基于用户签到和地理属性的个性化位置推荐算法研究

蔡海尼[1] 陈 程[1] 文俊浩[1] 王喜宾[2] 曾 骏[1]
(重庆大学软件学院 重庆 401331)[1] (重庆邮电大学软件工程学院 重庆 400065)[2]

摘 要 针对基于LBSNs (Location-based Social Networks)的位置推荐算法考虑因素单一且不能有效解决用户位于不同城市的位置推荐的问题,综合考虑潜在的社交影响、内容匹配影响和地理属性影响等因素,提出了基于用户签到和地理属性的个性化位置推荐算法SCL (Social-Content-Location)。该算法在协同过滤的基础上,引入了用户兴趣特征比较,改进了用户的相似度计算;同时,在分析位置的内容信息时,融入用户评论,缓解了位置标签的短文本特性对LDA (Latent Dirichlet Allocation)主题提取的影响,提高了用户兴趣和城市偏好主题提取的准确率。实验结果表明,SCL算法在本地城市召回率上较协同过滤算法U提高近65%,较LCA-LDA算法提高近30%;在异地城市召回率上,高于LCA-LDA算法近26%。这表明SCL算法在不同城市下的位置推荐具有一定的可行性。

关键词 潜在社交影响,内容匹配影响,地理属性影响,协同过滤,LDA主题提取
中图法分类号 TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2016.12.029

Personalized Location Recommendation Algorithm Research Based on User Check-ins and Geographical Properties

CAI Hai-ni[1] CHEN Cheng[1] WEN Jun-hao[1] WANG Xi-bin[2] ZENG Jun[1]
(School of Software Engineering, Chongqing University, Chongqing 401331, China)[1]
(School of Software Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)[2]

Abstract Since the consideration of location recommendation algorithms based on LBSNs (Location-Based Social Networks) is too single, and it couldn't effectively solve the problem of location recommendation for user in different cities, synthesizing the factors of potential social influence, content match influence and geographical property influence, the personalized location recommendation algorithm SCL (Social-Content-Location) based on user check-ins and geographical properties was proposed. SCL algorithm introduces the comparison of users' interest features based on the collaborative filtering, and it improves the similarity of users. At the same time, when the content information of location is analyzed, user's comments on location is integrated, and it alleviates the influence of the short text feature of location labels to LDA (Latent Dirichlet Allocation) topic extraction and improves the accuracy of user's interest and city preference topic in extraction. The experimental results show that, for the recall rate of residence city, algorithm SCL outperforms collaborative filtering algorithm U near 65%, and outperforms algorithm LCA-LDA near 30%. For the recall rate of new city, algorithm SCL outperforms algorithm LCA-LDA near 26%, which shows that algorithm SCL has certain feasibility for location recommendation under different cities.

Keywords Potential social influence, Content match influence, Geographical property influence, Collaborative filtering, LDA topic extraction

1 引言

智能终端和移动互联网的发展促进了基于位置的社交网络(LBSNs)的兴起,用户可以随时随地地签到,进行信息共享。分析用户在LBSNs中的签到数据,有助于推测用户对物理位置的访问偏好,进而实现用户的个性化位置推荐。文献[1-5]基于协同过滤为用户推荐感兴趣的位置,并取得了较好的效果。然而,协同过滤推荐的准确率严重依赖于用户签到数据的稠密度,为缓解数据稀疏对协同过滤位置推荐的准确度影响,文献[6-10]通过分析用户在LBSNs中访问过的位置信息,从语义层面提取出用户的兴趣特征,有效实现了在用户拥有少量访问记录情况下的位置推荐。同时,文献[1-4,6,7]发现,用户的签到行为呈现集聚分布的幂律特性,当用户与位置的距离超过20km时,用户的访问记录急剧下降。文献[11]研究了用户在Foursquare的签到数据,结果表明,45%的用户选择的位置距离不超过10英里,75%的用户选择的位置距离不超过50英里,可见距离也是影响用户位置选择的重要因素。

到稿日期:2016-02-23 返修日期:2016-03-23 本文受国家自然科学基金(61379158,61502062),科技支撑计划(2014BAH25F01),重庆市科技计划项目(cstc2014jcyjA40054)资助。

蔡海尼(1972-),女,博士,副教授,主要研究方向为个性化推荐,E-mail:hainic@tom.com;陈 程(1991-),女,硕士生,CCF会员,主要研究方向为个性化推荐,E-mail:20092154@cqu.edu.cn;文俊浩(1969-),男,博士,教授,博士生导师,CCF高级会员,主要研究方向为软件服务工程;王喜宾(1985-),男,博士,主要研究方向为机器学习;曾 骏(1984-),男,博士,主要研究方向为软件服务工程。

基于协同过滤的位置推荐在计算用户的相似度时，仅考虑了用户共同访问过的位置数量，忽略了由位置标签的多样化带来的影响，即兴趣不同的人也可能访问相同的位置，缺乏对用户兴趣相似度的定性分析，使得计算出的用户相似度与实际情况存在偏差。多数文献在分析位置的内容信息及为用户推荐感兴趣的位置时，仅考虑了位置的类别和标签。由于用户签到具有任意性，单从类别和标签分析的用户兴趣缺乏具体性和针对性，此外，也并未考虑其他因素的影响，尤其缺乏针对用户在不同城市下的有效位置推荐的研究。针对以上问题，本文提出基于用户签到和地理属性的个性化位置推荐算法，算法综合考虑潜在的社交影响、内容匹配和地理属性等因素的共同作用，并分别对用户在本地城市和异地城市的位置推荐进行了深入分析。

2 基于用户签到和地理属性的个性化位置推荐框架

传统基于协同过滤的位置推荐算法在分析用户间的相似性时，仅以两个用户访问过的共同位置数量为衡量标准，认为两个用户位置的重合数越多，则相似度越高[2,3,5]，其潜在的社交关系也越强，因而对目标用户的位置推荐的影响也越大。该方法仅对用户的访问历史进行了定量分析，忽略了由位置标签的多样化特点产生的影响，即访问过同一位置的不同用户不一定具有相同的兴趣爱好，缺乏用户相似性的定性分析。此外，即使两个用户共同访问的位置数量少，也并不表示其兴趣相似度也低。针对以上问题，本文在协同过滤的基础上引入用户的兴趣特征比较，提出改进后的用户相似度计算方法，以更准确地刻画用户间的相似性，并分析改进后的相似用户对目标用户位置推荐的影响，称之为潜在的社交影响。

由于用户在异地城市的签到数据稀少，协同过滤算法并不适用于该位置推荐场景，因此本文采取分析用户位置签到历史的内容信息的方法，挖掘出用户的兴趣主题特征，并为其推荐匹配的位置，以缓解协同过滤算法对用户签到数据稀疏的敏感问题。此外，为使异地城市的位置推荐更符合用户的签到习惯，本文综合考虑城市偏好对用户兴趣的影响。当前大多数文献在分析位置的内容信息时主要考虑其类别和标签，本文认为，出现在用户对位置评论中的关键词更能反映用户的兴趣；同理，分析某位置的所有用户的评论也能从用户体验的角度反映该位置的主题，从而有助于更准确地刻画城市偏好。综上所述，本文将位置的类别、标签和用户评论作为待分析的内容信息，从中提取出用户兴趣和城市偏好，并计算用户兴趣与城市偏好对待推荐位置的共同影响，称之为内容匹配影响。

某一位置的访问人数和访问次数越多，表明该位置的知名度越高，对用户的影响也越大，尤其是当用户对某个位置不了解时，位置的知名度更是影响用户选择的重要指标。此外，用户访问历史中位置与距离的关系表明，距离也是影响用户位置选择的重要因素，距离越近，位置被选择的概率越大，反之，则概率越小。因此，本文考虑位置的知名度和距离对推荐的共同影响，称之为地理属性影响。

不同因素对位置推荐的影响效果会因为用户是身处于本地城市还是异地城市而发生变化，单纯考虑某种因素不能有效地适应不同场景下的位置推荐。本文融合潜在社交影响、内容匹配影响和地理属性影响等因素，提出基于用户签到和地理属性的个性化位置推荐算法 SCL，实现用户在不同城市下的有效位置推荐，其推荐框架图如图 1 所示。

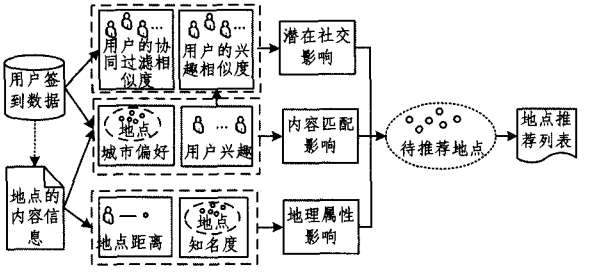


图 1 基于用户签到和地理属性的个性化位置推荐框架图

图 1 的推荐框架综合考虑了潜在社交影响、内容匹配影响和地理属性影响 3 种因素。其中，潜在社交影响衡量了相似用户对位置推荐的影响；内容匹配影响缓解了位置推荐对用户签到数据的依赖；地理属性影响则从位置的知名度和距离两方面分析其对用户位置选择的作用。本文认为，综合考虑以上 3 种因素，可实现各个因素之间的互补效果，因而可在不同场景下进行有效的位置推荐。

3 基于用户签到和地理属性的个性化位置推荐

3.1 LDA 主题模型和本文符号含义

分析内容信息以提取主题的常用方法包括基于 TF-IDF 的向量空间模型和 LDA 主题模型[12]。TF-IDF 方法综合词语在文本和文本集中出现的频率来表征其权重，由于文本的语义多样性，该方法无法获取具有相同语义的不同词语之间的联系，同时也无法区分相同词语在不同上下文中的不同含义。LDA 主题模型通过对文本集进行建模，利用文本的统计特性，将文本语料库映射到各个主题空间中，从而挖掘出隐藏在文本内的不同主题与词语之间的关系，被广泛用于文本主题分析。

LDA 模型作为一种对文本集进行建模的概率主题模型，通过主题关键信息对文本进行简短的描述，保留了本质信息，其模型示意图如图 2 所示。

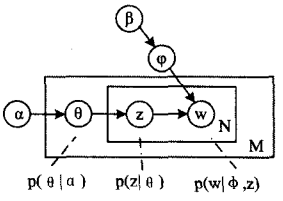


图 2 LDA 主题模型

图 2 中先验参数 α 反映了文本集中隐含主题间的相对强弱，β 刻画了所有隐含主题自身的概率分布，令 θ 和 φ 分别表示主题和词语的概率分布，z 表示特定主题，w 表示特定词语，N 表示特定文本中的词语数量，M 表示文本数量，则 LDA 生成文本的过程如下：

(1)根据先验参数 α，获取用户兴趣的主题概率分布 θ；

(2)根据先验参数 β，获得各主题对应的词语多项式分布 φ；

(3)从 θ 中随机选择主题 z，并根据该主题下的词语概率分布 φ 生成用户文档内容中的词语 w。

LDA 主题建模过程中最重要的是模型参数的估计，Gibbs 抽样的参数推理方法容易理解，并能有效地从文本集中提取出主题，使得其成为当前最流行的 LDA 主题模型抽样方法。

本文所涉及的符号及含义如表 1 所列。

表 1　符号及含义

符号	含义
α,β	LDA 先验参数
μ,ρ,γ,δ	参数因子
m	用户总数
n	位置总数
U	用户集合
V	位置集合
D	所有用户的位置访问历史集合
u_i	第 i 个用户
v_j	第 j 个位置
d_i	u_i 的位置访问历史
d_{ij}	u_i 对 v_j 的访问记录
d_{im}	u_i 的位置访问记录数
C	用户-位置签到矩阵
c_{ij}	u_i 对 v_j 的访问偏好
R	城市集合
r_i	第 i 个城市的位置集合
θ_i,θ_k	u_i 和 u_k 的兴趣主题概率分布
φ_i,φ_k	u_i 和 u_k 的主题词语概率分布
θ_{ri}	第 i 个城市偏好的主题概率分布
φ_{ri}	第 i 个城市偏好的主题词语概率分布
z	LDA 主题模型中的主题
w	词语
C_{vj}	v_j 的内容信息
K	主题数量
$f(u_i)$	u_i 的相似用户集合
U_{vj}	访问过 v_j 的用户集合
R	地球半径,6370996.81
π	圆周率
$C_{uk,vj}$	用户 u_k 对 v_j 的访问次数

3.2　个性化位置推荐算法

3.2.1　潜在社交影响

本文在协同过滤的基础上引入用户的兴趣特征相似性比较,提出改进后的用户相似度计算方法,并以此获取目标用户的潜在相似用户集合,将相似用户对待推荐位置的影响均值作为对目标用户位置选择的潜在社交影响。

根据所有用户的签到记录,借鉴文献[12]的函数$(1+x^{-1})^{-1}$对用户-位置签到矩阵进行归一化处理后,得到矩阵 C 如式(1)所示:

$$C=\begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m1} & c_{m2} & \cdots & c_{mn} \end{pmatrix} \tag{1}$$

定义 1　令 $\omega_{i,k}$ 表示用户 u_i 和 u_k 的协同过滤相似度,采用余弦相似性计算公式,有:

$$\omega_{i,k}=\frac{\sum_{v_j\in V} c_{ij}\cdot c_{kj}}{\sqrt{\sum_{v_j\in V} c_{ij}^2}\cdot\sqrt{\sum_{v_j\in V} c_{kj}^2}} \tag{2}$$

文献[6-10]利用 LDA 模型在文本主题提取方面的优势,根据用户访问过的位置信息,提取出用户的兴趣主题,有效地为用户推荐了感兴趣的位置。然而,由于用户的访问记录有限,且位置的类别和标签多表现为短文本,这为 LDA 主题模型有效地提取主题特征提出了挑战。为缓解短文本特性对 LDA 主题提取的影响,本文算法在位置的类别和标签的基础上融入用户的评论,利用出现在用户评论中的关键词以更准确地获取用户兴趣。

定义 2　令位置 $v_j=(id,name,city,lat,long,cate,tag,con,un,cn)$,其中,$id$ 表示唯一标识,$name$ 表示名称,$city$ 表示所在城市,lat 和 $long$ 分别表示经度、纬度,$cate$ 表示类别信息,tag 表示标签信息,con 表示用户的评论,un 表示签到人数,cn 表示签到次数;$d_{ij}=(u_i,v_j,c_j)$ 表示 u_i 对 v_j 的某次访问记录,其中,c_j 表示 u_i 对 v_j 的评论。根据 u_i 的位置访问历史,综合所有的位置类别、标签和 u_i 的评论,采用 LDA 模型获得用户 u_i 的兴趣主题概率分布为 θ_i,对应主题的词语概率分布为 φ_i;同样地,根据用户 u_k 的访问历史,获得其兴趣主题概率分布为 θ_k,对应主题的词语概率分布为 φ_k。

本文采用 JS 距离[13]根据用户主题的概率分布计算其兴趣相似度,JS 距离越大,表示用户的兴趣特征离异度越高,相似度越低;反之,JS 距离越小,表明用户兴趣特征的离异度越低,相似度越高。

定义 3　令 $inter_{i,k}$ 表示用户 u_i 和 u_k 的兴趣特征相似度,根据 JS 距离计算可得:

$$\begin{aligned} inter_{i,k}&=\frac{1}{D_{js}(\theta_i,\theta_k)} \\ &=\frac{1}{0.5*(D_{kl}(\theta_i,\frac{\theta_i+\theta_k}{2})+D_{kl}(\theta_k,\frac{\theta_i+\theta_k}{2}))} \end{aligned} \tag{3}$$

$$D_{kl}(\theta_i,\theta_k)=\sum_{n=1}^{K}\theta_{i_n}\ln\frac{\theta_{i_n}}{\theta_{k_n}} \tag{4}$$

定义 4　令 $\omega'_{i,k}$ 表示用户 u_i 和 u_k 改进后的相似度,综合式(2)和式(3),有:

$$\omega'_{i,k}=\mu*\omega_{i,k}+(1-\mu)*inter_{i,k} \tag{5}$$

定义 5　令 $S_{i,j}$ 表示用户 u_i 受到的潜在社交影响,在式(5)基础上可得:

$$S_{i,j}=\frac{\sum_{u_k\in f(u_i)}\omega'_{i,k}\cdot c_{k,j}}{\sum_{u_k\in f(u_i)}\omega'_{i,k}} \tag{6}$$

改进后的用户相似度融合了用户的协同过滤相似度和兴趣相似度,相较于单纯的协同过滤算法,缓解了位置标签的多样化对用户相似性的干扰;同时,由于用户偏向访问近距离的位置,考虑协同过滤相似性,可有效避免用户之间兴趣相似而物理距离较远的情况,提高了位置推荐的用户接受度。

3.2.2　内容匹配影响

基于协同过滤的位置推荐准确率依赖于用户签到数据的稠密度,在通常只有少量签到数据的异地城市的位置推荐中表现较差。为此,本文采用 LDA 主题模型,引入用户兴趣和推荐位置的内容匹配,以缓解推荐算法对用户签到数据的过度依赖。在分析位置信息时,本文综合考虑位置的类别、标签和用户评论,以提高主题提取的准确率。

定义 6　令 $uinter_{i,j}$ 表示用户 u_i 的兴趣和位置 v_j 的内容匹配度,采用 LDA 主题模型匹配度计算公式,有:

$$\begin{aligned} uinter_{i,j}&=p(u_i,v_j|\alpha,\beta) \\ &=p(\theta_i|\alpha)p(\varphi_i|\beta)p(u_i,v_j|\theta_i,\varphi_i) \\ &=\prod_{c\in C_{v_j}}\sum_{t=1}^{K}p(z_t|\theta_i)p(c|z_t,\varphi_i) \end{aligned} \tag{7}$$

定义 7　令 $pre_{i,j}$ 表示用户当前所在城市 r_i 的城市偏好和位置 v_j 的内容匹配度,则:

$$\begin{aligned} pre_{i,j}&=p(u_i,v_j|\alpha,\beta) \\ &=p(\theta_{r_i}|\alpha)p(\varphi_{r_i}|\beta)p(u_i,v_j|\theta_{r_i},\varphi_{r_i}) \\ &=\prod_{c\in C_{v_j}}\sum_{t=1}^{K}p(z_t|\theta_{r_i})p(c|z_t,\varphi_{r_i}) \end{aligned} \tag{8}$$

定义 8　令 $Con_{i,j}$ 表示用户兴趣和城市偏好对 v_j 的共同影响,综合式(7)和式(8),有:

$$Con_{i,j}(u_i,v_j|\theta_i,\theta_{r_i},\varphi_i,\varphi_{r_i})=\rho * u\ \text{int}\ er_{i,j}+(1-\rho) * pre_{i,j} \quad (9)$$

3.2.3 地理属性影响

定义 9 令 pop_j 表示位置 v_j 的知名度，利用信息熵[14]计算公式，有：

$$pop_j=[-\sum_{u_k\in U_{v_j}}(\frac{c_{u_k,v_j}}{cn_{v_j}} * \log\frac{c_{u_k,v_j}}{cn_{v_j}})] \quad (10)$$

定义 10 令 $dis(u_i, v_j)$ 表示 u_i 和 v_j 的物理距离，(x_1, y_1) 表示 u_i 的经纬度坐标，(x_2, y_2) 表示 v_j 的经纬度坐标，有：

$$dis(u_i,v_j)=R * \arccos(\sin(y_1 * \pi/180)\sin(y_2 * \pi/180)+\cos(y_1 * \pi/180)\cos(y_2 * \pi/180)\cos(x_1 * \pi/180-x_2 * \pi/180)) \quad (11)$$

定义 11 令 $Loca_{i,j}$ 表示 v_j 对用户 u_i 的地理属性影响，综合式(10)和式(11)，有：

$$Loca_{i,j}=pop_j * e^{-dis(u_i,v_j)} \quad (12)$$

3.2.4 基于用户签到和地理属性的个性化位置推荐算法

SCL 算法的伪代码如表 2 所列。

表 2 基于用户签到和地理属性的个性化位置推荐算法伪代码

```
算法 1 基于用户签到和地理属性的个性化位置推荐算法伪代码
输入：目标用户 u_i 和位置数量 N
输出：u_i 的 Top-N 位置推荐列表 L
1. function SCL_SpatialLocations(User u_i, Number N)
2. /* 数据预处理，除去签到次数小于 5 的用户和位置 */
3. /* 根据所有用户的签到数据构建签到矩阵 */
   building check-in matrix C;
4. /* 将推荐列表 L，目标用户 u_i 的候选相似用户集 SU 分别置为空集 */
   L←∅;
   SU←∅;
5. /* 计算各用户与目标用户 u_i 的协同过滤相似度 */
   for each u_k in U do
       calculating the similarity ω_{i,k} of collaborative filtering between u_i
       and u_k
       if (ω_{i,k} > 0) do
           insert u_k into SU;
       end
       end
6. /* 计算 u_i 和其相似用户的主题概率分布 */
   extracting the topic distribution θ_i of interest of u_i with LDA;
   for each u_k in SU do
       extracting the topic distribution θ_k of interest of u_k with LDA;
   end
7. /* 计算候选相似用户集中用户 u_k 与 u_i 的兴趣相似度 */
8.  for each u_k in SU do
        calculating the similarity of interests inter_{i,k} between u_i and u_k;
   end
9. /* 改进用户的相似度 */
10.  for each u_k in SU do
       calculating the improved similarity with w_{i,k} and inter_{i,k};
   end
11. /* 定位 u_i 当前所在城市 r_i */
12. /* 计算 r_i 的城市偏好 */
   extracting the topic distribution θr_i of city preference of r_i with LDA;
13.  for each v_j in r_i do
       calculating the social influence S_{i,j} by Equation 6;
       calculating the matching degree UP_{i,j} between u_i and v_j with θ_i and
       θr_i by Equation 9;
       calculating the geographical influence Loca_{i,j} by Equation 12;
       comprehensive calculating the recommendation probability by Equa-
       tion 13;
       insert v_j into L;
   end
14. order and return L.
```

潜在社交影响在协同过滤的基础上通过引入用户的兴趣主题相似度比较，改进了用户的最终相似度；内容匹配通过分析位置的内容信息，缓解了位置推荐对用户签到数据的依赖，同时，考虑城市偏好的影响，适用于用户出行到新城市的位置推荐；地理属性从位置的知名度和距离角度出发，分析了物理特征对用户位置选择的影响。由于单纯考虑某种因素缺乏对位置推荐的全面分析，不能有效地针对不同情况进行有效的推荐，因此本文提出基于用户签到和地理属性的个性化位置推荐算法 SCL，针对待推荐位置 v_j，其被推荐给目标用户 u_i 的概率计算公式如式(13)所示：

$$p(u_i,v_j)=\gamma * S_{i,j}+\delta * Con_{i,j}+(1-\gamma-\delta) * Loca_{i,j} \quad (13)$$

4 推荐性能测试与评价

4.1 数据集和参数优化

(1)数据集：本文的实验数据来源于 Foursquare 位置社交网站。由于可获得的公共数据集不包含位置的内容信息，本文借鉴文献[10]的方法，利用 Foursquare 的“Check-in API”抓取用户的签到数据，并根据用户签到记录中的位置 *id*，通过“Venue API”平台获取位置的详细信息。在数据预处理过程中，为减少用户和位置的冷启动影响，去除签到次数小于 5 的用户和位置。本文使用的数据集情况如表 3 所列，经统计其中共包括 12681 个用户、25283 个位置和 368745 个签到记录。

表 3 数据集描述

用户数	位置数	签到数	签到/用户	签到/位置	稀疏性
12681	25283	368745	29.08	14.58	99.95%

为验证本文推荐算法在用户本地和异地城市下的推荐性能，对实验数据集进行以下两种设置：1)本地城市：采用十折交叉验证法，将用户在本地城市签到数据的 10%作为测试数据集，余下的作为训练数据集，取 10 次结果的均值作为最后结果；2)异地城市：由于用户在异地城市的签到数据稀少，将用户在查询的异地城市的签到数据作为测试数据集，余下的作为训练数据集。本文取用户在待查询城市中历史签到位置的平均经纬度作为用户当前所在的物理位置。

(2)评价标准：采用准确率和召回率作为评价标准。

(3)对比算法：本文对比分析 SCL 算法和以下几种位置推荐算法的推荐性能。1)基于协同过滤的位置推荐算法 U[5]；2)结合协同过滤和用户兴趣比较的潜在社交影响 S，该算法是 SCL 的特例；3)地理属性 L，该算法是 SCL 的特例；4)基于内容的位置推荐 C，该算法是 SCL 的特例；5)基于用户兴趣和区域偏好的位置感知推荐算法 LCA-LDA[10]。

(4)参数优化：算法 SCL 中各种因素的影响效果因推荐城市的不同而产生变化，为提高算法的整体推荐效率，本文以 Top-N 中 N 取值为 5 时的 recall@5 作为评价标准，选取各权重因子的最优值。采取十折交叉验证法，获得潜在社交影响中 μ 的最优取值为 0.8；由于用户兴趣主题和本地城市偏好较相近，为计算异地城市偏好对用户的影响，选取用户在各异地城市对应最优 ρ 值的均值 0.7 作为最后取值；算法 SCL 在

本地和异地城市推荐场景中权重 γ 和 δ 的最优取值情况如表4所列。

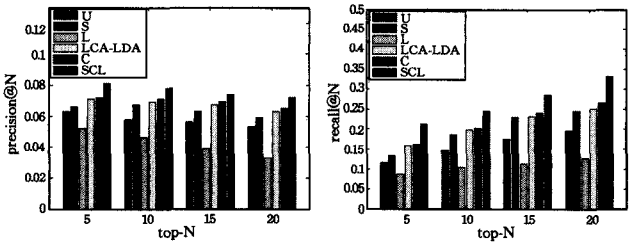
表4 参数优化情况

	γ	δ	$1-\gamma-\delta$
本地城市	0.7	0.2	0.1
异地城市	0.1	0.6	0.3

LDA主题模型建模时，本文取主题数量 $K=50$[9]，先验参数 $\alpha=50/K$，$\beta=0.01$[10]。

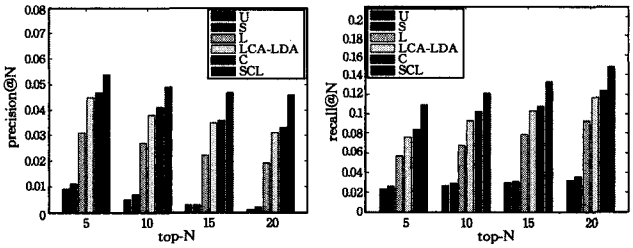
4.2 推荐性能评价

以准确率和召回率为标准，图3和图4分别示出了用户在本地和异地城市下各对比算法的位置推荐性能。



(a)准确率 (b)召回率

图3 本地城市下各算法的准确率和召回率对比



(a)准确率 (b)召回率

图4 异地城市下各算法的准确率和召回率对比

由图3可知，潜在社交影响 S 的位置推荐性能优于传统的协同过滤算法U，其中，在召回率上提高近13%，表明在协同过滤的基础上考虑用户的兴趣相似度，强化了对用户兴趣的分析，提高了用户相似度的准确率，进而提高了潜在相似用户对位置推荐的影响；算法C的推荐性能略优于LCA-LDA算法，原因可能在于本文仅是简单地对用户兴趣和城市偏好进行加权表示，并未改进LDA主题模型本身；同时，用户的评论信息相对较少，但结果仍表明研究用户的评论使得用户兴趣和城市偏好主题更具有针对性和真实性，提高了位置匹配的准确率；算法C的推荐性能优于算法S，本文认为可能是由于数据稀疏性造成的；算法SCL在各对比算法中性能表现最好，其原因可能如下：1)潜在的社交关系对用户的位置选择具有正向引导性，相似的用户具有相近的兴趣爱好，在位置内容匹配的基础上融入相似用户对位置的肯定，提高了位置推荐的准确率；2)用户签到活动的集聚特性反映了用户倾向于访问近距离的位置，以共同签到位置为基础的协同过滤方法可以挖掘出在物理距离上相近的用户，因此社交关系在某种程度上强化了对近距离位置的推荐，这正符合用户的签到习惯；3)算法SCL在社交推荐中考虑了内容匹配，可缓解数据稀疏性对协同过滤推荐的影响，同时综合考虑二者，从用户和位置内容上强化了对推荐位置的肯定；4)考虑了位置知名度的影响，更符合用户真实的位置选择倾向。算法SCL在位置推荐的召回率上，较算法U提高近65%，较算法LCA-LDA提高近30%。

由图4中可知，以协同过滤为基础的位置推荐算法表现较差，原因在于用户在异地城市中的签到数据稀疏，由算法U和S推测出的相似用户大都位于目标用户的本地城市，由于用户对近距离位置的访问偏好，由他们推荐的位置也大都位于本地城市，在物理距离上距用户当前所在位置较远，因而用户的接受率低；算法L在推荐效率上优于算法U和S，这是因为用户在异地城市的潜在社交关系较弱，所以其位置推荐效率较低；算法SCL和LCA-LDA的推荐效率优于算法U、算法S和算法L，表明分析位置的内容信息为用户推荐了与其兴趣爱好匹配的位置，优于异地城市下的潜在社交影响，同时相比于地理因素的影响，也更具有用户个性化；算法SCL在所有对比算法中表现最优，表明综合考虑各种因素提高了位置推荐的性能，其中，相较于算法LCA-LDA，算法SCL在召回率上提高近26%。

结束语 本文融合潜在社交影响、内容匹配影响和地理属性影响等因素，提出了基于用户签到和地理属性的个性化位置推荐算法SCL。实验结果表明，在协同过滤基础上引入用户的兴趣特征比较，改进了用户的相似度计算，提高了查找相似用户的准确率，进而提高了相似用户对位置推荐的影响力；分析位置的内容信息时融入用户评论，可以更具针对性地提取出用户兴趣和城市偏好主题；同时引入位置的地理属性，综合考虑以上因素，弥补了考虑单一因素对位置推荐的不足，有效适应了不同城市下的位置推荐。由于用户在受用户关系影响选择位置时，更容易选择好友曾经访问过的位置，而本文并未考虑用户的好友关系对位置推荐的影响；同时，分析用户评论信息时，忽略了用户的情感，所以可能将用户不喜欢的关键词误认为用户兴趣。因此，本文的下一步研究工作将从好友关系及其信任度[14,15]和用户的情感分析方面对位置推荐算法加以完善。

参 考 文 献

[1] Ference G, Ye M, Lee W C. Location recommendation for out-of-town users in location-based social networks[C]// Acm International Conference on Conference on Information & Knowledge Management. ACM, 2013: 721-726

[2] Cheng N H, Chang C H. Evaluation of Social, Geography, Location Effects for Point-of-Interest Recommendation[C]// Proceedings of the 2013 IEEE 13th International Conference on Data Mining Workshops. IEEE Computer Society, 2013: 766-772

[3] Ye M, Yin P, Lee W C, et al. Exploiting geographical influence for collaborative point-of-interest recommendation[C]// Proceedings of the 34th international ACM SIGIR conference on Research and Development in Information Retrieval. ACM, 2011: 325-334

(下转第178页)

learning algorithms [J]. Machine Learning, 2013, 90: 317-346

[8] Brzezinski D, Stefanowski J. Prequential AUC for Classifier Evaluation and Drift Detection in Evolving Data Streams [C]// Third International Workshop NFMCP 2014 Held in Conjunction with ECML(PKDD 2014). Heidelberg: Springer, 2015: 87-101

[9] Rutkowski L, Pietruczuk L, Duda P, et al. Decision trees for mining data streams based on the McDiarmid's bound [J]. IEEE Transactions on Knowledge and Data Engineering, 2013, 25(6): 1272-1279

[10] Domingos P, Hulten G. Mining High-Speed Data Streams [C]// Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2000: 71-80

[11] Magdalena. Batch Weighted Ensemble for Mining Data Streams with Concept Drift [C]// 9th International Symposium(ISMIS 2011). Heidelberg: Springer, 2011: 290-299

[12] Zhang Peng, Zhu Xing-quan, Shi Yong, et al. An Aggregate Ensemble for Mining Concept Drifting Data Streams with Noise [C]// 13th Pacific-Asia Conference, PAKDD 2009. Heidelberg: Springer, 2009: 1021-1029

[13] Wang Hai-xun, Fan Wei, Yu P S, et al. Mining Concept-Drifting Data Streams Using Ensemble Classifiers [C]// Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2003: 226-235

[14] Wozniak M, Kasprzak A, Cal P. Weighted Aging Classifier Ensemble for the Incremental Drifted Data Streams [C]// 10th International Conference, FQAS 2013. Heidelberg: Springer, 2013: 579-588

[15] Zhou Ji-xiang, Mao Shi-song. Statistical Methods for Quality Management [M]. Beijing: China Statics Press, 2008: 433-440(in Chinese)
周纪芗, 茆诗松. 质量管理统计方法[M]. 北京: 中国统计出版社, 2008: 433-440

[16] Wang Tao, Liu Ming-ju, Li De-ming. A Sstrong Chernoff Bounds Derived from Equitable Colorings of Graphs [J]. Journal of Mathematics, 2014, 34(6): 1015-1024

[17] Zliobaite I, Bifet A, Read J, et al. Evaluation methods and decision theory for classification of streaming data with temporal dependence [J]. Machine Learning, 2015, 98(3): 455-482

[18] Zhang Chen-guang, Zhang Yan. Semi-Supervised Learning [M]. Beijing: China Agriculture Scientech Press, 2013: 31-33(in Chinese)
张晨光, 张燕. 半监督学习[M]. 北京: 中国农业科学技术出版社, 2013: 31-33

[19] Li Pei-pei, Wu Xin-dong, Hu Xue-gang, et al. Learning concept-drifting data streams with random ensemble decision trees [J]. Neurocomputing, 2015, 166(c): 68-83

[20] Bofet A, Holmes G, Kirkby R, et al. MOA: Massive Online Analysis [J]. The Journal of Machine Learning Research, 2010, 11(2): 1601-1604

(上接第 167 页)

[4] 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

[5] Zhou D, Wang B, Rahimi S M, et al. A Study of Recommending Locations on Location-Based Social Network by Collaborative Filtering[M]// Advances in Artificial Intelligence. Springer Berlin Heidelberg, 2012: 255-266

[6] Hu B, Ester M. Social Topic Modeling for Point-of-Interest Recommendation in Location-Based Social Networks[C]// 2014 IEEE International Conference on Data Mining (ICDM). IEEE Computer Society, 2014: 845-850

[7] Jiang S, Qian X, Shen J, et al. Author Topic Model based Collaborative Filtering for Personalized POI Recommendation[J]. IEEE Transactions on Multimedia, 2015, 17(6): 907-918

[8] Liu B, Xiong H. Point-of-Interest Recommendation in Location Based Social Networks with Topic and Location Awareness[C]// SDM. 2013: 396-404

[9] Yin H, Zhou X, Shao Y, et al. Joint Modeling of User Check-in Behaviors for Point-of-Interest Recommendation[C]// Proceedings of the 24th ACM International on Conference on Information and Knowledge Management. ACM, 2015: 1631-1640

[10] Gao H, Tang J, Hu X, et al. Content-aware point of interest recommendation on location-based social networks[C]// Proceedings of the 29th AAAI Conference on Artificial Intelligence. 2015

[11] LI Gui, CHEN Sheng-hong, HAN Zi-yang, et al. Location-aware Recommendation Based on Collaborative Filtering[J]. Computer Science, 2014, 41(11A): 340-346(in Chinese)
李贵, 陈盛红, 韩子阳, 等. 基于协同过滤的位置感知推荐[J]. 计算机科学, 2014, 41(11A): 340-346

[12] Gao H, Tang J, Hu X, et al. Exploring temporal effects for location recommendation on location-based social networks[C]// Proceedings of the 7th ACM Conference on Recommender Systems. ACM, 2013: 93-100

[13] Wang Zhen-zhen, He Ming, Du Yong-ping. Text Similarity Computing Based on Topic Model LDA [J]. Computer Science, 2013, 40(12): 229-232(in Chinese)
王振振, 何明, 杜永萍. 基于 LDA 主题模型的文本相似度计算[J]. 计算机科学, 2013, 40(12): 229-232

[14] Zhou Er-chong, Huang Jia-jin, Xu Xin-xin. A Point-of-Interest Recommendation Method Based on User Check-in Behaviors in Online Social Networks[J]. Computer Science, 2015, 42(10): 232-234(in Chinese)
周而重, 黄佳进, 徐欣欣. 一种基于用户网络签到行为的位置推荐方法[J]. 计算机科学, 2015, 42(10): 232-234

[15] Zheng Jiong, Shi Gang. Recommender Algorithm Based on Dynamical Trust Relationship between Users[J]. Computer Science, 2015, 42(9): 230-234(in Chinese)
郑炅, 石刚. 基于用户间动态信任关系的推荐算法研究[J]. 计算机科学, 2015, 42(9): 230-234