

An Evaluation Method for Weibo User Influence with Adjustable Parameters

Hesong Jiang, Pinghua Chen

School of Computer, Guangdong University of Technology, GDUT, Guangzhou Guangdong
Email: jianghesong606@gmail.com

Received: Dec. 27th, 2018; accepted: Jan. 10th, 2019; published: Jan. 17th, 2019

Abstract

Aiming at the problem of PR value loss caused by the classic and improved PageRank algorithm in the calculation of microblog forwarding and comments, a parameter-adjustable Weibo user influence evaluation method RCRank is proposed. Based on the existing PageRank algorithm, RCRank adds microblog forwarding and comments to adjust the parameters that affect the influence calculation of Weibo users. The experimental results on the actual data set show that the RCRank method is closer to the real situation than the traditional PageRank and TunkRank algorithms when calculating user influence.

Keywords

Weibo User Influence, Directed Graph, PageRank, Adjustable Parameters

一种参数可调的微博用户影响力评价方法

江和松, 陈平华

广东工业大学计算机学院, 广东 广州
Email: jianghesong606@gmail.com

收稿日期: 2018年12月27日; 录用日期: 2019年1月10日; 发布日期: 2019年1月17日

摘要

针对经典和改进的PageRank算法在计算微博转发和评论时采取PR均值分配进而造成PR值损失问题, 提出了一种参数可调的微博用户影响力评价方法RCRank。RCRank在已有的PageRank算法基础上, 增加了微博转发和评论影响微博用户影响力计算值的可调节参数。在实际数据集上的实验结果表明, 在计算用户影响力时, 相比于传统的PageRank和TunkRank算法, RCRank方法更接近真实情况。

关键词

微博用户影响力, 有向图, PageRank, 参数可调

Copyright © 2019 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

Facebook、Twitter、微博、微信等流行的在线社交网络(Online Social Network, 简称 OSN)工具正在改变用户与用户之间的互联网通信和互动方式[1] [2], 网络社交、网上阅读、网上评论/转发正逐渐被人们接受, 成为人们主要的社交方式。据《中国互联网络发展状况统计报告》显示: 截止 2018 年 6 月, 中国网民规模达到 8.02 亿人, 其中微博用户规模为 3.37 亿人。以 2018 年 8 月份发生的 85 度 C 事件为例, 在强大网民的舆论下, 85 度 C 在台湾的股价连续 3 天下跌 21%, 足见网络传播的威力。由于微博用户量大、传播迅速, 微博传播规律的研究正成为政府和业界关注的焦点。研究 OSN 用户影响力, 挖掘网络传播关键路径和关键节点, 为政府就如何加快正面信息传播和控制负面新闻扩散提供决策支持, 这很有意义。

就微博用户影响力评价而言, 近几年, 业界涌现出多种方法, 典型的方法包括:

1) 基于 URL 追踪的评价方法。该方法通过追踪 URL 传播情况, 按一定方式分配相应的影响力进而进行评价[3]。

2) 基于用户关系的评价方法。该方法参考 Google 的 PageRank 算法或其改进算法[4]-[9], 但没有考虑到“僵尸粉”影响, 没有考虑到微博转发或者评论情况。

3) 基于用户行为的评价方法。微博用户行为包括微博转发、评论和点赞等, 研究微博用户行为, 主要是为排除“僵尸粉”干扰, 通过结合微博用户行为计算用户影响力, 能更符合真实情况[10] [11] [12] [13]。这些方法中, 有的考虑到用户粉丝数、用户是否认证、用户微博的转发和评论等指标, 有效地解决了“僵尸粉”干扰情况。但有的忽略了微博用户本身影响力的不同, 且单考虑用户行为因素, 忽略了用户之间本身存在影响力传播不均匀情况, 所以出现了基于用户行为权值的评价方法。

4) 基于 PageRank 和用户行为权值的评价方法[14] [15] [16]。近几年, 有学者提出了传统的 PageRank 算法应用在计算微博用户影响力时会存在 PR 值分配不合理问题。例如, 一个用户若经常转发或者评论某博主用户的微博, 那么在传递 PR 值(影响力值)时, 相比其他微博用户, 该用户理应给予更大 PR 值才较为合理一些。基于此, 提出引入权值思想的方法, 这些方法都得到较好效果, 但也存在一些不足。如下:

文献[14]为了弥补传统 PageRank 算法均值分配不足, 提出基于用户行为权值分配的 BWPR (PageRank based on behavior weight)算法, 给特定用户分配大一些 PR 值, 但该算法计算过程复杂, 不利于大规模数据计算。文献[15]采用加权 PageRank 算法结合 K-means 算法计算各个领域用户的影响力, 虽然节点影响力的传播按照边的权重分配, 但算法仅适用于特定领域加权计。文献[16]利用权重修正传统的 PageRank 模型的概率转移矩阵, 能取得较好的效果, 但没有考虑用户微博转发和评论等特征。

综合考虑以上存在的问题, 本文提出 RCRank 算法, 该方法不仅考虑到微博转发和评论特征, 而且考虑到传统 PageRank 算法存在均值分配不合理问题, 通过引入可调节参数进行修正, 且通过试验证明, 该方法计算过程简单有效。

2. 相关工作

2.1. 经典 PageRank 算法

PageRank 算法又称为佩奇排名算法, 是谷歌用来衡量网页重要性、标识网页等级、进行网页排序的一种方法。经典的 PageRank 算法如公式(1)所示。

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

其中: p_i 和 p_j 表示网页页面, N 表示网页页面的总数, $L(p_j)$ 表示页面 p_j 的出链数, q 表示阻尼系数。 q 一般取值 $q = 0.85$ 。

2.2. 改进的用户影响力评价算法

用户间微博评论关系示意图如图 1 所示。

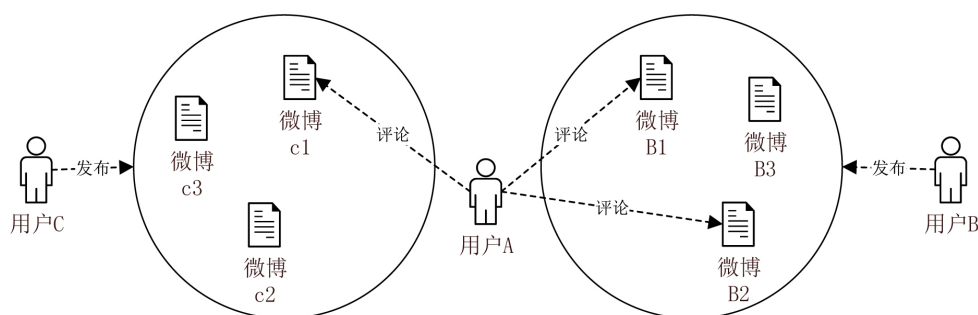


Figure 1. Schematic diagram of Weibo comment relationship between users
图 1. 用户间的微博评论关系示意图

在图 1 上, 用户 A 经常会评论用户 B 发布的微博。经典的 PageRank 算法在计算微博评论的 PR 值时, 使用了图 2 所示的有向图。

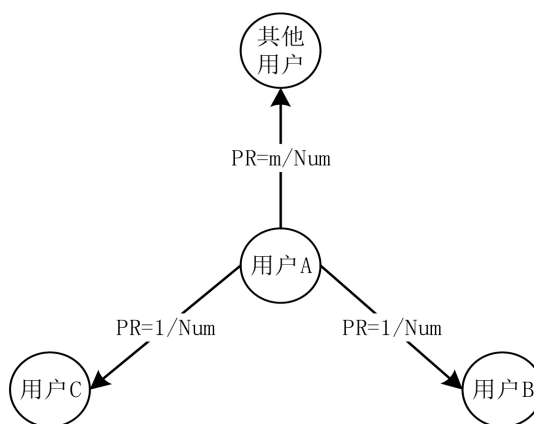


Figure 2. Traditional PageRank algorithm PR value distribution map
图 2. 传统 PageRank 算法 PR 值分配图

其中, Num 表示用户 A 评论的微博所涉及的博主总数, m 表示 B、C 之外的其他用户数量。尽管用户 A 评论用户 B 的微博次数多于评论用户 C 的微博次数, 但由于采取了 PR 值均值策略, 因此, 在计算

用户 A 的评论 PR 值时, 均取 $1/\text{Num}$ 值。这很显然与实际不符。实际上, 应该给用户 B 分配多一点 PR 值会更合理一些。

但传统 PageRank 算法模型采取的是均值分配方式, 所以如同 2 所示, 用户 C 和用户 B 传递得到的 PR 值是相等的($\text{PR} = 1/\text{Num}$)。这样相当于损失掉用户 A 传递给用户 B 的部分 PR 值。为了解决这个问题, 本文根据用户微博转发和评论情况引入了调节参数, 如图 3 所示:

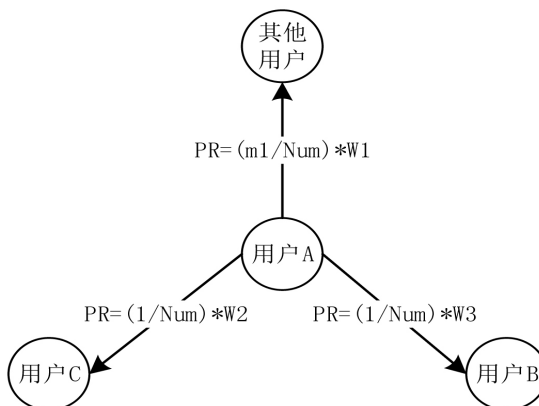


Figure 3. Improved PageRank algorithm comment assignment PR value directed graph
图 3. 改进后 PageRank 算法评论分配 PR 值有向图

其中 W_1 、 W_2 和 W_3 为调节参数。通过统计每个用户评论的情况, 可计算得到对应的 W_1 、 W_2 和 W_3 (具体计算见公式 2 和公式 3)。根据图 3 可知, 改进后的算法, 虽然分配思想上还是按传统 PageRank 算法模型进行分配, 但是通过统计用户评论情况, 每条边都增加对应的调节参数, 用户 A 传递给用户 C 和用户 B 的 PR 值大小与 W_2 和 W_3 有关, 而不是传统 PageRank 算法的均值分配。通过计算用户 A 对各个用户微博的评论情况, 使得用户 A 更加合理的传递 PR 值。

以下是计算微博用户影响力所涉及的几个定义和计算公式。

定义 1: 用户微博转发调节参数 W_{ij} 。计算微博转发影响力时, 根据微博用户实际转发情况合理调整 PR 调节参数 W_{ij} , 如公式(2)所示。

$$W_{ij} = \frac{r_{ij} * QR_i}{R_i} \quad (2)$$

其中: W_{ij} 表示用户 i 对用户 j 转发出链的调节参数, QR_i 表示用户 i 所有转发微博涉及到的微博用户总数, r_{ij} 表示用户 i 转发了用户 j 的微博总数, R_i 表示用户 i 所有转发微博总数。

定义 2: 用户微博评论调节参数。计算微博评论影响力时, 根据微博用户实际评论情况合理调整 PR 调节参数 u_{ij} , 如公式(3)所示。

$$u_{ij} = \frac{c_{ij} * QC_i}{C_i} \quad (3)$$

其中: u_{ij} 表示用户 i 对用户 j 评论出链的调节参数, 转发调节参数是在应用 PageRank 算法计算微博转发影响力时, 根据微博用户评论情况合理地分配 PR 值调节变量。

定义 3: 微博用户粉丝间影响力。微博用户之间关系记为有向图 $G = \langle V, E \rangle$, 其中 V 表示被研究的微博用户集合(即有向图的顶点集合), E 表示被研究用户之间粉丝关注情况(即有向图的边)。微博用户 i 粉丝间影响力计算如公式(4)所示。

$$F_Rank(i) = \frac{1-q}{N} + q \sum_{j \in G} \frac{F_Rank(j)}{L(j)} \quad (4)$$

其中: $F_Rank(i)$ 表示用户 i 的粉丝影响力, j 表示用户 i 的粉丝用户, $L(j)$ 表示用户 j 粉丝数, T 表示粉丝用户集合。

定义 4: 微博转发影响力。考虑微博用户影响不同, 在微博转发过程中所带来的影响力值也不一样, 所以传统算法使用微博的转发个数不能准确评价微博的影响力。RCRank 算法计算微博转发影响力是通过构建微博用户转发关系列表, 增加了微博转发调节参数。结合公式(2)用户 i 微博转发影响力计算如公式(5)所示。

$$R_Rank(i) = \frac{1-q}{N} + q \sum_{k \in T} \frac{R_Rank(k)}{L(k)} * W_{ik} \quad (5)$$

其中: $R_Rank(i)$ 表示用户 i 微博转发的影响力, k 表示转发用户, $L(k)$ 表示用户 k 转发微博的个数, T 表示转发用户集合。

定义 5: 微博评论影响力。同理, 计算微博评论影响力时, 也加入微博评论的调节参数 u_{ij} , 最终的微博评论影响力如公式(6)所示。

$$C_Rank(i) = \frac{1-q}{N} + q \sum_{e \in M} \frac{C_Rank(e)}{L(e)} * U_{ie} \quad (6)$$

其中: $C_Rank(i)$ 表示用户 i 微博评论影响力, e 表示评论用户, $L(e)$ 表示用户 e 评论微博的个数, M 表示评论用户集合。

定义 6: 微博用户影响力。通过增加微博转发和评论可调参数与博粉丝影响力相结合, 可得微博用户影响力计算公式, 如公式(7)所示。

$$Rank(i) = \alpha * F_Rank(i) + \beta * R_Rank(i) + \gamma * C_Rank(i) \quad (7)$$

其中: $Rank(i)$ 表示微博用户 i 的综合影响力, 由于得到的用户粉丝影响力、微博转发和评论影响力不能直接相加, 所以引入 α, β 和 γ 作为权重(且满足 $\alpha + \beta + \gamma = 1$), 目的是可以通过设置不同权重值, 得到更加合理的用户影响力。

实现算法流程:

第一步: 用户所发微博转发调节参数 W_{ij} 计算流程。

结合表 1 获取的数据格式(初始调节参数为 1.000, 暂为不需要调整的用户), 计算微博转发调节参数值的算法流程如下:

输入: 微博转发关系列表。

处理核心流程:

1. 连接数据库
2. result ← 查询微博转发关系表
3. While (result 记录非空)
4. 初始化 $r_{ij} = 0$, $QR_i = 0$, $R_i = 0$, $W_{ij} = 1$
5. if 用户 i 转发了用户 j 的微博
6. $r_{ij} = r_{ij} + 1$
7. if 转发用户 id == 用户 i
8. $R_i = R_i + 1$

9. if 转发用户 id == 用户 i && 博主 id 不同
10. $QR_i = QR_i + 1$
11. 计算 $W_{ij} = r_{ij} * QR_i / R_i$
12. 将调节参数计算结果 W_{ij} 更新到微博转发关系表中
13. 重复循环步骤 2
14. 循环体执行完, 则程序结束

Table 1. Weibo forwarding relationship**表 1.** 微博用户转发关系表

转发用户 id	博主 id	微博 id	调节参数
164**6540	263**6803	M_F9V1vkI0L	1.000
164**6540	263**6803	M_F9V1vkI0L	1.000
164**6540	263**6803	M_F9V1vkI0L	1.000
164**6540	263**6803	M_FyifsbegN	1.000
.....
597**1984	194**8655	M_F85qx42sC	1.000

第二步: 计算用户影响力算法

算法输入: 用户信息表, 用户关注粉丝表, 带有调节参数的微博转发和评论关系表

1. 构建有向图 $Graph_f=(VE)$ 、 $Graph_r=(VE)$ 和 $Graph_c=(VE)$ 。
2. $PR_f \leftarrow Graph_f.PageRank(0.01)$ 微博用户粉丝表使用传统 PageRank 算法
3. $PR_r \leftarrow runUntilConvergenceWithOptions(Graph_r, 0.01)$ 微博转发使用改进的算法
4. $PR_c \leftarrow runUntilConvergenceWithOptions(Graph_c, 0.01)$ 微博评论使用改进的算法
5. $Rank = \alpha * Rank_f + \beta * Rank_r + \gamma * Rank_c$

6. 通过计算的结果 Rank, 可得到微博用户最终影响力排名

算法输出: 最终的用户影响力排名和对应的影响力值

`runUntilConvergenceWithOptions()` //函数具体实现代码如下:

```
valpagerankGraph: Graph[(Double, Double), Double] = graph.outerJoinVertices(graph.outDegrees) {
  (vid, vdata, deg) => deg.getOrElse(0)}.mapTriplets {e=>(1.0/e.srcAttr)*e.attr.toString.toDouble} //修改后的
将每条边的调节参数加入且设定  $\delta = 0.01$ 。用户影响力计算过程, 主要通过用户关注粉丝、微博转发
和评论加权求综合排序, 得到最终排名。
```

3. 实验与结果分析

3.1. 实验数据收集

目前, 新浪微博的数据获取主流有两种方法: 一种是通过微博开发商提供的 API 获取, 这种方法首先需要申请开发者权限, 获取授权用户 API 接口, 且授权用户对接口访问频率有限制, 难于短时间获取大量的数据; 另一种是使用网络爬虫工具, 简单轻巧, 并且使用起来非常方便。

RCRank 算法因需要获取大量的微博转发和评论用户 id 列表, 考虑到微博官方提供的 API 接口访问频率受限, 为了快速获取数据信息, 本文采取 API 结合爬虫两种方式采集数据。

获取数据的环境为: 在虚拟机搭建 7 个 Ubuntu 系统, 其中分配一个主节点(Redis 服务器)用于管理爬虫请求队列, 剩下的 6 个作为子节点(Spider 爬虫服务器)用于爬取网页数据, 如图 4 所示。

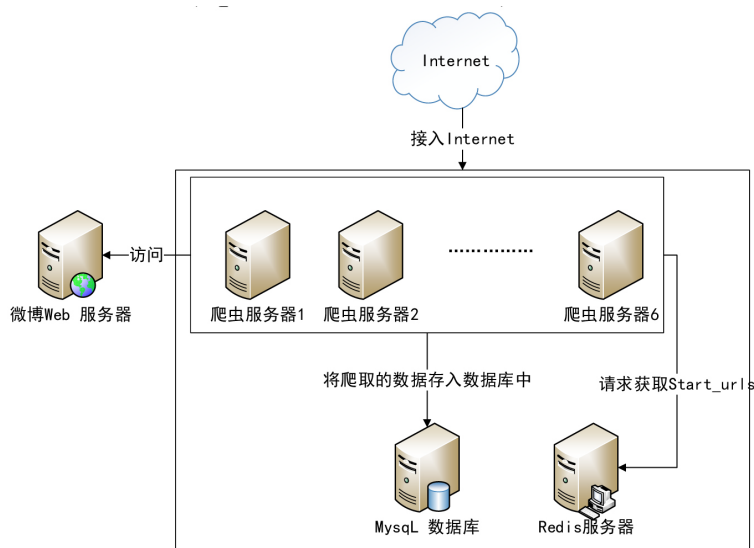


Figure 4. Distributed crawler data acquisition principle framework

图 4. 分布式爬虫数据获取原理框架

采用爬虫获取到 100,000 个微博用户数据, 包含有用户的粉丝关系数据和 2017 年 11 月至 2017 年 12 月份的微博转发和评论关系数据。数据清洗后, 得到 69,000 个用户, 然后随机截取出 5000 微博用户作为实验数据, 虽然用户数量少了, 但用户之间的关系保留完整。

3.2. 实验结果与分析

为了证明 RCRank 算法的有效性, 本文采用传统的 PageRank 算法、TunkRank 算法与 RCRank 算法在相同的数据集上进行对比实验。TunkRank 算法如公式(8)所示。

$$Influence(v) = \sum_{u \in Followers(v)} \frac{1 + p * Influence(u)}{\|Following(u)\|} \quad (8)$$

其中: $Influence(v)$ 表示 v 的所有粉丝集合, $Following(u)$ 表示关注用户 u 的用户集合。 p 表示假定用户 v 阅读一条用户 u 发布的微博后转发该微博的概率。

从公式(8)可以看出, TunkRank 算法在计算用户影响力是将微博用户的粉丝影响力 $1 / Following(u)$ 与粉丝转发微博所带来的影响力 $p * Influence(u) / Following(u)$ 相结合, 这样可以有效解决“僵尸粉”问题。

为了验证 RCRank 算法得到的用户影响力更接近真实影响力, 引入新浪微博官方影响力评价方法。该方法影响力构成: 用户粉丝质量, 短微博/长微博/视频发布数, 被转发, 评论, 赞次数; 被阅读, 打赏, 订阅, 曝光次数等。计算公式如公式(9)所示。

$$影响力 = a \times 活跃度 + b \times 互动力 + c \times 覆盖度 \quad (9)$$

满分为 100 分, 由活跃度, 互动力, 覆盖度三部分完成, 所占比例分别为: 25%, 55%, 20%。活跃度主要评测指标为短微博/长微博/视频发布量; 互动力主要评测指标为短微博/长微博/视频被转, 被评论, 被赞次数; 覆盖度主要评测指标有粉丝数, 阅读数, 打赏, 曝光, 篇均阅读等。

由于微博 V 认证用户与普通用户间影响力差距很大, 为方便观察和对比, 将所有实验计算得到的分数限制到一定区间, 活跃度, 互动力和覆盖度按上述比例采用归一化算法进行计算, 以 100 分为上限, 保留四位小数, 以四舍五入原则计入。如公式(10)所示。

$$X_{score} = \frac{X - X_{min}}{X_{max} - X_{min}} * 100 \quad (10)$$

其中: X 表示计算出的当前用户的影响力值, X_{min} 表示所有用户影响力的最小值, X_{max} 表示用户中影响力的最大值。

其中表 2 为微博官方影响力评价方法得到的用户影响力排名, 表 3 为使用 RCRank 算法得到的用户影响力排名。从图 5 发现四种方法计算得到的最后用户影响力排名有一定差别, 传统 PageRank 算法和 RCRank 算法计算得到的最后排名比较接近, 但 RCRank 算法在传统 PageRank 算法基础上结合微博转发和评论具体情况引入可调参数, 对传统的 PageRank 算法进行了改进, 使用非均值分配 PR 值思想代替传统 PageRank 算法的 PR 值的均值分配思想。

为了更加清楚看出实验对比结果, 引入均方根误差(RMSE)来准确度量三种算法来计算排名结果与真实排名的误差值。具体的评测公式如公式(11)所示。

$$RMSE = \sqrt{\sum_{i=1}^n (X_{model,i} - X_{real,i})^2} \quad (11)$$

Table 2. Ranking of influences based on official formulas

表 2. 根据官方公式得到的影响力排名

排名	微博用户名	粉丝数	微博数	评论转发总数	排名得分
1	coser 狱凰	1166	190	282	8.0222
2	甄洋洋洋	1197	110	131	5.3211
3	宋波小疯子	1163	56	100	4.4510
4	蕴华同人	1128	85	88	4.4510
5	抹茶阿姨 i-	1427	70	37	4.1916
6	狸奴_女	1346	12	67	4.1188
7	闪闪-Marbury	897	109	96	4.0765
8	魔王 1 号微	790	88	120	4.0683
9	护肤常识-	1714	5	8	4.0413
10	Cogitoergo	1270	34	55	3.8863

Table 3. RCRank algorithm calculation results

表 3. RCRank 算法计算结果

排名	微博用户名	粉丝数	转发总数	评论总数	排名得分
1	coser 狱凰	1166	117	165	2448.42
2	甄洋洋洋	1197	100	31	335.78
3	蕴华同人	1128	30	58	258.75
4	抹茶阿姨 i-	1427	0	37	169.33
5	闪闪-Marbury	897	28	68	158.62
6	宋波小疯子	1163	0	100	155.17
7	魔王 1 号微	790	67	55	137.51
8	护肤常识-	1714	1	7	134.64
9	Cogitoergo	1270	7	48	134.50
10	狸奴_女	1346	34	33	120.38

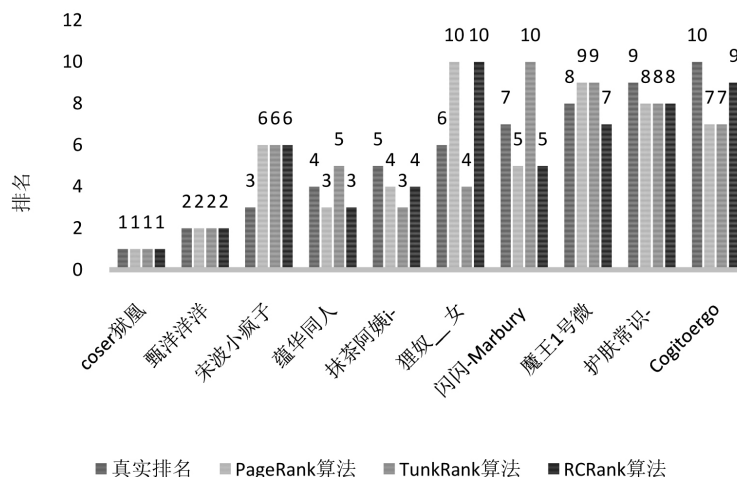


Figure 5. Weibo user ranking order calculated by different methods
图 5. 不同方法计算得出的微博用户排名顺序

其中: n 表示测量次数, $X_{model,i}$ 表示模型的计算值, $X_{real,i}$ 表示真实值。用真实排名位次表示 $X_{real,i}$ 数值, 使用算法计算出的排名位次来表示 $X_{model,i}$ 。

通过最后实验对比可知, RCRank 算法得出的 $RMSE = 5.83095189485$, 为最低值, 表明改进后的算法模型计算得到的结果排名和真实排名更接近, 如表 4 所示。

Table 4. RMSE values calculated by the three algorithms

表 4. 三种算法计算的结果 RMSE 值

算法名称	PageRank 算法	TunkRank 算法	RCRank 算法
RMSE	6.48074069841	6.0	5.83095189485

这也进一步体现了改进后的算法, 更加符合实际情况。综上所述, RCRank 算法是可以合理调整微博用户传递分配的 PR 值, 使每个微博用户所分配 PR 值更合理。

通过图 6 显示用户“闪闪-Marbury”微博被转发和评论的情况, 调节参数大于 1 表示这些用户转发或评论了博主多条微博。调节参数小于 1 表示这些用户更多的关注其他用户, 对用户“闪闪-marbury”只是偶尔转发或评论一次, 如果根据 RCRank 算法思想, 对关注度更大的用户传递更多 PR 值, 所以调节参数小于 1, 能够传递到“闪闪-marbury”的 PR 值相对更少。对于图 6, 调节参数等于 1 有两种情况: 第一, 表示这些用户只转发了或评论了“闪闪-marbury”的微博。第二, 表示这些用户对其它用户的微博转发或评论次数相同, 所以还是按原来的分配值进行分配, 不作调整。

通过引入公式来度量调节参数越大, 属于忠实粉丝的概率越大, 概率公式如公式(12)所示。

$$P_{\text{概率}} = 1 - \frac{1}{e^x} \tag{12}$$

其中 x 为调节参数值, 范围 $x \in [0, +\infty)$, 当 $x = 0$ 表示该用户没有评论, 假设该用户转发的调节参数值为 0, 综合考虑转发和评论情况, 可知该用户是博主忠实粉丝的概率为 0%。当 x 值越大, 表明该用户对博主用户发布的微博评论所占比重越高, 通过公式(12)可计算得出该用户是博主忠实粉丝概率越接近 100%。结合上述公式得到见图 7。

综合上述对图 6 解释, 结合观察图 7 可以在一定程度上挖掘出用户“闪闪-marbury”忠实粉丝有哪些, 例如用户“5848468419”, 对博主微博评论调节参数值达到 1.50, 说明相对其它用户, 更热衷于评

论“闪闪-marbury”所发布的微博。相反, 用户“1646406540”对“闪闪-marbury”微博转发调节参数为0.1, 说明相对其它用户基本不转发“闪闪-marbury”的微博, 很有可能只是一个路过者, 偶尔转发过一次。通过观察用户“闪闪-marbury”的粉丝列表发现用户“5848468419”在博主的粉丝列表中, 而用户“1646406540”不在博主的粉丝列表中。

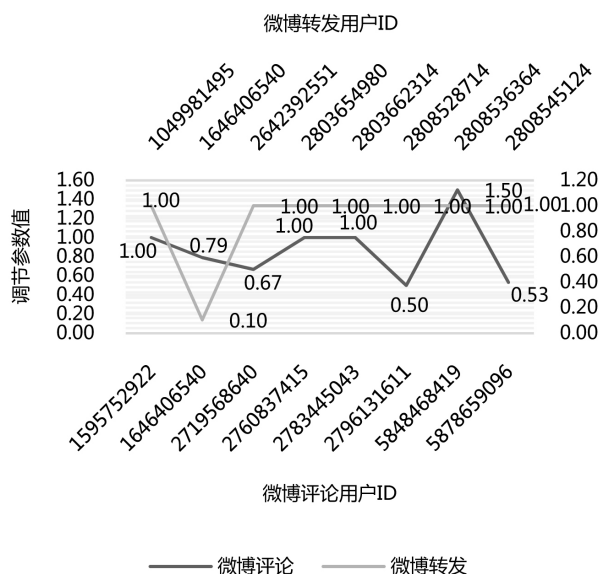


Figure 6. Forwarding and commenting on the distribution of adjustment parameters generated by the user’s “Sparkle-Marbury” microblog

图 6. 转发和评论用户“闪闪-Marbury”微博产生的调节参数分布

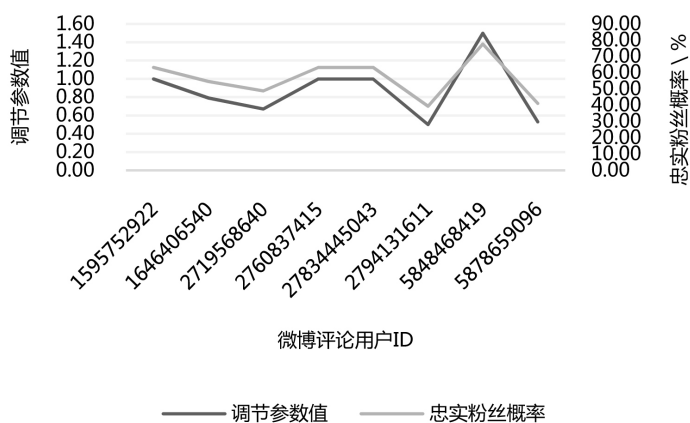


Figure 7. Comment on the distribution of adjustment parameters generated by the user “Sparkle-Marbury” microblog

图 7. 评论用户“闪闪-Marbury”微博产生的调节参数分布

4. 结束语

RCRank 算法中的可调节参数和近几年学者提出的边的权值, 在作用上是类似的, 都是为了解决传统基于 PageRank 计算方法中 PR 均值分配问题, 但是计算过程和原理不同, RCRank 算法主要从 PageRank 算法的底层结构分析, 从构建有向图的层面, 分析了有向图的入链和出链情况, 增加一个可以协调分配的调节参数, 从而解决合理分配 PR 值的问题。通过实验结果对比, RCRank 算法比传统 PageRank 算法在 PR 值分配上更加合理, 计算得到 RMSE 值比 TunkRank 算法更小。其次通过研究微博转发和评论的调

节参数, 发现调节参数大小在一定程度上可以作为挖掘用户列表中忠实粉丝的度量指标, 但怎么合理的找出是否为忠实粉丝的度量阈值, 提高挖掘的准确率, 这将成为以后研究的工作内容的一部分。

基金项目

国家基金项目(61572144); 广东省科技计划项目(No. 2016B030306002, No.2015B010110001, 2017B030307002)。

参考文献

- [1] Kwak, H., Lee, C., Park, H., *et al.* (2010) What Is Twitter, a Social Network or a News Media. *Proceedings of the 19th International Conference on World Wide Web*, New York, 26-30 April 2010, 591-600. <https://doi.org/10.1145/1772690.1772751>
- [2] Page, Lawrence, Brin, *et al.* (1998) The Pagerank Citation Ranking. Bringing Order to the Web. Stanford InfoLab.
- [3] Lamberti, F., Sanna, A. and Demartini, C. (2009) A Relation-Based Pagerank Algorithm for Semantic Web Search Engines. *IEEE Trans on Knowledge & Data Engineering*, **21**, 123-136. <https://doi.org/10.1109/TKDE.2008.113>
- [4] Jing, Y.S. and Baluja, S. (2008) Pagerank for Product Image Search. *Proceedings of the 17th International Conference on World Wide Web*, Beijing, 21-25 April 2008, 307-316. <https://doi.org/10.1145/1367497.1367540>
- [5] Tunkelang, D. A Twitter Analog to Pagerank. <https://thenoisychannel.com/2009/01/13/a-twitter-analog-to-pagerank>
- [6] 王顶, 徐军, 段存玉, 吴玥瑶, 孙静. 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.
- [7] Bakshy, E., Hofman, J.M., Mason, W.A., *et al.* (2011) Everyone's an Influencer: Quantifying Influence on Twitter. *Proceedings of the 4th ACM International Conference on Web Search and Data Mining*, Hong Kong, 9-12 February 2011, 65-74. <https://doi.org/10.1145/1935826.1935845>
- [8] Yamaguchi, Y., Takahashi, T., Amagasa, T., *et al.* (2010) Turank: Twitter User Ranking Based on User-Tweet Graph Analysis. *Proceedings of Web Information Systems Engineering*, Hong Kong, 12-14 December 2010, 240-253. https://doi.org/10.1007/978-3-642-17616-6_22
- [9] 张昊, 刘功申, 苏波. 一种微博用户影响力的计算方法[J]. 计算机应用与软件, 2015, 32(3): 41-44.
- [10] Liu, Y. and Lin, Y. (2007) Supervised Hits Algorithm for Medline Citation Ranking. 2007 *IEEE 7th International Symposium on BioInformatics and BioEngineering*, Boston, MA, 14-17 October 2007, 1323-1327. <https://doi.org/10.1109/BIBE.2007.4375740>
- [11] Asano, Y., Yu, T. and Nishizeki, T. (2008) Improvements of Hits Algorithms for Spam Links. *IEICE Transactions on Information & Systems*, **E91D**, 200-208.
- [12] Weng, J., Lim, E., Jiang, J. and He, Q. (2010) Twitter Rank: Finding Topic-Sensitive Influential Twitterers. *Proceedings of the 3rd ACM International Conference on Web Search and Data Mining*, New York city, New York, 4-6 February 2010, 261-270.
- [13] Cha, M., Haddad, H., Benevenuto, F., *et al.* (2010) Measuring User Influence in Twitter: The Million Follower Fallacy. *Proceedings of the 4th International AAAI Conference on Weblogs and Social Media*, Washington, DC, 10-17.
- [14] 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.
- [15] 刘金龙, 吴斌, 陈震, 沈崇玮. 基于领域划分的微博用户影响力分析[J]. 计算机科学, 2015, 42(5): 42-46.
- [16] 毛国君, 谢松燕, 胡殿军. PageRank 模型的改进及微博用户影响力挖掘算法[J]. 计算机应用与软件, 2017, 34(5): 28-32.

知网检索的两种方式：

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择：[ISSN]，输入期刊 ISSN：2161-8801，即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入，输入文章标题，即可查询

投稿请点击：<http://www.hanspub.org/Submission.aspx>

期刊邮箱：csa@hanspub.org