

DOI:10.13203/j.whugis20200355



文章编号:1671-8860(2021)05-0681-10

融入空间关系的矩阵分解POI推荐模型

魏海涛^{1,2} 李柯^{1,2} 赫晓慧^{1,2} 田智慧^{1,2}

1 郑州大学地球科学与技术学院,河南 郑州,450052

2 郑州大学中国气象科学研究院生态气象联合实验室,河南 郑州,450052

摘要:兴趣点(point of interest, POI)推荐是在基于位置的社交网络中流行起来的个性化服务。针对数据稀疏和隐性反馈的使用等问题,提出了一种关系型矩阵分解模型——合作竞争矩阵分解(cooperative competition matrix factorization, CC-MF)。该模型根据用户与POI间的相互关系建模,融入空间关系,并将空间关系细分为空间距离关系和空间拓扑关系,挖掘POI之间、POI与用户之间的空间关系,以缓解数据稀疏问题;同时使用加权最小二乘准则构建目标函数,缓解隐性反馈问题。在现实世界签到Foursquare数据集上进行实验,结果显示:(1)CC-MF模型显著提高了推荐结果的准确性;(2)考虑空间拓扑关系的空间距离因素能够进一步提升推荐系统的性能。因此,CC-MF模型具有良好的拓展性和解释性,且缓解了数据稀疏和隐性反馈使用问题。

关键词:兴趣点推荐;基于位置的社交网络;矩阵分解;空间关系;空间距离

中图分类号:P208

文献标志码:A

推荐系统是根据用户的某些行为收集用户的偏好,推荐给用户适合他们偏爱的服务或者物品^[1]。推荐系统在网上购物、浏览器搜索、外卖等领域具有广泛的应用。随着社交网络、全球定位技术和移动互联服务的发展,基于位置的社交网络(location-based social network, LBSN)服务也随之流行起来,例如Foursquare、Gowalla、Yelp^[2]等服务。这些LBSN服务通过用户的社交媒体将用户访问一个位置或签到的信息分享给社交网络好友,通过这种签到信息,可以发掘出用户与其签到位置的相互关系,并以此帮助人们寻找他们感兴趣的地方。兴趣点(point of interest, POI)推荐是根据用户的历史签到信息^[3]学习用户的访问偏好,推荐给人们感兴趣但没有访问过的位置。POI推荐不仅能够帮助人们发现未知的兴趣的地方,而且能够帮助营业者吸引潜在用户。

目前,POI推荐面临着数据稀疏和隐性反馈的使用这两大问题^[4]。针对数据稀疏性,很多研究者尝试使用加边信息来解决数据稀疏的问题^[5-10]。由于空间因素在POI推荐中能产生较大

影响,文献[11]研究了空间聚类现象以及空间距离对于用户签到行为的影响;文献[12-13]首先研究空间聚类现象,通过Tobler的地理学第一定律^[14]得出“一个用户倾向访问附近的POI”,然后根据用户签到数据,分析得出用户签到的概率服从幂律分布;文献[15]通过分析用户活动性提出了一种包括二维高斯核密度分析算法和一维幂律算法的推荐框架;文献[13]通过人类活动的阶段性表现研究时间对POI推荐的影响;文献[16]通过将时间的影响分为长期影响、短期影响和时变影响来分析时间对POI推荐的作用,取得了较好的结果。除了空间和时间因素外,还有学者考虑了社交关系^[6,17-20]。文献[6]提出了一种评价社交关系的一种度量,即信任度,由信任度来计算不同社交关系的影响程度;文献[17]提出了潜在的社交关系,认为用户潜在的社交关系会对用户的偏好产生影响,但是社交关系因素对于POI推荐系统性能的提升作用有限。也有研究提出使用社交评论信息来帮助推荐^[17,21-23],如文献[22]研究了用户评论的语义信息,并给出了这些信息对于POI推荐性能的提升;文献[24]使用卷积矩

收稿日期:2020-07-16

项目资助:国家重点研发计划(2018YFB0505000);河南省重点研发与推广专项(科技攻关)(192102210124)。

第一作者:魏海涛,博士,讲师,主要从事地理信息和智慧城市研究。zzu_wei@163.com

通讯作者:赫晓慧,博士,教授。hexh@zzu.edu.cn

阵分解(matrix factorization, MF)方式较好地融入了社交评论信息。针对隐性反馈使用问题,文献[25-27]采用加权矩阵分解(weighted matrix factorization, WMF)或负采样方法来解决该问题,分别是为每个元素增加一个权重或者概率的方式和用缺失数据建模的方式。这两种方法是目前针对隐性反馈使用问题的主要应对方法。

针对POI推荐系统中的数据稀疏性,本文采用融入空间关系的方法来缓解该问题。当前研究虽然考虑了地理因素的影响^[4,9,12-13,28-29],并且分析了地理因素的影响模式,但一般都是从空间距离的角度分析该因素的作用,较少考虑空间拓扑关系的影响,而空间拓扑关系是地理空间相关关系的重要组成部分。对于空间分析,空间关系的确立非常重要,因此考虑空间拓扑关系对于融入地理空间要素的POI推荐至关重要。本文在融入空间关系时,将空间关系进一步细分为空间距离关系与空间拓扑关系,缓解了数据稀疏的影响。

针对隐性反馈使用问题,本文采用WMF技术来缓解此问题,提出合作竞争矩阵分解模型(cooperative competition matrix factorization, CC-MF),首先通过分析影响POI推荐的相关关系,将空间关系融入到推荐模型中,从用户与POI关系的角度融入空间距离因素,从POI与POI关系的角度融入考虑空间拓扑关系的空间距离因素,然后综合它们间的合作关系和竞争关系寻找空间关系的影响。通过使用隐语义模型和WMF以及曝光矩阵分解模型完成推荐任务的启发^[27],从关系的角度分析空间关系因素的影响,并把空间关系融入到MF模型中,完成POI推荐。

1 兴趣点推荐

1.1 问题描述

在POI推荐中,极度的数据稀疏严重影响POI推荐结果的精度。例如,本文使用的用户POI签到Foursquare数据集的数据密度为0.81%。而在推荐系统方面流行的movielens数据集(small)的数据密度为1.7%,由此可见POI推荐任务形势更加严峻。POI推荐的数据稀疏性提高了推荐算法运算复杂性及正确推荐的难度。

根据数据反馈得到的信息类型可以将数据反馈分为显性反馈和隐性反馈^[26],显性反馈数据能够获取正例和负例,通过正例获得用户喜欢什么,通过负例获得用户不喜欢什么,因此只需要

通过这些显性反馈去推测用户的偏爱。但是,显性反馈数据获取很困难。而隐性反馈数据(例如签到数据)只有正例,没有负例,该类问题被称为单类问题^[25]。在用户-POI矩阵中,有签到信息的元素标记为1,没有签到信息的元素标记为0,如图1所示。对于没有签到信息的元素,不知道是用户不喜欢还是用户没接触过此POI,或是其他原因。隐性反馈不能直接得出用户的偏好,但是能利用它帮助推荐,因此隐性反馈值得深入考虑。

		POI →			
		1	0	1	1
User ↓	0	1	1	1	
	1	0	0	0	
	1	1	0	0	

图1 用户-POI矩阵

Fig. 1 User-POI Matrix

1.2 矩阵分解

MF技术是协同过滤(collaborative filtering, CF)技术中比较流行的一种处理方法。将MF运用到推荐算法上,可以将m个用户在n个POI上的原始用户签到矩阵 $R_{m \times n}$ 分解为一个用户特征矩阵 $P_{m \times k}$ 和物品特征矩阵 $Q_{k \times n}$, P 和 Q 的隐性特征空间的维度为k。用户原始签到矩阵 R 通常是一种非常稀疏的矩阵,且矩阵中元素的个数也能达到千万甚至上亿,使用MF技术可以有效降低数据的维度,合理有效地利用有限的信息。

奇异值分解(singular value decomposition, SVD)是矩阵分解的一种形式,不同于其他矩阵分解方法,SVD适用于任何矩阵。对于矩阵 $A_{m \times n}$,可以通过SVD得到 $A_{m \times n} \rightarrow X_{m \times k} \times Y_{k \times n}$ 。因为MF的扩展性极好和易于实现,研究人员提出了多种拓展MF算法,以适应不同推荐任务和改进原始MF模型。SVD++是MF中的一个经典模型,它是基于用户的某些隐性反馈信息能够从侧面反映用户的某些偏好的假设提出的。局部低秩矩阵近似(local low rank matrix approximation, LLORMA)是由文献[30]根据矩阵分解出来的矩阵需要满足局部低秩假设提出的。WMF是文献[25]提出用来解决单类问题的一种策略,使用的是加权低秩逼近技术;还有一种策略是负采样技术。社交正则化(social regularization,

SR_{ui})是由文献[31]提出的,它将隐性的社交关系融入到了MF中,认为用户的隐性社交关系也能影响人们的选择。除此之外,还有融合深层神经网络和MF的推荐技术^[32],如广义矩阵分解与多层次感知机结合等。

1.3 地理信息的作用

地理信息与人们的生活密切相关,它已经在人们的生活中变得常见,许多研究者已将其作为影响POI推荐的因素加入考虑。如文献[12]提出的统一协同POI推荐框架——基于用户偏好和社交影响以及地理影响的推荐(user preference/social influence/geographical influence based recommendation, USG),将用户偏好、社交信息、地理信息等因素加入到POI推荐模型中,帮助推荐。对于地理信息,他们先对数据进行统计分析,然后从分析结果中总结距离与用户签到之间的关系。文献[4]提出了特征空间分离因子分解模型(feature-space separated factorization model, FSS-FM)模型,加入了时间因素和空间因素的影响,它认为POI的隐性特征中包含时间序列特征和空间特征,并通过时间邻居和空间邻居来进行进一步的求解。文献[33]中提出了一个使用视觉内容、地理信息和社交关系的联合框架——视觉内容和地理影响POI推荐(visual contents and geographical influence for point-of-interest recommendation, VCG),使用图片挖掘更多的POI信息来增强或完善POI的隐性向量。文献[34]通过建立一个自动捕捉用户签到行为的时空上下文信息,且结合知识图谱嵌入的时空上下文感知和基于翻译推荐框架(spatiotemporal context-aware and translation-based recommender framework, STA),更加深刻地发掘了时空信息与用户签到行为的相关性。文献[35]将POI的地理

影响、POI的地理敏感性及其物理距离纳入学习框架,研究了POI特有的地理影响。

2 CC-MF 矩阵分解模型

针对数据稀疏和隐性反馈使用问题,基于经典的SVD矩阵分解模型,本文提出了一种关系型用户个性化推荐模型CC-MF。首先,CC-MF模型的目的是对于不同用户可以推荐给他们符合自身偏好的POI。其次,CC-MF模型是一种强调个性化推荐的推荐算法,根据合作关系和竞争关系来构建模型。本文关于合作关系的描述为:对于用户,一个用户可能访问某区域内的多个POI,因此,对于POI,它周围POI的被访问次数就会与该POI的访问之间产生一种合作关系,本文使用曝光量描述和计算这种关系的基础。因为在用户-POI签到矩阵中,用户没有签到记录的POI不一定是用户不喜欢的,还有可能是用户根本没有接触过的,所以可以从这方面考虑,假设用户接触POI,给POI赋予一个权重,根据考虑空间拓扑关系的空间距离因素,该值由此因素确定的地理邻居反映。竞争关系描述为:对于用户而言,(1)从时间角度上来讲,同一时间用户不能访问两个POI,这是带有时间的现实属性;(2)从POI角度来看,POI之间存在竞争关系,表现在大部分POI是想得到比较多的访问用户而不是较少的。所以POI之间既有增加该POI被访问概率的因素,又有阻碍该POI被访问的因素。据此,本文在添加影响因素时,通过赋予不同权重来表示合作-竞争关系。CC-MF模型是考虑合作关系和竞争关系而提出的个性化推荐模型,也是结合了空间拓扑因素、空间距离因素的拓展MF。

本文的整体推荐框架如图2所示,具体步骤如下:

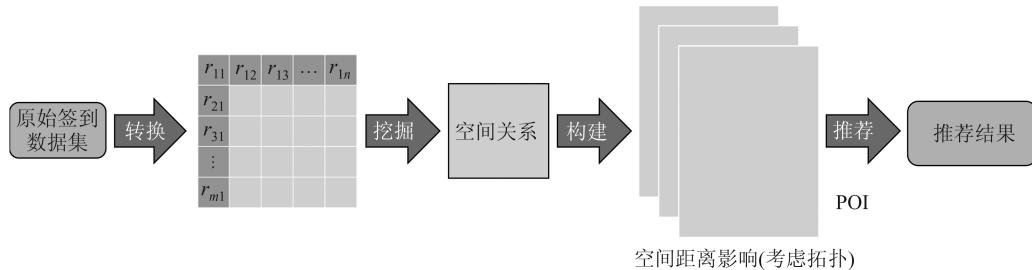


图2 整体推荐框架

Fig.2 Overall Recommendation Framework

1) 转换数据承载方式。对原始签到数据集进行处理,提取信息后转换成用户-POI签到

矩阵。

2) 挖掘空间关系影响。挖掘用户与POI之

间的空间距离因素影响和POI与POI之间的考虑空间拓扑关系的空间距离影响。

3) 矩阵分解。构建融入空间距离因素和考虑空间拓扑关系的空间距离因素的矩阵分解模型,获得用户与POI特征向量。

4) 个性化推荐。生成给定用户的Top- k 个POI推荐列表,实现个性化POI推荐。

对于考虑合作关系和竞争关系的推荐模型,参考文献[36],其基本的预测公式为:

$$\hat{r}_{ui} = b_i + b_u + \mathbf{p}_u \mathbf{q}_i^T \quad (1)$$

式中, \hat{r}_{ui} 是对 r_{ui} 预测的一个值, \hat{r}_{ui} 越大, 表示用户 u 访问POI i 的概率越大, 反之亦然; b_i 是POI i 的偏置项; b_u 是用户 u 的偏置项; \mathbf{p}_u 是用户 u 的特征向量; \mathbf{q}_i^T 是POI i 的特征向量的转置。

2.1 空间距离因素影响

生活圈就是根据人类行为的时空特征、公共资源划分出的一种活动区域,是人类生活在空间上的展开^[37]。从认知科学和地理学结合来看,2014年,诺贝尔生理学或医学奖获得者O'Keefe教授发现了位置细胞,Moser夫妇在O'Keefe教授的研究基础上发现了网格细胞,这两种细胞的作用是形成脑中位置信息。所以,人们能够对经常活动的区域在大脑中形成认知地图,依据认知地图,人们可以在脑中形成一种距离概念,判断人们与POI距离是否在可承受范围。而对于计算距离因素常用的幂律公式,由于收集到的原始签到数据中签到位置一般是由经纬度表示的,所以由此来计算两个POI间的转移概率,不过原始签到数据同一个人连续签到的间隔或一个POI连续有人在此签到的间隔较长,可能是几小时,也可能是几天或几星期^[16],这对于数据清洗及统计使用两个连续的签到行为是一个巨大的挑战。除此之外,使用幂律公式由距离求取转移概率所使用的距离变量,其取值范围能从0到几十千米,在半径几十千米的范围内包含上千个POI^[12-13],这会导致POI的转移概率量级差异巨大,从而使计算得到的较大的转移概率吞没掉较小的转移概率而产生与现实不符的情况。所以,本文首先将生活圈概念、特点及作用引入到POI推荐领域,在LBSN上添加生活圈,并根据以相关行为主义地理学、时间地理学、活动分析法等为核心的行为主义研究以及实际情况,选取生活圈半径。其次,计算用户以往的活动中心,统计用户访问过的POI位置(经纬度)、频率,假设每个POI的权重相同,计算出每个用户的活动中心。之

后,根据用户活动中心和生活圈,所有POI对于每个用户都会被分为两个部分,即生活圈内的POI和生活圈外的POI。通过生活圈来判断空间距离因素的影响,可以清晰地将POI划分为生活圈内和生活圈外,使得处理空间距离因素更加符合实际情况。为了满足上述计算需求,本文使用非线性函数给予空间距离因素影响一个值,计算公式为:

$$D_{ui} = \begin{cases} 1, & l(i|u) \leq d \\ 0, & l(i|u) > d \end{cases} \quad (2)$$

式中, $D_{ui} = 1$ 表示POI i 在以用户 u 活动中心为中心的生活圈范围内,反之则不在; l 是POI i 距离用户 u 往常活动中心的距离; d 是本文根据生活圈理论以及实验测试所得出的生活圈半径。

根据式(2)对每个用户、每个POI进行判断,计算得出对每个用户存在空间距离影响的POI集合,得到预测公式为:

$$\hat{r}_{ui} = b_i + b_u + \mathbf{p}_u (\mathbf{q}_i^T + \alpha_1 \sum_{i' \in I_{u,d}^D} w_{i'} \mathbf{q}_{i'}^T) \quad (3)$$

式中, α_1 为空间距离关系影响系数; $I_{u,d}^D$ 为距离用户 u 活动中心小于 d km 的POI集合; $w_{i'}$ 为 $I_{u,d}^D$ 集合长度的倒数。

2.2 考虑空间拓扑关系的空间距离因素影响

在地理学中,地理空间关系不仅包括空间距离关系,还有空间拓扑关系。空间拓扑关系对于地理学是一种重要的空间关系,对空间关系的建立和空间分析有着至关重要的作用。由于空间拓扑关系能显示和突出区域性、相邻性等关系,所以本文不仅融入了空间距离关系的影响,还考虑带有空间拓扑关系的空间距离因素的影响,并以此来帮助POI推荐。

根据之前提到的合作关系和竞争关系,POI的地理邻居会对其产生影响,所以本文使用POI地理邻居的曝光量来反映考虑空间拓扑关系的空间距离因素的影响。由于原始签到数据中是以POI的形式来记录空间实体的,这与承载着拓扑关系的矢量或栅格地图中以面状实体记录空间实体不同,不能直接判断相邻的空间实体,并且两个POI距离最近并不代表它们对应的空间实体是最近的。所以,本文是在空间拓扑关系约束下使用 k 近邻(k -nearest neighbor, k NN)的方法来寻找POI的地理邻居,并以此反映空间实体的关系。首先,在LBSN上根据POI的地理坐标选择离该POI最近的 k 个POI作为它的地理邻居(由于空间拓扑关系约束, k 不能太小),记录它们

在测试签到数据集上被访问频率的总和,之后对所有 POI 都计算其地理邻居;然后,对所有 POI 的地理邻居被访问频率求均值;最后,对于每个 POI 地理邻居被访问频率与均值求差,并乘以一个系数 μ ,再使用 sigmoid 函数将 POI 的考虑空间拓扑关系的空间距离影响进行归一化。根据式(4)~(7)计算 POI 的考虑空间拓扑关系的空间距离影响。

$$G'_i = \sum_{j \in I_i} F_j \quad (4)$$

$$\bar{G} = \frac{\sum_{i \in I} G'_i}{\text{card}(I)} \quad (5)$$

$$G_i = \sigma(\mu(G'_i - \bar{G})) \quad (6)$$

$$\sigma = \frac{1}{1 + e^{-x}} \quad (7)$$

式中, G'_i 为 POI i 的地理邻居被访问频率总和, 表示 POI i 的地理邻居在历史签到记录中的总体曝光量; I_i 为 POI i 的地理邻居集合; I 为所有 POI 集合; F_j 为 POI j 被访问的频率; \bar{G} 为所有 POI 地理邻居被访问频率总和的均值, 它是衡量一个 POI 地理邻居总体曝光量大小的指标, 类似于将 POI 分为位于繁华地区或不繁华地区, 此外, 由于采用的是自适应数据集的均值而不是固定值, 因此对于其他地区数据集也有一定的适用性; G_i 为 POI i 受到的考虑空间拓扑关系的空间距离影响; μ 为计算系数, 本文取 0.1。计算 G_i 的过程中使用 sigmoid 函数是为了增大地理邻居访问频率总和在 \bar{G} 附近的 POI 间差异, 而减小地理邻居访问频率总和远大于或远小于 \bar{G} 的 POI 间的差异, 可以强化考虑空间拓扑关系的空间距离因素对位于均衡地带 POI 的影响, 而弱化对位于特别繁华地带的 POI 或位于特别不繁华地带的 POI 的影响。

由于签到矩阵上大多数标记为 0 的元素并非用户不喜欢, 而是不知道, 所以本文的任务是模拟用户接触 POI, 使用用户地理邻居被访问的次数总和作为一个计算量, 并依靠此值计算考虑拓扑空间关系的空间距离因素影响。这样做的优势在于:对于那些用户没有签到的 POI, 它们也将会有个值参与后续的计算, 能够更好地提取 POI 的特征。加入考虑空间拓扑关系的空间距离影响的预测公式为:

$$\hat{r}_{ui} = b_i + b_u + p_u(q_i^T + \alpha_1 \sum_{i' \in I_{u,d}^D} w_{i'} q_{i'}^T + \alpha_2 \sum_{i'' \in I_{i,K}^G} w_{i''} q_{i''}^T) \quad (8)$$

式中, α_2 是考虑空间拓扑关系的空间距离影响系

数; $I_{i,K}^G$ 为 POI i 的 K 个地理邻居集合; $w_{i''}$ 为使用 softmax 函数根据空间拓扑关系影响 G_i 计算的权重。

2.3 计算优化策略

本文使用 WMF 中的加权最小二乘准则构建模型的目标函数, 它会对拥有更高频率的元素赋予更大的权重, 也是文献[26]提出的用来解决隐性反馈问题的一种方案。为了使模型既能取得好的性能, 又可以降低计算量, 在更新参数的过程中加入引导程序, 采用分块更新参数。使用这种处理方式可以有效地降低计算时间, 且能更快地使模型收敛。

原始目标函数为:

$$O_{\text{WMF}} = \min \sum_{u,i} w_{ui} (r_{ui} - \hat{r}_{ui})^2 \quad (9)$$

式中, w_{ui} 为与用户 u 在 POI i 的历史签到记录频率有关的信任度, 更多的签到记录意味着更大的 w_{ui} 。由于用户-POI 签到矩阵中的 0 元素可能是未知情况, 所以只有是 1 的元素才加入优化。

在本文的预测公式(8)下, 目标函数为:

$$O_{\text{GEO}} = \min \sum_{u,i} w_{ui} (r_{ui} - \hat{r}_{ui})^2 + \lambda \|\theta\|^2 \quad (10)$$

式中, λ 为正则化系数; θ 表示所有需要更新的参数向量。

对于式(10), 本文采用随机梯度下降法(stochastic gradient descent, SGD)更新参数:

$$\theta \leftarrow \theta - s \frac{\partial O_{\text{GEO}}}{\partial \theta} \quad (11)$$

式中, s 为学习率。

3 模型推荐性能实验

3.1 实验数据集与参数设置

本文使用真实签到 Foursquare 数据集^[13]来测试所提出的 CC-MF 模型的性能。该数据集包括 2 321 个用户、5 596 个 POI, 总共 194 108 个签到记录, 数据稀疏度为 0.81%。使用机器学习中常见的网格搜索法调整超参, 经过多次重复实验, 设置参数学习率 s 为 0.001, 正则化系数 λ 为 0.001, 空间距离关系影响系数 α_1 为 1.8, 考虑空间拓扑关系的空间距离影响系数 α_2 为 0.004 5, 生活圈半径 d 为 1.5 km, 地理邻居个数 k 为 10。实验使用计算机处理器(CPU)为 i7-8700, 内存为 32 GB。

实验中对数据集进行切片, 按时间顺序的前 80% 作为训练集, 后 20% 作为测试集。由式(8)计算出所有隐性变量, 之后根据计算得到的 \hat{r}_{ui} 按照大小顺序, 为每个用户推荐出 k 个 POI。

3.2 评价指标

评价指标取决于推荐任务,CC-MF模型的任务是推荐给用户Top- k 型推荐^[4]。本文使用精确率 P 和召回率 R 这两种评价指标来评估CC-MF模型的性能。精确率是指推荐正确的POI个数与推荐个数 k 的比值,召回率是指推荐正确的POI个数与用户在测试集中实际去过的POI个数的比值。

$$P = \frac{1}{\text{card}(U)} \sum_{u=1}^{\text{card}(U)} \frac{|S_u \cap M_u(k)|}{k} \times 100\% \quad (12)$$

$$R = \frac{1}{\text{card}(U)} \sum_{u=1}^{\text{card}(U)} \frac{|S_u \cap M_u(k)|}{\text{card}(S_u)} \times 100\% \quad (13)$$

式中, U 表示用户集合; S_u 表示用户在测试集中访问过的POI集合; $M_u(k)$ 表示CC-MF推荐算法得出的 k 个POI集合。

3.3 计算复杂度

假设本文使用 $m \times n$ 的用户签到矩阵,包含 m 个用户和 n 个POI。矩阵分解过程的计算复杂度为 $O(mn^2)$ 。计算空间距离因素影响的计算复杂度为 $O(n)$,考虑空间拓扑关系的空间距离影响的计算复杂度为 $O(n)$,本文采用离线计算,该过程的计算复杂度将不被算入到总计算复杂度中。因此,本文提出的推荐算法的计算复杂度为 $O(mn^2)$ 。

3.4 实验结果及分析

本文实验采用的基线方法包括:基于用户的协同过滤(user-based collaborative filtering, UCF)^[38],UCF是一种经典的基于用户协同过滤推荐方法;分解个性化马尔科夫链(factorized personalized Markov chain, FPMC)^[16],FPMC是将用户偏好转化为个性化的马尔科夫链;基于排序的地理因子分解方法(ranking based geographical factorization method, Rank-GeoFM)^[28],Rank-GeoFM是融合空间因素的MF模型,且运用了一种排序优化算法;特征空间分离因子分解模型(feature-space separated factorization model, FSS-FM/T)^[4],FSS-FM/T是融入空间因素的WMF模型,运用排序优化算法;基于矩阵分解的上下文感知POI推荐模型(general matrix factorization model, UCGSMF)^[23],UCGSMF是融合地理信息的矩阵分解上下文感知的推荐算法。在Foursquare数据集上使用本文提出的CC-MF方法与上述几种基线方法进行了对比实验,且测试了空间距离因素和考虑空间拓扑关系的空间距离因素的影响,实验结果分别如图3和图4所示。

由于模型是根据得分排序得到的推荐序列,所以排在最前面的几个推荐POI即推荐个数为5时,推荐算法得出的结果是最重要的^[28]。

1) 对比不同基线方法和CC-MF模型的推荐性能。由图3(a)可以得出,相较几种基线方法,CC-MF模型在 $k=5, 10, 20$ 时,推荐结果的精确率比各基线方法的精确率都要高,性能提升显著,这是因为CC-MF模型加入了考虑空间拓扑关系的空间距离因素的影响。在 $k=5$ 时,CC-MF模型的精确率为11.52%,比表现较好的FSS-FM/T提高了约10.33%;除FSS-FM/T以外的其他基线方法,精确率提升的幅度均超过15%。在 $k=10$ 或 20 时,CC-MF模型的精确率仍优于其他基线算法。由图3(b)可以得出,在 $k=5$ 时,CC-MF模型的召回率比其他基线方法提升约4%~9%,在 $k=10$ 或 20 时结果相似,即CC-MF模型的召回率也优于其他基线方法。这说明了空间关系对POI推荐有影响,且融入空间拓扑关系的空间距离因素能够更好地提升推荐算法的性能。

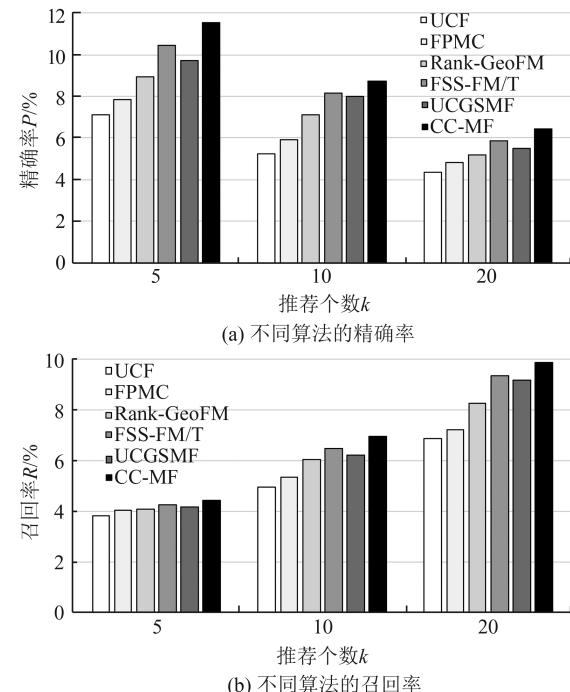


图3 使用Foursquare数据集的推荐性能
Fig.3 Recommendation Performance Using the Foursquare Data Sets

2) 对比不同因素的推荐性能影响。如图4所示,CC-MF模型同时计算空间拓扑因素和空间距离因素的影响,CC-MF/G只计算空间距离影响,CC-MF/D只计算考虑空间拓扑关系的空间距离因素影响,CC-MF/G/D为空间拓扑因素和

空间距离因素都不考虑。由图 4 可以得出,加入空间距离因素和考虑空间拓扑关系的空间距离因素都能提高推荐系统的性能。在 $k=5$ 时,对比 CC-MF/G/D,加入空间距离因素的 CC-MF/G 的精确率提升了约 5.38%,召回率提升了约 7.73%;加入考虑空间拓扑关系的空间距离因素的 CC-MF/D 的精确率提升了约 11.17%,召回率提升了约 11.08%;加入空间距离因素和考虑空间拓扑关系的空间距离因素的 CC-MF 的精确率提升了约 14.86%,召回率提升了约 13.66%。在 $k=10$ 或 20 时结果类似,此处不再赘述。

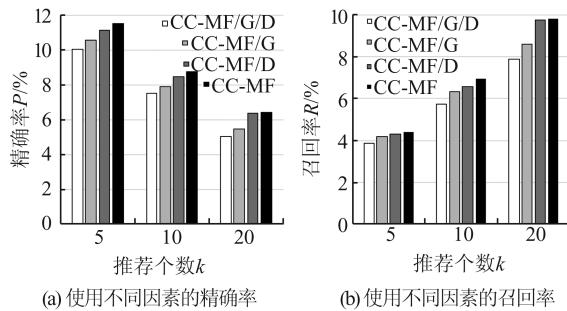


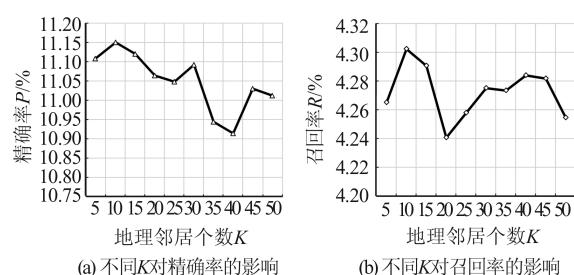
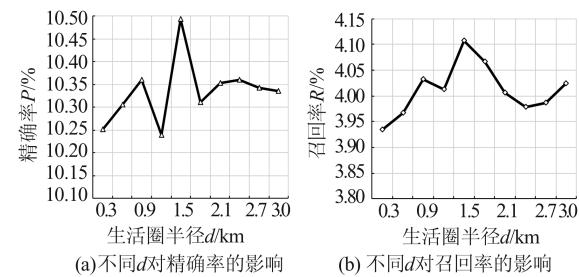
图 4 不同因素的推荐性能影响

Fig.4 Effects of Different Features on the Recommendation Performance

3) 对比不同参数 K 和 d 的影响。图 5 和图 6 展示了 POI 地理邻居个数 K 和生活圈半径 d 取不同值时对模型推荐性能的影响。可以看出, K 和 d 不同的取值对模型推荐性能的影响很大, 合适的取值能有效提升模型的推荐性能。从图 5 可以看出, 随着 K 的增大, POI 推荐的性能都有明显提升; 当 K 到达阈值时, 推荐性能不再提升并逐渐降低。综合考虑 POI 推荐的精确率和召回率, 当 $K=10$ 时, 模型性能达到最优。这表示地理邻居的个数不能过多, 也不能过少, 当地理邻居过少时, 可能会使 K 个地理邻居中没有包含真正相邻的空间实体, 导致与现实不符的情况, 不能很好地显示地理邻居的影响且产生偏差; 当地理邻居过多时, 可能会使地理邻居的影响差异缩小, 以至于不能产生良好的效果。从图 6 中也可以看出, d 的选择不能太大, 也不能太小。综合考虑 POI 推荐的性能, 当 $d=1.5 \text{ km}$ 时, 模型性能达到最优。这表示 d 过大或过小都不能很好地表示用户的生活圈范围。

通过多次实验, 得到如下结论:

1) 与多种基线方法对比, 从精确率和召回率上看, 本文提出的 CC-MF 模型比当前常用的几种模型的推荐效果都好。

(a) 不同 K 对精确率的影响 (b) 不同 K 对召回率的影响Fig.5 Effects of Different K on the Recommendation Performance(a) 不同 d 对精确率的影响 (b) 不同 d 对召回率的影响Fig.6 Effects of Different d on the Recommendation Performance

2) 考虑空间拓扑关系的空间距离影响能够进一步提升模型的性能。

4 结语

针对数据稀疏和隐性反馈使用问题, 本文提出了 POI 推荐模型 CC-MF, 首先从关系角度衡量用户、POI 间的相互关系, 并融入空间拓扑关系以及空间距离关系影响, 能够有效模拟空间关系的作用, 并且具有较好的解释性和拓展性。在 Foursquare 签到数据集上进行实验, 结果表明: (1) CC-MF 模型的性能优于当前常用的 POI 推荐算法; (2) 将空间关系细分为空间距离关系和空间拓扑关系并融入模型, 可以更有效地提升推荐算法的性能。在未来的工作中, 尝试将时间、社交关系等因素加入到 CC-MF 模型中, 研究时空、社交关系等因素的影响模式。

参 考 文 献

- [1] Bobadilla J, Ortega F, Hernando A, et al. Recommender Systems Survey [J]. *Knowledge-Based Systems*, 2013, 46:109-132
- [2] Lu J, Wu D, Mao M, et al. Recommender System Application Developments: A Survey [J]. *Decision Support Systems*, 2015, 74:12-32
- [3] Zhang C, Wang K. POI Recommendation Through

- Cross-Region Collaborative Filtering [J]. *Knowledge and Information Systems*, 2016, 46 (2) : 369-387
- [4] Cai L, Xu J, Liu J, et al. Integrating Spatial and Temporal Contexts into a Factorization Model for POI Recommendation [J]. *International Journal of Geographical Information Science*, 2018, 32 (3) : 524-546
- [5] Ning Jinsheng, Wu Xuequn, Liu Ziyao. Multi-user Location Recommendation Considering Road Accessibility and Time-Cost [J]. *Geomatics and Information Science of Wuhan University*, 2019, 44(5):633-639(宁津生,吴学群,刘子尧. 顾及道路通达性和时间成本的多用户位置推荐[J]. 武汉大学学报·信息科学版,2019,44(5):633-639)
- [6] Zhu Jinghua, Ming Qian. POI Recommendation by Incorporating Trust-Distrust Relationship in LBSN [J]. *Journal on Communications*, 2018, 39 (7) : 157-165(朱敬华,明骞. LBSN中融合信任与不信任关系的兴趣点推荐 [J]. 通信学报, 2018, 39 (7): 157-165)
- [7] Liao Guoqiong, Jiang Shan, Zhou Zhiheng, et al. Double Fine-Granularity POI Recommendation on Location-Based Social Networks [J]. *Journal of Computer Research and Development*, 2017, 54 (11) : 2 600-2 610(廖国琼,姜珊,周志恒,等. 基于位置社会网络的双重细粒度兴趣点推荐 [J]. 计算机研究与发展, 2017, 54(11): 2 600-2 610)
- [8] Baral R, Li T. Exploiting the Roles of Aspects in Personalized POI Recommender Systems [J]. *Data Mining and Knowledge Discovery*, 2018, 32 (2) : 320-343
- [9] Wang Yandong, Li Hao, Wang Teng, et al. The Mining and Analysis of Emergency Information in Sudden Events Based on Social Media[J]. *Geomatics and Information Science of Wuhan University*, 2016, 41(3):290-297(王艳东,李昊,王腾,等. 基于社交媒体的突发事件应急信息挖掘与分析[J]. 武汉大学学报·信息科学版,2016,41(3):290-297)
- [10] Yu Wenhao, Ai Tinghua, Yang Min, et al. Detecting “Hot Spots” of Facility POIs Based on Kernel Density Estimation and Spatial Autocorrelation Technique [J]. *Geomatics and Information Science of Wuhan University*, 2016, 41(2):221-227(禹文豪,艾廷华,杨敏,等. 利用核密度与空间自相关进行城市设施兴趣点分布热点探测[J]. 武汉大学学报·信息科学版,2016,41(2):221-227)
- [11] Hu L, Sun A, Liu Y. Your Neighbors Affect Your Ratings: On Geographical Neighborhood Influence to Rating Prediction [C]//The 37th International ACM SIGIR Conference on Research and Development in Information Retrieval, Gold Coast Queensland, Australia, 2014
- [12] Ye M, Yin P, Lee W-C, et al. Exploiting Geographical Influence for Collaborative Point-of-Interest Recommendation [C]//The 34th International ACM SIGIR Conference on Research and Development in Information Retrieval, Beijing, China, 2011
- [13] Yuan Q, Cong G, Ma Z, 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
- [14] Miller H J. Tobler’s First Law and Spatial Analysis [J]. *Annals of the Association of American Geographers*, 2004, 94(2): 284-289
- [15] Si Y, Zhang F, Liu W. An Adaptive Point-of-Interest Recommendation Method for Location-Based Social Networks Based on User Activity and Spatial Features [J]. *Knowledge-Based Systems*, 2019, 163: 267-282
- [16] Li X, Jiang M, Hong H, et al. A Time-Aware Personalized Point-of-Interest Recommendation via High-Order Tensor Factorization[J]. *ACM Transactions on Information System*, 2017, 35(4):1-23
- [17] Wang Nan, Li Jinbao, Liu Yong, et al. TPR-TF: Time-Aware Point of Interest Recommendation Model Based on Tensor Factorization [J]. *Journal of Jilin University (Engineering and Technology Edition)*, 2019, 49(3): 920-933(王楠,李金宝,刘勇,等. TPR-TF:基于张量分解的时间敏感兴趣点推荐模型 [J]. 吉林大学学报(工学版), 2019, 49(3): 920-933)
- [18] Ye M, Liu X, Lee W C. Exploring Social Influence for Recommendation a Generative Model Approach [C]//The 35th International ACM SIGIR Conference on Research and Development in Information Retrieval, Portland, USA, 2012
- [19] Guo J, Zhang W, Fan W, et al. Combining Geographical and Social Influences with Deep Learning for Personalized Point-of-Interest Recommendation [J]. *Journal of Management Information Systems*, 2018, 35(4): 1 121-1 153
- [20] Wu R, Luo G, Yang Q, et al. Learning Individual Moving Preference and Social Interaction for Location Prediction [J]. *IEEE Access*, 2018, 6: 10 675-10 687
- [21] Gao S. Spatiotemporal Analytics for Exploring Human Mobility Patterns and Urban Dynamics in the Mobile Age [J]. *Spatial Cognition and Computation*

- tion, 2014, 15(2): 86-114
- [22] Gao R, Li J, Du B, et al. A Synthetic Recommendation Model for Point-of-Interest on Location-Based Social Networks: Exploiting Contextual Information and Review [J]. *Journal of Computer Research and Development*, 2016, 53(4): 752-763
- [23] Peng Hongyuan, Jin Yuanyuan, Lü Xiaoqiang, et al. Context-Aware POI Recommendation Based on Matrix Factorization [J]. *Chinese Journal of Computers*, 2019, 42(8): 1 797-1 811(彭宏伟, 靳远远, 吕晓强, 等. 一种基于矩阵分解的上下文感知 POI 推荐算法 [J]. 计算机学报, 2019, 42(8): 1 797-1 811)
- [24] Xing S, Liu F, Zhao X, et al. Points-of-Interest Recommendation Based on Convolution Matrix Factorization [J]. *Applied Intelligence*, 2018, 48 (8) : 2 458-2 469
- [25] Pan R, Zhou Y, Cao B, et al. One-class Collaborative Filtering [C]// The 8th IEEE International Conference on Data Mining, Pisa, Italy, 2008
- [26] Hu Y, Koren Y, Volinsky C. Collaborative Filtering for Implicit Feedback Datasets [C]// The 8th IEEE International Conference on Data Mining, Pisa, Italy, 2008
- [27] Liang D, Charlin L, Mcinerney J, et al. Modeling User Exposure in Recommendation [C]//The 25th International Conference on World Wide Web, Montréal Québec, Canada, 2016
- [28] Li X, Cong G, Li X L, et al. Rank-GeoFM: A Ranking Based Geographical Factorization Method for Point of Interest Recommendation [C]//The 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, Santiago, Chile, 2015
- [29] 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, New York, USA, 2014
- [30] Lee J, Kim S, Lebanon G, et al. LLORMA: Local Low-Rank Matrix Approximation [J]. *The Journal of Machine Learning Research*, 2016, 17 (1) : 442-465
- [31] Hao M. An Experimental Study on Implicit Social Recommendation [C]//The 36th International ACM SIGIR Conference on Research and Development in Information Retrieval, Dublin, Ireland, 2013
- [32] He X, Liao L, Zhang H, et al. Neural Collaborative Filtering [C]//The 26th International Conference on World Wide Web, Perth, Australia, 2017
- [33] Zhang Z, Zou C, Ding R, et al. VCG: Exploiting Visual Contents and Geographic Influence for Point-of-Interest Recommendation [J]. *Neurocomputing*, 2019, 357: 53-65
- [34] Qian T, Liu B, Nguyen Q V H, et al. Spatiotemporal Representation Learning for Translation-Based POI Recommendation [J]. *ACM Transactions on Information System*, 2019, 37(2):1-24
- [35] Wang H, Shen H, Ouyang W, et al. Exploiting POI-specific Geographical Influence for Point-of-Interest Recommendation [C]//The 27th International Joint Conference on Artificial Intelligence, Stockholm, Sweden, 2018
- [36] Koren Y, Bell R, Volinsky C. Matrix Factorization Techniques for Recommender Systems [J]. *Computer*, 2009, 42(8): 30-37
- [37] Chai Yanwei, Zhang Xue, Sun Daosheng. A Study on Life Circle Planning Based on Space Time Behavioural Analysis: A Case Study of Beijing [J]. *Urban Planning Forum*, 2015(3): 61-69(柴彦威, 张雪, 孙道胜. 基于时空行为的城市生活圈规划研究:以北京市为例 [J]. 城市规划学刊, 2015(3): 61-69)
- [38] Zhou D, Wang B, Rahimi S M, et al. A Study of Recommending Locations on Location-based Social Network by Collaborative Filtering [C]//The 25th Canadian Conference on Artificial Intelligence, Toronto, Canada, 2012

Integrating Spatial Relationship into a Matrix Factorization Model for POI Recommendation

WEI Haitao^{1,2} LI Ke^{1,2} HE Xiaohui^{1,2} TIAN Zhihui^{1,2}

1 School of Geo-Science and Technology, Zhengzhou University, Zhengzhou 450052, China

2 Joint Laboratory of Eco-Meteorology, Zhengzhou University, Chinese Academy of Meteorological Sciences, Zhengzhou 450052, China

Abstract: Objectives: Point of interest (POI) recommendation is the prevalent personal service in location-based social network(LBSN), and aims to provide personalized recommendation services by using the information carried by LBSN. The utilization of spatial relationship information as the side information supplies a chance to product better POI recommend. However, thousands of users and POIs in the LBSN make the user-POI check-in matrix very large and sparse. In addition, check-in record data is typical implicit feedback data, which cannot directly reflect the user's preference. To tackle the aforementioned challenges, we propose a relational matrix factorization model based on cooperative competition matrix factorization (CC-MF). Methods: The CC-MF model can simulate the relationship between users and POIs, and divides spatial relationships into spatial distance relationship and spatial topological relationship. In order to alleviate the problem of data sparsity, the model excavates the spatial relationships among POIs, POIs and users by integrating spatial relationships. Firstly, we use nonlinear function to establish the spatial distance relationship between users and POIs, which can connect the relationship between users and POIs. Then, k -nearest neighbor (k NN) algorithm is used to calculate the geo-neighbors of POI by considering the spatial distance factor of spatial topological relationship, which can further alleviate the sparsity of data. Finally, the spatial relationship is integrated into the matrix factorization model. Meanwhile, the weighted least square method is used as the objective function of the CC-MF model to relieve the implicit feedback problem. Experiments are carried out on the real-world check-in Foursquare datasets. We test the recommendation performance of the proposed model and baseline methods, and analyze the crucial influence of different spatial relationships on POI recommendation. The precision and recall are used as evaluation metrics. Results: The results show that: (1) The CC-MF model significantly improves the precision and recall of the recommendation results. (2) Considering the spatial distance factor of the spatial topological relationship can further improve the performance of the recommendation system. Conclusions: Therefore, CC-MF model can make use of spatial relationship better and more comprehensive. The proposed CC-MF model has better scalability and better interpretability, and can alleviate the problems of data sparsity and implicit feedback usage.

Key words: point of interest (POI) recommendation; location-based social network(LBSN); matrix factorization; spatial relationship; space distance

First author: WEI Haitao, PhD, lecturer, specializes in geographic information and smart city. E-mail: zzu_wei@163.com

Corresponding author: HE Xiaohui, PhD, professor. E-mail: hexh@zju.edu.cn

Foundation support: The National Key Research and Development Program of China(2018YFB0505000); the Key Research and Development and Promotion Special Projects of Henan Province (Science and Technology Tackling Key Problems) (192102210124).

引文格式: WEI Haitao, LI Ke, HE Xiaohui, et al. Integrating Spatial Relationship into a Matrix Factorization Model for POI Recommendation[J]. Geomatics and Information Science of Wuhan University, 2021, 46(5):681-690.DOI:10.13203/j.whugis20200355(魏海涛,李柯,赫晓慧,等.融入空间关系的矩阵分解POI推荐模型[J].武汉大学学报·信息科学版,2021,46(5):681-690.DOI:10.13203/j.whugis20200355)