

# 基于联合概率矩阵分解的群推荐方法研究

王 刚 蒋 军 王含茹 杨善林

(合肥工业大学管理学院 合肥 230009)

**摘 要** 近年来随着互联网的快速发展,各种虚拟社区不断涌现,用户组成群共同活动的现象逐渐增多,人们开始逐渐关注面向群的推荐.已有的群推荐方法大多是在基于内存的协同过滤推荐方法上进行改进,或是认为组内成员相互独立,忽略了群内成员间的关联关系对群推荐结果的影响.为此,该文提出了一种基于联合概率矩阵分解的群推荐方法,更好地对群推荐问题进行建模.首先,利用用户加入的群的信息计算用户之间的相关性,其次,将用户相关性矩阵融入到概率矩阵分解过程中,得到个人预测评分,最后,利用面向群推荐问题中常用的合成策略对个人预测评分进行融合,得到群对项目的预测评分.进一步将该文提出的方法与现有常用的群推荐方法进行比较,在CiteULike数据集上进行实验,实验结果表明,该文所提出的方法在准确率、召回率等多种评价指标上都取得了更好的推荐结果.

**关键词** 群推荐;用户相关性;群组信息;概率矩阵分解;合成策略

**中图法分类号** TP18 **DOI号** 10.11897/SP.J.1016.2019.00098

## Study of Group Recommendation Based on Probabilistic Matrix Factorization

WANG Gang JIANG Jun WANG Han-Ru YANG Shan-Lin

(School of Management, Hefei University of Technology, Hefei 230009)

**Abstract** In recent years, a lot of virtual communities are emerging with the rapid development of the Internet. However, with the ever-increasing number of the users and generated information, there is difficult for users to find the valuable interesting information on the Internet. Recommendation system has become one of the most important tools to overcome these information overload problems. Meanwhile, the users on the virtual communities gradually intend to establish a group or join certain like-minded groups to facilitate their communication and sharing, which makes the group-oriented recommendation being hotter topics in these days. Researchers begin to pay more attentions to the group recommendation system. On the one hand, existing group recommendation methods are mostly improved by memory-based Collaborative Filtering (CF) method, but the memory-based CF method is seriously affected by the data sparse problem. On the other hand, the interactions among members in a group have not been effectively utilized in existing group recommendation methods, they ignore the influence of the relationships among group members since they just considered the group members were independent of each other. Actually, users in the same group should not be independent but have certain similarities in their preferences. To solve these problems, the model-based CF method, i. e., Probabilistic Matrix Factorization (PMF), is utilized to alleviate the data sparsity problem through adding side information into the model when predicting individual members' preferences. And the group information including

users' common group number and common group size are considered when measuring the users' interactions. Therefore, a novel group recommendation method based on PMF is proposed and presented as a prettier way to model the group recommendation problem in this paper. Firstly, the users' correlations are obtained by incorporating the group information into the measurement. Having assumed that the more common groups the users have and the smaller size of the common group, the higher similarity of users. Secondly, the users' correlations which contain the group information are incorporated into the PMF model to get a better individual prediction value. Finally, to aggregate the individual prediction values into the recommendation list for the group, the mostly used group aggregation strategies, such as the Average strategy, the Least Misery strategy, and the Most Pleasure strategy, are utilized in the aggregation phase. All items' aggregation value for a whole group are sorted in the descending order, and the top  $N$  items are selected and recommended to the group. To evaluate the effectiveness and the feasibility of the proposed method, the experiments were conducted on the CiteULike dataset. Specially, the results were evaluated in terms of Precision and Recall, together with two rank-sensitive metrics, i. e., Mean Average Precision ( $MAP$ ), and Mean Reciprocal Rank ( $MRR$ ). After that, several parameters were discussed including recommendation number, latent factor dimension, and regularization. The experimental results show that the proposed method in this paper has achieved better results at the evaluation metrics including  $Precision$ ,  $Recall$ ,  $MAP$ , and  $MRR$ . It is indicated that the proposed method which considering both the PMF model and the group information can efficiently improve recommendation performance.

**Keywords** group recommendation; user correlation; group information; probabilistic matrix factorization; aggregation strategy

## 1 引言

随着信息时代的发展, 各大社交网络越来越普及, 而用户在使用网络的同时, 每天会产生大量的用户自主生成内容, 用户很难从海量的信息中找到他们感兴趣的信息. 虽然搜索引擎是解决该问题的一种方法, 但它仅能做到搜索某个特定信息而并不能为每个用户都提供个性化主动信息服务<sup>[1]</sup>. 由此, 推荐系统应运而生, 以解决互联网时代背景下产生的信息过载问题. 推荐系统主要基于用户生成的相关数据, 以积极的方式为用户查找感兴趣的信息, 将其原本目标明确的信息搜索问题转换成更偏向用户使用习惯的信息发现问题, 近年来已在信息检索、电子商务以及网络广告等多个领域取得广泛应用, 正成为当前研究的热点<sup>[1-4]</sup>.

传统推荐系统更多的关注于为个体用户生成推荐列表. 然而, 随着互联网及社交网络的发展, 各大社交网络逐渐出现群组功能, 例如, 豆瓣社区有各种音乐组、电影组以及读书组等群组<sup>[5]</sup>. 社交网络上各

种群组的不断涌现, 使得用户根据自己的兴趣建立群组或选择群组加入的现象越来越多, 用户喜欢与同一群组内的好友共同参与某些活动. 这些群组可以增强用户间的交流, 提高用户间信息可共享性, 减少用户搜寻其感兴趣项目的的时间<sup>[6]</sup>. 值得注意的是, 在推荐过程中群组所扮演的重要角色往往被人们忽略. 由于用户通常根据专家的评论或就自己的意见找寻感兴趣的项目, 若推荐系统向用户所在的群组直接生成建议则又会在一定程度上简化用户的搜寻过程. 因此, 面向群组的推荐已成为当前一个新的挑战.

当前, 已有一些研究人员对群推荐问题进行了相关研究. 通过综合个体用户的偏好或合成个体用户的推荐结果, 可以获得最终群体的推荐结果<sup>[7]</sup>. 在针对个体用户的推荐中, 协同过滤被广泛使用, 因其已被验证是一种较有效的推荐方法. 由此, 研究者们为解决群推荐问题过程中也运用协同过滤方法. 例如, Kim 等学者提出两阶段混合推荐书籍方法, 先利用关键词技术寻找近邻, 再利用协同过滤方法生成推荐列表, 最后从草拟推荐列表中删除不相关的

书籍<sup>[8]</sup>. Ortega 等人利用协同过滤方法生成群组推荐结果,比较了在协同过滤的四个阶段分别进行合成的群推荐结果的不同<sup>[7]</sup>. Kassak 等学者提出一种将协同过滤方法与基于内容的方法相结合的混合方法对群组进行推荐,首先,为每个用户生成协同过滤推荐列表和基于内容的推荐列表,然后使用融合策略解决组冲突,最后得到的协同过滤推荐列表由基于内容的推荐列表重新排序<sup>[9]</sup>.

但是,上述群推荐方法仍有一些不足.一方面,传统的基于内存的协同过滤方法如 UserKNN、ItemKNN 等面临着数据稀疏问题,且可扩展性较弱;另一方面,大部分已有研究方法在进行群推荐时,群内成员间的相互作用关系没有得到很好的利用,使得为群组推荐的结果偏离实际群内成员的偏好.针对数据稀疏性问题,一种基于模型的协同过滤方法——矩阵分解,能够利用降维的思想有效缓解此问题<sup>[10]</sup>.然而,目前只有少数研究者在群推荐问题中引入矩阵分解的方法.其次,现有的群推荐方法假设群内成员是相互独立的,而基于群组的成员关系作为一种重要的信息却很少有研究考虑到.

基于上述两个问题,本文提出一种基于联合概率矩阵分解的群推荐方法来对群组进行推荐.首先,引入矩阵分解方法作为推荐框架的基础模型;其次,将用户对不同群组的兴趣偏好信息集成到矩阵分解的过程,通过联合概率矩阵分解方法生成个人预测评分;最后,选取群组推荐中的合成策略,合成各组的个人预测评分,产生群组推荐的结果.本文实验利用了 CiteULike 数据集,实验结果表明本文所提的方法取得了较好的效果,提升了群推荐的精度.

## 2 相关工作

### 2.1 协同过滤方法

协同过滤方法是目前推荐系统中最常用的推荐技术,其主要可以分为两大类:基于内存的方法和基于模型的方法<sup>[11]</sup>.

基于内存的协同过滤方法通过用户的历史信息找到用户或项目之间的相似度较高的邻居,根据邻居对项目的综合评价,预测用户对推荐项目的偏好<sup>[10-12]</sup>.相比之下,基于模型的协同过滤方法则主要是通过用户对项目的评分信息训练模型,以获取用户对项目的偏好程度<sup>[1,13]</sup>.其中,矩阵分解作为基于

模型的协同过滤方法中的一个代表性方法,因其能够利用降维的方法应对高维稀疏的数据,已被验证是一种表现良好的推荐方法且具有良好的可扩展性,逐渐成为协同过滤方法中较主流的方法<sup>[1,13]</sup>.其主要思想是将原始用户项目评分矩阵分解为用户的潜在特征矩阵与项目的潜在特征矩阵,减少原始矩阵的维数的同时,对得到的与原始评分矩阵近似的评分矩阵进行分析得出预测结果.目前,常用的矩阵分解方法有许多种,包括 NMF、SVD 等. Mnih 等人提出了一种概率矩阵分解(PMF)的方法,探讨了矩阵分解方法更深层的概率解释<sup>[14]</sup>.传统的矩阵分解方法虽然能够缓解数据稀疏问题,但是在矩阵分解中融入社会化信息可能会使该问题得到更进一步缓解.因此,在 Web2.0 环境下,一些学者将社交网络中广泛存在的朋友关系、信任关系、标签等社会化信息融入到矩阵分解方法当中,获得了一系列改进的概率矩阵分解方法.

虽然矩阵分解方法在个人推荐中被广泛的使用,并且取得了很好的推荐效果.但是,到目前为止,只有少数研究在群推荐问题中采用了矩阵分解方法,且大部分都没有考虑到群内成员相关性和群大小等群独有的特征对群推荐结果的影响.

### 2.2 群推荐方法

面向群组的推荐方法和面向个人的推荐方法在一定程度上存在着较大的差异.一方面,群推荐需要考虑群内每个成员的满意度,尽可能减少群成员的偏好冲突;另一方面,群推荐还需要找寻合适的融合策略,以使群组的意见能够被很好的表示.为此,已有一些研究者针对群推荐这个特殊的问题进行了相关的研究,主要将群推荐方法分为推荐方法融合和推荐结果融合<sup>[15]</sup>.

个人推荐系统的工作流程可以分为两个阶段:用户偏好获取和推荐生成.而已有的两类群推荐方法则正好与这两个阶段相对应,推荐方法融合在用户偏好获取阶段进行融合,其首先对群内用户的偏好信息单独进行建模,再将每个群内用户的偏好模型生成整个群的偏好模型,最后,基于群组偏好模型,使用个人推荐方法获得群推荐结果<sup>[7,13]</sup>.其具体过程如图 1 所示.而推荐结果融合则在推荐生成的阶段进行融合,首先所有成员的个人推荐结果由个体推荐方法生成,然后,将群组中所有成员的个人推荐结果融合,以获取群推荐结果<sup>[7,15]</sup>.其具体过程如图 2 所示.根据不同类型的个人推荐结果,推荐结果

融合又包含了评分融合与排序融合<sup>[16-17]</sup>. 其中, 评分融合是指合成表示为预测评分的个人推荐结果<sup>[18]</sup>, 而排序融合是指合成表示为排序列表的个人推荐结果<sup>[7,16]</sup>. 目前, 群推荐方法中推荐方法融合使用的较多, 而其在生成群组推荐结果的过程中受评分稀疏性影响较大, 对于不经常评分的用户来说, 很难捕捉其真实偏好, 因此合成的群偏好模型对该用户是有偏差的<sup>[15]</sup>. 而 Pessemier 等人在统一的平台上利用多个评价指标进行实验发现, 在使用基于项目的协同过滤和矩阵分解方法进行群推荐时, 推荐结果融合不仅能够有效地缓解数据稀疏性问题, 而且有效改善了群推荐的精度<sup>[19]</sup>. 由于本文提出的群推荐方法是在矩阵分解方法的基础上进行的, 所以本文在进行群推荐时采用推荐结果融合.

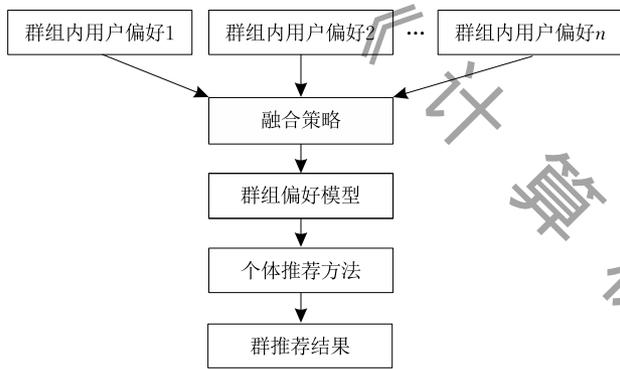


图1 推荐方法融合

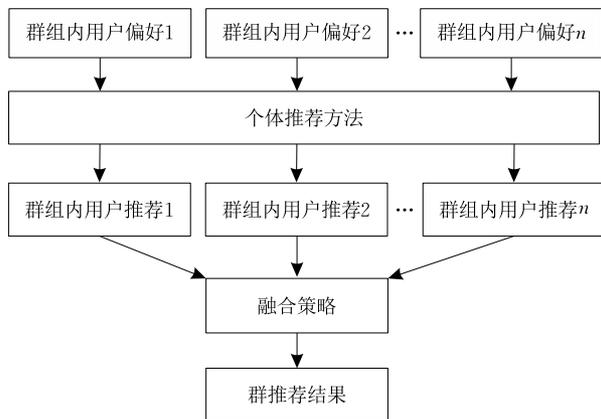


图2 推荐结果融合

尽管上述的推荐方法融合和推荐结果融合各有优缺点, 但是由图1和图2可知, 除了需要选择进行融合的阶段, 融合策略的选择与设计也会对群推荐产生影响<sup>[20-21]</sup>. 目前, 主要的融合策略包含均值策略、最小痛苦策略、Borda 计数策略和公平策略等<sup>[6,20-27]</sup>. 虽然已有很多研究者设计了很多新颖的融合策略, 但是 Pessemier 等众多研究者在实验过

程中发现, 群推荐中融合策略的选择不仅与使用的融合方法相关, 还与群推荐所涉及应用领域和实验过程中所用的数据集等其它众多因素相关.

综上所述, 在已有的协同过滤和群推荐方法的基础上, 仅有少部分研究在群推荐问题中用到矩阵分解方法, 例如, Ortega 等人提出基于协同过滤的矩阵分解方法为群进行推荐, 利用三种策略将用户评分数据合成群特征数据, 其中, 一种是对用户评分矩阵分解之后将用户特征矩阵合成为群组特征矩阵; 一种是在矩阵分解之前直接将用户的评分合成群的评分, 再对其进行分解; 而后者又可以衍生出另一种方法, 在将用户的评分合成群的评分之前, 为每个项目赋予不同权重, 最后对群组评分进行分解<sup>[28]</sup>. 还有部分研究着手于利用成员间的交互信息来进行群推荐, 但只从用户成员自身角度考虑, 没有从群组关系提供的信息出发, 例如, Liu 等人提出的 PIT 模型考虑了群成员的个人喜好和影响, 进一步加强了社交网络信息对个人影响的发现<sup>[29]</sup>. Yuan 等人提出 COM 概率模型也通过给群成员不同的权重来合成群的偏好, 认为与团队相关的专家通常更有影响力<sup>[30]</sup>. 基于已有的研究启发, 本文考虑了在矩阵分解的过程中融入群组的一些特有信息, 并选用了在以前的实验中取得了较好效果的推荐结果融合方法来进行群推荐的研究, 其中融合的过程中选用了多种融合策略进行比较, 以期实现更有效的群推荐方法.

### 3 问题定义

本文根据用户的偏好及用户加入群的信息对群组的偏好进行预测, 对本文问题给出形式化的定义如下. 首先, 系统中  $N$  个用户由  $\mathcal{U} = \{u_1, u_2, \dots, u_i, \dots, u_N\}$  表示,  $M$  个项目由  $\mathcal{V} = \{v_1, v_2, \dots, v_j, \dots, v_M\}$  表示,  $\mathbf{R} = \{R_{i,j}\}_{N \times M}$  指  $u_i$  对  $v_j$  的偏好值组成的矩阵. 另外, 有关群组信息, 系统中  $L$  个群组由  $\mathcal{G} = \{g_1, g_2, \dots, g_l, \dots, g_L\}$  表示, 用户加入群组信息的矩阵由  $\mathbf{G} = \{G_{l,i}\}$  表示, 其中, 如果  $G_{l,i} = 1$  表示用户  $u_i$  加入了群  $g_l$ , 否则  $G_{l,i} = 0$ .

为了对本文方法进行更好的阐述, 下面给出一些定义.

**定义 1.** 任意两用户  $u_i$  和  $u_m$  之间共同拥有的群组的集合, 称为他们的共同加入群集  $GI$ , 则  $GI_{i,m}$  可以表示为

$$GI_{i,m} = \{g_l \mid G_{l,i} = 1 \wedge G_{l,m} = 1, g_l \in \mathcal{G}\} \quad (1)$$

**定义 2.** 如果任意两用户  $u_i$  和  $u_m$  的共同加入群集  $GI_{i,m} \neq \emptyset$ , 那么就认为用户  $u_i$  和用户  $u_m$  之间有着一定的相关性. 根据群组结构信息衡量用户间的关联性, 以相关性矩阵  $S$  表示, 则其中任意两用户  $u_i$  和  $u_m$  的相关性  $S_{i,m}$  可以表示为

$$S_{i,m} = Sim(u_i, u_m), GI_{i,m} \neq \emptyset \quad (2)$$

**定义 3.** 将矩阵分解得到的两个低维矩阵分别称为用户潜在特征矩阵  $U$  和项目潜在特征矩阵  $V$ . 对于任意给定的用户  $u_i \in U$  和项目  $v_j \in V$ ,  $u_i$  对  $v_j$  的评分  $R_{i,j}$  可以计算为  $R_{i,j} \approx U_i^T V_j$ . 而用户潜在特征矩阵  $U$  由  $N$  个向量  $U_i$  构成, 项目潜在特征矩阵  $V$  由  $M$  个向量  $V_j$  构成, 则评分矩阵应满足:

$$R \approx UV \quad (3)$$

**定义 4.** 群预测评分矩阵  $RG$ , 用来表示群组对项目的预测评分, 行代表群组, 列代表项目. 给定一个群  $g_l$  和一个项目  $v_j$ , 选择合适的合成策略函数  $h(\cdot)$ , 对群内每一个用户  $u_i$  对项目  $v_j$  的预测评分进行处理, 获得群  $g_l$  对项目  $v_j$  的预测评分  $RG_{l,j}$ , 如下所示:

$$RG_{l,j} = h(U_i V_j), G_{l,i} = 1 \quad (4)$$

综上, 为了使文章符号表示更加清晰明确, 下面列出文中所用到的符号以及其含义, 如表 1 所示.

表 1 符号定义及含义

符号定义	含义
$U = \{u_1, u_2, \dots, u_i, \dots, u_N\}$	$N$ 个用户的集合
$V = \{v_1, v_2, \dots, v_j, \dots, v_M\}$	$M$ 个项目的集合
$R = \{R_{i,j}\}_{N \times M}$	$N \times M$ 阶用户-项目评分矩阵
$R_{i,j}$	用户 $u_i$ 对项目 $v_j$ 的评分
$\mathcal{G} = \{g_1, g_2, \dots, g_l, \dots, g_L\}$	$L$ 个群的集合
$ g_l $	群 $g_l$ 内包含的成员数目
$G = \{G_{l,i}\}$	群-用户信息矩阵
$G_{l,i}$	用户 $u_i$ 是否加入了群 $g_l$ , 1 是, 0 否
$ G_i $	用户 $u_i$ 加入的群组数量
$GI_{i,m}$	用户 $u_i$ 和用户 $u_m$ 的群组交集
$ GI_{i,m} $	用户 $u_i$ 和用户 $u_m$ 群组交集的数量
$S_{i,m}$	用户 $u_i$ 和用户 $u_m$ 的相似性
$U_i$	用户潜在特征向量
$V_j$	项目潜在特征向量
$RG = \{RG_{l,j}\}$	群预测评分矩阵
$RG_{l,j}$	群 $g_l$ 对项目 $v_j$ 的评分
$ARG$	由均值策略获得的群预测评分矩阵
$LRG$	由最小痛苦策略获得的群预测评分矩阵
$MRG$	由最大幸福策略获得的群预测评分矩阵

## 4 基于联合概率矩阵分解的群推荐方法

### 4.1 基于联合概率矩阵分解的群推荐框架

本研究在群组推荐过程中引入矩阵分解方法, 并以此为基础充分考虑群组的相关特征. 本文所提出的方法框架如图 3 所示.

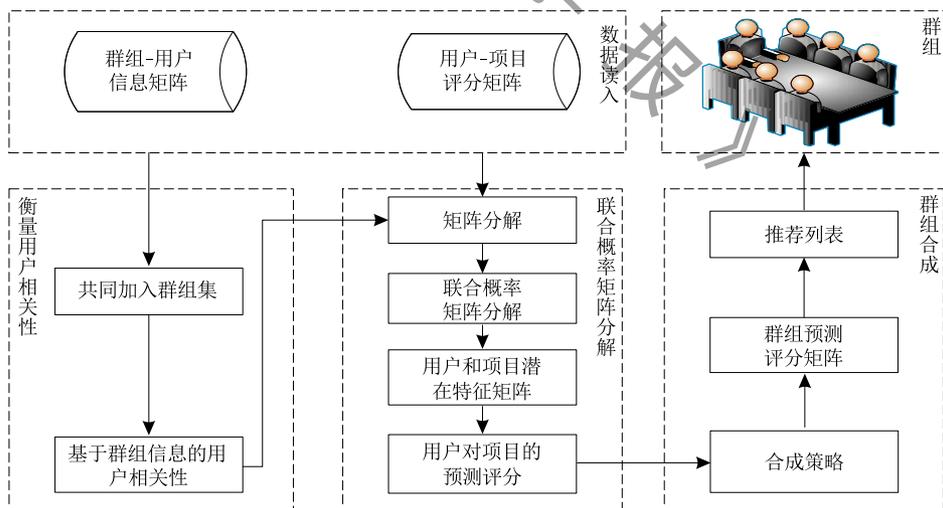


图 3 基于联合概率矩阵分解的群推荐方法框架

根据图 3 可以看出, 本文所提方法主要包括以下步骤:

(1) 衡量用户相关性. 用户间的相互关联性由群组及用户包含的信息获得. 其中, 用户拥有的群组、用户间拥有的共同群组以及共同群组的个数、规模等信息都被考虑在内.

(2) 联合概率矩阵分解过程. 将上述所得的用户的相关性集成到原始的评分矩阵分解过程中, 进行联合矩阵分解, 得到用户对项目的预测评分.

(3) 群组合成过程. 使用常用的合成策略生成群组对项目的预测评分矩阵, 根据评分高低选取最终的项目出现在群推荐列表中.

由此,整个过程通过输入的评分矩阵和用户加入群组的信息,最终为每个群组生成推荐结果.接下来将对上述过程进行详细介绍.

#### 4.2 建立基于群信息的用户相关性矩阵

在衡量用户的相关性时,现有的群推荐方法往往忽略了用户加入群组所产生的信息在一定程度上会对衡量用户的相关性起作用.比如说两个用户拥有的共同群组越多则可能越相似.为了更好的描述用户之间有关群组信息的关联性,构建群组-用户二部图,如图4所示.其中,用户间的连线表示两用户拥有着共同的群组,权重 $|GI_{i,m}|$ 则指向共同群组的个数.由此,用户相关性 $S_{i,m}$ 计算如下:

$$S_{i,m} = \frac{|GI_{i,m}|}{|G_i| + |G_m|} \quad (5)$$

其中, $|G_i|$ 和 $|G_m|$ 分别表示用户 $u_i$ 和用户 $u_m$ 加入的群数目.

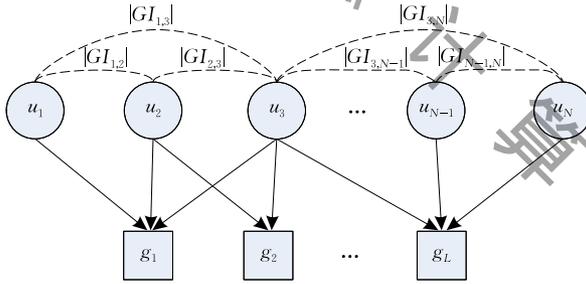


图4 群-用户二部图

此外,可能对用户相关性产生一定影响的因素还包括用户之间拥有的共同群组的规模大小.若用户 $u_1$ 和用户 $u_2$ 共同加入的某个群集,规模大小为10,而用户 $u_3$ 和用户 $u_4$ 共同加入的某个群集,规模大小为10000,那么用户 $u_1, u_2$ 应该比用户 $u_3, u_4$ 更具有相关性.因此,可以合理的认为,用户之间拥有的共同群组的规模越小,组内成员的相似度越高.由此,可以对用户相关性可以优化为

$$S_{i,m} = \frac{1}{|G_i| + |G_m|} \sum_{g_l \in GI_{i,m}} \frac{1}{|g_l| - 1} \quad (6)$$

其中, $|g_l|$ 表示群 $g_l$ 内拥有的成员数目.

#### 4.3 联合概率矩阵分解方法

本文提出一种联合概率矩阵分解方法,将基于群组结构信息的用户相关性集成到概率矩阵分解中,具体过程如下:

首先,假设评分矩阵 $\mathbf{R}$ 中的每个元素 $R_{i,j}$ 是相互独立的,且服从于均值为 $g(\mathbf{U}_i^T \mathbf{V}_j)$ ,方差为 $\sigma_{\mathbf{R}}$ 的高斯分布,据此得到评分矩阵 $\mathbf{R}$ 的概率分布如下:

$$p(\mathbf{R} | \mathbf{U}, \mathbf{V}, \sigma_{\mathbf{R}}^2) = \prod_{i=1}^N \prod_{j=1}^M (N(R_{i,j} | g(\mathbf{U}_i^T \mathbf{V}_j), \sigma_{\mathbf{R}}^2))^{I_{i,j}^{\mathbf{R}}} \quad (7)$$

其中, $N(x | \mu, \sigma^2)$ 表示 $x$ 的高斯分布,其中均值为 $\mu$ ,方差为 $\sigma^2$ . $I_{i,j}^{\mathbf{R}}$ 表示若用户 $u_i$ 在项目 $v_j$ 上的评分不为空,则 $I_{i,j}^{\mathbf{R}}=1$ ,否则 $I_{i,j}^{\mathbf{R}}=0$ ;  $g(x) = 1/(1 + \exp(-x))$ ,其目的是将 $\mathbf{U}_i^T \mathbf{V}_j$ 的值映射到 $[0, 1]$ 区间内.

其次,本研究考虑到用户特征向量要与其相似用户的特征向量近似,由此可得,用户特征矩阵 $\mathbf{U}$ 的高斯先验分布如下:

$$p(\mathbf{U} | \mathbf{S}, \sigma_{\mathbf{U}}^2, \sigma_{\mathbf{S}}^2) \propto p(\mathbf{U} | \sigma_{\mathbf{U}}^2) \times p(\mathbf{U} | \mathbf{S}, \sigma_{\mathbf{S}}^2) = \prod_{i=1}^N N(\mathbf{U}_i | 0, \sigma_{\mathbf{U}}^2 \mathbf{I}) \times \prod_{i=1}^N N(\mathbf{U}_i | \sum_{m=1}^N S_{i,m} \mathbf{U}_m, \sigma_{\mathbf{S}}^2 \mathbf{I}) \quad (8)$$

最后,假设项目特征向量 $\mathbf{v}_j$ 均服从高斯先验分布,均值为0,如下所示:

$$p(\mathbf{V} | \sigma_{\mathbf{V}}^2) = \prod_{j=1}^M N(\mathbf{V}_j | 0, \sigma_{\mathbf{V}}^2 \mathbf{I}) \quad (9)$$

综合以上定义,由贝叶斯推理可得后验概率分布为

$$p(\mathbf{U}, \mathbf{V} | \mathbf{R}, \mathbf{S}, \sigma_{\mathbf{R}}^2, \sigma_{\mathbf{U}}^2, \sigma_{\mathbf{S}}^2, \sigma_{\mathbf{V}}^2) \propto p(\mathbf{R} | \mathbf{U}, \mathbf{V}, \sigma_{\mathbf{R}}^2) p(\mathbf{U} | \mathbf{S}, \sigma_{\mathbf{U}}^2, \sigma_{\mathbf{S}}^2) p(\mathbf{V} | \sigma_{\mathbf{V}}^2) = \prod_{i=1}^N \prod_{j=1}^M (N(R_{i,j} | g(\mathbf{U}_i^T \mathbf{V}_j), \sigma_{\mathbf{R}}^2))^{I_{i,j}^{\mathbf{R}}} \times \prod_{i=1}^N N(\mathbf{U}_i | 0, \sigma_{\mathbf{U}}^2 \mathbf{I}) \times \prod_{i=1}^N N(\mathbf{U}_i | \sum_{m=1}^N S_{i,m} \mathbf{U}_m, \sigma_{\mathbf{S}}^2 \mathbf{I}) \times \prod_{j=1}^M N(\mathbf{V}_j | 0, \sigma_{\mathbf{V}}^2 \mathbf{I}) \quad (10)$$

为了便于求解,对上式两边取 $\ln$ 进行处理,可以得到式(11):

$$\begin{aligned} \ln p(\mathbf{U}, \mathbf{V} | \mathbf{R}, \mathbf{S}, \sigma_{\mathbf{R}}^2, \sigma_{\mathbf{U}}^2, \sigma_{\mathbf{S}}^2, \sigma_{\mathbf{V}}^2) = & -\frac{1}{2\sigma_{\mathbf{R}}^2} \sum_{i=1}^N \sum_{j=1}^M I_{i,j}^{\mathbf{R}} (R_{i,j} - g(\mathbf{U}_i^T \mathbf{V}_j))^2 - \\ & \frac{1}{2\sigma_{\mathbf{S}}^2} \sum_{i=1}^N ((\mathbf{U}_i - \sum_{m=1}^N S_{i,m} \mathbf{U}_m)^T (\mathbf{U}_i - \sum_{m=1}^N S_{i,m} \mathbf{U}_m)) - \\ & \frac{1}{2\sigma_{\mathbf{U}}^2} \sum_{i=1}^N \mathbf{U}_i^T \mathbf{U}_i - \frac{1}{2\sigma_{\mathbf{V}}^2} \sum_{j=1}^M \mathbf{V}_j^T \mathbf{V}_j - \frac{1}{2} (N \times K) \ln \sigma_{\mathbf{U}}^2 - \\ & \frac{1}{2} (\sum_{i=1}^N \sum_{j=1}^M I_{i,j}^{\mathbf{R}}) \ln \sigma_{\mathbf{R}}^2 - \frac{1}{2} (N \times K) \ln \sigma_{\mathbf{S}}^2 - \\ & \frac{1}{2} (M \times K) \ln \sigma_{\mathbf{V}}^2 + C \end{aligned} \quad (11)$$

其中, $K$ 为所求的用户和项目特征矩阵的潜在特征维度,并且 $C$ 为常数.对式(11)进行等价变换,如式(12)所示:

$$L(\mathbf{R}, \mathbf{S}, \mathbf{U}, \mathbf{V}) = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^M I_{i,j}^{\mathbf{R}} (R_{i,j} - g(\mathbf{U}_i^T \mathbf{V}_j))^2 +$$

$$\frac{\lambda_U}{2} \sum_{i=1}^N \mathbf{U}_i^T \mathbf{U}_i + \frac{\lambda_V}{2} \sum_{j=1}^M \mathbf{V}_j^T \mathbf{V}_j +$$

$$\frac{\lambda_S}{2} \sum_{i=1}^N \left( \left( \mathbf{U}_i - \sum_{m=1}^N S_{i,m} \mathbf{U}_m \right)^T \left( \mathbf{U}_i - \sum_{m=1}^N S_{i,m} \mathbf{U}_m \right) \right) \quad (12)$$

其中,  $\lambda_S = \frac{\sigma_R^2}{\sigma_S^2}$ ,  $\lambda_U = \frac{\sigma_R^2}{\sigma_U^2}$ ,  $\lambda_V = \frac{\sigma_R^2}{\sigma_V^2}$ . 上述目标函数的最小

值可由多种方式求得, 本文选择最常用的梯度下降方法进行求解, 由此, 参数  $\mathbf{U}_i$  和  $\mathbf{V}_j$  的梯度下降公式如式(13)和(14)所示:

$$\frac{\partial L}{\partial \mathbf{U}_i} = \sum_{j=1}^M \mathbf{I}_{i,j}^R (g(\mathbf{U}_i^T \mathbf{V}_j) - R_{i,j}) g'(\mathbf{U}_i^T \mathbf{V}_j) \mathbf{V}_j +$$

$$\lambda_S \left( \mathbf{U}_i - \sum_{m=1}^N S_{i,m} \mathbf{U}_m \right) - \lambda_S \sum_{m=1}^N S_{m,i} \left( \mathbf{U}_m - \sum_{t=1}^N S_{t,m} \mathbf{U}_t \right) + \lambda_U \mathbf{U}_i \quad (13)$$

$$\frac{\partial L}{\partial \mathbf{V}_j} = \sum_{i=1}^N \mathbf{I}_{i,j}^R (g(\mathbf{U}_i^T \mathbf{V}_j) - R_{i,j}) g'(\mathbf{U}_i^T \mathbf{V}_j) \mathbf{U}_i + \lambda_V \mathbf{V}_j \quad (14)$$

其中,  $g'(x)$  为  $g(x)$  的导数且  $g'(x) = \exp(-x) / (1 + \exp(-x))^2$ .

#### 4.4 评分合成

本文所提的群推荐方法最后一步则是要对已获得的个人预测评分进行合成, 以获取群组评分. 由上述联合概率矩阵分解方法可得, 用户  $u_i$  对项目  $v_j$  的预测评分  $\hat{\mathbf{R}}_{i,j}$  为

$$\hat{\mathbf{R}}_{i,j} = \sum_{k=1}^K U_{i,k} V_{j,k} \quad (15)$$

接着, 本文选取了均值、最小痛苦和最大幸福三种策略进行融合, 得到最终群推荐预测结果, 以评分矩阵  $\mathbf{RG}(\mathbf{ARG}, \mathbf{LRG}, \mathbf{MRG})$  表示. 其中, 均值策略是指该组所有成员预测评分的平均值作为该群组的预测评分, 其具体计算公式如式(16)所示:

$$\mathbf{ARG}_{l,j} = \frac{\sum_{u_i \in g_l} \hat{\mathbf{R}}_{i,j}}{|g_l|} \quad (16)$$

最小痛苦策略则指的是将群组预测评分与群组内最不满意成员的意见保持一致, 即将群组内所有成员评分的最小值当作群组评分, 其计算公式如式(17)所示:

$$\mathbf{LRG}_{l,j} = \text{Min}_{u_i \in g_l} (\hat{\mathbf{R}}_{i,j}) \quad (17)$$

而最大幸福策略则是将群组预测评分与群组内最满意成员的意见保持一致, 即将群组内所有成员评分的最大值当作群组评分, 其计算公式如式(18)所示:

$$\mathbf{MRG}_{l,j} = \text{Max}_{u_i \in g_l} (\hat{\mathbf{R}}_{i,j}) \quad (18)$$

最后, 由上述三种合成策略分别获取到  $\mathbf{ARG}$ 、 $\mathbf{LRG}$ 、 $\mathbf{MRG}$  后, 再分别形成推荐列表.

综上所述, 本文提出的方法如下所示.

**算法 1.** 基于联合概率矩阵分解的群推荐方法.

输入: 用户集合  $\mathcal{U}$ , 项目集合  $\mathcal{V}$ , 群集合  $\mathcal{G}$ , 用户-项目评分矩阵  $\mathbf{R}$ , 群用户信息矩阵  $\mathbf{G}$ , 最大迭代次数  $I$

输出: 群预测评分  $\mathbf{RG}$

1. FOR  $i=1, 2, \dots, M$
2. FOR  $m=1, 2, \dots, M$
3. IF ( $i \neq m$ )
4. 根据式(6)计算用户相关性
 
$$S_{i,m} = \frac{1}{|G_i| + |G_m|} \sum_{g_l \in G_{i,m}} \frac{1}{|g_l| - 1}$$
5. END FOR
6. END FOR
7. 初始化用户、项目潜在特征矩阵  $\mathbf{U}, \mathbf{V}$ , 生成随机矩阵  $\mathbf{U}$  和  $\mathbf{V}$
8. FOR  $iter=1, 2, \dots, I$
9. FOR each  $\langle i, j \rangle \in R$
10. 根据式(13)所求梯度更新  $\mathbf{U}_i = \mathbf{U}_i - \alpha \frac{\partial L}{\partial \mathbf{U}_i}$
11. 根据式(14)所求梯度更新  $\mathbf{V}_j = \mathbf{V}_j - \alpha \frac{\partial L}{\partial \mathbf{V}_j}$
12. END FOR
13. END FOR
14. 根据式(15)计算用户-项目预测评分矩阵
 
$$\hat{\mathbf{R}}_{i,j} = \sum_{k=1}^K U_{i,k} V_{j,k}$$
15. FOR  $l=1, 2, \dots, L$
16. FOR  $j=1, 2, \dots, N$
17. 根据式(16)利用均值策略计算群预测评分
 
$$\mathbf{ARG}_{l,j} = \frac{\sum_{u_i \in g_l} \hat{\mathbf{R}}_{i,j}}{|g_l|}$$
18. 根据式(17)利用最小痛苦策略计算群预测评分
 
$$\mathbf{LRG}_{l,j} = \text{Min}_{u_i \in g_l} (\hat{\mathbf{R}}_{i,j})$$
19. 根据式(18)利用最大幸福策略计算群预测评分
 
$$\mathbf{MRG}_{l,j} = \text{Max}_{u_i \in g_l} (\hat{\mathbf{R}}_{i,j})$$
20. END FOR
21. END FOR

#### 4.5 时间复杂度分析

本文所提方法的时间复杂度计算主要分为三个步骤. 首先, 衡量用户相关性时, 由于系统中有  $N$  个用户, 假定用户拥有的群组数为  $\bar{y}$ , 群规模大小为  $\bar{g}$ , 则计算用户相关性需要的时间为  $O(N \times \bar{g} \times \bar{y})$ . 然后, 联合概率矩阵分解时, 由于融入了用户相关性, 假定每个用户评价过  $\bar{t}$  个项目, 每个项目又被  $\bar{r}$  个用户评价, 则每次迭代所需要的时间为  $O(N \times \bar{t} \times K + N \times \bar{g}^2 \times \bar{y}^2 \times K + M \times \bar{r} \times K)$ . 最后, 合成评分

并生成群推荐列表时,由于个人评分矩阵生成的时间复杂度为  $O(N \times M \times K)$ ,而合成群组又需要  $O(L \times \bar{g} \times M)$ ,因此,这一步骤总共的计算时间复杂度应为  $O(N \times M \times K + L \times \bar{g} \times M)$ .由于本文方法计算复杂度主要来自于矩阵分解中的迭代过程,且由  $O(N \times \bar{l} \times K + N \times \bar{g}^2 \times \bar{y}^2 \times K + M \times \bar{r} \times K)$  可以体现,本文所提方法的时间复杂度的增加是线性的,因此其可扩展性较强,并且可以同样适用于数据量较大的情况.

## 5 实验设计

### 5.1 数据集

本研究的数据来源于 CiteULike 网站,它是能够帮助科研学者们增强学术交流的一个科研社交网站,该网站上的用户可以在浏览和阅读文章时,收藏感兴趣的论文.同时,该网站还允许用户创建科研小组或兴趣小组,邀请同领域或同爱好的其他用户加入,共同分享相关研究并进行交流.因此,CiteULike 网站所包含的信息与本研究所需的数据刚好契合,本研究从该网站爬取了用户 12 379 个、学术文章 1 343 257 篇、用户收藏学术文章的信息 1 563 838 次以及群组 4 748 个.为了保证数据的质量,我们对爬取的原始数据进行了一定的筛选,首先被用户收藏的次数小于 2 次的学术文章被剔除,进一步,收藏学术文章数少于 15 篇的用户以及包含用户数少于 2 个的群组都被剔除.最后,经过预处理,实验数据集包含用户数为 2 065 个、学术文章数为 85 542 篇以及用户收藏学术文章的信息 1 98 744 次和群组数 718 个.

### 5.2 评价指标

本研究选取了 *Precision*、*Recall*、*MAP* 和 *MRR* 作为评价指标.其中,*Precision* 指的是推荐列表中群真实喜欢的项目占推荐总数的比例.*Recall* 则是群真实喜欢的项目被推荐的比例.*MAP*,即平均准确率,在测量准确性的同时,也考虑了为该组推荐的项目的排序.*MRR*,即平均倒数排名,是指在返回的推荐列表中,符合群组偏好的项目位置越靠前,则推荐结果越好.下面给出了这 4 个评价标准的具体定义:

$$Precision = \frac{|R(g_l) \cap T(g_l)|}{|R(g_l)|} \quad (19)$$

$$Recall = \frac{|R(g_l) \cap T(g_l)|}{|T(g_l)|} \quad (20)$$

$$MAP = \frac{1}{|\mathcal{G}|} \sum_{l=1}^{|\mathcal{G}|} \frac{1}{|R(g_l)|} \sum_{k=1}^{|R(g_l)|} Pre(R_{lk}) \quad (21)$$

$$MRR = \frac{1}{|\mathcal{G}|} \sum_{l=1}^{|\mathcal{G}|} \frac{1}{rank(F_l)} \quad (22)$$

其中, $R(g_l)$ 表示群组  $g_l$ 的推荐列表集合, $T(g_l)$ 表示测试集中群内用户喜欢所有项目的集合, $Pre(R_{lk})$ 表示的该项目在推荐列表和群  $g_l$ 的测试集的交集集中的位置, $rank(F_l)$ 表示在推荐列表和测试集中第一个共同出现的项目在推荐列表中的排序.上述 4 个评价指标,可以分别衡量结果的准确性和排序情况,因此,同时考虑以上评价指标对本文所提的群推荐方法进行评价,可以得到相对客观的评价.

### 5.3 对比方法

本研究共选取了 12 种对比方法衡量本文方法的有效性.其中,UserKNN、SVD 以及 PMF 方法为预测个人评分阶段的对比方法;而在为群组合成个人预测评分的阶段,均值策略、最小痛苦策略和最大幸福策略也同样被作为合成方法.此外,为了与最新的群推荐方法进行对比,本文还选取了 Ortega 等人提出的 WBF 方法<sup>[28]</sup>、Kim 等人提出的 TSBook 方法<sup>[8]</sup>以及 Kassak 等人提出的 MultiRec 方法<sup>[9]</sup>加入对比实验.综上,本文提出的对比方法如下:以 UserKNN 为基础方法的群推荐方法 UK\_AVG, UK\_LM, UK\_MP;以 SVD 为基础方法的群推荐方法 SVD\_AVG;SVD\_LM, SVD\_MP;以 PMF 为基础方法的群推荐方法 PMF\_AVG, PMF\_LM, PMF\_MP;文献中最新的群推荐方法 WBF, TSBook, MultiRec.其中,UK\_AVG 表示利用 UserKNN 方法和均值策略相结合的群推荐方法;UK\_LM 表示利用 UserKNN 和最小痛苦策略相结合的群推荐方法;UK\_MP 表示利用 UserKNN 和最大幸福策略相结合的群推荐方法.对于其他的对比方法,本研究采用了与上述相同的命名方式.

### 5.4 实验流程

本研究的实验过程中,矩阵分解的迭代设置为 200 次,且在 UserKNN 方法中,目标用户的邻居数设置为 30 个.随后,对实验中的一些参数进行讨论,分别设置了不同的推荐个数、矩阵分解的潜在特征维度以及社会化参数  $\lambda_s$ ,讨论不同参数值对实验结果的影响.本研究将实验数据集进行随机划分,训练集和测试集的划分比例分别为 80%和 20%,最终的实验结果由十次实验结果的平均值获得.图 5 展示了本研究整体实验流程.

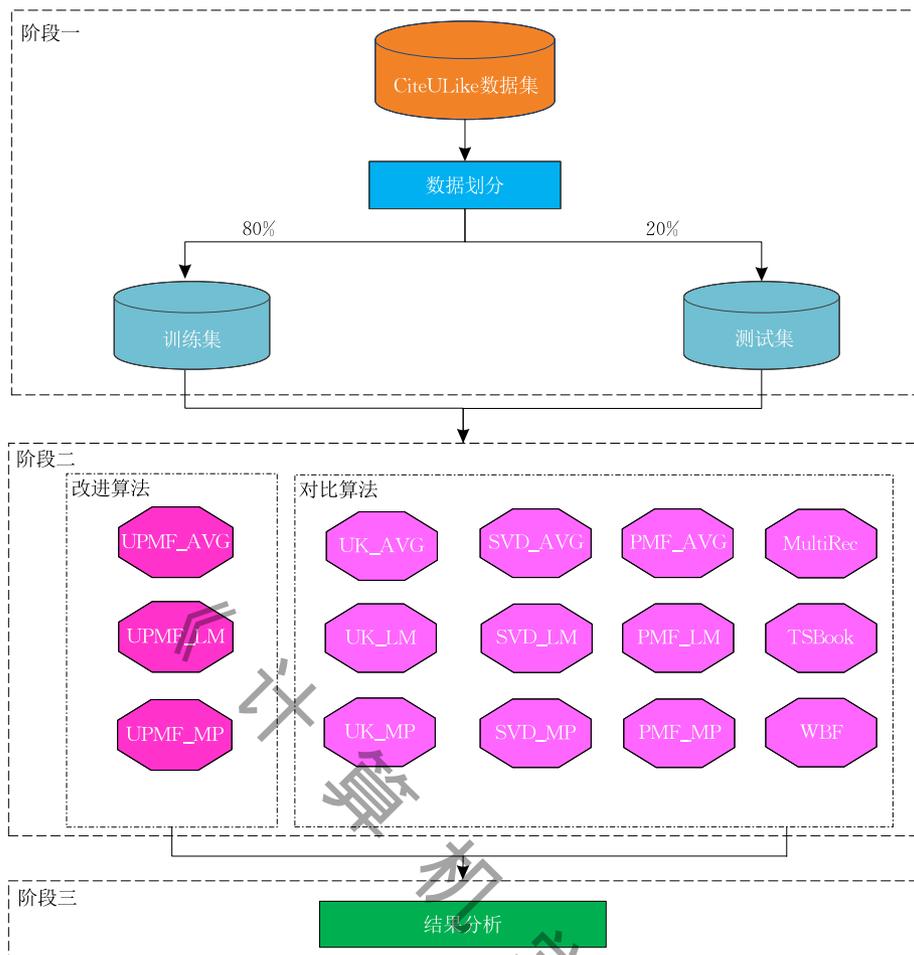


图 5 实验流程图

## 6 结果分析与讨论

### 6.1 实验结果

本研究的实验结果如表 2 所示,其中,取  $\lambda_U = \lambda_V = 0.001$ ,  $\lambda_S = 10$ ,  $K = 10$ , 且推荐个数  $d = 20$ , 表中加粗的数字表示相比之下更优的实验结果。

表 2 实验结果

方法	Precision/%	Recall/%	MAP/%	MRR/%
UK_AVG	4.50	0.95	12.75	15.57
UK_LM	2.98	0.84	11.46	13.29
UK_MP	4.92	0.93	13.45	15.68
SVD_AVG	4.29	0.82	12.52	14.78
SVD_LM	2.52	0.56	8.72	10.09
SVD_MP	4.14	0.87	13.21	15.32
PMF_AVG	4.35	0.84	12.62	14.90
MF_LM	2.56	0.57	8.81	10.14
PMF_MP	4.21	0.89	13.42	15.51
WBF <sup>[28]</sup>	5.11	0.95	15.21	17.44
TSBook <sup>[9]</sup>	4.37	0.87	13.46	15.82
MultiRec <sup>[10]</sup>	4.61	0.98	14.53	16.84
UPMF_AVG	<b>5.26</b>	<b>1.08</b>	14.92	17.69
UPMF_LM	5.16	1.05	14.56	17.32
UPMF_MP	5.21	<b>1.08</b>	<b>15.17</b>	<b>18.01</b>

表 2 所示的结果中, UPMF\_AVG、UPMF\_LM 和 UPMF\_MP 为本文提出的群推荐方法, 由表 2 可以看出: 首先, 本文所提的三种方法在多个评价指标下均表现较优, 进一步表明本文所提的群推荐方法在预测评分时考虑群组相关信息的有效性。其次, 在利用 UserKNN、SVD、PMF 和 UPMF 作为基础方法的群推荐方法中, 可以看出, 作为群合成策略的最小痛苦策略, 与其他两种群合成策略对比, 在大部分情况下均取得了较差的实验结果, 可能的原因是最小痛苦策略利用群内最不满意成员的意见当作群组意见会忽略群内大多数用户都真实感兴趣的文章, 因此最小痛苦策略的结果不太理想。此外, 本文的基础方法 UPMF 在三种群合成策略下的结果相差并不大, 进一步表明本文基础方法较 UserKNN、SVD 和 PMF 具有更好的稳定性。从 WBF 对比方法来看, 其 *Precision* 和 *Recall* 两个评价指标均低于本文提出的 UPMF 方法, 而 *MAP* 和 *MRR* 的值非常接近于 UPMF 方法, 甚至高于 UPMF。这表明本文的方法在推荐顺序满意度上效果与 WBF 相

当,而在推荐精度上明显优于 WBF 方法. MultiRec 方法在所有指标上优于基于 UserKNN、SVD 和 PMF 的群推荐方法,但略低于 WBF 方法,而 TSBook 方法则介于 WBF 方法与 MultiRec 方法两者之间. 综合来说,本文提出的方法整体优于基于 UserKNN、SVD 和 PMF 的群推荐方法以及近几年新提出的群推荐方法,具有更高的有效性.

### 6.2 参数分析

#### (1) 推荐个数 $d$

本实验分别将推荐个数设置为  $d = \{10, 20, 30, 40, 50\}$ , 观察实验结果,具体如图 6~图 9 所示.

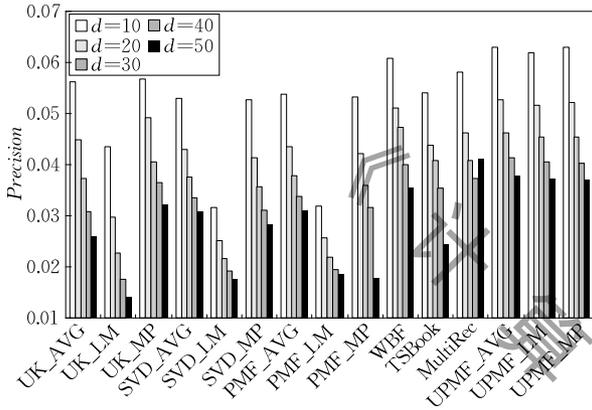


图 6 不同推荐个数  $d$  下的 Precision

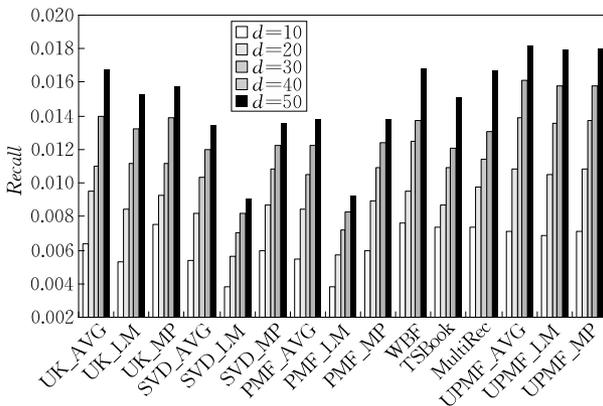


图 7 不同推荐个数  $d$  下的 Recall

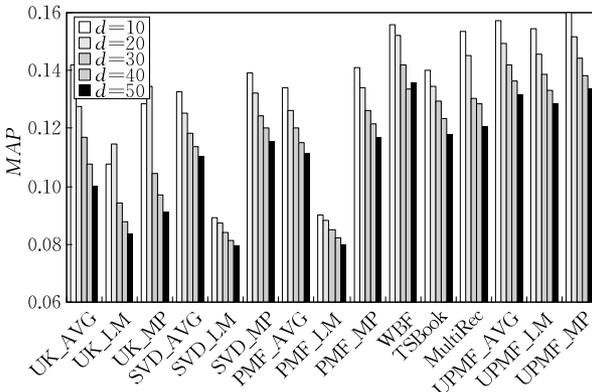


图 8 不同推荐个数  $d$  下的 MAP

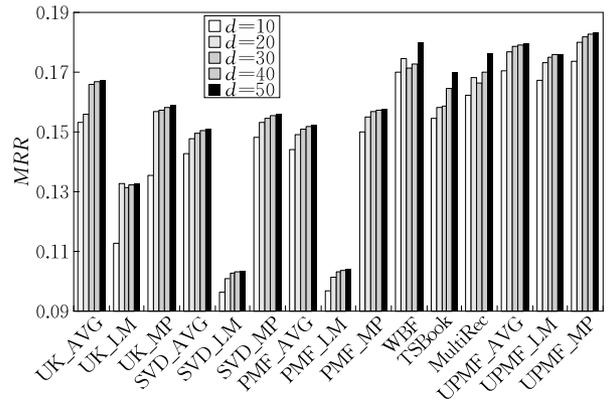


图 9 不同推荐个数  $d$  下的 MRR

图 6~图 9 展示了在不同的推荐个数下,所有方法在不同评价指标下的表现. 从图中可以看出, UPMF\_AVG, UPMF\_LM, UPMF\_MP 三种方法的推荐效果优于其他方法,即考虑群组结构信息的用户关联性可以有效提高群推荐的效果. 从图中还可以看出,推荐个数在 10 到 20 间所对应的值变化较快,而超过 20 以后,却变化得越来越慢,可能的原因是当推荐列表到达一定长度,群组对位于推荐列表尾部的项目之间偏好程度差距不大.

#### (2) 特征向量维度 $K$

矩阵分解中,  $K$  的值设置过小会导致潜在特征无法充分被表示,而设置过大则会增加计算复杂度. 另一方面,将推荐个数  $d$  固定为 10, 讨论不同  $K$  下的实验结果,取  $K = \{5, 10, 15, 20\}$ . 具体实验结果如图 10~图 13 所示.

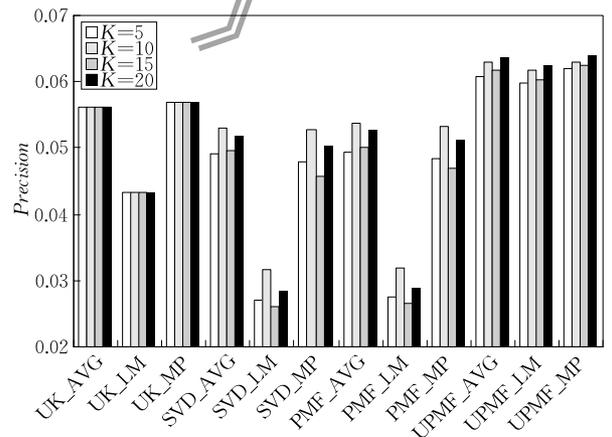


图 10 不同维度  $K$  下的 Precision

由图 10~图 13 可以看出,在基于 SVD 和 PMF 的群推荐方法下,实验结果均在  $K$  为 10 的时候表现较优,而当  $K$  大于 10,实验结果逐渐下降. 其推荐效果并没有随着特征维度的增加而不断提高,这可能是由于群推荐后期的合成过程在一定程度上影响

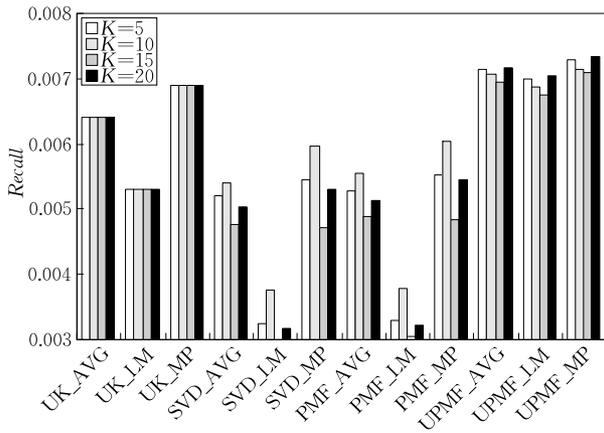


图 11 不同维度  $K$  下的 Recall

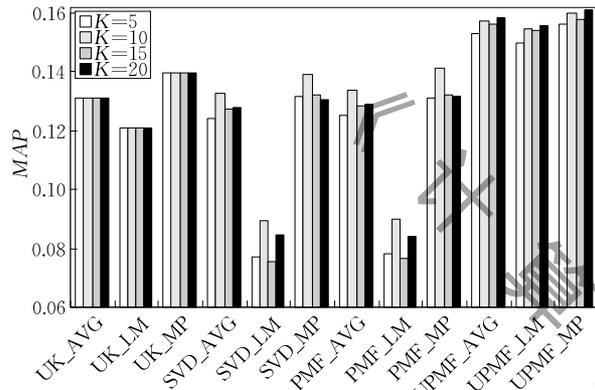


图 12 不同维度  $K$  下的 MAP

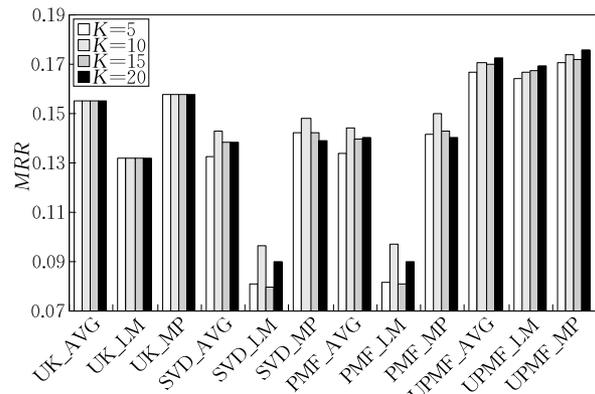


图 13 不同维度  $K$  下的 MRR

了矩阵分解的结果. 相比之下, 本文所提出的基于 UPMF 的群推荐方法, 在三种方法的四个评价指标下, 不同特征维度  $K$  之间的实验结果仅有很小的差异, 说明了 UPMF 方法较 SVD 与 PMF 具有更好的稳定性. 最后, 综合考虑以上分析, 本文的后续实验在  $K=10$  情况下进行.

(3) 社会化参数  $\lambda_s$

本文方法中的社会化参数  $\lambda_s$  可以表示用户相关性在矩阵分解中参与指数. 将  $\lambda_s$  的取值定为  $\lambda_s =$

$\{0.001, 0.01, 0.1, 1, 10, 20, 30, 40, 50\}$ . 具体实验结果如图 14~图 17 所示.

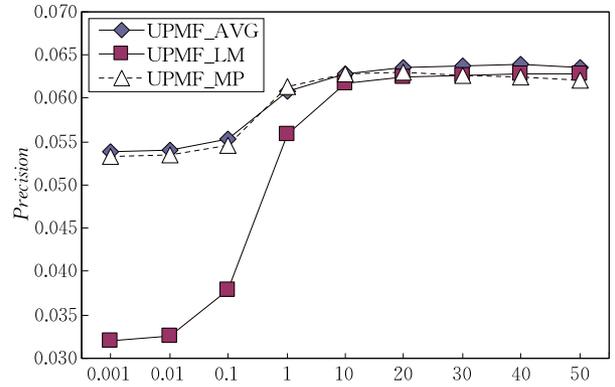


图 14 不同参数  $\lambda_s$  下的 Precision

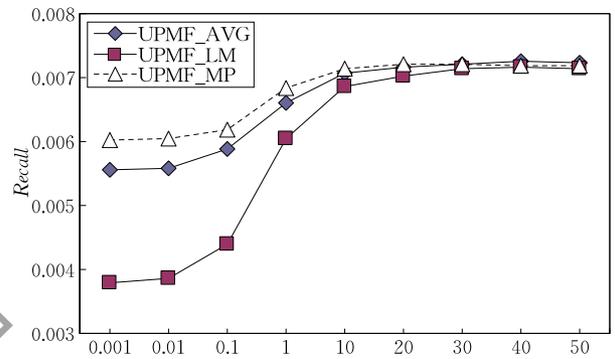


图 15 不同参数  $\lambda_s$  下的 Recall

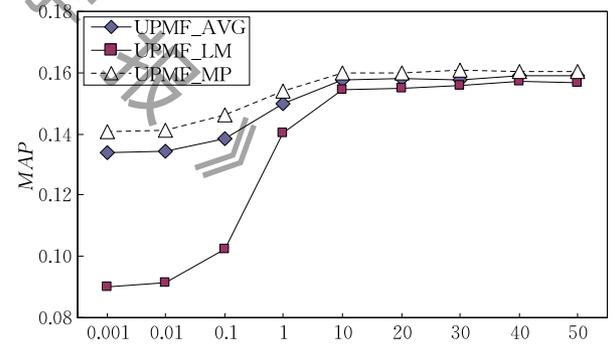


图 16 不同参数  $\lambda_s$  下的 MAP

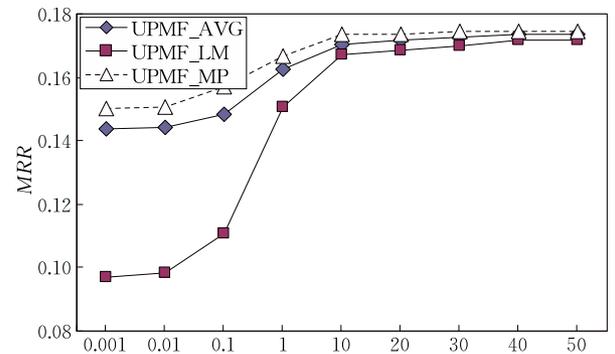


图 17 不同参数  $\lambda_s$  下的 MRR

由图 14~图 17 观察可得, 参数  $\lambda_s$  对方法产生了较为显著的影响. 而不同的参数  $\lambda_s$  对应的实验结果之间的变化不尽相同, 当  $\lambda_s$  取 0.001 和 0.1 时, 本文所提的三种方法在四项评价指标下均有很小的变化, 而随着  $\lambda_s$  不断增加至 10, 三种方法的结果值先迅速增加, 接着基本在  $\lambda_s$  为 10 时达到最大, 当  $\lambda_s$  的取值继续增加至大于 10 以后, 三种方法的结果基本保持不变. 该现象表明了基于群组信息的用户相关性的有效性, 其有关参数的设置会直接影响到群推荐的质量. 另外, 从以上分析可以看出, 当  $\lambda_s$  值为 10 时, 可以获得更好的推荐效果.

## 7 总结和工作展望

本研究提出了一种考虑群组信息的基于联合概率矩阵分解的群推荐方法. 首先, 考虑群组信息衡量用户相关性并融入联合概率矩阵分解方法, 对个人评分进行预测, 其次, 在为群组合成用户评分阶段, 选用群推荐中较常使用的三种融合方法进行合成, 最终为每个群组生成群推荐列表. 通过实验可以表明, 本文所提方法在多个评价指标下的表现均优于其他方法, 进一步验证了本文方法的有效性.

尽管本研究从联合概率矩阵分解的角度出发研究群推荐问题并取得了较好的结果, 但在未来的研究工作中仍需进行更深层次的探索. 一方面, 本文在进行群推荐时考虑了用户评分信息和群组结构信息, 而当前社交网络中还广泛存在众多其它社会化信息, 同样对群推荐有着一定的作用, 如何将多种社会化信息融合到群推荐问题中值得研究. 另一方面, 新的群推荐融合策略也值得研究者做出进一步探索.

## 参 考 文 献

- [1] Adomavicius G, Tuzhilin A. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge & Data Engineering*, 2005, 17(6): 734-749
- [2] Li Y M, Hsiao H W. Recommender service for social network based applications//*Proceedings of the 11th International Conference on Electronic Commerce*. Taipei, China, 2009: 378-381
- [3] Zanda A, Eibe S, Menasalvas E. SOMAR: A social mobile activity recommender. *Expert Systems with Applications*, 2012, 39(9): 8423-8429
- [4] Symeonidis P, et al. Geo-social recommendations based on incremental tensor reduction and local path traversal//*Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks*. Chicago, USA, 2011: 89-96
- [5] Tang J, Hu X, Liu H. Social recommendation: A review. *Social Network Analysis and Mining*, 2013, 3(4): 1113-1133
- [6] He Jun. Review of foreign aggregation strategy of group recommendation. *Library and Information Service*, 2013, 57(7): 127-133+88(in Chinese)  
(何军. 国外群推荐聚集策略研究综述. *图书情报工作*, 2013, 57(7): 127-133+88)
- [7] Ortega F, et al. Incorporating group recommendations to recommender systems: Alternatives and performance. *Information Processing & Management*, 2013, 49(4): 895-901
- [8] Kim J K, et al. A group recommendation system for online communities. *International Journal of Information Management*, 2010, 30(3): 212-219
- [9] Kassak O, Kompan M, Bielikova M. Personalized hybrid recommendation for group of users: Top-N multimedia recommender. *Information Processing & Management*, 2016, 52(3): 459-477
- [10] Sun Guang-Fu, et al. Recommendations based on collaborative filtering by exploiting sequential behaviors. *Journal of Software*, 2013, 24(11): 2721-2733(in Chinese)  
(孙光福等. 基于时序行为的协同过滤推荐算法. *软件学报*, 2013, 24(11): 2721-2733)
- [11] Schafer J B, et al. *Collaborative Filtering Recommender Systems*. The Adaptive Web. Berlin, Germany: Springer, 2007: 291-324
- [12] Cai Y, et al. Typicality-based collaborative filtering recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 2014, 26(3): 766-779
- [13] Liu Lu, Ren Xiao-Li. Review and perspective of the recommender systems. *Journal of Information System*, 2008, 2(1): 82-90(in Chinese)  
(刘鲁, 任晓丽. 推荐系统研究进展及展望. *信息系统学报*, 2008, 2(1): 82-90)
- [14] Mnih A, Salakhutdinov R R. Probabilistic matrix factorization // *Advances in Neural Information Processing Systems*. Vancouver, Canada, 2008: 1257-1264
- [15] Zhang Yu-Jie, Du Yu-Lu, Meng Xiang-Wu. Research on group recommender systems and their applications. *Chinese Journal of Computers*, 2016, 39(4): 745-764(in Chinese)  
(张玉洁, 杜雨露, 孟祥武. 组推荐系统及其应用研究. *计算机学报*, 2016, 39(4): 745-764)
- [16] Baltrunas L, Makcinskas T, Ricci F. Group recommendations with rank aggregation and collaborative filtering//*Proceedings of the 4th ACM Conference on Recommender Systems*. Barcelona, Spain, 2010: 119-126

- [17] Garcia I, Sebastia L, Onaindia E. On the design of individual and group recommender systems for tourism. *Expert Systems with Applications*, 2011, 38(6): 7683-7692
- [18] Qin D, Zhou X, Chen L, et al. Dynamic connection-based social group recommendation. *IEEE Transactions on Knowledge and Data Engineering*. DOI: 10.1109/TKDE.2018.2879658(in Press)
- [19] De Pessemier T, Doooms S, Martens L. Comparison of group recommendation algorithms. *Multimedia Tools and Applications*, 2014, 72(3): 2497-2541
- [20] Masthoff J. *Group Recommender Systems: Combining Individual Models*. *Recommender Systems Handbook*. Boston, USA: Springer, 2011: 677-702
- [21] Quijano-Sanchez L, et al. Social factors in group recommender systems. *ACM Transactions on Intelligent Systems & Technology*, 2013, 4(1): 1199-1221
- [22] Wang W, Zhang G Q, Lu J. Member contribution-based group recommender system. *Decision Support Systems*, 2016, 87: 80-93
- [23] Ardissono L, et al. Intrigue: Personalized recommendation of tourist attractions for desktop and hand held devices. *Applied Artificial Intelligence*, 2003, 17(8-9): 37-41
- [24] Jameson A. More than the sum of its members: Challenges for group recommender systems//*Proceedings of the Working Conference on Advanced Visual Interfaces*. Gallipoli, Italy, 2004: 48-54
- [25] Gorla J, et al. Probabilistic group recommendation via information matching//*Proceedings of the 22nd International Conference on World Wide Web*. Rio de Janeiro, Brazil, 2013: 495-504
- [26] Queiroz S R M, et al. Making recommendations for groups using collaborative filtering and fuzzy majority//*Proceedings of the Brazilian Symposium on Artificial Intelligence*. Berlin, Germany, 2002: 248-258
- [27] Guo J, et al. A social influence approach for group user modeling in group recommendation systems. *IEEE Intelligent Systems*, 2016, 31(5): 1
- [28] Ortega F, et al. Recommending items to group of users using matrix factorization based collaborative filtering. *Information Sciences*, 2016, 345: 313-324
- [29] Liu X, et al. Exploring personal impact for group recommendation//*Proceedings of the 21st ACM International Conference on Information and Knowledge Management*. Hawaii, USA, 2012: 674-683
- [30] Yuan Q, Cong G, Lin C-Y. COM: A generative model for group recommendation//*Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, USA, 2014: 163-172



**WANG Gang**, born in 1980, Ph. D., professor. His research interests include business intelligence and business analytics.

**JIANG Jun**, born in 1991, M. S. candidate. His research interests include data mining and recommender system.

**WANG Han-Ru**, born in 1994, M. S. candidate. Her research interests include data mining and recommender system.

**YANG Shan-Lin**, born in 1948, professor, Ph. D. supervisor, Academician of Chinese Academy of Engineering. His research interests include intelligent decision and information system.

## Background

Group recommendation has been an active area of research within the field of recommendation system. People in virtual communities tend to make up a group to participate in some activities together. These groups of people are in need of decision support. As a result, group recommendation came into being. Group recommendation systems are different from individual recommendation systems. The group recommendation aggregates individual prediction into a group prediction, and need to consider how to match with individual's tastes well.

Recently, some remarkable group recommendation systems have been developed, such as Polylens recommends movies, MusicFX selects a radio station among 91 stations. Providing personal recommendations to group users is becoming a new challenge in the recommended system area. However, there have been relatively a few number of studies on group recommendation systems so far. Most current

recommendation methods are improved by collaborative filtering based on memory, and ignored the relationship among group members by considered that members of group are independent of each other.

In our work, we proposed a group recommendation based on probabilistic matrix factorization, better to model the group recommended problem. The method combines user correlation matrix into probabilistic matrix factorization, and fuse the personal prediction by synthetic strategy commonly used in group-oriented recommendation. The experimental results show that the proposed method in this paper has achieved better results at many evaluation indicators.

This work is partially supported by grants from the National Natural Science Foundation of China (71471054, 91646111), and the Anhui Provincial Natural Science Foundation (1608085MG150).