

# 地理分层结构与社会关系的兴趣点推荐

李昱杭, 杨 艳, 高静远

(黑龙江大学, 黑龙江 哈尔滨 150000)

**摘 要:** 随着基于地理位置的社交网络的兴起, 兴趣点(POI)推荐引起了人们的许多关注。POI推荐向用户推荐他们可能感兴趣但没有访问过的地方, 从而解决用户“下一步去哪”的问题。本文提出新的用户相似性度量、全局影响力以及热门POI的概念。综合考虑了多个影响因素之间的关系, 以地理分层结构的矩阵分解模型(HGMF)为基础, 提出新的POI推荐算法HGS-MF。在Yelp和Gowalla社交网络数据集上对HGS-MF进行了评估。实验结果表明, HGS-MF方法的实验表现均优于传统的POI推荐算法。

**关键词:** 矩阵分解; 地理分层结构; 社会关系; 推荐; 兴趣点

**中图分类号:** TP311 **文献标识码:** A

## Geographical Hierarchy and Social Relations for POI Recommendation

LI Yuhang, YANG Yan, GAO Jingyuan

(Heilongjiang University, Harbin 150000, China)

**Abstract:** With the rise of location-based social networks, Points of Interest (POI) recommendations have attracted a lot of attention. The POI recommends places which users may be interested in but have not visited, thus addressing the user's issue about *Where to Go Next*. This paper proposes new user similarity measures, global influences, and the concept of popular POIs. Based on the relationship between multiple influencing factors, a new POI recommendation algorithm HGS-MF is proposed based on the Hierarchical Geographical Matrix Factorization model (HGMF). HGS-MF is evaluated on the social network datasets of Yelp and Gowalla. The experimental results show that the performance of the HGS-MF method is superior to that of the traditional POI recommendation algorithm.

**Keywords:** matrix factorization; geographical hierarchy; social relationship; recommendation; Points of Interest

### 1 引言(Introduction)

随着移动设备的普及与定位技术的进步, 越来越多的人通过社交软件分享生活。由于大规模移动记录的易得性和社交网络信息的广泛性, POI推荐成为热门的话题, 很多研究人员通过对用户的签到记录进行分析, 从而获取用户的偏好。例如: Zhao等人<sup>[1]</sup>、Lian等人<sup>[2]</sup>、Liu等人<sup>[3]</sup>对用户的签到

数据进行分析, 发现用户签到行为具有一定的地理规律, 通过签到行为可以将完全不吸引用户的地点与未被发现但可能吸引用户的地方进行区分。此外, POI的类别具有一定的层级关系, 某些类别的POI可以进一步地划分, 用户的内容偏好也同样呈现出一定的层次结构<sup>[4]</sup>。

然而, 当前的大多数矩阵分解模型都存在一个问题: 它

们只考虑用户和POI的偏好，没有考虑到用户的多种社交关系和最近一段时间热门的POI对用户的影响，例如，年轻人可能喜欢去他仰慕的公众人物签到过的POI进行签到，用户也有更大的可能性喜欢去与他具有相似行为的用户去过的POI进行签到。此外，年轻人可能会去最近比较火爆的店签到，因此，只考虑用户偏好和POI的属性，来推测这个用户的行为会降低POI推荐的准确率。

为了进一步提高实验表现，本文在之前的方法基础上提出一种新的用户的社交关系，并提出全局影响力与热门POI的概念，向用户进行更为准确的POI推荐，提出了一种地理分层结构与社会关系的推荐算法(Hierarchical Geographic and Social Matrix Factorization, HGS-MF)，实验结果表明，在Yelp和Gowalla两个真实数据集上与相关的算法相比较，该算法在召回率，准确率等评价指标上相对于传统方法都得到了显著提高。

## 2 相关工作(Related work)

近些年，随着地理位置的社会网(LBSN)的发展，研究者可以更方便地分析用户的签到记录进而分析用户的行为习惯，从而得到其偏好。许多研究者尝试提高POI推荐模型，从而提高实验效果。

在传统的基于用户的协同过滤模型基础上，文献[5]—文献[6]融合用户偏好、社会影响和地理影响因素，旨在提高推荐的准确性，但它没有考虑POI具有一定地理层级的关系，比如美术家更喜欢去美术博物馆签到，而不是普通博物馆；Li等人<sup>[7]</sup>根据社交朋友和类似用户的签到记录建模，从而提高模型的实验表现，但它考虑社交因素不够全面，并未考虑公众人物对用户的影响。

在传统的基于时间影响的推荐模型基础上，文献[8]—文献[9]提出时间感知的协同过滤推荐模型，通过结合时间因素和地理影响因素，根据人们日常生活习惯划分时间轴，并将中午时段作为午餐时段，但是每个人的生活作息不同，用部分人的日常生活习惯推测每个人的行为是不准确的，并且会降低POI推荐的效果。

在传统的基于矩阵分解的推荐模型基础上，Li等人<sup>[10]</sup>通过结合时间影响，考虑用户偏好、地理图形影响和用户移动性行为，利用地理因子分解方法实现的POI推荐，但是忽略了社交关系对用户的重要影响；Hu等人<sup>[11]</sup>根据用户签到频数，提取用户对POI的偏好，从而实现POI推荐，但是它没有考虑地理因素影响，并且在社交关系方面考虑的不够全面；Zhao等人<sup>[1]</sup>、Lian等人<sup>[2]</sup>分析用户的签到数据，发现用户签到行为具有地理规律，并且发现POI的类别具有一定的层级关系，但

其未考虑用户的社会关系，并且在POI的影响力方面考虑的不全面。

这些算法虽然在一定程度上解决了POI推荐领域上的一些问题，但是仍具有一定的局限性，例如：用户相似性的影响因素选取过于单一，没有充分挖掘POI的热门信息，没有充分利用用户的社交关系。本文基于分层地理矩阵分解模型。提出一种新的用户的社交关系，并提出全局影响力与热门POI的概念，从而向用户进行更为准确的POI推荐。

## 3 HGS-MF算法(HGS-MF algorithm)

### 3.1 基础模型

#### 3.1.1 加权矩阵分解模型

给定用户签到频数矩阵 $C \in \mathbb{R}^{M \times N}$ ，其中 $M$ 、 $N$ 分别为用户的个数和POI个数，矩阵中每一项 $c_{i,j}$ 表示用户 $u_i$ 在POI  $v_j$ 签到的次数。

在POI推荐中，由于用户去一个POI签到的频数反映用户对POI的喜好程度，所以签到频数越高表示该POI越受欢迎。然而用户签到频数矩阵仅提供可以观察到的正例，这是众所周知的单类协同过滤问题，该问题的处理方法是每个用户未访问过的POIs，即负例，给它们分配比正例小的权重<sup>[1]</sup>。设加权矩阵如下：

$$w_{i,j} = 1 + \log(1 + c_{i,j}/\varepsilon) \quad (1)$$

其中， $\log(1 + c_{i,j}/\varepsilon)$ 为 $c_{i,j}$ 的递增函数， $\varepsilon = 10^{-8}$ 。

结合传统的矩阵分解模型，把用户兴趣和POI特征抽象成 $K$ 个属性，映射为 $K$ 维潜在空间，其中 $K \ll \min(M, N)$ 。用户对POI的偏好可定义为它们潜在空间的内积，如公式(2)所示。

$$\min_{P,Q} \|W \odot (R - PQ^T)\|_F^2 + \gamma(\|P\|_F^2 + \|Q\|_F^2) \quad (2)$$

其中， $\odot$ 为Hadamard乘积， $W \in \mathbb{R}^{M \times N}$ 为加权矩阵，矩阵中每一项 $w_{i,j}$ 表示用户 $u_i$ 对POI  $v_j$ 偏好程度。 $R \in \mathbb{R}^{M \times N}$ 为用户偏好矩阵，矩阵中每一项 $r_{i,j}$ 表示用户 $u_i$ 是否访问过POI  $v_j$ ，根据用户签到频数矩阵 $C$ 获得，签到频数 $c_{i,j}$ 大于0时， $r_{i,j}$ 的值为1，否则， $r_{i,j}$ 的值为0。 $P \in \mathbb{R}^{M \times K}$ 为用户潜在偏好矩阵，矩阵中每一项 $p_{i,j}$ 表示用户 $u_i$ 对属性 $j$ 的偏好程度，其中 $K$ 为潜在空间维度， $Q \in \mathbb{R}^{N \times K}$ 为POI潜在属性矩阵，矩阵中每一项 $q_{i,j}$ 表示属性 $j$ 属于POI  $v_i$ 的概率。 $\|\cdot\|_F$ 表示矩阵的 $F$ -范数，即矩阵元素的绝对平方和的平方根， $\gamma$ 用来避免泛化误差，是控制用户内容和POI特征的权重参数。

#### 3.1.2 HGMF模型

文献[2]发现如果用户经常访问POI  $v_j$ 但是对于 $v_j$ 周围的POI一直没有访问行为，这些围绕 $v_j$ 但是未被访问的POI，用户的访问意愿较低。于是，Lian等人<sup>[2]</sup>改进加权矩阵分解模型，将整个地理区域划分为具有相同大小的 $L$ 个网格，表示为

$L = \{g_1, g_2, \dots, g_l\}$ , 并提出GeoMF模型如公式(3)所示。

$$\min_{P, Q} \|W \odot (R - PQ^T - XY^T)\|_F^2 + \gamma(\|P\|_F^2 + \|Q\|_F^2) + \lambda \|X\|_1 \quad (3)$$

其中,  $X \in \mathbb{R}^{M \times L}$  为用户的活动区域概率矩阵, 矩阵中每一项  $x_{i,l}$  表示用户  $u_i$  出现在网格区域  $g_l$  的概率。  $Y \in \mathbb{R}^{N \times L}$  为POI区域影响矩阵, 矩阵中每一项  $y_{j,l}$  表示POI  $v_j$  对网格区域  $g_l$  的影响力。  $\lambda$  是控制用户活动区域的权重参数,  $\|X\|_1$  用来限制矩阵的稀疏性。

分层地理矩阵分解算法(Hierarchical Geographical Matrix Factorization, HGMF): HGMF<sup>[1]</sup>是一个基于地理分层的加权矩阵分解模型, 文献[1]指出, 用户潜在偏好矩阵  $P$ , POI潜在属性矩阵  $Q$ , 用户活动区域概率矩阵  $X$ , 仍然具有一定的潜在属性, 可以被进一步的分解, 例如, 艺术家更喜欢去艺术博物馆签到, 而不是普通博物馆。

HGMF模型首先对POI潜在属性矩阵  $Q$  进行矩阵分解, 得到  $Q_1 \in \mathbb{R}^{N \times N_1}$  与  $\tilde{Q}_2 \in \mathbb{R}^{N_1 \times K}$ , 其中  $N_1$  为第二层POI的潜在子属性的数量, 从而进一步分解得到POI的第  $q$  层的潜在子类别矩阵  $Q_q \in \mathbb{R}^{N_q \times K}$ , 分解过程如公式(4)所示。

$$Q \approx Q_1 Q_2 \dots Q_q \quad (4)$$

同理, 对用户潜在偏好矩阵  $P$ , 用户活动区域概率矩阵  $X$ , 也可以进行相似分解, 具体分解过程, 如图1所示。其中矩阵  $\tilde{R}$  为预测用户偏好矩阵, 可模拟为用户潜在偏好矩阵  $P$  与POI潜在属性矩阵  $Q$  的内积与用户活动区域概率矩阵  $X$  与POI区域影响矩阵  $Y$  的内积之和。

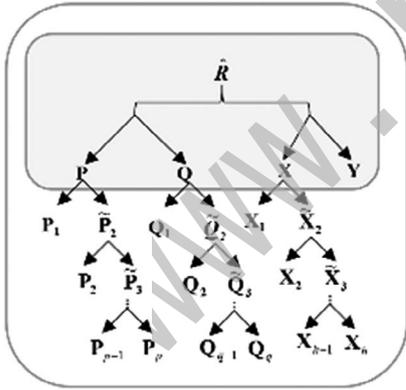


图1 HGMF和GeoMF的图形模型

Fig.1 The graphical model for HGMF and GeoMF

因此, 根据HGMF算法可得到如公式(5)所示的目标函数。

$$\min_{\substack{P_1, \dots, P_p, \\ Q_1, \dots, Q_q, \\ X_1, \dots, X_n}} \|W \odot \left( R - \prod_{i=1}^p P_i \prod_{j=1}^q Q_j^T - \prod_{t=1}^h X_t Y^T \right)\|_F^2 + \gamma \left( \sum_{i=1}^p \|P_i\|_F^2 + \sum_{j=1}^q \|Q_j\|_F^2 \right) + \lambda \sum_{t=1}^h \|X_t\|_1 \quad (5)$$

其中,  $\odot$  为Hadamard乘积,  $W$  为加权矩阵,  $R$  为用户偏好矩阵,  $P_i$  为用户潜在偏好的第  $i$  层的潜在子类别矩阵,  $Q_j$  为POI潜在偏好的第  $j$  层的潜在子类别矩阵,  $X_t$  为用户活动区域概率的

第  $t$  层的潜在子类别矩阵。

## 3.2 HGS-MF算法

本算法在HGMF模型的基础上进行改进, 提出新的用户相似性度量、全局影响力和热门POI的概念。

### 3.2.1 社会关系

在现实生活中, 用户在访问某POI时, 更倾向于倾听朋友的意见。用户的朋友往往与用户拥有相同的行为习惯, 在用户未访问过的POI, 用户往往会听从朋友的推荐<sup>[12]</sup>。

本文认为社交关系由用户相似度与用户的影响力两方面组成。在目前已有的用户相似度计算中, 只考虑在相同POI签到的用户的相似性, 以及拥有公共好友的用户相似性, 并没有考虑到用户的多种社交关系和最近一段时间热门的POI对用户的影响。本文在这两种相似度的基础上提出时间影响下的用户相似性, 并且在目前已有的POI推荐中, 没有在社交关系中考虑到用户的影响力因素, 本文提出了用户全局影响力, 并融入社交关系中。

定义1(用户好友关系矩阵): 设  $F$  为用户的好友列表, 满足  $F \in \mathbb{R}^{M \times N}$ , 矩阵中每一项  $f_{i,j}$  表示用户  $u_i$  与用户  $u_j$  的好友关系, 如果满足好友关系, 则  $f_{i,j}$  值为1, 否则为0。

定义2(时间影响用户行为矩阵): 设  $T$  为在时间影响下的用户的行为矩阵, 本文将时间分为24小时, 满足  $T \in \mathbb{R}^{M \times N \times 24}$ , 矩阵中每一项  $t_{i,j,t}$  表示用户  $i$  在  $(t-1, t)$  时间段内在POI  $v_j$  的签到情况, 如果签到则为1, 否则为0。

在用户相似度方面, 本文将其分为三类:

(1) 在相同POI签到的用户更相似, 由于用户可能多次去同一个POI签到, 为了更准确地展示相似度, 本文采用Pearson相似度计算, 用户  $u_x$  与用户  $u_y$  的相似度如公式(6)所示。

$$\text{corr}(x, y) = \frac{\sum_j^N ((c_{x,j} - \bar{c}_x)(c_{y,j} - \bar{c}_y))}{\sqrt{\sum_j^N (c_{x,j} - \bar{c}_x)^2} \sqrt{\sum_j^N (c_{y,j} - \bar{c}_y)^2}} \quad (6)$$

其中,  $c_{x,j}$  与  $c_{y,j}$  为用户  $u_x$  与用户  $u_y$  在POI  $v_j$  的签到次数,  $\bar{c}_x$  与  $\bar{c}_y$  为用户  $u_x$  与用户  $u_y$  的平均签到次数。

(2) 拥有公共好友越多的用户越相似, 本文采用Jaccard相似度计算, 用户  $u_x$  与用户  $u_y$  的相似度如公式(7)所示。

$$\text{Jacc}(x, y) = \frac{|F_x \cap F_y|}{|F_x \cup F_y|} \quad (7)$$

其中,  $F_x$  与  $F_y$  表示用户  $u_x$  与用户  $u_y$  的好友关系。

(3) 在相同时间段, 在同一POI签到的用户更相似具有更高的相似性, 因为他们具有更相似的生活作息, 本文采用余弦相似度计算, 用户  $u_x$  与用户  $u_y$  的相似度如公式(8)所示。

$$\text{cosine}(x, y) = \prod_{t=1}^{24} \frac{\mathbf{T}_{x,t} \cdot \mathbf{T}_{y,t}}{\left| \sum_k^N \mathbf{T}_{x,k,t} \right| \left| \sum_k^N \mathbf{T}_{y,k,t} \right|} \quad (8)$$

其中,  $\mathbf{T} \in \mathbb{R}^{M \times N \times 24}$ ,  $\mathbf{T}_{x,t}$ 与 $\mathbf{T}_{y,t}$ 表示用户 $u_x$ 与用户 $u_y$ 在 $(t-1, t)$ 时间段内所签到的POI向量。

在用户影响力方面, 本文考虑社交关系的有向性。现实生活中知名人物往往拥有较高的关注度, 即用户 $u_x$ 关注用户 $u_y$ , 但用户 $u_y$ 可能不关注用户 $u_x$ , 本文获取有向社会关系图, 并用矩阵 $G$ 存储, 部分社会关系图, 如图2所示。图中节点代表用户, 用户 $u_x$ 与 $u_y$ 之间的边表示用户 $u_x$ 关注用户 $u_y$ , 由于具有更多关注的知名人物具有更高的影响力, 本文为每个用户计算其全局影响力, 如公式(9)所示。

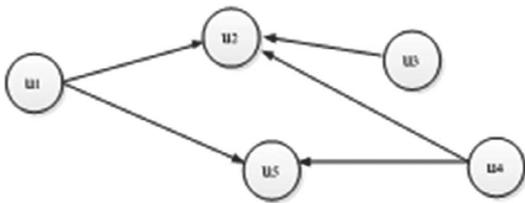


图2 部分有向社会关系图

Fig.2 Partially directed social relations graph

$$\text{inf}_x = \frac{\text{ind}_x}{M} \quad (9)$$

其中,  $\text{ind}_x$ 表示用户 $u_x$ 在社会关系图中的入度数,  $M$ 为用户数。

定义3(用户影响力矩阵): 设 $\mathbf{IN}$ 为用户影响力矩阵, 满足满足 $\mathbf{IN} \in \mathbb{R}^{M \times M}$ , 对于矩阵中每一项 $\text{in}_{x,y}$ , 如果 $G_{x,y} = 1$ , 即用户 $u_y$ 对用户 $u_x$ 有影响, 则 $\text{in}_{x,y} = \text{inf}_y$ , 如果 $G_{x,y} = 0$ , 即用户 $u_y$ 对用户 $u_x$ 没有影响, 则 $\text{in}_{x,y} = 0$ 。

将用户相似度, 用户影响力线性组合, 得到用户的社会关系, 如公式(10)所示。

$$\text{rel}(x, y) = \alpha \cdot \text{corr}(x, y) + \beta \cdot \text{Jacc}(x, y) + \varphi \cdot \text{cosine}(x, y) + \eta \cdot \mathbf{IN}_{x,y} \quad (10)$$

其中,  $\alpha + \beta + \varphi + \eta = 1$ , 均为可调节参数。

定义4(用户社会关系矩阵): 设 $S$ 为用户社会关系矩阵, 满足 $S \in \mathbb{R}^{M \times M}$ , 矩阵中每一项 $s_{i,j}$ 表示用户 $u_i$ 与用户 $u_j$ 的社会关系, 即 $\text{rel}(i, j)$ 。

### 3.2.2 热门POI

当代生活, 随着基于地理位置的社会网蓬勃发展, 更多的人会去近期人气火热的POI。例如近些年流行起来的蹦极运动、玻璃栈道等, 这些POI不仅局限于娱乐分类, 也可以是饭店等其他分类, 这些地点可以在一定程度上吸引更多的用户。

HGMF算法中的POI区域影响矩阵 $Y$ , 其行向量 $y_i$ 是POI  $v_j$ 对每一网格区域 $g_{i,t} \in L$ 影响力的向量, 它满足以POI  $v_j$ 为中心的正态分布, 然而它只考虑POI  $v_j$ 与网格区域 $g_{i,t}$ 距离因

素。本文在距离因素的基础上, 融入热门POI的影响力。

本文将热门POI定义为最近一段时间内频繁被签到的POI, 并且此热门POI的区域影响向量仍然满足以此热门POI为中心的正态分布, 于是修改POI区域影响矩阵 $Y$ , 矩阵中的每一项 $y_{j,i}$ 的值为 $y_{j,i} = \sqrt{2\pi}K\left(\frac{d(l,j)+h_j}{\sigma_c}\right)$ 。其中 $K(\cdot)$ 为标准正态分布,  $c$ 为POI  $v_j$ 的类别,  $d(l, j)$ 为网格区域 $g_{i,t}$ 与POI  $v_j$ 的距离, 最近一段时间内POI  $v_j$ 被访问的频率 $h_j$ 定义为 $h_j = \frac{\sum_i^M c_{i,j}}{\sum_k^N \sum_t^M c_{i,k}}$ 。其中,  $c_{i,j}$ 用户 $u_i$ 在POI  $v_j$ 的签到次数。

因此, 本文修改HGMF模型得到HGS-MF模型如公式(11)。

$$\hat{\mathbf{R}} = \prod_{i=1}^p P_i \prod_{j=1}^q Q_j^T + \prod_{t=1}^h X_t \hat{\mathbf{Y}}^T + S C \quad (11)$$

其中, 矩阵 $\hat{\mathbf{R}}$ 为预测用户偏好矩阵, 满足 $\hat{\mathbf{R}} \in \mathbb{R}^{M \times N}$ , 矩阵中每一项 $\hat{r}_{i,j}$ 表示用户 $u_i$ 对POI  $v_j$ 的预测偏好分数。矩阵 $\hat{\mathbf{Y}}$ 为融合热门POI因素的POI区域影响矩阵, 矩阵 $S$ 为用户社会关系矩阵, 矩阵 $C$ 为用户签到频数矩阵。

### 3.2.3 HGS-MF算法

综合考虑之前的两个因素, 本文将用户社交关系与热门POI的影响因素融合到HGMF模型中, 因此, 本文最终得到HGS-MF推荐算法的损失函数如公式(12)。

本文仍然采用HGMF模型的优化方法, 采用两步优化, 第一步优化只考虑平面结构, 利用交替最小二乘法进行优化, 其核心思想是固定其他参数使目标函数最小化。首先固定 $X$ 矩阵, 优化 $P$ 、 $Q$ , 得到 $P_i$ 的优化公式(13), 得到 $Q_j$ 的优化公式(14)。

$$\min_{\substack{P_i, Q_j \\ X_1, \dots, X_h}} \left\| \mathbf{W} \odot \left( \mathbf{R} - \prod_{i=1}^p P_i \prod_{j=1}^q Q_j^T - \prod_{t=1}^h X_t \hat{\mathbf{Y}}^T - S C \right) \right\|_F^2 + \gamma \left( \sum_{i=1}^p \|P_i\|_F^2 + \sum_{j=1}^q \|Q_j\|_F^2 \right) + \lambda \sum_{t=1}^h \|X_t\|_F^2 \quad (12)$$

s.t.  $X_t \geq 0, t \in \{1, 2, \dots, h\}$

$$P_i = (Q^T W^i Q + \gamma I)^{-1} Q^T W^i (r_i - Y x_i - C s_i) \quad (13)$$

其中,  $W^i$ 为 $N \times N$ 的对角矩阵,  $W^i_{j,j} = w_{i,j}$ 。

$$Q_j = (P^T W^j P + \gamma I)^{-1} P^T W^j (r_j - X y_j - S c_j) \quad (14)$$

其中,  $W^j$ 为 $M \times M$ 的对角矩阵,  $W^j_{i,i} = w_{i,j}$ 。

固定 $P$ 、 $Q$ , 利用梯度下降法优化 $X$ , 得到 $x_i$ 优化公式(15)。

$$x_i = H(x_i - \varepsilon(Q^T W^i (Y x_i - (r_i - Q P_i - C s_i)) + \lambda)) \quad (15)$$

其中, 参数 $\varepsilon$ 为学习率,  $H(\cdot)$ 为保证用户活动区域非负的函数, 如公式(16)所示。

$$H(x_i) = \begin{cases} x_i, & x_i > 0 \\ 0, & x_i \leq 0 \end{cases}, i \in \{1, 2, \dots, L\} \quad (16)$$

优化过程详见表格1。

表1 三个平面矩阵的优化算法

Tab.1 Optimization algorithm for three flat matrix

Algorithm 1: Optimization algorithm for three flat matrix, PQX

Input:  $K, k, \lambda, \gamma, \alpha, \beta, \varphi, \eta, C, \hat{Y}$ Output:  $P, Q, X$ 

```

1: Initialize  $P, Q, X$ 
2: repeat
3:   for each user  $u_i$  do
4:     update  $P_i$  as Eq.13
5:   end for
6:   for each POI  $v_j$  do
7:     update  $Q_j$  as Eq.14
8:   end for
9:   for each user  $u_i$  do
10:    update  $X_i$  as Eq.15
11:  end for
12: until reach the stop criterion

```

接下来，进行第二步分层优化，利用拉格朗日乘法，首先对 $P_i$ 进行优化，得到 $P_i$ 的优化公式(17)。

$$P_i(s,t) = \sqrt{\frac{(A_i^T (W \odot (R-B-SC) D_i^T))_{(s,t)}}{(A_i^T (W \odot (A_i D_i)) D_i^T + \gamma P_i)_{(s,t)}}} \quad (17)$$

其中， $A_i$ 、 $B$ 、 $D_i$ 满足如下公式：

$$\begin{aligned} A_i &= \prod_{j=1}^{i-1} P_j \\ B &= \prod_{t=1}^h X_t \hat{Y} \\ D_i &= \prod_{j=i+1}^p P_j \prod_{k=1}^q Q_k^T \end{aligned} \quad (18)$$

得到 $Q_j$ 的优化公式(19)。

$$Q_j(s,t) = \sqrt{\frac{(E_j^T (W \odot (R-B-SC) F_j^T))_{(s,t)}}{(E_j^T (W \odot (E_j F_j)) F_j^T + \gamma Q_j)_{(s,t)}}} \quad (19)$$

其中， $E_j$ 、 $F_j$ 满足如下公式：

$$\begin{aligned} E_j &= \prod_{i=1}^p P_i \prod_{k=q}^{j-1} Q_k^T \\ F_j &= \prod_{i=1}^{j-1} Q_i^T \end{aligned} \quad (20)$$

同理，得到 $X_k$ 的优化公式(21)。

$$X_k(s,t) = \sqrt{\frac{(N_k^T (W \odot (R-B-SC) O_k^T))_{(s,t)}}{(N_k^T (W \odot (N_k O_k)) O_k^T + \lambda X_k)_{(s,t)}}} \quad (21)$$

其中， $N_k$ 、 $O_k$ 满足如下公式：

$$\begin{aligned} N_k &= \prod_{i=1}^{k-1} X_i^T \\ O_k &= \prod_{t=k+1}^h X_t \hat{Y}^T \end{aligned} \quad (22)$$

HGS-MF模型的算法，详见表格2。

表2 HGS-MF模型

Tab.2 HGS-MF model

Algorithm 2: HGS-MF model

Input:  $K, k, \lambda, \gamma, p, q, h, \alpha, \beta, \varphi, \eta, C, \hat{Y}$ 

Output: top k POIs

```

1: Initialize  $\{P_i\}_{i=1}^p, \{Q_j\}_{j=1}^q, \{X_i\}_{i=1}^h$ 
2:  $\tilde{P}_i, \tilde{Q}_j, \tilde{X}_i \leftarrow PQX(K, k, \lambda, \gamma, \alpha, \beta, \varphi, \eta, C, \hat{Y})$ 
3: for i=1 to p-1 do
4:    $P_i, \tilde{P}_{i+1} \leftarrow MF(\tilde{P}_i, M_i)$ 
5: end for
6: for j=1 to q-1 do
7:    $Q_j, \tilde{Q}_{j+1} \leftarrow MF(\tilde{Q}_j, N_j)$ 
8: end for
9: for t=1 to h-1 do
10:   $X_t, \tilde{X}_{t+1} \leftarrow MF(\tilde{X}_t, K_t)$ 
11: end for
12: repeat
13:   for i=1 to p-1 do
14:     update  $A_i, B, D_i$  as Eq.18
15:     update  $P_i$  as Eq.17
16:   end for
17:   for j=1 to q-1 do
18:     update  $E_j, F_j$  as Eq.20
19:     update  $Q_j$  as Eq.19
20:   end for
21:   for k=1 to q-1 do
22:     update  $N_k, O_k$  as Eq.22
23:     update  $X_k$  as Eq.21
24:   end for
25: until reach the stop criterion
26:  $\hat{R} = \prod_{i=1}^p P_i \prod_{j=1}^q Q_j^T + \prod_{t=1}^h X_t \hat{Y}^T + SC$ 

```

首先，在算法的第2行中，需要对三个平面矩阵进行学习，并且获取POI区域影响矩阵 $Y$ ，如算法1所示。从第3行到第11行，进一步分解用户潜在偏好矩阵 $P$ ，POI潜在属性矩阵 $Q$ ，用户活动区域概率矩阵 $X$ 。从第12行到第24行，对每个层次的矩阵 $P_i$ 、 $Q_j$ 、 $X_k$ 进行学习，直到达到相应的停止标准，最后得到预测用户偏好矩阵 $\hat{R}$ 。

## 4 实验(Experiment)

### 4.1 数据集与实验环境

本文选用真实数据集来对HGS-MF模型进行训练并进行测试。Yelp与Gowalla数据集在POI推荐中广泛应用，它们不仅具有大量的位置信息与签到信息，还具有社交关系信息，数据集统计信息见表3。

表3 数据集

Tab.3 Datasets

数据集	签到数	用户数	POI数
Yelp	336443	33086	18993
Gowalla	651154	18737	32510

为克服数据稀疏性，本文只选用签到次数超过15次的用户，并从中随机选择80%作为训练集，20%作为测试集。

本文实验环境为macOS 10.14.4操作系统，8GB内存，Intel Core i7，实验程序为C++编写，使用Xcode10.2.1版本，实验使用开源的OpenBLAS<sup>[13]</sup>，此开发包对矩阵乘法运算进行优化，将矩阵乘法的时间复杂度优化为 $O(n^2 \log n)$ 。

### 4.2 评价指标

本文预测用户对未访问过的POI的偏好分数，将top- $k$  POI推荐给用户。利用准确率(precision@  $k$ )，召回率(recall@  $k$ )来衡量HGS-MF模型与其他模型，其公式表示如下：

$$\begin{aligned} \text{precision@ } k &= \frac{1}{M} \sum_{p=1}^M \frac{|S_p(k) \cap V_p|}{k} \\ \text{recall@ } k &= \frac{1}{M} \sum_{p=1}^M \frac{|S_p(k) \cap V_p|}{|V_p|} \end{aligned} \quad (23)$$

其中， $S_p(k)$ 是测试集中推荐给用户的 $k$ 个POI地点的集合， $V_p$ 是用户 $u_p$ 已经访问过的POIs的集合。

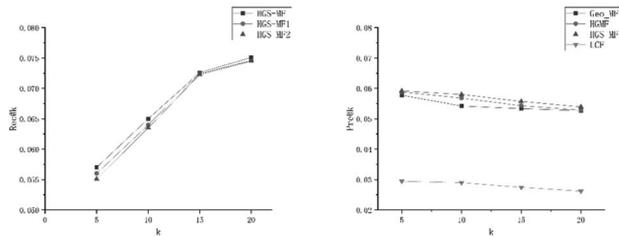
### 4.3 模型对比

本文将HGS-MF模型拆分为如下两个模型：

HGS-MF1：它是单独考虑社交关系影响的HGS-MF模型。

HGS-MF2：它是单独考虑热门POI影响的HGS-MF模型。

模型在Gowalla数据集上的对比，如图3所示。



(a)精确率@k-Gowalla (b)召回率@k-Gowalla

图3 owalla 数据集上的比较结果

Fig.3 Comparison on the Gowalla datasets

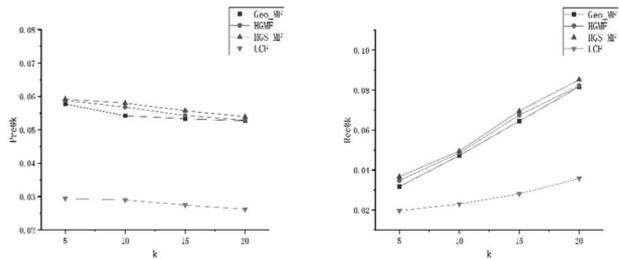
本文首先选取三种模型与HGS-MF模型进行对比，三种模型详细信息如下：

UCF<sup>[14]</sup>：它是一种面向用户的协同过滤模型，基于用户矩阵使用余弦相似度来衡量用户相似性。

GeoMF<sup>[2]</sup>：它是矩阵分解中较为先进的方法，它首先结合了地理信息，考虑了用户的因素和POI的影响因素，利用核密度估计观察空间聚集现象。

HGMF<sup>[1]</sup>：在GeoMF基础上，利用矩阵分解，考虑了用户的内容和空间偏好的层次结构。

经多次实验，选定各个模型最优参数， $k$ 的值分别设置为5、10、15、20。每次改变 $k$ 的值，则为每个算法计算准确率和召回率，考虑到实验的效果和有效性， $\alpha$ 、 $\beta$ 、 $\varphi$ 、 $\eta$ 为控制社交关系与热门POI影响的权重参数，通过多组参数交叉验证，将其设置为 $\alpha=0.3$ 、 $\beta=0.2$ 、 $\varphi=0.3$ 、 $\eta=0.2$ ，实验结果如图4与图5所示。



(a)精确率@k-Yelp (b)召回率@k-Yelp

图4 Yelp数据集上的比较结果

Fig.4 Comparison on the Yelp datasets

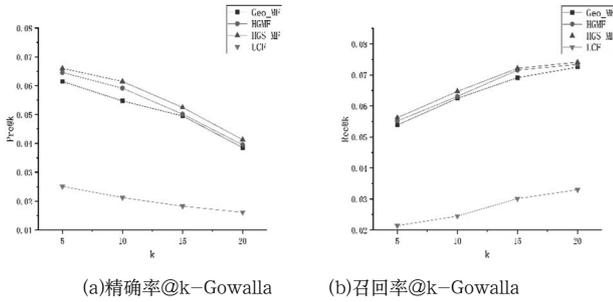


图5 Gowalla数据集上的比较结果

Fig.5 Comparison on the Gowalla datasets

根据对比可以看出,在参数均最优的情况下,HGS-MF模型在精确率与召回率两方面都高于其他算法。

## 5 结论(Conclusion)

本文利用用户的社会关系,提出了一个新的用户相似性度量、全局影响力以及热门POI的概念,在此基础上提出结合分层地理矩阵分解模型得到HGS-MF模型,利用两步优化,第一步优化采用交替最小二乘法和梯度下降方法求解单层模型,第二步采用KKT条件优化分层模型。实验结果表明,融合社会关系与热门POI影响力的加权矩阵分解能够有效地提高推荐性能,模型优于传统的POI推荐算法。未来将尝试将本文提出的模型和深度学习相结合,期待进一步提高兴趣点推荐性能。

## 参考文献(References)

- [1] Zhao P,Xu X,Liu Y,et al.Exploiting Hierarchical Structures for POI Recommendation[J].IEEE International Conference on Data Mining (ICDM),2017:655-664.
- [2] 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 & Data Mining.ACM,2014:831-840.
- [3] Liu B,Fu Y,Yao Z,et al.Learning geographical preferences for point-of-interest recommendation[C].Acm Sigkdd International Conference on Knowledge Discovery & Data Mining.ACM,2013:1043-1051.
- [4] Moreno-Jimenez J M,Vargas L G.A Probabilistic Study of Preference Structures in the Analytic Hierarchy Process with Interval Judgments[J].Mathematical and Computer Modelling,1993,17(4-5):73-81.
- [5] Ye M,Yin P,Lee W C .Location recommendation for location-based social networks[C].Acm Sigspatial International Symposium on Advances in Geographic Information Systems. DBLP,2010:458.
- [6] Ye M,Yin P F,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.ACM,2011:325-334.
- [7] Li H,Hong R,Zhu S,et al.Point-of-Interest Recommender Systems:A Separate-Space Perspective[C].2015 IEEE International Conference on Data Mining (ICDM).IEEE Computer Society,2015:231-240.
- [8] Yuan Q,Cong G,Ma Z,et al.Time-aware point-of-interest recommendation[C].International Acm Sigir Conference on Research & Development in Information Retrieval. ACM,2013:363-372.
- [9] Yuan Q,Cong G,Sun A.Recommendation with Geographical and Temporal Influences[C].Acm International Conference on Conference on Information & Knowledge Management. ACM,2014:659-668.
- [10] LiX T,CongG,et al.Rank-GeoFM:A Ranking based Geographical Factorization Method for Point of Interest Recommendation[C].Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval,2015:433-442.
- [11] Hu Y,Koren Y,Volinsky C.Collaborative Filtering for Implicit Feedback Datasets[C].Eighth IEEE International Conference on Data Mining.2008:263-272.
- [12] Wang H,Terrovitis M,Mamoulis N.Location recommendation in location-based social networks using user check-in data[C]. Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. ACM,2013:364-373.
- [13] Xianyi Z,Qian W,Yunquan Z.Model-driven Level 3 BLAS Performance Optimization on Loongson 3A Processor[C].2012 IEEE 18th International Conference on Parallel and Distributed Systems.IEEE,2012:684-691.
- [14] Sarwar B,Karypis G,Konstan J,et al.Item-based collaborative filtering recommendation algorithms[C].International Conference on World Wide Web,2001:285-295.

## 作者简介:

- 李昱杭(1997-),女,硕士生.研究领域:社会网,数据挖掘.  
 杨艳(1975-),女,博士,教授.研究领域:社会网,数据挖掘.本文通讯作者.  
 高静远(1994-),男,硕士生.研究领域:社会网,POI推荐.