

中国政府开放数据利用研究报告 (2020年)

国家社会科学基金重点项目(17ATQ006)课题组

华中师范大学信息管理学院

湖北省数据治理与智能决策研究中心

2020年4月

中国政府开放数据利用研究报告 (2020年)

国家社会科学基金重点项目(17ATQ006)课题组

发布机构

国家社会科学基金重点项目(17ATQ006)课题组

华中师范大学信息管理学院

湖北省数据治理与智能决策研究中心

版权声明

本报告版权属于国家社会科学基金重点项目“基于全生命周期的政府开放数据整合利用机制与模式研究”（项目编号：17ATQ006）课题组，华中师范大学信息管理学院，湖北省数据治理与智能决策研究中心。凡转载或引用本报告的观点、数据等，请注明“来源：国家社会科学基金重点项目“基于全生命周期的政府开放数据整合利用机制与模式研究”（项目编号：17ATQ006）课题组”。

前言

政府开放数据的有效利用是政府开放数据系统建设的核心目标，政府开放数据系统建设离不开数据开放平台的建设。政府开放数据平台的建设为政府数据赋能公众和企业提供了良好的平台，是挖掘数据动能，构建政府数据赋能体系的重要要素，也是数据利用的基础，2019年我国各级各类政府开放数据平台建设在政府政策驱动下，取得了显著成效，平台建设速度大幅加快，由2018年的62个增至108个（截止至2019年底），为完善政府开放数据体系建设夯实了基础。2019年是5G运用元年，“新基建”全面启动，移动网络不断迭代进化，数据流动和利用的技术环境发生了根本变化，公众与政府对数据利用和依赖性程度逐渐加强，政府、企业的数字化转型，数据赋能成为2019年的重要工作。在网络使用环境与专用平台进一步完善的情况下，我国政府开放数据是否已经形成“用数据说话、用数据决策、用数据管理、用数据创新”的新格局？政府开放数据的利用效能是否有所提升等问题势必成为政产学研更为关注的课题，在此背景下，弄清政府开放数据的利用现状，寻找提高数据价值的途径具有极大的紧迫性和必要性。

推进政府数据开放共享，提升数据资源价值势在必行，但数据价值表现多样，数据利用行为纷繁复杂，在对政府开放数据利用研究时可能挂一漏万，课题组希望能够做到去伪存真，删繁就简，通过科学分析由表及里。2019年7月，华中师范大学信息管理学院国家社科重点项目（17ATQ006）课题组通过抽样分析对2018年政府开放数据利用进行了初步的探索，并发布了首份《2018年度中国政府开放数据利用

报告》，主要聚焦于数据集的操作行为分析来发现问题，提出对策，尽管当时认真分析、小心求证，但还是在样本选择，分析方法，评价模型构建等方面存在不少问题。通过总结2018年研究报告制作经验，完善样本与数据采集方式，优化效能评价指标后，课题组于2020年4月完成了《中国政府开放数据利用研究报告（2020）》（以下简称报告），报告是课题组对2019年全国108个平台数据集监测分析制作完成的。报告主要从各级各类政府部门对政府开放数据的政策驱动、平台数据内容、数据利用、数据利用效能等方面来把握2019年我国政府开放数据的利用现状，同时与2018年相关数据进行比较分析，探讨我国政府开放数据利用存在的问题，并提出提高政府数据赋能的策略，但站在2020年的4月份的时间节点上完成此份报告，其难度超乎想象，特别是2020年初的新冠肺炎防疫会影响我们对2019年政府数据赋能的价值判断，为了实事求是，报告主要基于事实判断，以期数据赋能在探索中不断完善方法和路径，为相关政府部门、公众提供决策参考。

众所周知，目前新冠肺炎还在全球肆虐，各国政府在应对疫情防控的同时，已花费了大量时间精力来满足公众对政府开放数据的类型、时间、粒度等不同需求，政府开放数据从未如此受到公众关注，其范围之广、涉人之众、程度之强、影响之深、改变之大，目前还须拭目以待。面对新环境，新需求，新特征，新任务，课题组本着科学精神，立足我国数字中国战略总体要求，密切监控事态发展，拟在下一份研究报告中反映不同时间，不同环境下的新认识。

目录

1 政府开放数据政策驱动	1
1.1 政府开放数据政策发布	1
1.2 政府开放数据政策主题变化	1
1.3 政府开放数据政策价值取向	2
2 政府数据开放平台主要内容	4
2.1 平台建设	4
2.2 数据集主题	7
2.3 数据集数量	8
3 政府开放数据利用行为分析	13
3.1 政府开放数据利用行为概况	13
3.2 政府开放数据浏览行为分析	18
3.3 政府开放数据下载行为分析	20
3.4 政府开放数据利用行为综合分析	23
3.5 对比2018年报告的变化分析	34
4 政府开放数据利用效益分析	35
4.1 数据数量	35
4.2 数据标准	38
4.3 数据持续性	42
4.4 数据利用性	45
5 政府开放数据利用案例	51
5.1 交通层面	51
5.2 教育层面	52
5.3 营商环境层面	52
6 结论与对策	54
6.1 现状	54
6.2 存在问题	55
6.3 对策建议	57

1 政府开放数据政策驱动

1.1 政府开放数据政策发布

2019年我国政府开放数据的工作重点由2018年的建设与利用兼顾变为在建设的基础上强调政府开放数据的利用。我国政策主要围绕数据管理和数据利用制订。

数据管理政策方面，为规范电子政务建设与发展，推进政务数据共享与开放，提高政府服务与管理能力，我国多个省市出台了对政府开放数据管理的相关政策，如山东省出台的《山东省电子政务和政务数据管理办法》、上海市出台的《上海市加快推进数据治理促进公共数据应用实施方案》等。这些政策均有力支持了《中共中央关于坚持和完善中国特色社会主义制度推进国家治理体系和治理能力现代化若干重大问题的决定》中推进数字政府建设，加强数据有序共享的要求，也为政府开放数据的利用相关政策的实施提供了重要保障和推动力。

数据利用政策方面，国家层面上的政府开放数据利用在国家建设中体现出了其高应用性、高适用性的特点，如国务院颁布的《国务院关于实施健康中国行动的意见》、《交通强国建设纲要》、《国务院办公厅关于印发体育强国建设纲要的通知》等政策中均提出提升智慧化全民公共服务能力，实现资源整合、数据共享、互联互通，加强分析应用等要求。省级层面上的政府开放数据利用则应用更为广泛，电子商务、电子政务、在线政务服务、政府信息公开、互联网+政务等方面均有所体现，主要主题分布如图1-1所示。

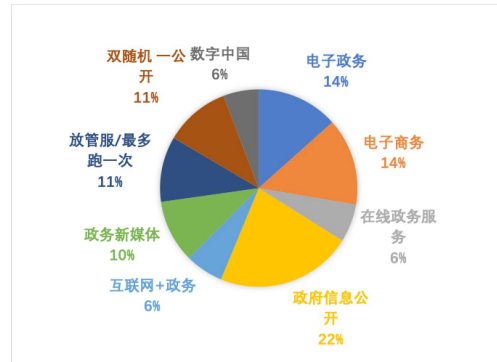


图1-1 2019年度我国政府开放数据相关政策主要主题分布

我国现阶段的政府开放数据利用在利用特征上呈现出由特殊到普遍、在利用方式上呈现出由复杂到简单、在利用程度上呈现出由低到高的趋势。从政策驱动实施分析，政策主要解决我国政府开放数据利用问题的发展方向由“如何利用”逐渐转变为“如何高效合理利用”，而对于其他领域中，政府开放数据的利用也逐渐成为不可缺少的一个环节，即由“特殊利用”走向“泛在利用”。

1.2 政府开放数据政策主题变化

2019年度我国政策文本出现的关键高频词包括“服务”、“公开”、“公共”、“监管”等词与2018年度的关键高频词较为相似，表明我国各省市政府在政策的制定上具有一定的连续性，关键词“加快加强”则体现出我国各省市对于政府开放数据政策的稳步实施。可见无论是“放管服”还是“最多跑一次”，其后续的政策颁布都能很好的围绕其展开。

2018年度对于政府开放数据的战略重点为政府开放数据的平台建设与管理，更加全面的进行政府信息公开，其主要目的为建设人民满意的服务型政府，维护政府的公信力。2019年度我国各省市的主要政策围

政策驱动

绕着建设与应用，而建设和应用的主要目的仍然为服务。相比于2019年度的高频词汇，2018年度的高频词汇“平台”、“网络”、“媒体”等词出现频率明显下降，结合政策的实际发布情况，报告认为我国政府开放数据平台建设已经相对完善，网络应用的主要方向也由如何建设和共享转变为如何利用。

1.3 政府开放数据政策价值取向

1.3.1 国家政策价值取向

2018年我国对政府开放数据的顶层设计趋于完善，制定了包括“互联网+政务服务”、“放管服”、“最多跑一次”等一系列政府开放数据利用相关政策以促进政务发展，为全面提升政务服务规范化、便利化水平，提高政府开放数据共享，更好的为企业和群众提供全流程一体化的政务在线服务，推动政府治理现代化打下了坚实基础。在顶层设计趋于完善后，2019年度我国出台的国家层面的政府开放数据相关政策数量共计条29条，同比去年增长了220%。政策内容主要包括全面及时准确了解企业和群众对政务服务的感受和诉求，接受社会监督，有针对性地改进政务服务，提升政府工作效能，优化营商环境，建设人民满意的服务型政府；进一步克服“准入不准营”现象，使企业更便捷拿到营业执照并尽快正常运营；持续深化“放管服”改革，进一步做好利用外资工作，稳定外资规模，优化外资结构；加强社会信用体系建设，深入推进“放管服”改革，进一步发挥信用在创新监管机制、提高监管能力和水平方面的基础性作用；加快推进公共法律服务体系建设，全面提升公共法律服务能力和水平；全面提升政务服务规范化、便利化水

平，为企业和群众提供高效、便捷的政务服务，优化营商环境；促进经济持续健康发展和社会大局稳定发挥积极作用；进一步推动全国政府网站和政府系统政务新媒体健康有序发展；持续深化“放管服”改革，推行部门联合“双随机、一公开”监管，实现市场监管领域全覆盖等。

2019年出台的政府开放数据保持了高度的政策连续性，如为持续2018年“放管服”相关政策出台的《国务院关于在市场监管领域全面推行部门联合“双随机、一公开”监管的意见》、为进一步深化“互联网+政务服务”出台的《国务院办公厅关于建立政务服务“好差评”制度提高政务服务水平的意见》、为实现乡村振兴战略规划出台的《数字乡村发展战略纲要》等。良好的连续性能够使政策的实施具有更高的效率和针对性，达到建设人民满意的服务型政府的目的。通过文本分析的方式分析2019年度我国政府出台的政府开放数据相关政策的主题，主题的价值取向主要以服务为中心，以建设与发展为目的，以信息为途径，以监管和监督为手段，完善各类社会体系。生成词云结果如图1-2所示。



图1-2 2019年我国国家级政府开放政策文本主要词云

1.3.2 地方政策价值取向

2018年我国各地方为响应国家政府开放数据战略纷纷出台了关于本地区的政府开放数据发展行动和规划，为政府开放数据的利用发展营造出了良好的氛围。在此基础上，2019年我国各省级政府逐渐完善各自的顶层设计，坚决的落实了国务院的政策，但是还没有达到国家政府开放数据“泛在利用”的战略高度，使政府开放数据的利用效果不能够完全体现，没有使政府开放数据的利用融入于其他政策之中。同时除个别省份外，总体上仍然停留在“如何利用”的问题上。2019年度我国个省份出台的省级政府开放数据相关政策共124条，同比去年增长了29%。我国各省份2019年度关于政府开放数据政策文件数量示意图如图1-3所示。

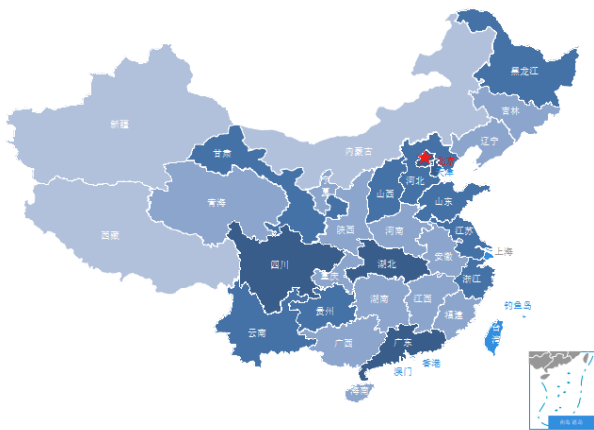


图1-3 我国各省份2019年度关于政府开放数据政策文件数量示意图

通过文本分析的方式分析2019年度我国政府省部级出台的政府开放数据相关政策的主题，主题的价值取向与国家政策价值取向相似。仍然是以服务为中心，以信息为途径，以建设和推进为主要目的。生成词云结果如图1-4所示。



图1-4 2019年我国省部级政开放政策文本主要词云

平台主要内容

2 政府数据开放平台主要内容

2.1 平台建设

政府数据开放平台是由各省级部门牵头，是支撑数字政府建设的数据资源服务架构系统，以大数据治理下数字政府建设为背景，以实现数字化政府服务形态的转变为目标，以创新政府数据开放的系统架构为依托，致力于各级政府部门可公开数据的下载与服务，能让公众在开展政务信息资源的社会化开发利用活动中，便捷地获取到易使用、高质量的政府开放数据，驱动重塑政务信息化架构下的新型政府形态，并推动信息资源增值服务业的发展及相关研究工作的开展。

课题组在2018年政府开放数据利用报告的基础上，增加了抽样规模，进一步扩大了平台搜索的范围。基于我国政府数据开放平台的类型划分，独立平台大多以“data.gov.cn”为域名，因此以“data.gov.cn”为域名对独立平台进行搜索；另外，针对嵌入在政府平台栏目中的数据，其平台类型为嵌入式平台，因此在对应的政府平台数据开放栏目中进行搜索。截止2019年12月，全国共有108个省、市或区政府建立了地方性政府数据开放平台(港澳台地区除外)，详细情况见附件3，2019年政府数据开放平台地理分布如图2-1所示。本节将从平台数量、类型、行政级别分布及地区分布四个方面分析2019年政府数据开放平台的建设情况。

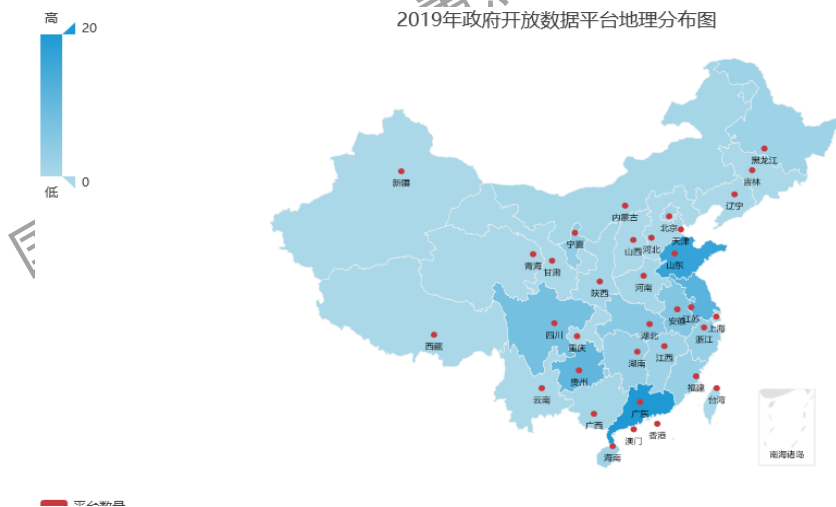


图2-1 2019年政府数据开放平台地理分布图

从平台上线数量来看，2019年我国新增政府数据开放平台46个，平台上线数量同比增长74.19%，新增平台增长速度较前几年处于中等水平，2019年政府数据开放平台的建设速度有所放缓，得益于前几年平台

建设的迅猛势头，我国大部分省市的政府数据开放平台都已陆续上线，因此平台增长率低于往年也情有可原。历年政府数据开放平台数量的具体增长情况如图2-2所示。

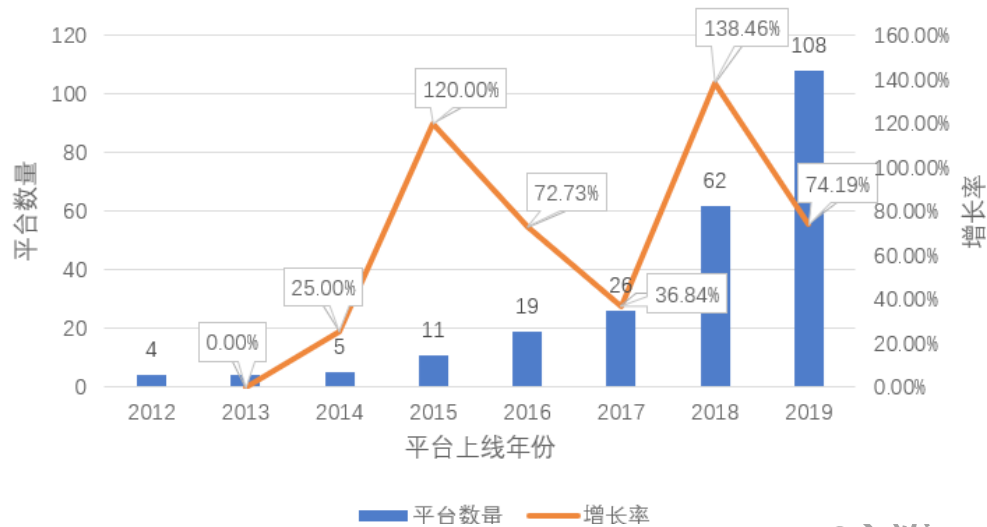


图2-2 2012年-2019年政府数据开放平台上线数量与增长率统计图

从平台所属的政府行政层级来看，我国省级、副省级与地市级政府的政府数据开放平台建设情况有所差异，2019年国内已上线的政府数据开放平台行政层级占比统计图如图2-3所示。据统中华人民共和国行政区划统计，全国共有31个省级行政区（港澳台除外）、15个副省级城市和333个地市级行政区。目前建有政府数据开放平台的副省级地区数量占市级总量的

66.67%，建有政府数据开放平台的省级行政区占比为51.61%，地市级城市中已建政府开放平台的地区占比最低，仅24.62%。由此可见，2019年，我国省级、副省级城市平台超过半数，总体上上线政府数据开放平台的地方占比有所提高，但是，地市级地区的平台建设占比依旧偏低，说明地方还有很大的发展空间。

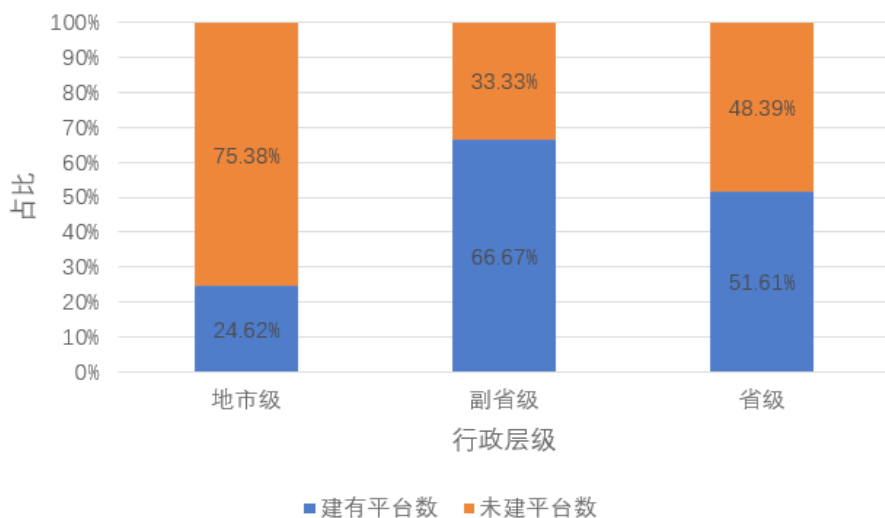


图2-3 2019年政府数据开放平台行政层级统计图

平台主要内容

从平台类型来看，我国82.41%的政府数据开放平台为独立平台，即数据统一汇聚在一个专门的平台上进行开放，仅雅安、阜阳等12个地区的政府数据开放平台为嵌入平台，这些地区的政府数据开放平台都嵌入在该地人民政府官网的栏目中，平台的建设情况也存在差异，部分嵌入式平台具有完备的开放数据主题分类、领域索引、数据接口以及数据应用详情页面，但也存在部分平台建设较为粗糙、功能不全等问题。

总得来说，独立的政府数据开放平台更加专业化，并且具备较好的用户体验，是未来政府数据开放平台的建设趋势。需要特别说明的是，虽然荆门市政府数据开放平台标题为“荆门市人民政府”，但就其内容和结构看，完全具备独立型平台的特征，故将其划分到独立型平台中。2019年政府数据开放平台类型统计详情如图2-4所示。



图2-4 2019年政府数据开放平台类型统计图

从平台所属地区来看，我国政府数据开放平台主要集中分布于华东、华南地区，其中华东地区的平台数量高达40个，占国内平台总量的37.04%，华南地区

的平台数量为27个，占国内平台总量的25%，华东华南地区的平台上线数在全国范围内处于领先地位。详情如图2-5所示。

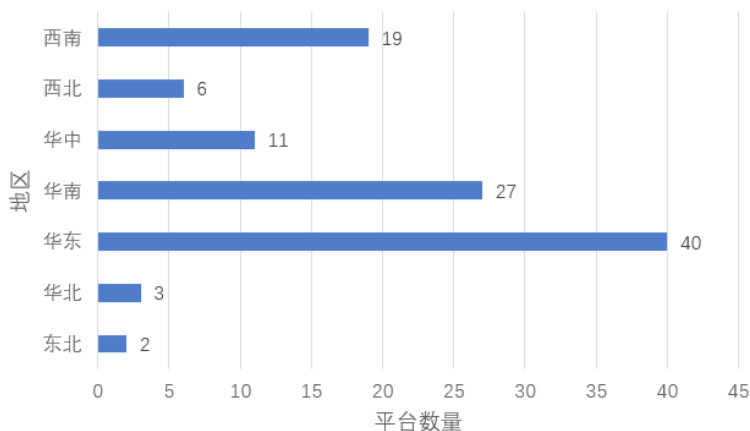


图2-5 2019年政府数据开放平台所属地区统计图

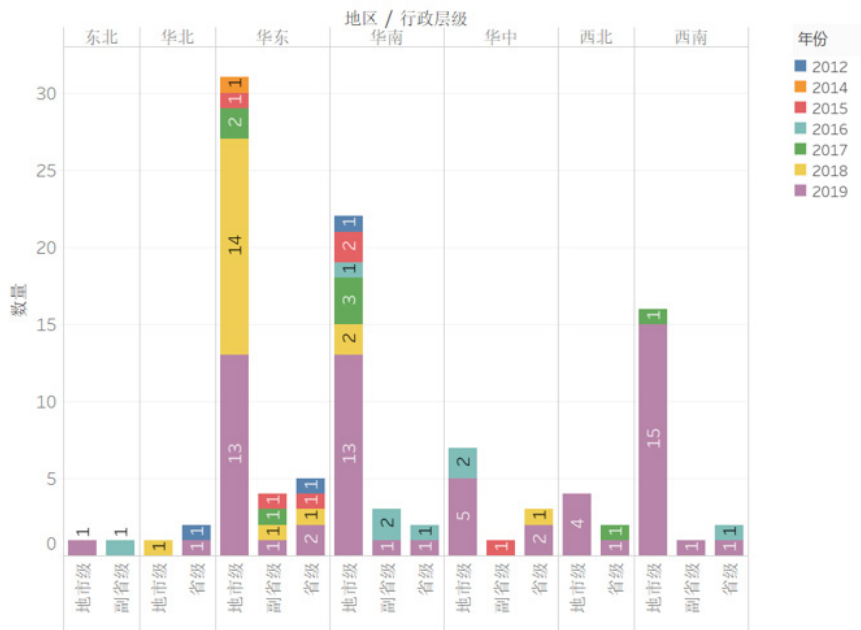


图2-6 2019年政府数据开放平台所属地区与行政层级统计图

结合现有平台的所属区域和行政层级可以了解到，平台总量排名靠前的地区分别是华东、华南，和西南地区。其中，华东地区的政府数据开放平台数量为40个，在全国范围内处于领先地位，2019年华东地区地市级市平台的增长量为13个，占2018年华东地区全年上线平台总量的81.25%，但地市级市平台增长率略低于2018年，总得来说，地市级平台建设速度最快，副省级、省级平台建设速度较缓，平台数量增长速率与2018年持平；华南地区现有政府数据开放平台共计27个，2019年该地区市级平台新增量（13个）高于前5年的市级政府数据开放平台的总量（9个），相比之下，省级、副省级平台建设的进展较为缓慢；西南地区的政府数据开放平台上线总量排名第三位，值得关注的是，2019年西南地区地市级市平台的增长最多（15个），是西南地区累计全年上线平台总量的7.5倍；说明西南地区地市级市平台建设速度迅猛。

综上，可以获悉2019年华东地区和华南地区的政府数据开放平台建设情况较好，各地区的新增地市级开放数据平台数量占比较大，备受重视，华东和华南地区表现最为明显。

2.2 数据集主题

为了对2019年已上线的108个政府数据开放平台的主题分类情况有一个直观的了解，对上述平台的数据集主题进行了统计与分析，数据集主题统计表如附件4所示，词云效果图如图2-7所示。

结合现有平台主题大类的关键词提取结果和词云来看，关键词权重排名靠前的词语分别是“服务”、“安全”、“资源”、“环境”、“团体”、“公共”、“机构”、“信用”、“监管”、“文化”、“经济”。一方面，“服务”一词的出现频率最高，

平台主要内容

体现了政府开放数据面向公众开放，为公众提供信息服务的宗旨；另一方面，通过高频词我们可以获悉当下政府开放数据主要集中在安全、环境、机构团体、信用、文化、经济等社会生活领域。



图2-7 2019年部分政府数据开放平台主题大类词云

通过对政府数据开放平台的持续性关注与分析，发现相比于过去各城市平台开放的数据主题的分类标准存在较大差异的情况。2019年政府数据开放平台的主题分类标准呈现由分散到集中的统一化趋势，如山东的政府数据开放平台统一使用“信用、交通运输、社会保障、地理、教育、生态环境、金融、医疗、卫生、就业、文化、科技、资源、农业、安全监管、质量、统计、气象、海洋、企业登记监管”这一分类体

系；广东省内的12个政府数据开放平台统一使用“资源环境、机构团体、城市建设、社会资源、经济建设、教育科技、道路交通、社会发展、公共安全、文化休闲、卫生健康、民生服务”作为政府数据开放平台主题分类。

综上所述，目前各地的政府数据开放平台使用的主题分类具备一定的地域统一性，同一地区内的平台数据建设与维护更加便利，方便实现了数据的互联互通与公众利用。但地区与地区之间还缺乏统一的主题分类标准，在一定程度上阻碍了不同地区之间政府开放数据的共享与联通。

2.3 数据集数量

从国内现有政府数据开放平台来看，数据集、数据接口和数据应用是平台开放的主要内容，可以直观反映出平台中政府开放数据的规模，2019年各平台的数据集情况如附件5所示。图2-8，2-9和2-10分别为2019年部分政府数据开放平台数据集数量统计图、2019年部分政府数据开放平台数据接口数量统计图、2019年部分政府数据开放平台数据应用数量统计图。

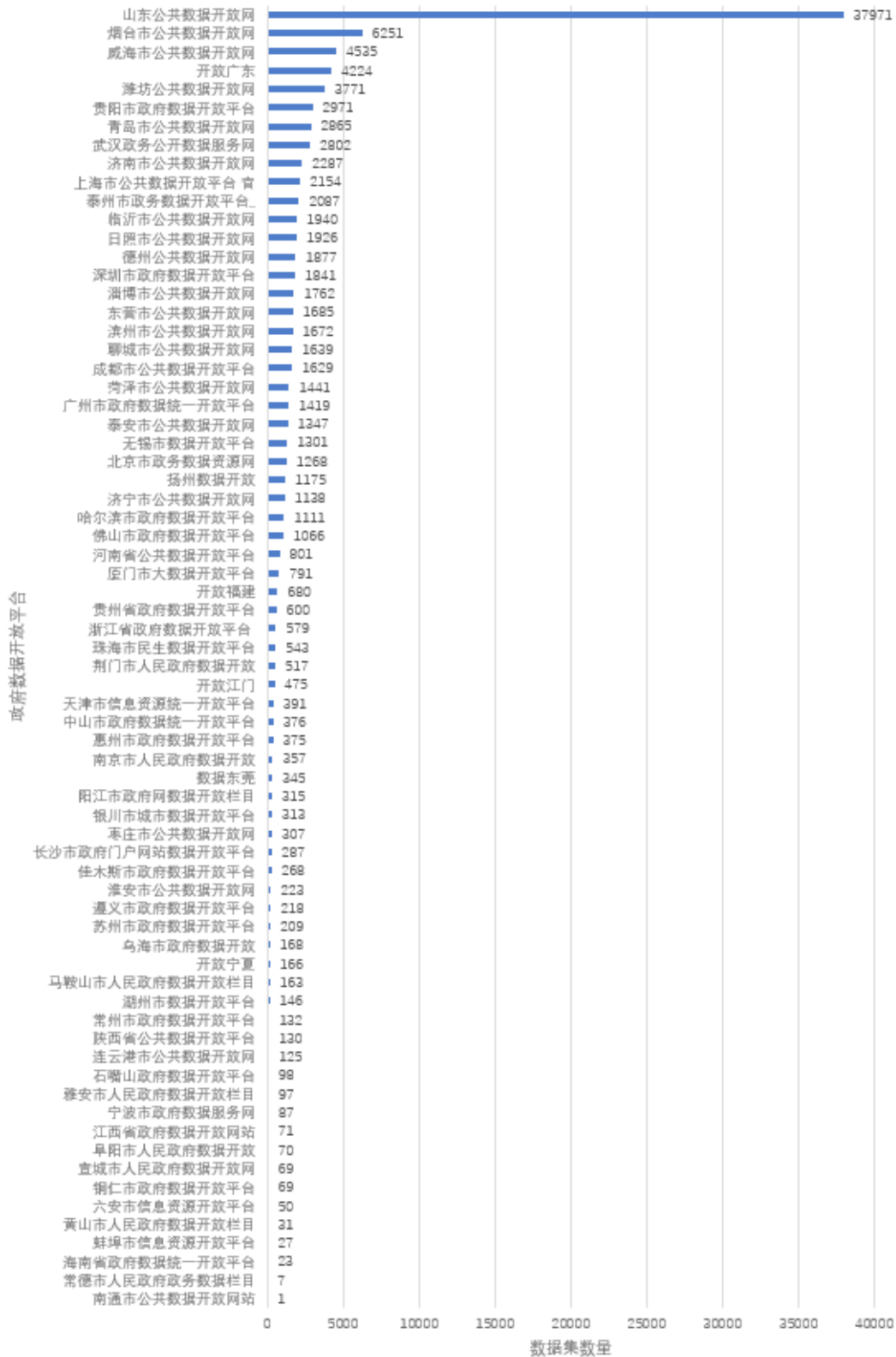


图2-8 2019年部分政府数据开放平台数据集数量统计图

平台主要内容

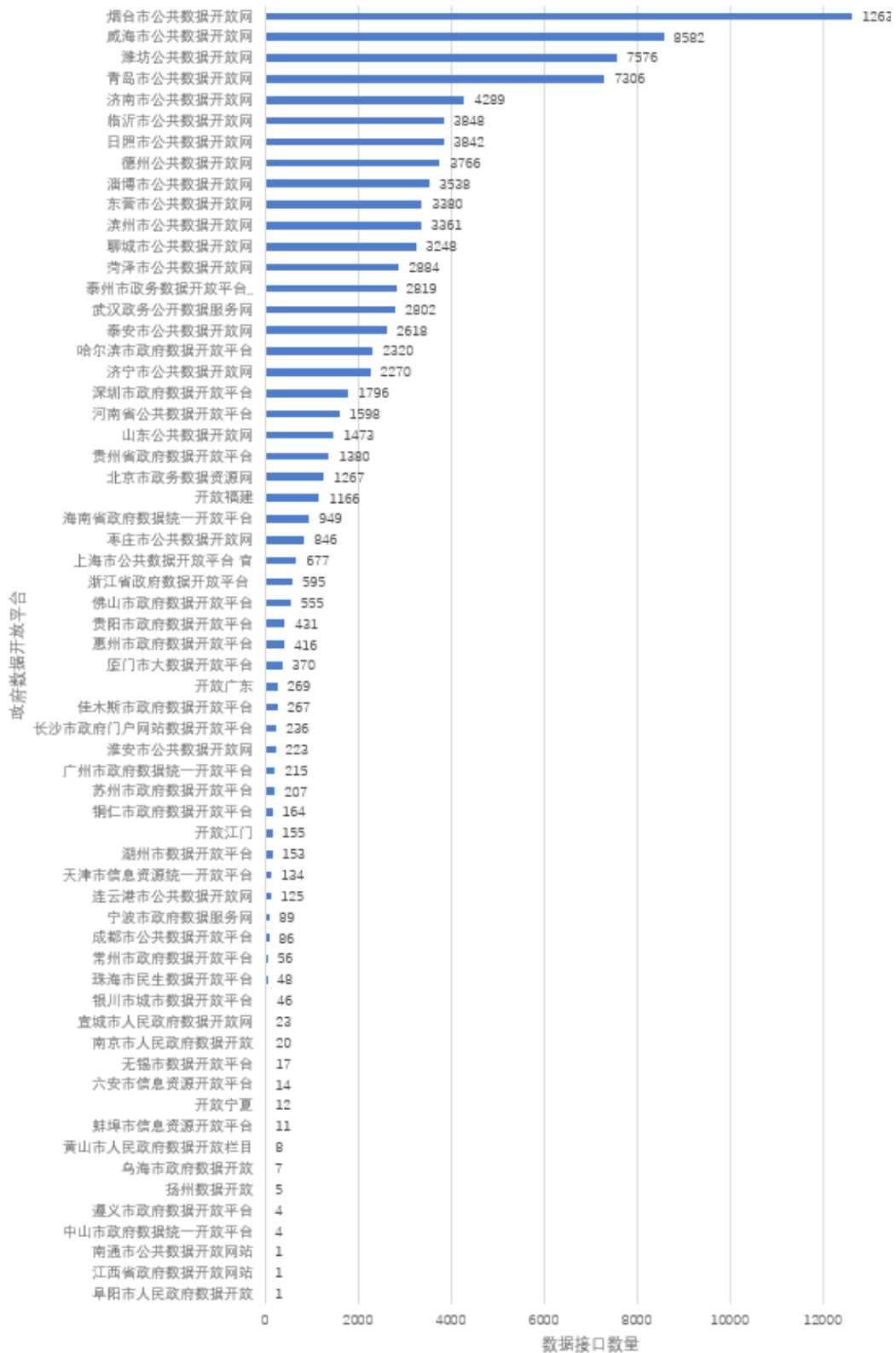


图2-9 2019年部分政府数据开放平台数据接口数量统计图

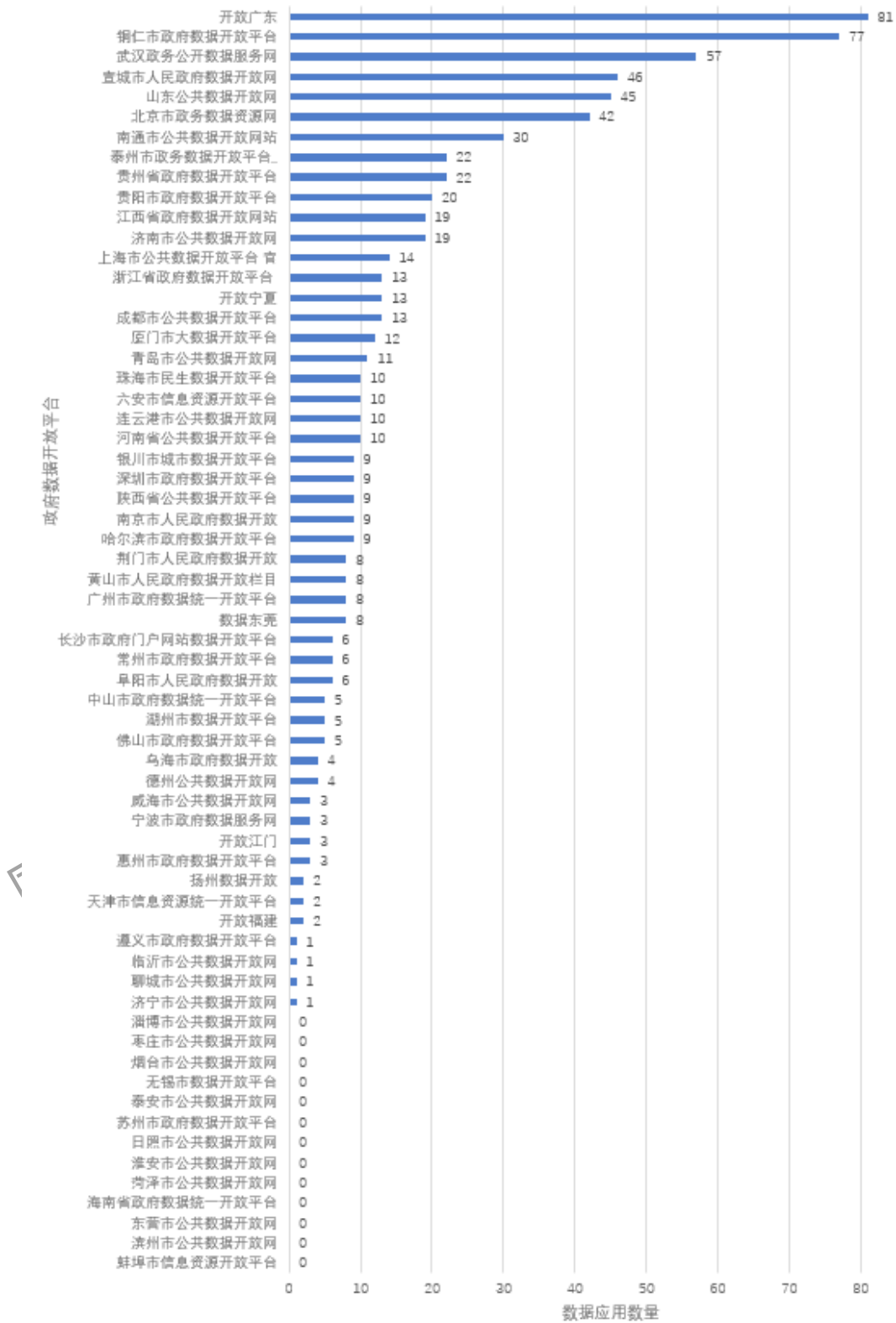


图2-10 2019年部分政府数据开放平台数据应用数量统计图

平台主要内容

从2019年现有政府数据开放平台的数据集数量来看，前三名的平台分别是山东公共数据开放网、烟台公共数据开放网和威海市公共数据开放网，在政府开放数据集数量排名前十的数据中，山东地区平台占比为60%。由此，我们可以获悉在全国范围内，山东地区开放的数据集数量的规模最大，其次是开放广东、贵阳市政府数据开放平台。相比之下，南通市、常德地区的政府数据开放平台开放的数据集不足十个，一定程度上受平台上线时间的限制，平台数据集的建设会随着时间的推移逐渐完善。

从2019年现有政府数据开放平台的数据接口数量来看，政府开放数据接口排名前十的数据都属于山东地区的平台，在全国范围内，山东地区开放的接口数量也呈现出领先优势。而东莞、阜阳、江西、南通等地的平台数据接口数量处于较低水平，此外，图2-9仅对政府开放数据中接口栏目中的现有接口量进行了统计，一方面，结合平台类型可以获悉，嵌入式平台大多未设立政府开放数据接口栏目，进一步凸显了独立式平台内容的全面性和专业性，部分独立平台也存在

缺乏接口数据的情况，政府数据开放平台的建设仍需不断推进；另一方面，部分平台存在接口数据嵌套在数据集栏目下的情况，如东莞的数据接口。为此，亟需对政府数据开放平台的栏目建设进行统一化管理，提升接口数据的可用性和易用性。

从2019年现有政府数据开放平台的数据应用数量来看，排名前十的平台分别为隶属于广东、贵州、湖北、安徽、山东、北京和江西省，说明在上述地区，有效利用政府开放数据资源，开发应用产品，数据应用的开发成果的数量领先其他地区。然而，目前已有13个地区存在数据应用数量为零的情况，其中存在一些建设较早的平台（如无锡市数据开放平台），因此，各地方应注重数据应用的建设与开发，便民利民，提高政府数据治理水平。

总得来说，政府相关部门应加快政府数据开放平台的数据集、数据接口的开放速度，完善平台建设规范，重视数据应用的开发，使用户能更加便捷、有效的利用政府数据，从中创造价值。

3 政府开放数据利用行为分析

政府开放数据的最终目的是实现和促进数据的使用与开发，本报告以数据本身作为切入点，分别从主题视角和区域视角两个层面出发，分析公众对不同主题开放数据的实际需求、了解我国各地域之间开放数据利用的差距；运用数据浏览率、下载率等指标对比分析开放数据的利用程度，通过聚类分析把握不同主题、不同区域政府开放数据的利用现状，以期促进我国政府开放数据工作的发展，提高我国政府开放数据的利用效率。

3.1 政府开放数据利用行为概况

3.1.1 主题视角下政府开放数据利用行为概况

为了揭示不同主题视角下开放数据的利用概况，首先统计各主题数据的浏览量、下载量，如图3-1所示；

此外，统计分析了各主题的单一样本开放数据平均浏览率、下载率和整体样本开放数据平均浏览率、下载率，分别如图3-2和图3-3所示。由图3-1可知，主题视角下浏览量排在前三位的是能源环境、社保就业、经济工商，主题视角下下载量排在前三位的是社保就业、经济工商、文体休闲，由此可知，经济工商和社保就业等主题的数据备受用户关注。此外，结合图3-2可知，主题视角下单一样本开放数据的平均浏览率均高于平均下载率，但仍然有部分主题在该指标的数值明显低于单一样本平均值，如信用服务主题、财税金融主题的单一样本平均浏览率均不足平均值608次/条，医疗健康主题、机构团体主题的单一样本平均下载率均不足平均值160次/条。结合图3-3可知，整体样本开放数据的平均浏览率和平均下载率之间有高有低，同样有部分主题在该指标的数值明显低于整体样本平均值，如信用服务主题的整体样本平均浏览率和平均下载率均不足0.071次/条，这表明各主题数据被关注的程度存在差异性。

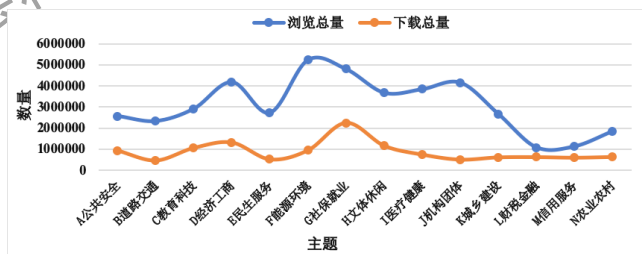


图3-1 主题视角下开放数据的浏览量和下载量

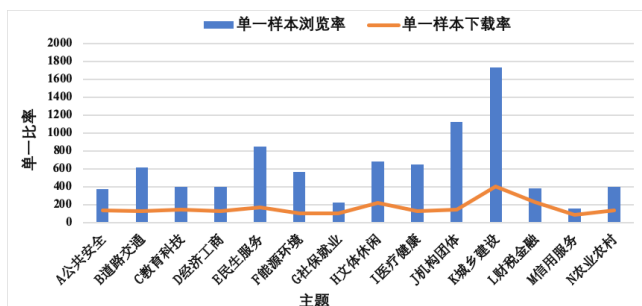


图3-2 主题视角下单一样本开放数据的平均浏览率和平均下载率

利用行为分析

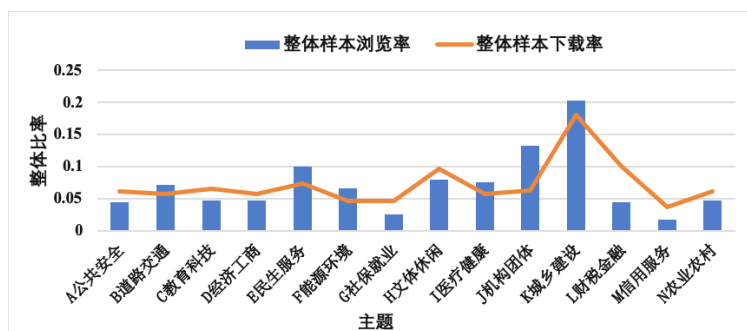


图3-3 主题视角下整体样本开放数据的平均浏览率和平均下载率

3.1.2 区域视角下政府开放数据利用行为概况

为了揭示区域视角下不同地区开放数据的利用概况，首先统计了各区域数据的浏览量和下载量，如图3-4所示；此外，统计分析了各区域的单一样本开放数据平均浏览率、下载率和整体样本开放数据平均浏览率、下载率，分别如图3-5和图3-6所示¹。由图3-4可知，区域视角下浏览量排在前三位的区域是浙江、深圳、上海；区域视角下下载量排在前三位的区域是浙

江、山东、上海，由此可知，上海、浙江、深圳等地区的政府开放数据备受用户关注。此外，结合图3-5可知，区域视角下，单一样本开放数据平均浏览率排在前三位的城市是浙江、黄山、深圳，单一样本平均下载率排在前三位的城市是黄山、浙江、上海。结合图3-6可知，区域视角下，整体样本开放数据平均浏览率排在前三位的城市是浙江、黄山、深圳，整体样本开放数据平均下载率排在前三位的城市是黄山、浙江、上海，这表明各区域数据被关注的程度存在差异性。

1 注：陕西、南京、石嘴山、无锡、铜仁五个地区的开放平台中并未提供数据的浏览量和下载量指标数值，故将其相应数值均记为0值。

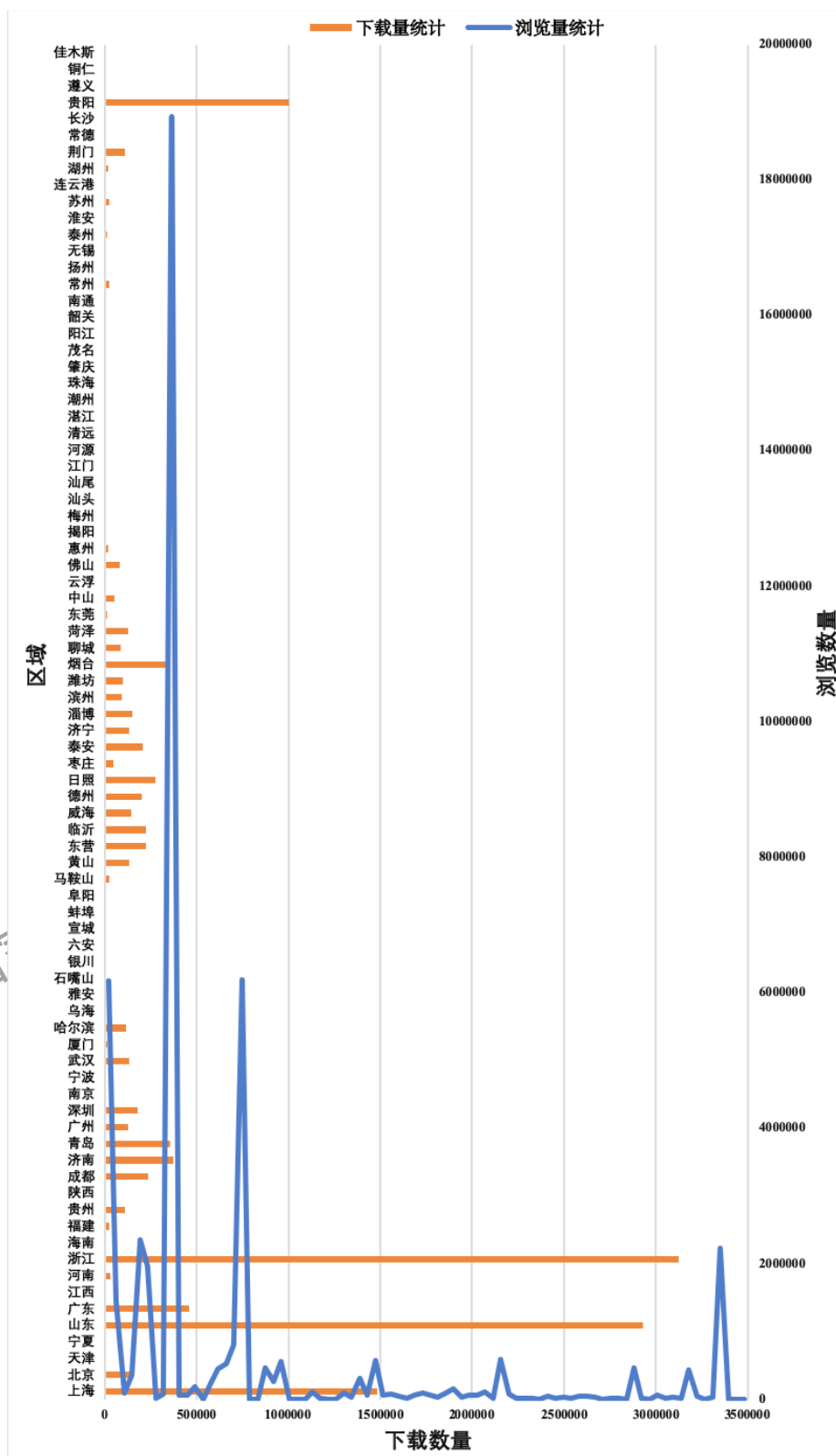


图3-4 区域视角下开放数据的浏览量和下载量

利用行为分析

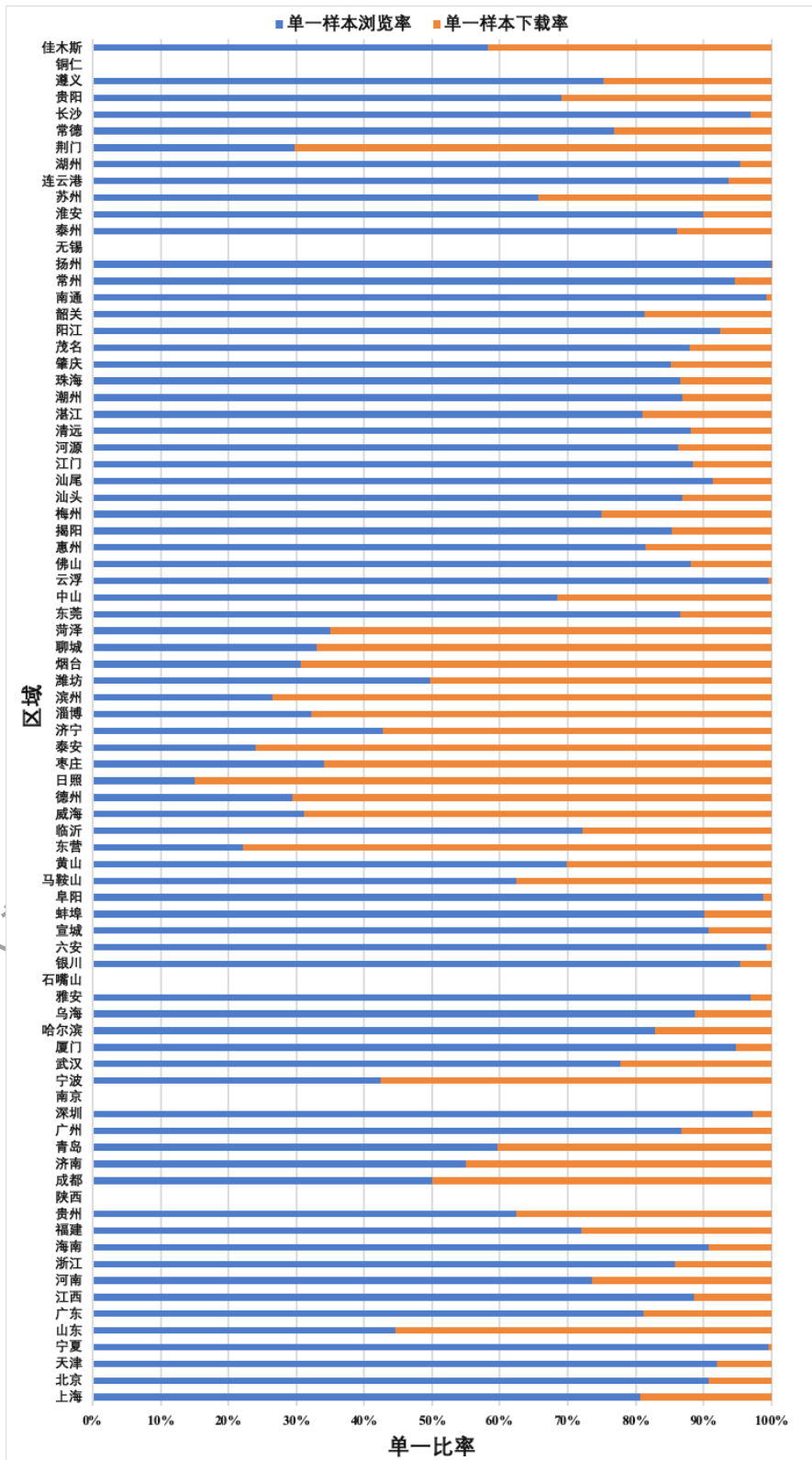


图3-5 区域视角下单一样本开放数据的平均浏览率和平均下载率

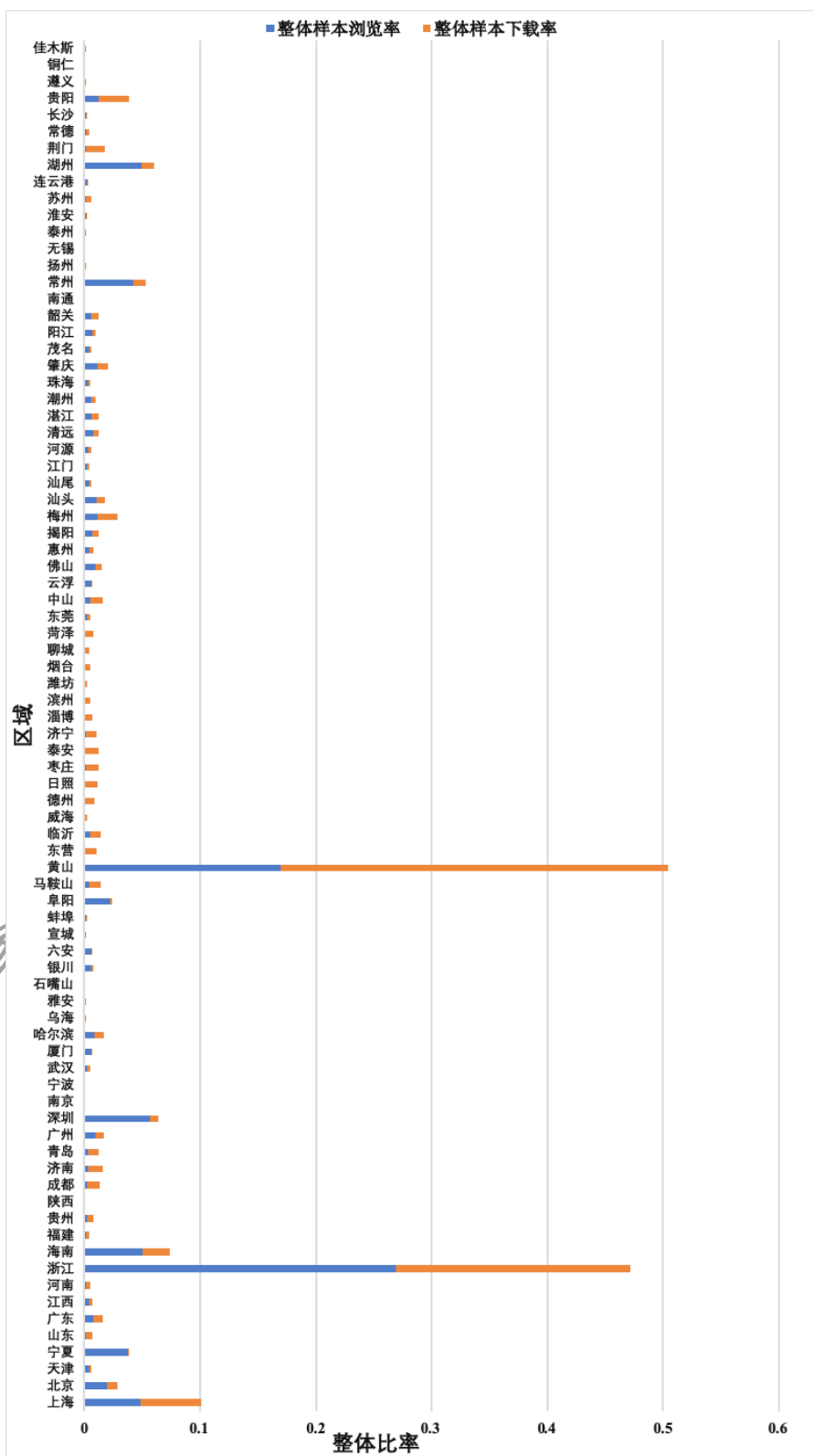


图3-6 区域视角下整体样本开放数据的平均浏览率和平均下载率

利用行为分析

3.2 政府开放数据浏览行为分析

3.2.1 主题视角下开放数据浏览率

浏览率能直观反映用户对某一主题数据的浏览行为和关注情况。首先绘制各主题大类开放数据的浏览率折线图和整体样本浏览率散点图，如图3-7所示。其

中，X轴代表14个主题大类，分别为：公共安全、道路交通、教育科技、经济工商、民生服务、能源环境、社保就业、文体休闲、医疗健康、机构团体、城乡建设、财税金融、信用服务、农业农村。此外，左侧Y轴代表开放数据的浏览率数值，右侧Y轴代表开放数据的整体样本浏览率数值。

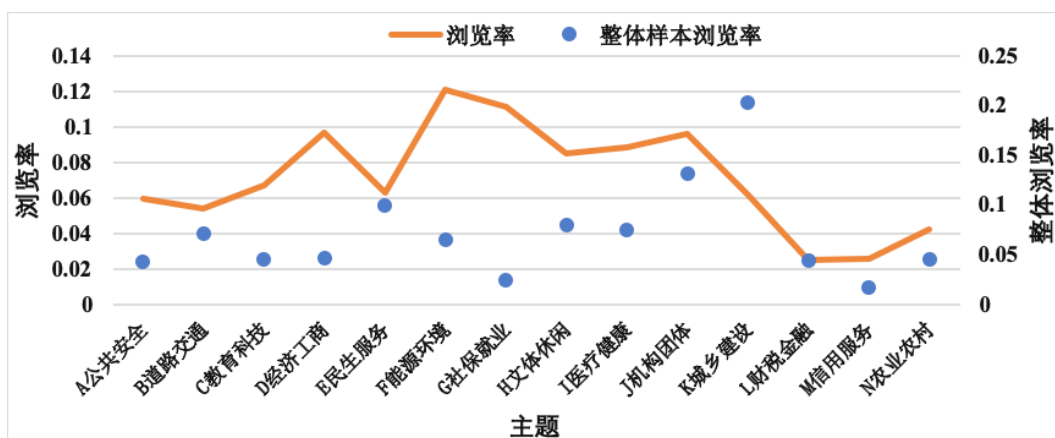


图3-7 主题视角下开放数据浏览率折线图和散点图

由图3-7可知，浏览率最高的主题是能源环境（约0.121），最低的主题是财税金融（约0.025），前者是后者的约4.84倍；且这14大主题中能源环境、社保就业、经济工商、机构团体、医疗健康、文体休闲主题数据的浏览率均高于平均值（约0.071）。整体样本浏览率最高的主题是城乡建设（约0.203），最低的主题是信用服务（约0.018），前者是后者的约11.28倍；且这14大主题中城乡建设、机构团体、民生服务、文体休闲、医疗健康、道路交通主题数据的整体样本浏览率均高于平均值（约0.071）。

此外结合图3-1、图3-2、图3-3，对比图3-7中开放数据的浏览率折线图和散点图可知，能源环境主题数据的浏览率（约0.121）虽然最高，但由散点图的分布可知，其整体样本浏览率并不高（约0.066）。而

城乡建设主题的数据却恰好相反，虽然该数据集的浏览率较低（约0.062），但其整体浏览率数值最高（约0.203），这表明各主题开放数据整体与部分的特征并不具有一致性。

3.2.2 区域视角下开放数据浏览率

同主题视角下开放数据的浏览率一样，将各区域的浏览率折线图与条形图绘制在一起，如图3-8所示，便于观察各地区开放数据的关注程度与状况。其中，左侧Y轴代表82个区域，分别为：上海、北京、天津、山东、广东、江西、.....、荆门、常德、长沙、贵阳、遵义、佳木斯。此外，右侧Y轴代表开放数据的浏览率数值，X轴代表开放数据的整体样本浏览率数值。

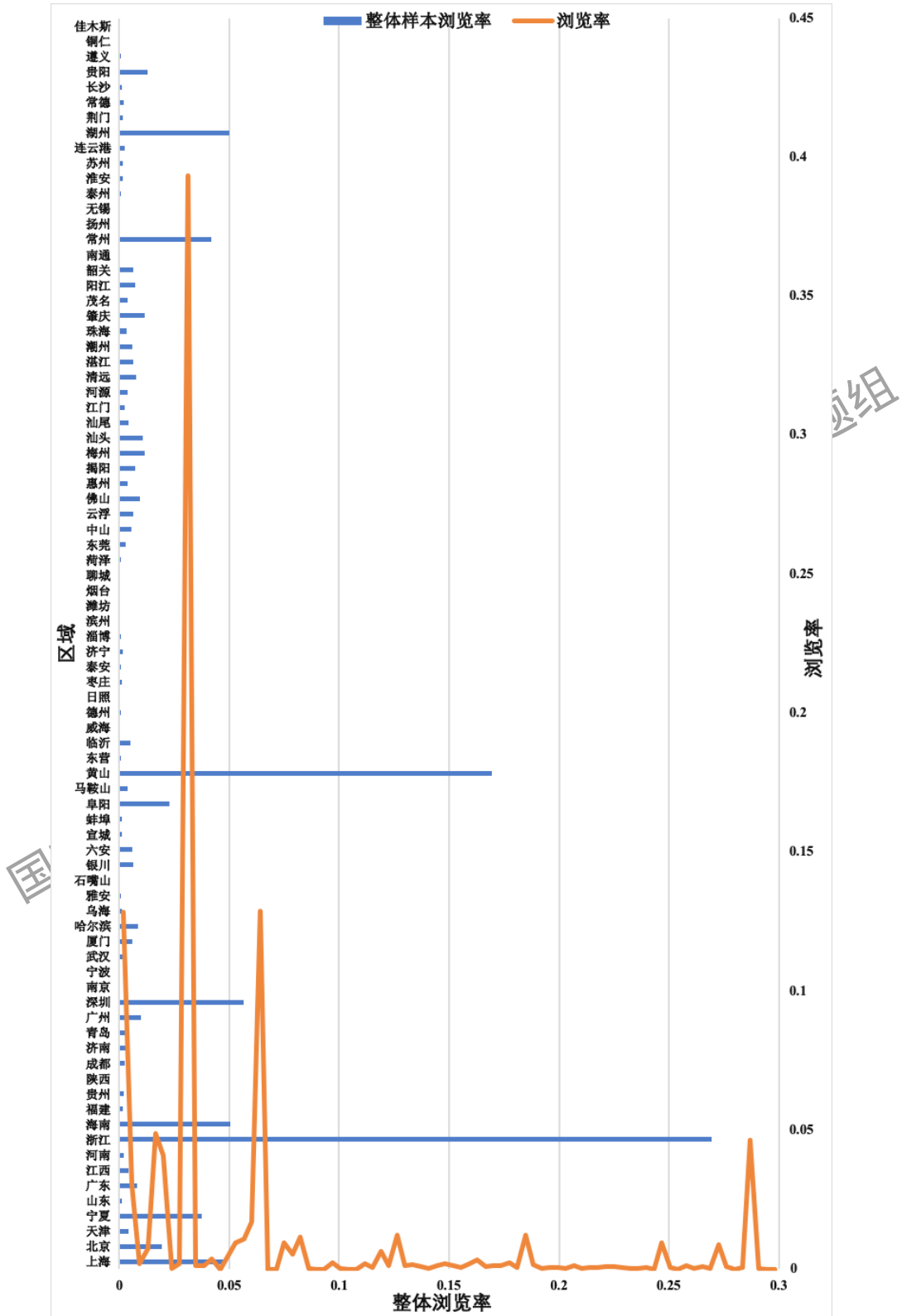


图3-8 区域视角下开放数据浏览率折线图和条形图

利用行为分析

由图3-8可知，区域视角下浏览率最高的地区是浙江、深圳、上海、山东等，整体样本平均浏览率最高的地区是浙江、黄山、深圳、上海等，说明浙江、深圳、上海地区开放的数据，备受用户关注。以上海为例，其数据浏览率达到0.129，远高于平均值（约0.012）；同时其整体样本平均浏览率（约0.048361）是威海（约0.000238）的203.197倍。与此形成对比的是，南通和宁波地区，在数据浏览率和整体样本平均浏览率这两项指标方面均排在末尾，其余各地排名稍有变化。

此外结合图3-4、图3-5、图3-6，对比图3-8中开放数据的浏览率折线图和条形图可知，黄山地区数据的整体样本平均浏览率虽然高达0.16932（排名第2），但其数据浏览率却只有0.00648，远低于平均值（约0.012）。而山东地区的数据却正好相反，虽然数据浏览率高达0.04899（排名第4），但其整体样本平均浏览

率只有0.00104，远低于平均值（约0.012）。这表明各地区开放数据整体与部分的特征并不具有一致性。

3.3 政府开放数据下载行为分析

3.3.1 主题视角下开放数据下载率

下载率是对浏览率的进一步说明与深化，它在很大程度上能反映出用户对某一数据的利用情况。同浏览率一样，图3-9所示的是各主题大类开放数据的下载率折线图和散点图。其中，X轴代表14个主题大类，分别为：公共安全、道路交通、教育科技、经济工商、民生服务、能源环境、社保就业、文体休闲、医疗健康、机构团体、城乡建设、财税金融、信用服务、农业农村。此外，左侧Y轴代表开放数据的下载率数值，右侧Y轴代表开放数据的整体样本下载率数值。

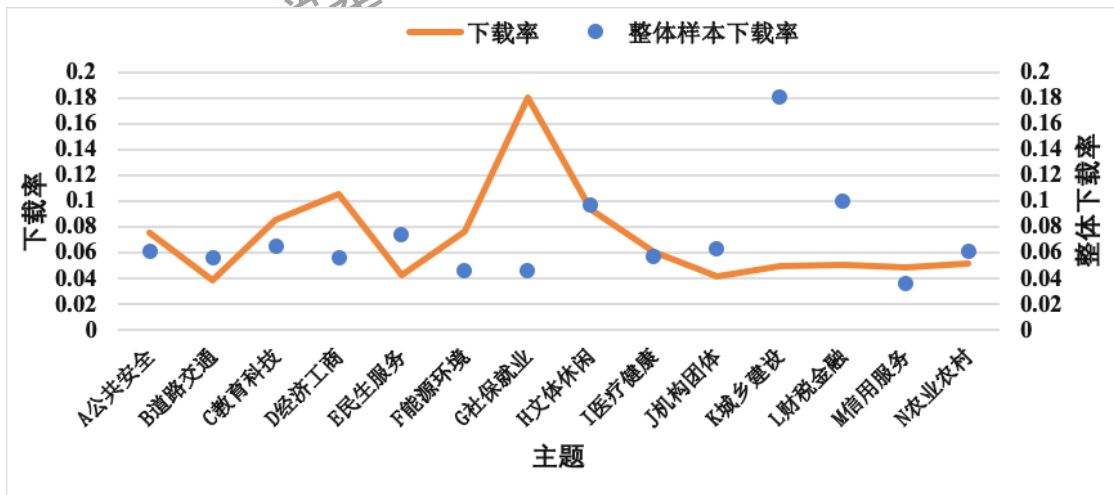


图3-9 主题视角下开放数据下载率折线图和散点图

由图3-9可知，下载率最高的主题是社保就业（约0.180），最低的主题是道路交通（约0.038），前者是后者的约4.74倍；且这14大主题中社保就业、经济

工商、文体休闲、教育科技、能源环境、公共安全主题数据的下载率均高于平均值（约0.071）。整体样本下载率最高的主题是城乡建设（约0.180），最低的

主题是信用服务（约0.036），前者是后者的约5倍；且这14大主题中城乡建设、财税金融、文体休闲、民生服务主题数据的整体样本下载率均高于平均值（约0.071）。这间接表现出用户对此类数据利用的真实性，表明被利用的程度较高，同时也反映出其余几类数据的利用程度有待提高。

此外结合图3-1、图3-2、图3-3，对比图3-9中开放数据的下载率折线图和散点图可知，社保就业主题数据的下载率（约0.180）虽然最高，但由散点图的分布可知，其整体样本下载率并不高（约0.046），远低于平均值（约0.071）。而城乡建设主题的数据却恰好相反，虽然该数据集的下载率较低（约0.049），远低于平均值（约0.071），但其整体浏览率数值最高（约0.180），这表明各主题开放数据整体与部分的特征并不具有一致性。

3.3.2 区域视角下开放数据下载率

各区域开放数据的下载率反映了用户对各城市开放数据的利用程度。同主题视角下开放数据的下载率一样，将各区域的下载率折线图与条形图绘制在一起，如图3-10所示，便于观察各地区开放数据的关注程度与状况。其中，左侧Y轴代表82个区域，分别为：上海、北京、天津、山东、广东、江西、……、荆门、

常德、长沙、贵阳、遵义、佳木斯。此外，右侧Y轴代表开放数据的下载率数值，X轴代表开放数据的整体样本下载率数值。

由图3-10可知，区域视角下下载率最高的地区是浙江、山东、上海、贵阳等，整体样本平均浏览率最高的地区是黄山、浙江、上海、贵阳等，说明上海、贵阳、浙江地区开放的数据，备受用户关注。以浙江为例，其数据下载率达到0.225758，远高于平均值（约0.012）；同时其整体样本平均浏览率（约0.202709）是宁波（约0.000282）的718.826倍。与此形成对比的是，南通和扬州地区，在数据浏览率和整体样本平均浏览率这两项指标方面均排在末尾，其余各地排名稍有变化。

此外结合图3-4、图3-5、图3-6，对比图3-10中开放数据的浏览率折线图和条形图可知，黄山地区数据的整体样本下载率虽然高达0.334655（排名第1），但其数据下载率却只有0.009767，远低于平均值（约0.012）。而山东地区却正好相反，虽然其数据下载率高达0.211765（排名第2），但其整体样本平均下载率只有0.005924，远低于平均值（约0.012）。这表明各地区开放数据整体与部分的特征并不具有一致性；此外，这一特征与区域视角下开放数据浏览率的特征相一致。

利用行为分析

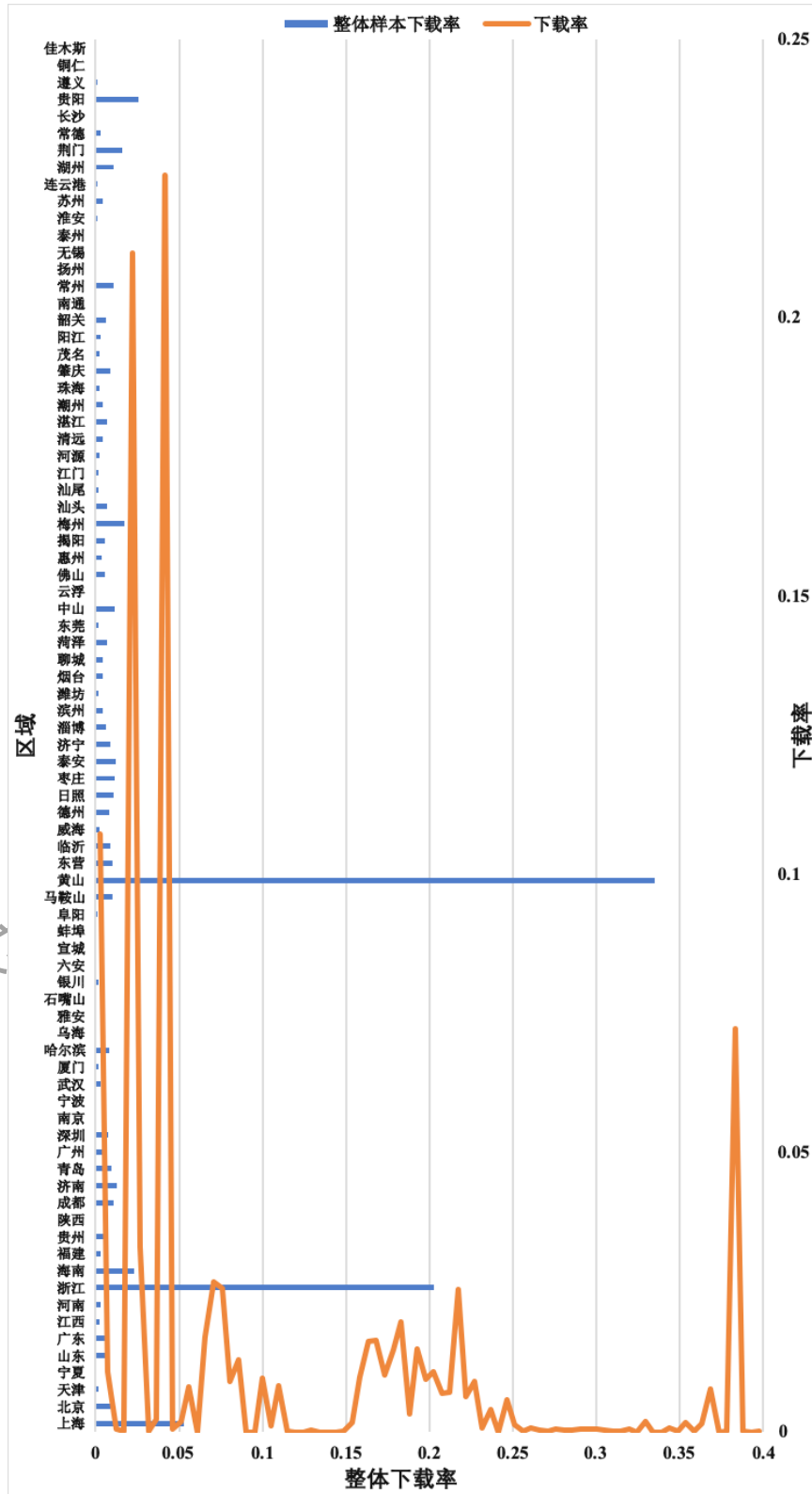


图3-10 区域视角下开放数据浏览率折线图和条形图

3.4 政府开放数据利用行为综合分析

3.4.1 主题视角下政府开放数据利用行为综合分析

为深入揭示各主题开放数据的利用现状与亲疏关系，对14个主题大类做聚类分析。聚类分析是指在事先不规定分组规则的情况下，将数据按其自身特征划分成不同的群组，各群组内部数据差距尽可能的小，

而各群组数据之间的差距尽可能的大。首先选择聚类指标，由于单一样本和整体样本的开放数据平均浏览率（下载率）的变化方向一致且一一对应，前者是测算后者的基础，因此聚类分析主要参考开放数据浏览率、下载率和整体样本开放数据平均浏览率、下载率4个指标的相关情况。与此同时，选用层次聚类法，其中个体距离采用平方欧式距离，类间距离采用Ward联接，最终聚类形成的层次分析图如图3-11所示。

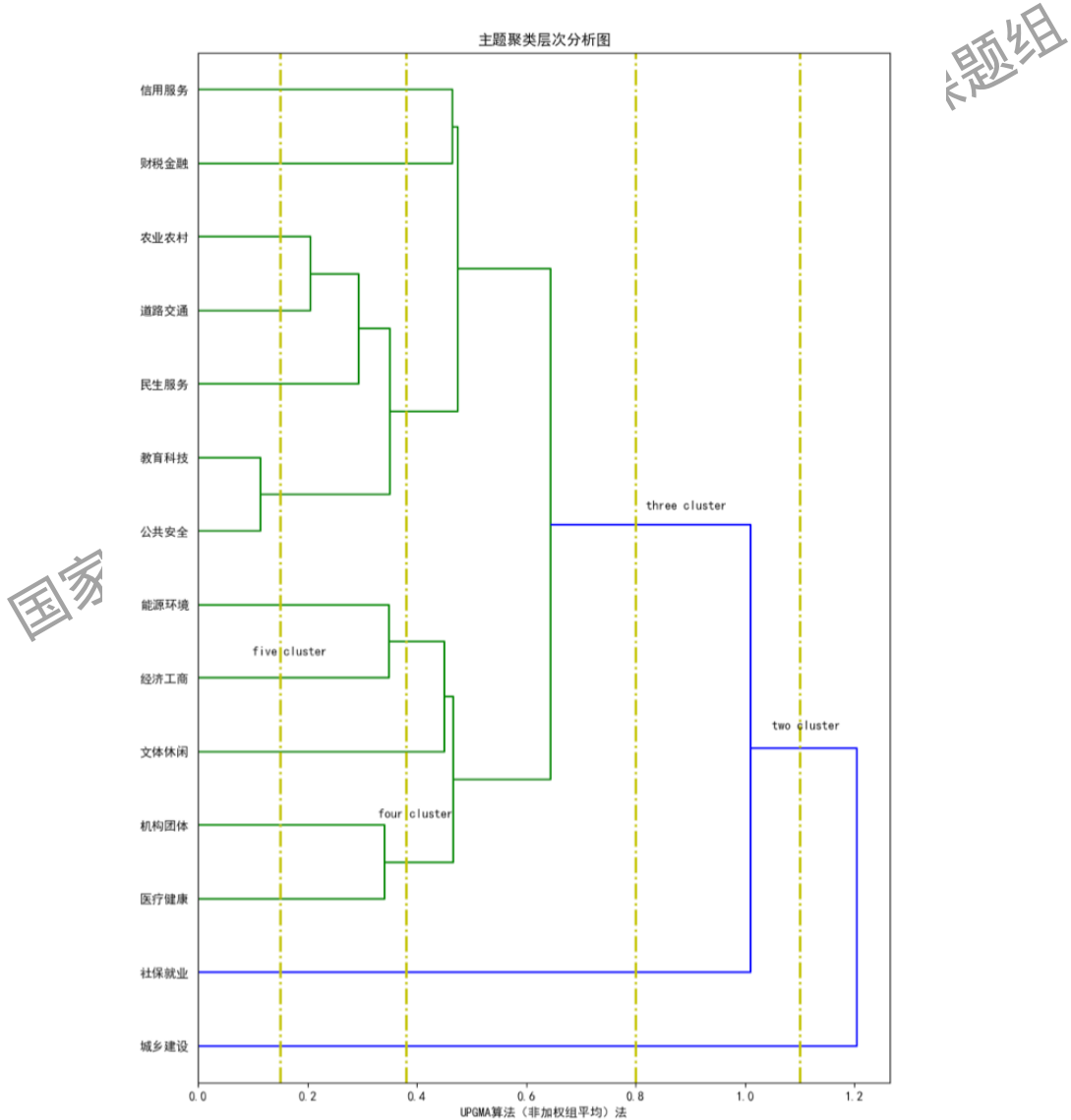


图3-11 主题视角下开放数据的层次聚类图

利用行为分析

为了更好地划分聚类数目，绘制了主题视角下开放数据聚类的系数图，如图3-12所示。由图3-12可知，随着类的不断凝聚和类目数量的不断减少，各类之间的距离迅速增大，碎石图逐渐趋于平坦。观察聚合系数图可知，当类别数为4或5时，折线的下降趋势趋缓，故将主题类别数设定为4或5类较好。根据冰柱图和树状图可知，当聚成4类时，“城乡建设”和“社保就业”聚为一类，“医疗健康”和“机构团体”聚为一类，“能源环境”“经济工商”和“文体休闲”聚为一类，“公共安全”“道路交通”“教育科

技”“民生服务”“农业农村”和“财税金融”“信用服务”全部聚为一类；当聚成5类时，“城乡建设”和“社保就业”聚为一类，“医疗健康”和“机构团体”聚为一类，“能源环境”“经济工商”和“文体休闲”聚为一类，“公共安全”“道路交通”“教育科技”“民生服务”和“农业农村”单独聚为一类，“财税金融”和“信用服务”单独聚为一类。为了更为细致的分析主题视角下开放数据利用的分布情况，本报告最终将14个主题的研究样本聚成5类。

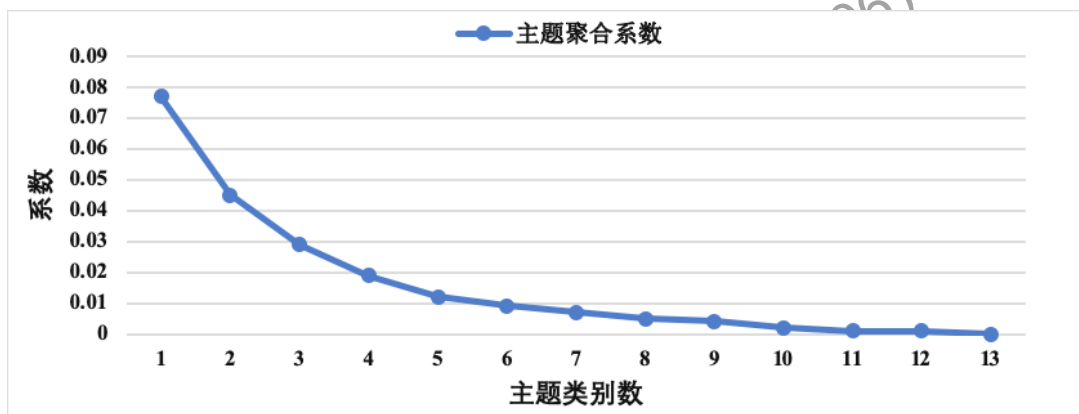


图3-12 主题视角下开放数据聚类系数图

第一类是“社保就业”和“城乡建设”。该类主题数据主要涉及某地区社会保障、社会发展、劳动人事、就业和城乡规划、住房保障、城市建设、村镇建设等方面的信息，其关注情况（浏览量）主要是由浙江、上海、贵阳、深圳、山东等地区的政府开放数据贡献的，其利用情况（下载量）主要是由山东、贵阳、浙江、上海、烟台等地区的政府开放数据贡献的。其中，“社保就业”主题数据在浏览率、下载率2项指标方面均超过相应平均值，但在单一及整体样本浏览率、下载率4项指标方面均未达到相应平均值；“城乡建设”主题数据在浏览率、下载率2项指标方面

均未达到相应平均值，但在单一及整体样本浏览率、下载率4项指标方面均超过相应平均值。

第二类是“医疗健康”和“机构团体”。该类主题数据主要涉及某地区医疗、卫生、健康，和政府机构、社会团体、社会组织等方面的信息，其关注情况（浏览量）主要是由深圳、浙江、上海、广东、贵阳等地区的政府开放数据贡献的，其利用情况（下载量）主要是由浙江、上海、山东、深圳、贵阳等地区的政府开放数据贡献的。其中，关注（浏览）行为方面，“医疗健康”和“机构团体”主题数据在浏览率和整体样本浏览率2项指标方面均超过相应平均值。利

用（下载）行为方面，“医疗健康”和“机构团体”主题数据在下载率和整体样本下载率2项指标方面均未达到相应平均值。

第三类是“经济工商”“能源环境”和“文体休闲”。该类主题数据主要涉及某地区经济建设、经济发展、商贸流通、商业贸易，资源能源、生态环境，文化、体育、休闲娱乐等方面的信息，其关注情况（浏览量）主要是由浙江、上海、深圳、广东、山东等地区的政府开放数据贡献的，其利用情况（下载量）主要是由浙江、山东、上海、贵阳、广东等地区的政府开放数据贡献的。其中，“经济工商”“能源环境”和“文体休闲”主题数据在浏览量、下载率2项指标方面均超过相应平均值；“经济工商”“能源环境”主题数据在整体样本浏览量、整体样本下载率2项指标方面均未达到相应平均值，“文体休闲”主题数据在整体样本浏览量、整体样本下载率2项指标方面均超过相应平均值。

第四类是“公共安全”“道路交通”“教育科技”“民生服务”和“农业农村”。该类主题数据主要涉及某地区生活安全、工作安全、人身安全，交通服务、交通运输，教育、科技、创新，民生生活、社区服务，农业服务、农村服务、农民服务等方面的信息，其关注情况（浏览量）主要是由浙江、上海、深圳、广东、北京等地区的政府开放数据贡献的，其利用情况（下载量）主要是由浙江、山东、上海、贵阳、济南等地区的政府开放数据贡献的。其中，关注（浏览）行为方面，“公共安全”“教育科技”“农业农村”主题数据在浏览率和整体样本浏览量2项指标方面均未达到相应平均值；“道路交通”“民生服务”主题数据在浏览率和整体样本浏览量2项指标方面均超过相应平均值。利用（下载）行为方面，“公共

安全”“道路交通”“教育科技”“农业农村”主题数据在下载率和整体样本下载率2项指标方面均未达到相应平均值；“民生服务”主题数据在下载率和整体样本下载率2项指标方面均超过相应平均值。

第五类是“财税金融”和“信用服务”。该类主题数据主要涉及某地区财政、金融，信用等方面的信息，其关注情况（浏览量）主要是由浙江、深圳、贵阳、山东、上海等地区的政府开放数据贡献的，其利用情况（下载量）主要是由浙江、山东、烟台、贵阳、成都等地区的政府开放数据贡献的。其中，“财税金融”“信用服务”主题数据在浏览量、下载率、整体样本浏览量3项指标方面均未达到相应平均值；“财税金融”主题数据在整体样本下载率1项指标方面超过相应平均值，“信用服务”主题数据在整体样本下载率1项指标方面未达到相应平均值。

3.4.2 区域视角下政府开放数据利用行为综合分析

由于区域视角下的开放平台数量较多，且各区域的经济、文化、资源环境等情况各有差异，故将82个政府数据开放平台分别按照省级开放平台、副省级开放平台、地市级开放平台三个层面，进行数据利用行为的聚类和对比较分析。

（1）省级开放平台——政府数据利用行为综合分析

为进一步揭示政府开放数据在省级区域视角下的利用现状，结合开放数据浏览量、下载率和整体样本开放数据平均浏览量、下载率4个指标，采用层次聚类法对各区域开放数据的利用状况做聚类分析。其中，个体距离采用平方欧式距离，类间距离采用平均组间联接，最终聚类形成的层次分析图如图3-13所示。

利用行为分析

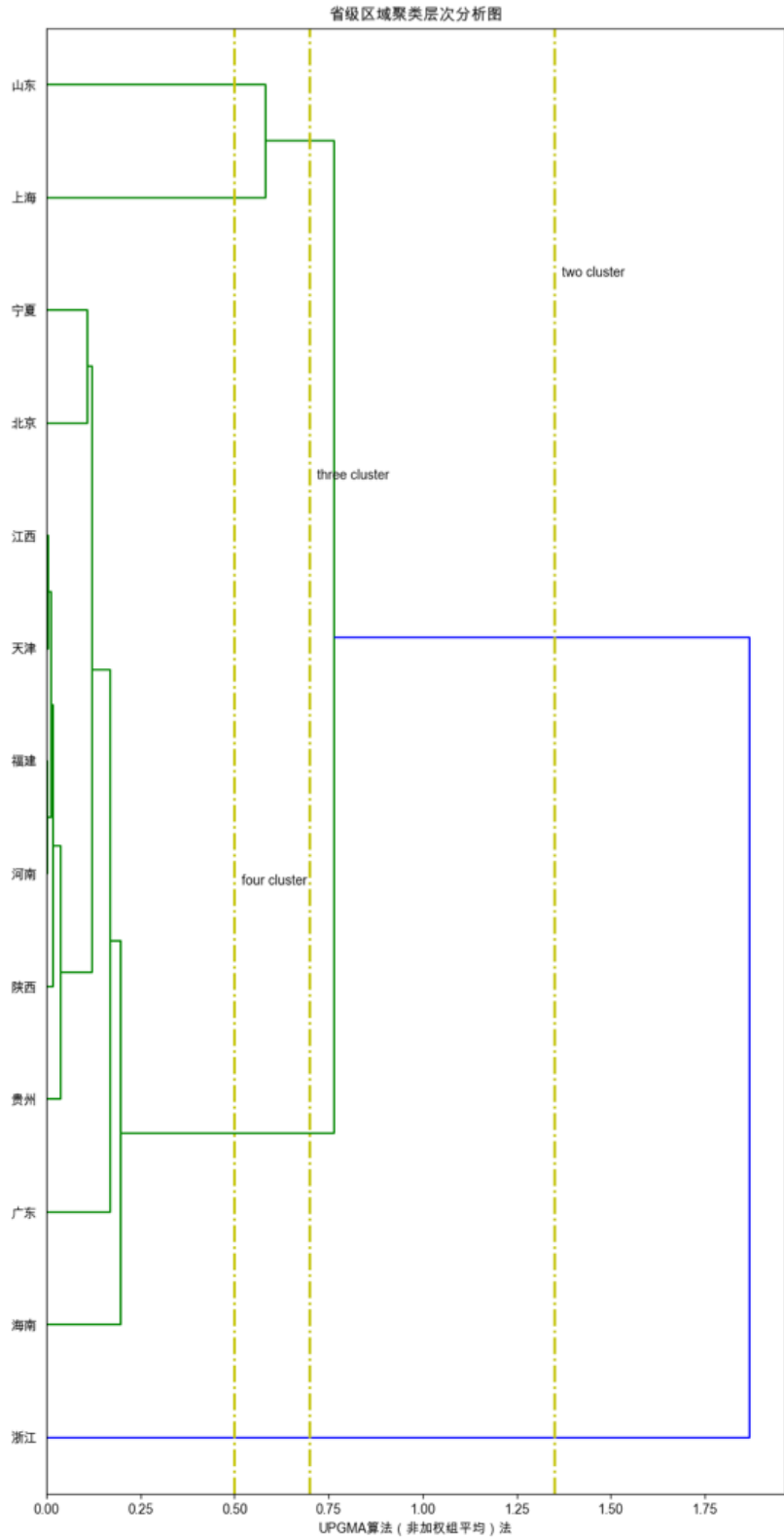


图3-13 省级开放平台—政府数据层次聚类图

为了更好地划分聚类数目，绘制了省级区域视角下开放数据聚类的系数图，如图3-14所示。由图3-14可知，随着类的不断凝聚和类目数量的不断减少，各类之间的距离迅速增大，碎石图逐渐趋于平坦。观察聚合系数图可知，当类别数为2或3时，折线的下降趋势趋缓，故将区域类别数设定为2或3类较好。根据冰柱图和树状图可知，当聚成2类时，浙江单独聚为一

类，山东、上海和北京、广东、宁夏、海南、贵州、陕西、江西、天津、福建、河南全部聚为一类；当聚成3类时，浙江单独聚为一类，山东、上海单独聚为一类，北京、广东、宁夏、海南、贵州、陕西、江西、天津、福建、河南单独聚为一类。为更细致地分析省级视角下不同地域之间的数据利用特征，本报告最终将13个区域的研究样本聚成3类。

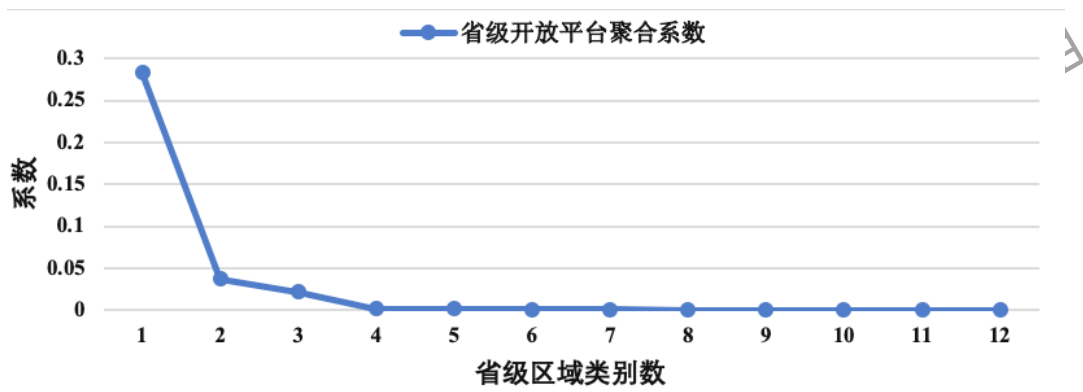


图3-14 省级开放平台—政府数据聚类系数图

第一类是浙江。该省级开放平台的数据利用现状，在四个指标的单独和综合排名最高。浙江在浏览率、下载率和整体样本浏览率、整体样本下载率这四个指标方面排名均为第1。

第二类是山东和上海。这类省级开放平台的数据利用现状，在四个指标的单独或综合排名较靠前。山东在浏览率和下载率两个指标方面排名很靠前（分别为第3和第2），但是在整体样本浏览率、整体样本下载率两个指标方面排名居中（分别为第12和第6）。上海在浏览率和下载率两个指标方面排名很靠前（分别为第2和第3），同样在整体样本浏览率和整体样本下载率两个指标方面排名也很靠前第3（分别为第3和第2）。

第三类是北京、广东、宁夏、海南和贵州、陕

西、江西、天津、福建、河南。这类省级开放平台的数据利用现状，在四个指标的单独或综合排名较居中、靠后。北京在浏览率、下载率和整体样本浏览率、整体样本下载率四个指标方面排名均居中（分别为第5、第5、第5、第4）。广东在浏览率、下载率和整体样本浏览率、整体样本下载率四个指标方面排名均居中（分别为第4、第4、第6、第5）。宁夏在四个指标的综合排名居中，其中在浏览率、整体样本浏览率两个指标方面排名居中（分别为第6和第4），但在下载率、整体样本下载率两个指标方面排名均靠后（均为第12）。海南在四个指标的综合排名居中，其中在整体样本浏览率和整体样本下载率两个指标方面排名较靠前（分别为2和第3），但浏览率、下载率两个指标方面排名均靠后（均为第10）。河南在四个指

利用行为分析

标的综合排名居中，其中在浏览率、下载率、整体样本浏览率、整体样本下载率指标方面分别排名第9、第7、第10、第8。贵州在四个指标的综合排名居中，其中在浏览率、下载率、整体样本浏览率、整体样本下载率指标方面分别排名第7、第6、第9、第7。陕西在四个指标的综合排名靠后，其中在浏览率、下载率、整体样本浏览率、整体样本下载率指标方面均排名第13。江西在四个指标的综合排名靠后，其中在浏览率、下载率、整体样本浏览率、整体样本下载率指标方面分别排名第12、第11、第7、第10。天津在四个指标的综合排名靠后，其中在浏览率、下载率、整体样本浏览率、整体样本下载率指标方面分别排名第8、第

9、第8、第11。福建在四个指标的综合排名靠后，其中在浏览率、下载率、整体样本浏览率、整体样本下载率指标方面分别排名第11、第8、第11、第9。

(2) 副省级开放平台——政府数据利用行为综合分析

为进一步揭示政府开放数据在副省级区域视角下的利用现状，结合开放数据浏览率、下载率和整体样本开放数据平均浏览率、下载率4个指标，采用层次聚类法对各区域开放数据的利用状况做聚类分析。其中，个体距离采用平方欧式距离，类间距离采用平均组间联接，最终聚类形成的层次分析图如图3-15所示。

国家社会科学基金重点项目(17ATQ006)课题组

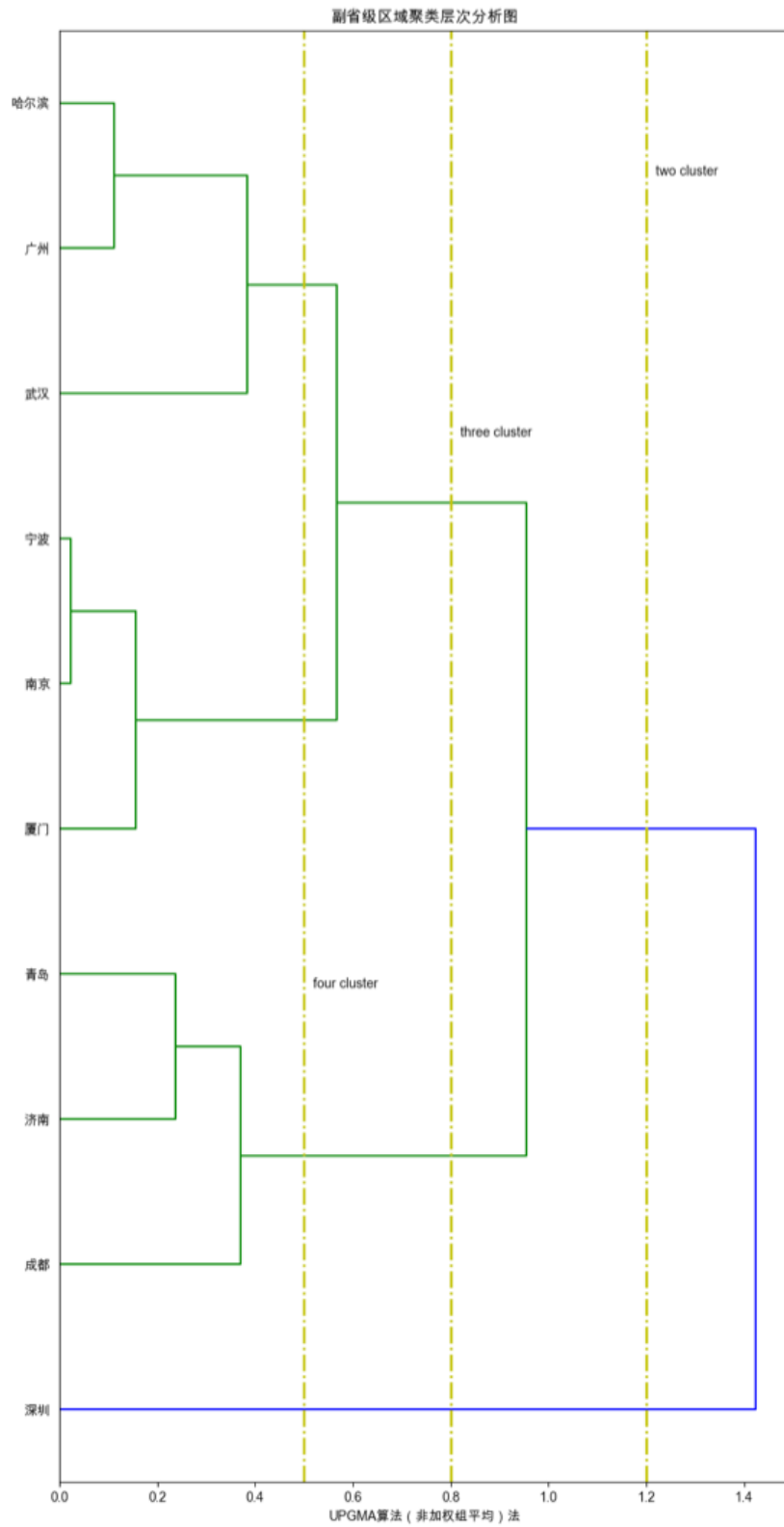


图3-15 副省级开放平台—政府数据层次聚类图

利用行为分析

为了更好地划分聚类数目，绘制了副省级区域视角下开放数据聚类的系数图，如图3-16所示。由图3-16可知，随着类的不断凝聚和类目数量的不断减少，各类之间的距离迅速增大，碎石图逐渐趋于平坦。观察聚合系数图可知，当类别数为2或3时，折线的下降趋势趋缓，故将区域类别数设定为2或3类较好。根据冰柱图和树状图可知，当聚成2类时，深圳单

独聚为一类，济南、青岛、成都和哈尔滨、武汉、南京、宁波、厦门、广州全部聚为一类；当聚成3类时，深圳单独聚为一类，济南、青岛、成都单独聚为一类，哈尔滨、武汉、南京、宁波、厦门和广州单独聚为一类。为更细致地分析副省级视角下不同地域之间的数据利用特征，本报告最终将10个区域的研究样本聚成3类。

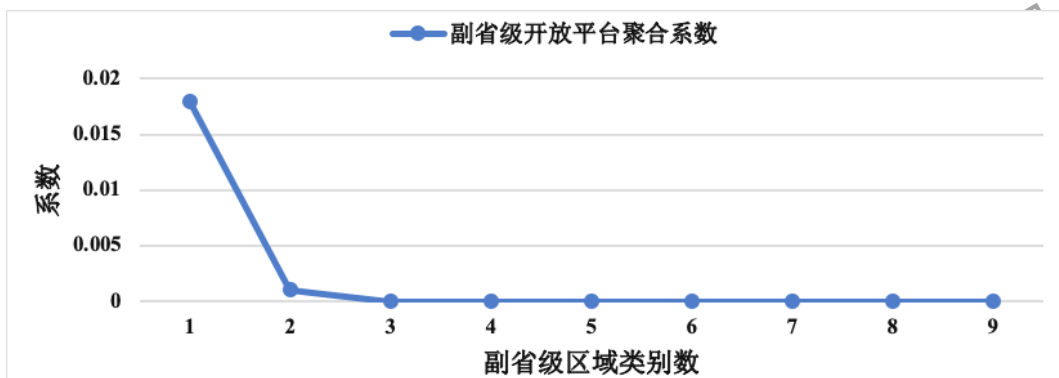


图3-16 副省级开放平台—政府数据聚类系数图

第一类是深圳。这类省级开放平台的数据利用现状，在四个指标的综合排名较靠前。深圳在浏览率和整体样本浏览率两个指标方面均排名第1，在下载率指标排名第4，在整体样本下载率指标排名第5。

第二类是济南、青岛和成都。这类副省级开放平台的数据利用现状，在四个指标的单或综合排名较居中。济南在浏览率和整体样本浏览率两个指标方面排名居中（分别为第6和第5），在下载率和整体样本下载率两个指标方面排名靠前（均为第1）。青岛在浏览率和整体样本浏览率两个指标方面排名居中（分别为第4和第6），在下载率和整体样本下载率两个指标方面排名靠前（分别为第2和第3）。成都在浏览率和整体样本浏览率两个指标方面排名靠后（分别为第8和

第7），在下载率和整体样本下载率两个指标方面排名靠前（分别为第3和第2）。

第三类是哈尔滨、武汉、南京、宁波、厦门和广州。这类副省级开放平台的数据利用现状，在四个指标的单或综合排名较靠后。哈尔滨在浏览率和整体样本浏览率两个指标方面排名靠前（均为第3），在下载率和整体样本下载率两个指标方面排名居中（分别为第7和第4）。武汉在浏览率、下载率两个指标排名居中（均为第5），在整体样本浏览率和整体样本下载率两个指标排名较靠后（分别为第8和第7）。南京在浏览率、下载率和整体样本浏览率、整体样本下载率这四个指标方面排名均靠后（均为第10）。宁波在浏览率、下载率和整体样本浏览率、整体样本下载率这

四个指标方面排名均靠后（均为第9）。厦门在浏览率、下载率和整体样本下载率三个指标方面排名均靠后（分别为第7、第8、第8），在整体样本浏览率指标排名居中（第4）。广州在下载率和整体样本下载率两个指标方面排名靠后（均为第6），在浏览率和整体样本浏览率两个指标方面排名靠前（均为第2）。

（3）地市级开放平台——政府数据利用行为综合分析

为进一步揭示政府开放数据在地市级区域视角下的利用现状，结合开放数据浏览率、下载率和整体样本开放数据平均浏览率、下载率4个指标，采用层次聚类法对各区域开放数据的利用状况做聚类分析。其中，个体距离采用平方欧式距离，类间距离采用平均组间联接，最终聚类形成的层次分析图如图3-17所示。

国家社会科学基金重点项目(17ATQ006)课题组

利用行为分析

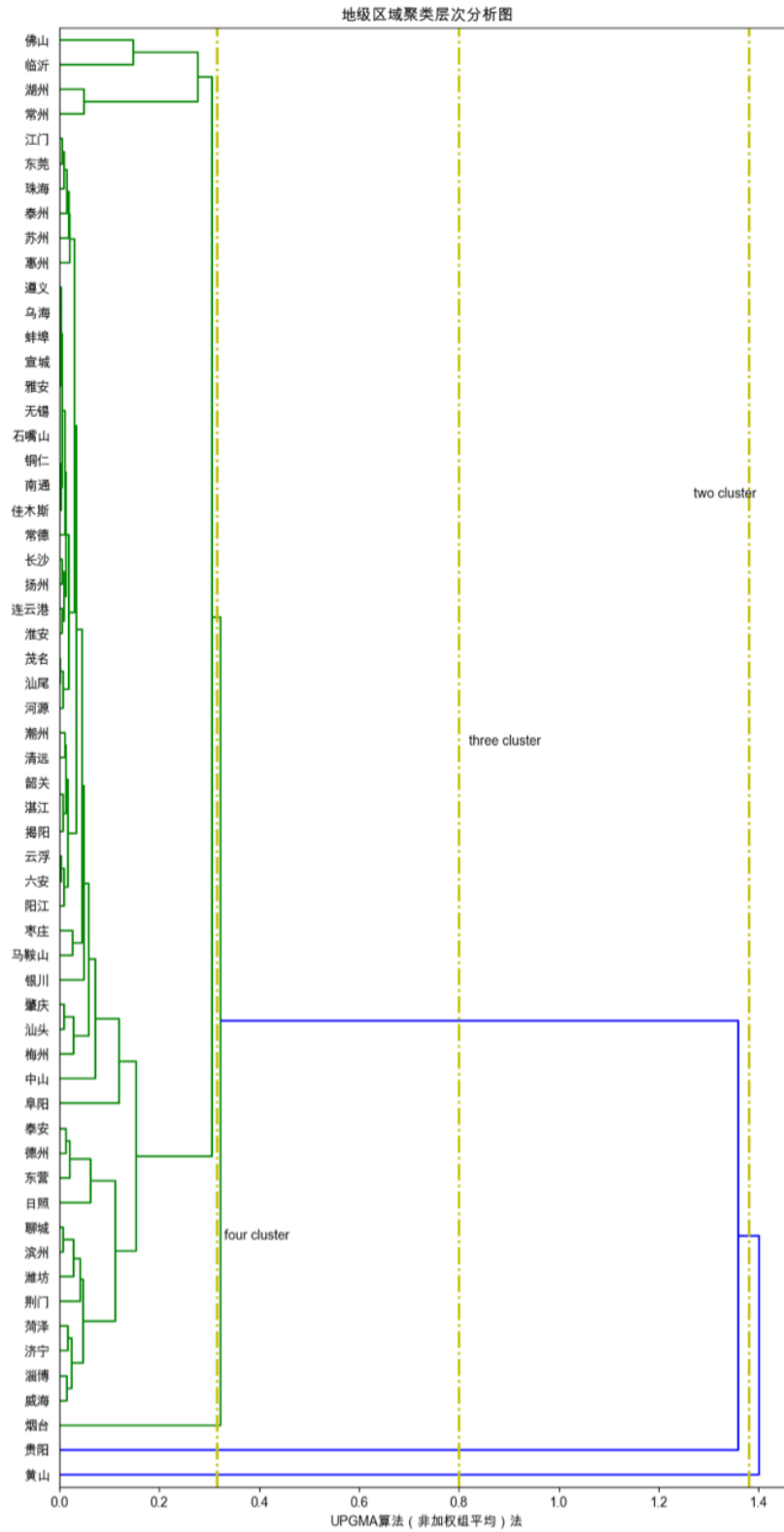


图3-17 地市级开放平台—政府数据层次聚类图

为了更好地划分聚类数目，绘制了地市级区域视角下开放数据聚类的系数图，如图3-18所示。由图3-18可知，随着类的不断凝聚和类目数量的不断减少，各类之间的距离迅速增大，碎石图逐渐趋于平坦。观察聚合系数图可知，当类别数为2或3时，折线的下降趋势趋缓，故将区域类别数设定为2或3类较好。根据冰柱图和树状图可知，当聚成2类时，除了贵阳、黄山以外的平台全部聚为一类；当聚成3类时，贵

阳和黄山单独聚为一类，泰安、德州、东营、日照、聊城、滨州、潍坊、荆门、菏泽、济宁、淄博、威海和烟台单独聚为一类，中山、梅州、汕尾、肇庆、银川等剩余的44个地区聚为一类。由于全样本数量较多，为更细致地分析地市级视角下不同地域之间的数据利用特征，本报告最终将59个区域的研究样本聚成3类。

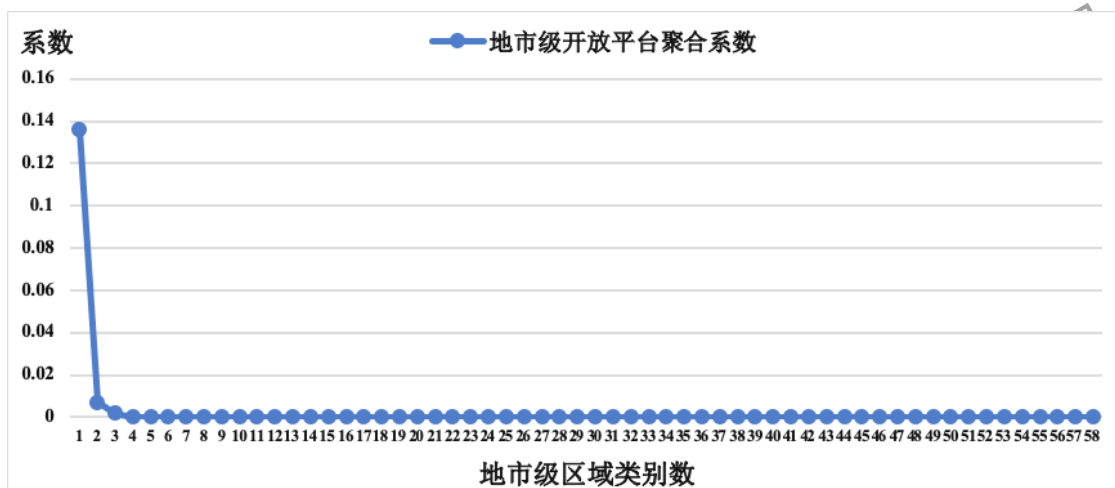


图3-18 地市级开放平台—政府数据聚类系数图

第一类是贵阳和黄山。这类地市级开放平台的数据利用现状，在四个指标的综合排名较靠前。贵阳在浏览率、下载率两个指标方面排名均为第1，在整体样本浏览率、整体样本下载率两个指标方面排名均靠前（分别为第5和第2）。黄山在整体样本浏览率、整体样本下载率两个指标方面排名均为第1，在浏览率、下载率两个指标方面排名均靠前（分别为第6和第10）。

第二类是泰安、德州、东营、日照、聊城、滨州、潍坊、荆门、菏泽、济宁、淄博、威海和烟台。这类地市级开放平台的数据利用现状，在四个指标的单独或综合排名较居中，且基本全部为山东省的地市级开放平台。其中，泰安在下载率、整体样本下载率

两个指标方面排名均靠前（分别为第6和第5），在浏览率、整体样本浏览率两个指标方面排名居中（分别为第19和第41）。德州在浏览率、下载率、整体样本下载率三个指标方面排名均靠前（分别为第13、第7和第16），在整体样本浏览率指标方面排名居中（为第43）。东营在下载率、整体样本下载率两个指标方面排名均靠前（分别为第5和第12），在浏览率、整体样本浏览率两个指标方面排名居中（分别为第21和第45）。

第三类是中山、梅州、汕尾、肇庆、银川等剩余的44个地区。这类地市级开放平台的数据利用现状，在四个指标的单独或综合排名较靠后。其中，汕尾在

利用行为分析

浏览率、下载率、整体样本下载率三个指标方面排名均靠后（分别为第47、第48和第39），在整体样本浏览率指标排名居中（第21）。肇庆在浏览率、整体样本浏览率两个指标方面排名均靠后（分别为第42和第38），在下载率指标排名居中（第18）。汕尾在下载率、整体样本下载率两个指标方面排名均靠后（分别为第34和第41），在整体样本浏览率指标排名居中（第25）。

3.5 对比2018年报告的变化分析

3.5.1 主题视角下的变化分析

《2018年度中国政府开放数据利用报告》中为10个主题大类，本报告新增了城乡建设、财税金融、信用服务、农业农村4个主题大类，共计14个主题大类。

对比《2018年度中国政府开放数据利用报告》和本报告中共有的10个主题大类发现：公众对教育科技、民生服务、经济工商等与社会民生领域相关的主题利用程度较高，这几类数据由于与公众日常生活联系密切，受到的关注度较高。而社保就业、能源环境、公共安全等主题虽然同样与用户的生活息息相关，但其利用状况并不理想，这与开放数据集的数量、用户需求等因素有关。此外，主题视角下开放数据的整体特征与部分特征具有不一致性，因此在改善开放数据整体利用状况的同时，还应注意其内部数据的增值。

而本报告新增的4个主题大类相较于去年的10个主题的整体利用情况较低，作为新近开放的主题数据，这与数据的时间累积性也有一定关联；且4个主题大类中，相较于城乡建设、农业农村这两个主题，公众

对信用服务、财税金融等领域的数据目前关注程度较低，这与用户意识、需求的紧急性等因素有关。

3.5.2 区域视角下的变化分析

《2018年度中国政府开放数据利用报告》中为6个区域样本，本报告为82个区域样本，且划分了行政级别进行聚类分析。

对比《2018年度中国政府开放数据利用报告》和本报告中共有的6个区域样本发现：①省级开放平台——上海。上海在2018年报告中的数据浏览率和下载率均居首位；在本报告中其浏览率、下载率和整体样本浏览率、整体样本下载率四个指标排名依旧靠前（分别为第2、第3和第3、第2），这与上海通过政府引导、提高数据质量、重视用户参与和数据创新等方式促进政府数据利用等行为密不可分。②副省级开放平台——哈尔滨、广州。对比2018年报告，哈尔滨和广州在数据利用情况方面这两个城市依旧聚为一类，两地具有较高的相似性；此外，两者之间的相似性还体现在其数据开放虽然起步较晚，但其开放数据的关注度和被利用情况尚可。③副省级开放平台——济南、武汉。2018年报告中，济南和武汉之间相似性较高，差距并不明显，聚为一类。而本报告中，济南在数据利用的综合排名中已经超过武汉。其中，济南在下载率和整体样本下载率两个指标方面排名靠前（均为第1），在浏览率和整体样本浏览率两个指标方面排名居中（分别为第6和第5）；武汉在浏览率、下载率两个指标排名居中（均为第5），但在整体样本浏览率和整体样本下载率两个指标排名较靠后（分别为第8和第7）。

4 政府开放数据利用效益分析

本次政府开放数据利用效益研究框架是在《2018年度中国政府开放数据利用报告》基础上，提出修改并增加数据容量、种类、开放授权、元数据完整性、发布与更新时间等指标，对研究框架进行完善。本报告利用爬虫工具及人工观察的方式获取包括数据集个数、数据容量、种类等指标的相关数据，从数据集数量、数据标准、数据持续性和数据利用性这四个维度对68个政府开放数据平台进行统计分析（平台选择见附录）。

4.1 数据数量

4.1.1 数据集个数

数据集个数指各平台每一主题类目下开放的数据集数量，不包含数据接口与数据应用数。数据集个数的统计是从平台上通过下载数据方式获取得到的数据集总和。数据集个数最多的十个政府开放数据平台如图4-1所示。目前共有29个政府开放数据平台的数据集个数超过1000个（即43%）。从各平台的数据集个数中位数来看，当前我国有50%的政府开放数据平台所开放的数据集个数低于548个。

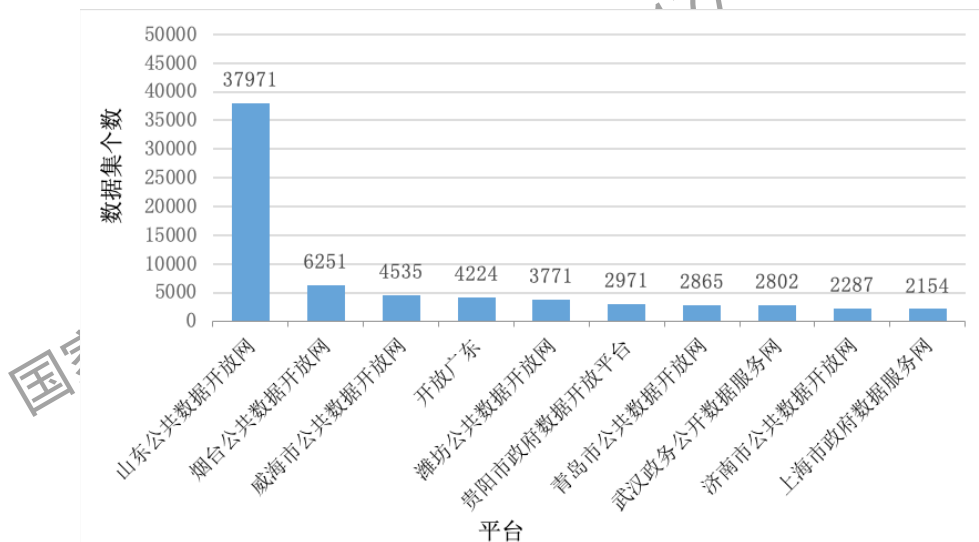


图4-1 数据集个数统计（前十名）

山东公共数据开放网所开放的数据集个数远大于其余平台，其所涉及的开放主题有20个，如图4-2所示。该平台涉及的主题较为广泛，但各个主题开放的数据集个数各不相同，社会保障类主题所开放的数据集个数远远超过其余19类主题，占据了24%的数据，依次为统计、信用、安全监管，这些主题都包含了2000

个以上的数据集，而气象类主题所开放的数据集个数最少，排在了末尾，仅包含94个数据集。通过查看气象类主题的相关数据，发现气象类主题的数据集有部分热度较高，具有较高的浏览量和下载量，而60%的数据不受公众欢迎，下载量为0。

利用效益分析

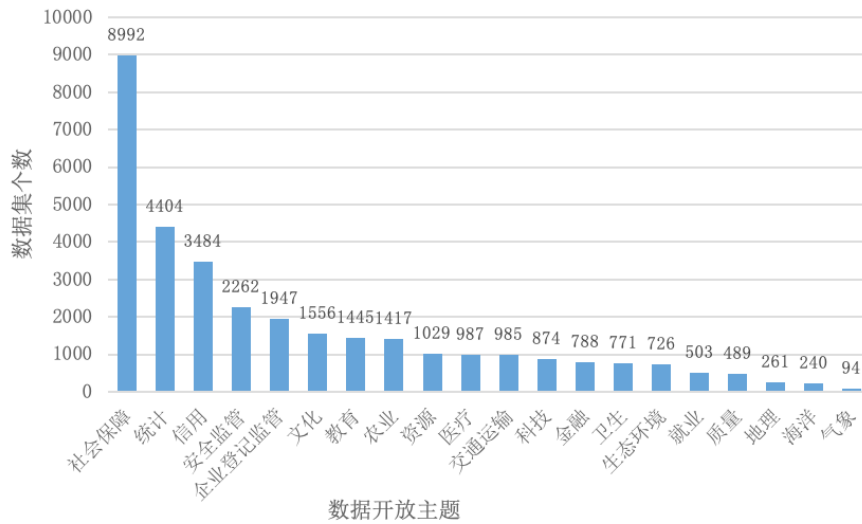


图4-2 山东省各主题下数据集个数统计情况

4.1.2 数据容量

数据容量指将每条数据集的条数（行数，即数据量）乘字段数（列数）并求和所得的数据总量，主要用来衡量各平台所提供数据集的实际数据量大小。目前，各政府开放数据平台间的数据容量差异较大，如

图4-3所示。在统计的68个政府开放数据平台中，有24个平台未提供数据项或数据量/资源记录条数，而扬州显示数据量，但所有数据集的数据量均显示为0条。在提供数据容量的这些城市中有20个平台提供的数据容量超过一亿，而有22个平台提供的数据容量低于6821.1万。

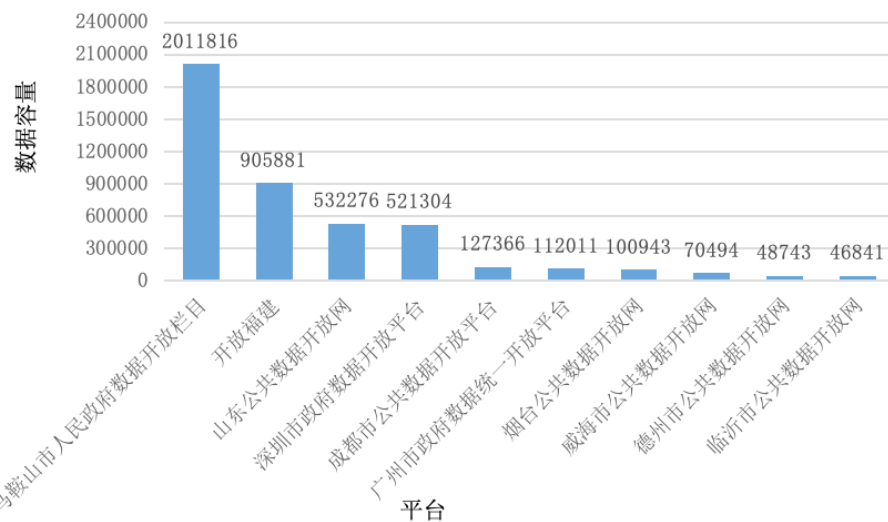


图4-3 可下载数据容量（前十名）单位：万

4.1.3 数据主题种类

政府开放数据的数据主题可以反映数据的内容范围和覆盖面，与用户的需求息息相关，主题数量的多少也可以反映该平台所开放数据的规模。图4-4展示了数据主题种类数前十的平台，从主题分类总数上来看，目前佛山、浙江、福建、青岛、厦门等政府开放

数据平台开放的主题类型相对丰富，均为20种以上，而有52.94%左右的平台所提供的数据主题种类不足18种。其中，佛山市政府的开放数据涵盖城市建设、道路交通、社会保障、服务、安全等25个领域，浙江省政府的开放数据包括生态环境、教育文化、市场监督管理、城建住房、交通运输、医疗卫生、财税金融等22个领域。

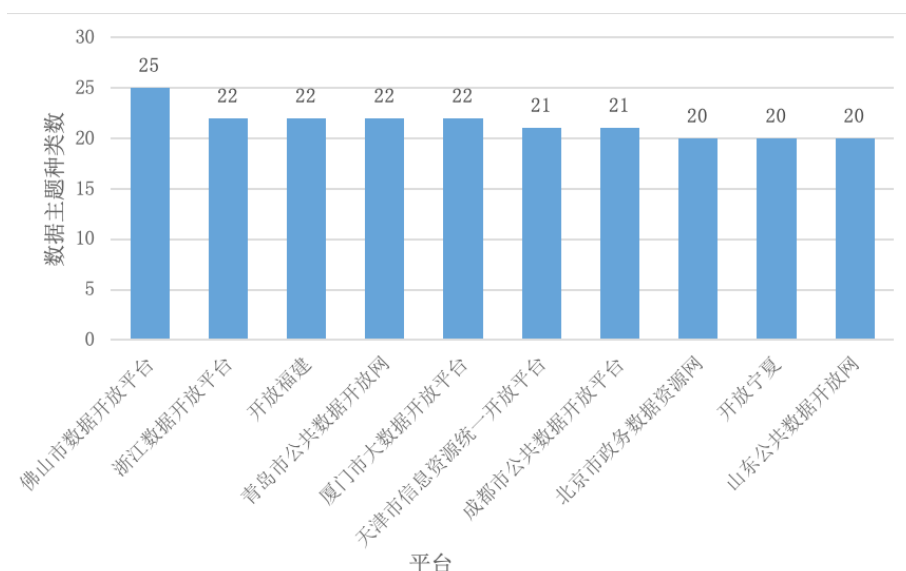


图4-4 数据主题种类数量（前十名）

从图4-5中可以看出，在佛山市政府开放数据平台上所开放的数据集主要以民生服务（10.13%）、公共安全（9.47%）类居多，而在社会保障类最少，仅为0.09%。从政府开放数据平台所开放的数据集个数和各

主题下开放的数据集个数来看，很多数据主题所开放的数据集个数仅为个位数，该平台上不同主题所开放的数据个数存在较大的差异，极不均衡，其资源建设迫切需要与用户需求进行紧密结合。

利用效益分析

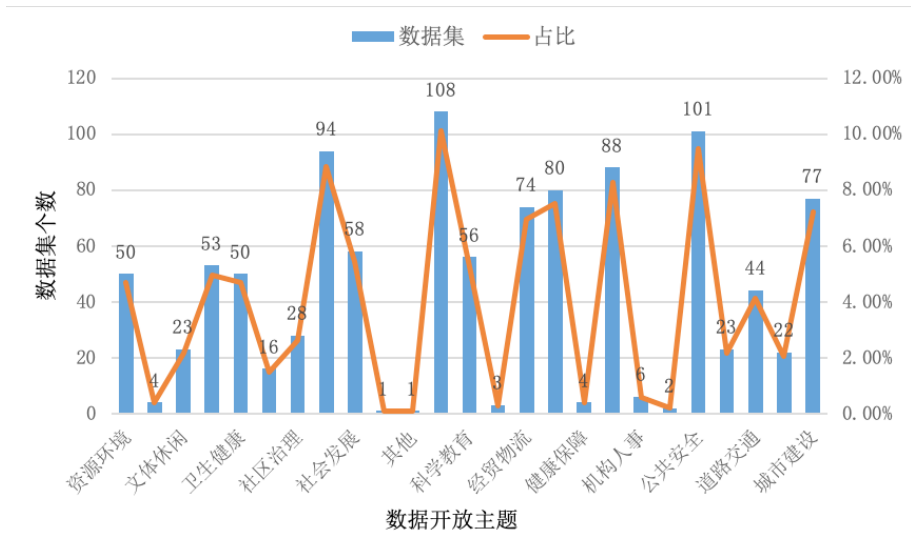


图4-5 佛山市各主题下数据集个数统计情况

4.2 数据标准

4.2.1 机器可读

关于开放数据的数据格式标准，最受国际认可的是万维网和关联数据创始人Tim Berners-Lee提出的开放数据的五星标准，其中，一星级数据指在互联网上开放授权以任意格式存在的数据，如DOC、PDF等格式的数据；二星级数据指可机读的数据，如XLSX、XLS格式的数据；三星级数据是指以非专属格式开放的数据，如CSV格式的数据；四星级数据指的是在三星级数据基础上利用W3C开放标准呈现的数据，如可以提供相应接口的数据；五星级数据则是在以上开放数据标准的基础上，提供数据发现服务。

数据格式是其进行评价的客观依据，直接影响

到数据利用的难易程度和效果好坏，本报告基于Tim Berners-Lee提出的开放数据五星标准，对68个政府开放数据平台上的数据格式进行统计分析，统计结果如图4-6所示。从统计的数据格式来看，主要涉及XLS、CVS、JSON、XML、DOC、DOCX、PDF、RAR、ZIP等数据格式，其中XLS、CSV、JSON、XML的平台覆盖率均为60%以上。XLS、XLSX等office办公软件由于存在使用习惯和兼容性等多方面的优势，故大部分的数据集都会提供两种或以上类型的数据资源格式。同时，CSV为自动化处理的结构化数据，故大多数政府开放数据平台都会提供该类数据资源格式。其中，广东、湖州、深圳、贵阳还广泛提供了RDF格式，RDF是为数据集设置固定的URL链接，有助于用户发现和链接到目标数据集的具体位置。

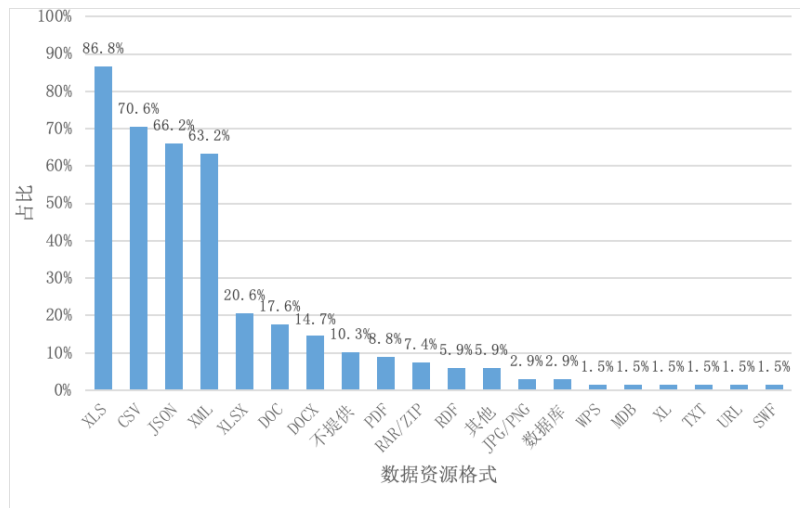


图4-6 数据集格式统计

数据若开放，则应是机器可读的格式。机器可读是用来衡量数据是否方便用户获取和利用、是否能被计算机自动抓取和处理，如XLS、CSV、TXT、XML、JSON、XLSX、WMS、RFT、XML、RDF等格式，而非DOC、PDF、JPG为机器不可识别读取的格式。68个政府开放数据平台提供的可机读格式比例如图4-7所示。其中，江西、南京、南通、苏州、无锡、雅安、阳江的政府开放数据平台不提供数据资源格式，而提供数据格式的网站的都提供了机器可读格式（如XLS、JSON、XML、CSV等），机器可读格式的平台覆盖率为100%。

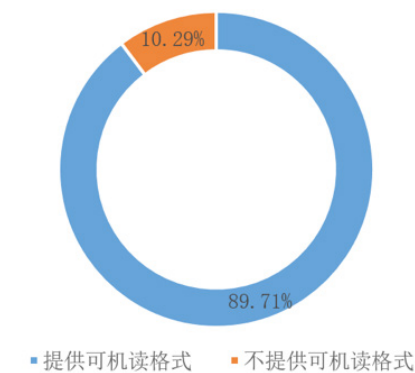


图4-7 各地提供机器可读格式比例

4.2.2 开放格式

开放格式是指可下载数据集应以开放的、非专属的格式提供，任何实体不得在格式上排除他人使用数据的权利，以确保数据无需通过某个特定（特别是收费的）软件或应用程序才能访问。如JSON、CSV、XML、TXT、KML、WMS、RTF等，而XLS、XLSX、DOC、DOCX、PDF、ODT不属于开放格式。68个政府开放数据平台提供的可机读格式比例如图4-8所示，其中，有19.12%的开放平台上数据集提供的是非开放格式，如XLS、XLSX、DOC、DOCX、PDF、JPG等格式，而有10.29%的网站不提供任何关于数据的格式信息。

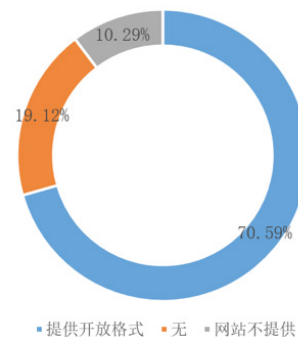


图4-8 各地提供开放格式比例

利用效益分析

经过统计，各平台开放的数据，满足开放格式的数据平均比例为76.20%。目前浙江、广州、六安、阜阳、马鞍山、东营、日照、枣庄、泰安、淄博、烟台、菏泽、菏泽、东莞、中山、淮安、连云港、湖州等政府开放数据平台100%提供了开放格式，其中湖州市同时提供了多种格式，如XLS、RDF、JSON、XML、CSV，以便更好地满足不同需求的数据使用者获取和利用数据。

4.2.3 数据接口

API接口指各平台向公众开放数据的API（应用程序接口）的情况，是否支持以数据调用接口的方式提供数据，满足部分用户开发应用程序的需求。除了提供数据下载服务外，大部分政府开放数据平台还提供了API接口（Application Programming Interface）。与通过下载数据方式获取数据集相比，API接口的数据使用者可通过参数有选择性地获取所需的数据用于特定的使用目的，以实时高效地满足不同场景下的数据需要。各地是否开放API接口情况如图4-9所示。其中有91.18%的平台开放了API接口，而8.82%的平台未开放API接口，其中包括荆门、陕西、石嘴山、阳江、常德、雅安等政府开放数据平台。

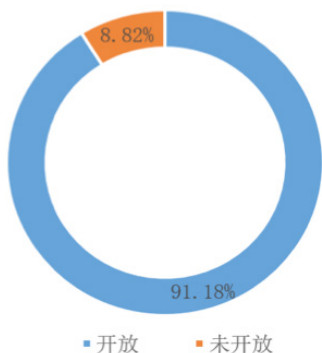


图4-9 是否开放API接口

4.2.4 开放授权

数据若开放，则其开放授权协议应明确授予数据开放主体和数据使用者自由免费访问、获取、非歧视获取、自由利用、自由传播与分享数据的权利。目前，各地的数据开放授权协议通常包含在政府开放数据平台中的服务/使用协议、网站/免责声明或服务条款中。本报告调查各数据门户上涉及到开放授权的内容，如图4-10所示。其中有73.53%的平台明确了数据开放的权利，而26.47%的平台上未能提供明确的开放数据授权协议，无法确保用户对于平台的相关权利，不符合政府开放数据的原则与要求。

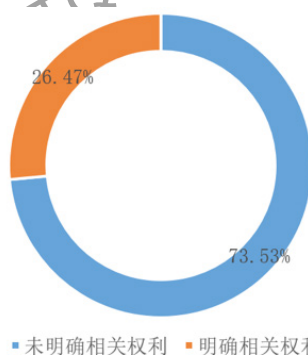


图4-10 开放授权统计情况

(1) 免费获取

免费获取是指用户获取和利用数据时不用支付费用，即授予用户免费获取和利用开放数据的权利，除非需要对数据进行额外的增值加工和针对少数人的个性化加工等。目前，对于免费获取，各地平台上的相应条款主要归为两类：1) 不限时免费；2) 现阶段免费，以后是否免费还不确定。从图4-11可以看出，仅有13.24%的平台用户保障了该权利，即为不限时免费，而48.53%的平台为现阶段免费。

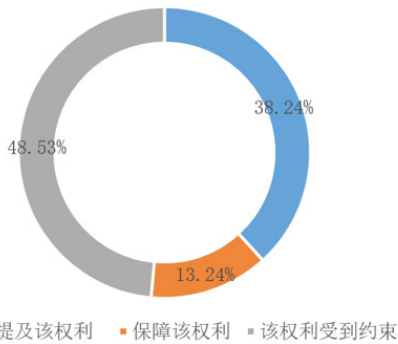


图4-11 “免费获取”授权统计情况

（2）非歧视性

非歧视性是指数据若开放，则其对任何人都平等地开放，即授予各类用户平等访问、获取、使用和分享“开放数据”的权利。在各网站条款中的文字描述主要有“用户享有数据资源的非排他使用权”、“不受歧视”等，从图4-12中可以看出，超过一半的平台保障了该项权利。

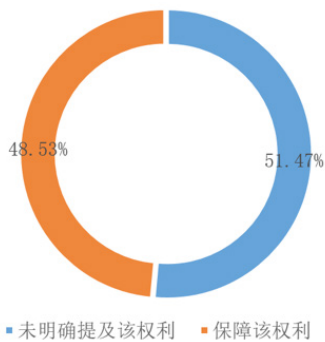


图4-12 “非歧视性”授权统计情况

（3）自由利用

自由利用是指用户可以根据自己的需要，不受限制地对数据进行商业或非商业的开发与利用。目前，各网站相应条款中对于该项规定主要分为“明确允许自由利用”和“未提及自由利用”两类，各网站条款中的描述为对数据“享有增值利用的权利”、“进行商业或非商业性的开发活动”、“可自由利用”等

不受限制的进行商业、和非商业性利用的权利。从图4-13中可以看出，仅有25%的平台保障了该项权利。

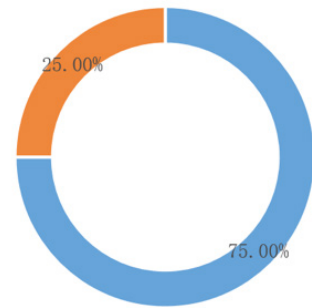


图4-13 “自由利用”授权统计情况

（4）自由传播与分享

自由传播是指用户可以自由传播和分享数据的权利，意味着不限制用户无偿转让和分享数据。目前各地平台上的相应条款分为“可自由传播”、“未提及自由传播与分享”、“自由传播受限”以及“禁止自由传播与分享”四类。“可自由传播”是指平台上的数据可自由传播与分享，不受任何限制；“未提及自由传播与分享”是平台上找不到相关声明或声明/条款中未明确提及该项权利；“自由传播受限”是平台给予用户可以自由传播与分享的权利，但用户不得有权或有偿转让在网站获取的各种数据资源；“禁止自由传播与分享”平台上未提及可自由传播与分享等字样，并规定“用户不得有偿或无偿转让在网站中获取的各种数据资源”。如图4-14所示，大部分的平台都未提及可自由传播与分享，并都规定了“用户不得有偿或无偿转让在网站中获取的各种数据资源”的权利，刺激了公众对政府数据资源的需求，不利于公众对政府数据资源的利用和产生价值。

利用效益分析

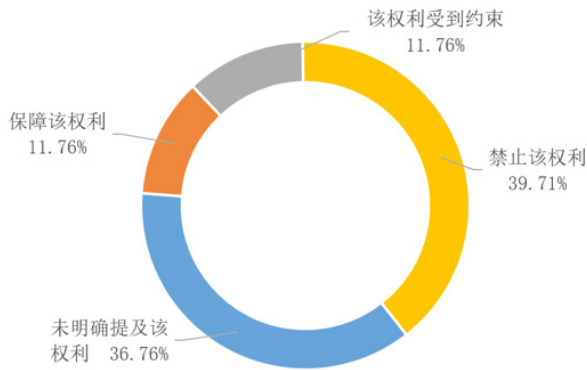


图4-14 “自由传播与分享”授权统计情况

4.2.5 元数据完整性

在平台上为每个数据集提供元数据描述，有助于用户了解数据集的内容和相关信息，对该数据集有个初步的判断，确定该数据是否符合自己的需求，从而

更好地利用数据。通过梳理国际开放数据平台上所提供的基本元数据条目，我国政务信息资源编目指南中关于核心元数据的说明以及68个平台上提供的元数据条目，本报告确定了以下十四项内容作为开放数据基本的元数据条目，包括标题、描述/摘要、关键字/标签、数据主题/领域、数据提供方、数据格式、更新频率、下载量、访问量、发布日期、更新日期、数据量、数据项、用户评分/评论。各平台提供数据基本元数据条目情况如图4-15所示，从图中可以看出，各平台对于元数据条目的提供尚未有统一的标准，出现有些平台有的元数据条目，而在其他平台中并未出现的情况，如标题在68个平台中都出现，而数据量条目仅在46个平台上出现，这些元数据条目的平台覆盖率为67.65%。

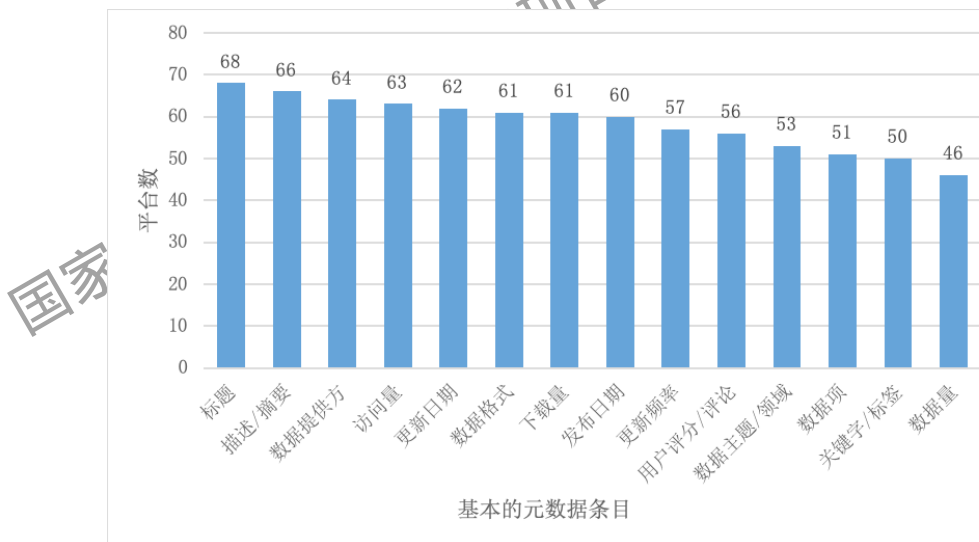


图4-15 各元数据条目在各平台的分布数量统计情况

4.3 数据持续性

数据持续性用来衡量各平台数据集自创建以来是否进行数据更新，主要以数据的发布时间、更新时间和更新类型来衡量。

4.3.1 发布时间

衡量政府开放数据平台上数据发布的持续性，主要从每月发布与更新的数据量即数据更新的视角来进行分析。本报告中将数据发布时间分为一周内、一

个月内、三个月内、半年内、一年内、一年前和不提供这七种类型（时间类型划分依据以平台最近发布数据集的时间与数据采集的时间距离），如图4-16所示。从图中可以看出，有29.41%平台的数据集最近发布时间是一周内，而有11.76%的平台不提供数据发布时间。现阶段，我国政府开放数据平台的数据开放工作尚未形成数据中后期的良好运行和高效的管理，发布时间不够明确，应加强发布数据以及发布时效性管理。

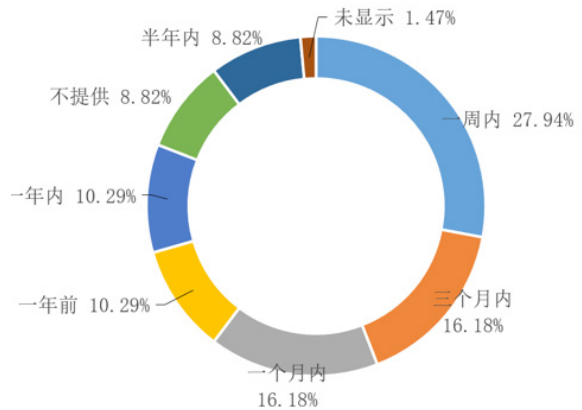


图4-17 各平台数据集更新时间统计情况

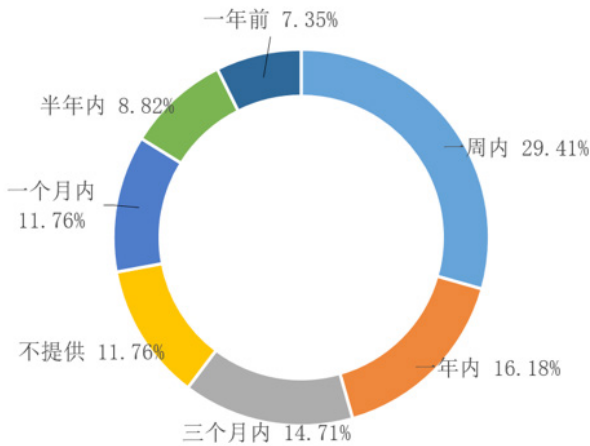


图4-16 各平台数据发布时间统计情况

4.3.2 更新时间

数据更新时间与数据发布时间相类似，也分为一周内、一个月内、三个月内、半年内、一年内、一年前和不提供这七种类型（时间类型划分依据以平台最近更新数据集的时间与数据采集的时间距离），如图4-17所示。其中，仅有27.94%平台数据集的最近更新时间为一周内，而有8.82%的平台不提供数据集的最近更新时间，聊城市政府开放数据平台虽提供数据更新时间，但网站上数据所显示的数据更新时间均为空。

4.3.3 更新频率

评估政府开放数据平台上数据是否“活跃”的标准之一就是数据的更新频率。本报告中将每季度更新、每月更新、每周更新、每日更新、实时更新等视为动态更新，将不更新/一次性更新、不定期更新、每年更新、每十年更新、每五年更新、每半年更新、按需更新、自定义更新等视为静态更新。动态更新的数据有利于反应事物的真实现状，从中挖掘出数据的价值，静态更新的数据不利于揭示事物的发展情况，更不利于数据价值的挖掘。部分政府开放数据平台所开放的数据集更新不及时或长时间不更新容易导致门户网站失去活力。从用户角度来看，若长时间在网站上找不到满足自身需求的数据，便会逐渐失去对政府的信心。

通过对各地政府开放数据平台数据集进行统计分析发现，仅有28.36%的数据可以每季度、每月、每周、每日、实时更新，其余71.64%的数据更新频率为每半年、每年、不定期、自定义或一次性更新，如图4-18所示。其中北京、陕西、南京、武汉、雅安、宣城、黄山、阳江、淮安、连云港、荆门等平台均没有更新频率

利用效益分析

的描述。图4-18说明了各平台上开放的数据集的更新频率汇总，动态、静态数据量比值约为0.3959。

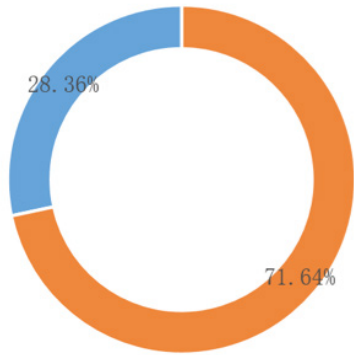


图4-18 数据更新频率动静态分布统计

同时，将各政府开放数据平台上开放的数据集的更新频率进行汇总，统计结果如图4-19所示。从图4-19中可以发现，数据集更新类型大多数为每年更新，占比为44.259%，而实时更新占比为15.332%。由图4-18和图4-19可知，当前在各地政府开放数据平台上已开放的数据多数为静态数据，数据更新不及时，严重限制了数据利用主题对各数据集的开发利用。

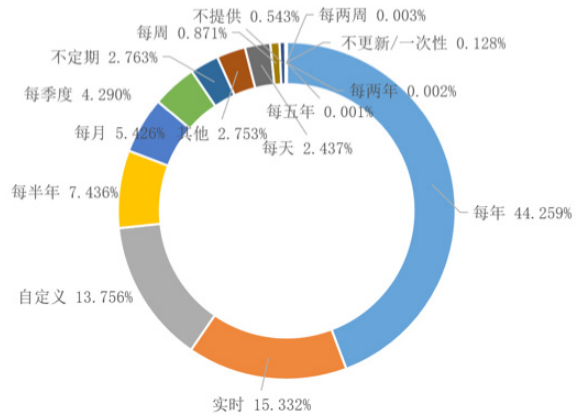


图4-19 数据更新频率分布统计

在统计中发现无锡市提供的数据集更新频率较为多样化，出现每季度、每两年、年度、年底、适时公布、有变动即更新、有变化后更新、每年更新等多种形式，其中每年又包含多种形式，将不定期/不定时/不固定，每年/年度/年底等多种同类归并后具体如图4-20所示。无锡市提供的数据集更新频率没有具体的标准，不利于促进政府数据开放共享，应事先从政府开放数据平台的角度出发，归纳总结数据集更新频率的标准，以期为自身制定政府开放数据更新频率标准提供有益参考。

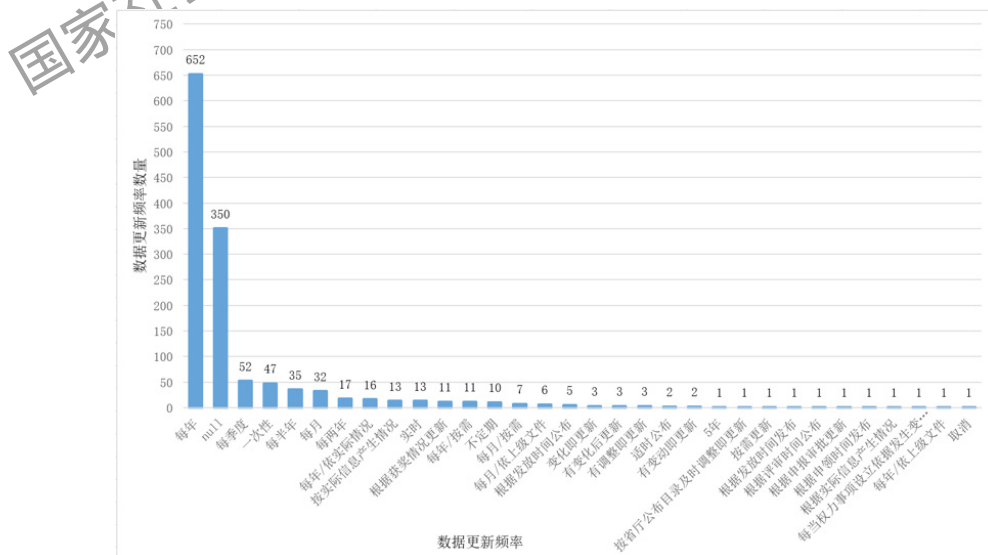


图4-20 无锡市数据开放平台数据更新频率统计

4.4 数据利用性

数据应用、数据集评价、数据的浏览量与下载量是体现用户对数据的关注度和数据价值的重要表现形式。对于数据利用性的评估，主要是从微观层面的用户视角来进行分析，当前研究大多数都从浏览量、下载量、评分量等角度来衡量开放数据的利用效果，故本报告主要是以数据应用、数据集评价、浏览量及下载量来衡量用户数据利用效果。

4.4.1 数据应用

数据应用是将政府开放数据转化为实际应用的情况，只有将这些政府数据转化为便于实际用户使用的应用，或者对数据进行进一步的挖掘分析出有用的价值，政府开放运动的意义才得以体现。数据应用主要是指数据集应用APP的数量（包括PC端应用和移动端应用）和适用于移动终端的应用等。如图4-21所示，其中广东省提供的数据应用最多，44个移动APP和37个网站应用，满足了用户的需求。而有67.64%的平台所提供的数据应用数量低于均值12，其中29.41%的平台不提供数据应用或显示提供数据应用的个数为0。

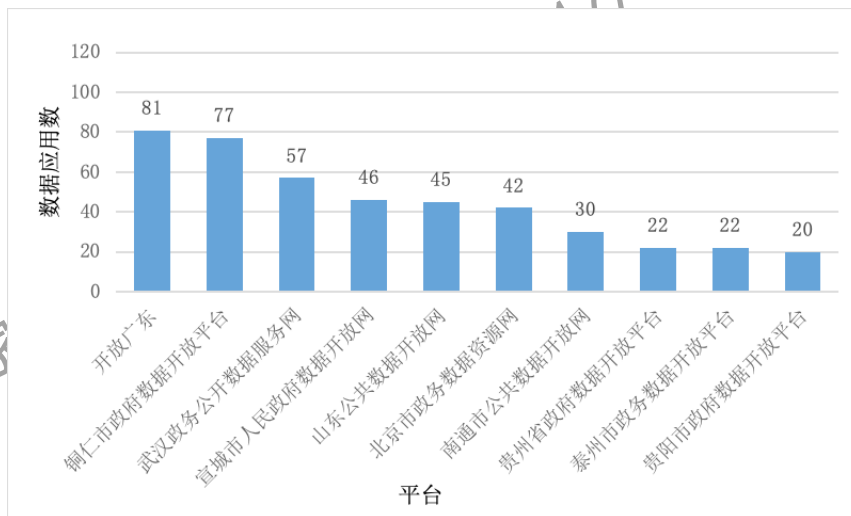


图4-21 数据应用数量统计（前十名）

4.4.2 数据集评价

用户的评价直接体现了用户对数据提供者以及数据本身的一种使用体验，是用户对数据质量等内容是否满意最直接的体现。从某种程度上反映了数据开放是否真的提高了公民参与社会治理的热情以及社会治

理的效率。从图4-22中可以看出，有82.35%的平台提供了数据集评价，而17.65%的平台不提供数据集评价。其中常州、广东、贵阳、湖州、江西、宁波、上海、深圳、浙江、天津、铜仁等平台均提供了数据集评分和数据集评论等多种评价方式，其中上海的打分选项更为详细，分为准确性、及时性、满意度、可用

利用效益分析

性四个维度，这种方式更易于看出用户对于该数据集的满意程度。而有部分平台虽然供了数据集评价，但全平台上的评分为0，如扬州。

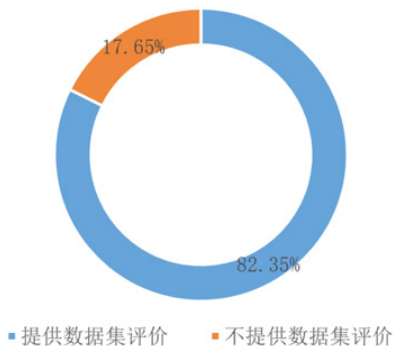


图4-22 数据集评价统计

4.4.3 浏览量与下载量

用户的浏览量和下载量从某种程度上体现了用户对于政府数据的关注程度和用户在政府开放数据平台上的参与程度，主要是用户对数据开放后的行为表现。浏览量是指用户对平台数据集的访问情况，下载量是指用户对平台数据集的下载情况。数据的浏览量和下载量的多少会影响用户对数据的关注和利用效果，本报告中主要以浏览量、下载量这两个指标来评估当前已经开放的数据集的利用情况。各地提供数据浏览与下载次数情况统计如图4-23所示，在68个平台中，并不是100%提供数据浏览和下载功能，仅有

91.18%的平台提供了这两项功能，8.82%平台仅提供数据下载或数据浏览功能，如陕西仅提供数据下载量，石嘴山仅提供数据浏览量，而南京、阳江、铜仁、无锡这四个平台则均不提供数据浏览和下载功能。

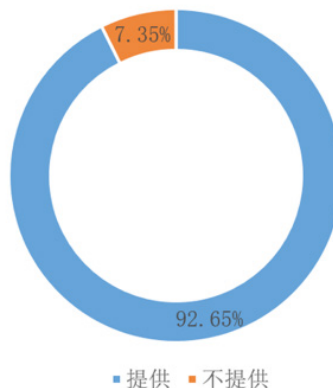


图4-23 各平台提供数据浏览与下载次数统计情况
(1) 数据平均下载量

数据集的下载量从某一方面能充分反映用户的数据需求、偏好、数据集质量以及平台的活跃程度。用户只有真正下载原始数据并加以利用，各平台开放数据的价值才能得到体现。由于各网站的数据开放时间不一致，开放的数据集个数间也存在较大的差异，单单比较数据集总下载量不能很好地比较出各网站的数据下载情况，故采用平均下载量，68个平台中平均下载量前十的数据集的平台如表4-1所示。总体来说，各平台数据集下载量差异较大，反映出各平台的实际使用情况间存在较大的差距。

表4-1 各平台数据集平均下载量（前十）

平台	数据集个数	数据集下载量	数据集平均下载量
黄山市人民政府数据开放栏目	31	135083	4358
浙江数据开放平台	579	2237909	3865
上海市人民政府数据服务网	2154	1485808	690
贵阳市人民政府数据开放平台	2971	1002557	337
海南省政府数据统一开放平台	23	6999	304
数据东莞	345	75094	218
荆门市人民政府数据开放	517	107353	208
常州市人民政府数据开放栏目	132	26663	202
贵州省政府数据开放平台	600	114141	190
济南市公共数据开放网	2287	371608	162

从统计结果上来看，上海和贵阳在数据集个数在大于1000的情况下，数据集平均下载量仍排名前五，说明这两个平台所开放的数据比较受欢迎。

(2) 数据平均浏览量

数据浏览量/访问量可以展现用户对于主题数据资源的关注程度，看出用户对于开放数据的需求，由于各平台上数据集个数存在差异，故采用平均浏览量进

行比较，如表4-2所示。从平均浏览量上来，浙江、黄山、东莞、深圳、海南、湖州、上海、宁夏、常州、阜阳这十个平台较高，其中浙江远远大于其余平台，各平台数据资源的平均浏览量参差不齐，这与各平台开放数据的数量、主题种类、价值、数据发布时间等各方面有着密切的联系。

表4-2 各平台数据集平均浏览量（前十）

平台	数据集个数	数据集浏览量	数据集平均浏览量
浙江数据开放平台	579	15507177	26783
黄山市人民政府数据开放栏目	31	311954	10063
数据东莞	345	1929354	5592
深圳市政府数据开放平台	1841	6203153	3369
海南省政府数据统一开放平台	23	69032	3001
湖州市公共数据开放网	146	436253	2988
上海市人民政府数据服务网	2154	6195209	2876
开放宁夏	166	367991	2217
常州市人民政府数据开放栏目	132	269313	2040
阜阳市人民政府数据开放	70	94430	1349

利用效益分析

浙江·数据开放提供的579个数据集中，大部分都提供了数据集浏览功能，数据集的浏览和下载次数情况如图4-24所示。数据浏览次数在10万次以上的数据集有13个，而下载次数在10万次以上的只有1个数据集，该条数据集属于“教育文化”主题，摘要内容为

“浙江省各类学校信息（学前、中小学、高校）”。41%左右的数据集浏览次数集中在10000-50000次之间，而浏览次数在100次以下的数据集涉及的数据领域主要为公共安全、信用服务和生态环境，内容主要与红十字会、机动车驾信息等相关。

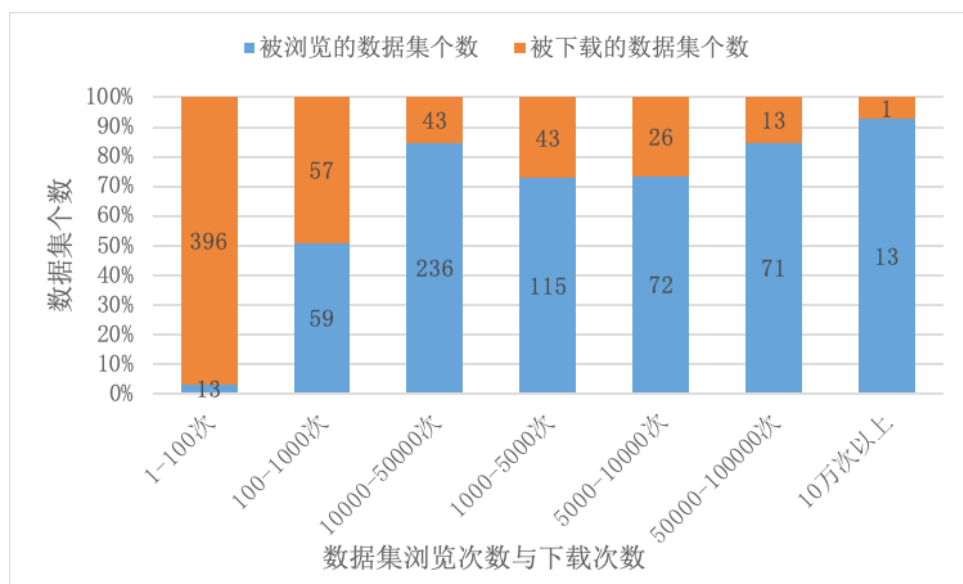


图4-24 浙江省数据集被浏览次数和下载次数统计

（3）数据下载率

用户往往会下载对自身有利用价值的数 据，为衡量各地政府开放数据平台数据集的下载利用情况，展示对数据集的整体关注程度，本报告中采用数据集的下载次数与浏览次数的比值（下载率），对有下载

次数和访问/浏览次数的各地政府开放数据平台进行综合分析。具体如图4-25所示。由于陕西、南京、石嘴山、阳江、无锡、铜仁这6个政府开放数据平台不提供数据浏览量或数据下载量，故不在图4-25中显示。

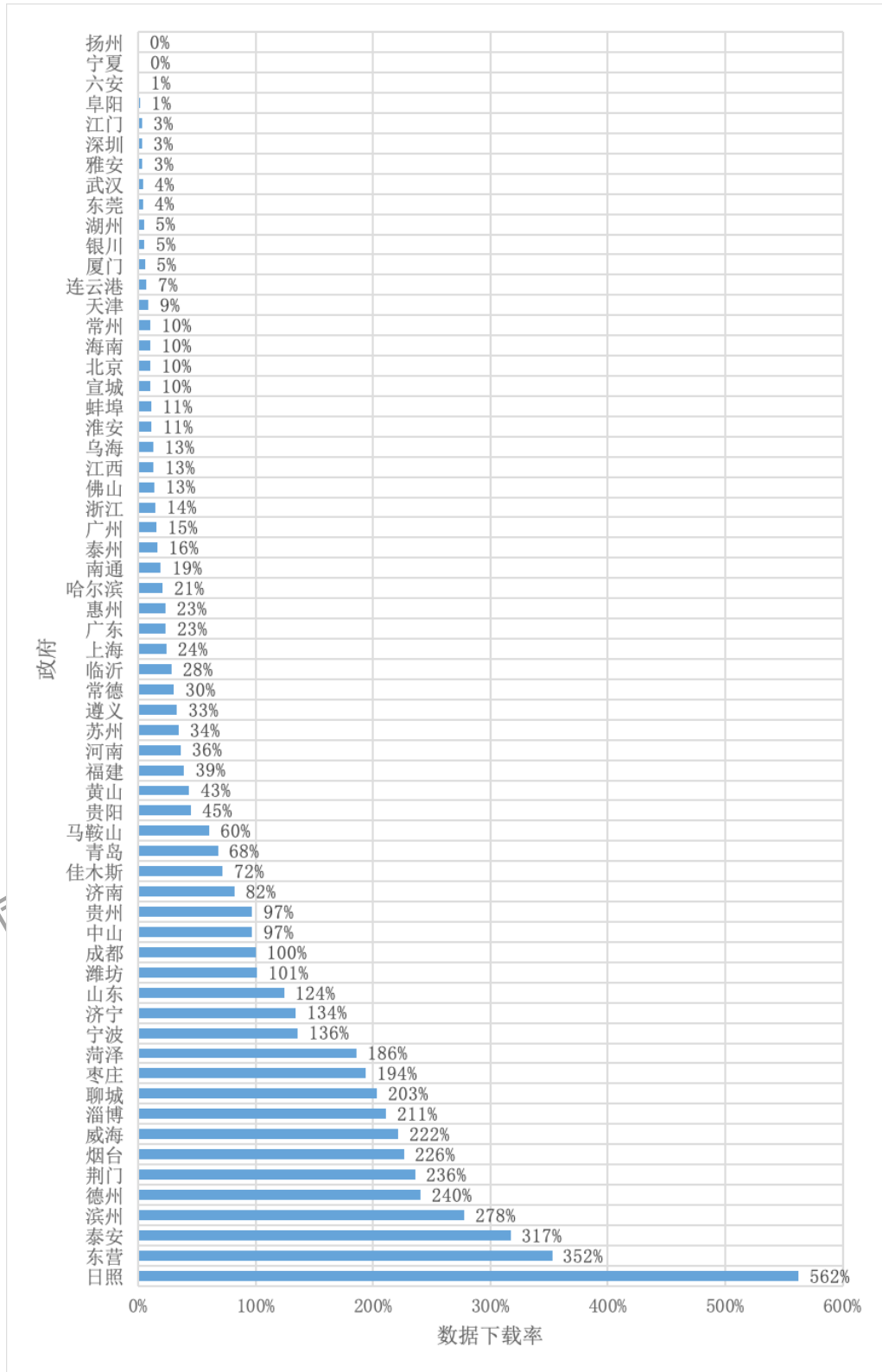


图4-25 各平台数据集下载率统计情况

利用效益分析

将各平台数据集的浏览量与下载量数据进行分析，发现大部分平台的数据集下载量小于浏览量，而日照、东营、泰安、滨州、德州、荆门、烟台等17个平台出现数据集下载量大于浏览量的情况，即“看得少，下的多”，表明这些开放的数据集满足用户的需求，从而激励用户进行获取和利用。目前，天津、连云港、厦门、银川、湖州、东莞、雅安、武汉、雅

安、深圳、江门、阜阳、宁夏、扬州等多个平台的数据集下载量均远低于数据浏览量，其中有39个平台的数据下载率低于50%（含未提供数据浏览量或数据下载量的平台）。很多用户查看了相关的数据后却没有下载，说明用户对于数据的利用并没有表现出很高的积极性，开放的数据也大多没有发挥到应有的价值，数据整体利用效果不佳。

国家自然科学基金重点项目(17ATQ006)课题组

5 政府开放数据利用案例

本节将通过交通、教育、营商环境三个层面的利用案例，介绍2019年政府开放数据的利用现状，以期为促进政府开放数据的深入开发与利用提供借鉴与参考。政府开放数据范围已经涉及公共安全、教育科技、民生服务、医疗健康、交通出行、经济工商、文体休闲等各个方面，而其中对公众影响最大、受关注程度最高的为交通、教育、营商环境三个方面。政府已逐步完善政府开放数据相关政策，为未来数据时代我国社会发展提前布局，推进政府开放数据的整合共享与高效利用，不断挖掘政府开放数据中蕴含的巨大价值。

5.1 交通层面

深化大数据在城市交通治理中的应用，向社会开放共享政府部门所拥有的交通出行数据是必不可少的关键一环。2019年11月4日，北京市交通委员会印发《北京市交通出行数据开放管理办法(试行)》的通知，将地铁公交拥挤度、实时路况、停车场泊位数等交通出行数据向社会开放¹。随着大数据技术蓬勃发展，越来越多的政府部门与企业携手合作，依靠大数据技术进行交通治理将是大势所趋。

案例1：基于大数据的贵阳公交站点位置合理性分析系统

（1）案例背景²

1 北京日报.交通出行数据向社会开放[EB/OL].(2019-11-05)[2020-02-17].http://www.beijing.gov.cn/zhengce/zcjd/201911/t20191105_483740.htm

2 人民网-贵州频道.2019贵阳开放数据应用创新大赛成

2019年首次举办的贵阳开放数据应用创新大赛，以“贵阳开放数据应用创新”为主题，旨在促进贵阳市大数据人才技术技能、创新思维、实践能力和协作能力的提升，为企业技术人员以及高校学生提供一个大数据创新应用设计竞技展示的平台，实现以赛促产、以赛促学、以赛促研、以赛促用的目标。此次大赛中，专注于公交领域深耕长达7年的武汉元光科技有限公司，通过贵阳市政府数据开放平台提供的相关领域开放数据，设计基于大数据的贵阳公交站点位置合理性分析系统，助力交通决策优化，实现交通政府开放数据的价值释放。

（2）利用实例³

为解决贵阳市站牌集中布置使公交车辆集中进站、单个站点公交班次短时间过于密集、部分公交站点设置不合理造成拥堵三大交通问题，基于大数据的贵阳公交站点位置合理性分析系统利用贵阳市公交到站、离站、车辆位置基本信息及运行数据等政府开放数据，进行站点平均途经班次分析、重点线路停靠时间分析及重点线路停靠时间分析，提高行车效率、提升乘客体验、降低事故风险。目前，武汉元光科技有限公司已通过国内最大实时公交APP“车来了”在国内200余个城市建立落地实施案例，实现实时查询公交位置功能，且同步推出了服务公交企业和交通主管部门的公共交通大数据平台，实现基于当地政府开放数据的交通决策优化。

（3）结论与建议

功落下帷幕[EB/OL].(2019-11-30)[2020-02-17].<http://gz.people.com.cn/n2/2019/1130/c222152-33591593.htm>

3 贵阳市政府数据开放平台.基于大数据的贵阳公交站点位置合理性分析系统[EB/OL].(2019-12-26)[2020-02-17].<https://data.guiyang.gov.cn/city/toAchievement.htm?pageNo=1>

利用案例

通过对政府部门的交通数据进行必要的清洗、脱敏和加工处理，以政企合作的方式进行开放利用，释放政府开放交通数据的价值。政府部门应引导与促进与企业的深度合作，为公众出行引导服务提供更好选择，向公众合理选择出行时间、出行方式和出行路线等交通出行需求提供高品质、精细化服务。

5.2 教育层面

新中国成立以来，党和国家始终把教育摆在优先发展地位，陆续推出《教育信息化“十三五”规划》、《教育部机关政府信息公开实施办法》、《教育部机关及直属事业单位教育数据管理办法》、《教育信息化2.0行动计划》等政策文件，对我国政府教育数据开放工作进行了初步规范与指导。但和发达国家的丰富实践相比，我国的教育数据开放还处于初级阶段，尚未形成一定的规模，缺乏完整的开放体系¹，对于政府开放教育数据的实践应用也处于起步阶段，仍需要政府、企业进一步深化合作，更好地服务于教育决策支持。

案例2：襄阳市义务教育招生平台APP

（1）案例背景²

2019年秋季，湖北省襄阳市樊城区试点试用“襄阳市义务教育招生平台”，即樊城区小学一年级、初中七年级所有新生统一通过“襄阳市义务教育招生平台（樊城区试点）”进行招录。襄阳市义务教育招生

平台APP2019是由湖北统讯智能科技有限公司专为湖北省襄阳市地区打造的一款招生报名服务平台，提供了义务教育阶段的招生政策、校情查询、咨询帮助等多项内容，综合全方位解决传统招生模式的“报名手续繁琐”问题。

（2）利用实例

为促进教育公平公正、提高便民化服务水平，襄阳市义务教育招生平台APP2019整合襄阳市的招生政策以及相关教育资讯，基于襄阳政府开放数据平台提供的普通中小学信息、志愿填报分数线等教育数据，提供学校信息查询与入学指导等服务。本地最新教育动态信息及时更新与校园招生信息公开收费透明，使得用户可通过该平台获取较为便捷的义务教育阶段入学服务，通过企业与政府合作的形式提供多渠道、集成化的政府开放数据服务，有助于政府开放教育数据开放范围地进一步扩大与更深层次地开发利用。

（3）结论与建议

目前较多政府开放教育数据仍停留在开放学校信息、招考分数线等基本数据层面，存在较为迫切的加大公开范围的用户需要，只有更为全面的教育数据开放才能促进深度利用。而如何真正实现跨系统的教育大数据采集和共享，开发利用蕴含于政府教育数据开放中的巨大价值，则需要政府与企业携手共建、互惠互利以实现合作共赢，打破教育数据孤岛，为教育决策提供科学支撑。

5.3 营商环境层面

近年来，围绕“放管服”工作部署，各地开展了一系列营商环境实践探索，如何利用开放政府数据，构建优质的营商便利服务体系，提升企业公众在政府

1 李青,王海兰.教育数据开放研究与实践现状述评[J].中国远程教育,2019(11):48-57.

2 楚天快报.义教招生平台已上线新生这样报名[EB/OL].(2019-07-02)[2020-02-17].
http://www.xf.gov.cn/sm/shkb/201907/t20190702_1743532.shtml

办事的便利度、满意度，是政府需不断探索的
提高行政运行效率及公共服务水平的重难点问题。

案例3：深圳市统一政务服务APP“i深圳”

（1）案例背景¹

2019年1月6日，开放广东政府开放数据平台上线
深圳市统一政务服务APP“i深圳”，该APP由中国平安
研发打造，以“市民用一个APP畅享全市综合服务”为
目标，建设全市统一的移动政务服务平台，助力深圳
市优化营商环境。“i深圳”的“营商通”板块提供政
策扶持、人才扶持等多个办事主题服务，助力政府激
发企业发展，优化营商环境。

（2）利用实例²

“i深圳”将面向企业和个人提供覆盖市级、区
级、街道、社区四级政府的政务服务导航、申办，其
最大亮点为针对市民与企业的办事需求，分别设置
“市民通”和“营商通”两大特色服务，利用“营商

通”为企业提供政务办事服务和企业经营公共服务，
让数据会说话、让数据“多跑路”。“i深圳”近期将
上线300多项服务事项，年底前将实现95%以上个人事项
和60%以上法人事项移动办理，并打造“千人千面”，
提供精准推送、个性化订阅等功能，真正实现“一屏
智享生活、一号走遍深圳”，助力深圳市营商环境优
化，提高城市竞争力。

（3）结论与建议

营商环境作为城市重要的软实力，直接影响当地
经济效益，提高城市竞争力，故营商环境相关政府开
放数据力度较大，实践探索相较其他行业走在前列。
城市营商环境持续优化需借力政府开放数据，促进政
务信息公开与共享，以政务大数据为核心全面建设
“互联网+政务服务”平台，实现政府开放数据资源赋
能营商环境。

1 中国证券报.深圳市统一政务服务APP“i深圳”正式发布
[EB/OL].(2019-01-14)[2020-02-17].

[http://www.ce.cn/xwzx/gnsz/gdxw/201901/14/
t20190114_31293350.shtml](http://www.ce.cn/xwzx/gnsz/gdxw/201901/14/t20190114_31293350.shtml).

2 深圳市人民政府办公厅.深圳市统一政务服务APP“i深
圳”正式上线：市民和企业用一个APP畅享全市政务服务[EB/
OL].(2019-01-14)[2020-02-17].

[http://www.drc.sz.gov.cn/yszw/szzw/201901/
t20190114_15302614.htm](http://www.drc.sz.gov.cn/yszw/szzw/201901/
t20190114_15302614.htm)

结论与对策

6 结论与对策

6.1 现状

基于本报告前几部分的分析内容，从数据平台建设状况、可供分析数据状况和数据利用状况三个方面对政府开放数据的现状进行分析与总结。

（1）数据平台建设状况

2019年，全国政府数据开放平台新增46个。相较于2015年至2018年平台建设的快速发展，2019年政府开放数据平台的建设速度有所放缓。政府数据开放平台所呈现的不均衡特征仍较为明显。①从行政层级来看，我国省级、副省级与地市级政府的政府数据开放平台建设情况有所差异。省级、副省级城市平台超过半数，总体上上线政府开放数据平台的地方占比有所提高但地市级地区的平台建设占比依旧偏低，仍有很大的发展空间。②从平台类型上，独立平台仍占据主体地位，仅乌海、雅安、阜阳等12个地区的政府开放数据平台类型为嵌入式。③从平台所属地区来看，平台总量排名靠前的地区分别是华东、华南，和西南地区，尤其是西南地区，平台建设发展势头更为迅猛。

（2）可供分析数据状况

数据集规模整体。仅有43%的地方平台开放的数据集个数超过1000个，总体开放的数据集数量较少，且存在着重复创建、无效的问题数据等现象。单个数据集中的数据量较少，部分平台不提供数据量或数据项。在提供数据容量的这些城市中有20个平台提供的数据容量超过一亿，大部分平台提供的数据容量还是处于偏低的水平。

数据主题方面。目前各地的政府开放数据平台使用的主题分类具备一定的地域统一性，同一地区内的

平台数据建设与维护更加便利，方便实现数据的互联互通与公众利用。不同主题之间数据开放不均衡，一些主题提供的数据集个数较少。部分地方平台提供的数据种类较少，目前佛山、浙江、福建、青岛、厦门等地方平台开放的主题类型相对丰富，涵盖的领域、行业较多，均为20种以上，而有37%左右的地方平台所提供的主题种类不足15种。

数据格式。各平台对于开放的数据格式没有统一的标准，格式较为单一。数据格式以XLSX、CSV、XML、JSON居多，平台覆盖率均为60%以上，而RDF、TXT等数据格式的平台覆盖率低于6%。提供数据格式的平台中，机器可读格式实现了全平台覆盖，而开放格式的平台覆盖率仅为70.59%。

数据持续性。在数据发布与更新时间方面，地方平台与数据采集时间距离一周内发布和更新数据占比不足30%。在数据更新频率方面，数据更新不及时，静态数据占比偏高，仅有28.36%的数据可以每季度、每月、每周、每日、实时更新。此外，部分数据并未按照规定的更新日期及时更新数据，按承诺更新比例较低。

数据利用性。在数据应用方面，有29.41%的地方平台无数据应用，其他平台开发的数据应用数量比较少，开放的数据应用多为APP，数据利用的形式不够丰富。在数据集评价方面，17.65%的地方平台并未提供数据集评价。广东、贵阳、上海、深圳、浙江等地方平台均提供了数据集评分和数据集评论等多种评价方式，其中上海的打分选项更为详细。从数据集的浏览量与下载量来看，有57.35%的平台数据集的下载率低于50%，由于用户在浏览数据集后并没有产生下载的行为，导致数据的利用率较低。

（3）数据利用状况

主题视角下，公众对教育科技、民生服务、经济工商等与社会民生领域相关的主题利用程度较高，这几类数据由于与数据用户工作生活联系密切，受到的关注度较高。而社保就业、能源环境、公共安全等主题虽然同样与用户的生活息息相关，但其利用状况并不理想，这与开放数据集的数量、用户需求等因素有关。用户尤其是企业用户对经济工商类数据的关注和需求更大，因此经济工商类数据各项指标都稳居高位。交通出行与用户的生活联系紧密，而医疗健康更是全社会关注的热点。用户对信用服务、财税金融等领域的问题目前关注还不太多，与用户意识、需求的紧急性等因素有关。各主题开放数据整体与部分的特征并不具有一致性。区域视角下，不同城市开放数据的利用程度不同，呈现出明显的高低之分。各地区开放数据整体与部分的特征不具有一致性。上海、贵阳、浙江等地区开放的数据，备受用户关注。

6.2 存在问题

基于对政府开放数据现状的总结和前期的分析内容，对政府开放数据目前存在的问题进行分析，主要包括专门性政策法规缺乏、数据获取平台问题、数据质量问题、社会参与度低、数据价值挖掘不充分、数据利用生态亟需构建等问题。

（1）缺乏专门针对政府数据开放的政策法规

在政府数据开放工作实施过程中，各级政府、数据管理局、信息化中心承担着政府数据开放的相关工作，但是中央并出台专门针对政府数据开放相关的法律法规。数据开放的范围也大多由各个平台自主确定，缺乏统一的标准。国家虽出台了政府信息资源管

理标准、大数据发展行动纲要等相关政策，但专门针对政府数据开放，确定数据开放范围，数据开放标准，明晰各方面权责分工的政策仍然缺乏。

（2）数据获取平台

①平台集约化程度不高

目前，地市级平台建设增长速度较快，且以专有式平台居多，从而形成了多个政府数据开放平台相互孤立的状况，对数据的流通和共享产生了阻碍。以省级或副省级平台作为基础性平台在一定程度上即可满足政府数据开放的需求，单独的政府数据开放平台的建设、运营和维护在一定程度上造成了资源的浪费。此外，由于各方面的信息力量缺乏协调和配合，海南等地区出现了新旧平台并存以及迁移后数据丢失的情况。广东、山东、贵州、陕西均存在多个市既有独立平台又汇集在省级平台的情况。

②平台数据推荐方式过于单一

政府数据开放平台的数据推荐目前大多数是基于数据的流行度和新颖度，在平台首页提供最新数据或根据浏览量提供最热数据。这种推荐方式相对较为单一。基于用户利用行为特征的推荐方式较少，而大部分的政府开放数据平台在数据申请、下载或评论时要求用户登录，对于用户的数据检索、浏览、下载等行为有清晰的记录，为基于用户行为特征的个性化推荐提供了基本的数据支撑。

③平台功能利用情况欠佳

大部分政府数据开放平台提供了数据反馈、数据申请功能以及数据需求调研等互动功能，但大多数的功能却并未得到应用。用户对于数据基本处于用而不评的状态，在极少数的用户评论中，所能提供的有价值的信息较少。致使政府数据开放缺乏来自用户层面

结论与对策

的反馈，从而造成对用户需求的精准定位不足。大部分嵌入平台仅起到了数据汇集对作用，且以类似于新闻的方式发布数据，不利用用户的利用。

④数据分析功能不够完善

政府数据开放平台提供了简单的数据统计分析功能，但基本是对网站开放数据的整体统计，并未为用户在线处理数据检索结果提供统计分析支持和可视化呈现，而这一功能往往可以进一步降低基于政府开放利用的门槛，能够更加有效吸引用户通过网站利用政府开放数据。

（3）可供分析数据

①缺乏统一的元数据标准

纵观全国已有的政府开放数据平台，虽然部分平台在省级范围内具有统一的元数据标准，但整体而言，全国平台的元数据标准各行其是，缺乏统一化和规范性。从细节层面来看，对于同一含义的元数据指标，不同平台具有不同的命名；对于同一元数据名称，不同的平台具备不同的含义。此外，部分平台并未提供数据利用的分析指标。在分析的网站中，陕西、南京、石嘴山、无锡、铜仁五个地区的开放平台中并未提供数据的浏览量和下载量指标数值。

②数据利用难度大

在数据格式方面，各地方平台上对于开放的数据格式没有统一的标准，格式较为单一，不利于公众对政府数据的再开发和再利用，在满足多样的用户需求方面尚存在一定的不足。在开放授权方面，有部分的网站数据开放许可协议及标准规范存在的问题，在服务条款、网站声明、免责声明中没有明确用户授权许可。大多数的数据在获取时需网站注册并进行实名认证，用户不能完全自由使用数据。

③数据可持续性不高

数据持续性。在数据发布与更新时间方面，其中有29.41%的地方平台最近一次数据集的发布时间为一周内，27.94%的地方平台数据集的最近更新时间为一周内。在数据更新频率方面，数据更新不及时，静态数据占比偏高，仅有28.36%的数据可以每季度、每月、每周、每日、实时更新。部分数据并未按照规定的更新日期及时更新数据，按承诺更新比例较低。

④缺乏统一的分类标准

在数据主题分类上，因目前尚无通用的政府公开数据主题分类标准体系。各平台所提供的政府公开数据主题划分亦不相同。过少或过多的主题划分均会给平台的数据分类管理工作、用户的数据检索和开发利用带来阻碍，数据主题划分的具体数量尚需根据数据内容信息进一步确定。同一开放平台内部，存在首页所提供的数据主题与数据详情页面不一致的情况。例如，宁夏平台首页共20个主题分类，但其数据集详情页面的主题分类多达51个。此外，存在政府数据开放平台例如雅安，其主题分类与其余各平台之间存在较大差异。

（4）数据开放社会参与度低

政府数据开放平台所发布的数据多来源于政府部门在管理和服务过程中直接产生或采集的数据或政府所采集的外部数据。而一些企业及其他社会组织成立的自主科研机构、研究院、数据开发中心等在进行研究过程中产生了大量与政府公共决策相关的数据，而这些机构往往不具备数据开放的意识，对政府数据开放的参与度较低。

（5）数据利用性不高

在数据应用方面，有大约三分之一对地方平台不

提供数据应用或显示提供数据应用的个数为0，开发的数据应用数量比较少，开放的数据应用多数以APP的形式呈现，数据利用的形式还不够丰富。从数据集的浏览量与下载量来看，由于数据质量偏低或数据更新频率较差等原因，用户在浏览数据集后并没有产生下载的行为，导致数据的利用率较低。用户对于政府开放数据的利用频度、深度和广度都有待进一步提升。

（6）数据利用促进活动较少

北京、广东、上海、贵州等地方政府开展了数据创新利用的促进活动，但只有上海的开放数据创新应用大赛和北京的全国高校数据驱动创新研究大赛等连续举办。大部分的地方政府并未开展数据利用相关的活动和比赛。致使政府和数据用户之间缺乏有效的双向互动。此外，数据利用活动的举办对于政府数据开放也可起到一定的宣传作用，各级政府在这一方面有待加强。

（7）数据利用成果科学性把控较难

用户对政府开放数据进行分析，由于自身的数据分析能力等方面的限制，对于分析结果的科学性和准确性难以有很好的把握。不仅如此，开放数据用户将所得的片面的分析结果通过社交平台等渠道进行发布，会对公众产生误导，从而造成决策失误和相应的损失。

（8）数据利用生态体系亟需构建

政府数据开放是一个从数据产生、开放、利用、价值实现、反馈的动态过程，其涉及政府、企业、公众、数据传递者等多个主体。政府数据开放工作各个流程的协调，各利益主体之间的利益协调、政府部分之间的分工协作等都需得到合理解决和配置。而目前，政府数据开放利用的生态体系还未形成。

6.3 对策建议

基于对政府开放数据有效利用的现状和问题分析，从政府开放数据参与主体视角出发，本报告提出充分发挥政府开放数据要素功能的具体实施路径为形成“释放数据价值”和“提升利用效能”两个中心，构建“三可两好三用”数据利用生态体系。

6.3.1 以释放数据价值为中心的建议

政府数据开放的目标是为了释放数据价值，赋能政府、企业和公众，各级各类政府部门是数据开放工作主体。基于此，以释放数据价值为中心的建议，从数据标准、数据全清单、数据汇聚、数据质量、数据确权和管理制度等基本点展开，如图6-1所示。从而综合提升用户体验、提高用户满意度、提高数据可获得性、提升数据安全和隐私。统一的数据标准体系和数据全清单提升了政府开放数据的工作效率，从而促进数据价值的释放。数据汇聚使不同部门、不同类型的的数据相互关联，从而发挥数据的最大价值。数据质量是数据价值创造的基础。数据确权保障了不同主体在不同阶段的权益，使数据价值产生的收益受益于用户。管理制度是对数据开放工作的约束和引导，为数据价值释放提供保障。

结论与对策



图6-1 释放开放数据价值路径结构模型

(1) 统一数据标准

要尽快建立统一的政府开放数据的元数据标准体系和主题分类体系，保障部门间数据格式、分类标准一致，以便于公众对开放数据的利用和反馈。著录的元数据项及其取值范围和元数据文件格式都需要进一步地统一和规范。对目前国际主流的元数据标准进行参考和整合，同时可根据实际情况添加符合需求的元数据项。建立通用的政府开放数据主题分类体系，设置合理的主题数量，既能满足数据开放的需求也避免为数据分类管理工作、用户的数据检索和开发利用带来阻碍。

(2) 建立开放数据全清单

建立政府开放数据全清单，包含开放清单和负面清单。清单设置的程序要公开化、科学化和民主化。开放清单注明开放数据的内容标准、公开方式、公开时限等所有公开要素，消除数据开放的灰色地带，每年对用户数据申请进行记录和分析，对开放清单进行全方位扩容，为用户提供更多领域和更大范围内的数据。设定负面清单，规定哪些数据不开放，明晰数据开放的禁区。设置后要根据经济社会发展形势的变化，在一定时间内进行适当的调整和补充。政府数据

开放部门可以根据开放数据全清单实行自检，从而提升数据开放的效率。

(3) 加强数据汇聚

加强数据汇聚，拓宽数据汇聚的广度和深度。在广度上，将不同侧面、不同部门、不同类型的数据汇聚起来并加以关联，构建统一的数据管理规则和标准，依据数据分级分类指南对数据进行划分，实现数据资源的分类管理、整合共享和有效配置。在深度上，数据并非简单的汇聚存储，要实现汇聚数据的可融合可共享，加强汇聚数据之间的关联，借助区块链等新一代信息技术，强化跨层级、跨地域、跨系统、跨部门、跨业务的数据融合共用。数据汇聚后期的数据运营和维护是一项较为繁重的工作，可设立专门机构进行数据的运营和维护，对政府开放数据的应用与挖掘进行专门的指导与规划。

(4) 提高数据质量

数据质量是开放数据价值创造的关键。数据质量包含数据完整性、准确性、及时性、一致性、可用性等多个方面。政府机构应建立和完善政府开放数据质量评估标准和质量保障机制，通过对数据采集、数据发布、数据流转等方面提升政府开放数据的质量。冗余、过时、无价值的暗数据，此类数据的价值尚未明确。应尽量最小化暗数据的比例。同时，政府可考虑将政府开放数据质量评估作为政府工作考核的绩效之一，也可借助第三方机构进行政府开放数据质量的评价与考核，为优化政府开放数据质量提供支撑。具体来说，政府在数据开放前应积极开展社会调查，了解用户实际的数据需求，有针对性地开放政府数据；在开放过程中应注重数据管理和评估，保证供给数据的质量，以激发公众使用数据的积极性，从而提高开放数

据的利用率，促进价值实现。应鼓励将开放数据集转化为数据应用，提供多种数据应用形式，吸引公众对于开放数据的利用，释放数据价值。

（5）加快数据确权

数据确权是数据共享、数据交易等的基础和保障，数据确权后可以划分职责，明晰责任，共享利益。要保障不同主体在不同阶段中对数据的不同权益，包括数据所有权、使用权和收益权等。围绕政府数据开放过程中的不同主体，探讨数据所有权、使用权和收益权的保护模式。此外，要保证在数据开放过程中，个人隐私不被泄露，数据不被恶意篡改，保障数据和隐私安全。制定行之有效的用户数据隐私保护与数据确权相平衡的法律规定。

（6）完善管理制度

加快制定和出台我国政府数据开放相关法规，使政府数据开放工作有规可循、有法可依。完善数据开放管理制度，明确政府数据开放的对象、内容、原则、边界，建立协调、监查、保障、绩效评估等制度，确定各级政府机构的数据权限、责任主体，切实保障数据开放执行力度。在政策制定过程中，不仅需要政府机构、领域专家的意见，还需要征求各方用户的真实需求，确保政策的科学性和合理性。此外，要注重法律法规和相关政策的执行，真正做到有法可依、有法必依、执法必严、违法必纠，从而使得政府数据开放更加规范有效。用户管理方面，相关部门可以组织专家学者组成数据开放利用培育团队，通过公开课、新媒体等渠道推进数据利用技能培训和教育，同时可在开放数据平台上提供数据利用和开发各类指南、工具使用手册等。此外，政府可以通过各种途径的宣传对公众利用开放数据加以引导，逐渐培育

起数据利用的社会氛围。工作评估方面，可以借鉴政府网站段评估方式，建立地方政府数据开放平台的绩效评估体系，定期对各地平台建设与数据开发情况进行评估，并按规定时间进行相关情况的发布。

6.3.2 以提升利用效能为中心的建议

政府数据开放工作是一个公共议题，其发展与每个组织和个体息息相关。企业、社会组织和公众作为政府开放数据的受益者，也应参与到政府数据开放进程中。对于用户来说，利用政府开放数据所能获取的收益是其最为关注的。基于此，以提升利用效能为中心的建议，从数据素养、数据思维、数据文化、数据融合和数据流通等基本点展开，如图6-2所示。数据素养、数据思维和数据文化是从意识和技术层面促使用户利用政府开放数据，从而提升数据利用效能。数据融合是打破数据孤岛释放数据价值提高数据利用效能的关键，数据通过流通才能实现价值。

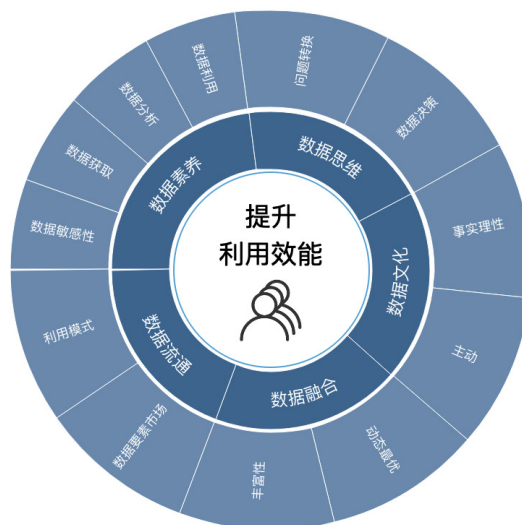


图6-2 提升政府开放数据利用效能路径结构模型

（1）提升数据素养

用户数据素养对政府开放数据的利用效果产生影

结论与对策

响。不断提升用户对于数据的敏感性、数据采集能力、数据分析和处理能力、利用数据进行决策的能力，用户可以公开课、专题讲座、交流协作团队等形式相互交流，不断提升数据素养。企业和公众采集政府开放数据，并加以分析、利用，对于数据使用和分析成果的管控也尤为重要。企业与公众要自觉遵守数据规范与伦理，合理使用数据、发布数据应用成果。用户在使用数据前应进行数据的质量检查工作，发现问题可及时向数据平台进行反馈。对于政府开放平台所出现的一些问题，诸如网站无法访问、数据内容不匹配、开放格式无法满足需求等，也应积极向有关政府部门进行反馈。用户在利用数据后，应主动对数据进行有效评论和打分，以提升政府开放数据的质量，为政府部门优化数据质量提供支撑。此外，用户也可通过政府开放数据监督政府行为，辅助政府决策，实现政府开放数据的全方位利用。

（2）培养数据思维

数据思维是对数据的一种全局性的认识，把问题转换为数据分析问题。用户将自身在特定场景下的问题转换为数据可分析的问题。一方面对于所收集到的数据，应具备一定的辨别力以及抽象概括能力，避免对数据的过度依赖和盲目信任。另一方面，能够有效地管理和使用数据，需要具备合理利用工具的思维方式和能力，发挥数据的价值作出最佳决策，同时避免数据的危害性，使数据的应用恰到好处，并把数据的分析结果应用到特定领域之中，创造价值。在产生数据需求时，多考虑具有较高价值的政府开放数据，使政府开放数据融合到工作生活的方方面面，在分析的过程中尊重事实、小心求证，逐步养成利用政府开放数据的习惯。同时鼓励企业利用政府开放数据为各行

业提供优质的数据服务，各方面合力共同为政府开放数据文化形成提供良好的氛围。

（3）形成数据文化

数据文化是尊重事实、强调精确、推崇理性和逻辑的文化。政府开放数据文化的形成离不开政府开放数据实践的发展。企业、社会组织和公众需要提高自身在政府开放数据实践中的参与意识和能力，提升自己参与公共事务的主人翁意识和责任感。政府开放数据用户应该积极主动参与政府开放数据的相关工作，在政策制定和数据开放过程中，充分发挥服务主体和传播主体的作用。积极表达自身的数据需求，推动政府的数据开放工作更具针对性，同时为数据开放政策和标准的制定提供参考和借鉴。对于数据开放平台提供的问卷调查和意见征集等内容，作为受益者群体有一定的义务作出相应的回应，为数据开放工作提分反馈和支持。此外作为传播主体，应积极主动向周围公众或组织宣传政府开放数据，形成共同应用开放数据的良好氛围。

（4）推动数据融合

数据融合是打破“数据孤岛”、释放“数据红利”、推进政府数据开放的关键。将用户自身数据和政府开放数据融合，充分提升数据利用效能。将不同来源的数据匹配和融合，不强调对数据的拥有，而是强调数据触及和返回的广度与丰富程度。海量、实时、多样的数据可以动态变化、扩展、演化，一旦相互聚合，就能相互作用、相互补充。构建元数据、主数据和全数据的分层分类存储架构实现数据多源对比和动态最优感知，用户可真针对具体业务和目标确定最优的数据指标组合，进而选择获取最优数据源。

（5）促进数据流通

数据只有流通才能实现价值。多维度相关数据的流通融合，才能实现数据价值最大化。政府数据开放后，要加快构建多主体共治共建共享的数据要素市场体系。积极发挥数据交易所等市场中介的作用。优化数据要素流通环境，加强数据流通管理，建立促进数据资源有效流动的的制度规范，推动政府开放数据产业发展，支持构建数据开发利用等场景。创新数据交易模式和数据产品定价方式，通过市场交易应对数据资产交易的需求，使数据真正依价值定价。数据产品一方面要以经济、教育等与用户密切相关的领域为重点，进一步拓展应用场景，另一方面也要利用现有数据预测社会需求，发现社会问题，以数据推动社会治理的精准化。用户应积极与政府构建G2B、G2O、G2C政府开放数据利用模式。对于企业用户，培育带动政府数据开发产业发展的龙头企业，明确其开发利用政府数据的服务方式和盈利模式，充分促进数据流通，

发挥政府开放数据的价值。

6.3.3 构建“三可两好三用”利用生态体系

政府开放数据重在对数据的利用，政府、企业、社会组织和公众应形成合力，共同提高政府开放数据利用率，构建“三可两好三用”利用生态体系。利用生态体系，如图6-3所示。建立政府开放数据利用生态体系的理论模型，在此基础上，相关部门可尝试围绕数据开放、数据整合、数据利用等方面建立政府开放数据利用试验区，通过不断总结可借鉴、可复制、可推广的实践经验，最终形成试验区的辐射带动和示范引领效应。本标准在2018年报告所提出的标准的基础上作出相应的修改和完善。三可侧重于政府在开放数据过程中对数据层面的把控，两好和三用则侧重于用户对于平台本身和利用的感知。

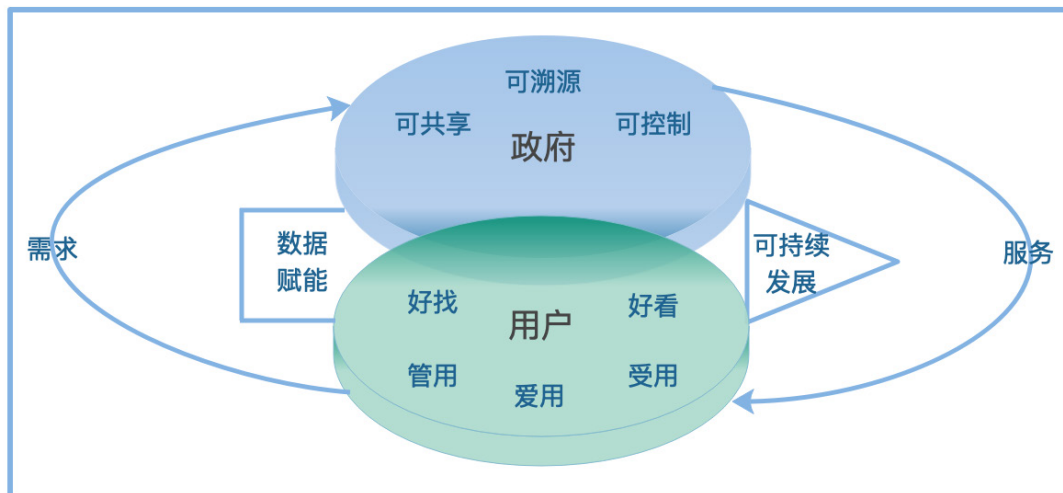


图6-3 政府开放数据利用生态体系

结论与对策

(1) 三可

①可共享

可共享的重点在于数据开放共享的标准和规范设立。一方面，政府应加快建立一套统一的、法定的数据开放共享标准规范，明确开放数据的技术标准和管理标准，同时建立健全政府开放数据共享机制，强化部门协同，以目标为导向，按需共享，规范建设过程，保证共享数据的质量，确保数据共享的安全。另一方面，在数据共享的过程中明确数据共享协议，说明分享哪些数据、分享的理由和方式、访问权限、风险控制和责任承担等，为数据可共享全面保驾护航。

②可溯源

数据可溯源是开放数据质量以及可信度的评估基础。要加强政府开放数据溯源元数据建设，确保政府开放数据有明确的数据来源、路径和时间，在元数据中设立明确的相关字段进行标注。目前的政府开放数据元数据中含有相应的溯源信息，但总体质量还不高，溯源能力较弱，某些字段值存在混乱和错误。因此，需要进一步加强基于全生命周期的政府开放数据利用溯源元数据体系建设，从标准化和规范化方面进一步提升元数据质量。

③可控制

建立对政府开放数据利用全过程的管理控制机制，从制度、技术和舆论三个维度对数据开放利用过程提供全方位的保障。制度维度，一方面政府应不断出台政府开放数据利用相关政策，明确数据利用、数据服务等方面的规范，并加强政策的推行力度，同时也要注意配套政策的协同发展；另一方面，要建立科学合理的开放数据利用管理制度，明确各过程的管理方式和各部门的管理职责。技术维度，理清数据开放

利用的业务流程，在数据筛选、开放、共享、下载等各个环节加强控制，明确各环节的技术规范，同时注重区块链、云计算等技术在政府开放数据利用和平台数据分析中的应用。此外，政府应加强开放数据利用的舆论引导，充分发挥公众开放数据利用传播主体的地位，鼓励公众参与到开放数据利用的管控过程中，实现开放数据利用的共治共建共享。

(2) 两好

①好找

好找体现在用户不仅能够找得到政府开放数据，而且可以找得快、找得准，着力点在于政府数据开放平台的功能性和便捷性。政府数据开放平台是用户获取开放数据的主要途径，在平台功能上，其设计应做到功能齐全、信息分类合理，必要时以问题为导向设置分类模块，要准确地传递数据信息，使用户可以精准获取信息内容，避免出现信息误解。检索方面，应丰富平台的检索方式和排序方式，满足用户在特定情境下的个性化检索需求，并提供对检索结果的分析与可视化。在平台操作上，尽量使用户在不需要学习的情况下就可以便捷地开展平台操作；同时，政府部门应逐步加强数据开放平台的人性化设计，减少用户注册和下载数据过程中的繁琐操作。相关部门应加大技术方面的投入，实现网站对于用户的快速响应，降低用户数据获取的时间成本。

②好看

好看是使政府开放数据具备较好的视觉体验，既能看得见，又能看得懂、看得舒适，着力点在于政府数据开放平台的设计性和美观性。视觉信息比文本信息更易于接收，数据可视化可以促进用户快速接收信息，在一定程度上增强了政府开放数据对用户的吸

引力。首先，要确保为用户提供看得见的政府开放数据，逐步加强政府数据开放的可视化建设。以图、表的形式展示数据动态变化信息，同时可考虑创造性地利用现代通信工具、多媒体技术或3D虚拟现实信息搜索环境等先进的技术来加强计算机生成的虚拟环境的浸入感，使用户获得更好的体验。其次，政府开放数据可视化的界面要简洁清晰，既符合用户的认知特性，也能引导视觉流向，让用户更流畅地接收信息，突出通过数据所要表达的主题，确保用户看得懂展现的数据内容。同时，数据可视化的设计要尽量符合美学的设计原则，在布局、配色、对比等方面增强视觉效果，为数据用户带来舒适的视觉体验。

（3）三用

①管用

促使政府开放数据在实战中管用。政府开放的数据应是用户关切或能满足用户信息需求的数据。政府部门应在确保信息安全的情况下开放多领域多主题的数据，并及时更新，保证数据的完整性和全面性。做好数据汇聚工作，充分发挥数据之间的关联关系，挖掘数据关联之中的巨大价值。此外，要遵循以问题为导向的原则，在数据开放过程中突出数据应用场景，开发在某一应用场景下相关的数据接口和应用，可尝试将应用场景或领域作为元数据字段归入元数据体系

中，方便用户基于元数据进行相关检索。

②爱用

做到用户爱用政府开放数据，这一标准体现用户对于政府开放数据的认可和忠诚度。一方面，政府应积极开放满足用户需求的高质量数据，提供丰富的数据开放格式，降低用户数据获取和利用成本的成本，从而增强用户利用政府开放数据的舒适感，便于培养用户长期利用政府开放数据的习惯。另一方面，政府应提升开放数据的透明度和完整度，同时做好数据安全保障工作，加强政府自身能力建设，增强用户对政府开放数据的信任度和认同感。综合提升用户对政府开放数据的粘度。

③受用

使用户对政府开放数据感到受用，增强用户利用开放数据的获得感。对于政府开放数据利用的获得感，既包括物质层面的也包括精神层面的，首先是要感受到数据利用带来的问题解决和价值增值。比如，企业通过利用政府开放数据从而增加了利润，科研人员通过利用政府开放数据提升了科研绩效，这些都是看得见摸得着的获得感。在精神层面，要让用户通过对政府开放数据利用而产生成就感，能够享受政府开放数据所带来的公平公正的权利。

联合发布方



华中师范大学信息管理学院



湖北省数据治理与智能决策研究中心

报告出品团队

段尧清 夏立新 李玉海 石义金 易明 娄策群

尚婷 陈玲 林平 刘宇明 何思奇 周密 王冰清 汤弘昊 聂楚函 何俊雨 宗莹萍