



开放科学
(资源服务)
标识码
(OSID)

“中科院之声”知识传播现状研究

曾灵

华南师范大学经济与管理学院 广州 510006

摘要: [目的/意义] 中国科学院作为我国最高学术科研机构, 一直活跃在国际科技前沿。本文以中国科学院官方微博为例, 探究其官方微博“中科院之声”在微博平台的知识传播现状, 把握知识传播过程中的主题和热点, 发现知识话题的传播规律和特征, 分析微博评论情感倾向, 为科学知识的网络化传播提供建议和参考。[方法/过程] 本文选取自“中科院之声”成立以来2013—2020年间发布的微博文本数据, 结合内容分析、词频分析等对目前发展的情况进行梳理, 并以微博评论为对象, 建立情感分析规则, 探究了该机构在微博平台进行知识传播过程中的受众情感倾向。[局限] 本研究暂时只考虑了中国科学院在微博平台的知识传播现状, 对其在其它社会网络平台的发展现状研究不足, 有待进一步的深入探讨; 本研究在进行情感分析过程中采用的是情感分析词典, 得到的结果不可避免存在一定的偏差, 如何通过机器来识别大脑中的复杂情感网络仍然有待科学家们的深入研究。[结果/结论] “中科院之声”通过微博平台积极传递着科学知识, 发布的微博内容以自然科学领域的最新研究进展、科学知识的科普等为主, 发布的内容得到了网友的积极关注和讨论, 发布微博的评论总体上呈现积极。但也存在内容题材不够“接地气”、负面新闻没有及时回应等问题, 建议增设实用性科学知识专题、加强与网友的交流和互动、及时关注传播数据的变化作出内容调整等。

关键词: 中科院之声; 知识传播; 现状研究; 情感分析

中图分类号: G35

作者简介 曾灵 (1996-), 硕士, 研究方向为信息计量、科学评价等, E-mail: 1350790827@qq.com。

引用格式 曾灵. “中科院之声”知识传播现状研究 [J]. 情报工程, 2021, 7(6): 14-30.

Research on the Current Situation of Knowledge Dissemination of “the Voice of the Chinese Academy of Sciences”

ZENG Ling

School of Economics and Management, South China Normal University, Guangzhou 510006, China

Abstract: [Objective/ Significance] As the highest academic and scientific research institution in China, the Chinese Academy of Sciences has been active in the forefront of international science and technology. Taking the official micro-blog of the Chinese Academy of Sciences as an example, this paper explores the status of knowledge dissemination of the official micro-blog “the Voice of the Chinese Academy of Sciences”, grasps the themes and hot spots in the process of knowledge dissemination, discovers the dissemination rules and characteristics of knowledge topics, analyzes the emotional tendency of micro-blog comments, and provides suggestions and reference for the network dissemination of scientific knowledge. [Methods/Process] This paper selects the micro-blog text data released from 2013 to 2020 since the establishment of “the Voice of the Chinese Academy of Sciences”, and uses content analysis, word frequency analysis and other methods to study the current development. Based on the establishment of sentiment analysis rules, this paper explores the audience’s sentiment tendency in the process of knowledge dissemination on micro-blog platform. [Limitations] For the time being, this study only considers the status of knowledge dissemination of the Chinese Academy of Sciences in micro-blog platform, but the research on its development status in other social network platforms is insufficient, which needs to be further explored. In this study, we use sentiment analysis dictionary in the process of sentiment analysis, and the results are inevitably biased. How to identify the complex emotional network in the brain by machine still needs to be further studied by scientists. [Results /Conclusions] The results show that : “the Voice of the Chinese Academy of Sciences” actively transmits scientific knowledge through the micro-blog platform. The micro-blog content mainly includes the latest research progress in the field of natural science, the popularization of scientific knowledge and so on. The content has been actively concerned and discussed by netizens, and the comments on micro-blog release generally show positive. However, there are also some problems, such as the content is not “grounded” enough, negative news not responding in time, and so on. It is suggested to add practical scientific knowledge topics , strengthen communication and interaction with netizens, and timely pay more attention to the changes of influential data to adjust the content.

Keywords: the Voice of the Chinese Academy of Sciences; knowledge dissemination; current situation research; Sentiment analysis

引言

随着现代技术的发展, 互联网平台已逐渐成为信息交流和文化传播的重要渠道, 线上学术交流也随之兴起, 一大批科研机构纷纷开设网络平台, 为知识传播搭建起桥梁, 学术信息

的在线交流模式开始蓬勃发展。在众多的网络平台上, 新浪微博以其低成本的信息获取、便捷的信息交互而获得众多科研机构的青睐。微博用户可以通过多种移动终端接入微博平台, 实现以文字、图片、视频等多媒体形式的信息即时分享和传播互动, 其具有使用门槛低、内

容丰富、传播快速、获取便捷、时效性高等特点。从2009年8月上线以来,新浪微博一直保持着爆发式的增长,根据《微博2020用户发展报告》可以得知,2020年微博月活用户达到了5.11亿,用户群体呈现年轻化趋势^[1]。微博平台的开放性使得它不仅具有社交网络的功能,其更加倾向于表现出社会媒体的功能。首先,该平台拥有大量的专家学者和丰富的学术资源,每天产生大量的学术信息,这是其他学术交流媒介所不能比拟的。其次,微博学术信息更新快,相比于期刊、会议等,该平台对于学术前沿信息的传播更加迅速。纵观以往的学术交流,数据信息的难以获取成为了学术交流的障碍之一,而微博的学术信息资源获取的成本低,用户可以便捷获取公开学术信息。更重要的是,无论是权威学者还是普通学者,均可以通过该平台进行平等的、“面对面”的学术交流,这让跨界交流变得频繁,更利于碰撞出智慧的火花。所以,综上所述,微博学术交流的优势逐渐突显^[2]。而中国科学院作为中国自然科学的最高学术机构,自建立以来即与科学共进,以国家富强为己任,培养了大批优秀的科研人才,为我国的科技进步、经济发展等作出了巨大的贡献。所以,自2013年5月中国科学院在微博平台开通官方微博“中科院之声”以来,迅速获得了大量关注。截止2020年底,“中科院之声”已发布微博31311条,吸纳的粉丝数将近450万。中国科学院作为国家战略科技的重要力量,其网络的传播影响力是其进行知识传播和社会服务的重要体现,所以研究其知识传播的发展现状对于提升其网络传播影响力、营造崇尚知识、尊重科学的交流氛围具有重要的意义^[3]。

1 相关研究综述

通过搜索相关的文献资料,笔者发现近几年关于学术科研机构的网络传播的研究主要集中在以下几个方面:王刚等^[4]认为在科研传播的过程中,传统的传播方式如纸质介质、电子介质等有一定的局限性,而科研社交网络则可以更快速传播,且具有交互性、共享性和可信性,能够满足用户的个性化需求,对于科技创新发展具有重要的意义,肯定了社交网络对于学术信息交流传播的重要性。国外著名的医学期刊非常注重利用自媒体来加强期刊的宣传,如*Nature*医学姊妹刊的每篇文章都提供给Connotea、Digg、Facebook、Newsvine、Delicious、Twitter等平台进行交互传播^[5]。并且,赵金萍等^[6]通过研究发现,在网络时代,公众对于与自身利益相关的科研成果表现出较为强烈的参与意愿,愿意积极参与到科学事件和科学信息的评议当中去。赵雅馨等^[7]通过获取微博用户关注的科研数据来探究微博用户进行科学关注的特点,对文献的类型、所属学科及研究主题等都展开了研究。其研究发现,论文(包括期刊论文、预印本、会议论文)是微博平台中进行科学传播的最主要的内容形式,其次是科学新闻,研究报告的形式也受到了微博用户的关注。医学和生物学是关注度较高的学科领域,而物理、化学等由于公众参与度较低而成为了微博关注中的“冷门”学科。该研究还发现,能够引起大量用户关注的科学主题主要是具备科学趣味性的文化话题、国家重大科研突破的新闻话题或是与生活息息相关的健康话题,且主题的专业性越强,关注度的消退就越快。

还有学者研究对比了国内和国外的学术社交媒体平台,发现我国的学术社交网络平台在定位、运营和发展中存在着缺乏全球化运营理念、用户黏性差、内容生产缺乏有效的激励手段等问题^[8],科研媒体平台对于科学事件的评议既不专业也不够理性,还容易造成负面的网络舆情^[5]。在如何更好促进科学传播和科技创新方面,学者胡昭阳^[9]提出众包科学的新思路,认为可以基于互联网和数字媒体向社会开放社会化协作的科研创新,通过利用公众的集体智慧,达到科学研究创新的目的,除了可以与公众建立较为紧密的联系之外,也可以更加低成本、低风险、创造性地实现科技创新。祝娜等^[10]研究了国外 Twitter 平台用户的传播数据,发现科研用户在科技创新中扮演着重要的角色,而且路径一般是从科学界到社会公众,从权威用户到普通用户等,建议增加有号召力的领袖信源的数量、增强网络社群中信源的信息交互能力等。除此之外,学者李杨等^[11]还分析了网络舆论的群体极化给科研人员参与科学传播活动所带来的影响,认为科研人员中的意见领袖对于大众舆论的引导作用并不明显,并且随着时间的推移,科学事件的多次极化讨论反而会降低各方对于这些事件的关注和参与,呼吁不同的机构应该就加强合作,让公众更好地参与到科学传播活动中来。而媒体平台也应该在尊重客观事实的基础上公正、理性报道学术事件,避免纯粹制造舆论来吸引眼球^[5]。

除此之外,此类平台文本内容的情感分析历来也是学者们关注的重点。马超等^[12]以上海公安机构微博为例,研究政务微博评论中的情感极性分析方法,探究微博已有属性和这些情

感极性之间的关系。国外学者 Holmberg 等^[13]研究了天文物理学家推特上的文本情感态度与链接内容的关系,认为学者们倾向于分享知识,而不是发表评论,对推荐内容的态度通常为积极的,但是不同微博提及的情感程度是不同的,因此需要根据特定语境分析不同的指标。Friedrich 等^[14]则分析了带有科学论文链接的推文的情感态度,发现目前的情感分析工具还不能精确的判断推文对科学论文的情感极性,应该提高假积极情感的识别能力,所以,情感分析方法的改进也是研究的热点之一。目前的情感分析方法可以总结为基于机器学习的方法和基于情感词典或者知识系统的方法。李岩等^[15]研究了基于微博短文本的情感分析,设计了针对用户文本评论情感分析的舆情分析系统用于识别情感倾向和演化规律来了解用户对于热点话题的观点和态度。王勇等^[16]则基于情感极性词典来对微博的文本内容进行情感分类的,并且在原有词典的基础上,不断改进情感词典,使得情感的识别和判断更加准确。除此之外,还有基于深度学习、基于半监督学习、基于混合向量模型、融合表情符号等的微博文本情感倾向分析,分析的内容包含了公共图书馆形象、网络股评、突发公共事件舆情分析等。

通过以上的内容梳理可以发现,目前对于科学知识的网络化的传播的研究逐渐成为学者们关注和讨论的重点。在此研究背景下,本文试图选取具有典型代表性的“中科院之声”作为本文的研究对象,关注此类学术科研机构在网络化时代进行知识传播的情况,通过分析现状揭示目前科学知识网络化传播的特征,并通过分析微博评论情感倾向,探究受众对于科学知

识传播的态度,为学术知识的网络传播贡献一份力量。

2 数据来源

在数据获取方面,本文选取中国科学院认证微博“中科院之声”的官方微博文本数据,该官方微博自2013年5月正式开通以来,受到了广大热爱科学、热爱知识的网友的关注。2013年5月27日,“中科院之声”正式加入微博平台,并在同年5月31日正式发布第一条官方微博,之后一直秉承着“以科学的眼光看世界,以全球的思维论科学”的理念积极进行学术知识的交流和传递。从“中科院之声”2013年正式发布第一条微博以来,到2020年12月31日,其发布的微博的数量已经达到了31311条。通过使用八爪鱼数据采集工具,本文将这些数据按照时间排序进行爬取,爬取的内容主要包括:微博原文发布的时间、发布的详细内容、每一条博文对应的转发数、评论数和点赞数等。本研究还选取了自2019年4月—2020年12月期

间微博原文的26581条评论作为本文的情感分析数据集,探究中国科学院在微博平台进行知识传播过程中的受众情感倾向。

3 传播内容的统计分析

3.1 发布微博的数量分析

“中科院之声”作为中国科学院的官方微博账号,代表的不仅仅是单纯的科学知识传播主体,其拥有着强大的“专家库”来确保传播内容的科学性、严谨性,是一个涵盖了科学家、新媒体工作者等专业人才的知识传播队伍^[17],所以,在网络时代,该官方微博账号肩负着知识传递的神圣使命。如图1所示,经过对31311条微博信息的不同年份的数量统计可以得知,除了2013年发布的微博数量为1812条,之后的每一年发布的微博数量均在3000条以上。经过计算,该官方微博平均每天发布的微博数量在9~12条之间。说明自中国科学院开通微博线上交流平台以来,积极发布微博,更新和传递科学知识。

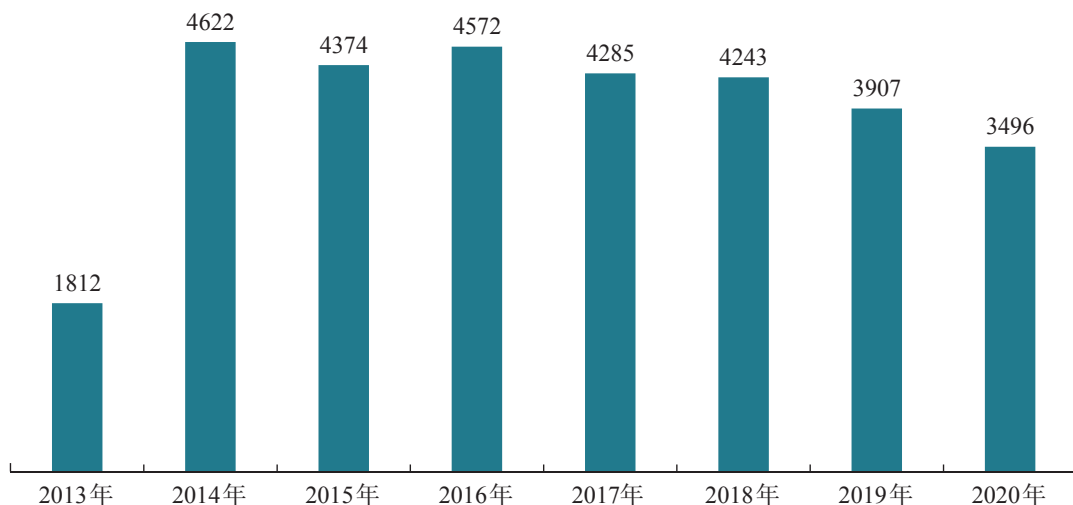


图1 不同年份发布的微博数量

3.2 发布微博的内容分析

经过对发布的 31311 条微博的粗略统计和分析，可以发现，发布的内容大致可以划分为以下几类：某领域最新研究进展或研究发现、科学史介绍、科学知识科普、博文转发或回复、其他，共五大类，详细含义如表 1 所示。同时，对不同类型的微博的数量进行统计可以观察到，

官方微博发布数量最多的内容是科学知识的科普，占据了发布微博数量的 51.40%。除此之外，发布数量较多的依次是不同领域的研究进展或研究发现、科学史介绍、科学知识的科普、博文转发或回复、其他，这些内容涵盖了自然科学各个领域的研究热点，构成了“中科院之声”科学知识传递的主要内容。

表 1 发布微博内容分类

发布的微博 内容分类	发布微博 数量（条）	含义解释
某领域的最新研究 进展或研究发现	10288	对近期国际或国内相关科研团队取得的某个领域的创新性突破和重大研究成果进行报道
科学史介绍	2717	陈述历史上某个时期发生的重大科学事件、介绍在不同领域对科学发展起推动作用的重要人物等，微博内容的开头通常带有“#科学史#”的话题标签
科学知识科普	16095	对科学领域的各类专业名词、不同事物的作用机理等进行介绍和描述，几乎涵盖了所有主题
博文转发或回复	1904	针对被提及的发布内容进行转发扩散或回复网友的评论
其他	343	重要科学人物的讣告、周年院庆、辟谣声明等

在 web3.0 时代，许多网站经常使用词云图来显示用户搜索过的关键词，从而给后来的用户提供搜索的参考指引。词云图作为文本可视化的重要工具之一，可以直观地体现出不同的词组在特定文本中出现的频率。它由众多的关键字或者标签组成，每个词的重要性程度可以根据字体的大小和字号的粗细进行判断。在词云图中，越是处于图片的中间位置、字体和字号越大，通常表明该词出现在文本中的概率就越大，也就越重要。本文在获取微博原文数据之后，编写了相应的 python 程序语言，在筛选、去除无意义词语的基础上，将微博原文的文本内容导入编程语言，得到文本内容的高频主题词及其频次。制作相应的词云图如图 2 所示，可以发现，“科学史”“材料”“生物”“植物”“系统”等高频词尤为凸显，说明这些内容在发布

微博的文本内容中出现的频率较高，同时也体现了这些主题的重要性。

3.3 高热度微博内容分析

本文选取评论数量最高的 8 条微博进行分析，可以看到，评论数量较高的微博内容主要涉及自然科学领域研究的重大性突破、公众关注的热点科学事件等。如评论数量最高的是近段时间关于天文学家捕获的首张黑洞照片的消息，这是人类历史上的重大突破，这一突破让人类对于此类天体有了更加清晰的认识，也标记着人类在对物质世界的探索不断取得新进展。这是一项由 200 多位科研人员进行国际合作所完成的非凡的科研成果，技术的突破、智慧的交流、创新的算法等汇聚到一起，打开了一个关于黑洞和全球视界的全新窗口。所以，这个

消息一经报道,就引起了网友们的广泛关注和讨论。网友表示:“见证历史了”“牛啊,有生之年系列!”“永远歌颂科学!”。其次,评论数较高的内容是关于重要科学人物逝世的讣告,于敏院士作为著名核物理学家、中国核武器事业重要奠基人,曾获得“两弹一星”功勋奖章、国家最高科技奖、全国道德模范称号等,他的逝世令人惋惜和悲痛,网友们纷纷表达对于老一辈科学家逝世的哀思和崇高敬意。如不少网友纷纷留言道:“国土无双,老先生千古!”“于敏爷爷走好,向您致敬,最伟大的人!”等。讨论度高的微博内容还包括一些不实信息的重要声明,如一些机构企图冒充中科院研究所发布一些不实内容来误导公众从而博得关注,中科院及时发博澄清,并附上官方的权威机构列表供公众参考识别,提醒公众注意甄别此类冒名机构。此外,还有一些重要论

文成果的科普专题,比如暗物质粒子卫星“悟空”工程团队最近在《Nature》杂志上发表了最新的重大突破性成果,对于这种专业性较强的知识内容,“中科院之声”微博官方为了让网友们能够理解这些晦涩难懂的专业术语和物理知识,特意采用生动形象的图画做了一期物理专题知识科普,并且以“孙悟空大战葫芦娃”的故事作为知识内容的叙述脉络,幽默风趣的语言、灵动飘逸的动画设计吸引了大批网友进行围观,引起了网友们的热烈讨论。有网友表示:“科普文案做得很生动,吸引人!从头看到尾,并且看懂啦,哈哈,‘悟空’好棒!”“我看懂了!我看懂了!!谢谢‘悟空’,谢谢文案绘图小哥哥小姐姐,我竟然看得懂!”这一期的科普取得了非常良好的效果。还有一些讨论度较高的内容是关于中科院对于新冠病毒谣言的澄清说明和其相关知识的科普等如表2所示。



图2 传播内容高频词云图

表2 高热度微博传播内容

序号	发布内容	转发数	评论数	点赞数
1	【天文学家捕获的首张黑洞照片公布】该黑洞图像揭示了室女座星系团中超大质量星系 Messier 87 中心的黑洞。该黑洞距离地球 5500 万光年, 质量为太阳的 65 亿倍。图中心的暗弱区域即为“黑洞阴影”。	87950	36212	207246
2	【讣告】著名核物理学家, 中国科学院院士, 中国核武器事业重要奠基人, “两弹一星” 功勋奖章、国家最高科技奖、改革先锋称号、全国道德模范称号获得者于敏同志因病医治无效, 于 2019 年 1 月 16 日 13 时 35 分在北京 301 医院逝世, 享年 93 岁。沉痛悼念于敏院士!	20694	14367	102638
3	【2018 年诺贝尔化学奖揭晓】Frances H. Arnold 和 George P. Smith、Sir Gregory P. Winter. 获得今年诺贝尔化学奖。	11923	12549	44858
4	【声明】中国科学院理化技术研究所杨冰同志不幸去世有关情况的说明(网页链接)关于我所杨冰同志不幸去世有关情况的说明	6908	12203	1737
5	【统一回复】感谢关注, 这里是“中国科学院”。	20653	10875	53666
6	【回复】无此研究所, 机构列表见——网页链接	29947	7920	149518
7	【或许, 你正在见证“物理课本”的改写!】暗物质粒子探测卫星“悟空”工程团队 2020-11-30 日在 Nature 发表论文。论文非常厉害, 但是怕你们看不懂论文, 我们做了个科普专题↓	22011	7073	61699
8	相关消息不实, 详见: 新冠病毒是人造? 科学解释为何阴谋论错得离谱, 新冠病毒基因组测序驳斥阴谋论, 持续寻找病毒 ... 科普可见——冠状病毒“这一家子”, 都是怎么回事?	2194	5234	33674

通常而言, 微博平台的“点赞”意味着网友对于博主发布的内容存在心理上的认同和赞赏, 是一种正向积极的情感表达; 而“评论”是针对博文的内容发表自己的见解和看法, 这种文字的表达有可能是积极正面, 也有可能是消极负面; “转发”在某种程度上也是一种对

于发布内容的认可, 且认可的程度较高, 通过转发这种方式提高微博内容在该平台的推荐度, 微博内容也因此可以获得很多的曝光度。按照表 2 的 8 条微博的评论数从高到低进行排序, 分别编号 1~8, 然后绘制这些评论数较高的微博传播数据图如图 3 所示。可以看到, 转发数和点

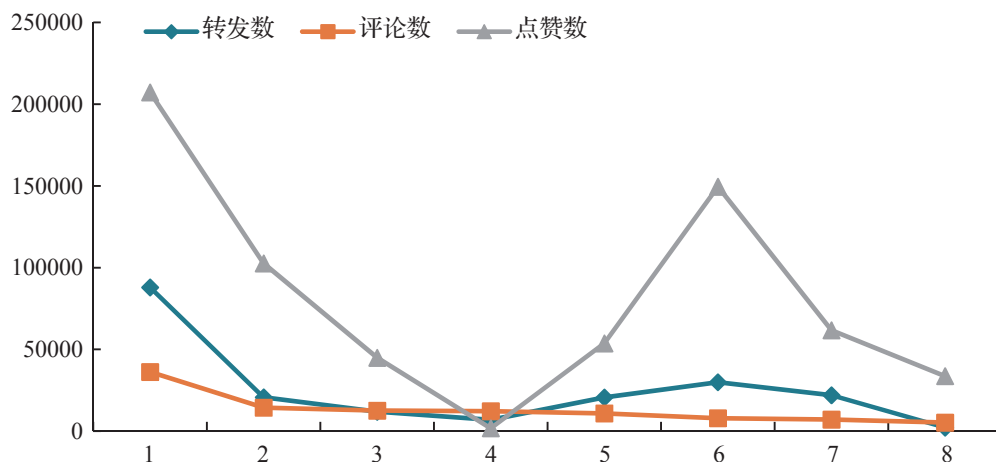


图3 高热度传播内容传播数据统计

赞数在整体趋势上较为一致,两者都表达对于微博内容的赞赏和认同的情感。通常评论数较高的微博,也会得到较多的点赞和转发。网友通过对这些内容的点赞、评论和转发,让这些内容获得了更多的推荐度,科学知识内容得到了更好的传递和扩散。

4 知识传播的主题变化情况

通过官方发布的微博内容结合 2013–2020 主题关键词及其频次如表 3 所示进行分析可以得知,从 2013 年开始,“中科院之声”已经开始发布

一些历史上某个时期发生的重大科学事件、介绍在某方面对科学技术的发展起重要推动作用的人物等内容。但从 2014 年开始,“中科院之声”开始正式设置了“科学史”专题系列的知识传播内容,所以,“科学史”一词贯穿了之后的每一年,成为了高频主题词。同时,“生物”“植物”“环境”等词始终保持着高热度,传播内容的主题主要围绕这几个方面展开。除此之外,“细胞”“疾病”“物理”“纳米”等词也成为传播的主题内容之一。2013–2020 年,传播的主题在大体上没有特别明显的变化,这也侧面说明“中科院之声”传播的核心主题较为稳定。

表 3 2013–2020 主题关键词及其频次 (Top 10)

2013年	频次	2014年	频次	2015年	频次	2016年	频次
嫦娥三号	191	科学史	404	科学史	415	科学史	385
月球	102	生物	234	生物	319	材料	316
机制	101	材料	210	材料	307	植物	278
系统	98	国际	207	植物	238	生物	262
生物	97	环境	204	环境	194	分子	230
地球	94	植物	191	分子	194	功能	199
植物	91	研究所	186	物理	174	环境	193
环境	89	分子	178	植物园	167	调控	193
工程	85	细胞	152	疾病	138	物理	191
实验	83	数据	137	纳米	138	细胞	190
2017年	频次	2018年	频次	2019年	频次	2020年	频次
科学史	355	科学史	395	科学史	367	科学史	384
材料	331	材料	352	材料	320	生物	279
生物	301	院士	301	分子	319	材料	253
分子	251	生物	290	环境	312	环境	251
环境	225	调控	246	生物	309	植物	248
调控	221	分子	244	调控	259	细胞	146
细胞	219	植物	236	植物	259	大气	127
植物	209	环境	227	细胞	230	病毒	117
量子	133	细胞	192	新进展	177	生态系统	112
纳米	133	纳米	191	数据	154	武汉	111

经过观察可以发现,“中科院之声”的知识传播也比较贴合时事,在不同时期及时传递着国内外发生的重大科学事件,搭建起信息交流和信息传输的桥梁。比如,2013年“嫦娥三号”探测器发射升空,实现了中国首次对地外天体直接探测的目标,这一创举极大地推动了国内外认识月球、研究月球和利用月球的探索热情,并且这个过程取得了大量创新成果。“中科院之声”对于“嫦娥三号”发射的过程进行了系列的跟踪报道,包括发射前的准备工作、需要具备的相关技术、可能面临的难题、取得成功的意义等,与网友们全方位分享了“嫦娥三号”各方面的资讯,让广大网友在“中科院之声”发布的信息中共同见证了这项伟大事业的完成。所以,在2013年的高频主题词中可以看到“嫦娥三号”“月球”“机制”等词频频出现。再如,2020年,新型冠状病毒在武汉爆发,成为举国关注的大事件,“中科院之声”开始发布关于新型冠状病毒的一系列研究进展。内容包括:科学家通过研究揭示了武汉新型冠状病毒进化来源、与其它病毒的基因组序列进行相似度比较、新冠冠状病毒论文的最新成果、哪些药物对新型冠状病毒有一定的抑制作用以及对于武汉病毒的网络谣言进行声明和辟谣等,还及时发布了新型冠状病毒信息库的链接地址供学术交流和研究使用。所以,2020年的主题词中,“病毒”“武汉”也成为了高频词。

5 知识传播中的情感分析

5.1 情感分析准备

社交网络情感分析^[18]是指利用自然语言处理技术来挖掘社交网络中用户的观点、态度和情感,主要用于网络的舆情监测和信息的预测等。在本文的研究中,采用基于情感词典的方法进行句子语义情感分值的测度,这种方法主要是对分词后的文本内容进行情感词的匹配,整合情感词进行不同权重的评分,最后得到文本的情感倾向分值。情感分析处理过程如图4所示,在得到相关的文本内容之后,先进行文本内容的预处理,去掉一些重复的内容、与情感分析无关的一些html标签等。为了判断句子中是否存在与各类词典相对应的词语,需要对这些句子进行分词,本文选择了python平台中的“结巴分词”作为本文的分词工具。情感词典主要分为:否定词词典、程度副词词典、积极情感词典、消极情感词典。为了得到更完整的词典,在结合需要分析的内容的基础上,本文还对不同词典的部分内容进行了整合和更新,使得得到的结果准确度更高。最后将需要进行情感测度的文本导入,匹配调整后的情感词典,最后计算不同文本的情感值即可得到结果。

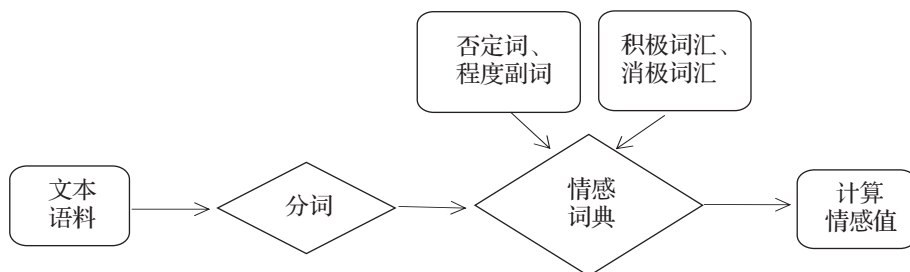


图4 情感分析处理过程

情感词典是本文情感分类最核心的部分, 本文采用的情感词典来源于 BosonNLP 情感词典, 这个词典来源于社交媒体文本, 是由波森自然语言处理公司推出的一款已经做好标注的

情感词典。词典中根据表达的情感极性的不同, 对每个情感词都进行了情感值的评分, 适用于进行社交媒体的情感分析, BosonNLP 情感词典样例如表 4 所示:

表 4 BosonNLP 情感词典样例

词语	分值	词语	分值
真无语	-5.60457743583	要淡定	0.0096523291731
T M	-5.60457743583	较量	0.749739904212
窝火	-4.39136025706	完好无损	1.11350709029
倒霉	-4.01036441213	初见成效	1.83923883486
透不过气	-2.88438335335	最开森	5.20945336563
华而不实	-0.140175366654	如虎添翼	6.37503943135

现有的 BosonNLP 情感词典虽然是针对社交媒体进行情感分析的词典, 但是由于时代的发展变化, 不同时期可能流行不同的“网络梗”, 包括词语、拼音、符号等多种组合, 在特定的网络媒介中传达特殊的含义, 久而久之形成特定的语言。这些网络流行语通常具有诙谐、幽默、新奇的特点^[19], 如“菜鸟”“辣鸡”“潜水”“我晕”等, 这些网络流行语经常被用于网民表达自己的观点和见解。为了使得情感的分类更加准确, 本文还根据得到的文本进行了词频分析和统计, 在此基础上, 筛选出现频次较高的带有主观性的情感词, 如“杠精”“键盘侠”“粉丝”等, 通过查找 BosonNLP 情感词典里意思相近的词

语, 取两者分值的平均值来人工赋予相应的分值, 并添加进该情感词典。

程度副词使用 hownet 情感分析词语集中的程度副词词集, 并且根据不同的预期强度进行了标记。如果大于 1 则表示情感的逐渐加强, 小于 1, 则表示情感的逐渐减弱, 本文按照“极其 1.8”“超 1.6”“较为 0.8”“稍 0.7”“微 0.5”进行了标记, 如表 5 所示。除了完善 BosonNLP 情感词典和根据不同强度标记程度副词词典, 在此基础上, 本文整合了包括哈工大停用词、四川大学机器智能实验室停用词、百度停用词等得到新的中文停用词表, 总共 2462 个, 为得到较为准确的情感分值打下了良好的基础。

表 5 程度副词词集标记

词语表示	极其	超	较为	稍	微
词语极性	1.8	1.6	0.8	0.7	0.5
代表词语	百分之百、彻头彻尾、充分、极、极度、极其、极为、绝、绝顶、绝对、完全、完完全全、之极、之至、至极	超额、过、过度、过分、过火、过劲、过了头、过猛、过甚、过于、强、溢、忒	多、更、更加、更较、进一步、愈、愈发、愈加、越... 越、越来越	点点滴滴、多多少少、或多或少、略微、稍稍、稍微、些微、些小、一点儿、一些、有些	半点、不大、不丁点儿、不甚、不怎么、没怎么、轻度、弱、丝毫、相对

在进行分词后,程序将自动遍历所有的文本内容,查找文本内容中的情感词、否定词和程度副词,并与不同的词典进行匹配。若情感词之间存在程度副词或否定词,则情感词之间的所有否定词或程度副词与后一情感词构成一个情感词组,例如在“今天很不开心”句子中“很”是程度副词,“不”是否定词,“开心”是情感词,那么得分就是 $SentiScore=(-1)^l \times 1.25 \times 0.77$, 其中 $(-1)^l$ 指句子中出现了 1 个否定词,而 1.25 和 0.77 则分别为程度副词、情感词词典的分值。所以,情感分值的计算可以表达为:

$$Si=(-1)^m \times Wi \times Ei$$

其中 Si 表示子句的情感得分; m 表示否定词的个数,由于汉语的特点,否定个数为单数时,表示否定,而当否定词个数为双数时,表示肯定; Wi 表示程度副词的权值; Ei 表示子句中情感词的权值。

5.2 情感分析结果

本文对选取时间内的学术微博的评论情感倾向进行识别,从评论的内容反映网友对于哪些科学知识表现出关注,关注的热点事件集中在哪里,中科院的知识传播达到了怎样的效果,这些评论会在微博形成强大的舆论场。在这些观点中,客观理性的观点与曲解误导的谣言并存^[20],稍有不慎,就会影响到中科院在学术界的声誉和权威,所以,对于其微博评论的情感进行分析是有必要的。由于在官方发布的微博中,部分微博原文底下并没有网友的评论和留言,为保证数据的科学准确性,本文选取了官方微博下发布的带有文字评论的相关数据,经

过初步筛选,共获取 26581 条有效数据,部分评论情感分值如表 6 所示。由于针对不同事件的评论可能存在较大的情绪差异,情感分值跨度范围较大,本文将通过进一步的细分来进行细腻情感的把握。根据情感的分值结果,结合情感词典,如果情感分值大于 0,可以认为该文本内容传递的情感为积极,根据不同的程度可以划分为:0~10 为一般积极情感,10~20 为中度积极情感,20 以上的分值可判断为高度积极情感。如果情感分值小于 0,可以认为该文本内容传递的情感为消极,根据不同的程度可以细分为:0~-10 为一般消极情感,-10~-20 为中度消极情感,-20 以下为高度消极情感。通过对样本数据集的统计,可以得到如图 5 所示的结果。

从得到的结果可以看到,这些评论绝大部分分值处于 0~10 分之间,属于一般积极情感,此类情感的评论数量为 17382 条,占到了总数的 65.39%。这些评论的内容主要是针对中科院发布的关于科学界发生的重大事件进行的讨论,比如,官方微博发布“为促进新型冠状病毒肺炎科研攻关,搭建了中国科学院新型冠状病毒肺炎科研文献共享平台”时,网友表示“这方面也一定要和美国那个数据库竞争一下”“共享很重要”“感谢科学家们付出的努力,希望疫情快点结束”等。再比如官方微博发布了“2020 年诺贝尔物理学奖揭晓”的报道,网友们表示“虽然看不懂也听不太懂,但很牛”“救命,真的是太伟大了”等表达对获奖者的敬佩、赞叹之情。而中度积极情感的分值范围在 10~20 之间,选取的样本中有 599 条属于该范围,这些评论的积极情绪的强度更大,表达的情感

也更加强烈。比如,当发布“中国科学家构建了76光子量子计算原型机,使得我国成功达到量子计算研究的首个里程碑”的消息时,网友们纷纷留言“微系统所牛逼!”“中科大是越来越优秀了噢”“虽然看不懂,但能感觉到是

个了不起的成就”“2020 科学界最激动人心的成果!”等,表达了对中国科学家在该领域取得的研究突破的激动、开心、自豪之情。高度积极情感的内容同样表达了对于祖国科技进步、日益强大的自豪和喜悦的情感。

表6 微博评论分值示例(部分)

评论内容	情感分值
生日快乐!很荣幸能成为中科院华南植物园的志愿者,四舍五入我也算中科院的人了哈哈哈	12.041030200283
感谢科学家们付出的努力,希望疫情快点结束。	6.477689991637
老院士一路走好!保家卫国铮铮风骨	4.9459588461685
GISAID篡改数据,我们要有自己的独立研究,加油	3.150635366827
期待近视可药物治愈的那天	0.590822761376
谁都会犯错,不是么	-1.07661253882
这种说话毫无逻辑和理论基础,甚至语言表达不文明的就别理了,出发点和态度就有问题,无药可救	-2.33455438796019
中科院发这个实在有点莫名其妙	-3.010823357297
唉。。。最近经常有这样的讣告,好难受	-5.97154103101
就很无语一堆话忽然觉得对着这样脑残的人说不出口	-6.4976749895

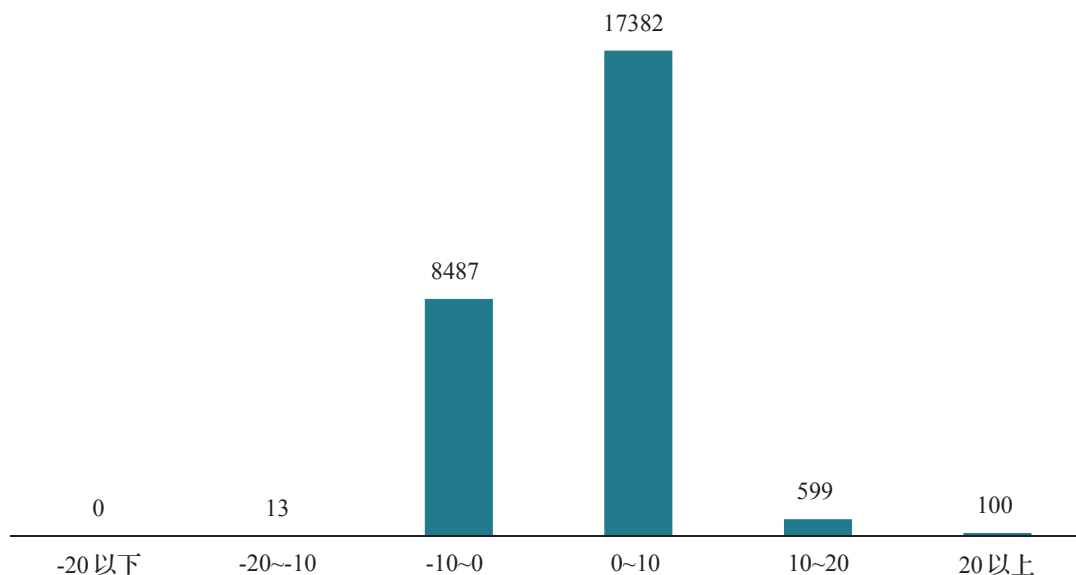


图5 微博评论情感分值数量统计

一般消极情感评论数量为8487条,占据了评论总数的31.93%,经过进一步的细分,可

以发现,8487条评论中有6397条评论的分值为处于-2~0分的区间,占据了一般消极情绪

的 75.37%，可以认为消极情感的情感强度非常微弱，基本上可以定义为“中性情绪”。而这些内容主要是网友对于发布信息内容的“调侃”和理性讨论。比如中科院之声发布的其中一条博文“在很多人的印象中，化肥不仅和土壤污染划等号，使用了化肥种出来的蔬菜、粮食也都不太健康。然而，事实果真如此吗？”网友们评论“事实是没有化肥我们不可能吃饱饭。”“合理使用不会造成伤害”“吃饱饭了就觉得化肥不健康了”。而在所有被认为带有消极情绪的评论中，仅有一小部分评论中带有相对明显的消极情绪。比如有言论是关于网友反映中国科学院招聘工作中存在“限制男女比例”的情况，于是网友们开始在发布的不同的博文下面提出“男女比例要求的事说说呗”“男女比例的事儿不回应一下吗？”“为什么土木院就没有女生，肯定是有人暗中搞歧视”“矿工男女比例失调那么严重，也没看见有女拳师站出来嚎几句，填补一下女矿工的缺”。由于官方微博没有及时对相关事件进行解释和回应，导致官方发布的某一些博文中虽然并没有涉及男女比例的事情，网友们也会自发地在不同博文下“刷屏”，希望得到官方的合理解释和回应。

6 讨论和建议

社会化媒体很重要的特点是和用户之间的相互交流，这也是其强劲的亲和性表现之一，因而信息的传播能力更强^[21]。这就意味着，在网络时代，脱离了学术“象牙塔”保护

的科研机构在网络平台中需要直面互联网上多样化、差异化的人群，这也对当今的科研机构的网络化传播提出了更高的要求。从上面的分析中可以得知，“中科院之声”发布内容占比最多的是对于各类自然科学知识内容的科普，从 2013 年加入微博平台以来，其一直致力于传播自然科学领域的科学知识内容。这些内容往往专业性强、涉及复杂的科学技术原理或物体作用机理，涵盖了大量的专业名词和行业术语。但对于普通网友而言，在没有专业知识基础的前提下，难以对这些复杂难懂的科学知识产生浓厚兴趣。本文随机选取了两条科普类的博文如表 7 所示进行观察。第一条微博内容是关于脱落酸（ABA）的科普，该博文在内容上采取了专业性的文字表达，并附上带有详细内容的网页链接进行进一步的介绍。但是该内容并没有受到网友较多的关注，除了 10 个转发和 17 个点赞以外，没有网友发表评论。另一条博文是与其类型相同的科普博文，主要介绍土壤质地与土壤肥力之间的关系，该内容则引起了较为热烈的反响，网友们围绕该内容进行了积极的讨论、转发、点赞。进一步观察可以发现，该博文的内容在题材上较为贴近日常生活，并且引用了接地气的俚语作为话题的引出，带给网友亲切感。其次，该博文在发布文字内容的同时，还配上了一张不同土壤的对比图，让人对即将阐述的知识主题和内容在视觉上有了更加直观的理解，也更容易引起网友进行阅读的兴趣，所以该博文收获的讨论度和关注度远远高于前者。

表7 科普博文传播效果对比示例

序号	发布内容	转发数	评论数	点赞数
1	脱落酸 (ABA) 作为一种重要的植物激素, 参与调控植物的生长发育、逆境响应。泛素介导的蛋白酶体降解途径, 在激素的信号转导过程中起着至关重要的作用。中科院遗传发育所发现泛素介导的脱落酸信号通路调控机制, 并且为研究不同细胞器之间信号转导提供了一些线索。(网页链接)	10	0	17
2	农民把大的土块称为“坷垃”, 俗话说: “庄稼就怕草, 更怕坷垃咬”。土壤质地与土壤性状如养分含量、通气透水性、保水保肥性以及耕作性状等都有很大的关系, 所以, 在说土壤肥力时, 质地是首先考虑的因素之一。下面的文章让你知道土壤质地与土壤肥力的关系。(网页链接)	3568	1042	346

从整体而言, “中科院之声”发布的不少微博内容均获得了网友较高的关注度, 如全球发展中的环境问题、科学界的重大技术突破等, 这些内容得到了网友们的积极评论、点赞或转发。但关注度低的内容也不在少数, 这些内容往往存在不够“接地气”、缺乏实用性、晦涩难懂等问题。当这些专业科学知识不再只是在专业圈子流传时, 应当注意进行在线交流时的信息通俗化, 让大家真正明白研究意义所在。科学需要社会的关注, 但是更需要的是对内容的认同和理解。因此, 本文认为: 作为开放性的网络科学交流平台, 应该尽可能避免局限于严肃、单一的题材, 加大对生活类知识内容的重视程度, 多从日常生活中获得科学知识传递主题的灵感, 调动网友们阅读和观赏的欲望, 这样既有利于引起网友们的关注, 又有利于他们进一步学习和掌握科学知识; 在传递专业性强、复杂难懂的科学知识时, 在保证内容科学性、真实性、权威性的前提下可以适当融入一些网络流行语或者方言俗语等来增加知识的趣味性, 通过诙谐幽默的语言拉近科学与公众的距离, 提升网友的兴趣度和生活贴近感, 避免对大量专业术语的平铺直叙或繁复数据的简单罗列和堆砌; 对于网友们“喜闻乐见”的内容, 可以

增设相应的专题系列等。

微博的精髓在于交流、互动, 在微博世界里, 与读者进行“朋友式的交流”是最为重要的, 经常与粉丝进行互动是一种很好的增加粉丝黏度的方法^[16]。网友在对微博进行评论的同时, 往往也会进行转发, 转发的次数越多, 这些知识内容越容易被推荐, 获得的曝光度也就越大, 使得科学知识得到更广范围的传播。经过进一步的观察发现, 如果官方账号参与了某一条微博的讨论和互动, 那么该微博的讨论度整体而言要比官方没有参与互动的微博的讨论度要高。

“中科院之声”可以抓住网友的心理需求, 就网友关注的动态和话题及时进行回复和转发, 让他们感受到较强的社群意识, 从而形成归属感。这种持续性的激励会使得关注了官方账号的网友愿意对“中科院之声”保持持续的关注和互动, 这也是科学知识传递的有效方式。再者, 虽然“中科院之声”发布的微博的评论总体上是积极的, 但是某些网友们会因为与该机构相关的一些负面事件而在内容不相关的微博下面进行“刷屏”来希望得到官方机构的重视和回应, 如上文提及的“男女比例事件”。说明在互联网高度发展的今天, 网友对于社会事件的参与意识越来越强烈, 越来越多的网民希望通

过互联网平台来发出自己的声音。学术科研机构在享受学术交流、知识传播快捷便利的同时,也必须要面临社会负面舆论带来的冲击,来鞭策自己“有则改之,无则加勉”。所以,舆论引导在学术界同样重要,在面对网友的质疑和批评时,此类机构应该积极主动表达官方立场和看法,掌握话语的主导权,积极评论转发,参与到与网民的互动和讨论中来,发挥意见领袖的作用,及时澄清解惑,引导舆论往更加积极的方向进行转变,制止谣言的发酵和传播,促使更多网友理性思考。

7 结语

本文通过对中国科学院的官方微博号“中科院之声”进行分析,探究作为重要学术科研平台的“中科院之声”在知识传播中所担任的角色定位和作用。研究发现,“中科院之声”自成立以来就积极搭建起科研知识和学术信息传递的桥梁。关注的主要领域为生物、环境、物理等,并且随着年份的变化,发布内容的核心主题并没有明显的变化,但是发布的具体内容会紧随时事,在不同时期及时传递着国内外发生的重大科学事件,让网友能够了解热点科学事件的最新进展并且持之以恒进行科学知识的科普。但是也存在着对于贴近生活的科学知识内容的关注较为欠缺、部分内容表达方式过于直板、严肃等问题。而在对在受众情感倾向的测度中可知,“中科院之声”发布的微博的评论总体上是积极的,但是某些网友们会因为与该机构相关的一些负面事件而在内容不相关的微博下面进行“刷屏”来希望得到官方机构

的重视和回应。舆论引导在学术界同样重要,在面对网友的质疑和批评时,机构应该积极主动表达官方立场和看法,参与到与网民的互动和讨论中来,制止谣言的发酵和传播。

最后,本研究还存在着以下局限:首先,本文只选取了中国科学院在微博平台的知识传播现状,但是仅局限于某一个平台并不能全面反映出其在社交媒体平台的发展全貌,需要对中科院在其它网络社交平台的知识传播发展现状进行全面追踪研究方可得出更加科学、严谨的结论。其次,本研究在情感分析中采用情感分析词典的方法,得到的结果不可避免存在一定的偏差,并且人类的语言系统是相当复杂的,文本和文本情感都是人类的产物,人类作为唯一的判别标准会导致文本的情感分类工作相当困难。词典模拟的是简单的思维定式,如何通过机器来识别大脑中的复杂情感网络仍然有待科学家们的深入研究。

参考文献

- [1] 微博数据中心.《2020 微博用户发展报告》[EB/OL]. [2020-12-31]. <https://weibo.com/ttarticle/p/show?id=2309404613871951282183>
- [2] 盛宇.基于微博的学术信息交流机制研究——以新浪微博为例[J].图书情报工作,2012(14):62-66.
- [3] 袁志彬,李猛.中国国立科研机构网络传播力的实证研究——以中国科学院北京分院所属研究院所为例[J].科普研究,2020(3):48-53+110-111.
- [4] 王刚,黄崑,贺曦冉,等.科研社交网络对科技论文传播的影响[J].中国高校科技,2017(6):43-45.
- [5] 刘冰,游苏宁.从中华医学会系列杂志初步使用博客谈医学期刊的自媒体应用[J].编辑学报,2012(4):364-368.
- [6] 赵金萍,戴晓晖,刘云章,等.新媒体主体在科学事件评议中的问题与责任——以韩春雨撤稿事件

- 网络评议为例[J]. 中国医学伦理学, 2019,(1):46-50.
- [7] 赵雅馨, 杨志萍, 陆颖. 基于替代计量学的微博用户科学关注特点研究[J]. 情报理论与实践, 2017(2):117-121.
- [8] 许志敏. 提高我国学术社交网络的国际传播能力——基于 ResearchGate 与“科研之友”等的比较研究[J]. 科技与出版, 2018(7):26-32.
- [9] 胡昭阳. 众包科学: 网络传播语境下的公众参与创新[D]. 合肥: 中国科学技术大学, 2015.
- [10] 祝娜, 王芳. 信息交互视角下 Twitter 用户的科技创新传播行为研究[J]. 情报学报, 2020(12):1340-1353.
- [11] 李杨, 金兼斌. 网络舆论极化与科研人员对科学传播活动的参与[J]. 现代传播(中国传媒大学学报), 2019(3):32-37+42.
- [12] 马超, 余辉, 夏文蕾, 管为栋. 政务微博评论中情感极性分析方法研究——以上海公安机构微博为例[J]. 现代情报, 2020(3):157-168.
- [13] Holmberg K, Bowman D, Haustein S, et al. Astrophysicists' conversational connections on Twitter[J]. PLOS ONE, 2014, 9(8):0106086.
- [14] Friedrich N, Bowman D, Stock G, et al. Adapting sentiment analysis for tweets linking to scientific papers[EB/OL]. [2016-05-16]. <http://arxiv.org/abs/1507.01967>.
- [15] 李岩, 韩斌, 赵剑. 基于短文本及情感分析的微博舆情分析[J]. 计算机应用与软件, 2013(12):240-243.
- [16] 王勇, 吕学强, 姬连春, 等. 基于极性词典的中文微博情感分类[J]. 计算机应用与软件, 2014(1):34-37+126.
- [17] 林荣耀. 抖音号“中科院之声”的科普情况分析[J]. 卫星电视与宽带多媒体, 2020(4):121-122.
- [18] 张琦, 张祖凡, 甘臣权. 融合社会关系的社交网络情感分析综述[J]. 计算机工程与科学, 2021(1):180-190.
- [19] 伍冠宇. 网络社会中“刷梗”的裂变式传播及其影响[J]. 新媒体研究, 2020(18):75-77.
- [20] 冯纓, 游苏宁. 传统医学期刊与新媒体微博共赢发展[J]. 编辑学报, 2012, 24(3):269-271.
- [21] 刘伟竹, 刘冰, 陈丽, 等. 顺风扬帆正当时——传统医学期刊与微博共谋新发展[J]. 中国科技期刊研究, 2013, 24(1):170-173.