

# 结合矩阵分解的混合型社会化推荐算法\*

杨丰瑞<sup>1,2</sup>, 刘彪<sup>1†</sup>, 杜托<sup>1</sup>

(1. 重庆邮电大学通信新技术应用研究中心, 重庆 400065; 2. 重庆重邮信科集团股份有限公司, 重庆 401121)

**摘要:** 推荐系统是用来解决当今时代信息过载的重要工具。随着在线社交网络的出现和普及,一些基于网络推荐算法研究的出现已经引起研究者的广泛关注。然而大多数信任感知的推荐系统忽略了用户有不同行为偏好不同的兴趣域。考虑用户间特定域信任网络,并且结合推荐项目之间特征属性信息,提出了一种新型社会化推荐算法(H-PMF)。实验表明,H-PMF算法在评分误差和推荐精度上都取得了很好的效果。

**关键词:** 信任网络; 协同过滤; 矩阵分解; 推荐系统

**中图分类号:** TP391      **文献标志码:** A      **文章编号:** 1001-3695(2018)06-1631-05

**doi:**10.3969/j.issn.1001-3695.2018.06.007

## Hybrid socialized recommendation algorithm based on matrix factorization

Yang Fengrui<sup>1,2</sup>, Liu Biao<sup>1†</sup>, Du Tuo<sup>1</sup>

(1. Communication Technology Applications Resarch Center, Chongqing University of Posts & Telecommunications, Chongqing 400065, China; 2. Chongqing Information Technology(Group) Co. Ltd, Chongqing 401121, China)

**Abstract:** Recommender systems (RSs) have become important tools for solving the problem of information overload. With the emergence and popularity of online social networks, some studies on network-based recommendation algorithm have emerged, raising the concern of many researchers. Trust is one kind of important information available in social networks and is often used for performance improvement in social-network-based RSs. However, most trust-aware RSs ignore the fact that the user has different preference in different domains of interest. This paper proposed a new social recommendation algorithm (H-PMF), which not only considered the user-specific domain trust network, but also combined the feature attribute information between recommended items. Experiments show that the H-PMF algorithm has better performance in both scoring error and recommendation accuracy.

**Key words:** trust networks; collaborative filtering; matrix factorization; recommender system

## 0 引言

随着 Web 2.0 和电子商务的快速发展,大量的互联网用户面临的问题是从如何找到更多的信息到如何找到更有效的信息,常规的推荐方法难以满足不同用户的需要。在这种背景下,为了满足不同用户的个性化需求,基于个性化的推荐算法成为一个热点的研究课题<sup>[1]</sup>。

传统的协同过滤推荐算法通常是收集和分析用户的评级信息来预测用户的兴趣<sup>[2]</sup>。协同过滤技术在推荐领域扮演重要的角色,并且在音乐、电影、电子学习和电子商务等领域已经得到了广泛的应用。近年来,随着社交网络的流行,用户在社交网络的信任关系变得非常重要,通常用于改善推荐性能和解决一些挑战性的问题,如数据的稀疏性和冷启动<sup>[3]</sup>。如 Ma 等人<sup>[4,5]</sup>利用概率矩阵分解技术对用户间的社会关系进行整合来提高推荐效果。

本文主要贡献是在已有社会化推荐算法的基础上提出结合矩阵分解的混合型推荐算法(hybrid recommendation algorithm for matrix factorization based on user and item, H-PMF)。其主要思想是结合用户之间的信任关系、推荐对象之间的特征属

性以及用户项目的评分矩阵,并将三者不同类型的信息关联到统一的推荐算法中。最后本文把测试数据应用到推荐系统上,分析此推荐算法的推荐精度和评分误差,以及在不同参数情况下对实验结果的影响。

## 1 相关工作

推荐系统按照使用的数据可以分为协同过滤、内容过滤和社会化推荐等几类。而社会化推荐系统是利用用户之间的信任网络或推荐项目之间的联系信息进行建模的算法。目前,社会化推荐算法主要是利用用户间的社会信任关系来提高推荐的性能。例如, Ma 等人<sup>[6]</sup>提出了基于矩阵分解的社会化推荐,其主要思想是基于矩阵降低维度的方法将用户之间的社会信任与评分矩阵结合在一起,最终取得了更好的效果。郭磊等人<sup>[7]</sup>在 Ma 等人研究的基础上提出了结合推荐项目间关联关系的社会化推荐算法(PMFUI),它是在矩阵分解的基础上利用推荐项目的关联信息更进一步地优化推荐系统。Jiang 等人<sup>[8]</sup>提出了一种基于社会上下文信息的推荐算法,对推荐项目间的内容相似性用目标函数来约束;其缺点是没有考虑项目间的关联信息。Liu 等人<sup>[9]</sup>提出一种 TruCom 的新型推荐方法。它首

收稿日期: 2017-01-15; 修回日期: 2017-02-23      基金项目: 重庆市研究生科研创新基金资助项目(CY15166)

作者简介: 杨丰瑞(1963-),男,重庆人,教授,博士,主要研究方向为通信新技术应用、数据处理;刘彪(1991-),男(通信作者),硕士研究生,主要研究方向为大数据、数据挖掘(1094210067@qq.com);杜托(1993-),男,硕士研究生,主要研究方向为数据挖掘。

先是生成用户间的特定信任网络,构建统一的目标函数;然后通过将直接和间接信任信息合并到矩阵分解推荐模型中来提高推荐精度。Jiang 等人<sup>[10]</sup>提出了一种结合稀疏矩阵填充的协同过滤算法,它是预先填充稀疏用户项目得分矩阵,实现准确的虚拟分数,并填写虚拟用户项目分数表。在 MovieLens 数据集上的测试表明,该算法有效提高了预测的准确性,在一定程度上解决了冷启动问题。

以上的社会化推荐算法虽然取得了不错的推荐效果,但大多是从用户间的信任网络出发;有些虽然考虑推荐项目的关联信息,如 PMFUI 算法,但是没有细化推荐项目之间的属性特征。因此,本文不仅考虑用户间的域信任网络,并结合项目间关联的特征属性,提出一种混合型的推荐算法。

## 2 H-PMF 算法流程

本章分别描述并细化评分矩阵、用户间社会信任网络矩阵和推荐项目关联关系矩阵的建立过程,并融合三个矩阵在统一框架下实现算法的更优推荐。

### 2.1 社会网络矩阵分解

在本文中使用的概率矩阵分解的方法(probabilistic matrix factorization, PMF)<sup>[11]</sup>作为推荐框架。假设用户的评分矩阵  $R$  中有  $m$  个用户,  $n$  个推荐项目且评分等级在  $[0, 1]$  内。在实际中,大多数推荐系统使用  $1 \sim R_{\max}$  的整数评分等级来表示用户对项目的评级信息。在本文中使用函数  $f(x) = (x - 1)/(R_{\max} - 1)$  将等级  $1 \sim R_{\max}$  映射到区间  $[0, 1]$  上。令  $R_{ij}$  表示用户  $i$  对推荐项目  $j$  的评分。 $U \in \mathbb{R}^{l \times m}$  和  $V \in \mathbb{R}^{l \times n}$  分别表示分解得到用户和项目相关的  $l$ -维特征矩阵,其中列向量  $U_i$  和  $V_j$  分别表示对应的潜在特征向量。因此可以把评分矩阵  $R$  的条件概率分布<sup>[12]</sup>定义为

$$p(R/U, V, \sigma_R^2) = \prod_{i=1}^m \prod_{j=1}^n [N(R_{ij}/g(U_i^T V_j), \sigma_R^2)]^{I_{ij}^R} \quad (1)$$

其中:  $N(x/\mu, \sigma^2)$  表示  $x$  服从均值为  $\mu$ 、方差为  $\sigma^2$  的高斯分布;  $I_{ij}^R$  是指示函数,如果用户  $i$  对项目  $j$  进行了评分,其值为 1,否则值为 0;逻辑函数  $g(x) = 1/(1 + \exp(-x))$ ,目的是将预测值 ( $U_i^T V_j$ ) 限定在区间  $[0, 1]$  上。假设  $U$  和  $V$  服从均值为 0 的球形高斯先验<sup>[13]</sup>,则

$$p(U/\sigma_U^2) = \prod_{i=1}^m N(U_i/0, \sigma_U^2 I),$$

$$p(V/\sigma_V^2) = \prod_{j=1}^n N(V_j/0, \sigma_V^2 I) \quad (2)$$

由贝叶斯推断,可得  $U$  和  $V$  的联合后验概率分布<sup>[11]</sup>:

$$p(U, V/R, \sigma_R^2, \sigma_U^2, \sigma_V^2) \propto P(R/U, V, \sigma_R^2) P(U/\sigma_U^2) P(V/\sigma_V^2) =$$

$$\prod_{i=1}^m \prod_{j=1}^n [N(R_{ij}/g(U_i^T V_j), \sigma_R^2)]^{I_{ij}^R} \times$$

$$\prod_{i=1}^m N(U_i/0, \sigma_U^2 I) \times \prod_{j=1}^n N(V_j/0, \sigma_V^2 I) \quad (3)$$

式(3)的 PMF 只利用了用户的评分矩阵对用户和项目的潜在特征进行评估,并没有考虑用户间的社会信任以及项目间的关联关系。

### 2.2 建立用户之间的社会信任域矩阵

社会信任网络的生成过程主要包括域分区和特定域信任网络生成两个步骤。

a) 域分区。用户通常在多个感兴趣的域之间表现不同。在大多数电子商务网站中,项目通常被分组不同的类别。这些类别在一定程度上表示用户感兴趣的域。因此,在本文中根

据项目的类别划分不同的域。

b) 信任网络生成。域分区意味着原始项目集将根据它们的域/类别被分割成多个项目子集。另外,如果用户已经对该域的一些项目进行了评级,则通过将用户分组到与特定域的相关集合的过程,原始用户集合也将被分割成多个相关用户子集。每个生成的项目集和相关用户集将用于生成特定域的信任网络。

信任网络是加权有向图,其中节点是用户,边是用户之间的信任关系。令公式  $G^d = \langle U^d, E^d, W^d \rangle$  是属于域  $d$  的信任网络,其中  $U^d$ 、 $E^d$  和  $W^d$  分别表示用户集、用户之间的信任关系以及域  $d$  关系权重的集合。信任网络的矩阵表示为  $T^d = [T_{u,v}^d]_{N^d \times N^d}$ ,其中:  $T_{u,v}^d$  分别表示用户  $u$  与  $v$  之间的信任关系;  $N^d$  表示域  $d$  的用户数量。

在图论中,节点之间的关系包括直接和间接邻域关联。因此,在特定域的信任网络中存在两种关系,即直接信任关系和间接信任关系。让  $TD^d$  和  $TI^d$  分别表示直接信任关系矩阵和间接信任关系矩阵。通过去除不属于该域的用户集合中的用户,可以容易地从原始信任网络中提取属于每个域的直接信任关系,如图 1 所示。

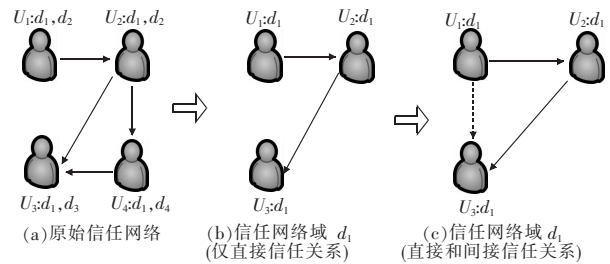


图 1 原始信任网络和信任网络  $d_1$

每个信任的权重值计算公式为

$$TD_{u,v}^d = T_{u,v} \times \frac{N_u^d}{|\cup_{x \in U^d} I_x^d|} \quad (4)$$

其中:  $N_u^d$  是用户  $u$  所评价的项目数;  $T_{u,v}$  是用户  $u$  与  $v$  之间的信任权重值;  $|\cup_{x \in U^d} I_x^d|$  是属于  $U^d$  的用户表示已经评级的项目数量。

间接信任关系可通过社会网络路径中信任传播的过程获得,如图 1 所示。当两个用户通过位于路径中的其他用户连接时,可以使用传播模型来计算它们之间的信任。用户  $u_1$  和  $u_3$  之间没有直接信任关系;然而它们通过用户  $u_2$  连接。沿着路径  $u_1, u_2$  和  $u_3$  计算信任值,如式(5)所示。

$$T_{u1,u3} = T_{u1,u2} \times T_{u2,u3} \quad (5)$$

一般而言,对于一个从  $w$  到  $u$  路径  $\text{path}_{u,w}$  的信任值计算公式如式(6)所示。

$$T(\text{path}_{u,w}) = \prod_{(v_k, v_l) \in (\text{path}_{u,w})} T_{v_k, v_l} \quad (6)$$

其中:对于具有循环路径的信任关系不能用于计算与该路径有关的信任权重。此外根据六度分离的著名理论<sup>[14]</sup>,本文将有效路径的最大长度限制为 6。因此用户  $u$  与  $w$  的间接信任值  $TI_{u,w}$  的计算公式如式(7)所示。

$$TI_{u,w} = \sum_{\text{all path}_{u,w}} T(\text{path}_{u,w}) \quad (7)$$

根据直接信任网络域和间接信任网络可得到用户间的信任网络矩阵  $T$ ,其中  $0 \leq T_{u,w} \leq 1$ 。

### 2.3 优化 SoRec 方法

利用 2.2 节得到的用户间的信任网络矩阵  $T$ 。  $T_{u,w}$  的值代

表示了社会网络中用户  $u$  和  $w$  的熟知程度或信任程度。本文利用 SoRec 的方法假设社会关系矩阵  $T$  代表了用户在社会信任网络的交互信息,通过奇异值分解可以得到用户社交行为的低维潜在特征矩阵。 $U \in \mathbb{R}^{l \times m}$  和  $Z \in \mathbb{R}^{l \times m}$  为分解后与用户相关的特征矩阵,列向量  $U_i$  和  $Z_j$  表示与它们相对应的潜在特征向量。矩阵  $T$  的条件概率分布为

$$p(T/U, Z, \sigma_T^2) = \prod_{i=1}^m \prod_{k=1}^m [N(T_{ij}/g(U_i^T Z_j), \sigma_T^2)]^{I_{ij}^T} \quad (8)$$

其中: $I_{ij}^T$  为指示函数,它表示用户之间存在社会信任关系时,其值为 1,否则为 0。假设  $U, Z$  服从均值为 0 的球形高斯先验:

$$p(U/\sigma_U^2) = \prod_{i=1}^m N(U_i/0, \sigma_U^2 I) \\ p(Z/\sigma_Z^2) = \prod_{k=1}^m N(Z_k/0, \sigma_Z^2 I) \quad (9)$$

此时,SoRec 方法是把用户间的社会信任网络与评级信息结合在一起进行联合分解,评判出具有社会信任关系的用户来进行推荐。其概率模型如图 2 所示。

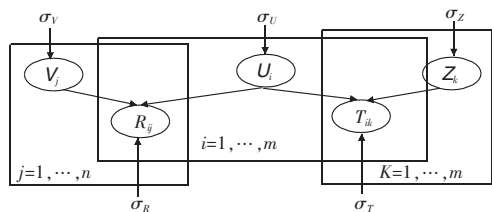


图2 优化 SoRec 的概率模型图

$U, V$  和  $Z$  的联合后验概率分布<sup>[15]</sup>可表示为

$$p(U, V, Z/R, T, \sigma_R^2, \sigma_T^2, \sigma_U^2, \sigma_V^2, \sigma_Z^2) \propto \\ p(R/U, V, \sigma_R^2) p(T/U, Z, \sigma_T^2) p(U/\sigma_U^2) p(V/\sigma_V^2) p(Z/\sigma_Z^2) = \\ \prod_{i=1}^m \prod_{j=1}^n [N(R_{ij}/g(U_i^T V_j), \sigma_R^2)]^{I_{ij}^R} \times \prod_{i=1}^m \prod_{k=1}^m [N(T_{ik}/g(U_i^T Z_k), \sigma_T^2)]^{I_{ik}^T} \times \\ \prod_{i=1}^m N(U_i/0, \sigma_U^2 I) \times \prod_{j=1}^n N(V_j/0, \sigma_V^2 I) \times \prod_{k=1}^m N(Z_k/0, \sigma_Z^2 I) \quad (10)$$

## 2.4 结合推荐项目属性特征的推荐方法

优化的 SoRec 方法更加精确了从用户角度方面的推荐,但认为推荐项目之间是独立的,没有考虑项目之间的关联特性的联系,而推荐项目之间的联系也是用户决策的重要因素。因此,为了把用户之间的内容特性考虑进去,本文在优化的 SoRec 方法的基础上对算法进行扩展,提出结合矩阵因子分解的混合推荐算法(H-PMF)。

构建基于推荐项目之间的关系矩阵  $C$ 。例如当推荐一部电影时,它会介绍电影的导演、主要演员及其电影的类型。电影分为喜剧还是悲剧等各种元素的混合风格,这些描述标签可以被认为项目的关键字。因此,对项目  $x_i = \{A_1, A_2, \dots, A_j, \dots, A_k\}$ ,其中  $A_j$  表示项目  $x_i$  的第  $j$  个特征。 $A_j$  是一个布尔值。如果项目  $x_i$  具有该特征,其值为 1,否则  $A_j = 0$ 。因此,项目集合可以形成一个基于内容特性的矩阵,如表 1 所示。

表1 基于特征项目的指示矩阵

项目	特征 $A_1$	特征 $A_2$	...	特征 $A_n$
项目 $x_1$	0	1	...	1
项目 $x_2$	1	1	...	0
...	...	...	...	...
项目 $x_n$	1	0	...	1

根据特征项目的指示矩阵,就可以计算推荐项目之间的关联关系,因此推荐项目  $x_i$  与  $x_j$  之间的关联关系权重可以表示为

$$C_{x_i x_j} = \sum_{j=1}^k (A_{x_i j} \times A_{x_j j}) / k \quad (11)$$

其中: $A_{x_i j}$  表示项目  $x_i$  有第  $j$  个特性其值为 1,否则为 0; $C_{x_i x_j}$  表示推荐项目  $x_i$  与  $x_j$  之间的特征关联关系,其值的大小代表了它们之间的特征关联程度。则  $C = \{C_{x_i x_j}\}$  表示一个  $(n \times n)$  的特征矩阵,且  $C_{x_i x_j} \in [0, 1]$  表示推荐对象之间的权重。若  $C_{x_i x_j} = 0$  表明两个推荐项目之间无特征关联。

通过对特征关系矩阵  $C$  进行分解,可以得到表示推荐项目之间特征关系的低维潜在特征矩阵。推荐项目  $x_i$  和  $x_k$  的特征关联关系是  $C_{x_i x_k}$ 。在本文中使用  $V$  与推荐项目有关的特征矩阵, $V$  矩阵分解后的低维特征向量为  $V_i$  和  $V_j$ ,因此,特征关系矩阵  $C$  的后验概率分布可表示为

$$p(C/V, \sigma_C^2) = \prod_{i=1}^n \prod_{j=1}^n [N(C_{x_i x_j}/g(V_i^T V_j), \sigma_C^2)]^{I_{ij}^C} \quad (12)$$

其中:指示函数  $I_{ij}^C$  表示推荐项目  $x_i$  和  $x_j$  的特征关联关系,即  $C_{x_i x_j} \neq 0$  时其值为 1,否则为 0。为了优化求解的过程,可以利用共享的潜在特征空间将推荐项目进行特征关联、用户间的社会信任网络与评分信息结合在一起,进一步提高推荐的精度。所用的概率模型如图 3 所示。

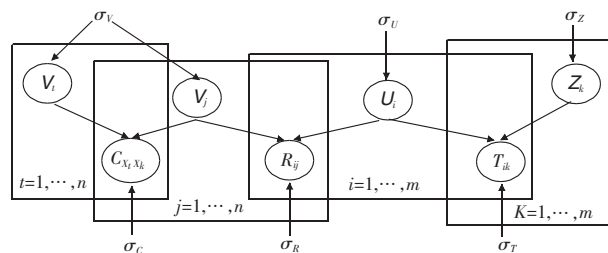


图3 H-PMF 概率模型图

$U, V, Z$  的对数联合后验概率进一步可以表示为

$$\ln p(U, V, Z/R, T, C, \sigma_R^2, \sigma_T^2, \sigma_C^2, \sigma_U^2, \sigma_V^2, \sigma_Z^2) = \\ -\frac{1}{2\sigma_R^2} \sum_{i=1}^m \sum_{j=1}^n [I_{ij}^R (R_{ij} - g(U_i^T V_j))^2] - \frac{1}{2\sigma_T^2} \sum_{i=1}^m \sum_{k=1}^m [I_{ik}^T (T_{ik} - g(U_i^T Z_k))^2] - \\ -\frac{1}{2\sigma_C^2} \sum_{i=1}^n \sum_{j=1}^n [I_{ij}^C (C_{x_i x_j} - g(V_i^T V_j))^2] - \frac{1}{2\sigma_U^2} \sum_{i=1}^m U_i^T U_i - \frac{1}{2\sigma_V^2} \sum_{j=1}^n V_j^T V_j - \\ -\frac{1}{2\sigma_Z^2} \sum_{k=1}^m Z_k^T Z_k - \frac{1}{2} \left( \sum_{i=1}^m \sum_{j=1}^n I_{ij}^R \right) \ln \sigma_R^2 - \frac{1}{2} \left( \sum_{i=1}^m \sum_{k=1}^m I_{ik}^T \right) \ln \sigma_T^2 - \\ -\frac{1}{2} \left( \sum_{i=1}^n \sum_{j=1}^n I_{ij}^C \right) \ln \sigma_C^2 - \frac{1}{2} (ml \ln \sigma_U^2 + nl \ln \sigma_V^2 + ml \ln \sigma_Z^2) + L \quad (13)$$

其中: $L$  是常数,则参数固定时  $U, V, Z$  的极大后验概率等效于最小化带正则项的误差平方和函数,如式(14)所示。

$$E(R, T, C, U, V, Z) = \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n [I_{ij}^R (R_{ij} - g(U_i^T V_j))^2] + \\ + \frac{\lambda_T}{2} \sum_{i=1}^m \sum_{k=1}^m [I_{ik}^T (T_{ik} - g(U_i^T Z_k))^2] + \frac{\lambda_C}{2} \sum_{i=1}^n \sum_{j=1}^n [I_{ij}^C (C_{x_i x_j} - g(V_i^T V_j))^2] + \\ + \frac{\lambda_U}{2} \|U\|_F^2 + \frac{\lambda_V}{2} \|V\|_F^2 + \frac{\lambda_Z}{2} \|Z\|_F^2 \quad (14)$$

其中: $\lambda_T = \sigma_R^2/\sigma_T^2$ ;  $\lambda_C = \sigma_R^2/\sigma_C^2$ ;  $\lambda_U = \sigma_R^2/\sigma_U^2$ ;  $\lambda_V = \sigma_R^2/\sigma_V^2$ ;  $\lambda_Z = \sigma_R^2/\sigma_Z^2$ ;  $\|\cdot\|_F^2$  为 Frobenius 范数。对式(14)所示的目标函数,可以通过  $U, V$  和  $Z$  上使用梯度下降的方法进行求解,使其达到局部最小值。

$$\frac{dE}{dU_i} = \sum_{j=1}^n I_{ij}^R g'(U_i^T V_j) (g(U_i^T V_j) - R_{ij}) V_j +$$

$$\lambda_T \sum_{k=1}^m I_{ik}^T g'(U_i^T Z_k) (g(U_i^T Z_k) - T_{ik}) Z_k + \lambda_U U_i$$

$$\frac{dE}{dV_j} = \sum_{i=1}^m I_{ij}^R g'(U_i^T V_j) (g(U_i^T V_j) - R_{ij}) U_i +$$

$$\lambda_C \sum_{i=1}^n I_{ij}^C g'(V_i^T V_j) (g(V_i^T V_j) - C_{x_i x_j}) V_i + \lambda_V V_j$$

$$\frac{dE}{dZ_k} = \lambda_T \sum_{i=1}^m f_{ik}^T g'(U_i^T Z_k) (g(U_i^T Z_k) - T_{ik}) U_i + \lambda_Z Z_k \quad (15)$$

其中:  $g'(x) = \exp(x) / (1 + \exp(x))^2$  是回归函数  $g(x)$  的导数。

### 3 实验分析与仿真

首先对本文研究的算法和已有的算法进行对比,比较它们的预测精度和推荐质量;然后进一步对 H-PMF 进行分析,比较在不同实验参数情况下对实验结果的影响。

#### 3.1 数据集的设置

本文选择 Epinions 作为社会推荐的数据集。Epinions.com 是一个共享和审查的网站。在此网站,用户需要免费注册,并且能够对其他用户发布的产品、公司、电影等主题提交自己的个人意见。用户对产品进行评论,评级是 1~5 的整数。这些评级和评论将影响未来的客户,决定着他们对于一个产品是否值得购买还是电影值得观看。此外,Epinions 中的每个成员维持一个信任列表用来呈现用户之间的信任关系。这个列表成为信任网,是由 Epinions 用于重新排序产品评论,便于用户首先看到他们信任的用户评价。因此,Epinions 是社会推荐实验的理想数据来源。

在本文的实验中使用的数据集 ([http://www.trustlet.org/wiki/Downloaded\\_Epinions\\_dataset](http://www.trustlet.org/wiki/Downloaded_Epinions_dataset)) 是由 Massa 在 Epinions 网站上收集的,其中包括 49 290 个用户总共有 139 738 个不同项目,每个用户至少有一次评论,总共有 664 824 条评论。用户和项目由匿名标志符表示。数据集由三部分组成,分别为评级数据集、信任数据集以及推荐项目之间关联信息数据集,例如对于评级数据集而言,[12 224 5]代表用户 12 对项目 224 的评级为 5;对于信任数据集而言,[11122 23412 1]代表用户 11122 完全信任用户 23412;对于推荐项目关联数据集[52 101 0.6]而言,代表项目 52 与 101 之间的关联程度为 0.6。

#### 3.2 评估指标

根据实验目的,本文采用平均绝对误差(mean absolute error, MAE)和均方根误差(root mean squared error, RMSE)<sup>[16]</sup>来为用户估计特定项目评分的准确性;查全率(recall)和查准率(precision)<sup>[17]</sup>是评价推荐分类列表的准确性,适用于具有二分喜好的用户系统。

a) MAE 是度量推荐算法预测值与真实值之间的平均绝对偏差。MAE 值越小,估计的准确度越高,定义为  $MAE = \frac{1}{n} \sum_{i \in U, j \in I} |x_{ij} - r_{ij}|$ 。其中: $n$  代表预测评分的个数; $x_{ij}$  代表用户  $i$  对项目  $j$  的预测评分; $r_{ij}$  为用户  $i$  对项目  $j$  的真实评分。RMSE 公式定义为  $RMSE = \sqrt{\frac{1}{n} \sum_{i \in U, j \in I} (x_{ij} - r_{ij})^2}$ , 且 MAE 和 RMSE 值越小,说明预测效果就越好。

b) 查全率(recall)是用来表示推荐列表中是否包含了用户偏好的全部项目。令  $L_i$  表示推荐系统为用户  $i$  产生的推荐列表,在本次实验中本文取  $L_i$  的长度为 5;  $R_i$  表示测试集中用户  $i$  的全部偏好项目,则 recall 的公式为  $recall = \frac{1}{n} \sum_{i=1}^n |L_i \wedge R_i| / |R_i|$ 。其中: $n$  表示训练集中用户的数目;  $|L_i \wedge R_i|$  表示推荐列表中用户实际偏好的项目数量。Precision 用来测量推荐列表中是否都是用户偏好的项目,其公式定义为  $precision = \frac{1}{n} \sum_{i=1}^n$

$|L_i \wedge R_i| / |L_i|$ 。

#### 3.3 仿真结果对比

在实验中本文比较了 PMF、SoRec、PMFUI、TruCom、H-PMF 和一种基于记忆类型的方法 ItemMean<sup>[18]</sup> 在未知数据上的推荐效果。其中参数设置如下: $\lambda_T = 10, \lambda_U = \lambda_V = \lambda_Z = 0.01$ , 且算法的最大迭代次数为 50。

表 2 给出在  $\lambda_c = 10$  时,潜在的特征向量的维度,本文选择推荐系统或学术实验常用的 5 维和 10 维情况下的实验结果。从表 2 中可以得出,在这两个不同维度情况下 H-PMF 方法在评分的误差率和分类的准确率上比其他(实验结果要优于 PMFUI、TruCom、SoRec 和 PMF 这四种方法)方法更有效果。这表明项目之间的属性特征和用户之间的信任网络,能够对推荐结果产生重要的影响。这也证明矩阵分解下融合用户间信任域及项目间关联特性的有效性。

表 2 实验结果对比( $\lambda_c = 10, L = 5$ )

方法	维度 = 5			
	MAE	RMSE	recall	precision
ItemMean	0.621 3	0.604 7	0.394 7	0.104 6
PMF	0.532 1	0.557 1	0.402 4	0.115 4
SoRec	0.523 7	0.547 1	0.422 1	0.123 1
PMFUI	0.515 6	0.535 1	0.435 7	0.127 0
TruCom	0.503 2	0.530 9	0.414 5	0.120 4
H-PMF	0.487 9	0.526 7	0.443 2	0.131 7

方法	维度 = 10			
	MAE	RMSE	recall	precision
ItemMean	0.647 2	0.626 3	0.393 8	0.102 3
PMF	0.532 8	0.557 8	0.403 5	0.107 1
SoRec	0.523 1	0.540 1	0.422 3	0.121 4
PMFUI	0.512 3	0.534 7	0.427 6	0.122 3
TruCom	0.500 4	0.529 6	0.416 8	0.119 3
H-PMF	0.478 3	0.523 1	0.442 1	0.130 1

Precision 和 recall 随推荐列表长度的变化情况如图 4 所示。从图 4 可以看出,当推荐列表的长度不断变化时,这四种推荐方法的 precision 和 recall 受列表长度有明显的变化,但 H-PMF 算法在推荐列表长度不同时,都能够取得较好的推荐结果。对 PMFUI 算法而言,它着重强调的是推荐项目间的关联关系;对 TruCom 算法而言,它着重强调用户之间的域信任网络。从图 4 可以看出,在推荐列表长度较短时,TruCom 算法在查全率和查准率上要优于 PMFUI;当推荐列表达到一定长度后,TruCom 算法在推荐精度要次于 PMFUI。这是由于 TruCom 算法随着推荐列表长度的增加,用户间的信任网络变得更加复杂,导致推荐精度下降。而 H-PMF 算法把这两者因素皆考虑在内,取得了更好的推荐质量。

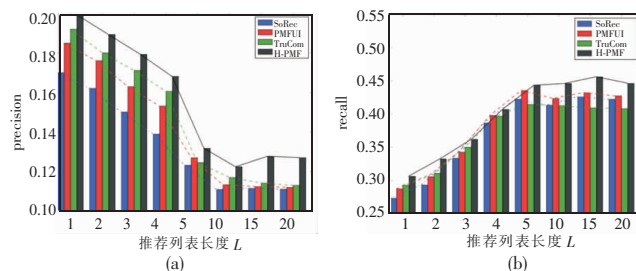


图 4 Precision 和 recall 随推荐列表长度的变化情况

#### 3.4 其他实验参数的影响

在 H-PMF 方法中,参数  $\lambda_c$  和  $\lambda_T$  对系统的推荐效果有着重要的作用。例如  $\lambda_c$  控制推荐项目之间的关联关系,当  $\lambda_c =$



0 时,此方法会只使用用户间的信任关系网络和用户项目的评分矩阵作为推荐的根据;当  $\lambda_c = \infty$  时,此方法仅使用了推荐项目间的联系信息作为推荐的根据。 $\lambda_r$  的值决定着用户之间的关系信任网络,其值的大小决定着用户之间的关系信任网络在推荐依据中所占的比重。例如,在以社交为主的电子网络中,可以把  $\lambda_r$  的值调大;在以项目推荐或销售为主的电子网络中,可以把  $\lambda_c$  的值调小;若一个网站中这两个兼而有之,可以在  $\lambda_c$  和  $\lambda_r$  的值选取中找到一个均衡,实现更为准确的推荐结果。

以  $\lambda_c$  的值为例如来探究对评价指标的影响,可以发现  $\lambda_c$  的取值对实验结果有明显的影响,当  $\lambda_c$  从 0 开始递增时 MAE 和 RMSE 的值逐渐降低,但超过某个阈值后 MAE 和 RMSE 的值就开始增大;而指标 recall 和 precision 开始随  $\lambda_c$  的增加而增加,达到某阈值后开始逐渐减小。这反映了推荐项目之间的关联信息在推荐系统所占的比重直接影响推荐的最终结果,如图 5 所示。

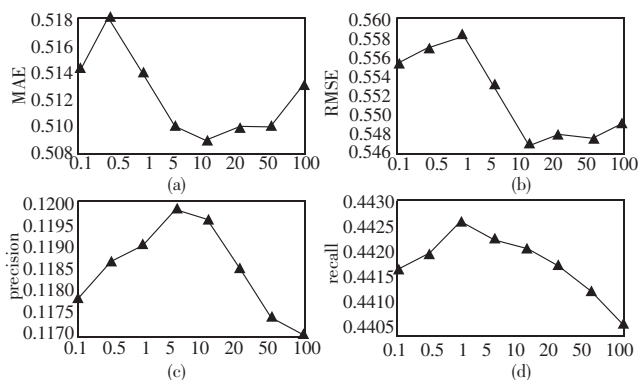


图5 实验参数  $\lambda_c$  的值对实验结果的影响(维度=10)

### 3.5 小结

上述实验分析的结果证明了 H-PMF 比其他算法在预测误差和推荐精度上取得了更好的效果;而实验着重对用户项目的评分矩阵和推荐项目的关联关系信息进行了数据仿真,证明了笔者预期的推荐效果。此外,算法考虑多种因素及其之间的联系,适合复杂网络环境下推荐系统的设计。未来要做的工作是如何建立用户之间的信任网络、数据评分矩阵及推荐项目之间的关联信息这三者之间的联系,以及如何权衡三者之间在特定网络环境中在推荐系统中所占的比重,真正实现推荐系统的智能化、精确化,更好地服务用户。

## 4 结束语

本文提出了结合矩阵分解的混合型社会化推荐算法,该算法在推荐过程中分别考虑了用户间的社会信任关系、评分矩阵以及推荐项目之间的关联特性这三部分信息,并将这三部分信息根据矩阵分解结合在同一框架下。笔者对公开的数据集进行实验,结果表明,推荐项目的关联特性对用户的推荐选择起到了重要的辅助作用。在不同的社会网络中可以调整推荐项目的关联特性在推荐系统所占的比重,使推荐系统达到最优的推荐效果。此外,在算法对比中,本文算法不仅在评分误差和推荐精度上取得了更好的效果,而且考虑了用户间信任关系,恰逢应对未来网络社交化的趋势;本文算法是把多种因素结合在同一框架下,能够通过参数的调整来适应多样的复杂网络类型,实现推荐系统的最优设计。

### 参考文献:

- [1] Duro F, Dolog P. Improving tag-based recommendation with the collaborative value of Wiki pages for knowledge sharing[J]. *Journal of Ambient Intelligence and Humanized Computing*, 2014, 5(1): 21-38.
- [2] Koren Y. Collaborative filtering with temporal dynamics[J]. *Communications of the ACM*, 2010, 53(4): 89-97.
- [3] Massa P, Avesani P. Trust-aware recommender systems[C]//Proc of ACM Conference on Recommender Systems. New York: ACM Press, 2007: 17-24.
- [4] Ma Hao, King I, Lyu M R. Learning to recommend with social trust ensemble[C]//Proc of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM Press, 2009: 203-210.
- [5] Ma Hao, King I, Lyu M R. Recommender systems with social regularization[C]//Proc of the 4th ACM International Conference on Web Search and Data Mining. New York: ACM Press, 2011: 287-296.
- [6] Ma Hao, Yang Haixuan, Lyu M R, et al. SoRec: social recommendation using probabilistic matrix factorization[C]//Proc of the 17th ACM Conference on Information and Knowledge Management. New York: ACM Press, 2008: 931-940.
- [7] 郭磊, 马军, 陈竹敏, 等. 一种结合推荐对象间关联关系的社会化推荐算法[J]. *计算机学报*, 2014, 37(1): 219-228.
- [8] Jiang Meng, Cui Peng, Liu Rui, et al. Social contextual recommendation[C]//Proc of the 21st ACM International Conference on Information and Knowledge Management. New York: ACM Press, 2012: 45-54.
- [9] Liu Haifeng, Xia Feng, Chen Zhen, et al. TruCom: exploiting domain-specific trust networks for multicategory item recommendation[J]. *IEEE Systems Journal*, 2017, 11(1): 295-304.
- [10] Jiang Wei, Yang Liping. Research of improved recommendation algorithm based on collaborative filtering and content prediction[C]//Proc of the 11th International Conference on Computer Science & Education. Piscataway, NJ: IEEE Press, 2016: 598-602.
- [11] Salakhutdinov R, Mnih A. Bayesian probabilistic matrix factorization using Markov chain Monte Carlo[C]//Proc of the 25th International Conference on Machine Learning. New York: ACM Press, 2008: 880-887.
- [12] Yin Dawei, Hong Liangjie, Davison B D. Structural link analysis and prediction in microblogs[C]//Proc of the 20th ACM International Conference on Information and Knowledge Management. New York: ACM Press, 2011: 1163-1168.
- [13] Dueck D, Frey B J. Probabilistic sparse matrix factorization, University of Toronto Technical Report Psi[M]. 2004.
- [14] Lawrence E E, Latha R. Analysis of six degrees of separation in facebook using ant colony optimization[C]//Proc of International Conference on Circuit, Power and Computing Technologies. Piscataway, NJ: IEEE Press, 2015: 1-5.
- [15] Massa P, Avesani P. Trust-aware collaborative filtering for recommender systems[C]//Proc of Confederated International Conferences on the Move to Meaningful Internet Systems. Berlin: Springer, 2004: 492-508.
- [16] Massa P, Bhattacharjee B. Using trust in recommender systems: an experimental analysis[C]//Proc of International Conference on Trust Management. Berlin: Springer, 2004: 221-235.
- [17] 杨博, 赵鹏飞. 推荐算法综述[J]. *山西大学学报: 自然科学版*, 2011, 34(3): 337-350.
- [18] Ricci F, Rokach L, Shapira B. Introduction to recommender systems handbook[M]. [S. l.]: Springer, 2011: 1-35.