

学校编码：10384

分类号_____密级_____

学号：24320101152263

UDC_____

廈門大學

硕士学位论文

协同过滤算法在移动电子商务推荐系统中的
应用研究

Research on Application of Collaborative Filtering Algorithm
in the Mobile E-commerce Recommendation System

杨波

指导教师：陈海山教授

专业名称：软件工程

论文提交日期：2013年4月

论文答辩日期：2013年5月

学位授予日期：_____年 月

指导教师：_____

答辩委员会主席：_____

2013年 月

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果,均在文中以适当方式明确标明,并符合法律规范和《厦门大学研究生学术活动规范(试行)》。

另外,该学位论文为()课题(组)的研究成果,获得()课题(组)经费或实验室的资助,在()实验室完成。(请在以上括号内填写课题或课题组负责人或实验室名称,未有此项声明内容的,可以不作特别声明。)

声明人(签名):

年 月 日

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文(包括纸质版和电子版)，允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

()1.经厦门大学保密委员会审查核定的保密学位论文，于
年 月 日解密，解密后适用上述授权。

()2.不保密，适用上述授权。

(请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。)

声明人(签名)：

年 月 日

摘要

协同过滤算法是电子商务推荐系统中应用最为成功的推荐算法,很多电子商务企业都采用协同过滤算法作为企业推荐系统的核心算法。推荐技术可以有效解决信息过载问题,增加网站的粘度和交叉销售能力。随着移动互联网的发展,移动电子商务也在逐步兴起,并逐渐占据了大量的市场份额,推荐系统也是必不可少的。与传统电子商务相比,移动电子商务具有移动性,可接受性,安全性,定位性等特点。因此将协同过滤算法应用到移动电子商务推荐系统中时,必须做相应的改进才能适用移动电子商务的特殊环境。

本文对电子商务个性化推荐系统进行了深入研究,认真分析了协同过滤推荐算法本身存在的数据稀疏性问题和冷启动问题。通过 Slope One 算法降低了数据的稀疏性,提高了推荐系统的执行效率,使推荐系统具有较好的推荐质量。冷启动问题可以通过用户的注册信息进行特征分类,根据不同的用户特征对用户进行推荐。

针对移动电子商务领域的特殊性,本文在协同过滤算法中引入了定位技术和时间信息。通过定位技术,移动电子商务在用户的交互过程中能达到很高的个性化程度,从而满足用户对服务的差异化需求。遗忘函数的应用则进一步提高了用户兴趣迁移中的预测精度,可以推荐给用户当前最感兴趣的项目。本文最后通过实验验证了这种改进思路的可行性,证明这种改进算法的确更适合在移动电子商务平台上运用。

关键字: 移动电子商务; 协同过滤算法; 推荐系统

Abstract

Collaborative filtering is the most successful e-commerce recommendation algorithm, many e-commerce companies are using collaborative filtering as the core algorithm of the enterprise recommendation system. Recommendation technology can effectively solve the problem of information overload and increase the viscosity of the website and the ability to cross-sell. With the development of mobile Internet, mobile e-commerce is gradually rising and gradually occupies a significant market share, so recommendation system is also essential. Compared with the traditional e-commerce, mobile e-commerce has the mobility, acceptability, safety, location and other characteristics. It does the corresponding improvement in order to apply the special environment of the mobile e-commerce.

The dissertation has in-depth study of e-commerce recommendation system and analyzes the data sparsity and cold start problems of the collaborative filtering algorithm carefully. The Slope One algorithm reduces the data sparsity and improves the efficiency and quality of the recommendation system. Cold start problems can be characterized through the user's registration information classification, recommended depending on the user characteristics.

Due to the particularity of mobile e-commerce, positioning technology and time information is introduced into the collaborative filtering algorithm. By positioning technology, mobile e-commerce can achieve a high degree of personalization to meet the different demands of the users. Time information applications to further improve the prediction accuracy of the migration of user interest; can recommend to the user the right item. The experiments verify the feasibility of the improvement algorithm, to prove that this improved algorithm is more suitable for mobile e-commerce platform.

Keywords : E-commerce; Collaborative Filtering Algorithm; Recommendation System

目录

第一章 绪论	1
1.1 研究背景	1
1.2 研究意义	2
1.3 研究现状及面临的挑战	3
1.3.1 推荐系统研究现状	3
1.3.2 协同过滤算法研究现状	5
1.4 本文研究内容	6
1.5 本文组织结构	6
第二章 相关理论	8
2.1 电子商务推荐系统	8
2.1.1 电子商务推荐系统的定义	8
2.1.2 电子商务推荐系统构成	8
2.2 常用推荐算法概述	11
2.2.1 基于内容的推荐算法	11
2.2.2 基于协同过滤的推荐算法	11
2.2.3 混合式推荐算法	12
2.3 评测指标	13
2.3.1 评分预测	13
2.3.2 TopN 推荐	14
2.3.3 评测指标讨论	15
2.4 移动电子商务	15
2.4.1 移动电子商务的定义	15
2.4.2 移动电子商务特点	17
2.5 基于位置的服务	18
2.6 本章小结	20

第三章 传统电子商务协同过滤推荐算法	21
3.1 用户行为数据的收集	21
3.2 基于用户的协同过滤推荐算法	23
3.2.1 评分预测	24
3.2.2 TopN 推荐	26
3.3 基于项目的协同过滤推荐算法	27
3.3.1 评分预测	28
3.3.2 TopN 推荐	30
3.4 UserCF 算法和 ItemCF 算法综合比较	31
3.5 传统电子商务协同过滤推荐算法面临的挑战	33
3.6 本章小结	34
第四章 移动电子商务协同过滤算法改进	35
4.1 协同过滤算法在移动电子商务平台的应用问题	35
4.1.1 冷启动问题	35
4.1.2 数据稀疏问题	36
4.1.3 定位问题	37
4.1.4 时间问题	38
4.2 冷启动问题解决方法	39
4.2.1 用户冷启动问题解决方法	39
4.2.2 项目冷启动问题解决方法	41
4.2.3 系统的冷启动问题解决方法	41
4.3 降低数据稀疏性的解决方法	41
4.4 引入定位技术的算法改进	43
4.5 引入时间信息的算法改进	43
4.5.1 遗忘函数技术	44
4.5.2 改进非线性遗忘函数	44
4.5.3 时间信息相关的 UserCF 算法	45
4.5.4 时间信息相关的 ItemCF 算法	46
4.6 推荐系统流程描述	47

4.7 本章小结	48
第五章 改进算法实验	49
5.1 实验数据集和环境	49
5.1.1 数据集	49
5.1.2 实验环境	50
5.2 推荐算法实验	50
5.2.1 实验步骤	50
5.2.2 评估标准	51
5.3 实验结果及分析	51
5.3.1 UserCF 算法实验	51
5.3.2 引入时间信息的 UserCF 实验	53
5.4 实验结论	56
5.5 本章小结	57
第六章 总结与展望	58
6.1 总结	58
6.2 展望	59
参考文献	60
致谢	64

Contents

Chapter 1 Introduction	1
1.1 Research Background	1
1.2 Implications for Research	2
1.3 Research Status and Challenges	3
1.3.1 Status of the Recommendation System	3
1.3.2 Status of the Collaborative Filtering Algorithm.....	5
1.4 Main Research Contents	6
1.5 Outline of the Dissertation	6
Chapter 2 Related Theories	8
2.1 E-commerce Recommendation System	8
2.1.1 Definition of E-commerce Recommendation System.....	8
2.1.2 E-commerce Recommendation System Constitutes.....	8
2.2 Overview of Common Recommendation Algorithm	11
2.2.1 Content-based Recommendation Algorithm.....	11
2.2.2 Collaborative Filtering Recommendation Algorithm.....	11
2.2.3 Hybrid Recommendation Algorithm	12
2.3 Evaluation Indicators	13
2.3.1 Score Predicted.....	13
2.3.2 TopN Recommendation	14
2.3.3 Comparison of Evaluation Indicators	15
2.4 Mobile E-commerce	15
2.4.1 Definition of Mobile E-commerce	15
2.4.2 Mobile E-Commerce Features	17
2.5 Location-based Services	18
2.6 Summary	20

Chapter 3 E-commerce Collaborative Filtering Algorithm	21
3.1 User Behavior Data Collection	21
3.2 User-based Collaborative Filtering Algorithm	23
3.2.1 Score Predicted.....	24
3.2.2 TopN Recommendation.....	26
3.3 Item-based Collaborative Filtering Algorithm	27
3.3.1 Score Predicted.....	28
3.3.2 TopN Recommendation.....	30
3.4 Comparison the UserCF and ItemCF	31
3.5 Challenges of E-commerce Collaborative Filtering	33
3.6 Summary	33
Chapter 4 CF for the Improvement of Mobile E-Commerce	35
4.1 Collaborative Filtering Problems of Mobile E-commerce	35
4.1.1 Cold Start Problem	35
4.1.2 Data Sparsity Problem	36
4.1.3 Location Problem	37
4.1.4 Time Problem.....	38
4.2 Solution of Cold Start	39
4.2.1 Solution of User Cold Start.....	39
4.2.2 Solution of Item Cold Start	41
4.2.3 Solution of System Cold Start.....	41
4.3 Solution of Data Sparsity	41
4.4 Positioning Technology	43
4.5 Time Information.....	43
4.5.1 Forgotten Function	44
4.5.2 Improved Nonlinear Forgotten Function	44
4.5.3 Time Information UserCF Algorithm.....	45
4.5.4 Time Information ItemCF Algorithm.....	46
4.6 Recommended System Process Description	47

4.7 Summary	48
Chapter 5 Improved Algorithm Experiments	49
5.1 Experimental Data Sets and Environment.....	49
5.1.1 Data Sets	49
5.1.2 Experimental Environment	50
5.2 Recommended System Experiments	50
5.2.1 Experimental Procedure.....	50
5.2.2 Evaluation Criteria.....	51
5.3 Experimental Results and Analysis	51
5.3.1 UserCF Experiments.....	51
5.3.2 Time Information Usercf Experiment	54
5.4 Experimental Results	56
5.5 Summary	57
Chapter 6 Conclusions and Future Work	58
6.1 Conclusions.....	58
6.2 Future Work	59
References	60
Acknowledgements	64

第一章 绪论

1.1 研究背景

随着网络技术的持续发展，政府对电子商务格外重视，电子商务成为了人们日益依赖的交易方式，从事电子商务营销业务的企业已是世界经济领域不可或缺的重要组成部分。在国内，电子商务虽然起步相对较晚，但随着 Internet 在我国快速发展，我国已有一大批从事电子商务营销业务的企业逐渐兴起，同时，随着我国经济社会的空前发展，消费市场也日益扩大，人们的消费方式与消费能力也得到了很大程度的扩展与提高，在这样的条件下，电子商务在国内实现了快速发展。电子商务已经成为人们生活中不可或缺的商务交易方式，它不仅便利了网上购物者，同时也是商家增加销售量的主要手段之一，另外，电子商务还可以为网购用户节省购物的时间和空间，为他们提供物美价廉、丰富多彩的商品和服务等^[1]。

通信技术的不断进步给互联网的发展注入了新鲜血液。如今通过智能手机、平板电脑随时随地接入互联网已经成为了人们的生活方式。传统的互联网企业也逐渐做出了各种改变，大部分的互联网企业都推出了适合移动客户使用的互联网产品以抢占移动互联网这块新兴市场。由于智能移动设备日益普及，信息资源的获取和推送可以发生在“任何时间、任何地点、任何方式”，为用户提供无处不在的信息内容已经成为可能，智能移动设备也逐渐成为人们获取信息的主要平台之一(如通过手机搜索网络信息、浏览新闻，使用平板电脑或者手机阅读电子书、听音乐、看视频，使用手机读写微博等)^[2]。移动互联网逐渐改变了人们的生活方式，人们随时随地都被一张无形的“网”联系在一起。

作为传统互联网领域中运营成熟的电子商务企业也开始在移动互联网布局。人们不再需要坐在电脑面前才能通过网络购物消费，这一方式的转变大大改变了大家的消费习惯，同时也增加了消费产品的类型。移动互联网信息内容和服务的不断增长将会逐渐超出移动用户所能接受的范围，同时，由于移动设备与传统的

PC 相比，界面显示、终端处理等能力都比较有限，所以移动互联网更容易产生信息过载问题，移动信息过载严重影响了移动网络资源的利用率，也使移动用户体验质量下降了^[3]。

传统电子商务也遇到过类似问题，导致电子商务的持续快速发展遇到了较大的压力。为了解决电子商务发展的这一难题，电商企业以及大学科研中心的研究人员进行了大量的探索研究。科研人员和工程师们寻求一种将用户感兴趣的信息主动推荐给用户的 service 方式，这便是个性化推荐服务。即：在线企业通过分析顾客的基本信息和网络活动记录来研究顾客的购买行为，并根据顾客的购物偏好等信息将顾客可能感兴趣的商品和服务自动推荐给用户。一方面可以帮助用户快速找到自己想要的商品，使用户感受到便利，这是实现用户重复网上购物的一个重要因素；另外，一些用户在闲暇之时抱着打发时间的心态浏览商品，原本没有购买商品的想法，正是由于推荐系统为其提供了一些其感兴趣的物品，可以挖掘出一些潜在用户，在满足顾客购买商品的需求的同时实现了在线企业实现增值的愿望^[4]。

移动电子商务在对待这样的问题时，也需要进行大量的研究去解决类似的问题。然而移动电子商务与传统的电子商务有很大的区别，在推荐系统的构建上需要考虑的因素也不尽相同。尽管很多推荐系统成功应用在 WebServices 这种传统的电子商务中，但这些推荐系统并不适合移动电子商务领域。除了可接入性和灵活性的问题，实时性和交互性也成为传统推荐系统应用在移动电子商务上的障碍。

推荐系统的核心是推荐算法。随着学者们研究的不断深入，越来越多的算法被应用在推荐系统中。在传统电子商务推荐系统中，应用最为广泛的推荐算法为协同过滤算法，但是经过分析它却存在着两大缺陷：冷启动问题和数据稀疏问题^[5]。本论文将对协同过滤算法进行深入的分析和研究，并对该算法进行改进以适应移动电子商务平台。

1.2 研究意义

互联网技术的不断发展让我们可以通过 PC 或者智能手机接触到海量的信

息。互联网的出现使人们步入一个信息爆炸的时代，因此如何筛选自己需要的信息是每个人都会面临的问题。电子商务作为对传统商务的补充，在互联网时代得到了快速的发展，越来越多的商品不断被搬上了互联网平台^[6]。这极大方便了人们购物和消费，人们足不出户就可以购买到自己心仪的产品。然而，网页上大量的商品信息在使客户有更多选择的同时也需要花费大量的时间去筛选和浏览，这无疑考验着客户的耐心。电子商务行业是个竞争十分激励的行业，各电子商务企业都想尽方法让用户留在自己的页面上，于是研究人员就将推荐系统应用到电子商务领域中。随着推荐系统的出现，用户获取商品信息的方式从简单的目标明确的数据的搜索转换到更高级更符合人们使用习惯的信息发现。一些新商品的出现也将通过推荐系统展现在用户的面前，这大大提高了电子商务网站的用户体验。

随着移动互联网时代的到来，移动电子商务正逐步兴起，各大电子商务网站也不断推出自己公司的手机终端以抢占市场。协同过滤算法是传统电子商务推荐系统中应用最为广泛和成功的算法，然而由于移动电子商务与传统电子商务有显著的区别，简单地将传统电子商务推荐系统直接应用到移动电子商务领域并不合适^[7]。因此，研究协同过滤算法在移动电子商务推荐系统中的应用具有十分现实的意义。

1.3 研究现状及面临的挑战

传统的电子商务发展已经相对比较成熟，当用户登录一个电子商务网站进行购物时就会有许多为用户量身打造的商品出现到用户的购物页面。

1.3.1 推荐系统研究现状

推荐系统是自动联系用户和物品的一种工具，它能够在信息过载的环境中帮助用户发现令他们感兴趣的信息，也能将信息推送给对它们感兴趣的用户。在电子商务平台上，推荐系统扮演了一个销售员的角色，它将用户感兴趣的物品推荐给有购买倾向的用户。因此，推荐系统不仅具有认知科学、模糊理论、信息检索、预测理论等领域的知识，并且跟管理科学和营销学有着密切的关系，是一套综合了各个学科的复合系统。个性化推荐系统作为一个独立的概念在 20 世纪 90 年代中期被提出来。近些年，随着电子商务的迅速发展，推荐系统在网络上的应用也

越来越多,国内外学术界也对推荐系统的研究投入了巨大的精力和热情。为了解决推荐系统中存在的问题,已经有无数科学家和工程师提出了很多天才的解决方案。在 2007 年,美国计算机协会(ACM)召开了第一届推荐系统领域的研究会议(ACM RecSys'07),这次会议对推荐系统的研究起到了积极的促进作用。在该届会议上,科研工作者们发表了大量高质量的关于推荐系统的论文和研究工作报告。这些研究工作和论文深入分析了当前推荐系统存在的问题,并针对这些问题从不同角度对推荐系统进行了改进。此次会议凸显出推荐系统在现代互联网时代,以及接下来的移动互联网时代的重要性。

推荐系统主要是通过特定的推荐算法对用户信息进行分析然后得出推荐信息,因此,一个完整的推荐系统主要由两个模块组成:(1)行为记录模块;(2)推荐算法模块。行为记录模块提供了推荐系统所需要的基本数据,推荐系统最终产生的推荐产品都是通过分析计算这些基本数据而得出来的。行为记录模块负责收集用户基本信息,如用户注册网站时的个人信息,性别,年龄,爱好等,同时也记录代表用户喜好的行为,例如购买、评分、问答、浏览、收藏、下载等。推荐算法模块是推荐系统最为核心的部分,推荐算法模块利用后台的推荐算法,对行为记录模块所提供的基本数据进行分析计算,实时地从产品集合中筛选出用户感兴趣的产品进行推荐^[8]。针对不同的应用场景和推荐时使用的数据不同,常见的推荐算法有三种:基于内容的推荐算法、协同过滤推荐算法、以及混合式的推荐算法^[9-12]。

目前,在国内外知名的大型电子商务平台上都有推荐系统,它们每时每刻都在为用户提供推荐服务,将不同的产品推荐给合适的用户。在美国,著名的电子商务网站亚马逊是个性化推荐系统的积极应用者和推广者,被 RWW(读写网)称为“推荐系统之王”。亚马逊的推荐系统深入到了其各类产品中,为其业务的发展做出了巨大的贡献。同时亚马逊的科学家们也投入了大量的人力物力去研究推荐系统,使其更精确地反应用户的真实需求,从而更好的为公司的业务服务。在国内,以当当网为代表的电子商务网站也采用了推荐系统,极大地方便了用户购物,同时也为公司的业务作出了巨大贡献。在移动电子商务领域,推荐系统也得到了广泛的应用,与传统电子商务相比移动电子商务领域的推荐系统会有显著的区

别，它会针对移动用户的特殊性，提供不同的推荐结果。目前，以亚马逊为主的电子商务网站在推荐系统的研究和应用方面还领先着国内，随着移动电子商务的不断发展，国内外的科学家和工程师们也逐渐将研究重点放在了移动电子商务推荐系统上。国内的移动电子商务才刚刚起步，移动电子商务推荐系统的研究也在逐步跟上，将传统电子商务推荐系统移植到移动电子商务领域的过程中，也要针对移动电子商务的特殊性做相应的改进，设计更好的推荐方法，产生更好的推荐效果。

1.3.2 协同过滤算法研究现状

随着 Web 2.0 的发展，Web 站点更加提倡用户参与和用户贡献，因此基于协同过滤的推荐机制应运而生。它的原理很简单，就是根据用户对物品或者信息的偏好，发现物品或者内容本身的相关性，或者是发现用户的相关性，然后再基于这些关联性进行推荐。

协同过滤算法是推荐系统中最基本的算法，该算法不仅在学术界得到了深入研究，而且在业界得到了广泛应用。协同过滤算法目前是电子商务推荐系统中应用最为成功的算法，协同过滤算法分为两大类，一类是基于用户的协同过滤算法，另一类是基于项目的协同过滤算法。其中基于用户的协同过滤算法是推荐系统中最古老的算法。该算法在 1992 年被提出，并应用于邮件过滤系统，1994 年被 GroupLens 用于新闻过滤。在此之后，基于用户的协同过滤算法都是推荐系统中最著名的算法。基于项目的协同过滤算法也是目前应用最广泛的算法之一，无论是亚马逊，还是 YouTube，其推荐算法的基础都是基于项目的协同过滤算法。这一算法就是由著名的电子商务公司亚马逊提出来的，目的是为了解决基于用户协同过滤算法的不足。目前，国内的许多电子商务网站也都使用基于协同过滤的推荐算法作为网站推荐系统的基础算法，如当当书城、京东商城和淘宝网等^[13]。

尽管已经有了大量关于协同过滤算法的研究工作，但是针对不同的应用场景，依然需要对协同算法进行研究和改进以提高协同过滤算法的推荐质量。协同算法存在的冷启动问题，数据稀疏性问题都是每个研究人员需要面对的，如何在应用协同过滤算法的时候解决或者降低这些问题对推荐质量的影响是目前研究协同过滤算法的重点。

Degree papers are in the "[Xiamen University Electronic Theses and Dissertations Database](#)". Full texts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to etd@xmu.edu.cn for delivery details.

廈門大學博碩士論文摘要庫