

目 录

摘 要.....	I
Abstract.....	II
第 1 章 绪 论.....	1
1.1 研究背景.....	1
1.2 研究目的和意义.....	2
1.3 研究现状.....	3
1.3.1 推荐算法发展.....	3
1.3.2 推荐算法的研究现状.....	3
1.3.3 国内外推荐算法应用现状.....	3
1.4 本文安排.....	4
第 2 章 相关技术介绍.....	5
2.1 推荐算法介绍.....	5
2.1.1 基于内容的推荐算法.....	5
2.1.2 基于内存的协同过滤推荐算法.....	5
2.1.3 基于模型的协同过滤推荐算法.....	5
2.1.4 混合推荐算法.....	6
2.2 相似度算法计算方式.....	6
2.2.1 基于距离计算相似度.....	6
2.2.2 基于系数计算相似度.....	7
2.3 推荐系统离线评测方案.....	7
2.3.1 离线实验评测概述.....	7
2.3.2 推荐算法评测指标.....	8
2.3.3 评分预测评价指标.....	9
2.4 本章小结.....	10
第 3 章 推荐算法选择和应用.....	11

3.1 基于用户的协同过滤算法介绍.....	11
3.1.1 User CF 原理.....	11
3.1.2 User CF 实现.....	11
3.2 基于项的协同过滤算法介绍.....	13
3.2.1 Item CF 原理.....	13
3.2.2 Item CF 实现.....	14
3.3 User CF 和 Item CF 的对比.....	15
3.4 本章小结.....	15
第 4 章 协同过滤推荐算法实验与应用.....	16
4.1 协同过滤推荐算法实验设计.....	16
4.1.1 实验 1 基于杰卡德相似度的 Item CF 实验.....	16
4.1.2 实验 2 基于余弦相似度的 Item CF 实验.....	18
4.1.3 实验 3 基于杰卡德相似度的 User CF 实验.....	19
4.2 协同过滤推荐算法实验结果.....	20
4.2.1 协同过滤相似性算法分析.....	20
4.2.2 推荐算法实验评测.....	21
4.3 本章小结.....	21
第 5 章 个性化视频推荐网站的设计.....	22
5.1 网站技术分析.....	22
5.1.1 后端技术介绍.....	22
5.1.2 数据库技术.....	22
5.1.3 网页前端展示技术.....	22
5.2 需求分析设计.....	23
5.2.1 视频播放网站前端.....	23
5.2.2 用户功能模块.....	23
5.2.3 推荐功能.....	24
5.2.4 管理员功能.....	24
5.3 数据库设计.....	25
5.3.1 用户数据表.....	25
5.3.2 电影数据表.....	26

5.3.3 电影播放记录表.....	26
5.3.4 缓存数据表.....	27
5.4 本章小结.....	29
第 6 章 个性化视频推荐网站的实现.....	30
6.1 开发环境搭建.....	30
6.2 用户网站功能实现.....	30
6.2.1 电影视频播放网站网页设计清单.....	30
6.2.2 用户个人信息管理.....	31
6.2.3 用户操作行为功能.....	33
6.2.4 站内搜索功能.....	33
6.2.5 推荐功能.....	34
6.3 后台管理功能实现.....	34
6.3.1 电影信息管理.....	34
6.3.2 电影资源管理.....	34
6.3.3 电影播放热度统计.....	35
6.3.4 用户信息管理.....	35
6.4 网站前端首页功能实现.....	36
6.4.1 前端柱形图绘制.....	36
6.4.2 前端分页功能实现.....	36
6.5 本章小结.....	36
结 论.....	37
参考文献.....	38
致 谢.....	40
附 录.....	41

第1章 绪 论

1.1 研究背景

近二十年，随着科技进步，互联网产业的发展，人们踏入大数据时代，互联网已经融入人们的日常生活。人们可以通过智能手机、平板电脑、个人计算机、智能手环及智能手表等网络设备接入到互联网，实现远程管理、远程监控、远程办公、移动支付、网络就医、在线学习、了解时事、信息交流及在线购物等操作，互联网将人类社会从独立的点联结成一张复杂、巨大的网。

由于互联网的网络节点繁多、拓扑结构复杂，因此信息流动复杂且迅速，就信息来源方面，有各大媒体、社交网络、电子商务网站、网络互娱公司、信息传媒平台、网民个人等源头提供大量的信息。在信息消费方面，有网民个体、各大数据分析公司、国家机关、行业公司等利用大量数据信息满足自身需要。例如亚马逊通过用户个人信息了解用户的个人偏好、收入水平等信息制定公司发展战略，信息数据俨然成为当前大数据时代最重要的战略资源。其次，用户经常会被各种推销消息所打扰，从推荐商品、购物、贷款、买房买车到各种诈骗、虚假信息，究其原因在于个人信息的泄露导致被垃圾数据所困扰。

如何合理、合法的利用数据信息资源，发展经济效益成为时代的主题。从用户个人角度，高效地查找所需信息可以节约个人时间，提高工作、生活效率；从公司发展角度，高效的信息过滤可以带给公司在同行业的竞争优势，提高公司商品的销售额。综上所述,商品和消费者之间如何高效、准确地匹配成为公司的财富密码，推荐系统成为了大数据时代发展方向。

结合电影观影网站的行业，上述问题引申为为电影和观众之间高效、准确地匹配问题，帮助用户找到心仪的电影带来经济效益。假如用户明确自己的需求，可以借助搜索引擎来查找得到自己喜欢的电影进行付费观影，例如用户甲想看《复仇者联盟3》时，可以在网站中搜索电影名进行付费观影；但是如果用户甲需求不明确时，应该如何找到喜欢的电影，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信息过载（Information Overload）问题。

为了处理信息过载需要有一个根据用户需求能有效快速地为推荐信息的系统，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 研究目的和意义

在日常生活中，可以发现视频网站方总能贴用户心意地推荐用户喜欢的电影、电视节目，这个过程背后原理是什么，作为内容提供者的视频网站是如何为作为消费者的用户推荐用户喜欢的电影。从信息获取的角度来看，人类获取信息的方法有主动搜索与被动推荐。

以实际经验举例：

主动搜索是：一个用户想要主动了解信息，会先从周围开始询问，再从周围了解到的线索拓展，继续向更大的范围询问，这也是搜索引擎的搜索过程。

被动推荐是：用户周围的朋友主动告诉用户相关信息，并由用户自行筛选，用户作为一个被动接收方不需要额外的搜索时间，为用户推荐信息是推荐系统需要处理的过程，这也是推荐系统的主要原理。

搜索引擎依靠用户手动在庞大的影视资源中搜索自己心仪的电影、电视节目，需要明确的目标，否则会陷入搜索困惑的地步，浪费大量无效时间。推荐系统通过将查找用户喜欢电影这一过程交付给系统进行，为用户节约搜索时间，通俗的说这是一个推荐系统根据条件猜测用户喜好的过程，目前研究旨在提高推荐系统猜测的准确率和效率。

个性化推荐系利用用户数据，使用推荐算法帮助用户查找和筛选电影，为用户提供充足、有效、合适、多样、新颖的个性化电影库。有效的推荐系统可以帮助减少搜索范围，节省搜索时间，提高用户体验。但推荐系统被要求有尽可能快的运行速度，尽可能少的资源占用，既要提高推荐准确率，又要提高推荐速度，这是推荐系统所面对的主要矛盾。

本设计通过建设一个小型的视频网站的过程，研究协同过滤推荐算法的原理和实现方法，利用推荐系统建立个性化视频网站，赋予个性化定制的功能，为不同的用户