

真假舆情，网络水军的社交媒体操纵检视

人民网舆情数据中心 朱雯琪 夏玘

随着互联网技术的进步和网络环境的变化，网络水军持续更新换代，愈发成为表现出与真人上网行为几近无异的智能化特征，亦有受资本付费雇佣操纵固有活跃帐号对指定内容进行转发评论的真人水军出现在公众视野。从多起网络舆情事件的复盘中发现，网络水军愈发成为网络舆情发酵，特别是反转舆情事件中不容忽视的一股力量。舆论呼吁严格审视网络水军在消费公众情绪、炮制虚假舆情乃至破坏互联网生态方面潜藏的社会危害，围绕该乱象的全方位检视和对症下药显得迫在眉睫。本文尝试结合国内学者长期以来对网络水军的观察和研究，从水军的组织形式、迭代和识别手段来一探网络水军的真面目。

一、网络水军组织形式更新 规模化特征显著

网络水军的行为受他人操纵，且具有区别于普通平台用户的特殊不正当目的。有学者对网络水军在国内的出现时间及不同时期水军内容发布技术的更新进行了梳理。网络水军由商业利益驱动，为达到影响网络民意、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¹，一般包括自然人和社交机器人两种类别，通过发布灌水信息并对相关话题进行炒作、引导来产生作用。²当下的网络水军呈现出职业化、规模化的特征。有研究表明，早期的网络“水军”大都单打独斗，没有形成特定的组织。2007年后开始出现专业化的网络营销公司，水军便受雇于网络公关或营销公司。而当下网络黑公关通过雇佣大量水军制造虚假评论，操纵网络舆论，是资本影响网络舆论最直接的一种方式。³流量经济的发展、泛娱乐化的兴起、议程设置的转移都是异常主体产生及存续的重要支撑。从网络职业传播者的营销需求、网络媒体自由度高且匿名性强的发展态势、网络受众多元复杂的心理特征来看，网络推手的出现满足资本推广需要、符合媒介传播特征、顺应受众心理趋向，具有必然性。⁴另外，有学者研究国外社交媒体平台受水军信息污染的

¹ 谢忠红,张颖,张琳.基于逻辑回归算法的微博水军识别[J].微型机与应用,2017,36(16):67-69+72.

² 邹仪,兰月新,张鹏,何巍,夏一雪.网络话题异常主体作用机理与仿真研究[J/OL].情报杂志:1-8[2021-07-08].<http://kns.cnki.net/kcms/detail/61.1167.G3.20210608.1324.028.html>.

³ 刘立荣.网络舆论场中资本逻辑的表现与特征[J].新闻知识,2021(04):28-31.

⁴ 彭媛,张曼玲.从传播学角度解读网络推手出现的必然性[J].新闻界,2008(03):26-28.

现状和受操纵的内容营销产业链，提出用户注意力作为稀缺资源的背景下，平台为增强用户黏性不会放任偏见性内容或虚假信息的病毒式传播，相反会为维护信息传播生态限制虚假新闻恣意泛滥的观点。她认为宣传者会因此故意生产迎合受众需求和心理的内容并利用漫游器和水军进行传播，以维护其在社交媒体上的活跃度和能见度，而这一内容又会被平台算法源源不断地推送给用户，操纵用户的实际行为。⁵

二、网络水军发布技术迭代，迷惑性加强

学者杨慧芸认为，早期的网络水军可以被当作是“发帖机”，它伴随着论坛而红火，被用于代替人工进行内容的规模推送。发帖机的后台操作者注册多个帐号进行运作，通过设置标题和内容、添加链接等操作，依托大量帐号在多个网络平台实现短时间、大规模的内容传播。2015年前后，随着各个网络平台风控系统的升级，发帖机基本不能使用，取而代之的是“群控”技术。群控系统是指采用一台电脑PC机，通过USB线控制多部手机，实现自动化脚本运作，通过不同的手机设备码骗过网络平台的风控系统，实现信息在网络上的海量投放。在自动化脚本运行的过程中，手机群控系统可以集中大数据的优势，编制出一定量的内容，有针对性地投放到社交媒体平台上。此外，另一种“云控”的方式也投入使用。云控是一种“群控”加上“云服务器”的技术，是指“通过无线连接，电脑、平板、手机通过后台发送指令到云端，云端的指令再发到手机群，继而执行任务。理论上，一台电脑可以控制上千台手机。”⁶可以看出，从早期原始的人工操纵多个帐号，到如今云端指令控制虚拟用户发布内容，网络水军为绕过社交网络平台的风险控制系统而不断进行技术更新，发布内容和整体用户行为逐渐智能化，具有愈发强烈的迷惑性。

三、通过技术手段挤掉网络水军的“水分”

已有较多研究成功从技术层面聚焦如何判别网络用户是否为水军：部分学者通过以贝叶斯模型为基础，研究水军判定方法。张艳梅选取特征属性，采用朴素贝叶斯模型和遗传算法相结合的方法，得出在朴素贝叶斯模型基础上增加阈值矩阵优化分类器，得到的概率矩阵较准确地识别水军用户的结论。⁷程晓涛研究发

⁵ 方师师. 社交媒体操纵的混合宣传模式研究[J]. 现代传播(中国传媒大学学报), 2018, 40(10):143-150.

⁶ 杨慧芸. 隐形操纵与数据污染:社交媒体中的机器人水军[J]. 新闻知识, 2020(01):3-10.

⁷ 张艳梅, 黄莹莹, 甘世杰, 丁熠, 马志龙. 基于贝叶斯模型的微博网络水军识别算法研究[J]. 通信学

现在朴素贝叶斯、贝叶斯网络和 C4.5 决策树三种算法下融合关系图特征、内容特征和行为特征的方法识别率较为准确。⁸另有学者采用分析用户社交关系的方式来判断其是否为水军。李大鹏等提出情感社会网络的概念，即传统社会网络统计成员交互频繁程度和情感倾向程度计算。该研究从微博的评论数据构建情感社会网络，并采用高斯混合模型发现情感社会网络中的团。⁹杨海梅等分别对电商领域、社交领域、微博领域检测新型网络水军所采用的识别方法进行分析，认为目前的网络水军识别研究都是基于水军行为特征、内容特征、关系特征与环境特征等，采用机器学习中的分类算法和深度学习模型等方法进行特征提取来发现水军，但仍应尝试更多领域的特征定义，以提高准确性和实用性。¹⁰

一段时间以来，网络水军通过有偿屏蔽删帖、虚假营销推广、舆情煽动炒作等方式谋取不当利益，所造成的不良影响日渐“浮出水面”，多方外部监管力量的行动也在逐步加码。今年以来，打击网络水军、流量造假和黑公关，成为国家网信办 2021 年“清朗”专项行动重点任务之一¹¹；微博安全中心官方微博日前发布《打击网络水军 微博在行动》的通告，表示将采取与公安机关建立水军线索上报机制等举措，把反水军工作向治本推进。但我们也必须认识到，打击如此隐秘、规模庞大且技术手段不断升级的“网络水军”及其背后的产业链，无法一蹴而就。就舆情工作而言，网络水军的参与损害网络舆论信息的真实性、危害网络舆论传播的公信力、误导舆情的监测判断和决策，亟需认知、技术、实践和理论多个层面的再探讨。¹²

报, 2017, 38(01):44-53.

⁸ 程晓涛, 刘彩霞, 刘树新. 基于关系图特征的微博水军发现方法[J]. 自动化学报, 2015, 41(09):1533-1541.

⁹ 李大鹏, 余宗健, 肖朝文, 邱江涛. 基于群体行为的社交媒体中水军分析[J]. 计算机工程与应用, 2018, 54(05):144-149.

¹⁰ 杨海梅, 王恒. 国内网络水军识别研究[J]. 网络安全技术与应用, 2021(02):152-154.

¹¹ 明乐齐. “网络水军”的危害界定及打防对策研究[A]. 公安部网络安全保卫局. 2019 互联网安全与治理论坛论文集[C]. 公安部网络安全保卫局:《信息网络安全》北京编辑部, 2019:3.

¹² 赵玉岗. 网络推手操纵参与式新闻的治理策略[J]. 青年记者, 2018(30):36-37.