

基于用户浏览行为的个性化推荐研究综述*

刘华真^{1a,1b}, 王巍^{1a,1b,2,3†}, 谷壬倩^{1a,1b}, 张屹晗^{1a,1b}, 郝亚奇^{1a,1b}

(1. 河北工程大学 a. 信息与电气工程学院; b. 河北省安防信息感知与处理重点实验室, 河北 邯郸 056038; 2. 江南大学 物联网工程学院, 江苏 无锡 214122; 3. 中国民用航空局第二研究所, 成都 610041)

摘要: 研究如何充分利用海量用户浏览行为数据, 构建更加精确的推荐算法和模型, 以提高推荐系统性能, 是目前个性化推荐领域研究的热点。针对这些问题, 首先对用户的浏览行为进行了简要概括表述, 给出了基于浏览行为推荐系统的总体框架, 回顾总结了基于用户浏览行为的推荐系统的发展历程。对其关键技术和单一浏览行为量化方法与混合浏览行为量化方法进行总结、对比和分析。最后讨论了结合多源异构数据的浏览行为推荐的最新成果, 总结了该领域未来研究难点和发展趋势。

关键词: 个性化推荐; 协同过滤; 神经网络; 兴趣度; 多源异构数据

中图分类号: TP391.3 **文献标志码:** A **文章编号:** 1001-3695(2021)08-004-2268-10

doi:10.19734/j.issn.1001-3695.2020.10.0347

Survey of personalized recommendation study based on user browsing behavior

Liu Huazhen^{1a,1b}, Wang Wei^{1a,1b,2,3†}, Gu Renqian^{1a,1b}, Zhang Yihan^{1a,1b}, Hao Yaqi^{1a,1b}

(1. a. School of Information & Electrical Engineering, b. Hebei Key Laboratory of Security & Protection Information Sensing & Processing, Hebei University of Engineering, Handan Hebei 056038, China; 2. School of Internet of Things Engineering, Jiangnan University, Wuxi Jiangsu 214122, China; 3. The Second Research Institute of Civil Aviation Administration of China, Chengdu 610041, China)

Abstract: In order to improve the performance of the recommendation system, the hot research of personalized recommendation field is how to build more accurate recommendation algorithms and models with massive user browsing behavior data. Firstly, this paper briefly described the browsing behavior of users, gave the overall framework of the browsing behavior recommendation system, and summarized the development of the recommendation system based on user browsing behavior. It summarized, compared and analyzed the key techniques and single browsing behavior quantization method and mixed browsing behavior quantization method. Finally, it discussed the latest achievements of browsing behavior recommendation combined with multi-source heterogeneous data and summarized future research difficulties and development trends in this field.

Key words: personalized recommendation; collaborative filtering; neural network; interest; multi-source heterogeneous data

0 引言

近些年, 社交网络媒体行业和电子商务行业的飞速发展, 使互联网用户数大大提高。面对用户浏览互联网时产生的海量多源异构数据, 如何对其进行精确有效地处理以提高推荐精度和用户满意度, 仍旧是个性化推荐服务的研究热点。目前, 冷启动和数据稀疏问题仍旧是个性化推荐服务面临的主要问题。用户—项目间交互产生的浏览行为等数据和浏览上下文等辅助信息为缓解该问题提供了一个方向。如何处理这些复杂多源的数据使其能准确地描述用户的兴趣偏好是目前研究的重点和难点。协同过滤是个性化推荐系统中应用最为广泛的算法, 而根据用户与项目间交互产生的浏览行为数据提取用户兴趣偏好的研究, 贯穿了协同过滤算法发展的整个历程。在基于邻域的协同过滤算法研究初期^[1-4], 提出将浏览时间、是否访问(点击)等浏览行为进行简单量化得到用户—项目间交互的隐式评分, 经实验分析表明: 简单处理浏览行为数据得到的隐式推荐与原有的显式评分推荐的效果相当; Koren 等人^[5]提出的矩阵因子分解协同过滤模型中, 集成用户显式评分、浏览行为布尔隐式反馈评分和用户属性布尔数据等, 更加深入地了解用户兴趣偏好, 解决因新用户增加导致的冷启动问题, 相

比邻域算法能够有效提高推荐系统的精度; 再到近几年发展十分火热的基于神经网络的协同过滤推荐算法^[6], 采用最基本的浏览行为二进制评分量化方法, 结合时间效应等其他辅助信息作为神经网络的输入数据。在个性化推荐服务和多元化社交网络蓬勃发展的今天, 用户浏览行为作为个性化推荐算法的关键输入数据, 对于其处理方式的技术的研究仍存在许多挑战, 需要进一步探索。

经搜索查阅相关文献发现, 目前大多数的推荐系统相关综述主要分为以下四个方面: 针对某种应用场景的综述, 如社会化推荐综述、旅游推荐综述、移动推荐综述; 针对某类或某种技术的综述, 如基于排序学习的推荐算法综述、基于深度学习的推荐系统综述; 针对推荐系统评估方法的综述; 针对推荐系统中某类普遍问题的综述, 如冷启动。没有专门的学者从用户浏览行为这一明确反映用户偏好的数据角度出发对推荐系统的研究进行相关综述。因此, 本文对基于用户浏览行为的推荐算法、研究现状和关键性技术进行比较、分析和总结。

1 用户浏览行为推荐概述

1.1 用户浏览行为

推荐系统利用整个用户群体在浏览互联网时产生的交互

收稿日期: 2020-10-23; **修回日期:** 2020-12-10 **基金项目:** 国家自然科学基金资助项目(61802107); 教育部—中国移动科研基金资助项目(MCM20170204); 河北省高等学校科学技术研究项目(ZD2020171); 江苏省博士后科研资助计划项目(1601085C)

作者简介: 刘华真(1995-), 女, 河南洛阳人, 硕士研究生, 主要研究方向为数据挖掘、隐式推荐系统; 王巍(1983-), 男(通信作者), 河北邯郸人, 副教授, 硕导, 博士, 主要研究方向为隐式人机交互、公共安全物联网(wangwei@hebeu.edu.cn); 谷壬倩(1996-), 女, 河北石家庄人, 硕士研究生, 主要研究方向为区域异构物联网、无人系统; 张屹晗(1996-), 男, 河北邯郸人, 硕士研究生, 主要研究方向为推荐系统; 郝亚奇(1993-), 男, 河北邯郸人, 硕士研究生, 主要研究方向为无人系统。

数据对海量信息进行过滤,为目标用户提供所需信息,既节省了自身运营成本,又可以提高用户的忠诚度和满意度。研究人员在20世纪90年代意识到用户浏览行为对于个性化服务的重要性^[7],用户无须主动参与,从其交互浏览行为数据就能提取到用户兴趣偏好。早期的推荐系统主要由用户主动提交自己的兴趣报告,例如评分、评论等通过协同过滤方法进行推荐,属于显式反馈推荐。该方法会造成用户额外的负担,降低用户体验,而且返回数据根据用户评分标准的不同易造成推荐准确度下降。此时,另一种基于浏览行为的隐式反馈推荐算法应运而生^[7],其与显式反馈推荐方法结合不仅大大地提高了推荐的准确性,推荐系统的应用领域也得到了进一步扩展。

浏览行为的研究应用最初是在信息检索领域,相较于其他反馈,能显著提高信息过滤的性能,从海量信息集合中快速过滤,为检索用户提供与其兴趣偏好相关性最高的信息集合^[8]。Morita等人^[9]通过将用户浏览时间偏好分析结果与用户显式评分对比之后,发现用户在阅读自己喜欢的新闻文章上花费的时间要比普通文章更长,表明用户浏览时间是反映用户兴趣偏好的一项有效数据。Konstan等人^[3]在1997年将基于浏览时间的协作过滤方法应用到Usenet新闻中。Oard等人^[7]在1998年进一步验证了打印、保存和添加书签等浏览行为能表明用户兴趣偏好,可用于弥补显式反馈评分数据不足的情况。浏览行为早期在推荐系统中的实际应用主要是单个用户浏览行为的偏好反馈结合显示反馈评分共同得出用户偏好,是基于用户的协同过滤推荐的早期简易模型。例如:Sakagami等人^[10]提出的根据用户浏览行为数据(滚动、放大窗口等行为)分析用户偏好的ANATAGONOMY服务器,集成用户显式反馈(用户对于文章的评分)和隐式反馈(用户浏览行为)应用于在线电子报纸。

随着互联网的快速发展,用户数量的增加、信息过载问题凸显,仅仅依靠显式反馈进行推荐的准确度稳定性下降,浏览行为等隐式反馈数据在个性化推荐模型中的重要性和需求量增加。众多研究者投入隐式推荐的研究中,在工业界也有简单应用,而隐式推荐中用户浏览行为的分析是其核心之一。在Oard等人^[7]以及Kelly等人^[11]关于用户浏览行为研究的成果上将用户浏览行为根据行为类型划分为三类:a)保存行为^[12],添加或删除书签、下载、收藏、打印、订阅;b)操作行为,点击、复制、拖动滚动条、调整页面大小、浏览时间、搜索;c)重复行为,重复访问、重复购买、重复点击。

1.2 用户浏览行为推荐的研究

推荐系统构建的第一步就是数据输入,互联网数据量和用户数量的急剧增加使显式评分数据的成本大大增加,数据稀疏问题更加严重,导致推荐准确度和用户满意度下降。用户一项目的隐式交互数据是解决该问题的一种理想方案^[3,4]。对用户浏览行为数据的采集,主要有服务器端、客户端和服务端与客户端综合三种采集方式。分析量化采集到的浏览行为数据得到用户对项目的隐式兴趣评分(用户兴趣度),根据用户的个性化需求与应用场景的不同,采用协同过滤等推荐算法得到个性化推荐列表。最终,通过排名列表、图片、链接等形式在页面上将结果推送给目标用户。根据用户和其浏览的项目属性信息、浏览行为数据和用户社交网络信息等多源异构数据,及其分析和处理技术,图1给出了浏览行为推荐的整体框架。

采用关键词 recommendation 和 recommender system 对推荐系统领域的国际顶级学术前沿会议进行题目和关键词检索统计:WWW2020,录用论文217篇,其中推荐系统相关论文38篇,占比17.5%;SIGIR2020录用长论文147篇,其中63篇与推荐相关,占比42.8%,是信息检索领域论文的2倍;KDD2020录用论文216篇,其中43篇与推荐相关,占比19.9%;人工智能顶会AAAI和IJCAI会议中推荐相关论文的数量相比前几年也有很大的增加。人工智能和大数据的流行,使个性化推荐领域的应

用研究变得更加火热。RecSys 近几年的挑战赛中,越来越重视用户交互行为的建模和分析,2017年Xing社交网络工作推荐挑战赛通过用户浏览行为数据(点击、添加书签、删除)与用户和项目的属性数据融合对求职者产生工作建议推荐^[13];2018年由Spotify在线音乐平台提供了一个“百万用户播放列表数据集(MPD)”,数据集由1000个稀疏播放列表组成,包括标题和元数据,以及用户收藏的播放列表数据,对用户推荐500首有序候选曲目列表^[14];2019年由trivago在线全球酒店搜索平台,提供涉及2018年11月一周内访问trivago网站的超过70万用户的近1600万会话交互匿名数据集,用于提供在线旅行推荐问题^[15];2020年由Twitter提供含有大约1.6亿条公共推文的大型公共数据集,含有转发、点赞、评论和转发加评论的社交浏览行为数据、用户数据和Tweet数据,且数据集会根据用户隐私保护条例对已删除推文和用户资料的数据集进行更新,使得挑战者必须不断地对数据测试集进行重新训练,用于对用户浏览行为参与度的预测^[16]。RecSys挑战赛作为推荐系统应用领域的赛事,自举办以来,有很大的影响力,2017—2019年都是对用户点击行为的预测,2020年是对用户浏览行为组合的预测,其根本目的都是对用户交互数据本身进行分析提取得到用户兴趣偏好,分析其在实际推荐系统中的应用性能。

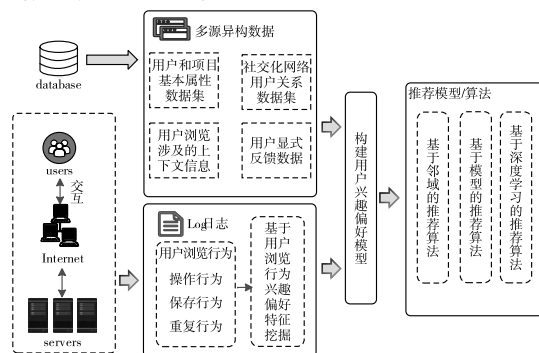


图1 基于浏览行为的个性化推荐系统框架
Fig.1 The framework of personalized recommendation system based on browsing behavior

1994年,Karlgren^[17]提出可以根据浏览行为进行用户聚类,探索了新闻推荐系统的设计,浏览行为开始逐渐运用到推荐系统领域。浏览行为早期在推荐系统的应用为单个浏览行为,主要用于信息过滤,为用户提供需要的项目,过滤掉不需要的信息。Stevens^[18]认为保存项目是一个很明显的选择行为,能够明确表明用户的偏好,表明用户对该项目的兴趣度很高,可有效缓解信息过载。文献[3,9]等发现在新闻系统中阅读时间与用户偏好存在正相关关系,实验结果表明基于浏览时间的隐式评分和基于显式评分的推荐预测准确度相当,而且基于用户浏览时间的用户偏好评分可以适应音频和视频等其他基于内容推荐不能应用的领域。Liang等人^[19]对阅读时间这一浏览行为在新闻个性化服务领域结合新闻内容用于确定用户兴趣作了进一步研究,将用户阅读时间、用户平均阅读速度和浏览新近度混合用于对用户兴趣度的分析和计算。

21世纪初,将隐式反馈如浏览行为等用于推荐系统的还很少,大多数推荐系统用的推荐方法都是基于显示反馈的。相关研究多为分析用户的浏览行为能否提取得到用户的兴趣偏好,进而用于个性化推荐。Sakagami等人^[10]开发了一个通过滚动、放大窗口等浏览行为学习用户阅读偏好的系统ANATAGONOMY,得到隐式评分,以生成个性化的网络新闻推荐。Kelly等人^[20]研究发现用户的浏览行为如添加书签、滚动和浏览时间的组合可以通过强化学习来学习用户的偏好用于信息过滤。

前文已经了解到用户在交互过程中会产生许多种浏览行为,但并不是所有浏览行为都能表明用户兴趣^[11],且表示的兴趣程度也不同。Claypool等人^[21]通过对70多个学生40多个小时的浏览行为数据的分析,发现时间能表明良好的隐式兴趣

评分,而鼠标移动和点击行为本身并不能表明隐性兴趣评分。但浏览行为组合,例如浏览时间和滚动的组合比单个浏览时间行为能更加精确地表示用户兴趣。相较于单个浏览行为对于表明用户兴趣的不足,或对不同类网站的重要性不同,浏览行为组合能更加精确地预测用户兴趣,提高用户对网站个性化服务的满意度。

1.3 用户浏览行为推荐的应用

经研究发现,浏览行为能够体现用户的兴趣偏好,随着互联网产业的快速发展和用户浏览行为获取方式的简化,基于浏览行为的隐式推荐系统已逐渐简单应用于大范围的互联网产业中。采用何种方式将浏览行为中的信息有效地应用于个性化推荐,相关研究者进行了大量的工作。

Xia 等人^[22]采用集群技术对 Web 服务器用户点击日志挖掘,获取用户点击数据,建立 userID-URL 关联矩阵,计算得到距离矩阵。根据日志挖掘算法和聚类将用户分到不同的组中,系统就可以将该组中其他用户浏览而该用户未浏览的项目向该用户进行推荐,以实现个性化服务。但对于此推荐方法的应用,适用场景有限,而且从日志中挖掘出来的数据类型有限,且存在大量噪声,分析处理复杂度较高。

根据用户兴趣偏好建立的个性化兴趣偏好模型是个性化服务的核心。Xing 等人^[23]提出了一种兴趣提取算法(BUIE),通过对浏览行为分析发现相对浏览时间(用户浏览文章的实际时间)对提取用户兴趣特征非常重要,因其反映兴趣的稳定性要远远高于打印、保存、复制等浏览行为。在关注用户浏览内容属性信息的基础上,根据用户浏览行为对用户兴趣项的权重进行优化,实验结果表明采用浏览行为优化得到的用户兴趣度与用户真实兴趣相似。但该算法仅仅是在小样本数据情况下的实验,在实际大数据复杂情况下的实用性和稳定性还未证实。

先前的研究仅仅是将用户的各种浏览行为等用户行为简单地组合或者只研究单个浏览行为,用于获得用户兴趣偏好。Zhao 等人^[24]提出了一种特别的行为矩阵分解协同过滤,在获取用户浏览行为之后,对用户的浏览行为通过潜在因子模型进行分解。将每种行为代表的用户兴趣偏好采用矩阵要素化技术单独建模,每种浏览行为代表用户不同的主题兴趣,再作为输入数据采用用户行为矩阵表示,用于建立用户个性化的主题兴趣档案。该算法在 Google+ 大规模数据集中进行运算,建立兴趣推荐器,实现了预测用户主题兴趣,建立用户兴趣档案的功能。而且,与普通用户行为组合方法相比,它能更加准确地预测用户个人兴趣偏好,实现了更大范围的项目覆盖。

推荐系统从提出至今虽然只有二十几年,但发展十分迅速,经历了从无到有,从简单应用到复杂应用的历程,将浏览行为推荐研究的关键阶段总结如图2所示。

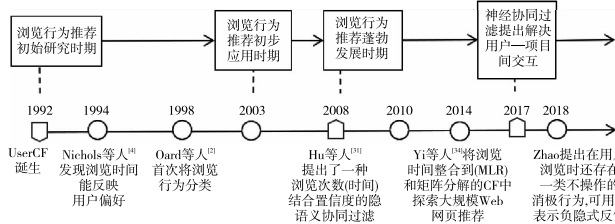


图2 基于浏览行为推荐的研究历程

Fig.2 Research history based on browsing behavior recommendation

2 基于浏览行为的推荐算法

基于浏览行为的隐式推荐引起广泛关注的主要原因是:它消除了显式评分数据收集和评估的成本,且含有丰富的用户一项目间的交互信息,内含用户的兴趣偏好^[4]。隐式反馈协同过滤推荐不仅在学术界有相当深入的研究,在工业界也有广泛的应用。协同过滤推荐的两个主要分支为邻域算法和隐语义

模型^[25],其中矩阵分解是隐语义模型中应用最多的模型^[5]。此外,近几年,深度学习的火热发展为协同过滤推荐系统原有问题的解决提供了新的思路。本章对基于浏览行为的邻域推荐算法、矩阵分解模型和深度学习推荐生成算法进行分类总结和特点对比分析,使读者细致地了解浏览行为推荐算法是如何对用户进行推荐的。

2.1 基于邻域的协同过滤推荐算法

2.1.1 基于用户的协同过滤

基于用户的协同过滤是推荐系统中经典的推荐算法^[26],与其他技术结合发展,应用场景最为广泛。核心理论是通过交互数据计算用户间的相似度,排名找出与目标用户高度相似的邻域用户,将相似用户兴趣排名最高的项目集合推荐给目标用户。Google 个性化新闻推荐系统^[27,28]是基于用户点击行为的隐式反馈协同过滤推荐的代表,用户点击一篇新闻的浏览行为,表达用户对该文章的兴趣偏好,通过二进制评分或者线性回归方程等方法量化得到用户对该文章或该类文章的兴趣度。根据上述分析可以把基于浏览行为的用户协同过滤分为三步:

- 通过对浏览行为量化得到用户对于项目的隐式兴趣评分。
- 计算用户之间的相似度,对用户进行聚类,划分兴趣相似的用户集合。
- 可将兴趣相似集合中除目标用户以外用户兴趣排名前 N ,但目标用户未与之产生交互行为的项目推荐给目标用户。

2.1.2 基于项目的协同过滤

基于项目的协同过滤与基于用户的协同过滤算法思想类似,主要应用于用户数量较多的大型网站。例如亚马逊网、Hulu 和 YouTube 的推荐系统都采用基于项目的协同过滤算法^[29]。根据用户浏览网站时间的迁移,跟踪用户实时兴趣演化,对用户兴趣模型进行实时更新,优化基于项目的协同过滤算法^[30]。如 YouTube 基于其视频网站的特点,利用用户观看次数和观看时间结合显式反馈等构造线性组合,量化用户兴趣偏好得到推荐排名,对用户进行推荐。该算法的优势在于可根据用户的历史浏览行为信息为推荐结果提供合理解释,每一个为用户推荐的项目都是有依据的,提升了用户对网站的忠诚度。根据以上描述分析,可以把基于浏览行为的项目的协同过滤分为三步:a) 根据项目聚类等方法,建立相似项目集合;b) 量化浏览行为得到用户对项目的隐式兴趣度评分;c) 根据项目相似度和用户对项目的兴趣度计算得到用户对未发生交互行为项目的兴趣度排名集合,推荐给目标用户。

2.2 基于模型的协同过滤推荐算法

基于模型的推荐是协同过滤衍生得到的一种最具发展潜力的推荐算法,其可扩展性强、精确度高。基于模型的推荐算法本质是一种机器学习方法,通过推荐训练模型,学习用户兴趣偏好进行推荐。矩阵分解模型通过各种量化计算方法将浏览行为融入推荐模型构建过程中,并利用矩阵分解模型对用户进行个性化推荐。Hu 等人^[31]采用二进制量化方法将用户产生浏览行为交互的项目评分置为 1,其他置为 0,并为其设置置信度权重,将用户浏览行为转换为 0-1 矩阵分解模型,解决了矩阵稀疏问题,提高了推荐的精度。Netflix 竞赛中提出的矩阵分解模型^[5]和加权矩阵的协分解模型^[32],将矩阵协分解结合丰富辅助信息和用户浏览行为,通过置信度对用户点击次数或其他用户浏览行为进行加权,将其整合到经典的矩阵分解模型中,构建的用户一项目矩阵分解模型优于经典的邻域推荐算法。Zheng 等人^[33]采用划分层级的模糊聚类算法获取用户兴趣聚类,得到兴趣主题分类。主要根据浏览时间联合数据挖掘中的置信度和支持等级计算用户对主题的兴趣度,再建立加权矢量模型 $U = (\text{兴趣主题}, \text{用户在主题上的兴趣度}, \text{主题的支持等级})$ 用于建模用户兴趣,结果与用户真实兴趣大部分相似,能够清楚地表示用户的偏好。

Yi 等人^[34]研究了基于浏览时间的模型推荐算法,将用户

的有效浏览时间作为反映用户兴趣偏好的一项指标,并将其整合到最新的矩阵分解模型中。相比于基于点击行为的矩阵分解模型,基于浏览时间的矩阵分解模型在克服矩阵稀疏问题的同时推荐性能更高,稳定性更强,可帮助网站获得更强的竞争力。Lyu^[35]分析了保存网页、收藏网页和网页浏览速度三种浏览行为,提出了一种新的计算用户兴趣度的方法,并将场景因子引入传统的二维向量空间模型中,将用户、主题和场景三个

因子构建为用户兴趣3D模型(user-theme-scene)。

未来,如何充分挖掘用户浏览行为数据背后深层次的兴趣偏好特征、基于矩阵分解扩展性强等特点,构建多浏览行为数据特征融合的矩阵分解模型是提升推荐系统性能的关键。

以上两类协同过滤推荐算法,通常都将用户浏览行为量化后用于表示用户的兴趣偏好,得到具体的隐式偏好评分来用于个性化推荐。两类算法的对比如表1所示。

表1 基于浏览行为的协同过滤推荐算法比较

Tab.1 Comparison of collaborative filtering recommendation algorithms based on browsing behavior

类型	浏览行为	核心理论、算法	特点	代表文献
基于邻域的推荐	点击、浏览时间	用户浏览行为偏好量化、余弦相似度、流行度惩罚等	用户浏览行为量化用于表示用户兴趣,缓解了数据稀疏问题,推荐可解释性高的同时计算复杂度也很高	文献[27~30]
基于矩阵分解模型的推荐	保存、收藏、点击次数、浏览时间、浏览速度	矩阵分解、最小二乘法、随机梯度下降、置信度加权、用户浏览行为偏好量化、聚类	矩阵分解等模型能够很好地融合显式反馈和隐式反馈的数据,解决矩阵稀疏问题的同时,通过矩阵分解降维,降低了计算复杂度,但相比邻域算法推荐可解释性较差	文献[5,31~39]

2.3 深度神经网络推荐算法

近几年,基于深度学习神经网络的推荐系统的相关研究十分热门,引起学术界众多学者投入相关方向的研究。如何采用深度学习的知识提高推荐算法的准确性是其研究的根本目的。神经网络引入推荐系统主要是通过非线性特征建模,以提高推荐结果的丰富性。目前,基于深度学习的推荐系统大多数都是基于显式反馈(例如评分)或者是对一些辅助信息(例如文本特征、图片)的学习,相对较少地探讨用户—项目隐式交互行为的核心部分(例如点击、浏览、购买等)^[40],尤其是含有较为丰富的浏览行为的隐式反馈推荐。如何对用户隐式交互行为数据进行学习是目前神经网络推荐研究的重点。

Shi等人^[41]提出的multiple feedback-based personalized ranking(MFPR)模型在原有的SVD++矩阵分解模型上,利用贝叶斯个性化排名算法(BPR)框架进行优化调整,将显式反馈与一种隐式反馈行为集成,得到单反馈个性化排名(SFPR)模型,然后扩展该模型集成更多的隐式反馈计算用户偏好,将用户的每种浏览行为用二进制评分来表示,集成用户显式反馈和多种隐式反馈,建立多隐式反馈个性化排名模型(MFPR)。虽然有针对隐式反馈评分的修正算法,提高了推荐精度,但是该模型的可解释性较差。Zheng等人^[42,43]提出的基于隐式反馈神经网络自回归模型,首先将用户点击行为转换为偏好二进

制向量和置信度向量,然后对偏好向量的概率进行神经网络建模,并由置信度向量进行加权。虽然该模型推荐效果要优于隐式矩阵分解模型(IMF),但该模型的扩展性较差,适用于特殊推荐场景。He等人^[6]在2017年提出的神经网络协同过滤(NCF)模型,标志着神经网络学习算法与推荐系统领域算法模型结合热潮的开始。学术界和工业界都开始研究通过神经网络学习用户项目的特征等辅助信息来提高推荐的精度。但对于用户—项目间的交互信息学习的研究仍很少。文献[41,44]提出的MFPR,融合显式反馈和多种隐式浏览行为集成计算得到用户偏好,利用BPR模型进行个性化排序,生成有序项目对作为标记数据用于top-N推荐,但随着矩阵维数的增加,计算的复杂度也随之增加。Wen等人^[40]在前人研究的基础上,应用矩阵分解(MF)和多层感知机(MLP)来捕获用户项目间交互复杂数据的线性和非线性关系,再通过深度神经网络(DNN)融合四种浏览行为和显式评分数据,提高了推荐质量。

总的来说,基于神经网络的用户浏览行为推荐模型,利用不同的神经网络与隐因子模型融合,学习用户与项目的隐表示。但不同类型的浏览行为数据组合,以及推荐场景等方面的差异,得到的神经网络模型也不同。表2列出了六个基于用户浏览行为的推荐模型,在所用神经网络模型、隐因子模型和应用场景,以及浏览行为类型等方面的区别。

表2 基于的浏览行为推荐模型比较

Tab.2 Comparison of recommendation models based on browsing behavior

模型	行为类型	神经网络模型	矩阵分解隐因子模型	特点
BPR-MF ^[45,46]	点击次数、购买次数	BPR	MF	采用成对排名损失方法,从单个行为中学习得到排名用于单一隐式反馈推荐场景
implicit CF-NADE ^[43]	点击	NADE	MF	应用于特殊的视频推荐场景
DeepFM ^[47]	点击、浏览时间	DNN	FM	单一点击行为神经网络学习推荐场景
NCF ^[6]	是否选择评分	MLP、DNN	GMF、NeuMF	单一浏览行为隐式反馈推荐场景
MFPR ^[41,44]	点击、点击后行为	BPR、SFPR	SVD++	多应用于点击行为或点击后浏览行为兴趣偏好权重占比高的推荐场景
MF&MLP ^[40]	点击、购买、加购、收藏	MLP、M-MLP、DNN	MF	混合多种类型丰富反馈数据的推荐场景

除此之外,随着深度学习的快速发展,近几年工业界中涌现出众多基于点击率(CTR)预估的深度学习演化模型。CTR预估模型在广告点击率预测和电子商务中拥有十分重要的地位和实用价值,CTR点击行为特征由三元组{clicked good_id, clicked shop_id, clicked category_id}表示。深度CTR模型一般是通过用户对用户点击行为数据中兴趣特征的挖掘与其他辅助信息特征融合,经由神经网络学习,拟合高阶的非线性特征后,通过一个sigmoid转换为0-1值,代表用户的点击概率。

阿里算法团队基于CTR模型提出了一个推荐系统模型:深度兴趣网络(DIN)^[48],通过观察收集到的实际线上数据,发现用户浏览网站时的兴趣十分多样,因此只有部分历史点击行为数据会影响用户的下一次点击行为,而不是所有的点击行为。而一般的CTR预估模型通常将用户所有点击行为同等对待,为了充分挖掘用户点击行为中的用户兴趣偏好和兴趣变化信息,对用户历史点击行为数据采用attention机制进行加权,

区分不同顺序的点击行为的重要性(权重)。一年后,阿里在DIN推荐模型的基础上提出了深度兴趣进化网络(DIEN)^[49],针对DIN模型中对于用户点击行为中兴趣动态演化过程捕捉的缺失,阿里提出将用户点击行为特征组成序列数据集的形式,而不是DIN模型中离散的用户点击行为特征;DIEN通过两层GRU和attention机制来建模点击行为序列特征,提取得到用户兴趣序列,捕捉相对于目标item的兴趣演化过程,与其他几类特征向量融合,通过MLP学习进行最终预测。通过观察得到用户在浏览时可能具有多个不同的兴趣,潜在兴趣可由浏览行为提取得到,因此阿里基于此问题 and 此前的研究提出了一种深度多兴趣网络(DMIN)^[50],通过用户浏览行为序列利用多头自我注意力机制对用户潜在的多个兴趣进行建模,提取多个用户兴趣,相当于一种类似于点击率预测的排序模型。

与阿里提出的一系列基于单一点击行为的深度CTR模型不同,Google、腾讯和京东则基于单一点击行为特征的CTR预

估模型,提出了多目标的 CTR 预估推荐模型。Google 提出的一种基于多目标学习模型 (MMoE) 的 YouTube 视频推荐模型^[51],以参与度和满意度两类指标为目标,其中参与度指标根据用户点击行为数据和视频观看数据分别预测用户的点击行为 (CVR) 和用户观看视频的时长 (CVR),两类指标的多种任务结合用于预测用户可能观看的下一个视频。多任务学习虽然可以通过多个目标间的协作提升预测性能,但如果多任务间结合作不当,甚至会产生冲突,产生跷跷板现象,引起负迁移,带来反向效果。为了解决此问题,腾讯提出了一种渐进分层的多任务学习个性化推荐模型:PLE^[52]。京东和百度则将多任务排序中的兴趣建模、多任务学习、偏差学习等部分融合,提出了 DMT 推荐模型^[53],对多个浏览行为序列建模,充分挖掘用户浏览行为中的用户兴趣偏好特征。表 3 是对六种最新的深度 CTR 演化模型的对比,基本都采用了深度学习的一些关键技术。除此之外,CTR 预估推荐模型中的数据往往含有一些偏差 (bias),最常见的就是位置偏差,还有一些诸如用户个人习性造成的浏览行为数据偏差等,消除浏览行为数据中的偏差是深度 CTR 推荐模型未来研究中的一个重点问题。

表 3 深度 CTR 演化模型比较
Tab. 3 Deep CTR evolution model comparison

模型	浏览行为	关键技术	序列关系
DIN ^[48]	点击	MLP、attention 机制	无
DEIN ^[49]	点击	MLP、attention_based GRU	点击序列
MRS_YT ^[51]	点击、浏览时间	MMoE、position bias	无
DMIN ^[50]	点击	MLP、multi-head self-attention	点击序列
PLE ^[52]	点击、浏览时间、分享、评论	customized gate control (CGC)、专家机制	无
DMT ^[53]	点击、加购、下单	MMoE、position bias、neighboring bias	多行为序列

2.4 混合推荐算法

协同过滤推荐算法目前已经非常成熟,但随着时代的发展,人们的需求也在增多,对推荐系统的时效性要求也逐渐提高。根据不同推荐方法的特点,互为补充混合得到的融合推荐算法相比单个推荐算法具有更高的精度和低复杂性。如在新闻推荐中最常用到将基于内容的推荐算法与协同过滤算法集成的混合推荐算法^[54~58,28],根据用户的点击行为等为每个文章生成基于内容的推荐分数 $CR(\text{article})$ 和协同过滤分数 $CF(\text{article})$,将两分数结合 $Rec(\text{article}) = CR(\text{article}) \times CF(\text{article})$,对要推荐的候选文章进行排名,推荐效果相比单协同过滤推荐算法,推荐精度提高的同时,推荐的内容也更丰富。

随着 Web 2.0 的快速发展和应用范围的扩展,推荐系统融入了更多的场景信息,如社交网络、地理位置等,造成用户项目交互相关数据噪声更高,但也为用户兴趣偏好提取提供了更多可能性。未来融合多源异构数据的推荐算法的研究将具有更加深远的现实意义。

3 用户浏览行为的量化

在个性化服务和信息技术蓬勃发展的今天,推荐系统的研究已经从简单的信息过滤转变为准确高效地为用户提供个性化服务,满足用户的个性化需求。因此,在推荐系统中建立一个能准确高效反映用户偏好的用户兴趣模型显得十分必要。根据用户的历史访问内容和用户访问时的行为分析来建立用户兴趣模型是目前研究学者们统一认同的有效方式^[58,59]。对用户访问时的浏览行为量化计算得到用户兴趣度,用来表示用户对一个项目的感兴趣程度。用户行为分为了显式和隐式两种,显式行为(如评分、调查表等)容易受噪声等不确定因素影响产生不准确或者大量重复无用的反馈数据,而隐式浏览行为既对用户没有负担,又可以获取大量相对准确多样的反馈数据。搜索研究发现,浏览行为用户推荐的主要方法是先预处理

量化,再融合其他数据挖掘或推荐模型算法进行推荐,是目前基于浏览行为推荐的主流方法。

3.1 用户兴趣度量化

用户浏览行为兴趣度客观地反映了用户的实际兴趣,属于隐式评分。用户浏览行为量化得到的隐式评分对于提高推荐算法的推荐性能有很好的提升效果和广泛的应用价值。例如面对海量的用户、商品数据,可以引入兴趣度用于提高聚类的准确性,降低推荐算法的复杂度^[60]。用户兴趣度一般用 0~1 的实数表示,根据适用场景、用户浏览行为对用户兴趣度贡献的差异等,计算方法种类繁多。

通过查阅大量最新文献,总结出用户浏览行为兴趣度的量化方法主要划分为单一浏览行为的兴趣度量化法和混合浏览行为兴趣度量化法两大类,单一浏览行为兴趣度的量化方法简单,可扩展性强,多浏览行为虽然计算复杂,但因其考虑全面,用于推荐算法后推荐性能要更好。以下是对两大类量化方法的简要概括和对比分析。

a)单一浏览行为兴趣度评分。用户浏览行为种类繁多,一些研究者认为查询、浏览、编辑等一系列浏览行为都可以通过网页浏览时间和翻页次数来间接反映^[61~63],而保存、标记书签等浏览行为则通过访问次数来体现^[61],或者仅单一根据用户访问时的浏览行为次数^[64]就能反映用户兴趣,计算复杂度也低。这种用户兴趣度量化方式为单一浏览行为兴趣度量化。

早期的隐式评分就是根据用户点击行为来构建的二进制评分,有点击行为兴趣度置为 1,无点击行为兴趣度置为 0。因其对数据类型要求单一,在更注重用户数据隐私保护的今天,该隐式评分方法仍然是推荐系统中应用最广的隐式评分方法,在基于深度学习^[6]和协同过滤^[27]的推荐中仍有用到。还有其他一些研究根据用户访问次数或浏览(页面/项目)的时间作为量化行为,通过简单的一元函数^[64]或回归函数^[65],如分析用户某段时间内访问某项目的次数与该段时间内的总访问次数关系计算得到用户对项目的兴趣度。

b)混合浏览行为兴趣度评分。采用数据挖掘方法得到用户访问日志中的浏览行为数据,区分不同浏览行为对用户兴趣度的贡献值,构建不同的浏览行为组合。排除一些高稀疏浏览行为和异常浏览行为的干扰,尽量估计获得与用户真实兴趣相符的用户兴趣度评分方式称为混合浏览行为兴趣度量化。

混合浏览行为兴趣度量化因浏览行为组合方式和量化计算方式的不同,推荐准确度也略有不同。文献[12,61,62]通过最小二乘法等方式计算出每种浏览行为的权重,构建多浏览行为的多元线性回归函数,将每种浏览行为的权重同其对应的浏览行为做内积,再求和来量化得到用户对项目的兴趣度。文献[66,67]在该方法的基础上对浏览时间和访问次数根据浏览项目内容大小进行简单计算处理之后,再与其对应权重做内积的方式,量化得到用户的兴趣度。文献[58,68,69]在单一浏览行为二进制评分的基础上,将能表达用户兴趣的浏览行为进行组合,得到用户浏览行为兴趣组合 F_i ,如果浏览页面时,一种行为发生就赋值为 1,未发生就赋值为 0,得到每种浏览行为的量化表示,并为每种行为赋予不同的权重 w_i ($\sum w = 1$)。

一些研究认为以上一些用户浏览行为兴趣度量化方法没有对一些特殊行为和干扰行为进行更加细致的分类和量化。如购买、收藏这类行为明确地表示了用户的兴趣偏好,是十分重要的浏览行为。而用户浏览时间则因为不同用户浏览速度的差异和浏览时间过长或过短等一些异常情况,导致不能进行有效的兴趣度量化,造成推荐结果不够准确,或推荐错误等情况。针对这些问题,文献[70~72]提出了采用分段函数和多元一次线性回归方程的方法来更加细致精确的量化用户的兴趣度。表 4 对浏览行为用户兴趣度量化的两大类方法在公式和量化效果等方面进行了对比分析。

表4 浏览行为量化方法对比分析
Tab.4 Comparative analysis of browsing behavior quantitative methods

行为类型	量化方法	公式	公式释义	文献	量化质量分析
二进制评分	二进制评分	$IR = \begin{cases} 1 & \text{有浏览行为} \\ 0 & \text{无浏览行为} \end{cases}$	二进制评分	[6, 27]	最简单基础的浏览行为量化方法,易代入推荐算法,但对推荐性能有一定的提升,但因含噪声多,导致推荐稳定性差
单一浏览行为	一元函数	$IR = \frac{N_i}{N_{sum}}$ $IR(x) = \frac{1}{1 + e^{-(x-\bar{x})}}$	N_i 表示访问页面 <i>i</i> 的次数或浏览时间, N_{sum} 表示该用户某时间段访问所有页面的总次数或总时间 logistic 回归函数,根据浏览次数或点击次数,预测用户对项目的兴趣度	[60,63,64,73] [65]	将单一浏览行为进行归一化,排除一部分噪声干扰,通过比重计算等将用户对每个项目的兴趣度进行区分,相比二进制评分量化,推荐稳定性有一定的提升
多元线性回归	多元线性回归	$IR = a \times T(P) + b \times V(P) + c$	a, b 为行为影响因子,根据历史行为数据,采用最小二乘法或主成分分析法估计每种浏览行为影响因素的权重	[12,61,62,66,67]	对浏览行为未处理或采用简单比值函数处理,无法精确表示用户兴趣
混合浏览行为	二值函数 & 权重	$IR_i = \sum_{v \in F} w_v \cdot f_v(di),$ ($\sum_{v=1}^F w_v = 1$)	f_v 是一个二值函数,用户对项目 <i>i</i> 有浏览行为 <i>v</i> 时函数值为1,否则为0, w_v 是每种浏览行为的权重,通过最小二乘法或熵权法估算,权重和为1	[58,68,74]	对浏览行为仅通过二值函数简单处理,与多元线性法一样计算了每种行为的权重,相比单一浏览行为的二值函数法,对推荐性能有很大的提升
分段函数	分段函数	$IR = \begin{cases} 1 & \text{if(重要浏览行为)} \\ f(w) & \text{else(其他浏览行为)} \end{cases}$	$f(w)$ 是由其他浏览行为组成的多元线性回归函数或分段函数	[70~72,75]	将用户浏览行为按重要程度分类,采用多元线性回归函数或深度学习等方法量化辅助浏览行为,提升用户兴趣评分准确度

3.2 混合浏览行为兴趣度量化模型

由第2章可知用户浏览行为是多种多样的,且每种浏览行为对于用户兴趣度的贡献程度也各不相同。经实验分析研究,多浏览行为相比单一浏览行为能构建更精确的用户兴趣模型,从而提高推荐系统的性能^[76]。如何量化不同浏览行为的贡献度即权重,用于计算用户兴趣偏好,是现在用户项目间交互数据挖掘的重点。针对混合浏览行为组合和量化方法模型,研究者提出了多种方法和模型对浏览行为进行细致量化。

李建廷等人^[72]通过浏览行为组合:操作行为组合(保存、打印、收藏、复制)、访问某一页面次数、用户在某一页面的浏览时间,分别计算出三种浏览行为的用户兴趣度,提出利用BP神经网络将三种兴趣度融合,得到用户综合兴趣度。夏义国等人^[71]采用浏览行为组合:操作行为组合(保存、打印、收藏)、访问某一页面次数、用户某一页面的浏览时间组成的最小用户兴趣浏览行为组合^[12],计算用户对项目的兴趣度。这两个研究对页面浏览时间和页面访问次数的处理方式相同:引入浏览速度代替原有的浏览时间,页面浏览率代替页面访问次数。但文献[71]将原页面与链接页面浏览时间根据页面主题相似度进行加权代替原有单一页面浏览时间,采用分段函数和多元一次线性回归函数计算用户对页面的兴趣度。因为浏览时间过短或过长都会对兴趣度计算造成干扰,文献[72]设置浏览时间阈值,尽量排除浏览时间异常的干扰。浏览行为处理优化函数如下:

$$BR(i) = \frac{\text{Freq}(i)}{\text{Freq}(s)} \quad (1)$$

其中: $BR(i)$ 为页面浏览率,统计某段时间用户访问某页面的次数与这段时间用户访问所有页面的次数的比值计算得到。

$$\text{Time}(i) = T_i + \sum_{j=1}^n \alpha_j T_j \quad (2)$$

其中: $\text{Time}(i)$ 为改进后的页面浏览时间; α_j 为链接页面的主题相似度; T_j 为点击进入链接的浏览时间。

$$\text{Speed}(i) = \frac{\text{Size}(i)}{\text{Time}(i)} \quad (3)$$

其中: $\text{Size}(i)$ 为页面*i*的大小; $\text{Time}(i)$ 为改进后浏览页面*i*的时间。

黄倩等人^[75]采用保存页面、收藏页面和页面浏览速度三种浏览行为组合,根据艾宾浩斯提出的记忆保存量与时间关系函数得到的兴趣衰减因子,得到与用户最近兴趣最匹配的兴趣度计算模型。浏览速度兴趣度公式和兴趣衰减因子函数^[77]如下:

$$\text{Speed}(i) = \frac{\text{Size}(i)}{\text{Time}(i)}, \min(T) < \text{Time}(i) < \max(T) \quad (4)$$

其中: $\text{Speed}(i)$ 为浏览速度; $\text{Time}(i)$ 为浏览时间。设置阈值排

除因浏览时间太短不感兴趣和浏览时间过长造成的异常情况。

$$K(x) = \frac{k}{(\lg T)^c + k} \quad (5)$$

邢玲等人^[70]认为当保存、打印、收藏、复制等浏览行为同时未发生时,浏览时间为衡量用户兴趣度的第二标准。当出现浏览时间过长,如用户忘记关闭页面等异常情况,再采用点击行为和拖动滚动条行为的组合来计算用户兴趣度,层层递进,保证用户兴趣度计算的准确性和评分数据足够密集。引入某页面预期阅读时间与实际阅读时间的比值*rtr*来代替浏览时间,消除不同用户浏览速度和异常浏览时间的影响。根据统计得到的一般人平均阅读速度,为*rtr*设置多重阈值,得到*f(rtr, sc, cc)*用户兴趣度分段函数如下:

$$f(rtr, sc, cc) = \begin{cases} 0 & rtr < 0.1 \\ 1 - \frac{(rtr - 0.75)^2}{0.65^2} & 0.1 \leq rtr \leq 0.75 \\ 1 & 0.75 < rtr \leq 1.25 \\ g(sc, cc) & rtr > 1.25 \end{cases} \quad (6)$$

$g(sc, cc)$ 根据浏览网页内容量*dl*,采用二元一次回归方程和分段函数得到,其函数表达式如下:

$$g(sc, cc) = \begin{cases} 1 & g_t \geq 1 \\ g_t = \frac{\alpha \cdot sc}{dl} + \frac{\beta \cdot cc}{dl} + \frac{\mu}{dl} & 0 \leq g_t < 1 \\ 0 & g_t < 0 \end{cases} \quad (7)$$

混合浏览行为组合的兴趣度量化模型对比分析如图3所示。

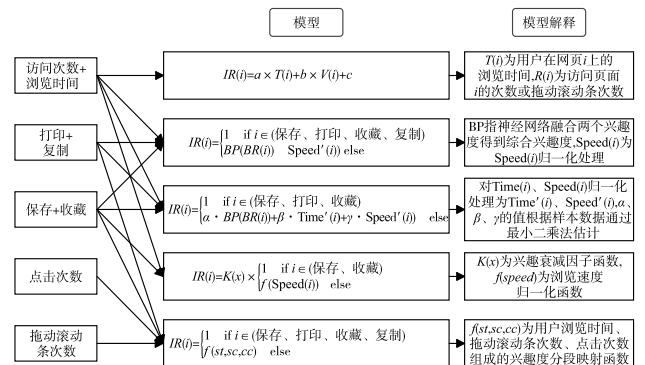


图3 混合浏览行为量化模型对比

Fig.3 Comparison of hybrid browsing behavior quantitative models

4 融合多源异构数据的推荐

互联网的高速发展为推荐系统提供了海量丰富的数据的同时,也造成了信息过载,同时数据稀疏问题仍旧存在,但也为

提高用户偏好预测精度和推荐系统性能提供了更丰富的辅助信息。如近几年,微博、B 站、YouTube 等社交网络应用的快速发展和广泛普及,越来越多的用户都或多或少地使用了其中一些社交软件。用户在社交网络上生成的交互信息多样,通过评论、收藏、分享、转发、发布帖子、视频等来实时表达对事件/主题的感受和看法的浏览行为,是能够预测用户实时兴趣偏好的有效数据来源^[76-78]。通过对多源异构数据挖掘,获取用户偏好,精准有效地建立用户兴趣模型,与推荐算法进行融合,能够有效缓解推荐系统存在的数据稀疏和冷启动问题。

4.1 用户浏览行为的上下文感知推荐

上下文信息一般指的是时间、位置、周围环境等,现有的推荐系统大多聚焦于“用户—项目”之间交互信息的挖掘,但在实际推荐场景中用户的兴趣会随时间、位置和环境等因素的变化而产生迁移,仅仅依靠“用户—项目”二维评分矩阵模型可能会导致推荐性能不稳定等问题。将上下文信息引入推荐系统,能够进一步提高推荐系统准确度和满意度,又兼具“普适计算”和“个性化”两种优势^[79]。在上下文环境下,将传统的“用户—项目”二维兴趣度预测模型扩展为包含上下文信息的多维预测模型,其结构化表示如下^[80]:

$$\begin{aligned} R &: User \times Context \times Item \rightarrow Rating \\ User &\subseteq Uname \times Age \times Address \times Profession \\ Item &\subseteq Iname \times Type \times Size \\ Context &\subseteq Location \times Time \times Environment \times Emotion \end{aligned} \quad (8)$$

用户的兴趣会随时间变化、衰减、迁移,所以时间上下文推荐是上下文感知推荐中应用最多的推荐系统。基于时间的上下文感知推荐对于用户兴趣偏好的准确获取有很重要的作用。Koren^[81]提出了一种融合时间上下文的 TimeSVD++ 算法,将时间上下文信息加入用户—项目二元矩阵中,提升了推荐的准确度和用户的满意度。涂丹丹等人^[82]基于概率矩阵分解提出了一种上下文广告推荐算法,根据访问网页次数、点击广告次数两种浏览信息和网页广告内容信息,量化得出用户对网页的兴趣度、用户对广告的兴趣度以及广告与网页的关联度,利用 logistic 函数将其三元线性组合映射到 $[0,1]$,最终得到广告点击率,用于解决数据稀疏造成的冷启动问题。

4.2 用户浏览行为的社交化网络推荐

近几年社交网络应用,如微博、知乎、豆瓣等的兴起,使社交网络与协同过滤等推荐算法融合,用于解决推荐系统冷启动问题和提高推荐性能的社交网络推荐模型,成为推荐系统研究的一个重要分支。蔡梦松等人^[83]提出了一种集成社交推荐与协同过滤的 top-N 推荐方法。首先根据二进制评分法将用户浏览行为构建为一个用户—项目的 $m \times n$ 阶评分矩阵 R ,采用协同过滤算法得到用户推荐集合 R_{CF} ;再根据社交关系数据构建用户信任集合,计算得到用户最近邻集合,以预测用户评分,得到社交网络关系的用户推荐集合 R_{SN} 。将两个推荐集合融合,即得到最终 top-N 推荐集合 $R = R_{CF} + R_{SN} - R_{CF} \cap R_{SN}$ 。李卫平等人^[84]将社交关系数据融入到用户评分矩阵,提出了一种矩阵分解社会推荐模型。根据社交关系网络图构建用户社交网络矩阵 $S_{n \times n}$ (n 为用户数),再与用户综合评分矩阵 $R_{n \times m}$ (n 为用户数, m 为项目数)合并为一个大的矩阵 $M = [S \ R]_{n \times (n+m)}$,最后采用随机梯度矩阵分解对该矩阵降维的推荐算法得出推荐集合。Yang 等人^[85,86]采用直接融合社交关系数据与用户浏览行为隐式评分数据的方式,提出了一种矩阵分解的社会化信任推荐模型 TrustPMF,将用户浏览行为偏好评分数据矩阵与用户社交信任数据矩阵融合为一个矩阵。将社交关系数据融入到矩阵分解模型中,提高推荐可靠性的同时,相比经典的矩阵分解模型推荐过程的可解释性更强。

随着大数据和人工智能时代的到来,传统基于显式评分和简单隐式评分的推荐系统正逐渐被融合多源异构数据的推荐系统所取代,多元化社会对个性化服务的要求更高。虽然前人

已经做了许多工作,但其仍面临从浏览行为中提取负反馈信息困难、用户隐私保护、用户兴趣动态更新和多源异构数据融合困难等问题。对于用户浏览行为这一明确反映用户兴趣偏好的数据处理,需要结合时间、地理位置、社交网络等情景信息,进行更加细粒度的量化处理,构建更加精确的用户兴趣模型,提高推荐系统的精度和推荐效果,仍是未来研究的重点方向。

5 问题及未来研究方向

用户与项目间的交互产生的用户浏览行为数据作为个性化推荐的重要依托,在当前互联网快速发展,信息严重过载的时代,社交网络、人工智能及数据挖掘等领域知识的引进,推动了基于浏览行为个性化推荐系统的进一步研究和发展,并取得了一定的成果,但交叉领域的融合仍存在许多问题和挑战亟需解决,主要有以下几方面:

a) 深度学习推荐算法性能问题。近几年,深度学习在推荐系统领域的应用成为了研究的热点,相关算法和模型层出不穷,但其研究百花齐放的背后,面临的实际情况是深度学习推荐算法的可复现性低的问题,模型有效性也广受质疑。推荐系统国际顶会 RecSys2019 最佳论文^[87]对近些年推荐系统领域相关顶会 KDD、RecSys、SIGIR、WWW 中深度学习推荐领域的 top-N 推荐算法进行复现分析,18 种推荐算法中只有 7 种能够有效复现,并采用 HR、NDCG 和 REC 等评价方法在不同 top-N 推荐需求下进行对比分析,结果显示,某些模型的推荐效果甚至低于最基本的邻域的推荐算法(KNN),相比传统优质模型,基于深度学习的推荐模型仍面临训练时间过长、难以复现、可移植性和应用性差等问题。

大多数深度学习推荐模型是对产生交互的用户和项目本身属性数据和社交网络、地理位置等辅助信息建模,很少涉及对浏览行为数据的建模,只是将用户浏览行为数据简单处理之后用于深度学习模型中,而对其交互行为背后的兴趣意图的学习很少,或只是对单一浏览行为(点击)背后意图的学习。而且,深度学习推荐算法相较经典的协同过滤算法有一个致命的问题:可解释性差。在小样本数据集中此问题可能还不显著,但在实际应用场景中用户的数量能达到亿级以上,推荐系统的可解释性对用户满意度和忠诚度的提升都很重要,提升深度学习推荐模型的可解释性,改进深度学习推荐算法的性能,使其能真正地应用于实际推荐系统中才是个性化推荐服务研究的本质,也是未来个性化深度学习推荐研究的目标之一。

b) 多源异构数据融合模型的建立。用户浏览行为背后的意图是相当复杂的,与推荐系统本身、环境、社交网络和时间上下文都有很大的关系。多元化的社交网络下,影响用户项目交互的因素也是多种多样的,现今的推荐模型或算法大多是针对一种具体情境下的推荐进行研究应用,如针对推荐系统本身的分辨用户推荐点击意图和真实兴趣点击意图;融合社交关系影响因子和信任度构建用户兴趣模型的推荐^[88];考虑用户项目交互时效性和用户兴趣转移的时间上下文推荐^[75,77,79,89],或是基于乘法原则将几种上下文信息类型的兴趣度简单融合的混合推荐^[90,91]。其原因主要有:多源异构数据的获取相对于传统推荐系统中单一领域数据的获取要复杂,尤其是跨媒体平台数据的获取,虽然可以改善数据稀疏问题,但其复杂程度也很高,也涉及到用户相关隐私问题;多源异构数据融合困难,现有的推荐模型大多为针对单一领域特定类型的数据的处理,对于跨领域多平台类型数据的融合十分困难。虽然目前已有 Google 和微软等公司在实际推荐系统中尝试通过深度学习等技术融合多源异构数据,但构建有效可行的多源异构数据融合模型仍处于初步研究阶段,未来一段时间仍是工业界和学术界研究的重点方向之一。

c) 浏览行为量化的复杂度与效率。浏览行为类型多样,在用户兴趣建模方面扮演着关键性角色,如广告推荐中最重要

的CTR预估模型,是对用户历史点击行为数据进行挖掘分析,预测用户下一次点击行为,提高广告的点击率;强化用户点击后的一系列浏览行为,将其作为表达用户兴趣偏好的指标高性能可用数据的新闻文章推荐算法,相比单一点击行为或基于点击后阅读时间的推荐系统有一定的提升,但对用户浏览行为数据进一步的挖掘研究还有很长的路要走。应用场景等的不同,使得每种浏览行为对用户兴趣偏好的贡献程度也不同,现有的推荐算法为了降低计算的复杂度,大都将用户浏览行为同质化为单一的选择行为,如基于神经网络协同过滤算法、图卷积神经网络等,都将用户浏览行为同质化为二进制选择行为^[6],未考虑用户浏览行为的多样性和贡献度的不同。虽然已经提出了相关浏览行为量化方法,但因其受多种因素影响,无固定量化模型和标准,导致其计算复杂度高、可扩展性差、运算效率低等问题。因此,在不同推荐应用场景和动态变化下快速提取和修正用户偏好、建立统一的用户浏览行为兴趣偏好量化模型,是下一步用户交互行为数据挖掘研究的重点。

d) 注意力机制和会话机制用于推荐。注意力机制与深度神经网络的结合主要应用于计算机视觉和自然语言领域,都取得了巨大的成果。针对深度学习推荐算法对于用户历史行为数据信息挖掘缺乏的问题,将注意力机制引入深度学习模型中,如阿里巴巴公司提出的DIN模型^[48],在深度神经网络中加入注意力机制学习用户历史行为,考虑不同时段用户浏览行为兴趣偏好的权重,解决了多用户浏览行为兴趣表示困难的问题;在此基础上又提出可建模用户点击行为序列兴趣的改进兴趣演化模型DIEN^[49],动态更新用户的兴趣。现代在线会话服务的应用,如淘宝客服、视频推荐等,需要根据用户的实时需求及时提供个性化服务,这就是基于会话的推荐。基于会话的推荐目前还处于初始发展阶段,大多是根据用户的实时点击数据构建的基于递归神经网络(RNN)的推荐模型,应用领域多为在线服务领域。推荐系统国际顶会举办的RecSys 2019挑战赛专注于在线旅游问题,由全球酒店搜索平台trivago提供了有关匿名数据集,该数据集涉及2018年11月一周内访问trivago网站的超过70万用户的近1600万会话交互,要求基于用户会话交互建立了点击预测模型,预测用户住宿推荐列表,实际推荐点击率最高的获胜^[15]。但是,目前相关领域的研究还很少,对于用户快速获取推荐反馈、推荐效率提升等问题还有许多问题值得研究。

e) 用户浏览行为数据偏差。用户在浏览网站时产生的浏览行为数据种类多样,Chen等人^[92]在关于偏差(bias)的研究中,认为用户浏览行为数据是观察性的,而不是实验性的,这使得数据中广泛存在各种偏差。其中最常见的偏差为曝光bias和位置bias,因为用户倾向于点击暴露在推荐页面的项目,推荐系统根据点击行为,又推荐更多的与点击项目相似的项目,造成循环反馈,影响了用户的行为,导致用户看到的内容丰富度和用户满意度下降;位置bias则是用户倾向去浏览点击或购买排序靠前的项目,对于排序靠后的项目则很少点击。除此之外用户浏览时还有可能会因为误点、时间等因素造成bias。如果对这些偏差处理不当,会大大影响推荐系统的长期性能。如何排除用户浏览行为数据中的多种偏差,使用户浏览行为特征与用户本身兴趣更贴合,是基于浏览行为推荐系统中必须考虑的难点部分之一。

f) 多浏览行为序列推荐。近两年,在工业界随着硬件技术的突破和深度学习技术的引入,基于用户行为序列的用户兴趣演化网络建模的研究增多,其中大多数研究为基于单一历史点击行为序列的兴趣演化推荐算法,对推荐系统的性能有很大的提升。但是京东和百度^[53]发表在CIKM 2020会议上的最新研究表明,多浏览行为序列如果结合得当,其推荐性能要高于单一点击行为序列推荐。而且,用户行为序列推荐对于用户动态兴趣的捕捉能力较强,如果结合会话机制等技术,在快速响应

用户需求方面对推荐系统性能的提升会更高。因此,为充分挖掘利用浏览行为数据中的隐含信息,选择合适的多个浏览行为构建多浏览行为序列推荐模型,为提升推荐系统响应速度和准确性又提供了一条新思路,十分值得进一步研究。

6 结束语

在信息和通信技术高速发展的时代,推荐系统作为解决信息过载问题的重要方法之一,一直是学术界和工业界研究的热点。当今多元化社交网络和互联网用户的爆炸式增长,为用户提供个性化服务,提高用户满意度是一项很大的挑战。用户浏览行为作为最能反映用户兴趣偏好的数据,其挖掘应用方式一直是推荐系统研究的热点方向之一。本文对浏览行为推荐进行概念梳理,对相关推荐算法和浏览行为量化方法进行了对比分析,回顾总结了目前融合辅助多源异构数据的浏览行为推荐的研究现状,并对未来该领域的难点和发展趋势进行了讨论分析。希望能为基于用户浏览行为的个性化推荐研究领域的研究学者提供有益的帮助。

参考文献:

- [1] Hill W, Terveen L. Using frequency-of-mention in public conversations for social filtering[C]//Proc of ACM Conference on Computer Supported Cooperative Work. New York: ACM Press, 1996: 106-112.
- [2] Oard D W, Marchionini G. A conceptual framework for text filtering[EB/OL]. 1996. <https://core.ac.uk/display/100121768>.
- [3] Konstan J A, Miller B N, Maltz D, et al. GroupLens: applying collaborative filtering to UseNet news[J]. *Communications of the ACM*, 1997, 40(3): 77-87.
- [4] Nichols D M. Implicit rating and filtering[C]//Proc of the 5th DELOS Workshop on Filtering and Collaborative Filtering. 1997: 10-12.
- [5] Koren Y, Bell R, Volinsky C. Matrix factorization techniques for recommender systems[J]. *Computer*, 2009, 42(8): 30-37.
- [6] He Xiangnan, Liao Lizi, Zhang Hanwang, et al. Neural collaborative filtering[C]//Proc of the 26th International Conference on World Wide Web. Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee, 2017: 173-182.
- [7] Oard D W, Kim J. Implicit feedback for recommender systems[C]//Proc of AAAI Workshop on Recommender Systems. Palo Alto, CA: AAAI Press, 1998: 83.
- [8] Seo Y W, Zhang B T. Learning user's preferences by analyzing Web-browsing behaviors[C]//Proc of the 4th International Conference on Autonomous Agents. New York: ACM Press, 2000: 381-387.
- [9] Morita M, Shinoda Y. Information filtering based on user behavior analysis and best match text retrieval[C]//Proc of the 17th Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval. Berlin: Springer-Verlag, 1994: 272-281.
- [10] Sakagami H, Kamba T. Learning personal preferences on online newspaper articles from user behaviors[J]. *Computer Networks and ISDN Systems*, 1997, 29(8-13): 1447-1455.
- [11] Kelly D, Teevan J. Implicit feedback for inferring user preference: a bibliography[J]. *ACM SIGIR Forum*, 2003, 37(2): 18-28.
- [12] 尹春晖, 邓伟. 基于用户浏览行为分析的用户兴趣获取[J]. 计算机技术与发展, 2008, 28(5): 37-39. (Yin Chunhui, Deng Wei. Extracting user interests based on analysis of user behaviors[J]. *Computer Technology and Development*, 2008, 28(5): 37-39.)
- [13] Proc of the Workshop on ACM Recommender Systems Challenge (RecSys Challenge'17)[C]. New York: ACM Press, 2017.
- [14] Proc of the Workshop on ACM Recommender Systems Challenge (RecSys Challenge'18)[C]. New York: ACM Press, 2018.
- [15] Proc of the Workshop on ACM Recommender Systems Challenge (RecSys Challenge'19)[C]. New York: ACM Press, 2019.
- [16] Proc of the Workshop on ACM Recommender Systems Challenge (RecSys Challenge'20)[C]. New York: ACM Press, 2020.
- [17] Karlgren J. Newsgroup clustering based on user behavior—a recommendation algebra[R]. [S. l.]: SICS, 1994.
- [18] Stevens C. Knowledge-based assistance for accessing large, poorly structured information spaces [D]. Denver: University of Colorado, 1993.
- [19] Liang Tingpeng, Lai H J. Discovering user interests from web browsing behavior: an application to internet news services [C]//Proc of

- the 35th Annual Hawaii International Conference on System Sciences. Piscataway, NJ: IEEE Press, 2002: 2718-2727.
- [20] Kelly D, Belkin N J. Reading time, scrolling and interaction: exploring implicit sources of user preferences for relevance feedback [C]//Proc of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM Press, 2001: 408-409.
 - [21] Claypool M, Le P, Wased M, *et al.* Implicit interest indicators [C]//Proc of the 6th International Conference on Intelligent User Interfaces. New York: ACM Press, 2001: 33-40.
 - [22] Xia Minjie, Zhang Jing. Research on personalized recommendation system for e-commerce based on Web log mining and user browsing behaviors [C]//Proc of International Conference on Computer Application and System Modeling. Piscataway, NJ: IEEE Press, 2010: 402-411.
 - [23] Xing Kun, Zhang Bofeng, Zhou Bo, *et al.* Behavior based user interests extraction algorithm [C]//Proc of International Conference on Internet of Things and the 4th International Conference on Cyber, Physical and Social Computing. Piscataway, NJ: IEEE Press, 2011: 448-452.
 - [24] Zhao Zhe, Cheng Zhiyuan, Hong Lichan, *et al.* Improving user topic interest profiles by behavior factorization [C]//Proc of the 24th International Conference on World Wide Web. New York: ACM Press, 2015: 1406-1416.
 - [25] Hofmann T. Latent semantic models for collaborative filtering[J]. *ACM Trans on Information Systems*, 2004, 22(1): 89-115.
 - [26] Goldberg D, Nichols D, Oki B M, *et al.* Using collaborative filtering to weave an information tapestry[J]. *Communications of the ACM*, 1992, 35(12): 61-70.
 - [27] Das A S, Datar M, Garg A, *et al.* Google news personalization: scalable online collaborative filtering[C]//Proc of the 16th International Conference on World Wide Web. New York: ACM Press, 2007: 271-280.
 - [28] Liu Jiahui, Dolan P, Pedersen E R. Personalized news recommendation based on click behavior[C]//Proc of the 15th International Conference on Intelligent User Interfaces. New York: ACM Press, 2010: 31-40.
 - [29] Liu N N, Zhao Min, Xiang E, *et al.* Online evolutionary collaborative filtering[C]//Proc of the 4th ACM Conference on Recommender Systems. New York: ACM Press, 2010: 95-102.
 - [30] Davidson J, Liebald B, Liu Junning, *et al.* The YouTube video recommendation system[C]//Proc of the 4th ACM Conference on Recommender Systems. New York: ACM Press, 2010: 293-296.
 - [31] Hu Yifan, Koren Y, Volinsky C. Collaborative filtering for implicit feedback datasets[C]//Proc of the 8th IEEE International Conference on Data Mining. Piscataway, NJ: IEEE Press, 2008: 263-272.
 - [32] Fang Yi, Si Luo. Matrix co-factorization for recommendation with rich side information and implicit feedback [C]//Proc of the 2nd International Workshop on Information Heterogeneity and Fusion in Recommender Systems. New York: ACM Press, 2011: 65-69.
 - [33] Zheng Ling, Cui Shuo, Dong Yue, *et al.* User interest modeling based on browsing behavior [C]//Proc of the 3rd International Conference on Advanced Computer Theory and Engineering. Piscataway, NJ: IEEE Press, 2010: 455-458.
 - [34] Yi Xing, Hong Liangjie, Zhong Erheng, *et al.* Beyond clicks: dwell time for personalization [C]//Proc of the 8th ACM Conference on Recommender Systems. New York: ACM Press, 2014: 113-120.
 - [35] Lyu Jian. User interest degree evaluation models [C]//Proc of the 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery. Piscataway, NJ: IEEE Press, 2017: 2855-2859.
 - [36] Pilászy I, Zibriczky D, Tikk D. Fast als-based matrix factorization for explicit and implicit feedback datasets [C]//Proc of the 4th ACM Conference on Recommender Systems. New York: ACM Press, 2010: 71-78.
 - [37] Yang Qinghong, Hu Hao, Xiong Neng. The research on user interest model based on quantization browsing behavior [C]//Proc of the 7th International Conference on Computer Science & Education. Piscataway, NJ: IEEE Press, 2012: 50-54.
 - [38] 王智圣, 李琪, 汪静, 等. 基于隐式用户反馈数据流的实时个性化推荐[J]. *计算机学报*, 2016, 39(1): 52-64. (Wang Zhisheng, Li Qi, Wang Jing, *et al.* Real-time personalized recommendation based on implicit user feedback data stream[J]. *Chinese Journal of Computers*, 2016, 39(1): 52-64.
 - [39] Koren Y. Factorization meets the neighborhood: a multifaceted collaborative filtering model [C]//Proc of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2008: 426-434.
 - [40] Wen Hongfa, Liu Xin, Yan Chenggang, *et al.* Leveraging multiple implicit feedback for personalized recommendation with neural network [C]//Proc of International Conference on Artificial Intelligence and Advanced Manufacturing. New York: ACM Press, 2019: 1-6.
 - [41] Shi Chuan, Liu Jian, Zhang Yiding, *et al.* MFPR: a personalized ranking recommendation with multiple feedback[J]. *ACM Trans on Social Computing*, 2018, 1(2): 1-22.
 - [42] Zheng Yin, Tang Bangsheng, Ding Wenkui, *et al.* A neural autoregressive approach to collaborative filtering[C]//Proc of the 33rd International Conference on International Conference on Machine Learning. [S. l.]: JMLR Press, 2016: 764-773.
 - [43] Zheng Yin, Liu Cailiang, Tang Bangsheng, *et al.* Neural autoregressive collaborative filtering for implicit feedback [C]//Proc of the 1st Workshop on Deep Learning for Recommender Systems. New York: ACM Press, 2016: 2-6.
 - [44] Liu Jian, Shi Chuan, Hu Binbin, *et al.* Personalized ranking recommendation via integrating multiple feedbacks [C]//Proc of Pacific-Asia Conference on Knowledge Discovery and Data Mining. Berlin: Springer, 2017: 131-143.
 - [45] Rendle S, Freudenthaler C, Gantner Z, *et al.* BPR: Bayesian personalized ranking from implicit feedback [C]//Proc of the 25th Conference on Uncertainty in Artificial Intelligence. 2009: 452-461.
 - [46] Lerche L, Jannach D. Using graded implicit feedback for Bayesian personalized ranking [C]//Proc of the 8th ACM Conference on Recommender Systems. New York: ACM Press, 2014: 353-356.
 - [47] Guo Huifeng, Tang Ruiming, Ye Yunming, *et al.* DeepFM: a factorization-machine based neural network for CTR prediction [C]//Proc of the 26th International Joint Conference on Artificial Intelligence. Palo Alto, CA: AAAI Press, 2017: 1725-1731.
 - [48] Zhou Guorui, Zhu Xiaoqiang, Song Chenru, *et al.* Deep interest network for click-through rate prediction [C]//Proc of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM Press, 2018: 1059-1068.
 - [49] Zhou Guorui, Mou Na, Fan Ying, *et al.* Deep interest evolution network for click-through rate prediction [C]//Proc of AAAI Conference on Artificial Intelligence. Palo Alto, CA: AAAI Press, 2019: 5941-5948.
 - [50] Xiao Zhibo, Yang Luwei, Jiang Wen, *et al.* Deep multi-interest network for click-through rate prediction [C]//Proc of the 29th ACM International Conference on Information & Knowledge Management. New York: ACM Press, 2020: 2265-2268.
 - [51] Zhao Zhe, Hong Lichan, Wei Li, *et al.* Recommending what video to watch next: a multitask ranking system [C]//Proc of the 13th ACM Conference on Recommender Systems. New York: ACM Press, 2019: 43-51.
 - [52] Tang Hongyan, Liu Junning, Zhao Ming, *et al.* Progressive layered extraction (PLE): a novel multi-task learning (MTL) model for personalized recommendations [C]//Proc of the 14th ACM Conference on Recommender Systems. New York: ACM Press, 2020: 269-278.
 - [53] Gu Yulong, Ding Zhuoye, Wang Shuaiqiang, *et al.* Deep multifaceted transformers for multi-objective ranking in large-scale e-commerce recommender systems [C]//Proc of the 29th ACM International Conference on Information & Knowledge Management. New York: ACM Press, 2020: 2493-2500.
 - [54] Miranda T, Claypool M, Gokhale A, *et al.* Combining content-based and collaborative filters in an online newspaper [C]//Proc of ACM SIGIR Workshop on Recommender Systems. New York: ACM Press, 1999.
 - [55] Carreira R, Crato J M, Gonçalves D, *et al.* Evaluating adaptive user profiles for news classification [C]//Proc of the 9th International Conference on Intelligent User Interfaces. New York: ACM Press, 2004: 206-212.
 - [56] Das A S, Datar M, Garg A, *et al.* Google news personalization: scalable online collaborative filtering [C]//Proc of the 16th International Conference on World Wide Web. New York: ACM Press, 2007: 271-280.
 - [57] Popescul A, Ungar L H, Pennock D M, *et al.* Probabilistic models for unified collaborative and content-based recommendation in sparse-data environments [C]//Proc of the 17th Conference in Uncertainty in Artificial Intelligence. San Francisco, CA: Morgan Kaufmann, 2001: 437-444.
 - [58] 张玉连, 王权. 基于浏览行为和浏览内容的用户兴趣建模[J]. *现代图书情报技术*, 2007(6): 52-55. (Zhang Yulian, Wang Quan. User profile mining of combining Web behavior and content analysis [J]. *Modern Library and Information Technology*, 2007(6): 52-55.)
 - [59] 徐科, 崔志明. 基于搜索历史的用户兴趣模型的研究[J]. *计算机*

- 技术与发展,2006,16(5):18-20. (Xu Ke, Cui Zhiming. User profile model based on user search histories[J]. *Computer Technology and Development*,2006,16(5):18-20.)
- [60] 崔春生,吴新宗,王莹. 用于推荐系统聚类分析的用户兴趣度研究[J]. 计算机工程与应用,2011,47(7):226-228. (Cui Chunsheng, Wu Qizong, Wang Ying. Study on user interest level for clustering analysis in recommender systems[J]. *Computer Engineering and Applications*,2011,47(7):226-228.)
- [61] 付关友,朱征宇. 个性化服务中基于行为分析的用户兴趣建模[J]. 计算机工程与科学,2005,27(12):80-82. (Fu Guanyou, Zhu Zhengyu. A user interest model based on the analysis of user behaviors for personalization[J]. *Computer Engineering & Science*,2005,27(12):80-82.)
- [62] 赵银春,付关友,朱征宇. 基于Web浏览内容和行为相结合的用户兴趣挖掘[J]. 计算机工程,2005,31(12):93-94,198. (Zhao Yinchun, Fu Guanyou, Zhu Zhengyu. User interest mining of combining Web content and behavior analysis[J]. *Computer Engineering*,2005,31(12):93-94,198.)
- [63] 陈冬林,聂规划. 基于商品属性隐性评分的协同过滤算法研究[J]. 计算机应用,2006,26(4):966-968. (Chen Donglin, Nie Guihua. Research on collaborative filtering algorithm based on item's attribute implicit rating[J]. *Journal of Computer Applications*,2006,26(4):966-968.)
- [64] 秦光洁,张颖. 基于综合兴趣度的协同过滤推荐算法[J]. 计算机工程,2009,35(17):81-83. (Qin Guangjie, Zhang Ying. Collaborative filtering recommendation algorithm based on comprehensive interest measure[J]. *Computer Engineering*,2009,35(17):81-83.)
- [65] 黄贤英,龙姝言,谢晋. 结合用户兴趣度聚类的协同过滤推荐算法[J]. 计算机应用研究,2019,36(9):2609-2612,2617. (Huang Xianying, Long Shuyan, Xie Jin. Collaborative filtering recommendation algorithm combined with user interest degree clustering[J]. *Application Research of Computers*,2019,36(9):2609-2612,2617.)
- [66] 葛桂丽,袁凌云,王兴起. 基于情境感知的用户个性化兴趣建模[J]. 计算机应用研究,2017,34(4):995-999. (Ge Guili, Yuan Lingyun, Wang Xingchao. Personalized user interest modeling based on context aware[J]. *Application Research of Computers*,2017,34(4):995-999.)
- [67] 宋泊东,张立臣. 基于改进协同过滤算法的用户页面兴趣度预测研究[J]. 计算机应用研究,2019,36(11):3266-3268. (Song Bodong, Zhang Lichen. Prediction of user page interest based on improved collaborative filtering algorithm[J]. *Application Research of Computers*,2019,36(11):3266-3268.)
- [68] 王微微,夏峰涛,李晓明. 一种基于用户行为的兴趣度模型[J]. 计算机工程与应用,2012,48(8):148-151,199. (Wang Weiwei, Xia Xiufeng, Li Xiaoming. Personal interest degree model based on consumer behavior[J]. *Computer Engineering and Applications*,2012,48(8):148-151,199.)
- [69] 屈娟娟. 大数据网络用户浏览隐式反馈信息检索仿真[J]. 计算机仿真,2019,36(9):430-433,468. (Qu Juanjuan. Big data network user browsing implicit feedback information retrieval simulation[J]. *Computer Simulation*,2019,36(9):430-433,468.)
- [70] 邢玲,宋章浩,马强. 基于混合行为兴趣度的用户兴趣模型[J]. 计算机应用研究,2016,33(3):661-664,668. (Xing Ling, Song Zhanghao, Ma Qiang. User interest model based on hybrid behaviors interest rate[J]. *Application Research of Computers*,2016,33(3):661-664,668.)
- [71] 夏义国,刘友华. 一种用户兴趣度计算与用户兴趣修正的改进方法[J]. 现代情报,2014,34(1):46-48,55. (Xia Yiguo, Liu Youhua. An improved method to calculate user's interest degree and amend user's interest[J]. *Journal of Modern Information*,2014,34(1):46-48,55.)
- [72] 李建廷,郭晔,汤志军. 基于用户浏览行为分析的用户兴趣度计算[J]. 计算机工程与设计,2012,33(3):968-972. (Li Jianting, Guo Ye, Tang Zhijun. User interest degree calculating based on analysis users' browsing behaviors[J]. *Computer Engineering and Design*,2012,33(3):968-972.)
- [73] 刘枚莲,刘同存,李小龙. 基于用户兴趣特征提取的推荐算法研究[J]. 计算机应用研究,2011,28(5):1664-1667. (Liu Meilian, Liu Tongcun, Li Xiaolong. Recommendation algorithm on feature extraction based on user interests[J]. *Application Research of Computers*,2011,28(5):1664-1667.)
- [74] 崔春生. 基于隐式浏览输入的用户聚类分析[J]. 计算机应用研究,2011,28(8):2862-2864. (Cui Chunsheng. User clustering analysis based on implicit navigation[J]. *Application Research of Computers*,2011,28(8):2862-2864.)
- [75] 黄倩,谢颖华. 一种基于网页浏览行为的用户兴趣度计算方法[J]. 信息技术,2015(5):184-186,191. (Huang Qian, Xie Yinghua. Method of computing for user interest degree based on user's browsing behaviors[J]. *Information Technology*,2015(5):184-186,191.)
- [76] Zhao Hua, Zou Ruofei, Duan Hua, et al. An online paper recommendation system driven by user's interest model and user group[C]//Proc of the 4th International Conference on Communication and Information Processing. New York:ACM Press,2018:141-144.
- [77] 孙海超. 基于自然遗忘的个性化推荐算法研究[D]. 长春:东北师范大学,2011. (Sun Haichao. Research on personalized recommendation algorithm based on natural forgetting[D]. Changchun:Northeast Normal University,2011.)
- [78] Zarrinkalam F, Fani H, Bagheri E. Social user interest mining: methods and applications[C]//Proc of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York:ACM Press,2019:3235-3236.
- [79] 王立才,孟祥武,张玉洁. 上下文感知推荐系统[J]. 软件学报,2012,23(1):1-20. (Wang Licai, Meng Xiangwu, Zhang Yujie. Context-aware recommender systems[J]. *Journal of Software*,2012,23(1):1-20.)
- [80] Adomavicius G, Tuzhilin A. Context-aware recommender systems[M]//Recommender Systems Handbook. Boston,MA:Springer,2011:217-253.
- [81] Koren Y. Collaborative filtering with temporal dynamics[C]//Proc of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York:ACM Press,2009:447-456.
- [82] 涂丹丹,舒承椿,余海燕. 基于联合概率矩阵分解的上下文广告推荐算法[J]. 软件学报,2013,24(3):454-464. (Tu Dandan, Shu Chengchun, Yu Haiyan. Using unified probabilistic matrix factorization for contextual advertisement recommendation[J]. *Journal of Software*,2013,24(3):454-464.)
- [83] 蔡孟松,李学明,尹衍腾. 基于社交用户标签的混合 top-N 推荐方法[J]. 计算机应用研究,2013,30(5):1309-1311,1344. (Cai Mengsong, Li Xueming, Yin Yanteng. Hybrid top-N recommendation method based on social user tag[J]. *Application Research of Computers*,2013,30(5):1309-1311,1344.)
- [84] 李卫平,杨杰. 基于随机梯度矩阵分解的社会网络推荐算法[J]. 计算机应用研究,2014,31(6):1654-1656,1664. (Li Weiping, Yang Jie. Stochastic gradient matrix factorization based on social network recommender algorithm[J]. *Application Research of Computers*,2014,31(6):1654-1656,1664.)
- [85] Yang Bo, Lei Yu, Liu Jiming, et al. Social collaborative filtering by trust[C]//Proc of the 23rd International Joint Conference on Artificial Intelligence. Piscataway,NJ:IEEE Press,2013:2747-2753.
- [86] Yang Bo, Lei Yu, Liu Jiming, et al. Social collaborative filtering by trust[J]. *IEEE Trans on Pattern Analysis and Machine Intelligence*,2016,39(8):1633-1647.
- [87] Dacrema M F, Cremonesi P, Jannach D. Are we really making much progress? A worrying analysis of recent neural recommendation approaches[C]//Proc of the 13th ACM Conference on Recommender Systems. New York:ACM Press,2019:101-109.
- [88] 陈婷,朱青,周梦溪,等. 社交网络环境下基于信任的推荐算法[J]. 软件学报,2017,28(3):721-731. (Chen Ting, Zhu Qing, Zhou Mengxi, et al. Trust-based recommendation algorithm in social network[J]. *Journal of Software*,2017,28(3):721-731.)
- [89] Hui Kangle, Hou Hong, Xue Siyu. An improved collaborative filtering algorithm based on user interest diffusion and time correlation[C]//Proc of the 5th International Conference on Mathematics and Artificial Intelligence. New York:ACM Press,2020:38-42.
- [90] 彭宏伟,靳远远,吕晓强,等. 一种基于矩阵分解的上下文感知 POI 推荐算法[J]. 计算机学报,2019,42(8):1797-1811. (Peng Hongwei, Jin Yuanyuan, Lyu Xiaoliang, et al. Context-aware POI recommendation based on matrix factorization[J]. *Chinese Journal of Computers*,2019,42(8):1797-1811.)
- [91] 任星怡,宋美娜,宋俊德. 基于位置社交网络的上下文感知的兴趣点推荐[J]. 计算机学报,2017,40(4):824-841. (Ren Xingyi, Song Meina, Song Junde. Context-aware point-of-interest recommendation in location-based social networks[J]. *Chinese Journal of Computers*,2017,40(4):824-841.)
- [92] Chen Jiawei, Dong Hande, Wang Xiang, et al. Bias and debias in recommender system: a survey and future directions[EB/OL]. (2020-10-07). <https://arxiv.org/abs/2010.03240v1>.