

融合用户信任度和相似度的基于核心用户抽取的鲁棒性推荐算法

赵明 闫寒* 曹高峰 刘昕鸿

(中南大学软件学院 长沙 410075)

摘要: 推荐系统可以方便地帮助人们做出决策, 然而, 目前很少有研究考虑到剔除不相关噪声用户的影响, 保留少量核心用户做推荐。该文提出基于信任关系和兴趣相似度的核心用户抽取的新方法。首先计算所有用户之间的信任度和兴趣相似度并且排序, 然后根据用户在最近邻列表中出现的频率和位置权重两种策略选择候选核心用户集合, 最后利用用户的推荐能力筛选出最终的核心用户并且做推荐。实验表明利用核心用户做推荐的有效性, 并且证明了利用20%的核心用户做推荐, 可以达到超过90%的准确性, 而且利用核心用户做推荐能很好地抵御托攻击对推荐系统造成的负面影响。

关键词: 推荐系统; 核心用户; 鲁棒性; 相似度; 信任度

中图分类号: TP391.3

文献标识码: A

文章编号: 1009-5896(2019)01-0180-07

DOI: 10.11999/JEIT180142

Robust Recommendation Algorithm Based on Core User Extraction with User Trust and Similarity

ZHAO Ming YAN Han CAO Gaofeng LIU Xinhong

(School of Software, Central South University, Changsha 410075, China)

Abstract: Recommendation systems can help people make decisions conveniently. However, few studies consider the effect of removing irrelevant noise users and retaining a small number of core users to make recommendations. A new method of core user extraction is proposed based on trust relationship and interest similarity. First, all users trust and interest similarity between pairs are calculated and sorted, then according to the frequency and position weight users travel in the nearest neighbor in the list of two kinds of strategies for the selection of candidate core collection of users. Finally, according to the user's ability the core users are sieved out. Experimental results show that the core user recommendation effectiveness, and verify that the core of user 20% can reach more than recommended accuracy of 90%, and through the use of core user recommendation the negative effects caused by the attacks on the recommendation system can be resisted.

Key words: Recommendation system; Core users; Robustness; Similarity; Trust

1 引言

随着互联网技术的发展, 各种各样的在线信息为人们提供了丰富多彩的便捷生活。然而, 随着信息的爆炸式增长, 越来越多的人发现很难迅速而有效地获得他们真正需要并且感兴趣的信息。在此背景下, 推荐系统作为解决“信息过载”问题的有效工具被提出并应用于多种在线平台。如今, 推荐系

统已经被广泛应用。例如, Amazon.com使用用户的购买记录为用户推荐图书。

推荐系统通常使用以下3种典型的推荐算法产生推荐列表: 基于内容的^[1], 协作过滤^[2,3]和情境感知方法^[4,5]。基于这些基本算法, 已经有很多扩展^[6,7]。与以往的研究和方法不同的是, 物理学的视角也被应用于推荐算法的设计中^[8-10], 其中最具代表性的是质量扩散算法和热传导算法。两种基于传播的方法的混合可以有效地解决推荐中的冷启动^[8]、多样性-准确性困境^[9]。

无论是传统的推荐算法, 还是基于二部图的推荐算法, 由于推荐系统自身的开发性和对数据的依赖性, 使得推荐系统非常容易受到恶意攻击。现如今对推荐系统的研究主要集中在设计新的推荐算法或是改进原有的推荐算法, 以提高推荐结果的准确

收稿日期: 2018-02-02; 改回日期: 2018-10-23; 网络出版: 2018-10-29

*通信作者: 闫寒 yanhan@csu.edu.cn

基金项目: 国家自然科学基金(61572526), 中南大学研究生创新项目(502211708)

Foundation Items: The National Natural Science Foundation of China (61572526), The Central South University Graduate Student Innovation Project (502211708)

率。最近的一些研究已经证明，推荐系统中存在一批核心用户，这些用户携带了更多的推荐信息。利用这些用户做推荐，已经可以产生令人满意的推荐结果。目前，对于抽取核心用户的研究中，主要有4种方法：(1)随机抽取；(2)基于节点的度的抽取；(3)基于资源的抽取；(4)基于相似度的抽取。

在推荐系统中，识别和抽取核心用户是非常有意义的，因为通过排除不相关和不可靠的用户，可以提高推荐系统的效率和鲁棒性。然而，目前，对于核心用户提取方法主要是基于图论中节点的出度，而不考虑节点之间的语义含义。此外，由于两个用户的相似度只有在对共同项目进行评分的情况下才能计算出来，因此，基于相似度的方法存在数据稀疏性问题。

为了解决这些问题，本文提出了基于用户兴趣相似度和信任关系以及利用用户的推荐能力来识别核心用户的新方法。核心用户的识别是通过3个步骤来实现的：首先，计算所有用户对之间的信任度和兴趣相似度。特别是对于兴趣相似度，本文提出基于K均值的聚类项目，然后基于相似的项目计算两个用户的兴趣相似度。因此，从信任关系和兴趣相似的角度，可以得到每个用户的最近邻居列表。其次，根据计算出的邻居列表，本文提出两种候选核心用户抽取策略。最后，利用用户的推荐能力筛选核心用户候选中真正的核心用户。本文的贡献包括：

(1) 提出了两种基于信任关系和用户兴趣相似度的核心用户提取策略。通过考虑项目之间的语义含义，提高了用户相似度的计算。同时考虑用户在社交网络中的交互数据来确定他们的信任关系。从信任关系和兴趣相似度的角度出发，根据每个用户的最近邻居列表，采取不同的策略，抽取核心用户。

(2) 通过分析核心用户候选集中用户的推荐能力，将小于推荐能力阈值的用户移除核心用户，得到真正的核心用户。通过利用少量的核心用户做推荐，大大提高系统的计算效率，并且通过剔除与目标用户不相关的用户提高系统的鲁棒性。

(3) 在实验中，对所提出的方法和现有的方法进行了比较，并从推荐的准确性，新颖性，长尾分布和用户程度分布以及对恶意攻击的抵御能力等方面对结果进行了综合分析。

本文的其余部分安排如下。第2节介绍核心用户的抽取策略。第3节介绍实验并分析实验结果。第4节总结全文并给出下一步的研究方向。

2 核心用户抽取策略

如上所述，本文的核心用户抽取方法由3个步骤组成。首先，通过计算所有用户对之间的信任度

和兴趣相似度来得到每个用户的最近邻居列表。其次，使用两种策略来选择核心用户候选集，并且通过分析用户的推荐能力来筛选真正的核心用户。第三，基于核心用户的推荐算法得到推荐结果。

一个推荐系统由一组用户 $\{u_1, u_2, \dots, u_m\}$ 和项目 $\{o_1, o_2, \dots, o_n\}$ 组成，并且可以由一个二部图网络^[8]来表示推荐系统。在最简单的情况下，每条边都具有相同的含义，表示用户和项目之间的一种交互。

2.1 基于信任度和相似度计算用户最近邻列表

2.1.1 用户信任度

首先使用社交网络中用户的信任关系来计算用户的信任度，得到用户的信任度最近邻列表。该方法不需要用户的评分数据，因此在一定程度上缓解数据稀疏问题。社交网络上的信息交流通常可以反映用户之间的信任关系，比如微信、QQ和短信等。一般来说，如果信息通讯量较大，通讯频率较高，则两个用户的信任度较高。换句话说，通信消息的长度和通信频率在一定程度上反映了用户之间的信任关系。在本文中，使用式(1)来计算两个用户之间的信任度。

$$t(u_1, u_2) = \lambda \times \frac{l_{u_1, u_2}}{\max_{u \in C_{u_1}} l_{u_1, u}} + (1 - \lambda) \times \frac{f_{u_1, u_2}}{\max_{u \in C_{u_1}} f_{u_1, u}} \quad (1)$$

$t(u_1, u_2)$ 表示用户 u_1 对用户 u_2 的信任程度。其中， C_{u_1} 表示与用户 u_1 有交流信息的用户的集合， λ 为调节因子，取值范围为 $[0, 1]$ ， l_{u_1, u_2} 表示用户 u_1, u_2 之间交流信息的长度， f_{u_1, u_2} 表示用户 u_1, u_2 之间交流的频率。不同用户之间的信任关系可能受到交流信息频率和交流信息长度的影响不同，所以利用调节因子 λ 可以对交流信息长度、交流频率赋予不同的权重来表示二者的不同影响程度。

2.1.2 基于项目聚类的用户相似度

对于传统的余弦或皮尔逊公式计算用户相似度时，根据用户-项目评分，只考虑到用户对共同项目的评分情况来计算用户之间的相似度。但是，如果两个用户不具有共同的评分，则不能计算它们的相似度。当评分数据稀少时，两个用户对相同项目的评价将会很少。显然，在这种情况下，判断这两个用户是否相似是不准确的。为了解决这个问题，本文结合项目聚类^[11]和传统的Pearson用户相似度公式来提高用户相似度的准确性。基于相似性项目的用户相似度计算为

$$u(u_1, u_2) = \frac{\sum_{k=1}^n (r_{u_1, o_k} - \bar{r}_{u_1})(r_{u_2, o_k} - \bar{r}_{u_2}) i(o_k, o_k)}{\sqrt{\sum_{k=1}^n (r_{u_1, o_k} - \bar{r}_{u_1})^2} \sqrt{\sum_{k=1}^n (r_{u_2, o_k} - \bar{r}_{u_2})^2}} \quad (2)$$

其中, r_{u_1, o_k} 是用户 u_1 对应用 o_k 的评分, r_{u_2, o'_k} 是用户 u_2 对应用 o'_k 的评分, $i(o_k, o'_k)$ 表示项目 o_k 与 o'_k 之间的相似度, \bar{r}_{u_1} 是用户 u_1 对曾经评价过的项目的平均评分, \bar{r}_{u_2} 是用户 u_2 对曾经评价过的项目的平均评分, 对于 $i(o_k, o'_k)$ 的计算, 本文中利用项目的描述信息, 通过利用Hownet语义计算工具计算项目之间的语义相似度。

2.2 融合用户的信任度和兴趣相似度

通过计算用户的信任度和相似度, 得到每个用户的信任度最近邻列表Au和相似度最近邻列表Bu。根据协同过滤算法的思想可知用户列表Au中排序越靠前的用户, 其影响力越大; 而Bu中排序越靠前的用户, 其包含的推荐信息越多。但同一用

户在两个列表中顺序可能不同, 因此需综合利用用户间的相似度和信任度, 重新排序用户在最近邻列表中的位置, 根据融合公式(3)得到新的最近邻列表Cu:

$$P_u = (1 - \alpha) \frac{N - l_A}{N - 1} + \alpha \frac{N - l_B}{N - 1} \quad (3)$$

其中, l_A 和 l_B 分别代表用户在Au和Bu中的排序序号, N 表示所有的用户数量。最终, 计算得到用户的 P_u 值, 然后根据 P_u 值大小降序排列用户, 得到列表Cu。

表1表示根据信任度公式计算的得到目标用户 u 的信任度最近邻列表Au和根据相似度公式计算得到目标用户 u 的相似度最近邻列表Bu融合得到最终的Cu。

表 1 融合结果表

Au (按照信任度大小排序)			Bu (按照相似度大小排序)			Cu (按照 P_u 值大小排序)		
序号	用户ID	信任度	序号	用户ID	相似度	序号	用户ID	$P_u (\alpha=0.4)$
1	user1	0.765210	1	user2	0.965210	1	user1	0.70
2	user4	0.582130	2	user5	0.882130	2	user2	0.60
3	user3	0.212420	3	user1	0.812420	3	user5	0.55
4	user5	0.200760	4	user3	0.700760	4	user3	0.35
5	user2	0.190855	5	user4	0.590855	5	user4	0.30

2.3 核心用户抽取策略

为了提取核心用户, 只有少数最相关的用户应该被考虑。这样, 可以减少推荐过程的计算, 并且可以减少与目标用户的相关度很小的噪声信息, 剔除部分不相关信息, 提高推荐系统的鲁棒性。本文利用两种策略^[8,9]选择核心用户。

第1种策略是选择所有在其他用户最近邻列表中出现的次数最多的用户, 并将其命名为基于频率的策略。具体过程为: 首先通过信任度、相似度和融合公式得到每个用户top-N最近邻列表Cu, 从最近邻列表Cu中选择前N邻居; 其次, 对于每个用户, 计算该用户在其他用户的前N最近邻列表中出现过多少次。最后, 选择出现频率最高的用户作为核心用户。

第2种策略是选择所有在其他用户最近邻列表中位置权重最高的用户, 并将其命名为基于位置权重的策略。该方法考虑每个用户在其他用户

的top-N的最近邻列表中出现的位置排序。假如用户 u_i 属于用户 u_j 的top-N最近邻列表Cu, 而且在最近邻列表中的位置是第 p 位, 则用户 u_i 将得到 $1/p$ 的得分。如果用户 u_i 也出现在其他用户的最近邻列表, 我们将用户 u_i 的每次得分相加得到最终的得分作为用户 u_i 的位置权重值, 具体计算公式为

$$W_{u_i} = \sum_{u_j | u_i \in N(u_j)} \frac{1}{p_{u_i, u_j}} \quad (4)$$

其中, $N(u_j)$ 表示用户 u_j 的最近邻列表(其中包含用户 u_i), p_{u_i, u_j} 表示用户 u_i 在用户 u_j 最近邻列表中出现的位置。最后, 选择用户位置权重最大的用户作为核心用户。

图1用一个例子来说明选择核心用户两种策略。对于每个用户, 选择Cu中前3名的邻居。核心用户数设置为2, 则基于频率的策略和基于位置权重的策略得到的核心用户分别为 $\{u_2, u_3\}$ 和 $\{u_2, u_1\}$ 。

最近邻列表	u_1	u_2	u_3	u_4	u_5	频率	核心用户	最近邻列表	u_1	u_2	u_3	u_4	u_5	位置权重	核心用户
u_1	u_2, u_3, u_5	0	1	1	0	1	u_2 4	u_1	u_2, u_3, u_5	0	1	1/2	0	1/3	u_2 5/2
u_2	u_1, u_3, u_4	1	0	1	1	0	u_3 4	u_2	u_1, u_3, u_4	1	0	1/2	1/3	0	u_1 7/3
u_3	u_4, u_2, u_5	0	1	0	1	1	u_1 3	u_3	u_4, u_2, u_5	0	1/2	0	1	1/3	u_3 4/3
u_4	u_1, u_2, u_3	1	1	1	0	0	u_4 2	u_4	u_1, u_2, u_3	1	1/2	1/3	0	0	u_4 4/3
u_5	u_3, u_2, u_1	1	1	1	0	0	u_5 2	u_5	u_3, u_2, u_1	1/3	1/2	1	0	0	u_5 2/3

图 1 基于频率和位置权重的核心用户抽取策略

对于抽取出的核心用户，本文通过利用用户的推荐能力进行筛选，剔除一部分推荐能力小于阈值的用户。用户推荐能力^[12]表示当前用户喜欢的项目同时也被其他用户喜欢的频率，衡量用户是否对流行、热门的项目感兴趣：

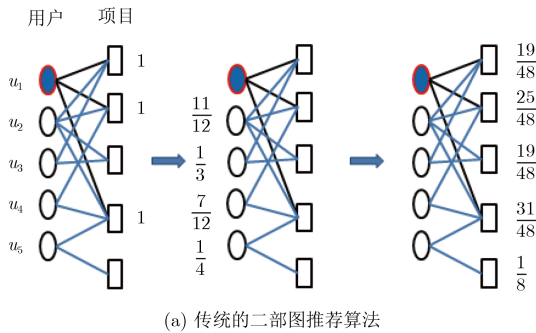
$$\text{AvgFr}(u_i) = \frac{\sum_{o_j \in I_{u_i}} \text{Fr}(o_j)}{\|I_{u_i}\|} \quad (5)$$

其中， $\text{Fr}(o_j)$ 表示评价过项目 o_j 的用户数，即项目 o_j 的流行度， I_{u_i} 表示用户 u_i 评价的项目集， $\|I_{u_i}\|$ 表示用户 u_i 评价的项目数量。将小于阈值的用户排除得到最终的核心用户集。

对于两种策略抽取出的核心用户，将形成两个核心用户集合，本文的实验部分将这两种策略进行对比分析，选出一种最优的核心用户抽取策略。

2.4 基于核心用户推荐算法

本文基于核心用户的推荐算法是利用基于二部图的推荐方法作为推荐算法。对于传统的基于二部图的推荐算法，假设每个项目分配一个单位资源，将资源分配给所有邻居用户，然后每个用户将收到的资源重新分配给所有与之相关联的项目。因此，



项目 o_α 从项目 o_β 接收到的资源是

$$S_{o_\alpha, o_\beta} = \frac{1}{k(o_\beta)} \sum_{i=1}^m \frac{1}{k(u_i)} \quad (6)$$

其中， $k(o_\beta)$ 表示项目 o_β 的度，即与项目 o_β 有关联的用户数目。 $k(u_i)$ 表示用户 u_i 的度，即与用户 u_i 有关联的所有项目数量。

对于向目标用户 u_i 做推荐时，将所有目标用户 u_i 的未关联的项目按照最终资源数降序排序后，为其推荐前 L 个项目。

对于基于核心用户的二部图推荐算法，将会有一点差异。即给定一组核心用户 C 和目标用户 u_i ，只有 C 中的用户才能从目标用户 u_i 相关联的项目中接收资源。其他非核心用户，即使与目标用户 u_i 关联相同或相似的项目也不会收到资源。然后，收到资源的用户将资源重新分配给项目端。

图2，是关于传统的基于二部图分推荐算法和基于核心用户的二部图推荐算法的对比。基于核心用户的推荐算法中，设定 u_2 和 u_4 为核心用户，从两个推荐结果来看，推荐结果是相同的，当时基于核心用户的推荐算法只使用了少量的用户做推荐。

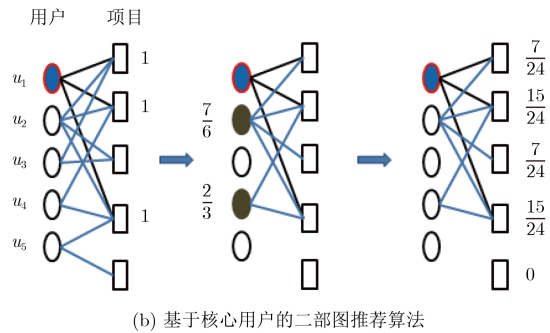


图2 传统二部图与基于核心用户的二部图推荐算法原理

3 实验结果与分析

本节介绍实验中使用的数据集，并分析实验结果。

3.1 数据集

本文使用的数据集是来自于Group Lens研究小组建立的MovieLens 100K公开数据集，此数据集是一个权威的实验数据集，被广泛应用于推荐系统领域。此数据集中包含了943个真实用户对1682个电影的真实评分数据，有100000条评分记录，并且为了减缓评分矩阵的稀疏性，记录中每个用户至少对20个电影进行评分，评分值的范围从1~5，且都是整数，1表示不喜欢此电影，5表示非常喜欢此电影。此外，通过编写爬虫程序，来获得关于电影介绍信息(用于项目相似性)和用户之间的沟通信息

(用于信任关系)的所需数据。该数据集包含4932个用户之间的19194条通信消息。为了验证推荐算法的性能和准确性，本文采用5折交叉验证的方式确定参数并对实验结果进行分析。

3.2 实验评价指标

针对推荐算法推荐结果的准确性，本节将使用召回率(R)进行评价。

$$R(L) = \frac{1}{m} \sum_{i=1}^m \frac{H_i(L)}{N_i} \quad (7)$$

召回率 R 表示推荐给目标用户的项目集合中，目标用户真正喜欢的项目占测试集中目标用户喜欢的所有项目的比例，其中， $H_i(L)$ 表示在推荐列表中，用户 i 真正喜欢的项目的个数， N_i 表示测试集中目标用户喜欢的项目个数， m 表示用户数目。

3.2.1 信任度公式中参数 λ 确定

首先要确定式(1)中计算两个用户之间信任度的值。实验中,将用户 u_i 的影响定义为与用户 u_i 关联的所有用户的信任度的平均值。对于943个用户,通过观察不同 λ 值中用户的平均影响力变化来确定参数 λ 。

如图3所示,当 $\lambda=0.5$ 时,所有用户的平均影响力达到最大值。这个结论一定程度上符合人们日常生活的情形,对于人们的日常交流行为中(比如手机,微信等交流),用户交流信息的长度和频率这两个方面,通常对用户信任关系的影响是相同的。一般来说,信任用户之间的交流信息的长度和频率呈现正相关关系,因此将权重参数 λ 设置为0.5是合理的,而且是最优的。

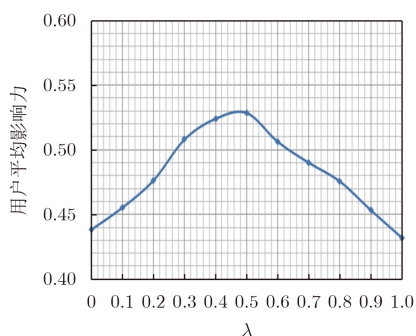


图3 不同参数 λ 值对所有用户平均影响力的影响

3.2.2 融合公式参数 α 的确定

利用信任度公式得到的信任度最近邻列表 A_u 和利用兴趣相似度得到的相似度最近邻列表 B_u ,将两个列表融合得到最终的最近邻列表 C_u ,在融合公式中,权重参数 α ,用来确定 A_u 和 B_u 对最终列表 C_u 的影响程度。

如图4所示,当 $\alpha=0.4$ 时,推荐算法的召回率达到最大值。这也表明,用户的相似度对推荐结果的影响会比用户的信任关系要大,同时说明用户的兴趣相似度在推荐系统中是至关重要的。

3.2.3 用户推荐能力阈值确定

在得到用户的最终的最近邻列表 C_u 后,我们通过分析用户的推荐能力,将推荐能力小于阈值的

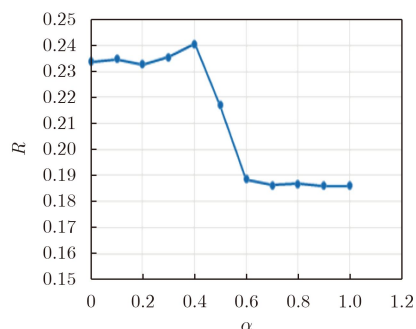


图4 不同参数 α 对推荐结果的影响

用户从最近邻列表 C_u 中剔除,图5是核心用户推荐能力的分布情况。

经统计分析,当核心用户的推荐能力不小于50时,核心用户推荐能力的分布情况比较分散,也就是说,可以用核心用户代表图5长尾分布的各个部分的用户信息(包括头部、中部和尾部)。此时核心用户最具有代表性。

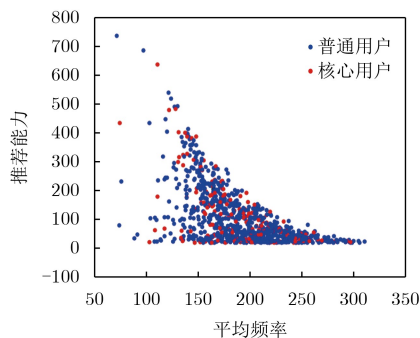


图5 核心用户分情况

3.3 实验结果与分析

(1) 分析算法的准确率: 本文实验是验证基于核心用户的鲁棒性推荐算法的效果。将本文中提出的基于核心用户的推荐算法与文献[6]中性能最好的方法进行对比,为了区分本文方法与其他方法,命名本文的方法为基于相似度-信任度的推荐算法,文献中的方法为传统的二部图推荐算法,在验证算法精确率的实验中只使用原始的数据集,并没有加入攻击用户概貌。

图6中 P 表示所有用户中核心用户的比例。当 $P=1.0$ 时,推荐算法将使用所有的用户,相当于传统的方法。一般来说,随着 P 的下降,推荐的准确率趋于下降,并且趋势是正相关的,因为基于二部图的推荐算法的可用信息越少, P 越小。在实验中,设置最近邻是top-10,推荐列表长度是15。

从图6可以明显看出,基于本文提出的相似度

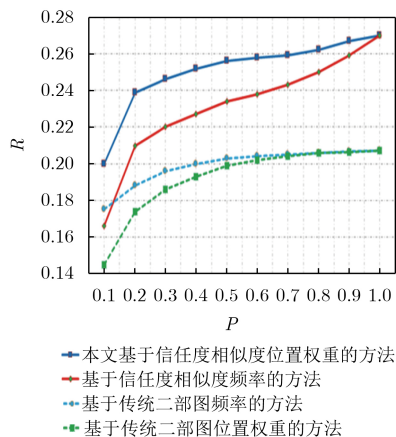


图6 算法召回率对比

和信任度融合的二部图推荐方法，推荐结果的召回率是最高的。经过统计分析，使用20%的核心用户 ($P=0.2$)做推荐，推荐结果的召回率为0.239，而使用全部的用户 ($P=1$)做推荐，相当于传统的协调过滤推荐算法，得到的推荐结果的准确率为0.264，可以看出仅仅使用20%的核心用户做推荐能达到全部用户做推荐准确率的90.39% ($0.239/0.264$)，同时，随着 P 值的增大，推荐结果的准确性增加得非常缓慢，这说明虽然利用少量的核心用户做推荐有一定的精度损失，但是推荐结果能达到90%以上的精度，可以令人满意，而且只利用少量用户算法的计算效率将成倍提高。因此，可以得出结论：利用少数携带更多推荐信息的用户做推荐，可以得到令人满意的推荐结果，同时将会大大提升推荐系统的效率。而且，通过对比分析，基于频率的策略而言，本文提出的推荐算法相比于文献中的方法平均能够提高20%的召回率，就基于位置权重的策略而言，本文提出的推荐方法相比于文献中的方法平均能够提高26%的召回率。

(2) 分析算法托攻击防御能力：就本文提出的基于核心用户抽取的推荐算法的鲁棒性分析，此实验是通过模拟实验进行的，实验中根据托攻击模型的原理，在原始数据集中人工生成托攻击用户数

据，其中，对目标项目的攻击意图是推攻击，攻击强度为10%，填充规模为5%。

如图7所示，本文提出的基于核心用户做推荐得到的推荐结果的召回率，通过对比系统中是否有托攻击用户，发现本文提出的基于核心用户的推荐算法对随机托攻击和均值托攻击能够起到很好的抵御作用。系统中是否存在托攻击用户，利用本文提出基于核心用户算法做推荐，得到推荐结果的召回率几乎没有差别，这是因为利用本文提出的基于信任度-相似度的核心用户抽取策略，能够有效地过滤托攻击用户，使得选取出来的核心用户中不包含托攻击用户，在利用核心用户做推荐时，托攻击(包括随机攻击，均值攻击，流行攻击和分段攻击)将不起作用，这也说明了本文算法对托攻击的防御效果是非常好的。

(3) 分析核心用户的合理性：在实验数据集中，共有943个用户，其中50%的用户流行度低于52，30%在52和138之间，其余的20%大于138。从图8可以看出，在高流行度占比下，基于二部图位置权重方法和频率方法选出的核心用户比基准值分别高出57.37%和42.11%，也就是说这两种方法更偏向于选择活跃的用户(观看电影次数多的用户)。对于本文提出的两种策略中，基于信任度和相似度的频

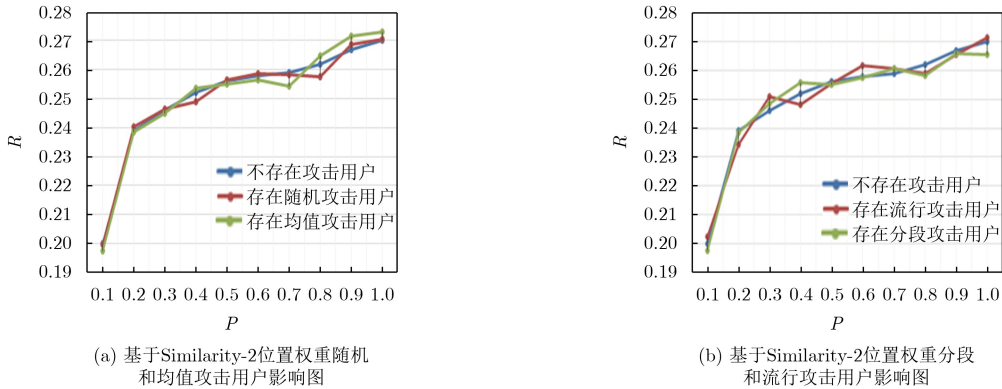


图7 攻击用户是否存在对推荐结果召回率的影响

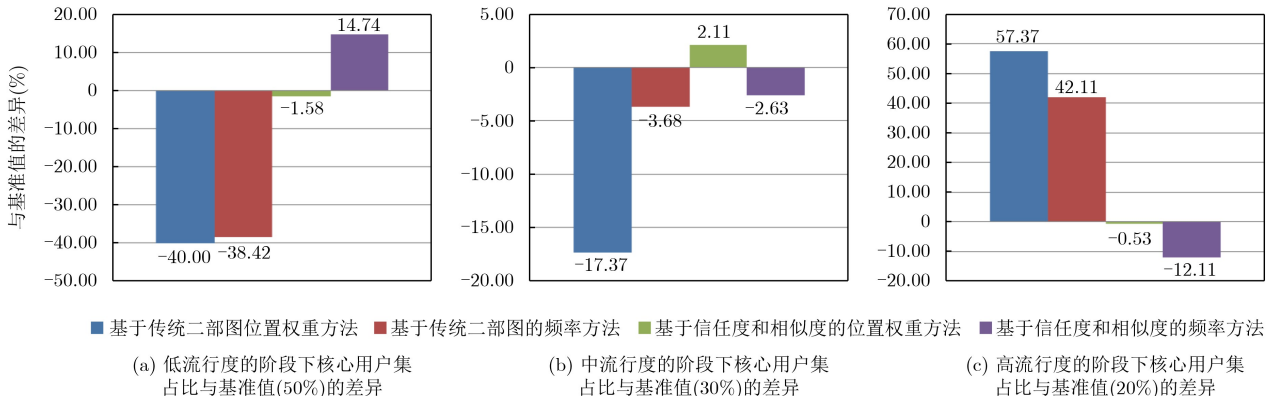


图8 核心用户在不同阶段不同方法的流行度与基准值的差异

率的方法选出的核心用户中,超过60%的用户流行度小于52,这是因为利用用户的推荐能力进行筛选操作之后,将一部分流行度大的用户剔除,使得基于信任度和相似度的频率的方法选出的核心用户分布不是非常合理。而基于信任度和相似度的位置权重排序的方法选出的核心用户是最合理的,因为选出的核心用户的流行度分布在3个流行度区间的比例分别接近50%,30%和20%。

4 结束语

本文提出了基于信任关系和兴趣相似度提取核心用户的新方法。首先定义两个用户之间的信任和相似度。对于信任度,本文从社交网络中提取用户之间的信任关系。对于相似度,将项目相似度合并到用户相似度的计算中。然后使用基于频率的和基于等级的策略来选择核心用户,并且通过用户的推荐能力进行最终的筛选过滤。发现核心用户通常出现在许多用户的排名靠前的前 N 位邻居列表中。核心用户通常携带大部分推荐信息,只能利用核心用户来达到令人满意的推荐精度,而且基于核心用户的推荐算法对托攻击具有一定的抵御能力。

未来的工作包括:(1)从多个方面寻找更多的方法来定义核心用户,例如考虑用户的兴趣迁移;(2)寻找一种相对稳定的核心用户生成方式,以减少更新频率;(3)对不同的数据集进行更多的实验,比较更多的方法。

参考文献

- [1] TAO Peng, WANG Wendong, GONG Xiangyang, *et al.* A graph indexing approach for content-based recommendation system[C]. IEEE International Conference on Multimedia & Information Technology, Kaifeng China, 2010: 93–97. doi: [10.1109/MMIT.2010.84](https://doi.org/10.1109/MMIT.2010.84).
 - [2] 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](https://doi.org/10.3724/SP.J.1146.2012.01743).
WANG Yubin, MENG Xiangwu, and HU Xun. A collaborative filtering recommendation algorithm based on information aging[J]. *Journal of Electronics & Information Technology*, 2013, 35(10): 2391–2396. doi: [10.3724/SP.J.1146.2012.01743](https://doi.org/10.3724/SP.J.1146.2012.01743).
 - [3] 王海艳, 张大印. 一种可信的基于协同过滤的服务选择模型[J]. 电子与信息学报, 2013, 35(2): 349–354. doi: [10.3724/SP.J.1146.2012.00946](https://doi.org/10.3724/SP.J.1146.2012.00946).
WANG Haiyan and ZHANG Dayin. A trusted service selection model based on collaborative filtering[J]. *Journal of Electronics & Information Technology*, 2013, 35(2): 349–354. doi: [10.3724/SP.J.1146.2012.00946](https://doi.org/10.3724/SP.J.1146.2012.00946).
 - [4] ADOMAVICIUS G and TUZHILIN A. Context-Aware Recommender Systems[M]. USA: Springer, 2015: 2175–2178. doi: [10.1007/978-0-387-85820-3_7](https://doi.org/10.1007/978-0-387-85820-3_7).
 - [5] PANNIELLO U, TUZHILIN A, and GORGOGLIONE M. Comparing context-aware recommender systems in terms of accuracy and diversity[J]. *User Modeling and User-Adapted Interaction*, 2014, 24(1/2): 35–65. doi: [10.1007/s11257-012-9135-y](https://doi.org/10.1007/s11257-012-9135-y).
 - [6] ZENG Wei, ZENG An, LIU Hao, *et al.* Uncovering the information core in recommender systems[J]. *Scientific Reports*, 2014, 4: 6140–6148. doi: [10.1038/srep06140](https://doi.org/10.1038/srep06140).
 - [7] 徐风苓, 孟祥武, 王立才. 基于移动用户上下文相似度的协同过滤推荐算法[J]. 电子与信息学报, 2011, 33(11): 2785–2789. doi: [10.3724/SP.J.1146.2011.00384](https://doi.org/10.3724/SP.J.1146.2011.00384).
XU Fengling, MENG Xiangwu, and WANG Licai. Collaborative filtering recommendation algorithm based on context similarity of mobile users[J]. *Journal of Electronics & Information Technology*, 2011, 33(11): 2785–2789. doi: [10.3724/SP.J.1146.2011.00384](https://doi.org/10.3724/SP.J.1146.2011.00384).
 - [8] LU Linyuan and LIU Weiping. Information filtering via preferential diffusion[J]. *Physical Review E*, 2011, 83(6): 066119. doi: [10.1103/PhysRevE.83.066119](https://doi.org/10.1103/PhysRevE.83.066119).
 - [9] ZHOU Tao, ZOLTÁN K, LIU Jianguo, *et al.* Solving the apparent diversity-accuracy dilemma of recommender systems[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2010, 107(10): 4511–4515. doi: [10.1073/pnas.1000488107](https://doi.org/10.1073/pnas.1000488107).
 - [10] XU Fuqing, WANG Zhiwu, TANG Li, *et al.* A mass diffusion-based interpretation of the effect of total solids content on solid-state anaerobic digestion of cellulosic biomass[J]. *Bioresource Technology*, 2014, 167(3): 178–185. doi: [10.1016/j.biortech.2014.05.114](https://doi.org/10.1016/j.biortech.2014.05.114).
 - [11] 黄武汉, 孟祥武, 王立才. 移动通信网中基于用户社会化关系挖掘的协同过滤算法[J]. 电子与信息学报, 2011, 33(12): 3002–3007. doi: [10.3724/SP.J.1146.2011.00364](https://doi.org/10.3724/SP.J.1146.2011.00364).
HUANG Wuhan, MENG Xiangwu, and WANG Licai. Collaborative filtering algorithm based on user social relationship mining in mobile communication network[J]. *Journal of Electronics & Information Technology*, 2011, 33(12): 3002–3007. doi: [10.3724/SP.J.1146.2011.00364](https://doi.org/10.3724/SP.J.1146.2011.00364).
 - [12] CHANG Na, IRVAN M, and TERANO T. An item influence-centric algorithm for recommender systems[C]. 11th International Conference on Distributed Computing and Artificial Intelligence, Salamanca, Spain, 2014: 553–560.
- 赵 明: 男, 1957年生, 博士后, 教授, 博士生导师, 研究方向为群智感知、无线传感器网络等。
闫 寒: 女, 1991年生, 硕士生, 研究方向为无线网络和数据挖掘等。
曹高峰: 男, 1991年生, 硕士生, 研究方向为推荐系统和数据挖掘等。
刘昕鸿: 男, 1995年生, 硕士生, 研究方向为机器学习和数据挖掘等。