

社会化媒体下的信息安全技术研究

□ 张思龙 / 南京政治学院上海校区军事信息管理系 上海 200433

摘要: 社会化媒体方兴未艾, 由此产生的信息安全问题也日渐显露。文章分析了社会化媒体的特点、面临的安全威胁, 并从信息监督、隐私保护等方面, 探讨了当前社会化媒体下信息安全的关键技术, 以期构建安全的社会化媒体环境提供技术参考。

关键字: 社会化媒体, 信息安全技术, 大数据, 隐私保护

DOI: 10.3772/j.issn.1673—2286.2014.02.005

1 引言

社会化媒体是一种区别于传统媒体的新兴媒体形式, 它依托于互联网, 将媒体、媒介以及受众之间的边界模糊化, 实现了媒体和受众的有机统一。社会化媒体与传统媒体的根本性区别在于网络用户的积极互动性。社会化媒体中, 用户之间逐渐产生了直接或间接的联系, 这就促使着网络媒体、社区网络的交叉与渗透, 形成了以互联网为载体的海量社会化媒体信息。在这种新的媒体形态下, 基于用户关联关系的多层次信息发布、传播以及获取形式逐渐将从媒体到受众单向传播的传统模式取代。这种多层次信息传播模式导致信息传播过程复杂化, 对网络媒体的安全性提出严峻挑战。同时社会化媒体也导致信息源激增且不可控, 社会化媒体信息安全保障显得格外迫切。如何利用信息安全技术, 保障社会化媒体信息安全, 对于提升媒体信息服务的可信性和可控性, 对于下一代互联网信息服务引擎的研究和应用具有十分重要的意义。

2 社会化媒体的特点

Antony Mayfield在提出社会化媒体概念的同时, 总结了社会化媒体的六个特征: 参与、公开、交流、对话、社区化和连通性^[1]。同时, 按照Kaplan等人的定义, 社会化媒体是“一组建立在Web 2.0技术和意识形态基础上的、基于互联网的应用, 允许用户生成内容 (User Generated Content, UGC) 的创造和交换。”当

然从表现形式上来看, 社会化媒体可以包括博客、维基百科、播客、论坛、社交网络和微博等, 并且各自的运作方式不尽相同。社会化媒体与传统的新闻媒体相比, 无论是信息产生、信息传播以及参与者等各方面都表现出新的特点。

2.1 社会化媒体存在信息碎片化特征

信息通过用户制造内容 (User Generated Content, UGC) 产生是社会化媒体的根本特征之一^[1]。信息生产、传播以及集中等, 都对用户没有任何身份门槛限制。同时, 社会化媒体鼓励和激发人们主动贡献和反馈内容, 并在媒体与用户之间双向传播、甚至是用户与用户之间的多对多传播, 社会化媒体和受众之间的界限也因此被极大的模糊。正是因为这种信息产生模式, 再加上社会化媒体的表现形式多种多样, 以及随处可见的移动智能终端使得用户可以使用更多的碎片化时间进行信息发布、交流, 交流的内容也不受限制, 从而使得信息的产生从发布时间到信息内容都具有碎片化特征。

2.2 在社会化媒体中存在社交化特征

在社会化媒体中, 允许用户之间以来源地区、兴趣爱好、关注主题等相同点建立密切联系, 并形成各种网络社区和用户群体^[2]。相比信息内容本身, 人们对于信息来源、用户关系等同等看待甚至更加重视。人们在日

常生活中不断创造信息、发布信息、交流信息的目的也是为了维护和拓展这种依靠信息分享而建立的关系网络,在关系网络中树立个体形象并强化自己的社区地位,而不仅仅是为了进行信息自我传播。社交化特征在一定程度上降低了社会化媒体内容的公共关注度和大众影响力,提高了用户归属感,这对研究社会化媒体的信息产生和传播规律,保障信息安全具有十分重要的意义。

2.3 社会化媒体具有大数据特征

社会化媒体所蕴含的信息量正以千万级的数量增加,只微博一种社会化媒体平台,每天产生的数据规模就达几千万条,社会化媒体数据正成为大数据的重要来源之一^[3]。

2.4 社会化媒体存在混沌化特征

正是社会化媒体的形式多种多样,并且其内部控制机制较少,才使得用户规模和信息规模非常庞大,各种信息高度透明且流动自由。社会化媒体的系统处于一种有序和无序之间的状态,其中各种热点事件的产生与发展在时间上是非周期性的、空间上也不规整,蕴含着丰富的随机性,难以进行事前预测和事后控制,但同时它也并非是完全无规律可言的混乱和无序,其内部有着社会化媒体独特的结构和发展规律。

正是基于这些特点,社会化媒体很容易吸纳众多社会普遍关注的热点新闻事件,并使得其扩散速度、影响范围以及影响力度都超乎想象。比如近几年出现的“西安药家鑫案”、河北大学“李刚门”案等热点事件,其中有保护弱者的正义之举,也不乏心怀不轨的故意煽动。对该类信息,如果不进行仔细辨别和监控,可能会使一些虚假、色情、暴力、赌博等有害和不良信息扩散,从而对信息内容安全造成威胁^{[4][5]}。

3 社会化媒体面临的信息安全挑战

社会化媒体产生出庞大的信息量,其中蕴藏着巨大潜力,随着快速提取、处理和分析技术的发展,可以快速捕捉到有价值的信息以提供参考决策,然而随之而来的是信息安全的挑战。

3.1 信息监管问题

庞大的用户群体和高度自由的信息发布,使得信息扩散无法得到有效监管。尽管通过信息内容分级可以对互联网信息传播进行引导和控制,但社会化媒体信息的碎片化、信息来源的草根性以及传播过程的复杂性使得对信息来源、传播的监管变得异常困难^[3]。

3.2 个人隐私问题

目前对社会化媒体的用户数据缺乏规范,更缺乏监管,大量事实表明,海量的用户数据如果不得到妥善处理,对用户隐私会造成严重威胁^{[4][5]}。同时社会网络分析相关研究显示,通过分析用户在社会化媒体中的群组关系,可以对用户的身份进行定位,这都间接造成用户身份等隐私信息的暴露^[6]。不仅如此,通过移动智能终端的使用,社会化媒体还包含了大量的用户位置信息等,少量的信息可能不会对用户隐私造成威胁,但是将多种行为聚集起来,其隐私就很有可能会暴露。这种隐含的数据暴露往往是人们无法预知和控制的,而且对个人的影响也是无法估量的。

3.3 虚假信息问题

由于信息生产环节对生产者没有任何限制,使得社会化媒体中充斥着各种各样的信息,而参与的用户往往缺乏甄别事实的能力,对信息本身的真实性和准确性无从知晓,这样,信息的客观性就会被忽视,虚假信息就乘虚而入^[7]。比如前一段时间出现的“微博打拐”、“北京大妈碰瓷老外”等事件,这都对媒体信息的可信性造成严重威胁。

4 社会化媒体信息安全保护关键技术

针对社会化媒体信息面临的信息监管、隐私保护、虚假信息的安全挑战,展开社会化媒体信息安全保护技术的研究,本文着重选取部分重要领域进行介绍。

4.1 社会化媒体主题挖掘技术

针对社会化媒体的深度语义挖掘,主题模型已逐渐成为主流的研究思想^[2]。在这里,主题的概念类似于话题,但又不等同于话题,主题往往包含比话题更丰富的信息,可以认为话题信息是包含在主题信息之中的^[9]。

主题模型的起源是隐性语义索引 (Latent Semantic Indexing, LSI), 隐性语义索引并不是概率模型, 因此也算不上是主题模型, 但是其基本思想为主题模型的发展奠定了基础。在LSI的基础上, Hofmann提出了概率隐性语义索引PLSA (Probabilistic Latent Semantic Analysis, PLSA), 该模型被看成是一个真正意义上的主题模型。而LDA (Latent Dirichlet Allocation, LDA) 等模型又在PLSA的基础上进行了扩展得到一个更为完全的主题概率生成模型。现在有很多基于主题模型对话题关联数据建模的研究工作。特别是针对海量的网络媒体数据噪声大、动态变化快等特点, 面向关联挖掘的话题识别研究工作具有广泛的应用背景。

4.2 信息过滤技术

信息过滤技术采取适当的技术措施, 对不良信息进行过滤, 既可阻止不良信息对人们的侵害, 适应社会对意识形态方面的要求, 同时, 通过规范用户的上网行为, 提高工作效率, 合理利用网络资源, 减少病毒对网络的侵害, 这就是信息过滤技术的根本内涵。

信息过滤技术主要包括基于内容的过滤技术和协同过滤技术^[5]。基于内容的信息过滤的理论基础主要是人工智能的理论和知识。日前基于内容的方法主要有基于统计、基于知识和二者结合的方法。基于统计的方法速度较快可以提高系统的实时性, 如决策树方法、支持向量机等, 而基于知识的方法能更好的提高系统的准确率, 如中国科学院声学研究所提出的语境框架模型等。

协同过滤, 又称社会过滤, 其基本思想十分直观, 认为人们往往会根据亲朋好友的推荐来做出一些日常生活选择, 比如购物、阅读、音乐等等。协同过滤系统就是将这一思想运用到信息安全保护中, 基于用户群体对某一信息的态度对该信息评价。

4.3 用户隐私保护技术

用户隐私保护首先要从负责用户数据管理的企事业单位着手, 建立健全用户数据收集、存储以及使用的相关规范, 并做好监督工作。从技术层面上讲, 实施数据隐私保护主要考虑以下两个方面, 第一, 确保在数据应用中不泄露用户隐私; 第二, 对海量社会化媒体数据进行有效利用^[6]。当前, 相关工作主要集中在设计隐私

保护原则和算法时既能保护用户隐私又能做到数据有效开发, 达到两者平衡状态。目前主要包括数据失真、数据加密、访问控制三类方法, 具体的有数据匿名保护技术、安全认证技术、基于角色的控制权限技术以及风险自适应的访问控制技术等等。社会化媒体中用户数据包含大量的关系型数据, 攻击者正是依靠用户的各种关系型数据, 包括标签、度数、中介度以及关系网络来识别用户的真实身份信息, 所以针对这一类数据需要在数据发布时对信息进行失真或隐藏, 包括用户标识和用户属性等数据信息。

4.4 信息真伪性辨别技术

信息真伪辨别技术能够提高虚假信息的识别能力。一方面, 社会化媒体数据具有大数据特性, 引进大数据分析的方法可以获得更高的识别准确率, 例如针对社会化媒体发布的特定信息, 可以通过分析发布者的位置分布信息、时间分布信息以及发布者历史行为信息进行信息的真伪性辨别。另一方面, 可以通过基于特征的方法对海量信息进行筛选, 研究表明虚假信息在发布者分布、发布时间分布以及内容分布上具有一定的趋同性, 可以通过采集信息的特征对信息分类, 比如S. Jamali等人^[10]将用户评论量和社会关系网络属性作为信息主题特征, 提出了信息分类预测算法等。

5 结语

社会化媒体带来了新的信息安全问题, 但其自身也是解决问题的重要手段。本文结合社会化媒体的特点, 从社会化媒体的信息监督、隐私保护、知识产权保护等方面对当前社会化媒体下的信息安全相关关键问题进行了阐述。总体而言, 当前国内外针对社会化媒体的研究还不充分, 一方面是它的产生时间相对较短没有深厚的研究积累, 另一方面相关政策法规还不尽完善, 只有通过技术和法规的共同完善, 才能更好的解决社会化媒体下的信息安全问题。

参考文献

- [1] KAPLAN A M, HAENLEIN M. Users of the world, unite! The challenges and opportunities of Social Media [J]. Business Horizons, 2010, 53(1): 59-68.

- [2] 杨士强,崔鹏.社区媒体信息传播机制[J].中国计算机学会通讯,2011,7(21):20-25.
- [3] 张银林.网络媒体发展中的信息安全问题探讨[J].信息安全与技术,2012,3(4):10-12.
- [4] FENG DENG GUO, ZHANG MIN, LI HAO. Big Data Security and Privacy Protection [J]. Chinese Journal of computers, 2013, 36(10): 1-13.
- [5] 李留英. Web2.0信息内容安全[J].数字图书馆论坛, 2009(9):16-22.
- [6] 朱克学.社交化媒体时代强力身份保护刻不容缓[J].中国传媒科技,2013(2):82-83.
- [7] 张尼,王志军.新媒体不良信息传播机制及对策研究[J].信息通信技术,2011(5):50-53.
- [8] 钟瑛.新媒体传播的社会问题及其规避[J].郑州大学学报(哲学社会科学版),2012(6):14-16.
- [9] 徐戈,王厚峰.自然语言处理中主题模型的发展[J].计算机学报, 2011,34(8):1423-1436.
- [10] JAMALI S, RANGWALA H. Digging Digg: Comment mining, popularity prediction, and social network analysis [C]// Proceedings of International Conference on Web Information Systems and Mining, Shanghai, 2009: 32-38.

作者简介

张思龙 (1987-), 硕士, 南京政治学院上海校区军事信息管理学助教, 研究方向: 信息安全。E-mail: ahang35@126.com

Research of Information Security Technology of the Social Media

Zhang Silong / Department of Military Information Management, Shanghai Branch of Nanjing Political College, Shanghai, 200433

Abstract: With the social media flourishing, the resulting security problem is more and more serious. In order to provide a frame of reference for guarantee of information security, this paper analyzes the characteristic and information security threat of social media, after that, this paper discusses the key technologies of social media information security in terms of information supervision and privacy protection, etc.

Keywords: Social media, Information security technology, Big data, Privacy protection

(收稿日期: 2014-01-08)