

基于新闻媒体的福建省科技创新大数据分析

陈健^{1,2}, 林洪^{1,2}

(1. 福建省科学技术信息研究所, 福州 350003; 2. 福建省信息网络重点实验室, 福州 350003)

摘要: 新闻媒体作为信息传播的重要渠道,反映了科技创新的动态。通过“新浪舆情通”平台分析了2023年1月—2024年6月期间福建省各地(市、区)的科技创新相关新闻报道。结果显示,报道数量与当地经济水平显著相关,厦门、福州和泉州分别位于前三名。此外,福州的数据波动性最大,各地(市、区)在不同媒体类型的占比也略有不同。时间序列分析表明,报道高峰通常与政策发布相关,并且报道量与高校和重点实验室数量存在一定的相关性。

关键词: 科技创新; 新闻媒体; 福建省; 大数据分析

中图分类号: G350 **文献标志码:** A **文章编号:** 1671-1807(2024)24-0302-06

当前,科技正迅猛发展,成为经济社会进步的核心驱动力,关乎企业兴衰与国家竞争力。因此,研究科技前沿并构建高效的创新情报机制对科学决策至关重要^[1]。尽管传统的专利分析等方法虽能提供一定洞察,但它们难以揭示创新与社会的深层联系。新闻媒体,作为信息流通的高速公路,不仅加速了科技创新成果的传播,还承载了丰富的创新资讯,直观反映了社会对新技术趋势的广泛关注与热议焦点。面对海量的新闻报道,利用文本挖掘和人工智能技术可以高效地识别和提取科技创新信息,洞察发展趋势,把握发展方向,为创新研究带来新的机遇^[2]。本文聚焦福建省内的科技创新媒体信息,旨在通过大数据分析来探索在地域分布、时间演变等多个维度上的创新规律与特征,力求描绘出科技创新活动的全景图谱。

1 相关研究工作

1.1 科技创新活动的社会认知与科技舆论

科技创新活动,源于社会需求,服务于社会发展,是推动人类文明进步的关键力量。新媒体时代,科技创新活动在海量数据的背景下成为社会关注的焦点。科技大数据不仅记录着科技创新的历程,也成为舆论热议的对象。因此,科技创新活动的社会认知与科技舆论研究成为研究的热点。随着互联网时代的到来,科技活动的相关信息得以通过自媒体进行传播,使得公众及媒体对科技活动的认知与宣传成为当前研究热点。祝娜和王芳^[3]基于

Twitter用户的真实传播数据,通过内容分析法将用户划分为七种类型,并可视化分析科技创新领域传播网络,从信息交互视角探索不同类型用户的传播行为特征;周茜和陈收^[4]通过媒体报道的情绪对媒体信息进行了区分,同时考虑了相关影响因素,对公众媒体对金融科技创新和监管的积极作用和消极影响进行了系统分析。

1.2 科技创新内容主题识别

关于创新主题内容识别的相关研究早已展开,祝娜等^[2]提出一种基于LDA(latent Dirichlet allocation)的科技创新主题语义识别方法,根据表征科技创新内容的关键词语义角色对应的上位词概率来识别科技创新主题。曹树金和闫颂^[5]以情报学期刊的科技论文全文为研究对象,通过BERT(bidirectional encoder representations from transformers)预训练模型识别出全文中的创新句,这些创新内容的提取主要源于科技文献。相较于科技文献,科技新闻在保持科学严谨和知识性的同时,也兼顾了内容的易读性和通俗性,是大众了解科技发展、认识科技前沿的重要途径之一。王曰芬等^[6]利用人工编码和文本挖掘方法识别科技新闻中科技成果转化主题,分析科技新闻的时空扩散的程度和特征;孔嘉等^[7]利用LDA模型,从主题类型和主题强度分析中、澳、英、美四国关于极地地区的科技新闻。

2 基于新闻媒体的科技创新活动

科技创新是在科学与技术日益融合的背景下

收稿日期: 2024-07-22

基金项目: 福建省科技计划项目(2023R1008001)

作者简介: 陈健(1971—),女,福建龙岩人,高级工程师,研究方向为电子政务;林洪(1998—),女,福建福州人,硕士,研究方向为遥感应用与科技信息可视化。

提出来的,它是一个将科学发现和技术发明成果应用于生产实践,创造新价值的过程^[8]。

2.1 科技创新活动在新闻媒体中的表达/表现

在搜索与科技创新相关的新闻媒体报道中,首先会选用“科技”“创新”“科技创新”或者一些与之表义相同的近义词来检索^[9],除此之外,还有一些关键词,如“最新成果”“最新进展”“新发现”或“新突破”等,在新闻媒体报道中都可以用来呈现科技创新的内容。同时新闻媒体也会运用简洁醒目的标题、图片、视频等方式,结合创新内容的不同特点和所在领域,选用不同的词汇,让科技创新的亮点内容更容易脱颖而出。例如,对于一项基础理论研究的重大突破,新闻媒体可能会使用“原创性理论”“首创新观点”“创新思路”等词语来强调其学术创新价值;而对于一项应用技术的新发明或新工艺,新闻媒体可能会用“首创新技术”“独特新方法”等词汇,凸显其创新实用价值。

2.2 新闻媒体数据来源及方法

科技媒体报道数据来源于“新浪舆情通”舆情监测平台,该平台汇集了海量的网络数据资源,能够及时捕捉互联网上出现的各类热点事件和舆论信息。针对“科技”“创新”“科技创新”这些核心关键词,结合“福建”“福州”“厦门”等省内各地市限定词,对该平台在2023年1月—2024年6月时间段内进行数据检索和挖掘,因平潭综合实验区(以下简称“平潭”)较为特殊,故平潭的数据将单独计算和呈现。所获取的数据种类丰富多样,包括微博、微信、新闻客户端、视频、网站、互动论坛和数字报纸等多个维度,涵盖从社交媒体到传统媒体的广泛信息来源。例如,微博和微信作为代表性的社交媒体平台,能够直接反映公众对科技创新话题的真实关注和讨论热度;新闻客户端则集中了各大媒体发布的权威报道和深度分析文章,这些内容不仅提供科技创新领域的最新动态和重大突破,还包含专家观点和政策解读;网站板块囊括政府机构、科研院所、企业官网等各类单位对科技创新的官方宣传;互动论坛和数字报纸则提供更多元化的观点和深度讨论。在数据处理过程中,使用文本挖掘和可视化分析的方法,后续将结合可视化分析的方法,尝试从多个角度对科技创新舆论进行深入解析,以揭示数据背后的趋势和规律。

3 从新闻媒体看福建省及地级市科技创新

3.1 地域差别

基于“新浪舆情通”监测平台,对福建地区“科

技创新”相关新闻媒体报道进行调查统计,结果显示,各地级市在报道数量上呈现明显的梯度分布(图1)。从各地级市的报道总量来看,厦门、福州和泉州位居前三,成为报道的热点地区。紧随其后的是宁德,虽然排名第四,但与前三名仍有一定差距。莆田、南平和龙岩的报道数量相近,处于中游偏下位置,而平潭和三明的报道量大致相当,排名最后。另外,全省的平均值约在129 724条(图1中红色虚线),数量最少的平潭与数量最多的厦门之间的差距达到10倍。这一数据分布反映福建省内在科技创新报道方面存在显著的区域差异且两极分化严重,主要体现在两个方面:一是厦门、福州、泉州三市与其他地(市、区)之间的明显差距;二是这三个领先城市之间的差距也远大于其他地级市之间的差距。

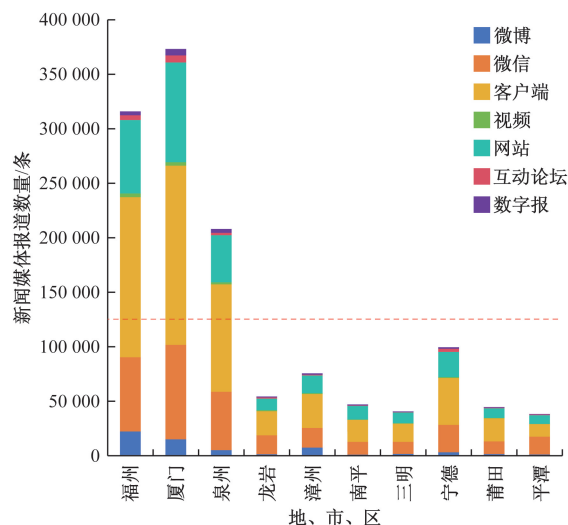


图1 各地(市、区)各类新闻媒体报道数量统计分布情况

为了进一步了解不同地(市、区)之间的数据整体分布特征,包括数据的集中趋势、离散程度、对称性以及异常值情况,绘制了10个地(市、区)的箱线图并进一步统计其均值和标准差(图2)。从整体来看,各地(市、区)的数据都呈正态分布,在数据离散程度方面,福州的数据显示最高的离散程度,这可能反映该市在研究期间经历了较大的波动或变化。紧随其后的是厦门,同样表现出较高的数据波动性。这两个城市的高离散度可能源于它们作为省内主要经济中心,更容易受到外部因素的影响。相比之下,三明的数据表现出最高的聚集程度,其次是龙岩,这种高度聚集的特征表明这些地(市、区)在研究期间保持了相对稳定的状态或者受到的外部影响较小。在异常值方面,箱线图清晰地展示

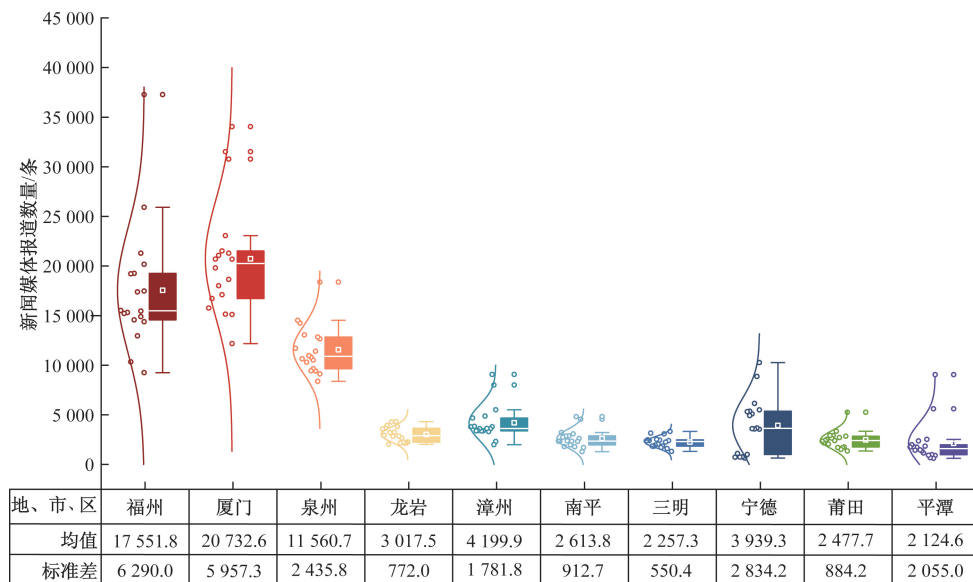


图2 各地(市、区)各类新闻报道数据分布情况箱线

超出上下极限的数据点,这些异常值的出现可能与特定的事件或政策变化有关,在后续的时序分析中将进一步探讨。

从整体结果中发现各地(市、区)在新闻媒体报道数量上的排名可能与其经济实力相关,通过对比2023年各地区的GDP总量、人均GDP和媒体报道数量排名更加验证了此想法(表1)。总体而言,GDP总量排名与新闻报道数量排名的相关性似乎高于人均GDP排名。然而,无论是GDP总量还是人均GDP,排名前三的地区都与新闻报道数量排名前三一致,即福州、厦门和泉州。对于其他地区,经济指标排名与媒体报道排名大体相近,通常相差不超过一两个位次。特别引人注目的是宁德市的表现,尽管其GDP总量排名第五,人均GDP排名第六,但在新闻媒体报道量中却跃居第四,且明显超出其他地(市、区),形成了一定的差距,这表明虽然经济因素可能影响媒体关注度,但并非唯一决定因素。探究其因,宁德时代新能源科技股份有限公司作为宁德市的明星企业和中国最大的动力电池制造商,对宁德市的科技创新确实做出了重要贡献,公司的重大技术突破、新产品发布等事件往往成为媒体关注的焦点,为其新闻报道提供丰富的素材,间接推动了宁德市整体科技创新环境的报道。

新闻媒体类型方面,客户端和网站是报道量最多的渠道,而视频、数字报和互动论坛则位居末尾(图3),不同地(市、区)在媒体类型的选择上也各具特色(图3),比如龙岩在视频这部分占比最高,为1.7%;漳州与“科技创新”相关的报道在微博这类是

表1 2023年福建省各地(市、区)GDP总量排名情况

地区 (按总量排名)	2023年GDP 总量/亿元	地区/ (按人均排名)	2023年人均 GDP/元	新闻媒体 报道数量 排名
福州	12 928.47	福州	153 036	厦门
泉州	12 172.33	厦门	151 969	福州
福建	54 355.10	泉州	137 091	泉州
厦门	8 066.49	福建	129 788	福建
漳州	5 728.43	三明	122 489	宁德
宁德	3 807.33	龙岩	122 163	漳州
龙岩	3 317.96	宁德	120 638	龙岩
三明	3 007.10	漳州	113 031	南平
莆田	3 070.73	平潭	97 163	莆田
南平	2 270.00	莆田	95 990	三明
平潭	370.19	南平	85 628	平潭

注:数据源自福建省统计局。

最多的,占比为9.9%;南平在网站和数字报这部分占比最高,分别为26.6%和2.0%;宁德在互动论坛这部分占比最高,为2.9%;莆田则在客户端这类占比最高,为47.5%;最后,平潭在微信这块的占比显著高于其他地市,约为42.8%。这种报道分布的不均衡反映各地区在科技创新实力、媒体关注度以及信息传播能力等方面的差异。

3.2 时序变化

为探究福建省内与“科技创新”相关的新闻媒体报道数量随月份变化的时间趋势,以福建全省检索结果为例,结果显示,2023年1月—2024年6月全媒体报道量呈现明显的波动趋势(图4)。统计期间的报道量高峰在2024年5月,达98 739条,这一现象可能是由于在5月福州举办了数字峰会,同时这两年的3月都是媒体报道的高峰期,这一激增可归

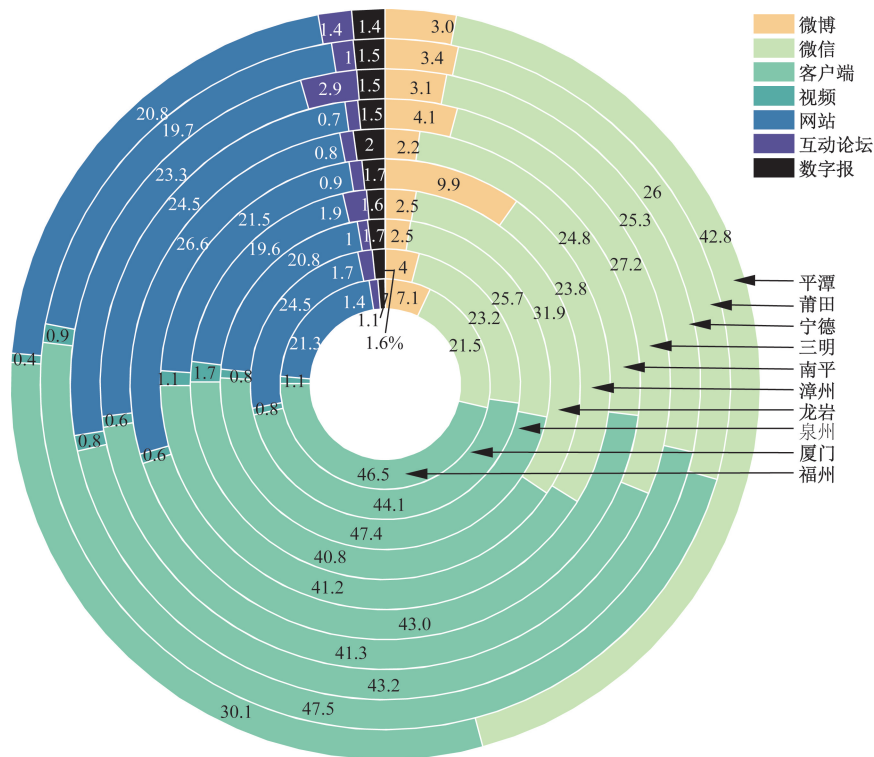


图 3 福建省各地(市、区)不同类型新闻媒体报道数量占比情况

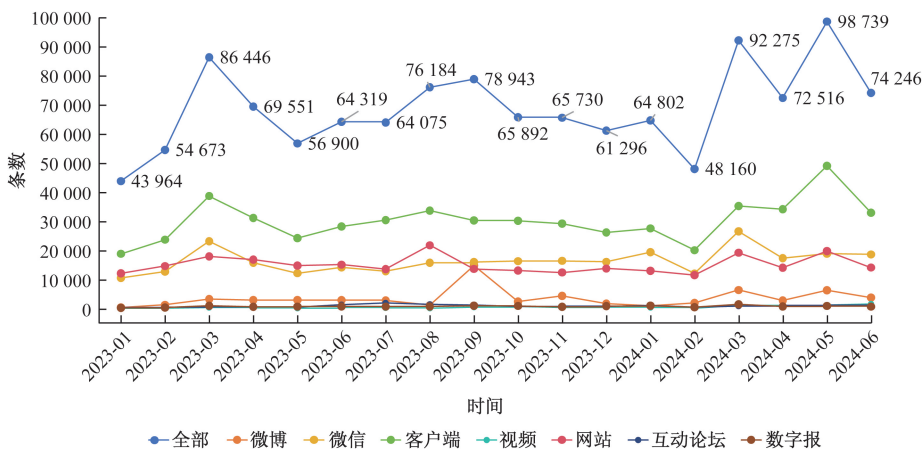


图 4 福建省 2023 全年“科技创新”相关各类新闻媒体检索情况

因于几个关键因素:首先,《党和国家机构改革方案》在 2023 年 3 月发布,其中涉及重要的科技领域改革,包括成立中央科技委员会和重组科学技术部,旨在加强党中央对科技工作的领导^[10];其次,每年 3 月举行的两会也进一步推动了相关报道的增加。相比之下,2023 年 1 月的报道量是统计期间内最低的,仅有 43 964 条。这种现象可能源于两个原因:一是根据历年数据,年初报道量普遍较少;二是 2023 年初仍受疫情影响,可能抑制了相关报道的产生。

为进一步深入研究福建省内各地级市的详细情况,对各地(市、区)的具体情况进一步分析,图 5 中的厦门、福州、泉州、宁德和平潭较其他地级市有

明显的峰值,用虚线强调表示。福州在 2024 年的 5 月达到统计期内的峰值,且明显远高于其他地(市、区),在这段时间内,第七届数字中国建设峰会在福州举办,本届峰会以“释放数据要素价值,发展新质生产力”为主题。平潭的峰值出现在 2023 年 9 月,对应时间段内,国务院发布了《中共中央 国务院关于支持福建探索海峡两岸融合发展新路建设两岸融合发展示范区的意见》,据统计,与该文件相关的报道占了当月所检索数据量的 80% 以上;泉州的高峰出现在 7 月,主要原因是福厦高铁泉州台商投资区段,CR450 动车组进行了一项重要的技术试验,而厦门和宁德的高峰也都与当时颁布的政策有关。

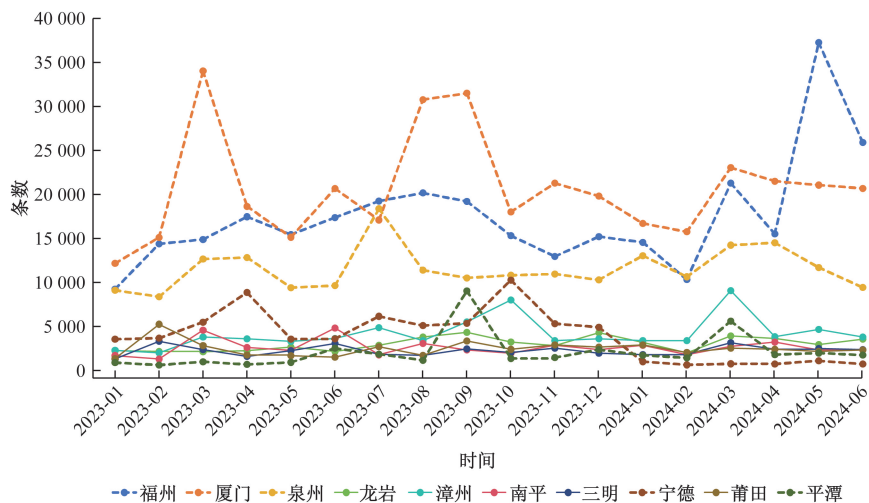


图5 福建省各地(市、区)2023 全年“科技创新”相关各类新闻媒体检索情况

3.3 与高校和省重点实验室数量的关系

高校是知识创新和人才培养的重要基地,在科技创新中扮演着核心角色。它们拥有多学科交叉的优势和丰富的研究资源,能够进行基础研究和应用研究的有机结合。高校不仅培养了大量高素质的创新人才,也通过产学研合作将科研成果转化为现实生产力。省级科技创新平台是区域科技创新体系的核心组成部分,它们聚焦于特定领域的前沿科学问题和关键技术难题。省重点实验室作为省级创新平台的典型代表,集中了优质科研资源,汇聚了领域内的顶尖人才和先进设备,能够开展高水平的科学研究和技术攻关。通过统计省重点实验室(以下简称“省重”)和各高校的数量,发现这二者的数量与“科技创新”的报道数量存在很大的相关性(图6)。福州、厦门和泉州在省重和高校数量上占据优势,尤其是福州表现突出,福州拥有133所高校和35家省级重点实验室,数量上超过厦门两倍以上,稳居首位。除了这三市和漳州,其他地级市的相关数量较为接近,这意味着拥有更多创新平台和教育资源的地区往往能够产生更多的创新成果,媒体关注和相关报道也更多;此外,福州和厦门也是省内唯二坐拥985和211的地市,在2023年度的省重考评结果为优秀的27家实验室中,其中26家皆来自福州和厦门。以厦门大学为例,作为福建省唯一一所985高校,拥有28家省级以上的重点实验室,在科技创新活动方面的报道数量占厦门市的19.87%,超过了除福州和泉州其他7个设区市的报道数量,约占全省总量的11%。

4 总结与展望

4.1 总结

通过“新浪舆情通”平台,获取了全网在2023年

1月—2024年6月期间福建省各地级市与“科技创新”相关的新闻媒体报道,并得出以下结论。

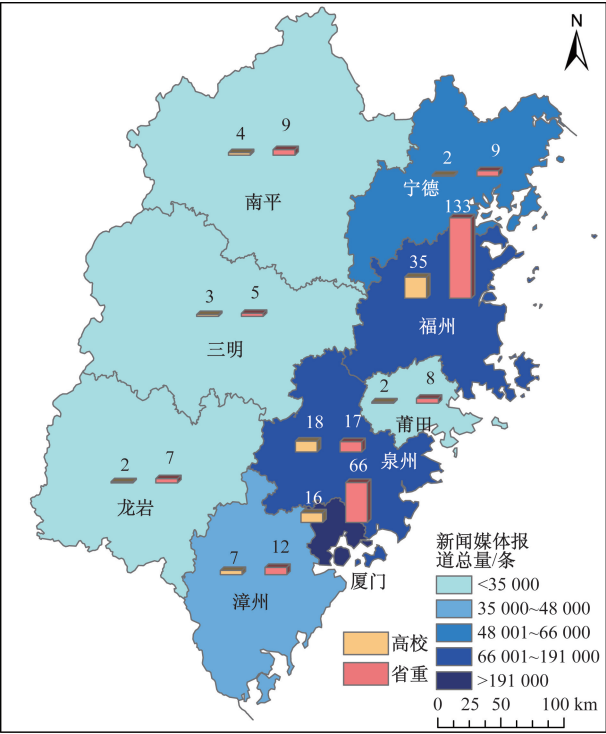
(1)从地域差别来看,新闻媒体的报道数量与各地市的经济水平有明显的相关性,皆以厦门、福州和泉州三地领先,且这三地与其他地级市差距较大,呈现明显的两极分化;福州的数据在所有地级市中的波动性最大,考虑其为省会城市的原因,受各种政策和外界因素影响较大;另外还发现不同地级市在媒体类型上的侧重点也各有不同,比如平潭在微信这块的占比要明显高于其他地级市。

(2)从时序变化来看,除了平潭,新闻媒体报道数量排名靠前的一般呈现较明显的高峰期。这些高峰期的出现,通常与国家或当地所颁布的各类政策密切相关。例如,平潭在2023年9月有明显的高峰,而对应时间段内,国务院发布了《中共中央 国务院关于支持福建探索海峡两岸融合发展新路建设两岸融合发展示范区的意见》。

(3)从省内各高校和省重分布情况来看,福州、厦门和泉州依旧是数量最多的,从侧面反映,具备丰富的教育资源和创新平台的地方才更容易催生丰富的创新成果,从而吸引更多媒体报道,获取更多的关注度。

4.2 展望

后续研究可针对福建省“六四五”产业新体系的划分,具体分析各个行业领域的媒体报道数量情况,探究不同产业与新闻媒体报道之间的潜在联系和影响。此外,在分析过程中,也发现了一些噪声信息,这些信息通常有着虚假或误导性的标题和与主题无关的内容,本文没有对这些噪声信息进行详细讨论,但它们的存在对数据分析的准确性和可靠性的潜在影响仍不容忽视。因此,未来研究可以考



平潭仅有一家省重,图中未展示平潭数据;基于审图号为 GS(2024)0650 号标准地图制作,底图无修改

图 6 福建省各地(市、区)高校和省重及新闻媒体报道总量分布情况

虑如何识别和过滤这些噪声信息,提高相关数据分析的准确性和可靠性。

参考文献

[1] 曹磊. 基于互联网科技媒体语料库大数据分析的前沿科技识别与跟踪方法初探[J]. 竞争情报, 2019, 15(3): 18-29.
[2] 祝娜, 王效岳, 杨京, 等. 基于 LDA 的科技创新主题语义识别研究[J]. 图书情报工作, 2015, 59(14): 126-134.
[3] 祝娜, 王芳. 信息交互视角下 Twitter 用户的科技创新传播行为研究[J]. 情报学报, 2020, 39(12): 1340-1353.
[4] 周茜, 陈收. 公众媒体参与金融科技创新监管的双刃剑效应[J]. 系统工程理论与实践, 2022, 42(7): 1782-1795.
[5] 曹树金, 闫颂. 基于语义角色信息的科技论文创新段落

定位及功能句识别方法研究——以中文情报学领域论文为例[J]. 情报理论与实践, 2022, 45(11): 1-9.

[6] 王曰芬, 王柳虹, 巴志超, 等. 政府科技新闻中科技成果转化主题识别与时空扩散分析[J]. 情报学报, 2023, 42(8): 939-951.
[7] 孔嘉, 邓三鸿, 张越, 等. Web 环境下的科技新闻主题挖掘——以国际极地考察领域为例[J]. 科技管理研究, 2020, 40(17): 173-179.
[8] 张来武. 科技创新驱动经济发展方式转变[J]. 中国软科学, 2011(12): 1-5.
[9] 徐雷, 张亚菲, 叶均玲. 科技文献创新内容的识别、组织与应用进展[J]. 情报学报, 2024, 43(2): 237-250.
[10] 陆成宽. 2023 年国内十大科技新闻解读[N]. 2023-12-26(002).

Big Data Analysis of Technology Innovation in Fujian Province Based on News Media

CHEN Jian^{1,2}, LIN Hong^{1,2}

(1. Fujian Institute of Scientific and Technological Information, Fuzhou 350003, China;
2. Fujian Key Laboratory of Information and Network, Fuzhou 350003, China)

Abstract: As an important channel for information dissemination, news media reflects the dynamics of technological innovation. Using the “Sina Public Opinion Monitoring” platform, an analysis was conducted on news reports related to scientific and technological innovation in various cities of Fujian Province from January 2023 to June 2024. It is found that the number of reports is significantly correlated with the local economic level. Xiamen, Fuzhou, and Quanzhou rank in the top three respectively. In addition, the data of Fuzhou shows the greatest volatility. It is also found that the proportion of each city in different media types is slightly different. Time series analysis indicates that the reporting peaks are usually related to policy releases, and there is a certain correlation between the reporting volume and the number of universities and key laboratories at the same time.

Keywords: science and technology innovation; news media; Fujian Province; big data analysis