

# 基于增量图形模式匹配的动态冷启动推荐方法

张亚楠<sup>1,2</sup>, 陈德运<sup>3</sup>, 王莹洁<sup>4</sup>, 刘宇鹏<sup>2</sup>

(1. 哈尔滨理工大学 计算机科学与技术博士后科研流动站, 哈尔滨 150080; 2. 哈尔滨理工大学 软件学院, 哈尔滨 150040;  
3. 哈尔滨理工大学 计算机科学与技术学院; 4. 烟台大学 计算机与控制工程学院, 烟台 264000)

**摘 要:** 针对忽视用户的社交关系变化可能得到不准确的推荐结果这一问题, 为冷启动用户基于社交网络拓扑结构增量更新相似用户, 并基于更新的相似用户给出准确的推荐. 用户社交关系是动态变化的, 然而现有基于社交网络的冷启动推荐却没有充分考虑社交关系的变更对推荐结果的影响. 为了给冷启动用户实时准确的推荐, 提出基于增量图形模式匹配的动态冷启动推荐方法(IGPMDCR), 增量地更新冷启动用户的相似用户, 为冷启动用户给出实时准确的推荐结果. 在真实社交网站的数据集的实验结果表明, IGPMDCR 可以在用户间社交关系变更的情况下, 为冷启动用户给出实时准确的推荐结果.

**关键词:** 冷启动推荐; 社交网络; 增量图形模式匹配; 社交网络拓扑结构

中图分类号: TP 391

文献标志码: A

文章编号: 1008-973X(2017)02-0408-08

## Incremental graph pattern matching based dynamic recommendation method for cold-start user

ZHANG Ya-nan<sup>1,2</sup>, CHEN De-yun<sup>3</sup>, WANG Ying-jie<sup>4</sup>, LIU Yu-peng<sup>2</sup>

(1. *Post-doctoral Research Station of Computer Science and Technology, Harbin University of Science and Technology, Harbin 150080, China*; 2. *Software school, Harbin University of Science and Technology, Harbin 150040, China*;  
3. *School of Computer Science and Technology, Harbin University of Science and Technology, Harbin 150080, China*;  
4. *School of computer and control engineering, YanTai University, YanTai 264000, China*)

**Abstract:** Since ignoring the change of user's social relationships would lead to inaccurate recommendations, similar users for cold start users were updated based on social network topology incrementally, and accurate recommendations were provided based on updated similar users. User social relationships might be dynamically changed, however, the existing social network based cold-start recommendation methods did not fully consider the impact caused by the change of social relationships on recommendations as time pass by. In order to give accurate and timely in manner recommendations for cold start users, an incremental graph pattern matching based dynamic cold-start recommendation method (IGPMDCR) was proposed, which could update similar users for cold-start user incrementally, and give accurate and timely in manner recommendations for cold-start user. Experimental results on the real social network websites datasets show that IGPMDCR can give cold-start user accurately and timely in manner recommendations when user's social relationships are changing.

**Key words:** cold-start recommendation; social network; incremental graph pattern matching; topology of social network

收稿日期: 2016-04-30.

浙江大学学报(工学版)网址: [www.zjujournals.com/eng](http://www.zjujournals.com/eng)

基金项目: 国家自然科学基金青年基金资助项目(61300115, 61502410).

作者简介: 张亚楠(1981—), 男, 讲师, 博士, 从事社交网络推荐等研究. ORCID: 0000-0002-0633-826X.

E-mail: [ynzhang\\_1981@163.com](mailto:ynzhang_1981@163.com)

冷启动用户的历史行为信息稀少,因此对冷启动推荐的研究一直围绕着挖掘、扩充冷启动用户的信息以及为冷启动用户发现相似用户<sup>[1-2]</sup>.国内外学者对冷启动推荐的研究,主要可以分为基于规则的推荐,基于内容的推荐,基于协同过滤的推荐,基于社交网络的推荐.1)基于规则的推荐如基于用户消费方式变化规律构建推荐规则,评估候选项目对规则的投影,并给出推荐<sup>[3]</sup>.通过抽取冷启动用户的特征,根据特征分簇并制定推荐规则,给出推荐<sup>[4]</sup>.基于规则的推荐只能将符合规则的信息推荐给用户,推荐信息的覆盖范围较小,且很难反映出由于用户间社交关系变化而导致的推荐结果变更.2)基于内容的推荐将与用户关注的关键词相似度较高的信息推荐给用户<sup>[5]</sup>.基于内容的推荐效果取决于准确的特征抽取,如 Yin 等<sup>[6]</sup>提出一种潜类统计混合模型.基于内容的推荐结果中包含大量与用户历史浏览信息重复的信息.3)基于协同过滤的推荐根据用户之间的相似程度或者项目之间的相似程度给出推荐<sup>[7]</sup>.为解决用户数据稀疏的问题, Jamali 等<sup>[8]</sup>提出非负矩阵分解(non-negative matrix factorization, NMF). Ma 等<sup>[9]</sup>提出基于概率矩阵分解(probabilistic matrix factorization, PMF)的推荐. Wu 等<sup>[10]</sup>提出在矩阵分解的过程中引入影响用户喜好的特征向量可以提高推荐的准确度. Koren<sup>[11]</sup>提出时间敏感的协同过滤推荐算法. Ren 等<sup>[12]</sup>提出平衡评分预测机制的协同过滤推荐方法.基于协同过滤的推荐效果受用户历史行为信息数量的影响非常明显.4)基于社交网络的推荐通过用户社交关系标签的方法为冷启动用户发现相似用户进而给出推荐<sup>[13-17]</sup>.准确地发现用户间信任关系是基于信任推荐的关键.国内外学者的研究主要包括信任关系的传递策略<sup>[18-19]</sup>、验证信任网络具有小世界特征<sup>[20]</sup>,基于社交网络的小世界特性和用户间弱关系构建信任网络<sup>[21]</sup>.如 Liu 等<sup>[22]</sup>提出传播过程中扩散的相对量与绝对量的权重博弈. Guha 等<sup>[23]</sup>提出用户间的信任关系可通过由用户给出少量不信任或信任实例预测. Deng 等<sup>[24]</sup>提出一种基于增强信任社交网络的服务推荐方法. Moradi 等<sup>[25]</sup>提出一种基于可信度的信任感知协同过滤方法.然而用户间社交关系不是静止不变的,冷启动用户本身及其他用户的社交关系会随着时间发生变化<sup>[26]</sup>,现有基于社交网络的冷启动推荐方法却没有充分考虑社交关系的变更对推荐结果的影响<sup>[27-28]</sup>.本文提出一种基于增量图形模式匹配的动态冷启动推荐方法(Incremental graph pattern matching based dynamic cold-start

recommendation, IGPMDCR),基于社交网络中用户间社交关系拓扑结构随时间变化的情况,实时地更新冷启动用户的相似用户,进而给出实时的推荐.

## 1 社交关系拓扑

社交网络中丰富的用户社交关系信息可以用来反映用户在现实社会中的偏好,通过为冷启动用户发现具有一定数量购物记录和评价信息的相似用户,可以为冷启动用户给出推荐.然而,社交网络中用户间的社交关系不是静止不变的,多数用户间的社交关系会随着时间的推移而发生变化,当与冷启动用户相关联的社交关系发生变化后,如果仍然基于变化前的相似用户为冷启动用户给出推荐,显然会产生差异.社交网络中用户的规模已经非常庞大,尤其在用户间社交关系频繁变更的情况下为冷启动用户重新匹配相似用户是代价昂贵的.因此,在冷启动用户已有相似用户的基础上,基于冷启动用户及其他用户社交关系拓扑结构的变化,为冷启动用户动态更新相似用户,给出实时的推荐是一种可行的方法.

社交关系拓扑是由表示用户的节点以及表示用户间社交关系的边组成的有向图.其中边的权重值可以表示社交关系的紧密程度,对边标记类型可以区分不同种类的社交关系.令  $u'$  表示与冷启动用户  $u$  存在某种社交关系的用户,  $u$  和  $u'$  之间的社交关系紧密程度用  $u$  和  $u'$  之间路径长度  $f_e(u, u')$  表示,路径长度越短则对应的社交关系越紧密,路径长度越长则对应的社交关系越疏远.  $f_e(u, u') = \infty$  表示用户  $u$  和  $u'$  之间不存在社交关系.  $u$  和  $u'$  间的社交关系类型用  $f_L(u, u')$  表示.通过社交关系类型和社交关系的路径长度可以描述用户在社交网络中的社交关系拓扑.令  $G = (V, E, f_A)$  表示社交网络的拓扑,其中 1)  $V$  表示有限的节点集合;2)  $E \subseteq V \times V$  表示从  $v$  到  $v'$  的边,其中  $v$  和  $v'$  表示社交网络中任意用户;3)  $f_A(\cdot)$  表示节点  $v$  的属性,如  $f_A(v) = (A_1 = a_1, \dots, A_n = a_n)$ ,  $A_i$  是用户  $v$  的一种属性,  $a_i$  是一个常量.以冷启动用户的社交关系拓扑为匹配模式,则匹配模式为  $Q = (V_Q, E_Q, f_L, f_e)$ ,其中  $V_Q$  是有限的点集,  $E_Q$  是有向边的集合.

## 2 基于增量图形模式匹配的动态相似用户更新

在冷启动用户已有相似用户的基础上,基于冷

启动用户及其他用户社交关系的变化,为冷启动用户动态更新相似用户,给出实时的推荐。

基于增量图形模式匹配的动态相似用户发现是基于冷启动用户已有相似用户的基础上,动态的为冷启动用户更新相似用户。因此首先要发现冷启动用户的相似用户。为冷启动用户发现相似用户的过程是在  $G$  中发现所有与匹配模式  $Q$  相匹配的节点集合,即基于有界图形模式匹配发现相似用户。令  $S \subseteq V_Q \times V$ , 基于有界图形模式匹配的相似用户条件: 1) 用户  $v$  的属性与用户  $u$  的属性相似。2) 对于  $(u, u')$ , 有向图中存在一个非空且长度不超过  $f_e(u, u')$  的路径  $v/\dots/v'$ , 且存在  $(u, v) \in S$ 。3) 对于  $(u, u')$ ,  $G$  中存在路径  $v/\dots/v'$ , 使  $f_L(v, v_1) \dots f_L(v_n, v')$  满足  $(u, u')$  上的关系  $f_L(u, u')$ , 其中路径  $v/\dots/v'$  中节点的顺序为  $(v, v_1, \dots, v_n, v')$ 。基于有界图形模式匹配的相似用户发现算法如算法 1 所示。其中,  $\text{mat}(u)$  记录  $G$  中可以匹配  $Q$  中节点  $u$  的节点,  $\text{demat}(u)$  记录  $G$  中不能匹配  $Q$  中节点  $u$  的节点。Anc() 和 desc() 分别表示节点前驱和后继节点。对于任意  $x \in V$  且  $(u', u) \in E_C$ ,  $\text{anc}(f_e(u', u), f_v(u'), x)$  表示  $x$  的符合如下条件的前驱节点  $x'$ ,  $x'$  满足 1)  $x'$  到  $x$  的距离不大于  $f_e(u', u)$ , 2)  $f_v(x')$  满足  $f_v(u')$ ;  $\text{desc}(f_e(u, u'), f_v(u'), x)$  表示  $x$  的符合如下条件的后继节点  $z$ ,  $z$  满足 1)  $z$  到  $x$  的距离不大于  $f_e(u', u)$ , 2)  $f_v(z)$  满足  $f_v(u')$ 。

**算法 1** 基于有界图形模式匹配的相似用户发现

输入:  $Q = (V_Q, E_Q, f_L, f_e), G = (V, E, f_A)$ 。

输出: 最大匹配集合  $S$

1. 计算  $G$  的距离矩阵  $M$ ;
2. for each  $(u', u) \in E_Q, x \in V$  do
3. 计算  $\text{anc}(f_e(u', u), f_v(u'), x), \text{desc}(f_e(u', u), f_v(u'), x)$ ;
4. for each  $u \in E_Q$  do
5.  $\text{mat}(u) := \{x \mid x \in V, f_A(x) \text{ 符合 } f_v(u), \text{ 且 } \text{out-degree}(a) \neq 0, \text{out-degree}(x) \neq 0\}$ ;
6. if  $(\text{out-degree}(a) \neq 0, \text{ 并且 } \exists (u', u) \in E_Q (x' \in \text{mat}(u), f_A(x') \text{ 满足 } f_v(u'), \text{ 并且 } \text{len}(x/\dots/x') \leq f_e(u', u))$   
 $\text{demat}(u) := \{x \mid x \in V, \text{out-degree}(x)\}$ ;
7. while  $(u \in V_Q \text{ 的 } \text{demat}(u) \neq \emptyset)$  do
8. for each  $(u', u) \in E_Q, z \in \text{demat}(u) \cap \text{mat}(u')$  do
9.  $\text{mat}(u') := \text{mat}(u') \setminus \{z\}$ ;
10. if  $(\text{mat}(u') = \emptyset)$  then return  $\emptyset$ ;
11. for each  $u''$  with  $(u'', u') \in E_Q$  do
12. for  $(z' \in \text{Anc}(f_e(u'', u'), f_v(u''), z) \wedge z' \in \text{demat}(u'))$  do
13. if  $(\text{desc}(f_e(u'', u'), f_v(u'), z') \cap \text{mat}(u') = \emptyset)$

14. then  $\text{demat}(u') := \text{demat}(u') \cup \{z'\}$ ;
15.  $\text{demat}(u) := \emptyset$ ;
16.  $S := \emptyset$ ;
17. for  $(u \in V_Q \text{ and } x \in \text{mat}(u))$  do  $S := S \cup \{(u, x)\}$ ;
18. return  $S$

当社交网络中用户间关系发生变化,即无论是冷启动用户本身发生变化,还是与冷启动用户关联的用户发生变化,都需要重新为冷启动用户计算相似用户。基于增量图形模式匹配为冷启动用户更新相似用户,根据  $G$  中用户间路径的变化,为冷启动用户删除或增加相似用户。其过程可表示如下,令  $M(Q, G)$  表示符合  $Q$  的最大匹配集合,令  $\Delta G$  表示在原  $G$  中删除或者添加边的集合,即  $G$  的更新,  $\oplus$  表示应用更新,令  $G'$  表示更新后的用户集合,则  $G' = G \oplus \Delta G$ ,  $G$  更新后符合  $Q$  的最大匹配记为  $M(Q, G \oplus \Delta G)$ , 令  $\Delta M$  表示变化的最大匹配,则  $M(Q, G \oplus \Delta G) = M(Q, G) \oplus \Delta M$ 。IGPMDCR 为冷启动用户更新相似用户可分为删除相似用户和增加相似用户。IGPMDCR 迭代的发现社交关系变更的节点的前驱节点的路径长度变化,进而判断社交关系变更的最大范围,再判断冷启动用户的已有相似用户是否处于社交关系变更的范围内。

1) 删除相似用户: 首先,确定由于删除用户节点,用户间社交关系变化的影响范围  $(\text{AFF})_s$ , 即用户退出以及用户间社交关系删除对原有用户间路径长度的影响。计算用户间最短路径长度的更改,以及用户间最短路径包含节点的变化。其次,在冷启动用户已有的相似用户中搜索并删除用户属性、路径长度不再匹配的用户,再分别以最短路径变更节点为起始节点,搜索其前驱节点,判断所经历路径是否既包含在匹配模式  $Q$  的路径中,又在符合  $Q$  的最大匹配集合  $M(Q, G)$  中仅存在一条与该路径匹配的路径,删除变更后的路径长度不满足限制条件的相似用户。动态删除相似用户算法 propCS 如算法 2 所示,其中  $G_r$  表示原相似用户集合,  $V_r$  和  $E_r$  分别表示相似用户节点和边的集合,  $G'_r$  表示动态更新后的相似用户集合,  $E_{cs}$  记录  $\text{AFF}_{cs}$  中的边,  $v'$  的前驱节点  $v''$ 。propCS 首先检测  $e = (v', v)$  是否为匹配模式的关键路径,如果不是关键路径则原相似用户集合  $G_r$  不改变。如果是关键路径,则递归的发现所有与  $v'$  匹配且由于删除  $e$  不再满足条件的边和节点。

**算法 2** 动态删除相似用户

Input:  $Q, G_r = (V_r, E_r), E_{cs}$ 。

Output:  $G'_r$ 。

1. if  $(E_{cs} := \emptyset \ \&\& \ e \notin E_r)$  then return  $G_r$ ;

```

2.  $E_{cs} : \text{push}(e)$ ;
3. while  $E_{cs}$  do
4.  $e := E_{cs} : \text{pop}()$ ;
5. for each  $e_Q = (u', u)$  match  $e = (v, v')$  do
6. if  $v'$  匹配  $u'$ ;
7. else  $v'$  不匹配  $u'$  then
8. for each  $e' = (v'', v')$  in  $E_r$  do
9.  $E_r := E_r \setminus \{e'\}$ ;  $E_{cs} : \text{push}(e')$ ;
10.  $V_r := V_r \setminus \{v'\}$ ;  $\text{mat}(u') := \text{mat}(u') \setminus \{v'\}$ ;
11. if  $\text{mat}(u') = \emptyset$  then return  $\emptyset$ ;
12. return  $G'_r$ .

```

2) 增加相似用户: 首先, 确定由于增加用户节点导致用户间社交关系变化的影响范围  $\text{AFF}_{cc}$ . 其次, 分别以最短路径变更节点为起始节点, 搜索其前驱节点, 判断所经历路径是否既包含在匹配模式  $Q$  的路径中, 又包含在符合  $Q$  的最大匹配集合  $M(Q, G)$  的路径中, 新增变更后的路径长度满足限制条件的用户作为相似用户. 动态增加相似用户算法  $\text{propCC}$  如算法 3 所示, 其中  $E_{cc}$  记录  $\text{AFF}_{cc}$  中的边.

#### 算法 3 动态增加相似用户

```

Input:  $E_{cc}, Q, G_r$ .
Output:  $G'_r$ .
1. if  $(E_{cc} := \emptyset \ \&\& \ e \notin E_r)$  then return  $G_r$ ;
2. for each  $e \in G_r$  do
3.  $E_{cc} := \{(v', v) \mid (v', v) \text{ 的长度小于 } (u', u) \text{ 的长度}\}$ ;
4. if  $E_{cc} \neq \emptyset$  then
5. for each  $u \in G_r$  do  $\text{mat}(u') := \text{mat}(u)$ ;
6. 为  $G_r$  计算匹配项;
7. if  $\text{mat}(u') \neq \emptyset$  then 更新  $G_r, E_{cc}$ ;
8. return  $G'_r$ ;

```

基于增量图形模式匹配为冷启动用户更新相似用户如算法 4 所示. 动态更新相似用户分别识别  $\text{AFF}_{cs}$  和  $\text{AFF}_{cc}$ , 当有新插入的边, 调用  $\text{propCC}$ , 删除边的过程调用  $\text{propCS}$ . 基于增量图形模式匹配为冷启动用户更新相似用户的时间复杂度与匹配模式  $Q$ , 更新社交关系  $E_{cc}$  和  $E_{cs}$  成正比. 即动态更新相似用户的时间复杂度为  $O((|E_Q| + |V_Q|)(|E_{cc}| + |E_{cs}|))$ , 展开  $(|E_Q| + |V_Q|)(|E_{cc}| + |E_{cs}|)$ , 得到  $(|E_Q||E_{cc}| + |E_Q||E_{cs}| + |E_{cc}||V_Q| + |V_Q||E_{cs}|)$ ,  $E_{cc}$  和  $E_{cs}$  近似同一个数量级. 在用户社交关系变化范围较小的情况下, 即  $E_{cc}$  远远小于  $E$  的情况下, 动态更新相似用户的时间复杂度为  $O(|E_Q||E_{cc}| + |E_{cc}||V_Q|)$  远小于重新计算相似用户的时间复杂度  $O(|V||V_Q| + |E_Q||V| + |E_Q||E| + |V_Q||E|)$ . 当社交关系发生显著变化时, 即  $E_{cc}$  与  $E$  近似为同一个数量级, 动态更新相似用户的时间复杂度为

$O(|E_Q||E_{cc}| + |E_{cc}||V_Q|)$  与重新计算相似用户的时间复杂度  $O(|V||V_Q| + |E_Q||V| + |E_Q||E| + |V_Q||E|)$  为同一个数量级, 所以社交关系发生显著变化后, 基于增量图形模式匹配更新冷启动用户的相似用户退化为重新计算冷启动用户的相似用户.

#### 算法 4 动态更新相似用户

```

Input:  $Q, G_r, E_{cc}, E_{cs}$ .
Output:  $G'_r$ .
1.  $E_{cs} := \{(v', v) \mid (v', v) \text{ 属于 } \text{AFF}_{cs} \text{ 对应存在 } (u', u) \in E_Q\}$ ;
2.  $\text{propCS}(E_{cs}; Q; G_r)$ ;
3.  $E_{cc} := \{(v_1', v_1) \mid (v_1', v_1) \text{ 属于 } \text{AFF}_{cc} \text{ 对应存在 } (u_1', u) \in E_Q\}$ ;
4.  $\text{propCC}(E_{cc}; Q; G_r)$ ;
5. return  $G'_r$ .

```

### 3 由冷启动用户的相似用户给出推荐

对冷启动用户的推荐实质是预测用户商品矩阵  $R = [r_{u,i}]_{m \times s}$  中的未知元素, 即冷启动用户对未知商品的评价值, 依据评价值的排序, 将 Top-N 评价值对应的商品推荐给冷启动用户. 令  $i$  表示商品,  $r_{u,i}$  表示冷启动用户  $u$  对商品  $i$  的评价值,  $m$  表示冷启动用户个数,  $s$  表示商品个数. 由于冷启动用户具有很少的购物记录和评价, 因此冷启动用户的用户商品矩阵是一个极其稀疏的矩阵. 为冷启动用户推荐, 其实质是一个矩阵补全过程, 即通过为冷启动用户发现相似用户, 再根据其冷启动用户的相似程度和其本身的评价值补全冷启动用户的用户商品矩阵. 为冷启动用户推荐的第一步是发现冷启动用户的相似用户. 第二步根据相似程度预测冷启动用户在用户商品矩阵中的未知项, 其中冷启动用户对未知项的评价值可由式(1)求得, 式(1)中  $w_{a,u}$  表示由第一步计算出的用户  $a$  对  $u$  的社交关系相似程度,  $w_{a,u}$  的值越大, 相应的推荐权重越大,  $r_{a,i}$  表示用户  $a$  对商品  $i$  的评价值,  $\bar{r}_a$  表示用户  $a$  的平均评价值,  $\bar{r}_u$  表示冷启动用户  $u$  的平均评价值.  $\sigma$  表示与冷启动用户相似且评价过商品  $i$  的用户数量. 为冷启动用户  $u$  给出  $N$  个推荐结果的形式化描述如式(2)所示, 其中  $I = \{i_1, i_2, \dots, i_s\}$  表示商品集合,  $N$  表示给出推荐的个数. 准确的为冷启动用户发现相似用户, 即可以准确地为冷启动用户给出推荐. 为冷启动用户  $u$  给出推荐的形式化描述如式(1)和(2)所示.

$$r_{u,i} = \bar{r}_u + \frac{1}{\sigma} \sum w_{a,u} (r_{a,i} - \bar{r}_a). \quad (1)$$

$$\text{Top}(u, N) := \max_{i \in I}^N r_{u,i}. \quad (2)$$

## 4 实验结果与分析

### 4.1 实验数据及评测方法

实验数据集取自 Epinions 和 YouTube 网站, 其中 Epinions 记录用户对商品的评价值, 以及用户间评价、分享、回复、转发评论关系, 用户对商品的评价值分为 5 个档次, 其中 1 表示不推荐, 5 表示非常推荐. 数据集中有 49 289 个用户对 139 544 个不同商品的评价, 评价总数达到 586 361 条. 其中有 35.8% 的用户至少评价过 8 个商品, 47.3% 的用户评价过 3~7 个商品, 16.9% 的用户至多评价过 2 个商品. 在实验中将评价个数少于 2 的用户作为冷启动用户, 测试集中冷启动用户的数量占 16.9%. 为了验证 IGPMDCR 能够实时反映用户社交关系变化并给出准确的推荐, 数据集收集了 Epinions 网站 640 d 的用户数据, 这些用户数据反映了社交关系随时间而发生变化的情况.

YouTube 是一个视频分享网站. YouTube 中包含不同的社交关系. 从 YouTube 爬取 30 522 个用户的配置文件. 基于所获取的信息, 构建 5 种不同的社交关系. 这些关系包括: 用户之间的联络关系; 共享好友关系; 用户之间共享订阅数量的关系; 用户间共享用户数; 共享喜爱视频的数量. 将有不超过约 3 个共享订阅和共享喜欢视频的用户作为冷启动用户. 在获取的 YouTube 数据集中, 冷启动用户占 27.37%. 同样收集了 YouTube 640d 的用户数据用来反映社交关系随时间而发生变化的情况.

实验结果比较方法采用均方根误差 (root mean square error, RMSE), RMSE 是评估算法计算的预测值与真实值之间差距的指标, 可用于衡量推荐效果<sup>[22]</sup>, RMSE 的定义如式 (3) 所示, 其中  $\hat{r}_{u,i}$  表示由算法预测的用户  $u$  对商品  $i$  的评价值,  $T$  表示用户评价的商品数目. RMSE 值越小, 则算法的推荐效果越好.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^T (r_{u,i} - \hat{r}_{u,i})^2}{T}}. \quad (3)$$

### 4.2 实验结果

分别与在实时冷启动推荐中取得较好效果的方法和效果较好的基于社交网络的推荐算法做对比试验.

在实时冷启动推荐中取得较好效果的方法包括: Bobadilla 等<sup>[1]</sup>提出一种基于神经学习优化的相

似度量方法, 动态的为冷启动用户给出推荐. Koren<sup>[11]</sup>提出时间敏感的协同过滤推荐算法, 在用户、项目特征向量中引入时间特征, 有效地区分瞬时效应和长期模式的用户兴趣. Lika 等<sup>[2]</sup>提出基于分类算法的协同过滤方法动态的发现相似用户, 进而为冷启动用户给出实时推荐. Ling 等<sup>[4]</sup>提出通过矢量余弦法获取用户相似矩阵, 实时的为各组用户给出推荐.

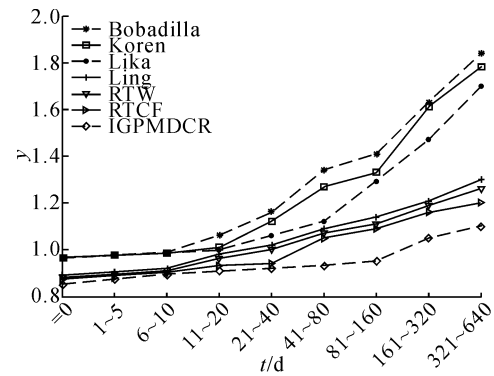
基于社交网络的推荐算法中推荐结果较好的方法包括: 基于增强信任社交网络的服务推荐方法和基于可信度的信任感知协同过滤方法. 基于增强信任社交网络的服务推荐方法, 又称为关联信任游走 (relevant trust walker, RTW). Deng 等<sup>[24]</sup>首先构建可获取用户在社交网络中信任度的矩阵分解方法, 然后改进随机游走方法获取推荐结果. 基于可信度的信任感知协同过滤方法 (reliability-based trust-aware collaborative filtering method, RTCF) 提升基于信任感知推荐的准确度. 由用户相似值和信任状态构建初始信任网络, 基于信任的可信度量初始评估构建信任网络的可信度, 重构可信度低于预设值的信任网络<sup>[25]</sup>.

由于推荐算法需要一定的训练集, 因此使用不同百分比的训练集, 分别占整个测试集的 10%, 20%, 40% 和 80%. 测试集中包含 640 d 的用户数据, 包含了用户社交关系随时间而发生变化的情况, 选取社交关系变动明显的 9 个阶段, 分别是: 第 1 个阶段初次使用数据集的时间; 第 2 个阶段距离起始时刻 1~5 d; 第 3 个阶段距离起始时刻 6~10 d; 第 4 个阶段距离起始时刻 11~20 d; 第 5 个阶段距离起始时刻 21~40 d; 第 6 个阶段距离起始时刻 41~80 d; 第 7 个阶段距离起始时刻 81~160 d; 第 8 个阶段距离起始时刻 161~320 d; 第 9 个阶段距离起始时刻 321~640 d. 分别在 Epinions 和 YouTube 测试集中测试 Bobadilla、Koren、Lika、Ling、RTW、RTCF 和 IGPMDCR 在不同时间阶段的推荐效果. 在 Epinions 测试集中的实验结果如图 1(a)~(d) 所示, 其中图 1(a)~(d) 分别对应测试集占 10%, 20%, 40% 和 80% 的情况, 横坐标  $t$  为社交关系变动明显的 9 个阶段距离起始时刻的时间, 纵坐标  $y$  表示由算法给出推荐结果的 RMSE 值, 在图 1(a)~(d) 中 IGPMDCR 的 RMSE 值始终小于 Bobadilla、Koren、Lika、Ling、RTW、RTCF 的 RMSE 值, 并且随着时间的推移 IGPMDCR 的 RMSE 值与 Bobadilla、Koren、Lika、Ling、RTW、RTCF 的 RMSE 值差距逐渐增加, 当超过 20 d 后, Bobadilla、Koren、

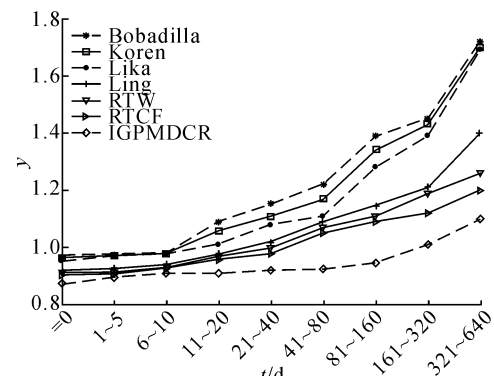
Lika、Ling、RTW、RTCF 的 RMSE 与 IGPMDCR 的 RMSE 差距开始加速扩大,到 321~640 d Bobadilla、Lika、Ling、RTW、RTCF 的 RMSE 与 IGPMDCR 的 RMSE 差距达到最大.这是由于 Bobadilla、Koren、Lika、Ling、RTW、RTCF 的方法对用户社交关系的变化不敏感,不能实时反映用户社交关系的变化和与其对应的用户偏好变化.

在 YouTube 测试集的实验结果如图 2(a)~(d)所示,其中图 2(a)~(d)分别对应测试集占 10%,20%,40%和 80%的情况.在图 2(a)~(d)中 IGPMDCR 的 RMSE 值始终小于 Bobadilla、Koren、Lika、Ling、RTW、RTCF 的 RMSE 值,并且随着时间的推移 IGPMDCR 的 RMSE 值与 Bobadilla、Koren、Lika、Ling、RTW、RTCF 的 RMSE 值差距逐渐增加,当超过 20 d 后, Bobadilla、Koren、Lika、Ling、RTW、RTCF 的 RMSE 与 IGPMDCR 的 RMSE 差距开始加速扩大,到 321~640 d Bobadilla、Koren、Lika、Ling、RTW、RTCF 的 RMSE 与 IGPMDCR 的 RMSE 差距达到最大.在 Epinions 和 YouTube 测试集的实验结果表明 IGPMDCR 在用户社交关系随时间而改变的过程中,可以为冷启动用户给出更准确的推荐.

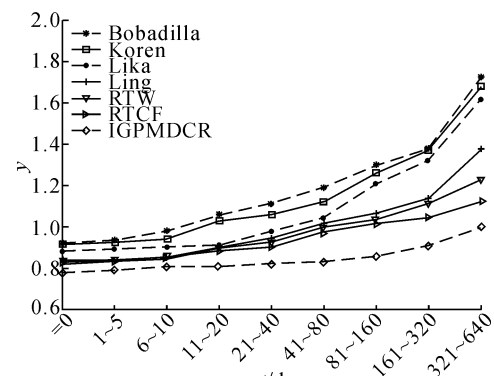
基于增量图形模式匹配为冷启动用户更新相似用户,基于变化前的社交网络  $G$  中用户  $u$  的最相似用户,然后根据增量挖掘技术增加或者删减最相似用户,生成社交关系变化后  $u$  的最相似用户.当社交关系发生显著变化时,通过实验来进行验证社交网络中社交关系发生变化的用户的百分比与基于增量图形模式匹配为冷启动用户更新相似用户的时间关系.如图 3 所示,横轴  $R$  为社交网络中社交关系变化的百分比,纵轴  $O$  为基于增量图形模式匹配为冷启动用户更新相似用户的时间与完全重新计算相似用户的时间的百分比.  $O$  取值越接近 0,基于增量图形模式匹配为冷启动用户更新相似用户的时间越少;  $O$  取值越接近 1,则基于增量图形模式匹配为冷启动用户更新相似用户退化为重新计算冷启动用户的相似用户.图 3 中,当社交网络中有 10%的用户社交关系变化时,基于增量图形模式匹配为冷启动用户更新相似用户的时间远小于重新计算相似用户的时间,随着用户社交关系变化比率的升高,基于增量图形模式匹配为冷启动用户更新相似用户的时间不断增加.当社交网络中有 20%以上的用户社交关系变化时,基于增量图形模式匹配为冷启动用户更新相似用户的时间加速上升,当有 30%以上用户社交关系变化时,基于增量图形模式匹配为冷启动用



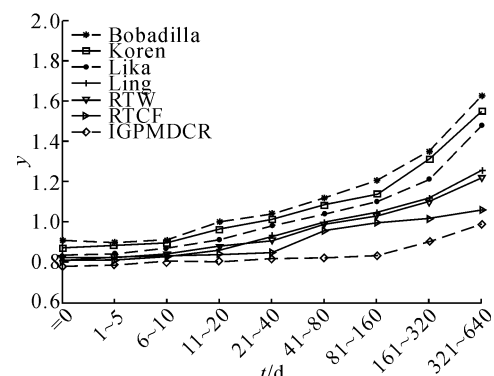
(a) 测试集占整个数据集的10%



(b) 测试集占整个数据集的20%



(c) 测试集占整个数据集的40%



(d) 测试集占整个数据集的80%

图1 在 Epinions 测试集的推荐效果比较

Fig. 1 Comparison of recommendation effectiveness against RMSE on Epinions

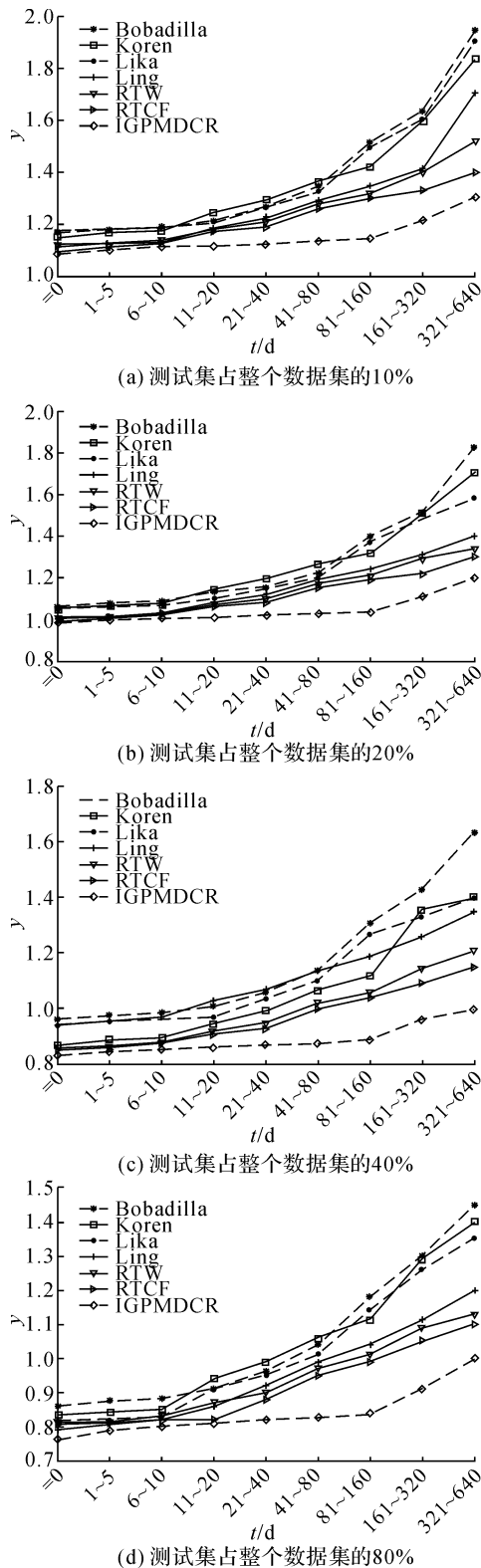


图2 在 YouTube 测试集的推荐效果比较

Fig. 2 Comparison of recommendation effectiveness against RMSE on YouTube

户更新相似用户的时间与用户社交关系变化百分比几乎成正比关系. 当有 90% 以上用户社交关系变化时, 基于增量图形模式匹配为冷启动用户更新相似

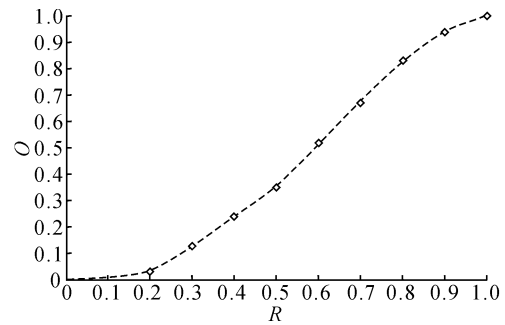


图3 社交网络中社交关系发生变化的用户百分比与基于增量图形模式匹配为冷启动用户更新相似用户的时间关系

Fig. 3 Relationship between percentage of users whose social relationship changed and time of updating similar users for cold-start user with incremental graph pattern matching

用户退化为重新计算冷启动用户的相似用户.

## 5 结 语

社交网络中冷启动用户不是孤立静止的, 随着时间的推移冷启动用户和其已有相似用户的社交关系都可能会发生变化, 如果不考虑用户间社交关系的变化, 还是基于原来的相似用户给出推荐则可能无法得到准确的推荐结果. 社交网络中用户的规模已经非常庞大, 无论是冷启动用户, 或者是其他用户的社交关系变更, 为冷启动用户重新匹配相似用户是代价昂贵的. 本文提出基于增量图形模式匹配的相似用户发现方法 (IGPMDCR), 在冷启动用户已有相似用户的基础上, 为冷启动用户动态更新相似用户, 进而给出实时、准确的推荐. 在包含大量冷启动用户且用户间社交关系随时间发生变化的数据集的实验结果表明, IGPMDCR 在较长的时间里可以动态的为冷启动用户更新相似用户, 为冷启动用户给出准确的推荐.

## 参考文献 (References):

- [1] BOBADILLA J S, ORTEGA F, HERNANDO A, et al. A collaborative filtering approach to mitigate the new user cold start problem [J]. *Knowledge-Based Systems*, 2012, 26(1): 225-238.
- [2] LIKA B, KOLOMVATSOS K, HADJIEFTHYMIADES S. Facing the cold start problem in recommender systems [J]. *Expert Systems with Applications*, 2014, 41(4): 2065-2073.
- [3] REN Y L, LI G, ZHOU W L. *PRICAI 2012: Trends in Artificial Intelligence* [M]. Berlin: Springer, 2012: 887-890.

- [4] LING Y X, GUO D K, CAI F, et al. User-based Clustering with Top-N Recommendation on Cold-Start Problem[C]// **Proceedings of the 2013 3rd International Conference on Intelligent System Design and Engineering Applications**. Hong Kong: IEEE Computer Society, 2013: 1585-1589.
- [5] LOPS P, DE GEMMIS M, SEMERARO G. **Recommender systems handbook** [M]. Berlin Heidelberg: Springer, 2011: 73-105.
- [6] YIN H, CUI B, CHEN L, et al. A temporal context-aware model for user behavior modeling in social media systems [C]// **Proceedings of the 2014 ACM SIGMOD international Conference on Management of Data**. Snowbird, USA: ACM, 2014: 1543-1554.
- [7] WANG J, DE VRIES A P, REINDERS M J T. Unifying user-based and item-based collaborative filtering approaches by similarity fusion [C]// **Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval**. Washington, USA: ACM, 2006: 501-508.
- [8] JAMALI M, ESTER M. A matrix factorization technique with trust propagation for recommendation in social networks [C]// **Proceedings of the 4th ACM Conference on Recommender Systems**. Barcelona, Spain: ACM, 2010: 135-142.
- [9] MA H, YANG H, LYU M R, et al. Sorec: social recommendation using probabilistic matrix factorization [C]// **Proceedings of the 17th ACM Conference on Information and Knowledge Management**. Napa Valley, USA: ACM, 2008: 931-940.
- [10] WU L, CHEN E H, LIU Q, et al. Leveraging tagging for neighborhood-aware probabilistic matrix factorization [C]// **Proceedings of the 21st ACM International Conference on Information and Knowledge Management**. Maui Hawaii, USA: ACM, 2012: 1854-1858.
- [11] KOREN Y. Collaborative filtering with temporal dynamics [J]. **Communications of the ACM**, 2010, 53 (4): 89-97.
- [12] REN L, GU J Z, XIA W W. An item-based collaborative filtering approach based on balanced rating prediction [C]// **Proceedings of 2011 International Conference on Multimedia Technology**. Hangzhou: IEEE, 2011: 3405-3408.
- [13] ZHAO W X, LI S, HE Y, et al. Connecting social media to e-commerce: cold-start product recommendation using microblogging information [J]. **IEEE Transactions on Knowledge & Data Engineering**, 2016, 28 (5): 1147-1159.
- [14] 李洋, 陈毅恒, 刘挺. 微博信息传播预测研究综述 [J]. **软件学报**, 2016(2): 247-263.  
LI Y, CHEN YH, LIU T. Survey on predicting information propagation in microblogs [J]. **Journal of Software**, 2016, 27(2): 247-263.
- [15] 赵泽亚, 贾岩涛, 王元卓, 等. 大规模演化知识网络中的关联推理 [J]. **计算机研究与发展**, 2016, 53(2): 492-502.
- ZHAO ZY, JIA YT, WANG YZ, et al. Link inference in large scale evolutional knowledge network [J]. **Journal of Computer Research and Development**, 2016, 53(2): 492-502.
- [16] REAFEE W, SALIM N, KHAN A. The power of implicit social relation in rating prediction of social recommender systems [J]. **Plos One**, 2016, 11(5): 1-20.
- [17] PENG F, LU J, WANG Y, et al. N-dimensional markov random field prior for cold-start recommendation [J]. **Neurocomputing**, 2016, 191(1): 187-199.
- [18] MA H, KING I, LYU M R. Learning to recommend with social trust ensemble [C]// **Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval**. Gold Coast, Australia: ACM, 2009: 203-210.
- [19] KIM Y A, SONG H S. Strategies for predicting local trust based on trust propagation in social networks [J]. **Knowledge-Based Systems**, 2011, 24(8): 1360-1371.
- [20] YUAN W W, GUAN D H, LEE Y K, et al. Improved trust-aware recommender system using small-worldness of trust networks [J]. **Knowledge-Based Systems**, 2010, 23(3): 232-238.
- [21] JIANG W J, WANG G J, WU J. Generating trusted graphs for trust evaluation in online social networks [J]. **Future generation computer systems**, 2014, 31(1): 48-58.
- [22] LIU R R, LIU J G, JIA C X, et al. Personal recommendation via unequal resource allocation on bipartite networks [J]. **Physica A: Statistical Mechanics and its Applications**, 2010, 389(16): 3282-3289.
- [23] GUHA R, KUMAR R, RAGHAVAN P, et al. Propagation of trust and distrust [C]// **Proceedings Of The 13th International Conference on World Wide Web**. New York, USA: ACM, 2004: 403-412.
- [24] DENG S, HUANG L, XU G. Social network-based service recommendation with trust enhancement [J]. **Expert Systems with Applications**, 2014, 41 (18): 8075-8084.
- [25] MORADI P, AHMADIAN S. A reliability-based recommendation method to improve trust-aware recommender systems [J]. **Expert Systems with Applications**, 2015, 42(21): 7386-7398.
- [26] LE H S. Dealing with the new user cold-start problem in recommender systems: A comparative review [J]. **Information Systems**, 2014, 58(1): 87-104.
- [27] WANG Y, YIN G, CAI Z, et al. A trust-based probabilistic recommendation model for social networks [J]. **Journal of Network & Computer Applications**, 2015, 55 (1): 59-67.
- [28] WANG Y, CAI Z, YIN G, et al. An Incentive Mechanism with privacy protection in mobile crowdsourcing systems [J]. **Computer Networks**, 2016, 102 (1): 157-171.