

# 一种基于二部图和节点角色划分的 社交网络推荐方案

肖云鹏, 刘瀚松, 刘宴兵

(重庆邮电大学网络与信息安全技术重庆市工程实验室, 重庆 400065)

**摘要:** 针对现有社交网络用户推荐方案中大规模网络个体相似性计算复杂度高以及个体节点无差异对待的问题, 本文提出一种基于二部图和节点角色划分的推荐方案. 首先, 通过划分重叠群体简化原生社交网络结构, 并进一步构建群体-个体二部图模型; 其次, 通过群体-个体二部图所反映的拓扑特征, 结合节点自身属性特征, 对个体进行角色划分, 提出一种基于群体-个体二部图的角色划分模型; 最后, 针对大规模网络中计算个体相似性复杂度高的问题, 构建基于角色差异下的个体-个体二部图模型, 实现层次化、个性化的推荐. 实验表明, 该方案适用于对社交网络中兴趣广泛度存在差异的个体间进行好友推荐, 并在较小规模的二部图上生成目标个体推荐列表, 降低了计算个体相似性的复杂度.

**关键词:** 社交网络; 二部图; 角色划分; 个体推荐

**中图分类号:** TP393      **文献标识码:** A      **文章编号:** 0372-2112 (2017)10-2425-09

**电子学报 URL:** <http://www.ejournal.org.cn>      **DOI:** 10.3969/j.issn.0372-2112.2017.10.016

## A Social Network Recommendation Scheme Based on Bipartite Graph and Node Role Division

XIAO Yun-peng, LIU Han-song, LIU Yan-bing

(Chongqing Engineering Laboratory of Internet and Information Security,  
Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

**Abstract:** In the view of the high complexity about similarity calculation and the indifference about individual nodes, a social network recommendation scheme based on bipartite graph and node role division is presented in this study. Firstly, the native social network structure is simplified by dividing overlapping groups. Furthermore, the bipartite graph model of group and individual is given. Secondly, the role division model is proposed by combining topological features of bipartite graph with node attributes. Finally, in order to resolve high computational complexity, the individual bipartite graph model is constructed based on user role difference. The model implements a hierarchical and personalized recommendation. Experiments show that the scheme can effectively recommend among social users who have different interests. In addition, the complexity of individual similarity computation is reduced because the target individual recommendation list is generated based on small scale bipartite graph.

**Key words:** social network; bipartite graph; role division; individual recommendation

### 1 引言

Internet 的迅猛发展使得信息大规模地出现在我们面前, 推荐系统顺势成为解决信息过载的重要技术, 如

国内外的新浪和 Amazon 等都不同程度使用了推荐系统帮助人们更准确、更高效地获取信息.

推荐系统经历了从非个性化到个性化的发展过程<sup>[1]</sup>, 个性化推荐方法被大致分为四类: 基于协同过滤

收稿日期: 2016-06-07; 修回日期: 2017-04-09; 责任编辑: 孙瑶

基金项目: 国家 973 重点基础研究发展计划 (No. 2013CB329606); 国家自然科学基金 (No. 61272400); 重庆市青年人才项目 (No. cstc2013kjrc-qn-rc40004); 教育部-中国移动研究基金 (No. MCM20130351); 重庆市研究生研究与创新项目 (No. CYS14146); 重庆市教委科学计划项目 (No. KJ1500425); 重庆邮电大学文峰基金 (No. WF201403)

的推荐<sup>[2]</sup>、基于内容的推荐<sup>[3]</sup>、基于网络结构的推荐<sup>[4]</sup>和混合推荐<sup>[5]</sup>. 部分研究者们把重心放在如何提高预测评分准确率却在一定程度上忽略了用户兴趣程度因素<sup>[6-8]</sup>, topN 推荐解决了预测评分准确而用户不感兴趣的困境.

研究者常常把网络抽象为图结构, 利用其结构特性分析网络. 例如, Zhou 等人<sup>[9]</sup>提出一种基于用户-物品二部图的物品相似度计算方法, 并据此为用户推荐其未收藏的物品. Xu 等人<sup>[10]</sup>利用三部图为交通网络建模, 提出一种新的基于数据驱动的方法以发现路网中重要的十字路口.

进一步, 在基于网络结构的 topN 推荐中, 常常分析社交用户的相似性, 一般而言, 用户节点相似性如果越高, 推荐就越容易成功<sup>[11]</sup>. 此外, 文献[12]探究了不相似用户是否有助于提高推荐效果. 将群体现象引入到社交网络并把用户划分到重叠或非重叠群体可降低计算复杂度. 张等人<sup>[13]</sup>基于链接相似性聚类识别重叠社区. 文献[14]提出一种动态贝叶斯非负矩阵分解方法以发现重叠群体. 吴等人<sup>[15]</sup>综合利用社交网络中个体的多维属性, 提出一种高效的社团发现算法用以发现动态网络中的群体. 王等人<sup>[16]</sup>设计一种启发式聚类模型, 在未设定类别数目下对用户自由聚类以客观确定用户邻居数.

此外, 在社交网络分析中一个有意义的研究点是分析社交用户的相似性或相异性, 据此发现节点重要性排名并划分出所有用户的角色. 对角色划分的研究已覆盖到舆情监测、链接预测等多个领域<sup>[17-20]</sup>. 综上, 考虑到二部图在描述两类不同对象及其关系的优势, 同时为形式化刻画社交网络群体, 本文构建群体-个体二部图模型, 综合该模型结构特征与节点自身所具有的属性特征分析节点重要性, 实现角色划分. 进一步地, 本文把角色划分应用于推荐, 最终提出一种基于二部图和节点角色划分的社交网络推荐方案.

## 2 相关知识

在社交网络中, 好友推荐作为信息推荐之一, 其本质是通过计算目标个体与其他个体的相似性实现信息过滤, 最终选择并推荐目标个体可能感兴趣的其他个体. 两个个体越相似则越有可能成为好友. 对于个体  $U_i$  和个体  $U_j$ , 常用符号  $\text{sim}(U_i, U_j)$  表示其相似性. 个体相似性计算的常用方法包括基于共同好友的相似性计算、基于共同群体的相似性计算等. 共同好友相似性定义为  $\text{sim}(U_i, U_j) = |F_{U_i} \cap F_{U_j}|$ , 其中  $F_{U_i}$  和  $F_{U_j}$  分别是个体  $U_i$  和  $U_j$  的好友集合. 共同群体相似性定义为  $\text{sim}(U_i, U_j) = |C_{U_i} \cap C_{U_j}|$ , 其中  $C_{U_i}$  和  $C_{U_j}$  分别是个体  $U_i$  和  $U_j$  所加入的群体集合.

同现实社会一样, 个体在社交网络中也扮演着不同角色. 角色是按照某一体系或标准划分的个体类型, 某类角色代表着具有某种相同特征的个体集. 对于个体  $U_i$ , 其所拥有的角色常用符号  $R_{U_i}$  表示. 角色划分是一个把全部个体节点划分为多类角色的过程. 为实现角色划分, 可分析社交网络所反映的结构特征, 比如分析全部节点的度分布规律, 得出节点在网络中所具有的重要程度, 根据程度差异把全部节点划分为重要节点、活跃节点、流失节点等角色. 同时, 角色划分也可通过相关的聚类方法如 K-Means、BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies) 以及图聚类等把相似节点聚成一类.

## 3 模型

### 3.1 模型框架

本模型框架可划分成三个关键部分, 包括: ①群体发现: 首先把多领域的物品类型抽象成多维度的兴趣群体, 再利用个体所收集的各类物品信息把个体归属到对应的兴趣群体, 最终为群体和个体构建群体-个体二部图模型; ②角色划分: 首先利用已形成的群体-个体二部图, 根据二部图所反映的拓扑信息结合节点自身的属性信息, 再综合本文定义的角色划分标准, 最终聚类全部个体, 实现角色划分; ③生成推荐列表: 首先对多类角色所对应的个体集两两组合, 构建多个角色差异下的个体-个体二部图, 再基于一个该类二部图生成目标个体的推荐个体列表, 最后基于多个该类二部图所生成的推荐个体列表, 综合生成目标个体的最终推荐个体列表.

### 3.2 问题建模

#### 3.2.1 群体-个体二部图模型

**定义 1** 兴趣群体集  $C = \{C_i | i = 1, 2, \dots, m\}$ , 对每一个  $C_i$ , 有  $C_i = \{C_{ij} | j = 1, 2, \dots, s_i\}$ .  $C_i$  表示如电影等各个领域下的兴趣群体集,  $m$  表示领域数,  $C_{ij}$  表示第  $i$  个领域下的第  $j$  个兴趣群体,  $s_i$  表示第  $i$  个领域下的群体数.

**定义 2** 兴趣群体-个体二部图  $G_{cu} = (C, U, E_{cu})$ , 其中  $C$  表示兴趣群体集,  $U$  表示个体集,  $E_{cu}$  表示边集, 且  $E_{cu}$  中任意一条边所关联的两个顶点分别属于  $C$  和  $U$ .

首先, 根据物品类型定义兴趣群体类型, 用  $\text{type}(I)$  表示物品类型. 如果个体在第  $i$  个领域下所收藏的第  $j$  类物品数占该个体所收藏的全部物品数比例  $u\text{Rat}_j$  大于等于该类物品数占全部物品数比例  $i\text{Rat}_j$ , 则把个体加入群体  $C_{ij}$ . 函数  $\text{isInCommunity}(U_k, C_{ij})$  决定  $U_k$  是否加入  $C_{ij}$ :

$$\text{isInCommunity}(U_k, C_{ij}) = \begin{cases} 1, & \text{uRat}_j \geq \text{iRat}_j \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

uRat<sub>j</sub> 计算公式如下:

$$\text{uRat}_j = |I_{U_k - C_j}| / |I_{U_k}| \quad (2)$$

(∀j, type(I<sub>U<sub>k</sub>-C<sub>j</sub>) = j)</sub>

其中, |I<sub>U<sub>k</sub>-C<sub>j</sub></sub>| 表示 U<sub>k</sub> 所收藏的第 i 个领域下第 j 类物品数, |I<sub>U<sub>k</sub></sub>| 表示 U<sub>k</sub> 所收藏的全部类型物品数.

iRat<sub>j</sub> 计算公式如下:

$$\text{iRat}_j = |I_j| / |I| \quad (3)$$

(∀j, type(I<sub>j</sub>) = j)

其中, |I<sub>j</sub>| 表示第 i 个领域下第 j 类物品数, |I| 表示全部类型物品数.

根据该模型获取全部个体应加入的群体, 构建群体-个体二部图 G<sub>cu</sub>. 该模型如图 1 所示, 其中椭圆圈内的物品属于同一类型.

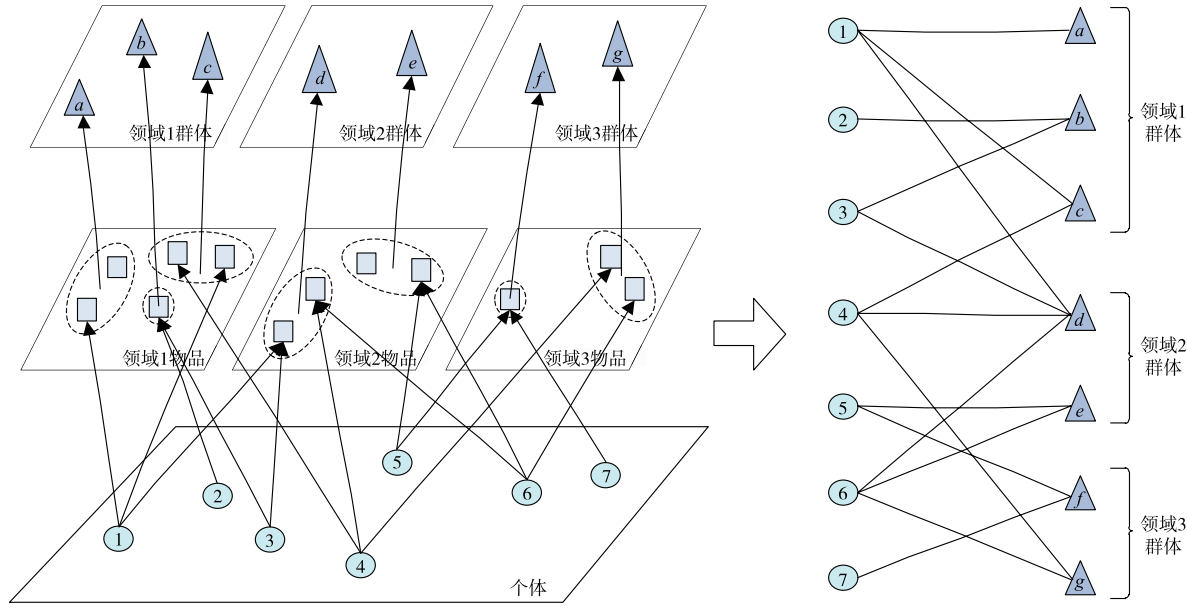


图1 群体-个体二部图模型

### 3.2.2 基于群体-个体二部图的角色划分模型

首先, 该模型把角色类别映射为具体的角色值, 其定义和计算如下:

**定义 3** 第 i 个领域的角色值 val(R<sub>C<sub>i</sub></sub>) (i = 1, 2, ..., m).

val(R<sub>C<sub>i</sub></sub>) 的计算公式如下:

$$\text{val}(R_{C_i}) = \log_2(|C_i| / |R|) \quad (4)$$

其中, |C<sub>i</sub>| 表示第 i 个领域群体集合数, |R| 表示角色数量, |R| 应满足 2 \* |R| < min(|C<sub>i</sub>|).

其次, U<sub>k</sub> 在第 i 个领域角色值计算如下:

$$\text{val}(R_{U_k - C_i}) = |C_{U_k - C_i}| / \text{val}(R_{C_i}) \quad (5)$$

其中, |C<sub>U<sub>k</sub>-C<sub>i</sub></sub>| 表示 U<sub>k</sub> 在第 i 个领域下所加入的群体数量.

最后, 综合 U<sub>k</sub> 在每个领域下的角色值, 得出其综合角色值 val(R<sub>U<sub>k</sub></sub>), 计算公式如下:

$$\text{val}(R_{U_k}) = \sum_{i=1}^m \text{val}(R_{U_k - C_i}) / |R| \quad (6)$$

**定义 4** 角色群体集 CR = {CR<sub>p</sub> | p = 0, 1, ..., |R|}. 角色群体是根据个体角色值所划分而成的群体, CR<sub>p</sub> 表示第 p 类角色群体, 群体内各个体角色值为 p, CR<sub>0</sub> 表示

空角色群体, 该群体中的所有个体未收藏任何物品, 角色值为 0.

不同个体因为所加入的兴趣群体数不同, 使得所计算的角色值不尽相同. 若把全部个体划分至 |R| 个角色群体, val(R<sub>U<sub>k</sub></sub>) 向上取整的值 cval(R<sub>U<sub>k</sub></sub>) 则作为个体 U<sub>k</sub> 的最终角色值, 把他加至第 cval(R<sub>U<sub>k</sub></sub>) 个角色群体.

由此, 通过该模型获取全部个体所属角色, 实现基于群体-个体二部图的角色划分, 形成全部个体与所属角色的关系矩阵. 角色划分如图 2 所示.

### 3.2.3 基于角色差异的个体-个体二部图推荐列表生成模型

考虑到每个个体所拥有的角色不同, 定义基于角色差异下的个体-个体二部图.

**定义 5** 基于角色差异的个体-个体二部图 G<sub>uu</sub> = (U<sub>CR<sub>p</sub></sub>, U<sub>CR<sub>q</sub></sub>, E<sub>uu</sub>), p 和 q 应满足 p, q = 0, 1, ..., |R| 且 CR<sub>p</sub> ≠ CR<sub>q</sub>. U<sub>CR<sub>p</sub></sub> 和 U<sub>CR<sub>q</sub></sub> 分别表示第 p 类和第 q 类角色群体中的个体集, E<sub>uu</sub> 表示边集且 E<sub>uu</sub> 中任意一条边所关联的两个顶点分别属于 U<sub>CR<sub>p</sub></sub> 和 U<sub>CR<sub>q</sub></sub>.

为便于下文描述, 将目标个体所在的个体集称为 A 部, 与之对应的关联个体集称为 B 部.

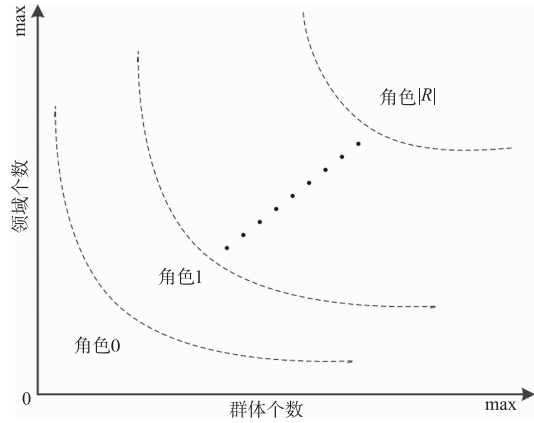


图2 角色划分图

首先,选取不同的两类角色,根据当前角色下两类个体间的共同兴趣是否超过兴趣阈值  $\text{intrest\_threshold}$  决定个体间是否连边,形成当前两类角色下的个体-个体二部图.共同兴趣值反映了两个个体间具有共同兴趣的程度,本文通过统计两个个体所共同收藏的物品数作为其共同兴趣值.如果两类个体共同兴趣值超过阈值则连边,否则不连边,计算兴趣阈值的公式如下:

$$\text{intrest\_threshold} = [\text{val}(\text{lowRole})]^4 \quad (7)$$

其中,  $\text{val}(\text{lowRole})$  表示选择的两类角色中较低的角色值.

接下来,基于一个角色差异的个体-个体二部图生成推荐个体列表.

生成推荐个体列表的过程首先为目标个体所关联的  $B$  部个体赋予初始值.考虑到不同个体兴趣差异,通过  $\text{setInitVal}(U_{\text{conn}})$  函数个性化赋予关联个体初始值,定义如下:

$$\begin{aligned} & \text{setInitVal}(U_{\text{conn}}) \\ &= \frac{|\{ \text{coll}(U_{\text{target}}) \} \cap \{ \text{coll}(U_{\text{conn}}) \}|}{\sum_{U_{\text{conn}} \in \text{conn}U} |\{ \text{coll}(U_{\text{target}}) \} \cap \{ \text{coll}(U_{\text{conn}}) \}|} \quad (8) \end{aligned}$$

其中  $\{ \text{coll}(U_k) \}$  表示某一个体  $U_k$  所收藏的物品集,  $U_{\text{target}}$  表示目标个体,  $U_{\text{conn}}$  表示与目标个体关联的个体,  $\text{conn}U$  表示在二部图中与  $U_{\text{target}}$  所关联的个体集.

考虑到用户和产品的相互影响类似于能量传递和扩散的过程,文献[9]把热传导和能量守恒扩散应用于个性化推荐,具体实施过程则利用二部图中用户、关联产品和目标产品的度定义相似性计算公式以计算产品相似度,最终选择相似度高的前  $N$  个产品推荐给目标用户.本文借鉴该方法,在个性化赋初始值后进行值的传递和扩散.具体实施过程则利用二部图中目标个体、关联个体和推荐个体的度计算关联个体与推荐个体相似性,并生成当前两类角色群体下的推荐个体列表  $L_{CR_{\text{target}}-CR_p}(U_{\text{recommend}})$ , ( $p=0,1,\dots,|R|$ ),  $CR_{\text{target}}$  表示目标

个体  $U_{\text{target}}$  所在的角色群体,  $U_{\text{recommend}}$  表示推荐个体.

最后,本文在为目标个体生成最终推荐列表的过程中,通过函数  $\text{getWtByRole}(\text{cval}(R_{U_{\text{target}}}), \text{cval}(R_{U_p}))$  计算目标个体所在角色群体  $CR_{\text{target}}$  与其它某一角色群体  $CR_p$  所生成的推荐列表权重  $w_{CR_{\text{target}}-CR_p}$ .当前推荐列表中应选择的推荐个体数通过函数  $\text{getNumByWt}(N, w_{CR_{\text{target}}-CR_p})$  计算并把在多个推荐列表中选出的个体以列表权重由大到小顺序排列,生成最终推荐个体列表  $L(U_{\text{recommend}})$ .上述两个函数定义如下:

$$\text{getWtByRole}(\text{cval}(R_{U_{\text{target}}}), \text{cval}(R_{U_p})) \quad (9)$$

$$= \frac{1}{|\text{cval}(R_{U_{\text{target}}}) - \text{cval}(R_{U_p})|}$$

$$\text{getNumByWt}(N, w_{CR_{\text{target}}-CR_p}) = \frac{N}{2} * w_{CR_{\text{target}}-CR_p} \quad (10)$$

### 3.3 算法设计

本文特点是将推荐同二部图结合,基于二部图实现角色划分并生成推荐列表,处理推荐过程中无差异化推荐问题,同时提高推荐的多样性以及降低个体相似性比较的复杂度,其对应的算法设计如算法1所示.

#### 算法1 构建群体-个体二部图模型算法

input: 社交网络中的物品数据  $I$ , 个体数据  $U$ , 个体-物品的收藏关系数据  $E_{ui}$

output: 群体-个体二部图  $G_{cu} = (C, U, E_{cu})$

step1: 统计多个领域下的物品类型,一种物品类型对应一类兴趣群体

$C = \{C_i | i=1, 2, \dots, m\}$ ,  $C_i = \{C_{ij} | j=1, 2, \dots, s_i\}$

step2: 根据函数  $\text{isInCommunity}(U_k, C_{ij})$  决定  $U_k$  是否加入  $C_{ij}$

step3: 得出  $U$  对  $C_{ij}$  加入与否的关系矩阵  $\text{Rel}_{U-C_{ij}}$

$$\begin{aligned} \text{Rel}_{U-C_{ij}} &= (\text{Rel}_{U-C_{ij}}) \\ &= (\text{Rel}_{U-C_{ij}}) \end{aligned}$$

$$(k=1, 2, \dots, |U|)$$

step4: 得出  $U$  对  $C$  加入与否关系矩阵  $\text{Rel}_{U-C}$

$$\text{Rel}_{U-C} = U \times C$$

step5: 根据 step5 所形成的关系矩阵,构建群体-个体二部图  $G_{cu} = (C, U, E_{cu})$

对算法1进行分析,设  $n$  代表个体数,在构建群体-个体二部图的过程中,首先需要形成个体集对某一群体加入与否关系矩阵,再形成个体集对全部群体加入与否的关系矩阵.该过程是逐步进行的,如算法2所示.算法时间复杂度为  $T(n) = O(n * |C|) \sim O(n)$ , ( $|C| \ll n$ ).

#### 算法2 基于群体-个体二部图的角色划分模型算法

input: 群体-个体二部图  $G_{cu} = (C, U, E_{cu})$

output: 个体与所属角色关系矩阵  $\text{Rel}_{U-CR}$

step1: 计算第  $i$  个领域下的角色值  $\text{val}(R_{C_i})$

$$\text{val}(R_{C_i}) = \log_2(|C_i|/|R|), (i=1, 2, \dots, m)$$

step2: 计算  $U_k$  在第  $i$  个领域下所对应的角色值  $\text{val}(R_{U_k-c_i})$

$$\text{val}(R_{U_k-c_i}) = |C_{U_k-c_i}| / \text{val}(R_{C_i})$$

$$(k=1, 2, \dots, |U|)$$

step3: 计算  $U_k$  在全部领域下所对应的角色值, 并得出  $U_k$  的综合角色值  $\text{val}(R_{U_k})$

$$\text{val}(R_{U_k}) = \sum_{i=1}^m \text{val}(R_{U_k-c_i}) / |R|$$

step4:  $\text{val}(R_{U_k})$  向上取整的值  $\text{cval}(R_{U_k})$  作为个体  $U_k$  的最终角色值, 并把  $U_k$  加至第  $\text{cval}(R_{U_k})$  类角色群体

step5: 重复 step3、step4, 直至计算得出全部个体各自的最终角色值, 形成个体与角色群体加入与否的关系矩阵  $\mathbf{Rel}_{U-CR}$

$$\mathbf{Rel}_{U-CR} = U \times CR$$

对算法 2 进行分析, 在基于群体-个体二部图实现角色划分的过程中, 首先需要计算多个领域下的角色标准值, 根据这些角色标准值计算一个个体在多个领域下的角色值并得出该个体的综合角色值, 最后计算出个体集在全部领域下的综合角色值. 该过程是逐步进行的, 如算法 3 所示. 算法时间复杂度为  $T(n) = O(n * m) \sim O(n)$ ,  $m \ll n$ .

**算法 3** 基于角色差异的个体-个体二部图目标个体推荐列表生成算法

input: 目标个体  $U_{\text{target}}$ , 全部个体与所属角色的关系矩阵  $\mathbf{Rel}_{U-CR}$ , 个体-物品的收藏关系数据  $E_{ui}$

output:  $U_{\text{target}}$  的推荐个体列表  $L(U_{\text{recommend}})$

step1: 选取  $CR_{\text{target}}$ , 未和  $CR_{\text{target}}$  组合过的某一角色群体  $CR_p$ , 根据个体间的共同兴趣形成当前两类角色群体下的个体-个体二部图  $G_{uu} = (U_{CR_{\text{target}}}, U_{CR_p}, E_{uu})$ , 其中  $p$  应满足  $p=0, 1, \dots, |R|$  且  $CR_{\text{target}} \neq CR_p$

step2: 生成基于当前两类角色群体下目标个体的推荐个体列表

$L_{CR_{\text{target}}-CR_p}(U_{\text{recommend}})$

step3: 重复 step1、step2, 最终生成基于角色差异的个体-个体二部图推荐个体列表  $L(U_{\text{recommend}})$ , 为社交网络用户推荐好友

对算法 3 进行分析, 在基于角色差异的个体-个体二部图生成推荐列表的过程中, 首先需要根据两类角色下个体间的共同兴趣形成个体-个体二部图, 此时的算法时间复杂度为  $T(n) = O(n_1 * n_2)$ ,  $n_1$  表示 A 部个体数,  $n_2$  表示 B 部个体数; 接下来生成目标个体在当前两类角色群体下的推荐列表, 通过个体权值和连边数计算目标个体与推荐个体的相似度, 此时的算法时间复杂度为  $T(n) = O(n_1 * n_2 * d) \sim O(n_1 * n_2)$ ,  $d$  表示目标个体所关联个体的平均数; 最终把在多个推荐列表中选出的个体以列表权值由大到小顺序排列, 算法整体时间复杂度为  $T(n) = O(l * ((n_1 * n_2 + n_1 * n_2 * d)) \sim O(n_1 * n_2)$ ,  $l = |R| - 1$ .

## 4 实验与分析

### 4.1 数据集

本文通过豆瓣网提供的 API 所获取的数据集验证

提出的推荐方案, 所使用的数据集中共包括 10740 个个体对 104278 部电影, 379710 本书和 292983 部音乐的收藏记录以及物品自身标签信息. 本实验从数据集中随机选择 80% 作为训练集, 20% 作为测试集.

### 4.2 实验方法与评价指标

本文重点研究降低计算个体相似性的复杂度以及实现层次化、差异化推荐, 因此实验围绕着计算效率和推荐效果两方面展开. 其中, 4.3 节是在不同角色划分和个体相似性计算方式下的计算效率对比; 4.4 节对比不同角色划分方式的推荐效果; 4.5 节是逐级与跨级推荐, 相互推荐的推荐效果对比; 4.6 节对比本文推荐方法和原有方法<sup>[9]</sup>的推荐效果. 本文选择准确率、召回率和排序准确度三个评价指标对推荐效果进行评估.

### 4.3 计算效率的实验对比分析

本文基于所划分的角色群体计算个体相似性以生成推荐个体列表. 图 3 对比了划分角色群体所用时间, 图 4 对比了计算个体相似性所用时间. 随着网络规模逐渐增加, 图 3 可以看出基于群体-个体二部图的角色划分所用时间远低于 K-Means 聚类, 同 BIRCH 聚类相比有所降低. 同时, 在 BIRCH 聚类中 CF-树结点因大小限制只可包含有限数目, 基于群体-个体二部图的角色划分则不存在该问题. 图 4 表明基于个体-个体二部图计算个体相似性所用时间低于其他两种方法. 实验结果说明本文所提出的方法可降低划分群体和计算个体相似性的复杂度.

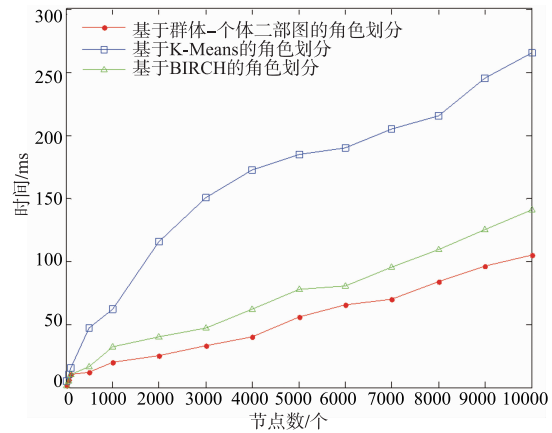


图3 划分角色用时对比

### 4.4 基于群体-个体二部图的角色划分实验对比分析

对于角色划分, 本文选择基于群体-个体二部图的角色划分、随机角色划分以及基于 K-Means 的角色划分三种方法进行对比, 统一使用基于二部图的推荐算法产生推荐列表, 该推荐算法即本文在 3.3 节所设计的算法 3. 在统一  $\lambda = 0.5$  的情况下, topN 推荐的  $N$  值每隔

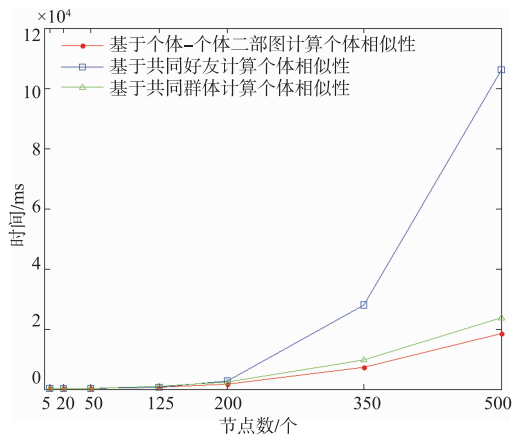


图4 计算个体相似性用时对比

一个从 10 取到 50 ( $\lambda$  是文献[9]相似性计算公式中参数之一,它是值在传递和扩散中的调节混合参数,在多个实验中参数配置情况和评价指标保持一致,后续不再特别说明). 实验共生成 5 个大类角色群体 (xLarge, large, medium, small, xSmall), 每个大类下又分为 5 个小类 (A, B, C, D, E). 考虑到部分个体对任何物品都不感兴趣等情况,实验设定一个空角色 null.

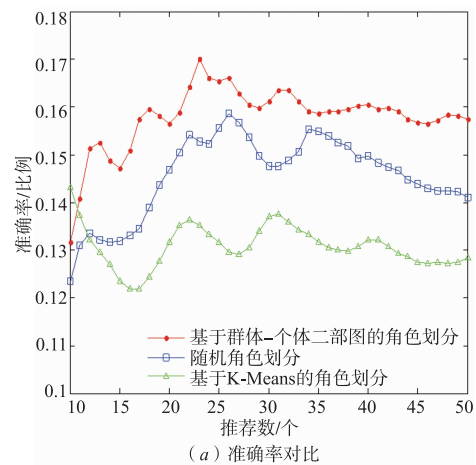
本节选取 medium (C, D) 两类角色,统一对 mediumC 角色所对应个体进行推荐. 图 5(a)(c) 分别可以看出基于二部图的角色划分在准确率和排序准确度上优于其他两种方法,图 5(b) 看出基于二部图的角色划分其召回率低于其他方式. 这说明基于二部图的角色划分在不同标准下有不同表现.

#### 4.5 基于个体-个体二部图的推荐列表生成实验对比分析

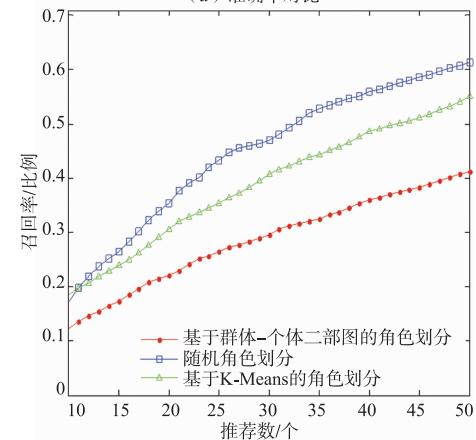
在生成推荐列表过程中,可按照角色由高至低逐级推荐,也可跨一级实现推荐,即所选择的两类角色其等级差的绝对值为 2. 在逐级推荐中,本文选择了 large (B, C, D), medium (B, C, D) 两组共六类角色. 在跨一级推荐中,本文选择了 large (B, D), medium (B, D) 两组共四类角色.

通过图 6 的实验结果可以直观地看出,由高到低逐级推荐与跨一级推荐相比效果差别不大. 这表明在进行层次化推荐时,可为某一类角色个体 A 以相近角色个体 B 作为载体推荐角色个体 C,也可直接为 A 推荐 C.

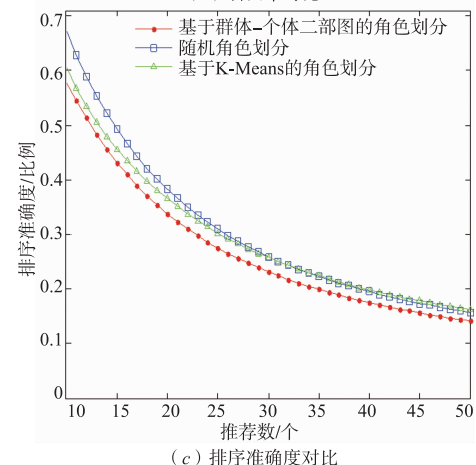
同时,推荐可在两部中相互进行. 通过实验可以看出,对于跨度小的 medium (B, C)、small (B, C) 两组角色,相互推荐效果并未出现差别较大情况. 对于跨度大的 mediumB 和 smallC 这组角色,却在准确率和排序准确度方面差别较大. 图 7 的实验结果表明不论由高到低逐级推荐还是由低到高逐级推荐,基本不会影响推荐效果.



(a) 准确率对比



(b) 召回率对比



(c) 排序准确度对比

图5 不同角色划分方式的推荐效果对比

#### 4.6 基于二部图和角色划分的推荐实验对比分析

文献[9]为关联对象赋予统一初始值,考虑到每个个体兴趣不一致,因此本文通过式(8)赋予各关联个体不同的初始值. 在  $\lambda = 0.0$  的情况下,实验以 medium (C, D) 两类角色群体为代表,统一对 mediumC 角色内个体进行推荐. 通过图 8 的实验结果可以看出,赋予个性化初始值后各推荐指标均好于统一赋予初始值,提升了推荐效果.

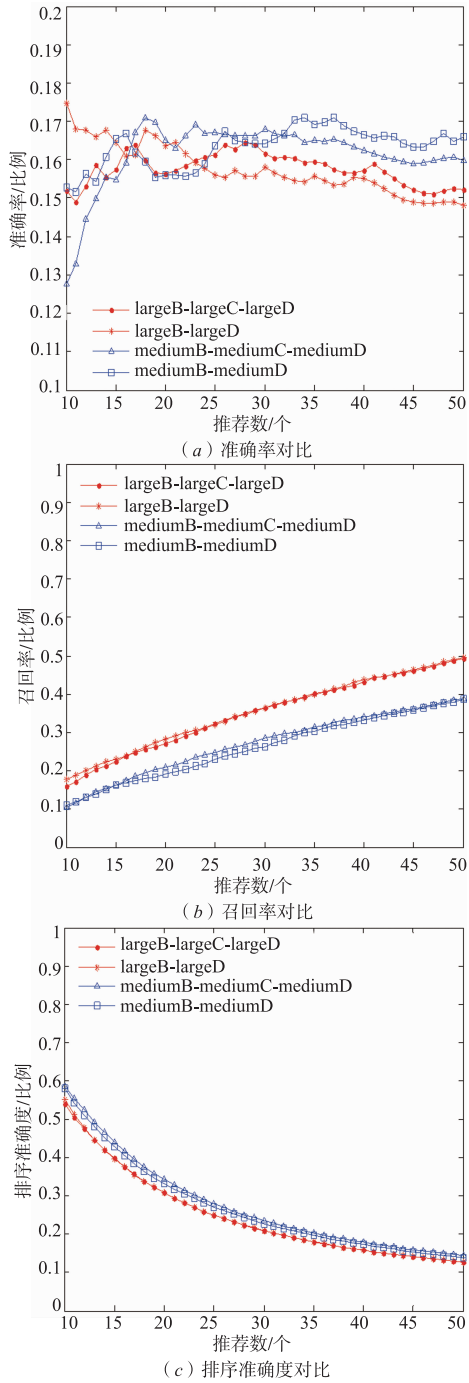


图6 逐、跨级推荐的推荐效果对比

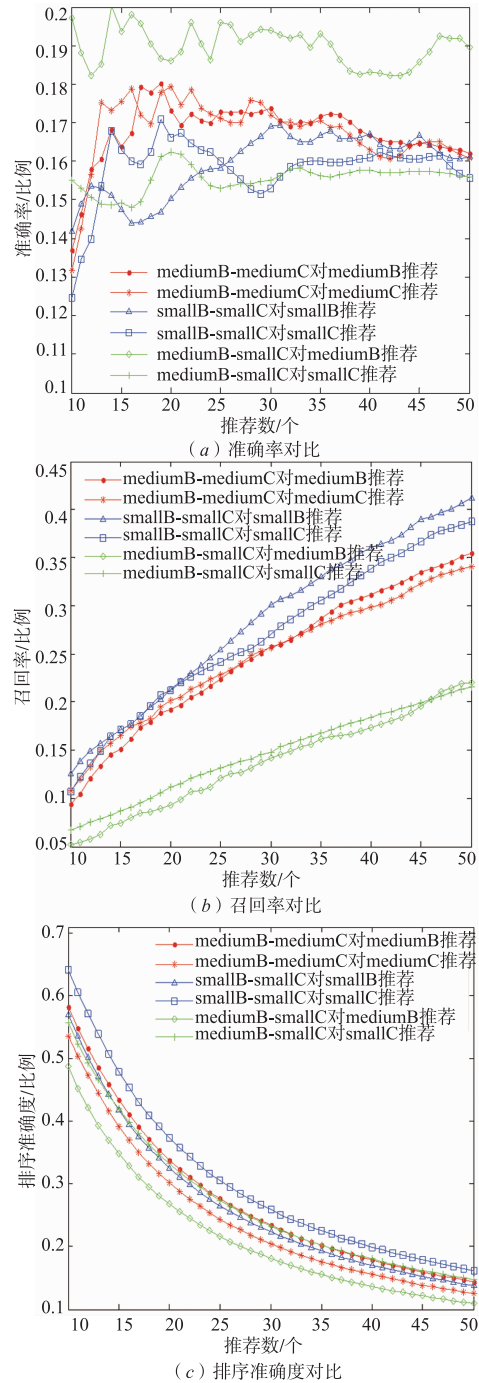


图7 相互推荐的推荐效果对比

### 5 总结

本文首先从个体兴趣和物品信息出发,挖掘多维兴趣群体并构建群体-个体二部图模型,接下来基于群体-个体二部图实现角色划分,并进一步构建基于角色差异的个体-个体二部图,最终基于多个角色差异下的个体-个体二部图实现好友推荐,形成一种基于二部图

和角色划分的层次化、个性化推荐方案. 实验表明,本方案可降低个体聚类与个体相似性比较的复杂度,基于二部图的角色划分整体上优于传统的角色划分方法. 同时,基于角色差异下的二部图不仅能实现差异化推荐,在个性化赋予个体初始值后也能明显提高推荐效果. 下一步工作可拟合赋初值与中间值的过程,进一步改善推荐效果.

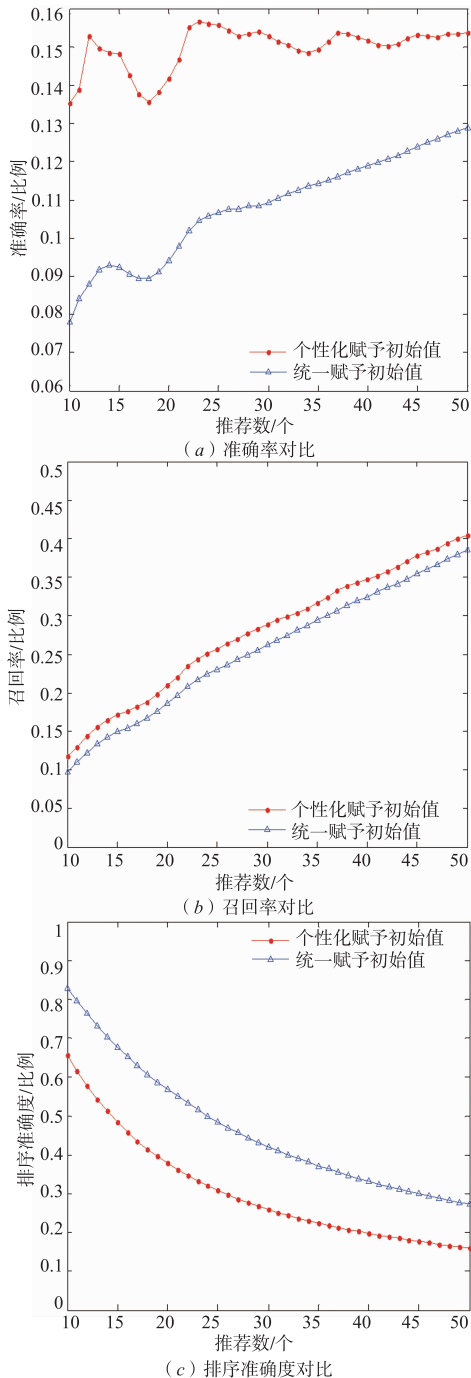


图8 两种赋初值法的推荐效果对比

## 参考文献

- [1] KLAŠNJA-MILČEVIĆ A, IVANOVIĆ M, NANOPOULOS A. Recommender systems in e-learning environments: a survey of the state-of-the-art and possible extensions[J]. Artificial Intelligence Review, 2015, 44(4): 571 – 604.
- [2] JIANG S H, QIAN X M, SHEN J L, et al. Author topic model-based collaborative filtering for personalized POI recommendations[J]. IEEE Transactions on Multimedia, 2015, 17(6): 907 – 918.
- [3] NARDUCCI F, GEMMIS M D, LOPS P. A general architecture for an emotion-aware content-based recommender system[A]. Proceedings of the 3rd Workshop on Emotions and Personality in Personalized Systems[C]. New York: ACM, 2015. 3 – 6.
- [4] CHOO E, Yu T, Chi M, et al. Revealing and incorporating implicit communities to improve recommender systems[A]. Proceedings of the Fifteenth ACM Conference on Economics and Computation[C]. New York: ACM, 2014. 489-506.
- [5] GHAZANFAR M A, PRUGELBENNETT A. Building switching hybrid recommender system using machine learning classifiers and collaborative filtering[J]. Iaeng International Journal of Computer Science, 2010, 37(3): 272 – 287.
- [6] STECK H. Evaluation of recommendations; rating-prediction and ranking[A]. Proceedings of the 7th ACM Conference on Recommender Systems[C]. New York: ACM, 2013. 213 – 220.
- [7] QIAO Z, ZHANG P, HE J, et al. Combining geographical information of users and content of items for accurate rating prediction[A]. Proceedings of the Companion Publication of the 23rd International Conference on World Wide Web Companion[C]. New York: ACM, 2014. 361 – 362.
- [8] HU L K, SUN A X, LIU Y. Your neighbors affect your ratings; on geographical neighborhood influence to rating prediction[A]. Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval[C]. New York: ACM, 2014. 345 – 354.
- [9] ZHOU T, KUSCSIK Z, LIU J G, et al. Solving the apparent diversity-accuracy dilemma of recommender systems[J]. Proceedings of the National Academy of Sciences, 2010, 107(10): 4511 – 4515.
- [10] XU M, Wu J P, Du Y M, et al. Discovery of important crossroads in road network using massive taxi trajectories[A]. Proceedings of the 3rd SigKDD Workshop on Urban Computing[C]. New York: ACM, 2014. 1 – 6.
- [11] TANG J L, HU X, LIU H. Social recommendation: a review[J]. Social Network Analysis & Mining, 2013, 3(4): 1113 – 1133.
- [12] ZENG W, SHANG M S, ZHANG Q M, et al. Can dissimilar users contribute to accuracy and diversity of personalized recommendation? [J]. International Journal of Modern Physics C, 2011, 21(10): 1217 – 1227.
- [13] 张桂杰, 张健沛, 杨静, 等. 基于链接相似性聚类的重叠社区识别[J]. 电子学报, 2015, 43(7): 1329 – 1335.

- ZHANG Gui-jie, ZHANG Jian-pei, YANG Jing, et al. Overlapping community detection based on link similarity clustering[J]. Acta Electronica Sinica, 2015, 43(7): 1329 - 1335. (in Chinese)
- [14] WANG W, JIAO P, HE D, et al. Autonomous overlapping community detection in temporal networks: a dynamic bayesian nonnegative matrix factorization approach[J]. Knowledge-Based Systems, 2016, 110: 121 - 134.
- [15] 吴斌, 王柏, 杨胜琦. 基于事件的社会网络演化分析框架[J]. 软件学报, 2011, 22(7): 1488 - 1502.  
WU Bin, WANG Bai, YANG Sheng-qi. Framework for tracking the event-based evolution in social networks[J]. Journal of Software, 2011, 22(7): 1488 - 1502. (in Chinese)
- [16] 王兴茂, 张兴明, 吴毅涛, 等. 基于启发式聚类模型和类别相似度的协同过滤推荐算法[J]. 电子学报, 2016, 44(7): 1708 - 1713.  
WANG Xing-mao, ZHANG Xing-ming, WU Yi-tao, et al. A collaborative recommendation algorithm based on heuristic clustering model and category similarity[J]. Acta Electronica Sinica, 2016, 44(7): 1708 - 1713. (in Chinese)
- [17] 朱江, 王柏, 吴斌, 等. 一种微博用户情感影响者发现模型[J]. 电子学报, 2015, 43(12): 2497 - 2504.  
ZHU Jiang, WANG Bai, WU Bin, et al. A model for finding emotional influencers in microblog[J]. Acta Electronica Sinica, 2015, 43(12): 2497 - 2504. (in Chinese)
- [18] ROSSI R A, AHMED N K. Role discovery in networks[J]. IEEE Transactions on Knowledge & Data Engineering, 2014, 27(4): 1112 - 1131.
- [19] CHENG D B, Lü L Y, SHANG M S, et al. Identifying influential nodes in complex networks[J]. Physica A: Statistical Mechanics and Its Applications, 2012, 391(4): 1777 - 1787.
- [20] ZHU T, WANG B, WU B, et al. Role defining using behavior-based clustering in telecommunication network[J]. Expert Systems with Applications, 2011, 38(4): 3902 - 3908.

#### 作者简介



肖云鹏 男. 1979年8月出生, 安徽蚌埠人. 副教授、硕士生导师, 主要研究方向为大数据和移动互联网.

E-mail: xiaoy@cqpt. edu. cn



刘瀚松 男. 1991年8月出生, 四川资阳人. 现为重庆邮电大学硕士研究生, 主要研究方向为推荐系统.



刘宴兵 男. 1971年4月出生, 四川遂宁人. 教授、博士生导师, 主要研究方向为网络分析和网络安全.