



湖北经济学院学术文库

# 电子商务智能推荐系统研究

DIANZI SHANGWU ZHINENG TUIJIAN  
XITONG YANJIU

余小高◎著

湖北长江出版集团  
湖北人民出版社



湖北经济学院学术文库

# 电子商务智能推荐系统研究

余小高◎著

湖北长江出版集团  
湖北人民出版社

鄂新登字 01 号

图书在版编目(CIP)数据

电子商务智能推荐系统研究/余小高著.  
武汉:湖北人民出版社,2012.2

ISBN 978 - 7 - 216 - 06983 - 0

- I. 电…
- II. 余…
- III. 电子商务—智能系统—研究
- IV. ①F713.36②TP393.07

中国版本图书馆 CIP 数据核字(2011)第 218014 号

电子商务智能推荐系统研究

余小高 著

出版发行: 湖北长江出版集团  
湖北人民出版社

地址:武汉市雄楚大道 268 号  
邮编:430070

印刷:武汉明天印务有限公司  
开本:787 毫米×1092 毫米 1/16  
字数:303 千字

经销:湖北省新华书店  
印张:17.5  
插页:1

版次:2012 年 2 月第 1 版  
书号:ISBN 978 - 7 - 216 - 06983 - 0

印次:2012 年 2 月第 1 次印刷  
定价:38.00 元

本社网址:<http://www.hbpp.com.cn>

## 总 序

说起大学，许多人都会想起教育家梅贻琦校长的名言：“所谓大学者，非谓有大楼之谓也，有大师之谓也。”大楼是一砖一瓦盖起来的，大师是如何成长起来的呢？

大师的成长，当然离不开种种主客观条件。在我看来，在诸多的因素中，一个十分重要甚至是不可或缺的就是学术思想、学术观点的充分展示与交流。纵观历史，凡学术繁荣、大师辈出时期，必有一个做学问者有可以互相炫技、彼此辩论、各显神通的大舞台。古代春秋战国时代诸子百家争鸣，离不开当时的客卿、门客制度，为不同的观点、流派彼此公开竞争并得到君主的采纳搭建了平台；学者阿英在论及晚清以来中国新思想、新艺术的繁荣时，列举了三条原因，其中的第一条原因“当然是由于印刷事业的发达，没有前此那样刻书的困难；由于新闻事业的发达，在应用上需要多量的产生”。北京大学之所以能成为中国新思想新文化的发源地，社团与杂志在其中发挥了巨大作用：1918年，《北京大学月报》成为中国最早的大学学报，加之《新青年》、新潮社等杂志、社团为师生搭建了一个有声有色的大舞台，开启了自己近一个世纪的辉煌。国外名牌大学的发展，无不伴随着一次又一次思想的激烈交锋、学术的充分争鸣，并且这些交锋和争鸣的成果都得到了最好的展示与传播，没有这些交锋与争鸣，就没有古老的牛津、剑桥，也没有现代的斯坦福、伯克利。近代以来的印刷技术、新的出版机制、文化传播业的发达，为新的思想、学术之间彼此的炫技和斗法提供了可能的舞台：伟大的舞台造就了伟大的演员。

在现代的大学中，如果说大楼是基础、大师是灵魂，那么大舞台就应该是机制——它意味着研究冲动的促动、言说欲念的激发、交流碰撞的实现。在这样的机制中，精神、灵魂得到孕育！思想、学术实现成长！大学本身就应该成为一个众声喧哗的大舞台，一个为大师成长提供基础和机制的场所。因此，大学出版自己的学术文库，运用现代传媒、现代出版为自己的教师提供思想碰撞、学术交流的平台，其意义绝不仅限于资金的支持和个人成果的

发表，它的意义更在于通过这种方式营造学术氛围、彰显学术精神，在学校形成言说、表达、交流的习惯和风尚，激励教师为了“台上一分钟”，做好“台下十年功”，凝聚大学理念、大学精神、大学风格。

湖北经济学院是一所年轻的大学，湖北经济学院的教师大多都还年轻，正逢为一所初创不久的大学承担奠基、开拓之责的难得人生机遇。我们中还没有大师，或者说在我们这里产生大师还有很长的路要走。然而，这并不等于我们不期待大师的产生，更不意味着我们不去为大师的成长做出努力、不去为大师的产生构筑平台。大师的成长是大学成长的永恒动力，对学术的追求是大学能够逾千年而长青的不竭源泉。现在，我们的平台也许还不够高，还不够大，但我们坚信：这个小平台也许就是未来大师的第一次亮相！因此，我们在这里鼓励每个人以充分的自信发出自己的声音，可以在众声喧哗中更加大声喧哗，在交流与碰撞中实现批判、被批判与自我批判，能够在这个平台上得到成长，收获乐趣，实现价值！

“湖北经济学院学术文库”就是这样一个为大师成长搭建的交流与对话的平台。每一本著作，都是我们的教师在各自学术领域中富有心得而最想表达的内容——他们渴望得到承认，也不怕获得批评；他们充满自信地言说，也将谦虚自谨地倾听。

愿“湖北经济学院学术文库”和湖北经济学院一同成长，愿它能成为一个大师初成的舞台，从中诞生出不朽的学术和永恒的大学精神！

湖北经济学院院长：



# 引 言

随着电子商务和互联网的迅猛发展，发布在互联网上的商务信息呈爆炸性增长。越来越多的企业和用户利用互联网发出信息和获取有价值的信息，互联网已成为人们发布、接收信息及交流的主要媒介，也逐渐成为各种信息流的主要传输渠道。面对越来越复杂的电子商务系统，越来越多的用户感觉到力不从心，并且这一严峻状况仍将持续发展下去。据统计，相对于99%的用户而言，99%的Web信息是无用的，每个用户真正需要的只是其中很小的一部分，大量的无关信息会干扰甚至淹没用户真正需要的内容。于是，在这个被称为信息爆炸的时代，信息过量几乎成为人人都需要面对的问题。如何有效地分析用户的需求，使用户不被信息的汪洋大海所淹没，帮助用户从互联网的商务信息海洋中发现他们所要查找或者感兴趣的资源，已经成为一个亟待解决的问题。

由此，电子商务智能推荐系统应运而生，它利用电子商务网站向客户提供商品信息和建议，帮助用户决定应该购买什么产品，模拟销售人员帮助客户完成购买过程。电子商务智能推荐系统从产生以来，就一直受到国内外用户和研究者越来越多的关注，一些试验型系统和应用系统相继被开发出来，其理论和应用都得到了普及，实现了客商的“双赢”。目前，推荐技术主要分为基于内容的推荐技术和协同过滤推荐技术。前者根据当前用户档案和商品档案的相似程度进行推荐；后者根据其他相似用户对商品的评分预测当前用户的评分，再进行推荐。同时，越来越多的其他推荐工具正逐步产生。

但是，随着电子商务系统的进一步复杂化，推荐系统逐渐面临一系列严峻挑战，包括实时性和可扩展性、稀疏性和“冷开始”等方面的问题。同时，推荐结果的可解释性也比较差，使用户对被推荐商品讨价还价这一符合市场特点的功能很难实现。这些问题影响了推荐系统的效率和推荐质量，影响了推荐系统对社会经济活动等的显著作用。

基于上述问题以及初步设想，针对目前电子商务智能推荐系统面临的现状，本书拟从解决主要挑战入手，对现有推荐系统进行研究，试图解决其几个主要方面的不足，并且支持对推荐结果的解释和讨价还价，最终目的是建

立推荐效率高、质量好的推荐系统。更为重要的是，通过电子商务智能推荐系统的研究，使得人们能借助有关的技术快速、灵活、准确地完成商务活动，达到双赢的局面，这对推动电子商务的进一步发展、促进国内经济和国际接轨，也有着重要的理论价值和实际意义。

# 目 录

|                           |    |
|---------------------------|----|
| 第一章 绪论                    | 1  |
| 第一节 研究背景                  | 1  |
| 第二节 国内外研究现状与分析            | 2  |
| 第三节 本书拟解决的关键问题            | 7  |
| 第四节 研究目的和意义               | 8  |
| 第五节 本书结构                  | 9  |
| 第二章 电子商务推荐系统简介            | 11 |
| 第一节 电子商务分析                | 11 |
| 第二节 信息检索和信息过滤             | 24 |
| 第三节 个性化推荐                 | 27 |
| 第四节 电子商务推荐系统模型            | 30 |
| 第五节 传统推荐算法简介              | 39 |
| 第三章 基于改进的 $k$ 最近邻协同过滤推荐算法 | 52 |
| 第一节 提出问题                  | 52 |
| 第二节 $k$ 最近邻搜索 (KNNs) 算法简介 | 53 |
| 第三节 相关研究                  | 55 |
| 第四节 P2PAKNNs 算法           | 58 |
| 第五节 BA-KNNs 算法            | 69 |
| 第六节 基于改进的 $k$ 最近邻协同过滤推荐   | 76 |
| 第七节 本章小结                  | 79 |
| 第四章 基于 KNDB 的协同过滤推荐算法     | 81 |
| 第一节 提出问题                  | 81 |



|            |                                 |            |
|------------|---------------------------------|------------|
| 第二节        | 聚类简介 .....                      | 83         |
| 第三节        | KNDB 聚类算法 .....                 | 87         |
| 第四节        | 基于 KNDB 的协同过滤推荐算法 .....         | 95         |
| 第五节        | 本章小结 .....                      | 97         |
| <b>第五章</b> | <b>基于关联规则的智能推荐技术 .....</b>      | <b>98</b>  |
| 第一节        | 关联规则在电子商务应用中存在的问题 .....         | 98         |
| 第二节        | 关联规则算法分析 .....                  | 99         |
| 第三节        | RSAA - BOUIGA 关联规则算法 .....      | 108        |
| 第四节        | 双阈值法 .....                      | 112        |
| 第五节        | 基于关联规则的智能推荐技术 .....             | 114        |
| 第六节        | 本章小结 .....                      | 118        |
| <b>第六章</b> | <b>基于隐式评分的智能推荐系统 .....</b>      | <b>119</b> |
| 第一节        | 提出问题 .....                      | 119        |
| 第二节        | BP 学习算法简介 .....                 | 121        |
| 第三节        | 体系结构 .....                      | 123        |
| 第四节        | 项档案的建立 .....                    | 125        |
| 第五节        | 用户档案的建立 .....                   | 126        |
| 第六节        | 协同过滤推荐的产生 .....                 | 134        |
| 第七节        | 本章小结 .....                      | 135        |
| <b>第七章</b> | <b>基于内容的智能推荐系统 .....</b>        | <b>137</b> |
| 第一节        | 提出问题 .....                      | 137        |
| 第二节        | 用户兴趣特征描述 .....                  | 138        |
| 第三节        | 体系结构 .....                      | 140        |
| 第四节        | 基于内容的推荐产生 .....                 | 142        |
| 第五节        | 本章小结 .....                      | 143        |
| <b>第八章</b> | <b>面向用户偏好和推荐结果的自动谈判协商 .....</b> | <b>144</b> |
| 第一节        | 研究现状 .....                      | 144        |

---

|             |                         |            |
|-------------|-------------------------|------------|
| 第二节         | 问题分析                    | 146        |
| 第三节         | 电子商务谈判简介                | 149        |
| 第四节         | 谈判模型                    | 152        |
| 第五节         | 谈判空间                    | 154        |
| 第六节         | 谈判策略                    | 155        |
| 第七节         | 本章小结                    | 157        |
| <b>第九章</b>  | <b>电子商务网站信息构建</b>       | <b>159</b> |
| 第一节         | 信息构建概述                  | 159        |
| 第二节         | WIESM 研究                | 161        |
| 第三节         | 电子商务网站 IA 设计分析          | 180        |
| 第四节         | 一个电子商务网站 IA 设计方案        | 187        |
| 第五节         | 本章小结                    | 190        |
| <b>第十章</b>  | <b>电子商务智能推荐系统的设计与实现</b> | <b>191</b> |
| 第一节         | 简介                      | 191        |
| 第二节         | BWARS 功能                | 193        |
| 第三节         | BWARS 的体系结构             | 197        |
| 第四节         | BWARS 主要模块设计与实现         | 201        |
| 第五节         | 系统优点                    | 249        |
| 第六节         | 本章小结                    | 250        |
| <b>第十一章</b> | <b>总结与展望</b>            | <b>251</b> |
| 第一节         | 全书总结                    | 251        |
| 第二节         | 展望                      | 253        |
| <b>参考文献</b> |                         | <b>254</b> |
| <b>后 记</b>  |                         | <b>270</b> |

# 第一章 绪论

## 第一节 研究背景

随着互联网的发展,电子商务规模越来越大。多样化和专用化产品逐步取代标准产品,厂家需要能生产多样化的产品来满足不同顾客的多种需求<sup>[1]</sup>。因此,电子商务平台应符合发展要求,为顾客提供更多的选择,这势必产生大量的信息。由于信息量的增大,导致如下两种结果:一方面,消费者失去了与商家的联系;另一方面,客户置身在大量的商品信息中束手无策,迷失在海量的商品信息空间中,无法顺利找到所需商品。电子商务站点采用推荐系统是解决信息量增多与客商双方的无奈这一矛盾的有效途径。

推荐系统能模拟销售人员向顾客推荐商品,帮助顾客顺利找到所需商品,从而完成购买过程。因此,它可以有效保留客户,提高销售水平。通过推荐系统,商家也可以保持与客户的联系,从而建立良好客户关系,提高客户回店率,所以推荐系统实现了客商的“双赢”。电子商务系统与推荐系统相互结合,能够帮助客户顺利找到所需商品。由于推荐系统在电子商务环境中能够方便地收集丰富的电子化数据,很好地检验推荐效果,因此推荐系统很适合在电子商务环境中实施。

电子商务推荐系统具有良好的发展和应用前景,成功的电子商务推荐系统能够产生巨大的经济效益。研究表明,使用个性化推荐系统后,电子商务销售行业的销售额能提高2%~8%<sup>[2]</sup>,尤其在日用百货、CD音像、电影、书籍等产品价格相对较为低廉、商品种类繁多、使用个性化推荐系统程度较高的行业,推荐系统能较大地提高企业的销售额。目前,几乎所有大型的电子商务系统都不同程度地使用了各种形式的推荐系统,如eBay、DangDang、Amazon和CDNOW等。因此,提供个性化服务的各种Web站点需要推荐系统的大力支持。

电子商务推荐系统应用越来越广泛,其理论也越来越成熟。但是,电子商务的规模随着客户和商品数量的增加越来越大,推荐系统将面对更大规模的

数据,需要及时有效地处理这些数据,实时地作出商品推荐,并有效地解释推荐的结果,从而说服客户。面对这些海量的数据,目前推荐系统的效率和准确度已不能适应实际的需求。因此,本书将进一步研究电子商务推荐系统,试图提高推荐系统的准确度和效率,并试图解决推荐结果可解释性差等问题,支持客商之间的讨价还价,提高推荐系统的有效性和准确性,以促成销售,实现推荐目的。

## 第二节 国内外研究现状与分析

### 一、研究现状简述

1997年,Resnick和Varian给出了推荐系统的定义:“利用电子商务网站向客户提供商品信息和建议,帮助用户决定应该购买什么产品,模拟销售人员帮助客户完成购买过程。”<sup>[3]</sup>该观点认为,用户在没有足够的经验或其他人建议的情况下面对大量的商品很难作出购买选择,该定义已被广泛引用。1998年7月,美国AAAI组织各学者在威斯康辛州专门召开了以推荐系统为主题的会议,集中讨论了推荐系统发展的问题。目前,推荐系统已广泛运用到各行业中,推荐对象包括文章、书籍、新闻、音像和网页等。

有效性和效率是衡量推荐系统优劣的重要指标。推荐系统主要包括两种:基于协同过滤(CF)和基于内容的推荐(CB)<sup>[4]</sup>。基于内容的推荐系统需要分析资源的内容信息<sup>[5][6]</sup>,根据用户兴趣建立用户档案(Profile),然后根据资源内容与用户档案之间的相似性向用户提供推荐服务<sup>[7]</sup>。基于内容的推荐技术是信息检索领域的重要研究内容<sup>[8][9]</sup>。内容分析技术<sup>[10][11][12]</sup>挖掘资源的特征,分析Web日志捕捉潜在规则。用户档案包括用户的注意点、兴趣点和偏好等。Salton等人提出根据用户反馈自动更新用户档案的方法<sup>[13][14]</sup>。Bayesian概率模型<sup>[9]</sup>、人工神经网络<sup>[1][15]</sup>、遗传算法<sup>[16]</sup>以及其他机器学习技术也被广泛应用于用户档案的建立和更新。

基于内容的推荐技术具有一定的局限性<sup>[17]</sup>。首先,资源内容必须以机器可以理解的格式表示,而很多信息例如视频、图像、音乐等多媒体信息很难做到这一点;其次,资源内容不能提供较多的建议,其分析范围比较小;第三,基于内容的推荐技术不能从样式、质量、审美等角度对项进行过滤。文献[18]认为,基于内容的过滤无法分析信息的质量,无法提供新颖的推荐。

针对上述存在的问题,很多研究者提出协同过滤推荐技术。协同过滤推荐

的主要思路是根据其他用户的观点产生对目标用户的推荐列表,它基于这样一个假设:如果用户对一些项(商品)的评分比较相似,则他们对其他项的评分也比较相似。协同过滤推荐算法中往往存在一个评分矩阵  $\text{user} \times \text{item}$ ,存储着用户对项的评分。协同过滤推荐系统使用统计技术等搜索目标用户的若干最近邻居,然后根据该最近邻居对项的评分预测目标用户对项的评分,再根据评分的高低选择前  $N$  个最高评分(Top -  $N$ ) 作为推荐列表。

在早期的协同过滤推荐系统中,用户之间需要相互了解对方的兴趣爱好,这给用户带来了很多的不便。因此,研究者又提出了自动协同过滤推荐技术<sup>[19]</sup>。

协同过滤推荐主要分为三类:一是基于内存的协同过滤(Memory-based Collaborative Filtering),先用相似统计等方法得到具有相似兴趣爱好的邻居用户,再产生推荐,所以该方法也称基于用户的协同过滤(User-based Collaborative Filtering)或基于邻居的协同过滤(Neighbor-based Collaborative Filtering)<sup>[20][21][22]</sup>。二是基于模型的协同过滤(Model-based Collaborative Filtering),先用历史数据得到一个模型,再用此模型进行预测<sup>[22]</sup>,然后产生推荐。基于模型的推荐中广泛使用的技术包括人工神经网络(Artificial Neural Network)等学习技术、潜在语义检索(Latent Semantic Indexing)和贝叶斯网络(Bayesian Networks)等。三是基于项的协同推荐算法(Item-based Collaborative Filtering Algorithm),由于基于用户的协同推荐算法随着用户数量的增多,计算量成线性加大,其性能越来越差,为此,在2001年 Sarwr 教授提出了该推荐算法。该算法通过先计算已评价项和待预测项的相似度,以相似度作为权重,加权各已评价项的评分,得到预测项的预测评分,并指出基于项的推荐算法比基于用户的推荐算法还要好,且能解决基于用户的协同推荐的两个问题。Breese 教授认为基于用户的协同推荐比基于模型的协同推荐方法更好<sup>[18]</sup>。Mild 教授从批判的角度指出基于项的协同推荐并不一定好,该算法准确度与采用的实验数据的规模有关,大多数情况下还是基于用户的协同推荐好<sup>[4]</sup>。

为了提高推荐系统的推荐效率和有效性,很多学者作了大量的研究。文献[23]提出通过图搜索(Horting)来计算用户最近邻居的优化算法。文献[24]提出对用户最近邻居和项均采用不同权重的方法改进推荐质量。文献[25]对各种用户间相似性度量方法进行了分析,提出了各种改进方法。文献[26][27]发现用户评分数据的稀疏性是导致推荐系统推荐质量下降的主要原因。文献[18]提出了基于项评分预测的 IRPRec 协同过滤推荐算法,该算法

首先计算项间相似性,由用户对相似项的评分预测用户对未评分项的评分,使得用户之间共同评分的项比较多,从而可以有效解决用户评分数据极端稀疏情况下传统相似性度量方法存在的不足,使得计算得到的目标用户的最近邻居比较准确,显著提高电子商务推荐算法的推荐精度。

传统的协同过滤推荐技术根据用户显式评分产生推荐结果,用户使用不方便。同时,由于项量的增大等一系列原因,用户也很难对所有的项一一评分。因此,许多研究者提出可以通过 Web 挖掘技术获取用户的隐式评分<sup>[28]</sup>,自动为用户给项评分<sup>[29][30]</sup>。文献[31]提出了基于分布式用户评分的推荐系统的研究。

随着电子商务规模的扩大,推荐系统的实时性研究逐渐成为研究热点,得到了越来越多研究者的关注。聚类技术(Clustering)将具有相似兴趣爱好的用户分配到相同的簇中<sup>[32][33]</sup>,聚类产生之后,根据簇中其他用户对商品的评价预测目标用户对该商品的评价。由于聚类过程离线进行,所以在线的推荐算法产生推荐的速度比较快。文献[34]中提出使用聚类分析技术将用户划分为不同的组,从而有效减小搜索空间。文献[35]中提出通过项聚类分析,从而在用户评分数据的子集上搜索最近邻居。文献[18]提出了基于项聚类的 ICRec 协同过滤推荐算法。该算法通过用户对项评分的相似性对项进行聚类并生成对应的聚类中心,在此基础上计算目标项与聚类中心的相似性,选择与目标项相似性最高的若干个聚类作为查询空间,在这些聚类中搜索目标项的最近邻居,从而能够在尽量少的项空间上搜索到目标项大部分的最近邻居,有效提高推荐系统的实时响应速度。为了提高实时性和可扩展性,文献[36]提出基于 P2P 结构的分布式推荐算法 PipeCF 以实现有效推荐。文献[37]提出 P2P 网络中基于内容的文档推荐。

协同过滤推荐技术也存在自身的不足<sup>[18][38]</sup>,主要表现在用户评分数据稀疏时推荐质量比较低、可扩展性和实时性不足、不能有效解释推荐结果等方面。因此,基于多种数据、多种技术的有效集成的推荐技术得到研究者的重视。文献[39]提出通过基于内容过滤和协同过滤的复合型推荐系统提高推荐质量。文献[1]提出元推荐系统以期提高推荐系统的推荐质量。文献[40]提出基于本体的个性化推荐技术。随着网格计算理论和实践的发展,网格技术和推荐系统的关系越来越紧密。文献[41]提出一个自适应的推荐系统信息网格体系结构。文献[42]作出了基于服务仲裁的推荐框架的研究。文献[43]提出面向数据挖掘的计算工作流的相关图形模型,采用基于得分结构的学习算法,以从科学数据和 Web 服务池中发现专业领域协同推荐的相关图形模型,同时面

向从科学数据和 Web 服务池中检索到的软件和数据集合,使用案例技术进行推理。网格和 Web Services 技术成功应用于真实世界,需要作为内部代理通信和服务集成的基础词汇表和元数据框架的发展,同时需要访问和使用解决问题的专业领域知识的知识库<sup>[44]</sup>。因此,文献[45]认为语义服务描述和基于专业领域知识的决策支持服务都是电子商务中资源集成的基本成分,提出了一种基于知识的语义服务合成方法及框架,以支持实时推荐。

## 二、推荐系统实例简介

推荐系统是信息检索和信息过滤领域的研究热点,得到了许多著名研究机构和研究者的关注,出现了大量研究型推荐系统实例<sup>[18]</sup>:

(1)TYPESTRY:是 Xerox PARC 研究中心提出的一个研究型协同过滤推荐系统,用于过滤电子邮件,推荐电子新闻。TYPESTRY 系统提供电子文档存储、用户评价存储和协同过滤推荐服务。在 TYPESTRY 系统中,设计了一种类似于 SQL 的查询语言 TQL,用户的查询请求中必须明确指出与自己兴趣爱好相似的其他用户。由于用户之间必须了解对方的兴趣爱好,因此 TYPESTRY 推荐系统只适用于用户群体比较小的场合。

(2)ACF:Active Collaborative Filtering 系统是 Carnegie-Mellon 大学开发的主动协同过滤推荐系统,用于电子文档推荐。ACF 系统通过指针实现协同过滤推荐服务,指针包含指向电子文档的超链接、电子文档的上下文信息以及用户撰写的电子文档评论。在 ACF 系统中,用户可以通过主动的方式将创建的指针推荐给其他可能感兴趣的用户,也可以将创建的指针保存在系统中供其他用户查看。ACF 系统也只适用于用户群体比较小的场合。

(3)GroupLens:由 MIT 开发的自动协同过滤推荐系统,用于新闻组信息推荐。GroupLens 系统通过用户的评分信息自动搜索用户的  $k$  个最近邻居,然后根据该  $k$  个最近邻居的评分信息产生最终的推荐结果,适合于用户数量比较大的场合。GroupLens 系统具有极好的开放性,用户可以通过 GroupLens 系统提供的 API 函数向 GroupLens 服务器提供评分信息,请求推荐结果。同时,GroupLens 系统提供三种客户端工具 Emacs Gnus、NN 和 NewsWatcher 达到上述目的。

(4)MovieLens:是 Minnesota 大学开发的研究型自动协同过滤推荐系统,用于推荐电影。与 GroupLens 不同,MovieLens 系统是一个基于 Web 的推荐系统,系统通过浏览器的方式进行用户评分数据收集与推荐结果显示,用户使用更加方便。

(5)Ringo:由 MIT 媒体实验室开发的研究型协同过滤推荐系统,用于提供个性化的音乐推荐服务。Ringo 系统可以向用户推荐用户最喜欢的音乐,预测用户最不喜欢的音乐,也可以预测用户对特定音乐的评分。

(6)Video Recommender:是 Bell Core 开发的协同过滤推荐系统,用于电影推荐。Video Recommender 系统通过电子邮件的方式收集用户评分数据,提供推荐服务。在 Video Recommender 系统中,不仅可以预测用户最喜欢的电影,也可以预测用户对特定电影的评分。在 Video Recommender 系统提供推荐结果的同时,还能向用户提供用户最近邻居的电子邮件联系方式,以及用户与最近邻居的相似度等信息。

(7)FAB:是 Stanford 大学数字图书馆项目组开发的基于内容的过滤和协同过滤的复合型推荐系统,用于推荐 Web 页面。其特点是综合了基于内容过滤的推荐和协同过滤推荐的优点,同时支持两种类型的推荐服务。FAB 系统主要包括页面收集代理、个人推荐代理和中心路由器。页面收集代理从 Web 上收集特定主题的页面,个人推荐代理从特定主题中选择用户感兴趣的页面推荐给用户。个人推荐代理根据文档内容信息建立用户的用户档案,然后根据用户档案之间的相似性搜索用户的最近邻居。产生的推荐结果可以基于用户档案中的文档内容信息,也可以基于用户最近邻居的评价信息(协同过滤)。

### 三、研究现状分析

电子商务推荐技术在实际应用中获得了极大的成功,许多大型的电子商务系统,如最大的网上书店 Amazon.com,最大的网上音乐商店 CDNow.com,最大的网上 CtoB 站点 eBay.com 等,都不同程度地提供了各种推荐服务。但是随着站点结构、内容的复杂度和用户人数的不断增加,电子商务推荐算法也面临着一系列挑战,主要包括<sup>[46]</sup>:

(1)实时性和可扩展性(Scalability)问题:“最近邻居”搜索算法(KNNs)的计算量随着用户和项数量的增加而大大增加,对于上百万之巨的数据,推荐系统必须快速处理高维 user × item 矩阵,实时搜索成百万邻居,在几毫秒内处理成千上万用户,提供推荐。传统的推荐算法将遭遇到严重的实时性和可扩展性问题。

(2)稀疏性(Sparsity)和“冷开始”(Cold Start)问题:在许多推荐系统中,每个用户涉及的信息量相当有限,在一些大的系统如 Amazon.com 中,用户最多不过就评估了上百万本书的 1% ~ 2%,造成评估矩阵数据相当稀疏。用少量的用户信息难以找到相似用户集,导致推荐效果大大降低。“冷开始



(Cold Start)”问题就是新推出商品由于没有得到客户的关注而导致始终得不到推荐。

(3) 精确性(Accuracy):通过寻找目标用户的若干相近用户来产生推荐集,在用户数量较大的情况下,推荐的可信度随之降低。

(4) 实时性与推荐质量之间的平衡:推荐系统的推荐精度和实时性是一对矛盾。大部分推荐技术在保证实时性要求的同时,是以牺牲推荐系统的推荐质量为前提的。在提供实时推荐服务的同时,如何有效提高推荐系统的推荐质量,需要做进一步深入的研究。

(5) 效率更好的数据挖掘算法的研究:更有效的  $k$ -最近邻搜索(KNNs)算法和聚类(Clustering)算法能够提高推荐的实时性和准确性。目前的推荐系统中的 KNNs 算法存在实时性不足等缺陷,难以快速处理大规模的数据;质量高的聚类算法能够有效分割用户群<sup>[17]</sup>,适合推荐的聚类算法的性能也有待提高。

(6) 新型电子商务推荐系统体系结构研究:当前大部分的电子商务推荐系统都只是一个个单一的工具,只能提供一种推荐模型。但由于电子商务系统本身的复杂性,不同场合需要不同类型的推荐,因此,需要研究新型电子商务推荐系统体系结构,以有效集成多种推荐工具,收集多种类型的数据,提供多种推荐模型,使得不同的推荐工具组合使用、互补长短,以满足不同类型的推荐需求。

(7) 对推荐结果进行解释和谈判协商的研究:电子商务推荐系统为了说服用户,需要向用户解释推荐产生的原因。目前的电子商务推荐系统只能通过简单的销售排行、向用户提供其他用户对商品的评价评分信息等方式来达到上述目的。为了适应市场经济的特点和卖方销售策略的需要,支持客商之间就推荐结果展开讨价还价等谈判协商是必要的。因此,需要进一步研究更加有效的方法,向用户解释产生推荐的原因,支持买卖双方的谈判协商,以增加用户对推荐系统的信任程度,说服用户听从推荐系统的推荐。

### 第三节 本书拟解决的关键问题

电子商务推荐系统的研究内容,主要是根据推荐系统的研究现状和应用发展的需要,通过有效的方法提高关键算法的性能,有效利用各种数据资源,合理构建推荐系统的体系结构,提高推荐系统的效率。同时,就推荐结果向客户进行解释,便于客商双方的磋商,加强双方的了解,从而实现推荐系统的促