

# 文本挖掘下的话语认知与构建

## ——以人工智能和 5G 领域白宫新闻文本为例

张誉曜, 陈媛媛\*

(新疆师范大学 计算机科学技术学院, 乌鲁木齐 830001)

**摘要:** [目的 / 意义] 政府新闻文本是政策文本的表达形式之一。对政府新闻文本进行挖掘分析, 能直观的展现表面动态, 揭示对话事件的立场及其背后关系, 能对情报分析工作、中国新型智库的发展建设, 提供有效的助力。[方法 / 过程] 文章利用 TextRank 算法进行核心句提取, 并结合批评话语分析理论框架来做评估, 以语料驱动从关键句、关键词两个宏观角度讨论政府决策态度, 以词性从话语策略的微观角度总结美国官方对人工智能和 5G 领域的话语认知形象。[结果 / 结论] 通过分析可以看出, 美国政府对于人工智能和 5G 领域的态度将持续保持在“政府引领下的美国需优先”这一局面, 面对发展中国家的科技竞争, 增加政府政策性文本的指导无疑是最好的对策建议。

**关键词:** 智库; 文本挖掘; TextRank 算法; 批评话语; 人工智能

**中图分类号:** G249.2      **文献标识码:** A      **文章编号:** 1002-1248 (2021) 04-0035-10

**引用本文:** 张誉曜, 陈媛媛. 文本挖掘下的话语认知与构建——以人工智能和 5G 领域白宫新闻文本为例[J]. 农业图书情报学报, 2021, 33(4): 35-44.

## Discourse Cognition and Construction Based on Text Mining: Taking the White House News Text in the Field of Artificial Intelligence and 5G as an Example

ZHANG Yuyao, CHEN Yuanyuan\*

(College of computer science and technology, Xinjiang normal university, Urumqi 830001)

**Abstract:** [Purpose/Significance] Government news text is one of the expression forms of [Purpose/Significance] Government news text is one of the forms of policy text. Text mining and analyzing the government news text can

收稿日期: 2020-07-28

基金项目: 2019 年度国家社科基金项目“美国著名智库网络影响力及其启示研究”(19BTQ067)

作者简介: 张誉曜 (ORCID: 0000-0003-1198-815X), 男, 硕士研究生, 新疆师范大学计算机科学技术学院, 研究方向为人工智能、数据挖掘与分析

\*通信作者: 陈媛媛 (1977-), 女, 博士, 硕士生导师, 新疆师范大学计算机科学技术学院, 研究方向为智库建设与评价、网络计量等研究。Email: 252019741@qq.com

intuitively show the dynamics, reveal the position of dialogue events and the relationship between them, and improve the intelligence analysis work and promote the development and construction of China's new think tanks. [Method/Process] This paper uses the TextRank algorithm to extract core sentences, and evaluates it with the theoretical framework of critical discourse analysis (CDA). Using the corpus, the paper discusses the government's decision-making attitude from two macro perspectives of key sentences and keywords, and sums up the American official discourse cognitive image in the field of artificial intelligence and 5G from the micro perspective of discourse strategy. [Results/Conclusions] The analysis shows that the attitude of the US government towards artificial intelligence and 5G will continue to maintain the situation that "a priority to the United States under the guidance of the government needs". In the face of technological competition in developing countries, depending on the guidance of the government in policy text is undoubtedly the best countermeasure.

**Keywords:** think tank; text mining; TextRank algorithm; critical discourse; artificial intelligence

## 1 引言

官方新闻是指权威机构所发布的新闻，所发布的内容能够代表其机构的意志。政府新闻文本是国家展现执政理念、国家战略、政策体系等描述政府工作方针、内容和计划的表现形式之一。白宫作为美国总统官邸，其发表的新闻文本具有即时性、广泛性以及权威性等特点，可以更好的引领、影响舆情。“智慧洞察，前瞻储备”为当今中国特色新型智库体系功能要求，研究分析政府新闻文本对情报分析、决策制定等方面具有重要引导作用<sup>[1]</sup>。

一直以来，推理法、内容分析法与量表法是各国研究人员对政府新闻文本态度识别研究中广泛使用的方法，研究人员在历史事件与政策方针的基础上进行逻辑推理，从而分析政府态度与变化情况；内容分析法，借鉴评价理论对新闻文本中的态度资源进行编码，进而识别新闻报道等文本数据中蕴含的政府态度；量表法，使用特定的态度量表进行针对性测评，以分析政府态度<sup>[1,2]</sup>。本文以传统新闻传播学的主要研究方法内容分析为参照，在样本、方法等方面与文本挖掘进行了对比：①在样本容量方面，内容分析法为小样本，文本挖掘法为大样本；②在抽取方法方面，内容分析法为随机抽样，文本挖掘法为方便抽样或者随机抽样；

③在内容方面，内容分析法为态度，文本挖掘法为态度和行为；④在分析单位方面，内容分析法为文章，文本挖掘法为符号和词句；⑤在编码方法方面，内容分析法为人工操作，文本挖掘法为人工和机器学习方法。对比发现，相比于内容分析法，文本挖掘法在对于大数据的处理方面更加有优势且适当增加了分析内容的宽度，对于结构化的分析来说文本挖掘法更加有效。

近年来，随着科技的发展，大数据在数据分析的广泛应用以及人工智能的普及，用传统的方式对海量数据文本进行处理时困难增加了许多，且速度较慢、准确率较低，越来越不适用。研究人员开始利用数据挖掘技术对文本进行挖掘分析，利用文本挖掘技术，可以更快速、更直接的获取所需核心信息，达到智能化的分析与决策，增加其科学性与实效性。本文利用成熟算法加持，将其扩展至新闻文本获取分析上，应用创新于智库文本研究，为智库研究学者扩展新的思路方法。

## 2 理论研究

### 2.1 文本挖掘理论依据

在常用的文本挖掘方法中，基于词典的文本分析使用较为广泛，基于词典的文本分析在前期需要学者

规划一些用于划分类的词语, 建立词典并用算法进行匹配以达到分类的目的。CONWAY 等利用此方法研究了 2012 年在美国总统竞选中媒体之间的议程设置, 比较了报纸和 Twitter 文本的相互关联<sup>[3]</sup>。对于关键词的提取, 目前分为两大块, 有监督的和无监督的。有监督的方法为利用标注好的训练集进行训练, 但由于需要人工预处理导致效率较低, 所以无监督的方法使用较为广泛。无监督的关键词提取主要分为基于 TF-IDF 模型、基于 LDA 模型以及基于词图模型 3 种方法。TF-IDF 是基于词频的模型, 并没有结合词性本身<sup>[4]</sup>; LDA 模型依据矩阵对关键词进行抽取, 抽取效果和训练集的主题息息相关<sup>[4]</sup>; 词图模型是将对文本的处理转化为网络图的连接分析, 该方法相比前两种无需批量训练数据集, 简单且效果较好, 例如对文本内的词语关系分析的 TextRank 算法等<sup>[5]</sup>。

而新闻文本, 作为一种新闻报道的存在形式, 其文本形式相对单一, 结构简单, 直入主题且表达明确。政府新闻文本, 作为新闻文本其中的特殊一类, 其特殊性在于政府利用直观、简洁、多层次以及“模糊语”来表达其政策性的文本<sup>[6]</sup>。政府新闻文本可看作一种特殊的政治语言, 其模糊语能有效的缓和问题双方的矛盾<sup>[7]</sup>。新闻文本主题多而杂, 一篇带有强烈感情色彩的文本通常会包含着大大小小数 10 个主题, 采用传统的内容分析法已然不适用。本文利用文本挖掘法提取文本关键句, 在新闻文本分析中, 通过文本挖掘技术, 可以更加智能化的洞悉政府的态度倾向, 为研究学者扩展了新思路。本文结合成熟算法 TextRank, 将算法应用于新闻文本数据获取方面, 进一步展现该领域的应用创新, 简要介绍算法如下。

## 2.2 TextRank 算法简要介绍

TextRank 算法是一种文本排序算法, 其是由 PageRank 演变而来。随着科技的进步, 在 20 世纪 90 年代末, 由 LARRY 和 SERGEY 提出的 PageRank 算法诞生<sup>[8]</sup>。PageRank 最初是为解决网页与网页之间的关系而提出的一种计算网页权重的算法, 它的思想核心为模拟使用者进行上网, 随机点开网页浏览并且随机

跳转任何网页, 若某网页被很多网页连接到, 那么其跳转到的概率会很大, 该网页 PR 值也相对较高<sup>[8,9]</sup>。PageRank 算法核心公式如 (1):

$$PR(*) = (1 - \delta) + \delta * \sum_{p \in Z_i} \frac{S_p}{B_p} \quad (1)$$

$\delta$  为阻尼系数, 通常设置为 0.85,  $Z_i$  为所有链接到网页  $i$  的集合,  $S_p$  为网页  $p$  的 PR 值,  $B_p$  为网页  $p$  的对外链出数, 由此, 可以有效的得出网页的重要性值。

PageRank 算法是根据网页之间的联系来构造网络, 而 TextRank 算法将网页换成文本, 文本即为句与句之间组成的有序结合, 将句子设为节点, 句子与句子之间的相似度类比于网页之间的链接关系, 也就是根据词之间的共现关系构造网络<sup>[10,11]</sup>。PageRank 算法构造的网络中的边为有向无权边, 而 TextRank 算法构造的网络中的边为无向有权边。

在这里, 本文先将 TextRank 详细提取关键词和关键句详细算法步骤总结如下。

- (1) 将所需文章内容整合成文本数据;
- (2) 将文本分割成整句形式, 如  $T=[j_1, j_2, j_3, \dots, j_n]$ ;
- (3) 依次将每个句子进行分词和词性标注, 删除停用词, 保留指定词性的词;
- (4) 向量化, 计算句子向量之间的相似性并存放在矩阵中, 如矩阵  $M = \begin{bmatrix} 0 & \dots & n-1 \\ \dots & & \\ n-1 & & \end{bmatrix}$ , 其中  $M[x, y]$  代表句子  $x$  和句子  $y$  之间的相似度;
- (5) 构建将相似矩阵转换为以句子为节点、相似性得分为边的图结构, 计算权重;
- (6) 根据排名合理提取。

TextRank 算法的核心公式如 (2):

$$WS(V_i) = (1 - \delta) + \delta * \sum_{p \in Z_i} \frac{\omega_{ip}}{\sum_{k \in B_p} \omega_{pk}} WS(V_p) \quad (2)$$

这里只是将 PageRank 算法公式略作修改, 表示两节点边的重要程度。

## 2.3 句子相似度

在从所需文本中提取核心关键句的情况下, 将所需文本中每个句子单独看做一个节点, 若有两个句子

有相似性，那么即可以认为这两个节点之间存在无向有权边<sup>[12]</sup>。令  $S_j$ 、 $S_k$  为两个句子，对其句子中词的个数求对数后再求和，并求在同一时刻出现在两个句子中的同一词的数量，具体公式如 (3) ~ (4) 所示。

$$\log(S_j) + \log(S_k) \quad (3)$$

$$|W_k| |W_k \in S_j \cap W_k \in S_k| \quad (4)$$

即，求句子相似度的公式为：

$$\text{Similarity} = \frac{|W_k| |W_k \in S_j \cap W_k \in S_k|}{\log(S_j) + \log(S_k)} \quad (5)$$

$W_k$  为句  $S_j$  与句  $S_k$  中共有的词，根据此公式，计算每个文本句子的相似度，随后通过设置阈值去掉最低值，构建图并重复迭代计算，依次排序得出所需文本句<sup>[13]</sup>。同理，将节点由句子换为词，即为词的提取，在这里就不在详细描述。

### 3 实证研究

#### 3.1 数据采集

考虑到政府新闻文本的实时性，本文结合当代中美两大科技问题进行数据采集分析。笔者使用“5G”“artificial intelligence”为搜索词，在美国白官网站

(www.whitehouse.gov) 官方新闻中进行数据获取，检索 2020 年、2019 年、2018 年、2017 年 4 年，共检索到相关新闻信息文章 556 篇，随后利用数据采集软件对上述检索出来的新闻文章进行采集收集，并且导入 Excel。结合本文所要突出的核心以及后续话语分析，笔者在这里进行了人工筛选对数据进行了清洗，去除与搜索词相关度不大的文章以及产品介绍等非政策性文本，最后得出所需新闻信息文本 266 篇 (图 1)。

以笔者人工筛选后的结果为例，分析后可以发现，作为政府部门，其在人工智能与 5G 领域 2017 年的上半年新闻发布量为 28 篇，下半年为 33 篇，上下半年从发文量来看基本持平；2018 年上半年新闻发布量为 35 篇，下半年为 44 篇，2018 年的发文量相比 2017 年略有增长，尤其在下半年的时候 5G 相关的新闻文本有了较大激增；2019 年上半年新闻发布量为 44 篇，下半年为 46 篇，发文量有了提高，此时人工智能领域的文本有了较大的占比；进入 2020 年，截止笔者搜索前，2020 年上半年的新闻发布量已经达到了 36 篇，人工智能领域的文本占 31 篇，占总发文量的 86% (图 2)。笔者结合时事政治因素，发现随着近两年华为与高通的 5G 之争、人工智能领域的火热，各国加大了对人工智能领域的研发主导权，相关发展中国家的创新应用

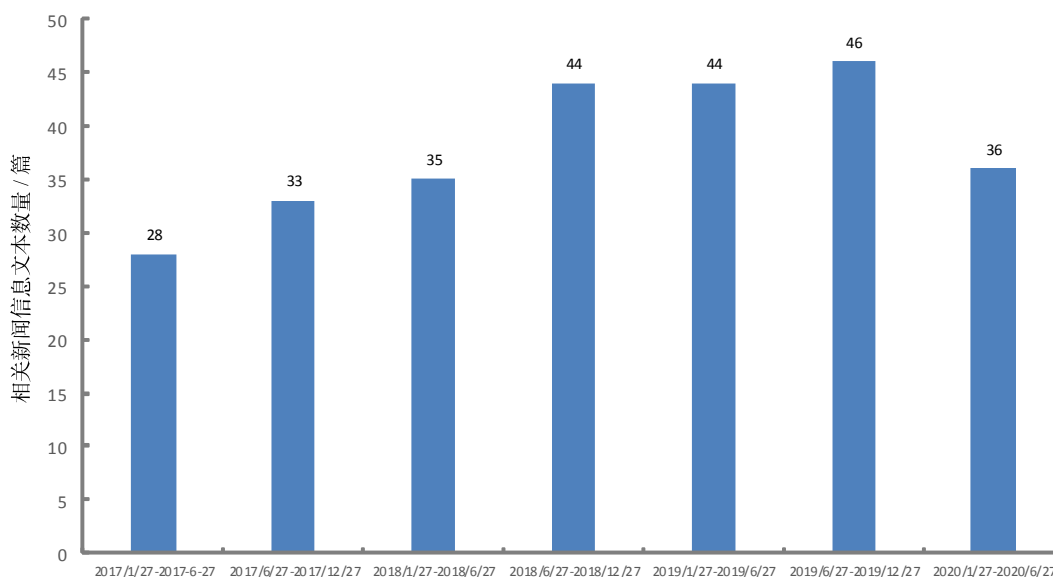


图 1 相关新闻数目统计图

Fig.1 Statistics of related news

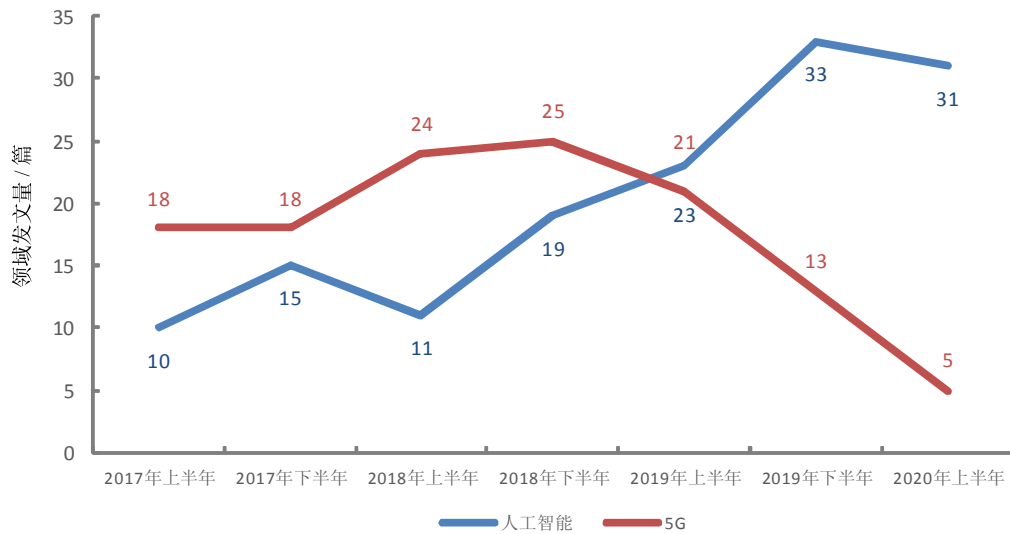


图 2 人工智能与 5G 领域发文量对比

Fig.2 Comparison of the number of papers published in the field of artificial intelligence and 5G

等也随之增加, 白宫新闻发布量也随之激增, 后文将结合词性微观角度对政府所发新闻文本进行话语分析。

### 3.2 词频统计

前文所述关键句获取后, 再以句子为素材进行关键词获取, 其具体步骤不在细述, 在这里利用 WordSmith7 软件对所获取数据文本进行统计, 排名如表 1 所示。

表 1 文本词频统计 (部分)

Table 1 Text word frequency statistics (part)

词	排名	频率	占比/%
American	3	73	1.46
China	5	57	1.12
Challenge	11	47	0.96
technology	26	39	0.87
artificial intelligence	37	25	0.78
country	44	14	0.69
future	46	11	0.42
.....			

表 1 中, “China” 与 “American” 都排名前 10, 且使用频率提及的较高, 说明在人工智能和 5G 领域, 中国具有较高的科技创新以及科研能力, 美国官方对中国的关注度远远高于其他国家。刨除 “China” 等专

有名词可以发现, 词 “Challenge” 名列前茅, 间接说明了美国官方关注点的根本旨意, 下面结合不同领域的关键词进行话语分析。

## 4 文本分析研究

### 4.1 批评话语对比分析

批评话语分析最早诞生于 20 世纪 70 年代, 其将社会权力和话语紧密结合, 运用话语分析的方式披露社会问题并给出解决办法<sup>[14]</sup>。批评话语认为文本结构与社会结构的关系是间接的, 通过媒介建构的批评话语, 分析的目的就是认清并且抗争不平等和不公平; 而不平等和不公平就涉及权力的滥用, 权利意味着控制, 即一部分对另一部分的控制, 故批评话语分析研究的一个重要问题就是权力<sup>[15,16]</sup>。批评话语分析的核心是权力、话语与意识形态, 下面结合词频从词性的微观角度进行话语分析。

### 4.2 美国政府对于 5G 及其发展的话语认知与构建

在 5G 主题下, 结合词间关系以及词性词频来分析例句中隐含的意识形态意义, 认知其话语的构建与背

后深意。该主题下,涉及到的高频实词名词有 country、future、United States、trade、economic、information 等,该类型的名词频率高关联紧密,比如词 country 的出现往往伴随着 United States、trade 的词频仅低于国

例 1: President Donald J. Trump is ensuring the United States leads in the global race to deploy secure and reliable 5G communications. This Administration's policies empower innovation and investment in America's 5G readiness, paving the way for what are sure to be some of the most substantial technological and economic advancements of the 21st century.

例 2: Information age, the nation that leads the world in wireless technology wins. To keep America's edge, we must accelerate our development and deployment of 5G.

例 3: America cannot risk lagging behind other countries. We must protect the economic and security advantages that come from the Nation's preeminence in wireless. The Administration has taken concrete steps to promote United States wireless leadership through smart tax and infrastructure policy, streamlining regulations, and developing stronger cybersecurity protections.

例 4: Wireless contributes \$475 billion to the economy every year and supports nearly 4.7 million jobs—many of which are high-wage and long-term. America's future will have connected homes and farms, autonomous vehicles, drones, and smart cities and communities. The Nation's wireless networks must be ready to support the foundation of America's future growth and prosperity. we will improve our national security through technological advancements. With 5G, we can better defend America with more resilient critical infrastructure, technologies like state-of-the-art radar systems, and cutting-edge communications on land and in space.

经济是外交政策的基础,在如今这个实力地位的国际关系中,科学技术已经成为了经济力量的强大助推器<sup>[17]</sup>。例 1 中带有强烈态度情感的词 ensuring、empower 以及 paving the way 进一步的展现出了核心观

点 in the global race to deploy secure and reliable 5G communications, 体现出了美国当下特朗普政府对于 5G 领域的大力创新支持以及不计余力的投资。例 2 由 must、accelerate 词可以看出美国政府对于 5G 领域的心态,实际上在该领域中国华为所展示的成果以及水平并不亚于美国,但作为政府新闻网站,模糊性与委婉性是它们的特色,通过 wins 可以看出美国的好胜心。科技本无国界,但科技领域利益巨大,美国政府无意“分享交流”,只为第一。例 3 通过具有强烈语气词 cannot、must、protections 来展示出美国政府对于无线领域第一的强硬态度,口吻上面更偏向与美国政府对美国民众的讲话,实际上侧面的反映出了美国政府对外的外交政策,无线领域的霸主地位和不可挑战性,美国政府表明其国家势在必得的态度。例 4 通过具体的数据向美国民众突出了占领无线领域尖端的好处,从工作数量、收入待遇、未来国民优惠政策方面来刺激美国私企与民众的心态。随后把 5G 与国家安全结合起来,情态动词 will 加上 improve 勾勒出了美国政府强力支持拿到 5G 技术的话语权是为了保护美国国土安全的画面。由此可见,一方面,美国政府针对 5G 这个信息科技领域的主动权是无可争议的,其主张强化本国利益优先,弱化全球合作,充分达到了既可以对美国利益的完美保护,又可以体现特朗普政府的美国需优先原则<sup>[18]</sup>。另一方面,从频繁出现的强烈态度情感词可以看出,美国政府对于 5G 领域的创新以及投入的第一准则就是美国单边利益至上。

### 4.3 美国政府对人工智能领域及其发展的话语认知与构建

借助高频词统计分析来看,针对人工智能领域,主要高频词有 China、economic、America、innovate、prospect、development 等。同理以高频词为节点,进一步考察人工智能领域的话语认知构建,举例如下。

例 5: President Trump released the National Strategic Overview for Quantum Information Science to guide Federal QIS actions, including the establishment of a Quantum Economic Development Consortium to build the

QIS industrial ecosystem.5G: The development and deployment of high speed, high capacity networks will spur innovation, enable cutting-edge technological advancements, and bring the benefits of connectivity to all Americans.

例 6: AI promises great benefits for American workers, with the potential to improve safety and increase productivity, maintaining our Nation's global leadership in AI will ensure that the technology is developed in a manner consistent with our Nation's values, policies, and priorities. AI must also be developed in a way that does not compromise our American values, civil liberties, or freedoms. Rapid technological innovation can lead to radical improvements in society and quality of life.

例 7: The Trump Administration recognizes the importance of securing our place as the world's innovation leader for generations to come. Today, the Administration is releasing a plan to help secure that future through STEM education: Charting a Course for Success: America's Strategy for STEM Education. In this time of rapid technological innovation, the United States finds itself in a global competition for STEM talent. Organizations from across the entire STEM ecosystem have been working to improve STEM education and training, with many examples of success upon which to build. Although Americans' basic STEM skills have modestly improved over the past two decades, there is still much room for improvement as America's adversaries work hard trying to surpass us.

例 8: Artificial Intelligence (AI) promises to drive growth of the United States economy, enhance our economic and national security, and improve our quality of life, Maintaining American leadership in AI requires a concerted effort to promote advancements in technology and innovation, while protecting American technology, economic and national security, civil liberties, privacy, and American values and enhancing international and industry collaboration with foreign partners and allies.

近年来, 发展中国家尤其是中国的快速崛起和在新兴技术领域不断取得突破, 引起了美国越来越大的警惕和抵制<sup>[19]</sup>。长期以来, 美国都是将国家安全建立在对前沿技术的垄断基础之上, 因此对于正在兴起的人工智能技术极为重视<sup>[20]</sup>。例 5 没有直接强烈的表明政府态度, 而是引用了特朗普政府发布的文件, 从另一个层面上增加了权威性和说服力。例 6 认为人工智能可以为美国带来巨大的利益, 要保持美国在人工智能领域的全球领导地位, 并用话语强化策略, 通过情态动词 will 表现了占领人工智能领域主导地位能给美国民众带来的根本改善。STEM 教育, 即科学 (Science), 技术 (Technology), 工程 (Engineering), 数学 (Mathematics), 人才一直是国家综合实力的保障与发展力量<sup>[21]</sup>。例 7 中特朗普政府认识到未来人才是确保美国作为世界创新领袖地位的重要性, 虽然在过去的年份里美国在 STEM 教育中一直走在前沿, 但随着其他国家的超越, 美国已经感觉到了危机, 在这个技术创新迅速的时代, 人才培养与竞争至关重要。例 7 指出, 政府部门已经将信息科学上升为国家战略意义, 通过政府部门的作用刺激创新以及尖端技术的进步, 以此来突出美国在全球信息领域的核心位置稳定不变。例 8 的背景是在中美重启贸易谈判之际, 在 2019 年特朗普签署的美国 AI 计划, 即“维持美国在人工智能领域领先地位”的总统令, 该令标注了美国政府扶持人工智能的五大重点, 推动基础发展、共享资源、建立规则、培育人才以及国际推广, 要求联邦政府加大资源投资用于人工智能的研究、推广, 影响美国市场创新活力, 该目的皆在推动美国在人工智能领域的发展, 同时削弱中国在这一领域的强劲势头。与 5G 领域相比, 美国政府将人工智能看作了下一代科技浪潮, 发言稿大幅较少但政策性文本却明显增大, 这是与 5G 领域话语对比最大的差异。中国曾提出要加强在前沿技术领域合作, 建设 21 世纪的“数字丝绸之路”的主张, 人工智能是其中的一项重要技术合作内容, 为此美国智库认为这是中国在“一带一路”的布局下通过投资实体和数字基础设施网络, 破坏美国与盟友间的网络体系, 重塑世界秩序和世界体系的行为<sup>[22,23]</sup>。随着

近年来发展中国家尤其是中国在人工智能领域的创新发展, 美国政府的科技霸权主义渐渐感觉到了危机, 增加政府的政策性文件指导无疑为最好的办法。

## 5 结论与展望

笔者在这里对白宫网站进行了信息查询, 所有来源均为美国政府官方文章, 所选题材为当下最热的科技尖端领域人工智能以及 5G。结合算法梳理后发现, 针对人工智能, 美国政府观念为“美国需优先”“政府引领”“数据要开放”, 特朗普政府对于此的表示并非为单纯的技术提升以及开发路线, 而是从国家战略角度集中调动联邦政府的资源, 由顶层引领, 从上至下推动人工智能的发展。算法、芯片和数据为人工智能的三大支柱, 美国在算法和芯片上保持着优势, 但是训练数据方面却是短板, 开放数据和计算资源为特朗普政府所倡导的特点, 但美国相关专家却表示此方案细节方面并不详细, 其执行的可能性较低<sup>[24]</sup>。

在 5G 方面, 特朗普政府的做法并不是由政府投资引导发展, 而是变为由私营部门驱动。美国联邦通信委员会宣布截止 2019 年底, 美国将会拥有 92 个商用 5G 网络, 计划投资 2 750 亿美元, 为自身创造 300 万个工作岗位。特朗普政府在对于 5G 竞争方面采取了十分大胆的行动, 其政府推出美国史上最大规模的频谱拍卖, 允许运营商竞标 37GHz、39GHz 和 47GHz 三个高频段, 并允许固网卫星运营商使用 50GHz 频段, 其目的是利用新增的频谱促进 5G、物联网和其他基于频谱的先进服务发展。美国抢着最先开通 5G, 竞争力方面由 2018 年的排名第三, 变为现在的一举超越韩国, 和中国并列第一<sup>[25]</sup>。

从理论方面出发, 本文丰富了对于政府文本态度研究的方法, 加强了对政府新闻文本分析的手段。当下, 对于国内外的研究学者来说, 对于政府文本态度的研究较常使用的为量表法和推理法以及内容分析, 效率较低且耗时耗力。本文以文本挖掘技术为引导, 结合批评话语分析理论框架作为评估, 对提取的核心句从关键词的微观方面来讨论政府决策态度问题, 此

举可为后续对于政府新闻文本的综合研究提供了新方法, 扩展了新思路。

另外研究存在一定不足, 一方面, 在对于所获取的新闻文本信息进行第一步人工筛选的时候, 可能会因为人工因素存在着一些偏差; 其次, 对于政府新闻文本的特征考虑的较为笼统, 由于政府新闻文本存在着简短性、多重性、模糊性等特征, 在对其进行宏观方面的讨论时应当综合考虑分析, 若考虑不周, 其种种特殊性很有可能会影响学者的最终判断。故在后续的研究中, 将会综合考虑新闻文本的特殊性, 针对性的提出解决方案, 在基于文本挖掘技术的前提下增加政府决策态度的识别性。

### 参考文献:

- [1] 段尧清, 何思奇, 林平. 基于新闻文本挖掘的政府态度识别实证研究[J]. 情报理论与实践, 2019(9): 86-92, 63.  
DUAN Y Q, HE S Q, LIN P. Empirical research on government attitude recognition based on news text mining[J]. Intelligence theory and practice, 2019(9): 86-92, 63.
- [2] 王世伟. 略论国家高端科技智库的功能定位[J]. 情报学报, 2018, 37(6): 590-599.  
WANG S W. On the functional orientation of national high-end science and technology think tanks[J]. Journal of information, 2018, 37(6): 590-599.
- [3] CONWAY B A, KENSKI K, WANG D. The rise of Twitter in the political campaign: searching for intermedia agenda-setting effects in the presidential primary[J]. Journal of computer-mediated communication, 2019.
- [4] 常耀成. 基于多特征融合的图模型关键词提取方法研究[D]. 天津: 中国民航大学, 2018.  
CHANG Y C. Research on keyword extraction method of graph model based on multi feature fusion[D]. Tianjin: Civil aviation university of China, 2018.
- [5] 刘奇飞, 沈炜域. 基于 Word2Vec 和 TextRank 的时政类新闻关键词抽取方法研究[J]. 情报探索, 2018(6): 22-27.  
LIU Q F, SHEN W Y. Research on keyword extraction method of current political news based on word2vec and textrank[J]. Intelli-



- gence exploration, 2018(6): 22-27
- [6] 杨保军. 新闻文本的个性特征[J]. 当代传播, 2004(2): 10-11.  
YANG B J. Personality of news text[J]. Contemporary communication, 2004(2): 10-11.
- [7] 田海龙. 语篇研究的批评视角: 从批评语言学到批评话语分析[J]. 山东外语教学, 2006(2): 40-47.  
TIAN H L. Critical perspective of discourse research: From critical linguistics to critical discourse analysis [J]. Shandong foreign language teaching, 2006 (2): 40-47.
- [8] 周锦章, 崔晓晖. 基于词向量与 TextRank 的关键词提取方法[J]. 计算机应用研究, 2019, 36(4): 1051-1054.  
ZHOU J Z, CUI X H. Keyword extraction method based on word vector and textrank[J]. Computer application research, 2019, 36(4): 1051-1054.
- [9] 宁建飞, 刘降珍. 融合 Word2vec 与 TextRank 的关键词抽取研究[J]. 现代图书情报技术, 2016(6): 20-27.  
NING J F, LIU J Z. Research on keyword extraction integrating word2vec and textrank[J]. Modern library and information technology, 2016(6): 20-27.
- [10] 夏天. 词向量聚类加权 TextRank 的关键词抽取[J]. 数据分析与知识发现, 2017, 1(2): 28-34.  
XIA T. Keyword extraction of word vector clustering weighted textrank[J]. Data analysis and knowledge discovery, 2017, 1(2): 28-34.
- [11] 李尚昊, 朝乐门. 文本挖掘在中文信息分析中的应用研究述评[J]. 情报科学, 2016, 34(8): 153-159.  
LI S H, CHAO L M. Review on the application of text mining in Chinese information analysis[J]. Information science, 2016, 34(8): 153-159.
- [12] 徐立. 基于加权 TextRank 的文本关键词提取方法[J]. 计算机科学, 2019, 46(s1): 142-145.  
XU L. Text keyword extraction method based on weighted textrank[J]. Computer science, 2019, 46(S1): 142-145.
- [13] 李志强, 潘苏含, 戴娟, 等. 一种改进的 TextRank 关键词提取算法[J]. 计算机技术与发展, 2020, 30(3): 77-81.  
LI Z Q, PAN S H, DAI J, et al. An improved Textrank keyword extraction algorithm[J]. Computer technology and development, 2020, 30(3): 77-81.
- [14] 赵芑. 2016 年国内批评话语分析研究综述[J]. 天津外国语学院学报, 2017(4): 72-79.  
ZHAO P. A review of critical discourse analysis in China in 2016[J]. Journal of Tianjin foreign studies university, 2017(4): 72-79.
- [15] 刘如, 张惠娜, 杜丽萍, 等. 基于情报 3.0 工作思路的自动简报系统设计及实现[J]. 情报学报, 2018, 37(2): 172-182.  
LIU R, ZHANG H N, DU L P, et al. Design and implementation of automatic briefing system based on intelligence 3.0 [J]. Journal of the China society for scientific and technical information, 2018, 37(2): 172-182.
- [16] 杨敏, 符小丽. 基于语料库的"历史语篇分析"(DHA)的过程与价值——以美国主流媒体对希拉里邮件门的话语建构为例[J]. 外国语(上海外国语大学学报), 2018, 41(2): 77-85.  
YANG M, FU X L. The process and value of corpus based "historical discourse analysis" (DHA): A case study of the discourse construction of Hillary Clinton's e-mail gate in American mainstream media[J]. Foreign languages (Journal of Shanghai foreign studies university), 2018, 41 (2): 77-85.
- [17] 李占芳, 彭梦莹. 美国主流新闻媒体对"中美贸易战"中双方形象的刻画——以《华盛顿邮报》为例[J]. 华北电力大学学报(社会科学版), 2019(5): 107-114.  
LI Z F, PENG M Y. The portrayal of the images of the two sides in the "Sino US trade war" by the mainstream news media in the United States - Taking the Washington Post as an example [J]. Journal of north China electric power university (Social science edition), 2019 (5): 107-114.
- [18] 刘文宇, 胡颖. 基于文本挖掘的非传统文本批评话语研究[J]. 天津外国语学院学报, 2020, 27(4): 29-41, 158-159.  
LIU W Y, HU Y. Research on non-traditional text critical discourse based on text mining[J]. Journal of Tianjin foreign studies university, 2020, 27 (4): 29-41, 158-159.
- [19] 彭红梅, 刘忠. 人工智能对国家安全的影响——基于美国四家智库的观点分析[J]. 社会科学文摘, 2020(8): 115-118.  
PENG H M, LIU Z. The impact of artificial intelligence on national security: An analysis based on the views of four American think tanks[J]. Social science abstracts, 2020(8): 115-118.
- [20] TEPEL T, GREGORY C. Allen: Understanding China's AI Strategy:

- Clues to Chinese strategic thinking on artificial intelligence and national security. Washington, D.C.: Center for a new American security February 2019[J]. SIRIUS – Ztschrift für strategische analysen, 2019, 3(3): 305–306.
- [21] 赵洪, 王芳, 王晓宇, 等. 基于大规模政府公文智能处理的知识发现及应用研究[J]. 情报学报, 2018, 37(8): 805–812.
- ZHAO H, WANG F, WANG X Y, et al. Research on knowledge discovery and application based on intelligent processing of large-scale government documents[J]. Journal of information, 2018, 37(8): 805–812.
- [22] 段伟文. 控制的危机与人工智能的未来情境[J]. 探索与争鸣, 2017(10): 7–10.
- DUAN W W. Crisis of control and future situation of artificial intelligence[J]. Exploration and contention, 2017(10): 7–10.
- [23] Amandeep Singh Gill. Artificial intelligence and international security: The long view[J]. Ethics & international affairs, 2019, 33(2): 12–20.
- [24] GUO L, VARGO C. The power of message networks: A big-data analysis of the network agenda setting model and issue ownership[J]. Mass communication & society, 2015, 18(5): 557–576.
- [25] ZHENG P, LI X. Is social media for everyone?: The role of social capital and communication anxiety in structuring use of social networking sites[J]. Journal of social media studies, 2016: 2.