

文章编号: 2095—0411 (2013) 03 - 0066 - 05

# 一种改进的新浪微博好友推荐算法<sup>\*</sup>

杨 晶, 杨长春, 丁 虹

(常州大学 信息科学与工程学院, 江苏 常州 213164)

**摘要:** 目前新浪微博的好友推荐机制存在一些缺点, 通过研究微博社区结构和权威用户对好友推荐的影响, 提出了一种改进的基于新浪微博的好友推荐算法。在同一微博社区中, 通过查找权威用户, 并结合用户之间的兴趣相似度和信任度进行好友推荐, 推荐过程中两次计算了用户的兴趣相似度并引入用户间信任度传播模型。选取微博社区中目标用户进行实验的结果表明, 权威用户在好友推荐中起了重要作用, 提高了好友推荐的效果。同时通过将该算法推荐的好友列表和原新浪微博推荐的好友列表作对比, 实验表明该算法具有较好的推荐效果。

**关键词:** 好友推荐; 微博社区结构; 权威用户; 兴趣相似度; 信任度

**中图分类号:** TP 301.6

**文献标识码:** A

**doi:** 10.3969/j.issn.2095—0411.2013.03.017

## A Modified Friend Recommending Algorithm Based on Sina Microblogging

YANG Jing, YANG Chang-chun, DING Hong

(School of Information Science and Engineering, Changzhou University, Changzhou 213164, China)

**Abstract:** There are some deficiencies in the current Sina microblogging friend recommending mechanism. By studying the influence of microblogging community structure and authority users on the friend recommendation, this paper proposed a modified friend recommending algorithm based on Sina Microblogging. In the same microblogging community, conducting friend recommendation by searching the authority should be combined with interest similarity and trust degree between users. In the friend recommending process, the algorithm calculated the interest similarity between users twice and introduces user trust degree propagation model. The results of selecting target users in real microblogging community to carry on the experiments show that the authority users play an important role in friend recommendation and can improve the effect of friend recommendation. At the same time, the results of contrasting the Sina microblogging friend recommending list and the proposed algorithm friend recommending list show that the proposed algorithm has a better recommending effect.

**Key words:** friend recommended; microblogging community structure; authority users; interest similarity; trust degree

微博已经成为时下最流行的大众化信息传播媒介, 通过微博平台用户可以轻易地完成各类信息

<sup>\*</sup> 收稿日期: 2013 - 03 - 07

基金项目: 国家自然科学基金项目资助 (61272367)

作者简介: 杨晶 (1989—), 女, 江苏镇江人, 硕士生; 通讯联系人: 杨长春。

的获取、生产、分享和传播。微博的另一大魅力是可以在这个虚拟网络中认识更多的朋友,以新浪微博为例,用户可以通过“你可能感兴趣的人”这个模块中由系统推荐的用户,找到志趣相投的人。

目前,新浪微博用于推荐好友的算法大致包括以下 5 种情况:①根据教育信息进行基本信息相似推荐。用户信息中有填写教育背景的,系统会根据入学时间、学校、所学专业等信息进行同类匹配,选择匹配度较高的进行推荐。②根据标签、关键词进行兴趣相似推荐。大多数用户都会填写具有自我个性化特色的标签,此类标签可以代表用户的兴趣爱好方向,系统会根据这些信息进行兴趣相似度查找,选择兴趣相似度高的进行推荐。③根据共同关注的人进行推荐。根据 2 个人之间所关注的共同好友的数量可以判断 2 人的共同兴趣。如果 2 人之间共同关注的人很多,那么此 2 人拥有相同兴趣的可能性就很大,那么可以将 2 人互相推荐为好友。④根据地理位置进行就近推荐。此处的地理位置包含两点:一是用户信息中的所在地;二是用户当前登录的 IP 地址。系统会根据这两者信息选择地理位置附近的用户进行推荐。⑤微博会员优先推荐。新浪微博系统推出了微博会员,用户一旦成为会员,将会得到系统优先将其推荐为好友的特权。

上述好友推荐算法是从全局的微博网络着手进行搜索得到好友推荐列表,具有一定的参考价值,但都没有考虑微博社区结构对好友推荐的影响,忽略了在同一社区内的用户更有可能成为好友的现象,也没有考虑到权威用户在推荐中的特殊影响力,因此推荐结果具有一定的盲目性和广泛性。本文研究了微博社区结构和权威用户对好友推荐的影响,在此基础上结合用户的兴趣相似度和用户之间的信任度,提出了一种改进的新浪微博好友推荐算法。

## 1 相关工作

目前,传统的个性化推荐系统所采用的推荐方法主要分为两种:协同过滤推荐算法和基于内容的推荐算法<sup>[1]</sup>。前者是根据相邻用户的资料对目标用户进行产品推荐,具体通过分析历史数据计算用户之间的相似度,生成与目标用户行为兴趣最相近的最近邻集合,根据最近邻对产品的评分预测目标用户对产品的评分,将评分最高的前  $N$  项产品推荐给目标用户。后者是推荐与目标用户之前选择的产品相似的产品,系统通过学习用户已经评价过的产品数据来获得用户的兴趣描述,再将待推荐产

品数据资料与用户兴趣描述匹配,决定是否推荐。随着个性化推荐系统的广泛应用,基于微博系统推荐的研究也越来越多,如标签推荐<sup>[2]</sup>、微博内容推荐<sup>[3]</sup>等。在这些研究中,用户的兴趣度和用户间的信任度等包含用户个性化特色的信息得到了广泛的关注。

社区结构是网络的一大特性,微博网络中也存在许多大大小小的社区,社区内的用户在某种意义上具有兴趣爱好相似的可能性更大<sup>[4]</sup>。另外不同的用户有不同的推荐效果,其中权威用户有更大的影响力。本文在同一微博社区结构中,基于用户的兴趣度和用户间的信任度研究及权威用户影响力的研究,提出了一种改进的新浪微博好友推荐算法,主要做了以下 5 个方面的研究工作:①微博社区模型的研究;②社区中的权威用户的搜索和权威用户推荐;③用户之间兴趣相似度的计算;④用户的信任度传播模型;⑤结合上述研究提出一种改进的新浪微博好友推荐算法。

## 2 好友推荐算法

### 2.1 微博社区模型

微博依靠其关注、转发、评论等功能建立了用户之间强大的互动关系。通过研究用户之间的交互频率可以发现微博网络中存在了许多复杂的社区结构,在同一社区里的用户相比于社区间的用户的交互更加频繁,联系更加紧密<sup>[4]</sup>。图 1 所示为简易的微博网络社区结构示意图,从图 1 中可以很容易的发现社区 A 和 B 内部节点联系非常密切,而社区之间联系比较稀疏。微博上的用户通过发布新博文、转发、评论他人的博文进行交互的,交互频繁的用户成为好友的可能性更大。从社区内的关系密切,社区间的关系稀疏,可以推断社区内的用户之间更有可能成为好友,这为向目标用户推荐好友提供了基础,缩小了搜索范围。因此,本文的好友推荐模型是在微博社区中进行的。

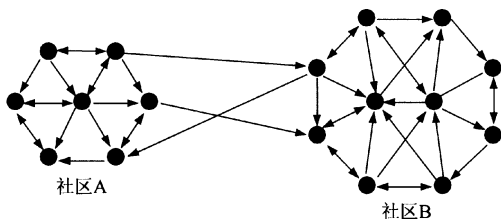


图 1 简易的社区结构演示

Fig. 1 Demo of simple community structure

2.2 社区中权威用户的搜索和推荐

权威用户（意见领袖），拥有大量的粉丝但其关注的人相对较少。他们经常发布博文，活跃度比较高。当其发布或转发一条博文时，该博文会被迅速扩散，影响力很大，这就是所谓的明星效应。例如新浪微博用户“吴奇隆”，他的粉丝数是 8 982 432，他关注的人数是 518，他发布的博文大多在十几分钟内就会发生上千次转发和评论。由于权威用户的巨大影响力，普通用户对他们感兴趣的可能性就比较高，很有可能去“关注”他们，而且普通用户对他们的信任度也会比其他普通用户高，考虑到权威用户在推荐过程中起的作用，本文将权威用户的搜索和推荐应用到好友推荐算法中。

肖宇<sup>[5]</sup>通过分析用户的内容信息或链接信息，对用户的权威度进行排序。文献 [6] 利用粉丝的数量和微博转发的数量对用户的影响力进行了衡量。文献 [7] 提出根据用户影响力和用户活跃度来挖掘权威用户。

综合对于权威用户影响力的研究发现，影响力排名靠前的用户几乎是各行业的名人。新浪微博中有一项官方认证即加“V”用户，这些用户是用自己的真实信息通过新浪微博系统验证的名人。根据这项功能，本文简化了权威用户的搜索过程，即：在一个社区的范围内，首先查找出所有的加“V”用户，然后利用兴趣相似度的计算方法（下文将介绍）提取与目标用户兴趣最相似的前 N 位权威用户成为我们算法中所需要的权威用户。而权威用户所推荐的是其所关注的用户。

2.3 用户兴趣相似度的计算

用户的兴趣相似度在好友推荐模型中占有很重要的地位。新浪微博中用户有“个人标签”，这是最具用户个性化特色的信息，最能表达用户的兴趣爱好。本文提取用户个人标签中的信息，利用语义分析工具进行词汇相似度分析，进而得到用户的兴趣相似度值。

刘群、李素建发表了基于《知网》的词汇语义相似度计算<sup>[8]</sup>，文中提出词语的距离和词语的相似度存在着密切的关系，二者是一对词语相同关系特征的不同表达形式。给定两个词语  $w1$  和  $w2$ ， $S_{(w1,w2)}$  表示二者的相似度， $D_{(w1,w2)}$  表示二者的词语距离，则词语间的相似度定义如下：

$$S_{(w1,w2)} = \frac{\alpha}{D_{(w1,w2)} + \alpha} \tag{1}$$

其中  $\alpha$  是一个可调节的参数。

本文使用了 WordSimilarity 软件来计算“个人标签”之间的相似度，它是根据上述刘群发表的文章中的原理编写的。通过其输入词语就可以直接得到两者之间的相似度（数值为 0 到 1 之间的实数），对于字数较多的字段可利用文本输入得到相似度。

2.4 用户的信任度传播模型

信任传播又称为“回馈”、“信息反馈”，提出此概念的最初目的是调查受众对大众传播媒介的信任度和对传播内容的反应，并以新闻或言论的形式发表受众对重大政治、经济、社会等各种问题的看法。信任传播的思想结合排序算法已经有效的运用到多种个性化推荐算法中。本文利用信任度传播的原理，将信任机制引入到微博好友推荐的模型中。

信任度传播原理可以简单的解释为，用户 a 与用户 b 是好友，则视为用户 b 取得了用户 a 的信任。而对于用户 b 的好友用户 c，c 取得了 b 的信任，但用户 a 对 c 的信任却会以一定的衰减因子  $\alpha$  减弱（ $\alpha$  为 0 到 1 之间的实数）。如果 a 的好友成员中同时与 c 是好友的人数越多，表示 c 取得了 a 的好友成员中越高的信任，那么 a 加 c 为好友的概率越大。同理，用户 a 对用户 c 好友的信任也会以一定的衰减因子  $\alpha$  传播。如果用户 c 有许多好友，很难保证他对每一个好友都有足够了解，根据信任平均分配原则，用户 c 的好友数量越多，用户 a 对其单个好友的信任会越低。

在微博系统中，普通用户一般对自己感兴趣的权威用户的信任度会比较高，通常会出现“爱屋及乌”的现象，他们会搜索这些权威用户的关注列表，从中再查找自己感兴趣的用户，进而关注他们。图 2 为信任度传播原理图。

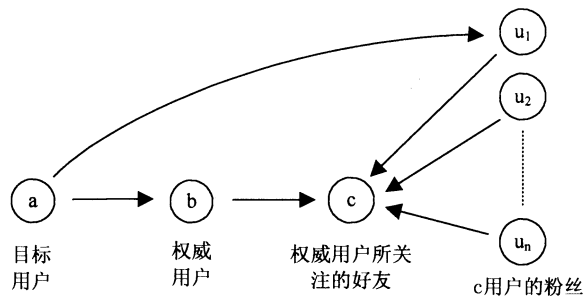


图 2 信任度传播模型  
Fig. 2 Trust propagation model

基于上述分析, 本文采用文献 [9] 中的用户可信度计算方法, 给定一个微博社区网络有  $n$  个用户, 目标用户  $a$  和另一用户  $c$  以及社区内关注  $c$  的用户列表  $u_1, u_2, \dots, u_k$ , 计算用户  $a$  对  $c$  的信任度公式为:

$$T_{a,c} = \frac{1-\alpha}{n} + \alpha \left( \frac{T_a(u_1)}{F(u_1)} + \frac{T_a(u_2)}{F(u_2)} + \dots + \frac{T_a(u_k)}{F(u_k)} \right) \quad (2)$$

其中,  $(1-\alpha)/n$  表示用户  $a$  对  $c$  的原有初始信任度;  $F(u_i)$  ( $i=1, 2, \dots, k$ ) 表示社区内  $u_i$  所关注的用户数量;  $T_a(u_i)/F(u_i)$  表示用户  $a$  对  $u_i$  的信任度会平均分给  $u_i$  所关注的用户。目标用户  $a$  对社区网络中所有用户信任度的总和为 1, 即:

$$\sum_{i=1}^n T_a(u_i) = 1 \quad (3)$$

## 2.5 好友推荐算法描述

根据本文的上述分析, 提出了一种改进的新浪微博好友推荐模型, 算法的流程如图 3 所示。

## 2.6 实验及分析

本文的实验数据来源于 <http://www.data-tang.com/data/11819> 提供的新浪微博用户数据。

实验设计: 利用文献 [4] 中实验结果, 在已发现的一个微博社区中选定目标用户, 为了验证权威用户对好友推荐的影响, 以及本文好友推荐算法对比原新浪微博好友推荐的效果, 进行了以下两组实验。

### 2.6.1 验证权威用户对好友推荐的影响

实验 1: 选取社区中用户“叮嚀丁丁”为目标用户。在整个社区用户集中随机选取 5 个非“V”认证用户→搜索他们关注的好友→计算目标用户对上述用户的信任度值→提取信任度值最高的前 25 位用户→再次与目标用户进行相似度计算→确定相似度值最高的前 12 位用户为最终推荐的好友。

实验 2: 用本文提出的好友推荐算法进行好友推荐。

以上两个实验分别记录目标用户和所推荐好友的兴趣相似度值, 用户兴趣相似度值越大, 则推荐效果越好, 图 4 是两个实验结果的对比折线图。

从图 4 非权威用户推荐和权威用户推荐的好友与目标用户的相似度对比图可以发现, 权威用户推荐的好友与目标用户的相似度值均高于非权威用户

推荐的好友与目标用户的相似度值, 这表明, 权威用户在好友推荐中有很大的影响, 本文将权威用户推荐考虑到好友推荐的算法中, 提高了好友推荐的效果。

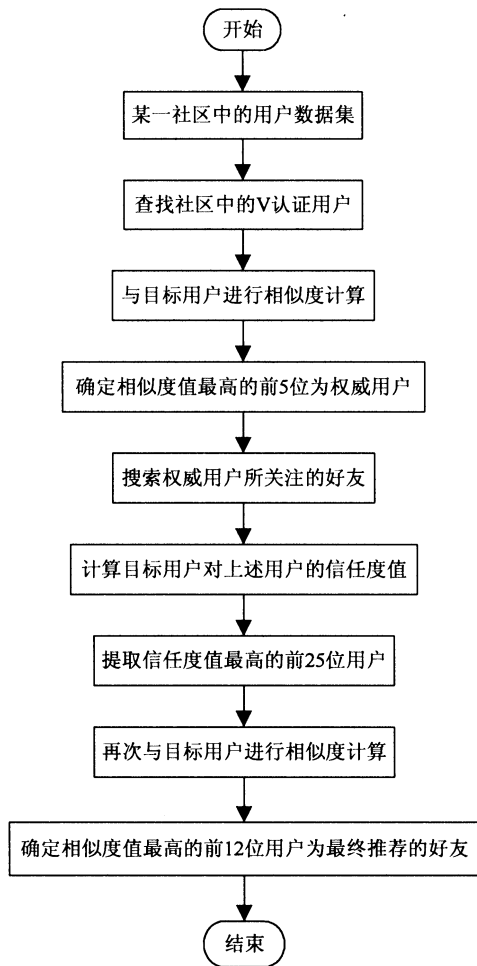


图 3 好友推荐算法流程图

Fig. 3 Flow chart of friend recommending algorithm

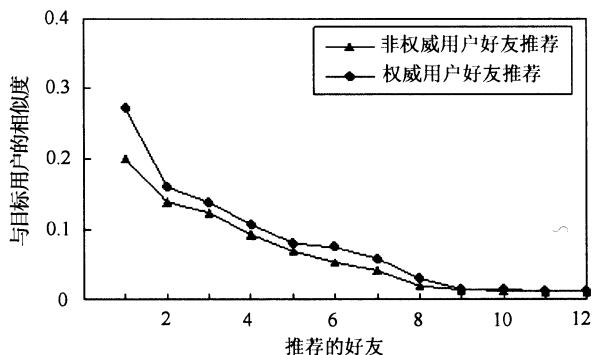


图 4 非权威用户推荐和权威用户推荐的好友与目标用户的相似度对比图

Fig. 4 Comparison graph of the similarity of friend recommendation by non-authoritative users and target user and that of friend recommendation by authoritative users and target user

### 2.6.2 验证本文好友推荐算法的效果

用本文好友推荐算法为目标用户推荐好友, 记



录目标用户与推荐好友的兴趣相似度值。然后计算原新浪微博推荐的好友与目标用户的兴趣相似度值并记录, 对比两种推荐, 用户兴趣相似度值越大, 则推荐效果越好。为此, 分别选取了社区中的两个目标用户“叮噹丁丁”和“北门清燕”。

实验 1: 目标用户“叮噹丁丁”的用户标签集合为  $L = \{\text{美食, 时尚, 明星控, 旅游}\}$ , 本文根据该标签集合对目标用户进行好友相似度匹配, 得到前 12 位用户成为被推荐的好友。新浪微博为该目标用户推荐了 28 位“可能感兴趣的人”, 为了对比方便, 将这 28 位用户与目标用户也进行了好友相似度匹配并排序取前 12 位用户。图 5 为两种算法推荐的好友与目标用户的相似度对比图, 从图 5 中的折线趋势可以看出本文算法推荐的好友与目标用户的相似度值平均高于新浪微博为其推荐的好友与目标用户的相似度值。

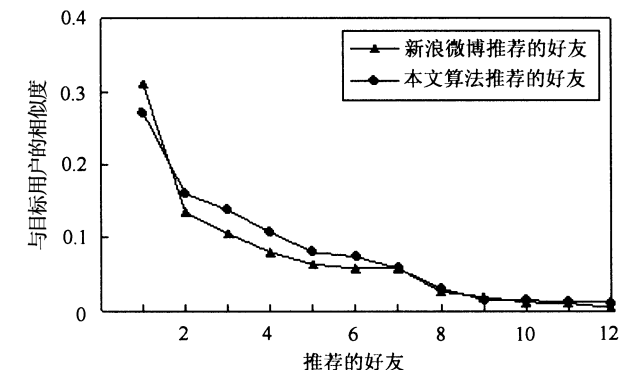


图 5 实验 1 两种算法推荐的好友与目标用户的相似度对比图  
Fig. 5 Experiment 1: Comparison graph of the similarity of friend recommendation by two algorithms and target user

实验 2: 目标用户“北门清燕”的用户标签集合为  $L = \{\text{宅死家中, 游戏, 上网, 听歌, 80 后, 文字, 心情}\}$ 。由于该用户的标签集合中的文本较多, 利用本文算法得到的兴趣相似度值略高, 即表明他们成为好友的可能性越大。图 6 是两种算法推荐的好友与目标用户的相似度对比图, 从图 6 中折线的分布可以发现, 本文算法推荐的好友与目标用户的相似度值均高于新浪微博推荐的好友与目标用户的相似度值, 说明目标用户的标签集合中的文本数量越多, 本算法得到的好友的优势越明显。

### 3 结 论

好友推荐作为新浪微博中最常用的一项应用, 其推荐的准确度直接关系到用户的体验舒适感。由于新浪微博目前的推荐机制的不足, 本文综合考虑了微博社区结构和权威用户对好友推荐的影响, 结

合用户兴趣度和用户间的信任度, 提出了一种改进的新浪微博好友推荐方法。在真实微博社区中选取目标用户进行实例分析的实验结果表明, 权威用户在好友推荐中起了重要作用, 提高了好友推荐的效果。同时通过对比本文算法推荐的好友和原新浪微博推荐的好友与目标用户的兴趣相似度, 体现了本文算法的推荐结果在一定程度上优于新浪微博的推荐结果, 尤其当目标用户的标签集中文本数量越多的情况下, 本文算法的优势越明显。

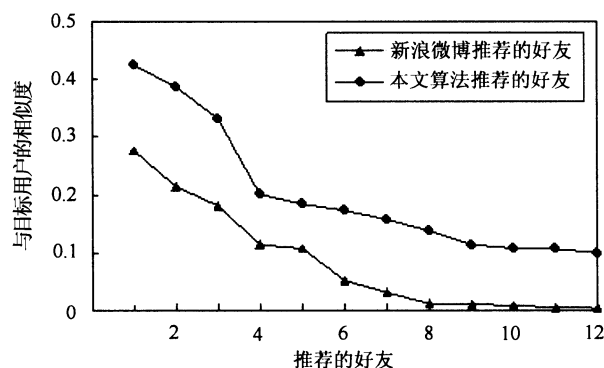


图 6 实验 2 两种算法推荐的好友与目标用户的相似度对比图  
Fig. 6 Experiment 2: Comparison graph of the similarity of friend recommendation by two algorithms and target user

### 参考文献:

- [1] 高琪, 辛乐. 基于用户偏好模型和情感计算的产品推荐算法 // 第 29 届中国控制会议论文集 [C]. 北京: 航空航天大学出版社, 2010: 2891 - 2896.
- [2] 陈渊, 林磊, 孙承杰, 等. 一种面向微博用户的标签推荐方法 [J]. 智能计算机与应用, 2011, 1 (3): 21 - 26.
- [3] 郑志娟. 微博个性化内容推荐算法研究 [J]. 电脑开发与应用, 2012, 25 (12): 23 - 25.
- [4] Yang Chang - chun, Ding Hong, Yang Jing, et al. Research of microblog community detection based on clustering analysis [J]. AISS, 2013, 5 (3): 25 - 31.
- [5] 肖宇, 许炜, 夏霖. 网络社区中的意见领袖特征分析 [J]. 计算机工程与科学, 2011, 33 (1): 150 - 156.
- [6] Meeyoung C, Haddadi H. Measuring user influence in Twitter: The million follower fallacy // Proceedings of International Conference on Weblogs and Social Media [C]. Washington, DC: AAAI Press, 2010: 10 - 17.
- [7] 刘志明, 刘鲁. 微博网络舆情中的意见领袖识别及分析 [J]. 系统工程, 2011, 6 (29): 8 - 16.
- [8] 刘群, 李素建. 基于《知网》的词汇语义相似度计算 [J]. 计算语言学及中文信息处理, 2007, 31 (7): 59 - 76.
- [9] 张中峰, 李秋丹. 社交网站中潜在好友推荐模型研究 [J]. 情报学报, 2011, 30 (12): 1319 - 1325.