

基于类别转移加权张量分解模型的兴趣点分区推荐

李胜* 刘桂云 何熊熊

(浙江工业大学信息工程学院 杭州 310023)

摘要: 基于位置社交网络的兴趣点(POI)推荐是人们发现有趣位置的重要途径,然而,现实中用户在不同区域的地点偏好侧重的差异,加之高维度的历史签到信息,使得精准而又个性化的POI推荐极富挑战性。对此,该文提出一种新型的基于类别转移加权张量分解模型的兴趣点分区推荐算法(WTD-PR)。通过结合用户连续行为和时间特征,来充分利用用户的历史访问信息,从而得到类别转移权重因子;接着改进用户-时间-类别张量模型,在此张量中加入类别转移权重,预测用户的喜好类别;最后,根据用户的历史访问区域划分出本地和异地,并基于用户的当前位置找出推荐区域范畴,进而引入位置因素和社交因素,结合候选类别作兴趣点分区推荐。通过在真实数据集上进行对比实验,实验结果表明,所提算法不仅具有通用性,而且在推荐性能上也优于其他对比算法。

关键词: 兴趣点推荐; 张量分解; 类别转移权重; 分区推荐

中图分类号: TP391

文献标识码: A

文章编号: 1009-5896(2022)01-0203-08

DOI: 10.11999/JEIT200934

A Recommendation Method for Point-of-Interest Partition Based on Category Transfer Weighted Tensor Decomposition Model

LI Sheng LIU Guiyun HE Xiongxiang

(School of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China)

Abstract: Point-Of-Interest (POI) recommendation in location-based social networks is an important way for people to find interesting locations. However, in reality, both the various user preference of locations in different regions and the high-dimensional historical check-in information make accurate and personalized POI recommendations extremely challenging. In this regard, a new type of recommendation algorithm for point-of-interest Partition Recommendation based on a category transfer Weighted Tensor Decomposition (WTD-PR) model is proposed. The proposed algorithm makes full use of the user's historical visit information by combining the user's continuous behavior and time characteristics to obtain the category transfer weight factor; Then, by improving the user-time-category tensor model and adding the category transfer weight to the tensor to predict the user's preference category; Finally, the local and remote locations are divided according to the user's historical access area, and the recommended areas are found based on the user's current location. After that, location and social factors are introduced and combined with the candidate categories to make the recommendation of points of interest. Through comparative experiments on real data sets, the proposed algorithm is proved not only to be universal, but also superior to other comparison algorithms in terms of recommendation performance.

Key words: Point-Of-Interest (POI) recommendation; Tensor decomposition; Category transfer weight; Partition recommendation

1 引言

随着社会信息化进程的不断加速,位置信息服

务在社会各行各业中越发普及,这使得基于位置的社交网络(Location Based Social Networks, LBSN)服务成为一项具有重要价值的应用,比如Foursquare、去哪儿网等都是提供类似服务为主的产品。POI推荐是基于位置的社交网络服务研究中的一项重要内容,它不仅在一定程度上解决大数据时代用户面临的信息过载问题,而且能够帮助商家快速找到目标用户,实现精准营销^[1]。

在如此良好的发展机遇面前,推荐系统面临着

收稿日期: 2020-11-02; 改回日期: 2021-04-19; 网络出版: 2021-08-26

*通信作者: 李胜 shengli@zjut.edu.cn

基金项目: 国家自然科学基金(61873239, 61675183), 浙江省重点研发计划(2020C03074)

Foundation Items: The National Natural Science Foundation of China (61873239, 61675183), The Key R&D Projects in Zhejiang Province (2020C03074)

许多挑战,比如用户的POI决策过程建模困难、数据稀疏等。传统的推荐算法主要是协同过滤算法,在此算法基础上衍生出了许多推荐模型,例如基于POI排名的因式分解^[2]、矩阵分解^[3]等。现有研究多利用高阶张量模型代替传统的用户定位矩阵,这有助于解决用户动态签到随时间变化的时间依赖性^[4]。但高维张量分解会增加系统的负担,如何权衡信息的维度成了研究的一大难题。

在推荐系统中,用户的显式偏好是结合用户历史行为和地点信息进行分析计算得到的^[3],如偏好的时间、类别、位置标签等。除此之外,还有一部分隐式信息需要挖掘,这些隐式反馈往往存在于交互关系中,如连续行为偏好^[5]、隐式社交关系^[6]等。为此,一些研究会把隐式反馈转化为权重以此降低矩阵维度^[6,7],或者把从信息中提取的影响因素纳入模型中减少数据稀疏^[8,9]。

事实上,使用隐式反馈来探索POI推荐是具有挑战性的,因为学习过程需要一种有效的策略。例如在挖掘用户连续行为偏好时,利用不同层次上的类别转换来对用户的偏好转换进行建模^[5],能减缓利用POI之间的转换进行建模^[10]带来的稀疏性。但是由于用户的移动性,推荐区域多有不同,此类模型在捕获POI的语义相关性上需要考虑更全面的影响因素才能给游客带来更多便利。

在一个不熟悉的区域中,旅行者和POI之间的交互是非常稀少的,但其兴趣在家乡与所到城市之间存在着漂移和转移的现象。将兴趣漂移和兴趣转移结合起来,利用用户跨城市的访问行为可以改善所到城市的POI推荐^[11,12]。其次,通过引入POI的显式信息^[4,8]或者借鉴朋友的偏好影响^[6],也可以缓解所到城市用户POI交互的稀疏性问题。那么,为了有效地利用社会和空间相关性,可以把兴趣的漂移和转移现象同显式和隐式信息中提取的影响因素结合起来,根据当前位置进行分区域推荐。

综上所述,对推荐系统的现有研究可以在模型、类别转化、影响因素以及分区域推荐4个方面进行突破。本文提出一种新型的基于加权张量分解模型的兴趣点分区推荐算法,主要贡献有以下3点:(1)对签到信息中蕴含的显式信息和隐式信息进行用户标签和地点标签的划分,改善时间、类别和朋友特征的提取方法,更好地把影响因子融合到不同的推荐区域中。(2)对张量构造进行改进,把隐式反馈中的用户连续行为偏好与类别信息结合作为模型的权重信息,加入到用户-时间-类别张量中,此做法不仅综合了用户、时间和类别这3维信息,又充分考虑到历史访问信息的权重影响。(3)基于用

户的移动性特征,把候选类别与提取的特征相融合进行分区POI推荐:依据用户当前位置与常驻地之间的距离把推荐场景细粒度化,划分出本地推荐和异地推荐,再匹配不同的影响因素做不同区域的推荐。发挥各类影响因素特点的同时,也使得预测用户的下一步偏好更精准。

2 预备知识

2.1节分析如何改善从显式和隐式信息中抽取特征的方法。2.2节是文章整体框架的图示。

2.1 相关性分析

2.1.1 时间特征分析

人们的行为在特定时间会有特定变化,即实时性,而且具有持续性^[13],因此在兴趣点推荐中加入时间分析是必要的。大多数文章对于时间区域的划分为均匀24段^[14],这忽略了用户出行习惯的不确定性及事件的持续性,而且会额外增加数据维度。从种类信息中选取6个具有时间象征性的类别,如图1(a)横轴坐标所示,分别对应为:1→专业场所(professional & other places)、2→办公楼(office)、3→博物馆(museum)、4→体育场(stadium)、5→户外休闲(outdoors & recreation)、6→夜生活场所(night-life spot),可以看出,用户在工作日和休息日的类别访问占比有所不同,这说明在工作日和休息日,用户的兴趣侧重不同。由Foursquare数据集中用户的访问分布可知,人们的活动轨迹大多集中在3个时间段,示例如图1(b)所示。本文把用户每天的时间活跃范围划分成3个时间段:“7:00~15:00”,“15:00~21:00”,“21:00~7:00(次日)”。综上所述,本文根据工作日、休息日和3个时间段划分6个时间标签,既减少数据稀疏性,又具有兴趣针对性。比如用户1会在休息日7:00~15:00时间段运动,用户2会在工作日21:00~7:00(次日)时间段现身酒吧。

2.1.2 类别特征分析

本文将类别信息和连续行为结合作为类别预测模型的权重信息,在捕获用户兴趣的同时可以简化信息维度。在连续POI推荐的问题上,以往的研究多使用1阶马尔科夫链来建模^[10],但此模型仅依赖用户最后访问的兴趣点,不考虑其历史访问。而实际中,每个用户有其独特的历史访问序列,对新兴趣点的访问概率也参考历史访问信息。因此本文引用一种基于高阶序列的模型——加权马尔科夫链模型来研究用户的连续行为,为此构造基于加权马尔科夫链的类别转移权重,定义为 \mathbf{W}^{TC} ,它将考虑签到历史中所有签到点对新兴趣点访问概率的影响。

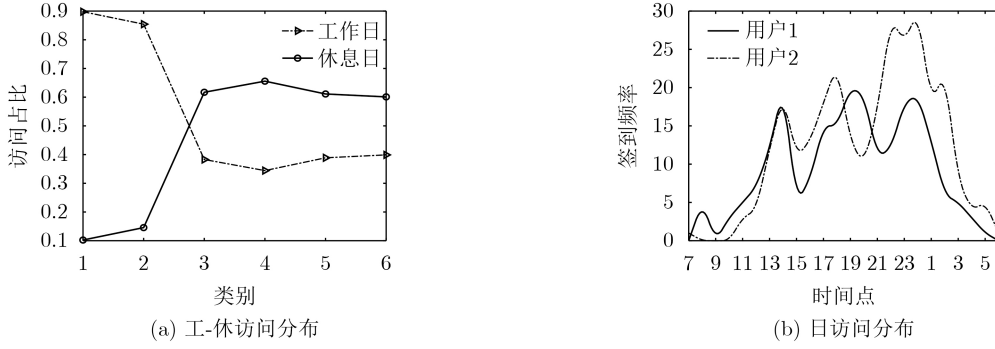


图1 时间特征划分示例

加权马尔科夫链是在转移图的基础上构造的，而基于类别访问序列的转移图相对于地点转移图更具有泛化度^[9]。在类别转移图中，把一个节点的转出频率定义为Cout，转移频率定义为CTrans。如图2所示，对于酒店这个节点，其转出频率有8次(Cout为8)，转入餐饮节点5次，转入大型交通工具节点2次(CTrans为7)。给出目标用户的类别访问序列，用户对新兴趣点的转移权重如式(1)所示

$$\begin{aligned}
 \mathbf{W}^{\text{TC}}(C_{n+1} | C_u) &= \sum_{i=1}^n f(C_{n+1}, C_i, n - i + 1) \\
 &= \sum_{i=1}^n \frac{w(C_i) \cdot \text{PTrans}(C_i \rightarrow C_{n+1})}{\sum_{i=1}^n w(C_i)}
 \end{aligned} \tag{1}$$

其中，函数 $f(C_{n+1}, C_i, n - i + 1)$ 是类别序列中访问点对 $\mathbf{W}^{\text{TC}}(C_{n+1} | C_u)$ 值的加权影响； $w(C_i) = 2^{-\alpha(n-t)}$ 为衰减参数 $\alpha(\alpha > 0)$ 的衰减权重； $\text{PTrans}(C_i \rightarrow C_j)$ 是基于转移图的转移概率，如式(2)所示

$$\text{PTrans}(C_i \rightarrow C_j) = \frac{\text{CTrans}(C_i, C_j)}{\text{Cout}(C_i)} \tag{2}$$

2.1.3 朋友特征分析

随着网络科技的发展，人们社交中“朋友”间的交互不断地丰富着个人的兴趣偏好，因此隐式社交关系在推荐系统中应用的越发频繁，且用户与“朋友”间的信任关系有助于提高推荐系统的性能^[15]。“朋友”的签到频率 z 遵循幂律分布^[6](Powerlaw Distribution, PD)，其概率密度函数如式(3)所示

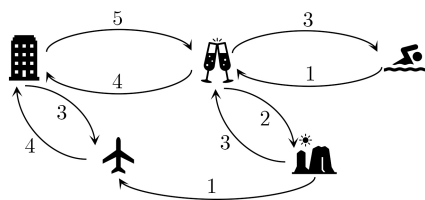


图2 类别转移图示例

$$f(z) = (\gamma - 1)(z + 1)^{-\gamma}, z \geq 0, \gamma \geq 1 \tag{3}$$

其中， γ 是训练出的参数。因社会影响力会随着朋友的签到频率增加而增加，所以最终通过 F 的累积分布函数得到社会关系的影响，如式(4)所示

$$\begin{aligned}
 F(z_{u,l}) &= \int_0^{z_{u,l}} f(z) dz \\
 &= 1 - (z_{u,l} + 1)^{1-\gamma}
 \end{aligned} \tag{4}$$

其中， $z_{u,l}$ 表示用户朋友在某个 l 上的签到频率。

为适应不同的推荐场景，本文把“朋友”群体划分为两类：邻居朋友(M_f^n)和兴趣朋友(M_f^h)，经过模型的训练，得到相应的影响权值为 F_n 和 F_h 。

邻居朋友指用户常驻地周围的人群，其具有潮流带动性。在本地推荐中结合的社交关系就是此类朋友，该集合的划分如式(5)所示

$$M_f^n \in \text{dist}(u_i, u_j) \cap \text{Sim}(u_i, u_j) \tag{5}$$

其中， dist 是用户 i 和用户 j 常驻地之间的距离， Sim 是由余弦相似度得到用户 i 和用户 j 之间的关联度。由式(5)划分邻居朋友集合再推出 $z_{u,l}$ ，通过式(4)计算出 F_n 值。

兴趣朋友指社交网络中兴趣相投的人群，其位置分布广，但在同一时间段有相似的兴趣。兴趣朋友的定义是为适应异地推荐的场景，如式(6)所示

$$M_f^h \in \text{time}(u_i, u_j) \cap \text{cat}(u_i, u_j) \tag{6}$$

其中， time 表示用户 i 和用户 j 出访的时间相似性， cat 是用户 i 和用户 j 历史访问的类别信息的交集程度。由该关系集合构造 $z_{u,l}$ ，代入式(4)求得 F_h 。

2.1.4 位置特征分析

有数据表明，人们短期内的活动范围不会太大，大多呈现多中心高斯分布^[16]，且有一个范围阈值 θ ，两地点间距离越近，被用户访问的概率就越高。在距离相关性中，两位置间的相关值 Val 如式(7)所示

$$\text{Val}_{i,j} = \frac{1}{\text{MIN}_{l_i \in L} \text{Dist}(l_j, l_i)}, \text{Dist} \leq \theta \tag{7}$$

地点流行度是指一个地理位置的访问频率，频

率越高则流行度越大。流行度跟时间息息相关^[17],因此,本文在Pop上加入时效性,如式(8)所示

$$\text{Pop}(l, t) = \frac{\text{Times}(l, t)}{\sum_{\tilde{l} \in L} \text{Times}(\tilde{l}, t)} \quad (8)$$

2.2 框架构造

本文的整体框架分为3部分,如图3所示。

3 类别预测

本文构造用户-时间-类别3维张量,并且从隐式反馈信息中提取用户连续行为偏好,将其作为类别转移权重加入张量分解中,然后优化分解过程得到类别预测列表。

3.1 张量构造

张量构造可解决用户动态签入行为的时间依赖性^[16],本文根据用户在某个时间场景下的特定喜好,分析出时间类别对,构造用户-时间-类别张量 \mathbf{R} 。

3.2 张量分解模型

本文采用张量分解中的Tucker分解,其原理是通过分解高维张量,生成稠密的预测张量来逼近原始张量,填补空缺值,从而生成推荐,如图4所示。

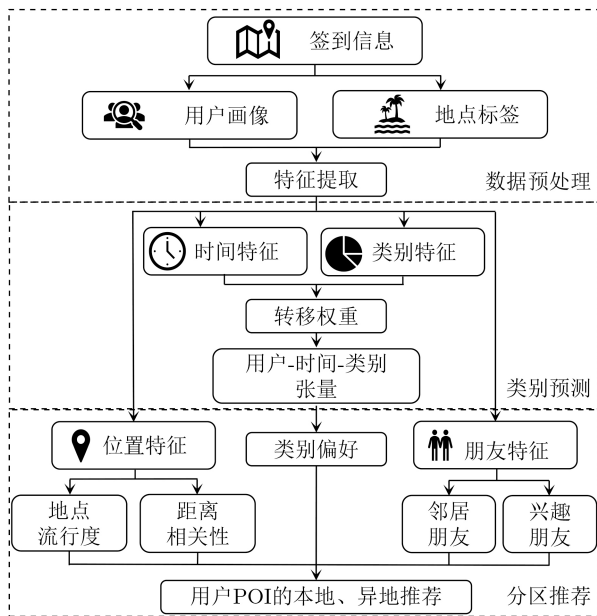


图3 框架图

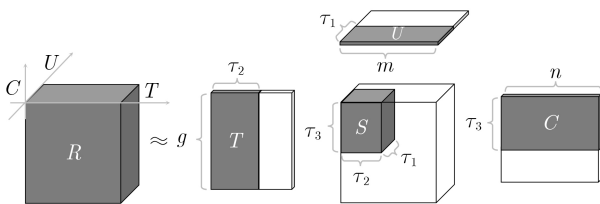


图4 截断Tucker分解: 秩-(τ_1, τ_2, τ_3)近似

(1)进行高阶奇异值分解(Higherorder Singular Value Decomposition, HOSVD),初始化张量:

首先把 \mathbf{R} 分解为3个因子矩阵 $\mathbf{U}, \mathbf{T}, \mathbf{C}$, 其中 $\mathbf{U} \in \mathbf{R}^{m \times \tau_1}, \mathbf{T} \in \mathbf{R}^{g \times \tau_2}, \mathbf{C} \in \mathbf{R}^{n \times \tau_3}$ 。再构造核心张量 $\mathbf{S} \in \mathbf{R}^{\tau_1 \times \tau_2 \times \tau_3}$ 去控制因子矩阵之间的交互。最后利用mode乘积去生成初始化张量,如式(9)所示

$$\hat{\mathbf{R}} = \mathbf{S} \times_{\mathbf{U}} \mathbf{U} \times_{\mathbf{T}} \mathbf{T} \times_{\mathbf{C}} \mathbf{C} \quad (9)$$

(2)模型训练采用随机梯度下降算法(Stochastic Gradient Descent, SGD)进行迭代优化,首先定义目标函数如式(10)所示

$$L(\mathbf{Y}, \mathbf{R}) = \frac{1}{2} \left\| \mathbf{Y} - \mathbf{W}^{\text{TC}} \cdot \hat{\mathbf{R}} \right\|_{\text{F}}^2 + \frac{\lambda}{2} \left(\|\mathbf{S}\|_{\text{F}}^2 + \|\mathbf{U}\|_{\text{F}}^2 + \|\mathbf{T}\|_{\text{F}}^2 + \|\mathbf{C}\|_{\text{F}}^2 \right) \quad (10)$$

其中,第1项是基于Frobenius范数的损失值计算, \mathbf{Y} 为最终的预测值, \mathbf{W}^{TC} 为类别转移权重;第2项为正则化项,用来解决众所周知的过拟合问题, λ 为调控参数。然后,把目标函数相对于因子矩阵 $\mathbf{U}, \mathbf{T}, \mathbf{C}$ 和核心张量 \mathbf{S} 求梯度,如式(11)所示

$$\left. \begin{aligned} \partial_{U_{i*}} L(\mathbf{Y}_{ijk}, \mathbf{R}_{ijk}) &= \partial_{Y_{ijk}} L(\mathbf{Y}_{ijk}, \hat{\mathbf{R}}_{ijk}) \mathbf{S} \times_{\mathbf{T}} \mathbf{T}_{j*} \times_{\mathbf{C}} \mathbf{C}_{k*} \\ \partial_{T_{j*}} L(\mathbf{Y}_{ijk}, \mathbf{R}_{ijk}) &= \partial_{Y_{ijk}} L(\mathbf{Y}_{ijk}, \hat{\mathbf{R}}_{ijk}) \mathbf{S} \times_{\mathbf{U}} \mathbf{U}_{i*} \times_{\mathbf{C}} \mathbf{C}_{k*} \\ \partial_{C_{k*}} L(\mathbf{Y}_{ijk}, \mathbf{R}_{ijk}) &= \partial_{Y_{ijk}} L(\mathbf{Y}_{ijk}, \hat{\mathbf{R}}_{ijk}) \mathbf{S} \times_{\mathbf{U}} \mathbf{U}_{i*} \times_{\mathbf{T}} \mathbf{T}_{j*} \\ \partial_{\mathbf{S}} L(\mathbf{Y}_{ijk}, \mathbf{R}_{ijk}) &= \partial_{Y_{ijk}} L(\mathbf{Y}_{ijk}, \hat{\mathbf{R}}_{ijk}) \mathbf{U}_{i*} \otimes \mathbf{T}_{j*} \otimes \mathbf{C}_{k*} \end{aligned} \right\} \quad (11)$$

其中,含有符号*的 U_{i*} 表示因子矩阵 \mathbf{U} 的第 i 行,其余类似表示同理。符号 \otimes 表示矩阵的外积。详细的类别预测算法过程如表1所示。

4 分区推荐

用户访问位置的概率不仅受距离影响,而且还受到位置固有特性的影响^[11]。因此,用户的实际签到位置分布在不同区域,且潜在影响因子多有不同。本文将POI推荐细粒度化:本地推荐和异地推荐。

本文用K-means聚类算法将每个用户签到信息中的经纬度聚类得到一个中心作为每个用户的常驻地loc;然后以此中心设置距离阈值 f ,若当前位置 cur 与常驻地 loc 的距离小于该阈值,则推荐场景为本地推荐,否则为异地推荐。如图5所示,这是某个用户的历史访问记录,虚线圈之内定义为此用户的本地范围,之外则为异地范围。

首先,将式(12)的类别预测分数进行排序,获得目标用户的类别预测列表 C 。再构造目标用户的二分图 $\text{BG} = (\tilde{U}, V_c, E)$,其中 \tilde{U} 表示目标用户的朋友集合用户, V_c 是 \tilde{U} 中用户在预测类别列表中访问过的位置, E 表示用户和位置之间的关系连接,如图6所示。

表 1 类别预测算法

输入：张量 \mathbf{R} ，权重 \mathbf{W}^{TC} ，误差阈值 ε
 输出：因子矩阵 $\mathbf{U}, \mathbf{T}, \mathbf{C}$ 和核心张量 \mathbf{S}

- (1) 通过HOSVD初始化张量 \mathbf{R} ，得到 $\hat{\mathbf{R}}$;
- (2) 赋给 $\mathbf{U} \in \mathbf{R}^{m \times \tau_1}$, $\mathbf{T} \in \mathbf{R}^{g \times \tau_2}$, $\mathbf{C} \in \mathbf{R}^{n \times \tau_3}$, $\mathbf{S} \in \mathbf{R}^{\tau_1 \times \tau_2 \times \tau_3}$ 随机数值，设置步长 φ ;
- (3) While $L_t - L_{t+1} > \varepsilon$:
- (4) For $\hat{\mathbf{R}}_{ijk} \neq 0$:
- (5) $\hat{\mathbf{R}}_{ijk} = \mathbf{S} \times_{\mathbf{U}} \mathbf{U}_{i*} \times_{\mathbf{T}} \mathbf{T}_{j*} \times_{\mathbf{C}} \mathbf{C}_{k*}$
- (6) $\mathbf{U}_{i*} \leftarrow \mathbf{U}_{i*} - \phi \lambda \mathbf{U}_{i*} - \phi (\mathbf{Y}_{ijk} - \mathbf{W}^{TC} \cdot \hat{\mathbf{R}}_{ijk}) \mathbf{S} \times_{\mathbf{T}} \mathbf{T}_{j*} \times_{\mathbf{C}} \mathbf{C}_{k*}$
- (7) $\mathbf{T}_{j*} \leftarrow \mathbf{T}_{j*} - \phi \lambda \mathbf{T}_{j*} - \phi (\mathbf{Y}_{ijk} - \mathbf{W}^{TC} \cdot \hat{\mathbf{R}}_{ijk}) \mathbf{S} \times_{\mathbf{U}} \mathbf{U}_{i*} \times_{\mathbf{C}} \mathbf{C}_{k*}$
- (8) $\mathbf{C}_{k*} \leftarrow \mathbf{C}_{k*} - \phi \lambda \mathbf{C}_{k*} - \phi (\mathbf{Y}_{ijk} - \mathbf{W}^{TC} \cdot \hat{\mathbf{R}}_{ijk}) \mathbf{S} \times_{\mathbf{U}} \mathbf{U}_{i*} \times_{\mathbf{T}} \mathbf{T}_{j*}$
- (9) $\mathbf{S} \leftarrow \mathbf{S} - \phi \lambda \mathbf{S} - \phi (\mathbf{Y}_{ijk} - \mathbf{W}^{TC} \cdot \hat{\mathbf{R}}_{ijk}) \mathbf{U}_{i*} \otimes \mathbf{T}_{j*} \otimes \mathbf{C}_{k*}$
- (10) End
- (11) 输出 $\mathbf{U}, \mathbf{T}, \mathbf{C}, \mathbf{S}$

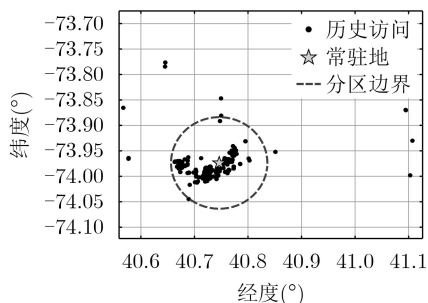


图 5 用户区域划分示例

然后在不同的区域，匹配不同权重的影响因素。最后，结合式(5)–式(8)和二分图BG，得到用户对下一个地点的访问概率Score如式(13)所示

$$\text{Score}_{(l,t+1)} = \alpha(\text{Val} + F_n) \cdot \text{Grade}_{(l,t)} + (1 - \alpha)(\text{Pop} + F_h) \cdot \text{Grade}_{(l,t)} \quad (12)$$

其中，Grade是从BG图中提取的目标用户对当前POI的偏好得分，取值范围归一到[0,1]之间，示例：在图7中， u_i 对地点 l_3 的当前偏好得分是由对类别 c_2 的兴趣分0.54乘上在上下文信息影响下属于类别 c_2 的 l_3 的权重分数0.89，最后进行归一化得到

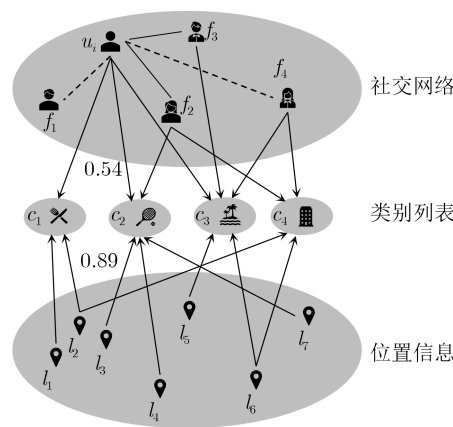


图 6 二分图示例

Grade值为0.79。 α 为分区推荐的可调控参数，由用户常驻地loc与其当前位置cur的距离来决定。

5 实验总结

5.1 数据集

Foursquare是一种基于地理位置信息的社交网络，它记录了用户的签到信息，例如用户信息、地点经纬度等，而且基于此网络的数据集多用于推荐

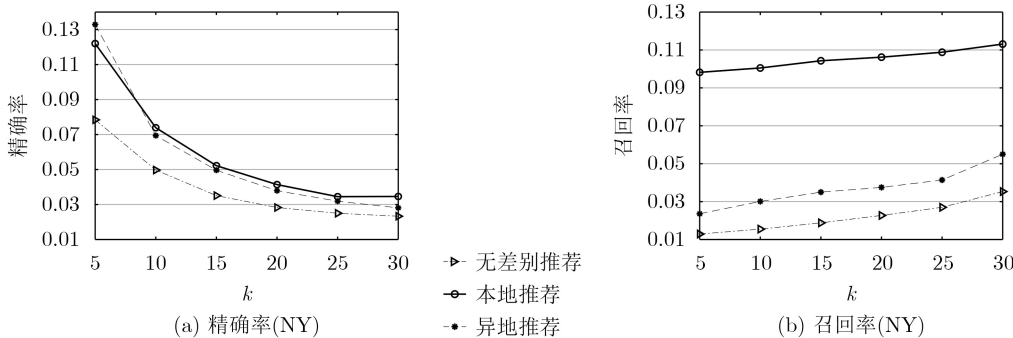


图 7 分区推荐的优势对比

系统的评测。本文的两个数据集是由Foursquare用户在加州(California, CA)和纽约州(New York, NY)两个特大地区生成的,包含了用户在2012年4月至2013年4月期间发布的信息,其中同时包含的类别信息和具体的时间信息,符合本文实验所需要的数据标签且方便数据预处理,分布如表2所示。在数据的预处理阶段,去掉签到记录少于10条的用户以及少于5个用户访问的地点,以此缓解用户的朋友关系。然后在每个用户中随机选择70%的签到信息作为实验的训练数据,30%作为实验的测试数据。

表2 数据集统计分布表

数据集	用户数量	地点数量	地点类别	签到记录
NY	972	5880	368	13258
CA	4018	71422	396	401555

5.2 评价指标

本文提及的算法采用以下两种评价指标:

(1) Precision: 精确率,在推荐的 k 个POI中有多大的概率是精确的推荐,公式如式(14)所示

$$\text{Pre}@k = \frac{1}{N} \sum_{u=1}^N \frac{|Z_u(k) \cap V_u|}{k} \quad (13)$$

(2) Recall: 召回率,在测试集中有多大的概率是精确的推荐,计算公式如式(15)所示

$$\text{Rec}@k = \frac{1}{N} \sum_{u=1}^N \frac{|Z_u(k) \cap V_u|}{|V_u|} \quad (14)$$

其中, Z_u 是测试数据集, V_u 为预测推荐集合。

5.3 实验结果对比分析

本文提出的基于类别转移加权张量分解模型的兴趣点分区推荐算法(Weighted Tensor Decomposition Partition Recommendation algorithm, WTD-PR)将与以下5种算法作对比:

(1) Augmented Square error based Matrix Factorization (ASMF)^[6]: 基于平方误差的矩阵分解算法,该算法融合了3种朋友影响以及地理影响。

(2) Geographical Factorization Model based on POI Ranking (Rank-GeoFM)^[2]: 基于POI排名的因式分解算法,并且融入地理和时间因素。

(3) POI recommendation based on Tensor Decomposition model (TD)^[4]: 基于张量分解模型并融合多维信息兴趣点推荐算法。

(4) A LOcation REcommendation algorithm that incorporates Geography, Friends and Popularity factors (GFP-LORE)^[8]: 融合了序列信息、地理、朋友和流行度因素的推荐算法。

(5) Location neighborhood-aware Weighted pro-babilistic Matrix Factorization (LWFM)^[7]: 位置邻域感知加权概率矩阵分解算法,从位置的角度挖掘地理特征,将其作为隐式反馈权重。

为了验证本文算法在兴趣点分区推荐上的有效性,我们进行了无差别推荐、本地推荐和异地推荐这3种验证实验,其在 k 为5、10、15、20、25、30的情况下的精确率和召回率如图7所示。相比于无差别推荐,本地推荐和异地推荐的推荐性能在不同的 k 值上都处于领先状态,这说明本文的模型在推荐层面更具有针对性,不同的场景中融入合适的影响因素,提高了推荐数据的有效性,缓解数据稀疏问题的同时提升人们的旅行体验度。

在不分区推荐中,本文训练出一个 α 值,满足与其他算法的性能对比方式, α 取0.69。实验数据如图8所示,无论是在哪一个区域,本文算法的性能都有较大的优势,且相对于POI的直接推荐,类别预测在系统冷启动问题上更有话语权。本文算法与LWFM和GFP-LORE这两种算法作对比,证明了高阶张量分解应用于推荐系统的优势:张量填补减缓了数据稀疏性,且降低了信息的维度。TD模型和WTD-PR模型的对比,充分地显示了权重的重要性,本文的权重分析是依据加权马尔科夫链的序列转移原理,这同时说明用户连续行为的特征和历史访问记录对于人们的下一步访问有着很大的主导作用。

除了以上两种实验场景,本文还做了消融实验。其中对比算法分为2类5种:(1)在类别预测模块消除信息维度:不融入时间信息(WTD-T),不融入类别信息(WTD-C);(2)在类别预测列表的基础上融合不同影响因素:融合朋友和距离因素(WTD-FL),融合朋友和地点流行度因素(WTD-FP),融合位置和地点流行度因素(WTD-LP)。对比结果如表3所示,第1类算法的性能指标没有第2类的高,这表明时间和类别信息对模型性能的影响较大,与现实中人们在不同的时段会有不同的喜好侧重现象相吻合。在第2类算法中,WTD-LP算法的指标相对较低,这说明社交关系在推荐模型中有很大的权重占比。相对于这两类算法,WTD-PR考虑多维上下文信息,分配其合适的权重,获得了最优的性能,这表明推荐系统中上下文信息的重要性。此外,在该实验中,去掉不同的数据标签,本文模型仍具有较好的性能,这说明本文模型能适应标签种类不同的相关数据集。

6 结束语

本文研究了一种基于类别转移加权张量分解模

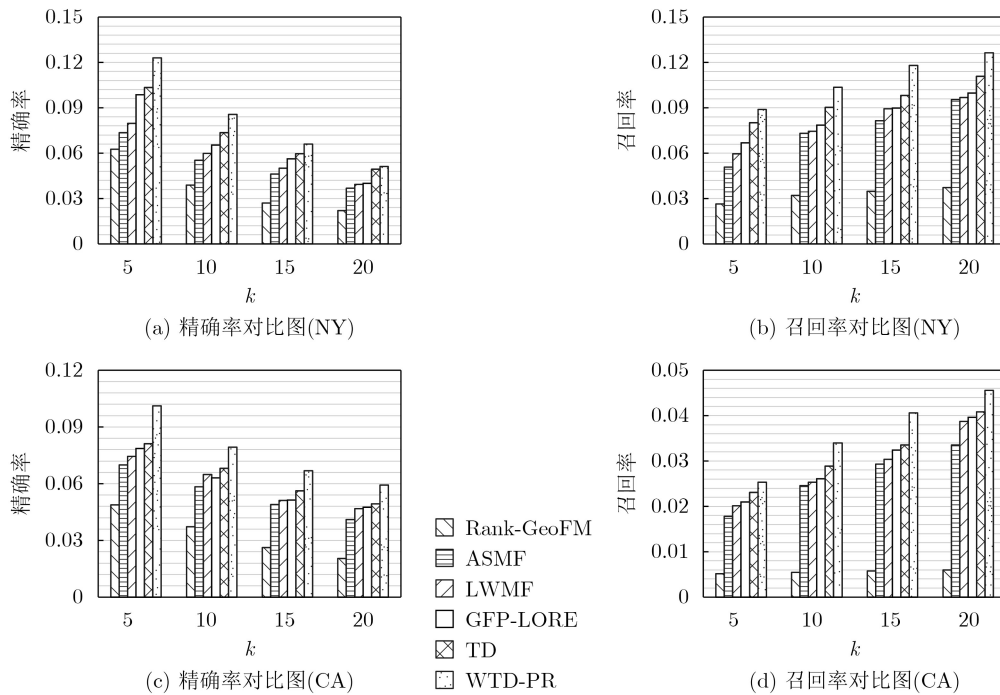


图 8 算法性能对比

表 3 多维因素有效性验证

数据集	评价指标	WTD-T	WTD-C	WTD-LP	WTD-FL	WTD-FP	WTD-PR
NY	Pre@10	0.0679	0.0685	0.0713	0.0739	0.0784	0.0810
	Rec@10	0.0815	0.0862	0.0977	0.0961	0.1005	0.1035
CA	Pre@10	0.0648	0.0653	0.0679	0.0685	0.0705	0.0794
	Rec@10	0.0257	0.0261	0.0273	0.0286	0.0291	0.0340

型的POI分区推荐算法。在类别因子上融入加权马尔科夫模型，生成类别转移权重；构造基于用户-时间-类别的加权张量，利用梯度下降算法进行迭代更新，推选出候选类别；再把位置信息中隐藏的距离因素、流行度因素以及朋友因素与候选类别进行融合，作基于用户当前位置的POI分区推荐，并在无差别推荐、不分区推荐和消融实验中做了对比分析。实验表明，本文算法在多维信息融合和张量改进方面做了很大的突破，提升了性能也增加了通用性。

参 考 文 献

[1] YU Fet, LI Zhijun, JIANG Shouxu, *et al.* Point-of-interest recommendation for location promotion in location-based social networks[C]. 2017 18th IEEE International Conference on Mobile Data Management, Daejeon, South Korea, 2017: 344–347. doi: [10.1109/MDM.2017.57](https://doi.org/10.1109/MDM.2017.57).

[2] LI Xutao, CONG Gao, LI Xiaoli, *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: 433–442. doi: [10.1145/2766462.2767722](https://doi.org/10.1145/2766462.2767722).

[3] LIAN Defu, ZHENG Kai, GE Yong, *et al.* GeoMF++: Scalable location recommendation via joint geographical modeling and matrix factorization[J]. *ACM Transactions on Information Systems*, 2018, 36(3): 33. doi: [10.1145/3182166](https://doi.org/10.1145/3182166).

[4] YAO Lina, SHENG Q Z, QIN Yongrui, *et al.* Context-aware point-of-interest recommendation using tensor factorization with social regularization[C]. The 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, Santiago, Chile, 2015: 1007–1010. doi: [10.1145/2766462.2767794](https://doi.org/10.1145/2766462.2767794).

[5] ZHANG Lu, SUN Zhu, ZHANG Jie, *et al.* Modeling hierarchical category transition for next POI recommendation with uncertain check-ins[J]. *Information Sciences*, 2020, 515: 169–190. doi: [10.1016/j.ins.2019.12.006](https://doi.org/10.1016/j.ins.2019.12.006).

[6] 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](https://doi.org/10.1145/2939672).

- 2939767.
- [7] GUO Lei, WEN Yufei, and LIU Fangai. Location perspective-based neighborhood-aware POI recommendation in location-based social networks[J]. *Soft Computing*, 2019, 23(22): 11935–11945. doi: [10.1007/s00500-018-03748-9](https://doi.org/10.1007/s00500-018-03748-9).
- [8] REN Yueqiang, WANG Ze, SUN Xiaona, *et al.* A multi-element hybrid location recommendation algorithm for location based social networks[J]. *IEEE Access*, 2019, 7: 100416–100427. doi: [10.1109/ACCESS.2019.2929313](https://doi.org/10.1109/ACCESS.2019.2929313).
- [9] YAO Lina, SHENG Q Z, WANG Xianzhi, *et al.* Collaborative location recommendation by integrating multi-dimensional contextual information[J]. *ACM Transactions on Internet Technology*, 2018, 18(3): 32. doi: [10.1145/3134438](https://doi.org/10.1145/3134438).
- [10] HE Jing, LI Xin, LIAO Lejian, *et al.* Inferring a personalized next point-of-interest recommendation model with latent behavior patterns[C]. The Thirtieth AAAI Conference on Artificial Intelligence, Phoenix, USA, 2016: 137–143. doi: [10.5555/3015812.3015833](https://doi.org/10.5555/3015812.3015833).
- [11] DING Jingtao, YU Guanghui, LI Yong, *et al.* Learning from Hometown and current city: Cross-city POI recommendation via interest drift and transfer learning[J]. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2019, 3(4): 131. doi: [10.1145/3369822](https://doi.org/10.1145/3369822).
- [12] YIN Hongzhi, ZHOU Xiaofeng, CUI Bin, *et al.* Adapting to user interest drift for POI recommendation[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2016, 28(10): 2566–2581. doi: [10.1109/TKDE.2016.2580511](https://doi.org/10.1109/TKDE.2016.2580511).
- [13] LIAN Defu, ZHANG Zhenyu, GE Yong, *et al.* Regularized content-aware tensor factorization meets temporal-aware location recommendation[C]. 2016 IEEE 16th International Conference on Data Mining, Barcelona, Spain, 2016: 1029–1034. doi: [10.1109/ICDM.2016.0131](https://doi.org/10.1109/ICDM.2016.0131).
- [14] GAO Huiji, TANG Jiliang, HU Xia, *et al.* Exploring temporal effects for location recommendation on location-based social networks[C]. The 7th ACM Conference on Recommender Systems, Hong Kong, China, 2013: 93–100. doi: [10.1145/2507157.2507182](https://doi.org/10.1145/2507157.2507182).
- [15] 赵明, 闫寒, 曹高峰, 等. 融合用户信任度和相似度的基于核心用户抽取的鲁棒性推荐算法[J]. 电子与信息学报, 2019, 41(1): 180–186. doi: [10.11999/JEIT180142](https://doi.org/10.11999/JEIT180142).
- ZHAO Ming, YAN Han, CAO Gaofeng, *et al.* Robust recommendation algorithm based on core user extraction with user trust and similarity[J]. *Journal of Electronics & Information Technology*, 2019, 41(1): 180–186. doi: [10.11999/JEIT180142](https://doi.org/10.11999/JEIT180142).
- [16] CHENG Chen, YANG Haiqin, KING I, *et al.* Fused matrix factorization with geographical and social influence in location-based social networks[C]. The Twenty-Sixth AAAI Conference on Artificial Intelligence, Toronto, Canada, 2012: 17–23. doi: [10.5555/2900728.2900731](https://doi.org/10.5555/2900728.2900731).
- [17] 司亚利, 张付志, 刘文远. 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[J]. 电子与信息学报, 2020, 42(3): 678–686. doi: [10.11999/JEIT190287](https://doi.org/10.11999/JEIT190287).
- SI Yali, ZHANG Fuzhi, and LIU Wenyuan. An adaptive point-of-interest recommendation method based on check-in activity and temporal-spatial probabilistic models[J]. *Journal of Electronics & Information Technology*, 2020, 42(3): 678–686. doi: [10.11999/JEIT190287](https://doi.org/10.11999/JEIT190287).
- [18] CHEN Jialiang, LI Xin, CHEUNG W K, *et al.* Effective successive POI recommendation inferred with individual behavior and group preference[J]. *Neurocomputing*, 2016, 210: 174–184. doi: [10.1016/j.neucom.2015.10.146](https://doi.org/10.1016/j.neucom.2015.10.146).
- 李 胜: 男, 1984年生, 博士, 副教授, 研究方向为大数据推荐、信号处理和医疗图像处理。
- 刘桂云: 女, 1995年生, 硕士生, 研究方向为兴趣点推荐。
- 何熊熊: 男, 1965年生, 博士, 教授, 研究方向为数据驱动迭代学习控制。

责任编辑: 马秀强