



计算机科学

COMPUTER SCIENCE

多因素特征融合的EBSN活动推荐方法

单晓欢, 宋瑞, 李海海, 宋宝燕

引用本文

单晓欢, 宋瑞, 李海海, 宋宝燕. 多因素特征融合的EBSN活动推荐方法[J]. 计算机科学, 2023, 50(7): 60-65.

SHAN Xiaohuan, SONG Rui, LI Haihai, SONG Baoyan. [Event Recommendation Method with Multi-factor Feature Fusion in EBSN](#) [J]. Computer Science, 2023, 50(7): 60-65.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[面向金融活动的复合区块链关联事件溯源方法](#)

Composite Blockchain Associated Event Tracing Method for Financial Activities
计算机科学, 2022, 49(3): 346-353. <https://doi.org/10.11896/jsjcx.210700068>

[基于耦合随机投影的张量填充方法](#)

Tensor Completion Method Based on Coupled Random Projection
计算机科学, 2021, 48(8): 66-71. <https://doi.org/10.11896/jsjcx.200900055>

[基于数据分布特征的线性孪生支持向量机](#)

Linear Twin Support Vector Machine Based on Data Distribution Characteristics
计算机科学, 2019, 46(6A): 407-411.

[面向大规模图数据的分布式子图匹配算法](#)

Distributed Subgraph Matching Algorithm for Large Scale Graph Data
计算机科学, 2019, 46(4): 28-35. <https://doi.org/10.11896/j.issn.1002-137X.2019.04.005>

[基于ASP模式的支持中小企业动态联盟使能服务系统的研究](#)

计算机科学, 2002, 29(10): 110-113.

多因素特征融合的 EBSN 活动推荐方法

单晓欢 宋 瑞 李海海 宋宝燕

辽宁大学信息学院 沈阳 110036

(shanxiaohuan@lnu.edu.cn)

摘 要 基于活动的社交网络(Event-based Social Network,EBSN)是一种新型的复杂异构社交网络,其中的个性化活动推荐具有一定的应用价值。近年来,随着 EBSN 的快速发展,传统方法利用数据挖掘技术有效解决了活动推荐的信息过载问题。然而,仅利用单特征属性或少量线性组合进行计算,且预定义固定权重将降低活动推荐的准确度,此外大多数方法忽略了用户反馈信息对后续推荐的影响。针对上述问题,提出了一种两阶段构成的多因素特征融合的活动推荐方法。查询预处理阶段,将 EBSN 中的活动、历史用户及其之间的关系抽象为有向异构图,并提取节点及边的特征信息进行辅助存储;利用该辅助数据过滤无效节点及边,进而获得相对较小的候选集;根据查询语境,将查询语义转化为查询图。在线查询阶段,融合潜在好友关系、基于活动的协同过滤以及用户对活动的兴趣这 3 方面特征进行活动推荐,并接收用户是否接受活动的反馈信息作为后续推荐的参考因素。在真实数据集和模拟数据集上进行了大量实验,结果表明所提方法相比对比算法在 EBSN 中活动推荐的精确度和用户的满意度方面更优。

关键词: 基于活动的社交网络;多因素特征融合;活动推荐;有向异构图;子图匹配

中图法分类号 TP311

Event Recommendation Method with Multi-factor Feature Fusion in EBSN

SHAN Xiaohuan, SONG Rui, LI Haihai and SONG Baoyan

College of Information, Liaoning University, Shenyang 110036, China

Abstract Event-based social network(EBSN) is a new kind of complex heterogeneous social network, the personalized event recommendation in it has certain application value. In recent years, with the rapid development of EBSN, the problem of information overload for event recommendation has been solved by data mining technology. However, it will reduce the accuracy of event recommendation by only using a single feature attribute or a small number of linear combinations for calculation, and predefining fixed weights. In addition, most approaches ignore the influence of user feedback information on subsequent recommendation. Aiming at the above problems, an event recommendation method fusing multi-factor features is proposed, which consists of two phases. In the query preprocessing phase, the events, historical users and their relationships in EBSN are abstracted as a directed heterogeneous graph, and the feature information of nodes and edges is extracted for auxiliary storage. A relatively small candidate set is obtained by filtering invalid nodes and edges with the auxiliary data. According to the query context, the query semantics are transformed into the query graphs. In the online query phase, it combines the characteristics of potential friends, event-based collaborative filtering and users' interests to recommend, and also receives feedback from users on whether they accept the event as a reference factor for subsequent recommendations. Large number of experiments on real datasets and simulated datasets verify the accuracy and user satisfaction of the proposed method in EBSN event recommendation.

Keywords Event-based social network, Multi-factor feature fusion, Event recommendation, Directed heterogeneous graph, Sub-graph matching

1 引言

近年来,随着互联网技术的飞速发展,一种新型的基于活动的社交网络(EBSN)^[1]逐渐兴起。与传统社交网络不同的是,EBSN 上的用户通过线上选取活动、线下参与的方式结识具有共同兴趣的用户^[2],打破了传统社交网络线上社交的局限性,其中比较典型的应用有 Meetup, Plancast, Douban

等^[3]。EBSN 上每天发布大量的活动,以 Meetup 为例,其拥有注册用户 4 千余万人,每月举办活动超过 300 万^[3]。随着用户和活动数量的日益增长,根据用户偏好挖掘其感兴趣的变得越来越困难,因此结合用户需求和活动特点进行有效的个性化活动推荐对 EBSN 的良性发展至关重要。

现有 EBSN 上的活动推荐方法已考虑了用户偏好、活动容量冲突、用户及活动的位置等特征属性^[4],用以提高活动

到稿日期:2022-09-05 返修日期:2022-12-24

基金项目:国家重点研发计划

This work was supported by the National Key Research and Development Program of China.

通信作者:宋宝燕(bysong@lnu.edu.cn)

推荐的准确度。然而在现实生活中,用户对活动推荐的要求各有不同,例如某些用户对活动举办地点更在意,而某些用户则更注重举办方的声誉或口碑,因此很难预先定义哪种因素有利于活动推荐。现有方法大多仅利用单个特征属性或少量线性组合进行计算,且预定义的固定权重会对大规模 EBSN 上的活动推荐质量产生影响。此外,现有研究较少考虑用户的反馈意见,即将用户接受与否的意见作为后续活动推荐的考虑因素,然而活动推荐后的用户反馈有利于我们更好地了解用户兴趣^[5],以便后续的个性化推荐。

基于上述问题,本文提出了一种 EBSN 中多因素特征融合的活动推荐方法,该方法由查询预处理和在线查询两阶段构成。本文的主要贡献点如下:

(1) 查询预处理阶段,由于 EBSN 是一种复杂的异构社交网络,因此本文排除冲突活动的影响,将 EBSN 中的活动、历史用户以及活动发生的先后顺序、用户参与活动情况抽象为有向异构图。为加速搜索用户感兴趣的活动的,本文提取节点及边的特征信息进行辅助存储,并利用该辅助数据过滤无效节点及边,获得相对较小的候选集。根据查询语境,将查询语义转化为查询图,进而将活动查询转化为子图匹配问题进行研究。

(2) 在线查询阶段,针对历史用户,融合用户间的好友关系、基于活动的协同过滤,以及用户对活动的兴趣值等多因素特征,加权计算用户对活动的兴趣程度,进而实现对用户的个性化活动推荐。针对初始用户,利用带有反馈的改进 UCB 算法,对影响活动的因素进行筛选,并根据多影响因素计算用户对活动的兴趣值,按兴趣值大小进行活动推荐。

2 相关工作

EBSN 推荐系统相较于传统推荐系统推荐内容更多,如活动推荐^[6]、群组推荐^[7-8]、朋友推荐^[9]及场地推荐^[10-11]等。本文针对活动推荐进行的深入研究发现,随着 EBSN 的兴起与发展,用户及活动数量日益增长,如何从大量活动中借助用户偏好、反馈信息等获得更准确的推荐结果变得愈发困难。

传统的推荐算法中,CB+CF 的混合推荐方法^[12]使用关联数据解决活动时效性引起的数据稀疏性问题,通过集成用户多样性模型来检测用户对特定主题的倾向,然而其忽略了活动本身特性对推荐结果的影响。

EBSN 上的活动推荐算法通常集中于将单个活动推荐给一组潜在用户,而忽略了不同活动之间的冲突问题,导致了不可行或冗余的推荐。为解决这一问题,She 等^[13]提出了对小规模数据有效的精确剪枝算法以及两种可应用于大规模数据集的近似算法,该方法仅考虑了活动的容量冲突,忽略了时间冲突造成的影响。Event2Vec^[14]是一种融合了时空信息的嵌入方法,该方法利用多任务学习和参数共享技术,将活动、位置和时间嵌入到基于活动序列数据的低维空间中,利用不同的 Event2Vec 变量来解决时空信息之间的交互影响。Trinh 等^[15]结合活动内容和用户特征来构造关联矩阵,再结合矩阵、时空信息和用户偏好,向关键用户的活跃好友推荐活动,其推荐对象具有明显的针对性,缺乏普适性。She 等^[16]提出了一种基于贪婪的启发式解决方案,该方法解决了早期方法限制每个用户只能参加一个活动的问题,考虑了时空冲突和旅行花费等因素,然而该近似方法的准确性无法得到保证。Liu 等^[17]提出了一种基于演化图的连续推荐算法,其使用

滑动窗口机制构造演化图,在演化图上执行重启随机游走为用户连续推荐新活动。Cheng 等^[18]通过约束问题,考虑活动参与上下限、时间冲突等因素,提出了两种近似算法。FASEA^[19]中除了考虑活动及用户容量限制、活动冲突,还允许用户提供是否接受推荐的反馈,以此提高推荐的准确性,文中分别利用 TS,UCB 以及 eGreedy 算法进行活动推荐,其中 UCB 算法效果最好。基于深度学习的方法^[20]通过建立深度学习模型来捕获相关信息,搜寻复杂的关系,通过学习用户和活动之间的关系,挖掘用户和活动的潜在特征,进而获得推荐列表,但其需要大量的训练数据以及复杂的参数调整。

3 基于有向异构图的活动推荐预处理

3.1 有向异构图构建

由于 EBSN 是一种复杂的异构社交网络,因此本文将 EBSN 转化为有向异构图 $G(V, E, L_V, W)$ 。其中 V 表示节点集合,由 EBSN 中的活动和历史用户两类节点构成;边集合 E 包含表示活动间发生先后顺序的有向边以及历史用户参与某一活动的有向边; L_V 则为节点标签属性集合,包含活动类别、举办时间、地点等属性; W 为边权值集合,仅存在于用户与活动间的有向边上。 $\langle F, I \rangle$ 二元组形式表示用户参与某一活动的信息,其中 F 表示用户参与该活动的次数,由用户历史信息获得; I 表示该活动对用户的影响力,其计算式如下:

$$I = \sum_{i=1}^n (1-D) \frac{C_i - R_i}{d} \quad (1)$$

其中, D 为衰减因子, C_i 表示当前活动时间, R_i 表示曾经参与过的某一活动的时间, d 表示一天 24 小时。随着时间的推移,用户对活动的兴趣程度会发生变化,活动对用户的影响力会下降,因此我们利用时间衰减函数计算每类活动的权值。

在构建有向异构图的过程中,将活动按举办日期及时间前后进行连边,早举办的活动指向晚举办活动。然而在 EBSN 大量待举办的活动中,举办日期及时间可能存在交集,即两个活动的举办日期相同且时间存在交叉,如两个活动举办日期均为 2020-03-02,举办时间分别为 8:00-11:00 及 9:00-12:00,则认为两个活动为冲突活动。本文对冲突活动间不做连边处理,以减少无效边数量,避免后续查询过程中再进行冲突检测等操作,有效提高查询效率。

图 1 为本文构建的具有 2 个用户节点(u_1, u_2)、15 个活动节点(v_1, v_2, \dots, v_{15})的有向异构图,不同阴影活动节点代表不同活动类别。

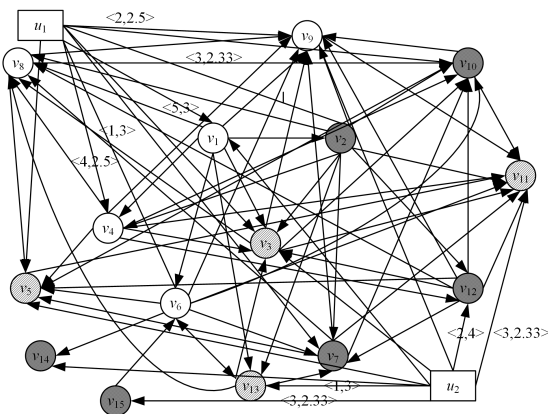


图 1 有向异构图

Fig. 1 Directed heterogeneous graph

3.2 异构图属性特征提取

有向异构图中活动节点及活动间的边上包含了大量有价值的信息,如活动节点的出入度(即其与其他活动的关联程度)、举办日期及时间、活动容量(活动可容纳参与用户数),以及活动类别等信息,这些信息对于后续的活动推荐具有良好的过滤作用,可用于快速筛选符合条件的活动。

为此本文通过深度优先遍历算法按照活动类别统计各节点的拓扑结构信息(入度、出度、活动举办日期及时间、活动容量),并将其保存到辅助二维表中。由于EBSN中的活动具有时序性,所以二维表中的活动信息根据活动节点举办日期顺序排序,当活动节点举办日期相同时,按照时间排序,若举办时间也相同,则按入度降序排序。同理将边按照两端点的活动类型进行分类,并统计每类边的数量及端点信息。

3.3 查询语义转换

在实际应用中,用户根据感兴趣的话题或者某些限制条件来搜索活动,本文按照有向异构图规则将查询语义转化为线性查询图,进而将活动查询转化为有向异构图上的子图匹配问题。例如,用户查询2021-03-17-2021-03-23期间举办的篮球(B)、演唱会(C)、话剧(D)3个类别的活动,且篮球早于话剧的活动,根据此语境转化的查询图如图2所示。

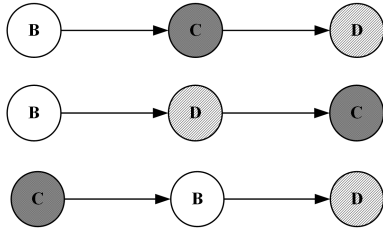


图2 查询语义转化为查询图

Fig. 2 Query semantics transforms into query graphs

3.4 基于属性特征的候选集过滤

子图匹配效率与候选集的大小及过滤程度紧密相关,利用高效的过滤策略可以获得较小的候选集,从而提高匹配速度。

根据查询日期及时间对候选节点进行过滤,日期可以是某个具体日期也可以是日期范围,利用辅助二维表将不满足查询日期的节点进行初始过滤;根据度信息进行二次过滤,异构图中活动节点的出入度应大于查询节点,进而得到最终的候选节点集。根据有向边的特性,在候选节点集的基础上进一步过滤,利用辅助边特征信息,根据边的类型检索,获得满足边类型且端点对应在对应节点候选集中的边。

4 多因素特征融合的活动推荐

4.1 潜在好友关系

传统社交网络中,用户间具有显性的好友关系,现有推荐算法常利用好友间存在相似偏好的特性来提高推荐效果。然而EBSN中用户的好友关系是隐式的,本文利用元路径思想挖掘用户间潜在的好友关系,作为活动推荐的考虑因素之一。

元路径是定义在网络模式上连接两类实体的一条特定路径,其形式化定义为 $A_1 \xrightarrow{R_1} A_2 \xrightarrow{R_2} \dots \xrightarrow{R_n} A_{n+1}$,表示实体类型之间的一种复合关系 $R = R_1 \circ R_2 \circ \dots \circ R_n$,其中 \circ 表示关系的复合操作^[21]。元路径描述了实体间的语义关系,例如本文

中 U 表示用户, E 表示活动,那么 UEU 就是一条元路径,表示两个用户共同参加某一活动。本文利用元路径相似度确定用户间的潜在好友关系,其计算式如式(2)所示:

$$S(x, y) = \frac{2 \times |\{P_{x \rightarrow y}; P_{x \rightarrow y} \in P\}|}{2 \times |\{P_{x \rightarrow x}; P_{x \rightarrow x} \in P\}| + 2 \times |\{P_{y \rightarrow y}; P_{y \rightarrow y} \in P\}|} \quad (2)$$

本文讨论的是对称路径,对象 x 和 y 是相同类型, $P_{x \rightarrow y}$ 表示一条 x 到 y 的路径。在给定 P 的路径模式下,对象 x 到 y 的总路径数越多,相似性越高。符合 P 的路径模式下,对象 x 和 y 到各自的总路径数是一个归一项,描述了对象 x 与 y 在 P 模式下自身的可达路径总数。该值越大,说明 x 和 y 自身在图中沿 P 路径的连接越发散,考虑相似度时,需要对这一项做归一约束。如用户 u_1 与 u_2 参加各个活动的次数分别为 $\{5, 3, 1, 4, 1, 1, 0, 0, 2, 2, 0, 0, 0, 0, 0\}$ 和 $\{4, 0, 2, 0, 1, 0, 1, 0, 1, 0, 3, 2, 1, 3, 3\}$,则利用式(2)计算 $S(u_1, u_2) = 0.431$ 。

交换矩阵^[21]是给定有向异构图和它的一条元路径 $p = (A_1 A_2 \dots A_n)$ 下的交换矩阵 $M = W_{A_1 A_2} W_{A_2 A_3} \dots W_{A_{n-1} A_n}$ 。其中 $W_{A_i A_j}$ 表示 A_i 类节点与 A_j 类节点之间的邻接矩阵; M 代表在元路径 p 下, A_i 类型中的第 i 个对象到 A_n 类型中的第 j 个对象之间的总路径数。邻接矩阵的存在使得元路径的相似度计算变得更容易,如式(3)所示:

$$S(x_i, x_j) = \frac{2M_{ij}}{M_{ii} + M_{jj}} \quad (3)$$

其中, M_{ij} 表示邻接矩阵第 i 行、第 j 列的数,同理 M_{ii} 和 M_{jj} 表示邻接矩阵第 i, j 行第 i, j 列的数。利用元路径,计算好友之间的相似度,并根据实际情况,选择top- k 个用户作为前 k 个潜在好友。本文选取 $k=10$,即最多有10个潜在好友对用户参加活动产生影响。

4.2 基于活动的协同过滤

基于活动的协同过滤是根据所有用户对活动的偏好,发现活动间的相似度,然后根据用户的历史偏好信息将类似的活动推荐给该用户^[22]。本文活动间的相似度通过修正余弦相似度来计算,通过减去用户对活动的平均评分,解决了余弦相似度忽略不同用户对活动评分尺度不同的问题。修正的余弦相似度计算式如下:

$$s(i, j) = \frac{\sum_{u \in U} (R_{u,i} - \bar{R}_u)(R_{u,j} - \bar{R}_u)}{\sqrt{\sum_{u \in U} (R_{u,i} - \bar{R}_u)^2} \sqrt{\sum_{u \in U} (R_{u,j} - \bar{R}_u)^2}} \quad (4)$$

其中, $s(i, j)$ 表示活动 i 和活动 j 的相似度, U 表示同时参加过活动 i 和 j 的用户集合。

本文利用活动相似性及用户参与活动的次数,计算用户参与活动的可能性,计算式如下:

$$P_{uj} = \sum_{i \in N(u) \cap SG(i, k)} \omega_{ji} r_{ui} \quad (5)$$

其中, P_{uj} 为用户 u 参与活动 j 的可能性; $N(u)$ 为用户参与活动集合; $SG(i, k)$ 为与活动 i 最相似的 k 个活动集合; ω_{ji} 表示活动 j 与活动 i 的相似度,利用式(4)计算获得; r_{ui} 表示用户 u 对活动 i 的评分,通过计算有向异构图中用户与活动间的边权值获得。

4.3 活动兴趣值计算

考虑到新用户没有参与活动的历史信息,无法通过潜在好友关系及协同过滤两种特征因素进行活动推荐,因此本文

将用户对活动的直接兴趣程度也作为活动推荐的参考因素之一。

由于每个活动均具有多个属性,无法准确判断哪个属性对用户的影响较大,因此本文采用引入用户反馈的改进 UCB 算法^[19]来计算活动兴趣值。该算法既能对重要特征进行选择,又能删除对因变量影响较小的特征,同时接收用户的反馈作为后续推荐特征选择的因素之一,有效提高了推荐的准确性。

EBSN 上的活动具有时序性,随着时间的推移,举办完的活动就失去了价值。对于新用户而言,他们在 EBSN 上完成活动推荐后便成为历史用户,当历史用户再次登陆平台参加活动时,此时历史用户的参与次数及活动影响力也会发生相应的变化。上述 EBSN 上的动态变化都将导致有向异构图拓扑结构及内容的动态演化,因此在有向异构图中需动态添加新增活动、新增历史用户及参与活动信息,删除已举办活动及其关联信息,更新用户与活动间的权值。本文考虑 EBSN 系统平台的维护规则,为了避免在整个动态变化过程中重新构建有向异构图,文中只针对局部变化的活动节点、历史用户与活动的有向边进行操作。采用滑动窗口的思想,设置时间间隔 t ,收集 t 时间段内的动态变化操作,每隔 t 时间对有向异构图进行统一更新。

4.4 在线活动推荐

本文多因素特征融合的活动推荐方法(Event Recommendation based on Multi-factor Feature Fusion, MFF_ER)综合潜在好友关系、基于活动的协同过滤,以及活动兴趣值这 3 种特征因素,计算用户对满足条件活动的综合分数 $score(u, e)$,并按照降序排序,依次推荐给用户,具体算法如算法 1 所示。

$$score(u_i, e_z) = \alpha S(u_i) + \beta P(u_i) + \gamma I(u_i) \quad (6)$$

其中, α, β, γ 为潜在好友关系 $S(u_i)$ 、基于活动的协同过滤 $P(u_i)$,以及用户对活动的兴趣值 $I(u_i)$ 这 3 种因素的权重因子。可根据实际情况,按各因素的重要程度进行设置,且保证相加之和为 1。

算法 1 MFF_ER Algorithm

输入: $\lambda_1, \epsilon, \lambda_2, V_{sum}, V_{sin}, C, v, \alpha, \beta, \gamma$

输出: A_t

1. $\mathbf{Y} \leftarrow \lambda_1 \mathbf{I}_{d \times d}$;
2. $\mathbf{b} \leftarrow \mathbf{0}$;
3. $\mathbf{M} = \mathbf{C} \times \mathbf{C}^T$;
4. for i to V_{sum} do
5. for j to V_{sum} do
6. $s(x_i, x_j) = \frac{2\mathbf{M}M_{ij}}{\mathbf{M}_{ii} + \mathbf{M}_{jj}}$;
7. $S_{ij} = s(x_i, x_j)$;
8. $P_{ij} = \sum_{i \in N(u) \cap S(i, k)} w_{ji} r_{ui}$;
9. end for
10. end for
11. for $t \leftarrow$ to T do
12. observe $\{x_{i,v}\}$;
13. $\hat{\theta}_1 \leftarrow \mathbf{Y}^{-1} \mathbf{b}$;
14. for $V_{sin} \in V_{sum}$ do

15. for $v \in V_{sin}$ do
16. $\tilde{r}_{t,v} \leftarrow \mathbf{x}_{t,v}^T \hat{\theta}_1$;
17. $\hat{r}_{t,v} \leftarrow \tilde{r}_{t,v} + \alpha \sqrt{\mathbf{x}_{t,v}^T \mathbf{Y}^{-1} \mathbf{x}_{t,v}}$;
18. end for
19. $\hat{r}_{t,v} \leftarrow \hat{r}_{t,v} + \hat{r}_{t,v}$;
20. end for
21. $r_{t,v} \leftarrow \alpha S + \beta P + \gamma \hat{r}_{t,v}$;
22. for $V_{sin} \in V_{sum}$ do
23. sort $r_{t,v}$ in descending order;
24. $A_t \leftarrow V_{sin}$;
25. end for
26. for $v \in V_{sin}$ do
27. $\mathbf{Y} \leftarrow \mathbf{Y} + \sum_{v \in A_t} \mathbf{x}_{t,v} \mathbf{x}_{t,v}^T$;
28. $\mathbf{b} \leftarrow \mathbf{b} + \sum_{v \in A_t} \hat{r}_{t,v} - \frac{\lambda_2}{2}$;
29. end for

在算法 1 中,输入参数 λ_1 表示初始化单位矩阵系数; ϵ 表示算法对活动推荐的探索系数; V_{sum} 表示活动候选集; V_{sin} 是候选集的一个子集,其活动数由用户选择接纳活动的数量决定; C 表示历史用户参与活动的次数, v 表示某一活动; 输出参数 A_t 则是最终推荐活动结果集。初始化属性值及兴趣值,如第 1—3 行所示; 第 4—10 行通过元路径计算潜在好友关系,通过活动间的相似性进行基于活动的协同过滤计算; 第 11—13 行进行权向量计算; 候选集中活动的兴趣值计算如第 14—20 行所示; 第 21 行获得融合多因素的活动综合评分; 第 22—25 行表示根据评分大小进行活动推荐; 更新 \mathbf{Y} 和 \mathbf{b} , 如第 26—29 行所示,并利用更新后的 \mathbf{Y} 和 \mathbf{b} 来计算权向量。算法的时间复杂度为 $O(n^2)$, 其中 n 为活动候选集 V_{sum} 中的活动数。

5 实验与分析

5.1 实验环境与数据集

本文实验在 Intel Core i7-1165G7 CPU @ 2.80GHz 4.70GHz 处理器、16GB 内存、512GB SSD 的计算机上完成。

本文采用真实数据集和模拟数据集进行实验验证。真实数据集爬取 Meetup 网站中 2016-08—2018-06 的 Meetup 数据,包含的用户数量为 32100,活动数量为 89450,其中用户属性主要包括用户位置、偏好,活动属性包括活动日期及时间、举办地点、类别等信息。模拟数据集中产生的 $\hat{\theta}$ 符合正态分布,生成一个具有 16 维的特征向量,且符合正态分布和均匀分布,活动容量符合正态分布,用户容量符合均匀分布,实验具体参数配置参照对比方法设置^[19]。

5.2 实验分析

本文从活动接受率、遗憾率,以及算法运行时间等方面,与同样采用 MAB(Multi-Armed Bandit)思想^[19]且考虑用户反馈的改进 TS,UCB,以及 eGreedy 算法进行对比分析。

5.2.1 活动接受率与遗憾率对比分析

接受率为用户接受活动的数量与推荐数量的比值,而遗憾率指受到活动容量限制或用户主观拒绝导致无法参与推荐的活动数与接受活动数的比值^[19]。本文通过接受率与遗憾率的实验对比来验证本文算法的准确性及用户的满意度。

如图 3 和图 4 所示, TS 相较于其他算法性能较差, 这是因为其参数设置是固定不变的, 无法通过前期推荐活动预估其对后续活动的影响。尤其针对新用户, 事先无法确定 θ 值, 只能根据观察到的样本进行估计。eGreedy 算法考虑了前期推荐对后续推荐的影响, 但由于随机选择概率小于参数 ϵ 时, 活动是随机选择推荐, 导致某些活动虽是最优的, 但初始化设置使得它被访问的概率较低, 因而失去了被推荐的可能。UCB 算法则改进了这一点, 其通过数据计算出所选择的臂, 避免随机选择造成的不确定性。本文 MFF_ER 算法融合多因素特征, 除了选择重要特征计算用户对活动的兴趣值外, 也将潜在好友关系及基于活动的协同过滤作为推荐参考因素, 实现用户的个性化推荐, 有效提高了用户的接受率。

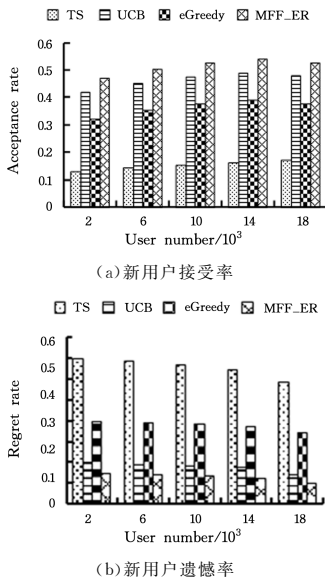


图 3 模拟数据集上的接受率及遗憾率对比

Fig. 3 Acceptance rate vs. regret rate on simulated datasets

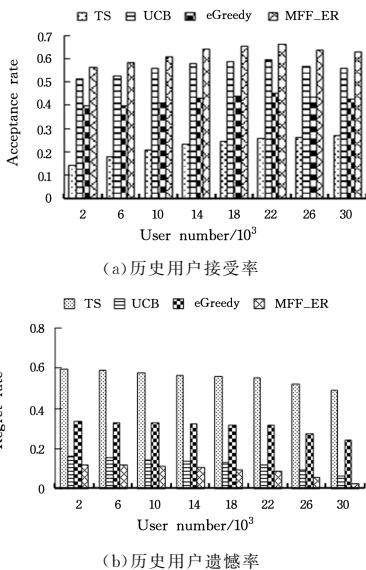


图 4 真实数据集上的接受率及遗憾率对比

Fig. 4 Acceptance rate vs. regret rate on real datasets

5.2.2 运行时间对比分析

图 5 给出了不同数据集上运行时间的对比情况, 图 5(a) 为模拟数据集上针对新用户的运行时间, 图 5(b) 为真实

数据集上针对历史用户的运行时间。

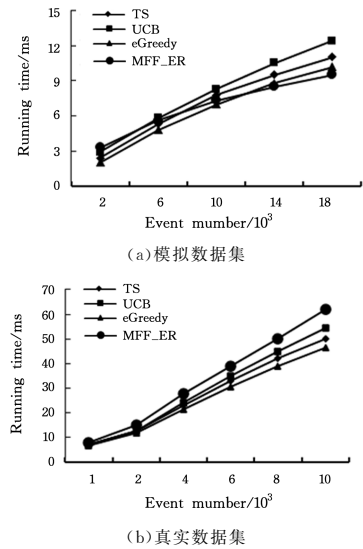


图 5 不同数据集上的运行时间对比

Fig. 5 Comparison of running time on different datasets

如图 5 所示, 本文算法针对历史用户的运行时间要多于新用户, 这是因为相较于新用户, 本文针对历史用户除了需要计算用户对活动的兴趣值外, 还需要计算用户的潜在好友以及基于活动的协同过滤。但在预处理阶段, 本文预先提取了活动及用户的属性特征信息作为辅助数据存储, 以尽量降低多因素计算造成的开销。UCB 算法耗时的原因是它需要为每个活动计算置信区间, 随着活动数的增加, 运行时间明显增加。TS 算法则需要通过采样来估计 θ 的取值, 这样也会耗费一定的时间。由于 eGreedy 算法存在一定数量的随机推荐, 因此活动数较少时, 其速度较快。

图 6 给出了当 α, β, γ 分别取不同值时对接受率的影响。从图 6 可以看出, 接受率随着活动数量的增加而有所提升。当 $\alpha=0.2, \beta=0.1, \gamma=0.7$ 时接受率最优, 这是因为用户对活动的感兴趣程度直接反映了用户参与某活动的欲望, 因此其对历史用户活动推荐的影响比其他两种因素更大。同时, 用户的潜在好友关系对推荐结果的影响大于基于活动的协同过滤, 这是因为用户在一定程度上更喜欢与好友一起参加活动, 且基于活动的协同过滤中活动间的相似性计算存在一定的误差。

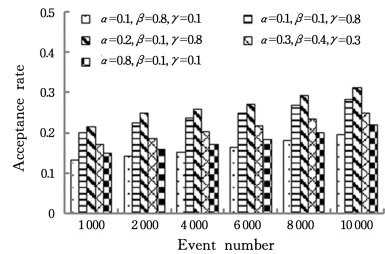


图 6 参数变化对接受率的影响

Fig. 6 Influence of parameter change on acceptance rate

结束语 本文对 EBSN 中的活动推荐问题进行了研究。将 EBSN 转换为有向异构图, 并通过提取的节点及边的属性特征对候选集进行剪枝过滤。在较小候选集上, 融合潜在好友关系、基于活动的协同过滤, 以及用户对活动的兴趣值进行

活动推荐,同时将用户是否接受推荐活动的反馈信息作为后续推荐的考虑因素。实验结果表明,所提方法能为用户快速准确地推荐活动,具有一定的实际应用价值。

参 考 文 献

- [1] WU G, LI L, LI X, et al. Graph Embedding Based Real-time Social Event Matching for EBSNs Recommendation [J]. World Wide Web, 2022, 25(1): 335-356.
- [2] KYOUNGSOO B, SUJI L, DOJIN C, et al. Recommending Personalized Events Based on User Preference Analysis in Event Based Social Networks [J]. Electronic Commerce Research, 2021, 21(3): 707-725.
- [3] LIAO G, LAN T, HUANG X, et al. Survey on Recommendation Systems in Event-based Social Networks [J]. Journal of Software, 2021, 32(2): 424-444.
- [4] WU G, LI X, CUI K, et al. A Graph Embedding Based Real-Time Social Event Matching Model for EBSNs Recommendation [C]//Proceedings of the 21st International Conference on Web Information Systems Engineering. Berlin: Springer, 2020: 41-55.
- [5] YUAN L. Modelling and Learning User Feedback in Event-based Social Networks [J]. International Journal of Performance Engineering, 2022, 18(1): 11-21.
- [6] LI B, WANG G, CHENG Y, et al. An Event Recommendation Model Using ELM in Event-based Social Network [J]. Neural Computing and Applications, 2020, 32: 14375-14384.
- [7] DU Y, MENG X, ZHANG Y, et al. GERF: A Group Event Recommendation Framework Based on Learning-to-rank [J]. IEEE Transactions on Knowledge and Data Engineering, 2019, 32(4): 674-687.
- [8] DU Y, MENG X, ZHANG Y. CVTM: A Content-venue-aware Topic Model for Group Event Recommendation [J]. IEEE Transactions on Knowledge and Data Engineering, 2019, 32(7): 1290-1303.
- [9] CHENG Y, YUAN Y, CHEN L, et al. Event-Participant and Incremental Planning over Event-Based Social Networks [J]. IEEE Transactions on Knowledge and Data Engineering, 2021, 33(2): 474-488.
- [10] PRAMANIKS, HALDAR R, KUMARA, et al. Deep Learning Driven Venue Recommender for Event-based Social Networks [J]. IEEE Transactions on Knowledge and Data Engineering, 2020, 32(11): 2129-2143.
- [11] VINH T L, NGUYEN P TA, TAY Y, et al. Interact and Decide: Medley of Sub-attention Networks for Effective Group Recommendation [C]//Proceedings of the 42nd ACM SIGIR International Conference on Research and Development in Information Retrieval. New York: Association for Computing Machinery, 2019: 255-264.
- [12] KHROUF H, TRONCY R. Hybrid Event Recommendation Using Linked Data and User Diversity [C]//Proceedings of the 7th ACM Conference on Recommender Systems. New York: ACM, 2013: 185-192.
- [13] SHE J, TONG Y, CHEN L, et al. Conflict-aware Event-participant Arrangement [C]//Proceedings of the 2015 IEEE 31st International Conference on Data Engineering. New York: IEEE, 2015: 735-746.
- [14] WANG Y, TANG J. Event2Vec: Learning Event Representations Using Spatial-temporal Information for Recommendation [C]//Proceedings of the 23rd Pacific-Asia Conference on Knowledge Discovery and Data Mining. Heidelberg: Springer, 2019: 314-326.
- [15] TRINH T, WU D, WANG R, et al. An Effective Content-based Event Recommendation Model [J]. Multimedia Tools and Applications, 2021, 80(11): 16599-16618.
- [16] SHE J, TONG Y, CHEN L. Utility-aware Social Event-participant Planning [C]//Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data. New York: ACM, 2015: 1629-1643.
- [17] LIU S, WANG B, XU M, et al. Evolving Graph Construction for Successive Recommendation in Event-based Social Networks [J]. Future Generation Computer Systems, 2019, 96: 502-514.
- [18] CHENG Y, YUAN Y, CHEN L, et al. Complex Event-participant Planning and Its Incremental Variant [C]//Proceedings of the 2017 IEEE 33rd International Conference on Data Engineering. New York: IEEE, 2017: 859-870.
- [19] SHE J, TONG Y, CHEN L, et al. Feedback-aware Social Event-participant Arrangement [C]//Proceedings of the 2017 ACM International Conference on Management of Data. New York: ACM, 2017: 851-865.
- [20] ZHANG S, YAO L, SUN A, et al. Deep Learning Based Recommender System: A Survey and New Perspectives [J]. ACM Computing Surveys, 2019, 52(1): 1-38.
- [21] YU Y X, ZHANG H J. Activity Recommendation Algorithm Based on Latent Friendships in EBSN [J]. Computer Science, 2018, 45(3): 198-205.
- [22] SHEN J, QIAO S J, HAN N, et al. Personalized Recommendation Model with Multiple Information Fusion [J]. Journal of Chongqing University of Technology (Natural Science), 2021, 35(3): 128-138.



SHAN Xiaohuan, born in 1987, Ph.D candidate, is a student member of China Computer Federation. Her main research interests include graph data processing technology and knowledge graph data management, etc.



SONG Baoyan, born in 1965, Ph.D, professor, Ph.D supervisor, is a member of China Computer Federation. Her main research interests include database techniques, big data management, etc.