

# 新时代科技新闻报道的主题研究

李云晓, 侯懿航

中国矿业大学(北京)文法学院, 北京

收稿日期: 2026年5月2日; 录用日期: 2026年5月22日; 发布日期: 2026年6月5日

## 摘要

融媒体时代背景下,科技作为第一生产力,科技新闻的话语传播与建构对于科技传播实效具有很大影响。近年来,越来越多学者开始关注科技新闻话语体系建构,但当前仍然面临科技新闻内容视角单一,深度不足等问题。本文以《中国日报》的科技新闻为语料借助自然语言处理技术中的主题建模方法,系统挖掘科技新闻的主题分布特征。研究表明,《中国日报》科技新闻聚焦“专业人才培养”、“市场规模扩大”、“数字经济发展”、“产业创新”、“文化传承与建设”、“场景应用需求与发展”、“技术研发”七大主题,具有需求导向鲜明,应用落点清晰,多元协同等特点。本研究结合文本量化和话语质化,完善了科技新闻传播体系建构研究,为后续科技新闻报道策略与导向提供了实证支撑。

## 关键词

科技, 新闻话语, 主题建模, 自然语言处理, 中国日报

# A Study on the Topics of Technology News Reporting in the New Era

Yunxiao Li, Yihang Hou

School of Law and Humanities, China University of Mining and Technology (Beijing), Beijing

Received: May 2, 2026; accepted: May 22, 2026; published: June 5, 2026

## Abstract

Against the backdrop of converged media era, science and technology are the primary productive force; the discourse communication and construction of science and technology news have a great impact on the effectiveness of science and technology communication. In recent years, an increasing number of scholars have begun to focus on the construction of technological news discourse systems. However, current tech news still faces challenges such as limited perspectives and insufficient depth in content. This study utilizes scientific and technological news from *China Daily* as the corpus,

employing topic modeling techniques from natural language processing technology, to systematically analyze the thematic distribution characteristics of scientific and technological news. This study reveals that *China Daily's* science and technology news coverage focuses on seven key themes: professional talent development, market scale expansion, digital economy growth, industrial innovation, cultural heritage preservation and development, application scenario demands and evolution, and technological R&D, characterized by distinct demand orientation, clear application focus, and multi-party collaboration. This study integrates textual quantification and discursive qualitative analysis to advance research on the construction of science and technology news communication systems, providing empirical support for subsequent reporting strategies and guidance in science journalism.

## Keywords

Technology, News Discourse, Topic Modeling, Natural Language Processing, China Daily

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

融媒体时代对科技新闻的生产、传播方式都有着十分深刻的影响,因此,科技新闻作为传递科技动态、阐释技术价值、促进科学普及的重要媒介,其话语表达、主题结构都会直接、实质性地影响着科技传播的效果。目前主流媒体的科技新闻存在内容同质化、深度不够、视角单一等诸种现实问题,也因此制约了科技传播质量的提高[1]-[4]。故而,利用文本挖掘结合跨学科理论的方法进行研究正是解决此问题的关键途径。

自20世纪50年代以来,自然语言处理(NLP)技术在文本分析领域已有十分成熟且充分的应用,可以以算法形式可靠、系统地对话言结构及语义加以处理[5]-[7]。与此直接关联的是,LDA主题模型作为典型的无监督学习方法,有着“文本-主题-主题词”的清晰结构,因此非常适合于从大规模语料中客观、自动地提取主要议题,有利于规避人工标注中所隐含的主观因素,适用于长时段、大容量新闻文本的研究中[8]-[10]。

本文以《中国日报》2013年1月1日至2025年4月10日科技新闻为研究样本,首先通过网络爬虫采集语料并加以规范化预处理,继而使用LDA模型进行主题建模,系统考查科技新闻的主题分布规律,充实完善科技新闻研究的方法体系,更有利于优化科技新闻报道策略,切实提高科技传播的实效性。

《中国日报》作为我国对外传播的重要窗口,在科技议题方面具有代表性地位,以其为例开展长时段系统分析能够更清晰、有效地呈现国家级外宣媒体的科技新闻话语构建特征,可为相关领域研究提供具有参考价值的实证依据。

但本研究仍有一定局限性,仅聚焦于《中国日报》这个国家级英文外宣媒体报道,暂未涉及国内主流媒体、海外媒体及自媒体等其他传播主体。受研究范围所限,本文揭示的主题分布与建模主要反映《中国日报》的科技新闻传播取向,尚不能完全代表我国科技新闻话语的整体面貌。

## 2. 研究设计

### 2.1. 语料库概况

本文以《中国日报》2013年1月1日至2025年4月10日年科技新闻文本作为主要研究语料,首先,

使用网络爬虫技术对目标文本做出定向、全量采集, 由此直接建成科技新闻专用语料库。更重要的是, 本研究明确限定了时间范围及主题边界, 只保留以“科技”为明确叙事焦点的新闻报道, 在采集原始数据之后又系统、有层次地进行人工核查以及机器清洗, 严格剔除重复文本及所有与科技主题无关的内容, 既确保了研究语料的纯粹性, 也更有利于进行后续研究。最终规范好的文本统一导入 Antconc 语料库软件, 形成标准化语料库, 形成了对于后面的主题挖掘、关键词搭配分析、情感判断诸种研究任务最可靠、最理想的数据基础。

## 2.2. 研究方法

自然语言处理技术是本文进行大规模文本分析所依赖的最基本、最核心的技术支撑, 即用计算机算法对人类语言的结构、词汇的意义、文本的逻辑进行自动化、可量化的解析, 能够高效地对海量文本进行数据清洗、特征提取及规律挖掘[5][6]。自然语言处理技术在新闻话语研究中已极好地弥补了传统人工阅读效率低、主观性强的种种不足, 因此成为处理长时段、大容量语料时的得力工具。LDA 隐含狄利克雷分布模型是自然语言处理体系中应用最广泛、最成熟的无监督学习方法之一, 具有明确、清晰的“文本-主题-主题词”三级结构, 在不依赖人工标注的情形下即可自动从语料中提取潜在主题, 并给出高权重关键词, 由此最大限度地减少研究者主观判断造成的偏差[9]。

## 2.3. 研究问题

1) 借助 LDA 主题建模工具, 探究 2013 年 1 月 1 日至 2025 年 4 月 10 日《中国日报》科技新闻中所呈现的核心主题类型与分布特征。

2) 基于关键词共现与语义关联, 揭示各主题内部的词汇结构、话语表达及话语建构模式。

## 3. 研究理论

本文依托自然语言处理技术(natural language processing, NLP), 构建科技新闻语料库, 对其进行文本挖掘, 利用 LDA 主题模型提取出《中国日报》2013 年 1 月 1 日至 2025 年 4 月 10 日有关“科技”的主题, 并进一步阐述科技新闻话语传播特点。

自然语言处理技术(NLP)自 20 世纪 50 年代开始发展, 逐渐拓展到不同的学科领域[5]。其主要作用是根据计算机算法开发模型对单词和语言结构进行理解用来研究人类语言, 通过算法分析人类语言, 进而给予反馈[6]。

隐含狄利克雷分布(Latent Dirichlet Allocation, LDA)模型是自然语言处理技术中较为常见的一种无监督文本学习方法。LDA 模型能够减少研究者对于语料的主观标记, 主要依赖自身算法对文档进行关键词提取[9][11]。LDA 模型具有明确的三级结构“文本-主题-主题词”, 使得文本主题分布可视化程度高。LDA 模型生成的关键词与主题之间的相关性具有高低之分, 通过这一可视化产出, 在进行主题阐释时能够直观地观察出主题分布的特点。

## 4. 结果与讨论

### 4.1. 主题建模总体解读

本文通过 LDA 模型得出 7 组关键词, 根据回溯原文详细分析关键词的语义联系并通过 Antconc 语料库软件进行关键词高频索引词搭配, 共提炼出“专业人才培养”、“市场规模扩大”、“数字经济发展”、“产业创新”、“文化传承与建设”、“场景应用需求与发展”、“技术研发”这 7 个主题(见表 1)。

**Table 1.** Topic modeling results of technology news in *China Daily*

**表 1.** 《中国日报》科技新闻主题建模结果

序号	主题命名	主题关键词
T1	专业人才的培养	0.012* “项目” + 0.011* “有限公司” + 0.007* “记者” + 0.006* “公司” + 0.006* “北京” + 0.006* “专业” + 0.005* “全国” + 0.005* “股份” + 0.005* “江西” + 0.005* “现场” + 0.005* “人才” + 0.005* “集团” + 0.005* “成都” + 0.005* “代表” + 0.004* “大赛” + 0.004* “创业” + 0.004* “员工” + 0.004* “职业” + 0.004* “学院” + 0.004* “团队”
T2	市场规模扩大	0.021* “企业” + 0.018* “产品” + 0.015* “品牌” + 0.014* “市场” + 0.011* “增长” + 0.011* “实现” + 0.010* “需求” + 0.009* “提升” + 0.008* “消费者” + 0.008* “技术” + 0.007* “行业” + 0.007* “公司” + 0.007* “用户” + 0.006* “超过” + 0.006* “显示” + 0.006* “全球” + 0.005* “规模” + 0.005* “提供” + 0.005* “中国” + 0.005* “达到”
T3	数字经济发展	0.075* “中国” + 0.024* “经济” + 0.018* “国际” + 0.018* “世界” + 0.016* “全球” + 0.015* “国家” + 0.013* “数字” + 0.012* “时代” + 0.010* “大会” + 0.009* “论坛” + 0.008* “社会主义” + 0.008* “举办” + 0.007* “人民” + 0.007* “创新” + 0.007* “表示” + 0.006* “文化” + 0.006* “主题” + 0.006* “博会” + 0.006* “来源” + 0.005* “成为”
T4	产业创新	0.023* “创新” + 0.014* “产业” + 0.014* “推动” + 0.013* “推进” + 0.010* “科技” + 0.010* “国家” + 0.009* “体系” + 0.009* “企业” + 0.009* “人才” + 0.009* “坚持” + 0.009* “提升” + 0.008* “高质量” + 0.008* “全面” + 0.008* “政策” + 0.008* “绿色” + 0.007* “加快” + 0.007* “实现” + 0.007* “经济” + 0.006* “促进” + 0.006* “战略”
T5	文化遗产与建设	0.015* “文化” + 0.009* “学习” + 0.008* “机器人” + 0.008* “问题” + 0.007* “精神” + 0.007* “空间” + 0.007* “传统” + 0.006* “航天” + 0.006* “城市” + 0.006* “体验” + 0.006* “时代” + 0.006* “解决” + 0.005* “学生” + 0.005* “中国” + 0.005* “价值” + 0.005* “人类” + 0.004* “成为” + 0.004* “品牌” + 0.004* “作为” + 0.004* “社会”
T6	场景应用需求与发展	0.018* “场景” + 0.010* “进行” + 0.005* “介绍” + 0.004* “院士” + 0.004* “现场” + 0.004* “视频” + 0.004* “分享” + 0.004* “教授” + 0.004* “技术” + 0.004* “获得” + 0.004* “总经理” + 0.004* “芯片” + 0.004* “科技” + 0.004* “嘉宾” + 0.004* “建筑” + 0.004* “团队” + 0.004* “项目” + 0.004* “系统” + 0.004* “专家” + 0.003* “社区”
T7	技术研发	0.023* “技术” + 0.022* “科技” + 0.018* “企业” + 0.017* “行业” + 0.017* “智能” + 0.015* “数字化” + 0.014* “应用” + 0.014* “平台” + 0.012* “数据” + 0.011* “领域” + 0.011* “智慧” + 0.010* “数字” + 0.010* “产业” + 0.009* “创新” + 0.009* “产品” + 0.009* “提供” + 0.008* “集团” + 0.007* “业务” + 0.007* “打造” + 0.007* “转型”

#### 4.2. 主题解读

##### 专业人才培养:

由高频词搭配表可知, 与“人才”的高频搭配实词以动词和形容词为主。两者在语义上略有差距。形容词主要聚焦人才类型, 重点体现出国家需要专业人才, 例如“高层次; 青年, 复合型”。索引行中以更专业化的形容词体现出对于专业人才的重视, 例如“消防人才, 网络安全人才, 技术人才”。而动词则主要体现对人才的行动措施的话语建构, 例如“引进, 培养, 市场”(见表 2)。

**Table 2.** Cultivation of professional talents

**表 2.** 专业人才的培养

Collocate	Rank	Likelihood
培养	1	25783.881
人才	2	25749.209
高层次	3	20730.638

续表

引进	4	13060.093
技能	5	7210.925
创业	6	6657.217
高素质	7	6383.251
领军	8	6196.138
青年	9	6038.436
高技能	10	6037.432
复合型	11	6020.634
培育	12	5022.679
高端	13	4941.011
集聚	14	4691.511

**市场规模扩大:**

由“高频词搭配表”可知,与“市场”的高频搭配实词主要聚焦市场的行动维度与空间维度。其中以形容词和动词占比最大。形容词诸如“国内”“海外”“全球”“广阔”聚焦市场的空间维度,通过空间的扩大体现出市场规模的扩大。而动词诸如“开拓”“扩展”“进入”则以物质过程更加直接体现中国为促进市场规模扩大做出的积极行动(见表3)。

**Table 3.** Expansion of market scale

**表 3.** 市场规模扩大

Collocate	Rank	Likelihood
资本	1	43929.494
规模	2	13755.589
中国	3	12112.359
开拓	4	10432.116
海外	5	9884.335
拓展	6	9725.407
国内	7	9605.351
全球	8	7854.100
竞争力	9	7111.122
进入	10	5074.079
市场	11	5028.814
广阔	12	4502.211
下沉	13	4296.541
经济体制	14	4045.163
细分	15	4037.158

### 数字经济发展:

由“高频词搭配表”可知,与“数字”的高频搭配实词以名词和动词为主,名词包含中性名词和消极性名词,中性名词如“经济”“技术”“产业”“贸易”等体现出不同领域的数字化转型,而消极名词,如“鸿沟”则体现出目前数字化仍然存在着挑战与风险,为下一步中国如何完善数字化提供导向。而动词则以表示物质过程的动词为主,“发展”“建设”“赋能”都体现出中国在数字化过程中所做的积极行动。从而展示出当前中国科技领域,数字化的进程整体呈现出积极发展的态势,但仍有问题等待解决或尚在解决现状(见表4)。

**Table 4.** Development of the digital economy

**表 4.** 数字经济发展

Collocate	Rank	Likelihood
经济	1	121907.175
孪生	2	30306.915
数字	3	26105.921
技术	4	12323.729
发展	5	12054.346
建设	6	8987.347
产业	7	8564.925
赋能	8	8031.293
鸿沟	9	6556.493
贸易	10	6511.123
时代	11	6431.052
人民币	12	6270.602
底座	13	5045.273
藏品	14	4950.773
金融	15	4577.206

### 产业创新:

科技新闻报道中所使用的“创新”关键词搭配有清楚可辨的逻辑:以“科技”“产业”“人才”诸种名词来描绘“技术驱动、要素完备”的发展图景,以协作模式类名词“协同”突出创新主体联动的特性,由此自然、妥帖地突出创新过程的合作性及系统性,也强化了创新生态中各要素协同共生的特征。继而以“驱动”“推动”等核心动力类词汇厘清创新在产业发展中的核心引擎作用,论证其赋能价值及推动力量。最后以“引领”“发展”等引领性动词建构创新引领未来的前瞻性论断,形成创新领先、方向明确的有力叙事(见表5)。

**Table 5.** Industrial innovation

**表 5.** 产业创新

Collocate	Rank	Likelihood
科技	1	120628.015
创业	2	49676.404

续表

驱动	3	31162.735
创新	4	25111.1
发展	5	24259.281
成果	6	19674.639
产业	7	16128.524
引领	8	15608.689
推动	9	14067.768
中心	10	13349.836
大赛	11	11852.198
人才	12	11533.779
融合	13	11329.293
协同	14	10766.523
自主	15	10316.957

**文化传承与建设:**

科技新闻中所见的“文化”的高频搭配词以名词、动词两类为主,且两者语义上彼此分明、各就其位:名词方面,“中华”“民族”“传统”自然、恰当地突出文化根基,“优秀”“红色”“内涵”层层强化文化内核,“旅游”“艺术”具体、明确地交代实践场景,从而自然地勾勒出“文化为基”的创新发展图景。动词方面,“传承”“弘扬”“传播”将创新与文化遗产的时间维度直接、有力地联结在一起,故而传统文化在科技赋能下的生命力得以极好呈现。“融合”“创意”着眼于创新实践路径,把文化与科技的跨界整合,也与文化与科技融合的发展导向形成完美呼应(见表6)。

**Table 6.** Cultural heritage and development

**表 6.** 文化传承与建设

Collocate	Rank	Likelihood
传统	1	41055.495
旅游	2	37389.894
传承	3	25595.444
中华	4	16172.185
自信	5	15415.743
历史	6	13203.725
优秀	7	12562
传播	8	12080.677
艺术	9	11958.142
创意	10	9836.091
弘扬	11	8787.64
内涵	12	7563.167
融合	13	7014.669
红色	14	6312.07
民族	15	6210.415

### 场景应用需求与发展:

科技新闻中所见到的“场景”高频搭配以名词、动词为主,二者有着十分清晰的语义分工,彼此呼应,形成完整、有层次的应用图景:名词方面,“应用”“业务”“解决方案”都明确说明落地领域,“中国”“生活”自然、妥帖地界定服务范围,故而有“需求导向、多元覆盖”的明确应用图景。动词方面,“落地”“发展”交代场景转化路径,“使用”“体验”突出用户价值,因此也自然地塑造出“技术赋能、体验优先”的建设形象,二者都契合场景化创新导向及用户需求本质(见表7)。

**Table 7.** Scenario application demand and development

**表 7.** 场景应用需求与发展

Collocate	Rank	Likelihood
应用	1	60060.78
消费	2	6105.193
业务	3	5502.553
落地	4	4840.892
发展	5	3882.655
生活	6	3424.51
使用	7	3421.392
智慧	8	3152.602
中国	9	3037.714
需求	10	2725.716
解决方案	11	2511.409
真实	12	2377.928
多种	13	2086.065
复杂	14	2000.915
体验	15	1880.158

### 技术研发:

科技新闻中所用的“技术”高频实词搭配词以名词、动词为主,但二者承载的语义信息却十分明确,且有层次的差别:在名词使用上,科技新闻从技术属性、应用场景、发展定位诸种角度系统、有逻辑地呈现技术特征,因此自然、妥帖地突出技术的领域归属(“人工智能”“数字”“新兴”),载体形态(“应用”“产品”)及层级优势(“先进”“领先”“优势”),形成“领域聚焦、优势凸显”的技术发展图景。在动词使用上,即用“研发”“攻关”“突破”诸词集中描写技术从积累到突破的动态过程,突出技术创新的突破性价值,与科技领域以技术创新驱动发展的基本逻辑形成完美契合(见表8)。

**Table 8.** Technological research and development

**表 8.** 技术研发

Collocate	Rank	Likelihood
应用	1	18463.746
人工智能	2	16998.368

续表

先进	3	15773.649
研发	4	12332.265
数字	5	10553.827
产品	6	8271.3
突破	7	5284.689
职业	8	5201.513
领先	9	5078.597
进步	10	4836.381
攻关	11	4749.244
积累	12	4530.534
路线	13	4497.761
优势	14	3981.803
新兴	15	3831.081

## 5. 结语

本文以 2013 年 1 月 1 日至 2025 年 4 月 10 日《中国日报》科技新闻为语料, 恰当运用自然语言处理技术、LDA 主题建模方法, 对新时代科技新闻的主题分布进行了系统、有层次的考察。首先用 LDA 模型提取出专业人才培养、市场规模扩大、数字经济发展、产业创新、文化遗产与建设、场景应用需求与发展、技术研发七大主题, 继而通过关键词搭配及索引进行分析, 厘清主流媒体科技新闻集中讨论技术突破、产业升级、创新实践此类要点, 传播时有明确的需求导向、应用落点及多元协同的特征, 由此有力地呈现出科技新闻中技术、社会、生活三者的共生关系以及相应价值导向。

本文把量化文本挖掘与质化话语阐释相结合, 较好地解决融媒体时代科技新闻内容同质化、深度不够、视角单一诸种现实问题, 也因而充实、完善了科技新闻传播研究的方法论体系, 对改进科技新闻报道策略提供了实证支撑。

本研究结论基于 2013~2025 年《中国日报》科技新闻单一语料得出, 暂无法全面呈现我国科技新闻语料的整体特征, 但《中国日报》作为国内最具国际影响力的英文外宣平台, 其科技新闻报道是展现中国科技发展与创新的重要载体, 对其进行专题研究有助于更加精准地把握外宣媒体视角下中国科技新闻报道的主题特征与话语策略。未来可以进一步扩大语料范围, 开展跨媒体对比研究, 系统展开学科交叉分析, 持续探索科技新闻话语的传播规律及实践路径, 也让科技新闻传播更专业、更深入、更有社会价值。

## 基金项目

本研究系中国矿业大学(北京)大学生创新训练项目的阶段性成果, 项目编号为 202508017。

## 参考文献

- [1] 刘新莉. 融媒体环境下科技新闻报道的创新策略[J]. 科技传播, 2020, 12(15): 89-90.
- [2] 张彬. 浅谈融媒体时代如何提升科技报道的传播力[J]. 新闻研究导刊, 2020, 11(2): 139+142.
- [3] 向一鹏. 融媒体背景下科技新闻传播创新探讨[J]. 中国报业, 2023(1): 102-103.

- 
- [4] 康家佳. 融媒体时代科技新闻报道面临的困境及应对之策探析[J]. 西部广播电视, 2024, 45(8): 49-52.
- [5] 熊建国. 自然语言处理工具在语料驱动批评话语分析中的运用——以英语洗钱新闻为例[J]. 解放军外国语学院学报, 2020, 43(5): 10-19+159.
- [6] 孙磊. 基于人工智能的自然语言处理技术分析[J]. 集成电路应用, 2025, 42(2): 146-147.
- [7] 冯志伟. 自然语言处理的历史与现状[J]. 中国外语, 2008(1): 14-22.
- [8] Blei, D.M., Ng, A.Y. and Jordan, M.I. (2003) Latent Dirichlet Allocation. *Journal of Machine Learning Research*, **3**, 993-1022.
- [9] 何伟, 程铭. 基于文本挖掘技术的美国清洁能源推文生态观研究[J]. 中国外语, 2024, 21(6): 26-34.
- [10] 逯万辉. 科学文献主题建模方法及其效果评估研究[J]. 现代情报, 2024, 44(4): 22-31.
- [11] 卫宣伶, 孙翔. 自然语言处理技术方法研究进展与发展趋势综述[J]. 中阿科技论坛(中英文), 2025(5): 84-88.