

基于签到活跃度和时空概率模型的自适应兴趣点推荐方法

司亚利^{①②③} 张付志^{*①③} 刘文远^{①③}

^①(燕山大学信息科学与工程学院 秦皇岛 066004)

^②(燕山大学里仁学院 秦皇岛 066004)

^③(河北省计算机虚拟技术与系统集成重点实验室(燕山大学) 秦皇岛 066004)

摘要: 针对现有兴趣点(POI)推荐算法对不同签到特征的用户缺乏自适应性问题, 该文提出一种基于用户签到活跃度(UCA)特征和时空(TS)概率模型的自适应兴趣点推荐方法UCA-TS。利用概率统计分析方法提取用户签到的活跃度特征, 给出一种用户不活跃和活跃的隶属度计算方法。在此基础上, 分别采用结合时间因素的1维幂律函数和2维高斯核密度估计来计算不活跃和活跃特征的概率值, 同时融入兴趣点流行度来进行推荐。该方法能自适应用户的签到特征, 并能更准确体现用户签到的时间和空间偏好。实验结果表明, 该方法能够有效提高推荐精度和召回率。

关键词: 基于位置社交网络; 兴趣点推荐; 用户活跃度; 隶属度; 高斯核密度估计

中图分类号: TP391

文献标识码: A

文章编号: 1009-5896(2020)03-0678-09

DOI: 10.11999/JEIT190287

An Adaptive Point-Of-Interest Recommendation Method Based on Check-in Activity and Temporal-Spatial Probabilistic Models

SI Yali^{①②③} ZHANG Fuzhi^{①③} LIU Wenyuan^{①③}

^①(School of Information Science and Engineering, Yanshan University, Qinhuangdao 066004, China)

^②(School of Liren, Yanshan University, Qinhuangdao 066004, China)

^③(Key Laboratory for Computer Virtual Technology and System Integration of Hebei Province (Yanshan University), Qinhuangdao 066004, China)

Abstract: Existing Point-Of-Interest (POI) recommendation algorithms lack adaptability for users with different check-in features. To solve this problem, an adaptive POI recommendation method UCA-TS based on User Check-in Activity (UCA) feature and Temporal-Spatial (TS) probabilistic models is proposed. The user check-in activity is extracted using a probabilistic statistical analysis method, and a calculation method of user's inactive and active membership is given. On this basis, one-dimensional power law function and two-dimensional Gaussian kernel density estimation combined with time factor are used to calculate the probability for inactive and active features respectively, and the popularity of POI is incorporated to recommend. This method can adapt to the users' check-in features and reflect the users' check-in temporal-spatial preferences more accurately. The experiments show that the proposed method can effectively improve the recommendation precision and recall.

Key words: Location-based social networks; Point-Of-Interest (POI) recommendation; User activity; Membership; Gaussian kernel density estimation

1 引言

近年来, 移动计算、无线通信和位置获取技术

的进步, 促进了基于位置社交网络(Location-Based Social Networks, LBSNs)的普及和发展。在社交网络应用平台中, 如Foursquare, Gowalla, Facebook, Twitter等, 用户不仅能够与其他用户建立社交关系、发布感兴趣的内容, 还可以通过签到的方式分享其在物理世界中的位置、访问时间、评论等信息。那么利用LBSNs中的多维信息从海量地点中为用户个性化地、准确地推荐其有可能感兴趣且从未访问过的地点^[1], 称为兴趣点(Point-Of-Interest,

收稿日期: 2019-04-25; 改回日期: 2019-10-29; 网络出版: 2019-11-11

*通信作者: 张付志 xjzfz@ysu.edu.cn

基金项目: 国家自然科学基金(61379116, 61772452), 河北省自然科学基金(F2015203046, F2015501105)

Foundation Items: The National Natural Science Foundation of China (61379116, 61772452), The Natural Science Foundation of Hebei Province (F2015203046, F2015501105)

POI)推荐。POI推荐能有效解决位置信息过载给用户带来的选择困扰问题,有助于提升用户在社交网络和现实生活中的体验,并能帮助商家分析和挖掘潜在的用户实现广告推送服务,为此兴趣点推荐成为了当前研究的热点问题^[2]。

目前,国内外学者对兴趣点推荐展开了深入研究,提出结合地理因素^[3,4]、时间信息^[5,6]、社交影响^[7]或内容语义^[8]的兴趣点推荐方法,通过加入1个或多个影响因素^[9,10]来提高推荐效果,并且采用基于用户的协同过滤^[11]、概率模型^[12]、矩阵分解模型^[13]、链接关系图^[14]和深度学习^[15]等方法进行建模和推荐。Zhang等人^[3,4]提出融合社交、类别、地理、序列和时间的联合位置推荐框架,采用2维核密度估计建模地理信息,使用签到序列中的 n 阶马尔可夫链和快速高斯传递进行概率推荐。Yuan等人^[5]提出一种基于用户协同过滤和时间感知的POI推荐方法,采用1维幂律函数模型和时间槽平滑技术来提高推荐性能。他们还融合用户签到、连续时间槽和地理信息到感知图模型中,利用广度优先搜索策略选择最佳路径作为推荐结果^[14]。Zhang等人^[6]提出一种利用时间相关性的概率推荐框架,划分用户在工作日和周末的签到活动,并使用核密度方法估计时间的概率密度。Zhao等人^[9]根据月份、星期和小时设计1个时间索引方案,在时间、用户和兴趣点交互的细粒度模型基础上,提出一种基于排序的对偶相互作用张量因式分解框架来进行兴趣点推荐。

兴趣点推荐已经有很多研究成果,但是推荐的精度和召回率仍然无法满足用户的要求,主要原因有两方面:一方面,现有POI推荐方法较少分析用户的签到行为特征。一部分比较活跃的LBSNs用户热衷于频繁签到和分享经验,而还有一部分活跃度较低的用户只有少量签到数据,因此对用户的不同签到行为需要进行深入的特征分析。另一方面,较少考虑为不同签到特征的用户设计适应性的推荐算法,导致推荐质量较差。如何将用户特征与其它影响因素相结合来设计自适应算法,是POI推荐研究的一个亟待解决问题。

本文提出一种基于用户签到活跃度特征和时间、空间概率模型的自适应POI推荐方法UCATS。首先利用概率统计提取并分析用户签到的活跃度特性,给出一种用户不活跃和活跃的隶属度计算方法。在此基础上,结合兴趣点流行度因素,对用户不同的活跃程度采用不同的时空模型方法,以实现兴趣点推荐的自适应性。最后在LBSNs数据集上,对提出算法和其它典型算法进行实验和性能分析。

2 基于签到活跃度特征的用户隶属度计算方法

2.1 问题描述

令 $U=\{u_1, u_2, \dots, u_m\}$ 表示所有用户的集合, $L=\{l_1, l_2, \dots, l_n\}$ 表示兴趣点的集合。 $UC_{all}=\{uc_1, uc_2, \dots, uc_N\}$ 是所有用户的签到记录集合,每条签到记录是一个3元组,即 $uc_k=(u_i, l_j, t)$,表示用户 u_i 在时间 t 签到了兴趣点 l_j 。把时间依据小时划分为24个离散时间槽 $T=\{0, 1, \dots, 23\}$,其中一个时间槽表示为 $t \in T$;用户 u 访问过的兴趣点集合表示为 $L_u \subseteq L$ 。因此,兴趣点推荐是在给定用户 $u \in U$ 和时间槽 t 的条件下,依据用户历史签到偏好以及签到的时空特征,从候选兴趣点集合 $L-L_u$ 中选出 $top-n$ 个最匹配用户偏好的兴趣点推荐给用户。

2.2 用户签到的活跃度特征提取和分析

针对用户签到的差异性,提取并分析签到记录数量和签到时间槽个数两个维度的用户签到活跃度特征,以此作为用户隶属度计算的基础。对于LBSN中的一个用户 $u \in U$,其签到数量 UCN_u 表示为用户 u 在LBSN数据集中的签到记录总数,其中函数 $g(u, uc_i)$ 用来判断是否为用户 u 的签到记录

$$UCN_u = \sum_{i=1}^{|UC_{all}|} g(u, uc_i) \quad (1)$$

$$g(u, uc_i) = \begin{cases} 1, & u \in uc_i \\ 0, & u \notin uc_i \end{cases} \quad (2)$$

在用户 $u \in U$ 的所有签到记录中,不同的签到时间槽数量 T_u 如式(3)所示,其中函数 $f(x)$ 是指示函数,表示用户在某个时间槽内是否有签到记录

$$T_u = \sum_{i=0}^{23} f(|L_{u,t_i}|) \quad (3)$$

$$f(|L_{u,t_i}|) = \begin{cases} 1, & |L_{u,t_i}| > 0 \\ 0, & \text{其它} \end{cases} \quad (4)$$

为了验证LBSNs用户签到活跃度的差异性,对Foursquare和Gowalla用户从签到记录数和签到时间槽个数两个维度进行分析。签到记录越多表明用户越活跃,而签到时间槽个数则反映了用户签到的广度,签到时间槽越多表明用户兴趣越广泛、越活跃。图1给出Gowalla和Foursquare用户签到次数的统计结果,可以看出用户的签到次数差别比较明显,有些用户只有几条签到记录,有的用户签到记录超过1000条。而且随机签到变量 UCN 服从重尾分布,说明有少数用户具有较多的签到记录,而大部用户签到较少。例如Gowalla共有10162个用户,仅有64个用户签到记录超过500条。由此说明

LBSNs中有一部分用户具有较多的签到记录, 其行为呈现活跃的状态。图2给出用户签到数量的概率质量函数(PMF)统计结果, 可以看出用户签到次数为10, 20和40时, Foursquare的PMF值分别为11.29%, 14.95%和19.60%, 有45.84%的Foursquare用户签到小于40; Gowalla的PMF值分别为26.94%, 26.39%和20.01%, 有73.40%的Gowalla用户签到次数小于40。由此说明, 有一部分用户具有较少的签到记录, 其行为呈现不活跃的状态。综合以上分析结果, 表明LBSNs用户的签到行为具有较大差异性, 有一部分用户比较活跃, 而还有一部分用户处于非活跃状态。

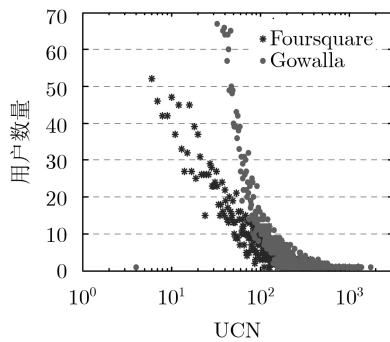


图1 两个数据集中UCN和相应用户数量统计图

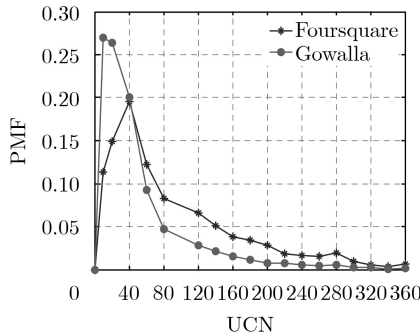


图2 用户签到数量的概率质量函数图

2.3 用户隶属度计算方法

本文采用模糊理论计算用户属于不同类别的相似程度, 即用户隶属度。令用户集合为 $U=\{u_1, u_2, \dots, u_{|U|}\}$, 每个用户 $u_j (j=1, 2, \dots, |U|)$ 具有签到数量和时间槽数量的2维签到特征: $u_j=(\text{userid}_j, \text{UCN}_j, T_j)$ 。把这些用户划分成活跃用户和不活跃两类, 两个类中心为 $C=\{c_1, c_2\}$, 每个样本 u_j 属于某一类 $c_i (i=1, 2)$ 的隶属度为 a_{ij} , 则目标函数定义为相应样本的隶属度与该样本到各个类中心的距离相乘, 如式(5)所示, 约束条件如式(6)所示

$$J = \sum_{i=1,2} \sum_{j=1}^{|U|} a_{ij}^m \|u_j - c_i\|^2 \quad (5)$$

$$\sum_{i=1,2} a_{ij} = 1, j = 1, 2, \dots, |U| \quad (6)$$

其中, m 是一个隶属度的因子, $m>1$; $\|\cdot\|$ 是标准欧式距离; 隶属度 $A=\{a_{ij}, j=1, 2, \dots, |U|\}$ 的计算如式(7)所示, 类中心的迭代如式(8)所示。经过反复迭代, 直到目标函数 J 收敛, 此时得到每个用户属于不活跃用户类和活跃用户类的最终隶属度值 $a_u=\{a_{1u}, a_{2u}\}$ 。

$$a_{ij} = \frac{1}{\sum_{k=1}^2 \left(\frac{\|u_j - c_i\|}{\|u_j - c_k\|} \right)^{\frac{2}{m-1}}} \quad (7)$$

$$c_i(k+1) = \frac{\sum_{j=1}^{|U|} ((a_{ij}(k))^m u_j)}{\sum_{j=1}^{|U|} (a_{ij}(k))^m} \quad (8)$$

3 基于用户签到活跃度和时空概率模型的自适应兴趣点推荐算法

本文把时间因素加入到1维幂律函数和2维高斯核密度估计模型中, 能够更准确反映用户在时间和空间上的签到关系和偏好。根据用户签到呈现出的不活跃和活跃特征, 给出一种采用时空概率模型的自适应兴趣点推荐算法。自适应性体现在算法能够分别考虑用户的不活跃和活跃程度, 相应地采用1维和2维时空模型进行概率推荐, 以提高推荐性能, 具体原因如下:

(1) 1维幂律函数是一种利用地点距离的参数估计方法, 需要利用历史签到数据作为先验知识, 并且假设用户签到的地点是相互独立的。把不活跃用户的所有签到地点作为参考点, 能全面体现签到行为特征和签到地点之间的距离关系。但是对签到记录较多的活跃用户, 使用1维幂律函数会产生难以选择参考点和计算量增大的问题。因此, 对签到不活跃的用户适合采用1维幂律函数模型。

(2) 标度不变性是幂律函数的一个重要特征, 含义是在复杂网络上任选一局部, 由于其自相似性, 局部网络的形态、规律、功能均与原网络不会发生变化。不活跃用户在其访问的所有地点中, 只进行少量的签到行为。因此, 对用户签到过的地点采用幂律函数进行概率分析, 能够代表用户整体访问地点的特征关系。

(3) 2维高斯核密度估计是一种利用空间地理经纬度的非参数估计方法, 无需先验知识和任何假定, 能够根据用户签到数据自身蕴含的特点和性质, 来拟合签到地点的分布特征。在大样本的用户

签到地点数据的情况下,核函数比较稳定,对推荐结果几乎没有影响。因此,2维高斯核密度估计模型适用于活跃用户。

3.1 基于时间因素的1维幂律函数模型

针对用户的不活跃属性,本文采用用户去过的历史地点和候选兴趣点之间距离的1维幂律函数,同时结合时间因素来计算推荐概率值。令 $l \in L - L_u$ 是用户 $u \in U$ 的候选兴趣点, $L_{u,t}$ 是用户 u 在时间槽 $t \in T$ 签到过的兴趣点集合,根据贝叶斯条件概率规则,用户 u 在时间槽 t 签到候选兴趣点 l 的概率 $p_1(l|L_{u,t})$ 为

$$\begin{aligned} p_1(l|L_{u,t}) &= \frac{p_1(l \cup L_{u,t})}{p_1(L_{u,t})} \\ &= \frac{p_1(L_{u,t}) \times \prod_{l_i \in L_{u,t}} p(l|l_i)}{p_1(L_{u,t})} \\ &= \prod_{l_i \in L_{u,t}} p(l|l_i) \end{aligned} \quad (9)$$

其中,把用户之前签到的兴趣点 $l_i \in L_{u,t}$ 当作参考点, $p(l|l_i)$ 表示基于参考点位置 l_i 的情况下,在候选兴趣点 l 的签到概率值,如式(10)所示。 $p(d_{li,l})$ 利用了1维幂律函数, α 和 β 两个参数的值由最大似然估计得到。 $d_{li,l}$ 是兴趣点 l_i 和 l 之间的距离, $l_i = \{\text{Lon}_i, \text{Lat}_i\}$ 表示兴趣点 l_i 的经度坐标和纬度坐标,地球半径 $R=6371$ km。

$$p(l|l_i) = \frac{p(d_{li,l})}{\sum_{l_j \in L, i \neq j} p(d_{li,l_j})} \quad (10)$$

$$p(d_{li,l}) = \alpha \times (d_{li,l})^\beta \quad (11)$$

$$d_{l_1,l_2} = R \cdot \arccos[\sin \text{Lat}_1 \cdot \sin \text{Lat}_2 + \cos \text{Lat}_1 \cdot \cos \text{Lat}_2 \cdot \cos(\text{Lon}_2 - \text{Lon}_1)] \quad (12)$$

3.2 基于时间因素的2维高斯核密度估计模型

针对用户的活跃属性,本文采用基于时间因素的2维高斯核密度估计方法来计算候选兴趣点的推荐概率值。令 $u \in U$ 是LBSN中的一个用户, $L_{u,t}$ 是该用户在时间槽 $t \in T$ 签到过的兴趣点集合, $|L_{u,t}|$ 是该用户在时间槽 t 签到地点的总数,兴趣点 l 的经度和纬度坐标表示为 $\mathbf{l}=(\text{Lat}, \text{Lon})^T$,那么对于用户 u 没有访问过的候选兴趣点 $l \in L - L_u$,采用基于时间的2维高斯核密度估计模型方法,计算的推荐概率值 $p_2(l|L_{u,t})$ 如式(13)所示

$$p_2(l|L_{u,t}) = \frac{1}{|L_{u,t}|H^2} \sum_{i=1}^{|L_{u,t}|} K\left(\frac{\mathbf{l} - \mathbf{l}_i}{H}\right) \quad (13)$$

其中, $K(\cdot)$ 是标准高斯核函数,如式(14)所示。式(15)计算用户 u 所有已经签到位置 L_u 的经纬度坐标的均

值 $\hat{\mathbf{u}}$,式(16)计算用户 u 历史签到位置经纬度坐标的方差 $\hat{\mathbf{h}}$,式(17)计算高斯核函数的最优带宽 H 。

$$K(\mathbf{X}) = \frac{1}{2\pi} \exp\left(-\frac{1}{2} \mathbf{X}^T \mathbf{X}\right) \quad (14)$$

$$\hat{\mathbf{u}} = \frac{1}{|L_u|} \sum_{i=1}^{|L_u|} \mathbf{l}_i \quad (15)$$

$$\hat{\mathbf{h}} = \frac{1}{|L_u|} \sum_{i=1}^{|L_u|} (\mathbf{l}_i - \hat{\mathbf{u}})^2 \quad (16)$$

$$H = \frac{1}{2} |L_u|^{-\frac{2}{3}} \sqrt{\hat{\mathbf{h}}^T \hat{\mathbf{h}}} \quad (17)$$

3.3 兴趣点流行度的计算方法

本文采用兴趣点的整体流行度和时间感知的流行度两部分因素,来计算兴趣点的流行度。对于一个兴趣点 $l \in L$,其整体的流行度可以表示为所有用户访问该兴趣点 l 的签到记录数与总签到记录数的比值

$$\text{OP}_l = \frac{\sum_{u \in U} n_{u,l}}{\sum_{u' \in U} \sum_{l' \in L} n_{u',l'}} \quad (18)$$

在一个时间槽 $t \in T$ 的情况下,兴趣点 $l \in L$ 的时间感知流行度可以表示为所有用户在 t 时刻访问 l 的签到记录数与所有时刻访问 l 的总签到记录数的比值

$$\text{TSP}_{t,l} = \frac{\sum_{u \in U} n_{u,t,l}}{\sum_{u' \in U} \sum_{t' \in T} n_{u',t',l}} \quad (19)$$

如果只使用一个目标时刻的信息,那么按小时划分签到数据将会导致数据更加稀疏。因此,本文采用连续时间槽的平滑技术来计算时间感知的流行度,再结合整体流行度,给出兴趣点 l 在 t 时间槽下的流行度

$$p_t(l) = \sum_{t'=t-k}^{t+k} \text{TSP}_{t',l} + \text{OP}_l \quad (20)$$

其中, t' 表示把 t 时刻同时向前和向后扩 k 个时间槽, $t'=[t-k, t+k]$ 。如果 $t'>23$,则 $t'=t'-24$;如果 $t'<0$,则 $t'=t'+24$ 。

3.4 自适应兴趣点推荐算法描述

结合用户签到活跃度(User Check-in Activity, UCA)特征和时空(Temporal-Spatial, TS)特征,构建一种用户活跃度和时空概率模型的自适应兴趣点推荐算法UCA-TS,以提高推荐性能。UCA-TS的实现过程如下:

步骤1 计算用户签到活跃度特征的签到数量 UCN_u 和签到时间槽数量 T_u ，并利用用户隶属度计算方法，把用户的不同签到特征量化为属于不活跃类和活跃类的隶属度 $a_u = \{a_{1u}, a_{2u}\}$ 。

步骤2 计算各个时间槽相关度 $w_{\Delta t}$ ，如式(21)所示。用户 u 首次签到兴趣点 l 的时间槽是 t ，与之后再签到该兴趣点的时间槽 t' 之间的时间差为 $\Delta t = t' - t$ ， $n_{u, \Delta t}$ 表示用户 u 签到相同兴趣点的时间差为 Δt 的个数。

$$w_{\Delta t} = \frac{\sum_{u \in U} n_{u, \Delta t}}{\sum_{\Delta t' = -11}^{12} \left(\sum_{u \in U} n_{u, \Delta t'} \right)} \quad (21)$$

步骤3 依据隶属度值，相应地对用户的不活跃程度计算各个时间槽相关度权重 $w_{\Delta t}$ 的1维幂律函数值 p_1 ，对用户的活跃程度采用各个时间槽相关度的2维高斯核密度估计值 p_2 ，同时结合兴趣点流行度 $p_t(l)$ ，计算用户 u 在 t 时刻访问兴趣点 l 的概率 $P_{u,t,l}$ ，进行top- n 推荐，如式(22)所示

$$p_{u,t,l} = a_{1u} \cdot \sum_{t'=0}^{23} w_{t'-t} \cdot p_1(l|L_{u,t'}) + a_{2u} \cdot \sum_{t'=0}^{23} w_{t'-t} \cdot p_2(l|L_{u,t'}) + p_t(l) \quad (22)$$

基于以上分析，给出UCA-TS算法的描述，如表1所示。

表1 自适应兴趣点推荐算法(UCA-TS)

输入：签到数据集 UC_{all} ，推荐的目标用户 u ，时间槽 t ；	(11) for each $l \in L - L_u$ do
输出：推荐的top- n 兴趣点	(12) $p_1(l L_{u,t}) \leftarrow 1$; $p_2(l L_{u,t}) \leftarrow 0$;
(1) 使用式(1)计算 UCN_u ，式(3)计算 T_u ；	(13) for each $l_i \in L_{u,t}$ do
(2) $k \leftarrow 0$ ；初始化 $C = \{c_1, c_2\}$ 和 $A = \{a_{ij}\}$ ；	(14) 计算 $d_{u,t}$ 和 $p(l_i)$ ；
(3) repeat	(15) 计算 $p_1(l L_{u,t}) \leftarrow p_1(l L_{u,t}) \times p(l_i)$ ；
(4) $k \leftarrow k+1$ ；	(16) 计算 $p_2(l L_{u,t})$ ；
(5) 更新聚类中心 $C(k)$ 和 $A(k-1)$ ；	(17) end for
(6) 更新隶属度矩阵 $A(k)$ 和聚类中心 $C(k)$ ；	(18) 使用式(20)计算 $p_t(l)$ ；
(7) until 式(5)收敛	(19) end for
(8) 返回 $a_u = \{a_{1u}, a_{2u}\}$ ；	(20) end for
(9) for $t'=0$ to 23 do	(21) 使用式(22)计算 $P_{u,t,l}$ ；
(10) 使用式(21)计算 w_{t-t} ；使用式(15)–式(17)计算 H ；	(22) 排序 $P_{u,t,l}$ 并返回top- n 兴趣点。

4 实验及分析

为了验证所提算法的有效性，本节在给出数据集和评估指标后进行实验和分析。针对算法的自适应性，在实验部分体现为：对Foursquare和Gowalla实验数据进行不活跃用户和活跃用户的多项数据统计，并且依据LBSNs签到实验数据，计算出每个用户属于活跃和不活跃的隶属度值；采用实现自适应的1维幂律函数和2维高斯核密度估计两种方法，对不活跃用户和活跃用户进行实验和结果对比分析，用以说明自适应方法的有效性；与其它典型算法进行实验和性能结果对比，分析本文自适应算法的有效性原因。

4.1 数据集

实验采用的LBSNs签到数据集是Foursquare和Gowalla，统计情况如表2所示。从表2可以看出，数据集包含大量的用户和兴趣点，但存在数据极为稀疏的问题，这是导致推荐算法性能较低的主要原因。对每个用户随机选取16%的签到记录作为测试集，其它的数据作为训练集，并且每个用户在测试集中的兴趣点，必须是训练集中没有签到的地点。体现了兴趣点推荐的本质特点，即为用户推荐以前没有去过的地方。

4.2 评估指标

为了评估兴趣点推荐方法的性能，本文使用精度、召回率和 F_β 指标3个评估指标，计算公式为

表2 LBSNs签到数据集的统计情况

数据集	签到数	兴趣点数	用户数	每个用户访问地点平均数	每个地点的平均访问数	每个用户平均签到数	签到密度
Foursquare	194108	5596	2321	46	19	84	0.0149
Gowalla	456905	24236	10162	30	13	45	0.0019

$$\text{pre}@n = \frac{\sum_{t=0}^{23} \sum_{u \in U} N_{u,t}}{N_{\text{rec}}} \quad (23)$$

$$\text{rec}@n = \frac{\sum_{t=0}^{23} \sum_{u \in U} N_{u,t}}{N_{\text{test}}} \quad (24)$$

$$F_\beta = (1 + \beta^2) \cdot \frac{\text{pre} \times \text{rec}}{\beta^2 \cdot \text{pre} + \text{rec}} \quad (25)$$

其中, N_{rec} 表示推荐的所有top- n 兴趣点数量, $N_{u,t}$ 表示用户 u 在 t 时刻, 同时在测试集和推荐结果集的兴趣点数量, N_{test} 表示测试集中的所有兴趣点数量。

4.3 实验结果和分析

为了评估兴趣点推荐算法的性能, 选取以下4个基于时空因素的兴趣点推荐算法和本文所提UCA-TS算法进行对比。

(1) SK算法^[4]是基于核密度估计的POI推荐算法, 融合社交、类别和时空因素。

(2) UTE+SE^[5]是基于1维幂律函数、流行度、时间因素和用户协同过滤的POI推荐算法。

(3) TPR+UM^[16]是利用时间和地理区域因素的用户移动性概率模型的POI推荐算法。

(4) SAMM^[17]是基于时空因素和多阶马尔可夫模型的自适应POI推荐算法。

4.3.1 时间槽参数 k 对推荐结果的影响

兴趣点流行度的式(20)中, 参数 k 表示把目标时间 t 向前和向后增加 k 个连续时间槽, 因此使用 $2k+1$ 个连续时间槽的平滑技术来计算流行度。使用合理的连续时间槽能够降低其它时间槽产生的噪音影响, 更好体现用户在时间上的签到特征, 从而提高推荐性能。为了选择最优的时间槽长度, 对UCA-TS算法的不同 k 值在Foursquare和Gowalla两个数据集上进行实验, top-5精度和top-20召回率的结果如图3所示。

从图3可以看出, 随着 k 值的增加, 即时间槽的扩增, 精度和召回率首先呈上升的趋势, 在 $k=2$ 时达到了最高的精度和召回率, 随后逐渐降低。具体来说, UCA-TS算法使用连续5个时间槽($k=2$)与仅仅使用1个目标时间槽($k=0$)相比, 在Foursquare和Gowalla上的精度分别提高了5.17%和6.56%, 召回率分别提高了4.57%和4.74%。基于以上实验结果和分析, 本文设置参数 $k=2$, 也就是以目标时刻 t 为中心的连续5个时间槽, 来计算兴趣点的流行度。

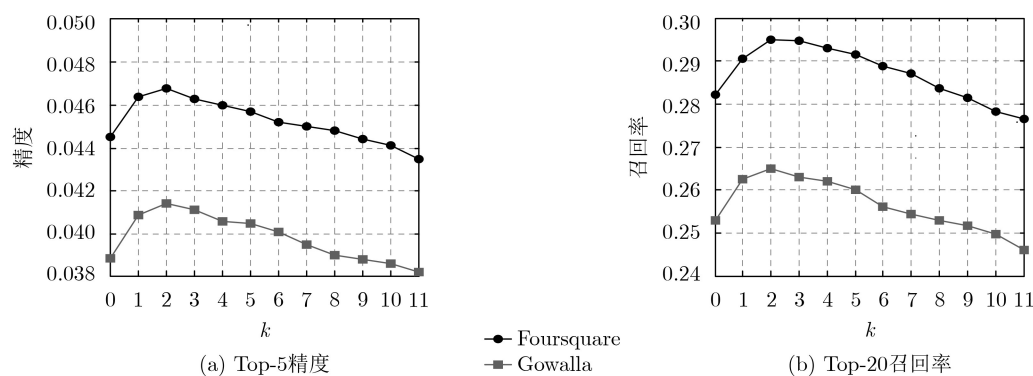


图3 k 对精度和召回率的影响

4.3.2 1维和2维空间概率模型的实验对比分析

为了对比空间的1维幂律函数概率模型方法(1D)和2维高斯核密度估计概率模型方法(2D)的推荐效果, 本文选取相对比较活跃或不活跃的用户, 分别进行推荐实验. 在Foursquare数据集中, 根据所有用户的隶属度矩阵 $A=\{a_{ij}, i=1,2, j=1,2, \dots, |U|\}$, 取出 $a_{1j} \geq 0.8$ 的用户生成活跃用户集合, $a_{2j} \geq 0.8$ 的用户生成不活跃用户集合。图4给出了Foursquare数据集上1维和2维两个模型对活跃用户和不活跃用户的top- n ($n=5, 10, 20$)精度和召回率。从图4可以看出, 对于不活跃用户, 使用1维幂率函数模型的top- n 推荐精度和召回率高于2维高斯

核密度估计模型相应的top- n 精度和召回率; 相反, 对于活跃用户, 使用2维高斯核密度估计模型的推荐效果具有明显的优势。主要的原因是: 1维方法是基于兴趣点之间的距离, 体现了兴趣点之间的地理直线距离, 更适用于签到记录较少的不活跃用户。2维方法是基于兴趣点经度和纬度的空间关系, 体现了兴趣点的空间分布特征, 更适用于签到记录大量且丰富的活跃用户。实验结果表明了针对用户的活跃度的两种方法的有效性。

为了更直观体现不活跃用户和活跃用户的差异性, 根据用户的隶属度值 $a_u=\{a_{1u}, a_{2u}\}$, 设定用户属于较大隶属度的一类。对Foursquare和Gowalla

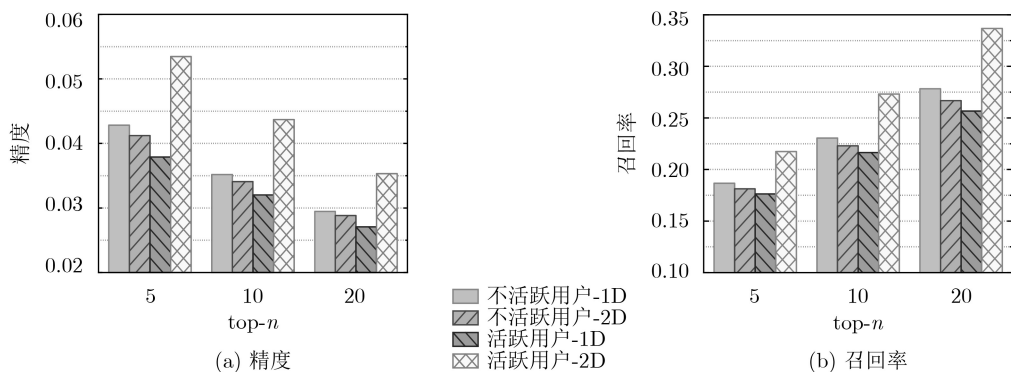


图4 1维和2维模型对Foursquare数据集中不活跃用户和活跃用户的精度和召回率

数据集的用户划分后, 分别进行不活跃用户和活跃用户的多项数据统计, 如表3所示。从表3可以看出, 不活跃用户和活跃用户在平均签到记录个数、

平均签到时间槽数量和平均签到地点数量等方面, 具有明显的差异性, 不活跃用户的统计结果均远远低于活跃用户的结果。

表3 Foursquare和Gowalla数据集的不活跃和活跃用户数据统计

数据集	用户类别	用户数量	签到记录总数	平均签到记录数	平均签到时间槽数	平均签到地点数
Foursquare	不活跃用户	1908	88181	46	12	28
	活跃用户	413	105927	256	19	114
Gowalla	不活跃用户	9756	304621	31	9	21
	活跃用户	406	152284	375	18	204

4.3.3 算法的推荐结果与分析

在Foursquare和Gowalla两个数据集上, 针对训练集和测试集中的所有数据, 采用5个兴趣点推荐算法分别进行实验。图5分别给出了Foursquare和Gowalla数据集上5种算法的top- n ($n=5, 10, 20$)精度和召回率结果。如图5所示, 本文提出的UCA-TS算法的精度和召回率在相同的条件下均优于其它4个算法。在Foursquare数据集上, UCA-TS算法与SK, UTE+SE, TPR+UM和SAMM算法相比, top-5精度分别提高了76.87%, 68.06%, 29.14%和12.50%; top-20召回率分别提高了68.57%, 62.32%, 27.93%和14.34%。在Gowalla数据集上, UCA-TS算法与SK, UTE+SE, TPR+UM和SAMM

算法相比, top-5精度分别提高了58.86%, 49.50%, 39.86%和10.40%; top-20召回率分别提高了60.61%, 53.48%, 21.34%和7.72%。

表4给出了5个兴趣点推荐算法在两个数据集上基于精度和召回率的 F_β 指标($\beta=1$)结果。从表4可以看出, UCA-TS算法的优于其它算法。具体地, F_β 指标在top-5, top-10和top-20情况下的结果平均值, UCA-TS算法比SK, UTE+SE, TPR+UM和SAMM算法在Foursquare数据集上分别提高了98.50%, 89.61%, 41.06%和11.69%; 在Gowalla数据集上分别提高了84.21%, 74.64%, 50.04%和9.98%。

以上结果表明, UCA-TS算法在两种数据集上

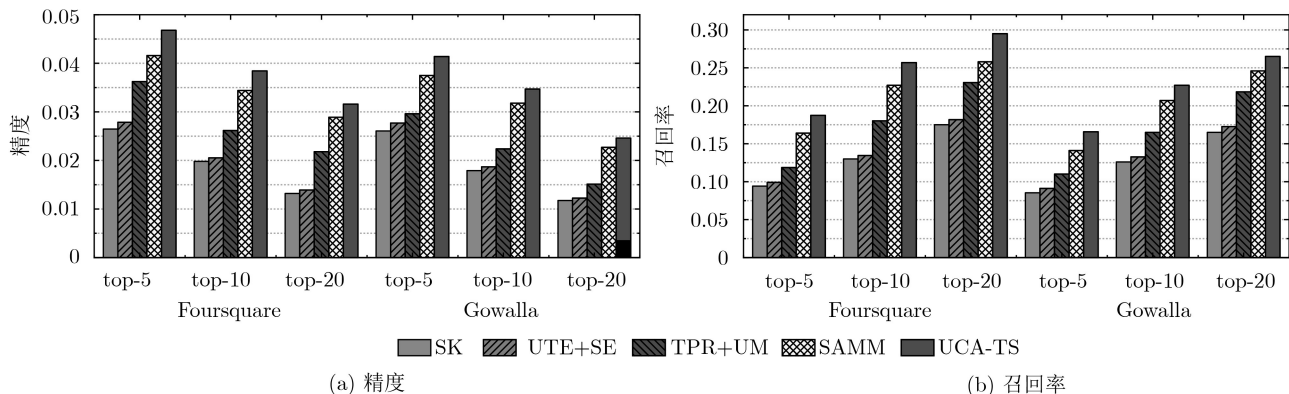


图5 兴趣点推荐算法在两个数据集上的精度和召回率

表4 兴趣点推荐算法在两个数据集上的 F_β 指标值($\beta=1$)

数据集	Top- n	SK	UTE+SE	TPR+UM	SAMM	UCA-TS
Foursquare	top-5	0.0413	0.0435	0.0555	0.0664	0.0749
	top-10	0.0344	0.0356	0.0457	0.0597	0.0669
	top-20	0.0245	0.0258	0.0398	0.0520	0.0571
Gowalla	top-5	0.0399	0.0425	0.0466	0.0592	0.0663
	top-10	0.0313	0.0328	0.0394	0.0551	0.0602
	top-20	0.0219	0.0229	0.0283	0.0416	0.0450

的推荐性能均优于其它对比算法的性能。分析原因,是因为:(1)研究并利用用户自身的签到活跃度特征,能够更深入、更准确地捕捉用户的签到行为偏好,从而进行有针对性的个性化推荐;(2)在空间的1维幂律和2维高斯核密度估计概率模型中,融入了时间因素,有效增强了兴趣点推荐的时空关联;(3)UCA-TS算法能够根据用户的签到活跃度特征,通过恰当采用不同的时空模型策略进行适应性的兴趣点推荐,充分利用了用户签到属性和时空模型的相关性,以达到最佳推荐效果;(4)采用了整体和时间感知相结合的兴趣点流行度计算方法,并采用时间槽平滑技术解决数据稀疏问题。综上,将用户签到活跃度特征和时空模型相结合的自适应兴趣点推荐算法能够获得更好的推荐性能。

从图5中还能看出,相同情况下5个算法在Foursquare数据集上的实验结果均高于Gowalla数据集上的实验结果。原因是Gowalla比Foursquare更加稀疏,这使得用户签到行为和兴趣点变得复杂和多样性,于是同样的算法在稀疏的数据信息中就更难进行准确的推荐。

5 结束语

本文提出了一种基于用户签到活跃度特征和时空概率模型的自适应兴趣点推荐方法UCA-TS。通过分析LBSNs签到数据,挖掘出用户的活跃度特征,并提出了用户隶属度计算方法来对不活跃和活跃程度进行量化。根据用户不同的活跃程度,采用基于时间因素的1维幂律函数方法和2维高斯核密度估计方法,进行自适应的兴趣点推荐,体现了用户的签到特征和兴趣点的时空关系。此外,提取并分析了兴趣点的流行度,包括整体流行度和基于时间的流行度两个部分,更合理地利用时间槽的相关性和空间的关系。实验结果表明,本文提出的自适应兴趣点推荐方法在精度、召回率和 F_β 指标方面均优于基线POI推荐方法。下一步的工作是挖掘兴趣点推荐的其他影响因素,加入到推荐算法中,以提高推荐性能。

参考文献

- [1] SI Yali, ZHANG Fuzhi, and LIU Wenyuan. CTF-ARA: An adaptive method for POI recommendation based on check-in and temporal features[J]. *Knowledge-Based Systems*, 2017, 128: 59–70. doi: [10.1016/j.knsys.2017.04.013](https://doi.org/10.1016/j.knsys.2017.04.013).
- [2] YU Yonghong and CHEN Xingguo. A survey of point-of-interest recommendation in location-based social networks[C]. *Trajectory-Based Behavior Analytics: Papers from the 2015 AAAI Workshop, Palo Alto, USA, 2015*: 53–60.
- [3] ZHANG Jiadong and CHOW C Y. CoRe: Exploiting the personalized influence of two-dimensional geographic coordinates for location recommendations[J]. *Information Sciences*, 2015, 293: 163–181. doi: [10.1016/j.ins.2014.09.014](https://doi.org/10.1016/j.ins.2014.09.014).
- [4] ZHANG Jiadong and CHOW C Y. Point-of-interest recommendations in location-based social networks[J]. *SIGSPATIAL Special*, 2015, 7(3): 26–33. doi: [10.1145/2876480.2876486](https://doi.org/10.1145/2876480.2876486).
- [5] YUAN Quan, CONG Gao, MA Zongyang, et al. Time-aware point-of-interest recommendation[C]. *The 36th International ACM SIGIR Conference on Research and Development in Information Retrieval, Dublin, Ireland, 2013*: 363–372. doi: [10.1145/2484028.2484030](https://doi.org/10.1145/2484028.2484030).
- [6] ZHANG Jiadong and CHOW C Y. 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. doi: [10.1109/TSC.2015.2413783](https://doi.org/10.1109/TSC.2015.2413783).
- [7] LI Huayu, GE Yong, HONG Richang, et al. Point-of-Interest recommendations: Learning potential check-ins from Friends[C]. *The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, USA, 2016*: 975–984. doi: [10.1145/2939672.2939767](https://doi.org/10.1145/2939672.2939767).
- [8] LIU Bin and XIONG Hui. Point-of-interest recommendation in location based social networks with topic and location awareness[C]. *The 2013 SIAM International Conference on Data Mining, Austin, USA, 2013*: 396–404. doi: [10.1137/1.9781611972832.44](https://doi.org/10.1137/1.9781611972832.44).

- [9] ZHAO Shenglin, ZHAO Tong, YANG Haiqin, *et al.* STELLAR: Spatial-temporal latent ranking for successive point-of-interest recommendation[C]. The 30th AAAI Conference on Artificial Intelligence, Phoenix, USA, 2016: 315–321. doi: [10.1007/978-981-13-1349-3_5](https://doi.org/10.1007/978-981-13-1349-3_5).
- [10] 任星怡, 宋美娜, 宋俊德. 基于位置社交网络的上下文感知的兴趣点推荐[J]. 计算机学报, 2017, 40(4): 824–841. doi: [10.11897/SP.J.1016.2017.00824](https://doi.org/10.11897/SP.J.1016.2017.00824).
REN Xingyi, SONG Meina, and SONG Junde. Context-aware point-of-interest recommendation in location-based social networks[J]. *Chinese Journal of Computers*, 2017, 40(4): 824–841. doi: [10.11897/SP.J.1016.2017.00824](https://doi.org/10.11897/SP.J.1016.2017.00824).
- [11] YIN Hongzhi, CUI Bin, CHEN Ling, *et al.* Modeling location-based user rating profiles for personalized recommendation[J]. *ACM Transactions on Knowledge Discovery from Data*, 2015, 9(3): 1–41. doi: [10.1145/2663356](https://doi.org/10.1145/2663356).
- [12] LIU Bin, XIONG Hui, PAPADIMITRIOU S, *et al.* A general geographical probabilistic factor model for point of interest recommendation[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2015, 27(5): 1167–1179. doi: [10.1109/TKDE.2014.2362525](https://doi.org/10.1109/TKDE.2014.2362525).
- [13] LIU Yiding, PHAM T A N, CONG Gao, *et al.* An experimental evaluation of point-of-interest recommendation in location-based social networks[J]. *Proceedings of the VLDB Endowment*, 2017, 10(10): 1010–1021. doi: [10.14778/3115404.3115407](https://doi.org/10.14778/3115404.3115407).
- [14] YUAN Quan, CONG Gao, and SUN Aixin. Graph-based point-of-interest recommendation with geographical and temporal influences[C]. The 23rd ACM International Conference on Information and Knowledge Management, Shanghai, China, 2014: 659–668. doi: [10.1145/2661829.2661983](https://doi.org/10.1145/2661829.2661983).
- [15] 冯浩, 黄坤, 李晶, 等. 基于深度学习的混合兴趣点推荐算法[J]. 电子与信息学报, 2019, 41(4): 880–887. doi: [10.11999/JEIT180458](https://doi.org/10.11999/JEIT180458).
FENG Hao, HUANG Kun, LI Jing, *et al.* Hybrid point of interest recommendation algorithm based on deep learning[J]. *Journal of Electronics & Information Technology*, 2019, 41(4): 880–887. doi: [10.11999/JEIT180458](https://doi.org/10.11999/JEIT180458).
- [16] ZHENG Haitao, ZHOU Yingmin, LIANG Nan, *et al.* Exploiting user mobility for time-aware POI recommendation in social networks[J]. *IEEE Access*, 2017, 99: 1–14. doi: [10.1109/ACCESS.2017.2764074](https://doi.org/10.1109/ACCESS.2017.2764074).
- [17] LIU Shudong and WANG Lei. A self-adaptive point-of-interest recommendation algorithm based on a multi-order Markov model[J]. *Future Generation Computer Systems*, 2018, 89: 506–514. doi: [10.1016/j.future.2018.07.008](https://doi.org/10.1016/j.future.2018.07.008).
- 司亚利: 女, 1981年生, 副教授, 研究方向为兴趣点推荐系统.
张付志: 男, 1964年生, 教授, 研究方向为推荐系统.
刘文远: 男, 1968年生, 教授, 研究方向为物联网系统.