

浙 江 大 学

本 科 生 毕 业 设 计 开 题 报 告



学生姓名：_____王莹_____

学生学号：_____3110103009_____

指导教师：_____郑小林_____

年级与专业：_____11 级软件工程_____

所在学院：_____计算机学院_____

一、题目： 电子产品在线交易平台的设计与实现

二、指导教师对开题报告、外文翻译和中期报告的具体要求:

外文翻译要求选择与电子商务平台相关的经典文献，翻译必须做到语句通顺，语义贴切。

在此基础上，开题报告要提出在线交易平台的设计与实现的具体方案，提出可行的技术路线，以及合理的研发计划。

指导教师（签名）_____

年 月 日

毕业设计开题报告、外文翻译和中期报告考核

导师对开题报告、外文翻译和中期报告评语及成绩评定：

王莹同学在阅读大量国内外相关文献的基础上，对电子商务平台的国内外研究现状以及存在的问题进行了深入的分析，外文翻译语义贴切，语句通顺。

在此基础上完成的开题报告，拟围绕电子科技产品在线平台展开研究，选题具有较好的理论意义和应用价值。开题报告中提出的研究目标明确，技术路线可行，研发计划合理。

成绩比例	开题报告 占（20%）	外文翻译 占（10%）	中期报告 占（10%）
分 值			

导师签名_____

年 月 日

答辩小组对开题报告、外文翻译和中期报告评语及成绩评定：

成绩比例	开题报告 占（20%）	外文翻译 占（10%）	中期报告 占（10%）
分 值			

开题报告答辩小组负责人（签名）_____

年 月 日

目录

- 1. 项目背景与意义 6
 - 1.1 项目名称 6
 - 1.2 项目背景 6
 - 1.3 意义与应用价值 7
- 2. 目标和任务 7
 - 2.1 目标 7
 - 2.2 任务 7
 - 2.2.1 Common 7
 - 2.2.2 B2C 7
 - 2.2.3 Admin 8
- 3. 可行性分析 8
 - 3.1 市场可行性 8
 - 3.2 技术可行性 8
 - 3.3 风险因素控制的可行性 8
 - 3.4 可行性分析结论 9
- 4. 初步技术方案和关键技术考虑 9
 - 4.1 项目总体构架分析 9
 - 4.2 Bootstrap+jQuery 9
 - 4.3 SpringMVC 9
 - 4.4 Hibernate 10
- 5. 预期工作结果 10
 - 5.1 功能实现 10
 - 5.2 测试完备 10
 - 5.3 发布成功 10
- 6. 进度计划 10
- 7. 参考文献 11

1. 项目背景与意义

1.1 项目名称

电子科技产品在线交易平台的设计与实现

1.2 项目背景

在信息时代，物美价廉的低附加值产业已渐渐失去其经济主力军的地位，高新电子信息产业已经异军崛起势不可挡。新兴国家市场将成为带动全球电子信息产业发展的新引擎，而中国有望在 3 到 5 年之内超过美国，成为全球第一大电子产品市场。开展网上销售，具有投入少、风险低、资金周转快、容易把握市场热点等，是企业转型的必然选择。

电子商务平台缩短了制造企业与终端消费者之间的距离，有利于制造企业更好的把握终端消费者的需求，设计更满足市场的产品，提高产品销量，从而大大提升企业利润。近年来，依靠电子商务平台，使行业内各企业间信息交流跨出了地区的限制，使各种信息更加快速、通畅的在国内各地传播，且传播费用极低，从而盘活了多个行业。把握住传统采购、分销渠道的同时，积极开拓并把握住发展迅速的电子商务市场，是制造企业得以快速发展的一个新的机遇。

我们所耳熟能详的电商平台，诸如美国的 eBay 和亚马逊，日本的乐天，还有我们国家的淘宝网，都已经发展为以开拓全面市场的准则(亚马逊早已不再仅仅是第一大网络书店)，涵盖了广大的销售商品范围，其顾客也没有年龄、性别、职业的界限。至于市场定位较为单一的商家，例如聚美优品、唯品会等，它们的面向用户有着较为明显的偏向与特点，在其销售产品的领域容易形成垄断的趋势，然而顾客源较单一却是不可否认的缺点。

如今，进入到电商行业的企业与日俱增，在同行业竞争压力不断增大的情况下，对电子商务的运营模式也不断革新来获得鲜活的生命力。大数据时代的到来，Web2.0 和移动互联技术的迅猛发展，移动电子商务发展迅速发展。碎片化时间的“闲逛”是消费者选择移动购物的主要原因之一。消费者在床上、在厕所、在地铁里、在等人过程中等任意碎片时间，为了打发无聊时光，在手机的电商 APP 中“闲逛”，不经意间发现一件很中意的产品，不假思索立刻在手机上购买。移动电商的崛起和成功毫无疑问是对传统电商的颠覆性“革命”，似乎已是大势所趋。

与移动电子商务相伴的是社会化电商的出现。社会化电子商务是电子商务的一种新的衍生模式。它借助社交媒介、网络媒介的传播途径，通过社交互动、用户自生内容等手段来辅助商品的购买和销售行为。在社会化电商中，除了产品本身的品质价格之外，用户的主观评分评价成为了其他用户选购商品的重要参考因素之一。深度分析社会化媒体上存在的海量的用户生成内容，挖掘这些信息中用户需求、观点和看法，对于电商的精准营销具有重要的指导意义。

一份来自尼尔森公司的调查报告显示，在中国，不论是网上浏览还是网上购物的渗透率都大大地超过了全球平均水平，中国的电子商务已经领跑全球。电商企业百舸争流的时代，认清当今的产业形势，深入剖析消费者心理，研究电商的发展方向，对于开发电商平台意义重大。

电子技术是十九世纪末、二十世纪初开始发展起来的新兴技术，二十世纪发展最迅速，应用最广泛，成为近代科学技术发展的一个重要标志。电子科技产品与传统产品相比，具有生命周期短、更新速率快、RD 强度大等特点，正是由于这些导致了开发高科技产品的高效益与高风险的伴随性、共生性。同时，电子科技产品并不同于廉价小商品，消费者在购置电子科技产品时更注重其质量和售后服务，因此在研发电子科技产品的交易平台时不仅需要对电商行业做深入调研，还需要针对电子产品的消费特点“量体裁衣”。

1.3 意义与应用价值

我们为电子科技产品量身定制并推出一个具有公信力的公共电子商务服务平台。在发展经济，加快行业发展的同时，又创建了一个高科技的电子商务平台，有利的拉动了相关地区的就业率，为地区的发展和稳定做出了相当大的贡献。这种结合政府牵头、高等院校智慧投入、企业参与产生的平台在电子商务领域是一种非常具有发展远景的创新。因电子商务具有跨地域性，在互联网上推广传统业务的交易模式，不仅在国内具有非常大的作用和效果，同样通过网上的信息展示，也能增进国内企业与国外的交流，提供更多让国内企业走出去的通道，从而进一步推动整合行业的发展，促进传统制造企业的升级转型。

2. 目标和任务

2.1 目标

构建一站式电子产品在线交易平台和现代物流服务平台，实现 B2B、B2C 交易模式，建立用户信用模型和适用电子商务平台的个性化推荐系统。

2.2 任务

为了便于开发和团队协作，我们将电子科技产品交易平台的开发分成 Common、B2C、Admin 三个部分。

2.2.1 Common

Common 指的是交易平台中的一些基础功能模块，包括注册认证登录模块、站内通信模块、内容管理模块、推荐模块。

- 注册认证登录模块：个人用户和企业用户的注册和登录，用户需要身份（营业执照）的认证，用户登录时若忘记密码可根据认证邮箱找回。
- 站内通信模块：站内用户之间的留言和私信联系。
- 内容管理模块：内容的添加编辑删除等，在内容上尽可能多样化，避免单一文字最好能加入视频元素。
- 推荐模块：推荐模块分为两个部分，一个部分是全局推荐流程，另一个是根据用户购买浏览评论记录进行个性化推荐。

2.2.2 B2C

B2C 交易模式是个人与企业之间的交易，包括卖家管理和买家管理。

- 买家管理：包括了基本信息管理、购物、订单管理、留言管理、退货管理、个人收藏等。

- 卖家管理：商品管理、店铺管理、店铺主题管理、物流管理、订单管理、退货管理。在基本的电商系统的框架上，我们为店铺设置了店铺主题管理这个流程，希望卖家能够适时地对店铺进行美化，以作为一种吸引买家的手段，同时也是满足买卖双方的一种审美需求。

2.2.3 Admin

Admin 部分是管理员模块，其中全局统计信息展示模块、用户管理模块、产品类别管理模块、店铺管理模块、订单查看模块、支付管理模块、物流管理模块、新闻管理模块、推荐管理模块、主题管理模块。

3. 可行性分析

3.1 市场可行性

虽然当下有许多电商平台，但是，我们的平台有明确的目标定位——电子科技产品。在这一范围内我们的平台将会有比较大的专业专一优势。同时，广东科技局将会配合我们项目组，对深圳的电子科技产品厂商们进行平台的推广，他们将成为这个平台的首批用户并进而为平台拓宽宣传渠道。

3.2 技术可行性

Java 是如今最流行的开发语言之一，它拥有一套庞大且完善的类库，内置了其他语言需要的类库和对操作系统功能的支持，并且 Java 开发团队对其的改善也日臻完美。MySQL 是一种关联数据库管理系统，关联数据库将数据保存在不同的表中，而不是将所有数据放在一个大仓库内，这样就增加了速度并提高了灵活性。在前端方面 jQuery 兼容各种浏览器，能够实现多样化动画效果并提供交互。

本平台的开发基于 J2EE 框架，J2EE 是一种利用 Java 语言的标准体系结构定义，利用它可以更为方便地在中间层加速分布式部署。J2EE 规范定义了如下组件：

客户端应用程序和 applet 是运行在客户端的组件；Java Servlet 和 Java Server Pages (JSP) 是运行在服务器端的 Web 组件；Enterprise Java Bean (EJB) 组件是运行在服务器端的业务组件。随着电子商务时代的到来，J2EE 已经成为开发电子商务应用的新型平台。它通过提供企业计算环境所必须的各种服务且更加易用，使得部署在 J2EE 平台上的多层应用可以实现更高的可用性、安全性、可扩展性。

3.3 风险因素控制的可行性

- 尽量保证准确或到位的成本和行动从而使项目规划更加准确，降低成本
- 期限不要过紧，保证项目质量
- 项目领导人应擅长技术工作，能界定、规划、评估、跟踪、传达以及控制项目，保证项目进行的高效率
- 项目规模本身就是风险，项目不要过大，根据实际情况确定项目规模
- 项目需要细致的编制文件，从而使整个团队具备确认项目计划的基础

- 增进团队协作，完成项目需要团队协作、信任以及成员之间互助的意愿

3.4 可行性分析结论

项目符合企业的发展方向，项目的功能非常实用且准备工作充分，项目风险以得到很好的控制和规避，项目可以进行建设。

4. 初步技术方案和关键技术考虑

4.1 项目总体构架分析

本系统主要以 Java 作为系统开发语言，以 bootstrap+jQuery 为设计框架开发前端界面，将用 Spring 框架和 Hibernate 技术将数据库隔离，从而独立出数据层，可以方便系统在不同的数据库之间的切换，这也就是本系统的 model 层。同时在 Control 层里通过 Struts 和 SpringSide 和 Spring Security 等技术在 Model 和 View 层中传递数据。最后在 View 层用 JSP 和 DWR 等技术同 Control 层进行交互。这样的层次结构，方便各层之间的接口设计，也方便今后系统的移植和维护。下文将对应用的主要技术进行介绍。

4.2 Bootstrap+jQuery

jQuery 是一套跨浏览器的 JavaScript 库，简化 HTML 与 JavaScript 之间的操作。jQuery 的语法设计使得许多操作变得容易，如操作文档对象（document）、选择 DOM 元素、创建动画效果、处理事件、以及开发 Ajax 程序。jQuery 也提供了给开发人员在其上创建插件的能力。这使开发人员可以对底层交互与动画、高级效果和高级主题化的组件进行抽象化。模块化的方式使 jQuery 函数库能够创建功能强大的动态网页以及网络应用程序。

Bootstrap 是 Twitter 推出的一个开源的用于前端开发的工具包。它由 Twitter 的设计师 Mark Otto 和 Jacob Thornton 合作开发，是一个 CSS/HTML 框架。Bootstrap 提供了优雅的 HTML 和 CSS 规范，它即是由动态 CSS 语言 Less 写成。Bootstrap 是基于 jQuery 框架开发的，它在 jQuery 框架的基础上进行了更为个性化和人性化的完善，形成一套自己独有的网站风格，并兼容大部分 jQuery 插件。

4.3 SpringMVC

Spring MVC 属于 SpringFrameWork 的后续产品，已经融合在 Spring Web Flow 里面。Spring 框架提供了构建 Web 应用程序的全功能 MVC 模块。使用 Spring 可插入的 MVC 架构，可以选择是使用内置的 Spring Web 框架还可以是 Struts 这样的 Web 框架。通过策略接口，Spring 框架是高度可配置的，而且包含多种视图技术，例如 JavaServer Pages (JSP) 技术、Velocity、Tiles、iText 和 POI。Spring MVC 框架并不知道使用的视图，所以不会强迫您只使用 JSP 技术。Spring MVC 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进行定制。

4.4 Hibernate

Hibernate 是一个开放源代码的对象关系映射框架，它对 JDBC 进行了非常轻量级的对象封装，使得 Java 程序员可以随心所欲的使用对象编程思维来操纵数据库。Hibernate 可以应用在任何使用 JDBC 的场合，既可以在 Java 的客户端程序使用，也可以在 Servlet/JSP 的 Web 应用中使用，最具革命意义的是，Hibernate 可以在应用 EJB 的 J2EE 架构中取代 CMP，完成数据持久化的重任。

Hibernate 的核心接口一共有 6 个，分别为:Session、SessionFactory、Transaction、Query、Criteria 和 Configuration。这 6 个核心接口在任何开发中都会用到。通过这些接口，不仅可以对持久化对象进行存取，还能够进行事务控制。

5. 预期工作结果

5.1 功能实现

实现 2.2 任务章节中描述的各个功能模块，正确无误的按照需求完成设计。

5.2 测试完备

- 完整的测试用例，完备的测试计划：根据项目进度指定出合理的测试计划，并根据功能编写覆盖完整的测试用例
- 开发者功能性测试：开发者根据测试用例完成交叉测试，确保各自的功能实现符合要求
- 系统集成测试：完成各个模块的集成，并进行系统集成测试，确保整个应用运行正常，符合用户要求
- 用户验收测试：编写详细的用户手册，并将应用提交给用户进行验收测试，并提供实时的支持，从而不断的改进应用

5.3 发布成功

制定详细的应用发布计划，在进行完测试后，将应用发布供电子产品厂商和个人消费者使用，完成应用的第一个版本的开发。

6. 进度计划

毕业设计将大致分为四个阶段：需求分析、概要设计与详细设计、系统实现与测试、毕业论文撰写。

目录	任务	开始时间	结束时间	工作日	优先级	备注
需求分析	需求分析与整理	2014/11/20	2014/12/10	20	中	

概要设计与详细设计	系统架构、功能模块详细设计	2014/12/11	2015/1/5	25	中	
	数据库设计	2014/1/6	2015/1/30	20	中	
	界面设计	2014/1/21	2015/2/20	30	中	
系统实现与测试	代码编写	2015/2/21	2015/5/5	60	中	
	系统测试	2015/4/21	2015/5/10	25	中	
毕业论文撰写	毕业论文撰写	2015/5/11	2015/5/30	15	中	

7. 参考文献

- [1] 李岩. 在线交易平台的设计与实现. 科学致富向导. 2013,(8): 47
- [2] 蔡永鸿, 刘莹. 基于大数据的电商企业管理模式研究, 中国商贸. 2014, (31): 74~75
- [3] 尼尔森公司. 中国领跑全球电商市场. 市场研究. 2014, (10): 11~12
- [4] 普索中国. 电商的“移动”物语--益普索 (Ipsos) 联合支付宝发布《2013 中国消费者如何看待移动支付报告》. 市场研究. 2013, (9): 9~10
- [5] 王洪艳. 社会化电商用户需求挖掘初探. 电子商务. 2014,(8): 33~34
- [6] 郭伟光. 我国 B2C 电子商务个性化商品推荐服务实证研究. 价值工程. 2014, (30): 25~27
- [7] 洪钟. 从电商发展看互联网未来的影响. 中国经贸导刊. 2014, (28): 48~51
- [8] 张琳琳. 浅谈垂直细分电子商务平台的运行模式. 才智. 2013, (26): 323
- [9] Xiaolin Zheng, Weifeng Ding, Jingnan Xu, Deren Chen. *Personalized recommendation based on review topics. Service Oriented Computing and Applications. Volume 8, Issue 1: 15~31*
- [10] Wenming Zuo, Qiuping Huang, Chang Fan, Zhenpeng Zhang. *Quality management of B2C e-commerce service based on human factors engineering. Electronic commerce research and applications. 2013, Volume 12, Issue 5: 309-320*
- [11] 马燕, 林传立. J2EE 在电子商务开发中的应用研究. 计算机应用研究. 2002, 19(9): 42~44
- [12] 曹鸣鹏, 赵伟, 许林英. J2EE 技术及其实现. 计算机应用. 2001, 21(10): 20~23
- [13] 张炜. 基于 J2EE 分布式架构的高性能电商交易接入平台研究与设计. 移动通信. 2014, (10): 90~96

软件工程本科生文献翻译

基于评论的个性化推荐算法研究

Personalized recommendation based on review topics

Xiaolin Zheng, Weifeng Ding, Jingnan Xu, Deren Chen

摘要

作为一个成功的个性化推荐技术,传统的协同过滤算法的核心思想在于从用户评分上计算用户相似度和产品相似度,从而就相似用户或相似产品的评分预测目标用户的评分并作出相关产品推荐。然而,在实际的应用中,我们面临数据稀缺的问题——大多数用户很少评分。因此,传统的协同过滤算法无法提供一个令人满意的推荐效果。

本文提出了一个全新的基于主题模型的协同过滤推荐机制和两个推荐算法:主题模型下基于用户的协同过滤算法(UCFTM)和基于产品的协同过滤算法(ICFTM)。基于主题模型的方法,首先对用户评论进行处理,计算评论的主题概率分布,表示用户对产品不同特征的偏好。其次在基于主题模型和用户的协同过滤算法中,综合了每个用户给出的所有评论的主题概率分布,计算用户最重视特征,表示用户最重视的产品特征。基于用户最重视特征,计算用户相似度。然后,根据相似用户的评分预测目标用户可能的评分。在基于主题模型和产品的协同过滤算法中,也综合了每个产品得到的所有评论的主题概率分布,计算产品最重视特征,表示产品最受用户重视的特征。进而基于产品最重视特征,可以计算产品相似度。最后,根据目标用户对相似产品的评分预测评分。

针对亚马逊上的六个真实数据集上实验结果表明,当大多数用户只给出一个评论和一个评分时,本文的算法相较于其他传统的协同过滤算法、解决数据稀疏性问题的算法和最新的基于主题模型的推荐算法都能取得更好的推荐效果。

关键词: 协同过滤, 推荐算法, 主题模型, 数据稀疏性

第 1 章 绪论

随着 web2.0 和社交网络的快速发展,越来越多的用户不再仅仅是信息的消费者,而更多的担当起信息的生产者。互联网上的绝大部分信息是来源于用户,每一个用户都可以产生自己的内容,这种现象也称为用户产生内容(User Generated Content, UGC),比如 Facebook、人人网、新浪微博等诸多社交应用都是用户产生内容的成功案例。

用户产生内容一方面给所有互联网使用者带来了丰富的信息,另一方面也产生了一个很严重的问题,那就是信息过载(Information Overloading):即大量的信息进入用户的大脑并已经远远超过了普通人大脑能够处理的范围。信息过载不仅造成大量的信息用户没法处理,更重要的是,大量无关的内容将会严重干扰到用户查找有用的信息。

对于一个在线购物网站,就拿亚马逊来说,解决信息过载就是要帮助它的用户找到他们所喜欢并愿意购买的产品。为了提高用户体验并促进产品销售,推荐系统

(Recommendation System) 应运而生, 进而成为在线购物网站不可或缺的一部分。

在过去十几年里, 传统的协同过滤算法 (Collaborative Filtering, CF) 得到了很大的发展和应用成为最成功的推荐技术。协同过滤是处理信息过载的一个有效方法, 它利用用户的社交网络信息来过滤掉无关的信息并产生推荐。

举个例子来说, 基于用户的协同过滤算法 (User-based Collaborative Filtering, UCF) 会搜索那些与目标用户 具有相似评分模式的用户作为相似用户, 然后利用这些兴趣相投的相似用户的评分来为预测特定产品的评分。

基于产品的协同过滤算法 (Item-based Collaborative Filtering) 则是从产品的角度出发来预测评分。它首先计算一个产品-产品的相似矩阵, 用来确定每一对产品之间的相似度。然后再根据目标用户对相似产品已有的评分, 预测目标用户对未评分产品可能的评分。

在传统的协同过滤算法中, 用户之间或者产品之间相似度的计算扮演着非常重要的角色。用来衡量这种相似性的方法有很多, 比如欧几里得相似度 (Euclidean Distance Similarity), 皮尔森相似度 (Pearson Correlation Similarity) 等等。

但是, 传统的协同过滤算法的可用性经常受限于数据稀疏性问题 (Data Sparsity) 和冷启动 (Cold Start) 问题。以基于用户的协同过滤算法为例, 如果一个用户只对很少几个产品给出评分, 或者是根本没有对任何产品给出评分的新用户, 基于用户的协同过滤算法往往没办法产生有效推荐。原因就在于此时计算用户相似度很困难。比如一个用户只对一个产品给出评分, 那么根据皮尔森相似度的定义, 此用户跟其他用户之间的相似度是没法计算的。即使用户相似度可以计算, 其结果也往往是不可靠的。因为用户之间只有少量共同评分的产品。对于基于产品的协同过滤算法来讲, 它也没办法比基于用户的协同过滤算法取得更好的推荐效果。原因是此时目标用户只有一个评分数据, 如果这个评分被用来做训练, 那么目标用户也就没有其他评分数据需要预测的了。而如果这个评分数据被用来做测试, 那么目标用户也没有其他评分数据能够用来做预测了。根据长尾理论[26], 这种数据稀疏性的问题在实际生活中是经常碰到的。

除了评分, 商家往往希望用户能够对他们所购买的产品给出评论。这些评论一般是以非结构化文本的形式保存的。当用户在决定是否购买某一产品时, 往往会参考其他用户对该产品给出的评论。

用户总是会在其对某一产品的评论中表达自己对这类产品的某种偏好。表 1 展示了一则来自亚马逊网站上一款笔记本电脑的评论例子。表中的内容描述了用户对该笔记本电脑的设计、响应速度和屏幕分辨率等几个特征比较满意, 同时还有关于屏幕特征负面的观点。需要特别注意的是, 用户能够针对产品的任何特征发表评论。这里说的产品特征不仅可以是产品固有的物理属性, 还可以是用户关注的任何对象, 比如笔记本电脑的价格、响应速度等等。

这些用户的评论对于一个在线购物网站来说, 是一笔无形的资产, 同时也同时也是商家用来挖掘潜在用户偏好、促进销售的新数据来源。在我们看来, 用户对产品给出的评论比评分更有价值, 因为评论给出了更多用户关于产品各个特征的语义信息, 体现了用户对产品不同特征的偏好, 而不仅仅只是笼统的评分。比如有些用户对笔记本电脑的屏幕更在意, 而有些用户则更注重笔记本电脑的价格。如果能够从用户的评论中理解并获取用户对产品不同特征的偏好, 就能给用户做出更加准确的推荐。

我们能利用主题模型来处理用户的评论, 并为每一个评论生成主题分布。然后基于评论的主题概率分布, 提出了两个不同的推荐算法。一个是基于主题模型和用户的协同过滤算法, 另一个是基于主题模型和产品的协同过滤算法。这两个算法都是传统协同过滤算法和主题模型方法的混合。它们在预测评分的过程中, 处理流程和传统协同过滤算法是一样的, 只是用户之间相似度的计算和产品之间相似度的计算不再是基于评分, 而是通过主题模型的方法计算得到的。

基于主题模型和用户的协同过滤算法通过评论主题概率分布，计算用户主题概率分布和用户最重视特征。然后基于用户最重视特征，计算用户相似度。最后，根据相似用户的评分预测目标用户可能的评分并产生推荐。

基于主题模型和产品的协同过滤算法通过评论主题概率分布，计算产品主题概率分布和产品最重视特征。然后基于产品最重视特征，计算产品相似度。最后，根据目标用户对相似产品的评分预测目标用户对未评分产品可能的评分并产生推荐。

本文的剩余部分构架如下：

第二章：评论发掘和协同过滤推荐的相关研究总览

第三章：基于话题模型的协同过滤框架和两个算法

第四章：六个真实数据集的实验记录和结果

第五章：总结

第五章 总结

在用户只给出一个评论和评分时，为了让个性化推荐更为准确，本文提粗了一个全新的推荐框架和两个相关算法。同时，用话题模型来解决计算用户相似度和物品相似度时的数据稀疏问题。在练歌算法中，用话题模型去处理每一个评论来生成评论主题分布。接着，我们重新设计并基于用户最重视特征重新计算用户相似度和物品相似度。最后，根据相似用户评分/相似物品评分预测目标物品评分并产生推荐。来自亚马逊网站的六个不同领域的真实数据集显示，当用户只给出一个评论和一个评分时，我们的算法比传统的协同过滤算法产生更准确的腿甲。不仅如此，在处理数据稀疏问题时，我们的算法也比其他同类基于主题模型的推荐算法有更好的表现。
