

# 电子商务个性化推荐技术分析 & 比较

## Analysis and Comparative Study of Personalized Recommendation Techniques in E - Commerce

马 丽 (西华师范大学 商学院 四川 南充 637002)

**摘 要:** 随着电子商务的迅猛发展,“信息超载”问题日渐突出,导致用户难以有效搜寻所需商品,个性化推荐技术由此在电子商务网站得到应用和普及。采用定性分析方法,对目前电子商务网站使用的信息检索、关联规则、基于内容的过滤和协同过滤等四种主要推荐技术作了应用层面的优缺点分析,并基于多个指标对其进行比较,以期为企业构建推荐系统提供决策参考。

**关键词:** 推荐技术 信息检索 关联规则 基于内容的过滤 协同过滤

### 1 引言

电子商务的迅猛发展不仅使企业节省了传统经营模式下必要的实体投资成本,而且消除了传统店面的商品陈列空间限制,为购物者提供了一个庞大的(也可说是无限大)商品柜台。但是由于顾客无法通过小小的计算机屏幕在短时间内浏览所有商品,且缺少促销人员的精心导购,从而使得用户面临“信息超载”(information overload)<sup>[1]</sup>。信息超载是指网站为用户提供的商品信息量过多,导致其难以迅速找到所需商品,从而容易使用户产生疲劳直至失去购物兴趣和离开。因此,电子商务网站面临一个严峻的问题:如何在用户浏览网站时将适合该用户的商品推荐到他/她面前,克服信息超载带来的不利影响,从而促成更多交易以增加企业销售额?

目前,信息检索(information retrieval)、关联规则(association rules)、基于内容的过滤(content-based filtering,CBF)和协同过滤(collaborative filtering,CF)已在电子商务推荐系统(E-commerce recommender systems)<sup>[2]</sup>中得到广泛使用。这四种技术都在一定程度上实现了面向用户的个性化推荐,但也都存在一些不足。本文通过对电子商务个性化推荐技术的分析和比较,以期为企业构建推荐系统提供决策参考。

### 2 电子商务推荐系统

电子商务推荐系统是解决信息超载问题的一种方案、一种帮助电子商务网站实现“一对一营销”(one-to-one marketing)战略的技术,同时也是电子商务网站进行自动化大规模定制的一条关键途径,被网站用作虚拟店员(virtual salespeople)向客户提供商品信息和建议,帮助用户决定应该购买何种商品<sup>[2-4]</sup>。电子商务推荐系统使得网站能适应每一个消费者并为其提供具有个性化的商品展现平台和购物体验。正如 Jeff Bezos (Amazon.com CEO) 所言:“如果我在网络上有三百万个用户,我就应有三百万个网上商店”<sup>[2]</sup>。电子商务推荐系统的作用表现在三个方面<sup>[2]</sup>:

(1) 将电子商务网站浏览者转变为购买者(converting browsers into buyers)。电子商务推荐系统能够帮助无购买意愿的浏览用户找到其感兴趣商品,从而将其转变为购买者。

(2) 提高电子商务网站交叉销售能力(increasing cross-sell)。交叉销售在现代商业中应用非常普遍。通过交叉销售,能够引导用户发现和购买自己确有潜在需求但在购买过程中未曾想到的商品。

(3) 建立客户忠诚度(building loyalty)。赢得客户忠诚度是一项基本的商业策略。在互联网上,用户只

需要点击一两下鼠标,便能从当前的电子商务网站转到其竞争对手网站那里。电子商务推荐系统通过建立网站与客户之间的增值关系 (value-added relationship) 来提高客户忠诚度,提高客户和网站之间的“粘性” (stickiness)。

时至今日,许多大型网站都已使用了电子商务推荐系统,例如全球最大的网上书店 Amazon、最大的网上拍卖站点 eBay、最大的网上音乐商店 CDNOW、主流门户网站 Yahoo、最大的搜索引擎 Google、最大的中文网上书店当当网等等。

### 3 推荐技术分析

#### 3.1 信息检索

信息检索在电子商务网站的应用点主要包括:

##### (1) 从用户角度看

① 商品分类浏览。商品分类浏览方法是基于主题分类的信息查找方法,其优点在于既方便网站进行商品信息管理,也符合人们认知事物的习惯。但是商品分类浏览存在以下缺点:①许多商品难以明确其类别。随着科学技术的不断发展,交叉学科越来越多,导致相当多的商品难以给出其明确分类,使得用户逐层浏览的结果是有可能找不到原以为应该属于该分类的商品。②缺乏统一的商品分类规则。电子商务网站没有类似图书分类法那样的统一商品分类规则,在分类名称上也存在较大差异。③商品分类浏览比较耗时。用户必须首先明确自身所需商品的分类,然后逐层查找。

② 关键字查询。关键词查询的优点是查全率较高,查询速度快,且查询不限于特定领域,用户使用方便。但是关键词查询的缺点也很明显:①查准率较差。用户提交的查询请求通常会由系统返回数以百计的结果,用户还需要进行二次乃至多次查找才能得到自己真正所需内容。②个性化程度低。任何用户只要给出相同的关键词,系统都会返回相同的查询结果。③不能实现新异发现 (serendipitous finds)。

##### (2) 从电子商务网站角度看

① 编辑推荐或专家推荐。例如定期推出由网站

编辑或专家撰写的关于某主题的商品目录及介绍。

##### ② 简单统计数据。例如近期的销售排行榜。

无论是用户使用的商品分类浏览、关键字查询,还是电子商务网站使用的编辑/专家推荐、统计数据,在面对日益增加的商品信息和用户迫切需要在最短时间找到最满意商品的矛盾时,都显得难以胜任。

#### 3.2 关联规则

关联规则<sup>[1,4]</sup>作为数据挖掘领域的重要技术,已在零售领域应用多年,通过用户交易数据产生关联规则,再结合用户当前购买行为作出推荐,主要用于实现交叉销售。最典型的关联规则应用是购物篮分析,即通过研究用户购物篮中商品之间的关系,发现同时被频繁购买的商品,从而帮助电子商务网站在用户下订单和付款时向其推荐相关商品。关联规则还有助于网站调整商品在页面上的陈列位置。e-VZpro<sup>[5]</sup>就是一个基于关联规则的推荐系统。

关联规则的缺点在于它是根据被购商品之间的关系来建立商品项之间的关联,因此个性化程度不高,且在数据集高维、稀疏的情况下会导致弱规则 (weak rules)<sup>[1]</sup>。Sarwar 等人<sup>[4]</sup>的实验证明关联规则的推荐效果不如协同过滤。此外,关联规则属于基于规则的推荐系统范畴,基于规则的系统的缺点是规则质量很难保证且不能动态更新,随着规则的数量增多系统将变得越来越难以管理<sup>[6]</sup>。

#### 3.3 基于内容的过滤

基于内容的过滤在技术方面与信息检索有许多相同点,不同之处在于其满足用户的长期信息需求。基于内容的过滤为每个用户建立用户描述 (user profile),记录用户所喜好/厌恶的内容,然后将其与项目 (例如商品、Web 页面、电影、音乐等) 的内容进行比较,把相似度较高的项目推荐给用户,由此提高了推荐的可测量性,并能对推荐结果作出较好解释。用户描述可以通过系统对用户喜欢的项目内容进行机器学习得到,或者通过用户的查询语句、问卷反馈等得到。基于内容的过滤系统有 LIBRA<sup>[7]</sup>、CiteSeer<sup>[8]</sup>、WebMate<sup>[9]</sup>等。但是基于内容的过滤存在两个主要缺点:

##### (1) 有限的内容分析 (limited content analysis)<sup>[10]</sup>。

由于基于内容的过滤要求对象结构性较好,因此主要

用于文本资源的过滤处理,而对图像、视频、音乐等结构复杂的、难以体现内容质量等特征的项目无法有效提取其信息进行推荐。即使对于文本资源,基于内容的过滤也只能反映其内容信息而无法判别文本资源的质量等特征。

(2) 不能提供新异推荐 (serendipitous recommendations)。这是由于基于内容的过滤推荐存在“过度专门化”(over-specialization) 的现象<sup>[10]</sup>,不利于挖掘用户潜在兴趣以推荐主题范围更广的商品。

### 3.4 协同过滤

为了克服基于内容的过滤所存在的缺点,Goldberg 等<sup>[11]</sup>提出了协同过滤。基于内容的过滤与协同过滤的工作机制比较如图 1 所示。

协同过滤根据与目标用户有相似兴趣偏好的其他用户对某信息的观点来判断该信息对目标用户是否有价值,进而决定是否将该信息推荐给目标用户,从而缓解信息超载。协同过滤的优势在于不需考虑被推荐项

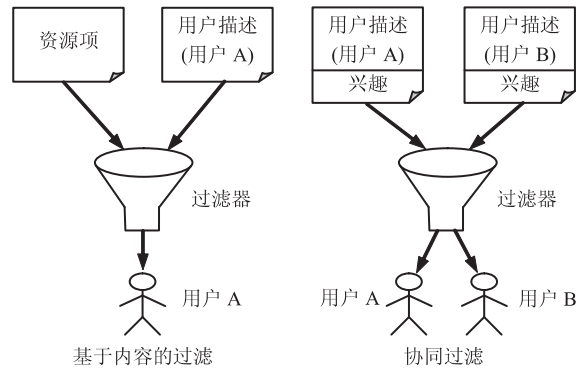


图 1 CBF 与 CF 的工作机制比较<sup>[6]</sup>

目的内容,因此不仅将过滤对象扩展到所有类型的资源,并且实现了新异推荐,从而成为目前电子商务推荐系统中广泛应用的、最成功的推荐算法。典型的协同过滤推荐系统有 GroupLens、MovieLens、WebSphere Personalization 等。协同过滤推荐流程见图 2:

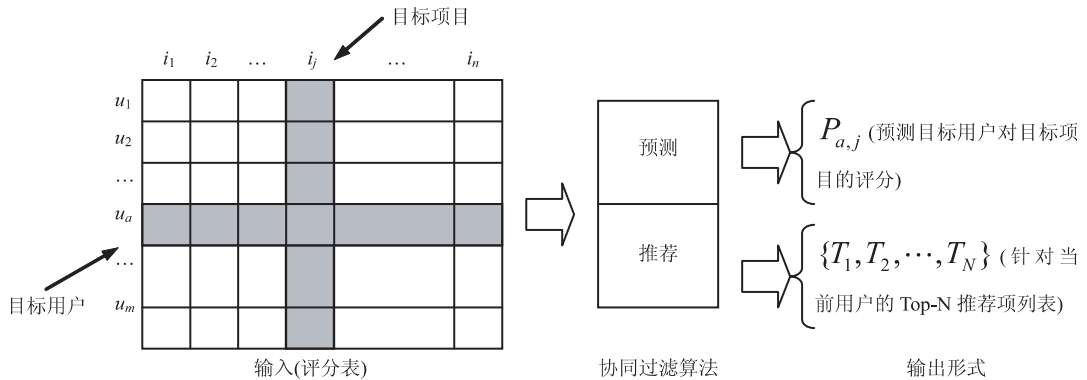


图 2 协同过滤推荐流程<sup>[12]</sup>

协同过滤在构建用户 - 项目评分矩阵基础上,使用统计技术寻找与目标用户有相同或相似兴趣偏好的邻居用户,再根据邻居用户对商品项的评分来预测目标用户对其未评分项的评分值,进而选择预测评分最高的前 N 项商品作为推荐集反馈给用户 (top - N 推荐),其基本思想是用户会对邻居用户所喜欢的商品产生兴趣。因此,用户评分数据收集越多,协同过滤算法的推荐质量越高。

但是,由于电子商务网站用户及商品项的数量庞大且不断增加,使得用户 - 项目评分矩阵成为高维矩

阵;同时用户给予评分的商品项很少,通常在 1% 以下<sup>[1]</sup>,导致评分矩阵数据极端稀疏。数据稀疏性 (sparsity) 问题<sup>[1]</sup>由此产生,并成为导致协同过滤推荐质量下降的主要原因,使得最近邻用户难以搜寻或搜寻准确度不高、在形成目标用户的最近邻用户集时丢失邻居用户的传递关系<sup>[4,13]</sup>等信息以及冷启动 (cold-start)<sup>[1]</sup>问题等。除了稀疏性问题外,协同过滤还存在其他一些不足:

(1) 可扩展性问题 (sparsity)<sup>[1]</sup>。即随着系统中用户和项目数量的增多,算法的计算复杂度急剧增加,导

致系统性能不断下降,直接影响推荐实时性。

(2) 同义性问题 (synonymy)<sup>[4]</sup>。在实际生活中,名称不同的商品可能是相似的。但是协同过滤算法不能发现这种潜在关联,因此对这些相似的商品进行分别对待。

(3) 特殊用户问题 (unusual user problem)<sup>[13]</sup>。即灰色绵羊问题 (gray sheep problem),指小部分用户的兴趣偏好相当特殊,与任何用户群体都不相同,因此很难得到准确的推荐。

#### 4 推荐技术比较

本文采用多个指标对信息检索、关联规则、基于内容的过滤和协同过滤四种推荐技术进行衡量,比较结果见表 1。

表 1 电子商务个性化推荐技术比较

	信息检索	关联规则	基于内容的过滤	协同过滤
自动化程度	低	高	高	高
持久性程度	低	低	高	高
个性化程度	低	低	高	高
推荐模式	被动推荐	主动推荐	主动推荐	主动推荐
用户是否参与	Yes	No	Yes	Yes
能否新异推荐	No	Yes	No	Yes
其他主要缺点	查准率较差	规则质量难以保证	有限的内容分析	数据稀疏性
典型系统	Google	e-VZpro	WebMate	WebSphere Personalization

下面对表 1 中前三个重要指标进行解释:

(1) 自动化程度 (degree of automation)<sup>[3]</sup>。自动化维度范围从完全自动推荐到完全手工推荐,自动化程度取决于客户为了得到推荐系统的推荐是否需要显式输入信息以及输入信息的多少等。

(2) 持久性程度 (degree of persistence)<sup>[3]</sup>。持久性程度维度范围从完全暂时性推荐到永久性推荐,暂时性推荐基于客户的单一会话 (session),永久性推荐

则基于客户先前的多个会话。

(3) 个性化程度 (degree of personalization)<sup>[14]</sup>。个性化程度用来反映推荐结果符合用户兴趣爱好的程度。

从表 1 可以看到,协同过滤和基于内容的过滤是最好的两种推荐技术,关联规则次之,最差的是信息检索。在缺乏足够的用户评分数据及商品项描述信息情况下,关联规则将是一个很好的辅助推荐途径。信息检索则应成为电子商务网站提供的最基本个性化服务技术。

#### 5 结束语

在电子商务环境下,企业面临的主要问题之一就是如何为用户提供个性化程度更高、更符合用户需求的商品和服务。这是电子商务企业价值链的源头和市场营销的起点。目前电子商务网站普遍使用了信息检索、关联规则、基于内容的过滤、协同过滤这四种主要推荐技术中的一种或多种。由于上述推荐技术各有其优势与不足,因此电子商务网站需要在构建其推荐系统时对各种推荐技术进行配合应用,以实现最佳的个性化推荐服务。

#### 参考文献

- 1 Sarwar BM. Sparsity, scalability, and distribution in recommender systems [ Ph. D. Thesis ]. Minneapolis: University of Minnesota, 2001.
- 2 Schafer JB, Konstan JA, Riedl J. E-commerce recommendation applications. Data Mining and Knowledge Discovery, 2001, 5(1-2): 115-153.
- 3 Schafer JB, Konstan JA, Riedl J. Recommender systems in e-commerce. In: Proceedings of the 1st ACM Conference on Electronic Commerce. New York: ACM Press, 1999: 158-166.
- 4 Sarwar B, Karypis G, Konstan J, Riedl J. Analysis of recommendation algorithms for E-commerce//Proceedings of the 2nd ACM Conference on Electronic Commerce. New York: ACM Press, 2000:158-167.
- 5 Demiriz A. Enhancing product recommender systems on sparse binary data. Data Mining and Knowledge Discovery, 2004, 9(2): 147-170.

(下转第 161 页)

(上接第 61 页)

- 6 曾春, 刑春晓, 周立柱. 个性化服务技术综述. 软件学报, 2002, 13(10): 1952 - 1961.
- 7 Mooney RJ, Roy L. Content - based book recommending using learning for text categorization//Proceedings of the 5th ACM Conference on Digital Libraries. New York: ACM Press, 2000: 195 - 204.
- 8 Bollacker KD, Lawrence S, Giles CL. Discovering relevant scientific literature on the web. IEEE Intelligence Systems, 2000, 15(2): 42 - 47.
- 9 Chen L, Sycara K. WebMate: a personal agent for browsing and searching//Proceedings of the 2nd International Conference on Autonomous Agents. New York: ACM Press, 1998. 132 - 139.
- 10 Adomavicius G, Tuzhilin A. Toward the next generation of recommender systems: a survey of the state - of - the - art and possible extensions. IEEE Transactions on Knowledge and Data Engineering, 2005, 17(6): 734 - 749.
- 11 Goldberg D, Nichols D, Oki BM, Terry D. Using collaborative filtering to weave an information Tapestry. Communication of the ACM, 1992, 35(12): 61 - 70.
- 12 Sarwar B, Karypis G, Konstan J, Riedl J. Item - based collaborative filtering recommendation algorithms//Proceedings of the 10th International Conference on World Wide Web. New York: ACM Press, 2001. 285 - 295.
- 13 Bridge D, Kelleher J. Experiments in sparsity reduction: using clustering in collaborative recommenders//Proceedings of the 13th Irish International Conference on Artificial Intelligence and Cognitive Science. London: Springer - Verlag, 2002: 144 - 149.
- 14 余力, 刘鲁. 电子商务个性化推荐研究. 计算机集成制造系统, 2004, 10(10): 1306 - 1313.