社交网络异常用户识别技术综述

仲丽君,杨文忠,袁婷婷,向进勇
ZHONG Lijun, YANG Wenzhong, YUAN Tingting, XIANG Jinyong

新疆大学 信息科学与工程学院,乌鲁木齐 830046
College of Information Science and Engineering, Xinjiang University, Urumqi 830046, China

ZHONG Lijun, YANG Wenzhong, YUAN Tingting, et al. Survey of abnormal user identification technology in social network. Computer Engineering and Applications, 2018, 54(16):13-23.

Abstract: With the rapid development of the Internet, social network has become an important social tool in daily life. However, the abnormal users in social networks emerge in an endless stream, and its harm is becoming increasingly serious. Therefore, identifying and detecting abnormal users in social networks plays an important role in improving the user experience and maintaining a good network environment. This paper introduces different types of abnormal social network users, and introduces the research progress of each type of abnormal user. Finally, it summarizes the anomaly detection methods, and divides the anomaly detection technologies in social networks into categories, clustering, statistics, information theory, hybrid and graphs, and the advantages and disadvantages of these six technologies are compared, which can help people understand abnormal users and anomaly detection technologies in social networks, and provides ideas for solving abnormal problems.

Key words: social network; abnormal user; abnormity recognition technology; machine learning

摘 要: 随着互联网的迅速发展,社交网络已经成为人们日常生活中的重要社交工具。然而,社交网络中的异常用户层出不穷,其危害也日益严重。因此,识别和检测社交网络中的异常用户对提高用户体验、保持良好的网络环境等具有重要作用。介绍了不同类型的社交网络异常用户,并对每种不同类型异常用户的研究进展进行了介绍;对异常检测方法进行了综述,将社交网络中的异常检测技术分为分类、聚类、统计、信息论、混合、图六大类,并对这六类技术各自的优缺点进行了比较,有助于人们了解社交网络中的异常用户、异常检测技术,为解决异常问题提供了思路。

关键词: 社交网络;异常用户;异常识别技术;机器学习
文献标志码: A **中图分类号:** TP391 doi:10.3778/j.issn.1002-8331.1804-0374

1 引言

随着互联网的迅速发展,社交网络在不同应用领域日益增多,已经成为人们日常生活中的重要社交工具。但与此同时,伴随着社交媒体的爆炸式增长,许多非法用户将其作为牟取利益的平台,国外的Twitter、Facebook,以及国内的网易、新浪等在线社交系统的许多用户经常受到各种异常用户的困扰。一些用户在社交网络中大肆传播虚假消息、垃圾信息等有害信息,向用户推送虚假广告,进行恶意传销,扰乱了社交平台的正常营销和推广,侵犯了公众的利益,污染了社交网络环境,对社会造成了不良的影响。因此,识别这些异常用户是社交网络研究领域的一个重要课题,对净化网络环境,维持网络秩序,提高用户上网体验,促进社会的和谐发展等具有重要作用。

异常检测在许多应用领域得到了广泛的应用,如垃圾邮件识别、入侵检测、欺诈检测、金融诈骗、故障检测、恶意用户识别、用户异常情绪、人类行为分析、网络诈骗、医学和公共卫生、工业损伤、图像处理、传感器网络

基金项目: 国家自然科学基金(No.U1603115);国家重点基础研究发展规划(973)(No.2014CB340500);国家自然科学基金重点项目(No.U1435215)。

作者简介: 仲丽君(1992—),女,在读硕士,研究领域为自然语言处理、信息安全,E-mail:1666042670@qq.com;杨文忠(1971—),通讯作者,男,博士,副教授,研究领域为网络舆情、情报分析、信息安全、无线传感器网络;袁婷婷(1993—),女,在读硕士,研究领域为自然语言处理、信息安全;向进勇(1992—),男,在读硕士,研究领域为自然语言处理、舆情分析。

收稿日期: 2018-04-28 **修回日期:** 2018-07-02 **文章编号:** 1002-8331(2018)16-0013-11

等。Markou等人[1]于2003年在监督学习的基础上对异常检测的新颖性进行了研究。文献[2-3]提出了有监督学习、无监督学习以及基于聚类技术的各种异常检测方法,但是没有对网络中的异常进行检测。文献[4]提出了一种基于一个样本类和两个类别的支持向量机的混合异常检测机制。文献[5]对金融领域中基于聚类的异常识别方法进行了总结,并从不同的角度对其进行了比较。Fire等人[6]根据社交网络自身的拓扑特征,对多种类型社交网络中的恶意行为进行检测,可以检测出多种类型的恶意配置文件,取得了良好的性能。

目前有关社交网络异常用户的识别研究,没有一个统一的归类,相关的综述性文章多是对其中一种类型进行总结,没有一个系统的介绍。因此本文将社交网络中的异常用户划分为具体的7个类别,并对每种不同类型异常用户的研究现状进行概要的介绍。然后,对异常用户识别技术进行了详细的阐述,并对它们的优缺点进行了比较。

2 介绍

2.1 异常的定义

异常是一种由各种异常活动产生的与常规不符的现象或事件,随着时间的演变,不同的作者对异常有不同的定义,表1给出了一些最常用的定义。结合已有定义以及对社交网络异常产生方式的分析,本文中的异常是指在社交网络中,个人或群体的行为不符合正常模式定义的特征行为或与其同龄人以明显不同的方式进行的互动,其表现为在同一结构中与其他用户行为不同的活动,比如异常的观点、情绪、行为等。

2.2 社交网络异常用户的定义

针对社交网络中存在的异常用户类型多样、边界不清、目的不同(骚扰、广告、引导舆论走向、欺诈等)等问题,本文从异常用户在社交网络上具有的属性特征出发,总结归纳了不同种类的异常用户的特点,将社交网络异常用户划分为7种独立不相交的类别。

(1)恶意用户,发布内容含指向病毒网页、钓鱼网站或恶意网站的有害链接,往往会给人们造成经济损失。

(2)僵尸用户,自动或手动地生产大量“僵尸”账号,增加粉丝量,提高影响力,主要意图是追随、热捧那些迫切需要成为热门用户、热门话题的用户,绝大多数情况下它们不会向其他用户发布垃圾信息,暴露自己,而是尽可能地将自己伪装成正常用户。

(3)垃圾用户,主要是利用社交平台频繁发布不请自来的大量相似或相同的信息,通常是单个用户,一般只是传播一些垃圾信息,不会对公司、组织造成很大的伤害。

(4)虚假用户,大量注册的伪造账户,发布虚假信息、虚假评论。

(5)网络水军,为了达到某种目的(营销、推广、上热搜、政治、公关等),大量发表、回复、转发、评论、提及他人,使目的信息大量传播,散播谣言,混淆事实,影响人们的判断,引导舆论走势,网络水军通常具有很强的群体特征,行为也更隐蔽,不易被察觉。

(6)异常情绪用户,从用户发布内容映射的情绪状态,找出隐藏在大量文本中的特定时间内情绪发生突然变化的用户。

(7)不良言论用户,该类用户不满足上述6类异常用户的任何特征,看似一切正常,但是在有关民族、宗教、党建政策等大是大非的问题上,发表含有反动、反党反人民、辱华、煽动民族团结、支持三股势力等言论的极端思想用户。

表2给出了这些异常用户的相关定义。

3 社交网络异常用户识别研究现状

本文将社交网络异常用户划分为7个类别:恶意用户、僵尸用户、垃圾用户、虚假用户、网络水军、异常情绪用户、不良言论用户,并分别对这7类异常用户识别的研究现状进行了分析。

3.1 恶意用户

无论是国外的Twitter、Facebook,还是国内的网易、腾讯、新浪等,近年来都饱受恶意用户的困扰。他们发布恶意链接、钓鱼网站、传播病毒程序,污染了网络环境,严重威胁到人们的财产安全,许多专家学者对恶意

表1 异常的不同定义

时间	作者	异常定义
1969年	Grubbs[7]	离群观察,即离群值,是与样本中其他成员明显偏离的事物
1980年	Hawkins[8]	异常是一种偏离其他观测的观察,以至于人们怀疑它是由不同的机制产生的
1994年	Bamet和Lewis[7]	观察结果(或观察子集)似乎与该组数据的其余部分不一致
1995年	John[7]	异常值也被认为是令人惊讶的真实数据,比如属于 A 类的点却被放置在了 B 类中,该点的真实分类令观察者惊讶
2001年	Aggarwal和Yu[7]	离群值被认为是一组定义在簇外的噪声点,也可以定义为位置外同时也与噪声分离的点
2009年	Chandola等人[7]	数据中的模式,但不符合正确行为的定义
2013年	Mohammad等人[9]	社交网络中的异常被定义为偏离大多数观察的观察结果
2014年	Savage等人[7]	网络的结构不同于正常模型的结构
2017年	Anand等人[10]	社交网络中偏离大多数用户的突发或不规则行为

表2 异常用户的不同定义

异常用户类型	异常用户定义
恶意用户	实施恶意行为、传播恶意信息的攻击性用户
僵尸用户	也叫僵尸粉,不参与正常的社交生活,以追随、热捧其他用户、话题为主要目的的用户
垃圾用户	发布垃圾信息、不良信息的用户
虚假用户	批量注册生成的虚假账户
异常情绪用户	一定时期内情绪波动较大的用户
网络水军[11]	一群有着特殊目的的用户,通过水军账号或程序机器人,在电商网站、论坛、微博等社交网络平台有组织、有计划地发表、传播信息的一类网络写手
不良言论用户	在微博、论坛等社交平台发表关于意识形态、社会制度、宗教、民族、地区等不利于社会稳定消息的用户

用户的识别展开了大量的研究。Stringhini等人[12]采用随机森林算法对三大社交网络上的恶意用户进行检测,取得了较好的分类准确率。Chu等人[13]使用贝叶斯对Twitter上发布恶意内容的用户进行检测。Zheng等人[14]使用支持向量机对新浪微博上的恶意用户进行检测,但时间开销较大,所提出的特征提取方法对于数据量巨大时并不适应。鉴于此,国内的谈磊等人[15]提出了一种改进的复合分类模型。Elmendili等人[16]使用基于内容和行为的技术检测恶意用户。微博由于文本较短,字数较少,从中可以提取的特征有限,单一地根据用户发布消息的内容来识别恶意用户具有局限性,不能很好地识别恶意用户。为了弥补这方面的不足,文献[17]通过在MySpace上设置蜜罐账户,从而获得恶意用户关注的方式,对这些恶意用户样本进行分析,提取特征,开发了一个基于机器学习的分类器。由于大多数的恶意检测技术是针对国外的Twitter、Facebook等社交网站进行的研究,不一定适用于国内的微博。因此,林成峰[18]在此基础上,将3种分类算法进行整合,检测新浪微博上的恶意用户。许翰林[19]根据动态信任模型检测社交网络中的恶意用户。刘佳[20]通过使用BP和RBF两种算法对网络中的文本进行分类来检测异常用户。Hong等人[21]提出了基于卷积神经网络的恶意用户检测方法。

3.2 僵尸用户

僵尸用户,最为熟知的就是僵尸粉,它们并不参与正常的社交活动,也不具备攻击性,而是追随、热捧其他用户、话题或者发布少量不良链接。僵尸粉由最早的“三无”,即无头像、无粉丝、无微博逐渐发展为“三有”,即有头像、有粉丝、有微博,其关注的用户有真实用户也有僵尸用户。随着微博用户基数的增加以及微博管理员管理力度的加大,这种僵尸粉已经不能满足互联网市场,随之而来的是人工添加的虚假粉丝,即“活粉”,资料、头像等信息完整,与真实的微博用户没有什么差别,每天定时更新微博,活跃度很高,粉丝数量不再是瞬间暴增而是稳定地增加,且多为真实用户。

由于僵尸粉通常是由机器自动产生的,其生成的注册名和人填写的注册名存在一定的不同。考虑到这一点,方明等人[22]提出了一种改进的分类方法进行僵尸用户的检测。针对新浪微博中伪装策略不断升级的虚假粉丝,文献[23]在用户的社交关系和关系属性两个特征的基础上,从图特征的角度,即双向关注的百分比和追随者数量与被追随者数量的比率,用来检测虚假粉丝。岳红等人[24]将静态与动态特征相结合对僵尸粉进行检测。王一博[25]使用支持向量机,根据用户的关系属性和内容属性实现异常用户的检测。Zhang等人[26]对爬取到的新浪数据进行分析,分析得到僵尸用户和正常用户不同的特征,手动将合法用户中的僵尸用户进行标注,再利用SVM机器学习分类器自动检测僵尸追随者。为了提高检测算法的准确性,文中增加了一些检测新特征,并将用户的配置文件和推文相结合,实验表明该方法比以前的检测方法更有效,准确率达到99.78%,误检率为11.57%。Li等人[27]使用图的方法根据用户的行为属性检测新浪微博上的僵尸账户。陶永才等人[28]在用户关系构成的社交网络上,对用户的粉丝进行聚类,然后与用户的社交网络关系相结合,建立僵尸用户的检测模型,结果表明,召回率和精确率良好,查全率相对偏低。王越等人[29]使用C4.5决策树对新浪微博中的僵尸粉进行检测。

3.3 垃圾用户

垃圾用户频繁发布大量重复的信息,对其他用户造成了骚扰。国内外有许多学者对垃圾用户的检测进行了研究。Wang等人[30]使用40多种分类算法对社交网站单个的垃圾信息进行检测,但是无法满足大量垃圾用户的检测,针对这个问题,Zhang等人[31]在信息论原理的基础上对垃圾用户进行分类,提出了一种基于URL驱动的评估方法。McCord等人[32]根据用户与内容两方面的特征,构建了4个分类器(分别是随机森林、SVM、朴素贝叶斯和KNN)来区分垃圾用户与正常用户,实验表明,随机森林的效果最好,精确度为95.7%。李赫元等人[33]使用支持向量机对中文微博垃圾用户进行检测。吴斌等人[34]考虑了微博垃圾用户的文本、行为以及社交网络信息并建模分析,识别垃圾用户。邹永潘等人[35]使用混合的方法从统计和语义两方面进行检测。赵斌等人[36]在对微博垃圾用户的检测中,提出了一种基于重用检测模型的过滤算法,在考虑文本特征的同时,将用户

的行为特征也考虑在内,从文本粒度的角度,分为语句级检测SRD和词项级检测TRD,SRD重于检测用户行为,TRD重于检测垃圾信息的主题。文献[37]在以往研究的基础上做了进一步的完善,将内容、用户行为与图形相结合,通过分析用户的行为及其与用户的关系,提出了一种基于图的检测方法。在两种情况下使用多种分类算法进行了实验,第一种是使用整个数据来构建和评估模型,误报率较高;第二种将普通用户与垃圾用户的数量比设为2∶1,与多种分类算法相比,逻辑回归算法准确率最高,达到了99.569%。

3.4 虚假用户

针对现有对虚假用户的研究,Cao等人[38]使用基于图的方法,开发了一种新工具SybilRank,结合社交属性,使在线社交网络根据用户感知到的虚假可能性对虚假账户进行排名,其中排名较低的多为虚假账户,通过对200 000个账户进行测试,准确率为90%。文献[39]将微博用户考虑在内,将微博自身的文本特征和发布者的社交网络特征相结合,形成一个特征向量,然后作为支持向量机的输入,从而进行分类,检测结果比单一的特征要好,准确率分别提高了13%和29%。文献[40]用混合的方法对社交网络中的虚假用户进行检测,准确率为98%。现有很多研究在采用机器学习方法检测虚假用户时需要人工标注大量的数据集进行训练,代价较大,而且存在样本不足的问题,鉴于这一点,方勇等人[41]使用层次聚类的方法对海量数据中的虚假用户进行检测,可以有效找出一定比例的虚假用户。Zhang等人[42]考虑到账户存在的目的,使用星型抽样的方法来发现所有高级账户的追随者列表中重复或相似的微博账户,实现虚假账户的检测。针对Twitter的研究较多,而Facebook相对较少,Gupta等人[43]使用基于规则的方法,收集了用户账户上的所有活动,根据用户个人资料及与其他用户的互动行为对Facebook上的虚假账户进行检测。Ersahin等人[44]利用信息论知识,提出了一种在Twitter社交网络上检测虚假账户的分类方法,准确率为90.41%。谭侃等人[45]使用基于*K*-means的主动学习方法对社交网络中的虚假用户进行检测。

3.5 网络水军

传统对网络水军的研究多是基于内容特征、黑名单、规则进行分析的,随着互联网的迅速发展,用户的防范意识不断增强,传统检测方法已经无法发现这些隐蔽的水军。现有对网络水军的研究中,Moh等人[46]考虑到用户具有的社交关系,如关注的好友、粉丝等,由特征矩阵计算该用户的可信任度,对该用户是否为网络水军进行识别。Wang等人[47]提出将社交网络水军间的关系作为特征,利用图模型对网络水军之间的关系进行建模。这种基于关系交互的方法在水军识别的准确率上有了一定的提高,但数据获取较困难。Husna等人[48]使用主成分分析以及*K*-means聚类算法对水军机器人进行识别。Bhat等人[49]提出了一种基于规则的水军检测方法,分析了许多邮件水军的行为,并根据其中两种行为特征构建了一个分类器。陈侃等人[50]通过构建决策树实现微博水军的检测。张良等人[51]使用逻辑回归算法对新浪微博中的网络水军进行检测,检测特征包括好友数、粉丝数、发文频率、所发博文数、离线时间、博文含URL率等。上述基于行为特征的识别方法,对隐藏很深的网络水军和复杂多变的特征属性,识别准确率较低,张艳梅等人[52]进行了改进,根据贝叶斯理论和遗传算法进行水军的识别。Subrahmanian等人[53]根据句法和语义特征、时间的行为特性、用户资料等特征,先进行最原始的检测,然后根据LDA算法进行聚类分析,再用SVM识别Twitter和Facebook上的机器人水军。Chen等人[54]使用改进的分类算法,将循环神经网络与自动编码器相融合,检测网络水军。

3.6 异常情绪用户

微博、Twitter等已经成为人们在线交流和传播情感的主要社交平台,用户可以在上面自由地发表个人的意见和观点,发泄自己的情感,表达他们对生活、服务、政策、产品、当前事件、热点话题和其他主题的看法。异常情绪用户的识别通常在语义特征的基础上对单用户或群体用户的情感进行研究,根据用户发布的内容,了解他们的情绪状态,从而可以分析网络舆情,及时发现异常情绪用户,并对他们进行开导,从而预防一些极端事件和群体事件的发生。Jain[55]设计了一个基于神经网络的跟踪单个用户情感的移动应用程序,将用户输入的文本、浏览器的历史记录以及社交媒体相结合进行情感分析,判断一个人是否表现出长期的负面情绪。Wang等人[56]提出了一种改进的情感分类方法,即将文本对象考虑在内,提高了分类的准确率,对Twitter数据进行异常检测,通过比较负面情绪所占的比例来观察一天的异常情况。刘翠娟等人[57]使用基于规则的方法,从多情感角度对社会群体情感的变化进行了分析,首先计算群体情感强度,然后通过人工方式标注情感词的强度,最后与句法分析相结合,计算出微博文本的情感类型和强度,采用可视化的方法对群体情感进行展示。熊建英等人[58]将基于规则的技术与图的方法相融合,在把句法分析和情感词典相结合的基础上,加入了社交网络互动关系,即计算与用户互动频繁的节点之间的信任值,从而通过可信反馈,识别出情绪异常节点。Sun等人[59]爬取了5年来100个用户的10 275个微博,按用户和月份分为5类:中立、快乐、惊讶、悲伤、愤怒,将这5类作为与用户情绪相关的变量进行建模,5种情感用五元高斯模型来模拟,检测出异常情绪状态。实验表明,单个用户的中性、快乐和悲伤情绪服从正态分布,群体的微博情绪服从幂律分布,个人用户异常检测准确率为83.49%,不

同月份为87.84%。文献[60]在此基础上提出了一个将统计与神经网络方法融合的改进模型用于检测微博异常情绪用户。

3.7 不良言论用户

由于微博、论坛等具有发布方式便捷多样，传播速度快等特点，对很多敏感和突发事件的传播更敏捷，成为很多舆情传播的载体。诸多社交平台的发布门槛极低，一些用户就一些政治问题，发表诸如暴力、反动、煽动民族团结、鼓吹民族分裂等敏感言论。比如2014年8月新疆网民发布暴恐言论，称武警轰炸莎车3个村，将妇女孩子全部击毙，达到煽动民族仇恨的目的，给人们造成了极大的恐慌；2017年1月乌鲁木齐市张某在“百度贴吧-乌鲁木齐吧”恶意攻击自治区维稳措施，给政府带来了负面的影响。虽然相关部门对其进行了管控和监测，但仍有一些用户会绕过这些敏感关键词，将它们以其他的方式表达出来。比如2018年4月发生的“洁洁良事件”，在微博使用辱华词语“支那”，对社会产生了十分恶劣的影响。

对不良言论用户的识别多是从用户发布的话题进行检测的，而且这些话题多是一些敏感话题。敏感话题是话题的一种，只是这些话题中含有一些敏感关键词，因而对话题的研究既涉及敏感话题的研究，又包括热点话题的研究。目前针对文本进行话题发现和跟踪的研究还不是很成熟，对社交网络敏感话题的检测与跟踪，一般是根据国家的需要制定，虽然有一些相关的产品应用，但是这方面的研究相对薄弱，参考文献也比较少。

翟东海等人[61]提出了基于CRF的敏感话题检测模型，对敏感话题具有的敏感性特征进行拟合和推断，与贝叶斯方法相比，具有较好的识别性能。王亮等人[62]考虑到图像、视频中含有的潜在信息，提出了基于SP&KM的改进的聚类算法检测敏感话题。潘大庆[63]使用层次聚类检测敏感话题。孙胜平[64]使用改进Single-Pass和层次聚类算法进行微博话题的发现。丁苳[65]通过Single-Pass算法进行修改，过滤掉与话题不相关的微博文本，最后使用朴素贝叶斯算法进行分类来实现对话题的跟踪。Ishikawa等人[66]针对同一个话题可能会使用不同的词汇来进行描述，提出一种与图算法混合的话题检测方法，该方法首先抽取一段时间和地域内的Twitter文本，然后抽取出文本中具有代表性的词，根据维基百科找到这些词的相关语义信息，并选择出表达意思相似的特征词作为主题词汇，最后根据这些新定义的词汇进行话题的检测。张越今等人[67]提出了一种基于相似哈希的增量型文本聚类算法。冯雪坪[68]根据微博具有时间周期的特点，在一定的时间窗口内先对微博进行分组，然后在每个分组中使用高频词对微博进行排序处理，最后使用改进的Sing-Pass算法对这些微博文本聚类。李勇等人[69]使用混合的方法检测话题，先使用LDA主题模型对微博文本进行建模，在检测出话题簇的基础上使用支持向量机发现话题。Ding等人[70]针对Twitter实时更新要求高以及内容少等特点，提出了一种混合的半监督狄利克雷处理过程来进行Twitter中的话题检测与跟踪。

4 异常识别技术

4.1 基于分类的技术

分类是从一组标记的数据实例中学习模型，然后用训练的分类器将测试实例分类到其中一个类中。分类属于监督学习的范畴，基于分类的异常检测技术分为两个阶段：首先用标记的训练数据来训练分类器，即训练阶段；其次用分类器将待测样本分为正常或异常两个类别，即测试阶段。以下是异常识别中常见的机器学习技术。

4.1.1 贝叶斯分类

贝叶斯分类器是基于贝叶斯理论的分类器，它首先计算特征值属于每个类别的先验概率，再根据人工标记文本类别的概率，计算其后验概率，把概率最大的那个类当作最终的分类结果。

在对恶意用户的识别中，文献[13]在传统检测方法的基础上，通过构造信息熵，结合机器学习技术，将用户分为三大类，提取用户的内容、属性和行为特征，构建了一个贝叶斯分类系统，在50万用户发布的4 000多万条推文上进行实验，结果表明检测准确率有了一定的提高。但是该方法计算量大，而且基于属性间的独立性假设，对真实社交网络进行检测时，分类准确率会受到影响。文献[30]提出了一个适用于所有社交网站的单个实例的垃圾信息检测框架，检测特征包括内容特征和行为特征，该方法可以实现单个垃圾用户的检测。文献[52]利用贝叶斯模型识别微博中的水军，选取了粉丝关注比、平均发布微博数、互相关注数、综合质量评价、收藏数等行为特征，与逻辑回归算法相比，该方法可以在学习一定量的样本后，对微博用户进行分类，即水军和非水军，准确率高达97%。

4.1.2 神经网络

神经网络是由多个神经元按照一定的层次结构连接起来的，也就是一组相互连接的节点，每个节点与相邻层中的其他几个节点有一个加权连接，单个节点将从连接节点接收到的输入数据与一个简单函数一起使用，来计算输出值。

在对僵尸用户的识别中，针对BP神经网络存在处理大量数据时收敛速度较慢、容易陷入局部极小等不足，文献[29]提出了一种将模拟退火算法与BP神经网络相结合的方法，从个人信息特征如关注数、粉丝数、人气指数等，和微博内容特征如微博数、微博转发率等，对新浪微博的5 000个用户进行检测，结果表明，该判别模型的准确率和召回率均为93%。文献[24]从静态与动态两

方面,将磷虾群优化算法和人工免疫算法的变异操作引入到网络连接权值和阈值的优化过程中,对神经网络进行训练,使用URL比值、用户关注度,将微博发布时间的随机性、用户被关注度等作为僵尸粉的检测特征。文献[21]提出了利用CNN提取用户发布内容的文本特征和图像特征,再用分类器进行分类,实现恶意用户的检测,实验平均精确率为79.93%。

4.1.3 支持向量机

支持向量机算法是建立在统计学习理论和结构风险最小化原则上的近似实现,其基本原理是找到一个超平面,实现正面和负面两类样本之间的分离边缘最大化。SVM算法既可以进行有监督的训练,也可以调整为无监督的学习。

在对僵尸用户的检测中,文献[71]将用户的行为特征关注数、微博数、转发数、微博中的被转发的微博数、微博中被评论的微博数等作为检测特征,准确率高于91%。文献[33]通过对垃圾用户行为的分析,从用户与微博内容这两方面,提出了适用于中文微博特点的7个用于检测垃圾用户的新特征,分别是用户权威度、用户关注度、纯粉丝度、用户头像特征、近期活跃度、用字多样性等,使用SVM对垃圾用户进行检测,分类准确率为94%,但该方法需要对数据进行人工标注,花费代价较大。文献[14]通过对新浪微博3万多名用户和1 600多万条消息的分析,提取出与消息内容和用户行为有关的18个特征,如点赞数、评论数、转发数、URL的平均数量等,使用SVM对恶意用户进行检测,实验表明,该方案分类准确度达99%,性能出色。文献[25]考虑到用户的关系属性和内容属性特征,选取5个特征值,分别是关注数、粉丝数、微博数、是否有评论和点赞,用SVM进行训练,结果表明,僵尸用户的测试精确度达到了96%。但是由于数据集是人工进行标注的,具有主观性,会对结果造成影响,不仅如此,对于一些伪装的高级用户,还应该考虑更多的特征。

4.1.4 基于规则的技术

基于规则的异常识别技术是学习并捕获系统正常行为规则的过程,任何不包含在系统中的东西都被认为是异常的。常见的算法有决策树、随机森林等。

在对恶意用户的检测中,文献[12]在Facebook、MySpace和Twitter上创建一组蜜罐账户,采集社交网络用户的信息,通过对数据进行分析,挖掘出其六大行为特征,分别是跟随者与关注者的比值、发送消息中的URL比、消息相似度即用户发送的消息之间的相似性、朋友选择、发送的消息数量、朋友数量,采用随机森林算法对用户分类,其准确性最好且误判率最低。在对虚假用户的识别中,文献[43]采用两种分类算法,即决策树与贝叶斯网络,检测Twitter上发布虚假消息的用户,提取了粉丝数、好友数、账户年龄、推文长度等用户和内容特征,结果表明,决策树的分类准确率更高,对文本特征的贡献度也更大。文献[72]使用C4.5决策树分类算法对垃圾用户进行检测,检测特征包括用户的关注粉丝比、链接比、互粉数、平均评论数等,检测准确率为92%。文献[29]使用C4.5决策树对新浪微博中的僵尸粉进行检测,准确率和召回率均为92.8%,检测特征包括用户个人信息、用户微博内容和用户链接关系等。

4.1.5 基于多种分类器的组合

单一的分类算法不能产生很好的结果,为了提高分类的准确率,可以将多种算法进行组合,弥补单一算法的不足,对异常用户进行检测。

由于朴素贝叶斯遵循类条件独立原则,在现实问题中不一定成立,因而对分类结果也会产生影响,而KNN算法准确率较好,但计算量大。为了弥补这两种算法的不足,文献[15]对恶意用户的识别进行了优化,提取用户属性和行为的特征,如注册时间、发文方式、好友数、粉丝数等,根据用户属性的相关性选择合适的分类算法,在确保分类准确率的同时,减少了时间开销。文献[18]在这些研究者的基础上,先收集大量的恶意用户数据,人工进行标注,对数据进行分析,将恶意行为分为3类,提取粉丝数、好友数、关注数等特征,对这3种恶意行为分别进行训练,再把这些分类器整合到一起,实现恶意行为和用户的自动检测。文献[22]从用户个人属性特征的角度,以注册的用户名为特征,采用SVM、ANN对僵尸用户进行分类,实现对僵尸用户的甄别,准确率均高于92%,但是该方法没有考虑到动态检测特征,因而对高级僵尸用户不能很好地识别。在对水军的识别中,Chen等人[54]考虑到微博评论中含有潜在的有价值信息,提出了新的检测特征,使用循环神经网络对评论中的特征进行时间序列的分析,把RNN模块的输出和微博本身的特征结合,作为自动编码器(AE)的输入;AE通过学习正常的行为模式产生输出,根据AE输入与输出的差值进行判断,实验准确率为92.49%,F1值为89.16%。

4.2 基于聚类的技术

聚类被定义为将大量数据划分为以相似度为基础的集合,每个类或簇由彼此相似且与其他群组不同的对象组成,属于无监督学习的范畴,不需要预先标记数据。根据聚类的不同标准,聚类过程会基于对象的相似性度量产生数据集的不同簇类,不在任何簇中的对象认为是异常的。聚类方法有多种,下面对一些常见的基于聚类的[73]异常识别技术进行介绍。

4.2.1 划分聚类

划分法要提前设定类簇的个数,然后通过反复地迭代将数据集划分为多个互斥的类簇,且同一类簇中对象相似度比较大,而不同类簇中的对象相似度比较小。

文献[48]通过对垃圾邮件发送者的行为特征进行分析,提取出邮件的发送频率、内容类型、到达时间、内

容长度等特征，使用*K*-means聚类算法进行分析，计算其相似性大小，实验结果精确度为90%。文献[45]使用排序与*K*-means聚类相结合的双层采样算法，提取用户的属性特征如关注数、粉丝数、账户年龄等，内容特征如URL数、评论数等，行为特征如发帖数目、文章转发数等，邻居特征如该用户邻居用户的粉丝数、关注数等，以及关系图特征如双向关注比、PageRank值等，对社交网络中的虚假用户进行检测。文献[62]提出了改进的识别方法，他们认为只分析用户发布的文字信息，不能得到精确的检测结果，因而在此基础上，对用户发布的图像、视频进行预处理，提取其中的信息，针对Single-Pass算法全局性能不好，*K*-means初始点选取困难等问题，提出了改进Single-Pass和*K*-means组合的敏感话题检测系统。先用Single-Pass算法进行聚类，形成一个个小簇，再利用*K*-means算法，挑选这些微型簇进行二次聚类，从而实现不良言论用户的检测。文献[74]为了提高*K*-means在聚类中的效果，根据最大距离选择初始聚类中心，并引入信息熵原理，计算各个属性的权重，使用赋权的欧式距离计算相似度。结果表明，该方法提高了异常检测率，降低了误报率。

4.2.2 层次聚类

层次聚类是通过对数据对象进行一层一层的聚类，最后形成一个聚类树的过程。根据聚类树的生成方式可以分为两类：凝聚和分裂的算法。凝聚式聚类方法是自底而上的，先将每个对象视为一个单一的聚类，再将其与最近的数据点迭代合并。分裂层次聚类方法与其相反，但计算上复杂度较高。

文献[63]提出了一种基于层次聚类的敏感话题检测算法，不仅可以实现对微博舆论的监测，还可以检测发布敏感话题的不良言论用户，结果表明，算法的检测精度为95.3%，话题误判率低于6%。文献[41]根据恶意注册的虚假账户具有很大相似性的特点，提出了使用层次聚类的方法，识别虚假用户。核心思想是先根据用户的字符串模式对海量数据进行分类，然后计算每个类别中元素的字符串相似度的大小，进行层次聚类，从而发现海量数据中的虚假用户。该方法不需要人工收集大量的数据，但是时间复杂度较高。

4.2.3 密度聚类

传统大部分的聚类算法是通过文本距离进行聚类的，而基于密度的聚类算法通过使用密度对不同的簇进行划分。该聚类算法可以克服基于距离的聚类算法只能发现球状簇的缺点，能够发现任意形状的簇。

Yoshida等人[75]提出了一种基于密度的垃圾用户检测方法，该方法需要从大量的电子邮件中提取信息，使用文档空间密度作为基本特征，实验表明，该方法处理速度快，准确率和召回率也非常高。Zabihi等人[76]使用基于密度的聚类算法，考虑到DBSCAN算法对维数较敏感，选择了4个特征识别异常用户，实验表明，该方法可以有效地区分正常用户和恶意用户，并且集群质量和准确性较高。Nguyen等人[77]提出了一种将分类与基于密度的聚类相结合的方法来检测网络中的僵尸用户。

4.2.4 增量式聚类

Single-Pass算法是TDT评测中应用最多的增量式聚类算法。核心思想是把到来的第一篇文档作为一个话题，分别计算后续输入的文本和已有话题的相似度。若相似度值大于给定阈值，则把该文本分到这个话题类别中；若小于给定阈值，表明该文本不属于已有的话题类别，需要新建一个话题类别。

在对不良言论用户的识别中，文献[64]针对微博内容更新快、文本字数受限等特点，提出了使用改进的Single-Pass算法和层次聚类算法进行微博话题的发现，该方法能够有效进行微博的话题发现。文献[67]也对Single-Pass算法进行了改进，提出了一种基于相似哈希的增量型文本聚类算法，利用相似哈希算法对文本向量进行降维，通过该算法得到的文本 Simhash指纹用海明距离来计算文本与文本簇之间的相似程度，在此基础上用改进的Single-Pass对新增的文本实时聚类。实验表明，该算法比于原Single-Pass在聚类效率上有了明显的提升，保证了时效性，有较高的实用价值。

4.3 基于统计的技术

统计的方法先假设数据样本符合某种分布，然后结合数理统计和概率论方法，根据样本中的正常和异常数据建立概率模型，判断给定的实例是否属于该模型。判断准则依赖于统计异常检测技术的如下假设：正常的数据实例通常分布在概率较高的区域，而异常数据则分布在概率较低的区域。用来拟合统计模型的方法有参数化方法和非参数化方法两种。参数化方法假设样本取自服从已知模型的某个分布，从给定的样本估计分布的参数，把这些估计放到假设的模型中，并得到估计的分布，然后使用它进行决策[78]。常见的参数化技术有高斯模型、卡方检验等。非参数化方法对数据的假设较少，不需要假定数据分布的先验参数，模型结构由给定的数据决定。常用的方法有直方图估计、核估计、KNN。

文献[32]从每个Twitter用户的帐户中提取出基于用户的特征，有好友数量、粉丝数量、用户信誉、在24小时内平均发布百分比，以及基于内容的特征，如发布的话题标签、推文中“@”其他人的数目、推文单词数、推文中URL数，构建了4个分类器来区分垃圾用户与正常用户，其中KNN的准确率和召回率均为92.8%。文献[79]通过卡方检验从话题标签、@其他用户的情况、URL、词语中找出了最具有代表性的10个特征，从内容、网络结构等方面利用Adaboost检测Twitter中的水军。Torabi等人[80]提出一种基于隐马尔可夫模型的检测方法，将内容特征和URL链接作为检测特征，从而实现垃圾发

送者的识别。文献[60]在此基础上,提出了一种基CNN-LSTM的混合模型,然后将其与多元高斯分布相结合,检测微博上的异常情绪用户,提出的混合模型可以获得更多有关上下文结构的信息,提高了泛化能力。

4.4 基于混合的技术

由于监督学习和非监督学习方法都存在各自的不足,单独使用任何特定的方法都不会得到精确的结果。为了克服这两种方法的缺点,提出了一种将两者相结合的混合方法。研究表明混合方法可以提高异常检测的准确率,能够获得更好的性能。

Chitrakar等人[81]将SVM算法和*K*-medoids聚类算法相结合进行异常检测,首先用*K*-medoids算法将数据进行分组,然后将得到的每一个簇使用SVM进行分类。实验结果表明,该方法提高了异常检测准确率,但在数据集非常大时,时间复杂度较高。大多数对虚假账户的检测是针对一个账户进行的,但在实际的社交网络中,每天都可能注册数十万个新账户,因此文献[40]首先用聚类算法将账户分组形成簇群,然后对这些分组形成的账户集群使用SVM、随机森林等算法分类为正常或虚假的用户,从而判断它们是否是由同一用户创建的。该方法中使用的主要特征是针对整个群集的,如名称、电子邮件地址、公司或大学等。文献[35]从统计特征和语义特征两方面对垃圾微博进行检测,先提取微博的显式特征(用户特征、内容特征);再利用LDA模型对隐含主题特征进行提取,构建特征向量,再用SVM进行分类。

4.5 基于信息论的技术

根据信息论的相关知识,如熵、条件熵、相对熵、相对条件熵、信息增益和信息成本等,可以描述数据集的相关特征并建立合适的异常检测模型。通常情况下,基于信息论的异常识别技术需要对训练数据集的特征进行研究,然后根据这个特征建立模型并用一个测试数据来评估这个模型的性能,从而利用信息理论测度来确定模型是否适合测试新数据集。

文献[50]根据社交网络中的用户与其他用户建立的交互关系(回复、评论、转发、点赞等)定义了3种6个特征,对传播行为进行量化,通过信息论原理计算样本节点的信息熵和各属性的信息增益比率,完成分类器的构建,该方法可用于多场景下的微博水军检测。在对Twitter上的垃圾用户检测的研究中,文献[31]提出了一种基于URL驱动的评估方法,利用香农信息论计算发布内容含URL的两个账户之间的相似性,将相似性较高的帐户关联起来,提取平均发布时间间隔、帐户发布的URL的平均数量等多个特征来进行分类,通过检测Twitter上的垃圾用户,验证了该方法的有效性。文献[44]研究了熵最小化离散化(EMD)的离散技术对虚假账户检测准确率的影响,实验通过EM对数据集进行处理,提取推文数量、朋友数量等行为特征,然后使用朴素贝叶斯进行分类,结果表明,对Twitter上的虚假账户的检测准确率从85.55%增加到90.41%,为其他社交媒体平台如Facebook、Instagram等提高虚假账户的检测准确率提供了新思路。

4.6 基于图的技术

社交网络可以看成是一个图形结构,用户以节点的形式表示,用户之间的交互行为用边表示。社交网络中的异常通常是基于行为、结构以及光谱进行的检测。

文献[19]从动态信任模型的角度出发,设计了一种基于信任计算和交互信息的社交网络信任模型,对用户节点进行评价,使社交网络的信任度计算更加有针对性,更适应于社交网站快速的发展变化,将提出的信任模型作为判断机制用于恶意用户检测。文献[16]使用一种基于协同过滤的安全策略来检测和识别行为突然发生变化的帐户,开发了一种工具来过滤数据,并根据用户的配置文件访问安全数据,确定社交网络中伪造身份的恶意用户,这种方法速度更快而且也更有效。文献[27]考虑到用户的行为属性,根据账户之间的社交关系构建社会关系矩阵,每个账户的PageRank值通过PageRank迭代计算得到,然后根据这个值对账户进行排名,从而检测出僵尸用户,两种测试情况下的准确率分别为84%和92%。该方法易于实现且只考虑账户之间的关系,没有考虑特征维度,因而排序结果的准确率会受到时间延迟的影响。Ying等人[82]根据恶意用户与正常用户处于不同的频谱空间区域,通过计算光谱或光谱坐标来识别正常和异常用户的特征值或特征向量,根据随机攻击链接来识别恶意节点,实现基于光谱技术的异常用户检测。

4.7 几种方法的比较

不同的方法具有不同的特点,表3给出了几种异常识别方法的优缺点[83]。

5 结束语

本文对社交网络的异常用户进行了分类,并对每种异常用户的发展现状进行了概述,同时对主要的异常识别技术进行了分类概括,并比较了它们的优缺点。最后,提出了异常用户检测中存在的问题:对一些高级伪装的异常用户进行异常检测相对比较困难;目前对异常用户的识别多是从一个或几个少量特征进行的检测,如何将内容、行为、属性、邻居等多个特征融合在一起,提高异常用户检测率是一个有价值的问题,与此同时,虽然有一些学者考虑到了时间特征,但没有受到广泛的关注,尤其是在动态社交网络中;对异常的检测还没有一个统一的标准,因为对不同的应用领域,异常的定义也是不同的,所以在一个领域中的异常检测技术不一定适用于其他领域,不仅如此,正常的行为模式是会随着时

表3 几种方法的比较

异常识别技术	优点	缺点
基于分类的技术	可以集成多个分类器对多类问题进行区分,效率比较高;测试阶段速度快	对训练数据以及各类数据标签的准确性较为依赖;对类别不平衡问题缺乏有效的解决方案
基于聚类的技术	不需要对数据进行标记,并且适用于其他复杂的数据类型;因为簇集数远小于数据对象的数量,所以当每个测试实例与簇集进行比较的时候,速度比较快	计算复杂度较高,当数据集较大时,花费代价较高;其性能主要依赖于聚类算法捕获正常实例的聚类结构的有效性;不属于任何簇的数据对象不一定就是异常的;只有当异常不在它们之间形成显著的聚类时,该方法才有效;由于聚类算法是强行将每个数据对象分配给某个簇,会导致异常分配到一个大簇群,从而将上述簇群中的对象视为正常
基于统计的技术	在假设数据样本符合某种分布成立的前提下,该方法为异常检测提供了一个新思路;统计方法得到的异常值与置信区间相关联,置信区间可作为附加信息并对所有的测试实例作出决策;若分布估计过程对数据中的异常具有鲁棒性,可以在无监督环境下运行	该方法依赖于数据样本符合特定分布的某种假设,但假设并不总是成立,尤其对高维数据集;即使假设合理并且有不止一个假设可以用于异常检测,但从中选出一个最佳的统计数据并不容易;直方图方法虽操作简单,但不易发现不同属性之间的关系,在一些特定的异常检测中无法实现
基于混合的技术	可以解决数据集类别不平衡问题;异常检测准确率高;误报率低	时间复杂度会增加,尤其是数据集巨大的时候
基于信息论的技术	可以在无监督环境下运行;不用对数据的基本统计分布作出任何假设	其性能依赖于信息论测度的选择,通常只有数据中存在大量异常时,该方法才能够检测到异常情况;应用于序列和空间数据集的信息论技术依赖于子结构的大小
基于图的技术	可以实现降维,适用于处理高维数据集,减少了运算量;可以在无监督的环境下运行	准确率低,计算复杂度高;对理论假设较为依赖

间的推移而变化的,当前的异常检测技术在未来不一定适用。今后的工作将主要集中在从大数据中发现数据新的特征以及正常和异常的行为模式;此外,将独立的异常检测技术进行连接并应用于复杂的多系统检测也是未来值得探索的问题。

参考文献:

[1] Markou M,Singh S.Novelty detection:a review part2:neural network based approaches[J].Signal Process,2003,83(12):2499-2521.

[2] Patcha A,Park J-M.An overview of anomaly detection techniques:existing solutions and latest technological trends[J].Computer Networks,2007,51(12):3448-3470.

[3] Hodge V,Austin J.A survey of outlier detection methoddologies[J].Artif Intell Rev,2004,22(2):85-126.

[4] Fu S,Liu J,Pannu H.A hybrid anomaly detection framework in cloud computing using one-class and two-class support vector machines[C]//International Conference on Advanced Data Mining and Applications.Berlin,Heidelberg:Springer,2012:726-738.

[5] Ahmed M,Mahmood A N,Islam M R.A survey of anomaly detection techniques in financial domain[J].Future Generation Computer Systems,2016,55(6):278-288.

[6] Fire M,Katz G,Elovici Y.Strangers intrusion detection detecting spammers and fake profiles in social networks based on topology anomalies[J].Human Journal,2012,1(1):26-39.

[7] Kaur R,Singh S.A survey of data mining and social network analysis based anomaly detection techniques[J].Egyptian Informatics Journal,2016,17:199-216.

[8] Hawkins D.Identification of outliers(monographs on statistics and applied probability)[M].Netherlands:Springer,1980.

[9] Doostari M A,Zeinali R,Lashkari H,et al.Anomaly detection in cliques of online social networks using fuzzy node-fuzzy graph[J].Journal of Basic and Applied Scientific Research,2013,3(8):614-626.

[10] Anand K,Kumar J,Anand K.Anomaly detection in online social network:a survey[C]//International Conference on Inventive Communication and Computational Technologies,2017.

[11] 莫倩,杨珂.网络水军识别研究[J].软件学报,2014(7):1505-1526.

[12] Stringhini G,Kruegel C,Vigna G.Detecting spammerson social networks[C]//Computer Security Applications Conference,2010:1-9.

[13] Chu Z,Gianvecchio S,Wang H,et al.Who is tweeting on Twitter:human,bot,or cyborg?[C]//Computer Security Applications Conference,2010:21-30.

[14] Zheng X,Zeng Z,Chen Z,et al.Detecting spammeon social networks[J].Neurocomputing,2015,159(C):27-34.

[15] 谈磊,连一峰,陈恺.基于复合分类模型的社交网络恶意用户识别方法[J].计算机应用与软件,2012,29(12):1-5.

[16] Elmendili F,Idrissi Y E B E,Chaoui H.Detecting malicious users in social network via collaborative filtering[C]//International Conference on Big Data,Cloud and Applications,2017:44.

[17] Webb S,Caverlee J,Pu C.Social honeypots:making-

friends with a spammer near you[C]//The Fifth Conference on Email and Anti-Spam, Mountain View, 21-22 August, 2008.
[18] 林成峰.新浪微博恶意用户研究及检测[D].上海:上海交通大学,2014.
[19] 许翰林.基于信任计算的社交网络恶意用户检测[D].南京:南京邮电大学,2016.
[20] 刘佳.基于人工神经网络的社交网站文章热度分类研究[J].通化师范学院学报,2015,36(12):56-59.
[21] Hong T, Chang C, Shin J.CNN-based malicious user detection in social networks[J].Concurrency & Computation Practice & Experience, 2017.
[22] 方明,方意.一种新型智能僵尸粉甄别方法[J].计算机工程,2013,39(4):190-193.
[23] Shen Y, Yu J, Dong K, et al.Automatic fake followers detection in Chinese micro-blogging system[C]//Pacific-Asia Conference on Knowledge Discovery and Data Mining.Springer International Publishing, 2014:596-607.
[24] 岳虹,张智,杨科,等.基于磷虾群免疫神经网络的微博僵尸粉检测[J].计算机应用与软件,2015,32(12):145-149.
[25] 王一博.基于Scrapy的社交网络异常用户检测系统研究与开发[J].信息与电脑,2016(14):97-98.
[26] Zhang Z, Zou F, Pan L, et al.Detection of zombie followers in SINA Weibo[C]//IEEE International Conference on Computer and Communications, 2017:2476-2480.
[27] Li S, Li X, Yang H, et al.A zombie account detection method in microblog based on the PageRank[C]//IEEE International Conference on Software Quality, Reliability and Security Companion, 2017:267-270.
[28] 陶永才,王晓慧,石磊,等.基于用户粉丝聚类现象的微博僵尸用户检测[J].小型微型计算机系统,2015,36(5):1007-1011.
[29] 王越,张剑金,刘芳芳.一种多特征微博僵尸粉检测方法与实现[J].中国科技论文,2014(1):81-86.
[30] Wang D, Irani D, Pu C.A social-spam detection framework[C]//CEAS, 2011:46-54.
[31] Zhang X, Zhu S, Liang W.Detecting spam and promoting campaigns in the Twitter social network[C]//IEEE International Conference on Data Mining, 2013:1194-1199.
[32] McCord M, Chuah M.Spam detection on Twitter using traditional classifiers[C]//International Conference on Autonomic and Trusted Computing, Banff, Sep 2-4, 2011:175-186.
[33] 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.
[34] 吴斌,李冠辰,刘宇,等.基于微博重复发送的垃圾用户甄别[J].数据采集与处理,2015,30(1):117-125.
[35] 邹永潘,李伟,王儒敬.基于多特征的垃圾微博检测方法[J].计算机系统应用,2017,26(10):184-189.
[36] 赵斌,吉根林,曲维光,等.基于重用检测的微博垃圾用户过滤算法[J].南京大学学报:自然科学,2013,49(4):456-464.
[37] Sarpiri M N, Gandomani T J, Teymourzadeh M, et al.A hybrid method for spammer detection in socialnetworks by analyzing graph and user behavior[J].Journal of Computers, 2017, 13(7):823-829.
[38] Cao Q, Sirivianos M, Pregueiro T, et al.Aiding the detection of fake accounts in large scale social online services[C]//USENIX Conference on Networked Systems Design and Implementation, 2012:15.
[39] Wang K, Wang Y, Li H, et al.A new approach for detecting spam microblogs based on text and user's social network features[C]//International Conference on Wireless Communications, Vehicular Technology, Information Theory and Aerospace & Electronic Systems, 2014:1-5.
[40] Xiao C, Freeman D M, Hwa T.Detecting clusters of fake accounts in online social networks[C]//ACM Workshop on Artificial Intelligence and Security, 2015:91-101.
[41] 方勇,刘道胜,黄诚.基于层次聚类的虚假用户检测[J].清华大学学报:自然科学版,2017(6):620-624.
[42] Zhang Y, Lu J.Discover millions of fake followers in Weibo[J].Social Network Analysis & Mining, 2016, 6(1):1-15.
[43] Gupta A, Lamba H, Kumaraguru P, et al.Faking sandy: characterizing and identifying fake images on Twitter during Hurricane Sandy[C]//International Conference on World Wide Web, 2013:729-736.
[44] Ersahin B, Aktas O, Kilinc D, et al.Twitter fake account detection[C]//International Conference on Computer Science and Engineering, 2017:388-392.
[45] 谭侃,高旻,李文涛,等.基于双层采样主动学习的社交网络虚假用户检测方法[J].自动化学报,2017,43(3):448-461.
[46] Moh T S, Murmann A J.Can you judge a man by his friends? Enhancing spammer detection on the Twitter microblogging platform using friends and followers[C]//Prasad S K, Vin H M, Sahni S.Proc of the Int'l Conf on Information Systems and Technology Management. Heidelberg:Springer-Verlag, 2010:210.
[47] Wang G, Xie S, Liu B, et al.Identify online store review spammers via social review graph[J].ACM Trans on Intelligent Systems and Technology, 2012, 3(4):61.
[48] Husna H, Phithakkitnukoon S, Palla S, et al.Behavior analysis of spam botnets[C]//International Conference on Communication Systems Software and Middleware and Workshops, 2007:246-253.
[49] Bhat V H, Malkani V R, Shenoy P D, et al.Classification of email using BeaKS: behavior and keyword stemming[C]//2011 IEEE Region 10 Conference, 2011:1139-1143.
[50] 陈侃,陈亮,朱培栋,等.基于交互行为的在线社会网络水

军检测方法[J].通信学报,2015,36(7):120-128.
[51] 张良,朱湘,李爱平,等.一种基于逻辑回归算法的水军识别方法[J].信息安全与技术,2015(4):57-62.
[52] 张艳梅,黄莹莹,甘世杰,等.基于贝叶斯模型的微博网络水军识别算法研究[J].通信学报,2017,38(1):44-53.
[53] Subrahmanian V S,Menczer F,Azaria A,et al.The DARPA Twitter bot challenge[J].Computer,2016,49(6):38-46.
[54] Chen W,Yan Z,Chai K Y,et al.Unsupervised rumor detection based on users' behaviors using neural networks[J].Pattern Recognition Letters,2017,105.
[55] Jain V.Abnormality detection in human emotional behavior through sentiment analysis:2014A7PS113P.Pilani:Birla Institute of Technology & Science.
[56] Wang Z,Joo V,Tong C,et al.Anomaly detection through enhanced sentiment analysis on social media data[C]//IEEE International Conference on Cloud Computing Technology and Science,2015:917-922.
[57] 刘翠娟,刘箴,柴艳杰,等.基于微博文本数据分析的社会群体情感可视计算方法研究[J].北京大学学报:自然科学版,2016,52(1):178-186.
[58] 熊建英.基于可信反馈的微博用户情绪异常预警模型研究[J].情报科学,2017(4):48-53.
[59] Sun X,Zhang C,Li G,et al.Detecting users' anomalous emotion using social media for business intelligence[J].Journal of Computational Science,2017.
[60] Sun X,Zhang C,Ding S,et al.Detecting anomalous emotion through big data from social networks based on a deep learning method[J].Multimedia Tools & Applications,2018(5):1-22.
[61] 翟东海,崔静静,聂洪玉,等.基于条件随机场的敏感话题检测模型研究[J].计算机工程,2014,40(8):158-162.
[62] 王亮,方勇.敏感话题收集相关技术的研究[J].现代计算机,2016(12):3-6.
[63] 潘大庆.基于层次聚类的微博敏感话题检测算法研究[J].广西民族大学学报:自然科学版,2012,18(4):56-59.
[64] 孙胜平.中文微博客热点话题检测与跟踪技术研究[D].北京:北京交通大学,2011.
[65] 丁蓉.微博热点发现技术的研究与实现[D].武汉:华中科技大学,2012.
[66] Ishikawa S,Arakawa Y,Tagashira S,et al.Hot topic detection in local areas using Twitter and Wikipedia[C]//ARCS Workshops,2012:1-5.
[67] 张越今,丁丁.敏感话题发现中的增量型文本聚类模型[J].信息网络安全,2015(9):170-174.
[68] 冯雪坪.微博话题检测与跟踪方法研究[D].武汉:华中科技大学,2013.
[69] 李勇,张克亮.面向LDA和VAM模型的微博热点话题发现研究[J].计算机应用,2015(6).
[70] Ding W,Zhang Y,Chen C,et al.Semi-supervised Dirichlet-Hawkes process with applications of topic detection and tracking in Twitter[C]//IEEE International Conference on Big Data,2017.
[71] 张锡英,车鑫,田宪允.一种基于微博用户行为的僵尸粉识别方法[J].黑龙江大学自然科学学报,2014,31(2):250-254.
[72] 孟祥飞,徐路,王思雨.基于新浪微博的社交网络垃圾用户分析与检测[J].科技与创新,2014(15):125-127.
[73] Ahmed M,Mahmood A N,Islam M R.A survey of anomaly detection techniques in financial domain[J].Future Generation Computer Systems,2016,55(6):278-288.
[74] 陈庄,罗告成.一种改进的*K*-means算法在异常检测中的应用[J].重庆理工大学学报,2015(5):66-70.
[75] Yoshida K,Adachi F,Washio T,et al.Memory management of density-based spam detector[C]//Symposium on Applications and the Internet,2005:370-376.
[76] Zabihi M,Jahan M V,Hamidzadeh J.A density based clustering approach for Web robot detection[C]//International Conference on Computer and Knowledge Engineering,2014:23-28.
[77] Nguyen T D,Cao T D,Nguyen L G.DGA botnet detection using collaborative filtering and density-based clustering[C]//International Symposium on Information & Communication,2015:203-209.
[78] Alpaydin E.机器学习导论[M].北京:机械工业出版社,2014.
[79] Ratkiewicz J,Conover M,Meiss M,et al.Detecting and tracking political abuse in social media[C]//International Conference on Weblogs and Social Media,Barcelona,2011.
[80] Torabi A A,Taghipour K,Khadivi S.Web spam detection:new approach with hidden Markov models[C]//Asia Information Retrieval Symposium,2013:239-250.
[81] Chitrakar R,Huang C.Anomaly detection using support vector machine classification with *k*-medoids clustering[C]//Asian Himalayas International Conference on Internet,2013:1-5.
[82] Ying X,Wu X,Barbara D.Spectrum based fraud detection in social networks[C]//IEEE International Conference on Data Engineering,2011:912-923.
[83] Chandola V,Banerjee A,Kumar V.Anomaly detection:a survey[J].ACM Computing Surveys,2009,41(3):1-58.