

基于用户行为的实时影响力算法

段婷婷¹ 高仲合² 郭文健³

(曲阜师范大学信息科学与工程学院)

摘要: 随着社交网络的发展, 用户影响力研究得到广泛关注。PageRank作为分析用户影响力的传统算法, 计算用户影响力时只考虑了用户的粉丝数和关注数, 并没有考虑用户本身的行为对影响力传播造成的影响。文章综合考虑了用户行为中的转发、评论、艾特数, 论证了随着时间的变化用户行为符合幂率分布, 以此定义了兴趣度模型, 用来研究用户行为随时间的变化规律。充分将用户个体的行为融入到传统的影响力评估模型中, 定义了活跃度的概念, 可剔除不活跃的用户。此外, 摒弃了PageRank中将影响力权值平均分配的方法, 重新定义了影响力分配因子, 对于信息传播过程中影响力贡献较大的用户给予更高的分配比例, 使得影响力的评估更加准确。实验结果表明, 本文提出的基于用户行为的实时影响力算法 (User Real-time Influence), 在提高影响力计算的准确性, 更加符合社交网络的特性。

关键词: 影响力; PageRank; 兴趣度; 活跃度; 分配因子

The real-time influence algorithm based on user behavior

Duan Tingting¹ Gao Zonghe² Guo Wenjian³

(School of Information Science and Engineering, Qufu Normal University)

Abstract: With the development of social network, the research of user influence has been widely concerned. PageRank as an analysis of the user influence of the traditional algorithm to calculate the impact of users only consider the number of fans and the number of concerns, and did not consider the impact of the user's own behavior on the impact of the spread. Considering the number of forwarding and comment, AIT in user behavior, user behavior changes demonstrated with time consistent with power-law distribution, thus defines the interest degree model used to study user behavior changes over time. By fully integrating the behavior of individual users into the traditional impact assessment model, the concept of active rate is defined, which can eliminate the non active users. In addition, PageRank will abandon the influence weight average distribution, redefined the influence of allocation factors, assign a higher proportion of the influence on the information dissemination process with a large user, the influence of the assessment more accurate. The experimental results show that the proposed real-time influence algorithm based on user behavior (User Real-time Influence), in improving accuracy of influence calculation, more in line with the characteristics of social networks.

Keywords: influence; PageRank; interestingness; Active rate; Distribution factor

0 引言

随着社交网络的发展, 社交网络中用户影响力的研究已经成为一个热门领域。该研究对于了解信息的传播机制, 舆情分析, 信息的传播具有重要的意义。目前对于影响力的评估方法主要有: 基于传统算法PageRank的评估方法, 该算法只考虑粉丝数量, 并没有考虑用户个体在产生影响力时的差异性; 基于用户行为的评估方法, 该方法只是简单的评估了用户的某种行为的影响, 并没有进行综合分析用户行为; 基于URL追踪的评估方法, 在多级联网页追踪中难度较大。

本文以新浪微博为背景, 提出基于用户行为的实时影响力算法, 解决了以下三个方面的问题: 定义了兴趣度的概念, 可以实时的评估粉丝对新消息的兴趣度, 能够有效识别僵尸粉。以活跃度替代传统算法中的阻尼系数, 能够更加准确的计算出粉丝对关注者的信息进行传播的概率。重新定义了权值

分配比例, 用影响力分配因子表示, 给予影响力传播贡献更大的用户更高的权值。避免了平均分配的不公平现象。

1 PageRank算法

PageRank算法^[2]是计算用户影响力中普遍应用的算法, 将社交网络看成是一张有向权值图 $G=(N, E, W)$, N 表示用户节点, 图中的边 E 表示的是用户之间的关系, 粉丝与关注者之间的兴趣度是图中的权值 W , 微博中的好友关系是单向的类似于PageRank中的链入链出关系, 粉丝表示链入关系, 关注者表示链出关系。因此用来计算网页之间的影响的PageRank算法便可以用来计算社交网络用户的影响力。如公式1-1所示:

$$PR(i) = (1-d) + d \sum_{(i,j) \in E} \frac{PR(j)}{O(j)} \quad (1-1)$$

$PR(i)$ 是网页 i 的PageRank值, d 表示阻尼系数,

(1-d)表示的是用户继续浏览页面的概率， $O(j)$ 表示网页j所有链出的个数。

2 基于用户行为的实时影响力算法

本文综合考虑用户的转发、评论、艾特的概率（以下统称这些行为数量为关系数），提出了兴趣和活跃度的概念，并对用户的传播贡献能力进行量化，添加了影响力分配因子，接下来详细叙述基于用户行为的实时影响力算法。

2.1 算法设计

2.1.1 兴趣度 $W(\tau)$

关注者和粉丝之间的关系程度，可以用兴趣度来表示，具体表现在粉丝对于关注者的消息存在的转发、评论、艾特的概率。如公式2-1所示：

$$W_{(j,i)}(\tau) = \frac{\sum R_{\tau} \cup C_{\tau} \cup M_{\tau}}{N} \quad (2-1)$$

其中 $W_{(j,i)}(\lambda)$ 表示用户在第 τ 周内粉丝的关系数与关注者所产生的总的评论、转发、艾特数量之比。这一指标能够看出粉丝对关注者的兴趣度。数据是利用关注者和粉丝在社交网络信息传播过程中产生的数据得来的。利用公式2-1计算多个关注者和粉丝之间的兴趣度随着时间变化的分布，经拟合这个概率分布可表达为： $W(\tau) = a \times \tau^{-a} + b$

2.1.2 活跃度 $\theta(t)$

关注者发布一条新消息，并不是所有粉丝都会对信息的传播做出贡献。 $\theta(t)$ 表示粉丝对新消息继续传播的概率。如公式2-2所示：

$$\theta(t) = \frac{\sum_{(i,j) \in E} W_{(i,j)}(t)}{N_E} \quad (2-2)$$

$W_{(j,i)}(t)$ 表示t时刻(i, j)之间的兴趣度。 N_E 为t时刻，关注者和粉丝之间所有的关系数。这个参数表明了该用户的活跃程度。可鉴别僵尸粉，保障有效粉丝的数量。

2.1.3 影响力分配因子计算

影响力分配因子能够衡量粉丝对影响力的传播贡献。在传统的PageRank算法中，考虑了网页链入链出的关系，对应到社交网络中也就是只考虑粉丝数和关注数，并没有考虑用户本身的行为对影响力传播的影响。采用平均分配的方式将影响力分配给粉丝。这一算法计算出的用户影响力并不准确。许多活跃的用户传播贡献非常大被无贡献的粉丝分走了影响力。因此本文在传统的PageRank模型中，将用户转发、评论、艾特概率重新进行建模， $I_{(j,i)}(t)$ 表示用户j分配给i的影响分配比。如公式2-3所示：

$$I_{(j,i)}(t) = \frac{W_{(j,i)}(t)}{\sum_{k:(j,k) \in E} W_{(j,k)}(t)} \quad (2-3)$$

2.2.4 基于用户行为的实时影响力算法URIRank将PageRank重新定义和建模之后模型如下：

$$URIR(i,t) = (1 - \theta(t)) + \theta(t) \times \sum_{k:(j,k) \in E} I_{(j,k)}(t) \times URIR(j,t) \quad (2-4)$$

$URIR(i,t)$ 表示t时刻，用户i的URIR值， $\theta(t)$ 表示粉丝对新消息继续传播的概率， $I_{(j,i)}(t)$ 表示用户j分配给i的影响分配比。

3 实验结果与分析

3.1 数据集

在本实验中，采用新浪微博API采集了500名用户信息，包括用户的粉丝数，微博数，转发评论艾特数。将这500名用户按照序号分成了5组分别用PageRank算法和URIR算法对500名用户进行影响力排名，最终选取了排名前20中具有典型意义的六名用户进行对比分析。

3.2 URIR的收敛性

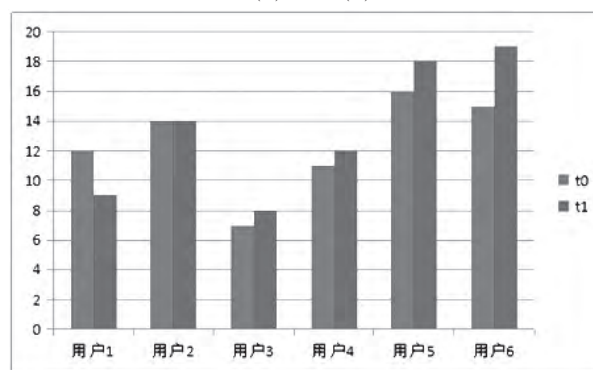
取初始时刻的数据为例，使用活跃度计算公式计算出初始时刻的 $\theta(t) = 0.013$ ；根据公式2-4计算URIR值，经过追踪发现算法在第五次的时候就显示为收敛状态。

3.2 URIR与PageRank的对比实验

表3-1 t1时刻 URIR算法和PageRank算法排名表

用户编号	URIR 排名	PageRank 排名	关系数变化 (t_0, t_1)
1	9	14	3412
2	14	9	87
3	8	11	2810
4	12	8	3
5	18	20	5138
6	19	18	17

图3-1 URIR(t_0)与URIR(t_1)排名对比



从表格1可以看出用户在(t_0, t_1)这段时间内关系数的变化，能够最直接的体现用户影响力的变化。用户2的关系数变化不大，因此在 t_1 时刻，URIR的排名降低，而PageRank排名却上升了，这是因为在这段时间内，用户2增加了许多的非活跃粉丝，也就是僵尸粉，这部分粉丝质量不高几乎没有对用户2的信息传播产生贡献，可以看出URIR算法的排名更加精确符合实际的社交网络实情。用户5在这个

时间段中产生了5138条关系数的变化,在PageRank算法下的排名依然降低了,该算法只考虑了少数粉丝数量的减少,却没有将大量优质粉丝产生的巨大的影响力考虑在内,URIRank算法真正挖掘出了具有影响力的用户。在图1中可看出URIRank算法可计算影响力随时间的变化,符合实时性。以上的图表表明URIRank算法比PageRank算法,更符合社交网络的特性,对影响力的评估更加精确。

4 结束语

本文将传统影响力评估算法PageRank与用户行为综合分析相结合,对传统模型进行重新建模,提出了兴趣度,活跃度以及影响力分配因子,降低了无效粉丝的影响力排名,且对贡献较大的用户给予了更高的影响力分配比值。通过对比实验验证了本算法在影响力的评估中,更适合社交网络的特点,取得的影响力排名具有实时性且更准确。

参考文献:

- [1] 吴柯. 基于交互行为的微博用户影响力评价研究[D]. 华南理工大学, 2014.
- [2] 黎明, 文海英, 杨杰, 陈旭日. 基于行为权值的微博用户影响力度量算法[J]. 计算机工程与应用, 2014, 17: 130-133.
- [3] 胡建华. 微博用户行为与影响力分析系统的研究[D]. 北京交通大学, 2013.
- [4] 康书龙. 基于用户行为及关系的社交网络节点影响力评价[D]. 北京邮电大学, 2011.
- [5] Shaozhi Ye, S. Felix Wu. Measuring message propagation and social influence on Twitter.com [M]//Social Informatics. Springer Berlin Heidelberg, 2010: 216-231.

作者简介:

段婷婷, 曲阜师范大学信息科学与工程学院,
硕士
电话: 13280086553
电子邮箱: tt_duan@126.com
联系地址: 山东省济宁市曲阜市曲阜师范大学
(273165)
高仲合, 曲阜师范大学, 教授, 硕士生导师
郭文健, 曲阜师范大学信息科学与工程学院,
硕士

(上接10页)

- [3] 彭金华. 反馈放大电路中反馈类型的分析与判断[J]. 宁波大学学报(理工版), 2005, 18(1): 87-89.
- [4] 李来运, 田永亮. 对负反馈电路判断的探讨[J]. 洛阳师范学院学报, 2006, 25(2): 48-50.
- [5] 张梅. 《模拟电子技术》课程中反馈概念及类型判断教法探讨[J]. 科技信息, 2009(22): 460, 463.
- [6] 杜玲云, 唐明涛. 放大电路中反馈类型的判别方法[J]. 信息系统工程, 2014(10): 85.
- [7] 汪涛. 基于放大电路中负反馈类型判定方法的研究[J]. 2015(2): 37-40.

(上接19页)

参考文献

- [1] 于治楼, 姜凯, 李峰基. 于FPGA的SOC验证平台的设计[J]. 信息技术与信息化, 2008年5期: 92-94.
- [2] 王阳赞. ARM CoreSight跟踪调试技术的研究与应用[D]. 武汉: 武汉理工大学, 2009年6月.
- [3] 施乐宁, 董金明. 基于ARM7TDMI的SOC的FPGA验证平台的设计[J]. 现代电子技术, 2007年10期: 72-73.
- [4] A.B.T.Hopkins, K.D.McDonald-Maier. Debug support strategy for systems-on-chips with multiple processor cores[J]. IEEE Transactions on Computers. vol.55, pp.174~184, 2006.
- [5] 虞致国, 魏敬和. 基于FPGA的ARM SOC原型验证平台设计[J]. 电子与封装, 2007年5月第49期: 25-28.

通信作者:

苗瑾超、新疆大学科学技术学院(阿克苏校区)、新疆维吾尔自治区阿克苏地区新疆大学科学技术学院(阿克苏校区)、843000、18709910867、282010436@qq.com

作者简介: 苗瑾超、1987、女、硕士研究生.