

# 一种基于协同上下文关系学习的同城活动推荐算法\*

赖奕安<sup>1,2</sup>, 张玉洁<sup>1,2</sup>, 杜雨露<sup>1,2</sup>, 孟祥武<sup>1,2</sup>



<sup>1</sup>(智能通信软件与多媒体北京市重点实验室(北京邮电大学),北京 100876)

<sup>2</sup>(北京邮电大学 计算机学院,北京 100876)

通讯作者: 张玉洁, E-mail: zhangyj@bupt.edu.cn

**摘要:** 新兴的基于活动的社交网络以活动为核心,结合线上关系与线下活动促进用户真实、有效的社交关系的形成,但过多的活动信息会使用户难以分辨和选择.结合上下文进行个性化同城活动推荐,是解决活动信息过载问题的一种有效手段.然而大部分现有的同城活动推荐算法都是从用户参与活动记录中间接统计用户对上下文信息的偏好,忽略了两之间潜在的交叉影响关系,从而影响了推荐结果的有效性.为了解决用户参与活动偏好与上下文信息潜在交叉影响关系利用不足的问题,提出了一种基于协同上下文关系学习的同城活动推荐算法(collective contextual relation learning,简称CCRL).首先,对用户参与活动记录和活动主办方、活动内容、活动地点、举办时间等相关上下文信息进行关系建模;然后,采用多关系贝叶斯个性化排序学习方法进行协同上下文关系学习及同城活动推荐. Meetup 数据集上的实验结果表明,该算法在多项指标上均优于现有的主流活动推荐算法.

**关键词:** 活动推荐;推荐算法;上下文感知;多关系学习;个性化

**中图法分类号:** TP311

中文引用格式: 赖奕安,张玉洁,杜雨露,孟祥武.一种基于协同上下文关系学习的同城活动推荐算法.软件学报,2020,31(2): 421-438. <http://www.jos.org.cn/1000-9825/5618.htm>

英文引用格式: Lai YA, Zhang YJ, Du YL, Meng XW. Local event recommendation algorithm based on collective contextual relation learning. Ruan Jian Xue Bao/Journal of Software, 2020,31(2):421-438 (in Chinese). <http://www.jos.org.cn/1000-9825/5618.htm>

## Local Event Recommendation Algorithm Based on Collective Contextual Relation Learning

LAI Yi-An<sup>1,2</sup>, ZHANG Yu-Jie<sup>1,2</sup>, DU Yu-Lu<sup>1,2</sup>, MENG Xiang-Wu<sup>1,2</sup>

<sup>1</sup>(Beijing Key Laboratory of Intelligent Telecommunications Software and Multimedia (Beijing University of Posts and Telecommunications), Beijing 100876, China)

<sup>2</sup>(School of Computer Science, Beijing University of Posts and Telecommunications, Beijing 100876, China)

**Abstract:** The newly emerging event-based social network (EBSN) based on the event as the core combines the online relationship with offline activities to promote the formation of real and effective social relationship among users. However, excessive activity information would make users difficult to distinguish and choose. The context-aware local event recommendation is an effective solution for the information overload problem, but most of existing local event recommendation algorithms only learn users' preference for contextual information indirectly from statistics of historical event participation and ignores latent correlations among them, which impacts on recommendation effectiveness. To take full advantage of latent correlations between users' event preference and contextual information, the proposed collective contextual relation learning (CCRL) algorithm models relations among users' participation records and related contextual information such as event organizer, description text, venue, and starting time. Then multi-relational Bayesian personalized ranking (MRBPR) algorithm is adapted for collective contextual relation learning and local event recommendation. Experiment results on

\* 基金项目: 北京市教育委员会共建项目

Foundation item: Mutual Project of Beijing Municipal Education Commission, China

收稿时间: 2017-12-25; 修改时间: 2018-03-22; 采用时间: 2018-06-18

Meetup dataset demonstrate that proposed algorithm outperforms state-of-the-art local event recommendation algorithms in terms of many metrics.

**Key words:** event recommendation; recommendation algorithm; context-aware; multi-relation learning; personalized

随着社交网络的进一步发展,人们不再满足于单纯的线上互动,越来越多的用户相约走下网络,在线下一起从事感兴趣的活动.由此衍生出了以活动为核心、连接线上和线下社交关系的社交网站,称为基于活动的社交网络(event-based social network,简称 EBSN)<sup>[1]</sup>.但随着 EBSN 上发布活动数量的增多,用户经常会感到难以分辨和选择,被庞大繁杂的信息淹没而不知所措,这就是典型的信息过载问题.考虑到用户日常的交通范围主要局限在城市之中,目前像 Meetup、豆瓣同城活动的众多 EBSN 均按活动举办城市对活动进行归类,在一定程度上缓解了信息过载问题.但这种方法仍然无法考虑用户不同的活动兴趣与需求,从用户角度出发帮助用户选择合适的同城活动.

个性化同城活动推荐正是解决上述问题的有力武器.它能够对用户的兴趣爱好进行建模,分析用户的个性化需求,帮助用户过滤掉自身不感兴趣的内容,并能向用户建议他可能感兴趣的内容<sup>[2]</sup>.目前,个性化推荐系统已在电子商务、社会化网络<sup>[3,4]</sup>等多个领域得到了充分应用.通过分析线下活动特征、活动参与用户的特点、用户之间的社交关系等各类 EBSN 数据,推荐系统能够发掘用户参与活动的个性化需求,为用户推荐符合其个人兴趣的活动,从而改善 EBSN 的用户体验,有利于促进 EBSN 更好、更快地发展.

目前,研究人员已经提出了多种个性化同城活动推荐算法.这些算法均以用户参与活动记录为基础进行用户偏好学习,其中,部分算法还会结合用户的社交关系、活动详情内容、举办时间、举办地点等上下文信息进行更深入的用户偏好建模.然而,大部分现有的同城活动推荐算法都是从用户参与活动记录中间接统计用户对上下文信息的偏好,再采用主题模型分析、相似度分析等方法计算得到用户的上下文偏好,融合得到用户偏好模型.但在实际情况中,用户往往综合考虑活动主办方、活动内容、活动时间、活动地点等多种因素,反复权衡后才决定是否参加某项活动.现有算法大都忽略了用户参与活动偏好与上下文信息之间、各种上下文信息相互之间潜在的交叉影响关系,导致未能准确把握用户参与活动的偏好,影响了推荐结果的有效性.

本文提出一种基于协同上下文关系学习的同城活动推荐算法 CCRL(collective contextual relation learning).该算法以用户参与活动记录为基础,结合用户加入活动群组关系和活动主办关系提取组织关系上下文,结合活动信息数据提取活动内容上下文、举办地点上下文、举办时间上下文等关系特征,最后采用多关系贝叶斯个性化排序算法进行协同上下文关系学习及个性化同城活动 TopN 推荐.与现有同城活动推荐算法相比,CCRL 算法充分利用了用户参与活动偏好与上下文信息之间、多种上下文信息之间的交叉影响关系,提高了算法的有效性.在 Meetup 数据集上的实验表明,CCRL 的推荐效果在准确率、召回率、nDCG 等指标上都要优于基准算法.

本文第 1 节回顾目前上下文感知推荐、社会化推荐、个性化同城活动推荐的研究成果.第 2 节构建基于活动的社交网络模型,并在此基础上定义个性化同城活动推荐问题.第 3 节介绍本文所提出算法的相关预备知识.第 4 节分析现有个性化同城活动推荐算法的不足,并提出一种基于协同上下文关系学习的同城活动推荐算法.第 5 节利用 Meetup 数据集对所提算法进行相关实验,并与其他个性化同城活动推荐算法进行对比分析.第 6 节总结本文的主要工作,并对下一步工作进行展望.

## 1 相关工作

传统的推荐系统以用户对物品的评分数据为基础,分析用户偏好并预测用户对于物品的评价或向用户推荐可能喜欢的物品.但用户偏好的形成并不是孤立的,而是会受到用户所处情境的影响.结合用户所处情境的上下文信息对用户偏好进行更加全面的分析,有利于提高推荐的精确度<sup>[5,6]</sup>.Yuan 等人<sup>[7]</sup>利用协同过滤模型将时间上下文信息引入到兴趣点(point of interest,简称 POI)推荐当中.Liu 等人<sup>[8]</sup>利用 POI 地点的分类信息学习用户的偏好迁移模式,采用矩阵分解方法对用户偏好在分类上的迁移以及用户在对应分类上的偏好进行预测.Zhang 等人<sup>[9]</sup>考虑了 POI 推荐中社交关系和时间信息的影响,提出了时空感知的社会化协同获取模型,以对地点、时

间、社交关系进行统一建模,并采用加权估计对级排序方法进行 TopN 推荐.Zhang 等人<sup>[10]</sup>综合位置关系、社交关系、分类关系进行特征融合,实现了 POI 推荐.

随着社交网络的出现和发展,通过用户间的社会关系信息,能够对用户偏好进行更准确的分析.这是由于用户偏好并不总是用户独立做出的判断,经常会受到用户身边好友的影响.社会化推荐中,主要应用评分网络和信任网络两种形式的社会关系网络<sup>[3]</sup>.Chaney 等人<sup>[11]</sup>结合基于偏好的推荐和社会化推荐,提出了社会化泊松分解模型,将社交网络信息融合到传统的矩阵分解模型当中.Tang 等人<sup>[12]</sup>发现,社交关系可能是由多种关系组合而成的异构关系,采用了基于高斯似然的矩阵分解模型来处理异构关系和弱依赖关系.邹本友等人<sup>[13]</sup>针对已有基于信任的推荐算法都是单一信任模型的问题,提出了基于主题的张量分解的用户信任推荐算法,能够发掘用户在选取不同物品时对好友的不同信任程度.胡勋等人<sup>[14]</sup>使用推土机距离在稀疏项目矩阵上计算跨项目的移动用户相似度,融合项目特征和移动用户信任关系进行协同过滤推荐.

作为从传统线上社交网络基础上演化而来的新型社交网络,EBSN 这一概念的首次提出可以追溯到 2012 年 Liu 等人在 KDD 会议上发表的文章<sup>[1]</sup>,其中,将以活动为核心、连接线上和线下社交关系的社交网站统称为基于活动的社交网络,即 EBSN.同年,Han 等人<sup>[15]</sup>以豆瓣同城活动网站为主要研究对象,分析了豆瓣同城活动网站上的活动属性、用户参与活动表现和社交关系影响.2013 年,Xu 等人<sup>[16]</sup>同样利用豆瓣同城活动网站的数据,分析活动人数和互动程度对社交网络行为的影响,考察用户关注行为与线下活动特征之间的关系,并结合行为设定理论、社会角色理论和用户采访数据对其进行解释.上述研究分析了 EBSN 与传统社交网络的异同点、EBSN 的数据特征、线上社交关系与线下活动之间的互相影响,表明活动信息特征和社交关系确实对于用户参与线下活动存在影响,为实现个性化同城活动推荐提供了一定的理论基础.

在数据分析和理论研究的基础上,2014 年,Du 等人<sup>[17]</sup>提出了结合矩阵分解方法和多因素邻域模型进行用户参与活动预测的 SVD-MFN 算法.该方法对用户参与活动矩阵进行矩阵分解,并利用 LDA 主题模型<sup>[18]</sup>对活动内容进行分析,从中提取用户的活动主题偏好,最后综合内容、时间、空间和社交关系因素进行用户参与活动预测.SVD-MFN 算法实际解决的是用户是否会参与某项活动的预测问题,更接近于二分类问题而非 TopN 推荐.

2015 年,Macedo 等人<sup>[19]</sup>提出了 MCLRE 算法,结合活动上下文,从社交关系、内容、地点和时间这 4 个方面提取活动特征,利用学习排序(learning to rank)方法<sup>[20]</sup>进行特征融合并实现同城活动 TopN 推荐.同年,Zhang 等人<sup>[21]</sup>为了解决同城活动推荐中的冷启动问题,提出的集体贝叶斯泊松分解(collective Bayesian Poisson factorization,简称 CBPF)算法融合了用户参与活动记录、活动主办方、活动内容、活动地点和起始时间等信息,从而在缺少用户参与活动记录的冷启动情景中仍然能够充分利用其他信息进行活动推荐.2016 年,Gu 等人<sup>[22]</sup>针对活动推荐中的冷启动问题和隐式反馈问题,提出了 AlphaMF 上下文感知矩阵分解(context aware matrix factorization)模型,结合隐式反馈矩阵分解模型和线性上下文特征模型进行活动推荐.2017 年,Jhamb 等人<sup>[23]</sup>提出的 GLFM 算法综合考察了用户兴趣、活动地点、活动流行度、位置距离、举办时间等可能影响用户参与活动偏好的因素,并特别从用户和活动两个角度考虑了活动主办群组因素的影响,综合进行同城活动推荐.2017 年,Liu 等人<sup>[24]</sup>分析了用户-活动-群组 and 用户-活动-地点这两种三角关系,建立了协同对级矩阵分解模型(collective pairwise matrix factorization,简称 CPMF),重点是协同利用偏好关系进行矩阵分解,从而对即将举办的同城活动进行推荐.

## 2 EBSN 模型与问题定义

个性化同城活动推荐以基于活动的社交网络为基础,充分利用 EBSN 中的各项数据挖掘用户的个性化需求并提供同城活动推荐.在 EBSN 中,通常包含用户、同城活动、活动主办方、活动举办地点等项目.在 Meetup EBSN 中,同城活动由多位用户组成的活动群组发起并举办,此时,活动主办方即为活动群组.活动本身还具有主办方、内容文本、举办地点、举办时间等信息,此外还包含用户参与活动记录数据、用户加入活动群组数据等.基于 EBSN 数据构建用户的兴趣模型,为每位用户推荐他可能感兴趣的即将举行的活动,是个性化同城活动推荐所要解决的问题.

## 2.1 基于活动的社交网络模型

基于活动的社交网络是个性化同城活动推荐的基础.通过对 Meetup.com、豆瓣活动等常见 EBSN 的分析,可将 EBSN 中用户、活动、活动群组、地点等项目的关系简化,如图 1 所示.

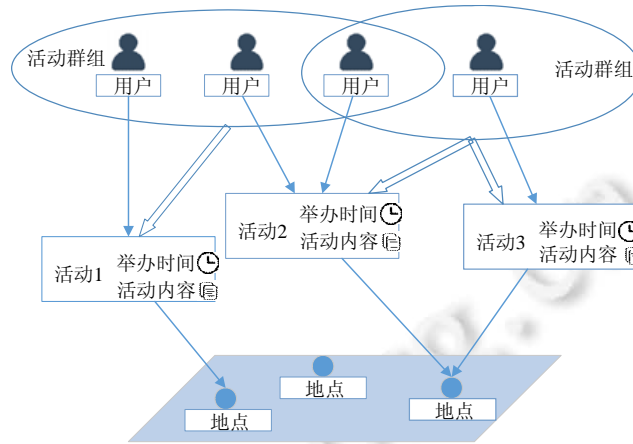


Fig.1 Simple diagram of relations among entities in EBSN

图 1 EBSN 中项目间关系简单示意图

其中,每项同城活动都由活动群组发起,而活动群组由若干用户组成,每名用户可以加入不同的活动群组.同时,用户既可以参与自己所在活动群组举办的活动,也可以参与其他活动群组举办的活动.每项同城活动除了活动主办方属性外,还包含活动内容和举办时间属性.每项活动还具有一个举办地点,一个地点也可以举办多项活动.

综合上述分析,给出基于活动的社交网络模型的形式化定义.

**定义 1(基于活动的社交网络(EBSN)模型).** 基于活动的社交网络可抽象成异构图  $G=(V,E)$ ,  $V=(U,E,G,L)$ ,  $E=(R_{UE},R_{UG},R_{EG},R_{EL})$ .其中,

- $U$  为 EBSN 中用户集合.
- $E$  为同城活动集合,  $e=(g_e, c_e, l_e, t_e) \in E$ , 其中,  $g_e$  为活动主办群组,  $c_e$  为活动内容文本,  $l_e$  为活动举办地点,  $t_e$  为活动举办时间.
- $G$  为活动群组集合.活动群组是活动的主办方.
- $L$  为活动地点集合.对于任一  $e \in E$ ,都必然有其活动地点  $l_e \in L$ .
- $R_{UE}$  为用户活动参与关系,  $(u, e) \in R_{UE}$  表示用户  $u$  参加了活动  $e$ .
- $R_{UG}$  为用户加入活动群组关系,  $(u, g) \in R_{UG}$  表示用户  $u$  加入了活动群组  $g$ .
- $R_{EG}$  为活动主办群组关系,  $(e, g_e) \in R_{EG}$  表示活动群组  $g_e$  主办了活动  $e$ .
- $R_{EL}$  为活动举办地点关系,  $(e, l_e) \in R_{EL}$  表示活动  $e$  在地点  $l_e$  举办.

## 2.2 问题定义

个性化同城活动推荐是在某一时间点上,根据用户参与活动记录向用户推荐近期即将举办的同城活动.为了准确地描述某一时间点上用户已经参与过的活动和稍后即将举办的活动,根据第 2.1 节给出的 EBSN 模型定义,给出个性化同城活动推荐中历史活动集合和候选活动集合的形式化定义.

**定义 2(历史活动集合).** 给定一时间点  $\tau$  并以此为分隔,截至该时间点已经结束的活动称为历史活动,构成历史活动集合  $E_{old}^{\tau}$ .

**定义 3(候选活动集合).** 给定一时间点  $\tau$  并以此为分隔,从该时间点起,已经发布但尚未举办的活动称为候

选活动,构成候选活动集合  $E_{cand}^r$ .

结合本文提出的 EBSN 模型,参考 Adomavicius 等人对于推荐系统的形式化定义<sup>[25,26]</sup>,定义个性化同城活动推荐问题如下.

**定义 4(个性化同城活动推荐问题).** 对任一用户  $u \in U$ ,求出候选活动集合  $E_{cand}$  中使评价函数  $\hat{s}$  值最高的  $n$  个候选活动,可表示为如公式(1)所示的形式.

$$\text{top-}n(u) = \arg \max_{e \in E_{cand}} \hat{s}(u, e, g_e, c_e, l_e, t_e) \quad (1)$$

评价函数  $\hat{s}$  结合上下文信息计算用户对活动的偏好评价价值,其形式如公式(2)所示.

$$\hat{s}: U \times E \times G \times C \times L \times T \rightarrow \mathbb{R} \quad (2)$$

其中,  $U$  为用户域,  $E$  为活动域,  $G$  为活动群组域,  $C$  为活动内容域,  $L$  为活动地点域,  $T$  为活动时间域.

### 3 预备知识

Krohn-Grimberghe 等人提出的多关系贝叶斯个性化排序(multi-relational factorization with Bayesian personalized ranking,简称 MRBPR)方法<sup>[27]</sup>基于协同矩阵分解(collective matrix factorization)方法<sup>[28]</sup>对多个存在交叉影响的关系进行联合建模,利用贝叶斯个性化排序(Bayesian personalized ranking)方法<sup>[29]</sup>从隐式反馈中学习用户偏好.该方法利用多种关系缓解稀疏性问题,尤其适合处理以隐式反馈为主的社交网络数据.

传统的单一矩阵分解方法<sup>[30]</sup>中,两种实体  $\varepsilon_1$  和  $\varepsilon_2$  之间的关系用矩阵  $R_{\varepsilon_1 \varepsilon_2}$  表示,称为关系矩阵.若有  $m$  个  $\varepsilon_1$  类实体与  $n$  个  $\varepsilon_2$  类实体,则有关系矩阵  $R_{\varepsilon_1 \varepsilon_2} \subseteq \mathbb{R}^{m \times n}$ .假设实体  $\varepsilon_1$  和  $\varepsilon_2$  之间的关系存在  $k$  个隐式影响因子,则可设低维矩阵  $E_1 \subseteq \mathbb{R}^{m \times k}$  表示  $\varepsilon_1$  类实体的隐式因子,  $E_2 \subseteq \mathbb{R}^{n \times k}$  表示  $\varepsilon_2$  类实体的隐式因子,使得有  $R_{\varepsilon_1 \varepsilon_2} \approx E_1 E_2^T$ .从而单一矩阵分解的优化目标如公式(3)所示.

$$\arg \min_{E_1, E_2} \ell(R_{E_1 E_2}, E_1 E_2^T) + \text{Reg}(E_1, E_2) \quad (3)$$

其中,  $\ell(\cdot)$  为损失函数,  $\text{Reg}$  为正则化项.

协同矩阵分解方法将单一矩阵分解方法扩展到多个关系矩阵上.对于实体  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m$  与关系矩阵  $R_{\varepsilon_1 \varepsilon_2}, \dots, R_{\varepsilon_1 \varepsilon_j}, \dots, R_{\varepsilon_i \varepsilon_j}$ , 令低维矩阵  $E_1, E_2, \dots, E_m$  分别对应  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m$  类实体的隐式因子,多种关系共享对应实体的隐式因子,则协同矩阵分解的优化目标如公式(4)所示.

$$\arg \min_{\Theta} \sum_R \omega_R \ell(R_{\varepsilon_i \varepsilon_j}, E_i E_j^T) + \text{Reg}(\Theta) \quad (4)$$

其中,  $\Theta = \{E_1, E_2, \dots, E_m\}$  为低维因子矩阵集合,  $\omega_R$  为关系矩阵权重,  $\sum_R \omega_R = 1$ .

考虑到用户参与活动记录  $r_{u,e} \in \{0,1\}$  属于隐式反馈,传统的矩阵分解方法并不能直接应用在用户参与活动记录矩阵  $R_{UE}$  上.不妨认为,在历史活动集合中,相对于未参与的活动,用户更偏好参与过的历史活动.设  $\hat{r}_{u,e}$  为用户  $u$  对于活动  $e$  的偏好程度,则应有:

$$r_{u,e_1} > r_{u,e_2} \Leftrightarrow \hat{r}_{u,e_1} > \hat{r}_{u,e_2} \quad (5)$$

通过学习  $\hat{r}_{u,e}$  的偏序关系,可以对用户参与活动的偏好进行预测.将公式(5)推广到一般的关系矩阵,化为公式(6).

$$r_{e_1, e_{2i}} > r_{e_1, e_{2j}} \Leftrightarrow \hat{r}_{e_1, e_{2i}} > \hat{r}_{e_1, e_{2j}} \quad (6)$$

其中,  $e_1 \in \varepsilon_1, e_{2i}, e_{2j} \in \varepsilon_2, r_{e_1, e_{2i}}, r_{e_1, e_{2j}} \in R_{\varepsilon_1 \varepsilon_2}$ .

贝叶斯个性化排序方法正是基于上述思想对类似于用户参与活动记录这样的隐式反馈进行排序学习的经典方法.令  $\hat{x}_{e_1, e_{2i}, e_{2j}}$  为用户偏好程度间的差值,  $\hat{x}_{e_1, e_{2i}, e_{2j}} = \hat{r}_{e_1, e_{2i}} - \hat{r}_{e_1, e_{2j}}$ ,其学习目标函数如公式(7)所示.

$$\text{BPR-OPT}(R_{\varepsilon_1 \varepsilon_2}, E_1 E_2^T) = \sum_{(e_1, e_{2i}, e_{2j}) \in S} \sigma(\hat{x}_{e_1, e_{2i}, e_{2j}}) \quad (7)$$

其中,  $S$  为训练样本集,  $S \subseteq \varepsilon_1 \times \varepsilon_2 \times \varepsilon_3$ ;  $\sigma$  为 Sigmoid 函数,  $\sigma(x) = \frac{1}{1 + e^{-x}}$ .

多关系贝叶斯个性化排序方法融合了协同矩阵分解方法与贝叶斯个性化排序方法. 以贝叶斯个性化排序目标函数为损失函数, 以  $L_2$  正则化方法为正则项, 其目标函数如公式(8)所示.

$$\arg \max_{\Theta} \sum_R \omega_R BPR - OPT(\mathbf{R}_{\varepsilon_i \varepsilon_j}, \mathbf{E}_i \mathbf{E}_j^T) + Reg(\Theta) \quad (8)$$

## 4 同城活动推荐算法

在个性化同城活动推荐中, 用户参与活动偏好与上下文信息之间、多种上下文信息之间存在着交叉影响关系. 用户在决定是否参与某项同城活动时, 通常会综合考虑活动内容、活动主办方、举办地点、活动时间等多项上下文因素进行权衡. 将这样的权衡视为多种上下文因素之间存在着潜在的交叉影响关系, 共同决定用户参与同城活动的偏好. 由于 EBSN 通常不包含用户对上下文的偏好反馈, 在进行个性化同城活动推荐时, 通常需要先从已知的用户参与活动记录中间接学习用户对上下文因素的偏好, 然后才能结合上下文信息进行用户偏好建模. 在间接学习过程中, 用户对活动偏好的差异会影响得到的上下文信息偏好, 结合上下文信息对用户参与活动偏好所存在的影响, 称用户参与活动与上下文信息之间存在潜在的交叉影响关系.

本文提出的基于协同上下文关系学习的个性化同城活动推荐算法以用户参与活动记录、活动组织信息、活动内容信息、举办地点信息、举办时间信息为基础, 充分利用了用户参与活动偏好与上下文信息之间、多种上下文信息之间的交叉影响关系进行用户偏好建模, 并采用了多关系贝叶斯个性化排序(MRBPR)方法进行协同上下文关系学习、实现个性化同城活动推荐.

### 4.1 现有算法的不足之处

在基于活动的社交网络中, 用户参与同城活动的信息属于隐式反馈. 例如在 Meetup.com、豆瓣活动等 EBSN 中仅含有用户参与活动记录, 而不记录用户表示不想参与的历史活动, 也不记录用户是否喜欢某项活动. 一般认为, 用户参与了某项活动即表明用户喜欢该项活动, 但对于用户没有参与的历史活动, 可能用户确实不喜欢该项活动而拒绝参与, 也可能是用户没有关注该项活动或者由于某些原因忽略了该项活动, 从而无法判断用户对于未参与的历史活动的偏好. 针对这样的隐式反馈数据, 通常认为用户对曾参与的历史活动的偏好程度要大于未参与的历史活动, 从而区分用户对活动的偏好差异.

由于用户参与活动记录所提供的用户偏好信息有限, 现有的个性化同城活动推荐算法大多考虑结合上下文信息辅助用户偏好建模, 如用户与活动主办方的关注关系、活动内容文本、活动举办地点、活动举办时间等可能影响用户参与活动偏好的因素. 这些上下文信息对用户参与活动偏好的影响是显而易见的, 但 EBSN 中并不包含用户对上下文信息的显式反馈, 既无法明确得知用户对各种上下文信息的偏好程度, 也不知道在用户综合考虑的过程中各种上下文信息所占有的权重, 这给利用上下文信息进行用户偏好建模带来了一定的困难.

目前, 大多数个性化同城活动推荐算法从用户参与活动记录中间接学习用户对上下文信息的偏好, 未能考虑后续用户偏好建模过程中出现的用户对活动偏好的差异的影响. 许多算法采用主题模型分析(常用于活动内容文本信息<sup>[17,19]</sup>)、相似度分析(常用于活动地点、时间等信息<sup>[17,19]</sup>)等方法计算用户的上下文信息偏好, 再尝试学习上下文信息对用户偏好的影响. 这样的间接学习过程通常认为用户对过往参与活动的偏好程度是均一的, 而忽略了用户对不同活动偏好的差异, 及其对用户上下文信息偏好的影响. 即使是在同一种上下文信息中, 用户对同一项目的偏好也会有所不同. 例如, 在同一活动群组举办的不同活动中, 受到用户对具体活动偏好的影响, 用户对于该活动群组的偏好可能不同. 现有算法未能考虑后续用户偏好建模过程中用户对不同活动的偏好可能存在的差异, 也就无法利用用户参与活动偏好对用户的上下文信息偏好的影响关系.

目前, 已有算法在一定程度上考虑了用户参与活动偏好与上下文信息之间的潜在影响关系, 但仍存在不足之处. 如, Macedo 等人<sup>[19]</sup>采用多关系贝叶斯个性化排序学习方法对用户参与活动记录、用户加入活动群组关系、活动主办群组关系进行了协同关系学习, 但该算法对于活动内容、举办地点、举办时间等上下文信息仍采用余

弦相似度或核密度估计等方法进行用户偏好程度统计,没有深入挖掘这些上下文信息与用户参与的历史活动之间的潜在影响关系.Liu 等人<sup>[24]</sup>采用对级排序学习技术对用户-活动-群组 and 用户-活动-地点这两种三角关系进行偏序关系学习,考虑到了用户参与活动记录与上下文信息之间的影响关系,但该算法仍然是通过统计用户参与活动记录来构建用户-群组/活动偏好矩阵;此外,也没有考虑活动内容、举办时间等上下文信息.

#### 4.2 基于协同上下文关系学习的个性化同城活动推荐算法

针对现有算法的不足,本文提出一种基于协同上下文关系学习的同城活动推荐算法 CCRL,充分利用了用户参与活动偏好与上下文信息之间、多种上下文信息之间的交叉影响关系.该算法主要考虑了以下几种潜在的交叉影响关系.

- (1) 活动主办方(活动群组)与用户偏好之间的潜在影响关系;
- (2) 活动内容与用户偏好之间的潜在影响关系;
- (3) 活动举办地点与用户偏好之间的潜在影响关系;
- (4) 活动举办时间与用户偏好之间的潜在影响关系;
- (5) 上述多种上下文信息之间潜在的交叉影响关系.

基于协同上下文关系学习实现用户偏好建模和同城活动推荐的主要过程如下.

- (1) 以用户参与活动记录为基础,结合用户加入活动群组关系和活动主办关系提取组织关系上下文信息,结合活动信息数据提取活动内容、举办地点、举办时间等上下文信息,构建上下文关系矩阵.
- (2) 将关系矩阵分解为对应实体的隐式因子矩阵的乘积,采用多关系贝叶斯个性化排序算法对多种上下文关系及其隐式因子进行协同学习,从而对用户偏好进行联合建模.
- (3) 基于协同上下文关系学习得到的用户参与活动偏好,实现同城活动的个性化 TopN 推荐.

对全体用户  $u \in U$  和活动  $e \in E$ ,由用户参与活动关系  $R_{UE}$  可建立用户参与活动矩阵  $R_{UE} = \{r_{u,e}\}_{|U| \times |E|}$ ,其中, $|U|$  为用户总数, $|E|$  为活动总数,且对任一  $(u,e) \in R_{UE}$ ,有  $r_{u,e} = 1$ .由多关系贝叶斯个性化排序算法可知,该用户参与活动矩阵可被分解为用户隐式因子  $U_{|U| \times k}$  和活动隐式因子  $E_{|E| \times k}$  两个低维矩阵,其中, $k$  为隐式因子维数.在此基础上,采用类似方法对组织关系、活动内容、举办地点、举办时间等上下文信息进行关系建模.

##### 4.2.1 组织关系上下文

用户与活动主办方的关系是影响用户参与活动偏好的重要因素.特别是在 Meetup EBSN 中,活动群组不仅是活动的发起者和主办方,更是用户群体的组织形式.加入活动群组一方面表明用户对该群组主办的活动感兴趣,另一方面表明用户与群组中的其他用户具有共同的兴趣爱好.同时,大多数基于活动的社交网络均具有活动推送功能:用户通过关注自己感兴趣的活动主办方(加入活动群组),在主办方发布新活动时获得活动推送,从而及时了解自己可能感兴趣的活动的信息,这进一步增加了用户参与活动的可能性.

对全体用户  $u \in U$  和活动群组  $g \in G$ ,由用户加入活动群组关系  $R_{UG}$  可建立用户加入活动群组矩阵  $R_{UG} = \{r_{u,g}\}_{|U| \times |G|}$ ,其中, $|G|$  为活动群组总数.对任一  $(u,g) \in R_{UG}$ ,有  $r_{u,g} = 1$ .该用户加入活动群组矩阵可分解为用户隐式因子  $U_{|U| \times k}$  和活动群组隐式因子  $G_{|G| \times k}$  两个低维矩阵,其中, $k$  为隐式因子维数.

对全体活动  $e \in E$  和举办该活动的活动群组  $g_e \in G$ ,由活动主办群组关系  $R_{EG}$  可建立活动主办群组矩阵  $R_{EG} = \{r_{e,g_e}\}_{|E| \times |G|}$ ,对任一  $(e,g_e) \in R_{EG}$ ,有  $r_{e,g_e} = 1$ .该活动主办群组矩阵可经贝叶斯个性化排序分解为活动隐式因子  $E_{|E| \times k}$  和活动群组隐式因子  $G_{|G| \times k}$  两个低维矩阵,其中, $k$  为隐式因子维数.

上述用户参与活动矩阵  $R_{UE}$ 、用户加入活动群组矩阵  $R_{UG}$ 、活动主办群组矩阵  $R_{EG}$  三者共享了用户隐式因子  $U$ 、活动隐式因子  $E$ 、群组隐式因子  $G$ .包含相同实体的关系之间共享对应的隐式因子,是协同上下文关系学习的基础.在利用多关系贝叶斯个性化排序方法进行协同上下文关系学习时,每一轮迭代过程中,3 种隐式因子共同得到更新,使得用户参与活动偏好与上下文信息之间以共享隐式因子为载体实现交叉影响.换言之,共享隐式因子在一定程度上表达了潜在的交叉影响关系.由此实现了对用户参与活动偏好与上下文信息之间潜在的交叉影响关系的学习,即协同上下文关系学习.后续引入的其他上下文关系和实体同理.

#### 4.2.2 活动内容上下文

此处提到的活动内容上下文信息包括活动标题、活动详情、活动须知等非结构化的文本类的描述信息.活动标题和活动详情充分描述了活动的具体内容,是吸引用户参与活动的主要因素.活动须知给出了用户参与活动时所需要注意事项,如活动费用、交通路线等,在一定程度上影响用户决策是否参与活动.

对活动内容文本进行分词和去除停用词之后,剩下的词汇称为有效内容单词  $c_e$ .对任一活动  $e \in E$ ,其全体有效内容单词构成活动内容单词集合  $C_e$ .全体活动内容单词集合记为  $C$ ,即  $C = \bigcup_{e \in E} C_e$ .

对全体活动  $e \in E$  和活动内容单词集合  $C_e \subseteq C$ ,建立活动内容单词矩阵  $R_{EC} = \{r_{e,c_e}\}_{|E| \times |C|}$ :对任一  $c_e \in C_e$ ,有  $r_{e,c_e} = 1$ .该活动内容单词矩阵可经贝叶斯个性化排序分解为两个低维矩阵:活动隐式因子  $E_{|E| \times k}$  和活动内容隐式因子  $C_{|C| \times k}$ ,其中,  $k$  为隐式因子维数.

#### 4.2.3 举办地点上下文

除了按城市对活动进行分类外,用户对于活动具体的举办地点依然有所偏好.从交通距离的角度考虑,用户更倾向于参与离自己居住地点较近的活动.同时,许多活动举办地点往往与特定主题的活动相关联,如音乐会活动通常选择在音乐厅、剧院等具有良好音响条件的场地举办.可见,用户的活动内容偏好与用户的举办地点偏好存在潜在的交叉影响关系.

由于缺少用户对举办地点的显式反馈信息,因此需要结合用户参与活动记录来学习用户对活动举办地点的偏好.对活动  $e \in E$  和活动举办地点  $l_e \in L$ ,由活动举办地点关系  $R_{EL}$  可建立活动举办地点矩阵  $R_{EL} = \{r_{e,l_e}\}_{|E| \times |L|}$ ,对任一  $(e,l_e) \in R_{EL}$ ,有  $r_{e,l_e} = 1$ .该活动举办地点矩阵可分解为两个低维矩阵:活动隐式因子  $E_{|E| \times k}$  和活动举办地点隐式因子  $L_{|L| \times k}$ ,其中,  $k$  为隐式因子维数.

#### 4.2.4 举办时间上下文

人类的日常行为模式经常体现出一定的周期性,例如,从一天来看,学生白天需要上课,职员白天需要上班,因此这些用户主要参加在晚上举办的活动;从一周来看,如职员或学生通常只有周末有时间参加活动,而自由职业者参加活动的时段则较为分散.与举办地点上下文类似,基于活动的社交网络中并不包含用户对于活动举办时间的显式偏好信息,需要从用户参与活动记录中间接学习用户对于举办时间的偏好.

对活动举办时间属性,从中提取星期特征  $w_e$  和小时特征  $h_e$  两项,其中,  $w_e$  从 0 到 6 分别表示从星期一到星期日的时间属性,  $h_e$  表示活动举办时间的小时值.由于小时维度的粒度过细,一般用户的活动偏好并不会精确到某一小时,因此,将一天 24 小时拆成 4 个时段进行分别分析:如 0 时~5 时视为深夜时段,6 时~11 时视为上午时段,12 时~17 时视为下午时段,18 时~23 时视为夜晚时段.从而定义活动的时段特征  $p_e \in \{1,2,3,4\}$ ,其值 1~4 分别表示上述从深夜到夜晚的 4 个时段.结合星期特征  $w_e$  和时段特征  $p_e$ ,定义活动举办时间  $t_e = 4w_e + p_e$ .易知,  $t_e$  的取值范围为  $\{1,2,\dots,28\}$ ,称为活动举办时间集合  $T$ .

由活动  $e \in E$  和活动举办时间  $t_e \in T$ ,可建立活动举办时间矩阵  $R_{ET} = \{r_{e,t_e}\}_{|E| \times |T|}$ ,对任一  $(e,t_e) \in R_{ET}$ ,有  $r_{e,t_e} = 1$ .该活动举办时间矩阵可分解为两个低维矩阵:活动隐式因子  $E_{|E| \times k}$  和活动举办时间隐式因子  $T_{|T| \times k}$ ,其中,  $k$  为隐式因子维数.

### 4.3 协同上下文关系学习

综合第 4.2.1 节~第 4.2.4 节所述,以用户参与活动记录为基础,结合用户加入活动群组关系和活动主办关系提取组织关系上下文,结合活动信息数据提取活动内容、举办地点、举办时间等上下文关系特征,共有用户参与活动矩阵  $R_{UE}$ 、用户加入活动群组矩阵  $R_{UG}$ 、活动主办群组矩阵  $R_{EG}$ 、活动内容单词矩阵  $R_{EC}$ 、活动举办地点矩阵  $R_{EL}$ 、活动举办时间矩阵  $R_{ET}$  等关系矩阵,涉及用户、活动、活动群组、活动内容单词、举办地点、举办时间等多个实体,分别对应用户隐式因子  $U$ 、活动隐式因子  $E$ 、活动群组隐式因子  $G$ 、内容隐式因子  $C$ 、位置隐式因子  $L$ 、时间隐式因子  $T$ .由多关系贝叶斯个性化排序方法可得公式(9).



$$\begin{cases} \mathbf{R}_{UE} \approx \mathbf{U}\mathbf{E}^T \\ \mathbf{R}_{UG} \approx \mathbf{U}\mathbf{G}^T \\ \mathbf{R}_{EG} \approx \mathbf{E}\mathbf{G}^T \\ \mathbf{R}_{EC} \approx \mathbf{E}\mathbf{C}^T \\ \mathbf{R}_{EL} \approx \mathbf{E}\mathbf{L}^T \\ \mathbf{R}_{ET} \approx \mathbf{E}\mathbf{T}^T \end{cases} \quad (9)$$

在上述关系矩阵的近似式中,用户参与活动矩阵  $\mathbf{R}_{UE}$ 、用户加入活动群组矩阵  $\mathbf{R}_{UG}$ 、活动主办群组矩阵  $\mathbf{R}_{EG}$  交叉共享了用户隐式因子  $\mathbf{U}$ 、活动隐式因子  $\mathbf{E}$ 、活动群组隐式因子  $\mathbf{G}$ ;用户参与活动矩阵  $\mathbf{R}_{UE}$ 、活动主办群组矩阵  $\mathbf{R}_{EG}$  又与活动内容单词矩阵  $\mathbf{R}_{EC}$ 、活动举办地点矩阵  $\mathbf{R}_{EL}$ 、活动举办时间矩阵  $\mathbf{R}_{ET}$  共享了活动隐式因子  $\mathbf{E}$ 。以上述共享隐式因子作为中间桥梁,承载了用户参与活动偏好与上下文信息之间、多种上下文信息之间潜在的交叉影响关系。在利用多关系贝叶斯个性化排序方法进行协同上下文关系学习的过程中,用户参与活动偏好主要由用户隐式因子和活动隐式因子结合表达,而更新多种上下文信息对应的隐式因子会影响用户隐式因子和活动隐式因子,利用了上下文信息对用户参与活动偏好的影响关系;上下文信息的隐式因子之间又通过用户隐式因子和活动隐式因子互相影响,利用到了上下文信息之间潜在的交叉影响关系;学习得到的用户隐式因子和活动隐式因子又反过来影响关联的上下文信息隐式因子,从而充分利用了用户参与活动偏好与上下文信息之间、多种上下文信息之间潜在的交叉影响关系。

基于多关系贝叶斯个性化排序方法,令  $\Theta = \{\mathbf{U}, \mathbf{E}, \mathbf{G}, \mathbf{C}, \mathbf{L}, \mathbf{T}\}$  为隐式因子集合,可以得到如公式(10)所示的目标函数。

$$\arg \min_{\Theta} \left( \omega_{UE} \ell(\mathbf{R}_{UE}, \mathbf{U}\mathbf{E}^T) + \omega_{UG} \ell(\mathbf{R}_{UG}, \mathbf{U}\mathbf{G}^T) + \omega_{EG} \ell(\mathbf{R}_{EG}, \mathbf{E}\mathbf{G}^T) + \omega_{EC} \ell(\mathbf{R}_{EC}, \mathbf{E}\mathbf{C}^T) + \omega_{EL} \ell(\mathbf{R}_{EL}, \mathbf{E}\mathbf{L}^T) + \omega_{ET} \ell(\mathbf{R}_{ET}, \mathbf{E}\mathbf{T}^T) + \lambda_U \|\mathbf{U}\|^2 + \lambda_E \|\mathbf{E}\|^2 + \lambda_G \|\mathbf{G}\|^2 + \lambda_C \|\mathbf{C}\|^2 + \lambda_L \|\mathbf{L}\|^2 + \lambda_T \|\mathbf{T}\|^2 \right) \quad (10)$$

其中,  $\ell(\cdot)$  为贝叶斯个性化排序损失函数,  $\omega$  为关系权重,  $\lambda$  为正则化参数。利用随机梯度下降的方法<sup>[31]</sup>对隐式因子参数  $\Theta$  进行学习,各次迭代中隐式因子参数的更新公式如公式(11)所示。

$$\mathbf{X} = \mathbf{X} - \alpha \frac{d\ell}{d\mathbf{X}} \quad (11)$$

其中,  $\mathbf{X}$  为隐式因子;  $\mathbf{U}, \mathbf{E}, \mathbf{G}, \mathbf{C}, \mathbf{L}, \mathbf{T}$ ,  $\alpha$  为学习率参数。各个隐式因子的具体更新公式参见第 4.4 节。

最终得到的用户对候选活动的偏好程度如公式(12)所示。

$$\tilde{r}_{u,e} = \tilde{\mathbf{u}} \cdot \tilde{\mathbf{e}} \quad (12)$$

这里估计的偏好程度并非用户对于候选活动的评分预测,而是一个用于排序比较的值。偏好程度值越大,表明该候选活动越应排在推荐列表的靠前位置。

对全体  $e \in E_{cmd}$  的  $\tilde{r}_{u,e}$  进行降序排列,取前  $N$  个活动构成推荐结果列表,即可实现个性化同城活动 TopN 推荐。

#### 4.4 CCRL算法描述

输入数据:

- a. EBSN 中用户集合  $U$ ;
- b. 同城活动集合  $E$ ;
- c. 活动群组集合  $G$ ;
- d. 活动地点集合  $L$ ;
- e. 用户活动参与关系  $\mathbf{R}_{UE}$ ;
- f. 用户加入活动群组关系  $\mathbf{R}_{UG}$ ;
- g. 活动主办群组关系  $\mathbf{R}_{EG}$ ;
- h. 活动举办地点关系  $\mathbf{R}_{EL}$ 。

输出数据:同城活动推荐列表。同城活动在推荐列表中越靠前,表明其预测偏好程度越高。

算法流程:

Step 1. 从输入数据构建用户参与活动矩阵  $R_{UE}$ 、用户加入活动群组矩阵  $R_{UG}$ 、活动主办群组矩阵  $R_{EG}$ 、活动内容单词矩阵  $R_{EC}$ 、活动举办地点矩阵  $R_{EL}$ 、活动举办时间矩阵  $R_{ET}$

Step 2. 初始化隐式因子矩阵  $U, E, G, C, L, T$

Step 3. repeat

Step 4. 从  $R_{UE}$  中提取  $(u, e_i, e_j)$

$$\text{Step 5. } \bar{u} \leftarrow \bar{u} + \alpha \left( \omega_{UE} \sigma(\hat{x}_{u, e_i, e_j}) e^{-\hat{x}_{u, e_i, e_j}} \cdot \frac{\partial}{\partial \bar{u}} \hat{x}_{u, e_i, e_j} + \lambda_U \bar{u} \right)$$

$$\text{Step 6. } \bar{e}_i \leftarrow \bar{e}_i + \alpha \left( \omega_{UE} \sigma(\hat{x}_{u, e_i, e_j}) e^{-\hat{x}_{u, e_i, e_j}} \cdot \frac{\partial}{\partial \bar{e}_i} \hat{x}_{u, e_i, e_j} + \lambda_E \bar{e}_i \right)$$

$$\text{Step 7. } \bar{e}_j \leftarrow \bar{e}_j + \alpha \left( \omega_{UE} \sigma(\hat{x}_{u, e_i, e_j}) e^{-\hat{x}_{u, e_i, e_j}} \cdot \frac{\partial}{\partial \bar{e}_j} \hat{x}_{u, e_i, e_j} + \lambda_E \bar{e}_j \right)$$

Step 8. 从  $R_{UG}$  中提取  $(u, g_i, g_j)$

$$\text{Step 9. } \bar{u} \leftarrow \bar{u} + \alpha \left( \omega_{UG} \sigma(\hat{x}_{u, g_i, g_j}) e^{-\hat{x}_{u, g_i, g_j}} \cdot \frac{\partial}{\partial \bar{u}} \hat{x}_{u, g_i, g_j} + \lambda_U \bar{u} \right)$$

$$\text{Step 10. } \bar{g}_i \leftarrow \bar{g}_i + \alpha \left( \omega_{UG} \sigma(\hat{x}_{u, g_i, g_j}) e^{-\hat{x}_{u, g_i, g_j}} \cdot \frac{\partial}{\partial \bar{g}_i} \hat{x}_{u, g_i, g_j} + \lambda_G \bar{g}_i \right)$$

$$\text{Step 11. } \bar{g}_j \leftarrow \bar{g}_j + \alpha \left( \omega_{UG} \sigma(\hat{x}_{u, g_i, g_j}) e^{-\hat{x}_{u, g_i, g_j}} \cdot \frac{\partial}{\partial \bar{g}_j} \hat{x}_{u, g_i, g_j} + \lambda_G \bar{g}_j \right)$$

Step 12. 从  $R_{EG}$  中提取  $(e, g_i, g_j)$

$$\text{Step 13. } \bar{e} \leftarrow \bar{e} + \alpha \left( \omega_{EG} \sigma(\hat{x}_{e, g_i, g_j}) e^{-\hat{x}_{e, g_i, g_j}} \cdot \frac{\partial}{\partial \bar{e}} \hat{x}_{e, g_i, g_j} + \lambda_E \bar{e} \right)$$

$$\text{Step 14. } \bar{g}_i \leftarrow \bar{g}_i + \alpha \left( \omega_{EG} \sigma(\hat{x}_{e, g_i, g_j}) e^{-\hat{x}_{e, g_i, g_j}} \cdot \frac{\partial}{\partial \bar{g}_i} \hat{x}_{e, g_i, g_j} + \lambda_G \bar{g}_i \right)$$

$$\text{Step 15. } \bar{g}_j \leftarrow \bar{g}_j + \alpha \left( \omega_{EG} \sigma(\hat{x}_{e, g_i, g_j}) e^{-\hat{x}_{e, g_i, g_j}} \cdot \frac{\partial}{\partial \bar{g}_j} \hat{x}_{e, g_i, g_j} + \lambda_G \bar{g}_j \right)$$

Step 16. 从  $R_{EC}$  中提取  $(e, c_i, c_j)$

$$\text{Step 17. } \bar{e} \leftarrow \bar{e} + \alpha \left( \omega_{EC} \sigma(\hat{x}_{e, c_i, c_j}) e^{-\hat{x}_{e, c_i, c_j}} \cdot \frac{\partial}{\partial \bar{e}} \hat{x}_{e, c_i, c_j} + \lambda_E \bar{e} \right)$$

$$\text{Step 18. } \bar{c}_i \leftarrow \bar{c}_i + \alpha \left( \omega_{EC} \sigma(\hat{x}_{e, c_i, c_j}) e^{-\hat{x}_{e, c_i, c_j}} \cdot \frac{\partial}{\partial \bar{c}_i} \hat{x}_{e, c_i, c_j} + \lambda_C \bar{c}_i \right)$$

$$\text{Step 19. } \bar{c}_j \leftarrow \bar{c}_j + \alpha \left( \omega_{EC} \sigma(\hat{x}_{e, c_i, c_j}) e^{-\hat{x}_{e, c_i, c_j}} \cdot \frac{\partial}{\partial \bar{c}_j} \hat{x}_{e, c_i, c_j} + \lambda_C \bar{c}_j \right)$$

Step 20. 从  $R_{EL}$  中提取  $(e, l_i, l_j)$

$$\text{Step 21. } \bar{e} \leftarrow \bar{e} + \alpha \left( \omega_{EL} \sigma(\hat{x}_{e, l_i, l_j}) e^{-\hat{x}_{e, l_i, l_j}} \cdot \frac{\partial}{\partial \bar{e}} \hat{x}_{e, l_i, l_j} + \lambda_E \bar{e} \right)$$

$$\text{Step 22. } \bar{l}_i \leftarrow \bar{l}_i + \alpha \left( \omega_{EL} \sigma(\hat{x}_{e, l_i, l_j}) e^{-\hat{x}_{e, l_i, l_j}} \cdot \frac{\partial}{\partial \bar{l}_i} \hat{x}_{e, l_i, l_j} + \lambda_L \bar{l}_i \right)$$

$$\text{Step 23. } \bar{l}_j \leftarrow \bar{l}_j + \alpha \left( \omega_{EL} \sigma(\hat{x}_{e, l_i, l_j}) e^{-\hat{x}_{e, l_i, l_j}} \cdot \frac{\partial}{\partial \bar{l}_j} \hat{x}_{e, l_i, l_j} + \lambda_L \bar{l}_j \right)$$

Step 24. 从  $R_{ET}$  中提取  $(e, t_i, t_j)$

$$\text{Step 25. } \bar{e} \leftarrow \bar{e} + \alpha \left( \omega_{ET} \sigma(\hat{x}_{e,t_i,t_j}) e^{-\hat{x}_{e,t_i,t_j}} \cdot \frac{\partial}{\partial \bar{e}} \hat{x}_{e,t_i,t_j} + \lambda_E \bar{e} \right)$$

$$\text{Step 26. } \bar{t}_i \leftarrow \bar{t}_i + \alpha \left( \omega_{ET} \sigma(\hat{x}_{e,t_i,t_j}) e^{-\hat{x}_{e,t_i,t_j}} \cdot \frac{\partial}{\partial \bar{t}_i} \hat{x}_{e,t_i,t_j} + \lambda_T \bar{t}_i \right)$$

$$\text{Step 27. } \bar{t}_j \leftarrow \bar{t}_j + \alpha \left( \omega_{ET} \sigma(\hat{x}_{e,t_i,t_j}) e^{-\hat{x}_{e,t_i,t_j}} \cdot \frac{\partial}{\partial \bar{t}_j} \hat{x}_{e,t_i,t_j} + \lambda_T \bar{t}_j \right)$$

Step 28. until 收敛

Step 29. 由  $\hat{R} = UE^T$  给出每位用户  $u \in U$  的推荐结果列表 top- $n(u)$

现有的个性化同城活动推荐算法在利用上下文信息时,大多都是预先从用户参与活动记录中间接学习用户上下文信息偏好,而没有考虑后续用户偏好建模过程中出现的用户对活动偏好的差异的影响.本文所提算法从用户参与活动记录出发,以关系矩阵的形式将组织关系、活动内容、举办时间、举办地点等相关上下文信息表达为上下文关系,并采用多关系贝叶斯个性化排序方法进行协同上下文关系学习,从而对用户偏好进行建模并实现个性化同城活动推荐.该算法充分考虑了用户参与活动偏好与上下文信息之间、多种上下文信息之间潜在的交叉影响关系,并将其应用到用户偏好建模和个性化推荐中,有利于提高推荐结果的有效性.

## 5 实验结果与分析

### 5.1 Meetup数据集

实验采用的数据集是文献[19]中提供的 Meetup 数据集.Meetup 数据集由 Macedo 等人使用 Meetup REST API,以美国亚利桑那州的 Phoenix 市、伊利诺伊州的 Chicago 市、加利福尼亚州的 San Jose 市作为采集目标,抓取了 2010 年 1 月~2014 年 4 月在 Meetup.com 上公开的这 3 个城市中的同城活动信息,其中包含了用户参与活动记录、用户加入的活动群组、活动群组主办的同城活动、活动的内容文本、举办地点、举办时间等信息.采集到的数据规模见表 1.

Table 1 Statistics of the Meetup dataset

表 1 Meetup 数据集规模统计

城市	用户数量 U	活动数量 E	活动群组数量 G	用户反馈数量	稀疏程度(%)
Chicago	207 649	190 927	2 321	1 375 154	99.99
Phoenix	117 458	222 632	1 661	1 209 324	99.99
San Jose	242 143	206 682	2 589	1 607 985	99.99

### 5.2 实验设计与基准算法

由于原数据集的整体规模较大,直接用于实验的计算时间较长,而且传统的单组训练集-测试集只能模拟 1 次推荐,不能模拟实际场景中用户在不同时间访问推荐系统获得不同推荐结果的情况,因此,参考 Macedo 等人的实验设置<sup>[19]</sup>,对 Meetup 数据集进行一定的处理.在 Meetup 数据集,在每个城市数据集中,采用滑动窗口的方法均匀提取 12 个时间点作为推荐时间点,即划分为 12 组数据集.在每组数据集中,将推荐时间点前 6 个月中的有效活动作为训练活动;将推荐时间点前 6 个月内创建,并且在推荐时间点之后举行的有效活动作为候选活动.将训练活动的用户参与情况、候选活动在推荐时间点前已有的用户参与情况作为训练集,将候选活动中在推荐时间点之后产生的用户参与情况作为测试集.由此可以模拟真实推荐系统中,用户在某一时刻登录推荐系统所看到的最近 6 个月内创建的、即将举办的同城活动的推荐结果,验证用户对推荐给他的活动的反馈.

每个城市数据集中,切分后得到的 12 组数据集的平均规模见表 2.

**Table 2** Statistics of partitioned dataset of each city**表 2** 各城市划分后的数据集规模统计

城市	用户数量 U	活动数量 E	活动群组数量 G	用户反馈数量	稀疏程度(%)
Chicago	24 317	19 043	2 321	86 478	99.99
Phoenix	15 143	22 099	1 661	74 389	99.99
San Jose	29 679	20 164	2 589	107 380	99.99

采用滑动窗口方法划分的 12 组数据集,分别反映 2010 年 1 月~2014 年 4 月期间,多个不同时间点上的用户参与活动情况.处理后的数据集有足够长的时间跨度,既符合用户实际使用 EBSN 和推荐系统的情景,也具有足够大的训练和测试规模.相对于将全部数据集切分为一组训练集-测试集,这种实验方式能够验证推荐算法在很长一段时间内的多个时间点上的推荐效果,从而比较推荐算法在很长一段时间内的推荐性能.

为了验证本文所提算法的有效性,采用以下算法作为基准算法,与本文所提算法进行比较.

- (1) **Most-Popular**(近期最流行)推荐算法<sup>[32]</sup>:以用户反馈数量作为流行度指标,向用户推荐近期最流行的同城活动.该方法被视为最基础的基准算法.
- (2) **MCLRE**算法<sup>[19]</sup>:由 Macedo 等人于 2015 年提出,结合活动上下文,从社交关系(主要是群组和关系模型方面)、内容、地点和时间等 4 个方面提取活动特征,利用学习排序方法融合上述特征并进行活动推荐.
- (3) **GLFM**算法<sup>[23]</sup>:由 Jhamb 等人于 2017 年提出,从用户和活动两个方面考察活动群组(主办方)对用户参与活动的影响,利用隐式因子模型分析活动群组(主办方)、举办地点、流行度、空间距离、举办时间等特征,基于贝叶斯个性化排序方法进行活动推荐.
- (4) **CPMF**算法<sup>[24]</sup>:由 Liu 等人于 2017 年提出,主要分析了用户-活动-群组和用户-活动-地点这两种三角关系,并建立了协同对级矩阵分解模型对偏好关系进行矩阵分解,从而对即将举办的同城活动进行推荐.

### 5.3 评价指标

准确率和召回率是进行推荐算法效果评价的常用指标,本文除了使用这两项指标外,还使用归一化折损累计增益(normalized discounted cumulative gain,简称 nDCG)<sup>[33]</sup>作为算法推荐效果的评价指标.因为对于进行 TopN 推荐的算法,用户感兴趣的项目在推荐结果列表中应当尽可能地靠前.为了更好地衡量排序学习推荐的效果,使用 nDCG 验证所提算法的有效性.

#### (1) 准确率

算法为用户  $u$  推荐的  $N$  个物品集合记为  $R_u$ ,用户  $u$  在测试集中实际偏好的物品集合记为  $T_u$ ,从而推荐的准确率定义如公式(13)所示.

$$precision @ N = \frac{\sum_u |R_u \cap T_u|}{\sum_u |R_u|} \quad (13)$$

#### (2) 召回率

算法为用户  $u$  推荐的  $N$  个物品集合记为  $R_u$ ,用户  $u$  在测试集中实际偏好的物品集合记为  $T_u$ ,从而推荐的召回率定义如公式(14)所示.

$$recall @ N = \frac{\sum_u |R_u \cap T_u|}{\sum_u |T_u|} \quad (14)$$

#### (3) 归一化折损累计增益

nDCG 指标被用来更好地衡量 TopN 推荐结果列表的准确性.在 nDCG 中,个性化推荐的准确率基于推荐项目在测试集中的真实顺序,用折损累计增益(discounted cumulative gain,简称 DCG)来表示,如公式(15)所示.

$$DCG_n = \sum_{i=1}^n \frac{r_i}{\log_2(i+1)} \tag{15}$$

nDCG 由推荐 DCG 值与理想 DCG 值的比值得到,如公式(16)所示.

$$nDCG_n = \frac{DCG_n}{IDCG_n} \tag{16}$$

### 5.4 实验结果与分析

参考 Macedo 等人在应用 MRBPR 算法学习活动组织关系中的参数设置<sup>[19]</sup>,设隐式因子维数为 300,学习率为 0.1,迭代次数设为 1 500 次,正则化系数设为 0.1.基准算法和提出的 CCRL 算法均在每个城市的 12 组数据集上进行实验,计算相关实验评价指标并取平均值作为算法性能的最终指标.

- 实验 1. 推荐准确率实验

推荐准确率指标侧重考查同城活动推荐算法的活动推荐结果中用户实际参与的活动数量.各项对比算法在 San Jose、Chicago、Phoenix 这 3 个城市数据集中的最终推荐准确率如图 2 所示.

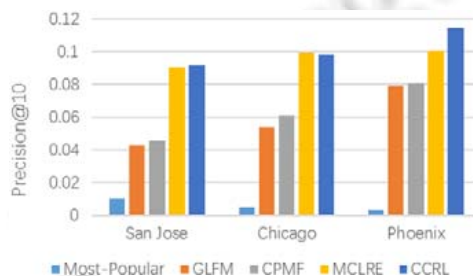


Fig.2 Precisions of different algorithms in different cities

图 2 不同城市中,不同算法的推荐准确率

从图 2 可以观察到:在 3 个城市数据集中,本文所提的 CCRL 算法的推荐准确率都要优于 Most-Popular 算法、GLFM 基准算法、CPMF 算法;在 San Jose 和 Phoenix 城市数据集中,也要优于 MCLRE 算法,这表明综合考虑多种上下文信息,通过协同上下文关系学习充分利用用户参与活动偏好与上下文信息之间的潜在影响关系,能够优化用户偏好建模的效果,提高推荐的准确性.

由 Most-Popular 算法较低的推荐准确率可以看到,活动流行度对用户参与活动偏好的影响较小.其他 4 种算法充分利用了用户参与活动记录和上下文信息进行用户参与活动偏好建模,特别是 GLFM 算法还从用户和活动两个角度来考虑活动群组信息对于用户参与活动偏好的影响.然而,GLFM 算法所采用的贝叶斯个性化排序方法从单一矩阵分解方法衍生而来,只能学习两种实体的隐式因子,如用户和活动.为了在贝叶斯个性化排序方法中融入上下文特征,GLFM 算法分别将其叠加到用户隐式因子、活动隐式因子、偏置项上,但依然无法脱离单一矩阵分解的框架.这种做法实际上压缩了数据维度,削弱了利用多种上下文信息进行用户偏好建模的优势,使其推荐准确率并不如基于协同矩阵分解思想的 CPMF 算法、MCLRE 算法、本文提出的 CCRL 算法.

CPMF 算法在推荐准确率上较优于 Most-Popular 算法和 GLFM 算法,但要低于 MCLRE 算法和本文提出的 CCRL 算法.由于 CPMF 算法采用协同概率矩阵分解方法对用户-活动-群组 and 用户-活动-地点这两种三角关系进行联合建模,其推荐效果比基于单一矩阵分解思想的 GLFM 算法要好.但 CPMF 算法没有考虑活动内容、活动时间等上下文信息,当输入数据中用户参与活动记录较少、活动群组和活动地点关系信息不足时,其推荐效果比采用了更多上下文信息的 MCLRE 算法要差.另外,CPMF 算法从用户-活动-群组、用户-活动-地点这两种三角关系中分别统计得到用户-群组偏好矩阵、用户-地点偏好矩阵,活动作为用户偏好的实际对象从这两个矩阵中消失,实际上是忽略了活动与群组、地点之间的相互影响关系;而本文提出的 CCRL 算法考虑了更多的上下文信息,并充分利用了用户参与活动偏好与上下文信息之间的相互影响关系,从而推荐准确率优于 CPMF

算法.

在 Chicago 城市数据集中,CCRL 算法的推荐准确率略低于 MCLRE 算法,但在其他两个城市中仍要优于 MCLRE 算法.这是由于 MCLRE 算法中也结合用户参与活动记录和组织关系上下文进行了用户偏好建模,而在 Meetup 网站上,用户加入的活动群组对用户参与活动行为具有较大影响,能够在很大程度上反映用户的参与活动偏好.此时引入其他上下文信息未必会明显提高推荐效果,反而可能引入用户偏好噪声,降低推荐准确率.

#### • 实验 2. 推荐召回率实验

推荐召回率指标考察用户实际参与的同城活动中被推荐算法正确推荐的数量,从另外一个角度考察推荐算法对用户偏好建模的效果.各项对比算法在 San Jose、Chicago、Phoenix 这 3 个城市数据集中的推荐召回率如图 3 所示.

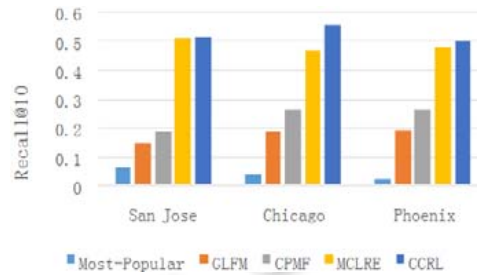


Fig.3 Recalls of different algorithms in different cities

图 3 不同城市中,不同算法的推荐召回率

从图 3 可以观察到,本文所提的 CCRL 算法在 3 个城市数据集中的推荐召回率都要优于其他基准算法.由于用户之间的个性化需求存在较大差异,最流行的活动并不能满足用户的个性化偏好,因此,Most-Popular 算法的推荐召回率是最低的.

从后 4 种算法的推荐召回率可以看出,充分利用上下文信息能有效提高用户偏好的学习准确性.GLFM 算法在 4 种结合上下文信息进行用户偏好建模的同城活动推荐算法中表现是最差的.尽管 GLFM 算法采用了多种上下文信息特征,如活动群组、活动地点、活动流行度、位置距离、活动时间等,还分别从用户和活动两个角度利用活动群组特征,但其所采用的贝叶斯个性化排序模型并不能充分利用这些特征的交叉影响关系,从而影响了其学习用户偏好、进行同城活动推荐的效果.

CPMF 算法在推荐召回率指标上的表现介于 GLFM 算法和 MCLRE 算法之间,远不如本文提出的 CCRL 算法.与 GLFM 算法采用贝叶斯个性化排序模型容纳多种上下文信息相比,CPMF 算法采用协同矩阵分解模型,利用多种上下文信息及其关系进行用户偏好学习.尽管 CPMF 算法提取的上下文信息特征较少,其推荐效果依然优于 GLFM 算法.MCLRE 算法一方面采用 MRBPR 算法处理用户-活动-群组关系,另一方面采用统计方法处理活动内容、活动地点、活动时间等上下文信息,融合多种特征数据进行用户偏好学习.这样得到的用户偏好模型更为精准,其推荐效果也比 CPMF 算法要好.本文提出的 CCRL 算法比 CPMF 算法考虑了更多的上下文信息,比 MCLRE 算法采用了统一的用户偏好模型并充分利用了其中的潜在影响关系,取得了良好的推荐效果.

#### • 实验 3. 推荐结果列表准确性实验

上述实验将推荐结果视为无序集合,尽管推荐结果在列表中的顺序不一致,但对于推荐指标的影响权重都是均一的.然而实际情况中,用户总是希望满足自己偏好的推荐结果在 TopN 推荐列表中尽量靠前.为了衡量推荐结果列表是否与用户偏好尽量匹配,采用 nDCG 指标对推荐结果列表进行评估,推荐结果数量为 10 个.评价结果如图 4 所示.



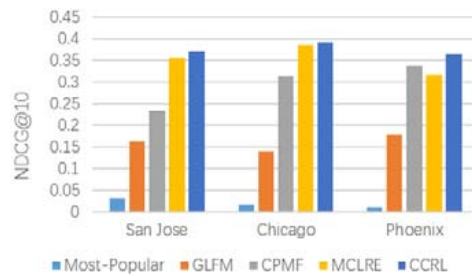


Fig.4 nDCGs of different algorithms in different cities

图 4 不同城市中,不同算法的 nDCG 指标

从图 4 可以观察到,无论是在哪一个城市数据集中,本文所提的 CCRL 算法的推荐结果列表都要优于其他基准算法的结果. Most-Popular 算法按活动流行度生成的推荐结果列表并不适用于大部分用户.特别是在用户历史参与活动记录极其稀疏的情况下,用户参与与流行度较高的活动的可能性很小.

GLFM 采用贝叶斯个性化排序方法作为核心思想,然而这种方法在利用多种上下文信息进行 TopN 推荐时存在明显不足:一方面,受单一矩阵分解方法的局限,无法充分利用多种上下文信息特征学习用户偏好;另一方面,作为对级排序学习方法,只能考虑两个同城活动的偏序关系,无法考虑最终推荐结果列表中活动的位置.因此, GLFM 算法利用多种上下文信息学习用户偏好的效果不佳,最终推荐效果远低于结合了协同矩阵分解和列表级排序学习技术的 MCLRE 算法、本文提出的协同上下文关系学习 CCRL 算法.

CPMF 算法在 nDCG 指标上的表现明显要优于 Most-Popular 算法和 GLFM 算法,在 Phoenix 城市的数据集上甚至略优于 MCLRE 算法,但依然不如本文提出的 CCRL 算法.由此可见,在基于多种特征信息进行用户偏好学习和同城活动推荐时,采用协同矩阵分解思想的方法比基于单一矩阵分解思想的方法具有明显的优势.通过分析不同城市数据集的数据规模发现,在 3 个实验城市数据集中, Phoenix 城市数据集的活动数量最多,而用户数量和活动群组数量最少,表明 Phoenix 城市数据集中的用户-活动-群组三角关系相对于其他数据集更加密集.尽管 CPMF 算法和 MCLRE 算法都考虑并且利用了用户-活动-群组三角关系,但 MCLRE 算法比 CPMF 算法引入了更多的上下文特征信息,可能对用户偏好学习带来了一定的干扰,因此, CPMF 取得了比 MCLRE 算法更好的推荐效果,但依然不如充分利用了上下文信息之间潜在影响关系的 CCRL 算法.

MCLRE 算法与本文提出的 CCRL 算法都充分利用了多种上下文信息,有效提高了推荐结果列表的质量,使真正符合用户偏好的活动能够进入推荐结果列表的上层,表明采用较多的上下文特征信息有利于用户偏好学习.但 CCRL 算法考虑到了用户对活动偏好中隐藏的用户参与活动与相关上下文信息之间、各种相关上下文信息之间的交叉影响关系,基于多关系贝叶斯个性化排序方法充分利用了这些影响关系学习用户对活动的偏好,并充分用于对候选活动的排序推荐上,从而获得了比其他 4 种基准算法更好的推荐效果.

## 6 结 论

基于活动的社交网络的迅速发展,在为用户提供更多有趣的活动的同时,也带来了显著的信息过载问题,从而推动同城活动推荐成为了研究热点.然而,现有同城活动推荐算法主要从用户参与活动记录间接统计用户对上下文信息的偏好,忽略了两之间潜在的交叉影响关系.本文对基于活动的社交网络进行建模,形式化定义了同城活动推荐问题,并在分析已有算法不足之处的基础上,提出了基于协同上下文关系学习的同城活动推荐算法(CCRL).该算法以用户参与活动记录为基础,融合分析了组织关系、活动内容、举办地点、举办时间等上下文信息,采用多关系贝叶斯个性化排序方法进行协同上下文关系学习,从而实现个性化同城活动推荐.在 Meetup 数据集上的实验结果表明, CCRL 算法的推荐结果在推荐准确率、召回率和 nDCG 等指标上都要优于现有主流算法.

针对多种上下文信息的协同关系学习,还有很多值得进一步研究的问题.例如,

- (1) 现有的多关系贝叶斯个性化排序方法以贝叶斯个性化排序方法这种对级排序学习方法为基础,为了取得更好的学习效果,探索采用列表级排序学习方法进行多种偏序关系的协同学习;
- (2) 进一步考察多种上下文信息之间的关联关系,发掘上下文关系之间的互相影响;
- (3) 考虑用户历史上在其他城市参与的活动信息,探索跨城市的用户偏好学习.

#### References:

- [1] Liu XI, He Q, Tian YY, Lee WC, McPherson J, Han JW. Event-based social networks: Linking the online and offline social worlds. In: Yang Q, Deepak A, Pei J, eds. Proc. of the 18th ACM SIGKDD Int'l Conf. on Knowledge discovery and data mining. New York: ACM, 2012. 1032–1040. [doi: 10.1145/2339530.2339693]
- [2] Resnick P, Varian HR. Recommender systems. Communications of the ACM, 1997,40(3):56–58. [doi: 10.1145/245108.245121]
- [3] Meng XW, Liu SD, Zhang YJ, Hu X. Research on social recommender systems. Ruan Jian Xue Bao/Journal of Software, 2015, 26(6):1356–1372 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4831.htm> [doi: 10.13328/j.cnki.jos.004831]
- [4] Yang XW, Guo Y, Liu Y, Steck H. A survey of collaborative filtering based social recommender systems. Computer Communications, 2014,41:1–10. [doi: 10.1016/j.comcom.2013.06.009]
- [5] Adomavicius G, Sankaranarayanan R, Sen S, Tuzhilin A. Incorporating contextual information in recommender systems using a multidimensional approach. ACM Trans. on Information Systems, 2005,23(1):103–145. [doi: 10.1145/1055709.1055714]
- [6] Wang LC, Meng XW, Zhang YJ. Context-aware recommender systems. Ruan Jian Xue Bao/Journal of Software, 2012,23(1):1–20 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4100.htm> [doi: 10.3724/SP.J.1001.2012.04100]
- [7] Yuan Q, Cong G, Ma ZY, Sun A, Magnenat-Thalmann N. Time-aware point-of-interest recommendation. In: Jones GJF, Sheridan P, Kelly D, Rijke M, Sakai T, eds. Proc. of the 36th Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval. New York: ACM, 2013. 363–372. [doi: 10.1145/2484028.2484030]
- [8] Liu X, Liu Y, Aberer K, Miao C. Personalized point-of-interest recommendation by mining users' preference transition. In: He Q, Iyengar A, Nejdi W, Pei J, Rastogi R, eds. Proc. of the 22nd ACM Int'l Conf. on Information & Knowledge Management. New York: ACM, 2013. 733–738. [doi: 10.1145/2505515.2505639]
- [9] Zhang W, Wang JY. Location and time aware social collaborative retrieval for new successive point-of-interest recommendation. In: Bailey J, Moffat A, Aggarwal CC, Rijke M, Kumar R, Murdock V, Sellis T, Yu XJ, eds. Proc. of the 24th ACM Int'l on Conf. on Information and Knowledge Management. New York: ACM, 2015. 1221–1230. [doi: 10.1145/2806416.2806564]
- [10] Zhang JD, Chow CY. GeoSoCa: Exploiting geographical, social and categorical correlations for point-of-interest recommendations. In: Baeza-Yates R, Lalmas M, Moffat A, Ribeiro-Neto B, eds. Proc. of the 38th Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval. New York: ACM, 2015. 443–452. [doi: 10.1145/2766462.2767711]
- [11] Chaney AJ, Blei DM, Eliassi-Rad T. A probabilistic model for using social networks in personalized item recommendation. In: Werthner H, Zanker M, Golbeck J, Semeraro G, eds. Proc. of the 9th ACM Conf. on Recommender Systems. New York: ACM, 2015. 43–50. [doi: 10.1145/2792838.2800193]
- [12] Tang J, Wang S, Hu X, Yin D, Bi Y, Chang Y, Liu H. Recommendation with social dimensions. In: Zilberstein S, Schuurmans D, Wellman M, eds. Proc. of the 30th AAAI Conf. on Artificial Intelligence. Palo Alto: AAAI, 2016. 251–257.
- [13] Zou BY, Li CP, Tan LW, Chen H, Wang SQ. Social recommendations based on user trust and tensor factorization. Ruan Jian Xue Bao/Journal of software, 2014,25(12):2852–2864 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4725.html> [doi: 10.13328/j.cnki.jos.004725]
- [14] Hu X, Meng XW, Zhang YJ. Recommendation algorithm combing item features and trust relationship of mobile users. Ruan Jian Xue Bao/Journal of Software, 2014,25(8):1817–1830 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4491.htm> [doi: 10.13328/j.cnki.jos.004491]
- [15] Han JW, Niu JW, Chin A, Wang W, Tong C, Wang X. How online social network affects offline events: A case study on douban. In: Proc. of the 2012 9th Int'l Conf. on Ubiquitous Intelligence & Computing and 9th Int'l Conf. on Autonomic & Trusted Computing (UIC/ATC). Fukuoka: IEEE, 2012. 752–757. [doi: 10.1109/UIC-ATC.2012.149]



- [16] Xu B, Chin A, Cosley D. On how event size and interactivity affect social networks. In: Mackay WE, Brewster S, Bødker S, eds. Proc. of the Extended Abstracts on Human Factors in Computing Systems (CHI 2013). Paris: ACM, 2013. 865–870. [doi: 10.1145/2468356.2468511]
- [17] Du R, Yu ZW, Mei T, Wang ZT, Wang Z, Guo B. Predicting activity attendance in event-based social networks: Content, context and social influence. In: Brush AJ, Friday A, Kientz J, Scott J, Song JH, eds. Proc. of the 2014 ACM Int'l Joint Conf. on Pervasive and Ubiquitous Computing. Seattle: ACM, 2014. 425–434. [doi: 10.1145/2632048.2632063]
- [18] Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. *Journal of Machine Learning Research*, 2003,3:993–1022.
- [19] Macedo AQ, Marinho LB, Santos RL. Context-aware event recommendation in event-based social networks. In: Werthner H, Zanker M, Golbeck J, Semeraro G, eds. Proc. of the 9th ACM Conf. on Recommender Systems. New York: ACM, 2015. 123–130. [doi: 10.1145/2792838.2800187]
- [20] Huang ZH, Zhang JW, Tian CQ, Sun SL, Xiang Y. Survey on learning-to-rank based recommendation algorithms. *Ruan Jian Xue Bao/Journal of Software*, 2016,27(3):691–713 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4948.htm> [doi: 10.13328/j.cnki.jos.004948]
- [21] Zhang W, Wang JY. A collective bayesian poisson factorization model for cold-start local event recommendation. In: Cao LB, Zhang CQ, Joachims T, Webb G, Margineantu DD, Williams G, eds. Proc. of the 21th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. New York: ACM, 2015. 1455–1464. [doi: 10.1145/2783258.2783336]
- [22] Gu YL, Song JX, Liu WD, Zou LX, Yao Y. Context aware matrix factorization for event recommendation in event-based social networks. In: Proc. of the 2016 IEEE/WIC/ACM Int'l Conf. on Web Intelligence (WI). Omaha: IEEE, 2016. 248–255. [doi: 10.1109/WI.2016.0043]
- [23] Jhamb Y, Fang Y. A dual-perspective latent factor model for group-aware social event recommendation. *Information Processing & Management*, 2017,53(3):559–576. [doi: 10.1016/j.ipm.2017.01.001]
- [24] Liu CY, Zhou C, Wu J, Xie HT, Hu Y, Guo L. CPMF: A collective pairwise matrix factorization model for upcoming event recommendation. In: Proc. of the 2017 Int'l Joint Conf. on Neural Networks (IJCNN). Anchorage: IEEE, 2017. 1532–1539. [doi: 10.1109/IJCNN.2017.7966033]
- [25] Adomavicius G, Tuzhilin A. Context-aware recommender systems. In: Ricci F, Rokach L, Shapira B, Kantor PB, eds. *Recommender Systems Handbook*. Boston: Springer-Verlag, 2011. 217–253. [doi: 10.1007/978-0-387-85820-3\_7]
- [26] Adomavicius G, Tuzhilin A. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Trans. on Knowledge and Data Engineering*, 2005,17(6):734–749. [doi: 10.1109/TKDE.2005.99]
- [27] Krohn-Grimberghe A, Drumond L, Freudenthaler C, Schmidt-Thieme L. Multi-relational matrix factorization using Bayesian personalized ranking for social network data. In: Adar E, Teevan J, Agichtein E, Maarek Y, eds. Proc. of the 5th ACM Int'l Conf. on Web Search and Data Mining. New York: ACM, 2012. 173–182. [doi: 10.1145/2124295.2124317]
- [28] Singh AP, Gordon GJ. Relational learning via collective matrix factorization. In: Li Y, Liu B, Sarawagi S, eds. Proc. of the 14th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. New York: ACM, 2008. 650–658. [doi: 10.1145/1401890.1401969]
- [29] Rendle S, Freudenthaler C, Gantner Z, Schmidt-Thieme L. BPR: Bayesian personalized ranking from implicit feedback. In: Proc. of the 25th Conf. on Uncertainty in Artificial Intelligence (UAI 2009). Arlington: AUAI, 2009. 452–461.
- [30] Koren Y, Bell R, Volinsky C. Matrix factorization techniques for recommender systems. *Computer*, 2009,42(8):30–37. [doi: 10.1109/MC.2009.263]
- [31] Bottou L. Large-scale machine learning with stochastic gradient descent. In: Lechevallier Y, Saporta G, eds. Proc. of the 19th Int'l Conf. on Computational Statistics (COMPSTAT 2010). Berlin: Physica-Verlag HD, 2010. 177–186. [doi: 10.1007/978-3-7908-2604-3\_16]
- [32] Gantner Z, Rendle S, Freudenthaler C, Schmidt-Thieme L. MyMediaLite: A free recommender system library. In: Mobasher B, Burke R, Jannach D, Adomavicius G, eds. Proc. of the 5th ACM Conf. on Recommender Systems. New York: ACM, 2011. 305–308. [doi: 10.1145/2043932.2043989]
- [33] Järvelin K, Kekäläinen J. Cumulated gain-based evaluation of IR techniques. *ACM Trans. on Information Systems (TOIS)*, 2002, 20(4):422–446. [doi: 10.1145/582415.582418]

## 附中文参考文献:

- [3] 孟祥武,刘树栋,张玉洁,胡勋.社会化推荐系统研究.软件学报,2015,26(6):1356-1372. <http://www.jos.org.cn/1000-9825/4831.htm> [doi: 10.13328/j.cnki.jos.004831]
- [6] 王立才,孟祥武,张玉洁.上下文感知推荐系统.软件学报,2012,23(1):1-20. <http://www.jos.org.cn/1000-9825/4100.htm> [doi: 10.3724/SP.J.1001.2012.04100]
- [13] 邹本友,李翠平,谭力文,陈红,王绍卿.基于用户信任和张量分解的社会网络推荐.软件学报,2014,25(12):2852-2864. <http://www.jos.org.cn/1000-9825/4725.htm> [doi: 10.13328/j.cnki.jos.004725]
- [14] 胡勋,孟祥武,张玉洁,史艳翠.一种融合项目特征和移动用户信任关系的推荐算法.软件学报,2014,25(8):1817-1830. <http://www.jos.org.cn/1000-9825/4491.htm> [doi: 10.13328/j.cnki.jos.004491]
- [20] 黄震华,张佳雯,田春岐,孙圣力,向阳.基于排序学习的推荐算法研究综述.软件学报,2016,27(3):691-713. <http://www.jos.org.cn/1000-9825/4948.htm> [doi: 10.13328/j.cnki.jos.004948]



赖奕安(1994-),男,广东茂名人,硕士生,主要研究领域为推荐系统.



杜雨露(1987-),男,博士,主要研究领域为个性化推荐系统.



张玉洁(1969-),女,副教授,主要研究领域为网络服务,用户需求,推荐服务.



孟祥武(1966-),男,博士,教授,博士生导师,CCF高级会员,主要研究领域为网络服务,用户需求,推荐服务.