软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2014,25(7):1505−1526 [doi: 10.13328/j.cnki.jos.004617]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

网络水军识别研究*

莫 倩, 杨 珂

(北京工商大学 计算机与信息工程学院,北京 100048)

通讯作者: 莫倩, E-mail: moqian@th.btbu.edu.cn, http://www.btbu.edu.cn

摘 要: 网络水军识别关键技术已成为当前数据挖掘领域最为活跃的研究之一.如何挖掘海量用户信息中潜藏的网络水军特征与行为模式,从而发现网络水军,以维护良好的网络环境,保障合理的网络秩序,已成为一项十分具有挑战性的工作.对比传统与新型网络水军识别研究,从识别特征角度对近几年内网络水军识别研究进展进行综述,对其关键技术和效用评价进行了前沿概括、比较和分析,并对网络水军识别中有待深入研究的难点和发展趋势进行了展望.

关键词: 网络水军识别;社交网络水军;电子商务水军;邮件水军;水军机器人

中图法分类号: TP393

中文引用格式: 莫倩,杨珂.网络水军识别研究.软件学报,2014,25(7):1505−1526. http://www.jos.org.cn/1000-9825/4617.htm

英文引用格式: Mo Q, Yang K. Overview of Web spammer detection. Ruan Jian Xue Bao/Journal of Software, 2014,25(7): 1505−1526 (in Chinese). http://www.jos.org.cn/1000-9825/4617.htm

Overview of Web Spammer Detection

MO Qian, YANG Ke

(School of Computer and Information Engineering, Beijing Technology and Business University, Beijing 100048, China)

Corresponding author: MO Qian, E-mail: moqian@th.btbu.edu.cn, http://www.btbu.edu.cn

Abstract: With its rising popularity, as evidenced in social networks, online shopping platforms and email systems, detection of Web spammer has already become one of the hottest topics in the data mining field. The main challenge of Web spammer detection is how to recognize spammer behavior patterns by examining spammer features and attributes from big dataset in order to limit the proliferation of Internet spam and insure quality of Internet service. This paper presents an overview of Web spammer detection, along with a comparison over the difference between traditional and burgeoning spammer detection approaches. The key techniques and evaluation methods are classified and discussed from several aspects. At last, the prospects for future development and suggestions for possible extensions are emphasized.

Key words: Web spammer detection; social network spammer; online shopping Websites spammer; E-mail spammer; spam bot

社会生活的高度信息化,使网络承载了蕴含价值的大数据,如新浪微博、大众点评网、豆瓣等拥有海量用户的社会化网络媒体,已经被组织和个人广泛地用来辅助决策.巨大的用户群与潜在的商机,使虚假意见和垃圾信息被广泛地制造和传播,该类危害的源头即俗称的网络水军.例如,“蒙牛陷害门”、“3Q 大战”、“王的盛宴”等事件背后,都隐藏着大量网络水军.网络水军形成巨大的虚假舆论场,影响网络民意、扰乱网络秩序、妨害经济利益,急需识别和治理.而网络水军识别研究[1−3]被认为可以有效解决此问题,得到学术界和工业界广泛关注和应用,并取得了一定的研究成果.网络水军识别通过挖掘用户信息中潜藏的水军特征和行为模式来实现.目前,网络水军识别研究在社交网络(如 Facebook,Twitter,MySpace,Weibo,RenRen)、电子商务(如 Amazon、eBay、阿

* 基金项目: 国家自然科学基金(61170112); 北京市属高等学校高层次人才引进与培养计划(CIT&TCD201304034); 民政部减灾和应急工程重点实验室开放基金(LDRERE20120105)

收稿时间: 2013-06-14; 修改时间:2014-01-21; 定稿时间: 2014-04-09

里巴巴、当当网)、邮件服务(如 Gmail,Yahoo E-mail,Hotmail)、网络论坛(如天涯、豆瓣)等众多领域都取得了较大进展[2,4−6].

以用户为中心的互联网信息服务,在表现形式、功能特性、服务质量等方面为用户提供了比传统网络更加丰富多彩的信息内容和网络服务.与此同时,网络水军规模大幅增加,其行为逐渐隐蔽和趋向正常用户.网络水军识别研究面临更大挑战.

近年来,网络水军识别研究[5−8]利用互联网海量数据的优势,并不断克服网络环境中的不利条件,通过分析大量用户信息发现水军特征和行为模式,识别出各个领域中潜藏的网络水军,解决网络水军危害互联网环境和秩序问题.许多大学和研究机构对网络水军识别展开了深入研究,如加州大学圣塔芭芭拉分校[2]、美国麻省理工媒体实验室[9]、宾州州立大学[10]、南阳理工大学[11]、伊利诺伊芝加哥大学[12]、伊利诺伊大学香槟分校[13]、卡耐基梅隆大学[14]、乔治理工大学[15]、德克萨斯 A&M 大学[16]、印度理工大学[17]、佛罗里达州立大学[18]等.国内的研究机构有香港科技大学[19]和清华大学[20]等.

本文对当前网络水军识别研究进展进行综述.第 1 节对网络水军识别研究进行概述.第 2 节重点介绍网络水军识别研究中若干关键技术,包括基于内容特征的网络水军识别研究、基于用户特征的网络水军识别研究、基于环境特征的网络水军识别研究、基于综合特征的网络水军识别研究、各领域网络水军识别研究对比以及网络水军识别效用评价等.第 3 节对网络水军识别中有待深入研究的难点和发展趋势进行展望.最后一节是结束语.

1 网络水军识别概述

1.1 网络水军识别研究的基本概念及其特点

网络水军是指那些由商业利益驱动,为达到如影响网络民意、扰乱网络环境等不正当目的,通过操纵软件机器人或水军账号,在互联网中制造、传播虚假意见和垃圾信息等网络垃圾意见产生者的总称[3,7,21−26].网络水军识别即在当前网络环境中运用 Web 信息挖掘技术[27],定义高区分度特征及行为模式发现潜藏的网络水军.网络水军也可以理解为整个网络用户中的离群点[28],但其特征与正常用户十分相近,因此其识别难度较高.

网路水军识别的形式化定义如下[21]:网络水军识别问题可以转换为一个二分类问题[29].设 U 表示访问某站点的用户集合:$U=\{u_1,u_2,\ldots,u_i,\ldots,u_{|U|}\}$,其中,$u_i$ 为第 i 个用户.设 A 为所有用户集合:$A=\{a_n,a_s\}$,其中,a_n 即正常用户的集合,a_s 为水军用户的集合,A 为若干个 U 的集合.目标函数为 $\Phi(u_i,a_j):U\times A\to\{0,1\}(1\leqslant i\leqslant U,j\in\{n,s\})$,其中,$\Phi(u_i,a_j)$是一个二分类函数,$\Phi(u_i,a_j)=\begin{cases}1, & u_i\in a_s\\ 0, & u_i\in a_n\end{cases}$.网络水军识别即发现用户 u_i 是否属于水军类别,因此,目标函数可以简化为 $\Phi(u_i)_s:U\to\{0,1\}$.

网络水军具备如下特点[1,3,8,30]:

(1) 目标相同:网络水军进行危害行为的目标大多都是获得经济利益或造成网络影响;

(2) 数量巨大:网络水军为达到其目的,造成网络影响,必然会大量利用水军软件机器人(后文简称为水军机器人)或傀儡账号;

(3) 行为异常:因其非正常动机,网络水军的行为模式区别于正常用户.

这些特点使得网络水军识别研究从统计角度具有可行性,为网络水军识别研究提供了基本的研究途径.网络环境的复杂和用户关注的增加,使水军行为模式隐蔽复杂化,并逐渐趋向于正常用户,也使得对其识别研究的难度加大.

1.2 传统网络水军识别研究

便捷邮件服务的流行,使互联网开始承载大量用户信息.早期网络环境中,获得用户邮箱和使用虚假邮箱的代价极小,初期邮件用户极易受影响,使得邮件领域网络水军泛滥.其运作方式主要表现为通过大量发送垃圾邮件引导用户前往商业性质网站,或通过水军机器人[31]发布海量垃圾邮件,以最大程度地传播垃圾信息.

上文所述即传统网络水军,其出现时间较早、数量规模相对较小、行为没有高度隐蔽性,产生的垃圾信息具有明显特征.因此,对其识别方法主要为基于垃圾信息内容分析,如邮件内容分析[32−37].同时,通过大量识别建立黑名单和白名单分别用来记录可疑用户信息和正常用户信息,以此提高水军识别效率及准确率[32].此外,邮件领域网络水军产生垃圾邮件所需资源类似,通过其使用资源及其网络层级特征能够很好地定位邮件水军.

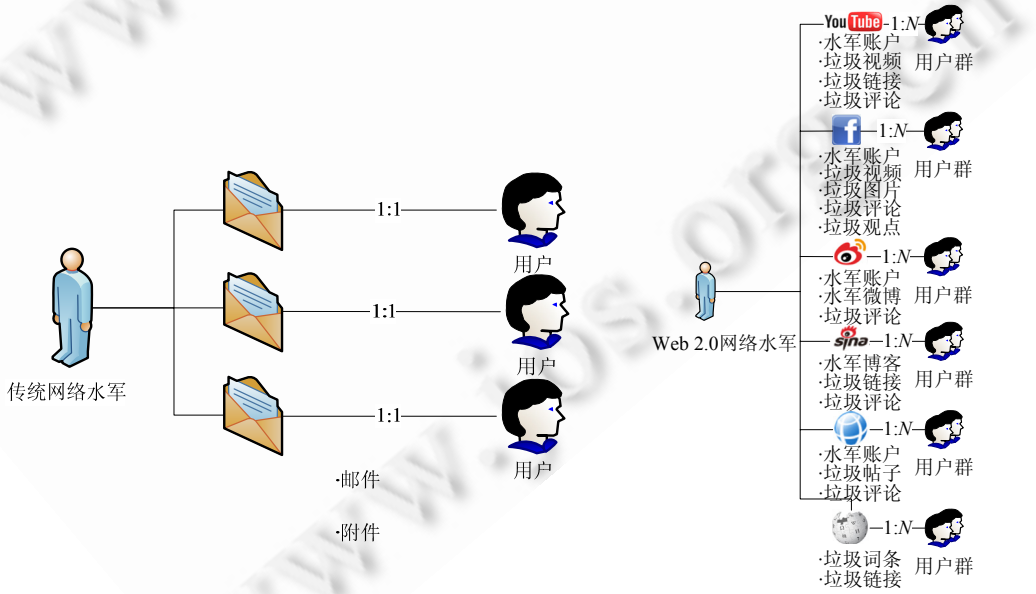
随着网络环境的复杂化和水军危害的增加,用户对其防范的能力也不断增强.为达到其目的,网络水军的行为逐渐复杂化并趋向于正常用户,传统邮件水军的识别方法无法发现这些隐蔽的网络水军.

1.3 Web 2.0网络水军识别研究

Web 2.0 是一种新兴的互联网方式,通过网络应用,促进网络中人与人之间的信息交换和协同合作,其模式以用户为中心.典型的 Web 2.0 站点有:网络社区、电商网站、社交网站、博客、Wiki、社会媒体等[38].Web 2.0 服务的盛行,使得具有极高价值的用户信息在网络上不断积累,开放的平台形成了以兴趣为聚合点的用户社群.这些特点同时也为网络水军提供了巨大的目标平台,刺激了水军的大量增加.其特点主要包括:目标范围广、危害影响大;关注点从目标产生内容转向目标用户本身;大量使用傀儡账号;行为具有高度隐蔽性;形成一定规模的网络水军团体,其团体内部具有紧密联系等[39].

当前,Web 2.0 网络水军识别研究按照目标领域的不同,可以分为邮件领域、电子商务领域、社交网络领域和论坛领域网络水军识别研究.网络水军识别研究按照研究方法的不同,可以分为基于用户产生内容特征、基于用户相关特征、基于环境特征的识别方法.本文将在第 2 节对不同目标领域中的网络水军识别研究进行分析和总结.

Web 2.0 网络水军识别研究是对传统水军识别研究的延伸与扩展,是网络环境变化衍生出的新型网络问题解决方案[40].同时,Web 2.0 环境下的网络水军涵盖范围较广,最初的垃圾邮件制造者和当前社交网络舆论影响者,都属于网络水军范围.近期互联网中的一些热点事件,如“蒙牛陷害门”、“3Q 大战”、“雇水军逛公园”、“艾滋女事件”等都属于网络水军造成的危害事件.Web 2.0 网络水军识别研究难度加大,较传统网络水军识别研究面临更大的挑战.图 1 显示了传统网络水军(如图 1(a)所示)与 Web 2.0 网络水军(如图 1(b)所示)目标用户及影响对比情况[41].



(a) 传统网络水军 (b) Web 2.0 网络水军

Fig.1 Traditional and Web 2.0 spammers' target and effect comparison

图 1 传统网络水军与 Web 2.0 网络水军目标用户及影响对比

此外,传统网络水军识别研究与 Web 2.0 网络水军识别研究的差异主要表现在目标对象、数量规模、识别难度等方面,详见表 1.

Table 1 Differences between traditional spammers detection and Web 2.0 spammers detection

表 1 Web 2.0 网络水军识别与传统网络水军识别的差异

	传统网络水军识别研究	Web 2.0 网络水军识别研究
水军目标对象	邮件用户	广大网络用户
水军影响程度	较小	极大
水军数量规模	正常	极其庞大
水军行为复杂性	较低	极高,趋向正常用户
水军伪装程度	较低	极高
识别需求	较小	迫切需要
准确性需求	一般	较高
识别难度	较小	较大
识别表现	一般	较好
识别自动化程度	较高	较低

2 网络水军识别的关键技术研究

Web 2.0 网络水军识别研究,是传统网络水军识别基础上的适应性识别研究.目前,国内外网络水军识别研究取得了较前几年更大的进展,但是仍然存在很多重要问题亟待解决.国外网络水军识别研究最初集中于邮件领域,并在近几年内迅速扩展到社交网络和电子商务领域中.国内网络水军识别研究相比之下较为缺乏.本文重点介绍基于内容特征、用户特征、环境特征以及综合特征的网络水军识别方法,并对邮件、电子商务、社交网络、论坛等互联网重要应用领域内的网络水军识别研究关键技术进行了对比分析和总结.

如图 2 所示,按照网络水军识别方法采用特征的不同,将网络水军识别方法分为基于内容特征、基于用户特征、基于环境特征和基于综合特征的识别.

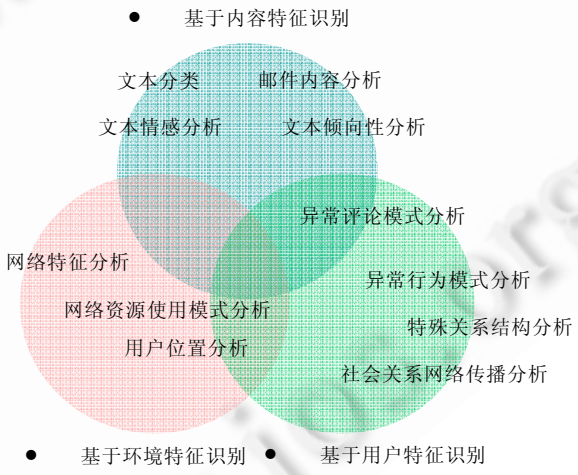


Fig.2 Comparison of different Web spammers detection methods

图 2 网络水军识别方法对比

2.1 基于内容特征的网络水军识别研究

早期网络水军识别研究着重分析网络水军产生的内容,这是由于早期网络环境中,网络水军产生的内容具有显著的可识别特征,如包含显著商业广告信息和垃圾邮件信息.并且早期网络环境中用户防范性较差,该类网络水军能够造成的影响巨大.基于内容特征的网络水军识别研究涉及机器学习中的自然语言处理分支.该类含

有观点的文本处理包括文本分类[42]、文本情感分析[43]以及文本倾向性分析[44]等方面.

早期邮件网络水军利用网络资源窃取用户邮箱信息的代价极小[45],其通过大量制造垃圾邮件引导用户点击商业广告站点,传播垃圾信息,以获得经济利益.对于垃圾邮件和邮件水军的检测研究很早就被重视[46,47].传统邮件水军识别研究主要基于垃圾邮件内容分析[48,49],该方法关注邮件水军制造的垃圾邮件本身.早期邮件领域中,具有显著商业特点的垃圾邮件能够引起用户注意,从而实现邮件水军影响目标用户的目的.邮件内容分析包括对邮件的贝叶斯分类[37]、基于关键词分类[50]、遗传算法分类[51]、神经网络分类[52]等方法.基于垃圾邮件内容特征的检测能够发现垃圾邮件和邮件水军,识别准确率较高.但是随着网络环境的复杂化,用户对于商业信息的反感不断增加,显著商业信息已经无法吸引用户.网络水军的策略随之变化,其制造的垃圾信息不再具有显著可识别的特征.因此,基于内容特征的传统网络水军识别研究不能有效发现新型网络水军.

Web 2.0 网络应用中,电子商务是发展最快的领域之一.电商平台中,用户对于商品的购买决定很大程度上依赖于商品的评论信息,如果某件商品拥有大量用户好评,用户会呈现出较大的购买倾向.各个商业组织和个人利用电子商务中用户产生内容辅助其决策,实现用户相关推荐等.电子商务领域成为众多网络水军目标领域之一.网络水军发布虚假评论来影响某件商品的评论走势,影响用户的购买决定,为其雇主或自身带来相应的商业利益.相比其他领域的网络水军行为,电子商务领域网络水军造成的影响最为严重.

传统电子商务领域网络水军识别方法主要依据评论内容相似性及其语言特征来发现虚假评论者,即电子商务网络水军[53−55].通过分析评论文本的倾向性,从而发现由网络水军发布、偏离正常用户评论的虚假评论[56],或通过虚假评论呈现出的不同于正常用户评论模式的特征[57],如重复使用大量无实义的形容词、语言多具有重复部分等,来识别潜藏于这些评论背后的网络水军.通过分析评论文本倾向性挖掘异常评论涉及一定的自然语言处理,其效率较低.而关注大量异常评论的统计模式避开了自然语言处理的瓶颈,利用统计学理论寻找异常评论,其效率和准确性都较高.文献[58]验证了利用评论自身的文本特征能够很好地识别虚假评论这一结论.但传统网络水军识别方法只能发现某一特定类型的网络水军,即,产生相似虚假评论的网络水军.但随着用户辨别力的增强,仅通过制造大量相似虚假评论的方式,网络水军能够造成的影响有限.因此,电子商务领域网络水军表现出更加趋向于正常用户的行为,其欺骗策略也逐渐多样化.传统电子商务网络水军识别方法对于越来越狡猾的水军并不适用.

网络论坛是用户发表自由言论并能够形成极大影响舆论的 Web 2.0 应用.论坛网络水军通过发布垃圾回帖或垃圾帖子提高其搜索引擎排名或影响用户意见导向.早期对于论坛网络水军识别的研究较少,较多的研究分析论坛上网络水军发布的垃圾内容,从而深入了解论坛水军的行为模式.文献[59]分别从用户浏览、论坛水军和论坛站点的角度对泛滥垃圾内容的论坛进行了分析,其发现网上论坛,包括网络环境较好的论坛,都存在大量由网络水军制造的垃圾内容.该文发现:论坛水军主要目的是为了提高其发布垃圾内容的搜索引擎排名,并提出基于内容特征的方法发现论坛水军制造的垃圾内容.该方法在发现论坛垃圾页面时较为有效,但其基于垃圾内容特征的方法无法高效地发现逐渐趋向正常用户的论坛水军.

本文作者研究团队分析股票领域的虚假股票评论观点,即股评观点,从而找出股票领域存在的网络水军.文献[60]提出一种混合的股评观点倾向性分析方法,找出偏离正常股评的股评信息,以最终找出利用这些不真实的股评观点影响股票价格的网络水军.文献[61]利用股票领域信息的篇章结构改进了对股评观点的分类,从而实现对股票评论更好的分析,并依据股票评论特征发现潜藏的网络水军.

2.2 基于用户特征的网络水军识别研究

随着网络环境逐渐复杂多样和用户辨别力的增强,使得制造传播具有显著特征内容的传统网络水军造成的影响不断降低.为了不断制造网络影响、妨害商业利益,网络水军逐渐衍生出多样的欺骗策略.其行为趋向于正常用户的行为,其发布内容也不再具有显著特征.通过分析变化的网络水军行为,基于用户特征的识别研究能够很好地发现潜藏的网络水军.因此,当前网络水军识别研究转向基于用户特征的识别,以实现从源头遏制网络水军和垃圾信息泛滥的目的.

2.2.1 基于用户行为特征的网络水军识别研究

分析当前网络环境和网络水军显示出的新特征,Prince 等人[45]首次利用“诱捕器”[62,63](即记录网络水军获取资源及制造垃圾信息过程的分布式系统)收集邮件水军行为数据,并对其如何获取用户邮箱的策略进行了分析.Bhat 等人[64]在分析大量邮件网络水军行为的基础上,利用两种行为特征构建邮件分类器.文献[64]采用的行为特征包括 URL 利用率和邮件发送时间.该文认为,邮件水军会利用非工作时间大量发送垃圾邮件,以此避免网络带宽的限制和被目标用户轻易发现.此外,该文还利用公开垃圾邮件数据集,能够与其他仅采用邮件内容特征的相似垃圾邮件过滤方法形成很好的对比.Stringhini 等人[65]通过人工主导邮件服务器发送错误反馈给邮件水军来遏制垃圾邮件的泛滥,他们认为:网络水军在发送大量垃圾邮件时,若收到邮件服务器关于目标地址不存在的反馈,则会降低网络水军对该目标地址的干扰.文献[65]从不同角度提出了一种遏制垃圾邮件蔓延的研究方法,但该方法需要极高的邮件水军识别准确性,以对其进行干扰.这一方法的前提条件在整个邮件网络水军识别研究中并不能保证完全满足,因此该方法的实际应用性较低,但该方案仍然为邮件网络水军的识别提供了一条新的思路.Husna 等人[66]分析了邮件水军机器人的行为特征,如其制造的邮件内容长度、邮件类型、垃圾邮件到达时间、垃圾邮件频率等特征,从而发现最为有效的邮件机器人行为特征.在此基础上,他们计算了大量邮件水军的相似性,聚类形成邮件水军团体,并对水军团体的共有行为进行分析,发现那些高度合作的邮件水军团体.Sawaya 等人[67]首次分析了复杂隐蔽性逐渐增加的邮件水军时间序列行为特征,并利用该类行为特征对其进行聚类分析.文献[67]通过识别移动服务商骨干网络中存在的邮件水军,按照其发送模式和聚类结果将其行为分为 3 类:突发性行为、周期性行为和持续性行为.与文献[45]采用的方法相似,该文也利用了诱捕账户来收集实验数据.

电子商务领域中,为影响和改变用户的购买决定,网络水军需对目标商品产生背离正常用户的评价.Lim 等人[12]捕捉了 Amazon 中几种具有代表性的网络水军行为,他们首先分析了 Amazon 中大量的产品评论,抽取出内容相近的评论.由于撰写让目标用户相信同时扭曲真实商品评论的虚假评论是极为繁复的工作,电子商务水军为了尽可能地减少自身工作量并获得最大的利益,其在发表虚假评论时会大量复制已发表或者别人的评论.因此,基于商品评价偏离的识别方法具有较高的准确率.此外,他们还发现:电子商务水军选取特定的目标商品,并不随机产生虚假评论,并且电子商务水军评分总是偏离目标商品评价均分.文献[68]进行了国内电子商务领域的网络水军识别研究,该文也是利用网络水军行为特征对其进行识别.Mukherjee 等人[69]利用电子商务领域网络水军可疑度作为隐性变量构建贝叶斯识别模型,他们分析了电子商务网络水军的各种行为,认为网络水军与正常用户具有极为不同的行为分布,如网络水平多具有评论集中突发性、评论极端性、发布早期产品评论等特点.文献[69]利用用户和其发表的评论特征构建分类器,利用上述水军行为特点将其与普通用户区分.该文首次利用评论自身语言特征客观验证了识别结果,并辅以人工评价,更好地验证了网络水军识别的准确性.该方法提高了电子商务网络水军识别评价的准确性,为电子商务网络水军识别结果评价提供了新的研究思路.文献[69]对电子商务网络水军的识别较文献[12]更加准确、高效,并实现了自动无监督识别,极大地提高了电子商务网络水军的识别效率,为该领域网络水军识别研究奠定了坚实的基础.

在新兴的 Web 2.0 网络服务中另一个急速扩张的领域是社交网络,社交网络包括各种各样不同类型的用户分享站点,如社交网站、视频分享网站、社会媒介站点以及社会化标签分享站点等.随着 Facebook、Tweeter、微博等国内外知名社交网站的兴起及火热,社交网络成为用户日常生活不可缺少的部分.社交网站中涉及到越来越多的用户信息,各个商业组织和个人利用这些信息辅助决策,因此,以妨害商业利益和扰乱网络秩序为目的,通过制造宣传垃圾观点的网络水军在社交网络中泛滥.社交网络领域的网络水军行为在各个目标领域中最为复杂,对其识别研究的难度也因此增大.尽管社交网络中水军行为具有很大的分散性,但其仍具有可建模识别的特征.社交网络中存储了海量用户观点,而网络水军也具有发表自身观点的权利.传统的基于内容特征的网络水军识别方法不可用于社交网络中.

但是,社交网络中,用户行为能够极大地反映出用户信息.例如,网络水军突发性行为模式与正常用户有明显区别.Benevenuto 等人[1]首先提出了在线视频分享站点中网络水军行为的统计数据,即,从著名互联网视频网

站 Youtube 收集的网络水军行为统计数据.这些数据证实了在社交网络中存在大量水军,同时,他们还定义了在线视频分享站点中的视频垃圾.他们采用人工标记方法建立训练数据集,即,数据集中的用户由人工标明是否为网络水军.之后分析已识别网络水军的行为,定义其特征,并利用 weka 中 3 种特征选择算法评价各个网络水军行为特征的分辨力.采用传统监督分类方法,判断未知用户是否为网络水军.该方法是社交网络领域水军识别研究的代表性方法,即,基于用户行为特征识别网络水军.之后的网络水军识别研究都以该方法为基础,增加特征或优化识别方法来提高社交网络中水军识别准确率.Parameswaran 等人[70]则提出了一种理论建模方法对水军行为及其偏离进行建模,他们发现:网络水军行为策略不断发生变化,并提出可以长期监控网络水军行为,以此建立黑名单来降低网络水军危害.

Webb 等人[15]将“诱捕器”[62,63]引入社交网络中,实现对网络水军的诱捕.他们利用 51 个诱捕器在 MySpace 上进行了 4 个月的实验数据采集,首次分析了社交网站中水军的地理特征,发现了网络水军地理分布现象,并对社交网络水军进行了分类.Lee 等人[16]证实了“诱捕器”能够收集网络水军行为模式信息,他们在 MySpace 和 Twitter 上进行了 10 种不同分类器的实验,评价了各个分类器的优劣,并在文献[71]中将该方法进行了优化.文献[71]通过 7 个月时间在 Twitter 上部署 60 个诱捕器收集网络水军行为数据,并通过分析这些数据定义了用户链路负载和粉丝关系等特征来提高识别表现.Langbehn 等人[26]使用多角度特征训练数据,以发现在线视频分享站点中的网络水军,减少基于单一特征识别所需的大规模数据集.Stringhini 等人[2]在 3 个主流社交网络平台:Facebook,MySpace 以及 Twitter 中使用“诱捕器”收集实验数据,分析不同平台的网络水军行为,提出相应的识别方法.此外,他们还分析了网络水军在社交网络中的影响力特征.由于社交网络水军的目的是达到影响最大化,因此,通过“诱捕器”获取网络水军行为数据的方法在社交网络水军识别中具有良好的应用性.相对上述研究,Ghosh 等人[17]采用不同方法收集网络水军行为数据,即,利用 Twitter 中已识别网络水军信息,即,Twitter 已封锁账户.他们分析了 Twitter 已封锁账户的行为策略,发现目标用户对网络水军的辨别力较低,更容易受到其欺骗.但该方法实验数据的全面性较上述识别方法要低,因此其召回率也相对较低.

Sureka 等人[72]提出视频分享站点中基于评论角度的网络水军识别方法,他们分析了 Youtube 中网络水军的评论模式.Yang 等人[73]首次深入分析了 Twitter 中网络水军的隐藏策略,并提出基于邻居节点特征发现 Twitter 中网络水军的方法.另外,还对比了文献[2]中提出的网络水军特征,评价了其在辨别网络水军行为时的分辨力.Tan 等人[74]在网络水军识别中引入了实时黑名单技术,对已确认的网络水军不再进行识别,同时维护记录正常用户信息的白名单.该方法对于总是使用新注册用户身份的网络水军无效,但具有这样特征的网络水军数量较少.由于网络水军的目的是最大化网络影响,并最大程度地减少工作量,所以网络水军会使用相对固定的水军账号.这些网络水军账号的注册时间都较长,并具有一定的用户影响力.Gargari 等人[75]在识别特征中加入网络资源层级特征实现网络水军识别,利用网络水军使用资源模式相似这一特征提高网络水军识别准确率.此外,它对社会化标签分享站点中的网络水军进行了分类.Zhu 等人[19]指出国内的社交网站——人人网中的网络水军识别缺陷,并提出复杂社交网络中的网络水军识别方法.

较短的帖子内容和自由的发帖规则使得网络论坛中存在大量由网络水军控制的水军机器人,该类水军机器人通过大量发帖和回复控制论坛的意见导向,同时使其制造的内容具有较高的搜索引擎排名,以污染整个网络内容.现有的对论坛水军机器人的防治多通过发帖回复时增加限制(如通过验证码等方式),或通过过滤论坛水军制造的垃圾内容等方式.但狡猾的论坛水军发展出各式策略避开该类限制,使得该类防治措施对论坛水军的防治效果较差,防治较为被动.

Hayati 等人[31]首次提出了 Web 2.0 水军机器人的概念,通过追踪其行为特征来发现 Web 2.0 平台中的水军机器人.该实验发现:水军软件机器人利用搜索引擎来寻找目标站点,并通过不断注册新用户账号,在短时间内发布大量垃圾信息.这些新注册账号的用户名利用机器生成,并且其个人主页只有较低的点击率,即,没有很好的交互行为.此外,Hayati 等人在文献[21]中提出了两种具体的行为特征:水军软件机器人的行为时间及其频率,实现对网络水军行为的建模.Hayati 等人在文献[76]中提出利用水军机器人进行水军行为时的特殊特征进行识别,并在文献[77]中综合早期研究提出利用自组织神经网络发现网络水军机器人的特征,并根据其不同的行为

将水军机器人分为 4 类,分别是:发布机器人、编辑机器人、传播机器人、综合机器人.2010 年,Hayati 等人在文献[78]中提出利用动态水军机器人的行为动作设计基于规则的动态识别方法,限制高度疑似的水军行为继续发生,从而在网络水军制造垃圾内容时制止其行为,实现从源头遏制网络水军泛滥的目的.

Raykar 等人[25]提出了众包服务中网络水军的定义,指出众包服务中网络水军行为具有其独特的特征.他们使用网络水军可疑度分数对用户排序,并通过经验贝叶斯算法来清除水军.同时,利用众包服务的质量分析众包服务站点中标签的一致性.

2.2.2 基于用户关系特征的网络水军识别研究

除基于内容和行为特征的网络水军识别研究方法之外,Brendel 等人[79]根据制造垃圾邮件的网络水军间关系特征来定位邮件水军.邮件往来中的用户关系具有稳定性,用户关系的变化能够反映用户行为.他们观察到:当用户关系发生某种改变时,极大的可能是由于网络水军大量制造垃圾邮件.同时,他们对用户关系依据图模型建模,并定义了相应的关系改变特征,发现突发的网络水军危害行为.此外,他们还对发现的可疑水军进行了隔离,实现垃圾邮件到达用户收件箱前的拦截,将网络水军造成的影响降至最低.该方法能够发现最近发生的网络水军行为,具有很好的时效性,从而在垃圾邮件开始泛滥时实现防治.但是该方法定义的特征对网络水军的分辨力较差,因此其准确率相对较低.Gammereldin 等人[5]根据邮件记录抽取出用户的社会网络,分析水军的社会关系特征,依此实现建模识别.Bouguessa 等人[22]分析了邮件记录中的用户社交网络结构,得到其中每个节点即用户的合法性分数,并按照混合β分布对其进行合法性建模,形成一种无监督网络水军识别方法.Sadan 等人[80]认为:评价邮件网络水军识别和垃圾邮件过滤的主要指标为其识别假阳性概率,因此他们分析了广泛运用的 URL 垃圾邮件过滤工具,提出利用邮件网络水军发送垃圾邮件时形成的发送域网络拓扑特征来修正过滤结果,以提高识别准确性并且降低识别假阳性概率.文献[80]经过实验发现:邮件网络水军形成的发送域网络拓扑中心性较低,反映出邮件网络水军通过使用大量易获得的网络资源来实现大规模垃圾邮件制造这一情况.

电子商务网络水军与目标产品和商店形成了一些特定关系.由于网络水军总是宣传其雇主的目标商品或商店,因此对于目标产品的选择并不具备随机性.Wang 等人[7]首次在电子商务网络水军识别中引入复杂关系模型,对网络水军、其产生评论、目标商店间的关系建模,分析节点间交互,以定位水军危害源头,并识别出可疑评论者.文献[7]是对现有网络水军识别研究的补充,它能够发现更加隐蔽的电子商务网络水军行为.此外,Wang 等人在文献[81]中使用图模型来表达网络水军可疑度、水军发布的评论质量以及评论所在商店的评分这三者之间的相互影响关系.其方法与 PageRank 中计算节点与边相互影响关系的方法相似,通过学习训练数据,迭代得到水军可疑度.该文利用图模型对网络水军间关系建模,很好地利用网络水军间关系特征来识别,具有很好的可推广性.与文献[7]关注网络水军、目标产品以及其发布评论三者之间的关系不同,Akoglu 等人[14]针对观察电子商务网络水军之间形成的网络体系,利用网络传播影响代替评论和用户特征来识别电子商务网络水军.他们首先对水军按照其特征进行可疑度打分,然后利用该水军与其他网络水军之间形成的关系对其进行可疑度修正.文献[14]是对手机 APP 评论中潜在网络水军进行的识别.不同于大型电子商务站点中网络水军发布的虚假产品评论,手机 APP 评论多具有短少杂的特点,并且其行为与电子商务站点网络水军也有一定的差别.因此,利用评论和用户特征无法很好地识别该类电子商务网络水军.该文很好地发现了该类型网络水军隐藏的关系特征,并发现该类网络水军多利用水军机器人发布大量虚假评论.此现象与移动客户端中网络水军识别研究还较为缺乏有关.随着移动互联网的兴起,移动客户端的网络水军识别研究是未来整个网络水军识别研究中的重要部分.Xu 等人[11]同样利用网络水军之间形成的组织关系,发现网络水军形成的网络水军团体,并在此基础上计算水军与其邻居间的相似性,利用 K 近邻算法修正对该网络水军的分类判定.该方法较其他运用关系特征构建单一分类器的网络水军识别研究更加准确.

利用网络水军形成团体关系的电子商务网络水军识别研究更多地利用图模型理论,结合网络水军独有的特点,不仅能够识别单个网络水军,并且能够发现网络水军形成的水军团体,因此能从源头更好地遏制网络水军的蔓延.

社交网络中,用户通过交互行为逐渐形成一个以用户为中心的社交圈子,用户间的社会关系蕴含着丰富的

用户信息.与正常用户相比,社交网络中的水军不具有正常的社会关系,其形成的关系网络结构特殊.例如,网络水军一般具有大量极度不平衡的关注粉丝比.因此,社交网络中的用户关系特征在识别社交网络领域网络水军时具有很好的区分性.

Song 等人[8]认为:基于用户行为特征的网络水军识别方法具有一定的滞后性和易伪造性,即,用户进行过类似水军的行为后,该类方法才能进行识别,并且该类用户特征很容易被网络水军修改掩饰.但是整个网络关系具有一定的稳定性,其特征不容易被用户行为所影响.因此,使用该类特征对网络水军识别具有更好的效果.文献[8]度量了用户间距离和用户联系紧密度等特征,并使用几种不同分类器进行分类学习.该文通过实验得出如下结论:在 Twitter 中,大多数网络水军发布的垃圾意见都只有少数接受者.Murmann[30]在 Twitter 中利用具有直接交互关系的邻居节点探测用户间信任关系,得到新的关系特征集.利用此特征集对用户进行可疑度排序,可疑度最高者即为网络水军.Moh 等人[82]同样利用 Twitter 中的用户社会关系,如其朋友及粉丝特征,通过不同特征矩阵得到该用户的可信任度,以此判断该用户是否为网络水军.Gayo-Avello 等人[83]依据图论利用 Twitter 中网络水军花费大量时间来粉目标用户或者等待目标用户回粉的特点来发现网络水军,并提出 Twitter 中网络水军最关注的主题排序.此外,他们还提出利用用户影响力特征以提高网络水军识别的准确率.Krestel 等人[84]利用网络水军可疑度会在社交网络中传播的特点,利用图模型上的传播发现标签分享站点中的网络水军.该方法实现了标签分享站点中网络水军、标签以及网络资源间关系结构的建模.通过给定一些种子节点的可疑度,依据种子节点向外传播可疑度的特点,从而计算整个图模型中所有节点的可疑度,发现可疑的用户节点.Bhat 等人[85]发现:与普通用户相似,社交领域网络水军也可形成一定程度的网络水军社区.因此,文献[85]从用户行为日志中抽取出用户交互图,发现其中形成的重叠社区图谱.在人工标记出一部分网络水军节点后,计算每一个待识别节点与已标记节点的社区关系,以此对未知节点进行分类.该文在真实的社交网络数据中加入模拟的网络水军用户,但人工模拟的网络水军行为往往与真实社交网络中泛滥的网络水军行为具有一定差别[86],因此,该方法对于广大真实社交网络中潜藏的网络水军识别表现如何仍有待实验评价.

Web 2.0 社交网络服务极大地反映了真实世界中人们的交往圈子,同时,该类型社会关系并不会随着网络水军日益复杂的躲避策略轻易变化,普通用户和网络水军都在该领域中形成一定的社会网络结构.因此,该类网络水军识别研究中利用用户关系特征挖掘社交领域网络水军有很好的可借鉴性.同时,该类网络水军识别研究中采用的网络结构特征不尽相同,因此其识别效率也不尽相同.对比现有研究来看,利用网络水军自身具有高度聚集性以及与普通用户关系稀疏性等特点,能够很好地发现社交领域的网络水军.

2.3 基于环境特征的网络水军识别研究

隐蔽的网络水军对用户展现出趋向于正常用户的特征,但其异常行为使其在网络环境层级表现出不同于正常用户的特点.Ramachandran 等人[87]在 2006 年首先提出基于水军网络级别特征的识别方法.他们从被水军污染领域中追踪收集了 17 个月共 1 000 万的垃圾邮件信息,将该数据与基于 IP 的黑名单信息、TCP 脚印信息、路由信息以及机器人网站命令追踪信息等联系起来对水军的网络级别特征进行分析,实现垃圾邮件的追踪.此外,他们还提出了邮件水军在网络级别的危害策略.

Tseng 等人[88]提出利用邮件服务器记录建立图模型,其中每个节点代表一个邮件账户,每条边代表一次邮件交互.计算图中每个节点的疑似水军分数,对其部分节点进行迭代排序,从而发现邮件水军.他们利用的邮件水军网络级别特征有 IP 地址和垃圾邮件频率等特征.

Hao 等人[89]也是利用水军网络级别特征进行识别,他们从网络特征一阶属性得到轻量级网络水军特征,这些轻量级特征不需要通过训练大量数据而得到,并且不需要对网络水军产生内容进行分析.同时他们还加入了黑名单技术作为首次过滤,提高了识别效率,避免了已确定网络水军的再次识别.文献[89]的数据来源为商业网络垃圾信息过滤系统.Uddin 等人[90]则是利用边缘路由器中 IP 冲突记录,即,网络流量记录来得到网络流量的统计信息,从而得到水军的网络级别行为特征(如用户 IP 间距离、实际物理距离等).此外,他们还提出了一个软件机器人网站的攻击行为探测机制,即,基于软件机器人攻击行为特征探测,并对软件机器人网站的攻击行为进行了分类分析.Ehrlich 等人[91]利用网络流量特征区分邮件水军机器人和正常用户,建立了首个 SMTP 流量模型,

发现了邮件水军在 SMTP 级别表现出的特征.

Schatzmann 等人[92]认为,基于水军网络协议级别行为特征的识别无法满足网络水军识别需求.他们提出从网络核心部分分析网络水军行为,以实现复杂网络水军行为的探测.他们利用一个国家级 ISP(Internet service provider,网络服务提供商)的网络水军行为记录,从 ISP 角度提出流量级别特征,实现对网络水军行为的建模.该方法使 ISP 能够监控其服务器,及时发现可能存在的水军行为并对其实施控制.针对文献[46]无法普遍适用性的缺点,Xu 等人在文献[93]中提出基于网络水军资源使用模式的识别方法.该文发现:网络水军总是利用可轻松获取的开放资源,因此,根据资源的突发性使用模式可以很好地识别网络水军.此外,该文对钓鱼水军和非钓鱼水军进行了区分,发现了一个具有高度行为相似性的资源捕获水军团体.该文发现的网络水军使用资源如下:某台代理服务器和获取大量邮件地址的开放网络服务等.该文不需要大量难以获取的实验数据,因此,该方法较文献[92]有更好的应用实验性.

Las-Casas 等人[6]提出从网络水军产生源头进行识别的方法,即,基于水军产生时的网络特征识别,并使用巴西宽带 ISP 的数据记录作为实验数据集.他们提出:仅根据单个网络水军行为很难精准发现网络水军,但从网络流量的角度却很容易发现网络水军特征.因为网络水军为尽可能减少工作负担并达到效益最大化,会集中在一段时间内大量制造垃圾意见,因此这段时间内的网络负载会突然加大,流量也会集中在某些链路中.与传统网络水军识别方法相比,该方法准确率较高.但该方法需要应用 ISP 数据,无法普遍推广.Duan 等人[18]实现网络水军机器人的识别,将水军机器人定义为那些被网络水军操作的傀儡机器账号.他们监控了一个美国校园网络中的出口信息,并进行了为期 2 个月的邮件追踪来收集实验数据,从实验数据中分析网络水军机器人特征,对其进行识别.

文献[94]提出了垃圾邮件产生的另一个来源,即,大规模僵尸网络.但僵尸网络与网络水军特征并不相同,僵尸网络的目的是窃取用户信息,对用户造成恶意危害;而网络水军并不刻意窃取用户信息,其旨在影响用户决定,妨害其经济利益.

基于环境特征的网络水军识别研究依据网络水军进行危害行为时产生的环境特征,该环境特征是无法被网络水军修改掩饰,因此其识别准确率较高.但基于环境特征的网络水军识别研究大多需要相应的实验数据集,因此其可推广性较其他网络水军识别研究方法要低.

2.4 基于综合特征的网络水军识别研究

如上文所述的网络水军识别研究多基于特定类型的网络水军特征,但基于特定类型网络水军特征的识别方法无法全面分析网络水军行为,因此其识别准确率具有瓶颈.在此基础上,综合多种特定类型的网络水军识别方法对于各个目标领域的网络水军都具有较高的识别准确率.

邮件领域中,结合垃圾邮件内容分析和网络水军显著特征,能够很好地提高邮件水军的识别,从而抑制大量垃圾邮件的产生.Li 等人[95]发现:含有相同 URL 的垃圾邮件是由同一个团体的邮件水军制造传播,该类 URL 是其共有的目标.根据该特征,Li 等人利用邮件水军所在团体信息,提高对危害极高的邮件水军,即,那些属于多个水军团体的邮件水军的识别,从而抑制了可能产生的大量垃圾邮件,提高了整个邮件水军的识别表现.

电子商务领域中,综合传统内容和用户特征的网络水军识别方法,能够提高电子商务领域网络水军识别表现.Gilbert 等人[13]分析了 Amazon 中一种特定类型水军,即,评论跟风者(发布与前者相同评论的评论者).他们对 Amazon 中丰富的评论信息进行了分析,发现了相似评论,从而找出评论跟风者.该方法只针对特定类型的网络水军,数据来源采用基于传统内容特征的方法,因此其识别表现有待提高,并且对于一般类型网络水军识别并不适用.Mukherjee 等人[4]首次提出了电子商务领域网络水军团体的识别方法,他们利用评论内容分析产生候选团体,再根据网络水军行为特征发现网络水军.在此基础上,并对网络水军团体、个人网络水军以及目标产品间关系建模,产生了一个人工标记的网络水军团体数据集.文献[4]是对文献[12]的进一步优化改进.Lu 等人[20]将网络水军用户特征与内容特征利用评论因子图模型相结合,并利用人工标记网络水军样本和可信度传播理论识别未知网络水军.文献[20]中建立的网络水军识别模型可同时挖掘出潜在的网络水军及其发布的虚假评论.但与文献[4]相似,识别中运用的网络水军样本都采用人工标记的方法,因此无法避免主观性因素,对识别准确率造成

一定的影响.与上述文献不同,文献[96]关注的是电子商务网络水军的行为和关系特征,其收集了中国亚马逊中百万用户评论数据、60 万评论者以及 13 万产品数据,分析用户行为特点,并评价了常见网络水军行为特征的识别有效性,选取最为有效的特征集构建分类器,同时挖掘其中的网络水军关系特征.在上述分类器的基础上,利用某个网络水军与其他网络水军形成的关系网络修正分类结果,以达到更好的识别效果.

电子商务网络水军受商业利益驱动,往往较其他领域网络水军表现出更强的隐蔽性,对其的识别研究也更为复杂,因此,利用电子商务网络水军综合特征对其进行的识别研究,较上述网络水军识别研究中使用单一分类器的方法更为严谨、准确,如文献[96]所采用的识别方法.该类方法扩展了以往网络水军识别的局限,充分利用网络水军的特点,以达到最好的识别效果.

基于用户行为和关系特征的网络水军识别方法捕捉用户在社交网络中的行为和关系,并不分析其发布的内容.传统网络水军识别方法即依据用户产生内容判断其是否为水军,因此结合用户行为、关系及其发布内容的综合特征网络水军识别方法很常见.

Zinman 等人[9]利用朴素贝叶斯以及神经网络的方法建模社交网络中的用户,按照其活跃程度将其分为 4 种类型.该文对社交网络中用户行为和关系特征进行了分析,依据网络水军显著行为模式进行识别,拓宽了社交网络用户的分类.Benevenuto 等人[97]将在线视频分享站点中的网络水军按目的分为两类细粒度的水军,并分别给出了这两类细粒度水军的定义.此外,该方法利用人工标记建立了一个 Youtube 用户标记数据集,为该类型网络水军的识别提供了测试数据集.根据网络水军以达到影响最大化为目的,Benevenuto 等人[3]分析了 Twitter 上 3 个最热门的主题,对其涉及到的用户进行标记,并利用 Twitter 用户的 Tweet 及其行为特征判断其是否为水军.但该方法使用的实验数据仅为部分热门主题参与用户,对用户的覆盖率有限,因此学习达到的效果有限,对网络水军的识别只有 70%的准确率,但其对于正常用户的识别准确率能够达到 96%.

Amleshwaram 等人[98]综合社交网络中各方面用户的特征,如行为、内容、用户间关系等,实现了社交领域网络水军的快速识别,且其识别所需时间和资源都大为降低.他们对未知水军进行了聚类分析,发现了 Twitter 上盛行的一些网络水军团体.他们发现:Twitter 上大多数网络水军只有很少的 Tweet,其主要目标是传播垃圾信息,制造网络影响.

Lin 等人[99]利用多种渠道收集了中国新浪微博中的大量网络水军数据,其中包括利用“诱捕器”和网络爬虫收集的 1 000 多个网络水军、利用关键字搜索并人工标记收集的网络水军以及直接购买的 8 600 个新浪微博僵尸粉丝.根据其行为目的不同,将该网络水军数据集中的网络水军分为 3 种类型,分别分析其行为关系特征并构建特定类型网络水军分类器.他们对社交网络水军按其行为特点进行细粒度划分,并构建相应的分类器识别,其识别结果具有较高的准确率.同时,由于网络水军数据来源广泛,涵盖了大多数新浪微博中可能存在的网络水军,因此,利用该数据集对新浪微博网络水军的识别具有良好的表现.在文献[99]将新浪微博网络水军进行细粒度划分的同时,有可能忽略某些职业水军,其表现为各种细粒度网络水军类型的综合.该文所采用的方法极有可能无法识别该类危害较大的职业网络水军.Yang 等人[100]对 Twitter 中大量网络水军的躲避策略进行了深入分析,并利用设计的爬虫工具深度挖掘 Twitter 数据,补充了仅利用 Twitter API 收集用户数据的不完整性,但因其数据来源相对单一,其网络水军识别实验数据准确性较文献[99]要低.同时,文献[100]利用网络水军对不同行为特征的隐藏成本和收益的差值评价了已有网络水军识别方法中所采用特征的有效性,如文献[2,3,10,16]中所采用的识别特征,按照其鲁棒性将其分为低效、居中、高效 3 类,并提出新的有效特征.此外,该文不同于以上社交领域网络水军识别研究的是:该文只关注某类特定的网络水军,即,发布钓鱼链接的恶性网络水军,因该类网络水军严重威胁着社交网络用户的隐私及安全.对其的识别研究,较其他危害性较低的广告垃圾型网络水军更为迫切.该文的识别方法对该类特定网络水军具有很好的识别效果,但对于社交领域广泛存在的大量广告垃圾型网络水军没有很好的识别表现,其方法具有一定的局限性.Wang 等人[101]收集了新浪微博中评论数超过 7 000 的 307 条热门微博以及其包含的所有用户及用户间联系和 4 000 万评论作为新浪微博网络水军识别数据,并人工标记出了 212 个网络水军和 732 个普通用户.利用可快速计算出的用户行为关系特征以实现新浪微博网络水军的高效实时识别,但文献[101]所采用的识别数据仅仅来源于新浪热门微博,虽可能包含大量的网络水军,但与文

献[99]相比,其数据完整性仍然有待提高.同时,其人工标记出的网络水军样本数目较少,加之受人工标记的主观性影响,其一定程度地影响了识别的准确性.

Aggarwal 等人[102]首次对移动客户端地理位置服务型社交网络,如 Foursquare(移动端地理位置服务型社交网站)中潜藏的网络水军行为进行了分析识别,利用 Twitter 中有关 Foursquare 的信息挖掘相应 Foursquare 用户数据,同时将 Foursquare 用户按其动机分为 4 类,并人为标记出其中的可疑网络水军.与文献[99]相似,分别对其构建分类器进行识别.移动端网络水军相较已有网络水军规模较小,但其行为特点近似于已识别出的大量互联网网络水军.因此,文献[99]无法覆盖某些职业网络水军并一定程度地影响了识别表现的问题在文献[102]中并不严重.此外,文献[102]是对于移动端社交领域网络水军识别的较新研究,其方法为移动端网络水军的识别研究奠定了基础.

因其行为与正常用户的极高相似性,论坛水军的识别具有相对较高的难度,现有研究多通过综合利用各类特征以提高其识别表现.文献[103]通过分析论坛内容并抽取其发布者关系,获悉论坛中存在的意见对峙结构,从而发现其异常意见持有者.该方法能够发现论坛中发布与目标帖子无关内容的论坛水军,但却无法区分已形成意见倾向的大规模论坛水军.而该类论坛水军是目前大量泛滥于各大论坛的主力水军,因此该方法具有较大的局限性.文献[104]利用网络水军行为特征及其制造的垃圾内容特征对国内门户网站论坛中的网络水军进行识别,利用网络水军的发布数量、回复间隔和活跃程度等行为特征,综合其制造相近垃圾内容的特点,发现潜藏的网络水军.此外,该文首次分析了国内论坛网络水军的地理分布情况.

2.5 各目标领域网络水军识别研究总结

邮件领域的网络水军识别研究在传统的基于内容特征的网络水军识别方法中引入用户行为、关系和环境特征,提高了网络水军识别表现.由于邮件领域网络水军表现出高度的行为相似性,因此,基于行为特征的识别方法能够较好地发现邮件网络水军,在邮件领域各类网络水军识别方法中平均表现最好.

电子商务领域网络水军识别研究具有极高的应用价值,一直是近几年研究的热点.按照其识别体系不同,又可将电子商务网络水军识别研究分为监督识别和非监督识别.电子商务网络水军监督识别方法主要包括依据评论内容特征的识别方法,如上文所述文献[55,57,58,105]所采用的方法.而非监督识别方法按照识别特征不同可分为 5 类,其识别特征主要包括个人网络水军特征[7,12]、网络水军团体特征[4,106]、时间序列特征[107]、评论模式特征(关联规则[57]、评论突发性[108])、行为分布特征[109].由于网络水军行为越来越隐晦,使得提高网络水军识别表现和评价识别结果面临极大挑战.传统的基于内容特征的网络水军识别方法对于新型隐蔽的电子商务网络水军识别效果较差,其识别性能具有瓶颈.电子商务领域网络水军识别方法主要为基于网络水军行为和关系特征的方法.网络水军目标的相同,使其行为和关系表现出高度可识别的特征,因此,基于综合特征的网路水军识别方法能够达到较高的准确率和召回率,相比基于特定类型特征的网络水军识别方法表现得更好.

社交网络领域网络水军识别研究难度大,识别周期长,结果不易评价,但结合内容、行为和关系等特征的综合特征网络水军识别方法能够提高网络水军的识别表现,对复杂社交网络中的网络水军识别表现最好.该领域网络水军特点繁多、数量庞大,其可造成的网络影响规模也日渐扩大.早期隐蔽性较差的社交领域网络水军通过伪造自身数据吸引或欺骗普通用户,该类水军可通过文献[2]中所采用的识别用户上下文特征的方法进行识别.但随着用户防范性能力的增强,网络水军危害策略不断变更,挖掘网络水军的本质性特征,即不易伪造特征,则成为识别网络水军的关键.该类社交领域网络水军识别特征包括周期性行为特征[1,15,70,73]、社会结构特征[8,82,85]、社会交往频率[30,83,84,86]等.采用该类特征进行的社交领域网络水军识别研究,比采用其他特征的识别研究更为有效、准确.

综上所述,传统网络水军集中于邮件领域并逐渐向电子商务和社交网络转移.这 3 个目标领域中的网络水军行为较为严重,对网络环境影响也较为严重.3 个目标领域中,网络水军识别都主要依据网络水军的自身特征,如内容、行为以及关系等特征.与传统的基于内容特征的网络水军识别方法相比,当前网络水军识别研究关注点转向网络水军的目的分析,实现从源头防治网络水军泛滥.其中,社交网络中的水军最为严重,且其行为较其他目标领域都更为复杂.基于特定类型特征的网络水军识别方法在邮件领域和电子商务领域可一定程度地发

现网络水军,但在复杂社交网络中,以上方法的网络水军识别表现较差.因此,社交网络水军识别研究基于综合特征进行识别,并不断适应性地增加网络水军识别新特征,以提高其识别准确率.

图 3 综合显示了网络水军不同目标领域和识别研究方法的综合对比,其中,X 轴为网络水军识别方法复杂性,Y 轴为不同目标领域中网络水军的行为复杂性.

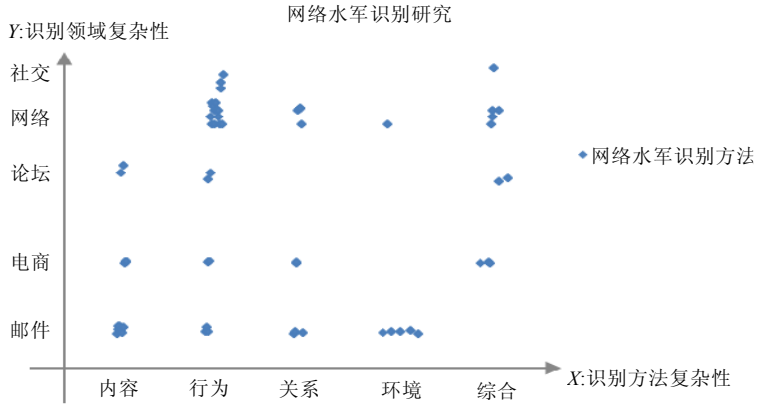


Fig.3 Domain and methods of Web spammers detection

图 3 网络水军识别研究领域方法图

2.6 网络水军识别的效用评价

效用评价对于检验网络水军识别表现和发现其存在的问题都十分重要,是网络水军识别不可缺少的步骤,而数据集和效用评价指标是其中两个重要的因素.

2.6.1 网络水军识别的相关数据集

目前,网络水军识别研究中没有公开可用的数据集,为验证网络水军识别结果的有效性,需从网络中收集相关领域实验数据集[1,15,18,71,73]或利用一些组织和公司提供的实验数据集[6,10,75,89].

文献[5]是邮件领域的网络水军识别研究,该文的数据集是校园网络 sustech.edu 中的邮件记录数据.该数据集包括 2009 年 11 月~2010 年 4 月间所有邮件的收发记录,共有 1 038 939 条记录,其中包括由 234 个账户产生的 20 963 个来往邮件地址对.

文献[6]也是邮件领域的网络水军识别研究,它使用协议解析器收集了巴西一个国家级 ISP 的出口信息,包括初始化时间、持续时间、使用协议、下载量、上传量、源 IP、目的 IP 以及 SMTP 协议中邮件地址等属性.每条记录的时间段都为 28 天(2009.3.1~2009.3.28 和 2010.6.12~2010.7.9),两个时间段内,SMTP 协议数据集的记录总数分别为 5 479 个用户 63 000 000 个会话和 5 389 个用户 5 000 000 个会话.

文献[7]是电子商务领域的网络水军识别研究,该文的数据集是从评论网站 www.resellerratings.com 收集的大量评论信息,其中包括评论者账号、评论内容、评论有用度分数以及评论发表时间和评论所在商店链接等属性.所抽取的评论数据都是在 2010 年 10 月 6 日中发布的,数据集共有 343 603 名评论者、14 561 个商店及 408 470 条评论.

文献[8]是社交网络领域网络水军识别研究,该文使用的数据是从 2011 年 2 月到 3 月间 Twitter 上发布的信息,共 148 371 名用户的 267 551 条 tweet、4 317 161 名粉丝和其 963 181 名关注的用户.

文献[104]是对国内网络水军的识别发现,主要针对国内门户网站中形成虚假舆论场的大量网络水军.该文收集了国内门户网站新浪和搜狐中关于“3Q”大战的新闻报道及用户评论,作为分析网络水军的实验数据.该数据集包括从 2010 年 9 月 10 日~2010 年 11 月 21 日两个月内新浪的 22 篇新闻报道和搜狐 24 篇新闻报道下的所有用户评论数据.进行数据预处理后,该评论数据包括 53 723 条新浪评论和 115 491 条搜狐评论.该数据集包

括手机发布用户和极少发布用户,这两类发布者都不太可能是大规模网络水军,因此去除数据集中这两类用户及其评论后,该数据集包括新浪 552 个用户和 20 738 条评论以及搜狐 223 个用户和 1 220 条评论.新浪和搜狐数据集的差异来源于搜狐允许用户匿名评论而新浪不允许,因此处理后的搜狐数据集较小.该差异说明,网络水军仍倾向于尽可能的匿名评论.

本文作者研究团队的网络水军识别实验数据集为淘宝中手机产品的 100 万条评论数据,这些数据包括评论者 ID、评论内容、评论有用度分数、评论发表时间、评论目标商品等信息,数据抽取时间为 2013 年 4 月 12 日~4 月 20 日.

2.6.2 效用评价指标

网络水军识别结果的主要评价指标为其识别准确性,主要包括 Precision[3,10,11,16,19,21,25,75,82,97], Recall[3,10,11,16,19,21,82,97],Accuracy[1,16,22,71],$F1$[1,3,10,11,16,19,21,22,25,71,73,75,82,97],FPR[1,8,16,21,22,71,73,74,89,98,110], ROC[8,16,25,63,71,75,82,89,98,110],NDCG@N[12,68,81]等.其中,NDCG@N 属于人工标记排序准确性指标,而其他几种属于分类精确度指标.NDCG@N 是在人工标记数据集后对数据集进行相应目标函数排序的准确性衡量指标,即,排序中目标函数较高的数据位于目标函数较低的数据之前.Precision 和 Recall 是数据挖掘领域的标准评价指标,评价数据挖掘实验的准确率和召回率.$F1$ 值为 Precision 和 Recall 的调和平均值,是对实验准确率和召回率综合表现的评价指标.Accuracy 指标评价一个分类器的表现,即,正确分类数所占比例.FPR(false positive rate)(假阳性)指标也是衡量预测分类准确率的指标,即,错误判断所占比例.ROC(receiver operating characteristic curve)曲线,即,将 FPR(假阳性)和 TPR(真阳性)作为 x/y 坐标轴生成的曲线,是描述预测假阳性和真阳性综合表现的评价指标.网络水军识别评价方法来自机器学习领域,网络水军识别算法在训练集上训练得到算法相应参数,在测试集上评价算法的性能.由于网络水军识别研究中,缺少公开可用的数据集,因此,通常通过人工评价相应数据集作为实验结果的比较基准[12,68,81],以此来评价识别结果的有效性和实用性.人工评价方法是网络水军识别研究评价的一种重要方法,但使用人工评价方法时评价人工成本较高并且评价结果主观性较强.当前,各目标领域的网络水军识别研究取得了不同程度的进展,因各目标领域网络水军目的影响的不同,其隐蔽程度各异,因此各目标领域网络水军的识别率和识别精度各异,见表 2 和表 3(其中,*表示评价同时使用识别率和识别精度).

Table 2 Different detection rate of Web spammers' target domain

表 2 各目标领域网络水军识别研究的识别率对比

文章\评价指标	发表时间	目标领域	研究方法	识别率(%)
74	2012.4	大型商业博客	基于用户行为特征	98.50
91	2010.4	邮件机器人	基于环境特征	98.20
1	2008.4	Youtube	基于用户行为特征	98.10
6	2013.2	邮件	基于环境特征	98.00
2	2010.12	Facebook	基于用户行为特征	98.00
84*	2008.6	BookmarkingSys	基于用户关系特征	97.47
3	2010.7	Twitter	基于综合特征	96.40
22*	2011.11	邮件	基于用户关系特征	95.40
79	2007.3	邮件	基于用户关系特征	95.00

Table 3 Different detection accuracy of Web spammers' target domain

表 3 各目标领域网络水军识别研究的识别精度对比

文章\评价指标	发表时间	目标领域	研究方法	识别精度(%)
16	2010.7	MySpace	基于用户行为特征	99.21
76	2010.4	论坛机器人	基于用户行为特征	96.24
22*	2011.11	邮件	基于用户关系特征	93.00
25	2012.2	众包服务	基于用户行为特征	91.00
104	2011.11	Sina+Sohu 论坛	基于综合特征	88.79
84*	2008.6	BookmarkingSys	基于用户关系特征	86.11
19	2012.1	人人网	基于用户特征	85.10

传统的网络水军识别研究主要使用识别准确性来评价识别表现,而 Web 2.0 网络水军识别研究中的识别效

率也很重要.当前,网络水军识别需分析海量数据发现高度隐蔽的网络水军,数据处理速度也是一项重要评价指标.效用评价的目的是评价网络水军识别表现,若评价结果不理想或不能让用户满意,应根据具体评价指标来分析识别研究需要改进的地方,以达到更好的识别效果.

3 网络水军识别研究发展的热点与难点

社交网络爆炸性的发展和海量数据处理能力的增加,使得网络水军识别研究取得了一定的进展,但作为一个新兴研究领域,其中需要深入研究并可能取得一定进展的部分有很多,主要包括:

(1) 电子商务领域网络水军识别研究

电子商务是较早发展的新兴网络商业模式,在传统零售业被大部分信息化的同时,网络购物成为 Web 2.0 网络服务中的主要部分.辅助用户购买决策的电子商务信息,成为网络水军的主要危害目标.准确定位电子商务网络水军,能够保证良好的电子商务环境,维护用户的切身经济利益.因与用户切身经济利益紧密相关,电子商务领域的网络水军识别研究较社交领域网络水军识别研究更为迫切.与此同时,电子商务网络水军识别研究因其独有的特点也较其他领域网络水军识别研究更为复杂:首先,电子商务领域用户信息较其他领域更为敏感,其直接涉及用户消费信息,关系到用户的切身经济利益,这些敏感的用户信息多掌握在大型电子商务公司中,网络水军识别研究必然需要以该类数据为基础,因此,准确权威的研究数据来源也成为困扰广大电子商务领域网络水军研究者的首要问题,该类研究数据同时有可能涉及用户的隐私安全信息,如何在保护用户隐私信息的同时保证研究的高效进行,是电子商务领域网络水军识别研究的首要问题;其次,近年来,电子商务领域网络水军盛行的一大原因即商家与网络水军的大量合作,在经济利益驱动下,越来越多的商家雇佣网络水军来影响商品的正常买卖,这因此也造成电子商务领域中不真实信息的日益增多;此外,电子商务领域网络水军识别研究的另一大难点即识别结果的评价,电子商务网络水军多具有与普通用户相似的特点,仅通过人工专家评价验证其识别结果具有很大的主观性.近年来,电子商务领域网络水军识别研究利用不同实验数据多角度挖掘网络水军行为特征,利用识别特征的不同对实验结果进行交叉对比,从而验证网络水军识别结果的准确性.该类方法弥补了人工评价的不足,对电子商务网络水军识别的评价具有重大意义.但与其他领域网络水军识别评价体系相比,电子商务网络水军的识别评价体系仍有待完善.

(2) 社交网络领域网络水军识别研究

社交网络中的水军识别是近几年网络水军识别研究的热点之一,由于社交网络蕴含着丰富的用户兴趣群体,潜藏着巨大商机,因此,社交网络领域成为网络水军重点侵袭的目标.社交网络中水军的识别和防范能够为用户提供一个良好的无内容污染的社交平台,维护合理的网络秩序,营造良好的网络环境.社交网络中水军识别必然是网络水军识别研究的重点和热门领域,且对其研究的突破能够为社交网络带来重大改变.

(3) 网络水军团体的发现

Web 2.0 网络水军与传统个人网络水军相比,已形成一定规模,并在网络水军间形成一定的组织结构.网络中普遍存在提供大规模网络水军服务的团体与组织,如何从海量数据中发现这些危害极高的网络水军团体,是网络水军识别研究极其重要的发展方向[4].这些网络水军团体之间具有紧密的联系,识别网络水军团体关键技术的突破能够极大程度地遏制网络中泛滥的水军数量.

(4) 网络水军特征的定义

网络水军识别研究中水军特征的定义是关键,识别特征的定义关系到网络水军识别的表现.由于不同目标领域网络水军行为不尽相同,因此网络水军表现出的特征也不尽相同.分析实验数据定义分辨力较高的网络水军特征,并提高网络水军识别准确率,是网络水军识别研究的难点之一.

(5) 网络水军识别结果的评价

网络水军识别表现主要由其评价指标来衡量.传统的数据挖掘研究的评价指标,如 Precision,Recall,Accuracy,ROC 等也被用来衡量网络水军识别的表现.为了评价网络水军识别结果,研究者经常使用人工评价方法作为评价标准.但是人工评价需要花费较大成本,而且具有一定主观性.因此,如何有效地评价网络水军识别

表现,是网络水军识别中需要研究的问题之一.

(6) 网络水军识别结果的解释

在传统网络水军识别的研究中,很少有研究人员关注网络水军识别结果的解释,但这方面的研究对于改进网络水军识别表现很重要.有效、正确的识别结果解释,可以使评价人员了解网络水军识别的工作流程,提高评价人员标记网络水军的准确率.同时,对使用网络水军识别结果的用户来说,合理的结果解释能够帮助用户评价网络水军识别的实用性,从而获得高质量的用户反馈,改善网络水军识别表现.

(7) 互联网隐私和安全问题

Web 2.0 互联网环境中,用户的隐私保护和信息安全问题[111−113]是网络水军识别研究发展的制约.网络水军识别研究是为了定位网络中存在的大量水军并对其采取防治措施,因此,网络水军识别研究必须分析用户产生内容、行为、关系等个人信息.但出于隐私与信息安全的考虑,用户不愿意为网络水军识别研究提供完整和准确的个人信息,因此,实验数据分析如何避免涉及用户个人信息并准确定位网络水军,是网络水军识别研究中的一个难点.

4 结束语

网络的急速发展和人们日常生活的高度信息化,使得网络服务已经成为人们获取信息、进行社会活动的一个主要方式.网络水军污染网络环境,扰乱网络秩序,妨害用户商业利益.而网络水军识别研究作为解决这一问题的有效手段,在数据挖掘领域得到了广泛关注和研究,并且由于网络水军妨害商业利益,能够为组织和个人带来经济效益,近年来其研究也得到了工业界的广泛关注和应用.同时,Web 2.0 网络水军识别研究仍然存在大量问题需要深入研究,以提高网络水军识别的准确性和实用性.因此,网络水军识别研究具有极其重要的研究意义和十分广阔的应用前景.

致谢 在此,我们向对本文的工作给予支持和建议的同行表示感谢.

References:

[1] Benevenuto F, Rodrigues T, Almeida V, Almeida J, Zhang C, Ross K. Identifying video spammers in online social networks. In: Castillo C, Chellapilla K, Fetterly D, eds. Proc. of the 4th Int'l Workshop on Adversarial Information Retrieval on the Web (AIRWeb 2008). New York: ACM Press, 2008. 45−52. [doi: 10.1145/1451983.1451996]

[2] Stringhini G, Kruegel C, Vigna G. Detecting spammers on social networks. In: Gates C, Franz M, McDermott J, eds. Proc. of the 26th Annual Computer Security Applications Conf. (ACSAC 2010). New York: ACM Press, 2010. 1−9. [doi: 10.1145/1920261.1920263]

[3] Benevenuto F, Magno G, Rodrigues T, Almeida V. Detecting spammers on Twitter. In: Proc. of the 7th Annual Collaboration Electronic Messaging, Anti-Abuse and Spam Conf. (CEAS 2010), Vol.6. 2010. 12−20. http://ceas.cc/2010/

[4] Mukherjee A, Liu B, Glance N. Spotting fake reviewer groups in consumer reviews. In: Mille A, Gandon F, Misselis J, Rabinovich M, Staab S, eds. Proc. of the 21st Int'l Conf. on World Wide Web (WWW 2012). New York: ACM Press, 2012. 191−200. [doi: 10.1145/2187836.2187863]

[5] Gammereldin SEA, Musa MEM. Analyzing and revealing spammer accounts in email social network: SUST Mail case. In: Proc. of the 2nd Int'l Conf. on Information and Communication Technology. Khartoum: Ministry of Communications and Information Technology, 2011. 3−11.

[6] Las-Casas PHB, Guedes D, Almeida JM, Ziviani A, Marques-Neto HT. SpaDeS: Detecting spammers at the source network. Computer Networks, 2012,57(2):526−539. [doi: 10.1016/j.comnet.2012.07.015]

[7] Wang G, Xie S, Liu B, Yu PS. Review graph based online store review spammer detection. In: Cook D, Pei J, Wang W, Zaiane O, Wu X, eds. Proc. of the 11th Int'l Conf. on Data Mining (ICDM 2011). Washington: IEEE Computer Society, 2011. 1242−1247. [doi: 10.1109/ICDM.2011.124]

[8] Song J, Lee S, Kim J. Spam filtering in Twitter using sender-receiver relationship. In: Sommer R, Balzarotti D, Maier G, eds. Proc. of the 14th Int'l Symp. on Recent Advances in Intrusion Detection (RAID 2011). Heidelberg: Springer-Verlag, 2011. 301−317. [doi: 10.1007/978-3-642-23644-0_16]

[9] Zinman A, Donath J. Is britney spears spam. In: Proc. of the 4th Conf. on Email and Anti-Spam (CEAS 2007). 2007. 1−10. http://ceas.cc/2007/

[10] Wang AH. Don't follow me: Spam detection in Twitter. In: Katsikas S, Samarati P, eds. Proc. of the 2010 Int'l Conf. on Security and Cryptography (SECRYPT 2010). Washington: IEEE Computer Society, 2010. 1−10.

[11] Xu C, Zhang J, Chang K, Long C. Uncovering collusive spammers in Chinese review Websites. In: He Q, Iyengar A, Nejdl W, eds. Proc. of the 22nd ACM Conf. on Information and Knowledge Management (CIKM 2013). New York: ACM Press, 2013. 979−988. [doi: 10.1145/2505515.2505700]

[12] Lim EP, Nguyen VA, Jindal N, Liu B, Lauw HW. Detecting product review spammers using rating behaviors. In: Huang J, Koudas N, Jones G, Wu X, Collins-Thompson K, An A, eds. Proc. of the 19th ACM Int'l Conf. on Information and Knowledge Management (CIKM 2010). New York: ACM Press, 2010. 939−948. [doi: 10.1145/1871437.1871557]

[13] Gilbert E, Karahalios K. Understanding deja reviewers. In: Inkpen K, Gutwin C, Tang J, eds. Proc. of the ACM Conf. on Computer Supported Cooperative Work (CSCW 2010). New York: ACM Press, 2010. 225−228. [doi: 10.1145/1718918.1718961]

[14] Akoglu L, Chandy R, Faloutsos C. Opinion fraud detection in online reviews by network effects. In: Kiciman E, Ellison NB, Hogan B, eds. Proc. of the 7th Int'l Conf. on Weblogs and Social Media (ICWSM 2013). Menlo Park: AAAI Press, 2013. 2−11.

[15] Webb S, Caverlee J, Pu C. Social honeypots: Making friends with a spammer near you. In: Proc. of the 5th Conf. on Email and Anti-Spam (CEAS 2008). 2008. 1−10. http://ceas.cc/2008/

[16] Lee K, Caverlee J, Webb S. Uncovering social spammers: social honeypots+ machine learning. In: Crestani F, Marchand-Maillet S, Chen HH, Efthimiadis EN, Savoy J, eds. Proc. of the 33rd Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval (SIGIR 2010). New York: ACM Press, 2010. 435−442. [doi: 10.1145/1835449.1835522]

[17] Ghosh S, Korlam G, Ganguly N. Spammers' networks within online social networks: A case-study on Twitter. In: Sadagopan S, Ramamritham K, Kumar A, Ravindra MP, Bertino E, Kumar R, eds. Proc. of the 20th Int'l Conf. on World Wide Web (WWW 2011). New York: ACM Press, 2011. 41−42. [doi: 10.1145/1963192.1963214]

[18] Duan Z, Chen P, Sanchez F, Dong Y, Stephenson M, Barker JM. Detecting spam zombies by monitoring outgoing messages. IEEE Trans. on Dependable and Secure Computing (TDSC 2012), 2012,9(2):198−210. [doi: 10.1109/TDSC.2011.49]

[19] Zhu Y, Wang X, Zhong E, Liu NN, Li H, Yang Q. Discovering spammers in social networks. In: Hoffmann J, Selman B, eds. Proc. of 26th AAAI Conf. on Artificial Intelligence (AAAI 2012). Menlo Park: AAAI Press. 2012. 1−7.

[20] Lu Y, Zhang L, Xiao Y, Li Y. Simultaneously detecting fake reviews and review spammers using factor graph model. In: Davis HC, Halpin H, Pentland A, eds. Proc. of the 5th Annual ACM Web Science Conf. (WebSci 2013). New York: ACM Press, 2013. 225−233. [doi: 10.1145/2464464.2464470]

[21] Hayati P, Chai K, Potdar V, Talevski A. Behaviour-Based Web spambot detection by utilising action time and action frequency. In: Taniar D, Gervasi O, Murgante B, Pardede E, Apduhan BO, eds. Proc. of the Computational Science and Its Applications (ICCSA 2010). Heidelberg: Springer-Verlag, 2010. 351−360. [doi: 10.1007/978-3-642-12165-4_28]

[22] Bouguessa M. An unsupervised approach for identifying spammers in social networks. In: Proc. of the 23rd IEEE Int'l Conf. on Tools with Artificial Intelligence (ICTAI 2011). Washington: IEEE Computer Society, 2011. 832−840. [doi: 10.1109/ICTAI.2011.130]

[23] Liu QW. Web spammers' detection and prevention. Press Outpost, 2012,6:021 (in Chinese with English abstract).

[24] Halpin H, Blanco R. Machine-Learning for spammer detection in crowd-sourcing. Human Computation AAAI Technical Report, WS-12-08, Menlo Park: AAAI Press, 2012. 85−86.

[25] Raykar VC, Yu S. Eliminating spammers and ranking annotators for crowdsourced labeling tasks. The Journal of Machine Learning Research, 2012,13:491−518.

[26] Langbehn HR, Ricci S, Gonçalves MA, Almeida JM, Pappa GL, Benevenuto F. A multi-view approach for detecting non-cooperative users in online video sharing systems. Journal of Information and Data Management, 2010,1(3):313−328.

[27] Russell MA. Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub, and More. 2nd ed., Sebastopol: O'Reilly Media, 2013. 5−43.

[28] Jiang F, Du JW, Sui YF, Cao CG. Outlier detection based on boundary and distance. Acta Electronica Sinica, 2010,38(3):700−705 (in Chinese with English abstract).

[29] Zhang L, Zhu J, Yao T. An evaluation of statistical spam filtering techniques. ACM Trans. on Asian Language Information Processing (TALIP), 2004,3(4):243−269. [doi: 10.1145/1039621.1039625]

[30] Murmann AJ. Enhancing spammer detection in online social networks with trust-based metrics [MS. Thesis]. San Jose: San Jose State University, 2009.

[31] Hayati P, Chai K, Potdar V, Talevski A. HoneySpam 2.0: Profiling Web spambot behaviour. In: Proc. of the Principles of Practice in Multi-Agent Systems. Heidelberg: Springer-Verlag, 2009. 335−344. [doi: 10.1007/978-3-642-11161-7_23]

[32] Wang JH. Social Network Analysis for E-mail Filtering. 2006. 69−78. http://www.im.cpu.edu.tw/cyber06/cyber06-a6.pdf

[33] Chang M, Yih W, Meek C. Partitioned logistic regression for spam filtering. In: Li Y, Liu B, Sarawagi S, eds. Proc. of the 14th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Minging (KDD 2008). New York: ACM Press, 2008. 97−105. [doi: 10.1145/1401890.1401907]

[34] Balaguer EV, Rosso P. Detection of near-duplicate user generated contents: The SMS spam collection. In: Proc. of the 3rd Int'l Workshop on Search and Mining User-Generated Contents (SMUC 2011). New York: ACM Press, 2011. 27−34. [doi: 10.1145/2065023.2065031]

[35] Sathawane KS, Tuteja RR. A robust spam detection system using a collaborative approach with an E-mail abstraction scheme and spam tree data structure. Int'l Journal of Computer Science and Applications, 2013,6(2):293−298.

[36] Tseng CY, Sung PC, Chen MS. Cosdes: A collaborative spam detection system with a novel e-mail abstraction scheme. IEEE Trans. on Knowledge and Data Engineering, 2011,23(5):669−682. [doi: 10.1109/TKDE.2010.147]

[37] Maan G, Tak G. Enhanced discussion on different techniques of spam detection. Int'l Journal of Computer and Technology, 2013, 4(2):248−253.

[38] Wirtz BW, Mory L, Piehler R. Web 2.0 and digital business models. In: Martinez-Lopez FJ, ed. Handbook of Strategic e-Business Management. Heidelberg: Springer-Verlag, 2014. 751−766. [doi: 10.1007/978-3-642-39747-9_31]

[39] Hayati P, Potdar V, Talevski A, Firoozeh N, Sarenche S, Yeganeh EA. Definition of spam 2.0: New spamming boom. In: Hussain FK, Chang E, eds. Proc. of the 4th IEEE Int'l Conf. on Digital Ecosystems and Technologies (IEEE DEST 2010). Washington: IEEE Computer Society, 2010. 580−584. [doi: 10.1109/DEST.2010.5610590]

[40] Hayati P, Potdar V. Spam 2.0 state of the art. Int'l Journal of Digital Crime and Forensics, 2012,4(1):17−36.

[41] Ridzuan F, Potdar V, Talevski A, Smyth WF. Key parameters in identifying cost of spam 2.0. In: Proc. of the 24th IEEE Int'l Conf. on Advanced Information Networking and Applications (AINA 2010). Washington: IEEE Computer Society, 2010. 789−796. [doi: 10.1109/AINA.2010.163]

[42] Sriram B, Fuhry D, Demir E, Ferhatosmanoglu H, Demirbas M. Short text classification in Twitter to improve information filtering. In: Crestani F, Marchand-Maillet S, Chen HH, eds. Proc. of the 33rd Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval (SIGIR 2010). New York: ACM Press, 2010. 841−842. [doi: 10.1145/1835449.1835643]

[43] Zhao YY, Qin B, Liu T. Sentiment analysis. Ruan Jian Xue Bao/Journal of Software, 2010,21(8):1834−1848 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/3832.htm [doi: 10.3724/SP.J.1001.2010.03832]

[44] Liu B. Sentiment analysis and subjectivity. In: Indurkhya N, Damerau FJ, eds. Handbook of Natural Language Processing. Boca Raton: CRC Press, 2010. 627−666.

[45] Prince MB, Holloway L, Langheinrich E, Dahl BM, Keller AM. Understanding how spammers steal your e-mail address: An analysis of the first six months of data from project honey pot. In: Proc. of the 2nd Conf. on Email and Anti-Spam (CEAS 2005). 2005. 1−7. http://ceas.cc/2005/

[46] Hayati P, Potdar V. Evaluation of spam detection and prevention frameworks for email and image spam—A state of art. In: Kotsis G, Taniar D, Pardede E, Khalil I, eds. Proc. of the 10th Int'l Conf. on Information Integration and Web-Based Applications & Services (iiWAS 2008). New York: ACM Press, 2008. 520−527. [doi: 10.1145/1497308.1497402]

[47] Chen ZX. Survey on spam filtering technology. Application Research of Computers, 2009,26(5):1612−1615 (in Chinese with English abstract).

[48] Almeida TA, Yamakami A, Almeida J. Probabilistic anti-spam filtering with dimensionality reduction. In: Shin SY, Ossowski S, Schumacher M, eds. Proc. of the 2010 ACM Symp. on Applied Computing (SAC 2010). New York: ACM Press, 2010. 1802−1806. [doi: 10.1145/1774088.1774469]

[49] Almeida TA, Yamakami A. Content-Based spam filtering. In: Proc. of the 2010 Int'l Joint Conf. on Neural Networks (IJCNN 2010). Washington: IEEE Computer Society, 2010. 1−7. [doi: 10.1109/IJCNN.2010.5596569]

[50] Li Z, Shen H. Soap: A social network aided personalized and effective spam filter to clean your e-mail box. In: Proc. of the 30th IEEE Int'l Conf. on Computer Communications (INFOCOM 2011). Washington: IEEE Computer Society, 2011. 1835−1843. [doi: 10.1109/INFCOM.2011.5934984]

[51] Lu KS, Chang CK. Using Web search results and genetic algorithm to improve the accuracy of Chinese spam email filters. In: Proc. of the 35th Annual IEEE Int'l Workshop on Computer Software and Applications Conf. (COMPSACW 2011). Washington: IEEE Computer Society, 2011. 286−291. [doi: 10.1109/COMPSACW.2011.56]

[52] Xu H, Yu B. Automatic thesaurus construction for spam filtering using revised back propagation neural network. Expert Systems with Applications, 2010,37(1):18−23. [doi: 10.1016/j.eswa.2009.02.059]

[53] Duh A, Stiglic G, Korosak D. Enhancing identification of opinion spammer groups. In: Proc. of the Int'l Conf. on Making Sense of Converging Media (AcademicMindTrek 2013). New York: ACM Press, 2013. 326−328. [doi: 10.1145/2523429.2523437]

[54] Lau RYK, Liao SY, Kwok RCW, Xu K, Xia Y, Li Y. Text mining and probabilistic language modeling for online review spam detection. ACM Trans. on Management Information Systems (TMIS), 2011,2(4):25.

[55] Li F, Huang M, Yang Y, Zhu X. Learning to identify review spam. In: Walsh T, ed. Proc. of the 22nd Int'l Joint Conf. on Artificial Intelligence (IJCAI 2011), Vol.3. Menlo Park: AAAI Press, 2011. 2488−2493. [doi: 10.5591/978-1-57735-516-8/IJCAI11-414]

[56] Liu HY, Zhao YY, Qin B, Liu T. Comment target extraction and sentiment classification. Journal of Chinese Information Processing, 2010,24(1):84−88 (in Chinese with English abstract).

[57] Jindal N, Liu B, Lim EP. Finding unusual review patterns using unexpected rules. In: Huang J, Koudas N, Jones G, eds. Proc. of the 19th ACM Int'l Conf. on Information and Knowledge Management (CIKM 2010). New York: ACM Press, 2010. 1549−1552. [doi: 10.1145/1871437.1871669]

[58] Ott M, Choi Y, Cardie C, Hancock JT. Finding deceptive opinion spam by any stretch of the imagination. In: Proc. of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Vol.1. Stroudsburg: ACL, 2011. 309−319.

[59] Niu Y, Wang YM, Chen H, Ma M, Hsu F. A quantitative study of forum spamming using context-based analysis. In: Proc. of the Network and Distributed System Security Symp. (NDSS 2007). 2007. 1−14. http://www.internetsociety.org/events/ndss-symposium-2007

[60] Mo Q, Zhang YJ, Hu HL, Zhang HP. Combined approach of polarity analysis on stock analyst. Computer Engineering and Applications, 2011,47(19):222−225 (in Chinese with English abstract).

[61] Hu HL, Mo Q. Research on opinion classification of stock recommendations based on discourse structure. Journal of Chinese Computer Systems, 2009,30(5):899−902 (in Chinese with English abstract).

[62] Bao J, Ji C, Gao M. Research on network security of defense based on Honeypot. In: Proc. of the 2010 Int'l Conf. on Computer Application and System Modeling (ICCASM 2010). Washington: IEEE Computer Society, 2010. 299−302. [doi: 10.1109/ICCASM.2010.5622780]

[63] Xie Y, Yu F, Achan K, Panigrahy R, Hulten G, Osipkov I. Spamming botnets: Signatures and characteristics. In: Bahl V, Wetherall D, Savage S, Stoica I, eds. Proc. of the ACM SIGCOMM 2008 Conf. on Data Communication (SIGCOMM 2008). New York: ACM Press, 2008. 171−182. [doi: 10.1145/1402958.1402979]

[64] Bhat VH, Malkani VR, Shenoy PD, Venugopal KR, Patnaik LM. Classification of email using BeaKS: Behavior and keyword stemming. In: Proc. of the Int'l Technical Conf. of IEEE Asia Pacific Region (IEEE TENCON 2011). Washington: IEEE Computer Society, 2011. 1139−1143. [doi: 10.1109/TENCON.2011.6129290]

[65] Stringhini G, Egele M, Kruegel C, Vigna G. Breaking the loop: Leveraging botnet feedback for spam mitigation. In: Proc. of the 7th Annual Graduate Student Workshop on Computing (GSWC 2012). Santa Barbara: University of California, 2012. 25−26.

[66] Husna H, Phithakkitnukoon S, Palla S, Dantu R. Behavior analysis of spam botnets. In: Proc. of the 3rd Int'l Conf. on Communication Systems Software and Middleware and Workshops (COMSWARE 2008). Washington: IEEE Computer Society, 2008. 246−253. [doi: 10.1109/COMSWA.2008.4554418]

[67] Sawaya Y, Kubota A, Yamada A. Understanding the time-series behavioral characteristics of evolutionally advanced email spammers. In: Yu T, Venkatakrishan VN, Kapadia A, eds. Proc. of the 5th ACM Workshop on Security and Artificial Intelligence (AISec 2012). New York: ACM Press, 2012. 71−80. [doi: 10.1145/2381896.2381908]

[68] Qiu YF, Wang JK, Shao LS, Liu DY. Research on product review spammer detection based on users' behavior. Computer Engineering, 2012,38(11):254−261 (in Chinese with English abstract).

[69] Mukherjee A, Kumar A, Liu B, Wang J, Hsu M, Castellanos M, Ghosh R. Spotting opinion spammers using behavioral footprints. In: Dhilon IS, Koren Y, Ghani R, eds. Proc. of the 19th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining (KDD 2013). New York: ACM Press, 2013. 632−640. [doi: 10.1145/2487575.2487580]

[70] Parameswaran M, Rui H, Sayin S. A game theoretic model and empirical analysis of spammer strategies. In: Proc. of the 7th Annual Collaboration, Electronic Messaging, Anti-Abuse and Spam Conf. (CEAS 2010), Vol.7. 2010. 1−7. http://ceas.cc/2010/

[71] Lee K, Eoff BD, Caverlee J. Seven months with the devils: A long-term study of content polluters on Twitter. In: Adamic LA, Baeza-Yates RA, Counts S, eds. Proc. of the 5th Int'l Conf. on Weblogs and Social Media (ICWSM). Menlo Park: AAAI Press, 2011. 185−192.

[72] Sureka A. Mining user comment activity for detecting forum spammers in YouTube. arXiv preprint arXiv: 1103.5044, 2011.

[73] Yang C, Harkreader RC, Gu G. Die free or live hard? Empirical evaluation and new design for fighting evolving Twitter spammers. In: Sommer R, Balzarotti D, Maier G, eds. Proc. of the 14th Int'l Symp. on Recent Advances in Intrusion Detection (RAID 2011). Heidelberg: Springer-Verlag, 2011. 318−337. [doi: 10.1007/978-3-642-23644-0_17]

[74] Tan E, Guo L, Chen S, Zhang X, Zhao Y. Spammer behavior analysis and detection in user generated content on social networks. In: Proc. of the 32nd Int'l Conf. on Distributed Computing Systems (ICDCS 2012). Washington: IEEE Computer Society, 2012. 305−314. [doi: 10.1109/ICDCS.2012.40]

[75] Gargari SM, Oguducu SG. A novel framework for spammer detection in social bookmarking systems. In: Proc. of the IEEE/ACM Int'l Conf. on Advances in Social Networks Analysis and Mining (ASONAM 2012). Washington: IEEE Computer Society, 2012. 827−834. [doi: 10.1109/ASONAM.2012.150]

[76] Hayati P, Potdar V, Chai K, Talevski A. Web spambot detection based on Web navigation behaviour. In: Proc. of the 24th IEEE Int'l Conf. on Advanced Information Networking and Applications (AINA 2010). Washington: IEEE Computer Society, 2010. 797−803. [doi: 10.1109/AINA.2010.92]

[77] Hayati P, Potdar V, Talevski A, Chai K. Characterisation of Web spambots using self organising maps. Int'l Journal of Computer Systems Science & Engineering, 2011,26(2):87−96.

[78] Hayati P, Potdar V, Talevski A, Smyth WF. Rule-Based on-the-fly Web spambot detection using action strings. In: Proc. of Collaboration Electronic Messaging Anti-Abuse and Spam Conf. (CEAS 2010). 2010. 13−14. http://ceas.cc/2010/

[79] Brendel R, Krawczyk H. Application of social relation graphs for early detection of transient spammers. WSEAS Trans. on Information Science & Applications, 2008,5(3):267−276.

[80] Sadan Z, Schwartz DG. Social network analysis of Web links to eliminate false positives in collaborative anti-spam systems. Journal of Network and Computer Applications, 2011,34(5):1717−1723. [doi: 10.1016/j.jnca.2011.06.004]

[81] Wang G, Xie S, Liu B, Yu PS. Identify online store review spammers via social review graph. ACM Trans. on Intelligent Systems and Technology (TIST), 2012,3(4):61. [doi: 10.1145/2337542.2337546]

[82] Moh TS, Murmann AJ. Can you judge a man by his friends? Enhancing spammer detection on the Twitter microblogging platform using friends and followers. In: Prasad SK, Vin HM, Sahni S, eds. Proc. of the Int'l Conf. on Information Systems and Technology Management (ICISTM 2010). Heidelberg: Springer-Verlag, 2010. 210−220. [doi: 10.1007/978-3-642-12035-0_21]

[83] Gayo-Avello D, Brenes DJ. Overcoming spammers in Twitter—A tale of five algorithms. In: Proc. of the Spanish Conf. on Information Retrieval (CERI 2010). 2010. 41−52. http://ir.ii.uam.es/ceri2010/en/

[84] Krestel R, Chen L. Using co-occurrence of tags and resources to identify spammers. In: Saeys Y, Liu H, Inza I, eds. Proc. of the Discovery Challenge Workshop at the European Conf. on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD 2008). Brookline: Microtome Publishing, 2008. 38−46.

[85] Bhat SY, Abulaish M. Community-Based features for identifying spammers in online social networks. In: Rokne JG, Faloutsos C, eds. Proc. of the 2013 IEEE/ACM Int'l Conf. on Advances in Social Networks Analysis and Mining (ASONAM 2013). New York: ACM Press, 2013. 100−107. [doi: 10.1145/2492517.2492567]

[86] Mukherjee A, Venkataraman V, Liu B, Glance N. What yelp fake review filter might be doing. In: Kiciman E, Ellison NB, Hogan B, eds. Proc. of the 7th Int'l AAAI Conf. on Weblogs and Social Media (ICWSM 2013). Menlo Park: AAAI Press, 2013. 409−418.

[87] Ramachandran A, Feamster N. Understanding the network-level behavior of spammers. In: Rizzo L, Anderson T, McKeown N, eds. Proc. of the 2006 Conf. on Applications, Technologies, Architectures, and Protocols for Computer Communications (SIGCOMM 2006). New York: ACM Press, 2006. 291−302. [doi: 10.1145/1151659.1159947]

[88] Tseng CY, Huang JW, Chen MS. ProMail: Using progressive email social network for spam detection. In: Zhou Z, Li H, Yang Q, eds. Proc. of the 11th Pacific-Asia Conf. on Advances in Knowledge Discovery and Data Mining (PAKDD 2007). Heidelberg: Springer-Verlag, 2007. 833−840. [doi: 10.1007/978-3-540-71701-0_92]

[89] Hao S, Syed NA, Feamster N, Gray AG, Krasser S. Detecting spammers with SNARE: Spatio-Temporal network-level automatic reputation engine. In: Proc. of the 18th USENIX Security Symp. Berkeley: USENIX Association, 2009. 101−118.

[90] Uddin AHMM. Detecting botnets based on their behaviors perceived from netflow data. Tartu: University of Tartu, 2009. 1−17. http://courses.cs.ut.ee

[91] Ehrlich WK, Karasaridis A, Liu D, Hoeflin D. Detection of spam hosts and spam bots using network flow traffic modeling. In: Proc. of the 3rd USENIX Conf. on Large-Scale Exploits and Emergent Threats: Botnets, Spyware, Worms, and More. Berkeley: USENIX Association, 2010. 7−15.

[92] Schatzmann D, Burkhart M, Spyropoulos T. Inferring spammers in the network core. In: Moon SB, Teixeira R, Uhlig S, eds. Proc. of the 10th Int'l Conf. on Passive and Active Network Measurement (PAM 2009). Heidelberg: Springer-Verlag, 2009. 229−238. [doi: 10.1007/978-3-642-00975-4_23]

[93] Xu KS, Kliger M, Hero III AO. Identifying spammers by their resource usage patterns. In: Proc. of the 7th Annual Collaboration Electronic Messaging, Anti-Abuse and Spam Conf. (CEAS 2010). 2010.

[94] Jiang J, ZhuGe JW, Duan HX, Wu JP. Research on botnet mechanisms and defenses. Ruan Jian Xue Bao/Journal of Software, 2012, 23(1):82−96 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/4101.htm [doi: 10.3724/SP.J.1001.2012.04101]

[95] Li F, Hsieh MH. An empirical study of clustering behavior of spammers and group-based anti-spam strategies. In: Proc. of the 3rd Conf. on Email and Anti-Spam (CEAS 2006). 2006. http://ceas.cc/2006/

[96] Xu C. Detecting collusive spammers in online review communities. In: Proc. of the 6th Ph.D. Students Workshop on Information and Knowledge Management (PIKM 2013). New York: ACM Press, 2013. 33−40. [doi: 10.1145/2513166.2513176]

[97] Benevenuto F, Rodrigues T, Almeida V, Almeida J, Goncalves M. Detecting spammers and content promoters in online video social networks. In: Allan J, Aslam J, Sanderson M, Zhai C, Zobel J, eds. Proc. of the 32nd Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval (SIGIR 2009). New York: ACM Press, 2009. 620−627. [doi: 10.1145/1571941.1572047]

[98] Amleshwaram AA, Reddy N, Yadav S, Gu G, Yang C. CATS: Characterizing automation of Twitter spammers. In: Proc. of the 5th Int'l Conf. on Communication Systems and Networks (COMSNETS 2013). Washington: IEEE Computer Society, 2013. 1−10. [doi: 10.1109/COMSNETS.2013.6465541]

[99] Lin C, Zhou Y, Chen K, He J, Yang X, Song L. Analysis and identification of spamming behaviors in Sina Weibo microblog. In: Zhu F, He Q, Yan R, eds. Proc. of the 7th Workshop on Social Network Mining and Analysis (SNAKDD 2013). New York: ACM Press, 2013. 5−13. [doi: 10.1145/2501025.2501035]

[100] Yang C, Harkreader R, Gu G. Empirical evaluation and new design for fighting evolving Twitter spammers. IEEE Trans. on Information Forensics and Security, 2013,(8):1280−1293. [doi: 10.1109/TIFS.2013.2267732]

[101] Wang K, Xiao Y, Xiao Z. Detection of Internet water army in social network. In: Proc. of the 2014 Int'l Conf. on Computer, Communications and Information Technology (CCIT 2014). Amsterdam: Atlantis Press, 2014. 189−192. [doi: 10.2991/ccit-14.2014.50]

[102] Aggarwal A, Almeida J, Kumaraguru P. Detection of spam tipping behaviour on foursquare. In: Carr L, Laender AHF, Loscio BF, eds. Proc. of the 22nd Int'l Conf. on World Wide Web Companion (WWW 2013 Campanion). New York: ACM Press, 2013. 641−648.

[103] Lee YJ, Shim JM, Cho HG, Woo G. Detecting and visualizing the dispute structure of the replying comments in the Internet forum sites. In: Proc. of the 2010 Int'l Conf. on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC 2010). Washington: IEEE Computer Society, 2010. 456−463. [doi: 10.1109/CyberC.2010.90]

[104] Chen C, Wu K, Srinivasan V, Zhang X. Battling the Internet water army: Detection of hidden paid posters. In: Rokne JG, Faloutsos C, eds. Proc. of the 2013 IEEE/ACM Int'l Conf. on Advances in Social Networks Analysis and Mining (ASONAM 2013). New York: ACM Press, 2013. 116−120. [doi: 10.1145/2492517.2492637]

[105] Jindal N, Liu B. Opinion spam and analysis. In: Najork M, Broder AZ, Chakrabarti S, eds. Proc. of the 2008 Int'l Conf. on Web Search and Data Mining (WSDM 2008). New York: ACM Press, 2008. 219−230. [doi: 10.1145/1341531.1341560]

[106] Mukherjee A, Liu B, Wang J, Glance N, Jindal N. Detecting group review spam. In: Srinivasan S, Ramamritham K, Kumar A, eds. Proc. of the 20th Int'l Conf. Companion on World Wide Web. New York: ACM Press, 2011. 93−94. [doi: 10.1145/1963192.1963240]

[107] Xie S, Wang G, Lin S, Yu PS. Review spam detection via temporal pattern discovery. In: Yang Q, Agarwal D, Pei J, eds. Proc. of the 18th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining (KDD 2012). New York: ACM Press, 2012. 823−831. [doi: 10.1145/2339530.2339662]

[108] Fei G, Mukherjee A, Liu B, Hsu M, Castellanos M, Ghosh R. Exploiting burstiness in reviews for review spammer detection. In: Kiciman E, Ellison NB, Hogan B, eds. Proc. of the 7th Int'l AAAI Conf. on Weblogs and Social Media (ICWSM 2013). Menlo Park: AAAI Press, 2013. 175−184.

[109] Feng S, Xing L, Gogar A, Choi Y. Distributional footprints of deceptive product reviews. In: Breslin JG, Ellison NB, Shanahan JG, eds. Proc. of the 6th Int'l AAAI Conf. on Weblogs and Social Media (ICWSM 2012). Menlo Park: AAAI Press, 2012. 98−105.

[110] Amleshwaram AA. Spammer detection on online social networks. [Ph.D. Thesis]. Texas: Texas A&M University, 2012.

[111] Kizza JM. Ethical and Social Issues in the Information Age. 2nd ed., New York: Springer-Verlag, 2003. 255−280. [doi: 10.1007/b98842]

[112] West W, Pulimood SM. Analysis of privacy and security in HTML5 Web storage. Journal of Computing Sciences in Colleges, 2012, 27(3):80−87.

[113] Zhou Y, Chen K, Song L, Yang X, He J. Feature analysis of spammers in social networks with active honeypots: A case study of Chinese microblogging networks. In: Proc. of the 2012 IEEE/ACM Int'l Conf. on Advances in Social Networks Analysis and Mining (ASONAM 2012). Washington: IEEE Computer Society, 2012. 728−729. [doi: 10.1109/ASONAM.2012.133]

附中文参考文献:

[23] 刘秋文.网络水军的识别和防范.新闻前哨,2012,6:021.

[28] 江峰,杜军威,眭跃飞,曹存根.基于边界和距离的离群点检测.电子学报,2010,38(3):700−705.

[43] 赵妍妍,秦兵,刘挺.文本情感分析.软件学报,2010,21(8):1834−1848. http://www.jos.org.cn/1000-9825/3832.htm [doi: 10.3724/SP.J.1001.2010.03832]

[47] 陈志贤.垃圾邮件过滤技术研究综述.计算机应用研究,2009,26(5):1612−1615.

[56] 刘鸿宇,赵妍妍,秦兵,刘挺.评价对象抽取及其倾向性分析.中文信息学报,2010,24(1):84−88.

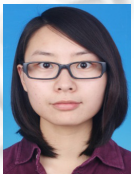
[60] 莫倩,张渝杰,胡航丽,张华平.一种混合的股评观点倾向性分析方法.计算机工程与应用,2011,47(19):222−225.

[61] 胡航丽,莫倩.利用篇章结构改进股评观点分类的研究.小型微型计算机系统,2011,30(5):899−902.

[68] 邱云飞,王建坤,邵良杉,刘大有.基于用户行为的产品垃圾评论者检测研究.计算机工程,2012,38(11):254−261.

[94] 江健,诸葛建伟,段海新,吴建平.僵尸网络机理与防御技术.软件学报,2012,23(1):82−96. http://www.jos.org.cn/1000-9825/4101.htm [doi: 10.3724/SP.J.1001.2012.04101]

莫倩(1972－),男,湖南东安人,博士,副教授,CCF 高级会员,主要研究领域为互联网信息挖掘,深度搜索引擎,知识管理.
E-mail: moqian@th.btbu.edu.cn



杨珂(1990－),女,硕士生,CCF 学生会员,主要研究领域为互联网信息挖掘,网络水军识别.
E-mail: christina900923@hotmail.com