

移动新闻推荐技术及其应用研究综述

孟祥武 陈 诚 张玉洁

(北京邮电大学智能通信软件与多媒体北京市重点实验室 北京 100876)
(北京邮电大学计算机学院 北京 100876)

摘 要 随着移动设备的普及和新闻读者群体人数的不断增加,移动新闻推荐已经成为移动推荐领域的热点之一.如何根据移动设备和新闻特点进行移动新闻推荐,以提高推荐性能和用户满意度,成为移动新闻推荐系统的主要任务.文中概括分析了移动新闻推荐的研究现状,并指出其与传统新闻推荐、其他移动推荐之间的区别.从新闻表示方法,移动用户新闻偏好获取,上下文感知的移动新闻推荐技术,基于社会化网络的移动新闻推荐技术,移动新闻展示以及典型应用等六个关键方面,对移动新闻推荐领域的最新研究成果进行了详细的比较、分析和总结.文中还从重点、难点两个方面讨论分析了移动新闻推荐系统面临的挑战,并指出为了进一步发展移动新闻推荐系统,未来还需要在数据集获取,效用评估,结合移动社会化网络,安全问题,形式化描述等方面深入开展一些研究工作.

关键词 个性化;上下文感知;移动社交网络;移动新闻推荐;社交媒体;数据挖掘

中图法分类号 TP18 **DOI 号** 10.11897/SP.J.1016.2016.00685

A Survey of Mobile News Recommend Techniques and Applications

MENG Xiang-Wu CHEN Cheng ZHANG Yu-Jie

(Beijing Key Laboratory of Intelligent Telecommunications Software and Multimedia,
Beijing University of Posts and Telecommunications, Beijing 100876)

(School of Computer Science, Beijing University of Posts and Telecommunications, Beijing 100876)

Abstract With the popularity of mobile devices and the increasing amount of news readers, mobile news recommendation has become one of the hottest topics in the recommendation field. It is the main task of mobile news recommender systems, how to improve recommendation performance and user satisfaction according to the characters of mobile devices and news. This paper summarizes the research status of mobile news recommender and distinguishes it from traditional news recommender as well as mobile recommender for other purposes. Specially, from six key aspects, including news articles representation, news preference acquisition for mobile users, context-aware mobile news recommendation techniques, mobile news recommendation techniques based on social networks, news presentation on mobile devices and some typical applications, this paper provides a deep comparison, analysis and peroration of the latest research in the field of mobile news recommendation. Furthermore, we discuss and analyse the challenges in the field of mobile news recommender from the perspectives of key points and difficulties. Finally, we point out some future work to be done to promote the further development of mobile news recommender, particularly the acquisition of mobile news data sets, performance evaluation, the methods of incorporating with mobile social networks, security problem and formal description and so on.

Keywords personalization; context-awareness; mobile social networks; mobile news recommendation; social media; data mining

收稿日期:2015-04-24;在线出版日期:2015-08-31. 本课题得到国家自然科学基金(60872051)、北京市教育委员会共建项目专项资助。
孟祥武,男,1966年生,博士,教授,博士生导师,中国计算机学会(CCF)高级会员,主要研究领域为 Web 服务、用户需求、推荐服务。
E-mail: mengxw@bupt.edu.cn. 陈 诚,女,1988年生,博士研究生,主要研究方向为数据挖掘、移动推荐服务. 张玉洁,女,1969年生,副教授,主要研究方向为智能信息处理、通信软件、网络服务.

1 引言

新闻传播是互联网的经典应用,随着互联网的迅速发展,上网浏览新闻已经成为许多网民的习惯. CNNIC 报告指出到 2014 年 12 月,网络新闻的使用率已达到 80.0%^①,远超其他网络应用. 网络新闻在传播深度和速度方面都远远领先于传统媒体,然而海量的新闻信息也让人们陷入面对大量垃圾信息的尴尬境地,因此新闻推荐(news recommendation)应运而生. 新闻推荐作为推荐系统^[1-4]研究领域的一个重要分支,就是指通过新闻内容和用户信息,尽可能多地帮助用户找到与用户兴趣偏好相匹配的新闻^[5]. 它不同于人工智能和信息检索的相关领域,现已成为一个独立的研究领域.

随着移动通信技术(如 3G、Wifi 和 4G 等)的发展,手机、平板电脑、PDA 等移动设备的迅速普及,移动信息服务正成为移动增值服务市场上新的热点^[6]. 与传统新闻推荐相比,由于受到移动设备固有的屏幕尺寸、存储空间、较低的评论查看率、计算能力、I/O 资源和高费用网络等因素的限制^[7],以及移动设备的绑定性、便携性、支持随时随地浏览新闻等因素的影响,移动新闻推荐在推荐效率、实时性、新闻内容展示和对时间、位置等上下文(context)^[2,8]的感知等方面的需求更高;与其他移动推荐对象相比,新闻具有来源广泛、数量庞大、更新速度快、老化现象明显等特性,因此移动新闻推荐还需要保证推荐结果的时效性,内容的新颖性和多样性等^[5]. 针对上述需求和特点,本文把焦点放在新闻领域,主要对移动新闻的推荐技术及其应用的研究现状进行了深入的讨论和分析.

移动新闻推荐作为移动推荐中的一个分支领域,也是近十年间发展起来的,虽然目前关于移动新闻推荐的研究还处在发展阶段,但它已经开始影响人们的生活方式,并引起了一些专家学者的注意. 因此,国内外越来越多的高校和研究机构加入到移动新闻推荐的研究中来,如英国的朴茨茅斯大学^[9]、美国的麻省理工大学^[10]、乔治亚理工学院^[7]、加拿大的加州大学欧文分校^[11]、瑞典皇家理工学院^[12]、日本的东京大学^[13]、韩国高级科学技术研究院^[14]、高雄大学^[15]、交通大学^[16-17]、政治大学^[18]、中国科学院^[19]、浙江大学^[20]、北京邮电大学^[6,21-22]、河海大学^[23]等.

本文第 2 节对移动新闻推荐进行概述;第 3 节重点分析讨论移动新闻推荐中的关键技术,包括新闻表示方法、移动用户偏好获取、上下文感知的移动新闻推荐、基于社会化网络的移动新闻推荐、移动新闻推荐结果显示等;第 4 节分类介绍几个典型的移动新闻推荐系统,对该领域的实际应用展开论述;第 5 节详细探讨移动新闻推荐领域的重、难点问题;最后是总结.

2 移动新闻推荐概述

2.1 新闻推荐

网络新闻是基础网络应用之一,用户通过一些专门的新闻网站(如 Google News^[24]等)阅读在线新闻,随着社会化网络服务(如 Twitter 等)的盛行,网民可以通过更多的渠道接触新闻资讯,并在对新闻的分享和转发过程中提升新闻的覆盖量. 传统新闻推荐中,用户通过台式机 etc 终端设备接入互联网,但是台式机与用户不存在绑定关系,因此用户如果要获得新闻网站提供的个性化推荐服务,需要注册登录网站进行新闻浏览,才能将用户与其浏览行为日志对应起来. 网站则可以通过用户的偏好档案^[25],订阅新闻或者用户的历史行为数据^[24,26],向用户进行个性化新闻推荐. 目前,已经出现一些传统新闻推荐系统,例如: SCENE^[5]、Athena^[25]、YourNews^[27](<http://ir.exp.sis.pitt.edu/gale/news>)等.

本文根据常见的推荐技术将新闻推荐分为以下 4 种并对其进行了比较分析:(1)基于内容的新闻推荐^[11]:该方法比较直观,易于理解,但是它对新闻特征的提取能力有限,存在着过度专门化的问题;(2)协同过滤新闻推荐^[28-29]:该技术可以为用户发现新偏好,适用于推荐一些复杂的非结构化的音、视频新闻^[7,20],但是可扩展性较差,存在“冷启动”、稀疏性等问题,而且由于新闻的新颖性(见 2.3 节),通常不能单独用于新闻推荐;(3)基于知识(语义)的新闻推荐^[8,25,30]:和上述两种方法相比,该技术可以直接将用户偏好映射到被推荐新闻上,不存在“冷启动”和稀疏性等问题,但是其推荐效果对知识库的依赖性较强,而专门领域的知识和推理规则通常较难获取;

① 第 35 次中国互联网络发展状况统计报告. 2015. <http://www.cnnic.net.cn/hlwzfzyj/hlwzxbg/hlwztjbg/201502/P020150203548852631921.pdf>

(4) 混合新闻推荐^[8,24,30]: 该技术采用一些策略(如加权、切换、特征放大等)整合多种推荐技术,以弥补上述推荐方法单独使用时存在的缺陷。目前,结合内容和协同过滤的混合推荐是使用最广泛的^[8-9,24,31]。

2.2 移动新闻推荐和传统新闻推荐异同点

传统新闻推荐基于台式机等非移动终端向互联网用户提供个性化新闻服务。由于台式机的计算、存储能力较强,接入网络稳定,操作便利,输入输出能力强,加之新闻本身对时间敏感,数量庞大,因此传统新闻推荐具有大规模、高时效、易交互、多元化等特点。随着手机等移动设备的日益普及,无线通信技术的不断发展,越来越多的人开始使用移动设备,通过无线网络浏览新闻,促进了传统新闻推荐系统向移动新闻推荐系统的过渡。

移动新闻推荐系统在复杂、融合、泛在的移动网络环境下,基于移动终端向用户提供个性化新闻服务。为了适应复杂多变的移动网络环境,合理利用移动设备有限的资源,移动新闻推荐系统通常采用 offline-online 工作模式,即离线建模-在线推荐。移动新闻推荐具有以下特点:

(1) 有限的计算和存储能力。设备的计算能力有限,能源和存储能力也很有限,且因为设备经常处于移动状态,所以其网络接入不稳定,具有较低的网络带宽。由于移动设备这些自身的限制,不能进行复杂的计算,因此移动新闻推荐的在线推荐过程一般要采用比传统新闻推荐更简单、高效、快速的计算方法。

(2) 屏幕尺寸小。移动设备通常屏幕尺寸较小,所以移动新闻推荐还需考虑如何根据不同的设备屏幕尺寸,调整新闻版面布局,即自适应性移动新闻推荐^[7]。

(3) 实时性。移动设备支持用户随时随地获取信息、浏览新闻,因此移动新闻推荐比新闻推荐更方便、快捷,相应的对移动新闻推荐的实时性要求更高。

(4) 个性化。移动设备的私有性、便携性,已经是一种个性化的标志,移动设备和移动用户之间存在着唯一的绑定关系,只要通过授权,就可以获取用户在移动设备注册入网时填写的人口统计学特征资料。此外,多样化的上下文(如用户情绪、位置等)也体现了移动新闻推荐的个性化。

(5) 移动性。移动新闻推荐的移动性表现在 3 个方面:用户的移动性(移动过程中,用户从不同的位

置接入服务)、设备的可携带性和无线接入(设备通过无线技术接入网络)。

移动新闻推荐是对传统新闻推荐的扩展和延伸。虽然两者的基本算法思想相似(如基于内容的推荐),但由于移动新闻推荐受到移动网络环境,移动设备自身的限制(处理能力差、屏幕小等)和特点(可移动性、便携性等)的影响,两者又存在本质上的区别,主要表现在新闻数据来源、设备性能、上下文感知等方面,详见表 1。

表 1 移动新闻推荐与新闻推荐比较

	新闻推荐	移动新闻推荐
新闻数据来源	新闻网站	RSS 订阅源、新闻网站
终端设备	个人电脑等	手机、平板电脑等
设备能源	充足	有限
存储资源	较大	很小
计算能力	强	弱
屏幕尺寸	较大	较小
接入网络	互联网,稳定	移动网络(GPRS、3G、4G),不稳定
网络带宽	较大	有限
实时性	较弱	较强
位置关联性	小	大
上下文信息	较少,语义运行上下文等	种类多样,如时间、地点、网络状况等

2.3 移动新闻推荐和其他移动推荐异同点

移动用户访问互联网主要进行的活动有资讯(以新闻定制为代表的媒体短信服务)、沟通(QQ 等聊天工具、移动社交网络等)、娱乐(音乐、游戏等)和移动应用程序下载,还有信息检索、移动电子商务等。如表 2 所示,针对移动用户不同的活动,相应地产生了一系列的移动推荐系统,例如移动旅游推荐,移动应用程序推荐,移动新闻推荐等。其中,COREs^[32]是一个普适的移动推荐系统,适用于多种移动服务推荐。从表 2 可以看出,大部分移动推荐系统都与上下文息息相关。

表 2 一些典型的移动推荐系统

推荐类型	典型系统	推荐方法	是否有上下文
移动新闻	MONERS ^[10]	混合推荐	有
	PPNews ^[9]	混合推荐	有
移动音乐	CoFoSIM ^[33]	协同过滤推荐	无
移动旅游	ReRex ^[4]	协同过滤推荐	有
移动电子商务	PR ^[34]	基于内容推荐	有
移动应用程序	TruBeRepec ^[35]	协同过滤推荐	有
移动服务	COREs ^[32]	混合推荐	有

由于新闻本身具有数量大、更新快、冗余度高以及地域性等特点,移动新闻推荐系统和其他移动推荐系统(如移动应用程序推荐、移动电子商务推荐)之间的差异性表现在 3 个方面:时效性(recency)^[9]、新

颖性(novelty)^[36]和位置感知。

(1) 时效性

移动新闻推荐区别于其他移动推荐的最大特点就是和时间上下文信息关联性很强,也就是通常所说的新闻时效性.因为新闻有很明显的老化现象,超过一定时间发布的新闻对用户来说是没有意义的,这一时间段和下一时间段的新闻推荐集不同,所以移动新闻推荐一定要实时.而其他的移动推荐,比如:移动购物推荐或者移动应用推荐,通常对时间上下文并不敏感.

(2) 新颖性

移动新闻推荐对推荐结果有新颖性的要求.新闻和其他被推荐对象(如电影、餐厅)不同,新闻来源非常广泛,数量庞大,同一个事件可能由多个记者报道,写成多篇内容相近的新闻,因此新闻中存在大量重复冗余的数据.在其他的移动推荐中,一个重要的推荐思想就是根据用户已知的感兴趣对象,向用户推荐与已知对象相似的潜在感兴趣对象.但是在移动新闻推荐中,用户并不愿意反复浏览相似的新闻,因此需要保证移动新闻推荐结果的新颖性,以提高用户体验.

(3) 位置感知

位置感知也是移动新闻推荐区别于其他移动推荐的一个特点.对移动新闻推荐来说,其推荐过程总是与用户所在的地理位置或者用户的运动轨迹有关,

这里通常需要使用一些定位技术来感知用户所在的地理位置,然后根据位置信息或者运动轨迹推荐相关新闻.移动设备定位技术通常分为两种:一种是GPS定位技术;另一种是基站定位技术.前者的定位精度较高,而后者不需要设备具有GPS定位能力.此外还有利用Wifi在小范围内进行定位的方式.除了旅游景点推荐和餐厅推荐等,其他的移动推荐与移动用户地理位置相关性并不明显,比如移动应用推荐.

3 移动新闻推荐系统关键技术

移动新闻推荐综合了传统新闻推荐和移动推荐两者的特点,同时又区别于这两者,成为一个独立的研究方向.目前移动新闻推荐还处于发展阶段.本节首先提出了移动新闻推荐的整体框架图,然后针对框架图中每一部分涉及的关键技术研究现状进行了分析、对比和概括总结.

3.1 移动新闻推荐整体框架

移动新闻推荐有3个组成要素:候选新闻、用户和推荐方法.图1给出了移动新闻推荐的通用模型.如图所示,该模型分为客户端和服务端两部分.客户端收集用户的原始数据(包括新闻浏览数据、传感器感知的上下文数据,如时间、运动状态等)上传至服务器端进行处理.服务器端从新闻源(RSS订阅

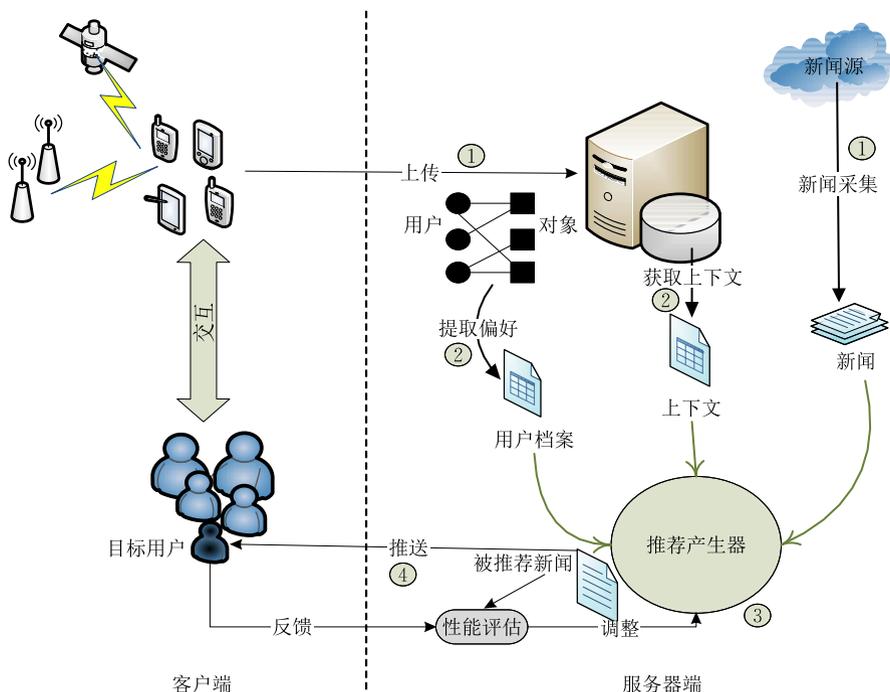


图1 移动新闻推荐通用模型

源等)进行新闻搜集;对移动用户进行偏好提取,建立用户档案;同时通过分析推理方法获取上下文信息,如利用 GPS 或者基站定位获取的经纬度信息分析得到移动用户的地理位置;最后综合上述信息,通过一个推荐产生器向目标用户推送其感兴趣的新闻并基于用户反馈对推荐性能进行评估,进一步调整推荐产生器,以适应用户偏好。

根据图 1 所示模型,移动新闻推荐过程可以划分为 4 个部分:

(1)数据采集和预处理. 该部分从 RSS 订阅源或新闻网站获取新闻集,并将非结构化的新闻文档采用统一的结构化方式进行描述;利用传感技术等获取移动用户浏览新闻的历史数据;通过显式、隐式或者推理的方法获取上下文信息^[2],并进一步对上下文信息进行量化或者计算。

(2)新闻偏好获取. 该部分根据移动用户历史行为数据以及上下文信息进行偏好提取. 偏好的获取方法可以分为显式获取(如手工定制兴趣类别或者由用户提供兴趣度评分)和隐式获取(即根据用户的浏览内容和浏览行为自动获取用户偏好,建立用户模型^[6,11],建模过程无需用户主动提供信息的方法). 由于新闻数量庞大、更新时间迅速加之移动设备屏幕小,不便于操作,显式的偏好获取方式会加重移动设备计算、通信负担,也会干扰用户正常的浏览行为. 因此,当前的研究主要使用的是隐式的获取方式. 此外,移动用户的新闻偏好是随时间动态变化的,因此需要适时地对用户偏好进行更新。

(3)移动新闻推荐. 该部分是移动新闻推荐的核心部分,除了考虑基于内容的推荐,协同过滤推荐等基本技术,还需要重点研究上下文感知的移动新闻推荐和基于社会化网络的移动新闻推荐. 其中,上下文感知的移动新闻推荐又分为单维上下文和多维上下文的移动新闻推荐. 基于社会化网络的移动新闻推荐可以分为基于社交网络和基于社交媒体两种推荐。

(4)推荐结果展示及效用评价. 移动新闻推荐过程中的最后一步是将推荐结果显示到移动设备上并对推荐性能进行评估. 由于移动设备屏幕尺寸比较小,需要采用适当的方式将移动新闻推荐结果集合展示给用户,并基于用户反馈,利用点击率、精确度、响应时间等评估指标评价移动新闻推荐性能。

3.2 移动新闻数据采集和预处理

移动新闻推荐系统整个过程由收集数据开始,

其需要采集的数据主要包括移动用户信息、新闻数据、上下文信息、移动用户浏览行为数据以及移动用户通信行为等信息. 其中,移动用户信息主要来自移动用户注册入网档案,新闻数据可以从新闻网站或 RSS 订阅源获取,上下文信息由用户自主给定或通过传感器直接、显式获取,而移动用户的浏览行为、通信行为等信息则可以从系统使用日志中提取^[6]。

数据采集阶段得到的初始数据可能存在脏数据、数据缺失等问题. 为了获得规范化的移动新闻推荐系统的输入数据,需要进一步对上述数据进行预处理(计算、量化、结构化等). 比如,时间上下文可以划分为上午、下午和晚上三个时间段并用唯一的量化值表示;用户的点击浏览行为也可以采用二进制量化表示. 显式获取的上下文信息一般较少,因此需要利用隐式、推理方法获取深层次的上下文信息. 如利用隐式方法从已知的坐标信息中获取移动用户位置信息;基于数据挖掘技术推理计算获取用户情绪上下文等. 通常采集到的新闻数据都是非结构化的,不利于处理. 下面我们着重介绍了新闻的结构化表示方法。

3.2.1 新闻表示方法

文本新闻是移动新闻推荐的主流研究,视频或音频新闻通常也可以由其附带的标题、类别属性、标签集合等信息构成相应的新闻文档. 由于新闻数量庞大,新闻文档没有统一的格式标准,因此为了便于存储和后续处理,需要将非结构化的新闻文档采用统一的结构化方法表示. 通常结构化新闻表示方法在移动新闻推荐和传统新闻推荐中是一致的。

目前常用的新闻表示方法有 VSM(Vector Space Model)、概率模型^[37]等. 概率模型表示方法即利用不同的概率生成模型获取新闻潜在的语义结构及相关分布,例如主题概率分布. LDA(Latent Dirichlet Allocation)^[38]、PLSA(Probabilistic Latent Semantic Analysis)^[37]、ESA(Explicit Semantic Analysis)^[39]等都是应用比较广泛的概率生成模型. PLSA 可以对高维空间进行降维,且模型比较稳定,但是缺乏处理新文档的能力,存在过拟合现象;LDA 克服了 PLSA 中存在的不足,更适用于大规模新闻集,但是不能对主题在新闻中出现的相关性进行建模;不同于前两种模型,ESA 采用可识别概念表示新闻,提高了其解释性,但是前期建立知识库比较复杂. VSM 是一种经典的结构化表示方法. 它通过用一组词项及

其对应权值构成的向量空间来描述新闻,即每篇新闻文档都可以用一个 $\{T, W\}$ 二元组来表示。 T 是所有词项(词语、主题或命名实体等)的集合,通过对新闻文档去停用词、分词和语义分析等获取。其中常用的分词工具有 Lucene、CKIP^[17]、ECSscanner^[18]等。和其他词项相比,命名实体可以更有效地捕捉到新闻主旨,如发生了什么事情,事件发生的时间、地点、涉及人员等。VSM 中常用的权值计算方法有频率统计^[40]、TF-IDF^[9,25]及其改进方法。其中使用最广泛的是 TF-IDF,其计算某个词项权重的公式如下:

$$\omega_{ij} = tf_{ij} \times idf_{ij} = \frac{\text{num}(T_{ij})}{\text{total}(D_i)} \times \log \frac{N}{n_{ij}} \quad (1)$$

$\text{num}(T_{ij})$ 表示词项 T_{ij} 在第 i 篇新闻中出现的次数, $\text{total}(D_i)$ 表示第 i 篇新闻包含的字词总个数; N 表示该新闻集中新闻总篇数, n_{ij} 表示出现字词 T_{ij} 的新闻篇数。

3.3 移动用户新闻偏好获取

移动用户对新闻的偏好不仅受到新闻内容影响还会受到上下文的影响,比如新闻的发布时间,用户所处的位置等。目前关于上下文的定义,还没有一个统一的说法,这里引用文献[32]中给出的定义“上下文就是指任何可以用来描述实体(如人、地点或者物等)情况的信息,其中实体是指用户与应用程序交互相关的客体,也包括用户和应用程序本身”。

3.3.1 移动用户新闻偏好表示方法

对应新闻的表示方法,上下文新闻偏好可以通过四元组 $\{U, C, T, P\}$ 表示; U, C, T 分别表示用户、上下文和对象的集合,这里对象可以是完整的一则新闻,也可以是用于表示新闻的词语、主题或者命名实体等。则移动新闻偏好中一个给定的具体实例 $\{u_i, c_j, t_k, p_{i,k}\}$ 代表的含义是移动用户 u_i 在上下文 c_j 下对对象 t_k 的偏好值为 $p_{i,k}$ 。

根据偏好的表示方法,本文将移动用户新闻偏好模型分成了单一建模和冗余建模两种方式:大部分研究都采用单一建模的方式,将获取的偏好用一种对象的确定值或分布来表示^[11,24,40]。例如 Billsus 等人^[11]只采用词语及其权值向量来表示用户新闻偏好,而 Liu 等人^[24]则仅利用主题及其概率分布来表示偏好;文献[40]分别根据 hashtag、实体和主题 3 种对象进行单一建模以比较不同的偏好模型。单一建模存储结构简单,可以节约存储资源,方便计算和处理,但推荐精确度受偏好获取方法影响较大;少数研究采用了冗余建模的方式,即将互有交集的多种对象融合建模。例如 Li 等人^[5]提出使用新闻的主

题分布、命名实体和相似用户列表共同构建用户偏好模型,这三者之间相互联系相互补充。冗余建模的方式存储结构比较复杂,增加了存储开销,但是容错性增强,有利于提高推荐精确度。

目前,移动用户新闻偏好获取主要有定量分析和定性分析两个方向。定量分析有利于偏好的微观比较形成偏好之间的全序关系,但不利于直观描述任意两个用户偏好之间的逻辑关系;定性分析可以直观地描述任何一对用户偏好实例之间的偏序关系,便于偏好之间的逻辑推理,但不利于计算。由于基于定性分析的新闻偏好获取方法^[21]比较少,所以本文主要是从定量分析的角度来描述移动用户新闻偏好的获取方法。

3.3.2 基于定量分析的移动用户新闻偏好获取方法

移动用户新闻偏好获取方法的定量分析,是指采用数值量化表示移动用户新闻偏好,并进行相关的数学计算。根据其实现的过程,定量的移动用户新闻偏好获取技术又可以分为两类:

(1) 启发式方法

基于启发式的方法就是指利用一些直观的启发式算法来获取移动用户新闻偏好,例如聚类分析、最近邻法^[21]、相似度计算^[34]等。Wei 等人^[26]根据用户的新闻点击历史,利用 Pearson 系数对用户的新闻偏好进行量化计算;Yang 等人^[34]则采用欧氏距离和相似度等计算用户偏好。通常,移动用户对新闻的偏好都是通过隐式方式从用户的浏览历史中挖掘获取的。由于移动设备有限的输入输出能力,移动用户很少在浏览新闻的时候给出具体的显式评分,因此新闻偏好在很多研究中采用二进制法表示^[24,26]:点击浏览记为 1,否则记为 0。但是这种量化方式获取的新闻偏好比较粗糙,只能简单地发现用户感兴趣的新闻,却无法精确地获取其偏好程度,也不能区分用户不感兴趣的新闻。

为了进一步细化用户对新闻的偏好,研究者深入分析了各种各样的用户浏览行为历史,以期获取更精细的新闻偏好。例如,通过用户是否打开新闻,浏览新闻的页数,是否下载新闻^[11],用户所在移动区,下载的新闻链接地址,每天下载某类新闻的次数,每次下载该类新闻的条数以及下载该类新闻时系统中新添加的新闻条数^[6]等来挖掘用户细粒度的偏好。移动新闻推荐中,用户的新闻偏好通常还受到上下文的影响^[8,41]。Cantador 等人^[8]利用余弦相似度来计算用户的新闻偏好,并根据偏好和本体概念之间的链接路径来获取用户语义上下文偏好;Wang

等人^[42]把上下文类型当作对象属性,利用多属性效用理论方法获取多维上下文移动用户偏好。

(2) 基于模型的方法

基于模型的方法就是指通过一些统计学模型、机器学习方法学习得到一个模型来获取移动用户的新闻偏好。目前,常用的模型有矩阵分解^[4]、贝叶斯网络^[9,24]、决策树^[15]、AHP (Analytic Hierarchy Process)^[9]等,这些模型的优缺点如表 3 所示。Baltrunas 等人^[4]提出采用结合上下文的矩阵分解评分预测技术来计算用户的偏好,并利用上下文之间的依存性来解决上下文和用户偏好冲突的情况;文献^[9]根据新闻领域的经验数据来选择相关变量和构建贝叶斯网络,并利用贝叶斯网络的快速推理能力预测用户感兴趣的新闻类别,以满足移动新闻推荐中实时个性化需求;Lee 等人^[15]提出使用多重决策树模型来预测用户偏好,将新闻分入预先定义的偏好类别中;Yeung 等人^[9]提出了一个 AHP 模型,通过用户分配和用户历史来计算层次分析上下文指标的相对权重,然后计算用户偏好。

表 3 基于模型的方法比较

模型	优点	缺点
贝叶斯网络	适用于不确定性推理,推理较精确	构造、训练过程较复杂
矩阵分解	有效缓解数据稀疏性,有较好的扩展性	解释性差
决策树	计算速度快,准确性较高	扩展性差,不适用大规模的偏好获取
AHP	适用于多属性影响的偏好获取,所需定量数据较少	定性成分多不易找到最优解

相比之下,启发式方法简单,容易实现,计算速度快,但是其效果不稳定,结果的好坏依赖于实际问题,因此无法保证偏好获取的准确性;而基于模型的方法提取用户偏好精确度较高,但是计算过程复杂,算法学习或者训练模型所需的时间较长。

3.3.3 移动用户偏好更新

由于用户的兴趣是随着时间不断变化的,偏好也会发生转移。为了获取用户最新的偏好以保证推荐精确性,需要实时对用户偏好进行更新,其相关技术主要有以下几种:

(1) 动态偏好建模技术。该技术在偏好更新过程中考虑了时间因素的影响,主要通过划分时间周期,时间片,滑动窗口等方法获取最新数据,以动态更新偏好模型。Gao 等人^[22]为每个用户维护了两个模型:兴趣模型和阅读历史模型,然后周期性地从阅读历史模型中选择前 n 个新闻标签及其权值动态更

新或者替换兴趣模型中的标签及权值;Li 等人^[43]通过分析范式将用户浏览历史数据分成多个时间片,然后利用时间权值方案动态地获取用户的长期偏好和短期偏好。动态偏好建模技术可以有效捕捉用户偏好随时间的变化,减少更新开销,但是其时间参数估计比较困难,既要避免引入过期偏好,增加计算复杂度,又要防止偏好丢失,影响偏好模型精确性。

(2) 自适应技术。它通过监测周围环境变化和用户反馈,自动地调整、更新用户偏好以适应这种变化。与动态建模技术不同,自适应技术只有在监测到变化且其会影响用户偏好的情况下,才会更新偏好模型,因此可以更有效地减少更新代价,但是需要进行复杂的离线训练,且无法保证模型的收敛性和稳定性。Shi 等人^[44]通过分析移动用户行为日志,检测其行为是否受上下文影响以及上下文移动用户偏好是否发生变化,然后利用基于上下文的最小二乘支持向量机对发生变化的上下文移动用户偏好进行学习。

目前,关于偏好更新的研究还比较少。在当前的大部分研究中,偏好更新意味着以最近的历史数据重新计算用户偏好。但是对于移动新闻推荐中庞大的新闻数量、多样的上下文信息、移动设备有限的计算能力和通信资源,这种更新需要付出更大的代价。因此关于移动用户新闻偏好的更新还需要进一步的研究。

3.4 移动新闻推荐

3.4.1 上下文感知的移动新闻推荐

“上下文感知计算”概念于 20 世纪 90 年代被提出来,但是目前“上下文感知”还没有公认的定义。Abowd 等人^[45]给出的定义被广泛采用:“如果系统使用上下文向用户提供相关的信息或服务,而相关性依赖于用户任务,则该系统认为是上下文感知的”。上下文感知是移动新闻推荐和传统新闻推荐之间最大的区别^[4,14]。与传统新闻推荐相比,移动环境中上下文具有细粒度(比如位置上下文在移动环境中可以分为办公室、餐厅、家等,而在传统新闻推荐中只能按城市、地区等划分),变化快,种类多样等特点,移动用户的新闻偏好受上下文的影响更大,因此需要实时感知多维动态上下文信息。加入上下文信息的推荐方法,从传统的“user-item”的二元形式,过渡到“user-item-context”的三元形式^[2]。移动新闻推荐使用了丰富的上下文信息(如表 4 所示)。本文重点介绍新闻区别于其他推荐对象的上下文(如位置、新颖性)以及多维上下文推荐等。

表 4 移动新闻推荐系统中的上下文信息

推荐系统	推荐内容	上下文	数据集
CARS ^[14]	移动内容和服务	位置,时间,设备,活动	韩国移动服务提供商提供的移动用户使用记录
ILNSS ^[18]	新闻和新闻摘要	位置	Google 新闻数据
PPNews ^[9]	移动新闻	情绪,用户状态,天气,位置,任务,附近的朋友和人口统计学上下文:性别,职业地位,年龄和社会经济等级	RSS 新闻网站数据和收集移动用户使用系统产生的数据
News@hand ^[8]	新闻项	用户兴趣的语义上下文	RSS 订阅源数据
EPG ^[23]	移动电视新闻	网络环境	DVB-H, IPTV 和互联网上的数据
ERS ^[13]	微博帖子	位于查询词前后的词语	Twitter 数据
SCPPS ^[12]	网络消息源内容	名字,生日,性别,当前位置,状态信息,宗教,政治观点,活动,群组,网络成员,参与事件,朋友列表和爱好;兴趣,电视剧,电影,书籍	移动用户使用系统产生的数据
MONERS ^[10]	移动新闻	新闻文章的重要性和时效性	韩国移动服务提供商提供的会员数据

(1) 基于时间的移动新闻推荐

基于时间的移动新闻推荐包括时效性和实时性(real-time)^[13,46-47]两个方面.时效性是新闻区别于其他对象的本质特征,发布时间较久的新闻就失去了它作为新闻的价值;由于移动设备具有便携性,可以支持移动用户随时浏览新闻,因此和传统新闻推荐相比,移动新闻推荐对实时性要求更高.本文按照时间上下文引入的方式,将基于时间的移动新闻推荐分为时间过滤和时间建模.

①时间过滤就是利用时间上下文对新闻进行过滤,筛选最近的新闻,又可以分为预过滤和后过滤两种.PADAC^[47]首先利用时间上下文对新闻集合进行预过滤,只缓存 30 min 内发布的新闻作为候选项,然后利用用户的个性化信息进行推荐,保证了新闻推荐的实时性;Karkali 等人^[46]提出了一个浏览器扩展组件 Match the News 用于新闻的实时推荐,该组件在移动用户登录浏览器浏览新闻时,预过滤出最近发布的新闻然后选择与用户当前浏览新闻最相关的进行推荐;SCENE^[5]利用时间对候选推荐列表进行后过滤,调整推荐列表的排序,保持了推荐结果的时效性和流行度之间的平衡.时间过滤对基于时间的新闻推荐进行了降维处理,实现起来比较简单,但是忽略了时间上下文和其他数据信息(如用户偏好)之间的关联关系,从而对推荐效果产生影响.

②时间建模就是将新闻的时间上下文进行建模,融合到偏好获取、推荐生成的整个过程中,处理方法比较复杂,但是保证了数据之间的关联性和完整性,有利于提高推荐精确度.新闻的时间属性指标(r_{ij})通常是用下面的公式来计算的:

$$r_{ij} = -(t_{\text{present}} - t_{\text{posted}} / t_{\beta})^2 + 1 \quad (2)$$

其中, t_{present} 表示当前的时间戳, t_{posted} 表示新闻发布时的时间戳, t_{β} 表示基准时间戳(一般取 24 h).当新

闻发布的时间已经超过 t_{β} 时, $r_{ij} = 0$. MONERS^[10]把新闻时效性当作上下文向量中的一维,单独进行了建模计算,然后结合其他上下文信息共同获取用户偏好,以提高新闻时效性;Yin 等人^[48]区分了面向用户的主题和面向时间的主题,并利用主题模型为时间上下文进行建模,然后基于用户固有兴趣和随时间变化的公众趋势进行推荐.

总的来说,基于时间的移动新闻推荐有利于提高新闻推荐的时效性,促进了最新发布的新闻传播,但是在一定程度上削弱了用户兴趣的主导地位.因此,在推荐过程中需要考虑新闻时效性和用户偏好之间的平衡.

(2) 基于新颖性的移动新闻推荐

基于新颖性的移动新闻推荐,其关键就是差异性度量,即通过比较待推荐新闻和用户浏览历史之间的差异度,将过于类似的新闻提前进行过滤删除,以保证所推荐新闻内容的新颖性和多样化.其中,常用于衡量单篇新闻文档和一组新闻集之间差异性的方法有距离公式^[36](如 KL 散度、Jensen-Shannon 散度、欧式距离等)和相似性度量^[9](如余弦相似度、Pearson 相似度等).

目前,基于新颖性的移动新闻推荐研究比较少.例如 Newsjunkie^[36]把新闻表示成词汇和命名实体,然后分别使用 KL 散度和自定义的 NE 方法(通过浏览历史中未出现的命名实体在待选新闻中的密度来衡量)计算新闻的新颖性,并通过新颖性算法分析了新闻文档间动态性和文档内动态性,但该方法仅基于用户当前浏览的新闻进行推荐,未考虑用户的长期偏好,不能全面地把握用户兴趣动向;Karkali 等人^[46]提出了一个 BM25H 权重函数用于抽取浏览历史和候选新闻的关键词,并将关键词不同于浏览历史关键词分布的新闻推荐给用户,以保证新闻

推荐的新颖性。Li 等人^[43]基于用户历史数据建立用户档案,然后根据用户档案过滤得到的“新数据”建立 user-item 关系图,并在该关系图上使用吸收随机漫步模型来选择新闻以提高推荐列表的新颖性。和传统的新颖性度量方法相比,该方法通过层次化新闻结构减少了比较计算量,但将新颖性信息限制在用户兴趣范围内,忽略了突发事件的影响。

基于新颖性的移动新闻推荐过滤掉重复冗余的信息,减轻了移动设备的处理负担,保证了推荐结果的多样性,有利于提高用户体验。但是容易出现过度处理的问题,即向用户推荐完全不符合其兴趣的新闻。因此,基于新颖性的移动新闻推荐,其未来研究重点应该是如何根据信息随时间的变化发展进行推荐。

(3) 基于位置的移动新闻推荐

新闻通常具有一定的地域性,即新闻在地理环境中的影响范围,不同的新闻其影响范围不同。例如,天气新闻、交通新闻其作用域通常是一个市或者区,而奥运会等全球体育盛事的相关新闻则会吸引世界各地的用户。在移动设备的支持下,用户可以在任何地方浏览新闻,其新闻偏好也会随着位置上下文而改变。比如,用户到 A 市出差,则他对 A 市的相关新闻更感兴趣。因此移动新闻推荐通常是位置感知的^[14,39,49-51]。本文根据推荐过程中新闻的位置属性和移动用户当前位置的匹配方式,将其分为两类:点匹配和范围匹配。

① 点匹配就是指位置属性严格匹配用户当前位置点的新闻才可以被推荐,其中位置点的划分可以是城市,街道等。Son 等人^[39]提出了 ELSA (Explicit Localized Semantic Analysis),从相关的地理标签文档中抽取各个位置的地理主题,将新闻和位置映射到同一个 Wikipedia 主题空间。该方法将不同属性的两种对象(新闻和位置)之间的匹配转换成简单的向量之间的比较,但是它没有考虑用户个体偏好;Yin 等人^[38]提出的 LCARS 系统基于对象共现模式和对象内容相似度来学习用户个人兴趣和位置的本地偏好,并结合它们向用户推荐与用户所处位置相关的事件。

② 范围匹配是指位置属于用户当前位置及其周边给定范围内的新闻都可以被推荐。这里的范围通常指的是以用户所在位置为圆心,给定半径的圆形区域内。Rosa 等人^[49]通过移动用户与新闻发生位置之间的欧式距离以及预先给定的范围计算新闻相应的位置评分,作为推荐新闻的指标;其中范围是

根据经验分析给出的,采用的交通工具(如步行、自行车等)不同,范围大小也不同;GeoFeed^[50]实现了新闻与用户位置的点匹配和范围匹配两种方式,并设计了一个智能模型来决定如何使用 3 种不同的策略 spatial pull、spatial push 和 shared push 进行位置感知的新闻流推荐。

此外,还可以根据用户的移动轨迹进行新闻推荐。Xu 等人^[51]提出了一个专为移动用户设计的推荐新闻流的框架 MobiFeed,它由 3 个关键功能组成:位置预测、相关性测量和新闻流调度。MobiFeed 将路径预测引入基于位置的移动新闻推荐中,然后根据用户的移动轨迹,实时地向用户推荐位置相关的新闻,为基于位置的新闻推荐提供了新的视角。

上述研究表明,基于位置的移动新闻推荐有利于用户发现与自身位置相关的新闻,捕捉到当下周围环境中所发生的事情,但是其研究趋向于提高位置匹配精确度,而忽略了位置感知的用户偏好研究,导致给定位置下推荐给所有用户的新闻都是一样的。此外,基于用户移动轨迹的新闻推荐值得进一步的研究。

(4) 多维上下文移动新闻推荐

移动环境中上下文信息丰富多样,新闻推荐通常受到多维上下文影响^[9,12]。例如,CARS^[14]使用了位置、时间、设备、活动等上下文;MONERS^[10]则在推荐时考虑了新闻的重要性和时效性上下文。实现多维上下文移动新闻推荐需要解决两个关键问题:① 上下文选择,即确定推荐过程受到哪些上下文影响;② 上下文对推荐过程的影响程度,不同上下文对移动新闻推荐的影响程度不同。

目前,多维上下文移动新闻推荐常用的方法和模型有:

① 上下文融合。Rosa 等人^[49]基于领域知识选择了位置、访问频率和分享频率作为移动新闻推荐上下文信息,并通过多种数据融合技术(如 CombSUM、CombMNZ 等)将多维上下文各自与新闻的关联性进行融合,得到总体关联度进行推荐。其中上下文对推荐的影响程度由用户个性化定制并通过用户反馈进行调节。该方法有利于降维,简化推荐过程,但选择的上下文信息较少,且对上下文信息只进行了特征层次的融合。

② FOAF 词汇表。Karapantelakis 等人^[12]基于 FOAF 方法为社会化网络中所有的上下文(如表 4 所示)建立 XML 格式的语义词汇表,然后基于本体匹配进行推荐。上下文在推荐中的影响力随着其与

被推荐项所属类别的似然度而改变。

③ AHP 模型. Yeung 等人^[9]在推荐过程中考虑了用户和新闻的所有上下文(见表 4),然后采用 AHP 模型对层次划分的上下文进行比较计算,确定其各自的影响权重用于推荐.其中上下文的影响权重可以根据个体或群组分配自动进行调整。

④ 子空间聚类. 根据经验选择, Mo 等人^[52]在推荐中使用了网络类型、设备类型、接入时间、离开时间和位置作为上下文信息,并采用属性聚类的子空间聚类方法对上下文进行分析.然后使用上下文聚类代替初始上下文,通过多维上下文和用户浏览行为的共现模式进行推荐。

上述技术各有利弊,其具体对比研究如表 5 所示。

表 5 多维上下文移动新闻推荐方法或模型对比

模型(方法)	优点	缺点
上下文融合	有利于处理多源异构的上下文信息,便于获得排序	处理过程耗时较长且处理效果非常依赖融合方法的选择
FOAF	从语义层面对多维上下文进行了结构化描述,便于扩展	缺乏对部分上下文的描述
AHP	适于多维上下文的综合评价,能够有效处理定性和定量相结合的问题	分析效果依赖于层次指标的划分
子空间聚类	可以有效实现多维上下文的聚类并具有良好的扩展性	聚类精确度较低

3.4.2 基于社会化网络的移动新闻推荐

社会化网络(social network)是一个由个体结点以及反映个体之间特定关系的边所组成的图.基于用户的社交行为(如通话,短信,信息分享)可以构建各种类型的社会化网络,例如用户间的通信网,社区成员网络.随着社会化网络服务(如 Facebook、微博等)的出现,用户与网络之间的关系有了翻天覆地的变化,它更侧重于用户的交互作用,用户既是内容的浏览者,也是内容的制造者.不同于传统意义上的新闻,社会化网络向用户提供了一个交友、分享资讯的平台,其在一定程度上起到了信息传播和流通的作用.因此,社交网络、博客和微博等内容也成为新闻的一大来源^[53-54].本文将社会化网络服务按功能分为社交网络和社交媒体两类,分别进行讨论分析。

(1) 基于社交网络的推荐技术

目前,国外比较流行的社交网络有 Facebook、MySpace 和 LinkedIn 等,而国内比较著名的社交网络有人人网等.不同的社交网络注重点并不一样,比如, Myspace 将自己定位为音乐型社交网络,而

LinkedIn 是一家面向商业客户的社交网络^[12]。

社交网络中互相加为好友的用户在现实生活中通常也是熟悉或了解的人,他们构成了一个社会化关系网,这个关系网是建立在信任的基础上,因此利用社交网络进行移动新闻推荐,其实就是利用用户之间建立的信任网络进行推荐,这不仅可以提高推荐内容的可信度,防止恶意节点散播有害信息,还可以帮助解决“冷启动”、数据稀疏等问题. Wang 等人^[41]把移动社交网络当作用户之间的信任关系网络,并基于信任网络的评价和协同过滤方法来进行推荐;文献^[26]结合了历史记录和社会化网络向用户推荐新闻,它可以不断地根据用户品味的相似性更新社会化网络的拓扑结构,以提高推荐性能; Lin 等人^[55]提出使用社交网络上潜在的影响力专家的意见等隐式社交因素作为推荐新闻的辅助资源.社交网络中可以挖掘的用户上下文信息很多.例如, Facebook 提供了人口统计资料上下文(如用户性别、宗教等),社交上下文(群组、网络成员信息和朋友列表等)以及位置上下文等^[12].因此,充分利用社交网络提供的丰富数据,深入挖掘更多潜在的用户信息,为用户提供可信度和准确度更高的个性化推荐,值得进一步研究。

(2) 基于社交媒体的推荐技术

社交媒体^[48,56](博客和微博等)与 Facebook、人人网等社交网络不同,它更多的是展示媒体的特性,作为资讯的来源,而不是关系的维护.从用户结构来看, Facebook 跟人人网的关注主要是家人、朋友、同学,而微博(博客)主要关注企业、名人、媒体等.社交媒体在因特网上产生了一种新的交流方式和出版媒体来传播有用信息.与传统新闻比起来,写微博(博客)的人没有专门的组织积极地将他们的内容宣传给用户,而是依靠网络用户的偶然访问、分享或者推荐^[56]。

针对社交媒体上信息数量多、更新快、内容稀疏和用户活跃等特点,已经出现了许多相关研究.根据社交媒体在移动新闻推荐中的作用,这些工作可以分为两类:

① 社会化过滤.利用社交媒体中用户之间的协同作用、社会化关系以及用户评论等信息进行筛选推荐,以解决社交媒体文本内容稀疏、数量多等问题.例如, Wang 等人^[56]在社交媒体中根据用户评论包含的结构化语义信息以及评论改变的思路进行新闻推荐.社交媒体中移动用户新闻偏好不仅由个人兴趣决定,还会受到热点话题(hot topic)的影响,其

通常是大量用户协同作用(如转发、评论等)的结果。因此,通过监测社交媒体上的信息内容和用户行为可以迅速发现热点话题,并向用户推荐。文献[19]基于相关主题模型从大量微博中抽取有限个隐含主题,帮助用户了解最近一段时间内的热点话题。但是该方法只能推荐高度概括的主题而不能推荐具体的相关微博,解释性较差;Sakaki 等人^[13]通过监测 Twitter 帖子和用户的协同作用,有效且迅速地发掘出目标事件相关话题及其发展轨迹,并将收集的信息实时地推送给用户。

② 媒体交互。事实上,许多社交媒体信息都和新闻事件相关, Twitter 上甚至有超过 85% 的内容是和新闻有关联的^[40]。因此可以将社交媒体和传统新闻交互进行新闻推荐^[40,47,57]。其中,两者的交互又可以分为两种情况。一种是利用传统新闻包含的丰富信息弥补社交媒体中短文本稀疏的内容,从而获取完整、全面的用户偏好。Abel 等人^[40]提出的个性化新闻推荐框架利用微博中的 URL 链接或相似度计算将微博与相关新闻联系起来,然后从新闻中抽取实体、主题等来丰富相关微博的语义信息,并分别建立了 3 种用户偏好档案用于新闻推荐:基于 hashtag、基于实体和基于主题的用户档案;另一种则是利用社交媒体上用户活跃的行为(如转发、评论等)以及丰富多样的上下文信息(如位置、社交关系等)为推荐传统新闻获取偏好。PADAC^[47]综合了多种用户社交足迹(如 Twitter 和 Facebook 上发布的信息等)进行实时的新闻推荐;Lee 等人^[57]利用 Twitter 中获取的信息,包括 tweets、re-tweets 和 hashtags,并从这些信息中抽取出重要的关键词来构建用户的偏好档案以提供个性化新闻推荐服务。

此外,针对博客^[16-17]也出现了一些个性化推荐研究,例如移动内容推荐系统(M-CRS)^[17],移动设备上的定制内容服务(m-CCS)^[16]等。

社会化网络促进了新闻的传播,并为移动新闻推荐提供了丰富的辅助信息(如帖子、关系网、用户评论等)。但是由于社会化网络中大部分的信息都是由数以亿计的移动用户提供的,数据量庞大,篇幅通常较短,缺乏标准或规范的格式,加大了数据挖掘的难度。此外,社会化网络中的信息更新速度比传统新闻更快,因此基于社会化网络的移动新闻推荐,对实时性的要求更高。

3.5 移动新闻推荐结果显示

将新闻推荐结果以合理、有效地方式展示给移

动用户是实现移动新闻推荐过程的最后一步。由于移动设备屏幕尺寸通常较小,不便于浏览新闻,而推荐结果在移动设备上的展示直接影响了用户对推荐效果的评价和整体满意度。因此,为了提高浏览效率和用户体验,需要考虑新闻推荐结果在移动设备上的显示。本文将新闻在移动设备上的展示分为以下两个方面:

(1) 显示内容

一条新闻通常由标题、图片、URL 链接和正文等部分组成。为了有效地显示新闻,需要将一些重要的有代表性的内容显示出来,而将次要的内容隐藏起来节约空间。例如:TNPYH^[58]只显示新闻的标题和导语;CityEvents^[49]首页则显示了标题、图片和摘要等。从上述研究可以发现,新闻摘要提取也是实现有效显示的关键。ILNSS^[18]提出了一个基于模糊综合评价方法的多文档新闻摘要方案,根据 3 项评估分数在每则新闻中定义一个关键段落作为摘要;Otterbacher 等人^[59]提出了分层文本摘要方法;Lee 等人^[60]提出了一个自适应的视觉块分割方法,将新闻中不必要的内容块过滤并抽取有用的内容进行展示。多文档摘要方法可以较全面地把握新闻主旨,但增加了计算复杂性;分层文本摘要方法便于计算和筛选,但是会丢失句子间的语义联系;视觉块分割方法比较容易实现,但是摘要质量并不理想。

(2) 显示方式

目前新闻推荐结果在移动设备上的显示主要有 4 种方式:列表、版块、地图和增强现实。其中,增强现实是一种将真实世界信息和虚拟世界信息无缝集成的新技术,它将新闻信息叠加到用户周边环境的现场摄像机视图上^[61],然后在移动设备上显示。Lee 等人^[10]使用被推荐的新闻、最近发布的新闻和新闻类别 3 种列表来显示新闻;CityEvents^[49]使用版块和地图两种方式显示新闻,并分别用排序和颜色来区分新闻的重要性;Goh 等人^[61]利用列表、地图和增强现实 3 种方式展示基于位置的信息,通过比较 3 种方式的性能,指出列表的用户体验最好,而地图最不适合浏览新闻。总的说来,列表和版块的显示方式类似,版块相对更灵活;地图常用在基于位置的新闻推荐中,有利于将新闻和位置直观地联系起来;与其他显示技术相比,增强现实的实现过程比较复杂,视图范围有限,不适于显示大规模的新闻信息,但是其显示方式更生动形象,可以在一定程度上减少用户的主动认知行为。随着智能移动设备(如智能手机、智能穿戴设备)的发展和普及,增强现实将是未

来新闻显示技术的研究重点。

3.6 移动新闻推荐效用评价

通过验证移动新闻推荐系统的性能,并进一步调整推荐产生器以适应用户新闻偏好具有实际意义。数据集和评价指标的选择是效用评价的两个关键因素。

3.6.1 移动新闻推荐数据集

移动新闻推荐中数据集的获取一直是一个难点问题。因为考虑到用户隐私、安全性等问题,有效且公开的数据集很少,这也是移动新闻推荐研究面临的一大挑战。目前,移动新闻推荐研究中数据集来源通常分为以下几类:

(1) 模拟生成。这类数据集通过一定的规则算法自动生成,有利于进行复杂的或者大规模的移动新闻推荐,但是由于数据由给定的规则产生,不具有普遍意义和真实性,因此实验结果不具有太大的公信力和说服力。例如,Wang 等人^[42]采用的 MobileServices 数据集。

(2) 调研(调查问卷)。采用调查问卷的形式从特定群体中收集相关数据构成数据集。这种方式构成的数据集具有一定的真实性,但是一般数据集规模较小,不适用于大规模的复杂推荐环境,数据的普遍客观性也不高,且采用这种方式获取数据集的成本较高。如文献^[34]收集了 136 个本科生和研究生的调查问卷构成了实验数据集。

(3) 网上随机抓取。通过爬虫程序或抓取软件从网络上随机抓取数据构成实验数据集。这种方式获得的数据集虽然具有一定的客观性,也比较有说服力,但是存在脏数据和数据冗余的情况,可能会对实验结果造成一定的误差。例如,Chen 等人^[18]从谷歌新闻网站上利用爬虫程序抓取的新闻数据;Li 和 Chu 等人^[62-63]从 Yahoo! 今日模块上抓取的新闻数据集;以及 Yeung 等人^[9]抓取了 24 个 RSS 新闻网站的数据。

(4) 公开数据集。这种数据集客观性最强,也最有说服力,但是来源却是最少的。因为目前公开的能够用于推荐系统的数据集很少,且一般这些公开数据集都具有一定的针对性,即适用于专门的推荐领域,比如 MovieLens 数据集通常被用于电影推荐等领域。新闻领域公开数据集目前有以下两类:

① 传统新闻数据。kosarak 数据集^[64]是一个匈牙利在线新闻网站的点击流数据集,其 items 数目为 41 270, transaction 数目为 990 002, 平均 transaction 长度为 8.1, 最大 transaction 长度为 2498, 该数据

集现在多应用于数据挖掘中频繁项目集算法的测试中,它并没有对用户的点击行为进行过多的数据采集,只记录了用户的新闻点击流。

② 社交媒体数据。2012 KDD Cup 公开了一个腾讯微博数据集^[65]。它由 7 个文件组成,分别记录了训练集数据、测试集数据、用户档案信息、对象信息、用户活动信息、用户社交活动数据和用户关键字。该数据集规模较大,并在多个领域如:用户档案、社交图谱、对象类别等提供了丰富的信息;Friendfeed 数据集^[19]包含了从 2009 年 9 月 1 日到 2009 年 9 月 30 日期间由 111 284 位不同用户所发表的 1 641 531 条帖子内容。

(5) 非公开数据集。这类数据集通常都是服务提供商为了实现某一领域研究提供的,但是出于对用户隐私的保护,只用于特定的论文实验中,并没有公开。例如, Lee 等人^[14]使用了韩国移动服务提供商提供的智能无线服务 200 名注册用户的数据集。该数据集收集整理了从 2005 年 3 月到 2005 年 5 月这 3 个月的用户使用日志,这些日志包含了新闻、电影和餐馆的服务信息。

(6) 构造数据集。利用一些已有的公开数据集,给定一系列规则对各个数据集进行数据提取、补充,融合成一个新的数据集。这类数据集数据来源公开可靠,并且可以由研究者自由调整构造规则,灵活度较高,方便生成大规模数据。在无法获得满意的有说服力的数据集时,该方法是很适用的,但是在构造数据集的时候一定要注意规则的合理性。

同时采用几类数据集进行实验,可以提高实验结果的可信度。例如, Yang 等人^[34]同时采用了模拟数据和调研数据,前者是随机生成的,后者调研收集了 2004 年 1 月到 2005 年 8 月间, 136 个本科生和研究生的真实数据;文献^[30]同时使用了真实用户的调研档案数据和基于 IMDb 和 MovieLens 的构造数据集。

3.6.2 移动新闻推荐评价指标

移动新闻推荐的性能主要从推荐结果准确性、系统响应时间和用户体验 3 个方面来体现。其采用的评价指标多种多样,本文将评价指标基于功能分成了 3 类:(1) 信息检索领域指标,如 Recall、Precision 和 F-measure 等^[43];(2) 统计学指标,如 t 检验^[33]、MAE^[16]等;(3) 个性化评估指标,比如实时性、点击率、阅读率^[10]、用户满意度^[12]等。

为了进一步探讨移动新闻推荐中的评价指标,我们对表 6 中几个典型的移动新闻推荐系统性能进

行了定量分析. 由于目前移动新闻推荐领域公开数据集很少, 很难获取到一个普适的数据集, 且不同的评价指标具有不同的适用范围(如阅读率只适用于文本新闻). 因此我们基于不同的新闻数据集就同一评估指标比较, 分析了不同系统的实验性能:

(1) 点击率. MONERS^[10] 平均点击率是 23.4% (假设新闻平均长度为 5 页), 相比其相关研究点击率有所降低; 而 m-CCS^[16] 平均点击率是 11.2%, M-CRS^[17] 平均点击率达到 27%, 两者均提高了相关推荐的点击率.

(2) 阅读率(阅读时长). MONERS^[10] 提高了新闻推荐的阅读率, 其平均阅读率可以达到 45.4% (假设新闻平均长度为 5 页); M-CRS^[17] 提高了推荐的阅读时长, 有 2% 的用户平均花 50~60 min 阅读其推荐的博客.

(3) 收听率(收听时长). iNewsBox^[66] 的平均收听时长达到 29.61 s, 延长了 12.18%; EagleRadio^[20] 在 top-10 推荐中的收听率能达到 75% 左右.

(4) 精确度. VCR^[7] 在半自动模式下其平均推荐精确度达到 80% 左右, 远远超过随机推荐模式; InfoSlim^[22] 只是定性地分析了其技术可以提高推荐精确度, 并没有进行定量的实验分析.

(5) 实时性. ERS^[13] 可以在 1 min, 甚至 20 s 以内向用户推送地震消息, 远超过日本气象局播送公告的速度; VCR^[7] 提供的多种预取方式可以有效地降低推荐时延, 最短时延可以缩小到 5 s 以内, 但是时延与其推荐精确度成反比.

4 移动新闻推荐应用研究

随着移动新闻推荐的不断发展, 目前国内外研究者也提出了一些有价值的实际应用如: PPNews^[9]、M-CRS^[17] 等. 表 6 列举了一些典型的移动新闻推荐系统, 并根据它们应用的领域分为了两类: 传统新闻和社会化网络. 下面本文对这两类系统分别进行比较分析.

表 6 典型的移动新闻推荐系统应用

应用领域	新闻类型	典型系统	推荐方法	数据集	推荐效果
传统新闻	文本新闻	PPNews ^[9]	混合推荐	RSS 新闻网站数据和收集移动用户使用系统产生的数据	能成功推荐满足用户的实时新闻
		InfoSlim ^[22]	基于知识	BBC, CNN 的新闻源	提高了推荐精确度
		MONERS ^[10]	基于内容	韩国移动服务提供商提供的会员数据	提高了新闻的全文阅读率
		Daily Learner ^[11]	基于内容	300 个用户使用系统的数据	提高了用户感兴趣新闻的排序
	视频新闻	EPG ^[23]	混合推荐	DVB-H, IPTV 和互联网数据	提供了无缝流畅的用户体验
		VCR ^[7]	投票推荐	CNN 网站的新闻视频	提高了推荐精确度, 降低了延迟
音频新闻	EagleRadio ^[20]	混合推荐	搜狐 RSS 新闻源	提高了新闻收听率并保证了新闻的多样性	
	iNewsBox ^[66]	基于内容	CNN, ABC 和 USA Today RSS 源	延长了收听时间, 减少用户操作	
社会化网络	微博帖子	ERS ^[13]	基于内容	Twitter 数据	发现目标事件的准确率达到 96% 并能提高传播速度
	博客文章	M-CRS ^[17]	混合推荐	3000 个博客网址的内容	提高了博客点击率和阅读时长
		m-CCS ^[16]	混合推荐	co-RSS feeds 博客文章	提高了博客文章的点击率
	文本新闻	SCPPS ^[12]	基于内容	462 个用户使用系统 6 个月的数据	减少了传播延迟并提高了用户满意度

4.1 传统新闻

这一节主要讨论传统新闻中的一些典型应用, 大部分移动新闻推荐系统都可以归为这一类. 这些系统在移动用户浏览新闻网站或者接收定制新闻时, 向用户进行推荐. 本文将传统新闻又分为以下 3 类:

(1) 文本新闻

文本新闻在移动设备上的推荐技术发展迅速, 是移动新闻推荐研究中的主流. 文本新闻移动推荐实现的关键是新闻和偏好的表示方法、上下文感知推荐等. PPNews^[9] 采用 Super Peer-Mobile Peer 模式, 基于用户的上下文信息主动推送新闻. Mobile Peer 主要负责获取传感数据, 然后由 Super Peer 处

理. 大部分复杂运算都在 Super Peer 上进行, 因此可以延长移动设备的电池使用时间; InfoSlim^[22] 使用了语义技术来注释新闻项和用户偏好, 这样处理可以使得新闻项属性和用户档案的相似性进行比较, 不仅是通过词汇水平的基于余弦的方法来实现, 也是通过语义层面的基于本体论的方法来实现的; MONERS^[10] 综合了新闻文章属性(如重要性、时间)和用户关于类别及新闻内容的偏好进行新闻推荐, 提高了新闻全文阅读率; Daily Learner^[11] 采用隐性方式获取用户偏好, 对用户个人兴趣具有自适应性, 减少了需要传输的信息量, 并且帮助用户以最少的代价获取相关的信息.

上述推荐系统从多个方面研究了文本新闻的

推荐,但是它们也存在一些不足。例如,PPNews 没有考虑相似新闻的过滤,在响应时间上还需进一步改善,以满足音频、视频等时间敏感内容的推送;InfoSlim 实现过程需要构建和查询本体树,会增大时间开销,而且没有对结果进行评估验证;使用 MONERS 推荐新闻,其点击率有所降低;由于 Daily Learner 缺乏显示评分计算,其最后的推荐列表排序可解释性较差。

(2) 视频新闻

除了新闻标题或者视频附带标签,通常很难获取视频新闻其他的内容信息,因此目前的视频新闻推荐技术都比较简单,个性化不强。此外,移动视频新闻推荐系统除了对推荐结果有个性化的需求,还需要保证新闻浏览的顺畅性。EPG^[23]可以感知网络环境上下文然后自动切换其结构来适应不同的环境以提供有效流畅的服务。比如,在移动设备上,该推荐系统内核可以独立工作;而使用家庭网络的时候,它可以利用计算机资源转换成混合模式;VCR^[7]为移动设备设计了一个混合的用户界面,其利用用户的相关反馈进行基于内容的推荐策略,并设计和实现了一个视频内容感知预取方案以降低客户端的用户感知延迟。

(3) 音频新闻

事实上,音频新闻和文本新闻两者的推荐技术基本一致。因为音频新闻通常都是通过语音合成技术从文本新闻转化过来的,因此可以利用相应的文本新闻内容分析、建模,然后用于音频新闻推荐。目前,关于音频新闻在移动设备上的推荐研究还比较少。EagleRadio^[20]是一个语音新闻推荐系统,它通过一个语音合成器为用户读取新闻,并根据用户的收听历史、和用户具有相似品味的邻居以及他们的偏好向用户进行推荐。EagleRadio 可以遵循多个用户兴趣,并且保证了推荐新闻列表的多样性。iNewsBox^[66]通过将收集的文本新闻转换成音频新闻,然后推荐给不便于浏览新闻的移动用户,以提高用户体验。

4.2 社会化网络

随着 Web2.0 的发展,社会化网络服务得到越来越多的人使用。移动社会化网络信息量大、更新速度快,其传播的深度和广度都要高于传统新闻,目前也是新闻的一大来源。此外,移动社会化网为新闻推荐提供了丰富的附加信息(如用户个人档案、社交关系等)。例如,SCPPS^[12]将社交网络中挖掘的上下文和第三方网站的信息结合起来创建用户档案,并

与适当标记的网络订阅源相匹配进行推荐。虽然 SCPPS 可以有效地利用社会化网络中的上下文信息,但是它对用户依赖性较大,需要用户在移动设备上安装系统并提供授权证书,系统才能访问用户的个人档案。因此,使用不够便利,也会过多占用移动设备有限的资源。

移动社会化网络中,用户偏好通常由长期偏好和短期偏好两部分组成。长期偏好由用户个人兴趣决定,通过长期的浏览历史反映。短期偏好则受到社会化网络中的热点话题或流行趋势影响。为了帮助移动用户快速了解近期内社会化网络中的热点话题,研究者们提出了一些解决方案。ERS^[13]具有较强的现实意义,它通过监测微博帖子来发现目标事件发生位置的中心和运行轨迹,因此可以迅速探测地震,并向注册用户发送邮件。实验证明其发送通知的速度比日本气象局播送的公告快的多;m-CCS^[16]根据对时间敏感的博客流行趋势来预测最近的流行博客主题,由于不同的移动用户对于流行博客主题有不同的偏好,因此进一步分析移动用户的浏览日志来确定其兴趣,然后将两者结合起来导出用户偏爱的博客话题和文章;M-CRS^[17]可以识别流行博客主题,并为新的博客文档计算用户偏好,然后进行相关博客推送服务。

从上述研究可知,和传统新闻不同,社会化网络是通过用户的关注、分享传播新闻的,大多数人分享或者关注的内容就会成为流行的趋势,形成热点话题。因此,移动社会化网络中用户的偏好是由流行主题和个人兴趣共同决定的。

5 移动新闻推荐中的重、难点问题

这部分我们主要讨论一下当前移动新闻推荐中存在的一些重点和难点问题,并分析今后移动新闻推荐研究的方向和热点。其中,移动新闻推荐中存在的重点问题主要包括以下 5 个方面:

(1) 移动新闻推荐的隐私问题

一个成功的移动新闻推荐系统需要建立相应的隐私保护机制,因为其涉及的用户个人信息(如用户档案、上下文等)较多。在移动新闻推荐系统中提供用户个人信息安全保障及保密措施,可以有效地解决隐私问题,提高用户对推荐系统的信任度。例如,文献[28]在用户给资源评级的时候使用假名以保护隐私。传统的离线隐私保护机制已经不能满足当前移动新闻推荐中实时获取、实时推荐的需求,因此,

在线的动态隐私保护机制是今后的研究重点. 此外, 混合保护机制以及根据信息重要程度划分等级的隐私保护策略也是值得研究的方向.

(2) 移动新闻推荐的安全问题

移动网络是个复杂的环境, 在对移动用户进行新闻推荐的时候需要考虑安全问题, 防止用户资料的恶意泄露和窃取, 以及一些恶意侵入, 破坏推荐结果, 将一些不良信息推荐给用户. 目前主要有两个研究方向: 建立鲁棒性移动新闻推荐系统和恶意节点监测. 其中, 恶意节点监测可以从两个方面来实现: ① 用户档案监测, 从根源上杜绝恶意节点加入推荐系统; ② 用户行为监测. 但是仅通过监测用户档案发现恶意节点比较困难, 因为其在建立档案时通常具有欺骗行为, 不易被系统发现. 因此, 目前的研究主要是针对用户行为进行监测的, 比如, 基于用户之间的信任行为建立声誉系统进行推荐^[35] 或者利用移动用户之间的社会化关系构建信任网络^[26,41,67], 以保证推荐结果的安全性和可信度. 其中, 后者实现的关键就是信任节点的选择和信任网络的构建策略, 还需进一步研究.

(3) 移动新闻推荐的计算复杂度和效率问题

在当前大数据时代的背景下, 新闻信息(如社交媒体)的爆炸式增长为移动新闻推荐带来了前所未有的挑战. 系统的可扩展性和实时性成为目前移动新闻推荐中亟待解决的问题. 大部分推荐系统都是将数据集分成了训练集和测试集两部分, 需要在线反复训练以得到推荐结果, 相比之下文献^[63]提供了一个轻量计算消耗的离线模型, 有利于提高推荐效率和减少资源损耗. 此外, 基于分布式计算技术(如 Map-Reduce)或云服务平台进行移动新闻推荐, 可以提高系统的可扩展性, 减少网络开销, 加速推荐过程和提高推荐效率^[52], 值得进一步研究. 但是在分布式处理中, 如何对新闻推荐任务进行有效地分割及融合以保证结果的完整性和准确性, 在未来的工作中还需进行重点研究.

(4) 移动社会化网络新闻推荐

随着移动社会化网络的迅速发展, 信息的传播深度和广度都达到了前所未有的规模, 因此基于移动社会化网络的新闻推荐也是目前的研究重点之一. 和传统媒体不同, 移动社会化网络上用户活跃、互动频繁、易形成流行信息, 而文本内容较短、数据稀疏, 因此热点话题推荐和如何解决短文本数据稀疏问题是移动社会化网络新闻推荐中两个重点的研究方向. 实现热点话题推荐的两个关键技术是: 话题

发现和话题跟踪. 目前, 研究者主要利用一些主题模型来学习、发现热点话题^[19]. 话题跟踪是基于内容或位置预测的话题推荐, 即跟踪发现话题的内容变化或者地理上的传播方向^[13]. 其中后者结合了移动用户的位置上下文, 目前相关研究还比较少, 应该是未来研究的趋势. 而短文本数据稀疏问题, 可以从下面两个方面入手: ① 利用移动社会化网络上丰富的可挖掘信息(如社会化关系)来弥补文本内容不足^[12]; ② 结合多种媒体资源, 如利用传统新闻来丰富短文本的语义信息^[40]. 其中, 后者的相关研究还比较少, 但它为解决短文本问题提供了新的思路, 值得进一步研究.

(5) 基于上下文的移动新闻推荐

移动新闻推荐通常受到不同类型上下文, 不同程度的影响^[9]. 目前基于新闻新颖性、用户移动轨迹以及联合多种上下文的相关研究还比较少, 因此如何根据用户偏好和新闻新颖性推荐新闻^[36], 利用轨迹预测进行新闻推荐^[51] 和多维上下文移动新闻推荐^[9] 均是未来的重点研究方向. 移动环境的多变性增加了上下文感知的难度, 既要防止引入噪音数据(不必要的上下文信息)、增加计算量, 又要避免上下文的缺失、影响推荐决策. 因此上下文选择策略、上下文推理技术(以补充缺失的上下文信息)以及提高系统的容错性是关键, 值得进一步的深入研究. 其中, 上下文选择策略可以从经验选择和系统自动选择两个方向来考虑^[12,52]; 而上下文推理技术目前主要有贝叶斯网络、本体或规则推理等^[2]. 此外如何利用丰富的上下文信息发现用户的潜在需求, 将目前简单的满足用户需求发展为发现用户需求, 也是未来上下文感知的移动新闻推荐目标.

目前, 移动新闻推荐中还存在一些难点问题, 如数据集获取、效用评估等, 主要有以下几点:

(1) 移动新闻推荐数据集获取困难

目前很少有公开可用的移动新闻推荐数据集. 出于隐私保护, 移动用户相关数据很难获取. 从大部分文献来看, 其实验数据都是来自对有限数量用户的调查, 或者某个地区运营商提供的部分用户的信息, 数据量少, 范围覆盖不广, 具有片面性, 实验结果不够有说服力, 而且通过对用户进行调查采集数据的花费比较多, 对于小型研究团队或个人研究者来说实验成本太高.

(2) 移动新闻推荐服务质量评估困难

评价移动新闻推荐性能, 验证数据集和评价指标的选择是关键. 但是如何衡量评估移动新闻推荐

服务的质量,还没有找到一个客观、公正、统一的量化评估标准.因此,各种各样的指标被应用到移动新闻推荐中进行性能评估,主要有信息检索领域指标,统计学指标,以及个性化评估指标.在缺乏有效验证数据集的情况下,很多推荐系统需要用户主动反馈^[34],比如给所推荐的新闻打分或者评级来反映用户对推荐结果的兴趣和满意度,或者通过对比用户在推荐前后的新闻浏览情况来衡量推荐性能^[11].但是现实情况中用户通常很少主动参与信息反馈.因此,一般通过开发推荐原型系统,并聘请志愿者使用系统以收集用户反馈信息.这种方式获取的验证数据集一般用户规模很小(少于300),实验成本高,耗时长,而且基于反馈的性能评估比较主观.

(3) 缺乏对移动新闻推荐结果的解释

随着移动新闻推荐的发展,其推荐过程已经越来越复杂,对用户来说也越来越不透明化.加之移动设备和新闻的自身特性,推荐过程绝大部分工作(例如偏好的隐式获取,新闻对象的加工和处理,产生候选推荐集等)都在服务器端完成.因此,用户往往很难理解最终的推荐结果.针对推荐结果作出合理有效的解释或者增加系统的透明性和可控性(例如向用户开放用户偏好模型^[27]),有利于提高用户对系统的信任度和满意度.

(4) 缺乏形式化分析、描述和验证

在移动新闻推荐中几乎没有文献给出了推荐算法的形式化描述和分析,也没有从理论的角度上验证所提出算法的收敛速度或者误差等.一般文献都是在设计实验的基础上,分析了算法的收敛速度、空间复杂度等,即对实验结果进行了验证分析,表明了算法的有效性.

6 总 结

随着手机等移动设备的日益普及,无线通信技术的不断发展,越来越多的人开始使用移动设备,通过无线网络浏览新闻等信息.由于新闻老化明显、数量庞大以及移动设备具有便携性、私有性、有限屏幕尺寸和计算存储能力等特性,信息过载、新闻展示、上下文感知以及社会化网络推荐都成为移动新闻推荐系统需要解决的问题.数十年来,推荐系统在学术界和工业界已经取得了一定的研究成果,但是移动新闻推荐的相关研究还需进一步发展.现在的移动新闻推荐还存在安全隐私、计算效率、数据集获取、推荐服务质量评估和形式化描述分析等重难点问

题,也是今后的研究热点方向.

参 考 文 献

- [1] Adomavicius G, Tuzhilin A. Towards 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
- [2] Wang Li-Cai, Meng Xiang-Wu, Zhang Yu-Jie. Context-aware recommender systems. *Journal of Software*, 2012, 23(1): 1-20(in Chinese)
(王立才, 孟祥武, 张玉洁. 上下文感知推荐系统. *软件学报*, 2012, 23(1): 1-20)
- [3] Meng Xiang-Wu, Hu Xun, Wang Li-Cai, Zhang Yu-Jie. Mobile recommender systems and their applications. *Journal of Software*, 2013, 24(1): 91-108(in Chinese)
(孟祥武, 胡勋, 王立才, 张玉洁. 移动推荐系统及其应用. *软件学报*, 2013, 24(1): 91-108)
- [4] Baltrunas L, Ludwig B, Peer S. Context relevance assessment and exploitation in mobile recommender systems. *Personal and Ubiquitous Computing*, 2012, 16(5): 507-526
- [5] Li Lei, Wang Dingding, Li Tao. SCENE: A scalable two-stage personalized news recommendation system//*Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*. Beijing, China, 2011: 125-134
- [6] He Yong-Chun, Zhou Yan-Quan, Wang Cong. Research on user's interests in mobile personalized news service system//*Proceedings of the 11th Chinese Artificial Intelligence Academic Annual Meeting*. Wuhan, China, 2005: 1460-1465(in Chinese)
(何永春, 周延泉, 王枫. 移动个性化新闻推荐系统中用户兴趣的研究//第十一届中国人工智能学术年会. 武汉, 中国, 2005: 1460-1465)
- [7] Lee J S A, Jayant N. Mixed-initiative multimedia for mobile devices: Design of a semantically-relevant low-latency system for news video recommendations//*Proceedings of the IEEE Southeastcon 2009*. Atlanta, USA, 2009: 267-272
- [8] Cantador I, Bellogín A, Castells P. Ontology-based personalised and context-aware recommendations of news items//*Proceedings of the International Conference on Web Intelligence and Intelligent Agent Technology*. Sydney, Australia, 2008: 562-565
- [9] Yeung K F, Yang Y Y, Ndzi D. A proactive personalised mobile recommendation system using analytic hierarchy process and Bayesian network. *Journal of Internet Services and Applications*, 2012, 3(2): 195-214
- [10] Lee H J, Park S J. MONERS: A news recommender for the mobile web. *Expert Systems with Applications*, 2007, 32(1): 143-150
- [11] Billsus D, Pazzani M J, Chen J. A learning agent for wireless news access//*Proceedings of the 5th International Conference on Intelligent User Interfaces*. New Orleans, US, 2000: 33-36

- [12] Karapantelakis A, Maguire G Q. Utilizing social context for providing personalized services to mobile users//Proceedings of the 5th European Conference on Smart Sensing and Context. Passau, Germany, 2010: 28-41
- [13] Sakaki T, Okazaki M, Matsuo Y. Earthquake shakes Twitter users real-time event detection by social sensors//Proceedings of the 19th International Conference on World Wide Web. Raleigh, USA, 2010: 851-860
- [14] Lee H J, Choi J Y, Park S J. Context-aware recommendations on the mobile web//Proceedings of the OTM Confederated International Workshops and Posters. Agia Napa, Cyprus, 2005: 142-151
- [15] Lee Wei-Po, Lu Cheng-Che. Customising WAP-based information services on mobile networks. *Personal and Ubiquitous Computing*, 2003, 7(6): 321-330
- [16] Liu Duen-Ren, Tsai Pei-Yun, Chiu Po-Huan. Personalized recommendation of popular blog articles for mobile applications. *Information Sciences*, 2011, 181(9): 1552-1572
- [17] Chiu Po-Huan, Kao Gloria Yi-Ming, Lo Chi-Chun. Personalized blog content recommender system for mobile phone users. *International Journal of Human-Computer Studies*, 2010, 68(8): 496-507
- [18] Chen Chih-Ming. Intelligent location-based mobile news service system with automatic news summarization. *Expert Systems with Applications*, 2010, 37(9): 6651-6662
- [19] Song Shuang-Yong, Li Qiu-Dan. Micro-blogging information recommendation system for mobile client. *Computer Science*, 2011, 38(11): 137-139(in Chinese)
(宋双永, 李秋丹. 面向移动终端的微博信息推荐方法. *计算机科学*, 2011, 38(11): 137-139)
- [20] Chen Wei, Zhang Li-Jun, Chen Chun, et al. A hybrid phonic web news recommender system for pervasive access//Proceedings of the International Conference on Communications and Mobile Computing. Yunnan, China, 2009: 122-126
- [21] Zhang Rui-Hua, Zhou Yan-Quan, Wang Cong, Li Lei. The research of news recommendation service based on mobile off-line reading system. *Journal of Beijing University of Posts and Telecommunications*, 2006, 29(6): 21-24(in Chinese)
(张瑞华, 周延泉, 王枫, 李蕾. 移动终端离线浏览系统的新闻推荐服务研究. *北京邮电大学学报*, 2006, 29(6): 21-24)
- [22] Gao Feng, Li Yuhong, Han Li. InfoSlim: An ontology-content based personalized mobile news recommendation system//Proceedings of the 5th International Conference on Wireless Communications, Networking and Mobile Computing. Beijing, China, 2009: 1-4
- [23] Zhu Chuan, Ma Matthew, Tan Chunguang. A recommender handoff framework for a mobile device//Proceedings of the 1st International Workshop on Mobile Multimedia Processing. Tampa, USA, 2008: 137-153
- [24] Liu J, Dolan P, Pedersen E R. Personalized news recommendation based on click behavior//Proceedings of the 14th ACM International Conference on Intelligent User Interfaces. Hong Kong, China, 2010: 31-40
- [25] IJntema W, Goossen F. Ontology-based news recommendation //Proceedings of the 2010 International Conference on Extending Database Technology. Lausanne, Switzerland, 2010: 1-6
- [26] Wei D, Zhou T, Cimini G. Effective mechanism for social recommendation of news. *Physica A: Statistical Mechanics and Its Applications*, 2011, 390(11): 2117-2126
- [27] Ahn J, Brusilovsky P, Grady J, et al. Open user profiles for adaptive news systems; Help or harm?//Proceedings of the 16th International World Wide Web Conference. Banff, Canada, 2007: 11-20
- [28] Resnick P, Iacovou N, Suchak M, et al. GroupLens: An open architecture for collaborative filtering of netnews//Proceedings of the ACM Conference on Computer Supported Cooperative Work. Chapel Hill, USA, 1994: 175-186
- [29] Das A, Datar M, Garg A. Google news personalization; Scalable online collaborative filtering//Proceedings of the 16th International World Wide Web Conference. Banff, Canada, 2007: 271-280
- [30] Cantador I, Bellogin A, Castells P. A multilayer ontology-based hybrid recommendation model. *AI Communications*, 2008, 21(2-3): 203-210
- [31] Claypool M, Gokhale A, Miranda T, et al. Combining content-based and collaborative filters in an online newspaper//Proceedings of the ACM SIGIR Workshop on Recommender Systems; Algorithms and Evaluation. Berkeley, US, 1999: 15-22
- [32] Costa A, Guizzardi R, Guizzardi G, Filho J. CORES: Context-aware, ontology-based recommender system for service recommendation//Proceedings of the 19th International Conference on Advanced Information Systems Engineering. Trondheim, Norway, 2007
- [33] Lee S K, Cho Y H, Kim S H. Collaborative filtering with ordinal scale-based implicit ratings for mobile music recommendations. *Information Sciences*, 2010, 180(11): 2142-2155
- [34] Yang Wan-Shiou, Cheng Hung-Chi, Dia Jia-Ben. A location-aware recommender system for mobile shopping environments. *Expert Systems with Applications*, 2008, 34(1): 437-445
- [35] Yan Z, Zhang P, Deng R H. TruBeRepec: A trust-behavior-based reputation and recommender system for mobile applications. *Personal and Ubiquitous Computing*, 2012, 16(5): 485-506
- [36] Gabrilovich E, Dumais S, Horvitz E. Newsjunkie: Providing personalized newsfeeds via analysis of information novelty//Proceedings of the 13th International Conference on World Wide Web. New York, USA, 2004: 482-490
- [37] Wu Y, Ding Y, Wang X. Topic based automatic news recommendation using topic model and affinity propagation//Proceedings of the 9th International Conference on Machine Learning and Cybernetics. Qingdao, China, 2010: 11-14

- [38] Yin H, Sun Y, Cui B, et al. LCARS: A location-content-aware recommender system//Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Chicago, USA, 2013; 221-229
- [39] Son Jeong-Woo, Kim A-Yeong, Park Seong-Bae. A location-based news article recommendation with explicit localized semantic analysis//Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval. Dublin, Ireland, 2013; 293-302
- [40] Abel F, Gao Q, Houben G J. Analyzing user modeling on Twitter for personalized news recommendations//Proceedings of the 19th International Conference User Modeling, Adaption, and Personalization. Girona, Spain, 2011; 1-12
- [41] Wang Yu-Xiang, Qiao Xiu-Quan, Li Xiao-Feng, et al. Research on context-awareness mobile SNS service selection mechanism. Chinese Journal of Computers, 2010, 33(11): 2126-2135(in Chinese)
(王玉祥, 乔秀全, 李晓峰等. 上下文感知的移动社交网络服务选择机制研究. 计算机学报, 2010, 33(11): 2126-2135)
- [42] Wang Li-Cai, Meng Xiang-Wu. A MAUT approach to elicitation of contextual user preferences. Advances in Information Sciences and Service Sciences, 2012, 4(5): 98-105
- [43] Li Lei, Zheng Li, Yang Fan, et al. Modeling and broadening temporal user interest in personalized news recommendation. Expert Systems with Applications, 2014, 41(7): 3168-3177
- [44] Shi Yan-Cui, Meng Xiang-Wu, Zhang Yu-Jie, Wang Li-Cai. Adaptive learning approach of contextual mobile user preferences. Journal of Software, 2012, 23(10): 2533-2549 (in Chinese)
(史艳翠, 孟祥武, 张玉洁, 王立才. 一种上下文移动用户偏好自适应学习方法. 软件学报, 2012, 23(10): 2533-2549)
- [45] Abowd G, Dey A, Brown P, et al. Towards a better understanding of context and context-awareness//Proceedings of the 1st International Symposium on Handheld and Ubiquitous Computing. Karlsruhe, Germany, 1999; 304-307
- [46] Karkali M, Pontikis D, Vazirgiannis M. Match the news: A firefox extension for real-time news recommendation//Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval. Dublin, Ireland, 2013; 1117-1118
- [47] Kang Dong-Yup, Han Dong-Kyun, Sim Gyumin, et al. PADAC²: Real-time news recommendation system with heterogeneous social footprints//Proceedings of the 16th International Conference on Advanced Communication Technology. PyeongChang, Korea, 2014; 79-82
- [48] Yin H, Cui B, Chen L, et al. A temporal context-aware model for user behavior modeling in social media systems//Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data. Snowbird, USA, 2014; 1543-1554
- [49] Rosa P M P, Rodrigues J J P C, Basso F. A weight-aware recommendation algorithm for mobile multimedia systems. Mobile Information Systems, 2013, 9(2): 139-155
- [50] Bao Jie, Mokbel M F, Chow Chi-Yin. GeoFeed: A location-aware news feed system//Proceedings of the 28th International Conference on Data Engineering. Arlington, US, 2012; 54-65
- [51] Xu Wenjian, Chow Chi-Yin, Yiu Man Lung, et al. MobiFeed: A location-aware news feed system for mobile users//Proceedings of the 20th International Conference on Advances in Geographic Information Systems. Redondo Beach, USA, 2012; 538-541
- [52] Mo Yijun, Chen Jianwen, Xie Xia, et al. Cloud-based mobile multimedia recommendation system with user behavior information. IEEE Systems Journal, 2014, 8(1): 184-193
- [53] Kwak H, Lee C, Park H, et al. What is Twitter, a social network or a news media?//Proceedings of the 19th International Conference on World Wide Web. Raleigh, USA, 2010; 591-600
- [54] Shima H, Youb K H, Lee J K, et al. Why do people access news with mobile devices? Exploring the role of suitability perception and motives on mobile news use. Telematics and Informatics, 2015, 32(1): 108-117
- [55] Lin Chen, Xie Runquan, Li Lei, et al. PRemiSE: Personalized news recommendation via implicit social experts//Proceedings of the 21st ACM International Conference on Information and Knowledge Management. Maui, USA, 2012; 1607-1611
- [56] Wang J, Li Q, Chen Y. User comments for news recommendation in social media//Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval. Geneva, Switzerland, 2010; 881-882
- [57] Lee Won-Jo, Oh Kyo-Joong, Lim Chae-Gyun. User profile extraction from Twitter for personalized news recommendation //Proceedings of the 16th International Conference on Advanced Communication Technology. Pyeongchang, Korea, 2014; 779-783
- [58] Tavakolifard M, Gulla J A, Almeroth K C, et al. Tailored news in the palm of your hand: A multi-perspective transparent approach to news recommendation//Proceedings of the 22nd International Conference on World Wide Web Companion. Rio de Janeiro, Brazil, 2013; 305-308
- [59] Otterbacher J, Radev D, Kareem O. News to go: Hierarchical text summarization for mobile devices//Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. Seattle, USA, 2006; 589-596
- [60] Lee E, Kang J, Park J, et al. Scalable web news adaptation to mobile devices using visual block segmentation for ubiquitous media services//Proceedings of the International Conference on Multimedia and Ubiquitous Engineering. Seoul, Korea, 2007; 620-625
- [61] Goh D H L, Lee C S, Razikin K. Comparative evaluation of interfaces for presenting location-based information on mobile devices//Proceedings of the 13th International Conference on Asia-Pacific Digital Libraries. Beijing, China, 2011; 237-246

- [62] Li L, Chu W, Langford J, Schapire R E. A contextual-bandit approach to personalized news article recommendation// Proceedings of the 19th International Conference on World Wide Web. Raleigh, USA, 2010; 661-670
- [63] Chu Wei, Park Seung-Taek. Personalized recommendation on dynamic content using predictive bilinear models// Proceedings of the 18th International Conference on World Wide Web. Madrid, Spain, 2009; 691-700
- [64] Bodon F. A fast apriori implementation//Proceedings of the IEEE ICDM Workshop on Frequent Itemset Mining. Melbourne, USA, 2003; 14-23
- [65] Niu Yanzhi, Wang Yi, Sun Gordon, et al. The Tencent dataset and KDD-Cup'12. KDD-Cup Workshop. Beijing, China, 2012
- [66] Xie Yanan, Chen Liang, Jia Kunyang, et al. iNewsBox: Modeling and exploiting implicit feedback for building personalized news radio//Proceedings of the 22nd ACM International Conference on Information and Knowledge Management. San Francisco, USA, 2013; 2485-2488
- [67] Qiao Xiu-Quan, Yang Chun, Li Xiao-Feng, Chen Jun-Liang. A trust calculating algorithm based on social networking service users' context. Chinese Journal of Computers, 2011, 34(12): 2403-2413(in Chinese)
(乔秀全, 杨春, 李晓峰, 陈俊亮. 社交网络服务中一种基于用户上下文的信任度计算方法. 计算机学报, 2011, 34(12): 2403-2413)



MENG Xiang-Wu, born in 1966, Ph. D., professor, Ph. D. supervisor. His research interests include web services, user requirements and recommender services.

CHEN Cheng, born in 1988, Ph. D. candidate. Her research interests include data mining and mobile recommendation.

ZHANG Yu-Jie, born in 1969, associate professor. Her research interests include intelligent information processing, communication software and web services.

Background

With the increasing use of mobile devices and the development of wireless web, more and more people prefer reading news on the handheld devices anytime and anywhere. To help mobile users improve the utilization of time fragments and choose interested news from huge amount of news sources rapidly, researchers make effort to study the mobile personalization news services and recommender systems. This paper gives a comprehensive overview of the current state of the art and provides a thorough analysis and comparison of the latest studies on mobile news recommendation, including the compare with web news recommendation

and mobile recommendation, general framework, key techniques of mobile news recommendation, typical applications and challenges in this field.

This work is supported by the National Natural Science Foundation of China under Grant No.60872051 and the Mutual Project of Beijing Municipal Education Commission. These projects mainly aim to study the methods of user requirement modeling and preference acquisition. This paper surveys the current research on mobile news recommendation which is a further study based on news preference acquisition.