

# 基于活动网络的微博用户影响力分析

张凤娟,王 濛,周 刚  
(信息工程大学,河南 郑州 450001)

**摘要:**传统对微博用户影响力的预测仅通过微博网络的静态拓扑作结构为分析基础,忽略了用户之间的真实交互情况或者仅利用了交互行为在单个目标上的聚合结果,没有考虑用户间交互行为的强弱和相互作用等问题。对此,设计并提出一种基于活动网络的微博用户影响力分析方法—ANR。ANR利用用户之间的交互行为构建微博用户活动网络,以活动网络为基础,考虑用户间的交互强度以及用户在网络中的活跃度,对用户属性特征和交互行为特征进行了量化,进而利用在该模型上的改进 PageRank 算法对用户影响力进行预测。实验结果表明,ANR 算法的预测结果比基于粉丝数排序的方法和基于好友关系网络的 PageRank 方法的预测结果更加准确和合理。

**关键词:**在线社交网络;微博;用户影响力;活动网络

中图分类号:TP391

文献标识码:A

文章编号:1673-629X(2018)09-0162-06

doi:10.3969/j.issn.1673-629X.2018.09.033

## Analysis of User Influence in Microblog Based on Activity Network

ZHANG Feng-juan, WANG Meng, ZHOU Gang  
(Information Engineering University, Zhengzhou 450001, China)

**Abstract:** The traditional ways of predicting the impacts of microblog users are through microblogging network static topology as the basis of analyzing, ignoring the real interaction between users or only using aggregation results of a single target without considering the interactions and their strength between users. For this, we propose a new method for analysis of user influence based on activity network, called ANR. It firstly uses the interaction behaviors between users to construct activity network model, based on which we quantize feature of individual attribute and interaction behavior considering the strength of interaction between the users and user activity, thus applying the improved PageRank algorithm to predict the user influence. The experiment shows that the ANR is more accurate and rational than the sorting method based on the number of fans and PageRank based on friend relationship network.

**Key words:** online social network; microblog; user influence; activity network

## 0 引言

微博作为一种在线社会网络以其实时性、高效性吸引了大量用户。与传统在线社会网络相比,微博用户间只需通过“关注”行为便可获得目标用户的微博信息,并以此构成具有广播性质的社会信息传播网络;用户间交互方便,可以实时地获取所关注用户的消息并参与讨论;而且微博具有多种信息发布平台和发布形式,已经逐渐成为人们传播消息和获取消息的重要媒介。

在微博中,有一些用户对信息传播和话题发展起到了极大的促进和推动作用,或者能够对其他用户产生很大的影响,这些用户被称为有影响力的用户,这些用户能够引导微博舆论的发展方向。分析并预测微博

网络中的用户影响力,有利于分析舆情在网络中的传播轨迹,追踪热点话题的关键人物,为舆情分析提供支持,准确判断影响力高的用户,以便进行网络舆情引导或者社会网络营销等。

影响力一般认为是一种以别人乐于接受的方式,改变他人情感、思想或行为的能力。在微博平台上,用户影响力本质上是某用户发布的帖子等信息引起其他用户行为改变的能力,一个用户对其他用户的作用越大,该用户的影响力也越大。用户通过发布微博消息、评论他人的消息等行为来表达自己的情感与观点,通过关注、转发、评论等与其他用户交互。文中认为那些影响力很大的用户,他们的微博信息,很多都会在微博网络中影响其他用户,与之交互的用户越多,受其影响

收稿日期:2017-06-29

修回日期:2017-11-07

网络出版时间:2018-04-28

基金项目:河南省科技计划项目基金(162102210036)

作者简介:张凤娟(1983-),女,硕士,讲师,研究方向为社交网络分析、大数据分析。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20180427.1626.004.html>

的用户也越多,表明该用户的影响力也就越大。

因此,做出两个假设:微博中用户通过发布信息和行为影响他人;信息和行为的传播依赖于用户间的交互行为。基于此,文中首先根据用户间的交互行为构建活动网络,在此基础上利用用户的行为特征结合用户间交互的强度改进 PageRank 算法,提出了一种新的基于活动网络的微博用户影响力分析算法—ANR,并通过实验进行验证。

## 1 相关工作

近年来,随着互联网的迅速发展,微博呈现爆炸式发展,对于微博的研究受到了科研人员的广泛关注。在微博网络中用户是处于核心地位的,许多文献都对微博网络尤其是微博 Twitter 中的用户影响力进行探索,主要侧重于对用户属性和网络结构两个方面的研究。Kwak 等<sup>[1]</sup>研究了社交网站 Twitter 中的数据,使用了粉丝数、微博被转发数以及 PageRank 算法对用户影响力进行了分析,对排序结果的相关性进行比较,实验结果说明基于粉丝数的排序结果和采用 PageRank 算法的排序结果具有较强的相关性,但是,基于粉丝数的排序结果却与基于转发数的排序结果有很大差别。Cha 等<sup>[2]</sup>分别从粉丝数、转发数和评论次数三个属性对微博用户影响力进行考量,通过实验对比了三种用户影响力排序方法,得出的结论与文献[1]类似,即粉丝数越多不代表具有越大的影响力,说明了粉丝数并不是衡量用户影响力的唯一与最好指标。Web Ecology Project 中对用户影响力的研究<sup>[3]</sup>同样表明粉丝数多并不代表具有较大的影响力,像转帖的次数、回复的次数、浏览的次数等这些代表用户发表话题的受关注程度的属性也是评价用户影响力的重要指标。上述对微博用户影响力的衡量方法大多依赖于主观感觉,文献[4]通过对微博中多种指标的回归分析详细研究了不同属性在信息传播中的作用,并以此来度量用户影响力,实验表明,用户行为和交互的特征比微博网络的静态拓扑特征(比如粉丝数)更能反映用户的影响力大小。

PageRank 算法<sup>[5]</sup>是 Google 创始人 Lary Page 和 Sergey Brin 提出的,是 Google 搜索重要的网页排名算法。它主要是通过网页之间的链接结构来确定网页的重要性,不仅可以衡量网页的重要度,由于微博用户节点间的交互行为类似网页的链接关系,因此该算法还可以用来评价社会网络中用户影响力的大小。Daniel Tunkelang<sup>[6]</sup>改进了 PageRank 算法,将随机游走的想法和用户微博数量等特征用于 Twitter 中用户影响力的排序上,提出了用户影响力计算的 TunkRank 算法。文献[1]不仅借鉴了 PageRank 算法,并与用户粉丝

数、被转发数度量用户影响力的方法进行了比较。Weng 等<sup>[7]</sup>对 PageRank 算法进行了改进,提出了判断用户在给定某个话题内的影响力的 TwitterRank 算法。Lü 等<sup>[8]</sup>以传统 PageRank 算法为基础提出了 LeaderRank 算法,该算法通过附加一个超级节点来处理 PageRank 算法中孤悬节点和网络不连通等问题。和 PageRank 算法相比,LeaderRank 算法的优势有无额外参数、适用范围广、收敛性好和排序结果稳定等。但是这些方法,都只是研究了网络结构而没有考虑用户属性和用户的交互行为。

综上所述,以往对微博用户影响力度量的方法无论是基于粉丝数还是基于对 PageRank 算法的改进,都只是分析了微博网络的不同拓扑特征或节点间的相互作用且多是直接使用微博网络的静态拓扑作结构作为分析的基础。但是微博网络的静态拓扑只关注了网络中的好友关系,而忽略了用户间的真实交互情况。文献[9-10]的研究表明,社交网络中的用户交互网络远小于基于好友关系的静态网络,网络中有大量“不活动”的用户,即“僵尸”用户,显然,使用包含大量“僵尸”用户的关系网来度量用户影响力是不够准确的。另外,还有研究仅使用被转发数、被评论数、用户被提及数等来度量用户影响力,这些指标虽然考虑到了用户之间的交互行为,但是仅利用了交互行为在单个目标上的聚合结果,没有考虑用户间交互行为的强弱和相互作用。

因此,文中将微博用户间的交互行为作为度量用户影响力的标准,首先基于用户交互行为建立活动网络(activity network)<sup>[11-13]</sup>,结合用户交互行为强度、用户在网络中的活跃度、PageRank 算法等因素,提出了一种基于活动网络来评估微博用户影响力的新算法。

## 2 建立活动网络

### 2.1 活动网络在微博中的运用

活动网络是在用户交往的基础上建立起来的网络,但是这种交往并不一定是友好关系,这种网络的拓扑结构称为活动拓扑结构<sup>[9]</sup>。例如,微博用户通常通过互相关注(follow)、互相提及(mention)、转发动态(retweet)和评论回复(remark)四种行为进行网络互动行为。通过这几种主要行为来构建活动网络。

### 2.2 微博活动网络构建算法

以三元组  $G_r = (V, E, D)$  形式表示活动网络  $G$ , 其中:

$r \in R = \{\text{follow, mention, retweet, remark}\}$ : 活动网络对应的用户交互行为;

$V = \{v_i | i = 1, 2, \dots, N\}$ : 活动网络的节点集合,每个节点对应一个微博用户,  $N$  为网络中的节点数目;

$E = \{(v_i, v_j, w_{ij}) \mid v_i, v_j \in V, i \neq j, v_i R v_j\}$  : 活动网络的边集合,  $v_i R v_j$  定位为节点  $v_i$  和节点  $v_j$  间产生了交互  $R$ ,  $w_{ij}$  为边的权值, 与交互  $R$  的次数、类型等相关;

$D = \{d_i \mid i = 1, 2, \dots, L\}$  : 活动网络中节点的属性向量, 对于每一个属性向量, 都有一个节点与之对应。

微博用户的所有交互活动内容都在微博文本上得以体现。结构化信息可以从半结构化微博文本中提取, 从而间接完成网络活动的构建, 这些结构化信息主要包括微博用户的网络节点及其相互之间的联系。

构建微博活动网络的主要过程为:

(1) 进行节点获取, 得到用户基本信息, 把微博用户的节点集结起来, 形成集合  $V$ 。

(2) 分析节点间的关系, 通过对微博用户互动行为的分析, 把用户节点间的关系集结起来, 形成集合  $E$ , 把用户交互行为的次数记录下来, 当作边的属性。

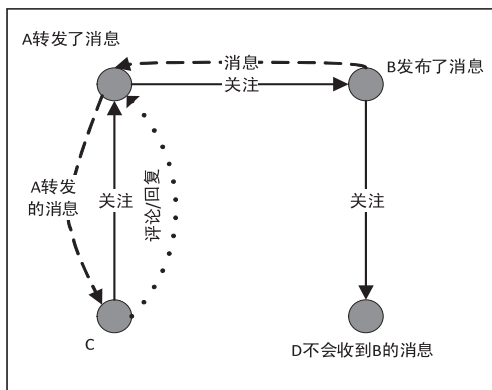
(3) 提取网络属性, 分析数据和网络需求, 通过对用户节点属性的构造, 得到边与节点之间的属性结果, 比如节点的访出和访问可以由用户的好友数和粉丝数来代表, 可以作为用户网络行为的特征表示, 用户相互间的行为边属性特征等也可以因此表示出来。选取属性时可以参考数据的储量和算法需求程度对其进行合理设置。

### 3 基于活动网络的微博用户影响力分析

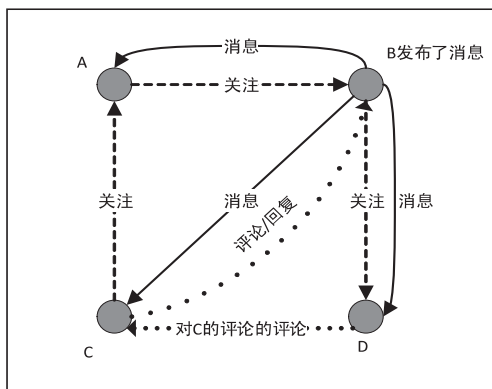
在微博网络中, 用户之间的联系是通过关注关系形成的, 但是这个关注是单方面的用户网络行为, 用户  $v$  可以单向关注用户  $u$ , 而且并不用得到用户  $u$  的同意确认, 甚至可以把用户  $u$  的动态随时随地发给其他关注了用户  $u$  的人  $\text{follow}(u)$ , 对用户  $u$  的评论、回复和转发等操作都有限制, 这样的信息交互过程如图 1(a) 所示。通过关注关系建立起来的为好友关系网络, 利用这个网络建立信息传播和交互机制, 用户可以方便地获取信息和发布消息, 并对其他用户产生影响。因此, 传统对用户影响力的评价都是通过分析好友关系网络的拓扑特征实现的。

但是, 还有一些特殊功能, 不同的用户分为了不同的圈子, 圈子内的用户可以随意互相评论其他用户的微博动态, 而不用互相关注。这种消息传播模式的一个应用如新浪微博微吧, 其中, 用户圈子内信息传播的主要途径不再是由关注行为构成的好友关系网络, 而是通过评论、转发、提及等交互行为构成的新网络, 如图 1(b) 所示。

所以, 文中首要任务就是在微博用户行为分析的基础上建立活动网络, 再借鉴 PageRank 算法的随机游走思想建立用户影响力分析模型, 最后结合节点的行为特征对链接分析过程进行优化。



(a) 依赖关注关系的信息传播模式



(b) 不依赖关注关系的信息传播模式  
(以新浪微博为例)

图 1 微博网络中信息的不同传播模式

#### 3.1 PageRank 算法

PageRank 算法是 Google 搜索重要的网页排名算法, 它主要是通过网页之间的链接结构来确定网页的重要性。由于微博网络中的用户节点链接关系与网页节点间的链接关系相似, 可以用 PageRank 算法来分析在线社会网络中重要用户、发现最优话语权的用户。PageRank 算法的基本原理如下: 凡是一个网页受到用户的拥护和追捧, 而得到了很高的排名, 那么它必然与许多其他网页相链接。PageRank 算法的两个基本假设: (1) 网页的重要程度是和页面的入链数呈正比的; (2) 网页的入链质量越高, 网页就越重要, 从而就会向其他页面传递更多的权重。PageRank 算法可以表述为:

$$\text{PageRank}(p_i) = \frac{1-q}{N} + q \sum_{p_j \in M(p_i)} \frac{\text{PageRank}(p_j)}{|L(p_j)|} \quad (1)$$

其中,  $p_1, p_2, \dots, p_N$  表示网页;  $M(p_i)$  表示页面  $p_i$  的页面链入数;  $L(p_j)$  表示页面  $p_j$  的页面链出数;  $N$  表示网络中的页面数量;  $\text{PageRank}(p_i)$  表示页面  $p_i$  的 PageRank 值, 所有页面的 PageRank 值构成网络的 PageRank 向量;  $q$  表示用户继续浏览该页面之后的页面的概率, 通常概率值为 0.85;  $1-q$  表示用户浏览该页面之后的页面的概率。



文中将 PageRank 算法的思想用于微博网络用户影响力排名,但是微博活动网络与互联网有很大不同,传统的 PageRank 算法有很大的局限性。

(1)传统 PageRank 算法只是在 Web 网络的基础上建立起来的页面链接关系,符合二元关系,这种关系大部分建立在微博好友关系的基础上,不过大多数时候,比如在热门话题的讨论、微吧、微群中,用户信息并不是沿着好友关系链传播而是通过特定的用户分组或用户群传播,用户间不需要添加为好友就可以直接进行交互。因此将 PageRank 算法用于基于回复/评论的活动网络,而不是好友关系网络。

(2)传统 PageRank 算法没有考虑到节点间的相互作用是不同的。微博用户的相互联系越强烈,相互间的影响就越大。在微博好友关系网中,边仅代表关注与被关注关系,一次即可完成,微博用户之间在互相评论、回复、转发等多次交互行为时产生了很多相互叠加的关系。PageRank 算法中在计算 PageRank 数值时,假设权值转移是相同的,并没有把节点的重要性考虑进去。

(3)传统 PageRank 算法只考虑了静态节点,没有考虑动态节点,只考虑了网页节点之间的关系对节点的重要影响,没有考虑微博用户的行为特征,这种行为特征是具有自主能动性的。在分析节点之间的传播上,PageRank 数值只考虑了节点的结构,没有考虑节点的非结构行为带来的活跃特征程度。在网络活动中用户的行为例如发布微博动态、评论回复、转发等都会在很大程度上影响信息的传播,PageRank 算法没有很好地体现这一特征。

3.2 在微博用户网络活动基础上的影响力分析算法

为了更准确地衡量微博用户的影响力,适应活动网络的特点,文中对 PageRank 算法进行了改进,设计了基于活动网络的用户影响力排序算法—ANR 算法。

已知活动网络  $G_r = (V, E, A)$ , 算法描述为:

$$r(u_i) = (1 - q)\lambda_i + q \sum_{u_j \in \text{fow}(u_i)} \mu_j^i r(u_j) \tag{2}$$

其中,  $r(u_i)$  表示对用户  $u_i$  的 rank 评价分数;  $\text{fow}(u_i)$  表示用户  $u_i$  的粉丝数;  $\text{foe}(u_i)$  表示用户  $u_i$  的关注数;  $N$  表示网络中用户数量;  $q$  表示阻尼系数,取值 0.85;  $\lambda_i$  表示在各个节点中用户  $u_i$  的相对重要性;  $\mu_j^i$  表示  $u_i$  在  $u_j$  的每一个关注用户集合中的相对重要性,为用户  $u_j$  传递给用户  $u_i$  的 rank 值的权重比。

定义  $h(u_i)$  为用户  $u_i$  在网络中的活跃度,  $g(u_j, u_i)$  为  $u_j$  与  $u_i$  之间的交互强度,则  $\lambda_i$  和  $\mu_j^i$  可由下列公式求得:

$$\lambda_i = \frac{h(u_i)}{\sum_{u_i \in V} h(u_i)} \tag{3}$$

$$\mu_j^i = \frac{g(u_j, u_i)}{\sum_{u_k \in \text{foe}(u_j)} g(u_j, u_k)} \tag{4}$$

同时定义用户  $u_i$  对用户  $u_j$  的交互序列为:  
 $\text{actions}_{i \rightarrow j} = r_1, r_2, \cdots, r_m, r_m \in R, m = 1, 2, \cdots, M$  (5)

用户节点活跃度  $h(u_i)$  为用户评论他人和发帖的总数:

$$h(u_i) = \sum_{t \in \text{topic}(i)} \lambda_t + \sum_{u_i \in V, k \neq i} (|\text{actions}_{i \rightarrow k}|) \tag{6}$$

用户间交互强度  $g(u_j, u_i)$  为用户间交互次数:  
 $g(u_j, u_i) = (|\text{actions}_{j \rightarrow i}| + |\text{actions}_{i \rightarrow j}|) \cdot \varphi_j^i \tag{7}$

其中,  $\text{topic}(i)$  表示用户  $u_i$  所发的帖子集合;  $\lambda_t$  表示帖子  $t$  的强度,文中不考虑帖子的规模对用户活跃度的影响,取  $\lambda_t \equiv 1$ ;  $|\text{actions}_{i \rightarrow j}|$  表示交互序列的长短;  $\varphi_j^i$  表示边强度与用户  $u_i$  与  $u_j$  间的交互变换系数,它同时受节点间的兴趣相似程度和关系亲密程度等方面的影响,并且展现了特殊节点。对特定节点间的交互强度网络影响进行了分析研究,取  $\varphi_j^i \equiv 1$ ,即不考虑节点间的兴趣相似性等因素对节点间交互强度的影响。

4 实验

4.1 数据集

文中实验数据来自于新浪微博中的“微吧”。微吧是一个基于主题讨论与交流的社区,上线于新浪微博,实现了微博内容的按主题聚合、深度讨论和潜在热点挖掘。微吧的显著特点为同一微吧内的用户都是关注同一主题的,具有较高的聚合度和活跃度,用户不需关注即可获取微吧内其他用户所发的消息,用户间的主要交互方式为评论或回复,适合进行影响力分析。

通过网络爬虫抓取了新浪微博郑州微吧部分帖子及其回复信息,抽取参与讨论的用户信息以及微吧内用户间的交互信息建立活动网络,并使用新浪微博 API 采集用户的属性特征,比如用户的朋友数量、粉丝数量、发帖数等。同时,可以通过对抓取的帖子的分析得到用户在微吧内的相应属性特征。其基本信息如表 1 所示。

表 1 基本信息

| 属性       | 值         | 属性        | 值       |
|----------|-----------|-----------|---------|
| 帖子数      | 15 631    | 用户数       | 54 194  |
| 交互用户数    | 3 592 792 | 好友关系数     | 621 982 |
| 最大人数参与量  | 937 015   | 最小人数参与量   | 0       |
| 用户发帖数最大量 | 1 685     | 用户发帖数最小量  | 0       |
| 用户发帖量平均量 | 0.3       | 帖子参与人数平均量 | 229.1   |
| 用户无交互数   | 231       | 用户非好友关系数  | 5 617   |

通过对表 1 的分析可以发现,微吧内每个用户交互量是用户关注关系量的 5.8 倍,相比于非好友用户数量,在 微博贴吧内完全没有交互关系的用户数量只占了 4.1%。图 2 和图 3 是用户在微吧内和在全网络的好友关系数量的 CDF 图,通过对所有用户微吧内和全网内关系数量的研究得出结论,微博贴吧内的朋友数量和粉丝数量仅占全网朋友数和粉丝数的一小部分,分别是 5.2% 和 1.2%。间接说明了基于微博用户交互的活动网络比仅使用好友关系网络来分析用户影响力更具有合理性。

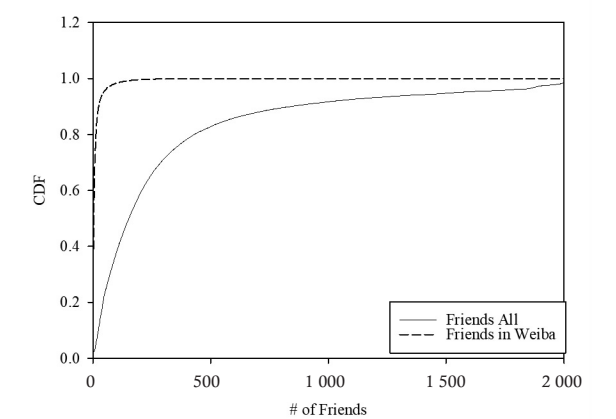


图 2 用户在微吧内的朋友数和全网的朋友数

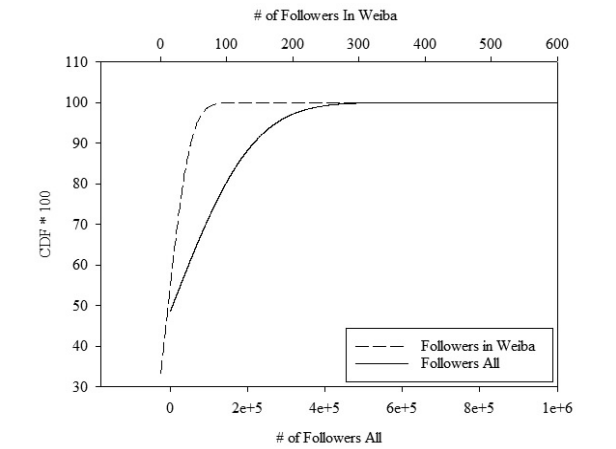


图 3 用户在微吧内的粉丝数和全网的粉丝数

4.2 实验及分析

4.2.1 对比算法

为验证 ANR 算法,运用以下两种常用影响力度量方法作为对照:

(1)PageRank 算法:PageRank 算法是经典的网络算法,微博中的用户影响力排序就是通过该算法进行的。另外,ANR 算法是在 PageRank 算法的基础上优化得来的,所以 PageRank 算法是基本对比算法之一,并且是在好友关系网络的基础上建立起来的。

(2)粉丝数:使用关注微吧中用户的粉丝数进行影响力排序,称为 FansC (fans count) 算法。Twitter 等

已经使用了该方法,它也被当作对比算法作为许多文献的参考。该方法也是文中的基准算法。

4.2.2 实验结果

表 2 左侧给出了在 ANR、PageRank 和 FansC 三种影响力排序算法的基础上得到的 Top-20 名用户。观察可得,PageRank 与 FansC 两种方法得到的 Top-10 有相同用户重叠(不考虑排名),Top-20 也有 14 名重叠用户—得到的结果与其他文献的研究结果相同<sup>[1]</sup>,PageRank 与 FansC 两种方法得到的用户排名结果是近似的。

表 2 三种排序方法的前 20 名用户

| 用户名       | Rank |    |       |     |        |       |           |
|-----------|------|----|-------|-----|--------|-------|-----------|
|           | ANR  | PR | FansC | 朋友数 | 粉丝数    | 发帖数   | 被评论数      |
| 郑州微吧      | 1    | 1  | 1     | 411 | 26 088 | 1 685 | 2 113 023 |
| 吧娘        | 2    |    |       | 94  | 330    | 105   | 4 509     |
| 一米 Andrew | 3    |    |       | 1   | 5      | 1     | 449       |
| 樱空旭       | 4    |    |       | 519 | 365    | 632   | 5 992     |
| 新浪河南      | 5    | 2  | 2     | 214 | 7 466  | 136   | 6 474     |
| _馨小馨钱多多_  | 6    |    |       | 9   | 11     | 1     | 282       |
| 花小呢       | 7    |    |       | 107 | 163    | 285   | 2 834     |
| 凡心        | 8    |    |       | 256 | 355    | 299   | 1 254 658 |
| 新浪郑州汽车    | 9    |    |       | 164 | 814    | 5     | 2 086     |
| 栗小穀       | 10   |    |       | 0   | 268    | 428   | 3 092     |
| 微探店       | 11   |    | 18    | 140 | 3 117  | 47    | 1 156     |
| Nice-u    | 12   |    |       | 29  | 47     | 1     | 363       |
| 翩翩先森      | 13   |    |       | 12  | 11     | 152   | 956       |
| 郑在发生      | 14   | 6  | 7     | 392 | 5 634  | 73    | 2 505     |
| 小韩喔喔      | 15   |    |       | 465 | 291    | 472   | 3 454     |
| 豫小浪       | 16   |    |       | 82  | 78     | 178   | 122 699   |
| _郭非非      | 17   |    |       | 35  | 50     | 169   | 1 717     |
| 郑州吧少      | 18   |    |       | 6   | 6      | 3     | 62        |
| 嘉惠有个小心思   | 19   |    |       | 2   | 12     | 1     | 53        |
| 文雯文-      | 20   |    |       | 11  | 29     | 1     | 53        |
| 大河报       |      | 3  | 3     | 156 | 7 337  | 5     | 18        |
| 微群小助手     |      | 4  | 6     | 8   | 6 159  | 0     | 0         |
| 薛蛮子       |      | 5  | 4     | 11  | 6 803  | 0     | 0         |
| 河南商报      |      | 7  | 9     | 153 | 4 494  | 5     | 102       |
| 打鱼晒网韩佳    |      | 8  | 5     | 84  | 6 209  | 1     | 310       |
| 平安中原      |      | 9  | 10    | 163 | 4 046  | 5     | 365       |
| 延参法师      |      | 10 | 8     | 24  | 5 034  | 1     | 228       |
| 微播河南      |      | 11 | 12    | 139 | 3 793  | 7     | 173       |
| 大河网       |      | 12 |       | 344 | 1 958  | 2     | 0         |
| 郑州晚报      |      | 13 | 11    | 140 | 3 888  | 5     | 60        |
| 杨华民       |      | 14 | 20    | 363 | 3 031  | 0     | 0         |
| 东方今报      |      | 15 | 16    | 221 | 3 180  | 0     | 0         |
| 微吧小管家     |      | 16 |       | 8   | 2 285  | 0     | 0         |
| 话题        |      | 17 |       | 4   | 2 498  | 0     | 0         |
| 郑州市城市管理局  |      | 18 |       | 386 | 2 414  | 0     | 0         |
| 郑州铁路局     |      | 19 |       | 36  | 2 632  | 5     | 9         |
| 汽车 912    |      | 20 | 14    | 84  | 3 229  | 0     | 0         |
| 成都微吧      |      |    | 13    | 25  | 3 396  | 0     | 0         |
| 郑州同城会     |      |    | 15    | 288 | 3 196  | 8     | 97        |
| 打鱼晒网      |      |    | 17    | 13  | 3 125  | 0     | 0         |
| 快乐郑州      |      |    | 19    | 64  | 3 082  | 0     | 0         |

表 3 给出了上述三种排序算法以及按发帖数、评

论数 Top-10 和 Top-20 的共同覆盖率。可以看出, ANR 算法得到的 Top-10/20 用户与用 PageRank 和 FansC 方法得到的用户的相同率仅有 20% 左右。ANR 算法是否可行, 可以从排名结果上看出, 排名结果可以使用发帖数和被评论数进行分析—这两个指标在一定程度上表现了用户行为的数量和质量, 文献[4,14]的结论表明这两项都能够较好地表示用户的影响力。

表 3 不同方法 TopN 共同覆盖用户比例

| TopN | PageRank | 粉丝数 | 发帖数  | 被回复数 |
|------|----------|-----|------|------|
| 10   | ANR      | 20% | 20%  | 50%  |
|      | PageRank |     | 100% | 10%  |
|      | 粉丝数      |     | 10%  | 20%  |
|      | 发帖数      |     |      | 40%  |
| 20   | ANR      | 15% | 20%  | 55%  |
|      | PageRank |     | 75%  | 10%  |
|      | 粉丝数      |     | 10%  | 20%  |
|      | 发帖数      |     |      | 65%  |

表 3 给出了 Top-20 用户的相关属性特征, 分析可得如下结论:

(1) ANR 算法预测得到的 Top-20 用户的发帖数和被评论数普遍较高; 在所预测的 Top-20 用户中, 三种算法的平均发帖数量分别为 233.7、96.5 和 98.9; 而对于不在 ANR Top-20 的用户中, 基于 PageRank 和基于粉丝数 FansC 两种算法的平均发帖量仅为 2.1 和 2.3。另外, 可以看出, 不在 ANR 算法 Top-20 的用户中有一部分没有或者很少在微吧中发帖子。

“影响”在维基百科和现代汉语词典中定义为“对别人的思想、情感或行为起作用”, 发帖数体现了用户的活跃程度, 被回复数则体现了该用户的影响力, 是否影响了别人。通过结论(1)的分析可知, ANR 算法预测的 Top-20 用户发帖数和被评论数都比 PageRank 和 FansC 算法预测的高, 也证明文中算法是有效的。

(2) 分析 ANR 算法得到 Top-20 用户, 发现一些发帖数较少的用户也有所出现。例如“一米 Andrew”、“\_馨小馨钱多多\_”、“Nice-u”等三个用户的发帖数仅为 1, 但仍然排进了 Top-20。分析这些用户的评论发现: 这些用户的评论中存在大量自己的回复, 而 PageRank 算法计算中无法识别这类问题, 会大幅度提高这类节点的得分值。另外还发现, 有些用户只活跃在早期, 但排名仍然靠前。典型用户为“新浪郑州汽车”, 只有 5 个帖子的发帖量。这 5 个帖子都是在微博建立之初发布的, 但是由于一些微吧活跃的用户, 如“樱空旭”、“凡心”等, 使得该用户的影响力比实际要高很多。

结论 2 表现了 ANR 算法的问题, 即 Native PageRank 算法本身具有的缺点, 在对时间的感觉上不是很

敏感。  
不过在预测用户影响力的排名时, PageRank 和 FansC 方法基本无法准确预测出微吧内的有影响力用户, 而 ANR 算法能较准确找出有影响力的用户, 也就是说文中算法对于预测用户影响力在像微吧这样的应用中是有效的。

5 结束语

通过分析现有对微博用户影响力的相关研究, 提出一种在活动网络基础上的新算法-ANR。该算法针对传统微博用户影响力评估的一些问题, 考虑用用户交互行为构建活动网络, 结合个人的属性特征和行为特征两方面体现在 PageRank 算法中, 进而预测用户影响力。通过对微吧数据的分析表明, 基于活动网络的微博用户影响力算法对影响力的识别准确性超过了基于好友关系网络的 PageRank 算法和基于粉丝数排序的算法。另一方面, ANR 算法的种种问题也在实验中表现出来, 忽视了用户的自我评论倾向和时间对用户评论行为预测的影响。

影响微博用户准确评估涉及的因素是多方面的<sup>[15]</sup>。文中单独举出了评论这一种网络行为, 其他的交互模式以及用户属性特征也会对用户影响力预测产生较大影响, 如果能将多种因素有效结合将大大提高用户影响力预测的准确度。

参考文献:

[1] KWAK H, LEE C, PARK H, et al. What is Twitter, a social network or a news media? [C]//Proceedings of the 19th international conference on world wide web. [s. l.]: ACM, 2010:591-600.

[2] CHA M, HADDADI H, BENEVENUTO F, et al. Measuring user influence in Twitter: the million follower fallacy[C]//International conference on weblogs and social media. Washington, DC, USA: [s. n. ], 2010.

[3] LEAVITT A, BURCHARD E, FISHER D, et al. The influentials: new approaches for analyzing influence on twitter[J]. Web Ecology Project, 2009, 4(2):1-18.

[4] 马俊, 周刚, 许斌, 等. 基于个人属性特征的微博用户影响力分析[J]. 计算机应用研究, 2013, 30(8):2483-2487.

[5] PAGE L, BRIN S, MOTWANI R, et al. The PageRank citation ranking: bringing order to the web[R]. [s. l.]: [s. n. ], 1999.

[6] TUNKELANG D. A twitter analog to pagerank[EB/OL]. 2009-01-13. <http://thenoisychannel.com/2009/01/13/a-twitter-analog-to-pagerank>.

[7] WENG Jianshu, LIM E P, JIANG Jing, et al. TwitterRank;



程操控对作物进行施肥;打药记录模块主要记录了农药的名称,喷洒农药的时间、地点、面积等信息,管理人员可以手动或远程操控对作物进行农药的喷洒,进而对作物的生长情况进行监控;采摘记录模块主要记录了作物采摘的情况,记录了采摘作物的名称,采摘的时间、地点、重量等信息,通过信息的记录,避免了未及时采摘和错误提前采摘情况的发生。

7.2 数据分析功能

通过各种集成技术,该农业物联网监控系统可对农作物周边环境进行空气温度、湿度,土壤和光照强度等信息的自动采集,并远程实时传输。

在该系统运行工作时,传感器技术起着绝对重要的作用,数据采集中的每一项参数都由安置在特定位置的特定传感器进行采集。若传感器非正常工作,那么通过该系统进行监管和决策的工作人员将会收到反馈。在数据分析的专栏中,系统可自动显示所有传感器在过去 24 小时内的工作状态,如果有传感器出现非正常运行状态,即出现故障,那么该系统将会第一时间出现警报提示,以提醒用户进行及时处理。

8 结束语

文中介绍了农业物联网的相关概念以及 MQTT 协议的协议解析、工作原理和消息推送机制。基于 MQTT 消息推送设计了农业物联网监控系统,系统除了数据采集以外还提供了远程控制、数据分析、农资和设备监管等功能,并时刻与用户保持联系。该系统能保证作物的顺利生长,方便了用户的监管与决策,为面向农业的互联网+应用提供了一些可行的参考。

参考文献:

[1] 李灯华,李哲敏,许世卫,等. 先进国家农业物联网的最新

进展及对我国的启示[J]. 江苏农业科学,2016,44(10):1-5.

[2] 毛燕琴,沈苏彬. 物联网信息模型与能力分析[J]. 软件学报,2014,25(8):1685-1695.

[3] 石东源,卢炎生,王星华,等. SVG 及其在电力系统软件图形化中的应用初探[J]. 继电器,2004,32(16):37-40.

[4] YANG S H. Internet of things [M]//Wireless sensor networks. London: Springer,2014:247-261.

[5] 刘爱军. 物联网技术现状及应用前景展望[J]. 物联网技术,2012,2(1):69-73.

[6] 李 瑾,郭美荣,冯 献. 农业物联网发展评价指标体系设计:研究综述和构想[J]. 农业现代化研究,2016,37(3):423-429.

[7] 李道亮. 物联网与智慧农业[C]//2017(第五届)全国水利信息化论坛. 出版地不详:出版者不详,2012.

[8] 张长利,沈维政. 物联网在农业中的应用[J]. 东北农业大学学报,2011,42(5):1-5.

[9] STANFORD-CLARK A,TRUONG H L. MQTT for sensor networks (MQTT-S) protocol specification[S]. [s. l.]: IBM,2013.

[10] 彭歆北. 物联网技术在现代农业中的应用[J]. 中国电信业,2013(8):64-66.

[11] 林 浒,张家铭,杨海波. 基于 MQTT 协议的即时消息业务设计与实现[J]. 计算机系统应用,2017,26(3):219-224.

[12] 姜 妮,张 宇,赵志军. 基于 MQTT 物联网消息推送系统[J]. 网络新媒体技术,2014,3(6):62-64.

[13] LIGHT R A. Mosquitto: server and client implementation of the MQTT protocol[J]. Journal of Open Source Software, 2017,2(13):265-266.

[14] 曾 昂,李 宁,严 俊. Mosquitto 大文件传输方式的研究与改进[J]. 计算机工程与应用,2017,53(4):123-127.

[15] 许金喜,张新有. Android 平台基于 MQTT 协议的推送机制[J]. 计算机系统应用,2015,24(1):185-190.

(上接第 167 页)

finding topic-sensitive influential twitterers[C]//Proceedings of the third ACM international conference on web search and data mining. New York, NY, USA: ACM, 2010: 261-270.

[8] LÜ Linyuan,ZHANG Yicheng,YEUNG C H,et al. Leaders in social networks, the delicious case[J]. PloS ONE,2011,6(6):e21202.

[9] 徐 恪,张 赛,陈 昊,等. 在线社会网络的测量与分析[J]. 计算机学报,2014,37(1):165-188.

[10] 张 赛,徐 恪,李海涛. 微博类社交网络中信息传播的测量与分析[J]. 西安交通大学学报,2013,47(2):124-130.

[11] CHUN H,KWAK H,EOM Y H,et al. Comparison of online social relations in volume vs interaction: a case study of cy-world[C]//Proceedings of the 8th ACM SIGCOMM confer-

ence on internet measurement. Vouliagmeni, Greece: ACM, 2008:57-70.

[12] JIANG Jing,WILSON C,WANG Xiao,et al. Understanding latent interactions in online social networks[J]. ACM Transactions on the Web,2013,7(4):18-19.

[13] NAZIR A,RAZA S,CHUAH C N. Unveiling facebook: a measurement study of social network based applications [C]//Proceedings of the 8th ACM SIGCOMM conference on internet measurement. New York, NY, USA: ACM,2008: 43-56.

[14] 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,2011,29(6):8-16.

[15] 丁兆云,贾 焰,周 斌,等. 社交网络影响力研究综述[J]. 计算机科学,2014,41(1):48-53.