# 第一章 绪论

## 1.1 研究背景

近年来，随着互联网产业的告诉发展，大数据时代随之应运而生，现今互联网已经融入到大众生活的点点滴滴中，成为人们生活中不可或缺的一部分，人们可以通过各种网络设备：智能手机、平板电脑、个人PC等接入到互联网中，人们借助互联网实现：远程办公、移动支付、网络购物、在线交易、网络娱乐、了解新闻时事等，互联网包罗万象，将人类社会各个方方面面联结在一起组成一张巨大的信息网。我们正处在信息化时代，网络上的信息种类越来越多、数量越来越庞大。社交网络、电子商务、网络互娱、信息传媒等每时每刻都在生产大量信息，人类被各种数据带入了大数据新时代，在互联网时代，人们越来越依赖通过互联网来获取信息、娱乐消遣，互联网丰富了业余生活；与此同时，互联网也充斥着大量无用信息、甚至垃圾信息，随之而来的问题是：用户如何在有限的时间内高效地找到自己的所需要的信息，内容提供者如何展示自己所提供的信息给最需要的人，即内容提供者和内容消费者之间的匹配，这成为了互联网时代所面临的问题。

在使用视频网站的过程中，假如用户明确知道自己的刚需，这个时候用户可以借助搜索引擎来查找得到自己所需要的信息，例如用户想看《复仇者联盟3》，此时用户可以在视频网站中搜索这部电影进行观影，但是如果用户比较无聊，处在一个需求不明确的时期应该如何处理，即用户想找一部自己喜欢的电影，但是不知道具体是什么，面对网站中数以万计的电影只能漫无目的的搜索，低效的搜索只会浪费用户的耐心导致用户放弃从而带来极差的用户体验，这类大部分用户都会遇到的问题叫：信息过载（Information Overload）。

为了处理信息过载问题，需要有一个根据用户需求能有效快速地为用户推荐信息的系统，它可以将网络信息过滤，将可能有用的信息分发给需要的用户，对于用户而言，可以将大海捞针变成湖里捞针，解决用户的选择困难问题，帮助用户获得良好的体验，从而获得心理上的满足，并且节省用户时间，帮助用户在有限的时间高效的完成目的；对于内容提供者来说，可以更加高效地分发内容，投其所好，赢得用户口碑，获得用户的青睐，节省系统的资源等不必要的消耗，在同行竞争者中占据先机，最终提高经济效益。这样的系统可以帮助内容提供者和内容消费者带来双赢。

## 1.2 研究意义

在实际使用中，我们发现视频网站总能贴用户心意地推荐合适的电影、电视节目，那么这些视频网站（内容提供者）是如何为用户（内容消费者）推荐喜欢的电影，这个过程的背后原理是什么，视频网站如何掌握用户喜好。首先，从信息获取的角度将人类获取信息的方法分为主动搜索与被动推荐的方式：首先以线下分析：一个人想要主动知道了解信息，首先会问周围认识的人是否知道，如果知道成功获取信息，如果不知道但知道可能知道就会得到一条线索，如果什么都不知道，那这条线索就断掉，将获取到的线索集合再次搜索；被动推荐是：你的朋友猜你想知道这个，问你我有一个消息你有没有兴趣知道来主动告诉你。

再类比到线上，主动搜索和被动推荐分别对应搜索引擎和推荐系统，搜索引擎是依靠用户主动在庞大的影视资源中搜索自己所心仪的电影、电视节目，但是主动查找需要很强的目的性，否则会浪费用户大量无效时间；而推荐系统虽然节省了用户时间，但是却将问题交付给内容提供者来处理，网站需要花大量时间获取用户数据、了解用户喜好，这个过程中充满各种问题。

个性化推荐系统通过用户数据用推荐算法来帮助用户进行查找和筛选信息，为用户提供充足、有效、合适、多样、新颖的信息帮助用户排忧解难，有效的推荐系统可以帮助减少用户的搜索时间，提高用户体验，但是推荐系统需要具有尽可能快的运行、响应速度、尽可能少的资源占用，既要提高效率，又要提高速度，因此对于推荐算法的优化成为解决这项问题的重要手段。

## 1.3 研究目的

通过建设一个小型的视频网站的过程来研究协同过滤推荐算法的原理和实现方法，利用推荐系统优化视频网站，赋予网站个性化的功能，能够为不同的用户定制不同的主页，推荐用户所喜欢的电影、电视节目。

在推荐算法系统中进行拓展：当使用不同的相似性算法时来验证出最适用的推荐算法，并尝试通过优化推荐算法或者提高计算能力来优化推荐系统的效率，一个优秀的推荐系统需要保证实时性、高效性、具有最高的利用率的系统才能具有商业价值。

研究推荐算法的评价性算法，在推荐系统的研究过程中，评价推荐结果的标准是一个非常基础且重要的前提，只有尽可能准确的评价才有推荐算法优化的前提。

# 第二章 研究现状及分析

## 2.1 推荐算法的发展

推荐算法和搜索引擎相似，但是搜索是主动的，而推荐系统根据用户的历史记录进行用户喜好推荐，用户处于一个被动的状态，推荐系统往往作为一个应用网站的子系统来进行设计，依赖网站的数据来进行推荐活动，而非单独存在。在互联网的各类网站中，推荐系统应用很广，从电商、视频、直播、新闻、媒体到广告、娱乐、游戏等都有推荐系统的影子，通过大量分析用户行为数据，给用户提供个性化服务来提高网站流量、吸引更多的用户、提高市场占有率，最终提高经济效益，个性化推荐系统应用于购物类网站、音乐类网站、电影和视频类网站、阅读类网站、社交和新闻类网站、小视频和直播类网站等。

个性化推荐算法的研究开始于上世纪90年代，随着互联网的发展开始应用于各行各业，随着行业发展需求不断的变化，对推荐系统的要求越来越高。个性化推荐系统按照内容可以分为（1）用户模块，即推荐的目标人群；（2）内容模块，即待推荐的主要内容；（3）推荐算法，将内容推荐给用户的规则，是连接用户与内容的核心桥梁，也是需要深入研究的内容。

## 2.1 推荐算法的研究现状

当前主流的推荐系统主要分为：基于内容的推荐系统（Content Based, CB）、协同过滤推荐系统（Collaborative Filtering,CF）、混合推荐系统，其中混合推荐系统是一种综合性推荐系统，是融汇各种推荐算法优点来设计，因为每种推荐算法都有其应用范围、限制，在不同的条件下，推荐效果各有差异，因此开发混合推荐系统成为当前的主流的研究方向。

协同过滤推荐算法（Collaborative Filtering,CF）目前应用较多，是一种成熟的推荐算法，目前在工业生产中被广泛推荐。其基本思想是根据用户历史行为数据推荐用户喜好。CF算法主要有三个流程：（1）数据处理，对于用户历史行为数据的数据处理；（2）相似性处理，根据历史数据确定相似性矩阵，对用户或者内容进行相似性计算；（3）相似性推荐，根据相似性矩阵为用户进行推荐。根据推荐系统要处理的问题可以分为：（1）评分预测；（2）Top N 推荐。

在推荐系统的发展中长期存在着一些问题：冷启动问题；数据稀疏问题；用户喜好偏移问题。冷启动问题即新用户、新项目没有历史数据的情况下如何加入到推荐系统中来；数据稀疏的问题是指海量的用户和项的情况下，用户数据稀少，导致用户相似性矩阵或物品相似性矩阵稀疏；用户兴趣漂移是指用户的兴趣变化，因为人是具有思维的个人，具有相当大的不确定性、偶然性有可能某一天用户的兴趣突然就变化180度，那么对于原先的相似性矩阵就无法适用了。

## 2.2 国内外推荐算法应用现状

个性化推荐系统在电影和视频网站中具有重要的应用地位，因为电影视频资源的数量庞大、不可遍历，导致用户对于海量信息的搜索存在巨大的困难。

腾讯视频网站的推荐系统，腾讯视频网站可以根据电影之间的相似性进行归类，根据用户播放记录帮助用户筛选用户可能感兴趣的电影；也可以根据用户好友的之间的联系，利用QQ、微信中的好友数据资源，来推荐用户好友圈子中受欢迎的电影节目。

QQ音乐平台的推荐，主要依据用户的历史记录和好友圈的历史记录来分析，通过比对所有人音乐的喜好来计算每首音乐之间的相似性来，通过用户的听歌记录来生产用户的私人歌单，通过所有用户的播放记录找到两个听歌风格相似的用户来互相推荐歌曲。

阅文网站的小说推荐，当用户浏览一本小说时，下面会给出推荐列表，从小说的相似性出发找到同类型的小说加入到推荐列表推荐给用户。

电商应用方面以亚马逊举例推荐系统在电商的发展主要得益于亚马逊，现在电商平台中的推荐系统就成了标配，通过在数以亿计的购买记录数据中提取用户喜好信息，用数据来刻画出来用户的喜好，对相似同类人、商品进行归类，掌握每一个人的性格、喜好等，通过用户喜好进行用户推荐。

在推荐列表应用中包含着商品的显性信息和隐形信息以及相关属性，其他用户对商品的评分，综合评价出两个商品的相关性。显性的信息是两个商品明面上的相似处，例如两本书之间的相似，而隐形的相似性是从用户的购买记录中发现的相似规律，例如笔记本和笔会经常被用户同时购买。在亚马逊的个性化推荐系统中，最重要的算法是一种基于物品的推荐算法，给用户推荐用户喜欢物品的相似物品，另一种推荐算法是基于用户的推荐算法，亚马逊是利用Facebook上的好友信息，判定好友之间的相似度来推荐商品。使用推荐系统的直接好处是给亚马逊公司带来直接的经济效益，据统计，亚马逊百分之三十左右的商品销售来自于推荐系统。

参考论文

1. 周春华, 沈建京, 李艳,等. 经典推荐算法研究综述[J]. 计算机科学与应用, 2019, 009(009):P.1803-1813.
2. 赵森. 融合用户偏好和项目关联的协同过滤算法研究[D].吉林大学,2020.