

# 符号学视角下对机器新闻写作的批判性解读

## ——以今日头条“Xiaomingbot”为例

陈炳宇

**【摘要】**机器新闻写作是指用人工智能进行的新闻写作，是近几年来传媒行业的又一新兴热点话题，“Xiaomingbot”则是其中的一个典型代表。符号学是一种以符号系统为研究对象，研究社会生活中符号生命的科学。通过在符号学视角下对“Xiaomingbot”的研究发现，新闻写作机器人存在着图文符号机械依存、语句符号独立僵硬和符号意义指代不明等问题。

**【关键词】**符号学；机器新闻写作；“xiaomingbot”；今日头条

DOI:10.13994/j.cnki.stj.2019.02.018

“Xiaomingbot”是今日头条媒体实验室和北京大学计算机所共同开发的写作机器人，主要从事体育方面的新闻写作。符号学是一种以所有符号系统为研究对象，研究社会生活中符号生命的科学。符号学研究符号的组成，探讨各种符号所代表的意涵，来建立一个“各种符号一般的科学：意义的系统”<sup>[1]</sup>。本文以“Xiaomingbot”为例，从符号学的理论角度来探析机器新闻写作的表达特征，并将机器新闻写作作品和传统的人工新闻作品进行对比，对机器新闻写作进行批判性解读。

### 一、“Xiaomingbot”的简介及写作特点

从2009年开始，美国西北大学智能信息实验室就将自己

开发研究的StatsMonkey系统投入了使用，发表了一篇关于美国职业棒球大联盟季后赛的新闻稿件，有关人工智能的写作自此开始被人们所熟知。之后，国内外的各大媒体也都在不断地探索机器人写稿的可能性，《洛杉矶时报》使用Quakebot系统，美联社等媒体使用Wordsmith作为平台进行新闻写作……进入到2015年以后，腾讯财经频道的新闻写作机器人“Dreamwriter”发表了一篇名为《8月CPI涨2%创12个月新高》的财经报道，开启了国内写作机器人进入新闻界的篇章。随后，新华社的“快笔小新”、今日头条的“Xiaomingbot”、南方报业集团的“小南”等采用人工智能技术的写作机器人也开始逐渐登上这个舞台。

写作机器人“Xiaomingbot”是今日头条实验室和北京大学计算机所万小军团团队产学合作的成果。“Xiaomingbot”正式投入使用是在2016年，所撰写的是针对里约奥运会的赛事简讯。在整个里约奥运会期间，“Xiaomingbot”撰写的稿件赢得了一百多万的阅读量，平均阅读率接近整个奥运会期间体育频道的阅读率。

在里约奥运会以后，“Xiaomingbot”将目标投向了欧洲冠军联赛（欧冠）、意大利足球甲级联赛（意甲）和美国职业篮球联赛（NBA）等各类体育赛事，同样取得了不错的成绩。

在写作特点上，“Xiaomingbot”首先是获取网上有关写作赛事的文字和讨论，并将其总结归纳成为一个相对长的赛事描述，是一

种基于 UGC 模式的信息采集手段；其次，“Xiaomingbot”采用先进的机器学习算法，能够学习图文语义匹配模型，在该模型的基础之上，用今日头条大量的图片和文本做语义匹配，挑选合适的图片，实现图文自动关联的发稿模式；此外，“Xiaomingbot”也运用了在机器新闻写作中广泛应用的模板生成技术，采用了依据句法结构的模板文本生成技术和摘要式生成技术，既能生成短资讯，也能够生成比较长的资讯。<sup>[2]</sup>

## 二、符号学视角下“Xiaomingbot”的缺陷

在先进技术的支持下，如果不仔细阅读，“Xiaomingbot”的新闻作品和传统的人工新闻作品并没有显著不同，都能够进行新闻报道，达到提供资讯的目的。但是，仔细分析，“Xiaomingbot”和传统人工的体育资讯新闻仍然有着一些较为显著的区别。

### （一）图片符号和文字符号的机械依存

因技术限制，“Xiaomingbot”并不能识别视频、动图和音频等符号形式，所以它的新闻作品都是由静态的图片符号和文字符号组成。尽管“Xiaomingbot”有着较为先进的图文语义匹配模型和大量可供选择的图片数据，但在图片符号和文字符号的匹配问题上仍然存在问题，给人明显的机械感。

“Xiaomingbot”在进行图文匹配的时候，通常会采用固定的模式。该模式的特征就是“一段一图”。在“Xiaomingbot”进行写作的时候，它首先会收集网络用户以及其他今日头条有权使用的信息，首先生成文字段落。随后，根据文字段落的意义，

在今日头条的图片库里面进行图文符号的匹配，再将图片插入到段落之前，形成“图片+段落”的作品呈现。并且，几乎是每一段话都有图片在前修饰。拿其报道意大利足球甲级联赛的案例《意甲第19轮恩波利0:1国际米兰遗憾失利》<sup>[3]</sup>来说，除最后两段对于两队阵容的介绍没有采用配图之外，其他所有的段落都采用了“一个图片+一个段落”的内容呈现模式，没有变化，显得非常机械化。它在有关其他赛事的报道也是如此，几乎都是采用“一个图片+一个段落”的呈现方式，很少会有变化。

而由人工编辑产生的对同样赛事进行报道的新闻作品，配图则是更加自由。拿腾讯体育作者冷雪写作的同样比赛内容的文章《意甲-国米1-0取连胜稳居三甲 凯塔进球被吹仍破门致胜》<sup>[4]</sup>来说，文章开头用一张典型的描绘比赛画面的图片来配图，随后在描绘比赛精彩部分的时候才去使用图片进行解释说明，并且会使用一些动图，让这个作品包含的符号更加多样，画面更加生动，更有一种生命感。

### （二）语句符号的僵化、独立

在语句的文本生成方面，许多写作机器人仅仅是依据事先写好的模板，将里面空缺的时间、地点、事件等信息补充完整，通过较为简单的填空作业实现新闻信息的生成。而“Xiaomingbot”则在此基础之上，应用了较为先进的文本生成技术，具有对用户发表的文字、讨论进行归纳整合的能力，并且学习多种赛事网络文字直播员的语句应用策略，实现较为高级的语句生成。但是，仍然难以避免语句符号的独立和僵化问题。

从“Xiaomingbot”报道2018年美网的比赛来看，这一段时间标题都是采用了“比赛名称+选手名和比分+不敌对手/获得胜利”的模式。内容上，第一段会采用与标题类似的结构，先是时间，然后是比赛名称，最后是选手和比赛结果。而在第一段的用词上面，如果说标题是某选手获得胜利，那么开头第一段结尾就会是“最终，某选手手感不俗，以几比几的比分获得胜利”；如果说标题是某选手不敌对手，那结尾就会变成“最终，某选手不敌对手，以几比几的比分遗憾止步”。在随后的段落里，“Xiaomingbot”则会对每局比赛过程的数据进行具体分析，而且在多数情况下会使用“比赛开始-随后-紧接着-最终”的逻辑结构，用词也很少会替换，呈现出单一、刻板、僵化的特点。

“Xiaomingbot”将文字符号组合形成的语句结构的确没有问题，但这种句子结构过分清晰和独立了。换句话说，“Xiaomingbot”所写出来的每一句话都可以非常轻松、明确地划分所有的句子成分，而且大部分的语句都采用“主谓宾”“主谓宾+状语”“主谓+状语”的简单句子结构，虽然结构的确很清晰、句子使用很正确，但给人以一种过分的正确感。每个文字符号的过分独立，让文章少了许多变化的美感，给人一种僵硬感。

而人工记者在进行新闻写作的时候，一般来说会采用更为多样和复杂的句式，让文字符号有更多的组合方式，让整篇文章的语句结构更加丰富，更有可读性。

### （三）对符号意义深度理解的困难

“Xiaomingbot”还存在的



▲ “Xiaomingbot” 今日头条账号

一个问题就是对于符号内涵和意义理解不够，会出现使用方面的问题。

比如说，其对于图片符号的意义理解就会出现偏差。在《NBA 今日看点：哈登 40+9+7 火箭力克步行者，武切维奇 21+14 魔术大胜尼克斯》<sup>[5]</sup>一文中，文章的第一张图使用了一张和NBA比赛毫无关联的一张美国士兵举着旗子的图，这张图会出现在东方IC图库中“火箭”一词的词库当中。这就是因为“Xiaomingbot”未能理解段落和图片的含义，造成图片使用方面的错误。

此外，“Xiaomingbot”还可能对文字符号的意义理解出错，其中最典型的就是其在之前里约奥运会报道上曾写的“失败女神朝其抛出了橄榄枝”<sup>[6]</sup>。“橄榄枝”是一个褒义的词语，一般会写“胜利女神朝其抛出了橄榄枝”，但不会将橄榄枝和失败女神联系起来。很显然，“Xiaomingbot”在文字符号的深层意义的理解上依然会出错。不过，针对于此，“Xiaomingbot”现在写作的时候都会直接写“取得胜利”“遗憾落败”这样虽然具有感情色彩，但相对简单、意义不深的词语，避免出错。

当然，人类记者在新闻写作的过程当中也难免会出现符号意义运用错误的问题。但是，人类记者发生此类错误的原因一般是粗心大意，而不是不能进行理解。

以上的三点缺陷是从符号学视角下对于机器新闻写作的批判性解读。这并不意味着机器新闻写作是无用的。相反，写作机器人因为有着发稿速度快、覆盖范围广、不会疲劳等诸多优势，在未来的新闻

写作中必将占有重要的地位。

### 三、对机器新闻写作进一步发展的建议

在符号学的研究视角之下，我们可以清晰地看到机器新闻写作存在着许多的问题。在此背景之下，笔者提出对于机器新闻写作发展的四点建议。

#### （一）符号间匹配关系的多样化发展

机器写作首先应该将自己语料库和数据库中的符号进行多样化匹配。例如，“Xiaomingbot”中“一个图片+一个段落”的匹配方式就显得十分的机械、刻板。即使人工智能难以理解段落的重要性，无法根据内容的重要性来选择是否需要匹配图片。不过，写作机器人的设计者可以尝试将随机化的思想引入到人工智能的写作过程当中，尝试基于大量文本数据的随机试验，并思考这些随机化的思想是否可以对人工智能的符号匹配运用产生积极影响。

另外来说，对于文本符号之间的匹配也应该向多样化的方向发展，改变刻板、僵硬的语句结构，尝试让文章每个符号变得不那么独立，让文字更加具有美感和可读性。

#### （二）语义理解的进一步加强

关于语义理解方面，也是在机器人新闻写作中必须要解决的一个问题。并且，这也是一个技术难题。

一方面，需要加强人工智能对于某些特定情感符号的理解程度，尽量理解词语的感情色彩和其他特殊的使用规则，防止出现词语的错用现象。另一方面，要研究如何让人工智能能够识别图像、声音和视频符号，并能够将这些符号更自如地运用到新闻写作当中。这样一来，不仅可以促使文章的符号多元化和阅读趣味的增加，也可以减少出现像“Xiaomingbot”对NBA比赛的报道中出现的图文匹配错误的情况。

#### （三）增强情感符号的运用

机器毕竟没有人那么充沛的情感，写出来的文字难免会有一些刻板和生硬。但是，为了追求更好的表达效果，即使是没有感情的AI机器人所写作的新闻作品，我们也应该要求它具有相对多的感情色彩，在语义理解的基础之上增强对情感符号的运用，从而增强文章的可读性，也让文章更加富有美感，达至更好的传播效果。

目前来看，“Xiaomingbot”已经会选用一些具有感情色彩的词来帮助文章更好地进行场景呈

现。它会使用一些具有感情色彩或者更加生动的词语,比如“遗憾”“速战速决”“势如破竹”等,这使得它写作的内容更加富有情感性。不过,比起其他体育频道的专业记者来说,这些用词仍然显得刻板和生硬。而且,有些时候也会出现错误,在使用此类带有感情色彩的词语的时候也可能发生类似于之前说的“失败女神”和“橄榄枝”相搭配的错误。人工智能对于情感符号的运用仍然需要进一步的研究和发展。

#### (四) 多元符号系统的创新性构建

新闻作品也是“人”的作品,是人类所创造符号的集合体。在某种意义上来说,人工智能也

是人类创造的一种符号表现形式,比起人类而言,人工智能更像是一个“符号的集合体”。但是,人工智能是一种具有智能的存在,这也就意味着人工智能同样具有创造性。并且,人工智能在某些方面也可以突破人类的局限性。

人工智能可以对人类现在构建的符号系统进行一些创新性的构建。写作机器人的实验者可以尝试更多的可能性,给人工智能一些“独立”的发展空间。并且,可以尝试跳出已有的新闻框架、逻辑和理念,从创新的角度来对机器新闻作品进行分析和解读,尝试让符号系统在人工智能语言下进行

多元化和创新性的发展。

#### 四、结语

现今,机器人新闻写作仍然处在一个快速发展的过程当中,正在对体育、财经领域的资讯类新闻写作造成不小的冲击。虽然不能否认机器人应用于资讯新闻写作,可以将记者和编辑从烦琐的基础工作中解放出来,更好地提升他们的创造力。不过,从符号学的视角下对于机器新闻作品进行研究,依旧能发现机器新闻写作的很多问题。即使不考虑深度报道,仅仅针对资讯类新闻写作方面而言,写作机器人仍然有许多需要改进的地方。

#### 注释:

- [1] 曾一果. 媒介文化理论概论 [M]. 中国人民大学出版社, 2014:93.
- [2] 刁毅刚, 陈旭管. “Xiaomingbot” 背后, 写稿机器人的技术探寻——专访北京大学计算机科学技术研究所万小军博士 [J]. 中国传媒科技, 2016(9): 7-11.
- [3] “意甲第19轮 恩波利 0:1 国际米兰 遗憾失利” [EB/OL]. 今日头条. <https://www.toutiao.com/i6640443736377524739/>.
- [4] “意甲-国米 1-0 取连胜稳居三甲 凯塔进球被吹仍破门致胜” [EB/OL]. 腾讯体育. <https://sports.qq.com/a/20181229/011006.htm>.
- [5] “NBA 今日看点: 哈登 40+9+7 火箭力克步行者, 武切维奇 21+14 魔术大胜尼克斯” [EB/OL]. 今日头条. <https://www.toutiao.com/i6622865527284957703/>.
- [6] “奥运会乒乓球男子团体半决赛 中国队(马龙/张继科/许昕)3:0 小试牛刀轻取韩国队(朱世赫/李相秀/郑永植)” [EB/OL]. 今日头条. <https://www.toutiao.com/a6319219544476877313/>.

#### 参考资料:

- [1] 曾一果. 媒介文化理论概论 [M]. 中国人民大学出版社, 2014:93.
- [2] 何芳明. 写稿机器人对新闻生产的影响及应用前景 [J]. 青年记者, 2018(33):77-79.
- [3] 孙愈中. 机器人写稿将取代记者工作的现实悖论 [J]. 新闻论坛, 2018(4):42-45.
- [4] 郝慧敏. 今日头条的算法生产新闻研究 [J]. 传媒论坛, 2018, 1(8):112-113.
- [5] 刁毅刚, 陈旭管. “Xiaomingbot” 背后, 写稿机器人的技术探寻——专访北京大学计算机科学技术研究所万小军博士 [J]. 中国传媒科技, 2016(9):10-11.
- [6] 刁毅刚, 陈旭管. 产学研合作路径下网红“Xiaomingbot”的前世今生——专访今日头条实验室技术总监李磊博士 [J]. 中国传媒科技, 2016(9):7-9.
- [7] 里约奥运机器人“记者”显身手 [J]. 中国记者, 2016(9):9.
- [8] 王悦, 支庭荣. 机器人写作对未来新闻生产的深远影响——兼评新华社的“快笔小新” [J]. 新闻与写作, 2016(2):12-14.
- [9] Van der Kaa H, Kraemer E. Journalist versus news consumer: The perceived credibility of machine written news[C]// Proceedings of the Computation+ Journalism Conference, Columbia University, New York. 2014, 24: 25.
- [10] Jung J, Song H, Kim Y, et al. Intrusion of software robots into journalism: The public's and journalists' perceptions of news written by algorithms and human journalists[J]. Computers in Human Behavior, 2017, 71: 291-298.

(陈炳宇: 苏州大学传媒学院新闻学专业学生)