

融合多因素的兴趣点协同推荐方法研究

陈 炯¹ 张 虎^{2,3} 曹付元^{2,3}

(山西职业技术学院计算机工程系 太原 030006)¹ (山西大学计算机与信息技术学院 太原 030006)²
(山西大学计算智能与中文信息处理教育部重点实验室 太原 030006)³

摘 要 兴趣点(Point-of-Interest, POI)推荐是为用户推荐可能感兴趣的地理位置的一项任务,是基于位置社交网络(Location-Based Social Networks, LBSN)服务中的重要研究内容。针对目前 POI 推荐准确率较低、推荐结果缺乏个性化、情感倾向因素融入差等问题,在综合分析兴趣点的地理位置、分类偏好、流行度、社交与情感倾向等相关影响因素的基础上,提出了融合多因素的兴趣点协同推荐模型(GCSR)。首先,根据 POI 地理位置数据计算地理相关分数;其次,根据用户的类别偏好,结合 POI 流行度定义分类偏好分数;然后,根据社交关系计算用户之间的社交关系强度,通过挖掘评论文本计算用户的情感倾向分数,并将二者与协同过滤推荐技术有效结合,从而得到社交情感分数;最后,将地理相关分数、分类偏好分数与社交情感分数有效融合,向用户推荐 Top-N 兴趣点。在 Foursquare 真实签到数据集上进行的多组对比实验显示,与基线模型中最好的 JRA 相比,GCSR 模型能够获得更好的推荐效果,准确率和召回率平均提高了 1.7% 和 0.6%。

关键词 基于位置的社交网络,兴趣点推荐,情感倾向,地理位置,社交关系

中图分类号 TP391 **文献标识码** A **DOI** 10.11896/jsjcx.180901757

Study on Point-of-interest Collaborative Recommendation Method Fusing Multi-factors

CHEN Jiong¹ ZHANG Hu^{2,3} CAO Fu-yuan^{2,3}

(Department of Computer Engineering, Shanxi Polytechnic College, Taiyuan 030006, China)¹

(School of Computer and Information Technology, Shanxi University, Taiyuan 030006, China)²

(Key Laboratory of Ministry of Education for Computation Intelligence and Chinese Information Processing (Shanxi University), Taiyuan 030006, China)³

Abstract Point-of-interest (POI) recommendation is a task to recommend geographical locations that users may be interested in. It is an important researches in location-based social networks (LBSN) services. For the existing problems that POI recommendation currently has lower recommendation precision, lacks of personalization in recommendation results, and has poor integration of sentimental orientation factors, etc., this paper proposed a POI collaborative recommendation model(GCSR) fusing multi-factors based on the comprehensive analysis of POI related influencing factors, such as geographical location, category preference, popularity, social and sentimental orientation and so on. Firstly, the geographical relevance score is calculated based on POI geographical location data. Secondly, category preference score is defined according to users' category preference and POI popularity. Then, the strength of the social relationships between users is calculated based on the social relationships, the sentimental orientation score of users is calculated by mining the comment text, and the two are effectively combined with the collaborative filtering recommendation technology to obtain the social sentiment score. Finally, geographical relevance score, category preference score and social sentiment score are effectively integrated to recommend Top-N POI. Multiple comparative experiments conducted on Foursquare's real check-in datasets demonstrate that the GCSR model achieves better recommendation effect, with an average improvement of 1.7% and 0.6% in precision and recall, compared with the best effective JRA in the baseline models.

Keywords Location-based social networks, Point-of-interest recommendation, Sentimental orientation, Geographical location, Social relationships

1 引言

随着社交网络、智能终端和自动定位技术的快速发展,基于位置的社交网络(LBSN)服务已成为一项具有重要价值的应用,Foursquare、Facebook Places、开开和街旁等都是提供类似服务为主的产品。兴趣点(POI)推荐是基于位置的社交网络服务研究中的一项重要内容,旨在为用户推荐新的感兴趣的位置,它不仅可以一定程度地解决大数据时代用户面临的信息过载问题,提升用户体验,而且能够帮助商家快速找到目标用户,实现精准营销。LBSN中的POI推荐已成为当前的研究热点^[1]。

近年来,国内外研究人员围绕POI推荐方法进行了大量的研究。大多数工作都是基于用户的签到历史和社交网络数据建模用户对兴趣点的个性化偏好,借助协同过滤等推荐技术进行兴趣点的推荐。丁勇等^[2]提出了一种融合相似性和好友信任的地点推荐方法,针对用户的非好友集和好友集分别设计了基于相似性和好友信任的地点推荐算法,推荐效果相比传统基于用户的协同过滤算法有明显提高。Ying等^[3]提出了一种称为UPOI-Walk的方法,该方法整合了用户的社交关系、用户偏好与流行度因素,使用基于HITS的随机游走模型进行兴趣点推荐,取得了较好的推荐效果。Ye等^[4]提出了一个统一的兴趣点推荐框架,将用户对POI的偏好、社交影响与地理影响相融合,基于协同过滤推荐技术进行兴趣点的推荐。Rahimi等^[5]提出了行为分解(BF)和潜在行为分析(LBA)两种位置推荐方法,利用用户移动行为和时空模式来进行位置推荐,取得了较好的效果。韩笑峰等^[6]提出了一种个性化兴趣点推荐的联合模型,综合考虑了POI的位置、类别、流行度、社交因素和用户行为,结合协同过滤技术生成推荐列表。该方法综合考虑了多个因素,推荐质量有了进一步的提升。然而,签到历史记录中积累的位置、类别、流行度、社交、评论等多源异构信息从不同角度描述了用户活动,隐藏着人们各式各样的生活模式和个人偏好,以上方法没有全面考虑这些因素的共同作用,特别是没有探索签到POI的评论文本对推荐结果的影响。

目前,一些研究已尝试将签到的POI评论文本引入兴趣点推荐。王啸岩等^[7]提出了一种融合用户社交网络数据、地理位置数据以及兴趣点评论文本这3个因素的POI推荐模型,分别使用协同过滤、幂律分布与潜在狄利克雷分配预测社交、位置与兴趣的概率,最后通过组合3个概率进行兴趣点的推荐。任星怡等^[8]提出了一种上下文感知的概率矩阵分解兴趣点推荐算法,将兴趣点的地理、文本、社会、分类与流行度信息进行有效融合,利用潜在狄利克雷分配模型挖掘兴趣点的文本信息,从而学习用户的兴趣话题。上述两种方法均使用了兴趣点的评论文本,但只是使用潜在狄利克雷分配模型学习用户的兴趣话题,与POI话题进行匹配,并没有挖掘评论文本的情感倾向。事实上,用户评论提供了明确的偏好信息,这是对签到行为的一种补充解释,通过分析评论文本的情感倾向,可以更加准确地了解用户的个性化偏好,并改善POI推荐的效果^[1]。

针对这些研究的不足,本文探索了一种改进的POI协同

推荐方法。该方法综合考虑了兴趣点的地理位置、分类偏好、流行度、社交与情感倾向因素,将不同因素的分数表示与协同过滤推荐框架相结合,并在此基础上提出了一种融合多因素的兴趣点协同推荐模型——GCSR模型。

2 融合多因素的兴趣点协同推荐模型

2.1 问题描述

LBSN上积聚着大规模的历史签到数据,签到数据中包含的多源异构信息为分析用户偏好、进一步推荐用户喜欢的POI提供了丰富的信息。LBSN包含用户和POI两类实体。用户之间存在社交关系,随着时间的推移,签到活动能够动态构建或扩充3种关系,即用户与位置之间的签到关系、用户与用户之间的社交关系、位置与位置之间的相关关系,同时生成照片、视频与评论等相关内容信息。例如,两个用户之间的社交关系因其访问了相同的POI而得到建立或加强;同样,两个POI之间的相关关系因其被相同的用户访问而得到强化或拓展。上述实体、关系、内容与时间4个层面的信息相互关联,共同组成一个多源异构的信息网络,如图1所示。本文提出的GCSR推荐模型,通过分析历史签到数据中隐含的模式和相关关系,可以很好地建模用户的签到行为,为用户推荐一个包含TOP-N未访问过的位置列表。

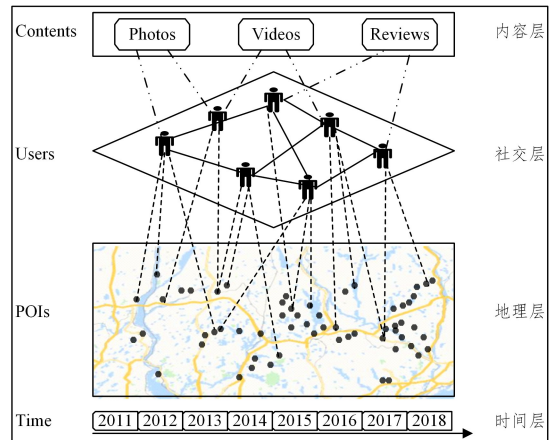


图1 基于位置的社交网络多源异构信息示意图

Fig.1 Sketch map of multi-side heterogeneous information in LBSN

签到历史数据中包含了签到时间信息、POI的位置信息、用户的社交信息、POI的类别信息与评论信息等,可以从签到数据中提取这些信息用于POI推荐。为方便说明,表1列出了本文所用符号及其含义。

表1 本文所用符号及其含义

Table 1 Symbol meanings used in this article

| 符号 | 含义 |
|-------|--|
| U | LBSN中所有用户组成的集合, $U=\{u_1, u_2, \dots, u_n\}$ |
| u_i | 用户集合 U 中的一个用户, $u_i \in U$ |
| F_i | LBSN中用户的好友组成的集合, $F_i \subseteq U$ |
| L | LBSN中所有POI组成的集合, $L=\{l_1, l_2, \dots, l_m\}$ |
| c_i | POI集合 L 中的一个POI, $c_i \in L$ |
| C | LBSN中所有POI类别组成的集合 $C=\{c_1, c_2, \dots, c_k\}$ |
| c_i | POI类别集合 C 中的一个POI类别, $c_i \in C$ |
| T | LBSN中所有签到记录组成的集合, $T=\{t_1, t_2, \dots, t_j\}$ |
| t_i | 签到记录集合 T 中的一次签到, $t_i \in T$ |
| R | LBSN中所有POI的评论组成的集合, $R=\{r_1, r_2, \dots, r_s\}$ |
| r_i | 评论集合 R 中的一个评论, $r_i \in R$ |

2.2 地理位置因素

用户在地理上受限的区域中活动,并且更喜欢访问附近的 POI,这种现象被称为空间聚集现象^[4]。Tobler 地理学第一定律指出“任何事物都相关,相近的事物关联更紧密”。在现实生活中,人们不会经常访问远离其活动区域的 POI,POI 的地理邻近性对用户的签到行为有重大影响。

由于人们访问较近的位置往往比较远的位置更加频繁^[9-10],本文通过计算候选 POI 与用户已访问 POI 之间的距离来预测推荐的可能性。 $L_i = \{l_1, l_2, \dots, l_n\}$ 表示用户 u_i 已访问过的 POI 集合, L 为所有 POI 的集合, $L_i \subseteq L$ 。用户 u_i 访问 POI $l_j \in L - L_i$ 的地理相关分数 $S_{i,j}$ 可以通过式(1)计算:

$$S_{i,j} = \frac{1}{\text{MIN}_{l_i \in L_i} \text{dist}(l_j, l_i)} \quad (1)$$

其中, $\text{dist}(l_j, l_i)$ 是 POI l_j 和 l_i 之间的地理距离。再将 $S_{i,j}$ 按照式(2)规范化为 0 到 1 之间的值,式(2)中分母代表用户 u_i 的所有候选 POI 中 $S_{i,j}$ 的最大值。通过式(2)能够得到用户 u_i 访问的候选 POI l_j 的地理相关分数 $S_{i,j}^g$:

$$S_{i,j}^g = \frac{S_{i,j}}{\text{MAX}_{l_k \in L - L_i} \{S_{i,k}\}} \quad (2)$$

2.3 分类流行度因素

LBSN 中的 POI 分为不同的类别,类别信息反映了 POI 提供的服务功能及服务风格。用户访问了某个 POI 代表用户对该类别 POI 的喜好,如果用户喜欢某个类别,他可能会访问属于该类别的更多 POI。然而,如果用户访问的有些类别的 POI 同时被很多用户频繁访问,那么这样的 POI 则不能代表该用户有别于其他用户的个性化偏好。因此,本文认为一个类别的 POI 被用户 u 访问的次数越多,同时被其他用户访问的次数越少,则该 POI 类别越能够代表用户 u 的个性化偏好。据此,可以从签到记录中分析用户访问过的 POI 类别信息来挖掘用户的个性化偏好。

Foursquare 签到数据集中的 POI 包括 8 大类,共计 240 个子类^[11],分别为 Arts & Entertainment(17),College & University(23),Food(78),Great Outdoors(28),Home & Work & Other(15),Nightlife Spot(20),Shop(45),Travel Spot(14),括号内的数字是各大类包括的子类数。

根据用户的签到历史记录,计算用户 u 在 c_i 类别的 POI 上的签到总数 $Totalch(u, c_i) = |\{t_k | t_k \in T \cap t_k. u = u \cap t_k. l. c = c_i\}|$,以及访问 c_i 类别 POI 的用户总数 $Totalpe(c_i) = |\{t_k. u | t_k. u \in U \cap t_k. l. c = c_i\}|$,并按照式(3)计算用户 u 访问 c_i 类 POI 的偏好分数 $cp(u, c_i)$,据此构建用户 u 的类别偏好向量 $CP(u) = \langle cp(u, c_1), cp(u, c_2), \dots, cp(u, c_8) \rangle$ 。

$$cp(u, c_i) = \frac{Totalch(u, c_i)}{\sum_{c_j \in C} Totalch(u, c_j)} * \lg \frac{|U|}{Totalpe(c_i) + 1} \quad (3)$$

其中, $|U|$ 为用户总数, C 为 POI 类别集合。为防止分母为零,第二项的分数表达式中分母加 1。

类别偏好向量 $CP(u)$ 代表了用户 u 对不同类别 POI 的偏好程度。然而, $CP(u)$ 只描述了用户 u 偏好的 POI 类别,每一类又包括多个 POI,很显然,在向用户推荐其偏好的 POI 时,

将这类所有 POI 同等重要地推荐给该用户是不合理的。事实上,同类别的各个 POI 因服务质量和用户体验的差异,对用户的吸引力是不同的,为捕捉同类别中不同 POI 的热门程度,引入 POI 流行度测度。POI 的流行度反映了 POI 的受欢迎程度,受欢迎程度可以在很大程度上影响用户的签到行为,同类别的 POI 流行度越高,则用户体验越好,推荐的优先级也应该越高。通常,流行度高的 POI 接待次数较多,而且访客人数也较多,因此使用 POI 的签到数与访客数来度量 POI 的流行度具有很大的合理性。

根据 POI 签到历史记录,计算 c 类 POI l_j 的签到总数 $Totalch(c, l_j)$ 和访客总数 $Totalpe(c, l_j)$: $Totalch(c, l_j) = |\{t_k | t_k \in T \cap t_k. l = l_j \cap t_k. l. c = c\}|$, $Totalpe(c, l_j) = |\{t_k. u | t_k. u \in U \cap t_k. l = l_j \cap t_k. l. c = c\}|$ 。对其标准化后,按照式(4)计算二者的算术平均值,得到类别 c 的 POI l_j 的流行度 $pop(c, l_j)$ 。

$$pop(c, l_j) = \frac{1}{2} \left\{ \frac{Totalch(c, l_j) - 1}{\text{MAX}_{l_k \in L} \{Totalch(c, l_k)\} - 1} + \frac{Totalpe(c, l_j) - 1}{\text{MAX}_{l_k \in L} \{Totalpe(c, l_k)\} - 1} \right\} \quad (4)$$

据此构建类别-兴趣点流行度矩阵 $CPP = [pop(c_i, l_j)]_{8 \times m}$, m 为 POI 的个数。

综合用户的类别偏好及 POI 流行度因素,可以计算出用户 u 对候选 POI 的偏好分数向量 $PS(u) = CP(u) * CPP$,向量 $PS(u)$ 中的各分量代表了用户 u 对各候选 POI 的偏好分数。用户 u_i 对候选 POI l_j 的分类偏好分数 $S_{i,j}^l$ 可通过式(5)计算获得。

$$S_{i,j}^l = \sum_{k=1}^8 cp(u_i, c_k) * pop(c_k, l_j) \quad (5)$$

2.4 情感倾向因素

用户访问 POI 时通常会留下体验评论,这些评论包含丰富的情感信息,反映了 POI 的服务质量和用户满意度,能够很好地代表用户的偏好。采用自然语言处理技术分析签到历史记录中的评论文本,进一步挖掘文本的情感倾向,能够准确了解用户的偏好,进一步提高 POI 的推荐质量^[7,12]。

文本由词组合而成,评论中情感词的极性代表了文本的倾向性,因此可以通过计算情感词的极性来获得评论的倾向性。情感极性计算通常假设:当一个词与积极词有强相关性时,它具有积极的情感极性,反之亦然。情感极性的计算有多种方法,比较简单有效的是基于点互信息(PMI)的方法^[13],通过计算目标情感词与情感词汇表中术语之间的 PMI,来确定情感词的极性。词 w_i 和 w_j 之间的 PMI 值定义如式(6)所示:

$$PMI(w_i, w_j) = \log_2 \frac{p(w_i, w_j)}{p(w_i) * p(w_j)} \quad (6)$$

其中, $p(w_i)$ 和 $p(w_j)$ 分别是 w_i 和 w_j 在语料库中的概率, $p(w_i, w_j)$ 是 w_i 和 w_j 共现的概率。而 $p(w_i) = \text{count}(w_i) / n$, $\text{count}(w_i)$ 为 w_i 在语料库中出现的次数, n 为语料库中包含的总词数。一个情感词的极性值 SO 是通过计算其与积极参

考词“excellent”和消极参考词“poor”的 PMI 值获得的,计算公式如式(7)所示:

$$SO(\text{phrase}) = \text{PMI}(\text{phrase}, \text{“excellent”}) - \text{PMI}(\text{phrase}, \text{“poor”}) \quad (7)$$

给定用户 u_i 对 POI l_j 的评论 $r_{i,j}$,按式(8)计算评论中所有情感词的 SO 分值的平均值,并将其作为评论 $r_{i,j}$ 的情感倾向分数 $so_{i,j}$ 。

$$so_{i,j} = \frac{1}{m} \sum_{\text{phrase}_j \in r_{i,j}} SO(\text{phrase}_j) \quad (8)$$

其中, m 为评论中包含的情感词的总数。

2.5 社交关系因素

LBSN 中包含了丰富的社交关系,朋友往往有类似的行为和很多共同的兴趣,用户在访问未知的 POI 之前通常会参考朋友的意见,朋友比普通用户更能够影响用户对 POI 的选择^[14]。研究表明,LBSN 中借助用户的社交关系,可以提高 POI 推荐的质量^[15]。

现实生活中,朋友之间关系的远近程度是不同的,有的好友仅仅是点头之交,而有的好友亲密无间,关系密切的朋友之间能够获得更大的信任;此外,朋友之间的兴趣也不是完全相同的,有的朋友志趣相投,而有的朋友各有所好,具有相同兴趣爱好朋友给出的推荐更有价值。对应到 LBSN 中,可以用朋友之间拥有的相同朋友的数量来描述朋友之间的远近程度,用朋友之间访问相同 POI 的数量来度量朋友之间兴趣的相似程度。因此,可以用式(9)计算用户及其朋友之间的社交关系强度。

$$sr_{i,k} = \frac{|F_i \cap F_j|}{|F_i \cup F_j|} * \frac{|L_i \cap L_j|}{|L_i \cup L_j|} \quad (9)$$

其中, F_i 与 F_j 分别代表用户 u_i 与用户 u_j 的朋友集合, L_i 与 L_j 分别代表用户 u_i 与用户 u_j 的签到 POI 集合。

2.6 GCSR 兴趣点推荐模型

基于用户的协同过滤推荐方法主要根据相似用户的兴趣向目标用户推荐其未知的项目,因其建立在用户关系的基础之上,而被广泛应用于 LBSN 的 POI 推荐服务中。

假设用户 $u_i \in U$ 在某 POI $l_j \in L$ 处的签到行为用 $t_{i,j}$ 表示,根据签到记录,若用户 u_i 在 POI l_j 处存在签到行为,则偏好估计 $t_{i,j} = 1$,否则 $t_{i,j} = 0$;可以根据式(10)预测用户 u_i 在未签到的 POI l_j 处的签到概率 $t_{i,j}^{\wedge}$ 。

$$t_{i,j}^{\wedge} = \frac{\sum_{u_k \in U} sim_{i,k} * t_{k,j}}{\sum_{u_k \in U} sim_{i,k}} \quad (10)$$

其中, $sim_{i,k}$ 表示用户 u_i 和用户 u_k 之间的相似度。 $sim_{i,k}$ 的计算方法有多种,余弦相似度方法是其中较为简单有效的方法,计算公式如式(11)所示:

$$sim_{i,k} = \frac{\sum_{l_s \in L} t_{i,s} * t_{k,s}}{\sqrt{\sum_{l_s \in L} t_{i,s}^2} * \sqrt{\sum_{l_s \in L} t_{k,s}^2}} \quad (11)$$

采用基于用户的协同过滤推荐方法进行 POI 推荐时,需要搜索 U 中的全部用户,并计算用户之间的相似度,计算开销较大,而且没有充分利用 LBSN 上的社交关系^[15]。此外,

签到活动仅代表用户对特定 POI 的一次访问,而看不出用户对 POI 的情感倾向。有时,用户对访问的 POI 给出了差评,而不是给予积极的评价。而且,用户对 POI 的偏好也分不同的程度,简单地认为用户 u_i 在 POI l_j 处存在签到就代表满意,进而令 $t_{i,j} = 1$,显然是不准确的。因此,本文对上述问题做出改进,将社交关系因素及情感倾向因素融入基于协同过滤的 POI 推荐中,使用朋友之间的社交关系强度 $sr_{i,k}$ 代替用户之间的相似度 $sim_{i,k}$,同时使用用户 u_i 在 POI l_j 处评论 $r_{i,j}$ 的情感倾向分数 $so_{i,j}$ 来调整签到偏好估计 $t_{i,j}$,这样不仅可以区分用户的情感倾向,还可以衡量情感的强弱。

现实生活中,并非所有的访客都对 POI 做出评论,有的用户在 POI 处签到后未做出评论,并不代表他对该 POI 不感兴趣。为方便统一处理,本文对签到 POI l_j 后未做评论的用户 u_i 设定情感倾向分数 $so_{i,j}$,其为该 POI 上所有用户评论情感分值的平均值,即 $so_{i,j} = \frac{1}{s} \sum_{r_{k,j} \in R} so_{k,j}$,其中, s 表示 POI l_j 上评论的总数。因此,本文将签到偏好估计调整为 $ts_{i,j} = so_{i,j} * t_{i,j}$ 。

基于以上分析,融入社交关系因素及情感倾向因素后用户 u_i 在未签到的 POI l_j 处的签到概率 $t_{i,j}^{\wedge}$ 可以根据式(12)计算。

$$t_{i,j}^{\wedge} = \frac{\sum_{u_k \in F_i} sr_{i,k} * ts_{k,j}}{\sum_{u_k \in F_i} sr_{i,k}} \quad (12)$$

其中, $sr_{i,k}$ 表示用户 $u_i \in U$ 及其朋友 $u_k \in F_i$ 之间的社交关系强度, $ts_{k,j}$ 表示用户 u_k 在 POI l_j 处的签到偏好分数。进一步,将计算得到的候选 POI 的签到概率 $t_{i,j}^{\wedge}$ 按式(13)标准化,得到用户 u_i 关于 POI l_j 的社交情感分数 $S_{i,j}^{\sigma}$ 。

$$S_{i,j}^{\sigma} = \frac{t_{i,j}^{\wedge}}{\text{MAX}_{u_i \in U \cap l_j \in L} t_{i,j}^{\wedge}} \quad (13)$$

LBSN 中的兴趣点推荐是一种基于异构信息和复杂关系的个性化推荐。本文综合利用 LBSN 的地理位置、分类偏好、流行度、社交与评论信息,来计算地理相关分数、分类偏好分数,并将情感倾向及社交关系融入基于朋友的协同过滤方法中,提出了融合多因素的 GCSR 兴趣点协同推荐模型。模型考虑了 5 个因素,融合了 3 个分数,通过分析用户的位置历史数据,计算地理相关性,得出用户对未签到的 POI 的地理相关分数 $S_{i,j}^g$;根据用户已签到 POI 的分类及未签到 POI 的流行度,计算分类相关性,获得用户对未签到 POI 的分类偏好分数 $S_{i,j}^c$;根据用户签到评论与签到历史数据,计算情感倾向分数与社交关系强度,并结合基于朋友的协同过滤方法推导出用户对未签到 POI 的社交情感分数 $S_{i,j}^{\sigma}$ 。

在 POI 推荐中,相比线性组合,使用乘法法则融合不同因素具有更高的鲁棒性^[8]。本文使用乘法法则融合不同分数,形成 GCSR 兴趣点协同推荐模型。最终,用户 u_i 对候选 POI l_j 的预测分数 $S_{i,j}^{\text{GCSR}}$ 可通过式(14)计算获得。

$$S_{i,j}^{\text{GCSR}} = S_{i,j}^g * S_{i,j}^c * S_{i,j}^{\sigma} \quad (14)$$

3 实验及分析

3.1 实验数据

Foursquare 是全球知名的基于地理位置的大规模社交服务网站,支持用户签到、发表评论,并与其他人分享当前所在地理位置及消费体验等信息。使用文献[16]中的数据集进行实验,数据收集于 2010 年 9 月至 2011 年 1 月,包括了来自 Foursquare 和 Twitter 等移动平台的 224 804 个用户和 22 388 315 条签到数据,72%的用户签到次数小于 100。本文按时间顺序从中选取一部分用户的 Foursquare 签到数据,并过滤掉签到次数少于 10 的用户和签到用户数不足 10 的 POI 数据,最终形成本文所需的实验数据集。实验数据集的统计信息如表 2 所列。实验中,随机选取数据集中 80%的用户签到数据作为训练集,其余 20%作为测试集。

表 2 实验用数据集的统计信息

Table 2 Statistical information of experimental dataset

| 用户数量 | 签到数量 | POI 数量 | POI 分类数 | 评论文本数量 | 社交关系数量 | 用户-POI 矩阵密度/% |
|-------|---------|--------|---------|---------|--------|---------------|
| 10462 | 1805955 | 13921 | 8 | 1661479 | 76879 | 1.24 |

3.2 评测指标

采用推荐系统中广泛使用的准确率($P@N$)和召回率($R@N$)来评价兴趣点推荐的性能。准确率指推荐结果中相关的兴趣点与推荐结果中兴趣点总数的比率,反映了推荐的准确性;召回率指推荐结果中相关的兴趣点与测试集中相关的兴趣点总数的比率,反映了推荐的全面性。相关计算公式如式(15)和(16)所示:

$$P@N = \frac{\sum_{u \in U} |Top_N(u) \cap L(u)|}{\sum_{u \in U} |Top_N(u)|} \quad (15)$$

$$R@N = \frac{\sum_{u \in U} |Top_N(u) \cap L(u)|}{\sum_{u \in U} |L(u)|} \quad (16)$$

其中, $Top_N(u)$ 表示算法执行后获得的 TOP_N 兴趣点推荐列表, $L(u)$ 表示测试集中用户去过的兴趣点列表。

3.3 性能对比

为了验证本文提出的 GCSR 模型的性能,选择当前较新的 4 个代表性模型进行性能对比,包括 HFC 模型、US 模型、USG 模型与 JRA 模型。

1)HFC 模型^[2]:融合了用户相似性与好友信任的地点推荐。

2)US 模型^[3]:考虑了社交关系、用户偏好与流行度因素,使用基于 HITS 的随机游走方法进行兴趣点推荐。

3)USG 模型^[4]:线性融合用户偏好、社交与地理信息的兴趣点推荐框架。

4)JRA 模型^[6]:融合了分类因素、流行度因素、位置因素、社交好友因素与用户历史签到行为的兴趣点推荐方法。

5)GCSR 模型:本文提出的融合了分类流行度偏好、地理位置、社交关系与情感倾向因素的兴趣点推荐模型。

实验中,推荐列表长度设定为 $N=5, 10, 15, 20$ 。实验结果如图 2 与图 3 所示。

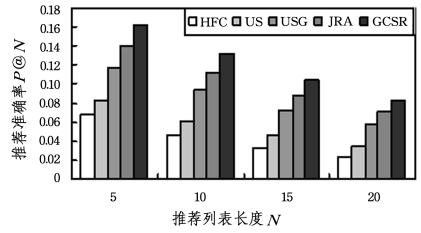


图 2 同类模型推荐的准确率对比

Fig. 2 Precision comparison of similar recommendation models

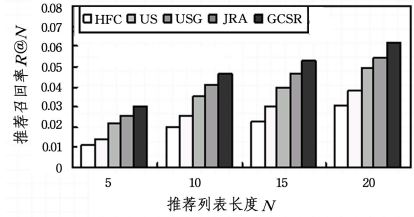


图 3 同类模型推荐的召回率对比

Fig. 3 Recall comparison of similar recommendation models

LBSN 中用户的签到矩阵密度很低,导致 POI 推荐的绝对准确率通常不高^[4,8,17],因此,本文重点比较相对准确率。从图 2 和图 3 可以看出,随着推荐列表长度 N 值的增大,推荐准确率减小,而召回率增大。这是由于推荐 POI 数目越多,用户在推荐列表中真实访问的 POI 的比例会降低,即有更多的推荐出的 POI 不是用户真正访问的 POI,导致推荐准确率下降;而当推荐 POI 数目增加时,有更多的用户真正访问的 POI 会进入推荐列表,导致推荐召回率有所增加。

此外,HFC 模型将用户集合划分为非好友集和好友集,根据用户签到历史行为与社交关系分别设计基于用户相似性和基于好友信任的兴趣点推荐算法,最后采用线性加权的方法组合得到推荐列表。HFC 模型由于仅考虑了用户签到历史行为与社交关系因素,没有结合 LBSN 中的地理信息、分类信息、流行度信息及用户的评论信息,因此模型最为简单,推荐准确率最低。US 模型将用户的社交因素、分类偏好因素和 POI 流行度因素结合到基于 HITS 的随机游走模型中,以估计用户签到 POI 的概率。US 模型由于考虑了社交因素、分类偏好因素与 POI 流行度因素,因此与 HFC 模型相比,准确率和召回率有了明显提升,但与 GCSR 模型相比,由于没有考虑地理位置因素和用户评论的情感倾向,因此准确率提升有限。USG 模型将用户偏好、社交影响和地理分布整合到基于用户和基于朋友的协同过滤推荐中,最后将 3 个因素的推荐分值进行线性组合以计算推荐结果。与 US 模型相比,USG 模型由于引入了地理因素,使得推荐准确率和召回率都有较大提升。但与 GCSR 模型相比,USG 模型没有利用分类信息、流行度信息和用户评论信息,因此模型的推荐性能还不能达到最佳效果。JRA 模型考虑了分类因素、流行度因素、位置因素、社交好友因素与用户历史签到行为,并与协同过滤推荐方法相结合进行推荐。与 USG 模型相比,JRA 模型增加了分类与流行度信息,使得推荐质量有所提升;但与 GCSR 模型相比,JRA 模型未考虑用户评论的情感倾向,因此推荐

效率较低。GCSR模型综合考虑了分类流行度偏好、地理位置、社交关系与情感倾向因素,并与协同过滤方法相融合进行推荐,因此推荐准确率最高。与推荐质量较好的JRA方法相比,GCSR模型平均提高了1.7%的准确率和0.6%召回率。以上数据说明,在兴趣点推荐中引入用户评论的情感倾向因素,能够明显提高推荐的效果。

3.4 不同因素的影响对比

为了分析GCSR模型中地理位置、分类流行度偏好、社交强度、情感倾向等因素在POI推荐中的重要程度,本文设计并对比了所提出的GCSR模型的4个简化模型,即GCS模型、CSR模型、GCR模型和GSR模型。其中,GCS模型是省略情感倾向因素后构建的测试模型;CSR模型是省略地理位置因素后的简化模型;GCR模型和GSR模型分别是移除社交因素和分类流行度因素后的简化模型。实验中设定推荐列表的长度 N 分别为5,10,15,20。实验结果如图4与图5所示。

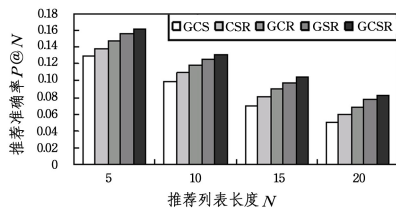


图4 不同因素影响推荐准确率的对比

Fig. 4 Precision comparison of recommendation models with different influential factors

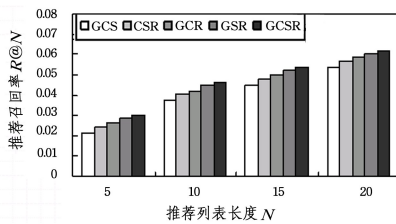


图5 不同因素影响推荐召回率的对比

Fig. 5 Recall comparison of recommendation models with different influential factors

从图4和图5可以看出,在推荐列表长度 N 取值相同的情况下,融合了地理位置、分类流行度偏好、社交强度、情感倾向等因素的GCSR模型的效果最好,推荐准确率和召回率均高于对比的GCS,CSR,GCR和GSR4种模型,说明融合了地理位置、分类流行度偏好、社交强度、情感倾向等因素的方法能够更好地建模用户的签到行为,有助于提高推荐的准确率和召回率。这与实际生活中人们的选择习惯相吻合,用户在选择自己感兴趣的地点时通常会考虑自己的喜好、地点的人气、路途的远近、朋友的选择与用户的评价。此外,地理位置、分类流行度偏好、社交强度与情感倾向因素在兴趣点推荐中所起的作用也是不同的,4种对比模型在推荐列表长度 N 取值相同的情况下,准确率和召回率按照从高到低的顺序依次是GSR,GCR,CSR及GCS,说明4种因素在提高兴趣点推荐质量中发挥的作用从大到小依次是情感倾向因素、地理位置因素、社交强度因素、分类流行度因素。

结束语 本文综合考虑了兴趣点的地理位置、分类偏好、流行度、社交与情感倾向因素,提出了一种有效融合多种因素的兴趣点协同推荐模型。该模型从大规模的签到历史记录中提取数据,根据POI地理位置数据计算地理相关分数;根据用户的分类偏好,结合POI流行度计算分类偏好分数;根据社交关系计算用户之间的社交强度,根据评论文本分析用户的情感倾向,并将二者与协同过滤推荐技术有效结合,从而得到社交情感分数;最后,将地理相关分数、分类偏好分数与社交情感分数有效融合,向用户推荐感兴趣的地理位置。在大规模Foursquare真实签到数据集上的多组对比实验,证明了本文方法的有效性。

目前,越来越多的研究利用时间影响来改进现有的位置推荐方法^[18-20]。时间影响拥有多维属性,如周期性、同构性、连续性与非均匀性等,然而,这些属性对POI推荐的影响尚未得到充分研究^[21]。此外,LBSN历史数据中的照片、视频等媒体内容提供了额外的偏好信息,有助于深入理解与建模用户行为,改善推荐效果^[20]。未来工作中,我们将进一步探索时间因素以及用户在POI生成的照片、视频等内容信息对用户签到行为的影响,以进一步提高推荐质量。

参考文献

- [1] ZHAO S, KING I, LYU M R. A Survey of Point-of-interest Recommendation in Location-based Social Networks[C]// Proceedings of the 29th AAAI Conference on Artificial Intelligence. California: AAAI Press, 2016: 53-60.
- [2] DING Y, LIU J. Venue Recommendation Algorithm Combined with Similarity and Friend-trust in LBSN[J]. Computer Engineering and Design, 2016, 37(8): 2120-2124. (in Chinese)
丁勇,刘菁. LBSN中融合相似性和好友信任的地点推荐算法[J]. 计算机工程与设计, 2016, 37(8): 2120-2124.
- [3] YING J C, KUO W N, TSENG V S, et al. Mining User Check-in Behavior with a Random Walk for Urban Point-of-interest Recommendations[J]. ACM Trans on Intelligent Systems and Technology, 2014, 5(3): 1-26.
- [4] YE M, YIN P F, LEE W C, et al. Exploiting Geographical Influence for Collaborative Point-of-interest Recommendation[C]// Proceedings of the 34th ACM SIGIR International Conference on Research and Development in Information Retrieval. New York: ACM Press, 2011: 325-334.
- [5] RAHIMI S M, WANG X, FAR B. Behavior-Based Location Recommendation on Location-Based Social Networks[C]// Proceedings of the 21th Pacific-Asia Conference on Knowledge Discovery and Data Mining. New York: Springer, 2017: 273-285.
- [6] HAN X F, NIU B N, YANG R. Personalized Point-of-interest Recommendation in Location-based Social Networks[J/OL]. [2018-03-09]. <http://www. arocmag. com/article/02-2019-05-011. html>. (in Chinese)
韩笑峰,牛保宁,杨草. 基于位置社交网络的个性化兴趣点推荐[J/OL]. [2018-03-09]. <http://www. arocmag. com/article/02-2019-05-011. html>.

- [7] WANG X Y, YUAN J L, QIN F. Point-of-interest Recommendation Based on Comment Text in Location Social Network[J]. *Computer Science*, 2017, 44(12): 245-248. (in Chinese)
王啸岩, 袁景凌, 秦凤. 位置社交网络中基于评论文本的兴趣点推荐[J]. *计算机科学*, 2017, 44(12): 245-248.
- [8] REN X Y, SONG M N, SONG J D. Context-aware Point-of-interest Recommendation in Location-based Social Networks[J]. *Chinese Journal of Computers*, 2017, 40(4): 824-841. (in Chinese)
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.
- [9] GAO H, LIU H. *Mobile Social Networking: Data Analysis on Location-Based Social Networks* [M]. New York: Springer, 2014: 165-194.
- [10] PHAM T A N, LI X, CONG G. A General Model for Out-of-town Region Recommendation [C]// proceedings of the 26th International Conference on World Wide Web. New York: ACM Press, 2017: 401-410.
- [11] BAO J, ZHENG Y, MOKBEL M F. Location-based and Preference-aware Recommendation Using Sparse Geo-social Networking Data [C]// Proceedings of the 20th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. New York: ACM Press, 2012: 199-208.
- [12] YANG D, ZHANG D, YU Z, et al. A Sentiment-enhanced Personalized Location Recommendation System [C]// Proceedings of the 24th ACM Conference on Hypertext and Social Media. New York: ACM Press, 2013: 119-128.
- [13] TURNEY P D. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews [C]// Proceedings of the 40th Annual Meeting on Association for Computational Linguistics. ACL, 2002: 417-424.
- [14] NJOO G S, KAO M C, HSU K W, et al. Exploring Check-in Data to Infer Social Ties in Location Based Social Networks [C]// Proceedings of the 21th Pacific-Asia Conference on Knowledge Discovery and Data Mining. New York: Springer, 2017: 460-471.
- [15] BARAL R, LI T. MAPS: A Multi Aspect Personalized POI Recommender System [C]// Proceedings of the 10th ACM Conference on Recommender Systems. New York: ACM Press, 2016: 281-284.
- [16] CHENG Z Y, CAVERLEE J, LEE K, et al. Exploring Millions of Footprints in Location Sharing Services [C]// Proceedings of the 5th International Conference on Weblogs and Social Media. California: AAAI Press, 2011: 81-88.
- [17] ZHANG J D, CHOW C Y. CoRe: Exploiting the Personalized Influence of Two-dimensional Geographic Coordinates for Location Recommendations [J]. *Information Sciences*, 2015, 29(1): 163-181.
- [18] HOSSEINI S, LI L T. Point-of-interest Recommendation Using Temporal Orientations of Users and Locations [C]// Proceedings of International Conference on Database Systems for Advanced Application. New York: Springer, 2016: 330-347.
- [19] LIAO D, ZHONG Y, LI J. Location Prediction Through Activity Purpose: Integrating Temporal and Sequential Models [C]// Proceedings of the 21th Pacific-Asia Conference on Knowledge Discovery and Data Mining. New York: Springer, 2017: 711-723.
- [20] LIU S D. User Modeling for Point-of-interest Recommendations in Location-based Social Networks: The State of the Art [J]. *Mobile Information Systems*, 2018, 2018(4): 1-3.
- [21] HOSSEINI S, YIN H, ZHANG M, et al. Jointly Modeling Heterogeneous Temporal Properties in Location Recommendation [C]// Proceedings of International Conference on Database Systems for Advanced Applications. New York: Springer, 2017: 490-506.