

# 基于WordScore原理的信息政策 价值评价模型与方法\*

□ 裴雷 / 南京大学信息管理系 南京 210093

孙建军 / 南京大学信息产业研究所 210093

**摘要:** 文章基于WordScore的基本原理和信息政策价值分类整合方法, 构建了政策价值TESCO模型及词表, 并以工信部和科技部的信息政策为样本, 将政策文本进行语词切分, 按照语词的价值属性建立“语词—政策价值”映射, 从而提取文本的政策价值取向, 并确定该文本或该领域的政策价值结构, 比较不同领域的政策价值差异。

**关键词:** 信息政策, 政策价值评价, WordScore, 词表, 词频

DOI: 10.3772/j.issn.1673-2286.2011.08.013

## 1 研究背景

信息政策是国家引导和调控信息工作的主要手段, 代表了信息技术和服务的战略需求与社会规范, 但其研究往往涉及整个社会经济系统, 并且研究者的主观意图和立场可能影响其学术观点, 因而建立客观的信息政策研究框架和方法一直是西方科学主义学者关注的领域。而政策文本相对具有规范性和相对独立性, 政见研究和政策分析研究者从上个世纪60年代开始就尝试运用科学计算和定量评估的方法, 建立客观政策分析框架, 并应用于政党分析和领域政策评价。

随着计算机技术的引入, 政策文本分析所能处理的素材量和处理精度得到了大幅提升。本世纪初, Michael Laver、Kenneth

Benoit和Will Lowe等提出的政策文本计算分析框架<sup>[1-3]</sup>, 广泛采纳了WordFrequency、WordScore、Wordfish、Text Encoding、Data Extraction等量化工具和方法, 使得政策文本分析方法得以成熟和完善, 并逐渐从政见分析拓展到具体政策分析领域。

在信息政策领域, 国内外虽然还没有完全量化的信息政策文本分析, 但信息政策文本的研究价值以及非纯人文的信息政策研究方法得以采纳和发展。同时, 目前信息政策文本分析的外部环境也已经成熟。第一, 信息政策与立法活动的发展, 使得可被分析的文本极为丰富, 可获取性也显著增加。仅以中国为例, 在国务院法制办登记的法律为57393件, 其中信息政策法律281件<sup>[4]</sup>。同时, 《政府信息公开条例》实

施以来, 各部委网站都公开了大量的政策文本, 其中也不乏信息政策文本。第二, 全球协作活动使得具体领域的政策比较和借鉴研究非常活跃。在我国近十年的信息政策研究中, 约有30%的文献为引介或比较研究, 为政策文本分析提供了广阔的比较应用空间<sup>[5]</sup>。第三, 在信息政策研究中, 学者群也自发引入对应分析、政策编码、社会网络等新的定量方法, 营造了政策分析的科学主义氛围。

此外, 目前国内资源政策和教育政策研究中也学者采纳该方法。基于上述背景, 本文采用WordScore的基本原理和信息政策价值分类整合方法, 以信息政策文本作为考察对象, 对工信部和科技部信息政策文本进行了探索性研究。

\* 本文系江苏省高校哲学社会科学基金重点项目“两化融合战略下江苏省信息资源开发利用的战略、思路、模式、途径与政策研究”(项目编号: 2010ZDXXM021)和江苏省教育厅哲学社会科学基金项目“基于概念统计的信息政策文本计算与实证”(项目编号: 2011SDJ870004)项目成果。

## 2 研究概念界定与处理

### 2.1 信息政策文本

从概念界定看, 信息政策文本所涉及的信息政策和文本两个概念并没有准确的界定方式。Browne就认为<sup>[6]</sup>, 信息政策是一个宽泛、多元和缺乏统一的范畴; Hernon和Relyea也注意到信息政策 (Information Policy) 常用单数形式, 而复数形式的信息政策 (Information Policies) 多探讨信息政策领域的具体问题<sup>[7]</sup>。因此信息政策的概念在研究过程中是不完整的、相互重叠的, 甚至是相互矛盾的。不仅如此, 在形式上有学者将政策界定为政府或公共机构发布的标准文档, 有学者将其与法律法规、标准规范和政府指令区分对待。而事实上, 信息政策存在于两个层面: 一种是明确的, 以文献形式记录下来; 另一种是不明确的, 以惯例形式表示出来, 如非文字的行为规范、预期和社会准则等。因此要准确界定“信息政策文本”非常困难, 本研究选择了由中华人民共和国全国人民代表大会、国务院以及各部委、省级政府或人民代表大会签发的规范法律、规章和指令作为分析样本, 以此获得国务院法制办登记在案的281份信息政策作为分析样本。

### 2.2 信息政策价值

按照《公共政策词典》的解释, 政策价值 (values) 是指“政策制定者以及其他涉及决策过程的人共有的偏好, 个人愿望

和目标; 价值观可能包括一个人的政治信条、个人偏好、组织目标以及政策取向”。因而, 信息政策价值作为政策系统行为的选择准则, 并非文本中语词或概念的单纯汇集。体现在研究方法上, 一种是通过专家意见调查或相关开放调查方法, 建立先验性政策分类词表 (prior policy vocabulary); 一种是通过语义聚类, 生成概念分层图 (policy conception hierarchy), 进而抽取出政策域。

本研究选择元信息政策价值 (meta-value) 的分析思路, 针对已有的信息政策词表和信息政策价值研究文献进行了梳理和归纳, 并对对应政策样本集进行映射和修正, 得到较为合理的信息政策价值分析模型。详细方法和结果参见第四节。

## 3 研究方法步骤

### 3.1 WordScore的基本思想

WordScore是由Laver、Benoit和Garry提出的一种基于文本价值倾向评估的加权评价方法<sup>[1]</sup>。其基本思想为:

①选定参考文本集 $R=\{r_1, r_2, \dots, r_n\}$ 。参考文本集为以往研究过的具有良好政策倾向定义的文库, 并给定不同参考文本 $r$ 在不同的政策领域 $d$ 一个参照的先验权值, 反映该文本在政策领域的权值。

②统计参考文本 $r$ 中每一个词语 $w$ 的分布频率 $F_{wr}$ , 即参考文本的词频分布。

③计算词语 $w$ 的条件分布概率 $F_{wr}$ , 计算方法为:  $F_{wr} = \frac{F_{wr}}{\sum_r F_{wr}}$ ,

即词语 $w$ 参考文本集 $R$ 中分布的条件概率。

④计算词语 $w$ 在政策领域 $d$ 的价值加权 $S_{wd}$ , 计算方法为:  $S_{wd} = \sum_r (P_{wr} \times A_{rd})$ , 为算术相乘的求和, 反映的是词语 $w$ 在政策领域 $d$ 的平均价值表现力。

⑤计算测试文本 $v$ 的词频分布 $F_{wv}$ 。

⑥计算测试文本 $v$ 在政策领域 $d$ 的价值得分 $S_{vd}$ ; 计算方法为:  $S_{vd} = \sum_w (F_{wv} \times S_{wd})$ 。

⑦评价测试文本 $v$ 中词语 $w$ 的政策价值变化, 测度其价值偏差方 $V_{vd}$ 与标准差 $S_{vd}^*$ 。计算方法为:

$$V_{vd} = \sum_w F_{wv} (S_{wd} - S_{vd})^2$$

$$S_{vd}^* = \sqrt{\frac{V_{vd}}{N_v}}$$

⑧逐一测算测试文本集 $V$ 中的政策价值得分, 进而求得其总的加权平均值。

### 3.2 改进与研究步骤

WordScore提供了一种定量研究词语—政策价值的映射方法, 能够区分出政策文本中每一个词汇的政策表现力, 即词汇的政策价值得分。在本研究中, 直接使用WordScore还有如下不足: 第一, 信息政策没有前导研究和先验权值, 无法估计先验权值 $A_{rd}$ 的分布, 必须引入主观干预的方法; 第二, 信息政策的政策领域 $d$ 的分布尚不明确, 需要先构建信息政策价值模型, 以区分政策领域 $d$ ; 第三, WordScore的测试样本一般在2000个用词, 总词量在50000以内, 大样本中需改进相关权值计算方法。

因而, 我们运用WordScore对词汇的政策价值表现力计量的思想, 但不考虑词语在政策领域价值的表现差异, 即不考虑先验权值 $A_{id}$ 的分布, 均为1, 那么可以直接用词频和专家小组赋值的方法进行价值评价和比较。依上述思想, 主要研究步骤设计如下:

①运用元信息政策价值分析方法构建信息政策价值模型, 确立信息政策领域 $d$ 的分布。

②建立样本集中文本的高频词语 $w$ 与信息政策价值 $d$ 之间的映射关系:

$M(w_i, d_k): w_i \xrightarrow{\text{mapping}} d_k | i \in [1, N_r]$ 。其中,  $N_r$ 为信息文本集 $R$ 中词语的总数量。  $M(w_i, d_k)$ 为一个二值函数, 其定义如下:

$$M_{wd} = M(w_i, d_k) = \begin{cases} 0, & \text{当 } w_i \text{ 不能表述 } d_k \text{ 价值时;} \\ 1, & \text{当 } w_i \text{ 能表述 } d_k \text{ 价值时。} \end{cases}$$

③运用对 $M(w_i, d_k)$ 赋值的方法, 测试信息政策价值模型可用性, 并设定为参考文本集 $R$ 。

④词语 $W$ 在政策领域 $d$ 的价值加权 $S_{wd}$ 由词语 $W$ 其在文本集中的赋值函数 $M(w_i, d_k)$ 累积值来表示, 即:  $S_{wd} = \sum_r (M_{wd} * F_{wr})$ 。

⑤计算参考文本集 $R$ 在政策领域 $d$ 的权值得分 $S_{md} = \sum_w S_{wd}$ 。

⑥计算测试文本 $v$ 的高频词语的词频分布 $F_{wv}$ , 统计测试文本 $v$ 在政策领域 $d$ 的价值得分 $S_{vd}$ 。

⑦测算文本集 $V$ 在政策领域 $d$ 的价值加权 $S_{md}$ , 为文本集中所有词汇的价值加总。

## 4 信息政策价值分析模型

信息政策价值分析模型是借鉴

已有信息政策价值分析和信息政策分类方式构建的一种较为均衡的政策价值分类词表。

### 4.1 信息政策价值分类整合

信息政策框架体系和信息政策分类在国内外一致比较活跃, 典型研究如Browne<sup>[6]</sup>、Hernon<sup>[7]</sup>、Burger<sup>[8]</sup>、Hill<sup>[9]</sup>、Moore<sup>[10]</sup>、Trauth<sup>[11]</sup>、Overman和Cahill<sup>[12]</sup>、Rowlands<sup>[13]</sup>、Orna<sup>[14]</sup>、Yusof<sup>[15]</sup>等的研究; 国内从1990年开始, 卢泰宏、蒋坡、朱庆华、周庆山、马海群、梁俊兰、罗曼、查先进、郑海燕、燕金武、杜佳等学者都撰写过相关专著, 探讨过信息政策框架体系和分类问题。从研究框架看, 学者多倾向从信息政策中产生的问题或活动要素划分(如表1)。

表1 信息政策分类框架与体系研究文献

研究者(时间)/模型	信息政策划分依据	分类体系
Lamberton (1974) 信息政策分类体系	信息政策活动领域和对象	科学技术信息政策; 社会科学信息政策; 服务信息政策; 信息技术需求政策 (in Browne <sup>[6]</sup> )
Karni (1983) <sup>[8]</sup> 信息政策范围/要素矩阵	信息政策的活动要素或对象	信息政策目标; 决定如何实现目标的方法; 政策实施、作用和变通方法; 信息服务系统
Trauth (1986) <sup>[11]</sup> IPO模型和信息政策研究框架	信息政策处理的时间轴和抽象层次	信息导入政策; 信息处理政策; 信息导出政策
Chartrand (1989) <sup>[16]</sup> 和 Milevki (1986) <sup>[17]</sup> 信息政策 分类体系	第95-98次美国国会信息政策立法工作及备忘录的内容分析	9类政策: 政府信息资源管理政策; 信息技术政策; 电信和广播政策; 国际通信政策; 信息披露、机密和隐私; 计算机管制和计算机犯罪; 知识产权; 图书和档案政策; 政府信息分发管理
Overman和Cahill (1990) <sup>[12]</sup> 信息政策价值模型	16份联邦信息政策的文本内容分析	7种信息政策价值: 信息获取与信息自由; 隐私权; 信息公开; 信息利用; 成本和收益; 保密和安全; 信息产权
Galvin (1992) <sup>[18]</sup> 信息政策价值观, 信息政策分类法		信息获取权、信息所有权和隐私权。以信息政策价值观为基础, 划分了价值权利实施、价值外部冲突解决和价值内部解决的三大类信息政策
Moore (1993) <sup>[10]</sup> 信息政策矩阵	信息政策内容体系和作用层次	产业、组织和社会3列, 信息技术、信息市场、信息工程、人力资源、法律法规5行



表1 信息政策分类框架与体系研究文献 (续)

研究者 (时间) /模型	信息政策划分依据	分类体系
Burger (1993) <sup>[8]</sup> 信息政策层次等级模型 Information policy hierarchy		①基础设施政策 (Infrastructural Information Policy); ②水平信息政策 (Horizontal Information Policy); ③垂直信息政策 (Vertical Information Policy)
Rowlands (1996) <sup>[13]</sup> 问题和选择性方法	波拉特价值政策框架 issues and options approach	第一部类信息政策、第二部类信息政策
Browne (1997) <sup>[6]</sup> 信息政策价值批判研究方法	信息政策的核心价值分类	技术性价值; 经济价值; 业务 (occupational) 价值; 空间价值; 文化价值
Hill (1994) <sup>[9]</sup> 信息政策问题类型	机构资源类型、文本类型、政策目标	政府对信息的获取, 政府中的信息管理, 政府对信皂和通信技术的利用, 信息技术产业, 电信和网络, 信息经济, 工业、农业和商业, 科学和技术信息, 信息产业, 图书和档案, 公共部门和私营部门问题, 向公众发布官方信息等。
Nilsen (2000) 信息通信政策模型		信息过滤、信息传播、版权、信息自由、政府信息、产业信息、文献、隐私权、科技信息 (in Browne <sup>[6]</sup> )
Orna (2008) <sup>[14]</sup> 信息政策问题空间	信息政策的问题导向, 分别归纳了国家信息政策层面的6项议题和组织层面的4项议题	国家信息政策、组织信息政策和国家机构共存信息政策
Yusof etc. (2010) <sup>[15]</sup> 信息政策问题分类	整合已有的信息政策分类研究	6类核心价值: 科技信息、图书馆事务、信息通信技术、信息社会、政府信息和经济价值
朱莉和朱庆华 (2003) <sup>[5]</sup> 信息政策研究问题分类	18篇文献综述总结	①信息资源保障政策; ②信息产业发展政策; ③信息交流与合作政策; ④通讯、广播政策; ⑤信息安全、保密政策
周庆山 (1995) <sup>[19]</sup> 国家信息法律体系	影响国家信息法体系形成的政治、经济、社会和法律因素	公民信息自由权法、商业信息保护法、大众传播与公共信息立法、政府信息保护法、信息产权法、网络信息法和国际信息法等
黄纯元 (1998) <sup>[20]</sup> 国家信息政策体系结构、信息政策问题群	现代社会宏观信息结构	国家信息政策体系结构由信息通讯、信息平台、信息应用三个基本部分所涵盖的信息政策问题群
马费成 (2000) <sup>[21]</sup> 信息政策内容框架	问题导向	信息资源管理、信息技术和信息产业发展、信息市场管理、知识产权保护、信息安全与计算机犯罪、信息利益分配和“信息公平”、信息国际化等问题
马费成 (2002)		信息内容政策、信息传输的组织管理和控制政策、技术政策、人才政策、投资政策、系统政策
于清文 (2002) <sup>[22]</sup>	信息流程	信息创造和生产政策、信息分配政策、信息交流政策以及信息消费政策
查先进 (2002) <sup>[23]</sup>	改进的Moore矩阵	信息技术、信息网络、信息市场、信息资源、信息人才和信息环境
杜佳 (2004) <sup>[23]</sup>	信息综合立法框架	信息资源、知识产权、信息产业、信息市场、信息技术、信息网络、信息安全、信息标准、信息人才、信息机构
燕金武 (2006)		“7P”信息政策框架 (Privacy, Piracy, Pornography, Pricing, Policing, Psychology, Protection)



此外,在政策分类中,MRG提出从外交关系(含军事)、自由和民主、政治系统、经济、福利和生活质量、社会结构和社会群体7个领域57个二级价值变量进行测定;涂端午在分析教育

政策时,引入了经济、权力、知识、技术、福利、专有称谓、名誉、意识形态和社会目标9个变量<sup>[25]</sup>。综合上述研究,技术价值(T)、经济价值(E)、社会价值(S)、文化价值(C)以及与

信息活动本身相关的价值(O)是涉及面最广泛的价值类型,因而提出相应的价值模型TESCO模型,并按一般的体系分类法,建立下述实验模型。

表2 TESCO信息政策价值模型

价值域 (Domain)	价值类 (Class)	价值基因词 (generic word)
技术价值 Technological	信息安全	密码、数字隐藏、可信网络、数字签名、认证、防火墙、病毒、木马……
	信息治理	IT标准、IT规划、架构、模块、嵌入……
	信息系统与应用	操作系统、管理信息系统、邮件、聊天室……
	信息传输与网络	网络、服务器、路由器、交换机、中继器、无线网、移动通信、无线电频率……
	信息基础设施	服务器、计算机、硬件、存储设备……
	数据和信息	数据库、存储系统、元数据、信息包……
	其他	云计算、网格、界面……
经济价值 Economic	信息市场管理	信息监管、信息机构、信息中介、经济信息交易、信息经纪人、客户隐私……
	信息产业	电信服务、出版、传媒、IT外包、咨询、电影、动漫、电子商务……
	金融信息与统计	证券信息披露、电子商务、网络银行……
	信息产品	数字产品、网络广告、数据库、即时通讯、网络搜索……
	其他	
社会价值 Social	信息化与信息社会	信息化、信息融合、信息社会、社会阶段、两化融合、信息资源开发、后工业化社会、信息环境政策……
	信息公平	数字鸿沟、数字桥接、均衡、信息承载力、信息生态……
文化和法律价值 Cultural&Law	知识产权	知识产权、著作权、商标权、专利权……
	知情权与隐私权	商业秘密、反不正当竞争、隐私、知情权、泄密……
	文化信息共享	文化信息共享、文化遗产、作品……
	图书馆和档案管理	图书馆、呈缴、档案馆……
	其他	信息福利、数字文化生活、博物馆、计算机犯罪……
信息活动价值 Occupational	信息开发	信息资源开发、信息采集、信息生成、学术文献、系统规划、需求分析、程序开发、系统采纳……
	信息组织	分类、标引、序化、标签……
	信息传播	信息公开、信息发布、上传、下载、转载、传输、馆际互借、共享、流通、获取、隐私、保密……
	信息利用	搜索、链接、引用、再利用、合作交流、侵权、评论……
	信息存储	档案管理、存储设备、应急管理、备案……
	信息回收或消亡	信息回收、归档、剔旧、替代、迁移……
	其他	信息组织创新、电子政务、电子商务、信息评估……

注:关于TESCO词表的设计和生成为词表原型的反复迭代过程,具体将另文介绍。



## 4.2 信息政策价值分类 模型TESCO的可用性评价

筛选Overman和Cahill 模型、

Browne模型和Yusof模型作为参照模型, 并从国务院法制办获取281份信息政策全文。通过反复的分词和设置过滤词, 最后将该文本集提

取出587,063个词语或短语, 其中有11,136个不同的词汇, 其中大多数为双字词, 见表3。

表3 信息政策文本样本集

文本数量	字符数	词语数量	平均每个词语 字数	词汇使用量
281	1,368,404	587,063	2.33	11,136
时间分布	1995年以前文 本数	1996-2000年文 本数	2001-2005年文 本数	2006-2010年文 本数
	17	52	105	107

对切分出的11,136个词汇筛选出4953个高频词汇(>5次), 并剔除无意义词语后得到1963个

实词, 累计词频203,563。采用编码小组对1963个实词按照不同的价值领域进行映射 $M(w_i, d_k)$ 的赋

值。在意见基本一致的情况下得到如下词频分布表4。

表4 信息政策价值的分布

N=203,563

模型与价值域 (Domain)	词数 ( $N_d$ )	累积词频得分 ( $\sum_r \sum_w M_{wd}$ )	政策价值加权 ( $S_{md}$ )	政策价值平均值 ( $S_{sd}$ ) $\times 1000$
Overman and Cahill模型	获取和自由价值	105	4957	0.024351
	隐私权和信息产权	24	1892	0.009295
	信息开放	61	13417	0.065911
	信息利用	180	6644	0.032639
	成本和收益	123	2429	0.011933
Browne模型	保密和安全	54	2797	0.01374
	技术价值	234	27602	0.135596
	经济价值	817	100940	0.495871
	信息活动价值	302	35719	0.175471
	空间价值	61	1872	0.009196
Yusof模型	文化价值	152	17681	0.086858
	科技信息	246	28934	0.142139
	图书馆	41	2793	0.013721
	ICT	107	37664	0.185026
	社会事务	112	14888	0.073138
TESCO模型	政府信息	121	27392	0.134564
	经济	817	100940	0.495871
	技术价值	234	27602	0.135596
	经济价值	817	100940	0.495871
	信息活动价值	302	35719	0.175471
TESCO模型	社会价值	112	14888	0.073138
	文化和法律价值	162	24799	0.121826
	文化和法律价值	162	24799	0.121826



由于在赋值过程中存在一词多值，或者一词不属于任何领域的情况，故而累积词数并不等于1963，累积词频也不等于总词频数203,563。从价值分类的词频分布看，Overman和Cahill模型分类过于具体，现有政策文本中大量高频词

汇赋值均为0，忽略了大量有用信息；Browne模型和Yusof模型基于我国信息政策而言，分布不均衡，显示价值层次不够合理。体现在词汇平均政策价值表现力测度上，TESCO模型更加均衡，词间差异更小（如表5）。

而按TESCO模型所揭示的信息政策价值分布，可知信息政策的经济价值是现有信息政策体系的主要关注领域，而社会价值关注度还不够（如图1所示）。

5 工信部和科技部信息政策价值萃取与比较

5.1 信息政策测试文本集的获取

分别从工信部门户网站和科技部门户网站采集信息政策文本47份和27份，进行词语分词处理，并逐步对过滤词进行过滤，得到测试文本集（如表6）。

5.2 信息政策测试文本集的处理

将工信部政策文本集切分筛选出的高频词语剔除无意义词语后，进行分类和赋值，得到如表7所示的词语加权得分表。

5.3 信息政策测试文本集的测试结果与比较

采用同样的方法，分别对各自的信息政策文本中的词语进行赋值评价和加权评价，最终计算各部委的政策价值领域分布（如表8），其中， $S_{md} = \sum \sqrt{S_{vs}}$ 。

表5 TESCO模型可用性评价

	均值	标准差
Overman and Cahill模型	3.721	3.285
Browne模型	4.979	1.740
Yusof模型	8.356	4.615
TESCO模型	6.345	0.645

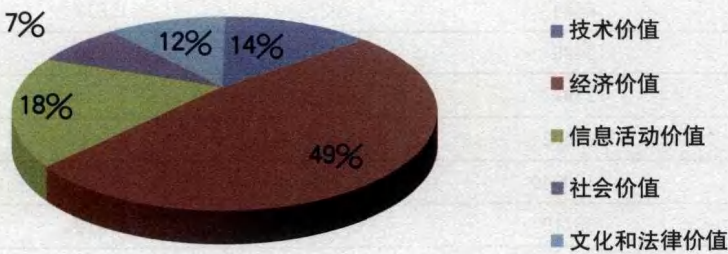


图1 我国信息政策价值分布

表6 信息政策测试文本集特征

	政策文本数	字符数	累积词频数	词语数量	高频词数量 (n>3)
工信部	47	469,173	94,817	4,346	2361
科技部	27	270,286	51,036	3,659	1826



表7 工信部信息政策测试文本集的词语价值加权得分表 (局部, Top20)

词语	$F_{wv}$	$S_{vf}$	$S_{vc}$	$S_{vs}$	$S_{ve}$	$S_{vo}$
电信	1577	0.018238	0.018238	0.000000	0.000000	0.018238
服务	1303	0.000000	0.015068	0.000000	0.000000	0.015068
业务	1136	0.012305	0.000000	0.000000	0.000000	0.000000
应当	1064	0.000000	0.010824	0.000000	0.010824	0.000000
管理	936	0.000000	0.010638	0.000000	0.000000	0.010638
规定	920	0.000000	0.000000	0.000000	0.010166	0.000000
通信	879	0.009497	0.000000	0.000000	0.000000	0.000000
机构	821	0.000000	0.000000	0.000000	0.000000	0.008072
信息产业	698	0.000000	0.007830	0.000000	0.000000	0.000000
部门	677	0.000000	0.000000	0.000000	0.000000	0.006799
电子	588	0.006635	0.000000	0.000000	0.000000	0.000000
单位	574	0.000000	0.000000	0.000000	0.000000	0.006470
互联网	559	0.006240	0.000000	0.006240	0.000000	0.006240
主管	539	0.000000	0.000000	0.000000	0.006119	0.006119
经营者	529	0.000000	0.004592	0.000000	0.000000	0.000000
网络	397	0.004496	0.000000	0.004496	0.000000	0.004496
域名	389	0.001974	0.000000	0.000000	0.000000	0.000000
经营	389	0.000000	0.028360	0.000000	0.000000	0.000000
标准	388	0.000000	0.000000	0.000000	0.004496	0.000000
国家	381	0.000000	0.000000	0.000000	0.000000	0.004488
.....	...	...	...	...	...	...
$\Sigma$	86458	$S_{vf}$	$S_{vc}$	$S_{vs}$	$S_{ve}$	$S_{vo}$

表8 工信部和科技部基于TEOSC模型的价值分布

		技术价值	经济价值	信息活动价值	社会价值	文化和法律价值
工信部	词语数量	340	333	631	255	257
	累积词频	12,364	12,585	18,314	6,096	9,867
	$S_{mt}$	0.14300	0.14556	0.21182	0.07051	0.11412
科技部	词语数量	243	378	654	258	327
	累积词频	4,913	5,483	9,096	4,571	6,344
	$S_{mt}$	0.13941	0.15558	0.25810	0.12970	0.18001



表8揭示：在自身价值结构中（如图2、图3所示），信息活动价值在两部委信息政策价值结构中所占比例均最大，而对比分析结果显示科技部信息政策对社会价值、文化和法律价值的关注度（36%）显著高于工信部（27%）；但工信部信息政策对技术价值和经济价值的关注度（42%）则高于科技部（34%）。

在词频统计过程中也发现：“机构”、“应当”、“国家”、“管理”等词汇均为两部委共同关注的高频词汇，而且两部委信息政策的主体围绕信息活动展开，并在“信息资源开发”、“信息化”等用词上具有相似的集中性。上述词频分布现象表明，两部委在信息化和信息资源开发等政策领域具有良好的政策合作空间。同时，在“电信”、“服务”、“发明”、“专利”等词频分布上的显著差异则揭示出二者今后协商问题的集中领域。

## 6 结论与展望

在以往政策定量分析中，多以政策文本或政策某一篇文章作为政策标引和分析的主体。但政策文本或

篇章往往具有多种价值属性，仅使用政策的“显示性”价值属性来解读和分析政策趋势容易忽视其他价值。WordScore及其代表的政策计算在能从最小的政策概念单元——词语的角度分析政策价值，提高了政策分析的精度和一致性，提供了一种从价值抽取和统计分析视角研究政策趋势的方法。同时，政策计算不能充分考虑到政策的社会背

景，可能出现政策的误读，也被国外一些政策分析学者所批评。

具体到本研究，通过整合分析创建了信息政策价值基本模型并选取了工信部和科技部的信息政策文本进行比较分析，达到了预期研究目的。如果能进一步引入词语 $W$ 在政策价值领域 $d$ 中的先验概率 $A_{wd}$ ，并挖掘政策文本的特征词，分析结果可能更加理想。

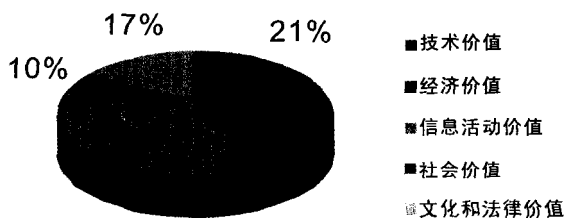


图2 工信部信息政策价值分布图

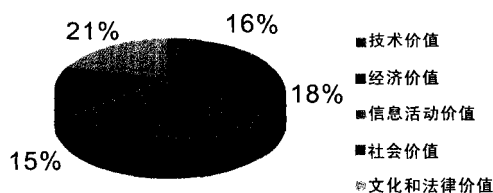


图3 科技部信息政策价值分布图



## 参考文献

- [1] LAVER, MICHEAL, GARRY J. Estimating policy positions from political texts[J]. American Journal of Political Science, 2000(44): 619-634.
- [2] LAVER, MICHEAL, BENOIT K, GARRY J. Placing political parties in policy spaces[J]. Unpublished paper, Trinity College Dublin, 2002.
- [3] Benoit K, LAVER M. Estimating party policy positions: comparing expert surveys and hand-coded content analysis[J]. Electoral Studies, 2007, 26 (1): 90-107.
- [4] 国务院法制办法律法规全文检索系统[OL]. [2011-02-24]. <http://search.chinalaw.gov.cn/search2.html>.
- [5] 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(4): 371-375.
- [6] BROWNE M. The field of information policy: 2. Redefining the boundaries and methodologies [J]. Journal of Information Science, 1997, 23(5): 339-351.
- [7] HERNON P, RELYEA H C. Information policy [M]// Drake M A. Encyclopedia of Library and Information Science. New York: Marcel Dekker, Inc., 2003: 1300-1315.
- [8] BURGER R H. Information Policy: A Framework for Evaluation and Policy Research[M]. New York: Ablex Pub. Corp., 1993.
- [9] HILL M W. National information policies and strategies: Overview and bibliographic survey[J]. West Sussex: Bowker Saur Limited., 1994.
- [10] MOORE N. Policy issues in the multimedia age[J]. Journal of Information Science. 1996, 22(3): 213-18.
- [11] TRAUTH E M. An integrative approach to information policy research[J]. Telecommunications Policy, 1986, 10(1): 41-50.
- [12] OVERMAN S E, CAHILL A G. Information policies: A study of values in the policy process[J]. Policy Studies Review, 1990, 9(4): 803-818.
- [13] ROWLANDS I. Understanding information policy: concepts, frameworks and research tools[J]. Journal of Information Science, 1996, 22(1): 13-25.
- [14] ORNA E. Information policies: Yesterday, today, tomorrow[J]. Journal of Information Science, 2008, 34(4): 547-565.
- [15] ZAWIYAH M. Yusof, Mokmin Basri and Nor Azan M. Zin. Classification of issues underlying the development of information policy[J]. Information Development, 2010(26): 204-213
- [16] Charand. Legislating Information Policy[R]. Bulletin of the American Society for Information Science, 1986(12).
- [17] MILEVSKI. Information Policy. Through Public Laws of the 95th—98th Congresses[C]//Proceedings of the American Society for Information Science Annual Meeting, 1986: 211-219.
- [18] GALVIN T J. Rights in conflict: public policy in an information age[M]//New worlds in information and documentation: Proceedings of the 46th FID Conference and Congress. The Netherlands: FID, 1994: 59-66.
- [19] 张宇文, 周庆山. 信息法学[M]. 北京: 法律出版社, 1995.
- [20] 黄纯元. 信息政策体系结构[J]. 情报资料工作, 1998(5): 8-13.
- [21] 陈亮, 马费成, 汪斌. 简论国家信息政策与法律法规体系[J]. 情报学报, 2002, 10: 554-560.
- [22] 于清文. 国家信息政策与立法[C]. 信息化与信息资源管理会议论文集. [出版者不详], 2002.
- [23] 查先进. 信息政策与法规[M]. 北京: 科学出版社, 2004: 8.
- [24] 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, 2005.
- [25] 涂端午. 政策生产: 价值的权威控制及其演变——1979~1998年中国高等教育政策文本分析[D]. 北京: 北京大学, 2009.

## 作者简介

裴雷 (1981-), 南京大学信息管理学讲师, 研究方向为信息政策与信息资源规划. E-mail: plei@nju.edu.cn

孙建军 (1962-), 南京大学信息管理学教授, 博士生导师, 研究方向为网络信息资源管理与计量学. E-mail: sjj@nju.edu.cn

## Evaluation Model of Information Policy Values Based on WordScore Theory: A Comparative Analysis of MIIT and MOST Information Policies

Pei Lei, Sun Jianjun / Nanjing University, Nanjing, 210093

**Abstract:** Taking policy documents as materials, this paper advocated a quantitative analysis method by word frequency calculating. Creating a information policy value model TESCO and its vocabulary, the paper took MIIT and MOST as examples, collected all information policies and divided the policy document into words and words, calculated the word frequency to mining potential rules. Then it mapped the words with value vocabulary, and examined the common values and differences. At last, it gave some advice to both information policy development and analysis methodology acceptance.

**Keywords:** Information policy, Policy values, Wordfrequency, WordScore, Vocabulary