高技术通讯 2021 年 第31卷 第11期:1126-1135

doi:10.3772/j.issn.1002-0470.2021.11.002

基于文本挖掘的国家创新型城市科技创新政策量化评价研究①

郄海拓② 陈雪迎 张志娟

(中国科学技术信息研究所 北京 100038)

摘 要 选取72个国家创新型城市作为研究对象,依据城市的经济发展能力和创新能力将样本划分为4个象限,形成包含全样本和4个象限在内的5个样本组。对各样本组2019年的科技创新政策进行文本挖掘,主要包括高频词、语义网络图、点度中心度、网络中心势和小团体分析。研究结果表明,我国国家创新型城市科技创新政策的重点突出,主要依托项目和企业开展研发活动,以实现科技和技术的创新,同时需要重视多渠道激励创新绩效的提升。本研究为政策优化与政策决策提供了参考和依据。

关键词 国家创新型城市;科技创新政策;政策量化;政策评价;文本挖掘

0 引 言

科技创新政策量化评价是政策评价的重难点问题,将文本数据转化成精准、客观、易读性和可读性强的客观数据和关系并识别政策的重点,同时挖掘核心关系并透视政策文本背后隐藏的关键信息具有研究价值。我国试点创新型城市形成的创新型城市发展创新网络为我国区域创新驱动发展提供了有力支撑[1],城市的科技创新政策对城市的建设和发展起到了引领作用。对国家创新型城市的科技创新政策进行研究有助于优化政策,使政策更好地为城市发展提供引导和支持。

1 研究背景和意义

科技部和国家发改委共支持了78个国家创新型城市,其中72个为地级市,其他为4个直辖市和2个县级市。国家创新型城市占据全国约八成的R&D经费支出,并拥有85%以上的有效发明专利[2],为创新型国家建设做出了很大贡献。国家创新型城市建设需要城市科技创新政策的引导和支持,因而实现政策优化、支持政策合理地进行立改废决策,成为国家创新型城市建设的重要一环。本研究基于文本挖掘,对国家创新型城市的科技创新政策予以挖掘和透视,分析隐藏在政策文本背后的关键信息,以支持政策的立改废决策。

为保证样本的一致性,本研究选取78个国家创新型城市中的72个地级市作为研究对象并以2019年科技创新政策作为研究范围,对国家创新型城市的科技创新政策进行文本挖掘,拟基于政策原文件挖掘和透视文本中隐含的关系与网络分布,使原本隐藏在大量文本背后的关系和数据得以清晰展现。除了对全样本进行文本挖掘,本研究还基于城市的经济发展能力和创新能力将72个样本划分到4个象限中,研究不同象限样本的科技创新政策特点。文本挖掘方法的运用,有助于将非结构化的文本转化为结构化的数据,对政策量化和政策评价以及基于此进行的政策立改废有重要的借鉴意义。

本研究的主要内容有两个:一是通过基于对原政策文件的文本挖掘更加深入地了解、评价和量化科技创新政策,加深政策研究的深度和客观程度;二是通过对政策的客观评价与量化,分析不同样本组

① 北京市科技计划(Z151100002615007)和中央级公益性科研院所基本科研业务费专项基金青年项目(QN2020-06)资助。
② 女,1988年生,博士;研究方向:国家创新型城市,科技创新政策,企业创新;联系人,E-mail:qiehaituo@sina.com。
(收稿日期:2020-12-30)

科技创新政策的不同特点并进行比较研究,最终对城市科技创新政策的立改废决策提供依据。

从理论意义角度来说,基于文本挖掘的国家创新型城市的科技创新政策量化评价综合采用 ROSTCM、Ucinet、Python 和 Gephi 软件实现了政策的深度挖掘,立足政策原文本的研究,增强了研究结论的客观性。另外,该量化评价还拓展了文本挖掘在政策研究领域的应用范围。从现实意义角度来说,研究得到的高频词、网络性、中心性和小团体分析有助于识别政策要点,加深对政策的理解,为政策制定者的立改废决策与政策受众更加深入地理解政策都有一定的现实意义。

2 科技创新政策量化评价文献综述

由于政策文本的数据形式比较特殊,最初常采用专家访谈[3]实现政策量化,虽然这种方法简单高效但难以避免主观性对结果的影响。文献[4]提出了用于政策量化的三维度操作手册,从政策力度、目标与措施维度实现政策量化评价。文献[5]基于内容分析从政策目标与政策工具角度法实现了区块链政策文本的量化。文献[6]提出 PMC 指数(policy modeling consistency index)法用于政策量化评价并被国内学者应用到我国科技创新政策的研究中[7-8]。文献[9]在此基础上提出衡量待评价政策与理想的完美政策间差异水平的“PMC 凹陷指数”来测度待评价政策的改进空间。文献[10]从政策工具角度构建基于政策工具与科技产业价值链的量化分析框架,实现了政策量化。在政策评价维度方面,文献[11]从政策目标、政策工具与政策力度 3 个维度评价我国高校创新创业政策。文献[12]基于政策工具视角,从 X、Y、Z 3 个维度评价双创政策,其中 X 维度包含供给型、环境型以及需求型政策,Y 维度包含技术布局、产业布局以及应用布局,Z 维度则包括了创新研发、创新成果转化与产业化以及创新推广与服务。

从总体趋势来看,科技创新政策量化是从主观逐渐走向客观的过程。科技创新政策量化评价的总体思路基本上都是首先基于一个评价框架,然后采用一定方法将政策文本与评价框架的关键要素予以对应,最后实现量化。评价框架的选择往往基于已有研究的经典框架,结合具体研究对象的具体情况进行设定,以更加契合研究对象的特征进而做出客观合理的评价。科技创新政策量化评价的下一步研究常常是研究政策与创新绩效之间的关系,如文献[5,13-14]的研究,因而客观合理地对政策进行量化评价是研究政策与创新之间关系的重要基础。

3 研究主体

对研究主体做简要描述,如图 1 所示。

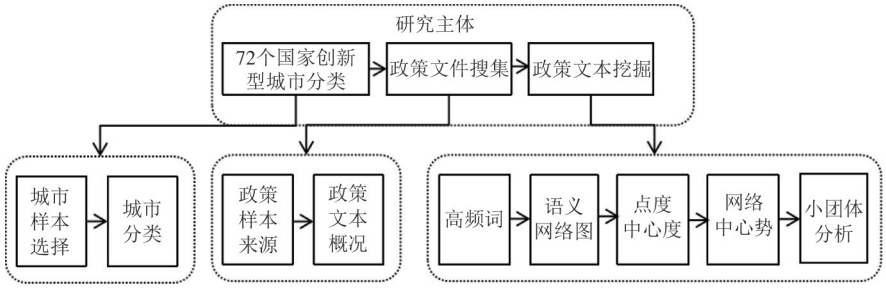

图 1 研究主体图

研究主体分为 72 个国家创新型城市分类、政策文本搜集和政策文本挖掘 3 个部分。第 1 个部分又分为城市样本选择与城市分类;第 2 部分分为政策样本来源和政策文本概况;第 3 部分对政策样本产生的高频词、语义网络图、点度中心度、网络中心势和小团体分析进行了各样本组的详细说明与比较。

3.1 72 个国家创新型城市分类

3.1.1 城市样本选择

本研究选取了 78 个国家创新型城市中的 72 个地级市,将北京市海淀区、上海市杨浦区等去掉,留

下《国家创新型城市创新能力评价报告》和《国家创新型城市创新能力监测报告》中进行完整评价的 72 个城市，即深圳、广州、杭州、南京、武汉、苏州、西安、长沙、成都、青岛、厦门、无锡、合肥、济南、宁波、常州、大连、郑州、沈阳、南昌、镇江、东莞、嘉兴、贵阳、昆明、芜湖、太原、福州、南通、湖州、佛山、哈尔滨、烟台、扬州、长春、泰州、洛阳、马鞍山、绍兴、兰州、株洲、潍坊、石家庄、徐州、南宁、金华、连云港、乌鲁木齐、海口、盐城、银川、秦皇岛、宜昌、东营、襄阳、呼和浩特、景德镇、济宁、泉州、唐山、西宁、龙岩、包头、宝鸡、拉萨、衡阳、萍乡、遵义、玉溪、南阳、吉林和汉中。

3.1.2　城市分类

按照人均 GDP 与《国家创新型城市创新能力评价报告》(2019 版)中国家创新型城市创新能力指数为划分象限的 2 个维度，分别以 50 百分位点为界，将 72 个城市划分到 4 个象限，方便后续进行比较研究，象限划分见图 2。

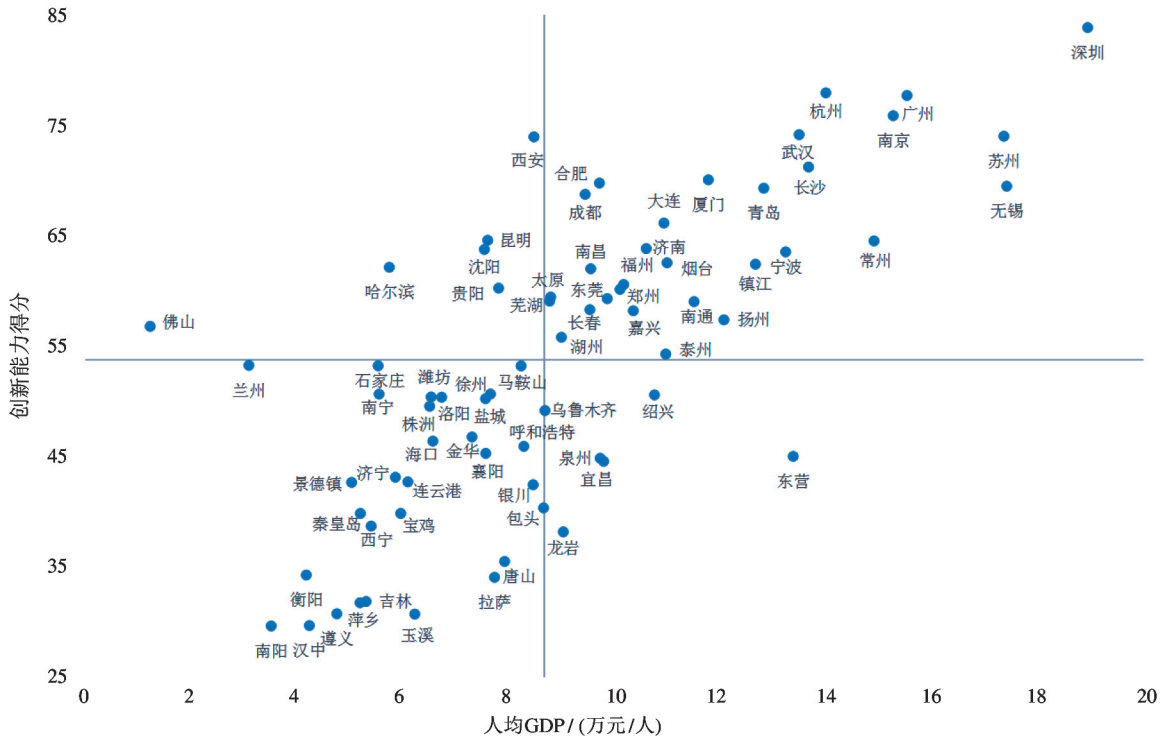

图 2　创新型城市四象限分类图

对 4 个象限的特点进行汇总，可以设定为：第一象限(高经济发展能力，高创新能力)；第二象限(低经济发展能力，高创新能力)；第三象限(低经济发展能力，低创新能力)；第四象限(高经济发展能力，低创新能力)。由象限划分结果可以看出，处于第一象限和第三象限的城市均有 30 个，样本数量较大，说明处于(高经济发展能力，高创新能力)和(低经济发展能力，低创新能力)两种状态的城市数量较多。

从样本城市来看，第一象限的城市是创新状态较优的，特别是处于右上的深圳，是资源和创新能力双优的城市。从象限来看，应实现从第二、三、四象限城市向第一象限的转移，如第二象限(低经济发展能力，高创新能力)的优化路径就是基于较强的创新能力提升经济发展能力实现象限转移，而第三象限(低经济发展能力，低创新能力)则有 2 条优化路径，分别是提升经济发展能力—提升创新能力和提升创新能力—提升经济发展能力。第四象限若实现象限转移则需要通过依靠自身高经济发展能力的优势提升创新能力。而在第一象限中城市优化的思路与象限转移的思路一致，即基于自身的经济发展能力条件和城市创新能力现状向整个图像的右上角进行转移。

3.2 政策文件搜集

3.2.1 政策样本来源

本研究选取72个国家创新型城市2019年的科技创新政策,来源主要包括城市人民政策官网、科技局、人力资源和社会保障局、知识产权局、工业和信息化局、财政局、发改委、金融监管局、经信委、北大法宝等网站。

3.2.2 政策文本概况

以上网站的科技创新政策经整理筛选形成政策文本全样本库,然后根据象限分类结果形成4个象限的分样本库。

政策数量排名前3的城市是南京、广州和长沙,从象限分布来看这3个城市均处于第一象限,科技创新政策数量体现城市对创新的重视程度,因而政策数量对于城市创新能力与城市经济发展均存在正向影响,可以作为研究下一步推进的方向。

对政策样本进行挖掘,进一步得到非结构化的文本数据背后隐藏的结构化数据与关系。

3.3 政策文本挖掘

3.3.1 高频词

首先对高频词进行挖掘和列示,为了兼顾高频词反映的信息全面程度和重点突出程度,研究选取30个高频词进行汇总,剔除没有实际意义的词,例如“非常”等词。其次,在高频词筛选中,由于共现关系较弱,排除了第二象限的科研诚信(词频169)和第三象限的孵化器(词频438)、基金(词频426)和团队(词频343),高频词汇总结果如表1所示。

表1 5个样本组高频词情况汇总表

	总样本	词频	第一象限	词频	第二象限	词频	第三象限	词频	第四象限	词频
1	项目	20 436	项目	13 151	项目	1967	企业	4358	项目	1779
2	企业	17 659	企业	10 997	企业	1129	项目	3539	企业	1175
3	服务	7880	服务	4905	基金	822	服务	2106	科技	468
4	技术	7313	技术	4606	资金	727	知识产权	1720	奖励	450
5	资金	6528	资金	4428	技术	601	科技	1682	资金	428
6	科技	5809	科技	3085	科技	574	技术	1680	技术	426
7	知识产权	4764	研发	3011	服务	464	创新	983	服务	405
8	创新	4584	创新	2991	创新	373	资金	945	补助	374
9	研发	4297	人才	2695	专家	328	研发	902	创业	348
10	人才	4061	知识产权	2572	风险	277	创业	882	知识产权	326
11	基金	4061	奖励	2549	资助	248	人才	870	机构	324
12	奖励	3659	基金	2529	贷款	240	专利	643	专家	305
13	创业	3433	资助	2408	科技创新	210	平台	578	专利	301
14	资助	3152	创业	2026	财政	222	体系	546	人才	297
15	补助	2815	领域	1997	补助	203	奖励	531	基金	284
16	专家	2786	专家	1832	人才	199	工业	527	资助	257
17	领域	2776	平台	1803	领域	181	领域	502	研发	246
18	平台	2610	补助	1801	创业	177	科技创新	491	经费	240
19	财政	2315	贷款	1606	政府	167	资源	485	创新	237
20	贷款	2303	财政	1522	经费	165	机制	463	贷款	155
21	经费	2176	补贴	1346	资源	165	政府	449	财政	172
22	专利	2146	经费	1336	立项	149	园区	440	科研	143
23	政府	1959	绩效	1235	知识产权	146	补助	437	绩效	125
24	科技创新	1890	政府	1223	科研	144	经费	435	政府	120
25	资源	1843	专利	1165	银行	143	金融	417	团队	120

续表 1

26	机制	1777	专项资金	1151	孵化器	139	科研	411	中小企业	119
27	绩效	1773	资源	1120	平台	139	财政	399	开发区	114
28	体系	1720	科技创新	1092	信贷	138	制度	375	金额	112
29	补贴	1717	机制	1086	研发	138	成果	362	补贴	103
30	工业	1674	工业	1049	科技企业	134	专家	321	大学生	101

从表 1 中可以看出，全样本和第四象限的高频词中“项目”、“企业”、“服务”、“技术”、“科技”和“创新”普遍是高频词中排名靠前的词，说明在各城市组成的样本组的科技创新政策中往往会突出上述相关内容，而从之后的高频词中可以看出科技创新政策的其他重点领域，如人才、补助、知识产权和研发等同样在城市的科技创新系统中非常重要。

3.3.2　语义网络图

基于政策文本进行进一步挖掘，语义网络图反映的是政策样本中高频词的关系与具体分布情况，其中词间连线代表的是两端的高频词同时出现在政策样本中。语义网络图能够将政策样本整体的分布情况以直观的方式体现出来，有助于将非结构化的文本转变成直观图像，挖掘文本背后的隐含关系与整个网络的分布态势。本研究中全样本和 4 个分样本的语义网络图如图 3 ~ 图 7 所示。

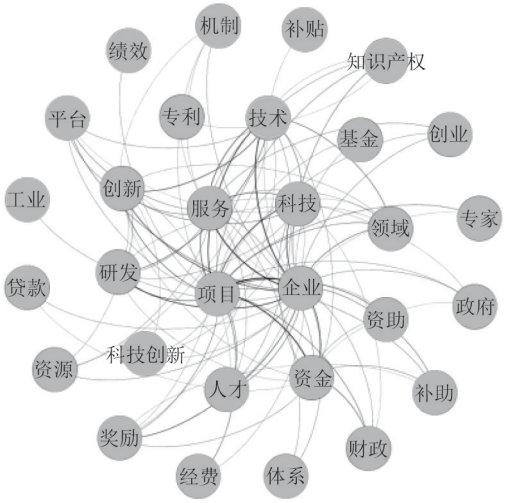

图 3　全样本语义网络图

从全样本来看，项目、企业、科技、服务、创新、研发、技术和资金是处于网络核心位置的高频词，且这些高频词之间的联系非常紧密，体现了科技创新政策总体的重点内容。从连线的情况来看，企业与项目、项目与资金、企业与服务、企业与技术、项目与创新、项目与研发等之间的关系非常紧密。这说明从总体来看，科技创新政策的重点主要集中在以企业和项目为依托的范围内的科技创新与技术研发，同时需要资金的支持，其语义网络图见图 3。

第一象限的语义网络图中处于中心的是企业、项目、资金、服务、创新、技术和科技，联系相对紧密的关键词是项目与资金、企业与项目、企业与服务、项目与技术、资金与企业、企业与研发以及企业与科技等。反映了第一象限（高经济发展能力，高创新能力）城市发展的重点在于依托企业和项目开展的研发活动，并且关注资金对于企业和项目的支持作用，其语义网络图见图 4。

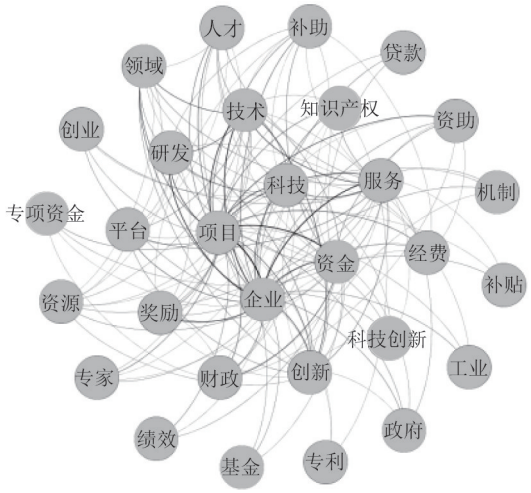

图 4　第一象限语义网络图

第二象限的语义网络图见图 5，以企业、项目、资金、创新、科技、技术为中心，其中联系较紧密的高频词包括项目与企业、服务与企业、技术与创新、企业与技术、企业与科技、项目与科技、项目与资金等。说明第二象限（低经济发展能力、高创新能力）城市发展重点在于基于项目和企业的科技创新和资金利

用,对围绕企业和项目提供服务是非常重要的。

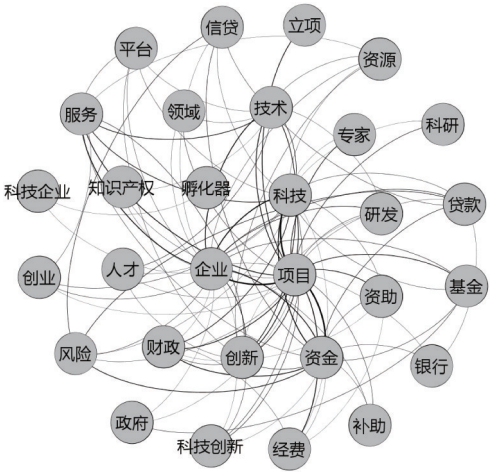

图 5　第二象限语义网络图

第三象限以企业、项目、服务、技术、研发、创新、科技为中心,其中联系较紧密的高频词包括企业与技术、企业与研发、企业与服务、技术与研发、技术与创新、企业与创新和企业与项目等。这说明第三象限(低经济发展能力、低创新能力)的城市发展重点在于基于项目和企业的技术创新与研发,其语义网络图见图 6。

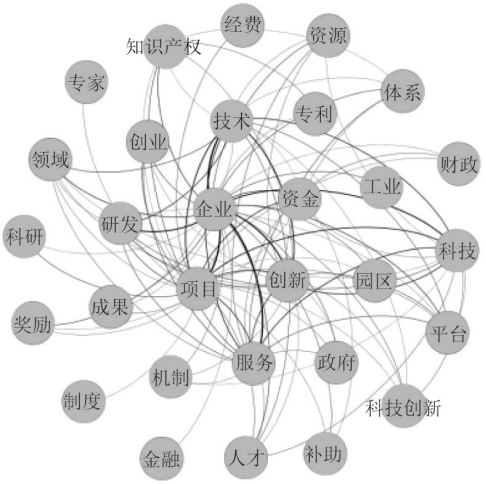

图 6　第三象限语义网络图

第四象限以项目、企业、资金、科技、技术、服务和奖励为中心,其中联系较紧密的高频词包括企业与奖励、项目与奖励、科技与项目、资助与奖励、企业与项目、企业与服务等。这说明第四象限(高经济发展能力、低创新能力)的城市发展重点在于基于项目和企业的科技与技术创新,且非常注重与之相关的资助与奖励,其语义网络图见图 7。

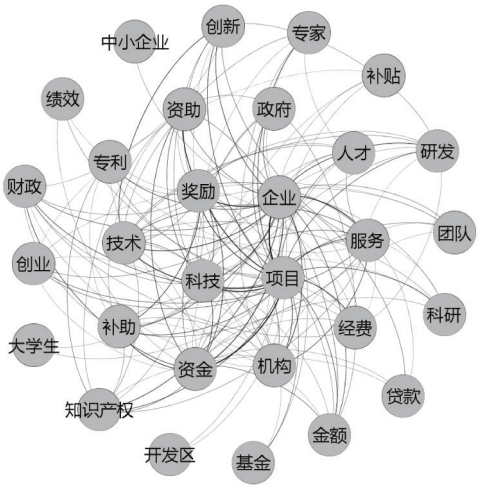

图 7　第四象限语义网络图

5 个样本组语义网络图的中心和高频词之间联系略有不同,但企业和项目以及与之相关的科技、创新、技术、服务、研发和资金总体来说是网络图共同的重点内容。

3.3.3　点度中心度

点度中心度反映的是特定高频词在网络中的重要程度,即与其他高频词的联系紧密程度,点度中心度的取值越高则代表该高频词(节点)在网络中的重要程度越高[15],在图中的节点越大、颜色越深。点度中心度的态势可以用图形表示,也可以用数据列示。全样本点度中心度图见图 8。从全样本点度中心度图中可以看出,企业、项目、服务、技术和资金是网络中重要性排名前 5 位的高频词。

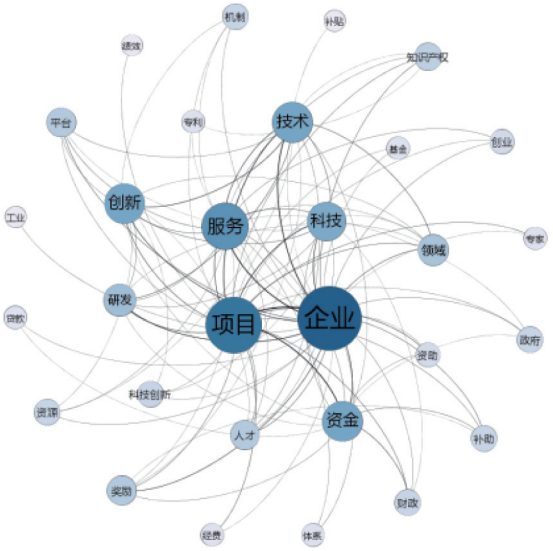

图 8　全样本点度中心度图

从第一象限点度中心度图(图 9)中可以看出,企业、项目、服务、技术和资金是网络中重要性排名前 5 位的高频词,与全样本情形非常相似,只是在具体数值上略有差别。

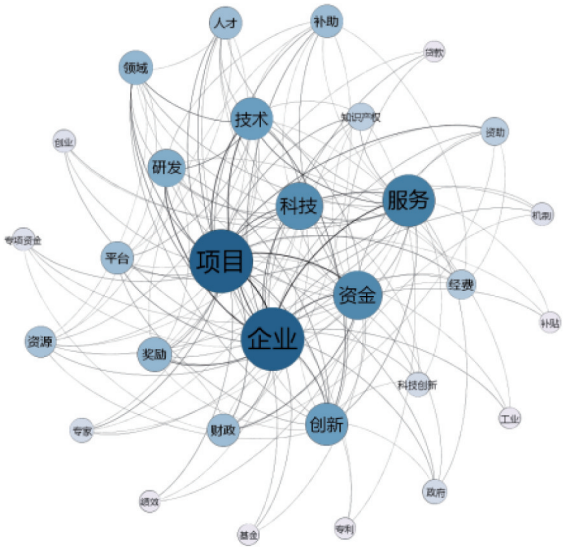

图 9　第一象限点度中心度图

从第二象限点度中心度图(图 10)中可以看出,项目、企业、科技、资金和创新是网络中重要性排名前 5 位的高频词,与第一象限情况略有差别,项目是最重要的高频词,科技的重要程度更为突出。

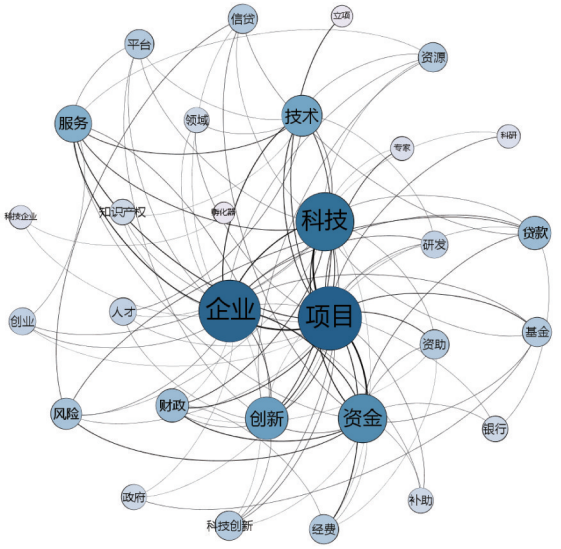

图 10　第二象限点度中心度图

从第三象限点度中心度图(图 11)中可以看出,企业、项目、服务、技术和创新是网络中重要性排名前 5 位的高频词,相比较而言服务在该象限的重要性尤为突出。

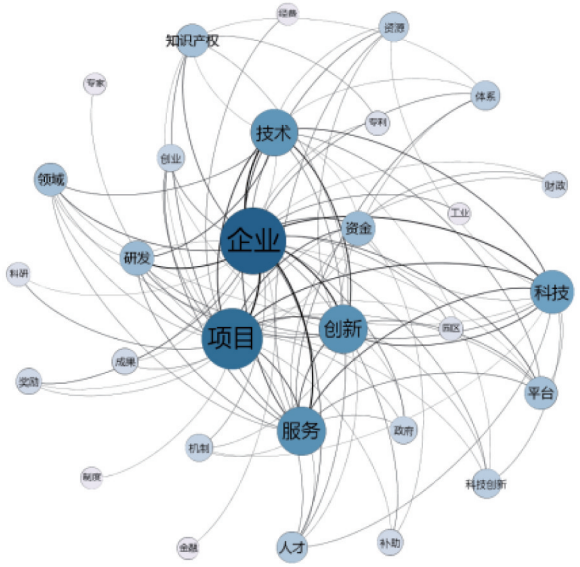

图 11　第三象限点度中心度图

从第四象限点度中心度图(图 12)中可以看出,项目、企业、资金、奖励和科技是网络中重要性排名前 5 位的高频词,相比较而言资金与奖励的重要性较为突出。

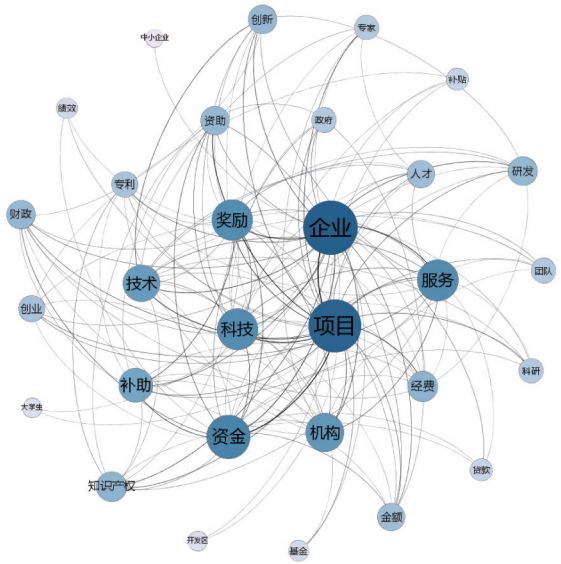

图 12　第四象限点度中心度图

从 4 个象限的点度中心度结果展示的情况可见,企业与项目以及围绕其的相关服务、资金、科技、创新、技术和研发是科技创新政策中重点关注的问题。

3.3.4　网络中心势

对 5 个网络的网络中心势进行测算,来了解各

网络的集中程度,网络中心势取值处于0% ~100%之间,网络中心势超过10%的网络即体现为一定的集中度,网络中心势体现的只是网络集中的程度,并不是优劣的绝对评价标准。5个样本组的网络中心势汇总如表2所示。

表2 网络中心势汇总表

序号	样本	网络中心势
1	全样本	31.03%
2	第一象限	28.71%
3	第二象限	28.79%
4	第三象限	36.26%
5	第四象限	29.00%

从网络中心势的取值来看,5个样本组的取值均高于28%,网络均存在一定的集中度。原因可能为城市按照国家-省-城市的主要路线作为制定城市创新科技政策的主要路径,因而城市间政策制定的思路和基础较为统一、语义表达准确、用词规范。另外,本研究在政策文本筛选阶段标准严格,选择的政策均为表达准确、重点突出的政策文件,对于一般性通知和没有重点内容的文件均在筛选阶段删除,因此样本质量较优。

从网络中心势的具体情况来看,第三象限的网络中心势取值最高,说明该象限内政策样本的集中程度较高,重点更为突出。而第二象限和第一象限的网络集中度略低,第二象限由于样本数较少,所以网络中心势取值较低;第一象限由于样本的城市创新能力和经济发展能力都较高,所以政策倾向多元化,重点相对分散。

3.3.5 小团体分析

小团体分析的结果有助于了解和透视隐藏在大量的文本信息后的关系,并以直观的方式体现出来。各样本组小团体分析的结果如表3所示,全样本的5个小团体分别是创新人才、企业创新、平台载体、创新生态和知识产权。

表3 全样本组小团体分析的结果表

小团体序号	小团体名称	小团体包含的关键词
1	创新人才	项目、科技、专家、人才、科技创新、绩效、资助、经费、基金
2	企业创新	企业、研发、技术、领域、工业、补贴、贷款
3	平台载体	服务、体系、创新、资源、创业、平台、机制
4	创新生态	资金、奖励、政府、补助、财政
5	知识产权	专利、知识产权

从表4中可以看出,第一象限的5个小团体分别是企业创新、创新人才、科技金融、创新生态和知识产权。

从表5中可以看出,第二象限的5个小团体分别是企业创新、创新生态、创新人才、科技金融和政府基金。

从表6中可以看出,第三象限的5个小团体分别是企业创新、创新人才、平台载体、创新机制和创新生态。

表4 第一象限小团体分析结果表

小团体序号	小团体名称	小团体包含的关键词
1	企业创新	企业、补贴、奖励、工业
2	创新人才	科技、平台、技术、研发、资源、人才、领域、创新
3	科技金融	项目、资金、贷款、经费、财政、基金、绩效、资助、补助、专项资金
4	创新生态	服务、科技创新、政府、机制、专家、创业
5	知识产权	专利、知识产权

表 5　第二象限小团体分析结果表

小团体序号	小团体名称	小团体包含的关键词
1	企业创新	企业、科技、孵化器、科技创新、创业、科技企业
2	创新生态	项目、财政、领域、专家、资助、科研、知识产权、经费、立项
3	创新人才	人才、创新、技术、服务、平台、研发、资源
4	科技金融	资金、信贷、风险、贷款、银行、补助
5	政府基金	基金、政府

表 6　第三象限小团体分析结果表

小团体序号	小团体名称	小团体包含的关键词
1	企业创新	企业、工业、创业、服务、专利、知识产权、政府、金融、体系
2	创新人才	项目、成果、经费、科研、专家、制度、园区、人才
3	平台载体	技术、平台、领域、研发
4	创新机制	创新、机制、资源、科技创新、科技
5	创新生态	资金、补助、奖励、财政

从表 7 中可以看出，第四象限的 5 个小团体分别是知识产权、创新生态、创新人才、技术研发和科技金融。

通过小团体分析的结果可知，4 个象限小团体均存在创新人才和创新生态小团体，说明创新人才和良好的创新生态是科技创新的重要因素。第三象限的样本对于科技金融对城市创新能力与经济发展能力的支持相对于其他样本来说稍弱，应予以重点关注。同理，第四象限对企业创新的关注程度稍弱，或者是对企业关注点略显分散，应予以重点关注；第二、三象限应关注知识产权领域。而从小团体间差异来看，第二象限的政府基金、第三象限的创新机制与平台载体和第四象限的技术研发都是这些象限各自关注的特殊领域，其他象限的城市也应以此为依据审视本城市的政策制定是否全面并突出重点，为城市科技创新政策的立改废提供依据。

表 7　第四象限小团体分析结果表

小团体序号	小团体名称	小团体包含的关键词
1	知识产权	企业、奖励、专利、贷款、补助、知识产权、大学生、开发区、中小企业
2	创新生态	项目、科技、资金、资助、科研、补贴、金额、政府、绩效
3	创新人才	服务、机构、团队、人才、基金、专家、创业
4	技术研发	技术、研发、创新
5	科技金融	财政、经费

4　结 论

基于 5 个样本组文本挖掘的结果可知，我国国家创新型城市的科技创新政策重点比较突出，主要依托项目和企业开展研发活动，实现科技和技术的创新发展，同时需要重视相关的支持服务和资金的支持与其他奖励来激励创新绩效的提升。在创新活动的落实方面来看，各组均认识到了创新生态和创新人才的重要性，并对创新活动相关的企业创新、科技金融、平台载体、知识产权等普遍有所关注，且在不同样本组体现出了一定差异。

对各样本组的科技创新政策进行文本挖掘有助于将文本背后隐藏的关键信息挖掘出来，并且摒弃研究人员的主观偏差，将非结构化数据转化为结构性的数据和关系，以透视政策的重点内容，也为城市科技创新政策的立改废提供了依据，为城市创新体

系建设提供了支持,为城市间相互借鉴提供了更清晰的思路和更突出的重点。

本研究下一步拟通过实证方法定量研究科技创新政策与创新型城市的经济发展和创新能力之间的关系,借此研究政策的效果、绩效与效率,并进一步为国家创新型城市的科技创新政策提供立改废决策支持。

参考文献

[1] 刘锴, 周雅慧, 王嵩. 城市科技创新效率与网络结构特征——对国家级创新型城市的实证分析[J]. 科技进步与对策, 2020,37(23):36-45

[2] 赵志耘. 创新型城市建设成效显著[J]. 科技中国, 2020,25(2):1-3

[3] McGrath L, Guglielmo L. Communities of practice and makerspaces: DMAC'S influence on technological professional development and teaching multimodal composing [J]. *Computers and Composition*, 2015, 9(1): 44-53

[4] 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

[5] 李芊, 李海芹, 郭建伟. 区块链政策文本量化分析——基于政策工具与政策目标视角[J]. 财会月刊, 2021(2): 148-154

[6] Ruiz Estrada M A, Yap S F, Nagaraj S. Beyond the ceteris paribus assumption: modeling demand and supply assuming Omnia Mobilis[J]. *International Journal of Economic Research*, 2008,5(2):185-194

[7] 张永安, 郄海拓. 金融政策组合对企业技术创新影响的量化评价——基于 PMC 指数模型[J]. 科技进步与对策, 2017, 34(2): 113-121

[8] 杜宝贵,王欣. 全面创新改革试验政策量化评价研究[J]. 中国科技论坛,2020,36(5):41-50

[9] 张永安, 郄海拓. "大众创业、万众创新"政策量化评价研究——以 2017 的 10 项双创政策情报为例[J]. 情报杂志, 2018, 37(3): 158-164

[10] 周楠. 天津科技创新政策文本量化分析——基于政策工具的视角[J]. 科技与经济, 2019,32(6):31-35

[11] 何继新,孟依浩,暴禹. 中国高校创新创业政策供给特征及组合评估:一个三维框架的量化分析[J]. 黑龙江高教研究, 2021,39(2):92-99

[12] 陈伟,王秀锋,孙伟男. 政策工具视角下"大众创业,万众创新"政策文本量化研究[J]. 科学管理研究, 2020,38(2):10-17

[13] 张永安,耿喆,李晨光, 等. 区域科技创新政策对企业创新绩效的影响效率研究[J]. 科学学与科学技术管理, 2016,37(8):82-92

[14] 闫瑾. 技术创新政策对高新技术企业创新绩效影响机理研究[D]. 北京:北京工业大学经济与管理学院, 2016

[15] 张永安,闫瑾. 基于文本挖掘的科技成果转化政策内部结构关系与宏观布局研究[J]. 情报杂志,2016,35(2):44-49

Research on quantitative evaluation of science and technology innovation policies of national innovative cities based on text mining

Qie Haituo, Chen Xueying, Zhang Zhijuan

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

Abstract

In this paper 72 national innovative cities are selected as the research samples. The samples are divided into four quadrants according to the cities' economic development ability and innovation ability, forming five sample groups including full sample and four quadrants. The text mining of science and technology innovation policies in 2019 is carried out, mainly including high frequency words, semantic network graph, point degree centrality, network centrality potential, and small group analysis. The result shows that the key points of the science and technology innovation policies of national innovative cities in Chinese are prominent, which mainly rely on projects and enterprises to carry out R&D activities in order to realize the innovation of science and technology. At the same time, it's necessary to pay attention to multi-channel method to improve the innovation performance. This research provides the reference and basis for policy optimization and decision-making.

Key words: national innovative city, science and technology innovation policy, policy quantification, policy evaluation, text mining