

LBSN 上基于兴趣圈中社会关系挖掘的推荐算法

李鑫^{1,2} 刘贵全¹ 李琳³ 吴宗大⁴ 丁君美¹

¹(中国科学技术大学计算机科学与技术学院 合肥 230022)

²(科大讯飞股份有限公司大数据研究院 合肥 230088)

³(武汉理工大学计算机科学与技术学院 武汉 430070)

⁴(温州大学瓯江学院计算机科学与技术学院 浙江温州 325035)

(xinli2@iflytek.com)

Circle-Based and Social Connection Embedded Recommendation in LBSN

Li Xin^{1,2}, Liu Guiquan¹, Li Lin³, Wu Zongda⁴, and Ding Junmei¹

¹(School of Computer Science and Technology, University of Science and Technology of China, Hefei 230022)

²(Big Data Research Institute, iFlytek Co., Ltd, Hefei 230088)

³(School of Computer Science and Technology, Wuhan University of Technology, Wuhan 430070)

⁴(School of Computer Science and Technology, Oujiang College Wenzhou University, Wenzhou, Zhejiang 325035)

Abstract With the pervasiveness of GPS-enabled smart phones, people tend to share their locations online or check in at somewhere by commenting on the merchants, thus arousing the prevalence of LBSN (location based social network), which takes POIs (point-of-interests) as the center. A typical application in social networks is the recommendation system, and the most common problem in recommendation system is cold start, that is, how to recommend for the users who rarely comment on the item or share comments. In this paper, we propose a recommendation algorithm based on circle and social connections in social networks. The circle is made up by all users who visit a particular category of items and their social connections. It means he is interested in this category that a user accesses the category of items. Our algorithm considers different social connections and circles on tradition matrix factorization. The social connections we use include the relationship between friends (explicit relation) and relevant experts(implicit), which are used as the rule to optimize the matrix factorization model. Experiments are conducted on the datasets from the 5th Yelp Challenge Round and Foursquare. Experimental results demonstrate that our approach outperforms traditional matrix factorization based methods, especially in solving cold-start problem.

Key words point-of-interests (POIs); recommendation; interest circle; social connection; cold-start

摘要 随着带有GPS定位功能的智能手机越来越普遍,人们喜欢分享他们的地理位置或者通过评论某个地方的商品从而留下用户的足迹,这引发了以共同的兴趣点(POIs)为中心,基于地理位置信息的社交网络研究(location based social network, LBSN). 社交网络中的一类典型应用是推荐系统,而推荐

收稿日期:2015-09-01;修回日期:2016-06-30

基金项目:国家自然科学基金项目(61202171);浙江省自然科学基金项目(LY15F020020);国家“八六三”高技术研究发展计划基金项目(2015AA015403)

This work was supported by the National Natural Science Foundation of China (61202171), the Natural Science Foundation of Zhejiang Province (LY15F020020), and the National High Technology Research and Development Program of China (863 Program) (2015AA015403).

通讯作者:刘贵全(gqliu@ustc.edu.cn)

系统中最常见的问题是冷启动,即在用户很少点评商家或分享评论时如何为他推荐感兴趣的商家.为解决冷启动问题,提出了一种在社交网络中基于兴趣圈的社会关系挖掘推荐算法.兴趣圈是由所有访问某一类别商品的用户群及他们之间的社会关系构成的社交联系,不同的用户访问同一类别商品表明他们对此类别具有相似兴趣.该方法在传统矩阵分解模型的基础上考虑不同的兴趣圈上的社会关系,使用的社会关系包括朋友关系(显性关系)和相关专家(隐性关系),并用它们作为规则化项来优化矩阵分解模型.实验数据集来自第 5 届 Yelp 挑战赛和自己爬取的 Foursquare 数据集,提出的方法与已有模型进行了充分的实验对比分析,结果表明,我们的模型特别是在解决冷启动问题方面优于多种现有的方法.

关键词 兴趣点;推荐;兴趣圈;社会关系;冷启动

中图法分类号 TP181

近年来兴起了基于地理位置信息服务的热潮,如 Foursquare 和 Yelp 挑战赛.用户在这种在线服务上可以通过评论或者给商家评分分享他们的经验.同时,他们可以加好友或者成为别人的粉丝,从而形成了社交网络.上面的 2 个特性是基于位置的社交网络(location based social network, LBSN)的基础.推荐算法可以帮助用户在 LBSN 中找到自己感兴趣的商家,成为学术界和工业界关注和研究的热点问题.

众所周知,影响推荐系统性能的 2 个主要问题是用户冷启动和数据稀疏,而冷启动问题研究的是在用户很少点评商家或分享评论时如何为他推荐感兴趣的商家.解决这类问题的可行性方案之一是将社会关系融入到推荐模型中.一个用户的品味跟与他有社会关系的人比较相似或受他们影响,这被称为社会学中的同质性.本文使用的社会关系包括朋友和专家关系.尽管现有的推荐系统在不断改进,但在 LBSN 上对社会关系的挖掘还存在很大的空间可以提升推荐系统的性能.其一,当位置信息嵌入到社交网络中形成新的 LBSN 时,社会联系将会扩大.除了现有的 2 个用户之间的关系外,另一个联系则是用户的位置信息.如 2 个用户访问了同一个商品时,说明他们共处于同一地理位置,那么他们很可能被联系起来^[1].其二,除了受自身朋友影响外,用户也可能受到访问过同类商品的专家的评分和评论影响.这是因为在某个商品的网页上,专家的评论在这类商品中总是排名很高,从而影响后面的用户.最后,传统利用社会关系的方法解决冷启动问题的假设是:虽然用户访问的商品比较少,但是他有足够的社会关系,我们可以通过学习他的朋友或者这方面专家的意见来给他们推荐.但当用户访问商品很少时,他很可能也没有丰富的社会关系.

本文针对上述冷启动和稀疏问题,提出一种基

于相同兴趣圈的推荐系统,使用社交网络为辅助信息来提高 POI 模型准确率.兴趣圈是由所有访问某一类别商品的用户群及他们之间的社会关系构成的社交联系,不同的用户访问同一类别商品表明他们对此类别具有相似兴趣.模型同时采用了朋友和专家 2 种社会关系作为矩阵分解的 2 个正则化约束项.对一个具体用户而言,朋友或者专家的正则化项对他的影响可能有所不同,所以对它们的处理也完全不同.尤其对于解决冷启动情况,用户评分数比较少很可能朋友数也不多,所以同一类商品下的专家用户可能对推荐有很大帮助.这些朋友和专家都根据历史评分记录以及评论被划分成不同的兴趣圈后,他们对系统推荐都发挥着重要的作用.我们根据 LBSN 信息提出基于兴趣圈中社会关系挖掘的 POI 推荐算法(circle-based and social connection embedded recommendation in LBSN, CSCR).CSCR 算法使用朋友和专家 2 种社会关系作为规则化项来优化目标函数的学习.在真实数据集上的实验结果表明我们的方法优于多种基于矩阵分解的经典方法.综上所述,本文的贡献如下:

1) 提出了一种基于不同兴趣圈的推荐算法,同时利用 LBSN 上显性和隐性的社会关系来提高推荐系统的性能,尤其对于解决冷启动问题效果显著.

2) 设计了基于朋友关系和基于专家用户的 2 个正则化项,用它们作为矩阵分解目标函数的约束条件,并考虑到它们对不同的用户的差异性影响.

3) 证明了 2 个等价定理.①无论 2 个用户的相似性是以独立方式还是平均值结合到正规化项中的,它们本质都是一样的;②目标函数中所有正则化项可以被视为是不同高斯分布的组合.

4) 在 Yelp 和 Foursquare 数据集上进行了充分的实验,结果表明 CSCR 算法优于现有的基于社会信息的推荐算法.

1 相关工作

在 LBSN 上对有相同兴趣点的推荐, 可以将 POIs 看做普通的项用通常的方法来解决. 基本的低秩矩阵分解技术由 Salakhutdinov 等人^[2]用概率给

出以及被 Koren^[3]巧妙地应用并赢得了 Netflix 奖. 低秩矩阵分解的前提是有一些潜在的因素分别控制用户和商品的属性. 预测的评分由 $R=UV^T$ 给出, U 和 V 分别是用户-特征矩阵和商品-特征矩阵. 模型训练是通过最小化含有 Frobenius 范数规则化项的目标函数的误差平方和.

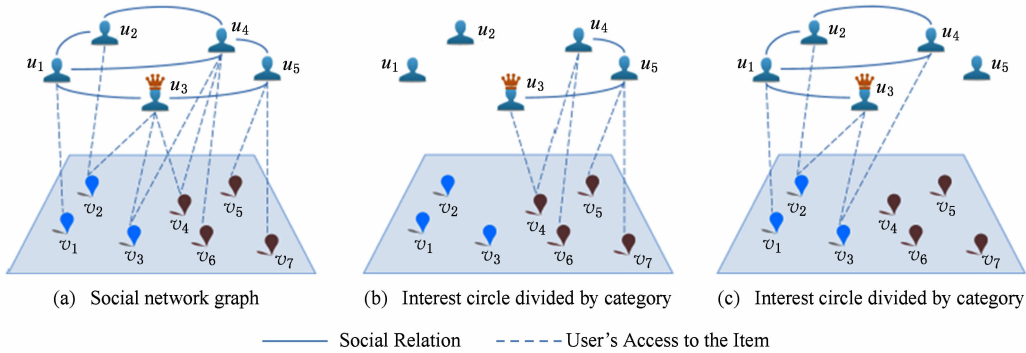


Fig. 1 Description of interest circle

图 1 在 POIs 不同类别下的基于兴趣圈模型描述

现在社交网络信息较易获得, 许多模型通过结合社会关系提高推荐系统准确率, 如 SocialMF^[4], SoReg (social regularization)^[5], hTrust (homophily trust)^[6], ASS (adaptive social similarity)^[7]. 这些模型假设用户的特征向量与他相似的邻居的特征向量相近, 即使用户只访问很少的商品也可以向邻居学习. 同时也有研究者将信任传播机制结合到模型中^[4,8], 或通过设计规则化项^[5-6] 以及通过调查不同的用户相似度度量方法和同质性系数从本质上区分目标函数. 然而这些方法中 2 个用户的相似性度量计算是通过他们访问的共同商品集得到的, 并不是所用的文章都采用这种方法, 如 Yu 等人^[7] 则没有使用用户潜在的特征向量来度量用户间的相似程度. 关于人际关系的影响和社会关系的研究^[9-10] 以及结合协同过滤的推荐算法的研究^[11-12] 也非常普遍.

然而, 以上方法忽略了用户社交关系的多类别性. Yang 等人^[13] 提出了基于兴趣圈的推荐方法, 他们通过推断特定类别的社交圈和设计这个兴趣圈中不同朋友的不同权重, 大量实验验证了这一方法优于 SocialMF 模型. 在此基础上 Zhao 等人^[14] 也基于兴趣圈研究了人际关系和人与人之间兴趣的相似度对模型的影响. 本文的特色是在每个兴趣圈内不仅利用了明确的社会关系(朋友), 而且考虑了隐含的社会关系(专家)来解决冷启动问题. Tang 等人^[15] 利用了局部和全局的社会关系, 同时考虑自己的朋友和所有用户中具有较高声誉的用户, 这篇文章的工作和我们的观点比较相近, 但其整个网络中有影

响力用户的影响是通过在目标函数中设置权重明确对模型的影响, 然后设计一个正则化项集成到模型中来. 而本文是通过对目标函数的优化学习得到这些权重.

2 问题定义

本文研究的问题是以社交网络信息为基础的 POI 推荐. 令 $U = \{u_1, u_2, \dots, u_M\}$, $V = \{v_1, v_2, \dots, v_N\}$ 分别是用户的集合和 POIs 的集合, 其中 M 和 N 分别指用户数和 POIs 数. 用户之间的关系可以用一个矩阵 $F \in \mathbb{R}^{M \times M}$ 表示, 当用户 u_i 是用户 u_j 的朋友时, 矩阵元素 f_{ij} 值为 1, 否则为 0. 所以 F 是一个对称矩阵, 这也使得社交网络中用户间的连接是一个无向图(如 Facebook). 每个 POI 属于 1 个类别 c , 例如饭店或者酒店. 根据每个用户的评级记录, 社交网络可以被划分成不同的兴趣圈, 而在每个兴趣圈中用户之间的社会连接是整个社交网络的子图, 如图 1 所示. 具体定义如下:

定义 1. 兴趣圈(circled social connection). 2 个用户 u_i 和 u_j 是在同一个兴趣圈里 C , 表示成 $(u_i, u_j) \in C$, 如果满足: 1) $f_{ij} = 1$; 2) $V_{u_i}^c \neq \emptyset$ 且 $V_{u_j}^c \neq \emptyset$, 其中 V_u^c 表示用户 u 在类别 c 中评分的一些 POIs.

有一些用户会被网站选中并标记成“专家”, 例如如图 1(a)中的用户 u_3 , 其表示 u_3 在某个领域或兴趣圈里经验很高. 而所有的专家用户组成的集合我们定义为 E .

我们目标是通过用户关系 F 、专家集合 E 和 POI 类别 c 作为辅助信息, 预测每一组 (u, v) 的评分. 尤其当用户的评分很少时即冷启动情况, 仍能给出令人满意的推荐结果.

3 模 型

本文在 LBSN 上基于兴趣圈中社会关系挖掘的 POI 推荐是在传统矩阵分解模型基础上引入朋友和专家关系作为规则化项来优化矩阵分解模型. 新模型也可以视为 CircleCon^[13] 增加了专家正则化项的延伸, 这也是为了更好地解决冷启动问题.

3.1 基本矩阵分解模型

基本的矩阵分解技术是将整个用户评分矩阵 R 分解成 2 个矩阵 $U \in \mathbb{R}^{M \times D}$ 和 $V \in \mathbb{R}^{N \times D}$. 通过对矩阵迭代学习, 最大限度地减少有正则化项的平方误差损失函数, 直到收敛. 其中 U 和 V 分别是用户-特征矩阵和 POI-特征矩阵, 特征维数 $D \ll M, N$. 因为本文采用了兴趣圈的概念, 所以用户评分矩阵 R 根据类别 c 被划分成更小的块, 即 R^c , 而目标函数根据每个类别修改后为:

$$L^c = \frac{1}{2} \|J \otimes (R^c - U^c V^{cT})\|_F^2 + \frac{\lambda}{2} (\|U^c\|_F^2 + \|V^c\|_F^2), \quad (1)$$

其中, $\otimes$ 是 Hadamard 乘积; U^c, V^c 分别是类别 c 下的用户-特征矩阵和 POI-特征矩阵; J 是与矩阵 R^c 大小相同的指示矩阵, 如果用户对 POI 给出了评分, 则 J 矩阵相应位置的值为 1, 否则为 0.

目前为止, 基本矩阵分解模型没有用任何的辅助信息, 对每个用户-商品对 (u_i, v_j) 通过 $\hat{R}_{ij}^c = U_i^c V_j^{cT}$ 进行预测.

3.2 基于朋友关系的正则化项

LBSN 在模型中使用的前提是用户与他的朋友有着相同的兴趣而且他容易受朋友的影响, 这分别被称为同质性^[16] 和社会影响力^[17]. 对于冷启动情况, 当用户评分很少但朋友较多, 那么相同兴趣圈内有相同爱好的朋友与他对某个商品的喜好程度可能比较相近, 所以利用朋友关系设计正则化项具有一定的合理性.

在此之前已经有不少利用朋友关系为用户设计相应的规则化方法^[4-5, 9, 18], 这里采用基于社交网络兴趣圈而设计的朋友关系规则化项^[13], 用信任划分实现朋友关系规则化. 信任划分是 1 种推断多个兴

趣圈下用户 u_i 到 u_j 的信任值的方法. 如果他们有对相同类别下的商品评分, u_i 对 u_j 的信任值与用户 u_j 评分的商品数成正比. 我们定义 S_{ij}^c 为在类别 c 下 u_i 到 u_j 的信任度, 具体如下:

$$S_{ij}^c = \begin{cases} \frac{|V_{u_j}^c|}{\sum_{(u_i, u_j) \in C^c} |V_{u_j}^c|}, & \text{if } (u_i, u_j) \in C^c. \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

通过归一化矩阵 S^c 中每一行记录得到新的矩阵 S^* , 类别 c 下的朋友关系规则化项为:

$$Reg_{\text{friends}} = \|U^c - S^{*c} U^c\|_F^2. \quad (3)$$

3.3 基于专家的正则化项

相对于朋友关系, 专家关系则是一种隐性的社会关系. 专家关系是指人们有时也会受到一个领袖人物(专家)的影响, 容易接受他们的观点并对用户自身行为产生影响. 如果假设用户的兴趣应该与他们兴趣圈中专家用户的较为接近, 那么我们就可以加入专家的影响. 而且, 当用户的朋友数少时专家关系的参考性将更大. 所以为了更好地解决冷启动问题, 我们在朋友关系的基础上又提出了基于专家关系的规则化项.

在传统的社交网络分析中, 往往通过影响力传播的方法研究专家的影响, 这种方法并不要求用户之间有直接关系. 然而 LBSN 可以提供额外的地理范围信息, 为我们研究专家的影响提供了新的方法. 我们知道用户在 LBSN 上的信息是以位置为中心的, 同时专家的评论和评分在每个商家的页面上等级很高, 从而反过来影响普通用户. 因此来自专家的影响有 2 个前提: 1) 专家用户影响普通用户需要他们访问了同一商品; 2) 一个专家不会受到其他专家的影响. 基于上述前提, 我们用矩阵 E^c 设计了 2 种定义专家对普通用户的影响, 定义如下:

定义 2. 专家对普通用户的影响. 在每个兴趣圈内, 当满足: 1) u_i 和 u_j 给同一商品评分了; 2) u_i 不是专家用户而 u_j 是专家用户, E_{ij}^c 值为 1, 否则为 0. 形式化表式如下:

$$E_{ij}^c = \begin{cases} 1, & V_{u_i}^c \cap V_{u_j}^c \neq \emptyset, u_i \notin E, u_j \in E; \\ 0, & \text{otherwise.} \end{cases} \quad (4)$$

因为有时候严格标记好的专家用户这个信息是无法获得的, 所以我们又提出了一种方法来估计用户的专业度. 这个想法是专家用户应该有更多的朋友, 可以吸引更多的关注. 我们使用每个用户对他朋友的平均吸引力来定义他的专业. 这种情况下的用户影响力定义如下:

$$E_{ij}^c = \begin{cases} \frac{|A_u^c|}{|F_u^c + \epsilon|}, & V_{u_i}^c \cap V_{u_j}^c \neq \emptyset; \\ 0, & \text{otherwise.} \end{cases} \quad (5)$$

其中 F_u^c 是类别 c 下用户 u 的朋友集合, A_u^c 是用户 u 吸引的用户集合. 我们的数据中, “吸引力”指的是用户朋友对其的赞美量或者投票数. ϵ 是一个常量, 它是为了避免 F_u^c 为空集使得分母为 0.

同样地, 我们要对 E^c 的每行进行归一化操作,

即 $\sum_{j=1}^{|V_u^c|} E_{ij}^{*c} = 1$, 所以关于专家用户的正则化项可定义如下:

$$Reg_{\text{experts}} = \|U - E^{*c} U^c\|_F^2. \quad (6)$$

3.4 有 2 个规则化项的基于兴趣圈的模型

目前为止, 我们通过式(1)(3)(6)可以得到最终的模型如下:

$$L^c = \frac{1}{2} \|J \otimes (R^c - U^c V^{cT})\|_F^2 + \frac{\lambda}{2} (\|U^c\|_F^2 + \|V^c\|_F^2) + \frac{\alpha}{2} \|U^c - S^{*c} U^c\|_F^2 + \frac{\beta}{2} \|U^c - E^{*c} U^c\|_F^2, \quad (7)$$

其中 α 和 β 分别是 2 个正则化项的参数.

式(7)将朋友和专家的影响融合到了模型中, 但这并不意味着它们两者是同时有效的. 如图 1(b) 中, u_3 是标记的专家用户, 与 u_4 和 u_5 分别有联系. 对 u_4 来说, 只用专家的规则化项对他有影响, 因为他与 u_3 用户访问了同一个地方但是与他又没有直接的联系. 而对 u_5 而言, 不管是专家的规则化项还是朋友的规则化项都对他有影响.

3.5 分析推断

由于模型已经被分解成不同的兴趣圈, 所以每个类别下的训练过程是独立的. 一般来说可以采用批处理梯度下降来最小化目标函数知道收敛. 学习过程如下:

$$\frac{\partial L^c}{\partial U^c} = (U^c V^{cT} - R^c) V^c + \lambda U^c + \alpha (I - S^{*c}) (U^c - S^{*c} U^c) + \beta (I - E^{*c}) (U^c - E^{*c} U^c), \quad (8)$$

$$\frac{\partial L^c}{\partial V^c} = (U^c V^{cT} - R^c)^T U^c + \lambda V^c, \quad (9)$$

其中 I 为单位矩阵. 一旦学习好每个类别 c 中的 U^c 和 V^c 之后, 就可以预测这个类别下用户-商品对 (u_i, v_j) 的评分. 模型同时考虑了 2 种正则化项, 即使对冷启动用户也可以给出有效的评分. 具体如算法 1 所示.

算法 1. 基于兴趣圈中社会关系的推荐模型.

输入: R^c, S^{*c}, E^{*c} ;

输出: U^c, V^c .

① 随机输入 U^c, V^c ;

② 循环从 $step=1$ 到 $MAXSTEP$;

③ $U^c = U^c - \eta \frac{\partial L^c}{\partial U^c}$;

④ $V^c = V^c - \eta \frac{\partial L^c}{\partial V^c}$;

⑤ $\eta = \frac{\eta}{1 + step / MAXSTEP}$;

⑥ 循环结束;

⑦ $\hat{R}^c = U^c V^{cT}$.

3.6 算法复杂度分析

模型的时间复杂度很大程度上取决于计算目标函数 L 以及计算它的梯度. 我们假设 $\bar{r}, \bar{f}, \bar{e}$ 分别是一个用户的平均评分数、朋友数和有影响力的专家数. 式(7)中 L^c 的计算代价是 $O(M \times (\bar{r} + \bar{f} + \bar{e}) \times D)$, 相应梯度的计算代价是 $O(M \times (\bar{r} + \bar{f}^2 + \bar{e}^2) \times D)$. 现实中 r 和 f 相对都很小, 且只有当用户访问的商品与专家用户访问商品一样时专家的评分及评论才会对用户产生影响, 所以 e 的值不会大于 r , 即 $\bar{e} \leq \bar{r}$. 因此模型整体的复杂度与总的用户数呈现出近似线性的关系, 这同时也说明我们提出的算法计算是有效的.

3.7 等价定理

现在已经有很多模型假设用户的兴趣可能与他的邻居比较接近, 并结合了社交网络信息设计不同的规则化项^[4-6,13]. 实际上, 无论 2 个用户的相似性是独立结合到正规化项中的还是以平均值结合到正规化项中的, 它们的本质都是一样的.

定理 1. 第一等价定理. 无论是使用独立的规则化项 $\frac{1}{2} \sum_{i=1}^M \sum_{j=1}^M S_{ij} \|U_i - U_j\|_F^2$ (参见文献[5-7]) 还是使用平均值的规则化项 $\frac{1}{2} \sum_{i=1}^M \|U_i - \sum_{j=1}^M S_{ij} U_j\|_F^2$ (参见文献[4,13]), 它们本质上是一样的.

证明. 像文献[6]中式(7)提到的一样, $L = D - S$ 是一个拉普拉斯矩阵, 其中,

$$D_{ii} = \sum_{j=1}^M S_{ij},$$

$$\frac{1}{2} \sum_{i=1}^M \sum_{j=1}^M S_{ij} \|U_i - U_j\|_F^2 = U^T (D - S) U = \text{tr}(U^T L U).$$

我们需要做的就是以平均为基础的正则化项与独立的有相同的表示形式.

$$\begin{aligned} & \frac{1}{2} \sum_{i=1}^M \left\| \mathbf{U}_i - \sum_{j=1}^M S_{ij} \mathbf{U}_j \right\|_F^2 = \\ & \frac{1}{2} \sum_{i=1}^M \sum_{i=1}^M \left\| \mathbf{U}_i \right\|_F^2 - \sum_{i=1}^M \sum_{j=1}^M S_{ij} \mathbf{U}_i \mathbf{U}_j^T + \\ & \frac{1}{2} \sum_{i=1}^M \sum_{j=1}^M S_{ij} \left\| \mathbf{U}_j \right\|_F^2 = \\ & \frac{1}{2} \sum_{i=1}^M \sum_{i=1}^M (1 + S_{ij}) \left\| \mathbf{U}_i \right\|_F^2 - \sum_{i=1}^M \sum_{j=1}^M S_{ij} \mathbf{U}_i \mathbf{U}_j^T = \\ & \sum_{k=1}^D \sum_{i=1}^M U_{ik} \left[\sum_{j=1}^M \left(\frac{1 + S_{ij}^2}{2} \right) U_{ik} - \sum_{j=1}^M S_{ij} U_{jk} \right] = \\ & \mathbf{U}^T (\mathbf{D}^* - \mathbf{S}) \mathbf{U} = \\ & \text{tr}(\mathbf{U}^T \mathbf{L}^* \mathbf{U}), \end{aligned}$$

其中 $\mathbf{D}_i^* = \frac{1 + S_{ij}^2}{2}$, $\mathbf{L}^*$ 也是拉普拉斯矩阵。证毕。

定理 2. 第二等价定理. 目标函数中的所有正则化项可以被视为是由各种估计用户兴趣标准差的高斯分布的组合。

证明. 在基本的矩阵分解模型中, 我们一般假设 $\mathbf{U}$ 和 $\mathbf{V}$ 矩阵都是标准差为 0 的高斯分布, 因此使用正则化项来避免过拟合问题, 所以我们有 $\mathbf{U}_{\text{basic}} \sim N(0, \sigma_1^2)$, 这里对 $\mathbf{U}$ 的讨论也同样适用于 $\mathbf{V}$ 矩阵. 已有的使用朋友关系规则化的模型使用邻居的平均兴趣作为均值, 如: μ_1 如果是用户兴趣的标准差, 那么朋友关系的规则化项可以表示为 $\mathbf{U}_{\text{friends}} \sim N(\mu_1, \sigma_2^2)$. 我们的模型中来自专家用户的影响也是一种对用户兴趣的估计, 同样地, 如果它的均值是 μ_2 , 则有 $\mathbf{U}_{\text{experts}} \sim N(\mu_2, \sigma_3^2)$.

在 i. i. d. (independent and identically distributed) 假设下, 这 3 个分布的线性组合是:

$$\begin{aligned} & \lambda \text{Reg}_{\text{basic}} + \alpha \text{Reg}_{\text{friends}} + \beta \text{Reg}_{\text{experts}} \leftarrow \\ & \lambda \mathbf{U}_{\text{basic}} + \alpha \mathbf{U}_{\text{friends}} + \beta \mathbf{U}_{\text{experts}} \sim \\ & N(\lambda 0 + \alpha \mu_1 + \beta \mu_2, \lambda^2 \sigma_1^2 + \alpha^2 \sigma_2^2 + \beta^2 \sigma_3^2), \end{aligned}$$

因此在我们的模型中, 用户的兴趣服从均值为 $\alpha \mu_1 + \beta \mu_2$ 的高斯分布。

4 实验结果

本节将本文提出的 CSCR 模型与已经存在的一些模型比较, 有 BaseMF^[2], PRM^[18], CircleCon3^[13], SocialMF^[4], SR^[19]. 所有的实验将在 Yelp 和 Foursquare 数据集上进行。

4.1 数据集

我们采用第 5 届 Yelp 挑战赛的数据集^①. 根据

商品类别可以划分成 10 个类: Activelife, Arts, Beauty, Education, Hotels, Nightlife, Pets, Restaurants, Services, Shopping. 数据集中所有用户数为 366 715, 每个用户或商品还包含很多信息, 我们只提取一些对我们有用的数据. 表 1 给出了 10 个类别下的用户数和商品数的统计量, 其中 r^c 指的是用户在类别 c 下的平均评分。

Table 1 Statistics of Various Categories on Yelp Dataset
表 1 Yelp 数据集上各个类别记录统计

Category	User Count	Item Count	Rating Count	Sparsity	r^c
Activelife	16 004	2 379	29 800	7.83E-4	3.968
Arts	30 212	1 900	79 087	1.38E-3	3.748
Beauty	21 622	4 521	39 835	4.08E-4	4.033
Education	2 523	480	3 067	2.53E-3	4.056
Hotels	25 676	1 987	58 541	1.15E-3	3.556
Nightlife	50 258	4 253	161 165	7.54E-4	3.668
Pets	6 044	1 103	8 985	1.35E-3	4.148
Restaurants	84 541	29 169	874 428	3.55E-4	3.735
Services	38 836	7 886	89 441	2.92E-4	3.722
Shopping	24 344	8 687	70 716	3.34E-4	3.744

另外, 实验还在自己爬取的 Foursquare 数据集的 4 个类别上做了对比. 这 4 个类别分别是 Hotels, Sports, Traffic, Arts. 其中, Hotels 类别下共有 4 728 个用户和 6 078 种商品, 用户评分的商品数共有 10 991 次; Sports 类别下含 3 443 个用户和 3 037 种商品, 用户评分的商品数共有 5 872 次; Traffic 类别下含 2 590 个用户和 2 791 种商品, 用户评分的商品数共有 4 894 次; Arts 类别下含 1 659 个用户和 1 291 种商品, 用户评分的商品数共 2 812 次。

4.2 评价指标

Yelp 数据集的每个类别下, 我们使用 5-折交叉验证训练模型. 每次使用 80% 的数据作为训练集, 剩下的 20% 作为测试集. 实验的度量标准采用的是平方根误差 (root mean squared error, RMSE) 和平均绝对误差 (mean absolute error, MAE), 这也是推荐系统中常用的 2 种度量方法. RMSE 的定义如下:

$$RMSE = \sqrt{\frac{\sum_{(u,i) \in \mathbf{R}_{\text{test}}} (R_{u,i} - \hat{R}_{u,i})^2}{|\mathbf{R}_{\text{test}}|}}.$$

其中, $\mathbf{R}_{\text{test}}$ 是用户-商品对 (u, i) 在测试集上的矩阵, $R_{u,i}$ 是实际用户 u 对商品 i 的评分, $\hat{R}_{u,i}$ 则是相对应

① <http://www.yelp.com/dataset/challenge>

的预测值. MAE 的定义表示如下:

$$MAE = \frac{\sum_{(u,i) \in R_{test}} |R_{u,i} - \hat{R}_{u,i}|}{|R_{test}|}.$$

4.3 对比实验

- 本节给出了我们的模型与已有模型的对比:
- 1) BaseMF^[2]. 基本的矩阵分解方法通过将用户-商品矩阵分解成用户-矩阵和商品矩阵, 该方法只用到了用户的评分矩阵, 没用到任何的社交网络信息.
 - 2) PRM^[18]. 这个模型考虑了用户的个性和社交网络信息, 考虑了用户和用户之间以及用户和商品之间的关系.
 - 3) CircleCon3^[13]. 基于兴趣圈的推荐模型将朋友圈的新特征与评分结合起来推断值得信任的兴趣圈.
 - 4) SocialMF^[4]. 基于信任传播的矩阵分解方法假设社交网络连接了所有的用户. 在模型中加入了信任信息来提高推荐系统的准确率.
 - 5) SR^[19]. 基于社交网络的推荐系统不仅使用了用户对商品的评分信息而且增加了社交网络中隐含的用户与商品间的关系, 而且包括了相同的和不相同的关系.
 - 6) CSCR. 这是我们提出的基于 2 种正则化项和兴趣圈的模型. 与 CircleCon3 相比增加了专家用户对推荐的影响.
- 设置矩阵分解维度 $d=10$, 正则化项参数 $\lambda=0.01, \alpha=\beta=1$. 在表 2 和表 3 中, 我们展示了不同模型在 Yelp 数据集上 2 种评价指标下的结果.

Table 2 MAE Metric on Ten Categories
表 2 10 个类别上的 MAE 度量结果

Category	BaseMF	PRM	CircleCon3	SocialMF	SR	CSCR
Activelife	1.150	0.831	0.743	0.853	0.981	0.705
Arts	1.061	0.928	0.829	0.774	0.830	0.742
Beauty	1.142	1.004	0.939	0.996	1.019	0.904
Education	1.011	0.963	0.768	0.989	0.935	0.783
Hotels	1.061	0.881	0.866	0.834	0.789	0.754
Nightlife	0.948	0.742	0.776	0.806	0.798	0.719
Pets	1.191	1.063	0.985	1.078	1.055	0.976
Restaurants	0.947	0.850	0.844	0.866	0.856	0.731
Services	1.235	0.872	0.837	0.877	0.849	0.775
Shopping	1.333	0.856	0.843	0.901	0.857	0.770
Average	1.108	0.899	0.843	0.896	0.897	0.786

Table 3 RMSE Metric on Ten Categories
表 3 10 个类别上的 RMSE 度量结果

Category	BaseMF	PRM	CircleCon3	SocialMF	SR	CSCR
Activelife	1.947	1.116	0.984	1.118	1.087	0.831
Arts	1.767	1.398	1.149	0.960	1.089	0.897
Beauty	2.114	1.630	1.543	1.575	1.613	1.412
Education	1.341	1.131	1.077	1.179	1.184	0.814
Hotels	1.834	1.236	1.284	1.120	1.015	0.928
Nightlife	1.381	1.088	1.015	1.129	1.101	0.892
Pets	2.251	1.790	1.779	1.653	1.795	1.627
Restaurants	1.323	1.185	1.116	1.154	1.123	0.886
Services	2.449	1.408	1.185	1.186	1.202	0.990
Shopping	2.640	1.176	1.209	1.247	1.257	0.969
Average	1.905	1.316	1.234	1.232	1.247	1.025

从表 2 可以看出在 MAE 评价指标下, CircleCon3 的准确率在前 5 个对比实验中是最高的. 但是文献 [18] 中 PRM 模型的效果优于 CircleCon3, 这是因为: 1) 我们发现尽管 PRM 用的也是 Yelp 数据集, 但是我们的数据量明显多于他们使用的数据量; 2) 从 PRM 数据集可以看出他们使用的用户量更少, 而商品数更多, 而且每个用户平均对很多商品有评分. PRM 方法更关注的是用户与用户之间以及用户与商品之间的关系, 所以在用户可以评分很多商品时模型表现效果会很好. 相应的 CircleCon3 考虑更多的则是用户信任的朋友, 所以当用户量很多时, 模型效果可能会较好. 结合我们的数据特点, 就可以理解为什么 CircleCon3 的效果好于 PRM 了. 但是我们也可以看出在某些类别, PRM 的性能也可能比 CircleCon3 要好 (如 Nightlife).

在表 3 的 $RMSE$ 度量标准下, CircleCon3, SocialMF 模型效果相对较好, 而 SocialMF 的整体效果可能稍微比 CircleCon3 好一些. 在 Arts, Hotels, Pets 这些类别下, SocialMF 优于 CircleCon3. 这 3 个类别的数据特点是数据稀疏度相对较低. 因为 SocialMF 使用的是所有可用的社会关系, 所以它可能会表现较好. 但是在其他的类别下, CircleCon3 的效果则更好.

我们的模型在 2 种度量标准下分别是最优的. 基于兴趣圈的方法使用信任的朋友及专家信息来辅助推荐. 通过实验结果, 可以看出加入专家群体的推荐效果是显著的. 因为当数据集中有更多的用户而每个用户平均评分的商品数又比较少时, 在相同圈子的专家可能会对推荐有很大的影响. 在表 2 中 MAE 的度量标准下我们方法的效果也并不总是最好的,

在 Education 类别下 CircleCon3 的准确率更高. 这可能是因为这个类别中的用户数和商品数都比较少, 所以受专家的影响并不明显. 但整体来说 CSCR 方法比前面提出的方法效果都好.

图 2(a) 是 baseMF, PRM, CircleCon3, CSCR 这 4 个算法在 Foursquare 数据集的 4 个类别上 MAE 度量结果, 图 2(b) 则是 RMSE 度量结果. 从图 2(b) 中可以看出除了在 Sports 类别的 RMSE 度量上 CircleCon3 效果最好外, CSCR 算法总体来说效果是最好的. 而 CircleCon3 之所以在 Sports 类别下 RMSE 值最小可能是因为该类别下用户的偏好更倾向于他们的朋友.

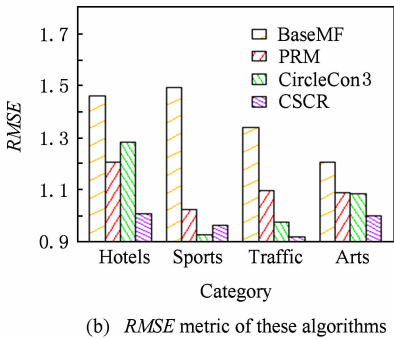
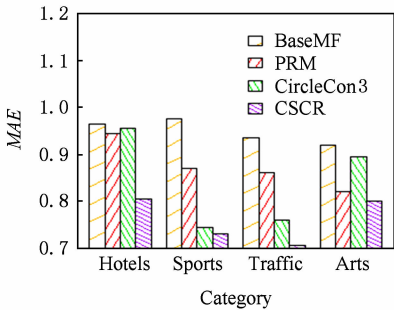
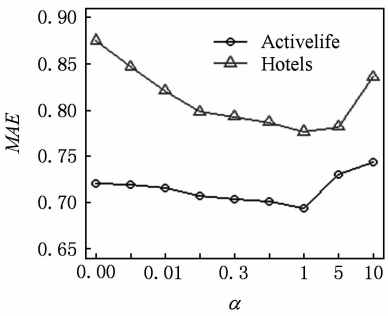


Fig. 2 Algorithm results on 4 categories of Foursquare
图 2 算法在 Foursquare 数据集 4 个类别上的结果

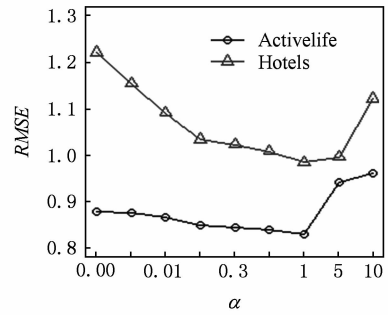
下面我们对模型中参数 α, β, λ 分别在 MAE 和 RMSE 度量下不同取值的模型效果. 实验选取了 Activelife 和 Hotels 两个类别的数据. 其中 α 取值为 $\{0, 0.001, 0.01, 0.1, 0.3, 0.5, 1, 5, 10\}$, β 取值为 $\{0, 0.001, 0.01, 0.1, 0.3, 0.5, 1, 5, 10\}$, λ 取值为 $\{0, 0.001, 0.01, 0.05, 0.1, 0.3, 0.5, 1, 10\}$.

图 3 展示了 α 变化对模型的影响. 可以看出当 $\alpha=1$ 时在 2 种度量标准下的准确率都是最高的. 当 $\alpha=0$ 时, 模型不考虑朋友的影响, 可以看出这时模型效果也不是最优的. 随着 α 值的增大, 可以看出加入朋友信息可以提高推荐性能, 直到达到最优点, 然后再增加 α 值, 模型效果将会下降, 甚至在一个类别

上会比不考虑朋友的影响还差. 这是因为模型过拟合了 U 矩阵而对 V 矩阵又估计不足.



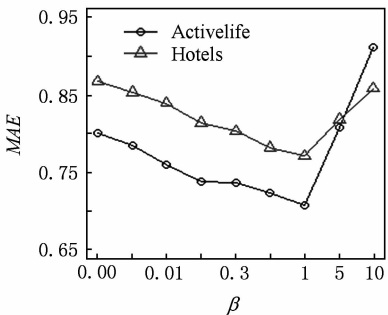
(a) MAE metric of CSCR algorithm



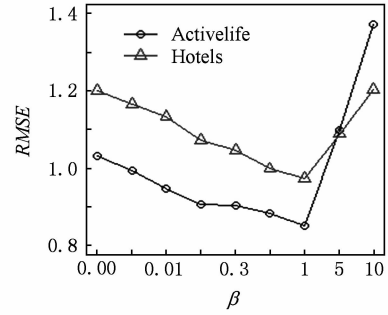
(b) RMSE metric of CSCR algorithm

Fig. 3 The results of CSCR algorithm on two categories with the change of α

图 3 CSCR 算法在 2 个类别上随 α 变化的结果



(a) MAE metric of CSCR algorithm



(b) RMSE metric of CSCR algorithm

Fig. 4 The results of CSCR algorithm on two categories with the change of β

图 4 CSCR 算法在 2 个类别上随 β 变化的结果

图 4 给出了 β 值对模型的影响. 当 $\beta=0$ 时, 模型不考虑专家群体对推荐的影响, 这也就是 CircleCon3 模型, 从图 4 中可以发现这时的模型效果并不是最优的; 随着 β 值的增加模型准确率越来越高直到最优, 这也说明了专家群体对推荐是有帮助的; 但是再增加 β 值, 模型准确率又会下降, 所以更大的 β 值并不一定会使推荐效果更好.

图 5 则给出了 λ 值对模型的影响. 从 2 个类别的数据分别在 MAE 和 RMSE 度量标准下的表现来说, 当 λ 值在 $0 \sim 0.01$ 过程中, 模型效果是越来越好的; 在 $0.01 \sim 1$ 的过程中, 模型对 λ 值变化并不敏感; 但是当其特别大 (如 $\lambda=10$), 模型的性能会迅速下降. 所以对 U, V 的规则化的系数不应太大.

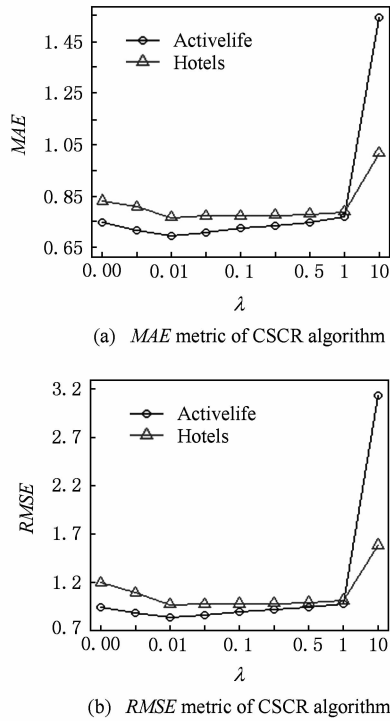


Fig. 5 The results of CSCR algorithm on two categories with the change of λ

图 5 CSCR 算法在 2 个类别上随 λ 变化的结果

4.4 用户冷启动问题

冷启动实验在 Yelp 数据集 Nightlife, Restaurants, Shopping 三个类别上进行, 对比用户评分的商品数变化时不同算法的效果. 最终给出的结果则是 3 个类别下的平均值. 测试集根据用户的评分分数被分成了 5 组, 分别是: $0 \sim 10$, $11 \sim 20$, $21 \sim 40$, $41 \sim 80$, >80 .

图 6(a)和图 6(b)分别表现了在不同的用户评分量下的这些算法效果. 图 6 横轴上的数字表示用

户给多少商品评分了, 如“ $0 \sim 10$ ”表示用户评分量在 $0 \sim 10$ 之间, “ $11 \sim 20$ ”表示用户评分量在 $11 \sim 20$ 之间, “ $80 \sim$ ”表示用户评分数超过 80, 以此类推. 图 6(a)的纵轴表示 MAE 值, 图 6(b)的纵轴表示 RMSE 值. 从这 2 幅图可以看出, 当用户评分的商品增加时, 算法的准确率也会提高; 当用户评分的商品很少时, BaseMF 算法效果很差, 但是随着评分数目的增加, 算法性能会变得越来越好. 不管怎样, 从图 6(a)和图 6(b)都可以看出, 我们提出的方法在 Yelp 数据集上 MAE 和 RMSE 的度量标准下值都最低. 这是因为基于兴趣圈的推荐不仅考虑了信任的朋友关系而且用到了专家群体的意见, 这在推荐过程中确实是有用的.

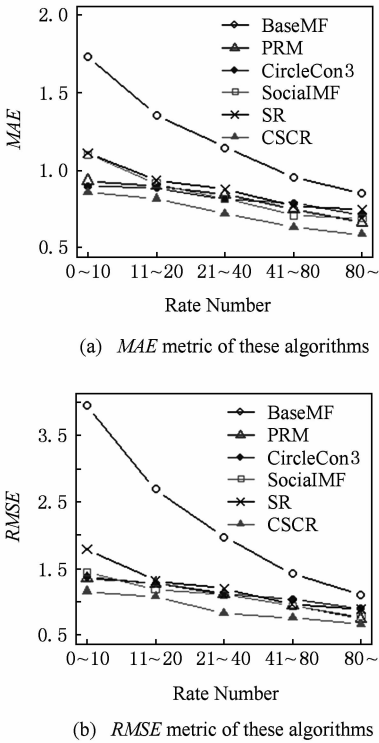


Fig. 6 The impact on the model with different user rating counts

图 6 用户评分的商品数对模型的影响

我们的方法在 5 组冷启动实验组效果始终是最好的. 因为我们的数据集中用户数很多, 商品数相对较少, 而大部分用户评分的商品数并不是很多, 在这种情况下我们提出的基于用户信任的朋友关系和专家的关系在推荐中就有重要的作用. 从图 5 中也可以得知, 当用户评分很多商品时, 算法的性能也会变得越来越好, 第 5 组实验的结果彼此就非常接近了. 所以用户评分更多的商品对模型是有帮助的, 但是也可以发现这个提升度是逐渐降低的.

5 结束语

本文的目标是解决冷启动问题,文中提出了一种在 LBSN 上基于兴趣圈中社会关系的模型来提高推荐效果。具体来说,社会关系有朋友关系和专家用户,由此设计的 2 个规则化项作为矩阵分解目标函数的约束项。在 CircleCon3 模型的基础上,我们不仅考虑了显性的用户信息而且考虑了隐含的关系来提高推荐系统的冷启动问题。朋友或者专家都有自己的评分记录,分别根据不同的兴趣圈做出划分,每个圈子中的所有商品都属于同一类别。在 Yelp 和 Foursquare 两个数据集上的实验表明本文提出的使用混合社交网络信息的方法整体优于已有方法。

参 考 文 献

- [1] Bao Jie, Zheng Yu, Wilkie D, et al. Recommendations in location based social networks: A survey [J]. *GeoInformatica*, 2015, 19(3): 525-565
- [2] Salakhutdinov R, Mnih A. Probabilistic matrix factorization [C] //Proc of the Int Conf on Machine Learning. New York: ACM, 2012: 880-887
- [3] Koren Y. Factorization meets the neighborhood: A multifaceted collaborative filtering model [C] //Proc of the 14th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2008: 426-434
- [4] Jamali M, Ester M. A matrix factorization technique with trust propagation for recommendation in social networks [C] //Proc of the 4th ACM Conf on Recommender Systems. New York: ACM, 2010: 135-142
- [5] Ma Hao, Zhou Dengyong, Liu Chao, et al. Recommender systems with social regularization [C] //Proc of the 4th ACM Int Conf on Web Search and Data Mining. New York: ACM, 2011: 287-296
- [6] Tang Jiliang, Gao Huiji, Hu Xia, et al. Exploiting homophily effect for trust prediction [C] //Proc of the 6th ACM Int Conf on Web Search and Data Mining. New York: ACM, 2013: 53-62
- [7] Yu Le, Pan Rong, Li Zhangfeng. Adaptive social similarities for recommender systems [C] //Proc of the 5th ACM Conf on Recommender Systems. New York: ACM, 2011: 257-260
- [8] Guo Lei, Ma Jun, Chen Zhumin. Trust strength aware social recommendation method [J]. *Journal of Computer Research and Development*, 2015, 52(9): 1805-1813 (in Chinese)
(郭磊, 马军, 陈竹敏. 一种信任关系强度敏感的社会化推荐算法[J]. *计算机研究与发展*, 2015, 52(9): 1805-1813)
- [9] Huang Junming, Cheng Xueqi, Guo Jiafeng, et al. Social recommendation with interpersonal influence [C] //Proc of the 19th European Conf on Artificial Intelligence. Amsterdam, Netherlands: IOS Press, 2010: 601-606
- [10] Jiang Meng, Cui Peng, Wang Fei, et al. Scalable recommendation with social contextual information [J]. *IEEE Trans on Knowledge and Data Engineering*, 2014, 26(11): 2789-2802
- [11] Yang Yongxiang. Recommendation algorithms based on combination of matrix factorization and random walk [D]. Beijing: Beijing Jiaotong University, 2014 (in Chinese)
(杨永向. 基于矩阵分解和随机游走相结合的推荐算法[D]. 北京: 北京交通大学, 2014)
- [12] Wei Suyun, Xiao Jingjing, Ye Ning. Collaborative filtering algorithm based on co-clustering smoothing [J]. *Journal of Computer Research and Development*, 2013, 50(1): 163-169 (in Chinese)
(韦素云, 肖静静, 业宁. 基于联合聚类平滑的协同过滤算法[J]. *计算机研究与发展*, 2013, 50(1): 163-169)
- [13] Yang Xiwang, Steck H, Liu Yong. Circle-based recommendation in online social networks [C] //Proc of the 18th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2012: 1267-1275
- [14] Zhao Guoshuai, Qian Xueming, Feng He. Personalized recommendation by exploring social users' behaviors [C] //Proc of the Multi Media Modeling. Berlin: Springer, 2014: 181-191
- [15] Tang Jiliang, Hu Xia, Gao Huiji, et al. Exploiting local and global social context for recommendation [C] //Proc of the Int Joint Conf on Artificial Intelligence. San Francisco, CA: Morgan Kaufmann, 2013: 264-269
- [16] McPherson M, Smith-Lovin L, Cook J M. Birds of a feather: Homophily in social networks [J]. *Annual Review of Sociology*, 2001, 27(1): 415-444
- [17] Sun Jimeng, Tang Jie. A survey of models and algorithms for social influence analysis [M] //Social Network Data Analytics. Berlin: Springer, 2011: 177-214
- [18] Feng He, Qian Xueming. Recommendation via user's personality and social contextual [C] //Proc of the 22nd ACM Int Conf on Information & Knowledge Management. New York: ACM, 2013: 1521-1524
- [19] Ma Hao. An experimental study on implicit social recommendation [C] //Proc of the 36th Int ACM SIGIR Conf on Research and Development in Information Retrieval. New York: ACM, 2013: 73-82



Li Xin, born in 1989. PhD. Received his PhD degree from the University of Science and Technology of China (USTC) in 2015. His main research interests include location based services, recommender system, machine learning.



Liu Guiquan, born in 1970. PhD. Associate professor in the University of Science and Technology of China (USTC). His main research interests include data and knowledge management systems, data and text mining, multiagent systems.



Wu Zongda, born in 1983. PhD. Associate professor in Wenzhou University. His research interests include information retrieval and personal privacy (zongda1983@163.com).



Li Lin, born in 1977. PhD. Professor in Wuhan University of Technology. Her current research interests include text mining, machine learning, information retrieval and recommender system (cathylilin@whut.edu.cn).



Ding Junmei, born in 1991. Master candidate in computer science at the University of Science and Technology of China (USTC). Her main research interests include machine learning and data mining.

2015 年《计算机研究与发展》高被引论文 TOP10

排名	论文信息
1	王继业, 孟坤, 曹军威, 程志华, 高灵超, 林闯. 能源互联网信息技术研究综述[J]. 计算机研究与发展, 2015, 52(5): 1109-1126 Wang Jiye, Meng Kun, Cao Junwei, Cheng Zhihua, Gao Lingchao, Lin Chuang. Information Technology for Energy Internet: A Survey [J]. Journal of Computer Research and Development, 2015, 52(5): 1109-1126 张焕龙, 胡士强, 杨国胜. 基于外观模型学习的视频目标跟踪方法综述[J]. 计算机研究与发展, 2015, 52(1): 177-190
2	Zhang Huanlong, Hu Shiqiang, Yang Guosheng. Video Object Tracking Based on Appearance Models Learning [J]. Journal of Computer Research and Development, 2015, 52(1): 177-190 王元卓, 贾岩涛, 刘大伟, 靳小龙, 程学旗. 基于开放网络知识的信息检索与数据挖掘[J]. 计算机研究与发展, 2015, 52(2): 456-474
3	Wang Yuanzhuo, Jia Yantao, Liu Dawei, Jin Xiaolong, Cheng Xueqi. Open Web Knowledge Aided Information Search and Data Mining [J]. Journal of Computer Research and Development, 2015, 52(2): 456-474 段洁, 胡清华, 张灵均, 钱宇华, 李德玉. 基于邻域粗糙集的多标记分类特征选择算法[J]. 计算机研究与发展, 2015, 52(1): 56-65
4	Duan Jie, Hu Qinghua, Zhang Lingjun, Qian Yuhua, Li Deyu. Feature Selection for Multi-Label Classification Based on Neighborhood Rough Sets [J]. Journal of Computer Research and Development, 2015, 52(1): 56-65 唐成华, 刘鹏程, 汤申生, 谢逸. 基于特征选择的模糊聚类异常入侵行为检测[J]. 计算机研究与发展, 2015, 52(3): 718-728
5	Tang Chenghua, Liu Pengcheng, Tang Shensheng, Xie Yi. Anomaly Intrusion Behavior Detection Based on Fuzzy Clustering and Features Selection [J]. Journal of Computer Research and Development, 2015, 52(3): 718-728 辛宇, 杨静, 汤楚衡, 葛斯乔. 基于局部语义聚类的语义重叠社区发现算法[J]. 计算机研究与发展, 2015, 52(7): 1510-1521
6	Xin Yu, Yang Jing, Tang Chuheng, Ge Siqiao. An Overlapping Semantic Community Detection Algorithm Based on Local Semantic Cluster [J]. Journal of Computer Research and Development, 2015, 52(7): 1510-1521 吴章玲, 金培权, 岳丽华, 孟小峰. 基于 PCM 的大数据存储与管理研究综述[J]. 计算机研究与发展, 2015, 52(2): 343-361
7	Wu Zhangling, Jin Peiquan, Yue Lihua, Meng Xiaofeng. A Survey on PCM-Based Big Data Storage and Management [J]. Journal of Computer Research and Development, 2015, 52(2): 343-361 秦兵, 刘安安, 刘挺. 无指导的中文开放式实体关系抽取[J]. 计算机研究与发展, 2015, 52(5): 1029-1035
8	Qin Bing, Liu An'an, Liu Ting. Unsupervised Chinese Open Entity Relation Extraction [J]. Journal of Computer Research and Development, 2015, 52(5): 1029-1035 陈世敏. 大数据分析 & 高速数据更新[J]. 计算机研究与发展, 2015, 52(2): 333-342
9	Chen Shimin. Big Data Analysis and Data Velocity [J]. Journal of Computer Research and Development, 2015, 52(2): 333-342 马思伟. AVS 视频编码标准技术回顾及最新进展[J]. 计算机研究与发展, 2015, 52(1): 27-37
10	Ma Siwei. History and Recent Developments of AVS Video Coding Standards [J]. Journal of Computer Research and Development, 2015, 52(1): 27-37