

中国信息政策研究主题的历时演进特征 (1986—2015年)*

裴雷, 李向举, 谢添轩, 成吉
(南京大学信息管理学院, 南京 210093)

摘要: 本文采用大样本数据分析方法, 对我国1986—2015年信息政策相关文献、热点主题的分布特征和规律进行探索性研究。研究发现, 政策主题历时演进存在典型的主题差异和涨落特征, 并且不同的主题区分尺度的研究结论具有较好的一致性和稳定性, 而从主题突变的时间分布中观测到中国信息政策研究的典型四个分期, 表明主题跟踪方法在政策比较分析、政策预测中具有一定的可行性。

关键词: 信息政策; 热点主题; 主题跟踪; 阶段划分

中图分类号: G20; G35

DOI: 10.3772/j.issn.1673-2286.2016.7.004

1 引言

信息政策的发展源于20世纪50年代初, 信息技术的引入和信息工作特定时期诞生了以推动信息利用和服务为目的国家政策。而热点体现为政策实践和理论研究两个层面: 我国信息政策实践最早可追溯到建国初期的文献保障服务政策; 而理论研究则是以1981年能源保密政策为代表的科技保密政策研究作为起始^[1]。1993年是中国信息政策研究的拐点, 随着20世纪90年代初的科技信息机构改革与转型、1994年以后的信息化推进, 信息政策理论和实践逐步成为研究热点, 关于信息化、信息产业、电子政务、政府信息公开、电子商务、信息安全等政策热点次第出现, 构成我国信息政策研究的主体^[2]。

为刻画信息政策研究的发展规律和主题演进特征, 研究者往往会提出信息政策发展的分期问题。在以往研究中, 发展阶段分期和世代分期是典型方法。在发展阶段分期中, 萌芽期、发展期、成熟期、衰退期是典

型的分期方法, 但发展阶段分期主要是基于标志性历史事件进行判断, 而不同研究者对标志性历史事件的主观研判具有一定偏差, 进而发展阶段分期在不同学者研究中并不完全一致^[3-5]; 世代分期则多以十年作为时间轴进行假想阶段区分, 但可能对某些政策研究主题的持续发展进行割裂^[6-8]。而事实上, 我国信息政策实践周期多与“五年规划”一致, 五年分期更符合政策实际。

本文主要根据我国信息政策研究的主题“涌现”, 采用五年分期方式描述个体差异, 聚焦刻画政策研究的主题演进。相对以往针对政策文献数量的计量和统计分析^[9-11], 本文更注重政策主题在时序上的演进和涨落特征。

2 数据采集与研究方法

2.1 中国信息政策文献的采集

一直以来, 信息政策是一个易于理解、却难以界定

* 本研究得到教育部人文社会科学青年项目“基于概念统计的信息政策文本计算与实证”(编号: 11YJC870020)和国家社会科学基金青年项目“信息政策扩散与转移研究”(编号: 12CTQ024)资助。

的术语,宽泛、多元和缺乏统一的范畴^[12],信息范畴和政策范畴并未形成固定概念,在具体研究环境中,或依据研究对象的指向进行界定,或根据政策制定机构的管辖范围进行界定^[13-14]。对政策的理解也存在狭义和广义的区分,广义的信息政策将信息法律也纳入研究范畴。

本文在文献采集中采纳广义概念,在主题范畴上则主要参照我国信息政策实践。根据对中国 59 751 项信息政策的主题标引(自建数据库),共涉及 47 个一级主题,其中信息公开、电子商务、电子政务、信息产业、保密政策、知识产权、信息安全、科技创新、文化产业、信

息化 10 个一级主题相关的信息政策占总量的 62%,为中国主要信息政策主题。因此,本文以“信息政策、信息法律、信息立法”为标题检索入口词进行第一轮检索,去重后得到 2 797 篇研究文献;并以“政策主题+(‘政策’ or ‘法律’)”作为标题检索入口词,将时间限定在 1986—2015 年进行联合检索。合并两轮检索结果,剔除没有实际意义和不相关文献,共得到 6 373 篇研究文献(见图 1)。文献分布表明,信息政策研究文献将保持持续增长趋势,证实中国信息政策领域的研究活力与成长潜力。

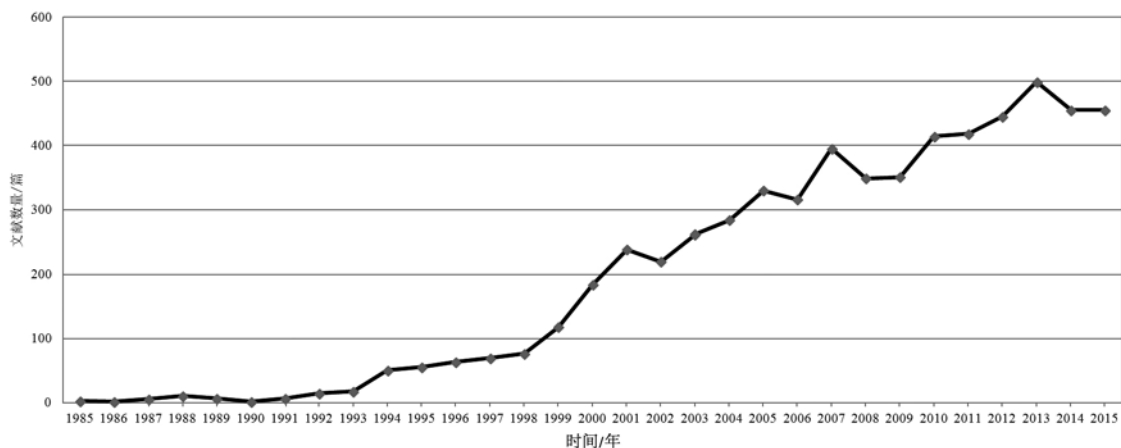


图1 信息政策研究文献的时间分布

2.2 信息政策研究主题跟踪研究方法

(1) 信息政策研究主题的定义。信息政策文献研究主题包括初始定义主题以及检索文献的描述主题,后者可通过关键词、标题和摘要等不同语料进行提取和清洗。具体而言,本文的研究主题定义包括三个层面。

第一,政策主题领域分类,即初始定义的10个主题领域,主题频数描述是文献数量,以该主题检索入口所得的有效文献数量表示。特别地,跨主题文献可重复计入不同主题统计口径,因而得到总频次 6 776 次。

第二,关键词主题,由研究文献的关键词描述。在研究中因关键词质量差异较大,需进行规范、合并和清洗,最终得到表意能力比较突出并具有一定排他性的 10 948 个关键词主题,共计词频 34 779 次。其中,排名前 40 位的关键词主题占总词频的 27.0%,词频大于 10 的前 247 个关键词主题占总词频的 41.2%,具有一定典型性。

第三,自然语言分词主题。即对关键词进行分词处理,以提取更小尺度的语义概念,构成信息政策的元主

题。因一些论文关键词质量较差,影响政策主题的语义表达,自然语言处理虽破坏了政策概念的完整性,但对更小尺度的概念分析具有一定显示度,其研究也具有启发性。按照分词方法,从文本集中提取到 5 781 个自然语言分词主题,共计词频 76 639 次。其中,前 10% 的高频主题占总词频的 79.1%,排名前 40 位的词频占总词频的 40.4%,而词频大于 20 的前 706 个自然语言分词主题占总词频的 78.1%,具有实际政策意义的自然语言分词主题共 390 个。

(2) 信息政策主题突变与跟踪方法。识别并跟踪学科发展主题 (Topic Detection and Tracking) 是学科前沿理解和政策决策分析的重要手段,也是当前信息检索研究的热点命题。从方法上看,存在大量统计学性质的方法识别^[15-18]、基于引文和耦合聚类的可视化方式^[19-21],还有大量主题识别和机器学习方法^[22-23]。本文并非探索主题跟踪方法的优劣,而主要是对两种方法的分析结果进行比较研究。

第一,采用累积增长曲线和频次热图两种可视化图形方法,跟踪和描述三类政策主题随时间的频度变

化,并根据累积增长曲线判断政策主题计量的极值和拐点,利用热图描述政策主题群的个体演进规律并进行辅助判断。

第二,运用时间序列聚类法测算政策主题的突变点或局部突变点^[18]。假定某一主题 n 年研究样本序列中第 i 年的频次为 A_i ,并且第 τ 年为突变点,则有

$$\Delta SA_{\tau-} = HSA_{\tau-} - USAV(A_{\tau-}) = \sum_{i=1}^{\tau} (A_i - \overline{A_{\tau-}})^2$$

$$\Delta SA_{\tau+} = HSA_{\tau+} + USAV(A_{\tau+}) = \sum_{i=\tau+1}^n (A_i - \overline{A_{\tau+}})^2$$

其中, $\overline{A_{\tau-}}$ 表示突变点之前序列集合中的平均频次, $\overline{A_{\tau+}}$ 表示突变点之后序列集合中的平均频次,并定义总离差: $S(A_{\tau}) = V(A_{\tau-}) + V(A_{\tau+})$ 。那么,当 $S(A_{\tau})$ 为极小

值时,所对应的 τ 时刻点可能为突变点。

3 中国信息政策研究主题历时演进特征描述

3.1 中国信息政策主题历时演进的可视化描述

分别对政策主题领域分类、典型关键词主题和自然语言分词主题进行历时跟踪,并绘制对应的累积增长曲线和频次热图(见图2),发现不同层次主题的统计中,信息政策主题均存在差异增长和涨落现象,并且呈现出相似的主题涨落趋势。

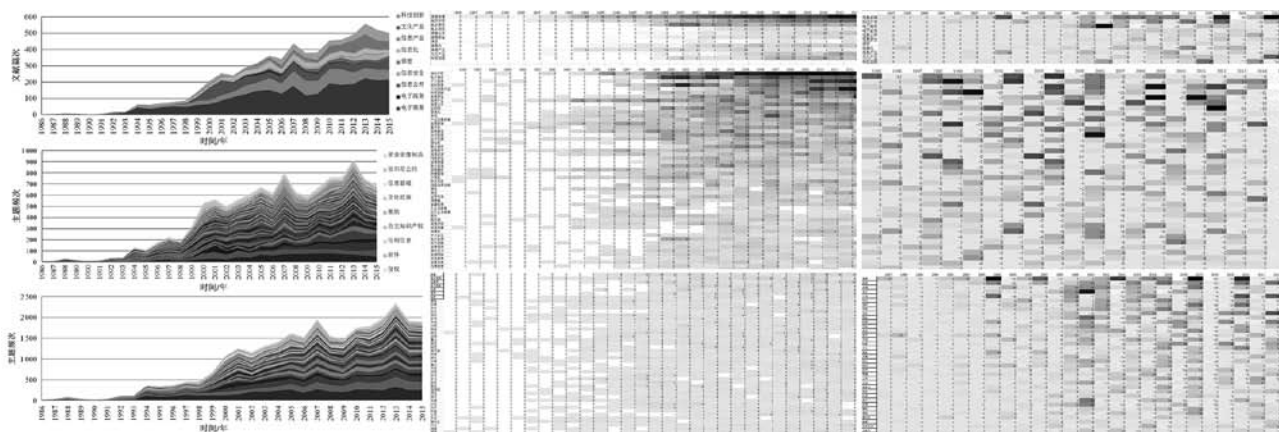


图2 信息政策主题累积增长曲线与频次热图

第一,不同主题分类方式的整体演进趋势一致。主题领域分类的累积增长曲线在1993—1994年、1998—2001年、2006—2007年以及2009—2010年存在显著的四个快速增长阶段;关键词主题累计增长曲线波动更加剧烈,但具有相似的高速增长时段,即1993—1994年、1998—2000年、2006—2007年以及2012—2013年,并新增1995—1997年的快速增长阶段;自然语言分词主题则形成1993—1994年、1998—2001年、2006—2007年和2012—2013年的四个典型增长高峰期,同时2002—2005年的增长也比较明显。从热图色温变化形状看,三者的分布热图和渐变热图基本一致,体现出相似的整体演进趋势。

第二,不同主题分类方式的分区政策热点基本一致。因政策主题的个体在表述和内涵上存在差异,因而分区热点的描述存在一定异同:从主题领域分类看,知识产权、电子商务、科技创新、文化产业和信息化的政策主题热度更高;在关键词主题中,知识产权、电子

商务、个人信息、税收政策和文化创意产业的政策主题热度更高,并且引入了更多细分主题,政策主题的观测范围更大;在自然语言分词主题中,信息、政策、法律、电子、产业、知识、保护、产权等主题片段提及率最高,虽在表述上破坏了关键词术语的完整性,但在语义层面的主题片段与关键词主题保持一致。

第三,不同主题分类方式的政策概念词频涨落基本一致。政策概念词频涨落,能够揭示政策主题的变动规律。如图2所示,无论是左半区峰值与谷值的分布,还是右半区黑区与白区的分布,都显示2000—2010年是波动比较剧烈的区间。虽然政策主题波动的个体差异性仍存在,但均显示出相似的整体涨落趋势。从政策主题的个体调查看,1994年以后个人信息、会计信息、隐私权、科技创新、网络安全、产业发展政策和信息安全政策等主题增长幅度比较明显,而信息化、税收政策、反垄断、信息市场、信息公开、信息通信等政策主题在部分年份出现了大的衰退。

3.2 中国信息政策的主题突变期测算与分布

分别对政策主题领域分类、典型关键词主题和自然语言分词主题进行主题突变期的测算，并统计政策主题突变期的时间分布，得到突变期的时间分布曲线（见图3）。从而发现，典型关键词主题和自然语言分词主题具有相似的峰值分布和增长轨迹。

如表1所示，自1992年开始，处于突变期的政策研究主题分布显著增多，说明信息政策研究开始加快发展；1993—2000年是信息政策研究高速发展时期，1999年达到峰值，说明此阶段新兴的研究概念大量涌现，抑或过去研究术语快速衰退；2001—2007年是稳定发展阶段，新兴研究术语的增长幅度显著降低，但总的概念频度增长明显；2007年以后进入发展稳定期，尚未出现衰退迹象。

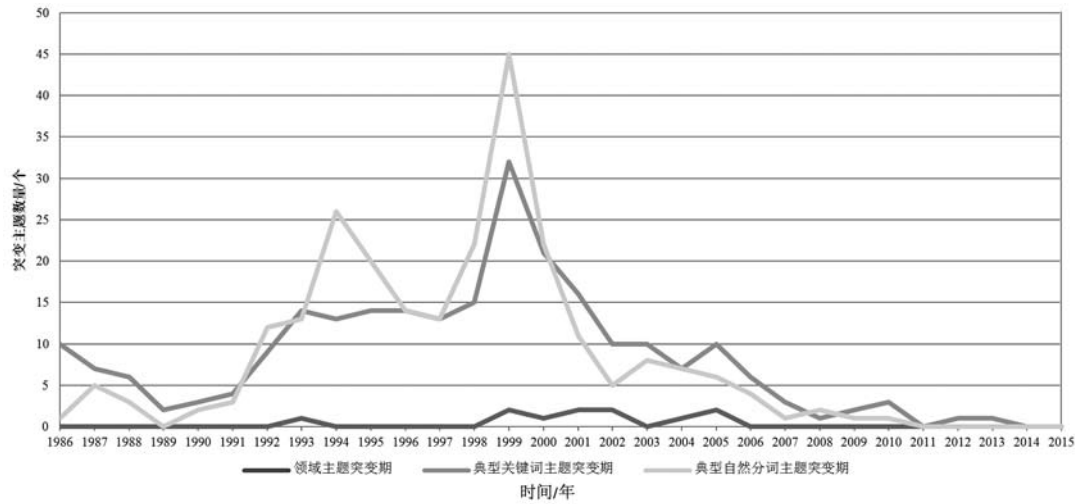


图3 信息政策主题突变期的时间分布

表1 信息政策研究主题突变期的时间分布

时间/年	Domain	Keywords	NaWords	时间/年	Domain	Keywords	NaWords	时间/年	Domain	Keywords	NaWords
1986	0	10	1	1996	0	14	14	2006	0	6	4
1987	0	7	5	1997	0	13	13	2007	0	3	1
1988	0	6	3	1998	0	15	22	2008	0	1	2
1989	0	2	0	1999	2	32	45	2009	0	2	1
1990	0	3	2	2000	1	21	22	2010	0	3	1
1991	0	4	3	2001	2	16	11	2011	0	0	0
1992	0	9	12	2002	2	10	5	2012	0	1	0
1993	1	14	13	2003	0	10	8	2013	0	1	0
1994	0	13	26	2004	1	7	7	2014	0	0	0
1995	0	14	20	2005	2	10	6	2015	0	0	0

注：“Domain”代表主题领域分类，“Keywords”代表典型关键词主题，“NaWords”代表典型自然语言分词主题。

3.3 中国信息政策频度涨落与主题突变期分布比较

从政策主题的频度涨落和突变主题的分布密度比较，具有三个特点。

第一，主题频度涨落比较明显的1993—1994年、1998—2000年，也是突变主题分布比较密集的时间阶

段，因而可以判定上述两个时间阶段是中国信息政策研究的关键转变时期；2006—2007年体现为主题频次的突增，而未体现出突变主题的明显增加；2012—2013年的突变主题密度不明显，可能因对典型政策主题进行了截频处理，从而使最新涌现的低频政策概念无法计量。

第二，不同主题的频度涨落与突变期分布关系并不

显著。以领域主题分类为例（见表 2），相对比较一致的是电子商务和电子政务主题，其突涨时间与突变点基本一致，其他主题的频度涨落与突变期关系并不显著。典型关键词主题和自然语言分词主题也具有类似的特征。

表 2 信息政策研究领域主题分类突变期与频次涨落比较

	信息政策	知识产权	电子商务	电子政务	信息公开	信息安全	保密	信息化	信息产业	文化产业	科技创新
频次最高	2013	2011	2001	2005	2011	2013	2011	2005	2004	2012	2013
增幅最大	2010	2007	1999	2003	2009	2013	2011	2005	2004	2012	2013
突变期	1999	2001	1999	2002	2004	2002	2000	2001	1993	2005	2006

第三，涨落与突变均具有一定的合理性。事件历史分析法（Event History Analysis）最早由Berry夫妇在彩票与税收政策扩散研究中引入^[24]，即以一段时间的典型政策事件为标志，对政策时间序列进行分析的方法^[25]。通常在研究中会引入重要政策事件进行对应分析，是国内政策阐释研究中的典型方法。以主题热度突变为例，热度突变反映在特定政策环境下，即某主题热度显著高于或低于上年度或上区间的热度现象，一般是政策事件延续效应、政策即将出台、政策危机或变革的标志。通过信息政策主题的跟踪分析，本文提取了 10 个典型的热度突变政策主题（见表 3），梳理并提取了相应时间段的典型政策事件。

表 3 信息政策研究主题突变与关联政策事件

政策主题	主题突变	典型关联政策事件
领域主题分类	信息产业1993年突变	1994年全国信息化领导小组成立，国家信息政策开始系统推动
	电子商务主题1999年、2000年突增与主题突变	业界认为1999年为电子商务元年，电子商务实践开始兴起
	信息公开主题2009年	2008年《政府信息公开条例》发布，2009年全面落实实施
	信息安全主题2013年	2012年，《全国人民代表大会常务委员会关于加强网络信息保护的決定》；2014年，中央网络安全和信息化领导小组成立
关键词主题	电子商务1999年	同上
	个人信息2004年、2009年	2003年国务院委托有关专家开始起草《个人信息保护法》，2005年专家建议稿完成
	信息安全2012年	同上
自然语言分词主题	1994年“信息”“产业”	同上
	2007年“公开”	同上
	2013年“安全”	同上

分析发现，不同主题分类方式的研究结论基本一致，而且研究主题突变与政策环境存在一定的对应关系，政策主题跟踪法在政策阐释层面具有一定的关联性和合理性。

4 信息政策主题历时演进特征分析

4.1 中国信息政策主题领域分类的历时演进特征分析

根据主题词频涨落特征，可将信息政策主题领域分为四种类型。

第一类，激活性政策研究，即从某一时间点“被激活”后，其文献数量显著多于以往并保持研究热度的研究类型。这类主题包括信息公开政策、信息化政策、信息安全政策、文化产业政策和科技创新政策等，如2007年的信息公开政策研究、2010年的文化产业政策以及2013年的科技创新政策研究。

第二类，持续激活性政策研究，即分阶段由几个“激活点”或“加速增长点”，推动该主题研究文献的持续增长，信息政策总体研究就体现出类似特征。如知识产权和电子商务政策研究，1995年是知识产权文献第一次加速增长阶段，2006年后进入第二次加速增长；电子商务政策研究从1998年开始启动，2000—2001年

是突破性发展阶段, 2014—2015年又开始加速发展。

第三类, 休眠性政策研究, 即具有休眠特征, 某一阶段激活成政策热点后开始冷却, 文献数量持续降低, 直至下一个阶段再次激活。这类主题如信息产业政策研究, 该研究大约每4—5年为一个周期, 1994年第一次爆发性增长, 1999—2002年为第二个高峰期, 2004年、2008年、2013年先后三次成为热点, 目前则趋于平缓。

第四类, 消褪性政策研究, 即过去某一时段为研究热点, 但高峰期过后文献数量显著下滑并持续低迷的研究类型。这类主题如电子政务宏观政策研究, 该研究于2002年启动, 2004—2006年是高峰发展阶段, 但目前宏观政策研究已不多, 或已转入具体政策

问题研究。

4.2 中国信息政策典型关键词主题的演进特征分析

通过主题频度的历时分布, 能够跟踪发现信息政策文献主题的演进与消亡趋势, 发现当前信息政策研究主题的变化特征。通过对典型关键词主题1986—2015年的频次分布及涨落特征余弦相似度计算, 大部分关键词分布的两两相似度极高, 差异性并不显著; 但以五年为区间跨度, 对政策概念进行分区后, 不同区间政策主题具有显著差异(见表4)。

表4 信息政策研究关键词的分区分布(前10名)

第一时区 (1986—1990年)		第二时区 (1991—1995年)		第三时区 (1996—2000年)		第四时区 (2001—2005年)		第五时区 (2006—2010年)		第六时区 (2011—2015年)	
主题	词频	主题	词频	主题	词频	主题	词频	主题	词频	主题	词频
信息技术	11	信息产业	44	信息产业	91	电子商务	215	知识产权	355	个人信息	328
计算机	9	科技信息	25	知识产权	86	知识产权	204	电子商务	189	知识产权	326
信息通信	5	信息政策	24	电子商务	70	税收政策	150	个人信息	149	文化创意产业	287
经济发展	5	信息服务	24	信息政策	64	信息产业	96	税收政策	129	电子商务	195
信息产业	4	著作权	16	信息通信	60	会计信息	87	科技创新	129	科技创新	156
个人信息	3	信息技术	14	税收政策	49	信息化	86	文化创意产业	129	税收政策	148
信息政策	3	专利	14	计算机	42	互联网	85	信息公开	107	信息安全	144
隐私权	3	计算机	13	经济发展	42	信息安全	83	信息安全	89	互联网	108
信息交流	3	信息通信	13	专利	40	信息通信	80	专利	83	信息公开	102
税收政策	2	信息市场	11	著作权	39	电子签名	72	信息化	78	产业发展政策	91

参考分类主题的类型区分方法, 关键词可区分为五类(见表5)。

第一类, 具有持续较高的词频分布, 即信息政策的持续研究热点, 贯通信息政策研究始终, 这类关键词共有48个, 典型的如信息产业、信息服务、信息基础设施、信息技术、保密工作和信息化等。

第二类, 关键词词频呈递增趋势并保持高频规模, 为信息政策当前研究热点, 从247个候选关键词中找到34个, 如信息安全、信息生态、信息环境、信息公开和信息伦理等。典型的如信息安全政策, 自第三时区(1996—2000年)大量出现(14篇), 第四时区(2001—2005年)增长到83篇, 此后保持持续高速增长, 并稳定高频规模。

第三类, 关键词词频呈递减趋势的, 为信息政策以往研究热点, 目前已经衰退的信息政策关键词, 如信息

系统、信息污染、信息机构和信息市场。关于信息系统的政策研究在第三时区和第四时区即1996—2005年达到峰值(35篇)后, 从第五时区开始迅速递减(15篇), 到第六时区已经很少有关于信息系统政策研究的文献(5篇)。信息污染则是在第三时区(1996—2000年)最多, 达到11篇, 此后迅速衰减, 目前已经没有关于信息污染的文献。信息机构, 在第二时区与第三时区之间, 相关的政策研究分别高达9篇和21篇, 而进入第四时区(2001年)后, 相关文献锐减为4篇, 而第五、第六两个时区仅1篇相关研究文献。

第四类, 出现在最近时区, 并迅速增长至关键词高频区, 为信息政策潜在研究热点。典型关键词是个人信息保护, 在第五时区之前鲜有相关研究, 但自从引入该关键词后, 均保持在高频词区, 年均相关论文都在40篇以上。

第五类，在多个时区出现，词频总量较高，但词频数并没有严格的时序变化规律，为波动性研究热点。这类关键词广泛分布于理论研究的核心议题和中频信息政策概念区间，如信息素养、信息交流、信息需求等。

表 5 信息政策研究的典型关键词类型

关键词类型	典型关键词示例
持续型	经济发展、电子商务、知识产权、信息化、信息服务、信息管理、信息基础设施、信息技术、保密工作、公共信息、技术进步、科技政策、电子信息、税收政策、档案信息
增长型	信息安全、信息公开、信息生态、信息伦理、知情权、信息环境、网络安全、文化产业、信息资源配置、信息意识、信息异化、网络信息资源、国家秘密
衰退型	信息系统、信息污染、信息市场、信息事业、信息资源开发、信息主体、信息源、信息时代、信息立法、电子政务、信息技术、信息通信政策、信息机构、信息活动、信息服务业、国家信息政策
潜在型	个人信息、隐私权、科技创新政策
波动型	信息共享、信息用户、信息需求、信息素养、信息权利、信息披露、信息内容、信息交流、信息犯罪、电子支付、信息产业

4.3 中国信息政策关键词主题突变演进特征分析

主题突变演进主要跟踪主题引入或衰退的时间节点，对分析政策研究的时期特色具有一定参考价值。依典型关键词的突变期分布可将中国信息政策研究分为1986—1992年的萌芽发展期、1993—2000年的加速发展期、2001—2007年的发展成熟期以及2008—2015年的稳定发展期四个典型时期。根据上述发展区间的突

变主题分布（见表6），发现如下特征。

第一，萌芽发展期的引入信息政策主题主要集中在知识产权相关术语和信息活动相关主题。在知识产权相关术语中，国际知识产权理论的引入是政策研究的亮点，比如巴黎公约、伯尔尼公约、涉外知识产权等术语的引入；在信息活动相关主题中，信息交流、信息市场与信息服务等政策主题的引入是规范信息活动的重要政策术语。此外，对于信息安全和保密的关注则主要以国际对抗和国家安全为主，具有典型的冷战特色。

表 6 信息政策研究典型关键词突变主题的分期分布

典型发展时期	典型关键词突变主题示例
萌芽发展期 (1986—1992年)	信息活动、技术进步、文学艺术作品、信息交流、国家主权、技术合同、危害国家安全、军事信息、巴黎公约、信息市场、信息保密、伯尔尼公约、信息产品、意思自治原则、信息技术、经济信息、信息权利、信息主体、国际条约、涉外知识产权、工业产权、录音录像制品、信息事业、信息内容、科技信息、信息服务、技术转移、产权保护、外观设计专利、技术信息、开发利用、国际比较
加速发展期 (1993—2000年)	信息产业、信息政策、计算机、商标、信息传播、信息管理、信息机构、信息社会、软件、信息通信、信息资源、金融信息、图书馆、信息法规、信息基础设施、信息犯罪、智力创作、软件著作权、专利、著作权、信息系统、不正当竞争、反不正当竞争、信息环境、版权、信息网络、商业秘密、网络服务、科技成果、信息产权、信息化、数据管理、数字签名、电子数据、政府信息资源、电子支付、域名注册、知识经济、网络交易、黑客攻击、电子商务、互联网、电子签名、电子合同、网络安全、数字内容、跨境电子商务、电子商务税收、WTO、信息自由、网络隐私、电子证据、隐私权、TRIPS协议、自主知识产权、网络信息资源、网络知识产权、域名
发展成熟期 (2001—2007年)	信息安全、信息披露、信息不对称、政务信息公开、信息共享、技术标准、信息隐私权、诚信、信息公开、电子政务、政务信息、教育信息化、信息公开法、文化政策、基因信息、知情权、信息网络传播、侵权、创新政策、保密、创新、文化创意产业、个人信息、科技创新、自主创新、区域创新、网络传播、定密、环境信息、三网融合、第三方平台、信息泄露、开放获取
稳定发展期 (2008—2015年)	金融支持、财税政策、质押融资、物联网、财政科技投入、云计算、信息消费、大数据

第二,加速发展期引入的政策主题最多,主要可分为三个部分:一是以信息技术和数字环境为主的信息活动或对象,如计算机、软件、互联网、信息通信、数据管理、网络服务、域名、黑客等主题;二是以电子商务和知识经济为对象的新经济活动研究,如电子商务、电子签名、电子合同、跨境电子商务、信息社会、不正当竞争、反不正当竞争等主题;三是知识产权的国际化 and 数字化延伸,如软件著作权、商业秘密、WTO、TRIPS协议、自主知识产权、网络知识产权等。

第三,发展成熟期引入的典型政策主题是电子政务与信息公开、文化创意产业与创新相关的主题。电子政务与信息公开主题是信息政策研究发展成熟期的典型应用主题,包括电子政务、政务公开、信息公开、信息披露、信息共享以及与之关联的定密、保密、个人信息、知情权、信息隐私权、信息安全、信息泄露等新的理论实践;文化创意产业与科技创新则是另一个政策研究创新领域。

第四,稳定发展期采集的高频政策主题并不多,仅就测算出的8个主题而言,主要体现为点状分布,如物联网、云计算、信息消费、大数据等与同时期国家发布或推进的相关国家信息政策密切吻合。此外,关于科技金融政策和财税政策也是同期的政策研究亮点,比如知识产权的质押融资、科技金融支持政策具有典型的时期特色。

5 结论与展望

本文通过对我国信息政策文献集的主题提取、跟踪统计以及主题突变测算,发现不同政策主题历时演进的趋同性、差异性以及主题突变的分期分布,也说明政策主题跟踪方法在政策分析、政策预测中的可行性。

首先,政策主题历时演进分析发现了典型的主题差异和涨落特征。从分类主题和关键词主题的特征分类看,主题差异和涨落主要体现在四个方面:第一,政策主题的快速增长,体现了政策热点的孕育或出现。比如政策主题历时跟踪发现1993—1994年是我国信息政策研究的突增点,这一判断与传统信息政策研究结论一致^[26];而2014年的“互联网+”和2015年的个人信息保护对当前信息政策的研究导向和政策实践的揭示也具有一定辅助和指引。第二,信息政策研究主题的快速衰退或因政策为局部热点、短期热点,或因替代性政策主题的出现,比如2004年信息通信政策、2005年

信息化政策、2006年电子政务等几组典型政策主题的衰退。第三,信息政策研究主题的涨落具有一定生命周期,体现了信息政策的周期性。如1998—2004年的信息通信政策,从快速增加到快速下降,经历了一个政策研究周期。第四,个别政策主题保持了比较稳定的热度增长,比如知识产权、信息技术等。

其次,政策主题突变期分布的分期方法,与我国信息政策实践具有一定趋同性,因而政策主题跟踪在政策发现、政策比较研究中具有一定的可用性。虽然有研究报告发现,信息政策的理论研究热点与政策实践热点并不完全一致,二者均可能存在一定的滞后或遗漏,但该方法关于电子商务、电子政务、WTO、大数据、云计算等政策主题的趋势研判还是比较准确的。

此外,从研究方法看,基于时间序列聚类方法的主题跟踪还存在一些不足。首先,时间序列聚类方法是一种历时性的回溯研究方法,需要基于数据积累,而对于最新涌现的低频政策概念无法计量,其研判精度不如主题频次统计跟踪;其次,时间序列聚类方法对封闭系统比较有效,但科学研究是开放系统,处于不同增长和发展过程中,其分期比较受局限;最后,时间序列聚类方法比较适合单纯逻辑增长的研究主题,进行逻辑上的两段区分,而对波动性或具有多阶段发展的研究主题无法准确区分。

总之,政策研究主题的历时跟踪更适合对热度涨落比较剧烈的政策主题预警,以及长期跟踪政策比较和政策学习研究。

参考文献

- [1] 周林彬.试论经济信息法律关系[J].法学杂志,1985(2):37-38,27.
- [2] 马费成,裴雷.我国信息资源政策与法律的实践与需求[J].图书馆论坛,2008(1):4-8.
- [3] 顾胜杰.我国信息政策研究综述[J].晋图学刊,1994(4):5-9.
- [4] 张新宇,尚萍.国外信息政策研究述评(1990—2001)[J].现代情报,2003,23(4):27-30.
- [5] 王素芳.我国信息资源开发利用政策法规初探[J].图书馆学刊,2004,26(4):16-17.
- [6] 朱烨.我国信息政策研究(1994—2003)综述[J].图书馆建设,2005(3):27-30.
- [7] 李杨.信息政策研究论文定量分析(1999—2008)[J].情报探索,2010(3):17-19.
- [8] 汪少敏,郑满满,王爱群.我国信息政策研究评述(1999—2008)[J].情报

- 杂志, 2009 (8):37-39.
- [9] 张璋.我国信息资源产业政策:现状、分析与前瞻[J].图书情报工作,2012,56(12): 85-89,137.
- [10] 刘君.中国信息政策研究论文(1980—1998年)定量分析[J].图书情报工作,2000 (2):21-25.
- [11] 朱莉,朱庆华.20世纪90年代以来我国信息政策与法规研究论文的定量分析[J].情报理论与实践,2003,26(4):371-375.
- [12] BROWNE M.The field of information policy:2.redefining the boundaries and methodologies[J].Journal of Information Science,1997,23(5): 339-351.
- [13] DRAKE M A,MAACK M N. Encyclopedia of Library and Information Science[M].2nd ed.New York:CRC Press,2003:1300-1315.
- [14] ROWLANDS I.Understanding information policy:concepts,frameworks and research tools[J].Journal of Information Science,1996,22(1):13-25.
- [15] POPESCU A,FLAKE G W,LAWRENCE S,et al. Clustering and identifying temporal trends in document databases[C]//Proceedings of IEEE Advances in Digital Libraries 2000,May 22-24,2000,IEEE Computer Society.Washington:[s.n.],2000:173-182.
- [16] ROY S,GEVRY D,POTTINGER W M. Methodologies for trend detection in textual data mining[C]// The 2nd SIAM International Conference on Data Mining,11-13,2002.Arlington:SIAM, 2002:158-174.
- [17] 王莉亚.基于关键词突变的主题突变研究[J].情报理论与实践, 2013,36(11):45-48.
- [18] 王莉亚.基于特征词突变的信息整合领域主题演化研究[J].图书情报研究,2014(3):75-78.
- [19] CHEN C M.CiteSpace II: detecting and visualizing emerging trends and transient patterns in scientific literature[J].Journal of the American Society for Information Science and Technology,2006,57(3):359-377.
- [20] SHIBATA N,KAJIKAWA Y,TAKEDA Y,et al.Detecting emerging research fronts based on topological measures in citation networks of scientific publications[J].Technovation,2008,28(11):758-775.
- [21] MORRIS S A,YEN G G.Crossmaps:visualization of overlapping relationships in collections of journal papers[J].Proceedings of the National Academy of Sciences,2004,101(suppl1):5291-5296.
- [22] 李湘东,张娇,袁满.基于LDA模型的科技期刊主题演化研究[J].情报杂志, 2014(7):115-121.
- [23] 杨星,李保利,金明举.基于LDA模型的研究领域热点及趋势分析[J].计算机技术与发展,2012(10):66-69,74.
- [24] GRAHAM E R,SHIPAN C R,VOLDEN C,et al.The diffusion of policy diffusion research in political science[J].British Journal of Political Science,2013,43:673-701.
- [25] BERRY F S.Sizing up state policy innovation research[J].Policy Studies Journal,1994,22(3):442-456.
- [26] 杜佳.中国国家信息政策法规体系构成研究——基于“国家信息政策法规数据库”的实证分析[D].武汉:武汉大学,2004.

作者简介

裴雷, 男, 1981年生, 南京大学信息管理学院副教授, 研究方向: 信息经济与信息政策分析、信息资源规划, E-mail: plei@nju.edu.cn。
李向举, 男, 1992年生, 南京大学信息管理学院研究生, 研究方向: 信息政策分析。
谢添轩, 男, 1994年生, 南京大学信息管理学院研究生, 研究方向: 网络信息资源管理。
成吉, 女, 1993年生, 南京大学信息管理学院研究生, 研究方向: 网络信息资源管理。

Evolution of Hot Topics of Information Policy Research Throughout Different Periods in China (1986-2015)

PEI Lei, LI XiangJu, XIE TianXuan, CHENG Ji
(School of Information Management, Nanjing University, Nanjing 210093, China)

Abstract: Based on text computing and historical analyzing methodology, this paper explored the distribution laws of both journal articles and topics frequencies occurred on information policy from 1986 to 2015 in China. Based on the analysis, a quite similar tendency of topics defined by different methods occurred all through the 30 years period, and four distinct phases can be divided according to burst terms' distribution, which demonstrated the possibility and reliability of topic tracking in policy analysis and forecast.

Keywords: Information Policy; Hot Topics; Topic Tracking; Periodic Division

(收稿日期: 2016-07-03)