

基于自动分词技术的国家宏观科技创新政策发展量化评价研究

刘 旭

(中国科学技术信息研究所, 北京 100038)

摘 要: 本研究基于自动分词技术, 通过对一段时间内国家科技管理部门发布的具有代表性的规范性文件进行量化评价研究, 在词频统计的基础上, 引入资讯检索与文本挖掘常用加权技术 TF-IDF, 评估单一字词对于文本整体的重要程度, 从而进一步整理出科技部党组 1 号文件中常年出现的高频词、体现时代特点的高频词和近年新出现的高频词, 对比具有特点的高频词, 从而深入梳理国家宏观科技创新政策的发展脉络, 对我国宏观科技创新政策进行深入研究。

关键词: 自动分词技术; 科技创新政策; 高频词; 量化评价研究

中图分类号: G311 **文献标识码:** A **DOI:** 10.3772/j.issn.1009-8623.2020.10.009

2020 年是全面建成小康社会、实现第一个百年奋斗目标之年, 是实现迈进创新型国家行列目标的决胜之年, 在我国科技发展历程中具有重要里程碑意义和承前启后作用。党的十八大以来, 党中央把科技创新摆在国家发展全局的核心位置, 全面实施创新驱动发展战略, 科技实力显著提升^[1]。站在这个特殊的时间节点回望, 我国科技创新取得一批重大成果。成绩背后, 离不开国家宏观科技创新政策的导向作用。本研究引入自动分词技术, 对一段时间内国家科技管理部门发布的具有代表性的规范性文件进行量化评价研究, 以尝试为科技创新主体解读国家宏观科技创新政策提供新的视角。

科技部作为国家科技管理部门, 每年以党组发文形式发布的国家科技创新工作的指导意见统称为“科技部党组 1 号文件”。科技部党组 1 号文件是全面阐述我国宏观科技创新政策的规范性文件, 通常在每年年初印发, 其内容涵盖当前和今后一个时期科技创新工作的总体思路、战略布局、政策制定、体制改革, 以及科技管理干部队伍建设等多个方面。

历年的科技部党组 1 号文件, 既是对上一年国家科技工作的全面总结, 又是对下一年国家科技工作的规划和部署。因此, 对科技部党组 1 号文件文本进行自动分词处理, 结合所处时代背景和国内外形势加以分析, 可以梳理形成国家宏观科技创新政策的发展脉络, 并为相关研究提供量化依据。

1 文献回顾

1.1 国家宏观科技创新政策研究综述

赵筱媛等^[2]基于供给、环境和需求三种政策工具(X 维度), 结合科技活动中的基础研究、研发和产业化三个阶段(Y 维度), 并依据其对第一产业、第二产业、第三产业的作用领域(Z 维度)不同, 构建了公共科技政策的三维分析框架。刘会武等^[3]面向创新政策评价, 对评价主客体以及相互间关系进行了界定分析, 提出了由评价目的、评价时间和创新活动构成的三维分析框架, 并以我国高新技术产业开发区为案例加以分析和佐证。随后, 刘会武等^[4]提出了由国家创新战略、政策供给、

作者简介: 刘旭(1992—), 男, 助理研究员, 主要研究方向为科技创新政策。

收稿日期: 2020-08-28

政策需求和评价判断 4 种要素及其相互关系构成的创新政策“钻石模型”分析框架,并从我国创新战略对政策演变的影响、政策供需关系以及政策评价等几方面进行了论述。赵伟等^[5]从科技政策工具的变化、科技资金投入方式转变和未来科技政策的趋势 3 方面阐述了创新驱动下科技政策由微观转向宏观的演变趋势。

1.2 国家宏观科技创新政策量化评价研究综述

刘凤朝等^[6]将统计分析方法引进创新政策研究领域,以《国家促进自主创新的政策分析与研究》为样本,依据创新政策内容类别和法律效力建立基本维度,分析了 1980—2005 年我国创新政策的演变趋势。彭纪生等^[7]提出了政策量化的操作标准,阐述了政策部门协同、目标协同、措施协同的演变路径以及对经济绩效的影响。Tanaka 等^[8]提出了一种新的定量政策评估文献计量方法,运用统计和定量校正的循证决策方法,探讨了财务支持的时机与研究人员进行重要研究的时机之间存在的差距问题。张永安等^[9]选取 2001—2016 年的国家科技创新政策文本作为研究对象,以创新投入和科技产出作为主要指标,对我国科技创新政策与国家创新能力的的作用机理进行检验。

可见,国家宏观科技创新政策的量化评价研究目前主要集中在分类标准、量化标准、统计方法和评价体系等几个领域。本研究引入新的量化评价方法,试图基于概率语言模型的自动分词技术,侧重单一字词对于文本整体的重要程度,一定程度上避免了在上述领域由于评价标准不统一而产生的结果差异。

2 研究设计

2.1 数据来源

本研究选取 2013 年至今科技部网站政府信息公开目录中科技部党组 1 号文件作为研究对象。值得说明的是,由于 2018 年国务院机构改革,将原科技部与外专局的职责整合,重新组建科技部,因此,科技部党组在 2018 年 1 月和 5 月两次发布党组 1 号文件。为保证研究结果的一致性,在此明确选用 2018 年 5 月的文本作为 2018 年科技部党组 1 号文件进行研究。

2.2 研究工具

自动分词技术是信息处理的基础,广泛应用于搜索引擎、机器翻译、语音合成、自动分类、自动摘要和自动校对等多个领域。jieba 是一款基于概率语言模型分词技术的 Python 中文分词组件。其原理是,利用统计词典构造前缀词典,然后通过前缀词典对输入文本进行切分,得到所有的切分可能,根据切分位置构造一个有向无环图(DAG),依据动态规划查找最大概率路径,找出基于词频的最大切分组合。对于未登录词,jieba 利用 HMM 模型进行分词,将分词问题视为一个序列标注(Sequence Labeling)问题,其中句子为观测序列,分词结果为状态序列。首先通过语料训练出 HMM 模型,然后利用 Viterbi 算法进行求解,最终得到最优的状态序列,并根据状态序列输出分词结果。

2.3 研究方法

本研究对 2013 年以来的 8 份科技部党组 1 号文件文本进行自动分词处理。在词频统计的基础上,引入资讯检索与文本挖掘常用加权技术 TF-IDF,评估单一字词对于文本整体的重要程度,从而进一步整理出科技部党组 1 号文件中常年出现的高频词、体现时代特点的高频词和近年新出现的高频词,对比具有特点的高频词,从而对我国宏观科技创新政策进行深入研究。

3 量化结果

3.1 常年出现的高频词

如图 1 所示,从词频统计来看,这类词常年大量出现在科技部党组 1 号文件中。时代在变,但这类词出现的频率不减,热度不退,见证了我国宏观科技创新政策的发展历程。与此同时,虽然这类词常年出现在科技部党组 1 号文件中,但时代赋予它们的内涵和外延可能已经不同。将这些常年出现的高频词还原到历年的科技部党组 1 号文件中,结合时代特点和语言环境加以分析,可以得到这类词内涵和外延的变化趋势。

如图 2、图 3 所示,“科技”和“创新”两词常年大量出现在科技部党组 1 号文件中。在此,将“科技”和“创新”两词一并阐述。伍蓓等^[10]对政策领域的“科技”和“创新”两词进行了界定,

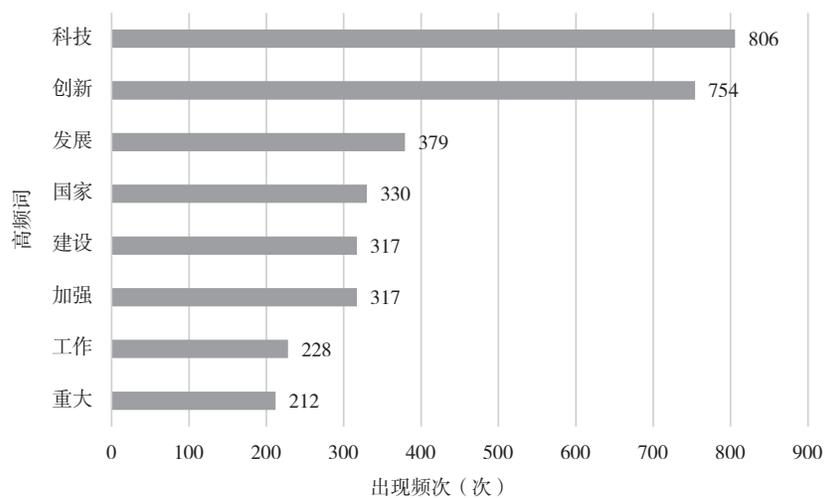


图 1 2013 年—2020 年高频词词频统计

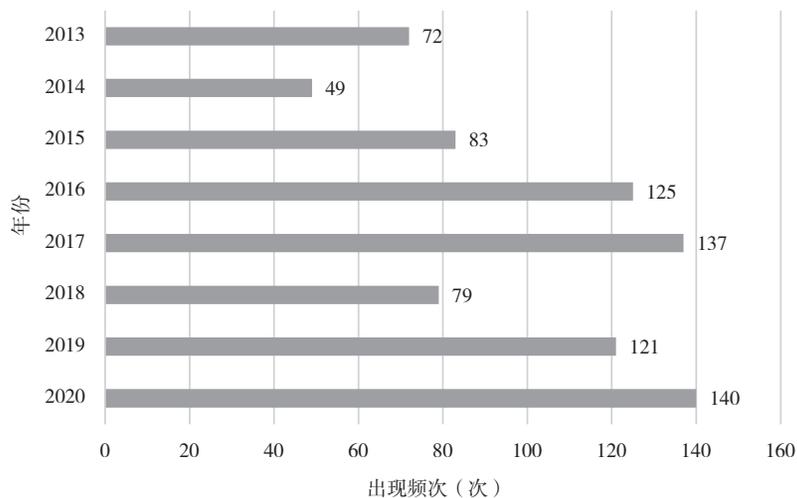


图 2 “科技” 一词的词频统计

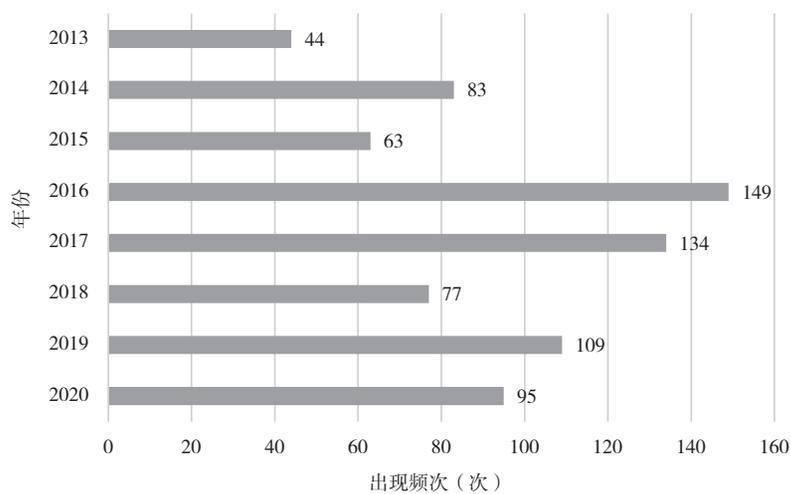


图 3 “创新” 一词的词频统计

认为“科技”一词侧重于科学前沿、重大理论研究和技术的实际应用等方面，而“创新”一词则侧重于营造创新环境和建设国家创新体系等方面。数据显示，在科技部党组1号文件中，“科技”一词的词频整体高于“创新”一词，但在2014年和2016年，“创新”一词的词频却超过“科技”一词，成为科技部党组1号文件中出现频次最高的词汇。

2014年科技部党组1号文件围绕“推动落实创新驱动发展战略”和“提高国家创新体系整体效能”做出了详细阐述，强调加大科技体制改革力度，进一步增强科技创新对经济社会发展的支撑引领作用。纵观全文，由于在上述2方面使用了较大篇幅，因此2014年科技部党组1号文件中“创新”一词的词频反超“科技”一词。

2015年6月，国务院印发《关于大力推进大众创业万众创新若干政策措施的意见》，从创新体制机制、优化财税政策等9方面提出30条具体举措，激发全社会创新潜能和创业活力^[1]。同年10月，党的十八届五中全会确立了“创新、协调、绿色、开放、共享”的发展理念，指出创新是引领发展的第一动力，必须摆在国家发展全局的核心位置。可见，由于“创新”一词在2015年国家层面发布的政策中频率上升明显，因此，这一现象也反映在2016年的科技部党组1号文件中。

值得一提的是，“创新驱动发展战略的顶层设计”的表述也仅出现在2014年和2016年的科技部党组1号文件中。区别在于，2014年科技部党组1号文件提出，加快实施创新驱动发展战略的顶层设计工作，要把做好实施创新驱动发展战略的顶层设计作为当年科技工作的首要任务；而2016年科技部党组1号文件明确，创新驱动发展战略的顶层设计工作已顺利完成。由此可见，2014年和2016年作为创新驱动发展战略顶层设计进程中2个重要的时间节点，也成为“创新”一词的词频超过“科技”一词的原因之一。

如图4所示，“发展”一词在历年的科技部党组1号文件中出现的频率同样值得关注。从党的十八大报告提出实施创新驱动发展战略，强调科技创新是提高社会生产力和综合国力的战略支撑，到党的十九大报告指出创新是引领发展的第一动力，是建设现代化经济体系的战略支撑，可以看出，表述在变，但在国家宏观战略中，“发展”与“科技”“创新”的关系始终密不可分。因此，“发展”一词在科技部党组1号文件中同样长期占据了重要位置。

如图5所示，“建设”一词在科技部党组1号文件中出现的频率基本呈现逐年递增状态，并在2017年出现明显增长。通过抽取2017年科技部党组1号文件中与“建设”一词相邻的关键词，“建

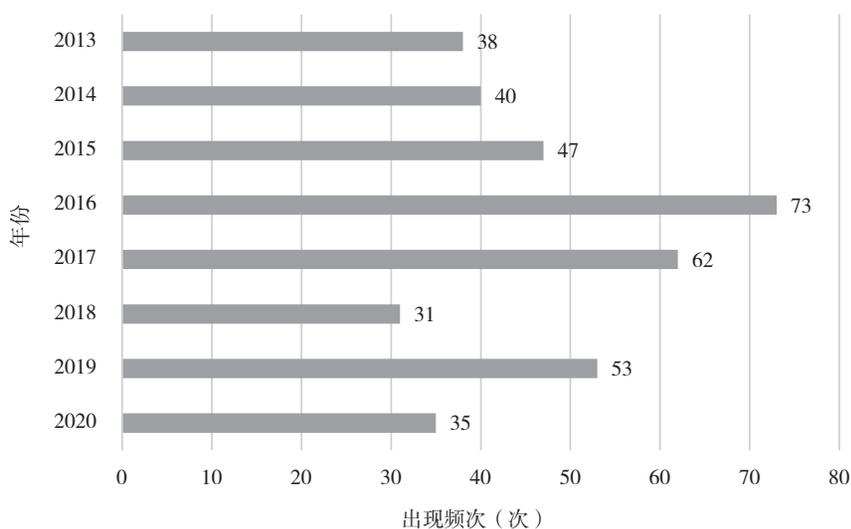


图4 “发展”一词的词频统计

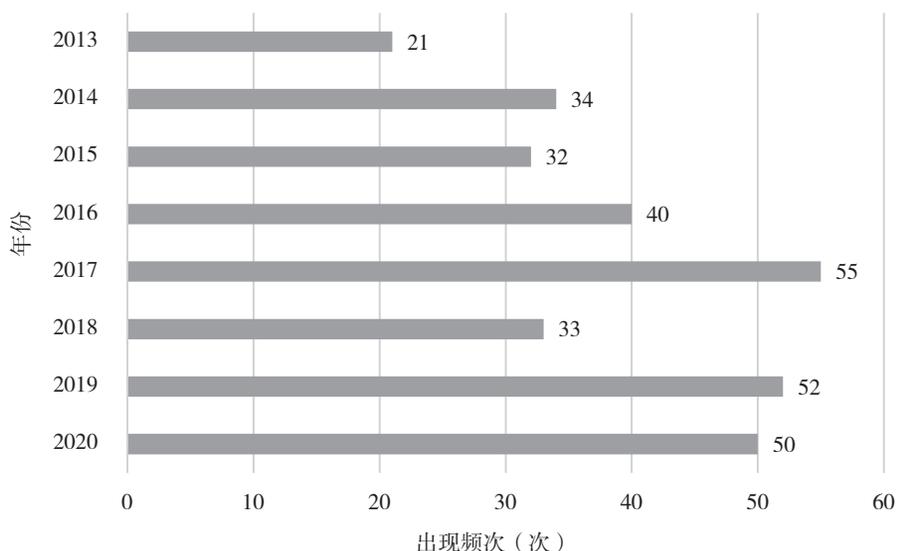


图5 “建设”一词的词频统计

设世界科技强国”这一表述引起了笔者的关注。统计发现，与“建设世界科技强国”有关的表述始于2017年，并在此后科技部党组1号文件中被持续提及。究其原因，早在2016年5月30日的全国科技创新大会、两院院士大会、中国科协第九次全国代表大会上，国家就号召广大科技工作者发扬追求真理、服务国家、造福人民的优良传统，勇担重任，勇攀高峰，为建设世界科技强国而奋斗。因此，“建设”一词作为科技部党组1号文件中常年出现的高频词，在2017年以后又被赋予了新的内涵和外延。

此外，通过观察2017年至2020年科技部党组1号文件对“建设世界科技强国”的描述，可以归纳其发展进程。从2017年的“坚持把建设世界科技强国作为奋斗目标”到2018年的“加快建设创新型国家和世界科技强国”，再从2019年的“系统谋划世界科技强国建设”到2020年的“以强有力的制度优势和治理效能保障我国跻身创新型国家前列和世界科技强国建设”，可见，一路走来，我国建设世界科技强国的目标更加坚定，道路更加清晰，任务更加明确。

3.2 体现时代特点的高频词

与常年出现的高频词不同，一部分高频词只集中出现在某个特定时期的科技部党组1号文件

中。在我国宏观科技创新政策发展的不同历史阶段，这类词具有所处时期的历史烙印，最能体现所处时期的突出特点。提取一段时间内科技部党组1号文件中权重最高的前50个词，并绘制词云图，可以发现党的十九大前后国家宏观科技创新政策的变化趋势。

3.2.1 2013年—2017年

如图6—图10所示，2013年—2017年，“计划”“企业”等词出现频次较高，占据较高权重。

“计划”一词在2013年—2017年的科技部党组1号文件中出现的频次分别为13、11、14、20和22，在体现时代特点的高频词当中具有一定代表性。词频变化趋势反映出两个主要特点，一是2014年—2017年“计划”一词词频逐年递增，且到2017年达到峰值，二是2017年以后，“计划”一词退出高频词行列，不再集中出现。2014年3月，国务院印发《关于改进加强中央财政科研项目资金管理若干意见》，要求实行科研项目分类管理，加强科研项目和资金配置的统筹协调^[12]。同年12月，国务院印发《关于深化中央财政科技计划（专项、基金等）管理改革方案的通知》，明确到2017年，中央财政科技计划全面按照优化整合后的5类科技计划（国家自然科学基金、国家科技重大专项、国家重点研发计划、技术创新



图 6 2013 年高频词词云图



图 7 2014 年高频词词云图



图 8 2015 年高频词词云图



图 9 2016 年高频词词云图



图 10 2017 年高频词词云图

引导专项、基地和人才专项)运行^[13]。上述科技部党组 1 号文件中“计划”一词的词频变化趋势的两个主要特点，刚好反映出科技计划管理改革的发展历程，即 2014 年—2017 年科技计划管理改革逐步深化，2017 年科技计划管理改革顺利完成。

在 2013 年—2017 年科技部党组 1 号文件中，“企业”一词出现的频次较高，分别为 14、20、11、12 和 15。2012 年 9 月，中共中央、国务院印发《关于深化科技体制改革加快国家创新体系建设的意

见》，要求强化企业技术创新主体地位，促进科技与经济紧密结合^[14]。2013 年 1 月，国务院办公厅印发了《关于强化企业技术创新主体地位全面提升企业创新能力的意见》，强调围绕促进科技与经济社会发展紧密结合，统筹发挥市场配置资源的基础性作用和政府的引导支持作用^[15]。此后，“企业”一词便在科技部党组 1 号文件中占据了重要位置。

3.2.2 2018 年—2020 年

如图 11—图 13 所示，2018 年—2020 年，“人才”“支撑”等词出现频次较高，占据较高权重。

“人才”一词自 2018 年开始集中出现在科技部党组 1 号文件中，且保持较高频次。2018 年—2020 年出现的频次分别为 25、13 和 19，内容主要包括培养本土高水平人才和引进国际高端人才等，一方面体现了科技部贯彻落实党的十九大报告中关于“加快建设人才强国”的战略部署，在国家宏观科技创新政策中对科技创新人才更加重视，另一方面表明在国务院机构改革后，科技部



图 11 2018 年高频词词云图

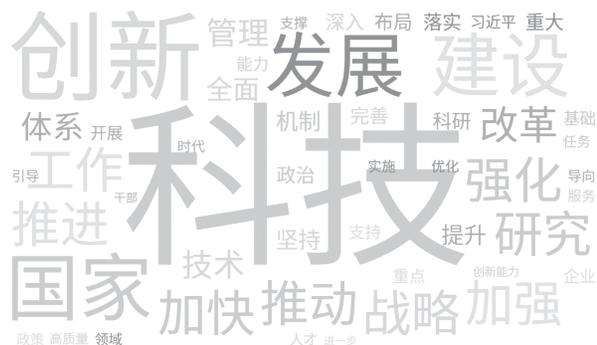


图 12 2019 年高频词词云图



图 13 2020 年高频词词云图

在引进国外智力、拓宽外国人才来华绿色通道等方面表现得更加积极主动。

此外，“支撑”一词在体现时代特点的高频词当中也具有一定代表性。2018年—2020年“支撑”一词出现的频次分别为9、13和14，通过抽取近3年科技部党组1号文件中与“支撑”一词相邻的关键词，“高质量发展”这一表述引起了笔者的关注。

“高质量发展”这一表述最早出现在党的十九大报告中。党的十九大报告指出，我国经济已由高速增长阶段转向高质量发展阶段。因此，围绕推动高质量发展要求，充分发挥科技创新的支撑作用，作为近3年我国科技工作的总体思路被写入科技部党组1号文件。“支撑”一词出现频次逐年增长，一定程度上也体现了科技创新在国家经济社会发展中正在发挥越来越重要的作用。

3.3 近年出现的新词

在近年的科技部党组1号文件中，出现了一些新词。这些词首次出现在近年来的科技部党组1号文件中，重要性较强，一定程度上能反映我国

宏观科技创新政策的发展趋势和主要特点。

“科学家精神”一词最早出现在2019年科技部党组1号文件中。文件明确，引导科研人员弘扬科学家精神，坚守科研伦理和道德，增强使命感，潜心研究、多出高质量原创成果。同年6月，中共中央办公厅、国务院办公厅印发《关于进一步弘扬科学家精神加强作风和学风建设的意见》，着力加强作风和学风建设，鼓励和引导广大科技工作者自觉践行、大力弘扬新时代科学家精神^[6]。此后，“科学家精神”一词在2020年科技部党组1号文件中再次出现，内容主要为重申改进科研作风学风问题。特别的是，“科学家精神”一词并非最早出现在中央层面的文件中，而是由科技部在2019年的党组1号文件中提出，并推动其成为中央层面的文件，进而持续强化落实。

2020年，“治理”一词开始集中出现在科技部党组1号文件中，且一跃成为高频词。通过抽取2020年科技部党组1号文件中与“治理”一词相邻的关键词，发现“体系”和“能力”两词与“治理”一词的关联性较强。2019年10月，党的十九届四中全会审议通过了《关于坚持和完善中国特色社会主义制度、推进国家治理体系和治理能力现代化若干重大问题的决定》，明确完善科技创新体制机制，健全符合科研规律的科技管理体制和政策体系。因此，2020年科技部党组1号文件聚焦治理体系和治理能力，更加注重以治理的思维和方法推进科技创新。

3.4 具有特点的高频词对比

本部分从历年的科技部党组1号文件中，遴

选了两组具有突出特点的高频词，依据近义、反义或相关性分组，结合所处时代背景下科技环境进行对比分析，探索这类词出现频次变化背后的深层次原因。

(1) “管理”与“服务”

“管理”与“服务”是公共行政的两项基本

职能，二者对立统一，相辅相成。如图 14 所示，2013 年以来，“管理”和“服务”两词的词频稳定保持在 5~25 之间，基本呈现“管理”与“服务”并重的特点。

影响“管理”一词词频的因素，除前述“科技计划管理改革”相关表述外，还体现在完善科

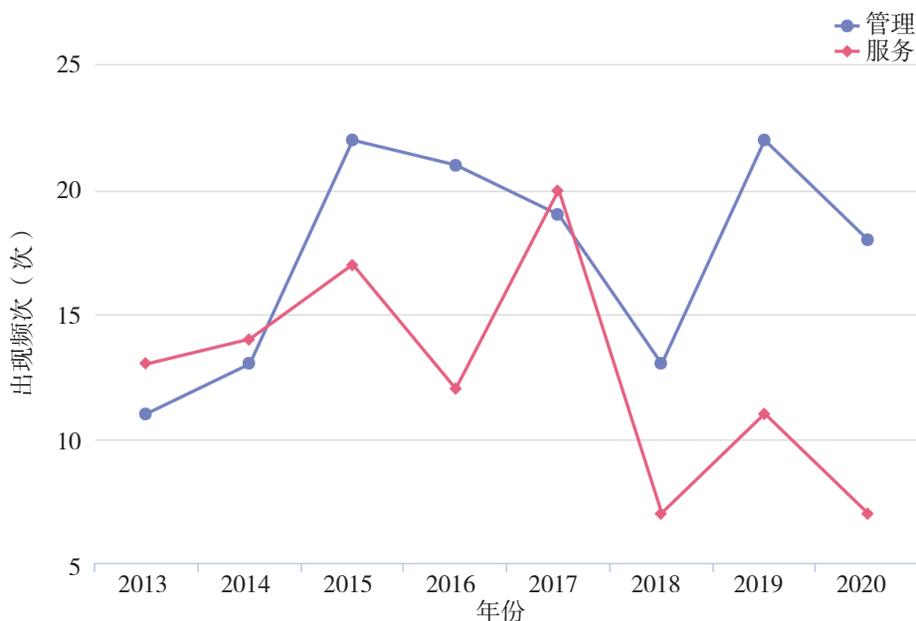


图 14 “管理”和“服务”两词的词频变化趋势

技政策法规体系方面。科技部作为国家科技管理部门，承担着制定并组织实施科技政策法规的职能。2013 年以来，在科技部党组 1 号文件中，明确表示推动出台了一系列管理条例、管理办法，进一步优化了科技创新政策法规环境。

除此之外，“科技管理干部队伍建设”相关表述也是历年科技部党组 1 号文件的重要组成部分。内容主要包括深化理论学习、落实教育培训、强化能力建设、改进工作作风、严格监督管理等多项要求，为国家科技改革发展提供了坚强保障。以上两方面内容，也使得“管理”一词在科技部党组 1 号文件中出现的频次显著增长。

近年来，随着“放管服”改革不断深化，政府职能更多地从“管理”向“服务”转变。2017 年，“抓战略、抓规划、抓政策、抓服务”（以下简称“四抓”）这一表述首次出现在科技部党组 1 号

文件中，并在此后被持续提及。值得一提的是，2017 年和 2018 年提出，按照“四抓”要求，推动政府职能从研发管理向创新服务转变。2019 年提出，突出“四抓”要求，坚持不懈把深化科技体制改革推向深入。2020 年提出，落实“四抓”要求，积极开展科技改革发展的前瞻谋划。由此可见，科技部持续强化政府职能转变，不断提高科技创新管理水平和服务能力，着力推动科技改革发展取得新成效。

(2) “体制”与“机制”

一般认为，“体制”是指社会生活中特定方面事务组织体系和管理制度的具体表现形式，而“机制”是指组织体系内部之间相互作用的过程和方式，二者相互作用。一方面，体制决定机制，是机制发挥作用的前提条件，另一方面，体制依赖与之相适应的机制实现。如图 15 所示，2013 年

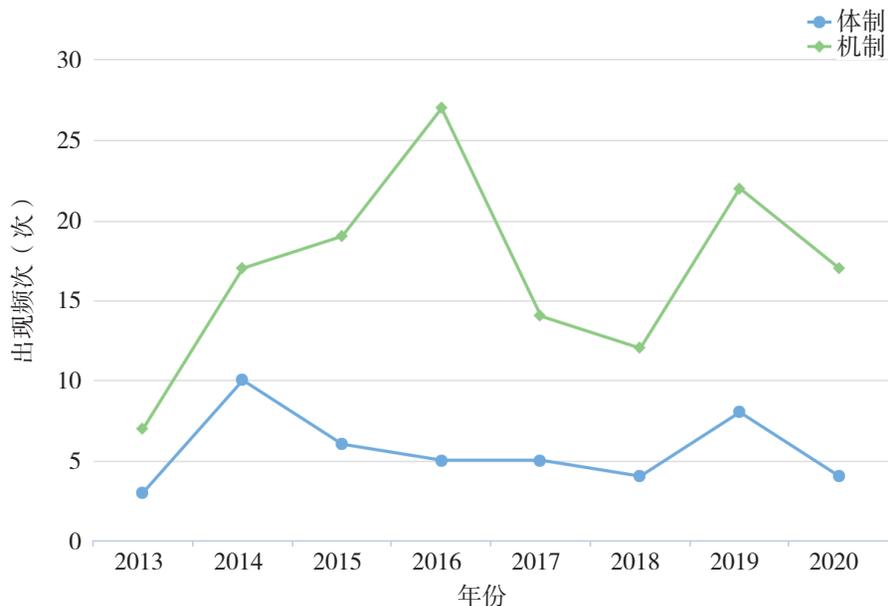


图 15 “体制”和“机制”两词的词频变化趋势

以来，“体制”一词的词频稳定保持在 10 以内，而“机制”一词的词频基本在 10~30 之间，远高于“体制”一词。

2015 年 3 月，中共中央、国务院印发《关于深化体制机制改革加快实施创新驱动发展战略的若干意见》，为科技体制机制改革提供重要遵循^[17]。同年 9 月，中共中央办公厅、国务院办公厅印发《深化科技体制改革实施方案》，旨在形成系统、全面、可持续的改革部署和工作格局，打通科技创新与经济社会发展通道，最大限度地激发科技第一生产力、创新第一动力的巨大潜能。方案包括建立技术创新市场导向机制，改革人才培养、评价和激励机制等多项任务^[18]。此后，科技体制机制改革便成为科技部党组 1 号文件中的重要内容，延续至今。

2016 年 3 月，中共中央印发《关于深化人才发展体制机制改革的意见》，提出了推进人才管理体制、改进人才培养支持机制，健全人才顺畅流动机制等多项改革措施，为形成具有国际竞争力的人才制度优势提供了重要遵循^[19]。随后，2017 年科技部党组 1 号文件提出，贯彻落实《关于深化人才发展体制机制改革的意见》，优化重大人才计划（项目、工程）组织实施机制，推动

形成创新人才蓬勃成长、创新成果竞相迸发的良好局面。

2016 年 7 月，国务院印发《“十三五”国家科技创新规划》，首次提出探索社会主义市场经济条件下科技创新的新型举国体制^[20]。2020 年，科技部党组 1 号文件在重申上述意见的同时，进一步提出凝练总结科技领域集中力量办大事的经验做法，以及研究提出构建新型举国体制的思路和框架。可见，2020 年已经成为构建新型举国体制进程中的重要时间节点，将在推进国家创新体系建设、跻身创新型国家前列和建设世界科技强国等方面发挥着关键作用。

上述“科技体制”“人才管理体制”“新型举国体制”等词反映出“体制”一词更为宏观，侧重于形式，而“机制”一词则显得相对微观，侧重于运行。也正因为如此，在科技部党组 1 号文件中，“机制”一词的频次远高于“体制”一词。2013 年以来，科技部多措并举，连续出台了一系列重要文件，投入更多精力、下更大气力狠抓改革任务落实，把深化科技体制改革推向深入。

4 研究结论

研究发现，国家宏观科技创新政策发展主要

呈现以下几方面特点:

(1) 总的看来,国家宏观科技创新政策紧随党中央、国务院重大决策部署。面向世界科技前沿、面向经济主战场、面向国家重大需求,国家科技管理部门从科技创新供给、科技资源配置和市场价值实现等方面协同发力,着力促进科技与经济紧密结合,推动经济实现质量变革、效率变革和动力变革,为经济高质量发展提供有力支撑。

(2) 党的十八大以来,国家宏观科技创新政策适应国际国内经济社会发展形势的需要,既有对国家层面战略规划延续和深化,也有对特定领域管理机制的细化和补充。当前,国家科技管理部门着力提升科技创新支撑引领作用,不断强化体系建设和能力建设,为我国奋力实现迈进创新型国家行列和建设世界科技强国提供坚实的制度保障。

(3) 党的十九大以来,国家宏观科技创新政策的重点逐渐由国家层面战略规划转向特定领域管理机制。国家科技管理部门落实科技体制机制改革和“放管服”改革等一系列重要举措,大力弘扬科学家精神,改进科研作风学风,完善科技计划管理和引才引智制度环境,着力推进国家科技创新治理体系和治理能力现代化。

此外,受到研究对象和研究方法限制,本研究不能穷尽2013年以来的全部国家宏观科技创新政策,对于科技部党组1号文件中未提及或词频较低的相关表述存在遗漏的可能。因此,本研究仅为相关研究提供量化依据,为科技创新主体解读国家宏观科技创新政策提供新的视角,并不能作为国家宏观科技创新政策落实成效的数据支持。■

参考文献:

- [1] 中共科学技术部党组.关于贯彻落实党的十九届四中全会精神奋力实现迈进创新型国家行列目标的意见[Z].2020-01-08.
- [2] 赵筱媛,苏竣.基于政策工具的公共科技政策分析框架研究[J].科学学研究,2007,25(1):52-36.
- [3] 刘会武,卫刘江,王胜光,等.面向创新政策评价的三维分析框架[J].中国科技论坛,2008(5):33-36.
- [4] 刘会武,王胜光.创新政策系统分析:钻石模型的提出及应用[J].科学管理研究,2009(4):6-9.
- [5] 赵伟,杨宇.创新驱动下科技政策演变趋势:由微观转向宏观[J].经济纵横,2015(16):76-77.
- [6] 刘凤朝,孙玉涛.我国科技政策向创新政策演变的过程、趋势与建议——基于我国289项创新政策的实证分析[J].中国软科学,2007(5):34-42.
- [7] 彭纪生,仲为国,孙文祥.政策测量、政策协同演变与经济绩效:基于创新政策的实证研究[J].管理世界,2008(9):25-36.
- [8] Tanaka K, Sakata I. New science and technology policy evaluation using bibliometric approach[C]. IEEE. 2017 Portland International Conference on Management of Engineering and Technology Portland: 2017: 1-5.
- [9] 张永安,鄯海拓,袁页.科技创新政策与国家创新能力作用机理研究[J].华东经济管理,2019,3(33):104-110.
- [10] 伍蓓,陈劲,王姗姗.科学、技术、创新政策的涵义界定与比较研究[J].科学学与科学技术管理,2007(10):68-74.
- [11] 国务院.关于大力推进大众创业万众创新若干政策措施的意见[EB/OL].[2020-08-19].http://www.gov.cn/zhengce/content/2015-06/16/content_9855.htm,2015-06-16/2020-08-10.
- [12] 国务院.关于改进加强中央财政科研项目 and 资金管理的若干意见[EB/OL].[2020-08-19].http://www.gov.cn/zhengce/content/2014-03/12/content_8711.htm,2014-03-12/2020-08-10.
- [13] 国务院.关于深化中央财政科技计划(专项、基金等)管理改革方案的通知[EB/OL].[2020-08-19].http://www.gov.cn/zhengce/content/2015-01/12/content_9383.htm,2015-01-12/2020-08-10.
- [14] 中共中央,国务院.关于深化科技体制改革加快国家创新体系建设的意见[EB/OL].[2020-08-19].http://www.most.gov.cn/yw/201209/t20120924_96972.htm,2012-09-24/2020-08-10.
- [15] 国务院办公厅.关于强化企业技术创新主体地位全面提升企业创新能力的意见[EB/OL].[2020-08-19].http://www.gov.cn/zwgk/2013-02/04/content_2326419.htm,2013-02-04/2020-08-10.

- [16] 中共中央办公厅, 国务院办公厅. 关于进一步弘扬科学家精神加强作风和学风建设的意见 [EB/OL]. [2020-08-19]. http://www.gov.cn/zhengce/2019-06/11/content_5399239.htm, 2019-06-11/2020-08-10.
- [17] 中共中央, 国务院. 关于深化体制机制改革加快实施创新驱动发展战略的若干意见 [EB/OL]. [2020-08-19]. http://www.most.gov.cn/yw/201503/t20150324_118707.htm, 2019-06-11/2020-08-10.
- [18] 中共中央办公厅, 国务院办公厅. 深化科技体制改革实施方案 [EB/OL]. [2020-08-19]. http://www.gov.cn/guowuyuan/2015-09/24/content_2938314.htm, 2015-09-24/2020-08-10.
- [19] 中共中央. 关于深化人才发展体制机制改革的意见 [EB/OL]. [2020-08-19]. http://www.mohrss.gov.cn/SYrlzyhshbzb/dongtaixinwen/buneyaowen/201603/t20160322_236103.htm, 2016-03-22/2020-08-10.
- [20] 国务院. 关于印发“十三五”国家科技创新规划的通知 [EB/OL]. [2020-08-19]. http://www.gov.cn/zhengce/content/2016-08/08/content_5098072.htm, 2016-08-08/2020-08-10.

Research on Quantitative Evaluation of the Development of National Macro Science and Technology Innovation Policy Based on Automatic Word Segmentation Technology

LIU Xu

(Institute of Science and Technology Information of China, Beijing 100038)

Abstract: The study is based on automatic word segmentation technologies to have quantitative evaluation research on the typical regulation documents issued by national managing bodies of science and technology. Based on word frequency statistics, information retrieval, text mining, and TF-IDF (a very common weighted technology) are adopted to evaluate the importance of a single word on the text as a whole. Then it summarizes the No. 1 Documents of the Ministry of Science and Technology, hence figuring out the high-frequency words seen in past years, the high-frequency words indicative of the characteristics of times, and the high-frequency words newly emerged in recent years. By comparing the high-frequency words with special characteristics, it probes into the development history and current status of national macroscopic technological innovative policies.

Key words: automatic word segmentation technology; science and technology innovation policy; high-frequency words; quantitative evaluation research