

# 营销科学学报

Journal of Marketing Science

# JMS

清华大学经济管理学院  
北京大学光华管理学院

- 个性化商品推荐：基于“最近邻居”的加权协同过滤推荐方法/ 孙鲁平 汪平 苏萌 1
- 消费者选择结果效价、放弃方案信息、满意和后悔/ 李东进 武瑞娟 李研 15
- 跨文化视角下的仿冒奢侈品购买动机研究/ 卢晓 Michel C.T. Phan 29
- 旅游业顾客参与对服务失误归因及行为意向的影响研究——以旅游形式为调节变量/ 彭艳君 42
- 折中效应和相似效应对消费者积分兑换奖品的偏好影响/ 陈荣 苏崧 黄劲松 59
- 到“实地”去：营销问题的人类学视角分析/ 王朝辉 沙振权 程瑜 72
- 服务品牌内化：结构、组成要素与概念界定/ 邱玮 白长虹 87
- 服务价值层级与银行客户情感反应研究/ 唐小飞 钟帅 梅发贵 陈滨桐 96
- 基于网页版式的网页营销策略研究/ 盛金根 左娟 夏曦 李永建 115

F713.50-55

20121

26

# 营销科学学报

## Journal of Marketing Science

第 7 卷第 4 辑

(总第 26 辑)

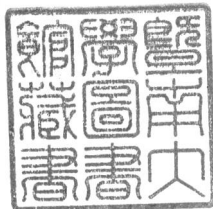
2011 年 12 月

Vol. 7, No. 4, December 2011

主办单位

清华大学经济管理学院

北京大学光华管理学院



科学出版社

北 京

## 内 容 简 介

《营销科学学报》是由清华大学经济管理学院和北京大学光华管理学院联合主办,由中国内地和中国香港 20 余所研究型大学的管理类学院共同协办的中国第一本市场营销领域的学报。《营销科学学报》作为中国市场营销学术研究的理论阵地,为海内外营销学者提供了一个进行创新性研究的交流平台,也获得了海内外营销学者的广泛认同。

《营销科学学报》已经连续出版了 7 卷 26 辑,它们在一定程度上反映了中国市场营销学科理论研究和应用研究的最新进展,适合从事市场营销相关研究的人员阅读,也可供对市场营销感兴趣的人员参考。

### 图书在版编目(CIP)数据

营销科学学报.第 7 卷,第 4 辑/清华大学经济管理学院,北京大学光华管理学院主编.—北京:科学出版社,2012

ISBN 978-7-03-033331-5

I. ①营… II. ①清… ②北… III. ①市场营销学-丛刊 IV. ①F713.50

中国版本图书馆 CIP 数据核字(2012)第 005032 号

责任编辑:陈 亮 张 宁 / 责任校对:邹慧卿

责任印制:张克忠 / 封面设计:陈 敬

科学出版社出版

北京东黄城根北街16号

邮政编码:100717

<http://www.sciencep.com>

双青印刷厂印刷

科学出版社发行 各地新华书店经销

\*

2012年1月第一版 开本:850×1168 1/16

2012年1月第一次印刷 印张:7 3/4

字数:220 000

定价:32.00 元

(如有印装质量问题,我社负责调换)

## 大 事 记

“2012年中国营销科学学术年会”将于2012年8月11~12日在大连召开。本次年会的主题是“经济发展方式转变与市场营销”，本次年会由《营销科学学报》主办，由大连理工大学和东北财经大学共同承办。

“2012年中国营销科学博士生论坛”将于2012年8月10日在大连理工大学举行。论坛将评选优秀博士论文一等奖1名，二等奖2名，三等奖3名。

2012年中国高等院校市场学年会将于2012年7月13~15日在内蒙古召开，本次会议由内蒙古财经学院承办。

中国高等院校市场学研究会2011年教学研讨会于10月29日在湖北武汉召开。本次教学年会由中国高等院校市场学研究会主办，高等教育出版社和中南财经政法大学联合承办。来自全国50多所高校的160多位从事市场营销教学和研究的教师就特色专业建设和教学方法创新两个议题展开了热烈的讨论。

西安交通大学教授、《营销科学学报》副主编庄贵军及其团队申请的“网络环境下基于新一代交互技术的客户关系与营销渠道管理”课题于2011年获国家自然科学基金重点项目资助，资助期限5年。

---

Personalized Product Recommendation: a Weighted Collaborative Filtering Algorithm Based on the Nearest Neighbors	1	Sun Luping, Wang Ping, Su Meng
The Chosen Outcome Valence, the Forgone Alternative Information, Satisfaction and Regret	15	Li Dongjin, Wu Ruijuan, Li Yan
Counterfeit Luxury Goods Purchase Motivation Study: a Cross Culture Perspective	29	Lu Xiao, Michel C. T. Phan
A Study on the Impact of Customer Participation on Attribution of Service Failure and Behavior Intention—Moditating by Tourism Type	42	Peng Yanjun
Effects of Compromise Effect and Similarity Effect on Consumers' Preference for Points Rewards Redemption	59	Chen Rong, Su Song, Huang Jinsong
To the "Field" : Marketing Analysis of Anthropology	72	Wang Zhaohui, Sha Zhenquan, Chen Yu
Service Brand Internalization: Structure, Component Elements, and Conceptual Definition	87	Qiu Wei, Bai Changhong
A Research on Service Value Hierarchy and Emotion Response of Bank Customers	96	Tang Xiaofei, Zhong Shuai, Mei Fagui, Chen Bintong
Study of Web Marketing Based on Web Page Layout	115	Sheng Jingen, Zuo Juan, Xia Xi, Li Yongjian

孙鲁平<sup>①</sup>,汪平<sup>②</sup>,苏萌<sup>③</sup>

**摘要** 个性化商品推荐是实现个性化营销的重要工具,网站可以根据消费者对商品的评分数据进行推荐,但评分数据的非随机缺失严重影响了现有推荐方法的效果。本文在“最近邻居”法的基础上,提出了加权协同过滤推荐(weighted collaborative filtering, WCF)方法。该方法将商品评分的缺失模式相似性视为重要的消费者偏好信息加入商品相似性的计算,改进了现有推荐方法只采用共同评分相似性的不足,同时该方法还根据共同评分的多少调整相似性以及合理选取“最近邻居”。采用电影评分数据的实证研究发现,在相似性调整参数 $\theta$ 和“最近邻居”数量 $K$ 的合理取值下,考虑了缺失模式的WCF方法均比传统方法的推荐效果好。而且,WCF方法还比常用于处理数据缺失的SVD(奇异值分解)方法有更高的推荐精度。不仅如此,当我们采用不同的方法计算缺失模式相似性及采用缺失程度不同的数据集时,WCF方法的推荐效果仍比传统推荐方法和SVD方法好。最后本文讨论了WCF方法对购物网站的实用价值。

**关键词** 个性化营销,个性化商品推荐,缺失模式,加权协同过滤推荐方法

## 个性化商品推荐:基于“最近邻居”的加权协同过滤推荐方法<sup>④</sup>

### 1 引言

随着互联网的普及,越来越多的消费者选择在网上购物。消费者购物方式的改变催生了大量购物网站,如淘宝商城、凡客诚品、红孩子商城等,中国的B2C(business-to-consumer)市场发展潜力巨大。根据中国互联网络信息中心(CNNIC)的调查,截止到2010年6月,中国网民已达4.2亿人,世界排名第一(赵光霞,2010)。而且,在网上购物的网民数量年增长率高达30%。艾瑞咨询的调查发现,2011年第一季度中国第三方网上支付交易规模达3650亿元,并预测第二季度的网上支付交易规模将达到4543

亿元,较2010年第一季度增长24.5%(程善宝,2011)。在网络购物中,网络口碑和个性化推荐对消费者的选择和购买发挥着越来越大的作用。前者是消费者发起的对产品或企业的评论;后者是企业发起的,根据每个消费者偏好的不同推荐相应的商品,实现个性化营销。相对于企业难以掌控的网络口碑,个性化推荐可以更直接地为企业所用。目前很多大型电子商务系统都有某种程度的商品推荐,亚马逊是其中的典范,其创始人杰夫·贝佐斯说:“100万个用户就应该有100万个亚马逊。”基于这种营销理念,亚马逊为用户提供“买过这本书的用户还买过什么”、“浏览过这本书的用户还浏览过什么”,这虽未完全实现个性化,但仍能有效地挖掘消费者的潜在需求。

① 孙鲁平,北京大学光华管理学院市场营销系博士研究生,E-mail: sunluping@gsm.pku.edu.cn。

② 汪平,北京大学光华管理学院市场营销系博士研究生,E-mail: wangping2@gsm.pku.edu.cn。

③ 苏萌,北京大学光华管理学院市场营销系副教授,E-mail: sumeng@gsm.pku.edu.cn。

④ 真诚感谢匿名评审专家为本文的改进和完善所提供的建设性意见。本研究得到国家自然科学基金项目(70802004)的资助。

相关研究表明,只有 16% 的人逛亚马逊时知道买什么,亚马逊 35% 左右的销售额正是来自其推荐系统。此外,美国 ChoiceStream 公司的用户调研也表明:71% 的用户愿意接受个性化的内容,45% 的用户更愿意访问提供个性化推荐的网站,其中 56% 的用户更愿意回访这些网站;在提高商品销售种类方面,37% 的音乐购买者、42% 的电影购买者和 57% 的移动内容购买者会因为适合的推荐而购买更多商品 (Cosgrove and Fougere, 2007),这意味着个性化推荐蕴藏着巨大商机。

在营销领域,很多研究已经发现网络口碑的数量、评分情况及在不同社区间传播的离散度对产品的销量、收益等有显著影响 (Chen et al., 2003; Chevalier and Mayzlin, 2006; Duan et al., 2005; Godes and Mayzlin, 2004; Hu et al., 2008; 邱凌云, 2008; 王晓玉和晁钢令, 2008; 袁喜娜等, 2010),而关于个性化商品推荐的研究却相对较少。Senecal 和 Nantel (2004) 发现个性化推荐比专家推荐和其他消费者推荐更能促进购买,同时,Howard 和 Kerin (2004) 发现个性化的商品推荐可以提高直邮广告的回复率。不仅如此,个性化商品推荐还可以降低消费者的搜寻努力并改善决策质量 (Häubl and Trifts, 2000),甚至可以影响消费者的长期偏好 (Häubl and Murray, 2003)。此外,通过提供个性化商品推荐,购物网站能增大消费者的转换成本、培养顾客忠诚 (Häubl et al., 2004)。然而,个性化商品如果推荐得不够准确,则将产生严重的负面影响。美国 ChoiceStream 公司的用户调研发现,39% 的用户不愿意再回到推荐质量差的网站,35% 的用户不愿意从推荐质量差的网站买东西 (Cosgrove and Fougere, 2007)。因此,个性化商品推荐是把双刃剑,研究如何精准地给消费者推荐商品对购物网站开展个性化营销起着至关重要的作用。

目前购物网站在准确地进行个性化推荐方面还面临很多困难。首先,独立的购物网站很难获得消费者完整的兴趣图谱。其次,每个购物网站都有很多种商品,提取、归纳商品的特征很困难,尤其对于体验型的产品(如电影、视频等),我们很

难制定统一的标准概括它们的特征,因此也很难量化消费者对这些商品的偏好。最后,即使不利用消费者和商品的特征,仅采用消费者对商品的评分数据进行推荐也存在技术难题。每个购物网站都有很多种类的商品,每个消费者仅可能给很少一部分商品进行评分,在现有推荐方法中,这种评分的缺失一般被视为随机的。但研究发现,消费者倾向于发表或关注极端评论而非中间评论 (郝媛媛等, 2010; Dellarocas et al., 2007),对一件商品的喜爱程度显著地影响消费者是否对其评分 (Marlin and Zemel, 2007),因此,评分的缺失在大多数情况下都是非随机的,简单的假设评分随机缺失可能产生不好甚至错误的推荐。

本文提出基于“最近邻居”的加权协同过滤推荐 (weighted collaborative filtering, WCF) 方法,该方法借用物以类聚的原理,根据消费者对与某个商品类似的商品群的偏好预测消费者对该商品的偏好。该方法不需要消费者和商品的特征,仅采用消费者对商品的评分数据,因此目前很多购物网站都可以采用。而且,在寻找与某商品相似的商品(即“最近邻居”)时,该推荐方法不仅考虑了两个商品在所有消费者中共同评分的相似性,还考虑了两个商品评分的缺失模式相似性,大大减少了非随机性评分缺失对推荐造成的不良影响。本文利用 MovieLens 的电影评分数据对 WCF 方法进行了实证研究,发现该方法比传统的协同过滤方法以及奇异值分解 (single value decomposition, SVD) 方法有更高的推荐质量,同时,对于缺失程度不等的多个数据集,我们都发现 WCF 方法较其他方法有更好的推荐效果。WCF 方法适用于中国很多购物网站,对竞争激烈的电子商务企业有重要的营销启示。

## 2 研究背景

### 2.1 个性化营销与个性化商品推荐

个性化营销,又称一对一营销,是指商家根据每个消费者的特殊偏好来相应调整自己的经营行为的营销理念,如针对每个消费者提供不同的促销方案、旗帜广告等,这些特殊偏好可能是

消费者提出的,也可能是企业从各种渠道主动搜集到的(Imhoff et al., 2001; Peppers and Rogers, 1995; Wind and Rangaswamy, 2001)。在个性化营销中,企业与市场将逐步建立一种新型关系,企业尽可能按每个顾客的偏好进行生产,以多品种、中小批量的混合生产取代过去的大批量生产,这有利于节省中间环节,降低销售成本,提高企业营销业绩(强海涛和蒋缨, 2002)。个性化营销的核心在于强调企业对个体的关注,注重建立良好的客户关系,企业通过建立消费者的个人数据库和信息档案,及时了解每位顾客的最新需求,向顾客提供个性化的产品和服务。Vargo 和 Lusch(2004)提出,关系营销是一个新的发展趋势,而个性化营销的重要目标就是培养客户关系,尤其在竞争激烈的网络零售中,个性化营销更是购物网站进行差异化的重要组成部分。

个性化营销的概念被提出后,随着购物网站逐渐引入数据挖掘技术,相关研究和实践也越来越多。目前将个性化营销理论应用于营销实践的研究多围绕 4P 策略展开,如个性化促销或定价(Chen et al., 2002; Liu and Zhang, 2006; Shaffer and Zhang, 2002; Vidyanand et al., 2005)和个性化广告(Xu et al., 2008)。而随着网络技术的发展,个性化商品推荐已成为营销与信息科学领域重叠的新研究方向,也是电子商务中非常重要的个性化营销工具。

根据实施技术的不同,个性化推荐可分为基于内容的推荐、协同过滤推荐和组合推荐。基于内容的推荐只根据商品的特征和消费者的历史偏好进行推荐,而不依赖其他消费者的评价,多被用于推荐含文本的商品,如网页(Pazzani and Billsus, 1997)和新闻(Centeno et al., 1999)等,常用词频-倒排文档频率(TF-IDF)(Resnick and Hal, 1997)、聚类分析、贝叶斯分类算法(Mooney and Bennett, 1998)等技术。这种推荐可以向用户推荐新商品,但不适合构造非常复杂的商品。协同过滤是目前最成功的推荐方法,它可以只利用其他消费者的评价进行推荐。协同过滤常采用“最近邻居”(Resnick et al., 1994; 邓爱林等, 2003)、聚类分析(O'Conner and Herlocker,

1999)、数据降维(Sarwar et al., 2000)以及贝叶斯网络(Breese et al., 1998)等方法。该方法的推荐质量容易受到数据缺失的影响,且对“新商品”和“新用户”的推荐效果不佳。组合推荐方法综合以上两种方法产生推荐,分为三种类型(许海玲等, 2009):①后融合,即直接融合两种方法产生的推荐结果(Claypool et al., 1999; Pazzani and Billsus, 1997);②中融合,指以一种方法为框架,融合另一种方法(郭艳红和邓贵仕, 2008);③前融合,指将两种方法整合到统一框架下,一般用基于模型的方法实现。组合推荐方法兼具两种方法的优点,但运算量大、过程复杂。

## 2.2 国内外管理领域的相关研究

协同过滤法是最常用的个性化推荐方法,分为基于模型的方法和基于记忆的方法。在管理领域,国外关于协同过滤推荐的学术研究主要利用消费者和商品的信息进行建模。Ansari 等(2000)将消费者和商品的异质性加入分层贝叶斯模型,取得了比基于记忆的方法更好的推荐质量。Ying 等(2006)认为评分缺失是非随机的,他们采用基于模型的方法,不仅考虑了消费者和产品的异质性,还估计了评分数据缺失与否的概率分布,提高了推荐质量。Marlin 和 Zemel(2007)也发现评分的缺失是非随机的,并采用 CPT- $v$  缺失数据模型提高了推荐质量。国外的很多其他研究都没有考虑评分缺失的非随机性,而 Ying 等(2006)、Marlin 和 Zemel(2007)虽然考虑了这一点,但他们的模型需要大量的消费者和商品信息,而且计算复杂,这对于中国目前尚无完备的消费者和商品信息的购物网站都是不实用的。

国内管理领域关于协同过滤推荐的研究多为基于记忆的方法,尤其是“最近邻居”法。这些学者的贡献在于将“最近邻居”法与营销领域里的 RFM (recency, frequency, monetary value) 模型、CRM(customer relationship management) 等进行了结合。赵晓煜和丁延玲(2006)提出采用顾客的历史购买数据而非商品评分进行推荐,孙玲芳和张婧(2010)对此进行了实证研究,首先



利用 RFM 模型对购买数据进行分析,推断出消费者对商品的偏好,再对偏好采用“最近邻居”法。而王卫平和吴伦(2007)则先识别出 CRM 中的优质客户群,然后用交易数据进行“最近邻居”法的销售机会推荐,发现采用“最近邻居”法推荐比 Apriori 算法更能促进交叉销售。此外,国内还有管理学者研究个性化文档的推荐(庞秀丽等,2008),以及对个性化推荐的技术进展进行回顾和述评(李聪等,2011;刘建国等,2009;朱岩和林泽楠,2009)。

关于“最近邻居”法的技术创新集中在信息科学领域,其中多数研究通过改进“最近邻居”的寻找方法(即相似性度量)减少数据稀疏性对推荐的影响。目前常用的相似性度量有 Pearson 相关系数,受限制的 Pearson 相关系数以及余弦相似性等(李聪,2010;刘建国等,2009),学者们对这些方法进行了各种改进。例如,邓爱林等(2003)用两个用户评分的并集而非交集计算用户相似性,提出了基于项目评分预测的协同过滤算法;李聪等(2008)也采用两个用户评分的并集计算相似性,但他们将用户进一步区分为无推荐能力和有推荐能力,只对有推荐能力的用户计算相似性,改进了推荐质量。此外,数据降维也是常用的处理数据稀疏性的推荐方法(李聪等,2011;Cao et al., 2011; Adomavicius and Tuzhilin, 2005), SVD 是最常用的方法,孙小华等(2006)提出的 Pear-After-SVD 方法先用 SVD 对缺失数据预估,再用 Pearson 相似性寻找“最近邻居”,提高了“最近邻居”法的推荐质量。近几年,学者们还提出了一些新的数据降维方法,如概率矩阵分解(probabilistic matrix factorization, PMF)(Salakhutdinov and Mnih, 2008a)、贝叶斯概率矩阵分解(Salakhutdinov and Mnih, 2008b)等,但这些方法都采用了概率模型,预测结果很容易受到数据特征和先验参数的影响,而且计算相对复杂,因此本文只对传统的协同过滤方法和未采用概率模型的 SVD 方法进行推荐效果对比。

无论是管理领域还是信息科学领域,数据的稀疏性都是“最近邻居”法的瓶颈。虽然学者们

对改进推荐质量做出了突出贡献,但这些改进仍然没有突破数据随机缺失的假设,目前国内外的很多研究仍假设评分的缺失是随机的。

### 2.3 缺失数据处理

虽然国外已有研究发现评分的缺失是非随机的(Marlin and Zemel, 2007; Ying et al., 2006),国内相关研究仍假设缺失是随机的,对这种非随机性缺失如果不予以考虑,则可能产生不准确甚至错误的推荐。

根据 Little 和 Rubin(1987)的缺失数据统计分析理论,数据的缺失可分为三类:

(1) 完全随机缺失(missing completely at random, MCAR),这种数据缺失不会对统计分析造成偏差,因此可以被忽略。

(2) 随机缺失(missing at random, MAR),这种数据缺失依赖于数据中可观测的变量,因此可以采用多重填补的方法减低甚至消除缺失对统计分析造成的偏差。

(3) 非随机缺失(missing not at random, MNAR),这种缺失依赖于数据中不可观测的变量(即缺失变量本身),目前尚没有统一的方法来处理这类缺失。

面临非随机缺失,数据中可观测的信息将无法预测缺失机制,此时可以采用缺失模式(missing-indicator)方法(Donders et al., 2006; Miettinen, 1985),即将变量的缺失与否编译成一个哑变量(变量缺失则取值为 0,非缺失则取值为 1),在回归中将这个哑变量作为一个控制变量。这种方法可以在一定程度上减少非随机缺失造成的不良影响。

在协同过滤推荐中,评分的缺失可能有两个来源:一个是消费者从来没有购买过某个商品,因此没有给出评分;另一个是消费者购买过某个商品但感觉很一般,并不特别喜欢也没有极端的厌恶,因此也没有给出评分。无论哪种情况,评分的缺失与否都包含了一部分消费者的偏好信息,因此在个性化推荐中应充分利用这部分信息。而国内目前常用的“最近邻居”法以及专门用于处理数据稀疏性的 SVD 方法等都没有考虑



加权协同过滤推荐方法认为商品的相似性有两个来源:一是商品共同评分的相似性[见式(2)],另一个是商品被评分与否的相似性,即缺失模式相似性。缺失模式相似性  $\text{sim}_{\text{missing}}(s, s')$  用缺失模式矩阵  $M$  中商品  $s$  和  $s'$  所对应行的 Pearson 相关系数度量,计算方法与式(2)类似,则

$$\text{sim}_{\text{missing}}(s, s') = \frac{\sum_{c \in C, s \in S} (m_{c,s} - \bar{m}_s)(m_{c,s'} - \bar{m}_{s'})}{\sqrt{\sum_{c \in C, s \in S} (m_{c,s} - \bar{m}_s)^2 \sum_{c \in C, s' \in S} (m_{c,s'} - \bar{m}_{s'})^2}} \quad (4)$$

其中,  $m_{c,s}$  和  $m_{c,s'}$  分别表示用户  $c$  对商品  $s$  和  $s'$  的评分是否存在,存在则取值为 1, 否则取值为 0。  $\bar{m}_s$  和  $\bar{m}_{s'}$  分别表示  $M$  矩阵中第  $s$  和  $s'$  行的行均值。此外,在数据挖掘领域,计算两个 0-1 向量(即元素只包括 0 和 1 的向量,其中,1 表示数据非缺失,0 表示数据缺失)之间的相似性,还常用 Hamann 系数、Yule 系数、Rogers & Sneath 方法以及匹配法等(Cheetham and Hazel, 1969),这些方法均通过两个 0-1 向量共同拥有 1 与共同拥有 0 的次数来度量向量间的相似性,因此原理非常类似。为了检验 WCF 方法的优越性是否在不同的缺失模式相似性的计算方法下都成立,本文还选取应用较为广泛的 Hamann 系数与 Pearson 相关系数作了对比分析。Hamann 系数的计算公式如下:

$$\text{sim}_{\text{missing}}^H(s, s') = \frac{(P + A) - (E_s + E_{s'})}{N_{s'} + A} \quad (5)$$

其中,  $P$  为商品  $s$  和  $s'$  的评分同时为非缺失的次数;  $A$  为两个商品的评分都为缺失的次数;  $E_s$  为商品  $s$  的评分非缺失而商品  $s'$  的评分缺失的次数;  $E_{s'}$  为商品  $s$  的评分缺失而商品  $s'$  的评分非缺失的次数;  $N_{s'}$  为商品  $s$  或  $s'$  的评分非缺失的总次数。

传统方法只考虑共同评分的相似性,而本文用商品  $s$  和  $s'$  的缺失模式相似性  $\text{sim}_{\text{missing}}(s, s')$  与其共同评分相似性  $\text{sim}_{\text{initial}}(s, s')$  加权,得一个更好的商品相似性。此外,当采用 Pearson 相关系数度量两个商品的共同评分相似性时,一般需要两个商品有较多的共同评分,如果共同评分过少,就会得到不可靠的商品相似性,直接影响推

荐质量。而且,在预测评分时,“最近邻居”的个数也可能显著地影响推荐质量。因此,本文提出的 WCF 方法将同时考虑以下三个方面。

(1) 根据共同评分多少调整商品相似性。两个商品的共同评分越多,它们之间的相关系数就越可靠。假设  $\text{sim}_{\text{initial}}(s, s')$  为由式(2)得来的商品共同评分的相似性,则共同评分相似性应根据两个商品的共同评分个数进行如下调整:

$$\text{sim}'_{\text{initial}}(s, s') = k \times \text{sim}_{\text{initial}}(s, s'), k = \begin{cases} \frac{n}{\theta}, n < \theta \\ 1, n \geq \theta \end{cases} \quad (6)$$

其中,  $k$  为调整权重;  $n$  为商品  $s$  和  $s'$  共同评分的个数。在现有研究中,  $\theta$  常取 20~60 不等,本文将尝试不同的  $\theta$  值,检验 WCF 方法在各种  $\theta$  值下的推荐效果。

(2) 考虑缺失模式的相似性。由于评分缺失往往不是随机的,因此我们计算商品  $s$  和  $s'$  缺失模式的相似性  $\text{sim}_{\text{missing}}(s, s')$ , 并与  $\text{sim}'_{\text{initial}}(s, s')$  加权平均,得到综合相关系数  $\text{sim}_w(s, s')$ ,

$$\text{sim}_w(s, s') = \alpha \text{sim}_{\text{missing}}(s, s') + (1 - \alpha) \text{sim}'_{\text{initial}}(s, s') \quad (7)$$

其中,  $\alpha$  为缺失矩阵相似性所占的权重。

(3) 合理选取“最近邻居”。推荐某商品时,与之很相似的商品才有参照价值。“最近邻居”的数量可以直接给定,也可以通过相似性阈值来确定(李聪, 2010),本文通过规定“最近邻居”的数量选取“最近邻居”,即与某个商品最相似的前  $K$  个商品被选为该商品的“最近邻居”。在现有研究中,  $K$  的选取依照数据中的商品数量和缺失程度而有所不同,本文将尝试不同的“最近邻居”数量。我们将商品  $s'$  与其他所有商品的相关系数中最大的  $K$  个相关系数保留,其他的设为零,得到  $s'$  与任意商品  $s$  的最终相关系数  $w\text{sim}(s, s')$ :

$$\begin{aligned} w\text{sim}(s, s') &= I_{x_{(K)}}[\text{sim}(s, s')_w] \\ &= \begin{cases} 0, \text{sim}(s, s')_w < x_{(K)} \\ \text{sim}(s, s')_w, \text{sim}(s, s')_w \geq x_{(K)} \end{cases} \end{aligned} \quad (8)$$

其中,  $x_{(i)}, i=1, 2, \dots, S$  为列向量  $\text{sim}_w(S, s')$  的

次序统计量,即  $x_{(K)}$  为所有商品与商品  $s'$  的相关系数中排名为  $K$  的相关系数。

综合以上考虑,则

$$w\text{sim}(s, s') = I_{x_{(K)}} \left[ \alpha \text{sim}_{\text{missing}}(s, s') + (1 - \alpha) \frac{k \times \sum_{c \in C_{s'}} (r_{c,s} - \bar{r}_s)(r_{c,s'} - \bar{r}_{s'})}{\sqrt{\sum_{c \in C_{s'}} (r_{c,s} - \bar{r}_s)^2 \sum_{c \in C_{s'}} (r_{c,s'} - \bar{r}_{s'})^2}} \right] \quad (9)$$

WCF 方法预测用户  $c$  对商品  $s'$  的评分为  $\tilde{r}(c, s')$ :

$$\tilde{r}(c, s') = \frac{\sum_{s \neq s'} w\text{sim}(s, s') \times r_{c,s}}{\sum_{s \neq s'} |w\text{sim}(s, s')|} \quad (10)$$

即

$$\tilde{r}(c, s') = \frac{\sum_{s \neq s'} \left\{ I_{x_{(K)}} \left[ \alpha \text{sim}_{\text{missing}}(s, s') + (1 - \alpha) \frac{k \times \sum_{c \in C_{s'}} (r_{c,s} - \bar{r}_s)(r_{c,s'} - \bar{r}_{s'})}{\sqrt{\sum_{c \in C_{s'}} (r_{c,s} - \bar{r}_s)^2 \sum_{c \in C_{s'}} (r_{c,s'} - \bar{r}_{s'})^2}} \right] \times r_{c,s} \right\}}{\sum_{s \neq s'} \left| I_{x_{(K)}} \left[ \alpha \text{sim}_{\text{missing}}(s, s') + (1 - \alpha) \frac{k \times \sum_{c \in C_{s'}} (r_{c,s} - \bar{r}_s)(r_{c,s'} - \bar{r}_{s'})}{\sqrt{\sum_{c \in C_{s'}} (r_{c,s} - \bar{r}_s)^2 \sum_{c \in C_{s'}} (r_{c,s'} - \bar{r}_{s'})^2}} \right] \right|} \quad (11)$$

## 4 实证研究

为了验证本文提出的 WCF 方法的优越性,我们利用电影评分数据比较 WCF 方法和传统 CF 方法以及 SVD 方法的推荐效果。

### 4.1 数据描述

本文采用 MovieLens 的电影评分数据,包括 943 名用户对 1682 部电影的 100 000 个评分。其中每个用户至少对 20 部电影评分,评分采用 5 分量表,越高代表用户越喜欢这个电影。我们将这评分随机分成估计样本(90 000 个)和测试样本(10 000 个),估计样本用于优化参数,测试样本用于检验各方法的预测效果。

若用户对每个电影评分,应有 1 586 126 (943×1682)个评分,而实际只有 100 000 个,数据缺失高达 93%。描述性分析显示,数据缺失模式可能包含着重要的偏好信息,与 Ying 等(2006)的发现类似,一些电影类型如黑色电影,虽然评分数很少,但平均评分却是 18 类电影中最高的(图 1)。这说明,可能只有非常喜欢这种

电影的用户看了这部电影并且给出了评分。这种情况对于很流行的电影不太明显,但在偏冷门的电影中很严重。基于这种考虑,本文提出 WCF 方法来更好地寻找电影的“最近邻居”,然后对测试样本中的 10 000 个评分进行预测。

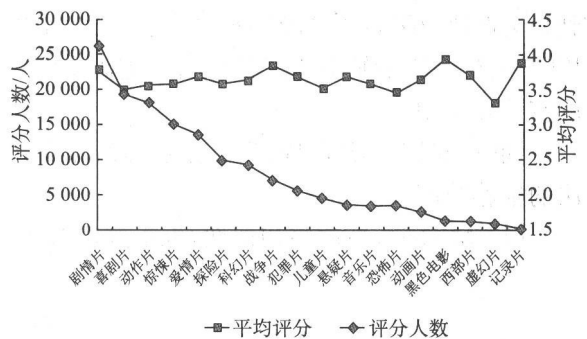


图 1 各类型电影的评分人数及平均评分

### 4.2 WCF 及与不同 CF 算法的比较

推荐中常用平均绝对偏差(mean absolute error, MAE)衡量推荐质量,MAE 值越小,推荐质量就越高。

$$MAE = \frac{\sum_{i=1}^N |\tilde{r}_i - r_i|}{N} \quad (12)$$

其中,  $N$  为测试样本中的评分数量;  $\tilde{r}_i$  为预测评分;  $r_i$  为真实评分。

为检验 WCF 方法对推荐质量的改进在统计上是否显著, 我们采用 Bootstrap 的原理 (Little

and Rubin, 1987), 从原始数据中随机抽取估计样本 (90 000 个评分) 和测试样本 (10 000 个评分) 50 次, 使用每次随机抽取的估计样本进行参数优化, 测试样本用于计算推荐精度 (MAE)。表 3 中的结果均为 50 次计算的均值。

表 3 不同模型不同权重  $\alpha$  下的 MAE 值

模型/ $\alpha$	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Model1	1.4056	1.3162	1.2311	1.1517	1.0780	1.0111	0.9530	0.9054	0.8724	0.8606	0.8828
Model2	1.2061	1.0916	0.9951	0.9206	0.8707	0.8431	0.8315	0.8311	0.8497	0.8565	0.8810
Model3	0.9488	0.9158	0.9132	0.9084	0.8977	0.8718	0.8289	0.7689	0.7384	0.7312	0.7345
WCF	0.8028	0.7902	0.7782	0.7682	0.7589	0.7490	0.7419	0.7342	0.7291	0.7279	0.7319

注: 此处取“最近邻居”的数量  $K=50$ , 共同评分对的调整参数  $\theta=30$ 。

为验证缺失模式的重要作用, 我们分别对缺失模式相似性在商品相似性中的权重  $\alpha$  为 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0 的情形进行了计算 (表 3 最后一行)。结果发现, 对于 WCF 方法, 当  $\alpha=0$  时, 即完全不考虑缺失模式相似性时, MAE 是最高的 (0.8028), 随着  $\alpha$  逐渐增加, MAE 逐渐降低, 即推荐精度越来越高。当  $\alpha=0.9$  时, WCF 方法的 MAE 达到最小 (0.7279), 比  $\alpha=0$  时降低了 9.33%, 推荐精度得到显著改进 ( $t(49)=88.53, p<0.001$ )。这说明 WCF 方法考虑的缺失模式信息对于提高推荐精度有重要作用。

同时, 我们比较了 WCF 方法和传统推荐方法的推荐精度:

(1) Model1, 不调整相似性, 也不选取“最近

邻居”;

(2) Model2, 调整相似性 ( $\theta=30$ ), 但不选取“最近邻居”;

(3) Model3, 不调整相似性, 但选取“最近邻居” ( $K=50$ )。

而 WCF 方法则既调整了相似性又选取了“最近邻居”。对这三个对比模型, 我们同样分别计算了缺失模式相似性占不同权重时的 MAE (表 3 和表 4), 发现 Model1 在  $\alpha=0.9$  时达到最优 MAE (0.8606); Model2 在  $\alpha=0.7$  时 MAE 最优 (0.8311); Model3 的最优权重  $\alpha=0.9$ , 最优 MAE 为 0.7312。三个模型的最优 MAE 都比相应的  $\alpha=0$  时的 MAE 值有显著降低。这再次验证了考虑缺失模式对提升推荐精度的重要作用。

表 4 模型对比

模型	“最近邻居”选取	相似性调整	缺失模式				
			MAE( $\alpha=0$ )	最优权重	最优 MAE	MAE 改进*	改进显著性
Model1	不考虑	不考虑	1.4056	0.9	0.8606	0.5450	$p<0.001$
Model2	不考虑	考虑	1.2061	0.7	0.8311	0.3750	$p<0.001$
Model3	考虑	不考虑	0.9488	0.9	0.7312	0.2176	$p<0.001$
WCF	考虑	考虑	0.8028	0.9	0.7279	0.0749	$p<0.001$

\* 指  $\alpha=0$  时 (即不考虑缺失模式) 的 MAE 减去最优 MAE。

我们还发现,即使在都不考虑缺失模式相似性( $\alpha=0$ )时,WCF方法的MAE(0.8028)在四个模型中仍是最底的,即WCF方法相对于三个对比模型推荐精度更高。这说明,WCF方法比三个对比模型的推荐精度更高的原因,不仅仅在于考虑了缺失模式相似性,合理地选取“最近邻居”和根据共同评分调整相似性也发挥了重要作用。而在WCF方法考虑了缺失模式相似性后,最优MAE降为0.7279,比不考虑缺失模式相似性的三个对比模型的推荐精度改进更大。而且,即使在三个对比模型中也考虑缺失模式相似性,对比四个模型的最优MAE(表4),WCF方法的MAE(0.7279)同样是最底的,而且在统计上是显著的( $p<0.001$ )。而对比四个模型考虑缺失模式前后的MAE变化,发现Model1和Model2的MAE改进最大(分别为0.5450和0.3750),而Model3和WCF方法的MAE改进较小,分别降低了0.2176和0.0749(表4)。这说明合理选取“最近邻居”对推荐精度的提升远大于相似性调整。

同时,采用Hamann系数[见式(5)]计算缺失模式相似性,我们得到了类似的结论。我们发现WCF方法的最优MAE相对于传统推荐方法仍然是最低的,此时 $\alpha=0.6$ ,最优MAE为0.7659。而Model1,Model2和Model3的最优MAE均在0.82以上,都比WCF方法的最优

MAE高。而且,我们发现采用Hamann系数计算缺失模式相似性,使得每种推荐方法的最优MAE都提高了,因此用Pearson相关系数计算缺失模式相似性在本研究中可能是更好的选择。

为了检验WCF方法(相对于传统方法)的推荐效果对“最近邻居”的数量 $K$ 和共同评分对数 $\theta$ 的敏感性,我们分别选取 $K=50, K=100, K=150, K=200, K=250, K=300$ 以及 $\theta=20, \theta=30, \theta=40, \theta=50, \theta=60$ ,对WCF方法及传统方法的MAE进行了计算,发现 $K$ 的选取对WCF方法及传统方法推荐效果的影响远大于 $\theta$ ,不管 $\theta$ 取哪个值, $K$ 值对各个模型推荐效果的影响都是非常类似的,我们以 $\theta=30$ 为例作讨论(图2)。我们发现,在 $K$ 的每个取值下,WCF方法和Model3的最优MAE都是最小的,其次是Model2,最优MAE最大的是Model1。同时,相对于 $K$ 的其他可能取值, $K=50$ 时每个模型的最优MAE都是最低的。因此,对于本文的数据,我们采用的“最近邻居”数量 $K=50$ 作进一步讨论。在给定 $K$ 取值的条件下,当 $\theta$ 取值从20到60变化时,由于Model1和Model3没有考虑相似性调整,因此其最优MAE保持不变;而对于WCF方法和Model2,我们发现 $\theta$ 取各种值时,它们的最优MAE变化很小。而且不管 $\theta$ 取多少,WCF方法的最优MAE总是最低的,本文取WCF方法的最优MAE相对较低的 $\theta=30$ 。

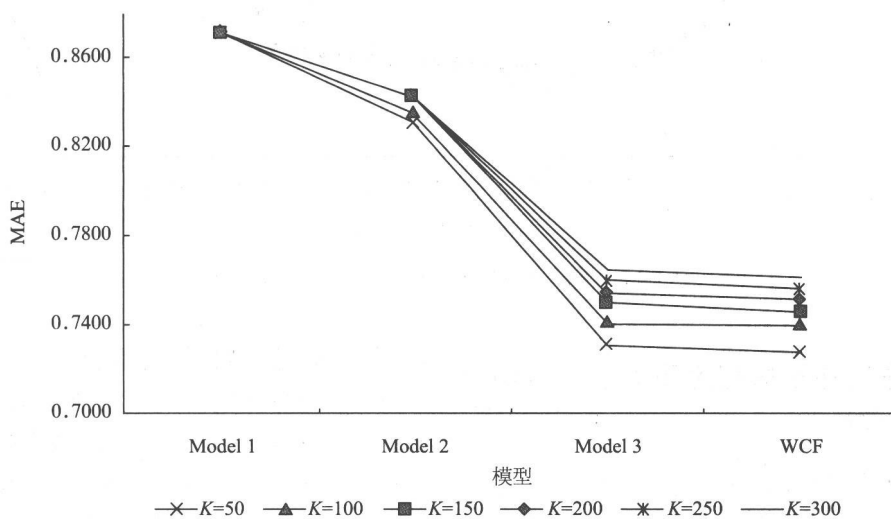


图2 不同“最近邻居”数量 K 下各模型的 MAE

### 4.3 WCF 方法与 SVD 方法的比较

SVD 方法是目前个性化推荐中最常用的数据降维方法(Adomavicius and Tuzhilin, 2005; Deerwester et al., 1990; Strang, 2009)。采用 SVD 方法对稀疏矩阵进行压缩,可以更好地处理评分数据的大量缺失,达到比传统推荐方法更高的推荐质量(Sarwar et al., 2000)。虽然近几年有些学者还提出了基于概率模型的 PMF、BMF 等方法,发现它比 SVD 方法的推荐质量更高(Salakhutdinov and Mnih, 2008a, 2008b),但这些方法都假设商品评分服从一定的概率分布,而且计算相对复杂,因此本文主要将 WCF 方法与业界最常用的不基于概率模型的 SVD 方法进行比较。

Deerwester 等(1990)提出的 SVD 方法如下:

- (1) 使用 SVD 方法将原始评分矩阵分解为  $U, S$  和  $V$ ;
- (2) 将矩阵  $S$  降为  $k$  维;
- (3) 计算降维后的矩阵  $S_k$  的平方根  $S_k^{1/2}$ ;

(4) 分别计算矩阵  $U_k S_k^{1/2}$  和  $S_k^{1/2} V_k'$ ;

(5) 用户  $c$  对电影  $s'$  的预测评分为  $\hat{r}(c, s') = \bar{C} + U_k \sqrt{S_k}(c) \cdot \sqrt{S_k} V_k'(s')$ 。

我们分别用行均值和列均值进行了计算,结果发现采用列均值时 SVD 的推荐质量更好,这与 Sarwar 等(2001)的研究结果是一致的,因此本文只报告采用列均值得出的结果。为保证可比性, SVD 方法将采用与 WCF 方法相同的电影评分数据,并且随机抽样并计算 50 次。每次计算都分别取  $k=1, 2, \dots, 30, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550$  和 600,重复步骤(1)~(5)。

结果显示,  $k=12$  时 SVD 方法推荐质量最好(图 3),此时  $MAE_{SVD}=0.7659$ ,与 WCF 方法的最优  $MAE(\alpha=0.9$  时,  $MAE_{WCF}=0.7279$ )相比, SVD 方法的推荐精度比 WCF 方法低 5.22%。这表明,即使是目前常见的处理数据稀疏性的 SVD 方法,也没有考虑了缺失模式相似性的 WCF 方法推荐精度高,这再次印证了 WCF 方法的优越性。

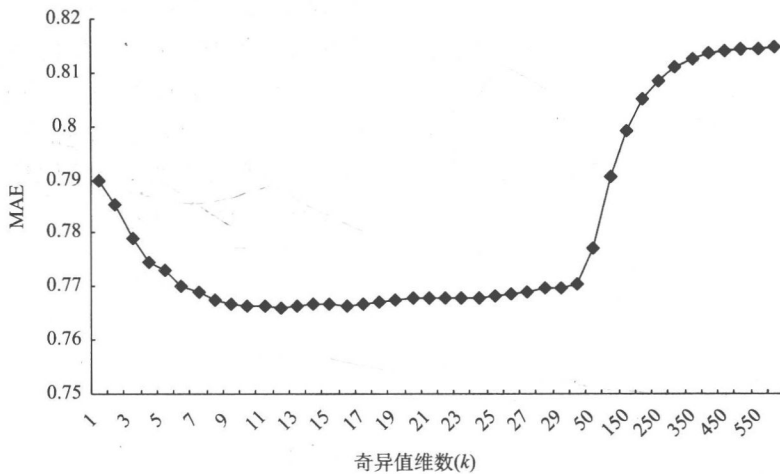


图 3 不同维数的 SVD 算法的 MAE 值

### 4.4 WCF 方法在缺失程度不同的数据中的推荐效果

本文除了利用 MovieLens 的完整数据集,还设计了四个缺失程度不同的数据来进一步验证 WCF 方法的推荐效果。沿用 Cao 等(2011)的做

法,我们从完整数据中抽出评分不超过 30、40、50、60 的电影分别组成新的数据集  $N=30, N=40, N=50, N=60$ ,这四个数据的缺失程度逐渐降低。我们用每个数据分别计算 Model1、Model2、Model3、SVD 方法以及 WCF 方法的最优 MAE(表 5)。

表 5 缺失程度不同的数据集中各推荐方法的效果比较

数据集	Model1	Model2	Model3	SVD	WCF
N=30	0.9921	0.9794	0.8672	0.8961	0.8451
N=40	0.9305	0.9204	0.8311	0.8852	0.8212
N=50	0.9231	0.9090	0.8408	0.8682	0.8187
N=60	0.8917	0.8793	0.8106	0.8708	0.8002

注:每个数据集在最优 MAE 时的参数可能不同,但对于同一数据集,不同模型采用的参数是相同的。

我们发现,对于每个数据集,WCF 方法的最优 MAE 都是最低的,即对于缺失程度不等的四个数据集,WCF 方法的推荐效果都是最好的。而且随着缺失程度的逐渐降低,WCF 方法的最优 MAE 也逐渐降低,从 0.8451 变为 0.8002,而 Model1、Model2、Model3 以及 SVD 方法等推荐方法的最优 MAE 虽然也逐渐降低,但都比 WCF 方法的最优 MAE 高。因此,这进一步验证了 WCF 方法在推荐效果上的优越性。

## 5 结论和讨论

随着网络零售竞争的日益加剧,个性化营销对于争取新客户、培养忠诚客户、提高销量变得越来越重要。个性化促销、定价、广告等在实施的过程中尚存在很多困难,鉴于单个购物网站很难获取全面的消费者信息,同时结构化地归纳和提炼商品的特征也非常困难,因此个性化的商品推荐很难做好。而且,即使不用消费者信息和商品特征,仅采用消费者对商品的评分信息,评分缺失的非随机性也严重影响了现有推荐方法的效果。本文在“最近邻居”法的基础上,提出了 WCF 方法,在计算商品之间的相似性时考虑了缺失模式相似性,改进了传统推荐方法只采用共同评分相似性的不足。采用 MovieLens 的电影数据,我们发现考虑缺失模式相似性可以显著地提高推荐精度,在缺失模式相似性的权重为 0.9 时 WCF 方法达到最优 MAE(0.7279),而根据共同评分的多少调整相似性以及合理选取“最近邻居”也可以显著提高推荐精度,但合理选择“最近邻居”对改进推荐精度的作用更大。此外,为进一步验证 WCF 方法的优越性,本文还与传统推

荐方法及 SVD 方法进行了对比,发现 WCF 方法的推荐精度比传统推荐方法及 SVD 方法有显著提高。不仅如此,当我们采用不同的方法计算缺失模式相似性,以及采用缺失程度不同的数据集时,WCF 方法的推荐效果均比传统的协同过滤方法和 SVD 方法好。

本文的主要贡献在于将评分数据的缺失视为非随机的,并借助缺失数据分析理论,将评分的缺失模式视为有用的偏好信息加入个性化商品推荐,不仅如此,我们还根据商品共同评分的多少调整相似性以及合理的选取“最近邻居”,并最终提出了 WCF 方法。而且实证研究发现 WCF 方法比传统推荐方法以及 SVD 方法的推荐效果更好。该方法计算简便,对拥有消费者评分数据的网站具有很高的实用价值,能够帮助这些网站提升个性化营销的效果。该方法还可以很方便地被应用于其他数据类型,目前很多购物网站只拥有消费者的浏览和购买数据,这些隐性评分数据也可以被用于 WCF 方法的个性化商品推荐,不仅能帮助网站提高在线成交量,还可以给顾客提供附加价值,让顾客对网站的个性化推荐产生长期依赖,进而形成顾客忠诚。

本文的局限在于,采用的数据是国外的电影评分数据,将来希望能够获得国内购物网站的商品评分数据,用国内的数据来验证 WCF 方法的推荐质量。而且,将来还可以利用购物网站的消费者购买数据,采用 WCF 方法进行个性化商品推荐,进一步验证该方法的推荐质量。最后,如果购物网站可以提供消费者与商品的特征信息,WCF 方法还可以结合其他的缺失数据处理方法,帮助购物网站进一步提高个性化商品的推荐质量。



## 参考文献

- [1] 程善宝. 2011. 艾瑞咨询:2011Q1 中国第三方网上支付交易规模达 3650 亿元[EB/OL]. <http://ec.iresearch.cn/54/20110419/137647.shtml> [2011-11-11].
- [2] 邓爱林,朱扬勇,施伯乐. 2003. 基于项目评分预测的协同过滤推荐算法[J]. 软件学报, 14(9): 1621-1628.
- [3] 郭艳红,邓贵仕. 2008. 协同过滤系统项目冷启动的混合推荐算法[J]. 计算机工程, 34(23):11-13.
- [4] 郝媛媛,叶强,李一军. 2010. 基于影评数据的在线评论有用性影响因素研究[J]. 管理科学学报, 13(8):78-96.
- [5] 李聪. 2010. 电子商务协同过滤可扩展性研究综述[J]. 现代图书情报技术, (11):37-41.
- [6] 李聪,梁昌勇,马丽. 2008. 基于领域最近邻的协同过滤推荐算法[J]. 计算机研究与发展, 45(9): 1532-1538.
- [7] 李聪,梁昌勇,杨善林. 2011. 电子商务协同过滤稀疏性研究: 一个分类视角[J]. 管理工程学报, 25(1):94-101.
- [8] 刘建国,周涛,汪秉宏. 2009. 个性化推荐系统的研究进展[J]. 自然科学进展, 19(1):1-15.
- [9] 庞秀丽,冯玉强,姜维. 2008. 电子商务个性化文档推荐技术研究[J]. 中国管理科学, 16(S1):581-586.
- [10] 强海涛,蒋纓. 2002. 个性化营销的理念与策略[J]. 重庆商学院学报, (4):50-52.
- [11] 邱凌云. 2008. 网上口碑的信息效价与情感线索对说服效果的影响机制研究[J]. 营销科学学报, 4(4):32-44.
- [12] 孙玲芳,张婧. 2010. 基于RFM模型和协同过滤的电子商务推荐机制[J]. 江苏科技大学学报(自然科学版), 24(3):285-289.
- [13] 孙小华,陈洪,孔繁胜. 2006. 在协同过滤中结合奇异值分解与最近邻方法[J]. 计算机应用研究, (9): 206-207.
- [14] 王卫平,吴伦. 2007. 协同过滤在CRM交叉销售中的应用研究[J]. 管理学报, 4(4):436-441.
- [15] 王晓玉,晁钢令. 2008. 产品危机中口碑方向对消费者态度的影响[J]. 营销科学学报, 4(4):1-12.
- [16] 许海玲,吴潇,李晓东,等. 2009. 互联网推荐系统比较研究[J]. 软件学报, 20(2):350-362.
- [17] 袁喜娜,张海林,姜旭平,等. 2010. 网络口碑营销分析——基于韩国网上影评和票房收入的实证研究[J]. 营销科学学报, 6(1):41-58.
- [18] 赵光霞. 2010. 中国网民数量达 4.2 亿手机网民 2.77 亿[EB/OL]. <http://media.people.com.cn/GB/40606/12156872.html> [2011-11-11].
- [19] 赵晓煜,丁延玲. 2006. 基于顾客交易数据的电子商务推荐方法研究[J]. 现代管理科学, (3):93-94.
- [20] 朱岩,林泽楠. 2009. 电子商务中的个性化推荐方法评述[J]. 中国软科学, (2):183-192.
- [21] Adomavicius G, Tuzhilin A. 2005. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions[J]. IEEE Transactions on Knowledge and Data Engineering, 17(6):734-749.
- [22] Ansari A, Essegai S, Kohli R. 2000. Internet recommendation systems [J]. Journal of Marketing Research, 37 (3):363-375.
- [23] Breese J S, Heckerman D, Kadie C. 1998. Empirical analysis of predictive algorithms for collaborative filtering [C]. Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence:43-52.
- [24] Cao B, Yang Q, Sun J T, et al. 2011. Learning bidirectional asymmetric similarity for collaborative filtering via matrix factorization[J]. Data Mining and Knowledge Discovery, 22(3):393-418.
- [25] Centeno V L, Panadero C F, Kloos C D. 1999. Personalizing your electronic newspaper[C]. Proceedings of the 4th Euromedia Conference:26-28.
- [26] Cheetham A H, Hazel J E. 1969. Binary (presence-absence) similarity coefficients[J]. Journal of Paleontology, 43 (5):1130-1136.
- [27] Chen Y, Fay S, Wang Q. 2003. Marketing implications of online consumer product reviews[D]. University of Florida.
- [28] Chen Y, Iyer G. 2002. Consumer addressability and customized pricing [J]. Marketing Science, 21 (2): 197-208.
- [29] Chevalier J, Mayzlin D. 2006. The effect of word of mouth on sales: online book reviews [J]. Journal of Marketing Research, 43 (3):345-354.
- [30] Claypool M, Gokhale A, Miranda T, et al. 1999. Combining content-based and collaborative filters in an online newspaper [C]. Proceedings of ACM SIGIR'99 Workshop Recommender Systems: Algorithms and Evaluation.