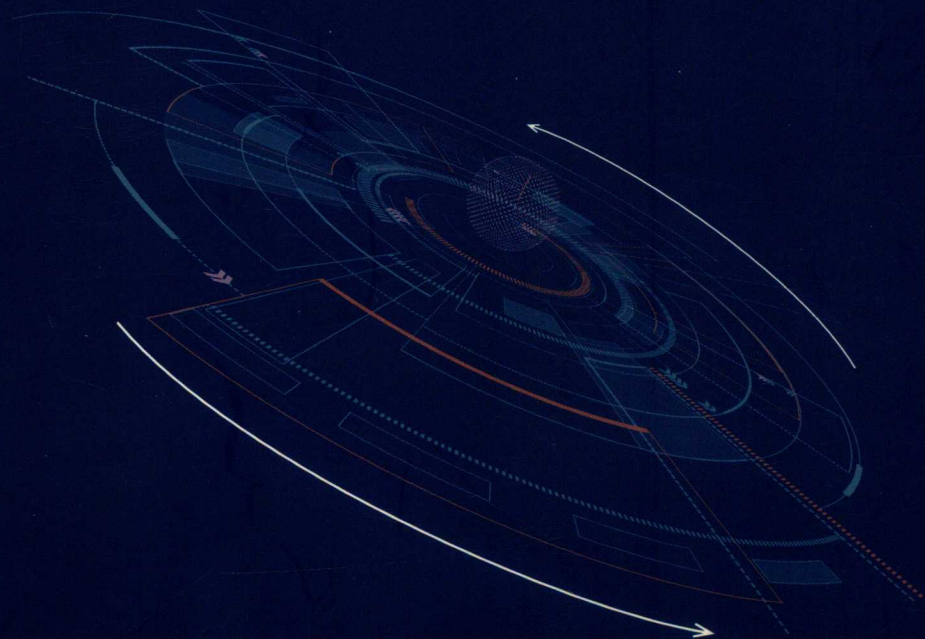


# 「大数据技术 专利分析」

»»»» » 黄 斌◎著



科学出版社

# 大数据技术专利分析

黄 斌 著

本书获得贵州省科技厅专利战略研究计划项目  
(黔科合基础〔2017〕1710)资助



科 学 出 版 社

北 京

## 内 容 简 介

本书基于专利数据分析大数据技术，主要阐述大数据技术市场格局、大数据技术专利竞争态势、大数据技术专利风险预警、大数据重点专利技术、大数据专利技术布局和大数据领域相关政策研究等。本书注重运用多学科的理论与方法，从国际和国内各区域层面对大数据专利技术进行分析，为我国大数据技术发展提供决策支撑。

本书既可作为高等院校相关专业的教学用书，又可作为企业单位、研究机构等从事大数据技术专利开发或分析的相关人员的参考书。

### 图书在版编目(CIP)数据

大数据技术专利分析/黄斌著. —北京: 科学出版社, 2019.6

ISBN 978-7-03-061143-7

I. ①大… II. ①黄… III. ①专利-分析-数据处理 IV. ①G306-39

中国版本图书馆 CIP 数据核字 (2019) 第 083641 号

责任编辑: 孙露露 秦景曼 / 责任校对: 陶丽荣  
责任印制: 吕春珉 / 封面设计: 东方人华平面设计部

科学出版社 出版

北京东黄城根北街 16 号

邮政编码: 100717

<http://www.sciencep.com>

北京虎彩文化传播有限公司 印刷

科学出版社发行 各地新华书店经销

\*

2019 年 6 月第 一 版 开本: 787×1092 1/16

2019 年 6 月第一次印刷 印张: 10 1/4

字数: 249 000

定价: 83.00 元

(如有印装质量问题, 我社负责调换〈虎彩〉)

销售部门电话 010-62136230 编辑部电话 010-62135763-2010

版权所有, 侵权必究

举报电话: 010-64030229; 010-64034315; 13501151303

# 前 言

随着互联网和智能硬件的快速普及，数据以爆炸方式增长。大数据是信息化发展的产物，它已成为国家重要的基础性战略资源，是新的基础设施。大数据能力是一个国家或区域的新型竞争力。党中央、国务院高度重视大数据在经济社会发展中的作用，党的十八届五中全会提出“实施国家大数据战略”，国务院印发《促进大数据发展行动纲要》，全面推进大数据发展，加快建设数据强国。全球新一代信息产业处于加速变革期，大数据技术和应用处于创新突破期，国内市场需求处于爆发期，我国大数据产业面临重要的发展机遇。

大数据技术为决策提供依据，在政府、企业、科研项目等的决策中扮演着重要的角色，在社会治理和企业管理中起到了不容忽视的作用。然而，当前大数据技术研究仍存在诸多问题，如大数据技术有着怎样的市场态势、竞争态势和发展阶段？哪些国家、哪些区域及哪些单位机构处于优势地位？大数据技术创新存在着什么样的风险，以及如何评价？大数据重点技术及其特性是什么？大数据技术布局的相关性和关联性又如何？国际上主要国家的相关政策有哪些？我国各区域相关政策的目标定位是什么？作为我国首个国家大数据综合试验区的贵州省目前处于什么状态，又如何发展？对于这些问题的思考对当前我国尤其是贵州省相应政策的制定具有重要意义，要解决这些问题，势必要求对大数据技术发展态势进行较为全面的分析。

加强知识产权保护成为新时代中国扩大开放的重大举措。以专利为突出代表的知识产权作为一种新兴生产要素，在世界范围内的产业竞争中发挥着越来越重要的战略性作用。国家建立专利制度，是为了保护发明创造专利权，鼓励发明创造，有利于发明创造的推广应用，促进国家科学技术的发展，增强国家的科技创新能力。专利是世界上最大的技术信息源，包含了世界科技信息的90%~95%。近年来，全球大数据技术专利的申请数量增长迅速，大数据技术呈现蓬勃发展的趋势。

本书的主要特色和贡献是应用跨学科知识与技术，丰富和发展专利分析理论。本书运用专利信息学、计量经济学、统计学、数据挖掘和社会网络分析等理论，采取门限回归、面板数据模型、主成分分析、主路径分析、典型相关分析、聚类分析、关联分析等方法，从国际和国内各区域层面对大数据技术市场格局、大数据技术专利竞争态势、大数据技术专利风险预警、大数据重点专利技术、大数据专利技术布局及大数据领域相关政策进行分析。

作者在撰写本书的过程中得到了贵州省科技厅专利战略研究计划项目（黔科合基础（2017）1710）的资助，同时得到了贵州民族大学的支持与协助，在此一并表示感谢。

由于作者水平有限，加之撰写时间仓促，书中难免存在不足之处，希望读者给予批评指正。

黄 斌

2019年1月于贵州民族大学

# 目 录

第 1 章 绪论	1
1.1 大数据的内涵	1
1.1.1 大数据	1
1.1.2 大数据技术	2
1.1.3 国家大数据战略的内在要求	3
1.2 大数据研究概况	3
1.3 本书的研究内容与结构安排	4
1.3.1 专利与技术的对照	4
1.3.2 本书研究对象、视角、思路与目标	5
1.3.3 本书研究内容与框架	5
1.3.4 本书研究方法	6
1.3.5 本书数据来源与处理	6
1.3.6 本书研究特色与创新	7
参考文献	8
第 2 章 大数据技术市场格局	10
2.1 大数据市场背景	10
2.2 大数据技术进展	10
2.2.1 大数据采集与预处理	10
2.2.2 大数据存储与管理	11
2.2.3 大数据计算模式与系统	11
2.2.4 大数据分析可视化	12
2.3 大数据技术趋势	12
2.3.1 大数据处理模式多样化	12
2.3.2 大数据与人工智能深度融合	13
2.3.3 大数据安全与隐私保护	13
2.4 大数据技术市场态势	13
2.4.1 国外大数据市场竞争态势	13
2.4.2 国内大数据市场竞争态势	18
2.5 小结	21
参考文献	22
第 3 章 大数据技术专利竞争态势	23
3.1 理论基础	23

3.2 研究设计	24
3.2.1 分析思路	24
3.2.2 方法与数据	25
3.2.3 数据来源与处理	30
3.3 结果与分析	31
3.3.1 国际大数据技术专利竞争态势	31
3.3.2 国内大数据技术专利竞争态势	37
3.4 小结	46
参考文献	47
<b>第4章 大数据技术专利风险预警</b>	<b>50</b>
4.1 诉讼专利态势	50
4.1.1 理论基础	50
4.1.2 研究设计	50
4.1.3 结果与分析	51
4.2 专利涉诉风险评价	61
4.2.1 理论基础	61
4.2.2 研究设计	61
4.2.3 结果与分析	63
4.3 专利技术创新风险评价	68
4.3.1 理论基础	68
4.3.2 研究设计	68
4.3.3 结果与分析	71
4.4 小结	74
参考文献	75
<b>第5章 大数据重点专利技术</b>	<b>78</b>
5.1 理论基础	78
5.2 研究设计	78
5.2.1 研究方法	78
5.2.2 数据来源与处理	80
5.3 结果与分析	80
5.4 小结	95
参考文献	95
<b>第6章 大数据专利技术布局</b>	<b>96</b>
6.1 理论基础	96
6.2 研究设计	97
6.2.1 研究方法	97



6.2.2 数据来源与处理	99
6.3 结果与分析	99
6.3.1 国际专利技术布局的关联性与相关性	99
6.3.2 国内专利技术布局的关联性与相关性	110
6.4 小结	118
参考文献	119
<b>第7章 大数据领域相关政策研究</b>	<b>120</b>
7.1 理论基础	120
7.2 研究设计	120
7.2.1 分析方法	120
7.2.2 分析思路	121
7.3 大数据政策环境	121
7.3.1 大数据国家战略政策比较	121
7.3.2 我国大数据政策概况	126
7.3.3 贵州省大数据政策分析	137
7.4 小结	140
参考文献	140
<b>第8章 总结与展望</b>	<b>141</b>
8.1 总结	141
8.1.1 主要分析成果	141
8.1.2 建议	142
8.2 展望	143
附录	144
附录1 OECD划分的IPC分类号	144
附录2 国际大数据技术专利基础竞争态势(门限回归分析结果)	146
附录3 国际大数据技术核心专利竞争态势(门限回归分析结果)	147
附录4 国际大数据技术专利综合竞争态势分析(面板数据模型分析结果)	148
附录5 我国大数据技术专利基础竞争态势(门限回归分析结果)	149
附录6 我国大数据技术核心专利竞争态势(门限回归分析结果)	150
附录7 我国大数据技术专利综合竞争态势分析(面板数据模型分析结果)	151
附录8 国民经济行业分类和代码(A~E)	153

# 第1章 绪 论

大数据正在改变着人类思维、工作、生活和学习的方式，引领社会发生新的变革。大数据作为基础性战略资源被各国提升至国家战略，大数据产业已成为各国经济社会发展的新引擎。我国大数据产业形成了以京津冀、长江三角洲（以下简称长三角）、珠江三角洲（以下简称珠三角）和中西部及东北地区为集聚发展区的发展格局，产业生态日渐成熟。2013年被称为“大数据元年”，也是作为全国首个大数据综合试验区的贵州省的大数据产业奠基之年。贵州省抓住了大数据产业发展的重大机遇。发展大数据产业是贵州省实现弯道超车、后发赶超的有效途径。而大数据技术正是大数据产业发展壮大的重要支撑，可以为社会经济活动提供决策依据，提高各个领域的运行效率，提升社会经济的集约化程度，对经济发展转型具有重要的推动作用。

专利是表征技术的合理的、可靠的、可行的指标，是最有效的技术信息载体<sup>[1-6]</sup>。专利作为一种新兴的生产要素，在产业格局方面发挥着越来越重要的作用。专利不仅能够影响技术研发策略，而且能够影响技术发展路线选择；不仅能够影响产业的竞争与合作态势，而且能够影响产业生态系统的构成。专利的刚性钳制作用是影响甚至决定产业竞争成败的关键<sup>[7]</sup>。

然而，国内外关于大数据技术的研究较多从期刊论文、会议论文或研究报告出发，较少从专利视角进行探讨<sup>[8-10]</sup>，这不利于为促进大数据技术发展的政策制定提供更多参考。因此，对大数据技术进行专利分析具有重要的现实意义和理论意义。

本书以大数据技术专利分析为视角开展研究，研究意义在于：一方面，能够丰富和发展专利分析理论与方法；另一方面，基于专利信息对大数据技术展开多维度分析，剖析国家和区域大数据技术态势，希望能够为促进我国及贵州省大数据技术领域的发展提供决策支撑。

## 1.1 大数据的内涵

### 1.1.1 大数据

虽然信息技术领域已有“海量数据”“巨量资料”“大规模数据”等概念，但这些概念着眼于数据规模本身，未能充分反映数据爆发背景下的数据处理与应用需求。而“大数据”这一概念不仅指规模庞大的数据对象，也包含对这些数据对象的处理和应用活动，是数据对象、技术与应用三者的统一<sup>[11]</sup>。

大数据是指无法在可承受的时间范围内用常规软件工具进行捕捉、管理和处理的数据集合，其基本特征如表 1-1 所示。



表 1-1 大数据的基本特征

特征	说明
规模性	大数据的特征首先体现为“数量大”，存储单位从过去的吉字节（GB）到太字节（TB），直至拍字节（PB）、艾字节（EB）或泽字节（ZB） <sup>①</sup> 。随着信息技术的高速发展，数据呈爆发式增长
多样性	广泛的数据来源，决定了大数据形式的多样性。大数据大体可以分为 3 类：一是结构化数据；二是非结构化数据；三是半结构化数据
高速性	大数据的交换和传播是通过互联网、云计算等方式实现的，远比传统媒介的信息交换和传播速度快捷
价值性	大数据最大的价值在于从大量不相关的各种类型的数据中，挖掘出对未来趋势与模式预测分析有价值的信息

大数据可分成大数据技术、大数据工程、大数据科学和大数据应用等领域。其中，大数据技术是解决大数据问题的核心。大数据技术的战略意义不在于掌握庞大的数据信息，而在于对这些含有意义的数据进行专业化处理，即在于提高对数据的“加工能力”，通过“加工”实现数据的“增值”。

### 1.1.2 大数据技术

大数据技术是指从各种各样类型的巨量数据中，快速获得有价值信息的技术<sup>[12]</sup>。大数据技术有别于传统的数据处理技术。大数据技术与传统的数据处理技术的比较如表 1-2 所示。大数据的技术体系通常可以分为大数据采集与预处理、大数据存储与管理、大数据计算模式与系统、大数据分析与管理、大数据可视化分析及大数据隐私安全与保护等方面。其中，大数据采集一般是通过大数据智能感知层，主要包括数据传感体系、网络通信体系、传感适配体系、智能识别体系及软硬件资源接入系统，实现对结构化、半结构化、非结构化的海量数据的智能化识别、定位、跟踪、接入、传输、信号转换、监控、初步处理和管理等；大数据预处理主要完成对已接收数据的辨析、抽取、清洗等操作<sup>[13]</sup>。大数据存储与管理是指用存储器把采集到的数据存储起来，建立相应的数据库，并进行管理和调用<sup>[13]</sup>。大数据计算模式与系统是根据大数据的不同数据特征和计算特征，从多样性的大数据计算问题和需求中提炼并建立的各种高层次抽象模型及其运行的软件环境。大数据分析与管理是指从大量的、不完全的、有噪声的、模糊的、随机的实际应用数据中，提取隐含在其中的、人们事先不知道的但又是潜在有用的信息和知识的过程<sup>[13]</sup>。大数据可视化分析是指在利用计算机自动化分析能力的同时，充分挖掘人对于可视化信息的认知能力优势，将人、机的各自强项进行有机融合，借助人机交互式分析方法和交互技术，辅助人们更为直观和高效地洞悉大数据背后的信息、知识与智慧<sup>[14]</sup>。大数据安全与隐私保护对传统的安全机制提出了新的挑战，同时也为信息安全领域带来了新的发展契机。传统的数据安全与隐私保护机制是针对小规模、静态的数据而定制的，对于大数据来说并不适合<sup>[15]</sup>。

表 1-2 大数据技术与传统的数据处理技术的比较

内容	大数据技术	传统的数据处理技术
采集	传感器、社交网络	人工输入
存储	虚拟化	磁盘

<sup>①</sup> 1GB=1024MB, 1TB=1024GB, 1PB=1024TB, 1EB=1024PB, 1ZB=1024EB。

续表

内容	大数据技术	传统的数据处理技术
传输	无线移动	有线
处理	OLTP、OLAP <sup>①</sup>	单机
分析	大规模数据挖掘算法	小规模数据统计、挖掘
解读	可视化	图表、文字

### 1.1.3 国家大数据战略的内在要求

只有全面准确理解国家大数据战略的内在要求，才能形成广泛的社会共识，充分调动社会资源，完成构建国家大数据体系的各项任务。国家大数据战略的内在要求主要体现在5个方面：一是推动大数据技术产业创新发展；二是构建以数据为关键要素的数字经济；三是运用大数据提升国家治理现代化水平；四是运用大数据促进保障和改善民生；五是切实保障国家数据安全与完善数据产权保护制度<sup>[16]</sup>。

## 1.2 大数据研究概况

2011年，麦肯锡公司的《大数据：创新、竞争和生产力的下一个前沿》报告较为突出地强调了大数据的重要性<sup>[17]</sup>。大数据技术是实现一切大数据采集与预处理、存储管理、大数据分析挖掘、大数据安全和大数据可视化的技术手段，是大数据价值实现的重要条件。大数据技术框架主要分为存储管理、计算、分析和可视化技术，而存储、计算和分析技术是大数据的关键技术，例如，Google的文件系统、MapReduce处理算法和分布式数据库；雅虎公司基于MapReduce编程模型开发的Hadoop开源框架；IBM和Microsoft的并行处理计算机系统（parallel computer system, PCS）及大规模并行处理（massive parallel processing, MPP）等，都是具体的大数据核心和关键技术<sup>[10, 18]</sup>。

在Web of Science数据库中，以Ts=“big data”作为检索策略，检索到16 885篇文献，检索日期为2017年8月23日。这些文献涉及计算机科学与技术、信息与通信工程等100多个学科，不仅包括大多数自然科学领域，也包括教育学、法学和历史学等领域。多数文献是关于大数据技术和方法的应用。还有一些文献关注大数据时代的隐私与信任机制，以及大数据对当前社会的影响。从总体上来看，基于专利视角分析大数据技术的文献不多。在读秀数据库检索出题名包含“大数据”的著作有952本，检索日期为2017年8月23日；在中国知网数据库检索出主题为“大数据”的文献有91 470篇，检索日期为2017年8月26日。从研究的主题来看，多数是关于大数据技术态势、影响和应用的文献。大数据研究文献集中在2011年之后。从总体上来看，基于专利视角分析大数据技术的文献也不多。

学者们从不同的角度对大数据技术进行了探讨，取得了丰富的成果。例如，Gossett分析了大数据技术对人类生活、工作和思维的变革<sup>[19]</sup>；Chen等论述了大数据及其相关技术发展的挑战和前景<sup>[20]</sup>；Baaziz和Quoniam探讨了如何利用大数据技术优化运营石油产业的上游企业<sup>[21]</sup>；Hu等提出由大数据的生成、获取、存储和分析技术组成的大数据技术系统框架<sup>[22]</sup>；

① OLTP（on-line transaction processing，联机事务处理），OLAP（on-line analytical processing，联机分析处理）。

Arias 和 Bae 利用大数据技术对电动汽车充电需求进行预测<sup>[23]</sup>；Ilyasova 等分析了大数据技术在医学诊断系统中的应用<sup>[24]</sup>；Thota 等探讨了在不同的云部署模型中用于大数据存储的体系结构<sup>[25]</sup>；Storey 和 Song 分析了大数据概念建模的作用，并对大数据概念建模的有效性提出建议<sup>[26]</sup>；马费成等探讨了大数据对情报学研究的影响<sup>[27]</sup>；李扬等分析了融合统计思想的大数据算法<sup>[28]</sup>；杨慧和许欢探究了作为理性工具的大数据技术在保障政治生态环境过程中实现公共诉求和修复制度功能的重要价值<sup>[29]</sup>；苏玉娟探讨了大数据技术在高新技术企业数据治理创新方面的应用<sup>[30]</sup>；康红霄和王爱冬提出基于大数据技术的公共危机预测模式和管理策略<sup>[31]</sup>；邱仁宗等分析了大数据技术的伦理问题<sup>[32]</sup>。余芳东分析了大数据在政府统计中的应用、瓶颈及融合路径<sup>[33]</sup>；虞铭明等探讨了医疗健康大数据分析的关键技术与决策支持<sup>[34]</sup>；王旻和蔡淑琴探讨了社会化媒体平台大数据资源模型<sup>[35]</sup>。虽然，学者们在大数据技术领域取得了丰富的研究成果，但是多数研究侧重于大数据技术的发展、影响、方法和应用等方面，而从专利视角分析大数据技术的研究成果较少<sup>[8-10]</sup>，这不利于为促进大数据技术发展提供更多的决策支撑。鉴于此，本书基于专利数据分析大数据技术市场格局、大数据技术专利竞争态势、大数据技术专利风险预警、大数据重点专利技术和大数据专利技术布局及大数据领域相关政策等方面，以期为我国及贵州省大数据技术发展提供决策支撑。

## 1.3 本书的研究内容与结构安排

### 1.3.1 专利与技术的对照

专利是指专有的权利和利益。技术是制造一种产品的系统知识，是所采用的一种工艺或提供的一项服务，不论这种知识是否反映在一项发明专利、一项外观专利、一项实用新型专利或者一种植物新品种中，或者反映在技术情报或技能中，或者反映在专家为设计、安装、开办、维修一个工厂或为管理一个工商业企业及其活动而提供的服务、协助等方面<sup>[36]</sup>。通常，一个专利分类号被界定为一种技术，通过专利分类号可以了解专利所属的技术领域<sup>[37]</sup>。不同的专利分类体系又有不同的专利分类号。常见的专利分类体系有德温特专利分类体系、国际专利分类体系、美国专利分类体系和日本专利分类体系等。其中，国际专利分类（International Patent Classification, IPC）是唯一国际通用的，从技术的功能和应用相结合的角度进行分类的专利文献分类和检索工具。世界知识产权组织（World Intellectual Property Organization, WIPO）提供的《IPC 技术对照表》将全部专利技术划分为 5 大技术类型、35 种技术领域（表 1-3）及其对照的 757 个 IPC 分类号（附录 1），能够较为全面地反映专利类型及其所属技术领域<sup>①</sup>。

表 1-3 专利与技术对照

类型	技术领域
T <sub>1</sub> 电气工程	F <sub>1</sub> 电机、电气装置、电能；F <sub>2</sub> 音像技术；F <sub>3</sub> 电信；F <sub>4</sub> 数字通信；F <sub>5</sub> 基础通信程序；F <sub>6</sub> 计算机技术；F <sub>7</sub> 计算机技术管理方法；F <sub>8</sub> 半导体
T <sub>2</sub> 仪器	F <sub>9</sub> 光学；F <sub>10</sub> 测量；F <sub>11</sub> 生物材料分析；F <sub>12</sub> 控制；F <sub>13</sub> 医学技术

① 世界知识产权组织，IPC 技术对照表[EB/OL]. (2016-02-26) [2018-02-20]. <http://www.wipo.int/ipstats/zh/index.html>.

续表

类型	技术领域
T <sub>3</sub> 化工	F <sub>14</sub> 有机精细化学; F <sub>15</sub> 生物技术; F <sub>16</sub> 药品; F <sub>17</sub> 高分子化学、聚合物; F <sub>18</sub> 食品化学; F <sub>19</sub> 基础材料化学; F <sub>20</sub> 材料、冶金; F <sub>21</sub> 表面加工技术、涂层; F <sub>22</sub> 显微结构和纳米技术; F <sub>23</sub> 化学工程; F <sub>24</sub> 环境技术
T <sub>4</sub> 机械工程	F <sub>25</sub> 装卸; F <sub>26</sub> 机器工具; F <sub>27</sub> 发动机、泵、涡轮机; F <sub>28</sub> 纺织和造纸机器; F <sub>29</sub> 其他特殊机械; F <sub>30</sub> 热工过程和器具; F <sub>31</sub> 机器零件; F <sub>32</sub> 运输
T <sub>5</sub> 其他	F <sub>33</sub> 家具、游戏; F <sub>34</sub> 其他消费品; F <sub>35</sub> 土木工程

### 1.3.2 本书研究对象、视角、思路与目标

本书研究的对象、视角、思路和目标分别如下:

- 1) 研究对象为大数据技术及相关政策。
- 2) 研究视角为专利信息视角。
- 3) 研究思路为分析大数据技术市场格局、大数据技术专利竞争态势、大数据技术专利风险预警、大数据重点专利技术、大数据专利技术布局 and 大数据领域相关政策, 为我国及贵州省大数据技术发展提出建议。
- 4) 研究目标为通过对大数据技术市场格局、大数据技术专利竞争态势、大数据技术专利风险预警、大数据重点专利技术、大数据专利技术布局 and 大数据领域相关政策分析, 以期为我国及贵州省大数据技术领域的发展方向和推进方式提供决策支持。

### 1.3.3 本书研究内容与框架

本书主要研究内容如下:

第1章, 绪论, 包括研究背景与意义、相关概念界定、国内外研究现状和研究设计(包括研究方法、数据来源与处理、技术路线及创新点)等内容。

第2章, 大数据技术市场格局, 包括大数据技术的进展与趋势、大数据行业的市场环境和大数据技术的市场竞争格局等内容。

第3章, 大数据技术专利竞争态势, 从国家与区域层面, 比较分析国家之间和区域(我国各省市)之间的专利竞争态势, 包括基础竞争态势、核心专利竞争态势和总体竞争态势等内容。

第4章, 大数据技术专利风险预警, 从国家与区域层面, 对国家之间和区域(我国各省市)之间的专利风险预警进行分析, 包括涉诉专利态势、涉诉专利关联性和专利风险评价等内容。

第5章, 大数据重点专利技术, 包括重点专利识别、专利引文网络主路径构建和重点专利技术分析等内容。

第6章, 大数据专利技术布局, 利用技术关联性和技术相关性分析专利技术布局设计。

第7章, 大数据领域相关政策研究, 对国内外主要国家或地区大数据领域相关政策进行分析, 对区域(我国各省市)大数据领域相关政策进行比较分析, 对贵州省大数据产业发展现状及政策进行分析。

第8章, 总结与展望, 在前几章研究内容的基础上进行总结, 为大数据技术发展提出建议, 同时也对今后专利分析进行展望。

本书研究框架如图1-1所示。

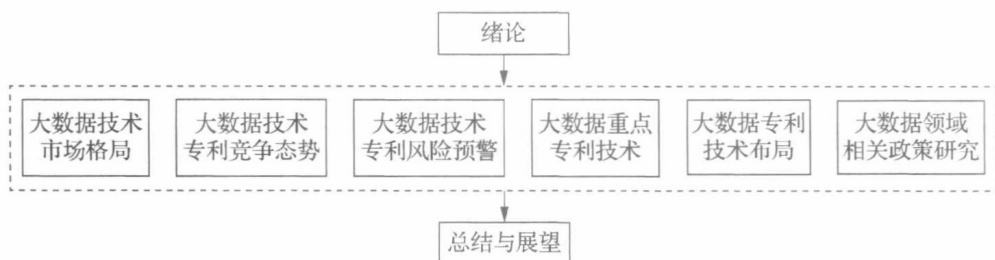


图 1-1 本书研究框架

### 1.3.4 本书研究方法

本书基本研究方法及其对应章节如表 1-4 所示。具体研究方法详见各章所述。

表 1-4 本书基本研究方法及其对应章节

研究方法	对应章节
文献查阅法	第 1 章、第 2 章、第 3 章、第 4 章、第 5 章、第 6 章和第 7 章
专利指标组合分析法	第 3 章、第 4 章、第 5 章和第 6 章
综合评价法	第 3 章、第 4 章和第 5 章

### 1.3.5 本书数据来源与处理

本书数据来源与处理方式说明如下。

1) 数据来源。鉴于 Derwent Innovation (DI, 德温特创新平台) 和 IncoPat 科技创新情报平台的数据范围与特点 (表 1-5 和表 1-6), 本书收集的大数据技术专利数据主要来源于德温特创新平台和 IncoPat 科技创新情报平台。

表 1-5 DI 专利数据范围与特点

项目	说明
范围	1) 德温特世界专利索引 (Derwent World Patents Index, DWPI), 包含世界各地 50 家专利授予机构的增值专利信息, 涵盖 6 000 多万个专利和近 3 000 万个同族专利。增值的专利信息包括经技术专家改写的标题和摘要, 德温特专利分类、手工代码, 德温特专利家族及标准专利权人代码 2) 德温特专利引文索引 (Derwent Patents Citation Index, DPCI), 包含 1 900 多万个同族专利的增值专利引用信息 3) 专利全文: 美国、欧洲、WIPO、英国、加拿大、法国、德国、日本、韩国等专利全文数据 4) 全球 100 多个专利授予机构的标准化专利数据 5) INPADOC 著录项目、法律状态和专利家族数据 6) 美国专利法律状态信息
特点	1) 提供多种检索方式, 包括智能检索、表单检索、专家检索和公开号检索 2) 可视化和多样化的分析工具, 包括专利地图、文本聚类、引证分析, 以及对专利记录进行分组分析的预定义图表和自定义图表等 3) 提供专利原文下载、数据导出、多级文件夹、自动预警和监控等功能, 实现便捷的信息存储、共享、追踪

表 1-6 IncoPat 专利数据范围与特点

项目	说明
范围	<p>1) IncoPat 完整收录全球 102 个国家/组织/地区 1 亿多个基础专利数据, 对 22 个主要国家的专利数据进行特殊收录和加工处理, 数据字段更完善, 数据质量更高</p> <p>2) 对主要国家的题录、摘要进行机器翻译, 提供可供检索的多语种标题摘要信息</p> <p>3) 对重点企业和机构的不同别名、译名、母公司和子公司名称, 建立标准化的申请人名称代码表</p> <p>4) 对国内外专利的法律状态、同族信息、引证信息进行深度加工, 丰富字段信息</p> <p>5) 将中美专利诉讼、转让、许可、质押、复审无效等法律信息与专利文献相关联, 实现大数据融合, 便于进行专利的多维分析和价值挖掘</p> <p>6) 每周将国内外最新发布专利更新入库, 速度领先于国内同类系统, 支持用户及时掌握最新技术</p>
特点	<p>1) 我国官方提供的专利信息中可供检索的字段仅有 20 个左右, IncoPat 针对全球数据进行深度整合后, 可供检索的字段达到 206 余个。检索界面包括简单检索、高级检索、批量检索、引证检索和法律检索 5 种形式, 可满足普通用户和专业用户的不同检索需求</p> <p>2) IncoPat 提供了数千万个国外专利的中文标题和摘要, 以及我国专利的英文标题和摘要, 支持中英文双语检索和浏览全球专利, 同时支持用小语种检索和浏览小语种专利原文, 语言不再是用户获取专利智慧的障碍</p> <p>3) 建立了各行业龙头企业或机构的名称代码表, 收录了数万家公司的母公司和子公司名称, 以及中英文别名和译名等信息, 尤其对在我国进行专利布局的重要企业进行了系统的收录, 可以帮助用户全面网罗竞争对手的专利</p> <p>4) 针对一项技术或一个竞争对手进行专利检索时, 检索结果经常达到上万个专利, 想在其中找到价值度较高的专利, 要耗费大量的时间。IncoPat 利用数据挖掘、迭代优化的方法, 创建了一套客观的价值度评价体系。通过专利价值度排序, 在对数千万个专利信息进行批量加工后, 我们发现专利中的很多参数与价值度相关</p> <p>5) 系统对专利和专利族的前后引证信息和非专利引证信息进行深度加工, 可按专利号和申请人检索前后引证信息, 也可对检索结果的前后引证信息进行批量检索。检索结果可以按照被引证频次进行排序和限定, 可以查看专利引证树, 对重点专利进行多级引证分析, 帮助用户快速聚焦核心专利</p>

2) 检索表达式。大数据技术专利数据的相关检索表达式: 在 DI 专利数据库中检索大数据技术相关专利的表达式为“(all=(Big ADJ Data) OR all=(MapReduce) OR all=(BigTable) OR all=(MPP ADJ Comput\*) OR all=(Hadoop) OR all=(GoogleFileSystem)) AND (IC=(G06F\*) OR IC=(H\*))”；在 IncoPat 专利数据库中检索大数据技术相关专利的表达式为“TIABC=(大数据) OR TIABC=(Big Data) OR TIABC=(MapReduce) OR TIABC=(BigTable) OR TIABC=(Hadoop) OR TIABC=(GoogleFileSystem) OR TIABC=(MPP Comput\*) AND (IPC=(G06F\*) OR IPC=(H\*))”<sup>[8,10]</sup>。其中, 对于涉诉专利的检索, 在 IncoPat 专利数据库中检索, 检索表达式为“lgi-flag=1”；对于典型大数据企业的个案分析, 如 Google、IBM 和 HP 等, 将检索词的概念范围扩大, 由“big data”扩大为“data”, 以期能够较为全面地了解这些企业在数据分析与处理方面的技术优势与实力。

3) 数据处理。利用电子表格中 VBA 程序语言, 将专利数据中专利分类号与《IPC 技术对照表》中 757 个技术一一对照, 并利用 Derwent Data Analyzer (DDA) 软件对专利数据进行清洗、筛选、整理和分析。

由于各章专利分析内容不同, 具体的专利数据检索与处理见各章所述。

### 1.3.6 本书研究特色与创新

结合定量、定性与可视化方法, 从国家和区域层面分析大数据领域的市场竞争格局、专利竞争态势、专利风险预警、重点专利技术、专利技术布局 and 大数据政策环境, 为我国



及贵州省大数据技术领域发展方向和推进方式提供决策支持。

### 参 考 文 献

- [1] GRILICHES Z. Patent statistics as economic indicators: a survey[J]. NBER working papers, 1990, 28(4): 1661-1707.
- [2] DALTON D M, BURKE T P, KELLY E G, et al. Quantitative analysis of technological innovation in knee arthroplasty: using patent and publication metrics to identify developments and trends[J]. The journal of arthroplasty, 2016, 31(6): 1366-1372.
- [3] PANTANO E, PRIPORAS C V, SORACE S, et al. Does innovation-orientation lead to retail industry growth? Empirical evidence from patent analysis[J]. Journal of retailing and consumer services, 2017, 34(1): 88-94.
- [4] ARDITO L, D'ADDA D, PETRUZZELLI A M. Mapping innovation dynamics in the internet of things domain: evidence from patent analysis[J]. Technological forecasting and social change, 2017, 5(4): 1-14.
- [5] RAITERI E. A time to nourish? Evaluating the impact of public procurement on technological generality through patent data[J]. Research policy, 2018, 47(5): 936-952.
- [6] FUJII H, MANAGI S. Trends and priority shifts in artificial intelligence technology invention: a global patent analysis[J]. Economic analysis and policy, 2018, 58(6): 60-69.
- [7] 贺化. 专利与产业发展系列研究报告[M]. 北京: 知识产权出版社, 2013.
- [8] 李鹏飞, 卢瑾, 辛一. 基于专利的大数据技术发展情报分析及战略研究[J]. 情报杂志, 2014, 33 (9): 45-50.
- [9] 党倩娜, 曹磊, 罗天雨. 全球大数据产业技术创新态势及相关政策研究[M]. 上海: 上海科学普及出版社, 2015.
- [10] 李文娟, 刘桂锋, 卢章平. 基于专利分析的我国大数据产业技术竞争态势研究[J]. 情报杂志, 2015, 34 (7): 65-70.
- [11] 李宇. 信息资源管理与应用[M]. 北京: 国家行政学院出版社, 2015.
- [12] 常晋义, 高燕. 计算机科学导论[M]. 2版. 北京: 清华大学出版社, 2014.
- [13] 娄岩. 医学计算机与信息技术应用基础[M]. 北京: 清华大学出版社, 2015.
- [14] 任磊, 杜一, 马帅, 等. 大数据可视分析综述[J]. 软件学报, 2014, 25 (9): 1909-1936.
- [15] 宋智军. 深入浅出大数据[M]. 北京: 清华大学出版社, 2016.
- [16] 郑迦元. 实施国家大数据战略的内在要求和价值体现[J]. 重庆行政 (公共论坛), 2018, 19 (1): 49-51.
- [17] MANYIKA J, CHUI M, BROWN B, et al. Big data: the next frontier for innovation, competition, and productivity[R/OL]. (2011-05-26) [2018-10-12]. <https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation>.
- [18] CHANG F, DEAN J, GHEMAWAT S, et al. Bigtable: a distributed storage system for structured data[J]. ACM transactions on computer systems (TOCS), 2008, 26(2): 1-26.
- [19] GOSSETT E. Big data: a revolution that will transform how we live, work, and think[J]. Perspectives on science and christian faith, 2015, 67(2): 156-158.
- [20] CHEN M, MAO S, ZHANG Y, et al. Big data: related technologies, challenges and future prospects[M]. Heidelberg: Springer, 2014.
- [21] BAAZIZ A, QUONIAM L. How to use big data technologies to optimize operations in upstream petroleum industry[J]. Social science electronic publishing, 2014, 1(1): 19-25.
- [22] HU H, WEN Y, CHUA T S, et al. Toward scalable systems for big data analytics: a technology tutorial[J]. IEEE access, 2014: 652-687.
- [23] ARIAS M B, BAE S. Electric vehicle charging demand forecasting model based on big data technologies[J]. Applied energy, 2016, 183(12): 327-339.
- [24] ILYASOVA N, KUPRIYANOV A, PARINGER R, et al. Particular use of big data in medical diagnostic tasks[J]. Pattern recognition and image analysis, 2018, 28(1): 114-121.
- [25] THOTA C, MANOGARAN G, LOPEZ D, et al. Architecture for big data storage in different cloud deployment models[M]//Handbook of research on big data storage and visualization techniques. IGI Global, 2018: 196-226.
- [26] STOREY V C, SONG I Y. Big data technologies and management: what conceptual modeling can do[J]. Data & knowledge engineering, 2017, 108(3): 50-67.
- [27] 马费成, 张瑞, 李志元. 大数据对情报学研究的影响[J]. 图书情报知识, 2018, 185 (5): 4-9.
- [28] 李扬, 张长, 朱建平. 融合统计思想的大数据算法[J]. 统计研究, 2018, 35 (7): 125-128.

- [29] 杨慧, 许欢. 大数据技术优化政治生态发展路径研究[J]. 行政论坛, 2018, 25 (4), 38-42.
- [30] 苏玉娟. 大数据技术与高新技术企业数据治理创新: 以太原高新区为例[J]. 科技进步与对策, 2016, 33 (6): 47-52.
- [31] 康红霄, 王爱冬. 基于大数据技术的公共危机预测研究[J]. 科技管理研究, 2015, 35 (6): 175-178.
- [32] 邱仁宗, 黄雯, 翟晓梅. 大数据技术的伦理问题[J]. 科学与社会, 2014, 4 (1): 36-48.
- [33] 余芳东. 大数据在政府统计中的应用、瓶颈及融合路径[J]. 调研世界, 2018, 26 (11): 3-11.
- [34] 虞铭明, 张迺英, 李月娥. 医疗健康大数据分析的关键技术与决策支持[J]. 中国科技论坛, 2018, 34 (11): 53-62.
- [35] 王昉, 蔡淑琴. 社会化媒体平台大数据资源模型研究[J]. 管理学报, 2018, 15 (10): 1064-1071.
- [36] 魏巍, 冯琳. 国际经济合作[M]. 大连: 东北财经大学出版社, 2009.
- [37] 黄斌, 黄鲁成, 吴菲菲, 等. 度量专利表征技术的差异[J]. 情报杂志, 2015, 34 (10): 34-37.

## 第 2 章 大数据技术市场格局

大数据可以释放出巨大价值，给每个行业带来新的发展机遇。例如，政府可以利用大数据让政务更加公开透明；企业可以利用大数据分析用户消费习惯，创造更大的商业价值等。

### 2.1 大数据市场背景

大数据以爆炸式的发展速度蔓延至各行各业，随着各国不断加大对大数据的扶持力度，加之资本的青睐，全球大数据市场规模保持着高速增长的态势。前瞻产业研究院《2017—2022 年中国工业大数据产业发展前景与投资战略规划分析报告》显示，2016 年全球大数据市场规模达到 281 亿美元，同比增长 22%；其中，硬件、软件和服务所产生的收入分别为 83 亿美元、91 亿美元、107 亿美元，占比分别为 30%、32%和 38%。中国信息通信研究院《中国大数据发展调查报告（2017 年）》显示，2016 年我国大数据市场规模为 168 亿元人民币。大数据正与传统产业进行融合，对传统产业进行从设计研发、生产管理到售后维护全流程的改变。大数据应用的市场非常广阔。例如，零售业可以运用大数据调整销售策略，制造业可以运用大数据加强售后维护，农业可以运用大数据制定收割路线，电信业可以运用大数据加强精准营销等。

### 2.2 大数据技术进展

大数据领域每年都会涌现出大量的新技术，成为大数据获取、存储、处理分析或可视化的有效手段。近些年来，大数据技术在以下方面获取了较大进展。

#### 2.2.1 大数据采集与预处理

大数据的采集是在确定用户目标的基础上，针对与该目标相关的所有结构化、半结构化和非结构化数据的采集，采集后对这些数据进行处理，从中分析和挖掘有价值的信息。大数据采集的主要特点和面临的挑战往往是成千上万的用户同时进行访问和操作而引起的高并发量。传统的数据采集与大数据的数据采集相比，在数据来源、数据类型和数据处理等方面具有差异性，如表 2-1 所示<sup>[1]</sup>。在大数据的生命周期中，数据采集处于第一个环节。根据 MapReduce 产生数据的应用系统分类，大数据的采集主要有 4 种来源：管理信息系统、Web 信息系统、物理信息系统和科学实验系统。对于不同的数据集，可能存在不同的结构和模式，如文件、XML 树、关系表等，表现为数据的异构性。对多个异构的数据集，需要做进一步集成处理或整合处理，将来自不同数据集的数据进行收集、整理、清洗、转换，生成到一个新的数据集，为后续查询和分析处理提供统一的数据视图。针对管理