

基于 HITS 算法的微博用户可信度评估

吴树芳^{1 2} 徐建民^{3*}

(1. 天津大学管理与经济学部, 天津 300072; 2. 河北大学管理学院, 河北 保定 071000;

3. 河北大学计算机科学与技术学院, 河北 保定 071000)

摘要: 以新浪微博为研究平台, 在 HITS(hyperlink-induced topic search) 算法的基础上, 提出融合用户交互行为和博文内容的微博用户可信度评估算法。分别构建基于交互行为和基于博文内容的微博用户有向链接图, 图中节点表示用户, 有向边体现用户基于交互行为或基于内容的指向关系; 依据 HITS 算法计算两种拓扑结构下微博用户的权威度和中心度; 以融合的权威度作为度量评估用户可信度。试验采用从新浪微博采集的数据作为测试集合, 通过反复训练法获得可信度阈值, 绘制不同可信度算法的用户可信度曲线, 验证了算法的可行性和有效性。

关键词: HITS 算法; 微博用户; 可信度; 交互行为; 博文

中图分类号: TP391 **文献标志码:** A

Evaluation of microblog users' credibility based on HITS algorithm

WU Shufang^{1 2}, XU Jianmin^{3*}

(1. College of Management and Economics, Tianjin University, Tianjin 300072, China;

2. College of Management, Hebei University, Baoding 071000, Hebei, China;

3. School of Computer Science and Technology, Hebei University, Baoding 071000, Hebei, China)

Abstract: Based on Sina-Microblog and HITS(hyperlink-induced topic search) algorithm, a new user's credibility algorithm that merged user interactions and blog contents was putted forward. The new algorithm firstly constructed two directed connection graphs based on user interactions and blog contents respectively, where nodes represented users and arcs embodied the direction relationship between users. Authority and hub of these two connected graphs was computed. The fusion authority was adopted as measurement to evaluate user's credibility. The data collected from Sina-Microblog as test set was used to conduct experiments. Threshold of credibility was obtained by repeated training, and then credibility curves of different algorithms were drawn to verify the feasibility and effectiveness of the new algorithm.

Key words: HITS algorithm; microblog users; credibility; interaction; blog

0 引言

随着互联网的发展, 出现了各种社交网络平台, 如英文的 Twitter、Facebook, 中文的新浪微博和腾讯微博等。这些社交平台已经渗透到人们生活的各个方面, 成为人们获取信息的主要渠道之一。据统计, 2015 年 6 月我国微博用户已达 6.68 亿, 截止 2015 -

09-30, 微博月活跃用户数为 2.12 亿人, 有超过 62% 的微博用户处于非活跃状态。如何辨别其中的虚假用户, 是搭建一个诚信、公平、健康的社交平台亟需解决的问题之一^[1-2]。尽管新浪微博社区管理中心发布了一系列微博用户信誉约束规则, 但这些规则仍然存在一些缺陷, 导致微博用户可信度排名不合理^[3]。解决微博用户的可信度问题不仅可以帮助人们找到真实可靠的微博信息, 也有利于发现领域专家, 某个

收稿日期: 2016-03-31; 网络出版时间: 2016-09-18 22:48:39

网络出版地址: <http://www.cnki.net/kcms/detail/37.1391.T.20160918.2248.002.html>

基金项目: 河北省社会科学基金资助项目(HB15TQ013)

作者简介: 吴树芳(1979—), 女, 河北邯郸人, 副教授, 博士, 主要研究方向为信息检索与不确定信息处理。E-mail: shufang_44@126.com

* 通讯作者: 徐建民(1966—), 男, 河北邯郸人, 教授, 博士, 主要研究方向为信息检索与不确定信息处理。E-mail: hbuxjm@hbu.edu.cn

领域的专家一般是微博用户中的高可信度用户^[4-5]。现有的微博用户可信度多依据微博注册用户信息的完整性和社会地位来评估^[6],事实上,注册信息完整或社会地位高的微博用户其可信度不一定高,故仅以此来判断用户可信度存在不足。

目前对微博用户可信度的研究尚处于起步阶段,研究内容较少。CASTILLO C 等人提出了话题可信度评估方法,首先提取和话题相关的 Tweets 用户信息、Tweets 传播的特征和 Tweets 内容的特征,然后采用决策树估计话题的可信度,该项研究针对的是内容可信度,并没对微博用户可信度进行深入研究^[7]。闫光辉等人在 PageRank 算法的基础上,参考网络群体结构平衡理论,结合微博用户本身的属性因素,提出了基于链接的微博用户可信度评估模型,该模型忽略了微博用户所发博文对其可信度的影响^[8]。GUPTA M 等人提出一个用户可信度评估算法,该算法依据用户所发博文的可信度均值获得用户的可信度,忽略了微博用户本身属性及用户行为对其可信度的影响^[9]。MUKHERJEE A 等人提出 GS-Rank 算法用于检查欺骗群,即一群用户互相协作,进行系列虚假评论^[10]。CHU Z 等人将信息熵与机器学习方法结合,提出了微博虚假用户识别算法,取得了一定成效^[11]。徐建民等人通过对微博用户行为进行分析,运用逻辑回归算法,提出了一个基于逻辑回归的用户可信度评估模型^[12]。

上述研究内容在对微博用户可信度进行评估时,或者基于微博用户本身属性、行为,或者基于用户所发博文内容。事实上,二者对微博用户可信度的评估贡献度不分上下,如果能将二者有效融合,则将对微博用户可信度的评估更为合理。本文在 HITS 算法^[13]的基础上分别获得基于用户交互行为^[14]和基于内容的微博用户权威度和中心度,并通过线性调和的方法将二者融合,获得每个微博用户的最终权威度/中心度,实现对微博用户可信度评估。试验采用从新浪微博获取的用户信息验证文章提出的评估方法的可行性和有效性。

1 HITS 算法

HITS 算法是文献[15]提出的一个网页重要性评判算法,是链接分析中非常基础且重要的算法,目前已广泛应用于自然语言处理相关领域。该算法采用中心度(Hub)和权威度(Authority)这两个重要的指标来计算检索到的网页的重要性。一个网页的权威度高,说明该网页是与某个主题相关的高质量网

页,例如与“搜索引擎”主题相关的网页中,谷歌和百度属于权威度高的网页。一个网页的中心度高,说明该网页包含了很多指向高权威度网页的链接,例如中文语言资源联盟(Chinese Linguistic Data Consortium, CLDC)页面中包含一些具有权威性的开源软件和中文语料库,并通过链接的方式指向了这些资源,则这个页面可以认为是“自然语言处理”主题下的一个高中心度网页。HITS 算法中隐含有以下两个假设:(1) 一个高权威度网页会被很多高中心度的网页指向;(2) 一个高中心度的网页会指向很多高权威度的网页。

依据上述假设,可知中心度和权威度两个指标是相互增强的关系^[16],一个页面的重要性通过权威度和中心度两个值进行度量。假设 $A(i)$ 表示网页 i 的权威值, $H(i)$ 代表网页 i 的中心度,图1所示的网页集合中,有1个网页有链接指向页面1,同时页面1通过3个链接指向其它3个页面,则页面1的权威度依据指向其节点的网页的中心度计算,即: $A(1) = H(1) + H(2) + H(3)$,页面1的中心度通过它指向的节点的权威度计算,即 $H(1) = A(1) + A(2) + A(3)$ 。综上可知,节点中心度和权威度的计算公式为:

$$A(i) = \sum_{j=1}^n H(j), \quad (1)$$

$$H(i) = \sum_{k=1}^m A(k), \quad (2)$$

式中 $H(j)$ 为指向节点 i 的节点的中心度; $A(k)$ 为节点 i 指向的节点的权威度。式(1)、(2)可看出:中心度和权威度是相互增强、不断迭代的关系。

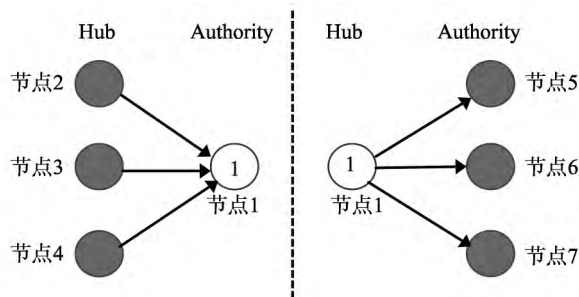


图1 网页中心度、权威度

Fig. 1 Hub and authority of web pages

在微博中,一个虚假的微博用户常常通过人为增加自己的关注量来提高影响力^[17]。结合 HITS 算法,从微博用户关注行为的角度看,人为增加关注量仅可能增加虚假用户的权威度,无法增加其中心度,故依据中心度和权威度两个指标来评估微博用户的可信度具有可行性。此外,对微博用户可信度的评估,除依据微博用户本身的特性外,还可依据微博用

户所发博文。虚假微博用户所发博文内容一般是机械式的,固定格式且没有实质内容,故其他微博用户对该博文几乎没有交互行为(关注、转发、评论、赞)。反之,如果微博用户 $User(i)$ 所发博文 $Blog(i)$ 被多个高中心度的博文引用,或者该博文引用过一些高权威度的博文,则博文 $Blog(i)$ 的中心度和权威度就高,相应地 $User(i)$ 的可信度就高。

本研究在 HITS 算法的基础上,从微博用户的交互行为和其所发博文两个角度出发,对微博用户的可信度进行评估,提出一个新的微博用户可信度评估算法 HITS-UC。

2 微博用户可信度评估

2.1 基于交互行为的用户权威度/中心度

在微博网络中,若不涉及博文内容,微博用户的交互行为主要包括关注、收藏、转发、评论、赞,由于收藏数无法获得,所以本研究在计算交互度时不考虑收藏行为。HITS 算法的核心原理是依据网页的链入和链出单向边体现网页间的引用关系,并以此计算每个网页的中心度和权威度,得出网页重要性排序。在微博网络中,如果任意 2 个微博用户间有某种交互行为,则认为二者存在单向的指向关系,如图 2 所示。从图 2 中可以看出,存在自 $User(a)$ 指向 $User(f)$ 的边,则二者之间存在交互行为。依据 HITS 算法,则可计算出图中所示环境下 $User(a)$ 的权威度 $A_{User(a)}$ 和中心度 $H_{User(a)}$ 。

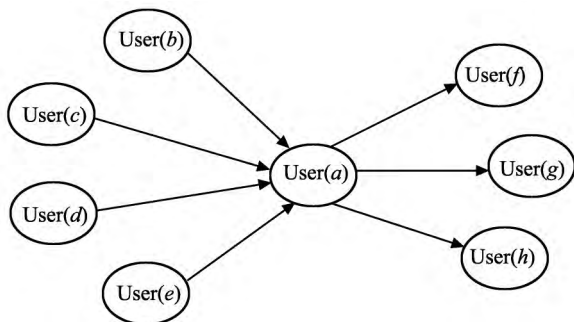


图 2 基于交互行为的微博用户链接图

Fig. 2 Micro-blog users link graph based on interactions

上述思想存在以下缺点:若任何 2 个微博用户 $User(i)$ 和 $User(j)$ 之间有任何一种交互行为,就绘制一条有向边,导致最终绘制的拓扑图中有很多无效边,从而使得用户权威度和中心度不合理,故论文做出如下限制:首先计算 $User(i)$ 对 $User(j)$ 的交互度 $Idegree(User(i) \rightarrow User(j))$,其计算方法如公式(3)所示,如果其值大于指定的阈值 ε ,则绘制二者之间的有向边,否则,认为 $User(i)$ 对 $User(j)$ 的交

互比较微弱,可忽略。

$$Idegree(User(i) \rightarrow User(j)) =$$

$$\alpha_1 \omega_{Follow} + \alpha_2 (\omega_{Retweet} + \omega_{Comment} + \omega_{Praise}), \quad (3)$$

式中 ω_{Follow} 、 $\omega_{Retweet}$ 、 $\omega_{Comment}$ 、 ω_{Praise} 分别为 $User(i)$ 对 $User(j)$ 关注、转发、评论、赞的权重。为了将交互度 $Idegree$ 的值限定在 $0 \sim 1$,假定 $\alpha_1 + \alpha_2 = 1$,参数 α_1 , α_2 用于调节关注和其余 3 种(评论、转发、赞)交互行为对 $Idegree(User(i) \rightarrow User(j))$ 的贡献度。在微博中,用户评论、转发、赞的微博一般来自于其关注的微博用户,经验上,关注和其余 3 种行为影响的和对最终交互度的影响相当,故假定 $\alpha_1 = \alpha_2 = 0.5$,在未来的研究中将对不同交互行为对交互度的影响程度做进一步的深入研究。关注行为的权重 ω_{Follow} 只有 2 个取值,要么关注,要么不关注,故:

$$\omega_{Follow} = \begin{cases} 1, & \text{如果 } User(i) \text{ 关注 } User(j); \\ 0, & \text{否则。} \end{cases} \quad (4)$$

观察、分析新浪微博可以发现,其余 3 种交互行为的难易程度为赞 < 转发 < 评论,从而导致微博中赞的个数大于其余 2 种,转发数大于评论数。理论上,三者对交互度的贡献程度为评论最大,转发次之,赞最小。如果直接依据交互行为的交互次数计算各自的权重,则赞的权重会远远大于其余两种,这显然与实际不符。此外,为了保证 $\omega_{Retweet}$ 、 $\omega_{Comment}$ 、 ω_{Praise} 的和不超过 1,故将分母固定为 $2N$,最终给出下述的交互行为权重计算方法:

$$\omega_{Retweet} = \frac{N - \text{Num}_{retweet}}{2N}, \quad (5)$$

$$\omega_{Comment} = \frac{N - \text{Num}_{comment}}{2N}, \quad (6)$$

$$\omega_{Praise} = \frac{N - \text{Num}_{praise}}{2N}, \quad (7)$$

式中, $N = \text{Num}_{Retweet} + \text{Num}_{Comment} + \text{Num}_{Praise}$,其中 Num_i 为对应交互行为的交互次数。

综上,依据微博用户间的有向交互度,可以得到微博用户基于交互行为的有向交互图(如图 2 所示),并依据式(1)、(2)可计算出微博用户 $User(i)$ 基于交互行为的权威度 $A_1(User(i))$ 和中心度 $H_1(User(i))$ 。

2.2 基于博文内容的用户权威度/中心度

微博属于非结构化的短文本,长度不可以超过 140 字。微博中除了有文本、图片、视频外,还有 2 个很特殊的符号“#”和“@”。“#”表示微博所属的话题,“@”是在博文内容中体现的用户交互关系。如在新浪微博中,用户“微相册”(U_1)发表了一篇博文,博文中“#helo 一月#”表示该博文属于话题“hel-

lo 一月”“@ 微博相机”(U_3)、“@ 微笑的阿榕”(U_4)表明该微博用户和用户“微博相机”和“微笑的阿榕”有交互关系。本节将基于博文内容中这 2 个典型特征 绘制微博用户间的有向链接关系图 进而计算基于内容的用户权威度和中心度。详细规则为: 如果用户 $User(i)$ 和 $User(j)$ 在博文的“#”里提到了相同的标题, 则同时存在 $User(i) \rightarrow User(j)$ 和 $User(j) \rightarrow User(i)$ 的有向边, 如果 $User(i)$ 在博文里“@”了 $User(j)$, 则仅存在有向边 $User(i) \rightarrow User(j)$ 。依据该规则, 针对新浪微博中用户“微相册”(U_1)和“Limm”(U_2)发表的 2 篇博文, 可以绘制出图 3 基于博文内容的用户链接关系图。在图 3 所示拓扑结构的基础上, 依据式 (1)、(2) 可以得到微博用户 $User(i)$ 的基于博文内容的权威度 $A_c(User(i))$ 和中心度 $H_c(User(i))$ 。

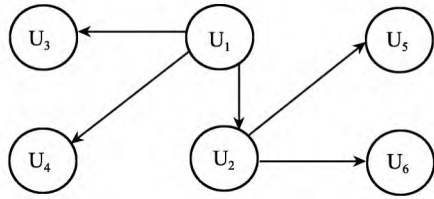


图 3 基于内容的用户链接图

Fig. 3 Users link graph based on blog

2.3 微博用户可信度评估算法 HITS-UC

在上文中, 依据用户交互关系和博文内容绘制了两类有向链接关系图, 最终计算可信度时将二者融合, 即对于任意 2 个微博用户 $User(i)$ 和 $User(j)$, 如果二者存在满足阈值的交互关系和内容关系中的一种, 则二者之间存在有向边。融合微博用户交互行为和博文内容的用户权威度、中心度

$$A(User(i)) = \frac{A_1(User(i)) + A_c(User(i))}{2}, \quad (8)$$

$$H(User(i)) = \frac{H_1(User(i)) + H_c(User(i))}{2}, \quad (9)$$

依据式 (8) (9) 计算出待评微博用户的中心度和权威度后, 按照权威度降序对微博用户进行排序, 若权威度小于阈值 δ , 则认为该微博用户的可信度比较低, 属于虚假用户。基于 HITS 的用户可信度评估算法 (HITS-UC) 实现过程为:

输入: 热门榜单中的微博用户;

输出: 拓展后的可信微博用户;

实现过程:

(1) 运用随机函数 $random()$ 从微博热门榜单中随机选出几个权威度较大的微博用户节点作为根集合 S ;

(2) 依据用户间的交互行为和交互度阈值 ε , 通

过向 S 中加入被 S 中节点指向的微博用户和指向 S 中节点的微博用户来拓展集合 S , 将 S 拓展成一个更大的集合 T_1 , 节点间以有向边相连, 形成边集 E_1 ;

(3) 初始化 T_1 中节点的权威度、中心度为: $A_1(User(i)) = H_1(User(i)) = 1 (1 \leq i \leq n)$;

(4) 重复执行下面的操作, 直至集合 T_1 中的所有微博用户 $User(i) (1 \leq i \leq n)$ 的权威度值、中心度值趋于稳定:

$$A_1(User(i)) = \sum H_1(User(k)) \text{ 规范化为}$$

$$A_1(User(i)) = A_1(User(i)) / (\sum \frac{A_1(s)}{2});$$

$$H_1(User(i)) = \sum A_1(User(v)) \text{ 规范化为}$$

$$H_1(User(i)) = H_1(User(i)) / (\sum \frac{H(s)}{2});$$

(5) 依据用户所发博文内容, 通过向 S 中加入被 S 中节点指向的微博用户和指向 S 中节点的微博用户来拓展集合 S , 将 S 拓展成一个更大的集合 T_2 , 节点间以有向边相连, 形成边集 E_2 ;

(6) 针对 T_2 重复步骤 (3) (4), 得到微博用户基于博文内容的权威度 $A_c(User(i))$, 中心度 $H_c(User(i))$ 。

(7) 将步骤 (4) (6) 得到的结果线性调和, 得到用户最终的权威度 $A(User(i))$, 中心度 $H(User(i))$ 。

(8) 依据权威度评估微博用户的可信度, 对其进行降序排序;

(9) 如果微博用户的权威度小于阈值 δ , 则将其视为虚假用户。

3 试验

3.1 测试集合

由于目前没有统一的、权威的微博数据集供使用, 因此本文通过新浪微博提供的 API 接口和爬虫工具, 随机爬取了 16 487 名用户、165 859 条微博, 以及他们之间关注、评论、转发、赞的数据, 经人工辨别后发现其中有 9 652 个真实用户和 6 835 个虚假用户, 最终将这些数据作为试验测试集合。

3.2 评测标准

评估一个微博用户是否为可信用用户, 实际是二分类问题, 其性能可采用查全率 P_{recall} 、查准率 $P_{precision}$ 和 $F_{measure}$ 来衡量, 计算公式为:

$$P_{recall} = \frac{a}{a+b}, \quad (10)$$

$$P_{precision} = \frac{a}{a+c}, \quad (11)$$

式中 a 表示命中的正样本数; b 表示遗漏的正样本

数; c 表示计算出为正样本实际上是负样本的数量。查全率 P_{recall} 和查准率 $P_{\text{precision}}$ 有时候会出现矛盾, 这样就需要将二者归一化, 最常见的方法就是将 P_{recall} 和 $P_{\text{precision}}$ 加权调和平均, 得到 F_{measure} (又称为 F_{score}) 评测指标, F_{measure} 值越高性能越好, 计算方法如下:

$$F_{\text{measure}} = \frac{(1 + \beta^2) P_{\text{recall}} P_{\text{precision}}}{\beta^2 (P_{\text{recall}} + P_{\text{precision}})}, \quad (12)$$

当参数 $\beta = 1$ 时, 就是最常见的 F_1 , 即

$$F_1 = \frac{2 P_{\text{recall}} P_{\text{precision}}}{P_{\text{recall}} + P_{\text{precision}}}. \quad (13)$$

3.3 试验结果

对于交互度阈值 ε , 观察式 (3) 与 (4) 可以发现, 只要微博用户 $\text{User}(i)$ 关注了 $\text{User}(j)$, 则 $\text{Idegree}(\text{User}(i) \rightarrow \text{User}(j)) \geq 0.5$, 故将 ε 设置为 0.5。对于微博用户可信度阈值 δ , 采用反复训练的方法获得其较优值。将试验数据中的 9 652 个真实用户和 6 835 个虚假用户无规律混合, 然后运用本文章提出的算法计算 $\delta = 0.1, 0.2, 0.3, \dots, 0.9, 1.0$ 时的性能指标, 结果如表 1 所示。观察表 1 可以发现, 当 $\delta = 0.6$ 时, F_1 最高, 故将 0.6 作为算法中的可信度阈值 δ 的较优取值。如果想得到更为合理的 δ 值, 可以进一步细化训练 0.55 ~ 0.65 范围内的 F_1 值。当 $\delta = 0.6$ 时, 依据 HITS-UC 算法, 正确判定了测试集中 14 962 名微博用户, 准确率为 0.907 5, 后续实验中将采用 0.6 作为微博用户可信度阈值。

表 1 不同阈值 δ 下的 F_1 值比较
Table 1 The comparison of F_1 value based on different threshold δ

δ	Recall	Precision	F_1
0.1	1.000 0	0.585 4	0.738 5
0.2	0.935 7	0.598 2	0.729 8
0.3	0.905 3	0.689 4	0.782 7
0.4	0.783 2	0.855 6	0.817 8
0.5	0.796 1	0.863 9	0.828 6
0.6	0.794 5	0.907 5	0.847 2
0.7	0.756 1	0.913 5	0.827 3
0.8	0.702 1	0.917 4	0.795 4
0.9	0.653 1	0.958 1	0.776 7
1.0	0.518 1	0.962 4	0.673 6

为了验证本研究提出的微博用户可信度评估算法的可行性, 从测试集中随机抽取 100 名微博用户, 做了 2 项实验。第一项试验用于比较本研究提出的单独依据基于交互行为 (HITS-U)、基于博文内容 (HITS-C) 和二者融合的 HITS-UC 算法的性能, 图 4 为该试验的结果, 即三者在不同查全率下的查准率曲线, 曲线距离坐标原点越远, 其性能越优。

观察图 4 可以发现, 融合交互行为和博文内容的 HITS-UC 算法性能较优, 产生该结果的原因是因为该算法有效考虑了虚假用户的两个典型特征: 缺少用户间的交互行为和缺少博文内容中的交互。

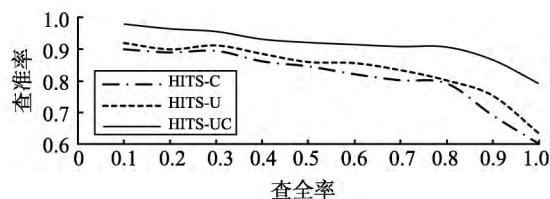


图 4 不同查全率下的查准率曲线

Fig. 4 Precision curves under different recall

第二项试验首先采用较为经典的算法 Basic-CA 算法^[9]、User-Rank^[18]算法和本研究提出的算法 HITS-UC 计算上述 100 个微博用户可信度值, 然后绘制三者对应的可信度曲线, 其结果如图 5 所示。图中横轴为用户编号 (共 100 个用户, 编号为 1 ~ 100), 纵轴为用户对应的可信度。观察图 5 可以发现, 虽然同一个微博用户采用 3 种算法计算的信任度不同, 但是 3 种算法下, 100 个微博用户的信任度曲线走势基本相同 (即信任度排序基本相同), 从而验证了基于 HITS 的微博用户可信度评估算法的可行性。从图 5 看出, HITS-UC 算法计算出的用户信任度大致上介于 Basic-CA 算法和 User-Rank 算法之间, 产生这种结果的原因是因为 Basic-CA 算法主要是基于博文内容计算用户可信度, User-Rank 算法主要是基于用户本身计算用户可信度, 而 HITS-UC 算法采用了等比例调和的方法同时考虑了用户和博文内容, 故理论上其计算出的可信度值也应基于二者之间。

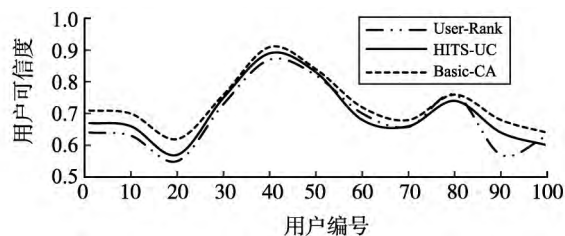


图 5 不同算法下 100 个用户的信任度曲线

Fig. 5 Credibility curves of 100 users based on different algorithms

4 结语

本研究参考了 HITS 算法的基本原理, 并将其应用到微博用户可信度评估, 提出了融合微博用户交互行为和博文内容的用户可信度评估算法 HITS-UC。试验采用从新浪微博采集的数据作为测试集, 首先获得了交互度阈值 ε 和信任度阈值 δ 的较优取

值,然后通过和 Basic-CA 算法、Page-Rank 算法比较,验证了本文提出算法的可行性。在未来的研究中,将进一步改进现有工作,使得现有算法更加完善、合理。

参考文献:

- [1] SONG J, LEE S, KIM J. Spam filtering in Twitter using sender-receiver relationship [M]. Berlin, German: Springer, 2006: 301-317.
- [2] 王越, 张剑金, 刘芳芳. 一种多特征微博僵尸粉检测方法 with 实现[J]. 中国科技论文, 2014, 9(1): 81-86.
WANG Yue, ZHANG Jianjin, LIU Fangfang. Detection of micro-blog zombie fans based on multi-features [J]. China Science Paper, 2014, 9(1): 81-86.
- [3] 刘晓飞. 基于链接分析的微博用户可信度研究[D]. 兰州: 兰州交通大学, 2015.
LIU Xiaofei. Research on credibility of microblog users based on link analysis [D]. Lanzhou: Lanzhou Jiaotong University, 2015.
- [4] 蒋盛益, 陈东沂, 庞观松, 等. 微博信息可信度分析研究综述[J]. 图书情报工作, 2013, 57(12): 136-142.
JIANG Shengyi, CHEN Dongyi, PANG Guansong, et al. A review of micro-blog information reliability analysis [J]. Library and Information Service, 2013, 57(12): 136-142.
- [5] 毛佳昕, 刘奕群, 张敏, 等. 基于用户交互行为的微博用户社会影响力分析[J]. 计算机学报, 2014, 37(4): 791-880.
MAO Jiaxin, LIU Yiqun, ZHANG Min, et al. Social influence analysis for micro-blog user based on user behavior [J]. Chinese Journal of Computers, 2014, 37(4): 791-880.
- [6] Wikipedia Inc. Credibility [EB/OL]. (2013-01-20) [2015-01-20]. <http://en.wikipedia.org/wiki/Credibility>.
- [7] CASTILLO C, MENDOZA M, POBLTETE B. Information credibility on Twitter [C]//Proceedings of Information International Conference on World Wide Web. New York, USA: ACM Press, 2011: 675-684.
- [8] 闫光辉, 刘晓飞, 王梦阳. 基于链接的微博用户可信度研究[J]. 计算机应用研究, 2015, 32(10): 2910-2917.
YAN Guanghui, LIU Xiaofei, WANG Mengyang. Research on credibility of microblog users based on link [J]. Application Research of Computers, 2015, 32(10): 2910-2917.
- [9] GUPTA M, ZHAO P, ZHAO J. Evaluation event credibility on Twitter [C]//Proceedings of the 2012 SIAM International Conference on Data Mining. California, USA: SIAM Press, 2012: 153-164.
- [10] MUKHERJEE A, LIU B, GLANCE N. Spotting fake reviewer groups in consumer reviewer [C]//Proceedings of the 21st International Conference on World Wide Web. New York, USA: ACM Press, 2012: 191-200.
- [11] CHU Z, GIANVECCHIO S, WANG H, et al. Detecting automation of twitter accounts: are you a human, bot, or cyborg? [J]. IEEE Transactions on Dependable and Secure Computing, 2012, 9(6): 811-824.
- [12] 徐建民, 栗武林, 吴树芳, 等. 基于逻辑回归的微博用户可信度建模[J]. 计算机工程与设计, 2015, 36(3): 772-777.
XU Jianmin, SU Wulin, WU Shufang, et al. Modeling user reliability based on logistic regression in Micro-blog [J]. Computer Engineering and Design, 2015, 36(3): 772-777.
- [13] 苗家, 马军, 陈竹敏. 一种基于 HITS 算法的 Blog 文摘方法[J]. 中文信息学报, 2011, 25(1): 104-109.
MIAO Jia, MA Jun, CHEN Zhumin. A new HITS-based summarization approach for Blog [J]. Journal of Chinese Information Processing, 2011, 25(1): 104-109.
- [14] 周小平, 梁循, 张海燕. 基于 R-C 模型的微博用户社区发现[J]. 软件学报, 2014, 25(12): 2808-2823.
ZHOU Xiaoping, LIANG Xun, ZHANG Haiyan. User community detection Micro-blog using R-C model [J]. Journal of Software, 2014, 25(12): 2808-2823.
- [15] KLEINBERG J M. Authoritative sources in a hyper-linked environment [C]//Proceedings of the 9th Annual ACM-SIAM Symposium on Discrete Algorithms. New York, USA: ACM Press, 1998: 668-677.
- [16] 田中生. 基于影响力的社会网络关键用户识别方法研究[D]. 长春: 吉林大学, 2015.
TIAN Zhongsheng. Research on key user identification method based on influence in social networks [D]. Changchun: Jinlin University, 2014.
- [17] 李赫元, 俞晓明, 刘悦, 等. 中文微博客的垃圾用户检测[J]. 中文信息学报, 2014, 28(3): 62-67.
LI Heyuan, YU Xiaoming, LIU Yue, et al. Research on detecting spammer in Micro-blogs [J]. Journal of Chinese Information Processing, 2014, 28(3): 62-67.
- [18] 王峰, 余伟, 李石君. 新浪微博平台上的用户可信度评估[J]. 计算机科学与探索, 2013, 7(12): 1125-1134.
WANG Feng, YU Wei, LI Shijun. Evaluation of user credibility based on Sina weibo platform [J]. Journal of Frontiers of Computer Science and Technology, 2013, 7(12): 1125-1134.

(编辑: 胡春霞)