

中国区域科技创新政策量化评价研究

郝海拓, 张志娟

(中国科学技术信息研究所, 北京 100038)

摘要: 本研究旨在基于1506项区域科技创新政策标题对区域科技创新政策进行量化研究,从高频词、主题关键词、主题关键词的内容词项及相关度等方面分析我国东中西部三大经济区的创新政策情况。基于R语言和文本挖掘,借助主题模型将1999—2017年我国1506项区域科技创新政策标题进行深度解析。研究表明我国三大经济区区域科技创新政策具有不同特征;试点仍是各地区最常用的政策推广方法;我国各地方政府正在向服务型政府转变且初见成效。

关键词: LDA主题模型; 主题关键词; 科技创新政策; R语言; 政策量化

中图分类号: G322.0 **文献标识码:** A **DOI:** 10.3772/j.issn.1009-8623.2019.10.007

根据CB Insights的榜单,2017年中国占据了全球36%的新生“独角兽”份额,中国独角兽企业总数为59家^[1]。我国创新活动进入了高速发展期,党和国家对创新和创业的强力支持是创新创业工作开展的动力源泉,而各地方政府结合本地区的具体特点对创新创业予以的政策支持对企业来说同样重要。在十九大报告中,习总书记强调“创新驱动发展战略”,并“鼓励更多社会主体投身创新创业”,区域科技创新政策是国家实施宏观调控最主要的手段,对国家经济和发展起着重要的引领作用,而非结构化或半结构化政策文本的量化和评价一直是学术界研究的重点,随着学科交叉的深化,各学科间的相互渗透也为政策研究提供了重要的研究思路 and 工具。

相比于以往的研究,本文在方法和研究内容上都有所创新。方法上,本文旨在运用R语言将政策文本转化成可供分析的各类客观数据,以加深区域科技创新政策的研究深度,更深入、多维度地对其进行剖析。另外,在选择最佳主题数量时,以往的大多数研究都是作者主观节选一定数目的关键词,而在本研究中,运用R语言进行计算分析,

计算得到的最优主题关键词数为7,因而本研究严格客观执行计算结果,选择7个主题关键词进行后续分析。研究内容上,以往对区域科技创新政策进行的研究往往集中在政策文本内容的研究上,如通过词频确定重点研究内容。但某一项具体政策的研究内容往往会在政策文本中大量重复出现,这就使得通过正文词频的统计方法无法剔除政策文本大量重复所带来的结果偏差。而新兴起的概念或热点也很容易埋没在大量的重复数据中。因而本文选取1999—2017年间涉及区域科技创新政策的1506项政策标题,避免正文大量重复出现关键词给研究结果带来的偏差,政策标题是整篇政策文本的高度凝练,不仅能有针对性地反映政策的核心内容,还能有效避免对核心词的重复计算。因而本研究在方法和研究内容上都有一定创新。

1 文献回顾

1.1 区域科技创新政策分类及评价研究综述

Rothwell等^[2]最早于1985年将科技创新政策划分为供给、需求和环境三种类型,该分类方式较经典,许多专家学者的研究均基于此分类开展。

第一作者简介:郝海拓(1988—),女,助理研究员,主要研究方向为科技创新政策、企业创新。

收稿日期:2019-09-20

Boschma^[3]以区域创新政策为研究对象,提出区域科技创新政策的区域创新战略、政策及活动的三维划分标准。Freitas等^[4]则认为可以按照使命型、扩散型、一般型、特定型、地方主导、中央主导型对政策进行类型划分。Schramm等^[5]认为立法监管、财政与技术经济和公开披露等维度对科技创新政策的评价和分类而言至关重要。Carreira等^[6]提出应该依据立法监督、信息量、科技经济效应、决议和刺激性四维度来划分政策类型。李晨光等^[7]认为集群创新科技政策可以划分为紧迫性、权威性、指导性、阶段性以及长期性政策,并结合中关村为案例进行研究。张永安等^[8]从政策力度、政策目标和政策措施三个维度测度科技创新政策与国家创新能力之间的关系。贺德方等^[9]基于国家创新体系理论从要素、主体、关联、产业、区域、环境、开放和反馈等方面对科技创新政策体系框架进行分析。

分类及评价研究是对区域科技创新政策进行的基础性研究,首先对政策进行分类,然后通过不同维度对政策进行评价,评价后进行量化研究,使复杂的非结构化和半结构化文本转化为直观可读的结构化数据,用于对政策立改废提供可量化的依据,这是政策研究的主要思路。

1.2 区域科技创新政策量化挖掘研究综述

区域科技创新政策量化挖掘研究是基于政策分类和评价的进一步探索,旨在将复杂的文本文件转化为便于比较的量化数据,国内外专家学者在此方面都进行了各类有意义的探索。Green等^[10]认为应从创新模式、功能与政策效应三方面测度区域科技创新政策。我国学者彭纪生等^[11]提出涉及力度、措施与目标三维度的政策评价手册,并在每个维度下设二级指标,结合政策文本实现政策量化,其方法为政策量化提供了系统的标准化依据,成为后续许多学者研究的参照标准。刘会武等^[12]将“钻石模型”引入创新政策测度体系。Estrada^[13]提出PMC指数模型方法,为政策的量化与评价提供了专业的方法和系统的依据。张永安等^[14]即基于此PMC指数模型对2014—2016年国务院的创新政策进行了量化。盛亚等^[15]基于京、沪、浙、粤、苏五地的区域创新政策文本,实现了政策目标和政策措施的量化和比较研究。张炜等^[16]采用政策赋值

量化法以及模糊数学方法构建基于创新政策的协同度、政策强度和政策完善度的多维度评价指标,实现区域创新政策的量化。张永安等^[17]使用R语言结合政策文本实现我国区域科技创新政策的文本挖掘,对政策主题、内容及相关性进行了深入的研究。王卉彤等^[18]以北京、上海和深圳为例研究区域创新政策中的金融服务政策,并通过量化结果提出三个地区不同的创新创业政策侧重点。国内外学者对科技创新政策的量化主要基于指标及各类量化工具,随着研究的深入,政策量化结果越来越能够精准地体现政策的具体特征,为政策的立改废提供可行的依据。

2 研究方法数据来源

2.1 研究方法

本文采用R语言作为文本转化为结构化数据的最主要工具,结合文本挖掘方法实现政策标题文本的数据转化。Prior等^[19]首次将政策文本和语义分析结合,对政策基本要素予以界定。随着学者研究的深入,主题挖掘逐渐成为研究重点内容之一,我国学者常用的研究方法主要包括共词矩阵、社会网络和主题模型等^[20],通常是对主题的内涵、词频、强度与关联关系等的分析。本文运用的潜在狄利克雷分布(Latent Dirichlet Allocation, LDA)主题模型常被用于非结构和半结构化的文本数据的主题和知识的挖掘、文本情感挖掘等领域。

2.2 数据来源

本研究选取北大法宝数据库作为主要的政策文本来源,共选取1506项区域科技创新政策标题,其中东部955,占总数的63.41%;中部276,占比18.33%;西部地区275项,占比18.26%。各城市的政策数量也在表1中予以列示。

以表1中统计的区域科技创新政策标题为样本,分别对我国中东西三大经济区的区域科技创新政策情况进行研究。

3 基于R语言的分析流程

结合已有研究及本研究思路,基于R语言使用原理建立政策标题挖掘模型,模型主要分为5部分。第一步,建立区域科技创新政策标题库。从北大法宝数据库中筛选出符合标准的政策标题,将这

表 1 政策文本选取明细表

全国总计: 1 506					
东部:	955	中部:	276	西部:	275
北京	30	安徽	36	甘肃	24
福建	28	河南	106	广西	27
广东	215	湖北	54	贵州	13
海南	4	湖南	18	内蒙	30
河北	37	江西	41	宁夏	43
黑龙江	22	山西	21	青海	4
吉林	27			陕西	31
江苏	193			四川	41
辽宁	54			西藏	2
山东	43			新疆	12
上海	171			云南	12
天津	15			重庆	36
浙江	116				

些标题放置在区域科技创新政策标题库中。第二步, 对政策标题进行预处理, 设定词库以便设定分词和停用等规则。第三步, 基于 R 语言和文本挖掘方法实现标题量化, 通过高频词、云图 (Word Cloud) 实现文本的数据化, 第四步, 通过 R 语言计算得出最优主题关键词个数并进行主题关键词筛选, 这也是本研究的创新点之一, 避免以往直接凭感觉进行选择的主观性。第五步, 通过 R 语言得到主题关键词词项、相关度及主题词矩阵等量化结果。具体步骤如图 1 所示。

3.1 建立标题库

本文筛选区域科技创新政策标题的具体步骤是在北大法宝数据库中设定具体要求, 设定研究对象为区域政策, 选择关键词科技、创新等分别对备选政策进行筛选, 选择这些关键词下的政策文本, 然后根据政策标题进行筛选, 剔除无关或已废止或失效的政策, 最终形成包含 1 506 项区域科技创新政策标题的标题库。

3.2 预处理

根据标题库中的标题, 自行设定词库并结合

Rwordseg 进行中文分词和停用词处理, 完成对区域科技创新政策标题的预处理。

3.3 高频词处理

在对政策文本标题进行分词和停词等预处理后, 要通过 N-gram 模型对核心词进行认定, 假定某句话 m 由特定的 n 项词语组合而成, 为方便说明将其依次命名为 $w_1, w_2, w_3, \dots, w_n$, 假设每一词语出现的概率或可能性仅与其前 $n-1$ 项有关, 则将第 i 项的概率表达为:

$$P(w_i | w_1, w_2, w_3, \dots, w_n) = P(w_{i-n+1}, w_{i-n+2}, w_{i-n+3}, \dots, w_{i-1}, w_i) = \frac{c(w_{i-n+1}, w_{i-n+2}, w_{i-n+3}, \dots, w_{i-1}, w_i)}{c(w_{i-n+1}, w_{i-n+2}, w_{i-n+3}, \dots, w_{i-1})} \quad (1)$$

核心词汇与核心词组均可通过式 (1) 予以计算, 据此筛选出高频词。

3.4 主题关键词确认

LDA 模型是由 Blei 首次提出的^[21], 使复杂的文本转化为概率分布成为可能, 并将语义关联加入传统分析, 通过主题关键词代表词语所构成的概率分布, 保证了文本分析的客观合理性。LDA 主题

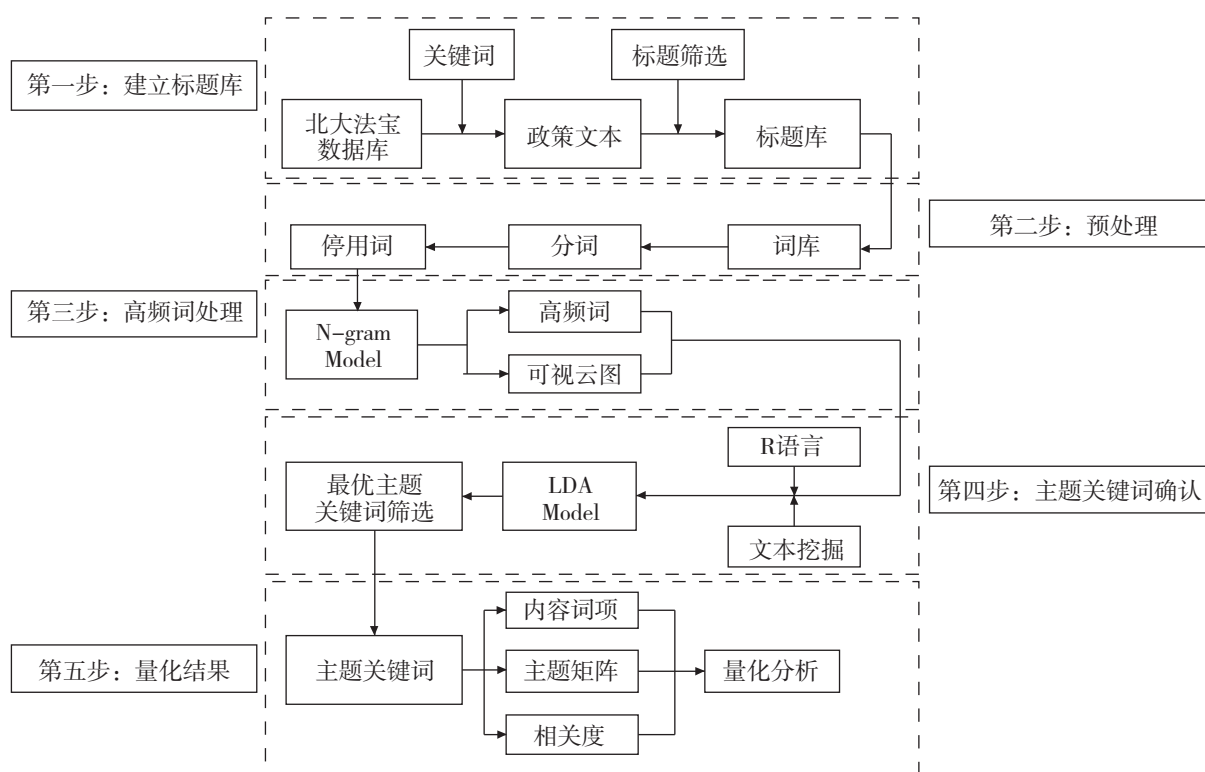


图1 基于R语言的区域科技创新政策分析流程图

模型的建立综合了 Bayes 概率理论、EM 算法以及 Dirichlet 分布理论等，为政策以及其他文本的量化提供了工具和方法。LDA 模型原理已有很多文献进行了详细的论述，本研究不再赘述，而将研究重点放在研究的结果以及结果分析上。在这一步中的关键结论是根据计算结果确定最优的主题关键词。

3.5 量化结果

通过以上几步的处理，确定主题关键词，以及各主题关键词的内容词项、主题矩阵和相关度矩阵，作为后续量化分析的基础性依据。

4 区域科技创新政策量化结果

4.1 政策标题高频词分析

词频分析是已有文献中最常用的方法，为了研究不同区域的创新政策特点，本研究将区域科技创新政策划分为东部、中部和西部三个单位，作为独立的研究对象进行研究，再将三个地区进行横向比较，以增加对比的精度。

首先将三大经济区的高频词进行列示，如表2所示。

本研究分别选取了东部、中部和西部三大经济区的区域科技创新政策标题进行词频结果汇总。首先观察东部地区，结果表明，“科技”和“创新”的词频是最高的，分别有1126和973。此外前20的高频词涉及“通知”“印发”“人民政府”“委员会”“年度”“深圳市”“上海市”“项目”“科学技术”“工作”“建设”“意见”“办法”“行动计划”等管理内容，这些信息从不同角度帮助本研究做政策标题的深度解读。从地区方面而言，出现在高频词中的地区包括“深圳市”“上海市”“江苏省”。说明这三个地区是东部地区区域科技创新政策关注度最高的三个地区，或者是政策试点地区，对东部地区甚至全国都具有一定的参考价值。从政策主体角度而言，东部高频词中较频繁出现“人民政府”“委员会”“办公室”，说明区域科技创新政策无论是制定还是执行和指导都需要地方政府及其相关部门的大力支持。在政策类型上，“通知”排名第一，词频729，另外有词频为111的“意见”和词频为107的“办法”。体现了我国区域科技创新政策的效力力度分布情况。我国学者彭纪生等^[20]

表 2 东中西部高频词统计表

序号	东部		中部		西部	
	词汇	词频	词汇	词频	词汇	词频
1	科技	1 126	科技	309	科技	299
2	创新	973	创新	275	创新	270
3	通知	729	通知	209	通知	199
4	印发	277	人民政府	151	人民政府	131
5	人民政府	252	河南省	100	印发	96
6	委员会	226	印发	84	宁夏回族自治区	60
7	年度	209	办公室	70	重庆市	54
8	深圳市	207	年度	54	青少年	51
9	上海市	196	计划	43	办公厅	47
10	项目	172	湖北省	41	科学技术	43
11	科学技术	142	工作	40	办公室	40
12	工作	137	企业	39	工作	40
13	建设	112	青少年	38	意见	40
14	意见	111	科学技术	36	发展	34
15	办法	107	决定	35	决定	34
16	行动计划	98	办法	32	办法	33
17	管理	97	管理	32	自治区	33
18	办公室	95	建设	32	农业	32
19	江苏省	94	人才	32	甘肃省	31
20	指南	92	南昌市	31	市长	30

曾提出以政策力度、目标和措施为三大评价维度的政策评价体系,其中政策力度从低分到高分代表政策强制力的增强,彭纪生将通知划分为强制力度最弱的级别,得分为 1,意见、办法为第二级别,得分为 2。结合各评分体系,可以看出我国的区域科技创新政策更多采用通知形式下达,着重于对政策受众的鼓励和引导,较少运用国家强制力来提升政策受众的参与积极性。另外,从高频词项目和科学技术可以看出,区域科技创新政策鼓励和支持以科学技术为导向的创新,更加看重集约型创新发展,体现了我国生产方式的转变。而项目是实现集约型

创新发展的最主要途径,以项目的形式来增加创新的活力和灵活性。

再来看中部地区的高频词分布情况,与东部相同或相近的词不再赘述,只观察能体现地方特色的高频词。从地区来看,高频词中多次提到“河南”“湖北”“南昌”,说明这三个地区是区域科技创新政策的重点关注区域。另外“企业”和“人才”也进入了中部地区前 20 位的高频词列表中,这说明中部地区的区域科技创新政策非常重视企业和人才对区域发展的重要作用。中部地区相对于东部而言资源相对匮乏,对人才的吸引力度有限,地方政府基于

此种情况制定相关人才引进政策，结合科学技术的引进和高新技术企业的开办，增强区域活力，带动区域经济发展。

最后看西部地区，西部地区的“宁夏回族自治区”“重庆市”“自治区”“甘肃省”是区域科技创新政策最集中的领域。另外，“科学技术”在西部地区高频词排名中高于其在东部和中部地区的排名，这说明西部地区非常重视科学技术，这对于资源相对匮乏的西部地区而言是一个利好信息。最后，对“农业”的重视也是西部地区与东中两区的最大区别，这说明西部地区的区域科技创新政策结合了本地区的实际情况，从最根本处进行了有针对性的引导，政策不再是宏观层面的，而是落实到了生产生活的方方面面，指导精度较高。

而通过横向比较可以看出,东中西部的区域科技创新政策标题词频中排在前三位的均为“科技”“创新”“通知”。这说明我国三大经济区对于科技创新都是非常重视的,而政策多以通知方式下达,给了受众更多自主权,降低其逆反心理,而在整理政策文本内容的过程中,本研究发现,区域科技创新政策多从补贴、优惠和财政支持等各方面对区域经济进行支持,而较少地对政策受众进行限制,这说明我国的政策体系正在逐渐走向完善,政府也正在转变成服务型政府且初见成效。图2将三个经济区的词云结果予以列示,有助于更为直观地了解三大经济区政策标题的重点内容。从左到右依次为东部、中部和西部的词云情况,字体大小代表的是词频的多寡。



图2 东中西三大经济区词云分布情况

4.2 政策主题强度及三大经济区主题关键词列表

4.2.1 主题强度计算方法

主题强度考察的是三大经济区的区域科技创新政策标题文本中各个主题占语料库的比例,计算方法可表示为式(2)。其中 P_k 为待求的第 k 个主题关键词强度, N 为政策标题库中所有政策的总数, θ_{ki} 为第 k 个主题关键词出现在第 i 个政策标题中的可能性大小,即概率。为了避免主观性,本文运用R语言

对最优主题关键词个数进行计算分析,发现三大经济区均在主题关键词为7时出现最优,因而本文选择7个主题关键词进行后续的研究,结果见表3。

$$P_k = \frac{\sum_i^N \theta_{ki}}{n} \quad \text{式(2)}$$

结合式(2)对三大经济区前7个主题词进行统计,结果见表3。

表 3 三大经济区科技创新政策主题关键词强度统计表

序号	东部		中部		西部	
1	科技	0.044 702	科技	0.046 339	科技	0.048 182
2	创新	0.749 776	创新	0.070 554	创新	0.065 919
3	通知	0.041 560	通知	0.096 100	通知	0.609 274
4	印发	0.042 660	人民政府	0.052 992	人民政府	0.070 705
5	人民政府	0.038 340	河南省	0.634 152	印发	0.066 763

续表

序号	东部		中部		西部	
6	委员会	0.048 315	印发	0.030 107	宁夏回族自治区	0.064 792
7	年度	0.034 648	办公室	0.069 756	重庆市	0.074 365

从表 3 的统计结果中可以看出, 东部排名前三位的主题关键词分别是“创新”“委员会”“科技”, 中部地区排名前三位的主题关键词分别是“河南省”“通知”“创新”, 而西部地区排在前三位的分别是“通知”“重庆市”“人民政府”。值得特别说明的是中国东部地区的“创新”强度占比超过了 7 成, 这说明人才和资源相对较集中的东部地区把创新放在了区域发展的战略地位上, 而从全国范围内来看, 东部又引领着中部和西部的发

展, 中部和西部“创新”的强度排名也非常靠前, 所以说创新在我国区域发展中的地位越来越重要, 地方政府对其重视程度也越来越高。

4.2.2 三大经济区主题词项列表

为了对主题关键词进行深入研究, 本研究还对主题关键词内容词项及相关度进行了深入分析, 对东部、中部和西部三大经济区分别进行主题词划分, 每个主题下设立 10 项进行列示, 东部地区结果如表 4 所示。

表 4 东部地区主题关键词及内容词项汇总表

序号	主题 1 科技	主题 2 创新	主题 3 通知	主题 4 印发	主题 5 人民政府	主题 6 委员会	主题 7 年度
1	宁波市	科技	第一批	改革	机构	暂行办法	成立
2	合作	创新	落实	转化	高等学校	财政厅	批转
3	科学技术	通知	山东省	大会	国土资源	国际	批复
4	信息化	印发	财政	领导小组	科技厅	上海市	提高
5	徐州市	人民政府	上海	推广	示范	优秀	闸北区
6	任务	委员会	区域	园林绿化	扶持	技术	资金
7	情况	年度	重点项目	单位	丹东市	市长	中共南京市委
8	山东省	深圳市	环保	教育局	服务体系	政策措施	中国
9	黑龙江省	上海市	经费	鼓励	英才	促进	贡献奖
10	精神	项目	社会	广东	软科学	林业	鼓励

在根据强度进行的选择中, 本研究选择了“创新”“委员会”“科技”作为东部地区作为最关键的关键词, 因而首先观察主题关键词“创新”。从“创新”主题关键词内容词项可以看出, 其主要围绕“科技”“创新”“项目”等内容展开, 且试点主要围绕深圳和上海市两个地区。内容词项反应了与“创新”主题关键词关系密切的各项重点内容。

第二个主题关键词是“委员会”, 通过“财政厅”“政策措施”等内容词项来反应此主题的关键内容。第三个主题关键词“科技”的内容词项信息量也较丰富, 如“合作”“信息化”“科学技术”“任务”等都反应了“科技”的关注要点。

表 5 和表 6 分别是中部和西部地区主题关键词及内容词项汇总表。

表 5 中部地区主题关键词及内容词项汇总表

序号	主题 1 科技	主题 2 创新	主题 3 通知	主题 4 人民政府	主题 5 河南省	主题 6 印发	主题 7 办公室
1	减排	加快	办公厅	转发	科技	活动	高校
2	许昌市	创业	南昌市	办公室	创新	获得者	通报
3	意见	支持	全省	政策	通知	颁发	资金
4	评审	开展	节能	建设	人民政府	南阳市	公布
5	人才队伍	试行	江西省	奖励	河南省	农业	重点
6	新兴产业	实施方案	安徽省	南昌市	印发	提升	濮阳市
7	加强	改革	山西省	委员会	办公室	暂行办法	若干
8	科技局	江西省人民政府	湖南省	专项	年度	长治市	漯河市
9	科技厅	焦作市	科技厅	表彰	计划	第五届	安徽省
10	政策措施	决定	农业	湖北省人民政府	湖北省	黄石市	服务

表 6 西部地区主题关键词及内容词项汇总表

序号	主题 1 科技	主题 2 创新	主题 3 通知	主题 4 人民政府	主题 5 印发	主题 6 宁夏回族自治区	主题 7 重庆市
1	通报	管理	科技	财政厅	年度	推进	实施方案
2	成立	政策	创新	加快	德阳市	促进	规划
3	规划	资金	通知	申报	组织	兰州市	失效
4	科技局	新疆维吾尔自治区	人民政府	教育厅	补助	领导小组	内蒙古自治区
5	企业	重庆市	印发	工程	银川市	转发	年度
6	增强	百色市	宁夏回族自治区	科学技术委员会	创新奖	平台	评选
7	能力	加快	青少年	基金	改革	研究	单位
8	达州市	供给	办公厅	若干	贵州省	暂行办法	信息化
9	第四届	攀枝花市	重庆市	改革	团队	计划	专项资金
10	获奖	条例	科学技术	广西	包头市	试行	创新

“河南省”“通知”“创新”是中部地区最主要的三个主题关键词,“河南省”下的内容词项集中突出了区域科技创新政策的要点,如“科技”“创新”“计划”等等。河南省在我国中部地区中起着至关重要的指引作用。第二个主题关键词是“通知”,内容词项包含很多对区域科技创新政策落实而言起到关键作用的地区和部门,如“安徽省”“山西省”“湖南省”“科技厅”“办公厅”等等。第三个主题关键词是“创新”,内容词项包含“加快”“创业”“支持”“开展”“试行”“实施方案”“改革”等具有指导性的词汇。另外从这些词汇也可以看出,区域科技创新政策更趋于引导、鼓励和支持而非强制,说明中部地区的政策体系正在逐渐走向完善。

“通知”“重庆市”“人民政府”是西部地区强度排名前三的主题关键词。其中主题关键词“通知”下的内容词项包含“科技”“创新”“科学技术”等要点,也包含“人民政府”“办公厅”这类政策执行和落实单位和政策执行和试点地区,如“重庆市”和“宁夏回族自治区”。第二个主题关键词“重庆市”代表的是西部地区政策执行和响应最快的地区,其内容词项反应了政策落实到具体地区的要点,如“实施方案”“规划”“信息化”“专项

资金”“创新”。第三个主题关键词是“人民政府”,其内容词项反应了人民政府为推进区域科技创新政策的落实所开展的具体工作,涉及“申报”“改革”等各个方面,并对相关“基金”支持进行了描述。

4.3 主题关键词内容词项及相关度分析

上部分对三大经济区各自的主题关键词及其内容词项进行了列示,本部分将对内容词项及相关度进行数据列示,但由于全部列示数据量非常大,而且容易将关键信息埋没在大量的数据中,因而本文只选取部分较为重要的关键词(上一部分中分别选取的东中西三大经济区的前三个主题关键词)及其内容词项的相关度予以说明,具体见表7列示。

表7是东中西三大经济区主题关键词及其内容词项的相关度汇总表,下面逐区进行分析。

东部地区的第一个关键词是“创新”,与之相关度最高的内容词项是“科技”,相关度0.116 519,说明东部地区的区域科技创新政策在创新方面多强调科技创新。与“创新”的相关度排名第二的内容词项是“创新”本身,关联度0.100 688。第三名是“通知”,关联度0.075 234。说明区域科技创新政策的创新意图多以通知方式自上而上传达。排在后面的内容词项还有“印发”“人民政府”“委

表 7 东中西三大经济区主题关键词及其内容词项的相关度汇总表

区域	主题关键词		主题关键词		主题关键词	
	创新		委员会		科技	
东部	内容词项	相关度	内容词项	相关度	内容词项	相关度
	科技	0.116 519	暂行办法	0.026 040	宁波市	0.024 873
	创新	0.100 688	财政厅	0.023 313	合作	0.021 964
	通知	0.075 234	国际	0.017 860	科学技术	0.021 964
	印发	0.028 362	上海市	0.017 860	信息化	0.019 055
	人民政府	0.025 671	优秀	0.016 496	徐州市	0.017 600
	委员会	0.023 395	技术	0.015 133	任务	0.016 145
	年度	0.021 636	市长	0.015 133	情况	0.013 236
	深圳市	0.021 119	政策措施	0.015 133	山东省	0.013 236
	上海市	0.018 946	促进	0.013 770	黑龙江省	0.011 782
	项目	0.017 290	林业	0.013 770	精神	0.011 782

续表

区域	主题关键词		主题关键词		主题关键词	
	主题关键词		主题关键词		主题关键词	
	河南省		通知		创新	
	内容词项	相关度	内容词项	相关度	内容词项	相关度
中部	科技	0.127 186	办公厅	0.056 437 8	加快	0.045 148 9
	创新	0.113 196	南昌市	0.039 335 5	创业	0.038 744 8
	通知	0.086 039	全省	0.034 449 1	支持	0.038 744 8
	人民政府	0.060 939	节能	0.029 562 7	开展	0.035 542 7
	河南省	0.041 188	江西省	0.027 119 5	试行	0.035 542 7
	印发	0.034 605	安徽省	0.024 676 3	实施方案	0.025 936 6
	办公室	0.023 495	山西省	0.024 676 3	改革	0.022 734 6
	年度	0.022 261	湖南省	0.022 233 1	江西省人民政府	0.022 734 6
	计划	0.017 734	科技厅	0.022 233 1	焦作市	0.022 734 6
	湖北省	0.016 089	农业	0.022 233 1	决定	0.022 734 6
	通知		重庆市		人民政府	
	内容词项	相关度	内容词项	相关度	内容词项	相关度
西部	科技	0.134 881	实施方案	0.039 020	财政厅	0.057 556
	创新	0.119 949	规划	0.029 345	加快	0.050 825
	通知	0.089 634	失效	0.029 345	申报	0.044 093
	人民政府	0.058 866	内蒙古自治区	0.022 896	教育厅	0.037 361
	印发	0.043 482	年度	0.022 896	工程	0.033 995
	宁夏回族自治区	0.027 193	评选	0.022 896	科学技术委员会	0.030 629
	青少年	0.023 121	单位	0.019 671	基金	0.027 264
	办公厅	0.019 954	信息化	0.019 671	若干	0.023 898
	重庆市	0.019 049	专项资金	0.019 671	改革	0.020 532
	科学技术	0.018 596	创新	0.016 446	广西	0.020 532

员会”“年度”“深圳市”“上海市”“项目”，关联度从 0.028 362 下降到 0.017 290，由于篇幅所限，对于这些内容词项不再详述。东部地区的第二个主题关键词是“委员会”，与其相关度最高的内容词项是“暂行办法”，相关度 0.026 040；相关度第二的是“财政厅”，相关度 0.023 313；“国际”“上

海市”“委员会”相关度并列第三位，为 0.017 860，另外还有与之关系较密切的内容词项包括“优秀”“技术”“市长”“政策措施”“促进”“林业”。东部地区第三个主题关键词是“科技”，与之相关度最高的前三项内容词项为“宁波市”“合作”“科学技术”，相关度分别为 0.024 873、0.021 964 和

0.021 964。另外还有“徐州市”“任务”“情况”。“山东省”“黑龙江省”“精神”与之相关度也较高。

中部地区第一个主题词是“河南省”，与之相关度最高的内容词项是“科技”，相关度 0.127 186；相关度次高的是“创新”，相关度 0.113 196，足以见科技创新在中部地区的重要地位。第三位是“通知”，相关度 0.086 039，另外还有“人民政府”“河南省”“印发”“办公室”“年度”“计划”“湖北省”与“河南省”关系密切。第二个主题关键词是“通知”，与之相关的前三项内容词项分别是“办公厅”“南昌市”“全省”，相关度分别是 0.056 437 8、0.039 335 5 和 0.034 449 1，主要涉及各区域和相关方面，另外的内容词项还包括“节能”“江西省”“安徽省”“山西省”“湖南省”“科技厅”“农业”。中部的第三个主题关键词是“创新”，相关度排名前三位的内容词项包括“加快”“创业”“支持”，相关度分别是 0.045 148 9、0.038 744 8 和 0.028 744 8。另外还有“开展”“试行”“实施方案”“改革”“江西省人民政府”“焦作市”“决定”与之关系密切。从内容词项可以看出，与主题关键词“创新”相关的内容词项较多集中在具体工作开展和落实方面，印证了前文中提到的中部地区的创新政策较详细，对具体工作的指导和支持作用较强。

西部地区的第一个主题关键词是“通知”，与之相关的前三项内容词项分别是“科技”“创新”“通知”，相关度分别是 0.134 881、0.119 949 和 0.089 634。说明西部地区的区域科技创新政策也以通知为主，且非常强调科技和创新，另外的相关性较强的内容词项还包括“人民政府”“印发”“宁夏回族自治区”“青少年”“办公厅”“重庆市”“科学技术”几项内容。西部地区第二项主题关键词是“重庆市”，与之相关性最强的两个内容词项是“实施方案”“规划”，相关度分别为 0.039 020 和 0.029 345，说明西部地区的区域科技创新政策与重庆市相关的政策非常重视政策的具体实施方案以及规划。排名第三的内容关键词是“失效”，作为标题是将以前的区域科技创新政策进行废止，并执行新的政策，说明西部地区非常注重区域科技创新政策与现实情况适应，对已经不适合具体环境的政策予以更新。另外与重庆市相关性较高的内容词项还包括“内蒙古自治区”“年度”“评选”“单位”“信息化”“专

项资金以及创新”。西部地区的第三个主题关键词是“人民政府”，与之相关性最高的内容词项是“财政厅”，相关度 0.057 556；紧随其后的是“加快”和“申报”，相关度分别是 0.050 825 和 0.044 093。主要体现了对人民政府相关工作的要求，如加快审批流程、简化申报手续等行政支持。另外的内容词项还有“教育厅”“工程”“科学技术委员会”“基金”“若干”“改革”“广西”，主要内容都是侧重于对人民政府的要求。这说明我国的区域科技创新政策不仅在实施形式上从强制政策转向鼓励、支持和引导的政策，而且更加强调地方政府及相关部门对政策的落实所起的重要作用，逐步向服务型政府转变。

5 研究结论

根据实证结果可以得到以下结论：

(1) 我国东部、中部和西部三大经济区的区域科技创新政策各有特点。东部凭借资源优势已经将科技创新作为推动区域发展的主要驱动力，同时对科学技术、信息化以及合作等非常重视，在全国范围中，东部地区对中部和西部地区起着重要的指引作用。从主题关键词来看，东部地区的创新活动主要依靠政策受众自发响应区域科技创新政策，已经不再主要依靠人民政府的指引作用。中部地区和西部地区对创新和科技要点也较为重视，但程度较东部地区略低，这是由地区具体情况决定的。中部地区的主题关键词中包含“创新”，这证明中部地区对创新非常重视。中部地区的另一个特点就是多次提到各个省市，说明中部地区的政策非常具体，较有针对性，地方政府服务意识较强。在主题关键词的划分上，西部地区对创新重视程度稍弱，政府政策及相关部门的具体行为依然是推动区域发展的最主要动力，其主题关键词中依然存在“人民政府”。

(2) 试点仍是各地区常用的政策推广方式。不论从三大经济区的主题关键词来看，还是从东中西各区域的具体情况来看，区域科技创新政策实证结果中出现了非常多的地区，这说明试点依然是政策推行的一种主要形式，选择本区域或本地区较有代表性的地区进行政策试点，有助于政策由点到面的推广，也有利于积累优秀经验，避免人力的浪费。

(3) 各地方政府正在向服务型政府转变且初见成效。从政策的强度和与政策相关主题关键词的内容词项可以看出,我国地方政府已经逐步实现了服务型政府的转变。在政策强度上较多地采用鼓励、支持和引导的方式,不再由上而下强制执行;在内容词项上多出现“加快”“支持”“基金”“促进”等字眼,而很少有体现强制性的要求和规定。中央政府和地方政府正在千方百计调动政策受众的积极性,并在各方面予以支持,来提高大众创业、万众创新的积极性,也在大力落实服务型政府的切实转变。

6 研究展望

本研究通过对我国三大经济区的1506项区域科技创新政策标题进行挖掘,对词频、主题关键词强度、内容词项及相关性 etc. 进行了深入分析,研究发现科技创新已经受到重视,而且会成为未来很长一段时间的区域发展的核心驱动力,而我国政府也在逐步实现服务型政府的转变。下一步的研究将会集中在从时间维度纵向观察政策演变的进程,拟采用仿真方式将服务型政府的特征予以实现,模拟和比较服务型政府与强制型政府政策执行效果。■

参考文献:

- [1] 观察者网. 全球独角兽企业版图: 中国超欧洲 美国地位动摇 [EB/OL]. [2019-08-31]. <https://www.myzaker.com/article/59e3f4aa9490cbb50a00000f/>.
- [2] Rothwell R, Zegveld W. Reindustrialization and technology[M]. Longman: M.E. Sharpe, 1985: 13-37.
- [3] Boschma R A. Regional Innovation Policy[J]. European Planning Studies, 2005, 13(4): 495-496.
- [4] Freitas I M B, Tunzelmann N V. Mapping public support for innovation: A comparison of policy alignment in the UK and France[J]. Research Policy, 2008, 37(9): 1 446-1 464.
- [5] Schramm M E, Trainor K J, Shanker M, et al. An agent-based diffusion model with consumer and brand agents[J]. Decision Support Systems, 2010, 50(1): 234-242.
- [6] Carreira C, Teixeira P. Entry and exit as a source of aggregate productivity growth in two alternative technological regimes[J]. Structural Change & Economic Dynamics, 2011, 22(2): 135-150.
- [7] 李晨光, 张永安. 集群创新科技政策“效用—响应”测度分类研究[J]. 中国科技论坛, 2012, 28(10): 24-30.
- [8] 张永安, 鄢海拓, 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, 33(3): 104-110.
- [9] 贺德方, 唐玉立, 周华东. 科技创新政策体系构建及实践[J]. 科学学研究, 2019, 37(1): 3-10.
- [10] Green D G, Sadedin S. Interactions matter—complexity in landscapes and ecosystems[J]. Ecological Complexity, 2005, 2(2): 117-130.
- [11] 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, 24(9): 25-36.
- [12] 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, 27(4): 6-9.
- [13] Estrada M A R. Policy modeling: Definition, classification and evaluation[J]. Journal of Policy Modeling, 2011, 33(4): 523-536.
- [14] 张永安, 鄢海拓. 国务院创新政策量化评价——基于PMC指数模型[J]. 科技进步与策, 2017, 34(17): 127-136.
- [15] 盛亚, 孙津. 我国区域创新政策比较——基于浙、粤、苏、京、沪5省(市)的研究[J]. 科技进步与对策, 2013, 30(6): 99-103.
- [16] 张炜, 费小燕, 方辉. 区域创新政策多维度评价指标体系设计与构建[J]. 科技进步与对策, 2016, 33(1): 142-147.
- [17] 张永安, 马昱. 基于R语言的区域技术创新政策量化分析[J]. 情报杂志, 2017, 36(3): 113-117.
- [18] 王卉彤, 韩天实. 全国重点区域创新创业金融服务政策量化分析与对比研究[J]. 经济研究参考, 2018, 40(44): 32-41.
- [19] Prior L, Hughes D, Peckham S. The Discursive Turn in Policy Analysis and the Validation of Policy Stories[J]. Journal of Social Policy, 2012, 41(2): 271-289.
- [20] 杨慧, 杨建林. 融合LDA模型的政策文本量化分析——基于国际气候领域的实证[J]. 现代情报, 2016, 36(5): 71-81.
- [21] Blei D M. Latent dirichlet allocation[J]. Journal of Machine Learning Research, 2003, 3(4): 993-1 022.

Quantitative Evaluation of Regional Science and Technology Innovation Policy of China

QIE Hai-tuo, ZHANG Zhi-juan

(Institute of Scientific and Technical Information of China, Beijing 100038)

Abstract: The purpose of this study is to quantify the regional science and technology innovation policy based on 1 506 regional science and technology innovation policy headlines. It analyzes the innovation policies of the three major economic zones of China from the aspects of high frequency words, key words, content words and correlation. Based on the R language and text mining, the titles of the 1 506 regional science and technology innovation policies of China in 1999-2017 are deeply analyzed with the help of the theme model. The results show that there are different characteristics of regional science and technology innovation policy in the three major economic zones of China. The pilot is still the most commonly used policy promotion method in various regions, and local governments in China are transforming to service-oriented government and have achieved initial success.

Key words: LDA theme model; theme key words; science and technology innovation policy; R language; policy quantification

欢迎订阅《全球科技经济瞭望》

《全球科技经济瞭望》(中文版)为月刊,每月30号出版,每期定价20.00元,全年240.00元(含邮资),邮发代号:82-537;

汇款方式:银行汇款

户名:中国科学技术信息研究所(请务必注明《全球科技经济瞭望》订刊费)

帐号:0200232109200900593

开户行:中国工商银行北京玉渊河潭支行

联系电话:(010)68511566

联系人:赵君虹

Fax:(010)68511566

E-mail:liaowang69@126.com

主页:<http://www.kjliaoawang.com.cn>