

# 标签价值与劳动转型：推荐新闻的算法责任

慕海昕 新闻学院 2018213494

**摘要：** 算法推荐成为新闻业新的分发模式，在取代人的工作时，算法也应当承担原有的责任。推荐系统是机器学习问题，标签与特征是其中的核心要素，其中蕴含着对新闻价值的判断，应当由总编辑进行把关。算法推荐还可能让写作者屈从于推荐原理与用户兴趣，这需要平台对推荐机制进行改进。同时，算法推荐背后存在大量以内容管理员为代表的劳动者，他们的价值也应当被肯定。

**关键词：** 推荐新闻；算法；机器学习；传播劳动

## 1.引言

推荐算法的应用为传统新闻业的分发环节带来了颠覆性影响。2017年1月，中国互联网络信息中心（CNNIC）发布了《2016年中国互联网新闻市场研究报告》，指出主要的新闻媒体和社交媒体几乎均在算法业务上加大了投入，主打个性化、智能推荐的“算法分发”正在成为互联网新闻的主要分发方式。无独有偶，猎豹全球智库以猎豹移动大数据平台 libra 的数据为参考，发现2016年美国新闻APP市场中以 Flipboard、News Republic 为代表的算法推荐类新闻软件所占的份额正在上升。

在传统新闻业关于“把关后移”“信息茧房”等话题的反思声浪中，新崛起的互联网巨头们也没有那么轻松。2016年，在接受《财经》记者专访时，创始人张一鸣尚在表态今日头条是科技公司而非媒体，不需要总编辑，并抛出“今日头条不模拟人性，也不引导人性，你们文化人给了我们太多深刻的命题”这样的论调；而两年后，当旗下产品“内涵段子”被关停，张一鸣则写出了“一直以来，我们过分强调技术的作用，却没有意识到，技术必须要用社会主义核心价值观来引导”的道歉信。同样地，Facebook 创始人扎克伯格在面临假新闻泛滥的指责时，也以“我原先以为 Facebook 只是技术公司，现在我意识到我们应当承担更多的责任”的表态寻求公众的谅解。

不论是西方媒体标榜的新闻专业主义，还是中国媒体奉行的马克思主义新闻观，毫无疑问，新闻工作从来都是有价值指导的。当人从分发环节的操作者这一角色中逐渐退场，原有的责任并不由此灭失，而是应当转移给算法——占据了这一流程的新员工。

## 2.标签价值：推荐算法核心技术中的价值判断

2018年1月，今日头条秉持“算法透明”的原则，由算法架构师曹欢欢公开了推荐系统使用的算法原理。曹欢欢指出，推荐系统的本质是解决用户、环境和资讯三者匹配的问题，并给出了  $y = F(X_i, X_u, X_c)$  的推荐系统公式，进入公式的便是机器提取出的上述三者的标签。这与学者王茜记录了今日头条对不同用户推送的信息并对其进行内容分析后总结出的“今日头条算法价值观念的四要素”即场景、内容、用户偏好和平台优先级是基本一致的。也就是说，在已经成熟的推荐系统中，输入用户、环境和资讯内容，系统可以对这三类信息分别提取出不同的标签（资讯的标签中可能包含其来源的平台优先级），再通过标签的适配决定为处在特定环境中具有特定属性的用户推荐特定类别的资讯内容，完成输出。显然，对于推荐系统来说，最重要的价值是标签——用户、环境和资讯内容都包含着大量的数据信息，哪些信息应当成为被提取出的“标签”作为决定推荐的指标呢？

曹欢欢给出的推荐系统公式是一个典型的监督学习问题。监督学习（supervised learning）的目的是构造一个函数，可以根据示例输入-输出对（example input-output pairs）将输入映射到输出，而示

例的输入-输出对则意味着监督学习的过程中需要大量带标记的训练数据。以大数据专家周志华在《机器学习》中的“判断好瓜”问题为例，收集每个瓜的色泽、根蒂、敲声等数据输入机器作为每个瓜的特征（feature），由人给出其中一些瓜是“好瓜”还是“坏瓜”的标记（即设定标签，label）作为训练集（training set），机器从中构建由特征到标签的映射模型完成学习，并将这一模型应用于有数据但未标注的瓜，对其做出“好瓜”或是“坏瓜”的预测（prediction），通过这些机器给出的标签的准确性可以判断模型的精确程度。由此可见，已标注数据构成了监督学习的基础，而由人设定的标签正是实现监督学习的核心，也正是“监督”这一概念的来源。

用户的兴趣类别划分是从文本的类别划分中对应得来的，因此，推荐系统的基础是确定文本的标签。确定设置哪些标签、以及何种文本应当被打上这样的标签，事实上是对内容进行分类并制定分类的标准，在实际中相当于确定如何设置报纸版面或电视频道，并决定一则新闻应在何处刊载、一档节目应在何处播出，而这些是传统媒体中编辑、甚至总编辑的工作。新闻编辑专家蔡雯指出，新闻编辑工作不仅关注方法论层面的“如何做”，更关注价值观层面的“为何做”和“为谁做”，思想和价值是编辑理念的根基。而算法标签的价值，自然是这种编辑理念对机器的移植。

另一个问题则在于算法提取标签的过程。如上文所述，推荐系统是通过特征对标签进行计算的，而这些特征数据也并不是完全中立的。大数据专家周涛指出，算法的设计人员在对特征进行选择时，做出的往往是“自以为正确”的选择，而这些选择事实上是受到初始数据的样本与个人经验的偏见共同制约的，这些制约条件为算法的偏见赋予了“第一推动力”，并最终随着数据的积累与算法的运行被逐渐放大。

总而言之，整个算法推荐系统的构建过程中都蕴含着价值判断，这些标签的选择与形成本来是传媒从业者基于对新闻价值的判断而完成的本职工作，现在则被写入代码中由算法承担责任。

### 3.劳动转型：更精确的生产者与风险中的管理员

作为今日头条的站内内容，曹欢欢的这篇《今日头条算法原理》同样被推荐系统打上了标签。在这些标签中，“足球”“中国足球甲级联赛”“皇家马德里足球俱乐部”显得有些格格不入。为什么会有这些标签？在对算法原理进行解释的过程中，曹欢欢以足球为例，介绍了层次化文本分类算法，提及了一些足球话题，而这些文本中的关键词显然被机器识别了，因此这篇文章也就获得了类似的标签，有更高的概率被推荐给足球爱好者阅读——尽管文章内容并不关系到任何一场足球赛。

这似乎是进一步提升算法准确性可以解决的问题。但另一个视角是，使用能够精确到关键字的算法进行推荐，对这一机制有所了解的写作者完全可以预测自己的推荐结果，并为了扩大作品在算法下被推荐的范围，对内容进行有倾向性的调整。今天的新闻教育中算法类课程已经占据一席之地，各大高校纷纷将媒介编程、计算传播作为新时代新闻传播学科建设的一个发力方向，可以判断，在今天与未来的新闻传播从业者心中，理解算法逻辑已经成为了必要技能之一。这当然会帮助写作者在现有的分发机制下获得更好的传播效果，但或许会在根源上无意识中限制内容生产的丰富性。

而算法分发在某种程度上正鼓励内容生产者更“精确”地完成工作。自媒体浪潮中，一批写作者已经把握了用户情绪规律，并以此指导自己的写作。现象级自媒体咪蒙直言，自己的文章写作思维其实是在为用户做产品，做产品的关键是了解用户痛点，并总结出生存、关系、生活方式三级痛点以及与之相应的分析方式。面对《一个出身寒门的状元之死》之类的文章，传统媒体的编辑当然会对事实性信息进行核查，或许还会对其“操纵大众情绪”的行文手法给出批评意见；但在算法分发系统中，这类与用户情绪契合的文章则会因为初始被推荐用户的高点击率进一步获得更大的流量。企鹅号“三表龙门阵”被盗号后更名“娱乐与露露”，后者以《热巴在机场摔坏了粉丝手机，脱口而出八个字，网友：活该你那么火》等一系列标题党娱乐类文章在天天快报等算法推荐平台中三个月获得了七万元收入，与原号在科技领域数年未获得任何收益形成了鲜明的对比。显然，推荐系统大

多数时候是在迎合用户兴趣，而非引领时代价值。那么为推荐系统提供内容的写作者们，完全有理由在此做出最经济理性、最符合市场价值的选择。

在内容生产者的劳动在算法中转型的同时，推荐系统也孕育了内容管理员这一全新工种。由于机器对于低质量内容的判断水平还比较低，今日头条采用人工复审的模式，由管理员对平台上的内容进行审核，确认达到标准后再上线，否则会对假新闻、题文不符等不达标的内容进行删除。事实上这正是科技巨头们采用的一致方法，但内容管理员的劳动却暴露在风险之中。传播政治经济学著名学者文森特·莫斯可研究了商业内容管理员的工作，发现他们大部分是亚洲人，工作内容是长时间盯着屏幕上那些丑陋堕落的东西，却仅能赚取低薪，他们的劳动被商品化，并且彻底降格(degraded)。微软“在线安全团队”的成员 Soto 和 Blauert 因长期审查谋杀儿童的视频、儿童色情和兽交图片等内容患上创伤后应激障碍(post-traumatic stress disorder, PTSD)，两人因此对微软提起了诉讼；Facebook 同样有大量员工从事类似的工作，该领域学者 Sarah Roberts 认为，人们会受到高度影响并因此脱敏(desensitized)，但目前公司提供何种精神保障措施尚不明确。

丹尼尔·贝尔早在 20 世纪末便提出，伴随着一个依赖技术、特别是依赖信息的生产和传播的社会的崛起，科学和技术工作者组成的真正的知识阶层会占据新的统治地位。而莫斯可和他的太太凯瑟琳共同编著的《信息社会的知识劳工》中，二人进一步指出，技术融合打破媒介壁垒、公司并购导致跨媒体集团的崛起，在这种时代背景下，许多劳动者只是从事着与知识相关的工作，这些看上去体面的工作并没有帮助他们占据统治地位，相反，这些“知识劳工”正处于相当不利的境地。

## 4. 算法责任：内容总编辑与劳动者价值

算法新闻应当承担编辑责任。2016 年，《人民日报》发表评论文章《算法盛行更需总编辑》，提出在这个由算法主导的时代，“更需要把关、主导、引领的‘总编辑’，更需要有态度、有理想、有担当的‘看门人’”。今天的今日头条、天天快报等算法新闻平台已经有了总编辑，从工作内容来说，他们应当既懂新闻价值又懂算法逻辑，从前者出发，对推荐算法的特征选择与标签确定给出明确的框架，并在算法运行与改进的过程中始终确保没有偏离应有的价值方向。同时，推荐系统的算法应当透明化、公开化，接受社会的监督。

在对内容生产的影响上，算法新闻也可以发挥更为积极的作用。新媒体专家彭兰指出，如果运用得当，算法甚至可能成为刺破信息茧房的武器，打破原有的信息泡泡。从这个意义上讲，推荐系统不能只以用户兴趣和点击率作为唯一标准，更应该兼顾信息传播的社会效益，适当对重大社会议题、科学知识普及等内容进行推荐，同时对优质平台、优质原创内容生产者予以倾斜性的保护，从经济补贴与算法推荐两方面提供支持。

算法新闻有机会将劳动者从低端工作岗位中解放出来吗？这个答案或许并不乐观。一方面，在对低质量内容的审核上，提高机器学习的准确程度在技术难度上比较困难，对虚假信息、低俗信息等的判断相当程度上依赖于人的常识，而常识，事实上正是今天机器学习研究中最难以让机器掌握的部分；另一方面，即使在算法上有了较大的突破，完成机器学习依然离不开大量的标注数据，而这背后将是无数的众包平台。算法可能加剧的劳动异化与阶级分化是另一个值得讨论的话题，但无论如何，每一家标榜算法推荐的新媒体公司都应当承认人的劳动价值，为他们提供更好的工作保障；社会也应当看到人工智能背后的人。

## 参考文献：

- [1] 中国互联网信息中心（CNNIC）. 2016 年中国互联网新闻市场研究报告[EB/OL].  
<http://www.cac.gov.cn/files/pdf/cnnic/newsmarket.pdf>. 2019 年 4 月 14 日.
- [2] 猎豹全球智库. 美国新闻 app 最新排名：新闻个性化推荐，美国人买单吗[EB/OL].  
<https://www.jiemian.com/article/910015.html>. 2019 年 4 月 14 日.
- [3] Ballon Pieter. Platform types and gatekeeper roles: The case of the mobile communications industry[C]. In Summer Conference on CBS-Copenhagen Business School, Denmark. 2009.
- [4] 宋玮. 张一鸣：今日头条不模拟人性，也不引导人性，你们文化人给了我们太多深刻的命题[EB/OL].  
<https://www.huxiu.com/article/174510/1.html>. 2019 年 4 月 14 日.
- [5] 张一鸣. 致歉和反思[EB/OL]. <http://tech.163.com/18/0411/07/DF31T0P300097U7R.html>. 2019 年 4 月 14 日.
- [6] 央视网. 一图解读 | 扎克伯格道歉史[EB/OL].  
<http://news.cctv.com/2018/04/13/ARTIpM5mkdA6bxkGica5ByUk180413.shtml>. 2019 年 4 月 14 日.
- [7] 曹欢欢. 今日头条算法原理[EB/OL]. <https://www.toutiao.com/a6511211182064402951/>. 2019 年 4 月 14 日.
- [8] 王茜. 打开算法分发的“黑箱”——基于今日头条新闻推送的量化研究[J]. 新闻记者, 2017（09）：7-14.
- [9] Stuart J, Peter N. Artificial Intelligence: A Modern Approach[M]. 3rd Ed, New Jersey: Prentice Hall, 2009. 695-697.
- [10] Mohri M, Rostamizadeh A, Talwalkar A. Foundations of Machine Learning[M]. Massachusetts: The MIT Press, 2012.7.
- [11] 周志华. 机器学习[M]. 北京：清华大学出版社, 2016. 1-4.
- [12] 蔡雯. 马克思主义新闻观对于新闻编辑的指导意义——编写马克思主义理论研究和建设工程教材的一点思考[J]. 当代传播, 2017(5): 10-12.
- [13] 周涛. 数据的偏见[J]. 金融博览, 2017(5): 22-23.
- [14] 咪蒙. 咪蒙是怎样炼成的? [EB/OL]. [http://www.sohu.com/a/210304291\\_464025](http://www.sohu.com/a/210304291_464025). 2019 年 4 月 22 日.
- [15] 人民日报微博. 人民微评：自媒体不能搞成精神传销[EB/OL].  
[https://weibo.com/2803301701/HeDcswEpu?type=comment#\\_rnd1556006587153](https://weibo.com/2803301701/HeDcswEpu?type=comment#_rnd1556006587153). 2019 年 4 月 22 日.
- [16] 三表龙门阵. 河南女孩露露给我上了一堂七万的课[EB/OL].  
<https://mp.weixin.qq.com/s/uKBTm8IWS6zObog-x6Q00g>. 2019 年 4 月 22 日.
- [17] [加拿大]文森特·莫斯科. 数字劳工与下一代互联网[J]. 徐偲骥, 张岩松译. 全球传媒学刊, 2018(4): 127-139.
- [18] Levin Sam. Moderators who had to view child abuse content sue Microsoft, claiming PTSD[EB/OL]. The Guardian, 2017-1-11. <https://www.theguardian.com/technology/2017/jan/11/microsoft-employees-child-abuse-lawsuit-ptsd>. 2019 年 4 月 23 日.
- [19] Solon Olivia. Facebook is hiring moderators. But is the job too gruesome to handle?[EB/OL]. The Guardian, 2017-5-4. <https://www.theguardian.com/technology/2017/may/04/facebook-content-moderators-ptsd-psychological-dangers>. 2019 年 4 月 23 日.
- [20] [美国]丹尼尔·贝尔. 后工业社会的来临——对社会预测的一项探索[M]. 高铨, 王宏周, 魏章玲译. 北京：新华出版社, 1997. 391-397.
- [21] [加拿大]凯瑟琳·麦克切尔, 文森特·莫斯科. 信息社会的知识劳工[C]. 曹晋, 罗真, 林曦译. 上海：上海译文出版社, 2008. 1-8.
- [22] 姚建华. 编者按：传播政治经济学与数字时代的劳动转型[J]. 全球传媒学刊, 2018(4): 92-96.
- [23] 人民日报评论部. 算法盛行更需总编辑[EB/OL]. [www.opinion.people.com.cn/n1/2016/1223/c1003-28970666.html](http://www.opinion.people.com.cn/n1/2016/1223/c1003-28970666.html). 2019 年 4 月 23 日.
- [24] 彭兰. 假象、算法囚徒与权利让渡：数据与算法时代的新风险[J]. 西北师大学报（社会科学版）, 2018(5): 20-29.