



南京理工大学
NANJING UNIVERSITY OF SCIENCE & TECHNOLOGY

毕业设计说明书

作 者： 张振源 学 号： 913106840514

学 院： 计算机科学与工程学院

专业(方向)： 计算机科学与技术

题 目： 商品移动推荐系统的研究与开发

指导者： 李泽超
(姓 名) (专业技术职务)

(姓 名) (专业技术职务)

评阅者：
(姓 名) (专业技术职务)

2017 年 5 月

声 明

我声明，本毕业设计说明书及其研究工作和所取得的成果是本人在导师的指导下独立完成的。研究过程中利用的所有资料均已在参考文献中列出，其他人员或机构对本毕业设计工作做出的贡献也已在致谢部分说明。

本毕业设计说明书不涉及任何秘密，南京理工大学有权保存其电子和纸质文档，可以借阅或网上公布其部分或全部内容，可以向有关部门或机构送交并授权保存、借阅或网上公布其部分或全部内容。

学生签名：

年 月 日

指导教师签名：

年 月 日

毕业设计说明书中文摘要

随着我国经济社会现代化发展水平的提高，商场上的商品种类繁多，消费者应接不暇，在挑选商品的过程中往往浪费了许多时间，甚至不利于市场经济的发展。同时，基于 android、ios 系统等移动终端的消费形式有着 PC 端消费不可比拟的便利性，快捷性等，“掌上消费”在人们的消费形式中的占据的比重越来越大。于是，推荐系统如何在繁多的商品中，挑选出具有用户个性化特征的商品是一个值得关注的问题。这种消费形式不仅可以为企业节省金钱和时间、人力，也提高了客户的购物环节的效率 and 满意度。尤其在生活节奏比较快的大都市，人们不希望在挑选商品的过程太缓慢，特别是工作忙碌的上班族人群。

本文中的推荐系统通过用户对以往商品的评分数据，采用 BPFM 贝叶斯概率矩阵分解算法，对数据稀疏的用户商品评分的矩阵进行因子分解，最后将用户可能感兴趣的商品推荐给用户。本文中的推荐系统是基于 Android 平台开发的 APP。推荐系统将用户信息保存到服务器上的 MongoDB 数据库中。服务器与 android 移动终端的网络通信架构采用 AsyncHttp，服务器将用户最可能感兴趣的商品通过 servlet 发送给移动终端，移动终端并将这些数据反馈给用户。

本文研究工作的成果及意义在于，设计及实现的 android 终端应用能在信息量巨大的商品数据库中，在用户可接受的时间范围和推荐满意度的范围内，快速、有效地进行商品推荐，从而节省商家和消费者的时间和成本。

关键词 BPFM android MongoDB 移动终端 商品推荐

毕业设计说明书外文摘要

Title RESEARCH AND IMPLEMENTATION of A PRODUCT RECOMMENDATION
SYSTEM on MOBILE PHONES

Abstract

With the development of China's economic and social modernization level, a wide range of goods are displayed on the shopping malls, and consumers overwhelmed in the process of selecting goods often waste a lot of time, and even not conducive to the development of market economy. At the same time, based on the android, IOS system and other forms of consumption of mobile terminals have an incomparable convenience with PC-side consumption etc., "handheld consumption" in people's consumption in the form of occupying an increasing proportion. So, how to recommend in a wide range of goods, and select the goods with personalized features is a matter of concern. This form of consumption can not only save money and time for enterprises, manpower, but also improve the customer's shopping efficiency and satisfaction. Especially in the metropolis, people especially the busy working group of people, do not want to slowly pick the goods.

By the user's score of the previous products, the recommendation system in this paper uses the BPMF Bayesian probability matrix decomposition algorithm and finally recommend the goods which the users may be interested of. The recommended system in this article is based on the Android development platform. The recommended system saves the user information to the Mongo database on the server. Server and Android mobile terminal network communication architecture use AsyncHttp, the server will sent the goods which users most likely interested of through the servlet to the mobile terminal, and mobile terminal back the data to the user.

The results and significance of this research work is that the design and

implementation of the android terminal application can quickly and effectively carry out the product recommendation in the commodity database with huge amount of information, within the range of user acceptable time range and recommendation satisfaction, so as to save time and money for businesses and consumers.

Keywords [BPMF](#) [android](#) [MongoDB](#) [mobile terminal](#) [recommend system](#)

目 录

1	绪论	1
1.1	工程背景及意义	1
1.2	国内外研究现状	1
1.3	总体技术方案及其社会影响	2
1.4	技术方案的经济因素分析	3
1.5	论文章节安排	3
2	基础知识	5
2.1	商品移动推荐系统机制	5
2.2	商品移动推荐系统开发工具介绍	6
3	需求分析	8
3.1	功能需求	8
3.2	接口需求	9
3.3	非功能需求	9
4	系统设计与实现	10
4.1	支撑环境	10
4.2	总体设计方案	10
4.3	类的设计	11
4.4	数据库设计	12
4.5	具体实现	15
	结 论	22
	致 谢	23
	参 考 文 献	24

1 绪论

我国商品经济发展迅速,商品种类也是越来越多,然而人们的生活节奏也快,尤其是在北上广这样的大都市,人们在购买商品时,越来越不愿意在挑选商品的过程中花费过多的时间,基于此背景及现代社会手机人手一部的背景下,本文提出商品移动推荐系统。

1.1 工程背景及意义

据中国零售行业市场竞争力分析报告,中国 2012 年网络销售总额已经达到 2000 亿美元,这一数据非常接近世界最大的网络销售市场——美国。中国网络销售的发展潜力非常巨大。截止 2011 年,中国的网民数量已位居全球首位,宽带用户的数量达到 1.29 亿人,远远超过美国的 8100 万人,但是中国的宽带普及率才只有 30%。此外,中国的网络零售的增长速度也大大超过其他国家和地区,从 2003 年开始平均增长率达 120%,按照这个速度发展,到 2020 年中国的网络购物市场规模可能达到 5000 亿美元,这一数额可能相当于美日英德法目前世界前五的规模总和^[1]。

互联网上的商品信息量巨大,并且现代社会人人基本上都离不开手机等移动终端,其便利性和快捷性在互联网消费上展示的淋漓尽致。消费者们往往不可能花费大量的时间去挑选自己满意的商品。消费者们如何合理且高效率地选择出自己满意的商品,成为了商品成功售出的关键问题。本文描述的基于 android 移动终端的商品推荐系统的提出试图为广大的手持移动终端的消费者们解决这个商品挑选困难的问题。

1.2 国内外研究现状

近些年来随着电子商务的普及,互联网上移动终端消费者的迅速增多,电子商务呈现出一篇欣欣向荣的景象。同时,这也意味着商品信息的“大爆炸”,人们担心的不再是商品信息太少无从购选,而是担心信息太多使人难以购选。于是,如何快速、有效地进行商品推荐成了一个具有很大现实意义的话题。越来越来的学者投身其中,提出了很多推荐算法,但可大致分为以下三类:基于内容的推荐算法、协同过滤推荐算法和基于知识的推荐算法。

1.2.1 国内研究现状

近些年来,随着互联网的深入发展,越来越多的信息在互联网上传播,产生了严重的信息过载。如果不采用一定的手段,用户很难从如此多的信息流中找到对自己有价值的信息,

于此，我国涌现了很多购物推荐的网站。如淘宝网、当当网、卓越网等 B2C 网站和百度糯米、美团网等 O2O 网站。他们主要是根据用户以往搜索过的内容以及购买记录、收藏的商品或商家等推荐相关的商品，达到了较好的用户体验。

百度糯米、美团网这类团购网站现在也已经是广为人知，很多商家也已经有美团、百度糯米优惠折扣，它们会根据你选择的所在地、分类情况（如美食）等进行相应的推荐。另外，网易云音乐、虾米网也有了付费音乐（随着知识产权法力度的普及，越来越多的音乐人的作品得到版权保护，很多作品也因此收费）的推荐，这些网站则是根据用户的兴趣向用户推荐音乐，而且其推荐的音乐歌单，得到了大多数用户的好评。

1.2.2 国外研究现状

互联网最先是在美国出现的，与其相应的，网络销售也是最先在美国出现。对于商品推荐算法，也是国外研究的比较早。推荐系统最早被应用在电子商务网站中，由于用户的需求往往是模糊的、不明确的，此时如果商家根据用户的兴趣、爱好准确地预测出用户可能感兴趣或者正在寻找的商品，那么商家就能够把用户模糊的潜在需求转化为现实需求，从而实现商品销售量的提高^[10]。在美国，最著名的购物网站当属亚马逊了，总所周知 Amazon 是以音像制品和书籍起家，也是全球最早也是最大的书籍经销商。如今已成长为包罗万象的超级综合网站。就亚马逊书城而言，其会根据用户浏览过的商品和收藏的书籍或商家进行相关的书籍推荐。在学术界，国际上对于推荐技术相关的研究也如雨后春笋般出现。自 1999 年开始，ACM 每年召开的 ACM-EC（电子商务研讨会），推荐系统的研究文章逐步增大；ACM 的 SIGKDD 小组设立 WEBKDD 组，主题集中在电子商务中的 Web 挖掘技术和推荐技术。自 2001 年开始，ACM 下面的 SIGIR 开始专门把推荐系统作为一个研讨主题。21 世纪以后，数据挖掘、机器学习、人工智能等领域中相关的重要国际学术会议中，如 AAAI、IDCM、ICML、SIGIR、ECML 等，都包含推荐系统或相关算法技术的主题内容^[11]。

1.3 总体技术方案及其社会影响

本系统主要采用了 BPMF 贝叶斯概率矩阵分解算法、servlet 技术（一种基于 Java 编程语言的 Web 服务器端编程技术）、安卓网络请求技术 AsyncHttp（一款为 android 网络请求的框架，采用最少的线程，基于 AsyncPool 线程池实现的 http 请求，支持 post 和 get、delete 等多种请求，实现请求任务和结果异步获取^[12]），用以 Android 客户端和数据库通过 Web 服务器进行间接连接，以减少 Android 端处理器运行等的压力，增加系统的可扩展性、可用性等。本文

通过将用户-商品-评分数据矩阵采用 BPMF 贝叶斯概率矩阵分解算法, 将得到具有用户隐含特征的矩阵和商品隐含特征的矩阵, 再将这两个矩阵行列向量相乘, 得到多个特定用户-商品-评分矩阵, 对预测评分进行排序后, 将预测评分前几的商品通过手机终端上的 APP 展示给用户。对于广大的用户而言, 该技术大大减少了用户挑选商品的时间, 将消费者的潜在购物需求激发出来, 有力地促进了市场经济的发展, 使消费者-生产者达成一个良好的商业循环, 甚至促进 GDP 的增长。

1.4 技术方案的经济因素分析

本系统是基于 Android 平台, 利用 MATLAB、MyEclipse、Android Studio、MongoDB 等工具等进行开发实现的。在本系统的商品推荐环节, 采用的是 BPMF (贝叶斯概率矩阵分解) 算法, 从经济的角度分析, 该推荐算法成熟而且是公开的、免费的, 因此在推荐算法的成本上, 可以说是零成本。该 BPMF 算法相对成熟, 其预测评分还是具有一定的可靠性的, 所以其性价比还是很高的。另外, 现实生活中, 大多数人用的手机/平板系统是 android 系统, 因此本系统具有广阔的用户前景。还有, 在开发本系统的过程中, 我们所用到的工具等, 大多数是免费使用的 (MyEclipse 是收费的, 当然也可以使用免费的 Eclipse 替代), 因此在开发环节, 成本还是较低的。

1.5 论文章节安排

本文的开头介绍了全球网络零售市场的发展状况和前景, 以及现在市场上确实很需要该类的商品移动推荐系统去节省人们在挑选商品过程中所花费的时间。接着, 本文针对商品移动推荐系统的需求, 分析国内外研究现状, 将退出商品移动推荐系统的软件架构设计并做出详细的研究。

本文的章节安排和组织结构如下:

第一章介绍了全球市场上网络销售的良好发展状况和广阔的前景以及国内外商品推荐移动应用的现状, 以及本文将要研究实现的目标。

第二章探讨了设计该商品移动推荐系统所需的基础知识: 系统的开发工具、开发技术等, 以及为何用这些工具及技术, 使用这些工具和技术的益处等。

第三章是对该商品移动推荐系统进行需求分析, 分为功能需求、接口需求、非功能需求三个部分。在此章中主要探讨功能需求, 以满足用户对该商品移动推荐系统的需要, 提高该系统设计的有效性。

第四章主要介绍了本文的系统的设计与实现的细节。在系统的设计方面，根据第三章对目标系统的需求分析，对系统所应满足的功能需求等进行详细的设计。在系统的实现方面，主要介绍了系统在实现中应该注意的关键部分以及主要的实现步骤。

2 基础知识

当今，互联网发展迅速，其所带来的信息爆炸，人们现在担心的不再是信息量太少，而是担心信息量太大，导致人们在挑选货物时无所适从，而在此购物环节中花费了太多的时间。于此背景下，现在已经出现了很多的商品推荐系统。本文通过研究 BPMF 贝叶斯概率矩阵分解算法并将其应用到商品移动推荐系统中，试图帮助人们节省在挑选商品过程中所付出的时间。另外，很多时候用户的意图并不是很明确，或者很难用清晰的语义表达，有时甚至连用户自己都不清楚自己的需求，这个时候推荐系统的出现，也很好地满足了用户这种“逛”的需求。

本文所研究的商品移动推荐系统的数据是采用 Netflix prize 中公开的数据集，可将其简化成用户-商品-评分的矩阵，其列向量为 3。在 MATLAB 上运行 BPMF 算法，将其分解成具有隐含特征的用户矩阵和具有隐含特征的商品矩阵，再将这两个矩阵的行列向量相乘，得到许多具体的用户-商品-预测评分矩阵。将其在 MATLAB 上的矩阵存储到服务器的数据库中。数据库采用的是 MongoDB，通过架设 Apache Tomcat 服务器，使用一种 android 网络请求架构 Asynchttp，将所要推荐的商品在 android app 上展示给用户。

2.1 商品移动推荐系统机制

在商品移动推荐系统中的实现过程中，该系统不建议直接连接 MongoDB 数据库，而是通过服务器进行排序之后将结果通过 Servlet 输出到 Web Browser 中，android 端通过 AsyncHttp 将内容解析展示在 APP 的 UI 中。该系统的通信架构如图 2.1 所示。

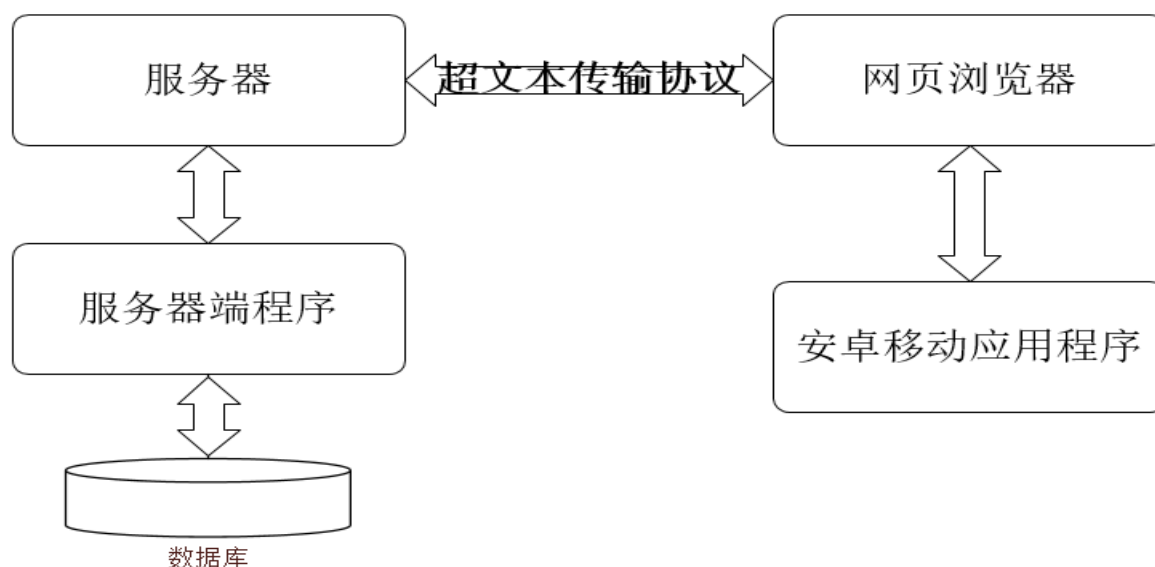


图 2.1 servlet 架构和 android 通信图

2.2 商品移动推荐系统开发工具介绍

为了实现该系统，本文采用了多种开发工具：MATLAB、MongoDB、Android studio、MyEclipse。接下来，我们主要介绍用这些工具的好处及用处。

2.2.1 MATLAB

MATLAB 是美国 MathWorks 公司出品的商业数学软件，用于算法开发、数据可视化、数据分析以及数值计算的高级技术计算语言和交互式环境，主要包括 MATLAB 和 Simulink 两大部分。MATLAB 是 matrix&laboratory 两个词的组合，意为矩阵工厂（矩阵实验室）。它将数值分析、矩阵计算、科学数据可视化以及非线性动态系统的建模和仿真等诸多强大功能集成在一个易于使用的视窗环境中，为科学研究、工程设计以及必须进行有效数值计算的众多科学领域提供了一种全面的解决方案，并在很大程度上摆脱了传统非交互式程序设计语言（如 C、Fortran）的编辑模式，代表了当今国际科学计算软件的先进水平^[2]。我们的 Netflix prize 公开的数据集用矩阵的形式在 MATLAB 中表示。由上可知，MATLAB 很适合进行矩阵相关的运算，在 MATLAB 中进行贝叶斯概率矩阵分解，也是更好地提高矩阵分解、得到结果集的速度。

2.2.2 MongoDB

当前，互联网用户多达几十亿，传统的 RDBMS 数据库解决方案难以满足快速增长的海量数据处理需求，人们越来越多地采用专用数据库，它们不受制于传统 SQL 数据库的限制和开销。这些数据库被称为 NoSQL，意思是“不仅仅是 SQL”；它们并非要取代 SQL 数据库，而是旨在提供另一种数据存储方式。而本文中所用的 MongoDB 正是当前使用最为广泛的 NoSQL 数据库^[3]。

网络上商品种类繁多，可能同种类型不同商品并不是具有相同的特征分类，所以并不是每件商品都是适合用相同的表（例如 MySQL 中的表）去存储商品信息，而 MongoDB 的文档结构则很适合存储这种种类繁多且杂乱的信息。另外，MongoDB 是市面上性能最高的数据库之一。商品信息量巨大，所以要求数据库的性能也是越高越好。同时，MongoDB 具有高扩展性，其结构使得能够将数据分不到多台服务器中，从而轻松地实现横向扩展。

2.2.3 Android Studio

Android Studio 是 Google 于 2013 年 I/O 大会上针对 Android 开发退出的新的开发工具，是 Google 专门为 Android “量身定制”的官方 IDE 工具，而且 Google 建议广大 Android 开发者尽快从 Eclipse+ADT 的开发环境改为使用 Android Studio^[6]。Android Studio 不再基于 Eclipse，

而是基于 IntelliJ IDEA 的 Android 开发环境。实际上，IntelliJ IDEA 一直都是一款非常优秀的 Java IDE 工具，只是因为 IntelliJ IDEA 是一款商业的 IDE 工具（虽然也有免费的社区交流版，但功能相当有限），因此影响了 IntelliJ IDEA 的广泛应用。现在，Google 以 IntelliJ IDEA 为基础推出的 Android Studio 同样可以免费使用，因此具有非常大的吸引力。从成本和便利性上考虑，本商品移动推荐系统采用 Android Studio 进行开发。

2.2.3 MyEclipse

MyEclipse，是在 Eclipse 基础上加上自己的插件开发而成的功能强大的企业级集成开发环境，主要用于 Java、Java EE 以及移动应用的开发^[7]。MyEclipse 是目前 Java 开发人员最喜欢的开发工具之一，本商品移动推荐系统采用 MyEclipse 以实现服务器端上的 Web 开发。服务器通过 Web 上的内容与移动端进行交互。

3 需求分析

在互联网技术高速发展的今天，越来越多的商品经过网络的途径售出，无论是以 B2B（企业对企业）、B2C（企业对个人）、C2C（个人对个人）如闲鱼（一个二手货物网络销售平台）、O2O（网上购买线下消费）如美团和百度糯米等，人们的消费形式渐渐地传统的线下消费转变成线上消费，而且，随着手机普及的今天，相比于 PC 端，人们更喜欢在移动端上进行网上消费。但伴随着高速发展的互联网技术而来的，也有着商品数据量过于巨大、繁杂的问题，人们挑选处理起来很费时间，于此背景下，人们更希望有那么一个助手，能帮助人们根据自身喜好进行货物推荐、帮助人们节省时间和精力。这不仅是消费者希望得到的目标，也是商家更希望看到的。

3.1 功能需求

就实际意义而言，商品移动推荐系统用户希望得到自己可能感兴趣的商品内容，而该系统对每一个不同的用户其推送的内容都应该是不同的。这样的系统应该包括用户管理模块，商品信息推送模块。

3.1.1 用户管理需求

首先，该商品移动推荐系统应该是基于用户的，用户起码可以注册账号或者通过登录之前已注册过的账号进入该系统。

1) 注册功能。在打开 APP 的第一个界面中，应该有明显的注册的按钮，可以让新用户简单地单击该按钮地进入账号注册界面。在该注册界面中，应该有用户名、密码、再次确认密码（为了防止注册用户之前输入的密码输错了）、邮箱地址、手机号码。其次，邮箱地址和手机号码应该是只有正确的邮箱格式和手机号格式才能成功注册。

2) 登录功能。登录按钮也应该是在打开 APP 的第一个界面中。因为每一个用户的特征（之前对于各种商品的评分、可能感兴趣的物品）都不一样，所以必须得有用户登录的功能，在用户登录之后，才能基于用户进行相应的商品内容推送。

3) 密码找回功能。大千世界，在信息爆炸的今天，人们要记忆的事情很多，所以很有可能因为太久没登录该商品移动推荐系统而忘记了自己的密码。所以在 APP 的登录界面中，应该有找回密码的按钮。该系统可以用邮箱地址和手机号码（两个都填对时）才能修改密码。

3.1.2 商品信息推送模块

在用户登录成功后，在主界面的工具栏中应该至少有一个刷新按钮进行用户可能感兴趣的商品内容进行刷新和显示在主界面列表中。

3.2 接口需求

该系统的用户界面应该是友好的，对于账号和密码填入的文本框应该有明显的提示，可采用 `android.support.design.widget.TextInputLayout` 下的 `AutoCompleteTextView`，这样的文本框在获得焦点后，会将文本框提示 `android: hint` 的内容变小并在该文本框上方显示。

另外，商品移动推荐系统不应是应用 Android 自带的 SQLite 数据库，所以该系统应能与服务器进行通信，则需要在 Android Studio 的环境下，在 Android 应用的清单文件中 `AndroidManifest.xml` 文件中添加 Android 允许应用程序打开网络 Socket 权限 `uses-permission android:name="android.permission.INTERNET"`。

3.3 非功能需求

在信息爆炸的时代下，提出的商品移动推荐系统其推荐商品内容的速度应该是能在人们接受的时间范围内的，也就是说应该快速地得到推荐的内容。所以得采用高性能的、检索查询速度快的数据库。在安全需求方面，APP 最好采用用户数据与移动终端相分离的形式，用户数据（如用户密码等）应该存储在服务器上，用户成功登录之后，再将用户所需的内容通过 Android 网络请求框架 `Asynhttp` 返回给用户。

4 系统设计与实现

根据上一章对该商品移动推荐系统的功能需求分析，我们可以将该系统分成 3 个模块：用户-商品数据处理模块（预处理模块）、用户管理模块、商品信息推送模块。首先，在预处理模块中，我们将 Netflix prize 的测试数据集在 MATLAB 上进行 BPMF（贝叶斯概率矩阵分解）、再得到用户-商品-评分矩阵，再将这些矩阵数据导入到 MongoDB 中。然后，我们在用户管理模块中可进行注册或者登陆等操作。在登录成功后，进入到商品信息推送模块，该系统将该用户相应的商品信息在主界面的列表中显示出来。

4.1 支撑环境

4.1.1 软件环境

进行矩阵分解：MATLAB R2014b；

数据库：MongoDB v3.4.3；

Web project IDE: MyEclipse 10；

App IDE：Android Studio 2.1.2。

4.1.2 硬件环境和网络环境

服务器配置：

CPU: Intel(R) Core(TM) i7-3630QM @2.40GHz；

内存: 4GB DDR3；

操作系统: 64位的 windows10；

Android 真机配置：

Android 版本：5.1.1，API22

处理器：Qualcomm Technologies, Inc. MSM8939

内存：1917MB

Smartisan OS（锤子手机系统）版本：v2.6.4

网络环境：该系统用无线局域网，移动端与服务器通过局域网环境进行数据交互。

4.2 总体设计方案

在本商品移动推荐系统中，根据需求分析，进行功能模块的划分，下面给出相应的总体流程图如图 4.2.1 和系统架构图如图 4.2.2。

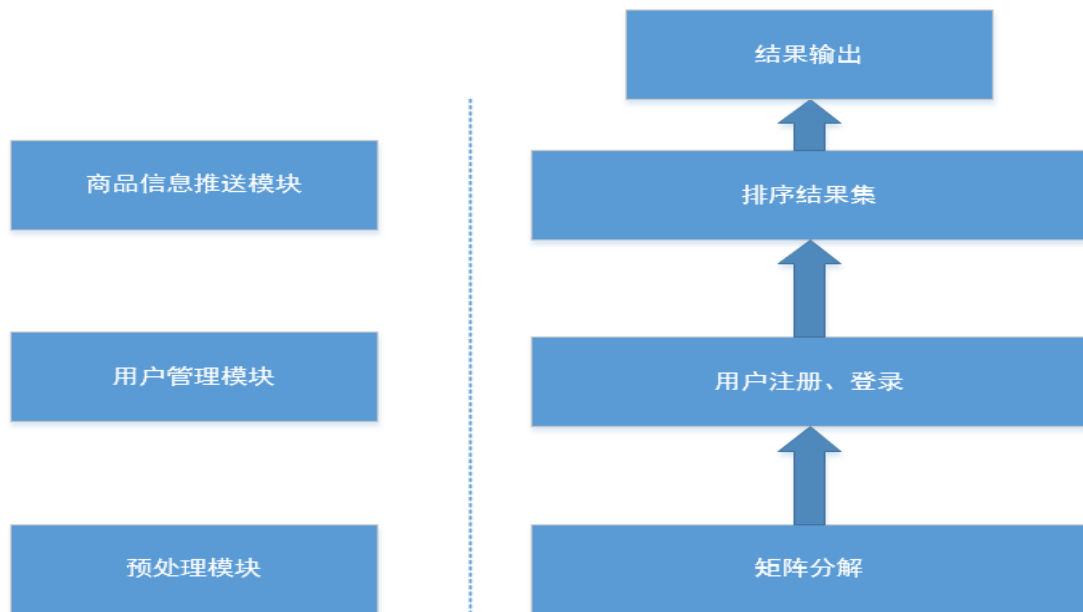


图4.2.1 总体流程图

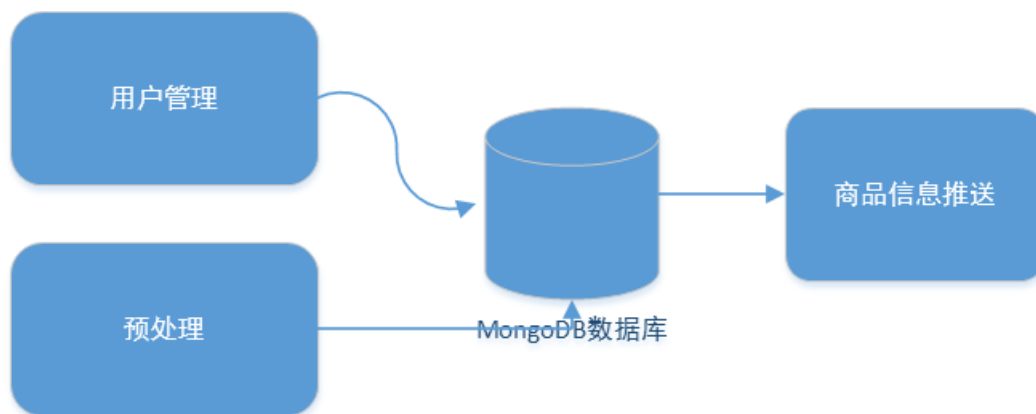


图4.2.2 系统架构图

4.3 类的设计

在本节中，针对上面的需求分析和总体设计方案，可设计了以下的类图，描述所有类及其属性、方法的含义，并用类图表示。本商品移动推荐系统的 UML 类图如图 4.3 所示。

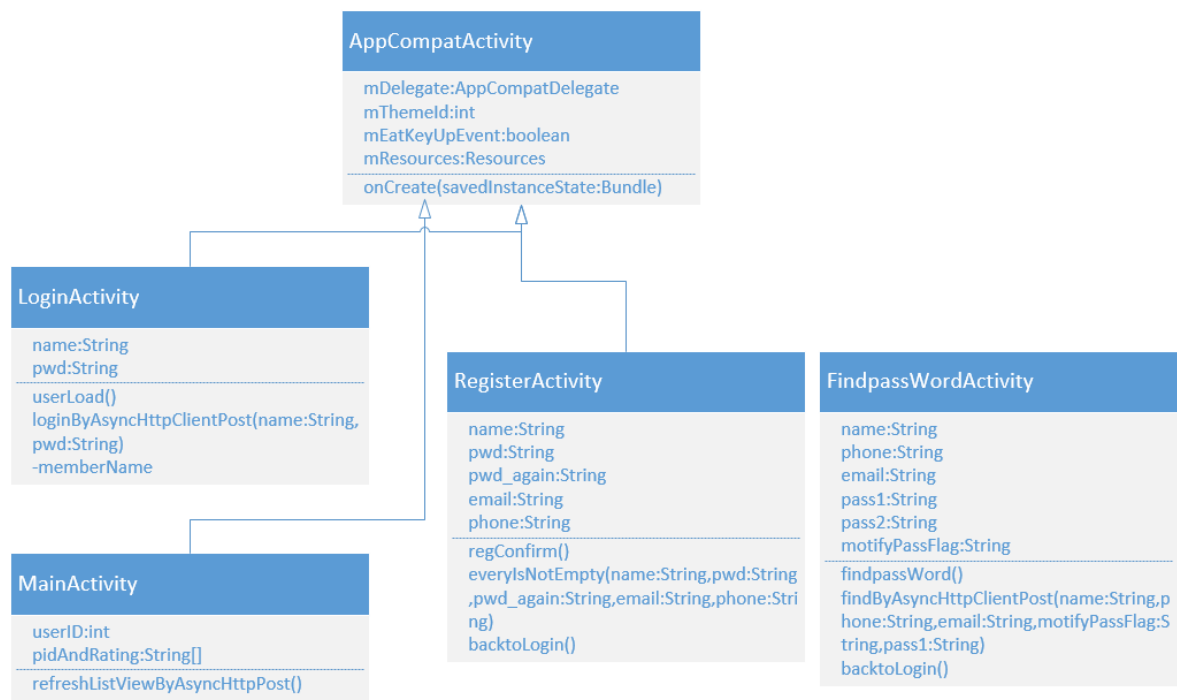


图4.3 系统 UML 类图

4.4 数据库设计

4.4.1 数据库 E-R 图

针对以上的需求分析，本系统的数据库 E-R 图如图4.4.1所示。

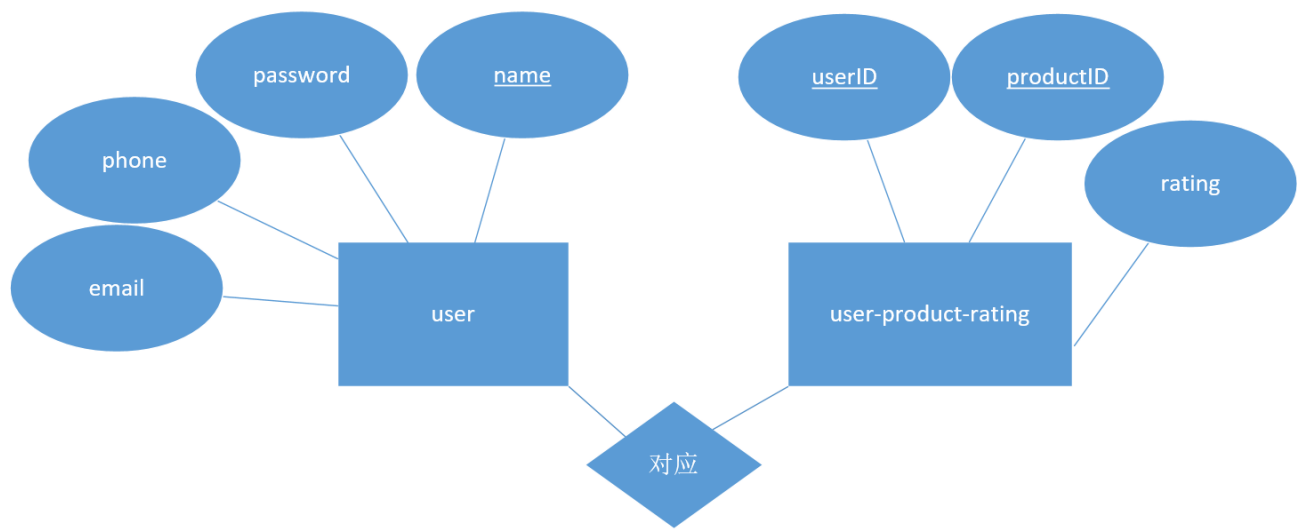


图4.4.1 系统数据库 E-R 图

4.4.2 数据库关系设计

在数据库关系的设计中，本系统可设计系统数据库用户表如表 4.4.2.1 所示和系统数据

库评分表如表 4.4.2.2 所示。

表 4.4.2.1 系统数据库用户表

Name	Password	Phone	Email
String	String	String	String

表 4.4.2.2 系统数据库评分表

UserID	ProductID	Rating
String	String	String

4.4.3 存储过程

在 MATLAB 中运行 BPMF 矩阵分解的代码后，即预处理模块后，在 MATLAB 中生成一个矩阵文件 `pmf_weight.mat`，里面包含有 `w1_M1(3952x10 double)` 和 `w1_P1(6040x10 double)` 这两个矩阵，分别表示商品矩阵和用户矩阵。3952 是商品数，6040 是用户数，这里的 10 表示的是商品/用户的隐含特征因子个数。`w1_M1` 矩阵的每一行表示的是一件商品，及其各个隐含特征因子的值，表示的是该件商品对于各个隐含特征因子的包含度。`w1_P1` 矩阵的每一行表示的是一个用户，及其各个隐含特征因子的值，表示的是用户对于各个隐含因子的喜好度。为了得到用户-商品-评分矩阵，将用户矩阵 `w1_P1` 的行向量（1x10 double）乘以商品矩阵 `w1_M1` 的行向量的转置（10x1 double），得到具体的该用户对应该商品的一个预测评分值。接下来，我们将生成得到的矩阵 `result_rating.mat` 文件。运算得到生成评分矩阵的代码如下：

```
num_m = 3952; %number of products
num_p = 6040; %number of users
num_raw = num_m*num_p;
num_col = 3;
rating_matrix = zeros(num_raw,3); %d {userid,productid,rating}
ra_raw_cur = 1;
for iStart = 1:num_p
    for jStart = 1:num_m
        rating = double( sum(w1_P1(iStart,:) * (w1_M1(jStart,:))) );
        rating_matrix(ra_raw_cur,1) = iStart;
        rating_matrix(ra_raw_cur,2) = jStart;
        rating_matrix(ra_raw_cur,3) = rating;
        ra_raw_cur = ra_raw_cur + 1;
    end
end
```

```
end  
  
end  
  
if (iStart*jStart == num_raw)  
    save result_rating rating_matrix  
end
```

至此，在 MATLAB 中，生成了 result_rating.mat 文件，里面包含有 rating_matrix (23870080x3 double) 矩阵，由于矩阵过大，不能将其直接另存为一个 Excel 表格。于是将矩阵分割保存在 60 个 Excel 文件中。分割保存预测评分矩阵代码如下：

```
str1 = 'C:\Users\张振源\Documents\MATLAB\storeAddress\user';  
userid = 1;  
str2 = '_matrix.xlsx';  
num_m = 3952; %number of movies  
num_p = 6040; %number of users  
num_p1 = 6000;  
matr_test = zeros(num_m*100,3);  
for iStart = 1:num_p1/100  
    for jStart = 1:num_m*100  
        matr_test(jStart,:) = rating_matrix((iStart-1)*num_m*100+jStart,:);  
    end  
    userid = iStart;  
    useridStr = num2str(userid);  
    storeAddress = strcat(str1,useridStr);  
    storeAddress = strcat(storeAddress,str2);  
    success = xlswrite(storeAddress,matr_test,"");  
    strdisp = 'this is document:'  
    strdisp = strcat(strdisp,num2str(iStart));  
    disp(strdisp);  
end  
matr_add = zeros(num_m*40,3);
```

```
for addi = 1:40*num_m  
    matr_add(addi,:) = rating_matrix(6000*num_m+addi,:);  
end  
success1=xlswrite('C:\Users\张振源\Documents\MATLAB\storeAddress\useradd.xlsx',matr_add,'');  
disp('over')
```

至此，MATLAB 中的预测评分矩阵已成功保存在 Excel 中。接下来我们需要将这些 Excel 文件的第一行加上列名 `userid`, `productid`, `rating`。再将这些 `xlsx` 格式的 Excel 文件另存为 `csv` 格式的文件。

最后再将这些 Excel 文件导入 MongoDB 数据库中，在 CMD 窗口下用以下命令进行导入（为了表示简便，已将 Excel 文件移至 E 盘根目录下）：

```
mongoimport -d mySystem -c myRating --type csv --headerline --file "E:\user1_matrix.csv"
```

命令中的 `-d` 后接本系统数据库名，`-c` 后接该数据库里的集合名，`--type` 后接导入的文件类型，`--headline` 表示将表格中的第一行作为数据库文档中的属性名，相当于 MySQL 数据库里的列名。

4.5 具体实现

4.5.1 客户端文件组织

在该系统的文件组织设计中，具体文件组织形式，如图 4.5.1 所示。Activity 是 Android 应用中负责与用户交互的组件，通过 `setContentView(View)` 来显示指定组件。Activity 为 Android 应用提供了可视化用户界面，如果该 Android 应用需要多个用户界面，那么这个 Android 应用将会包含多个 Activity，多个 Activity 组成 Activity 栈，当前活动的 Activity 位于栈顶。LoginActivity.java 运行登录界面的 Activity，用以显示布局文件 `activity_login.xml` 中指定的组件和用以与服务器上的 web 的 `/servlet/DBUtil?` 建立连接。RegisterActivity.java 运行注册界面的 Activity，用以显示布局文件 `activity_register.xml` 和用以与服务器上的 web 的 `/servlet/DBUtilReg?` 建立连接。content_main.xml 是显示推荐商品内容的列表的布局文件，包括在 `app_bar_main.xml` 中，nav_header_main.xml 是导航栏布局文件，在主界面的屏幕左侧往右滑就能使其显现出来。MainActivity.java 运行推荐商品内容主界面的 Activity，用以显示布局文件 `activity_main.xml` 中指定的组件和用以服务器上 web 的 `/servlet/ProductRecommend?` 建立连接。FindpassWordActivity.java 运行找回/修改密码的界面的 Activity，用以显示布局文件 `activity_findpassword.xml` 和用以与服务器上的 web 的 `/servlet/FindPass?` 建立连接。

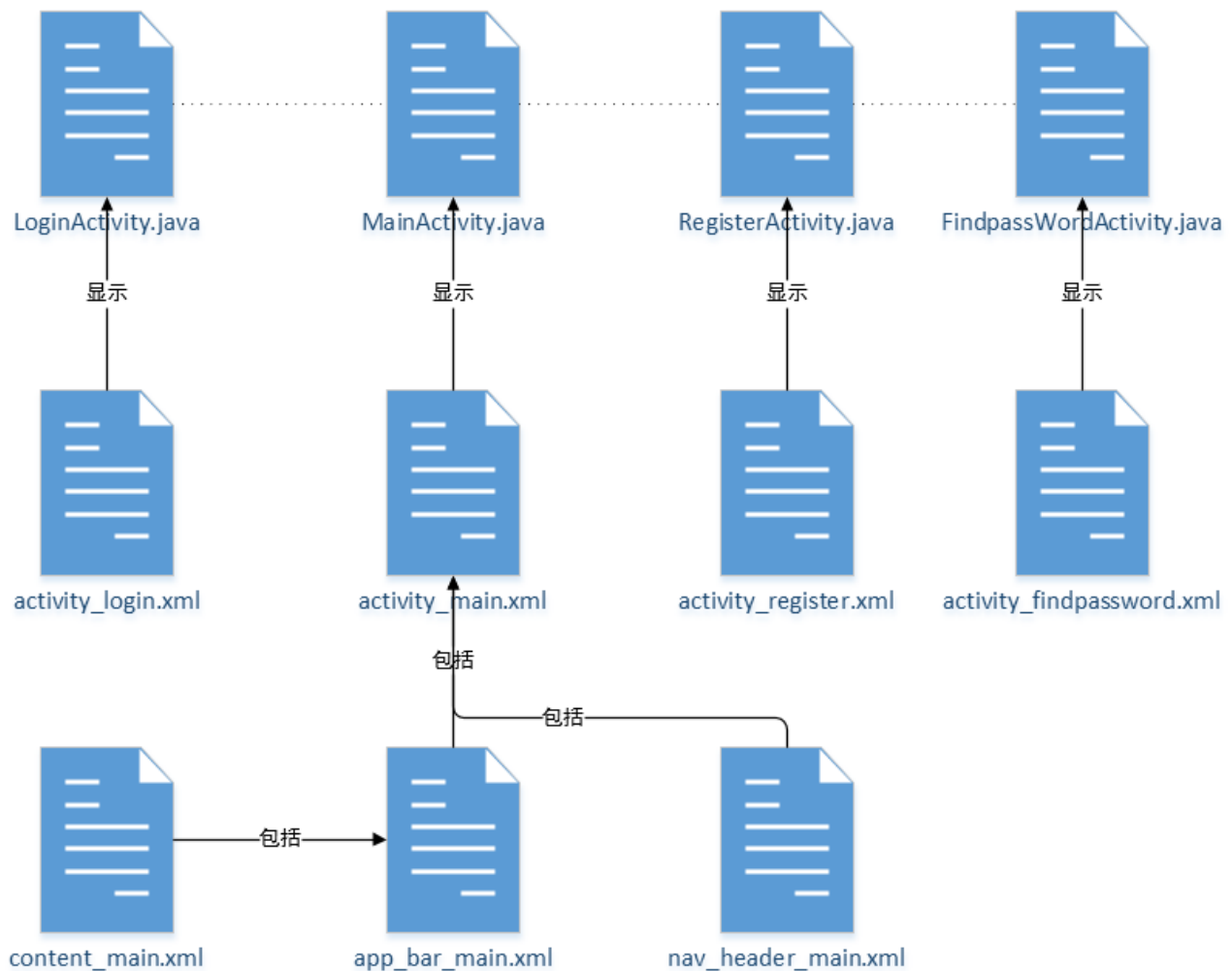


图4. 5. 1 系统组件图

4. 5. 2 算法

本系统在预处理模块中使用了 BPMF 算法。所谓 BPMF 算法就是在 PMF 算法^[4]的基础上用贝叶斯估计的方法去估计基础 PMF 中的超参数，使降低过拟合现象，使精度有所提高，BPMF 是 PMF 的一个优化方案。在介绍 BPMF 之前，先介绍 PMF 算法。

在此，我们设有 N 个用户， M 部电影。一个评分系统可以用 $N \times M$ 矩阵 R 来表示。推荐系统问题如下： R 矩阵中只有部分元素是已知的（用户只给一部分电影打过分），且 R 往往非常稀疏，要求出 R 缺失的部分。而用户和电影之间的关系（即用户对电影的偏好）可以由较少的几个因素的线性组合决定。用矩阵语言来描述，就是评分矩阵可以分解为两个低维矩阵的乘积 $R=U^T V$ ，其中 $D \times N$ 矩阵 U 描述 N 个用户的属性， $D \times M$ 矩阵 V 描述 M 部电影的属性。实际上，由于系统噪音存在，不可能做出这样的完美分解，另外 R 包含很多未知元素。

用贝叶斯观点来说, R 是观测到的值, U, V 描述了系统的内部特征, 是需要估计的^[8]。

BPMF 的模型和与 PMF 模型类似, 但在求解时, 是从贝叶斯角度而不是传统概率角度出发, 不再把系统参数当做一个固定值估计, 而是作为一个服从某种分布的随机变量, 转而估计该分布的参数^[9]。

4.5.3 服务器端实现

在服务器端, 用 MyEclipse10 编写 Web Project, 命名为 MongoDBcon, 其文件组织如图 4.5.3 所示。

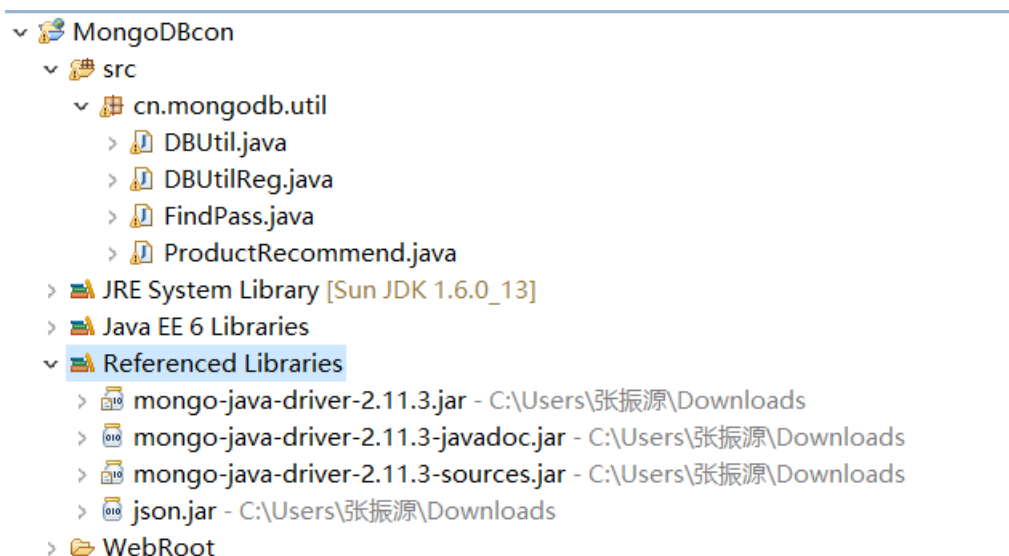


图 4.5.3 服务器端文件组织

在建好 Web Project 之后, 需要导入 MongoDB 对 java 支持的驱动包, DBUtil.java 文件是用以支持客户端的登录界面对于 MongoDB 数据库的查询操作, 从客户端获取到相应的参数后进行登录查询, 如查询账号/密码是否正确等, 再反馈给客户端相应的应答语句。

DBUtilReg.java 文件是用以支持客户端的注册账号界面对于 MongoDB 数据库的查询及增加记录操作。在客户端输入完想要注册的账号及其他相关信息之后, 服务器获得相应的参数值, 若当想要注册的账号已存在时, 或者可以成功注册此账号时, 服务器端都会在通过 `response.getWriter.write()` 中反馈给客户端。

FindPass.java 文件是用以支持客户端的密码找回和密码修改界面对于 MongoDB 数据库的查询和更新操作。在客户端输入注册时的账号、手机号和邮箱, 三者均正确时, 用户可选择找回旧密码, 如果填写了新密码, 则是直接将旧密码放弃, 修改成了新密码。比如用 `String userName = request.getParameter("username")` 从客户端获取相应的用户名参数等, 根据参数名/值的, 接下来服务器进行相应的反馈。

ProductRecommend.java 是用以支持客户端的商品推荐界面商品列表刷新功能的实现,

其对 MongoDB 的商品 collection 集合进行 sort 降序排序。比如用 `DBCursor cursor = collection.find(query,field).limit(limitDocument).sort(new BasicDBObject("rating",-1))`。其中 -1 表示将结果集降序排列。

4.5.4 客户端实现

1) 用户管理。包括有用户登录、注册、密码找回/修改三大功能。如图 4.5.4.1 用户登录界面, 图 4.5.4.2 用户注册界面, 图 4.5.4.3 密码找回/修改界面。为了实现如图 4.5.4.2 中的将“请输入用户名”这句提示语在编辑框上方缩小显示功能, 需在 `activity_register.xml` 文件中的 `EditText` 控件外部用

```
<android.support.design.widget.TextInputLayout
```

```
</android.support.design.widget.TextInputLayout>
```

包起来, 而且一个 `TextInputLayout` 只能包含一个 `EditText` 控件, 另外, 在 `LoginActivity` 需要继承 `AppCompatActivity` 方能实现此功能。

在图 4.5.4.3 密码找回界面中, 当输入的用户名存在且手机号和邮箱均正确时, 点击找回/修改密码按钮后, 密码会以半透明对话框的形式显示出来。若未在下区域填写新密码, 则直接找回旧密码, 若两次填写新密码且两次填写相同, 则也会以对话框的形式告知用户密码以使用户确认。



图 4.5.4.1 用户登录



图 4.5.4.2 用户注册



图 4.5.4.3 找回/修改密码

2) 商品推荐。如图 4.5.4.4 商品推荐列表所示。在商品推荐这个界面中, 我们可以看到, 其上方的工具栏右侧有一按钮, 单击后可看见修改密码功能按钮和刷新商品列表按钮。

若单击修改密码，则会进入用户管理的修改/找回密码的 Activity，与图 4.5.4.3 无异。刷新商品列表就是将数据库中的预测该用户评分值前十的商品排列出来。客户端关键代码如下：

```
private LinkedList<String> data = new LinkedList<String>();  
for(int ij=0; ij<pidAndRating.length; ij++) {  
    data.add("productid:"+pidAndRating[ij]+" "+"rating:"+pidAndRating[ij+1]);  
    ij++;  
}  
arrayAdapter = new ArrayAdapter(MainActivity.this, android.R.layout.simple_list_item_1, data);  
listView.setAdapter(arrayAdapter);
```

每一次刷新列表，都需要先将 data 里的数值清空，用 data.clear()语句清空数据。

另外，如图 4.5.4.5 导航抽屉视图所示，显示用户的头像，用户名和用户的邮箱。用户名和用户的邮箱是从登录界面在数据库查询后得到的用户信息中获得的。即从 LoginActivity 传递数值给 MainActivity，此处用到了 Intent 传递数值的方法。如在 LoginActivity 中编写 intent.putExtra()方法，在 MainActivity 中编写 Bundle bundle = intent.getExtras(); String nav_username = bundle.getString("username22");即可在 MainActivity 中得到数值。接下来，还得绑定 NavigationView 中的 TextView 控件的 ID，不同于一般的控件 ID 绑定，首先得先绑定 NavigationView 控件的 ID，再从其控件容器中查找绑定其中的 TextView 控件。如用以下代码设置 TextView 控件显示文本内容：

```
nav_navi = (NavigationView)findViewById(R.id.nav_view);  
tv_nav_username =  
(TextView)nav_navi.getHeaderView(0).findViewById(R.id.nav_header_username);  
tv_nav_email = (TextView)nav_navi.getHeaderView(0).findViewById(R.id.nav_header_email);  
tv_nav_username.setText(nav_username);  
tv_nav_email.setText(nav_email);
```

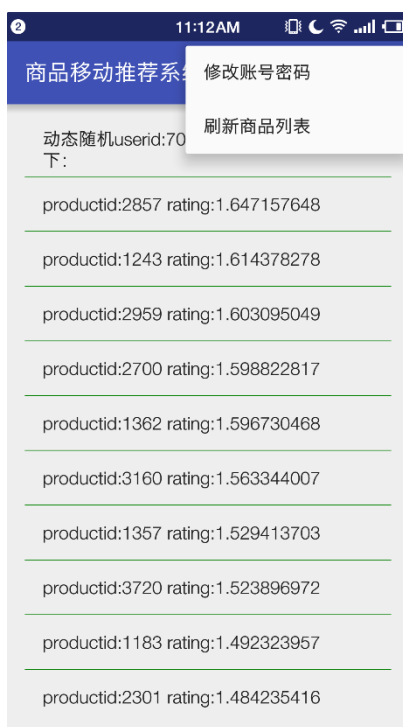


图 4.5.4.4 商品推荐界面

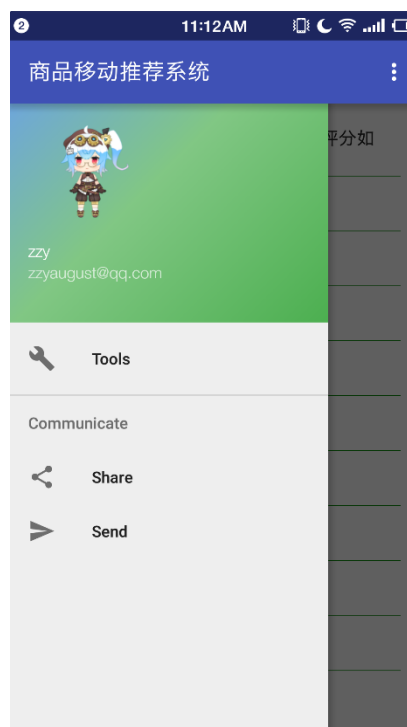


图 4.5.4.5 导航抽屉视图

4.5.5 Android 真机连接测试

此处 Android 真机测试，首先将手提电脑(服务器)上的猎豹无线 WIFI 打开，打开 CMD，运行 ipconfig 命令，查看无线适配器的 IPv4 地址，如图 4.5.5.1 无线局域网适配器 ip 所示。

无线局域网适配器 本地连接* 3:

```
连接特定的 DNS 后缀 . . . . . :  
本地链接 IPv6 地址. . . . . : fe80::d061:d22c:7857:3b84%19  
IPv4 地址 . . . . . : 192.168.191.1  
子网掩码 . . . . . : 255.255.255.0  
默认网关. . . . . :
```

图 4.5.5.1 无线局域网适配器 ip

将 Android Studio 中连接服务器的 url 中的 ip 改成图 4.5.5.1 中对应的地址：192.168.191.1，然后将 Android 项目打包下载到真机上。在服务器端，将 MongoDB 服务器打开，运行 CMD，输入 mongod 回车，运行成功。在服务器上确定将 Web Program 部署在 Tomcat 后，运行 Tomcat（此处可用 MyEclipse 打开 Tomcat 也可用 CMD 的方法运行，建议用 CMD 的方法，关闭 MyEclipse 以节省内存的使用）。打开真机上的浏览器，输入 http://192.168.191.1:8080 后回车，可看见如图 4.5.5.2 真机访问 Tomcat 所示。当然也可以对服务器中编写好的 doGet 方法测试，在浏览器中输入 http://192.168.191.1:8080/MongoDBcon/servlet/DBUtil?username=zzy&userpass=zzq 后回车，即可如图 4.5.5.3 真机测试 doGet 方法所示。

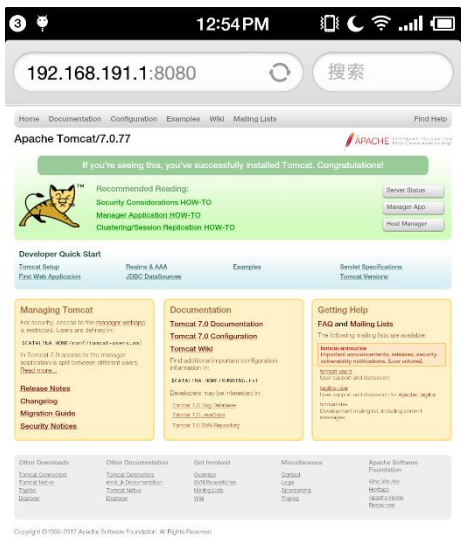


图 4.5.5.2 真机访问 Tomcat 主页

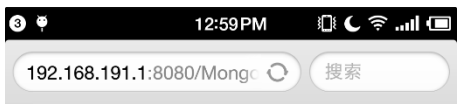


图 4.5.5.3 真机测试 doGet 方法

至此，真机测试成功，打开商品推荐系统 APP 自然也是能成功登陆运行的。另外，在服务器上的浏览器也是能成功访问 Tomcat 主页的。在建工程的过程中，用 doGet 的方法测试，可使程序的测试更加便利，减少程序 BUG 的数量和减少了程序调试的时间。

结 论

为了让用户有效地节省在购买商品的过程中挑选商品的时间，本文的研究目的采用 Netflix prize 公开的数据集，使用 BPMF 算法进行预处理，对移动终端的用户进行商品推荐服务。

根据文中的需求分析，完成了用 MATLAB 进行数据集的预处理、服务器端 Web Project 的开发和客户端 Android 工程的开发。采用客户端（移动终端）到服务器到数据库间接访问数据库的方式，也可以节省移动终端的流量和电量。

在服务器端的实现部分，本文着重研究了服务器连接、操作 MongoDB 数据库的方法以及连接、获取 Android 端请求的方法。服务器端的实现，正是商品移动推荐系统的后台实现。

在 Android 端的实现部分，本文着重研究了 Android 端连接服务器、与服务器进行通信的方法以及对 Android 端 UI 界面的美化，及非需求功能的完善。

本文所涉及的商品移动推荐系统在推荐商品给特定用户的服务让用户较为满意，但还可以继续改进，可改进的地方如下：

- （1）采用商品种类数量更加大、用户更多的数据集以便商品移动推荐系统的测试；
- （2）采用更加成熟和更加优秀的商品推荐算法，使得用户对于推荐的商品更加满意；
- （3）在 Android 端的用户管理模块加入短信通知或邮件通知用户，告知用户所推荐的商品的详细信息；
- （4）可收集 Android 端用户对于现所推荐的商品进行反馈并做出评分，在服务器端接收反馈数据，在后台动态修改数据库内容，最好能进行实时商品推荐。

Android 平台的移动终端是很普及的，几乎是人手一部。因此若在 Android 移动终端上进行商品推荐，则是具有很大的推广市场。同时，社会生活节奏的加快，以及 Android 移动终端携带的便捷性，使人们愿意使用这样的商品移动推荐系统，为生活购物等提供很大的便利。

致 谢

历时将近四个月的时间终于完成了本篇论文，我在做此工程项目和写作过程中遇到了很多的障碍和苦难，正是导师李泽超老师耐心的指导和不畏困难的精神感动了我，使我无畏地做下去，勇于挑战困难。李老师的博学才多、求实工作的作风以及丰富的实战经验潜移默化地影响了我。从最初的开题，到中期考核，到最终的完成，李老师经常在百忙之中回复我的邮件，也经常现场给与我耐心的指导，对我的研究起到了很大的帮助，使我在学习 Android 和设计商品推荐系统的过程中受益匪浅。在此向我的导师致以最诚挚的感谢和最衷心的祝福。

还要感谢我的母校南京理工大学，感谢母校提供的良好的学习氛围和硬件设施，无论是自习室和图书馆安静的学习环境，还是中国知网、万方论文的免费下载阅读等，都对我的研究提供了很多帮助。

最后，感谢我的家人，以及所有关心我、鼓励过我、帮助过我的朋友们！

参 考 文 献

- [1] 中国网络零售行业发展情况分析报告,
<http://www.chinairn.com/news/20130322/163418999.html>
- [2] 百度百科. MATLAB. <http://baike.baidu.com/item/MATLAB>
- [3] Brad Dayley 著, 米爱中译. MongoDB 入门经典[M]. 北京: 人民邮电出版社, 2015.
- [4] Ruslan Salakhutdinov, Andriy Mnih. Probabilistic Matrix Factorization. NIPS, 2007.
- [5] 项亮. 推荐系统实践. 人民邮电出版社. 2012 年 6 月
- [6] 李刚. 疯狂 Android 讲义[M]. 北京: 电子工业出版社, 2015 年 6 月
- [7] 百度百科. MyEclipse. <http://baike.baidu.com/item/MyEclipse>
- [8] 推荐系统算法 PMF,
<http://blog.csdn.net/shenxiaolul984/article/details/50372909>
- [9] 推荐系统算法 BPMF,
<http://blog.csdn.net/shenxiaolul984/article/details/50405659>
- [10] 李大为. 基于 Hadoop 的推荐系统研究与开发[D]. 北京: 华北电力大学, 2015
- [11] 博客园. 推荐系统领域的相关会议与期刊,
<http://www.cnblogs.com/guolei/archive/2013/02/21/2856397.html>
- [12] 开源项目大全. Android 网络请求框架 asyncHttp,
<https://www.oschina.net/p/asynchttp-for-android>