

基于社交信任和标签偏好的景点推荐方法^①

陈烨天, 米传民, 肖琳

(南京航空航天大学 经济与管理学院, 南京 211106)

通讯作者: 肖琳, E-mail: xiaolin@nuaa.edu.cn

摘要: 针对现有的景点推荐算法在处理用户关系时忽视了用户隐性信任和信任传递问题, 以及当用户处于新城市时由于缺乏用户历史记录无法做出准确推荐的情况, 本文提出一种综合用户信任关系和标签偏好的个性化景点推荐方法. 在仅仅考虑用户相似度时推荐质量差的情况下引入信任度, 通过挖掘用户隐性信任关系解决了现有研究在直接信任难以获取时无法做出推荐的情况, 有效缓解了数据稀疏性和冷启动问题. 同时在用户兴趣分析过程中将景点和标签的关系扩展到了用户、景点和标签三者的相互关系, 把用户的兴趣偏好分解成对不同景点标签的长期偏好, 有效地缓解了缺乏用户历史游览记录时推荐质量不佳的问题. 通过在 Flickr 网站上收集的数据进行实验验证, 结果表明本文提出的混合推荐算法有效地提高了推荐精度, 在一定程度上缓解了冷启动和新城市问题.

关键词: 个性化推荐; 信任度; 标签; 用户兴趣; 景点推荐

引用格式: 陈烨天, 米传民, 肖琳. 基于社交信任和标签偏好的景点推荐方法. 计算机系统应用, 2018, 27(9): 10-17. <http://www.c-s-a.org.cn/1003-3254/6503.html>

Attraction Recommendation Algorithm Based on Social Trust and Tag Preferences

CHEN Ye-Tian, MI Chuan-Min, XIAO Lin

(College of Economics and Management, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China)

Abstract: In view of that the existing attractions recommendation algorithm ignores the user's implicit trust and trust transfer when dealing with users' relationships, and the difficulties of making accurate recommendation for users in the new city due to the lack of user history records, this paper presents a personalized attraction recommendation algorithm based on users' social trust and tag preferences. According to user's rating behavior and context information, user's implicit trust is tapped, and the trust among users is obtained through trust transfer, which effectively alleviates the data sparsity. Then, by analyzing the relationships among users, attractions, and tags, the user's preference is decomposed into the preference of different attraction labels to further explore the user's long-term interest preferences. Experimental results on the data collected on Flickr website show that the hybrid recommendation algorithm proposed in this study effectively improves the accuracy of recommendation and relieves cold start and new city problems to a certain extent.

Key words: personalized recommendation; trust; tag; user's interest; attraction recommendation

随着国民生活水平的提高和互联网的快速发展, 我国在线旅游发展迅猛, 2016年中国在线旅游市场交易规模已达6026亿, 同比增长34%^[1]. 人们可以在各种在线旅游平台上挑选自己喜欢的旅游产品并制定相应

的出行计划. 然而严重的信息过载问题, 使得用户很难便捷地从海量的检索结果中获取自己真正需要的旅游信息^[2]. 此外, 目前国内的在线旅游电商平台提供的服务项目单一, 旅游路线、景点介绍几乎都是固定内容,

① 基金项目: 国家社会科学基金 (17BGL055)

Foundation item: National Social Science Foundation of China (17BGL055)

收稿时间: 2018-01-05; 修改时间: 2018-01-31; 采用时间: 2018-02-06; csa 在线出版时间: 2018-07-26

推荐系统的搜索和推荐结果多以产品热度排序,无法满足用户个性化的需求^[3]。尤其是在旅游景点的推荐上,景点热度和时间因素会对用户造成极大地干扰。用户缺乏旅游经验,面对过于广泛和缺乏人性化的推荐信息,难以获取较高的满意度^[4]。

为解决上述问题,学者们针对个性化旅游推荐技术做了广泛的研究,主要集中在基于社交媒体的推荐上^[5-7]。但区别于图书、音乐等物品的推荐^[8],由于旅游景点众多,用户购买频次低,导致数据稀疏性极大,传统的推荐方法难以解决冷启动和“新城市”等问题^[9]。本文通过引入用户信任关系来缓解数据稀疏性和冷启动问题,根据用户社交行为和影响力衡量显性信任,利用上下文信息挖掘隐性信任,并通过信任传递计算用户间接信任,构建完整的信任网络。在信任推荐的基础上,进一步融入用户-标签兴趣信息,提出一种结合社交信任和地理标签的 STGT 推荐算法 (recommended algorithm combined with Social Trust and Geo-Tags)。实验结果表明,本文提出的算法能在一定程度上提高推荐精度,同时缓解新用户和新城市等问题。

1 相关工作

个性化旅游推荐系统通过挖掘用户数据,获得用户的潜在需求和未来喜好,从而主动为用户推荐满足需求或感兴趣的信息,不仅有效解决了信息过载问题,而且满足了用户的个性化需求^[10]。不同于其他个性化推荐,个性化旅游推荐面临的挑战更大^[11]: (1) 旅游数据更复杂; (2) 用户更难准确表达自己的需求; (3) 大多数用户的历史旅游记录和偏好信息都很少; (4) 旅游数据更稀疏。传统的推荐方法如协同过滤推荐等效果不佳,数据稀疏性、冷启动问题严峻,还有新城市推荐问题^[5]。

为解决上述问题,学者们在传统的推荐方法上做了不少改进。从获取用户个性化信息方式角度可分为两类: (1) 用户交互方式。如 Ahas 等人^[12]通过邀请用户填写问卷的方式获取用户偏好; 麻风梅^[13]以会话的方式在线获取用户偏好和需求,结合用户浏览行为得到综合兴趣度进行推荐。这类基于知识、会话式的推荐方法不需要用户历史数据,故不存在冷启动问题,对非注册用户也能做出个性化推荐,但数据稀疏性问题依旧无法解决。另一方面,这种方式无法得到大量用户的数据,时间和操作成本高,能使用的用户训练样本小,

无法满足日益庞大的数据量处理需求。(2) 用户产生的社交媒体数据。这些社交媒体数据包括签到数据 (check-in)^[14]、GPS 轨迹信息^[15]、旅游博客^[6]等。这类方法能获得大量的用户历史数据和消费行为数据,覆盖面广,得到的结果更具有普适性。近年来,随着互联网技术的不断提高,带有地理信息的图片数据越来越受到研究者的关注^[7]。以 Flickr、Panoramio 为代表的图片分享网站的兴起,为研究者提供了海量的用户旅游照片。这些照片附带地理位置、拍摄时间、标签等重要信息,本身就体现了用户的旅游兴趣偏好,为个性化推荐带了极大地便利^[16]。

随着用户个性化需求的不断提高,学者们利用地理图片社交媒体这一数据时,分析用户偏好的方法主要分成两类: (1) 利用上下文信息分析用户偏好,包括用户年龄等个人信息、消费成本、天气、时间和季节等时空信息^[17]。如 Memon 等人^[18]综合考虑用户偏好和时间因素,根据用户过去时间内在旧城市的旅游偏好推荐新城市的旅游产品。Yin 等人^[15]进一步区分了游客偏好和当地居民偏好,利用当地居民偏好缓解数据稀疏性和新城市问题。这些研究在传统用户相似性度量假设用户未评分项为 0 的基础上,通过添加附加约束条件缓解数据稀疏性问题,但没有从根本上解决这个问题。因此有研究者提出在计算用户相似性时,先用相似性度量方法计算未评分稀疏项和其他项目的直接相似性,增加共同评分项来缓解稀疏性。但这类研究集中在电影、音乐数据,在旅游景点推荐方面的效果有待检验。(2) 利用标签信息。之前的研究大多使用用户评分数据,但相同的评分不能代表兴趣真正相同。标签信息可以进一步直观地揭示用户对景点本身的兴趣点。如 Lu 等人^[7]结合地理手册从地理信息图片中得到目的地后,引入目的地风格标签,结果比只用流行游览时间等上下文信息更精确。虽然这些研究通过标签细化了用户兴趣偏好,但大多简单地考虑了标签-景点关系,进行景点预过滤筛选处理,忽视了标签-标签关系、用户-标签关系。

此外,部分学者还引入了社交信任关系来缓解传统协同过滤算法的弊端,通过综合信任度和相似度,提高推荐精度^[19]。如史一帆^[20]提出的基于用户社交关系和物品标签的协同过滤推荐,把与目标用户相似度大于 0 或信任度大于 0 的用户当做相邻用户。但这些研究大多聚焦在图书、音乐、电影等推荐领域,涉及旅

游照片、游记等类型数据的旅游推荐研究极少. 刘艳等人^[3]通过用户景点照片矩阵计算用户偏好, 结合好友亲密信任关系计算相似度, 但只是简单判断了好友间是否有关关注关系, 忽视了用户交互行为和信任的传递问题. 在现实推荐中, 社交用户间的直接信用值获取也是一个难题.

因此, 本文基于图片数据, 根据用户社交行为、社交影响力和上下文信息获取用户直接信任和隐性信任关系, 在考虑信任传递的基础上挖掘用户间接信任, 最后综合相似度和信任度得到推荐结果. 为缓解用户信任好友稀缺时推荐效果不佳的问题, 引入用户标签偏好, 提出一个混合社交信任和标签兴趣的 STGT 景点推荐算法. 实验结果表明, 该算法能够有效利用多维信息, 提高推荐精度.

2 基于社交信任的景点推荐模型

2.1 构建信任网络

利用图片媒体数据挖掘用户信任关系的关键是构建信任网络: 给定一个用户-景点评分模型, 如何表示用户集合中用户间的直接信任关系. 直接信任指直接相连的用户间的信任关系, 可进一步分为显性信任和隐性信任.

(1) 显性信任

显性信任可通过用户社交互动行为和社交影响力获取. 在本文中, 当用户 i 和用户 j 存在好友关系, 且用户 i 对用户 j 有点赞等互动行为, 则认为两个用户间存在显性信任关系, 用 $trust_{i,j}^{ex}$ 表示, 计算公式为:

$$trust_{i,j}^{ex} = \alpha_1 \frac{r_j}{\min(R)} + (1 - \alpha_1) \frac{f_j}{\min(F)} \quad (1)$$

其中, $R = (r_1, r_2, \dots, r_n)$, r_j 表示用户 i 对用户 j 发布的照片等状态的点赞数; $F = (f_1, f_2, \dots, f_n)$, f_j 表示用户 j 在社交网络中的被关注数. α_1 取 0.75, 将较大权重分配给用户社交互动行为.

(2) 隐性信任

当两个用户为非好友时, 即 $trust_{i,j}^{ex} = 0$, 但不代表他们是非信任关系, 因此需要判断计算他们之间是否有隐性信任关系. 本文通过计算用户共同游玩景点的上下文情景信息相似度, 来推导用户隐性信任, 计算如下:

$$trust_{i,j}^{im} = \alpha_2 sim_{time}(i, j) + (1 - \alpha_2) sim_{weather}(i, j) \quad (2)$$

其中, $trust_{i,j}^{im}$ 表示用户 i 和用户 j 之间的直接隐性信任

关系, sim_{time} 是指游玩时间相似度, 而 $sim_{weather}$ 是指游玩天气相似度. α_2 取 0.5.

$$sim_{time}(i, j) = \frac{\sum_{k \in L(i) \cap L(j)} \frac{1}{|t_{ik} - t_{jk}| + 1}}{\sqrt{|L(i)||L(j)|}} \quad (3)$$

其中, $sim_{time}(i, j)$ 表示用户 i 和 j 的游玩时间相似度, $L(i)$ 表示用户 i 游玩过的景点集合, $L(j)$ 表示用户 j 游玩过的景点集合, $k \in L(i) \cap L(j)$ 表示用户 i 和 j 共同游玩过的景点, t_{ik} 表示用户 i 游玩景点 k 的时间, t_{jk} 表示用户 j 游玩景点 k 的时间.

$$sim_{weather}(i, j) = \frac{\sum_{k \in L(i) \cap L(j)} equal(i, j, k)}{\sqrt{|L(i)||L(j)|}} \quad (4)$$

$$equal(i, j, k) = \begin{cases} 1 & CI_i = CI_j \\ 0 & CI_i \neq CI_j \end{cases}$$

其中, $sim_{weather}(i, j)$ 表示用户 i 和 j 的游玩天气相似度, 上下文天气信息用 $CI = (Temperature, Condition)$ 表示, 包含气温 $Temperature = \{hot, warm, cold\}$, 天气状况 $Condition = \{rainy, sunny, cloudy\}$.

综合前面的显性信任和隐性信任关系, 得到用户直接信任值 $trust_{i,j}^{dir}$ 计算方式如下:

$$trust_{i,j}^{dir} = \begin{cases} trust_{i,j}^{ex} & trust_{i,j}^{ex} \neq 0 \\ trust_{i,j}^{im} & trust_{i,j}^{ex} = 0 \end{cases} \quad (5)$$

得到 $trust_{i,j}^{dir}$ 就能构建信任网络, 过程如下: 遍历每个用户和其他用户间的直接信任值, 与全局信任阈值 $threld_{all}$ 比较, 若 $trust_{i,j}^{dir}$ 大于 $threld_{all}$, 说明用户间直接相连, 反之则说明没有直接相连, 由此检查每个用户的连接情况. 网络中的每个节点分别代表一个用户, 连接节点每条边代表用户间相互连接, 边上的权值代表直接信任度, 边的方向代表信任方向.

2.2 计算间接信任

信任具有传递性, 即当用户 a 信任 b , 用户 b 信任 c 时, 认为用户 a 对 c 也存在着信任关系^[21]. 因此仅仅考虑用户直接信任值构建的信任网络并不能完全反应用户的信任关系, 还需要考虑信任在网络中传播时的间接信任值. 关于信任传播比较经典的算法是 TidalTrust^[22] 和 MoleTrust^[23] 算法. 本研究借鉴 TidalTrust 算法思想, 做出了改进. 通过考虑隐性信任解决了用户显性信任难以获取时无法进行信任度计算的问题, 并引入信任最长传播距离解决了 TidalTrust 算法中舍弃高信任值的长距离路径问题.

首先需要为每个节点用户设置一个路径信任阈值, 将目标用户对所有邻居用户的信任值的平均值作为路径信任阈值 $pthreld_s$:

$$pthreld_s = \frac{\sum_{f \in TF(s)} trust_{s,f}^{dir}}{|TF(s)|} \quad (6)$$

其中, $TF(s)$ 表示用户 s 的所有直接信任好友, $trust_{s,f}^{dir}$ 表示用户 s 对直接信任好友 f 的直接信任值. 在信任传递过程中, 这个路径信任阈值将过滤所有信任值低于此阈值的路径.

TidalTrust 算法中取目标节点到源节点的最短路径 d 作为信任最长传播距离, 这有可能舍弃路径长但信任度高的路径. 因此, 文献[24]中信任传播距离的研究:

$$L = \frac{\ln n}{\ln m} \quad (7)$$

其中, n 表示信任网络中的用户总数, m 表示网络中用户的邻居用户的平均数, 即平均度数. 本文在信任最长传递距离的选择上取 $MD = \max(L, d)$. 如果用户 A 不能到达用户 B , 则 B 对 A 就没有信任价值. 但如果 A 能在最大传递距离内达到 B , 则意味着 B 对 A 来说具有一定的信任价值[25].

选择完信任路径和传递距离后, 通过加权源节点与直接信任节点间的信任值, 得到源节点对目标节点的间接信任值, 公式如下:

$$trust_{s,e}^{indir} = \frac{\sum_{\substack{i \in TF(s) \& \\ trust_{s,i}^{dir} > pthreld_s}} \omega \times trust_{s,i}^{dir}}{\sum_{i \in TF(s) \& trust_{s,i}^{dir} > pthreld_s} trust_{s,i}^{dir}} \quad (8)$$

其中, 加权因子是通过比较直接信任节点到目标节点的最大路径信任值确定的, 如公式 (9) 所示. $route(i, e)$ 表示用户 i 到用户 e 的所有路径集合, $(x, y) \in r$ 指路径 r 上的任意两个用户 x 和 y ,

$\min_{\substack{(x,y) \in r \& \\ trust_{x,y} > paththreld_x}} (trust_{x,y}^{dir})$ 表示单条路径 r 上的两个用户间信任值的最小值.

$$\omega = \max_{r \in route(i,e)} \left\{ \min_{\substack{(x,y) \in r \& \\ trust_{x,y} > paththreld_x}} (trust_{x,y}^{dir}) \right\} \quad (9)$$

综合上述计算, 信任网络中任意两个用户间的传递距离在信任最长传播距离内时, 他们之间存在信任关系: 当距离为 1 时, 两用户直接相连, 取直接信任值

作为信任度; 当距离大于 1 且不超过最长传播距离时, 取间接信任值作为用户间信任度. 当用户间距离大于最长传播距离时, 信任度为 0.

2.3 结合相似度和信任度的推荐策略

在传统的协同过滤推荐算法中, 目标用户对目标项目的评分可以通过加权计算目标用户的相邻用户打出的评分来进行预测. 但随着系统中用户和项目数量的增加, 数据稀疏性越来越高, 以及新用户和新项目带来的冷启动问题大大降低了传统协同过滤算法的推荐精度[26]. 研究证实在推荐过程中引入信任关系能缓解上述协同过滤推荐算法中的主要问题, 有效地提高推荐结果的准确度. Massa[23]提出的 MoleTrust 模型则采取通过信任关系过滤相邻用户, 信任度替代用户相似度的策略. 经典的 TidalTrust 算法则利用信任关系过滤相邻用户, 再通过信任权重加权评分的方式进行评分预测.

这些基于社交信任关系的推荐算法虽然提高了推荐结果的准确度, 但推荐结果的覆盖率却比不上经典的协同过滤推荐计算方式. 因此本文采用综合协同过滤推荐和信任推荐的推荐策略:

$$P_{trust}(a, i) = \bar{r}_a + \frac{\sum_{u \in N^T} t_{a,u}(r_{u,i} - \bar{r}_u) + \sum_{u \in N^+ \setminus N^T} sim_{a,u}(r_{u,i} - \bar{r}_u)}{\sum_{u \in N^T} t_{a,u} + \sum_{u \in N^+ \setminus N^T} sim_{a,u}} \quad (10)$$

其中, $P_{trust}(a, i)$ 表示目标用户 a 对景点 i 的预测评分, N^+ 表示和目标用户 a 相似度高的相邻用户集, $t_{a,u}$ 表示用户 a 和用户 u 间的信任度, N^T 表示对目标项目 i 有过评分且信任度 $t_{a,u}$ 超过阈值的用户集合, $N^{T+} = N^T \cap N^+$. $r_{u,i}$ 表示用户 u 对项目 i 的评分, \bar{r}_u 是用户 u 的项目平均评分, \bar{r}_a 是用户 a 的项目平均评分. $sim_{a,u}$ 表示用户 a 和用户 u 间的用户相似度, 通过 Pearson 相关系数计算得到.

当目标用户和其他用户之间有直接或间接的信任关系时, 则使用信任度代替相似度作为加权计算的权重, 通过优先考虑信任用户的方式来保留信任推荐的优势. 另一方面, 覆盖率可以通过考虑无法获取信任信息的那部分相似用户保证.

3 考虑信任关系和标签偏好的 STGT 景点推荐算法

虽然基于信任的推荐往往能获得较高的满意度和

覆盖率,但现实中信任关系一般比较稀疏,好友数量不足,导致推荐效果不理想.因此本文在考虑信任关系的基础上,融入用户偏好,通过计算用户对标签的偏好来分析用户兴趣,进一步挖掘了用户的个性化需求.

3.1 基于标签的用户偏好建模

传统的协同过滤通过用户评分来表现用户兴趣偏好,受到数据稀疏性影响推荐结果不佳.标签信息的引入可以很好地细分用户兴趣点,通过计算用户对标签的偏好度能很好地表示用户的兴趣点分布,缓解数据稀疏问题,同时在面对新用户时也能通过获取标签偏好来缓解冷启动问题.

(1) 计算景点-标签关联度

首先我们定义用户集合 $U = \{u_1, u_2, \dots, u_n\}$, 景点集合 $L = \{l_1, l_2, \dots, l_m\}$, 标签集合 $T = \{t_1, t_2, \dots, t_k\}$, L_u 表示用户 u 浏览过的景点集合, T_u 表示用户 u 使用过的标签集. 一个景点会被很多用户标记标签,在得到用户的标签偏好前需要在景点和标签之间建立联系,评价一个标签在景点资源中的重要性.文献[27]表明, sigmoid 函数能最好地表示标签的权重,用 $W(l, t)$ 表示景点-标签关联度,即:

$$W(l, t) = \frac{1}{e^{-tag_quality(l, t)}} \quad (11)$$

其中, $tag_quality(l, t)$ 可用 TF 算法算得,表示标签 t 在景点 l 的标签集中出现的比重.

(2) 计算用户-标签关系

本文从用户标签行为和用户评分行为两方面衡量用户标签喜好度:一方面考虑了用户标记标签时间,另一方面当标签在评分越高的景点出现,表示用户对标签的兴趣越高.用 tag_pref_{ini} 表示用户 u 对标签 t 喜好度,计算如下:

$$tag_pref_{ini}(u, t) = \frac{\sum_{l \in L_u} W(l, t) \times (r_{u, l} - \bar{r}_u) \times time}{\sum_{l \in L_u} W(l, t) \times time} \quad (12)$$

其中 L_u 表示用户 u 浏览过的景点集合.

$$time = \frac{1}{1 + |time_period| \cdot \varphi} \quad (13)$$

$time_period$ 表示到用户标记标签的时间到现在的时间间隔, φ 取 5×10^{-10} .

(3) 计算标签-标签关系

当用户标签行为和评分较少时,标签稀疏性较高,

为解决这个问题,本文采取了拓展标签的方法,为每个用户的标签集增加相关度高的相关标签.首先计算标签相关度:

$$rel(t, t') = \frac{|user(t) \cap user(t')|}{N(t, t')} \quad (14)$$

其中, $|user(t) \cap user(t')|$ 表示同时使用过标签 t 和 t' 的用户数, $N(t, t')$ 表示标签集中同时含有标签 t 和 t' 的景点数.按相关度对标签排序,选取 Top M 个相关标签作为相关标签,得到新的用户标签集 T'_u .

用户对相关标签的喜好度可通过加权相关标签喜好度得到:

$$tag_pref_{ex}(u, t) = \frac{\sum_{t_j \in T_u} rel(t, t_j) \times tag_pref_{ini}(u, t_j)}{\sum_{t_j \in T_u} rel(t, t_j)} \quad (15)$$

综合 tag_pref_{ini} 和 tag_pref_{ex} 可得用户的标签喜好度 $tag_pref_{u, t}$.

3.2 综合信任关系和用户偏好的推荐策略

由于基于信任的推荐结果未充分考虑用户兴趣,本文通过对用户标签喜好度的分析进一步进行了用户兴趣建模.因此在最终的推荐策略上,本文选择加权混合推荐的方式融合两种推荐方法的优势,通过加权因子 λ 将两个预测结果线性组合, λ 的取值由实验结果确定.预测公式如下:

$$P(a, i) = \lambda P_{trust}(a, i) + (1 - \lambda) P_{tag}(a, i) \quad (16)$$

其中, $P_{trust}(a, i)$ 指基于信任关系的推荐得到的预测评分,计算方法在公式(10)中已给出. $P_{tag}(a, i)$ 指基于标签兴趣的推荐得到的预测评分,计算公式如下:

$$P_{tag}(a, i) = \frac{\sum_{t \in T_{(i)}} W(i, t) \times tag_pref_{a, t}}{\sum_{t \in T_{(i)}} W(i, t)} + \bar{r}_a \quad (17)$$

其中, $T_{(i)}$ 表示景点 i 包含的标签集.最后根据 $P(a, i)$ 对推荐景点列表进行降序排序,选取 Top K 个结果作为推荐结果.

4 实验验证和分析

4.1 实验数据集

Flickr 是一个提供免费及付费数位照片存储、分享服务的图片网站,重要特点就是基于社会网络的人际关系拓展与内容的组织.在当前的景点推荐领域中,很多推荐方法都会把 Flickr 数据集作为实验数据

集进行研究. 本文的数据集就是选用 Flickr 公开的图像数据集. 城市的天气信息通过第三方天气服务网站的 API 爬取. 对得到的数据集进行预处理, 删除那些元数据中的地点信息和实际经纬度不符以及带有错误时间的照片. 利用文献[28]中的 P-DBSCAN 方法对照片的经纬度进行聚类得到景点, 去除游览景点数少于 3 个的用户, 去除包含图片数少于 10 张的景点, 最后的统计结果如表 1 所示. 为检验算法在解决“新城市”问题上的效果, 选择至少去过四个城市中两个城市的用户, 且其在训练集中的游览景点数不少于 5 个. 实验时随机将数据集的 80% 作为训练集, 20% 作为测试集.

表 1 实验数据集统计信息

城市	照片数	用户数	标签数	景点数
北京	211 608	43 465	225 362	438
上海	236 331	72 831	247 864	512
杭州	64 358	2476	71 771	168
南京	48 537	2110	49 411	139

4.2 评价指标

本研究采用的评价指标是推荐系统中常用的精确度 (Precision) 和平均绝对偏差 (MAE). MAE 用来衡量推荐算法的预测评分和实际评分之间的差异程度, 用户对景点的评分可通过用户在景点拍摄的照片数量占比表示. MAE 的值越小, 代表推荐的准确度越高, 推荐效果越好, 定义如下:

$$MAE = \frac{\sum_{i=1}^n |p_{a,i} - r_{a,i}|}{n} \quad (18)$$

其中预测用户 a 对景点 i 的评分为 $p_{a,i}$, 用户对景点的实际评分是 $r_{a,i}$, n 为测试集的景点数量.

精确度是 Top- n 推荐问题的重要衡量指标, 用 $R(u)$ 表示根据训练集预测的推荐列表, $T(u)$ 表示用户在测试集上的实际访问景点列表, 计算公式如下:

$$Precision = \frac{\sum_{u \in U} |R(u) \cap T(u)|}{\sum_{u \in U} |R(u)|} \quad (19)$$

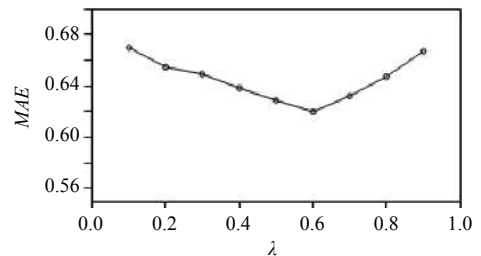
4.3 实验结果分析

实验一. 确定加权因子 λ

为了选择最佳的加权参数使得混合推荐的效果达到最好, 我们先固定推荐数量和用户最近邻数都为 25 进行实验, 通过改变 λ 的取值观察 MAE 指标的变化. λ 从 0.1 取到 0.9, 依次递增 0.1. 实验结果如图 1 所示.

从图 1 中可看出, 随着加权因子 λ 的逐渐增大,

MAE 的值先逐渐减小再逐渐增大. 当 λ 的取值为 0.6 时, MAE 的值达到最小, 推荐质量达到最佳. 因此我们在后面的实验中取 $\lambda=0.6$, 来确定用户最近邻数量和不同算法推荐效果的比较.

图 1 λ 取值对实验结果的影响

实验二. 和其他推荐算法的比较

为了验证本文提出融合信任关系和用户偏好的 STGT 推荐算法的准确性和推荐效果, 将本文提出的算法和考虑标签信息和社交关系的 UKS 算法^[20]、基于信任和项目偏好的 TIPCF 算法^[26]以及 MFT-Co^[29]信任推荐算法进行对比. 固定推荐列表长度为 15, 拓展标签的数量取 20, 通过改变最近邻数量观测 MAE 和 Precision@15 值的变化. 实验结果如图 2、图 3 所示, 横坐标为最近邻数目个数, 纵坐标为 MAE 和 Precision@15 的值.

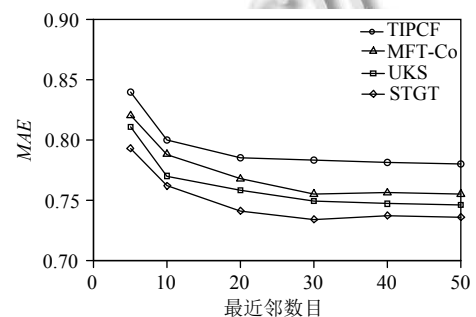


图 2 各算法 MAE 值对比

从图 2 实验结果中可看出, 最近邻数目确实是影响推荐质量的因素, 随着最近邻数目的增大, 各推荐算法的 MAE 值下降明显, 当数目达到 30 左右时 MAE 的值趋于稳定. 说明最近邻用户与目标用户的相似度和信任度已经接近阈值, 此时继续增大最近邻数目, 增加的近邻用户已经不具备较好的推荐参考价值 (相似度或信任度低于阈值), 导致推荐效果趋于平缓. 从整体上看, STGT 算法的推荐效果明显优于其他三种算法, 而

UKS 的推荐质量比 MFT-Co 和 TIPCF 好。

从图 3 实验结果可知, 本文提出的 STGT 算法在精确度上有明显提高, 比 TIPCF、MFT-Co 和 UKS 算法分别提高了 14.01%、10.07%、6.62%。从增长趋势看, 随着最近邻数目的增长, 精确度的提升趋势放缓并趋于稳定, 继续加入相似度或信任度低的近邻用户无法再提高推荐精确度。

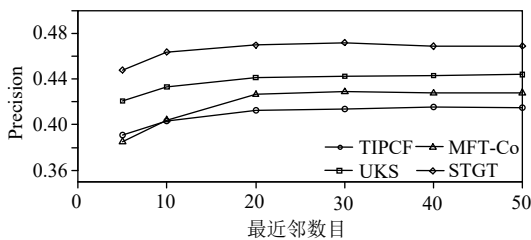


图 3 各算法 Precision@15 值对比

为验证本文提出的算法在缓解冷启动问题上的效果, 随机抽取 10 000 名用户进行实验, 并随机去除 150 个景点在训练集中的用户访问信息, 模拟冷启动环境。从图 4 的实验结果可知, STGT 算法对于系统中新项目的应对能力和推荐效果明显优于其他三种算法, 随着最近邻数目的增加, 算法性能趋于稳定, 有效地缓解了冷启动问题。

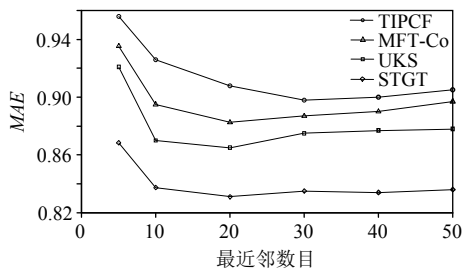


图 4 冷启动环境下各算法推荐效果对比

从以上分析可得出以下结论: (1) 由于本文提出的 STGT 算法更准确地衡量了用户间的信任关系, 通过综合考虑用户评分行为和兴趣品偏好挖掘出用户隐性信任关系, 缓解了数据稀疏性的同时获得了更准确的近邻用户。(2) 本文不仅考虑了其他算法使用的标签-景点关系, 还进一步考虑了标签-标签关系和用户-标签关系, 将用户对景点的兴趣偏好分解成用户对标签的偏好, 有效地解决了在新城市中无法获取用户兴趣偏好的问题, 提高了推荐准确性。(3) 当新用户或新项目进入系统时, 通过利用用户信任关系和标签偏好, 在一定

程度上缓解了冷启动问题。

5 结论与展望

现有的景点推荐算法在考虑用户间关系时主要利用了社交网络或好友关系获取用户间直接信任, 忽视了用户隐性信任和信任传递问题; 而在考虑用户和景点关系时主要从用户评分和天气、时间等上下文信息计算用户相似性进行推荐, 当用户处于新城市时由于缺乏用户历史记录无法做出准确推荐。针对这些问题, 本文提出一种综合信任关系和标签偏好的景点推荐方法 (STGT 方法)。首先结合用户评分行为和上下文信息挖掘用户隐性信任, 构建信任网络, 融合相似度和信任度做出推荐。然后通过全面考虑用户、景点和标签之间的关系, 把用户的兴趣偏好分解成对不同景点标签的偏好, 更准确地获取了这些代表用户长期兴趣的偏好, 有效地缓解了数据稀疏性, 在新城市问题中能做出更准确的推荐。通过在 Flickr 网站上收集的数据, 设计了对比实验。实验结果表明本文提出的算法有效地提高了推荐准确度, 在一定程度上缓解了冷启动和新城市问题。

未来的工作包括考虑用户间信任的动态变化, 以及用户标签存在的同义、多义问题, 标签本身存在的问题可能会影响推荐质量。此外, 可以获取其他类型的旅游信息, 如游记等, 进一步挖掘用户兴趣, 提高用户最近邻搜索精度。

参考文献

- 1 杨玉琼. 中国在线旅游市场的发展研究. 中国商论, 2017, (5): 57-58. [doi: 10.3969/j.issn.1005-5800.2017.05.026]
- 2 Ricci F, Rokach L, Shapira B, et al. Recommender Systems Handbook. Berlin, Heidelberg: Springer, 2011. 1-35.
- 3 刘艳, 潘善亮. 基于 LBSN 好友关系的个性化景点推荐方法. 计算机工程与应用, 2015, 51(8): 117-122. [doi: 10.3778/j.issn.1002-8331.1409-0181]
- 4 Bobadilla J, Ortega F, Hernando A, et al. Recommender systems survey. Knowledge-Based Systems, 2013, 46: 109-132. [doi: 10.1016/j.knosys.2013.03.012]
- 5 Majid A, Chen L, Chen GC, et al. A context-aware personalized travel recommendation system based on geotagged social media data mining. International Journal of Geographical Information Science, 2013, 27(4): 662-684. [doi: 10.1080/13658816.2012.696649]
- 6 Pang YW, Hao Q, Yuan Y, et al. Summarizing tourist

- destinations by mining user-generated travelogues and photos. *Computer Vision and Image Understanding*, 2011, 115(3): 352–363. [doi: [10.1016/j.cviu.2010.10.010](https://doi.org/10.1016/j.cviu.2010.10.010)]
- 7 Lu X, Wang CH, Yang JM, *et al.* Photo2Trip: Generating travel routes from geo-tagged photos for trip planning. *Proceedings of the 18th ACM International Conference on Multimedia*. New York, NY, USA. 2010.
- 8 Mi CM, Shan XF, Ma J. Improved network-based recommendation algorithm. *I-Business*, 2014, 6(3): 109–116.
- 9 Zhang CY, Wang K. POI recommendation through cross-region collaborative filtering. *Knowledge and Information Systems*, 2016, 46(2): 369–387. [doi: [10.1007/s10115-015-0825-8](https://doi.org/10.1007/s10115-015-0825-8)]
- 10 Borràs J, Moreno A, Valls A. Intelligent tourism recommender systems: A survey. *Expert Systems with Applications*, 2014, 41(16): 7370–7389. [doi: [10.1016/j.eswa.2014.06.007](https://doi.org/10.1016/j.eswa.2014.06.007)]
- 11 Kabassi K. Personalizing recommendations for tourists. *Telematics and Informatics*, 2010, 27(1): 51–66. [doi: [10.1016/j.tele.2009.05.003](https://doi.org/10.1016/j.tele.2009.05.003)]
- 12 Ahas R, Aasa A, Roose A, *et al.* Evaluating passive mobile positioning data for tourism surveys: An estonian case study. *Tourism Management*, 2008, 29(3): 469–486. [doi: [10.1016/j.tourman.2007.05.014](https://doi.org/10.1016/j.tourman.2007.05.014)]
- 13 麻风梅. 基于游客综合兴趣度的旅游景点推荐. 测绘与空间地理信息, 2014, 37(3): 55–56, 61. [doi: [10.3969/j.issn.1672-5867.2014.03.016](https://doi.org/10.3969/j.issn.1672-5867.2014.03.016)]
- 14 Liu JJ, Huang Z, Chen L, *et al.* Discovering areas of interest with geo-tagged images and check-ins. *Proceedings of the 20th ACM International Conference on Multimedia*. New York, NY, USA. 2012.
- 15 Yin HZ, Sun YZ, Cui B, *et al.* LCARS: A location-content-aware recommender system. *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA. 2013.
- 16 Jiang SH, Qian XM, Mei T, *et al.* Personalized travel sequence recommendation on multi-source big social media. *IEEE Transactions on Big Data*, 2016, 2(1): 43–56. [doi: [10.1109/TBDATA.2016.2541160](https://doi.org/10.1109/TBDATA.2016.2541160)]
- 17 时念云, 李秋月. 基于情境感知的个性化推荐算法. *计算机系统应用*, 2017, 26(9): 135–139. [doi: [10.15888/j.cnki.csa.005931](https://doi.org/10.15888/j.cnki.csa.005931)]
- 18 Memon I, Chen L, Majid A, *et al.* Travel recommendation using geo-tagged photos in social media for tourist. *International Conference on Computer & Emerging Technologies*, 2014.
- 19 林建辉, 严宣辉, 黄波. 融合信任用户的协同过滤推荐算法. *计算机系统应用*, 2017, 26(6): 124–130. [doi: [10.15888/j.cnki.csa.005805](https://doi.org/10.15888/j.cnki.csa.005805)]
- 20 张朝恒, 何小卫, 陈勇兵. 基于社交网络信息的协同过滤推荐算法. *计算机技术与发展*, 2017, 27(12): 28–34. [doi: [10.3969/j.issn.1673-629X.2017.12.007](https://doi.org/10.3969/j.issn.1673-629X.2017.12.007)]
- 21 时念云, 张芸, 马力. 基于综合相似度和社交标签的推荐算法. *计算机系统应用*, 2017, 26(10): 178–183. [doi: [10.15888/j.cnki.csa.006025](https://doi.org/10.15888/j.cnki.csa.006025)]
- 22 Golbeck JA. Computing and applying trust in web-based social networks [Ph.D thesis]. College Park, MD: University of Maryland, 2005.
- 23 Massa P, Avesani P. Trust metrics on controversial users: Balancing between tyranny of the majority and echo chambers. *International Journal on Semantic Web and Information Systems*, 2007, 3(1): 1–21. [doi: [10.4018/IJSWIS](https://doi.org/10.4018/IJSWIS)]
- 24 Yuan WW, Guan DH, Lee YK, *et al.* The small-world trust network. *Applied Intelligence*, 2011, 35(3): 399–410. [doi: [10.1007/s10489-010-0230-7](https://doi.org/10.1007/s10489-010-0230-7)]
- 25 彭鹏, 米传民, 肖琳. 基于用户信任和兴趣的概率矩阵分解推荐方法. *计算机系统应用*, 2017, 26(9): 1–9. [doi: [10.15888/j.cnki.csa.005933](https://doi.org/10.15888/j.cnki.csa.005933)]
- 26 郑洁, 钱育蓉, 杨兴耀. 基于信任和项目偏好的协调过滤算法. *计算机应用*, 2016, 36(10): 2784–2788, 2798. [doi: [10.11772/j.issn.1001-9081.2016.10.2784](https://doi.org/10.11772/j.issn.1001-9081.2016.10.2784)]
- 27 Sen S, Vig J, Riedl J. Tagommenders: Connecting users to items through tags. *Proceedings of the 18th International Conference on World Wide Web*. New York, NY, USA. 2009.
- 28 Kisilevich S, Mansmann F, Keim D. P-DBSCAN: A density based clustering algorithm for exploration and analysis of attractive areas using collections of geo-tagged photos. *Proceedings of the 1st International Conference and Exhibition on Computing for Geospatial Research & Application*. New York, NY, USA. 2010.
- 29 陈梅梅, 薛康杰. 基于标签簇多构面信任关系的个性化推荐算法研究. *数据分析与知识发现*, 2017, 1(5): 94–101.