

基于用户轨迹数据的移动推荐系统研究*

孟祥武^{1,2}, 李瑞昌^{1,2}, 张玉洁^{1,2}, 纪威宇^{1,2}



¹(智能通信软件与多媒体北京市重点实验室(北京邮电大学),北京 100876)

²(北京邮电大学 计算机学院,北京 100876)

通讯作者: 孟祥武, E-mail: mengxw@bupt.edu.cn

摘要: 近年来,随着移动智能设备的普及,移动社交网络方兴未艾,用户习惯和朋友分享自己的精彩经历,因此产生了大规模具有时空属性的用户轨迹数据.从狭义的角度来看,轨迹数据是指连续采样的 GPS 数据.从广义的角度来看,在时空域存在连续性的序列,都可以称作轨迹.例如:在社交网络上的用户签到序列就可以认为是粗粒度的轨迹数据.广义轨迹数据具有时空异构性、连续与离散并存、时空项目的层次性不明显和分类不明确等特点,但是相比于 GPS 轨迹数据,广义轨迹数据来源广泛,蕴含丰富的信息,这给传统的移动推荐系统带来了巨大的机遇.与此同时,广义轨迹数据规模大、结构丰富,这也给传统的移动推荐系统带来了巨大的挑战,如何利用广义用户轨迹数据来提升移动推荐系统的性能,已成为学术界和产业界共同关注的重要课题,以轨迹数据特征作为切入点,对近年来基于广义用户轨迹数据的移动推荐系统的主要模型方法和推荐评价指标进行了系统综述,阐述了与传统移动推荐系统的联系和区别.最后,对基于广义用户轨迹数据的移动推荐系统有待深入研究的难点和发展趋势进行了分析和展望.

关键词: 广义轨迹;推荐系统;应用;综述

中图法分类号: TP311

中文引用格式: 孟祥武,李瑞昌,张玉洁,纪威宇.基于用户轨迹数据的移动推荐系统研究.软件学报,2018,29(10):3111-3133.
<http://www.jos.org.cn/1000-9825/5608.htm>

英文引用格式: Meng XW, Li RC, Zhang YJ, Ji WY. Survey on mobile recommender systems based on user trajectory data. Ruan Jian Xue Bao/Journal of Software, 2018,29(10):3111-3133 (in Chinese). <http://www.jos.org.cn/1000-9825/5608.htm>

Survey on Mobile Recommender Systems Based on User Trajectory Data

MENG Xiang-Wu^{1,2}, LI Rui-Chang^{1,2}, ZHANG Yu-Jie^{1,2}, JI Wei-Yu^{1,2}

¹(Beijing Key Laboratory of Intelligent Telecommunications Software and Multimedia (Beijing University of Posts and Telecommunications), Beijing 100876, China)

²(School of Computer Science, Beijing University of Posts and Telecommunications, Beijing 100876, China)

Abstract: In recent years, with the popularity of mobile smart devices, location based social networks are on the rise. Users trend to share their wonderful experiences with their friends, resulting in producing large-scale user trajectory with temporal and spatial attributes. From a narrow perspective, the trajectory data refers to continuously sampled GPS data only. From a broad perspective, it can be called trajectory data as long as the data has sequential characteristic. Thus, the check-ins, acquired from a social network, can also be considered coarse-grained trajectory data. The generalized trajectory data has the characteristics of spatiotemporal heterogeneity, continuous and discrete coexistence, and containing temporal-spatial items with unclear hierarchy and classification. However, compared to the GPS trajectory data, the generalized trajectory data source is extensive and contains rich information, which brings great opportunity to the traditional mobile recommender system. At the same time, the generalized trajectory data has big scale and diversity

* 基金项目: 北京市教育委员会共建项目专项

Foundation item: The Mutual Project of Beijing Municipal Education Commission, China

收稿时间: 2017-05-15; 修改时间: 2018-03-27; 采用时间: 2018-05-17; jos 在线出版时间: 2018-06-07

CNKI 网络优先出版: 2018-06-07 14:53:51, <http://kns.cnki.net/kcms/detail/11.2560.TP.20180607.1453.009.html>

structure, which also presents great challenges to the system. It has become an important issue how to use the generalized trajectory data to improve the performance of mobile recommender system in academia and industry. This paper takes the trajectory data characteristics as the focal point to analyze and survey main recommender methods and evaluation metrics based on generalized user trajectory data. Further, it expounds the relationships and differences between traditional mobile recommender systems and the mobile recommender systems based on user trajectory data. Finally, the paper discusses the difficulty and development trend of mobile recommender systems based on generalized user trajectory.

Key words: generalized trajectory; recommender system; application; survey

随着下一代网络技术的飞速发展,移动通信网络与计算机网络逐渐融合,形成了丰富多彩的移动网络服务^[1].国际电联的 2016 版 ICT 事实与数字显示,移动电话的覆盖现已几近无处不在.据估计,约 95% 的全球人口(约 70 亿人)已得到 2G 蜂窝移动网络的覆盖.在过去的 3 年间,先进移动宽带网络(LTE)的发展如火如荼,目前已覆盖近 40 亿人,相当于全球人口的 53%^[2].使用 LTE 的每部智能手机都具有产生数字轨迹的功能,因此,手机可以当作移动传感器,可以实时监测移动用户的实际位置.用户在现实世界中移动,移动设备可以采集用户的运动轨迹,相关的应用可以利用轨迹数据来提高用户的服务质量.伴随着如 GPS 和无线传感等技术的快速发展,用户可以轻松地获取自己的当前位置信息,为用户进行位置共享提供了支撑.除此之外,社交网络(LBSNs)的兴起也激发了广大用户分享个人经历的热情,产生了许多与时空位置相关的社交数据,如旅游历程、消费点评等^[3].

丰富的轨迹数据使得挖掘用户时空和序列偏好并生成精准推荐结果成为可能,因而,基于用户轨迹数据的移动推荐已成为重要的推荐应用^[4,5].由于轨迹数据具有时序性,用户轨迹序列中不同位置节点具有不同的含义,不同位置节点的顺序对挖掘用户偏好的作用和影响也不相同,因此,与传统的移动推荐有明显区别.当前,基于用户轨迹数据的移动推荐的核心是对轨迹数据进行深度信息挖掘^[6],获取轨迹数据的时空语义特性和轨迹相关的用户属性、社会属性等信息,综合考虑各种信息生成推荐结果^[7-9].

本文对基于用户轨迹数据的移动推荐系统研究进行综述,相应的逻辑结构图如图 1 左侧所示.

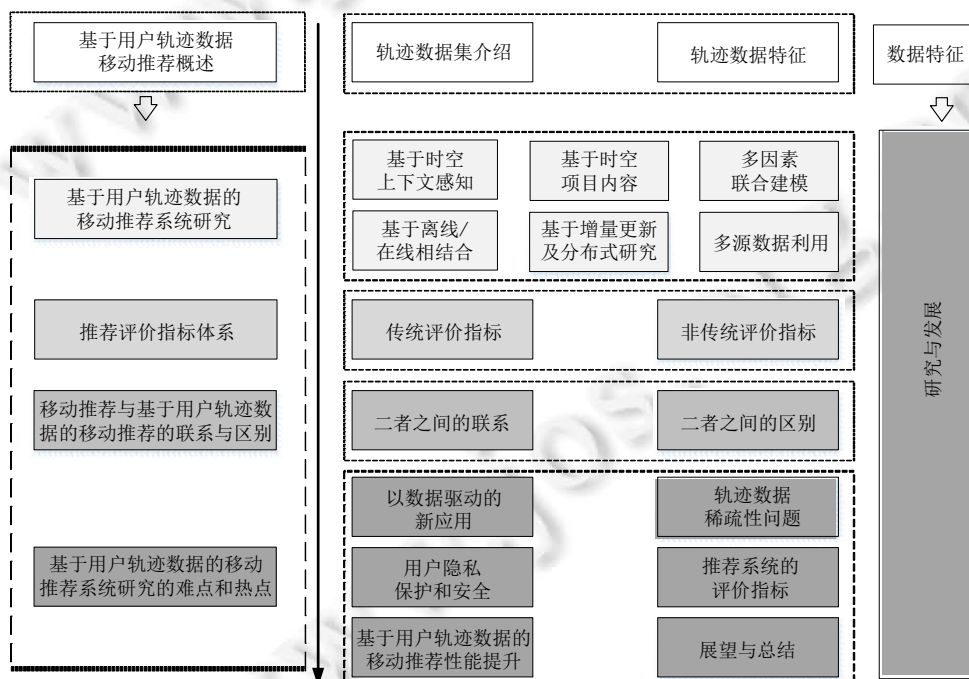


Fig.1 Survey framework of mobile recommender systems based on user trajectory data (MRS-TD)

图 1 基于用户轨迹数据的移动推荐系统研究综述框架

本文逻辑上主要分为 5 个部分:第 1 节为系统概述,主要对轨迹数据分类情况进行分析,展示当前主要的轨迹数据集,分析轨迹数据的特点和带来的机遇和挑战,第 2 节重点介绍基于用户轨迹数据的移动推荐系统研究情况,涉及如何利用轨迹数据以及如何生成推荐结果,第 3 节介绍基于用户轨迹数据的移动推荐系统评价指标,第 4 节对移动推荐系统与基于用户轨迹数据的移动推荐系统的联系和区别进行总结分析,第 5 节对有待深入研究的难点和发展趋势进行分析和展望,最后是结束语,由图 1 右侧还可以看出:轨迹数据在基于用户轨迹数据的移动推荐中处于一个非常重要的地位,其使用的推荐技术与轨迹数据特征息息相关。

1 基于用户轨迹数据的移动推荐系统概述

基于用户轨迹数据的移动推荐系统是指充分利用轨迹数据特有的时空和序列属性,采用机器学习、大数据分析等手段提取轨迹数据的语义信息和移动对象的概貌信息,如轨迹数据的地理语义、移动对象的移动和停止状态、个性偏好和社会关系信息等,进而生成符合用户行为时空特征的推荐结果。本文提到的轨迹数据是一种广义的轨迹数据,不仅包含时空信息,还包含丰富的用户行为和项目内容信息,如用户分享的地理图片、用户对项目的评分和评价以及项目的属性和分类等。广义轨迹采集的数据可能是文本、图片或数值型的,数据的结构可能是结构化、半结构化或非结构化的,因而广义的轨迹数据是异构的,总体上来看具有空间分布不均匀、时间间隔不等、项目层次性不明显和分类不明确等特征。依据这种广义的轨迹数据生成的推荐结果,最大的特点是与时间和空间有密切关联。为表述方便,在不产生歧义的情况下,后续部分提到的轨迹数据就是指广义轨迹数据。

1.1 移动轨迹数据的分类和当前的轨迹数据集

轨迹就是一种印迹,总体可分为隐含轨迹、地理轨迹和时空轨迹^[10]。隐含轨迹表示一种演化,这种演化与实际运动无关,如从教育界到研究机构工作,这种隐含轨迹的含义与移动对象在抽象空间中的想法相关。地理轨迹描述从一个地方到另外一个地方,如“我从北京到天津,然后又去了济南”,虽然这种轨迹没有使用空间坐标来定义,但是具有明显的地理空间属性。时空轨迹表示移动对象在地理空间中的位置变化,如果使用空间坐标来表示移动对象的位置,则轨迹具有时空属性,本文涉及的移动轨迹数据就属于时空轨迹。根据采样方式的不同,文献[11]将轨迹数据分为以下 3 类。

- 1) 基于时间的轨迹数据,即,采用相等时间间隔对移动对象进行采样生成的轨迹;
- 2) 基于位置的轨迹数据,即:当移动对象位置发生变化时,记录下移动对象的位置生成的位置轨迹;
- 3) 基于事件触发的轨迹数据,即移动对象的活动触发传感器事件后,由传感器触发产生的轨迹。

在现实生活中,

- 第 1 种对应的是如 GPS 这类轨迹生成设备生成的轨迹,具有明显的位置、时间和运动速度等属性,连续性强;
- 第 2 类对应着诸如居民出行调查、人口迁徙统计,用户到达某个地点,就给用户添加一条位置转移记录,主要依靠人工操作来完成。虽然这类位置具有时空属性,但是很难获取对象的在某个位置的速度和运动方向等信息;
- 第 3 类对应着诸如刷卡数据、WIFI 定位数据、签到数据等。记录的位置不一定是移动对象的真实位置,只有满足某个触发条件,移动对象的相关信息才被记录,显然无法描述移动对象详尽的移动信息。

除此之外,为了研究方便,还有根据已知轨迹数据特征人工合成的轨迹数据,如根据 Foursquare 的特征^[4],主要用于推荐算法分析和验证^[12]。

通常来说,移动轨迹数据包含移动对象的位置、时间、运动速度等属性信息。从时空角度可以看成是一种具有连续性的序列,这种序列就是符合一种序列性规则的时空痕迹。任何对象在空间中的位置,只要能够按照合适的时间和空间粒度,可以依据某种规则关联起来形成有意义的序列,就可以看作是轨迹数据。显然,轨迹数据是一种广义的定义。按照这种广义的分类方式,当前在基于用户轨迹数据的移动推荐中涉及到多种轨迹数据集,如 GoeLife、Foursquare、Gowalla 等,各种数据集的详细情况见表 1。

Table 1 An introduction to moving trajectory data sets**表 1** 移动轨迹数据集介绍

数据集	特征	数据集描述
GeoLife ^[4,13-22]	密集连续运动轨迹,缺少语义	微软亚洲研究院向外公开的数据集,主要是包含经纬度和时间戳的 GPS 轨迹数据.数据内容包括两个部分:用户户外活动轨迹和 119 辆汽车的年轻驾驶员的驾驶记录,轨迹数据大部分采集于北京市
Nokia MDC ^[17]	密集连续运动轨迹,语义相对丰富	诺基亚在瑞士洛桑举办移动数据挑战赛生成的数据,近 200 名志愿者参与了这次活动,收集了近两年(2009~2012)的连续数据.数据集包括由 GPS、蜂窝网和 WLAN 生成的位置信息、多媒体的创建和使用数据以及用户行为数据
Foursquare ^[8,9,23-43]	稀疏离散签到轨迹,语义丰富	基于用户地理位置信息(LBS)的手机服务网站,鼓励手机用户与他人分享自己当前所在地理位置等信息,Foursquare 数据集即来自该网站
Gowalla ^[23,25,27,32-34,39,43-45]	稀疏离散签到轨迹,语义丰富	与 Foursquare 类似的应用,可以在朋友间分享地点、活动、旅行线路等信息.另外,Gowalla 还提供商家、竞选和组织以独特的方式赢得忠诚度,拓展受众面,并建立难忘的体验,目前属于 Facebook
Twitter ^[8,36,41,46]	稀疏离散轨迹,语义丰富	Twitter 是一家美国社交网络及微博客服务的网站,虽然其空间项目不包含类别和标签信息,但是它支持第三方的位置分享服务,如 Foursquare 和 Gowalla.可以从 Foursquare 上挖掘信息弥补缺少的信息
Flickr 数据集 ^[5,12,47-50]	图像构成的轨迹	包含具有地理标签的照片数据集,由通过用户自己拍摄和分享的照片构成
MIT reality mining ^[51]	离散轨迹,隐含语义丰富	MIT 多媒体实验室于 2004 年~2005 年从个人手机收集的数据,包括通话记录、邻近的蓝牙设备、手机信号塔 ID、应用程序的使用以及手机状态(如充电和空闲).这些数据主要从上下文应用程序中产生
Yelp ^[33,35,37,40]	评论丰富,以天为粒度的轨迹数据	Yelp 是美国著名商户点评网站,创立于 2004 年,囊括各地餐馆、购物中心、酒店、旅游等领域的商户,用户可以在 Yelp 网站中给商户打分,提交评论,交流购物体验等
RFID 数据集 ^[52]	离散访问轨迹,范围小,语义丰富	用户访问主题公园生成的离散序列数据
出租车数据集 ^[29,53-55]	连续密集的移动轨迹,无语义	从车联网或出租车公司获取的数据,主要以 GPS 轨迹方式呈现
合成数据集 ^[12,51,55]	人工定制细节的轨迹	根据算法验证的需求,根据轨迹特点和一系列规则生成的数据,主要用于推荐算法的理论研究

1.2 用户轨迹数据集的特征

能够产生轨迹数据的移动设备有多种,如 GPS、RFID、WSN、WIFI、智能手机传感器、GSM 信号、红外线或超声波系统等,这些设备时刻不停地产生各式各样的轨迹数据.移动互联网和物联网的快速普及,给用户分享个人数据提供了极大的方便,促进轨迹数据来源向着多样化发展,如研究机构共享数据 GeoLife、MDC 等,商业公开数据 Foursquare、Gowalla 等.显然,轨迹数据完全符合大数据的 5V 特征,即:Volume(大体量)、Velocity(速度)、Variety(多样性)、Veracity(真实性)和 Value(大价值).轨迹作为一种大数据资源,有自己独有的特征,具体内容如下.

- 时空序列性

轨迹数据由包含空间坐标和时间戳的位置序列构成,在一定的空间粒度和时间粒度下,不论是密集的还是稀疏的轨迹数据,数据当中的位置点都具有很强的空间毗邻性和时间连续性,时空序列性是其最核心的特征.如图 2 所示:在轨迹映射图 B 中,轨迹 1~轨迹 3 包含的位置点不同,是 3 条不同的轨迹;相同的位置点在不同的轨迹中的顺序不同,序列性反映到轨迹含义层面会有很大的区别.

- 轨迹数据异构性

不同类的轨迹数据包含的数据类型不同,如 GeoLife 数据集包含的是经纬度坐标、海拔和日期时间等^[13], Gowalla 包含的是用户 ID、日期时间、经纬度和签到位置 ID 等^[23],而 Flickr 包含的是带地理标签的照片集

合^[5].从采集时间间隔看,MDC、GeoLife 数据采集间隔短,数据密集而连续;而像 Gowalla 签到数据的时间间隔从几分钟到几天、几个月都有可能,数据稀疏且离散^[45].由此可以看出,轨迹数据在内容和时空上都存在着异构性.如图 2 中的 A 和 B 分别代表着离散的轨迹和连续的轨迹,对这两种不同的轨迹处理方式也不相同.

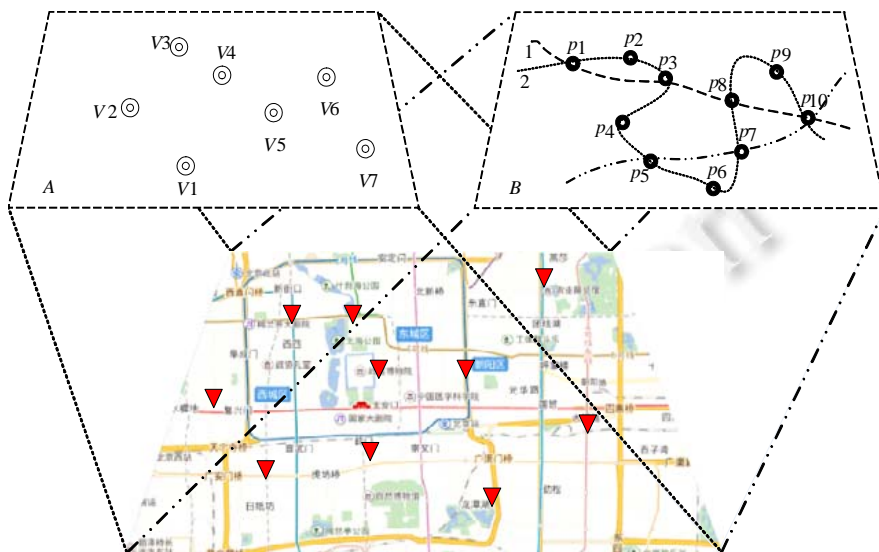


Fig.2 Characteristics of moving trajectory datasets

图 2 移动轨迹数据集的特征

- 语义信息规模大与语义信息密度低并存

轨迹数据属于大数据,语义信息的总体数量非常大,但是语义信息密度比较低.从数据本身来看,特别是社交网络类的轨迹数据,如通过对 Gowalla 的分析显示^[45],该数据集在时间和空间上都非常稀疏.例如,一年内签到超过 58 条的用户仅占 10%左右.虽然用户签到的每条数据都带有相应的语义信息,但是由于签到的时间间隔比较大,而且间隔还不统一,所以在一定的时间区间或在一个空间区域内,平均语义密度比较低.从用户角度来看,在考虑推荐时效性的前提下,也只会参考当前时间前的一小段和用户当前位置附近的区域的部分数据.本来用户自身的数据就不多,在进行时空细分的情况下,语义信息密度会更低.

由此看来,轨迹数据是一种特殊的大数据,大数据特性和轨迹数据自身的特性为基于用户轨迹数据的移动推荐带来一些挑战和机遇.这意味着传统的移动推荐思路、方法不太适合基于用户轨迹数据的推荐,但是这样的挑战也决定着基于用户轨迹数据的移动推荐有一个广阔的发展空间.

1.3 基于用户轨迹数据的移动推荐系统

人们在现实世界中运动会以轨迹的形式留下丰富的信息,由于这些信息蕴含着人们的移动特征、个性偏好以及与相关时空项目的关联信息,对用户的轨迹数据进行挖掘和分析可以获取经济效益,所以分析和利用轨迹数据引起了学术界和产业界的广泛关注.例如,早在 GPS 应用开始时,就有研究人员投身基于 GPS 数据的移动推荐^[56],特别是 Zheng 等人在文献[13]中结合 GeoLife 项目,详细地阐述了用户、用户位置和用户轨迹之间的关系,为我们展示了基于轨迹分享个人生活经历、生成旅游推荐、个性化朋友推荐和位置推荐等应用场景,这极大地鼓舞了研究者们对移动推荐研究的热情.但是,GeoLife 数据本身缺少相应的语义信息,特别是用户的概貌信息和相应的评价信息,所以它在推荐领域的应用还存在着一些局限性.随着具有丰富语义信息的社交网络和个性化分享网络的出现,为推荐领域带来了各式各样的轨迹数据,从此才算是给移动推荐带来了前所未有的春天.这些轨迹数据彼此之间异构性比较强,数据之间存在互补关系,使得传统的移动推荐跳出了仅参照位置进行推荐的限制,扩展到如异质网络应用^[57]、多数据集联合^[29,53]、时空全面考虑^[9]、潜在信息分析^[46,58]和联合建模^[59]

等,实现数据集之间进行互补,基于用户轨迹数据的移动推荐研究和应用才逐渐走上快速发展的道路。

在我们的现实生活中已经存在不少基于用户轨迹的移动推荐应用,如地图服务中的路况预测和实时路况应用,就是依赖各种数据终端,如车载或手持的 GPS 导航设备等。利用回传的轨迹数据,根据历史轨迹数据可为用户提供未来时段的路况预测,根据实时的轨迹数据可提供当前的路况推荐,方便用户提前规划行程,避开拥挤时段或拥挤区域。除此之外,还有城市道路规划方面,道路拥挤一般是由于当前的道路已经不能满足城市发展对交通的需求,可以根据车流轨迹数据获取拥挤路段分布信息,为城市道路规划部门推荐最需要解决的道路区域。

2 基于用户轨迹数据的移动推荐系统的研究

基于用户轨迹数据的移动推荐利用轨迹数据为用户提供与轨迹相关的时空项目推荐、兴趣点推荐、旅游推荐、路线推荐(智能交通推荐)、活动推荐等。一方面,用户访问时空项目需要经济支持,如旅游是一种消费活动,经济承受能力限制了用户访问时空项目的频率。另外,再加上用户对隐私保护的需求,阻碍其上传或共享自己的精确、完整的轨迹数据,所以轨迹数据稀疏性比较大^[60,61]。由于轨迹数据的时空特性决定了基于用户轨迹数据的移动推荐与时空关联度非常大,所以需要充分考虑充分利用轨迹数据的大数据特性和与时空和序列关联的特性,为用户提供个性化的推荐结果^[23]。

基于用户轨迹数据的移动推荐扩展了移动推荐的关注的因素,从当前基于用户轨迹数据的移动推荐研究情况可以看出:基于用户轨迹数据的移动推荐使用的数据集比移动推荐使用的数据集更加广泛,并且强调多信息融合,重视数据的“广度分析”。充分考虑轨迹数据的时空属性给推荐带来的影响,我们可以从多个角度对基于用户轨迹数据的移动推荐进行分析。

- 1) 从对影响用户行为因素建模的角度出发,基于用户轨迹数据的移动推荐的研究可以划分为基于轨迹时空上下文感知、基于时空项目内容以及多因素联合建模的研究;
- 2) 轨迹数据迅速增长既给推荐提供了大量的信息,同时也增加了数据处理和信息更新的负担。从提升推荐实时性和降低处理消耗的角度出发,可以划分为基于离线/在线相结合和基于增量更新的研究;
- 3) 大多数方法都是对单一数据源进行挖掘,例如 Geolife、Foursquare、Gowalla 等,但是少量文献已经对同一个用户账号关联的多源数据进行挖掘,以此弥补单一数据源上信息的缺失和噪音问题。因此,从不同轨迹数据集存在关联的角度出发,可以划分为多源数据利用的研究。

2.1 基于轨迹时空上下文感知

轨迹数据是一种具有时空属性的大数据,包含着移动用户的时空概貌信息。以往基于位置的移动推荐主要是根据用户当前位置和空间项目信息,研究用户行为的空间特征,探索移动用户的空间偏好,旨在为其实现个性化的空间项目的查询和推荐服务等^[62]。虽然存在极少数研究关注数据蕴含的时间和空间因素,但仍然忽视了序列性。在基于用户轨迹数据的移动推荐过程中,广泛地分析了轨迹数据中包含的丰富信息,除了利用轨迹数据包含的如位置信息、空间项目信息和用户评价信息等基本信息外,还需要获取相应的时空上下文^[25,26,37,44,50,53,63]和序列上下文^[8,14,16,22,23,32,47,52,64]信息。这两种上下文信息主要体现了时空属性,如时间变化、空间运动状态和时空序列等。其中,时间上下文和空间上下文被统一视作时空上下文,结合序列上下文,能够刻画用户行为在时空域上的连续性,有助于从深层次理解和掌握用户行为的规律。

• 时空上下文研究

广义的轨迹数据是一种时空属性明显的信息,包含了许多时空上下文信息。时空上下文主要涉及时间变化、运动方向和速度变化。其中,时间属性是轨迹数据的一个重要属性,所以在基于用户轨迹数据的推荐中扮演着重要的角色。例如:用户倾向于在不同的时间访问不同的地方,如中午喜欢自助餐、晚上喜欢看电影。时空项目的可用性有时间限制,如博物馆 9 点钟才开门、不同的用户访问时空项目的时间消耗不同、时空项目在可用期间也会受到外在因素的影响等等。另外,移动用户的空间运动状态,如速度和方向等,也会对推荐产生影响。

用户的习惯与时间存在密切的关系,针对用户倾向于在不同的时间访问不同的地方,一些研究充分利用这种关系提升推荐效果。文献[25]结合用户签到数据的时间因素,考虑用户与时间相关的偏好,开发了一个纳入时

间信息的协同推荐模型,在考虑地理信息的基础上,为用户在一天的特定时间内推荐一个 POIs 集合.通过对数据分析,发现用户在不同的时间具有不同的签到偏好,且在相近的时间区域上用户偏好相似.针对用户行为的时间特征,文献[26]将一天划分成 24 小时,并提出具有时间效应的位置推荐框架(LRT)用于 LBSNs 中的位置推荐,取得了较好的效果.这说明时间上下文能够有效提升推荐的性能.

除了用户个性化的习惯外,群体用户活动的聚集现象与时间也有密切关系,如游客喜欢在 9:00 左右涌向景点.这种倾向往往会对时空项目的可用性造成影响,同时还影响用户的个人体验.如文献[50]提出一种人群感知旅行推荐算法(PersCT),挖掘 POI 拥挤时间对旅行的影响,提升用户的旅行兴趣和体验.类似地,虽然城市交通状态的动态变化和乘客出现时间具有随机性,但是通过分析也存在统计学规律性.文献[53]研究出租车在城市中不同街道的载客率和收益间的动态关系,提出一个名为 HUNTS 载客路线推荐系统,该系统利用一个动态评分系统从载客率和效益角度对每段道路进行实时评分,充分考虑时间影响,并基于启发式方法和天际线技术提出一种轨迹缝合的新方法,为出租车实时地提供近似最优载客路线推荐.

考虑到时空项目的可用性有时间限制,文献[37]综合分析存在的时间限制,提出一种名为 TripRec 的推荐系统以精简推荐搜索空间,采用基于前缀的深度优先搜索策略(PDFS)生成满足时间约束的最优旅行路径.鉴于用户访问相邻的空间项目过程中停留时间的不确定性,文献[44]提出一个称作位置和时间感知的协同的检索模型(LTSCR),同时对位置、时间等信息建模,采用利用加权近似配对排序(WARP)损失函数的协同检索模型生成连续的 POI 序列(TOP-N)推荐.

用户行为在时间域和空间域上的交互可以定义为其运动状态,即运动速度和方向,用户的运动状态变化也影响着推荐结果.虽然在移动旅游推荐应用中也经常利用空间上下文,但是没有考虑用户的运动状态转换信息.一些基于轨迹数据的移动推荐系统采用了这种状态转换信息,并获得了良好的效果.例如,文献[63]利用用户运动速度和轨迹,依据用户的运动轨迹的方向设置兴趣区域的中心,沿着用户的移动方向向前移动,忽略用户当前位置后方的 POI,结合用户的偏好,快速找到最有趣的 POIs.然而,获取用户的运动状态需要用户的轨迹数据相当密集,否则,如速度或方向的实际意义会打折扣.例如,用 10 个位置点表示用户在 10 分钟内移动了 400m 与在 1 天内移动了 4 000KM 相比,前者的精确度更高,可以充分考虑运动状态对推荐带来的影响.显然,如果数据像后者那样过于稀疏,这类推荐方法的推荐性能会严重下降.

- 序列上下文研究

人类活动在时空上存在明显的序列性:按照时间排序,用户访问的项目存在先后顺序;在空间上,用户访问的项目存在远近关系.在广义轨迹数据上进行的序列上下文研究,主要关注于用户轨迹数据的序列性所反映的用户偏好的变化以及推荐结果和用户近期访问项目序列上存在的潜在关系.

在序列性对推荐影响的研究方面,为了有效地对广义轨迹中的序列上下文建模,也产生了一系列方法,例如,文献[8]设计了一种顺序个性化空间项目推荐框架 SPORE,其引入潜在概率生成算法融合了序列影响和个人兴趣,用与马尔可夫链不同的模型解决了传统移动推荐引入序列上下文时伴随的非个性化问题.文献[14]提出一种基于时间线序列建模的方法来捕获与用户行为模式相关联的顺序和与用户活动相关的上下文,以更好地捕捉用户当前和过去的活动模式之间的相似之处,结合时间线匹配方法,为用户提供精准的交通方式推荐服务.文献[16]提出一种基于频繁路径挖掘的方法,利用滑动时间窗口和频繁轨迹门限值,从群体用户的历史轨迹数据中挖掘用户频繁路线,用固定停止率(FSR)识别不同的行程模式,为有相似出行线路的群体用户提供出行共乘推荐.文献[22]利用用户运动序列属性、地理空间层次属性和访问位置的流行度判断用户的相似性,为用户提供朋友和位置推荐.文献[23]提出一种基于图嵌入的通用模型(GE),该模型以 POI 与 POI、区域、时间和内容这 4 种关系图的方式捕获顺序效应和时间循环效应,将序列因素引入空间项目的表示学习中,并利用负采样方式快速完成学习,提升了 POI 的推荐效果.文献[32]提出名为 LORE 的推荐方法,LORE 逐步从位置序列挖掘序列模式,将用户行为序列模式表示为动态位置-位置转换图(L²TG),采用加法马尔可夫链(AMC)预测用户通过 L²TG 访问相应位置的概率,最后将地理影响力和社会影响力的顺序影响融入统一的推荐框架.上述基于用户轨迹数据的移动推荐方法都取得了不错的效果,这说明序列因素的引入有助于提高空间项目推荐的性能.

在连续兴趣点推荐或路线推荐研究中,推荐项目之间存在序列性是最显著的特征.如文献[47]中提出一个基于网格的时空轨迹兴趣区域挖掘和序列模式挖掘框架,引入3D技术对热门区域和序列模式加以展示,从照片集合所反映的轨迹序列中预测用户接下来可能去的地方和出发时间.文献[52]以RFID系统为基础构建了一个针对主题公园参观路径的推荐系统,该系统通过RFID系统记录游客进入公园参观的路径、主题区域、参观顺序等信息,根据历史游客参观序列和停留时间长度的相异性,把参观序列进行分段聚类处理,构建历史游客的游乐偏好矩阵和利用当前参观者的偏好向量作为推荐标准,如预定出发时间、喜欢的主题区域和优先次序,为参观者生成一个合适的参观路径推荐.文献[64]提出利用一系列照片中携带的GPS数据,扩展数据挖掘理论中的关联规则挖掘和序列挖掘对风景行程进行挖掘,实现根据当前位置和以前访问的景点序列为用户提供个性化风景旅游行程推荐列表.除此之外,神经网络也被用于推荐中,例如:文献[65]采用RNN获取用户序列行为之间的关系,为用户提供下一位置推荐;文献[66]根据时间间隔的长度,采用RNN和GRU获取用户运动模式中的序列关联特征,然后融合社会关系进行位置推荐.

从表2可以看出,采用的技术有改进传统技术和大量利用空间模型两个发展趋势.改进传统技术相当常见,如在文献[14,25,44,63]中,可以找到协同过滤、基于知识、基于内容和混合推荐的影子;像关联规则挖掘等,在文献[64]中有所应用.利用空间模型是一个比较新颖的方向,如在文献[16,47]中用到网格技术,文献[23]中用到图嵌入模型,文献[65,66]使用循环神经网络获取用户相关的序列特征.从这些方法可以看出:利用上下文信息的推荐方法比较多,但是各种方法特征各异.如果能够形成一个通用的或统一的框架会更有利于推荐,其相关研究有待进一步拓展.

Table 2 Trajectory context perception

表2 轨迹上下文感知

名称	索引代表	特点
时间 上下文 研究	UTP ^[25] ,LRT ^[26] ,PersCT ^[50]	用户偏好与时间相关,相邻时间段的签到分布具有相似性,用户会选择合适的时间去访问空间项目
	PDFS ^[37] ,LTSCR ^[44] ,HUNTS ^[53]	时空项目在不同时间具有不同的意义
	CARS ^[63]	移动对象的速度、方向对候选推荐结果存在影响
序列 上下文 研究	SPORE ^[8]	用户移动序列效应影响推荐结果
	SeqNCRec ^[14]	过去运动的序列性与推荐结果序列性相关
	GRT ^[16] ,HGSM ^[22]	过去运动具有相同序列性的用户相似度较高
	GE ^[23]	顺序效应与时间循环相关影响推荐
	LORE ^[32]	位置的序列性对用户的签到行为存在影响
	SPM ^[47]	离散的、具有地理标识的图像利用时空
	MSTravel ^[64]	位置序列化可以形成时空轨迹
	轨迹聚类算法 ^[52]	不同的序列具有不同的含义,路径与序列相互印证
ST-RNN ^[65] ,RNN,GRU ^[66]	采用循环神经网络获取序列行为模式	

2.2 基于时空项目内容的推荐研究

轨迹数据除了具有独特的时空特性外,还具有可以称作内容的信息,如时空项目的语义、类别以及其他一些描述性的信息等,我们把这些可以直接获取的内容信息称作显性内容.然而,经常存在轨迹数据缺少显性内容的情况,如Geolife数据集,空间信息密集,但是缺少语义.除此之外,还存在依靠轨迹数据所具有的显性内容还不足以满足推荐需求,如Foursquare等.为了给用户提供更好的推荐结果,有些研究采用技术手段对轨迹数据进行处理,获取一些与显性内容相近的信息或者与显示内容相关的更深层的内容信息,为了方便,我们称其为隐式内容.针对基于时空项目的内容推荐,主要包括语义信息挖掘和潜在主题挖掘两个方面,对比分析见表3.

采用语义信息挖掘研究主要针对缺少语义信息的轨迹数据,如GPS类数据的语义信息稀缺,由于位置点比较稠密,很难依据具体的位置点进行基于内容的推荐,所以需要利用不同的粒度对位置点进行分层聚类处理,从一个较大的空间上获取相应的语义信息.如文献[15,21]基于多用户的GPS轨迹数据提出一种树形层次图的框架模型(TBHG),按照不同的空间区域粒度对属于轨迹数据中的位置点进行分层聚类,获取停留点或停留区域,用停留点或停留区域代替具体的位置点,弥补了具体位置点的内容信息不足的缺陷.文献[17]提出一种称作上

下文树(context tree)的层次性数据结构,利用地面使用数据来增强地理空间轨迹,实现挖掘隐藏在空间轨迹数据中的隐藏内容信息,可以更好地理解用户属性和发现用户与活动的关联。

Table 3 Recommendation based on spatio-temporal content

表 3 基于时空项目内容的推荐

名称	索引代表	特点
语义信息挖掘	TBHG ^[15,21]	树的不同层次表示不同的空间粒度,层次越低表示空间粒度越小.例如,最底层聚类对象为位置点,中间层或最顶层的聚类对象可以覆盖多个位置点的区域,如一个商业区或一个城市等
	Contex tree ^[17]	利用上下文树结构从多尺度分层揭示地理空间轨迹潜在知识,了解个人或团体活动,用于理解用户的属性和用户的活动模式
潜在主题挖掘	ST-LDA ^[33] ,LA-LDA ^[38] ,GIB-FR ^[58]	利用朴素贝叶斯潜在因子模型进行改进,实现用户与项目关联信息的挖掘
	Hidden Markov model ^[45]	利用混合隐含马尔可夫模型对用户活动模式进行建模,挖掘用户的活动偏好,提升活动推荐的准确性
	CLSA,ALSA,DLSA ^[46]	地理相关的主题模型,挖掘位置携带的潜在信息
	PLSA ^[48]	利用马尔可夫链和基于概率的潜在语义分析获取位置序列的主题,提供旅游路径推荐

潜在主题挖掘主要针对轨迹数据已有的内容信息不足以满足推荐需求的情况,如签到数据、携带地理标签的图像数据等.主要特点是总体内容信息丰富,直接应用这些内容信息还不能生成让用户非常满意的推荐结果,所以需要分析数据获取隐式内容.目前,隐含狄利克雷分布(LDA)是一种比较成熟的文档主题生成模型,也称为三层贝叶斯概率模型,所以常见在此模型基础上进行改进应用.如文献[33]提出一种时空潜在概率生成模型(ST-LDA),根据用户在每个区域中的签到内容来学习用户与区域依赖的个性化兴趣,在 ST-LDA 的基础上建立一种社会空间协同推理框架,通过有效地利用社会和空间相关信息,用于增强推断与区域相关的个人兴趣.文献[38]提出一种基于位置感知的潜在信息分析框架(LA-LDA),获取用户和项目的空间信息对生成推荐的潜在影响,捕获用户偏好信息,在 TOP-N 推荐和冷启动推荐中获得较好的效果.文献[58]提出一种基于贝叶斯的潜在因子方法(GIB-FR),对地理信息、隐含的用户评级和用户行为这 3 种信息联合处理,实现更好的事件推广和用户体验的朋友推荐.文献[45]提出一种混合隐含马尔可夫模型,利用空间项目类别信息对用户下一步活动类型进行预测.利用签到数据蕴含的位置信息获取用户潜在的移动信息,获取用户的活动偏好,根据给定的估计类别分布情况预测用户最可能去的位置.文献[46]提出基于聚类的局部语义分析(CLSA)、基于自动编码器的局部语义分析(ALSA)、深度局部语义分析(DLSA)这 3 个地理主题特征模型,对潜在信息进行分析,解决高维度、稀疏性和信息冗余问题,为用户提供基于位置的精准新闻推荐.文献[48]提出一种摄影师行为模型,该模型把马尔可夫链与主题模型统一到一个概率框架中,利用 PLSA 模型获取照片标签信息与旅游路线主题信息的相似性,推荐符合用户偏好的旅行计划.

基于内容的推荐方法是一种比较成熟的方法,主要是利用项目的内容信息为用户推荐与自己过去偏好相近的项目.从其目前在基于用户轨迹数据的移动推荐中的应用情况可以看出:这里所提到的基于内容的推荐主要体现在对轨迹数据内容获取的方式上,然后才是利用内容和其他的生成推荐结果.

2.3 多因素联合建模

轨迹数据是一种独特的时空数据,在一定的时空区域内,有效的数据非常稀疏.如签到类轨迹数据,假如有 100 个用户和 1 000 个空间项目,每个用户在 24 小时内签到了 20 个 POI,要想完整地表示用户与这些 POI 之间的关系,需要 $100 \times 1000 \times 24$ 的矩阵,但是有效信息仅仅是 100×20 .除此之外,在本地签到的机会远比在外地签到的机会要大,一旦用户身处不熟悉的地方,也会面临数据稀疏的问题.稀疏的数据会带来严重的冷启动问题,带来高昂的处理消耗,并且会影响推荐结果的生成.但是从另外一个角度来看,稀疏的数据也是一种非常重要的数据,例如用户没有签到的 POI 可能是非常好的候选推荐结果,所以在推荐过程中需要解决数据稀疏性问题,还需要利用稀疏性带来的机遇.

虽然上下文感知模型能够缓解数据稀疏性,但是还存在一些缺陷,例如:社会化上下文可在一定程度上克服

用户冷启动问题,却不能解决项目冷启动问题;基于项目内容的方法可以缓解项目冷启动问题,却不能有效克服数据稀疏性问题带来的负面影响.目前,有些研究发现,联合多种因素对推荐结果产生的影响超出各种因素对推荐结果产生影响的总和,所以多因素联合建模也成为当前基于用户轨迹数据的移动推荐中被广泛研究的方法.结合当前的研究情况,多因素联合建模可以分为多属性结合^[4,9,19,20,23,28,30,34,39,49,54,59]和跨区域联合^[35,40,41],具体情况见表4.

Table 4 Approach of joint modeling

表4 联合建模的方法

类型	索引代表	特点
多属性结合	STS ^[4]	相关度模型和基于用户的协同过滤,考虑空间时间、语义
	SocialTrail ^[9]	综合考虑空间、时间和流行度
	PTG-Recommend ^[19]	考虑时间、流行度和地理信息
	GTS-LP ^[20]	无参数分层聚类算法、概率模型和基于前缀树的索引结构,综合考虑时间、地理和语义
	GE ^[23]	包含 POI-POI 图、POI-区域图、POI-时间图和 POI-语义图这 4 种图,综合考虑多种因素的影响
	SSTREC ^[28]	基于主题模型,考虑时间及其他因素
	JIM ^[30]	考虑内容影响、时间影响、地理影响和流行度影响,综合各种因素的影响
	Geo-PFM ^[34]	个性偏好、地理影响、用户移动行为
	TRM ^[39]	TRM(topic-region model),同时发现用户签到活动中的语义、时间和空间模式,同时针对这 3 种模式对用户做出 POIs 选择的综合影响进行建模
	Tensor Factorization ^[49]	对位置、活动和时间等维度进行协同处理
	PRR ^[54]	结合反向轨迹索引、网格索引和地图匹配技术,考虑时间和空间的影响
跨区域联合	GeoMF ^[59]	联合 POI 携带的地理信息和用户的空间聚集现象进行建模,结合矩阵分解提供 POI 推荐
	CRCF ^[35]	区别对待用户的长期内容偏好和短期位置偏好
	GeoSoCa ^[40]	考虑 POI 的地理、社会和类别的关联
	gSCorr ^[41]	考虑地理与社会相关性,解决空间对推荐的影响

• 多属性结合

单独考虑时间、空间、状态和序列虽然可以提升基于用户轨迹数据移动推荐的效果,但是提升效果毕竟有限,所以,当前的研究存在利用时空属性与其他相关属性结合的发展趋势,如语义、流行度等.语义信息在轨迹数据中是一种稀缺信息,充分利用语义可以提高推荐效果.例如,文献[4]利用 GPS 轨迹时空属性结合语义特征属性,设计了一种时空语义感知位置推荐算法(STS),为用户提供准确的、时间感知的和有意义的位置推荐,并取得了较好的效果.流行度反映了空间项目的冷热程度,文献[9]提出一种基于空间、时间、流行感知的社会轨迹推荐系统 SocialTrail,该系统利用兴趣区域的空间一致性,采用最佳访问时间过滤策略对推荐的兴趣区域序列进行过滤,逐步生成与用户旅行持续时间相匹配的最优兴趣区域序列.类似地,文献[19]提出一种基于密度聚类 and 贝叶斯规则的推荐框架(PTG-recommend),同时考虑语义信息、流行度信息、时间信息和地理信息抽取候选兴趣点,提升了兴趣点推荐准确性.文献[20]提出一种基于地理-时间-语义的位置预测推荐方法(GTS-LP),详细分析了地理位置、时间和语义对用户访问下一位置的触发作用,对用户最可能访问的位置进行推荐.除此之外,采用成熟的模型框架融合多因素的研究也很常见,如图模型、主题模型和概率模型等.如文献[23]提出一种基于图的嵌入模型(GE),该模型以 4 种 POI 关系图表示和捕获顺序效应、地理影响、时间循环效应和语义效应对推荐结果的影响,解决稀疏性对推荐带来的影响.文献[28]提出一种基于主题模型,结合时间序列、用户社交网络的影响、用户的局部偏好和 POIs 的流行情况等特征,生成一系列让游客感兴趣的旅行序列作为旅行线路.概率模型应用得比较广泛,文献[30]利用联合概率生成模型,在获取用户兴趣和空间项目流行度后,模拟用户在决策过程中的签到行为.利用内容、时间等多种因素相结合,解决了用户-POI 矩阵的稀疏性,并生成 POI 推荐.类似地,文献[34]提出一种通用的概率因素模型框架(Geo-PFM),把地理影响和用户的移动性都考虑到推荐过程中,利用泊松概率矩阵分解(Geo-PFM)获取用户签到数据的隐性反馈,获得更好的推荐效果.不同模型的组合同样也有不俗的表现,如文献[39]提出了概率生成模型-主题区域模型(TRM)来发现用户签到活动中的语义、时间和空间模式,并同时对这 3 种模式给用户选择 POIs 造成的综合影响进行建模,提供实时 POI 推荐.文献[54]利用网格索

引管理路网信息,建立反向轨迹索引管理 GPS 轨迹;然后,通过时间过滤、空间过滤和偏好过滤来进行参考轨迹检索;最后,采用局部路径推荐和全局路径推荐相结合,生成基于个人喜好的个性化最短驾驶路径推荐.文献[59]结合用户在地理区域上的聚集现象和地理信息进行建模,结合地理建模和加权矩阵分解,解决了用户-POI 矩阵稀疏的问题,生成 POI 推荐.

- 跨区域联合

跨区域联合包括同类区域和不同类区域之间的联合,如空间区域和社会关系区域,二者内部关联属于同类区域联合,二者之间关联就属于不同类区域关联.同类区域关联研究主要体现在地理空间之间,如文献[35]将整个地理区域划分为若干个位置区域,首先使用与位置无关的内容推荐框架对用户的长期内容偏好建模,获取用户对 POI 内容的评级,然后使用与位置相关的推荐框架对短期位置偏好进行建模,预测用户对 POI 位置的评级,然后结合用户对内容和位置的评级情况生成推荐.除此之外,用户之间的社会关系与空间项目之间地理关系相结合也是一种可行的方案,如文献[40]提出一种名为 GeoSoCa 的推荐方法,该方法在用户签到的 POI 之间利用地理相关、社会相关和类别相关进行建模,降低了用户签到数据稀疏性对推荐造成的影响.类似地,文献[41]针对新签到推荐的冷启动问题提出一种地理社会相关模型(gsCorr),把地理空间和社会关系进行关联分析,取得了较好的效果.

综上所述,多因素联合的方法在利用轨迹数据空间属性的基础上,充分利用轨迹数据的语义等信息和成熟的数据模型,对轨迹数据进行综合分析,并取得了较好的效果.虽然方法各异,但也证实,除了依赖时空上下文信息之外,多属性结合和跨区域联合也可以很好地缓解数据稀疏性.

2.4 离线/在线相结合

推荐实际上是虚拟与现实相结合的过程,用户在虚拟世界留下丰富的广义轨迹数据和相关的附属信息,推荐的时效性要求推荐必须快速响应推荐要求.从另外的角度来看,推荐需要稳定老用户和开发新用户,并且要依靠老项目引进新项目,这就不可避免地涉及到与新用户、新项目相关的冷启动问题.显然,解决冷启动需要对轨迹数据进行深度挖掘和分析,这是一项极为耗时的操作.要满足以上需求,推荐系统就必须能够快速处理历史数据和当前的需求.系统的性能受到系统的硬件限制,性能尽管在不断地提升,但是相对来说是一个定值,所以推荐对系统的需求远远超过系统的负载能力.面对这个似乎不可调和的矛盾,基于用户轨迹数据的移动推荐也呈现出采用离线和在线相结合的趋势.

在现有的研究成果中,主要在离线部分处理耗时的操作,在线部分负责快速响应推荐请求.根据离线部分是否采用确定的模型匹配技术,可以把离线在线结合分为离线建模类^[5,12,30,31,42]和无模型类^[55,63].

离线建模类最大的特征是在离线部分采用历史数据建立模型,然后对模型进行训练,并依据新到的数据对模型进行更新.在线部分负责轻量级的操作,如响应用户的请求和推荐结果显示等.如文献[5]提出一个主题包离线模型(TPM),根据旅游记录和旅游景点照片离线构建旅行主题包空间,该空间包括用户的兴趣、消费能力和对旅行时间和旅行季节的偏好.在线模块专注挖掘用户旅游偏好和基于用户旅游偏好推荐个性化 POI 序列.文献[12]所述的离线部分利用轨迹数据构建知识模型用于计算推荐,该离线模型在有新的 GPS 数据可用时自动执行;在线部分使用当前用户旅游访问的路径信息和离线训练的知识模型生成旅游推荐列表.文献[30]首先对联合建模模型 JIM 进行离线训练,从轨迹数据中获取包含用户兴趣、移动模式和 POI 属性的知识模型.大量的计算过程都放在离线计算中,在线推荐的计算过程仅仅是一个简单的组合过程,极大地提高了推荐效果和效率.文献[31]也提出了包含离线建模和在线推荐部分的位置内容感知推荐系统(LCARS),离线部分通过考虑用户的兴趣和位置吸引力对用户的空间项目偏好建模.在线部分把查询用户和查询的城市当作输入,自动结合用户的兴趣和城市吸引力生成 TOP-N 推荐.文献[42]所述的离线部分利用树形层次结构对用户访问的位置类别进行加权处理,实现对个体用户的偏好建模和发现本地旅行专家.采用迭代学习模型,从用户历史位置中提取用户偏好;在线推荐部分利用用户偏好和本地旅行专家的意见,为用户挑选最优的候选结果.可以看出:采用离线训练模型和在线推荐的方式,可以极大地提高推荐的响应效率,所以越来越多的对时效性要求较高的系统喜欢采用这种模式.但是我们也发现,这个离线在线结合模式也存在一定的弊端,如随着轨迹数据的更新,模型也需要更

新,尽管可以在离线部分采用高性能的处理硬件和策略,但还是存在计算消耗过大的问题,这也给模型的更新提出了更高的要求。

与离线建模对应的就是无模型类,其最大的特点是在离线处理中没有采用模型.虽然离线部分也是针对推荐过程处理轨迹的高消耗操作,但在离线部分不需要训练模型,主要采用特定的算法对轨迹数据进行直接处理.如文献[55]提出一种基于动态规划的反向路径增长算法(BP-Growth),离线生成潜在的候选位置选序列,然后结合离线预处理和在线查询,以较低的计算复杂度为出租车司机推荐最优载客点序列,减少了司机的漫游寻客花费.类似地,文献[63]基于知识与协同过滤构建一个瀑布式混合旅游推荐系统,该推荐系统的逻辑推荐部分运行在远程固定的服务器上,负责确定用户的物理位置、速度和轨迹,并使用它们来指导推荐过程;手持的客户端应用程序负责与用户交互,处理上下文条件,并在需要时从服务器请求推荐结果。

伴随着轨迹数据集体量的越来越大,用户对推荐实时性和推荐结果个性化需求越来越苛刻,将使离线与在线相结合的推荐研究和应用越来越常见。

2.5 基于增量更新和分布式研究

在现实世界中,新轨迹数据增加的速度非常快,虽然离线/在线相结合提高了推荐响应速度,但是大多数离线模型并不支持实时更新^[56].这就意味着我们需要根据新采集的数据来重新训练模型参数,显然,这是一项高消耗的工作,严重阻碍了推荐效率的提高.面对新数据对模型带来的影响,已有学者从多个角度展开了研究.就当前的研究情况来看,主要可以分为增量更新和分布式相关的研究.增量更新的核心思想是:在原始模型的基础上,根据新产生的数据对模型进行部分更新.如文献[67]提出增量张量分解模型用于基于位置的推荐中,当新数据到达时,可以在先前获得结果的基础上自适应地自动更新模型,避免了对整个数据的重新计算.类似地,文献[68]提出增量奇异值分解用于张量分解,解决新数据增加带来的模型更新问题.通过对这两种研究情况进行分析可以看出:二者的核心是用到了矩阵分解的降维能力,通过低维修改实现高维的变动.这是一个值得深入研究的方向。

虽然增量更新可以解决新数据对模型更新带来的压力,但是处理能力总有一个限度.针对这种情况,有研究者从信息处理的角度看待这个问题.从信息处理的流程来看,原始数据需要采用一定的语法规则来组织,然后对数据进行挖掘和分析获取语义信息,最后从语义信息中获取有用的信息,通常被称作语用,语用是数据处理中需要获取的核心信息.将大规模的广义轨迹数据直接传输到某个研究中心越来越不现实.采用分布式的“移动计算力”代替“大量移动数据”也是一个潜在的发展趋势^[69].研究人员只需把自己的程序或处理脚本上传到轨迹数据分布存储地服务器,在存储地进行计算,然后把结果传到研究地进行汇总,这将极大地提高对广义轨迹数据进行信息提取和分析的能力,但是同时,也给数据存储和本地计算提出了新的要求。

2.6 多源数据利用

广义轨迹数据包括多种类型的数据,不同的数据集虽然在结构和内容上具有不同的特性,但是不同的数据集在时空上也可能存在着某些关联,如 GPS 轨迹数据集的某些位置与 Foursquare 数据集中的位置在时空上存在重合现象,那么这两个数据集就可以借助位置进行联合应用.如文献[4]利用位置关联,利用 Foursquare 数据的类别信息为缺少语义的 Geolife 打上标签,结合时空和语义信息为用户提供位置推荐.类似地,文献[5]把 Flickr 数据集和 IgoUgo.com 上的旅游游记数据相结合,根据 Flickr 数据集中的旅游景点照片和 IgoUgo.co 旅游记录获取用户旅游偏好,为用户提供个性化 POI 序列推荐.文献[29]联合公共交通数据集、Foursquare 数据集和微软亚洲研究院的 T-drive 数据集,利用数据集中地理位置的类别层次进行推理,采用蚁群算法进行泛化的行程推荐,满足用户多条件限制需求.文献[53]提出一个名为 HUNTS 载客路线推荐系统,联合历史和在线的 GPS 数据和交易数据,实时地提供近似最优载客路线推荐.除此之外,还存在两个不同的应用支持同一账号登录,这样两个数据集就可以通过账号进行关联,或者是一种应用支持间接的第三方应用提供辅助性的功能.如 Twitter 就是采用 Foursquare 和 Gowalla 提供位置共享服务,使得 Twitter 用户可以分享自己的签到信息^[8]。

从应用效果可以看出:联合利用不同的数据集可以充分发挥每个数据集的优点,通过利用数据关联互补,不

但可以给用户提供更加个性化、人性化的推荐结果,还可以满足用户更加苛刻的要求,如支持用户的模糊需求推荐,极大地提升了用户满意度。

3 推荐系统评价指标

推荐系统的评价可以通过在线或离线方式进行.在线方式根据用户使用系统的实时反馈或填写调查问卷的方式对系统进行评价;离线方式通过把数据集分为训练集和测试集,在训练集上利用推荐方法或训练模型获取推荐结果,利用测试集对推荐结果进行评价^[70].不论是在线还是离线评价,都需要采用一系列的指标进行衡量.文献[71]对推荐系统的评价指标进行了全面的综述,把评价指标分为准确度、基于排序加权指标、覆盖率、多样性和新颖性这五大类共 27 种评价指标,这里称作传统评价指标.在基于用户轨迹数据的移动推荐领域中,存在部分传统评价指标和一些特殊的非传统指标.

3.1 传统评价指标

在基于用户轨迹数据的移动推荐中,大多数研究采用把轨迹数据分成训练集和测试集的方式进行,评价方式也是采用离线的方式.评价指标一般可以分为推荐效率和推荐效果两大类:推荐效率主要衡量推荐算法的时间复杂度,而推荐效果主要是推荐的准确度等.这里,我们只对使用最多的推荐效果指标进行简单分析.在基于轨迹数据的移动推荐中,常用 Precision^[4,15,19,20,25-27,32,34,40,42,44,46,48,50,51,57-59], Recall^[4,19,25-27,32,34,35,40,42,44,46,50,51,57,59]和 F-measure^[4,20,50]对推荐结果的准确性进行度量.在评价过程中,我们希望 Precision 和 Recall 的数值同时越高越好,但是实际上二者经常出现一高一低的情况,这时就要采用 F-measure 综合考虑它们,对 Precision 和 Recall 进行加权调和平均.

另外,为了克服前 3 个指标单点值的局限性,更好地反映系统的全局性能,还用到 MAP(mean average precision)^[5,15,21]指标.另外,有时还需要通过计算预测评分和实际评分之间的差距来判断系统的预测性能,这类指标对评分预测的精准度进行度量,常用的指标有 MAE 和 RMSE.由于涉及到评分,所以这类指标常用于数据集中有用户评分的情况或可以把隐式反馈转换成评分的情况.除此之外,当推荐的结果以一种序列的方式出现时,时常希望最相关的结果排到最前面,通过采用 nDCG(normalized discounted cumulative gain)^[15,28,49]对排序的效果进行度量.以上常见的评价指标的详细情况见表 5.

Table 5 Traditional evaluation metrics

表 5 传统评价指标

指标类别	评价指标	描述
推荐准确性指标	$Precision = \frac{TP}{TP + FP}$	TP为推荐命中数目,FP为推荐未命中数目.对于TOP-N推荐而言,TP+FP就等于N,而TP为前N个项目在测试中被命中的数目
	$Recall = \frac{TP}{TP + FN}$	FN为在测试数据中,用户实际访问过但是不在TOP-N列表中的项目数,TP+FN等于测试项目总数
	$F - Measure_{\beta} = \frac{(\beta^2 + 1) \times (P \times R)}{\beta^2 P + R}$	β 是参数,P是准确率,R是召回率,当 $\beta=1$ 时,就是常见的F1-Measure
	$MAP@n = \frac{\sum_{i=1}^m AP_i}{m}, AP_k@n = \frac{\sum_{j=1}^n rel_{j,k}}{n}$	$AP_k@n$ 为推荐给用户k共n个POI的平均Precision, $rel_{j,k}$ 表示相关值,如果用户 u_k 访问的 POI_j 出现在测试集中, $rel_{j,k}$ 等于1;否则,等于0.MAP@n表示m个用户的MAP.MAP的效果虽然不错,但是不够直观,需要结合Precision-Recall曲线图(PR),实际上,AP也就是PR曲线与X轴围成的面积
预测评分精准度	$RMSE = \sqrt{\frac{1}{N} \sum_{u,i} (r_{ui} - r_{ui}^*)^2}$	r_{ui} 为预测评分, r_{ui}^* 为实际评分,N为评分项数目
	$MAE = \frac{1}{N} \sum_{u,i} r_{ui} - r_{ui}^* $	r_{ui} 为预测评分, r_{ui}^* 为实际评分,N为评分项数目
排序精准度	$NDCG = \frac{DCG}{DCG^*}, DCG = \frac{1}{N} \sum_{u=1}^N \sum_{j=1}^j \frac{g_{uj}}{\max(1, \log_b j)}$	NDCG是一种来自信息检索的方法,其中,位置以对数形式衰减.假设推荐物品i给用户u,会得到一个“收益” g_{ui} ,物品集合J的平均折扣累积收益(DCG)如公式所示.在公式中, DCG^* 为理想的DCG

3.2 非传统评价指标

在基于用户轨迹数据的移动推荐中,除继承了部分传统指标外,还包括一些衍生于传统评价指标和一些独特的专用指标,这里统称为非传统指标,包括基于传统指标的衍生指标和专用指标。

具体情况见表 6,衍生指标包括 Accuracy@k^[8,33,38,39,44]、Prediction Rate^[12]、Modified Accuracy^[12]、Average Error^[12]、S@N^[28]、Crowdedness (Crd)^[50]、Popularity (Pop)^[50]、User Interest (Int)^[50]和 User Happiness(F)^[37]等。

Table 6 Nontraditional evaluation metrics

表 6 非传统评价指标

类别	评价指标	描述
衍生评价指标	准确率 $Accuracy@k = \frac{\#hit@k}{ D_{test} }$	#hit@k代表在测试集中所有的命中数, D _{test} 代表所有的测试用例数,相当于Precision
	Prediction rate	系统能够做出预测的轨迹的百分比,相当于Precision
专用指标	修改准确率 Modified accuracy	一个区域的游客可能会转移到同一区域或另一个区域的另一个POI的可能性程度
	平均错误 $Average Error = \frac{a}{n-1}$	计算每个轨迹的平均误差百分比,n个区域的轨迹分割n-1次,然后作n-1次比较,如果轨迹推荐列表包含下一个区域,则比较结果为真,反之则为假,设为假的次数为a
	成功概率 S@N	在TOP-N的POI中,用户至少访问了其中一个POI的概率
	拥挤度 $Crd = \frac{U(i,t)}{\max(U(i,t \in T_p))}$	i为POI的编号,t为时间,T _p 为一个时间段,U为在时间段T _p 内编号为i的POI附近的拥挤程度
	流行度 $Pop(i) = \frac{Ocr(i)}{\max(Ocr(j, j \in W))}$	Ocr(i)为编号为i的POI出现的次数
	用户兴趣度 Int(u,i)	在Int(u,i)中u为用户,i为POI编号,Int(u,i)计算用户POIi与用户过去访问的POI的相似度,结果是一个向量,最终把u用户对行程中所有POI的用户兴趣Int相加就是用户的兴趣值
	用户快乐程度 $F(P,u) = \sum_{i \in P} r_{ui}^*$	在路线P上,把用户对路线上所有的POI评分r _{ui} [*] 进行求和,最终的结果作为用户对这条线路的满意程度

从表 6 可以看出:非传统评价指标并不多,Accuracy@k 和 Prediction Rate 相当于传统准确度评价指标中的 Precision 与 Precision 的计算方法类似.Modified Accuracy 衡量把用户当前位置添加到推荐结果中对推荐效果的提升;Average Error 衡量推荐列表不包含下一区域的比例;S@N 指的是 Success at rank N,在 TOP-N 排名集合中找到至少一个真正访问的 POI 的概率.Crd、Pop 和 Int 反映的是 POI 的拥挤度、流行度和用户对 POI 的兴趣.User Happiness 指标是一个非常罕见的指标,主要衡量用户对路径的满意程度。

综上评价指标可以看出,基于用户轨迹数据的移动推荐系统的评价指标主要还是传统的评价指标.这主要是因为当前基于用户轨迹数据的移动推荐的研究,其核心思想与传统的推荐系统推荐流程基本一致,所以评价指标有极大的相似性.非传统评价指标虽然不太多,但从中也可以看出,学术界也在尝试新的评价指标.显然远远不够。

4 移动推荐系统与基于用户轨迹数据的移动推荐系统的联系与区别

移动推荐系统是传统推荐系统在移动互联网领域的延伸,基于用户轨迹数据的移动推荐系统衍生于移动推荐系统,所以二者之间存在着一些联系.然而近年来,工业界和学术界对轨迹数据的深入挖掘和应用,促使基于轨迹数据的移动推荐成为一个特色鲜明的热门研究领域.继续把基于用户轨迹数据的移动推荐归到宽泛的移动推荐研究中显然不太合适.所以,了解二者之间的联系和区别,有利于从根源上掌握基于轨迹数据的移动推荐的精髓。

4.1 移动推荐系统与基于用户轨迹数据的移动推荐系统的联系

它们都属于移动推荐,由于二者都处在一个移动的环境中,所以都具有移动属性,如用户的移动性、设备的移动性和接入的移动性^[72].用户的移动性指的是用户在使用移动设备时时空位置不固定,具有显著的“3A”特征

(anytime,anywhere 和 anyhow),如果考虑使用移动设备的人可能也不固定(anyone),就构成了“4A”特征(任何人、任何时间、任何地点和任何方式).设备的移动性是指设备便于携带,可以随用户移动.显而易见:由于设备是移动的,造成设备的接入也是移动的.由此可见,用户的移动性、设备的移动性和接入的移动性是二者的核心属性.二者共有的核心属性决定了推荐可用的上下文信息变化很快,上下文信息对用户偏好需求影响更大,需要利用数据蕴含的各类信息才能生成满足移动用户需求的推荐结果^[73].

4.2 移动推荐系统与基于用户轨迹数据的移动推荐系统的区别

移动推荐与基于用户轨迹数据的移动推荐系统的区别主要体现在是否与位置相关以及如何对待位置的时空属性、序列性和状态转换等属性上,二者最大的区别如图 3 所示.

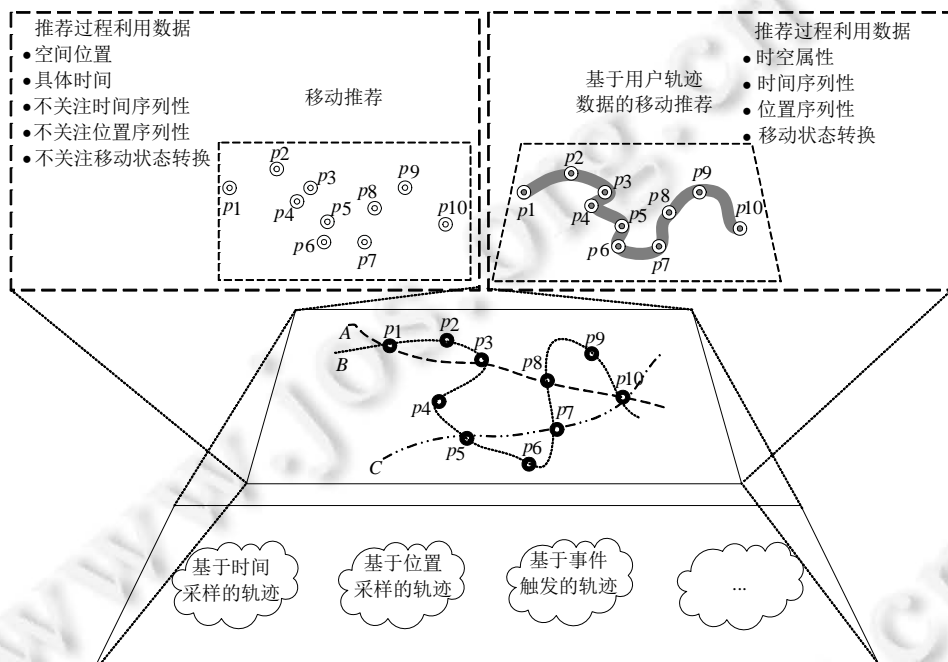


Fig.3 Differences between mobile recommender systems and mobile recommender systems based on trajectory data

图 3 移动推荐系统与基于轨迹数据的移动推荐系统之间的区别

• 与移动位置的相关性

移动推荐系统可以分为传统的移动推荐系统和基于位置的移动推荐系统,前者指的是针对非移动推荐的桌面应用开发出移动应用版本,推荐过程根本没有利用用户的位置信息,或者应用的数据就不是具有时空属性的轨迹数据,如网络购物推荐;后者才考虑到用户的位置对推荐结果的影响^[72].显然,基于用户轨迹数据的移动推荐系统与用户的位置密切相关.从这里我们可以看出:掌握基于轨迹数据的移动推荐最重要的是要认识到它与基于位置的移动推荐的联系和区别,否则,很容易把基于轨迹数据的移动推荐等同于基于位置的移动推荐或移动推荐.

• 与空间序列的相关性

移动推荐系统在利用位置信息时,更多地是考虑位置的空间属性,如地理位置,与时间关联不大.虽然在出现了 GPS 数据后也有考虑到位置和时间,但忽略了空间项目之间的序列性对推荐的影响,如位置序列性和时间序列性.移动推荐中主要的应用为 POI 或 POI 序列推荐,推荐的结果更多地以离散点或点集合的方式呈现,序列性不是太明显^[72].如仅用一个 POIs 集合表示用户下一个可能感兴趣的 POI,集合中的 POIs 之间没有顺序.

基于用户轨迹数据的移动推荐系统在推荐过程中通常会考虑位置的时间属性和空间属性^[23,26],考虑位置序列性和时间序列性对推荐结果的影响^[22,23].由于轨迹数据本身蕴含着移动对象的移动速度和移动方向,这也是基于用户轨迹数据的移动推荐系统在推荐过程中考虑的因素^[63].显然,移动推荐没有考虑用户的速度和方向因素.

我们认为,移动推荐与基于用户轨迹数据的移动推荐系统最大的区别在于如何看待移动位置的时空属性和序列性上.文献[8]的研究结果显示,地理影响、内容影响、时间影响和序列性影响对推荐结果的影响依次递增,所以把时空序列性作为区分二者的依据是可行的.也就是说,判断推荐系统是属于移动推荐还是属于基于用户轨迹数据的移动推荐不在于推荐过程是否使用了轨迹数据,而在于是否考虑到轨迹数据的时空序列属性并应用在推荐中,或者体现在推荐结果中.

综上所述可以看出:轨迹数据的特性促进了基于用户轨迹数据的移动推荐的发展,同时催生了与轨迹数据相关的研究问题,如序列性分析、多属性联合、离线/在线结合、增量更新、分布式研究和多数据源利用等,同时还对推荐结果评价指标提出了新的需求.

5 基于用户轨迹数据的移动推荐系统研究的难点和热点

作为从移动推荐衍生出来的一个新兴的研究领域,基于用户轨迹数据的移动推荐研究已经取得了一定的进展.在发展的过程中面临着众多的挑战和机遇,还存在着可以深入研究并可能取得突破性的研究方向,主要包括以数据驱动的新应用、轨迹数据稀疏性问题、用户隐私保护和安全、推荐系统评价指标和基于用户轨迹数据的移动推荐性能提升.

5.1 以数据驱动的新兴应用

移动应用领域丰富、多样的数据催生了基于用户轨迹数据的移动推荐应用.因此,根据数据源对此应用进行划分,将其分为两类研究,即单数据源上的应用和多源数据上的应用^[23].其中,目前大多数研究基于单数据源,又可以进一步划分为3类.

- 1) GPS 定位设备产生的数据,具有高采样率和明显的时空关联性,但缺乏相应的用户行为及项目内容信息,如 GeoLife、MDC 等;
- 2) 各种位置社交网络产生的数据,采样率低,用户行为间隔较长,但有丰富的用户行为及项目内容信息,如 Foursquare、Gowalla、Flickr 等;
- 3) 模拟/半模拟数据集.为了进行研究,根据一定的规则,对现有的数据集进行扩展而形成的数据.

考虑到用户在不同应用的社交平台展现不同类型的偏好,如果能够根据用户账号关联及行为地理空间的约束,将同一个用户在多个数据源的信息进行整合,则能够进一步获得更多有效的信息.例如:根据账户关联,将 Foursquare 和 Flickr 上的纽约城市旅游信息映射起来;类似地,将 Foursquare 和 Twitter 的相应数据有效整合^[8].在大数据应用时代,多种数据源的关联能够有效地揭示用户潜藏的价值,进一步提高推荐的性能^[74].然而,由于用户隐私安全保护及公司价值壁垒,多数研究仅仅能够单独分析多个单源数据,当数据中有效信息均被研究后,将形成瓶颈,阻碍研究进一步提升推荐性能^[60,61].

每种活跃的移动应用均会实时产生大规模的轨迹数据,新旧轨迹间蕴含着丰富的用户偏好变化信息.例如,新产生的轨迹中能够计算出当前用户的偏好;同时,用户历史轨迹中获取用户的长期偏好.那么两者间如何缓解用户偏好冲突、整合用户长期偏好和短期偏好、更精准地计算用户偏好是亟待解决的问题^[35].换句话说,我们可将新旧轨迹的应用问题归纳为两个层次.

- 1) 在早期用户数据的基础上,如何快速处理轨迹大数据,实现对移动用户偏好的修正和准确获取;
- 2) 对于历史位置变迁引入的噪音问题,如何构建模型提取有效的用户行为特征.

综上所述,从数据处理角度来看,移动用户轨迹的应用可从3个方面开展.

- 多源数据关联应用;
- 轨迹大数据的快速处理;

- 非个性化轨迹数据的应用.

第3种应用主要包括各种非个性化的轨迹数据,例如出租车GPS轨迹.这类数据以出租车为载体,结合其计费 and 油耗等信息,为司机提供最优的载客路线规划.总体而言,轨迹数据的研究应该充分以数据为驱动,探索多元因素对用户行为的影响,揭示广义轨迹数据中多种信息的内在联系,从而逐步提高推荐性能,缓解数据稀疏问题带来的性能瓶颈.

5.2 轨迹数据稀疏性问题

数据稀疏性是持续影响推荐性能的关键问题.目前,数据稀疏性问题的主要来源有如下几类.

- 1) 用户稀疏的签到行为和大规模的时空项目间的矛盾.鉴于时空项目的访问需要更多的花费,又因为用户隐私安全保护的需求,目前的研究仅仅能够获取非常稀疏的用户签到行为,造成了严重的数据稀疏性问题;
- 2) 冷启动用户和项目引入的稀疏性问题.活跃的社交网络应用时刻会吸引大量的新用户加入,同时会使大量新的时空项目加盟,这些新加入的用户几乎没有签到信息,却是平台非常希望吸附和激活的用户.这些用户的引入,进一步加重了数据稀疏性问题;
- 3) 新城市问题带来的数据稀疏性问题^[36,43].在新的区域,用户几乎没有签到信息,而在用户活动区域的签到信息难以帮助用户在新区域做出准确决策,这种用户行为在区域上的不平衡会引起数据稀疏性问题,从一定程度上来看,其与前两种类型有一定的重叠;
- 4) 引入多维上下文造成的数据稀疏性.为了缓解数据稀疏性,一些方法引入了多种上下文约束,将“用户-项目”矩阵扩展为“用户-项目-多维上下文”张量,利用数据间的相关性,缓解数据稀疏性.但在一定程度上,扩展后的张量可能会比原矩阵更加稀疏;
- 5) 路径推荐引入的数据稀疏性.作为基于轨迹数据的移动推荐应用,路径推荐需要计算多个时空项目的有序组合的评分,假设有 n 个项目,常见的兴趣点推荐只需要推荐下一个时刻用户需求的项目,假设进行top- k 的推荐,则时间复杂度是 C_n^k ;而假设推荐下来 k 个景点的路线,则时间复杂度是 n^k ,庞大的组合空间会导致严重的稀疏性问题^[8].

目前,对于数据稀疏性问题的解决方法,主要集中于如下几类.

- 1) 多元因素联合建模.

对于上述原因1)~原因3)引起的数据稀疏性问题,通过将多种上下文、项目语义信息整合到推荐模型,揭示更多的因素对用户行为的联合约束,从而缓解数据稀疏性问题.

- 2) 数据驱动的方法革新.

对于上述原因4)引起的数据稀疏性,则根据数据特性,选择合适的方法.例如,文献[23]发现,用户行为在时间上有聚类性,即时间上下文中用户签到数据密集,则首先选择时间感知的张量分解方法;同时,用项目的分类信息取代时空项目的数量,进一步缩小了张量的规模,从而缓解了传统张量分解数据稀疏问题.而文献[24]发现,用户行为的序列信息较为密集,则以序列信息为主,将诸多元素压缩到统一低维空间,利用嵌入图学习的方法缓解数据稀疏性问题.文献[8]发现,用户行为在区域上聚类,因此使用潜在主题模型,将区域视作影响用户行为的潜在因素,从而缓解了数据稀疏性.

- 3) 引入约束条件缓解轨迹推荐的数据稀疏性.

针对轨迹推荐中独有的组合空间规模过大的问题,现有的方法引入大量约束条件,如花费、时间、项目有效开放时间等约束条件,结合个性化的用户偏好获取方式,缓解路线推荐中的数据稀疏性问题.例如,用户希望获取一条从广州到北京的旅游路线,同时明确指出,不想再次途径武汉(由于时间问题),希望中途路过河南少林寺,并且希望最终品尝各种北京小吃,总共时间花费2天,旅游结束时希望购买全聚德的北京烤鸭,有一定的旅游预算.这种情况下,需要将多种限制条件和用户偏好相结合.

5.3 用户的隐私保护和安全性问题

在大数据时代,无论是通过告知用户取得许可,还是把关键信息模糊化或匿名化,就算消灭了个人敏感信息,也不能消灭数据中存在的社会关联关系,这三大隐私保护策略都不能真正地保护用户的隐私^[74].尽管用户对个人隐私泄露的担心阻碍了其分享详细轨迹的热情,但是随着数据量和数据种类的增多,综合分析数据会泄露个人隐私.对各类数据进行综合分析是推荐过程必经流程,例如对轨迹数据进行聚类、时空分析等操作就有可能获取用户的家庭位置、工作地点、兴趣点、热门路径和轨迹出现规律等等隐含信息.另外,用户分享的轨迹数据还蕴含着用户的当前位置和用户的社会属性等等.显然,用户位置和社会属性是用户的核心信息,属于用户隐私^[60,61].虽然在研究中时刻强调用户隐私保护和安全问题,但在实际应用中却又迫切需要用户的这些信息^[7].除此之外,当前智能移动设备非常普及,用户的移动设备在无线通信时会泄露位置信息,例如接入基站的位置、WIFI 接入点位置、用户报告(注册移动应用和服务填写的信息)等,综合分析就可以得出用户精确的位置和姓名、职业、电话等社会属性.显然,用户的隐私保护也是当前基于轨迹数据的精确移动推荐研究中的难点和热点,如何在保护用户隐私的前提下提供个性化的精准推荐,是一个重要的研究方向.

5.4 基于用户轨迹数据集的移动推荐评价问题

在推荐系统领域,评价指标有推荐效果和推荐效率两种指标,常见的推荐效果评价指标是 Precision、Recall 等,相应的推荐效率指标主要是指推荐算法的时间复杂度.这些指标继承于基于互联网信息检索领域^[71],并且目前基于用户轨迹数据的移动推荐评价基本上也是这些指标.

然而,基于轨迹数据的移动推荐系统使用的是广义轨迹数据,轨迹数据特有的时空性、序列性等特性对推荐过程和推荐效果产生了很大的影响,但是并没有合适的指标对这些影响进行量化和对比.例如,POIs 序列推荐,序列反映着结果的重要性或紧迫性,但是当前的准确率评价指标却没有考虑序列因素对推荐结果的影响.综上所述可以看出:除了继承移动推荐的部分评价指标外,还需要根据基于轨迹数据的移动推荐制定更加合适的专用评价指标.

5.5 基于用户轨迹数据推荐的推荐性能提升

近年来,基于用户轨迹数据的移动推荐研究取得了较大的进展,随着研究的向前推进,人们越来越注重推荐效果和推荐效率等推荐性能的提升.显然,持续产生的新数据和用户对推荐响应时间和准确性的需求是最主要的推动力.具体来说,可以分为两个问题:(1) 实时产生的新数据,如何通过合适的途径快速引入推荐模型,并且提取用户偏好的变化,高效计算出改变后的用户偏好;(2) 如何根据用户的历史数据,快速生成推荐结果.在这种情况下,算法的效率、精准性成为研究的重点.为了实现算法的高效性,目前研究显示主要有 3 类方式.

- (1) 建立离线/在线框架.将复杂的计算复杂度剥离到离线模块;而推荐生成过程使用轻量级的算法,放置在线模块.这就要求设计模型时,高计算量的部分和轻微计算量的部分有明显的划分,并且在划分之间的耦合度要低^[30];
- (2) 增量更新的方式.每次只需要计算新增数据对用户偏好的影响,这又需要思考如下两个子问题:(a) 如何快速计算新数据中用户偏好的变化量;(b) 如何解决用户原有偏好和更新后偏好的冲突^[68];
- (3) 并行化方式.这种方法主要包括两类.
 - 一类是普通方法的并行化^[69],主要是 Hadoop、stream 平台.但是学术界对其研究成果并不多,主要集中在工业应用领域;
 - 另一类是 GPU 并行化.随着深度学习的火热,由于 GPU 的高矩阵计算性能,使得 GPU 上的并行化变得流行.

由此可以看出:应用离线/在线框架的推荐系统将越来越普遍,最核心的还是模型实时更新的问题计算速度提升的问题.虽然通过硬件可以提升一定的性能,但是作为学术研究,主要还是需要研究更加实用的更新模型和设计更好的算法.

除此之外,用户的宽泛性意向还体现在对智能化推荐的需求上,如在城市智能导航应用中,需要用户自己了

解一些必要的信息(旅行攻略等),指定多种推荐约束后才能获得推荐结果.更多的情况是用户自己没有具体的计划,当前也没有自动化系统可以为旅行者提供个性化的行程推荐.对计算机而言,实现满足这样模糊又宽泛要求的综合推荐非常困难,所以在基于轨迹的移动推荐领域,满足宽泛的多维推荐需求也是一个热点问题.

6 结束语

随着移动设备和通信网络的进步和普及,基于位置的移动应用方兴未艾,促使移动推荐服务持续高速发展.鉴于用户行为数据的稀疏性,传统的移动推荐方法陷入性能瓶颈.以数据为驱动的思想催生了广义轨迹数据的定义和使用,即将用户的历史记录视作序列、语义和上下文的组合,这样的定义使得广义的轨迹有独特的序列信息,蕴含着强烈的时空关联性.相比于传统的 GPS 数据,广义的轨迹数据还拥有丰富的项目内容和用户行为信息.它不仅能够在物理世界中构建用户间的联系,还能够清晰地刻画用户在时空领域上的偏好,有助于研究者揭示用户行为演化的规律.因此,基于用户轨迹数据的移动推荐系统成为移动推荐领域研究的热点,并涌现出许多与移动轨迹相关的应用服务,如精准的个性化兴趣点推荐、旅游路线规划等.虽然目前已经取得了一定的成果,但仍然存在大量需要深入和扩展研究的内容,例如,如何有效缓解轨迹数据稀疏性问题、如何从广度和深度上有效利用轨迹数据、如何建立更有效的评价指标,等等.因此,基于用户轨迹数据的移动推荐系统具有重要的研究意义和广阔的应用场景.

References:

- [1] Meng XW, Hu X, Wang LC, Zhang YJ. Mobile recommender systems and their applications. *Ruan Jian Xue Bao/Journal of Software*, 2013,24(1):91–108 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4292.htm> [doi: 10.3724/SP.J.1001.2013.04292]
- [2] http://www.itu.int/net/pressoffice/press_releases/2016/pdf/30-zh.pdf
- [3] Kefalas P, Symeonidis P, Manolopoulos Y. A graph-based taxonomy of recommendation algorithms and systems in LBSNs. *IEEE Trans. on Knowledge and Data Engineering*, 2016,28(3):604–622.
- [4] Abdel-Fatao H, Li J, Liu J. Unifying spatial, temporal and semantic features for an effective GPS trajectory-based location recommendation. In: Mohamed A, ed. *Proc. of the Databases Theory and Applications, Australasian Database Conf.* Cham: Springer-Verlag, 2015. 41–53. [doi: 10.1007/978-3-319-19548-3_4]
- [5] Jiang S, Qian X, Mei T, *et al.* Personalized travel sequence recommendation on multi-source big social media. *IEEE Trans. on Big Data*, 2016,2(1):43–56. [doi: 10.1109/TBDATA.2016.2541160]
- [6] Kim J, Mahmassani HS. Spatial and temporal characterization of travel patterns in a traffic network using vehicle trajectories. *Transportation Research Part C: Emerging Technologies*, 2015,59:375–390. [doi: 10.1016/j.trc.2015.07.010]
- [7] Saxena AS, Goyal V, Bera D. Mintra: Mining anonymized trajectories with annotations. In: *Proc. of the 20th Int'l Database Engineering & Applications Symp.* New York: ACM Press, 2016. 105–114. [doi: 10.1145/2938503.2938551]
- [8] Wang W, Yin H, Sadiq S, *et al.* Spore: A sequential personalized spatial item recommender system. In: *Proc. of the 32nd IEEE Int'l Conf. on Data Engineering (ICDE)*. Helsinki: IEEE, 2016. 954–965. [doi: 10.1109/ICDE.2016.7498304]
- [9] Zhang Q, Yu L, Long G. SocialTrail: Recommending social trajectories from location-based social networks. In: Mohamed A, ed. *Proc. of the Databases Theory and Applications, Australasian Database Conf.* Cham: Springer-Verlag, 2015. 314–317. [doi: 10.1007/978-3-319-19548-3_26]
- [10] Spaccapietra S, Parent C, Damiani ML, *et al.* A conceptual view on trajectories. *Data & Knowledge Engineering*, 2008,65(1): 126–146. [doi: 10.1016/j.datak.2007.10.008]
- [11] Ting LI, Pei T, Yuan Y, *et al.* A review on the classification, patterns and applied research of human mobility trajectory. *Progress in Geography*, 2014,33(7):938–948 (in Chinese with English abstract). [doi: 10.11820/dlkxjz.2014.07.009]
- [12] Baraglia R, Frattari C, Muntean CI, *et al.* A trajectory-based recommender system for tourism. In: Huang RH, ed. *Proc. of the 8th Int'l Conf. on Active Media Technology*. Berlin, Heidelberg: Springer-Verlag, 2012. 196–205. [doi: 10.1007/978-3-642-35236-2_20]

- [13] Zheng Y, Xie X, Ma WY. GeoLife: A collaborative social networking service among user, location and trajectory. *Bulletin of the Technical Committee on Data Engineering*, 2011,33(2):32–39.
- [14] Kumar G, Jerbi H, O'Mahony MP. Personalised recommendations for modes of transport: A sequence-based approach. In: *Proc. of the Int'l Workshop on Urban Computing*. San Francisco: ACM Press, 2016.
- [15] Zheng Y, Zhang L, Xie X, *et al.* Mining interesting locations and travel sequences from GPS trajectories. In: *Proc. of the 18th Int'l Conf. on World Wide Web*. New York: ACM Press, 2009. 791–800. [doi: 10.1145/1526709.1526816]
- [16] He W, Li D, Zhang T, *et al.* Mining regular routes from GPS data for ridesharing recommendations. In: *Proc. of the ACM SIGKDD Int'l Workshop on Urban Computing*. New York: ACM Press, 2012. 79–86. [doi: 10.1145/2346496.2346510]
- [17] Thomason A, Griffiths N, Sanchez V. Context trees: Augmenting geospatial trajectories with context. *ACM Trans. on Information Systems (TOIS)*, 2016,35(2):14. [doi: 10.1145/2978578]
- [18] Ye A, Zheng Y. Large scale trajectory data management. In: *Proc. of the 2015 Int'l Conf. on Logistics, Informatics and Service Sciences (LISS)*. IEEE, 2015. 1–5. [doi: 10.1109/LISS.2015.7369700]
- [19] Liu Y, Seah HS. Points of interest recommendation from GPS trajectories. *Int'l Journal of Geographical Information Science*, 2015, 29(6):953–979. [doi: 10.1080/13658816.2015.1005094]
- [20] Ying JC, Lee WC, Tseng VS. Mining geographic-temporal-semantic patterns in trajectories for location prediction. *ACM Trans. on Intelligent Systems & Technology*, 2013,5(1):328. [doi: 10.1145/2542182.2542184]
- [21] Zheng Y, Xie X. Learning travel recommendations from user-generated GPS traces. *ACM Trans. on Intelligent Systems and Technology (TIST)*, 2011,2(1):2. [doi: 10.1145/1889681.1889683]
- [22] Zheng Y, Zhang L, Ma Z, *et al.* Recommending friends and locations based on individual location history. *ACM Trans. on the Web (TWEB)*, 2011,5(1):5. [doi: 10.1145/1921591.1921596]
- [23] Li X, Jiang M, Hong H, *et al.* A time-aware personalized point-of-interest recommendation via high-order tensor factorization. *ACM Trans. on Information Systems (TOIS)*, 2017,35(4):31. [doi: 10.1145/3057283]
- [24] Xie M, Yin H, Wang H, *et al.* Learning graph-based POI embedding for location-based recommendation. In: *Proc. of the 25th ACM Int'l on Conf. on Information and Knowledge Management*. New York: ACM Press, 2016. 15–24. [doi: 10.1145/2983323.2983711]
- [25] Yuan Q, Cong G, Ma Z, *et al.* Time-Aware point-of-interest recommendation. In: *Proc. of the 36th Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval*. New York: ACM Press, 2013. 363–372. [doi: 10.1145/2484028.2484030]
- [26] Gao H, Tang J, Hu X, *et al.* Exploring temporal effects for location recommendation on location-based social networks. In: *Proc. of the 7th ACM Conf. on Recommender Systems*. New York: ACM Press, 2013. 93–100. [doi: 10.1145/2507157.2507182]
- [27] Cheng C, Yang H, Lyu MR, *et al.* Where you like to go next: Successive point-of-interest recommendation. In: *Proc. of the 23rd Int'l Joint Conf. on Artificial Intelligence*, Vol.13. 2013. 2605–2611.
- [28] Rakesh V, Jadhav N, Kotov A, *et al.* Probabilistic social sequential model for tour recommendation. In: *Proc. of the 10th ACM Int'l Conf. on Web Search and Data Mining*. New York: ACM Press, 2017. 631–640. [doi: 10.1145/3018661.3018711]
- [29] Meng XX, Wang XD, Zhou XM. Intention oriented itinerary recommendation by bridging physical trajectories and online social networks. *Ruan Jian Xue Bao/Journal of Software*, 2012,23(Suppl.(1)):159–168 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/12018.htm>
- [30] Yin H, Zhou X, Shao Y, *et al.* Joint modeling of user check-in behaviors for point-of-interest recommendation. In: *Proc. of the 24th ACM Int'l Conf. on Information and Knowledge Management*. New York: ACM Press, 2015. 1631–1640. [doi: 10.1145/2806416.2806500]
- [31] Yin H, Cui B, Sun Y, *et al.* Lcars: A spatial item recommender system. *ACM Trans. on Information Systems (TOIS)*, 2014,32(3):11. [doi: 10.1145/2629461]
- [32] Zhang JD, Chow CY, Li Y. Lore: Exploiting sequential influence for location recommendations. In: *Proc. of the 22nd ACM SIGSPATIAL Int'l Conf. on Advances in Geographic Information Systems*. New York: ACM Press, 2014. 103–112. [doi: 10.1145/2666310.2666400]
- [33] Yin H, Zhou X, Cui B, *et al.* Adapting to user interest drift for POI recommendation. *IEEE Trans. on Knowledge and Data Engineering*, 2016,28(10):2566–2581. [doi: 10.1109/TKDE.2016.2580511]

- [34] Liu B, Xiong H, Papadimitriou S, *et al.* A general geographical probabilistic factor model for point of interest recommendation. *IEEE Trans. on Knowledge and Data Engineering*, 2015,27(5):1167–1179. [doi: 10.1109/TKDE.2014.2362525]
- [35] Zhang C, Wang K. POI recommendation through cross-region collaborative filtering. *Knowledge and Information Systems*, 2016, 46(2):369–387. [doi: 10.1007/s10115-015-0825-8]
- [36] Pham TAN, Li X, Cong G. A general model for out-of-town region recommendation. In: *Proc. of the 26th Int'l Conf. on Republic and Canton of Geneva: Int'l World Wide Web Conf.* 2017. 401–410. [doi: 10.1145/3038912.3052667]
- [37] Zhang C, Liang H, Wang K. Trip recommendation meets real-world constraints: POI availability, diversity, and traveling time uncertainty. *ACM Trans. on Information Systems (TOIS)*, 2016,35(1):5. [doi: 10.1145/2948065]
- [38] Yin H, Cui B, Chen L, *et al.* Modeling location-based user rating profiles for personalized recommendation. *ACM Trans. on Knowledge Discovery from Data (TKDD)*, 2015,9(3):19. [doi: 10.1145/2663356]
- [39] Yin H, Cui B, Zhou X, *et al.* Joint modeling of user check-in behaviors for real-time point-of-interest recommendation. *ACM Trans. on Information Systems (TOIS)*, 2016,35(2):11. [doi: 10.1145/2873055]
- [40] Zhang JD, Chow CY. GeoSoCa: Exploiting geographical, social and categorical correlations for point-of-interest recommendations. In: *Proc. of the 38th Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval*. New York: ACM Press, 2015. 443–452. [doi: 10.1145/2766462.2767711]
- [41] Gao H, Tang J, Liu H. Addressing the cold-start problem in location recommendation using geo-social correlations. *Data Mining and Knowledge Discovery*, 2015,29(2):299–323. [doi: 10.1007/s10618-014-0343-4]
- [42] Bao J, Zheng Y, Mokbel MF. Location-Based and preference-aware recommendation using sparse geo-social networking data. In: *Proc. of the 20th Int'l Conf. on Advances in Geographic Information Systems*. New York: ACM Press, 2012. 199–208. [doi: 10.1145/2424321.2424348]
- [43] Ference G, Ye M, Lee WC. Location recommendation for out-of-town users in location-based social networks. In: *Proc. of the 22nd ACM Int'l Conf. on Information & Knowledge Management*. New York: ACM Press, 2013. 721–726. [doi: 10.1145/2505515.2505637]
- [44] Zhang W, Wang J. Location and time aware social collaborative retrieval for new successive point-of-interest recommendation. In: *Proc. of the 24th ACM Int'l on Conf. on Information and Knowledge Management*. New York: ACM Press, 2015. 1221–1230. [doi: 10.1145/2806416.2806564]
- [45] Ye J, Zhu Z, Cheng H. What's your next move: User activity prediction in location-based social networks. In: *Proc. of the 2013 SIAM Int'l Conf. on Data Mining*. Bangkok: Siam Society, 2013. 171–179. [doi: 10.1137/1.9781611972832.19]
- [46] Chen C, Lukasiewicz T, Meng X, *et al.* Location-Aware news recommendation using deep localized semantic analysis. In: *Proc. of the Int'l Conf. on Database Systems for Advanced Applications*. Cham: Springer-Verlag, 2017. 507–524. [doi: 10.1007/978-3-319-55753-3_32]
- [47] Bermingham L, Lee I. Spatio-Temporal sequential pattern mining for tourism sciences. In: *Proc. of the Int'l Conf. on Computational Science*, Vol.29. Leiden: Elsevier, 2014. 379–389. [doi: 10.1016/j.procs.2014.05.034]
- [48] Kurashima T, Iwata T, Irie G, *et al.* Travel route recommendation using geotagged photos. *Knowledge and Information Systems*, 2013,37(1):37–60. [doi: 10.1007/s10115-012-0580-z]
- [49] Bhargava P, Phan T, Zhou J, *et al.* Who, what, when, and where: Multi-Dimensional collaborative recommendations using tensor factorization on sparse user-generated data. In: *Proc. of the 24th Int'l Conf. on World Wide Web*. New York: ACM Press, 2015. 130–140. [doi: 10.1145/2736277.2741077]
- [50] Wang X, Leckie C, Chan J, *et al.* Improving personalized trip recommendation by avoiding crowds. In: *Proc. of the 25th ACM Int'l Conf. on Information and Knowledge Management*. New York: ACM Press, 2016. 25–34. [doi: 10.1145/2983323.2983749]
- [51] Yu X, Pan A, Tang LA, *et al.* Geo-Friends recommendation in GPS-based cyber-physical social network. In: *Proc. of the Int'l Conf. on Advances in Social Networks Analysis and Mining*. Washington: IEEE Computer Society, 2011. 361–368. [doi: 10.1109/ASONAM.2011.118]
- [52] Tsai CY, Chung SH. A personalized route recommendation service for theme parks using RFID information and tourist behavior. *Decision Support Systems*, 2012,52(2):514–527. [doi: 10.1016/j.dss.2011.10.013]

- [53] Ding Y, Liu S, Pu J, *et al.* Hunts: A trajectory recommendation system for effective and efficient hunting of taxi passengers. In: Proc. of the IEEE Int'l Conf. on Mobile Data Management. Piscataway: IEEE, 2013. 107–116. [doi: 10.1109/MDM.2013.21]
- [54] Dai J, Yang B, Guo C, *et al.* Personalized route recommendation using big trajectory data. In: Proc. of the 31st IEEE Int'l Conf. on Data Engineering (ICDE). Washington: IEEE Computer Society, 2015. 543–554. [doi: 10.1109/ICDE.2015.7113313]
- [55] Huang J, Huangfu X, Sun H, *et al.* Backward path growth for efficient mobile sequential recommendation. IEEE Trans. on Knowledge and Data Engineering, 2015,27(1):46–60. [doi: 10.1109/TKDE.2014.2298012]
- [56] Zheng VW, Cao B, Zheng Y, *et al.* Collaborative filtering meets mobile recommendation: A user-centered approach. In: Proc. of the 24th AAAI Conf. on Artificial Intelligence, Vol.10. AI Access Foundation, 2010. 236–241.
- [57] Pham TAN, Li X, Cong G, *et al.* A general recommendation model for heterogeneous networks. IEEE Trans. on Knowledge and Data Engineering, 2016,28(12):3140–3153. [doi: 10.1109/TKDE.2016.2601091]
- [58] Lu Y, Qiao Z, Zhou C, *et al.* Location-Aware friend recommendation in event-based social networks: A Bayesian latent factor approach. In: Proc. of the 25th ACM Int'l Conf. on Information and Knowledge Management. New York: ACM Press, 2016. 1957–1960. [doi: 10.1145/2983323.2983883]
- [59] Lian D, Zheng K, Ge Y, *et al.* GeoMF++: Scalable location recommendation via joint geographical modeling and matrix factorization. ACM Trans. on Information Systems (TOIS), 2018,36(3):33. [doi: 10.1145/3182166]
- [60] Chow CY, Mokbel MF. Privacy of spatial trajectories. In: Proc. of the Computing with Spatial Trajectories. New York: Springer Science+Business Media, 2011. 109–141. [doi: 10.1007/978-1-4614-1629-6_4]
- [61] Xue AY, Zhang R, Zheng Y, *et al.* Destination prediction by sub-trajectory synthesis and privacy protection against such prediction. In: Proc. of the 29th IEEE Int'l Conf. on Data Engineering. Washington: IEEE Computer Society, 2013. 254–265. [doi: 10.1109/ICDE.2013.6544830]
- [62] Lathia N. The anatomy of mobile location-based recommender systems. In: Recommender Systems Handbook. 2nd ed., New York: Springer Science+Business Media, 2015. 493–510. [doi: 10.1007/978-1-4899-7637-6_14]
- [63] Barranco MJ, Noguera JM, Castro J, *et al.* A context-aware mobile recommender system based on location and trajectory. In: Proc. of the Management Intelligent Systems. Berlin, Heidelberg: Springer-Verlag, 2012. 153–162. [doi: 10.1007/978-3-642-30864-2_15]
- [64] Tai CH, Yang DN, Lin LT, *et al.* Recommending personalized scenic itinerary with geo-tagged photos. In: Proc. of the IEEE Int'l Conf. on Multimedia and Expo. Piscataway: IEEE, 2008. 1209–1212. [doi: 10.1109/ICME.2008.4607658]
- [65] Liu Q, Wu S, Wang L, *et al.* Predicting the next location: A recurrent model with spatial and temporal contexts. In: Proc. of the 30th AAAI Conf. on Artificial Intelligence. Palo Alto: AAAI Press, 2016. 194–200.
- [66] Yang C, Sun M, Zhao WX, *et al.* A neural network approach to jointly modeling social networks and mobile trajectories. ACM Trans. on Information Systems (TOIS), 2017,35(4):36. [doi: 10.1145/3041658]
- [67] Zou B, Li C, Tan L, *et al.* Location-Based recommendation using incremental tensor factorization model. In: Proc. of the Int'l Conf. on Advanced Data Mining and Applications. Cham: Springer-Verlag, 2014. 227–238. [doi: 10.1007/978-3-319-14717-8_18]
- [68] Symeonidis P, Papadimitriou A, Manolopoulos Y, *et al.* Geo-Social recommendations based on incremental tensor reduction and local path traversal. In: Proc. of the 3rd ACM SIGSPATIAL Int'l Workshop on Location-Based Social Networks. New York: ACM Press, 2011. 89–96. [doi: 10.1145/2063212.2063228]
- [69] Pankratius V, Li J, Gowanlock M, *et al.* Computer-Aided discovery: Toward scientific insight generation with machine support. IEEE Intelligent Systems, 2016,31(4):3–10. [doi: 10.1109/MIS.2016.60]
- [70] Xiang L. Recommendation System Practice. Beijing: People's Posts and Telecommunications Press, 2012. 19–24 (in Chinese).
- [71] Zhu YX, Lu LY. Evaluation metrics for recommender systems. Journal of University of Electronic Science and Technology of China, 2012,41(2):163–175 (in Chinese with English abstract). [doi: 10.3969/j.issn.1001-0548.2012.02.001]
- [72] Ricci F. Mobile recommender systems. Information Technology & Tourism, 2010,12(3):205–231. [doi: 10.3727/109830511X12978702284390]
- [73] Meng XW, Wang F, Shi YC, Zhang YJ. Mobile user requirements acquisition techniques and their applications. Ruan Jian Xue Bao/Journal of Software, 2014,25(3):439–456 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4521.htm> [doi: 10.13328/j.cnki.jos.004521]

- [74] Mayer-Schönberger V, Cukier K, Wrote; Zhou T, Trans. Big Data: A Revolution That Will Transform How We Live, Work, and Think. Hangzhou: Zhejiang People's Publishing House, 2013. 197–198 (in Chinese).

附中文参考文献:

- [1] 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. <http://www.jos.org.cn/1000-9825/4292.htm> [doi: 10.3724/SP.J.1001.2013.04292]
- [11] 李婷,裴韬,袁焯城,等.人类活动轨迹的分类、模式和应用研究综述.地理科学进展,2014,33(7):938–948. [doi: 10.11820/dlkxjz.2014.07.009]
- [29] 孟祥旭,王晓东,周兴铭.基于物理轨迹数据和社会网络的泛化行程推荐.软件学报,2012,23(增刊 1):159–168. <http://www.jos.org.cn/1000-9825/12018.htm>
- [70] 项亮.推荐系统实践.北京:人民邮电出版社,2012.19–24.
- [71] 朱郁筱,吕琳媛.推荐系统评价指标综述.电子科技大学学报,2012,41(2):163–175. [doi: 10.3969/j.issn.1001-0548.2012.02.001]
- [73] 孟祥武,王凡,史艳翠,张玉洁.移动用户需求获取技术及其应用.软件学报,2014,25(3):439–456. <http://www.jos.org.cn/1000-9825/4521.htm> [doi: 10.13328/j.cnki.jos.004521]
- [74] Mayer-Schönberger V, Cukier K,著;周涛,译.大数据时代:生活、工作与思维的大变革.杭州:浙江人民出版社,2013.197–198.



孟祥武(1966—),男,山东招远人,博士,教授,博士生导师,CCF 高级会员,主要研究领域为网络服务,用户需求,推荐服务.



张玉洁(1969—),女,副教授,主要研究领域为网络服务,用户需求,推荐服务.



李瑞昌(1982—),男,博士生,CCF 专业会员,主要研究领域为社会化网络分析,数据挖掘,推荐服务.



纪威宇(1987—),男,博士生,CCF 学生会会员,主要研究领域为网络分析,数据挖掘,推荐服务.