



一种改进的微博用户影响力评估算法

黄贤英, 阳安志, 刘小洋, 刘广峰

(重庆理工大学 计算机科学与工程学院, 重庆 400054)

摘 要: 在已有 PageRank 算法构建的微博用户影响力评估模型中, 存在用户自身属性信息欠缺以及在用户不活跃期间其影响力被误判下降的问题。为此, 综合考虑用户自身的属性, 基于用户的活跃度、认证信息及博文质量来确定其自身的基本影响力, 通过引入用户博文的传播率挖掘用户的潜在影响力, 结合用户不同好友的质量, 基于改进的 PageRank 算法构建微博用户影响力评估算法。实验结果表明, 与改进 BWPR 算法相比, 该算法准确率、召回率和 F 值分别提高 13.5%、10.1% 和 12.3%, 能准确、客观地反映微博用户的实际影响力, 可为社交网络中的意见领袖挖掘、信息传播和舆论引导等研究提供参考。

关键词: 微博; 社交网络; 影响力; PageRank 算法; 传播率

开放科学(资源服务)标志码(OSID):



中文引用格式: 黄贤英, 阳安志, 刘小洋, 等. 一种改进的微博用户影响力评估算法[J]. 计算机工程, 2019, 45(12): 294-299.

英文引用格式: HUANG Xianying, YANG Anzhi, LIU Xiaoyang, et al. An improved algorithm for microblog user influence evaluation[J]. Computer Engineering, 2019, 45(12): 294-299.

An Improved Algorithm for Microblog User Influence Evaluation

HUANG Xianying, YANG Anzhi, LIU Xiaoyang, LIU Guangfeng

(College of Computer Science and Engineering, Chongqing University of Technology, Chongqing 400054, China)

[Abstract] The current PageRank algorithm based microblog user influence evaluation model has many problems, such as the lack of users' attributes and the incorrect evaluation of users' influence during their inactive period. In this paper, the attributes of users are comprehensively considered, and their basic influence is confirmed based on users' activeness, authentication information and blog quality. Then, according to the blog dissemination rate, this paper further excavates the potential influence of users. Ultimately, with the evaluation of the quality of different microblog friends, a microblog user influence evaluation algorithm is established on the basis of improved PageRank algorithm. Experimental results show that compared with BWPR improved algorithm, the accuracy, recall rate and F value of the proposed method are increased by 13.5%, 10.1% and 12.3% respectively. It can reflect the actual influence of users in a more accurate and objective way, which can provide references to the researches on opinion leader identification, information dissemination and the guidance of public opinions in social networks.

[Key words] microblog; social network; influence; PageRank algorithm; dissemination rate

DOI:10.19678/j.issn.1000-3428.0053029

0 概述

随着 Web 技术和移动互联网的广泛应用, 在线社交网络得到快速的发展, 其在推动信息传播方面

起到了重要的引导作用^[1]。新浪微博是国内最为活跃的社交网络之一, 《第 40 次中国互联网络发展状况统计报告》和《2017 微博用户发展报告》表明, 截至 2017 年 9 月, 新浪微博月活跃人数达 3.76 亿, 在

基金项目: 国家社会科学基金(17XXW004); 教育部人文社科青年基金(16YJC860010); 重庆市教育委员会人文社会科学研究项目(17SKG144, 18SKGH100); 2018 年重庆市科委技术创新与应用示范项目(cstc2018jscx-msybX0049); 2017 年度重庆市高校网络舆情与思想动态研究咨政中心开放课题(KFJJ2017024)。

作者简介: 黄贤英(1967—), 女, 教授, 主研方向为社交网络、传播模型; 阳安志(通信作者), 硕士研究生; 刘小洋, 副教授、博士后; 刘广峰, 硕士研究生。

收稿日期: 2018-10-30

修回日期: 2018-12-21

E-mail: pureyangcry@foxmail.com

国内微博社区处于领先地位。微博社交网络是社会关系维系和信息传播的重要渠道,对国家安全与社会发展都会产生深远的影响^[2]。

在社交网络中,部分用户节点对其他用户节点有极大的影响力,被称为“意见领袖”,他们针对舆论等相关事件发表言论,与网民、媒体之间形成互动,其观点往往影响着大批粉丝和舆论走向。如何有效地评估微博社交网络中用户的影响力,挖掘“意见领袖”,对于信息传播、舆论引导、广告推荐、网络营销等方面都有极其重要的作用,引起了众多领域研究人员的兴趣^[3-5]。

目前研究利用链接关系分析微博用户影响力的方法较多,即基于用户间的交互关系,通过改进 PageRank 算法建立具有影响力的评估方法^[6-7]。改进的方法主要包括基于用户节点的属性、基于用户的交互行为和结合用户的综合信息等。基于用户节点的属性主要是基于用户粉丝数、博文数、是否认证等信息,如文献[8-9]基于用户自身的信息确定不同的基本分数来评估影响力。文献[10]结合用户的粉丝数和是否认证等信息来评估用户的影响力。基于交互行为方法是结合用户博文的转发、评论关系网络,如文献[11]基于微博用户交互关系来挖掘意见领袖。以上2种用户影响力评估方法仅考虑单方面的因素,但有潜在误差过大、评估结果不准确等问题。此外,微博中有大量假冒的粉丝,作为“僵尸粉”存在,所以结合用户的粉丝数会影响评估结果准确性^[12-13]。为了更加合理地评估微博用户的影响力,研究人员开始结合用户自身属性和交互行为建立综合模型,如文献[14-15]基于用户的属性,通过博文的转发、评论和提及的交互关系信息来评估用户的影响力,文献[16]提出了基于用户节点的行为和交互网络的拓扑结构来挖掘 Twitter 上的意见领袖用户。

综合用户的自身属性和交互行为,通过改进 PageRank 算法能比较客观、合理地评估微博用户的影响力,但是基于用户交互关系,已有改进的 PageRank 算法多数为用户设置相同的初始影响力分数,利用用户的自身属性进行加权,通过迭代计算来修正不同用户的影响力分数,进而挖掘出意见领袖^[17]。若某个意见领袖用户一段时间未发表博文,或未参与交互,此类方法计算出的用户影响力会很低,而用户实际的影响力不一定急剧下降,忽视了影响力较大的用户的潜在传播力。

本文综合考虑用户自身的属性,在改进 PageRank 算法的基础上提出一种微博用户影响力评估算法。通过确定用户自身的基本影响力,同时引入用户博文的传播率并结合用户的交互关系,以提升用户的实际影响力。

1 相关研究

1.1 PageRank 算法

用户在社交网络中的交互可以被视为有向链接,常用的基于链接的排序算法之一是 PageRank 页面排序算法,该算法由 Google 的两位创始人提出,最初是为了实现网页排名,其在搜索引擎中被广泛使用。页面的分数通过不断迭代计算,直至收敛到一个固定值,但当某些页面只存在入链或出链时,迭代结果会出现“排名泄漏”和“排名下沉”,得到不合理的排名结果。为了解决这个问题,引入了随机浏览模型,即每个页面都可以随机地访问其他页面。PageRank 算法计算公式如式(1)所示。

$$PageRank(x) = (1-d) + d \sum_{y \in L(x)} \frac{PageRank(y)}{N_y} \quad (1)$$

其中, $PageRank(x)$ 、 $PageRank(y)$ 分别表示页面 x 和 y 的排名分数, $L(x)$ 表示页面 x 的链入页面集合, N_y 表示页面 y 总的链出数, d 是阻尼系数,表示一个页面被其他页面随机访问的概率,即使没有被其他页面引用,页面也能获得基本分数,从而保证页面的迭代分数收敛。在计算时,页面的初始基本分数为 $1-d$,所有页面的基本分数一致。

PageRank 算法本质是对有向图的节点进行排序,所以它也能被应用到其他领域。微博社交网络中用户与其他用户的关注、交互关系可以视为链接,通过迭代计算可以得到用户的影响力分数。在网页排名中,从页面获得的分数与该页面输出链接的数量成反比,因为页面排名的目的是评估页面被访问的可能性,这不符合社交网络用户影响力的评估机制,所以很多研究都对 PageRank 算法进行了改进。

1.2 BWPR 算法

为分析本文提出算法的效果,将 BWPR 算法^[15]作为对比算法之一。该算法与其他结合用户自身属性信息和交互关系的微博用户影响力评估模型类似,利用用户的自身属性确定加权因子,基于用户博文的转发、评论和提及的交互关系网络,通过改进 PageRank 算法来迭代计算用户的影响力分数,进而挖掘出意见领袖。该算法具体计算公式如式(2)所示。

$$BWPR(u_i) = d + (1-d) \sum_{v_j \in F_o(u_i)} B_w(u_i, v_j) \cdot BWPR(v_j) \quad (2)$$

其中, u_i 、 v_j 表示微博用户, $F_o(u_i)$ 是用户的粉丝集合, d 表示用户的基本分数, $B_w(u_i, v_j)$ 是加权因子。与原始的 PageRank 算法相比, BWPR 算法将所有用户的基本分数初始化为 d ,再结合用户的交互关系,通过用户自身属性确定的加权因子 $B_w(u_i, v_j)$ 进行迭代计算得到用户的影响力分数。

微博用户发表博文有一定的时间间隔,若用户一段时间不发博文,不参与交互,则通过类似 BWPR

的改进算法得到的用户影响力将会很低,而在实际情况中,用户的影响力不一定会急剧下降。

2 用户影响力评估算法

2.1 用户自身影响力

PageRank 算法认为所有用户是绝对平等的,这与社交网络中用户的实际情况不符。结合前人的工作,本文对微博中影响力较大的意见领袖进行分析,从而确定用户自身的基本影响力。

1) 活跃度

影响力较大的用户对微博事件的关注和参与度较高,他们积极地发表言论,对博文传播有极大的推进作用。用户的活跃度可以通过用户一段时间发表博文的数量来度量,计算公式如式(3)所示。

$$u_{\text{active}}(x) = \frac{\sum_{t \in T} N_t}{T} \quad (3)$$

其中, $u_{\text{active}}(x)$ 表示用户 x 的活跃度, $\sum_{t \in T} N_t$ 表示用户 x 在时间段 T 内所发表的博文数量。

2) 认证信息

通过官方认证的用户会成为加 V 用户,更易受到其他用户的关注与认可,他们对某个事件的评价和看法会被较多的用户查看,用户整体的知名度会上升,认证用户一般具有较大的影响力。计算公式如式(4)所示。

$$V(x) = \begin{cases} 1, & x \text{ 是认证用户} \\ 0, & x \text{ 不是认证用户} \end{cases} \quad (4)$$

其中, $V(x)$ 表示用户 x 的认证信息。

3) 博文质量

随着微博的流行,用户的规模急剧增大,影响力较大的意见领袖的博文受到极大的关注。意见领袖用户通常有大量的粉丝,先前的一些研究将粉丝数作为用户影响力评估的一个重要因素,但微博中存在虚假粉丝,文献[4,14]的研究表明,用户的影响力和粉丝数没有直接关系。

对微博社交网络中意见领袖用户的特点进行分析后,发现微博用户博文的质量价值越大,博文被转发、评论的可能性就越高,用户对博文互动的次数越多,博文的传播也就越广。于是,本文利用用户博文的转发、评论和点赞信息来确定博文的质量,计算公式如式(5)所示。

$$Q_{\text{blog}}(x) = \alpha_1 \cdot \text{avg}_{\text{retweet}} + \alpha_2 \cdot \text{avg}_{\text{comments}} + \alpha_3 \cdot \text{avg}_{\text{like}} \quad (5)$$

$$\alpha_1 + \alpha_2 + \alpha_3 = 1$$

其中, $Q_{\text{blog}}(x)$ 表示用户 x 的博文质量, $\text{avg}_{\text{retweet}}$ 、 $\text{avg}_{\text{comments}}$ 和 avg_{like} 表示在时间 T 内,用户博文的平均转发数、评论数和点赞数, α_1 、 α_2 和 α_3 是各部分的加权重。

因为用户不同自身属性信息的特征数值区间相差非常大,不能直接线性相加,于是对这些信息进行数值归一化,如式(6)所示。

$$F_{\text{norm}} = \frac{F - F_{\min}}{F_{\max} - F_{\min}} \quad (6)$$

其中, F_{norm} 表示归一化处理后的特征值, F 表示该特征的原数值, F_{\min} 表示所有用户该特征的最小值, F_{\max} 表示所有用户该特征的最大值。

本文对微博用户自身特征信息进行归一化处理,定义了用户自身影响力分数,计算公式如式(7)所示。

$$\text{Base}(x) = u_1 \cdot U_{\text{active}}(x) + u_2 \cdot V(x) + u_3 \cdot Q_{\text{blog}}(x) \quad (7)$$

$$u_1 + u_2 + u_3 = 1$$

$\text{Base}(x)$ 是用户 x 的自身影响力的基本分数,是对用户的活跃度、认证信息、博文质量进行线性加权确定的。

2.2 用户博文的传播率

在微博社交网络中,用户的影响力主要是通过博文的转发、交互进行传播体现的^[4],为修正先前的研究在评估用户影响力时,在用户不活跃期间,其影响力被误判下降的问题,引入用户博文的传播率,挖掘用户的潜在传播力,计算公式如式(8)所示。

$$\text{spread}(x) = \sum_{\tau \in M_t} \frac{|\varphi_{\tau}|}{|\varphi_v|} \frac{\sum_{\gamma \in \varphi_{\tau}} |\text{reweets}_{\gamma}|}{\sum_{\gamma \in \varphi_{\tau}} |\text{interactive}_{\gamma}|} \quad (8)$$

其中, $\text{spread}(x)$ 表示用户 x 博文的传播率, M_t 表示用户 x 在时间 t 内参与的博文交互集合, φ_{τ} 表示用户在时间 t 内转发的博文集合, φ_v 表示用户在时间 t 内所有博文(原创/转发)集合, reweets_{γ} 表示博文 γ 的转发数, $\text{interactive}_{\gamma}$ 表示博文 γ 的交互(转发/评论/点赞)数。

用户博文的传播率综合考虑了用户的活跃度和交互等信息,高影响力用户的传播率远大于一般用户,他们发表的博文更易受到转发和评论,传播更广。引入传播率,减少了传统改进 PageRank 的算法通过用户自身属性(结合粉丝数)来确定加权因子进行计算用户影响力带来的误差,即使意见领袖用户一段时间不参与交互,其潜在影响力值也不会消失。相比于 BWPR 改进算法,引入博文的传播率后,能更好地识别有影响力的用户。

2.3 UserRank 算法

本文通过改进 PageRank 算法,结合用户的自身影响力,引入博文的传播率后重新构建了 UserRank 用户影响力评估算法,计算公式如式(9)所示。

$$\text{UserRank}(x) = (1 - d) \cdot \text{Base}(x) + d \cdot \sum_{y \in B(x)} \text{spread}(y) \cdot \text{UserRank}(y) \quad (9)$$

其中, $\text{UserRank}(x)$ 、 $\text{UserRank}(y)$ 表示用户 x 和 y 的影响力值, $\text{Base}(x)$ 表示用户 x 的自身影响力值, $\text{spread}(y)$ 表示用户 y 的博文传播率, $B(x)$ 表示关注用户 x 的好友集合, d 表示加权重。

原始的 PageRank 算法和 BWPR 算法^[15]在计算时,每个用户的初始分数值相同,即认为每个用户同等重要,每个粉丝都将自己的分数平均分配给其关注用户,该方法误差较大。而本文的 UserRank 算

法,一方面综合考虑了不同用户的自身信息,为其确定不同的基本影响力值,另一方面区别用户不同的好友对其影响力提升的贡献大小,并引入博文传播率,避免用户影响力在不活跃期间被误判下降,能更加合理地评估微博用户的实际影响力。

3 实验结果与分析

3.1 数据获取与处理

以新浪微博的数据进行实验,数据主要来源于自然语言处理和信息搜索平台(www.nlpir.org)发布的 NLPir 微博博主语料库^[8],该数据集包括 100 万用户的详细信息及交互信息,进行分析筛选后,选取了其中的 13 万个用户,对这些数据分析处理、重新采集后,最终得到的用户信息有:

- 1) 用户信息主要为用户 ID、用户名、性别、认证信息、微博总数、关注数、粉丝数。
- 2) 最近博文信息主要为博文的内容、点赞数、评论数、转发数。
- 3) 关注信息主要为关注列表、交互关系列表。用户关注关系网络的相关信息如表 1 所示。

表 1 用户关系网络指标

节点数	边数	平均度	平均路径长度	平均聚类系数
130 000	589 137	3.781	2.503	0.019

首先对用户间的交互关系网络进行分析,结果如图 1 所示。

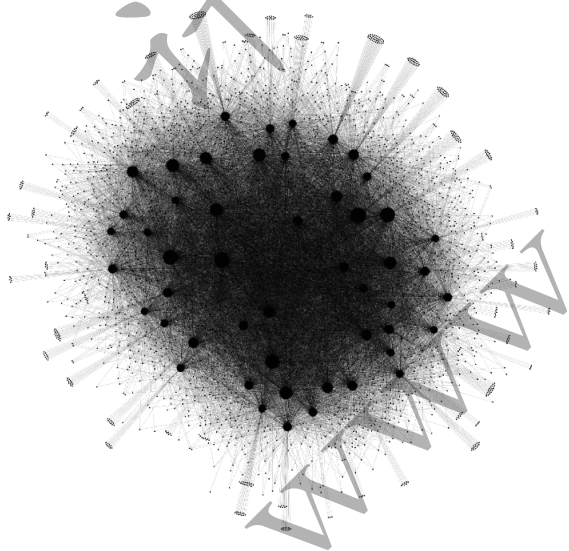


图 1 用户关注关系

图 1 是被关注排名前 50 个用户的交互关系,图中的每个节点代表一个用户,每 2 个用户的交互关系形成一条边,用户节点参与的交互次数越多,图中其节点就越大。实验结果表明,用户的影响力在博文的转发、交互中体现,相比于一般用户,影响力较大的意见领袖用户的博文被大量用户转发,转发其博文的用户越多,该用户的影响力就越大。

3.2 结果分析

为验证 UserRank 算法的有效性,本文采用以下 3 种评估算法进行对比分析:

1) 按粉丝数排名。按照微博用户粉丝数的多少来对用户影响力排名。

2) PageRank 排名。PageRank 算法是利用用户的交互关系来对用户排名。

3) BWPR 算法排名。基于用户博文的转发、评论和提及网络对用户排名。

参考文献[4]定义的算法的准确率、召回率和 F 值来评估各个算法的效果,计算公式如式(10)所示。

$$S_2 = (S_{\text{UserRank}} \cap S_{\text{PageRank}}) \cup (S_{\text{UserRank}} \cap S_{\text{BWPR}}) \cup (S_{\text{UserRank}} \cap S_{\text{FansRank}}) \cup (S_{\text{PageRank}} \cap S_{\text{BWPR}}) \cup (S_{\text{PageRank}} \cap S_{\text{FansRank}}) \cup (S_{\text{BWPR}} \cap S_{\text{FansRank}}) \quad (10)$$

其中, S_{UserRank} 表示本文的 UserRank 算法用户排名集合, S_{PageRank} 表示 PageRank 算法的用户排名集合, S_{BWPR} 表示 BWPR 算法的用户集合, S_{FansRank} 表示按粉丝数排名的用户集合, S_2 表示各个算法总的用户排名集合。

UserRank 算法的准确率定义如式(11)所示。

$$\text{Precision}_{\text{UserRank}} = \frac{|S_{\text{UserRank}} \cap S_2|}{|S_{\text{UserRank}}|} \quad (11)$$

其中, $\text{Precision}_{\text{UserRank}}$ 表示 UserRank 算法的准确率,其他对比算法的准确率定义类似。

UserRank 算法的召回率定义如式(12)所示。

$$\text{Recall}_{\text{UserRank}} = \frac{|S_{\text{UserRank}} \cap S_2|}{|S_2|} \quad (12)$$

其中, $\text{Recall}_{\text{UserRank}}$ 表示 UserRank 算法的召回率,其他对比算法的召回率定义类似。

综合指标 F 值定义如式(13)所示。

$$F\text{-measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (13)$$

其中, $F\text{-measure}$ 表示准确率和召回率的调和平均值,其值越大,表明该方法的效果越好。

各个算法的准确率对比结果如图 2 所示。

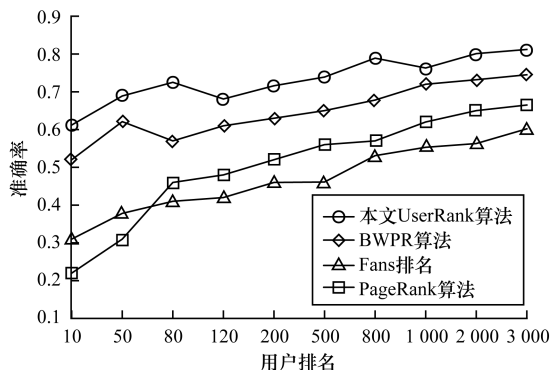


图 2 不同算法的准确率对比结果

图 2 是各个算法将用户在不同的 topN 排名下其准确率结果。可以看出,随着统计排名人数的增加,各个算法的准确率都在提高,但本文的 UserRank 算法明显优于其他对比算法,BWPR 算法次之;粉丝数排名算法在 top10 的准确率高过 PageRank 算法,但随着统计人数的增多,按用户的粉丝数排名的准确率最差。

各个算法召回率对比的实验结果如图 3 所示。

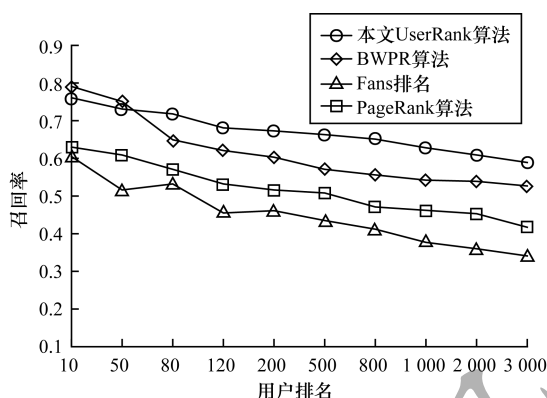


图 3 不同算法的召回率对比结果

图 3 是各个算法在不同 topN 排名下的召回率结果。用户的排名人数在 top10 时,BWPR 算法的召回率是最高的,但是在 top50 后,BWPR 算法的召回率结果低于本文的 UserRank 算法,整体看来,本文的 UserRank 算法召回率效果相对较好。

F 值是准确率和召回率的综合指标,各个算法的实验结果如图 4 所示。

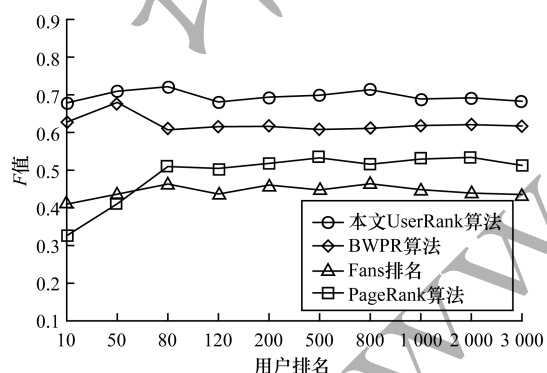


图 4 不同算法的 F 值对比结果

从图 4 可以看出,UserRank 综合指标是最好的,接着是 BWPR 算法和 PageRank 算法,最差的是按粉丝数排名的结果。这表明微博用户的影响力与用户的自身传播力和博文的质量密切相关,本文的 UserRank 算法在评估微博用户的影响力以及挖掘意见领袖用户有更好的效果。此外,单一地利用用户粉丝数量作为用户影响力大小的评估指标,实际效果较差。

表 2 列出了用户按粉丝数排名、BWPR 算法和本文的 UserRank 算法中用户影响力的前 10 名用户。

表 2 不同算法的用户影响力 top10

用户排名	UserRank	BWPR	粉丝数排名
1	人民日报	谢娜	微博小秘书
2	央视新闻	赵丽颖	微博客服
3	人民网	人民日报	新手指南
4	头条新闻	何炅	微博管理员
5	新京报	李开复	谢娜
6	谢娜	央视新闻	何炅
7	新浪新闻	杨幂	赵薇
8	赵丽颖	薛之谦	人民日报
9	何炅	头条新闻	央视新闻
10	新浪娱乐	angelababy	李开复

从表 2 可以看出,这些算法挖掘出的意见领袖用户都偏向于综艺、娱乐、明星和新闻等领域,这些领域的用户影响力较高,说明很多微博用户倾向于关注娱乐、明星的生活动态以及时事新闻,这些意见领袖用户在微博中的信息传播、舆论导向起着重要的作用。从粉丝数排名结果可以看出,微博小秘书、微博客服等微博官方博客主要由于粉丝数众多(均超过 2 亿),排名靠前,但实际中这些用户的活跃度较低,用户与之交互较少,故没有出现在 BWPR 算法和本文的 UserRank 算法的前列;因为 BWPR 算法更加关注用户之间的交互,所以挖掘出用户都有频繁的博文交互关系,如主持人谢娜、演员赵丽颖,因其粉丝基数较大,所以博文的转发、评论和交互较多,故在 BWPR 算法前列,但如果一段时间用户不参与交互,这些用户影响力会被误判而急剧下降;本文的 UserRank 算法引入博文的传播率后,能挖掘出一些潜在影响力较大的用户,如人民日报、头条新闻、新浪新闻等用户,这些用户的博文质量和活跃度较高,相比与综艺、明星用户,他们粉丝的组成是不同的。当前有很多明星用户,为了快速提升自己影响力,会花钱购买用户来增加自己的粉丝数,即所谓的“僵尸粉”,此外,他们的粉丝所在兴趣领域相对较狭窄。人民日报、头条新闻作为官方媒体,其权威性较高,发布的话题具有社会性,其粉丝所在领域是多元的,故其博文的传播涉及面较广,潜在传播影响力远大于一般明星用户。

社交网络中的意见领袖能够影响大量的用户,基于这个特点,引入用户影响力的覆盖率作为一个评估指标。用户影响力覆盖率是指在网络中,有影响力的用户节点直接或间接影响其他用户节点的覆盖范围的比率,计算公式如式(14)所示。

$$I(k) = \frac{\sum_{k=1}^N P(k)}{N} \quad (14)$$

其中, $I(k)$ 为前 k 个用户的影响力覆盖率, N 表示该网络中的所有用户, $P(k)$ 表示排名前 k 个用户影响的其他用户节点的个数。

实验统计了各个算法挖掘出的前 1 000 名用户的博文在转发、评论和交互中所涉及的其他用户,作为被影响力的用户,实验结果如图 5 所示。

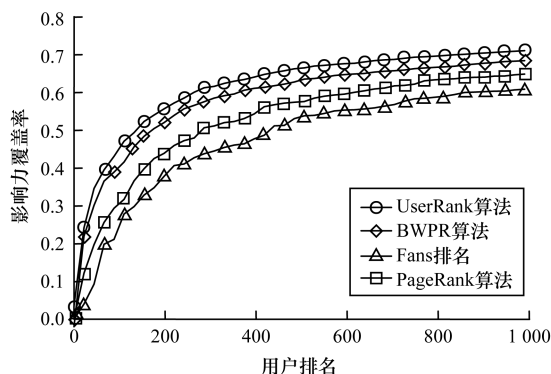


图5 不同算法的影响力覆盖率对比

从图5可以看出,在13万个用户节点构成的网络中,本文的UserRank算法挖掘出的前1000个用户的影响力具有最好的覆盖率,达到了70%,与BWPR算法相比在一定程度上更有优势。按粉丝数排名的用户影响力覆盖率最低,表明在该网络中粉丝数较多的用户节点,粉丝在该用户博文的转发、交互中参与率较低,活跃度不高,用户的实际影响力较低。

4 结束语

用户影响力的评估作为社交网络数据挖掘的一个研究热点,对信息传播、舆论导向和相关推荐等领域都有积极的作用。已有的改进PageRank算法在评估微博用户影响力研究中,存在用户在不活跃期间其影响力被误判下降的问题。本文分析微博意见领袖用户的特点,考虑用户的自身影响力和好友质量,并引入用户的博文传播率,基于用户的交互关系信息,通过改进PageRank算法构建一个新的微博用户影响力评估算法。在新浪微博数据集上的实验结果表明,该算法能较好地评估微博用户的实际影响力,挖掘出意见领袖用户。下一步将尝试挖掘用户更丰富的自身信息,如不同年龄段、不同职业人群的信息并且基于不同主题来评估微博用户影响力,以提高算法的准确性和实用性。

参考文献

- [1] SAITO K, KIMURA M, OHARA K, et al. Super mediator—a new centrality measure of node importance for information diffusion over social network [J]. *Information Sciences and International Journal*, 2016, 329 (C): 985-1000.
- [2] The 40th China statistical report on internet development [R]. Beijing: China Internet Network Information Center, 2017. (in Chinese)
第40次中国互联网络发展状况统计报告 [R]. 北京: 中国互联网络信息中心, 2017.
- [3] LI Kan, ZHANG Lin, HUANG Heyan. Social influence analysis: models, methods, and evaluation [J]. *Engineering*, 2018, 4 (1): 40-46.
- [4] ZHAO Jiangqiang, GUI Xiaolin, FENG Tian. A new method of identifying influential users in the Micro-blog networks [J]. *IEEE Access*, 2017 (99): 1.
- [5] ZHA Cong, LYU Y, YIN Hao, et al. UCPR: user classification and influence analysis in social network [C]// *Proceedings of IEEE International Conference on Distributed Computing Systems Workshops*. Washington D.C., USA: IEEE Press, 2017: 311-315.
- [6] RIQUELME F, GONZALEZ-CANTERGANI P. Measuring user influence on twitter: a survey [J]. *Information Processing and Management*, 2016, 52 (5): 949-975.
- [7] HAN Zhongming, CHEN Yan, LIU Wen, et al. Research on node influence analysis in social networks [J]. *Journal of Software*, 2017, 28 (1): 84-104. (in Chinese)
韩忠明, 陈炎, 刘雯, 等. 社会网络节点影响力分析研究 [J]. *软件学报*, 2017, 28 (1): 84-104.
- [8] LI Hukang, HUANG Sigi, SUN Guozi. An opinion leader perceptual model based on PageRank algorithm [C]// *Proceedings of IEEE International Conference on Behavioral, Economic and Socio-Cultural Computing*. Washington D.C., USA: IEEE Press, 2015: 150-155.
- [9] ZHANG Yangsen, ZHENG Jia, TANG Anjie. A quantitative evaluation method of micro-blog user authority based on multi-feature fusion [J]. *Acta Electronica Sinica*, 2017, 45 (11): 2800-2809. (in Chinese)
张仰森, 郑佳, 唐安杰. 基于多特征融合的微博用户权威度定量评价方法 [J]. *电子学报*, 2017, 45 (11): 2800-2809.
- [10] WANG Ding, XU Jun, DUAN Cunyu, et al. Improved user influence evaluation algorithm based on PageRank [J]. *Journal of Harbin Institute of Technology*, 2018, 50 (5): 60-67. (in Chinese)
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.
- [11] ZHANG Chenghong, LU Tian, CHEN Shoucong, et al. Integrating ego, homophily, and structural factors to measure user influence in online community [J]. *IEEE Transactions on Professional Communication*, 2017, PP (99): 1-14.
- [12] CRESCI S, PIETRO R D, PETROCCHI M, et al. Fame for sale: efficient detection of fake twitter followers [J]. *Decision Support Systems*, 2015, 80: 56-71.
- [13] CASTELLINI J, POGGIONI V, SORBI G. Fake Twitter followers detection by denoising autoencoder [C]// *Proceedings of International Conference on Web Intelligence*. New York, USA: ACM Press, 2017: 195-202.
- [14] YI Xiushuang, HAN Yeting, WANG Xingwei. The evaluation of online social network's nodes influence based on user's attribute and behavior [J]. *Communications in Computer and Information Science*, 2013, 401: 9-20.
- [15] QI Chao, CHEN Hongchang, YU Hongtao. Method of evaluating micro-blog users' influence based on comprehensive analysis of user behavior [J]. *Application Research of Computers*, 2014, 31 (7): 2004-2007. (in Chinese)
齐超, 陈鸿昶, 于洪涛. 基于用户行为综合分析的微博用户影响力评价方法 [J]. *计算机应用研究*, 2014, 31 (7): 2004-2007.
- [16] ZEYNEP Z A, SULE G O. Identifying topical influencers on Twitter based on user behavior and network topology [J]. *Knowl-Based Systems*, 2018, 141: 211-221.
- [17] SEN T, CHAUDHARY D K. Contrastive study of simple PageRank, HITS and weighted PageRank algorithms: review [C]// *Proceedings of IEEE International Conference on Cloud Computing, Data Science and Engineering Confluence*. Washington D.C., USA: IEEE Press, 2017: 721-727.