公众关注的重点。观察发现，今日头条作为国内有代表性的内容创作与分发平台，越来越重视社会责任和行业领导者的责任。从创立之初，内容安全就一直存在于最高优先级队列中，并设有专门的审核团队负责内容安全，避免因 1% 推荐内容出现问题而造成较大影响事件的发生。现在，今日头条的内容主要来源于两部分，一是具有成熟内容生产能力的 PGC 平台，一是 UGC 用户内容，如问答、用户评论、微头条等。这两部分内容需要通过统一的审核机制。如果是数量相对少的 PGC 内容，会直接进行风险审核，没有问题会大范围推荐。UGC 内容需要经过内容识别技术的过滤，有问题的会进入二次风险审核。审核通过后，内容才会被真正进行推荐。这时如果收到一定量以上的评论或者举报负向反馈，还会再回到复审环节，有问题直接下架。整个机制相对而言比较健全，作为行业领先者，在内容安全上，今日头条一直用最高的标准要求自己。

在关于内容安全的技术方面，今日头条采用的内容识别技术主要有鉴黄模型、谩骂模型以及低俗模型。低俗模型通过深度学习算法训练非常大的样本库，图片和文本。这部分模型更注重召回率，准确率甚至可以牺牲一些。谩骂模型的样本库超过百万，给发布不当评论的用户设置一些惩罚机制。泛低质识别涉及的情况非常多，像虚假新闻、黑稿、题文不符、标题党、内容质量低等等，这部分内容由机器理解是非常难的，需要大量反馈信息，包括其他样本信息比对。实践中，不能仅通过低质模型的召回，还需要结合人工复审，将阈值进一步提高。目前最终的召回已达到 95%，这部分其实还有非常多的工作可以做。今日头条人工智能实验室李航老师在和密歇根大学共建科研项目，设立谣言识别平台。

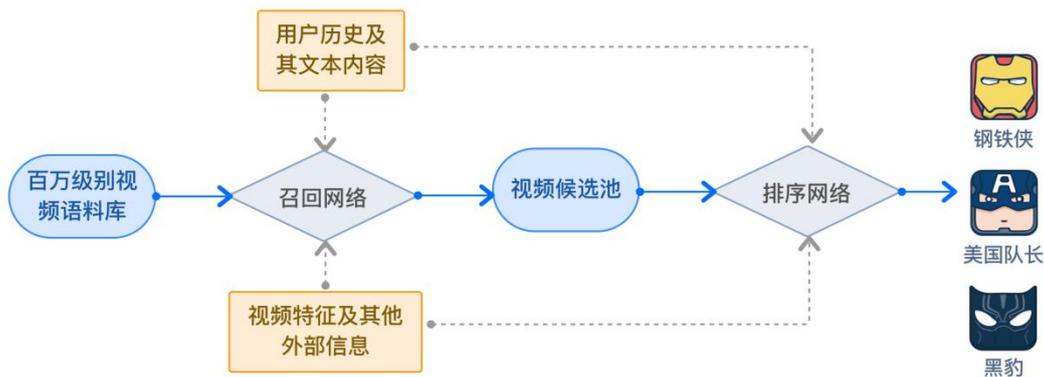


图 12 YouTube 推荐系统示意图

**推荐系统在视频、音乐场景下的应用**。随着人们对于娱乐生活的需求提高，视频网站不断发展，特别是 UGC 网站的发展让用户逐渐自产内容，其中发展迅速的有国外的 YouTube、国内的哔哩哔哩以及风靡全球的 TikTok 和抖音。作为全球最大的视频分享网站，YouTube 平台中几乎所有的视频都来自 UGC，它的视频基数高达十亿级别，为了让用户能够发现自

已感兴趣的事物，YouTube 在推荐系统上的探索很早就开始了，可谓经典。接下来就以 YouTube 的推荐算法为例进行简要分析。

整个系统可以分为三个部分，第一部分是召回网络：此阶段的目的是从百万级的视频中挑选小部分的视频用于之后的排序，这部分需要处理的数据量非常大，速度要求快。召回网络会根据用户的历史信息进行召回，这一阶段召回的视频满足用户泛化的兴趣，用户之间的相似度则通过粗略的特征来表示，如用户观看视频的 ID，搜索请求和用户特征。第二部分为排序网络：此阶段会使用更加丰富和详尽的用户和视频特征，并对于多目标进行优化，一方面要预测用户的满意程度，另一方面则是参与度指标，例如用户的观看时长、评论、转发等，最终加权输出总分数，然后根据分数进行排序，依次展示给用户。第三部分，进行线下评估，评估指标有准确度、召回率等，最终效果还是需要线上做 A/B 测试，考察点击率、观看时间等指标。

**推荐系统在电商场景下的应用。**如今推荐系统已经成为各大电商平台的主要流量入口，是做到比用户更懂用户的关键，也是掌握零售时代主动权的重要手段。国内外大型的电商平台如 Amazon、淘宝、京东、拼多多等都离不开推荐系统的应用。淘宝作为我国最大的电商平台之一，它的手机移动端的淘宝 APP 的推荐是典型代表。以手淘推荐为例，该推荐算法需要考虑的问题主要有以下几点：第一是购物决策周期，用户的决策周期需要经过发现需求、获取信息、商品对比和下单决策的过程，因此决策周期较长，针对不同购物状态，系统需要做出合适的系统推荐。具体来说，所有用户的购买行为都是由内部或外部的刺激引起的，比如外部刺激有产品的质量、款式、服务、广告等，内部因素有动机、个性、态度、观念、习惯等。用户在种种刺激因素的作用下，经由复杂的心理活动过程，产生购买动机，在动机的驱使下，做出购买决策，采取购买行动，并进行购买评价，由此完成一次完整的购买行为；第二是实效性，有些商品属于低频，有些属于紧急物品，因此系统需要快速地感知和捕获用户的实时兴趣和探索未知需求，在这一方面，阿里开发了 Blink 处理任意的流数据；第三是用户群体复杂性，对于未登录用户、新用户和活跃度不同的用户，需要制定差异化的推荐策略；第四是多场景，手淘推荐覆盖场景高达几百个，每个场景条件不同，模型的超参数也不同，而人工逐个优化效率较低，因此需要模型能够迁移学习。

一个经典推荐系统分为召回和排序两个部分。召回指根据用户的兴趣从海量的商品中去检索出相关候选商品，满足推荐相关性和多样性需求；排序则是依据用户兴趣会对候选集进行排序截取优先级最高的 k 个商品，最终给用户推荐。手机淘宝的召回技术是动态实时多兴趣表达(MIND)，采用胶囊网络的动态路由算法来获得用户多兴趣表示，将用户的历史行为聚集起来，每一组历史行为都用于推断对应特定兴趣的用户表示向量，输出多个表示向量以代表特定用户的不同兴趣，从而对用户的多兴趣以及兴趣的动态性进行刻画。在此过程中，系统将学习一个函数，可以通过用户的行为得到用户的兴趣，并基于此兴趣计算用户对候选集产生的购买行为的概率。在排序阶段，相比于传统基于打分做一个贪心排序，手淘推荐使

用了生成式排序模型，对于给定用户，从候选集中，选出若干个物品组成集合，将传统的 Top-K 推荐转化为 Exact-K 推荐，将排序优化问题转化为集合优化，以实现精准推荐。

在实际场景中的淘宝推荐举例，淘宝采用了图嵌入技术 (graph embedding)，对每个用户基于其历史数据构建一个有向图，将用户历史点击过的商品表示为图中的节点，在两个被连续访问的商品节点之间建立一条边，且边的权重表示从头部商品节点到尾部商品节点出现的次数，比如边 A ->B。然后通过随机游走 (random walk) 生成商品序列，在利用 skip-gram 算法生成商品的表示向量。通常在短时间内访问的商品之间的关联度较高，而物品之间不同的访问顺序可以获得不同的商品表示，能够从用户的历史购买行为中挖掘商品的相似性。但是图嵌入方法适用于稠密图，即对存在较多交互商品的用户友好，存在冷启动问题。为解决冷启动问题，即准确学习到与用户没有交互的的商品的特征表示，比如上新物品等，系统会给此类物品添加一些额外辅助信息 (side information)，包括类别、品牌、商店、价格等等。例如喜欢佳能相机的用户也可能喜欢佳能新出的一款滤镜。综上所述，商品的嵌入表示最终由商品本身的嵌入表示和其额外辅助特征表示两部分集成。

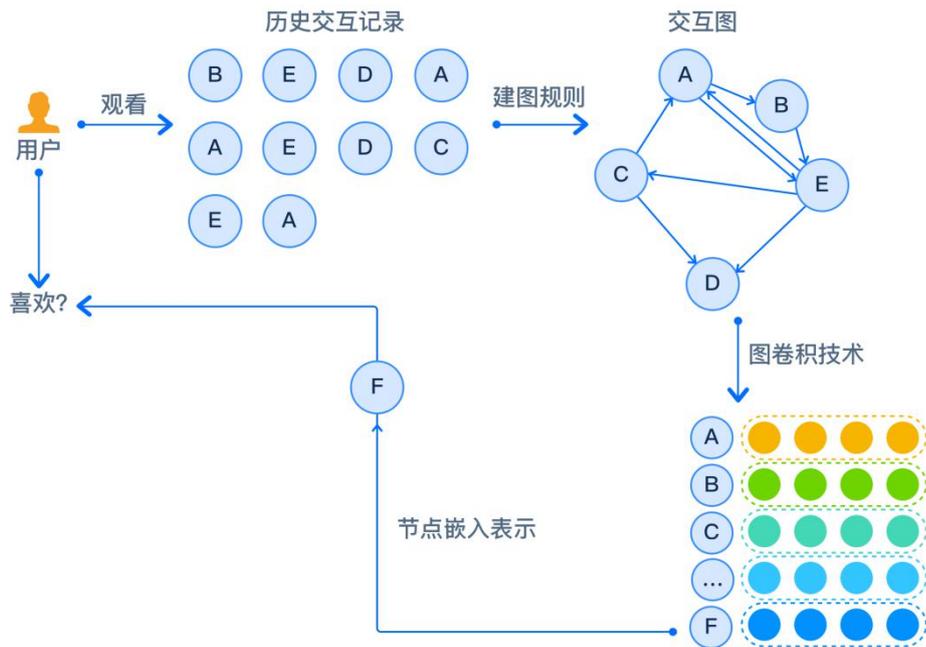


图 13 基于图嵌入技术的手淘推荐流程

**推荐系统在社交场景下的应用。**在如今的互联网时代，线上社交在人们的生活中占据了十分重要的位置。随着技术的发展，许多企业也看到了社交类应用在推荐中的优势——社交关系数据的社交同质性和社交影响力，即好友之间兴趣的相似性和好友对用户的影响力的价值。以微信为例，微信在精准推荐中采取的算法是 RALM，即基于 look-alike 模型的实时推荐算法。广告主会给系统提交一批用户名单，该用户名单是广告主的目标人群，称为种子用户，然后根据历史数据挑选部分对该类广告不感兴趣的用户作为负样本，使用某种计算用户

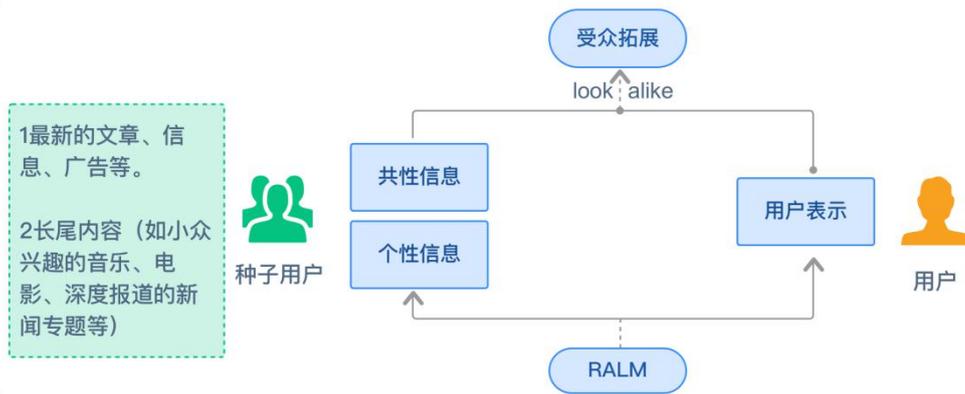


图 14 微信精准推荐算法流程图

相似度的方法，找到和种子用户最相似并且和负样本相差最大的目标用户，随后把广告推给目标用户。在此过程中，对于实时更新的文章、广告等，RALM 会实时对种子用户进行拓展，学习准确性和多样性的用户表达方式。

微信精准推荐中的推荐算法第一阶段会对用户的表示进行学习，得到兼顾所有用户的多样性和准确性的向量嵌入表达，第二阶段是根据上述兼顾所有用户的共性嵌入表达和用户个性信息组合形成对种子用户整体的表达。其中局部信息使用注意力机制提取和目标用户相关的部分，而全局特征使用全局注意力机制实现，其所做的就是捕捉用户群体自身内部的兴趣分布。为了减少线上开销，腾讯在该部分采取了 k-means 聚类的方法。就整个推荐流程来讲，系统大致分为三模块，第一部分是底下的离线训练，学习表示和 look-alike；第二部分在线异步处理，定时更新减约线上计算；最终是在线服务，根据线上请求以及前两部分准备的内容计算相似度。

## 第三章 推荐领域面临的问题

### 3.1. 信息茧房

“信息茧房”概念是由美国学者凯斯·桑斯坦在 2006 年出版的著作《信息乌托邦》<sup>[36]</sup>中提出的，指主动或被动地关注自己感兴趣的信息，久而久之形成信息的壁垒。具体来说，用户一旦形成阅读习惯，就会被自己的兴趣限定于特定领域，对其他问题和知识缺乏关注，使人们的注意力及兴趣在较小的范围内。如图 14 所示，比如用户 A 初期喜欢观看体育、历史、娱乐和音乐四个类型的新闻，推荐系统根据用户的观看记录将四类新闻推荐给用户，用户根据当下的兴趣偏好选择性的观看了体育和历史类的新闻。接着，推荐系统根据用户的浏览记录捕捉到用户近期浏览的体育和历史类的新闻较多，推测用户可能比较喜欢这两类的新闻，将推荐结果中的体育和历史类的新闻数量占比提高以满足用户的兴趣偏好。用户在此推荐结果下根据自己的喜欢继续浏览。久而久之，用户 A 获取的信息大多是跟体育相关的，很难获取音乐或者娱乐等其它相关的资讯。随着时间的推移，以用户 A 为代表的人群对外界的信息接收范围变得越来越小。

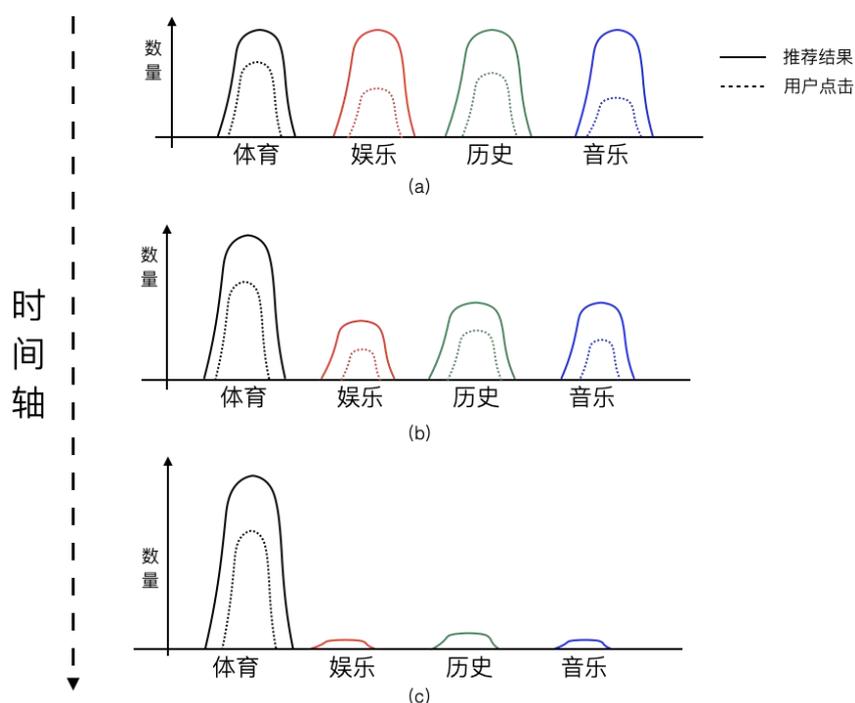


图 15 信息茧房图例。随着时间推移，用户兴趣范围由体育、娱乐、历史、音乐逐渐限定到体育领域。

信息茧房从来都有，并非新现象。从传统媒体时代到如今的智能传播时代，一直存在于我们的身边。如图 15 所示，在大众传媒时代，尽管传媒的内容和形式多种多样，用户依然选择看自己感兴趣的刊物、栏目或者频道等。进入门户网站时代，用户也只关注自己感兴趣的频

道，比如音乐、体育等。只不过进入智能传播时代，随着科学技术的发展，推荐算法为信息茧房提供了新的形成机制。在信息爆炸的今天，推荐系统作为一种信息过滤系统，根据用户的属性和历史行为记录，学习出用户的兴趣爱好，预测用户对给定事物、内容、物品或者服务等喜爱程度而进行相关的推荐。根据兴趣爱好提供智能化、个性化、精准化的推送是如今普遍的媒介传播机制，可能使一部分人陷入“信息茧房”效应的主要原因。信息茧房对人们的日常生活和文化行为产生深远影响，弱化“信息茧房”效应可以让用户克服“信息茧房”带来的局限性，尝试接收来自平台其他的信息，从“小圈子”中走出“信息茧房”效应带来的困境，拓展信息范围，使算法能够更好地服务用户。

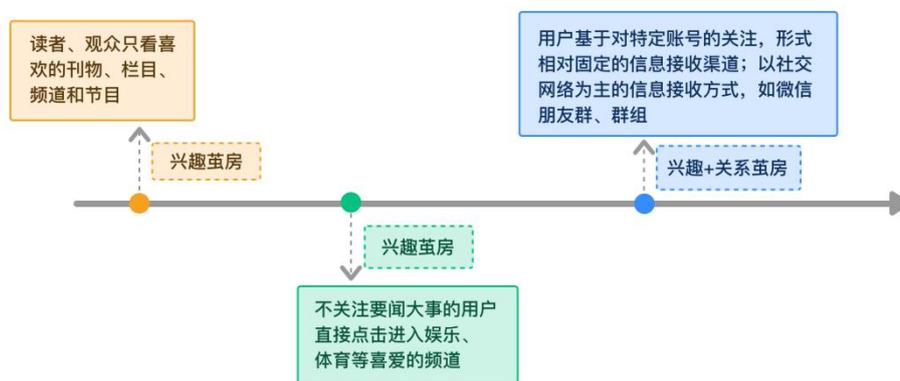


图 16 信息茧房在不同时期的表现举例

### 3.2. 算法黑盒

在人工智能时代，推荐系统的技术发展突飞猛进，以深度学习、强化学习或复杂的集成模型等为代表的先进技术方案给推荐系统的提升带来了显著收益。这类模型通常具有很高的准确性。然而，随着深度学习等方案的应用，模型复杂度越来越高，整个推荐系统越来越变成一种黑盒，以至于这些模型的内部工作机制难以理解，无法估计每个特征对模型预测结果的重要性，更不能理解不同特征之间的相互作用关系。试想，如果推荐系统不仅仅是建模用户兴趣特征以提供个性化推荐结果，还能将用户兴趣白盒化地给出一些人能理解的解释，比如给出较为简单的文字标签解释推荐结果，而不是给出一串无法理解的数字。如果推荐系统能够解释为什么会有相应的推荐结果，那么推荐系统就可以实时地去预估用户未来的兴趣，还能指导内容生产者去生产更多满足用户需要的内容，将整个系统的推荐价值推到更高。同时，推荐系统推荐某个物品、内容或者服务的逻辑可以显式的透传给用户，让用户给出对这个推荐逻辑的反馈，而不是仅仅给出对推荐结果的反馈，这样推荐系统能更接近感知用户真实的兴趣，也会变得更精准。

可解释性推荐系统正是为了应对推荐系统黑盒问题而提出的。无论模型如何复杂，可解释性推荐模型都能通过旁路系统的方式对原有的复杂黑盒推荐系统进行拆解，来解释推荐的

原因。可解释的推荐系统能够更好地帮助我们理解整套复杂的推荐系统，并且能为偶尔出现的 bad case 进行合理解释和改进。模型的可解释性是大多数机器学习系统都需要的一种能力，也是推荐系统在人工智能时代研究的主要趋势。

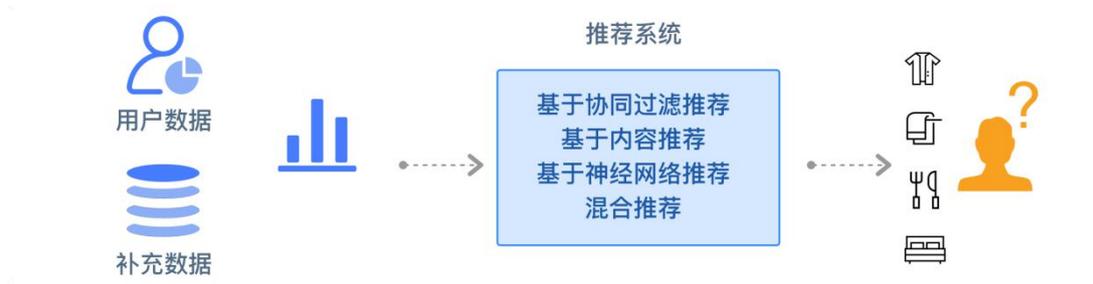


图 17 算法黑盒

### 3.3. 价值观及伦理问题

在报纸作为主要信息传播模式的时代，信息传播是中心化的，大众能看到什么信息，取决于媒体的编辑。而在智能手机普及的今天，大众获得信息的来源已经从报纸、电视转移到了手机上。目前市场上一些知名的信息类软件，比如今日头条、腾讯新闻、新浪看点、北京时间、一点咨询，UC 等，会采集用户点击的新闻类型，浏览的时长，甚至是在某一个广告图片上停留的时间等各种数据，通过这些数据，使用个性化算法来为用户推荐可能感兴趣的个性化资讯。

如果用户将自己的时间和阅读渠道都全部交给了一个应用，那么这个应用应当担负起一些信息传播的伦理责任吗？下面给出一些非常有趣的例子。如果用户持续地在阅读一些误导性很强的文章，比如“读书无用论”，“各种偏方秘方”等，此时的推荐算法实际上会给用户推荐更多这样的内容，从而将用户带入一个更偏执的信息环境中。这里面一个核心的伦理问题就是“用户长期存在于这样的环境中而形成的错误价值取向，是否需要可能对造成的社会危害负责”。Luck Dormehl 在《算法时代》<sup>[37]</sup>一书中也讲到一个例子，美国政府根据姓名、出生地、宗教信仰、人脸识别算法、历史行为数据等，会对每一位航空旅客是恐怖分子的嫌疑度进行打分，一些无辜的人因为疑似恐怖分子，而经常在机场被羁留检查，甚至多次错过飞机。这里面一个核心的伦理问题就是“我们是否应该为尚未发生的一种可能性付出代价”。面对上述类似的伦理问题，人们不得不思考推荐系统在其中到底扮演着什么角色？需要承担什么责任？是否需要推荐算法进行规制以规避上述问题的产生？因此，本小节就推荐系统目前面临的典型且公众较为关注的价值观和伦理问题做如下总结。

#### 3.3.1. 虚假新闻与低俗内容

虚假新闻和低俗内容是推荐算法面临的客观现象。从算法设计的初衷看，其本身并不会提倡标题党、煽情和低俗化内容。智能推荐算法从海量数据中挖掘特征来训练模型，为用户呈现推荐结果。客观上，造成虚假新闻和低俗内容泛滥的原因主要如下：（1）在互联网的海量数据中往往已经存在大量低质化信息，推荐算法直接在包含噪声的数据基础上进行建模，在大众审美品位没有得到根本提升的前提下，以及好奇、猎奇心理的驱使下，低俗内容和虚假新闻会广泛阅读、浏览和传播。为满足用户的兴趣偏好，推荐系统可能将点击和浏览量高的虚假和低俗内容继续推荐给用户；（2）推荐算法的设计和推荐过程都有人为参与，设计推荐算法的工程师未接受专业化新闻伦理训练的情况下，在算法发展的初期主要关注内容和用户兴趣的匹配度等指标，没有将对社会价值的导向要求需要内化为算法的具体规则，在客观上呈现出推荐内容的低质化倾向。（3）不同于人工编辑和人工筛选，纯粹的个性化推荐系统缺乏对文章的质量与内容的把关，终极目标就是实现流量最大化。因此，用户在猎奇心理驱使下点开的虚假新闻、低俗内容就会在网络环境中广泛地传播，如早孕网红、暴力血腥等内容严重破坏了网络生态环境。

推荐算法并不必然助长内容生态的低质化。这要从算法设计的主要特征分析，分为个体、群体、整体三个层次的特征：对个体用户，算法一般通过对内容特征、人的特征、环境特征三个维度指标的分析，在特定人和特定内容之间做出力求精准的匹配。内容特征可能包括领域分类、主题词、实体词、来源、质量评分、相似文章等指标，人的特征包括兴趣、年龄、性别、职业、使用行为、机型等指标，环境特征包括时间、地点、天气和网络类型等。在群体层面，算法通过寻找不同用户在兴趣分类、主题、实体词和使用行为上的相似性，将一个用户感兴趣的内容推荐给另一个人，这已不是基于用户自己的历史行为，而是基于群体隐性关联之上的协同推荐。就网民整体，算法则基于内容的热度特征，包括全平台的热点文章或不同类别、主题和关键词的热点内容，在“冷启动”阶段对新用户进行初步推荐。

随着推荐系统的流行，信息分发逐渐从编辑中心转向机器分发，在信息环境的活力和丰富程度获得加强的同时，信息环境的不可控性亦因此加强。出于内容生态和社会责任的考量，像低俗内容的打压，标题党、低质内容的打压，重要新闻的置顶、加权、强插，低级别账号内容降权都是需要信息分发系统进行干预，使得信息成为历史的精华，而不是一种被操纵和营销的产品。

### 3.3.2. 算法偏见和歧视

智能算法推荐是否会加剧全球不平等现象与社会冲突？《自然》杂志曾用 BIBO 表示 Bias In, Bias Out（即偏见进，偏见出）。可见，偏见从人类社会移植而来，隐秘地隐藏在算法之中。可能人们会说，算法是一种数学表达，是很客观的，不像人类那样有各种偏见、情绪，容易受外部因素影响，怎么会产生歧视呢？一些生活中的现象说明算法歧视确实存在。比如，一些图像识别软件之前还将黑人错误地标记为“黑猩猩”或者“猿猴”；微软公司在

美国 Twitter 上的线上聊天机器人 Tay 在与网民互动过程中，成为了一个集性别歧视、种族歧视等于一身的“不良少女”；在亚马逊公司曾开发的一款用于筛选简历的算法中，意外被“培养”出歧视女性应聘者的偏好。显然，算法从受相同偏见影响的内容中学习更多偏见，这些认知偏见经网上广泛传播后导致更为广泛的社会冲突。更进一步，如果一个 AI 系统依赖多数学习，自然不能兼容少数族裔的利益。比如，有一小部分用户，他们的喜好很特别，但是现有算法无法有效地将这一小部分用户和其他大部分用户区分开。那么这些用户的数据无法很好地影响算法的训练，导致模型最终决定给一小部分用户推荐的结果和大部分用户的推荐结果一样。由于推荐系统的优化目标是整体的准确率或者收益，这一小部分用户给出的负向反馈会被忽略掉，慢慢的这一小部分用户和平台的交互可能变得越来越少，最终被推荐系统成功地挤走。而这一切发生得是很隐蔽的，可能系统都没意识到自己丢失了这部分价值。从道德上讲，这样对少数派的忽略是不公平的。从商业价值上讲，可能短期看上去仅仅失去占比很少的少数派代表的价值，然而如果推荐系统这样循环下去，其实是在源源不断的失去一小部分价值，累积起来是非常大的损失。随着算法决策越来越多，类似的歧视也会越来越多。而且，算法歧视会带来危害。一方面，如果将算法应用在犯罪评估、信用贷款、雇佣评估等关切人身利益的场合，一旦产生歧视，必然危害个人权益。另一方面，深度学习是一个典型的“黑箱”算法，连设计者可能都不知道算法如何决策，要在系统中发现有没有存在歧视和歧视根源，在技术上是比较困难的。

那么，为什么算法可能暗藏歧视？算法决策在很多时候其实就是一种预测，用过去的的数据预测未来的趋势。算法模型和数据输入决定着预测的结果。因此，这两个要素也就成为算法歧视的主要来源。一方面，算法在本质上是“以数学方式或者计算机代码表达的意见”，包括其设计、目的、成功标准、数据使用等等都是设计者、开发者的主观选择，设计者和开发者可能将自己所怀抱的偏见嵌入算法系统。另一方面，数据的有效性、准确性，也会影响整个算法决策和预测的准确性。比如，数据是社会现实的反映，训练数据本身可能是歧视性的，用这样的数据训练出来的 AI 系统自然也会带上歧视的影子；再比如，数据可能是不正确、不完整或者过时的，带来所谓的“垃圾进，垃圾出”的现象；此外，算法歧视可能是具有自我学习和适应能力的算法在交互过程中学习得到的，AI 系统在与现实世界交互过程中，可能没法区别什么是歧视，什么不是歧视。

更进一步，算法有可能将歧视固化或者放大，使歧视自我长存于整个算法里面。算法决策是在用过去预测未来，而过去的歧视可能会在算法中得到巩固并在未来得到加强，因为错误的输入形成的错误输出作为反馈，进一步加深了错误。最终，算法决策不仅仅会将过去的歧视做法代码化，而且会创造自己的现实，形成一个“自我实现的歧视性反馈循环”。包括预测性警务、犯罪风险评估、信用评估等都存在类似问题。相比于人的能动性，算法决策其实缺乏对未来的想象力，而人类社会的进步需要这样的想象力。

### 3.3.3. 算法公平性

公平，这个词大家经常听到，存在于日常生活的方方面面。比如，找工作时，有的公司不看能力只看学历，有的单位不愿招女员工而倾向于招男员工等等。类比上述性别偏见、种族歧视等不公平问题，推荐系统中同样存在这样的不公平问题。那么，公平性之于推荐系统又是什么？如同没有绝对的正义，也没有绝对的公平。所谓的公平，只是站在不同的角度定义而已。比如招聘网站上，求职者方的公平可能定义为：相同的学历和能力可以被推荐类似的工作，而不受国籍、种族和性别等偏见；招聘方的公平可能定义为：系统可以将他们的招聘信息推给优质的求职者，而不会因为公司的背景等因素，只将招聘信息推给不太合适的求职者。在电商场景下亦是如此，购买者的公平可能定义为：商品推荐应该多样化，而不受物品流行度、物品曝光量、价格、厂商等的影响；物品的公平可能定义为：物品应具有相同的曝光率，而不应受物品流行度等影响。因此，从公平性关注的角度出发，将公平性分为2个方面：一是从用户角度；二是从物品角度。

用户偏好放大是针对用户方（user）的公平性问题，体现为用户搜索了一样东西后，接下来一段时间的推荐都是关于这个物品的，哪怕你已经购买或点击了这个物品。比如某用户购买了3条裙子和一个鞋子，由于推荐系统放大了用户对裙子的偏好，推荐系统很大可能只给用户推荐裙子，从而忽略了用户对鞋子或其他物品的喜爱。推荐系统的这种偏好放大现象将逐渐缩小用户的兴趣范围，向用户推荐的东西越来越单一，丧失了多元化，让用户看不到本来想要或者未来会感兴趣的東西。那么，为什么会出現用户偏好放大的现象呢？从数据层面看，推荐系统建立在类别分类不平衡的输入数据上，在没有其他可用信息的情况下，推荐系统将结果全部预测为最多的那一类就可以得到最高的准确率。从推荐模型层面看，推荐系统采样正样本时倾向于采用数量占比较多的类别，在训练过程中将这种偏差带入模型中，并在推荐结果中放大这种偏差。

流行度偏差通常是针对物品方（items）的公平性问题，体现为推荐系统更多偏爱推荐较为流行的物品，而不流行的物品有较少的机会被推荐或展示。长此以往，流行的物品变得越来越流行，不流行的却变得越来越不流行，这就是推荐系统中的马太效应。首先是对不流行物品的不公平，不流行物品可以类比为新开的店铺，所卖物品也许质量很好，却迟迟得不到推荐系统的推荐，让店家不得不选择别的平台。其次是对推荐平台也是不利的，提高不流行物品的曝光率，一方面可以为用户得到更多的选择，提高推荐列表的多样性，让推荐系统更好地实现个性化，另一方面可以吸引更多新的店家，扩大平台规模。反之，推荐结果如果包含大量的流行物品，结果趋同，用户得不到好的个性化体验，而且这些流行物品可能并不需要推荐系统就能被用户发现，其实也是一种资源的浪费。

### 3.4. 用户隐私与安全问题

用户的隐私和成安全问题是推荐系统面临的一对主要矛盾。一方面由于数据稀疏，需要获取尽可能多的数据。相反，在拿到越来越多用户数据后，用户数据的隐私和安全问题就会显

现。研究表明<sup>[38]</sup>，70%至89.5%的互联网用户认为个人隐私信息面临泄露风险；利用项目评分和用户相似度矩阵来进行推演攻击的可行性，这导致了用户对个人隐私信息的普遍担忧。通常，用户的历史信息越详尽，推荐结果则越精准。然而，对用户数据的深度分析与挖掘会对用户的隐私造成严重的威胁。因此，如何做好缓解数据稀疏与保护用户隐私安全之间的平衡是一个很严峻的研究问题，也使得在保护用户隐私的前提下实现准确的推荐成为目前推荐系统领域的一个研究热点。

很多 AI 系统，包括深度学习技术，大多都是依赖大量的数据来训练学习算法。数据已经成了 AI 时代的“新石油”，这带来新的隐私忧虑。一方面，如果在深度学习过程中使用大量的敏感数据，这些数据可能会在后续被披露出去，对个人的隐私会产生影响。所以 AI 研究人员已经在提倡如何在深度学习过程中保护个人隐私。另一方面，考虑到各种服务之间大量交易数据，数据流动不断频繁，数据成为新的流通物，可能削弱个人对其个人数据的控制和管理。当然，现在已经有一些可以利用的工具来在 AI 时代加强隐私保护，诸如经规划的隐私、默认的隐私、个人数据管理工具、匿名化、假名化、差别化隐私、决策矩阵等等都是在不断发展和完善的一些标准，值得在深度学习和 AI 产品设计中提倡。

## 第四章 推荐领域研究趋势

### 4.1. 信息茧房

从某种层面上讲，信息茧房效应是推荐信息的不平衡性造成的。信息茧房效应导致用户长期只能浏览限制领域的信息，从而失去了对事实的全面的认识和客观的评价，造成“真相稀缺”。信息茧房问题在工业界相关的推荐产品中体现为向用户重复推荐相似的物品，影响用户体验度，从而降低对平台的满意度。在工业界，主要从召回、精排和重排三个阶段缓解信息茧房问题。聚焦到在学术界，破除信息茧房问题的主要手段是对推荐结果进行多样化，通过对用户兴趣多样性的探索，将多样化的结果推荐给用户。较为典型的方法 MMR<sup>[25]</sup>提出最大边际相似性规则，在每次迭代中，将未选中的项目与已有的推荐列表中的项目的相关性的线性组合计算一个得分，将得分最高的项目添加到推荐列表中。这个得分表示未选中的项目与每个已经选择项目之间的差异值，得分越大，差异值越小，相反，得分越小差异值越大。IDRAR<sup>[26]</sup>利用马尔可夫随机游走，在推荐算法中设计探索多样性。DPP<sup>[27]</sup>是一种优雅的模式，引入了以更好地评估多样化算法的多样性指标。DPP 提出了一种加速算法来大大加快推荐结果推断的过程，使其适用于大规模场景。同时 DPP 模型中的可调参数允许用户平稳地控制多样性水平。IARD<sup>[28]</sup>利用商品流行度对推荐结果进行重排。COUSIN<sup>[29]</sup>构建用户和物品的相似性矩阵，采用回归模型来量化用户和物品之间的关联强度。T-RD<sup>[30]</sup>结合用户兴趣和用户之间的信任度来提高推荐系统的精度和增强推荐结果的多样化。

为规避信息茧房的影响，国内外各大互联网公司和研究机构积极响应，参与到弱化信息茧房的研究中来。举例来说，《纽约时报》作为最早的推荐系统，通过不断改进推荐算法以更好地推断用户兴趣的完整性和多样性，从而更精准地推送给用户相应的新闻。早期的《纽约时报》基于每篇文章的主题、作者、相关关键词标签等来推荐内容。随后，《纽约时报》通过协同过滤技术基于读者历史阅读的内容相似性来推荐。为提高推荐算法的公平性和推荐内容的多样性，《纽约时报》将其算法优化为“协作主题建模”（collaborative topic modeling, CTM），使得推荐系统能够推荐新发布、未探索过的文章。具体过程为在：（1）对每篇文章的内容建模；（2）对读者偏好内容项进行建模；（3）结合读者的偏好和内容之间的相似性作出推荐。这种算法跨越了基于内容的推荐算法和协同过滤的算法，解决了前两种算法的问题。《纽约时报》在 2018 年又推出了“您的每周版”（Your Weekly Edition），使用编辑和算法的混合方法推送新闻时事简讯。另外，《华尔街日报》于 2016 年创设了一个“红推送，蓝推送”（Red Feed, Blue Feed），将 Facebook 上同类内容的自由倾向、保守倾向的信息并列呈现给用户，以此提醒用户其偏向性，并推荐另一观点相左的内容，帮助用户平衡、多元化其新闻消费。《卫报》在 2016 年新设专栏“戳破你的泡泡”（Burst your bubble），每周选取 5 篇值得一读的保守派文章，拓宽读者视野。Facebook 从个人主题列表中删除了个性化内容，使相关文章功能由从前用户阅读及分享文章后继续发布相似新闻报道到根据同一

主题的不同角度发布文章。新闻聚合平台 BuzzFeed 为了缓解平台制造的“过滤气泡”，在新闻板块引入了名为“泡泡之外” (Outside Your Bubble) 的新功能，将发表在 Twitter、Facebook、Reddit 等平台上的观点呈现出来，试图让用户了解自身社交媒体空间之外发生的事情及看法。《瑞典日报》(NZZ) 在谷歌数字新闻倡议下，研发了“陪伴” (Companion) App，通过机器学习生成算法，确保推送的一系列新闻中包含一个“惊喜”，既能将阅读习惯和用户偏好考虑在内，同时也不会丢失任何重要内容。《金融时报》构建“知识构建者”系统，并针对新闻报道设计两个评分标准：一是该新闻报道在多大程度上代表了《金融时报》的整体报道；二是如果该新闻涉及本报既有的新闻观点和主题，它的得分会更高。这些标准一方面可以将同一个主题的各种报道推荐给读者，增加用户接触信息的多元化；另一方面如果一篇报道包含了用户之前阅读过的大量内容，那么该文章的评分就会较低，因此该系统可以让用户接触更多之前从未涉及的新颖性新闻。观察发现，今日头条是国内较早运用算法推荐的客户端，会去全面地分析用户的兴趣爱好，并且能够随着用户的持续使用，越来越懂用户的多样化需求。

从客观上来说，个性化推荐就像漏斗一样，会将推荐内容与用户相匹配，倾向于向用户推荐高度符合其偏好的内容，致使推荐的内容越来越窄化，使用户停留在推荐系统构造的信息茧房中。在外界的印象里，互联网应用为了追求精准推荐、吸引用户，客观上会缩小用户的内容或商品消费范围。但在行业实践中，与外界固有认知相反，互联网应用、特别是位于头部的大型平台有追求算法多样性的内在动力。对行业内代表性应用的数据分析表明，阅读内容的类型数量是否够多、所阅读内容类型的分散程度是否够高，与用户是否能长期留存关联密切，呈正相关。上述两项指标对用户长期留存的作用，可以与信息的展现总量、用户的停留时长、用户阅读量等指标的影响相媲美。头部平台正是认识到了这一点，在实践中积极探索促进算法多样性的做法。例如，通过多样性提升、内容品类打散等方式，对提升用户的留存周期、次日留存率、多日留存率、人均活跃天数以及停留时长都有较好的促进作用。以字节跳动的两款软件抖音短视频和今日头条为例，分析企业如何打破信息茧房的壁垒，为用户提供多样化的推荐结果。在抖音 APP 中，除了在浏览界面的推荐窗口中为用户提供其感兴趣的短视频，还在主界面提供同城窗口和关注窗口，为用户提供从信息茧房中走出去的机会。同时抖音是一个具有温度的 App，如果用户在凌晨还在刷抖音，会有“刷很久抖音了吧，需要休息啦”的提醒用户早点休息的防沉迷短视频。为解决信息茧房问题，今日头条建立了一个通用信息平台以提供足够丰富的内容题材，足够多样的分发方式，让人们看到更大的世界。围绕内容体裁和分发方式两个维度，今日头条已建成标准意义上的通用信息平台，既涵盖图文、视频、问答、微头条、专栏、小说、直播和音频等内容体裁，也囊括内容运营、推荐引擎、关注订阅和搜索引擎等多种分发方式。聚焦推荐算法，今日头条使用非常流行的 LR 和 DNN 技术建立了一套强大的算法推荐系统，根据业务场景的不同调整模型架构。通过对内容（图文、视频、问答、微头条）、用户特征（各种兴趣标签，职业、年龄、性别）和

环境特征（工作场合、通勤、旅游地）三个维度建模，推荐模型会给出一个预估，即推测推荐内容在这一场景下对这一用户是否合适。一方面，今日头条 App 采用树状、较为细化的内容分类机制，在推荐界面提供多元兼容和多观点并存的信息资讯。另一方面，今日头条也推出“千人万元计划”，“创作者收益计划”等，继续加大优质内容扶持力度，以提高推荐内容的质量。从非算法的角度来讲，今日头条的召回级别是千万级，候选池较大，且内容更加多元。从算法角度看，今日头条采用以下几种方式破除信息茧房：（1）提供热榜模块：用户可以从热榜模块中知道周围世界共同关注的事情，比如国家大事、社会新闻等，使用户不仅仅局限于自己的个人兴趣；（2）对理性作者的支持：现实世界中的许多事情不是非黑即白的，针对同一件事情的不同观点都有其各自的理由，加大对理性观点作者的扶持，增加其发布内容的曝光流量，在让更多的用户能观看到其新闻文章的同时，也能使用户接触到更多不同的观点、以新视角从不同层面了解和认知世界，避免陷入极端情绪之中；（3）优质内容扶持：今日头条平台会根据用户的历史记录评分，奖励提供优质内容的用户，惩罚提供低俗内容的用户，将低俗内容过滤、打散；（4）长尾推荐：今日头条平台会根据用户的历史长尾搜索记录进行推荐，不仅仅局限于用户近期的兴趣爱好，同时从用户的长期兴趣的角度进行多元化推荐。除了在技术层面采取上述措施规避信息茧房问题，今日头条所在公司字节跳动在管理中也积极制定相应措施，重点关注多目标综合。虽然信息茧房在短期内可以提高用户的平均使用时长，实现短期流量增长，但是平均使用时长的延长可能会驱赶体验差的用户。因此，长期来看，仅提高用户平均使用时长是不可取的。字节跳动采用的多目标综合机制侧重发展多元化且有潜力的产品，重视长期利益，看中产品未来的发展潜力。字节跳动将多目标综合机制渗透到绩效考核中，从整体上寻求大系统的平衡，从而引导产品开发者 and 设计者对产品内容多元化发展的关注。因此，从字节跳动自身的目标以及产品的定位来说，其存在巨大的内驱力解决信息茧房问题。

**规避信息茧房的措施。**信息茧房问题会使人们关注的信息领域习惯性的被自己的兴趣所引导，从而将自己的生活和看到的世界禁锢在茧房一样的空间。面对推荐算法带来的信息茧房问题，未来的推荐系统可以从以下几个方面开展研究：

（1）提高企业责任感，注重用户算法素养的培养。首先，平台企业作为算法推荐技术的实际应用者，应当认识到自己对于信息茧房问题所应负的责任。要站在用户的角度，采取必要的技术手段规避和打破信息茧房，为用户提供多样化的推荐结果。其次，用户算法素养的提高是打破信息茧房非常关键的一步。用户必须对于自己正在获取信息保持清醒地认知，拒绝沉浸于算法推荐的同质化信息内容当中。与此同时，用户还需要提高对推荐算法技术了解的主动性，避免其被束缚在信息茧房中，无法获取到自己领域之外的信息。

（2）增加推荐多样性，优化算法推荐机制以及多领域审核机制。建立多样性的审核机制和优化推荐技术有利于打破信息茧房的封闭空间，建立整个平台空间的优质氛围。针对推荐多样性，要突破现有的个性化推荐环节单一的推荐模型，采用个性化搜索和个性化推荐相

结合混合模式，从更多角度刻画用户画像以满足用户个性化需求，实现推荐结果的多元性和丰富性。针对算法优化机制，要实现推荐算法的多目标优化。通过算法分析和内容提取，提高对用户最有价值的优质信息的推送权重；通过挖掘现实世界中复杂的社会关系框架提供的相关知识和关系图谱，满足人们对推荐模式更深层次的认知需求，使推荐实现公共价值、专业水准、多元化均衡目标的统一。针对多领域审核机制，应组建由多领域的专业人士和专家组成的审核团队，根据不同的专业领域知识从不同角度衡量推荐结果的多样化以及合理性。这里的领域包括但不限于人工智能、法学、伦理学等。尤其在新闻行业中，必须要具有专业素养的人进行审核和把关，不能完全依赖于算法运行来实现信息的传播，以保证公共价值得以体现。

(3) 增加人工在推荐状态中的干预，提高推荐结果的质量。人工编辑在算法推荐中的干预较少，通常只能决定内容的推荐状态。例如人工无法决定用户最终能否看到此条内容、以及在什么位置能看到这个内容等。特别地，人工智能时代信息分发标准发生了变革，由传统媒体时代在编辑和分发新闻时所依据的“时效性、重要性、接近性、显著性、趣味性”五要素正转变为算法对用户关联度、新闻实用性和用户兴趣度的高度关注和依赖。此外，在互联网时代，以算法技术为驱动的应用软件发展迅速，逐渐渗透到人类生活的各个方面，开发此应用软件的技术部门逐渐成为公司核心利益的重要支撑。因此，企业需要采取一些创新举措，更好发挥人工编辑筛选和整合信息的作用，最终形成消息或者内容的整合推送。

(4) 建立和规范算法推荐机制的监管制度。推荐算法根据用户的兴趣偏好自主的选择物品或者服务推送给用户，针对信息茧房问题，除了提高企业责任感和相关技术人员的素养、优化算法推荐技术、加强人机协调等措施，还需要政府制定相关的法律制度予以规制。将算法纳入法律管理，政府和有关部门能从宏观上加强对算法自身的法律监管。同时，相关监管部门可以通过分析算法推荐技术中的推荐倾向与推荐标准，对其推荐可能产生的一系列后果及影响进行评估，设置一条算法不能跨越的底线。结合以上分析，在法律监管中应从算法、企业平台和用户的角度对推荐系统进行监管，比如确定算法决策规则标准、算法的审查程序及规则；设置算法利用用户信息和用户隐私安全的底线；应用平台在利用算法推荐技术前的备案；相关部门不定期抽检算法推荐技术的运行情况，以及结合用户的投诉举报，在发现推荐决策规则异常后，立刻要求相关平台对异常进行解释说明，并对其违规、不法行为进行惩治。

## 4.2. 算法黑盒

在推荐系统领域，不论是工程实践还是学术研究，总体上还比较缺乏足够的推荐解释上的研究和投入，在真实的推荐产品落地上对推荐的解释也还比较初步。这种情况发生的原因：主观上企业和研究者都更加注重推荐系统精准性的提升，较少地站在用户的角度来思考问题，从而忽略了对推荐的解释；客观上受到国家或商业机密、普通用户缺乏技术专业基础、

基于深度学习算法推荐难以用人可以直观理解的方式说明等因素的限制。聚焦到学术界，利用知识图谱的可推理、可解释特性，可以有效解决推荐系统的黑盒问题。这类借助知识图谱实现可解释性推荐的方法一般利用知识图谱的路径推理技术，通过强化学习或者深度神经网络探索用户与候选物品之间的图谱连通路径实现可解释性推荐。用户与候选物品之间的图谱路径就是向用户推荐该候选物品的理由。较为前沿的技术 PGPR 以用户实体为起点，通过定义强化学习中政策网络、动作空间、状态转换函数以及回报函数等探寻知识推理过程，为推荐结果形成清晰的推荐理由。KGAT<sup>[31]</sup>使用图卷积神经网络以用户和项为中心，迭代的集成多跳邻居以获得用户与项的综合嵌入表示。最后计算用户与项的嵌入表示之间的相似度衡量用户对该物品的喜爱程度。其中用户与项之间由权重较高的边链接而成的路径是该推荐的解释理由。ERSR<sup>[32]</sup>修改了传统的图卷积以区分来自不同层的信息，并将表示向量分解为与数据中的某个语义方面相关的几个部分。

对用户来说，他不仅仅希望能从系统中得到合适的推荐，还想要清楚为什么系统做出这样的推荐，如此用户才会更加认可和信赖推荐系统。为了消除大众对算法黑盒的担忧，目前各大互联网公司和研究机构希望能够通过公布推荐算法和原理，以及相关研究方案，来提高算法的透明度，从而增加用户对产品的信任。比如，在 GitHub 平台上，程序员可以分享项目代码并与其他人在线合作；国内外知名研究团队和企业 在 GitHub 上保留开源数据和代码，以供其他人查阅；《纽约时报》设置了“开放” (Open) 专栏用于新闻技术文章共享；BBC 发表了许多学术论文来介绍其算法工具细节；谷歌曾公开其早期的 PageRank 排名算法的排序标准，并定期发布透明度报告；Facebook 在世界各地都会雇佣当地专业公司来审核其运行机制；苹果公司每年会发布两期透明度报告；字节跳动的抖音、头条等平台上已经增加或者正在增加推荐理由的功能，使用户可以从使用者的角度更好的了解到收到某种推荐的原因，比如抖音在给用户推荐某短视频时，会有类似于“80%的用户浏览了此视频”的易于理解和接受的推荐理由。同时，国际版 TikTok 设立透明度中心，允许特定监管部门和专家等部分审查其算法，并测试源代码。这是向公众解释说明其做出推荐的原因和准则，这是破除算法黑箱和增加算法可解释性的一种尝试。

然而，推荐系统作为一个包含非常多功能模块的工程体系，非常复杂，设计一个可解释的推荐系统并不是一件简单的事情，需要投入大量的研发力量和研发费用。整体而言，一个推荐系统主要包括三要素——用户、项、用户对项的操作行为(点击、购买、观看、浏览等)，其中项具备一些显式特征，用户对项的行为则代表了用户某种程度对具备该特征的项的兴趣。基于以上几点分析，目前的推荐系统的可解释性主要从如下几个方面实现：

**以用户为媒介**来做推荐解释用一句话表述是“跟您相似的用户都喜欢该推荐物品”。这类推荐解释的形式丰富多样，生成解释时往往会用到用户社交信息。微软研究员 Amit Sharma 等人把社交理由大致分成了四类，并分别进行了用户测试，这四类理由是整体热度 (2,612,211 位 Facebook 用户喜欢它)、朋友热度 (您的朋友中有 7 人喜欢它)、好朋友 (张三喜欢它)、



图 18 四类可解释推荐方法总结

好朋友及个数（张三和其他 5 位您的朋友喜欢它）。最终研究结果显示，后两种解释表达效果可以有效提高说服力。同时在使用这两种解释方式时，如果没有目的性地随机选择一个朋友，说服力会降低，因此选择一个正确的朋友来解释原因是非常重要的。以用户为媒介的推荐解释，与以物品、特征为媒介的推荐解释相比，更能推进用户对推荐物品进行评估，但是对于用户对于该物品喜爱程度在使用前后变化不大。如果推荐的物品用户不够喜欢，还反而降低了用户的满意程度，现实生活中已经有了相关场景的实现。例如在社交类推荐场景中，如微信、脉脉等，好友关系是可以十分自然地用于推荐解释的，微信“看一看”中的“朋友在看”就是一种基于社交关系的推荐解释，抖音屏幕下方“你的好友”标记也在提醒用户这是熟人发布的视频，当你看到给你推荐的信息有好友看过，你一般不会对该推荐产生反感的情绪，因为一眼就知道微信是因为你的好友才给你推荐这个，这种非常直接、一目了然的推荐解释，在朋友的亲密程度越高的推荐下，被用户接受的可能性越大；而在非社交类产品中，我们可以基于用户在产品上的行为相似度来构建用户关系，如果两个用户喜欢同样的项，有着同样的行为特征，即使可能他们根本不认识，我们也可以认为这两个用户是相似的用户。我们可以用“跟你兴趣相似的用户也喜欢 B”或者“喜欢 A 的用户也喜欢 B”等文字描述来进行推荐解释，这类的推荐解释应用范围非常广。此外，如果企业对用户有比较好的了解，可以通过构建一套完善且较为精准的用户画像系统，基于用户画像来找到与该用户相似的用户，从而利用相似用户的兴趣也可能相似来做推荐解释。比如，抖音会根据你关注的抖音号、地理位置给你推荐有着相同特征用户所看的视频，美团会基于你当前所在的位置在给你推荐餐厅，在这个场景中，你当前所在的位置就是一类画像信息。同时，企业还可以基于用户浏览项后的评论来做推荐解释，目前很多互联网产品都具备评论的功能，例如点评、弹幕等，如果将每个项的评论信息收集起来，通过 NLP 技术进行处理，从中提取出关键词，并根据用户的正负反馈及评论次数，可以找出提取出的关键词的重要程度，赋予不同权重，那么就

可以利用权重占比较大的关键词来作为项的推荐解释,因为这些关键词标签是用户在使用产品之后的真实反馈,能够比较精确的反映用户对于该产品的态度,从而形成推荐理由。

**以项为媒介**来做推荐解释用一句话表达则是“该物品和您喜欢的其他物品相似”。如果用户喜欢的某个产品 A 跟 B 相似,那么就可以根据 A 与 B 的相似性来预测该用户可能也会喜欢产品 B,这时给用户推荐解释的逻辑就是基于假设——“因为你喜欢 A, A 和 B 是相似的,所以我们猜你也喜欢 B”。各大网站上如亚马逊、Netflix 等都较为广泛地应用了该推荐逻辑,它的展现方式常常采用曾经购买、评论或者浏览物品的列表形式,列表中的物品和目前推荐的物品相似。相关研究表明这类解释方法能够增加系统的被接收程度,但是也存在问题,即用户可能难以找到物品之间的关联,另外如果用户希望看到一些自己没有购买过的、不一样的物品,此时推荐解释力不足。对于项相似性的建立,可以利用 TF-IDF、主题模型等算法对项的标题,使用场景,标签等内容特征来构建项之间的相似关系;也可以基于用户的行为构建项之间的关系,比如可以利用嵌入模型对用户操作(观看、购买、点击等)行为建模构建项之间的相似关系,除此之外,从用户的评论信息中提取出标签、用户态度、关键词等信息,利用 TF-IDF、主题模型等算法构建项之间的相似关系。

**以特征为媒介**的推荐解释用一句话表述是“您可能会喜欢推荐物品的这些(某某)特征”。这类推荐解释常见的表现通常是一一列举推荐物品中用户感兴趣的主要特征,还可以基于用户对项的行为为用户打上相对应地偏好标签,通过显示的标签来建立用户和项之间的联系,进而通过这些标签来做推荐解释,比如用户喜欢看科幻、恐怖电影,那么可以将“科幻”、“恐怖”作为用户的兴趣标签,并将科幻、恐怖类的电影推荐给用户。研究表明,与前两类推荐解释相比,这类推荐解释可以帮助用户准确判断是否喜欢推荐物品,用户对这类推荐解释的满意度往往更高。这类推荐要求建模粒度更细,通过提取物品特征并判断用户对不同特征的感兴趣程度,从而找到最适合用于解释的特征,因此,这类推荐也是目前的研究热点。

其他方式。除了以上几类常见的推荐解释,还有一些更细的推荐解释。基于环境的推荐解释,这类往往结合时间、天气、地域、用户心情、场所、上下文等各类环境因素来做推荐解释。比如“您目前位于王府井附近,给您推荐附近的火锅店”、“适合跑步时听的音乐”等,这类推荐解释就是**基于环境**的解释;基于科学知识、科学实验结果的解释,有科学证据表明晚上“睡前喝点红酒有助于睡眠”,可以用这类被验证或者被大多数人认可的科学知识来作为推荐解释。比如用户晚上在浏览淘宝,淘宝给用户推荐了一瓶红酒,其实可以用“晚上喝一点红酒、有利于睡眠”来作为推荐解释。这样的解释帮助增强用户对推荐的接受程度;基于**权威人士、明星效应**的解释,有时大家更愿意信任权威人士的观点,这就是为什么很多牙膏广告不是请明星,而是让一位穿白大褂的医生打扮的人对牙齿护理进行讲解说明,这是为了让大家注意到该牙膏效果所具有的权威性,从而提高用户的购买意愿,同样,明星光环效应,可以吸引许多粉丝甚至普通人的追随,因为大家会愿意相信公众人物欺骗用户的可能性低,因此我们可以利用权威人士及明星的这种效应来做推荐解释。当你女朋友在浏览淘宝

看到一条裙子，系统可以用“迪丽热巴同款日常穿搭”作为推荐解释，这样你的女朋友会愿意认为迪丽热巴同款上身效果会很不错，从而可能接受推荐；基于大众行为的推荐解释，从心理学角度来看，人类社会的羊群效应普遍存在，即人都是有从众心理的，大家都喜欢的信息、产品，你也会倾向于信任这类信息和产品，利用该原则进行推荐解释，例如我们可以用项被用户喜欢(观看视频、商品销量、浏览次数等)的次数来作为推荐解释的理由，“这部电影有 2000 万人看过，所以推荐给你”，“这篇文章已经有 3000+人浏览，所以推荐给你”，“该商品有 3 万人购买，所以推荐给你”等。

**克服算法黑盒的措施。**在构建推荐系统解释时，特别是在现实业务中实现推荐解释时，会遇到很多现实问题，为了让推荐解释不仅仅局限于理论价值，同时能够产生业务价值，则需要考虑如何将工程落地，如何在实际应用场景中将可解释性的推荐系统用于提升用户体验和用户满意度是当下推荐系统较为关注的问题。随着国家对可解释性人工智能的大力布局，目前各个行业也都注意到了用户对算法黑盒问题的关注，正在积极进行可解释性的探索。聚焦到推荐系统领域，利用深度学习技术建模推荐过程而带来的算法黑盒问题是人类进入人工智能时代必然面临的现象。面对算法黑盒问题，未来的推荐系统可以从以下几个方面开展研究：

(1) 推行算法披露机制，提高算法透明度。数据是算法运行的前提条件和基本要素，技术公司需要对数据质量、可信度、准确性、误差范围、采样范围、缺失值等进行必要说明。对于模型输入的变量及其特征，特别是模型运行时的关键变量值等信息均有必要向用户披露，从而使公众对算法的建模过程有较为全面的了解。此外，技术公司应就算法的运算和决策过程予以必要解释，特别是算法程序的推理规则以及相关的评测指标代表的含义，如准确度、基准值、置信水平等统计数据。

(2) 建立算法透明性法律规制，为算法透明提供现实可能性和外部环境条件。数据是人工智能时代的炼金石，媒体机构和技术公司对海量数据进行抽取、清洗、筛选和挖掘，实现对用户特征的建模和对产品内容的智能分发与推荐。按照算法透明性原则，平衡掌握数据的媒体机构和技术公司对数据披露的成本投入和现实收益比例，增加其披露数据的主动性。建立保护数据安全的法律法规，明确企业或政府工作人员数据泄露的责任并追责。同时，加大政府和机构行使监督和问责职能，提供安全有保障的外部环境。

(3) 加强社会监督，提高社会公众对算法透明化的认知。首先，算法开发和利益相关者应遵循伦理规范，兼顾平台效率和公共利益，向公众公布算法原理和运行机制，实现算法透明化。其次，鼓励社会公众应多学习和了解必要的算法常识，能清醒认识到算法带来的负面问题，对算法的系统性影响给予审慎的评价和认知。最后，建立第三方独立组织对算法透明进行全方位监督，支持学术性组织、非营利机构或自媒体的适当介入。

(4) 建立通用的可解释推荐框架及评测标准。目前来说，不同的推荐模型通常会采取特定的可解释性推荐系统，使得这种可解释推荐系统的拓展性较弱。对于一些新兴的推荐模

型，例如含有复杂深度神经网络的混合模型缺乏足够的解释能力。如果能够建立独立于模型的可解释推荐框架体系，就可以避免针对每个推荐系统分别设计解释方案，从而提高方法的可拓展性和适用范围。另外，可解释推荐如何评测一直是困扰可解释性研究员的一个难题，线上测试和人工标注条件要求较为苛刻，线下的评测标准不够成熟。在这个方面，如果能够降低可解释性研究的门槛，建立可解释性的基础理论，可以进一步提高推荐解释的实用性。

(5) 利用知识图谱或者生成模型的对话式推荐增强算法解释能力。首先，知识图谱本质上是以节点和边组成的结构图，形式化的描述了自然世界客观存在的事实。知识图谱作为可读性高的外部知识载体，以其结构化、可推理等特征有着极大的可能性来提高算法解释能力。比如，通过知识图谱包含的异质信息可以发现并打通不同媒介之间的关联，进一步增强推荐结果的可解释性。现有的可解释性推荐工作所生成的推荐解释往往局限于以物品为媒介、以用户为媒介或者以特征为媒介中的某一种，对这三类直接的相关性以及组合影响探索不够深入。如何能够利用知识图谱，将这三类媒介之间的壁垒打通，从而能够实现灵活选择其中最合适的媒介或媒介组合来适应具体情况，以对用户进行推荐与解释，是后续研究的热点。随着政府、企业和用户对于可解释人工智能的重视程度不断提高，企业和相关研究者对于该领域越来越关注，将知识图谱和深度学习结合，有着极大的发展潜力。其次，目前的推荐解释形式普遍都是预先设定，在沟通方式上过于呆板，能够根据用户心理进行推荐的情况有限，继而无法回应用户的个性化问题。如果能用生成模型让推荐系统根据用户问题或回答自行组织语言，以拉家常的方式探索用户的兴趣偏好，就可以在与用户交互的过程中进行实现灵活、多变的推荐解释。

#### 4.3. 价值观及伦理问题

人工智能的不断发展和日常生活中的广泛应用将会推动人类社会发展历程，但是，人工智能的发展也是一把双刃剑，在它服务于社会的同时，人工智能背后的伦理问题也逐渐显现了出来。现在的人工智能更多是计算机领域的工程师在参与，哲学、伦理学、法学等其他社会学科的参与较少，未来可以加强研究跨学科的人工智能伦理测试。如今，人工智能快速发展，人类在诸如围棋、图像识别、语音识别等等领域开始落后于人工智能，这提醒需要重视对人工智能进行的伦理测试，包括道德代码、隐私、正义、有益性、安全、责任等等，这对于人工智能最终究竟是服务社会，还是危害社会十分重要。但这并不是要求完全拒绝技术的发展，而是要重视技术带来的问题，保持必要的技术敏锐度，将算法掌握在人类能够控制的范围内。目前算法面临的价值观和伦理问题主要包括虚假新闻与低俗内容、算法偏见与歧视以及算法带来的不公平性问题。面对大众对这些问题越来越关注的需求，国内外互联网企业如字节跳动、腾讯、阿里、谷歌等都在理智积极解决算法带来的价值观和伦理问题。

值得注意的是，在学术界，已有一部分工作在解决算法偏见与歧视问题，避免因算法带来的不公平现象，从而缓解算法带来的价值观和伦理问题。研究<sup>[34]</sup>认为用户行为数据是观察

性的，而不是实验性的。这使得数据中广泛存在各种偏差，包括但不限于选择偏差、位置偏差、暴露偏见和流行偏见。盲目拟合数据而不考虑固有偏差会导致许多严重的问题，例如离线评估和在线指标之间的差异，损害用户对推荐服务的满意度和信任度等。其中选择偏见、曝光偏见、位置偏见和一致性偏见是用户的历史交互数据自身携带的属性，推荐算法在建模过程中会将数据本身的偏见放大，以显示的方式呈现给用户，而流行度和不公平性偏见是推荐算法自身携带的问题，比如对流行度偏见来说，推荐系统倾向于将流行度高的物品推荐给用户，就会出现流行的物品越流行，而不流行的物品越来越不流行。为克服上述推荐算法带来的偏见与歧视问题，比如 MF-IPS<sup>[34]</sup>借鉴无监督学习训练框架里的 tri-training 方式，设计了模型不可知的元学习方法解决选择偏差问题。该方法使用了三个对称的预测器，其中两个预测器用于生成具有伪评分的可靠数据集，另一个预测器在生成的伪评分数据集上执行最后的预测；SERec<sup>[35]</sup>假设人们的购物偏好是从在线朋友那里获得的产品信息，但这些用户不必共享相似的偏好。该方法设计社交正则和社交增强驱动的协同过滤算法解决曝光偏差问题。

聚焦到工业界，各大公司和企业目前相继开展对如何应对推荐算法带来的伦理问题的研究。具体来说，主要从以下几方面展开：

(1) 优化以算法为核心的分发模式，在网络世界中建立人类可以自由行走的“人行道”。部分机构发起“戳泡”运动，希望借助技术手段，来帮助人们摆脱思维禁锢，更加开放包容地倾听全方位的声音、观察世界不同的风景。例如，Google 推出了“逃离泡沫”的插件，可以根据用户的阅读偏好进行反向推荐，主要推荐一些格调积极、易于接受的内容。该模式不但可以有效处理类似于“信息茧房”等问题，同时还能够在感官上给用户带来更好的体验效果，实现双目标。哈佛大学教授凯斯·桑斯坦认为，网络传播的“回音室”局限同样会产生群体“极化”的问题。因此，必须创建“人行道”，让用户能够在互联网中自由行走，发现不同的“风景”，为他们分发各种不在计划内的信息，让社会中不同领域、不同年龄、不同地区的人都可以对自己所不了解的信息有所接触，从而摆脱“信息茧房”带来的信息范围不断收缩，最终局限于自身领域或自身思想的问题。

(2) 实现算法规范化，建立人工与机器结合的双重审核机制。在算法技术的完善层面，由人工对机器进行必要训练，结合语义分析、关键词分析等方法建立可以被人工智能识别的正能量模型、社会负面模型以及反低俗模型数据库，并由算法技术实现过滤和筛选平台的内容，从而有效剔除标题党、抄袭、虚假新闻和恶意信息等极度劣质信息。此外，一些互联网平台媒体采取多元共治，联合建立并加强外部审核和监督机制。通过邀请政府机构、传播学界、新闻媒体等社会领域知名专家组成专家团，考虑不同行业、不同领域的专家意见来保证内容审核的有效程度，同时还需进一步扩大意见领袖和热心网友的参与度。在虚假新闻和低俗内容的治理上，抖音作为一款已经全方位渗透到人类日常生活的 App，肩负着重要责任和使命。抖音在这一方面投入了大量的精力和资金以治理违规、违法、低俗、涉黄等不良内容。另一方面，抖音设置了极高的医疗广告门槛，要求医疗视频的信息发布者进行资格审查，也

组建了专业的医疗团队对医疗视频进一步筛选。同时，抖音平台也引入了一些官方账号和权威账号，提升视频内容生态系统整体的质量。还有国外巨头公司苹果已经聘请了许多传统记者对其订阅新闻应用 Apple News 进行编辑、策划以及监督，职业社交网站 LinkedIn 聘用了记者来从事内部产品编辑工作，字节跳动聘请专业人士对图文资讯、视频内容进行严格审核和筛选，对虚假新闻和低俗内容进行打压，向社会输出正能量。因此，在算法技术之下，需要坚持“人工+技术”的双重审查机制，要能够确保受众信息接收渠道和形式是多样的，从而防止算法主宰新闻领域。总之，要实现传统媒体、自媒体与算法平台和谐共存的优质大众传播体系的构建。

(3) 提升算法开发者的“算法素养”，设计者要内心自律并明晓自身承担的社会责任。算法的迅速出现和大量运用，使得绝大多数人对它的认知都停留在“知其然而不知其所以然”的层面。提升算法素养，一方面需要大力推动算法教育，使与算法有关的科普书籍进学校、进社区，让广大受众对算法的机制与运作原理有一个基本的认识；另一方面，算法驱动的信息提供商要采取措施以弥补信息消费中的裂痕，例如多种标注提醒。如今一些国外新闻 APP 会根据新闻倾向和用户阅读习惯标注出“左倾”和“右倾”的程度并对用户做出警示。具有社会影响力的算法驱动的信息平台，理应具有内心自律和社会责任，清晰自身作为具有一定意义上公共属性的机构，承担着重要的社会责任。当前，我国应积极建设算法新闻专业主义，加强对信息推送算法运行团队新闻价值观的培养。将公共理性置于工具理性之上，确立一种新的公共传播指导哲学。

**算法伦理问题解决措施。**随着推荐技术在人类日常生活中应用越来越广泛，推荐算法带来的伦理问题一直是社会各界不断讨论的话题。面对推荐系统带来的各种各样的伦理问题，未来的推荐系统可以从以下几个方面开展研究：

(1) 破除伦理困境。比如人工智能中非常经典的伦理窘境，在来不及刹车的情况下，如果自动驾驶汽车往前开就会把三个站在通道中的人撞死，但如果转向就会使另一个人死亡，陷入两难时，车辆应当如何选择？自动驾驶汽车出现事故，到底该由车的购买者负责还是由自动驾驶汽车的应用程序负责？在面对类似电车困境的问题时，功利主义和绝对主义会给出基于其立场的不同道德选择，在人类社会的基础上，破除伦理困境，寻找人工智能时代伦理问题的最优解是解决伦理问题的基础。

(2) 在算法研发中贯彻伦理原则。一方面，算法研发人员在进行算法研发时需要遵守一些基本的伦理准则，坚持对社会有益、坚持不利用算法作恶、采取包容性的设计、支持多样性、透明性，以及重视用户隐私的保护等。另一方面，还需要建立跨学科、多样性的算法伦理审查制度，将不同领域的专家聚集在一起，对算法技术和产品的伦理影响进行评估，并提出建议。同时发挥各领域专家的优势，积极参与算法落地过程。

(3) 对算法进行必要的监管。随着推荐算法技术越来越复杂，对于社会生活的影响越来越大，其重要性也逐渐提升。未来需要对算法进行一定程度的监管，可以设立的监管措

施包括标准制定，包括分类标准、性能标准、预测标准、设计标准、责任标准，算法开发标准，算法透明标准等等。此外，还需要建立严格的审批制度，对于一些算法以及用户数据的使用，未来可能需要监管部门进行事先审批，如果没有经过审批就不能向市场推出。最后，可能需要建立由计算机、法律、人伦道德等各界人员组成的第三方机构，为算法带来的不公平性问题进行专业鉴定，向社会公众披露不公平性问题的原因，以此加强对掌握算法的技术公司的监管。

(4) 针对算法决策和歧视，以及由此造成的人身财产损害，政府和有关部门需要提供法律救济，确定事故主体的法律责任。首先，对于算法决策，需要告知用户是否用自动化的手段进行决策决定。如果有，那么决策的机制又是什么，用户对于推荐算法的决策有知情权；其次，法律需要健全申诉的机制。对于算法造成的人身财产损害，一方面，无辜的受害人理应得到救助；另一方面，对于自动驾驶汽车、智能机器人等人工智能带来的责任挑战，需要严格责任、差别化责任、强制保险和赔偿等。

#### 4.4. 用户隐私与安全问题

针对这一问题，一系列隐私保护方法近年来被提出，主要包括随机干扰方法和分组匿名方法。这些方法的核心思想是把推荐服务提供者(service provider, SP)作为不可信的实体，用户的信息必须经过相应的处理后才能提交给 SP。然而，虽然随机干扰方法能够有效保护用户的历史信息，但却无法防止 SP 根据产生的推荐结果来推测用户的行为特征；分组匿名方法虽然能够将个人的行为模式泛化，但却需要用户之间建立信任关系并充分交换信息。因此，这些方法在实际应用中都具有一定的局限性。为克服上述问题，学术界和工业界开展了研究，并将目前较为流行的方法分为：(1) 基于数据扰动的方法。在基于数据扰动技术实现推荐系统隐私保护方面，用户利用各类加法扰动或乘法扰动方法对其拥有的数据集实现隐私保护，按照用户的意愿决定是否对隐私进行保护。(2) 基于公钥全同态加密的方法。在利用公钥全同态加密技术实现推荐系统的隐私保护方面。主要思想是利用公钥全同态对用户的历史数据训练集加密后上传到推荐服务器，后者利用其全同态性质在密文域上进行预测模型的建立和推荐结果计算，因此能在一定程度上解决推荐系统可用性与隐私性的统一问题。

(3) 基于多用户、多数据模型的隐私保护协同过滤推荐算法得到了学术界的广泛关注。除了产业界一直所关注的用户隐私和安全问题，实际上内容的隐私安全性问题也是一个值得关注的方向。在实际的场景中，一个文章结合文字和图片等多元信息隐藏传播关于他人的虚假新闻，并向公众暴露他人隐私；或者在视频中公然揭露他人隐私。这类安全隐私造成了极大的安全隐患，一旦传播开来，将难以控制对他人和社会造成的不良影响。目前各公司对这方面的处理有一些初步探索。例如，今日头条在审核时目前只允许评价公众人物，限制对普通人隐私的侵犯；抖音采取导出音频判断是否提及他人姓名的方式来控制这类信息的传播。对

于多元化信息带来的隐私与安全问题，还需要学术界、产业界做更系统的研究。同时对于公众人物的评价和讨论是否也侵犯了他人的隐私，这个伦理问题依然需要更深入的研究。

**保护用户隐私与安全措施。**在人工智能时代下，为保护用户数据的隐私和安全，研究标准模型或通用组合模型下适用于推荐系统隐私保护的新型加密安全模型是亟需进行的。推荐系统的隐私保护需要从三个方面进行构建：（1）用户数据隐私保护。在单用户、多数据模型下，用户历史数据训练集作为推荐系统预测模型建立与推荐结果计算的输入，是防止用户隐私与安全泄露的第一道防线。对用户数据的保护要求其能有效抵抗由恶意推荐服务器、本域未授权用户、不同域中的推荐服务器以及不同域中的未授权用户之间发起的合谋攻击；（2）推荐结果隐私保护。推荐结果一定程度上体现了用户的个性化偏好，对推荐结果的隐私保护就是对用户个人信息的保护。可以通过仅用户本身或推荐结果授权用户可以成功解密推荐结果的方式保护用户隐私；（3）预测模型隐私保护。推荐系统根据用户的历史训练数据挖掘用户兴趣偏好，利用用户特征进行推荐。推荐模型内部工作机制可以有效反应用户特征获取的过程。如果推荐过程或者推荐算法被泄漏，用户的个人数据也会被泄漏。因此，对预测模型的保护是预防用户隐私与安全问题的另一有效手段。

## 第五章 推荐算法面临的问题总结与思考

推荐系统是互联网行业中随处可见的产品形态。从淘宝、京东、微信、爱优腾、哔哩哔哩、今日头条到抖音都有推荐系统的影子。可以说推荐系统已经完全融入了我们的生活之中。然而随着推荐系统的流行，越来越多的问题被人们所关注，比如推荐算法的公平性，对低俗内容的容忍度和传播性以及推荐算法的黑盒问题等。但究其根本，产生这些问题的原因是推荐系统本身的设计缺陷还是推荐数据的马太效应和稀疏性，值得人们深思。这里，我们就推荐系统面临的典型问题给出客观的分析。

**推荐算法产生信息茧房？** 信息茧房从来都有，并非新现象。从传统媒体时代到如今的智能传播时代，一直存在于我们的身边。只不过进入智能传播时代，随着科学技术的发展，推荐算法为信息茧房提供了新的形成机制，并让我们得以显性化审视这一古老问题。从某种层面上讲，信息茧房效应是因为推荐信息的不平衡性，导致用户长期只能浏览限制领域的信息，从而失去了对事实的全面的认识和客观的评价，造成“真相稀缺”。探其根本，造成信息茧房的原因既包括用户自身追求个性化的主观需求造成的茧房，也包括推荐算法造成的茧房。对于用户主观偏好造成的信息茧房，用户倾向于选择自己爱买的东西或者浏览自己爱看的新闻，并乐于沉浸在与自己观点相吻合的世界里，而对观点相违背的世界本能的排斥和抵触，这种带有偏好的选择就会造成信息茧房。与推荐算法有关的信息茧房实际上与用户的个性化需求有关，为满足用户的需求，推荐算法根据用户画像和历史记录按照某种逻辑被不断的开发，给用户精准推荐某类固定信息以增强用户黏度。久而久之，用户就会存在于推荐算法造成的信息茧房中，造成认知结构单一，丧失对其他信息的知情权。为规避信息茧房的影响，通过不断改进推荐算法在用户数据上的建模能力，以及从用户和物品角度设计多样化推荐算法以更好地推断用户兴趣的完整性和多样性，从而更精准的为用户做推荐以破除信息茧房。

**推荐算法黑盒问题带来可解释性差？** 推荐算法的可解释性差是人工智能时代进化的必然现象，是科学技术高速发展的中间状态。以深度学习、强化学习或复杂的集成模型等为代表的先进技术方案给推荐系统的性能提升带来了显著收益，但由于这些模型自身带有的隐变量的不确定性，使得推荐系统越来越变成一种黑盒，无法向人们非常直观地呈现推荐理由及其合理性。不论是推荐系统的工程实践还是学术研究，在推荐解释上的研究和投入较少，在真实的推荐产品落地也不太关注推荐解释。幸运的是，近年来，政府、工业界和学术界对为人工智能赋予更好的可解释性给予了更广泛的关注。在后深度学习时代人工智能时代，研究数据驱动与知识引导深度融合的推荐模型将会是提高可解释性的关键手段。

**推荐算法带来虚假新闻和低俗内容泛滥？** 推荐算法的设计并不会带来虚假新闻和低俗内容的泛滥。推荐算法是一种过滤系统，根据用户信息、用户阅读的内容以及用户所处的环境等因素，从网上海量数据来训练和生成满足用户偏好的内容。具体来说，一则资讯的传播要经历创作者的创作和发布、内容生产者的抓取和分发，以及内容消费者的点击和阅读。其中，推荐算法充当内容分发者的角色。究其根本，虚假新闻和低俗内容的传播是由三者共同

造成的，而推荐算法的设计本身并不会带来虚假新闻和低俗内容的传播。一方面，由于追求流量关注的创作者发布大量低质化的内容以及内容消费者由于猎奇和好奇心理对低质化内容的大量阅读和传播，严重破坏了网络生态环境。另一方面，算法工程师未进行专业的新闻伦理培训，缺乏对文章的质量与内容的把关，仅关注内容和用户兴趣的匹配度指标，没有将对社会价值的导向要求内化为推荐的具体规则。所设计的推荐系统捕捉到用户对此低俗内容的偏好，为实现流量最大化，继而将此类满足用户偏好需求的相关内容推荐给用户，在客观上呈现出推荐内容的低质化倾向，造成虚假新闻和低俗内容的传播。因此，通过建立“人机结合”双重审核机制以及改进以算法为核心的分发模式，是避免低俗化内容更广泛传播的有效途径。

**推荐算法存在偏见和歧视，造成不公平问题？** 推荐算法在一定程度上带来的偏见和歧视追根溯源是由于数据对现实世界的不完全解读造成的。推荐算法的基础是大数据的挖掘和处理，通过挖掘用户点赞、转发、收藏等数据，来判断用户的兴趣偏好。样本数据的有限性和可伪造性影响了推荐算法的准确性和公正性。一方面，由于用户隐私的保护规定和技术手段的有限性等因素，加之每时每刻都在产生万量级规模的数据，全样本数据的获得并不具备可行性。另一方面，信息热度数据的可伪造性也为算法准确性带来遮蔽，可人为操纵的点赞量、转发量等虚假流量对于算法规则的欺骗，其带来的推荐结果必然包含误差和偏见。同时，推荐算法在数据选取标准、数据模型的权重设定各环节都贯穿着人为因素，算法不可避免地会反映设计者对于世界的认识。而当设计者将自身固有的社会成见嵌入规则之中时，智能算法在反映这种偏见的同时，也可能放大歧视倾向。再者，由于训练数据本身分布不均衡以及存在不同的流行度，推荐算法会受到流行性偏见、选择偏见以及曝光偏见的影响。随着时间的推移，这些偏见会被加强，使得流行和热门的项目越来越流行，被推荐的几率越来越大，而不流行或者冷门的项目则受到不公平待遇，被推荐的几率越来越小。因此，提高推荐算法对推荐结果的多样化，建立监管机制以及提升算法开发者的“算法素养”是减轻算法带来偏见和歧视现象的有效手段。

**推荐算法造成用户数据泄漏和安全隐患？** 推荐算法的设计本身不会造成对用户数据的泄漏和隐私威胁。但随着网上用户的行为越来越多，个性化推荐通常需要收集用户的个人数据以进行处理和分析，会让用户怀疑个人隐私遭到严重侵犯，对用户隐私造成威胁。为加强推荐系统对用户个人数据的控制和管理，使用密码技术来保护推荐系统中的用户隐私已经成为学术界研究的热点之一。

## 参考文献

- [1] Yehuda Koren, Robert M. Bell, and Chris Volinsky. Matrix factorization techniques for recommender systems. *IEEE Computer*, 42(8):30–37, 2009.
- [2] Ruslan Salakhutdinov and Andriy Mnih. Probabilistic matrix factorization. In *NIPS*, 2007.
- [3] Ian Porteous, Arthur U. Asuncion, and Max Welling. Bayesian matrix factorization with side information and dirichlet process mixtures. In *AAAI*, 2010.
- [4] Yong-Deok Kim and Seungjin Choi. Scalable variational bayesian matrix factorization with side information. In *AISTATS*, pages 493–502, 2014.
- [5] Sheng, L.; Kawale, J.; and Fu, Y. 2015. Deep collaborative filtering via marginalized denoising auto-encoder. 811–820.
- [6] Wu, Y.; Dubois, C.; Zheng, A. X.; and Ester, M. 2016. Collaborative denoising auto-encoders for top-n recommender systems. In *ACM International Conference on Web Search and Data Mining*, 153–162.
- [7] Xue, H. J.; Dai, X. Y.; Zhang, J.; Huang, S.; and Chen, J. 2017. Deep matrix factorization models for recommender systems. In *International Joint Conference on Artificial Intelligence*, 3203–3209.
- [8] I.Soboroff and C.Nicholas, Combining Content and Collaboration in Text Filtering, *Proc. Int'l Joint Conf. Artificial Intelligence Workshop: Machine Learning for Information Filtering*, Aug. 1999.
- [9] Cheng, H. T.; Koc, L.; Harmsen, J.; Shaked, T.; Chandra, T.; Aradhye, H.; Anderson, G.; Corrado, G.; Chai, W.; and Ispir, M. 2016. Wide & deep learning for recommender systems. 7–10.
- [10] Guo, H.; Tang, R.; Ye, Y.; Li, Z.; and He, X. 2017. Deepfm: A factorization-machine based neural network for ctr prediction. 1725–1731.
- [11] Zhang, L.; Liu, P.; and Gulla, J. A. 2018. A deep joint network for session-based news recommendations with contextual augmentation. In *Proceedings of the 29th on Hypertext and Social Media*, 201–209.
- [12] Shoujin Wang, Liang Hu, Longbing Cao, Xiaoshui Huang, Defu Lian, and Wei Liu. 2018, Attention-based transactional context embedding for next-item recommendation. In *Proceedings of AAAI*, 2532–2539.
- [13] Xu Chen, Hongteng Xu, Yongfeng Zhang, and et al. 2018. Sequential recommendation with user memory networks. In *Proceedings of WSDM*, 108–116.
- [14] Fuzheng Zhang, Nicholas Jing Yuan, Defu Lian, Xing Xie, and Wei-Ying Ma. 2016. Collaborative knowledge base embedding for recommender systems. In *Proceedings of KDD*, 353-362.
- [15] Xiang Wang, Dingxian Wang, Canran Xu, Xiangnan He, Yixin Cao, Tat-Seng Chua. 2019. Explainable Reasoning over Knowledge Graphs for Recommendation. In *Proceedings of AAAI*, 5329-5336.
- [16] Hongwei Wang, Fuzheng Zhang, Jialin Wang, Miao Zhao, Wenjie Li, Xing Xie, and Minyi Guo. 2018. Ripplenet: propagating user preferences on the knowledge graph for recommender systems. In *Proceedings of CIKM*, 417-426.
- [17] Hongwei Wang, Miao Zhao, Xing Xie, Wenjie Li, and Minyi Guo. 2019. Knowledge graph convolutional networks for recommender systems. In *Proceedings of WWW*, 3307-3313.
- [18] Thomas N Kipf and Max Welling. 2016. Variational graph auto-encoders. *arXiv preprint arXiv:1611.07308* (2016).
- [19] Yikun Xian, Zuohui Fu, S. Muthukrishnan, Gerard de Melo, Yongfeng Zhang. 2019, Reinforcement Knowledge Graph Reasoning for Explainable Recommendation. In *Proceedings of SIGIR*, 285-294.
- [20] Ruining He, Wang-Cheng Kang, and Julian McAuley. 2017. Translation-based recommendation. In *Proceedings of the Eleventh ACM Conference on Recommender Systems*. In *Proceedings of ACM*,

161–169.

- [21] Lei Zheng, Vahid Noroozi, and Philip S. Yu. 2017. Joint deep modeling of users and items using reviews for recommendation. In WSDM.
- [22] Yongfeng Zhang, Qingyao Ai, Xu Chen, and W Bruce Croft. 2017. Joint representation learning for top-n recommendation with heterogeneous information sources. In Proceedings of ACM, 1449–1458.
- [23] Hongwei Wang, Fuzheng Zhang, Xing Xie, Minyi Guo, 2018. DKN: Deep Knowledge-Aware Network for News Recommendation. In Proceedings of WWW , 1835-1844.
- [24] Qiannan Zhu, Xiaofei Zhou, Zeliang Song, Jianlong Tan, Li Guo. 2019. DAN: Deep Attention Neural Network for News Recommendation. In Proceedings of AAAI, 5973-5980.
- [25] Jaime Carbonell and Jade Goldstein. 1998. The use of MMR, diversity-based reranking for reordering documents and producing summaries. In Proceedings of ACM, 335–336.
- [26] Xiaojin Zhu, Andrew B Goldberg, Jurgen Van Gael, and David Andrzejewski. 2007. Improving Diversity in Ranking using Absorbing Random Walks. In Proceedings of HLT-NAACL, 97–104.
- [27] Laming Chen, Guoxin Zhang, Hanning Zhou. 2017. Improving the Diversity of Top-N Recommendation via Determinantal Point Process. CoRR abs/1709.05135
- [28] G. Adomavicius, Y.O. Kwon, 2012, Improving aggregate recommendation diversity using ranking-based techniques, IEEE Transactions on Knowledge and Data Engineering 24 (5) 896–911.
- [29] M. Gan, COUSIN: a network-based regression model for personalized recommendations, Decision Support Systems 82 (2016) 58–68.
- [30] Ting Yu, Junpeng Guo, Wenhua Li, Harry Jiannan Wang, Ling Fan: Recommendation with diversity: An adaptive trust-aware model. Decis. Support Syst. 123 (2019).
- [31] Xiang Wang, Xiangnan He, Yixin Cao, Meng Liu, Tat-Seng Chua: KGAT: Knowledge Graph Attention Network for Recommendation. KDD 2019: 950-958
- [32] Ninghao Liu, Yong Ge, Li Li, Xia Hu, Rui Chen, Soo-Hyun Choi. 2020. Explainable Recommender Systems via Resolving Learning Representations. CIKM . 895-904
- [33] Jiawei Chen, Hande Dong, Xiang Wang, Fuli Feng, Meng Wang, Xiangnan He: Bias and Debias in Recommender System: A Survey and Future Directions. CoRR abs/2010.03240 (2020)
- [34] Y. Saito, “Asymmetric tri-training for debiasing missing-not-atrandom explicit feedback,” in SIGIR, 2020, 309–318.
- [35] M. Wang, X. Zheng, Y. Yang, and K. Zhang, “Collaborative filtering with social exposure: A modular approach to social recommendation,” in AAAI, 2018.
- [36] Fung A . Infotopia[J]. Politics & Society, 2013, 41.
- [37] L Dormehl. The Formula: How Algorithms Solve All Our Problems. and Create More[J]. perigee books, 2015.
- [38] Aher R K , Adik A R . Privacy –Enhanced Web Personalization. 2007, 628-670.