

# 算法推荐新闻：技术困境与范式变革

贾 军

**[提要]** 本文结合业界实践案例,通过对当前移动新闻分发领域中个性化推荐算法的技术原理、应用现状、发展前景以及对新闻业带来的挑战和影响等方面进行多角度考察,认为个性化推荐算法技术本身会带来“信息茧房”等技术困境,其本身是媒体智能化发展在当前阶段的一种形态,符合信息传播趋势,也将改变传统新闻把关的范式和准则。传统新闻编辑应以人机互动、人机协同进化的方式将新闻专业主义延续至算法把关中。

**[关键词]** 算法推荐; 新闻分发; 智能媒体; 范式变革

中图分类号: G206.2

文献标识码: A

文章编号: 1004—3926(2019)05—0152—05

**基金项目:** 国家社科基金项目“三网融合”背景下中国报业组织结构转型研究(10BXW013)阶段性成果。

**作者简介:** 贾军,河南财经政法大学文化传播学院讲师,博士,研究方向:媒介发展。河南 郑州 475001

个性化推荐算法(PREA, Personalized Recommendation Algorithms, 以下简称算法)是当前人们在信息爆发增长趋势下获取信息的一个越来越重要的方式。随着近两年算法在移动信息分发中的广泛应用,其对新闻传播行业的新闻生产和业态变迁产生的影响已引起业界和学界的广泛关注和深度思考。与传感器新闻、虚拟现实等人工智能技术驱动新闻生产方式的创新相类似,算法推荐新闻对传统新闻行业在编辑室内已完成全部新闻把关的生产方式带来了颠覆性冲击。同时,这种方式对传统新闻行业形成的行业共识和职业规范也形成了较大的冲击。算法推荐新闻广泛应用于新闻传播业,媒体将新闻把关交由机器和计算机程序,对既有新闻范式会产生什么样的影响?越来越多的传媒科技公司布局算法推荐新闻,这对新闻行业将会产生什么样的影响?业界或者市场在获得这种技术带来的益处之外,应当对这种技术保持什么样的警惕和反思?这是在当前移动资讯分发领域内算法推荐新闻热度不减的情况下,亟需思考的问题。

本文主要结合近两年采用个性化推荐算法分发新闻其若干应用的分析,阐述当前阶段算法分发新闻的基本原理和特征、发展现状和趋势,重点观察算法对传统新闻业生产流程和范式的冲击,以及可能面临的局限和困境,并尝试对算法、人工

在新闻生产变革中功能和角色的转换进行展望。

## 一、什么是算法推荐新闻?

算法推荐新闻,在业内一般代指通过计算机程序设定,实现内容到用户的精准传播和个性化推荐,其主要特点在于根据用户的个人偏好数据,主动向用户推荐其可能感兴趣的信息。理想的算法中,计算机信息传播系统可以全方位、立体式满足人的信息需求。就像凯文·凯利在《必然》一书中描绘的,人可以像插电一样随时接入智能流,一走进办公室,办公室内桌椅、电脑会识别后调整好姿势,电脑已打开常用软件、网站等。计算机像人类一样阅读新闻信息,基于无处不在的连接和定位服务,记录、分析甚至通过脑电波、眼球转动等识别人的信息需求,根据用户的兴趣、场景多维画像,主动、即时、准确为用户提供信息等。

个性化推荐算法最早是在1995年的美国人工智能协会上由科学家Robert Armstrong等人首次提出<sup>[1]</sup>。至1992年,以协同过滤技术的出现和应用为标志,个性化信息推荐系统成为一门独立的研究,成为解决“信息过载”的重要工具之一<sup>[2]</sup>。它建立用户和被推荐对象之间的连接关系,在海量数据中挖掘用户潜在的兴趣偏好,为用户呈现个性化的推荐列表,能够起到一种过滤信息的作用。该算法较早用于商业领域,例如亚马逊电商网站在顾客的个性化营销中运用算法挖掘用户潜

在感兴趣的商品,获得了很好的收益。但是由于计算机的文本识别技术以及智能设备、移动互联网等普及程度所限,算法较少运用于新闻分发领域。较早的是 2004 年谷歌通过聚类分析用户的浏览记录,发现用户兴趣模式,据此为用户提供信息过滤和信息推荐服务。

如果说在 2015 年以前,算法分发主要应用于科技互联网公司,在 2015 年之后,算法对新闻生产制作的影响则开始延伸到传统媒体中。在国外,脸书 New Feed 的历次算法调整、改进都引起了其合作传统新闻媒体的流量波动,对其依赖性越强的媒体遭受的流量打击就越大,最高流量下滑速度达 42.7%,而幅度上升最大者为 27.3%。<sup>[3]</sup>就国内而言,截至 2016 年 5 月,入驻今日头条头条号的自媒体账号数量达到 8.5 万,另有 3.5 万的其他类别头条号(含传统媒体、政府机构、公益机构和科研机构)。<sup>[4]</sup>这显示着算法平台在新闻传播渠道方面日益强大的控制能力,以及向上游新闻内容生产领域的布局。本文以当前国内今日头条等移动资讯分发应用的实践,试图探讨算法推荐新闻为新闻业发展产生的影响。

二、2016: 算法推荐新闻的突破发展年

2016 年是算法推荐新闻取得突破性进展的一年。这并不是指这一年算法取得技术上的突破,而是当前市场上作为主要新闻传播渠道的新闻客户端——每天使用移动终端时间在 1 小时以上的用户比例为 81.5%,3 小时的为 46.6%,其中,将新闻类网站、客户端作为阅读新闻首选的用户达到 63%,以社交应用首选的比例是达到 49.4%<sup>[5]</sup>——在这一年里增加技术算法投入,改进原有新闻分发方式,或直接投资新型技术算法公司。例如,搜狐等增加了推荐频道或个性化头条的显著度;国内互联网巨头腾讯、百度将移动资讯分发视为战略级移动互联网流量入口,并加大技术算法投入,如下表所示<sup>[6]</sup>。截至 2016 年 5 月,腾讯新闻技术算法已超过人工推送,其相对市场渗透率达到 59.6%,今日头条和搜狐新闻市场渗透率均超过 30%,网易新闻为 26%,包含技术算法类的应用已占据该领域主流。根据第三方观测机构易观《中国移动资讯信息分发市场专题研究报告 2016》显示,在 2016 年后以算法推荐的资讯信息将在和人工推荐的比例持平后首次超过后者。<sup>[7]</sup>

国内主要新闻客户端引入个性化推荐算法

|            | 产品       | 措施  |
|------------|----------|---|
| 传统门户网站 APP | 新浪新闻 APP | 2015 年发布新版本,推荐频道位置靠前                        |
|            | 网易新闻 APP | 2015 网易新闻客户端升级,推出个性化头条                      |
|            | 搜狐       | 2014 年年底,客户端 5.0 版本上线,主打智能混合推荐模式            |
| 社交平台       | 微信       | 2016 年 6 月,上线“朋友圈热文”,通过用户图像的大多数数据分析进行内容精准推荐 |
|            | 微博       | 上线“头条文章”,给予全站最高信息流权重                        |
| 电商         | 京东       | 2015 年成立实验室,主动机器学习算法                        |
|            | 淘宝       | 借助精确算法打造“淘宝头条”                              |
| 浏览器        | UC 浏览器   | 2015 年上线新版本,资讯推送采用人工推荐和机器推荐制相结合             |
|            | 360 浏览器  | 2016 年最初版本强调“智能个性化推荐”                       |
| 视频         | 爱奇艺      | 2015 年推出“Vedio out”技术,运用算法识别视频内容            |
|            | 优酷       | 2014 年首推个性版优酷,算法推荐是亮点                       |
| 其他         | 百度“百家号”  | 2016 年 6 月上线采用机器+人工方式进行审核和发布                |

今日头条是国内外在个性化移动资讯分发领域起步较早的移动端。算法一般指为解决问题而设计一整套由计算机语言表示的多个步骤构成的执行序列,基于移动端的个性化新闻推荐算法步骤非常独特,即使目的和原理相同,具体构成也不尽相同。以今日头条为例,本文根据其负责人的公开演讲资料,大概将其算法归为以下几个步骤:

一是机器读入海量新闻并实现文本的结构化

和标签化。机器不关心新闻具体内容是什么,只是将文本分解为机器可以识别以及计算异同性的字符串和字符节。

二是对文本进行分类聚合,算法测量文本相似性。聚合和热点计算是机器对文本内容(包括主题词、兴趣标签)、阅读数据(包括点赞、互动、转发、阅读时长等)等两类标签标记和计算,以机器学习实现海量数据标记,即计算机从预先标注好

类别的数据样本之间的关系中学习预测,将已标注节点的标签按照相似度传递给其他未标注的文本或网页的节点<sup>[8]</sup>,并进行打分排序。

三是基于用户注册初始数据、历史阅读数据和场景数据对用户构建多维画像,建立用户兴趣模型,并根据用户的阅读反馈,不断更新模型。

四是根据新闻文本标签和用户画像标签的匹配程度,完成推荐,在实际中多采取算法组合,种类以基于内容相似性的算法和基于协同过滤的算法等为主。在协同过滤算法中,假设具有相似兴趣偏好的用户具有相似的信息需求<sup>[9]</sup>,计算机根据用户数据计算相似用户作为邻近用户,从邻近用户群所点击阅读的内容中选择该用户感兴趣的文章,并排序呈现。

五是过滤和消重,并通过少量的人工编辑,以弥补机器忽略的少量具有新闻价值的重要新闻等。

今日头条作为一款移动推荐引擎应用,凭借其技术优势成为该领域内的领跑者。从2012年到2016年年均新增用户1亿多,月增用户1000多万,每个月客户增长量几乎相当于其他类似机构一年甚至几年的增量,显示了移动资讯个性化分发领域巨大的市场潜力。最近,今日头条宣称其内部60%的员工都是产品技术人员,只有极少数的编辑人员,内容分发几乎完全通过算法技术实现。

而搜狐新闻、凤凰新闻等这种从传统门户网站发展起来的移动新闻客户端也在2015年起将算法引入新闻编辑。搜狐在内容分发上采取“编辑流+推荐流”的模式,其中“推荐流”即为基于下拉推荐以后产生的内容,来自机器和数据计算,在组织构成上调整内容管理部门架构,加强新闻编辑流内容选择的水平功力和价值观,对新闻编辑的职责进行重新定位<sup>[10]</sup>。凤凰新闻推出的一点资讯(凤凰网旗下),在技术驱动高度数据化的同时,注重人机配合为用户提供价值更高的阅读体验。百度、UC等浏览器,在移动版本中上线类似“无分类信息流”板块、“猜你喜欢”之类的个性化推荐栏目,这类应用在浏览器用户积累数据方面具备较多优势。腾讯的天天快报对腾讯用户的画像更为准确。

算法个性化推荐新闻,不仅应用在移动资讯分发中,在类似谷歌的新闻搜索引擎和类似脸书的以社交关系分发新闻中也有较多应用。谷歌新

闻作为一个新闻内容聚合产品,其中个性化推荐这一块,根据用户的搜索和点击行为完成用户聚类,预测用户新闻事件兴趣,协同过滤生成新闻推荐列表,列表内各新闻网页排序也以一系列算法而定,其公布的一项专利显示算法公式包括新闻网站定期产生的新闻数量、新闻来源的平均长度、新闻来源新闻报道的重要性、突发新闻得分、使用模式、舆论意见、流通量、写作风格等。<sup>[11]</sup>脸书公司新闻推送的核心业务 News Feed 中决定用户看到哪些动态信息的算法要素包括用户关系、用户明确或含蓄的兴趣偏好、页面关系、时间衰变等。2016年4月美国皮尤研究中心发布的《2016年美国媒体发展状况报告》报告显示,64%的美国成年人从社交媒体获取新闻,谷歌、脸书等5家科技巨头公司在数字广告市场中占比达65%<sup>[12]</sup>。基于社交关系和搜索引擎的新闻算法分发已经成为新闻业务的重要环节。

影响算法推荐新闻发展的关键因素还包括人工智能技术在图文识别、机器学习、专家系统等领域的突破。谷歌、脸书、微软以及国内的百度等相继成立人工智能实验室。与算法推荐新闻密切相关的人工智能技术在2016年得到了比较大的突破,例如计算机识别图像中特定物体的精确度不断提高,错误率下降到5%以下,与人类持平;机器深度学习的相关数据量、软件和技术大幅增长,这些软件可以教会计算机自己思考,快速学习新事物,及时解决各种普遍问题。

在算法推荐新闻实现所需要的两个基本条件中,除了人工智能技术能够准确识别图文信息、个人偏好这一条件之外,用户获取计算机分发信息的速度应足够快捷,成本应足够低。计算机、智能设备、移动互联网等设施的快速更迭,大数据、云计算等支撑技术的演进,互联网向泛在化、广域化、智能化的发展,人工智能生态圈的逐步发展,为人工智能从实验室的技术研究应用于市场、应用于新闻推荐提供了可能。

### 三、技术困境与反思

新闻传播学界对由机器、算法把关新闻这一发展趋势持谨慎的态度。毫无疑问,算法推荐新闻的出现和发展过程正是智能手机从工具应用向内容应用转变这一颠覆性变化发生的过程,符合用户资讯阅读向移动端转化的发展趋势。但是算法推荐新闻尚处于发展初期,其能够在提高信息分发效率的同时,是否会对传统新闻行业的生产

流程和规范构成颠覆性冲击?这种冲击会为我们接受阅读信息产生什么样显著和潜在的效果?甚至连公众本身对机器决定自己看什么新闻的态度也相当矛盾。2015年爆发的脸书“偏见门”事件中,脸书在其新闻应用 News Feed 新增加的“趋势话题”(Trending Topics)模块中,组建人工编辑团队把关新闻推送结果,公众对科技公司内存在传统媒体编辑把关这一环节表示广泛的关注和质疑,认为人工编辑介入破坏了脸书这样超大资讯平台应持的客观立场<sup>[13]</sup>。事实上,新闻把关无处不在,并不存在一个绝对客观、中立的新闻信息平台。计算机和机器看似按照既定程序,过程和结果不受人干预。事实上,算法是人们接触信息平台后获取信息的背后操作者,算法的权重要素如何构成?这些要素由谁决定?这些决定者是否拥有该项权力?其背后的标准和价值观是否符合新闻规范和价值标准?

从用户的角度来说,算法推荐新闻主要采用基于内容相似性的匹配和协同过滤,其假设用户当前的信息需求和用户长期的个人兴趣相关,具有相似性特征的群体内成员具有相似的信息需求。用户基于个人兴趣选择性接受新闻信息,基于文本相似性和用户相似性推荐的新闻算法,长期为用户提供用户关注和愉悦的信息,像“过滤气泡”<sup>[14]</sup>一样,将多样化信息、与用户兴趣、意见相背离的信息过滤排除在用户的信息世界内。随着时间的推移,是否会像陷入茧房的蚕茧,即“信息茧房”效应<sup>[15]</sup>?

同时,计算机程序在推荐应用的背后,持续地记录用户的阅读数据,挖掘用户与用户之间隐藏的相似性和相关性。根据协同过滤算法,具有相似性特征的用户通过点击,共享具有相似性和一致性的信息和意见,越多人点击的新闻在被推荐列表中的排名越靠前,个人兴趣经由平台的偏向作用,被迫追随大众的兴趣,长期通过该平台阅读新闻资讯的用户是否存在兴趣和视野固化和狭窄的可能?“互联网可以把头脑相似的甚至是微弱的偏好发展成同质的小团体,其内部交往大过群外交往”,一项基于十多个国家的实证研究显示,“极化群体”现象,即“团体成员一开始即有某种偏见,在协商后,人们朝偏向的方向继续移动,最后形成极端的观点”,“在团体成员以匿名身份相会的情况下更为严重”<sup>[16]</sup>,算法作为一只看不见的手,是否会导致用户群体极化现象的出现?

计算机算法本身存在的局限,以及可能导致的技术困境,将会是新闻业中的一个核心问题,业界对此的态度也相当谨慎。例如,苹果公司的新闻服务 Apple News 业务在 2015 年发出新闻编辑的招聘广告,认为算法在突发新闻、优质新闻内容和小众新闻方面与人工编辑相差较大,引入人工编辑能够提高用户体验,并避免漏掉“原创、生动的新闻事件”。<sup>[17]</sup> Twitter 在新闻推荐上舍弃了原有的标签算法,推出了新闻聚合功能插件 Moments,以人工编辑筛选、整合信息,形成当天热点事件的整合推送<sup>[18]</sup>,在新闻价值判断等方面可以有效提高用户体验。

#### 四、新闻把关的范式变革与突破

纵观近两年国内外媒体机构在个性化算法推荐所取得的发展,包括今日头条在内的一批应用受到不同程度的诟病。但是它代表着一种无法忽视的发展趋势。伴随着算法个性化推荐新闻这一趋势的还有用户智能化、核心内容生产智能化、平台智能化等趋势,这些现象推动着传媒向智能化发展。根据摩尔定律预测计算,到 2020 年,全世界传感器、制动器等占据全球物联网设备的主导地位,人均对应 9.4 台物联网设备。通过系列传媒资源的发展,将达到传播介质的界限完全消失,达到视频、音频、图像、文字等不同产品以数据信息流的形态在可穿戴设备、云终端、网站及相关社交媒介的自由匹配与分享,以提供个性化、场景化、定制化的内容、社交、关系服务。最终,社会传播系统向在线社会信息传播系统发展,生成新型的智能传媒系统。<sup>[19]</sup> 算法个性化推荐只是未来智能媒体传播系统在当前的一个阶段性发展状态。

2006 年、2007 年新京报分别对 TOM 网站、浙江在线提起著作权侵权诉讼,认为后者“实质上替代了报纸网站作为网络服务者向公众提供新闻作品”<sup>[20]</sup>。2014 年搜狐网站对今日头条提起著作权侵权诉讼,理由同样是认为后者的深度链接服务替代了搜狐客户端向公众提供新闻。仅 7 年之隔,门户网站由被口诛笔伐的侵权者变为“以战求和”的被侵权者。这背后折射了信息传播渠道控制权的变迁。传统媒体、网络媒体等依赖于计算机算法平台这一赛博空间内的虚拟渠道和介质实现内容的售卖。在整个新闻传播业务流程内,算法推荐新闻应用持有“传播”环节的同时,通过版权交易和扶持内容生产等向“内容生产”领域发展,而传统媒体则从新闻生产传播被迫向“新闻报

道”领域退守。为了应对危机,传统媒体将工作重心向“新闻传播”环节转移,媒体机构对纯技术或者技术背景的数据工程师、可视化编辑等岗位的需求量增加,对非技术类的传统采编岗位也提出了融合性人才的要求。<sup>[21]</sup>而据招聘网站统计,资讯类客户端对数据人才的需求量以及支付薪酬均是人工编辑的数倍。

以机器、计算机算法为代表的机器“劳动力”的发展在一定程度上,将人工编辑从机械重复性的工作中解救出来,但确实增加了传统编辑的职业危机感,传统编辑的工作内涵和要求会发生新的变化,主要包括三个层面的变化:一是在新闻传播教育和职业培训中,懂新闻的算法工程师和懂算法的新闻编辑将在行业内取得优势地位,技术素养、技能与团队协作精神将成为传统人工编辑的核心技能之一;二是计算机算法通过机器学习、内容识别等技术的应用提高对新闻内容和人类主观世界的了解和判断,传统人工编辑需要在理念与实践上与机器、计算机形成对话和互动,才有可能将新闻专业主义规范在计算机算法语言和系统中;三是计算机和算法技术对当前编辑工作形成了挑战,但也将是一种自我定位升级的驱动力,积极面对挑战,媒体人在技术压力前完成职业提升,才能与机器算法形成技术协同进化。

#### 参考文献:

- [1]李洁. 基于 Android 智能终端的个性化新闻推荐系统[J]. 微处理机, 2015(3).
- [2]吴登能. 面向移动互联网的个性化新闻推荐算法研究[D]. 杭州: 杭州师范大学, 2013.
- [3]脸书阅读平台导流能力再度下滑[EB/OL]. <http://news.qq.com/original/quanmeipai/lianshuliuliang.html>.
- [4]今日头条头条号. 2016年7月头条号大数据月度报告[EB/OL]. <http://www.toutiao.com/i6319648658346213889/>.

- [5]智媒来临: 2016 中国新媒体趋势报告[EB/OL]. <http://tech.qq.com/a/20161115/003171.htm#p=18>.
- [6]严九元. 算法时代到来, 工程师能取代人工编辑吗?[EB/OL]. <http://www.tmtpost.com/2440711.html>.
- [7]易观. 中国移动资讯信息分发市场专题研究报告 2016[EB/OL]. <https://zhuanlan.zhihu.com/p/22034532>.
- [8]张俊丽. 标签传播算法理论及其应用研究综述[J]. 计算机应用研究, 2013(1).
- [9]刘青文. 基于协同过滤的推荐算法研究[EB/OL]. 合肥: 中国科学技术大学, 2013.
- [10]搜狐发声明否认裁撤内容部不再设编辑岗位传闻[EB/OL]. <http://tech.163.com/16/1214/14/C88KQ5LD00097U7R.html>.
- [11]36 氪. Google 新闻排名算法透视[EB/OL]. <http://industry.caijing.com.cn/2013-02-23/112524822.html>.
- [12]方师师. 平台崛起成定论 业务创新带探索——2016 年皮尤美国新闻业报告[J]. 新闻记者, 2016(7).
- [13]杰罗姆. Facebook “新闻偏见门”是怎么回事?[EB/OL]. <http://jerome.baijia.baidu.com/article/452043>.
- [14]话媒糖团队. 社交平台决定你看什么新闻, 算法中立只是个传说[EB/OL]. <http://www.jzwcom.com/jzw/e6/13769.html>.
- [15][美]凯斯·R·桑斯坦. 信息乌托邦: 众人如何生产知识[M]. 毕竞悦译. 北京: 法律出版社, 2008.
- [16]转引胡泳. 新词探讨: 回声室效应[J]. 新闻与传播研究, 2015(6).
- [17]新浪科技. 苹果: 人比算法更懂新闻, 你要不要投个简历?[EB/OL]. <http://tech.sina.com.cn/it/2015-06-16/doc-ifxczqar0947273.shtml>.
- [18]杨斯钧. 算法也有“不靠谱”的时候, 人工编辑再度占领社交媒体[EB/OL]. <http://www.tmtpost.com/1492692.html>.
- [19]吕尚彬. 传媒智能化与智能媒体[J]. 当代传播, 2016(4).
- [20]邵国松. 新闻聚合的版权问题研究[J]. 南京社会科学, 2015(5).
- [21]蔡雯, 翁之颢. 融合转型的传媒业需要什么样的新闻传播人才? ——对近年传媒业人才需求状况的观察与分析[J]. 新闻记者, 2016(12).

收稿日期 2018-10-10 责任编辑 吴定勇