

● 裴 雷 (南京大学 信息管理学院, 江苏 南京 210093)

我国科学数据共享政策概念一致性与政策质量评估*

摘要: 文章构建了上下位政策概念的一致性指标测算框架, 对我国气象、地震、农业等 12 个领域的科学数据共享政策文本进行内容编码和实证分析, 讨论了当前我国科学数据共享政策在吸收、扩散和创新过程中的政策文本质量, 并提出改进建议。

关键词: 科学数据; 数据共享; 政策一致性; 质量评估

Abstract: This paper constructs a framework for the calculation of index consistency of upper and lower policy concepts, and makes content encoding and empirical analysis of the text of the scientific data sharing policy in the 12 fields such as meteorology, seismology, agriculture, and so on. The paper discusses the policy text quality of the present scientific data sharing policy in the absorption, diffusion and innovation process in China, and offers some suggestions for improvement.

Keywords: scientific data; data sharing; policy consistency; quality evaluation

政策扩散和传播中存在两个典型问题: 一类是下位类政策对上位类的承继, 一类是下位类政策对上位类政策的创新。而政策之间的承继和创新直接表现为政策文本对政策术语或政策概念的采纳或创新。换言之, 若下位政策的政策概念或术语与上位政策基本一致, 则下位政策体现较好的承继性; 反之, 则体现为创新性。而政策文本之间的政策概念一致性, 不仅包含术语采用的一致性, 也包括术语频次的一致性。

国外在立法论证、区域政策制定、大型项目规划中广泛应用“政策一致性”评估, 确保政策、计划或法律不与现行政策、法律或项目冲突, 提高其可执行力。这类政策一致性研究方法主要通过“专家”的理解, 通过对法条的逐一比对进而论证新订法律、政策、规划或方案的合理性, 是典型的内容分析法^[1]。典型的研究如欧盟委员会每两年发布的《发展政策一致性》报告, 该报告对欧盟及其成员国在各个领域实施的政策进行一致性评估^[2], 2012 年报告重点在贸易和金融、气候变化、食品安全、人口流动和安全 5 个领域进行了分析, 进而指出今后欧盟在政策设计上应更多地考虑发展的广度, 重点应放在实际利益和协同一致上。美国加利福尼亚公共事务部在《加利福尼亚州环境质量法案 (CEQA)》制定过程中, 系统采用一致性评估方法^[3], 对与其相关的区域规划、空气和水保护法案、农业保护法案等多种法案的政策目标和措施

进行一致性评估。其他如 South Lake Tahoe 市的环境评估法案则与当地已有相关政策法案中逐条进行了比对和分析, 按迁移、一致、相关、减轻、不相关、不冲突、冲突等进行逐条评估^[4]; 拉斯维加斯的土地使用规划^[5]、林肯市的公园环境管理法案^[6]、101 RANCH 特别规划案^[7]均采用类似的一致性评估方法。但尚未出现借助计算机的政策一致性自动分析方法。

1 基于政策文本的政策概念一致性测度方法

政策一致性可视为政策文本的语义相似特征, 也从一定程度反映下位政策对上位政策的吸收和承继质量以及上位政策的扩散质量。基于此, 本文设计了基于政策文本的政策概念一致性测度方法, 包括 4 个基本步骤。

1) 政策概念的萃取。将政策文本进行自然语言的分词处理, 然后比对自主开发的政策主题词表, 萃取可用的政策术语。将可用的政策术语按语义相似性进行归并, 得到政策概念序列和相应的频度属性。

2) 政策概念分区。采用类似高频词计算方法, 分别区分政策概念的高频区、中频区和低频区。高频区为频次 n 满足:

$$\left| \sum_{i=n}^{FW_{\max}} FW_i - 50\% \right| \leq \min \left\{ \left| 50\% - \sum_{i=n-1}^{FW_{\max}} FW_i \right|, \left| 50\% - \sum_{i=n+1}^{FW_{\max}} FW_i \right| \right\}$$

中频区则为频次 n 到频次 n' , 其中频次 n' 满足:

$$\left| \sum_{i=n}^{FW_{\max}} FW_i - 80\% \right| \leq \min \left\{ \left| 80\% - \sum_{i=n-1}^{FW_{\max}} FW_i \right|, \left| 80\% - \sum_{i=n+1}^{FW_{\max}} FW_i \right| \right\}$$

* 本文为教育部青年项目“基于概念统计的信息政策文本计算与实证”(项目编号: 11YJC870020) 和国家社会科学青年项目“我国信息政策转移与扩散研究”(项目编号: 12CTQ024) 的成果。

$$| 80\% - \sum_{i=n+1}^{FW_{\max}} FW_i |$$

剩余政策概念为低频区。

3) 政策概念采纳一致性。采用类似 Holie 信度测试的方法, 假定 A 政策文本在某一区间共有 N_1 个政策概念, B 政策文本在相应区间共有 N_2 个政策概念, 其中有 n 个概念为两个政策文本共有的政策概念。那么, 两个政策文本的政策概念一致性 C_{AB} 表述为:

$$C_{AB} = \frac{2n}{N_1 + N_2}$$

4) 政策概念频度分布一致性测度。通过计算上位政策与下位政策的共同政策概念频度序列的相关系数来描述政策内容的一致性, 即: $r_{AB} = \text{Cov}(F_A, F_B)$, 采用统计软件的 CORREL 函数处理。

2 政策数据采集及其文本特征

目前, 我国共有 23 个相应的科学数据共享项目^[8], 先后发布 40 项科学数据共享规章制度和方法, 81 项共享标准和规范, 263 项相关行业标准^[9]。通过中国科技资源共享网和科学数据共享工程网, 共检索到相关科学数据共享政策和标准规范 120 余件次。以《中华人民共和国科学数据共享条例》为上位类政策, 选取 12 份专项共享政策作为下位类政策用于内容分析。分别为《气象资料共享管理办法》、《测绘科学数据共享实施办法》、《林业科学数据(资料)共享管理办法》、《地震科学数据共享管理办法(试行)》、《水利科学数据共享管理办法(试行)》、《医药卫生数据共享管理细则》、《基础科学数据中心及共享服务网项目管理办法》、《中国科学院资源环境科学数据中心章程》、《材料环境腐蚀国家野外科学观测研究试验站网数据共享与管理办法》、《交通运输科学数据共享管理办法》、《农业科学数据共享管理办法》和《地质资料管理条例》。

对政策样本进行词频统计并进行语义化归并, 作为参照样本的共享条例包括 586 个词汇, 2 836 个词频; 通过语义处理后, 涵盖 472 个政策概念, 总频度 2 165。对 12 份政策文本进行分词、清洗和概念归并, 梳理其政策概念分布。剔除专业指向度较高的“科学”、“数据”和具体的领域概念, 发现高频概念集中于“共享”、“管理”、“使用”、“办法”、“服务”、“用户”、“提供”、“规定”、“汇交”、“资料”、“加工”、“活动”、“平台”、“信息”、“资源”; 以及若干表示范围和机构的高频概念, 如“单位”、“国家”、“部门”、“主管”、“机构”、“负责”、“组织”、“中心”等。目标文本与采选文本具有显著的相似性。

3 政策概念一致性分析

3.1 政策概念分区

按政策概念分区计算方法, 划分政策概念的高频区、中频区和低频区, 结果如表 1 所示。

表 1 我国科学数据共享政策的政策概念分区

文本	高频区		中频区		低频区	
	频次	数量	频次	数量	频次	数量
国家科学数据	11	25	3	140	< 3	307
气象	7	23	2	128	1	141
测绘	4	31	2	69	1	154
林业	12	15	3	50	< 3	194
地震	8	23	2	127	1	188
水文	7	37	2	139	1	226
医药	7	51	2	172	1	266
基础	6	50	2	147	1	247
环境	6	27	2	89	1	165
材料	10	36	3	123	< 3	321
交通	6	23	2	63	1	156
农业	4	26	2	63	1	153
地质	10	27	2	153	1	248

3.2 政策概念的采纳一致性

采纳一致性是指政策概念的延续和扩散, 指下位类政策完全采纳上位类政策概念的频率。对采纳一致性的测算结果如表 2 所示。

表 2 我国科学数据共享政策的政策概念采纳一致性

文本	整体采纳一致性	高频区	中频区	
		高频一致性	中频一致性	累积一致性
气象	0.283	0.375	0.144	0.247
测绘	0.317	0.250	0.137	0.317
林业	0.301	0.300	0.119	0.278
地震	0.412	0.500	0.229	0.406
水文	0.300	0.323	0.161	0.175
医药	0.307	0.395	0.150	0.289
基础	0.312	0.293	0.128	0.254
环境	0.268	0.308	0.134	0.256
材料	0.380	0.230	0.209	0.272
交通	0.323	0.375	0.091	0.287
农业	0.328	0.431	0.081	0.236
地质	0.304	0.269	0.167	0.313

从表 2 可以看到: ①在分区测算中, 累积一致性、高频一致性与整体一致性基本一致。②《地震科学数据共享管理办法(试行)》、《材料环境腐蚀国家野外科学观测研究试验站网数据共享与管理办法》和《农业科学数据共享管理办法》与《中华人民共和国科学数据共享条例》(专家建议稿)最为相似, 分布承继约 41.2%、38% 和 32.8% 的政策概念, 具有较好的政策概念覆盖面。③对高频核心政策概念, 则是《地震科学数据共享管理办法(试行)》、《农业科学数据共享管理办法》和《医药卫生

数据共享管理细则》与《中华人民共和国科学数据共享条例》(专家建议稿)最为相似。

3.3 政策概念的频度分布一致性

频度分布一致性是指政策概念重要性的显示,指上下位类政策对重要政策概念的频度差异。对频度分布一致性的测算结果如表3所示。

表3 我国科学数据共享政策的政策概念频度一致性

文本	整体一致性	高频一致性	中频一致性
气象	0.127	0.158	-0.057
测绘	0.880	0.908	0.076
林业	0.795	0.912	0.892
地震	0.916	0.925	0.151
水文	0.645	0.914	0.342
医药	0.845	0.819	0.010
基础	0.312	0.347	-0.238
环境	0.715	0.479	0.268
材料	0.705	0.936	0.232
交通	0.921	0.951	-0.065
农业	0.934	0.963	-0.447
地质	0.105	-0.351	0.075

从表3可以看到:《农业科学数据共享管理办法》、《交通运输科学数据共享管理办法》和《地震科学数据共享管理办法(试行)》与《中华人民共和国科学数据共享条例》(专家建议稿)最为相似,相同政策概念的频度分布相关系数分别达到0.934、0.921和0.916。

综合比较表2和表3,说明政策概念分区对政策概念采纳一致性和频度分布一致性的影响是显著的,发现高频区和中频区的一致性差异显著:①政策概念采纳高频一致性与中频一致性几乎完全独立,其相似相关系数仅0.0008,政策概念频度分布高频一致性与中频一致性的相似相关系数也仅0.167。②政策概念采纳整体一致性与中频概念的关联度更大(高频为0.329,中频为0.610),而政策概念频度分布一致性与高频概念关联度更大(高频为0.907,中频为0.118)。

4 我国科学数据共享政策的质量评估

4.1 我国科学数据共享政策一致性的二维尺度评估

一般认为,采纳一致性和分布一致性均较高的可视政策质量较高,说明下位政策结构完整,内容一致;反之,下位类政策质量存在偏差或不足。从政策概念采纳和政策频度两个维度进行综合分析,其结果如图1所示。

从图1可知,按上述质量标准可分为典型的三类政策:第Ⅰ类为政策内容及频度分布均与上位类政策相似的,为《地震科学数据共享管理办法(试行)》和《材料环境腐蚀国家野外科学观测研究试验站网数据共享与管理

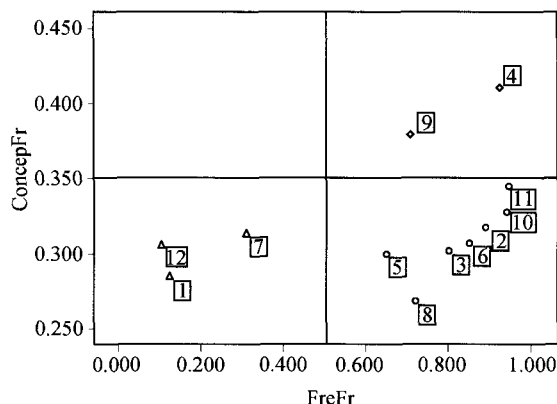


图1 政策一致性的二维尺度分析

办法》。第Ⅱ类为政策频度分布与上位类政策相似,但政策内容采纳偏低,说明下位政策与上位政策结构相似,但政策范围过于专业或狭窄,包括《农业科学数据共享管理办法》、《交通运输科学数据共享管理办法》等7部政策。第Ⅲ类为政策内容及频度分布均与上位类政策相差较大的,说明政策结构和内容均与上位政策不一致,包括《地质资料管理条例》、《基础科学数据中心及共享服务网项目管理办法》和《气象资料共享管理办法》。

4.2 我国科学数据共享政策核心内容一致性的二维尺度评估

由于下位政策应用场景不同于上位政策,其中频区的采纳一致性和频度一致性都相对较低,采选样本也支持上述假定。因此,在政策承继和扩散过程中,高频政策概念更能代表政策制定过程中上下位政策的一致性。对高频政策概念的采纳一致性和频率分布一致性进行分析,结果如图2所示。

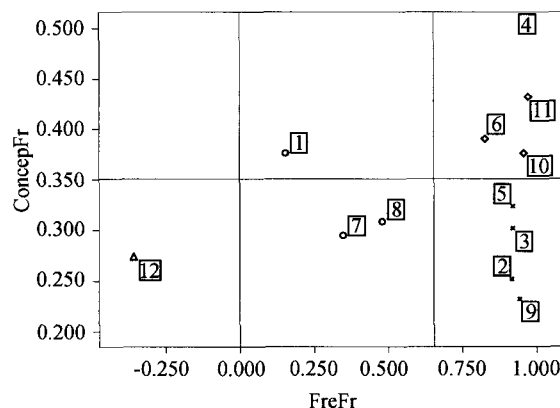


图2 高频政策概念一致性的二维尺度分析

从图2可知,对高频政策概念采纳和政策频度两个维度进行分析,则体现出4个典型分类:第Ⅰ类为高相似、高承继的政策,包括4、6、10、11号政策,即《地震科

学数据共享管理办法(试行)》和《医药卫生数据共享管理细则》、《交通运输科学数据共享管理办法》和《农业科学数据共享管理办法》。第Ⅱ类为高相似、低承继的政策,或者是比上位政策更突出专业性内涵的政策,包括2、3、5、9号政策,即《测绘科学数据共享实施办法(报批稿)》、《林业科学数据(资料)共享管理办法(报批稿)》、《水利科学数据共享管理办法(试行)》和《材料环境腐蚀国家野外科学观测研究试验站网数据共享与管理办法》等政策。第Ⅲ类为结构和内容适中,但政策结构和内容可进一步优化的政策,包括1、7、8号政策,即《气象资料共享管理办法》、《基础科学数据中心及共享服务网项目管理办法》和《中国科学院资源环境科学数据中心章程》,其中1号政策与7、8号政策略有不同,1号政策主要体现为结构,即政策概念的详略程度,后者表现为内容承继,即政策概念的覆盖面。第Ⅳ类为12号政策,距离上位政策有不小差距,需要政策更新或改进。

5 分析

政策一致性分析运用情报分析方法对政策质量和政策内涵具有一定的评估作用,通过计算化的软件方法对政策备案、政策扩散工作进行快速检评,使非专业政策和法规研究人员也能针对下位政策制定提出一定建议,对于我国立法管理和政策管理具有一定的价值。同时,政策一致性测度结果取决于因素很多,既有政策内容和结构设计的原因,也有语言描述方式的影响。但通过研究有如下基本结论。

1) 政策概念的承继主要体现于高频区,政策特性概念则集中于中频区。因而,通过高频区政策概念的测度能较好地反映两个政策之间的政策概念承继与扩散。

2) 通过政策一致性的测算,对政策改进和完善具有一定指导。因采纳整体一致性与中频概念的关联度更大,频度分布一致性与高频概念关联度更大,故当采纳一致性偏低,则意味着需调整政策结构,优化中频区的政策概念结构;当频度分布一致性偏低,则意味着需突出政策重点,加强政策概念高频区与上位政策的一致性。

3) 二维尺度测算提供了相对可行的政策文本评估方法,分类具有一定参考价值。为进一步测试上述方法的可信度,笔者随机找了10位研究生对上述政策效果进行一致性的5分制评定,结果显示地震、材料、农业、交通平均得分在4.5分以上,地质得分在3分左右,与评测结果基本吻合。从政策文本的发生看,也与评测结果基本一致:12号政策是在科学数据共享项目启动前就制定的政策,1号政策则是项目开展早期制定的开创性政策,并未考虑完整的政策范围。而国家法案的起草则有地震局李志

雄等专家的参与,因而与地震4号政策具有较高的一致性。所以,从政策概念一致性评测角度研究政策文本质量,具有一定可信度。

同时,针对我国科学数据共享政策评估结果,笔者建议:2、3、4、5、6、9、10和11号政策是相对质量较高的政策文本,尤其是《地震科学数据共享管理办法(试行)》和《材料环境腐蚀国家野外科学观测研究试验站网数据共享与管理办法》,不论从结构上还是内容上,与《中华人民共和国科学数据共享条例》保持了较高的一致性;而1、7、8号政策,即《气象资料共享管理办法》、《基础科学数据中心及共享服务网项目管理办法》和《中国科学院资源环境科学数据中心章程》还有一定的改进空间,其中1号政策主要体现为政策结构需要优化,即政策概念的详略程度,7号和8号表现为内容承继,即政策概念的覆盖面还不够,尤其对数据汇交和有偿使用方面关注还不够;而急需升级改造的是12号政策,即《地质资料管理条例》,与共享政策不论从内涵还是结构上均不一致。

本文的分类分区主要以欧拉距离作为聚类的参量,图中的参照线仅仅是经验获得,因此在评定优质政策文本的一致性指标的标准值或基准值目前并没有确定,有待今后更多经验研究的积累。□

参考文献

- [1] NUTTALL S. "Consistency" and the CFSP: a categorization and its consequences [M] //EFPU Working Paper 2001/3. London: London School of Economics and Political Science, 2001.
- [2] GASPERS J. The quest for European foreign policy consistency and the treaty of Lisbon [J]. Humanitas Journal of European Studies, 2008, 2 (1): 19-53.
- [3] ftp://ftp.epuc.ca.gov/gopher-data/enviro/tehachapi_renewables/FINALPolicy_ALL.pdf.
- [4] http://www.trpa.org/RPUEISComments/1_Public%20Agencies/City%20of%20South%20Lake%20Tahoe.pdf.
- [5] <http://sbcountyplanning.org/projects/05TPM-00002/Documents/DEIR/TOC.pdf>.
- [6] <http://www.ci.lincoln.ca.us>.
- [7] <ftp://ftp.co.imperial.ca.us/icpds/specific-plan/101-ranch/09consistency.pdf>.
- [8] 路鹏,苗良田,李志雄,等.我国科学数据共享现状的调查与分析[J].地震,2007(7):125-130.
- [9] 刘润达,彭洁.我国科学数据共享政策法规建设现状与展望[J].科技管理研究,2010(13):40-45.

作者简介:裴雷,男,1981年生,博士,副教授。

收稿日期:2013-02-26