https://doi.org/10.1051/jnwpu/20193761294

基于用户兴趣领域中可信圈挖掘的推荐模型

白昀, 蔡皖东

(西北工业大学 计算机学院, 陕西 西安 710072)

摘 要:基于信任的推荐系统通过系统评分数据和用户信任关系为用户推荐所需资源。现有相关工作中在考虑信任关系时,通常考虑的是一种泛化的信任关系,尚未充分挖掘信任关系信息与特定兴趣领域之间的关系,对推荐的准确性和可靠性会产生一定的劣化影响。考虑到以上问题,提出基于用户兴趣领域的信任圈模型,针对不同兴趣领域分层挖掘用户间潜在的隐形信任关系;并充分融合显性信任关系为用户资源进行综合评分。该模型不仅考虑信任信息与领域的匹配关系,而且能够挖掘在具体领域下用户间的隐性信任关系,能够进一步提高评分预测的精确度和覆盖率。通过在 Epinions 数据集上的实验,证明了所提出的基于用户兴趣领域可信圈挖掘的推荐模型与基于泛化信任关系的传统推荐算法相比可以有效提高推荐评分预测的准确度和覆盖率。

关 键 词:信任关系;兴趣领域;推荐算法;可信圈;社会网络

中图分类号:TP393 **文献标志码:**A **文章编号:**1000-2758(2019)06-1294-08

推荐系统是解决信息过载问题的有效方法[1]。基于社会网络的推荐算法通过将社会网络中用户间的信任关系引入推荐算法从而提高推荐质量[2-5]。目前学者大都基于泛化信任关系(over-generalized trust relationship)(即用户 A 信任用户 B 表示用户 A 在各个领域均信任用户 B)进行推荐研究[6-10],并未考虑用户在不同领域上会信任不同的用户集这一现象,以及忽略了评分数据多类别、多领域混合这一事实,在预测评分时会存在误用其他领域中的信任信息和评分数据的现象,影响评分预测的准确性。为此,一些学者提出将用户及其社交关系按圈子划分,基于圈子进行推荐以期提高推荐效果[11-16]。Mcauley 等[11]提出基于网络结构和用户个人资料信息的社交圈发现模型,考察个体属性对社交圈形成的影响。Burton 等[12]提出 SCFinder 社交圈发现算法,通过向社交圈添加相邻节点或定期删除社交圈中的节点形成社交圈,该算法关注隐性社会关系对社交圈形成的影响。Zhong 等[13]提出基于用户移动数据的社交圈推理模型,Lan 等[14]提出基于多视图网络结构的可自动检测社交圈,以上几种模型都忽略了用户社交关系的领域属性对社交圈形成的影响。Yang 等[15]将用户和社会关系根据商品类别划分“圈”,提出基于“圈”的推荐算法,但该模型中并未考虑信任关系的领域区分性。Yin 等[16]通过发现用户的兴趣点结合 PMF 模型提出基于兴趣圈的 Top-N 推荐,但其侧重研究用户兴趣点的发现,并未研究社会关系对推荐系统效果的影响。以上“社交圈”形成算法未将用户历史信息和社会关系按圈子进行划分,都存在误用其他“圈”中信息的现象,推荐系统的评分准确度仍可进一步提高。因此,挖掘用户在具体兴趣领域中的信任关系对于提高该情境下的推荐准确性与覆盖率将起到关键性作用。本文重点关注用户在具体兴趣领域中的信任关系构建问题,提出基于用户兴趣领域可信圈挖掘的推荐算法。

1 基于兴趣领域可信圈挖掘的推荐模型

问题 给定网络信任图 $G=(V, E)$,以及用户对各物品的评分 R、其他用户对物品评分 R 的反馈评分 R_{feedback}(feedback rating)和物品的领域分类 C,

收稿日期:2019-02-27 **基金项目:**榆林市科技计划项目(2016-24-4)与陕西省教育厅专项科学研究计划项目(19JK0526)资助
作者简介:白昀(1985—),女,西北工业大学博士研究生,主要从事社交网络的信任模型、概率图研究。

如何建立面向目标用户兴趣领域的 M 层可信圈,并依据可信圈预测目标物品评分。

在产品评论网站中,各物品的组织形式往往使用类别的方式,同类别的物品具有相似的特征。在本文中将类别视为领域。

本文的目标是通过用户的评论记录和用户间的显性信任关系,首先判断出用户的兴趣领域,寻找出在兴趣领域中目标用户的 M 层可信圈,挖掘用户间的潜在信任关系,综合预测目标用户感兴趣的物品评分,以期提高评分预测的准确度及覆盖率。基于兴趣领域的可信圈模型(trust circle model based on interest domain,简称 ID-TrustCircle)是以目标用户为中心,寻找其兴趣领域中的可信用户集,即可信圈。即首层可信圈 $\boldsymbol{T}^{(1)}_{\mathrm{circl}_i}$ 为用户本身,再以 $\boldsymbol{T}^{(1)}_{\mathrm{circl}_i}$ 为中心,寻找可信用户集,迭代 M 次,每一次迭代搜索到的可信用户集组成目标用户兴趣领域的 M 层可信圈。可信圈的每层可信用户集由二部分组成:① 显性领域信任用户,由兴趣领域内的显性信任关系挖掘而得;② 隐性领域可信用户,通过在兴趣领域内挖掘用户间的隐性信任关系推理而得。表 1 列出来本文使用的所有符号。

表 1　符号集

符号	描述 的用户集合
$N_{u_i}(C_k)$	表示用户 u_i 在领域 C_k 中撰写的评论的数量
$N(C_k)$	表示在领域 C_k 中发布的评论总数目
F_i	表示用户 u_i 的擅长领域,本文假定每个用户仅存在一个擅长领域
I_i	表示用户 u_i 的兴趣领域
$C_k = F_i$	表示用户 u_i 的擅长领域是 C_k
$B_i(C_k)$	表示用户 u_i 在领域 C_k 上的活跃度
$A_i(C_k)$	表示用户 u_i 在领域 C_k 上发表评论的认可度
$R^{(i)}_{\mathrm{feedback}}$	表示物品 i 所对应的评论的反馈评分集
$u_i \xrightarrow{T} u_j$	表示用户 u_i 信任用户 u_j
$u_i \xrightarrow{C_k} u_j$	表示用户 u_i 在领域 C_k 中信任用户 u_j
$\boldsymbol{T}^{(v)}_{\mathrm{circl}_i}(C_k)$	表示目标用户 u_i 在领域 C_k 中前 v 层(包括第 v 层)的可信用户群
$\boldsymbol{T}^{(v)}_i$	表示目标用户 u_i 第 v 层的显性领域信任用户集
$\boldsymbol{S}^{(v)}_i$	表示目标用户 u_i 第 v 层的领域相似用户集
$\boldsymbol{R}^{(v)}_i$	表示目标用户 u_i 第 v 层的隐性领域可信用户集

1.1　基于泛化信任关系的显性领域信任用户挖掘

通常来说人们往往习惯于在他人所擅长的领域信任对方,譬如人们遇到法律问题往往信任的是律师,健康疾病问题更愿意信任医生,教育问题更愿意信任教育从业者。因此本文假设:在未指明信任领域时,用户 u_i 信任用户 u_j,表示用户 u_i 在用户 u_j 擅长的领域中信任用户 u_j,即 $u_i \xrightarrow{F_j} u_j$。根据此条假设,将用户 u_i 在哪些领域信任用户 u_j 的问题转化为求用户 u_j 的擅长领域 F_j。

1.1.1　根据用户行为判断用户擅长领域

定义 1　用户的擅长领域:指用户在某一领域具有一定的了解,其见解具有参考价值。

本文通过用户活跃度和其评论认可度来判断用户是否在某一领域中擅长。

用户活跃度计算公式

$$B_i(C_k) = \frac{N_{u_i}(C_k)}{N(C_k)} \quad k \in [1, C]$$

评论认可度计算公式

$$A_i(C_k) = \frac{\sum_{i \in C_k} R^{(i)}_{\mathrm{feedback}}}{N_{u_i}(C_k)} \quad N_{u_i}(C_k) \neq 0$$

即当用户 u_i 在领域 C_k 中发布过评论,其评论的认可度为评论过的所有物品的反馈评分的均值。因此,用户擅长领域的判断公式如下

$$\text{若} \begin{cases} B_i(C_k) \geqslant \mathrm{avg}(B(C_k)) \\ A_i(C_k) \geqslant \mathrm{avg}(A(C_k)) \end{cases}, \text{则 } C_k \in F_i \quad (1)$$

用户 u_i 在领域 C_k 中的活跃度超出平均水平,且其他用户的平均认可度大于该领域中所有物品的平均认可度,则领域 C_k 为用户的擅长领域。

1.1.2　根据用户行为判断用户兴趣领域

用户会对其感兴趣领域内的物品给予较多关注。本文中通过用户对不同领域内物品的评论数量来判断其兴趣领域。具体规则如下:

定义 2　兴趣领域:用户在某一领域中发布评论数大于等于阈值 θ 且其评论数量最多的领域是用户的兴趣领域,即满足公式

$$I_i = \max(N_{u_i}(C_k)) \text{ 当 } N_{u_i}(C_k) \geqslant \theta, k \in [1, c] \quad (2)$$

式中,c 表示领域的种类数,θ 是阈值。若用户在所有领域的评论数均小于阈值 θ,那么该用户将无法确定兴趣领域。

1.1.3 显性领域信任用户

当用户 u_i 对用户 u_j 有泛化信任,而用户 u_j 擅长的领域 C_k 又与用户 u_i 感兴趣的领域一致,那么根据生活经验,人们往往习惯于信任对方擅长的领域,可以得出用户 u_i 在用户 u_j 擅长的领域 C_k 中信任 u_j。因此,将寻找用户 u_i 在其兴趣领域中的领域信任用户问题,分解为 3 个小问题:1) 求用户 u_i 的兴趣领域 I_i;2) 求用户 u_j 的擅长领域 F_j;3) I_i 与 F_j 是否相等;即满足公式

$$\begin{cases} 1.\text{若 } I_i = F_j \text{ 且 } u_i \xrightarrow{T} u_j,\text{则有 } u_i \xrightarrow{C_k} u_j \\ 2.\text{若 } I_i \neq F_j \text{ 且 } u_i \xrightarrow{T} u_j,\text{则 } u_i \text{ 在领域 } C_k \text{ 中不信任 } u_j \end{cases} \tag{3}$$

式中,$i \in [1,N]$,$k \in [1,c]$,$i \neq j$,N 表示用户数,C 表示领域数。

由(3) 式中情况 1 可得,用户 u_i 在领域 C_k 中的显性领域信任用户是满足情况 1 的所有用户集合;当符合情况 2 时,用户 u_i 在领域 C_k 中的显性领域信任用户集中不包括 u_j。

1.2 基于相似性的隐性领域可信用户挖掘

在兴趣领域中根据相似度寻找目标用户的领域相似用户集,然后根据信任传播机制挖掘隐性领域可信用户(在领域内与目标用户存在信任关系的用户)。可信圈由显性领域信任用户与隐性领域可信用户组成。

1.2.1 领域相似用户

在协同过滤中,用户相似度的主要计算方法有皮尔逊相关系数(pearson correlation coefficient, PCC)和余弦相似度(cosine-based similarity)算法,这 2 种算法都是基于用户共同评分项目来计算用户间的相似度。由于仅通过共同评分项目求得的相似度误差比较大,本文计算无信任关系 2 个用户在领域 C_k 中的领域相似度算法在传统方法上进行改进,考虑共同评分项目数量对相似度的影响。领域相似度为

$$S_{C_k}(u_i,u_j) = \frac{(N_{C_k})^2}{N_{C_k}(u_i) \times N_{C_k}(u_j)} \times S_{C_k}^{\cos}(u_i,u_j) \tag{4}$$

式中,N_{C_k} 是在领域 C_k 中用户 u_i 和 u_j 共同评价项目的数量,$N_{C_k}(u_i)$ 和 $N_{C_k}(u_j)$ 分别表示用户 u_i 和 u_j 在领域 C_k 中评价的项目数量,$S_{C_k}^{\cos}(u_i,u_j)$ 是评分相似度,即根据用户 u_i 和 u_j 在领域 C_k 中共同评价的物品评分情况,使用余弦相似度计算的领域 C_k 中用户 u_i 和 u_j 的用户相似度。该公式充分考虑共同评价项目数量对领域相似度的影响,使得当评分相似度相同时,该领域中共同评价项目数量越多则领域相似度越大。

领域相似用户集计算公式为

$$\text{若 } S_{C_k}(u_i,u_j) > \sigma,\text{则 } u_j \in \boldsymbol{S}_i^{(v)} \tag{5}$$

式中,σ 是阈值。即在领域 C_k 中,用户 u_i 与 u_j 的相似度大于阈值 σ,则认为用户 u_j 是用户 u_i 在领域 C_k 中的领域相似用户。

1.2.2 隐性领域可信用户

本节致力于通过用户的评分信息和已知的信任关系寻找目标用户在兴趣领域中的间接信任用户,这类用户称为隐性领域可信用户,其定义如下:

定义 3 隐性领域可信用户:在领域 C_k 中,若 $u_x \in \boldsymbol{S}_i^{(v)}$,且 $u_i \xrightarrow{T} u_x$,则 $u_x \in \boldsymbol{R}_i^{(v)}$。

根据定义可知,满足以下情况之一可证明目标用户与领域相似用户间存在间接信任关系。

情况 1

$$\boldsymbol{R}_i^{(v)} = \{u_x \mid u_y \xrightarrow{T} u_x, u_y \in [\boldsymbol{T}_{\text{circl}_i}^{(v-1)} \cup \boldsymbol{T}_i^{(v)}], u_x \in \boldsymbol{S}_i^{(v)}\}$$

在领域 C_k 中,若目标用户 u_i 已确认的 $v-1$ 层可信圈或第 v 层的显性领域信任集中,存在任一用户 u_y 信任第 v 层的领域相似用户 u_x(即满足图 1a)所示关系),那么 u_x 是目标用户 u_i 第 v 层的隐性领域可信用户。

证明 $[\boldsymbol{T}_{\text{circl}_i}^{(v-1)} \cup \boldsymbol{T}_i^{(v)}]$ 表示在领域 C_k 中目标用户 u_i 已确认 $v-1$ 层可信圈和第 v 层的显性领域信任用户集,即目前确认的可信圈中用户集。由可信圈的定义可知,$u_y \in [\boldsymbol{T}_{\text{circl}_i}^{(v-1)} \cup \boldsymbol{T}_i^{(v)}]$,则 $u_i \xrightarrow{T} u_y$,又 $u_y \xrightarrow{T} u_x$,根据信任传播的传递性可得 $u_i \xrightarrow{T} u_x$,即 $u_x \in \boldsymbol{R}_i^{(v)}$。

情况 2

$$\boldsymbol{R}_i^{(v)} = \{u_x \mid u_z \xrightarrow{T} u_y \text{ 或 } u_y \xrightarrow{T} u_z,\text{且 } u_y \xrightarrow{T} u_x, u_x \in \boldsymbol{S}_i^{(v)}, u_y \in \boldsymbol{S}_i^{(v)}, u_z \in [\boldsymbol{T}_{\text{circl}_i}^{(v-1)} \cup \boldsymbol{T}_i^{(v)}], x \neq y\}$$

在领域 C_k 中,若目标用户 u_i 已确认的 $v-1$ 层可信圈或第 v 层的显性领域信任集中,存在任一用户 u_z 信任第 v 层的领域相似用户 u_y,且 u_y 信任第 v 层的领域相似用户 u_x(即满足图 1b) 所示关系),那

么 u_x 是目标用户 u_i 第 v 层的隐性领域可信用户。

证明　$u_z \in [\boldsymbol{T}_{\mathrm{circl}_i}^{(v-1)} \cup \boldsymbol{T}_i^{(v)}]$,则 $u_i \xrightarrow{T} u_z$,又 $u_z \xrightarrow{T} u_y$ 且 $u_y \xrightarrow{T} u_x$,根据信任传播的传递性可得 $u_i \xrightarrow{T} u_x$,即 $u_x \in \boldsymbol{R}_i^{(v)}$。

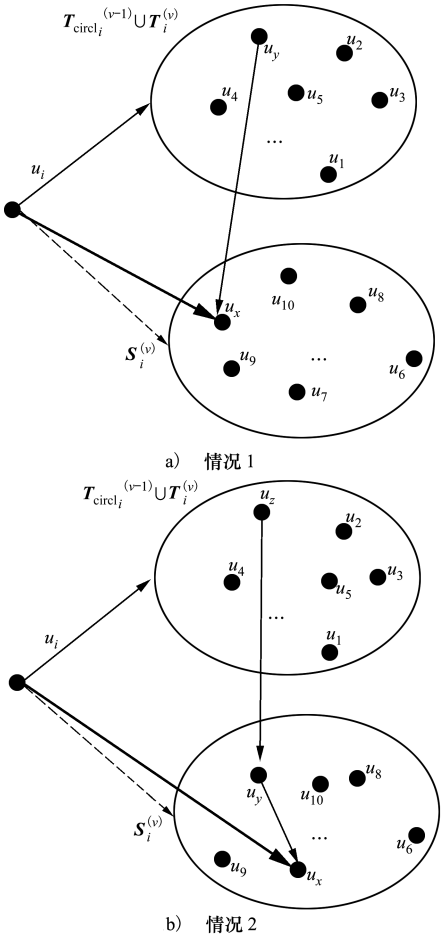

图 1　隐性领域可信用户

1.3　构建面向目标用户兴趣领域的 M 层可信圈

针对目标用户 u_i 的兴趣领域的 M 层可信圈算法描述如下:

输入　网络信任图 $G = (V, E)$,以及用户对各物品的评分 R、其他用户对物品评分的反馈评分 R_{feedback},和物品的领域分类 C

输出　目标用户 u_i 的兴趣领域 I_i 和兴趣领域中的 M 层可信圈

算法步骤:

1) 根据(2)式确定目标用户 u_i 的兴趣领域 I_i

2) 确定兴趣领域 I_i 中的 M 层可信圈

3) for $v = 1$ to M

　　根据(3)式计算用户 u_i 的显性领域信任用户集 $\boldsymbol{T}_i^{(v)}$

　　根据定义 2 计算用户 u_i 的隐性领域可信用户集 $\boldsymbol{R}_i^{(v)}$

　　形成目标用户的可信用户集:

　　$\boldsymbol{T}_{\mathrm{circl}_i}^{(v)} = \boldsymbol{T}_i^{(v)} + \boldsymbol{R}_i^{(v)}$

　　$u_i = \boldsymbol{T}_{\mathrm{circl}_i}^{(v)}$

end

在规模为 n 个用户、e 条信任关系的社交网络中,根据上述算法,计算用户 u_i 的显性领域信任用户集 $\boldsymbol{T}_i^{(v)}$ 时其所需时间复杂度为 $o(n)$,计算用户 u_i 的隐性领域可信用户集 $\boldsymbol{R}_i^{(v)}$,其本质为广度搜索遍历,时间复杂度为 $o(n+e)$;则本文中 M 层可信圈算法的时间复杂度为 $o(n+e)$。

1.4　评分预测

本文所提可信圈评分预测方法:目标用户可信圈中所有用户对项目 i 的评分与经归一化处理后的领域相似度乘积的总和。计算公式如下

$$\hat{r}_{ui} = \sum_{t \in \boldsymbol{T}_{\mathrm{circl}_i}^{(M)}} (n(S(u,t)) \times r_{ti}) \tag{6}$$

式中,$\hat{r}_{ui}$ 表示目标用户 u 对项目 i 的预测评分,r_{ti} 表示用户 t 对项目 i 的评分,$t \in \boldsymbol{T}_{\mathrm{circl}_i}^{(M)}$ 表示用户 t 是目标用户 u 的可信圈中用户,$n(S(u,t))$ 表示经归一化处理的用户 u 与用户 t 的领域相似度。归一化公式为

$$n(S(u,t)) = \frac{S(u,t)}{\sum_{t \in \boldsymbol{T}_{\mathrm{circl}_i}^{(M)}} S(u,t)} \tag{7}$$

2　实验分析

2.1　实验数据集

为了验证本文提出推荐模型的有效性,在 Epinions 数据集进行实验。该数据集共包含 22 166 个用户、296 277 件物品、355 813 条信任关系、922 267 条用户对物品的评分、922 267 条用户对评分质量的反馈评分及 27 种物品的分类。采用 1~5 分制的评分机制,代表评分质量或认可程度从低到高。

本文在用户集中随机选择 10% 的样本作为测试集,即选择了 2 216 个样本,评分预测使用留一法(leave-one-out)来评估模型,保留一条评分信息,利用剩余的评分信息和信任网络来预测该评分。

2.2 实验参数

图 2 分析了用户评论数量分布规律。由图可知不足 6%的用户发布的评论数少于 10,而发布评论数在[11,30]条之间的用户约占总用户的 65%,即大多数用户发布评论的积极性不明显。约 90%的用户发布的评论条数范围是[11,100],仅 6%的用户发布的评论大于 100 条,其中不足 1%的用户呈现超出 1 000 条的海量评论(因超出 100 条评论数的用户所占比例较小,并未在图中显示)。根据分析,为使尽可能多的用户可判定其兴趣领域,在可信圈模型中,设定参数 $\theta = 10$,即用户评论数量需达到 10 条以上。

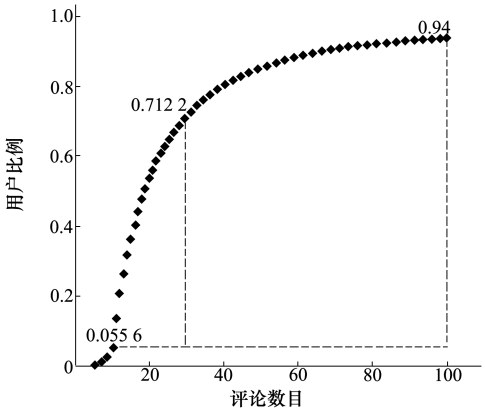

图 2 用户评论数量分布

参数 $\sigma = 0.85$,即 $S_{C_k}(u_i, u_j) > 0.85$,认为用户 u_j 是用户 u_i 在领域 C_k 中的领域相似用户。

2.3 评价指标

本节采用平均方根误差(RMSE)、覆盖率(coverage),精确度(precision)以及 F 指标(F-measure)作为算法的平均指标。

r_{ui} 是用户 u 对项目 i 的真实评分,$\hat{r}_{ui}$ 是系统的预测评分,r_{test} 是所有的测试样例。

覆盖率表示评分预测模型能够预测出评分的用户物品对占测试集的比例。r_{success} 表示测试集中能够预测评分的用户物品的集合。

$$L_{\text{RMSE}} = \sqrt{\frac{\sum_{(ui) \in r_{\text{test}}} (r_{ui} - \hat{r}_{ui})^2}{|r_{\text{test}}|}} \tag{8}$$

$$C = \frac{|r_{\text{success}}|}{|r_{\text{test}}|} \tag{9}$$

$$P = 1 - \frac{L_{\text{RMSE}}}{4} \tag{10}$$

$$F_1 = \frac{2 \times P \times C}{P + C} \tag{11}$$

2.4 实验结果与分析

2.4.1 算法性能分析

为了验证文中提出的基于兴趣领域可信圈挖掘的推荐算法的有效性,在 Epinions 数据集上与经典的随机游走算法(RandomWalk)、基于用户的协同过滤推荐(User-based CF)模型和 SCFinder 模型[12]。进行比较,同时也将基于兴趣领域的可信圈模型(ID-TrustCircle)按照信任与领域是否关联分为基于泛化信任关系的可信圈模型(trust circle based on over-generalized trust relationship, OG-TrustCircle)和基于信任关系领域区分的可信圈模型(trust circle based on domain-specific trust relationship, DS-TrustCircle)进行比较。

图 3 展示了 4 种算法的各个指标对比。从图 3 中精确度和覆盖率 2 项评估指标可以看出,相对于 Item-based CF 模型,其他 4 个模型引入信任关系缓解了数据稀疏问题,明显提升了推荐评分的准确率,除 SCFinder 模型以外的 3 个模型的覆盖率得到显著提升。SCFinder 模型根据兴趣标签构建社交圈时,因删除了兴趣不同的信任用户导致其覆盖率不高。RandomWalk 模型和 DS-TrustCircle 模型在覆盖率评估指标上没有明显的性能差别,因为,根据领域中用户评论物品的相似性融合筛选规则获取的可参考用户弥补了因删减信任关系导致的数据信息减少问题。OG-TrustCircle 模型不仅有泛化信任关系还有基于用户相似性推理出的可参考用户,大大缓解了数据稀疏问题,因此,覆盖率评估指标比其他 3 种算法高出 20%以上。SCFinder 模型比 RandomWalk 模型和基于泛化信任关系的可信圈模型(OG-TrustCircle)的精确度指标低,是因为 SCFinder 模型对社交圈规模进行设定(本文采用原文设置规模,设置社交圈用户规模为 100)使得一些超级用户社交圈规模设置较小具有参考价值的用户并未包含在社交圈,而普通用户其中又包含了许多不具有参考价值的用户,导致精确度降低。RandomWalk 模型因为随机游走的次数达万次,得到的预测参数值远远超过 OG-TrustCircle 模型和 SCFinder 模型得到的预测参考值,因此 RandomWalk 模型的精确率比 OG-TrustCircle 模型和 SCFinder 模型高。DS-TrustCircle 模型并未限制信任圈规模,另外,根据领域构成信任圈使得与预测项目具有相关性的用户都包含在信任

圈中,因此 DS-TrustCircle 模型的精确度指标最高,其预测的准确率最佳。

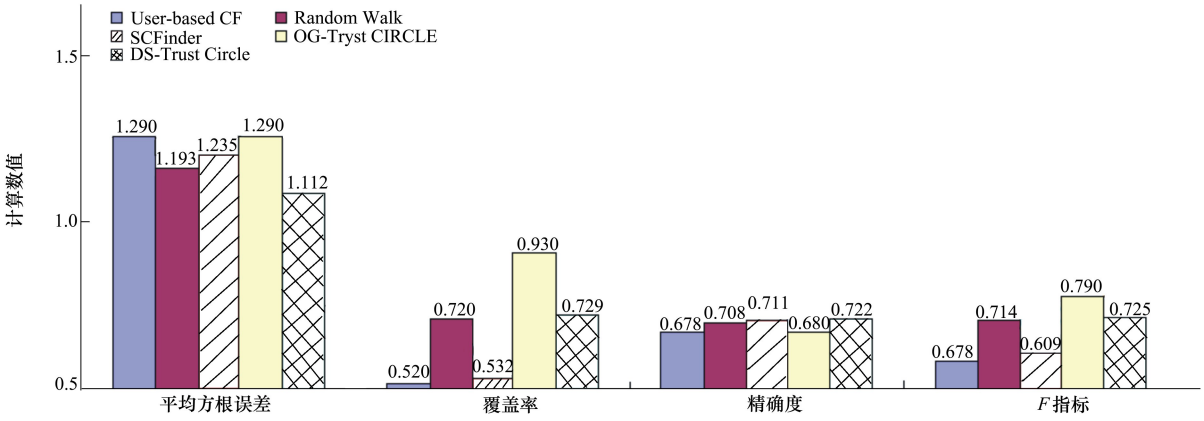

图 3　不同算法的评估指标比较

从实验结果中关于 RMSE 和 F-measure 2 项评估指标可以看出,数据的稀疏性导致 Item-based CF 算法的整体性能较差;RandomWalk 模型和 SCFinder 模型利用泛化信任关系寻找预测项目的关联数据,从一定程度上缓解数据稀疏问题,因此 3 项指标均明显优于 User-based CF 模型;基于信任关系领域区分的可信圈模型(DS-TrustCircle)将泛化的信任关系进行领域区分形成 n 层可信圈,其可信用户与目标用户在物品所属领域内的相关性更为紧密,用户对目标物品的评分也将能够为目标用户对目标物品的评分提供更有价值的参考,因此, DS-TrustCircle 模型的整体性能比 RandomWalk 模型、User-based CF 模型和 SCFinder 模型提高了 7.6%~20.2%。基于泛化信任关系的可信圈模型(OG-TrustCircle)因其覆盖率非常高,其整体性能达到最好; DS-TrustCircle 模型在目标物品所属领域下寻找与目标用户在该领域中具有潜在信任关系的用户,从而使用于预测的参考评分的可参考价值的准确率得到提升,提升了预测的精确度,但因将信任关系进行细分,覆盖率评估指标相比基于泛化信任关系的可信圈有所下降,F_1 指标有所下降。总体而言,在利用用户间信任关系的推荐模型中,DS-TrustCircle 模型预测评分的精确度较高,但若考虑整体覆盖情况,则 OG-TrustCircle 模型预测效果较好。本实验结果同时也说明信任关系与领域有关,不同的领域下,原有的信任关系可能不具有存在意义。在具体的领域下研究信任关系比直接使用泛化信任关系在提高推荐效果方面具有积极作用。

2.4.2　可信圈模型相关指数分析

本节主要对可信圈层数与预测精度的关系进行分析,并分析领域相似度对评分预测的影响。

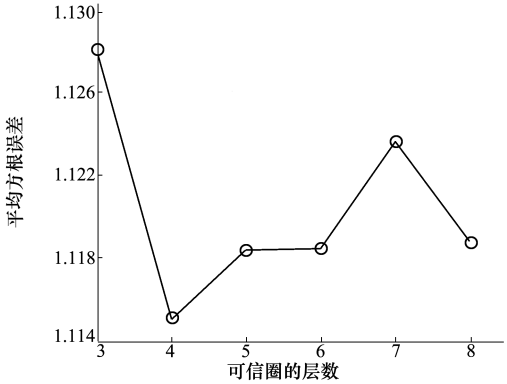

图 4　可信圈层数对平均方根误差(RMSE)的影响

图 4 展示了可信圈层数和预测精度的关系,由图可知可信圈层数为 4 时,RMSE 指标达到最小值,即预测评分效果最佳。评分预测准确性需综合考虑信任衰减和预测所需参考值找的数量 2 项因素。信任衰减主要与信任层数有关,根据三度理论,当信任关系超过三跳信任强度会急剧衰减;根据图 5 中评分预测所需参考值分布,预测评分所需样本数在 4 层可信圈内可穷尽的用户占 80.6%而 3 层可信圈内穷尽的用户只占 50%左右。在第三层预测评分,因有不到 50%的用户其参考值样本不全面而降低准确率,而在第五层之后预测由于大量信任关系较弱的参考值样本的加入降低了预测的准确率。因此,综合考虑两方面的因素,在第四层预测评分准确率较高。

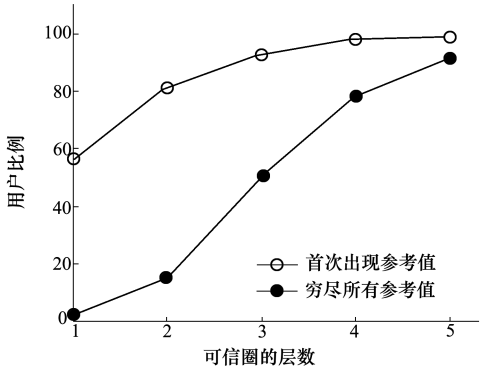

图 5 预测评分所需参考值在可信圈中的层数分布图

2.4.3 算法复杂度分析

基于实验结果,在规模为 n 个用户和 e 个信任关系的社交网络中,User-based CF 模型、Random walk 模型和 SCFinder 模型的时间复杂度均为 $o(n^2)$,可信圈模型时间复杂度为 $o(n+e)$,通过比较发现,可信圈模型在进行评分预测时,不仅显著地降低了算法复杂度,还提高了预测精确度。

3 结 论

本文提出基于用户兴趣领域的信任圈模型,针对不同兴趣领域分层挖掘用户间潜在的隐形信任关系;充分融合显性信任关系为用户资源进行综合评分。该模型不仅考虑信任信息与领域的匹配关系,而且能够挖掘在具体领域下用户间的隐性信任关系,以提高评分预测的精确度和覆盖率。

通过公开数据集上的实验表明本文提出的基于用户兴趣领域可信圈挖掘的推荐模型在准确率及覆盖率上优于对比的协同过滤推荐算法和基于泛化信任关系的随机游走推荐算法。同时还发现基于 4 层可信圈模型进行评分预测其效果最优。

参考文献:

[1] SU X, KHOSHGOFTAAR T M. A Survey of Collaborative Filtering Techniques[J]. Advances in Artificial Intelligence, 2009,12(1): 1-19

[2] YANG X, GUO Y, LIU Y, et al. Bayesian-Inference-Based Recommendation in Online Social Networks[J]. IEEE Trans on Parallel and Distributed Systems, 2013, 24(4): 642-651

[3] LIU F, LEE H J. Use of Social Network Information to Enhance Collaborative Filtering Performance[J]. Expert Systems with Applications, 2010, 37(7): 4772-4778

[4] DELIC A, MASTHOFF J, NEIDHARDT J, et al. How to Use Social Relationships in Group Recommenders: Empirical Evidence[C]//Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization, 2018: 121-129

[5] TAHERI S M, MAHYAR H, FIROUZI M, et al. Extracting Implicit Social Relation for Social Recommendation Techniques in User Rating Prediction[C]//Proceedings of the 26th International Conference on World Wide Web Companion, 2017: 1343-1351

[6] JAMALI M, ESTER M. Trustwalker: a Random Walk Model for Combining Trust-Based and Item-Based Recommendation[C]//Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2009

[7] MASSA P, AVESANI P. Trust-Aware Recommender Systems[C]//Proceedings of the 2007 ACM Conference on Recommender Systems, 2007: 17-24

[8] ZHANG B, HUANG Z, YU J, et al. Trust Computation for Multiple Routes Recommendation in Social Network Sites[J]. Security and Communication Networks, 2014, 7(12): 2258-2276

[9] MA H, KING I, LYU M R, et al. Learning to Recommend with Explicit and Implicit Social Relations[J]. ACM Trans on Intelligent Systems and Technology, 2011, 2(3): 1-19

[10] MA H. An Experimental Study on Implicit Social Recommendation[C]//Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2013: 73-82

[11] MCAULEY J, LESKOVEC J. Discovering Social Circles in Ego Networks[J]. ACM Trans on Knowledge Discovery from Data, 2014, 8(1): 1-28

[12] BURTON S H, GIRAUDCARRIER C G. Discovering Social Circles in Directed Graphs[J]. ACM Trans on Knowledge Discovery from Data, 2014, 8(4):1-27

[13] ZHONG T, LIU F, ZHOU F, et al. Motion Based Inference of Social Circles via Self-Attention and Contextualized Embedding [J]. IEEE Access, 2019, 7: 61934-61948

[14] LAN C, YANG Y, LI X, et al. Learning Social Circles in Ego-Networks Based on Multi-View Network Structure[J]. IEEE Trans on Knowledge and Data Engineering, 2017, 29(8): 1681-1694

[15] YANG X, STECK H, LIU Y. Circle-Based Recommendation in Online Social Networks[C]//Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2012

[16] YIN B, YANG Y, LIU W. ICSRec: Interest Circle-Based Recommendation System Incorporating Social Propagation[C]//4th IEEE International Conference on Information Science and Technology, 2014: 250-255

Recommendation Model for Trust Circle Mining Based on Users′ Interest Fields

BAI Yun, CAI Wandong

(School of Computer Science and Engineering, Northwestern Polytechnical University, Xi′an 710072, China)

Abstract: A trust-based recommendation system recommends the resources needed for users by system rating data and users′ trust relationship. In current relevant work, an over-generalized trust relationship is likely to be considered without exploiting the relationship between trust information and interest fields, affecting the precision and reliability of the recommendation. This research, therefore, proposes a users′ interest-field-based trust circle model. Based on different interest fields, it exploits potential implicit trust relationships in separated layers. Besides, it conducts user rating by combining explicit trust relationships. This model not only considers the matching between trust information and fields, but also explores the implicit trust relationships between users do not revealed in specific fields, thus it is able to improve the precision and coverage of rating prediction. The experiments made with the Epinions data set proved that the recommendation model based on trust circle exploiting in users′ interest fields proposed in this research, is able to effectively improve the precision and coverage of the recommendation rating prediction, compared with the traditional recommendation algorithm based on generalized trust relationship.

Keywords: trust relationship; interest field; recommendation algorithm; trust circle; social network

© 2019 Journal of Northwestern Polytechnical University.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.