

LBSN 中融合信任与不信任关系的兴趣点推荐

朱敬华, 明骞

(黑龙江大学计算机科学与技术学院, 黑龙江 哈尔滨 150080)

摘 要: 兴趣点 (POI, point of interest) 推荐是位置社交网络 (LBSN, location-based social network) 重要的个性化服务, 广泛用于热门景点推荐和旅游线路规划等。传统的基于协同过滤的推荐算法根据用户相似性和位置相似性进行推荐, 未考虑推荐用户与目标用户间的信任关系, 而信任关系有助于提高推荐系统的准确性、顽健性和用户满意度。首先分析了信任与不信任关系的传播特征, 然后给出了信任度的表示和计算方法, 最后提出了融合用户相似性、地理位置相似性和信任关系的混合推荐模型。实验结果表明, 与传统协同过滤推荐方法相比, 融合信任关系的混合推荐方法显著提高了推荐结果的准确性和用户满意度。

关键词: 位置社交网络; 兴趣点推荐; 协同过滤; 信任关系

中图分类号: TP311

文献标识码: A

doi: 10.11959/j.issn.1000-436x.2018117

POI recommendation by incorporating trust-distrust relationship in LBSN

ZHU Jinghua, MING Qian

Department of Computer Science and Technology, Heilongjiang University, Harbin 150080, China

Abstract: POI (point of interest) recommendation is an important personalized service in the LBSN (location-based social network) which has wide applications such as popular sights recommendation and travel routes planning. Most existing collaborative filter algorithms make recommendation according to user similarity and location similarity, they don't consider the trust relationship between users. And trust relationship is helpful to improve recommendation accuracy, robustness and user satisfaction. Firstly, the propagation property of trust and distrust relationship was analyzed. Then, the measurement and computation method of trust were given. Finally, a hybrid recommendation system which combined user similarity, geographical location similarity and trust relationship was proposed. The experiments results show that the hybrid recommendation is obviously superior to the traditional collaborative filtering in terms of results accuracy and user satisfaction.

Key words: LBSN, POI recommendation, collaborative filtering, trust relationship

1 引言

近年来, 随着智能手机的普及和移动互联网技术的飞速发展, 人们使用智能终端来获取位置信息变得越来越方便、准确。基于位置的社交网络 (LBSN,

location-based social network) 如 Foursquare、Gowalla 等逐渐兴起, 且受到广大用户的喜爱^[1]。与传统的社交网络相比, LBSN 中引入了地理位置信息, 用户可以对当前访问的餐厅、电影院等兴趣点 (POI, point of interest) 签到, 并与好友分享位

收稿日期: 2017-11-08; 修回日期: 2018-05-18

通信作者: 朱敬华, zhujinghua@hlju.edu.cn

基金项目: 国家自然科学基金资助项目 (No.61100048, No.61370222); 黑龙江省自然科学基金资助项目 (No.F2016034)

Foundation Items: The National Natural Science Foundation of China (No.61100048, No.61370222), The Natural Science Foundation of Heilongjiang Province (No.F2016034)

置等信息。LBSN 中包含用户和兴趣点这 2 类节点，如图 1 所示。其中， $u_1 \sim u_4$ 代表用户， $l_1 \sim l_6$ 代表兴趣点，从用户指向兴趣点的有向边表示用户在该地点的签到行为。LBSN 中主要包含 3 种关系：用户的友邻关系、用户与兴趣点的签到关系、兴趣点间的地理位置关系。

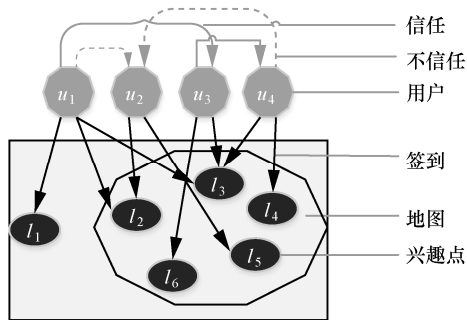


图 1 用户—兴趣点签到模型

LBSN 中的 POI 推荐是为用户推荐新的感兴趣位置，并为用户提供可信、可靠和高满意度的旅游体验信息。

传统的基于协同过滤 (CF, collaborative filtering) 的 POI 推荐方法存在数据稀疏、恶意推荐和扩展性问题^[2]。研究表明，社交网络中的信任关系能够体现用户间的兴趣相似性和影响力，而兴趣和信任度是正相关的^[3]。与系统的推荐相比，人们更喜欢好友的推荐。将社交信任关系融入推荐系统中，能够解决传统推荐方法的冷启动和数据稀疏问题。例如，文献[4]提出基于信任关系的推荐方法 CoSRA+T，该方法是对基于用户相似性的推荐方法 CoSRA 的改进，将用户的信任关系融入推荐系统中，能提高推荐的精度。GUO 等^[5-6]提出利用信任关系进行电影推荐，解决传统推荐方法的数据稀疏和冷启动问题。在 LBSN 中，除了社交信任关系外，用户的地理位置和偏好对 POI 推荐也有影响。例如，文献[7]提出基于用户偏好和背景嵌入的深度神经网络架构，学习用户和 POI 相关的各种背景环境来预测用户对目标 POI 的喜好程度。

虽然现有的 POI 推荐算法中结合了社交信任，但是并没有考虑用户间的不信任关系，关于不信任关系对 POI 推荐的作用研究很少。但是实际的应用系统如 Ebay、淘宝等的运营经验表明：在信任模型中引入不信任机制能够有效抑制用户的不诚信和恶意推荐。文献[8]指出不信任关系对推荐起重要的作用，并提出根据用户的偏好和信任与不信任关系

的推荐系统，实验结果表明，不信任关系对于推荐质量的提高有显著的作用。

综上所述，现有的 POI 推荐方法缺少对信任关系和不信任关系传递特性的研究，因此本文重点研究 LBSN 中基于信任和不信任关系的 POI 推荐方法。

如图 1 所示，用户间的实有向边表示信任关系， u_1 信任 u_3 ， u_3 信任 u_4 ；虚有向边表示不信任关系， u_1 和 u_4 都不信任 u_2 。本文研究了信任和不信任关系的传播特征，给出了信任度和不信任度的计算方法。在推荐系统中融入信任关系，改善推荐质量，规避不信任用户的恶意推荐。结合用户相似性、地理位置和信任关系构建了一个混合的 POI 推荐模型。实验结果表明，有信任关系的推荐结果的准确度相较传统的基于用户相似性的协同过滤方法有很大的提高，并且保证了推荐结果的可信度，且能够有效地抵御恶意推荐。

本文的主要贡献如下。

- 1) 将信任关系融合到兴趣点推荐中，分析了信任与不信任关系的传播特征，给出了用户间信任度和不信任度的计算方法。
- 2) 结合用户相似性、地理位置和信任关系构建了一个混合兴趣点推荐模型，计算目标用户在兴趣点的签到概率从而生成推荐列表。

2 相关工作

2.1 基于协同过滤的推荐

该类推荐方法利用历史签到信息进行推荐，协同过滤推荐可分为基于记忆的推荐和基于模型的推荐。Ye 等^[9]认为用户签到受个人喜好、好友影响和兴趣点远近距离 3 个方面影响，综合用户偏好、好友和地理信息提出混合推荐算法，明显提高了推荐结果的精准度。基于模型的推荐通过构建用户评分模型，预测和计算评分，生成推荐。Liu 等^[10]认为用户评分由用户偏好、兴趣点属性、兴趣点间距离共同决定，构建了贝叶斯图模型计算用户对候选兴趣点的喜好程度。Yang 等^[7]提出基于用户偏好和背景嵌入的深度神经架构，通过学习用户和 POI 相关的各种背景环境来预测用户对目标 POI 的喜好程度。利用泛化的矩阵分解方法和拉普拉斯正则化将协同过滤方法和半监督学习 (SSL, semi-supervised learning) 进行有效桥接。

上述推荐方法均需要用户的历史签到数据，且数据稀疏性对推荐质量有较大影响。

2.2 基于信任关系的推荐

Guo 等^[5-6]在推荐系统中融入了用户间的信任关系, 解决传统 CF 方法数据稀疏和冷启动问题。Zhang 等^[11]基于社交关系和地理因素构建了一个推荐模型, 根据用户签到记录、社交影响和地理影响建立组合模型, 利用 PageRank 算法改进模型的推荐精度。Lee 等^[8]在 CF 方法中融入信任和不信任关系, 利用矩阵分解和 k 近邻相结合的方法最大化用户评级信息和信任信息的效用, 实验结果表明, 不信任信息对预测用户评级有相当好的作用。Victor 等^[12]在基于信任的推荐系统中引入显式不信任, 可提高推荐结果的精准度。Ma 等^[13]研究推荐系统中隐式信任关系和不信任关系的预测问题。关于信任度的计算, Zhou 等^[14]选取了 6 个现实社交网络进行对比, 衡量利用局部信息相似性预测链路的效果, 结果显示共同邻居方法表现最佳。Symeonidis 等^[15]提出 FriendT-NS 推荐算法, 基本思想是使用邻居节点的度信息进行推荐, 比随机游走和最短路径等方法准确度高。上述方法为基于节点相似性计算直接信任度, 基于路径的方法主要用于计算间接信任度, 需要考虑单一路径的信任传递问题和多信任路径的信任合并问题。Tobler 等^[16]提出 TidalTrust 算法, 采用广度优先搜索计算信任度, 寻找源用户与目标节点之间的所有最短路径, 通过信任度权重计算用户对不直接相连目标用户的信任度。

上述工作与本文工作不同之处在于, 本文研究了信任与不信任关系的传递特征, 计算信任和不信任度的方法开销低。另外, 本文提出的混合推荐系统通过考虑多因素、用户相似性、地理位置邻近性和信任关系, 线性组合后得到推荐评级, 能更好地反映实际情况、提高推荐结果质量。

3 预备知识

3.1 基于用户相似性的推荐

一般而言, 相似用户的行为也有很高的相似性, 因此 CF 推荐方法根据与目标用户相似的用户行为预测目标用户潜在的偏好。在基于位置的社交网络中, U 和 L 分别表示用户集合和兴趣点集合, 用户 $u_i \in U$ 在地点 $l_j \in L$ 的签到记录为 c_{ij} 。 $c_{ij}=1$ 表示用户 u_i 在地点 l_j 签到过, $c_{ij}=0$ 表示 u_i 未去过 l_j 。根据相似用户的历史签到记录推测目标用户在候选兴趣点的签到概率, 用 $\hat{c}_{i,j}$ 表示用户 u_i 在候选兴趣点 l_j 的签到概率, 计算式为^[17]

$$\hat{c}_{i,j} = \frac{\sum_{u_k} w_{i,k} c_{k,j}}{\sum_{u_k} w_{i,k}} \quad (1)$$

其中, $w_{i,k}$ 表示用户 u_i 和 u_k 的相似性。

用户相似性的计算方法很多, 如余弦相似度、皮尔逊相关系数等, 本文采用余弦相似度计算用户 u_i 和 u_k 的相似性, 计算式^[9]为

$$w_{i,k} = \frac{\sum_{l_j \in L} c_{i,j} c_{k,j}}{\sqrt{\sum_{l_j \in L} c_{i,j}^2} \sqrt{\sum_{l_j \in L} c_{k,j}^2}} \quad (2)$$

3.2 基于地理位置的推荐

地理学第一定律认为, 人类的行为与所处的地域有很大的关系, 人们之间的距离越近, 关系就越紧密^[16]。在 LBSN 中, 用户的签到在地理上会有聚集的现象, 这种现象主要是因为用户在选择签到位置时会有 2 种倾向: 1) 倾向于去离家或公司较近的地点; 2) 倾向于去感兴趣的位置附近的地点, 即使这些地点可能离家较远。

本文利用 Power Law 分布和朴素贝叶斯方法计算兴趣点间距离在推荐中的影响。对于用户 u_i 和签到过的地点集合 L_i , L_i 内任意一对签到位置的概率为 $\Pr[d(l_m, l_n)] = a \cdot d(l_m, l_n)^b$, 其中, $d(l_m, l_n)$ 为兴趣点 l_m 和 l_n 之间的距离, 则用户 u_i 签到集合 L_i 中所有位置的概率为

$$\Pr[L_i] = \prod_{l_m, l_n \in L_i, l_m \neq l_n} \Pr[d(l_m, l_n)] \quad (3)$$

因此, 对于一个候选的兴趣点 l_j , 用户 u_i 在 l_j 签到的条件概率^[18]为

$$\Pr[l_j | L_i] = \prod_{l_y \in L_i} \Pr[d(l_j, l_y)] \quad (4)$$

计算除 L_i 集合外的所有候选兴趣点的签到概率为 $\Pr[l_j | L_i], (l_j \in L - L_i)$, 将最高概率的候选兴趣点推荐给用户。

4 基于局部信任关系的推荐

信任网络记录了用户间的信任关系, 用有向图 $D = \langle V, E \rangle$ 表示。 V 代表用户集合, E 代表信任关系即有向边集合。信任网络如图 2 所示, 其中, A, B, C, D 代表 4 个用户, 6 条有向边表示用户间的信任关系, 边上的数值表示信任度。信任网络与社交网络不同, 信任网络中的信任关系是单向信任的, A

信任 B ，不代表 B 也信任 A 。而社交网络中的好友关系是双向的， A 是 B 的朋友，则 B 也是 A 的朋友。

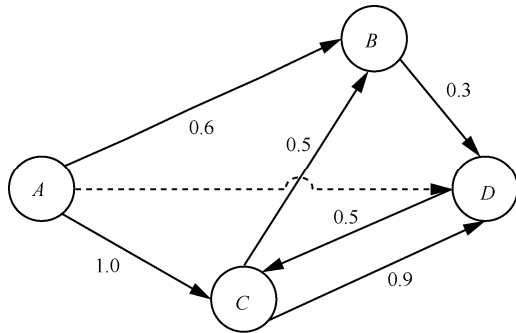


图 2 信任网络

在信任网络中，信任关系分为 2 种，直接信任和间接信任。在图 2 中，实线边表示用户间的直接信任关系，可以通过信任评价获得；虚线边表示间接信任关系，可以通过直接信任的传递获得。从图 2 可以看出，用户 A 直接信任用户 B 和用户 C ，根据信任的传递性^[17]推断出，用户 A 间接信任用户 D 。

本文同时考虑信任和不信任关系，下面分析它们的传递特性，给出信任度和不信任度的计算方法。

4.1 局部信任及其传递

根据用户信任度计算时涉及的相关用户的范围，信任分为局部信任和全局信任 2 种。与全局信任相比，局部信任关系由于具有个性化特征，因此更适合度量用户间的信任度^[10]。但是，计算局部信任度时需要计算所有用户对的信任度，全局信任度计算只为每个用户计算一次，因此局部信任的计算开销更大。在系统安全性方面，尽管 2 种信任都可能受到恶意攻击，但是由于局部信任的传递路径的起始节点由用户自主选取，因此可以降低受攻击的可能性^[16]。

本文考虑了 2 种类型的局部信任，局部信任关系和不信任关系。信任关系具有传递性，可以在社交网络中从一个用户传递到另一个用户。分析信任关系的传递，可以发现用户更多潜在的信息，从而对用户进行更准确的评估，以提高推荐的准确性。如果用户 u 信任用户 k ，用户 k 信任用户 v ，而 u 和 v 没有直接信任关系，则用户 u 对用户 v 的信任度 $W_{u,v}$ 为

$$W_{u,v} = \max_{k \in O(u)} W_{u,k} \times \frac{|O(u) \cap O(v)|}{|O(v)|} \quad (5)$$

其中， $O(u)$ 为被用户 u 信任的邻居用户集， $O(v)$ 为

被用户 v 信任的邻居用户集。

图 3 是一个简单的局部信任传播计算实例，用户间的连线表示信任关系。在本例中， $O(A)=\{B,C\}$ ， $O(S)=\{B,D\}$ ， $\max W_{A,k}=1$ ，则用户 A 对用户 S 的信任度为

$$1 \times \frac{|(B,C) \cap (B,D)|}{|(B,C)|} = 1 \times \frac{2}{3} = 0.666$$

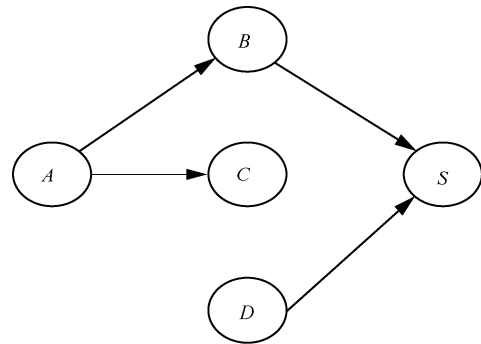


图 3 局部信任传播计算实例

4.2 局部不信任及其传递

社交网络中信任问题的研究主要集中于信任的计算和信任的传递，然而关于不信任关系的研究才刚刚起步，因此对不信任问题的研究相对较少。在推荐系统中利用不信任关系的算法更是少之又少，主要原因包括：1) 在社交网络中，很难获得不信任关系的数据；2) 不信任关系的传递性与信任关系的传递性不同，人们对于怎样利用不信任关系辅助推荐意见并不统一。

本节研究用户间的不信任关系及其传递特征。不信任关系与信任关系的传播理论相似，基本思想是利用直接的不信任用户计算非直接的不信任关系。采用 Model Trust 算法^[17]计算间接不信任关系时，为了计算不信任路径较长的间接不信任度，设置一个影响因子 d 来获得较长路径的不信任度，

$$ut_{u,v} = \frac{1}{d} ut'_{u,v} \quad (6)$$

其中， $ut'_{u,v}$ 是 Model Trust 算法计算得到的间接的不信任度， d 是通过宽度优先搜索方法获得的用户之间的最短路径长度。可以看出，不信任的影响力随传递路径长度的增加而衰减。不同数据集上的实验研究表明，2~3 级的传递适合进行传播效应的研究^[7]。社交网络中不信任关系传递超过 2 个用户后，潜在的不信任链已无法识别。因此，本文仅考虑长度为 2 的不信任关系传递路径，以减少计算的开销。

不信任的传递如图 4 所示, 以用户 A 到 C 的不信任传递为例。图 4 中 x_i 为每条传递路径的中间用户, $+$ 表示信任关系, $-$ 表示不信任关系。在信任社交网络中, 用户 A 到 C 的不信任传递表明在 A 和 C 之间存在一条潜在的、非直接的不信任链。而 A 与 C 之间可能存在多条二级通路, 因此计算不信任度时需要综合所有二级通路的结果。图 4 中包含 2 种类型的二级通路, 第一种类型的通路 A 不信任 B , 显著强化了 A 到 C 的不信任关系; 第二种类型通路 A 信任 B , 还需考虑 B 对 C 的信任关系, 才能正确计算 A 到 C 的间接不信任度。

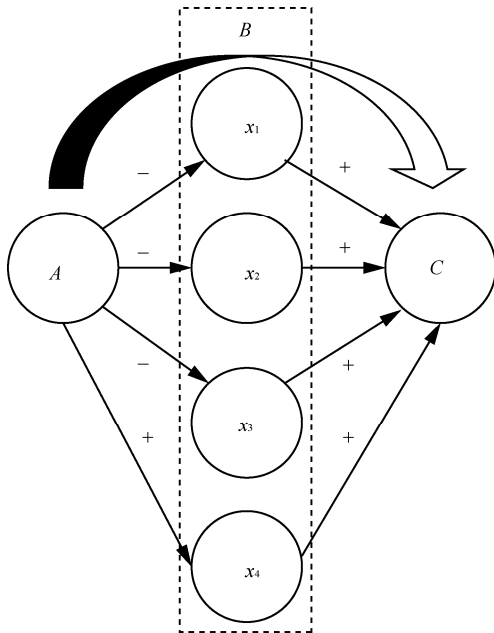


图 4 用户 A 到用户 C 的简单二级通路传递模型

社交平衡理论^[18]研究了 3 人间的友好和敌对关系。该理论认为, 类似于“我朋友的朋友是我的朋友”的社会关系比“我朋友的朋友是我的敌人”更为常见, 也更为稳定。则由社交平衡理论可知, 在只有 3 个用户的社交结构中, 用户 A 到用户 C 的二级通路中只有 2 种类型是可以稳定存在的, 第一种类型是用户 A 不信任中间用户 B , 中间用户 B 信任用户 C , 则 A 到 C 潜在的是不信任关系; 第二种类型是用户 A 信任中间用户 B , 中间用户 B 信任用户 C , 则 A 到 C 潜在的是信任关系。

根据上述分析可知, 当用户 A 到用户 C 之间不存在直接信任和信任关系时, 用户 A 对用户 C 的不信任度计算式为

$$W_{A,C}^- = \min_{x \in O_{out}(A)} W_{A,x}^- \frac{|O_{out}(A) \cap O_{in}(x^+)|}{|O_{out}(A^+)|} - \max_{x \in O_{out}(A)} W_{A,x}^+ \frac{|O_{out}(A^+) \cap O_{in}(x^+)|}{|O_{out}(A^+)|} \quad (6)$$

其中, x 表示用户 A 与用户 C 之间的中间用户; $O_{out}(A)$ 、 $O_{out}(A^+)$ 、 $O_{out}(A^-)$ 分别表示 A 的出邻居、 A 信任的邻居和 A 不信任的邻居; $O_{in}(x^+)$ 是信任中间用户 x 的用户集合。式(6)右边的第一个式子中的 \min 运算表示 A 不信任的邻居中的最小信任度 (不信任度为负值, 该值表示对最不信任用户的信任度), 第二个式子中的 \max 表示 A 信任的邻居用户中的最大信任度。

4.3 综合 2 种局部信任关系

综上所述, 在信赖朋友的影响下, 用户 u_i 去 l_j 的概率为

$$\hat{c}_{i,j} = \frac{\sum_{u_k \in F_i} W_{i,k} \times c_{k,j}}{\sum_{u_k \in F_i} W_{i,k}} \quad (7)$$

其中, 集合 F_i 表示用户 u_i 信赖的朋友集合, $W_{i,k}$ 表示用户 u_i 对用户 u_k 的信任度, $c_{k,j}$ 表示用户 u_k 在 l_j 的历史签到信息, $c_{k,j}=1$ 表示 u_k 去过地点 l_j , $c_{k,j}=0$ 表示其未去过地点 l_j 。

一般地, 推荐时人们更倾向于考虑直接信任, 因此当两用户之间同时存在 2 种信任关系时, 只利用直接信任关系进行推荐。

在信任社交网络中, 根据 Jaccard 系数方法可以计算出相邻用户间的信任度和不信任度。具有传递条件的非相邻用户, 根据社交平衡理论, 寻找到所有符合条件的二级通路, 计算信任和信任关系的传递, 即得非相邻用户间的信任度。非相邻用户 u_i 对用户 u_k 的信任度计算式为

$$W_{i,k} = \sum_{x \in F_i} W_{i,x}^+ + W_{i,k}^- \quad (8)$$

其中, $W_{i,x}^+$ 表示信任通路的信任值, 根据式(5)可以得到; $W_{i,k}^-$ 表示综合所有通路的不信任值, 根据式(6)可以得到; x 是用户 u_i 和用户 u_k 的中间用户。

5 融合信任关系的混合推荐系统

5.1 混合框架

在第 3 节中, 本文详细介绍了基于用户相似性和基于地理位置这 2 种推荐方法。这 2 种方法虽然

能够为用户推荐兴趣点，但是推荐的质量并不理想。现实中，人们更倾向于接纳信赖的朋友的推荐，因此本文在第 4 节详细分析了信任与不信任关系的传播和计算方法。在 LBSN 中，用户根据其所信任朋友的签到历史考虑去一个新地点的可能性，更符合实际。因此，本文将用户相似性、地理位置和信任关系这 3 种影响进行线性组合，确定用户 u_i 到兴趣点 l_j 的概率为

$$S_{i,j} = (1 - \alpha - \beta)S_{i,j}^s + \alpha S_{i,j}^g + \beta S_{i,j}^t \quad (9)$$

其中， $S_{i,j}^s$ 、 $S_{i,j}^g$ 、 $S_{i,j}^t$ 分别表示基于用户相似性、地理位置和信任关系得到的用户 u_i 去兴趣点 l_j 的概率。系数 α 和 β ($0 \leq \alpha + \beta \leq 1$) 为调节因子， $\alpha = 1$ 、 $\beta = 0$ 表示推荐系统仅考虑地理位置的影响； $\alpha = 0$ 、 $\beta = 1$ 表示推荐系统仅考虑信任关系的影响；其他情况表示推荐系统同时考虑用户相似性、地理位置和信任关系这 3 种影响。

对于伪造签到记录的恶意用户，即使该用户的虚假签到记录与目标用户高度相似，也不能对目标用户的 POI 推荐产生决定性的影响。根据式(9)，对目标用户的 POI 推荐依赖于用户相似性、地理相似性和信任关系。一个恶意用户无法通过建立直接和间接的朋友关系获得目标用户的足够信任，因此也无法对目标用户进行恶意推荐。

5.2 单因素签到可能性

根据上述混合框架，为了得到用户 u_i 去兴趣点 l_j 的可能性 $S_{i,j}$ ，需要分别计算 $S_{i,j}^s$ 、 $S_{i,j}^g$ 和 $S_{i,j}^t$ 的值，利用式(10)的 $p_{i,j}^s$ 、式(11)的 $p_{i,j}^g$ 和式(12)的 $p_{i,j}^t$ 来得到上述 3 个值。

根据式(1)可得基于用户相似的推荐概率为

$$p_{i,j}^s = \frac{\sum_{u_k} w_{i,k} c_{k,j}}{\sum_{u_k} w_{i,k}} \quad (10)$$

根据式(4)可得基于位置相似的推荐概率为

$$p_{i,j}^g = \Pr[l_j | L_i] = \prod_{l_y \in L_i} \Pr[d(l_j, l_y)] \quad (11)$$

根据式(7)可得基于信任关系的推荐概率为

$$p_{i,j}^t = \frac{\sum_{u_k \in F_i} w_{i,k} c_{k,j}}{\sum_{u_k \in F_i} w_{i,k}} \quad (12)$$

因此，有

$$S_{i,j}^s = \frac{p_{i,j}^s}{z_i^s}, \text{ 其中, } z_i^s = \max_{l_j \in L-L_i} \{p_{i,j}^s\}$$

$$S_{i,j}^g = \frac{p_{i,j}^g}{z_i^g}, \text{ 其中, } z_i^g = \max_{l_j \in L-L_i} \{p_{i,j}^g\}$$

$$S_{i,j}^t = \frac{p_{i,j}^t}{z_i^t}, \text{ 其中, } z_i^t = \max_{l_j \in L-L_i} \{p_{i,j}^t\}$$

5.3 生成推荐

根据上述构建的混合框架，生成推荐的具体步骤如下。

步骤 1 分别实现基于用户相似性、地理位置和信任关系的推荐。

步骤 2 使用步骤 1 的 3 个推荐，分别筛选出 top- k 个候选兴趣点集合。

步骤 3 根据混合框架线性组合的 3 组推荐结果，得到最终的推荐序列。

5.4 复杂性分析

本文提出的混合推荐系统需要计算 $S_{i,j}^s$ 、 $S_{i,j}^g$ 和 $S_{i,j}^t$ 。 $S_{i,j}^s$ 利用式(1)计算，用户的相似性采用余弦相似性计算方法，计算开销和传统协同过滤方法一样，为 $O(n^2d)$ ，其中， n 为用户数量， d 为候选兴趣点数量。 $S_{i,j}^g$ 利用式(4)计算，考虑所有兴趣点对之间的距离，因此计算开销为 $O(nd^2)$ 。 $S_{i,j}^t$ 为基于信任关系的推荐概率，需要计算直接和间接信任强度以及不信任度，根据式(5)、式(6)和式(8)可以得到，计算开销为 $O(n^3d)$ 。由于 3 种概率的计算是独立的，因此可以采用并行方式计算，推荐系统的总开销为 $O(n^3d)$ 。

6 实验对比与分析

本节通过实验分析，验证提出的混合推荐算法的性能：1) 分析推荐算法中的不同因素，如用户相似性、地理位置和信任关系，在推荐中各自的影响权重；2) 对比融合信任关系的推荐算法和传统算法的推荐结果的准确性以及算法的时间效率。

6.1 实验数据和参数设置

本文使用 2 个具有代表性的位置社交网络 (Foursquare 和 Gowalla) 的数据进行实验。数据集中主要有用户信息、签到记录、朋友关系以及兴趣点信息。为减少数据的稀疏性，挑选 2 个热门城市 (New York 和 Austin) 的数据进行实验。另外，对数据进行了预处理，从中剔除了签到次数少于 10 次且

好友数少于 5 人的用户，剔除被签到次数少于 5 的兴趣点。处理后的数据集，Foursquare 的数据集中包含 9 800 个用户、4 626 个兴趣点和 45 711 条签到记录；Gowalla 的数据集中包含 3 112 个用户、3 298 个兴趣点和 27 149 条用户的签到记录。从数据集中随机抽取 80% 的数据作为训练集，其余 20% 的数据作为测试集进行实验。实验赋予 α 和 β 不同的值，取值满足 $0 \leq \alpha + \beta \leq 1$ ，观察平均绝对误差 (MAE) 随 α 和 β 的变化情况，最终选取使 MAE 最小的值。当 $\alpha = 0.1$ ， $\beta = 0.2$ 时，推荐结果最优。若减少训练集规模， α 和 β 的最优配比略有不同。

6.2 评价指标

为评价兴趣点推荐算法的效果，本文采用推荐问题通用的指标：准确率和召回率。准确率指推荐结果中用户将来真正去的数量占推荐总数的比例，反映了推荐的准确率。假设在向目标用户推荐 N 个兴趣点后，目标用户去了推荐兴趣点中的 k 个兴趣点，则准确率 (precision) = $\frac{k}{N}$ 。

召回率指推荐结果中用户将来真正去的数量占用户将来访问兴趣点总量的比例，反映了推荐的全面性。假设目标用户将来共访问了 m 个兴趣点，该兴趣点可能是推荐的兴趣点，也可能不是推荐的兴趣点，则召回率 (recall) = $\frac{k}{m}$ 。实验测试了 $N=5$ 、10 和 20 情况下的

推荐结果的准确率和召回率， N 的默认值设为 5，准确率和召回率是在相同数据集上执行本文推荐算法 100 次得到的所有值的平均值。

6.3 实验结果分析

本文将融合用户相似性、地理位置和信任关系的混合兴趣点推荐方法 (TSG) 与以下 3 种推荐方法进行比较。

1) user-based CF(U)：基于用户相似性的推荐。

2) preference and context embedding(PACE)^[7]：基于用户偏好和背景嵌入的深度神经网络推荐。

3) user social and geographical model (USG)^[8]：基于用户社交关系和地理因素的推荐。

图 5~图 7 显示了在 Foursquare 和 Gowalla 数据集上运行 4 种推荐算法的实验结果。分别考察兴趣点个数 $N=5$ 、10 和 20 时，4 种推荐算法的准确率和召回率。

图 5 显示了在 $N=5$ 、10、20 情况下，4 种推荐算法的准确率和召回率。从图 5 可以看出，在推荐

过程中，考虑多个因素比单因素得到更好的推荐效果，随着 N 值的增大，推荐的准确率会有一定程度的下降。但本文的 TSG 算法在各种情况下都显示出最佳的推荐性能。这也验证了融合信任关系后，避免了恶意推荐，提高了推荐结果的准确性和可靠性。

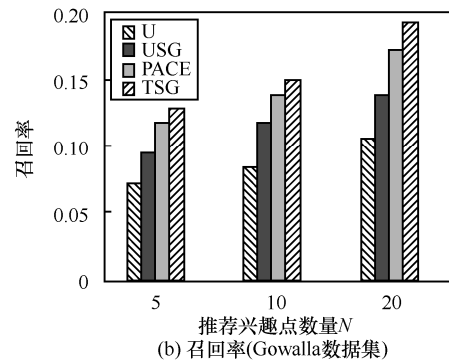
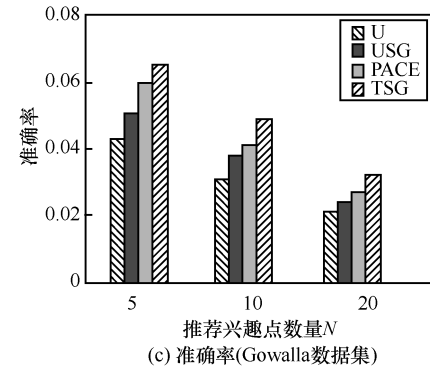
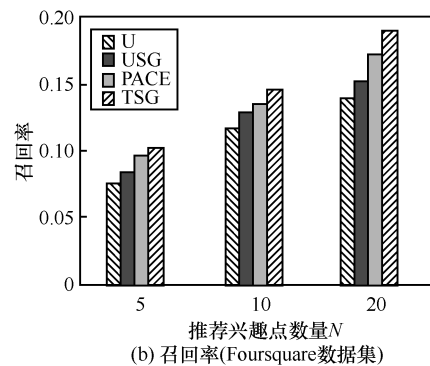
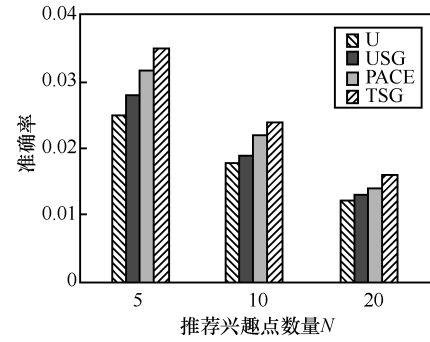
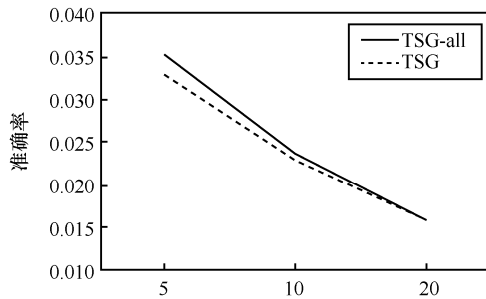
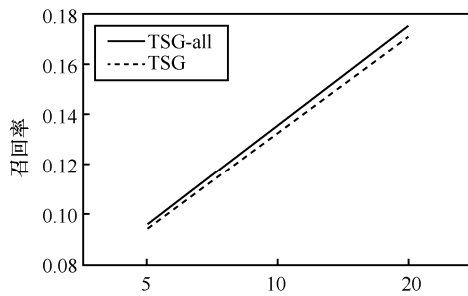


图 5 推荐性能比较

如图 6 所示, 实验对比了混合推荐算法中只考虑信任关系和同时考虑信任与不信任关系情况下的推荐结果质量。TSG 代表在混合推荐算法中只融合信任关系, TSG-all 代表信任和不信任关系同时考虑的推荐算法。



(a) 准确率(Foursquare数据集)



(b) 召回率(Foursquare数据集)

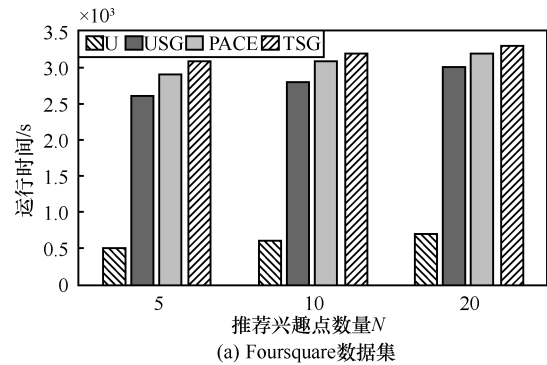
图 6 不信任关系在推荐中的影响

为了说明不信任关系能够抵制恶意推荐, 本文在真实数据集中注入了虚假签到记录。TSG 算法仅考虑信任关系, TSG-all 算法考虑了不信任关系。

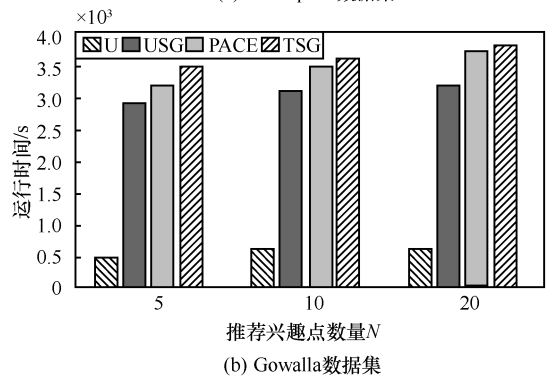
从图 6 可以看出, 融合了不信任关系之后, TSG-all 推荐算法的准确率和召回率都比只考虑信任关系的 TSG 算法有改善, 这是因为虚假签到的用户是目标用户不信任的, 在形成推荐的过程中过滤了虚假用户的推荐信息。这也证明了引入不信任关系对推荐质量的提高是有效的, 能够抑制恶意推荐。通过研究和合理利用不信任关系及其传递可以发现更多用户间潜在的信息, 以此来提升推荐的性能。

如图 7 所示, 实验对比了本文提出的混合推荐算法与单因素的推荐算法的运行时间。从图 7 可以看出, 在 2 种数据集上, 融合了信任关系的混合推荐算法的运行时间要稍高于考虑单因素的传统推荐算法。这是因为混合推荐系统要分别计算基于用户相似性、基于地理位置信息以及基于信任关系和不信任关系的签到概率, 然后进行线性组合获得最

终推荐结果, 时间开销分析详见 5.4 节。但是 3 种推荐系统相对独立, 可以并发运行, 因此时间开销的增加不是非常大。



(a) Foursquare数据集



(b) Gowalla数据集

图 7 运行时间比较

7 结束语

本文通过对基于位置的社交网络用户的信任与不信任关系及其传递对兴趣点推荐的影响研究发现, 将用户间的社交影响细化为信任关系, 可以改善推荐的性能。基于此, 提出了一个融合用户相似性、地理位置和信任关系的兴趣点推荐框架。与基于用户的协同推荐以及基于地理位置的推荐相比, 混合推荐算法的推荐结果更准确, 用户的满意度更好。

POI 推荐可能由于敏感信息(如地理位置等的泄露)面临隐私保护问题。未来将研究 POI 推荐中的隐私保护和推荐准确度之间的平衡问题。

参考文献:

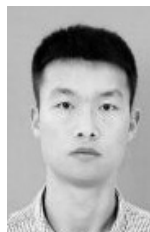
[1] ZHENG Y. Location-based social networks: users[M]. Springer New York, 2011: 243-276.
 [2] 张富国. 基于社交网络的个性化推荐技术[J]. 小型微型计算机系统, 2014, 35(7): 1470-1476.
 ZHANG F G. Survey of online social network based personalized recommendation [J]. Journal of Chinese Computer Systems, 2014, 35(7): 1470-1476.

- [3] BHUIYAN T. A survey on the relationship between trust and interest similarity in online social networks[J]. Journal of Emerging Technologies in Web Intelligence, 2010,2(4):291-299.
- [4] CHEN L J, GAO J. A trust-based recommendation method using network diffusion processes[J]. arXiv:1803.08378. 2018:1-14.
- [5] GUO G. Integrating trust and similarity to ameliorate the data sparsity and cold start for recommender systems[C]//ACM Conference on Recommender Systems. 2013: 451-454.
- [6] GUO G, ZHANG J, THALMANN D. Merging trust in collaborative filtering to alleviate data sparsity and cold start[J]. Knowledge-Based Systems, 2014, 57(2): 57-68.
- [7] YANG C, BAI L, ZHANG C, et al. Bridging collaborative filtering and semi-supervised learning: a neural approach for POI recommendation[C]//ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2017: 1245-1254.
- [8] ZHANG D C, LI M, WANG C D. Point of interest recommendation with social and geographical influence[C]// IEEE International Conference on Big Data. 2017:1070-1075.
- [9] LEE W P, MA C Y. Enhancing collaborative recommendation performance by combining user preference and trust-distrust propagation in social networks[M]. Elsevier Science Publishers BV, 2016.
- [10] YE M, YIN P, LEE W C, et al. Exploiting geographical influence for collaborative point-of-interest recommendation[C]//International ACM Sigir Conference on Research & Development in Information Retrieval. 2011:325-334.
- [11] LIU B, FU Y, YAO Z, et al. Learning geographical preferences for point-of-interest recommendation[C]//ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2013: 1043-1051.
- [12] VICTOR P, VERBIEST N, CORNELIS C, et al. Enhancing the trust-based recommendation process with explicit distrust[J]. ACM Transactions on the Web, 2013, 7(2): 6.
- [13] MA X, LU H, GAN Z. Implicit trust and distrust prediction for recommender systems[J]. 2015: 185-199.
- [14] ZHOU T, LU L, ZHANG Y C. Predicting missing links via local information[J]. European Physical Journal B, 2009, 71(4): 623-630.
- [15] SYMEONIDIS P, TIAKAS E, MANOLOPOULOS Y. Transitive node similarity for link prediction in social networks with positive and negative links[C]//ACM Conference on Recommender Systems. 2010: 183-190.
- [16] TOBLER W R. A computer movie simulating urban growth in the detroit region[J]. Economic Geography, 1970, 46(s1): 234-240.
- [17] MASSA P, AVESANI P. Trust metrics on controversial users[J]. International Journal on Semantic Web & Information Systems, 2009, 3(1): 39-64.
- [18] CARTWRIGHT D, HARARY F. Structural balance: a generalization of heider's theory[J]. Psychological Review, 1956, 63(5): 277.

[作者简介]



朱敬华 (1976-), 女, 博士, 黑龙江大学教授、硕士生导师, 主要研究方向为社会网络推荐、传感器网络、数据挖掘。



明寒 (1991-), 男, 黑龙江大学硕士生, 主要研究方向为基于位置社交网络的个性化推荐。