LBSN中基于元路径的兴趣点推荐

曹玖新 董 羿 杨鹏伟 周 涛 刘 波

(东南大学计算机科学与工程学院计算机网络和信息集成教育部重点实验室(93K-9) 南京 211189)

摘 要 兴趣点(Point-Of-Interest,POI)推荐是基于位置的社交网络(Location-Based Social Networks,LBSN)中的一项重要个性化服务.由于LBSN中数据的极度稀疏性,基于协同过滤的算法推荐精度不高,文中提出基于元路径的兴趣点推荐算法.首先根据LBSN结构特征构建带权异构网络模型,其次引入元路径来描述节点间不同类型关联关系,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.文中在3个真实LBSN签到数据集上进行了实验,结果表明该算法可以有效缓解LBSN中的极度稀疏性问题,比传统推荐算法有更好的推荐效果.

关键词 基于位置的社交网络;异构网络;兴趣点推荐;元路径;数据挖掘;社会媒体

中图法分类号 TP393 **DOI号** 10.11897/SP.J.1016.2016.00675

POI Recommendation Based on Meta-Path in LBSN

CAO Jiu-Xin DONG Yi YANG Peng-Wei ZHOU Tao LIU Bo

(*School of Computer Science and Engineering, Key Laboratory of Computer Network and Information Integration of MoE of China under Grants No.93K-9, Southeast University, Nanjing* 211189)

Abstract POIs recommendation is a crucial personalized service of Location-Based Social Networks (LBSN). The extreme data sparsity in LBSN presents a big challenge to traditional collaborative filtering recommendation algorithms. To this end, we propose a novel POIs recommendation algorithm based on meta-path. Firstly, LBSN is considered as a weighted heterogeneous network according to its structure attributes, and the meta-path is introduced to describe different relations between nodes. Then the set of meta-paths between user and POI nodes is given on the basis of three degrees of influence. In addition, we calculate the eigenvalues of meta-path by random walk to measure the relevancy between the head and end nodes in a path instance, and determine the weight of each meta-path feature by supervised learning. At last, the check-in probability to each candidate POI is calculated to achieve the recommendation result. To evaluate our algorithm, we implement experiments on three real datasets. The results prove that our method can effectively alleviate the sparsity problem and improve the accuracy of recommendation system.

Keywords location-based social networks; heterogeneous networks; POI recommendation; meta-path; data mining; social media

---

收稿日期:2014-07-17;**在线出版日期**:2015-07-16.**本课题得到国家自然科学基金**(61272531,61202449,61272054,61370207,61370208,61300024,61320106007,61472081)、**国家"九七三"重点基础研究发展规划项目基金**(2010CB328104)、**国家"八六三"高技术研究发展计划项目基金**(2013AA013503)、**高等学校博士点学科专项科研基金**(20110092130002)、**江苏省科技计划项目基金**(SBY2014021039-10)、**江苏省网络与信息安全重点实验室**(BM2003201)、**计算机网络和信息集成教育部重点实验室(东南大学)**(93K-9)**资助**.**曹玖新**,**男**,1967**年生**,**博士**,**教授**,**博士生导师**,**中国计算机学会**(CCF)**会员**,**主要研究领域为服务计算、网络安全、社会计算**. E-mail: jx.cao@seu.edu.cn. **董 羿**,**男**,1991**年生**,**硕士研究生**,**主要研究方向为社会计算**. **杨鹏伟**,**男**,1989**年生**,**硕士研究生**,**主要研究方向为社会计算**. **周 涛**,**男**,1989**年生**,**博士研究生**,**主要研究方向为社会计算**. **刘 波**,**女**,1975**年生**,**博士**,**副教授**,**中国计算机学会**(CCF)**会员**,**主要研究方向为普适计算、社会计算**.

1 引 言

近年来,卫星通信、GPS 设备、无线传感器网络、物联网通信等移动互联网技术不断进步,人们日常使用的智能终端提供的位置定位功能越来越精确、便捷,如智能手机、平板电脑等.在此背景下,基于位置的社交网络[1](Location-Based Social Networks,LBSN)服务得到迅速发展,且受到广大用户的喜爱,如国外的 Foursquare、Gowalla,国内的街旁、嘀咕等.LBSN 与传统在线社交网络(Online Social Networks,OSN)的主要区别在于其增加了地理位置信息,用户可以对当前访问的兴趣点(如餐厅、电影院、旅游景点等)签到,并与好友分享自己的签到信息.在 LBSN 中,兴趣点推荐服务旨在为用户推荐一些新的可能感兴趣的位置,促使用户更好地了解其所在城市,提高平台的用户体验.

目前个性化推荐技术在不同应用领域受到工业界和学术界的广泛关注[2],如电子商务网站为用户推荐商品[3]、视频网站为用户推荐电影[4].随着社交网络和智能终端的普及,基于社交网络的推荐系统[5]和移动推荐系统[6]的研究得到越来越多的关注.推荐技术较好地解决了互联网信息过载的问题,方便用户快速有效地找到自己需要的内容、信息和服务.LBSN 同时包含用户和兴趣点节点,并且兴趣点规模巨大,而每个用户访问的兴趣点数量很有限,所以签到数据极度稀疏.LBSN 的异质性和签到数据的极度稀疏性给兴趣点推荐问题带来了新的挑战.

基于以上考虑,本文提出了一种基于元路径的兴趣点推荐算法.首先根据 LBSN 结构特征构建带权异构网络模型,其次引入元路径来描述节点间不同类型关联关系,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,并利用逻辑回归方法获得特征的权值,最后计算特定用户将来在各候选兴趣点的签到概率从而生成推荐列表.针对用户签到数据稀疏性问题,本文引入元路径概念,通过好友关系和兴趣点相关属性来增加元路径数量,从而丰富有效数据,缓解稀疏性问题,提高了推荐的精确度,且时间复杂度较低.

本文第 2 节介绍 LBSN 中兴趣点推荐技术的相关工作;第 3 节介绍针对 LBSN 构建的异构网络模型;第 4 节提出异构网络中基于元路径的兴趣点推荐算法;第 5 节介绍实验的数据集和方案设计;第 6 节对算法进行横、纵向的对比分析,验证算法的有效性;第 7 节总结并探讨将来的研究工作.

2 相关工作

近年来,LBSN 中的推荐技术备受国内外学者的关注.兴趣点推荐所采用数据集可以分为基于 GPS 的轨迹数据和 LBSN 中的签到数据.文献[7-12]均基于 GPS 轨迹数据进行推荐研究,其首要工作就是从轨迹数据中挖掘出兴趣点.而 LBSN 中的签到数据已经包含带有语义信息的兴趣点,不再需要兴趣点挖掘,并且具有丰富的用户、兴趣点属性以及好友关系,因此受到研究者的青睐.文献[13-19]属于基于签到数据的兴趣点推荐研究.

推荐技术发展至今已取得不少成果,从技术方法角度主要可分为以下 3 个方面:

(1)基于上下文的推荐.其特点在于利用用户属性(如性别、爱好等)和兴趣点属性(如类别、标签等)构建推荐模型.Bao 等人[13]将推荐系统分成线下、线上模块,线下模块利用用户、兴趣点属性准确挖掘用户个人偏好和本地专家信息,线上模块快速选取推荐候选兴趣点集合和计算兴趣点评分,从而进行推荐.

(2)基于链路分析的推荐.特点在于从网络拓扑结构角度挖掘节点关联性.Long 等人[7]在分析了签到和好友之间的关系后,提出基于 HITS 的兴趣点推荐算法.Cheng 等人[14]将个人签到序列看作马尔科夫模型并考虑地理区域限制,实现短期的兴趣点推荐.李雯等人[15]借鉴马尔可夫模型对移动对象的历史轨迹建模,并结合对象的运动趋势综合推荐.

(3)基于协同过滤的推荐.特点在于从用户历史签到记录角度进行推荐,可进一步分为基于记忆的推荐和基于模型的推荐.基于记忆的推荐就是传统的基于用户的协同过滤和基于项的协同过滤.Zheng 等人[8]首次运用基于用户的协同过滤方法进行兴趣点推荐,之后在文献[9-10]中参考基于项的协同过滤思想提出基于兴趣点的协同过滤推荐算法.Ye 等人[11]认为用户签到受个人偏好、好友关系和兴趣点距离三方面影响,综合基于用户的协同过滤、基于好友的协同过滤和基于地理信息的推荐方法提出混合协同过滤算法,大大提高了推荐精度.基于模型的推荐通过构建用户评分生成模型实现评分

预估以完成推荐. Liu 等人[16]认为用户评分由用户属性、兴趣点属性、兴趣点间距离共同决定，在此基础上构建贝叶斯图模型来预估用户对新兴趣点的喜好程度. Hu 等人[17]首次将时空话题模型运用于兴趣点推荐，将用户属性与签到的时空规律相结合来提高推荐精度.

以上推荐技术均基于用户历史记录数据，数据的稀疏性对推荐精度都具有一定程度的影响. 有些学者专注于解决数据稀疏性问题的研究，典型的方法有矩阵的奇异值分解[18-19]，在此基础上，Lian 等人[12]对其改进并提出加权矩阵分解算法. 文献[20]运用 PageRank 思想以及改进的 cosine 相似度度量方法来缓解数据稀疏性.

本文提出一种基于元路径的兴趣点推荐算法. 其优势在于，不仅考虑了用户与兴趣点的直接关系，还考虑了网络中各类节点之间的潜在关联性. 在兼顾用户的签到记录、好友关系和兴趣点相关的基础上，本文引入元路径来刻画节点间关系类型，利用元路径特征计算用户签到概率. 该算法通过考虑多种元路径来丰富目标用户的有效数据，从而缓解 LBSN 中数据极度稀疏的问题，因此可以较好地提高推荐效果.

3　LBSN 异构网络模型

异构网络是指存在不同性质节点的网络，一般定义为三元组 $G\langle V,E,A\rangle$，其中 V 是节点集合，E 是边的集合，A 是节点类型集合. LBSN 是典型的异构网络，包括两类节点：用户和兴趣点节点以及三类边：用户-用户边、用户-兴趣点边以及兴趣点-兴趣点边. 其中用户-用户边表示用户之间的好友关系；用户-兴趣点边表示用户对兴趣点的访问签到；兴趣点-兴趣点边表示兴趣点之间的相关性. 其网络结构示意图如图 1 所示.

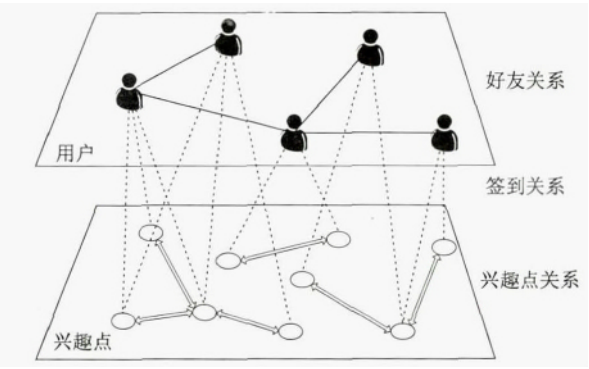


图 1　LBSN 的用户、兴趣点结构图

本文将 LBSN 建模为带权值的异构网络，用四元组 $G\langle U,L,E,W\rangle$ 表示，其中：

(1) $U=\{u_1,u_2,\cdots,u_n\}$ 为用户节点集合.

(2) $L=\{l_1,l_2,\cdots,l_m\}$ 为兴趣点节点集合.

(3) $E=E_{UU}\cup E_{UL}\cup E_{LL}$ 是网络中所有边的集合，本文认为 LBSN 中所有边都是无向的，其中，$E_{UU}=\{(u_i,u_j)\mid u_i\in U,u_j\in U\}$ 表示用户-用户边集合，即好友关系，$E_{UL}=\{(u_i,l_j)\mid u_i\in U,l_j\in L\}$ 表示用户-兴趣点边集合，即用户签到行为，$E_{LL}=\{(l_i,l_j)\mid l_i\in L,l_j\in L\}$ 表示兴趣点-兴趣点边集合，即兴趣点相关.

(4) $W=W_{UU}\cup W_{UL}\cup W_{LL}$ 是网络中边的权值集合，其中，$W_{UU}=\{w(u_i,u_j)\mid(u_i,u_j)\in E_{UU}\}$ 为用户-用户边的权值集合，$w(u_i,u_j)$ 表示用户之间的亲密程度，$W_{UL}=\{w(u_i,l_j)\mid(u_i,l_j)\in E_{UL}\}$ 为用户-兴趣点边的权值集合，$w(u_i,l_j)$ 表示用户对兴趣点的偏好程度，$W_{LL}=\{w(l_i,l_j)\mid(l_i,l_j)\in E_{LL}\}$ 为兴趣点-兴趣点边的权值集合，$w(l_i,l_j)$ 表示兴趣点之间的相关度.

在异构网络 LBSN 的三类边中，兴趣点-兴趣点边所体现的兴趣点相关性较难度量. 本文从用户行为角度考虑，认为经常短时间内被连续先后签到的两兴趣点在现实世界中是相互关联的. 比如，一个学校的多个校区之间距离可能很远，但是经常会被该校师生连续先后访问. 兴趣点相关的定义如下.

定义 1.　兴趣点相关. 给定时间阈值 Δt，若存在用户对兴趣点 l_i 和 l_j 进行连续签到且签到时间间隔小于 Δt，则认为兴趣点 l_i 和 l_j 是相关的.

本文采取如下策略给定三类边的权值. 对于用户-用户边，即用户之间存在好友关系，权值均为 1. 对于用户-兴趣点边和兴趣点-兴趣点边的权值，考虑使用多种方法计量，并在实验部分确定最优方法，包括：

(1) 计数计量. 用户-兴趣点边的权值为该用户对该兴趣点的签到总次数；兴趣点-兴趣点边的权值为所有用户的签到记录中，两兴趣点在时间阈值内被连续先后访问的总次数.

(2) 二值计量. 用户-兴趣点边，即用户、兴趣点之间有过签到，权值均为 1；兴趣点-兴趣点边，即两兴趣点间存在相关性，权值均为 1.

(3) 对数计量. 对计数计量法中的总次数求对数作为权值.

4 基于元路径的兴趣点推荐算法

4.1 元路径

元路径(meta-path)主要用来描述异构网络中任意两节点间的不同路径类型[21]. 在 LBSN 中,两个节点之间可以存在不同类型不同长度的路径. 比如,两个用户之间的路径可以存在以下形式:用户-用户-用户、用户-兴趣点-用户以及用户-兴趣点-兴趣点-用户等. 不同路径代表的物理意义不同,所体现出的节点间关联程度也不同. 这样的路径可以用元路径来描述,元路径的相关定义如下.

定义 2. 元路径. 在异构网络 $G\langle V,E,A\rangle$ 中,元路径定义为如下形式的路径 $P=A_1\xrightarrow{R_1}A_2\xrightarrow{R_2}\cdots\xrightarrow{R_n}A_{n+1}$,其中 $A_i\in A$,R_i 表示不同节点类型间的关系.

在 LBSN 中,节点类型分为用户节点 U 和兴趣点节点 L,关系类型 $R_i\in\{U-U,U-L,L-L\}$,由于认为 LBSN 中所有边都是无向的,所以关系 $U-L$ 与关系 $L-U$ 是一致的,反映的都是用户签到行为.

定义 3. 实例路径. 在异构网络 $G\langle V,E,A\rangle$ 中,对于元路径 $P=A_1\xrightarrow{R_1}A_2\xrightarrow{R_2}\cdots\xrightarrow{R_n}A_{n+1}$,若存在真实路径 $p=(v_1,v_2,\cdots,v_{n+1})$,$v_i\in V$,其中对任意 i,节点 v_i 的类型为 A_i,v_i 与 v_{i+1} 之间关系类型为 R_i,那么路径 p 称为元路径 P 的一条实例路径,所有这样的真实路径称为元路径 P 的实例路径集 P'.

元路径本质上描述的是两节点之间路径的类型,不同的元路径体现两节点间不同类型的关联性. 为了方便描述,本文给出元路径特征值的概念,定义如下.

定义 4. 元路径特征值. 若网络中节点 v_1 与 v_{n+1} 间有元路径 $P=A_1\xrightarrow{R_1}A_2\xrightarrow{R_2}\cdots\xrightarrow{R_n}A_{n+1}$,则元路径 P 的特征值为其所体现出的节点 v_1 与 v_{n+1} 间的关联程度,用 $Eig(P)$ 表示.

4.2 基于元路径的兴趣点推荐算法

兴趣点推荐是为用户推荐未曾访问过但可能感兴趣的兴趣点. 对于该问题,本文提出基于元路径的兴趣点推荐算法,基本思想为:首先确定始于用户节点类型且终于兴趣点节点类型的元路径集,以此作为预测的特征集;然后定义元路径特征值的计算方法,进而使用监督学习获得不同元路径特征的相应权值,据此计算出用户将来在各兴趣点的签到概率从而生成推荐列表.

4.2.1 LBSN 中元路径集的确定

基于 LBSN 的网络模型,确定始于用户节点类型且终于兴趣点节点类型的元路径集,如表 1 所示. 根据小世界现象[22]和三度影响力理论[23]可以推断,长度大于 3 的元路径体现的关联关系非常弱,所以本文只考虑路径长度为 2、3 的元路径. 其中,$U-U-L$ 体现了基于用户的协同过滤思想,$U-L-L$ 体现了基于项的协同过滤思想. 以往的研究大多基于这两类路径,当用户签到很少时,有效数据的不足将大大降低传统推荐算法的准确度. 本文考虑了 3 跳以内的另外 4 条元路径,可以通过好友关系和兴趣点相关大大丰富有效数据,从而缓解数据稀疏性.

表 1 用户-兴趣点之间的元路径集

元路径	元路径体现的语义
$U-L-L$	用户可能喜欢与访问过的兴趣点相关的兴趣点
$U-U-L$	用户可能喜欢好友喜欢的兴趣点
$U-L-U-L$	用户可能喜欢与其有相同签到偏好的用户喜欢的兴趣点
$U-U-U-L$	用户可能喜欢与其有相同交友偏好的用户喜欢的兴趣点
$U-L-L-L$	用户可能喜欢与访问过的兴趣点多层相关的兴趣点
$U-U-L-L$	用户可能喜欢与好友访问过的兴趣点相关的兴趣点

4.2.2 元路径特征值的计算

基于给定的元路径集,本文提出考虑边上具有权值情况下的元路径特征值计算方法. 给定 LBSN 网络 $G\langle U,L,E,W\rangle$,针对元路径 P 及其实例路径集 P',元路径 P 的特征值为所有实例路径体现的关联程度之和,计算公式为

$$Eig(P)=\sum_{p\in P'}cor(p) \tag{1}$$

其中 $cor(p)$ 为实例路径 p 体现出的节点间关联度.

对实例路径 $p=(a_1,a_2,\cdots,a_{n+1})$,$a_1\in U$ 为用户节点,$a_{n+1}\in L$ 为兴趣点节点,其他 a_i 为实例路径中的 1 个中间节点. 路径 p 体现的节点间关联度 $cor(p)$ 基于随机游走思想计算[24],假设一个粒子从节点 a_1 出发,在网络中随机游走,定义 $cor(p)$ 为粒子按照实例路径 p 游走到节点 a_{n+1} 的概率. 随机游走中每一跳游走的概率都认为是相互独立的,因此粒子按照实例路径 p 游走的概率等于各跳的概率之积,$cor(p)$ 的计算公式如下:

$$cor(p)=\prod_{i=1}^{n}Pro(a_i,a_{i+1}) \tag{2}$$

其中 $Pro(a_i,a_{i+1})$表示随机游走过程中从节点 a_i 直接到节点 a_{i+1} 的概率.

在带权值的异构网络中，本文定义 $Pro(a_i,a_{i+1})$ 的计算公式如下：

$$Pro(a_i,a_{i+1})=\frac{w(a_i,a_{i+1})}{\sum_{v\in N(a_i)}w(a_i,v)} \tag{3}$$

其中 $N(a_i)$表示是节点 a_i 的邻居且与节点 a_{i+1} 的类型一致的节点集合.

4.2.3　兴趣点推荐算法

根据之前确定的用户-兴趣点之间的元路径集 $M=\{P_1,P_2,\cdots,P_6\}$，任意用户兴趣点对 (u,l) 都可以通过元路径特征值计算方法得到一个特征值向量 $\vec{\alpha}=\{Eig(P_1),Eig(P_2),\cdots,Eig(P_6),1\}$，其中常数 1 仅为了使用逻辑分布公式而添加. 之后，基于逻辑回归的逻辑分布公式可知，用户 u 去兴趣点 l 签到的概率为

$$\rho=\frac{1}{1+e^{-\vec{\alpha}\vec{\theta}}} \tag{4}$$

其中向量 $\vec{\theta}$ 为特征值向量中各个特征在概率预测中相应的权值，可以基于训练集运用监督学习方法获得. 在此基础上依据概率大小选取 TOP-N 的兴趣点集作为推荐结果.

综上，基于元路径的兴趣点推荐算法描述如算法 1 所示. 其中，L_C 表示为用户 u_0 推荐的候选兴趣点集合，L_R 表示为用户 u_0 最终推荐的兴趣点集合.

算法 1.　基于元路径的兴趣点推荐算法.

输入：元路径集 $M=\{P_1,P_2,\cdots,P_6\}$，LBSN 网络 $G\langle U,L,E,W\rangle$，目标用户 u_0

输出：兴趣点推荐结果 L_R

```
1. L_C = initializeCandidateSet(u_0);
2. FOREACH POI l∈L_C Do
3.   i=0;
4.   FOREACH meta-path P∈M Do
5.     eig_P=0;
6.     P'=findInstancePathSet(u_0,P);
7.     FOREACH instance-path p∈P' Do
8.       cor_p=ComputeCor(p);
9.       eig_P+=cor_p;
10.    END FOR
11.    α[i++]=eig_P;
12.  END FOR
13.  α[i]=1;
14.  ρ_l=e^{αθ}/(e^{αθ}+1);
15. END FOR
16. L_R=TOP-N(L_C,ρ);
17. RETURN L_R;
```

算法第 1 行：依据给定元路径集初始化目标用户的推荐候选兴趣点集合；第 6 行：在给定元路径 P 的基础上，确定对应的实例路径集；第 7～10 行：通过计算每个实例路径反映的关联度获得元路径 P 的特征值；第 14 行：根据学习得到的权值向量 $\vec{\theta}$ 计算签到概率；第 16～17 行，从候选集中选取概率最大的 N 个兴趣点作为最终推荐结果.

假设网络 $G\langle U,L,E,W\rangle$ 有 n 个用户，r 个兴趣点，m 条边，基于元路径的兴趣点推荐算法时间复杂度分析如下：算法时间复杂度主要体现在 3 层 FOR 循环，第 1 层时间复杂度为 $O(r)$，第 2 层时间复杂度为 $O(c)$，c 为元路径集大小，本文为常数 6，第 3 层时间复杂度为 $O(m)$，所以，该算法总的时间复杂度为 $O(rcm)$，即 $O(rm)$.

5　实验设计

5.1　数据集

本文实验部分将在 3 个数据集上验证算法的有效性，这 3 个数据集是来源于两个典型 LBSN 服务平台(Foursquare 和 Gowalla)的真实签到数据：

(1) Gowalla 公开数据集①. 该数据集为斯坦福大学研究人员在 Gowalla 平台抓取的真实签到数据. 该数据集中兴趣点和用户分布于世界各地，未经过处理公开在互联网上.

(2) Foursquare 全球数据集(FoursquareAll). 该数据集是本文项目组通过公开的 API 抓取的，是遍布全球的 Foursquare 签到数据的随机采样.

(3) Foursquare 纽约数据集(FoursquareNY). 所有的签到兴趣点都在纽约范围内. 对全球数据集过滤得到在纽约范围内的签到记录，再通过已有用户集合进一步抓取签到记录而获得.

每个数据集都包括好友关系和签到记录. 好友关系由若干条边组成，每条边的形式为：〈用户 ID，好友用户 ID〉；每条签到记录的形式为：〈签到序号，用户 ID，兴趣点 ID，签到时间〉. 在获得数据集之后，对短时间内进行了大量签到的虚假行为进行了过滤. 预处理后，3 个数据集的基本信息如表 2 所示. 其中，签到密度计算公式如式(5)所示：

$$\text{签到密度}=\frac{\text{签到矩阵中非零元素个数}}{\text{签到矩阵中元素总个数}} \tag{5}$$

① https://snap.stanford.edu/data/loc-gowalla.html

表 2 数据集基本信息

数据集	用户数	兴趣点数	好友边数	签到数	签到密度
Gowalla	96283	1016059	837988	4516034	4.6E-5
FoursquareAll	109030	1278774	2680516	12733427	9.1E-5
FoursquareNY	23781	77950	517960	1570325	8.5E-4

可见 3 个数据集的签到密度都较低，反映了 LBSN 中兴趣点推荐的极度稀疏性问题.

5.2 训练集与测试集的划分

本实验在获得数据集后，需要划分训练集和测试集，从训练集中通过逻辑回归学习方法获得推荐算法的权值向量 $\vec{\theta}$，然后基于测试集评价推荐算法的效果. 首先根据签到时间先后将签到记录分为 3 个集合：T_1、T_2 和 T_3，使它们数量比约为 8∶1∶1. 此外，实验中假设用户的好友关系是不变的. 生成训练集和测试集的方法如下：

(1) 训练集. 基于签到数据 T_1 和好友关系数据构建 LBSN 网络，对于每个用户确定其通过元路径可达的兴趣点集合. 然后在 T_2 中查找该用户是否对各兴趣点进行了签到，若存在签到，则生成正例样本，否则生成负例样本.

(2) 测试集. 与获取训练集方法类似，基于签到数据 T_1、T_2，对用户确定可达兴趣点集合. 若用户在 T_3 中存在签到，则生成测试样本.

基于以上划分，计算所有正负例、测试样本的元路径特征值向量，获得训练集正负例样本的信息包括：用户 ID、兴趣点 ID、元路径特征值向量以及正负例标识.

6 实验结果分析

6.1 评价指标

为了评价兴趣点推荐算法的效果，本文选用推荐问题通用的指标：准确率、召回率. 准确率指推荐结果中用户将来真正去的数量占推荐总数的比例，反映了推荐的准确性. 召回率指推荐结果中用户将来真正去的数量占用户将来访问兴趣点总量的比例，反映了推荐的全面性. 对用户 u 进行兴趣点推荐的准确率和召回率定义如下：

$$\text{准确率} = \frac{|R(u) \cap T(u)|}{|R(u)|} \tag{6}$$

$$\text{召回率} = \frac{|R(u) \cap T(u)|}{|T(u)|} \tag{7}$$

其中，$R(u)$ 为对用户 u 进行推荐的兴趣点集合，$T(u)$ 为用户 u 在测试集上实际的签到集合. 准确率和召回率相互制约，综合利用二者可以对预测结果做出客观的评价.

6.2 不同推荐算法的对比

为验证本文提出的基于元路径的兴趣点推荐算法的推荐效果以及引入兴趣点相关的合理性，本实验将与文献[11]中的多个典型兴趣点推荐算法进行比较和分析，各对比算法介绍如表 3.

表 3 各对比算法描述列表

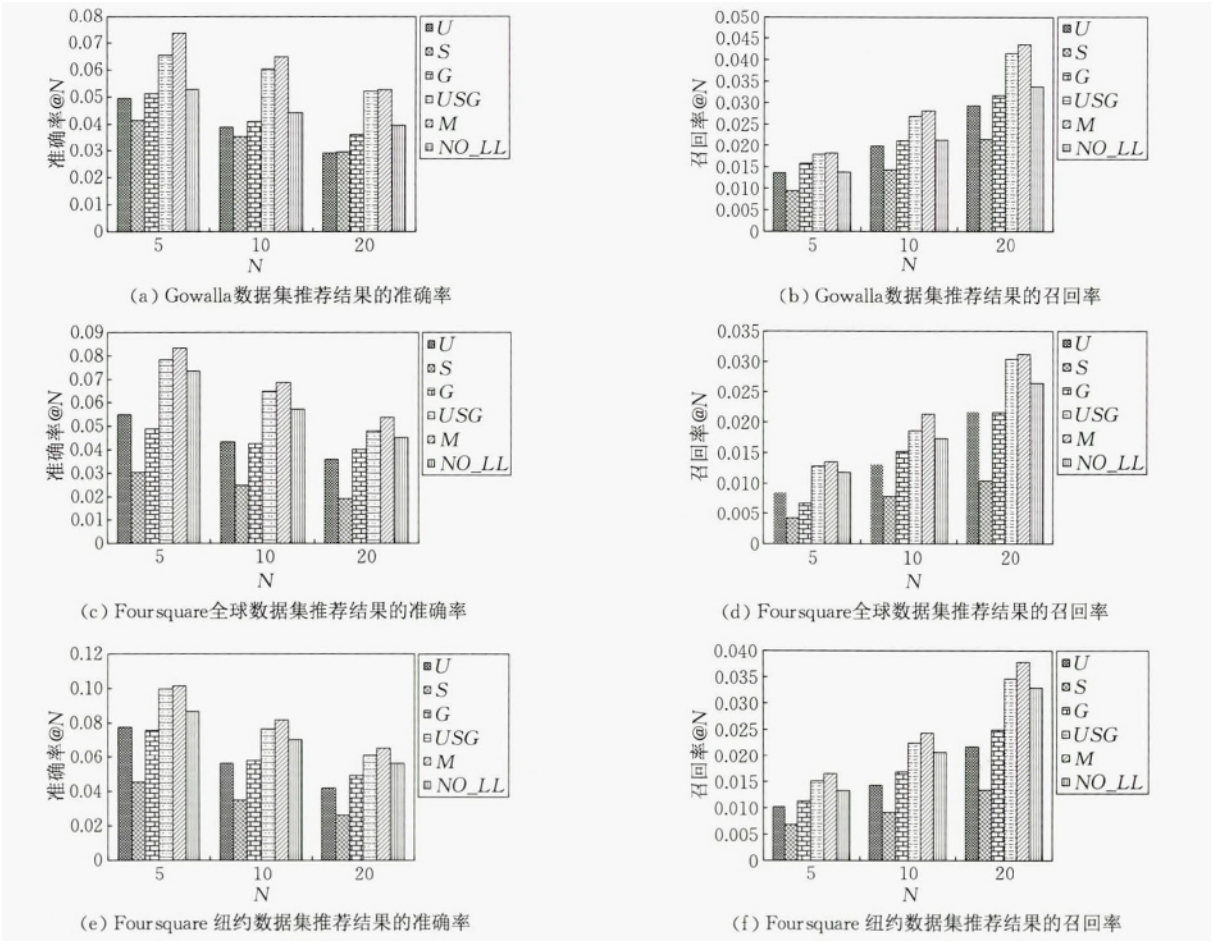
算法	算法描述
User-based CF(U)	基于签到记录计算用户相似度，再运用传统协同过滤进行推荐
Friend-based CF(S)	基于用户共同好友和签到计算用户相似度，再运用传统协同过滤进行推荐
GI-based $R(G)$	用户签到兴趣点距离符合幂律分布，从地理位置影响角度计算签到的概率
Mixed $R(USG)$	同时考虑用户偏好、好友关系和地理因素，以上 3 种方法得到的签到概率的加权和即为最终所求概率
Meta-path based $R(M)$	本文提出的基于元路径的兴趣点推荐算法
Meta-path based R without $L-L(NO\_LL)$	不考虑兴趣点相关的基于元路径的兴趣点推荐算法

其中 U、S、G 和 USG 均为文献[11]涉及算法. 另外，根据后文实验结果分析，M 和 NO_LL 选取最优参数设定：将兴趣点相关定义中的时间阈值设为 4 h，使用二值计量方法来表示 LBSN 中用户-兴趣点边和兴趣点-兴趣点边的权值.

针对 3 个数据集，在进行相同预处理的前提下，本文实现并比较了以上 6 个推荐算法，推荐结果的准确率和召回率分别如图 2 所示. 其中(a)(b)对应 Gowalla 数据集，(c)(d)对应 Foursquare 全球数据集，(e)(f)对应 Foursquare 纽约数据集. 图 2 中展示了各算法 TOP-N(N=5,10,20)的推荐性能.

由图 2 可以看出，无论 N 取何值，本文所提出的基于元路径的推荐算法在 3 个数据集上准确率、召回率普遍优于其他推荐算法. 此外，该算法相比于不考虑兴趣点相关的情况效果提升显著，说明根据用户行为给出的兴趣点相关定义符合实际，对推荐效果有较大的影响.

比较各算法在 Foursquare 不同的数据集上的推荐结果可以发现，相比于纽约数据集，所有算法在全球数据集上的推荐精度都有一定程度的降低，如表 4 所示. 本文提出的基于元路径的推荐算法推荐精度降低比例在 15%～20%之间，而其他算法的推荐精度普遍降低 20%以上. 由此可见，本文所提出的基于元路径的兴趣点推荐在一定程度上缓解了数据稀疏性对推荐结果带来的不利影响，更适用于 LBSN 这种极度稀疏的复杂网络.



(a) Gowalla数据集推荐结果的准确率　(b) Gowalla数据集推荐结果的召回率

(c) Foursquare全球数据集推荐结果的准确率　(d) Foursquare全球数据集推荐结果的召回率

(e) Foursquare 纽约数据集推荐结果的准确率　(f) Foursquare 纽约数据集推荐结果的召回率

图 2　3 个数据集的算法对比结果

表 4　各推荐算法在 Foursquare 不同数据集上的准确率

推荐算法	N	Foursquare 纽约	Foursquare 全球	准确率下降比例/%
U	5	0.0775	0.0548	29.3
	10	0.0562	0.0433	23.0
	20	0.0419	0.0359	14.3
S	5	0.0453	0.0303	33.1
	10	0.0348	0.0248	28.7
	20	0.0261	0.0192	26.4
G	5	0.0757	0.0488	35.5
	10	0.0579	0.0425	26.6
	20	0.0491	0.0401	18.3
USG	5	0.0996	0.0783	21.4
	10	0.0764	0.0647	15.3
	20	0.0612	0.0479	21.7
M	5	0.1014	0.0833	17.9
	10	0.0817	0.0687	15.9
	20	0.0653	0.0536	17.9

6.3　本算法中不同参数的对比

6.3.1　兴趣点相关定义中时间阈值的选取

时间阈值的大小决定着兴趣点之间相关性的判断，进而影响推荐算法的推荐结果. 该部分实验基于 Foursquare 纽约数据集，权值采用二值计量，观测在时间阈值取不同值时基于元路径的推荐算法的效果，结果如图 3 所示. 在另外两个数据集上也有类似结果.

由图 3 可知，当时间阈值为 4 h 左右时，可以同时获得最高的推荐精度和召回率，即最准确地反映两个兴趣点之间的相关性.

6.3.2　权值计量方法的选取

在 LBSN 网络模型中，用户-兴趣点边和兴趣点-兴趣点边的权值计算方法对推荐算法的效果也有一定的影响. 该部分实验在 Foursquare 纽约数据集上，时间阈值设为 4 h，运行推荐算法，观测采用不同的权值计量方法时的推荐效果，如图 4 所示. 在另外两个数据集上也有类似结果.

由图 4 可知，二值表示权值的方法获得了最佳的推荐结果. 二值表示权值本质上其实是忽略了边上的权值信息，但反而更好地刻画了复杂网络并提高了推荐精度. 目前的研究[25-26]发现，在复杂网络的链路预测问题上使用权值信息并不能带来预测精度的显著提升，而本文的结果和其研究是相符的. 此外，本文还使用基于好友的协同过滤对 3 种权值计量方法进行了实验比较，实验结果相同，也是二值表

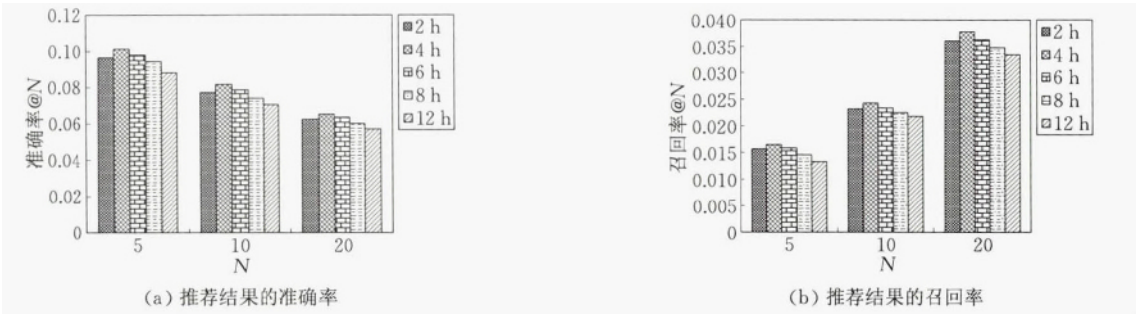


图 3 Foursquare 纽约数据集上，兴趣点相关定义中不同时间阈值的推荐结果对比

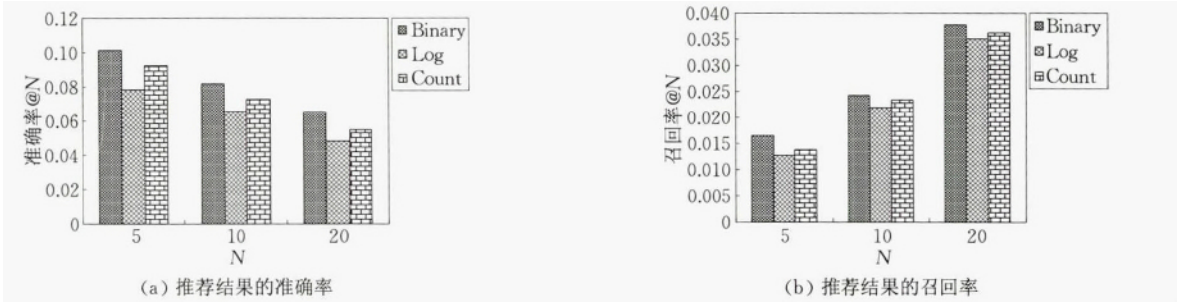


图 4 Foursquare 纽约数据集上，不同权值计量方法的推荐结果对比

示权值的方法获得了最佳推荐精度. 二值计量相比于其他两种计量方法，实质是削弱了强连接的影响，间接表现出复杂网络中弱连接的作用.

7 总 结

本文针对 LBSN 中的兴趣点推荐问题，提出基于元路径的兴趣点推荐算法. 将 LBSN 视为异构网络，引入元路径描述用户与兴趣点节点间的关联性，并给出元路径特征值的定义及其计算方法，再通过逻辑回归方法学习各特征权重，最后计算用户将来在各兴趣点的签到概率从而生成推荐列表. 实验表明，本文提出的算法在 LBSN 真实数据集上相比于其他推荐算法有更高的准确率和召回率，有效地缓解了数据的稀疏性问题.

将来的工作主要在以下几个方面进行. 首先，我们将考虑如何有效地运用数据集中的时间因素，挖掘用户行为的时空模式，以提高推荐效果. 其次，我们将在推荐兴趣点时考虑用户所处的情景上下文信息，如季节、天气等，使得推荐结果更加人性化.

致 谢 审稿专家和编辑提出了宝贵意见和建议，在此表示感谢！

参 考 文 献

[1] Zheng Y. Location-based social networks: Users//Zheng Y, Zhou X eds. Computing with Spatial Trajectories. New York: Springer, 2011: 243-276

[2] Adomavicius G, Tuzhilin A. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Transactions on Knowledge and Data Engineering, 2005, 17(6): 734-749

[3] Schafer J B, Konstan J, Riedl J. Recommender systems in e-commerce//Proceedings of the 1st ACM Conference on Electronic Commerce. Denver, USA, 1999: 158-166

[4] Miller B N, Albert I, Lam S K, et al. MovieLens unplugged: Experiences with an occasionally connected recommender system//Proceedings of the 8th International Conference on Intelligent User Interfaces. Miami, USA, 2003: 263-266

[5] Li Shan-Tao, Xiao Bo. Recommendation system based on social network. Software, 2013, 34(12): 41-45(in Chinese)
(李善涛，肖波. 基于社交网络的信息推荐系统. 软件，2013，34(12)：41-45)

[6] Meng Xiang-Wu, Hu Xun, Wang Li-Cai, et al. Mobile recommender systems and their applications. Journal of Software, 2013, 24(1): 91-108(in Chinese)
(孟祥武，胡勋，王立才等. 移动推荐系统及其应用. 软件学报，2013，24(1)：91-108)

[7] Long X, Joshi J. A hits-based POI recommendation algorithm for location-based social networks//Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. Niagara, CAN, 2013: 642-647

[8] Zheng Y, Zhang L, Ma Z, et al. Recommending friends and locations based on individual location history. ACM Transactions on the Web (TWEB), 2011, 5(1): 5-48

[9] Zheng Y, Xie X. Learning location correlation from GPS trajectories//Proceedings of the 2010 Eleventh International Conference on Mobile Data Management (MDM). Kansas City, USA, 2010: 27-32

[10] Zheng Y, Xie X. Learning travel recommendations from user-generated GPS traces. ACM Transactions on Intelligent Systems and Technology (TIST), 2011, 2(1): 2-30

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

[12] Lian D, Zhao C, Xie X, et al. GeoMF: Joint geographical modeling and matrix factorization for point-of-interest recommendation//Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, USA, 2014: 831-840

[13] Bao J, Zheng Y, Mokbel M F. Location-based and preference-aware recommendation using sparse geo-social networking data//Proceedings of the 20th International Conference on Advances in Geographic Information Systems. Redondo Beach, USA, 2012: 199-208

[14] Cheng C, Yang H, Lyu M R, et al. Where you like to go next: Successive point-of-interest recommendation//Proceedings of the 23rd International Joint Conference on Artificial Intelligence. Beijing, China, 2013: 2605-2611

[15] Li Wen, Xia Shi-Xiong, Liu Feng, et al. Location prediction algorithm based on movement tendency. Journal on Communications, 2014, 35(2): 46-53(in Chinese)
(**李雯, 夏士雄, 刘峰等. 基于运动趋势的移动对象位置预测. 通信学报**, 2014, 35(2): 46-53)

[16] Liu B, Fu Y, Yao Z, et al. Learning geographical preferences for point-of-interest recommendation//Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Chicago, USA, 2013: 1043-1051

[17] Hu B, Jamali M, Ester M. Spatio-temporal topic modeling in mobile social media for location recommendation//Proceedings of the 2013 IEEE 13th International Conference on Data Mining (ICDM). Dallas, USA, 2013: 1073-1078

[18] Ma H, Yang H, Lyu M R, et al. Sorec: Social recommendation using probabilistic matrix factorization//Proceedings of the 17th ACM Conference on Information and Knowledge Management. Napa Valley, USA, 2008: 931-940

[19] Zheng V W, Zheng Y, Xie X, et al. Collaborative location and activity recommendations with GPS history data//Proceedings of the 19th International Conference on World Wide Web. Raleigh, USA, 2010: 1029-1038

[20] Yildirim H, Krishnamoorthy M S. A random walk method for alleviating the sparsity problem in collaborative filtering//Proceedings of the 2008 ACM Conference on Recommender Systems. Lausanne, Switzerland, 2008: 131-138

[21] Sun Y, Han J, Yan X, et al. Pathsim: Meta path-based top-k similarity search in heterogeneous information networks. Proceedings of the VLDB (Very Large Data Bases) Endowment, 2011, 4(11): 235-246

[22] Watts D J, Strogatz S H. Collective dynamics of 'small-world' networks. Nature, 1998, 393(6684): 440-442

[23] Christakis N A, Fowler J H. Connected: The Surprising Power of Our Social Networks and How They Shape Our Lives. London: Little, Brown Book Group, 2009

[24] Tong H, Faloutsos C, Pan J Y. Fast random walk with restart and its applications//Proceedings of the 6th International Conference on Data Mining. Washington, USA, 2006: 613-622

[25] Lü L, Zhou T. Link prediction in weighted networks: The role of weak ties. Europhysics Letters, 2010, 89(1): 18001

[26] Wind D K, Morup M. Link prediction in weighted networks//Proceedings of the 2012 IEEE International Workshop on Machine Learning for Signal Processing (MLSP). Reims, France, 2012: 1-6

CAO Jiu-Xin, born in 1967, Ph. D., professor, Ph.D. supervisor. His research interests include service computing, network security and social computing.

DONG Yi, born in 1991, M. S. candidate. His research interest is social computing.

YANG Peng-Wei, born in 1989, M. S. candidate. His research interest is social computing.

ZHOU Tao, born in 1989, Ph. D. candidate. His research interest is social computing.

LIU Bo, born in 1975, Ph. D., associate professor. Her research interests include pervasive computing and social computing.

Background

In recent years, location-based social networks (LBSN) become increasingly popular among young people. In LBSN, users can check in which means sharing their location to their friends. People are pleased to visit some point-of-interests (POIs) and check in there. Recommending POIs is a crucial personalized service which makes it more convenient for users

to find where they want to go. User-based and item-based collaborative filtering algorithms are normally very effective in general recommendation systems, such as recommending films and recommending goods. However, as a user can only visit a few POIs in LBSN, the extreme sparsity of check-in data invalidates traditional collaborative filtering recommendation algorithms.

Therefore, we propose a novel location recommendation algorithm based on meta-path. Firstly, we consider the location-based social network as a weighted heterogeneous network, and introduce the meta-path to describe different relations between nodes. Then the set of meta-paths between user and POI nodes is given on the basis of three degrees of influence. In addition, we define the eigenvalue of meta-path and its calculation method, and determine the weight of every meta-path feature by supervised learning. At last, the probability of checking in to each candidate POI is calculated to achieve the recommendation result.

This work is supported by the National Key Basic Research Program (973 Program) of China under Grant No. 2010CB328104, the National Natural Science Foundation of China under Grant Nos. 61272531, 61202449, 61272054, 61370207, 61370208, 61300024, 61320106007, 61472081, the National High Technology Research and Development Program (863 Program) of China under Grant No. 2013AA013503, the Specialized Research Fund for the Doctoral Program of Higher Education of China under Grant No. 2011009213002, the Jiangsu Provincial Science and Technology Plan Program under Grant No. SBY2014020139-10, the Jiangsu Provincial Key Laboratory of Network and Information Security under Grant No. BM2003201 and the Key Laboratory of Computer network and Information Integration of Ministry of Education of China under Grant No. 93K-9.