

融合隐语义和邻域算法的兴趣点推荐模型*

吴海峰^{1†}, 张书奎^{1,2}, 林政宽¹, 贾俊铖¹

(1. 苏州大学 计算机科学与技术学院, 江苏 苏州 215006; 2. 江苏省无线传感网高技术研究重点实验室, 南京 210003)

摘要: 随着众多具有传感功能的智能手机和可穿戴设备的普及, 基于位置的服务得到了快速发展, 其中基于位置的社交网络(location-based social networks, LBSN)逐渐被大多数人所接受, 基于位置社交网络可以为人们提供兴趣点推荐服务。为了提供更加精准的兴趣点推荐服务, 提出了一种融合的算法模型。通过隐语义分析算法来充分挖掘用户的历史行为, 使用基于邻域的方法结合好友和地理位置等因素, 然后在统一的框架中融合这两种推荐方式的结果, 实现了对用户行为更好的预测。实验结果表明, 提出的兴趣点推荐方法拥有较好的准确率和召回率。

关键词: 基于位置的社交网络; 兴趣点推荐; 隐语义; 信息融合

中图分类号: TP301 **文献标志码:** A **文章编号:** 1001-3695(2018)07-1955-05

doi: 10.3969/j.issn.1001-3695.2018.07.007

Synthetic recommendation model for point-of-interest: fusion latent factor and neighborhood-based algorithm

Wu Haifeng^{1†}, Zhang Shukui^{1,2}, Lin Zhengkuan¹, Jia Juncheng¹

(1. School of Computer Science & Technology, Soochow University, Suzhou Jiangsu 215006, China; 2. Jiangsu High Technology Research Key Laboratory for Wireless Sensor Networks, Nanjing 210003, China)

Abstract: Recently, with the popularity of wearable devices and smart phones which have sensing capabilities, mobile positioning technique is gradually mature. Meanwhile, location-based services have been developed rapidly and the social network included in it, called location-based social networks (LBSN), has gradually been accepted by most people. It can provide point of interest recommendation service. In order to provide a more accurate point of interest recommended service, this paper presented a fusion algorithm model. It excavated the previous behaviors of users sufficiently through latent factor algorithm, and used the neighborhood-based algorithm considering other factors such as friends and geographical position. And then fused results of this two recommended ways based on the unified framework which achieved a better prediction of user behavior. The experimental results show that the point of interest recommendation method has better precision and recalling rate.

Key words: LBSN; point of interest recommendation; latent factor; data fusion

0 引言

当今时代, 移动互联网技术飞速发展, 推动了智能手机和各种智能可穿戴设备的普及, 带动了移动社交网络技术走向成熟。各种定位技术的进步, 催生了基于位置的社交网络(location-based social networks, LBSN)服务。LBSN允许用户以签到的方式分享自己的位置信息, 从而产生大量的签到数据, 其中包括用户的位置信息和情境信息如用户的社交关系、时间、类别等, 如图1所示。目前国外比较主流的LBSN有Foursquare、 Gowalla、 Geolife^[1], 国内的有嘀咕、街旁、四方^[2]。

相对于传统的社交网络, LBSN引入了用户的位置信息, 能够帮助人们更方便地了解周围的信息以探索周围的环境从而辅助自己的决策。为了达到这种目的, 兴趣点推荐就成为了一种重要的方式。Foursquare利用用户实时的位置和当前的分类信息为用户推荐兴趣点, 如用户可能想要知道自己周围有哪些可以就餐的中式餐馆^[3]; Yu等人^[4]通过来自位置社交网络的数据来分析用户的消费模式, 从而判断某一区域最适合开设商店的类型; Zhao等人^[5]通过移动终端和服务网站收集用户

的位置和社交信息, 对基于位置的服务进行评级打分; He等人^[6]通过语义标注技术自动地为地点进行语义标注, 从而为用户推荐兴趣点。本文通过研究用户的历史信息, 挖掘用户与地点之间隐式存在的关联关系, 然后为用户推荐他们可能会感兴趣的地点信息。

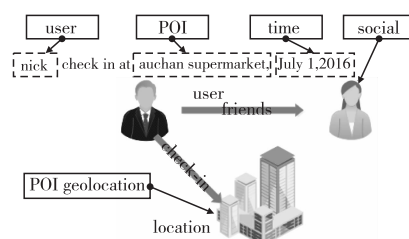


图1 基于位置的社交网络

当前的位置社交网络兴趣点推荐系统大多采用基于邻域的或者基于模型的方法。基于邻域的算法简单易于实现, 它只是简单地对照用户之间共同访问地点的数目, 没有考虑用户之间偏好的相似度可能仅仅集中于某一个分类。例如用户1喜欢中餐也喜欢西餐, 用户2也喜欢中餐但不喜欢西餐,

收稿日期: 2017-03-20; **修回日期:** 2017-05-18 **基金项目:** 国家自然科学基金资助项目(61201212); 江苏省自然科学基金资助项目(BK2011376); 江苏省“六大人才高峰”项目(2014-WLW-010); 苏州市融合通信重点实验室项目(SKLC2013XX); 江苏省产学研前瞻性项目(BY2012114); 软件新技术与产业化协同创新中心部分资助项目

作者简介: 吴海峰(1993-), 男(通信作者), 河南人, 硕士研究生, 主要研究方向为社交网络推荐系统(hfwu@stu.suda.edu.cn); 张书奎(1966-), 男, 内蒙古人, 教授, 博导, 博士, 主要研究方向为无线传感器网络、分布式计算、智能信息处理、并行和分布式系统等; 林政宽(1974-), 男, 中国台湾人, 副教授, 硕导, 博士, 主要研究方向为并行与分布式系统、互连网络、无线网络、无线传感器网络、算法设计与分析、图论等; 贾俊铖(1973-), 男, 江苏人, 副教授, 硕导, 博士, 主要研究方向为移动互联网、物联网、无线通信网络。

如果两个人访问了很多相同的中餐馆,那么他们通过邻域的方式得到的兴趣相似度就会相当高,算法就会向用户 2 推荐西餐厅,这样的推荐方式显然是不合理的。另外,基于邻域的模型对功能极为相似的地点缺乏语义识别功能。基于模型的方式虽然可以解决这一问题,但需要评分数据作为样本,而签到数据中通常是不包含用户评分数据的。针对以上问题,本文以采集负样本形式为模型提供评分数据,使用 LFM (latent factor model) 充分挖掘用户历史行为,结合社交网络中好友影响和地理位置因素,提出了一种融合隐语义和邻域算法的兴趣点推荐模型。

本文的主要贡献如下:

a) 通过负样本采样的方式补充评分数据,以使用 LFM 模型充分挖掘用户的历史信息,解决了用户评分数据缺失和地点相似性问题,为最终推荐结果的产生提供更加准确的用户历史偏好因素。

b) 提出了一种线性的基于协同过滤的好友影响因素分析算法,全面考虑了社交因素对用户行为的影响。

c) 融合 LFM 挖掘的用户历史偏好、好友影响和地理位置因素形成推荐结果,发挥了各自的长处,实现了对用户行为更好的预测。

d) 在真实的 Gowalla 和 Foursquare 数据集上对论述的方法进行了充分的实验,验证了方法的有效性。与其他方法相比,本文提出的模型在推荐的准确率和召回率等评测指标上得到了提高。

1 相关工作

当前基于位置的社交网络推荐服务主要有地点推荐^[7-9]、好友推荐^[10-12]和社会活动推荐^[13,14]。其推荐方式主要有基于邻域的推荐和基于模型的推荐。

基于邻域的兴趣点推荐一般采用协同过滤算法^[15],以余弦相似度来计算用户之间的兴趣相似性。Ye 等人^[16]使用协同过滤的算法来提高兴趣点推荐算法的准确率,他们通过幂律分布来挖掘用户的签到特征,并提出了一种基于朴素贝叶斯的算法来融合用户偏好、地理位置和好友关系的推荐模型。Wang 等人^[17]深入分析了个性化地点推荐算法面临的数据稀疏、冷启动和签到地点距离用户常住地远近这三个问题,提出了一种改进的自适应地点推荐算法。该算法融合基于用户协同过滤、社交影响和朴素贝叶斯分类算法,并动态适应用户当前所在的具体位置,力求为用户推荐最适合用户当前状态的地点。Nunes 等人^[9]认为过去访问相似位置的用户在未来也倾向于访问相同的位置,而且用户喜欢在一个新的兴趣区域内集中签到。Nunes 等人基于这种猜想使用随机游走算法,通过一个融合的模型来探索地点的位置、距离与用户的连续签到行为之间的相互影响关系,从而产生推荐服务。

基于模型的方法大多采用奇异值分解的算法 (singular value decomposition, SVD),基本思想是把高维向量空间模型中的文档映射到低维的潜在语义空间中。隐语义模型是由 Funk 在 Netflix Prize 后公布的一种算法。该算法利用矩阵分解技术把用户评分矩阵分解为两个低维矩阵,然后用这两个低维矩阵来估计用户对项目的评分。隐语义模型经过多年的发展产生了很多有效的模型和方法,其中主要的技术有 PLSA、LDA、隐含类别模型 (latent class model)、隐含主题模型 (latent topic model) 等。隐语义分析技术有很多的优势,可以基于用户行为统计进行自动聚类,可以动态调整分类的细粒度;可以发现用户与商品之间隐含的关联关系等。Cheng 等人^[18]将地理位置因素融入到用户签到概率模型中,得到一个多中心高斯模型,将社交因素和地理位置因素融合到一个广义矩阵分解框架中来提高推荐的准确率。Hu 等人^[19]基于 NMF 矩阵分解模型融合用户的兴趣类别和地理位置信息进行兴趣点推荐,采用用户的邻居影响因素对地理位置信息进行建模。

除了上面两种推荐系统的典型算法类型,还有一些文章使用图论等方法来实现兴趣点的推荐。例如曹玖新等人^[20]基于三度影响力设置用户—兴趣点间元路径特征集;然后通过随机游走方法计算元路径特征值,以度量实例路径中的首尾节点间关联度,并利用监督学习方法获得各特征的权值;最后计算特定用户将来在各兴趣点的签到概率,从而生成推荐列表。Xu 等人^[21]通过情感监督随机漫步算法将用户的签到和评论信息融入到一个三方图中来预测用户的地点偏好,从而实现个性化的兴趣点推荐。

上文提到的这些模型都取得了不错的效果,但它们都由于侧重的方向单一而存在一些问题。基于邻域的方式不能充分挖掘用户的历史信息,在推荐过程中可能会出现较大的偏差;一些基于模型的推荐将所有的影响因素都融入在一起,但它们都没有考虑负样本的问题,而且缺乏很好的推荐解释。本文首先使用负样本采样来填充用户的历史偏好信息;然后使用隐语义模型来充分挖掘用户的历史偏好信息;接着使用基于协同过滤的邻域算法来分析用户的好友影响因素,使用基于类簇信息的邻域算法来分析地理位置自身属性对用户的影响因素;最后使用线性融合的框架将几种影响因素综合考虑得出最终的推荐结果。

2 推荐模型

本章详细介绍本文提出的融合模型中的算法,并给出了详细的推导过程和伪代码描述。

2.1 历史行为影响

用户的历史偏好是影响用户签到行为的一个重要因素。从用户的历史签到记录中可以得到用户—地点的签到数据矩阵,通常这样的矩阵是稀疏的。LFM^[22]通过将签到矩阵映射到一个低维的隐空间中来解决这一问题。LFM 一般是使用在具有评分数据的显性反馈数据集中,而本文中的用户签到数据只有隐性反馈数据,只有用户的签到记录而没有评分数据,可采用负样本采样的方式解决这一问题。负样本采样的原则是将那些很热门用户却没有行为的地点作为负样本。负样本采样的个数要与正样本的数目大致相似 (正样本表示用户感兴趣的地点,负样本表示用户不感兴趣的地点)。算法过程描述如下:

LFM 将签到矩阵 C_{uv} 分解为两个相乘的低维矩阵:

$$C_{uv} = P_u^T Q_v = \sum_{f \in F} P_{uf} Q_{vf} \quad (1)$$

其中: F 表示隐类集合; P 为用户—隐类矩阵; Q 为地点—隐类矩阵。如果 (u, v) 为正样本 $C_{uv} = 1$, 否则 $C_{uv} = -1$, 这样就可以得到一个具有正负样本的用户—签到集合。

通过优化下面的损失函数寻找到合适的 P 和 Q :

$$S = \min \sum_{(u,v) \in K} (C_{uv} - \sum_{f=1}^F P_{uf} Q_{vf})^2 + \lambda \|P_u\|^2 + \lambda \|Q_v\|^2 \quad (2)$$

其中: $\lambda \|P_u\|^2 + \lambda \|Q_v\|^2$ 是用来防止过拟合的正则化项; λ 可通过训练获得。然后对 P_{uf} 和 Q_{vf} 使用梯度下降法找到最快下降方向:

$$\frac{\partial S}{\partial P_{uf}} = -2Q_{vf}e + 2\lambda P_{uf} \quad (3)$$

$$\frac{\partial S}{\partial Q_{vf}} = -2P_{uf}e + 2\lambda Q_{vf} \quad (4)$$

其中: e 表示实际值与预估值之间的误差。对 P_{uf} 和 Q_{vf} 利用最速下降进行迭代更新:

$$P_{uf} = P_{uf} + \alpha(eQ_{vf} - \lambda P_{uf}) \quad (5)$$

$$Q_{vf} = Q_{vf} + \alpha(eP_{uf} - \lambda Q_{vf}) \quad (6)$$

其中: α 是算法的学习速率,它的选取可通过实验迭代过程动态调整。通过迭代得到 P_{uf} 和 Q_{vf} 矩阵,两者相对应的项相乘可以得到用户对地点的历史偏好兴趣度。其算法描述如算法 1 所示。

算法 1 隐语义分析算法 (latent factor analysis algorithm)

输入: 用户—地点签到矩阵 C , 参数 α, λ, n, F 。

```

1  随机初始化  $P, Q$  矩阵//根据正太分布进行初始化
2  for step in range(0, n) do
3    for user  $u$  in  $C$  do
4      为用户  $u$  采集负样本
5      for local  $v$  in  $C$  do
6         $e_{uv} = C_{uv} - E_{uv}$  //  $C_{uv}$  表示用户的真实评分  $E_{uv}$  表示
          预测的评分
7        for  $f$  in (0,  $F$ ) do //  $F$  表示隐类的数量
8           $P[u][f] + = \alpha * (e_{uv} * Q[v][f] - \lambda * P[u][f])$ 
9           $Q[v][f] + = \alpha * (e_{uv} * P[u][f] - \lambda * Q[v][f])$ 
10       end for
11     end for
12   end for
13    $\alpha$  衰减
14 end for
输出:  $P, Q$  矩阵

```

2.2 好友因素的影响

一般来说,在社交网络中的用户与和他有关联的好友有着相似的兴趣,并且直接或间接地受到好友偏好的影响,在社会学中被称为趋同性和社会影响力。通过对 Gowalla 数据集的分析也发现了这一关联。本文认为基于社交网络的用户好友关系极大地影响了用户基于兴趣点的签到行为,可利用用户的社会网络关系来提升兴趣点推荐算法的性能。

通过分析发现,用户好友之间的影响因素主要集中在两个方面:a)两者之间共同好友的数量;b)用户与好友共同访问过的地点个数。其影响因子计算方法如式(7)所示。

$$S_{u_1 u_2} = \eta \frac{|F_{u_1} \cap F_{u_2}|}{|F_{u_1} \cup F_{u_2}|} + (1 - \eta) \frac{|V_{u_1} \cap V_{u_2}|}{|V_{u_1} \cup V_{u_2}|} \quad (7)$$

其中: $S_{u_1 u_2}$ 表示用户 u_1 对用户 u_2 的影响因子; η 是一个可调节参数,取值为(0,1]; F_{u_i} 表示用户 u_i 的好友集合; V_{u_i} 表示用户 u_i 签到地点集合。本文综合共同好友数量和共同访问的地点数来计算用户之间的好友影响因子,然后选取 top- N 个分数最高的用户组成用户好友子集 U^* 。对于候选地点列表 V^n 计算用户 u_i 对于地点 v_j 的感兴趣程度,如式(8)所示。

$$I(u_i, v_j) = \frac{\sum_{(u_k \in U^*)} S_{u_i u_k} W_{u_k v_j}}{\sum_{(u_k \in U^*)} S_{u_i u_k}} \quad (8)$$

其中: $W_{u_k v_j}$ 表示用户 u_k 在地点 v_j 的签到记录, $W_{u_k v_j} = 0$ 为没有签到记录,否则 $W_{u_k v_j} = 1$ 。其算法描述如算法2所示。

算法2 好友因素分析算法 (friendship factor analysis algorithm)

输入:用户—好友关系矩阵 P ,用户—地点签到矩阵 C ,参数 η 。

```

1  初始化好友影响因子矩阵  $S$ 
2  for user  $u_1$  in  $P$  do
3    for user  $u_2$  in  $u_1$ . friend_list do
4      计算  $u_1$  与  $u_2$  之间共同好友的数量记为  $n$ 
5      计算  $u_1$  和  $u_2$  共同访问的地点数目记为  $m$ 
6      计算  $u_1$  和  $u_2$  所有好友的数量之和记为  $S_1$ 
7      计算  $u_1$  和  $u_2$  所有签到地点数目之和记为  $S_2$ 
8       $S[u_1][u_2] = \eta \frac{n}{S_1} + (1 - \eta) \frac{m}{S_2}$ 
9    end for
10 end for
11 初始化用户—地点兴趣度矩阵  $I$ 
12 for user  $u_1$  in  $S$  do
13   for local  $v$  in  $C$  do
14     for  $u_2$  in  $u_1$ . friend_list do
15        $I[u_1][v] + = \frac{S[u_1][u_2] * C[u_2][v]}{S[u_1][u_2]}$ 
//如果  $u_2$  在地点  $v$  签到过  $C[u_2][v] = 1$ , 否则为 0
16     end for
17   end for

```

```

18 end for
19 最后根据结果的正确率调整参数  $\eta$  的值
输出:矩阵  $S$ 。

```

2.3 地理位置因素的影响

在位置社交网络中,地点与用户之间的距离会对用户的选择产生重要影响。因此,在推荐中除了要分析用户自身的偏好,还要考虑地理位置自身属性的问题。为了更好地分析兴趣点自身的属性对用户的影响,本文对 Gowalla 数据集中的用户签到行为进行了分析,发现用户倾向于在已签到地点的附近进行签到,这种签到行为具有地点类簇的现象,换句话说就是用户的已签到地点都集中于一个或者几个区域。

对于候选地点集合 V^n 中的元素来说,所在区域的用户签到地点数目越多,用户在此地点签到的概率就越大。本文定义一个直径大小为 D_q 的区域(经过测试发现 D_q 为 1 KM 时最为合理)。文献[23]提出两个 GPS 坐标之间的距离采用大圆距离的计算方式,也就是地球的球面距离。设两个地点 l_1 和 l_2 的经纬坐标分别为 $(\text{lan}_1, \text{lon}_1)$ 、 $(\text{lan}_2, \text{lon}_2)$,则两个地点之间距离为

$$d(l_1, l_2) = 2R \times \arcsin \sqrt{L} \quad (9)$$

$$L = \sin^2 \left(\frac{\text{lan}_1 - \text{lan}_2}{2} \right) + \cos(\text{lan}_1) \cos(\text{lan}_2) \sin^2 \left(\frac{\text{lon}_1 - \text{lon}_2}{2} \right) \quad (10)$$

其中: R 为地球的半径。

所以兴趣点 v 自身的属性对用户 u 的影响可表示为

$$\text{sum}(u, v) = \sum_{l \in V^l} W_{ul} \quad (11)$$

$$W_{ul} = \begin{cases} 0 & d_{ul} - D_q > 0 \\ 1 & d_{ul} - D_q \leq 0 \end{cases} \quad (12)$$

将上面公式得出的结果进行归一化处理,就得到了兴趣点 v 自身的属性对用户 u 的影响因素。

$$l(u, v) = \frac{\text{sum}(u, v) - \min(\text{sum})}{\max(\text{sum}) - \min(\text{sum})} \quad (13)$$

其具体描述如算法3所示。

算法3 地理位置因素分析算法 (geographic factor analysis algorithm)。

输入:用户—地点签到矩阵 C ,地点—坐标矩阵 G 。

```

1  初始化用户—地点兴趣度矩阵  $I$ 
2  for user  $u$  in  $C$  do
3    for local  $v_1$  in  $u$ .not_check_in_list do
4      for local  $v_2$  in  $u$ .check_in_list do
5        if  $d(v_1, v_2) > D_q$ 
6           $I[u][v_1] + = 1$ 
7        else continue
8      end for
9    end for
10 end for
11 对矩阵  $I$  中数据进行归一化处理
输出:用户—地点兴趣度矩阵  $I$ 。

```

2.4 融合过程

上文分别从用户历史偏好、好友影响、地理位置等因素分析了用户在某个地点签到的概率,本节将通过线性融合框架来集成这三种算法。

设 H_w 、 I_w 和 L_w 分别为用户 u 在地点 v 签到的用户历史偏好、好友影响和地理位置因素影响的可能性, R_w 表示用户 u 在地点 v 签到的可能性,融合框架如式(14)所示。

$$R_w = (1 - \alpha - \beta) H_w + \alpha I_w + \beta L_w \quad (14)$$

其中: α 和 β 是两个权重参数($0 \leq \alpha + \beta \leq 1$),这两个参数的值说明了用户偏好、好友影响和地理位置这三个因素对于用户签到行为的影响程度。 $\alpha = 1$ 表明 R_w 完全依赖于好友影响, $\beta = 1$ 表明 R_w 完全依赖于地理位置影响, $\alpha = \beta = 0$ 则表示 R_w 完全依赖于用户的历史偏好。

算法4 多因素融合模型 (multi-factor fusion model)。

输入:算法1得到的 P, Q 矩阵,算法2得到的好友影响因子矩阵

S, 算法3得到的用户—地点兴趣度矩阵 I , 参数 α, β, F 。

```

1  for each user  $u$  do
2    for each local  $v$  do
3      for  $f$  in  $(0, F)$  do
4         $H_{uv} + = P_{uf} Q_{vf}$  //历史偏好影响
5      end for
6       $I_{uv} = S_{uv}$  //好友影响
7       $L_{uv} = I_{uv}$  //地理位置影响
8       $R_{uv} = (1 - \alpha - \beta) H_{uv} + \alpha I_{uv} + \beta L_{uv}$ 
9    end for
10  end for
11 选取 Top- $N$  生成推荐列表, 然后根据推荐准确度调整参数  $\alpha, \beta$ 
    输出: 推荐列表。

```

设有 M 个用户和 N 个地点, 基于用户的协同过滤算法来说, 要存储用户—地点签到矩阵需要 $O(M \times N)$ 的空间, 而 LFM 需要的存储空间是 $O(F \times (M + N))$ 。好友影响计算的空间复杂度与好友关系数目和每个用户的好友数量相关。若有 M 个用户而每个用户的平均好友数量为 n , 则空间复杂度为 $O(M \times N)$ 。

3 实验结果

3.1 实验数据及环境

本文的实验代码都是通过 Java 实现。实验机器为一台 Pentium CPU 1.8 GHz, 2 GB 内存的戴尔台式计算机, 编译环境为 Eclipse。

本文在实验中采用了 Gowalla^[24] 和 Foursquare 的真实数据集。以 Gowalla 为例说明, 该数据集有两部分, 一部分是用户签到数据, 另一部分是用户社交关系数据。签到数据和社交关系数据如表 1 所示, 签到数据集的数据格式如表 2 所示。

表1 数据统计

项目 (event)	统计 (statistics)
用户数目 (nodes)	196 591
地点数目 (edges)	950 327
签到数据 (check-ins)	6 442 890
朋友关系 (friendship)	1 900 654
用户平均签到次数 (average check_in)	60.162 2
地点平均被签到次数 (location check by user)	5.029 7

表2 数据格式

用户	签到时间	经度	纬度	地点
196 514	2010-07-24T13:45:06Z	53.364 811 9	-2.272 346 583	145 064

由表 1 可看到用户的平均签到次数为 60.162 2 次, 而地点的平均被签到次数为 5.029 7 次, 相对于庞大的地点数目来说签到数据是相当稀疏的, 其中大部分地点可能仅被一两个用户签到过。稀疏数据会对结果产生较大的偏差, 需要对其进行预处理, 仅保留每天活跃次数超过 5 次的活跃用户。而且其中有很多用户对同一地点只有过一次签到行为, 这种重复率较低不能明显表现出用户喜好的数据会对本文的计算结果产生负面影响, 所以本文把用户只有一次签到行为的记录也进行清理。本文将一年签到数据的前十个月作为训练样本, 后两个月的数据作为测试数据。

3.2 评价指标

本文采用准确率 (precision)、召回率 (recall)、平均准确率 MAP (mean average precision)^[25] 作为度量指标。设 $R(u)$ 为算法给出的推荐列表, $T(u)$ 为用户在测试集上的行为列表, 那么推荐结果的召回率为

$$\text{recall} = \frac{1}{|U|} \sum_{u \in U} \frac{|\sum_{u \in U} |R(u) \cap T(u)|}{|\sum_{u \in U} |T(u)|} \quad (15)$$

推荐的准确率计算方式为

$$\text{precision} = \frac{1}{|U|} \sum_{u \in U} \frac{|\sum_{u \in U} |R(u) \cap T(u)|}{|\sum_{u \in U} |R(u)|} \quad (16)$$

MAP 是一种有序序列的评价指标, 一般在信息检索中用来评测搜索引擎的效果, MAP 的值越高, 表明越相关的地点排名越靠前, 推荐的结果越合理, 在本文中用于检测是否预测得分越高准确性也越高。MAP 的计算方式如下:

$$\text{MAP} = \frac{1}{|U|} \sum_{u \in U} \frac{1}{k} \sum_{i=1}^k \text{precision}_u(k) \quad (17)$$

其中: U 是用户数量; u 表示一个具体的用户; k 是推荐列表数目, 在实验中 k 分别取 5、10、20, 其中 MAP 的计算中 k 的取值为 1~20。

3.3 实验结果与分析

本文选择了三种基于不同算法的且具有代表性的模型进行比较。

a) PR^[26]。该模型给出了本地兴趣点的推荐方法, 把地点推荐问题转换为概率问题。PR 模型考虑了用户类别偏好、时间、距离和地点的流行度。PR 模型在本地推荐中提供了很好的效果。

b) USG^[17]。一种改进的自适应地点推荐模型。该模型融合用户协同过滤、社交影响和朴素贝叶斯分类, 并适应用户的当前位置, 力求为用户推荐最适合用户当前位置的地点。USG 模型有效地解决了推荐过程中数据稀疏性的问题, 在个性化的地点推荐中取得了不错的性能。

c) NCPD^[19]。基于 NMF 矩阵分解模型融合用户的兴趣类别信息和地理位置信息进行兴趣点推荐, 采用用户的邻居影响因素对地理位置信息进行建模。NCPD 模型创新性地考虑了用户的邻居影响因素, 得出了很多创新的结论, 取得了很好的推荐准确率。

k 值为推荐地点列表数目, 根据实际情况为用户产生的推荐列表一般不会超过 20 个, 所以本文对 k 分别取 5、10、20 来进行测试。上文中 λ (算法 1) 的值取 0.007, α (算法 1) 的值取 0.03。

图 2~4 为模型在 Gowalla 数据集上的结果, 图 5~7 为模型在 Foursquare 数据集上得到的结果。

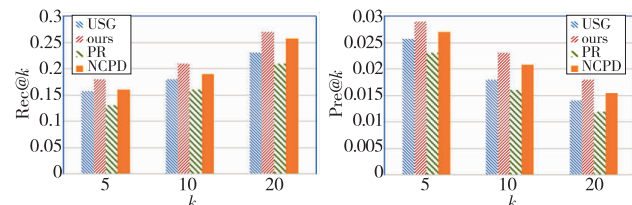


图2 recall (Gowalla 数据集)

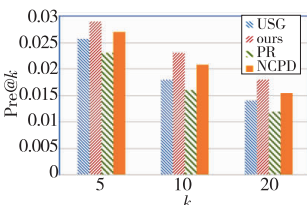


图3 precision (Gowalla 数据集)

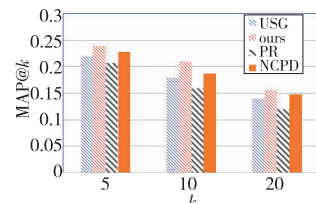


图4 MAP (Gowalla 数据集)

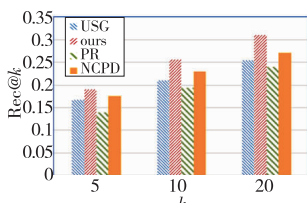


图5 recall (Foursquare 数据集)

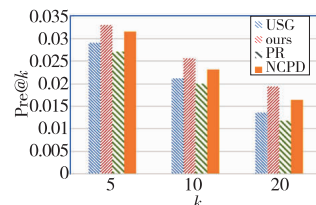


图6 precision (Foursquare 数据集)

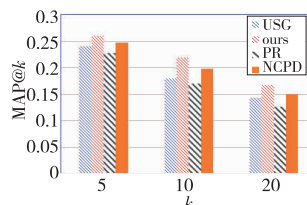


图7 MAP (Foursquare 数据集)

从图中可以看出,四种模型之间的性能具有明显的差距,本文提出的融合模型比其他三种模型具有更好的性能,其中USG和PR这两种使用贝叶斯来进行预测的模型之间性能的差别比较小,而使用NMF算法的NCPD模型效果明显要强过一筹;但NCPD模型由于没有考虑用户的社交关系对推荐结果的影响,所以比本文提出的模型效果要稍差,而USG和PR模型使用传统的聚类算法没有很好地解决地点的语义相似性问题,所以效果最差。当推荐的数目增多时,本文的融合模型相对于其他三种模型的优势更加明显,这个结果证明从历史偏好、社交关系、地理位置这三个方面对用户的行为进行分析建模是有效的。从图中可以看出算法的准确率是随着 k 值的增大而减小的,这说明算法对于用户兴趣的广度挖掘得还不够充分。而在Foursquare数据集上的实验,由于Foursquare数据集用户签到的稠密度要高于Gowalla数据集,所以各算法在Foursquare数据集上的表现要比在Gowalla数据集上好一些,但总体趋势来说各模型在两个数据集上的性能差别大致相同。

4 结束语

在位置社交网络推荐系统的发展中,如何将多元的情境信息与用户的位置信息联系起来一直是研究的一个重点问题。本文提出了一种融合模型,该模型将用户的历史偏好、社交关系和地点的地理位置信息融合在一起进行有效的推荐,并通过负样本采样的方式补充了用户的评分数据,较好地解决了在充分挖掘用户的历史偏好的同时结合其他影响因素进行兴趣点推荐的问题。真实的数据集表明本文的融合模型相对于其他的主流模型具有更好的准确率、召回率和MAP值。

本文为基于位置的社交网络推荐与传统的推荐系统之间的结合作出了一步大胆的常识,传统的推荐系统与基于位置的社交网络推荐之间的最大区别就在于有没有完整的评分细则,基于位置的推荐服务为了方便用户没有设置评分系统,本文的负样本采样策略在某种程度上弥补了这一空缺。在以后的工作中可以进一步细化样本的层次,争取可以按照当前流行的五分评价系统对用户签到数据进行评分,这样一些在传统推荐系统中已经成熟的模型和方法就可以直接运用到基于位置的社交网络推荐服务中,进一步打破传统推荐系统与基于位置的社交网络推荐系统之间的壁垒。

在以后的工作中还可以将时间因素与影响用户签到行为的其他因素一起融合到这个模型中去,进一步提高兴趣点推荐的用户满意度。

参考文献:

- [1] Zheng Yu. Location-based social networks: users[M]//Computing with Spatial Trajectories. New York: Springer, 2011: 243-276.
- [2] 翟红生,于海鹏. 在线社交网络中的位置服务研究进展与趋势[J]. 计算机应用研究, 2013, 30(11): 3221-3227.
- [3] Zhou Dequan, Wang Bin, Rahimi S M, et al. A study of recommending locations on location-based social network by collaborative filtering[C]//Proc of the 25th Canadian Conference on Artificial Intelligence. Berlin: Springer-Verlag, 2012: 255-266.
- [4] Yu Zhiwen, Tian Miao, Wang Zhu, et al. Shop-type recommendation leveraging the data from social media and location-based services[J]. ACM Trans on Knowledge Discovery from Data, 2016, 11(1): 1-9.
- [5] Zhao Guoshuai, Qian Xueming, Kang Chen. Service rating prediction by exploring social mobile users' geographical locations[J]. IEEE Trans on Big Data, 2017, 3(1): 67-78.
- [6] He Tieke, Yin Hongzhi, Chen Zhenyu, et al. A spatial-temporal topic model for the semantic annotation of POIs in LBSNs[J]. ACM Trans on Intelligent Systems and Technology, 2016, 8(1): 12-19.
- [7] Yuan Quan, Cong Gao, Ma Zongyang, et al. Time-aware point-of-interest recommendation[C]//Proc of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM Press, 2013: 363-372.
- [8] Yuan Quan, Cong Gao, Sun Aixun. Graph-based point-of-interest recommendation with geographical and temporal influences[C]//Proc of the 23rd ACM International Conference on Conference on Information and Knowledge Management. New York: ACM Press, 2014: 659-668.
- [9] Nunes I, Marinho L. A personalized geographic-based diffusion model for location recommendations in LBSN[C]//Proc of the 9th Latin American Web Congress. Washington DC: IEEE Computer Society, 2014: 59-67.
- [10] Pan Guo, Xu Yuming. Friends prediction based on fusion of topology and location in LBSN[J]. Computer Science, 2014, 9: 24.
- [11] Bagci H, Karagoz P. Context-aware friend recommendation for location based social networks using random walk[C]//Proc of the 25th International Conference Companion on World Wide Web. 2016: 531-536.
- [12] Su Chengcheng, Yu Yaxin, Sui Mingfei, et al. Friend recommendation algorithm based on user activity and social trust in LBSNs[C]//Proc of the 12th Web Information System and Application Conference. Piscataway, NJ: IEEE Press, 2015: 15-20.
- [13] Bagci H, Karagoz P. Random walk based context-aware activity recommendation for location based social networks[C]//Proc of IEEE International Conference on Data Science and Advanced Analytics. Piscataway, NJ: IEEE Press, 2015: 1-9.
- [14] Xu Huang, Yu Zhiwen, Feng Yun, et al. LBSN-based personalized travel package recommendation system[J]. Computer & Modernization, 2014(1): 186-191.
- [15] Herlocker J L, Konstan J A, Borchers A, et al. An algorithmic framework for performing collaborative filtering[C]//Proc of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM Press, 1999: 230-237.
- [16] Ye Mao, Yin Peifeng, Lee W C, et al. Exploiting geographical influence for collaborative point-of-interest recommendation[C]//Proc of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM Press, 2011: 325-334.
- [17] Wang Jingjin, Lin Kunhui, Li Jia. A collaborative filtering recommendation algorithm based on user clustering and slope one scheme[C]//Proc of the 8th International Conference on Computer Science & Education. Piscataway, NJ: IEEE Press, 2013: 1473-1476.
- [18] Cheng Chen, Yang Haiqin, King I, et al. Fused matrix factorization with geographical and social influence in location-based social networks[C]//Proc of the 26th AAAI Conference on Artificial Intelligence. 2012: 17-23.
- [19] Hu Longke, Sun Aixun, Liu Yong. Your neighbors affect your ratings: on geographical neighborhood influence to rating prediction[C]//Proc of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval. 2014: 345-354.
- [20] 曹玖新,董昇,杨鹏伟,等. LBSN中基于元路径的兴趣点推荐[J]. 计算机学报, 2016, 39(4): 675-684.
- [21] Xu Guandong, Fu Bin, Gu Yanhui. Point-of-interest recommendations via a supervised random walk algorithm[J]. IEEE Intelligent Systems, 2016, 31(1): 15-23.
- [22] Pan Rong, Zhou Yunhong, Cao Bin, et al. One-class collaborative filtering[C]//Proc of the 8th IEEE International Conference on Data Mining. Washington DC: IEEE Computer Society, 2008: 502-511.
- [23] Distance calculation algorithms[EB/OL]. (2013-05-08). <http://www.ga.gov.au/earth-monitoring/geodesy/geodetic-techniques/distance-calculation-algorithms>. html.
- [24] <http://snap.stanford.edu/data/loc-gowalla>. html[EB/OL].
- [25] 姜超. 基于位置的社交网络中高效的地点推荐方法研究[D]. 杭州: 浙江大学, 2012.
- [26] 张森. 基于位置社交网络的兴趣点推荐方法研究[D]. 重庆: 西南大学, 2015.