

一种基于社交网络社区的组推荐框架

刘宇* 吴斌 曾雪琳 张云雷 王柏

(北京邮电大学智能通信软件与多媒体北京市重点实验室 北京 100876)

摘要: 面向用户群组的推荐主要面临如何有意义地对群组进行定义并识别, 以及向群组内用户进行有效推荐两大问题。该文针对已有研究在用户群组定义解释性不强等存在的问题, 提出一种基于社交网络社区的组推荐框架。该框架利用社交网络结构信息发现重叠网络社区结构作为用户群组, 具有较强的可解释性, 并根据用户与群组间的隶属度制定了考虑用户对群组贡献与用户从群组获利的4种聚合与分配策略, 以完成组推荐任务。通过在公开数据集上与已有方法的对比实验, 验证了该框架在组推荐方面的有效性和准确性。

关键词: 社交网络; 组推荐; 重叠网络社区; 非负矩阵分解

中图分类号: TP391; TP393

文献标识码: A

文章编号: 1009-5896(2016)09-2150-08

DOI: 10.11999/JEIT160544

A Group Recommendation Framework Based on Social Network Community

LIU Yu WU Bin ZENG Xuelin ZHANG Yunlei WANG Bai

(Beijing Key Laboratory of Intelligence Telecommunication Software and Multimedia, Beijing University of Posts and Telecommunications, Beijing 100876, China)

Abstract: Group recommendation confronts two major problems, *i.e.*, unambiguous definition and identification of groups and efficient recommendation to users in groups. To tackle the two problems, a group recommendation framework based on social network community is proposed. The framework takes into account social network structural information to identify overlapping groups, which is well interpreted; and fulfills the task of recommending to groups by performing aggregation and allocation strategies using the membership of users related to groups, which considers how much users contribute to groups and benefit from groups. Experimental results on publicly open datasets demonstrate its efficiency and accuracy on the task of group recommendation.

Key words: Social network; Group recommendation; Overlapping network community; Non-negative matrix factorization

1 引言

随着 Web 2.0 技术的快速发展, 用户接入互联网后更倾向于在诸如 Epinions, Ciao, 豆瓣, 大众点评等针对商品、服务的在线评价网站寻求购买意见和建议, 以帮助他们选择适合的商品、服务等项目(如电影、音乐、餐厅等, 下文的“项目”均指商品或服务)。根据用户以往对项目的评分和评价信息, 针对个人的个性化推荐系统能够为用户提供他们所感兴趣的项目。在现实世界中, 个人因不同原因聚集而形成群组。尽管有大量面向个人的个性化推荐

研究, 但针对群体进行推荐的研究还较少。此外, 针对用户群组的组推荐也有其实际意义^[1]。根据研究^[2,3], 组推荐主要面临两大问题: (1)如何定义“群组”的概念并确定群组; (2)如何根据每个用户的个人偏好进行适当的聚合而完成向用户群组推荐。

第1个问题可理解为对用户的“分组策略”, 现有方案大致可分为3种: (1)根据用户的角色身份, 形成固定用户群组^[2,4,5], 如朋友、家人等, 或组成临时用户群组^[6,7], 如在同一固定时间正在健身房运动的顾客等。(2)随机形成的用户群组^[8,9], 如收听音乐直播的人, 成员可随意加入或退出群组。(3)根据一定规则确立的群组^[10,11], 如根据用户对项目喜好形成的兴趣群组。确定群组方法的不同, 影响着对用户组进行推荐的有效程度。当前研究中的大多数分组策略, 都或多或少存在着可解释性差的问题。

第2个问题可理解为对群组内用户进行“推荐

收稿日期: 2016-05-27; 改回日期: 2016-07-18; 网络出版: 2016-08-09

*通信作者: 刘宇 liuyu@bupt.edu.cn

基金项目: 国家重点基础研究发展计划(2013CB329606), 北京市共建项目专项

Foundation Items: The National Key Basic Research and Department Program of China (2013CB329606), Special Fund for Beijing Common Construction Project

决策”。基于用户偏好，即用户对项目的评分，针对个人的个性化推荐方法有基于内容的^[12]，基于协同过滤的^[13,14]，以及利用社交信息改进推荐效果的^[15,16]等。而在组推荐中，同一群组中不同用户对同一项目可能有不同的偏好^[2]。因此，需要通过一种映射方法，将多个用户的偏好信息映射为用户群组偏好，即制定对群组内用户偏好的“聚合策略”。常用的聚合策略有平均策略、随机策略、最小痛苦策略等^[1]。此外，如前所述，不同的分组策略也会导致不同的聚合策略。

针对群组推荐中这两个问题，本文提出了一种基于社交网络社区的组推荐框架。框架将利用社交网络结构信息发现的重叠网络社区确认为用户群组，并根据每个用户在不同社区中的贡献与获利程度提出4种策略，最后由聚合策略和分配策略给出面向群组的推荐结果。

当前，互联网中很多在线评价网站都允许用户间建立关系，如“关注”或“信任”。基于用户关系，一些针对个人的个性化推荐系统利用社交信息达到较好的推荐效果^[15,16]。根据用户关系网络，连接紧密的用户自然而然地聚集在一起形成用户组，不同用户组间连接稀疏，从而形成社区结构^[17]。这种由社交网络关系形成的社区结构，是一种天然的分组方式，具有很好的可解释性。因此，本文提出的采用社交网络结构信息发现的重叠网络社区确认为用户群组的方法，具备较好的可解释性；结合重叠分组的特性，可以为组推荐提供多种具有合理解释的聚合及分配策略。实验表明，本文提出的框架在用户关系信息稠密的情况下，能达到更好的组推荐效果。本文第2节对组推荐问题进行定义；第3节提出一种基于社交网络社区的组推荐框架；第4节利用公开数据集进行了实验并对实验结果进行分析和讨论；第5节进行总结和工作展望。

2 组推荐问题定义

在组推荐中，推荐系统需要完成两方面的任务，即有意义地完成用户分组并向不同用户群组有效地推荐项目。以往的组推荐研究仅利用了用户对项目的评分去完成这两个任务。值得注意的是，众多在线评价网站引入了社交网络模块，允许用户间建立关系。根据网络社区理论^[17]，网络社区是用户在社交网络中天然形成的群组，其中也蕴含着用户与用户间、用户与社区间、社区与社区间的深层次互联关系。

因此，在基于社交网络的组推荐中，可利用的信息有用户对项目的评分和用户社交关系网络。以

m 和 n 表示用户数和项目数。用户对项目的评分通常以评分矩阵 $\mathbf{R} = [r_{ij}]^{m \times n}$ 表示，其中 r_{ij} 是用户 p_i 对项目 q_j 的评分， $r_{ij} = 0$ 表示缺失数据。用户间关系网络通常以邻接矩阵 $\mathbf{G} = [g_{ij}]^{m \times m}$ 表示，其中 g_{ij} 代表用户 u_i 和 u_j 的关系。若用户 u_i 和 u_j 是朋友关系，定义 $g_{ij} = 1$ ，否则 $g_{ij} = 0$ 。为简化问题，本文仅考虑了对称的用户关系网络。图1展示了一个本文所研究的基于社交网络的组推荐示例的用户关系网络，其具有5个用户；表1是其关系网络的邻接矩阵，表2是他们对5个项目的评分。这5个用户形成2个重叠分组，分别是{用户1，用户2，用户3}和{用户2，用户4，用户5}，用户2同时归属于两个社区分组。

本文所研究的组推荐问题即：根据评分矩阵 \mathbf{R} 获得用户偏好；根据邻接矩阵 \mathbf{G} ，获得基于社交网络社区的用户分组 \mathbf{U} ；根据分组 \mathbf{U} 中用户对所属社区的贡献和获利程度，提出基于关系网络结构信息的聚合策略和分配策略；根据聚合和分配策略，由用户偏好完成对群组的推荐，产生预测的评分矩阵 $\hat{\mathbf{R}}$ 。

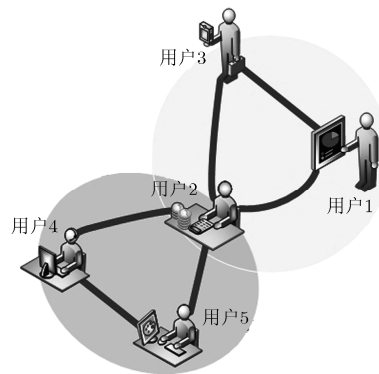


图1 基于社交网络的组推荐示例的用户关系网络

表1 图1中用户关系网络的邻接矩阵

	用户1	用户2	用户3	用户4	用户5
用户1	0	1	1	0	0
用户2	1	0	1	1	1
用户3	1	1	0	0	0
用户4	0	1	0	0	1
用户5	0	1	0	1	0

表2 图1中用户对项目的评分矩阵(采用了5分制)

	项目1	项目2	项目3	项目4	项目5
用户1	3	0	4	0	0
用户2	3	0	4	0	4
用户3	2	2	5	0	3
用户4	0	4	3	3	4
用户5	3	0	4	3	3

3 基于社交网络社区的组推荐框架

针对上一节提出的组推荐问题, 本文提出了一种基于社交网络社区的组推荐框架, 包含以下3个部分: 基于对评分矩阵分解的用户偏好获取, 基于重叠网络社区发现的用户分组, 以及基于社区结构的聚合和分配策略。其中前两个部分均采用了非负矩阵分解作为基本方法, 以将两部分进行整合。

3.1 基于对评分矩阵分解的用户偏好获取

用户偏好直接体现在用户对项目的评分上。但在真实数据中, 评分矩阵存在大量的“零值”, 即缺失了大量的评分数据, 具有很高的稀疏性^[18]。因此, 采用评分矩阵作为用户偏好的推荐结果效果并不好^[10]。流行的推荐系统通常采用低秩矩阵分解(low-rank matrix factorization)将评分矩阵 R 分解为两个低秩矩阵 P 和 Q , 通过这两个矩阵的乘积 \hat{R} 作为评分矩阵的近似, 以对缺失值进行估计^[19]。

值得注意的是, 矩阵 P 和 Q 又分别代表用户隐因子矩阵和项目隐因子矩阵, 即将用户的评分在隐含的因素上予以表示。与文献[10]中方法类似, 本文采用矩阵分解的方法获得用户偏好, 即以矩阵 P 的行向量(隐式偏好)来表示。为结合采用基于非负矩阵分解的重叠网络社区发现来进行用户分组的方法, 本文使用非负矩阵分解(non-negative matrix factorization)得到分解矩阵, 即可在最小化以下目标函数时得到

$$l = \frac{1}{2} \|R - PQ^T\|_F^2 + \frac{\lambda_P}{2} \|P\|_F^2 + \frac{\lambda_Q}{2} \|Q\|_F^2, \quad (1)$$

s.t. $P \geq 0, Q \geq 0$

其中, $\|\cdot\|_F$ 是 Frobenius 范数, 矩阵 P 和 Q 分别为 $m \times f$ 和 $n \times f$ 的用户隐因子矩阵和项目隐因子矩阵, f 是隐因子的数量, λ_P 和 λ_Q 是防止过拟合的参数。式(1)中的目标函数可以通过乘法更新规则迭代得到。根据文献[20]中的方法, 得到更新矩阵 P 和 Q 的规则:

$$P_{ik} \leftarrow P_{ik} \sqrt{\frac{[RQ]_{ik}}{[PQ^T + \lambda_P P]_{ik}}} \quad (2)$$

$$Q_{jk} \leftarrow Q_{jk} \sqrt{\frac{[R^T P]_{jk}}{[QP^T + \lambda_Q Q]_{jk}}} \quad (3)$$

其中, $[\cdot]_{ij}$ 和 \bullet_{ij} 表示相应矩阵的第 i 行, 第 j 列的元素。式(2)和式(3)的正确性和收敛性可以通过满足 KKT 条件来进行证明, 证明过程与文献[20]中所述相似, 本文不再赘述。基于对评分矩阵分解的用户偏好获取算法如表3所示。在算法中, 更新矩阵 P 时保持矩阵 Q 不变, 更新矩阵 Q 时保持矩阵 P 不变。

表3 基于对评分矩阵分解的用户偏好获取算法

算法 getUserPref()	
输入:	用户评分矩阵 R , 隐因子数量 f , 规则化参数 λ_P 和 λ_Q
输出:	用户隐式偏好矩阵 P , 项目隐因子矩阵 Q
1	将矩阵 P 和 Q 以非负随机数初始化
2	while 未收敛 do
3	根据式(2)更新矩阵 P
4	根据式(3)更新矩阵 Q
5	end while
6	return $[P, Q]$

3.2 基于重叠网络社区发现的用户分组

在可通用化的用户定义分组研究中, 文献[10]采用 k -means 将用户偏好进行聚类而得到用户分组; 虽然这种分组可解释为基于用户兴趣的社团, 但具有一定的模糊性。文献[21]采用由用户相似性生成的网络结构发现兴趣群组; 但这种方法未考虑社交网络中真实存在的用户关系网络, 因此其得到的分组意义仍不明确。如前所述, 用户在社交网络中通过用户关系天然聚集形成社区, 即分组。这种社区结构可以认为是一种具有良好解释意义的分组。因此, 本文所研究的基于社交网络的组推荐, 将社交网络社区发现作为分组策略。

与基于对评分矩阵分解的用户偏好获取目的相似, 本文采用了非负矩阵分解来发现重叠社区结构, 这样, 非负矩阵分解技术有机地将本文提出的组推荐框架整合起来。这种方法发现的分组具有重叠结构。即某个用户可以同时归属于多个分组, 这也与现实情况相同, 即某个用户拥有多种兴趣或属于多个真实群体。

研究表明, 非负矩阵分解在重叠社区发现任务上具有很好的准确性^[22]和良好的解释性^[23]。给定用户关系邻接矩阵 G , 通过最小化以下目标函数进行矩阵分解, 可以得到一个网络划分(分组)结果。

$$l = \frac{1}{2} \|G - UU^T\|_F^2, \quad \text{s.t. } U \geq 0 \quad (4)$$

其中, 矩阵 U 是 $m \times d$ 的网络分组隶属度矩阵, d 是分组数量。与式(4)不同, 本文加入了可调节所发现网络社区重叠程度的参数, 修改后的目标函数为

$$l = \frac{1}{2} \|G - UU^T\|_F^2 + \frac{\lambda_{UU}}{2} \|U^T U - I\|_F^2, \quad (5)$$

s.t. $U \geq 0$

其中, I 是单位矩阵。式(5)中第2个规则化项用来控制所发现分组的重叠程度。对于重叠分组划分来说, U 的内积即 $U^T U$ 是对角占优矩阵; 对于没有重叠结构的划分来说, 该内积则是对角矩阵。引入参数 λ_{UU} , 当 $\lambda_{UU} = 0$ 时, 使社区分组结果是最大程

度重叠的；当 $\lambda_{UV} > 0$ 且增大时，使 $U^T U$ 更逼近对角矩阵，划分的结果逐渐变得不重叠。与 3.1 节类似，式(5)的更新规则见式(6)。表 4 列出了用户分组算法。

$$U_{ik} \leftarrow U_{ik} \sqrt{\frac{[GU + \lambda_{UV}U]_{ik}}{[UU^T U + \lambda_{UV}UU^T U]_{ik}}} \quad (6)$$

表 4 基于重叠网络社区发现的用户分组算法

算法 partGroup()	
输入：	用户关系网络邻接矩阵 G ，社区分组数量 d ，规则化参数 λ_{UV}
输出：	用户网络分组隶属度矩阵 U
1	将矩阵 U 以非负随机数初始化
2	while 未收敛 do
3	根据式(6)更新矩阵 U
4	end while
5	return U

3.3 基于社区结构的聚合和分配策略

已有研究中通常仅对用户偏好进行聚合生成群组推荐结果，如采用平均策略、随机策略等^[11]，如文献[10]验证了平均策略在组推荐中的效果。基于其工作，本文提出了一些用于组推荐的聚合策略。

一般来说，社交网络用户在群组决策中具有不同的影响力。不仅不同用户在同一群组中的影响力不同，同一用户在不同群组中的影响程度也是不同的。并且，在考虑对用户偏好进行聚合的同时，同样应考虑聚合结果对用户选择的最终影响。例如，一组基于朋友关系的用户，经讨论决定选择一部电影。其中，有的用户是大多数人的朋友，对选择结果影响很大，而有的用户仅是少部分人的朋友，对最终选择的影响很小。因此，本文提出采用分配策略，对聚合结果进行再分配，从而达到更好的组推荐效果。本文所提出的聚合及分配策略，均结合了重叠网络社区分组信息中每个用户与社区分组间的隶属度。

若行向量 $P_{u_j} = [p_{u_j,1} \ p_{u_j,2} \ \dots \ p_{u_j,k}]$ 表示用户 u_j 的偏好，其中 $p_{u_j,r}$ 表示用户 u_j 在第 r 个方面的偏好或兴趣。以用户隐式偏好为例，用户 u_j 在 k 个隐因子的偏好即为长度为 k 的向量 P_{u_j} 。若行向量 $P_{c_i} = [p_{c_i,1} \ p_{c_i,2} \ \dots \ p_{c_i,k}]$ 表示群组 c_i 的偏好，其中群组 c_i 包含 g 个用户 $\{u_1, u_2, \dots, u_g\}$ 。偏好聚合即通过某个函数 $F(\cdot)$ ，使 $P_{c_i} = F(P_{u_1}, P_{u_2}, \dots, P_{u_g})$ 。若行向量 $\hat{P}_{u_j} = [\hat{p}_{u_j,1} \ \hat{p}_{u_j,2} \ \dots \ \hat{p}_{u_j,k}]$ 表示对用户 u_j 的组推荐预测结果，而用户 u_j 隶属于一组社区 $\{c_1, c_2, \dots, c_s\}$ ，则偏好

分配即通过某个函数 $F'(\cdot)$ ，使 $P_{u_j} = F'(P_{c_1}, P_{c_2}, \dots, P_{c_s})$ 。

采用非负矩阵分解作为用户分组方法的一个重要原因是，如式(4)所示，矩阵 U 可以表示用户在重叠分组上的隶属度：其列向量表示每个用户对某分组贡献的程度，行向量表示某个用户从重叠分组获利的程度。这种隶属度可衡量一个用户在其所属群组中的“参与强度”（文献[23]中称为“participation strength”），可视作用户对分组的贡献和从分组的获利。从重叠分组隶属度的角度，本文给出了一组聚合与分配策略：

分组平均策略 在聚合时，对分组内每个用户的偏好进行平均；在分配时，将某用户所属的重叠群组偏好平均分配给该用户。这种策略，平均看待每个分组内的用户对群组偏好的贡献，并将每个用户所属的重叠分组的偏好平均地分配给该用户。

分组贡献策略 在聚合时，考虑分组内每个用户不同的贡献程度；在分配时，将某用户所属的重叠群组偏好平均分配给该用户。由于每个用户属于每个分组的隶属度不同，“分组贡献策略”以其作为权重，将每个分组内的用户偏好进行加权平均后作为该分组的群组偏好。

分组获利策略 在聚合时，对分组内每个用户的偏好进行平均；在分配时，考虑用户从一个或多个不同分组的获利程度。与“分组贡献策略”类似，“分组获利策略”将每个用户所属的重叠分组的偏好进行加权平均后分配给该用户。

分组权重策略 在聚合时，考虑分组内每个用户不同的贡献程度；在分配时，考虑用户从一个或多个不同分组的获利程度。

3.4 基于社交网络社区的组推荐框架

基于前述 3 个部分，即基于对评分矩阵分解的用户偏好获取，基于重叠网络社区发现的用户分组，以及基于社区结构的聚合和分配策略，本文提出的基于社交网络社区的组推荐框架可通过表 5 所示的算法表示，下文以 SocoGrec 表示该框架。此外，本文所提出的框架也可以整合更多的聚合和分配策略。

4 实验与结论

4.1 数据集

本文选取了可公开获取的数据集 FilmTrust 和 CiaoDVD^[24]，均含有用户关系网络 and 用户评分数据：

FilmTrust 是一个允许用户对电影进行评分，并可以与其他用户建立信任关系的在线网站数据。用

表 5 基于社交网络社区的组推荐框架

算法 SocoGrec()	
输入:	用户评分矩阵 \mathbf{R} , 关系网络邻接矩阵 \mathbf{G} , 隐因子数量 f , 社区分组数量 d , 正则化参数 λ_P, λ_Q 和 λ_{UU} , 策略 method
输出:	预测评分矩阵 $\hat{\mathbf{R}}$
1	$\mathbf{U} \leftarrow \text{partGroup}(\mathbf{G}, d, \lambda_{UU})$
2	$[\mathbf{P}, \mathbf{Q}] \leftarrow \text{getUserPref}(\mathbf{R}, f, \lambda_P, \lambda_Q)$
3	初始化贡献矩阵 \mathbf{U}_C 和获利矩阵 \mathbf{U}_B , 矩阵大小与 \mathbf{U} 相同
4	switch (method)
5	case 分组平均策略:
6	$\mathbf{U}_{Cij} \leftarrow 1$, 如果 $\mathbf{U}_{ij} > 0$; 0, 其他。将 \mathbf{U}_C 按列标准化
7	$\mathbf{U}_{Bij} \leftarrow 1$, 如果 $\mathbf{U}_{ij} > 0$; 0, 其他。将 \mathbf{U}_B 按行标准化
8	case 分组贡献策略:
9	将 \mathbf{U}_C 按列标准化
10	$\mathbf{U}_{Bij} \leftarrow 1$, 如果 $\mathbf{U}_{ij} > 0$; 0, 其他。将 \mathbf{U}_B 按行标准化
11	case 分组获利策略:
12	$\mathbf{U}_{Cij} \leftarrow 1$, 如果 $\mathbf{U}_{ij} > 0$; 0, 其他。将 \mathbf{U}_C 按列标准化
13	将 \mathbf{U}_B 按行标准化
14	case 分组权重策略:
15	将 \mathbf{U}_C 按列标准化
16	将 \mathbf{U}_B 按行标准化
17	end switch
18	$\hat{\mathbf{R}} \leftarrow \mathbf{U}_B \mathbf{U}_C^T \mathbf{P} \mathbf{Q}^T$
19	return $\hat{\mathbf{R}}$

户间的信任关系形成了社会信任网络, 表示了用户与用户的关系。

CiaoDVD。Ciao 是一个在线商品评价网站, 也引入了社会信任关系网络。其中, CiaoDVD 是包含在“DVD”分类下的用户评分, 数据集中包含了用户信任关系。

为了较为准确地发现社区分组, 实验选取了公开数据集的子集。即所用数据集的用户既拥有评分信息, 也拥有社交网络关系; 没有任何评分或没有任何社交关系的用户则被排除。数据集特征如表 6 所示。

表 6 数据集特征

	FilmTrust	CiaoDVD
用户数量	427	733
项目数量	1827	10305
评分数量	11848	19621
评分矩阵稀疏度 (%)	98.48	99.74
用户关系数量	807	23582
用户关系网络密度	0.0089	0.0879

4.2 对比方法及实验设置

LGM 方法^[10]通过对用户隐式偏好采用 k -means 方法进行聚类得到隐式分组, 并将群组内所有用户偏好进行平均作为群组偏好。其应用场景与本文类似, 且都利用了用户对项目的评分矩阵来提取信息, 如用户偏好。因此, 实验中将 LGM 方法与本文所提框架中的 4 种策略方法进行比较。

对比实验采用了 5-折交叉验证方法, 被随机平均分为 5 份的数据中, 4 份作为训练集, 1 份作为测试集。由于隐因子数量一般不会对推荐效果产生较大影响, 因此在实验中, 设置隐因子的数量为 $f = 10$ 。以最小取 1, 最大取数据集内用户个数, FilmTrust 数据集的分组数量分别设定为 {1, 5, 10, 20, 100, 200, 427}, CiaoDVD 数据集的分组数量设定为 {1, 10, 20, 100, 200, 500, 733}。对于 SocoGrec, 设定 $\lambda_{UU} = 0$, 以保证分组发现结果是以最大程度重叠的。简单起见, 防止过拟合的规则化项参数 λ_P 和 λ_Q 均设置为 0.01。在实验中, 将本文提出的 4 种策略分别表示为 SocoGrec-M, SocoGrec-C, SocoGrec-B 和 SocoGrec-W。

4.3 评价指标

本文选取了绝对平均误差 (Mean Absolute Error, MAE) 作为评价指标, 其定义如下:

$$\text{MAE} = \frac{1}{N} \sum_{i,j} |r_{ij} - \hat{r}_{g_{ij}}| \quad (7)$$

其中, r_{ij} 表示了原始数据集中所观测到的用户 i 对项目 j 的评分, $\hat{r}_{g_{ij}}$ 是用户 i 在组推荐系统中给予项目 j 的预测评分, N 是作为测试集的评分数量。由其定义可知, 指标 MAE 越小, 则所预测的推荐结果越准确。

4.4 实验结果与讨论

4.4.1 在数据集 FilmTrust 上的对比结果 在数据集 FilmTrust 上的对比结果如表 7 所示。从结果可以看出 SocoGrec 所提出的 4 种策略方法的评价指标 MAE 随着分组数量的增加而降低。SocoGrec-B 和 SocoGrec-W 的效果总是比 SocoGrec-M 和 SocoGrec-C 要好。SocoGrec-B 总是比 SocoGrec-C 效果好, 在一些分组数量时比 SocoGrec-W 效果好。也就是说, 在采用平均的聚合策略情况下, 考虑用户从一个或多个分组中的获利程度更能提高推荐预测的效果; 与 SocoGrec-B 相比, 同时考虑了用户对分组贡献和从分组获利的 SocoGrec-W 效果与其相近, 也说明了在数据集 FilmTrust 上, 与带权重的聚合策略相比, 采用带权重的分配策略使推荐预测效果提升更多。对比方法 LGM 的评价指标 MAE 则随着分组数量的增加而上升, 直到其效果变得比

表 7 在数据集 FilmTrust 上组推荐算法的 MAE 对比

用户社区分组数量	SocoGrec-M	SocoGrec-C	SocoGrec-B	SocoGrec-W	LGM
1	1.52674	1.52593	1.32401	1.32267	0.95536
5	1.41679	1.42074	1.32721	1.33180	1.00121
10	1.38919	1.39289	1.32203	1.32510	1.02864
20	1.32006	1.32169	1.31516	1.31347	1.03851
100	1.16278	1.16393	1.13223	1.13100	1.06932
200	1.05752	1.07085	1.04416	1.05344	1.05765
427	1.01865	1.02086	1.01973	1.01558	1.06085

SocoGrec 更差。值得注意的是，当分组数量为 1 时，LGM 仅对数据集中所有用户的偏好进行平均，而 SocoGrec 仍然考虑了用户与群组间的贡献与获利程度，具有较好的解释性；尽管此时 SocoGrec 的评价指标并不比 LGM 好。

4.4.2 在数据集 CiaoDVD 上的对比结果 在数据集 CiaoDVD 上的对比结果如表 8 所示。与在 FilmTrust 数据集上的结果类似，随着分组数量的增加，SocoGrec 的预测效果变好，而 LGM 的预测效果则变差。然而与之不同的是，在 CiaoDVD 数据集上，SocoGrec 提出的 4 种策略的 MAE 指标总是优于 LGM 方法；随着分组数量的增加，策略 SocoGrec-M 和 SocoGrec-B 改善较大，而 SocoGrec-C 改善较小，SocoGrec-W 的效果几乎不变，而且 SocoGrec-C 总是优于 SocoGrec-B。即在 CiaoDVD 数据集上，与带权重的分配策略相比，采用带权重的聚合策略使推荐预测效果提升更多。在两种数据集上产生不同效果的原因可能在于，尽管两个数据集的评价稀疏度相近，但用户关系网络稠密程度相差了一个数量级(参考表 6)。这也说明了用户关系网络的稠密程度对几种 SocoGrec 策略的影响是不同的，关系网络更稠密，本文所提出的组推荐框架能够获得更好的效果。

4.4.3 在数据集 FilmTrust 上重叠分组的影响 值得注意的是，由于分组可以重叠，本文所提出的框架允许分组数量多于用户数量。从图 2 中可以看出，在数据集 FilmTrust 上，重叠分组对组推荐影响是较大的。首先，一旦分组数量增加，对比方法 LGM

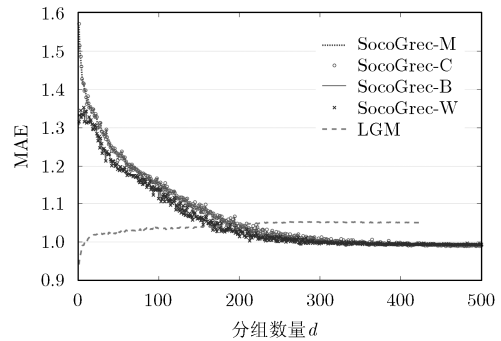


图 2 不同社区分组数量在数据集 FilmTrust 上的算法的 MAE 对比

的效果迅速变差；SocoGrec 策略则逐渐变好，分组数量升至约 170 至 190 后，其效果比 LGM 方法好。其次，SocoGrec-M 与 SocoGrec-C 的效果相近，SocoGrec-B 与 SocoGrec-W 效果相近，再次印证了前述对几种策略效果的讨论。最后，基于非负矩阵分解，本文所提出的框架可以发现重叠分组，具有一定的现实意义。

5 结束语

本文提出了一种基于社交网络社区的组推荐框架。框架采用非负矩阵分解方法将用户偏好获取与用户分组有机地整合起来，并利用用户-分组隶属度制定了相应的偏好聚合与分配策略。与现有方法的对比实验展示了该框架的优越性和有效性。此外，该框架在获取用户偏好、确定分组以及聚合与分配策略方面具有良好的解释性。并且，该框架可以发现重叠分组，对分组定义和组推荐都有一定的现实意义。

表 8 在数据集 CiaoDVD 上组推荐算法的 MAE 对比

用户社区分组数量	SocoGrec-M	SocoGrec-C	SocoGrec-B	SocoGrec-W	LGM
1	2.00683	1.99314	1.99868	1.98487	1.99728
10	1.99181	1.98598	1.99109	1.98565	2.02620
20	1.99218	1.98489	1.99234	1.98501	2.03498
100	1.98877	1.98574	1.98948	1.98495	2.06730
200	1.98781	1.98644	1.98711	1.98490	2.06314
500	1.98658	1.98671	1.98714	1.98685	2.07713
733	1.98627	1.98562	1.98730	1.98546	2.07981

基于本文的研究,未来可以针对以下几点进行扩展工作。可将框架适配到有向网络、双向网络,使分组策略、聚合与分配策略可利用更丰富的用户关系信息,进一步改善组推荐效果。其次,可引入其他聚合与分配策略。另外,本文使用乘法更新来进行非负矩阵分解,实际实现中包含大量耗时的矩阵乘法运算,因此可以采用其他算法时间复杂性较低的非负矩阵分解方法或将算法并行化,以提高框架算法运行效率。

参考文献

- [1] MASTHOFF J. Recommender Systems Handbook[M]. Boston, MA: Springer US, ch. Group Recommender Systems: Combining Individual Models, 2011: 677-702.
- [2] JAMESON A and SMYTH B. The Adaptive Web: Methods and Strategies of Web Personalization[M]. Berlin, Heidelberg: Springer Berlin Heidelberg, ch. Recommendation to Groups, 2007: 596-627.
- [3] AMER-YAHIA S, ROY S, CHAWLAT A, et al. Group recommendation: Semantics and efficiency[J]. *Proceedings of the VLDB Endowment*, 2009, 2(1): 754-765. doi: 10.14778/1687627.1687713.
- [4] O'CONNOR M, COSLEY D, KONSTAN J, et al. ECSCW 2001: Proceedings of the Seventh European Conference on Computer Supported Cooperative Work 16-20 September 2001[M]. Bonn, Germany. Dordrecht: Springer Netherlands, ch. PolyLens: A Recommender System for Groups of Users, 2001: 199-218.
- [5] DE CAMPOS L M, FERNANDEZ-LUNA J M, HUETE J F, et al. Group recommending: a methodological approach based on bayesian networks[C]. IEEE 23rd International Conference on Data Engineering Workshop, Istanbul, Turkey, 2007: 835-844.
- [6] O'HARA K, LIPSON M, JANSEN M, et al. Jukola: democratic music choice in a public space[C]. Proceedings of the 5th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques, New York, USA, 2004: 145-154.
- [7] SPRAGUE D, WU F, and TORY M. Music selection using the partyvote democratic jukebox[C]. Proceedings of the Working Conference on Advanced Visual Interfaces, New York, USA, 2008: 433-436.
- [8] CHAO D L, BALTHROP J, and FORREST S. Adaptive radio: achieving consensus using negative preferences[C]. Proceedings of the 2005 International ACM SIGGROUP Conference on Supporting Group Work, New York, USA, 2005: 120-123.
- [9] MCCARTHY J F and ANAGNOST T D. Musicfx: An arbiter of group preferences for computer supported collaborative workouts[C]. Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work, New York, USA, 1998: 363-372.
- [10] SHI J, WU B, and LIN X. A latent group model for group recommendation[C]. 2015 IEEE International Conference on Mobile Services, New York, USA, 2015: 233-238.
- [11] BALTRUNAS L, MAKCINSKAS T, and RICCI F. Group recommendations with rank aggregation and collaborative filtering[C]. Proceedings of the Fourth ACM Conference on Recommender Systems, New York, USA, 2010: 119-126.
- [12] PAZZANI M J and BILLSUS D. The Adaptive Web: Methods and Strategies of Web Personalization[M]. Berlin, Heidelberg: Springer Berlin Heidelberg, ch. Content-Based Recommendation Systems, 2007: 325-341.
- [13] SHI Y, LARSON M, and HANJALIC A. Collaborative filtering beyond the user-item matrix: a survey of the state of the art and future challenges[J]. *ACM Computing Surveys*, 2014, 47(1): 1-45. doi: 10.1145/2556270.
- [14] 王玉斌, 孟祥武, 胡勋. 一种基于信息老化的协同过滤推荐算法[J]. 电子与信息学报, 2013, 35(10): 2391-2396. doi: 10.3724/SP.J.1146.2012.01743.
WANG Y, MENG X, and HU X. Information aging-based collaborative filtering recommendation algorithm[J]. *Journal of Electronics & Information Technology*, 2013, 35(10): 2391-2396. doi: 10.3724/SP.J.1146.2012.01743.
- [15] 邢星. 社交网络个性化推荐方法研究[D]. [博士学位论文], 大连海事大学, 2013.
XING X. Research on recommendation methods in social networks[D]. [Ph.D. dissertation], Dalian Maritime University, 2013.
- [16] 涂丹丹, 舒承椿, 余海燕. 基于联合概率矩阵分解的上下文广告推荐算法[J]. 软件学报, 2013, 24(3): 454-464. doi: 10.3724/SP.J.1001.2013.04238.
TU D, SHU C, and YU H. Using unified probabilistic matrix factorization for contextual advertisement recommendation [J]. *Journal of Software*, 2013, 24(3): 454-464. doi: 10.3724/SP.J.1001.2013.04238.
- [17] GIRVAN M and NEWMAN M E. Community structure in social and biological networks[J]. *Proceedings of the National Academy of Sciences*, 2002, 99(12): 7821-7826. doi: 10.1073/pnas.122653799.
- [18] BORATTO L and CARTA S. Using collaborative filtering to overcome the curse of dimensionality when clustering users in a group recommender system[C]. Proceedings of 16th International Conference on Enterprise Information Systems, Lisbon, Portugal, 2014: 564-572.
- [19] 方耀宁, 郭云飞, 丁雪涛, 等. 一种基于局部结构的改进奇异值分解推荐算法[J]. 电子与信息学报, 2013, 35(6): 1284-1289. doi: 10.3724/SP.J.1146.2012.01299.

- FANG Y, GUO Y, DING X, *et al.* An improved singular value decomposition recommender algorithm based on local structures[J]. *Journal of Electronics & Information Technology*, 2013, 35(6): 1284–1289. doi: 10.3724/SP.J.1146.2012.01299.
- [20] DING C, LI T, and PENG W, *et al.* Orthogonal nonnegative matrix t-factorizations for clustering[C]. Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 2006: 126–135.
- [21] CANTADOR I, and CASTELLS P. Extracting multilayered communities of interest from semantic user profiles: application to group modeling and hybrid recommendations [J]. *Computers in Human Behavior*, 2011, 27(4): 1321–1336. doi: 10.1016/j.chb.2010.07.027.
- [22] SHI X, LU H, HE Y, *et al.* Community detection in social network with pairwise constrained symmetric non-negative matrix factorization[C]. Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, Paris, France, 2015: 541–546.
- [23] PSORAKIS I, ROBERTS S, EDBEN M, *et al.* Overlapping community detection using Bayesian non-negative matrix factorization[J]. *Physical Review E*, 2011, 83(6): 066114. doi: 10.1103/PhysRevE.83.066114.
- [24] LibRec. Datasets[OL]. <http://www.librec.net/datasets.html>, 2016.
- 刘宇：男，1986年生，博士生，研究方向为数据挖掘、复杂网络、社交网络分析。
- 吴斌：男，1969年生，教授，博士生导师，研究方向为复杂网络、社交网络分析、数据挖掘和智能信息处理、大数据分析。
- 曾雪琳：女，1991年生，硕士生，研究方向为数据挖掘、社交网络分析、推荐系统。
- 张云雷：男，1983年生，博士生，研究方向为数据挖掘、复杂网络、社交网络分析。
- 王柏：女，1962年生，教授，博士生导师，研究方向为电信系统软件、分布式计算技术、数据挖掘和智能信息处理。