

协同过滤推荐技术综述^{*}

冷亚军¹ 陆 青¹ 梁昌勇^{2,3}

¹(上海电力学院 经济与管理学院 上海 201300)

²(合肥工业大学 管理学院 合肥 230009)

³(合肥工业大学 过程优化与智能决策教育部重点实验室 合肥 230009)

摘 要 协同过滤是推荐系统中广泛使用的推荐技术,研究人员对如何完善协同过滤推荐技术开展大量工作,但是相应的研究总结较少.文中对协同过滤的相关研究进行全面回顾,首先阐述协同过滤的内涵及其存在的主要问题,包括稀疏性、多内容及可扩展性,然后详细介绍国内外学者针对以上问题的解决方案,最后指出协同过滤下一步的研究重点.文中介绍一个相对完整的协同过滤知识框架,对理清协同过滤的研究脉络,为后续研究提供参考,推进个性化信息服务的发展具有一定意义.

关键词 个性化服务,推荐系统,协同过滤,信息超载

中图法分类号 TP 311

Survey of Recommendation Based on Collaborative Filtering

LENG Ya-Jun¹, LU Qing¹, LIANG Chang-Yong^{2,3}

¹(College of Economics and Management, Shanghai University of Electric Power, Shanghai 201300)

²(School of Management, Hefei University of Technology, Hefei 230009)

³(Key Laboratory of Process Optimization and Intelligent Decision-Making, Ministry of Education, Hefei University of Technology, Hefei 230009)

ABSTRACT

Collaborative filtering is a widely used technique in recommender systems. Extensive studies are carried out on collaborative filtering. However, systematic summary of this field is scarce. In this paper, research of collaborative filtering is reviewed. The meaning and key issues of collaborative filtering, including sparsity, multiple-content and scalability, are described firstly, and then the solutions to the above key issues are introduced in detail. Finally, the future work of collaborative filtering is pointed out. The knowledge framework of collaborative filtering is introduced. It makes the research clues of collaborative filtering clear, provides a reference to other scholars, and improves the performance of personalized information services.

^{*} 国家自然科学基金项目(No. 713311002, 71271072)、高等学校博士学科点专项科研基金项目(No. 20110111110006)、上海高校选拔培养优秀青年教师科研专项基金项目(No. sdl10021) 资助

收稿日期: 2013-11-06; 修回日期: 2014-04-29

作者简介 冷亚军(通讯作者), 男, 1985 年生, 博士, 讲师, 主要研究方向为电子商务、数据挖掘. E-mail: huayi2001@163.com. 陆青, 男, 1982 年生, 博士, 讲师, 主要研究方向为进化计算、数据挖掘. 梁昌勇, 男, 1965 年生, 教授, 博士生导师, 主要研究方向为智能决策支持系统、企业信息化.

Key Words Personalized Service , Recommender System , Collaborative Filtering , Information Overload

1 引言

随着移动互联网、物联网、云计算等技术的快速发展,全球数据量呈爆炸式增长,大数据时代已经到来。全球数据量级从 TB 发展至 PB 乃至 ZB,可称为海量、巨量甚至超量。相对于以往便于存储的以文本为主的结构化数据,音频、视频、图片、地理位置信息等非结构化数据的比例逐步提升,达到 80% 左右^[1]。互联网上信息的急剧增长,一方面使人们获取的信息资源越来越丰富,给人们带来极大的便利;另一方面,面对海量的信息资源,人们不得不花费更多的时间和精力去搜寻对其有帮助的信息,因此“信息超载(Information Overload)”^[2-3]现象越来越严重。

推荐系统(Recommender System)^[4-7]是解决信息超载问题的有效方案,它根据用户特征推荐满足用户需求的对象,实现个性化服务。推荐系统的优点在于,它能主动收集用户的特征资料,通过对用户个性、习惯、偏好的分析,为用户定制并提供其感兴趣的信息;同时能及时跟踪用户的需求变化,根据变化自动调整信息服务的方式和内容。与搜索引擎提供的“一对多”式的信息服务不同,推荐系统输出的结果更符合用户的个性化需求,实现“一对一”式的信息服务,同时用户的参与程度也更低,从而大为降低用户搜寻信息的成本^[8]。推荐系统作为一种新近的智能信息服务方式,在电子商务、社会网络、数字化图书馆、视频/音乐点播等领域得到广泛应用。特别是在电子商务领域,其作用尤为突出^[6]: 1) 将电子商务网站浏览者转变为购买者(Converting Browsers into Buyers); 2) 提高电子商务网站的交叉销售能力(Increasing Cross-Sell); 3) 建立客户忠诚度(Building Loyalty)。

目前推荐系统所采用的推荐技术主要包括关联规则(Association Rules)^[9]、基于内容的推荐(Content-Based Recommendation)^[10]、协同过滤(Collaborative Filtering)^[3]和混合推荐(Hybrid Approach)^[11]。协同过滤根据其他用户的偏好向目标用户推荐,它首先找出一组与目标用户偏好一致的邻居用户,然后分析该邻居用户,把邻居用户喜欢的项目推荐给目标用户^[12-15]。协同过滤的优势在于: 1) 不需考虑被推荐项目的内容^[16-18]; 2) 可为用户提供新异推荐(Serendipitous Recommendation)

tions)^[16-17]; 3) 对用户访问网站时的干扰较小; 4) 技术易于实现^[19]。因此它成为一种较流行的推荐技术。自 1992 年美国施乐公司 PARC 研究中心(Xerox Palo Alto Research Center)正式提出协同过滤以来,国内外众多研究人员对如何完善协同过滤推荐技术开展大量的工作,但少有相应的研究总结,尤其是国内学者工作更是较少。本文系统归纳协同过滤领域的研究工作,全面回顾协同过滤的内涵、发展、存在问题、研究现状及未来研究热点,为后续研究提供参考,促进我国个性化信息服务更好、更快地发展。

2 协同过滤及其存在的问题

推荐系统的根源最早可追溯到认知科学、近似理论、信息检索、管理科学等领域的研究^[7],自 20 世纪 90 年代初期第一个推荐系统 Tapestry^[20]提出以来,推荐系统开始成为一个独立的研究领域,并一直保持着较高的研究热度。众多组织和学者对推荐系统展开广泛、深入的研究,并从不同角度定义推荐系统。本文综合众多文献的描述,将推荐系统定义为:推荐系统采用统计分析和机器学习技术识别用户的兴趣偏好,向用户提供满足其需求的信息、商品和服务的推荐,从而减少用户的搜寻成本,增加网站运营商的收益。

对于推荐系统的任务,李聪^[3]给出如下形式化的描述。

Given:

对于网站 w , 令其用户集合

$$U = \{u_1, u_2, \dots, u_m\},$$

项目集合

$$I = \{i_1, i_2, \dots, i_n\}.$$

存在一个兴趣度函数 $\delta(u, i)$ 用以计算用户 u 对项目 i 的兴趣度,即 $\delta: U \times I \rightarrow \mathbf{R}^w$, \mathbf{R}^w 是一定范围内的非负实数。任取一用户 $u_a \in U$ 。

Goal:

- 1) 对任一项目 $i_j \in I$, 计算 $\delta(u_a, i_j)$;
- 2) 求得项目集合

$$I(u_a) = \arg \max_{i_j \in I} \delta(u_a, i_j).$$

目前各 Web 站点用以满足用户个性化需求的推荐技术主要包括关联规则、基于内容的推荐、协同过滤和混合推荐。表 1 列出各种推荐技术的主要优

缺点.

表 1 个性化推荐技术对比

Table 1 Comparison of personalized recommendation techniques

推荐技术	主要优点	主要缺点
关联规则	自动化程度高 新异推荐	个性化程度低 规则难以提取 规则质量难以保证
基于内容的推荐	个性化程度高 自动化程度高 结果直观,易解释	有限的内容分析 不能实现新异推荐
混合推荐	个性化程度高 自动化程度高 推荐准确性高 新异推荐	有限的内容分析 技术实现难度高 执行效率低于单一 推荐技术
协同过滤	个性化程度高 推荐领域广 新异推荐 技术易于实现	稀疏性问题 多内容问题 可扩展性问题

“协同过滤”一词最早由 Goldberg 等在文献 [20] 中提出,文献 [20] 详细介绍第一个协同过滤推荐系统 Tapestry. Tapestry 是一个电子文档过滤系统,它要求用户评价已阅读过的文档,其他用户根据这些评价选择适合自己的文档. 用户需输入一些查询语句(语句内容包括文档发送人、发送日期、文档主题等),系统处理查询语句后给出过滤结果. Tapestry 的缺点是不能自动完成推荐过程,需用户手动输入查询语句;而且用户对越多的成员有所了解,系统给出的过滤结果越精确,这并不适合在 Internet 这种用户群体庞大且相互陌生的环境中应用.

Resnick 等^[21] 提出第一个自动化协同过滤系统 GroupLens. GroupLens 的作用是帮助用户在大量的网络新闻中找到自己喜欢的新闻,其基本思想是如果用户过去对新闻的偏好是一致的,将来他们的偏好仍将一致. GroupLens 采用评分的方式搜集用户偏好,用户对已浏览过的文章评分(评分值为 1~5 的整数,分数越高表明用户越喜欢某文章),系统计算用户间的评分相似性,选出相似性较高的一组用户预测当前用户对新文章的偏好程度. 后来, Konstan 等^[22]、Miller 等^[23] 丰富和扩展 GroupLens 的功能,使得 GroupLens 成为一个基于开放式架构的分布式系统. GroupLens 的体系结构如图 1 所示. GroupLens 的提出为协同过滤技术的发展奠定坚实的基础,此后对于协同过滤的大量研究大都遵循文献 [21] 中算法的思想. GroupLens 研究小组还成功地将协同过滤应用于 Internet,由此推动协同过滤在各种网站中的广泛应用.

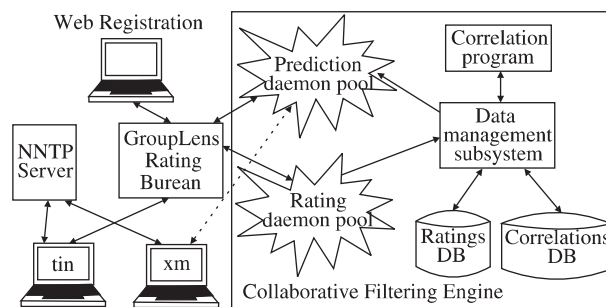


图 1 GroupLens 的体系结构^[3]

Fig. 1 Architecture of GroupLens^[3]

自 GroupLens 提出后,协同过滤取得较大发展.越来越多的研究人员对协同过滤展开研究,国际顶级学术会议和期刊发表的协同过滤的论文逐年增加.许多 Web 站点使用协同过滤推荐系统向用户提供个性化信息服务,如网上商城 Amazon、电影网站 Netflix、网上音乐社区 Last.fm 等,其中 Amazon 公司对协同过滤推荐系统的研究和应用已达十余年.

协同过滤通常可分为两类^[24]: 基于记忆(Memory-Based)的协同过滤和基于模型(Model-Based)的协同过滤. 基于记忆的协同过滤利用整个用户-项目评分数据集进行计算,每个用户都是评分预测过程的组成部分. 基于记忆的协同过滤为目标用户选择一部分兴趣相近的邻居用户,根据邻居用户的评分预测目标用户对项目的评分值. 典型的基于记忆的协同过滤有最近邻(Neighbor-Based)协同过滤及其改进算法. 基于模型的协同过滤根据训练集数据学习得出一个复杂的模型,然后基于该模型和目标用户已评分数据,推导出目标用户对未评分项目的评分值. 典型的基于模型的协同过滤有基于聚类技术的协同过滤、基于概率方法的协同过滤、基于矩阵分解的协同过滤等.

协同过滤技术具有其他推荐技术无法比拟的优势,具体如下.

1) 推荐对象可为任何类型的资源. 由于协同过滤基于相似用户的兴趣偏好产生推荐,因此它的关键是找出与目标用户偏好相似的邻居用户,而不必分析、提取项目内容信息. 这使得协同过滤技术可将任何类型的资源项推荐给用户.

2) 产生新异推荐. 协同过滤不对比项目内容与用户描述间的相关性,所推荐项目不完全局限于用户的历史偏好范围内,这将有助于发现用户的潜在兴趣,实现跳跃式推荐.

3) 对用户干扰较小. 协同过滤一般使用评分向

量作为用户偏好的表示方式, 用户登录系统后只需对一些项目评分, 便可获得推荐服务, 用户使用系统的整个过程中受到的干扰较小。

4) 技术易于实现. 协同过滤的用户偏好信息搜集容易, 算法程序并不复杂且可扩展性较好, 因此整个协同过滤系统的实现较为容易。

尽管协同过滤在个性化推荐方面取得较大成功, 但其本身存在的一些关键问题制约着其进一步发展, 本文将这些问题概括如下。

1) 稀疏性(Sparsity) ^[25-28]. 实际的网站中用户和项目数量庞大, 而用户通常只对一小部分项目评分, 可用于计算用户/项目间相似性的数据非常有限, 使得最近邻搜寻不够准确, 推荐质量较差. 经常可看到的现象是两个用户/项目间没有任何共同评分项, 导致相似性无法计算. 即使有的用户/项目间相似性可计算, 可靠性也难以保证。

2) 多内容(Multiple-Content) ^[29]. 传统的协同过滤没有考虑项目类别的影响, 当网站中项目类别的内容完全不同时, 传统的协同过滤算法搜寻出的最近邻往往与目标用户仅在个别项类上偏好相似, 导致推荐结果不够合理。

3) 可扩展性(Scalability) ^[3, 25, 30-31]. 网站中用户和项目的数量庞大(如 Amazon 商城拥有数以百万计的商品), 且不断增加, 这使得用户-项目评分矩阵成为高维矩阵, 由此产生协同过滤的可扩展性问题, 即随着用户和项目数量的增多, 算法的计算复杂度急剧增加, 严重影响系统推荐的实时性。

3 协同过滤研究现状

国内外研究人员针对协同过滤中存在的稀疏性、可扩展性、多内容等问题进行广泛而深入的研究, 本文将这些研究归纳如下。

3.1 稀疏性问题研究现状

实际网站中项目的数量庞大且不断增加, 而用户通常只对一小部分项目评分(一般不超过项目总数的 1% ^[25]) 感兴趣, 导致用户的评分数据非常稀疏. 数据稀疏性从以下两方面对协同过滤产生不利影响: 1) 使得最近邻搜寻不够准确, 从而影响协同过滤的推荐质量; 2) 搜寻到的最近邻用户评分过少, 影响推荐准确性并导致推荐覆盖率降低. 研究人员提出大量解决稀疏性问题的方法, 本文将这些方法归纳为 6 类: 空值填补、新相似性方法、结合基于内容的推荐、推荐结果融合、图论和其他方法。

3.1.1 空值填补

采用有效的预测方法填补评分矩阵中的未评分项, 可提高矩阵中的评分密度, 降低数据稀疏性对协同过滤推荐的影响。

最简单的做法是将未评分项设定为一个缺省值, 这个值可以是评分中值、众数、用户评分均值、项目评分均值等^[32]. 由于未评分项获得的评分值不可能完全相同, 因此缺省值方法的可信度不高. 邓爱林等^[33]提出基于项目评分预测的协同过滤算法(Item Rating Prediction, IRP), 该算法采用基于项目的协同过滤方法填补用户评分项并集中的评分空值, 在填补后的并集上计算用户间的相似性. 李聪等^[34]认为基于项目的协同过滤方法本身存在的稀疏性问题影响填补效果, 他们改进文献[33]的方法, 采用领域最近邻方法预测评分项并集中的未评分值. 文献[35]根据用户评分向量交集大小选择候选邻居集合, 采用 BP 神经网络预测用户对项目的评分, 减小候选邻居评分数据的稀疏性。

Zhang 等^[36]指出, 基于用户的协同过滤方法搜寻出的最近邻绝大多数对目标项目没有评分, 从而严重影响评分预测的准确性. 为此他们设定一个迭代次数阈值, 循环执行基于用户的协同过滤方法以填补最近邻的未评分项, 在得到所有最近邻对目标项目的评分后, 采用传统的协同过滤方法生成最终的预测评分. 文献[37]提出一种基于近邻评分填补的混合协同过滤推荐算法, 该算法对原始评分矩阵全局降维, 在低维的主成分空间中计算用户相似性, 减少算法复杂度. 采用奇异值分解(Singular Value Decomposition, SVD) 填补近邻评分缺失值, 降低近邻评分的稀疏性. Jeong 等^[12]将基于用户和基于项目的协同过滤方法互相融合, 使用该混合方法循环填补用户-项目评分矩阵中的未评分元素, 直到评分矩阵稳定为止. 文献[38]采用均值填补、线性回归预测、预测均值匹配、贝叶斯分类预测等多种方法填补原始评分矩阵, 并对比各填补方法的准确性。

3.1.2 新相似性方法

传统的相似性度量方法对稀疏数据的处理能力较差, 研究人员提出一些新的相似性度量方法来解决这一问题。

周军锋等^[39]使用一种修正的条件概率方法计算项目间的相似性, 该方法体现共同评分数、评分值差异和项目所属类别对项目相似性计算的共同影响. 张光卫等^[40]设计一种基于云模型的相似度计算方法(Likeness Comparing Method Based on Cloud Model, LICM). LICM 根据用户的评分情况计算用

户的评分频度向量,利用逆向云算法将评分频度向量转换为评分特征向量,评分特征向量反映用户对项目的平均满意度、评分集中度和评分不确定度。LICM 通过计算评分特征向量间的余弦夹角得到用户的相似性。Luo 等^[41]提出局部用户相似性和全局用户相似性的概念,分别计算用户间的局部相似性和全局相似性,根据相似性大小选出局部最近邻和全局最近邻,然后计算两种最近邻的预测评分,并混合评分,使用一个静态的权重控制参数,平衡两种预测评分的贡献度。Anand 等^[42]拓展 Luo 的工作,他们考虑评分数据的整体稀疏度和局部稀疏度,以此估算权重控制参数,使参数的确定动态化。Choi 等^[43]计算用户相似性时考虑所有项目与目标项目的相似程度,与目标项目越相似,它在最近邻搜寻过程中所起的作用就越大。Kaleli^[44]综合评分相似性、评分不确定度差异两种信息,搜寻目标用户的最近邻。首先根据用户评分向量信息熵计算每个用户的评分不确定度;然后计算所有用户与目标用户的评分相似性、评分不确定度差异;最后将最近邻的选择过程转化为 0-1 背包问题,即满足所有近邻用户与目标用户不确定度差异之和小于常数 C ,且这些用户与目标用户的相似性之和达到最大。文献[45]采用近邻传播算法更新传统方法计算得到的用户/项目相似性,算法收敛后可得到最优的相似性值。文献[46]提出一种两阶段最近邻选择算法(Two-Phase Neighbor Selection, TPNS)。TPNS 首先计算用户间的近邻倾向性,选择近邻倾向性较高的用户组成初始近邻集合;然后根据初始近邻集合计算目标用户与其他用户间的等价关系相似性,使用等价关系相似性修正目标用户的初始近邻集合,得到最近邻集合。

3.1.3 结合基于内容的推荐

基于内容的推荐采用内容描述信息表示用户偏好,在协同过滤中适当结合基于内容的推荐可丰富用户的偏好信息,降低协同过滤对用户评分的过度依赖。这种方法实际上是一种混合推荐方法(如第2节所述)。

高滢等^[47]提出结合似然关系模型和用户等级的协同过滤算法(Collaborative Filtering Combining Probabilistic Relational Models and User Grade, PRMUG),该算法根据用户的评分数量确定用户等级,分别采用基于用户等级的协同过滤和基于似然关系模型的推荐方法预测目标项目的评分,将用户等级作为权重参数加权两种预测结果。Melville 等^[48]使用基于内容的方法填补用户评分向量中的空值,缓解评分数据的稀疏性,从而提高协同过滤的

推荐质量。Lee 等^[49]将时间信息融入到协同过滤中,将购买时间和项目推出时间分为若干阶段,每一阶段赋予相应的权重值,最后基于这些时间权重完成整个推荐过程。Liu 等^[50]同样也考虑时间信息对推荐质量的影响,提出一种基于序列规则和协同过滤相混合的推荐方法。在序列规则推荐中,将用户的交易数据聚类,根据不同时期的交易聚类挖掘用户购买商品的序列规则,将目标用户前 $T-1$ 时期的交易数据与序列规则匹配产生推荐结果。在协同过滤推荐中,以所有用户 T 时期(当前时期)的交易数据为基础计算目标用户的最近邻,根据最近邻的意见完成推荐。最后线性组合两种推荐结果,实现对目标用户的最终推荐。文献[51]考查社会化标签对推荐结果的影响,根据用户和项目的标签特征向量构建用户项目偏好矩阵、项目相似性矩阵,由用户历史行为及项目相似性计算用户对未使用项目的偏好程度。

3.1.4 推荐结果融合

在用户-项目评分矩阵上,采用不同方法计算用户对项目的偏好,将各种计算结果融合,可提高结果的可靠性。

文献[52]提出一种基于影响集的协同过滤算法,同时筛选出目标项目的最近邻和逆最近邻,采用最近邻和逆最近邻产生预测结果,并设计 4 种方法融合预测结果。文献[53]使用 Bregman 联合聚类方法聚类评分矩阵,确定出目标用户和目标项目所在聚类,融合类内评分均值、目标用户评分均值、类内所有用户评分均值、目标项目评分均值和类内所有项目评分均值,以此得到目标用户对目标项目的预测评分。Lee 等^[54]分别使用基于用户和基于项目的协同过滤产生两种预测结果,以共同评分数作为权重参数混合预测结果。Wang 等^[55]提出一种基于相似性融合的协同过滤算法(Similarity Fusion, SF),结合邻居用户对目标项目的评分、目标用户对邻居项目的评分、相似用户对相似项目的评分 3 种数据,并通过评分标准化及概率方法完成评分预测。Liang 等^[56]使用信息论联合聚类对用户和项目同时聚类,计算用户/项目间的类别偏好相似性;然后混合用户/项目间的类别偏好相似性和评分相似性,根据混合相似性分别采用基于用户和基于项目的协同过滤方法预测未知评分;最后线性加权两种预测结果,获得最终预测评分。

3.1.5 图论

Aggarwal 等^[57]提出基于 Horting 图的协同过滤推荐算法。Horting 图中的节点表示用户,节点之间

的边表示用户之间的相似度,在图中搜索近邻节点,综合各近邻节点的评分形成最后的推荐. Papagelis 等^[26]使用社会网络模型缓解评分数据的稀疏性,社会网络由用户空间、项目空间、用户对项目的评分活动及用户间的相似关系构成,经过网络上一系列的信任推导(Trust Inference),可在用户之间建立新的关联,从而使无共同评分项的用户也能互相推荐项目. 文献[58]总结基于社会网络的协同过滤研究,将社会网络协同过滤划分为基于矩阵分解和基于近邻推荐两种类型,并对比两种类型包含的代表性算法,得到的结论如下:基于矩阵分解的社会网络协同过滤具有较高的推荐质量,而基于近邻推荐的方法更易于被实现和执行. Zhou 等^[59]将对分网络(Bipartite Network)引入到个性化推荐中,提出一种基于网络推理的推荐方法(Network-Based Inference, NBI). NBI 将用户和项目映射为对分网络中的两组节点,节点间的边表示用户对项目有所关注. NBI 为目标用户关注过的每个项目分配一个单位的资源,这些资源通过边传递给用户,再从用户重新传回给项目,最后获得资源最多的前 N 个项目被推荐给目标用户. Zeng 等^[60]认为积极评分和消极评分对资源传递的影响是不同的,他们在 Zhou 研究的基础上,将用户项目间的边区分为积极边和消极边,并对每种边赋予相应的权重,资源在边上传递时要考虑所在边的权重值.

3.1.6 其他方法

除以上五类方法外,研究人员还提出一些其他用于解决稀疏性问题的方法. 如潘红艳等^[61]使用分类算法划分项目,然后把对同一类项目感兴趣的用戶聚集到一起,这样评分矩阵便被划分为若干较密集的矩阵块,提高局部评分密度. Goldberg 等^[62]从 Jester 数据集中抽取一部分项目作为给定集(Gauge Set),用户使用系统时必须对给定集中的全部项目评分. Goldberg 以增加用户-系统交互的方式来降低评分数据的稀疏性. 文献[63]改进传统的评分预测公式,将预测公式拆分成评分均值和邻居用户评分偏差两部分. 对于评分均值,该文通过线性组合项目评分均值、目标用户评分均值、邻居用户评分均值得到. Yang 等^[64]使用矩阵分解技术将评分矩阵 R 分解为用户特征矩阵 U 和项目特征矩阵 V ,将信任关系矩阵 T 分解为信任特征矩阵 B 和被信任特征矩阵 W . 分别提出信任模型 $B^T V$,被信任模型 $W^T V$ 和混合模型

$$g\left(\left(\frac{B^r + W^e}{2}\right)^T \left(\frac{V^r + V^e}{2}\right)\right) R_{\max}.$$

并用实验证明 3 种模型优于现有的协同过滤算法,且所提出的混合模型具有最高的推荐准确性. Koren 等^[65]同样使用矩阵分解技术将评分矩阵分解为用户特征矩阵和项目特征矩阵,并根据评分数据均值、用户评分偏差和项目评分偏差建立一个基本预测模型,将矩阵分解预测模型与基本预测模型混合,使用混合模型预测评分空值. 此外还将隐式评价信息、时间信息分别融入到混合模型中,进一步提高算法的准确性. 随后, Koren 等^[66]又把以上混合模型分别与基于用户的协同过滤、基于项目的协同过滤融合,在不损害最近邻协同过滤准确性的基础上,提高其运行效率.

3.1.7 稀疏性问题研究现状小结

空值填补对数据稀疏性的改善直观、显著,它利用用户已有评分数据填充全部或部分评分缺失值,不会给用户使用系统带来额外负担. 但空值填补本身是对评分缺失值的一种预测,预测值并不完全代表用户真实偏好,预测偏差将会影响后续计算的准确性.

新相似性方法充分利用用户评分数据包含的各种有用信息(如评分值、评分数量、评分信息熵等),与传统仅考虑评分值差异的相似性方法相比,具有更高的稳定性. 新相似性方法的不足在于它仍然基于用户已有的有限评分数据进行计算,没有从本质上改变评分数据的数量,对系统推荐质量的改善程度有限.

基于内容的推荐方法建立用户描述文件和项目内容文件,将与用户描述匹配度最高的项目内容推荐给用户. 结合基于内容的推荐可丰富用户偏好信息或推荐结果,但网站中多数项目为非结构化的,难以进行内容分析,因此限制该方法的应用.

推荐结果融合是将不同方法计算出的用户对项目偏好进行综合,实验结果显示,该方法比单一推荐结果的计算方法具有更高的推荐质量. 但该方法的计算结果可解释性差,另外该方法的准确性受融合机制的影响较大.

图论方法用图的方式表达用户与项目间的关系,由于图论方法表达直观、形象,通常可挖掘出新的关系. 图论方法的缺点与新相似性方法类似,并没有改变偏好数据的数量,没有从本质上解决稀疏性问题.

本节从稀疏性改善程度、推荐质量、算法实现难度这 3 个方面对比上述稀疏性问题解决方案,详细内容见表 2.

表2 稀疏性问题的解决方案对比

Table 2 Comparison of different solutions to sparsity problem

	空值 填补	新相似性 方法	结合基于 内容的推荐	推荐结果 融合	图论
稀疏性 改善程度	高	低	高	低	低
算法 推荐质量	低	中	高	中	中
算法 实现难度	中	低	高	低	高
典型算法	IRP ^[33]	TPNS ^[46]	PRMUG ^[47]	SF ^[55]	NBI ^[59]

3.2 可扩展性问题研究现状

协同过滤算法最耗时的环节是相似性计算,计算目标用户与所有用户的相似性,复杂度为 $O(mn)$,其中 m 表示网站中用户数量, n 表示项目数量.实际网站中,用户和项目的数量通常可达到上百万,使得协同过滤算法的可扩展性受到严峻挑战.研究人员针对这一问题展开广泛研究,并取得不少成果,本文将改善协同过滤可扩展性问题的方法分为数据集缩减、聚类、矩阵分解、主成分分析和增量更新5类.

3.2.1 数据集缩减

采用适当的方法缩减评分数据集,可降低数据集的规模,提高算法的运行速度.最简单的做法是删除评分低于一定数目的用户和项目,在剩余的评分矩阵上计算用户的相似性^[67].这样做的缺点是会导致部分用户得不到推荐及部分项目无法被推荐给有需求的用户.Yu等^[68]指出可从整个用户评分数据库中选出一个小子集,即描述空间(Profile Space),在描述空间上搜寻目标用户的最近邻,以提高算法的推荐速度.Russell等^[69]提出基于离散小波变换的协同过滤算法(Discrete Wavelet Transformation, DWT),该算法使用小波数据压缩项目空间,使项目评分向量的数目成倍减小,在缩小后的项目空间中使用传统的协同过滤算法产生推荐.Acilar等^[70]提出一种基于人工免疫网络(Artificial Immune Network)的协同过滤模型,使用人工免疫网络算法压缩用户评分矩阵,压缩后矩阵中的用户数目明显减少,评分数据稀疏度大为降低.随后采用2种方式计算推荐结果:1)直接在压缩矩阵上产生推荐;2)聚类压缩矩阵中的用户,根据聚类产生推荐.

3.2.2 聚类

Sarwar等^[71]使用K-means聚类算法聚类用户-项目评分矩阵,选择目标用户所在聚类的用户作为最近邻集合,通过减小最近邻搜寻空间的方式提高协同过滤的可扩展性.李涛等^[72]同样使用K-

means聚类算法聚类用户,生成若干聚类中心;然后计算每一用户与聚类中心的相似性,生成类别所属程度矩阵;最后在类别所属程度矩阵上搜寻目标用户的最近邻.郁雪等^[73]认为,由于实际网站中项目的数量巨大,直接聚类对用户-项目评分矩阵会非常耗时.使用主成分分析降维项目空间,在维数减少的主成分空间上使用自组织神经网络聚类算法聚类用户.文献[74]采用改进的近邻传播聚类算法聚类用户,根据灰关系等级和Jaccard系数定义用户相似度矩阵,缓解评分数据稀疏性,提高聚类质量.对于产生的 n 个初始聚类,保留Silhouette值较大的前 k 个聚类,将剩余聚类中的用户按所属类的Silhouette值分配到相应的聚类中,产生指定类数的聚类,加快聚类速度.

OConner等^[75]提出基于项目聚类的协同过滤算法,首先聚类项目,然后在相应的聚类中搜索目标用户的最近邻.邓爱林等^[76]采用K-means聚类算法聚类项目,生成相应的聚类中心,在此基础上计算目标项目与聚类中心的相似性,从而只需在与目标项目最相似的若干个聚类中就能寻找到目标项目的大部分最近邻居并产生推荐列表.文献[77]使用内容信息扩充项目评分向量,基于扩充的评分向量聚类项目,然后采用一种概率模型预测用户在项目上的评分值.

George等^[53]使用Bregman联合聚类方法同时聚类评分矩阵中的用户和项目,确定出目标用户和目标项目所在聚类,融合类内评分均值、目标用户评分均值、类内所有用户评分均值、目标项目评分均值和类内所有项目评分均值,以此得到目标用户对目标项目的预测评分.Wu等^[78]使用内容信息构建用户-属性矩阵、项目-属性矩阵,在以上两个矩阵和用户-项目评分矩阵基础上,使用信息论联合聚类同时聚类用户和项目,把用户、项目划分到相应的类别中.随后根据George方法的思想,利用用户、项目、用户所属类、项目所属类、用户-项目所属联合聚类的评分均值计算目标用户对目标项目的预测评分.吴湖等^[79]提出一种两阶段联合聚类的协同过滤算法,使用BlockClust联合聚类算法划分用户-项目评分矩阵,在划分后的子评分矩阵上通过加权非负矩阵分解(Weighted Non-negative Matrix Factorization)预测未知评分.由于George、Wu和吴湖的方法直接采用类内评分信息产生预测结果,不需要搜寻目标用户的最近邻,因此具有更快的推荐速度.

3.2.3 矩阵分解

通过矩阵分解可将用户评分矩阵分解为几个维度较小的子矩阵, 有利于提高协同过滤算法的实时性. 文献[80]使用奇异值分解(SVD)将 $m \times n$ 阶用户评分矩阵 R 分解为 3 个低阶矩阵:

$$R \approx U_k \times S_k \times V_k^T, k < \min\{m, n\},$$

然后基于低阶矩阵 $U_k S_k^{1/2}$ 进行协同过滤推荐. Voza-lis 等^[81]认为单纯使用用户的评分信息, 难以保证协同过滤算法的推荐质量, 他们提出一种 SVD 结合人口统计学数据的协同过滤算法, 使用人口统计学数据提高 SVD 推荐算法的推荐精度. Chen 等^[82]使用正交非负矩阵分解将评分矩阵分解为 3 个子矩阵. 矩阵分解过程就是对用户和项目同时聚类的过程, 然后将最相似的若干聚类作为目标用户/项目的候选邻居集合, 缩小最近邻的搜寻范围. Luo 等^[83]指出当前绝大多数基于矩阵分解的协同过滤算法都不具备增量更新能力, 即无法利用数据库中最新产生的评分数据进行推荐, 他们提出一种基于正则化矩阵分解的增量更新推荐算法, 设计 2 种更新机制: 增量正则化矩阵分解方法、考虑线性偏差的增量正则化矩阵分解方法, 以完成对评分数据的及时更新.

3.2.4 主成分分析

主成分分析(PCA)^[84]是一种有效的维数约简技术, 它通过线性变换将多个有相关性的变量转化为较少数个互不相关的新变量, 新变量保留原始数据的主要信息, 消除原始数据的信息重叠. Goldberg 等^[62]将 PCA 应用于 Jester 推荐系统中, 利用 PCA 对项目给定集进行全局降维, 在转换后的主成分空间聚类用户, 产生推荐结果. Kim 等^[85]认为, Goldberg 的方法要求用户必须对给定集中的所有项目评分, 这在实际应用中是不可行的, 因此他们提出不需要设置给定集的迭代主成分分析方法(Iterative PCA, IPCA), 采用 IPCA 填补用户未评分项并降维整个用户评分矩阵. Lee 等^[86]对购物篮数据(二元评分矩阵)中的项目维进行主成分分析, 以主成分向量作为预测变量建立二元 logistic 回归模型, 预测用户对项目的购买概率. 郁雪等^[73]提出一种基于 PCA 和自组织神经网络(Self-Organization Mapping Net, SOM)聚类的混合协同过滤模型, 使用基于项目的协同过滤填补电子政务数据库中的评分空值, 对填补后数据的项目维进行 PCA, 然后在主成分空间上使用 SOM 聚类用户, 把相似性最大的聚类作为最近邻的搜寻空间.

3.2.5 增量更新

Papagelis 等^[87]采用增量更新方法(Incremen-

tal) 提高协同过滤的可扩展性. 将相似性计算公式的分子和分母拆分为 B 、 C 、 D 三个独立因子, 目标用户提交新的评分后, B 、 C 、 D 会产生相应的增量值 e 、 f 、 g . 这样计算目标用户与其他用户的相似性, 只需通过很小的计算量求得 e 、 f 、 g 即可, 不用再循环项目空间中的所有项目, 依次对比每个项目上用户间的评分差异. 采用增量更新机制后, 目标用户所有相似性的计算复杂度由 $O(mn)$ 减少为 $O(m)$, 其中 m 为用户数量, n 为项目数量. Luo 等^[88]分析认为 Papagelis 方法的空间复杂度较高, 且推荐质量较差. Luo 变换 Pearson 相关系数公式, 生成新的相似性计算公式广义 Dice 系数, 并使用相似性支持度调整 Dice 系数. 在评分预测阶段, 考虑用户和项目的评分偏差. 复杂度分析和实验结果表明, Luo 方法较 Papagelis 方法具有更小的存储开销和更高的预测准确性. 李聪等^[89]受 Papagelis 方法的启发, 提出一种项目相似性增量更新机制, 提高基于项目协同过滤的可扩展性.

3.2.6 可扩展性问题研究现状小结

数据集缩减算法实现难度较小, 但删除大量用户的评分数据, 会造成信息丢失, 很多潜在最近邻用户被排除在搜寻范围之外.

聚类技术是解决可扩展性问题的最有效手段, 它的优点是划分后的每一子数据集中用户/项目都是相似的, 且对可扩展性问题的改善程度较显著. 其缺点是受数据稀疏性影响, 聚类算法的质量急剧下降, 使后续最近邻搜寻不够准确, 且有些聚类算法的实现难度较大.

矩阵分解与 PCA 对可扩展性问题的改善程度也较显著, 但二者都会造成信息丢失. 矩阵分解的效果与数据集密切相关, 在项目空间维数很高的情况下, 矩阵分解的效果难以保证.

增量更新将相似性公式拆分为 3 个独立因子, 目标用户提交新评分后, 算法只需在原有相似性基础上计算独立因子的增量即可, 无需扫描整个项目空间, 算法处理速度大为提高, 且不损害推荐准确性. 但该方法不适用于系统新用户, 新用户由于不存在任何相似性数据, 仍需在所有项目上计算与其余用户的相似性. 另外当系统中用户评分活动较频繁, 目标用户与其他用户同时提交评分时, 增量更新的处理过程会变得非常复杂.

本节从可扩展性改善程度、推荐质量、算法实现难度 3 方面对比上述可扩展性问题解决方案, 详细内容见表 3.

表3 可扩展性问题的解决方案对比

Table 3 Comparison of different solutions to scalability problem

	数据集 缩减	聚类	矩阵 分解	主成分 分析	增量更新
可扩展性 改善程度	高	高	高	高	中
算法 推荐质量	低	中	中	低	高
算法 实现难度	低	中	高	中	低
典型算法	DWT ^[69]	K-means ^[71]	SVD ^[80]	IPCA ^[85]	Incremental ^[87]

3.3 多内容问题的研究现状

多内容问题首先由 Yu 等^[29]提出,他们指出传统的协同过滤算法没有考虑项目类别的影响,当网站中项目类别的内容完全不同时,传统的协同过滤算法搜寻出的最近邻往往不够合理,导致推荐质量较差.针对这一问题,他们首先计算目标项目与其他项目的相似性,确定目标项目所属类别,然后在目标项目所属类别中搜寻目标用户的最近邻.由于所有最近邻对被预测项目的内容都较熟悉,因此推荐精度较高.

但是 Yu 的方法仅在一个项目类别中搜寻最近邻,实际网站中用户的评分非常稀疏,一个项目类别中的评分更稀疏,使得他们的方法在实际的应用中效果很差.文献[90]针对以上问题提出一种基于项类偏好的协同过滤推荐算法,首先为目标用户找出一组项类偏好一致的候选邻居,然后在候选邻居中搜寻目标用户的 k 最近邻,从整体上提高最近邻搜寻的准确性.

与多内容问题相关的研究还包括:李聪等^[34]也在一个项目类别中搜寻最近邻(称为领域最近邻),采用领域最近邻的评分预测用户评分项并集中的评分空值,该方法与 Yu 的方法存在同样问题.张海鹏等^[91]构建用户-项目类别矩阵,计算项目与所有类中心的相似性,选择相似性较大的类中心作为最近邻,预测项目评分,并填补用户-项目评分矩阵.由于类中心的个数太少,最近邻选择范围较窄,该算法的填补效果难以保证. Papagelis^[92]根据用户在项类上积极评分数与评分总数的比值构建用户-类别矩阵,在用户-类别矩阵上计算用户间的类别偏好相似性,直接加权类别偏好相似性产生预测结果.无论是积极评分还是消极评分都表达用户对项目所属类别的关注, Papagelis 过分强调积极评分的作用,其计算得到的类别偏好值不够合理.另外 Papagelis 使用用户间的类别偏好相似性,而不是评分相似性进行评分预测,结果的可靠性值得商榷.

对于多内容问题,国内外学者的研究尚处于探索阶段,并不像稀疏性问题、可扩展性问题的研究那样成熟.在多内容问题研究方面,学者们更多采用单独项类搜寻最近邻的方法,单独项类评分数据异常稀疏,这在现实系统中是不可行的.可考虑设计一些方法,如首先计算用户间的项目类别偏好相似性,然后在项目类别偏好相似性较高的用户中搜寻目标用户的最近邻,或结合基于内容的推荐,对目标用户感兴趣项类中的特定项目进行内容分析,找出与这些内容匹配度较高的备选项目,然后根据目标用户最近邻的意见,从备选项目中产生 top- N 推荐集.

除以上问题外,研究人员还研究协同过滤中存在的同义词、托攻击、冷启动、推荐信任度、显示跟踪等问题.

同义词问题是指当两个(多个)项目内容相同或相近,名称不同时,协同过滤推荐系统无法挖掘这种内在关联,仍将这项目区别对待,使得推荐效果变差.一些研究人员建立特征词词典或采用自动特征词扩展解决同义词问题^[93].文献[94]指出潜在语义检索技术是解决同义词问题的有效方案.

托攻击问题指某些网站操纵用户,令他们故意高分或低分地评价一些项目,以使某些项目变得畅销,另一些项目不受欢迎.托攻击导致项目的销售情况被人为干扰,影响所有用户对协同过滤推荐系统的满意度. Lam 等^[95]发现基于项目的协同过滤比基于用户的协同过滤更能抵御恶意用户的攻击. Mobasher 等^[96]指出混合推荐方法或基于模型的协同过滤可部分解决托攻击问题. Bell 等^[97]定义全局影响,该方法在数据标准化阶段移除部分全局影响,利用剩余的全局影响搜寻最近邻,将恶意用户排除在最近邻搜寻范围之外.

冷启动问题是稀疏性的极端情况,也称为“早期评价人问题(Early-Rater Problem)”,具体可分为“新用户问题(New User Problem)”和“新项目问题(New Item Problem)”,即当一个新用户(新项目)刚进入推荐系统不久,由于其提供(接受)的评价信息极为有限,系统难以向新用户推荐其可能喜欢的项目或将新项目推荐给可能喜欢它的用户^[3,94].文献[98]设计相似性度量方法(Proximity-Impact-Popularity, PIP)来解决冷启动问题, PIP 计算用户共同评分的近似度、影响度和流行度,将三者的乘积作为搜寻最近邻的基础. Liu 等^[99]指出 PIP 有 3 点不足: 1) 只考虑用户评分值差异,未考虑用户评分数量信息; 2) 仅从局部语境中确定用户相似性,未考虑用户的全局偏好; 3) PIP 非标准化的公式难以与其他相似

性算法结合. 因此他们改进 PIP, 形成新的启发式相似性度量方法(New Heuristic Similarity Model, NHSM). NHSM 分别计算用户间的 PSS(Proximity-Significance-Singularity) 相似性、Jaccard 相似性和 URP(User-Rating-Preference) 相似性, 将三者乘积作为用户间最终的相似性. Kim 等^[18] 提出一种偏差映射模型, 采用协同过滤方法预测用户的已有评分, 将预测偏差存储到一个矩阵中; 对于用户的评分空值, 计算出预测结果, 并使用矩阵中的预测偏差修正预测结果. Kim 的方法适用于评分数据极端稀疏的情况, 当用户评分数量普遍小于 5 时, 方法具有更高的准确性. Li 等^[77] 提出一种基于项目的概率模型, 使用内容信息扩充项目评分向量, 缓解冷启动中的新项目问题.

显示跟踪是用户偏好获取阶段面临的挑战, 由于系统让用户评价项目会给用户带来额外负担, 用户通常只提供少部分评价信息, 导致评价数据过于稀疏, 推荐质量下降. Schafer 等^[100] 指出协同过滤推荐系统的维护人员可采取适当的激励措施, 提高用户评分的积极性, 增加系统显示评价信息的数量.

另外还有研究人员指出协同过滤系统在进行项目推荐时, 不仅要根据预测评分的高低选择项目, 还应考虑预测评分的可信度. Schafer 等^[100] 的研究表明, 推荐项目时综合考虑预测评分和信任度比单纯采用预测评分更合理, 他们分别给出基于用户和基于项目协同过滤的信任度计算方法.

4 未来研究方向

尽管研究人员研究协同过滤的稀疏性问题、多内容问题和可扩展性问题, 并取得一些成果, 然而对于协同过滤的研究还有待不断完善, 今后的进一步研究工作包括如下方面.

1) 新用户偏好模型. 用户偏好信息是对用户提供个性化推荐服务的基础, 系统只有充分掌握用户偏好, 才有可能提供高质量的推荐. 目前协同过滤推荐系统主要让用户评分项目, 以评分信息反映用户的偏好. 评分方式简单、直接, 易于推荐模型的建立, 同时对用户操作系统时的干扰较小, 但评分数据所包含的信息量有限, 仅依靠评分难以提供令用户满意的推荐. 另外很多项目不适合用纯定量化的方式评价, 这给用户表达偏好/系统分析用户偏好增加难度. 因此有必要尝试构建新的用户偏好模型, 可从用户对项目的评论信息中提取用户偏好, 建立基于文本信息的用户偏好模型. 借鉴文献[101]中的 PR-

CUM(Personalized Recommendations via Collaborative User Model) 挖掘用户评论项目的文本信息, 采用特征词模式、特征词权重表示用户偏好, 将用户偏好表达拓展到文本信息空间. 用户感兴趣的特征词数量远大于评分项目数量, 可降低偏好数据的稀疏性. 词语等文本信息与评分值相比具有更丰富的含义, 能提高偏好表达的准确性.

2) 改进评价方法. 评价方法可用于评估协同过滤系统的性能, 确定系统是否很好地满足用户需求^[100]. 好的评价方法可跟踪用户使用系统满意程度, 及时发现系统不足, 不断完善协同过滤系统. 目前绝大多数协同过滤算法的评价采用离线分析的方式, 即将用户已评分的项目划分为训练集和测试集, 采用训练集中的数据预测用户测试集中的偏好, 将测试集中的真实偏好与预测偏好对比评估算法优劣. 该方法以用户评价过的项目为基础评估算法, 是对真实情况的一种模拟. 离线分析的优势在于它能在任何数据集上快速、经济地评价算法, 尤其是规模很大的数据集. 但该方法有 2 点明显不足: (1) 评价时所用项目仅局限于用户已有评分的项目, 由于数据稀疏性, 多数项目不能用于评估算法; (2) 仅局限于对预测结果的客观评价, 不能反映用户对协同过滤系统的主观偏好. 一种变通的方法是在线对用户访问、调查, 收集用户对推荐项目集合的评论信息, 或直接获取用户使用系统的满意程度^[102]. 在线评价通过与用户互动得到其真实感受, 较离线分析更准确, 是未来评价协同过滤系统的主要趋势.

另外要扩展系统的评价准则. 最常用的评价准则是准确性准则, 如评价预测准确性的平均绝对误差(Mean Absolute Error, MAE)、均方根误差(Root Mean Squared Error, RMSE) 等, 评价推荐准确性的 Recall、Precision、F1 等. 准确性反映的是系统的预测评分能力, 预测准确性高并不能保证系统用户有一个满意的使用经历. 协同过滤系统不仅要提供准确的预测, 更要提供有用的推荐. 因此用于衡量系统有用性的准则需不断被采用, 如覆盖率准则、新颖性准则、学习率准则、满意度准则等^[100, 102].

3) 推荐解释. 协同过滤系统的界面多向用户展示项目的推荐列表(有时伴随项目的预测评分), 采用的是黑箱机制(Black Box), 缺乏对推荐原理的介绍. 解释推荐产生的原因可使系统的工作机理更透明, 从而帮助用户了解系统的优缺点. 对推荐过程中的细节和推荐结果做出解释, 使用户知道怎样获得的推荐, 为何会得到如此推荐, 用户会更信任系统. Herlocker 等^[103] 探索性研究协同过滤的推荐解释问

题 提出一个推荐解释概念模型,使用易用性评价方法测试几种不同的解释类型.他们的研究表明:邻居用户怎样评价项目、协同过滤系统以往的表现、新项目与用户评高分项目的相似性、喜欢的演员是否在项目中出现等因素都为用户所关心,影响用户对推荐结果的信任程度.未来的工作可探究更丰富的推荐解释体系,设计并开发合理的、符合用户使用习惯的推荐解释界面等.

4) 隐私保护.为向用户提供个性化的商品或服务,协同过滤系统需了解用户的个人信息.事实上,系统对用户了解的越多,越能为用户提供高质量的推荐.随着系统存储的用户信息不断增多,用户愈加关心自己的何种信息被采集,这些信息存放在哪里,被用来做些什么.资源库在协同过滤系统中居于核心地位,它存储用户的评分及其他相关信息,一旦遭到攻击,用户的信息安全将受到威胁.近年来,协同过滤中的隐私保护问题引起越来越多研究人员的兴趣.隐私保护最常用的技术大致可分为两类:基于加密的技术和随机扰乱技术.文献[100]、[104]~[106]初步研究协同过滤推荐的隐私保护问题,但该问题的研究还有待进一步深化.

5) 协同过滤与云计算技术的结合.协同过滤推荐系统常需要处理上百万乃至上亿级的数据记录,如何存取这些海量数据以及快速计算出推荐结果,成为协同过滤面临的巨大挑战.云计算采用分布式技术,云端系统构建在大量廉价的服务器集群之上,通过大规模集群获得与高性能计算机相当的计算能力和存储能力,集群的管理提高系统的可扩展性^[107].可尝试将协同过滤与云计算技术相结合,使协同过滤推荐系统具有更强的并行计算能力、容错能力和更高的推荐实时性.

6) 领袖客户的识别.领袖客户又称核心客户,通常以其亲身经历、体验和专业化的视角,在网络社区中发表一些评价商品(服务)性能和功能的言论.该类客户的言论说服力强,内容容易被理解、认同和接受,对社区中其他成员的消费观念和购买决策有着重要影响^[108].如何区分领袖客户与一般客户,合理确定领袖客户意见(偏好)的权重,在协同过滤推荐的研究中具有重要意义.

5 结 束 语

推荐系统帮助用户在海量的信息资源中搜寻真正有价值的信息,节约用户的搜寻成本,同时也提高

用户对网站的忠诚度,增加网站收益.由于巨大的应用需求,推荐系统得到学术界和企业界的广泛关注.美国计算机学会 ACM(Association for Computing Machinery) 多次把推荐系统列为研讨主题,国内外期刊也纷纷将推荐系统作为研究专题^[8].推荐系统在电子商务、社会网络、数字化图书馆、视频/音乐点播等领域得到广泛应用,未来的社会化网站将由推荐系统所驱动.

协同过滤是推荐系统中应用较成功的推荐技术,自 1992 年美国施乐公司 PARC 研究中心正式提出协同过滤以来,国内外众多研究人员对如何完善协同过滤推荐技术开展大量的工作.本文对协同过滤领域的相关研究进行系统归纳,介绍一个相对完整的协同过滤知识框架.这对于理清协同过滤的研究脉络,为后续研究提供参考,推进我国个性化信息服务的发展具有一定意义.

参 考 文 献

- [1] Chen R M. Challenge, Value and Coping Strategy in the Big Data Era. *Mobile Communications*, 2012, 36(17): 14 - 15 (in Chinese)
(陈如明. 大数据时代的挑战、价值与应对策略. *移动通信*, 2012, 36(17): 14 - 15)
- [2] Borchers A, Herlocker J, Konstan J, et al. Ganging up on Information Overload. *Computer*, 1998, 31(4): 106 - 108
- [3] Li C. Research on the Bottleneck Problems of Collaborative Filtering in E-commerce Recommender Systems. Ph. D Dissertation. Hefei, China: Hefei University of Technology, 2009 (in Chinese)
(李 聪. 电子商务推荐系统中协同过滤瓶颈问题研究. 博士学位论文. 合肥: 合肥工业大学, 2009)
- [4] Resnick P, Varian H R. Recommender Systems. *Communications of the ACM*, 1997, 40(3): 56 - 58
- [5] Zenebe A, Norcio A F. Representation, Similarity Measures and Aggregation Methods Using Fuzzy Sets for Content-Based Recommender Systems. *Fuzzy Sets and Systems*, 2009, 160(1): 76 - 94
- [6] Schafer J B, Konstan J A, Riedl J. E-commerce Recommendation Applications. *Data Mining and Knowledge Discovery*, 2001, 5(1/2): 115 - 153
- [7] Adomavicius G, Tuzhilin A. Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. *IEEE Trans on Knowledge and Data Engineering*, 2005, 17(6): 734 - 749
- [8] Xia P Y. Research on Collaborative Filtering Algorithm of Personalized Recommendation Technology. Ph. D Dissertation. Qingdao, China: Ocean University of China, 2011 (in Chinese)
(夏培勇. 个性化推荐技术中的协同过滤算法研究. 博士学位论文. 青岛: 中国海洋大学, 2011)
- [9] Su X N, Yang J L, Deng S H, et al. Theory and Technology of Data Mining. Beijing, China: Scientific and Technical Documentation

- Press, 2003 (in Chinese)
(苏新宁 杨建林 邓三鸿, 等. 数据挖掘理论与技术. 北京: 科学文献出版社, 2003)
- [10] Baeza-Yates R, Ribeiro-Neto B. Modern Information Retrieval. New York, USA: ACM Press, 1997
- [11] Xu H L, Wu X, Li X D, *et al.* Comparison Study of Internet Recommendation System. Journal of Software, 2009, 20(2): 350 – 362 (in Chinese)
(许海玲 吴 潇 李晓东, 等. 互联网推荐系统比较研究. 软件学报, 2009, 20(2): 350 – 362)
- [12] Jeong B, Lee J, Cho H. An Iterative Semi-explicit Rating Method for Building Collaborative Recommender Systems. Expert Systems with Applications, 2009, 36(3): 6181 – 6186
- [13] de Campos L M, Fernández-Luna J M, Huete J F, *et al.* Combining Content-Based and Collaborative Recommendations: A Hybrid Approach Based on Bayesian Networks. International Journal of Approximate Reasoning, 2010, 51(7): 785 – 799
- [14] Karypis G. Evaluation of Item-Based Top-N Recommendation Algorithms // Proc of the 10th International Conference on Information and Knowledge Management. Atlanta, USA, 2001: 247 – 254
- [15] Liang C Y, Leng Y J, Wang Y S, *et al.* Research on Group Recommendation in E-commerce Recommender Systems. Chinese Journal of Management Science, 2013, 21(3): 153 – 158 (in Chinese)
(梁昌勇 冷亚军 王勇胜, 等. 电子商务推荐系统中群体用户推荐问题研究. 中国管理科学, 2013, 21(3): 153 – 158)
- [16] Herlocker J L, Konstan J A, Borchers A, *et al.* An Algorithmic Framework for Performing Collaborative Filtering // Proc of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. Berkeley, USA, 1999: 230 – 237
- [17] Herlocker J, Konstan J A, Riedl J. An Empirical Analysis of Design Choices in Neighborhood-Based Collaborative Filtering Algorithms. Information Retrieval, 2002, 5(4): 287 – 310
- [18] Kim H N, El-Saddik A, Jo G S. Collaborative Error-Reflected Models for Cold-Start Recommender Systems. Decision Support Systems, 2011, 51(3): 519 – 531
- [19] Leung C W K, Chan S C F, Chung F L. A Collaborative Filtering Framework Based on Fuzzy Association Rules and Multiple-Level Similarity. Knowledge and Information Systems, 2006, 10(3): 357 – 381
- [20] Goldberg D, Nichols D, Oki B M, *et al.* Using Collaborative Filtering to Weave an Information Tapestry. Communications of the ACM, 1992, 35(12): 61 – 70
- [21] Resnick P, Iacovou N, Suchak M, *et al.* GroupLens: An Open Architecture for Collaborative Filtering of Netnews // Proc of the ACM Conference on Computer Supported Cooperative Work. Chapel Hill, USA, 1994: 175 – 186
- [22] Konstan J A, Miller B N, Maltz D, *et al.* GroupLens: Applying Collaborative Filtering to Usenet News. Communications of the ACM, 1997, 40(3): 77 – 87
- [23] Miller B N, Riedl J T, Konstan J A. Experiences with GroupLens: Making Usenet Useful Again // Proc of the USENIX Winter Technical Conference. Anaheim, USA, 1997: 219 – 231
- [24] Breese J S, Heckerman D, Kadie C. Empirical Analysis of Predictive Algorithms for Collaborative Filtering // Proc of the 14th Conference on Uncertainty in Artificial Intelligence. San Francisco, USA, 1998: 43 – 52
- [25] Sarwar B, Karypis G, Konstan J, *et al.* Item-Based Collaborative Filtering Recommendation Algorithms // Proc of the 10th International Conference on World Wide Web. Hong Kong, China, 2001: 285 – 295
- [26] Papagelis M, Plexousakis D, Kutsuras T. Alleviating the Sparsity Problem of Collaborative Filtering Using Trust Inferences // Proc of the 3rd International Conference on Trust Management. Paris, France, 2005: 224 – 239
- [27] Barragáns-Martínez A B, Costa-Montenegro E, Burguillo-Rial J C, *et al.* A Hybrid Content-Based and Item-Based Collaborative Filtering Approach to Recommend TV Programs Enhanced with Singular Value Decomposition. Information Sciences, 2010, 180(22): 4290 – 4311
- [28] Kim H N, Ji A T, Ha I, *et al.* Collaborative Filtering Based on Collaborative Tagging for Enhancing the Quality of Recommendation. Electronic Commerce Research and Applications, 2010, 9(1): 73 – 83
- [29] Yu L, Liu L, Li X F. A Hybrid Collaborative Filtering Method for Multiple-Interests and Multiple-Content Recommendation in E-commerce. Expert Systems with Applications, 2005, 28(1): 67 – 77
- [30] Albadvi A, Shahbazi M. A Hybrid Recommendation Technique Based on Product Category Attributes. Expert Systems with Applications, 2009, 36(9): 11480 – 11488
- [31] Su J H, Wang B W, Hsiao C Y, *et al.* Personalized Rough-Set-Based Recommendation by Integrating Multiple Contents and Collaborative Information. Information Sciences, 2010, 180(1): 113 – 131
- [32] Sun X H. Research of Sparsity and Cold Start Problem in Collaborative Filtering. Ph. D Dissertation. Hangzhou, China: Zhejiang University, 2005 (in Chinese)
(孙小华. 协同过滤系统的稀疏性与冷启动问题研究. 博士学位论文. 杭州: 浙江大学, 2005)
- [33] Deng A L, Zhu Y Y, Shi B L. A Collaborative Filtering Recommendation Algorithm Based on Item Rating Prediction. Journal of Software, 2003, 14(9): 1621 – 1628 (in Chinese)
(邓爱林 朱扬勇 施伯乐. 基于项目评分预测的协同过滤推荐算法. 软件学报, 2003, 14(9): 1621 – 1628)
- [34] Li C, Liang C Y, Ma L. A Collaborative Filtering Recommendation Algorithm Based on Domain Nearest Neighbor. Journal of Computer Research and Development, 2008, 45(9): 1532 – 1538 (in Chinese)
(李 聪 梁昌勇 马 丽. 基于领域最近邻的协同过滤推荐算法. 计算机研究与发展, 2008, 45(9): 1532 – 1538)
- [35] Zhang F, Chang H Y. Employing BP Neural Networks to Alleviate the Sparsity Issue in Collaborative Filtering Recommendation Algorithms. Journal of Computer Research and Development, 2006, 43(4): 667 – 672 (in Chinese)
(张 锋 常会友. 使用 BP 神经网络缓解协同过滤推荐算法的稀疏性问题. 计算机研究与发展, 2006, 43(4): 667 – 672)

- [36] Zhang J Y, Pu P. A Recursive Prediction Algorithm for Collaborative Filtering Recommender Systems // Proc of the ACM Conference on Recommender Systems. Minneapolis, USA, 2007: 57 – 64
- [37] Leng Y J, Liang C Y, Lu Q, *et al.* Collaborative Filtering Recommendation Algorithm Based on Neighbor Rating Imputation. Computer Engineering, 2012, 38(21): 56 – 58, 66 (in Chinese)
(冷亚军 梁昌勇 陆青, 等. 基于近邻评分填补的协同过滤推荐算法. 计算机工程, 2012, 38(21): 56 – 58, 66)
- [38] Su X Y, Khoshgoftaar T M, Greiner R. A Mixture Imputation-Boosted Collaborative Filter // Proc of the 21st International Florida Artificial Intelligence Research Society Conference. Coconut Grove, USA, 2008: 312 – 316
- [39] Zhou J F, Tang X, Guo J F. An Optimized Collaborative Filtering Recommendation Algorithm. Journal of Computer Research and Development, 2004, 41(10): 1842 – 1847 (in Chinese)
(周军锋 汤显 郭景峰. 一种优化的协同过滤推荐算法. 计算机研究与发展, 2004, 41(10): 1842 – 1847)
- [40] Zhang G W, Li D Y, Li P, *et al.* A Collaborative Filtering Recommendation Algorithm Based on Cloud Model. Journal of Software, 2007, 18(10): 2403 – 2411 (in Chinese)
(张光卫 李德毅 李鹏, 等. 基于云模型的协同过滤推荐算法. 软件学报, 2007, 18(10): 2403 – 2411)
- [41] Luo H, Niu C Y, Shen R M, *et al.* A Collaborative Filtering Framework Based on Both Local User Similarity and Global User Similarity. Machine Learning, 2008, 72(3): 231 – 245
- [42] Anand D, Bharadwaj K K. Utilizing Various Sparsity Measures for Enhancing Accuracy of Collaborative Recommender Systems Based on Local and Global Similarities. Expert Systems with Applications, 2011, 38(5): 5101 – 5109
- [43] Choi K, Suh Y. A New Similarity Function for Selecting Neighbors for Each Target Item in Collaborative Filtering. Knowledge-Based Systems, 2013, 37: 146 – 153
- [44] Kaleli C. An Entropy-Based Neighbor Selection Approach for Collaborative Filtering. Knowledge-Based Systems, 2014, 56: 273 – 280
- [45] Jeong B, Lee J, Cho H. Improving Memory-Based Collaborative Filtering via Similarity Updating and Prediction Modulation. Information Sciences, 2010, 180(5): 602 – 612
- [46] Leng Y J, Liang C Y, Ding Y, *et al.* Method of Neighborhood Formation in Collaborative Filtering. Pattern Recognition and Artificial Intelligence, 2013, 26(10): 968 – 974 (in Chinese)
(冷亚军 梁昌勇 丁勇, 等. 协同过滤中一种有效的最近邻选择方法. 模式识别与人工智能, 2013, 26(10): 968 – 974)
- [47] Gao Y, Qi H, Liu J, *et al.* A Collaborative Filtering Recommendation Algorithm Combining Probabilistic Relational Models and User Grade. Journal of Computer Research and Development, 2008, 45(9): 1463 – 1469 (in Chinese)
(高滢 齐红 刘杰, 等. 结合似然关系模型和用户等级的协同过滤推荐算法. 计算机研究与发展, 2008, 45(9): 1463 – 1469)
- [48] Melville P, Mooney R J, Nagarajan R. Content-Boosted Collaborative Filtering for Improved Recommendations // Proc of the 18th National Conference on Artificial Intelligence. Edmonton, Canada, 2002: 187 – 192
- [49] Lee T Q, Park Y, Park Y T. A Time-Based Approach to Effective Recommender Systems Using Implicit Feedback. Expert Systems with Applications, 2008, 34(4): 3055 – 3062
- [50] Liu D R, Lai C H, Lee W J. A Hybrid of Sequential Rules and Collaborative Filtering for Product Recommendation. Information Sciences, 2009, 179(20): 3505 – 3519
- [51] Cai Q, Han D M, Li H S, *et al.* Personalized Resource Recommendation Based on Tags and Collaborative Filtering. Computer Science, 2014, 41(1): 69 – 71, 110 (in Chinese)
(蔡强 韩东梅 李海生, 等. 基于标签和协同过滤的个性化资源推荐. 计算机科学, 2014, 41(1): 69 – 71, 110)
- [52] Chen J, Yin J. A Collaborative Filtering Recommendation Algorithm Based on Influence Sets. Journal of Software, 2007, 18(7): 1685 – 1694 (in Chinese)
(陈健 印鉴. 基于影响集的协作过滤推荐算法. 软件学报, 2007, 18(7): 1685 – 1694)
- [53] George T, Merugu S. A Scalable Collaborative Filtering Framework Based on Co-clustering // Proc of the 5th IEEE International Conference on Data Mining. 2005. DOI: 10.1109/ICDM.2005.14
- [54] Lee J S, Olafsson S. Two-Way Cooperative Prediction for Collaborative Filtering Recommendations. Expert Systems with Applications, 2009, 36(1): 5353 – 5361
- [55] Wang J, de Vries A P, Reinders M J T. Unifying User-Based and Item-Based Collaborative Filtering Approaches by Similarity Fusion // Proc of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. Seattle, USA, 2006: 501 – 508
- [56] Liang C Y, Leng Y J. Collaborative Filtering Based on Information-Theoretic Co-clustering. International Journal of Systems Science, 2014, 45(3): 589 – 597
- [57] Aggarwal C C, Wolf J L, Wu K L, *et al.* Horting Hatches an Egg: A New Graph-Theoretic Approach to Collaborative Filtering // Proc of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Diego, USA, 1999: 201 – 212
- [58] Yang X W, Guo Y, Liu Y, *et al.* A Survey of Collaborative Filtering Based Social Recommender Systems. Computer Communications, 2014, 41: 1 – 10
- [59] Zhou T, Ren J, Medo M, *et al.* Bipartite Network Projection and Personal Recommendation. Physical Review E, 2007. DOI: 10.1103/PhysRevE.76.046115
- [60] Zeng W, Zhu Y X, Lü L Y, *et al.* Negative Ratings Play a Positive Role in Information Filtering. Physica A: Statistical Mechanics and Its Applications, 2011, 390(23/24): 4486 – 4493
- [61] Pan H Y, Lin H F, Zhao J. Collaborative Filtering Algorithm Based on Matrix Partition and Interest Variance. Journal of the China Society for Scientific and Technical Information, 2006, 25(1): 49 – 54 (in Chinese)
(潘红艳 林鸿飞 赵晶. 基于矩阵划分和兴趣方差的协同过滤算法. 情报学报, 2006, 25(1): 49 – 54)
- [62] Goldberg K, Roeder T, Gupta D, *et al.* Eigentaste: A Constant Time Collaborative Filtering Algorithm. Information Retrieval,

- 2001, 4(2): 133 – 151
- [63] Hatami M, Pashazadeh S. Enhancing Prediction in Collaborative Filtering-Based Recommender Systems. *International Journal of Computer Sciences and Engineering*, 2014, 2(1): 48 – 51
- [64] Yang B, Lei Y, Liu D, *et al.* Social Collaborative Filtering by Trust // *Proc of the 23rd International Joint Conference on Artificial Intelligence*. Beijing, China, 2013: 2747 – 2753
- [65] Koren Y, Bell R, Volinsky C. Matrix Factorization Techniques for Recommender Systems. *Computer*, 2009, 42(8): 30 – 37
- [66] Koren Y, Bell R. Advances in Collaborative Filtering // Ricci F, Rokach R, Shapira B, *et al.*, eds. *Recommender Systems Handbook*. New York, USA: Springer, 2011: 145 – 186
- [67] Linden G, Smith B, York J. Amazon. com Recommendations: Item-to-Item Collaborative Filtering. *IEEE Internet Computing*, 2003, 7(1): 76 – 80
- [68] Yu K, Schwaighofer A, Tresp V, *et al.* Probabilistic Memory-Based Collaborative Filtering. *IEEE Trans on Knowledge and Data Engineering*, 2004, 16(1): 56 – 69
- [69] Russell S, Yoon V. Applications of Wavelet Data Reduction in a Recommender System. *Expert Systems with Applications*, 2008, 34(4): 2316 – 2325
- [70] Acilar A M, Arslan A. A Collaborative Filtering Method Based on Artificial Immune Network. *Expert Systems with Applications*, 2009, 36(4): 8324 – 8332
- [71] Sarwar B M, Karypis G, Konstan J, *et al.* Recommender Systems for Large-Scale E-commerce: Scalable Neighborhood Formation Using Clustering [EB/OL]. [2013 – 10 – 25]. http://www.grouplens.org/papers/pdf/sarwar_cluster.pdf
- [72] Li T, Wang J D, Ye F Y, *et al.* Collaborative Filtering Recommendation Algorithm Based on Clustering Basal Users. *Systems Engineering and Electronics*, 2007, 29(7): 1178 – 1182 (in Chinese)
- (李涛, 王建东, 叶飞跃, 等. 一种基于用户聚类的协同过滤推荐算法. *系统工程与电子技术*, 2007, 29(7): 1178 – 1182)
- [73] Yu X, Li M Q. Effective Hybrid Collaborative Filtering Model Based on PCA-SOM. *Systems Engineering – Theory & Practice*, 2010, 30(10): 1850 – 1854 (in Chinese)
- (郁雪, 李敏强. 基于 PCA-SOM 的混合协同过滤模型. *系统工程理论与实践*, 2010, 30(10): 1850 – 1854)
- [74] Leng Y J, Liang C Y, Lu W X. Clustering of Web Users Based on an Improved Affinity Propagation Algorithm. *Journal of the China Society for Scientific and Technical Information*, 2012, 31(9): 993 – 997 (in Chinese)
- (冷亚军, 梁昌勇, 陆文星. 基于改进近邻传播算法的 Web 用户聚类. *情报学报*, 2012, 31(9): 993 – 997)
- [75] O'Connor M, Herlocker J. Clustering Items for Collaborative Filtering [EB/OL]. [2013 – 10 – 25]. http://www.cin.ufpe.br/~idal/rs/oconner_m.pdf
- [76] Deng A L, Zuo Z Y, Zhu Y Y. Collaborative Filtering Recommendation Algorithm Based on Item Clustering. *Mini-Micro Systems*, 2004, 25(9): 1665 – 1670 (in Chinese)
- (邓爱林, 左子叶, 朱扬勇. 基于项目聚类的协同过滤推荐算法. *小型微型计算机系统*, 2004, 25(9): 1665 – 1670)
- [77] Li Q, Myaeng S H, Kim B M. A Probabilistic Music Recommender Considering User Opinions and Audio Features. *Information Processing and Management*, 2007, 43(2): 473 – 487
- [78] Wu M L, Chang C H, Liu R Z. Integrating Content-Based Filtering with Collaborative Filtering Using Co-clustering with Augmented Matrices. *Expert Systems with Applications*, 2014, 41(6): 2754 – 2761
- [79] Wu H, Wang Y J, Wang Z, *et al.* Two-Phase Collaborative Filtering Algorithm Based on Co-clustering. *Journal of Software*, 2010, 21(5): 1042 – 1054 (in Chinese)
- (吴湖, 王永吉, 王哲, 等. 两阶段联合聚类协同过滤算法. *软件学报*, 2010, 21(5): 1042 – 1054)
- [80] Sarwar B, Karypis G, Konstan J, *et al.* Application of Dimensionality Reduction in Recommender System – A Case Study [EB/OL]. [2013 – 10 – 25]. <http://www.dtic.mil/get-tr-doc/pdf?AD=ADA439541>
- [81] Vozalis M G, Margaritis K G. Using SVD and Demographic Data for the Enhancement of Generalized Collaborative Filtering. *Information Sciences*, 2007, 177(15): 3017 – 3037
- [82] Chen G, Wang F, Zhang C S. Collaborative Filtering Using Orthogonal Nonnegative Matrix Tri-factorization. *Information Processing and Management*, 2009, 45(3): 368 – 379
- [83] Luo X, Xia Y N, Zhu Q S. Incremental Collaborative Filtering Recommender Based on Regularized Matrix Factorization. *Knowledge-Based Systems*, 2012, 27: 271 – 280
- [84] Yu J H, Yang W Q. *Multivariate Statistical Analysis and Its Application*. Guangzhou, China: Sun Yat-Set University Press, 2005 (in Chinese)
- (余锦华, 杨维权. *多元统计分析与应用*. 广州: 中山大学出版社, 2005)
- [85] Kim D, Yum B J. Collaborative Filtering Based on Iterative Principal Component Analysis. *Expert Systems with Applications*, 2005, 28(4): 823 – 830
- [86] Lee J S, Jun C H, Lee J, *et al.* Classification-Based Collaborative Filtering Using Market Basket Data. *Expert Systems with Applications*, 2005, 29(3): 700 – 704
- [87] Papagelis M, Rousidis I, Plexousakis D, *et al.* Incremental Collaborative Filtering for Highly-Scalable Recommendation Algorithms // *Proc of the 15th International Symposium on Foundations of Intelligent Systems*. Saratoga, USA, 2005: 553 – 561
- [88] Luo X, Xia Y N, Zhu Q S, *et al.* Boosting the K-Nearest-Neighborhood Based Incremental Collaborative Filtering. *Knowledge-Based Systems*, 2013, 53: 90 – 99
- [89] Li C, Liang C Y. Incremental Updating Mechanism of Collaborative Filtering in Accordance with User Interest Changes. *Journal of the China Society for Scientific and Technical Information*, 2010, 29(1): 59 – 66 (in Chinese)
- (李聪, 梁昌勇. 适应用户兴趣变化的协同过滤增量更新机制. *情报学报*, 2010, 29(1): 59 – 66)
- [90] Leng Y J, Liang C Y, Zhang E Q, *et al.* A Collaborative Filtering Recommendation Algorithm Based on Item-Class Preference. *Journal of the China Society for Scientific and Technical Information*, 2011, 30(7): 714 – 720 (in Chinese)

- (冷亚军,梁昌勇,张恩桥,等. 基于项类偏好的协同过滤推荐算法. 情报学报, 2011, 30(7): 714-720)
- [91] Zhang H P, Li L B, Li X, *et al.* Collaborative Filtering Recommendation Algorithm Based on Item-Class Rating. *Journal of the China Society for Scientific and Technical Information*, 2008, 27(2): 218-223 (in Chinese)
- (张海鹏,李烈彪,李仙,等. 基于项目分类预测的协同过滤推荐算法. 情报学报, 2008, 27(2): 218-223)
- [92] Papagelis M, Plexousakis D, Rousidis I, *et al.* Qualitative Analysis of User-Based and Item-Based Prediction Algorithms for Recommendation Systems [EB/OL]. [2013-10-25]. http://www.ics.forth.gr/isl/publications/paperlink/hdms04_camera_ready_submitted.pdf
- [93] Jones K S. A Statistical Interpretation of Term Specificity and Its Application in Retrieval. *Journal of Documentation*, 1972, 28(1): 11-21
- [94] Su X Y, Khoshgoftaar T M. A Survey of Collaborative Filtering Techniques [EB/OL]. [2013-11-01]. <http://www.hindawi.com/journals/aai/2009/421425/>
- [95] Lam S K, Riedl J. Shilling Recommender Systems for Fun and Profit // *Proc of the 13th International Conference on World Wide Web*. New York, USA, 2004: 393-402
- [96] Mobasher B, Burke R, Bhaumik R, *et al.* Effective Attack Models for Shilling Item-Based Collaborative Filtering Systems // *Proc of the 11th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Chicago, USA, 2005: 13-23
- [97] Bell R M, Koren Y. Improved Neighborhood-Based Collaborative Filtering // *Proc of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Jose, USA, 2007: 7-14
- [98] Ahn H J. A New Similarity Measure for Collaborative Filtering to Alleviate the New User Cold-Starting Problem. *Information Sciences*, 2008, 178(1): 37-51
- [99] Liu H F, Hu Z, Mian A, *et al.* A New User Similarity Model to Improve the Accuracy of Collaborative Filtering. *Knowledge-Based Systems*, 2014, 56: 156-166
- [100] Schafer J B, Frankowski D, Herlocker J, *et al.* Collaborative Filtering Recommender Systems // Brusilovsky P, Kobsa A, Nejdl W, eds. *The Adaptive Web*. New York, USA: Springer, 2007: 291-324
- [101] Kim H N, Ha I, Lee K S, *et al.* Collaborative User Modeling for Enhanced Content Filtering in Recommender Systems. *Decision Support Systems*, 2011, 51(4): 772-781
- [102] Herlocker J L, Konstan J A, Terveen L G, *et al.* Evaluating Collaborative Filtering Recommender Systems. *ACM Trans on Information Systems*, 2004, 22(1): 5-53
- [103] Herlocker J L, Konstan J A, Riedl J. Explaining Collaborative Filtering Recommendations // *Proc of the ACM Conference on Computer Supported Cooperative Work*. Philadelphia, USA, 2000: 241-250
- [104] Zhang F, Sun X D, Chang H Y, *et al.* Research on Privacy-Preserving Two-Party Collaborative Filtering Recommendation. *Acta Electronica Sinica*, 2009, 37(1): 84-89 (in Chinese)
- (张锋,孙雪冬,常会友,等. 两方参与的隐私保护协同过滤推荐研究. 电子学报, 2009, 37(1): 84-89)
- [105] Polat H, Du W L. Privacy-Preserving Collaborative Filtering Using Randomized Perturbation Techniques // *Proc of the 3rd IEEE International Conference on Data Mining*. Melbourne, USA, 2003: 625-628
- [106] Canny J. Collaborative Filtering with Privacy // *Proc of the IEEE Symposium on Security and Privacy*. Berkeley, USA, 2002: 45-57
- [107] Lü X J. Research on Intelligent Recommender System Based on Cloud Computing. Master Dissertation. Hefei, China: Anhui University, 2012 (in Chinese)
- (吕雪骥. 基于云计算平台的智能推荐系统研究. 硕士学位论文. 合肥: 安徽大学, 2012)
- [108] He J M. Research on Method of Listening to the Voice of Customers from the Web Community in Internet. Ph. D Dissertation. Hefei, China: Hefei University of Technology, 2010 (in Chinese)
- (何建民. 面向网络社区聆听客户声音方法研究. 博士学位论文. 合肥: 合肥工业大学, 2010)