

# 基于微博行为数据的不活跃用户探测

刘 晶<sup>1,2</sup>, 王 峰<sup>1</sup>, 胡亚慧<sup>1,3</sup>, 李石君<sup>1</sup>

(1. 武汉大学计算机学院 武汉 430072; 2. 中南民族大学计算机科学学院 武汉 430074; 3. 空军预警学院计算机教研室 武汉 430000)

**【摘要】**随着微博注册用户的增长,探测不活跃账号,自动判定用户活跃度有重要的商业价值。该文提出了一种自动检测算法并通过实验验证。算法核心是提出的影响用户活跃度的4个判定因子,可由用户行为计算得到。算法包含用户活跃度概率层次模型(ADPHM)和用户评分模型(USM)。ADPHM模型计算用户是不活跃用户的概率;USM模型计算用户活跃度得分。实验数据集包含了新浪微博2 316 281个用户信息和141 322 019条微博内容。实验结果表明,该算法能在线性时间复杂度下自动检测出不活跃账号,完善用户可信度评估体系。

**关键词** 活跃度; 自动识别; 不活跃用户; 微博; 社交网络

中图分类号 TP182

文献标志码 A

doi:10.3969/j.issn.1001-0548.2015.03.016

## Detecting Inactive Users from Behavior Data Based on Weibo

LIU Jing<sup>1,2</sup>, WANG Feng<sup>1</sup>, HU Ya-hui<sup>1,3</sup>, and LI Shi-jun<sup>1</sup>

(1. Computer School, Wuhan University Wuhan 430072;

2. College of Computer Science, South-Central University for Nationalities Wuhan 430074;

3. Computer Teaching and Research section, Air Force Early Warning Academy Wuhan 430000)

**Abstract** With the growth of registered users in microblog, how to detect inactive accounts and automatically judge the user activity have an important commercial value. To meet this need, an automatic detection algorithm is proposed and experimentally tested. The kernel of automatic detection algorithm is four determining factors of inactive users we defined, which can be calculated by user's behavior. The algorithm contains User Active Degree Probability Hierarchical Model (ADPHM) and User Scoring Model (USM). The ADPHM is employed to estimate the probability of inactive user; the USM is used to give a user's activity score. Experiment data contains 2 316 281 users' information and their 141 322 019 tweets crawled from Sina-Weibo. Experimental results show that this method can detect inactive users automatically and improve user confidence evaluation system in linear time complexity.

**Key words** activity; automatic identification; inactive users; microblog; social network

自2006年Jack Dorsey创建Twitter以来,以微博(Microblog)为载体的社交网络席卷互联网领域。微博以其快速便捷的特性风靡全球,迅速成为新一代信息传播形式。随着微博、微信等社交网络,移动互联网,智能终端的快速发展,出现了大量的与人密切相关的数据。如何从这些大数据中更好地理解用户和为用户服务是信息产业中一个重要研究方向。

近十年来,在线社交网络飞速发展。国内各主流门户网站也纷纷推出各自的微博产品,微博在中国呈指数级扩张,用户数量与日俱增。腾讯公司拥有7亿多用户,而新浪公布的最新数据表明,新浪的注册用户已超过5.6亿。在线社交网络已经成为连接物理社交世界和虚拟网络空间的桥梁。网络用户产

生的信息和用户与用户之间的交互在社交网站上留下了各种足迹,直接促成了网络大数据时代的到来。

新浪微博是国内最大最具有影响力的微博网站。而最近,香港大学研究人员进行的一项研究显示,新浪微博注册用户数量与真实用户严重不符。57%的新浪微博用户的信息流上没有任何内容,其帐户是非活跃帐户或由营销公司创建的所谓“僵尸号”。从商业角度来看,社交网站活跃用户越多,业务增长潜力越大。虽然新浪微博官方提出了一些发现和解决不活跃用户的措施和方法,但这些措施并不是那么行之有效。

目前仅有为数不多的方法提到了不活跃用户的探测和消除问题。文献[1]提出了一种基于统计特征

收稿日期: 2014-12-29; 修回日期: 2015-03-05

基金项目: 国家自然科学基金(61272109); 中央高校基本科研业务费专项资金(CZY15006)

作者简介: 刘晶(1981-),女,博士生,主要从事社会网络分析与数据挖掘方面的研究。

与双向投票算法来发现“完全非活跃”状态的隐式垃圾用户; 文献[2]通过统计用户发帖行为的显式特征来识别垃圾用户; 文献[3]利用Twitter中垃圾用户与正常用户的发帖行为和交友行为的不同特性来分析垃圾用户。

本文对新浪微博2 316 281个用户的141 322 019条微博行为进行了分析, 提出了用户活跃度因子概念, 建立用户活跃度概率层次模型和评分模型, 采用层次分析方法为微博用户划分等级, 描述用户从正常用户到不活跃用户所经历的几个阶段。实验表明利用该模型能有效探测出不活跃用户账号。

## 1 数据

本文通过新浪微博开放平台提供的应用程序接口(API)抓取用户基本信息, 用户发表的微博和用户关注/粉丝关系网络。随机选择了100个用户作为种子, 沿关注关系滚雪球(snowball)式爬取用户信息, 剔除重复账号后, 共获取了2 316 281个用户信息。爬取他们最近的1 000条微博行为(部分用户能获取的公开微博可能不足1 000)。最终用于分析的数据集包含2 316 281个用户信息和141 322 019条微博数据。实验数据集包含3部分: 用户基本信息、关注/粉丝网络和微博。

### 1) 用户基本信息

包含用户、性别、帐号创建时间、位置信息(省/市)、已发表微博数、粉丝数、关注数、描述及认证信息。

### 2) 关注/粉丝网络

新浪微博设置了3种用户关系网络: 关注、双向关注和粉丝, 其中, 关注和粉丝是一对相互关系。如果A关注了B, 则B在A的关注列表中, 而A在B的粉丝列表中。如果A关注了B并且B也关注了A, 则称A、B双向关注。双方互相出现在对方的关注和粉丝列表中。

一旦用户A关注了用户B, B发表的所有公开微博都会显示给粉丝A, 因此用户可以通过建立关注关系来订阅关注用户发表的微博。

### 3) 微博

微博信息集包含: 微博创建时间、转发数、评论数、被“赞”数及微博内容, 如果该条微博是转发, 则还包括被转微博的用户ID、被转微博ID、创建时间及微博内容。由于新浪微博把评论不作为一个独立微博消息, 而是关联到被评论的微博中, 故而本文抓取的用户微博中不包含评论内容。

新浪微博事实上是一个虚拟社交圈, 圈中的注册用户正常情况下应该发或者转发微博, 探讨和参与一系列微博活动。但某用户的微博账号如果自注册以来只有关注其他用户的行为, 那么该账号很有可能就是僵尸用户账号。此外, 各种用于社交网络的信息同步软件直接整合和同步国外微博上的信息, 再加上新浪微博用户以认证名人、机构和媒体为主的特性, 使得普通老百姓基本处于收听状态, 原本的发言平台演变为“只读不写”的收听平台, 也是导致正常用户逐步演变成不活跃用户的原因之一。

## 2 用户活跃度模型

如果一个用户在较长时间内没有发微博, 该用户就是不活跃用户。基于此, 本文提出微博变化率的概念。

**定义 1** 微博数变化率 $Tw(\Delta)$ 指用户在某段时间内的微博行为变化情况:

$$Tw(\Delta) = \frac{Tw(i) - Tw(j)}{t} = Tw' \quad (1)$$

式中,  $i, j$ 分别表示时间段的起止时间;  $t$ 表示从 $i$ 到 $j$ 的时间间隔。如果用户的微博数在一定时间间隔内发生变化, 那么该用户的活跃度就会增大。类似的, 分别定义用户的关注数变化率、收藏数变化率和互粉数变化率。

**定义 2** 关注数变化率 $Fo(\Delta)$ 指用户在某段时间内的关注数量变化情况:

$$Fo(\Delta) = \frac{Fo(i) - Fo(j)}{t} = Fo' \quad (2)$$

如果某用户的关注数间隔一段时间会增加或减少, 说明该用户在新浪微博上和其他用户间发生了互动行为。

**定义 3** 收藏数变化率 $Fa(\Delta)$ 指用户在某个时间段内收藏数量的变化情况:

$$Fa(\Delta) = \frac{Fa(i) - Fa(j)}{t} = Fa' \quad (3)$$

**定义 4** 互粉数变化率 $Bf(\Delta)$ 指用户在某个时间段内互粉数的变化情况:

$$Bf(\Delta) = \frac{Bf(i) - Bf(j)}{t} = Bf' \quad (4)$$

称式(1)~(4)中的4个变化率:  $Tw'$ 、 $Fo'$ 、 $Fa'$ 和 $Bf'$ 为模型中的4个关键因子。如果某用户逐渐趋向于不活跃, 这4大因子中若干项的变化率在某个时间段内必然会越来越小, 直至为零。当某个用户的4项指标全部为零时, 该用户则成为极度不活跃用户。

本文没有选择用户的粉丝数变化率作为衡量因子,原因在于粉丝是其他用户对某用户的单向关注行为,并不能体现该用户的个人动态和他从事的微博活动。

本文将定义的4个因子作为判断用户活跃度的初始判断依据。变化率可以被量化为得分 $S$ 和概率 $P$ 。其中, $P$ 表明用户是不活跃用户的概率。而根据得分 $S$ 可以计算出活跃用户的一个可能排序。为了表述方便,把这4个因子统一定义为通配影响因子 $Z$ 的4个特征向量:

$$Z = (Tw', Fo', Fa', Bf')$$

相应地, $Z$ 对时间的一阶偏导数和二阶偏导数可分别表示为 $Z' = \frac{\partial Z}{\partial t}$ ,  $Z'' = \frac{\partial^2 Z}{\partial t^2}$ 。

## 2.1 用户活跃度概率层次模型

为了清楚地划分不活跃用户的级别,体现出层次关系,文中建立了用户活跃等级的划分标准。当用户的概率 $P$ 在某个划分区间内时,就可以和划分标准形成对应的关系。由活跃度模型计算得出的概率 $P$ 大于等于设立的阈值或者得分 $S$ 低于阈值时,则用户趋近于不活跃用户。本文给出不活跃用户级别的判定标准如下:

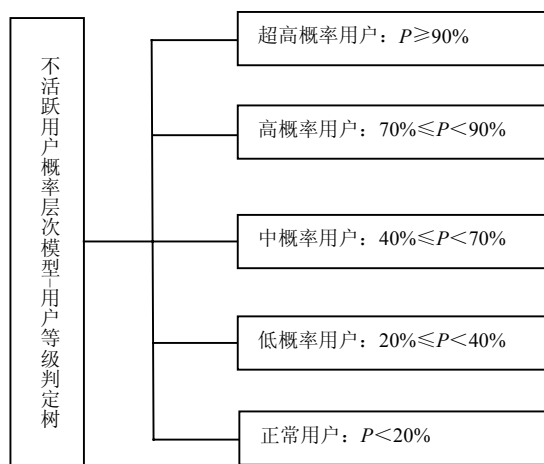


图1 不活跃用户概率层次模型-用户等级判定树

由图1可以看出, $P$ 高于90%的用户是不活跃用户的可能性最大; $P$ 介于70%~90%之间的用户为高概率用户,如果该用户继续停止在微博上的社交活动,其 $P$ 值会继续攀升,用户会演变成超高概率用户; $P$ 介于40%~70%之间的用户为中概率用户,其活跃度一般; $P$ 介于20%~40%之间的用户活跃度较高,是不活跃用户的可能性很低;而低于20%的用户为正常用户,一般来说,正常用户积极进行微博社交活动。

本文根据用户在新浪微博上行为特征的变化情

定义4个影响因子在某段评估时间内的变化趋势:

1) 上升的U型曲线:通配影响因子 $Z$ 对时间的一阶导数大于零且对时间的二阶导数大于零。说明 $Z$ 呈现出增长趋势,且增长速度越来越快,从而该用户被评估为不活跃用户的概率就会越来越小。

2) 上升的倒U曲线:通配影响因子 $Z$ 对时间的一阶导数大于零且对时间的二阶导数小于零。表明 $Z$ 呈现出增长趋势,但增长速度越来越慢,从而该用户被评估为不活跃用户的概率就会越来越大。

3) 下降的倒U曲线:通配影响因子 $Z$ 对时间的一阶导数小于零且对时间的二阶导数大于零。表明 $Z$ 呈现出下降趋势,且下降速度越来越快,从而该用户被评估为不活跃用户的概率就会越来越小。虽然这种情况下,该用户的活跃度很大,实际上这是一种严重失去粉丝的情况,这种情况的持续发展并不利于用户在微博平台的虚拟社交活动。当失去粉丝到一定程度之后,用户的4个影响因子就有可能趋于平衡状态,当这种情况出现时,该用户的影响因子对评估时间段的变化曲线与“下降的U型曲线”等同。当然这种情况在微博平台上的不活跃用户探测过程中并不常见。

4) 下降的U型曲线:通配影响因子 $Z$ 对时间的一阶导数小于零且对时间的二阶导数小于零。表明 $Z$ 呈现出下降趋势,但下降速度越来越慢,从而该用户被评估为不活跃用户的概率就会越来越大。

特别地,在上述情况2)和4)中,如果这种变化趋势最终导致 $Z$ 对时间的一阶导数等于零且对时间的二阶导数等于零,表明此时该用户在这段时间间隔内, $Z$ 既没有上升也没有下降( $Z' = \frac{\partial Z}{\partial t} = 0$ ),即这段时间内,用户未从事任何微博活动。此时该用户开始进入评估蛰伏期,所谓蛰伏期指的是用户成为不活跃用户的时刻 $i$ 开始到它摆脱不活跃状态呈现复苏状态的时刻 $j$ 为止的时间差值。 $Z$ 对时间 $t$ 的一阶导数为零时,表明该用户进入不活跃阶段,一阶导数为零的瞬间表示蛰伏期的开始,一阶导数不为零的瞬间表示蛰伏期的结束。如果用户的蛰伏期超过判定阈值,则该用户为不活跃用户的概率为100%,即用户被判定为不活跃用户。可表述为, $Tw' \cup Fo' \cup Fa' \cup Bf' = 0$ ,即 $Z' \cap Z'' = 0$ 。

## 2.2 不活跃用户评分模型

根据4个影响因子对时间的二阶导数的计算公式: $Tw'' = \frac{Tw'(i) - Tw'(j)}{t}$ ,  $Fo'' = \frac{Fo'(i) - Fo'(j)}{t}$ ,

$Fa'' = \frac{Fa'(i) - Fa'(j)}{t}$ ,  $Bf'' = \frac{Bf'(i) - Bf'(j)}{t}$ , 不活跃用户判定算子如下:

$$\eta = \begin{cases} 0 & Z'' = 0 \\ \sum \frac{Z'}{Z''} & Z'' \neq 0 \end{cases} \quad (5)$$

$\eta$ 等于0时,表明用户从该时刻开始进入蛰伏期,成为候选不活跃用户;  $\eta$ 大于0时,表明用户属于上述情况1)和3);  $\eta$ 小于0时,表明待评估用户属于上述情况2)和4)。

为了对不活跃用户进行排序,定义用户的行为特征变化曲线在 $t$ 时刻与 $X$ 轴形成的夹角 $\theta$ 的正弦值 $\sin\theta$ 为该用户在时间 $t$ 时刻的活跃度得分,用 $S(t)$ 表示,即 $S(t)=\sin\theta$ 。 $S(t)$ 表示其在微博平台中从事社交活动的活跃程度。得分越高,该用户4个影响因子的变化率之和越大,反之4个影响因子的变化率之和越小。基于 $S(t)$ ,给出不活跃用户的概率模型:

$$P(t) = (1 - \sin\theta) \times 100\% \quad (6)$$

$P(t)$ 表示用户在某一时刻 $t$ 被判定为不活跃用户的概率。特别的,虽然被判定为不活跃的用户在蛰伏期内活跃度为零,但这种判定是有时效性的。即使某用户已经被判定为不活跃用户,但只要在某时刻该用户再次活跃起来,那么他的判定概率将小于100%。

为了与2.1节中的不活跃用户等级判定标准相对应,根据用户的活跃度评分划分用户等级如图2所示。

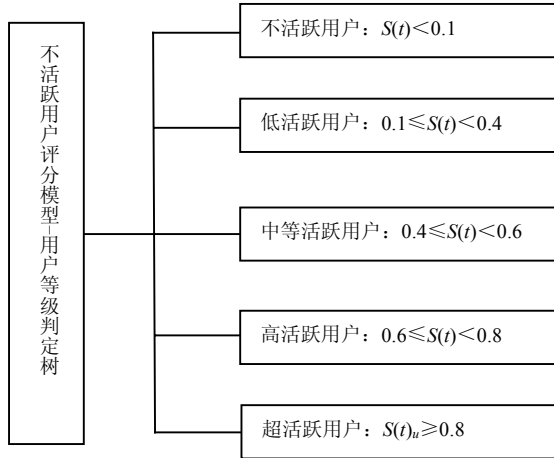


图2 不活跃用户评分模型-用户等级判定树

由图2可以看出,活跃度得分低于0.1的用户活跃度非常低,为不活跃用户。这类用户与超高概率不活跃用户相对应;得分0.1~0.4之间的用户为低活跃用户,与高概率不活跃用户相对应;得分介于0.4~0.6之间的用户为中等活跃用户,表明用户的微博活动较为活跃,与中等概率的不活跃用户相对应;

得分介于0.6~0.8之间的用户为高活跃用户,表明用户的微博活动较为频繁,与低概率的不活跃用户相对应;而得分高于0.8的用户为超活跃用户,表明用户的活跃程度非常频繁,与正常用户相对应。

### 2.3 不活跃用户的探测算法

本节介绍不活跃用户探测模型相关的算法,并对算法进行分析。

算法 1 不活跃用户评分模型算法

- 1) 输入: 用户的微博数 $Tw$ , 关注数 $Fo$ , 收藏数 $Fa$ 和互粉数 $Bf$ , 用户数 $i$ , 随机时间段 $\Delta t$ ;
- 2) 输出: 用户的活跃度得分 $S(t)$ ;
- 3) 初始化:  $i \leftarrow 1$ ;
- 4) 针对每个用户 $i$ 执行以下步骤:
  - ① 根据式(1)计算 $Tw'(i)$ ;
  - ② 根据式(2)计算 $Fo'(i)$ ;
  - ③ 根据式(3)计算 $Fa'(i)$ ;
  - ④ 根据式(4)计算 $Bf'(i)$ ;
  - ⑤ 根据式(5)计算判定算子 $\eta$ ;
- 5) 重复步骤4),直到处理完所有用户;
- 6) 根据判定算子筛选出不活跃用户;
- 7) 根据4个影响因子计算用户的活跃度得分;
- 8) 算法结束。

该算法在执行完待评估的 $n$ 个的用户,并计算出他们的活跃度得分之后终止,算法的时间复杂度为 $O(n)$ 。

将由算法1得到的用户活跃度得分作为不活跃用户概率层次模型算法的初始输入,详见算法2:

算法 2 不活跃用户概率层次模型算法

- 1) 输入: 用户数 $i$ ;
- 2) 输出: 不活跃用户的判定概率;
- 3) 初始化:  $i \leftarrow 1$ ;
- 4) 针对每个用户 $i$ 执行以下步骤:
  - ① 导入由算法1得到的该用户活跃度得分;
  - ② 根据式(6)计算不活跃用户的判定概率;
- 5) 重复步骤4),直到处理完所有用户;
- 6) 返回每个用户的概率;
- 7) 算法结束。

本文依据算法2计算每个用户是不活跃用户的概率,并结合2.1节给出的不活跃用户概率层次模型,将评估用户与用户活跃度等级相对应。

## 3 实验分析

首先根据数据集分析用户2次微博行为之间的平均时间间隔 $\Delta t$ 分布情况,如图3所示。

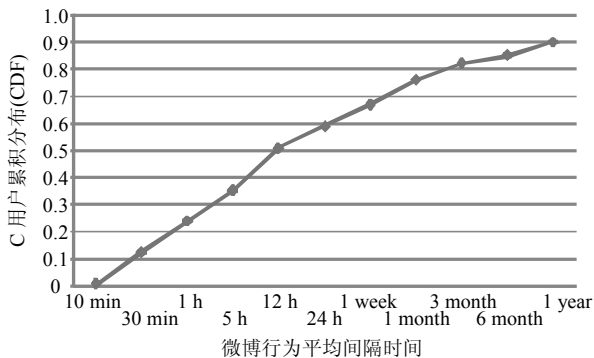


图3 用户平均微博行为时间间隔 $\Delta t$ 分布

结果表明约60%的用户每天都会刷新微博，约10%的用户在一年的时间内没有任何微博行为。

从数据集中抽样部分验证算法的有效性。根据上面的实验结果，选取24 h作为分析用户活跃度的时间段间隔 $\Delta t$ ，即以自然天数为分析计量单位。图4表示由不活跃用户概率模型计算得到的实验集中所有用户是不活跃用户的概率分布。

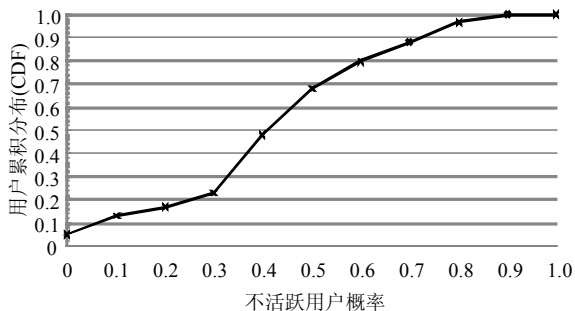


图4 不活跃用户概率分布

从图4可以看出，超高概率用户和低概率用户的用户数目相对较少，而中概率用户和高概率用户的用户数目相对较多。用户活跃度概率在区间0.3~0.6的用户占了全部用户的60%左右，用户的增长速度最快。整个不活跃用户的分布情况呈现出中间高，两边低的类高斯分布。实验提取不活跃度概率在0.6以上的高概率用户，分析其账号创建时间 $ctime$ 和最后一次发微博时间 $ltime$ 的关系，其结果如图5所示。

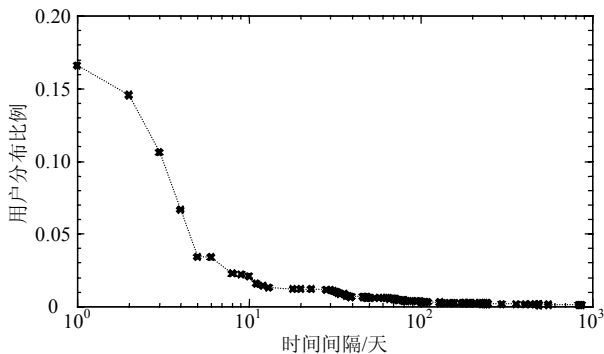


图5 不活跃用户生存时间间隔分布

结果表明不活跃用户大部分在创建账号后就几乎没有新的微博行为，有56.47%的用户在注册后一周内没有任何微博行为。根据实验的结果可以推断检测出的不活跃用户包括大部分未被使用或很少使用的账号。

实验结果表明不活跃用户探测模型能有效发现微博社区中的不活跃用户，能在一定程度上发现处于“完全不活跃”状态的用户。

### 4 结束语

本文分析了新浪微博用户的虚拟社交行为特征，提出4个关键因子作为不活跃用户的初始判定依据，同时为微博用户的社交活动勾勒出了一套从正常用户到不活跃用户的演变历程。在此基础上建立不活跃用户探测模型，初步解决了新浪微博平台上的不活跃用户探测问题。1) 提出了基于新浪微博平台上的不活跃用户探测模型：用户活跃度评分模型和不活跃用户概率层次模型；2) 提出了一套划分用户等级和用户活跃度的划分标准。

实验表明，基于新浪微博平台的不活跃用户探测模型能有效探测出不活跃账号，且探测算法的时间复杂度仅为 $O(n)$ 。本文的论述为新浪微博上不活跃用户的探测和判定工作，提供了一条新颖的思路，以供研究者将来开展进一步的研究工作。

感谢澳洲国立大学的Lexing, Paul博士对收集数据作出的贡献，以及对本项目的研究给予的支持和帮助。

### 参 考 文 献

[1] 丁兆云, 周斌, 贾焰, 等. 微博中基于统计特征与双向投票的垃圾用户发现[J]. 计算机研究与发展, 2013, 50(11): 2336-2348.  
 DING Zhao-yun, ZHOU Bin, JIA Yan, et al. Detecting spammers with a bidirectional vote algorithm based on statistical features in microblogs[J]. Journal of Computer Research and Development, 2013, 50(11): 2336-2348.

[2] STRINGHINI G, KRUEGEL C, VIGNA G. Detecting spammers on social networks[C]//Proceedings of the 26th Annual Computer Security Applications Conference. New York, USA: ACM, 2010: 1-9.

[3] WANG A H. Don't follow me: Spam detection in twitter[C]//Proceedings of the 2010 International Conference on Security and Cryptography. Washington, USA: IEEE Press, 2010: 1-10.

(下转第444页)

- networks[C]//Proc of the 5th WSEAS Intl Conf on Applied Computer Science. Hangzhou: [s.n.], 2006.
- [5] SALIM S, JAVED M, AKBAR A H. A mobile agent-based architecture for fault tolerance in wireless sensor networks [C]//Communication Networks and Services Research Conference (CNSR). Montreal, QC, Canada: IEEE, 2010.
- [6] QI Hai-rong, XU Ying-yue, WANG Xiao-ling. Mobile-agent-based collaborative signal and information processing in sensor networks[J]. Proceedings of the IEEE, 2003, 91(8): 1172-1183.
- [7] CHEN M, KWON T, YUAN Y, et al. Mobile agent based wireless sensor networks[J]. Journal of Computers, 2006, 1(1): 14-21.
- [8] HEINZELMAN W R, CHANDRAKASAN A, BALAKRISHNAN H. Energy-efficient communication protocol for wireless microsensor networks[C]//Proceedings of the 33rd Annual Hawaii International Conference on System Sciences. Hawaii: USA, IEEE, 2000.
- [9] HUANG C F, TSENG Y C, LO L C. The coverage problem in three-dimensional wireless sensor networks[J]. Journal of Interconnection Networks, 2007, 8(3): 209-227.
- [10] LIU B, TOWSLEY D. A study of the coverage of large-scale sensor networks[C]//IEEE International Conference on Mobile Ad-hoc and Sensor Systems. Fort Lauderdale, FL, USA: IEEE, 2004.
- [11] YANG T, YONG T. Short life artificial fish swarm algorithm for wireless sensor network[C]//International Conference on Computational Problem-solving (ICCP). Beijing: IEEE, 2013.
- [12] WANG Yi-yue, LIAO Hong-mei, HU Heng-yang. Wireless sensor network deployment using an optimized artificial fish swarm algorithm[C]//International Conference on Computer Science and Electronics Engineering (ICCSEE). Hangzhou: IEEE, 2012.

编辑 张俊

-----  
(上接第414页)

- [4] SOFUS A, MACSKASSY. On the study of social interactions in twitter[C]//Proceedings of the 6th International AAAI Conference on Weblogs and Social Media. Dublin, USA: AAAI Press, 2012.
- [5] REZA Z, MOHAMMAD-AMIN J, HAMIDREZA B, et al. A novel approach for social behavior analysis of the blogosphere[C]//Proceedings of the 21st Conference of the Canadian Society for Computational Studies of Intelligence. Windsor: Springer-Verlag, 2008: 356-367.
- [6] BOYD D, GOLDER S, LOTAN G. Tweet, tweet, retweet: Conversational aspects of retweeting on twitter[C]//Proceedings of the 43rd Hawaii International Conference on System Sciences. Honolulu, USA: IEEE Press, 2010.
- [7] LAS-CASAS PHB, GUEDES D, ALMIDA J M, et al. SpaDeS: Detecting spammers at the source network[J]. Computer Networks, 2012, 57(2): 526-539.
- [8] 莫倩, 杨珂. 网络水军识别研究[J]. 软件学报, 2014, 25(7): 1505-1526.
- MO Qian, YANG Ke. Overview of web spammer detection[J]. Journal of Software, 2014, 25(7): 1505-1526.
- [9] LIM E P, NGUYEN V A, JINDAL N, et al. Detecting product review spammers using rating behaviors[C]//Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM 2010). New York, USA: ACM Press, 2010: 939-948.
- [10] AKOGLU L, CHANDY R, FALOUTSOS C. Opinion fraud detection in online reviews by network effects[C]//Proceedings of the 7th International Conference on Weblogs and Social Media (ICWSM 2013). Menlo Park: AAAI Press, 2013: 2-11.
- [11] DUAN Z, CHEN P, SANCHEZ F, et al. Detecting spam zombies by monitoring outgoing messages[J]. IEEE Transactions on Dependable and Secure Computing, 2012, 9(2): 198-210.

编辑 蒋晓