

情报理论与实践

Information Studies: Theory & Application

ISSN 1000-7490, CN 11-1762/G3

《情报理论与实践》网络首发论文

题目：我国大数据政策主题分析及发展动向研判
作者：张涛，马海群
网络首发日期：2021-10-11
引用格式：张涛，马海群. 我国大数据政策主题分析及发展动向研判[J/OL]. 情报理论与实践. <https://kns.cnki.net/kcms/detail/11.1762.G3.20211011.1454.002.html>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

●张 涛¹，马海群²

(1. 黑龙江大学信息管理学院，黑龙江 哈尔滨 150080；2. 黑龙江大学信息资源管理研究中心，黑龙江 哈尔滨 150080)

我国大数据政策主题分析及发展动向研判*

摘 要： [目的/意义]自国务院颁布《促进大数据发展行动纲要》以来，我国大数据产业经历了快速发展阶段，但新型冠状病毒肺炎疫情影响下国际形势发生新变化及数据正式成为新型生产要素后，给未来大数据产业发展带来了诸多不确定性，因此对我国现有大数据政策的主题分析及未来发展动向研判具有重要的现实意义。 [方法/过程]文章从团队自建语料库中选取 367 条大数据政策作为样本，通过构建以政策发布时间和层级为纵轴，以主题聚类、主题词共现强度、主题相似度为横轴的二维分析框架来对我国大数据政策进行综合分析。 [结果/结论]分析结果表明，未来我国大数据政策可能出现六大发展动向：数据安全将成为大数据政策发展的核心问题；大数据与各领域深度融合将成为政策发展新方向；关键核心技术研发与攻关将是大数据政策持续关注点；数据智能将引领军民融合的深度发展；跨领域的大数据人才培养将是政府关注重点；构建联防联控与应急管理协同机制将以大数据为重要支点。

关键词： 大数据政策；主题分析；数据安全；政策研判

Topic Analysis and Developing Trends of Big Data Policy in China

Abstract: [Purpose/significance] Since the State Council promulgated the “Action Plan for Promoting the Development of Big Data”, China's big data industry has experienced rapid development. However, new changes have taken place in the international situation under the influence of the COVID-19 epidemic and data has become a new factor of production have brought many uncertainties to the future development of the big data industry. Therefore, it is of great practical significance to analyze the topic of current big data policy and judge its developing trends. [Method/process] This paper selects 367 big data policies from the self-built corpus of the team as samples, and conducts a comprehensive analysis of China's big data policies by constructing a two-dimensional analysis framework with the policy release time and hierarchy as the vertical axis, and the topic clustering, theme co-occurrence intensity and topic similarity as the horizontal axis. [Result/conclusion] The analysis results show that there may be six major developing trends in China's big data policy in the future: data security will become the core issue of big data policy development. Deep integration of big data and various fields will become a new direction for policy development. Core technology research and development will be the continuous focus of big data policy. Data intelligence will lead the deep development of military-civilian integration. Cross-field big data talent training will be the focus of the government. Establishing a coordination mechanism for joint prevention, joint control and emergency management will take big data as an important fulcrum.

Keywords: big data policy; topic analysis; data security; policy research and judgment

大数据驱动下新一轮科技革命催生出新的产业模式和服务业态，给全球社会发展和人类生产生活带来了翻天覆地的变化，世界各国不断通过制定政策的方式来促进大数据产业发展^[1]。2015 年 8 月国务院颁布的《促进大数据发展行动纲要》（以下简称《纲要》）是我国大数据发展的顶层设计和总体部署，此后国家各部委、地方政府先后出台了一系列政策来推动大数据产业的落地。进入 2020 年，在新型冠状病毒肺炎疫情（以下简称新冠疫情）带来国

*本文为国家社会科学基金重点项目“总体国家安全观下的国家情报工作制度创新研究”（项目编号：20ATQ004）和黑龙江大学研究生创新科研项目（项目编号：YJSCX2021-010HLJU）的成果之一。

际形势新变化、数据正式成为新型生产要素等外部因素的催化下,大数据成为了推动我国经济高质量发展的新动能,2021年3月国家公布的“十四五”发展规划中“数据”相关表述达到53处,这表明推动大数据产业发展是国家中长期战略的重要组成部分,而制定具有中国特色的大数据战略不仅可促进大数据产业的可持续发展,还可提升我国在大数据领域的国际话语权和规则制定权,随着2021年5月《安徽省大数据发展条例》的实施,我国大数据政策还将掀起新一轮规划高潮,因此对我国大数据政策未来发展动向研判具有重要的现实意义。

1 研究综述

对国内外大数据政策文献分析发现,国内学者研究集中在以下三个视角。

1) 国家级大数据政策研究。周京艳等^[2]采用内容分析方法,从政策工具的视角,对18份国家级大数据政策进行文本量化分析。李樵^[3]通过政策工具构建三维分析框架对63项大数据政策进行分析。徐蕾等^[4]利用可视化分析工具获取共词网络对9个国家级大数据综合试验区的政策进行研究。刘亚亚等^[5]从政策主题词和政策工具两个维度对94项国家层面大数据政策进行分析。胡峰等^[6]通过计算11项国家级大数据政策的PMC指数值和凹陷程度来评价每项大数据政策的优劣。段尧清等^[7]采用社会网络分析法和主题分析法对213份政务大数据政策扩散特征进行分析。

2) 地方政府大数据政策研究。范梓腾等^[8]从政策“目标—工具”视角分析了地方政府大数据政策的主要内容、基本特征和发展趋势。丁文姚等^[9]采用内容分析法、社会网络分析方法对我国各省市68条地方大数据政策的扩散模式及转移特征进行分析。陈玲等^[10]从政策工具维度和政策主题维度构建了政府开放数据政策的实施框架。王长征等^[11]运用扎根理论和社会网络分析从时间、空间和各部门间的协作网络等方面来探究地方政府大数据治理注意力的演变。蒋天骥等^[12]以我国29个省级地方政府的198篇大数据政策为样本,将政策作用、目标主体和政策工具作为政策分析框架进行文本统计和分析。张涛等^[13]对《纲要》和我国22个地区大数据政策进行比较分析,发现了广东省和福建省大数据政策最为完整和全面,内蒙古自治区和四川省大数据政策中区域特色较为突出。

3) 贵州省大数据政策研究。季飞等^[14]利用NVivo11对贵阳市发展大数据产业的42份文件做信息编码和文本量化分析。安小米等^[15]构建政府大数据治理体系,并采用内容分析方法对贵州省政府大数据治理体系利益相关方访谈文本和政府大数据治理相关政策文本进行了实证研究。张维冲等^[16]以贵州省大数据政策为样本数据,对大数据政策关键表述进行知识表示与知识抽取。谭海波等^[17]以贵州省为例,运用政策内容量化分析方法,对2014—2018年贵州省政府出台的大数据产业政策文本进行分析。

此外国外学者也对此方向展开了研究,如J. Roski等^[18]提出需要大数据政策的调控才可以使大数据在医疗保健领域创造巨大价值。S. Giest等^[19]分析了当前研究中缺乏大数据在数据治理及其在政策过程中的应用,并提出了未来研究的方向。Choi Bong等^[20]基于首尔公共大数据的现状,结合大数据校园和城市数据科学实验室的实际数据,提出了振兴公共大数据的政策方向。L. C. Mahrenbach等^[21]探讨了巴西、印度和中国政府计划利用大数据来构筑由数据驱动政府决策构想,并提出构建大数据时代政府和公民角色的新思路。

综上所述,对现有成果分析如下:一是国内学者从国家级、地方政府、区域三个视角对大数据政策进行了较为全面的分析;二是国外学者研究成果集中在基于大数据的政府决策分析层面。以上研究为本文思路的提出奠定了坚实的基础,当前研究成果虽已较为全面,但多以2020年前发布的政策为主,研究未能反映出未来我国大数据政策发展动向,在新冠疫情影响下国际形势发生新变化和数据正式成为新型生产要素后,尚缺少对我国大数据政策未来发展动向研判的研究成果,而此研究对推动我国大数据产业发展起到重要的现实意义。因此文本尝试利用LDA主题聚类、主题词共现强度、主题相似度等政策文本计算方法^[22]对我国

367 部大数据政策进行不同时间、不同层级的主题分析，以深入挖掘我国大数据政策的发展趋势。

2 研究框架

由于广义上的大数据政策包含信息、数据、知识、智慧、情报等关键词，但如果对此类大数据政策聚类会导致主题过于分散，因此文章在选取样本时精确查找自建语料库^[23]中标题带有“大数据”字样的政策文本 367 条，政策发布时间范围为 2013—2020 年，主要包括规划纲要、实施意见、实施方案、条例、管理办法、通知、规定等，政策文本语句片段共 47485 条，字符数共 2784216 个，利用 ICTCLAS 进行新词提取及结合自建语料库中政策词表共形成 1078 条政策词语。利用 Python 中 Gensim 的 LdaModel 函数建立主题模型^[24]，将预处理后的文本数据进行主题聚类，拟定 2~30 区间内的整数作为候选主题数，通过调用 LDA 主题模型的 Log_Perplexity 方法^[25]得出不同主题对应的困惑度数值，困惑度最小值为最优主题数，在确定最优主题后，提取最优主题所生成的“主题—词”和“文档—主题”两个概率分布，通过“主题—词”的分布，确定各个主题出现概率较高的关键词，计算每个主题中关键词在所有政策文本语句片段中的共现强度，形成主题词矩阵，导入 Gephi 中生成主题词共现矩阵图，在该软件中词与词之间的连线越粗代表其共现强度值越大，以此综合识别大数据政策中的主题内容。研究从政策发布时间和层级两个维度^[26]，利用 LDA 主题聚类、主题词共现强度计算、主题相似度计算等方法对大数据政策进行主题识别，主题时空扩散分析及主题演变分析，不同阶段大数据政策主题相似度计算能揭示其相似性和差异性，根据已有文献研究^[27]，确定 0.3 为相似度阈值，如相邻阶段主题相似度越大主题间演进关系越显著，本文采用桑基图实现大数据政策主题演进路径的可视化呈现。基于对大数据政策主题综合分析结果，最终对我国大数据政策未来发展动向做出研判，研究框架如图 1 所示。



图 1 大数据政策主题分析框架图

3 基于时间维度的大数据政策主题分析

3.1 时间维度

对我国大数据政策年代划分可以挖掘不同时间阶段的政策关注主题，通过不同阶段政策主题的发展变化，可以对大数据政策扩散及演变进行分析^[28]。本文将《纲要》出台的年份和大数据政策发布数量减少拐点年份作为阶段划分依据，分为三个阶段，如图 2 所示，萌芽期（2013—2014 年）共 13 条政策，该阶段政策发布较为稀疏，是大数据政策发展的初始阶段；扩散期（2015—2017 年）共 271 条政策，自 2015 年《纲要》颁布后人工智能政策发布进入了密集阶段，该阶段政策发布较为密集，从政策发布数量上看呈现出快速扩散趋势，该阶段

政策占总数的 73.84%；趋缓期（2018—2020 年）共 83 条政策，2018 年大数据政策发布数量逐渐下降，尤其是近两年大数据政策发布数量较少。从时间维度看，在萌芽期上海、广东、贵州、重庆等地区发布大数据政策较早，由此推动了国家大数据政策的出台，这符合创新政策自下而上的吸纳辐射扩散模式^[29]。而进入扩散期扩散效应较为明显，趋缓期政策逐渐减少，趋于平稳，政策扩散曲线大致呈 S 型分布，符合政策创新扩散的一般规律^[30]。

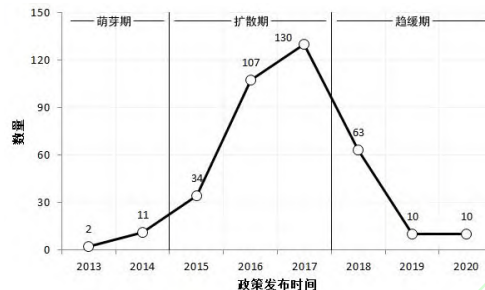


图 2 大数据政策发布时间阶段划分图

3.2 主题聚类及共现强度分析

萌芽期（2013—2014 年）政策数量为 13，最优主题数量为 7，主题如下：主题 1，云计算、大数据中心、数据资源建设；主题 2，电子商务、信息安全、大数据基础设施建设；主题 3，大数据产业园建设，大数据人才引进、人才培养；主题 4，大数据工程、技术、服务体系、数据中心、数据资源建设；主题 5，大数据农业生产经营应用；主题 6，大数据工程技术行业标准；主题 7，大数据应用范例。在主题中除去“大数据 产业”“大数据 发展”“大数据 服务”等共性词语，“智慧 旅游”“高新技术开发区”“广东省 应用范例”等主题词在 Gephi 中共现强度较高。分析结果如下：此阶段是大数据政策发展初始阶段，大数据政策发布数量相对较少，只在上海、广东、贵州、重庆等地区出台大数据发展相关规划，主题集中在大数据基础设施，数据资源，大数据技术服务体系，云计算、大数据中心等内容，尚未呈现扩散趋势，其中 2013 年 7 月上海市科学技术委员会发布的《上海推进大数据研究与发展三年行动计划》（2013—2015 年）在所收集的政策中发布时间最早，属萌芽期中较有代表性的政策。

扩散期（2015—2017 年）政策数量为 271，最优主题数量为 10，主题如下：主题 1，大数据技术服务体系、数据资源建设；主题 2，大数据公共服务监管与创新；主题 3，大数据安全与保护、技术标准实施；主题 4，建立并完善大数据信用与失信机制；主题 5，健康医疗大数据；主题 6，农业大数据助推脱贫攻坚、扶贫；主题 7，政务大数据与网上办事大厅；主题 8，大数据市场主体服务和监管；主题 9，大数据综合治税；主题 10，大数据人才引进与奖励。在主题中除去“大数据 产业”“大数据 发展”“大数据 服务”等共性词语，“健康医疗”“移民 搬迁”“市场 主体”“失信 惩戒”“综合治税 涉税信息”等主题词在 Gephi 中共现强度较高。分析结果如下：扩散期大数据政策除了在大数据技术服务体系、数据资源建设等内容持续关注外，还更加关注数据安全与保护，并且逐渐向健康医疗、农业农村、综合治税等领域迈进。2015 年起在国务院发布的《促进大数据发展行动纲要》《关于运用大数据加强对市场主体服务和监管的若干意见》《关于促进和规范健康医疗大数据应用发展的指导意见》和工业和信息化部发布的《大数据产业发展规划（2016—2020 年）》等政策引领下，政策发布数量及政策主题呈现出中央向地方快速扩散趋势。此阶段政策扩散趋势与“地方政府的政策扩散受中央政府影响，中央政府可以通过强制或激励推进政策扩散”现象较为契合^[9]。

趋缓期（2018—2020 年）政策数量为 83，最优主题数量为 6，主题如下：主题 1，智能制造、智慧城市、大数据与区块链；主题 2，大数据技术服务体系、政府大数据开放与共享；

主题 3，人工智能、大数据与工业大数据；主题 4，健康医疗大数据；主题 5，交通运输大数据、军民融合大数据；主题 6，大数据安全与技术服务标准，数据泄露。在主题中除去“大数据 产业”“大数据 发展”“大数据 服务”等共性词语，“健康 医疗”“网信 军民融合”等主题词在 Gephi 中共享强度较高。分析结果如下：随着各地区纷纷建立并完善大数据政策，2018 年起政策发布数量逐渐减少，部分地区大数据 3~5 年规划接近尾声，大数据政策扩散进入趋缓期，但大数据产业发展已经大势所趋，近年来，新兴信息技术、军民融合等关键词不断在大数据政策中涌现，尤其是 2020 年新冠疫情期间人工智能与大数据为应对疫情发挥了重要作用，且数据正式成为新型生产要素后，这将极大促使新兴技术与大数据深度融合并使其在各领域中发挥重要作用。

3.3 主题相似度

通过对不同阶段政策主题相似度计算能揭示其相似性和差异性，如图 3 所示。

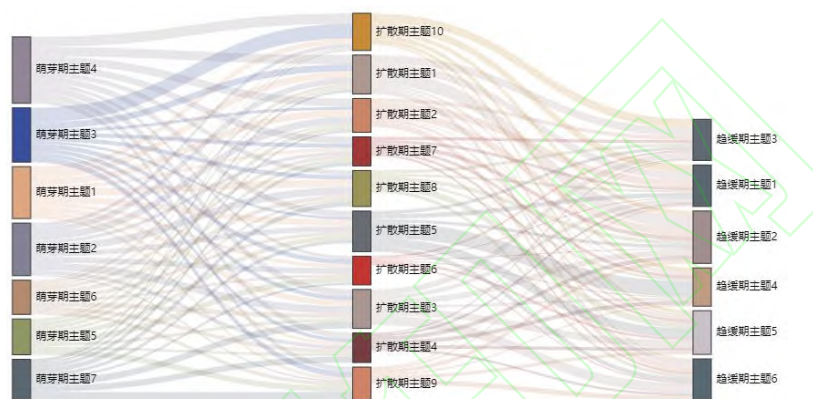


图 3 不同阶段主题相似度图

在萌芽期—扩散期主题相似度数值较低，主题间差异性较大，其中萌芽期主题 3 和扩散期主题 10 的相似度最高为 0.6523，萌芽期主题 3 中主要提及贵阳市大数据产业园建设，大数据人才引进等，扩散期主题 10 中主要提及贵州省大数据人才引进与奖励、大数据创新服务平台建设等，这两个不同阶段的主题对大数据人才引进都尤为关注。萌芽期主题 4 和扩散期主题 1 的相似度为 0.4100，萌芽期主题 4 中提及大数据资源建设，扩散期主题 1 中主要提及大数据技术服务体系、数据资源建设，呈现出一定相似性。分析结果如下：大数据人才引进与培育及大数据资源建设主题相似性相对较高，这说明了不同发展阶段的大数据政策中对该主题关注的延续性。

在扩散期—趋缓期主题相似度数值要高于萌芽期—扩散期，扩散期主题 5 与趋缓期主题 4 相似度最高为 0.6921，其中它们主题均为健康医疗大数据。扩散期主题 1 与趋缓期主题 2 相似度为 0.5990，其中扩散期主题 1 为大数据技术服务体系、数据资源建设，趋缓期主题 2 为大数据技术服务体系、政府大数据开放与共享，在大数据技术服务体系方面具有一定相似性。扩散期主题 3 与趋缓期主题 6 相似度为 0.4314，其中扩散期主题 3 为大数据安全与保护、技术标准实施，趋缓期主题 6 为大数据安全与技术服务标准，数据泄露等，在大数据安全与保护方面呈现出相似性。分析结果如下：健康医疗大数据主题相似度最高，结合 2016 年国务院发布《关于促进和规范健康医疗大数据应用发展的指导意见》，可以发现在该政策发布后健康医疗大数据作为持续关注主题，而新冠疫情期间健康医疗大数据更是在智能诊治、物资调配、药物研发等方面发挥了重要作用，因此未来政策中应重点关注大数据在常态化疫情防控中的应用；而大数据技术服务体系建设，大数据安全与保护主题相似度也较高，这与《促进大数据发展行动纲要》发布后国家对大数据技术服务和大数据安全保护等主题持续关注相契合。

3.4 主题扩散及演进分析

从政策主题扩散视角来看，2015—2017 年随着一系列中央政策的出台我国大数据政策经历了快速扩散阶段，从 2018 年起大数据政策发布数量逐渐减少，进入平稳阶段，但在此过程中人才培养与引进、大数据技术服务体系是持续关注主题。在人才培养与引进方面，贵阳市早在 2015 年就发布《关于加快大数据产业人才队伍建设的实施意见》提出要创新大数据人才培养、引进和储备、集聚机制，大力加强大数据产业人才队伍建设，由此可见，贵阳市对大数据人才十分重视，而其他领域或地区大数据政策都将人才培养与引进加入到重点保障措施中来。在大数据技术服务体系方面，技术是大数据产业发展的重要要素，大数据政策中对关键核心技术的研发与攻关尤为重视，其中贵州省 2020 年 6 月发布的《贵州省大数据标准化体系建设规划》是大数据技术服务体系建设中较有代表性的政策，该政策提出要从提高大数据标准供给能力、提升大数据标准化水平方面为大数据发展及融合应用提供技术支撑。

从政策主题演变视角来看，不同阶段关注的主题有所区别，但大数据资源建设是长期关注的主题，尤其是数据正式被纳入新型生产要素范围后，数据成为基础性资源和战略性资源，数据资源建设的重要性凸显。在扩散期国务院发布一系列大数据政策后，大数据安全与技术服务标准、健康医疗大数据、大数据安全与保护成为政策中主题关注热点，对趋缓期主题分析发现，工业大数据、智能制造、新兴技术与大数据深度融合、军民融合大数据等将在未来一段时间内成为大数据政策的发展趋势。

4 基于层级维度的大数据政策主题分析

4.1 层级维度

按大数据政策发布机构的层级划分可以挖掘不同层级的政策关注主题，通过对不同政策层级主题相似度计算能揭示政策扩散的内容和规律，从而对当前大数据政策主题扩散进行分析。将现有大数据政策分为根政策、干政策和枝政策三个层级^[31]，根政策是指国务院颁布的 3 条政策法规；干政策是指国家各部委颁布的 42 条政策；枝政策是指各省、市、区县等地方政府所颁布的 322 条政策，从政策数量上看符合自上而下层级扩散模式。

从枝政策的地域划分可以看出，如图 4 所示，贵州省（77 条），广东省（33 条）政策发布数量要远超过其他省份，这符合不同发展水平区域间政策跟进扩散模式。贵州省十分重视大数据产业在本区域的发展，作为我国云计算及大数据发展战略布局中的重要省份，由于地缘与环境上具有得天独厚的优势，且省、市、区县政府都有较强的执行力，大数据政策在贵州省的各级政府中落实较好。目前国家大数据中心及国内众多优秀互联网企业的数据中心都相继落户贵州，因此完善的政策体系也促使了大数据产业在贵州省的快速发展。由于贵州省大数据政策文本相对齐全，且在区域大数据政策研究文献综述中多以贵州省大数据政策为样本进行分析^[14-17]，基于此本文也将贵州省政策作为主题分析对象。

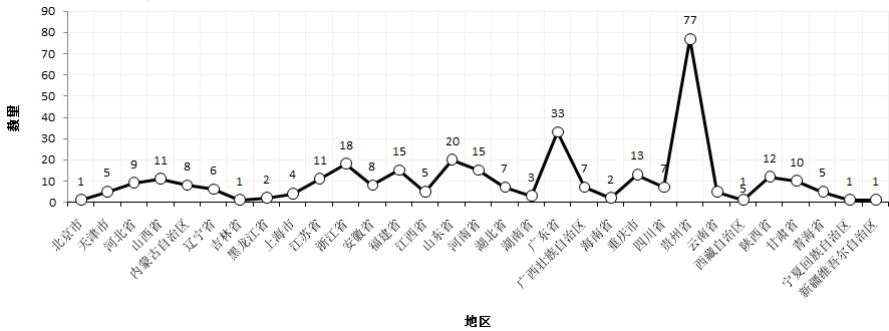


图 4 大数据政策发布地区图

4.2 主题聚类及共现强度分析

由于根政策数量较少，分别为《促进大数据发展行动纲要》《关于运用大数据加强对市

场主体服务和监管的若干意见》《关于促进和规范健康医疗大数据应用发展的指导意见》，且主题较为明确，因此文章主要对干政策、枝政策和贵州省的政策文本进行分析。干政策数量为 42，最优主题数量为 10，主题如下：主题 1，国土资源大数据与地质灾害防治；主题 2，智慧城市、时空大数据云平台建设；主题 3，智慧城市建设、大数据安全、数据泄露与隐私保护；主题 4，交通运输大数据联防联控；主题 5，农业农村大数据、健康医疗大数据、生态环境大数据；主题 6，智慧城市建设、测绘地理信息；主题 7，大数据安全技术标准，气象大数据；主题 8，工业大数据；主题 9，大数据安全与隐私保护；主题 10，大数据技术、服务创新体系、水利大数据。在主题中除去“大数据 产业”“大数据 发展”“大数据 服务”“大数据 技术”等共性词语，“地质 灾害防治”“国家 战略”“智慧 城市”“云 平台”“联防联控”等主题词在 Gephi 中共现强度较高。分析结果如下：在国务院发布 3 部根政策后，大数据政策快速扩散到健康医疗、生态环境、农业农村、交通运输、气象、工业、水利等领域，未来大数据将会在更多领域发挥重要作用，其中大数据安全与隐私保护、数据资源建设在很多主题中都有不同程度出现，这也是大数据应用过程中主要关注的问题。

枝政策数量为 322，最优主题数量为 14，主题如下：主题 1，大数据创新创业产业园区；主题 2，大数据综合治税；主题 3，大数据人才培养与引进；主题 4，大数据安全保障；主题 5，区块链与大数据社会信用体系监管；主题 6，工业大数据；主题 7，公共服务、数据资源、大数据开放与共享；主题 8，农业农村大数据；主题 9，健康医疗大数据；主题 10，农业大数据助推脱贫攻坚；主题 11，大数据在物流、旅游、电子商务、智能制造的应用；主题 12，加快大数据与实体经济相融合；主题 13，大数据技术服务创新、数据资源建设；主题 14，运用大数据加强市场主体服务和监管。在主题中除去“大数据 产业”“大数据 发展”“大数据 服务”“大数据 技术”等共性词语，“建立 机制”“综合治税 涉税信息”“资源 共享”“健康 医疗”“移民 搬迁”等主题词在 Gephi 中共现强度较高。分析结果如下：枝政策中涉及到的主题领域较为广泛，较为全面地识别了当前我国各地区大数据政策主题分布情况，在主题 5 中主题词包含贵州省、贵阳市、贵安、区块链、信用大数据等关键词，这说明贵州省大数据政策发布的数量在全国范围内占比较高。

贵州省政策数量为 77，最优主题数量为 14，主题如下：主题 1，大数据安全与保护；主题 2，大数据资源建设、数据开放与共享；主题 3，健康医疗大数据；主题 4，5G 通信服务、军民融合大数据；主题 5，大数据产业融合创新；主题 6，主题大数据与人工智能、物联网、区块链等新技术深度融合、农业产业脱贫攻坚；主题 7，大数据技术研发、服务体系建设；主题 8，推进大数据社会信用体系；主题 9，大数据人才培养；主题 10，大数据综合治税；主题 11，大数据在物流、旅游、电子商务、智能制造领域中的应用；主题 12，大数据人才队伍建设；主题 13，工业大数据；主题 14，政府大数据聚集共享开放。在主题中除去“大数据 产业”“大数据 发展”“大数据 服务”“大数据 技术”等共性词语，“违反 条例”“云 平台”“开放 共享”“健康医疗 医疗卫生”“农业 产业脱贫攻坚”“实体经济 深度融合”，“重点实验室 战略”，“人才 培养”，“智能 制造”等主题词在 Gephi 中共现强度较高。分析结果如下：由于贵州省大数据政策占枝政策的 23.9%，因此在枝政策中主题 1 和主题 5 分别从不同程度上体现出贵州省大数据政策的主题。贵州省是大数据政策的领跑者，在大数据规划、开放与共享、数据安全保障、人才培养、智能制造、新兴技术融合等领域制定的政策都早于国内其他地区，示范性较强，因此对贵州省大数据政策主题分析具有一定意义。

4.3 主题相似度

通过对不同层级政策主题相似度计算能揭示其相似性和差异性，如图 5 所示。

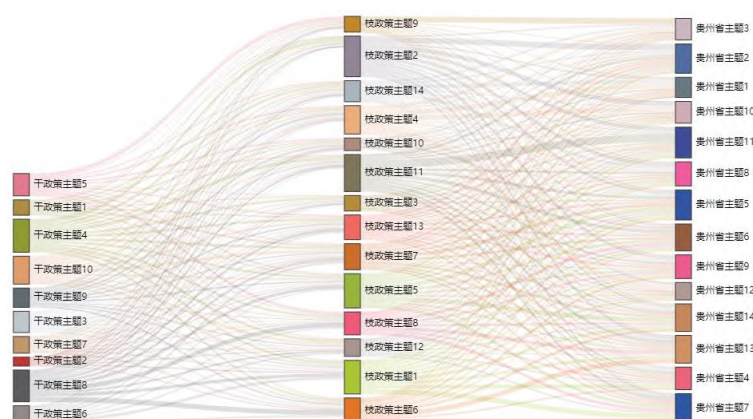


图 5 政策层级主题相似度图

在干政策—枝政策主题相似度数值较低，干政策中主题 8 和枝政策主题 6 的相似度为 0.5579，主要关注工业大数据。2020 年 4 月工业和信息化部发布《关于工业大数据发展的指导意见》，该意见将在促进我国工业数字化转型，激发工业数据资源要素潜力，加快工业大数据产业发展过程中发挥重要作用。干政策主题 5 和枝政策主题 9 的相似度为 0.4948，这两个主题中只有健康医疗大数据的内容具有一定相似性。分析结果如下：由于干政策和枝政策属于不同层级的政策，虽然部分主题中呈现出一定相似性，但总体来说政策关注的范围有所差别，因此整体相似性较低。

从枝政策—贵州省来看，贵州省的大数据政策主题虽然从数量上和枝政策的主题相同，而且部分主题相似性较高，但是从主题关注内容上看差异性较大。枝政策中主题 11 和贵州省主题 11 的相似度为 0.8650，关注主题为大数据与物流、旅游、智能制造等行业深度融合；枝政策中主题 9 和贵州省主题 3 的相似度为 0.6287，关注主题为大数据在健康医疗、医疗卫生等领域应用；枝政策主题 2 和贵州省主题 2 的相似度为 0.6237，其中关注主题是贵阳市贵安新区委员会关于大数据的举措，贵安新区已经成为南方数据中心核心区，同时也是全国大数据产业集聚区、全国大数据应用与创新示范区。分析结果如下：枝政策和贵州省的政策主题数相同，主题内容也具有较高的相似性。在大数据在物流、旅游、智能制造、健康医疗、医疗卫生等领域都是各地区政策中共同关注的主题。

4.4 主题扩散及演进分析

从政策主题扩散视角来看：数据安全在各阶段主题词中关注较高，而健康医疗大数据在两个阶段主题相似度计算中数值较高，因此在数据安全、健康医疗等领域大数据政策扩散效果较好。数据安全主题，在国务院发布的 3 部根政策中都将数据安全保障作为重要部分，这足以凸显其重要性，2017 年起全国信息安全标准化技术委员会连续两年发布《大数据安全标准化白皮书》，提出了大数据安全标准框架，2019 年 8 月贵州省发布的《贵州省大数据安全保障条例》是全国首部省级层面大数据安全保障法规，在很多方面进行了积极尝试和有效探索。健康医疗大数据主题，在 2016 年国务院发布《关于促进和规范健康医疗大数据应用发展的指导意见》后，国家卫健委又印发了《国家健康医疗大数据标准、安全和服务管理办法（试行）》，随后各省市地区共发布 33 条关于健康医疗大数据的政策，由此可见，政府通过强制或激励推进健康医疗大数据政策得到较好的扩散效果。

从政策主题演变视角来看，干政策和枝政策在工业大数据，健康医疗大数据、农业农村大数据等主题延续性较好，但也存在从干政策中大数据在国土资源、地质灾害防治、生态环境、交通运输联防联控等领域的应用演变到枝政策的大数据在综合治税、信用体系监管、物流、旅游等领域的应用，由不同政策层级的主题演变可以发现干政策中的主题多为国家整体规划层面的政策，而枝政策的主题多为区域实施落实层面的政策。枝政策和贵州省从主题数

量到主题内容上一致性均较好。贵州省大数据政策在数据开放与共享、数据安全与保护、大数据与各领域创新融合、人才队伍建设、大数据技术研发等方面具有完善的体系结构。贵州省即集成了枝政策的主题,又在大数据与新兴技术产业创新融合、军民融合大数据等主题与枝政策有所差别,因此贵州省的大数据政策对我国未来大数据政策发展研判有一定借鉴意义。

5 我国大数据政策发展动向研判

基于大数据政策文本内容及政策文本计算分析结果对我国大数据政策未来发展动向做出以下 6 个方向的研判。

1) 数据安全将成为大数据政策发展的核心问题。基于对政策主题扩散的研究发现,数据安全主题在政策扩散期和各层级的大数据政策中均为关注热点。长期以来我国政府高度重视数据安全问题,《纲要》是我国大数据产业布局的战略性政策,在该政策中将强化数据安全保障作为主要任务之一。此后我国发布的一系列大数据政策中,数据安全都作为政策中重要内容,它成为大数据政策发展演变过程中不变的主题。目前国家正逐步构建数据安全保护制度框架,其中已经实施的《中华人民共和国网络安全法》《中华人民共和国数据安全法》和即将正式实施的《个人信息保护法》将从网络数据、数据及个人信息三个维度共同构建我国数据安全保护制度框架的顶层设计^[32],未来各领域各地区将围绕三部法规在大数据隐私保护制度和审查制度、大数据分级分类保护制度建设与完善等方面出台相应政策。2020 年 9 月国务委员王毅在国际研讨会高级别会议上提出《全球数据安全倡议》,体现我国政府在数据安全问题上兼具国际化视野与全局策略,一系列举措足以凸显出我国政府在未来一段时间内对数据安全的战略布局,因此数据安全与隐私保护将成为大数据政策中的核心关切。

2) 大数据与各领域深度融合将成为政策发展新方向。从政策层级的主题演变视角分析发现,我国大数据政策在农业、交通运输、健康医疗、工业、气象、水利、旅游、物流等领域发挥了重要作用,未来大数据与各领域的深度融合将成为政策发展的新方向。从政策主题扩散和演变角度看,健康医疗大数据主题关注度较高:2020 年初突如其来的新冠疫情使健康医疗大数据在智能诊治、物资调配、药物研发等领域发挥了重要作用^[33],进而加速了健康医疗大数据的快速发展,未来大数据也将在常态化疫情防控中发挥重要运用。从政策层级主题演变中发现,工业大数据主题有较好的延续性,工业大数据是智能制造的基石,2020 年 5 月工信部在发布的《关于工业大数据发展的指导意见》中明确了工业大数据是推动制造业数字化、网络化、智能化发展的关键生产要素^[34]。随着中国制造 2025 战略的推进,大数据与智能制造技术融合发展将推进工程生产过程智能化,进而全面提升企业研发、生产、管理和服务的智能化水平。随着大数据基础设施的完善和大数据分析技术的成熟,大数据除在原有领域继续深度融合外,还将在社会信用、综合治税、军民融合等更多领域发挥其重要的价值。

3) 关键核心技术研发与攻关将是大数据政策持续关注点。基于对大数据政策时间维度的主题分析发现,大数据技术服务体系作为持续关注的主题,而在趋缓期智能制造,人工智能、区块链等主题词不断涌现,这将使新兴技术与大数据深度融合后关键核心技术研发与攻关成为未来政策持续关注点。随着大数据、人工智能、区块链等技术的发展,新一轮科技革命和产业变革给全球发展和人类生活带来翻天覆地的变化^[35],我国也通过制定新兴技术国家战略加速新业态、新产业的发展。2020 年 4 月《关于构建更加完善的要素市场化配置体制机制的意见》发布后,数据成为新型生产要素,人工智能、区块链等则成为新型生产工具,而将新型生产要素和生产工具相融合,关键核心技术研发与攻关是首要任务,这些关键核心技术源于基础研究,但对关键核心技术的研究具有多学科综合、高度复杂的特征,只有把这些关键核心技术牢牢掌握在自己手中,才能从根本上解决大数据产业发展过程中的瓶颈问题。因此政府要从政策保障、产权保护、项目驱动、合作研发等方式发挥知名高校、科研院所的科研优势,以问题为导向全面增强技术创新能力,进而实现关键共性核心技术体系的构

建。

4) 数据智能将引领军民融合的深度发展。基于对大数据政策时间维度趋缓期和贵州省大数据政策分析发现,军民融合都是主题之一。军民融合发展过程中产生的海量数据具备大数据特征,将大数据技术应用于军民融合工作中,既是国家战略和国民安全的需要,也是军民融合发展的必然趋势。国家应以大数据、人工智能为引领,着力建设军民融合创新示范体系,细化军民融合深度发展的组织管理体系、工作运行体系、制度保障体系^[36]。其中2018年8月天津市委网信办发布的《关于加快推动大数据产业发展和网信军民融合配套政策及实施细则》较有代表性,该政策较为全面地从培育壮大军民融合产业、大力推动军民融合发展、加快引进军民融合培育人才、优化对军民融合企业的服务、加强军民融合发展统筹协调等方面制定了大数据与军民融合的发展计划。在新冠疫情对国内外安全形势影响较大的情况下,未来我国政府应充分发挥军民融合与大数据对国防建设和经济社会发展的双向支撑拉动作用,形成全要素、多领域、高效益的军民融合深度发展格局。

5) 跨领域的大数据人才培养将是政府关注重点。人才队伍建设贯穿于大数据政策主题时间与层级维度分析中,在《促进大数据发展行动纲要》、《大数据产业发展规划(2016—2020年)》等重要政策中都明确指出要加强大数据人才培养,推动建立健全多层次、多类型的大数据人才培养体系。大数据人才的突出特征是具备复合能力,具有跨领域的思维。因此国家要继续加大大数据人才培养的顶层设计,加强战略扶持且适当的政策倾斜,我国大数据专业建设要有跨领域、跨学科视野,这对培养新一代大数据专业人才意义重大^[37]。各行业、各地区要营造跨领域大数据人才成长所必需的学术环境和较好的生活环境,充分发挥高校、科研院所的优势,加强大数据学科专业一体化建设,实现大数据与不同领域的产学研一体化,进而形成健全的跨领域大数据人才培养体系,在培养大数据人才的同时,政府要积极促进大数据产业高质量发展吸引人才留住人才,而人才加盟也将进一步推动各行业、各地区大数据产业的高质量发展。因此跨领域的人才培养与引进将是政府关注的重点。

6) 构建联防联控与应急管理协同机制将以大数据为重要支点。在干政策中灾害防治、交通运输联防联控是大数据政策关注的主题之一,这些都是应急管理的重要组成部分。在新冠疫情期期间,应急管理部门在数据监测、物资供给等方面能够联防联控、及时调配,大数据都发挥了重要作用。在2019年4月国务院安全生产委员会办公室等部门在《关于加强应急基础信息管理的通知》中提出构建一体化全覆盖的全国应急管理大数据应用平台。2020年2月国家卫健委在《关于加强信息化支撑新型冠状病毒感染的肺炎疫情防控工作的通知》中提出利用大数据技术对疫情发展进行实时跟踪、重点筛查、有效预测,为科学防治、精准施策提供数据支撑。同年3月习近平总书记指出要鼓励运用大数据、人工智能、云计算等数字技术,在疫情监测分析、病毒溯源、防控救治、资源调配等方面更好地发挥支撑作用。一系列重磅政策凸显国家对基于大数据应急管理的重视程度,长远来看,应用大数据是把我国应急管理制度优势转化为治理效能的重要支点,要利用大数据技术深入研究突发事件发生发展规律,构建完善趋势分析、应急预测、动态演化模型,加快建立大数据自动预警平台,加快建立以大数据为支点的联防联控与应急管理协同机制^[38]。

6 结束语

数据要素是经济长期增长的重要动力,在一定程度上决定未来经济发展的质量。当前以大数据为代表的信息资源已经演进为新型生产要素,这助推了我国大数据政策持续发展的趋势。文章通过构建以政策发布时间和层级为纵轴,以主题聚类、主题词共现强度、主题相似度为横轴的二维分析框架来对我国大数据政策进行综合分析,最终从六个方向对大数据政策发展动向进行研判:数据安全将成为大数据政策发展的核心问题;大数据与各领域深度融合将成为政策发展新方向;关键核心技术研发与攻关将是大数据政策持续关注点;数据智能将引领军民融合的深度发展;跨领域的大数据人才培养将是政府关注重点;构建联防联控与应

急管理协同机制将以大数据为重要支点。通过对我国大数据政策发展动向的研判,以期由政府制定具有中国特色的大数据战略提供决策支持,并促使我国大数据产业的可持续发展。□

参考文献

- [1]张勇进,王璟璇.主要发达国家大数据政策比较研究[J].中国行政管理,2014(12):113-117.
- [2]周京艳,张惠娜,黄裕荣,李梦辉,刘如.政策工具视角下我国大数据政策的文本量化分析[J].情报探索,2016(12):7-10,16.
- [3]李樵.我国促进大数据发展政策工具选择体系结构及其优化策略研究[J].图书情报工作,2018,62(11):5-15.
- [4]徐蕾,李庆,肖相泽.基于扎根理论的大数据政策共词网络研究[J].现代情报,2018,38(6):157-164.
- [5]刘亚亚,曲婉,冯海红.中国大数据政策体系演化研究[J].科研管理,2019,40(5):13-23.
- [6]胡峰,温志强,沈瑾秋,姚缘.情报过程视角下大数据政策量化评价——以11项国家级大数据政策为例[J].中国科技论坛,2020(4):30-41,73.
- [7]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.
- [8]范梓腾,谭海波.地方政府大数据发展政策的文献量化研究——基于政策“目标—工具”匹配的视角[J].中国行政管理,2017(12):46-53.
- [9]丁文姚,张自力,余国先,韩毅.我国地方大数据政策的扩散模式与转移特征研究[J].大数据,2019,5(3):76-95.
- [10]陈玲,段尧清.我国政府开放数据政策的实施现状和特点研究:基于政府公报文本的量化分析[J].情报学报,2020,39(7):698-709.
- [11]王长征,彭小兵,彭洋.地方政府大数据治理政策的注意力变迁——基于政策文本的扎根理论与社会网络分析[J].情报杂志,2020,39(12):111-118.
- [12]蒋天骥,张瑶,周庆山.基于文本量化分析的我国地方政府大数据产业政策的完善策略研究[J].现代情报,2021,41(2):132-140,161.
- [13]张涛,马海群,易扬.文本相似度视角下我国大数据政策比较研究[J].图书情报工作,2020,64(12):26-37.
- [14]季飞,李韦,吴水叶.西部地方政府发展大数据产业的政策工具选择——基于贵阳市42份文件的文本量化分析[J].政府治理评论,2018,4(2):31-49.
- [15]安小米,白献阳,洪学海.政府大数据治理体系构成要素研究——基于贵州省的案例分析[J].电子政务,2019(2):2-16.
- [16]张维冲,王芳,黄毅.基于图数据库的贵州省大数据政策知识建模研究[J].数字图书馆论坛,2020(4):30-38.
- [17]谭海波,郑清清,王海函.地方政府大数据产业政策:工具偏好及其匹配——基于贵州省政策文本的分析[J].中国行政管理,2021(1):52-58.
- [18]ROSKI J, BO-LINN G W, ANDREWS T A. Creating value in health care through big data: opportunities and policy implications[J]. Health Affairs, 2014, 33(7):1115-1122.
- [19]GIEST S, NG R. Big data applications in governance and policy[J]. Politics and Governance, 2018, 6(4):1-4.
- [20]CHOI Bong, YUN Jongjin; UMTAEHYEE. A Study on policies to revitalize the public big data in Seoul[J]. Knowledge Management Review, 2019, 20(3):73-89.

- [21] MAHRENBACH L C, MAYER K. Framing policy visions of big data in emerging states[J]. Canadian Journal of Communication, 2020, 45 (1) : 129-141.
- [22] 吴杨. 大数据政策文本与现实的偏差及完善路径研究[J]. 公共管理学报, 2020(1): 31-46, 169-170.
- [23] 马海群, 张涛. 文献信息视阈下面向智慧服务的语料库构建研究[J]. 情报理论与实践, 2019, 42(6): 124-130.
- [24] 张涛, 马海群. 一种基于 LDA 主题模型的政策文本聚类方法研究[J]. 数据分析与知识发现, 2018(9): 59-65.
- [25] GRIFFITHS T L, STEYVERS M. Finding scientific topics[J]. PNAS, 2004, 101(6) : 5228-5235.
- [26] 张涛, 马海群. 基于政策文本计算的开放数据与数据安全政策协同研究[J]. 情报理论与实践, 2020, 43(6): 149-155, 141.
- [27] 刘自强, 王效岳, 白如江. 多维度视角下学科主题演化可视化分析方法研究——以我国图书情报领域大数据研究为例[J]. 中国图书馆学报, 2016, 42(6): 67-84.
- [28] 赵蓉英, 张心源, 张扬, 等. 我国“五计学”演化过程及其进展研究[J]. 图书情报工作, 2018, 62 (13): 127-138.
- [29] 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.
- [30] 刘自强, 许海云, 岳丽欣, 等. 面向研究前沿预测的主题扩散演化滞后效应研究[J]. 情报学报, 2018, 37(10): 979-988
- [31] 汪涛, 谢宁宁. 基于内容分析法的科技创新政策协同研究[J]. 技术经济. 2013(9): 22-28.
- [32] 马海群, 张涛. 从《数据安全法(草案)》解读我国数据安全保护体系建设[J]. 数字图书馆论坛, 2020(10): 44-51.
- [33] 柴扬帆, 孔桂兰, 张路霞. 医疗大数据在学习型健康医疗系统中的应用[J]. 大数据, 2020 (5): 29-44.
- [34] 孙立. 工业大数据驱动下“制造业+互联网”融合与创新研究[J]. 电子商务, 2020(9): 19-20.
- [35] 张涛, 马海群. 国际数据政策文献研究现状、前沿热点与演化趋势分析[J]. 情报科学, 2020, 38(9): 149-156.
- [36] 蔡建峰, 张芳. 大数据视角下“互联网+军民融合”生态体系构建研究[J]. 西北工业大学学报(社会科学版), 2018(4): 95-99.
- [37] 宁慧聪. 中国大数据专业建设的跨学科模式研究[J]. 计算机科学, 2019, 46(11): 159-162.
- [38] 彭宗超, 黄昊, 吴洪涛. 新冠肺炎疫情前期应急防控的“五情”大数据分析[J]. 治理研究, 2020(2): 6-20.

作者简介: 张涛 (ORCID: 0000-0002-3367-4541, 通信作者, Email: zhangtao@hlju.edu.cn) 男, 1981 年生, 博士生, 副教授, 硕士生导师。研究方向: 文本计算与数据分析。马海群 (ORCID: 0000-0002-2091-7620), 男, 1964 年生, 博士, 教授, 博士生导师。研究方向: 信息政策与法律。

录用日期: 2021-09-07