

基于信任环的用户冷启动推荐

杨圩生 罗爱民 张萌萌

(国防科学技术大学信息系统与管理学院 长沙 410073)

摘要 近年来,为了解决推荐系统的用户冷启动问题,信任推荐技术得到了长足发展。然而,传统的信任推荐技术在处理信任关系上比较粗糙。基于信任环的推荐思想严格控制了信任度对推荐结果的影响。实验结果表明,该方法能有效解决用户冷启动问题,并提高推荐的准确率。

关键词 信任环,用户冷启动,推荐

中图分类号 TP31 **文献标识码** A

Trust-circle Based Recommendation on User Cold-start

YANG Wei-sheng LUO Ai-min ZHANG Meng-meng

(College of Information Systems and Management, National University of Defense Technology, Changsha 410073, China)

Abstract In recent years, to solve the problem of user cold-start on recommendation system, the technology of trust-based recommendation has got rapid development. However, in transacting the trust relationship, traditional trust-based recommendation technologies seem to be comparatively rough. The idea of trust-circle based strictly controls the effects of trust degree on the results of recommendation. As experimental results show, this method can not only efficiently settle the user cold-start problem, but also improve the precision of recommendation.

Keywords Trust-circle, User cold-start, Recommendation

1 引言

随着互联网和信息技术的不断发展和普及,人们通过各种终端设备随时随地地参与网络活动,使得互联网信息的数量呈指数级增长,导致了信息爆炸。信息爆炸使得人们很难从海量的信息中寻找符合自身个性化需求的信息。因此,个性化推荐应运而生。

传统的推荐技术主要包括基于内容的推荐^[1]和协同过滤推荐^[1-3]两类。基于内容的推荐技术是根据历史信息,如评价、分享、收藏过的文档等,构造用户偏好模型,并将属性相似度高的产品向用户做出推荐。协同过滤推荐技术是根据用户-项目效用矩阵,寻找相似用户,并将相似用户评分高的产品向用户做出推荐。可以看出,基于内容的推荐技术从产品角度出发,寻找相似产品;而协同过滤推荐技术从用户角度出发,寻找相似用户。然而,传统的推荐技术面临着数据稀疏性、冷启动等问题^[4]的考验。

冷启动问题包括项目冷启动问题和用户冷启动问题两个方面,可以说是数据稀疏性问题的两个极端。简言之,项目冷启动是指在网络商品中由于被选择次数少,人气度不高而难以向用户做出推荐的新商品。用户冷启动是指刚参与网络活动,没有可追溯的网络行为信息的新用户。由于缺少用户对项目的评分信息,传统的协同过滤推荐技术无法得知目标用户的相似用户,因此,很难对其进行预测评分和推荐。

为了解决用户冷启动问题,近年来,一些学者提出了基于用户之间信任关系的推荐思想^[6-14]。这种推荐思想考虑了网络用户之间的联系,根据用户的直接或间接信任用户预测其对项目的评分。然而,目前的信任推荐算法忽视了用户之间信任关系的紧密程度对推荐结果的影响。

为了更好地实现个性化推荐,充分利用网络社会用户之间丰富的关系数据,针对用户冷启动问题,本文提出了基于信任环的推荐思想。该方法在分析网络社会用户之间信任关系的基础上,利用信任度这一定量化的指标构建多层信任环,根据信任用户对产品项目的评分信息来推测目标用户对产品项目的评分,解决了用户冷启动问题。

2 相关研究

2.1 用户冷启动

协同过滤推荐技术考虑了用户与用户之间的关系,这样的关系是通过分析不同用户对相同产品的评分信息,来确定具有兴趣相似性的用户。然而,通过分析用户的网络行为,发现大部分的用户很少对其浏览过的产品进行评价,即使有少部分的用户存在较多的评分信息,但用户共同评分的产品也是少之又少,有些用户甚至没有任何评分信息,这种数据的缺失即是用户冷启动问题的表现。如表1所列,用户 u_4 没有对任何项目进行评分,即为冷启动用户。

表1 用户冷启动(u_i)

$r(u_i, i_j)$	i_1	i_2	i_3	i_4	i_5
u_1	3		5	4	
u_2	1	4		2	4
u_3	3		4	4	3
u_4					

孙冬婷等^[5]分析了冷启动问题的产生原因,重点总结了解决冷启动问题的算法现状,分析比较了不同算法的性能差异和各自存在的优缺点,为研究者解决冷启动问题提供了参考。

2.2 信任推荐

互联网上的用户有着千丝万缕的联系,朋友之间的信任关系为解决用户冷启动问题提供了可能。

Paolo Massa 等^[6]通过挖掘 epinions 网站上用户之间的信任关系,利用信任用户对产品的评分信息预测新用户对产品的可能评分,根据评分高低决定是否对用户做出推荐。然而,该方法所考虑的信任关系仅包含信任与不信任两种类型,即信任度为 1 或 0。

Simon Meyffret 等^[7]考虑了信任关系和用户相似度对预测结果的联合影响,将信任度的取值归一化为 $[0, 1]$,但是没有考虑信任度低的用户对预测结果的负面影响。

Samaneh Moghaddam 等^[9]建立了一个二维的动态信任模型,从可探索性和可信任性两个角度分别测度用户对产品项目评分的可得性以及用户之间的信任关系。同时,通过用户反馈机制来修改用户之间的信任关系,以达到优化模型的目的。但是,在实际的电子商务网站中,用户提供的反馈信息是极其稀少的。

Hao Ma 等^[10]提出了一种基于要素分析的最优化算法,考虑了信任关系和不信任关系对推荐结果的联合影响。但是,如何获取并表示不信任关系是该方法得到实际检验与应用的瓶颈之一。

总的来说,国外学者对信任推荐的研究比较深入,充分挖掘了网络社会用户之间的信任关系。但是,对于信任关系的定量化运用还有待进一步研究。

3 信任环

不同领域的学者对信任的定义不尽相同。在电子商务网站中,信任关系不仅仅表现为网络个体之间片面的情感联系,还表现为某个网络个体对另一网络个体行为的可接受程度,一般是通过分析网络交互行为形成的。信任关系是解决冷启动问题、实现推荐的重要依据,信任度越高,推荐的结果越符合顾客需求。现给出如下定义:

定义 1 信任度是指用户之间信任关系的紧密程度,是一个用户对另一个用户的信任等级。本文将信任度的取值规定为 $[0, 1]$ 。

定义 2 信任路径长度是指在信任关系中,从用户 u_i 到用户 u_j 的边数。两个用户之间的可达路径可能有多种,本文只考虑最短信任路径长度,记为 $d(u_i, u_j)$ 。

若用户 u_i 对 u_j 存在信任关系,则表示为 $t(u_i, u_j)$ 。本文将信任关系分为直接信任关系和间接信任关系,直接信任关系表示用户对其朋友的信任度,间接信任关系表示用户对朋友的朋友的信任度,如图 1 所示。

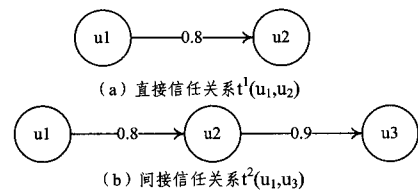


图1 信任关系分类

根据信任关系的分类和信任路径长度的大小,将信任路径长度为 k 的用户之间的信任度记为 $t^k(u_i, u_j)$ 。当 $k=1$ 时,表示具有直接信任关系的用户之间的信任度,其表示为 $t^1(u_i, u_j) = t(u_i, u_j)$; 当 $k>1$ 时,用户 u_i 和 u_j 之间的间接信任度表示为:

$$(1) \text{ 当 } k=2 \text{ 时, } t^k(u_i, u_j) = t(u_i, u_1) \cdot t(u_1, u_j)。$$

$$(2) \text{ 当 } k>2 \text{ 时, } t^k(u_i, u_j) = t(u_i, u_1) \cdot \left(\prod_{m=1}^{k-2} t(u_m, u_{m+1}) \right) \cdot t(u_{k-1}, u_j)。$$

其中,用 $F(u_i, u_j) = \{u_1, u_2, \dots, u_{k-1}\}$ 表示在最短路径长度为 k 的前提下,用户 u_i 到用户 u_j 之间的朋友集合。

以图 1 为例,图 1(a)中用户 u_1 对用户 u_2 的直接信任度 $t(u_1, u_2) = 0.8$; 图 1(b)中用户 u_1 对用户 u_3 的间接信任度 $t^2(u_1, u_3) = 0.8 \times 0.9 = 0.72$ 。

定义 3 信任环是指以用户 u_0 为中心,以信任关系 (u_0, u_j) 为纽带,按照信任路径长度由小到大的顺序,由信任路径长度相同的用户构成的环。根据信任环的规模,可以分为单层信任环和多层信任环。用 u_{ij} 表示第 i 层信任环的第 j 个用户,如图 2 所示。

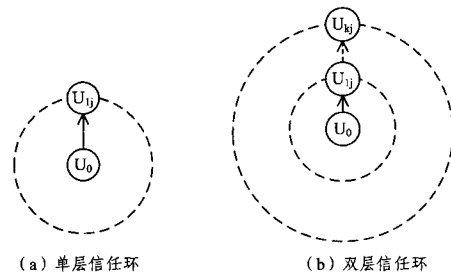


图2 信任环

为了提高信任推荐的准确率,根据信任关系的紧密程度,本文为每一层信任环都设立了信任度阈值,将第 i 层的信任度阈值记为 T^i ,即只考虑用户 u_0 的第 $i-1$ 层朋友对第 i 层朋友的信任度不小于信任度阈值的用户。由于信任的可靠性会随着信任的传递逐渐衰减,信任度阈值应满足递增性。

现将本文用到的二元关系定义如下:

(1) 用户 u 与其朋友 f 之间的信任关系表示为 $t(u, f)$;

(2) 若用户 u 对项目 i 存在评分,则表示为 $r(u, i)$;

(3) 若用户 u 对项目 i 不存在评分,可通过信任关系预测评分,则表示为 $\hat{r}(u, i)$ 。

其中,将朋友分为“有效朋友”和“无效朋友”两类。有效朋友是指 u 对 f 的信任度不小于信任阈值,且 f 对项目 i 存在评分或可得到预测评分的朋友;无效朋友是指 u 对 f 的信任度小于信任阈值或不可得到 f 对项目 i 评分的朋友。

根据信任环的定义及其结构,用户对项目的预测评分分为以下 3 种情况:

(1) 当用户对项目存在评分时, $\hat{r}(u_0, i_0) = r(u_0, i_0)$;

(2) 当用户对项目不存在评分,且存在对项目进行评分或可得到预测评分的信任度不小于信任阈值的用户时,考虑信任长度为 k ,从 u_0 的信任用户 u_{1j} 评估 $\hat{r}(u_0, i_0)$;

(3) 当不满足以上两种情况时,无法得知用户对项目的评估,记为 \perp 。

根据上文分析,将第 i 层的有效朋友表示为:

$$\Phi^i(u_0, t, i_0) = \{f \in F_{u_0}^i \mid \exists r(f, i_0) \wedge t(F_{u_0}^{i-1}, f) \geq T^i\}$$

式中, $F_{u_0}^i$ 表示用户 u_0 的第 i 层信任环上的朋友集合。

根据信任环采取由内向外分析、由外向内计算的方法,可得第 i 层、第 j 个用户 u_{ij} 对项目 i_0 的预测评分如式(1):

$$\hat{r}^i(u_{ij}, i_0) = \begin{cases} r(u_{ij}, i_0), & \exists r(u_{ij}, i_0) \\ \frac{\sum_{f \in \Phi^{i+1}(u_0, t, i_0)} t^{i+1}(u_{ij}, f) \cdot \hat{r}^{i+1}(f, i_0)}{\sum_{f \in \Phi^{i+1}(u_0, t, i_0)} t^{i+1}(u_{ij}, f)}, & \exists r(u_{ij}, i_0) \wedge \Phi^{i+1}(u_0, t, i_0) \neq \emptyset \\ \perp, & \text{otherwise} \end{cases} \quad (1)$$

那么,根据式(1),可以得到第一层信任环的有效朋友对项目的预测评分。则在信任路径长度为 k 的情况下,用户对项目的评分预测如式(2):

$$\hat{r}_k(u_0, i_0) = \begin{cases} r(u_0, i_0), & \exists r(u_0, i_0) \\ \frac{\sum_{f \in \Phi^1(u_0, t, i_0)} t^1(u_0, f) \cdot \hat{r}^1(f, i_0)}{\sum_{f \in \Phi^1(u_0, t, i_0)} t^1(u_0, f)}, & \exists r(u_0, i_0) \wedge \Phi^1(u_0, t, i_0) \neq \emptyset \\ \perp, & \text{otherwise} \end{cases} \quad (2)$$

当 $k=1$ 时,即为利用直接信任关系得到用户 u_0 对项目 i_0 的评分。

4 案例

以图3所示的信任环模型为例,其信任路径长度 $k=2$,各信任环的信任度阈值设定为 $T^1=0.7, T^2=0.8$,各用户 u 对项目的评分如表2所列,现预测用户 u_0 对项目 i_0 的评分。

表2 $r(u, i_0)$

u	u_0	u_{11}	u_{12}	u_{13}	u_{14}	u_{15}	u_{21}
$r(u, i)$	x	0.8	\perp	\perp	0.9	\perp	0.6
u	u_{22}	u_{23}	u_{24}	u_{25}	u_{26}	u_{27}	u_{28}
$r(u, i)$	1	0.8	0.9	0.8	0.8	0.7	0.8

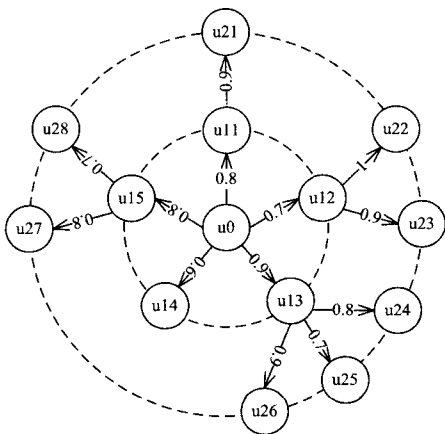


图3 信任环案例模型

结合信任度阈值的设定,用户 u_0 的第 2 层信任环上的有效朋友集合为 $\{u_{21}, u_{22}, u_{23}, u_{24}, u_{26}, u_{27}\}$, 第 1 层信任环上的有效朋友集合为 $\{u_{11}, u_{12}, u_{13}, u_{15}\}$ 。

根据式(1)得到第一层用户对项目的预测评分,如表3所列。

表3 $\hat{r}(u, i_0)$

u	u_{11}	u_{12}	u_{13}	u_{14}	u_{15}
$\hat{r}(u, i_0)$	0.8	0.91	0.85	\perp	0.7

其中,由于 $t(u_0, u_{14}) < T^1$,即用户 u_0 对其朋友 u_{14} 的信任度小于信任阈值,故 u_{14} 为无效朋友,不考虑其对项目的评分。

根据式(2)得到用户 u_0 对项目 i_0 的评分为:

$$\hat{r}(u_0, i_0) = \frac{0.8 \times 0.8 + 0.91 \times 0.7 + 0.85 \times 0.9 + 0.7 \times 0.8}{0.8 + 0.7 + 0.9 + 0.8} = 0.81$$

5 实验与评价

5.1 Epinions

文献[5]提供了采集于电子商务网站 www. Epinions. com 的数据集及其特征。本实验使用的 Epinions 数据集包含有 49290 个用户、139738 个不同的产品项;用户之间的二元信任关系共有 487181 条;用户对产品项的评分信息共有 664824 条。该数据集有超过 53% 的用户对产品项的评分信息少于 5 条,具有较强的用户冷启动性。

5.2 RMSE

本文将均方根误差 RMSE 作为检验预测质量的评价指标, RMSE 定义为:

$$RMSE = \sqrt{\frac{\sum_i (\hat{r}(u_j, i) - r(u_j, i))^2}{N}}$$

式中, $r(u_j, i)$ 表示用户对项目的实际评分, $\hat{r}(u_j, i)$ 表示用户 u_j 对项目 i 的预测评分, N 表示测试的总评分数。

5.3 实验结果

由于 Epinions 数据集中用户之间的信任关系只存在信任与不信任,即信任度为二元值 1(信任)和 0(不信任),故设定各层信任度阈值均为 1。当最短信任路径长度 $k=\{1, 2, 3\}$ 时,实验结果如图4所示。

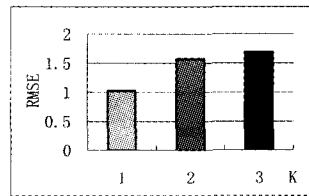


图4 $RMSE-k=\{1, 2, 3\}$

通过实验发现:利用直接信任关系($k=1$)进行预测评分的误差较小,推荐的准确性较好,而随着信任路径长度 k 的增加,均方根误差越大,即用户预测评分与实际评分相差越大,则推荐的准确性越差。分析原因可知,利用间接信任关系预测评分,随着信任路径长度 k 的增大,由朋友集合得出的用户之间的信任关系,即信任的传递性越不可靠,导致预测评分的准确性降低。举例而言,用户 A 信任用户 B,而用户 B 信任用户 C,通过用户 B 得到用户 A 对用户 C 的信任关系是不可靠的。因此,由用户 C 通过用户 B 得到用户 A 对某一项目的预测评分与用户 A 对该项目的实际评分可能存在较大差异。

(下转第 397 页)

2.3 控制变更的策略

1)做好需求分析,在需求调研阶段,软件开发人员必须认真分析和整理用户的要求,明确用户的需求后,一定要将其编写成软件需求说明书,并对说明书进行评审,让双方负责人签字确认,以免在项目实施的过程中频繁进行需求变更。

2)加大评审力度,不论在申请阶段还是在执行阶段都要对变更申请单进行评审,确定哪些变更是必须做的,哪些变更可以不做,总之,要尽量减少项目的变更。

3)沟通交流,项目组内部要定期开会,当项目进展到一定阶段要积极和用户方交流,保证尽早发现问题。

结束语 软件配置管理工作是一个不断改进的过程,要求管理者、项目经理、配置管理人员等相关人员在长期的工作过程中关注和掌控配置管理的实施细节,在实践中不断积累、调整和完善。在实施过程中认真总结经验教训,持续改进,只有这样才能提高软件过程能力成熟度,保证软件的开发工作更加有序和高效,才能使软件产品的质量得到最大程度的保障。

随着军用软件质量管理体系的日渐成熟,武器装备研制方已逐步意识到,武器装备软件在研制和信息化管理过程中面临的不仅仅是单纯的技术问题,更多的是管理上的问题。同时在生产实践中意识到:在软件开发过程中实施配置管理不能一蹴而就,要有计划、有步骤、有监控地进行,在现在的环境基础上选取容易上手的切入点,同时要为今后的扩展留下空间,即“近期着手、远期着眼”。

(上接第 365 页)

结束语 本文提出的基于信任环的推荐思想主要有以下两点特性:

(1)利用网络社会用户之间丰富的关系数据,在一定程度上解决了由于缺少评分信任而导致的用户冷启动问题。

(2)通过设定信任度阈值,充分考虑了用户间信任关系的紧密程度对推荐结果的影响,信任度阈值越大,推荐的准确性就越高。

同时,本文的研究存在以下几点不足:

(1)本文只考虑了最短信任路径为一条的情况,对于多条最短信任路径的情况有待分析。

(2)本文只分析了用户之间信任关系对推荐结果的影响,而对于网络社会中信任的特征没有深入思考,如对称性、传递性等。

(3)用户信任度和用户相似度对推荐结果的联合影响需进一步研究。

参考文献

- [1] Bunke H, Wang P S P. Personalization Techniques and Recommender Systems [M]. Singapore: World Scientific Publishing Co. Pte. Ltd., 2008
- [2] You Wen, Ye Shui-sheng. A survey of collaborative filtering algorithm applied in E-commerce recommender system [J]. Computer Technology and Development, 2006, 16(9): 70-72
- [3] Herlocker J, Konstan J, Terveen L. Evaluating collaborative filtering recommend system [J]. ACM Trans on Information system, 2004, 22(1): 5-53
- [4] 周涛. 个性化推荐的十大挑战 [J]. 中国计算机学会通讯, 2012, 8(7): 48-61
- [5] 孙冬婷, 何涛, 张福海. 推荐系统中的冷启动问题研究综述 [J].

参考文献

- [1] 徐晓春, 李高健. 软件配置管理 [M]. 北京: 清华大学出版社, 2002
- [2] 胡斌. 变更信息管理机制及其在 PDM 变更管理中的应用 [J]. 中国全学位数据库, 2004, Y581843
- [3] Donaldson S E, Siegel S G. 成功的软件开发 [M]. 蔡愉祖, 邓本江, 译. 北京: 机械工业出版社, 2003
- [4] Green, Charles T. A study of a practical method for software process improvement using a process change management system, Dissertation Abstracts International, Section: B, Director: Kenneth M. Anderson, 2003; 3364
- [5] Yang J. 软件配置管理基础 [DB/OL]. PMT 评论, <http://www.sawin.com.cn/satech.asp?class=SCM>
- [6] 陈越. 软件配置管理过程及其关键活动 [DB/OL]. <http://www.sawin.com.cn/satech.asp?class=SCM>
- [7] Software Configuration Management Process, Warren Mosely, Texas Instruments Resident Affiliate at the Software Engineering Institute Carnegie Mellon University [DB/OL]. http://www.sweforum.net/scm/ieee_1074_cm_process.pdf
- [8] Software Configuration Management Guidebook NASA-GB-9503 [S]. National Aeronautics and Space Administration, 1995, 8
- [9] 祝军生. 装备软件研制质量管理中存在的问题及对策 [J]. 军用标准化, 2005 (1)
- [10] 张永锋. 变更全攻略 [J]. 软件世界, 2007(2.5)

计算机与现代化, 2012, 5: 59-63

- [6] Massa P, Avesani P. Trust-aware recommender systems [C]// Proceedings of the 2007 ACM Conference on Recommender Systems. New York, ACM, USA, 2007; 17-24
- [7] Meyffret S, Medini L, Laforest F. User-centric trust-based recommendation [J]. IEEE computer society, 2012; 707-713
- [8] Meyffret S, Medini L, Laforest F. Trust-based Local and Social Recommendation [C]// Proceedings of the 2012 ACM Conference on Recommender Systems. Dublin, Ireland, ACM, 2012; 53-60
- [9] Moghaddam S, Jamali M, Ester M. FeedbackTrust: Using Feedback Effects in Trust-based Recommendation Systems [C]// Proceedings of the 2009 ACM Conference on Recommendation Systems. New York, USA, ACM, 2009; 269-272
- [10] Ma Hao, Lyu M R, King I. Learning to Recommendation with Trust and Distrust Relationships [C]// Proceedings of the 2009 ACM Conference on Recommendation Systems. New York, USA, ACM, 2009; 189-196
- [11] Zhou Li-xin. Trust based recommendation system with social network analysis [J]. IEEE, 2009
- [12] Ma Hao, Zhou Deng-yong, Liu Chao. Recommender systems with social regularization [C]// Proceedings of the 2011 ACM Conference on Recommender Systems. HongKong, ACM, 2011; 287-296
- [13] Yang Xi-wang, Steck H, Liu Yong. Circle-based recommendation in online social networks [C]// Proceedings of the 2012 ACM conference on Recommender systems. ACM, Beijing, China, 2012; 1267-1275
- [14] Yuan Wei-wei, Han Y, Guan D. Efficient routing on finding recommenders for trust-aware recommendation systems [C]// Proceedings of the 2012 ACM conference on Recommender systems. ACM, Kuala Lumpur, Malaysia, 2012