

文章编号: 2095-2163(2022)03-0001-05

中图分类号: TN929.5

文献标志码: A

地理社交网络中群体推荐算法的综述

俞菲^{1,2}, 姜守旭²

(1 常熟理工学院 计算机科学与工程学院, 江苏 常熟 215500; 2 哈尔滨工业大学 计算学部, 哈尔滨 150001)

摘要: 随着移动互联网和智能手机的广泛应用,地理社交网络已经成为人们日常生活不可或缺的社交信息平台。由于人是一种具有典型社会属性的群体,群体活动是人在现实世界中的日常生活、工作的状态写照。因此,在这种新型的社交信息平台上,关于群体以及群体活动相关的推荐服务需求越来越成为当前推荐系统中的重要任务。虽然,目前已有大量研究来解决面向群体活动的群体推荐问题,但在群体推荐的研究中仍然存在一些问题:(1) 群体活动决策过程中群体成员对决策的贡献程度不能直接从数据中获得;(2) 现有的关于群体推荐的算法大都强调被推荐的群体是当前存在的,然而,事实上参加活动的群体大都是应活动的发起而组建的。因此,本文对地理社交网络中群体推荐算法的相关研究进行了系统地分析和整理;主要从当前的已有研究中采用的方法和研究相关的数据集两方面进行总结;最后,提出了该领域的几个潜在研究方向。

关键词: 地理社交网络; 群体推荐; 群体活动

The survey of group recommendation on Geo-Social Networks

YU Fei^{1,2}, JIANG Shouxu²

(1 School of Computer Science and Engineering, Changshu Institute of Technology, Changshu Jiangsu 215500, China;

2 Department of Computer Science, Harbin Institute of Technology, Harbin 150001, China)

[Abstract] With the wide application of mobile Internet and smart phones, geographic social network has become an indispensable social information platform in people's daily life. Since people are groups with typical social attributes, group activities are the portrayal of people's daily life and work in the real world. Therefore, in this new type of social information platform, the demand for recommendation services related to groups and group activities has increasingly become an important task in current recommendation systems. Although a lot of research has been done on the group recommendation problem of group activities, there are still some problems in the research of group recommendation: (1) the contribution of group members to decision-making in the process of group activity decision-making cannot be directly obtained from the data; (2) the most existing groups recommendation algorithms emphasize that recommendation groups currently exist. However, most of the groups that participate in the campaign are realitely formed in response to the campaign's initiation. Therefore, this paper mainly systematically analyzes and organizes related researches on group recommendation algorithms in heterogeneous geographic social networks from the methods and datasets used in existing research. Finally, several potential research directions in this field are proposed.

[Key words] Geo-Social Networks; group recommendation; group activity

0 引言

地理社交网络(Geo-Social Networks, GeoSN)是当前移动互联网不断发展的衍生网络。当前广泛应用的地理社交网络主要包括基于位置的社交网络(Location-Based Social Network, LBSN)和基于活动的社交网络(Event-Based Social Networks, EBSN)。当前基于位置的社交网络比较熟知的有:Foursquare、Gowalla、Meituan等;基于事件(活动)的社交网络当前应用广泛的有:Meetup、DoubanEvent、Plancast等。地理社交网络相比传统的社交网络主要区别在于:GeoSN将物理世界和虚拟的社交网络

融合在一起,记录了人们线上线下的活动行为,推动位置相关的推荐服务的产生。当前在地理社交网络中的推荐服务主要是根据推荐内容分为潜在好友推荐、位置推荐、媒体推荐、活动推荐等;然而,这些推荐服务的共同特点是:服务对象都面向单个用户。事实上,人是具有社会属性的,即人都是作为每个群体的成员在生活、工作。

面向群体活动的推荐服务在近些年也逐渐成为社交媒体中主要的应用服务。在已有的研究中,群体根据构成的方式可分为:临时用户群体和稳定用户群体。临时用户群体是指群体中的用户是因为活动而临时组建的,当活动结束后,用户群体就会随着

作者简介: 俞菲(1987-),女,博士,讲师,主要研究方向:知识图谱、推荐系统、时空数据挖掘;姜守旭(1968-),男,博士,教授,主要研究方向:软件定义无线网、智能交通、物联网等。

通讯作者: 姜守旭 Email: jsx@hit.edu.cn

收稿日期: 2021-10-25

活动的完成而解散,这种群体典型的特点就是群体中的用户的不确定性;稳定用户群体是指群体中的用户大都是具有相同且稳定的偏好而组建的,这种偏好是一个长期的喜好,这种群体典型的特点是参加的活动大都是群体共同的偏好相关的。然而,现有关于群体推荐的研究不仅没有给出群体活动中群体的组建方式,也没有考虑不同的组建方式会导致推荐的内容不同。除此之外,群体推荐的研究中仍然存在一些问题:

(1) 群体活动决策过程中群体成员对决策的贡献程度不能直接从数据中获得;

(2) 已有的关于群体推荐的算法大都强调被推荐的群体是当前存在的。

本文将提供一个全面的文献回顾,主要从当前的已有研究中采用的方法和数据集两方面进行总结,并且对这些方法之间的异同给出详细的分析说明。

1 群体推荐算法

为了便于对地理社交网中群体推荐问题的理解,首先给出该问题的相关定义和符号说明。

定义 1 地理社交网 (Geo-Social Network, GeoSN) 是一个有向加权图:

$$g = \langle v, \varepsilon, w \rangle$$

其中, v 是由用户节点集合 U 和位置节点集合 L 组成; ε 表示相同类型节点间的链接和不同类型节点间的链接,包括: $\varepsilon_{u,u} = E_{u,u} \cup E_{l,l}$, $\varepsilon_{heter} = E_{u,l} \cup E_{l,u}$; $E_{u,u}$ 和 $E_{l,l}$ 分别表示用户间的社交关系、位置之间的序列关联关系; $E_{u,l}$ 和 $E_{l,u}$ 分别表示用户与位置之间的签到关系、位置与用户的时空关联关系; w 表示相同类型节点间的链接和不同类型节点间的链接上的影响力权值: $w = \{w_{homo}, w_{heter}\}$ 。

定义 2 地理社交网络中的群体: 已知地理社交网络 $g = \langle v, \varepsilon, w \rangle$, 群体 v_g 由 v 中的 n_{v_g} 个节点组成, 节点的种类 $\Gamma(v_g) = \{user, location\}$, $G(v_g)$ 是由 v_g 生成的子图, 即 $G(v_g) \in g, v_g \subseteq v$, 以及 $E_{v_g} \subseteq \varepsilon$ 。本文主要研究的推荐服务对象是用户群体 $U_g = v_g \subseteq U$ 。

定义 3 地理社交网络中的群体推荐: 已知地理社交网 $g = \langle v, \varepsilon, w \rangle$, 目标用户群体 U_g 以及群体中用户的签到活动记录 $D(U_g)$, 常整数 $k \in N^+$, 群体推荐是为目标用户群体 U_g 推荐 $Top - K$ 个活动构成的列表 S_{rec} , 使得被推荐的活动是目标用户群体可能喜欢、感兴趣的活动。

已有的地理社交网络中的面向群体活动的群体推荐算法, 主要可以分为基于记忆的方法 (memory-based methods) 和基于模型的方法 (model-based methods)。

1.1 基于记忆的群体推荐算法

现有, 基于记忆的方法可分为基于记忆和基于模型两种, 其中, 基于记忆的方法又可分为偏好聚合方法和分数聚合方法。

偏好聚合方法是基于群体中所有用户偏好为用户群体推荐符合群体偏好的活动。McCarthy 等人^[1]推出了一个在公共环境中协调组成员对音乐的偏好的系统, 利用组成员偏好代理选择最适合组成员偏好的音乐; Yu 等人^[2]通过合并用户资料提供给多个观众, 从而确保合并后的结果接近大多数用户的偏好; Wang 等人^[3]设计了一种基于信任感知的虚拟协调器的社会群体推荐方法, 集成了结果聚合和汇总聚合两种不同的聚合策略; Zou 等人^[4]主要解决的问题是消除社会群体推荐过程中的偏差, 通过利用子群体偏好计算项目相关性, 设计了一个群体推荐的局部优化框架; Minz 等人^[5]提出了一种社会群体推荐的共识方案, 将多个专家的共识汇集到一个整体推荐列表中, 每个专家代表群体中的一个人; Du 等人^[6]将群体推荐问题形式化为一个排序问题, 提出了一种基于学习排序技术的群体事件推荐模型。

分数聚合方法是先预测群体中每个用户在候选项目上的得分, 通过预定义的策略来聚合群体内成员的预测得分, 代表群体的偏好。两种最受欢迎的得分汇总算法是 Average (AVG), Least Misery (LM)。AVG 算法的核心思想是将群体中所有用户个体对群体活动 (item) 打分取均值, 将均值作为用户群体对群体活动的打分, 该算法将群体的打分均值作为目标函数; LM 算法是在用户群体中选择对活动 (item) 的最低打分作为整个用户群体的打分, 核心思想是让群体中最可能不满意的用户都尽可能的满意。

以上两种方法中存在不足。具体地, AVG 算法可能会返回一些对群体中某些成员有利但对其他成员不利的活动, 而 LM 最终可能会推荐一些没有人喜欢的活动。Baltrunas 等人^[7]指出, 任何一种算法的表现都取决于群体规模和群体内用户间相似性; Yahia 等人^[8]考虑了群体成员在每个项目上的偏好分歧是不可避免的, 即群体中用户之间的相关性和不一致性, 从而提高了 AVG 和 LM 算法的推荐质量。

1.2 基于模型的群体推荐算法

基于模型的方法对群体推荐项目的决策过程建模。Liu 等人^[9]提出了一个用于群体推荐的个人影响的主题模型,假设最有影响力的用户应该代表群体,并对群体的决策有很大的影响;Yuan 等人^[10]认为群体活动的决策过程是受到群体中每个用户对活动(item)的偏好主题以及整群体关于活动(item)的偏好主题的影响。因此,提出一种综合考虑活动(item)主题的共识模型来解决群体推荐问题,Hu 等人^[11]引入了深度学习建模框架,该算法对用户群体中所有用户的偏好进行综合特征的学习与提取,该算法的优点是避免数据的脆弱性。

随着注意力机制在深度学习以及认知科学中的广泛应用,关于群体推荐研究中也开始将群体活动的决策过程通过注意力机制来体现和表示。TRAN 等人^[12]首次提出在群体活动中,群体中的成员往往可能只遵循少数用户的决策,这些用户是群体的领导者或专家,为了解决这一挑战,TRAN 等人提出使用注意机制来捕捉组中每个用户的影响;Yin 等人^[13]基于在群体活动中,每个用户在不同活动以及群体中对活动的决策作用应该是不同的,因此提出了一种新的群体推荐系统,以注意机制和二部图嵌入模型 BGEM 为构建模块,该算法采用关注机制来学习每个用户的社会影响力,使其适应不同群体的社会影响力,并开发出一种新的深度社会影响力学习框架,挖掘和整合用户的全局和局部社会网络结构信息,进一步提高对用户社会影响力的估计。

表1 群体推荐主要使用的实验数据集

Tab. 1 The main experiment datasets in Group Recommendation Research

数据	#User	#Group	#Item	#U-I interaction	#G-I interaction
Weeplace	8,643	22,733	25,081	1,358,458	180,229
Yelp ^[1]	34,504c	24,103	22,611	482,365	26,995
Mafengwo	5,275	995	1,513	39,761	3,595
CAMRa2011 ^[2]	602	290	7,710	11,344	145,068
Douban	2,964	2,630	39,694	823,851	463,040
Gowalla	25,405	44,565	38,306	1,025,500	154,526

注:#U-I和#G-I表示用户与item的交互次数、用户群体与item的交互次数。

3 群体推荐的评价指标

现有研究中,面向群体推荐的评价指标主要有:准确率、召回率、标准化贴现累积、命中率以及平均倒数排名。

准确率 Precision Rate (Pre),式(1)~(2):

$$Pre_u @ K = \frac{|S_{rec} \cap I_u^{test}|}{|S_{rec}|} = \frac{|S_{rec} \cap I_u^{test}|}{K} \quad (1)$$

虽然现有研究已经对群体活动中的决策过程进行了分析和表示学习,采用启发式或基于注意的偏好聚合策略来合成群体偏好,但是这些模型主要侧重于用户之间的成对连接,而忽略了群体内外复杂的高阶交互。此外,由于用户群体-活动(item)之间的交互非常稀疏,使得群推荐存在严重的数据稀疏性问题。Zhang 等人^[14]提出了一种用于群体推荐的自监督超图学习框架,通过捕获用户群体内和群体间的交互来缓解原始数据本身的数据稀疏问题;Guo 等人^[15]提出了一种基于群体之间的相似性的群体推荐 HyperGroup,通过基于学习到的群体成员个人偏好的超边缘嵌入技术学习群体偏好,该方法主要目的也是克服群体活动中用户群体与活动之间的交互信息稀疏的问题。

2 群体推荐相关的数据集

当前,地理社交网络已经成为人们的主要社交平台,主要可以分为两种社交网络:基于位置的社交网络(LBSN)以及基于事件(活动)的社交网络(EBSN)。其中,EBSN是一种典型的面向群体活动的社交平台,因此,在异构地理社交网中做群体推荐相关研究时,主要在EBSN的数据集上,由于部分研究需要考虑到EBSN关于单个用户的个性化偏好分析,LBSN可以为单个用户的偏好建模提供相对完备的数据基础。本文给出在异构地理社交网络中主要使用的公开数据集:Weeplace、Yelp^[1]、Mafengwo、CAMRa2011^[2]、Douban、Gowalla,以及数据中的统计内容。具体信息见表1。

$$Pre @ K = \frac{\sum_{u \in U} Pre_u @ K}{|U|} \quad (2)$$

其中,Pre@K表示用户u的准确率; S_{rec} 表示被推荐的内容(item); I_u^{test} 表示在测试集中用户u真实交互过的内容(item);K表示推荐列表中内容的数量。

召回率 Recall Rate (Rec)^[12]:

$$Rec_u @ K = \frac{|S_{rec} \cap I_u^{test}|}{|I_u^{test}|} \quad (3)$$

$$Rec_u@K = \frac{\sum_{u \in U} Rec_u@K}{|U|} \quad (4)$$

其中, $Rec_u@K$ 表示用户 u 的召回率; S_{rec} 表示被推荐的内容 (item); I_u^{test} 表示在测试集中用户 u 真实交互过的内容 (item); K 表示推荐列表中内容的数量。

标准化贴现累积 Normalized Discounted Cumulative Gain (nDCG), 式(5)~(7):

$$nDCG_u@K = \frac{DCG_u}{\max DCG_u} \quad (5)$$

$$DCG_u = rel_1 + \sum_{i=2}^K \frac{rel_i}{\log_2(i)} \quad (6)$$

$$nDCG@K = \frac{\sum_{u \in U} DCG_u@K}{|U|} \quad (7)$$

其中, rel_i 表示对目标用户 u 是否接受被推荐的内容: 如果接受, 则 $rel_i = 1$, 反之, $rel_i = 0$; $\max DCG_u$ 表示对目标用户 u 的最优推荐列表的 DCG 值, 这样做的目的是为了保证 $nDCG_u@K$ 的取值范围在 0 ~ 1 之间。

命中率 Hit Ratio (HR) [12-13]:

$$HR@K = \frac{\#hit@K}{|D_{test}|} \quad (8)$$

其中, 对于测试集中的每一个组-项交互 (U_g, v), $\#hit@K$ 是指在测试集中 hit 的数量, $|D_{test}|$ 表示测试用例的总数。

首先计算项目 v 和所有候选项目的排名得分; 选取排名分数最高的 K 个项目作为 Top- K 推荐列表。如果项目 v 出现在这个列表中, 就有一个命中。否则, 就错过了。

平均倒数排名 Mean Reciprocal Rank (MRR), 表示正确检索结果值在检索结果中的排名, 用来评估检索系统的性能, 式(9):

$$MRR = \frac{1}{|D_{test}|} \sum_{(U_g, v) \in D_{test}} \frac{1}{rank(v)} \quad (9)$$

其中, $|D_{test}|$ 表示测试用例的总数, MRR 表示测试用例中用户群体交互过的项目, 在推荐列表中第一个在 ground-truth 中的项目所在的排列位置。

一个好的推荐模型应该有较大的 MRR 值。

4 未来的方向

根据本文对现有群体推荐相关研究的分析, 发现大部分已有研究忽略了群体活动中用户群体的组建方式对推荐结果的影响, 以及用户在个体活动和

群体活动中的偏好区别上的分析。事实上, 在地理社交网络中面向群体活动的群体推荐应该考虑现实生活中, 许多参加活动的群体都是因为活动本身而组建的。因此, 在面向这种情况的群体推荐, 应该考虑群体组件与活动本身之间的相互影响、相互作用以及群体中的用户与活动之间历史的交互行为。

基于以上分析, 关于地理社交网络中的群体推荐中关于群体组建过程与活动之间的因果关系的建模将成为接下来的亟待解决的问题。

5 结束语

本文对地理社交网络中群体推荐相关研究进行了全面的综述。展示了在这个新兴的研究领域中, 大量最近的研究论文的深刻发现; 基于群体推荐算法的数据建模的机理, 将群体推荐算法分为基于记忆的方法和基于模型的方法; 介绍了群体推荐算法相关研究的数据集以及广泛使用的评价指标; 最后, 基于对已有研究的分析, 本文提出了面向群体推荐, 群体组建过程与活动之间的因果关系的建模将成为接下来的亟待解决的问题。

参考文献

- [1] MCCARTHY J F, ANAGNOST T D. MusicFX: an arbiter of group preferences for computer supported collaborative workouts[C]// in Proceeding of the ACM conference on computer supported cooperative work, 1998:363-372.
- [2] YU Z, ZHOU X, HAO Y, et al. TV Program Recommendation for Multiple Viewers Based on user Profile Merging[J]. User Modeling and User-Adapted Interaction, 2006, 16(1):63-82.
- [3] WANG Q, YIN H, CHEN T, et al. Next Point-of-Interest Recommendation on Resource-Constrained Mobile Devices[C]// in Proceedings of The Web Conference, 2020:906-916.
- [4] YIN H, ZOU L, NGUYEN Q V H, et al. Joint event-partner recommendation in event-based social networks[C]// in IEEE international conference on Data Engineering, 2018:929-940.
- [5] XIAO L, MIN Z, YONGFENG Z, et al. Fairness-Aware Group Recommendation with Pareto-Efficiency[J]. ACM, 2017:107-115.
- [6] DU Y, MENG X., ZHANG Y, et al. GERF: A group event recommendation framework based on learning-to-rank[J]. IEEE Transactions on Knowledge and Data Engineering, 2019, 32(4): 674-687.
- [7] Linas Baltrunas, Tadas Makcinskas, Francesco Ricci. Group recommendations with rank aggregation and collaborative filtering[C]// In Proceedings of the fourth ACM conference on Recommender systems, 2010:119-126.
- [8] ROY S B, THIRUMURUGANATHAN S, AMER-YAHIA S, et al. Exploiting group recommendation functions for flexible preferences[C]// in IEEE international conference on Data Engineering, 2014:412-423.