

融合社交地理信息加权矩阵分解的 兴趣点推荐算法

何颖¹, 王卓然², 周旭³, 刘衍珩^{1,2}

(1. 长春财经学院信息工程学院, 长春 130122; 2. 吉林大学计算机科学与技术学院, 长春 130012; 3. 吉林大学公共计算机教学与研究中心, 长春 130012)

摘要: 针对用户-兴趣点矩阵稀疏以及难于从隐反馈中获取用户对未访问位置的偏好而影响兴趣点推荐准确度的问题, 本文提出了一种融合社交地理位置信息的加权矩阵分解兴趣点推荐算法(SGWMF)。首先, 通过用户之间的相关性对社交信息进行幂律分布建模, 基于用户好友的签到信息计算用户访问位置概率; 其次, 利用地理信息符合幂律分布特点重构用户访问位置偏好矩阵, 缓解矩阵数据稀疏性问题; 再次, 为了增强加权矩阵分解模型的有效性, 通过建模社交信息和地理信息挖掘出用户对未访问位置的偏好, 并以隐反馈项的形式改进加权矩阵分解的目标函数; 最后, 在两个真实数据集上对算法性能进行对比验证, 结果表明本文算法的性能要优于其他兴趣点推荐算法, 推荐结果的准确性有明显提高。

关键词: 计算机应用; 社交地理信息; 加权矩阵分解; 兴趣点推荐

中图分类号: TP391 **文献标志码:** A **文章编号:** 1671-5497(2023)09-2632-08

DOI: 10.13229/j.cnki.jdxbgxb.20211201

Point of interest recommendation algorithm integrating social geographical information based on weighted matrix factorization

HE Ying¹, WANG Zhuo-ran², ZHOU Xu³, LIU Yan-heng^{1,2}

(1. College of Information and Enignerring, Changchun University of Finance and Economics, Changchun 130122, China; 2. College of Computer Science and Technology, Jilin University, Changchun 130012, China; 3. Center for Computer Fundamental Education, Jilin University, Changchun 130012, China)

Abstract: The point-of-interest (POI) recommendation services provided by the location-based social network (LBSN) have become an important means of mining users' preference for POIs. The sparsity of user-POI matrix is the primary problem to be solved, and a large number of unknown values in implicit feedback cannot reflect user preferences. To improve recommendation precision, this paper proposes a point of interest recommendation algorithm integrating social geographical information based on weighted matrix factorization (SGWMF). The social information is modeled through the power-law distribution.

收稿日期: 2021-11-15.

基金项目: 国家自然科学基金项目(61806083); 中央高校基本科研业务费项目(93K172021Z02).

作者简介: 何颖(1982-), 女, 副教授, 博士. 研究方向: 移动网络应用. E-mail: yinghe@ccufe.edu.cn

通信作者: 周旭(1988-), 女, 副教授, 博士. 研究方向: 社交网络分析, 计算智能. E-mail: zhoux16@jlu.edu.cn

The check-in information of the user's friends is converted into the user's visit location preference. Secondly, the power-law distribution of geographical information is used to construct the user's visit location preference matrix to alleviate the data sparsity problem. Thirdly, in order to extend the effectiveness of the model, we improve the objective function by adding implicit feedback term. Finally, the experimental results on two datasets show that it has better performance than other POI recommendation algorithms and can improve the accuracy of recommendation results.

Key words: computer application; social geographical information; weighted matrix factorization; point-of-interest (POI) recommendation

0 引言

近年来,基于位置的社交网络(Location based social networks, LBSN)发展迅速,如 Gowalla、Foursquare、Yelp 和 Brightkite 等。人们可以使用手机、电脑等移动设备以签到的形式与好友分享自己的地理位置、访问兴趣点等。兴趣点推荐旨在分析用户签到信息、挖掘用户兴趣爱好,从而更好地理解用户行为,为用户推荐他们感兴趣的位置。

用户签到的数据会随着时间的增长变得越来越多,用户和兴趣点的数量也越来越大,会出现用户访问地点不均衡的情况。由于用户仅访问少部分地点,导致签到矩阵极度稀疏,给兴趣点推荐的研究带来了极大挑战。目前应用范围最广的推荐算法是协同过滤(Collaborative filtering, CF)算法,其主要分为基于内存的CF算法和基于模型的CF算法^[1]。基于内存的CF又称为基于近邻的CF,它利用用户的签到矩阵计算用户或地点之间的相似度,计算并预测用户对未知地点的评分,为用户推荐前 n 个地点。

矩阵分解模型^[2]是典型的基于模型的CF方法,它通过降维以及迭代的方法补全预测评分矩阵,能在一定程度上缓解数据稀疏问题,目前应用较广泛。大多数研究是通过改进的用户相似度、信任传递等方法研究社交关系,将好友关系应用于兴趣点推荐。张青博等^[3]构建了用户-项目异构网络挖掘用户间的相似性,采用矩阵分解实现目标用户的兴趣点推荐;Ma等^[4]在矩阵分解中引入用户相似度计算,提出了基于个人正则化和基于平均值正则化的两种模型;Xu等^[5]利用直接信任关系和间接信任关系计算用户之间的信任值,将用户之间相似性以正则化项的形式加入到矩阵分解中进行兴趣点推荐。任星怡等^[6]通过对地理

信息建模得到相关分数,并融合概率矩阵分解提出了基于上下文信息的概率矩阵分解推荐算法;李昆仑等^[7]利用改进的皮尔逊相似度算法计算出用户间的信任值并构建信任网络,将社交信息与文本信息融合纳入概率矩阵分解中以提高算法的推荐性能。

用户签到数据中含有小部分可直接反映用户偏好的已知值和大量不能反映用户偏好的未知值(即隐反馈数据)。虽然上述方法采用矩阵分解模型在一定程度上缓解了矩阵稀疏性问题,但忽略了未知值对推荐性能的影响,因为未知值可能是用户不喜欢的兴趣点,也可能是用户不曾知道但喜欢的兴趣点,不能将所有未知值都假设为用户不喜欢。加权矩阵分解(Weighted matrix factorization, WMF)通过赋予未知值统一的较小权重减少未知值对推荐的影响,能有效挖掘用户的隐藏偏好,具有降低时间和空间复杂度的特点^[8]。Lian等^[9]提出了一个基于地理影响的加权矩阵分解模型,在因子分解模型中分别构建用户的活动区域向量和兴趣点影响区域向量扩展用户和兴趣点的潜在向量;Guo等^[10]只考虑了地理影响,对地理信息进行建模,将兴趣点之间的相关性作为正则化项与加权矩阵分解融合,向用户进行位置推荐;夏英等^[11]计算了用户之间的相似性并且利用局部活动区域建立地理相关模型,最后采用矩阵分解获取用户对位置的预测评分以实现推荐;Cheng等^[12]观察到用户签到通常分布在几个中心附近,将其建模为多中心高斯模型,并结合泊松因子模型进行推荐;Liu等^[13]认为用户的地理偏好和兴趣偏好是相互影响的,将地理影响与泊松因子模型结合提出了地理概率因素模型;Liu等^[14]基于用户对近处的兴趣点有相似的偏好以及同一区域有相似用户偏好,利用加权矩阵分解进行位置推荐研究。现有研究仅通过单一地引入地理信

息或者社交关系实现推荐,没有充分地利用网络中包含的上下文信息,在一定程度上制约了推荐结果的准确性。

为了深入挖掘未知值中隐藏的用户偏好,提高推荐结果的准确性,本文提出了融合社交地理信息加权矩阵分解的兴趣点推荐算法(SGWMF)。首先,基于用户喜欢访问邻近兴趣点的特点,对用户访问兴趣点的地理信息进行建模,通过密度聚类算法和幂律分布得到用户签到概率矩阵;其次,基于用户好友影响用户访问偏好这一理念,利用幂律分布从用户好友的历史签到数据中学习用户位置访问偏好矩阵;再次,将社交信息和地理信息进行建模挖掘用户对未访问位置的偏好并以隐反馈项的形式改进加权矩阵分解目标函数;最后,在两个真实数据集 Gowalla 和 Brightkite 上进行实验,结果表明本文算法比其他兴趣点推荐算法具有更优的性能表现。

1 加权矩阵分解

加权矩阵分解模型根据用户对不同兴趣点的访问次数给予不同权重表征用户对兴趣点偏好的置信度,可以减少未知值对推荐结果的影响力。

$$W_{u,l} = \begin{cases} r \cdot r_{u,l} + 1, & r_{u,l} > 0 \\ 1, & r_{u,l} \leq 0 \end{cases} \quad (1)$$

式中: $W_{u,l}$ 为置信权重矩阵; $r \cdot r_{u,l} + 1$ 为一个单递增函数, r 为一个常数; $r_{u,l}$ 为用户 u 访问兴趣点 l 的频率。

加权矩阵分解的原始目标函数 $F(P, Q)$ 为:

$$F(P, Q) = \frac{1}{2} \sum_{u=1}^N \sum_{l=1}^M W_{u,l} (C_{u,l} - P_u Q_l^T)^2 + \frac{\lambda_P}{2} \|P_u\|_F^2 + \frac{\lambda_Q}{2} \|Q_l\|_F^2 \quad (2)$$

式中: C 为用户签到的 0/1 矩阵; $P \in R^{|N| \times K}$ 为用户特征矩阵, $Q \in R^{|M| \times K}$ 为地点特征矩阵,两者为低维矩阵, K 为潜在因子的维数, N 为用户的数量, M 为地点的数量; $\|\cdot\|_F^2$ 为 Frobenius 范数; λ_P, λ_Q 分别为。

2 本文算法

2.1 问题描述

在基于位置的社交网络中, $U = \{u_1, u_2, \dots, u_{|N|}\}$ 为所有用户的集合, $u \in U$; $L = \{l_1, l_2, \dots, l_{|M|}\}$ 为兴趣点(地点)的集合, $l \in L$ 且 l

包含经纬度(lat, lon); R 为用户-兴趣点签到频率矩阵; S 为一个 $N \times N$ 的社交关系矩阵,其中每个元素 $s_{u,u'}$ 表示用户 u 和用户 u' 是否是好友,如果是好友,那么 $s_{u,u'} = 1$, 否则 $s_{u,u'} = 0$; z 为用户好友的签到频率。兴趣点推荐就是给定用户 u , 向其推荐 Top- n 个兴趣点即 $\{l_1, l_2, \dots, l_{top}\}$ 。本文主要参数如表 1 所示。

表 1 参数表

Table 1 Parameters table

参数	定义	参数	定义
U	数据集中所有用户的集合	R	用户-地点访问频率矩阵
u	某个用户 $u \in U$	$r_{u,l}$	用户 u 在地点 l 的访问频率
L	数据集中所有地点的集合	O_u	用户 u 的地点簇类
l	某个地点 $l \in L$	B_u	用户 u 的中心位置集
C	用户签到 0/1 矩阵	b_n	用户 u 在簇 n 下的中心位置
S	用户社交关系矩阵	z	用户好友的签到频率
$s_{u,u'}$	用户 u 与用户 u' 有好友关系		

2.2 地理信息建模

根据地理学第一定律:空间相关性、地物之间的相关性与距离有关,一般来说,距离越近,事物间的相关性越大;距离越远,事物间相异性越大。用户一般都是居住在某一城市,通常会有一个主要的活动范围。用户会在活动范围内去访问地点,通常用户喜欢访问中心位置附近的地方。

为了计算用户的活动中心位置,以往方法通常是在用户的移动轨迹中选取一点,以该点为圆心,以 R_u 为半径画圆。与用户访问的圆形区域中所有位置频率之和最大的圆心位置即为用户中心位置,但是在寻找中心位置过程中会遇到圆形区域中频率之和相似或相等的情况,此时则无法确定中心位置。

为了获得更精准的结果,本文首先通过聚类找到用户的中心位置,然后在用户活动区域内运用幂律分布对新位置进行推荐,这样更加符合用户的生活习惯和实际情况。具体地,通过基于密度聚类的 DBSCAN 算法对每个用户访问的地点进行聚类,得到用户访问地点的簇 $O_u = \{o_{u,1}, o_{u,2}, o_{u,3}, \dots, o_{u,n}\}$, O_u 为用户 u 的地点簇, n 为该用户的地点簇个数; $B_u = \{b_1, b_2, \dots, b_n\}$, B_u 为用户 u 的中心位置集, b_n 为用户 u 在簇 n 下的中

心位置。

$$b_n = \left(\frac{\sum_{l \in o_{u,n}} r_{u,l} \cdot \text{lat}_l}{\sum_{l \in o_{u,n}} r_{u,l}}, \frac{\sum_{l \in o_{u,n}} r_{u,l} \cdot \text{lon}_l}{\sum_{l \in o_{u,n}} r_{u,l}} \right) \quad (3)$$

式中: lat_l 为地点 l 的纬度; lon_l 为地点 l 的经度。

从地理角度分析用户在两个数据集上的签到数据,由于用户连续两次签到的距离近似服从幂律分布。本文采用幂律分布对用户签到概率进行建模,其概率密度函数为:

$$y = \alpha \cdot x^\beta \quad (4)$$

式中: x, y 分别为同一用户访问两点之间的距离和该距离出现的频率; α 和 β 均为参数。

对式(4)进行对数操作,即 $\lg y = \omega_0 + \omega_1 \lg x$ 。设置 $\alpha = 10^{\omega_0}, \beta = \omega_1, y' = \lg y, x' = \lg x$ 。进一步可以得到:

$$y'(x', \omega) = \omega_0 + \omega_1 x' \quad (5)$$

为了避免过拟合,加入了正则化项,损失函数 $E(\omega)$ 为:

$$E(\omega) = \frac{1}{2} \sum_{n=1}^N \{ y'(x'_n, \omega) - t_n \}^2 + \frac{\lambda}{2} \|\omega\|^2 \quad (6)$$

$$\omega_0 \leftarrow \omega_0 - \eta [(\omega_0 + \omega_1 x'_n - t_n) + \lambda \omega_0] \quad (7)$$

$$\omega_1 \leftarrow \omega_1 - \eta [(\omega_0 + \omega_1 x'_n - t_n) \cdot x'_n + \lambda \omega_1] \quad (8)$$

采用梯度下降法最小化误差的平方和,找到最佳参数值 α 和 β 。

确定用户中心位置后,以半径 R_a 构建用户的活动区域,该半径是用户可能访问地点的最远距离,设定 $R_a = 60 \text{ km}$; 然后,在区域内部寻找用户中心位置和用户未访问位置之间的最小距离 d :

$$d = \min_{b_n \in B_n} \text{dis}(l_{\text{new}}, b_n) \quad (9)$$

式中: l_{new} 为活动区域内部用户未访问过的地点,即新位置。

最终使用幂律分布建模每个用户在活动区域内的签到距离 d 与签到概率 G 之间的相关性为:

$$G = \alpha \cdot d^\beta \quad (10)$$

2.3 社交信息建模

在实际生活中,朋友们通常有相似的兴趣爱好,朋友关系是一种双向强连接,用户的朋友会对用户访问的地点产生影响。由于在两个真实数据集上好友的签到频率符合幂律分布,本文根据用户与其亲密朋友之间的社交相关性,利用幂律分布将朋友签到频率转化为用户访问地点的偏好概率:

$$z_{u,l} = \sum_{u' \in U} s_{u,u'} r_{u',l} \quad (11)$$

式中: $r_{u',l}$ 为用户 u' 在地点 l 的访问频率; $z_{u,l}$ 为用户好友的签到频率。

按照 Gowalla 和 Brightkite 数据集中已包含的用户间好友信息,本文通过数据集中用户的好友关系构建用户社交关系矩阵 S 。如果两个用户是好友,则 $s_{u,u'} = 1$ 。

好友的签到频率服从幂律分布,即:

$$Z = (\gamma - 1)(z + 1)^{-\gamma} \quad (12)$$

运用最大似然估计得到:

$$\gamma = 1 + |U| |L| \cdot$$

$$\left[\sum_{u' \in U} \sum_{l' \in L} \ln \left(1 + \sum_{u'' \in U} s_{u',u''} r_{u'',l'} \right) \right]^{-1} \quad (13)$$

式中: $\sum_{u'' \in U} s_{u',u''} r_{u'',l'}$ 为用户 u' 的好友 u'' 在地点 l' 的访问频率。

由此得到累积分布函数 H 为:

$$H = \int_0^z Z(a) da = 1 - (z + 1)^{1-\gamma} \quad (14)$$

该模型反映了好友对用户偏好的影响。

2.4 改进的加权矩阵分解

由于未知值(即用户未访问过的兴趣点)包含两种情况,可能是用户完全不喜欢的兴趣点,也可能是那些符合用户偏好但尚未访问的兴趣点,WMF 通过赋予未知值统一的较小权重减少未知值对推荐结果的影响,但未知值不完全是符合用户偏好的兴趣点。本文基于用户签到位置的地理分布关系和用户之间的社交关系,对未知值分别进行地理和社交信息上的建模。类比于置信权重矩阵 W 定义了未知值的权重矩阵 D ,根据未知值权重矩阵 D 为未知值分配权重,减少未知值对推荐结果的影响,该方法很好地融合各种影响因素,具有一定的扩展性。本文将设计的隐反馈项 $D(\mathbf{0} - PQ^T)$ 添加到传统加权矩阵分解的目标函数式(2)中,提出了改进的加权矩阵分解目标函数。

本文通过对地理信息和社交信息建模挖掘未知值中用户的隐藏偏好,得到未知值权重矩阵,以隐反馈项的形式构建的新目标函数为:

$$F'(P, Q) = \frac{1}{2} \sum_{u=1}^N \sum_{l=1}^M W_{u,l} (C_{u,l} - P_u Q_l^T)^2 + \frac{1}{2} \sum_{u=1}^N \sum_{l=1}^M D_{u,l} (\mathbf{0} - P_u Q_l^T)^2 + \frac{\lambda_p}{2} \|P_u\|_F^2 + \frac{\lambda_q}{2} \|Q_l\|_F^2 \quad (15)$$

$$D_{u,l} = \eta \cdot G + (1 - \eta) \cdot H \quad (16)$$

通过有效地融合社交信息和地理信息,构建新的加权矩阵分解的目标函数 $F'(P, Q)$, 并采用最小二乘法, 最小化目标函数, 迭代计算出 P, Q 。

$$\frac{\partial F}{\partial P_u} = \sum_{l=1}^M W_{u,l} (C_{u,l} - P_u Q_l^T) \cdot (-Q_l) + \sum_{l=1}^M D_{u,l} (0 - P_u Q_l^T) \cdot (-Q_l) + \lambda_p P_u \quad (17)$$

$$\frac{\partial F}{\partial Q_l} = \sum_{u=1}^N W_{u,l} (C_{u,l} - P_u Q_l^T) \cdot (-P_u) + \sum_{u=1}^N D_{u,l} (0 - P_u Q_l^T) \cdot (-P_u) + \lambda_q Q_l \quad (18)$$

令 $\frac{\partial F}{\partial P_u} = 0, \frac{\partial F}{\partial Q_l} = 0$, 即得到 P_u 和 Q_l 分别为:

$$P_u = W_{u,l} C_{u,l} Q_l (Q_l^T W_{u,l} Q_l + Q_l^T D_{u,l} Q_l + \lambda_p I)^{-1} \quad (19)$$

$$Q_l = W_{u,l} C_{u,l} P_u (P_u^T W_{u,l} P_u + P_u^T D_{u,l} P_u + \lambda_q I)^{-1} \quad (20)$$

PQ^T 则为最终用户在兴趣点的签到概率。

3 实验及结果分析

3.1 实验数据集

在两个大规模基于位置的社交网络数据集上对本文算法性能进行验证, 选择的两个数据集分别为 Gowalla 和 Brightkite, 它们包含了用户、地点的经纬度、用户社交关系等属性。每条用户签到记录包括用户、地点 ID、纬度、经度、签到时间、好友关系。数据集信息如表 2 所示。

表 2 数据集信息

Table 2 Statistics of two datasets

参数	数据集	
	Gowalla	Brightkite
签到数量	1 278 274	4 747 281
用户数量	18 737	51 406
兴趣点数量	32 510	772 966
好友关系数量	86 985	428 156
用户-兴趣点的矩阵密度/ 10^{-5}	130	2.7084

Gowalla 数据集包含 1 278 274 条签到记录, Brightkite 数据集包含 4 747 281 条签到记录。为了清除较少发生的异常数据, 减少空间浪费, 提高推荐的效率, 本文对 Gowalla 数据集进行了预处理, 过滤掉少于 10 次签到次数的用户和访问次数少于 10 次的兴趣点。经处理后的数据集有 2150 个用户、6668 个兴趣点、209 195 条签到数据。对于 Brightkite 数据集, 过滤掉签到次数低于 10 次的用户和访问次数少于 15 次的兴趣点, 处理后的

数据集有 2376 个用户、9041 个兴趣点、799 229 条签到数据。最后, 将两个数据集随机选取分成两部分: 80% 作为训练数据, 20% 作为测试数据。

3.2 实验结果的评价指标

评价指标是评价推荐质量好坏的重要指标, 根据得到的用户对兴趣点的偏好评分, 采用两个指标准确率 Precision@n 和召回率 Recall@n 对推荐性能做评估:

$$\text{Precision}@n_0 = \frac{1}{N} \sum_{u=1}^{N_u} \frac{|R_u \cap T_u|}{n} \quad (21)$$

$$\text{Recall}@n_0 = \frac{1}{N} \sum_{u=1}^{N_u} \frac{|R_u \cap T_u|}{|T_u|} \quad (22)$$

式中: $N_{\text{总}}$ 为用户的总数; R_u 为用户推荐 Top- n 个兴趣点; T_u 为用户 u 在测试集中真实访问的位置; n_0 为推荐列表的长度。

准确率表示命中兴趣点数量与为用户推荐兴趣点数量的比率; 召回率表示命中兴趣点数量与被预测用户实际访问过兴趣点的比率。

3.3 对比算法

将本文融合社交地理信息加权矩阵分解的兴趣点推荐算法与其他算法进行比较。此外, 为了探讨不同因素对推荐结果的影响, 将 SGWMF 拆分成融合社会信息的加权矩阵分解推荐算法 S-WMF 和融合地理位置信息的加权矩阵分解推荐算法 G-WMF。

UCF (CF): 考虑用户之间的关联, 通过余弦相似度计算用户之间的相似性, 为用户进行兴趣点推荐^[1]。

WMF: 该方法基于矩阵分解模型, 对未知值分配较小的统一权重, 采用加权矩阵分解技术对用户-兴趣点矩阵进行矩阵分解, 根据预测值得到推测结果^[8]。

GeoMF: 采用二维核密度估计捕捉地点空间聚集现象, 构建用户活动区域和地点影响向量, 并结合矩阵分解算法实现兴趣点推荐^[9]。

L-WMF: 该方法从位置角度考虑地理因素对用户签到行为的影响, 结合加权矩阵分解实现兴趣点推荐^[10]。

SLGMF: 该方法对用户相似性进行建模并且利用局部活动区域建立地理相关模型, 利用矩阵分解获得预测评分进行兴趣点推荐^[11]。

SGWMF: 本文提出的融合社交地理信息加权矩阵分解的兴趣点推荐算法, 通过增加隐反馈项对

目标函数进行改进,向用户推荐符合偏好的位置。

S-WMF:本文方法只采用所提出的社交建模方法构建隐反馈项,利用改进加权矩阵分解实现推荐。

G-WMF:本文方法只采用所提出的地理信息建模,并以隐反馈项的形式改进加权矩阵分解目标函数进行兴趣点推荐。

3.4 结果分析

3.4.1 参数设置

加权矩阵分解中的参数 K (即潜在因子的维数)会影响推荐的结果, K 值不同,准确率也不同。本文以 Top-5 为基础,确定 K 的最佳值。

如图 1 所示,对于 Gowalla 数据集,当 $K < 100$ 时,准确率大幅度上升,当 $K \approx 100$ 时,准确率开始趋于平稳,当 $K = 200$ 时达到峰值;当 $K > 200$ 时,虽然图像中显示下降,但其实波动很小。由于 $K > 100$ 时,基本趋于平稳, K 值越大运算时间越长,在 Gowalla 数据集下,本文设置 K 值为 150,保证模型的性能并降低运算时间。

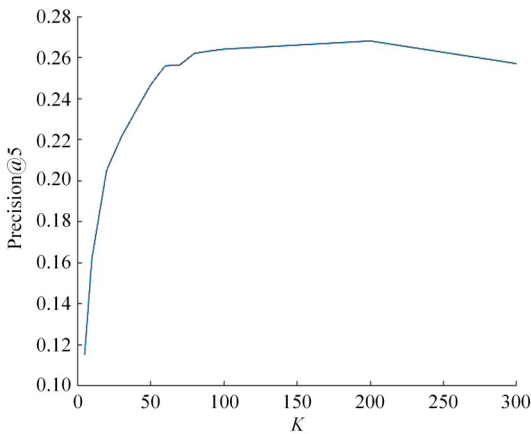


图 1 参数 K 在 Gowalla 数据集上对准确率的影响

Fig. 1 Influence of K for accuracy on Gowalla dataset

如图 2 所示,对于 Brightkite 数据集, K 值越大,准确率越高,当 $K < 90$ 时,准确率提升幅度较大,当 $K = 150$ 时准确率开始趋于平稳,此后精度介于 0.56~0.57 之间。在 Brightkite 数据集下,本文设置 $K = 150$ 。

3.4.2 性能比较

图 3 和图 4 分别为本文算法与其他算法(UCF、WMF、GeoMF、L-WMF、SLGMF、S-WMF 和 G-WMF)在 Gowalla 数据集上向用户推荐的 Top- n 个兴趣点的准确率和召回率对比结果,推荐列表长度 n_0 分别取 5、10 和 15。

针对于 Gowalla 数据集,以 $n = 10$ 为例,UCF

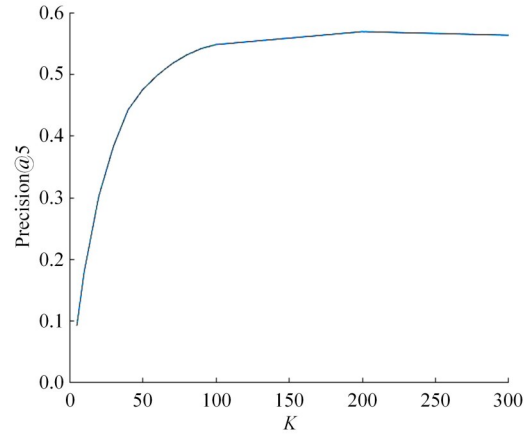


图 2 参数 K 在 Brightkite 数据集上对准确率的影响

Fig. 2 Influence of K for accuracy on Brightkite dataset

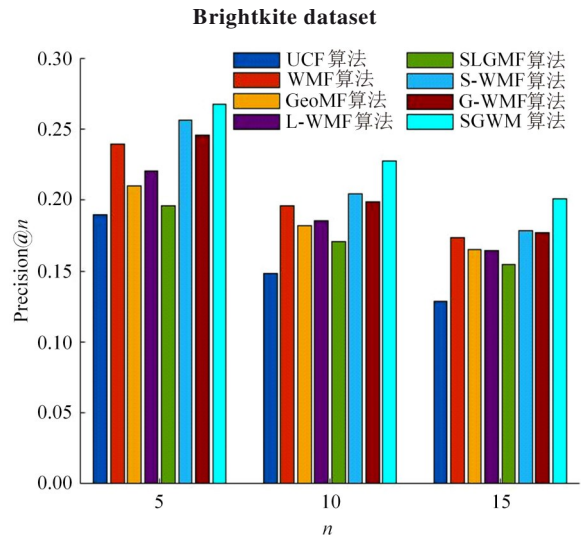


图 3 各算法在 Gowalla 数据集上的准确率对比结果

Fig. 3 Precision of different algorithms on Gowalla dataset

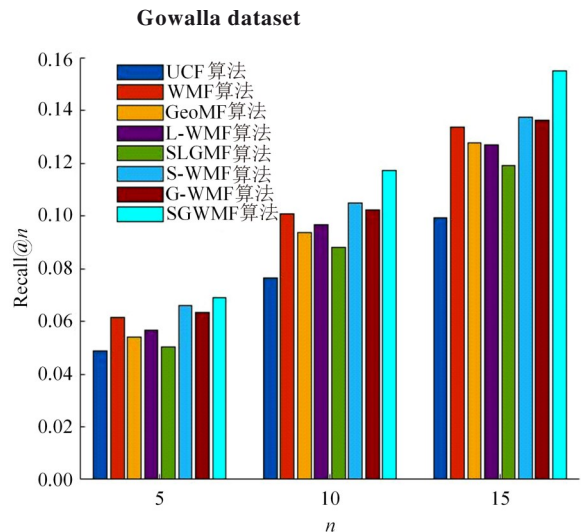


图 4 各算法在 Gowalla 数据集上的召回率对比结果

Fig. 4 Recall of different algorithms on Gowalla dataset

Gowalla dataset

算法考虑了用户间的社交关系,采用基于余弦相似度计算用户之间的相似性,对目标用户进行推荐。UCF 算法的准确率为 0.15,召回率为 0.08;而 WMF 算法的准确率为 0.20,召回率为 0.10,WMF 算法与 UCF 相比准确率提升了 33%。通过数值可以看出,加权矩阵分解的性能表现优于协同过滤算法。原始 WMF 的目标函数对未知值分配的权重相等且都为 1,但未知值不完全是符合用户偏好的地点。为了对未知值分配不同的权重,本文算法 SGWMF 设计了隐反馈项,它可以通过社交地理信息挖掘用户访问位置偏好,能很好地区分未知值中哪些是符合用户偏好的地点,哪些是用户不喜欢的地点。由实验结果可见,SGWMF 的准确率为 0.23,相比于 WMF 提高了 15%。这说明构建隐反馈项对于推荐准确率的提升产生了积极影响。GeoMF 算法的准确率和召回率分别为 0.18 和 0.09,相比于 UCF 算法准确率提升了 20%;L-WMF 算法的准确率和召回率分别为 0.19 和 0.10,相比于 UCF 算法分别提升了 27% 和 25%;以 $n=10$ 为例,SLGMF 算法的准确率为 0.17,而本文算法准确率为 0.23,与 UCF、WMF、GeoMF、L-WMF 和 SLGMF 算法相比分别提升了 53%、15%、28%、21% 和 35%。此外,通过对比 S-WMF 和 G-WMF 算法可以发现,仅考虑社会信息的 S-WMF 算法的推荐结果优于仅考虑地理信息的 G-WMF 算法,这表明社会信息结果的影响要大于地理信息。由实验结果可见,本文算法的性能要明显优于其他对比算法。

图 5 和图 6 分别为 7 种算法在 Brightkite 数据集上的准确率和召回率比较结果。由图 5 和图 6 可见:随着推荐长度 n 由 5 增加到 15,所有算法的推荐准确率值都下降,推荐召回率值增大,整体上本文算法性能要优于其他算法。

以 $n=5$ 为例,UCF 算法的准确率为 0.35,UCF 算法相对于其他算法,其推荐效果并不理想。本文 SGWMF 算法在 Brightkite 数据集上准确率和召回率分别为 0.56 和 0.04,而 WMF、GeoMF 和 L-WMF 算法的准确率分别为 0.49、0.42 和 0.50,本文算法 SGWMF 的准确率比 WMF、GeoMF 和 L-WMF 算法分别提升了 14%、33% 和 12%;召回率比 WMF 算法的 0.037 和 GeoMF 算法的 0.031 分别提升了 8% 和 29%,比 L-WMF 算法的 0.037 提升了 8%。GeoMF 算

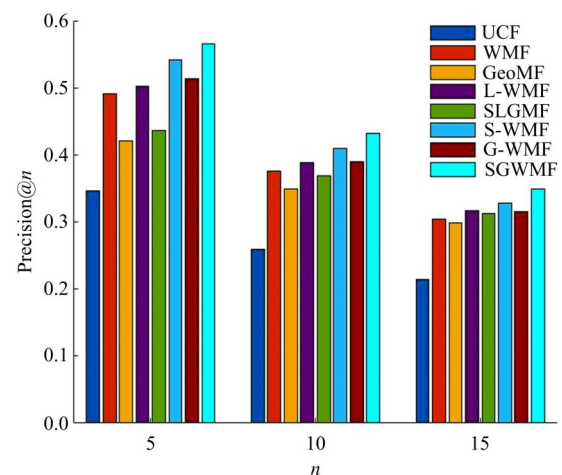


图 5 各算法在 Brightkite 数据集上准确率对比结果

Fig. 5 Precision of different algorithms on Brightkite dataset

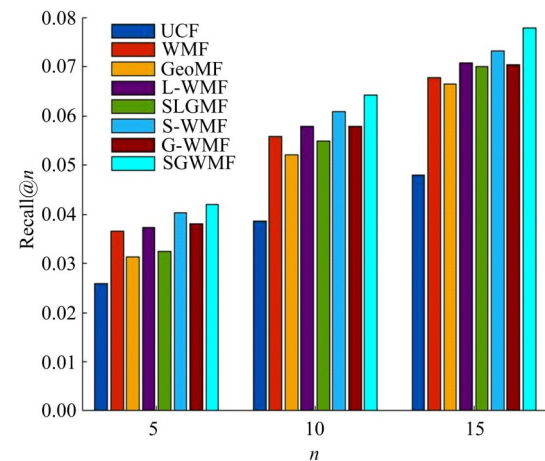


图 6 各算法在 Brightkite 数据集上的召回率对比结果

Fig. 6 Recall of different algorithms on Brightkite dataset

法和 L-WMF 算法的表现性能低于 SGWMF 算法主要是由于它们仅考虑了地理信息的影响,而忽略了社交信息的影响力。

从图 5 可以看出,以 $n=5$ 为例,SLGMF 算法的准确率为 0.44,SGWMF 算法相比 SLGMF 算法在准确率上提高了 27%,由实验结果可知,SGWMF 算法的性能表现更优,具有更好的推荐效果。这是由于本文同时考虑了社交信息和地理信息两方面影响,并引入隐反馈项更好地解决用户签到矩阵中未知值对推荐结果的影响,对未知值分配不同的权重,提高推荐结果的准确率。将 WMF 与 S-WMF、G-WMF 算法比较时,发现只考虑社会关系的 S-WMF 算法准确率为 0.54,只考虑地理关系的 G-WMF 算法准确率为 0.51,其结果都要高于 WMF 算法,说明考虑社交信息和地理影响均会提高推荐结果的准确性。

4 结束语

为解决矩阵的稀疏性问题以及更好地挖掘用户对未访问位置的隐藏偏好,提高推荐算法的性能,本文一方面从用户之间的社交信息考虑朋友对用户签到产生的影响,利用幂律分布建模用户访问偏好;另一方面,从地理角度考虑用户更倾向于访问邻近的兴趣点,利用聚类方法先找到用户的中心位置,再在用户活动区域内通过幂律分布建模用户对位置的访问情况;最后将地理信息和社交信息进行综合分析挖掘用户隐藏的位置偏好,设计新的加权矩阵分解目标函数以获得用户对地点的偏好矩阵。在两个真实数据集 Gowalla 和 Brightkite 上的实验结果表明,相比于其他算法,本文算法具有更高的准确率和召回率。

参考文献:

- [1] Zhao S, King I, Lyu M R. A Survey of Point-of-interest Recommendation in Location-based social Networks[J/OL]. [2021-11-05]. <https://doi.org/10.48550/arXiv.1607.00647>
- [2] 程龙, 李涵. 基于矩阵分解的推荐算法研究综述[J]. 北京信息科技大学学报: 自然科学版, 2021, 36(2): 38-45, 51.
Cheng Long, Li Han. A review of recommendation algorithms based on matrix factorization[J]. Journal of Beijing Information Science & Technology University, 2021, 36(2): 38-45, 51.
- [3] 张青博, 王斌, 崔宁宁, 等. 基于注意力机制的规范化矩阵分解推荐算法[J]. 软件学报, 2020, 31(3): 778-793.
Zhang Qing-bo, Wang Bin, Cui Ning-ning, et al. Attention-based regularized matrix factorization for recommendation[J]. Journal of Software, 2020, 31(3): 778-793.
- [4] Ma H, Zhou D, Liu C, et al. Recommender systems with social regularization[C] //Proceedings of the Forth International Conference on Web Search and Web Data Mining, WSDM, Hong Kong, China, 2011: 287-296.
- [5] Xu C, Ding AS, Zhao K. A novel POI recommendation method based on trust relationship and spatial-temporal factors[J]. Electronic Commerce Research & Applications, 2021, 48: 101060.
- [6] 任星怡, 宋美娜, 宋俊德. 基于位置社交网络的上下文感知的兴趣点推荐[J]. 计算机学报, 2017, 40(4): 824-841.
Ren Xing-yi, Song Mei-na, Song Jun-de. Context-aware point-of-interest recommendation in location-based social networks[J]. Chinese Journal of Computers, 2017, 40(4): 824-841.
- [7] 李昆仑, 翟利娜, 赵佳耀, 等. 融合信任关系与评论文本的矩阵分解推荐算法[J]. 小型微型计算机系统, 2021, 42(2): 285-290.
Li Kun-lun, Zhai Li-na, Zhao Jia-yao, et al. Matrix factorization recommendation algorithms by exploiting trust relationship and review text[J]. Journal of Chinese Computer Systems, 2021, 42(2): 285-290.
- [8] Hu Y, Volinsky K Y. Collaborative filtering for implicit feedback datasets[C]//2008 8th IEEE International Conference on Data Mining, Pisa, Italy, 2008: 263-272.
- [9] Lian D, Zhao C, Xie X, et al. GeoMF: joint geographical modeling and matrix factorization for point-of-interest recommendation[C]//ACM Sigkdd International Conference on Knowledge Discovery and Data Mining, New York, USA, 2014: 831-840.
- [10] Guo L, Wen Y, Liu F. Location perspective-based neighborhood-aware POI recommendation in location-based social networks[J]. Soft Computing, 2019, 23(22): 11935-11945.
- [11] 夏英, 张金凤. 融合社交关系和局部地理因素的兴趣点推荐[J]. 计算机工程与应用, 2021, 57(15): 133-139.
Xia Ying, Zhang Jin-feng. POI Recommendation fusing social relations and local geographic factors[J]. Computer Engineering and Applications, 2021, 57(15): 133-139.
- [12] Cheng C, Yang H, King I, et al. Fused matrix factorization with geographical and social influence in location-based social networks[C]//AAAI Conference on Artificial Intelligence, Ontario, Canada, 2012: 17-23.
- [13] Liu B, Xiong H, Papadimitriou S, et al. A general geographical probabilistic factor model for point of interest recommendation[J]. IEEE Transactions on Knowledge & Data Engineering, 2015, 27(5): 1167-1179.
- [14] Liu Y, Wei W, Sun A, et al. Exploiting geographical neighborhood characteristics for location recommendation [C]//CIKM'14: Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management, Shanghai, China, 2014: 739-748.