

事件社交网中基于有向标签图及用户反馈的活动推荐方法

单晓欢, 张志国, 宋宝燕, 任成林*

(辽宁大学 信息学院, 沈阳 110036)

(* 通信作者电子邮箱 243100038@qq.com)

摘要: 由于基于事件的社交网络(EBSN)中的活动具有时效性, 传统社交网络推荐算法无法适用于EBSN。此外, 大多数算法忽略了能影响后续推荐质量的前用户是否接受活动的反馈意见。为此, 提出一种EBSN中基于有向标签图及用户反馈的活动推荐方法。首先, 将EBSN抽象为有向标签图, 并抽取图节点及边的属性特征信息, 构建有向图结构特征(DGSF)索引, 该索引由节点属性特征索引、有向边属性特征索引以及时间特征索引构成, 利用该索引对节点及边进行初次过滤。其次, 提出基于DGSF索引的多属性候选集过滤策略, 利用时间、节点的出入度、标签类型等特征的限制, 实现对查询图候选集的进一步剪枝, 避免冗余计算。然后, 提出一种具有用户反馈的改进UCB(Upper Confidence Bound)活动推荐算法——EN_UCB, 通过引入弹性网回归, 根据多影响因素计算用户对活动的兴趣值, 为用户推荐兴趣值高的活动, 同时接收用户是否接受该活动的反馈, 以优化后续用户的推荐。大量实验结果表明, EN_UCB算法的接受率高于TS(Thompson Sampling)、UCB以及eGreedy算法, 遗憾率远远低于TS和eGreedy算法, 且运行效率高于TS、UCB以及eGreedy算法, 活动数越大, 优势越明显。所提算法能有效实现EBSN上的在线活动推荐。

关键词: 基于事件的社交网络; 有向标签图; 用户反馈; 活动推荐; 弹性网回归

中图分类号: TP311 **文献标志码:** A

Activity recommendation method based on directed label graph and user feedback in event-based social network

SHAN Xiaohuan, ZHANG Zhiguo, SONG Baoyan, REN Chenglin*

(School of Information, Liaoning University, Shenyang Liaoning 110036, China)

Abstract: Due to the timeliness of activities in Event-Based Social Network (EBSN), the traditional social network recommendation algorithms cannot be applied to EBSN. In addition, most of the traditional recommendation algorithms ignore the feedback that can affect whether the previous users accept the recommendation, which influences subsequent recommendation quality. Therefore, an activity recommendation method based on directed label graph and user feedback in EBSN was proposed. Firstly, EBSN was abstracted into a directed label graph, and a Directed Graph Structure Feature (DGSF) index was construction by extracting the property feature information of nodes and edges to filter nodes and edges for the first time. DGSF index consists of node property feature index, directed edge property feature index and time feature index. Secondly, a multi-attribute candidate set filtering strategy based on DGSF index was proposed. By using the limits of time, in-degrees and out-degrees of nodes, and label types, the further pruning of the candidate sets was realized to avoid redundant computation. Thirdly, an improved UCB (Upper Confidence Bound) activity recommendation algorithm with user feedback was put forward, namely EN_UCB (Elastic Net UCB). In EN_UCB, with the introduction of the elastic net regression, the interest values of the user to the activities were calculated according to many influencing factors, and the activities with high interest values were recommended to the user. At the same time, the feedback whether the user accepted the activities was received to optimize the subsequent user recommendation. Experimental results show that EN_UCB has the accept rate higher than TS (Thompson Sampling), UCB and eGreedy, the regret rate far lower than TS and eGreedy, the running time superior to TS, UCB and eGreedy, and the larger the number of activities, the more obvious the advantages. The proposed method implements online activity recommendation in EBSN effectively.

Key words: Event-Based Social Network (EBSN); directed label graph; user feedback; activity recommendation; elastic network regression

收稿日期: 2019-08-12; **修回日期:** 2019-09-12; **录用日期:** 2019-10-24。 **基金项目:** 国家自然科学基金资助项目(61472169, 61502215, 61802160, 51704138); 辽宁省重点研发计划项目(2017231011); 沈阳市中青年科技创新人才支持计划项目(RC180244); 辽宁省公共舆论与网络安全大数据系统工程实验室资助项目(04-2016-0089013); 辽宁省教育厅科学研究项目(LYB201617)。

作者简介: 单晓欢(1987—), 女, 辽宁沈阳人, 实验师, 博士研究生, CCF会员, 主要研究方向: 图数据处理、数据库、大数据管理; 张志国(1991—), 男, 山东临沂人, 硕士研究生, 主要研究方向: 大数据管理; 宋宝燕(1965—), 女, 辽宁开原人, 教授, 博士, CCF高级会员, 主要研究方向: 数据库、射频识别事件流处理、大数据管理、图数据管理; 任成林(1967—), 男, 辽宁沈阳人, 高级工程师, 主要研究方向: 通信传输、数据处理。

0 引言

近年来,信息技术及互联网技术的飞速发展,推动了社交网络的广泛应用,其中以 Meetup^[1]、Plancast^[1]、Douban^[2]和 Google+Event^[3]等为代表的基于事件的社交网络(Event-Based Social Network, EBSN)^[4]作为一种新型的社交网络应用得到了快速的发展。用户可以在EBSN平台上注册、创建、发布和组织社交活动,如篮球、音乐会、社团活动、公益招募等,用户可以根据自己的需求通过限制某些条件在线搜索并在线下参加其选择的活动。

EBSN平台会不定期地发布各种活动,同时也会将过期的活动下线。随着社交网络的不断兴起,EBSN上活跃用户数以及活动数量日益增长。因此,用户找到其感兴趣的活动变得越来越困难,研究有效的EBSN活动推荐方法以提高推荐质量将面临巨大的挑战。

为此,本文提出一种EBSN上基于有向标签图及用户反馈的活动推荐方法。因为EBSN是一种异构的复杂社交网络,因此本文利用活动发生的时间先后顺序将平台上的活动抽象为有向标签图,其中节点表示活动,有向边则表示两个活动发生的先后以及发生地之间的距离,活动的属性(如活动类型,发生的时间、地点等)则通过节点标签进行表示,两个活动是否同一天进行等属性则通过边标签表示。同时将EBSN上的活动推荐转换成子图查询问题进行研究。本文的主要内容如下:

1) 提出一种有向图结构特征(Directed Graph Structure Feature, DGSF)索引,该索引由节点属性特征(Node Property Feature, NPF)索引、有向边属性特征(Directed Edge Property Feature, DEPF)索引以及时间特征(Time Feature, TF)索引构成。利用NPF索引,根据时间、节点的入度及出度等信息过滤掉无效节点,以获得较小的节点候选集;同理,利用DEPF索引及TF索引,根据边标签属性以及活动发生的时间过滤掉无效边,以获得较小的边候选集。

2) 提出基于DGSF索引的多属性候选集过滤策略,利用时间、节点的入度、出度、标签类型以及相隔天数等特征的限制,实现对查询图候选集的进一步剪枝,避免过多的冗余计算。

3) 在获得的候选集基础上,提出一种带有用户反馈的改进UCB(Upper Confidence Bound)活动推荐算法——EN_UCB(Elastic Net UCB),在UCB算法中引入弹性网回归,根据多影响因素计算用户对活动的兴趣值,并按兴趣值从大到小向用户进行推荐,同时接收用户的反馈,实现有效的在线活动推荐。

1 相关工作

随着EBSN中发布的活动越来越多,为用户找到最符合其兴趣的活动变得越来越困难^[5],同时EBSN中的活动推荐具有冷启动、实时、容量限制及冲突限制等特性,这导致传统的协同过滤、矩阵分解等推荐方法无法直接应用于EBSN的活动推荐。

文献[6]在进行活动推荐时,通过分析用户的社会角色以及位置属性,对现有的协同过滤算法进行扩展,定义了四种活动类别,提出一种基于社交及位置属性的活动分类机制。一旦不同的类别被确定,利用适当的排名为新用户进行推荐,解决了冷启动问题。文献[7]提出了一种EBSN中基于上下文的推荐方法,该方法利用用户的喜好和位置关系等多种信息进行综合评估,进而推荐出最合适的活动。文献[8]分析用户对某个主题相关的活动感兴趣,那么其有可能会参加该主题相关的后续活动,通过分析时间、空间等特征,对用户进行监督学习,根据提取的特征进行活动推荐。该算法明确了每个特征对参与活动的影响,算法相对简单,但是它忽略了用户间的社会属性关系对活动推荐的影响。为解决活动推荐的冷启动问题,一种贝叶斯泊松分解模型CBPF(Collective Bayesian Poisson Factorization)^[9]被提出,将其贝叶斯泊松因式分解作为基本单元,用于建模用户对活动、社会关系以及活动内容的响应;然后利用标准矩阵分解模型的思想进一步连接这些基本单元。此外,该模型中活动内容、组织者以及位置信息被用来表示预测用户对冷启动活动的响应。SIARS(Social Information Augmented Recommender System)^[10]充分利用活动组织者及群组的社会影响力,并结合基本的上下文信息进行活动推荐。该系统结合EBSN及其他社交网络的信息描绘活动组织者的社会影响力,并考虑群组成员之间的互动以进行活动推荐。此外,文献[10]还提出了一种利用主题模型寻找活动所属最相似主题的内容感知推荐模型,该模型结合地点知名度及其分布的位置进行活动推荐。该方法考虑了活动的组织者和群组的社会影响力对活动的影响,寻找与活动属性相似的主题并结合地点分布进行活动推荐,提高了活动推荐的准确性;然而该模型未考虑活动冲突、活动容量等影响因素。

上述推荐方法虽然考虑了部分约束条件,但仍未达到全局最优推荐,同时忽略了用户是否接受推荐活动对后续推荐的影响。文献[11]考虑了活动冲突、位置信息以及活动开销等因素,提出了一种启发式算法,考虑多种因素对活动推荐的影响,提高了活动安排的效率;然而该算法没有考虑在线互动安排。文献[12]提出了两种利用剪枝技术的近似算法及精确算法解决不同活动之间存在冲突的问题,从而避免冗余安排,该算法实现了线上活动推荐。文献[13]考虑了诸多限制因素,实现了线上活动推荐,同时为解决现有方法衡量事件属性仅利用单个值和少量属性的线性组合、权重采用预定义和固定值以及不考虑用户是否接受推荐活动等问题,提出了一种新的在线活动推荐策略,引入MAB(Multi-Armed Bandit)问题^[14],分别利用TS(Thompson Sampling)算法、UCB算法^[15]以及eGreedy算法进行活动推荐。

2 有向图结构特征索引

2.1 EBSN与有向标签图转换

本文将EBSN抽象成有向标签图 $G(V, E, LV, LE)$,其中将EBSN中活动抽象为图中节点, V 表示节点集合;活动发生的先后顺序则通过有向边表示, E 则为边集合;活动的属性,如

活动类型、举办时间、地点等抽象为节点标签, LV 则为节点标签属性集合; 两个活动之间的距离以及举办时间利用边标签表示, LE 为边标签属性集合。本文将满足距离小于等于 m (单位: m) 的活动按举办时间的先后进行连边, 先举办的活动指向在其之后举办的活动。

在 EBSN 平台中, 不同的活动之间可能存在一定的冲突, 如两活动在同一时间举办或时间存在交叉等, 为避免在后续查询推荐过程中进行冲突检测等操作, 本文将 EBSN 抽象为有向标签图的过程中已对冲突事件进行过滤, 有效提高了查询效率。

下面以列举的 15 个活动为例, 如表 1 所示, 将其抽象为 15 个节点, 同时按活动类型分为足球 A、电影 B、音乐会 C。其

表 1 节点属性标签表
Tab. 1 Node property label table

节点编号	活动类型	举办地点	举办单位	举办日期	举办时间
v_1	A	大东足球场	大东足球协会	2019-03-17	8:00—11:00
v_2	B	大东大影院	南风大影院	2019-03-17	11:10—13:00
v_3	C	乐天新影城	乐天新影城	2019-03-19	17:00—20:00
v_4	A	铁西足球场	铁西足球协会	2019-03-18	11:00—13:00
v_5	C	卢米万象汇影城	卢米万象汇影城	2019-03-22	18:00—20:00
v_6	A	和平足球场	和平足球协会	2019-03-19	15:00—17:00
v_7	B	中华大影院	中华大影院	2019-03-19	17:20—19:00
v_8	A	城北足球场	城北足球协会	2019-03-20	9:00—12:00
v_9	A	铁东足球场	铁东足球协会	2019-03-22	15:00—18:00
v_{10}	B	铁东音乐厅	铁东音乐厅	2019-03-20	12:30—14:00
v_{11}	C	金逸影城	金逸影城	2019-03-23	18:00—20:00
v_{12}	B	盛京大影院	盛京大影院	2019-03-18	13:10—15:20
v_{13}	C	实维影院	实维影院	2019-03-18	14:00—16:00
v_{14}	B	南湖音乐厅	南湖音乐厅	2019-03-19	20:00—22:00
v_{15}	B	大众音乐厅	大众影视	2019-03-19	8:00—10:00

用户在进行活动搜索时, 可以根据用户的查询限制条件, 将活动搜索转换为查询图表示, 以实现将推荐活动转换为子图查询。例如, 用户要系统为其推荐 2019-03-17—2019-03-23 期间的足球、音乐会和电影活动, 同时足球和电影活动要相差一天。根据查询需求可转换为图 1(b) 所示的 6 种查询图表示, 其中边上的属性 b 表示活动不在同一天进行, “1” 则表示相差一天。

2.2 DGSF 索引

2.2.1 NPF 索引

本文有向标签图中节点的类型、时间、出入度等属性标签

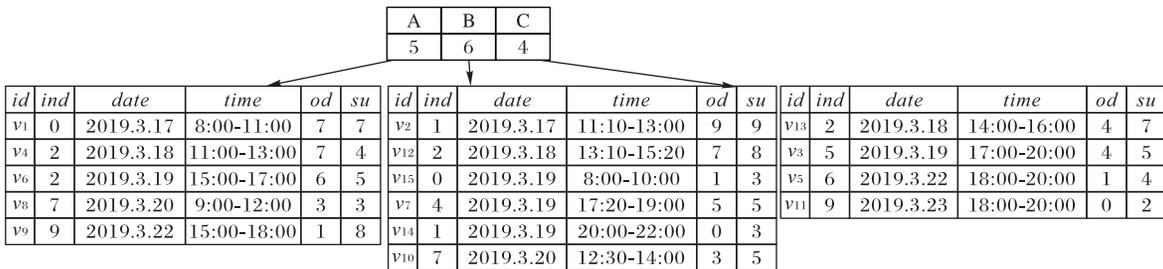


图 2 NPF 索引示意图

Fig. 2 Schematic graph of NPF index

2.2.2 DEPF 索引

DEPF 索引同样包含两级结构: 顶层结构为边标签类型; 底层结构则为每种边标签类型所包含的边, 由 $\langle id1, id2 \rangle$

中如 v_3 与 v_6 因时间有交集, 则认为两活动为冲突活动, 无边相连。转换后的有向标签图如图 1(a) 所示。

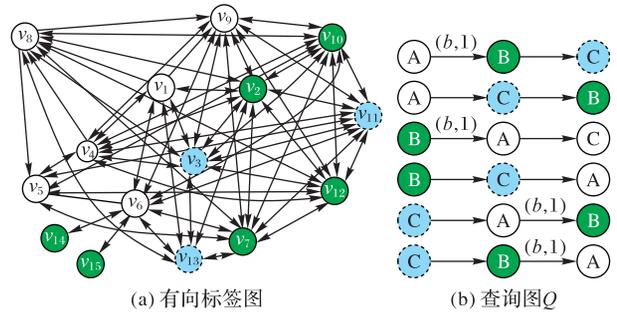


图 1 有向标签图及查询图示例

Fig. 1 Examples of directed label graph and query graph

具有一定的标志性和可辨别性, 因此遍历有向图, 获取节点的属性标签信息构建 NPF 索引。该索引由两级结构构成: 顶层结构索引节点类型, 由 \langle 节点类型, 所属类型数量 \rangle 构成; 底层结构则由 \langle 节点编号, 入度, 日期, 时间, 出度, 活动容纳人数 \rangle 构成, 通过广度优先遍历, 统计各个节点的上述信息。以图 1 为例, 有向标签图 G 中包含了 A、B、C 三种类型的节点, 其 NPF 索引如图 2 所示, 其中: id 表示节点编号, ind 表示节点入度, $date$ 表示活动举办日期, $time$ 表示时间, og 表示节点出度, su 则表示活动能容纳的用户数量。

成。值得注意的是, 因为本文研究的是有向图, 所以 AB 和 BA 不是同一种类型。仍以图 1 为例, 图 G 包含 AA、AB、AC、BA、BB、BC、CA、CB、CC 共 9 种类型的边, 其 DEPF 索引如图 3 所

11)行所示。用户可以选择接受或是拒绝,用户的反馈结果将影响后续的推荐。

5 实验与结果分析

5.1 实验环境及配置

本文实验环境为 Intel Core i7-8550U CPU @1.80 GHz 2.00 GHz 处理器,16 GB 内存,256 GB SSD+1 TB 硬盘,编程语言为 Java。

实验分别在真实数据集和仿真数据集上完成。真实数据集来自 Meetup 社交网站中美国旧金山附近用户及其相关活动的数据,时间范围是 2016-08-31—2018-06-31。其中用户数据包含了用户的位置、偏好等信息,活动数据包括活动的属性、举办日期、时间、举办人、地点等信息。在仿真数据集中,产生的 θ 符合正态分布,特征向量符合正态分布和均匀分布,生成一个具有 16 维的特征。实验具体参数配置如表 2 所示。

表 2 实验参数配置

Tab. 2 Experimental parameter configuration

参数	描述	设定值	默认值
dis	距离	50, 100, 150, 200	100
$ V $	活动数	1 000, 2 000, 4 000, 5 000	1 000
D	向量维数	4, 8, 12, 16, 20	16
$ U $	用户数	400, 600, 800, 1 000, 20 000	20 000
C_v	活动容纳量	$N(50, 100), N(200, 100)$	$N(50, 100)$
C_u	用户容纳量	$U:[1:8]$	$U:[1:8]$
θ	系数	$N(0, 2), U:[-2, 2]$	$N(0, 2)$
X	特征向量	$N(0, 2), U:[-2, 2]$	$N(0, 2)$
λ_1	算法参数	0.5, 1, 2	1
λ_2	算法参数	0.5, 1, 2	0.5
α	算法参数	0.5, 1, 2	2

5.2 结果分析

本节将从接受率、遗憾率^[12]以及查询时间等方面进行实验,验证本文方法的有效性和可行性。

如图 5 所示, $k=10^3$, 随着用户数量的增加,接受率在逐渐上升而遗憾率逐渐降低,这是因为算法的估计 θ 经过若干次修正后变得更加精准。TS 算法性能最差,其接受率较低而遗憾率很高,这是因为在具有特征的 MAB 的情况下,所有活动通过共享 θ 相关联,TS 不能通过前期推荐的活动而预估其对后续活动的影响。eGreedy 算法虽然会在每次推荐后,改善其后续推荐,但由于小于参数 ϵ 时,活动是随机安排的,这导致用户的接受率和遗憾率会相对低于 UCB。本文 EN_UCB 算法引入弹性网回归对特征进行筛选,提高了推荐的准确性,因此用户接受率更高。在仿真数据集及真实数据集中,接受率及遗憾率出现突然下降的情况,这是受活动容量限制的影响;同时接受率与遗憾率出现上升或下降的波动则是受用户反馈的影响。

图 6 显示了各算法的运行时间对比情况,随着活动数的增加,运行时间逐渐增加。其中:UCB 算法最耗时,这是因为它需要为每个活动计算置信区间;TS 算法则需要通过采样获得 θ ,同样需要花费一定的时间;当活动数量较少时,eGreedy 算法运行速度较快,这是因为当小于参数 ϵ 时,活动随机安排;本文构建有向标签图以及 DGSF 索引的过程均在线下进行

行,因此不计入 EN_UCB 算法的运行时间,EN_UCB 算法仅在过滤后的较小候选集上进行查询,因此运行时间最短。

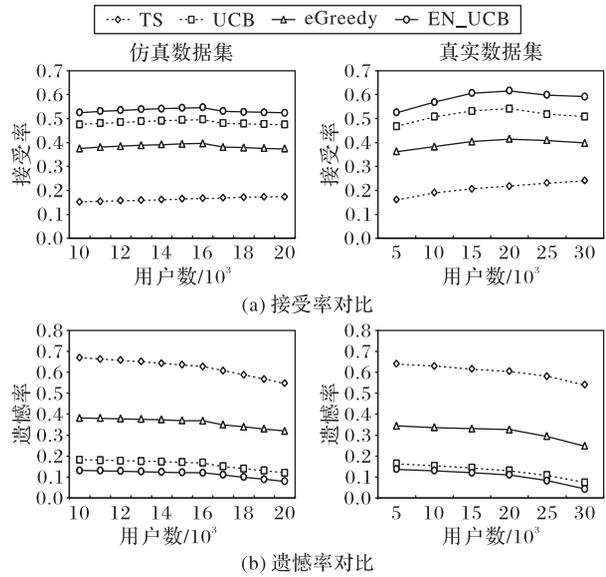


图 5 不同数据集上接受率及遗憾率对比

Fig. 5 Comparison of accept and regret rates on different datasets

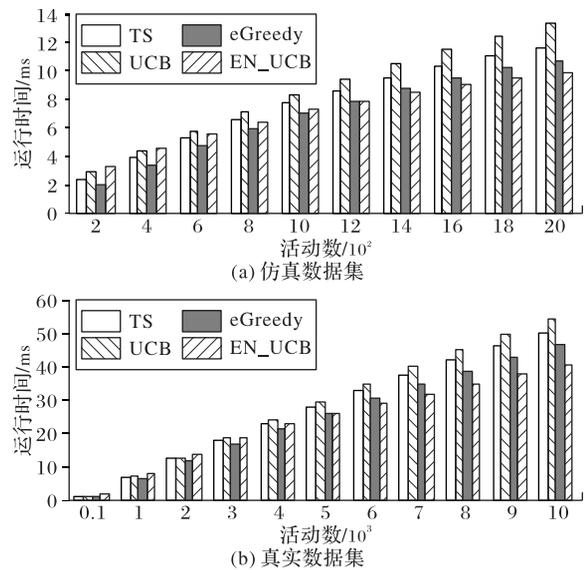


图 6 不同算法在不同数据集上的运行时间对比

Fig. 6 Running time comparison of different algorithms on different datasets

6 结语

本文针对 EBSN 上的活动推荐问题展开研究,首先将 EBSN 转换为有向标签图,并在转换过程中对冲突活动进行过滤;其次,提取节点及边的属性特征信息,构建 DGSF 索引;然后,提出基于 DGSF 索引的多属性候选集过滤策略,利用时间等特征限制进行剪枝过滤,以获得查询候选集;提出一种具有用户反馈的改进 UCB 活动推荐算法,引入弹性网回归以提高推荐准确性,同时接收用户反馈,以优化后续活动推荐。实验结果表明,本文提出的方法能快速准确地为用户推荐活动,具有一定的实际应用价值。

然而,在本文的研究中未考虑 EBSN 中活动动态变化的情况,因此在未来工作中将对动态变化下的索引维护问题进

行研究,使得本文方法更加完善。

参考文献 (References)

- [1] AGRAWAL S, GOYAL N. Thompson sampling for contextual bandits with linear payoffs [C]// Proceedings of the 30th International Conference on Machine Learning. [S. l.]: JMLR. org, 2013, 28: III-1220-III-1228.
- [2] XU B, CHIN A, COSLEY D. On how event size and interactivity affect social networks [C]// Proceedings of the 2013 Extended Abstracts on Human Factors in Computing Systems. New York: ACM, 2013: 865-870.
- [3] BACKSTROM L, HUTTENLOCHER D, KLEINBERG J, et al. Group information in large social networks: membership, growth, and evolution [C]// Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2006:44-54.
- [4] LIU X, HE Q, TIAN Y, 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. New York: ACM, 2012: 1032-1040.
- [5] QIAO Z, ZHANG P, ZHOU C, et al. Event recommendation in event-based social networks [C]// Proceedings of the 28th AAAI Conference on Artificial Intelligence. Palo Alto: AAAI, 2014: 3130-3131.
- [6] DALY E M, GEYER W. Effective event discovery: using location and social information for scoping event recommendations [C]// Proceedings of the 5th ACM Conference on Recommender Systems. New York: ACM, 2011: 277-280.
- [7] WANG Z, HE P, SHOU L, et al. Toward the new item problem: context-enhanced event recommendation in event-based social networks [C]// Proceedings of the 2015 European Conference on Information Retrieval, LNCS 9022. Cham: Springer, 2015:333-338.
- [8] ZHANG X, ZHAO J, CAO G. Who will attend? — predicting event attendance in event-based social network [C]// Proceedings of the 16th IEEE International Conference on Mobile Data Management. Piscataway: IEEE, 2015:74-83.
- [9] ZHANG W, WANG J. A collective Bayesian Poisson factorization model for cold-start local event recommendation [C]// Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2015: 1455-1464.
- [10] WANG Z, ZHANG Y, LI Y, et al. Exploiting social influence for context-aware event recommendation in event-based social networks [C]// Proceedings of the 2017 IEEE Conference on Computer Communications. Piscataway: IEEE, 2017:1-9.
- [11] SHE J, TONG Y, CHEN L, et al. 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.
- [12] SHE J, TONG Y, CHEN L, et al. Conflict-aware event-participant arrangement and its variant for online setting [J]. IEEE Transactions on Knowledge and Data Engineering, 2016, 28 (9) : 2281-2295.
- [13] SHE J, TONG Y, CHEN L, et al. Feedback-aware social event-participant arrangement [C]// Proceedings of the 2017 ACM SIGMOD International Conference on Management of Data. New York: ACM, 2017:851-865.
- [14] JONES P W. Bandit problems, sequential allocation of experiments [J]. Journal of the Operational Research Society, 1987, 38 (8):773-774.
- [15] AUER P. Using confidence bounds for exploitation-exploration trade-offs [J]. Journal of Machine Learning Research, 2002, 3: 397-422.

This work is partially supported by National Natural Science Foundation of China (61472169, 61502215, 61802160, 51704138), the Key Research and Development Program of Liaoning Province (2017231011), the Shenyang Young Science and Technology Innovation Talents Support Program (RC180244), the Project of the Liaoning Public Opinion and Network Security Big Data System Engineering Laboratory (04-2016-0089013), the Scientific Research Project of Liaoning Provincial Education Department (LYB201617).

SHAN Xiaohuan, born in 1987, Ph. D. candidate, experimentalist. Her research interests include graph data management, database, big data management.

ZHANG Zhiguo, born in 1991, M. S. candidate. His research interests include big data management.

SONG Baoyan, born in 1965, Ph. D., professor. Her research interests include database, RFID event stream processing, big data management, graph data management.

REN Chenglin, born in 1967, senior engineer. His research interests include communication transmission, data processing.