

时间感知的兴趣点推荐方法

王莹丽¹ 姜聪聪¹ 冯小年² 钱铁云¹

¹ 武汉大学计算机学院 武汉 430072

² 中国电力财务有限公司 北京 100005

(481446779@qq.com)

摘要 在基于位置的社交网络(Location-based Social Networks,LBSN)中,用户共享位置与与位置信息相关的内容。兴趣点推荐是LBSN的重要应用,根据用户历史访问签到记录推荐其可能感兴趣的位置。与其他推荐问题(如产品推荐或电影推荐)相比,用户对兴趣点的偏好在时间感知特征上尤为凸显。文中探索了时间感知特征对兴趣点推荐任务的影响,提出了时间感知的兴趣点推荐方法TAPR(Time Aware POI Recommendation)。该算法基于不同的时间尺度构建不同的关系矩阵,并且利用张量分解将构建出的多个关系矩阵分解从而得到用户与兴趣点的表示。最后,该算法利用余弦相似性计算用户与未访问POIs的相似性得分,并结合用户偏好建模的算法得到最终推荐分数。在两个公开数据集上的实验结果表明,TAPR算法比其他基于兴趣点推荐算法表现更好。

关键词: 兴趣点推荐; 时间感知特征; 张量分解; 表示学习

中图分类号 TP311

Time Aware Point-of-interest Recommendation

WANG Ying-li¹,JIANG Cong-cong¹,FENG Xiao-nian² and QIAN Tie-yun¹

¹ School of Computer Science,Wuhan University,Wuhan 430072,China

² China Power Finance Co.,Ltd.,Beijing 100005,China

Abstract In location-based social networks (LBSN), users share their location and content related to location information. Point-of-interest (POI) recommendation is an important application in LBSN which recommends locations that might be of interest to users. However, compared with other recommendation problems (such as product and movie recommendation), the users' preference for POI is particularly determined by the time feature. In this paper, the influence of time feature on POI recommendation task is explored, and a time-aware POI recommendation method is proposed, called TAPR (Time Aware POI Recommendation). Our method constructs different relation matrices based on different time scales, and uses tensor decomposition to decompose the constructed multiple relation matrices to obtain the representation of the user and the POI. Finally, our method uses cosine similarity to calculate similarity scores between users and non-visited POIs, and combines the algorithm of user preference modeling to obtain the final recommendation score. Experimental results on two public datasets show that the proposed TAPR performs better than other POI recommendation methods.

Keywords POI recommendation, Time aware feature, Tensor decomposition, Representation learning

1 引言

随着智能手机等移动设备的广泛使用和无线网络技术的快速发展,一些基于位置的社交网络(Location-based Social Network,LBSN)服务,如Foursquare和Gowalla等越来越受欢迎。在此类服务中,用户可以在感兴趣的位置签到或者与好友分享自己的位置信息。该类用户感兴趣的位置被称为兴趣点(Point of Interest,POI)。随着用户数量和位置信息的增多,为不同的用户推荐POI的方法得到了广泛关注。POI推

荐从两个方面创造价值:1)帮助用户发现城市中感兴趣的位置和服务;2)帮助LBSN中的企业为目标客户发布广告。

POI推荐的潜在价值引起了学术界的诸多关注,POI推荐系统层出不穷^[1-2]。为了建立实用的POI推荐系统,地理位置和时间因素起着重要作用。研究表明,LBSN中的用户移动性地表现出明显的时间感知特征^[1-3]。例如,有的用户总是在周一下午留在办公室,晚上在酒吧娱乐。总而言之,用户签到数据的时间感知特征可以从以下3个方面进行概括。

(1)周期性:用户共享相同的周期模式,在同一时间访问

到稿日期:2021-04-14 返修日期:2021-06-01

基金项目:国家自然科学基金(61572376);国家电网有限公司科技项目(5700-202072180A-0-00-00)

This work was supported by the National Natural Science Foundation of China(61572376) and State Grid Technology Project(5700-202072180A-0-00-00).

通信作者:钱铁云(qty@whu.edu.cn)

相同或类似的 POI^[2-3]。例如,用户总是在中午访问餐馆,其他用户也是如此。

(2)连续性:用户当前的签到位置主要与最近的签到^[1,4]相关。Gao 等^[1]假设用户偏好在连续两个小时内相似并对此属性进行建模。

(3)不统一性:用户的签到偏好在一天的不同时间发生变化^[1]。例如,在中午用户可以访问餐馆,而在晚上用户可以访问酒吧。通过捕获观察到的时间感知特征,已有研究提出了不同方法来增强 POI 推荐性能^[1-2,4]。

之前的工作^[1-2,4]无法将上述 3 个特征统一建模。并且,文献^[5]提到先前的工作忽略了一个重要的事实,即时间影响存在于不同的时间尺度。例如,以一日为尺度,用户一般会在白天访问办公室周围的地方,但是早晨和晚上会在住处周围活动。以一周为尺度,用户一般在工作日留在城市工作,并在周末去郊外游玩。因此,从不同的时间尺度去捕获时间感知特征非常必要。

基于以上想法,本文提出了一种利用时间感知特征构建关系张量的方法,即时间感知的 POI 推荐算法(Time Aware POI Recommendation, TAPR)。该算法利用时间特性构建关系张量,并采用 RESCAL 算法^[6]对张量进行分解,以学习得到更好的用户表示和 POI 表示,并有效地结合用户偏好建模,使其能够更好地向用户推荐感兴趣的 POI。

本文针对用户签到访问记录存在的不同时间尺度的特征,为每个时间尺度构建对应的关系矩阵,将用户与 POI 以及多个关系矩阵形成一个三维张量,然后利用张量分解学习得到用户与 POI 的表示。本文主要将带有时间的签到访问信息分为两类时间尺度。

(1)以日为尺度,将每日根据工作状态和休息状态划分为 5 个不同时段,即凌晨、上午、中午、下午和晚上。

(2)以周为尺度,将每周划分为两个不同的时段,即工作日和周末。

最后,对于通过算法学习到的用户与 POI 的潜在表示,本文利用余弦相似度度量用户与 POI 的相关性,结合用户偏好建模向用户推荐未访问的 POI。

2 相关工作

2.1 兴趣点推荐

POI 推荐算法大多数是基于协同过滤(Collaborative Filtering, CF)方法,这些方法一般分为两种类型:基于内存和基于模型。一方面, Ye 等^[7]提出了基于用户的协同过滤方法来解决 LBSN 中的 POI 推荐问题,并通过线性组合地理影响、社会影响和偏好相似性来进一步完善系统。另一方面,基于模型的 CF 被提出用于解决从其可扩展性中受益的 POI 推荐问题。Cheng 等^[8]提出了一个多中心高斯模型来捕捉用户的地理影响,并将其与社会矩阵分解(Matrix Factorization, MF)模型^[9]结合起来推荐 POI。Gao 等^[10]提出了一种基于 MF 的模型,将位置推荐框架与时间效应(LRT)结合起来,利用时间相邻的框架之间的相似性来提高推荐性能。Ma 等^[11]提出了一种话题-位置-类别感知的协同过滤兴趣点推荐算法 TGC-CF。该算法利用文本信息、地理信息和类别信息,有效地融合了兴趣话题、地理影响和类别偏好因素,从而在一定程

度上缓解了数据稀疏性问题。

2.2 基于时间信息的兴趣点推荐

用户在选择兴趣点进行签到时,难免会受到时间感知特征的影响,已有大量研究工作使用时间信息进行兴趣点推荐。2011 年, Cho 等^[4]提出了 LBSN 中登记数据的周期性。人们总是在中午去餐馆,所以他们根据周期性,在中午为用户推荐没有去过的餐馆。Pearson 相关性可以帮助实现在同一时间段推荐类似的 POI。2013 年, Yuan 等^[2]结合时间相似性和非时间相似性来更好地测量相似性。同年, Gao 等^[4]观察了非均匀性属性(用户的登记偏好在一天的不同时间变化)和连续性(用户在这一时间点的偏好与上一时间点的相似),提出了基于 MF 技术的 LRT 模型来模拟非均匀性和连续性。同时, Cheng 等^[10]提出了具有局部区域约束(FPMC LR)的因式个性化马尔可夫链模型^[12]来捕获连续性,假设两个连续登记的 POI 之间存在强相关性。ATTF 模型^[5]将 3 种时间感知特征融合到一起,对不同尺度的时间影响进行建模。

目前,有很多方法将时间信息作为辅助信息融入 POI 推荐系统中。Han 等^[13]提出利用 POI 的时间感知特征向量来计算 POI 间的相似度,以此为基础构建 POI 相似度图,并利用图方法进行学习。MCAGE 模型^[14]和 STP-UDGAT 模型^[15]也利用时间信息来构建 POI 图。HME 模型^[16]将两个被访问时间间隔小于阈值的 POI 视为有关联。总体而言,这些方法主要利用时间信息来构建 POI 之间的关联,并利用其他方法来对有关联的 POI 进行建模。因此,尽管上述方法隐式融入了时间信息,但没有考虑不同时间段之间和时段与 POI 之间的显式关系。

3 基于时间感知特征的兴趣点推荐

3.1 问题陈述

为便于说明,令 $U = \{u_1, u_2, \dots, u_n\}$ 为 LBSN 的用户集合, n 代表用户的数量; $P = \{p_1, p_2, \dots, p_m\}$ 为兴趣点集合, m 代表兴趣点的数量; $T = \{T_1, T_2\}$ 是时间感知特征的集合。其中 $T_1 = \{t_1, t_2, t_3, t_4, t_5\}$ 是以日为尺度划分的 5 个时间段,分别为凌晨(0 点-8 点)、上午(8 点-12 点)、中午(12 点-15 点)、下午(15 点-19 点)和晚上(19 点-24 点); $T_2 = \{t_{work}, t_{wkd}\}$ 是以周为尺度划分的 2 个时间段,分别是工作日和周末。根据 LBSN 上用户签到到兴趣点的历史信息,可以得到用户-兴趣点交互张量 $\mathbf{X} = \{\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4, \mathbf{X}_5, \mathbf{X}_{work}, \mathbf{X}_{wkd}\}$ 。根据不同的时间感知特征和用户的签到时间, \mathbf{X} 有不同的划分方式:根据 T_1 的 5 个时间段, $\mathbf{X}_{T_1} = \{\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4, \mathbf{X}_5\}$ 对应 5 个交互矩阵;根据 T_2 的 2 个时间段, $\mathbf{X}_{T_2} = \{\mathbf{X}_{work}, \mathbf{X}_{wkd}\}$ 对应着 2 个交互矩阵。表 1 列出了本文中的关键符号及其说明。

表 1 关键符号

Table 1 Key symbols

符号	说明
U, n	用户集合及其数量
P, m	兴趣点集合及其数量
T, T_1, T_2	时间感知特征集合及其两种时间感知特征
u_i	某个用户, $u_i \in U$
p_j	某个兴趣点, $p_j \in P$
\mathbf{X}	交互张量
\mathbf{X}_k	某个交互矩阵, $\mathbf{X}_k \in \mathbf{X}$

兴趣点推荐问题的定义:对于目标用户 $u_i \in U$, 给定该用户的历史签到信息(包括签到的兴趣点和签到时间), 计算用户 u_i 未来会访问的其未访问过的兴趣点 p_j 的感兴趣分数 S_{ij} , 然后为该用户推荐 top- k 个感兴趣分数最高的兴趣点。

3.2 整体框架

本文根据每周以及每日的两个时间感知特征构建用户与 POI 的关系矩阵。我们通过充分利用用户签到信息, 为在两个时间感知特征上具有相同关系的用户和 POI 构建关系矩阵, 进而提出一种支持 POI 推荐任务的多尺度时间关系构建方法。本文的方法主要由两个步骤组成: 1) 关系矩阵的构建; 2) 学习用户与 POI 的表示。

本文首先基于不同时间感知特征分别构建用户 u_i 和兴趣点 p_j 的关系矩阵。然后, 构造一个三维张量 \mathbf{X} , \mathbf{X} 中的每个切片 \mathbf{X}_k 表示每个时间感知特征下的关系 k 中用户和 POI 之间的连接关系。其次, 我们通过使用算法 RESCAL 对张量 \mathbf{X} 进行分解。最后, 通过学习得到用户与 POI 的表示, 并利用余弦相似性计算用户与未访问 POI 的相似性得分, 即可获得此 POI 被推荐给用户的相应分数。

3.3 基于时间感知特征构建关系张量

在兴趣点推荐系统中, 有用户和 POI 两个实体。每个用户在不同时间段对 POI 具有不同的活动记录, 例如, 用户会在工作日访问工作地点以及工作地点附近的位置, 而在周末则会选择访问住处附近或者在商场周围的兴趣点。在不同的时间段, 每个用户的表现特性不一样。因此本文定义了各个时间段用户与 POI 的关系矩阵, 用于记录每个时间段内用户与 POI 之间的关系。故本文基于时间感知特征, 根据用户签到记录的时间信息构造了三维的用户-兴趣点交互张量 \mathbf{X} 。

\mathbf{X} 包含两个不同时间感知特征对应的三维张量 \mathbf{X}_{T_1} 和 \mathbf{X}_{T_2} 。由上述可知: \mathbf{X}_{T_1} 包含工作日关系矩阵 \mathbf{X}_{work} 和周末关系矩阵 \mathbf{X}_{week} 两个切片; \mathbf{X}_{T_2} 包含每天的时间段的 $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4, \mathbf{X}_5$ 关系矩阵。三维张量中的每一个张量切片含有表 2 列出的关系。

表 2 用户与 POI 的关系矩阵

Table 2 Relation matrix between user and POI

	U	P
U	$\mathbf{U-U}$	$\mathbf{U-P}$
P	$\mathbf{P-U}$	$\mathbf{P-P}$

其中, $\mathbf{U-U}$ 表示用户与用户之间的关系, 即同一时间段两个用户共同访问 POI 的个数, 相同的 POI 个数越多表示用户 a 与用户 b 具有的兴趣更相似, 在利用张量分解时用户 a 与用户 b 的特性会相互影响, 使得到的潜在用户表示更接近真实用户。

$\mathbf{U-P}$ 表示用户与 POI 之间的关系, 即同一时间段内用户 a 访问兴趣点 j 的次数与其所有签到次数的比例。其中, $\mathbf{U-U}$ 关系矩阵为频数矩阵, 而 $\mathbf{U-P}$ 矩阵中的元素为比例数, 因此我们将其同比例扩大 10 倍作为签到次数。

$\mathbf{P-U}$ 表示 POI 与用户之间的关系, 即同一时间段内用户 a 访问 POI j 的次数与在该 POI 签到的所有用户数的比例。其中, $\mathbf{U-U}$ 关系矩阵为频数矩阵, $\mathbf{P-U}$ 矩阵中的元素为比例数, 因此我们将其同比例扩大 10 倍作为签到次数。

$\mathbf{P-P}$ 表示 POI 与 POI 之间的关系, 即同一时间段两个 POI 被共同用户访问的个数, 用户相同个数越多表示 POI i 与 POI j 的相似性越高, 同样在利用张量分解时 POI i 与 POI j 的特性会相互影响, 使得到的潜在 POI 表示更接近真实 POI。

在 Gowalla 与 Foursquare 两大真实数据集中存在大量用户以及 POI, 签到记录比较稀疏。在构建用户与 POI 的关系矩阵时, 由于绝大多数人保持相对稳定的日常生活状态(例如, 早上去办公室, 下班后在餐厅吃饭), 当我们将时间分成基于小时的时段以及将每周划分两个时间状态, 会使用户签到记录数据更加稀疏。为了解决矩阵稀疏性, 我们发现如果两个用户在相同时间段具有相似的偏好行为, 则他们可能同时访问类似的 POI。因此, 我们通过利用用户与用户的关系 $\mathbf{U-U}$, 采用协同过滤计算用户 a 与用户 b 之间的相似性, 基于相似用户的相似性设置阈值(如设置为 0.8), 将用户 b 中用户 a 未访问的 POI 的访问记录添加到用户 a 的访问记录中, 以填充用户 a 的访问记录, 即填充 $\mathbf{U-P}$ 与 $\mathbf{P-U}$ 两个关系矩阵。

根据上面提到的时间感知特征以及每个时间感知特征具有相同的关系, 本文分别在不同的时间感知特征下建立用户与 POI 的关系矩阵。为了将不同矩阵的值统一到参考系统, 我们使用 sigmoid 函数进行标准化, 因此, 值“0”将标准化为“0.5”。然后, 我们将这些矩阵联合起来形成相邻的张量, 每个矩阵都是张量的一部分。用户和 POI 被视为张量中的相同实体。我们分别为每周 T_1 以及每日 T_2 建立两个独立的张量 \mathbf{X}_{T_1} 和 \mathbf{X}_{T_2} , 并利用张量分解分别学习用户与 POI 的表示。

3.4 学习用户与 POI 的表示

在一般情况下, 签到信息包含用户在不同时间点的访问记录, 包含用户在某个时刻访问某个 POI 的记录。基于 3.3 节所述的关系, 我们可以构造两个邻接张量 \mathbf{X}_{T_1} 和 \mathbf{X}_{T_2} , 其次, 通过采用文献[6]的 RESCAL 分解算法将不同时间段的关系信息相关用户 ID 和 POIID 转换为对应的向量表示。给定张量 $\mathbf{X}_{n \times n \times m}$, RESCAL 算法具有秩 r 近似, 其中切片 \mathbf{X}_k 被分解为:

$$\mathbf{X}_k \approx \mathbf{A} \mathbf{R}_k \mathbf{A}^T \quad (1)$$

其中, k 的范围是从 1 到 m ; \mathbf{A} 是 $n \times r$ 矩阵, 其中第 i 行表示第 i 个实体; \mathbf{R}_k 是非对称 $r \times r$ 矩阵, 其描述了第 k 个关系的潜在分量的相互作用。

虽然每个切片中的 \mathbf{R}_k 不同, 但 \mathbf{A} 保持不变。通过最小化式(2)所示的损失函数得到 \mathbf{A} 和 \mathbf{R}_k 的表示:

$$f(\mathbf{A}, \mathbf{R}_k) = \frac{1}{2} \left(\sum_k \|\mathbf{X}_k - \mathbf{A} \mathbf{R}_k \mathbf{A}^T\|_F^2 \right) \quad (2)$$

在本文方法中, 切片 \mathbf{X}_k 是关系张量的第 k 个关系。第 i 个实体是第 i 个用户或 POI。如 3.3 节所述, 为了自动获取更多有用的全局信息, 我们构建不同时间段下用户与 POI 的关系矩阵。因此, 我们可以通过张量分解将多重时间段中的信息嵌入到最终学习的表示中。正如 Nickel 等^[6]证明, 所有关系对第 i 个实体的学习潜在分量表示具有决定性影响。它通过学习全局损失函数来消除原始数据的噪声。因此, 我们通过学习所有关系矩阵来进一步优化用户和 POI 的表示。由 3.3 节可知, 本文分别为每周 T_1 以及每日 T_2 两个不同时间尺度建立两个独立的张量, 并分别得到执行张量分解后学

习到的用户(U^{T_1} 和 U^{T_2}) 与 POI(P^{T_1} 和 P^{T_2}) 的表示。随后利用余弦相似度度量计算用户(u) 对于未访问 POI(j) 的感兴趣程度 $S_{uj}^{T_1}$ 和 $S_{uj}^{T_2}$, 计算方法如式(3)、式(4)所示:

$$S_{uj}^{T_1} = \cos(U^{T_1}, P^{T_1}) \quad (3)$$

$$S_{uj}^{T_2} = \cos(U^{T_2}, P^{T_2}) \quad (4)$$

其中, \cos 表示余弦相似度函数。

本文根据每周以及每日两个不同时间尺度分别得到了用户对于未访问 POI 的感兴趣程度。虽然同为时间周期性, 在不同时间尺度下每个用户的表现可能相似, 但也可能有很大差别。例如, 一种情况是用户在周末也会经常工作, 那么周一至周日上午的签到模式相似, 即在周末和工作日其关系也会相同, 因此两个尺度得到的用户与 POI 表示会很相似。另一种情况是在周一至周五的签到模式相似, 而周末签到模式与工作日大相径庭, 因此, 两种尺度下的时间感知特征并不相同。此时将两种状态相互结合, 能向用户提供更好的 POI 推荐。

基于上述分析, 本文采用线性组合方式将两个时间感知特征组合在一起, 作为用户(u) 对于未访问的 POI(j) 的最终感兴趣程度, 如式(5)所示:

$$S_{uj}^T = \alpha_1 * S_{uj}^{T_1} + \alpha_2 * S_{uj}^{T_2} \quad (5)$$

3.5 偏好建模和 POI 推荐

为了推断用户对目标 POI 的偏好, 本文综合考虑了用户偏好和时间信息对 POI 推荐的影响, 使用线性融合框架将基于用户偏好的用户感兴趣的 POI 排名列表和基于时间感知特征得到的用户感兴趣的 POI 排名列表整合成最终排名列表。通过整合两种因素, 令 S_{uj}^U 和 S_{uj}^T 分别表示 POI(j) 处的用户(u) 的签到分数, 分别对应于基于用户偏好和时间影响的推荐得分, 如式(6)所示:

$$S_{uj} = \beta_1 * S_{uj}^U + \beta_2 * S_{uj}^T \quad (6)$$

其中, S_{uj}^U 为基于协同过滤, 通过聚合类似用户的行为来发现用户的隐含偏好得分。

综上所述, 本文提出的 TAPR 算法如算法 1 所示。

算法 1 TAPR 算法

输入: 用户集合 \mathbf{U} , 兴趣点集合 \mathbf{P} , 用户签到的时间和兴趣点信息, 时间感知特征集合 \mathbf{T} ;

输出: 每位用户对未访问的 POI 的感兴趣得分;

1. $\mathbf{T} \leftarrow \{\mathbf{T}_1, \mathbf{T}_2\}$ // 将时间划分成不同的时间段;
2. $\mathbf{X} \leftarrow \{\mathbf{X}_{T_1}, \mathbf{X}_{T_2}\}$ // 根据不同的时间段得到用户-兴趣点交互张量;
3. $f(\mathbf{A}, \mathbf{R}_k) = \frac{1}{2} (\sum_k \|\mathbf{X}_k - \mathbf{A}\mathbf{R}_k\mathbf{A}^T\|_F^2)$ // 张量分解, 得到用户(U^{T_1} 和 U^{T_2}) 与 POI(P^{T_1} 和 P^{T_2}) 的表示;
4. $S_{uj}^{T_1} \leftarrow \cos(U^{T_1}, P^{T_1})$ // 基于时间感知特征 T_1 , 用户对未访问 POI 的感兴趣得分;
5. $S_{uj}^{T_2} \leftarrow \cos(U^{T_2}, P^{T_2})$ // 基于时间感知特征 T_2 , 用户对未访问 POI 的感兴趣得分;
6. $S_{uj}^T \leftarrow \alpha_1 * S_{uj}^{T_1} + \alpha_2 * S_{uj}^{T_2}$ // 基于时间感知特征 T_1 和 T_2 , 用户对未访问 POI 的感兴趣得分;
7. $S_{uj} = \beta_1 * S_{uj}^U + \beta_2 * S_{uj}^T$ // 将基于协同过滤和时间感知特征的感兴趣得分。

4 实验及分析

4.1 实验设置

4.1.1 实验数据集

本文采用文献[17]中使用的两个真实数据集 Gowalla 和

Foursquare 作为实验数据集。

Gowalla 数据集包含 2009 年 2 月至 2010 年 10 月的用户签到信息。与文献[17]相同, 本文筛选出登记 POI 少于 15 的用户以及访客少于 10 人的 POI, 过滤后的数据集包括 18737 个用户、32510 个 POI、1278274 个用户访问记录。

表 3 Foursquare 和 Gowalla 的数据统计

Table 3 Data statistics of Foursquare and Gowalla datasets

	Foursquare	Gowalla
# Users	24941	18737
# POIs	28593	32510
# check-ins	1196248	1278274

Foursquare 数据集包含 2012 年 4 月至 2013 年 9 月的用户签到数据。Foursquare 数据集消除了签到 POI 次数少于 10 的用户以及少数 POI 的访客超过 10 位的 POI, 过滤后的数据集包含 24941 个用户、28593 个 POI 和 1196248 个用户访问记录。

我们将每个数据集划分为训练集、验证集和测试集。对于每个用户, 我们使用早期的 70% 的访问记录作为训练数据, 最近的 20% 的访问记录作为测试数据, 剩余 10% 的记录作为验证数据。此外, 我们通过训练数据中的用户签到信息作为用户偏好建模, 利用所有用户含有时间点信息的访问记录来学习用户与 POI 的表示。再利用张量分解学习得到用户与 POI 的表示, 并计算用户对于未访问 POI 的相似性作为推荐访问得分, 然后结合用户偏好建模计算每个用户的不同候选 POI 的感兴趣得分, 选择前 K 个候选 POI, 并将它们与测试数据中的真实的签到 POI 作比较。我们在 Gowalla 和 Foursquare 的数据集上评估了 K 分别取 5, 10 以及 20 的实验效果。

4.1.2 评价指标

本文使用文献[17]提出的 3 种广泛使用的度量来评估模型, 即精确度(Precision)、召回率(Recall)以及归一化折损累积增益(NDCG)。NDCG 是衡量排序质量的指标, 能够较好地表示实验效果, 我们选择推荐列表的前 K 个 POI 进行计算。评价指标分别表示为 $Pre@K$, $Re@K$, $NDCG@K$, 其中 K 是推荐列表中的 POI 个数。

Precision 和 Recall 沿用文献[5]中的定义, 具体如下:

$$Pre@K = \frac{1}{|U|} \sum_{u \in U} \frac{|P_u^K \cap T_u|}{|P_u^K|} \quad (7)$$

$$Re@K = \frac{1}{|U|} \sum_{u \in U} \frac{|P_u^K \cap T_u|}{|T_u|} \quad (8)$$

其中, U 表示用户集合, K 是推荐列表中的 POI 数, P_u^K 表示用户 u 推荐列表中 POI 的前 K 个推荐, T_u 表示测试集中用户 u 的真实 POI 推荐列表。对于每个度量, 我们在实验中将 K 设置为 5, 10 及 20。

NDCG^[8] 的定义如式(9)所示:

$$NDCG@K = \frac{DCG@K}{idealDCG@K} \quad (9)$$

其中, $DCG@K = \sum_{i=1}^{p \leq K} \frac{1}{np \cdot \log(i+2)}$, p 是推荐 K 列表中推荐正确的 POI 个数, $idealDCG@K$ 是理想状态下的 DCG。

4.1.3 对比方法

我们将本文方法(TAPR)与 5 种较为先进的 POI 推荐方

法以及利用时间信息的 LRT 方法进行对比,以证明本文模型利用时间信息的有效性。

(1)USG^[7]:USG 探索用户偏好、社会影响力和 POI 的地理特征对用户访问 POI 的影响。此外,USG 采用一个统一的 POI 推荐框架,它将用户偏好融合到了具有社会影响力和地理影响力的 POI 推荐过程中。

(2)LRT^[1]:基于从现实世界 LBSN 数据集观察到的用户移动的时间属性,LRT 引入了一种新颖的 POI 位置推荐框架。LRT 中的实验结果表明了时间模式在解释用户行为方面的重要性。

(3)LORE^[18]:LORE 方法除了考虑社会和地理影响外,还考虑了连续影响。LORE 利用了 POI 推荐的顺序影响,并且将地理影响和社会影响的连续影响融合到一个统一的推荐框架中;特别地,地理影响被建模为二维 POI 签到概率分布而不是现有研究中的一维距离概率分布。

(4)GeoMF^[19]:GeoMF 通过增加原始用户和 POI 的潜在因素来扩展 MF,包括用户的活动区域和 POI 影响区域。

(5)RankGeoFM^[20]:RankGeoFM 是一种基于排名的 MF 模型,它通过考虑邻近 POI 的吸引力包括地理影响并根据 POI 的最近评论来进行 POI 推荐。

(6)GeoIE^[21]:GeoIE 利用 POI 特定的地理影响来改进 POI 推荐,并使用 3 个因素来模拟两个 POI 之间的地理影响,即 POI 的地理影响、POI 的地理敏感性以及它们的物理距离。

4.2 实验结果

在将本文方法与基线方法进行比较之前,我们首先使用两种不同的时间感知特征来评估本文方法的性能,如图 1 所示。3.3 节中的两个时间感知特征每日 T_1 时间段和每周 T_2 时间段分别被命名为 time_hour 和 time_week。两个时间感知特征的线性结合表示为 time_hour + time_week。因为 Foursquare 数据集和 Gowalla 数据集上具有相同的实验效果趋向,所以本文仅在 Gowalla 数据集上对本文划分的时间感知特征进行相关实验比较。从图 1 观察到,对于单个时间感知特征因素, T_2 明显优于 T_1 。其原因在于:在 T_1 时间段,时间被细分为不同小时的时间段,使得数据集中签到的用户与 POI 的关系矩阵更加稀疏,即使利用用户相似性填充了部分用户访问记录的用户与 POI 的块矩阵,但用户与 POI 的关系矩阵相对每周状态下的时间段还是比较稀疏。

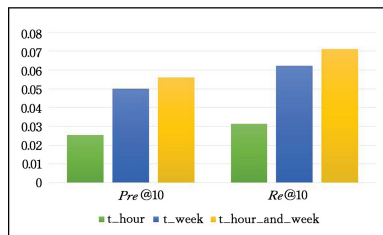


图 1 时间感知特征在 Gowalla 数据集上的比较

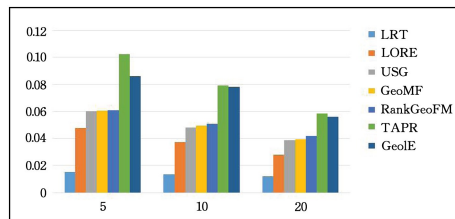
Fig. 1 Comparison of time features in Gowalla

从图 1 中可以看出,两个特征的线性结合使实验效果达到更好。每日不同时间段的用户特征充分表明了用户在当下时间段的偏好特征,相对而言,每周划分周末与工作日的状态,工作日关系矩阵更能体现用户工作日的特性,周末关系矩

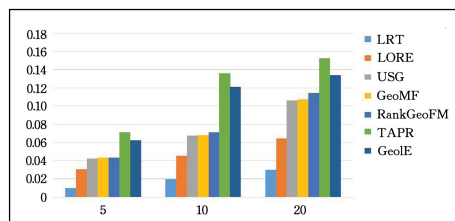
阵则体现了用户周末的特性,因此实验结果表明将两者线性结合时效果达到最好。

因此,本文提出将 T_1 时间尺度特征与 T_2 时间尺度特征线性结合作为本文利用时间信息的模型,并结合用户偏好一同建模,得到最终的用户对候选 POI 的感兴趣程度。

我们将本文方法与 4.1.3 节中的基线方法进行对比,实验结果如图 2、图 3 所示。其中,TAPR 表示本文方法,其他六种为 POI 推荐基线方法。



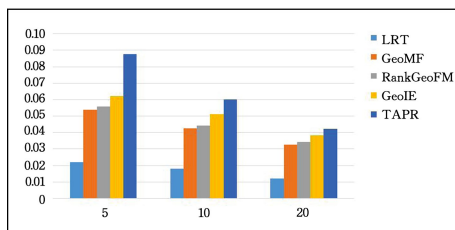
(a) Pre@K-Gowalla



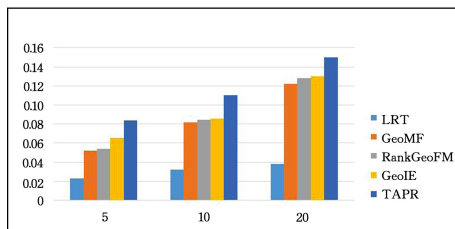
(b) Re@K-Gowalla

图 2 Gowalla 数据集上的基线方法对比

Fig. 2 Comparison of baseline methods on Gowalla



(a) Pre@K-Foursquare



(b) Re@K-Foursquare

图 3 Foursquare 数据集上的基线方法对比

Fig. 3 Comparison of baseline methods on Foursquare

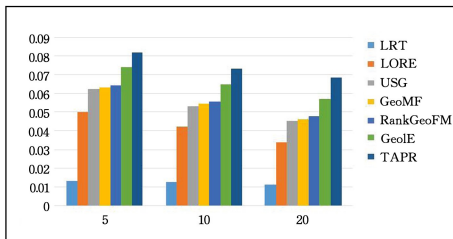
首先,从图 2、图 3 中可以看出,本文方法在 Gowalla 数据集和 Foursquare 数据集上的 Precision 和 Recall 值明显优于 GeoMF 以及 GeoIE。

其次,本文方法相比于 LRT 等较为传统的方法,性能提高非常明显。USG 与 LORE 均使用了朋友信息而未使用时间信息,这说明合理利用时间信息能够很好地提高推荐性能。LRT 在这些极限模型中表现最差,因为它只考虑了时间信息,而未对地理信息和社会影响进行建模。虽然 LRT 方法也考虑了事件周期性,但基于时隙划分签到矩阵使得数据更加

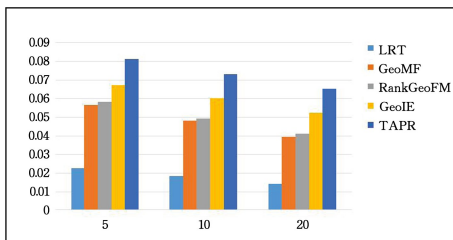
稀疏,对学习用户偏好造成了负面影响。而本文方法采用张量分解更好地解决了处理时间信息时的高维数据以及矩阵稀疏性的问题,使其 POI 推荐性能得到提升。

类似地,从图 4 可以看出,本文方法在 Gowalla 数据集和 Foursquare 数据集上的 NDCG 值明显优于 GeoMF 以及最先进的 GeoIE 方法。

总体而言,对于 POI 推荐任务,本文提出的 TAPR 方法在两个数据集上的整体表现要比现有 POI 推荐方法更好,这表明时间信息的利用对提高推荐预测的准确性具有重要作用。



(a) NDCG@K-Gowalla



(b) NDCG@K-Foursquare

图 4 两个数据集上的 NDCG@K 结果对比

Fig. 4 Comparison of NDCG@K results

结束语 本文提出了一个利用时间尺度特征构建用户与 POI 关系矩阵的算法,并利用张量分解得到用户和 POI 的向量表示,再根据上述表示得到推荐得分。最后,我们将提出的方法在 Foursquare 以及 Gowalla 两个大型数据集上与其他 POI 推荐基线方法进行了对比实验,实验结果表明本文方法相对于其他方法表现更好。

本文提出的模型目前仅利用了时间信息以及用户偏好信息,在今后研究中,将同时考虑地理因素、朋友关系等附加信息,并将其与时间感知特征更好地结合。

参 考 文 献

[1] GAO H, TANG J, HU X, et al. Exploring temporal effects for location recommendation on location based social networks [C]//Proceedings of the 7th ACM Conference on Recommender Systems. ACM, 2013:93-100.

[2] YUAN Q, CONG G, MA Z, et al. Time-aware point-of-interest recommendation[C]// Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 2013:363-372.

[3] MUHAMMAD A S, ROHIT K, TOON C, et al. Location influence in location-based social networks[C]// Proceedings of the Tenth ACM International Conference on Web Search and Data Mining. 2017:621-630.

[4] CHO E, MYERS S A, LESKOVEC J. Friendship and mobility: user movement in location-based social networks[C]// Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2011:1082-1090.

[5] ZHAO S, LYU M R, KING I. Aggregated Temporal Tensor Factorization Model for Point-of-interest Recommendation [C]//Neural Information Processing—23rd International Conference ICONIP. 2016:450-458.

[6] NICKEL M, TRESP V, KRIEGEL H P. A Three-Way Model for Collective Learning on Multi-Relational Data[C]// International Conference on Machine Learning. Omnipress, 2011.

[7] YE M, YIN P, LEE W C, 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. ACM, 2011:325-334.

[8] CHENG C, YANG H, KING I, et al. (2012) Fused matrix factorization with geographical and social influence in location-based social networks [C]// Proceedings of the Twenty-sixth AAAI Conference on Artificial Intelligence. 2012:17-23.

[9] MA H, ZHOU D, LIU C, et al. Recommender systems with social regularization[C]// Proceedings of the Fourth ACM International Conference on Web Search and Data Mining. ACM, 2011:287-296.

[10] CHENG C, YANG H, LYU M R, et al. Where you like to go next: successive point-of-interest recommendation [C]// Proceedings of the Twenty-third International Joint Conference on Artificial Intelligence. AAAI Press, 2013:2605-2611.

[11] MA L B, QIN X J. Topic-Location-Category Aware Point-of-interest Recommendation [J]. Computer Science, 2020, 47(9): 81-87.

[12] RENDLE S, FREUDENTHALER C, SCHMIDT-THIEME L. Factorizing personalized Markov chains for next-basket recommendation[C]// Proceedings of the 19th International Conference on World Wide Web. ACM, 2010:811-820.

[13] HAN P, LI Z, LIU Y, et al. Contextualized Point-of-Interest Recommendation[C]// Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI. 2020: 2484-2490.

[14] CHEN J, MENG X W, JI W Y, et al. POI Recommendation Based on Multidimensional Context-aware Graph Embedding Model[J]. Journal of Software, 2020, 31(12):3700-3715.

[15] LIM N, HOOI B, NG S K, et al. STP-UDGAT: Spatial-Temporal-Preference User Dimensional Graph Attention Network for Next POI Recommendation[C]// The 29th ACM International Conference on Information and Knowledge Management. CIKM, 2020:845-854.

[16] FENG S, TRAN V T, CONG G, et al. HME: A Hyperbolic Metric Embedding Approach for Next-POI Recommendation[C]// Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. 2020: 1429-1438.

[17] LI X, GAO C, LI X L, et al. Rank-geofm: A ranking based geographical factorization method for point of interest recommenda-

tion[C] // Proceedings of the 38rd International ACM SIGIR Conference on Rresearch and Development in Information Retrieval. 2015;433-442.

- [18] ZHANG J D, CHOW C Y, LORE Y L. Exploiting sequential influence for location recommendations[C] // Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. 2014;103-112.
- [19] LIAN D, ZHAO C, XIE X, et al. Geomf: joint geographical modeling and matrix factorization for point-of-interest recommendation[C] // The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2014;831-840.
- [20] LIU Y, PHAM T A N, CONG G, et al. An experimental evaluation of point-of-interest recommendation in location-based social networks[C] // Proc. VLDB Endow, 2017.
- [21] WANG H, SHEN H, OUYANG W, et al. Exploiting POI-Specific Geographical Influence for Point-of-Interest Recommenda-

tion[C] // Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence(IJCAI 2018). 2018; 3877-3883.



WANG Ying-li, born in 1992, master. Her main research interests include recommendation systems and so on.



QIAN Tie-yun, born in 1970, Ph.D, professor, Ph.D supervisor, is a member of China Computer Federation, ACM and IEEE. Her main research interests include Web mining and natural language processing.