

## 基于时间效应的兴趣点推荐混合模型

张岐山<sup>1</sup>, 李 可<sup>1</sup>, 林小榕<sup>2</sup>

(1. 福州大学 经济与管理学院, 福州 350108; 2. 北京交通大学 下一代互联网互联设备国家工程实验室, 北京 100044)

**摘 要:** 在基于位置的社交网络中, 兴趣点实时推荐数据和用户签到数据存在高稀疏性问题。提出一种基于时间效应的混合推荐模型。通过用户潜在兴趣点数据模型计算用户时间行为影响分数和地理位置影响分数, 并用线性统一模型进行处理, 选取 Top  $S$  个兴趣点作为用户的潜在兴趣点。将用户的潜在签到记录引入基于时间效应的矩阵分解模型中, 考虑时间差异性和连续性对推荐结果的影响, 在此基础上进行优化求解, 提出推荐策略。实验结果表明, 与 LRT 模型、UTE + SE 模型相比, 该模型的推荐效果较好, 其准确率和召回率最高可达 0.103 4 和 0.111 8。

**关键词:** 基于位置的社交网络; 时间信息; 地理位置信息; 矩阵填充; 矩阵分解; 实时推荐; 兴趣点

**中文引用格式:** 张岐山, 李可, 林小榕. 基于时间效应的兴趣点推荐混合模型[J]. 计算机工程, 2019, 45(8): 203-209.

**英文引用格式:** ZHANG Qishan, LI Ke, LIN Xiaorong. Hybrid model for point-of-interests recommendation based on time effect[J]. Computer Engineering, 2019, 45(8): 203-209.

### Hybrid Model for Point-of-Interests Recommendation Based on Time Effect

ZHANG Qishan<sup>1</sup>, LI Ke<sup>1</sup>, LIN Xiaorong<sup>2</sup>

(1. School of Economics and Management, Fuzhou University, Fuzhou 350108, China;

2. National Engineering Laboratory for Next Generation Internet Interconnection Devices, Beijing Jiaotong University, Beijing 100044, China)

**[Abstract]** In Location-Based Social Networks (LBSNs), real-time recommendation data of Point-of-Interests (POIs) and check-in data of users are highly sparse. Therefore, a hybrid recommendation model based on time effect is proposed. Through the data model of potential POIs of users, the time behavior influence scores and geographical location influence scores are calculated, and the linear unified model is used for processing. Top  $S$  POIs are selected as the potential POIs of users. Incorporate the user's potential check-in recorders into a time-based matrix factorization model, taking into account the influence of time difference and continuity on the recommendation results. On this basis, optimization solution is carried out, and the recommended strategy is proposed. Experimental results show that compared with the LRT model and UTE + SE model, the proposed model has better recommendation effect, and its precision and recall rate can reach up to 0.103 4 and 0.111 8.

**[Key words]** Location-Based Social Networks (LBSNs); time information; geographical location information; matrix filling; matrix factorization; real-time recommendation; Point-of-Interests (POIs)

**DOI:** 10.19678/j.issn.1000-3428.0051987

## 0 概述

近年来, 基于位置的社交网络 (Location-Based Social Networks, LBSNs) 得到迅猛发展, 例如 Loopt、Yelp、Foursquare、Whrrl 等<sup>[1]</sup>。在这些 LBSNs 中, 用户访问餐馆、电影院、博物馆等线下兴趣点 (Point-of-Interests, POIs) 时, 可在线进行“签到”活动并分享他们访问兴趣点时的建议与感受<sup>[2]</sup>。兴趣点推荐可以减少用户的搜寻时间, 为商家提供精准的营销策略。一些学者将日常模式纳入推荐系统中, 可以更好地

了解用户的移动行为, 在正确的时间为目标用户推荐适合的兴趣点集。兴趣点实时推荐更容易激发用户的消费欲望, 同时也面临更严重的数据稀疏问题, 影响推荐效果。

LBSNs 中的偏好矩阵 (用户-兴趣点矩阵) 存在严重的数据稀疏问题<sup>[3-5]</sup>, 若根据时间槽划分偏好矩阵, 会使数据变得更加稀疏, 对学习用户偏好造成负面影响。文献[3]将时间信息纳入矩阵分解过程中, 生成具有实时推荐效果的似然比检验模型 (Likelihood Ratio Test, LRT), 文献[6]采用融合时

**基金项目:** 国家自然科学基金 (61300104); 福建省自然科学基金 (2018J01791)。

**作者简介:** 张岐山 (1962—), 男, 教授、博士、博士生导师, 主研方向为数据挖掘、推荐系统、系统优化与仿真; 李 可、林小榕, 硕士研究生。

**收稿日期:** 2018-07-02    **修回日期:** 2018-08-10    **E-mail:** 411700408@qq.com

间信息的协同过滤模型,2种方法均容易受到数据稀疏性的影响。

同时,兴趣点实时推荐面临准确率和召回率低的问题。文献[4]将LRT模型同USG、GeoMF、IRenMF等9种主流模型进行比较,发现LRT模型的表现较差。这是因为数据高稀疏性导致兴趣点集实时推荐的准确率和召回率较低,而实时推荐的要求比常见的推荐要高,难度更大,并且LRT模型中没有对地理位置信息进行建模。

本文构建一种混合推荐模型ATG(Aware-Time and Geography),融合签到信息、地理位置信息和时间信息的影响,实现兴趣点集的实时推荐。针对兴趣点实时推荐中数据高稀疏性的问题,ATG模型采用平滑相似度和矩阵填充的方式,以考虑用户相似性,同时缓解数据稀疏的影响,改善模型的推荐效果。

## 1 相关工作

用户在LBSNs中会产生大量的签到信息、朋友信息、地理位置信息、时间信息和评论信息等。如何利用这些信息,提高兴趣点推荐的效果是研究的重点。

协同过滤<sup>[7-8]</sup>和矩阵分解<sup>[9-10]</sup>是兴趣点推荐研究中的2种主流算法。协同过滤算法容易受到数据稀疏性的影响,不能很好地实现对隐式反馈数据集的挖掘。矩阵分解算法不仅可以缓解数据稀疏性的影响还可以融合异构数据源,考虑隐式反馈数据集,但其未考虑用户之间的相似性。为了克服这2种算法的不足,研究人员提出混合算法。文献[11]提出了“两步走”的框架。第1步设计基于线性聚集和基于随机游走的2种方法,为每个用户学习一组潜在兴趣点。第2步用基于平方误差的损失函数和基于排名误差的损失函数来模拟签到。

在兴趣点推荐模型中,对地理位置的研究较为成熟。通过对地理位置信息的研究发现,人们往往喜欢访问离自己近的兴趣点。文献[12]将地理位置影响引入LBSNs的兴趣点推荐中,采用幂律概率模型捕捉兴趣点对之间的地理位置影响,通过朴素贝叶斯方法计算拟推荐兴趣点的地理位置对目标用户的影响分数,再使用统一的协同过滤方法进行兴趣点推荐。文献[13]将地理位置信息纳入矩阵分解算法中,根据地理特征,将领域分为实例级别领域和区域级别领域2个层次,利用领域的特征进行个性化推荐。文献[11]考虑用户“家”的地理位置,当仅考虑地理位置的影响时,家与兴趣点之间的距离同其访问该兴趣点的概率呈幂律分布。

签到行为中的时间信息也引起了研究人员的广泛关注。在LBSNs中,用户的每次签到、评论都有时间戳。时间信息在兴趣点推荐中的应用主要包括

时间序列模式、时间周期模式和时间感知主题。时间序列模式指沿着时间轴,在LBSNs中连续的、个性化的推荐。文献[14]将马尔可夫链纳入矩阵分解算法中,通过用户的移动约束完成LBSNs中连续的、个性化的兴趣点推荐任务,预测用户第2天甚至是未来几天的移动偏好。文献[15]认为人类的偏好会随着时间的推移而变化,提出STG模型来表示用户的长期偏好和短期偏好,以签到时间和推荐时间之间的间隔作为衰减因子,离推荐时间越远的兴趣点,其权重越小。时间周期模式是指用户的偏好会随时间呈周期性变化,不同的人会有不同的时间行为。例如,以小时为单位,一天则为一个周期。在兴趣点推荐中引入时间周期模式,可以更好地了解人类的移动行为,在正确的时间推荐合适的兴趣点,促进用户消费。文献[3]依据时间的非均匀、连续特性,研究用户签到的时间周期模式,将时间信息纳入矩阵分解中,生成具有实时推荐效果的LRT推荐模型,并使用4个时间聚合策略整合用户对不同时态的偏好。文献[6]将时间周期模式融合到基于用户的协同过滤框架中,通过用户之间的时间相似行为进行实时推荐。时间感知主题是指用户在相同或相近的时间槽内会访问相似的兴趣点,这也称为签到内容与签到时间的“同现模式”。基于这一特点,可根据兴趣点的签到时间来推断该兴趣点的语义标签。例如,某一兴趣点的签到时间主要集中在11:00—13:00,17:00—19:00,则其就极有可能是餐饮类的兴趣点。文献[16]将时间信息标记为特征,提出一种利用标记方法来推荐兴趣点的语义标签。文献[17]提出时间感知兴趣点模型,用LDA(Latent Dirichlet Allocation)方法判断主题的时间分布,从而推断主题。

为了在LBSNs中获得实时推荐兴趣点集,考虑时间差异性和连续性对兴趣点推荐的影响,并缓解数据稀疏性,本文提出基于时间效应的兴趣点推荐混合模型ATG,采用平滑相似度和矩阵填充的方式提高模型的推荐效果。

## 2 用户潜在兴趣点数据模型

### 2.1 推荐模型框架

本文ATG模型采用“两步走”的推荐框架,如图1所示。首先,建立用户潜在兴趣点数据模型,采用协同过滤算法,基于用户的签到信息、地理位置信息和时间信息计算时间行为影响分数和地理位置影响分数,进而发现目标用户的潜在兴趣点和可能签到的时间。将潜在签到信息(用户ID-兴趣点ID-时间)填充到偏好矩阵中,缓解矩阵的数据稀疏性问题。其次,建立一个基于时间效应的矩阵分解模型,考虑时间差异性和连续性的影响,训练用户特征矩阵和兴趣点特征矩阵。最后,估算出目标用户对于

某一特定兴趣点的访问概率,选取 Top  $k$  个兴趣点进行推荐。

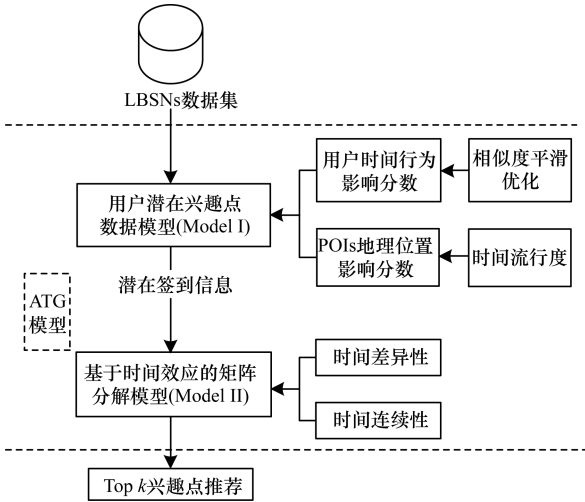


图 1 本文兴趣点推荐体系的结构框架

**定义** 假设  $u_i$  为目标用户,  $l_j$  为拟推荐兴趣点,  $U = \{u_1, u_2, \dots, u_n\}$  为用户集,  $L = \{l_1, l_2, \dots, l_m\}$  为兴趣点集。  $n$  个用户对  $m$  个兴趣点的偏好构成偏好矩阵  $P$ ,  $P_{i,j}$  表示  $u_i$  对  $l_j$  的偏好,  $P_{i,j} \in P$ 。当  $u_i$  访问过  $l_j$  时,  $P_{i,j} = 1$ , 反之  $P_{i,j} = 0$ 。  $P_{i,j} = 0$  并不意味着  $u_i$  不喜欢  $l_j$ , 还有可能是用户不知道这个兴趣点。因此, 本文运用 ATG 模型计算出  $u_i$  访问每一个未访问过的 POIs 的概率, 选取 Top  $S$  作为  $u_i$  的拟推荐兴趣点集。

### 2.2 基于时间行为影响的模型构建

用户在 LBSNs 中有大量的签到信息。其中, 时间是一个重要的信息。由于大多数人的日常行动保持相对固定的模式, 且每个人在不同的时间点会有不同的偏好, 为了捕捉用户时间行为的相似性, 本文将时间分成以小时为单位的 24 个时间槽, 通过用户在各时间槽中的历史签到记录构建时间偏好模型, 计算用户之间的时间行为相似度, 达到实时推荐兴趣点集的目的。采用余弦相似度来度量 2 个用户之间的时间行为相似度, 建模如下:

$$W_{u_i, u_u}^t = \left( \sum_{i=1}^T \sum_{z \in M_i^o} r_{i,t,z}^2 \cdot \sum_{i=1}^T \sum_{z \in M_u^o} r_{u,t,z}^2 \right)^{-\frac{1}{2}} \cdot \sum_{i=1}^T \sum_{z \in M_i^o \cap M_u^o} r_{i,t,z} r_{u,t,z} \quad (1)$$

其中,  $r_{i,t,z}$  表示  $u_i$  在时间槽  $t$  内到兴趣点  $l_z$  的签到次数,  $r_{u,t,z}$  表示  $u_u$  在时间槽  $t$  内到兴趣点  $l_z$  的签到次数,  $M_i^o$  表示  $u_i$  访问过的兴趣点集合,  $M_u^o$  表示  $u_u$  访问过的兴趣点集合。

然而, 该模型存在 2 个问题:

1) 根据 24 个时间槽划分签到信息导致数据更加稀疏, 容易影响推荐效果。

2) 稀疏的签到信息容易导致一些无意义的相似度计算。例如, 用户 A 在 12 点时访问过兴趣点  $a$ , 在

18 点时访问过兴趣点  $b$ ; 而用户 B 在 18 点时访问兴趣点  $a$ , 在 12 点时访问过兴趣点  $b$ 。若不考虑时间信息, 用户 A 和用户 B 的相似度为 1; 如果考虑时间信息, 两者的相似度则为 0。

为了解决这 2 个问题, 本文使用时间槽之间的相似度值  $\rho$  来平滑用户间的时间行为相似性  $w$ , 建模过程如下:

令  $R_{u,t} = (r_{u,t,l_1}, r_{u,t,l_2}, \dots, r_{u,t,l_m})$  为用户  $u$  在时间槽  $t$  内的签到向量。  $R_{u,t_i}$  和  $R_{u,t_j}$  分别表示用户  $u$  在时间槽  $t_i$  和  $t_j$  的签到向量, 具体如下:

$$R_{u,t_i} = (r_{u,t_i,l_1}, r_{u,t_i,l_2}, \dots, r_{u,t_i,l_m})$$

$$R_{u,t_j} = (r_{u,t_j,l_1}, r_{u,t_j,l_2}, \dots, r_{u,t_j,l_m})$$

对于用户  $u$  而言, 时间槽  $t_i$  和  $t_j$  之间的相似度  $\rho_u(t_i, t_j)$  的计算公式如下:

$$\rho_u(t_i, t_j) = \left( \|R_{u,t_i}\|_F^2 \cdot \|R_{u,t_j}\|_F^2 \right)^{-\frac{1}{2}} (R_{u,t_i} \cdot R_{u,t_j}) \quad (2)$$

取所有用户的相似度平均值作为时间槽  $t_i$  和  $t_j$  之间的相似度值  $\rho(t_i, t_j)$ , 计算过程如下:

$$\rho(t_i, t_j) = \frac{\sum_{u=1}^n \rho_u(t_i, t_j)}{n} \quad (3)$$

利用不同时间槽之间的时间相似度  $\rho(t_i, t_j)$  平滑元素  $r_{u,t,l}$ , 得到新值  $\tilde{r}_{u,t,l}$ , 建模如下:

$$\tilde{r}_{u,t,l} = \frac{\sum_{t'=1}^T \rho(t, t') r_{u,t',l}}{\sum_{t'=1}^T \rho(t, t')} \quad (4)$$

把经过平滑处理的  $\tilde{r}_{u,t,l}$  带入式(1), 计算用户间的时间行为相似度, 计算过程如下:

$$W_{u_i, u_u}^{te} = \left( \sum_{i=1}^T \sum_{z \in M_i^o} \tilde{r}_{i,t,z}^2 \cdot \sum_{i=1}^T \sum_{z \in M_u^o} \tilde{r}_{u,t,z}^2 \right)^{-\frac{1}{2}} \sum_{i=1}^T \sum_{z \in M_i^o \cap M_u^o} \tilde{r}_{i,t,z} \tilde{r}_{u,t,z} \quad (5)$$

时间行为对于兴趣点的影响分数  $C_{i,t,l}^{te}$  为:

$$C_{i,t,l}^{te} = \frac{\sum_u w_{i,u}^{te} \sum_{t'} \tilde{r}_{u,t',l} \cdot \rho(t, t')}{\sum_u w_{i,u}^{te}} \quad (6)$$

### 2.3 基于地理位置影响的模型构建

人们往往喜欢访问地理位置离自己较近的兴趣点, 单考虑地理位置影响的因素, 用户访问兴趣点的概率与其距离遵循幂率分布, 模型如下:

$$\Pr(d) = a \cdot d^b \quad (7)$$

其中,  $d$  表示用户到兴趣点的距离,  $a$  和  $b$  均为幂律分布的参数。

在本文中只有用户历史签到记录信息, 没有用户的实时地理位置信息, 因此, 无法算出用户与某一兴趣点之间的准确距离。为了解决这一问题, 本文采用基于朴素贝叶斯规则的模型来计算地理位置相似度  $P(l_j | L_{u_i})$ , 建模如下:

已知用户在兴趣点  $l_i$  签到, 则其访问兴趣点  $l_j$  的概率如式(8)所示。

$$P(l_j | l_i) = \frac{\Pr(d(l_i, l_j))}{\sum_{l_k \in L_{u_i}, l_k \neq l_i} \Pr(d(l_i, l_k))} \quad (8)$$

已知  $u_i$  的全部历史签到记录  $L_{u_i}$ , 可计算  $u_i$  访问兴趣点  $l_j$  的概率  $P(l_j | L_{u_i})$ , 其地理位置影响分数与概率呈正比, 根据朴素贝叶斯规则, 该分数计算如下:

$$C_{i,j}^g \propto P(l_j | L_{u_i}) = \frac{P(l_j) P(L_{u_i} | l_j)}{P(L_{u_i})} \propto P(l_j) P(L_{u_i} | l_j) \prod_{l_k \in L_{u_i}} P(l_k | l_j) \quad (9)$$

其中,  $P(L_{u_i})$  是一个定值,  $P(l_j)$  是兴趣点  $l_j$  被签到的先验概率, 即所有用户访问  $l_j$  的先验概率, 而不是特指用户  $u_i$  访问  $l_j$  的概率。值得注意的是, 本文假定  $L_{u_i}$  中兴趣点的签到概率相互独立。

#### 2.4 基于时间流行度影响的模型构建

根据经验可得, 不同兴趣点存在不同的高峰期, 如餐饮类兴趣点的高峰期集中在 11、12、17、18 这些时间槽内。本文引用“时间流行度”, 即某一兴趣点在时间槽  $t$  内的签到次数与该时间槽内签到总次数的比值来判断兴趣点在这一时间槽内的受欢迎程度。图 2 给出最受欢迎的 5 个兴趣点在各个时间槽内的时间流行度分布情况, 由图 2 可知, 每个兴趣点的时间流行度在时间槽之间都有各自的分布。

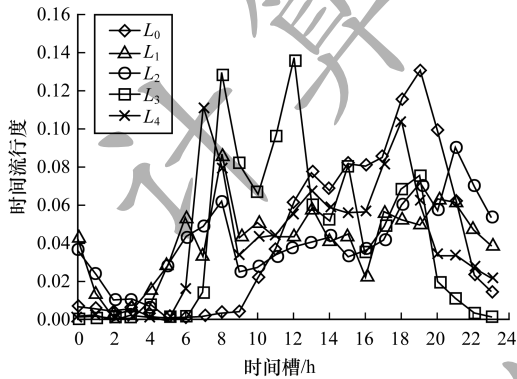


图 2 最受欢迎的 5 个兴趣点的流行度分布情况

在用户访问兴趣点时, 用户行为不仅受地理位置和时间行为的影响, 还受到时间流行度的影响。将时间流行度的影响纳入 ATG 模型, 建模为:

$$P_i(l_j) = \beta T_{i,j} + (1 - \beta) P(l_j) = \beta \frac{|Rt_{j,t}|}{\sum_{l \in L} |Rt_{l,t}|} + (1 - \beta) \frac{|Rt_j|}{\sum_{l \in L} |Rt_l|} \quad (10)$$

其中,  $T_{i,j}$  表示兴趣点  $l_j$  在时间槽  $t$  内的时间流行度,  $Rt_{j,t}$  表示兴趣点  $l_j$  在时间槽  $t$  内的签到次数,  $Rt_j$  表示兴趣点  $l_j$  在各时间槽内的签到次数总和。时间流行度的影响分数计算如下:

$$C_{i,t,j}^g \propto P_i(l_j) P(L_{u_i} | l_j) = P_i(l_j) \prod_{l_k \in L_{u_i}} P(l_k | l_j) \quad (11)$$

#### 2.5 统一模型框架

本文采用线性统一模型综合考虑时间、地理因素对兴趣点推荐的影响。由于这些因素是通过不同

的方法来衡量的, 具有不同的价值范围, 因此本文采用最小-最大归一化的方法进行处理, 然后再进行聚集。选取分数最高的 Top S 个兴趣点作为  $u_i$  的潜在兴趣点。

$$\hat{c}_{i,t,l}^t = \frac{c_{i,t,l}^t - \min_{l'}(c_{i,t,l'}^t)}{\max_{l'}(c_{i,t,l'}^t) - \min_{l'}(c_{i,t,l'}^t)} \quad (12)$$

$$\hat{c}_{i,t,l}^{gt} = \frac{c_{i,t,l}^{gt} - \min_{l'}(c_{i,t,l'}^{gt})}{\max_{l'}(c_{i,t,l'}^{gt}) - \min_{l'}(c_{i,t,l'}^{gt})} \quad (13)$$

$$score_{i,t,j} = \zeta \hat{c}_{i,t,l}^t + (1 - \zeta) \hat{c}_{i,t,l}^{gt} \quad (14)$$

### 3 基于时间效应的矩阵分解模型

#### 3.1 时间差异性

时间差异性指用户在不同的时间会有不同的偏好, 例如, 用户 A 喜欢下午访问 KTV, 用户 B 喜欢晚上访问 KTV。为了模拟这个属性, 本文以小时为单位, 将一天划分为 24 个时间槽,  $t = \{0, 1, \dots, 23\}$ , 如果签到时间为 12:00—12:59 则  $t = 12$ 。用户  $u_i$  对于兴趣点  $l_j$  的偏好程度受用户潜在特征和兴趣点潜在特征的影响。令在时间槽  $t$  内的用户特征矩阵为  $U_t$ , 兴趣点特征矩阵为  $V$ , 偏好矩阵为  $P$ , 则有如下公式:

$$P_t \approx \hat{P}_t = U_t^T V \quad (15)$$

考虑时间差异性, 则损失函数如下:

$$\min_{U, V, Q} \sum_{t=1}^T \|W_t \odot (P_t - \hat{P}_t)\|_2^2 + \frac{\lambda_u}{2} \|U_t\|_2^2 + \frac{\lambda_v}{2} \|V\|_2^2 \quad (16)$$

其中,  $\lambda_u, \lambda_v$  为正则化常数,  $W_t$  表示在时间槽  $t$  内的权重矩阵,  $\odot$  表示同或运算,  $w_{i,t,j}$  表示  $u_i$  在时间槽  $t$  内访问  $l_j$  的重要度量, 考虑用户签到次数的影响, 本文采用平方根的方法计算  $w$ , 过程如下:

$$w_{i,t,j} = \begin{cases} 1 + \sqrt{1 + \gamma \times r_{i,t,j}}, & j \in M_i^o \\ 1, & \text{其他} \end{cases} \quad (17)$$

#### 3.2 时间连续性

时间连续性指用户在连续的时间槽内有相似的偏好, 例如, 一般用户在 11:00—12:00 和 12:00—13:00 这 2 个时间槽内倾向于访问餐厅类的兴趣点。为了在模型中纳入时间连续性的影响, 本文提出一个时间规则, 其损失函数如下:

$$\min \sum_{t=1}^{23} \sum_{i=1}^n \rho_i(t, t-1) \|U_i^t - U_i^{t-1}\|_2^2 \quad (18)$$

其中,  $\rho_i(t, t-1)$  根据式(3)计算得到, 表示连续时间的相似程度。

#### 3.3 模型构建

为了缓解数据稀疏问题, 本文从用户潜在兴趣点数据模型中挖掘用户的潜在签到记录, 并填充在偏好矩阵中。然而, 用户对于潜在兴趣点和已签到过的兴趣点的偏好有差异, 因此, 本文将二元偏好变量  $P_{i,j}$  扩充为三元值, 具体如下:

$$P_{i,t,j} = \begin{cases} 1, j \in M_i^o \\ \alpha, j \in M_i^p \\ 0, \text{其他} \end{cases} \quad (19)$$

其中,  $\alpha \in [0, 1]$ ,  $M_i^o$  表示目标用户  $u_i$  访问过的兴趣点集,  $M_i^p$  表示目标用户  $u_i$  的潜在兴趣点集。

综合考虑时间差异性和时间连续性的影响, 可以通过求解以下优化问题来获得用户时间行为偏好:

$$\min_{U, V, Q} \sum_{t=1}^T \|W_t \odot (P_t - U_t^T V)\|_2^2 + \lambda_u \sum_{t=1}^{23} \|U_t\|_2^2 + \lambda_v \|V\|_2^2 + \lambda_l \sum_{t=1}^{23} \sum_{i=1}^n \rho_i(t, t-1) \|U_t^i - U_{t-1}^{i-1}\|_2^2 \quad (20)$$

本文采用变更最小二乘法优化损失函数, 训练出用户特征矩阵  $U_t$  和兴趣点特征矩阵  $V$ , 其更新过程如下:

$$\tilde{U}_t^i = U_t^i \left( \frac{(W_t \odot P_t) V + \lambda_l \sum_{t'} U_{t'}^{i-1}}{(W_t \odot U_t^T V) V + \lambda_l \sum_{t'} U_{t'}^i + \lambda_u U_t^i} \right)^{\frac{1}{2}} \quad (21)$$

$$\tilde{V}_i = V_i \left( \frac{\sum_{t=0}^{23} (W_t \odot P_t)^T U_t^i}{\sum_{t=0}^{23} (W_t \odot U_t^T V)^T U_t^i + \lambda_v V} \right)^{\frac{1}{2}} \quad (22)$$

## 4 实验结果与分析

### 4.1 实验数据集

本实验的数据来自 Foursquare 真实数据集<sup>[18]</sup>, 采集的是 2012 年 4 月 12 日—2013 年 2 月 16 日东京的长期(约 10 个月)签到数据, 包括用户 ID、兴趣点 ID、兴趣点经纬度及时间信息。预处理后的数据集中含有 2 292 名用户, 5 268 个兴趣点及 385 704 条签到记录。用户-兴趣点矩阵密度为 0.031 9。由于 LBSNs 存在严重的数据稀疏性, 因此 LBSNs 背景下的推荐模型准确率和召回率普遍较低。数据集的相关信息如表 1 所示。

表 1 Foursquare 数据集的相关信息

信息类别	数量
用户	2 292
兴趣点	5 268
签到记录	385 704
测试集	321 209
训练集	64 495

为了验证 ATG 模型的准确性, 对 Foursquare 数据集做如下处理:

- 1) 剔除访问兴趣点少于 10 的用户。
- 2) 剔除用户访问少于 10 的兴趣点。
- 3) 将数据集中 80% 的数据作为训练集, 剩余的 20% 作为测试集。

### 4.2 评价指标

为了评估模型的推荐效果, 本文采用时间求和的聚合策略计算最终的用户签到偏好, 以准确率和召回率为指标来评估算法性能, 计算如下:

$$P@k = \frac{1}{N} \sum_{i=1}^N \frac{\sum_{t=0}^{23} T(u_{i,t}) \cap \sum_{t=0}^{23} H(u_{i,t})}{k} \quad (23)$$

$$R@k = \frac{1}{N} \sum_{i=1}^N \frac{\sum_{t=0}^{23} T(u_{i,t}) \cap \sum_{t=0}^{23} H(u_{i,t})}{\sum_{t=0}^{23} T(u_{i,t})} \quad (24)$$

其中,  $P@k$  表示向目标用户推荐 Top  $k$  个兴趣点时的准确率,  $R@k$  表示向目标用户推荐 Top  $k$  个兴趣点时的召回率,  $N$  为用户数,  $T(u_{i,t})$  表示在测试集中  $u_i$  在时间槽  $t$  内访问过的兴趣点,  $H(u_{i,t})$  表示在推荐的  $k$  个兴趣点中命中的兴趣点。在本文实验中,  $k$  分别设置为 5、8、10、12、15、20。

### 4.3 推荐模型对比

为了评估 ATG 模型的性能, 本文选取以下 2 个实时推荐的经典模型与本文模型进行对比:

1) LRT<sup>[3]</sup> 是一个时间增强的矩阵分解模型, 根据时间槽将偏好矩阵分成 24 个子矩阵, 在不同的时间槽内对用户的偏好建模, 计算出所有的潜在向量, 从而得到最终的推荐分数, 进行实时推荐。

2) UTE + SE<sup>[6]</sup> 是融合时间感知的协同过滤推荐模型, 综合考虑用户行为偏好、时间信息和空间信息对兴趣点推荐的影响, 采用时间平滑技术计算用户之间的行为相似度, 实现在特定时间内为用户推荐兴趣点。

通过超参寻优发现, 当  $k$  设置为 5、8、10、12、15、20, 特征矩阵维度设置为 10, 用于调整时间流行度与兴趣点先验概率之间的权重  $\beta$  设置为 0.5, 用于调整用户时间行为影响与地理位置影响之间的权重  $\zeta$  设置为 0.8, 潜在兴趣点的偏好常量  $\alpha$  设置为 0.6, 正则化常数  $\lambda_u$ 、 $\lambda_v$ 、 $\lambda_l$  分别设置为 2、2、1,  $\gamma$  设置为 10 时, 推荐效果最佳。

### 4.4 结果分析

为了评估 ATG 模型的性能, 将其与 LRT 模型、UTE + SE 模型进行对比分析。

#### 4.4.1 参数 $\alpha$ 的取值

ATG 模型首先通过用户潜在兴趣点数据模型挖掘潜在签到记录, 然后将其与用户签到记录一并填充至偏好矩阵中, 以降低数据稀疏性的影响。然而, 用户对于潜在兴趣点和已签到过的兴趣点的偏好是不同的, 因此, 在式(19)中设置参数  $\alpha$  作为用户对于潜在兴趣点的偏好值。下文将着重分析参数  $\alpha$  的值对推荐效果的影响。

图 3、图 4 分别给出推荐结果的准确率和召回率随  $\alpha$  值的变化情况。由图 3 和图 4 可知, 准确率和召回率会随着  $\alpha$  的不同而发生改变, 当  $\alpha = 0.6$  时,

两者达到最大值,当  $\alpha$  大于 0.6 时,推荐效果逐渐下降。这是因为潜在签到记录是根据用户潜在兴趣点数据模型学习而来的,和实际数据存在差距,当  $\alpha$  的值太大时,填充潜在签到记录不但不能提高推荐效果,反而产生了噪音,影响推荐效果。当  $\alpha$  的值太小时,潜在签到记录无法产生很大的作用,失去其意义。因此,参数  $\alpha$  的取值对本文模型的推荐效果有重要影响。

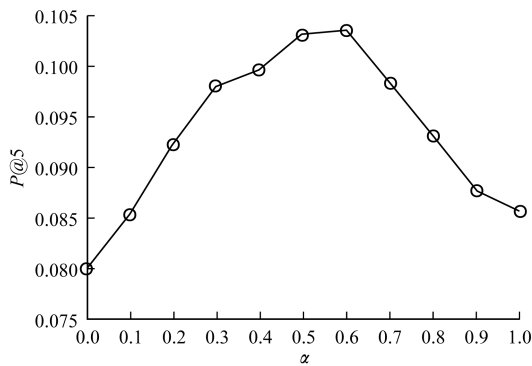


图3  $\alpha$  值对 Top 5 推荐准确率的影响

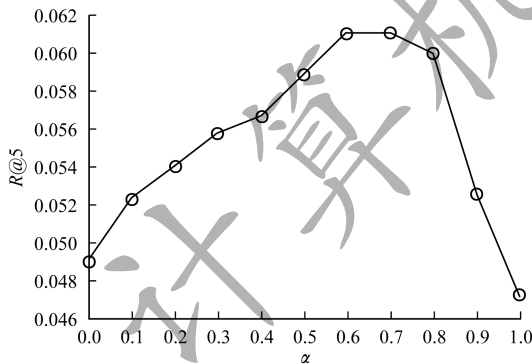


图4  $\alpha$  值对 Top 5 推荐召回率的影响

#### 4.4.2 推荐模型比较与分析

在  $k=5, 8, 10, 12, 15, 20$  的情况下,准确率和召回率分别用  $P@k, R@k$  表示,各模型的准确率和召回率如表 2、图 5、图 6 所示。

表 2 3 种模型在 Foursquare 数据集上的推荐效果对比

指标	LRT 模型	UTE + SE 模型	ATG 模型
$P@5$	0.054 6	0.087 8	0.103 4
$P@8$	0.046 4	0.064 7	0.077 4
$P@10$	0.042 3	0.056 8	0.068 3
$P@12$	0.037 6	0.051 3	0.061 0
$P@15$	0.032 6	0.045 6	0.053 3
$P@20$	0.028 1	0.039 2	0.046 1
$R@5$	0.034 4	0.055 3	0.062 0
$R@8$	0.046 1	0.065 9	0.075 6
$R@10$	0.053 0	0.072 8	0.082 7
$R@12$	0.057 0	0.078 8	0.088 5
$R@15$	0.061 7	0.087 1	0.096 5
$R@20$	0.070 6	0.098 9	0.111 8

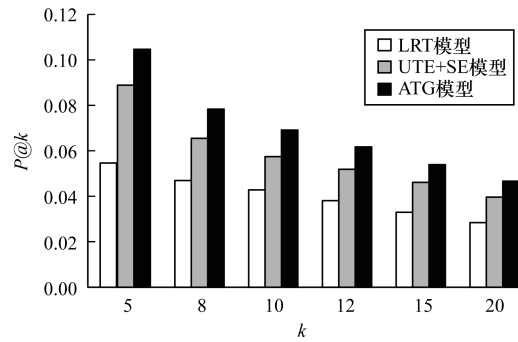


图5 3 种模型基于 Foursquare 数据集的准确率对比

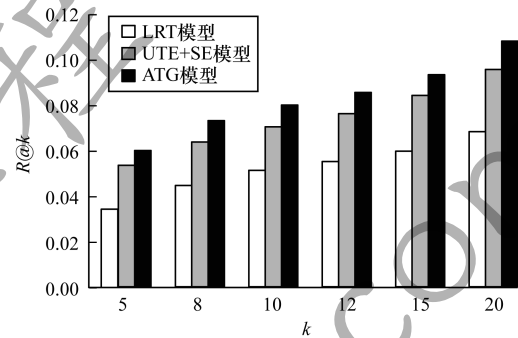


图6 3 种模型基于 Foursquare 数据集的召回率对比

从图 5、图 6 可知,在 3 种模型中,LRT 模型的推荐效果较差,ATG 模型的推荐效果较好,其原因如下:

1) LRT 是一个时间增强的矩阵分解模型,其只考虑了时间信息的影响,没有考虑地理位置和社会关系的影响。同时,简单地根据时间槽划分偏好矩阵,易使数据更加稀疏,对学习用户偏好造成负面影响。因此,LRT 模型表现出较差的推荐效果。

2) UTE + SE 是一个融合时间感知、地理位置因素的协同过滤模型,其采用了平滑相似度的方法,可有效缓解数据稀疏性及相似度无效计算的问题,改善推荐效果,但由于协同过滤算法易受数据稀疏性的影响,因此 UTE + SE 模型的推荐效果介于 LRT 模型和 ATG 模型之间。

3) ATG 是一个融合时间信息和地理位置信息的混合推荐模型,将协同过滤算法和矩阵分解算法有效结合,既考虑了用户时间行为的相似程度,又采用平滑相似度和矩阵填充的方法,有效地缓解数据稀疏性问题,考虑了时间行为相似性、时间流行度、时间差异性、时间连续性和地理位置对兴趣点实时推荐的影响,因此,ATG 模型的推荐效果较好。

## 5 结束语

本文提出一种混合推荐模型 ATG,通过平滑用户时间行为相似度和矩阵填充的方式降低数据稀疏性的影响。综合考虑时间相似度、时间流行度、时间差异性和时间连续性的因素,实现在正确时间内为目标用户推荐合适的兴趣点集。实验结果表明,与

LRT模型、UTE+SE模型相比,ATG模型的准确率和召回率较高。下一步将对异地场景推荐的问题进行研究,同时考虑评论信息等上下文信息对推荐效果的影响,进一步提高推荐准确率和召回率。

#### 参考文献

- [1] 曹玖新,董羿,杨鹏伟,等. LBSN中基于元路径的兴趣点推荐[J]. 计算机学报,2016,39(4):675-684.
- [2] 余永红. 融合多源信息的推荐算法研究[D]. 南京: 南京大学,2017.
- [3] GAO Huiji, TANG Jiliang, HU Xia, et al. Exploring temporal effects for location recommendation on location-based social networks [C]//Proceedings of ACM Conference on Recommender Systems. New York, USA: ACM Press,2013:93-100.
- [4] LIU Yiding, CONG Gao, YUAN Quan, 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.
- [5] 宋亚伟,司亚利,刘文远,等. 融合时间特征和协同过滤的兴趣点推荐算法[J]. 小型微型计算机系统,2016,37(6):1153-1158.
- [6] YUAN Quan, CONG Gao, MA Zongyang, et al. Time-aware point-of-interest recommendation[C]//Proceedings of International ACM SIGIR Conference on Research and Development in Information Retrieval. New York, USA: ACM Press,2013:363-372.
- [7] ZHENG Ning, JIN Xiaoming, LI Lianhao. Cross-region collaborative filtering for new point-of-interest recommendation[C]//Proceedings of International Conference on World Wide Web. New York, USA: ACM Press, 2013:45-46.
- [8] LIAN Defu, ZHAO Cong, XIE Xing, et al. GeoMF: joint geographical modeling and matrix factorization for point-of-interest recommendation [C]//Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, USA: ACM Press,2014: 831-840.
- [9] LI Xutao, CONG Gao, LI Xiaoli, et al. Rank-GeoFM: a ranking based geographical factorization method for point of interest recommendation [C]//Proceedings of International ACM SIGIR Conference on Research and Development in Information Retrieval. New York, USA: ACM Press,2015:433-442.
- [10] 高榕,李晶,杜博,等. 一种融合情景和评论信息的位置社交网络兴趣点推荐模型[J]. 计算机研究与发展,2016,53(4):752-763.
- [11] LI Huayu, GE Yong, HONG Richang, et al. Point-of-interest recommendations: learning potential check-ins from friends[C]//Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, USA: ACM Press,2016:975-984.
- [12] YE Mao, YIN Peifeng, LEE W C, et al. Exploiting geographical influence for collaborative point-of-interest recommendation [C]//Proceeding of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York, USA: ACM Press,2011:325-334.
- [13] LIU Yong, WEI Wei, SUN Aixin, et al. Exploiting geographical neighborhood characteristics for location recommendation [C]//Proceedings of the 23rd ACM International Conference on Information and Knowledge Management. New York, USA: ACM Press, 2014: 739-748.
- [14] CHENG Chen, YANG Haiqin, LYU M R, et al. Where you like to go next: successive point-of-interest recommendation [C]//Proceedings of International Joint Conference on Artificial Intelligence. Reston, USA: AIAA Press,2013:2605-2611.
- [15] XIANG Liang, YUAN Quan, ZHAO Shiwan, et al. Temporal recommendation on graphs via long-and short-term preference fusion [C]//Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, USA: ACM Press,2010:723-732.
- [16] YE Mao, JANOWICZ K, LEE W C. What you are is when you are: the temporal dimension of feature types in location-based social networks[C]//Proceedings of ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. New York, USA: ACM Press,2011:102-111.
- [17] 任星怡,宋美娜,宋俊德. 基于用户签到行为的兴趣点推荐[J]. 计算机学报,2017,40(1):28-51.
- [18] YANG Dingqi, ZHANG Daqing, ZHENG V W, et al. Modeling user activity preference by leveraging user spatial temporal characteristics in LBSNs [J]. IEEE Transactions on Systems, Man and Cybernetics Systems, 2014,45(1):129-142.