



吉林大学学报(工学版)

Journal of Jilin University(Engineering and Technology Edition)

ISSN 1671-5497,CN 22-1341/T

《吉林大学学报(工学版)》网络首发论文

题目: 融合社交地理信息加权矩阵分解的兴趣点推荐算法
作者: 何颖, 王卓然, 周旭, 刘衍珩
DOI: 10.13229/j.cnki.jdxbgxb20211201
收稿日期: 2021-11-15
网络首发日期: 2022-03-19
引用格式: 何颖, 王卓然, 周旭, 刘衍珩. 融合社交地理信息加权矩阵分解的兴趣点推荐算法[J/OL]. 吉林大学学报(工学版).
<https://doi.org/10.13229/j.cnki.jdxbgxb20211201>



网络首发: 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

出版确认: 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约, 在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

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：基于位置的社交网络提供的兴趣点推荐服务成为挖掘用户兴趣偏好的重要手段。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；然后，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的实验结果表明该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

DOI：10.13229/j.cnki.jdxbgxb20211201

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

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

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

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资助项目(93K172021Z02)。

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

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

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、访问兴趣点的经历。兴趣点推荐旨在分析用户签到的信息、挖掘用户的兴趣爱好, 从而更好地理解用户的行为, 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巨大的挑战。目前应用范围最广的推荐算法是协同过滤算法 (Collaborative Filtering, CF), 主要分为基于内存的 CF 和基于模型的 CF^[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者地点之间的相似度, 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, 目前应用很广泛。Zhang 等人^[3]构建用户-项目异构网络挖掘用户间的相似性, 采用矩阵分解对目标用户实现推荐; Ma 等人^[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; Li 等人^[5]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到概率矩阵分解中以提高推荐性能。

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的对推荐的影响, 能有效地挖掘用户的隐藏偏好, 具有降低时间和空间的复杂度的特点^[6]。现有研究仅通过单一的引入地理信息或者社交关系来实现推荐, 而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。Lian 等人^[7]提出一个

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的影响区域向量来扩展用户和兴趣点的潜在向量; Guo 等人^[8]只考虑了地理影响, 对地理信息建模, 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。夏英等人^[9]计算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实现推荐。

为了深入挖掘未知值中隐藏的用户偏好, 提高推荐结果准确性, 本文研究融合社交地理信息的加权矩阵分解算法实现兴趣点推荐。主要贡献包括:

(1) 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; (2) 基于用户的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对位置访问偏好矩阵; (3) 将社交信息和地理信息进行建模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目标函数; (4) 在两个真实数据集 Gowalla 和 Brightkite 上进行实验, 结果表明该算法比其他兴趣点推荐算法具有更优的性能表现。

1 相关工作

1.1 基于社交信息的推荐

社交信息在兴趣点推荐中起到了重要的作用, 用户与自己的朋友通常具有相似的兴趣爱好, 人们在购买或者访问地点做决策时通常会询问朋友的想法^[10]。多数研究将好友关系应用于兴趣点推荐中, 主要是通过改进的用户相似度、信任传递等方法研究社交关系。Deng 等人^[11]对未知值设计非均匀加权法, 考虑社交信息, 提出融合社交信息的全加权矩阵分解模型, 提高算法性能; 王凯慈^[12]将社交信息通过幂律分布建模应用到重力模型中并与加权马尔科夫链结合进行推荐; 张志军^[13]提出基于概率矩阵分解模型和直接信任、间接信任和信任传播的方法结合的位置推荐算法, 改善推荐算法的性能; Qian 等人^[14]基于概率矩阵分解模型, 将个人兴趣、人际兴趣相似性和人际影响力三个社交因素有效融合构建推荐模型; Li 等人^[15]对皮尔逊相似度计算方法进行改进, 计算用户的相似度构建相似性网络, 提出基于用户相似性传递的 CF 算法; Pan 等人^[16]采用

随机游走法确定用户之间的相似性，并采用矩阵分解算法进行用户兴趣点推荐。Xu 等人^[17]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，将用户之间的相似性以正则化项的形式加入到矩阵分解中进行推荐。

1.2 基于地理信息的推荐

地理信息影响用户访问兴趣点的概率，利用地理信息可以提高基于位置社交网络中兴趣点推荐的质量。Ren 等人^[18]对地理信息建模得到相关分数，并融合概率矩阵分解提出基于上下文信息的概率矩阵分解推荐算法；Liu 等人^[19]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；Cheng 等人^[20]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 等人^[21]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 等人^[22]提出一个基于贝叶斯非负矩阵分解的地理概率因素框架，可以捕获用户签到行为的地理影响。Du 等人^[23]利用地点经纬度以及流行度对地理信息建模以预测用户对地点的偏好，再与矩阵分解得到的预测结果进行相乘，为用户提供推荐列表。

1.3 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的权重来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。

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

$\alpha \cdot r_{u,l} + 1$ 为是一个单调递增的函数， α 是一个

常数。 $r_{u,l}$ 表示用户 u 访问兴趣点 l 频率。

加权矩阵分解的原始目标函数为：

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

其中， \mathbf{C} 是用户签到的 0/1 矩阵，用户特征矩阵

$\mathbf{P} \in R^{|N| \times K}$ 和地点特征矩阵 $\mathbf{Q} \in R^{|M| \times K}$ 是低维矩阵，

K 是潜在因子的维数； N 是用户的数量， M 是地

点的数量。 $\|\cdot\|_F^2$ 是 Frobenius 范数。

2 融合社交地理信息加权矩阵分解的兴趣点推荐

2.1 问题描述

在基于位置的社交网络中， $U = \{u_1, u_2, \dots, u_{|N|}\}$ 表示所有用户的集合， $u \in U$ ， N 代表用户的数量； $L = \{l_1, l_2, \dots, l_{|M|}\}$ 表示兴趣点（地点）的集合， M 代表地点的数量； $l \in L$ 且 l 包含经纬度 (lat, lon)； \mathbf{R} 代表用户-兴趣点签到频率矩阵； \mathbf{S} 是一个 $N \times N$ 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 表示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 $\{l_1, l_2, \dots, l_{top}\}$ 。本文主要符号如表 1 所示。

表 1 符号表

Table 1 Notation

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

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，事物间向异性越大。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，以往的方法是通过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。用户访问的圆形区域中所有位置的频率之和最大的圆心位置即为用户的中心位置，

但是在寻找中心位置的过程中会遇到圆形区域中的频率之和相似或相等的情况，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，在用户的活动区域内运用幂律分布对新位置进行推荐，则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 表

示用户 u 的地点簇， n 为该用户的地点簇的个数； $B_u = \{b_1, b_2, \dots, b_n\}$ ， B_u 表示用户 u 的中心位置集， b_n 表示用户 u 在簇 n 下的中心位置。

$$b_n = \left(\frac{\sum_{l \in O_{u,n}} r_{u,l} \cdot lat_l}{\sum_{l \in O_{u,n}} r_{u,l}}, \frac{\sum_{l \in O_{u,n}} r_{u,l} \cdot lon_l}{\sum_{l \in O_{u,n}} r_{u,l}} \right) \quad (3)$$

其中， lat_l 表示地点 l 的纬度， lon_l 表示地点 l 的经度。

本文将用户在两个数据集上的签到数据从地理角度进行分析，从图 1 和图 2 可以看出用户连续两次签到的距离近似的服从幂律分布。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：

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

其中， x, y 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 为参数。

对公式(4)进行对数操作，即 $\lg y = w_0 + w_1 \lg x$ 。设置 $\alpha = 10^{w_0}$ ， $\beta = w_1$ ， $y' = \lg y$ ， $x' = \lg x$ 。进一步可以得到：

$$y'(x', w) = w_0 + w_1 x' \quad (5)$$

为了避免过拟合加入了正则化项，损失函数 $E(w)$ 如公式(6)所示，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 。

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

$$w_0 \leftarrow w_0 - \eta[(w_0 + w_1 x'_n - t_n) + \lambda w_0] \quad (7)$$

$$w_1 \leftarrow w_1 - \eta[(w_0 + w_1 x'_n - t_n) \cdot x' + \lambda w_1] \quad (8)$$

在确定了用户的中心位置后，我们以半径 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的位置之间的最小距

离 d ：

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

l_{new} 表示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：

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

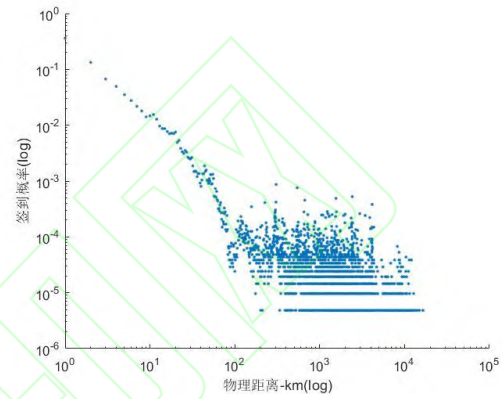


图 1 Gowalla 数据集地理签到距离分布

Fig.1 Geographical check-in distance distribution in Gowalla dataset

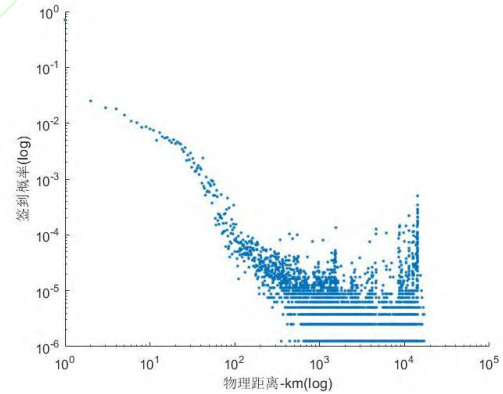


图 2 Brightkite 数据集地理签到距离分布

Fig.2 Geographical check-in distance distribution in Brightkite dataset

2.3 社交信息建模

在实际生活中，朋友们通常有相似的爱好，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。如图 3 和图 4 所示，可以看出在两个真实数据集上好友的签到频率是符合幂律分布。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

的偏好概率为:

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

按照 Gowalla 和 Brightkite 数据集中已包含的用户间好友信息, 我们通过数据集中用户的好友关系构建用户社交关系矩阵 \mathbf{S} 。如果是两个用户是好友, 则 $s_{u,u'} = 1$, $r_{u',l}$ 表示用户 u' 在地点 l 的访问频率, $z_{u,l}$ 表示用户好友的签到频率。

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:

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

其中, $\sum_{u'' \in U} s_{u',u''} r_{u'',l'}$ 表示用户 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:

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

该模型反映了好友对用户偏好的影响。

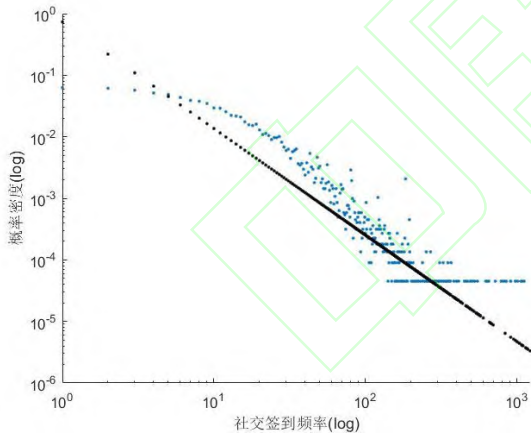


图 3 Gowalla 数据集社交签到频率分布

Fig.3 Social check-in frequency distribution in Gowalla dataset

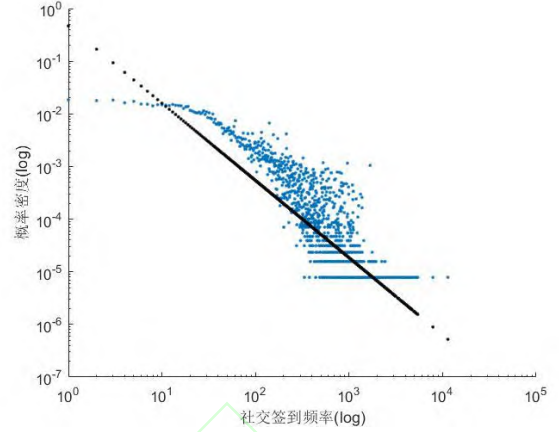


图 4 Brightkite 数据集社交签到频率分布

Fig.4 Social check-in frequency distribution in Brightkite dataset

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建模, 类比于置信权重矩阵 \mathbf{W} 定义了未知值的权重矩阵 \mathbf{D} , 根据未知值权重矩阵 \mathbf{D} 为未知值分配权重, 减少未知值对推荐结果的影响, 能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 $D(0 - PQ^T)$ 添加到传统加权矩阵分解的目标函数公式(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:

$$F(\mathbf{P}, \mathbf{Q}) = \frac{1}{2} \sum_{u=1}^N \sum_{l=1}^M W_{u,l} (C_{u,l} - P_u Q_l^T)^2 + \frac{1}{2} \sum_{u=1}^N \sum_{l=1}^M D_{u,l} (0 - P_u Q_l^T)^2 + \frac{\lambda_P}{2} \|\mathbf{P}_u\|_F^2 + \frac{\lambda_Q}{2} \|\mathbf{Q}_l\|_F^2 \quad (15)$$

其中, $W_{u,l}$ 的计算如公式(1)所示, $D_{u,l}$ 的计算方法如公式(16)所示。

$$D = \eta \cdot \mathbf{G} + (1 - \eta) \cdot \mathbf{H} \quad (16)$$

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 $F(\mathbf{P}, \mathbf{Q})$, 并采用最小二

乘法，最小化目标函数，迭代计算出 \mathbf{P}, \mathbf{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 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 则表示最终用户在兴趣点的签到概率。

3 实验结果分析

我们在两个大规模的基于位置的社交网络数据集上进行实验，来验证所提出方法的性能。

3.1 实验数据集

本文选择两个 LBSN 数据集为 Gowalla^[8]和 Brightkite^[24]，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

数据集	Gowalla	Brightkite
签到数量	1278274	4747281
用户数量	18737	51406
兴趣点数量	32510	772966
好友关系数量	86985	428156
用户-兴趣点的矩阵密度	1.3×10^{-3}	2.7084×10^{-5}

Gowalla 数据集包含 1278274 条签到记录，Brightkite 数据集包含 4747281 条签到记录。为了清除较少发生的异常数据，减少空间的浪费，提高推荐的效率。本文对于 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 个用户和 6668 个兴趣点，共 209195 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、799229 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：

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

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

其中， N 是用户的总数， R_u 表示为用户推荐 Top- n 个兴趣点， T_u 表示用户 u 在测试集中真实访问的位置， n 表示推荐列表的长度。准确率表示命中的兴趣点数量与为用户所推荐的兴趣点数量的比率，召回率表示命中的兴趣点数量与被预测用户实际访问过的兴趣点的比率。

3.3 对比算法

本文将所提方法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，我们将 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推荐^[1]；

WMF：该方法基于矩阵分解模型，通过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，根据预测值得到预测结果^[6]；

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^[7]；

L-WMF：该方法从位置角度考虑地理因素对用户签到影响，结合加权矩阵分解实现兴趣点推荐^[8]；

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^[9]。

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。

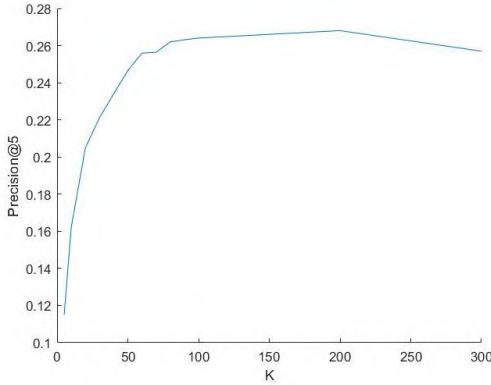


图 5 参数 K 在 Gowalla 数据集上对准确率的影响

Fig.5 The influence of K for accuracy on Gowalla dataset

如图 5 所示，对于 Gowalla 数据集，当 $K < 100$ 时，准确率大幅上升，约 100 维的时候，准确率开始趋于平稳，在 200 维时候到达峰值；超过 200 维虽然图像中显示下降，但其实波动很小。由于 K 值大于 100 维时，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。

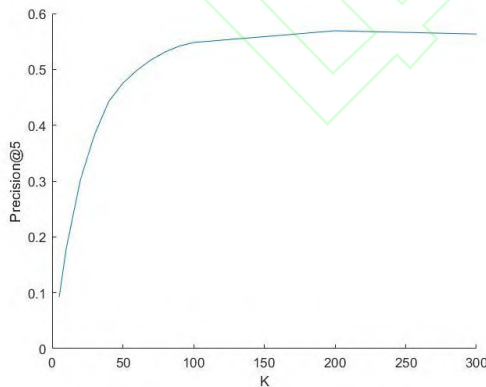


图 6 参数 K 在 Brightkite 数据集上对准确率的影响

Fig.6 The influence of K for accuracy on Brightkite dataset

3.4.2 性能比较

如图 6 所示，对于 Brightkite 数据集， K 值越大，准确率越高，当 K 值小于 90 时，准确率提升

幅度较大，当 $K=150$ 时准确率开始趋于平稳，此后精度位于 0.56~0.57 之间。在 Brightkite 数据集下，本文设置 $K=150$ 。

图 7、图 8 显示本文提出的推荐算法与其他算法 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 分别取 5、10 和 15。

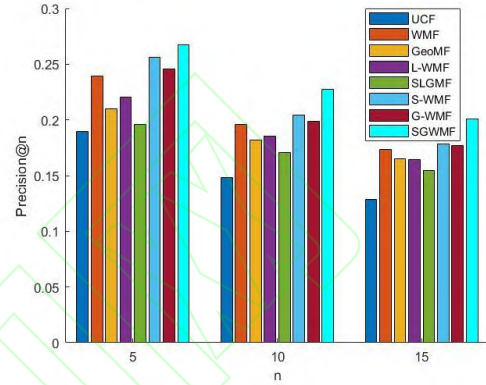


图 7 各算法在 Gowalla 数据集上的准确率对比结果

Fig.7 The precision of different algorithms on Gowalla dataset

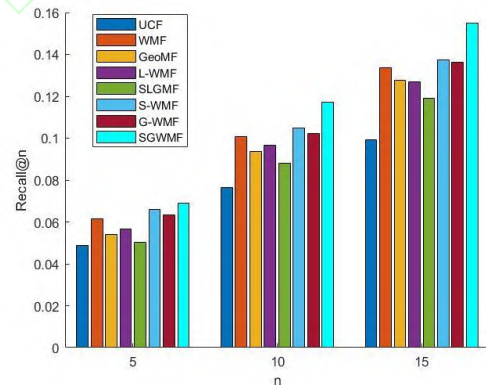


图 8 各算法在 Gowalla 数据集上的召回率对比结果

Fig.8 The recall of different algorithms on Gowalla dataset

针对于 Gowalla 数据集，以 $n=10$ 为例，UCF 考虑用户间的社交关系，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 算法的准确率是 0.15，召回率是 0.08；而 WMF 算法的准确率是 0.20，召回率是 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，我们算法 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。从实验结果上可见，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 的准确率和召回率分别为 0.18 和 0.09，相比于 UCF 准确率提升了 20%；L-WMF 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，而本文提出的算法准确率为 0.23，与 UCF、WMF、GeoMF、L-WMF 和 SLGMF 相比分别提升 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 的推荐结果优于仅考虑地理信息的 G-WMF，这表明社会信息结果的影响要大于地理信息。从实验结果上可见，本文所提算法的性能要明显优于其他对比算法。

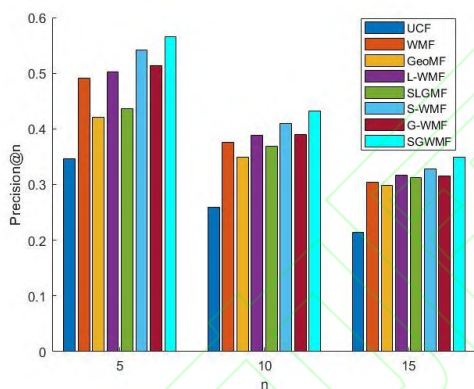


图 9 各算法在 Brightkite 数据集上准确率对比结果

Fig.9 The precision of different algorithms on Brightkite dataset

图 9 和图 10 分别显示了七种算法在 Brightkite 数据集上的准确率和召回率比较结果。从图中可见，随着推荐长度 n 由 5 到 15 的增加，所有算法的推荐准确率值在下降，推荐的召回率值在增加，整体上本文提出的算法要优于其他算法。

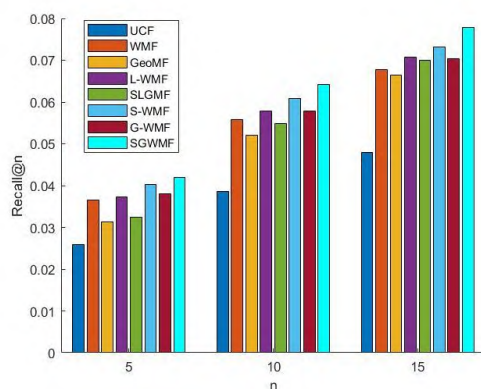


图 10 各算法在 Brightkite 数据集上的召回率对比结果

Fig.10 The recall of different algorithms on Brightkite dataset

以 $n=5$ 为例，UCF 的准确率是 0.35，UCF 算法相对于其他算法，其推荐效果并不理想。本文提出的算法在 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 的准确率分别为 0.49、0.42 和 0.50，我们的算法 SGWMF 的准确率比 WMF、GeoMF 和 L-WMF 分别提升了 14%、33% 和 12%；召回率比 WMF 的 0.037 和 GeoMF 的 0.031 分别提升了 8% 和 29%，比 L-WMF 的 0.037 提升了 8%。GeoMF 算法和 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 9 可以看出，以 $n=5$ 为例，SLGMF 算法的准确率为 0.44，所提算法 SGWMF 比 SLGMF 提高了 27%，从实验结果上可知所提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，并通过引入隐反馈项后能更好地解决用户签到矩阵中未知值对推荐结果影响问题，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 和 G-WMF 算法相比较时，从实验结果上发现只考虑了社会关系的 S-WMF 的准确率为 0.54，只考虑了地理关系的 G-WMF 的准确率为 0.51，其结果都要高于 WMF，这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的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,本文的算法与其他算法相比具有更高的准确率和召回率。在以后的工作中,我们将进一步研究社交网络中影响用户间关系的其他因素,并结合文本信息,图像信息等优化位置推荐算法,进一步提高推荐性能。

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], arXiv:1607.00647, 2016:1-30
- [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 Qian-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] 李昆仑,翟利娜,赵佳耀,王萌萌.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, Wang Meng-meng. Matrix Factorization Recommendation Algorithms by Exploiting Trust Relationship and Review Text[J]. Journal of Chinese Computer Systems, 2021, 42(2): 285-290.
- [6] Hu Y, Koren Y, Volinsky. Collaborative filtering for implicit feedback datasets[C]. 2008 eighth IEEE international conference on data mining. IEEE, 2008: 263-272
- [7] 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: ACM, 2014: 831-840.
- [8] 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.
- [9] 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.
- [10] Cho E, Myers S A, Leskovec J. Friendship and mobility: user movement in location-based social networks[C]// Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2011: 1082-1090.
- [11] 邓秀勤,刘太亨,刘富春,等.基于全加权矩阵分解的用户协同过滤推荐算法[J].计算机科学,2019,46(A2):199-203+219.
- Deng Xiu-qin, Liu Tai-heng, Liu Fu-chun, et al. User Collaborative Filtering Recommendation Algorithm Based on All Weighted Matrix Factorization[J]. Computer Science, 2019, 46(A2): 199-203+219.
- [12] 王凯慈.基于LBSN签到数据的用户行为预测与兴趣点推荐[D].南京:南京邮电大学,2016.
- Wang Kai-ci. User Behavior Prediction and Point of Interest Recommendation Based on LBSN Check-in Data[D]. Nanjing: Nanjing University of Posts and Telecommunications, 2016
- [13] 张志军.社交网络中个性化推荐模型及算法研究[D].山东:山东师范大学,2015.
- Zhang Zhi-jun. Research on Personalized Recommendation Model and Algorithm in Social Network[D]. Shandong: Shandong Normal University, 2015
- [14] Qian X, Feng H, Zhao G, et al. Personalized recommendation combing user interest and social circle[J]. IEEE Transaction on Knowledge Data Engineering, 2014, 26(7): 1763-1777.
- [15] 李征,段垒,王亚鲁.基于用户相似性传递的协同过滤推荐方法[J].河南大学学报(自然科学版),2019,49(1):61-68.
- Li Zheng, Duan Lei, Wang Ya-lu. Collaborative Filtering Recommendation Method Based on User Similarity Propagation[J]. Journal of Henan University(Natural Science), 2019, 49(1): 61-68.

[16]Pan Z, Cui L, Wu X, et al. Deep potential geo-social relationship mining for point-of-interest recommendation[J], IEEE Access, 2019, 7: 99496–99507.

[17]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.

[18]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.

[19]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: ACM, 2014:739-748.

[20]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: AAAI, 2012: 17-23.

[21]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.

[22]Liu B, Fu Y, Yao Z, et al. Learning Geographical Preferences for Point-of-Interest Recommendation [C]// Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining, Chicago: KDD, 2013: 1043-1051.

[23]杜亚洲, 张书钦, 王金洋等. 融合地理信息的兴趣点推荐[J]. 计算机科学与应用, 2020,10(4):629-640.

Du Ya-zhou, Zhang Shu-qin, Wang Jin-yang et al. Point of Interest Recommendation Based on Geographic Information[J]. Computer Science and Application, 2020,10(4):629-640.

[24]Brightkite. Accessed: February. 2020. [Online]. Available: <http://snap.stanford.edu/data/loc-brightkite.html>.