

政府科技新闻中科技成果转化的主题识别与时空扩散分析

王曰芬 王柳虹 巴志超 岑咏华 王琦

【摘要】政府科技新闻已成为发布与推广科技政策、传播与扩散科技成果的重要媒介。当前中国政策研究主要关注科技政策文本数据,对科技新闻这一载体在科技政策扩散过程中的作用关注较少。结合科技政策扩散过程的复杂性与科技新闻的时效性、简明性、广泛性等特征,本文试图挖掘并分析政府科技新闻在科技政策扩散过程中的作用特点及规律。首先,利用抽样与数据爬取方法采集并建立科技新闻数据集;其次,利用人工编码和文本挖掘方法识别出科技成果转化主题,并结合政策实施过程要素构建主题关联理论框架;最后,从扩散强度、扩散广度、扩散爆发与扩散分布四个维度研究有关科技成果转化的科技新闻时空扩散过程和特征。科技新闻主题与科技政策主题扩散特征具有一致性;科技新闻政策扩散作用特征凸显,不仅能即时而全面地关注科技成果转化政策落实,而且能简洁而动态地反映科技部与各个省市贯彻落实科技政策的侧重点及其关联事件的影响。同时,本文提出的科技新闻主题识别与时空扩散分析方案具有通用性,能够弥补基于政策文本数据研究政策扩散的不足。

【关键词】科技新闻;科技政策;主题识别;时空扩散;科技成果转化

【作者简介】王曰芬(1963-),女,教授,博士生导师,天津师范大学管理学院(天津 300387),南京理工大学经济管理学院(南京 210094),研究方向为知识服务与数据挖掘,E-mail:yuefen163@163.com;王柳虹(1997-),女,硕士研究生,南京理工大学经济管理学院,研究方向为数据挖掘与知识发现(南京 210094);巴志超(1990-),男,助理教授,研究员,南京大学数据智能与交叉创新实验室,研究方向为知识图谱与知识发现(南京 210023);岑咏华(1979-),男,教授,天津师范大学管理学院,研究方向为知识网络与情报分析(天津 300387);王琦(1992-),男,博士研究生,南京理工大学经济管理学院,研究方向为信息计量与数据科学(南京 210094)。

【原文出处】《情报学报》(京),2023.8.939~951

【基金项目】天津市科技智库项目“天津高质量发展的重点科技指标对比和评价研究”(21ZLZKZF00130);江苏省研究生科研与实践创新计划项目“图书情报领域学术资源平台传播与利用的对比研究”(SJCX21_0169)。

0 引言

在创新驱动发展战略实施过程中,政府部门通过发布科技新闻,及时传播科技政策、推广科技创新活动和推介先进科技成果,以推动科技创新转化为社会生产力。科技新闻不仅是政策宣传与推广的重要手段,而且是反映政策制定、发布、实施过程及其相关活动的重要途径。相比于科技政策文件传播,

科技新闻以叙事为主,简明扼要,在信息传播时效性、公开性、广泛性与变动性等特点上表现更加突出。因此,挖掘科技新闻传播要素与分析所报道的政策关联事件状态及变化是政策扩散研究的重要组成部分,更是作为政策舆情文本研究科技政策对于科技创新发展作用的重要工具和载体。

已有政策扩散研究着重于关注公共政策扩散模

式与机制^[1]、科技政策扩散路径^[2]、政策知识转移与扩散^[3]，并强调通过政策文本案例研究得出具有普遍性结论等。科技新闻传播特点导致信息采集处理的非结构化特征与烦琐程度显著，使得学术界较少深究科技新闻这一载体的科技政策扩散作用。随着科技新闻数字化、网络化与社会化传播的日益普及和文本挖掘技术等技术发展，通过政府网站获取大量科技新闻数据并加工处理的难度大大减少，使得通过对科技新闻的挖掘与分析进而多途径揭示科技政策扩散机理，从中探究推动科技创新活动举措与保障实施的过程和效果成为可能。而科技新闻主题是传播者关注点与传播主要内容的体现，是探究科技政策传播扩散的主要切入点。

综上所述，针对科技新闻数据集，如何识别具有可靠性的主题并建立主题关联理论框架，创新扩散的哪些指标可用于衡量科技新闻传播时空特征及其变化，不同层级或地区科技新闻报道具有怎样的主题差异性，科技新闻主题布局与科技政策主题布局的对比程度如何，这些问题的探究与解决对于丰富政策扩散理论与拓展信息传播知识体系具有重要学

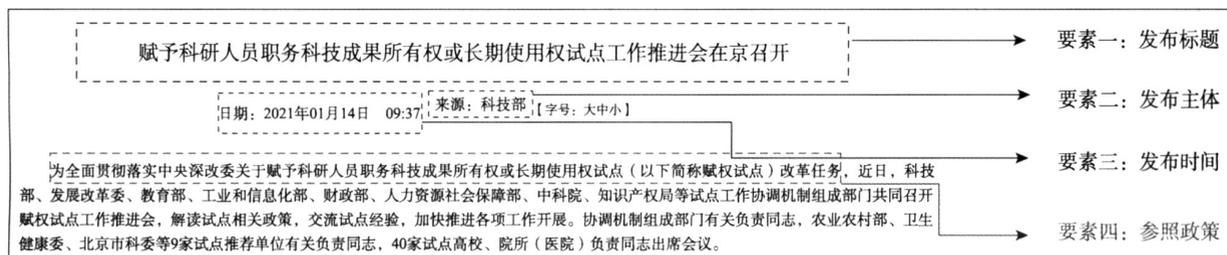
术价值，对于提升政府科技新闻传播影响、促进国家政策实施效果和提升信息资源管理功能具有重要实践意义。因此，本文尝试以政府网站科技新闻作为数据来源，将扎根理论编码、文本挖掘与可视化等方法结合，构建科技新闻主题识别与时空扩散分析的研究框架和指标算法，针对上述问题展开探索，为学术研究与实践应用奠定基础。

1 相关研究工作

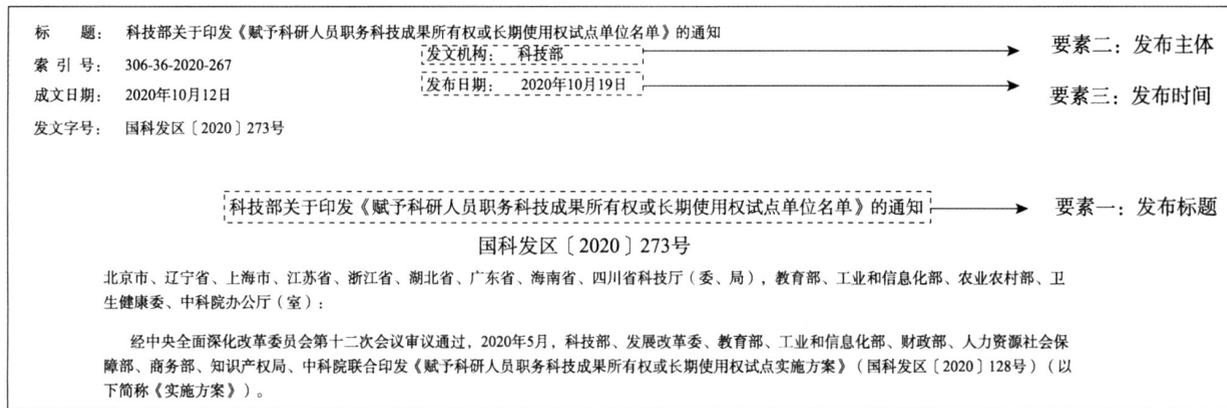
1.1 科技新闻与科技政策

科技新闻文本涵盖政策采纳主体、社会新闻媒体与社会公众等多类主体，是积极传播科技政策、普及科技成果的重要媒介。政策文本是政策法规及条文存在的物理载体^[4]，包括政策原文本、政策执行文本、政策宣传文本、政策分析文本与政策阅读文本等^[5]。从传播学角度看，科技新闻与科技政策都具有发布标题、发布主体、发布时间、发布内容等基本传播要素；同时，科技新闻包含参照科技政策要素，将科技新闻与科技政策形成关联。科技新闻文本与科技政策文本的对照如图1所示。

从图1可发现，在表达形式上，科技新闻文本与



a. 科技新闻文本



b. 科技政策文本

图1 科技新闻与科技政策的传播要素对比图

科技政策文本构成的基本特征大致相同,这为通过文本挖掘方法提取科技新闻主题特征并开展时空扩散分析奠定研究基础;在表达内容上,科技新闻文本不仅包含对政策发布基本情况的及时报道与政策要点的重点解读,而且涵盖对政策实施过程与效果的追踪反映,这为持续而全面研究政策扩散相关联的主要事件或关键影响因素等提供可行性;在表达效果上,科技新闻文本相对篇幅短小且易于被公众接受与理解,这使得借助科技新闻的功能深化政策扩散机理分析并推动科技创新活动发展具有先决条件。

1.2 政策主题识别与演化分析

政策主题识别与演化规律是政策扩散研究的重点议题。政策主题识别是政策演化分析的基础,相关学者们主要采用人工编码^[6]、主题概率模型^[7]、神经网络^[8]等方法识别政策文本主题。陈慧茹等^[9]引入扎根理论编码方法选取科技创新政策文本的关键词;Benites-Lazaro等^[10]结合LDA(latent Dirichlet allocation)主题建模和大数据计算方法识别巴西报纸、政府文件等关键主题,研究巴西乙醇生产相关的社会政策话语变化及其与气候变化和食品安全的关系。然而,单纯依靠人工编码容易导致客观性缺失,而现有智能主题识别方法用于政策文本分析存在局限,难以考虑到政策文本语境变化的影响^[11]。

政策文献计量是一种量化分析政策文本结构属性的研究方法,通过获取可复现、可验证结果从宏观层面明晰政策演进规律、影响范围及发展趋势^[12]。黄萃等^[13]采用社会网络分析研究1978-2013年科技政策文献,分析不同阶段政府部门合作关系的整体网特征、局部网特征和单节点子网特征;段尧清等^[14]采用社会网络分析和主题分析法,基于政策间参照网络分析政务大数据政策扩散时间特征、空间特征和主题特征。然而,国内研究主要关注政策原文本政策演化特点与规律,受政策文本载体限制,难以实现对政策实施过程、相关影响及反馈的充分探索。

1.3 政策扩散测度与时空扩散规律

自20世纪60年代政策扩散研究引发学者的关注以来^[15],政策扩散测度与时空扩散规律一直是国内外学者探索的重要话题。早期Brown等^[16]提出政

策扩散在时间维度上呈S曲线,空间维度上表现为“邻近效应”,以及区域内出现“领导者-追随者”的层级效应。主要的政策扩散测量方法或指标有Walker的政策扩散指数^[15]、Savage的政策创新指数^[17],以及政策承继、吸收与扩散质量^[18]。借鉴知识扩散,张剑等^[19]提出公共政策扩散的文献量化研究维度与方法,并以中国科技成果转化政策文本为例,从强度、广度、速度与方向四个指标研究科技成果转化政策扩散过程和特点。然而,鲜有研究结合科技新闻文本特征多维度描述和分析政策扩散过程。

综上,现有研究大多关注政策原文本的政策扩散研究,而缺乏对科技新闻政策扩散的关注。科技新闻可全方位而动态地反映科技政策的传播、扩散及实施过程,在文本表达形式、内容与效果上,具有研究政策扩散的基础、可行性与先决条件。因此,探索科技新闻主题政策扩散的特点与演化规律,可为追踪科技政策主题演化和路径提供更多样化的研究方法支持与更丰富的数据资料参考。

2 研究设计

2.1 研究思路与分析框架

借助已有研究成果并结合科技新闻特征,本文提出如下研究思路:首先,选择具有代表性的政府网站科技工作动态栏目等科技新闻作为数据来源,采用目的性与理论性抽样相结合方式,爬取原始数据并建立科技新闻数据集;其次,将扎根理论编码与文本挖掘相结合,处理与归并汇集原始数据,将通过饱和度和检验而涌现的科技成果转化主题范畴作为研究对象,建立科技成果转化新闻数据库,识别核心主题并建立关联,提取主题特征要素;再其次,借助创新扩散理论,从扩散强度、广度、爆发、分布四个维度进行指标界定与算法设计及测量;最后,借助科学计量与可视化方法,从主题维度扩散状态、主题时序与空间演化、科技新闻主题与科技政策主题布局对比四个方面进行时空扩散分析。研究框架如图2所示。

2.2 科技新闻主题识别方法与流程

为尽可能准确地识别科技新闻主题,本文综合运用扎根理论编码与文本挖掘技术。其中,扎根理论方法包括实质性编码和理论性编码^[20],所设计的识别方法与流程如下。

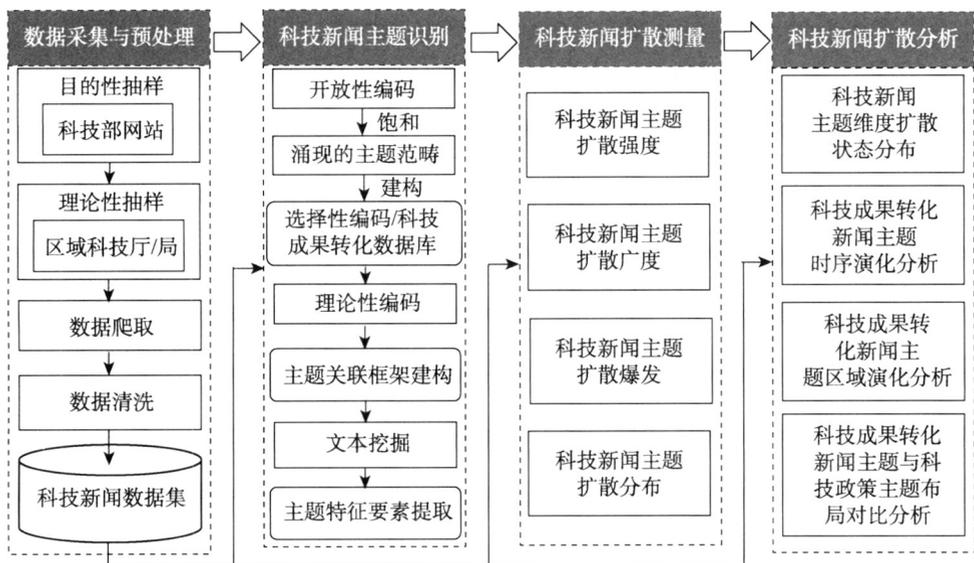


图2 科技新闻主题识别与扩散分析的研究框架

(1) 实质性编码

实质性编码包括开放性编码和选择性编码,通过不断比较数据以概念化和抽象化数据。开放性编码阶段将数据资料逐句编码、逐层概念化并形成参考点,逐级提升概念化程度后形成概括性范畴;选择性编码阶段对范畴产生足够关联的数据进行编码,选取频繁重现、与其他数据及其属性相关联的范畴支撑核心主题的饱和和检验。

(2) 理论性编码

在实质性编码基础上,组织核心主题以构建理论框架。概念化核心范畴间隐含的关联以实现理论的饱和,如并列、因果和递进等,聚集研究过程中形成的思想、概念和范畴,形成完整的理论。若发现理论无法饱和,则需要追溯整个研究过程再次选择性编码,或者重新进行理论性抽样,重新补充数据,以实现理论的饱和。此外,对初步构建理论与已有研究进行不断比较,可发现和补充已有概念、范畴及理论的不足。当与现有研究文献的比较不能产生新的概念或范畴时,理论达到饱和,理论框架构建工作完成。

(3) 文本挖掘

文本挖掘方法主要用于科技新闻的采集与清洗、数据集与关系矩阵构建、主题词频统计与分布计量等,以提取理论框架中科技成果转化新闻主题要

素并建立关联。首先,采集与存储科技新闻案例的发布信息、文本内容,并构建科技新闻案例数据集;进而提取案例中发布机构或信息来源、发布时间等发布信息,构建科技新闻发布信息的文档数据集;同时,根据科技新闻中“贯彻落实……”“根据……”等提示词,提取案例的参照政策要素,建立科技新闻-科技政策的关系矩阵。其次,统计扎根分析获得的参考点与核心主题,进而构建科技新闻案例-核心主题-参考点关系矩阵。最后,提取并统计科技新闻核心主题案例数量、发布主体与时间、参考点等要素,识别出科技成果转化新闻主题并构建数据库。

2.3 科技新闻主题扩散测量方法

科技新闻扩散与学科知识扩散相类似,借鉴创新扩散理论^[21]与已有研究^[19,22],分析科技新闻主题传播扩散过程中的强度特征、结构特征、变化趋势及其联系^[22]。从传播扩散视角看,科技新闻主题扩散的频次和关注覆盖面体现了传播扩散过程中的强度、结构特征;同时,科技新闻主题扩散是随时间发展的进程^[21],科技新闻主题扩散爆发可测度主题随时间变化的趋势;此外,科技新闻主题扩散分布可分析不同科技新闻发布主体在科技新闻主题扩散过程中的联系、关注主题侧重点的相似度^[23]。因此,本文从频次、广度、时间、分布四个维度对科技新闻主题扩散的特征和过程开展研究,所设计测度指标阐述如下。

(1) 科技新闻主题扩散强度

科技新闻主题扩散强度(S&T News Topic Diffusion Intensity, STNTDI)是通过科技新闻主题扩散频次来描述主题扩散,以揭示科技新闻关注点数量及强度。某个主题被提及案例数量越多,强度越大。科技新闻扩散强度分为绝对强度(Absolute S&T News Topic Diffusion Intensity, ASTNTDI)和相对强度(Relative S&T News Topic Diffusion Intensity, RSTNTDI)。其中,绝对强度为科技新闻样本集中单个主题的累计案例数 r_i ,相对强度为绝对强度 r_i 与科技新闻样本集案例数总和 D 的比值:

$$ASTNTDI=r_i \quad (1)$$

$$RSTNTDI=r_i/D \quad (2)$$

(2) 科技新闻主题扩散广度

科技新闻主题扩散广度(S&T News Topic Diffusion Breadth, STNTDB)是从覆盖范围的角度描述主题扩散,以揭示科技新闻关注点的覆盖面。某个主题在案例中被提及次数越多,则该主题扩散广度越大。扩散广度分为绝对广度(Absolute S&T News Topic Diffusion Breadth, ASTNTDB)和相对广度(Relative S&T News Topic Diffusion Breadth, RSTNTDB)。其中,单个科技新闻主题的绝对广度为单个科技新闻主题的累计新闻参考点数 r_i ,相对广度为科技新闻主题参考点数与科技新闻样本集主题参考点数总和 T 的比值:

$$ASTNTDB=r_i \quad (3)$$

$$RSTNTDB=r_i/T \quad (4)$$

(3) 科技新闻主题扩散爆发

科技新闻主题扩散爆发(S&T News Topic Diffusion Burst Index, STNTDBI)是从时间角度描述科技新闻主题扩散状态及演化,以揭示科技新闻关注主题的时序性聚集分布及其变化程度。若关注主题爆发次数较多,爆发时间较长,则该主题具有较高与较持续的扩散突变性。借鉴 Kleinberg^[24]的“bursty and hierarchical structure in streams”,使用爆发检测技术检测科技新闻关注的核心主题何时处于爆发状态及爆发程度。科技新闻发布月份为时间点 t ,每个时间点 t 的主题总数用 d 表示,目标核心主题用 r 表示。 R 是每个时间点目标核心主题数总和, D 是每个时间点所有主题数总和。每个时间点目标主题的基线比率定义为

$$P_0=\frac{R}{D} \quad (5)$$

同时,爆发状态比率 P_1 为基线比率乘以常数 s_0 。目标核心主题爆发状态的预期比率定义为

$$P_1=s_0 \times P_0 \quad (6)$$

核心主题在给定时间内状态的转换,取决于拟合优度 σ 和转换成本 τ 。主题在某时间点 t 的实际比率与预期比率之间拟合优度用 σ 表示,状态转换难度用 τ 表示。其中, i 表示状态, $i=0$,表示处于基线状态, $i=1$,表示处于爆发状态; n 是时间点数量; γ 是过渡到更高状态的难度,转换到较低状态或保持相同状态时成本为零。总成本函数为 σ 和 τ 的总和,定义为

$$C(i, r_t, d_t)=\sigma+\tau=-\ln\left[\left(\frac{d_t}{r_t}\right)p_i^{r_t}(1-p_i)^{d_t-r_t}\right]+(i_{next}-i_{prev})\times \gamma \times \ln(n) \quad (7)$$

基于成本函数,选择 $t=n$ 时成本最小的状态与从 $t=n$ 转换到 $t=m$ 时成本最低的状态,最后找到最小化总成本的状态序列。由此,科技新闻主题爆发强度函数(从 t_n 开始至 t_m 结束)定义为

$$STNTDBI(\text{weight})=\sum_{i=t_n}^{t_m}(C(0, r_i, d_i)-C(1, r_i, d_i)) \quad (8)$$

(4) 科技新闻主题扩散分布

科技新闻主题扩散分布(S&T News Topic Diffusion Distribution, STNTDD)从空间角度描述科技新闻主题的空间分布及与邻近区域的差异性,以揭示科技新闻主题扩散的区域分布特征。不同区域的科技新闻主题分布指标越大,区域间主题分布越相似。不同区域用 a, b 表示,目标科技新闻主题总数用 R 表示,采用 Jaccard 系数计算主题分布的相似性:

$$STNTDD(a, b)=\frac{|R_a \cap R_b|}{|R_a \cup R_b|} = \frac{|R_a \cap R_b|}{|R_a| + |R_b| - |R_a \cup R_b|} \quad (9)$$

3 实证分析

3.1 数据来源、编码处理与数据分布

(1) 数据来源

科技部属于新闻发布国家级层级,视为信息流的上行层;而京津冀、长三角经济圈是我国科技发展代表性区域,其省市级科技厅/局的新闻发布属于省市级层级并视为下行层,而省市间为同行层。由此,确定科技部为目的性样本,京津冀、长三角区域

为理论性样本。从相应网站上针对科技要闻与科技动态等栏目采集数据并预处理,筛选贯彻落实科技创新方针政策相关工作的重要新闻,并建立科技新闻数据集。数据采集时间范围为2017年10月1日至2021年12月31日,获取关于科技创新的新闻共4486篇。

(2)编码处理

对采集数据开放性编码,经过饱和度检验,发现“科技成果转化”“区域技术转移”范畴初步涌现。根据研究中形成的范畴进行理论性抽样,清洗和筛选

已涌现的主题范畴,得到科技成果转化新闻案例共1214篇,案例数据包含案例标题、发布时间、机构与内容等信息。按层级、年份分组得到科技成果转化新闻数据集,主题基本信息如表1所示。

(3)数据分布特征

为整体了解科技新闻发布状态,对科技成果转化新闻进行案例数与国家/区域数的时间统计,如图3和图4所示,案例数在时间分布上每一年的态势与数量相接近。其中,2020年新闻的数值达到302篇,是历年来最大值。在国家/区域分布上,科技部科技新

表1 科技成果转化主题的基本信息

层级	科技部门	案例数量(篇)	网址
国家	科技部	154	https://www.most.gov.cn/kjbgz/
京津冀区域	北京	150	http://kw.beijing.gov.cn/col/col6382/index.html
	天津	141	http://kxjs.tj.gov.cn/XWDT4045/BMXW1075/
	河北	183	https://kjt.hebei.gov.cn/eportal/ui?pageId=91573
长三角区域	上海	137	https://stcsn.sh.gov.cn/xwzx/gzdt/
	江苏	186	http://std.jiangsu.gov.cn/col/col82536/index.html
	浙江	136	http://kjt.zj.gov.cn/col/col1228971342/index.html
	安徽	127	http://kjt.ah.gov.cn/kjzx/gzdt/index.html

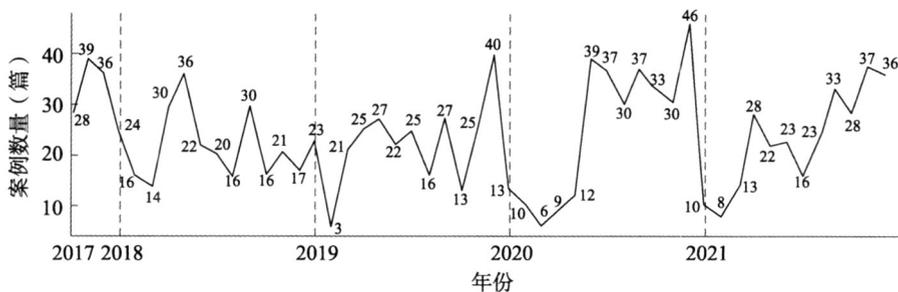


图3 科技成果转化新闻案例数的时间分布

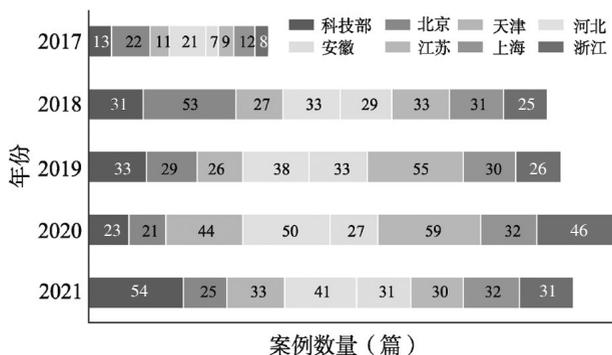


图4 科技成果转化新闻案例所属国家/区域的时间分布(彩图请见 <https://qbx.istic.ac.cn/>)

闻数量于2021年达到峰值,北京市科技新闻在2018年激增并达到最高峰,天津市、江苏省、河北省、浙江省在2020年达到峰值,上海市从2018年开始保持稳定的新闻数量。

3.2 科技新闻主题识别、关联框架构建与特征提取

第一,开放性编码,对科技新闻案例逐级编码形成参考点,如科研机构、科技创新联盟等。对初步涌现的科技成果转化案例数据进一步编码并经过饱和和梯度检验,共获得16个范畴,包括产学研合作、创新创业和梯度技术转移等。

第二,选择性编码,发现产学研合作、创新创业等范畴支持核心主题高校与科研机构并判定饱和,梯度技术转移、科技扶贫等范畴支持核心主题区域技术转移并判定饱和。科技成果转化中9个核心主题获得5级参考点或范畴支持判定饱和,包括区域技术转移、科技成果产业、科技企业、科技成果转化服

务人员、高校与科研机构、科技创新平台、科技成果市场、科技需求对接和科技成果推介应用。

第三,理论性编码,参考《中华人民共和国促进科技成果转化法》《国家科技成果转化引导基金管理暂行办法》和相关科技政策研究,并按照“需求-供给-服务”思路,总结核心主题之间关联,凝练三个维度并构建理论框架,如图5所示。以反映科技成果转化新闻三类维度下的主题关联,为后续的科技新闻主题特征提取和时空分析提供依据。

第四,采用文本挖掘技术提取科技新闻核心主题的参照政策、案例及参考点等特征要素并加以统计。所得科技新闻核心主题特征如表2所示。

3.3 科技成果转化新闻主题时空扩散分析

3.3.1 科技新闻主题扩散状态分布

主题识别为主题扩散分析提供锚点。依据科技成果转化新闻主题关联框架,结合扩散强度和广度测量指标分析不同维度主题侧重点和扩散状态。

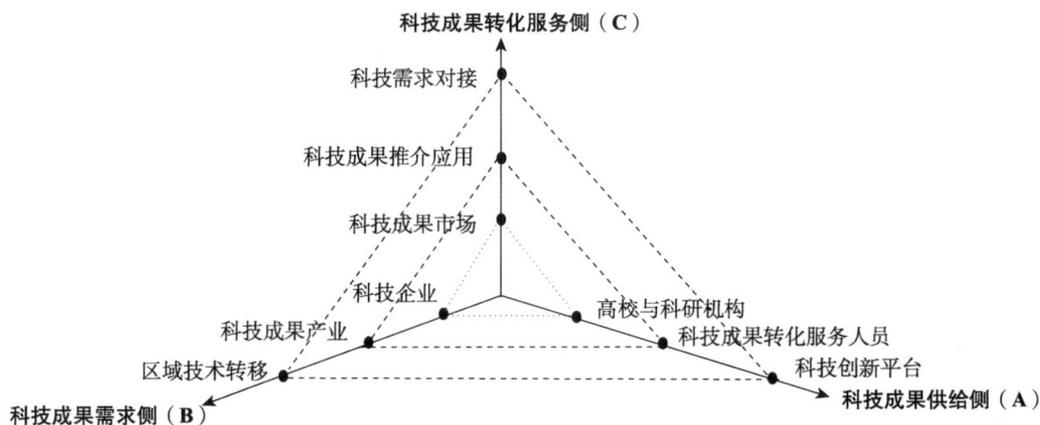


图5 科技成果转化新闻主题关联框架

表2

科技成果转化新闻核心主题特征

科技新闻核心主题	案例数	参考点数	参照政策(部分)
高校与科研机构(a1)	182	744	《关于举办第七届中国国际“互联网+”大学生创新创业大赛的通知》
科技成果转化服务人员(a2)	133	570	《国家技术转移专业人员能力等级培训大纲》
科技创新平台(a3)	283	1224	《关于推进国家技术创新中心建设的总体方案(暂行)》
科技企业(b1)	163	646	《关于科技企业孵化器大学科技园和众创空间税收政策的通知》
科技成果产业(b2)	193	791	《关于开展重大科技成果产业化专题债有关工作的通知》
区域技术转移(b3)	184	973	《关于进一步加强东西部扶贫协作工作的指导意见》
科技需求对接(c1)	225	730	《关于科技扶贫精准脱贫的实施意见》
科技成果转化推介应用(c2)	278	1117	《关于举办2020年全国科技活动周的通知》
科技成果市场(c3)	305	1232	《关于技术市场发展的若干意见》

(1)不同维度主题扩散广度和强度计算所得结果如表3所示,从中可知不同维度或三个维度上核心主题被关注的侧重点及其差异。

由表3可知,在科技成果供给侧中,科技创新平台在本维度中扩散广度和扩散强度都高且在三个维度中相对广度和相对强度最高;高校与科研机构和科技成果转化服务人员相比,绝对广度比较高而相对广度和相对强度要弱一些,高校与科研机构作为创新的主要供应者总体上在新闻中受到更多关注。在科技成果需求侧中,区域技术转移的绝对广度和绝对强度都比较高,科技企业相对广度在本维度最高且在三个维度中处于次高序列,科技成果产业的绝对广度居中而相对广度、绝对强度和相对强度都是最弱的。在科技成果转化服务侧中,科技成果市场在本维度的扩散广度和扩散强度都高且在三个维度中的绝对广度最高,说明科技成果市场深受关注且被提及的范围最广,其次是科技成果推介应用,最弱的是科技需求对接。

(2)不同维度主题按照时间序列的相对扩散广度

和强度的计算结果如图6所示,图中呈现了科技新闻主题的时序性分布特征。从整体上看,9个核心主题除了2017年外总体分布相对比较均匀,可见这些核心主题在科技成果转化中一直受到高频关注,并都有较广的覆盖面。例如,科技成果供给侧的科技创新平台(a3)主题在2018-2020年逐渐上升,且在2020年呈现较显著的聚集分布态势,扩散相对广度达到峰值0.3108,扩散相对强度亦达到峰值0.2816;科技成果转化服务人员(a2)和科技成果推介应用(c2)在2020年处于高数值,科技企业(b1)和科技需求对接(c1)在2021年上升为峰值。这些核心主题的分布变化,与2020年10月发布的第十四个五年规划中所提出的“布局建设综合性国家科学中心和区域性创新高地”“促进各类创新要素向企业集聚”等政策要点相切合。

3.3.2 科技新闻主题扩散时序演化

从时序维度分析核心主题的扩散爆发,以揭示科技新闻主题扩散时序特征。

(1)科技新闻主题扩散爆发检测

主题爆发检测可识别科技成果转化政策的主题

表3 科技成果转化主题扩散的广度与强度

科技新闻维度	科技新闻核心主题	绝对广度	相对广度	绝对强度	相对强度
科技成果供给侧(A)	高校与科研机构(a1)	724	0.0810	225	0.1154
	科技成果转化服务人员(a2)	571	0.0938	275	0.1411
	科技创新平台(a3)	1266	0.1795	293	0.1503
科技成果需求侧(B)	科技企业(b1)	646	0.1490	188	0.0965
	科技成果产业(b2)	798	0.0690	146	0.0749
	区域技术转移(b3)	973	0.1437	273	0.1401
科技成果转化服务侧(C)	科技需求对接(c1)	790	0.0787	160	0.0821
	科技成果推介应用(c2)	1203	0.0920	183	0.0939
	科技成果市场(c3)	1552	0.1132	206	0.1057

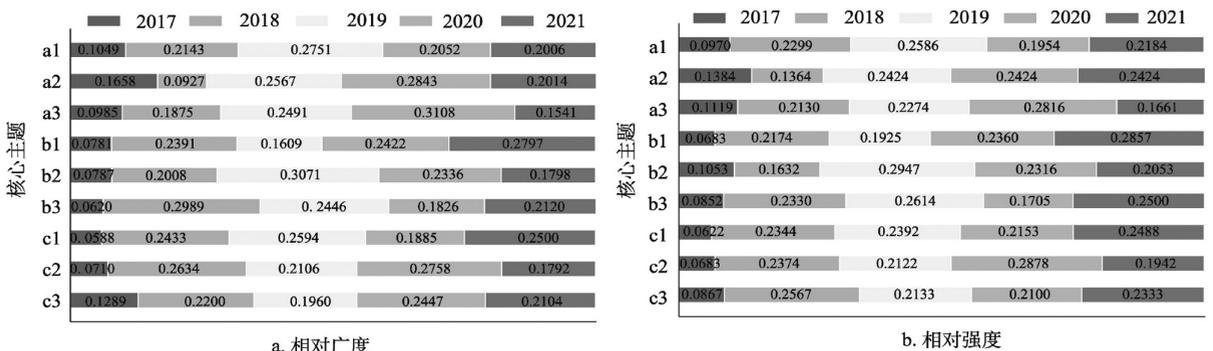


图6 科技成果转化主题扩散的相对广度和相对强度的时序分布(彩图请见 <https://qxb.istic.ac.cn/>)

转变和新趋势。利用上文提出的STNTDBI算法检测主题时序演化及主题爆发状态,设 $s=2, \gamma=0.05$ 。其中,科技部的科技新闻主题出现12个爆发状态,京津冀区域出现12个爆发状态,长三角区域出现15个爆发状态,位列Top 5的结果如表4所示。其中c3的单次爆发强度达到峰值5.825585,进入爆发状态共9次。由此可见,c3表示的科技成果市场在科技成果转

化新闻中多次被掀起高潮并多次有新趋向的出现。

(2)科技新闻主题爆发周期分析

为进一步识别主题爆发周期,绘制整个时间段内爆发主题的标准化比率,值等于1表示等于基线比率,小于1(浅色)表示比率小于基线,大于1(深色)则表示大于基线,在科技部、京津冀区域、长三角区域各部分主题的爆发周期性分布结果如图7所示。

表4 科技成果转化主题扩散爆发检测结果(Top 5)

科技部		京津冀区域		长三角区域	
爆发点	爆发强度	爆发点	爆发强度	爆发点	爆发强度
a1	4.542378	c2	5.526063	c3	5.825585
a3	3.906900	b2	5.122160	b2	5.367751
b1	2.690763	a2	2.977340	b3	5.083172
c3	1.513124	b1	2.288622	c2	4.043153
b2	1.403196	c3	2.236538	c1	2.762978

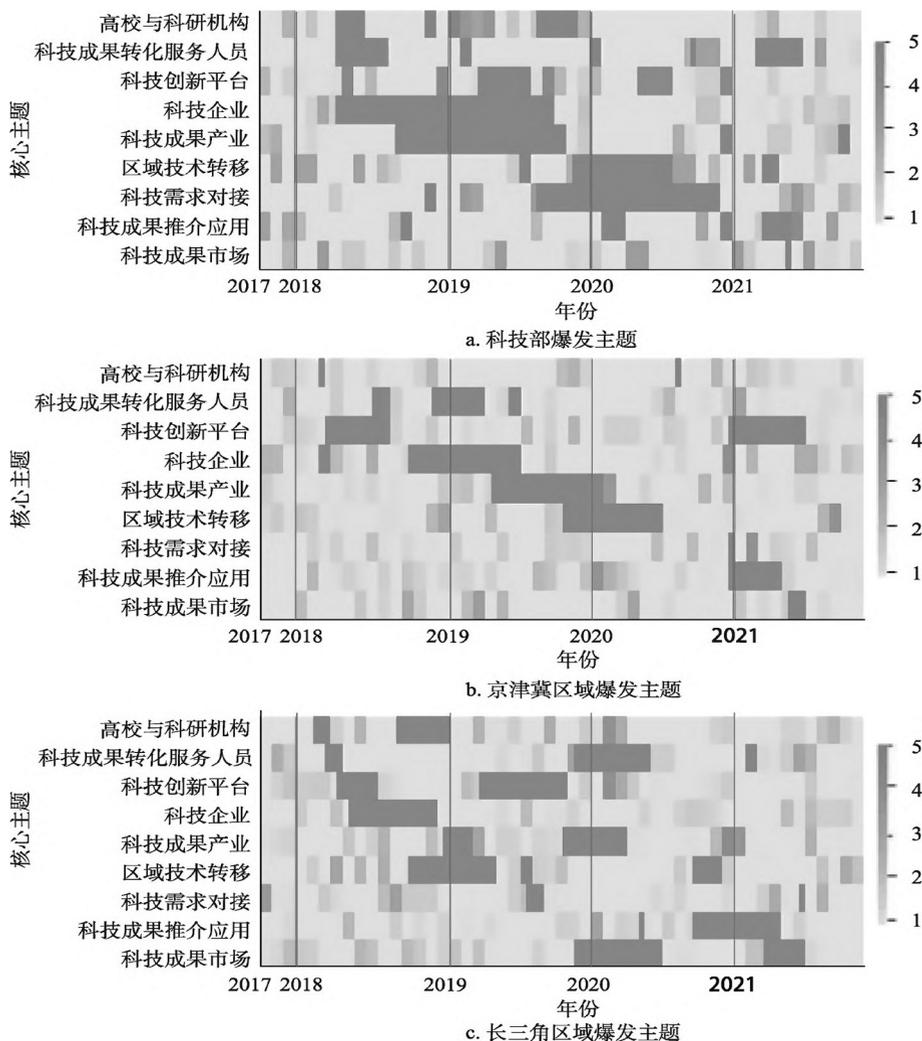


图7 科技部与各区域爆发的主题及其周期性分布(爆发强度 > 1)

由图7可以发现,主题爆发周期呈现四种状态:其一,整体的每一个部分都出现高强度连续的爆发,如科技企业、区域技术转移;其二,整体的每一个部分都出现多次爆发,有高有低,不连续,如科技创新平台、高校与科研机构等;其三,某个部分出现多次高强度的爆发,如科技成果市场整体出现次数多、分布较零散,而在长三角区域于2020-2021年2次连续高强度爆发,科技需求对接整体出现次数少而在科技部出现2次高强度爆发,且1次跨越2019-2020年延续较长时间;其四,不同部分同一时间段爆发重合度较高,如科技企业和科技成果产业、区域技术转移和科技需求对接在科技部中分别于2018-2019年、2019-2020年有较高重合度,科技创新平台和科技成果推介应用在京津冀区域于2021年有较高重合度,高校与机构和科技成果转化服务人员、科技成果产业、科技创新平台、科技成果市场在长三角区域于2020年有较高重合度。上述状态表明科技新闻主题爆发状态变化是多样的,但是在不同空间都会有相关主题发生着共同的变化。

(3)主题爆发与科技政策关联

结合科技新闻所参照的政策,分析主题爆发与科技政策的关联,以科技创新平台与科技企业主题为例,具体如下。

首先,科技创新平台频繁爆发且持续时间稳定,在科技部样本多次爆发,在2019年、2020年各有持续的高强度爆发,而该时间段内恰好科技部、财政部于2019年6月发布了《国家科技资源共享服务平台优化调整名单》;在京津冀区域也是多次爆发,且在2018年、2021年各有持续的高强度爆发,对应时间段内,2018年6月北京市科学技术委员会印发《北京市科技创新基地培育与发展工程专项管理办法》,2021年1月国务院新闻办公室召开“推进北京市国际科技创新中心建设发布会”;在长三角区域亦多次爆发,持续高强度爆发时间集中在2018年、2019年,对应时间段内,2018年10月,科技部与上海市在沪举行2018年部市工作会商会议,专题研究加快推进上海具有全球影响力的科技创新中心建设;2019年3月,上海发布《关于进一步深化科技体制机制改革增强科技创新中心策源能力的意见》。这表明科技创新

平台主题爆发充分反映了相关政策发布落实情况及其实施过程中的关联事件或活动。

其次,科技企业主题爆发始于2018年京津冀样本,随后在京津冀、科技部样本持续爆发到2019年,后来断续较弱地爆发,2020年年底在京津冀出现过短暂强度爆发。结合表3科技企业扩散的绝对扩散值较低而相对扩散值较高以及图6相对扩散值在2021年达到峰值,科技企业和科技成果产业扩散广度并不高。这表明一方面,2019年9月科技部支持科技型中小企业创新发展的政策措施、2020年科技创新支撑科技企业复工复产措施、2021年出台促进各类创新要素向企业集聚等举措在科技新闻中多次得到充分反映;另一方面,由于当时的新冠疫情、国际科技竞争加剧等环境变化,科技企业创新发展受到较大影响。

3.3.3 科技新闻主题扩散空间演化

从空间维度分析核心主题发布频次及其变化,以揭示科技新闻主题扩散空间特征。科技新闻主题的国家级和区域省市级分布表明政策知识、信息等在国家与区域、区域间、区域内流动状态。

(1)科技新闻主题扩散空间分布相似性

采用科技新闻扩散分布指标,分别计算9个核心主题在科技部与7个省市的空间分布相似性,如图8所示。从上下层级的信息流看,科技部与7个省市主题分布相似性都不高,在0.41~0.54。从区域内同层级的信息流看,长三角区域4个省市的相似性在0.61~0.76,其中上海与浙江、上海与江苏以及江苏

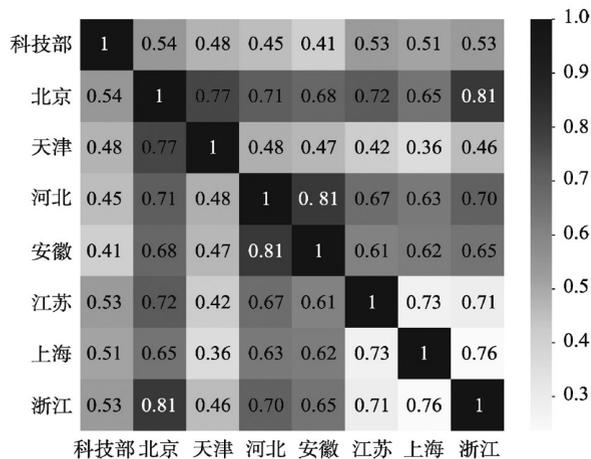


图8 科技新闻主题扩散分布在空间上的相似性

与浙江都高于0.70,而安徽与其他3个省市都低于0.70且相差幅度不大。京津冀区域3个省市的相似性在0.48~0.77,相差幅度比长三角大,而天津、河北分别与北京的相似性远高于两者之间的相似性。从区域间同层级的信息流看,北京与其他6个省市的相似性在0.65~0.81。而除天津外,省市间的相似性在0.61~0.81,尤其是浙江与北京、河北与安徽相似性达到0.81。由此可见,上下层级主题分布相似性低于同层级,说明地方政府在科技新闻报道时保持地方关注点同时彼此相互借鉴;长三角区域城市间相似性较高,而跨区域间浙江与北京、河北与安徽所关注主题的接近程度最高。

(2)科技新闻主题扩散空间分布状态及其可视化

按照“主题维度-主题发布-核心主题”思路,设计主题扩散空间分布状态可视化图谱,如图9所示。左侧第一列表示维度分布,右侧第一列表示核心主题,线条粗细与主题扩散强度成正比。结合主题扩散空间分布相似性,可表达在空间上科技新闻重点主题及占比情况:①北京、上海、浙江重点关注科技创新平台,以倡导依托科技创新平台积极研发产业关键技术;②江苏重点关注科技成果市场,以倡导落实技术要素市场化配置;③天津、河北、安徽重点关注科技需求对接和区域技术转移,反映“京津研发、河北转化”协同创新模式,以及长三角技术市场一体化发展的倡导。这表明各个省市在科技成果转化上

都各有关注与反映的侧重点。

3.4 科技新闻与科技政策主题布局对比分析

为呈现国家与区域、区域间、区域内侧重点差异是否与相应的科技政策相关,对比分析科技新闻与科技政策的主题布局,并采用基于权重的Jaccard系数^[25]进行科技新闻与科技政策的主题分布相似度测算,进而分析验证科技新闻主题扩散分布的可靠性。以国家、北京和上海政策主题对应科技部、北京和上海的科技新闻为例,根据2021年6月北京市促进科技成果转化议事协调联席会议办公室发布的《国家及各省市促进科技成果转化政策汇编(目录)》,包括国家政策176条、北京政策94条、上海政策23条。按照9个核心主题分别对国家及区域的科技政策进行人工编码、分类,科技新闻与科技政策的主题分布映射与对比分析如图10所示。

科技部科技新闻与国家科技政策在供给侧、服务侧分布总体相似度较高。其中,需求侧的科技成果产业基本一致,而科技企业、区域技术转移相似度稍低;北京市三侧分布不均衡较显著,供给侧的主题相似度达到0.92,而服务侧的只有0.33;上海市三侧分布较均衡,最高的是供给侧,最低的是需求侧。同时,供给侧的主题相似度都很高,但是国家相似度最高的是高校与科研机构,北京与上海都是科技成果转化服务人员。进一步结合不同层面政策或者新闻条款及主题相似度,可见在国家层面上呈现高度重

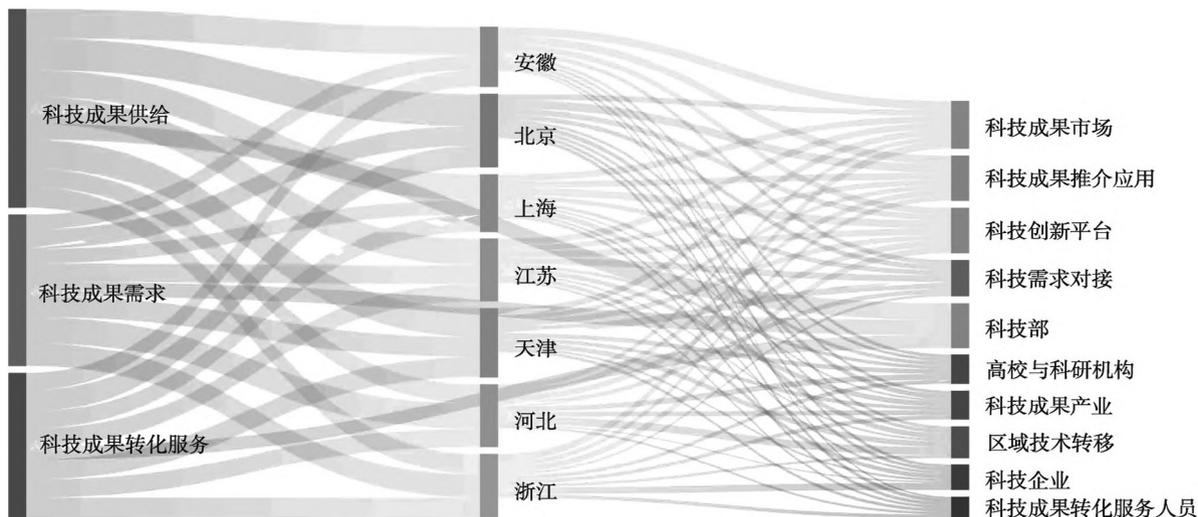


图9 科技成果转化主题扩散的空间分布状态可视化图谱



图 10 科技成果转化政策与科技成果转化新闻主题分布对比
框图中的数值表示政策或新闻的条数,箭头中间的数值表示两两之间的主题相似度。

视科技成果转化和保持新闻报道与政策发布三个侧面主题较高一致性的特点,在区域层面上呈现新闻报道与政策发布对于所侧重主题具有较高一致性,且同时通过差异性表达地方举措的特点。总体来说,科技新闻与科技政策的主题分布总体相似度较高,分布基本一致,但不同层级、城市的科技新闻侧重点有所不同。

4 结论与讨论

为深入探究科技新闻政策扩散的作用,本文围绕如何从科技新闻中提取主题并建立关联、如何选择并界定创新扩散指标测度主题扩散、如何分析主题时空扩散的分布特点等问题,将扎根理论、文本挖掘、科学计量分析、创新扩散理论等理论与方法结合,构建具有可行性与通用性的科技新闻主题识别及其结构关联的方案,构造科技新闻主题扩散的强度、广度、爆发、分布测量指标。同时,针对涌现并通过饱和度检验的科技成果转化新闻主题范畴进行实证研究,分析主题时空扩散状态、主题扩散时序演化与空间演化,并对比分析科技新闻与科技政策在主题布局上的关联,证实基于科技新闻的主题扩散测度是科技政策扩散研究的补充性指标与方法,揭示出科技新闻政策扩散作用具有及时性、动态性、覆盖范围广、关联内容多等特点。

相关研究仍存在诸多值得探索之处,未来建议加强不同类型科技新闻扩散的差异、科技新闻与科技政策的映射与解析方式、科技新闻对于政策传播扩散的多样化表征等方面的研究,并考虑进一步结合新闻传播、政策研究等相关领域的研究成果进行交叉融合与扩展,更进一步地提升扩散指标的准确性与科学性。

参考文献:

[1]王浦劬,赖先进.中国公共政策扩散的模式与机制分析[J].北京大学学报(哲学社会科学版),2013,50(6):14-23.

[2]许乾坤,刘耀.科技政策隐性扩散路径自组织研究[J].情报资料工作,2022,43(1):61-70.

[3]Newmark A J. An integrated approach to policy transfer and diffusion[J]. Review of Policy Research, 2002, 19(2): 151-178.

[4]黄萃,任弢,张剑.政策文献量化研究:公共政策研究的新方向[J].公共管理学报,2015,12(2):129-137,158-159.

[5]刘小年.中国农民工政策研究[M].长沙:湖南人民出版社,2007.

[6]Theis D R Z, White M. Is obesity policy in England fit for purpose? Analysis of government strategies and policies, 1992-2020[J]. The Milbank Quarterly, 2021, 99(1): 126-170.

[7]王芳,徐路路.基于智能化公文主题分析的我国政策层级扩散倾向性研究[J].情报学报,2021,40(4):387-401.

[8]Dominguez-Morales J P, Jimenez-Fernandez A F, Dominguez-Morales M J, et al. Deep neural networks for the recognition and classification of heart murmurs using neuromorphic auditory sensors[J]. IEEE Transactions on Biomedical Circuits and Systems, 2018, 12(1): 24-34.

[9]陈慧茹,肖相泽,冯锋.科技创新政策加权共词网络研究——基于扎根理论与政策测量[J].科学学研究,2016,34(12):1769-1776.

[10]Benites-Lazaro L L, Giatti L, Giarolla A. Topic modeling method for analyzing social actor discourses on climate change, energy and food security[J]. Energy Research & Social Science, 2018, 45: 318-330.

[11]Törnberg A, Törnberg P. Muslims in social media discourse: combining topic modeling and critical discourse analysis [J]. Discourse, Context & Media, 2016, 13: 132-142.

[12]李江,刘源浩,黄萃,等.用文献计量研究重塑政策文本数据分析——政策文献计量的起源、迁移与方法创新[J].公共管理学报,2015,12(2):138-144,159.

[13]黄萃,任弢,李江,等.责任与利益:基于政策文献量化分析的中国科技创新政策府际合作关系演进研究[J].管理世界,2015(12):68-81.

[14]段尧清,尚婷,周密.我国政务大数据政策扩散特征与主题分析[J].图书情报工作,2020,64(13):133-139.

[15]Walker J L. The diffusion of innovations among the American states[J]. The American Political Science Review, 1969, 63(3): 880-899.

[16]Brown L A, Cox K R. Empirical regularities in the diffusion of innovation[J]. Annals of the Association of American Geographers, 1971, 61(3): 551-559.

[17]Savage R L. Policy innovativeness as a trait of American states[J]. The Journal of Politics, 1978, 40(1): 212-224.

[18]施茜,裴雷,李向举,等.信息化政策理论与实践的交互扩散研究——以江浙信息化政策样本为例[J].情报学报,2016,35(10):1081-1089.

[19]张剑,黄萃,叶选挺,等.中国公共政策扩散的文献量化研究——以科技成果转化政策为例[J].中国软科学,2016(2):145-155.

[20]吴毅,吴刚,马颂歌.扎根理论的起源、流派与应用方法述评——基于工作场所学习的案例分析[J].远程教育杂志,2016,35(3):32-41.

[21]E.M. 罗杰斯.创新的扩散[M].唐兴通,郑常青,张延臣,译.北京:电子工业出版社,2016:128-165.

[22]邱均平,瞿辉,罗力.基于期刊引证关系的学科知识扩

散计量研究——以我国“图书馆、情报、档案学”为例[J].情报科学,2012,30(4):481-485,491.

[23]陈启明,王效岳,白如江,等.多源数据融合下突发公共事件社会关注与政策趋向研究——以新冠肺炎疫情为例[J].情报探索,2022(6):15-25.

[24]Kleinberg J. Bursty and hierarchical structure in streams [J]. Data Mining and Knowledge Discovery, 2003, 7(4): 373-397.

[25]潘磊,雷钰丽,王崇骏,等.基于权重的Jaccard相似度度量的实体识别方法[J].北京交通大学学报,2009,33(6):141-145.

Spatiotemporal Diffusion Analysis of Achievement Transformation Topics in Government Science and Technology News

Wang Yuefen Wang Lihong Ba Zhichao Cen Yonghua Wang Qi

Abstract: Government news on science and technology (Sci. & Tech. News) has become an important medium for publishing, promoting, disseminating, and diffusing scientific and technological policies. Currently, Chinese policy research focuses primarily on the policy itself and less on the role of the news in analyzing policy diffusion. Combining the complexity of the policy diffusion process and characteristics of Sci. & Tech. News, such as timeliness, conciseness, and extensiveness, this study attempts to explore the role of the government in Sci. & Tech. News diffusion, analyzing and characterizing the laws governing the policy diffusion process. First, sample and data crawling methods were used to collect and establish a database of Sci. & Tech. News. Subsequently, using manual coding and text mining methods, topics of science and technology achievement transformation were identified and combined with the elements of policy implementation to build a theoretical framework of topic association. Finally, the processes and characteristics of the spatial and temporal diffusion of Sci. & Tech. News on achievement transformation were studied considering the four diffusion dimensions of intensity, breadth, explosiveness, and distributiveness. It was determined that Sci. & Tech. News and policy topics display consistent proliferation characteristics. The role of Sci. & Tech. News in policy diffusion is significant, not only in terms of the immediate and comprehensive implementation of science and technology achievement transformation policies, but also in terms of reflecting the focus of the Ministry of Science and Technology on the concise and dynamic implementation of science and technology policies across provinces and cities. The proposed analysis is a generic scheme for the identification of spatial and temporal diffusion of Sci. & Tech. News topics, and the research was conducted to compensate for the shortcomings of policy diffusion studies that are based on policy text.

Key words: Science and technology news; Science and technology policy; Topic identification; Spatiotemporal diffusion; Science and technology achievement transformation