

# 议题凸显与关联构建:Twitter 社交机器人对新冠疫情讨论的建构<sup>\*</sup>

■ 师 文 陈昌凤

【内容摘要】在海外社交平台 Twitter 上,社交机器人参与到新冠疫情的英文讨论中,并进行议题的凸显和关联构建。结构化主题模型和语义相似度计算的结果显示,就讨论的主题而言,虽然人类更关注美国国内疫情,但是社交机器人热衷于发布世界上其他国家的疫情;社交机器人还被用于发布疫情流行的动态。就关联构建而言,在社交机器人的语料中,与“coronavirus”(冠状病毒)最相似的词汇为“wuhancoronavirus”(武汉冠状病毒),而在人类语料中,最相似的词汇为“virus”(病毒),这反映出社交机器人试图在病毒起源问题上进行倾向性影射,而人类的表述则更为客观中性。此外,社交机器人常常标榜其发布的内容为“coronavirustruth”(新冠疫情的真相),但是人类则更常使用“rumor”(谣言)一词,表达对信息的批判性审视。

【关键词】 计算传播学; 社交机器人; 新型冠状病毒; 社交媒体

## 一、引言

在世界范围内,社交媒体被认为是可以有效促进民主化及社会运动的平台<sup>①</sup>,社交媒体上的文本数据也被视作推测公众意见的“传感器”<sup>②</sup>。2020年初,新型冠状病毒肺炎在全球范围内的流行成为重大的风险事件,在疫情风险充斥现实空间的背景下,线下对话的空间进一步被挤压。对于公众而言,社交媒体成为理解危机事件的重要渠道;对于政策制定者而言,社交媒体上因承载海量公共讨论,因此既扮演公众认知的“传感器”,也成为各利益相关方角逐的焦点,实则扮演着反映和塑造公众认知的渠道性工具。基于此背景,社交媒体上公共讨论的真实性与透明性,关乎公众是否在免于操纵的情况下理性构建对新冠疫情的认知,也关乎政策决策者对公众意见的掌握是否真实。

但是,越来越多的证据显示,社交媒体上大量信息的作者并非人类,而是由被自动化软件控制的社交机器人所生产。<sup>③</sup>纵然社交平台一直在试图识别并清除社交机器人,但社交机器人可以进行行为的策略性调整以规避检测。<sup>④</sup>活跃的 Twitter 帐户中约有 9% 到 15% 属于社交机器人。<sup>⑤</sup>社交机器人往往被批量生产并投入使用,在自动化脚本的控制下模仿人类行为,比如发布内容并与其他用户互动社交。实证研究发现,社交机器人已经具备激发人类用户与之互动的能

力<sup>⑥</sup>,简单的策略化运营方案即可将机器人帐户提升为有影响力的用户。<sup>⑦</sup>少量社交机器人(占用户总量的 5%—10%)即足以引发沉默的螺旋,从而改变公众舆论的走向。<sup>⑧</sup>在 Facebook 上,社交机器人对社交关系网渗透的成功率高达 80%。<sup>⑨</sup>

此前的研究和报道已经发现,在新冠疫情相关的社交媒体讨论中存在大量的社交机器人,在此基础上,本研究试图进一步探讨社交机器人在疫情讨论中扮演的角色:(1) 新冠疫情的流行与防控是兼具政治属性和公共卫生属性的复杂公共事件,公众对新冠疫情的认知理应整体、系统,在新冠疫情的讨论中,社交机器人和人类在话题倾向上有何种差异? 社交机器人如何构建其他概念与新冠病毒的关联?(2) 新冠疫情在全世界范围内的扩散使其与国际政治发生了密切关联,中国和美国曾就疫情问题在舆论场上数次交锋,在涉及到中国时,社交机器人和人类在话题倾向上有何种差异? 社交机器人如何构建其他概念与中国的关联?

本文采用计算传播学方法回应以上问题,通过对海外社交平台 Twitter 上与新冠疫情相关的英文推文进行文本挖掘,将作者身份作为协变量引入文本的结构化主题模型以探讨作者身份对话题流行度的影响。此外,我们计算了机器人和人类语料库中的特定概念之间的语义相似度,以期厘清自动化舆论操纵在新冠

<sup>\*</sup> 本文系国家社科基金重大项目“智能时代的信息价值观引领研究”(项目编号:18ZDA307)的研究成果。

疫情讨论中所扮演的角色。

## 二、相关文献

### 1. 社交机器人与健康传播

此前研究发现,社交机器人积极介入了公共健康卫生事件的讨论。有些社交机器人被用于实现良善目的,如科学家曾设计了个性化社交机器人用于科普,以期对吸烟者进行个性化干预,提升公共健康意识。<sup>①</sup>但在更多情况下,机器人被认为是公共健康事件讨论中的恶意行动者,对公共健康构成威胁。此前研究发现,社交机器人会蓄意扩大健康议题中的某些侧面,而回避另一些侧面。比如,社交机器人倾向于强调大麻在治疗身体和心理疾病方面的医学效用,但较少谈及大麻的危害性<sup>②</sup>,机器人还声称电子烟是有效的戒烟辅助工具,并出于商业目的对某些品牌的电子烟进行推销,以及在疫苗接种的辩论中,社交机器人高频转发与其意见一致的用户的信息<sup>③</sup>,支持疫苗和反对疫苗的机器人策略性地同时扩大对立双方的声音<sup>④</sup>,通过“错误平衡”(false balancing)在疫苗有效性方面制造争议,以动摇公众接种疫苗的决心。

目前,新冠疫情的公共讨论中使用社交机器人进行舆论操纵的现象已经引发广泛关注,但是目前各方尚未就机器人的角色达成共识。此前,有媒体称中国社交机器人参与推广了 Twitter 上的话题标签#grazieCina,以期宣传中国在新冠疫情中对意大利的帮助。<sup>⑤</sup>《纽约时报》也发文指控社交机器人参与转发剑桥大学关于疫情的论文,该论文对病毒的中国起源论表示质疑;同时,由于 Twitter 上有两条称赞中国对疫情作出快速响应的推文得到大量转发,《纽约时报》推断其背后有社交机器人的贡献。<sup>⑥</sup>但是,学术界的发现却有所不同,有学者通过关键词的词频统计,发现在关于新冠肺炎的讨论中,社交机器人常常触及香港修例、台湾问题、维吾尔族人权问题,但是该学者认为社交机器人在新冠疫情危机中对此类问题的讨论可使中国侵犯言论自由和人权行为得到关注,并称社交机器人“促进了民主讨论”。

### 2. 社交机器人与中国议题

社交机器人被广泛应用于国内及国际政治博弈中。社交机器人往往以较高的频次发布大量信息,制造对某一方的虚假支持或抹黑。例如,在英国脱欧公投中,支持脱欧与反对脱欧的最活跃用户均为社交机器人,机器人用户贡献了英国脱欧讨论总流量的 32%<sup>⑦</sup>,它们还相互勾连、快速形成中小型的机器人转发网络以更快地扩散信息。<sup>⑧</sup>在日本选举中,社交机器人大规模重复发布和转发同一条消息,降低信息环境的多样性。<sup>⑨</sup>尤其值得注意的是,社交机器人发布内容

的真实性有待商榷,例如在委内瑞拉政治议题中,社交机器人伪装成政治领袖、政府机构以推广特定政治信息。<sup>⑩</sup>

在 2016 年美国大选中,IRA(俄罗斯互联网研究机构)曾被指控系统性地使用虚假社交媒体账号干预美国政治议程<sup>⑪</sup>,并影响特定意识形态的用户对信息的转发。<sup>⑫</sup>在法国大选的线上讨论中,大量社交机器人积极参与到“马克龙泄密”事件的传播中,而这些机器人的个人简介中却包含美国大选的竞选口号,研究者认为甚至存在专门进行社交机器人交易的黑市,使机器人在世界范围内的各个政治事件中被重复使用。<sup>⑬</sup>

社交机器人对中国相关问题的操纵也引发了国内外学者的关注。此前研究已经发现,大量社交机器人和虚假账户活跃于与“中国”相关的内政外交讨论中,有的机器人使用中文,针对使用 VPN 登录海外社交平台的中国网民进行自动化宣传。<sup>⑭</sup>也有大量机器人用英文发布内容,意图操纵英文社区内对中国相关议题的讨论,在 Twitter 上对中国相关议题的英文讨论中,社交机器人生产的内容占比超过五分之一,相比经济议题,机器人更愿意讨论中国的政体和人权问题,在涉港、澳、台、疆议题中,发表反对中国政府的立场的言论。<sup>⑮</sup>此外,在中美贸易战等事件中,学者也发现了社交机器人的身影。<sup>⑯</sup>在香港修例风波中,大量社交机器人扩散《纽约时报》对此事的报道,支持修例运动的关键词在机器人的用户名和推文中高频出现。<sup>⑰</sup>

### 3. 关联认知

虽然世界卫生组织倡导避免将病毒与特定地区进行关联,但是新冠疫情与中国的关联一直是科学界和舆论场上的焦点议题。一方面,作为首先在中国得到报告的全球公共卫生事件,新冠疫情中的中国所扮演的角色为全世界所关注。另外一方面,虽然病毒的起源尚在研究中,目前来自自然科学界的证据并未将中国与病毒起源进行关联<sup>⑱</sup>,但是新冠病毒源于中国的论调仍广为流传,美国总统特朗普用“Chinese virus”指代新冠病毒,美国国务卿蓬佩奥也多次公然声称新冠病毒源自武汉病毒研究所。<sup>⑲</sup>

认知科学认为,人类对事物的认知并非基于线性结构,而是呈现网络状态<sup>⑳</sup>,并据此提出“关联记忆”和“关联学习”模型<sup>㉑</sup>,即强调人们通过关联的方式对事物进行学习和记忆,使人类对于特定事物的认知与对头脑中其他事物的认知相互关联。<sup>㉒</sup>在认知网络中,节点代表一个个特定的概念,并代表概念之间的关联,如果两个概念总是被同时提及,则它们在人类头脑中的关联会被创立或强化,如果两个概念之间已经形成了关联,则刺激会更容易地从一个事物传导至另外一个

事物,使两个事物更容易在脑海中被同时激活。<sup>②</sup>麦库姆斯在提出议程设置的第三层——网络议程设置时也借鉴了关联记忆和关联认知模型。他认为鉴于个体基于关联关系对事物形成认知,那么媒体上对事物关联的呈现会影响用户构建概念之间的绑定关系,即媒体可以通过强调某些概念之间的关联,使它们在人类头脑中的关联更加显著,进而将对事物之间关联的塑造引入效果研究视野。<sup>③</sup>

虽然人类头脑中对事物中的关联建构难以被直接测量,但是语言学家认为,通过观察人类语言中对事物之间关联性的呈现,可以反向推测头脑中对于事物之间关联的认知。<sup>④</sup>有学者通过对大数据文本中的词汇共现关系进行语义网络分析,呈现围绕特定议题的网络化认知。<sup>⑤</sup>但是随着自然语言处理技术的发展,词汇之间的关联不仅可由显式的共现关系得以呈现,也可以在将词语向量化处理之后,通过计算词向量之间的距离,揭露概念之间的隐秘关联。顶级学术期刊 Science 上曾发表过一篇论文,其使用词向量计算人类语言中潜在的性别偏见和种族偏见<sup>⑥</sup>,并发现女性名字与家庭关系更紧密,而男性的名字与职业词汇关系更紧密;欧洲裔美国人的名字与“令人愉悦”关联更紧密,而非洲裔美国人的名字与“令人不悦”关联更紧密。有学者认为,隐式关联比基于共现关系获得的显式关联更能反映词汇之间的深层关系。<sup>⑦</sup>

### 三、数据

本研究选择海外社交平台 Twitter 作为研究对象。我们参考世界卫生组织官方网站上发布的时间线<sup>⑧</sup>,以世界卫生组织的行为为基准,挑选出三个重要时间点收集数据:(1)1月22日,世卫组织驻中国代表团发表声明称,有证据表明武汉的疫情存在人传人现象,但需要更多的调查以了解传播的全部范围,这标志着世界卫生组织开始注意并评估对新冠疫情的潜在传播能力。(2)1月30日,WHO 宣布将新冠肺炎疫情列为“国际关注的突发公共卫生事件”(PHEIC),标志着新冠疫情在世界范围内引发关注。(3)3月11日,WHO 宣布新冠肺炎疫情已经构成全球性大流行病(Pandemic),标志着新冠疫情在全球范围内的爆发。

由于本研究涉及疫情在世界范围内的流行,因此遵循 Twitter 平台数据收集规则,以事件发生时刻对应的 GMT+0 时区时间为标准确定检索日期,获取事件发生当天的推文。世界卫生组织将由新型冠状病毒引发的疾病命名为缩写 COVID-19,以表示“Corona Virus Disease 2019”,但是该缩写的确定日期为2月11日,而本研究的时间区间包含该缩写确定之前和缩写确定之后。为保证数据的一贯性,我们仍选择新冠病

毒的全称 #coronavirus(不区分大小写)作为检索关键词。

我们使用自主设计的 Python 脚本,抓取指定时间发布、含有上述关键词的原创英文推文。值得注意的是,该抓取方法要求推文作者亲自直接提及“#coronavirus”一词,如果作者转发了一条含有“#coronavirus”的推文,但转发时并未亲自提及该词,则本条转发推文不在抓取之列,该抓取逻辑确保了含有“#coronavirus”的文本由作者本人作出,与本研究试图探究作者身份与文本关系的研究意图相吻合。并且,我们清洗掉文本中的表情、符号、超链接,并将文本统一成小写。

### 四、方法

#### 1. 机器人识别

近年来,计算机学界已经意识到社交机器人对社交媒体公共讨论的干扰,大量算法被用于自动化地检测社交机器人。有学者利用社交机器人发帖频次较高的特征,将每日发帖频率超过某阈值的账户识别为机器人;<sup>⑨</sup>也有学者利用机器人与人类用户在文本特征、账户特征以及行为特征等方面的差异确定机器人用户与人类用户之间的界限。在目前的算法实践中,印第安纳大学开发的综合性机器人识别工具 Botometer(旧称 BotorNot)的使用最为广泛<sup>⑩</sup>,曾被用于检测美国大选、法国总统选举、英国脱欧事件中的机器人。<sup>⑪</sup> Botometer 将社交网络、朋友、用户个人资料、语言情感、时间活动模式等维度的信息注入机器学习模型,提取账户特征,生成一个位于 0—1 之间的机器人评分(bot score),评估特定账号为机器人的概率。机器人评分越接近于 1,则该账号有越大概率有自动化操纵的痕迹;机器人评分越接近于 0,则该账号有越大概率被人类所使用,我们参考此前研究的惯例,以 0.5 作为区分机器人和人类的临界值。<sup>⑫</sup>如果用户设置了隐私权限、已注销或被关停、推文数量不足以支撑分析,Botometer 则无法对其进行机器人概率评估,我们将这类用户标记为未知用户,排除在本研究的分析之外。

本研究共获取 195201 条推文,经检测,其中 18497 条(9.27%)由机器人生产,176166 条(90.25%)由人类用户生产,535 条(0.27%)由未知用户生产。以上推文共由 118720 名不同用户生产,其中,10865 名(9.15%)是机器人用户,107478 名(90.53%)是人类用户,375 名(0.32%)是未知用户。

#### 2. 结构化主题模型(STM)

在主题模型研究领域,LDA(Latent Dirichlet Allocation)自 2003 年被提出以来得到广泛应用<sup>⑬</sup>,但是 LDA 的局限性在于其只能输出主题分布,而不生成进行假设检验的置信区间。如果研究者希望探究元数据

对于主题分布的影响,进而对元数据中的变量和 LDA 输出主题分布的关系进行统计分析,这一结果会缺乏统计置信度。

为解决将主题模型应用于社会科学研究时面临的统计分析需求,2013 年,Robert 等人在 LDA 的基础上提出基于无监督学习的 STM(Structural Topic Model),将文档级别的协变量引入主题模型,在模型计算过程中输出一个或多个协变量与主题分布、词汇分布的关系。<sup>[4]</sup>对比实验证明,STM 主题模型中对变量计算的准确性高于将传统统计方法额外加诸 LDA 主题模型。<sup>[5]</sup>此外,STM 还可计算变量对于特定主题中词汇分布的影响<sup>[6]</sup>,该优点使 STM 在近年来的计算传播研究中得到越来越多的关注。有学者在分析“女司机”相关讨论的话题时,使用 STM 研究不同平台(报纸、微博)对于话题分布产生的影响。<sup>[7]</sup>也有学者<sup>[8]</sup>使用 STM 分析了媒体将网站上不同主题的新闻同步至社交媒体时所遵循的把关原则。

具体而言,我们使用 R 语言中的 STM 包对文本展开结构主题模型分析。在本研究中,我们训练了两个 STM 模型:(1)在 STM1 中,我们将机器人和人类的所有推文输入模型,并将作者身份(社交机器人、人类)作为主题流行度协变量引入模型,意在探究作者身份对于疫情讨论主题的影响。(2)在 STM2 中,我们仅将机器人和人类生产的含有“china”的推文输入模型,并将作者身份(社交机器人、人类)作为主题流行度协变量引入模型,意在探究作者身份如何影响疫情讨论中涉及“中国”的推文主题分布。STM 的主题个数 K 需要用户自行指定,根据 STM 模型的原理,K 并无绝对最优解,需要由使用者基于可解释性进行判断<sup>[9]</sup>。对于 STM1 和 STM2,我们均在  $K=6$  至  $K=14$  之间进行尝试,经过比较,我们将 STM1 中的 K 值确定为 12,将 STM2 中的 K 值确定为 8。

### 3. 词向量化及语义相似度

在本研究中,我们首先对语料库中的单词进行向量化。传统的热独编码思路为对每个词进行稀疏的 N 维热独(one-hot)编码,此处 N 为语料库的词汇量,相应地,语料库中每个词的索引为  $0 \sim N-1$ ,通过建立一个 N 维的向量,将每个词转化为一个独特的 N 维向量,在该向量中,只有该词的索引位为 1,其余位均为 0。然而,在这种传统的向量表示里,维度等于词典大小,而实践中词典大小往往在万词量级,这会导致维数灾难,矩阵过于稀疏导致空间利用率较低。此外,由于热独编码中任意两个向量两两正交,所以任意两个词的余弦相似度固定为 0,热独编码无法表达词汇之间的关系。在本研究中,我们采用更为前沿的 word2vec

技术,该技术由 Google 团队于 2013 年发布,其采用无监督神经网络,基于连续词袋模型(CBOW)或者跳字模型(Skip-gram),借助神经网络中的隐藏层将热独编码下的高维词语文本映射到较低维空间,通过训练获取新的向量表示,词与词之间的关联可以通过两个词向量之间的余弦相似度计算得出。<sup>[9]</sup>

具体而言,我们使用 Python 语言,调用 gensim 库中的 word2vec 接口,选择连续词袋模型,将机器人语料和人类语料分别输入模型,获取基于机器人语料的词向量和基于人类语料的词向量。随后我们使用 most-similar 接口,分别获取机器人和人类语料中,与“coronavirus”“china”关联最为紧密的 30 个词汇。

总而言之,在对推文作者身份进行识别之后,我们使用以上方法进行了以下四组分析:(a)在疫情讨论中,作者身份(社交机器人、人类)是否影响了话题的分布?(b)在社交机器人和人类的语料中,哪些词与“coronavirus”(新冠病毒)具有最紧密的关联?(c)在与中国相关的疫情讨论中,作者身份(社交机器人、人类)是否影响了话题的分布?(d)在社交机器人和人类的语料中,哪些词与“china”(中国)具有最紧密的关联?

## 五、结果

### 1. 疫情话题在机器人和人类中的呈现

表 1 呈现了 STM1 输出的 12 个主题及其高区别度词汇。除了主题 1 没有明确的意义指向之外,其他的 11 个主题涵盖了英国、美国、中国的疫情,疫情对经济、学校的影响,疫情的预防措施,医疗状况,公共健康等方面。

如上文所述,我们将作者身份(机器人和人类)作为主题流行度协变量引入 STM1 模型,用以探究人类和机器人在选择讨论主题时遵循的不同逻辑。图 1 展示了机器人和人类对每个主题的倾向。图中的零点表示标准化之后的各主题在语料库中的分布,如果话题分布位于零点左侧,则该话题在作者身份为机器人的推文中得到更多呈现;如果话题分布位于零点右侧,则该话题在作者身份为人类的推文中得到更多呈现。如果一个话题对应的横坐标值为 0.1,则表明该话题在人类推文中的呈现比其在机器人推文中的呈现多 10%。据图 1 显示,虽然大多数主题分布在零点附近,表明机器人和人类讨论中的疫情主题分布差异有限。相较而言,机器人更多发布英国疫情、中国疫情的信息以及疫情动态、预防措施,但人类更关注美国的疫情和疫情带来的计划变动。

表 1 疫情讨论中 12 个主题的意义及代表性词汇

主题序号	意义	高概率词汇
1	\	people time good flu thing call today month start coming ive thought die long best
2	英国等地疫情爆发	coronavirus ,covid ,coronavirusupdate ,corona ,well ,coronavirusoutbreak ,coronavirusk ,coronavid ,coronavir-ususa ,coviduk ,officially ,comoravirusk ,coronaoutbreak ,coronavirusupdates ,pandemia
3	美国疫情	coronavirus trump going cant realdonaldtrump panic american toilet paper ,president ,real ,medium ,house ,a-merica ,word
4	疫情带来的计划变动	coronavirus ,event ,cancelled ,march ,game ,cancel ,concern ,fan ,conference ,big ,canceled ,meeting ,decision ,postponed going
5	疫情动态	coronavirus ,case ,health ,italy ,death ,confirmed ,country ,state ,emergency ,number ,breaking ,update ,positive ,total ,tested
6	公共健康	coronavirus ,help ,news ,read ,live ,great ,watch ,question ,situation ,community ,expert ,latest ,free ,public ,advice
7	中国疫情	coronavirus ,china ,spread ,outbreak ,virus ,wuhan ,travl ,ncov ,quarantine ,chinese ,flight ,international ,iran ,city ,novel
8	医疗状况	pandemic ,coronavirus ,test ,testing ,disease ,hospital ,doctor ,cdc ,patient ,system ,spread ,healthcare ,vaccine ,classifies ,action
9	疫情的经济影响	coronavirus ,work ,business ,sick ,government ,crisis ,plan ,online ,student ,impact ,market ,budget ,economy ,class ,employee
10	疫情对学校的影响	coronavirus ,day ,school ,week ,close ,spreading ,fast ,social ,nation ,shut ,hour ,kid ,ago ,open ,area
11	预防措施	coronavirus ,hand ,virus ,dont ,stay ,keep ,corona ,wash ,safe ,mask ,face ,hope ,avoid ,kill ,worried
12	疫情风险	coronavirus ,life ,year ,care ,risk ,better ,symptom ,lot ,family ,told ,serious ,bad ,sure ,contact ,making

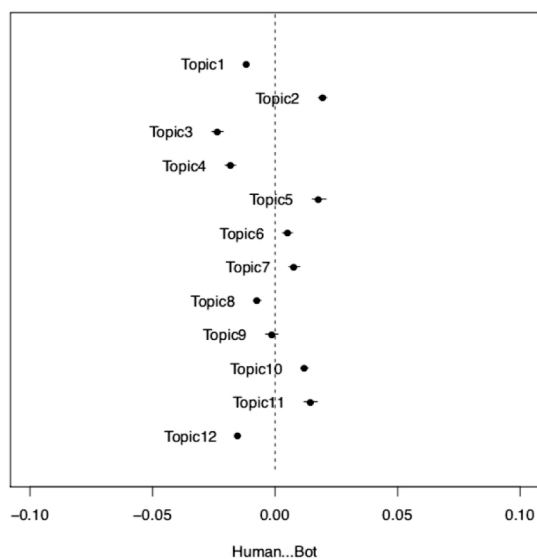


图 1 作者身份对疫情讨论主题的影响

## 2. 中国相关讨论在机器人和人类中的呈现

表 2 呈现了 STM2 输出的 8 个主题及其高区别度词汇,涵盖疫情预防治疗、各国疫情动态、疫情跨国传播、食用野生动物、政治阴谋等方面。

如上文所述,我们将作者身份(机器人和人类)作为主题内容协变量引入 STM2 模型,用以探究疫情讨论中,人类和机器人谈及中国时的主题分布。相较于图 1,机器人和人类在中国相关的主题讨论上有更大差异,机器人热衷于发布疫情动态,如死亡数、确诊数等;也更倾向于讨论疫情对于经济、贸易带来的影响。相对而言,人类更多地讨论疫情的预防治疗和跨国传播。

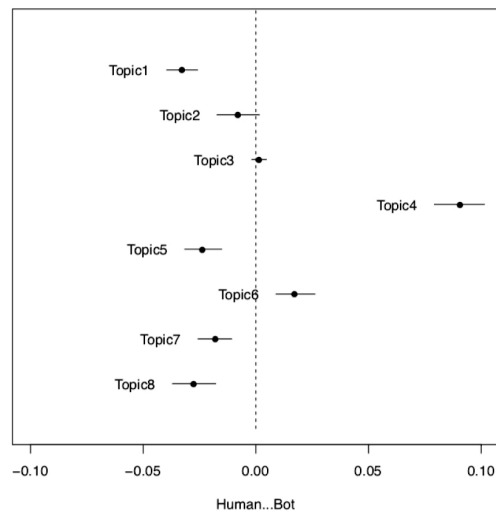


图 2 作者身份对中国相关的疫情讨论主题的影响

## 3. 机器人和人类语言中的语义关联

基于对词向量的相似度计算,我们获取了机器人和人类分别如何围绕疫情、中国构建语义关联,并将结果呈现于表 3 中。比较机器人和人类语料中与“coronavirus”相似度排名前 30 的词汇,我们发现,在机器人的语料中,“wuhancoronavirus”(武汉冠状病毒)一词与目标词汇相似度最高,该词汇以“武汉”这一地区名对病毒进行命名,意在地区进行污名化,并试图构建中国武汉与病毒起源的关联。相比之下,在人类的语料中,与目标词汇相似度最高的是“virus”(病毒),该词是对冠状病毒这一事物的客观表达,未含有污名化意图。值得注意的是,“wuhancoronavirus”“wuhanvirus”

(武汉病毒)、“wuhannpneumonia”(武汉肺炎)在机器人语料中与“coronavirus”相似度分别位于第1名、第9名和第25名,但是二者在人类语料库中,与“coronavir-

us”的相似度均未进入前30名,这表明机器人有更明显的意图将“武汉”这一意象与病毒名称关联起来,在病毒起源这一敏感议题上进行倾向性引导。

表2 与中国相关的疫情讨论中8个主题的意义及代表性词汇

主题序号	意义	高区别度词汇
1	预防与治疗	mask ,american ,cant ,month ,work ,face ,vaccine ,working ,west ,feel ,ill ,treatment ,german ,lesson ,wearing
2	各国疫情动态	covid ,italy ,iran ,coronavirusupdate ,france ,rate ,spain ,slows ,coronavid ,coronavirusupdates ,covidchina ,chinalies ,mortality ,coronavirustality ,coviduk
3	亚洲疫情增长	pray ,russia ,contained ,alarming ,grow ,china ,growing ,passed ,close ,coronavirus ,inaction ,scary ,asia ,approaching ,full
4	疫情规模	case ,death ,confirmed ,total ,taiwan ,number ,reported ,worldwide ,toll ,update ,province ,transmission ,hubei ,rise ,mainland
5	政治与阴谋	communist ,hand ,blame ,best ,job ,regime ,shit ,democrat ,created ,left ,drug ,party ,bioweapon ,body ,weapon
6	经济影响	economy ,crisis ,economic ,true ,student ,money ,rapidly ,god ,cure ,asian ,aid ,step ,online ,oil ,demand
7	食用野生动物	human ,animal ,originated ,symptom ,resident ,center ,bat ,cdcgov ,respiratory ,wild ,wildlife ,prevention ,sign ,returned ,seattle
8	跨国传播	health ,emergency ,travel ,flight ,public ,declared ,international ,declares ,organization ,concern ,airport ,passenger ,continues ,screening ,thursday

表3 机器人和人类语料中,与“coronavirus”和“china”相似度排名前30的词汇

	coronavirus		china	
	bot	human	bot	human
1	wuhancoronavirus	virus	outside	chinese
2	covidindia	covid	iran	wuhan
3	coronavirususa	globally	southkorea	europa
4	atlantic	extent	date	ccp
5	coronaalert	further	toll	country
6	coronavirustruth	rumor	japan	chinas
7	coronavid	healthanalytics	slows	hubei
8	coronavirusec	handful	korea	iran
9	wuhanvirus	growing	mainland	taiwan
10	codvid	mend	deaths	korea
11	coronaoutbreak	worldwide	death	illegal
12	coronavirusindia	reagan	surpassing	epicenter
13	novelcoronavirus	stave	panama	nigeria
14	coronavirusinindia	corona	wuhan	hk
15	coronavirussenya	interpreted	total	african
16	coronavirusus	hopefully	soar	originated
17	covidespana	becoming	surpassed	daegu
18	combsdaelp	unknown	hubei	africa
19	coronaindia	undetected	rise	germany
20	forbes	bermuda	usa	hongkong
21	coronaviruscanda	amongst	indonesia	islamic
22	coronavirუსnyc	easily	vietnam	abroad
23	covid	jab	ratio	worstaffected
24	wuhanoutbreak	potential	italy	epicentre
25	wuhanpneumonia	possibility	raising	lombardy
26	gocorona	lengthen	south	fold
27	coronvirusuk	fast	recovery	battle
28	cdc	wider	globally	southkorea
29	trumpantichrist	elder	totaling	europes
30	covidusa	paranoia	discharged	origin

此外,在机器人语料中,大量高相似度词汇由“coronavirus”和国家名先后组合而成,用于表达特定国家的疫情状况,这与上文所述,社交机器人热衷于发布各国疫情信息相吻合。结合此前研究<sup>③</sup>,我们认为这反映出大量新闻机器人参与到疫情讨论中,被用于发布疫情动态信息及警报。

值得注意的是,在机器人语料中,“coronavirus-truth”(新冠病毒真相)一词与目标词汇的相似度排名高居第6名,而在人类语料中,“rumor”(谣言)一词与目标词汇的相似度排名位于第6名。这一差异表明,社交机器人发言时经常标榜其发布的新冠相关内容为真相或内幕真相,而人类用户则更多提及谣言这一概念,表达对信息环境的批判性审视。

在社交机器人语料中,国外地名、疫情增长等词汇与“china”关联程度较高;相较而言,在人类语料中,出现更多的是与“china”相关联的中国国内地名。这一差异说明社交机器人更关注各国疫情增长,而人类更关注中国作为疫情“震中”的情况。

## 六、讨论与结论

作为一个全球性风险事件,新冠疫情具有多面性、复杂性,其在社交媒体上的呈现对形塑公众的认知有重要潜在影响,但是社交机器人的活动使社交媒体的信息环境更加复杂。本研究聚焦海外社交平台 Twitter 上关于新冠疫情的讨论,借助结构化主题模型及语义相似度计算,探究社交机器人如何凸显特定话题并构建事物之间的关联,以期影响新冠疫情在社交媒体上的整体呈现。我们发现,社交机器人与人类用户对

新冠疫情的讨论侧重点存在差异;即便当社交机器人和人类提及同样的概念,他们对概念之间关联的构建也存在差异。

此前对于社交机器人与健康传播的研究认为,社交机器人大量发布与科学共识相反的健康信息,增加了对公共健康的威胁。<sup>②</sup>但是在本研究中,我们发现社交机器人在健康讨论中的作用具有多面性。一方面,社交机器人被应用于进行疫情新闻速报,随着新冠疫情在全球范围内的流行,社交机器人大量发布世界各地的疫情发展动态,客观上起到了传播新闻的作用;但另外一方面,社交机器人发布信息的质量值得商榷,虽然社交机器人标榜其发布的内容为“coronavirustruth”,这一词汇暗示其试图揭露虚假现象幕后的真相,但是根据此前的研究<sup>③</sup>,机器人本身就是助推低质量信息扩散的始作俑者,他们热衷于转发来源于低可信度来源的消息,却甚少转发事实核查网站的消息。相比之下,人类用户在讨论新冠病毒时,常常提及“rumor”一词,反而折射出人类在面对信息洪流时有更强的批判意识,试图积极识别虚假信息。

社交机器人对于不同国家疫情的讨论强度差异也值得思考。根据权威流量统计网站 Statista 最新发布的 Twitter 用户构成,用户数量位居前列的国家分别是

美国、日本、俄国、英国、沙特阿拉伯<sup>④</sup>。同时,考虑到本研究将目标推文的语言限定为英文,在这种情况下,社交机器人对于英国、中国等国家疫情的关注和人类用户对于美国国内的疫情关注形成了微妙的反差。虽然计算传播学工具尚无法窥见 Twitter 社交机器人背后的动机,但机器人是否扮演了转移公众对于美国国内疫情关注的角色值得怀疑。同时,语义相似度计算的结果指出,社交机器人试图通过对病毒进行策略性命名的方式强化武汉与病毒起源之间的关联——这一关联本不存在于人类语义中,以期促成舆论场上对中国的不利讨论。这一发现与此前研究的发现不谋而合,共同指向了社交机器人被用于国际舆论场上的政治角逐,以期营造对中国的负面舆论。<sup>⑤</sup>

鉴于新冠疫情兼具公共卫生属性和国际政治维度,本研究聚焦于文本层面,试图探究社交机器人对以上两个纬度的话题凸显和关联构建。但是,由于社交机器人系以内容发布、转发、关注、点赞等多种方式对社交平台进行深度参与,本文难以呈现社交机器人在新冠疫情讨论中的角色的复杂全貌。在计算文本研究的基础上,未来研究或可基于社群传播、人机交互等视角对社交机器人在疫情讨论中的角色进一步挖掘,亦或对机器人和人类行为之间的因果关系进行统计学检验。

#### 注释:

- ① Meraz S. *Hashtag Wars and Networked Framing: The Private/Public Networked Protest Repertoires of Occupy on Twitter*. Between the Public and Private in Mobile Communication vol. 36 2017. p. 303.
- ② Kirilenko A. P., T. Molodtsova and S. Stepchenkova. *People as Sensors: Mass Media and Local Temperature Influence Climate Change Discussion on Twitter*. Global Environmental Change – human and Policy Dimensions vol. 30 2015. p. 92.
- ③ Wojcik S. *5 Things to Know About Bots on Twitter*. 2018; Howard P. N., B. Kollanyi and S. Woolley. *Bots and Automation Over Twitter During the US Election*. Computational Propaganda Project. Working Paper Series 2016; Schuchard R. et al. *Bots Fired: Examining Social Bot Evidence in Online Mass Shooting Conversations*. Palgrave Communications vol. 5 no. 1 2019. p. 5.
- ④ Luceri J. et al. *Evolution of Bot and Human Behavior During Elections*. First Monday vol. 24 no. 9 2019.
- ⑤ Varol O. et al. *Online Human – Bot Interactions: Detection, Estimation, and Characterization*. in Eleventh international AAAI conference on web and social media. 2017.
- ⑥ Shao J. et al. *The Spread of Low – Credibility Content by Social Bots*. Nature Communications vol. 9 issue 1 2018. p. 4792.
- ⑦ Messias J. et al. *You Followed My Bot! Transforming Robots Into Influential Users in Twitter*. 2013.
- ⑧ Cheng C., Y. Luo and C. Yu. *Dynamic Mechanism of Social Bots Interfering With Public Opinion in Network*. Physica A – statistical Mechanics and Its Applications 2020. p. 124167.
- ⑨ Boshmaf Y. et al. *The Socialbot Network: When Bots Socialize for Fame and Money*. in Proceedings of the 27th annual computer security applications conference. 2011. ACM.
- ⑩ Deb A. et al. *Social Bots for Online Public Health Interventions*. in advances in social networks analysis and mining. 2018.
- ⑪ Allem J. – P. P. Escobedo and L. Dharmapuri. *Cannabis Surveillance With Twitter Data: Emerging Topics and Social Bots*. American Journal of Public Health vol. 110 no. 3 2020. p. 359.
- ⑫ Yuan X., R. J. Schuchard and A. T. Crooks. *Examining Emergent Communities and Social Bots Within the Polarized Online Vaccination Debate in Twitter*. Social Media + Society vol. 5 no. 3 2019. p. 6.
- ⑬ Broniatowski D. A. et al. *Weaponized Health Communication: Twitter Bots and Russian Trolls Amplify the Vaccine Debate*. American journal of public health vol. 108 no. 10 2018. p. 1382.
- ⑭ Gabriele F. B. e. *How China Unleashed Twitter Bots to Spread COVID – 19 Propaganda in Italy*. Formiche. <https://formiche.net/2020/03/china-unleashed-twitter-bots-covid19-propaganda-italy/>. Accessed on June 26th.
- ⑮ Qiqing W. Y. L. *Behind China's Twitter Campaign: A Murky Supporting Chorus*. 2020.
- ⑯ Howard P. N., Kollanyi B. *Bots #Strongerin and #Brexit: Computational Propaganda During the UK – EU Referendum*. arXiv: Social and Information Networks. 2016.
- ⑰ Bastos M. T. and D. Mercea. *The Brexit Botnet and User – Generated Hyperpartisan News*. Social Science Computer Review vol. 37 no. 1 2019. p. 46.
- ⑱ Schaffer F. S. Evert and P. Heinrich. *Japan's 2014 General Election: Political Bots Right – Wing Internet Activism and Prime Minister Shinzō Abe's Hidden Nationalist Agenda*. Big data vol. 5 no. 4 2017. p. 301.
- ⑲ Forelle M. C. et al. *Political Bots and the Manipulation of Public Opinion in Venezuela*. arXiv: Social and Information Networks. 2015. p. 6.

- ②① Linvill J. D. L. and P. L. Warren. *Troll Factories: Manufacturing Specialized Disinformation on Twitter*. Political Communication. 2020. p. 16.
- ②② Badawy A. K. Lerman and E. Ferrara. *Who Falls for Online Political Manipulation?* in Companion Proceedings of The 2019 World Wide Web Conference. 2019.
- ②③ Ferrara E. *Disinformation and Social bot Operations in the Run up to the 2017 French Presidential Election*. First Monday vol. 22 no. 8. 2017.
- ②④ Bolsover G. and P. Howard. *Chinese Computational Propaganda: Automation Algorithms and the Manipulation of Information About Chinese Politics on Twitter and Weibo*. Information Communication & Society vol. 22 no. 14 2019. p. 2071.
- ②⑤ 师文、陈昌凤《分布与互动模式: 社交机器人操纵 Twitter 上的中国议题研究》,《国际新闻界》2020 年第 5 期,第 69 页。
- ②⑥ 张洪忠、赵蓓、石韦颖《社交机器人在 Twitter 参与中美贸易谈判议题的行为分析》,《新闻界》2020 年第 2 期,第 50 页。
- ②⑦① 师文、陈昌凤《社交机器人在新闻扩散中的角色和行为模式研究——基于〈纽约时报〉“修例”风波报道在 Twitter 上扩散的分析》,《新闻与传播研究》2020 年第 5 期,第 16、13 页。
- ②⑦ Forster P. et al. *Phylogenetic Network Analysis of SARS – CoV – 2 Genomes*. Proceedings of the National Academy of Sciences vol. 117 no. 17, 2020. p. 9241.
- ②⑧ David E. Sanger. *Pompeo Ties Coronavirus to China Lab Despite Spy Agencies' Uncertainty*. New York Times. 2020. May. 3rd.
- ②⑨ Downs R. M. and D. Stea. *Image and Environment: Cognitive Mapping and Spatial Behavior*. Transaction Publishers. 2017. p. 73.
- ③⑩ Lang A. *The Limited Capacity Model of Mediated Message Processing*. Journal of Communication vol. 50 no. 1 2000. p. 65.
- ③⑪ Santanen E. L. R. O. Briggs and G. – J. De Vreede. *The Cognitive Network Model of Creativity: A new Causal Model of Creativity and a new Brainstorming Technique*. in Proceedings of the 33rd Annual Hawaii International Conference on System Sciences. 2000. IEEE; Alonso E. and E. Mondragón. *Associative Learning and Behaviour: An Algebraic Search for Psychological Symmetries*. Language Representation and Reasoning: Memorial Volume to Isabel Gómez Txurruka 2007. p. 35.
- ③⑫ Collins A. M. and E. F. Loftus. *A Spreading – Activation Theory of Semantic Processing*. Psychological review vol. 82 no. 6 1975. p. 407.
- ③⑬ Guo L. H. T. Vu and M. McCombs. *An Expanded Perspective on Agenda – Setting Effects: Exploring the Third Level of Agenda Setting*. Revista de comunicación no. 11 2012. p. 55.
- ③⑭ Wettler M. and R. Rapp. *Computation of Word Associations Based on Co – occurrences of Words in Large Corpora*. in VERY LARGE CORPORA: ACADEMIC AND INDUSTRIAL PERSPECTIVES. 1993. p. 9.
- ③⑮ Meraz S. *Hashtag Wars and Networked Framing: The Private/public Networked Protest Repertoires of Occupy on Twitter*. Between the Public and Private in Mobile Communication 2017. p. 309; Kitzie V. and D. Ghosh. *# Criming and # Alive: Network and Content Analysis of two Sides of a Story on Twitter*. Proceedings of the Association for Information Science and Technology vol. 52 no. 1 2015. p. 3.
- ③⑯ Caliskan A. J. J. Bryson and A. Narayanan. *Semantics Derived Automatically From Language Corpora Contain Human – like Biases*. Science, vol. 356 no. 6334 2017. p. 184.
- ③⑰ 王晗啸、李成名、于德山、巴志超《基于上下文语义的网络议程设置研究——以红黄蓝事件为例》,《国际新闻界》2020 年第 5 期,第 88 页。
- ③⑱ WHO Archived: WHO Timeline – COVID – 19. 2020. <https://www.who.int/news-room/detail/27-04-2020-who-timeline---covid-19>. Accessed on June 26th.
- ③⑲ Howard P. N. B. Kollanyi and S. Woolley. *Bots and Automation over Twitter During the US Election*. Computational Propaganda Project: Working Paper Series 2016; Howard P. N. and B. Kollanyi. *Bots # StrongerIn and # Brexit: Computational Propaganda During the UK – EU Referendum*. Available at SSRN 2798311 2016.
- ④⑩ Davis C. A. et al. *Botornot: A System to Evaluate Social Bots*. in Proceedings of the 25th International Conference Companion on World Wide Web. 2016. International World Wide Web Conferences Steering Committee.
- ④⑪ Luceri L. et al. *Evolution of Bot and Human Behavior During Elections*. First Monday vol. 24 no. 9 2019; Shao C. et al. *The Spread of Low – credibility Content by Social Bots*. Nature communications vol. 9 no. 1 2018. p. 4782; Ferrara E. *Disinformation and Social Bot Operations in the Run up to the 2017 French Presidential Election*. First Monday vol. 22 no. 8 2017; Badawy A. E. Ferrara and K. Lerman. *Analyzing the Digital Traces of Political Manipulation: The 2016 Russian Interference Twitter Campaign*. in 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining(ASONAM). 2018. IEEE.
- ④⑫ Shao C. et al. *The Spread of Low – credibility Content by Social Bots*. Nature communications vol. 9 issue. 1 2018. p. 4788; Ferrara E. *Disinformation and Social bot Operations in the Run up to the 2017 French Presidential Election*. First Monday vol. 22 no. 8 2017; Badawy A. E. Ferrara and K. Lerman. *Analyzing the Digital Traces of Political Manipulation: the 2016 Russian Interference Twitter Campaign*. in 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining(ASONAM). 2018. IEEE.
- ④⑬ Blei D. M. A. Y. Ng and M. I. Jordan. *Latent Dirichlet Allocation*. Journal of machine Learning research vol. 3 2003. p. 993.
- ④⑭ Roberts M. E. et al. *The Structural Topic Model and Applied Social Science*. in Advances in neural information processing systems workshop on topic models: computation application and evaluation. 2013. Harrahs and Harveys Lake Tahoe.
- ④⑮ Wesslen R. *Computer – Assisted Text Analysis for Social Science: Topic Models and Beyond*. arXiv: Computation and Language. 2018.
- ④⑯⑰ Roberts M. E. B. M. Stewart and D. Tingley et al. *R Package for Structural Topic Models*. Journal of Statistical Software vol. 91 no. 2 2019. p. 3 p. 12.
- ④⑰ Li M. and Z. Luo *The ‘Bad Women Drivers’ Myth: the Overrepresentation of Female Drivers and Gender Bias in China’s Media*. Information Communication & Society vol. 23 no. 5 2020. p. 782.
- ④⑱ Pak C. . *News Organizations’ Selective Link Sharing as Gatekeeping: A Structural Topic Model Approach*. Computational Communication Research, vol. 1 no. 1 2019. p. 57.
- ⑤⑩ Mikolov T. et al. *Efficient Estimation of Word Representations in Vector Space*. in international conference on learning representations. 2013.
- ⑤⑫ Sutton J. *Health Communication Trolls and Bots Versus Public Health Agencies’ Trusted Voices*. American Journal of Public Health vol. 108 no. 10 2018. p. 1281; Allem J. and E. Ferrara. *Could Social Bots Pose a Threat to Public Health*. American Journal of Public Health vol. 108 no. 8 2018. p. 1005.
- ⑤⑬ Shao C. et al. *The Spread of Low – credibility Content by Social Bots*. Nature communications vol. 9 no. 1 2018. p. 4789.
- ⑤⑭ Statista *Leading Countries Based on Number of Twitter Users as of April 2020*. <https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/>. Accessed on June 26th.
- ⑤⑮ Bolsover G. and P. Howard. *Chinese Computational Propaganda: Automation Algorithms and the Manipulation of Information About Chinese Politics on Twitter and Weibo*. Information Communication & Society vol. 22 no. 14 2019. p. 2068; 师文、陈昌凤《分布与互动模式: 社交机器人操纵 Twitter 上的中国议题研究》,《国际新闻界》2020 年第 5 期,第 70 页; 师文、陈昌凤《社交机器人在新闻扩散中的角色和行为模式研究——基于〈纽约时报〉“修例”风波报道在 Twitter 上扩散的分析》,《新闻与传播研究》2020 年第 5 期,第 14 页。

(作者师文系清华大学地球系统科学系博士研究生; 陈昌凤系清华大学新闻传播学院常务副院长、教授、博士生导师)

【责任编辑: 张毓强】