

# 在线社会网络环境下基于朋友圈的推荐

张 舒<sup>1</sup>, 王成强<sup>1</sup>, 李 强<sup>1</sup>, 李 慧<sup>2</sup>

(1. 淮海工学院商学院, 江苏 连云港 222005)  
(2. 淮海工学院计算机工程学院, 江苏 连云港 222005)

[摘要] 在线社交网络信息的主要作用是提高推荐系统的精度,但仅仅依靠传统的评分推荐系统(RS)是无法实现的.现在,为了更好地为用户在各活动中提供服务,许多在线社交软件支持一种称为“朋友圈”的新功能,该功能对“朋友”重新进行了定义.论文提出了一种基于朋友圈的推荐系统.该推荐系统旨在处理可用的评分数据并结合社交网络数据推断出特定类别领域的社会信任圈子.主要根据系统预测出的专业水平对圈内朋友划分不同的等级.通过对公开的数据进行验证实验,验证了本文所提出的基于朋友圈的推荐模型可以更好地利用用户的社会信任信息,从而有效地提高推荐系统的准确性.

[关键词] 在线社交网络,评分推荐系统,信任,朋友圈

[中图分类号] TP391 [文献标志码] A [文章编号] 1001-4616(2018)04-0072-07

## Circle-based Recommendation in Online Social Networks

Zhang Shu<sup>1</sup>, Wang Chengqiang<sup>1</sup>, Li Qiang<sup>1</sup>, Li Hui<sup>2</sup>

(1. School of Business, Huaihai Institute of Technology, Lianyungang 222005, China)  
(2. Department of Computer Science, Huaihai Institute of Technology, Lianyungang 222005, China)

**Abstract:** Online social network information promises to increase recommendation accuracy beyond the capabilities of purely rating/feedback-driven recommender systems(RS). As to better serve users' activities across different domains, many online social networks now support a new feature of "Friends Circles", which refines the domain-oblivious "Friends" concept. RS should also benefit from domain-specific "Trust Circles". This paper presents an effort to develop circle-based RS. We focus on inferring category-specific social trust circles from available rating data combined with social network data. We outline several variants of weighting friends within circles based on their inferred expertise levels. Through experiments on publicly available data, we demonstrate that the proposed circle-based recommendation models can better utilize user's social trust information, resulting in increased recommendation accuracy.

**Key words:** online social networks, recommender systems, trust, circle

推荐系统<sup>[1]</sup> (Recommender Systems, RS) 通过向用户推荐他们感兴趣的潜在项目来处理信息过载. 传统的协同过滤方法通过挖掘用户评分历史数据<sup>[1]</sup> 来预测用户的兴趣. 日益流行的在线社交网络提供额外的信息, 以提高基于单纯评分的 RS. 一些基于社会信任的 RS 可以提高推荐准确性<sup>[2]</sup>. 为了更好地服务于不同领域的用户活动, 许多在线社交网络现支持“朋友圈”, 改进了领域中无所不在的“朋友”概念. Google+ 是第一个引入“圈子”功能的, 它允许用户将同学、家人、同事和其他人分配到不同的群组. 长期拥有朋友列表的 Facebook 也推出了“群组”功能, 可将用户分配到不同群组, 从而获得更精细的信息共享, 用户可以与不同的群组共享不同的信息. 在 Twitter 中, 用户可以将关注他们的人(粉丝)组织成“列表”. 当用户查看列表时, 他会看到粉丝发布的一系列推文.

基于圈子的推荐就是应用于这种对多人进行推荐的技术. 推荐产生的结果应尽量满足圈子内所有个体的偏好. 目前, 对于圈子群组推荐系统一般有以下两种实现方案: 一是通过建模将群组内所有个体用户

收稿日期: 2018-05-10.

基金项目: 国家自然科学基金(61403156、61403155)、连云港市科技计划项目(JC1608、CG1611)、淮海工学院自然基金项目(Z2017012、Z2015012)、淮海工学院教学改革项目(XJG2017-2-5)、教育部协同育人项目(201702134005、201701028110).

通讯联系人: 李慧, 博士, 副教授, 研究方向: 数据挖掘, 智能信息处理. E-mail: shufanzs@126.com

聚合为一个虚拟的个体用户,再对此虚拟用户进行推荐以作为对原始的群组推荐<sup>[3]</sup>;二是对于群组内每个用户先进行个性化推荐,再使用某种聚合手段进行聚合,这样就形成了对原始群组推荐的一个单一的推荐列表<sup>[4]</sup>.

对群组推荐的研究主要集中在以下方面:聚合群组内个体用户的推荐列表所用的聚合函数的研究<sup>[5]</sup>;群组内个体用户对于群体的影响<sup>[6]</sup>;对个体进行情感建模以提高群组推荐结果的满意度的研究<sup>[7]</sup>;个性化推荐算法用于群组推荐算法的研究以及尝试利用群组推荐系统对个性化推荐系统相关难题的解决<sup>[8]</sup>.

本文旨在开发一个基于朋友圈的推荐模型 CRS(circle-based recommendation system, CRS). 我们的目标是在可用的评分数据基础上结合社会网络数据来推断特定类别的社会信任圈. 由于在信任圈中所有类别的社会信任链混合在一起,因此我们提出一种新颖的算法来推断特定类别的朋友圈,并根据每个类别中的用户评分活动来推断每个链接上的信任值. 推断一个圈中的链接信任值,首先要评估用户基于自己的评分活动以及所有信任他的用户在一个类别中的专业水平. 然后,向用户分配与他们的专业水平成正比的信任值,最终利用重构的信任圈对推荐模型进行低秩矩阵分解. 通过在公开数据集的验证实验,表明本文提出的基于圈子的推荐模型由于有效地利用了圈子的信任信息,因此可以获得更加精准的推荐效果.

## 1 相关工作

在本文中,我们利用低秩矩阵分解模型进行推荐,因为该模型被认为是协同过滤的最精确的单一模型之一<sup>[9-10]</sup>. 下面首先对本文涉及的相关技术进行介绍.

### 1.1 矩阵分解(Matrix Factorization, MF)

首先介绍基本的低秩矩阵分解(matrix factorization, MF)方法,可以将预测评分矩阵用  $\hat{R} \in \mathbf{R}^{u_0 \times i_0}$  表示,其中  $u_0$  表示用户数量,  $i_0$  表示项目数量,模型可表示为:

$$\hat{R} = r_m + PQ^T, \quad (1)$$

式中,矩阵  $P \in \mathbf{R}^{i_0 \times d}$  和  $Q \in \mathbf{R}^{u_0 \times d}$ ,其中  $d$  是潜在空间的维数,且  $d \leq i_0, u_0$  和  $r_m \in \mathbf{R}$  是(全局)偏移量.

通过最小化平方误差(通常使用 Frobenius/L2 范数正则化)<sup>[11]</sup>,在观察到的评分数据上对该模型进行如下训练:

$$\frac{1}{2} \sum_{(u,i) \in \text{obs.}} (R_{u,i} - \hat{R}_{u,i})^2 + \frac{\lambda}{2} (\|P\|_F^2 + \|Q\|_F^2). \quad (2)$$

式中,  $\hat{R}_{u,i}$  表示等式(1)给出的预测模型. 而  $R_{u,i}$  是表示用户  $u$  对项目  $i$  的训练数据中的实际评分值. 对该目标函数可以使用梯度下降法有效地求解<sup>[12]</sup>. 一旦学习到了低秩矩阵  $P$  和  $Q$ ,任何  $(u,i)$  项目模式都可以根据等式(1)来预测评分值.

### 1.2 MF 和社交网络

社交网络数据的有效利用可以大大提高评分预测的准确性,并且为了将这两个数据源进行整合,已经提出了多种训练模型,例如 SocRec 算法(social recommendation, SoRec)<sup>[13]</sup>, STE 算法(social trust ensemble, STE)<sup>[14]</sup>, 社会正规化推荐系统,推荐系统的自适应社会相似性,其中 SocialMF 模型<sup>[15]</sup>可以达到较低的均方根误差(root mean square error, RMSE)值,因此在我们的实验中,将其作为基准模型进行对比实验.

#### 1.2.1 SocialMF 模型

SocialMF 模型在 RMSE 方面要优于 SoRec 和 STE. 社交网络信息用矩阵  $S \in \mathbf{R}^{u_0 \times u_0}$  表示,其中  $u_0$  是用户数. 用户  $u$  与用户  $v$  (例如用户  $u$  信任/知道/遵循用户  $v$ ) 的有向加权社交关系由  $S_{u,v} \in (0,1]$  表示. 缺席或不可观察的社会关系由  $S_{u,v} = s_m$ , 其中通常  $s_m = 0$  的社交网络矩阵  $S$  的每一行被归一化为 1, 导致每个用户  $u$  具有  $S_{u,v}^* \propto S_{u,v} = s_m$  和  $\sum_v S_{u,v}^* = 1$  的新矩阵  $S^*$ .

SocialMF 的基本思想是社交网络中的邻居可能有类似的兴趣. 这种相似性由等式(3)中的目标函数中第二项进行约束,表示用户模型  $Q_u$  应当类似于他朋友的  $Q_v$  (根据平方误差测量)的(加权)平均值:

$$\frac{1}{2} \sum_{(u,i) \in \text{obs.}} (R_{u,i} - \hat{R}_{u,i})^2 + \frac{\beta}{2} \sum_{\text{all } u} ((Q_u - \sum_v S_{u,v}^* Q_v)(Q_u - \sum_v S_{u,v}^* Q_v)^T) + \frac{\lambda}{2} (\|P\|_F^2 + \|Q\|_F^2) \quad (3)$$

等式(3)可以通过梯度下降法进行优化(参见参考文献[12]中的更新方程).

## 2 基于朋友圈的推荐模型

基于朋友圈的推荐(circle-based recommendation system, CRS)模型可以被看作是 SocialMF 模型到社交网络与朋友圈推断的扩展.

### 2.1 信任圈推断

我们可以从被划分为不同类别(或类型等)的项目评分(或其他反馈)数据进行朋友圈的推断. 基本思想是用户可以信任每个朋友涉及其专业的某些项目类别,但不一定涉及其他类别. 例如,关于汽车的朋友圈可能与儿童电视节目的朋友圈有很大不同. 为此,我们将社交网络  $S$  中的信任关系划分成几个子网  $S^{(c)}$ , 每个子网仅涉及单个项目类别  $c$ .

**定义 1** 推断圈 对于每个项目类别  $c$ , 用户  $v$  在用户  $u$  的推荐圈中, 如在  $C_u^{(c)}$  中, 当且仅当以下两个条件成立:

- 在初始的社会网络中  $S_{u,v} > 0$ , 并且
- 评分数据中  $N_u^{(c)} > 0$  并且  $N_v^{(c)} > 0$

其中  $N_u^{(c)}$  表示在项目类别  $c$  中用户  $u$  已经为项目评分的数量. 否则, 用户  $v$  将不在用户  $u$  关于项目类别  $c$  的圈子里, 即  $v \notin C_u^{(c)}$ .

### 2.2 信任值分配

在社会网络矩阵  $S^{(c)}$  中捕获相同信任圈中朋友之间的信任值(基于项目类别  $c$ ), 使得如果  $v \notin C_u^{(c)}$ ,  $S_{u,v}^{(c)} > 0$  则  $S_{u,v}^{(c)} = 0$ ; 如果  $v \in C_u^{(c)}$ , 则  $S_{u,v}^{(c)} > 0$ . 下面, 我们给出用户  $v$  针对类别  $c$  在用户  $u$  的推断圈中定义正值  $S_{u,v}^{(c)} > 0$  的三种情况. 在第 3 节中将给出验证实验.

#### 2.2.1 CC1: 对等信任

我们首先给出从项目类别  $c$  的推断圈内定义信任值  $S_{u,v}^{(c)} > 0$  的最简单形式: 用户  $u$  的推断圈中的每个用户  $v$  被分配相同的信任值, 即  $S_{u,v}^{(c)*} = \text{const}$ , 如果  $v \in C_u^{(c)}$ . 常数由归一化约束  $\sum_{v \in C_u^{(c)}} S_{u,v}^{(c)*} = 1$  确定, 相当于  $S_{u,v}^{(c)*} = 1/|C_u^{(c)}|$ ,  $\forall u \in C_u^{(c)}$ .

#### 2.2.2 CC2: 基于专业的信任

在本节中, 我们概述了在信任圈内向朋友分配不同信任值的两种形式. 目标是向圈子中的专家朋友分配更高的信任值. 我们使用专家为某项目类别中给出的评分数目作为他们专业水平的近似值. 因为在某项目类别中的专家会比其他人更有可能对该类别中的项目进行评分.

我们将信任值分配给  $C_u^{(c)}$  中的  $u$  粉丝, 与他们在  $c$  类中的专业水平成比例. 基于这个想法, 我们再考虑下面两个形式:

- 变式 a: 在这种情况下, 类别  $c$  中的用户  $v$  的专业级别等于在类别  $c$  中用户  $v$  所给出的评分数量, 即  $E_v^{(c)} = N_v^{(c)}$ . 因此,

$$S_{u,v}^{(c)} = \begin{cases} N_v^{(c)} & \text{if } v \in C_u^{(c)}, \\ 0 & \text{otherwise.} \end{cases}$$

然后我们对  $S^{(c)}$  矩阵的每行进行归一化, 公式为:

$$S_{u,v}^{(c)*} = \frac{S_{u,v}^{(c)}}{\sum_{v \in C^{(c)}_u} S_{u,v}^{(c)}}, \quad (4)$$

其确保将每个圈子中的所有用户  $v$  的权重归一化为 1:

$$\sum_{v \in C^{(c)}_u} S_{u,v}^{(c)*} = 1.$$

- 变式 b: 在这种情况下, 类别  $c$  中的用户  $v$  的专业水平是两个分量的乘积: 第一分量是在类别  $c$  中分配的评分的数量, 第二分量是类别  $c$  中他所有粉丝的投票值  $F_v^{(c)}$ . 可直观认为: 如果大多数  $v$  的粉丝在类别  $c$  中有很高的评分, 并且他们都信任  $v$ , 可以说明  $v$  是  $c$  类中的专家.

我们用  $\mathcal{O}_v^{(c)}$  表示  $c$  类中  $v$  的粉丝投票值. 每个粉丝  $w \in F_v^{(c)}$ , 我们计算它在每个单独类别中的评分分布. 我们用  $D_w$  表示所有类别的分布向量,

$$\mathbf{D}_w = \left( \frac{N_w(1)}{N_w}, \frac{N_w(2)}{N_w}, \dots, \frac{N_w(m)}{N_w} \right). \quad (5)$$

用户  $v$  的专业水平可通过下式进行计算:

$$E_v^{(c)} = N_v^{(c)} \sum_{w \in F_v^{(c)}} \mathbf{D}_w(c),$$

因此,信息值可表示如下:

$$\mathbf{S}_{u,v}^{(c)} = \begin{cases} N_v^{(c)} \sum_{w \in F_v^{(c)}} \mathbf{D}_w(c) & \text{if } v \in C_u^{(c)}, \\ 0 & \text{otherwise.} \end{cases}$$

与上述情况相同,我们最终也在这里对  $\mathbf{S}^{(c)}$  矩阵的每一行进行归一化处理:

$$\mathbf{S}_{u,v}^{(c)*} = \mathbf{S}_{u,v}^{(c)} / \sum_{v \in C_u^{(c)}} \mathbf{S}_{u,v}^{(c)}.$$

### 2.2.3 CC3:信任分裂

从本质上来讲以前的基于圈子的推断和信任值赋值,假设如果  $u$  向  $v$  发出信任语句,并且  $u$  和  $v$  同时在类别  $c$  中具有评分,则  $u$  信任  $v$ . 信任值分配在每个圈中单独计算. 在实践中,用户  $u$  向  $v$  发出信任语句,可能仅仅是因为  $v$  在其同时具有评分的类别子集中的评分. 在  $c$  类中,  $u$  对  $v$  的信任值应该反映这种情况的可能性. 一个简单的方法是使似然率与类别  $c$  中  $v$  的评分成比例. 换句话说,假设  $u$  信任  $v$ , 如果  $v$  在类别  $c_1$  中比在  $c_2$  中具有更多的评分,则  $u$  更可能信任  $v$ , 因为  $v$  在  $c_1$  中的评分比在  $c_2$  的评分高. 如果  $u$  和  $v$  同时在多个类别中具有评分,则  $u$  对  $v$  的信任值应当被那些共同评分类别进行分割.

现在通过  $c$  进行信任值的规范化:

$$\mathbf{S}_{u,v}^{(c)} = \begin{cases} \frac{N_v^{(c)}}{\sum_{c: v \in C_u^{(c)}} N_v^{(c)}} & \text{if } v \in C_u^{(c)}, \\ 0 & \text{otherwise.} \end{cases}$$

像以前一样,规范化  $\mathbf{S}^{(c)}$  矩阵(跨  $v$ )的每一行,以使信任值独立于每个圈中用户的活动水平:

$$\mathbf{S}_{u,v}^{(c)*} = \mathbf{S}_{u,v}^{(c)} / \sum_{v \in C_u^{(c)}} \mathbf{S}_{u,v}^{(c)}.$$

我们注意到,在  $c$  和  $v$  之间的归一化可以被看作迭代过程的第一步,称为迭代比例拟合<sup>[15]</sup>. 当迭代此过程收敛时,其中对于每个  $v$  产生的关于  $c$  和  $v$  的精确联合归一化的结果:  $\sum_c \sum_v \mathbf{S}_{u,v}^{(c)} = \text{const}$ ,  $\sum_c \mathbf{S}_{u,v}^{(c)} = 1$ , 并且每个  $\sum_v \mathbf{S}_{u,v}^{(c)} = \text{const}$ . 虽然迭代过程产生精确的归一化,但是计算成本非常昂贵. 由于这个原因,实验部分中的结果仅在一次迭代后获得.

## 2.3 模型训练

### 2.3.1 每个类别的评分训练

我们使用如上定义的(归一化的)信任网络  $\mathbf{S}^{(c)*}$ , 为每个类别  $c$  训练单独的矩阵分解模型. 对于每种类型涉及的朋友圈,为每个  $c$  获得单独的用户简档  $\mathbf{Q}^{(c)}$  和项目简档  $\mathbf{P}^{(c)}$ . 类似于 SocialMF 模型,但与使用推断社交圈的朋友的关键差异在于我们使用以下模型训练目标函数

$$L^{(c)}(\mathbf{R}^{(c)}, \mathbf{Q}^{(c)}, \mathbf{P}^{(c)}, \mathbf{S}^{(c)*}) = \frac{1}{2} \sum_{(u,i) \text{ obs.}} (R_{u,i}^{(c)} - \hat{R}_{u,i}^{(c)})^2 + \frac{\beta}{2} \sum_{\text{all } u} ((\mathbf{Q}_u^{(c)} - \sum_v \mathbf{S}_{u,v}^{(c)*} \mathbf{Q}_v^{(c)}) (\mathbf{Q}_u^{(c)} - \sum_v \mathbf{S}_{u,v}^{(c)*} \mathbf{Q}_v^{(c)})^T) + \frac{\lambda}{2} (\|\mathbf{P}^{(c)}\|_F^2 + \|\mathbf{Q}^{(c)}\|_F^2). \quad (6)$$

在类别  $c$  中,我们只使用评分  $R_{u,i}^{(c)}$ ;  $\hat{R}_{u,i}^{(c)}$  是类别  $c$  中项目  $i$  的预测评分,

$$\hat{R}_{u,i}^{(c)} = r_m^{(c)} + \mathbf{Q}_u^{(c)} \mathbf{P}_i^{(c)T}. \quad (7)$$

其中我们将全局偏差项  $r_m^{(c)}$  定义为类别  $c$  中观察到的训练评分的平均值. 等式(6)在所有观察到的用户项对  $(u,i)$  上延伸,其中用户项  $i$  属于类别  $c$ . 注意,该模型仅捕获类别  $c$  中的用户和项目简档,即  $\mathbf{Q}^{(c)}$  和  $\mathbf{P}^{(c)}$ ,  $\mathbf{P}^{(c)} \in \mathbf{R}^{i_0^{(c)} \times d}$ , 其中  $i_0^{(c)}$  是类别  $c$  和  $\mathbf{Q}^{(c)} \in \mathbf{R}^{u_0^{(c)} \times d}$  中项目的数量. 等式(6)可以通过梯度下降方法进行求解<sup>[13]</sup>:



$$\frac{\partial L^{(c)}}{\partial Q_u^{(c)}} = \sum_{i: \text{cat}(i)=c} I_{u,i} R^{(c)} (r_m^{(c)} + Q_u^{(c)} P_i^{(c)T} - R_{u,i}^{(c)}) P_i^{(c)} + \lambda Q_u^{(c)} + \beta (Q_u^{(c)} - \sum_{v \in C_u^{(c)}} S_{u,v}^{(c)} * Q_v^{(c)}) + \beta \sum_{v: u \in C_v^{(c)}} S_{u,v}^{(c)} * (Q_v^{(c)} - \sum_{w \in C_v^{(c)}} S_{v,w}^{(c)} * Q_w^{(c)}). \quad (8)$$

其中  $\text{cat}(i)$  是项  $i$  的类别.

$$\frac{\partial L^{(c)}}{\partial P_i^{(c)}} = \sum_{\text{all } u} I_{u,i} R^{(c)} (r_m^{(c)} + Q_u^{(c)} P_i^{(c)T} - R_{u,i}^{(c)}) Q_u^{(c)} + \lambda P_i^{(c)}. \quad (9)$$

对于每个类别  $c$ , 一旦知道了  $Q^{(c)}$  和  $P^{(c)}$ , 则该模型可以根据等式(7)来预测用户对  $(u, i)$  的等级.

3 实验

在本节中, 我们评估本文所提出的推荐算法性能, 并将其在 Epinions 数据集上与现有的基于圈子的推荐的方法进行比较.

3.1 数据集

Epinions 是一个消费者观点网站, 用户可以在其中查看项目 (如汽车, 电影, 书籍, 软件等), 也可以为项目给出 1~5 以内的评分. 发现评论者的评价和评分是有价值的时候, 用户还可以向其他用户表示信任. 每个用户都有一个受信任的用户表. 通过将用户添加到她的信任表来向另一个用户发出信任声明. 在 Epinions 数据集中, 用户之间的信任当否用 0,1 表示: 如果用户 B 在用户 A 的信任列表中, 则用户 A 对 B 的信任值为 1, 否则为 0. 实验中我们使用 Epinions 数据集中 71 002 名用户的评分数据, 其中 451 类别共计 104 356 个不同的项目. 评分的总数是 571,235. 大量评分落入少数大类别中. 前 5 个类别中用户和项目的分布情况如表 1 所示.

表 1 Epinions 数据: 前十大类别统计  
Table 1 Top 10 category statistics in Epinions dataset

类别	用户数量	物品数量	评分数	稀疏性	信任分数
视频和 DVD	17 312	10 065	94 261	0.999 459	44.64%
书籍	11 296	21 662	47 889	0.999 804	36.09%
音乐	10 188	14 905	43 079	0.999 716	21.48%
视频游戏	9 124	2 389	29 661	0.998 639	13.32%
玩具	6 373	3 344	26 789	0.998 743	21.49%

3.2 评估指标

我们在实验中进行 5 次交叉验证. 在每个验证中, 我们使用 80% 的数据作为训练集, 剩余的 20% 作为测试集.

实验使用平均绝对误差 (MAE), 均方根误差 (RMSE) 作为算法性能的评价指标.

MAE (mean absolute error, MAE) 表示所有单个观测值与算术平均值的偏差的绝对值, 该值越小, 表明推荐算法的性能越好, 其定义如下:

$$MAE = \frac{\sum_{i,j} |r_{i,j} - \bar{r}_{i,j}|}{N}. \quad (10)$$

$r_{i,j}$  的含义为用户  $i$  向项目  $j$  做出的具体评分,  $\bar{r}_{i,j}$  的含义为基于应用推荐模型, 用户对项目评分的预期值,  $N$  为评分样本个数.

RMSE (root mean square error, RMSE) 表示观测值与真值偏差的平方与观测次数  $N$  比值的平方根, 该值越小, 表明推荐算法的性能越好, 其定义如下:

$$RMSE = \sqrt{\frac{\sum_{i,j} (r_{i,j} - \bar{r}_{i,j})^2}{N}}. \quad (11)$$

3.3 评价基准

为了验证本文所提出的基于圈子推荐方法的有效性, 我们求推荐法与如下方法进行比较:

- Base MF: 这种方法是矩阵分解方法的基准, 没有考虑社交网络.

• Social MF:该方法在文[13]中提出.它利用了用户之间的社会信任来提高 Base MF 的推荐精度.它使用数据集中提供的所有社交链接.

• CC1:对于在一个类别  $c$  中的推荐,当且仅当通过链接连接的用户对都在类别  $c$  中具有评级时,将使用社交链接.社会信任值的计算如第 2.2.1 节所述.

• CC2:它和 CircleCon1 使用相同的社交链接子集.但社会信任值是根据 2.2.2 节提出的两种基于专业知识的归一化算法计算的.

• CC3:它和 CircleCon1 使用相同的社交链接子集.但是社会信任值是根据第 2.2.3 节提出的信任分裂算法计算的.在所有的实验中,我们设置低秩矩阵分解的维数为  $d=10$ ,正则化常数  $\lambda=0.1$ .

### 3.4 实验结果分析

比较表 2 和表 4(也是表 3 和表 4)中 BaseMF 的性能,我们可以看到,更多的评分信息使 BaseMF 能够更准确地掌握用户的兴趣.当比较表 2 和表 4(也是表 3 和表 4)中的 SocialMF 的性能时,结论同样成立.比较表 2 和表 4(以及表 3 和表 4)中的 CircleCon3,我们可以看到所有类别评分与每类别 Circle 信任关系相结合,在所有情况下提供了最好的预测准确性.这是因为所有类别评分为我们提供了更多的信息来掌握用户的兴趣,为少数类别重建的圈减轻了跨越不同类别发生的混合信任推测所导致的模糊性.

因此,我们发现,通过所有类别评分所提出的变体 CC3 可以得到最低的 RMSE 和 MAE 值.

表 2 使用所有评级(维度  $d=10$ )的 10 个最大类别的 RMSE 比较

Table 2 Ten largest categories of RMSE comparisons using all ratings(dimension  $d=10$ )

类别	Base MF	Social MF	CC1	CC2	CC3
视频和 DVD	0.931(7.66%)	0.890(3.44%)	0.863	0.874	0.860
书籍	0.862(7.36%)	0.834(4.23%)	0.812	0.809	0.799
音乐	0.849(7.14%)	0.806(2.27%)	0.794	0.798	0.788
视频游戏	0.861(9.49%)	0.809(3.67%)	0.786	0.781	0.780
玩具	0.859(10.31%)	0.808(4.69%)	0.777	0.782	0.770

表 3 使用所有评级(维度  $d=10$ )的 10 个最大类别的 MAE 比较

Table 3 Ten largest categories of MAE comparisons using all ratings(dimension  $d=10$ )

类别	Base MF	Social MF	CC1	CC2	CC3
视频和 DVD	0.931(7.66%)	0.890(3.44%)	0.863	0.874	0.860
书籍	0.862(7.36%)	0.834(4.23%)	0.812	0.809	0.799
音乐	0.849(7.14%)	0.806(2.27%)	0.794	0.798	0.788
视频游戏	0.861(9.49%)	0.809(3.67%)	0.786	0.781	0.780
玩具	0.859(10.31%)	0.808(4.69%)	0.777	0.782	0.770

表 4 仅对给定类别的评分进行训练时的绩效比较

Table 4 Training performance comparison given category of ratings

类别	Base MF	Social MF	CC1	CC2	CC3
视频和 DVD	0.931(7.66%)	0.890(3.44%)	0.863	0.874	0.860
书籍	0.862(7.36%)	0.834(4.23%)	0.812	0.809	0.799
音乐	0.849(7.14%)	0.806(2.27%)	0.794	0.798	0.788
视频游戏	0.861(9.49%)	0.809(3.67%)	0.786	0.781	0.780
玩具	0.859(10.31%)	0.808(4.69%)	0.777	0.782	0.770

## 4 结论

在本文中,我们提出了一种通过引入“朋友圈”的概念来提高推荐精度的新颖方法.该方法通过确立用户朋友圈的最佳子集实现在感兴趣的项目类别中进行推荐.由于这些推断的朋友圈是针对各种项目类别定制的,并为每个圈子内的朋友分配不同的权重.在公开数据集上进行验证实验,结果显示该方法相比使用混合社交网络信息的现有方法有显著的优势.

[参考文献]

- [1] ZHANG D, HSU C H, CHEN M, et al. Cold-start recommendation using bi-clustering and fusion for large-scale social recommender systems[J]. IEEE transactions on emerging topics in computing, 2017, 2(2): 239–250.
- [2] WU H, YUE K, PEI Y, et al. Collaborative topic regression with social trust ensemble for recommendation in social media systems[J]. Knowledge-based systems, 2016, 97(C): 111–122.
- [3] ZHAO Z L, WANG C D, WAN Y Y, et al. FTMF: Recommendation in social network with Feature Transfer and Probabilistic Matrix Factorization[C]//International Joint Conference on Neural Networks. New York, USA: IEEE, 2016.
- [4] GOHARI F S, HAGHIGHI H, ALIEE F S. A semantic-enhanced trust based recommender system using ant colony optimization[M]. Holland: Kluwer Academic Publishers, 2017.
- [5] ROY S B, THIRUMURUGANATHAN S, AMER Y S, et al. Exploiting group recommendation functions for flexible preferences[C]//IEEE, International Conference on Data Engineering. USA: IEEE, 2016: 412–423.
- [6] FANG A. Examining information on social media: topic modelling, trend prediction and community classification[C]//International ACM SIGIR conference on Research and Development in Information Retrieval. USA: ACM, 2017.
- [7] CREMONESI P, GARZOTTO F, NEGRO S, et al. Looking for “good” recommendations: a comparative evaluation of recommender systems[C]//Human-Computer Interaction-INTERACT 2011–, Ifip Tc 13 International Conference, Lisbon, Portugal, September 5–9, 2011, Proceedings. Germany: DBLP, 2017.
- [8] AGGARWAL C C. An introduction to recommender systems[M]//Recommender Systems. Germany: Springer International Publishing, 2016.
- [9] ZIĘBA M, TOMCZAK J M, GONCZAREK A. RBM-SMOTE: restricted boltzmann machines for synthetic minority oversampling technique[C]//Asian Conference on Intelligent Information and Database Systems. Germany: Springer International Publishing, 2015: 377–386.
- [10] KUMAR V, PUJARI A K, SAHU S K, et al. Collaborative filtering using multiple binary maximum margin matrix factorizations[J]. Information sciences, 2017, 380(C): 1–11.
- [11] WONG T L, LAM W, XIE H, et al. A joint framework for collaborative filtering and metric learning[J]. Lecture notes in computer science, 2016: 184–196.
- [12] YIN G S, ZHANG Y N, DONG Y X, et al. A constrained trust recommendation using probabilistic matrix factorization[J]. Acta Electronica Sinica, 2014, 42(5): 904–911.
- [13] MA H, YANG H, LYU M R, et al. SoRec: social recommendation using probabilistic matrix factorization[C]//Acm Conference on Information & Knowledge Management. USA: ACM, 2008: 931–940.
- [14] MA H, KING I, LYU M R. Learning to recommend with social trust ensemble[C]//International ACM SIGIR Conference on Research and Development in Information Retrieval. USA: ACM, 2009: 203–210.
- [15] JAMALI M, ESTER M. A matrix factorization technique with trust propagation for recommendation in social networks[C]//ACM Conference on Recommender Systems. USA: ACM, 2010: 135–142.

[责任编辑: 黄 敏]