

一种改进的微博用户影响力分析算法

李 勇

(解放军信息工程大学,河南 郑州 450000)

摘要:针对现有用户影响力分析算法的不足,基于网页结构与微博用户网络的相似性,通过分析微博用户之间的行为交互以及传统的 PageRank 算法,提出一种新的 MBUInfluence 算法来对用户影响力进行分析。该算法结合微博用户的活跃度、积极性和传播度等特点,将微博用户影响力定义为由用户自身行为权重和粉丝的影响力构成。通过分析 PageRank 算法,结合微博转发率、评论率、点赞率等特征,制定了微博用户的影响比例函数,形式化定义了微博用户行为权重和粉丝影响力。通过采集新浪微博实验数据,从转发数、粉丝数、新增粉丝数等方面与传统的 FansRank、ForwardRank、PageRank 等影响力排名算法进行对比,从不同角度分析该算法的实际应用效果,并得出微博的用户影响力与单位时间内新增粉丝的数量和质量相关的结论。

关键词:微博;用户影响力;粉丝影响力;PageRank;MBUInfluence

中图分类号:TP391

文献标识码:A

文章编号:1673-629X(2020)08-0027-07

doi:10.3969/j.issn.1673-629X.2020.08.005

An Improved Algorithm of Microblog User Influence Analysis

LI Yong

(PLA Information Engineering University, Zhengzhou 450000, China)

Abstract: Aimed at the deficiencies of conventional user influence analysis algorithms, and based on the similarity between web structure and microblog user network, we propose a novel MBUInfluence algorithm to analyze user influence by analyzing microblog user interactions and the conventional PageRank algorithm. In this algorithm, the influence of microblog user is defined as the combination of the weight of their own behaviors and the influence of their fans in combination with such key microblog user features as activeness, vibrancy and popularity. Combining with the characteristics of microblog forwarding rate, comment rate and thumbing up rate, the influence proportion function of microblog user is formulated by analyzing PageRank algorithm, and the behavior weight of microblog users and the influence of fans are formally defined. Comparing with the traditional ranking algorithms such as FansRank, Forward Rank and PageRank in terms of number of forwards, number of fans, number of new fans and so on, the practical application effect of the proposed algorithm is analyzed from different angles, and the conclusion that microblog user influence is related to the quantity and quality of new fans in a unit time is drawn.

Key words: microblog; user influence; fans' influence; PageRank; MBUInfluence

0 引言

微博(Microblog)是一种新型的网络传媒,它通过140字符左右的文字、图片和链接等信息形式,实现信息的分享。当前中国正处于社会转型时期,虽然经济持续发展,但社会矛盾日益凸显。无论是新疆暴乱、昆明暴恐,还是南海仲裁、钓鱼岛争端等,这些重大突发事件无一不受到人民的强烈关注。通过分析微博网络中用户的影响力,实时掌握准确舆情、制定解决方案、积极引导社会舆论,维护社会和谐稳定。

《现代汉语词典》对“影响力(Influence)”的定义

为:“对别人的思想或行为起作用的能力”。学界对影响力的定义也不尽相同。刘耀庭^[1](2008)认为用户的影响力指的是改变他人思想和行为的能力,其可以通过传播实现。Benevenuto^[2](2009)认为用户影响力是用户之间的相互作用力,这种作用力产生的效果越大,其影响力也就越大。

文中将“用户影响力”定义为对别人的思想或行为产生作用甚至使之改变的能力,也就是指用户在微博消息传播过程中所发挥的重要作用。主要从微博用户影响力的两类因素开展研究,第一类是用户关系^[3],

包括用户之间的关注和粉丝关系,可以认为是它的静态属性;第二类是用户行为^[4],包括用户的发布、转发、评论、点赞、@提及等操作,可以认为是它的动态属性。如果微博用户发布的信息有足够的吸引力,那么他的微博将会被更多的人评论和转发,也就会影响更多的用户,他的影响力就会更大,同时该用户所发布的信息就会传播得越快越远。

1 PageRank 算法分析

PageRank 算法是由 Google 公司的拉里·佩奇和谢尔盖·布林提出并发表。Google 使用 PageRank 算

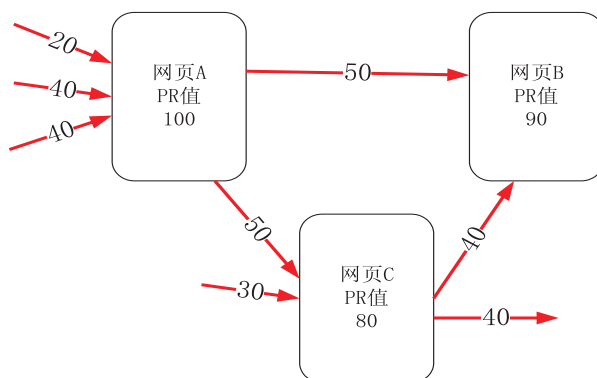


图 1 PageRank 算法 PR 值分配图

PageRank 算法的 PR 值计算公式如下所示:

$$PR(x) = (1 - d) + d \sum_{y \in Y(x)} \frac{PR(y)}{N(y)} \quad (1)$$

其中, $PR(x)$ 表示 x 的 PR 值, $Y(x)$ 表示链入 x 的网页集合, $PR(y)$ 表示 y 的 PR 值, $N(y)$ 表示网页 y 的链接数量, d 是规范化因子。经过大量的实验, Google 认为一般取 $d = 0.85$ ^[3]。通过以上公式可知:页面的 PageRank 值是由所有链入它的页面的重要性决定的,并经过算法迭代实现。

PageRank 算法主要依靠历史节点(例如旧网页)来进行判断,并且由于不同节点(网页)的类型不同,所以还存在一些缺陷。例如:PageRank 算法仅仅依靠网页反向链接的数量来决定 PR 值,链接数量决定了页面的排名,并没有区分反向链接页面的质量,显然垃圾网页和权威页面是不同的;同时 PageRank 算法将网页权值平均分配给它的链出网页,没有区分其链出网页的不同,同样影响了页面的排名质量。因此在将该算法应用到微博用户影响力计算的同时,需要对以下两个问题^[6]进行重点解决:

(1) 解决用户粉丝的“陈旧”问题。

如果“旧粉丝”微博更新不及时,内容陈旧,这些微博用户的影响力可参考性不高;“新粉丝”更新微博及时,能够增加所关注的用户影响力。

(2) 解决用户关注度权重平均的问题。

由于社交网络中节点之间存在差异性,不同的微

法进行网页排名,标识页面等级,继而评价网站的质量高低。PageRank 算法通过网页的链入数量和链入质量来标志页面,并对其进行了量化。网页的重要性^[5]采用 PR 值来表示,每个网页的 PR 值不仅仅要考虑被链接网页的数量,还要考虑链接到该网页的网页质量和重要性的影响,如图 1 所示。

PR 值为 100 的网页 A 将 PR 值平均分配给了 2 个网页 B 和 C,每个网页获得 50;网页 B 获得网页 A 和 C 的 PR 值,所以 $PR_B = 90$;网页 C 获得网页 A 的 PR 和其他网页的 PR,所以 $PR_C = 80$ 。

博用户对于其关注者的关注度之间也存在明显的差异。

2 MBUInfluence 算法

2.1 算法理论依据

在微博用户网络中,每个用户都有自己的关注列表与粉丝列表,其中关注列表表示该用户关注了哪些人,粉丝列表表示有多少粉丝关注该用户,并能够通过微博超级链接进行访问和浏览^[7-9]。基于上述分析,文中提出了基于 PageRank 思想的微博用户影响力排名算法(microblog user influence, MBUInfluence),其理论依据如下:

(1) 每个微博用户的关注数量类似于 Web 页面中的出度,粉丝数量类似于 Web 页面的入度,因此微博具有与 Web 页面相似的网络结构。

(2) 用户在社交网络中的影响力的评价,主要考虑用户之间的关注和粉丝关系,通过计算影响力值来对社交网络中的用户进行排名。因为涉及的参考因素少,所以效率和结果较好。

2.2 算法基本思想

文中对 PageRank 算法进行了改进,MBUInfluence 算法涉及到的因素包括以下两个方面:

(1) 用户自身行为权重:指的是算法中通过分析用户自身行为(包括用户自身微博总数、粉丝总数、是

否被认证、所发微博被评论数和被转发数等方面)综合得到的用户影响力指标。

(2) 粉丝影响力:指的是粉丝自身对所关注的微博用户影响力产生的贡献,粉丝自身作为微博用户,其影响力可以由下级粉丝的影响力来贡献和衡量。

如果将 PageRank 算法用于此处,需要增加用户关注度分配比例函数,对用户节点的关注度分配进行修改。定义符号 $\text{Inf}(M,N)$ 表示 M 分配给 N 的影响力。用户 A 关注了用户 B,C,D ,用户 A 将作为 B,C,D 的粉丝对其产生影响,用户 A 分配给 B 的影响力分配比例为 $\frac{\text{Inf}(A,B)}{\text{Inf}(A,B) + \text{Inf}(A,C) + \text{Inf}(A,D)}$,用户 A 分配给 C 的影响力分配比例为 $\frac{\text{Inf}(A,C)}{\text{Inf}(A,B) + \text{Inf}(A,C) + \text{Inf}(A,D)}$,用户 A 分配给 D 的影响力分配比例为 $\frac{\text{Inf}(A,D)}{\text{Inf}(A,B) + \text{Inf}(A,C) + \text{Inf}(A,D)}$ 。因为 $\text{Inf}(A,B), \text{Inf}(A,C), \text{Inf}(A,D)$ 的不同,用户 B,C 与 D 将从用户 A 得到不同的影响力,表示用户 A 对用户 B,C,D 的关注程度也不同。通过以上分析可以得出,微博粉丝对关注者的关注程度是不一样的,这通常受很多的因

素影响,粉丝也将分配不同的影响力值给不同的关注者。

2.3 算法设计

在 MBUInfluence 算法中,用户的影响力由用户自身行为权重和粉丝影响力组成,前者主要考虑粉丝数量、评论数、转发数和是否认证用户等因素,全面衡量得到微博自身的行为权重。后者主要考虑用户粉丝的影响力分配,粉丝的重视程度决定分配多少影响力权重给该用户^[10-11]。基于上述分析,定义用户 m 的影响力值 MBUInf 计算公式如下:

$$\text{MBUInf}(m) = \text{SelfInf}(m) + (1 - d_m) + d_m \sum_{n:(n,m) \in E} A_{(n,m)} \text{MBUInf}(n) \quad (2)$$

其中, $\text{SelfInf}(m)$ 表示用户 m 自身的行为权重,公式的剩余部分基于 PageRank 算法改进而来, $\text{MBUInf}(n)$ 表示用户 m 的粉丝用户 n 的影响力, $n:(n,m) \in E$ 表示 n 是 m 的粉丝之一, E 表示 m 的粉丝集合。MBUInf 算法模型中的参数描述如下:

(1) $\text{SelfInf}(m)$:表示用户 m 自身的行为权重,其可以由用户的活跃度、用户的积极度、用户的传播度三个指标来衡量^[12],具体如图 2 所示。

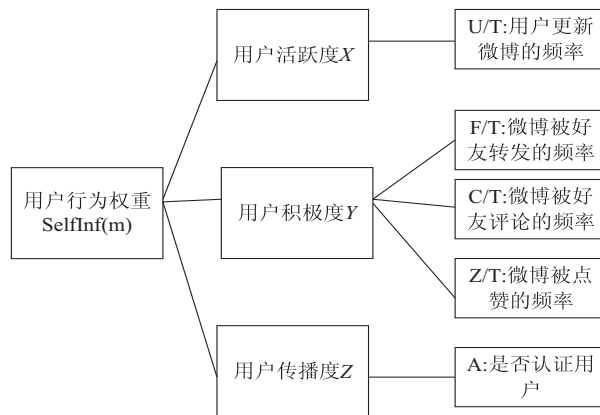


图 2 微博用户的行为权重

图中, $\text{SelfInf}(m)$ 表示用户 n 的行为权重,同时考虑用户的活跃度、积极度和传播度等三方面的信息,得到用户 m 的行为权重计算公式:

$$\text{SelfInf}(m) = aX_m + bY_m + cZ_m \quad (3)$$

将 X_m, Y_m, Z_m 根据定义得到:

$$\text{SelfInf}(m) = a \frac{U_m}{T} + b(d \frac{F_m}{T} + e \frac{C_m}{T} + f \frac{Z_m}{T}) + cA \quad (4)$$

其中, T 表示一个时间段, U_m 表示用户更新微博的数量, F_m 表示用户转发微博的数量, C_m 表示用户评论微博的数量, Z_m 表示用户点赞微博的数量, A 表示用户是否是认证用户。通过 U_m 解决了 PageRank 算法的问题 1 (微博更新不及时问题)。 a, b, c, d, e, f 表示加权系

数,文中做如下设定: $a = 0.4, b = 0.6, c = 0.1, d = 0.5, e = 0.3, f = 0.2$,然后通过上述公式可以计算 W_i 。

(2) 参数 d_m : d_m 表示用户 m 的微博消息被粉丝转发的平均概率,数学公式定义如下:

$$d_m = \frac{\sum_{n:(n,m) \in E} W_{(n,m)}}{N_m} \quad (5)$$

其中, $W_{(n,m)}$ 表示某时刻用户 n 对 m 的关注度, $\sum_{n:(n,m) \in E} W_{(n,m)}$ 表示 m 的粉丝对节点 m 的关注度之和, N_m 表示用户 m 在某时刻的粉丝总数, E 表示用户 m 的粉丝集合^[13-14]。 d_m 类似于 PageRank 中的规范化因子,表明用户对自己粉丝的影响程度。例如:如果当前许多新粉丝关注了用户 m ,那么这些新粉丝浏览和转发

用户 m 微博消息的概率就比较大,公式中的参数 d_m 值自然就会比较大。

(3) 影响力分配比例 $A_{(n,m)}$ 。

$A_{(n,m)}$ 表示用户 n 的 MBUInf 值分配给其关注者 m 的比例,通过分配比例的不同,解决了 PageRank 算法的问题 2(关注度权重平均问题), $A_{(n,m)}$ 的数学公式定义如下:

$$A_{(n,m)} = \frac{W_{(n,m)}}{\sum_{n:(n,i) \in F} W_{(n,i)}} \quad (6)$$

其中, $W_{(n,m)}$ 表示用户 n 对 m 的关注度, $\sum_{n:(n,i) \in F} W_{(n,i)}$ 表示用户 n 对其所有关注者的关注度之和, $n:(n,i) \in F$ 表示 i 是 n 的关注者之一, F 表示 n 的关注者集合。

通过上面的三个公式可以看出,关注度 $W_{(n,m)}$ 的计算是一个关键,关注度表示用户 n 对 m 的微博消息的关注程度,体现了用户 n 和 m 之间的亲密关系。文中的关注度定义由用户转发关注者的微博、评论关注者的微博以及点赞关注者的微博构成。具体如下所示:

$$W_{(n,m)} = p \frac{F_{(n,m)}}{T} + q \frac{C_{(n,m)}}{T} + r \frac{Z_{(n,m)}}{T} \quad (7)$$

用户之间的关注度 $W_{(n,m)}$ 的影响因素包含用户 n 对 m 所发微博的转发数、评论数和点赞数,用户自身的行为权重 $\text{SelfInf}(m)$ 的影响因素包括所有粉丝(包括 n)对 m 所发微博的转发数、评论数和点赞数,因此关注度 $W_{(n,m)}$ 的加权系数 p 、 q 、 r 认为与 $\text{SelfInf}(m)$ 的系数 d 、 e 、 f 相同,因此系数 $p = 0.5$, $q = 0.3$, $r = 0.2$ 。

某时刻用户 m 影响力值 MBUInf 的具体算法如下所示:

算法 1: MBUInf 微博用户影响力算法

输入: 用户 m 的元数据系列信息

输出: 用户 m 的影响力值

过程: (1) 根据式(4), 计算出用户 m 的自身行为权重 $\text{SelfInf}(m)$ 。

(2) 根据式(7), 计算出用户 m 的所有粉丝(包括 n) 对其的关注度 $W_{(n,m)}$ 。

(3) 根据式(5)和步骤(2)所得结果, 计算当前时刻, 用户 m 所有粉丝转发其微博信息的概率 d_m 。

(4) 初始化用户的影响力, 设置用户 MBUInf 的初始值为 1。

(5) 根据式(6)和步骤(2)所得结果, 计算当前时刻, 用户 m 的每个粉丝给其关注度的分配比例 $A_{(n,m)}$ 。

(6) 根据式(2), 计算用户 m 新的 MBUInf 值, 并作为下次迭代的初始 MBUInf 值。

(7) 重复步骤(6), 直到用户的影响力值收敛就停止计算, 收敛条件设置为两次迭代的 MBUInf 值的差小于极小值 ε 。

(8) 得到用户 m 的影响力值 MBUInf。

通过对 MBUInfluence 算法的设计, 将用户的影响力与用户自身行为权重和用户粉丝的影响力结合起

来, 其中后者的计算需要进行迭代, 当粉丝的迭代结果趋于收敛时, 将得到稳定的微博用户影响力值。该算法以社交网络理论为基础, 结合 PageRank 算法, 既考虑微博用户网络特征, 充分考虑了粉丝的影响力, 又考虑微博用户的属性特征, 以及微博转发率、评论率、点赞率等特征, 能够更好地反映用户的综合影响力。

3 实验及分析

3.1 实验数据

实验主要完成用户影响力算法的评价与分析。实验数据来源于新浪微博, 首先设置采集条件进行微博搜索, 然后采集发布这些微博的用户详细信息, 采集结果数据使用 json 格式进行存储。该实验用户数据采集详细条件如表 1 所示。

表 1 微博用户数据采集条件

条件	内容
微博平台	新浪微博
采集关键字	土耳其、俄罗斯、撞机
采集时间	2015. 11. 15 ~ 2015. 12. 15
粉丝数	>100
发帖数	>20
采集用户数量	80 000 个左右
数据格式	json 格式
采集用户信息	用户 ID、用户名称、所发微博总数、更新微博、评论微博、点赞微博、是否认证、关注数、粉丝数、粉丝列表等

3.2 评价标准

本小节将对 MBUInfluence 算法通过实验进行对比与分析, 对比的对象是用户影响力研究领域中常见的算法和排名方法, 它们是:

(1) FansRank 算法^[15]: 该算法主要通过比较微博用户的粉丝数量, 来衡量用户在某时刻的影响力大小。特点直观, 因此大多数学者都将它作为重要的算法来进行对比。

(2) PageRank 算法: 经典的 PageRank 算法将微博用户看作 Web 页面, 规范化因子 d 取 Google 的经验值 0.85, 计算用户在某个时刻的影响力。

(3) ForwardRank 算法^[16-17]: 该算法主要通过用户微博的总转发量, 来衡量微博用户在该时刻的影响力, 原因在于转发数表示了其他用户对于该用户所发微博的认可。

(4) 微博风云榜: 微博风云榜是一家面向社交媒体平台来做数据统计、监测和挖掘分析的网站, 提供微博和微信账号的影响力与价值排名服务。

根据 3.1 节所述, 从用户数据库中随机抽取 10 个

用户并按照粉丝数排名,计算 2015 年 12 月 15 日用户 所示。
的影响力值并进行对比。对比用户的基本信息如表 2

表 2 用户基本信息

用户 ID	用户名称	微博	转发	评论	点赞	关注	粉丝	认证
5591596171	奉命归来	1 657	76 832	972	3 762	198	33 459	false
1870236943	陕西网	67 758	121 215	4 443	2 310	1 673	30 697	false
2085811113	深夜走过长安街 001	697	109 450	22 983	10 718	457	28 907	true
5051675817	介草君	32 808	101 982	2 789	2 332	2 957	25 442	true
3602486214	方新侠 258	7 596	22 872	2 363	6 987	161	21 720	true
3624248184	时事新闻快报	13 707	432 734	22 700	9 743	291	17 553	true
1649523107	9imybaby	5 877	38 056	1 543	3 272	469	14 501	true
2447267944	军史作家罗建华	10 665	2 467	2 517	1 945	2 152	10 950	false
2139796227	中国军事天地	4 753	142 778	13 677	107 372	1 454	8 360	true
1885217285	黄昌荣_三峡西北望	62 312	57 283	2 476	6 839	1 637	5 439	false

3.3 结果与分析

下面将 MBUInfluence 算法从不同的角度与其他排名方法进行对比。

(1)同 FansRank 和 ForwardRank 算法的对比。

微博的用户影响力体现了用户节点对其他节点能够施加的影响,具体表现形式包括评论、转发或关注。其中评论动作表示粉丝对用户发布的信息认同或者否定,同用户之间进行的互动;转发动作表示粉丝对用户发布的信息认可,并且希望与自己的粉丝分享这条消

息,客观上加速了信息的传播;关注表示粉丝浏览用户的信息后,希望跟踪该用户发布的全部消息,所以粉丝会将该用户加入关注列表。例如,在微博平台上,许多人对于明星、意见领袖、知名媒体或者兴趣相同的人感兴趣,他们就会加这些用户为关注者,继而及时得到他们的相关信息。

通过 MBUInfluence 算法将得到用户的 MBUInf 值,该算法的排名结果与 FansRank 和 ForwardRank 算法的排名结果对比分析情况如表 3 所示。

表 3 MBUInf 排名同 FansRank、ForwardRank 排名对比

用户 ID	粉丝个数	FansRank	转发总次数	ForwardRank	MBUInf 排名
5591596171	53 459	1	86 832	6	5
1870236943	30 697	2	121 215	3	3
2085811113	28 907	3	119 450	4	4
5051675817	23 442	4	91 982	5	6
3602486214	21 720	5	22 872	9	9
3624248184	17 553	6	432 734	1	1
1649523107	14 501	7	38 056	8	8
2447267944	10 950	8	2 467	10	10
2139796227	8 360	9	142 778	2	2
1885217285	5 439	10	57 283	7	7

通过表 3 可以看到,MBUInf 值的排名结果与 ForwardRank 值的排名结果基本相同,但是与 FansRank 值的排名结果具有显著差异,说明用户的影响力与微博转发次数具有强联系,而与用户粉丝数具有弱联系。在相同的时间段内,MBUInf 排名仅有第 5、6 名的结果与 ForwardRank 存在差异,ID 为 5591596171 的用户其

微博被转发的次数低于用户 5051675817,但他的 MBUInf 排名比后者靠前,经过进一步分析得到,用户 5591596171 的粉丝数量比用户 5051675817 多了 2 万多名。MBUInf 算法在分配 MBUInf 值的过程中,用户将从自己的粉丝处得到被分配的 MBUInf 值,如果粉丝数量存在差异,那么得到的 MBUInf 值也会存在差

别,所以如果两个用户的粉丝数如果存在巨大悬殊,那么几乎可以肯定粉丝数高的用户其 MBUInf 值也必然大。由于用户 5591596171 的粉丝数高于用户 5051675817,那么如果他发一条微博消息,将会被更多人看到,他的信息将传播的更广,他的影响力也就相对更大。

(2)同 PageRank 算法、微博风云榜排名方法的对比。

根据式(1),用户的粉丝数表示入链,关注数表示出链,能够计算表 2 用户的 PageRank 数值,然后进行

排名,并与 MBUInf 排名进行对比。从表 4 可以看出,MBUInf 的排名结果与 PageRank 排名完全不同,说明 MBUInf 虽然是基于 PageRank 思想,但是其排名与 PageRank 排名并没有太大的相关性。

同时 MBUInf 的排名结果与用户的粉丝数排名也没有必然联系,说明 MBUInf 不完全依赖于粉丝数。MBUInf 的排名同微博风云榜的排名基本相同(除部分风云榜无法提供查询用户的排名外),说明该排名方法基本反映了微博用户的实际影响力,具体排名信息如表 4 所示。

表 4 MBUInf 排名同 PageRank 排名对比

用户 ID	粉丝个数	粉丝排名	PageRank 排名	风云榜排名	MBUInf 排名
5591596171	33 459	1	3	/	5
1870236943	30 697	2	5	3	3
2085811113	28 907	3	10	4	4
5051675817	25 442	4	6	/	6
3602486214	21 720	5	7	/	9
3624248184	17 553	6	1	1	1
1649523107	14 501	7	4	/	8
2447267944	10 950	8	2	/	10
2139796227	8 360	9	9	2	2
1885217285	5 439	10	8	/	7

(3)不同时刻的 MBUInf 值对比。

MBUInfluence 算法增加了时间因素,为了验证不同时刻粉丝因素的影响,分别选取一个月前后用户的

MBUInf 值进行对比,设定 t_1 为 2015 年 12 月 15 日, t_2 为 2015 年 11 月 15 日。实验结果如表 5 所示:

表 5 不同时刻的 MBUInf 值对比关系

用户 ID	新增粉丝数	MBUInf(t_1)	MBUInf(t_2)	新增粉丝最高的 MBUInf
3624248184	113	1	2	54 801
2139796227	122	2	1	231
1870236943	97	3	6	31 995
2085811113	89	4	4	36
5591596171	59	5	5	34
5051675817	42	6	3	345
1885217285	39	7	7	67
1649523107	50	8	8	234
3602486214	20	9	9	56
2447267944	5	10	10	87

下面分析不同时刻用户影响力的变化,主要从新增粉丝的数量、质量(通过新增粉丝最高的 MBUInf 值来体现)来比较。

首先分析用户 3624248184 和用户 2139796227,从 t_2 时刻到 t_1 时刻,第一二名的位置发生了变化,除了上

文分析过的新增粉丝数量比较接近外,用户 3624248184 有高影响力的粉丝关注他,由于该粉丝的影响力比较高,所以分给用户 3624248184 的影响力也比较高。因此导致用户 3624248184 的排名经过 1 个月超过了 2139796227 的排名。

用户 2085811113 和用户 5591596171, 其名次没有发生变化, 虽然两个用户粉丝最高 MBUInf 比较接近, 但是由于用户 2085811113 新增的粉丝较多, 所以其排名仍然领先于用户 5591596171。

综合以上分析, 可以得出如下结论: 微博用户影响力在不同时刻的变化同新增粉丝的数量以及新增粉丝的质量相关。上述结论与实际情况相符合, 在某个时间段内, 如果用户 A 得到高影响力的用户关注, 那么用户 A 的影响力也将变高, 成为“明星宠儿”; 如果用户 A 得到许多用户的关注, 即使没有高影响力的用户, 用户 A 的影响力也会变高, 成为“草根领袖”。

4 结束语

通过考虑用户之间的行为交互对 PageRank 算法进行改进, 提出一个新的 MBUInfluence 算法来对用户影响力进行分析, 并将该算法得到的 MBUInf 排名与 FansRank、ForwardRank、PageRank 等排名方法进行比较, 进一步分析它们之间的差异。

实验结果表明, 提出的算法所获得的排名相对准确和高效。根据综合分析可以得到以下结论: 用户影响力与用户粉丝的绝对数量之间不存在必然联系; 用户影响力与一段时间内新增粉丝的数量以及新增粉丝的质量存在联系, 上述两种如果突出就将改变用户的影响力排名; 用户影响力和用户与粉丝之间的交互行为密切相关, 这些交互行为包括转发、评论等, 不同交互行为对他人影响力的贡献也不一样。

参考文献:

- [1] 刘耀庭. 社交网络结构研究[D]. 杭州: 浙江大学, 2008.
- [2] BENEVENUTO F, RODRIGUESY T, CHA M. Characterizing user behavior in online social networks[C]//IMC. Chicago: ACM, 2009: 54-58.
- [3] 王晓光. 微博社区交流结构及其特征研究[D]. 上海: 华东师范大学, 2011.
- [4] 李 勇, 张克亮, 李伟刚. 基于微博的网络舆情分析系统设

- 计[J]. 计算技术与自动化, 2013, 32(2): 123-127.
- [5] 王晓彤. 基于 PageRank 的微博用户影响力度量[J]. 广东工业大学学报, 2016, 33(3): 49-54.
- [6] 王 珏, 曾剑平, 周葆华, 等. 基于聚类分析的网络论坛意见领袖发现方法[J]. 计算机工程, 2011, 37(5): 44-46.
- [7] BADASHIAN A S, STROULIA E. Measuring user influence in GitHub: the million follower fallacy[C]//Proceedings of the 3rd international workshop on crowdsourcing in software engineering. Austin: ACM, 2016: 15-21.
- [8] BASKY E, HOFMAN J M, WATTS D J, et al. Identifying “influencers” on Twitter[C]//Proc of WSDM. [s. l.]: ACM, 2011: 111-117.
- [9] CHA M, HADDADI H, BENEVENUTO F, et al. Measuring user influence in twitter: the million follower fallacy[C]//Proceedings of the fourth international conference on weblogs and social media. Washington: AAAI, 2010: 10-17.
- [10] CHANG Y, WANG X, MEI Q, et al. Towards Twitter context summarization with user influence models[C]//ACM international conference on web search and data mining. Rome: ACM, 2013: 527-536.
- [11] BAKSHY E, HOFMAN J M, MASON W A. Everyone’s an influencer: quantifying influence on Twitter[C]//Proc of WSDM. [s. l.]: ACM, 2011: 65-74.
- [12] 吴 渝, 马璐璐, 林 茂, 等. 基于用户影响力的意见领袖发现算法[J]. 小型微型计算机系统, 2015, 36(3): 561-565.
- [13] 王 顶, 徐 军, 段存玉, 等. 基于 PageRank 的用户影响力评价改进算法[J]. 哈尔滨工业大学学报, 2018, 50(5): 60-67.
- [14] 何 跃, 尹小佳, 朱 超. 基于情感及影响力的微博用户群体特征分析——以 A 手机为例[J]. 数据分析与知识发现, 2017(10): 12-20.
- [15] 陈 浩. 基于 Hadoop 的微博用户影响力排名算法研究[D]. 上海: 华东理工大学, 2014.
- [16] 刘 玲, 杨长春. 一种新的微博社区用户影响力评估算法[J]. 计算机应用与软件, 2017, 34(7): 212-216.
- [17] 张凤娟, 王 濛, 周 刚. 基于活动网络的微博用户影响力分析[J]. 计算机技术与发展, 2018, 28(9): 162-167, 171