

文章编号: 1003-0077(2019)08-0121-11

基于 ESU 图的活动社交网络用户参加活动推荐

仲兆满^{1,2}, 戴红伟¹, 管燕¹

(1. 淮海工学院 计算机工程学院, 江苏 连云港 222005;

2. 江苏金鸽网络科技有限公司 大数据事业部, 江苏 连云港 222005)

摘要: 活动社交网络(EBSNs)为用户提供了方便的组织、参加和分享社交活动的平台。该文面向 EBSNs 活动推荐问题,提出了包含活动(Event)、主办方(Sponsor)和用户(User)的 ESU 图模型,深入揭示了 EBSNs 的实体及其社交关系。因为用户参加活动受多个因素影响,我们提出了基于 ESU 图的活动推荐多因素决策模型,包括社交影响力、活动内容、活动地点及活动时间。根据 ESU 图特点,提出了基于双向重启随机游走算法 BD-RWR 的实体重要度计算方法。选取真实的 EBSNs 平台—豆瓣同城验证所提方法的有效性。实验结果表明,该文提出的 ESU 图模型及融合了多因素的活动推荐模型,与已有最新方法相比,有效地提升了用户参加活动的推荐效果。

关键词: 活动社交网络;活动推荐;图模型;多因素推荐模型;双向重启随机游走算法

中图分类号: TP391

文献标识码: A

ESU Based Event Recommendation in Event-Based Social Networks

ZHONG Zhaoman^{1,2}, DAI Hongwei¹, GUAN Yan¹

(1. School of Computer, Huaihai Institute of Technology, Lianyungang, Jiangsu 222005, China;

2. Department of Big Data, Jiangsu Jingge Network Technology Co., Ltd.,
Lianyungang, Jiangsu 222006, China)

Abstract: Event-based social networks (EBSNs) provide convenient online platforms for users to organize, attend and share social events. Focusing on recommending upcoming events for users in EBSNs, we present a hybrid social model of Event-Sponsor-User graph by incorporating events, event sponsors (event groups) and users, to best capture the entities and their complex social relations in EBSNs. Due to the fact that users' interests are motivated by a complex set of factors, we propose a model of event recommendation with multiple factors based on ESU graph, which includes social influence, event content, location and time. According to the characteristics of ESU graph, the calculation of entity importance is estimated by bidirectional random walk with restart (BD-RWR). A comprehensive performance evaluation on real-world data sets collected from DoubanEvent shows the proposed method is more effective than state-of-the-art methods.

Keywords: event-based social network; event recommendation; graph model; with multiple factors recommendation model; bidirectional random walk with restart

0 引言

近几年,活动社交网络(Event-based Social Networks,EBSNs)引起了研究者的关注。这主要是因为其通过在线方式为用户提供组织、参加以及

分享社交活动的平台,且用户还可以参加真实的线下活动。比如,Meetup,Plancast,Facebook Event 以及豆瓣同城等等。面对 EBSNs 上,在不同时间和地点举办的诸多活动,用户需要花费大量时间才能寻找到自己感兴趣的活动。在 EBSNs 上,实现活动自动向用户精准推荐,进而让用户丰富业

收稿日期: 2018-09-07 定稿日期: 2018-10-29

基金项目: 国家自然科学基金(61403156);江苏省第五期 333 高层次人才培养工程;江苏高校品牌专业建设工程资助项目(PPZY2015A038);连云港 521 高层次人才培养对象资助项目

余生活、拓展社交关系和享受团队娱乐具有重要的现实意义。

与以往的社交网络不同的是,EBSNs有其独有的特点:①活动多由主办方举办。主要包含活动类型、活动内容、活动时间、活动地点以及对活动感兴趣/参加的人等等信息,活动内涵更加丰富;②EBSNs上有大量实体。包含活动、用户和主办方等,且这些实体之间构建了特有的复杂社交关系;③用户参加活动不仅受活动内容的吸引,而且受社交关系、活动的时间和地点等因素的影响。

已有的EBSNs用户参加活动推荐的研究主要是围绕EBSNs上的活动及用户的属性展开。根据研究成果利用EBSNs上的实体及其关系深度的不同,现将已有工作总结如下:

2012年,Liu等^[1]第一次定义了活动社交网络EBSNs的概念,认为其连接了线上和线下的社交世界,是一种新型的社交媒体。在EBSNs的社群检测时,侧重于用户和活动之间的关联分析。

文献[2-3]利用了EBSNs上朋友间的关系进行活动参加预测。Chin等^[2]调研了用户参加线下活动的行为习惯,发现EBSNs上社交影响存在。并且用户选择参加活动在一定程度上受朋友的影响,但仅仅使用关注构建社交关系。Xu等^[3]验证了EBSNs上朋友间的相互动态影响力对活动的参加有较大的作用,考虑到了用户的偏好和基于朋友的社交关系。

Jamali等^[4]使用社交矩阵分解技术MF(Matrix Factorization),提出了整合用户—活动参加结果和简单社交关系的SocialMF方法,社交影响力的计算仅考虑了用户和活动间的关系。

Li等^[5]讨论了社交活动组织(SEO)问题,而不是活动的检测,但其仅仅考虑了属性间的相似度及用户间的朋友关系两个因素。

文献[6-7]以豆瓣同城为例,研究了EBSNs的活动参加预测问题,侧重于邀请有影响力的关注者(followers)。Yu等^[6]扩展了信任分布模型识别最具影响力的被邀请者,考虑了用户对活动的偏好及社交影响最大化。Du等^[7]提出了奇异值分解方法用于活动参加的预测,考虑到了活动的内容、上下文(时空)及社交影响。

Zhang等^[8]研究EBSNs活动参加预测问题的主要思想是:如果一个用户对一个主题的活动感兴趣,他可能参加与这个主题相关的新活动。对用户参加活动的分析包括语义、时间和空间特征。

Li等^[9]提出了混合协作过滤模型MF-EUN,融合了活动和用户的邻居。为了解决社交影响矩阵的稀疏性,提出了基于附加信息的邻居发现方法。其对EBSNs的用户之间、用户与活动之间、活动与举办方之间的建模不够准确,且未考虑时间要素对活动推荐的影响。

Qiao等^[10]提出了贝叶斯潜在因素模型用于活动推荐,该模型融合了异构社交关系、地理特征和潜在排名。Zhang等^[11]侧重于EBSNs上的冷启动活动的推荐研究。

Liu等^[12]提出了上下文(静态的和动态的)感知推荐方法SoCo,使用随机决策树用于用户-活动分组,利用朋友关系推导出用户的兴趣偏好,但其构建的社交网络图仅仅是用户之间的。

Tong等^[13]提出了瓶颈感知社交活动安排模型(BSEA)用于EBSNs的全局推荐,使用贪婪、随机贪婪和基于局部搜索的最优化技术解决BSEA的NP问题。

Müngen等^[14]将Meetup社交网络形式化为二部图,重点利用用户和活动的地点进行活动推荐。

Liu等^[15]将豆瓣社交网络形式化为9元组混合网络,包括用户U、活动E、群组G,发起者H和标签T。其中,活动E包括活动时间Em、地点El、成本Ec、类型Et。使用带重启的随机游走算法预测用户参与的活动,并考虑了用户和活动的相似性。

已有研究存在的不足及本文的创新点:①EBSNs的建模不够准确,多强调用户与活动之间的关系描述,例如,文献[1,4,8],而忽略了用户与主办方、活动与主办方之间的关系。本文提出的ESU图模型,更精准的揭示了EBSNs的实体及其复杂社交关系,对面向EBSNs的后续研究有重要的参考价值;②用户是否参加活动是受多个因素影响的,已有方法对这些因素利用的不够,比如文献[2-4,6-7,9]。本文提出的基于ESU图融合了活动对用户的社交影响力、活动内容、地点和时间多因素的活动推荐模型,能全面的捕捉EBSNs上活动推荐的本质特征;③由于对EBSNs建模多侧重用户和活动的关联,因此活动与用户之间的影响力计算多基于矩阵分解技术,未能考虑到活动主办方的作用。例如,文献[4,6,7,9]。本文根据ESU图实体独有的特点,提出了基于双向重启随机游走算法BD-RWR的实体重要度计算方法,能更好的利用ESU图的整体结构,更合理地计算不同实体的影响力。

1 ESU 图模型

在 EBSNs 中,实体主要包括活动、主办方和用户,这三者之间存在复杂的社交关系,每个实体又包含了多种属性。全面地理清这些实体之间的关系及每个实体包含的属性,是面向 EBSNs 开展后续研究工作的基础。

本文在深入调研 EBSNs 特点的基础上,提出的活动—主办方—用户 ESU(Event-Sponsor-User)图模型,如图 1 所示。

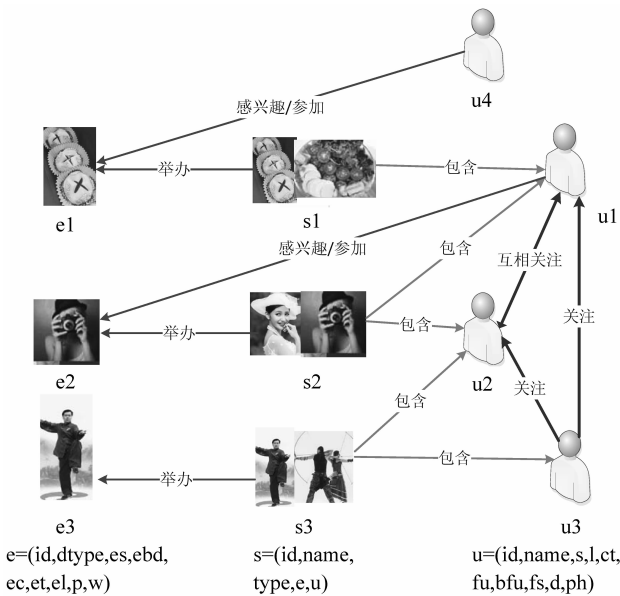


图 1 ESU 图模型

图 1 所示的 ESU 图模型中,节点 e 代表活动(图左), s 代表活动主办方(图中)和 u 代表用户(图右),图下方是对每个实体的属性描述。主办方 s 会举办多个活动,与其举办的活动是单向关联;主办方 s 会包含多个用户,与其包含的用户是单向关联;用户之间通过关注(follow)建立单向关联,两者之间如果相互关注,则形成双向关联,并且双向的权重不同;用户 u 可以感兴趣/参加活动 e ,用户与活动是单向关联。ESU 图具有较高的稀疏性,体现在以下 4 点:(1)活动之间没有连接关系;(2)主办方之间没有连接关系;(3)用户之间双向关系较少;(4)一些用户可能不属于任何的举办方,不参加任何活动。

ESU 图模型的相关概念描述如下:

定义 1 ESU 图,描述为一个二元组: $ESU = (Entity, Relation)$ 。其中,Entity 代表实体,包括活动集 E 、主办方集 S 和用户集 U ;Relation 代表关系,有

主办方-活动关系集 SE 、主办方-用户关系集 SU 、用户-活动关系集 UE 以及用户之间关系集 UU 。

定义 2 ESU 图活动 e ,描述为一个九元组:

$e = (id, etype, es, ebd, ec, et, el, p, w)$ 。其中, id 为活动唯一标识, $etype$ 是活动类型标签, es 为活动发起人, ebd 为活动简述,类似于活动名称, ec 为活动详情, et 为活动时间, el 为活动地点, p 为要参加活动的用户, w 为对活动感兴趣的用户。

定义 3 ESU 图主办方 s ,描述为一个五元组:

$s = (id, name, type, e, u)$ 。其中, id 为主办方唯一标识, $name$ 为主办方名称, $type$ 为主办方的活动类型, e 为举办的活动, u 为关注主办方的用户。不同的 EBSNs 对活动举办者的称谓有所不同,常见的有社区、社群、组织者、举办方、主办方等等,本文统一称为主办方。

定义 4 ESU 图用户 u ,描述为一个十元组:

$u = (id, name, s, l, ct, fu, bfu, fs, d, ph)$ 。其中, id 为用户唯一标识, $name$ 为用户名, s 为用户兴趣签名, l 为用户常居地, ct 为账号创建的时间, fu 为他关注的用户, bfu 为关注他的用户, fs 为他关注的主办方, d 为用户日记, ph 为用户相册。

定义 5 ESU 图主办方—活动关系集,描述为 $SE = \{se = (s_i, e_j) \mid s_i \rightarrow e_j \mid s_i \in S, e_j \in E\}$,主办方 s_i 通过“举办”的方式与活动 e_j 建立单向关系。

定义 6 ESU 图主办方—用户关系集,描述为 $SU = \{su = (s_i, u_j) \mid s_i \rightarrow u_j \mid s_i \in S, u_j \in U\}$,主办方 s_i 通过“包含”的方式与用户 u_j 建立单向关系。

定义 7 ESU 图用户—用户关系集,描述为 $UU = \{uu = (u_i, u_j) \mid u_i \rightarrow u_j \mid u_i \in U, u_j \in U\}$,用户 u_i 和 u_j 通过“关注”的方式建立单向关系,如果相互关注,则建立双向关系(好友关系)。

定义 8 ESU 图用户—活动关系集,描述为 $UE = \{ue = (u_i, e_j) \mid u_i \rightarrow e_j \mid u_i \in U, e_j \in E\}$,用户 u_i 通过“感兴趣/参加”的方式与活动 e_j 建立单向关系。

基于 ESU 图,在全面利用实体及其关系的基础上,可以进行重要活动、主办方及用户的挖掘,用户参加活动倾向性的挖掘,活动社群检测,信息流分析,活动到用户的推荐或预测,用户的活动行为分析及主办方举办活动的行为分析等等系列研究。

2 基于 ESU 图的用户参加活动推荐

2.1 活动推荐多因素决策模型

定义 9 用户参加活动推荐,描述为 $f(EF)$

$(e_i), EF(u_j)) \Rightarrow YN(e_i, u_j)$, $EF(\cdot)$ 为特征抽取函数, $f(\cdot)$ 为根据活动 e_i 和用户 u_j 的特征决策是否将活动 e_i 推荐给用户 u_j , $YN(e_i, u_j)$ 为决策结果, 其值为“是(Yes)”和“否(No)”。

在 EBSNs 上, 用户 u_j 参加活动受到多种因素的影响, 比如用户 u_j 的好友参加了某个活动 e_i , 则用户 u_j 可能会参加活动 e_i ; 用户 u_j 经常参加主办方 s_k 的活动。当主办方 s_k 再次举办活动时, 用户 u_j 可能参加 s_k 举办的活动, 这都体现为社交影响因素。当一个新活动 e_i 到来时, 如果活动 e_i 与用户 u_j 曾经参加过的活动的相似度非常高, 用户 u_j 可能参加活动 e_i , 这体现为活动的内容因素。但 EBSNs 上活动的时间和地点与用户参加活动的地点和时间又有着密切关系。比如, 用户 u_j 习惯周末参加聚会活动, 即使活动内容非常接近, 但时间不吻合, 用户 u_j 也不一定有时间参加; 又如用户 u_j 经常参加某个地点的活动, 即使活动内容非常接近但地点相距较远, 用户 u_j 也不一定愿意参加远距离的活动。

基于 EBSNs 的 ESU 图模型, 本文提出活动推荐多因素决策模型, 如图 2 所示。

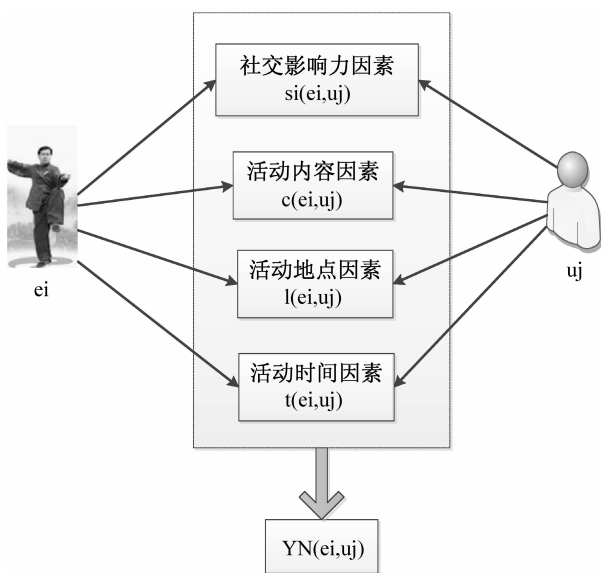


图 2 基于 ESU 图的活动推荐多因素决策模型

图 2 所示模型中, 活动 e_i 通过 ESU 图的活动—主办方—用户关系对用户 u_j 产生的社交影响力记为 $si(e_i, u_j)$, e_i 的活动内容与用户 u_j 曾经参加过的活动的内容的相关性记为 $c(e_i, u_j)$, e_i 的活动地点与用户 u_j 曾经参加过的活动地点的相关性记为 $l(e_i, u_j)$, e_i 的活动时间与用户 u_j 曾经参加过的活动的时间的相关性记为 $t(e_i, u_j)$, 最终在四个因素的基础上得到活动 e_i 对用户 u_j 的推荐结果为 $YN(e_i, u_j)$ 。

下面详细的介绍活动推荐多因素决策模型中的四个因素的计算方法。

2.2 社交影响力因素

ESU 图包含活动、主办方和用户三类实体, 以及实体之间的各种关系, 活动对用户的影响力受三类实体的共同作用。

对于构建了活动和用户二者之间的关系模型, 使用潜在因子模型可以直接挖掘活动与用户之间的潜在主题, 包括矩阵分析、概率主题模型及 LDA 等技术。而对于包括了多个实体及其关系的 ESU 图模型而言, 随机游走算法具有明显的优势。随机游走算法具有可解释性强、缓解稀疏性、逻辑简洁、易于实现等优点, 其在信息检索和推荐系统中已得到了广泛的应用。

为了计算 ESU 图上活动对用户的影响力, 需要计算主办方—活动、主办方—用户、用户—用户以及用户—活动的关系权重, 进而选取合适的随机游走算法计算活动 e_i 对不同用户的影响力。

2.2.1 主办方—活动关系权重

令 $UN(s_j)$ 为主办方 s_j 包含的用户数量, $UN(e_i)$ 为 s_j 中对活动 e_i 感兴趣/参加的用户数量。则主办方 s_j —活动 e_i 的关系权重如式(1)所示。

$$se(s_j, e_i) = \frac{UN(e_i)}{UN(s_j)} \quad (1)$$

式(1)利用了主办方 s_j 中参加活动 e_i 的用户数量, 及 s_j 中所有的用户数量度量主办方对活动的关系权重。 $se(s_j, e_i)$ 值越大, 说明主办方 s_j 中有越多的用户参与 s_j 举办的活动。

2.2.2 主办方—用户关系权重

令 $SN(u_i)$ 为用户 u_i 加入的主办方数量, $EN(s_j)$ 为主办方 s_j 曾经举办的活动数量, $SEN(u_i)$ 为用户 u_i 曾经参加的主办方 s_j 的活动数量, 则主办方 s_j —用户 u_i 的关系权重如式(2)所示。

$$su(s_j, u_i) = \frac{SEN(u_i)}{EN(s_j)} * \ln \frac{1}{SN(u_i)} \quad (2)$$

式(2)前半部分说明用户 u_i 参加主办方 s_j 的活动越多, 主办方 s_j 对用户 u_i 的影响越大; 式(2)后半部分考虑到了用户 u_i 加入的其他主办方数量, 直观的来讲, 一个用户参加的主办方越多, 主办方 s_j 对该用户的影响力就越小。

2.2.3 用户—用户关系权重

令 $E(u)$ 为用户 u 参加的活动集合, $E(v)$ 为用户 v 参加的活动集合, $S(u)$ 为用户 u 加入的主办方

集合, $S(v)$ 为用户 v 加入的主办方集合, 则用户 u -用户 v 的关系权重如式(3)所示。

$$uv(u, v) = \alpha \frac{|E(u) \cap E(v)|}{|E(u)|} + \beta \frac{|S(u) \cap S(v)|}{|S(u)|} \quad (3)$$

式(3)综合的考虑了用户 u 和用户 v 参加的共同活动和加入的共同主办方的情况, α 和 β 是平衡共同活动和共同主办方之间权重的参数。

需要注意的是, 反过来, 用户 v 对用户 u 的权重是不同的, 因为 v 和 u 可能对不同数量的主办方感兴趣, 也可能参加不同数量的活动, 但计算方法和式(3)类似。

2.2.4 用户—活动关系权重

令 $E(u_j)$ 为用户 u_j 参加过的活动集合, s_j 为活动 e_i 所在的主办方, $ES(u_j)$ 为用户 u_j 参加过 s_j 举办的活动集合。则用户 u_j -活动 e_i 的关系权重如式(4)所示。

$$ue(u_j, e_i) = \frac{|ES(u_j)|}{|E(u_j)|} \quad (4)$$

式(4)说明用户 u_j 参加过活动 e_i 所在的主办方的活动越多, 用户 u_j 对活动 e_i 的影响就越大。

2.2.5 基于双向重启随机游走算法的节点影响力计算

基于随机游走的 PageRank 算法通过网页间的超链接分析, 计算每个网页的重要性。计算方法如式(5)所示。

$$R_k = dWR_{k-1} + \frac{1-d}{N}I \quad (5)$$

其中, R_k 为第 k 次迭代时页面的重要度, W 表示页面之间的连接权重, d 是一个阻尼系数, 取值范围为 0-1, 通常取 $d=0.85$, 向量 $I=(1, 1, \dots, 1)^T$, N 为网页的数量。

重启型随机游走算法 RWR^[15] 主要用于计算图中任意两个节点间的结构相关性。计算图中各个节点与节点 j 的结构相关性方法如式(6)所示。

$$R_k = cWR_{k-1} + (1-c)e_j \quad (6)$$

其中, $1-c$ 为返回节点 e_j 的概率, e_j 为第 j 维为 1 的单位向量, 初始时 $R_0=e_j$, R_k 和 W 的含义和式(5)相同。

ESU 图为有向图, 三类实体的重要性体现方式不同。主办方 s_i 的重要度取决于 s_i 举办的所有活动及包含的所有用户对其产生的影响, 因此 s_i 的计算需要考虑其出度(Hubs), PageRank 考虑的是页面的入度。活动 e 节点的重要度取决于主办方以及

用户对其产生的影响。用户 u 节点的重要度取决于主办方和关注他的用户对其产生的影响。因此, ESU 图的活动和用户重要度的计算, 应该考虑其入度(Authorities)。

因此, 在计算 ESU 各类节点的重要度时, 需要从出度和入度两个角度出发进行迭代计算。为此, 我们提出了 ESU 图上基于双向重启随机游走算法(BD-RWR)的节点重要度计算方法。活动主办方的重要度、用户和活动的重要度迭代计算分别如式(7)~式(9)所示。

$$S_k = c(W_{SE}E_{k-1} + W_{SU}U_{k-1}) + (1-c)se_j \quad (7)$$

$$U_k = c(W_{SU}S_{k-1} + W_{UU}U_{k-1}) + (1-c)ue_j \quad (8)$$

$$E_k = c(W_{SE}S_{k-1} + W_{UE}U_{k-1}) + (1-c)ee_j \quad (9)$$

其中, se_j 、 ue_j 和 ee_j 分别为第 j 维为 1 的单位向量。 $1-c$ 为返回节点 se_j 、 ue_j 和 ee_j 的概率。 W_{SE} 、 W_{SU} 、 W_{UU} 和 W_{UE} 分别为主办方—活动、主办方—用户、用户—用户和用户—活动的关系权重。

经过若干次迭代后, 得到每个用户 u_i 的节点重要度。节点重要度越大, 说明活动 e_i 在 ESU 图上对用户产生的社交影响力越大, 用户越有可能参加活动 e_i 。

为保证随机游走算法迭代结果的收敛, 不同算法采用了不同的策略。经典的 PageRank 在计算节点重要度时引入阻尼系数 d , 确保迭代过程中不会出现节点重要度为 0 的情况, 即认为图是强连通的, 对应的矩阵是不可约的。对于重启型随机游走算法 RWR 而言, 通过添加返回起始节点的概率, 同样达到了迭代结果收敛的目的。

定理 1 双向重启随机游走算法 BD-RWR 是收敛的。

证明: 在计算主办方的重要度时, 考虑的是节点出度, 添加了返回到自身节点概率 $(1-c)$ 。式(7)的转移概率矩阵 W_{SE} 和 W_{SU} 分别通过 2.2.1 节中的主办方—活动权重、2.2.2 节中主办方—用户权重计算方法得到, 在迭代过程中保持不变。活动节点初始权重 $E_0=1/|E|$, E 是活动集合, 用户节点初始权重 $U_0=1/|U|$, U 是用户集合。式(7)的主办方重要度计算转化为单向的重启型随机游走算法, 只是与传统的随机游走算法的方向相反而已。可见, 基于出度的主办方重要度计算的迭代过程是收敛的。

在计算活动和用户节点的重要度时, 考虑到节点入度, 添加了返回到自身节点概率 $(1-c)$ 。式(8)的转移概率矩阵 W_{SU} 和 W_{UU} 分别通过 2.2.2 节中的主办方—用户权重、2.2.3 节中的用户—用户

权重计算方法得到。式(9)的转移概率矩阵 W_{SE} 和 W_{UE} 分别通过 2.2.1 节中的主办方—活动权重、2.2.4 节中的用户—活动权重计算方法得到。主办方节点初始权重 $S_0 = 1/|S|$, S 是主办方集合。式(8)的用户重要度计算和式(9)的活动重要度计算转化为单向重启型随机游走算法。可见,基于入度的活动和用户重要度计算的迭代过程是收敛的。

由以上分析可见,双向重启随机游走算法 BD-RWR 是收敛的。

2.3 活动内容因素

给定一个特定活动 e_i 后,面向用户推荐时,需要考虑 e_i 的活动内容与用户 u_j 曾经参加过活动的相似性。假设用户 u_j 曾经参加过的活动集合记为 $E(u_j)$ 。

已有方法在计算活动 e_i 与用户参加过的活动 $E(u_j)$ 的相似度时,多是使用 LDA 模型从活动中提取主题,进而使用余弦相似度度量活动的相似性。

但由于 EBSNs 活动资料往往比较简练,这加重了文本资料的稀疏性,不容易从文本中准确的提取出主题。而且把所有的活动混在一起提取主题,容易把用户参加量小的活动淹没。

对从豆瓣同城采集到的 17 822 个活动(实验 3.1 节)进行统计发现,其中 85.7% 的活动都带有标签,这为活动的相似度计算提供了便利。

使用活动的类型标签计算活动的相似性,活动 e_i 与 $E(u_j)$ 的相似度计算方法,如式(10)所示。

$$c(e_i, u_j) = c(e_i, E(u_j)) = \frac{|CF(e_i) \cap CF(E(u_j))|}{|CF(e_i) \cup CF(E(u_j))|} \quad (10)$$

其中, $CF(e_i)$ 表示活动 e_i 的类型标签集合, $CF(E(u_j))$ 表示用户 u_j 参加过活动的类型标签集合。

使用活动标签的方法计算活动间的相似性不仅计算量小,而且效果也较好。实验 3.6 节的结果表明,从活动内容的角度出发,使用 LDA 模型与使用活动标签相比, $F1$ -measure 仅仅相差 0.01。

2.4 活动地点因素

活动地点指活动线下举办的真实地点。和传统社交媒体不同的是,线下活动的举办是 EBSNs 独有的特点。因此,地点因素对决定是否向用户推荐活动有重要的作用。

基于位置的社交网络(Location-based Social Networks, LBSNs)相关的研究[引用]已经表明,随

着活动地点与用户地点距离的增加,用户参加活动的可能性会降低。通过对采集的豆瓣同城用户参加活动地点的统计发现,用户所在地点与活动举办地点服从幂率分布。即大多数用户参加的活动,其地点都是离用户地点是比较近的,在对数坐标下表现为一条直线。

图 3 显示了北京城市用户参加活动与用户地点之间的关系(数据描述详见实验 3.1 节,对数坐标)。

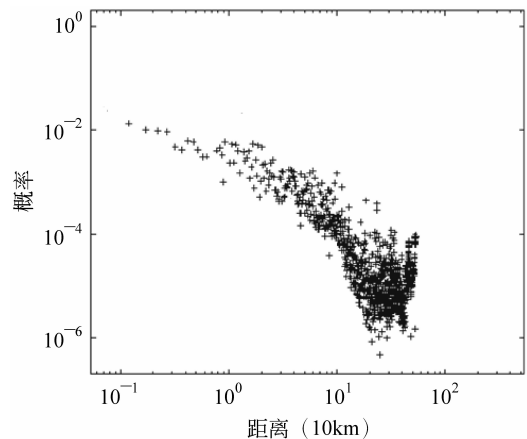


图 3 用户参加活动的距离概率分布

活动 e_i 地点与用户 u_j 地点的相关性计算如式(11)所示。

$$l(e_i, u_j) = \exp \left\{ -\frac{dis(l_{e_i}, l_{u_j})^2}{2} \right\} \quad (11)$$

式(11)中, l_{e_i} 表示活动 e_i 的地点, l_{u_j} 表示用户 u_j 的地点, $dis(l_{e_i}, l_{u_j})$ 表示 l_{e_i} 与 l_{u_j} 的距离。地点都有一个经度和纬度的数值,容易计算两个地点的距离。

有的 EBSNs 在用户注册时,要求通过定位的方式输入用户的地点。例如,Meetup;有的 EBSNs 在用户注册时,地点是用户选取的,而且只是到了地市级别,不够具体,比如豆瓣同城。而用户在注册时通过定位方式输入的注册地点,并不一定就是用户的工作或者居住地点。因此,我们使用用户曾经参加过的活动地点估算用户的具体地点。

EBSNs 上举办的活动地点一般都比较具体。例如“东城区,东直门南大街 14 号,保利剧院”,通过地图定位容易确定活动的经纬度值。假设用户参加过的活动集合为 u_j , $E = \{e_1, e_2, \dots, e_m\}$ 。每个活动的地点用经纬度表示,记为 u_j , $e_k = (lon, lat)$ ($1 \leq k \leq m$),用户参加过活动的中心地点记为 $l_{u_j} = (lon, lat)$ 。 u_j , E 包含活动的纬度的最大和最小值计算分别如式(12)、式(13)所示。

$$u_j.E.maxlat = \max(u_j.E.e_1.lat, u_j.E.e_2.lat, \dots, u_j.E.e_m.lat) \quad (12)$$

$$u_j.E.minlat = \min(u_j.E.e_1.lat, u_j.E.e_2.lat, \dots, u_j.E.e_m.lat) \quad (13)$$

$u_j.E$ 包含活动经度的最大和最小值计算分别如式(14)、式(15)所示。

$$u_j.E.maxlon = \max(u_j.E.e_1.lon, u_j.E.e_2.lon, \dots, u_j.E.e_m.lon) \quad (14)$$

$$u_j.E.minlon = \min(u_j.E.e_1.lon, u_j.E.e_2.lon, \dots, u_j.E.e_n.lon) \quad (15)$$

则中心地点 $l_{u_j} = (lon, lat)$ 的经度和纬度的计算方法分别如式(16)、式(17)所示。

$$l_{u_j}.lon = (u_j.E.maxlon - u_j.E.minlon)/2 + u_j.E.minlon \quad (16)$$

$$l_{u_j}.lat = (u_j.E.maxlat - u_j.E.minlat)/2 + u_j.E.minlat \quad (17)$$

2.5 活动时间因素

人们参加活动受限于时间因素的影响,一般表现为天和周的周期性。例如,习惯每天下班后参加活动,或者每个周末参加活动。因此,时间因素对用户能否参加活动有重要的影响。

将用户曾经参加过的与活动 e_i 同类型的活动集合记为 $E(u_j)$ (每个活动都有类型标签),活动 e_i 的举办时间记为 t_{e_i} ,用户参加过的同类型活动 $e_k' \in E(u_j)$ 的举办时间记为 $t_{e_k'}$,则活动 e_i 的时间与同类型活动 e_k' 的时间偏差如式(18)所示。

$$t(e_i, e_k') = e^{-\frac{(t_{e_i} - t_{e_k'})^2}{2}} \quad (18)$$

活动 e_i 与用户 u_j 的时间偏差计算方法如式(19)所示。

$$t(e_i, u_j) = \min(t(e_i, e_1'), t(e_i, e_2'), \dots, t(e_i, e_{|E(u_j)|}')) \quad (19)$$

式(19)在计算活动 e_i 与用户 u_j 的时间偏差时,一方面考虑活动的类型,因为不同的时间,用户参加的活动是不同的。例如,晚上参加聚会,而周末参加旅游;另一方面,使用用户 u_j 参加过的活动中与 e_i 举办时间最靠近的活动时间偏差作为度量标准,可以避免求平均值带来的偏差。

在计算活动时间的偏差时,不考虑年、月和日的

差别,只考虑小时的偏差情况。例如,活动 e_i 的举办时间 t_{e_i} = “2017 年 7 月 1 日 15 点”,活动 e_k' 的举办时间 $t_{e_k'}$ = “2017 年 7 月 23 日 14 点”,则 $(t_{e_i} - t_{e_k'})^2 = (15 - 14)^2 = 1$ 。

3 实验及分析

3.1 实验数据及评价指标

本文选取豆瓣同城的数据作为实验语料。豆瓣同城拥有超过 3 千万用户,且线下活动的开展活跃度也很高。近期的不少研究成果也都使用豆瓣作为实验数据的来源,包括文献[2,6,7,9,12]。

使用豆瓣提供的开放接口“API V2”获取采集数据。例如,通过接口“https://api.douban.com/v2/event/28684165”可以获取一个活动 JSON 格式的信息。在采集用于分析的数据的时间段上,Zhang 等^[8]实验结果表明,使用 60 天的数据进行分析效果已经足够好。因此,我们采集了豆瓣同城北京市 2017 年 3 月 1 日—2017 年 4 月 30 日,共计 61 天的数据。为了避免一些不活跃用户或者活动对总体实验结果的影响,删除其中的一些非常不活跃的用户及活动。感兴趣/参加活动少于 5 个的用户认为是冷启动用户,占有用户总数的 9%;如果某个活动感兴趣/参加的人数少于 8 个就认为是冷启动活动,占有活动总数的 6%。经过对 2 个月的数据进行预处理之后,共得到 8 663 个用户,17 822 个活动,294 个主办方。

在 EBSNs 中,用户对于某个活动用参加或者不参加来表示,可以将活动参加问题转换为一个二分类问题。实验时,随机选取数据集中的 80% 用于训练,其他的 20% 用于测试。

实验方法通过 WEKA 平台实现,WEKA 是一个基于 JAVA 公开的数据挖掘工作平台,集成了大量数据挖掘和机器学习算法,包括数据的预处理、分类、回归、聚类、关联规则和交互式界面上的可视化等等。

使用 $F1\text{-measure} = (2 \times P \times R) / (P + R)$ 作为评测指标,其中 P 是推荐用户参加活动的准确率, R 是推荐用户参加活动的召回率。

3.2 实验方法

本文共选用了 6 种实验方法用于实验的分析和比较,分别介绍如下:

- 基本的矩阵分解方法,它只考虑用户-活动

的参加结果,矩阵中元素的权值使用本文式(4)计算,简记为 MF;

- 使用了矩阵分解相关的信任传播,整合了用户-活动参加结果和简单社交关系,用户社交关系的构建使用本文提出的式(3),类似于文献[4]提出的方法,简记为 SocialMF,参数设置, $\lambda_u=\lambda_v=0.1, \lambda_T=5$;
- 文献[9]提出的方法,简记为 MF-EUN-ER,影响力的阈值设置为 0.5,活动的邻居 $k_1=50$,用户的邻居 $k_2=60$,区域聚类个数 $t=20$;
- 由于决策树方法可读性好,有助于人工分析,所以使用经典的 J48 决策树方法检验本文所提 4 个因素的效果。四个因素的计算方法如第 2 节所述,其中用户间关系权重参数 $\alpha=\beta=0.5$ (经过实验确定),简记为 F4-DT-ER;
- 本文所提方法的变形,活动内容主题的提取使用 LDA 模型,其他因素计算方法不变,参考文献[9],对 LDA 涉及的参数分别设置为 $\alpha=50/k, \beta=0.01$, k 为主题个数, $k=70$,简记为 F4-C-DT-ER;
- 本文所提方法的变形,社交影响力的计算使用经典的 PageRank 方法,其他因素计算方法不变。类似于文献[6-7]所提方法,简记为 F4-SI-DT-ER。

3.3 总体推荐结果

对于用户参加活动的推荐总体结果,使用四种方法进行比较,分别是 MF、SocialMF、MF-EUN-ER 和 F4-DT-ER。四种方法获取的结果如表 1 所示。

表 1 四种方法获取的总体推荐结果

方法	P	R	$F1\text{-measure}$
MF	0.68	0.66	0.67
SocialMF	0.73	0.71	0.72
MF-EUN-ER	0.82	0.79	0.80
F4-DT-ER	0.87	0.85	0.86

由表 1 可见,本文提出的方法 F4-DT-ER 获取的结果最为理想。 $F1\text{-measure}=0.86$,其次是 MF-EUN-ER 方法, $F1\text{-measure}=0.80$ 。四种方法中, MF 的效果最差, $F1\text{-measure}=0.67$ 。主要原因是其仅仅从用户和活动之间的关系出发,忽略了 EBSNs 上的很多其他关系,类似于传统用户-项目

之间的推荐方法,这也说明传统的推荐方法用于 EBSNs 的用户参加活动推荐效果是比较差的。方法 SocialMF 的效果有所提升, $F1\text{-measure}=0.72$,原因是其考虑到了社交关系,考虑到了用户间的信任传播,但其仅仅是考虑了用户之间的社交关系,对其他实体间的关系利用不够。方法 MF-EUN-ER 根据 EBSNs 具有的特点,考虑到了活动内容、活动地点及社交关系,得到的结果有了很大程度的改善, $F1\text{-measure}$ 值比 NMF 提高了 0.13,比 SocialMF 提高了 0.08。本文所提方法 F4-DT-ER 深入的分析利用了 EBSNs 独有的社交关系特点,提出了 ESU 图的用户影响力计算方法,很好的利用了活动的内容、地点及时间因素,取得了最好的效果。

这说明,对 EBSNs 而言,需要抓住其独有的特点,综合的考虑多个方面的因素,才能取得更好的用户参加活动推荐结果。而本文所提的 ESU 图表示模型、社交影响力计算方法、活动推荐多因素决策模型都是 EBSNs 所独有的特点。

3.4 不同特征的贡献

为了检验 4 类因素对用户参加活动推荐结果的影响,使用经典 J48 决策树,对 4 类因素进行考察。仅仅使用内容因素 C、空间因素 L、时间因素 T 和社交影响因素 SI 的方法,分别记为 C-DT-ER、L-DT-ER、T-DT-ER 和 SI-DT-ER。不考虑社交影响因素 SI,只使用内容因素 C、空间因素 L 和时间因素 T,该方法记为 CLT-DT-ER。只使用社交影响因素,不考虑内容因素 C、空间因素 L 和时间因素 T,该方法记为 SI-DT-ER。共计使用 6 种方法进行实验的对比和分析。

对每种方法,进行 4 次交叉检验,最后取平均值。6 种方法得到结果如表 2 所示。

表 2 六种方法获取的推荐结果

方法	P	R	$F1\text{-measure}$
T-DT-ER	0.34	0.31	0.32
L-DT-ER	0.36	0.31	0.33
C-DT-ER	0.58	0.53	0.55
SI-DT-ER	0.7	0.68	0.69
CLT-DT-ER	0.73	0.72	0.72
F4-DT-ER	0.87	0.85	0.86

由表 2 可见,方法 T-DT-ER、L-DT-ER、C-DT-ER 和 SI-DT-ER 都使用单一的因素进行用户参加

活动推荐,效果最好的是 SI-DT-ER, $F1\text{-measure}$ 为 0.69。这说明在 EBSNs 上,有效的利用 ESU 图的社交关系,进行合理的影响力计算便可以取得较好的效果。单一的使用时间 T 效果最差, $F1\text{-measure}$ 为 0.32。因为时间 T 虽然靠近,但活动的内容用户不一定感兴趣,地点可能离用户较远,不方便参加。单一的使用地点 L 效果也很差, $F1\text{-measure}$ 为 0.33。因为地点 L 虽然靠近,但活动的内容用户不一定感兴趣,用户不一定有时间参加。单一的使用内容 C 效果有所提升, $F1\text{-measure}$ 为 0.55,这说明内容因素发挥的作用较大。方法 CLT-DT-ER 使用了内容 C 、地点 L 和时间 T 因素, $F1\text{-measure}$ 为 0.72,效果已经较好。方法 F4-DT-ER 有效的利用了 EBSNs 上的四类因素,取得了最好的效果。

3.5 社交影响力计算的影响

本文在 EBSNs 上,根据 ESU 图实体独有的入度和出度的特点,分实体类型使用双向重启随机游走算法迭代计算节点的重要度。为了检验本文所提社交影响力计算方法的有效性,使用两种方法进行实验分析和比较,分别是 F4-DT-ER 和 F4-SI-DT-ER。

对每种方法,进行四次交叉检验,最后取平均值。方法 F4-DT-ER 和 F4-SI-DT-ER 得到的推荐结果如表 3 所示。

表 3 方法 F4-DT-ER 和 F4-SI-DT-ER 获取的推荐结果

方法	P	R	$F1\text{-measure}$
F4-SI-DT-ER	0.76	0.75	0.75
F4-DT-ER	0.87	0.85	0.86

由表 3 可见,方法 F4-SI-DT-ER 的效果也比较理想, $F1\text{-measure}$ 为 0.75。PageRank 方法强调整节点的入度,在 ESU 图上,对活动、用户两类实体计算是合理的。但对主办方而言,其实体影响力的计算应该使用其出度。方法 F4-DT-ER 在计算节点影响力时,采用分而治之的策略,对不同实体采用不同的计算方法,体现出了 ESU 图独有的社交特性,因此取得的效果最好。

3.6 内容特征提取的影响

活动内容主题的提取,对基于内容进行活动推荐有重要的影响。已有的方法多是使用 LDA 模型提取活动主题,本文根据 EBSNs 上的活动标签描述

比较丰富(约 85.7%的活动有标签)的特点,使用标签计算活动之间的相似性。为了检验所提方法的有效性,使用两种方法 F4-DT-ER 和 F4-C-DT-ER 进行实验比较。对采集到的 17 822 个活动,删除没有标签的活动,还剩下 15 273 个活动。对这些活动,同样是随机选取 80%进行训练,20%进行测试。

对每种方法,进行四次交叉检验,最后取平均值。方法 F4-DT-ER 和 F4-C-DT-ER 得到的推荐结果如表 4 所示。

表 4 方法 F4-DT-ER 和 F4-C-DT-ER 获取的推荐结果

方法	P	R	$F1\text{-measure}$
F4-C-DT-ER	0.86	0.85	0.85
F4-DT-ER	0.87	0.85	0.86

由表 4 可见,方法 F4-C-DT-ER 的效果非常理想, $F1\text{-measure}$ 为 0.85。这说明使用 LDA 模型提取活动的主题是能够获取较好的效果,相比方法 F4-DT-ER 使用活动的标签仅相差 0.01。但基于 LDA 模型的活动内容相关度计算的复杂度远远高于基于活动标签的方法。因此,对 EBSNs 而言,使用活动标签计算内容相似度,不但计算工作量小,而且能够获取较好的效果。

4 总结及展望

本文以活动社交网络 EBSNs 的用户参与新活动推荐为出发点,在总结已有研究方法的优点及不足的基础上,提出了 EBSNs 的活动—主办方—用户(ESU)图表示模型,揭示了 EBSNs 上的实体及其复杂的关联关系。由于用户参加活动受多方面因素的影响,在 ESU 图的基础上,提出了活动推荐多因素决策模型。进而,根据 ESU 图的特点,提出了分实体类型的基于双向重启随机游走算法的节点重要度计算方法。选取豆瓣同城进行了数据的采集、推荐模型、各个因素计算方法的对比分析。

虽然本文在 EBSNs 的用户参加活动推荐方向取得了一定的进展,但针对如下问题还需要进一步提升:①本文的活动推荐面向的主要是 EBSNs 的活跃用户,冷启动用户参与活动的推荐需要借鉴已有的方法并结合 EBSNs 独有的特点加以解决;②EBSNs 活动内容的描述体现在活动名称(活动概述)、活动详情、活动标签、活动类型等等方面,如何全面的衡量活动之间的相似度需要进一步研究;

③用户在 EBSNs 上,除了有与活动相关的信息外,还有用户喜欢、用户日记、用户相册、用户评论等等信息。这些信息对用户参加活动的推荐作用如何,需要进一步研究;④面向 EBSNs,除了进行用户参加活动的推荐研究之外,其他内容也需要深入研究。含主办方和用户行为分析,社区检测(含重叠社区),社区及用户的倾向性分析等等。

致谢

感谢江苏金鸽网络科技有限公司为本研究提供的实验数据集。

参考文献

- [1] 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 conference on Knowledge Discovery and Data Mining, 2012: 1032-1040.
- [2] Chin A, Tian J, Han J, et al. A study of offline events and its influence on online social connections in Douban [C]//Proceedings of the GreenCom and iThings/CP-SCOM, 2013: 1021-1028.
- [3] Xu T, Zhong H, Zhu H, et al. Exploring the impact of dynamic mutual influence on social event participation [C]//Proceedings of the 2nd International Conference on Sustainable Design and Manufacturing, 2015: 262-270.
- [4] Jamali M, Ester M. A matrix factorization technique with trust propagation for recommendation in social networks [C]//Proceedings the fourth ACM conference on Recommender systems, 2010: 135-142.
- [5] Li K, Lu W, Bhagat S, et al. On social event organization [C]//Proceedings of the 19th ACM SIGKDD conference on Knowledge Discovery and Data Mining, 2014: 1206-1215.
- [6] Yu Z, Du R, Guo B, et al. Who should I invite for my party? combining user preference and influence maximization for social events [C]//Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing, 2015: 879-883.
- [7] Du R, Yu Z, Mei T, et al. Predicting activity attendance in event-based social networks: content, context and social influence [C]//Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing, 2014: 425-434.
- [8] Zhang X M, Zhao J, Cao G H. Who will attend? - predicting event attendance in event-based social network [C]//Proceedings of the 16th IEEE International Conference on Mobile Data Management (MDM), Pittsburgh, PA, USA, 2015: 74-83.
- [9] Li X, Cheng X, Su S, et al. A hybrid collaborative filtering model for social influence prediction in event-based social networks [J]. Neurocomputing, 2017, 230: 197-209.
- [10] Qiao Z, Zhang P, Cao Y, et al. Combining heterogeneous social and geographical information for event recommendation [C]//Proceedings of the 28th AAAI, 2014: 145-151.
- [11] Zhang W, Wang J. A collective bayesian poisson factorization model for cold-start local event recommendation [C]//Proceedings of the 21th ACM SIGKDD conference on Knowledge Discovery and Data Mining, 2015: 1455-1464.
- [12] Liu X, Aberer K. SoCo: a social network aided context-aware recommender system [C]//Proceedings of the 22nd international conference on World Wide Web, 2013: 781-802.
- [13] Tong Y X, She J Y, Meng R. Bottleneck-aware arrangement over event-based social networks: the max-min approach [J]. World Wide Web, 2016, 19 (6): 1151-1177.
- [14] Mungen A A, Kaya M. A novel method for event recommendation in Meetup [C]//Proceedings the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 2017: 959-965.
- [15] Liu S H, Wang B, Xu, M H. Event recommendation based on graph random walking and history preference reranking [C]//Proceedings of the SIGIR'17, Shinjuku, Tokyo, Japan, 2017: 861-864.
- [16] Tong H H, Faloutsos C, Pan J Y. Fast random walk with restart and its applications [C]//Proceedings of the 6th IEEE International Conference on Data Mining. Hong Kong, China: IEEE Press, 2006: 613-622.
- [17] Yang D, Zhang D, Yu Z, et al. Fine-grained preference-aware location search leveraging crowdsourced digital footprints from LBSNs [C]//Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing, 2013: 479-488.



仲兆满(1977—), 博士, 副教授, 主要研究领域为
信息检索、人工智能和社交网络分析。
E-mail: zhongzhaoman@163.com



戴红伟(1975—), 博士, 教授, 主要研究领域为人
工智能。
E-mail: 3974523025@qq.com



管燕(1976—), 硕士, 讲师, 主要研究领域为模式
识别与人工智能。
E-mail: gy764@sohu.com

第二届大数据安全与隐私保护学术会议在兰州成功举办

由中国中文信息学会大数据安全与隐私保护专业委员会主办, 兰州理工大学承办、华中科技大学协办的第二届大数据安全与隐私保护学术会议于 2019 年 7 月 13 日—14 日在兰州圆满举行。中国中文信息学会副理事长兼秘书长孙乐研究员、兰州理工大学副校长王全进教授、学会大数据安全与隐私保护专委会主任李风华研究员, 以及来自全国 41 家单位的 120 多位代表参加了本次会议。

兰州理工大学副校长王全进教授在开幕式上致辞, 欢迎大家到兰州来研讨交流大数据有序共享和隐私保护。中国中文信息学会副理事长兼秘书长孙乐研究员在致辞中代表理事长方滨兴院士对会议的召开表示祝贺, 并指出, 大数据安全与隐私保护是关系到国计民生、经济发展、社会稳定的核心关键技术, 学术界、工业界以及政府机构都高度关注该问题, 专委会要加强并促进大数据安全与隐私保护的学术交流、技术进步, 扩大会的影响力, 推动大数据安全与隐私保护的技术和产业发展。孙乐研究员还对专委会未来工作提出了建议。

大数据安全与隐私保护专委会主任、中国科学院信息工程研究所李风华研究员代表专委会感谢主办方对本次会议举办的大力支持, 以及学会对专委会成立近两年来工作的指导与支持, 并介绍了下一步工作思路。李风华指出, 首先要把专委会这个学术交流的平台办成国内大数据安全和隐私保护的最高学术平台, 交流最新成果, 锻炼组织与交流沟通能力, 组建联合攻关队伍, 促进原始创新, 推动学术研究, 支撑国家的大数据战略; 其次, 要做好宣传, 把专委会网站建成开放学术平台, 建设隐私计算开源社区, 专委会网站还将加强与国内学报的联动, 共同建立发表原创成果、公证首创权的首选平台; 第三, 持续推进隐私计算研究, 中国科学院信息工程研究所、西安电子科技大学等单位联合于 2015 年在国内外首次提出隐私计算, 专委会将促进国内外有志于隐私计算的研究人员共同参与, 不断完善隐私计算的基础理论和技术体系, 并解决国家关心的问题; 第四, 隐私计算已经得到了国家科技管理部门和业界越来越多的关注, 国家重点研发计划、国家自然科学基金重点项目都开始支持隐私计算方面的研究, 算力智库研究院发布的白皮书也引用了隐私计算的论文, 隐私计算研究由国家需求牵引, 产业需求迫切, 具有越来越广阔的原始创新研究前景; 第五, 专委会要不断扩大国际影响力, 计划明年将 PRICOM 学术研讨会扩大为 PRICOM 国际学术会议, 吸引更多的国际学术同行参与; 最后, 专委会要为青年人成长搭建平台, 鼓励年青人以感恩的心态、成就他人的心怀、构建广义“团队”的视野, 共同进步, 报效国家, 最终成就自己。本次会议的主题是“受控共享为数据应用排忧解难, 隐私计算为个人信息保驾护航”, 会议对接互联网+、大数据安全、数字经济、人工智能、5G 安全等国家战略, 为支撑大数据有序共享和隐私保护提供学术交流的平台。

中国科学院信息工程研究所李风华研究员做了题为“隐私计算: 理论、计算框架及其未来发展趋势”的特邀报告, 回顾了 2015 年首次提出隐私计算概念的背景和过程, 介绍了 2016 年和 2019 年发表的隐私计算论文的框架和内涵, 以及隐私计算框架落地的实例, 指出了隐私计算未来的研究方向; 来自悉尼科技大学何祥建教授做了题为“A Mobile Multimedia Data Collection Scheme for Secured Wireless Multimedia Sensor Networks”的特邀报告; 来自纽卡斯尔大学的董长宇博士做了题为: “Secure Computation&Data Sketching: A happy Marriage”的特邀报告。会议还邀请了华中科技大学邹德清教授、中国科学院信息工程研究所王明生研究员、数篷科技吴烨博士、西安电子科技大学朱辉教授、国防科技大学李爱平教授等 7 位专家围绕大数据安全和隐私保护方面做了“虚拟计算环境隐私侵犯取证分析”、“隐私保护中的密态数据处理”、“基于轻量虚拟化和软件定义边界的企业大数据保护实践”、“面向互联网服务的数据隐私保护”、“大数据隐私保护技术概述”、“机器学习中的大数据安全与隐私保护”、“软件漏洞挖掘与数据泄露分析”主题报告。会议还安排了 13 个论文分组报告, 并进行了研讨。