

引用格式:冯堂林,何亮.一种基于元路径的社交事件推荐方法[J].微电子学与计算机,2021,38(8):73-79. [FENG T L, HE L. Meta-path based social event recommendation[J]. Microelectronics & Computer, 2021, 38(8): 73-79.]

一种基于元路径的社交事件推荐方法

冯堂林,何 亮

(西安交通大学 计算机学院,陕西 西安 710049)

摘 要: 基于事件的社交网络为用户提供线上线下互动模式的社交服务,用户在线上组织事件,线下参与事件.以往的研究基于因素或基于图实现社交事件推荐,难以充分挖掘社交网络中的语义信息.为解决该问题,提出了基于元路径的社交事件推荐方法,元路径包含丰富的语义信息,是异质信息网络的重要特征.首先构建了一个异质社交网络,接下来利用元路径计算用户与事件之间的相关性矩阵,在此基础上,通过最优化问题求解元路径权重,最后将相关性矩阵和元路径权重线性组合实现面向用户的社交事件推荐.该算法在应用最为广泛的基于事件的社交网络平台—Meetup 网站真实数据集上进行了实验,采用不同度量指标与其它推荐方法进行了对比,实验结果表明本算法优于其它对比方法.

关键词: 基于事件的社交网络;异质信息网络;元路径

中图分类号: TP274

文献标识码: A

文章编号: 1000-7180(2021)08-0073-07

Meta-path based social event recommendation

FENG Tanglin, HE Liang

(School of Computer Science and Technology, Xi'an Jiaotong University, Xi'an 710049, Shaanxi, China)

Abstract: Event based social networks (EBSNs) provides users with online to offline social services. Users can organize events online and participate in them offline. Particularly, social event recommendation is a hot issue in this field. The existing studies have mainly focused on factors-based or graph-based social event recommendation, which can hardly exploit the semantic information in the social networks thoroughly. To address this problem, a meta-path based social event recommendation method is proposed. The meta-path contains rich semantic information, being an important feature of the heterogeneous information networks. Firstly, a heterogeneous network is constructed, and then various candidate meta-paths are utilized to calculate the correlation matrix between the users and events. On this basis, the weight of each meta-path is solved by optimization method. Accordingly, the social event recommendation is realized in terms of the linear combination of correlation matrix and meta-path weights. The algorithm is validated on the real dataset of Meetup, the most popular event-based social network platform. Meanwhile, various metrics are explored to compare the recommendation methods. The experimental results show that the proposed algorithm outperforms the comparative ones.

Key words: event based social networks; heterogeneous information networks; meta-path

1 引言

基于事件的社交网络(Event-based Social Networks, EBSNs)提供了线上组织发布事件,线下参

与事件的新型社交模式,这种同时兼顾线上线下的社交网络吸引了众多参与者加入. Meetup、豆瓣同城等 EBSNs 网站得到了广泛关注.随着用户和事件数量的增多, EBSNs 平台的数据规模日益增长,这

使得用户发掘感兴趣的社交事件变得困难. 如何为用户推荐其感兴趣的社交事件, 提高社交事件的参与度, 成为 EBSNs 研究的热点问题. 以往的社交事件推荐方法主要包括基于因素和基于图的推荐. 基于因素的推荐方法将影响事件推荐的潜在因素作为评分计算模型的输入, 再将输出结果作为事件推荐的评判标准^[1-4]; 基于图的推荐方法将任务转换为节点邻近问题^[5-7]. 这些方法未考虑社交网络的异质特性, 在社交网络语义信息挖掘方面存在着不足.

异质信息网络 (Heterogeneous Information Network, HIN) 是一种含有多种节点类型、多种边类型的复杂网络, 已在相似性搜索、聚类 and 分类等方面得到了广泛应用. EBSNs 的用户社交关系具有显著的异质网络特征, 这些特征为事件推荐提供了新的研究思路. 在应用最为广泛的 EBSNs 平台 Meetup 中, 用户形成了各种线上群组, 每个用户可以根据自己的偏好加入不同群组. 一方面, 同一群组的用户会有相似的兴趣; 另一方面, 用户通过参加社交事件来进行线下交流, 而参加同一事件的用户可能会表现出类似的偏好, 在线上线下社交网络中都有密切交流的人群可能对未来事件有相似的偏好. 因此, 在 Meetup 平台中, 用户、事件和群组等多种角色赋予了社交网络天然的异质特性. 本文尝试为 Meetup 平台构建异质信息网络, 并在此基础上实现面向用户的社交事件推荐.

2 基于元路径的社交事件推荐

2.1 异质社交网络模式构建

在 Meetup 平台中, 主要有用户 (User)、群组 (Group)、事件 (Event)、标签 (Tag)、地理位置 (Location) 等各类信息, 其中标签是描述用户或群组偏好的词组, 用户及群组可以具有多个标签. 地理位置表示用户所在地或线下事件举办地点的经纬度信息, 基于这些信息可建立 Meetup 平台的异质网络模式. 由于标签文本及地理位置信息数据量过大, 直接基于标签及地理位置建立异质网络模式将显著增加计算复杂度, 不利于后续的事件推荐工作, 因此需要首先完成标签及地理位置信息的预处理.

2.1.1 标签聚类

标签经过聚类后可形成若干不同的主题 (Subject), 每个主题包含不同数量的标签, 总体上主题的数量将明显少于标签数量. 标签聚类基于 U-T (user-tag)、U-G (user-group) 和 G-T (group-tag) 三种关系, 假设 UT 表示 U-T 关系邻接矩阵, UG 表示

U-G 关系邻接矩阵, GT 表示 G-T 关系邻接矩阵, 矩阵中的元素取值 1 表示节点间存在连边关系, 0 表示不存在连边关系. 可得到如下矩阵 M .

$$M = UT + UG \cdot GT$$

M 中的 UT 部分表示用户直接包含的标签关系; $UG \cdot GT$ 表示用户已加入群组所包含的隐含标签, 这一部分标签通过 U-G 关系映射到用户. 因此 M 表示新的 U-T 关系矩阵, 其中标签包含用户的直接和隐含标签. M 中的每个元素 m_{ij} 可以解释为用户 u_i 拥有标签 t_j 的频次. 在此基础上, 可对 M 矩阵进行聚类以形成主题.

2.1.2 地理位置聚类

Meetup 平台中, 地理位置以经纬度信息表示, 细粒度的位置信息同样会导致网络稀疏问题, 需要对其进行更高层次的抽象. 对 Meetup 平台数据的分析表明绝大部分群组内社交事件的组织均在同一城市内完成, 因此本文仅考虑城市内的事件推荐. 可将一个城市举办过社交事件活动的地理位置通过聚类方式划分为区域 (Region), 每个特定区域在异质网络中代表一个区域节点实例, 每个地理位置则属于某一个特定区域.

2.1.3 异质网络模式

经过以上聚类预处理后, 可建立 Meetup 平台的异质网络模式, 包含用户 U、事件 E、主题 S、群组 G 和区域 R 等五种节点类型及 U-R、U-E、R-E、U-G、U-S、G-S 和 G-E 等七种边类型, 如图 1 所示.

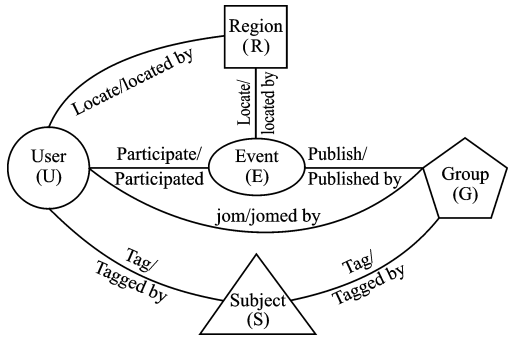


图 1 异质社交网络模式

2.2 社交事件推荐

元路径在异质网络中具有丰富的语义信息, 本小节基于 2.1 节建立的异质网络模式实现社交事件推荐算法. 首先搜索出 Meetup 平台异质网络模式中的可用元路径; 再根据每条元路径计算用户-事件对之间的相似度; 接下来构建损失函数计算出各元路径权重; 在此基础上, 基于元路径组合完成面向用户的社交事件推荐.

2.2.1 元路径搜索

本文实现的社交事件推荐算法基于元路径计算用户与事件的相似性,因此需要首先确定所有可用的元路径,且这些元路径以用户 U 为起始节点,事件 E 为目标节点类型. 对图 1 所示网络模式进行搜索,可获得指定长度的可用元路径. 算法 1 所示为元路径搜索方法,搜索过程从起始节点类型 U 开始,向相邻节点类型逐层进行搜索,寻找在指定最大长度内且目标节点为 E 的路径. 算法输入为异质社交网络模式的邻接矩阵(即图 1 所示网络模式对应的邻接矩阵)、起始节点类型和目标节点类型,预定义的元路径最大长度,输出为最大长度内的所有可用元路径. 其中 $\text{next_list}()$ 函数用于进行下一步迭代,搜索与 first_list 相邻的边, $\text{merge_list}()$ 函数用于将迭代后的结果与 first_list 合并.

算法 1 元路径搜索

```
输入:
     $M_G$ : 异质信息网络模式构成的邻接矩阵
    start_type: 元路径起始节点类型
    end_type: 元路径目标节点类型
     $l_{\max}$ : 元路径最大搜索长度
输出:
    MetaPathSet =  $\{P_1, P_2, \dots\}$ : 可用元路径集合
begin
    MetaPathSet  $\leftarrow$  new_empty_list()
    // 创建元路径空集合
    first_list  $\leftarrow$  new_list(start_type)
    // 创建以 start_type 为起点的链表
    for  $i \leftarrow 1$  to  $l_{\max}$  do
        second_list  $\leftarrow$  next_list(first_list)
        // 调用 next_list 函数, 进行下一步迭代
        com_list  $\leftarrow$  merge_list(first_list, second_list)
        // 调用 merge_list 函数, 将迭代结果与 first_list 合并
        first_list  $\leftarrow$  com_list // 将 com_list 赋值给 first_list
        for  $j \leftarrow 1$  to length(first_list) do
            // 将新得到的元路径合并到 MetaPathSet
            if 元路径 first_list[j] 后一个元素与 end_type 一致 then
                MetaPathSet  $\leftarrow$  MetaPathSet 与 first_list[j] 元路径
            end if
        end for
    end for
end
```

例如,当元路径搜索最大长度 $l_{\max}=4$ 时,可获得长度为 2、3、4 的不同元路径,如 $U-R-E$ 、 $U-G-E$ 、 $U-R-U-E$ 、 $U-E-R-E$ 、 $U-R-U-R-E$ 、 $U-R-U-G-E$ ……等. 长度为 1 的元路径表示用户已参与事件,因此不在推荐时使用.

2.2.2 基于元路径的相关性计算

在获得所有可用元路径后,接下来将根据每条元路径计算出起始节点与目标节点的相关性,该过程基于相关性度量方法 $\text{DPRel}^{[8]}$ 进行. 为了利用 DPRel 度量两个节点之间的相关性,需将异质社交网络转化为一个仅由起始节点对象和目标节点对象组成的带权路径矩阵^[9],如定义 1 所示.

定义 1(带权路径矩阵) 在一个异质信息网络中,元路径 $P = (A_1 A_2 \cdots A_{t+1})$ 所对应的带权路径矩阵表示为:

$$M = W_{A_1 A_2} \times W_{A_2 A_3} \times \cdots \times W_{A_t A_{t+1}}$$

式中, $W_{A_i A_j}$ 是对象类型 A_i 与 A_j 间的邻接矩阵,矩阵元素 $M[x_i, y_j]$ 表示节点 $x_i \in A_1$ 与 $y_j \in A_{t+1}$ 之间元路径 P 的实例数量,且 $M[x_i, y_j] = M[y_j, x_i]$.

如图 2 所示,该 Meetup 异质社交网络片段包含三个用户 $u1$ 、 $u2$ 和 $u3$,两个区域 $g1$ 和 $g2$,三个事件 $e1$ 、 $e2$ 和 $e3$,以元路径 UGE 为例计算带权路径矩阵,计算过程及结果如图 3 所示.

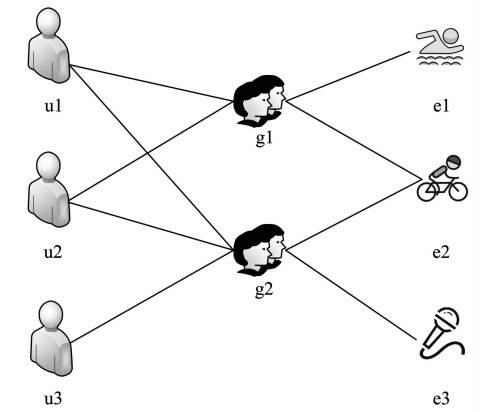


图 2 Meetup 异质社交网络片段

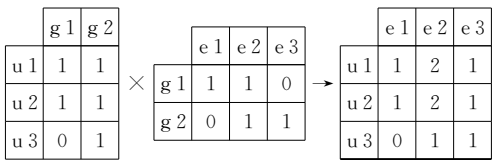


图 3 带权路径矩阵计算过程

在带权路径矩阵基础上可进一步计算 DPRel

相关性,其定义如下:

定义 2(DPRel 相关性) 给定一条元路径 $P = (A_1 A_2 \cdots A_{t+1})$, 节点类型 A_1 与 A_{t+1} 具有不同类型, 在一个只包含节点类型 A_1 与 A_{t+1} 的二分异质网络中, 起始节点 $a_{1i} \in A_1$ 与目标节点 $b_{(t+1)j} \in A_{t+1}$ 之间的 DPRel 相关性计算如下:

$$DPRel(a_{1i}, b_{(t+1)j} | P) = \frac{w(a_{1i}, b_{(t+1)j}) \left(\frac{1}{\deg(a_{1i})} + \frac{1}{\deg(b_{(t+1)j})} \right)}{\frac{1}{\deg(a_{1i})} \sum_j w(a_{1i}, b_{(t+1)j}) + \frac{1}{\deg(b_{(t+1)j})} \sum_i w(a_{1i}, b_{(t+1)j})}$$

式中, $w(a_{1i}, b_{(t+1)j})$ 是带权路径矩阵 $\mathbf{M}$ 中的值 $M[a_{1i}, b_{(t+1)j}]$, 表示起始节点 $a_{1i} \in A_1$ 与目标节点 $b_{(t+1)j} \in A_{t+1}$ 之间特定元路径实例的数量; $\deg(a_{1i})$ 和 $\deg(b_{(t+1)j})$ 分别表示节点 a_{1i} 和 $b_{(t+1)j}$ 在二分异质网络中的度. 基于元路径 P_t 的相关性矩阵记为 $\mathbf{S}(P_t)$.

仍以图 2 所示的 Meetup 异质社交网络片段为例, 图 3 是其对应的带权路径矩阵, 由此可计算出节点 u_1 和 e_1 基于元路径 $P = UGE$ 的相关性:

$$DPRel(u_1, e_1 | P = UGE) = \frac{1 \times (\frac{1}{3} + \frac{1}{2})}{(\frac{1}{3} \times 3 + \frac{1}{2} \times 2)} \approx 0.83$$

2.2.3 元路径权重计算

元路径的权重学习非常重要, 一方面每个元路径都有其特定的语义含义, 另一方面不同元路径对社交事件推荐的作用程度也不尽相同. 例如, 元路径 URE 是指用户所在地与事件举办地位于同一个区域, 而 UGE 则表示用户加入的群组与发布事件的群组是同一个群组. 同时, 这两个元路径对推荐的影响度也不一样. 因此, 有必要通过算法学习出每条元路径的重要性, 将更高的权重分配给对推荐精确度更有帮助的元路径.

为了学习元路径的权重, 本文将权重学习转化为一个最优化问题. 首先, 构建对应的损失函数如下:

$$L(\beta) = \frac{1}{2} \sum_{u_i \in U, e_j \in E} \left\| 1 - \mathbb{I}(u_i, e_j) \sum_{t=1}^T \beta_t DPRel(u_i, e_j | P_t) \right\|_2^2 + \frac{\alpha}{2} \|\beta\|_2^2$$

s. t. $\beta_t \geq 0, \forall t = 1, \dots, T$ (1)

式中, $\beta = \{\beta_1, \beta_2, \dots, \beta_T\}$ 表示 T 个元路径构成的权

重向量; $\beta_t, t = 1, 2, \dots, T$ 表示第 t 个元路径的权重; U 表示所有用户构成的集合; E 表示所有事件构成的集合; u_i 表示起始用户节点; e_j 表示目标事件节点; P_t 表示第 t 个元路径, $DPRel(u_i, e_j | P_t)$ 函数表示基于元路径 P_t 计算起始用户节点 u_i 和目标事件节点 e_j 的 DPRel 相关性, 相关计算方法已在 2.2.2 节介绍; α 表示避免过拟合的正则化参数; $\|\cdot\|$ 表示 L_2 范数, $\mathbb{I}(\cdot)$ 指示函数定义如下.

$$\mathbb{I}(u_i, e_j) = \begin{cases} 1, & (u_i, e_j) \in R_{UE} \\ -1, & \text{otherwise} \end{cases}$$

式中, R_{UE} 表示基于用户—事件对的训练集. 该指示函数表示如果节点 u_i 和 e_j 之间在训练集中存在直接连边, 就返回 1, 否则返回 -1.

构建完损失函数之后, 该问题就可以转换为如下最优化问题:

$$\beta^* = \underset{\beta = \{\beta_1, \beta_2, \dots, \beta_T\}}{\operatorname{argmin}} L(\beta)$$

该最优化问题的思想是, 如果一条连接起始用户节点和目标事件节点的元路径具有高的相关性, 同时该起始用户节点和目标事件节点在测试集中存在直接连边, 那么这条元路径就会被分配更高的权重, 即该元路径对事件推荐更有帮助. 因此, 最优化问题的目标是最大化具有直接连边的用户事件对之间的关系, 最小化没有直接连边的用户事件对之间的关系. 对该最优化问题进行求解, 就可以得到每条元路径的权重 β_t .

取损失函数关于 $\beta_t, t = 1, 2, \dots, T$ 的偏导数, 令其等于 0, 则可得到各权重参数 β_t 的值如下:

$$\beta_t = \frac{\sum_{u_i \in U, e_j \in E} \mathbb{I}(u_i, e_j) DPRel(u_i, e_j | P_t) (1 - \mathbb{I}(u_i, e_j) \sum_{k \neq t}^T \beta_k DPRel(u_i, e_j | P_k))}{\alpha + \sum_{u_i \in U, e_j \in E} \mathbb{I}(u_i, e_j)^2 DPRel(u_i, e_j | P_t)^2}$$

(2)

根据上述等式, 可使用算法 2 求解 β_t , 算法的输入包括异质社交网络, 用于推荐的 T 条元路径, 正则化参数 α , 基于元路径 $P_t, t = 1, 2, \dots, T$ 计算所得的 DPRel 相关性矩阵, 以及收敛精度参数 δ . 为了保证 β_t 在迭代过程中不出现负值, 算法参数每次迭代后的值取 0 与本次迭代计算所得值中的较大者. 算法的收敛条件为参数 β 新值向量与上一次迭代值向量的绝对差小于 δ .

为了更好的表示每个元路径的权重, 算法最后

对权重向量进行了归一化。

算法 2 元路径权重学习

输入:
 $G = (V, E)$:异质社交网络
 $P = \{P_1, P_2, \dots, P_T\}$: T 条元路径的集合
 $S(P_t)$: 基于 DPRel 的第 t 条元路径的相关性矩阵, 其中
 $t = 1, 2, \dots, T$
 α : 正则化参数
 δ : 收敛系数
输出:
 $\beta = \{\beta_1, \beta_2, \dots, \beta_T\}$: 元路径权重集合
Begin
 将 $\beta = \{\beta_1, \beta_2, \dots, \beta_T\}$ 初始化为大于 0 的向量
 do
 $\beta^{last} = \beta$
 按照公式(2) 计算新的 β
 $\beta_t = \max(0, \beta_t), t = 1, 2, \dots, T$
 $\beta = \{\beta_1, \beta_2, \dots, \beta_T\}$ // 得到新的权重向量
 while $|\beta - \beta^{last}| < \delta$
 // 权重向量差小于收敛系数时迭代终止
for $t \leftarrow 1$ to T do // 权重向量归一化
 计算 $w_t = \frac{\beta_t}{\sum_{i=1}^T \beta_i}, \forall t = 1, \dots, T$
end for
end

2.2.4 社交事件推荐

得到元路径权重后,即可计算出用户与事件之间的组合相关性,该组合相关性是依据每条元路径预测的相关性及其对应的权重来综合计算的. 每个用户 - 事件对 u_i 和 e_j 的综合性相关性计算公式如下:

$$DPRel(u_i, e_j | P) = \sum_{t=1}^T w_t DPRel(u_i, e_j | P_t) \tag{3}$$

式中, P 表示所有可用元路径的集合; P_t 表示第 t 条元路径.

那么,所有用户 - 事件对之间的综合相关性矩阵 $S(P)$ 可以用以下公式来表示:

$$S(P) = \sum_{t=1}^T w_t S(P_t) \tag{4}$$

式中, $S(P_t)$ 表示基于元路径 P_t 的 $DPRel$ 相关性矩阵.

接下来,将基于综合相关性矩阵进行用户事件

推荐. 对于每个用户,推荐前 K 个综合相关性最大且与该用户没有直接连边的事件列表.

本算法的时间复杂度主要由元路径权重学习过程决定,由公式(2)可知,每次权重计算过程需遍历一次 $DPRel$ 相关性矩阵,因此每次权重计算的时间复杂度为 $O(|U| * |E|)$,其中 $|U|$ 表示用户数, $|E|$ 表示事件数. 权重个数为 T ,每次迭代的时间复杂度为 $O(T * |U| * |E|)$,设迭代次数为 n ,则元路径权重学习过程的总体时间复杂度为 $O(n * T * |U| * |E|)$.

3 实验结果与分析

3.1 实验设置

本节通过 Meetup 社交网络平台的真实数据集对所实现的基于元路径的社交事件推荐算法进行实验验证,并与基于多元马儿可夫链的 HeteRS^[7]、基于偏好传播的 BPP^[10] 和基于反向随机游走的 RWRR^[11] 等社交事件推荐算法进行对比分析. 实验采用的评估指标包括: AUC (Area Under ROC Curve)、P@n (Precision at Position n)、MAP (Mean Average Precision) 和召回率 (Recall) 等. 实验所用工作站配置为: CPU: Intel Core i5-4590, 主频 3.3GHz, 内存 32GB, Windows Server 64 位操作系统.

实验所用数据集由 Liu 等人^[12] 整理,爬取自 Meetup 社交平台,数据集中包含用户、事件、群组、位置和标签等五类节点信息,用户-事件、用户-标签、用户-位置、用户-群组、事件-位置、群组-标签和群组-事件等七类边信息. 本实验提取了亚特兰大、芝加哥、休斯顿、旧金山和洛杉矶等五个典型城市的数据,并过滤掉了未参加活动的用户及没有参与者的活动,五城市的数据情况如表 1 所示.

表 1 五城市数据集统计信息

	亚特兰大	芝加哥	休斯顿	旧金山	洛杉矶
用户(U)	10 618	13 279	9 872	16 890	7 930
事件(E)	8 734	11 547	10 414	6 551	7 672
群组(G)	2 806	3 228	2 071	5 855	4 917
标签(T)	7 514	11 387	7 175	10 560	9 747
位置(L)	42 936	43 903	38 494	34 904	43 432

如 2.1 节所述,本文首先对五城市数据集的文本标签及地理位置进行了 K-means 聚类预处理,以获得粒度更高的主题及区域节点信息.

3.2 实验结果与分析

为了便于表述,本实验将基于元路径的社交事件推荐算法记为 MetaPath,以下采用不同度量指标对 MetaPath 及对比方法的实验结果进行分析。

(1)AUC

表 2 所示为不同推荐算法在五城市的 AUC 结果.从图中可以看出,与 RRWR 算法相比,MetaPath 算法除了芝加哥的 AUC 相接近外,其他城市的 AUC 都明显较高;与 HeteRS 和 BPP 算法相比,MetaPath 算法在所有城市的 AUC 评价都明显更优;在 AUC 平均值上,Metapath 算法亦优于其他算法。

表 2 各推荐算法五城市 AUC 值

	HeteRS	RRWR	BPP	MetaPath
亚特兰大	0.881 1	0.939 6	0.913 3	0.955 9
芝加哥	0.928 9	0.973 4	0.963 8	0.978 8
休斯顿	0.893 7	0.941 8	0.895 8	0.968 3
旧金山	0.919 7	0.953 9	0.924 6	0.977 7
洛杉矶	0.903 5	0.948 9	0.918 2	0.964 2
平均 AUC	0.905 4	0.951 5	0.923 1	0.969 0

MetaPath 算法之所以在 AUC 上优于其他算法,是因为 MetaPath 算法充分考虑了 Meetup 社交网络的语义信息,而对比算法对语义信息挖掘不足,基于传统随机游走的思想也带入了一定的噪声路径,影响了算法的性能。

(3)P@n

表 3 所示为不同推荐算法在各城市的 P@1 和 P@3 值,可以看到,在五个城市中,MetaPath 算法明显优于其它算法。

表 3 各推荐算法五城市 P@1 和 P@3 值

	P@n	HeteRS	RRWR	BPP	MetaPath
亚特兰大	P@1	0.04	0.22	0.08	0.31
	P@3	0.08	0.17	0.08	0.28
芝加哥	P@1	0.08	0.27	0.12	0.38
	P@3	0.06	0.28	0.11	0.37
休斯顿	P@1	0.05	0.19	0.08	0.25
	P@3	0.07	0.18	0.11	0.23
旧金山	P@1	0.07	0.27	0.11	0.35
	P@3	0.05	0.26	0.09	0.32
洛杉矶	P@1	0.08	0.27	0.07	0.38
	P@3	0.1	0.19	0.1	0.29

(3)MAP

表 4 所示为不同算法在五个城市的 MAP 值,可以看到,MetaPath 在各城市的 MAP 值均优于其

它方法,MetaPath 的整体推荐准确率有着显著提升,也表明了语义信息在社交事件推荐中的重要性。

表 4 各推荐算法五城市 MAP 值

	HeteRS	RRWR	BPP	MetaPath
亚特兰大	0.081 2	0.269 7	0.131 1	0.349 5
芝加哥	0.118 9	0.391 6	0.151 1	0.412 1
休斯顿	0.080 2	0.259 2	0.120 4	0.395 0
旧金山	0.110 0	0.372 1	0.149 7	0.440 3
洛杉矶	0.120 2	0.291 3	0.121 2	0.355 8

(4)召回率

图 4 描述了召回率与推荐列表长度 K 的关系曲线.在本实验中,仅考虑 MetaPath 和 RRWR 算法,因为这两个算法在之前的实验中分别表现最优和次优.从结果可以看出:随着 K 的增加,召回率一直在增大,这意味着更多的有效事件出现在推荐列表中;无论长度 K 是多少,MetaPath 的召回率始终优于 RRWR,表明 MetaPath 提高了有效事件的命中率,也验证了本文所提推荐方法的有效性。

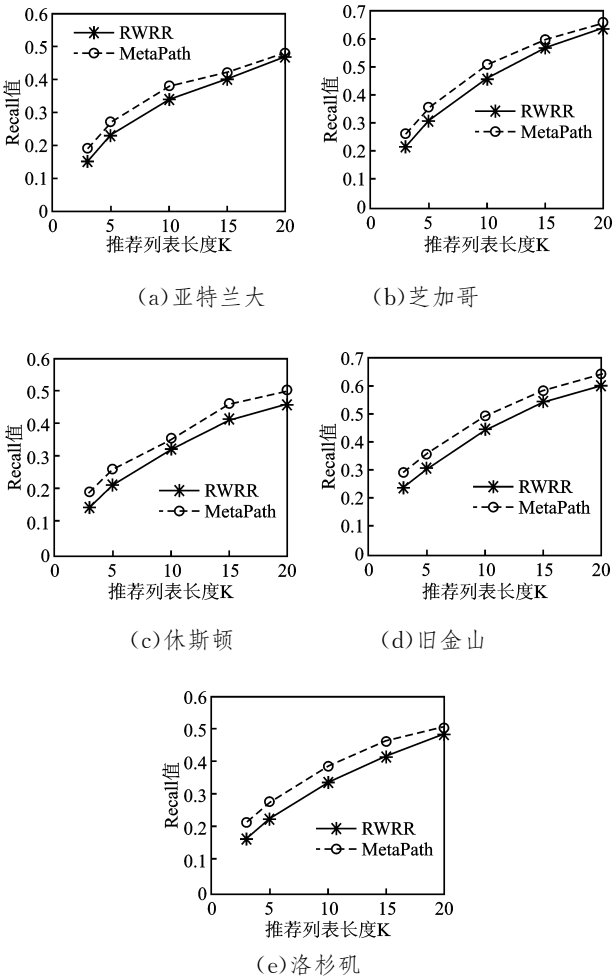


图 4 五城市召回率实验结果

4 结束语

本文提出了一种基于元路径的社交事件推荐算法,首先为 EBSNs 构建异质信息网络,再通过异质网络中的元路径挖掘社交网络中的语义信息,构建用户与事件之间的相关性矩阵. 由于不同元路径对事件推荐的影响程度不同,本算法通过优化问题对元路径的权重进行求解,得到各元路径权重. 在此基础上,利用相关性矩阵与元路径权重实现社交事件推荐. 最后,本文采用应用广泛的 EBSNs 平台 Meetup 社交网络真实数据集进行了实验分析,与其他推荐算法的对比验证了本文基于元路径推荐算法的有效性. 本文将异质信息网络与元路径方法引入到 EBSNs 中,深入挖掘社交网络语义信息,为 EBSNs 的事件推荐研究提供了新的思路. 如何结合不同用户的特性实现个性化元路径选择的社交事件推荐是该方法今后值得研究的课题.

参考文献:

- [1] QIAO Z, ZHANG P, CAO Y N, et al. Combining heterogenous social and geographical information for event recommendation[C]//Proceedings of the 28th AAAI Conference on Artificial Intelligence. Québec City, Québec, Canada: AAAI, 2014: 145-151.
- [2] CHEN HH, CUI X L, JIN H. Top-kfollowee recommendation over microblogging systems by exploiting diverse information sources[J]. Future Generation Computer Systems, 2016, 55: 534-543. DOI: 10.1016/j.future.2014.05.002.
- [3] CHEN CC, SUN YC. Exploring acquaintances of social network site users for effective social event recommendations[J]. Information Processing Letters, 2016, 116(3): 227-236. DOI: 10.1016/j.ipl.2015.11.013.
- [4] YIN H Z, SUN Y Z, CUI B, et al. LCARS: a location-content-aware recommender system[C]//Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Chicago, IL, United States: ACM, 2013: 221-229. DOI: 10.1145/2487575.2487608.
- [5] LEE S, SONG SI, KAHNG M, et al. Random walk based entity ranking on graph for multidimensional recommendation[C]//Proceedings of the 5th ACM Conference on Recommender Systems. Chicago, IL, United States: ACM, 2011: 93-100. DOI: 10.1145/

2043932.2043952.

- [6] GAO B, LIU TY, WEI W, et al. Semi-supervised ranking on very large graphs with rich metadata[C]//Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Diego, CA, United States: ACM, 2011: 96-104. DOI: 10.1145/2020408.2020430.
- [7] PHAM T-AN, LI X, CONG G, et al. A general graph-based model for recommendation in event-based social networks[C]//Proceedings of the IEEE 31st International Conference on Data Engineering. Seoul, Korea (South): IEEE, 2015: 567-578. DOI: 10.1109/ICDE.2015.7113315.
- [8] GUPTA M, KUMAR P, BHASKER B. A new relevance measure for heterogeneous networks[C]//Proceedings of the 17th International Conference on Big Data Analytics and Knowledge Discovery. Valencia, Spain: Springer, 2015: 165-177. DOI: 10.1007/978-3-319-22729-0_13.
- [9] GUPTA M, KUMAR P, BHASKER B. HeteClass: a meta-path based framework for transductive classification of objects in heterogeneous information networks[J]. Expert Systems with Applications, 2017, 68: 106-122. DOI: 10.1016/j.eswa.2016.10.013.
- [10] YUAN Q, CONG G, SUN A X. Graph-based point-of-interest recommendation with geographical and temporal influences[C]//Proceedings of the 23rd ACM International Conference on Information and Knowledge Management. Shanghai, China: ACM, 2014: 659-668. DOI: 10.1145/2661829.2661983.
- [11] MO Y J, LI B X, WANG B, et al. Event recommendation in social networks based on reverse random walk and participant scale control[J]. Future Generation Computer Systems, 2018, 79: 383-395. DOI: 10.1016/j.future.2017.02.045.
- [12] LIU X J, HE Q, TIAN YY, et al. Event-based social networks: linking the online and offline social worlds[C]//Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Beijing, China: ACM, 2012: 1032-1040. DOI: 10.1145/2339530.2339693.

作者简介:

冯堂林 男,(1992-),硕士.研究方向为社交网络.

何亮(通讯作者) 男,(1975-),博士,讲师.研究方向为数据挖掘与机器学习. E-mail: lhc@mail. xjtu. edu. cn