

# 基于社区发现的兴趣点推荐

龚卫华 沈松

(浙江工业大学计算机科学与技术学院 杭州 310023)

**摘要** 近年来,LBSN(Location-based Social Networks)作为一种典型的异质信息网络越来越受到大众的关注。针对 LBSN 中用户签到信息十分稀疏的情况,文中提出了一种基于社区发现的兴趣点推荐算法 CBR(Community-Based Recommendation)。该算法首先在社交媒体层上计算目标用户与聚类后的兴趣主题簇的相似度;其次通过兴趣主题簇与地理位置簇之间的关联矩阵  $R$  计算用户在地理位置簇上的隶属度;然后进一步融合用户的社交关系,从而得到用户对各个兴趣点的偏好分数;最后按照兴趣点的分数进行排序,以实现 Top- $k$  推荐。实验结果表明,该算法可以明显提高兴趣点的推荐质量。

**关键词** 多维关系,社区发现,兴趣点推荐,地理位置簇

**中图分类号** TP391 **文献标识码** A **DOI** 10.11896/jsjcx.190400440

## Community Detection Based Point-of-interest Recommendation

GONG Wei-hua SHEN Song

(School of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China)

**Abstract** In recent years,LBSN (Location-based Social Networks) has attracted more and more attention as a typical heterogeneous information network. In view of the sparse check-in information of users in LBSN,this paper proposed a recommendation algorithm CBR (Community-Based Recommendation) based on community detection. It first calculates the similarity between the target user and the clustered interest topic cluster on the social media layer,and then calculates the user's membership degree on the geographic cluster through the association matrix  $R$  between the interest topic cluster and the geographic cluster. Then it further integrates user's social relationship to get user's preference scores for each point of interest,and finally sorts according to the interest scores to achieve the Top- $k$  recommendation. The experimental results show that the proposed algorithm can significantly improve the recommendation quality of the points of interest.

**Keywords** Multidimensional relationship,Community detection,Point-of-interest recommendation,Geographic cluster

## 1 引言

近年来,在线社交网络与位置服务的深度融合与发展,催生了基于位置的社交网络(LBSN)。这种新型网络不仅包含传统社交网络中用户间的社交关系,还包含用户在不同地点的签到关系以及用户的社交媒体信息,因此具有更复杂的关系结构特征,能够提供更广泛的应用服务。例如 FourSquare, FacebookPlaces, Yelp 等不仅类似于传统社交网络具有社交功能,还能提供信息共享与各种推荐等应用服务,兴趣点推荐就是其中之一。

然而,虽然 LBSN 包含了用户丰富的线上线下社交信息,现实中大部分用户的签到数据却十分稀疏,这对于分析预测用户将要访问的兴趣点具有极大的阻碍。但现在流行的兴趣

点推荐算法都只能通过数据处理缓解签到密度的稀疏性问题,无法对签到密度近乎为零的个别用户进行推荐。大多数现有的兴趣点推荐算法缺乏对用户的适应性,对几乎所有用户使用相同的算法,因此对签到密度比一般用户更加稀疏的用户进行推荐是目前兴趣点推荐中少有人关注的难题。

经过大量研究,LBSN 的线上兴趣主题社区与线下的地理位置簇存在着某种关联,本文设定该矩阵为  $R$ ,通过目标用户与兴趣主题簇的相似度得到其在兴趣主题簇的分布,再通过该关联矩阵间接得到用户在地理位置簇上的分布,从而弥补用户签到信息稀疏的不足。

本文第 2 节介绍了社区发现的研究现状;第 3 节给出了基于社区发现的兴趣点推荐算法;第 4 节进行仿真实验及结果分析;最后总结全文。

收到日期:2019-04-17 返修日期:2019-06-13 本文受国家自然科学基金青年项目(61502420),浙江省自然科学基金项目(LY13F020026, LY16F020032),中国博士后科学基金项目(2015M581957),浙江省教育厅科研项目(Y201840116)资助。

龚卫华(1977-),男,博士,副教授,主要研究方向为社交网络、数据挖掘,E-mail:619626727@qq.com(通信作者);沈松(1993-),男,硕士生,主要研究方向为社交网络。

## 2 相关工作

兴趣点推荐研究发展至今,从技术角度主要分为以下4种:协同过滤方法、矩阵分解模型、概率模型和基于链接的方法。

协同过滤是推荐系统中使用最广泛的技术,也适用于LBSN中的兴趣点推荐。它假设具有相似访问历史的用户在将来也会访问类似的兴趣点,并通过寻找与目标用户具有类似偏好的其他用户来推荐兴趣点。文献[1]通过基于用户及其朋友的协同过滤,融合用户偏好和社会影响,提出了统一的兴趣点推荐框架。该研究还观察到用户可能会前往不同的地方并创建多个签到空间,并且用户的签到到兴趣点往往相距不远,因此采用幂律函数来捕捉兴趣点之间的地理影响。文献[2]利用基于朋友的协同过滤和二维核密度估计,将社会影响力和地理位置影响力整合到统一的兴趣点推荐框架当中。文献[3]针对活跃用户和非活跃用户提出了两种基于用户的协同过滤方法,该方法利用了时空特征与签到序列来进行兴趣点推荐。文献[4]提出基于元路径的兴趣点推荐算法,引入元路径来描述节点间不同类型的关联关系,通过随机游走方法计算元路径的特征值,以度量实路径中首尾节点间的关联度,最后根据用户和兴趣点的关联度进行推荐。文献[5]认为用户评分由用户属性、兴趣点属性、兴趣点距离共同决定,在此基础上构建了贝叶斯模型来预估用户对新兴兴趣点的评分。

矩阵分解模型的推荐方法十分灵活,可以扩展为包含很多潜在影响因子的算法。文献[6]将用户的历史签到信息细分为24个小时,并利用高阶奇异值分解算法得到不同时间段下用户主题的兴趣点分布,以此进行兴趣点推荐。文献[7]融合了兴趣点的地理、文本、社会、流行度等信息,提出了一种上下文感知的概率矩阵分解兴趣点推荐算法TGSC-PMF。文献[8]提出了联合加权矩阵分解和二维核密度估计来构建地理形态的模型,有助于解决矩阵的稀疏性问题,提高推荐性能。文献[9]采用贝叶斯非负矩阵分解模型来编码空间影响和个性化偏好,其假设这些位置聚集在许多高斯地理分布的潜在区域中,并且使用幂律参数项来模拟距离因子。文献[10]从位置角度利用两级地理邻域特征,提出了一种使用加权矩阵分解的交替优化算法。

在兴趣点推荐中,概率模型可以描述多种因素影响之下的用户兴趣点概率分布。文献[11]使用高斯混合模型来探索其登记历史记录,并自动学习用户的活动规律。文献[12]提出了一个概率模型来计算用户标签和位置偏好关键字之间的映射关系,用于预测上下文相关的位置,缓解了数据稀疏问题。文献[13]提出了一种时间匹配泊松分解模型(TM-PFM)来描述兴趣点的流行度,模拟用户的规律性,并将用户和兴趣点之间的时间匹配纳入整体考虑。文献[14]提出了一种统一的概率生成模型,用于识别用户签到活动的语义、时空模式,以支持兴趣点推荐。文献[15]提出了一种概率框架,利用核密度估计来计算用户和位置的时间影响相关性。

基于链接的模型通过用户与兴趣点之间的拓扑图链接来进行推荐。文献[16]进一步包括了用户之间的关系,在拓扑图中加入了用户之间的关联边并将其用于推荐。文献[17]提

出了地理时间影响感知图来建模签到记录,再利用广度优先传播算法进行有效推荐。文献[18]提出了融合多方面特征的兴趣点推荐模型MAPS,考虑了分类、时间、社会 and 空间等四维因素。文献[19]在规范化的登记网络上执行基于Hits的登记游走,以捕获用户意图对登记行为的关联关系。文献[20]借鉴马尔可夫模型对移动对象的历史轨迹建模,并结合对象的运动趋势进行综合推荐。

综上所述,现有的兴趣点推荐算法虽然考虑了多维因素,但总体来说都是对单个用户的处理,即便考虑了用户之间的社交以及兴趣联系,也没有对用户社区群体进行处理,提出的方案也只能缓解用户签到数据的稀疏性,而无法有效提高推荐性能,因此研究基于社区发现的兴趣点推荐方法具有十分重要的意义和价值。然而现在流行的兴趣点推荐算法都只能通过数据处理来缓解签到到密度的稀疏性问题,无法对签到密度近乎为零的个别用户进行推荐。因此,对签到密度比一般用户更加稀疏的用户进行推荐是目前兴趣点推荐的难题,也是本文需要解决的问题。

## 3 兴趣点推荐模型的建立过程

### 3.1 问题描述

将已得到的兴趣主题簇与地理位置簇进行进一步建模,假设模型中有 $d$ 个用户 $UM = \{um_1, um_2, \dots, um_d\}$ , $m$ 个主题 $T = \{t_1, t_2, \dots, t_m\}$ , $x$ 个兴趣点 $L = \{l_1, l_2, \dots, l_x\}$ ,以及这些实体之间的关系矩阵。关系矩阵具体如下。

(1)矩阵 $F_{d \times m}$ :表示模型中的用户与主题之间的兴趣关系矩阵,其中 $f_{ij}$ 表示用户 $um_i$ 发布了主题 $t_j$ 。

(2)矩阵 $G_{d \times x}$ :表示模型中的用户与兴趣点之间的签到关系矩阵,其中 $g_{ij}$ 表示用户 $um_i$ 在兴趣点 $l_j$ 上签到。

(3)矩阵 $P_{x \times g}$ :表示兴趣点对地理位置簇的隶属矩阵,其中 $p_{ij} = 1$ 表示兴趣点 $l_i$ 隶属于地理位置簇 $j$ 。

(4)矩阵 $S_{m \times c}$ :表示兴趣主题簇上主题的分布矩阵,其中 $s_{ij}$ 表示兴趣簇 $i$ 上主题 $t_j$ 所占的频率。

(5)矩阵 $R_{c \times g}$ :表示兴趣主题簇与地理位置簇的关联矩阵。

(6)矩阵 $O_{d \times d}$ :表示模型中的用户之间的社交关系矩阵,其中 $o_{ij} = 1$ 表示用户 $um_i$ 与用户 $um_j$ 存在社交关系, $o_{ij} = 0$ 则表示不存在社交关系。

假设目标用户集 $UT = \{ut_1, ut_2, \dots, ut_n\}$ ,其与模型中实体的关系矩阵表示如下。

(1)矩阵 $W_{n \times d}$ :表示模型中的用户与目标用户的社交关系矩阵,其中 $w_{ij} = 1$ 表示目标用户 $ut_i$ 与模型中的用户 $um_j$ 存在社交关系, $w_{ij} = 0$ 则表示不存在社交关系。

(2)矩阵 $B_{n \times m}$ :表示目标用户与主题之间的兴趣关系矩阵,其中 $b_{ij}$ 表示用户 $ut_i$ 发布了主题 $t_j$ 。

本文的任务为:向任意目标用户 $ut_i \in UT$ 推荐Top- $k$ 个兴趣点集合 $POI_i$ ,即 $POI_i = \{l_{i1}, l_{i2}, \dots, l_{iN}\}, l_{ij} \in L, j = 1, 2, \dots, K$ 。

### 3.2 用户与兴趣主题簇的相似性计算

在LBSN的社交媒体层中,为了获得目标用户在每个兴

趣主题簇上的分布,首先需要计算目标用户与兴趣主题簇的主题相似性,具体公式如下:

$$q_{ij} = \frac{B_i \cdot S_j}{|B_i| |S_j|} \quad (1)$$

其中,  $B_i$  表示目标用户  $ut_i$  发布的主题列表,  $S_j$  表示兴趣主题簇的主题频率分布。计算这两项的余弦相似度,就可以得到目标用户  $ut_i$  与兴趣主题簇  $j$  之间的相似性。

对矩阵  $Q$  进行标准化处理:

$$q_{ij} = \frac{q_{ij}}{\sum_{j=1}^c q_{ij}} \quad (2)$$

其中,  $q_{ij} \in Q_{n \times x}$  表示目标用户  $ut_i$  与兴趣主题簇  $j$  的相似性占其与整个社交媒体层主题相似性的比例,可以用来表示用户  $ut_i$  在兴趣主题簇  $j$  上的隶属度。

### 3.3 用户与兴趣点的相似性评分

大量的研究表明,线上兴趣主题簇与线下地理位置簇之间存在一定程度的关联性,假设它们之间的关联矩阵为  $R$ 。在此前提下,为了实现对签到稀疏目标用户的兴趣点推荐,在计算目标用户与兴趣主题簇的相似性后,再借助关联矩阵  $R$  可以间接预测目标用户对地理位置簇上各兴趣点的评分。

目标用户对地理位置簇的隶属矩阵  $Z$  可由式(3)计算得到:

$$Z = Q \cdot R \quad (3)$$

其中,  $Q$  是由式(2)得到的目标用户与兴趣主题簇的隶属关系矩阵,  $R$  则是兴趣主题簇与地理位置簇的关联矩阵。显然,式(3)表明目标用户对地理位置的兴趣偏好是通过用户与主题簇关系矩阵以及主题簇与地理位置簇关联矩阵连乘传递而得。

目标用户与兴趣点的相似性评分与目标用户的朋友在其所属地理位置簇中兴趣点上的签到频率相关。目标用户所属地理位置簇中的兴趣点如式(4)所示:

$$WP = Z \cdot P^T \quad (4)$$

其中,  $wp_{ij} \in WP$  表示兴趣点  $l_j$  是目标用户  $ut_i$  所属地理位置簇中的兴趣点。矩阵  $Z$  表示式(3)计算得到的目标用户对地理位置簇的隶属矩阵,矩阵  $P$  表示兴趣点对地理位置簇的隶属矩阵,将其连乘得到目标用户与兴趣点的地理位置关联矩阵。

另一方面,用户  $ut_i$  与模型中的用户  $um_j$  的社交相似度为:

$$ws_{ij} = \begin{cases} \frac{W_i \cdot O_j}{|W_i| |O_j|}, & \text{if } w_{ij} = 1 \\ 0, & \text{else} \end{cases} \quad (5)$$

其中,向量  $W_i \in W$  表示目标用户  $ut_i$  的社交关系列表,  $O_j \in O$  表示模型中的用户  $um_j$  的社交关系列表,计算这两项的余弦相似度可得到用户之间的社交相似度。目标用户在加权社交关系影响下的兴趣点偏好矩阵如下:

$$V_i = WS_i \cdot G \quad (6)$$

其中,矩阵  $G$  表示模型中的用户对兴趣点的签到频率矩阵,  $WS_i$  是式(6)计算得到的目标用户  $ut_i$  与模型中用户的关系列表。两两之间的关系通过关联矩阵传递得到目标用户与兴趣点之间的关联关系。

通过对目标用户在社交媒体层以及地理位置层上的分析,得到了用户在兴趣层面上可能访问的兴趣点,然后将其通过用户的社交关系加以约束,具体公式为:

$$pr_{ij} = \omega p_{ij} \cdot v_{ij} \quad (7)$$

其中,  $\omega p_{ij}$  是由式(5)得到的用户所属地理位置簇中的兴趣点,  $v_{ij}$  表示目标用户与兴趣点之间的关联关系。

向量  $pr_i$  表示目标用户的兴趣点评分列表,将向量  $pr_i$  中的每个元素进行排序,取 Top- $k$  个兴趣点,最后将这  $k$  个兴趣点推荐给目标用户。兴趣点推荐算法的具体过程如算法 1 所示。

#### 算法 1 兴趣点推荐算法 (POI Recommendation)

输入: 兴趣主题簇矩阵  $S$ , 地理位置簇矩阵  $P$ , 社交关系矩阵  $W$  和  $O$ , 主题发布矩阵  $B$ , 兴趣点签到矩阵  $G$ , 主题相似性矩阵  $A$ , 模型用户兴趣主题簇分布矩阵  $F$ , 兴趣主题簇与地理位置簇关联矩阵  $R$

输出: 对每个目标用户  $ut_i$  推荐 Top- $k$  个兴趣点集合  $POI_i$

1. FOR EACH  $ut_i \in UT$  // 对于每个目标用户
2. FOR EACH  $j \leq c$  // 对于每个兴趣主题簇
3. 由式(1)得到  $q_{ij}$  // 用户与兴趣主题簇的相似度
4. 通过式(2)进行标准化,得到用户在兴趣簇上的隶属度  $q_{ij}$
5. END FOR
6. END FOR
7. 通过式(3)计算  $Z$  // 通过关联矩阵得到用户在位置簇上的隶属度
8. 通过式(4)计算  $WP$  // 目标用户所属地理位置簇中的兴趣点集合
9. FOR EACH  $ut_i \in UT$  // 对每个目标用户
10. FOR EACH  $um_k \in UM$  // 对每个模型中的用户
11. 通过式(5)计算  $ws_{ij}$  // 两个用户之间的社交关系相似性
12. END FOR
13. END FOR
14. FOR EACH  $ut_i \in UT$  // 对每个目标用户
15. 通过式(6)计算  $V_i$  // 目标用户和每个兴趣点的关联性
16. 通过式(7)计算  $pr_i$  // 得到每个用户的兴趣点评分列表
17. 对  $pr_i$  进行排序,实现 Top- $k$  推荐。
18. END FOR

算法 1 中的目标用户有  $n$  位,模型中的用户有  $d$  位,有  $c$  个兴趣主题簇,  $x$  个兴趣点。算法 1 中第 1—6 步计算了每个用户在兴趣主题簇中的隶属度,总共运行了  $cn$  次;第 9—13 步计算了目标用户与模型中用户的社交相似性,需要运行  $nd$  次;第 14—18 步对每个用户进行相似性评分,并进行兴趣点推荐,采用堆排序,每次选出 Top- $k$  个用户,运行次数近似为  $(\text{Top-}k) \cdot \log x$ ,总共需要运行  $n \cdot (\text{Top-}k) \cdot \log x$  次。因为  $c$  与 Top- $k$  是常数,所以算法 1 的时间复杂度为  $O(nd) + O(n \log x)$ 。

## 4 仿真与实验结果分析

### 4.1 实验数据集描述

本文实验使用的两种 LBSN 数据集分别是 Foursquare (NYC) 和 Yelp,实验之前先过滤数据集以移除一些异常数据。对于 Foursquare(NYC)数据集,筛选出超过 2 人签到的地理位置、评论数超过 5 条的用户及其所拥有的社交关系。

对于 Yelp 数据集,筛选出签到数超过 50 的地理位置数、评论超过 5 条的用户及其所拥有的社交关系。预处理完成后,将每个数据集按照 4:1 分成独立的训练集与测试集两部分。将训练集作为模型中的信息集合,进行社交媒体层与地理位置层的聚类分析,而测试集则作为目标用户的信息集合。预处理完成后,各数据集中的用户数、位置数、评论数以及社交关系和签到关系等统计信息如表 1、表 2 所列。

表 1 Foursquare(NYC)数据集的基本信息

Table 1 Basic information of Foursquare(NYC) dataset basic information

类别	用户数	签到数	社交链接数	位置数	评论数
测试集	963	12246	9146	27303	4356
训练集	3848	46781	31475	27303	16972

表 2 Yelp 数据集的基本信息

Table 2 Basic information of Yelp dataset

类别	用户数	签到数	社交链接数	位置数	评论数
测试集	3236	104228	65899	22005	21328
训练集	12948	411335	226658	22005	454086

由表 1、表 2 可知,Foursquare(NYC)测试集的平均链接数为 9.50,Yelp 数据集的平均链接数为 20.36,这表明 Yelp 数据集的社交关系紧密度高于 Foursquare。

#### 4.2 对比算法

为了验证本文提出的兴趣点推荐算法的有效性,选取了以下对比算法进行实验,对比算法都来自文献[21]。

Topic-based CF(T):基于评论主题计算兴趣相似度,再运用协同过滤算法进行兴趣点推荐。该算法只考虑了用户之间的兴趣主题相似性。

Friend-based CF(F):基于用户的共同好友计算社交相似度,再运用协同过滤算法进行兴趣点推荐。该算法只考虑了用户的社交关系。

Mixed R(TF):融合了上述两种算法,既考虑了用户之间的主题相似性,也考虑了用户之间的社交关系。

CBR(Community-Based Recommendation):本文提出的基于社区发现的兴趣点推荐方法,充分考虑并融合了用户之间的主题相似性、社交关系相似性以及用户的地理位置信息。

#### 4.3 评价指标

本文使用准确率(precision)、召回率(recall)、F-度量量和 MAP 来评估兴趣点推荐方法的性能。

准确率指推荐结果中目标用户真正签到的兴趣点占推荐其访问兴趣点的比例。召回率指推荐结果中目标用户真正签到的兴趣点占目标用户访问兴趣点总量的比例。这两个指标分别反映了推荐性能的准确性和全面性,具体定义如下:

$$precision = \frac{1}{|UT|} \cdot \sum_{u_i \in UT} \frac{|POI_i \cap RP_i|}{|RP_i|} \quad (8)$$

$$recall = \frac{1}{|UT|} \cdot \sum_{u_i \in UT} \frac{|POI_i \cap RP_i|}{|POI_i|} \quad (9)$$

其中, $POI_i$ 表示目标用户  $u_i$  的推荐列表, $RP_i$ 为假设其实际访问的兴趣点列表。

F-度量(F-measure)是准确率和召回率的调和平均值,其定义如下:

$$F_\beta = (1 + \beta^2) \cdot \frac{precision \times recall}{\beta^2 \cdot precision + recall} \quad (10)$$

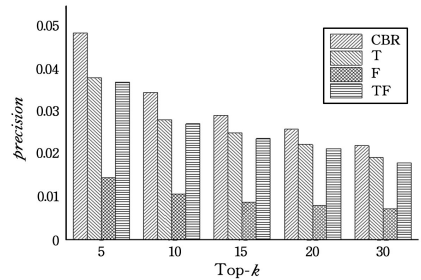
本文中  $\beta$  被设定为 1。当  $\beta=1$  时,F-度量在准确率和召回率中取得了平衡。

以上指标都与参数 Top-k 有关,而 MAP 指标则描述了在所有参数下的准确率的平均值,可以从整体上把握算法的性能。

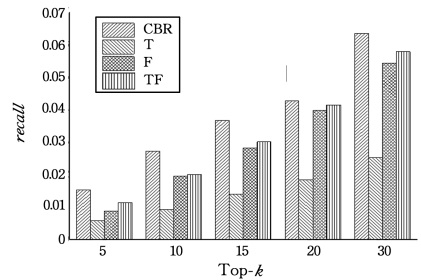
$$MAP = \frac{1}{|UT|} \cdot \sum_{u_i \in UT} ( \int precision d(recall) ) \quad (11)$$

#### 4.4 实验结果分析

本文比较了 4 种算法分别在 Foursquare(NYC)与 Yelp 数据集上的准确度,设定推荐的兴趣点个数 Top-k 分别为 5, 10, 15, 20 和 30。在图 1(a)所示的 Foursquare 数据集中可以看出,无论 Top-k 取何值,本文算法 CBR 都明显优于其他算法,最多比第二位的 T 算法的准确率高出 28%左右,即使在最差时也高 14%左右。算法 TF 融合了用户之间的主题相似性与用户之间的社交关系,但并没有优于算法 T。由此可知,在 Foursquare 数据集中,用户的社交关系对用户的兴趣点推荐过程并没有明显的作用,算法 F 的准确率只有算法 T 的 40%左右,这也间接论证了该结论。实验结果表明,本文算法 CBR 融合了用户的地理位置簇信息,具有明显的优势。



(a) Foursquare (NYC)



(b) Yelp

图 1 不同参数下推荐准确率的对比

Fig. 1 Comparison of recommended precision under different parameters

在图 1(b)中,本文算法 CBR 的准确率依然优于其他算法。不同的是,在 Yelp 数据集中,算法 TF 优于算法 T 和算法 F,并且算法 TF 的准确率分别比算法 T 和算法 F 高出 119%和 11.8%。由此可见,用户的主题相似性与社交关系相似性对提高用户的兴趣点推荐的准确率都具有促进作用。

其深层次原因是,在 Yelp 数据集中用户的社交链接密度远高于 Foursquare(NYC)数据集。实验结果表明,CBR 算法较好地融合了用户的兴趣相似性、社交关系与地理位置簇信息,比其他推荐算法更准确。

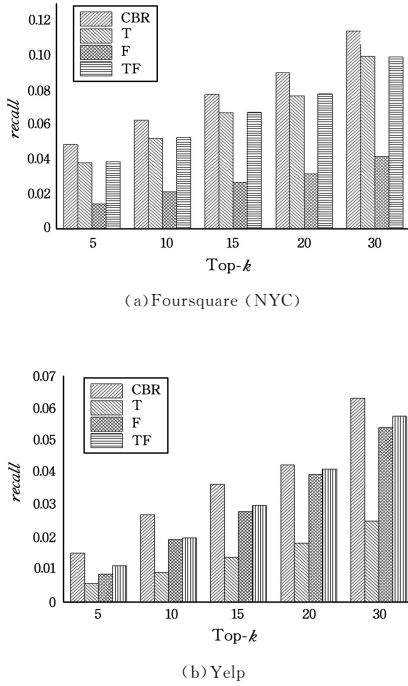


图 2 不同参数下推荐召回率的对比

Fig. 2 Comparison of recommended recall under different parameters

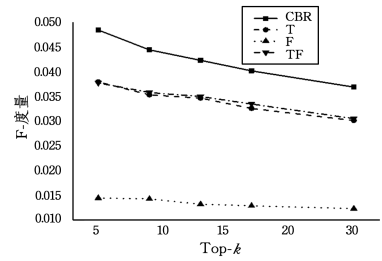
在图 2(a)中,随着参数的增长,各算法的召回率排名不变,本文算法一直优于其他所有算法,其召回率至少高于排名第二的 T 算法 15% 以上,具有明显的优势。

在图 2(b)中,本文算法具有相同的趋势,高于排名第二的算法 TF 10% 以上。值得一提的是,在 Yelp 数据集上,算法 TF 的召回率也高于算法 T 与算法 F。这说明用户社交关系不但提高了兴趣点推荐的准确率,还提高了其召回率。实验结果证明,本文算法 CBR 较好地融合了用户的社交关系,提高了其推荐召回率。

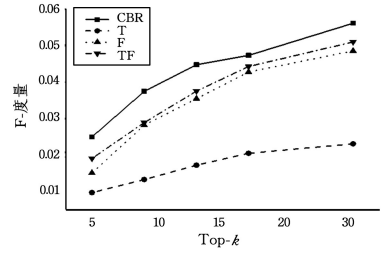
由此可知,随着推荐兴趣点个数的增多,准确率不断下降,召回率不断上升。这是因为当推荐的兴趣点个数增多后,用户将访问的兴趣点和不再访问的兴趣点都会增多,可见推荐算法的准确率和召回率是互相制衡的两个指标。 $F_{\beta}$  作为准确率和召回率的调和平均值,可以比较全面地反映算法的性能随着推荐个数增多的变化规律。

在图 3(a)中,随着兴趣点推荐个数的增加,各算法的 F-度量指标不断下降。CBR 算法和 T 算法的  $F_{\beta}$  分别从 0.048 和 0.038 下降到 0.037 和 0.030,但 CBR 算法的性能仍高于 T 算法 20% 以上。

与 Foursquare 的情况相反,在图 3(b)的 Yelp 数据集中,F-度量指标随着推荐个数的增多而不断上升,CBR 算法和 TF 算法的  $F_{\beta}$  分别从 0.025 和 0.019 上升到 0.056 和 0.051,但 CBR 算法的性能仍高于 TF 算法 10% 左右。



(a) Foursquare (NYC)



(b) Yelp

图 3 不同参数下 F-度量的对比

Fig. 3 Comparison of F-measure under different parameters

由实验结果可得,与其他 3 种算法相比,在不同的数据集与参数下,本文算法的推荐性能都具有明显的优越性,但这只能反映有限 Top-k 值下的推荐性能,无法体现算法的总体性能。MAP 指标解决了这一矛盾,它实质上是在不同 Top-k 下的平均准确率。实验具体结果如表 3 所列。

表 3 MAP 值对比

Table 3 Comparison of MAP value

数据集	CBR	T	F	TF
Foursquare(NYC)	0.045	0.040	0.015	0.041
Yelp	0.040	0.015	0.028	0.032

表 3 中,在 Foursquare(NYC)数据集上,本文算法 CBR 的 MAP 值高于第二位的 TF 算法 9.8%,而在 Yelp 数据集上则高出 TF 算法 25% 左右。可见,本文算法 CBR 的 MAP 值在两个 LBSN 数据集中都明显高于其他 3 种算法。因为 MAP 指标体现的是在不同的 Top-k 参数下,准确率对召回率的积分,所以 CBR 算法在整体情况下都优于其他 3 种算法。

**结束语** 本文针对 LBSN 兴趣点推荐中签到密度稀疏的问题,提出了一种基于社区发现的兴趣点推荐算法。该算法先根据目标用户发布的主题计算用户与兴趣主题簇之间的相似性,进一步得到用户的兴趣主题隶属度,再通过关联矩阵  $R$  获得用户与地理位置簇的关系,最后融合社交关系得到用户对兴趣点的评分,按照评分高低对用户进行推荐。在两种 LBSN 数据集上的实验结果表明,只融合用户之间兴趣主题相似性的算法 Topic-based CF(T) 与只融合了用户之间的社交关系的算法 Friend-based CF(F) 在该信息较为丰富时具有明显的效果,如在 Foursquare(NYC)数据集上的算法 T 和在 Yelp 数据集上的算法 F 具有较高的准确率与召回率;但若数据稀疏,则推荐的准确率与召回率就会有明显的下降,如 Foursquare(NYC)数据集上的算法 F,其主要原因是该数据集的平均社交链接密度远远低于 Yelp 数据集,社交关系稀

疏。而融合了社交关系与兴趣主题相似性的算法 TF 在准确率与召回率上相比单一信息的算法有所提升,但都不如本文提出的基于社区发现的兴趣点推荐算法 CBR,该算法融合了兴趣主题相似性、社交关系相似性和地理位置信息,比其他算法具有更加准确、全面的推荐结果。但本文只考虑了基于静态社区的兴趣点推荐方法,在日后的工作中拟引入时间戳发现动态社区,并观察社交网络的动态演化规律,以此对用户进行实时的兴趣点推荐。

## 参考文献

- [1] YE M, YIN P, LEE W, et al. Exploiting geographical influence for collaborative point-of-interest recommendation[C] // Proceedings of the 34th ACM SIGIR international conference on Research and Development in Information Retrieval. Beijing: ACM, 2011: 325-334.
- [2] ZHANG J, CHOW C, LI Y, LORE: Exploiting sequential influence for location recommendations[C] // Proceedings of the 22nd ACM International Conference on Advances in Geographic Information Systems, SIGSPATIAL. Dallas: ACM, 2014: 103-112.
- [3] SI Y, ZHANG F, LIU W. CTF-ARA: An adaptive method for POI recommendation based on check-in and temporal features [J]. Knowl. -Based Syst, 2017, 25(4): 59-70.
- [4] DONG J X, DONG Y. POI Recommendation Based on Meta-Path in LBSN[J]. Chinese Journal of Computers, 2016, 39(4): 675-684. (in Chinese)  
曹玖新,董羿. LBSN 中基于元路径的兴趣点推荐[J]. 计算机学报, 2016, 39(4): 675-684.
- [5] LIU B, FU Y, YAO Z, et al. Learning geographical preferences for point-of-interest recommendation[C] // Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Chicago: ACM, 2013, 128815: 1043-1051.
- [6] ZHENG Y, ZHANG L, MA Z, et al. Recommending friends and locations based on individual location and history[J]. ACM Transactions on the Web, 2011, 5(11): 5-48.
- [7] REN X Y, SONG M N. Contest-Aware Point-of-Interest Recommendation in Location-Based Social Networks[J]. Chinese Journal of Computers, 2017, 40(4): 824-841. (in Chinese)  
任星怡,宋美娜. 基于位置社交网络的上下文感知的兴趣点推荐[J]. 计算机学报, 2017, 40(4): 824-841.
- [8] LIAN D, ZHAO C, XIE X, et al. GeoMF: Joint geographical modeling and matrix factorization for point of interest recommendation[C] // Proceedings of the 20th ACM International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2014: 831-840.
- [9] ZHANG J, CHOW C. Point-of-interest recommendations in location-based social networks[J]. Sigspatial Spec, 2016, 7(3): 26-33.
- [10] LIU Y, WEI W, SUN A, et al. Exploiting geographical neighborhood characteristics for location recommendation[C] // Proceedings of the 23th ACM International Conference on Information and Knowledge Management. Shanghai: ACM, 2014: 739-748.
- [11] LIU B, XIONG H, PAPADIMITRIOU S. A general geographical probabilistic factor model for point of interest recommendation[J]. IEEE Transactions ON Knowledge and Data Engine, 2015, 27(5): 1167-1179.
- [12] ALIANNEJADI M, CRESTANI F. Personalized Context-Aware Point of Interest Recommendation[J]. ACM Transactions on Information Systems, 2018, 36(4): 45-72.
- [13] YAO Z. Exploiting Human Mobility Patterns for Point-of-Interest Recommendation[C] // Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining. Marina Del Rey: ACM, 2018: 757-758.
- [14] YIN H, ZHOU X, SHAO Y, et al. Joint modeling of user check-in behaviors for point-of-interest recommendation[C] // Proceedings of the 24th ACM International on Conference on Information and Knowledge Management. Melbourne: ACM, 2015: 1631-1640.
- [15] ZHANG J, CHOW C. TICRec: A probabilistic framework to utilize temporal influence correlations for time-aware location recommendations[J]. IEEE Transactions on Services Computing, 2016, 9(4): 633-646.
- [16] WANG H, TERROVITIS M, MAMOULIS N. Location recommendation in location based social networks using user check-in data[C] // Proceedings of ACM Sigspatial International Conference on Advances in Geographic Information Systems. Orlando: ACM, 2013: 374-383.
- [17] YUAN Q, GONG G, SUN A. Graph-based point-of-interest recommendation with geographical and temporal influences[C] // Proceedings of the 23rd ACM International Conference on Information and Knowledge Management. Shanghai: ACM, 2014: 659-668.
- [18] BARAL R, LI T. MAPS: A Multi Aspect Personalized POI Recommender System[C] // Proceedings of the 10th ACM Conference. Boston: ACM, 2016: 281-284.
- [19] YING J, KUO W, TSENG V, et al. Mining user check - in behavior with a random walk for urban point-of-interest recommendations[J]. ACM Transactions on Intelligent Systems and Technology, 2014, 5(3): 40.
- [20] LI W, XIA S, LIU F, et al. Location prediction algorithm based on movement tendency[J]. Journal on Communications, 2014, 35(2): 46-53.
- [21] YE M, YIN P, LEE W, et al. Exploiting geographical influence for collaborative point-of-interest recommendation[C] // Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval. Beijing: ACM, 2011: 325-334.