

基于混合好友路径信任度的社交好友推荐算法

陈思憬¹ 骆冰清² 孙知信¹

(1.南京邮电大学 物联网学院,江苏 南京 210000;

2.南京邮电大学 通信与信息工程学院,江苏 南京 210000)

摘要:为了解决以往好友信任度计算准确性不足的问题,提出依据三度以内好友路径信任度的好友推荐算法。与传统的依据好友信任度的好友推荐算法相比,传统的信任度算法仅仅考虑二度好友路径对用户之间好友信任度的影响,而忽略了三度好友路径的作用。根据三度影响力理论,在以往依据二度好友路径信任度推荐算法的基础上,引入三度好友路径信任度对好友推荐效果的影响能得到更准确的信任度计算效果。计算用户间的混合好友路径信任度,并依据混合好友路径信任度进行好友推荐,其好友推荐效果更准确。实验结果表明,比起没有引入三度好友路径的信任度算法,引入三度好友路径信任度的混合好友路径信任度算法其F1-measure评价指标增加了3到4个百分点,能有效提高好友信任度计算的准确性以及好友推荐的有效性。

关键词:社交网络;三度影响力;好友路径;信任度;好友推荐

中图分类号:TP391

文献标识码:A

文章编号:1673-629X(2018)02-0074-04

doi: 10.3969/j.issn.1673-629X.2018.02.017

Social Friend Recommendation Algorithm Based on Trust of Paths between Mixed Friends

CHEN Si-jing¹ LUO Bing-qing² SUN Zhi-xin¹

(1.School of Internet of Things, Nanjing University of Posts and Telecommunications, Nanjing 210000, China;

2.School of Communication and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210000, China)

Abstract: To address the previous inaccurate calculation of trust level between friends, we propose a friend recommendation algorithm according to trust of paths between friends within three degrees. Traditional trust-based friend recommendation algorithm only considered the influence of two-degree path on the trust between users and ignored the function of three-degree path of trust. Based on the three degrees of influence rule and the trust of two-degree friend path, this algorithm introduces the influence of three-degree trust path on friend recommendation, calculating the trust between mixed friends and performing friend recommendation in accordance with the trust of path between mixed friends. Compared to the case without introducing the new algorithm, the experiment shows that there is a three to four percentage increase of recommendation in the evaluation index of F1-measure for the social friend recommendation algorithm based on trust level of path between mixed friends, which can improve the accuracy of calculating trust between friends and the effectiveness in friend recommendation.

Key words: social media; three degrees of influence; friend path; trust; friend recommendation

0 引言

近年来,中国互联网发展迅猛,腾讯、新浪等社交网络应用的成功案例层见叠出。面对庞大的数据,如何为用户推荐好友成为社交网络应用市场角力的重点,衍生出了许多实用技术^[1]。目前好友推荐算法的主要依据有两类:基于用户之间的相似性和基于用户

之间的信任度。对于如何依据用户之间的信任度进行好友推荐,学界进行了大量研究^[2-11]。

唐晓波等^[2]提出了一种基于复杂信任网络的社会化媒体好友推荐模型,通过构建用户的信任网络体系,推荐信任度高的用户潜在好友。朱强等^[5]提出了一种基于信任度的协同过滤推荐算法,在算法中融入信任

收稿日期:2017-01-23

修回日期:2017-05-25

网络出版时间:2017-10-19

基金项目:国家级大学生创新训练计划支持项目(SZDG2015025)

作者简介:陈思憬(1994-),男,研究方向为基于网络的计算机软件技术;骆冰清,博士,讲师,研究方向为无线传感器网络、信息网络;孙知信,博士,教授,研究方向为计算机网络与安全。

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1450.TP.20171019.1624.034.html>

度的影响力。Yao 等^[7]将用户好友信任度分成了主动信任与被动信任,并对两种情况进行了加权计算。Wang 等^[8]提出应该依据用户评分和好友信任度来估计用户间的好友强度,结合用户好友强度以及偏好相似性来进行好友推荐。文献[2,7]的算法的关注重点是如何更加精确地计算好友信任度,其他算法的重点则在于如何合理地平衡好友信任度与用户相似度之间的关系。然而,以上算法均忽略了三度好友路径对于用户信任度计算的影响。

1967年,Stanley Milgram提出了六度分割理论^[9],认为一个人跟任何一个陌生人的距离都不会超过6个人。在此基础上,Nicholas A Christakis提出了3度影响力规则(three degrees of influence rule),认为每一个人的行为与情感会对三度以内的好友产生影响^[10]。2015年,王名扬等^[11]将三度影响力理论引入好友推荐算法,提出了基于三度影响力理论的好友推荐算法,以用户之间三度以内的好友路径数量为依据进行好友推荐,拓展了好友推荐算法的关系连接范围。然而该算法没有考虑到用户对其好友路径的信任度的影响。

基于三度影响力理论(three-degree influenced),提出一种采用三度以内好友路径计算好友信任度的社交网络好友推荐法,并通过实验证明该算法的准确性。

1 混合好友路径信任度算法

1.1 好友信任的度量

信任是通过搜集、分析用户之间的交互行为获得的用于衡量二人好友关系密切程度的度量。用户之间的信任包括:

(1) 认识信任: 用户不会添加自己不了解的陌生人为好友,因此用户之间建立好友关系时,可以认为两个用户之间已经产生了一定的信任关系;

(2) 交互信任: 用户之间的交互行为越多,可以认为两个用户之间的信任关系越强。

文中采用交互信任进行信任度计算,当用户 a_i 参与评论、转发好友用户 a_j 的发布内容时,其对于好友用户 a_j 的交互信任度会随之增加。在此定义交互信任度算法如下:

$$T(a_i, a_j) = \frac{\text{interactive}_{a_i a_j}}{\sum_{j=1}^n \text{interactive}_{a_i a_j}} \quad (1)$$

其中, $T(a_i, a_j)$ 表示用户 a_i 与用户 a_j 之间的交互信任度; $\text{interactive}_{a_i a_j}$ 表示用户 a_i 参与评论、转发用户 a_j 发布内容的次数; $\sum_{j=1}^n \text{interactive}_{a_i a_j}$ 表示用户对所有好友的参与评论、转发信息的次数。

1.2 二度好友路径信任度

在现实生活中,人们通常会在他人的介绍下结交新好友,这说明好友之间的信任是可以传递的,只是好友信任会在传递的过程中逐渐削弱。只要获取好友用户之间的信任度,就能够依照某种信任传递规则获知这种信任对于他人的影响^[12]。假设用户 a_i 对于其一度好友用户 a_j 存在信任关系 $T(a_i, a_j)$,而用户 a_j 对于用户 a_i 经过 a_j 的二度好友 a_k 也存在信任关系 $T(a_j, a_k)$,那么用户 a_j 对于用户 a_k 的信任度可以在一定程度上作为用户 a_i 对于用户 a_k 的参考。基于以上认识,定义信任传递规则:

$$\Gamma_2(a_i, a_j, a_k) = T(a_i, a_j) * T(a_j, a_k) \quad (2)$$

其中, a_i 为该路径的起始端点,即该算法的源用户; a_k 为目的用户; a_j 为该路径经过的好友用户; $\Gamma_2(a_i, a_j, a_k)$ 为从用户 a_i 出发,经过用户 a_j 到达用户 a_k 的二度好友路径信任度。

在实际应用中,一般情况下两个用户之间总是存在多条用户好友路径。要实现信任度计算,需要聚合多条路径的好友信任度。文中定义的二度好友路径信任度聚合方法为:

$$\rho_2(a_i, a_k) = \frac{\sum_{a_j \in T_{a_i}} \Gamma_2(a_i, a_j, a_k)}{N} \quad (3)$$

其中, T_{a_i} 为用户 a_i 的好友集合; $\rho_2(a_i, a_k)$ 为从用户 a_i 到达用户 a_k 的所有二度好友路径的信任值聚合; N 为从源用户 a_i 到达目标用户 a_k 的二度好友路径数量。

1.3 三度好友路径信任度

计算三度好友路径信任度,首先要找出从源用户到目标用户的三度以内好友路径。社交网络常用的表现形式为拓扑图,通常用节点表示用户,用两个节点间的边表示两个用户之间存在的好友关系。以用户 a_1 为核心的虚拟社交网络拓扑如图1所示。

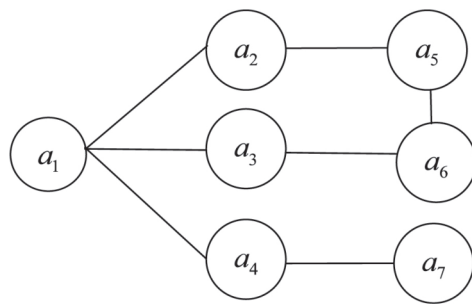


图1 虚拟社交网络拓扑

根据图1可以得到各个用户的好友列表集合,即 $T_{a_1} = \{a_2, a_3, a_4\}$, $T_{a_2} = \{a_1, a_5\}$, $T_{a_3} = \{a_1, a_6\}$, $T_{a_4} = \{a_1, a_7\}$, $T_{a_5} = \{a_2, a_6\}$, $T_{a_6} = \{a_3, a_5\}$, $T_{a_7} = \{a_4\}$ 。每个用户的好友列表里任意两个元素组成元素对,均可视为经过该用户的二度好友路径。得到二度好友路

径后,只要将其端点用户好友列表中一度好友路径与该二度好友路径进行匹配,即可得到三度好友路径。在算法实际进行过程中,需要考虑到产生回路的可能性,将用户的二度好友同时也是用户一度好友的情况筛选除去。

寻找源用户通往目的用户三度好友路径的具体步骤如下:

(1) 通过好友列表寻找源用户 a_i 的一度好友;

(2) 源用户 a_i 的一度好友的好友列表中除了源用户 a_i 之外的好友用户 a_e 与源用户 a_i 组成二度好友对;

(3) 筛选二度好友对,除去路径目的用户 a_e 是源用户 a_i 的一度好友的好友对;

(4) 将好友推荐目的用户 a_k 的好友列表与源用户 a_i 的二度好友 a_e 进行匹配,得到源用户通往目的用户的三度好友路径。

文中定义的单条路径三度好友路径信任度如下:

$$\Gamma_3(a_i, \mu_j, \mu_e, \mu_k) = T(a_i, \mu_j) * T(a_j, \mu_e) * T(a_e, \mu_k) \quad (4)$$

其中, a_i 为源用户; a_k 为目的用户; a_j, a_e 为该路径经过的好友用户; $\Gamma_3(a_i, \mu_j, \mu_e, \mu_k)$ 为从用户 a_i 出发,经过用户 a_j, a_e 到达用户 a_k 的三度好友路径信任度。

多条路径好友信任度算法如下:

$$\rho_3(a_i, \mu_k) = \frac{\sum_{a \in T_a, b \in T_a} \Gamma_2(a_i, a, b, \mu_k)}{M} \quad (5)$$

其中, T_{a_i} 表示用户 a_i 的好友集合; $\rho_3(a_i, \mu_k)$ 表示从用户 a_i 到达用户 a_k 的所有三度好友路径的信任值聚合; M 表示从源用户 a_i 到达目标用户 a_k 的三度好友路径数量。

1.4 混合好友路径信任度

在社交网络中,两个用户之间通常不止存在一条好友路径,因此需要将其之间的二、三度好友路径信任度聚合起来以得到其混合好友路径信任度,定义如下:

$$\omega(a_i, \mu_k) = \sum_{n=2}^3 \rho_n(a_i, \mu_k) \quad (6)$$

文中提出的混合好友路径信任度好友推荐算法的处理方法如图 2 所示。

该算法通过分析用户好友信息以及用户交互信息得到用户之间的好友路径及其信任度,对用户之间信任度进行排序从而得到好友推荐结果。

2 实验

2.1 实验数据及依据

实验将文中算法与依据三度以内好友路径数量(三度无信任)的好友推荐算法、基于共同二度以内好

友路径信任度(二度信任)的好友推荐算法进行比较分析。

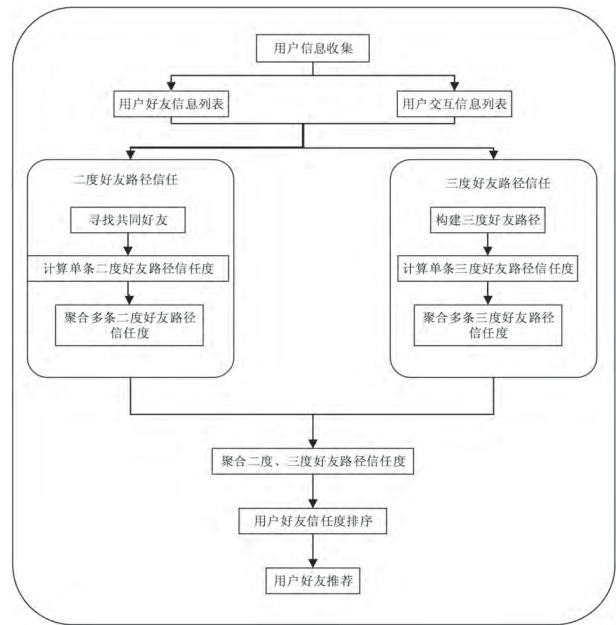


图 2 混合好友路径信任度好友推荐算法

目前新浪微博有很多开放 API,为原始实验数据的采集提供了便利^[13]。实验采用新浪微博用户数据作为原始实验数据,收集、挑选了 63 641 个用户信息,1 028 331 个用户互粉数据以及 1 055 043 条用户交互行为数据。实验将用户数据分为包含 80% 用户的测试集以及 20% 的训练集。先隐藏用户部分好友信息,将依据其他好友信息得到的推荐好友集 Top- K 与隐藏的好友信息进行比对以验证算法的有效性。其中 K 表示推荐好友数量。采用准确率(precision)以及召回率(recall)两个常用指标作为验证好友推荐算法有效性的依据^[14]。F1-measure 是综合了 precision 和 recall 的信息检索评价指标。分别定义如下:

$$\text{precision} = \frac{H(a_i)}{K(a_i)} \quad (7)$$

$$\text{recall} = \frac{H(a_i)}{G(a_i)} \quad (8)$$

$$\text{F1-measure} = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}} \quad (9)$$

其中, $H(a_i)$ 表示对用户 a_i 的推荐好友是用户 a_i 的隐藏好友的数量; $K(a_i)$ 表示向用户 a_i 推荐的好友总数量; $G(a_i)$ 表示用户 a_i 的隐藏好友用户总数量。

2.2 实验结果与分析

实验将推荐好友 Top- K 的数量 K 分别设定为 5、10、20,应用 precision、recall 和 F1-measure 分别将基于二度以内好友路径信任度的好友推荐算法以及基于三度以内好友路径数量的好友推荐算法^[11]与文中基于三度以内好友路径信任度的好友推荐算法进行性能比较,结果如图 3~5 所示。

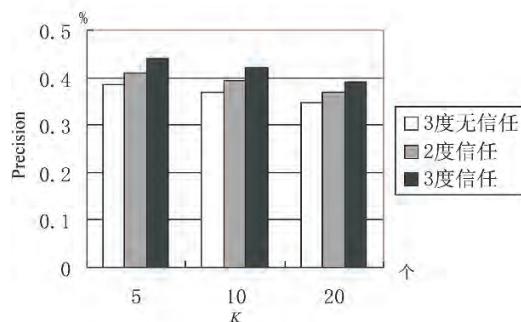


图3 三种算法在不同K值下的 precision 值

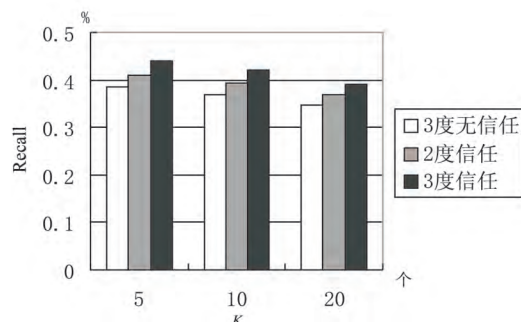


图4 三种算法在不同K值下的 recall 值

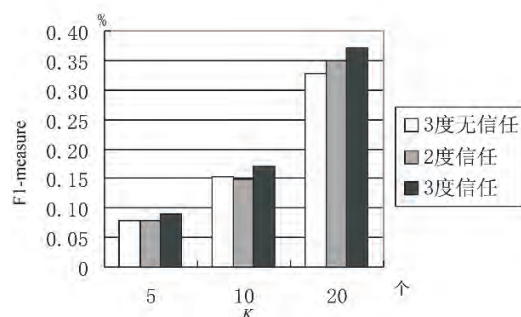


图5 三种算法在不同K值下的 F1-measure 值

由图3~5可以看出,三度好友信任推荐要比三度好友无信任推荐具有更高的 precision 与 recall,其 F1-measure 性能指标比三度好友无信任推荐得到的要高。这证明了当好友推荐考虑三度以内的好友路径影响时,考虑好友路径的信任度影响的推荐效果要比仅仅依靠好友路径数量进行好友推荐的效果好。而在考虑好友路径信任度对好友推荐效果的影响时,引入三度好友路径的影响,其三项好友推荐性能指标比仅仅考虑二度好友路径信任度的影响均有所提升。这证明了进行好友推荐时引入三度好友路径的影响能有效提高好友推荐效果,同时也证明了三度影响力的正确性。

3 结束语

文中在三度影响力理论下对以往的好友推荐算法

进行了改进,拓宽了好友推荐的参考因素范围,引入三度好友路径对好友推荐的影响。经实验证明,引进三度好友的影响力之后,好友推荐算法的准确性要比仅仅考虑二度好友影响的算法高,而在三度好友影响力的基础上加上好友路径的信任度影响因素,其算法的准确性更高。该算法对于好友推荐系统的性能有一定提高。三度影响力理论也为以后的好友推荐算法提供了新的发展思路,从而能更有效地向用户推荐好友。

参考文献:

- [1] 刘建国,周涛,汪秉宏.个性化推荐系统的研究进展[J].自然科学进展,2009,19(1):1-15.
- [2] 唐晓波,孙飞.基于复杂信任网络的社会化媒体好友推荐研究[J].情报理论与实践,2015,38(11):96-102.
- [3] 张怡文,岳丽华,张义飞,等.基于共同用户和相似标签的好友推荐方法[J].计算机应用,2013,33(8):2273-2275.
- [4] 王涛,覃锡忠,贾振红,等.基于相似度和信任度的关联规则微博好友推荐[J].计算机应用,2016,36(8):2262-2267.
- [5] 朱强,孙玉强.一种基于信任度的协同过滤推荐方法[J].清华大学学报:自然科学版,2014,54(3):360-365.
- [6] PAN W, XIA S, LIU Z, et al. Mixed factorization for collaborative recommendation with heterogeneous explicit feedbacks[J]. Information Sciences, 2016, 332: 84-93.
- [7] YAO W, HE J, HUANG G, et al. Modeling dual role preferences for trust-aware recommendation[C]//Proceedings of the 37th international ACM SIGIR conference on research & development in information retrieval. Gold Coast, Australia: ACM, 2014: 975-978.
- [8] WANG M, MA J. A novel recommendation approach based on users' weighted trust relations and the rating similarities[J]. Soft Computing, 2015, 20(10): 3981-3990.
- [9] KE X. A social networking services system based on the "six degrees of separation" theory and damping factor[C]//Proceedings of the 2010 2nd international conference on future networks. Washington, DC: IEEE Computer Society, 2010: 438-441.
- [10] WALKER S K. Connected: the surprising power of our social networks and how they shape our lives[J]. Journal of Family Theory & Review, 2011, 3(3): 220-224.
- [11] 王名扬,贾冲冲,杨东辉.基于三度影响力的社交好友推荐机制[J].计算机应用,2015,35(7):1984-1987.
- [12] 蒋黎明,张琨,徐建,等.证据信任模型中的信任传递与聚合研究[J].通信学报,2011,32(8):91-100.
- [13] 黄延伟,刘嘉勇.新浪微博数据获取技术研究[J].信息安全与通信保密,2013(6):71-73.
- [14] 朱郁筱,吕琳媛.推荐系统评价指标综述[J].电子科技大学学报,2012,41(2):163-175.