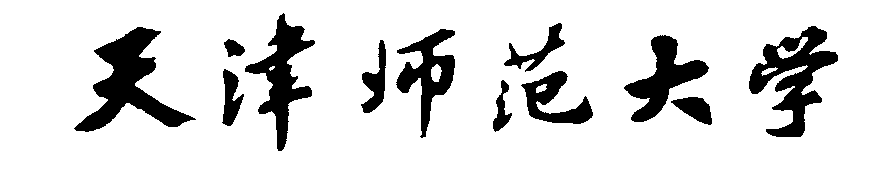
分类编号：\_\_\_\_\_\_\_ 单位代码： 10065

密 级：\_\_\_\_\_\_\_ 学 号： 1410030164



研究生学位论文

论文题目：基于协同过滤算法的学习资源个性化推荐

系统设计与实现

学 生 姓 名： 章 伟 申请学位级别： 硕 士

专 业 名 称： 教育技术学

研 究 方 向： 远程教育技术开发及应用

指导教师姓名： 于 鷃 专业技术职称： 副教授

提交报告日期：

天津师范大学学位论文原创声明

本人郑重声明：此处所提交的学位论文《基于协同过滤算法的学习资源个性化推荐系统设计与实现》，是本人在导师指导下，在天津师范大学攻读硕士学位期间独立进行研究工作所取得的成果。据本人所知，论文中除已注明部分外不包含他人已发表或撰写过的研究成果。对本文的研究工作做出重要贡献的个人和集体，均已在文中以明确方式注明。本声明的法律结果将完全由本人承担。

作者签名： 日期： 年 月 日

天津师范大学学位论文使用授权书

《基于协同过滤算法的学习资源个性化推荐系统设计与实现》系本人在天津师范大学攻读学位期间在导师指导下完成的学位论文。本论文的研究成果归天津师范大学所有，本论文的研究内容不得以其他单位的名义发表。本人完全了解天津师范大学关于保存、使用学位论文的规定，同意学校保留并向有关部门送交论文的复印件和电子版本，允许论文被查阅和借阅，同意学校将论文加入《中国优秀博硕士学位论文全文数据库》和编入《中国知识资源总库》。本人授权天津师范大学，可以采用影印、缩印或其他复制手段保存论文，可以公布论文的全部或部分内容。

本学位论文属于（请在以下相应方框内打“√");

保密□,在 年解密后适用本授权书

不保密

作者签名: 日期: 年 月 日

导师签名: 日期: 年 月 日

原创性声明

本人声明所呈交的论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得**天津师范大学**或其它教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

论文作者签名： 日期： 年 月 日

研究生学位论文使用授权说明

本人完全了解天津师范大学关于收集、保存、使用研究生学位论文的规定，即：

* 按照学校要求向图书馆提交学位论文的印刷本和电子版本；
* 图书馆有权保存学位论文的印刷本和电子版，并通过校园网向本校读者提供全文与阅览服务。
* 图书馆可以采用数字化或其它手段保存论文；
* 因某种特殊原因需要延迟发布学位论文，按学位论文保密规定处理，保密论文在解密后遵守此规定。

论文作者签名： 导师签名：

日期： 年 月 日

关于同意使用本人学位论文的授权书

中国科学技术信息研究所是国家科技部直属的综合性科技信息研究和服务机构，是国家法定的学位论文收藏单位，肩负着为国家技术创新体系提供文献保障的任务。从六十年代开始，中国科学技术信息研究所受国家教育部、国务院学位办、国家科技部的委托，对全国博/硕士学位论文、博士后研究工作报告进行全面的收藏、加工及服务，迄今收藏的国内研究生博/硕士论文已经达到100多万册。

学位论文是高等院校和科研院所科研水平的体现，是研究人员辛勤劳动成果的结晶，也是社会和人类的共同知识财富。为更好的利用这一重要的信息资源，为国家的教育和科研工作服务，在国家科技部的大力支持和越来越多的专家学者提议下，中国科学技术信息研究所和北京万方数据股份有限公司承担并开发建设了《中国学位论文全文数据库》的加工和服务任务，通过对学位论文全文进行数字化加工处理，建成全国最大的学位论文全文数据库，并进行信息服务。

本人完全了解《中国学位论文全文数据库》开发建设目的和使用的相关情况，本人学位论文为非保密论文，现授权中国科学技术信息研究所和北京万方数据股份有限公司将本人学位论文收录到《中国学位论文全文数据库》，并进行信息服务（包括但不限于汇编、复制、发行、信息网络传播等），同时本人保留在其他媒体发表论文的权利。

论文题目：基于协同过滤算法的学习资源个性化推荐系统设计与实现

毕业院校： 天津师范大学

毕业时间： 2017年6月

论文类型：博士论文　　　　□　　　　硕士论文

　　　　　博士后研究报告　□　　　　同等学力论文　　　□

授权人签字：

日期： 年　　月　　日

# 摘 要

当前，互联网技术与现代远程教育技术正处于高速融合期，以学习者为中心的网络学习以其不受时空限制、费用成本低、服务功能完善等优势迅速成为现代教育教学体系中重要的形式之一。然而，学习者在海量学习资源中却面临着无从选择的境地，这就是人们所说的“信息过载”问题。

“信息过载”问题一直困扰着学习者，如何解决它迫在眉睫。借鉴个性化推荐系统在音乐网站、视频网站、新闻媒体网站等的成功应用，教育领域的研究学者也开始将目光从传统教学转向互联网环境下混合式教学。本课题的任务是研究学习资源个性化推荐系统的设计与实现，提出运用协同过滤算法来优化推送信息，方便学习者更加有针对性且高效地享受互联网环境所带来的便利，同时考虑到当前互联网学习者多平台学习用户体验的要求，本课题设计开发的系统将支持主流三大应用平台：PC、手机、PAD。

本文研究了互联网中常见的几个学习资源系统，并归纳出个性化推荐在学习资源系统中存在的一些不足。针对这些不足，以天津师范大学的大学生群体为研究对象，设计并开发了基于协同过滤推荐算法的学习资源个性化推荐系统CFSYSTEM，并对系统的前后台进行了相应的测试工作。系统的初步试用在一定程度上能够帮助天津师范大学的大学生获取个性化的学习资源。

**关键词：**协同过滤；推荐系统；个性化推荐；学习资源；冷启动；数据稀疏

# ABSTRACT

At present the time has come that Internet technology are becoming increasingly connected with modern distance education. Internet learning, centered on learners, with no limit of time and space, low cost and updated service function, soon turns into one of the main patterns of modern teaching system. However, faced with tremendous resources of knowledge, learners feel at loos to make a choice. This phenomenon is so called ‘overloaded information’.

To properly address the learners’ trouble from it, effective ways are extremely in need. With the help of e-business, as well as successfully applied special recommended system, researchers are trying to explore the personal technology in the education area. This paper is to study the special recommendation on learning resources, proposing that applying collaborative filtering to improve the advertisement of learning resources, which can help learners study more efficiently and in the meantime realize the support of PC and mobile phones and PAD with regard to the publication of Internet.

This paper introduces several learning resource systems as well as some weaknesses and shortcomings for personalized recommendation under the systems. In response to these problems, choosing students of Tianjin Normal university as the target group, we design and develop a individualized resource system-CFSYSTEM on the basis of collaborative filtering algorithm，along with correspondingly testing work. Preliminary application of the system to some degree may assist college students of TJ with an access to personalized learning resources.

Keywords：collaborative filtering; recommendation system; specialization recommendation learning resources; data deficiency issue; cold boot; data sparsity

**目 录**

[摘 要 I](#_Toc481496852)

[ABSTRACT II](#_Toc481496853)

[第一章 绪论 1](#_Toc481496854)

[1.1 研究背景 1](#_Toc481496855)

[1.2 国内外现状研究 2](#_Toc481496856)

[1.2.1 国外研究现状 2](#_Toc481496857)

[1.2.2 国内研究现状 3](#_Toc481496858)

[1.3 研究意义 5](#_Toc481496859)

[1.4 论文结构 5](#_Toc481496860)

[第二章 理论基础与技术支持 7](#_Toc481496861)

[2.1 相关概念界定 7](#_Toc481496862)

[2.1.1协同过滤 7](#_Toc481496863)

[2.1.2学习资源 7](#_Toc481496864)

[2.1.3个性化推荐 7](#_Toc481496865)

[2.1.4学习资源个性化推荐 8](#_Toc481496866)

[2.1 联通主义理论 8](#_Toc481496867)

[2.2 个性化推荐算法 9](#_Toc481496868)

[2.2.1 协同过滤推荐算法 9](#_Toc481496869)

[2.2.2 基于内容过滤推荐算法 13](#_Toc481496870)

[2.2.3 基于关联规则的推荐算法 14](#_Toc481496871)

[2.2.4 常用几种推荐算法的比较 15](#_Toc481496872)

[2.3 相似度计算方法 16](#_Toc481496873)

[2.3.1 Pearson相似度 16](#_Toc481496874)

[2.3.2 余弦相似度 17](#_Toc481496875)

[2.3.3 欧氏距离相似性 18](#_Toc481496876)

[2.4 相关技术 18](#_Toc481496877)

[2.4.1 SSM框架 19](#_Toc481496878)

[2.4.2 JSON 19](#_Toc481496879)

[2.4.3 AJAX技术 20](#_Toc481496880)

[第三章 系统分析与设计 22](#_Toc481496881)

[3.1 常见学习资源系统分析 22](#_Toc481496882)

[3.1.1百度传课 22](#_Toc481496883)

[3.1.2网易云课堂 23](#_Toc481496884)

[3.1.3天津师范大学网络教学平台 24](#_Toc481496885)

[3.2 用户需求分析 25](#_Toc481496886)

[3.2.1 业务需求分析 25](#_Toc481496887)

[3.2.2 性能需求分析 26](#_Toc481496888)

[3.2.3 系统用例分析 27](#_Toc481496889)

[3.3 系统设计指导思想 28](#_Toc481496890)

[3.4 系统功能设计 30](#_Toc481496891)

[3.4.1 个人中心设计 30](#_Toc481496892)

[3.4.2 资源中心设计 31](#_Toc481496893)

[3.4.3 日志中心设计 31](#_Toc481496894)

[3.4.4 管理中心设计 31](#_Toc481496895)

[3.5 推荐算法设计 32](#_Toc481496896)

[3.6 数据库设计 34](#_Toc481496897)

[3.6.1 数据实体设计 34](#_Toc481496898)

[3.6.2 E-R图设计 38](#_Toc481496899)

[3.6.3 数据库表结构设计 38](#_Toc481496900)

[第四章 系统实现与测试 43](#_Toc481496901)

[4.1 系统实现总体过程 43](#_Toc481496902)

[4.2 系统核心模块实现 43](#_Toc481496903)

[4.2.1 登录模块 44](#_Toc481496904)

[4.2.2 最新学习资源推荐模块 46](#_Toc481496905)

[4.2.3 热门学习资源推荐模块 47](#_Toc481496906)

[4.2.4 学习资源审核模块 48](#_Toc481496907)

[4.2.5 我的评论模块 50](#_Toc481496908)

[4.2.6 学习资源个性化推荐模块 51](#_Toc481496909)

[4.3 系统运行测试 53](#_Toc481496910)

[4.3.1 单元测试 53](#_Toc481496911)

[4.3.2 功能测试 55](#_Toc481496912)

[4.3.3 界面测试 56](#_Toc481496913)

[第五章 总结与展望 58](#_Toc481496914)

[5.1 总结 58](#_Toc481496915)

[5.2 展望 59](#_Toc481496916)

[参考文献 60](#_Toc481496917)

[致谢 63](#_Toc481496918)

# 第一章 绪论

## 1.1 研究背景

2017年1月22日，中国互联网络信息中心（CNNIC）发布第39次《中国互联网络发展状况统计报告》[1]，文件再一次对中国网民的数量进行了一次统计，统计结果让人震惊，网民人数已经超过了7亿大关，成为世界上网民数量最多的国家，互联网普及率达到53.2%。面对互联网如此强悍的势头，教育领域研究者也不约而同地将目光锁定到互联网学习环境下教育的发展。

互联网行业在自身高速发展的同时带动了一批批其他传统行业的发展，传统行业纷纷跨界拥抱互联网。当然，互联网能够取得今天这样的成就，WEB 2.0的支持功不可没。WEB 2.0技术以用户为中心，数据作为核心驱动力，用户同时扮演着数据消费者和数据生产者的角色。教育行业的跨界也正悄然进行，将教育行业的价值逻辑与互联网行业的运转逻辑有效结合势在必行。互联网环境中的学习者数量出现井喷式地增长，他们在互联网上获取学习资源，发布学习资源，在这样一个相对自由的学习环境中，随之而来的是学习资源数量也呈现指数增长，产生“信息过载”。“信息过载”即指用户在海量信息中无法快速准确定位到自己所需信息。目前，解决“信息过载”有两种主要技术：信息检索技术和信息过滤技术。信息检索技术主要以搜索引擎为核心，其中，耳熟能详的有Google、Bing、Baidu、360和Sougou。搜索引擎要求用户提供关键字对所需信息进行描述，然后，根据所提供的关键字做分词处理，传入后台服务器，查找数据库，最后返回相关信息列表。然而，在一些情况下，用户对自己需求并不是十分明确，无法提供准确的关键字进行搜索，这样也导致了搜索结果差强人意。信息过滤技术主要以推荐系统（Recommender System）为核心，其中，耳熟能详的推荐系统主要分布在电子商务、音乐、图书、视频、社交领域，比如Amazon、eBay、Taobao、Emusic、YouTube和Facebook。推荐系统主要根据用户在系统上操作的历史记录，采用特定算法对记录分析，预测该用户在系统上的行为习惯，以此来推荐用户可能需要的信息。

协同过滤推荐算法以其对研究对象广泛适用性，以及能够挖掘用户潜在需求的优势，被大力推广，现在已经成为个性化推荐领域的核心算法。尽管协同过滤算法有其在个性化推荐领域的天生优势，但是马克思的辩证唯物主义向世人揭示了任何事物总是具有它的两面性的真理，协同过滤算法也不例外，它的缺陷在于系统运行初期新用户冷启动问题，新项目冷启动问题[2]，以及贯穿在系统运行整个生命周期中的数据稀疏问题。借助教育技术学的专业优势，采用协同过滤算法作为主要计算方法，开发学习资源个性化推荐系统。考虑到系统的健壮性，本课题研究中提出了最新学习资源榜单、热门学习资源榜单、用户-资源标签匹配来应对系统运行初期的新用户和新项目问题，增加用户浏览记录矩阵和用户下载记录矩阵来应对整个系统运行周期的数据稀疏问题。

## 1.2 国内外现状研究

互联网技术在国外的发展领先于国内。国外研究者借助互联网技术，采集用户信息、分析用户行为习惯、预测用户需求，个性化推荐系统随即孕育而生。近年来，国内大数据技术被广泛应用，个性化推荐系统初露锋芒，尤其在电子商务领域，阿里巴巴主席马云先生更是宣布世界从IT（Information Technology）时代即将进入DT（Data Technology）时代。个性化推荐系统在电子商务领域的成功也引起了其他各个领域研究者的关注。在教育领域，研究者通过对协同过滤算法在学习资源个性化推荐方面的深入研究，为学习者能够更好获得个性化学习资源提供了强大的支持。

### 1.2.1 国外研究现状

20世纪90年代中期，国外一些前沿研究领域的文章中陆续出现“协同过滤”的概念，它标志着在个性化推荐方向上的研究正式开始。

Amazon第一个将个性化推荐技术带到人们的日常生活中，它采集系统上所有用户的消费数据，运用个性化推荐算法计算商品的相似度，并将计算结果应用到每位用户上，为其推荐商品，激发购买欲望，为其带来了巨额的经济收益。Amazon的个性化推荐不是一种简简单单商业意义上的营销手段，而是其在电子商务领域的未来战略部署。电子商务平台最重要一个因素是用户粘合度，个性化推荐技术相当于电子商务平台和用户之间的粘合剂，牢牢绑定了平台上的用户。结合六度空间理论在现实中的应用情况，老用户为平台源源不断地带来新用户。基于项目协同过滤算法[3]（Item-based Collaborative filtering Recommendation Algorithm）就是由Amazon提出的。Amazon开创了个性化推荐技术的新时代，它的成功也引起了其他各个领域研究者的关注。不同领域中杰出的研究者将个性化推荐技术在各自领域中不断地完善和发展，并逐渐形成群集效应。个性化推荐技术开始跨越单个领域的限制，在不同的平台中的广泛被应用。

20世纪初期，国外教育领域的研究者敏锐地嗅到个性化推荐技术在网络学习中的机遇，纷纷进行个性化推荐技术应用到网络学习中的研究。1992年Xerox PARC研究中心以Goldberg为代表的研究者针对垃圾邮件过滤创建了基于协同过滤的Tapestry。紧接着，1993年，卡内基梅隆大学（Carnegie Mellon University）和麻省理工学院（Massachusetts Institute of Technology）分别发布WebWatcher和LIRA。随后，和麻省理工学院（Massachusetts Institute of Technology）在IJCAI再次发布个性化导航智能体Letizia推荐系统。WebWatcher、LIRA、Letizia标志着个性化推荐服务的诞生。荷兰开放大学（Open Universiteit Netherlands，缩写为OUNL）教育技术实验中心（Educational Technology Experiment Center，缩写为ETEC）通过对个性化推荐系统进行研究，总结了不同的个性化推荐算法的优缺点，适用场景及使用范围，最终以此结论设计并开发了综合性的个性化推荐系统。与此同时，加拿大萨斯卡通大学（University of Saskatchewan）的教育学者对于个性化推荐的研究也取得了突破性的进展，完成了智能网络学习推荐系统的设计与开发工作。智能网络学习推荐系统对学习者和互联网环境都有很强大的自适应性，分析学习者学习行为和互联网环境的动态变化进行即时的更新。

### 1.2.2 国内研究现状

国外特别是欧美地区的研究者在个性化推荐技术领域取得了惊人的成就，而且个性化推荐技术在各行各业的成功应用上也证实了它确实有很大的潜在价值。国内个性化推荐技术的正式起步源于2000年路海明等所著的一篇论文——《基于多混合智能实现个性化网络信息推荐》[4]。

国内个性化推荐技术比国外个性化推荐技术起步晚，但是拥有一个共同点就是国内外个性化推荐技术均于电子商务领域首先被应用。国内的美团、淘宝、天猫、京东、当当等企业间的竞争非常激烈，平台上用户的数量、平台与用户的粘合度、用户体验决定着企业的未来发展和生存，而个性化推荐技术能够很好的为平台增加用户数量，紧密平台和用户的粘合，提高用户体验。这些电子商务平台在业务上引入个性化推荐技术后，当用户登录系统，平台会根据用户之前浏览过的商品信息推送类似商品给用户，而且还会根据与该用户购买过相同商品的用户的兴趣，推荐商品给该用户。如今，国内电子商务巨头京东更是与国内第一社交平台腾讯展开战略合作，当用户在京东上浏览商品后，退出京东系统，进入社交平台腾讯，官方页面推送信息上将会出现之前用户在京东上浏览商品信息以及类似商品信息。可见，国内个性化推荐服务已经无处不在。

国内网络学习平台[32]如网易云课堂、百度传课、腾讯课堂、中国MOOC学院等也引入了个性化推荐技术，但是通过用户体验反馈以及论坛上评价发现，国内网络学习平台学习资源数量并不全面，推送信息不够准确，用户体验不高。个性化推荐技术在国内教育领域的主要一些研究如下：2002年，国内教育领域学者开始思考如何将个性化推荐技术引入到教学中来。丁琳[31]等人率先把个性化推荐技术整合到网络远程教育中，克服了一道道难关之后，完成了远程教育个性化的服务模型，这在国内教育领个性化推荐领域是一大创举。数字化教学发展在21世纪初期取得了不错的发展，但是在这个过程中，学习者开始对数字化教学中一成不变的模式不满，王艳芳[5]对学习者在数字化学习中所面临的问题进行深入分析，总结原因并确定为数字化学习中个性化得不到满足，于2008年构建出一套能够支持学习者的个性化推荐的数字化学习系统。2009年，何玲等提出基于学习者学习进度确定近邻学习者方法，跟踪WEB学习者学习行为，动态，动态推荐学习资源。2010年，互联网环境下的学习模式开始风靡国内，赵薇等[6]将互联网环境下的学习模式与当前成熟的数字化学习模式进行融合，构建符合用户学习习惯的互联网学习个性化推荐服务。2012年，虚拟学习社区中学习者对个性化推荐需求越来越迫切，杨丽娜等将尝试个性化推荐技术融入虚拟学习社区，并为此设计了建立虚拟社区的基本架构和技术流程。

总体来说，国内有一些对学习资源个性化推荐的研究，但是都处于起步阶段并不是十分的成熟。随着互联网学习者对于网络学习平台学习资源的要求越来越严苛，“信息过载”现象越来越普遍，学习资源个性化推荐服务终将成为决定网络学习平台兴衰的主导力量。

## 1.3 研究意义

互联网技术和大数据技术在世界掀起了一阵热潮，影响范围之广，层面之深前所未有，并由此推动了第三次工业革命和第四次工业革命。在教育领域中，高等院校率先顺应潮流，建立自己的学习资源系统，借此传递学校的教学风格，教学理念和教学文化。高等院校建立的学习资源系统由学校相关负责人发布相关学习资源，信息更新速度较慢且相对陈旧，互联网学习者往往不能获取到自己需要的学习资源。由此，导致了学习资源系统上用户粘合度不高、用户体验不佳、最终学习资源系统形同摆设，不了了之。当前，互联网学习者如何能够更好地从学习资源系统中获取适合自身的学习资源，系统设计者使用何种个性化推荐[20]机制，开发工程师运用怎样的科技力量搭建系统是教育领域研究中的一个新的热点。

本课题研究基于对当下互联网上学习资源个性化推荐系统（包括商业性质的百度传课、腾讯课堂、网易云课堂和公益性质的一些高等院校学习资源系统）研究的背景下，分析当前学习资源个性化推荐系统自身存在的一些不足，提出基于过滤协同算法进行个性化推荐。然而，单一的推荐算法自身也存在的一些不足，为了提高推荐结果的精确度，采用混合推荐算法和标签进行优化。“信息过载”背景下的网络学习平台学习资源个性化推荐已经大势所趋，具有广泛的应用价值。

## 1.4 论文结构

第一章：绪论。主要分析了学习资源个性化推荐的背景、国内外现状、以及深入地探索该领域研究所取得的成果，得出本次研究的意义。最后，将本课题论文结构进行概括总结。

第二章：理论基础及技术支持。重点介绍了作为本次课题研究核心理论基础的联通主义理论；详细分析了个性化推荐算法中的主流推荐算法、协同过滤推荐算法、基于内容过滤推荐算法和基于关联规则推荐算法，并对其进行比较后确定协同过滤推荐算法作为本次研究课题的核心算法；研究核心算法中相似度计算方法，包括Pearson相似度、余弦相似度和欧氏距离相似度，分析总结后选定Pearson相似度；最后，对系统设计和开发过程中所涉及到的技术做了简要说明。

第三章：系统分析与设计。分析当前互联网上现有常见的学习资源系统，总结当前学习资源个性化推荐系统存在的一些不足，寻求解决策略。针对天津师范大学的大学生群体进行用户需求分析，明确系统设计的指导思想，最后具体到系统功能模块，推荐算法以及数据库的设计。

第四章：系统实现与测试。确定系统实现的总体过程，进而详细阐述了其中较为核心的功能模块，展示核心代码的实现过程以及页面效果。系统完成后，运行系统并测试。

第五章：总结与展望。对整个研究的过程以及所得到的结果进行总结，为后期其他研究人员提供参考，并对未来下一步工作进行展望。

# 第二章 理论基础与技术支持

## 2.1 相关概念界定

### 2.1.1 协同过滤

1992年Goldberg等最先提出协同过滤这一概念，并加其应用到垃圾邮件滤和新闻推荐的Tapestry系统中。1994年，美国学者Paul Resnick等发表《GroupLens:An Open Architecture for Collaborative Filtering of Netnews》将协同过滤思想正式提出。直到今天，协同过滤技术已经成为当前算法水准最高，使用最普遍的个性化推荐算法。该类方法可以说是使用群体智能发掘用户潜在兴趣，这一点正是“协同”的体现[35]。协同过滤，顾名思义就是在将具有相同属性的因素过滤出来。

### 2.1.2 学习资源

从教育理论、教育技术学思维和资源建设方式的角度来讲，AECT(美国教育与传播学会)对学习资源做出的定义，被大众视为标准。随着社会的进步，AECT对学习资源的定义也在发生着变化。AECT’77的定义中，学习资源被分为“设计的资源”和“利用的资源”[36]。AECT’94定义中，学习资源是支持学习的资源，包括教学材料、支持系统、学习环境，甚至可以包括能帮助个人有效学习和操作的任何因素[37]。学习资源是构建教学的核心要素。

### 2.1.3 个性化推荐

个性化推荐，即根据用户属性的差异为用户提供符合其个人兴趣的推荐[38]。个性化推荐的概念最早是由 Resnick 在 20 世纪 90 年代提出，之后被Amazon第一次应用到实践上。如今，随着网络教育的发展及“信息过载”问题的出现，个性化推荐在各个领域的迫切需求引起了研究者的关注，表2-1对信息检索技术与个性化推荐进行了对比。

表2-1传统信息检索和个性化推荐对比表格

|  |  |  |
| --- | --- | --- |
| 内容 | 信息检索技术 | 个性化推荐 |
| 任务理解程度 | 清晰 | 模糊 |
| 自动化实现程度 | 低 | 高 |
| 动态支持 | 被动支持 | 主动支持 |
| 实现难易程度 | 简单 | 困难 |

个性化推荐能够为用户推荐其潜在兴趣的信息，从而提高信息获取的效率[39]。此外，个性化推荐还能够引导用户的信息需求，充分挖掘用户潜在兴趣，帮助其发现新的兴趣。通过主体对推荐信息的反馈情况可以动态地不断更新主体的兴趣模型，从而保持为主体提供最能够满足其需求的各种信息产品。

### 2.1.4 学习资源个性化推荐

学习资源个性化推荐的概念目前在学术界并没有给出统一的解释和定义，文献[40]提出，个性化在线教育推荐系统能够自适应用户的学习风格、知识水平的，而且能够智能向用户推荐其感兴趣或可使其提高知识水平的学习资源或策略。Chen Y J根据学习资源所处的环境，将学习资源的个性化推荐分为基础资源库的推荐、虚拟学习社区中的资源推荐、虚拟学习环境的资源推荐和学习系统中的资源推荐[41]。但是Chen Y J认为的前三种类型只是从用户兴趣出发进行推荐，并没有对用户的学习过程进行分析，而学习资源个性化推荐系统中主要以用户的学习过程为参考，向用户推荐与某具体学习过程有关的资源。

本课题研究的学习资源个性化推荐是基于协同过滤算法将系统中的学习资源有效的推荐给用户，让他们在海量学习资源面前不再迷航。

## 2.2 联通主义理论

2005年，乔治西蒙斯颇具预见性的发表了一篇主旨为数字时代学习理论的文章[17]，借着互联网的风口，吸引了无数前沿教育研究者和互联网学习者的关注，联通主义学习理论也由此在庞大的学习理论的体系中崭露头角，成为当下热点。乔治西蒙斯提出联通主义学习理论的基本原则在于知识的流通，指出知识是分散在社会各行各业中，并且在不同的意识形态中含义不尽相同。而后，在联通主义学习理论发展和完善的过程中，乔治西蒙斯也扮演了重要推动者的角色，积极对其进行补充和说明。其中有最为著名的三点补充，分别是：正确地认识和理解学习活动是怎样发生比学习活动本身更为重要，学习的最终目标是将所学到的应用的实际的生活场景中；随着社会的发展和科技的进步，学习的渠道已经突破了传统课堂的限制，互联网线上线下结合的学习方式已经成为重要力量；学习不单单是指学会和掌握知识，而是在于分享知识和创造知识的过程[18]。

联通主义学习理论指出学习活动是在学习者之间或者学习者与学习资源之间的联通中产生。互联网学习活动的过程中，学习者之间，学习者与学习资源之间如何进行有效的联通是该学习活动成功与否的关键。在学习者之间的联通中，学习资源个性化推荐系统通过记录学习者的行为习惯并分析学习者的学习数据，运用特定的个性化推荐算法，将系统中有类似特征的互联网学习者进行有效联通并聚类，为其推荐学习资源。在学习者和学习资源之间的联通中，学习资源个性化推荐系统通过记录当前学习者行为习惯，分析学习者的历史学习数据，将与当前学习者历史记录中学习资源有类似特征的学习资源与学习者进行有效联通并聚类，为其推荐学习资源。

## 2.3 个性化推荐算法

本课题将对时下流行的协同过滤推荐算法（Collaborative Filtering）、基于内容（Content-based）的推荐算法和基于关联规则（Rule-based）的推荐算法进行研究，分析其优缺点、适用场景和实践中的应用情况，最终选择符合课题研究的个性化推荐算法。

### 2.3.1 协同过滤推荐算法

协同[19]过滤推荐算法和其他推荐算法不同之处在于它可以适用非结构化复杂对象。假设一个大学网上学习资源下载系统的场景，学生A下载了编号为a、c和d的学习资源，学生B下载了编号为a和b的学习资源，学生C下载了编号为a和c的学习资源。这时可以近似认为学生C和学生A具有相同的兴趣，然后将编号为d的学习资源推荐给学生C（如图2-1所示）。这个过程体现了基于用户的协同过滤算法的思想。

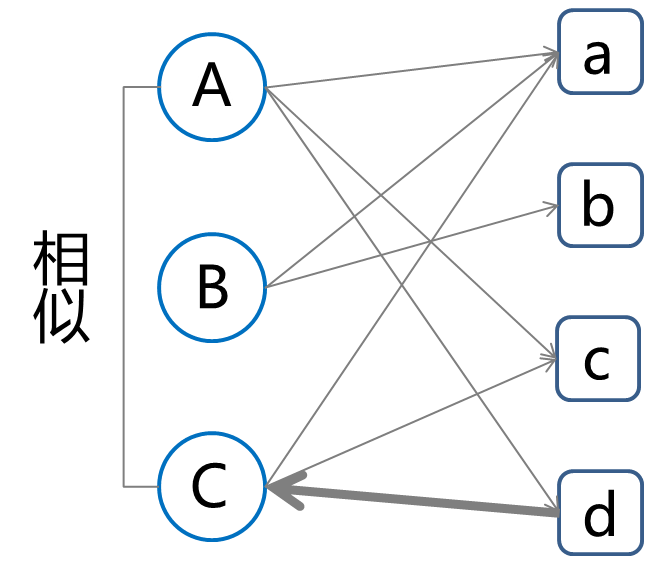


图2-1基于用户的协同过滤算法图示

同样的，从另外一个视角观察，学生A下载了编号为a、c和d的学习资源，学生B下载了编号为a和d的学习资源，学生C下载了编号为a和c的学习资源。通过仔细分析发现，下载编号为a和c的学习资源具有捆绑性，下载编号为a的学习资源的一般也会选择下载编号为c的学习资源。学生B下载了编号为a的学习资源，这时就会将编号为c的学习资源推荐给B同学（如图2-2所示）。这个过程体现了基于项目的协同过滤算法的思想。

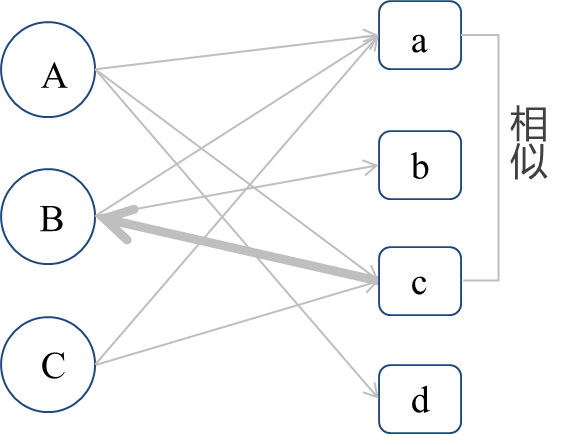


图2-2基于项目的协同过滤算法图示

1. 基于用户的协同过滤算法实现

基于用户的协同过滤算法是指具有相同兴趣群体中的一个用户如果对某件事物产生兴趣，系统就将认定这件事物对群体中其他用户同样具有吸引力，进而将其推荐给群体中其他用户。常见的“N-近邻”[7]是指系统根据当前用户信息计算出和当前用户具有类似兴趣的用户群体，并根据兴趣的相似程度进行排序。排在前N位的用户对项目的评分数据，将作为系统为当前用户进行个性化推荐的依据。

第一步：提取数据库中用户对于所有项目的评分数据，并将其组成一个m\*n的二维矩阵（m表示用户数量，n表示项目数量）。项目（Item）数量为m，集合表示为：Im = {I1，I2，……，Im}。用户（User）数量为n，集合表示为Un = {U1，U2，……，Un}，矩阵中第i行第j列表示的是Ui对Ij的评分Vij如表2-2所示。目前，大多数推荐系统中将评分的数值设在0-5的区间，通常用五角星进行表示。

表2-2用户-项目的评分矩阵

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| User Item | I1 | I2 | …… | Ij | …… | Im |
| U1 | V11 | V12 | …… | V1i | …… | V1m |
| U2 | V21 | V22 | …… | V2i | …… | V2m |
| …… | …… | …… | …… | …… | …… | …… |
| Uj | Vj1 | Vj2 | …… | Vji | …… | Vjm |
| …… | …… | …… | …… | …… | …… | …… |
| Un | Vn1 | Vn2 | …… | Vni | …… | Vnm |

第二步：根据第一步提供的用户-项目矩阵来计算用户之间的相似度得到N-近邻用户，最终将预测值高的项目推荐给目标用户。首先，计算用户间的相似性，得到与目标用户具有相似度所有近邻。然后，考虑近邻中的用户的兴趣相同程度。接着，将这些值的加权平均数充当目标用户对该项目偏好的预测值。最后，以项目的预测值为依据，将高分项目推荐用户。其中，每个偏好程度的权重取决于目标用户与近邻用户的相似度，用户之间越相似，偏好程度的权重越大。反之，用户之间差异越大，偏好程度的权重越小，计算流程如表2-3所示。本系统中，计算相似度的算法将在下一节中介绍。

表2-3基于用户的协同过滤算法流程

|  |
| --- |
| 基于用户的协同过滤算法 |
| 1. 根据用户-项目矩阵，计算系统中其他用户others与目标用户user的相似度值u\_similarity 2. 将相似度值u\_similarity由大到小依次排列，选择相似度值u\_similarity大的作为N-近邻用户 3. 统计N-近邻用户所有评价过的项目，然后将目标用户user评论过的项目过滤出去，列出候选推荐项目的集合V 4. 从N-近邻用户中找出对集合V已经评论的用户，计算其与目标用户user间的相似度i\_similarity 5. 按偏好程度权重i\_similarity将目标用户user对候选推荐项Item的偏好并入平均值，计算项目Item的最后预测值，将预测值高的项目推荐给目标用户user |

1. 基于项目的协同过滤算法

一个正常运行的系统，用户的数量往往会随着时间而增长，当用户数量达到一定程度时，这时系统采用基于用户的协同过滤算法进行推荐所需要的时间会越来越长，严重影响性能。因此，在2001年Sarwar[21]提出了基于项目的协同过滤算法。相较于基于用户的协同过滤算法根据用户相似度间接推荐项目，基于项目的协同过滤算法则比较直接，程序计算速度也更加快速。此算法直接根据目标用户已经评分的项目和未评分的项目间的相似度进行推荐[8]。

第一步：提取数据库中用户对于所有项目的评分数据，并将其组成一个m\*n的二维矩阵（m表示用户数量，n表示项目数量）。

第二步：根据第一步提供的用户-项目矩阵来计算项目之间的相似度得到N-近邻项目，最终将预测值高的项目推荐给目标用户，计算流程如表2-4所示。

表2-4基于项目的协同过滤算法流程

|  |
| --- |
| 基于项目的协同过滤算法 |
| 1. 获取目标用户未评分的所有项目集合I 2. 获取目标用户已经评论的所有项目集合J 3. 计算集合I与集合J的相似度similarity 4. 按偏好程度权重similarity将目标用户对集合J的偏好并入平均值，计算集合J的预测值，将集合J中分数高的项目推荐给目标用户 |

协同过滤推荐算法应用到个性化推荐服务时，并不需要关注系统项目的类型、属性、结构，这是该算法的一大优势，这也让协同过滤算法获得了一致好评。然而，随着互联网的迅猛发展，系统对用户体验要求越来越高，协同过滤算法想要适应目前大环境，就必须正视自身存在的冷启动问题和数据稀疏。

协同过滤推荐算法计算出推荐结果的数据是来自于系统所有用户对于项目的评分记录。在实际应用中，随着系统用户越来越多，系统项目越来越多，用户-项目评分矩阵却会越来越稀疏。根据数据显示，用户-项目的评分矩阵的稀疏度通常达到了99%[9]。从上面的介绍可以得出，用于协同过滤推荐算法相似度计算的数据主要就是来来自于用户-评分矩阵。无论是基于用户的协同过滤算法还是基于项目的协同过滤算法，当评分数据并不够大的时候，对于用户和项目的相似度计算结果也往往不够精确，进而影响个性化推荐结果。

数据稀疏的极端表现形式——冷启动问题，主要分为“新用户问题”和“新项目问题”，它们是个性化推荐系统建立初期都将会面临的问题[22]。“新用户问题”是指当前用户注册登录个性化推荐系统，由于系统中还没有当前用户对与项目评分数据，所以系统无法进行相似度计算，也就无法得到N-近邻用户，最终也就不能将N-近邻用户感兴趣的项目推荐给当前用户。“新项目问题”是指当前用户登录个性化推荐系统，虽然当前用户对项目有过评分历史记录，但是系统中其他用户均未对项目评分，也就是说系统中的项目评分仅仅来自于当前用户，所以协同过滤算法也就无法计算出类似当前用户已经评分的项目。这样也导致了个性化推荐系统无法为当前用户推荐项目。

冷启动问题和数据稀疏是制约协同过滤推荐算法进一步推广的两大因素，本课题也将对如何解决这两大难点提出相应的解决方案。对于冷启动问题，系统的设计中加入最新学习资源榜单、热门学习资源榜单和用户-学习资源标签匹配三大方案加以解决。对于数据稀疏问题，系统将分析用户浏览学习资源历史情况、用户下载学习资源历史情况配合学习资源评分加以解决。

### 2.3.2 基于内容过滤推荐算法

基于内容的过滤推荐算法结合信息检索提供的数据，进行信息的筛选，最终达到给用户推荐项目的目的。举例说明，百度搜索中，用户在搜索栏中输入所要查询信息的关键字，页面将会根据关键字信息提供信息列表。用户下一次登录百度搜索页面时，在百度搜索栏下面新闻列表中，发现除了百度发布的一些广告信息外，还有一些与用户之前在搜索栏中输入关键字相关的新闻推送，这就是典型的基于内容过滤推荐算法。基于内容过滤推荐算法实质就是解析文本内容，利用信息检索、信息聚类、资源分类等技术对内容分析后，进行推荐。基于内容的推荐算法的优势是研究信息的主体是文本内容，而文本内容在计算机中计算的速度相对其他类型资源更加快速，相对而言处理速度也会更快，推荐内容更准确，硬件环境要求更低。但是，对于文本的处理恰恰也是基于内容的推荐算法的不足之处。当今，互联网技术发展迅速，用户与机器交互的方式呈现多样化趋势，基于内容的推荐算法越来越无法满足互联网用户的需求，它主要的不足呈现在以下两个方面：

（1）视频资源、音频资源、图片资源等非文本类型资源没办法通过基于内容的算计进行个性化推荐。

（2）基于内容的推荐算法在进行推荐是存在这样一个问题，那就是它基于用户存储的历史信息进行推荐，对用户数据中未存储的信息无法进行数据挖掘[34]再推送给用户。在电商系统中，以Alibaba的天猫平台为例，基于内容的推荐算法根据用户之前输入的关键字信息来推荐相关商品，而这些商品有一个共同点，就是在系统中的评分都相当高，不具有个性化。

基于内容的推荐与基于协同过滤推荐存在差异的是，基于内容的推荐[10]并不是依赖用户对项目的历史评分或反馈信息，而是依据用户曾经感兴趣的项目的内容特征信息为用户建立兴趣特征模型，然后从系统中选择与用户过去喜欢的项目最为相似的产品。简单而言，就是将用户过去喜欢的项目与所有候选项目进行对比，推荐匹配程度高的项目[11]；或者直接推荐与用户偏好相似度高的项目。基于内容的推荐算法基于这样的假设：用户现在感兴趣的项目与其过去喜欢的项目相类似。

### 2.3.3 基于关联规则的推荐算法

计算机专业数据挖掘领域中有一些著名的算法，基于关联规则的推荐算法就是其中之一。系统对所有用户的历史数据进行提取，计算项目集合中具有关联规则的项目组合，将这些项目组合存放在数据库中。当系统发现当前用户正在操作项目组合中的某一个项目时，会将项目组合中其他的项目推荐给当前用户。关于基于关联规则的推荐算法有一个经典案例—啤酒与尿布，它说的是美国沃尔玛超市通过数据发现，男性顾客在购买婴儿尿片时，经常会买几瓶啤酒来犒赏自己，于是商店营销人员想出了将啤酒和尿片放在一起的促销手段，事实证明尿布和啤酒的销量大大提升。传统的关联规则挖掘算法，例如Apriori和FP增长都有被应用到学习资源推荐领域。文献[12]是基于Moodle 框架系统挖掘用户历史行为数据，提出了结合K-means聚类方法和Apriori 关联规则方法混合的方法推荐课程。文献提出了基于协同过滤和序列模式混合的学习资源推荐方法事实上也取得了较好的实验结果。文献[13]构建了一个集合挖掘序列模式、资源的多项特征和用户评分的推荐模型。挖掘序列模式的方法是对原来的Apriori 进行了改进，然后结合了 Prefix Span寻找模式树结构。

基于关联规则的推荐算法[14]即使有很强悍的扩展性，而且推荐结果也常常能够为用户所接收，但是它并没有真正达到高程度个性化的要求，而且形成关联规则往往需要海量数据作为支撑，这在系统初期系统项目往往很难达到海量的要求。就算系统项目达到了海量数据级别，关联规则的生成也往往很难，准确度也未必很高。

### 2.3.4 常用几种推荐算法的比较

表2-5常用几种推荐算法的比较

|  |  |  |
| --- | --- | --- |
| 推荐算法名称 | 优点 | 缺点 |
| 协同过滤推荐算法 | 自动化程度、个性化程度高 | 冷启动问题（新用户问题和新项目问题） |
| 能够处理非结构化的复杂项目 |
| 随着系统运行时间的推移，推荐结果更加准确 | 数据稀疏 |
| 挖掘用户潜在兴趣 |
| 基于内容过滤推荐算法 | 文本结构类项目推荐准确度、效率较高 | 推荐形式单一，对于非结构化的复杂对象无法进行操作 |
| 对用户潜在兴趣的挖掘能力不够 |
| 基于关联规则的推荐算法 | 推荐结果扩展性高 | 个性化程度低 |
| 关联规则推荐准确率不高 |
| 关联规则生成难 |

## 2.4 相似度计算方法

在对协同过滤推荐算法、基于内容的推荐算法和基于关联规则的推荐算法三种算法研究的前提上，结合本课题的研究内容，最终将协同过滤推荐算法作为学习此次研究的主要算法。在上一小节的协同过滤推荐算法的说明中可以了解到，无论是基于用户的协同过滤推荐算法还是基于项目的协同过滤推荐算法，它们关键步骤都是寻找N-近邻用户集合或者N-近邻项目集合，而寻找N-近邻用户集合或者N-近邻项目集合的前提是计算出用户间或者项目间的相似度[15]。根据相似度计算的结果直接确定最近邻，并决定推荐结果。接下来就对当前流行的三种计算相似度的算法进行详细介绍。

### 2.4.1 Pearson相似度

Pearson相似度中一个重要概念是Pearson系数。Pearson相关系数（Pearson correlation coefficient）也称皮尔森积矩相关系数(Pearson product-moment correlation coefficient) ，是一种线性相关系数，它是在-1到1区间的一个数值，可以反映两个变量之间相关的程度。通过对数学概念中线性相关系数的理解得出，具有线性相关的两个变量，当其中一个变量变大另外一个变量也会随之变大，反之，当其中一个变量变小另外一个变量也会随之变小。在协同过滤算法相似度的计算中，需要从系统中提取两个用户变量或者两个项目变量，当两个变量数值变化呈现正相关时，表示两个变量具有线性关系，用户间或者项目间的相似值高，而当两个变量数值变化成负相关时，表示两个变量不具有线性关系，用户间或者者项目间的相似值低。Pearson系数的计算公式为：

在基于用户的协同过滤算法中应用Pearson系数时，Pearson系数的计算公式将做出与之相对的改变，相似度计算公式如下所示。是用户u和用户v基于Pearson系数计算得到的相似度。其中，I(u)为用户u评论过的项目集合， I(v)是用户v评论过的项目集合。是用户u对于项目i的偏好度，是用户v对于项目i的偏好度，和分别表示了用户u和用户v在全部项目上偏好值的平均水平。

同样地，在基于项目的协同过滤算法中应用Pearson系数时，Pearson系数的计算公式也要做出与之相对的改变，相似度计算公式如下所示。代表项目u和v基于Pearson系数计算得到的相似度。其中，I(i)和I(j)分别表示评论过项目i和评论过项目j所评论的项目集合。表示用户u对项目i的偏好度，表示用户v对项目i的偏好度，和分别表示了用户u和用户v在全部项目上偏好值的平均水平。

### 2.4.2 余弦相似度

余弦相似度[16]是借助计算两个向量的夹角余弦值来估算用户的相似度。基于用户的协同过滤算法，用户对项目的偏好值抽象为m维空间中的点，其他用户和目标用户所抽象成的点分别与原点进行连接，两条线段交叉后形成夹角。这个夹角运用数学知识中的余弦定理，计算其余弦值，并以余弦值用来表示用户间相似度；基于项目的协同过滤算法，将系统中其他所有用户对所有项目评价信息的偏好值和目标用户对所有项目评价信息的偏好值抽象为m维空间中的点，其他用户所有项目历史评价信息偏好值、目标用户所有项目历史评价信息偏好值分别与原点连接，两条线段形成夹角的余弦值表示项目间的相似度。计算余弦相似度的公式为：

根据数学知识中对于余弦值的理解，两个角余弦值的范围在-1到1区间。当夹角为0°数值为1，当夹角为90°数值为0，当夹角为180°数值为-1，当夹角为270°时数值为0，当夹角重返360°即0°时数值为1。将数学知识带入余弦相似度中来，总结，夹角在0°-180°区间中，夹角越小，用户间或者项目间的相似度越高，相反，夹角越大，用户间或者项目间相似度越低；空间角度以360度为标准，夹角在180°-360°区间中，夹角越小，用户间或者项目间相似度越低，相反，夹角越大，用户间或者项目间相似度越高。

### 2.4.3 欧氏距离相似性

欧式距离也称之为欧几里德度量（Euclidean Metric），指在m维空间中的两个点之间物理距离或向量的自然长度。在个性化推荐算法中欧式距离计算相似度简单易行，将系统中的用户或项目抽象成m维空间中的点，m为用户数量或项目数量，偏好度为维度坐标。项目的相似度的计算除了上述Pearson相似度和余弦相似度也可以计算m维空间中两点的欧式距离来推导。欧式距离计算公式为：

d所计算出来的数值能够对用户间或者项目间相似度作为参考。d的值越大代表两个点在m维空间中距离越远，用户间或者项目间越不相似；d的值越小代表两个点在m为空间中距离越近，用户间或者项目间越相似。d代表距离长度，所以数值永远为一个大于0的数字。为了让欧式距离d更形象体现出用户间或项目间相似度，会将其抽象成一个在0到1区间的数值。转换公式为：

使用上述表达式即可以得到用户间或者项目间相似度。当用户或项目的偏好值完全相同的时候，d等于0，w等于1；当用户或项目的偏好值不完全相同时，偏好越不同，d值越大，w值越趋向于0。w的值越大，相似度越高，相反，w的值越小，相似度越低。

## 2.5 相关技术

系统开发过程中笔者主要采用JAVA编程语言，Windows 10专业版操作系统，WEB服务器使用的是Apache开源社区的Tomcat 8.0版本，浏览器支持IE8.0及以上，FireFox，Google Chrome和Opera。

表2-6开发工具列表

|  |  |
| --- | --- |
| **开发工具类型** | **开发工具名称** |
| IDE编辑器 | IntelliJ IEDA 2016 |
| 数据库 | MySQL5.0 + Redis |
| 数据库图形化工具 | Navicat Premium 11 |
| 数据库设计 | Power Designer 15 |
| 用例图设计 | Visio 2010 |
| 图像处理软件 | PhotoShop CS6 |
| 文本编辑器 | Notepad++ |
| 反编译器 | Jd-gui |

### 2.5.1 SSM框架

SSM框架是当前Java EE开发领域中主流三大框架Sring、SpringMVC和MyBatis的简称。Spring相当于一个大型工厂，负责将SpringMVC和MyBatis进行有效整合，完成开发过程中显示层、业务逻辑层、数据访问层有机调用。

Spring是一个开源框架，它是为了解决企业应用开发的复杂性而创建的，使用基本的JavaBean来完成以前只可能由EJB完成的事情。然而，Spring的用途不仅限于服务端的开发，从简单性、可测试性和松耦合的角度而言，任何Java应用都可以从Spring中受益。简单来说，Spring是一个轻量级的控制反转（IOC）和面向切面（AOP）的容器框架。

SpringMVC属于SpringFrameWork的后续产品，融合在Spring Web Flow中，SpringMVC分离了控制器、模型对象、分派器以及处理程序对象的角色，这种分离让它们更容易定制。

Mybatis本是apache开源社区中著名开源项目iBatis，2010年这个项目由apache software foundation迁移到了google code，并改名为MyBatis。MyBatis是一个基于Java的持久层框架。iBatis提供的持久层框架包括SQL Maps和Data Access objects(DAO) MyBatis消除了几乎所有JDBC代码和参数的手工设置以及结果集的检索。MyBatis使用简单的XML或者注解用于配置和原始映射，将接口和Java的POJO映射成数据库中的记录。

### 2.5.2 JSON

JSON全称**J**ava**S**cript **O**bject **N**atation，一种轻量级数据交换格式，适合不同操作系统包括当前主流的Windows、Linux、Unix、MacOS、Windows Phone、Android等WEB服务器与JavaScript的数据交互。JSON是在传统数据交换格式XML的基础上优化发展而来。主要优化有：降低交互数据量的大小；JSON能够在JavaScript和Jquery充当对象使用，而且原生支持，不需要第三方工具辅助；运行过程中数据解析速度快，提高开发效率。

### 2.5.3 AJAX技术

AJAX全称**A**synchronous **J**avascript **A**nd **X**ML，一种用于创建交互式网页应用的开发技术，它能够在静态网页上通过HTTP请求连接后台服务器将其变成动态网页。HTTP请求发送至后台服务器，后台服务器接收数据并进行业务处理后返回相应的数据至前台，完成页面的异步更新。

本次系统开发过程中，主要使用的是Jquery操作AJAX，简单易用高效，跨越不同内核的浏览器版本。Jquery操作AJAX主要有以下几种方法：$.load()、$.get()、$.post、$.getScript和$.ajax()。$.ajax()是Jquery底层最为原生的AJAX实现方法，以下用$.ajax()一段代码进行说明。

|  |
| --- |
| $.ajax(){  [url : localhost:”8080/cfs/index/login.do](url:localhost:)”,  type : ”POST”,  data : {“username” : ”1410030164” , ”password” : ”1410030164”},  dataType : ”json”,  beforeSend : function() {  alert(“正在加载，请稍等”);  },  success : function() {  alert(“成功”);  },  error : function() {  alert(“失败”);  }  } |

表2-7 Jquery AJAX 主要参数说明

|  |  |
| --- | --- |
| 参数 | 参数说明 |
| url | 发送相关后台方法的相对地址或绝对地址。 |
| data | 发送到后端服务器的数据，必须为Key/Value格式，后台自动将请求参数解析为字符串格式。 |
| type | 路由器请求后台方式，一般为POST和GET，默认为GET。随着REST编码风格的盛行，部分浏览器也开始支持PUT和DELETE。 |
| dataType | 后端服务器返回到前端的数据类型。可用值有xml、html、script、json、jsonp和text，一般开发环境中使用较多的是json和text两种类型。 |
| beforeSend | 发送请求前的事件处理。一般用于提高用户体验，在事件的处理结果返回前，弹出友好界面提示。 |
| success | 请求成功后的回调函数。 |
| error | 请求失败后的回调函数。 |

# 第三章 系统分析与设计

进入21世纪，大学生的学习方式变得越来越丰富，除了传统的学校教室外，PC端互联网学习环境也越来越成熟。借助手机、PAD在大学生群体中风靡之势，各大互联网学习平台也将目光延深到手机和PAD等移动媒介上，移动学习进入学习者的生活。本课题研究基于对现有互联网上学习平台现状分析总结，设计并开发以协同过滤算法为核心算法的学习资源个性化推荐系统CFSYSTEM，服务于天津师范大学的大学生群体，它也将顺应移动学习的浪潮，兼容PC、手机和PAD[23]。

## 3.1 常见学习资源系统分析

目前，网络学习平台众多，其中最具有代表性的是百度传课和网易云课堂。本课题研究过程中为了更加贴近天津师范大学的大学生群体，也重点分析了天津师范大学网络教学平台。以下是几大网络学习平台的详细介绍。

### 3.1.1百度传课

百度传课属于商业性质网络学习平台，它致力于打破地区教学资源不平衡，直播课是其特色。百度传课的用户按照自身需求购买相应课程，预约直播上课时间，进行学习。课程负责人可以通过实名认证后创建学校，招收互联网学习者，发布特色课程。



图3-1百度传课页面

该网络学习平台能够满足大多数用户的学习需求，但是并没有提供学习者之间的学习资源共享的功能，而且平台推荐的课程大多数是以课程播放量进行推送，不具有个性化特征。

### 3.1.2网易云课堂

网易云课堂属于商业性质的网络学习平台，它拥有海量的课程，涵盖IT互联网、英语四六级、小语种、茶艺、居家、母婴教育、亲子教育、中小学教育、高考辅导等等，课程教师也大多数为该领域名师。网易云课堂的用户根据自身需求购买相应的课程，在“我的学习”中，根据自身学习程度安排自主学习。网易云笔记，进度管理与学习监督和问答是网易云课堂的特色。



图3-2网易云课堂页面

该网络学习平台有一些课程，教师只是将学习资源上传到平台，真正的问答环节中教师没办法即使解决学生的疑问，时效性存在一定缺陷。在课程类别处理上，有一些课程定位不清，以至于用户到达该目录时无法找到相应课程，或者到达该目录发现完全不相干的课程。

### 3.1.3天津师范大学网络教学平台

天津师范大学网络教学平台属于公益性质网络学习平台，它凝聚了天津师范大学一批批优秀教师的教育生涯的结晶，课程网站上有网络课程，推荐课程和精品课程。天津师范大学用户根据自身专业选择相应课程，公告通知中会对课程发布情况进行即时说明。



图3-3天津师范大学网络教学平台主页面

该网络学习平台具有天津师范大学特色，对天津师范大学大学生群体在专业课程进修上有一定指导意义。然而，学习平台上的课程数量并不多，涵盖领域也并不全面，用户有些时候并不能找到与自身需求相符合的课程。推荐课程模块，只是简单的按照点击量进行推荐，并不具有个性化推荐特点。

## 3.2 用户需求分析

### 3.2.1 业务需求分析

系统业务主要分为两个部分。第一部分为学习资源共享业务，主要包括学习资源的浏览，上传，下载，评分，评论和删除，第二部分为学习资源的个性化推荐业务。学习资源共享业务作为基础平台为学习资源个性化推荐业务提供数据，助力协同过滤计算[33]。

天津师范大学学生处负责人指定系统管理员，系统管理员根据校办提供的新生名单，以学号为用户名，默认密码为用户名的方式将学生用户注册到系统中。用户登录系统，首页会为学生用户展示最新学习资源榜单和热门学习资源榜单（热门学习资源榜单的顺序依据学习资源的下载量与浏览量的总和从高到低排列）。系统设计了用户中心、日志中心、资源中心和管理中心用于支持学习资源共享业务和学习资源个性化推荐业务，。

### 3.2.2 性能需求分析

学习资源个性化推荐系统旨在为天津师范大学大学生群体提供一个高效便捷的学习资源个性化推荐环境，工程师评估用户数量以及学习资源存储所需容量，提供相应硬件支持。目前，天津师范大学全日制大学生据2016年官方统计为28785人，将这些大学生信息全部导入系统中，预估系统上同时在线用户数和资源操作情况。虽然用户数量相对于企业级互联网开发设定的用户数量而言并不算很多，但这也对系统性能提出了相当高的要求。下面将性能需求进行以下四个方面的分析。

（1）交互性：系统开发的最终目的是为天津师范大学学生用户提供学习资源共享和学习资源个性化推荐而服务。当今，苹果公司设计的IPHONE系列为互联网用户带来了高质的用户体验，这也使得互联网用户对交互式体验要求越来越高。工程师在系统的设计和开发过程中需要重视用户体验，故该系统在界面设计借助具有极佳用户体验的Bootstrap框架及其他优秀插件进行辅助，做到了对PC、手机和PAD完美兼容。

（2）安全性：近年来，随着互联网、大数据的爆发，导致了很多因为系统安全性缺陷造成的数据泄漏事件。其中熟为人知的包括，苹果的iCloud数据泄漏，Lindin用户数据泄漏，谷歌公司用户信息泄漏等等。数据安全已经成为工程师最关心的问题。基于当前互联网行业发展趋势，该系统对用户的隐私信息都进行的MD5混合算法加密，即使系统管理员也无法查询用户的隐私信息，确保数据安全。

（3）稳定性：天津师范大学2016官方统计在校全日制学生数量为28578人，作为系统设计者，必然要考虑高访问量极端情况下系统的负载能力。针对可能出现的高并发量，该系统使用nginx搭建WEB服务器，对HTTP请求进行方向代理实现负载均衡；数据请求端，使用Redis和MySQL混合使用，Redis进行数据缓存处理，MySQL生成本地数据文件；运用JAVA原生多线程编程技术，对并发进行处理，防止数据阻塞。

（4）健壮性：设计过程使用的是原型开发法，系统会随着用户需求的增加，对性能要求越来越高。所以，在系统运行的整个生命周期内，工程师都需要不断进行功能完善，迭代开发。基于此，系统的设计必须为后期扩展留下开发空间，以适应变化。

### 3.2.3 系统用例分析

CFSYSTEM的系统用例主要是消息用例，日志用例、管理用例和资源用例。其中，消息相关的用例有发布公告用例和查看公告用例。日志相关用例有用户日志用例和系统日志用例。管理用例中包含添加用户用例、批量添加用户用例、删除用户用例和批量删除用户用例。资源用例包括浏览资源用例、搜索资源用例、上传资源用例、下载资源用例、删除资源用例、发布评论用例、删除评论用例、最新资源推荐用例、热门资源推荐用例和个性化资源推荐用例。这些用例解释了CFSYSTEM的使用细节，如图3-4所示。

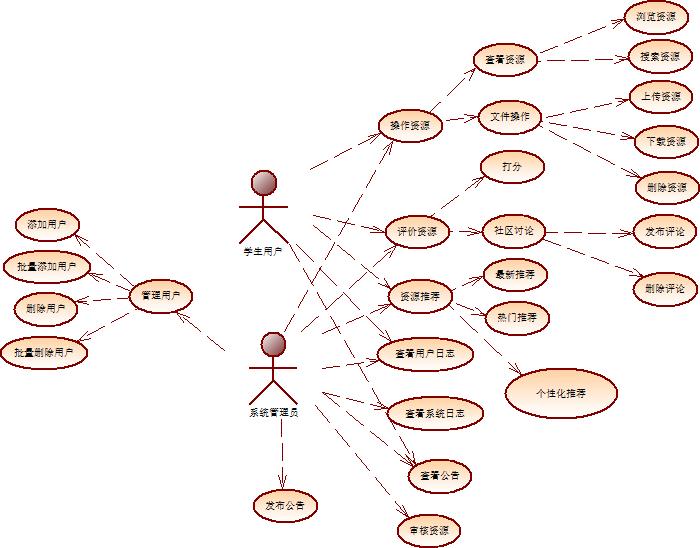


图3-4系统用例图

## 3.3 系统设计指导思想

MVC架构设计思想是当前互联网企业项目架构的主流思想，它将庞大的互联网项目解耦成各个相对独立的模块，团队开发中能够减少工程师间代码编程的冲突，提供团队开发效率，特别是在SCRUM敏捷开发团队中效果更加明显。MVC架构将系统分为了三个基本部分：模型（Model），视图（View）和控制器（Controller）。典型的JAVA EE项目开发过程中，模型对象将映射数据库设计中的实体字段，存储数据信息；视图是应用程序中处理数据显示的部分，通常从模型数据中创建视图，以Java Servlet Page形式展示；控制器是MVC架构中的核心部分，它连通模型和视图，当用户与系统的发生交互行为，系统通过交互信息判断用户选用的控制器，然后该控制器提取模型数据并转发给相应视图，控制器类型通常为Servlet。

三层架构设计思想与MVC架构思想一样，是当前互联网企业家头的主流思想，它们并不是相互独立的两种架设计思想，而是相互作用相互融合的两种思想。项目架构阶段也通常将它们融入项目架构体系中，配合使用。三层架构主要是数据层，业务逻辑层和显示层。系统中清晰的层级，能够让工程师在开发过程中更容易理解系统的架构，提高开发效率。

本次系统架构设计融入了MVC架构思想和三层架构思想，这样笔者在开发过程中思路更加清晰，明白当前阶段任务的完成情况以及下一个阶段需要完成的任务，IDEA IntelliJ 2016中系统的架构显示如图3-5所示。

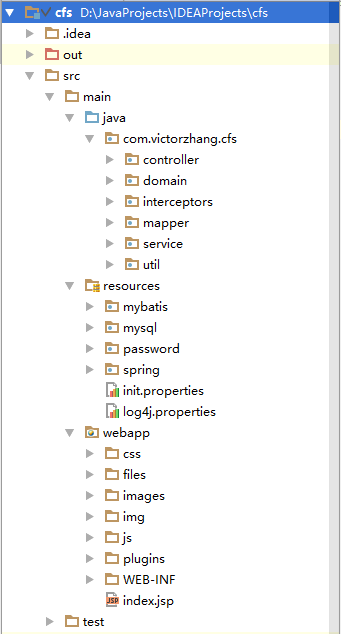


图3-5 IDEA IntelliJ 2016系统架构图

## 3.4 系统功能设计

CFSYSTEM共设计了四大中心模块，分别是：个人中心、资源中心、日志中心和管理中心。个人中心下包括用户信息、发布公告、最新公告和全部公告；资源中心下包括用户资源、系统资源、个性化推荐资源、用户评论、资源审核、最新资源排行榜和热门资源排行榜；日志中心下包括系统日志和用户日志；管理中心包括添加用户、批量添加用户、删除用户和批量删除用户。

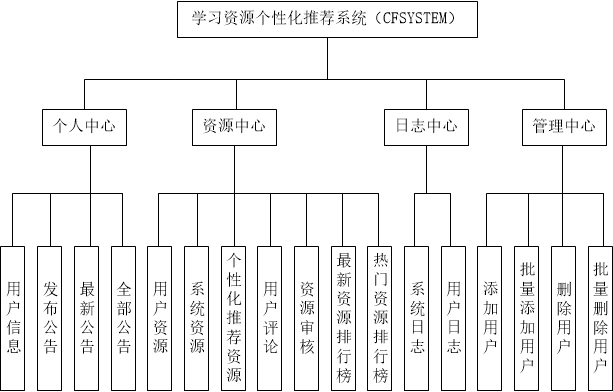


图3-6系统结构设计图

### 3.4.1 个人中心设计

用户被系统管理员注册进入系统后，需要完善个人用户信息。个人中心支持用户信息修改，用户根据提表单信息对真实姓名，身份证号码，手机号，邮箱地址，性别，个人标签等进行修改。其中，邮箱地址为防止用户忘记密码而设计，如果用户忘记密码需要将其找回就可以通过系统给用户信息中填写的邮箱发送找回密码链接，学生用户登陆邮箱查收邮件，点击系统发送邮件中提供的密码链接便可以找回密码。这里需要重点说明的是，邮件中的找回密码的链接是经过MD5加密处理，保证用户隐私安全。发布公告业务的权限设置为系统管理员，系统管理员成功登录系统后，可以对系统所有用户发布公告消息，该业务设计的目的可以作为系统维护使用。公告栏的功能是显示系统管理员发布的最新公告小心，历史公告消息可以通过首页页面顶层的消息按钮查询，并且该消息按钮会提示用户未读消息的数量。当用户点击消息按钮时，页面显示最新10条未读消息的列表以及全部消息查看按钮，点击全部消息查看便可进入历史消息页面。

### 3.4.2 资源中心设计

用户登录系统后，可以根据资源名称、资源类型、开始日期和结束日期四个限定条件查询和删除自己上传至系统的所有学习资源，，对于共享学习的环境，上传、下载业务也必不可少。对于系统中所有的学习资源，用户也可以根据资源名称、开始日期和结束日期来查询，也可以对其进行评分、下载和查看，需要指出的是评分功能需要跳转到学习资源详情页面进行操作。当系统收集了当前用户的数据后，会进行协同过滤推荐计算，将计算结果推荐给当前用户。对于评论信息，用户可以按资源名称、开始日期和结束日期来查询和删除。学习资源成功上传至系统需要经过系统管理员的审核，系统管理员根据资源名称、开始日期和结束日期查询待审核的学习资源，根据对其详细情况的判断，决定是否能够通过。

### 3.4.3 日志中心设计

系统对用户的隐私安全十分关注，用户的登录、注销、更改学习资源信息、更新用户信息等关键操作都会被记录到日志中。当用户发现自己的账号存在异常情况时，通过查看日志即可发现异常操作的类型、时间、具体说明以及IP地址。系统管理员对日志拥有最高权限，他不仅能够查看自己的日志信息，而且能够查看系统中所有用户的日志信息，这也主要是为了系统管理员对系统使用情况进行数据统计。

### 3.4.4 管理中心设计

系统管理员拥有添加用户、批量添加用户、个别删除用户和批量删除用户的权限。天津师范大学每年9月的新生入学，学校都会为每一名学生提供学号，而所有学号有一个特征就是它们都连续，基于此设计批量添加和批量删除功能，提高系统管理员工作效率。目前，将Excel导入功能的新需求写入了计划表，后期依据系统功能实现情况决定是否需要实现。如果加入Excel导入功能的话，系统管理员可以直接将Excel表格数据导入至数据库。

## 3.5 推荐算法设计

CFSYSTEM中设计四个子模块的学习资源推荐场景。用户登录系统进入首页，页面将显示最新学习资源榜单和热门学习资源榜单为其推荐学习资源。在学习资源个性化推荐模块中，后台会根据当前用户对所有学习资源打分的历史记录，用户浏览学习资源的历史记录，以及用户下载学习资源的历史记录为其个性化推荐学习资源，这部分也是本课题研究的重点。然而，系统设计时，考虑用户可能没有对任何学习资源进行打分，没有下载过任何学习资源，甚至没有浏览过任何学习资源，从而无法为系统个性化推荐提供个人信息数据的情况，设计了用户-项目标签匹配功能来推荐相应的学习资源。还有一点则是用户对某一学习资源进行了评分，而系统其他用户均未对该学习资源进行评分，导致用户数据稀疏无法被个性化推荐。笔者对此提出添加用户下载学习资源记录和用户阅读学习资源记录作为协同过滤算法的数据进行个性化推荐。

学习资源个性化推荐模块主要采取了基于用户的协同过滤算法和基于项目的协同过滤算法作为主要推荐算法，其中，对协同过滤推荐算法的数据输入做了严格的层级限制[24]。本系统的输入终端分别是用户浏览记录矩阵，用户下载记录矩阵和用户评分矩阵。

用户浏览矩阵

用户下载矩阵

用户打分矩阵

基于用户的协同过滤算法

基于项目的协同过滤算法

用户推荐列表矩阵

图3-7推荐算法结构图

（1）基于用户的协同过滤算法和基于项目的协同过滤算法是协同过滤算法体系中应用最多的算法[25]，它们往往被混合起来使用。基于用户的协同过滤算法在系统面临新项目冷启动问题时，受到的影响较小，这是因为基于用户的协同过滤算法会通过计算找出近邻用户，参考近邻用户的偏好度，为当前用户推荐项目[26]。相反，基于项目的协同过滤算法在系统面临新用户冷启动问题时，受到的影响较小，这是因为基于项目的协同过滤推荐算法会通过计算出近邻项目，为当前用户推荐项目。如果同时出现新用户冷启动和新项目冷气问题，参考最新学习资源榜单、热门学习资源榜单和用户标签。

（2）用户浏览记录矩阵和用户下载记录矩阵式是为了防止数据稀疏问题而设计。在实际应用中，有很多用户对学习资源的打分比较随意，或者根本很少对学习资源打分，这对系统进行学习资源个性化推荐产生很大的影响。为了解决该问题，系统收集用户浏览项目的信息和用户下载项目的信息，最终形成用户浏览记录矩阵和用户下载记录矩阵。这些数据相对评分数据更容易获得，也能够在一定程度上代表用户的偏好属性，这对减少数据稀疏问题有很大帮助。

## 3.6 数据库设计

### 3.6.1 数据实体设计

（1）用户实体描述。用户实体是对系统当前所有用户的信息的描述，并保存这些信息。需要注明的是，考虑到用户的隐私安全，系统将密码和用于找回密码的随机数做了MD5加密，即使用户管理员也无法获得明文信息。用户实体的属性信息如图3-8所示。

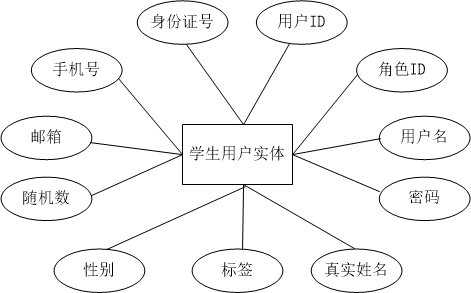


图3-8用户实体属性图

（2）角色实体描述。角色实体是对系统当前所有权限角色的信息的描述，并保存这些信息。角色实体的属性信息如图3-9所示。



图3-9角色实体属性图

（3）日志实体描述。日志实体是对系统当前所有日志的信息的描述，并保存这些信息。日志实体的属性信息如图3-10所示。

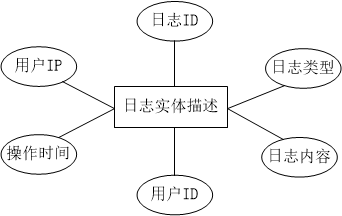


图3-10日志实体属性图

（4）消息实体描述。消息实体是对系统当前所有消息的信息的描述，并保存这些信息。消息实体的属性信息如图3-11所示。

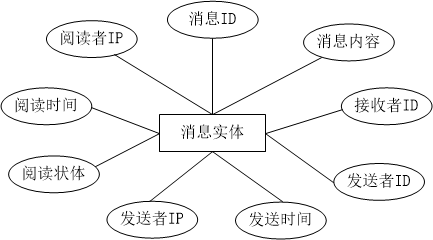


图3-11消息实体属性图

（5）资源实体描述。资源实体是对系统当前所有资源的信息的描述，并保存这些信息。资源实体的属性信息如图3-12所示。

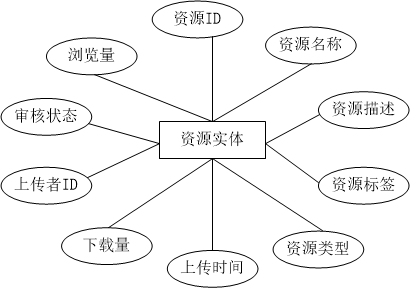


图3-12资源实体属性图

（6）资源评论实体描述。资源评论实体是对系统当前所有资源评论信息的描述，它保存了所有资源评论的信息。资源评论实体的属性信息如图3-13所示。

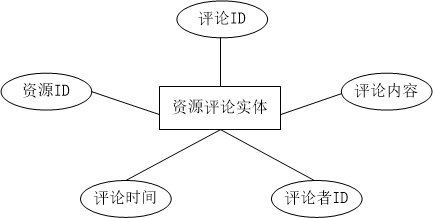


图3-13资源评论实体属性图

（7）资源下载记录实体描述。资源下载记录实体是对系统当前所有资源下载记录的信息的描述，它保存了所有资源下载记录的信息。资源下载记录实体的属性信息如图3-14所示。

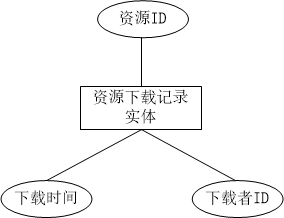


图3-14资源下载记录实体的属性图

（8）资源浏览记录实体描述。资源浏览记录实体是对系统当前所有资源浏览记录的信息的描述，它保存了所有资源浏览记录的信息。资源浏览记录实体的属性信息如图3-15所示。

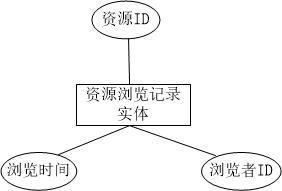


图3-15资源浏览记录实体的属性图

（9）资源评分实体描述。资源评分实体是对系统当前所有资源评分的信息的描述，并保存这些信息。资源评分实体的属性信息如图3-16所示。

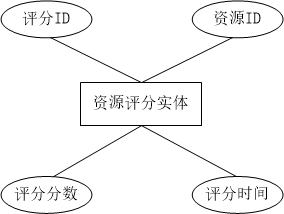


图3-16资源评分实体的属性图

### 3.6.2 E-R图设计

E-R图，又名实体-联系图，它用来描述系统中存在的各个实体间相互联系的关系图。在CFSYSTEM中，实体数量相对较多，实体间的联系也相对复杂。鉴于此，在本小节中只对系统中的学习资源个性化推荐模块进行了E-R图的绘制工作，如图3-17所示。

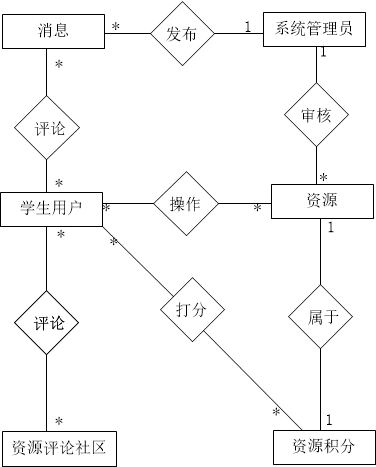


图3-17学习资源个性化推荐模块E-R图

### 3.6.3 数据库表结构设计

CFSYSTEM服务于的是天津师范大学大学生特定用户群体，对于提供的学习资源个性化推荐功能进行详细分析后，设计如下数据库方案。数据库方案设计的有两条重要原则，一是保证数据表设计的合理性，二是有效利用物理存储空间。这里，将重点介绍CFSYSTEM的MySQL数据库的表格设计。笔者在设计过程根据实际生活常识和以往企业级开发经验对表格字段的数据类型、字符长度，主键设置，非空处理做了严格的控制。

（1）学生用户user表。user表中存储系统用户信息，id为主键，是以UUID生成的32位大写英文字母或者数字，作为唯一标识符。如表3-1所示。

表3-1 学生用户user表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名称 | 数据类型及长度 | 主键 | 非空 | 文字说明 |
| id | char(32) | Y | Y | 用户id |
| role\_id | char(32) | N | Y | 角色id |
| username | varchar(40) | N | Y | 用户名 |
| password | char(32) | N | Y | 登陆密码  （MD5加密） |
| realname | varchar(40) | N | Y | 真实名字 |
| user\_idcard | char(18) | N | N | 用户身份证 |
| user\_mobile | varchar(11) | N | N | 用户手机 |
| user\_email | varchar(32) | N | N | 用户邮箱 |
| random\_code | varchar(32) | N | N | 邮箱激活验证码 |
| gender | tinyint(1) | N | N | 性别  （0男；1女） |
| tag | varchar(40) | N | N | 用户标签 |
| gmt\_create | datetime | N | Y | 记录创建时间 |
| gmt\_modify | datetime | N | N | 记录修改时间 |

（2）角色role表。role表中存储系统角色信息，id为主键，是以UUID生成的32位大写英文字母或者数字，作为唯一标识符。如表3-2所示。

表3-2 角色role表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名称 | 数据类型及长度 | 主键 | 非空 | 文字说明 |
| id | char(32) | Y | Y | 角色id |
| roledname | varchar(40) | N | Y | 角色名称 |
| gmt\_create | datetime | N | Y | 记录创建时间 |
| gmt\_modify | datetime | N | N | 记录修改时间 |

（3）日志log表。log表中存储系统日志信息，id为主键，是以UUID生成的32位大写英文字母或者数字，作为唯一标识符。如表3-3所示。

表3-3 日志log表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名称 | 数据类型及长度 | 主键 | 非空 | 文字说明 |
| id | char(32) | Y | Y | 日志id |
| log\_type | varchar(40) | N | Y | 日志类型 |
| log\_content | varchar(200) | N | Y | 日志内容 |
| user\_id | char(32) | N | Y | 日志操作用户id |
| user\_date | datetime | N | Y | 日志操作时间 |
| user\_ip | varchar(15) | N | Y | 日志操作用户ip |

（4）消息message表。message表中存储系统消息信息，id为主键，是以UUID生成的32位大写英文字母或者数字，作为唯一标识符。如表3-4所示。

表3-4 消息message表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名称 | 数据类型及长度 | 主键 | 非空 | 文字说明 |
| id | char(32) | Y | Y | 消息id |
| msg\_content | text | N | N | 消息内容 |
| receive\_user\_id | char(32) | N | Y | 消息接收者id |
| send\_uesr\_id | char(32) | N | Y | 消息发送者id |
| send\_time | datetime | N | Y | 发送时间 |
| send\_user\_ip | varchar(15) | N | Y | 消息发送者ip |
| is\_read | tinyint(1) | N | Y | 阅读状态 |
| read\_time | datetime | N | N | 阅读时间 |
| read\_ip | varchar(15) | N | N | 阅读者ip |

（5）资源resource表。resource表中存储资源信息，id为主键，是以UUID生成的32位大写英文字母或者数字，作为唯一标识符。如表3-5所示。

表3-5 资源resource表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名称 | 数据类型及长度 | 主键 | 非空 | 文字说明 |
| id | char(32) | Y | Y | 资源id |
| resource\_name | varchar(40) | N | Y | 资源名称 |
| res\_description | text | N | N | 资源描述 |
| resource\_tag | varchar(200) | N | N | 资源标签 |
| resource\_type | tinyint(4) | N | Y | 资源类型 |
| uploadtime | datetime | N | Y | 上传时间 |
| download\_count | int | N | Y | 下载量 |
| browse\_count | int | N | Y | 阅读量 |
| server\_path | varchar(200) | N | Y | 资源存储地址 |
| user\_id | char(32) | N | Y | 上传者id |
| verify\_type | tinyint(4) | N | Y | 审核状态 |
| gmt\_create | datetime | N | Y | 记录创建时间 |
| gmt\_modify | datetime | N | N | 记录修改时间 |

（6）资源评论resource\_comment表。resource\_comment表中存储系统评论信息，id为主键，是以UUID生成的32位大写英文字母或者数字，作为唯一标识符。如表3-6所示。

表3-6 资源评论resource\_comment表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名称 | 数据类型及长度 | 主键 | 非空 | 文字说明 |
| id | char(32) | Y | Y | 评论id |
| comment\_content | text | N | N | 评论内容 |
| user\_id | char(32) | N | Y | 评论者id |
| comment\_time | datetime | N | Y | 评论时间 |
| resource\_id | char(32) | N | Y | 评论资源id |

（7）资源下载resource\_download表。resource\_download表中存储系统资源下载信息，id为主键，是以UUID生成的32位大写英文字母或者数字作为唯一标识符。如表3-7所示。

表3-7 资源下载resource\_download表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名称 | 数据类型及长度 | 主键 | 非空 | 说明 |
| id | char(32) | Y | Y | 资源评论id |
| resource\_id | char(32) | N | Y | 资源id |
| download\_time | datetime | N | Y | 下载时间 |
| user\_id | char(32) | N | Y | 下载者id |

（8）资源浏览resource\_browse表。resource\_browse表中存储系统资源浏览信息，id为主键，是以UUID生成的32位大写英文字母或者数字作为唯一标识符。如表3-8所示。

表3-8 资源浏览resource\_browse表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名称 | 数据类型及长度 | 主键 | 非空 | 文字说明 |
| id | char(32) | Y | Y | 资源浏览id |
| resource\_id | char(32) | N | Y | 资源id |
| browse\_time | datetime | N | Y | 浏览时间 |
| user\_id | char(32) | N | Y | 浏览者id |

（8）资源评分resource\_score表。resource\_score表中存储系统资源评分信息，id为主键，是以UUID生成的32位大写英文字母或者数字作为唯一标识符。如表3-9所示。

表3-9 资源评分resource\_score表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名称 | 数据类型及长度 | 主键 | 非空 | 文字说明 |
| id | char(32) | Y | Y | 资源评分id |
| user\_id | char(32) | N | Y | 用户id |
| resource\_id | char(32) | N | Y | 资源id |
| rating | int | N | Y | 评分 |
| ratingtime | datetime | N | Y | 评分时间 |

# 第四章 系统实现与测试

## 4.1 系统实现总体过程

系统的开发分为一下三个阶段：

第一阶段：确定系统的总体功能。论文的选题后，笔者查阅了大量文献，仔细分析互联网现有的学习资源个性化推荐系统中存在的问题，进而选定协同过滤算法为核心算法，并对其应用到CFSYSTEM中能够产生的效果进行实验预测。

第二阶段：确定系统各个分模块的细节。CFSYSTEM目的在于解决天津师范大学大学生群体面临的学习资源“信息过载”问题，各个分模块都将为学习资源个性化推荐模块提供数据服务，所以严格地对各个模块之间的数据交换进行架构分层，并最终确定使用MVC三层架构模式。数据库为Redis和MySQL混合使用，数据库连接池采用Alibaba在GitHub分享的Durid。系统后台框架为Spring、SpringMVC、MyBatis，编程语言主要为JAVA，前端使用Bootstrap响应式布局。

第三阶段：确定系统各个模块使用的技术，主体编程语言与应用平台。在第三章—系统总体设计中，CFSYSTEM采用Bootstrap进行界面响应式处理，兼容PC、手机和PAD平台。前端使用Jquery（JS库的一种）+ Ajax完成页面局部刷新和异步响应，采用JSON数据格式完成前后台数据交换。找回密码模块使用JAVA EE原生JAVA MAIL支持开发，日志模块采用Apache开源社区Log4j框架支持开发，学习资源上传模块采用Apache开源社区里轻量级commons-fileupload库支持上传，学习资源个性化推荐模块中，采用Pearson系数计算相似度，协同过滤算法计算推荐集合，等等。CFSYSTEM融合了软件工程理论、数学体系中相关算法和现代远程教育思想，克服开发中遇到的各个难点，突破技术障碍，并最终实现自身功能。

## 4.2 系统核心模块实现

CFSYSTEM主要组织结构设立了四大中心，分别为用户中心、资源中心、日志中心和管理中心。

用户中心，导航列表中包含用户信息，发布公告和公告栏。点击用户信息，进入用户信息模块，页面上依次显示当前用户的所有信息，能够修改的信息在对应的选项中可以进行操作，禁止修改的信息在对应的选项中不可以进行操作，例如角色选项。重点需要解释的是，页面中的用户标签选项，它是设计用来辅助解决协同过滤算法中存在的冷启动问题，即当前用户在系统中还没有进行交互操作，历史数据为空，此时系统会根据提供的用户标签，查找合适学习资源，按照热门程度排序为用户个性化推荐；点击发布公告，进入发布公告模块（权限仅为系统管理员），页面显示文本编辑器以及发布按钮，输入需要发布的公告信息，点击发布按钮即可发布公告；点击公告栏，进入公告栏模块，页面显示当前发布的最新公告信息，当然，用户如果想查询所有的公告信息则需要进入消息模块中去查询。

日志中心，它主要针对用户信息安全，系统便于管理而设计，导航列表包含用户日志和系统日志。点击用户日志，进入用户日志模块，页面显示当前用户日志信息，包括操作类型、操作说明、操作日期、操作的IP地址，以表格形式；点击系统日志，进入系统日志模块（权限仅为系统管理员），页面显示系统所用用户日志信息。

管理中心（权限仅为系统管理员），导航列表包含添加用户、批量添加用户、删除用户和批量删除用户。系统管理员根据学校相关部门下发的指示，将天津师范大学相关学生用户信息录入系统中，考虑到系统默认用户名和密码均为学生学号，学号的一个特征就是号码连续性，故在管理中心设计并实现用户批量添加模块。然而，有时候需要添加特定用户时，又设计并实现个别用户添加模块；个别用户删除模块和用户批量删除模块是系统管理员参考系统日志信息后，对于个别或者一些学生用户发布敏感信息，屡教不改，会将其从系统中删除。

最后，资源中心的各个模块为本课题的核心模块，将分为以下5个小节进行详细介绍。

### 4.2.1 登录模块

用户输入CFSYSTEM的网站地址进入登陆页面，此时需要输入用户名和密码进行登陆验证，若用户名和密码与数据库中记录相一致，验证成功进入系统，否则验证失败显示友好提示信息无法进入系统。登陆模块的核心代码如下：

|  |
| --- |
| **//用户登陆核心代码**  **public** User doLoginByUsernameAndPassword(String username, String password, HttpServletRequest request) **throws** Exception {  **//后台验证用户名和密码是否为空**  **if** (StringUtils.*isNotEmpty*(username) && StringUtils.*isNotEmpty*(password)) {  **//根据用户名和密码获取用户对象**  User userObj = **new** User(username, **new** MD5Utils().getMD5ofStr(password));  ***logger***.info(username + ***TRY\_LOGIN***); **//加入Log4j日志**  User user = **userMapper**.get(userObj);  **if** (user != **null**) {  request.getSession().setAttribute(***USER\_ID***, user.getId());**//USER\_ID加入session中**  request.getSession().setAttribute(***ROLE\_ID***, user.getRoleId());**//ROLE\_ID加入session中**  **//将日志加入数据库中**  **logService**.saveLogByLogTypeAndLogContent(***LOGIN\_SYSTEM***, request.getHeader(***USER\_AGENT***));  ***logger***.info(username + ***LOGIN\_SUCCESS***); **//加入Log4j日志**  **return** user;  }  ***logger***.error(user + ***LOGIN\_FAIL***); **//加入Log4j日志**  }  **return null**; } |

登录成功后进入系统首页，如图4-1所示。页面顶端分别显示了系统Logo，当前用户消息提示、用户名、以及退出按钮。页面左端的导航栏分别显示了快速导航、个人中心、资源中心、日志中心和管理中心。页面右端主内容区显示了四大中心的各个子业务模块，分别用不同颜色的磁贴来显示以及最新学习资源榜单和热门资源榜单。

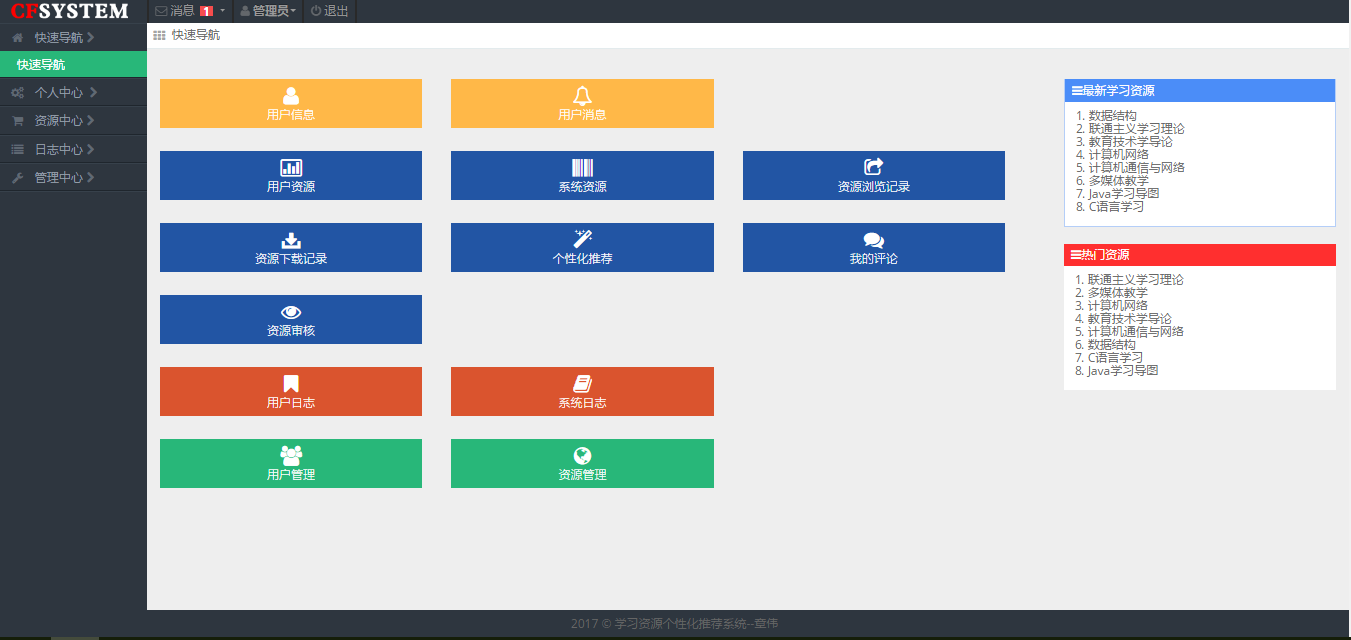


图4-1系统主页面

### 4.2.2 最新学习资源推荐模块

最新学习资源榜单在一定程度上能够缓解学习资源个性化推荐模块中冷启动的问题。当系统中用户数量或学习资源数量尚不足够多的时候，协同过滤算法不能准确地计算用户间，用户与学习资源间的相似度，进而无法对当前用户推荐相应的学习资源。此时，用户可以根据最新最细资源榜单侧面了解CFSYSTEM上用户的最新学习趋势，如果发现最新学习资源榜单中的某项学习资源满足自身的学习兴趣，用户可以点击该学习资源名称进入该项学习资源的详细信息页面。用户从本地上传学习资源的时候，数据库resource表中存储该学习资源的各项数据同时会自动填充gmt\_create字段，值为当前时间的时间戳。最新学习资源榜单列出最新学习资源的时候将按照gmt\_create倒序排列的方式进行显示，核心代码为：

|  |
| --- |
| @RequestMapping(**"/listNewestResource.do"**) @ResponseBody  **public** Map<String, Object> listNewestResource() **throws** Exception {  Map<String, Object> map = **new** HashMap<>();  **//获取最新学习资源列表数据**  map.put(**"data"**, **resourceService**.listNewestResource());  **return** map; }  **//数据库端SQL语句，按照时间顺序返回前8项资源id和资源名称**  <**select id="listNewestResource" resultType="java.util.Map"**>  SELECT id, resource\_name FROM resource  WHERE verify\_type = ‘1’ **//审核已通过**  ORDER BY gmt\_create DESC LIMIT 8 </**select**> |

用户进入系统后，主页显示最新学习资源榜单，榜单上列举了最新的八条学习资源记录。最新学习资源榜单如图4-2所示。



图4-2最新学习资源推荐页面

### 4.2.3 热门学习资源推荐模块

热门学习资源榜单与最新学习资源榜单一样能够在一定程度上缓解学习资源个性化推荐模块中冷启动的问题。当系统中用户数量或学习资源数量尚不足够多的时候，协同过滤算法不能准确地计算出用户间，用户与学习资源的相似度，进而无法对当前用户推荐相应的学习资源。此时，用户可以根据热门学习资源来选取适合自己的学习资源。从心理学的研究分析中得知，大多数人都有随众心理，这也是热门学习资源推荐模块设计的理论依据。热门学习资源榜单列表中的学习资源是按照用户的下载量和浏览量的数量总和进行排序显示。在第三章数据库设计中，resource表中包含了两个字段分别是download\_count和browse\_count，它们分别代表下载量和浏览量数据。当用户对学习资源进行下载操作时，根据resource\_id更新download\_count字段，数值加1，而browse\_count字段会在用户进入学习资源详情页面触发，根据resource\_id进行自动更新，数值加1。每次当用户进入首页时，后台会对数据库resource表中的download\_count和browse\_count做加法运算，根据得到的数值，进行热门学习资源榜单排序。考虑到用户有可能频繁的切换到首页，而且是多用户并发操作时，后台的计算有可能导致系统崩溃。鉴于此，对首页进行缓存处理，设置一个合适时间值，在这个时间值内，首页的热门学习资源推荐将取Redis数据库中数据，超过这个时间值将重新对MySQL数据库进行读写，这样既提高了用户体验，也减轻了服务器压力。热门学习资源的核心代码为：

|  |
| --- |
| @RequestMapping(**"/listHotResource.do"**) @ResponseBody **public** Map<String, Object> listHotResource() **throws** Exception {  Map<String, Object> map = **new** HashMap<>();  **//获取热门学习资源列表数据**  List<Map<String, Object>> list = **resourceService**.listHotResource();  **//如果资源id或资源名称为null，则返回前台null**  **if** (list.get(0).get(**"id"**) == **null** || list.get(0).get(**"resource\_name"**) == **null**) {  **return null**;  }  map.put(**"data"**, list);  **return** map; }  **//数据库端SQL语句，按照浏览量和下载量总和大小的顺序返回前8项资源id和资源名称**  <**select id="listHotResource" resultType="java.util.Map"**>  SELECT id, resource\_name FROM resource  WHERE verify\_type = ‘1’ **//审核已通过**  GROUP BY id  ORDER BY *SUM*(resource\_download\_count+resource\_browse\_count) DESC  LIMIT 8 </**select**> |

用户进入系统后，主页显示热门学习资源榜单，榜单上列举了热门的八条学习资源记录。热门学习资源榜单如图4-3所示。



图4-3热门学习资源推荐页面

### 4.2.4 学习资源审核模块

学习资源审核模块相对比较特殊，它是为了防止系统中学习资源出现敏感信息而设计的，权限为系统管理员。用户上传学习资源后，并不能立即发布到系统中，而是进入系统管理员的学习资源审核模块中，此时系统管理员查看上传的学习资源是否符合上传要求，通过审核或者拒绝。学习资源上传至服务器并在数据库resource插入该条记录，verify\_type字段自动填充0（0代表待审核，1代表审核通过，2代表审核不通过），这些verify\_type字段值为0和2的记录无法发布到系统中。同时，需要说明的一点是，系统中所有展示的学习资源均为系统管理员通过审核的资源。所以，考虑到一个系统管理员可能无法及时处理如此多学习资源的审核，系统将设立多个系统管理员来协同工作。学习资源审核模块的核心代码为：

|  |
| --- |
| @RequestMapping(**"/listVerifyResource.do"**) @ResponseBody **//显示管理员需要审核的学习资源，参数\_page和\_pageSize用于分页，其他参数用于查询**  **public** Map<String, Object> listVerifyResource(String \_page, String \_pageSize, String resourceName, String resourceType, String startDate, String endDate) **throws** Exception {  **//后台验证该权限为系统管理员****if** (StringUtils.*equals*(CommonUtils.*sesAttr*(**request**, ***ROLE\_ID***), ***ADMIN\_ROLE\_ID***)) {  **//获取资源对象**  Resource resource = **new** Resource(resourceName, resourceType, ***RESOURCE\_VERIFYING***);  **return resourceService**.listPaging(resource, \_page, \_pageSize, startDate, endDate, **null**);  }  **throw new** IllegalAccessException(***NO\_ACCESS\_PERMISSION***); **//抛出权限异常** } |

系统管理员进入学习资源审核模块，如图4-4所示。学习资源审核页面会依次显示所有待审核学习资源的详细信息，默认按照上传时间的先后顺序进行排序。页面顶端有三个搜索限定条件，分别为资源名称，开始日期和结束日期，系统管理员按照个人需求可以快速定位到待审核学习资源。待审核学习资源表格中从左到右依次显示学习资源名称，学习资源描述信息，类型，上传用户，上传时间，只有当学习资源符合系统上传的要求时则才通过审核，否则将拒绝学习资源上传。



图4-4学习资源审核页面

### 4.2.5 我的评论模块

我的评论模块是系统为所有用户提供交流的地方，当用户下载并使用了系统中特定的学习资源后，想要对其发表感想，便可以将之写入该学习资源的评论中。

用户对某个学习资源进行评论后，评论信息会存储到数据库的resource\_comment表中，表中记录评论内容、评论人id、资源id以及评论时间。这些评论信息将为其他浏览到该学习资源或准备下载该学习资源的用户借鉴。我的评论模块核心代码为：

|  |
| --- |
| @RequestMapping(**"/listPaging.do"**) @ResponseBody  **//显示评论信息，参数\_page和\_pageSize用于分页，其他参数用于查询** **public** Map<String, Object> listPaging(String \_page, String \_pageSize, String resourceName, String resourceType, String startDate, String endDate) **throws** Exception {  **//获取评论对象**   Comment comment = **new** Comment(CommonUtils.*sesAttr*(**request**, ***USER\_ID***));  **//加入搜索参数，封装成GenericQueryParam对象**  GenericQueryParam param = **new** GenericQueryParam();  param.fill(**"resourceName"**, resourceName);  param.fill(**"resourceType"**, resourceType);  **return commentService**.listPaging(comment, \_page, \_pageSize, startDate, endDate, param); } |

用户进入我的评论模块，页面内容如图4-5所示。我的评论页面显示的评论列表均为当前用户的评论信息，默认按照当前用户评论时间的先后顺序进行排序。页面顶端有三个搜索限定条件，分别是资源名称，开始日期和结束日期，这样当前用户按照个人需求便可以快速定位到特定的评论信息。评论信息表格从左到右依次显示资源名称，评论时间和评论内容。当前用户如果想再次查看该学习资源评论的详细信息，点击对应所在行的详情按钮，即可跳转到详细页面。当前用户如果认为对该学习资源的评论不合适也可点击对应所在行的删除操作按钮，即可删除对该学习资源的评论内容。

图4-5我的评论页面

### 4.2.6 学习资源个性化推荐模块

学习资源个性化推荐模块是系统的重点模块，它是由两个子模块组成。一个是打分子模块，另一个是个性化推荐列表子模块。

（1）学习资源打分子模块。当用户进入学习资源详细页面，根据其他用户对该学习资源的评论信息以及自身的使用体验，对学习资源进行打分（学习资源评分为5分制）。用户评分后，评分信息会存储到数据库的resource\_score表中，表中记录评分人id，学习资源id，分值以及评分时间。resource\_score表中的记录将为协同过滤算法提供数据支持。学习资源打分模块核心代码为：

|  |
| --- |
| **//存储学习资源评分信息**  **private void** saveResourceScore(HttpServletRequest request, String resourceId, String rating) **throws** Exception {  String resourceScoreId = CommonUtils.*newUuid*();**//获取学习资源评分id**  String resourceId = resourceId; **//赋值学习资源id**  String rating = rating; **//赋值评分数值**  String ratingTime = CommonUtils.*getDateTime*();**//获取评分时间**  String userId = CommonUtils.*sesAttr*(request, ***USER\_ID***); **//获取评分人id**  **//插入数据库中**  **int** successCount = **resourceMapper**.saveResourceDownload(**new** ResourceScore(resourceScoreId, resourceId, rating, ratingTime, userId));  **if** (successCount <= 0) {  **throw new** SQLException(***INSERT\_ERROR***); **//抛出数据库异常**  } } |

用户进入学习资源打分页面，如图4-6所示。系统设计的打分为5分制，以空心五角星进行表示，鼠标从左往右划，划过的五角星会从空心变成实心，鼠标从右往左划，划过的五角星会从实心变成空心，实心五角星的个数就代表当前用户所打的分数。页面中间显示的是之前所有用户对该学习资源进行评论的信息，页面最下面也有输入框供用户进行评论，输入的评论信息点击提交后会显示到当前页面评论区的最上端。



图4-6学习资源打分页面

（2）学习资源个性化推荐列表子模块。三种情况的分析，第一，当系统的用户数量相对较少的时候，基于用户的协同过滤算法并不能十分准确的推荐个性化的学习资源给当前用户时，采用基于项目的协同过滤算法；第二，当系统的项目数量相对较少的时候，基于项目的协同过滤算法并不能十分准确的推荐个性化的学习资源给当前用户时，采用基于用户的协同过滤算法；第三，当系统用户和项目达到一定数量时，不可避免的会产生数据稀疏问题，采用学习资源下载矩阵和学习资源浏览矩阵进行学习资源的个性化推荐。本子模块的核心代码依据第二章的2.3节和2.4节的理论实现，这里不做重复论述。主要公式为：

用户成功登录系统后进入学习资源个性化推荐子模块，页面显示的推荐列表为系统根据当前用户的历史评分数据进行协同过滤算法获得。系统为了避免冷启动问题，采用了用户标签与学习资源标签匹配的方式进行个性化推荐，用户贴标签即在用户信息模块中，用户可以添加个人标签，代表个人兴趣爱好。另外，为了避免数据稀释问题，系统采用了学习资源浏览矩阵和学习资源下载矩阵的方式，为系统提供数据。对于新用户冷启动问题和新项目冷启动问题，系统将基于用户的协同过滤算法和基于内容的协同过滤算法混合起来，进行计算。学习资源个性化推荐列表中从左到右依次显示学习资源名称，学习资源描述，类型，上传用户，上传时间。根据这些信息，当前用户判断推荐的学习资源是否确实符合个人兴趣，若符合，则进行下载，若还想进一步观察可点击该学习资源所在行的详情按钮进入详情页面了解。学习资源个性化推荐列表如图4-7所示。



图4-7学习资源个性化推荐列表

## 4.3 系统运行测试

### 4.3.1 单元测试

在软件工程领域，单元测试作为白盒测试的一种常用方法，指的是工程师对系统中最小可测试单元编写测试代码进行验证[27]。当前，互联网行业中流行的测试驱动开发模式TDD，单元测试就是其主要的测试手段。JAVA WEB项目中，工程师进行单元测试时使用最多的是Kent Beck和Erich Gamma创建的Junit框架。

系统在开发过程中使用的是编辑器是IntelliJ IDEA 2016，项目构建工具为Apache社区的Maven 3.1.1。为了提高工程师的开发效率，通常会将Maven3.1.1整合到IntelliJ IDEA2016中，集成之后的项目代码结构如图4-8所示src/main/java中存放项目功能代码，src/main/resources中存放项目资源文件，src/main/webapp中存放web项目文件，如html文件、javascript文件、typescript文件、css文件、jsp文件等，而src/test/java和src/test/resources中存放的则是项目的功能测试文件和资源测试文件。与src/man中文件命名不同，src/test中文件命名往往在文件名结尾加Test，以作区分。

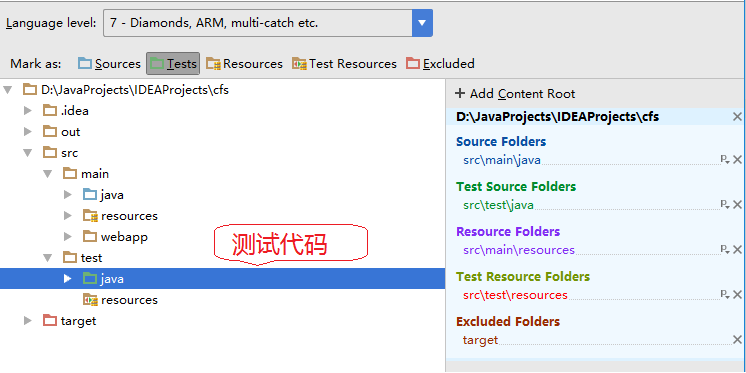


图4-8项目代码结构

笔者在CFSYSTEM的开发过程中，对用户登录、学习资源上传、发送邮件，学习资源下载等都进行了严格的单元测试。这里仅贴出用户登录的JUnit完整测试代码，以供参考。

|  |
| --- |
| **public class** UserTest {   @Autowired  @Qualifier(**"userService"**)  **private** UserService **userService**; **//注入UserService**   **private** MockHttpServletRequest **request**; **//Mock模拟HttpServletRequest**  **private** MockHttpServletResponse **response**; **//Mock模拟HttpServletResponse**   @Before  **public void** before() {**//测试进行前执行的方法**  **request** = **new** MockHttpServletRequest();**//实例化MockHttpServletRequest**  **request**.setCharacterEncoding(**"UTF-8"**);**//设置请求编码格式为UTF-8**  **response** = **new** MockHttpServletResponse();**//实例化MockHttpServletResponse**  }   @Test  **public void** testGetUserByUsernameAndPassword() **throws** Exception {**//测试登录**  **//doLoginByUsernameAndPassword参数分别为用户名和密码**  **userService**.doLoginByUsernameAndPassword(**"1410030164"**, **"1410030164"**, **request**);  } } |

### 4.3.2 功能测试

在软件工程领域，功能测试作为功能测试的一种常用方法，指的是对系统开发工作中已经完成的功能模块进行用户层面的测试，在这个过程中往往将要系统部署到服务器上运行，测试人员进行测试时只关注测试结果与此功能预期结果是否一致。如果测试结果与预期所要达到的结果一致，测试通过，反之，测试不通过。在CFSYSTEM的登陆功能测试过程中，测试人员进入系统登录页面，在表单对应的输入框中填写开发人员所提供测试账号的用户名和密码，点击登陆按钮后，如果页面提示登录成功的信息则说明登录功能测试通过，反之，登录功能测试失败。当测试人员完成测试，将把测试结果反馈给开发人员，开发人员根据测试结果，进行下一步开发工作。表4-1显示CFSYSTEM主要模块的功能测试，页面相应情况以及测试结果。

表4-1系统主要模块的功能测试情况及结果

|  |  |  |  |
| --- | --- | --- | --- |
| 模块 | 功能测试 | 页面响应情况 | 测试结果 |
| 登录模块 | 正常登录 | 成功进入系统 | 测试通过 |
| 用户名或密码为空 | 显示友好错误提示用户名或密码不能为空 |
| 用户名不存在 | 显示友好错误提示：用户名不存在 |
| 用户名和密码不匹配 | 显示友好错误提示：用户名和密码不匹配 |
| 忘记密码模块 | 邮箱地址正确 | 系统显示邮件发送成功提示，进入该邮箱查收邮件，修改密码，用新密码正常登陆 | 测试通过 |
| 邮箱地址错误 | 页面显示邮箱地址不存在 |
| 用户信息模块 | 修改用户信息 | 修改用户信息不符合该项输入要求，输入框变红，并显示友好错误提示 | 测试通过 |
| 保存用户信息 | 提示信息保存成功 |
| 发布公告模块 | 输入公告信息 | 公告信息只支持文本输入，不支持输入格式 | 测试通过 |
| 发布公告 | 提示公告发布成功 |
| 学习资源上传模块 | 上传学习资源 | 提示学习资源上传成功 | 测试通过 |
| 查看上传的学习资源信息 | 显示上传的学习资源信息列表 |
| 学习资源下载模块 | 下载学习资源 | 提示下载成功，文件存放于本地指定文件夹内 | 测试通过 |
| 学习资源审核模块 | 通过审核 | 系统可查询到审核通过学习资源 | 测试通过 |
| 拒绝通过审核 | 系统中无法查询该学习资源 |
| 学习资源删除模块 | 删除学习资源 | 提示删除成功，系统中无法查询该学习资源 | 测试通过 |
| 学习资源个性化推荐模块 | 查看推荐的学习资源 | 显示学习资源详情页面 | 测试通过 |
| 下载推荐 | 提示下载成功，文件存放于本地指定文件夹内 |
| 退出系统 | 注销用户 | 退出登录，返回系统登录也，清除cookie、session | 测试通过 |

### 4.3.3 界面测试

界面测试，主要测试系统界面布局是否合理，页面元素是否发生错位，不同页面的风格是否保持一致，页面特效展示是否流畅等等。CFSYSTEM在界面设计上，采用了当前最流行的响应式框架Bootstrap3.0，对PC，手机和PAD进行完美兼容。页面图标字体风格采用font-awesome，页面弹出框选自github上的dialog，日期显示调用My97DatePicker，还有一些主流插件如jquery-ui、select2、validate、jquery-form等进行搭配使用。对于界面测试，笔者特地邀请惠普公司美工设计师进行体验并测试，美工设计师针对界面主色调给出建议，并给出色调调整方案，其他界面测试均已通过。

# 第五章 总结与展望

## 5.1 总结

在互联网技术与现代远程教育技术高速融合的背景下，互联网上的学习资源数量呈指数型增长，出现“信息过载”现象。这样，学习者想要获取能够符合自己的学习资源，就会存在一定的困难。因此，设计并实现了协同过滤算法的学习资源个性化系统，以达到为互联网学习者提供学习资源个性化推荐服务的目标。以下是本课题进行探究的几个方面内容：

（1）通过研究大量关于学习资源个性化推荐的专业文献，结合当前背景形势，以及国内外相关动态，确定本课题研究目标。参考现有互联网学习环境中学习资源个性化推荐的不足和由此引发的问题，探索解决方法，提出解决方案。

（2）在对基于协同过滤算法的学习资源个性化推荐系统设计和实现相关理论以及技术进行深入研究的前提下，针对本课题选定的协同过滤算法中存在的新用户冷启动、新项目冷启动和数据稀疏问题，提出具体化的解决方案。

（3）本课题将基于协同过滤算法的学习资源个性化推荐系统的设计与实现作为核心。CFSYSTEM为本次课题所设计和实现的系统，系统设计过程中从需求分析到系统结构设计再到推荐算法应用设计最后到数据库设计，科学严谨的设计过程就是为了CFSYSTEM开发阶段能够更有效的进行。系统后端开发主要采用JAVA作为系统编程语言，数据库采用Redis和MySQL，前端引进Bootstrap响应式框架为PC、手机、PAD平台兼容提供支持。CFSYSTEM中学习资源个性化推荐模板使用协同过滤算法中基于用户的协同过滤和基于项目的协同过滤混合算法。推荐模块中，首先需要从MySQL数据库中获得相应评分数据集，其次计算出用户或者项目相似度，得出TOP-N近邻推荐集。最后，将推荐结果返回给CFSYSTEM用户。这样，可以为互联网学习者节省寻找学习资源的时间，也能在互联网学习者需求不明确或者没法清晰表达的时候为其进行个性化推荐。

## 5.2 展望

学习资源个性化推荐系统是现代互联网学习环境中的一个子系统。虽然个性化推荐技术在互联网电商平台被应用的非常成熟，但是在互联网学习平台上还几乎相当于一个空白区。所以，本人在研究过程中遇到相当多难题，提出了相应的解决方法，也起到了一定的积极结果。本次研究还可以从以下几个方面进行拓展，列举几处：

（1）本课题侧重学习资源个性化推荐，然而，互联网学习环境千变万化，影响个性化推荐的因素的考虑随之越来越复杂。进一步完善用户模型中学习者特征设计尤为重要，下一步要做的就是引入更多学习者的特征，比如学习动机、认知水平、学习风格、心理发展状况。

（2）CFSYSTEM的学习资源个性化推荐模块只有在学习资源数量足够大时，协同过滤算法发挥的作用才越大，推荐准确度越高。但是，在研究过程中使用的所能获取最大的数据量也并没有达到这一前提，这局限了协同过滤算法实验验证时在准确性。下一步需要将系统推广给更多用户使用，丰富CFSYSTEM中的学习资源。

# 参考文献

1. CNNIC发布第37次《中国互联网络发展状况统计报告》[J].国家图书馆学刊,2016,(02):76.
2. 王春才,邢晖,李英韬.个性化推荐系统冷启动问题研究[J].现代计算机(专业版),2015,(29):36-38+60.
3. Weiquan Wang and IzakBenbasat. Recommendation Agents for Electronic Commerce: Effects of Explanation Facilities on Trusting Beliefs[J]. Journal of Management Information Systems,2007,(4): 217-246
4. 廖竞.基于网络结构的学习推荐技术研究[D].西南科技大学,2011.
5. 黄昌勤.网络课程资源支持下基于语义Web的个性化主动学习服务研究与实现[J].电化教育研究,2012,(03):49-55+60.
6. 孙励.个性化推荐技术及其在远程教育中的应用[J].学周刊,2015,(28):14.
7. 杨现民,余胜泉,张芳.学习资源动态语义关联的设计与实现.中国电化教育,2013(01):70-75.
8. 姚曜等.协同过滤技术研究综述.装备指挥技术学院学报,2011(05):81-88.
9. 赵向宇.Top-N 协同过滤推荐技术研究[D],2014,北京理工大学
10. 沈键. 电子商务的个性化协同过滤推荐算法研究[D].上海交通大学,2013.
11. Lops P,De Gemmis M,Semeraro G. Content-based recommender systems: State of the art and trends[M]. Recommender systems handbook. Springer US, 2011: 73-105.
12. Aher S B,Lobo L. Combination of machine learning algorithms for recommendation of courses in e-learning system based on historical data[J]. Knowledge-Based Systems, 2013,51: 1-14.
13. Salehi M,Nakhai Kamalabadi I. Hybrid recommendation approach for learning material based on sequential pattern of the accessed material and the learner’s preference tree[J].Knowledge-Based Systems,2013,48: 57-69.
14. 杨麟儿.基于用户兴趣的个性化推荐系统的研究与设计[D].北京交通大学,2008.
15. 孔维梁.协同过滤推荐系统关键问题研究[D].华中师范大学,2013.
16. 许景楠.基于评论和评分的个性化推荐算法研究[D].浙江大学,2013.
17. 郭永志.基于学习共同体理论的网络学习模式研究[J].中国电化教育,2011(8).
18. 田富鹏,田斐予.基于联通主义的个人学习环境(PLE)研究综述[J].电化教育研究,2014(7).
19. 刘强.协同过滤推荐系统中的关键算法研究[D].浙江大学,2013.
20. 查文琴. 基于用户聚类的项目多内容协同过滤推荐[D].合肥工业大学,2009.
21. B Sarwar,G Karypis,J Konstan,et al. Item-based collaborative filtering recommendation algorithms[C].//Processdings of the 10th International conference on World Wide Web. ACM,New York, 2001:285-295.
22. Lam X N,Vu T,Le T D,et al. Addressing cold-start problem in recommendation systems[C]. Proceedings of the 2nd international conference on Ubiquitous information management and communication. ACM, 2008: 208-211.
23. 王永固,邱飞岳,赵建龙,刘晖.基于协同过滤技术的学习资源个性化推荐研究[J].远程教育杂志,2011,(03):66-71.
24. 徐天伟,宋雅婷,段崇江.基于协同过滤的个性化推荐选课系统研究[J].现代教育技术,2014,(06):92-98.
25. 陈彦萍,王赛.基于用户-项目的混合协同过滤算法[J].计算机技术与发展,2014,(12):88-91+95.
26. 王金辉.基于标签的协同过滤稀疏性问题研究[D].中国科学技术大学,2011.
27. 顾宗林.IPv6网络数据流监测技术的分析与测试[D].北京交通大学,2008.
28. 王志军,陈丽.联通主义学习理论及其最新进展[J].开放教育研究,2014(5).
29. 王佑镁,祝智庭.从联结主义到联通主义:学习理论的新取向[J].中国电化教育,2006(3).
30. 张乐乐,黄如民.联通主义视域下的移动学习环境设计[J].现代教育技术,2013(2).
31. 丁琳,旲长永.数据挖掘在远程教疗个性化服务中的应用[J].电化教育研究,2002(9):43-46
32. 祝智庭.网络教育应用教程[M].北京:北京师范大学出版社,2007:27
33. 易明.基于WEB挖掘的个性化信息推荐[M].北京:科学出版社,2010,33-34
34. Han JW, Kamber M.数据挖掘概念与技术[M].范明,孟小峰译.北京:机械工业出版社,2001:45-47
35. 刘强.协同过滤推荐系统中的关键算法研究[D].浙江大学,2013.
36. 孟红娟,郑旭东.对AECT2005教育技术定义的批判分析与思考[J].电化教育研究,2005,(06):34-37.
37. 罗永祥. AECT’94定义对中国教育技术发展的影响——以相关论文为研究对象的内容分析[J]. 中国电化教育,2004,(05):14-18.
38. 查文琴.基于用户聚类的项目多内容协同过滤推荐[D].合肥工业大学,2009.
39. 任宇. E-Learning个性化系统的推荐策略研究——来自电子商务推荐系统的启示[D].东北师范大学,2010.
40. Klašnja-Milićević A, Vesin B, Ivanović M, et al. E-Learning personalization based on hybrid recommendation strategy and learning style identification[J]. Computers & Education, 2011,56(3): 885-899.
41. Chen Y J, Chu H C, Chen Y M,et al. Adapting domain ontology for personalized knowledge search and recommendation[J]. Information & Management, 2013, 50(6): 285-303.

# 致谢

天津师范大学的学生生涯，慢慢地已经接近了尾声。回首过去，思绪万千。在这里我度过人生中青春的七年，一届又一届老生的离去换来一届又一届新生的到来，这次即将离开母校的换成了自己，不禁内心涟漪。此刻，我只想说感谢，感谢一路有你。

首先，对于我的导师于鷃老师在这三年来给予的指导与帮助、关怀与鼓励，我想在这里真诚的表示感谢。即使有繁忙的教学任务，于老师对于我们的论文也是一丝不苟。不论是从开始的论文选题还是到后来的论文撰写，总是给我们提出建设性的意见和指导。

其次，我还要感谢教育科学学院教育技术学专业的各位硕士生导师，包括王志军老师、姜艳玲老师、卢丽萍老师、柳彩志老师等三年来在学业上对我的教诲与启迪，在此，谨向各位老师表示诚挚的谢意。学院的石丽老师、杨绍峰老师、张麦玲老师等多位老师在平时学习和生活上也给予了许多的帮助，让身处他乡的我感受到了家的温暖。

然后，我要衷心感谢2014届教育技术学专业的全体同学，正是大家在学习和生活上给予我莫大的帮助与支持，使我的求学之路充满欢乐。你们对教育事业的满腔热血，对理论学习孜孜不倦，让我看到了即将出现在下一个教育时代中佼佼者的身影。衷心的祝愿你们工作顺利，事业有成。

最后，我要感谢我的父母。感谢你们给予了我生命，感谢你们养育了我成长。每次我回家你们都得千叮嘱万嘱咐，怕我一个人在外吃不好穿不好，怕我受到委屈，而我总说你们的担心是多余的，但心里却异常温暖。即将步入工作岗位，愿在以后的生活中我能够成为你们坚强的精神后盾。