

文章编号: 1003-0077(2017)04-0184-07

微博社交网络的用户影响力评价方法

吴 慧, 张绍武, 林鸿飞

(大连理工大学 计算机科学与技术学院, 辽宁 大连 116024)

摘 要: 该文主要研究在微博社交网络中怎样评价用户的影响力。在影响用户影响力的众多因素中, 该文认为用户的传播能力越强, 用户的信息便可以更快地在网络中扩散, 其影响力也越大。和传统的用户影响力评价方法相比, 该文综合考虑用户的活跃度和用户所发微博质量两个方面的因素, 得到用户的影响力权重, 然后把每一个用户作为社交网络中的节点, 计算其在社交网络中的影响力。通过在公开语料集和真实数据中的实验, 表明该方法是可行的, 比传统的用户影响力评价方法更能客观、真实地反映用户的实际影响力。

关键词: 社交网络; 用户影响力; 活跃度; 微博质量

中图分类号: TP391

文献标识码: A

Evaluation of the User's Influence on Microblog

WU Hui, ZHANG Shaowu, LIN Hongfei

(School of Computer Science and Technology, Dalian University of Technology, Dalian, Liaoning 116024, China)

Abstract: This paper investigates the evaluation of the user influence on Sina microblog. Among various factors, a user is considered as more influential if his information is disseminated faster to a larger extent. Compared with traditional methods, the user's active degree and the quality of posts are both taken into consideration. Treating each user as a node in the social network, the final user influence is estimated. The experiments on both public dataset and real dataset from Sina microblog show the validity of the method.

Key words: social network; user influence; active degree; microblog quality

1 引言

微博作为一种新型的在线媒介形态,正在悄无声息地改变着人们的生活和思考方式。在微博时代,每个人好像都有一个麦克风,在 140 字的信息中分享着自己心情和故事,同时表达着自己对世界的认识,微博因此成为舆论的重要发源地之一,在我们所形成的社交网络中扮演着越来越重要的作用^[1-2]。微博具有评论、转发、关注等多种功能,可以加快信息的传播。每一个用户所发的微博都可以引起其他用户的关注、转发和评论,从而在微博社交网络中产生一定的影响。当然,每一个用户的影响力也是截然不同的。一般来

说,拥有较多粉丝的用户影响力也较大,其言论和观点可以被更多的人看到,在微博中产生一定的影响,可以推动舆论的高潮,甚至对舆论的走向起到推波助澜的作用。然而由于僵尸粉的存在,使得简单的用粉丝数来衡量用户的影响力在实际的应用中存在弊端。用户微博的转发数作为衡量用户影响力的因素之一,可以反映用户所发微博的质量及用户的影响力情况,微博被广泛地转发,相应的微博信息可以更快地在网络中传播,也可以反映出用户的影响力。将用户放入到整个社交网络中,用户表示节点,用户之间的互粉表示链接边,可以形成整个网络图,可以通过网络的拓扑结构得到每个节点的影响力。对用户的影响力评估有不同的方法,每一种方法都有其侧重点,在本

收稿日期: 2014-09-25 定稿日期: 2015-04-23

基金项目: 国家自然科学基金(61562080, 61632011, 61572102); 国家重点研发计划(2016YFB1001103); 教育部留学回国人员科研启动基金和高等学校博士学科点专项科研基金资助课题(20090041110002)

文中,影响力的衡量主要是信息在网络中的传播速度。用户越活跃,其所发的微博内容被更多的人转发和评价,该用户通过微博所承载的信息也能够更快地在网络中扩散,广为人知,那么该用户的影响力也越大。

在实际的应用领域,对用户的影响力进行合理的评估,可以在微博社交网络中找到影响力较高的用户,可以对影响力较高的用户进行广告推荐或植入,该用户的信息在网络中可以更快地传播,进而可以花费最小的代价在一个社交网络中进行广告宣传。在舆情分析中,可以找到影响力较大的用户,因该用户的言论在社交网络中起到领袖的作用,通过对该用户言论的分析和引导,就可以对舆情的控制起到积极的作用。

本文在评价用户影响力方面,考虑制约用户影响力的诸多因素,比如用户的粉丝数、用户微博的转发数等,得到用户在微博社交网络中的活跃积极性及其所发微博的质量,计算每一个用户的权重,根据用户的权重进行影响力排名。将此排名和传统的方法排名进行比较和分析。

本文的结构安排如下:第二节介绍了在微博影响力研究方面的相关工作;第三节介绍了加入用户权重的影响力评价方法,同时说明了在评价微博社交网络中一般的传统评价方法;第四节介绍了实验部分,包括实验数据的来源和实验的结果,并对结果进行了分析和对比;第五节介绍了对本文的总结,以及下一步的工作设想。

2 相关工作

对社交网络中用户影响力的研究,大体上可以从以下几个方面进行。

(1) 从用户的静态属性来研究。Meeyoung Cha 等^[3]在对 Twitter 社交网络中用户影响力评价的研究中,从用户的粉丝数、用户的转发数等排名来研究用户的影响力,其中粉丝数较多的一般为知名公众人物,反映了现在的流行趋势;转发数较多的一般是媒体,其微博的内容质量较高。但是该方法所得到的排名只是根据粉丝数或转发数,没有综合考虑包括用户行为在内的一系列指标,使得排名有一定的限制,不能普遍地反映用户的实际影响力。考虑到微博是一种互动的社交网络,石磊等^[4]提出了用户活跃度模型,通过考虑用户粉丝、用户发微博的频率等计算用

户的活跃指数,从而得到用户的活跃度排名。用户活跃度虽然可以作为衡量微博用户影响力的一项指标,但并不能说明用户在微博中参与积极性越高、越活跃,该用户就越具有影响力。Danah Boyd 等^[5]将 Twitter 用户的转发、回复等行为表示成用户的行为权重,在权重的基础上计算用户的影响力。该文章对用户的行为考虑得较周全,但是又忽略了用户的粉丝数等因素。张华平、孙梦姝等^[6]通过分析用户所发的微博数、粉丝数、用户关注数的数值特征,得到用户的影响力模型。但是该论文得到的是用户群体的影响力,比如加 V 用户的影响力比普通用户的影响力高,没有得到个体用户的影响力。

(2) 从社交网络中形成的拓扑结构的角度来研究。Yuto Yamaguchi 等^[7]通过分析用户之间的关注关系,评估每一个用户在社交网络中的影响力。但是该方法仅仅考虑用户被关注边的多少,在网络图中所考虑的因素太少。Weng 等^[8]基于 PageRank 算法,根据用户之间的粉丝联系所形成的网络关系,得到用户的影响力排名。该方法用粉丝数量和质量来衡量用户的影响力,较为直观,实现也较为简单,在 Twitter 社交网络中取得较好的效果,但实际的微博社交网络中存在僵尸粉等现象,粉丝作为微博用户影响力的评价指标并不是很全面。王琛、陈庶樵^[9]改进了传统的 PageRank 算法的用户影响力评价方法,引入了微博传播能力这一概念,加入用户的行为可以更真实地反映用户的影响力。但是该方法提出的微博传播能力只包括了用户评论等很少量的信息,有一定的缺陷。

针对以上两个方面的主要研究,不同的影响力因素有其不同的侧重点,也有将两者结合起来考虑,将用户的属性和用户所在社交网络中的拓扑结构同时进行研究,能较为普遍地反映用户的影响力^[10-11]。

通过对以上各种算法的研究,本文综合考虑反映用户影响力的指标,用户所发微博积极性越高,所发的微博质量越高,微博将被越广泛的转发和评价,其信息也相应地在网络中传播得更快,其影响力也越大。用户的积极活跃性考虑了用户在一段时间内所发原创微博的频率、转发的频率,用户所发微博的质量考虑了用户所发微博在这一段时间内被转发次数和评价次数,从而得到用户的权重,该权重反映了用户的影响力。和传统的用户影响力评价指标比较,本文所提出的加入用户权重的影响力排名更具

合理性,并能客观、真实地反映微博用户的影响力。

3 用户影响力评价方法

3.1 加入权重的用户影响力评价方法

本文所提出的加入用户权重的影响力评价方法主要考虑两个方面的因素:用户的活跃度和用户所发微博的质量。其中用户的活跃度包括用户所发微博和用户转发的微博,活跃度反映了用户参与微博互动的热情和积极性,用户的活跃度越高说明其与其他博主的互动越频繁,更新微博越快,更新的微博的信息可以引起其他用户的好奇而关注围观,进而加快该博主的微博信息传播。用户所发的微博质量包括用户微博被转发和被评论的次数,用户所发的微博质量越高,越容易引起大众的转发和评论,该博主的微博信息也被传播得更快,如图 1 所示。

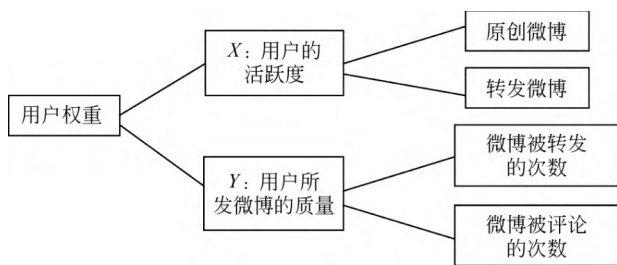


图 1 用户的权重及其相关因素

每一个用户的权重计算公式如式(1)。

$$W_i = X_i + Y_i \quad (1)$$

X_i 是指用户 i 的活跃度。计算用户 i 的活跃度 X_i 时考虑该用户在一段时间 T 内的原创微博数量 P_i 和转发微博数量 R_i , 具体的计算公式如式(2)。

$$X_i = \frac{P_i}{T} + \frac{R_i}{T} \quad (2)$$

Y_i 是指用户 i 所发微博的质量。 Y_i 是用户 i 在一段时间 T 内所有微博影响力的平均值,评价每一条微博的影响力主要考虑微博被评论次数和微博被转发次数两个方面的因素,根据参考文献[12]计算每一条微博的影响力公式如式(3)。

$$y_{i,j} = \sqrt[3]{MR_{i,j}} + \sqrt{MC_{i,j}} \quad (3)$$

其中 $y_{i,j}$ 表示用户 i 的第 j 条微博的影响力, $MR_{i,j}$ 表示用户 i 的第 j 条微博被转发的次数, $MC_{i,j}$ 表示用户 i 的第 j 条微博被评论的次数。得到用户 i 的微博影响力公式如式(4)。

$$Y_i = \frac{\sum_{j=1}^n y_{i,j}}{n} \quad (4)$$

其中, n 是指用户在时间段 T 内所发微博的数量,包括原创微博和转发微博。

3.2 用户影响力排名的评价指标

在对用户影响力排名评价时,常用的评价指标有排名的 Spearman 序列相关系数、Kendall 序列相关系数、重叠率、计算代价等。

其中, Spearman 相关系数反映的两组排名之间的线性相关,该值越接近 +1 或 -1, 两组排名之间呈线性相关, Spearman 相关系数的符号反映了两组排名之间正相关和负相关的关系。符号为正号, 两组排名呈正相关, 符号为负号, 两组排名呈负相关。具体 Spearman 相关系数排名评价方法如式(5)所示。其中, x_i 和 y_i 分别表示在两组排名中的排名序号, N 表示排名总数。

$$\rho = 1 - \frac{6 \sum (x_i - y_i)^2}{N^3 - N} \quad (5)$$

Kendall 相关系数反映了一组排名相对于另一组排名的分歧。其中,如果两组排名是完全吻合的,该值为 +1;如果两组排名分歧最大,该值为 -1;两组排名越一致,其值也越大。具体 Kendall 相关系数排名方法如式(6)所示。其中, P 表示两组排名一致的对数, n 表示排名总数。

$$\tau = \frac{4P}{n(n-1)} - 1 \quad (6)$$

重叠率(overLap)主要是指两组排名在前 N 名重叠的次数(记为 $\text{top } N$), N 的取值可以根据具体的数据集合理选择。具体 overLap 计算方式如式(7)所示。其中, $\text{top } N(x)$ 和 $\text{top } N(y)$ 分别表示两组排名中 $\text{top } N$ 的次数。

$$\text{overLap}(x, y) = |\text{top } N(x) \cap \text{top } N(y)| \quad (7)$$

计算代价是指根据该方法得到的用户影响力评估所消耗的时间、空间复杂度,以便于在具体的研究中根据应用和环境选择理性的评估方法。

4 实验结果与分析

4.1 语料来源

本实验包括两个语料集,第一个是在公开的数据集上的实验。第二个是在真实的数据集上的实验。

在公开数据集中, 本文选取的是 2012 年 *kdd-cup track1* 上的数据集, 该数据集是从腾讯微博爬取的共 90 天内的相关数据, 将得到的数据和数据间的相互关系从以下两个方面概括:

(1) 用户属性。包括用户所发原创微博数、用户转发微博数、每一篇微博被评论的次数、每一篇微博被转发的次数;

(2) 用户关系。包括用户之间的关注、用户的粉丝。

在真实数据集中, 本文选取的是 2012 年新浪微博名人影响力榜 9 月份的数据。其中名人堂中的数据基本上是经过认证的, 数据信息比较真实可信, 数据主要是用户的属性, 包括用户所发微博、转发微博、用户在这段时间内所发微博被转发和评论的次数等。

4.2 对比实验的描述

本文所选取的对比实验是用户的粉丝数、转发数排名及参考文献[13]在网络拓扑结构中得到每一个用户的领袖排名, 即 *leaderRank*, 其中领袖是指有影响力的用户。

(1) 粉丝数。虽然用户的粉丝数多, 不一定说明其影响力大, 但是粉丝数也反映了用户的人气和流行度, 如果微博用户被很多的粉丝跟随, 其微博动态也可以被更多的人看到, 信息可以得到一定的传播。

(2) 转发数。用户的转发数多, 其信息可以更快地在网络中扩散, 可以在一定程度上反映用户的影响力。

文献[13]是基于网络拓扑结构得到的一个领袖排名 *leaderRank*。*leaderRank* 主要是对 *PageRank* 算法^[14]的改进。在微博社交网络中, 用户可以看作是节点, 用户之间的“互粉”形成一条边, 如图 2 所示, 节点 A 和 B 分别表示用户, A 到 B 的边表示 A 是 B 的粉丝。



图 2 用户之间的关系

LeaderRank 算法提出了一个虚拟的根节点 (*ground node*), 该根节点和网络中的节点形成双向的链接, 如图 3 所示。其中, 实线的单向边表示节点之间的粉丝跟随, 虚线所形成的双向边表示根节点和网络中所有节点的链接。

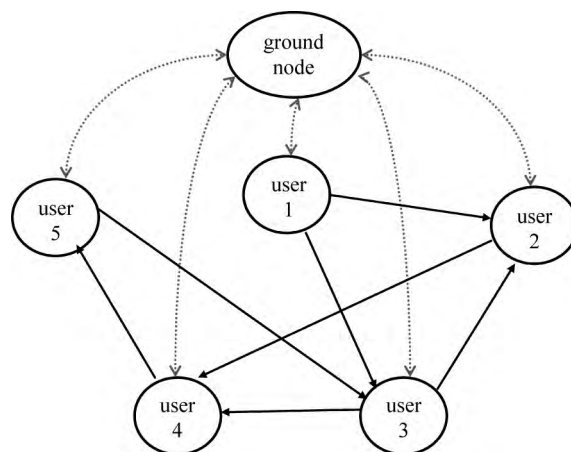


图 3 加入 *ground node* 节点后的网络图

计算节点影响力的公式如式(8)。其中, u, v 表示网络中的节点, $M(u)$ 是指向节点 u 的所有节点集合, $N(v)$ 是节点 v 的出度。从式(8)可以看出, 计算节点的 *leaderRank* 值的过程是一个迭代的过程, 其中初始化网络中每一个节点的 *leaderRank* 值(LR)为 1.0, 而根节点的初始值为 0.0。

$$LR(u) = \sum_{v \in M(u)} \frac{LR(v)}{N(v)} \quad (8)$$

得到每一个节点的 *leaderRank* 值后, 和根节点的 *leaderRank* 值归一化进行累加, 得到最终的 *leaderRank* 值, 计算公式如式(9)所示。其中, N 是根节点的出度, 即网络中的所有节点数。

$$LR(u) = LR(u) + \frac{LR(g)}{N} \quad (9)$$

4.3 实验说明

本文选取了两个数据集, 在公开数据集和在真实数据集中, 针对不同的数据集, 为了说明本文所提出的加入用户权重的方法的效果, 将其与各个影响力排名进行了比较。根据数据集不同的特点, 实验也有所区别。

在公开数据集 *kddcup track1* 的实验中, 分别对用户的粉丝数、用户转发数、基于 *leaderRank* 算法的用户影响力排名和加入用户权重的影响力方法排名, 比较在不同的评价指标中各个排名方法的异同。其中, Spearman 相关系数分别比较两组排名占总排名的 1%、10% 和总排名的相关性。重叠率比较 top 10、top 20、top 50、top 100 出现相同节点的个数。

在真实数据集新浪名人微博社交网络中, 分别对用户的粉丝数、用户转发数及加入权重的影响力方法排名。其中 Spearman 系数比较占总排名 10%

和总排名的线性相关性。因为在名人排名榜中,占总排名 1% 的排名太少,所以没有比较 Top 1% 的 Spearman 系数。重叠率比较 Top 10、Top 20、Top 50 的出现重复用户,真实的名人微博排行榜中数据集较小,所以没有比较 Top 100 出现重复的用户个数。在该数据集中的实验没有基于 leaderRank 算法的用户影响力排名,因为在特定的真实数据集中,给出了用户,用户之间的相互关注链接太少,所得到的节点之间的网络拓扑图也很稀疏,网络结构信息太少,故没有从网络拓扑方面比较各个排名的异同。

4.4 实验结果

分别在公开的语料集和真实数据集中实验,可以得到每种影响力排名下的前 10 名用户排名,从不同的影响力评价方法中,比较任意两组排名。具体的实验结果如表 1~8 所示。

表 1 在 kddcup track1 中各个算法的 Spearman 相关系数

Correlation	Top 1/%	Top 10/%	All
粉丝数 vs. 转发数	0.136 5	0.192 6	0.571 8
粉丝数 vs. 基于 leaderRank 算法	0.903 0	0.792 4	0.860 0
粉丝数 vs. 加入用户权重	0.102 4	0.420 4	0.439 5
转发数 vs. 基于 leaderRank 算法	0.967 7	0.769 4	0.337 8
转发数 vs. 加入用户权重	0.425 6	0.989 2	0.962 4
基于 leaderRank 算法 vs. 加入用户权重	0.697 8	0.755 8	0.297 6

表 2 在 kddcup track1 中各个算法的 Kendall 系数

	Kendall 系数
粉丝数 vs. 基于 leaderRank 算法	0.678 8
粉丝数 vs. 加入用户权重	0.267 3
转发数 vs. 基于 leaderRank 算法	0.066 5
转发数 vs. 加入用户权重	0.316 6

表 3 在 kddcup track1 中各个算法的重叠数 overLap

	Top 10	Top 20	Top 50	Top 100
粉丝数 vs. 转发数	1	2	4	10
粉丝数 vs. 基于 leaderRank 算法	4	13	34	43
粉丝数 vs. 加入用户权重	2	3	8	17
转发数 vs. 基于 leaderRank 算法	0	1	6	35

续表

	Top 10	Top 20	Top 50	Top 100
转发数 vs. 加入用户权重	1	7	20	52
基于 leaderRank 算法 vs. 加入用户权重	1	2	6	36

表 4 在新浪微博名人排名榜中各个算法的 Spearman 系数

Correlation	All	Top 10/%
粉丝数 vs. 转发数	0.542 1	0.539 4
粉丝数 vs. 加入用户权重	0.514 5	0.642 4
转发数 vs. 加入用户权重	0.946 8	0.563 6

表 5 在新浪微博名人排名榜中各个算法的 Kendall 系数

	Kendall
粉丝数 vs. 加入用户权重	0.374 1
转发数 vs. 加入用户权重	0.823 0

表 6 在新浪微博名人排名榜中各个算法的重叠数 overLap

	Top 10	Top 20	Top 50
粉丝数 vs. 转发数	3	8	31
粉丝数 vs. 加入用户权重	3	7	31
转发数 vs. 加入用户权重	7	14	48

表 7 在 kddcup track1 各个算法得到的 Top 10 用户排名

排名	粉丝数排名	转发数排名	基于 leaderRank 算法的排名	加入用户权重的排名
1	1 760 378	2 121 077	1 774 462	2 012 081
2	1 774 717	377 344	330 571	2 105 405
3	1 760 642	1 579 498	1 774 457	1 185 040
4	1 760 350	1 100 559	1 774 520	1 412 782
5	647 356	1 181 393	1 902 321	2 122 259
6	1 419 930	463 485	1 774 705	741 560
7	1 837 210	1 118 081	1 760 654	1 774 586
8	1 774 505	866 409	346 092	1 760 310
9	1 928 678	830 738	1 954 203	1 341 177
10	682 877	414 978	1 774 687	1 774 815

4.5 实验结果分析

由表 1、表 2、表 4、表 5 在不同的数据集中各个排名评价指标中可以看出不同的影响力排名符号都为正号,都是正相关的,这说明了虽然影响力的评价

指标不同,但是各个方法之间也不是毫无关联的,比如粉丝数多并不一定影响力大,但是粉丝数多其影响力不会太小。

表 8 在新浪微博名人排名榜中各个算法得到的
Top 10 用户排名

排名	粉丝数排名	转发数排名	加入用户权重的排名
1	姚晨	加措活佛-慈爱基金	阿信
2	小 S	阿信	刘忻
3	谢娜	吴奇隆	王力宏
4	王力宏	刘忻	加措活佛-慈爱基金
5	何炅	李承鹏	吴奇隆
6	赵薇	张杰	蔡康永
7	杨幂	蔡康永	张歆艺
8	陈坤	张小娴	张杰
9	蔡康永	谢娜	小 S
10	林心如	王力宏	郭德纲

表 1 中粉丝数和转发数相关性较低,粉丝多的其转发不一定高;粉丝数和 leaderRank 算法相关度较高,根据粉丝数得到的影响力较高的,leaderRank 算法得到的影响力也较高,因为 leaderRank 算法是对 PageRank 算法的改进,而 PageRank 算法是与节点的入度(即粉丝数)相关的;转发数和加入用户权重排名相关性较大,两者都反映了用户的信息在微博中传播的速度,而这也说明了加入用户权重的影响力评价方法可以加快微博信息的传播。

表 2 也说明了基于 leaderRank 算法的影响力评价方法和粉丝数的评价方法更一致,而加入用户权重的评价方法和转发数的评价方法更一致。

从表 3 和表 6 的各个排名重叠数可以看出虽然各种排名有所区别,但是无论哪一种排名,随着 top N 中 N 的增大,重叠数也逐渐增大,各种排名之间有一定的相互联系。同时从表 3 中可以看出各个排名侧重点有所不同,排名很少重合。

通过表 4 和表 5,可以看出在真实数据集中加入用户权重的影响力评价方法和用户的转发数相关性较大,用户的权重越大,用户的转发数越多,在微博中也传播得更快。而信息更快更广的传播,使用户的实际影响力也越大。

表 7 和表 8 是在不同的评价方法中得到的 Top 10 的用户排名,每种排名算法得到的排名侧重点不

同,排名也有差异。由表 8 可以看出粉丝数较多的用户一般是活跃流行的明星,转发数较多的一般是微博所发的质量较高、引起共鸣和关注的用户,比如知名媒体等。而加入用户权重的用户影响力排名综合考虑了各种因素,将用户的活跃度和用户所发微博的质量两个方面的因素结合起来,用户越活跃,所发的微博被转发和评论的次数也越高,该用户的信息也可以更快地在社交网络中传播,更能为人所知。

在计算代价上,粉丝数影响力排名最简单直接,只需要统计每一个节点的入度即可。转发数需要累加用户在这一段时间内每条微博的转发数。leaderRank 算法则是基于网络拓扑结构计算每一个用户的 leaderRank,得到每一个节点的入度、出度等,还需要迭代的过程,相比较而言其计算成本较大。加入用户权重的评价方法,需要得知每一个用户所发的微博数和转发数,用户的微博被转发和评论的次数,计算代价适中。

5 结论与未来工作

本文针对微博社交网络中用户影响力排名进行研究,微博用户的信息可以更快地在网络中传播,其影响力也越大。通过分析得到一种加入用户权重的用户影响力算法,该方法考虑了用户本身的活跃积极性和所发微博的质量两个方面。

比较传统的对用户属性的评价方法包括用户粉丝数和用户转发数等,可以使信息在网络中得到更快的传播,在基于网络拓扑结构对用户影响力的评价方法中,本文通过给每一个用户加入一个权重,在计算其在微博社交网络中的影响力,可以突出每一个用户的活跃度和所发微博质量两方面的因素,更加合理地反映用户的实际影响力。

通过本文的研究,可以知道用户影响力包括用户的主动行为和用户所发微博被转发评论等被动行为(用户所发微博质量)。在实际的微博社交网络中,为了提高用户的影响力,用户可以通过活跃的参与微博的发帖、评论、转发加强和其他用户的互动,这样可以引起更多人的关注,增加更多的粉丝,同时可以发布更多高质量的微博,引发大家对微博的转发、评价,引起更多的围观,让信息在网络中更快的传播,提高用户在微博社交网络中的影响力。

评价微博用户的影响力可以从不同的角度给出不同的影响力排名,每一种影响力因素的侧重点有所不同,怎样给出合理的大众接受的影响力排名需

要在不同的应用环境中区分考虑。下一步的工作可以从实际微博社交网络中影响微博用户排名的因素,比如微博认证、微博标签等方面来研究微博用户的影响力评价方法。

参考文献

- [1] Java A, Song X, Finin T, et al. Why we twitter: understanding microblogging usage and communities [C]//Proceedings of knowledge discovery and data mining. 2007: 56-65.
- [2] Zhao D, Rosson M B. How and why people Twitter: the role that micro-blogging plays in informal communication at work[C]//Proceedings of international conference on supporting group work, 2009: 243-252.
- [3] Cha M, Haddadi H, Benevenuto F, et al. Measuring user influence in Twitter: the million follower fallacy [C]//Proceedings of international conference on weblogs and social media, 2010: 10-17.
- [4] 石磊,张聪,卫琳. 引入活跃指数的微博用户排名机制[J]. 小型微型计算机系统, 2012(1): 110-114.
- [5] Boyd D, Golder S A, Lotan G, et al. Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter[C]//Proceedings of hawaii international conference on system sciences, 2010: 1-10.
- [6] 张华平,孙梦姝,张瑞琪,等. 微博博主的特征与行为大数据挖掘[J]. 中国计算机学会通讯, 2014(6): 36-43.
- [7] Yamaguchi Y, Takahashi T, Amagasa T, et al. TURank: twitter user ranking based on user-tweet graph analysis [C]//Proceedings of web information systems engineering, 2010: 240-253.
- [8] Weng J S, Lim E P, Jiang J, et al. TwitterRank: finding topic-sensitive influential twitterers[C]//Proceedings of the 3rd ACM International Conference on Web Search and Data Mining (WSDM 2010). New York, USA. ACM 2010: 261-270.
- [9] 王琛,陈庶樵. 一种改进的微博用户影响力评价算法[J]. 信息工程大学学报, 2013(6): 380-384.
- [10] Ye S, Wu S F. Measuring message propagation and social influence on Twitter. com[C]//Proceedings of social informatics, 2010: 216-231.
- [11] Romero D M, Galuba W, Asur S, et al. Influence and passivity in social media[C]//Proceedings of european conference on principles of data mining and knowledge discovery, 2011: 18-33.
- [12] 原福永,冯静,符茜茜. 微博用户的影响力指数模型[J]. 情报分析与研究, 2012(6): 60-64.
- [13] Lu L, Zhang Y, Yeung C H, et al. Leaders in Social Networks, the Delicious Case [J]. PLOS ONE, 2011,6(6): e21202.
- [14] Page Lawrence, Brin Sergey. The PageRank citation ranking: bring order to the web[R]. Technical report, Stanford Digital Library Technologies Project. 1998.



吴慧(1987—), 硕士, 主要研究领域为文本挖掘、社会关系网络分析和社交媒体处理。

E-mail: 925836442@qq.com



林鸿飞(1962—), 博士, 教授, 博士生导师, 主要研究领域为自然语言处理、情感分析与观点挖掘、信息检索与信息推荐、社会计算与舆情分析、面向生物医学领域的文本挖掘等。

E-mail: hflin@dlut.edu.cn



张绍武(1967—), 博士, 教授, 主要研究领域为文本挖掘、信息检索、自然语言处理、情感计算和观点挖掘、社会计算和舆情分析。

E-mail: zhangsw@dlut.edu.cn