1. **国、内外研究现状和发展动态**

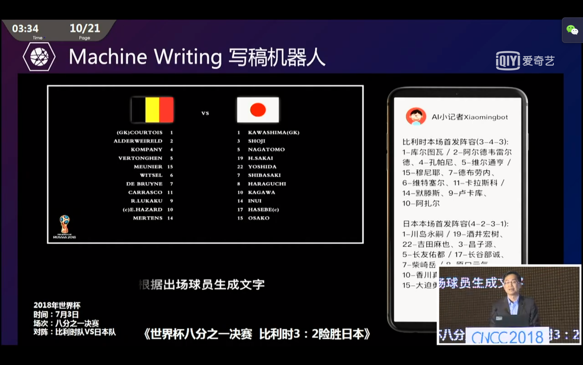
随着人工智能技术的不断发展进步，其应用也逐渐渗透到日常生活中的方方面面。其中，“机器新闻写作”就是其在新闻传播领域的重要应用。2016年里约奥运会上，由北大计算机所万小军团队研发的“张小明”新闻机器人横空出世，13天内，撰写了457篇关于乒乓球、羽毛球、网球等项目的消息简讯和赛事报道，一时间吸引了世人的眼球；而实际上，在“张小明”出现之前，新闻机器人就已经被不少世界上的主流媒体关注并加以使用——



其实，早在2006年，美国商业信息集团汤姆森金融就已经开始运用电脑程序自动生成部分财经新闻，可在上市公司公布业绩短短0.3秒内就发布出一篇盈利报道。这也被看作是“机器新闻写作”的初探，只可惜，在更加注重质量而非速度的报纸时代，它并未引起太大关注。

在那之后，西北大学的科技精英们研制出了新一代智能写作软件Narrative Science，这时的机器写作已经不仅仅局限于对数据的呈现，还已经具有了强大的运算能力，展示出一篇具有“人情味”的新闻报道——此时，已经具有了如今“机器写作”的雏形。

随着社会的飞速发展，传统的报纸、电视等传播媒介日渐衰微，数字化、碎片化的阅读成为人们最主要的信息来源。基于人们这样的需求，新一代“新闻机器人”应运而生，在不断的改进和充实中，逐渐满足了人们对新闻报道的各种需求——如今的“新闻机器人”，已经可以结合最新的自然语言处理、机器学习、大数据分析和视觉图像处理技术，通过语法合成与排序，基于现有的文本模板生成完整的新闻——此处所说的“完整”，不仅仅是指可以生成一篇完整的文本，更可以通过图片检索自己选择合适的配图，甚至模仿人类的语气，使新闻文本读起来更加亲切、抓人眼球。



（新一代的机器新闻成品，已经可以利用成熟的图片、视频抓取与识别技术自动生成语义明确、内容完整连贯的文章）

如今，新一代的机器写作的应用更加广泛。由最开始的集中于数据化、模板化的财经新闻和体育新闻报导为主，到现在，已经可以根据数据的宏观对比，得出变化、趋势等结论并加以呈现；同时，同一则新闻可以生产出多种个性化的版本。而国外由于机器写作研究的起步较早，拥有相对多的技术优势和实战经验，除了能够实现国内的应用外，还可以在房地产、商业等更多领域得到使用，对数据的追踪、复杂信息的消化、关系的挖掘等方面也要比国内更加完备与成熟，甚至还可以写出极具个性化的新闻报道。



（“张小明”可以自己检索图片，还可以使用“笑到最后”“实力不俗”这种极具个性化的词语）

不过，由于相关技术仍然不够完备，当下的机器写作依然有一系列问题亟待解决，比如机器人对信息的理解深度还不够、新闻成品模板化明显，扁平而又千篇一律的文章依旧是绝大多数机器写作产品的常态，缺乏亮点；对信息的提炼和概括能力也还有所不足。这些也是研究人员们接下来将要致力于解决的方向。



（新闻扁平化、千篇一律，是当今机器新闻的最大问题之一）

一言以蔽之，当前，“新闻机器人”是目前机器写作在我们日常生活中最为常见的应用，且已经证明了其高效性和可信赖性，但由于相关技术还不够成熟，其依旧有十分广阔的发展前景，有望在未来成为跨领域的多面手、人类记者编辑的好帮手，而本课题团队也想要以目前已经比较完备的基于模板生成完整新闻这一技术为基础，完成活动类微信公众号推文的实现，提高生产推文的效率，使机器写作这一技术在新媒体领域得到更加普遍的应用。

参考文献：《新闻写作机器人的应用及前景展望——以今日头条新闻机器人张小明（xiaomingbot）为例》

《今日头条的算法生产新闻研究》，郝慧敏，《传媒论坛》，2018.08

《“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》，刁毅刚、陈旭管，《中国传媒科技》，2016.09

相关网页链接：http://media.people.com.cn/n1/2017/0111/c409691-29014245.html

爱奇艺视频：《字节跳动副总裁人工智能实验室主任马维英：AI赋能内容创作和交流》

**（七）已有基础**

* 1. **与本项目有关的研究积累和已取得的成绩**

1. 本课题团队已研究今日头条、新华社万小军等机器新闻写稿前例，对基本流程有一定掌握；同时，在新媒体传播形式、内容及影响中有相关的研究经验，例如通过对微信公众号平台内容在一定群体中的传播及影响的研究。
2. 小组成员可以熟练掌握并应用定量分析、采访访谈技巧等研究方法。
3. 能够使用爬虫爬取网页，并对爬取下来的内容进行分词处理、词频分析等，即已经可以抓取文本并且对文本进行最基本的分析操作。
4. 数理统计知识准备充足，已经可以对大量数据进行结构化的归类和统计分析，整理出想要的有用数据信息。
   1. **已具备的条件，尚缺少的条件及解决方法**

已具备条件：

1. 掌握MySQL基础知识与操作，能够熟练使用数据库并进行数据库的管理等，比如建立图片库，为文章生成合适的配图；
2. 具备Python编程基础，为对NLTK类库和SnowNLP类库的学习和实际应用做好了准备；
3. 有扎实的数学、概率、数理统计基础，对人工智能领域的基本算法已有初步了解并能够合理应用，为更进一步的研究打好了坚实的理论基础；
4. 初步学习过机器学习相关知识，为日后生成模板文本和完整连贯文章等研究成果提供理论基础；

缺少条件：

1. 课题团队目前缺少对市场需求细化的把握、对活动类微信推文大数据资源的掌握，尤指论文资料、课程书籍等；对用户的具体需求了解还有所不足。
2. 目前尚缺乏对NLP技术的了解，暂时难以通过自然语言处理的现有手段对文本进行分析；

解决方法：

针对目前问题，团队将通过市场调查分析及模式探索创新解决已有问题，采用问卷、当面采访等方式，了解用户需求，以实时更新团队的具体研究方向，呈现出符合用户需求的研究成果。

1. 针对目前问题，团队将通过市场调查分析及模式探索创新解决已有问题，采用问卷、当面采访等方式，了解用户需求，以实时更新团队的具体研究方向，呈现出符合用户需求的研究成果。
2. 学习NLP自然语言处理技术、进一步深化机器学习领域的探索和研究，以解决生成文本模板、自动生成连贯的文章、完成完整排版的技术；

**成员参与科研情况：**

王泽宇：

1.曾搭建过一台服务器，至今仍在供自己及同伴使用；

2.熟悉C++、Java、Python编程，曾参与设计过较大的数据管理系统，能够对大量数据进行储存、管理和分析；

3.掌握新闻的基本知识，曾参与制作数据新闻《分析了3220个地名，我们寻到了中国地名的文化旧根……》