

一种基于用户网络签到行为的地点推荐方法

周而重¹ 黄佳进¹ 徐欣欣²

(北京工业大学国际 WIC 研究院 北京 100124)¹ (北京工业大学材料科学与工程学院 北京 100124)²

摘要 基于位置的地点推荐服务日益强调用户的个性化需求。对此,可通过社交网站上用户与其好友之间、用户与签到地点以及地点与地点之间的关联,来从用户的网络签到行为中总结出用户的出行特点,从而提出一种基于用户网络签到行为的个性化地点推荐方法。该方法通过融合用户对地点的个人偏好程度、地点自身属性对用户的影响程度以及用户好友对地点的推荐程度,来筛选出候选地点中满足用户个性化需求的地点。实验结果验证了该方法在一定场景下的可行性和有效性。

关键词 地点推荐,个人偏好,地点影响力,好友推荐

中图法分类号 TP391 文献标识码 A DOI 10.11896/j.issn.1002-137X.2015.10.047

Approach to Place Recommendation Based on User Check-in Behavior in Online Network

ZHOU Er-zhong¹ HUANG Jia-jin¹ XU Xin-xin²

(International WIC Institute, Beijing University of Technology, Beijing 100124, China)¹

(College of Material Science and Engineering, Beijing University of Technology, Beijing 100124, China)²

Abstract The location-based service for place recommendation pays more attention to the user's personalized need. Hence, the characteristics of user trip were extracted through the links between entities in online social network, such as the social tie between users, interaction between users and places, and proximity between places. An approach to personalized place recommendation based on user check-in behavior in online social network was consequently proposed. The approach ranks the candidate places to meet user's personalized need by combining the user preference to the place, influence of the place on the target user, and social recommendation from user's friends. Experimental results show that the proposed approach is feasible and effective in a given context.

Keywords Place recommendation, User preference, Influence of place, Friend recommendation

1 引言

随着手机与互联网的完美结合,人们的社会需求促使基于位置的服务软件逐渐应用到生活的各个方面。例如手机商务软件已可以帮助用户寻找周边适合的消费场所;社交软件可以根据用户的好友在社交平台上的网络签到数据,帮助用户获知其好友当前的位置和行为。由于困扰上述软件服务性能的瓶颈主要来自于用户不断增长的个性化诉求,而凭借用户有限的历史信息又不足以预测用户的潜在偏好,因此研究者们开始尝试借助用户的社交网络来推测用户的偏好,从而事先主动为用户进行个性化的信息筛选。针对地点推荐问题,由于用户对出行地点的选择会受外界及个人偏好的影响,因此用户的出行模式是有规律的。用户在社交媒介上的签到行为在一定程度上反映了用户在各种因素共同作用下对出行地点某些属性或特征的偏好。对此,本文依据用户在基于位置的社交媒介上的签到行为及其社交圈的组成,提出一种基于用户网络签到行为的地点推荐方法。

2 相关工作

与传统在线社交网络相比,基于位置的社交网络的显著特点是可以从时间和空间(经度和纬度)属性上来展现每位用户的出行路线。针对基于位置的社交网络,如何为用户进行个性化的地点推荐的研究已经取得了很大的进展。在基于位置的社交网络中,人与人、人与地点以及地点与地点间的关联已被研究者作为地点推荐的重要因素进行评估。立足于人与人以及人与地点间的关联,基于用户的协同过滤方法依据相似用户对同一地点有相同爱好这一假设,通过发现与目标用户兴趣相似的若干用户来帮助其筛选出可能感兴趣的地点^[1]。研究者 Symeonidis 等发现用户更倾向于采纳好友的意见而非外界全部意见^[2]。基于朋友间的社交影响,研究者 Ye 等提出了基于好友的地点推荐方法,即在目标用户的朋友圈内采用基于用户的协同过滤方法来评估目标用户对其未曾访问过的地点的喜好程度^[3]。由于大多数用户访问的地点非常有限,因此数据稀疏性问题严重制约了协同过滤方法的推

到稿日期:2014-10-11 返修日期:2015-03-16 本文受国家自然科学基金项目(61420106005)资助。

周而重(1981—),男,博士,助理研究员,主要研究方向为 Web 智能、机器学习, E-mail: zez2008@emails.bjut.edu.cn; 黄佳进(1977—),男,博士,讲师,主要研究方向为 Web 智能、数据挖掘; 徐欣欣(1980—),女,博士,讲师,主要研究方向为电磁功能材料设计。

荐效果。为应对数据稀疏性问题,研究者 Ye 等发现用户的活动轨迹多集中在其曾经访问过的地点周围^[4],而且用户的每次活动范围都是有限的,因此地点间的距离也被作为地点推荐的重要评估因素^[5]。同时当访问陌生区域时,用户不会轻易选择那些流行度不高和朋友没有推荐过的地点。依据该假设,越来越多的研究者从用户的个人偏好、地点的流行性及朋友推荐 3 个方面,来为用户推荐其未曾访问过的地点^[6-8]。本文提出的个性化地点推荐方法也从个人偏好、地点的影响力及朋友推荐 3 个方面来评估用户访问某一地点的可能性。与同类方法的不同之处是:评估地点对用户的影响程度时,本方法考虑了用户对出行区域的熟悉性;评估社交圈对用户地点选择的影响程度时,考虑了人的遗忘特点对其朋友推荐作用的影响。

3 基于用户网络签到行为的地点推荐方法

本文提出的方法通过分析用户自身的签到特点来评估其个人对访问过的地点的喜好程度,通过分析其好友的签到特点来评估好友们对地点的推荐程度,并根据全体用户的签到特点以及地点之间的邻近性来评估地点对用户的影响程度,从而筛选出候选地点中满足用户个性化需求的地点。

3.1 用户个人偏好评估

越来越多的研究发现,用户的个人偏好在地点选择时要明显强于其它因素^[4,6,8],个人偏好评估因此成为地点推荐方法中的重要一环。在基于位置的社交网络中,用户在某一地点的签到次数会在一定程度上体现出用户对该地点的偏好程度。同时用户签到行为的时间分布以及地点被全体用户访问的情况同样会反映出目标用户对该地点的偏好,即用户在目标地点的签到次数和签到天数越多,地点为其他用户签到的次数越少,越能体现出用户对目标地点的喜爱。本方法采用研究者任克江在其地点推荐方法中提出的用户偏好评估公式^[9]:

$$Pe(u_k, l_i) = \frac{C_{k,i}}{\sum_{l_j \in L_k} C_{k,j}} \times \frac{d_{k,i}}{d_k} \times \log \frac{\sum_{u \in U_i} C_{u,i}}{|U_i|} \quad (1)$$

式中, $C_{k,i}$ 为用户 u_k 访问候选地点 l_i 的次数, L_k 为用户 u_k 访问过的地点集合, l_j 表示 L_k 中第 j 个地点, $d_{k,i}$ 为用户 u_k 访问地点 l_i 的天数, d_k 为用户 u_k 出行的总天数, U_i 为访问过地点 l_i 的用户集, u 表示用户集 U_i 中的用户, $|U_i|$ 为访问过地点 l_i 的人数。

3.2 地点影响力评估

在出行路线安排上,用户未曾访问过的候选地点的流行度越高,用户越容易选择该地点。同时,用户的出行会明显受到其居住地和日程的约束,候选地点离用户熟悉区域越近,用户访问该地点的可能性越大。对此,本方法计算地点对用户的影响力时,重点评估地点的流行度、出行距离以及地点所在区域的用户熟悉度。信息熵已成为评估地点流行度的一项重要指标,其计算公式如下^[10]:

$$E(l_i) = - \sum_{u \in U_i} \left(\frac{C_{u,i}}{\sum_{u \in U_i} C_{u,i}} \times \log \frac{C_{u,i}}{\sum_{u \in U_i} C_{u,i}} \right) \quad (2)$$

地点的信息熵值越小,地点属于私人场所的概率就越大,流行度也越低。在度量用户对访问地区的熟悉程度时,本方

法计算用户访问过的全部地点与指定候选地点间的距离,来度量候选地点是否超出用户的日常活动范围。度量公式为:

$$D(u_k, l_i) = \prod_{l_k \in L_k} e^{-dis(l_i, l_k)} \quad (3)$$

式中, $dis(l_i, l_k)$ 为候选地点 l_i 和地点 l_k 间的距离, l_k 表示用户 u_k 访问过的地点集合 L_k 中的第 k 个地点。式(3)的值越大,出行地点离用户日常活动区域越远。

另一方面,离候选地点最近的用户访问过的地点的用户偏好程度值越大,用户对候选地点周边区域的熟悉程度也会越高。依据地点的流行性和用户对出行地区的熟悉度,地点 l_i 对用户 u_k 的影响力评估公式为:

$$Pop(u_k, l_i) = E(l_i) \times \frac{Pe(u_k, l_i)}{e^{dis(l_i, l_j)}} \times D(u_k, l_i) \quad (4)$$

式中, $l_j = \arg \min_{l_k \in L_k} dis(l_i, l_k)$, 即用户 u_k 访问过的地点集合 L_k 中距离候选地点 l_i 最近的地点。

3.3 好友推荐程度评估

用户好友在网络分享其出行经历的行为,同样会影响用户的出行选择。研究者 Cho 等发现用户容易在好友签到某地后拜访该地点^[11]。对此,基于好友的地点推荐方法采用了用户与其好友间的兴趣越相似,好友对地点兴趣值越高,则地点被推荐给目标用户的概率越大的思想。该方法中用户 u_k 的好友们对候选地点 l_i 的推荐程度计算公式如下^[3]:

$$R_{u_k, i} = \frac{\sum_{u_j \in U_k'} w_{j,k} \times C_{j,i}}{\sum_{u_j \in U_k'} w_{j,k}} \quad (5)$$

式中, U_k' 为用户 u_k 的朋友集合, u_j 为 U_k' 中的第 j 个用户, $C_{j,i}$ 为用户 u_j 访问候选地点 l_i 的次数, $w_{j,k}$ 为用户 u_j 与用户 u_k 间的兴趣相似度。

本方法同样借鉴基于好友的地点推荐方法,即在计算用户与好友间的兴趣相似度时,采用双方访问过的相同地点越多,则双方的兴趣越可能相似的原则。另一方面,用户的遗忘性会使得好友推荐的效果随时间的流逝而发生衰减,因此本方法评估用户好友在不同时间段内的推荐效果。用户 u_k 的好友们对候选地点 l_i 的推荐程度的计算公式为:

$$Sr(u_k, l_i) = \frac{\sum_{u_j \in U_k'} \frac{|L_j \cap L_k|}{\min(|L_j|, |L_k|)} \times \sum_{m=1}^{C_{j,i}} e^{-\Delta t_m}}{\sum_{u_j \in U_k'} \frac{|L_j \cap L_k|}{\min(|L_j|, |L_k|)}} \quad (6)$$

式中, $|L_j|$ 为用户 u_j 访问过的地点总数, $|L_j \cap L_k|$ 为目标用户 u_k 与其好友 u_j 共同访问过的地点数, $\min(x, y)$ 为数值 x 与 y 中的最小值, Δt_m 为用户 u_j 第 m 次访问候选地点 l_i 的时间距当前时间的间隔数。鉴于目标用户好友的朋友也可能与目标用户间存在相同的爱好,而且研究已发现双方的距离越近,双方成为朋友的概率越大^[3],因此用户 u_k 的朋友集合 U_k' 还包含了其全部好友的朋友。此外,在线社交网站是记录用户生活日志的重要媒介,其历史签到数据也是用户规划出行路线时的重要参考依据,因此本方法还考虑用户的历史签到数据对其自身的推荐作用,即用户 u_k 的朋友集合 U_k' 也包含了用户 u_k 自己。

3.4 个性化地点推荐模型

基于上述个人偏好、地点影响力及朋友推荐的评估公式,将 3 个公式进行线性组合后得到的地点推荐模型如下:

$$R(u_k, l_i) = (1 - \alpha - \beta) \times Pe(u_k, l_i) + \alpha \times Pop(u_k, l_i) + \beta \times Sr(u_k, l_i) \quad (7)$$

式中,系数 α 与 β 的取值范围为0至1,且 α 与 β 之和必须小于1。

由于各个评估公式计算出的数值处在不同的数量级上,因此本方法需对由式(1)计算所得的用户个人偏好、式(2)计算得到的地点流行度、式(3)评估出的地点偏离程度以及由式(6)计算得到的好友推荐程度值进行标准化。标准化的前提条件是各项指标的原始数值大于0。数值标准化方法采用如下所示的逻辑回归公式^[12]:

$$norm(x) = \frac{1}{1 + e^{-x}} \quad (8)$$

4 实验

实验数据为 Gowalla 社交网站上的用户签到数据和朋友圈数据^[11]。实验提取网站用户从2009年2月至2010年10月间在洛杉矶地区签到的154803条记录,涉及的用户人数为6550,同时根据数据集中提供的朋友圈数据来确定每位用户的好友。用户签到数据的属性包括:用户编号、签到时间、签到地点的经纬度及签到地点编号。朋友圈数据的属性包括:用户编号和其好友的用户编号。实验从访问过洛杉矶地区的6550人中选取了429名满足以下条件的用户:在洛杉矶地区签到次数超过5次;在当地最后一次出行时至少签到2个地点;最后签到的2个地点间距离不超过5千米。实验从对应洛杉矶地区的数据集中抽取该429名用户最后一次出行时的最后签到地点作为测试数据,其余数据作为训练数据,来验证本文提出的地点推荐方法的可行性。实验结果的评价指标为查全率和查准率,评估公式如下^[8]:

$$recall@N = \frac{\sum_{u \in U_t} |Q_u \cap M_u|}{\sum_{u \in U_t} |Q_u|} \quad (9)$$

$$precision@N = \frac{\sum_{u \in U_t} |Q_u \cap M_u|}{\sum_{u \in U_t} |M_u|} \quad (10)$$

式中, Q_u 为用户 u 真实签到的地点, M_u 为推荐模型为用户 u 推荐的前 N 个候选地点, U_t 为参与测试的用户集。实验中 N 的取值为3,5,10,20,30和50。

表1 本文提出的地点推荐方法的参数取值

参数组编号	α	β
1	0.1	0.3
2	0.3	0.6
3	0.4	0.4
4	0.6	0.3
5	0.5	0.1
6	0.3	0.1

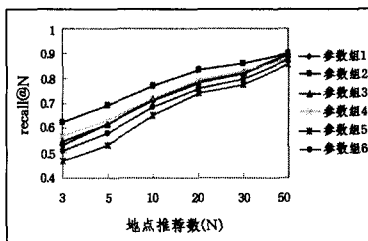


图1 参数调整时本文提出的方法的查全率

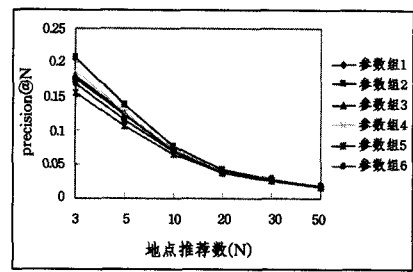


图2 参数调整时本文提出的方法的查准率

为确保本文提出的地点推荐方法的性能,实验首先需对式(7)中地点影响力和朋友推荐两项对应的系数 α 与 β 进行优化。系数 α 与 β 的取值见表1,不同参数值时地点推荐的查全率和查准率见图1和图2,从中可见参数组2对应的地点推荐效果为最佳,即 α 取值为0.3, β 取值为0.6。

为验证本文提出的方法的有效性,实验所选用的对比方法1为采用式(5)进行地点评估的基于好友的地点推荐方法。该方法中用户与其好友的相似度计算公式如下^[4]:

$$w_{j,k} = \eta \times \frac{|F_j \cap F_k|}{|F_j \cup F_k|} + (1 - \eta) \times \frac{|L_j \cap L_k|}{|L_j \cup L_k|} \quad (11)$$

式中, F_j 为用户 u_j 的朋友集合, L_j 为用户 u_j 访问过的地点集合,系数 η 取值在0至1之间。实验中系数 η 取值为0.5,目标用户的朋友集合除了包含其全部好友外,还包含了其好友的朋友。对比方法2为采用贝叶斯分类原理的基于距离的地点推荐方法^[5]。该方法将用户已签到的地点作为目标类,从而将地点推荐问题转化为候选地点的分类问题,即计算每个候选地点属于目标类的概率。用户 u_i 访问地点 l_j 的概率计算公式如下^[5]:

$$\Pr(l_j | L_i) = \prod_{l_y \in L_i} \frac{1}{dis(l_y, l_j)} \quad (12)$$

式中, L_i 为用户 u_i 访问过的全部地点。

3种方法在查全率和查准率上的实验结果如图3和图4所示。由于多数用户的活动范围有限,因此用户的朋友能推荐的地点十分有限,而且用户朋友的签到地点中往往还包含了私人社交场所,导致可供目标用户参考的地点更为稀少。同时用户好友们活动区域的不同,又严重降低了候选地点被推荐的概率。本文提出的方法考虑了候选地点对用户的影响,从而降低了数据稀疏性对推荐方法的制约。基于距离的地点推荐方法仅考虑了用户的出行距离和活动区域的影响,而本文提出的方法综合考虑了用户的个人偏好、地点的影响和朋友推荐等重要因素在其出行地点选择时的作用。因此本文提出的方法的地点推荐效果优于2种对比方法的地点推荐效果。

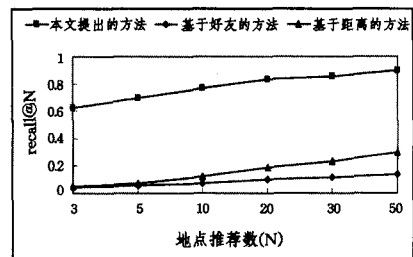


图3 3种地点推荐方法的查全率

(下转第261页)

decomposition recommender algorithm based on local structures [J]. Journal of Electronic & Information Technology, 2013, 35 (6):1284-1289

- [10] Cai Rui, Zhang Chao, Zhang Lei, et al. Scalable Music Recommendation by Search [C] // International Multimedia Conference, 2007:1065-1074
- [11] Andoni A, Indyk P. Nearest-optimal hashing algorithms for approximate nearest neighbor in high dimensions[J]. Communications of the ACM, 2008, 51(1):117-122
- [12] 高豪林, 彭天强, 李弼程, 等. 近似最近邻搜索算法——位置敏感哈希[J]. 信息工程大学学报, 2013, 14(3):332-340
- Gao H L, Peng T Q, Li B C, et al. Approximate nearest neighbor

searching algorithm—locality sensitive hashing [J]. Journal of Information Engineering University, 2013, 14(3):332-340

- [13] Slaney M, Casey M. Locality-sensitive Hashing for Finding Nearest Neighbors [J]. IEEE Signal Processing Magazine, 2008, 8 (3):128-131
- [14] Indyk P, Motwani R. Approximate nearest neighbors: towards removing the curse of dimensionality[C] // The Symposium on Theory of Computing, 1998:604-613
- [15] Datar M, Indyk P, Immorlica N, et al. Locality-Sensitive Hashing scheme Based on p-stable Distributions [C] // Proceedings of the Twentieth Annual Symposium on Computational Geometry, 2004:253-262

(上接第 234 页)

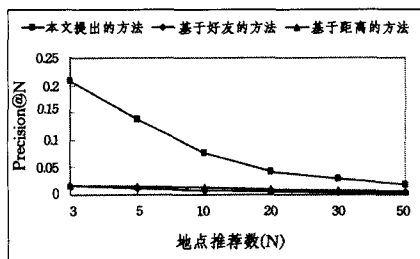


图 4 3 种地点推荐方法的查准率

结束语 本文根据用户在基于位置的社交网络上的签到行为及社交圈的组成,通过综合考虑用户的个人偏好、地点对用户的影响及其好友的推荐 3 个因素,来为用户进行个性化的地点推荐。实验结果说明,考虑用户对出行区域的熟悉性以及人的遗忘特点对其好友推荐作用的影响有助于提高地点推荐的效果。但本文提出的方法没有对用户的出行动机进行预测,即识别用户所在区域是否为其社交、购物或旅游观光的场所,以致不能事先对候选地点进行筛选,从而在一定程度上影响了方法的地点推荐效果。

参 考 文 献

- [1] Jia-Ching Y, Huan-Sheng C, Kawuu W L, et al. Semantic trajectory-based high utility item recommendation system[J]. Expert Systems with Applications, 2014, 41(10):4762-4776
- [2] Panagiotis S, Antonis K, Yannis M. GeoSocialRec: explaining recommendations in location-based social networks[C] // Proceedings of the 17th East-European Conference on Advances in Databases and Information Systems, Germany: Springer Verlag, 2013:84-97
- [3] Mao Y, Pei-feng Y, Wang-Chien L. Location recommendation for location-based social networks[C] // Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems, New York: Association for Computing Machinery, 2010:458-461
- [4] Mao Y, Pei-feng Y, Wang-Chien L, 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, New

York: Association for Computing Machinery, 2011:325-334

- [5] Quan Y, Gao C, Zong-yang M, et al. Time-aware point-of-interest recommendation[C] // Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval, New York: Association for Computing Machinery, 2013:363-372
- [6] Jia-Ching Y, Eric Hsueh-Chan L, Wen-ning K, et al. Urban point-of-interest recommendation by mining user check-in behaviors[C] // Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York: Association for Computing Machinery, 2012:63-70
- [7] Nai-Hung C, Chia-Hui C. Evaluation of social, geography, location effects for point-of-interest recommendation[C] // Proceedings of IEEE 13th International Conference on Data Mining Workshops, Los Alamitos: IEEE Computer Society, 2013:766-772
- [8] Jia-Ching Y, Wen-ning K, Vincent S T, et al. Mining user check-in behavior with a random walk for urban point of interest recommendations[J]. ACM Transactions on Intelligent Systems and Technology, 2014, 5(3):1-26
- [9] 任克江. 基于地理信息的检索和用户数据挖掘[D]. 大连:大连理工大学, 2013
- Ren Ke-jiang. Information retrieval and user data mining based on geographic information [D]. Dalian: Dalian University of Technology, 2013
- [10] 潘果, 徐雨明. LBSN 中位置信息与网络拓扑相融合的好友预测[J]. 计算机科学, 2014, 41(9):115-118
- PAN Guo, Xu Yu-ming. Friends predication based on fusion of topology and location in LBSN[J]. Computer Science, 2014, 41 (9):115-118
- [11] Eunjoon C, Seth A M, Jure L. Friendship and mobility: user movement in location-based social networks[C] // Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York: Association for Computing Machinery, 2011:1082-1090
- [12] 李秀艳. 多生物特征身份识别方法研究[D]. 天津:天津大学, 2010
- Li Xiu-yan. Research on personal identity recognition method based on multi-biometric[D]. Tianjin: Tianjin University, 2010