计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.181202253

融合用户自身因素与互动行为的微博用户影响力计算方法

王新胜　马树章
江苏大学计算机科学与通信工程学院　江苏 镇江 212013

摘　要　由于微博高影响力用户在商品营销、社会舆论引导等方面起着重要的作用,因此挖掘高影响力用户成为了微博社交网络中的热点研究问题。针对微博用户影响力计算中存在交互行为与用户自身因素分析不全面的问题,提出了微博用户影响力计算方法 MBUI-SFIM(Micro-blog userinfluence based on user's self-factors and interaction computing model)。该方法考虑了微博用户直接影响力和间接影响力两个方面:在用户直接影响力计算中,通过对用户的自身因素如微博用户粉丝数、用户活跃度、近期微博质量等的分析,计算出用户的初始影响力,然后分析用户互动行为如用户的微博可见率、微博用户互动系数,计算出用户传播能力,最后将初始影响力与用户传播能力相结合,基于改进 PageRank 算法计算出用户直接影响力;在用户间接影响力计算中,通过对用户网络图连接结构进行分析,根据不相邻用户连接路径的不同,将用户间接影响具体分为简单路径、重复路径、复杂路径 3 种情况进行讨论,从而计算出用户间接影响力。实验结果表明,相比 PageRank 算法和 MR-UIRank 算法,所提算法在用户排名准确性上分别提高了 14.8%和 8.3%。

关键词:微博;自身因素;交互行为;PageRank;直接影响力;间接影响力
中图法分类号　TP393

Method of Weibo User Influence Calculation Integrating Users' Own Factors and Interaction Behavior

WANG Xin-sheng and MA Shu-zhang
Department of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang, Jiangsu 212013, China

Abstract　Weibo users with high-impact play an important role in commodity marketing and social publicity guidance, so mining high-impact users becomes a hot research issue in Weibo social networks. As for the problems of incomplete behavior analysis of interaction behavior and user's own factors in calculation of micro-blog user influence, the micro-blog user influence based on user's self-factors and interactiont computing model was proposed. This method considers the direct influence and indirect influence of Weibo users. In the user's direct influence calculation phase, the initial influence of the user is calculated by analyzing the user's own factors such as the number of fans of Weibo users, user activity, and recent microblog quality. Then the user interaction behavior is analyzed, such as the user's microblog visibility rate, microblog user interaction coefficient, so as to calculate the user communication ability. Finally, by combining the initial influence with the user communication ability, the user's direct influence is cakulated based on the improved PageRank algorithm. In the calculation of user indirect influence phase, through the analysis of the connection structure of the user network diagram and according to the different connection paths of non-adjacent users, the indirect impact of the user is divided into three categories: simple path, repeated path and complex path, then the user indirect influence is calculated. The experimental results show that the proposed algorithm is 14.8% and 8.3% higher than the PageRank algorithm and the MR-UIRank algorithm in terms of the user ranking accuracy.

Keywords　Microblog, Self-factor, Interaction behavior, PageRank, Direct influence, Indirect influence

1　引言

随着信息时代的来临,互联网已经进入 Web2.0 时代,成为人类社会关系维系和信息传播载体的工具。一些新的社交网络应用的出现,打破了长期以来信息一直被传统门户垄断的格局[1]。微博应用作为社交网络中出现的新一代信息交流与获取平台,以其开放性高、信息传播快、交互性强等特点,成为了较受欢迎的社交网络应用之一。微博应用凭借其自身的信息传播能力,已经成为当前热点事件讨论的首发地。在微博社交网络中,如果某个用户发表的观点和想法总能被其他

到稿日期:2018-12-03　返修日期:2019-05-01　本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。
基金项目:中国博士后科学基金(2015M571688);江苏大学高级技术人才科研基金(12JDG104)
This work was supported by the China Postdoctoral Science Foundation (2015M571688) and Research Fund for Advanced Technical Talents of Jiangsu University (12JDG104).
通信作者:王新胜(wxs@ujs.edu.cn)

用户所认同，或者某个用户总能对当前热点事件给出自己独特的评价，并引起微博用户直接讨论和关注，则称此类用户为微博应用中具有影响力的用户[2]。对于高影响力用户的挖掘，在微博社交平台就显得更为重要。

针对影响力分析中常常忽略间接影响力的问题，文献[3]通过单一路径与多路径分析了移动用户的间接影响力，但是该方法对用户连接路径问题考虑得不够全面，未考虑到复杂路径，因此具有一定的局限性。为了识别某一话题内最有影响力的用户，Weng 等[4]提出了 Twitterrank 算法以计算 Twitter 中的用户影响力，但是该算法只考虑了相似话题之间的用户影响力，不具有一般性。文献[5]针对微博社交网络用户关系网络中存在的僵尸粉现象，在计算用户影响力过程中分析了微博用户的实际行为，引入了微博用户权重来计算用户影响力，但是该算法并没有考虑微博用户的初始影响力。本文提出的 MBUI-SFIM 算法在微博用户影响力计算过程中，综合考虑影响微博用户影响力的影响因素，利用改进的 PageRank 算法[6]计算微博用户直接影响力，根据不相邻用户连接路径的不同分类处理用户间接影响力，然后得到用户总影响力。最后，实验证明了 MBUI-SFIM 算法的有效性。

本文第 2 节详细探讨近年来用户影响力的相关研究；第 3 节提出本文的算法模型；第 4 节通过实验验证算法的有效性；最后总结全文并展望未来。

2 相关工作

用户影响力是社交网络分析的重要内容，也是社交网络中的一个重要研究点。为了提高用户影响力计算的准确性，Yamaguchi 等通过加入微博发布和转发关系来对 PageRank 算法进行改进，提出了 TUrank 算法[7]，解决了传统算法在用户影响力计算时只考虑用户关注关系的问题，但是该方法均衡分配各种关系的边权，具有一定的不足之处。传统的文献在评价用户影响力时，通常认为粉丝量大的用户的影响力大。文献[8]在研究微博用户影响力传播时，把微博用户粉丝数量和微博转发数量进行对比，发现粉丝数多的微博用户所发微博并不一定被其他用户转发，从而说明粉丝数量多的用户的影响力不一定会很大，但该算法没有综合考虑微博粉丝的互动行为。文献[9]针对影响力研究集中于单一网络的现象，通过扩展基于树的算法模型来解决多社交网络用户影响力最大化问题。Liu 等[10]将识别用户影响力强度与用户节点间的话题分布相结合，基于概率模型来计算用户影响力间的影响力强度。毛佳昕等[11]在微博用户影响力分析上结合微博用户行为与网络结构两个要素来分析用户影响力，解决了算法在分析影响力时忽略用户阅读行为的问题。为了更加准确的获取影响力与时间之间的关系，Zhou 等[12]使用回归模型分析用户影响力，揭示了用户影响力随时间变化的规律，但是文献[11-12]忽略了对用户微博内容的考虑，因此具有一定的局限性。文献[13]基于用户的显式用户关系和回复内容用户关系两种关系多视角评估用户的影响力。文献[14]将口碑效应原理与信息传播模型相结合计算用户影响力。但是文献[13-14]在用户影响力分析过程中，未考虑到用户间接影响力问题。文献[15]通过转发、评论等行为分析用户间的直接互动，将用户间的直接交互纳入到矩阵分解目标函数中，解决了推荐算法中用户影响力的计算问题，但是其仅考虑到直接交互，并未考虑间接交互。

以往的用户影响力研究工作大多关注微博网络结构与用户属性。但是在网络结构分析上对用户间接影响力的分析不够全面，在用户属性分析方面不能系统地分析用户的个人属性与行为属性，导致影响力计算稍有误差。因此，本文从直接影响与间接影响两方面分析用户影响力，在直接影响分析中系统地分析了用户的自身因素与互动行为，并根据不相邻用户不同的连接路径详细介绍了用户间接影响力的计算方法，最终融合用户自身因素与互动行为提出了 MBUI-SFIM 算法。

3 微博用户影响力计算方法 MBUI-SFIM

3.1 构造社交网络图

在计算微博用户影响力时需要构造社交网络关系图，本文使用有向图 $G(V,E)$ 来建模社交网络图。其中，V 是微博网络中用户顶点集合，E 是直接相连用户边集合。如果用户 v 关注用户 u，那么从用户 v 有一条边指向用户 u。本文根据爬取到的微博用户互粉关系，构造了微博用户连接样例图(见图 1)，图 1 中含有 400 个用户和 2000 个关系连接。

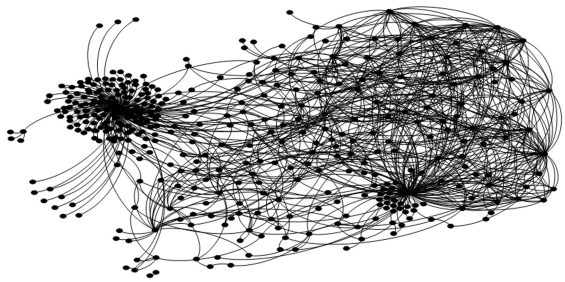

图 1 微博用户关系图

Fig. 1 Relationship diagram of Weibo user

3.2 MBUI-SFIM 算法

MBUI-SFIM 算法将微博用户影响力分成直接影响力和间接影响力两部分。用改进的 PageRank 算法计算用户直接影响力，根据微博用户间接相连路径不同计算用户间接影响力，最终得出用户总影响力。用户总影响力的计算式如式(1)所示：

$$I_{(u)}=\omega_1 DI_{(u)}+\omega_2 II_{(u)} \tag{1}$$

其中，u 为微博社交网络中某一特定用户，$DI_{(u)}$ 是用户 u 的直接影响力计算结果，$II_{(u)}$ 表示用户 u 的间接影响力计算结果，$I_{(u)}$ 为用户的总影响力计算结果。对参数 ω_1 和 ω_2 的约束如下：$\omega_1+\omega_2=1$，且 $\omega_1>\omega_2$。$DI_{(u)}$ 用户直接影响力和 $II_{(u)}$ 用户间接影响力的具体介绍如下。

3.2.1 MBUI-SFIM 用户直接影响力的分析

在微博社交网络中，用户直接影响力的定义为：若存在节点 a 和 b 直接相连，且节点 a 有能力改变或者影响节点 b，则称节点 a 对节点 b 有直接影响力。MBUI-SFIM 算法基于用户直接相连关系，考虑微博用户的自身因素与互动行为，分析并提出了用户直接影响力的计算方法。

(1)基于用户自身因素的直接影响力计算

在微博影响力分析中，用户的自身因素对于用户影响力的分析至关重要。在微博应用平台中，微博用户自身因素常

常体现在用户粉丝量、活跃度、近期微博质量与用户可信度4个方面。

1)微博粉丝数。微博粉丝数是直观体现用户影响力的一个重要因素。微博中一个微博用户拥有的粉丝数和该微博用户所对应的影响力是正相关关系,正如微博网络中出现的"富者越富"的现象:普通微博用户会选择关注名气大的微博用户;而对于刚加入微博的用户,微博平台会为其推荐名气大、粉丝多的用户。因此,粉丝数是直观衡量微博用户影响力的一个重要因素。

2)用户活跃度。用户活跃度是最近一段时间内用户的活跃程度。在微博社交网络中,主要选取用户评论量、转发量、点赞量与近期发布微博总数来衡量用户活跃度。综合这些条件,用户 u 的活跃度计算式如式(2)所示:

$$Act(u)=\frac{T_{count(u)}+T_{repost(u)}+T_{comment(u)}+T_{praise(u)}}{t} \tag{2}$$

其中,$Act(u)$表示用户 u 在时间 t 的活跃程度,$T_{count(u)}$ 表示用户 u 在 t 时间内发布的微博总数,$T_{repost(u)}$ 表示用户 u 在 t 时间内转发的微博总数,$T_{comment(u)}$ 表示用户 u 在 t 时间内的评论总数,$T_{praise(u)}$ 表示用户 u 在 t 时间内的点赞微博总数,t 表示一段时间。

3)近期微博质量。近期微博质量反映了用户近期微博被其他用户所认可的程度,该指标以近期微博被评论量、被转发量、被点赞量进行分析量化,其表达式如式(3)所示:

$$Q(u)=\frac{\sum_{i=1}^{N}F_i(u)}{T_{F(t)}}+\frac{\sum_{i=1}^{N}C_i(u)}{T_{C(t)}}+\frac{\sum_{i=1}^{N}P_i(u)}{T_{P(t)}} \tag{3}$$

其中,$Q(u)$是用户 u 的近期微博质量,$F_i(u)$是用户 u 的第 i 条博客被转发数,$C_i(u)$是用户 u 的第 i 条博客被评论数,$P_i(u)$是用户 u 的第 i 条博客点赞数,N 是用户 u 的近期微博总数,$T_{F(t)}$,$T_{C(t)}$,$T_{P(t)}$是 t 时间内所有用户总的微博转发数、评论数与点赞数。

4)用户信用度。信用度是用户在微博平台上的认证程度。新浪微博为用户认证提供了4种认证方式,对于不同种类的微博认证,本文给出计算式如式(4)所示:

$$Credit(u)=\begin{cases}0.5, & \text{自媒体认证用户}\\ 1, & \text{身份认证或自媒体用户}\\ 1.5, & \text{官方认证用户}\\ 0, & \text{其他无任何认证用户}\end{cases} \tag{4}$$

对微博用户自身要素进行分析,用户初始影响力的计算表达式如式(5)所示:

$$MUR_{initial}(u)=\alpha_1 Fans(u)+\alpha_2 Act(u)+\alpha_3 Q(u)+\alpha_4 Credit(u) \tag{5}$$

其中,$MUR_{inital}(u)$是用户的初始影响力,$Fans(u)$是用户 u 的粉丝数,$Act(u)$是用户 u 的活跃度,$Q(u)$是用户 u 的近期微博质量,$Credit(u)$是用户 u 的信用度。

(2)基于用户自身因素与互动行为的用户直接影响力计算

在社交网络中,用户互动行为是计算用户影响力的关键要素之一,从用户互动系数、微博用户可见率、影响强度与传播能力方面分析微博用户直接影响力,并基于改进的 PageRank 算法计算用户直接影响力。

1)互动系数。转发、评论、点赞行为是微博的主要互动行为。互动系数具体反映了两个用户之间的互动行为是否强烈,系数越大,用户之间的关系越紧密。互动系数如式(6)所示:

$$Interactive(u,v)=\alpha\cdot\frac{T_{repost(u,v)}}{T_{repost(v)}}+\beta\cdot\frac{T_{comment(u,v)}}{T_{comment(v)}}+\chi\cdot\frac{T_{praise(u,v)}}{T_{praise(v)}} \tag{6}$$

其中,$Interactive(u,v)$是粉丝用户 v 与其关注用户 u 的互动系数,$T_{repost(u,v)}$表示粉丝 v 转发用户 u 的微博数,$T_{Comment(u,v)}$表示粉丝 v 评论用户 u 的微博数,$T_{prase(u,v)}$表示粉丝 v 点赞用户 u 的微博数,$T_{repost(v)}$,$T_{comment(v)}$,$T_{prase(v)}$分别表示粉丝的转发微博、评论微博、点赞微博的总数。

2)微博可见率。微博信息可见率是指某个用户发表的内容被其他用户查看到的概率。在某段时间内,用户 u 的微博被其粉丝用户 v 浏览到的概率取决于用户 u 更新微博的频率与粉丝用户 v 的关注数,因此微博可见率的计算式如式(7)所示:

$$F_{see(u,v)}=\frac{T_{count(u)}}{t}*\frac{1}{T_{followers(v)}} \tag{7}$$

其中,$F_{see(u,v)}$表示微博用户 u 对于粉丝用户 v 的微博可见率,$T_{count(u)}$表示用户 u 在该段时间内发布的微博总数,t 表示一段时间,$followers(v)$表示粉丝用户 v 的关注用户,$T_{followers(v)}$表示粉丝用户 v 的关注总数。

3)影响强度。当用户浏览其他用户博客,并与其产生互动行为时,用户之间的影响力也在此时产生。由于用户交互与微博信息的可见率具有方向性,因此影响强度也具有方向性,本文定义微博用户影响强度如式(8)所示:

$$Affect(u,v)=F_{see}(u,v)\cdot Interactive(u,v) \tag{8}$$

其中,$Affect(u,v)$表示用户 u 对粉丝 v 的影响强度。

4)传播能力。对于用户 u 而言,用户 v 是用户 u 的粉丝用户之一。因此,本文定义用户 u 对于用户 v 的影响传播能力如式(9)所示:

$$Spread(u,v)=\frac{Affect(u,v)}{\sum_{i\in Followers(v)}Affect(i,v)} \tag{9}$$

其中,$Followers(v)$表示用户 v 的关注用户,i 为用户 v 所关注的用户之一,$Spread(u,v)$是具体量化结果。对于用户 v 所关注的用户,其影响力和传播能力之和为1,具体如式(10)所示:

$$\sum_{i\in Followers(v)}Spread(i,v)=1 \tag{10}$$

MBUI-SFIM 以用户自身因素与用户互动行为为依据计算用户直接影响力,最终将用户传播能力与用户初始影响力代入改进的 PageRank 算法中,用户直接影响力的计算式如式(11)所示:

$$DI_{(u)}=(1-d)+d\cdot\sum_{v\in Fans(u)}Spread(u,v)\cdot MUR_{initial}(v) \tag{11}$$

其中,$DI_{(u)}$表示用户 u 的直接影响力,$MUR_{inital}(v)$表示其粉丝用户 v 的初始影响力。

3.2.2 用户间接影响力计算

在微博用户影响力研究中,研究者往往把目光聚集于通过与用户直接相关联的用户互动信息、用户粉丝数等因素来计算用户影响力,而往往忽略掉用户间接影响力的计算。由图2可以看出,一个原创微博除了可以影响与它相连的第一层

微博博主,还可以通过用户互动行为影响第 2 层、第 3 层、…、第 n 层博主。因此,用户间接影响力计算也是非常重要的。MBUI-SFIM 算法根据用户间连接路径的不同,将用户间接影响力分成 3 种情况讨论。

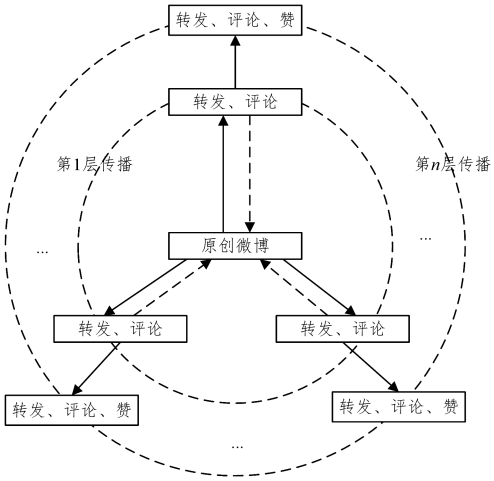

图 2 微博信息传播图

Fig. 2 Dissemination diagram of Weibo information

在微博中,用户 a 与用户 b 间接相连,根据用户 a 到用户 b 连接路径不同,本文将用户间接影响力分为简单路径、重复路径、复杂路径 3 种路径进行分析讨论,具体分析如下。

(1)简单路径。微博社交网络中,存在不相邻用户节点 A 与节点 D,A 到 D 的路径为 i_k 与 i_v,其中 $i_k=\{A,B,D\}$,$i_v=\{A,C,D\}$,i_k 与 i_v 的路径长度相同,且 $node(i_k)\bigcap node(i_v)=\{A,D\}$,简单路径如图 3 所示。

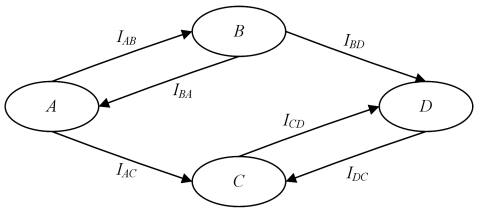

图 3 简单路径图

Fig. 3 Simple path diagram

MBUI-SFIM 算法对微博用户间接影响力简单路径的计算如式(12)所示:

$$\begin{aligned}IAffect(A,B)&=\max\{I_{AB}\times I_{BD},I_{Ac}\times I_{CD}\}\\&=\max\{DI_A\times DI_B,DI_A\times DI_C\}\end{aligned}\tag{12}$$

(2)重复节点路径。在微博社交网络中,微博用户 A 与微博用户 E 间接相连,且其间存在多个相交节点。定义用户 A 到用户 E 的路径集合为 I,$\exists\, i_k,i_v\in I$,i_k 路径为 $i_k=\{A,B,C,E\}$,i_v 路径为 $i_v=\{A,D,C,E\}$,i_k 与 i_v 的路径长度相同,且 $node(i_k)\bigcap node(i_v)=\{A,C,E\}$,重复节点路径如图 4 所示。

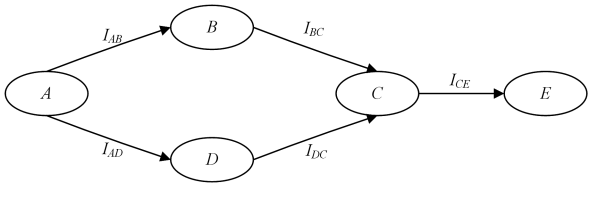

图 4 重复节点路径图

Fig. 4 Repeated node path diagram

MBUI-SFIM 算法对微博社交网络中重复节点路径间接影响力的计算如式(13)所示:

$$\begin{aligned}IAffect(A,E)&=\max\{I_{AB}*I_{BC},I_{AD}*I_{DC}\}*I_{CE}\\&=\max\{DI_A*DI_B,DI_A*DI_D\}*DI_C\end{aligned}\tag{13}$$

(3)复杂路径。复杂路径是指微博用户 A 到用户 E 的间接相连路径比较复杂。在复杂路径中,$\exists\, i_k,i_v,i_j\in I$,$i_k$ 和 i_j 的路径长度相同,其中,$i_k=\{A,B,C,E\}$,$i_j=\{A,G,H,E\}$,i_k 与 i_j 路径长度不等于 i_v,$i_v=\{A,D,E\}$,$node(i_k)\bigcap node(i_v)\bigcap node(i_j)=\{A,E\}$。复杂节点路径如图 5 所示。

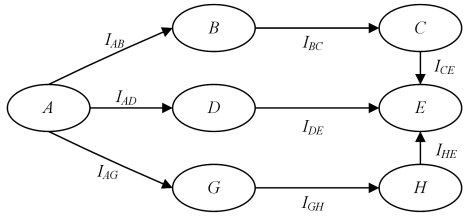

图 5 复杂路径图

Fig. 5 Complex path diagram

MBUI-SFIM 微博社交网络中复杂节点路径间接影响力的计算式如式(14)所示:

$$\begin{aligned}IAffect(A,E)&=\max\{I_{AD}*I_{DE},\max\{I_{AB}*I_{BC}*I_{CE},\\&\quad I_{AG}*I_{GH}*I_{HE}\}\}\\&=\max\{DI_A*DI_D,\max\{DI_A*DI_B*\\&\quad DI_C,DI_A*DI_G*DI_H\}\}\end{aligned}\tag{14}$$

本文通过用户关注关系对微博社交网络进行建模,在间接影响力计算方面,基于用户直接影响力计算结果,并根据用户连接路径不同,对微博用户间接影响进行分类处理,最终得出微博用户间接影响力的计算式如式(15)所示:

$$II_{(u)}=\frac{\sum\limits_{v\in node(u)}IAffect(u,v)}{T_{node(u)}}\tag{15}$$

其中,$node(u)$是与用户 u 间接相连的节点,$T_{node(u)}$是与用户 u 间接相连的节点总数,$II_{(u)}$是用户 u 的间接影响力。

3.3 算法性能分析

MBUI-SFIM 算法考虑了微博用户自身因素与互动行为,如用户的粉丝数、用户活跃度、用户信用度、用户互动系数、微博可见率等,利用改进的 PageRank 算法计算微博用户的直接影响力,并根据不相邻节点连接路径不同计算了用户的间接影响力。该算法对微博用户影响力计算方面的分析更加全面,符合实际,能更准确地衡量微博用户的实际影响力。

由于 MBUI-SFIM 算法对微博用户影响力计算方面的分析更加全面,需要增加计算近期微博质量 $Q(u)$、微博初始影响力 $MUR_{inital}(u)$,以及用户间接影响力 $II_{(u)}$ 的值,与经典的 PageRank 算法相比,本文算法的计算时间复杂度略有增加,但更加准确地计算出了用户影响力,这个代价是值得的。

4 实验结果分析与讨论

4.1 实验数据采集

本文在获取用户数据的过程中,选取新浪微博作为数据来源,由于从新浪微博 API 接口获取数据的限制较多,因此本文选用 Python 技术[16]从新浪移动端爬取数据信息。在爬取过程中,随机选定一个中心用户,从该用户出发,逐层爬取

用户信息，爬取的数据信息如下：

(1)用户 ID，用户姓名，用户关注数，粉丝数，微博总数量；

(2)用户关注信息，主要有用户关注 ID、用户 ID、用户粉丝 ID；

(3)用户微博信息，用户 ID，用户发微博内容，时间，评论次数，转发次数，点赞次数，微博是否原创；

(4)用户互动信息，其中包括用户微博 ID，转发者 ID，评论者 ID，评论内容，点赞者 ID。

由于微博爬虫爬取到的是近六个月的微博数据，旧的微博数据体现不出微博用户当前的影响力，因此在数据选用方面，本文采用近 3 个月的数据作为实验数据。对于微博数据中存在僵尸粉的现象，本文根据上文提出的用户活跃度定义做了剔除僵尸粉操作。

4.2 实验结果与分析

本文通过与文献[5]提出的 MR-UIRank 和经典的PageRank 算法进行对比，从用户影响力排名准确性与影响覆盖范围两个方面对 MBUI-SFIM 算法进行了性能分析。

(1)用户影响力排名准确性对比

使用 MBUI-SFIM 算法得到影响力前 10 名的用户信息，如表 1 所列，表中包含用户粉丝数、用户互动量与具体的用户微博数量。对表 1 中影响力排名前十的用户进行分析，前 10 名用户均为新浪微博会员用户，且用户等级相对较高。用户微博包含大量的与商品营销、舆论引导相关的内容，以“无为有为”“澄净如歌”用户为例，在 2018 年金马奖“台独”风波中，他们发布了关于“中国一点也不能少”的中国地图，表达出了自己坚定的爱国立场。同时，该微博内容在微博网络中，通过粉丝一层层传播，其微博点赞、评论、转发量均相对较高。该博主在自己的朋友圈中发挥着引导舆论走向的作用。因此，从用户影响力应用方面进行分析，本文得出的影响力排名前十的用户排名比较合理。从表 1 也可以看出，粉丝数大的用户的影响力并不一定高，MBUI-SFIM 算法得出的用户影响力高低取决于对微博互动行为与自身因素的综合分析。

表 1 用户影响力前十名的具体信息表

Table 1 Top 10 specific information tables for user impact

用户名	互动量	粉丝数	微博数
ynkhsya	14594	264	3266
澄净如歌	3791	178	1268
506070 后聚集地	12966	588	141
安晶的摩羯座	7692	1050	69554
蘇宅火火貓	2114	156	11237
无为尚有为	64784	948	1206
毓轩时	1030	189	6322
慕容小歌	62485	382	1590
云在青天水在瓶	2654	735	5296
古月听雨 920	14488	538	7608

微博用户互动量和用户微博近期质量分别如图 6、图 7 所示，用户互动具体包括用户微博被转发、被点赞、被评论行为。在微博社交网络中，当用户微博被其他用户所认可时，就会发生互动行为，因此互动量最能够具体体现用户影响力。而互动量与用户影响力成正相关关系，用户影响力排名越靠前，用户互动量就越高，因此互动量最能体现用户影响力排名的准确性。近期微博质量能够表现用户近期所发微博是否被其他用户所认可，通常影响力大的用户近期微博质量较高，因此近期微博质量也能体现用户影响力。图 6 中，纵坐标为微博用户互动数量，横坐标为影响力排名 Top-k 的用户。图 7 中，纵坐标为用户近期微博质量值，横坐标为影响力排名 Top-k 的用户。从图 6 中可以看出，MBUI-SFIM 算法与 MR-UIRank 算法得到的微博用户排名前十的用户微博互动量与近期微博质量相差不大，排名前十的用户相似度很高。图 6 中，当横坐标分别为 20 和 25 时，相比于 PageRank 算法和 MR-UIRank 算法的计算准确性分别提高了 14.8%和 8.3%。图 7 中，MBUI-SFIM 算出的 Top-k 个用户的近期微博质量高于 MR-UIRank 与 PageRank 算法所算出的结果。这是因为 MBUI-SFIM 算法在用户影响力计算过程中综合考虑了用户自身因素与用户互动行为，因此结果值比 MR-UIRank 与 PageRank 算法的结果值高，此结果也表明了 MBUI-SFIM 算法用户影响力的准确性更高。

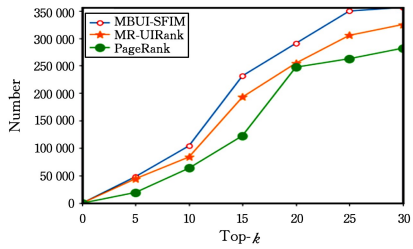

图 6 用户微博互动量

Fig. 6 Interaction diagram of user Weibo

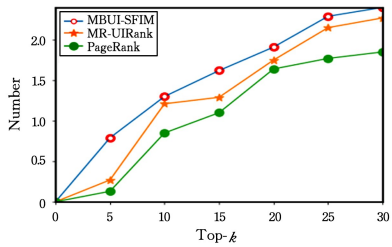

图 7 用户微博近期质量

Fig. 7 Recent quality diagram of user Weibo

(2)影响覆盖范围对比

本文采用独立级联模型(Independent Cascade Model IC)[17-18]来测试微博用户影响力的传播覆盖范围。在微博应用平台中，当微博用户 i 发布一条微博信息时，该条微博信息被用户 j 浏览到，并且用户 j 对于该微博信息做出了互动行为(如转发，点赞，评论)，互动行为意味着该条微博对粉丝产生了影响。这种传播行式符合独立级联模型(IC)中的激活行为，因此本文选用 IC 级联模型做覆盖范围对比。

图 8 为影响范围覆盖对比图，选取影响力排名 Top-k 的用户作为种子节点用户，剩余用户为待激活用户。图 8 中，横坐标代表影响力排名 Top-k 的种子节点用户，纵坐标代表被影响的用户数量。从图可得，MBUI-SFIM 得出的影响力大的用户，在 IC 独立级联模型中的影响范围大于 MR-UIRank 和 PageRank 算法得出的影响力最大用户的影响范围，且种子节点在 500 左右范围时，600 个用户均被影响，而 MR-UIRank 算法与 PageRank 算法得出的种子节点为 550 左右时，节点集合才均被影响。当所有用户均被影响时，通过 MBUI-SFIM 算法得到的种子节点集合人均影响力范围大约

为 1.2，而 MR-UIRank 算法与 PageRank 算法得到的种子节点集合人均影响力范围大约为 1.09。这表明 MBUI-SFIM 算法得出的影响力大的用户影响范围广且影响人数增长速度快。

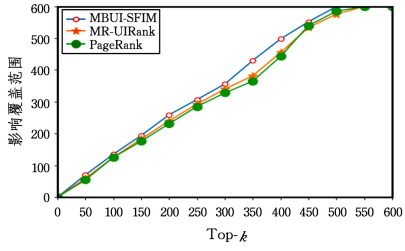

图 8 影响范围覆盖对比图

Fig. 8 Comparison diagram of impact range coverage

结束语 用户影响力已经成为当前社交网络中的一个热点研究问题，微博作为一种获取信息快速、准确的平台出现，方便了人们的生活，并且已经成为人们获取信息的主要方式之一。本文提出的 MBUI-SFIM 算法在微博用户影响力计算的过程中，从直接影响力与间接影响力两个方面考虑，在直接影响力计算中，通过量化用户自身因素与用户行为因素，使用改进的 PageRank 算法得到用户直接影响力计算结果，在计算用户间接影响力时，根据不相邻节点连接路径不同，将用户间接影响力具体分成 3 类进行处理，最终得到用户间接影响力。最后将本文提出的 MBUI-SFIM 算法在覆盖率和准确率两方面与 MR-UIRank 算法和经典的 PageRank 算法进行比较，验证本文算法的有效性。

高影响力用户的挖掘在微博网络平台中能够引导社会舆论的走向，帮助商家进行商品营销。因此，通过 MBUI-SFIM 算法挖掘出的微博高影响力用户，在微博网络平台中能发挥重要实际应用。但是，本文并未对微博用户积极影响与消极影响做处理，因此在后续工作中将对此进行进一步的研究。

参 考 文 献

[1] BINGOL K, ERAVCI B, ETEMOGLU C O, et al. Topic-based influence computation in social networks under resource constraints[J]. IEEE Transactions on Services Computing, 2016, 12(6): 970-986.

[2] LIU Q, XIANG B, YUAN N J, et al. An influence propagation view of pagerank[J]. ACM Transactions on Knowledge Discovery from Data (TKDD), 2017, 11(3): 1-30.

[3] PENG S, YANG A, CAO L, et al. Social influence modeling using information theory in mobile social networks[J]. Information Sciences, 2017, 379: 146-159.

[4] WENG J, LIM E P, JIANG J, et al. Twitterrank: finding topic-sensitive influential twitterers[C]// Proceedings of the Third ACM International Conference on Web Search and Data Mining. ACM, 2010: 261-270.

[5] SUN H, ZUO T. Research and Realization of Influence Optimization in Cloud Computing Environment[J]. Journal of Chinese Computer Systems, 2018, 39(1): 42-47.

[6] PAGE L, BRIN S, MOTWANI R, et al. The PageRank citation ranking: Bringing order to the Web[J]. Stanford Digital Libraries Working Paper, 1998, 9(1): 1-14.

[7] YAMAGUCHI Y, TAKAHASHI T, AMAGASA T, et al. Turank: Twitter user ranking based on user-tweet graph analysis[C]// International Conference on Web Information Systems Engineering. Springer, Berlin, Heidelberg, 2010: 240-253.

[8] MEEYOUNG C, HAMED H, FABRICIO B, et al. Measuring user influence in twitter: The million follower fallacy[C]// Fourth International AAAI Conference on Weblogs and Social Media. Menlo Park: AAAIPress, 2010: 10-17.

[9] LI G L, CHU Y P, FENG J H, et al. Influence maximization on multiple social networks [J]. Chinese Journal of Computers, 2016, 39: 643-656.

[10] LIU L, TANG J, HAN J, et al. Mining topic-level influence in heterogeneous networks[C]// Proceedings of the 19th ACM International Conference on Information and Knowledge Management. ACM, 2010: 199-208.

[11] MAO J X, LIU Y Q, ZHANG M, et al. Social influence analysis for micro-blog user based on user behavior[J]. Chinese Journal of Computers, 2014, 37(4): 791-798.

[12] ZHOU J, ZHANG Y, WANG B, et al. Predicting user influence in microblogs[C]// 2016 First IEEE International Conference on Computer Communication and the Internet (ICCCI). IEEE, 2016: 292-295.

[13] WU L, YANG B, JIAN M, et al. MPPR: Multi Perspective Page Rank for User Influence Estimation[C]// 2018 IEEE International Conference on Big Data and Smart Computing (BigComp). IEEE, 2018: 25-29.

[14] BAKSHY E, HOFMAN J M, MASON W A, et al. Everyone's an influencer: quantifying influence on twitter[C]// Proceedings of the Fourth ACM International Conference on Web Search and Data Mining. ACM, 2011: 65-74.

[15] LI C, XIONG F. Social recommendation with multiple influence from direct user interactions[J]. IEEE Access, 2017, 5: 16288-16296.

[16] YANG L U, HUAKANG L I, GUOZI S. Distributed microblog crawler system based on P2P[J]. Journal of Jiangsu University, 2016, 37(3): 296-301.

[17] WEN Z, KVETON B, VALKO M, et al. Online influence maximization under independent cascade model with semi-bandit feedback[C]// Advances in Neural Information Processing Systems. 2017: 3022-3032.

[18] KEMPE D, KLEINBERG J, TARDOS É. Maximizing the spread of influence through a social network[C]// Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2003: 137-146.

WANG Xin-sheng, born in 1972, Ph.D, associate professor, is member of China Computer Federation (CCF). His main research interests include wireless sensor network and social network.