

可视化视域下我国科普政策文本内容挖掘分析

孔德意 *

(河南工程学院人文社会科学学院, 郑州 451191)

[摘要] 为了挖掘出科普政策文本的核心内容及其存在问题, 本研究利用 ROSTCM 词频分析软件对科普政策文本关键词进行了提取, 构建了关键词共词矩阵、相关矩阵和相异矩阵, 之后运用 Ucinet 和 SPSS 软件对关键词进行了聚类分析、多维尺度分析和社会网络分析。从聚类分析得出, 科普政策内容关键词可分为八大类, 分别为青少年和农民的培训和指导、科普资源共享、基础设施建设与经费投入等, 表明科普政策内容涉及的主题广泛。在多维尺度分析中, 将我国科普政策内容分为四大词团, 我国科普政策内容聚焦点清晰可见。对关键词进行了网络密度分析得出, 科普政策内容设计存在缺陷。最后提出加强政策内容的层次性、提高政策内容的系统性等对策建议。

[关键词] 可视化 科普政策 文本内容挖掘 共词分析

[中图分类号] G322.0 **[文献标识码]** A **[DOI]** 10.19293/j.cnki.1673-8357.2018.03.002

本研究在对科普政策进行系统梳理和深入研读的基础上, 以 1994—2016 年的 511 项国家层面科普政策为分析样本, 以 ROSTCM、Ucinet 和 SPSS 软件为工具支撑, 对我国国家层面科普政策文本内容进行了深入挖掘, 探求科普政策文本中难以直接观察到的问题, 以期对我国科普政策制定提供理论依据, 为提高科普政策质量和提升公民科学素质提供理论支撑。

1 样本选择与筛选

1.1 样本来源

在搜索政策文本的来源方面, 本研究的科普政策文本数据主要来源于“北大法律信息网”中的“中国法律法规规章司法解释全

库”, 同时选取政府官方网站公布的政策文本为补充, 以防止遗漏, 进而保证政策文本选取的权威性和全面性。样本的选择是公开颁布的, 不包括未公开的政策文本。

1.2 检索关键词

“1994 年 12 月 5 日发布的《关于加强科学技术普及工作的若干意见》, 是新中国成立以来党中央和国务院共同发布的第一个全面论述科普工作的纲领性文件, 也是我国有史以来第一个公布于众的指导科普工作的官方文件”^[1], 因此, 本研究将检索时间起点设定为 1994 年, 时间终点设定为 2016 年。在检索政策文本过程中, 首先以“科学技术普及”为关键词进行“全文”模糊查找, 共搜索到科学技术普及相关法律法规 2 936 项。为了保

收稿日期: 2018-03-08

* 通信作者: E-mail: kongdeyi_kdy@sohu.com。

证研究的精确性，进一步缩小范围，在上述检索结果中，以“科普”为关键词进行二次检索，共检索到相关政策文本 733 项。

1.3 政策文本的筛选和整理

为确保所选取的政策文本具有准确性和代表性，提高研究针对性，本研究按照以下原则对检索到的政策文本进行遴选和整理。首先，只采用国家层面发布的政策文本，即发文单位为中共中央、全国人大、国务院及其所属机构。其次，所检索的政策类型主要是法律法规、纲要、通知、意见、规划、公告等文件，如复函、批复等其他文件不计入在内。再次，科普内容相关性较低的政策文本，以及在政策条文中未直接规定和明确体现科普措施，仅体现发文单位对科普发展所持态度的政策文本不纳入统计范围。最后，将转发以及重复的政策文本给予剔除。对研究意义不大的政策文本，如科普节目播出时间通知、个人获奖名单通报、科普基地命名公告、会议和培训时间通知等不予考虑。通过对政策文本的整理和遴选以及对政策文本收集结果的调整和完善，最终梳理有效政策文本 511 份。

2 关键词提取与整理

2.1 关键词提取

本研究首先删除了政策文本中的一些无实际意义的内容，如文件的发文机关、文件的接受单位以及政策发布时间等。然后运用 ROSTCM 词频分析软件中的批量文件处理功能将 50 份研究样本归并为一份样本，即“母文本”，之后根据分析需要，运用 ROSTCM 分析软件对“母文本”进行“分词”和“词频分析”等试探性分析，初步统计情况见表 1。

2.2 关键词整理

从表 1 中可以看出，有些分词后词组并非能表达出文件里所要传达的原始含义，如

表 1 政策文本词频统计

序号	关键词	词频	序号	关键词	词频	序号	关键词	词频
1	科普	3 809	21	普及	301	41	技术	186
2	工作	1 103	22	设施	297	42	有关	183
3	科学	892	23	提高	285	43	国土	180
4	科技	866	24	青少年	275	44	建立	179
5	活动	844	25	服务	270	45	示范	178
6	资源	719	26	公众	267	46	推动	174
7	建设	574	27	国家	262	47	相关	174
8	开展	537	28	农村	260	48	管理	171
9	教育	461	29	实施	254	49	支持	171
10	社会	435	30	部门	242	50	开发	167
11	发展	415	31	中国	229	51	重要	166
12	科协	413	32	项目	218	52	积极	164
13	全国	388	33	科学技术	217	53	进行	163
14	基地	383	34	培训	210	54	发挥	161
15	单位	361	35	能力	210	55	纲要	160
16	组织	359	36	计划	203	56	共享	159
17	加强	339	37	环境	200	57	重点	157
18	宣传	330	38	内容	200	58	充分	154
19	素质	322	39	知识	200	59	结合	150
20	社区	309	40	基础	187	60	参与	149

资料来源：根据“北大法律信息网”的“中国法律法规规章司法解释全库”下载数据整理统计而成。

“基础设施”被拆分为“基础”“设施”，“国土资源”被拆分为“国土”“资源”等。另外，与研究主题关联性不大的词组也充斥在其中，如“以上”“为了”“相关”等。因此，为了保证关键词的有效性，必须通过人工手动处理，对文本进行进一步整理。首先，通过该软件自带的分词过滤功能，将与研究主题关

表 2 关键词自定义词表

关键词名称	关键词名称	关键词名称
社区科普	保护环境	科普能力
科普事业	科学素质	科普人才
科学知识	资源共享	科普教育
科技知识	科学文化素质	技术创新
科普工作	科普作品	可持续发展
科普活动	公众参与	经济发展
科普宣传	科技创新	科技进步
科普示范基地	科学技术创新	科普服务
科普队伍	工作规划	技术创新
人才培养	发展规划	科普项目
社会发展	经费支持	科普产业
社会进步	科普人员	资金支持
资源共建	科普基地	科学文化
国土资源	基础设施	农村发展
志愿者	经费投入	
环境保护	科学技术普及	

资料来源：根据“北大法律信息网”的“中国法律法规规章司法解释全库”下载数据整理统计而成。

联性不强的词组过滤掉。之后利用 ROSTCM 分析软件中的自定义词表功能,在默认词表的基础上,根据样本内容自定义词组,对 ROSTCM 分析系统分词的计算方法进行重新定义,增加一部分新生词组,如表 2 所示。由于我国语言文字含义极其丰富,有些词组存在意思相近或具有包含关系,从而导致 ROSTCM 分析软件无法识别,这就需要进行人工判断和处理。例如“经费投入”“经费支持”“资金投入”等词组意思十分接近,“科普基地”包含“科普示范基地”等,这些词组都可以进行归并处理。

通过对关键词的提取与整理,并重新定义分词过滤词表和自定义词表,得到过滤后的关键词有效词表,由于篇幅所限,表 3 中只列出排名前 30 位的关键词,从表 3 中可以大致了解科普政策文本中的核心内容。

表 3 政策文本关键词有效词表(前 30 位)

序号	关键词	词频	序号	关键词	词频	序号	关键词	词频
1	科普	1 495	11	科学素质	235	21	环境	144
2	科学技术	1 342	12	服务	226	22	申报	144
3	科普资源	430	13	经费	226	23	参与	139
4	科普教育	434	14	培训	210	24	共享	139
5	科普活动	410	15	科普宣传	207	25	基层	137
6	社会	405	16	规划	203	26	社区科普	128
7	基础设施	279	17	项目	199	27	指导	124
8	青少年	275	18	社区	179	28	科技馆	123
9	农村	258	19	科普基地	176	29	农民	121
10	公众	257	20	能力	171	30	人员	119

资料来源:根据“北大法律信息网”的“中国法律法规规章司法解释全库”下载数据整理统计而成。

3 共词矩阵构建

3.1 原始矩阵生成

“共词分析是通过对文献中集中出现的词汇进行统计,分析某一学科领域研究主题或研究方向的专业术语共同出现在一篇文献中的现象,从而判断学科领域中各个研究主题间的关系,纵向和横向展示出学科领域的研究热点和发展进程,展现该领域的研究历史和现状。”^[1] 本研究将表 3 中的 30 个高频关键词输入 EXCEL 表中,统计每个关键词与其他关键词

在政策文本中共同出现的次数,即两两统计它们在同一篇政策文本中出现的次数,如果两个关键词在政策文本中同时出现的频率越高,说明它们之间的关系越为密切,这样就形成了一个 30×30 的共词矩阵,也可称为原始矩阵,见表 4(为方便显示,将横向表头“科普”至“人员”等关键词字样分别用序号 1—30 表示)。共词矩阵中对角线上数字为相应关键词出现的总频次,非主对角线单元格中的数据为两个关键词共同出现的频次。如关键词“科学素质”出现了 235 次,关键词“科普基地”出现了 176 次,而“科学素质”与“科普基地”两个关键词共同出现的频次为 16 次,说明有 16 项政策文本同时使用了这两个关键词。

表 4 部分共词矩阵列表

	1	2	3	4	5	6	7	8	9	...	29	30
科普	1 495	41	21	24	32	27	20	17	18	...	11	23
科学技术	41	1 342	20	26	33	24	18	18	20	...	11	24
科普资源	21	20	434	15	19	18	12	10	9	...	6	8
科普教育	24	26	15	430	21	20	14	14	13	...	8	19
科普活动	32	33	19	21	410	27	20	17	17	...	11	20
社会	27	24	18	20	27	405	13	15	14	...	9	16
基础设施	20	18	12	14	20	13	279	9	12	...	6	15
青少年	17	18	10	14	17	15	9	275	11	...	9	19
农村	18	20	9	13	17	14	12	11	258	...	8	11
公众	27	22	11	14	20	20	11	12	13	...	7	17
科学素质	32	29	15	20	24	22	13	14	20	...	8	15
服务	9	11	5	9	6	6	5	3	6	...	3	6
经费	21	20	9	14	16	12	11	10	11	...	8	12
培训	15	17	9	10	14	15	10	11	13	...	6	19
科普宣传	23	25	13	15	21	16	13	13	16	...	11	20
规划	38	31	17	22	28	25	18	18	21	...	10	23
项目	17	18	14	11	16	11	10	8	9	...	3	10
社区	13	10	7	10	9	10	8	9	18	...	3	10
科普基地	21	21	9	12	15	18	25	11	10	...	7	14
能力	27	23	17	15	23	19	16	11	12	...	7	7
环境	16	14	7	9	14	12	9	8	7	...	6	10
申报	11	8	5	5	9	6	6	4	5	...	3	3
参与	17	19	13	17	19	23	11	13	9	...	8	16
共享	7	8	8	7	5	5	3	4	5	...	1	2
基层	13	14	8	9	12	10	8	7	12	...	8	8
社区科普	5	5	3	4	4	3	3	5	10	...	3	3
指导	16	15	12	7	16	16	6	7	8	...	5	10
科技馆	8	9	4	6	7	9	12	6	6	...	2	4
农民	11	11	6	8	11	9	6	9	8	...	121	10
人员	23	24	8	19	20	16	15	19	11	...	10	119

资料来源:根据“北大法律信息网”的“中国法律法规规章司法解释全库”下载数据整理统计而成。

3.2 相关矩阵及相异矩阵构造

“在实际的计量化共词分析的过程中，关键词共现频次为绝对值，难以真正反映词与词之间的相互依赖程度，要想真正揭示关键词之间的共现关系，有必要对其进行包容化处理，构造相关矩阵。”^[2] 本研究运用 Ochiai 系数对共词矩阵进行处理，具体的处理方法是 将共词矩阵中 X、Y 两关键词共同出现的次数除以 X、Y 两关键词各自出现频次开方的乘

积，将多值矩阵转化为在 [0，1] 区间取值的相关矩阵，如公式 1 所示。

Ochiai 系数=
$$\frac{X、Y\text{关键词同时出现的频次}}{\sqrt{X\text{关键词出现的总频次}} \cdot \sqrt{Y\text{关键词出现的总频次}}}$$
 (1)

利用 Ochiai 系数计算公式将共词矩阵（表 4）中的数值转换成相关矩阵，见表 5（为方便显示，将横向表头“科普”至“人员”等关键词分别用序号 1—30 表示）。

“相关矩阵中的数值表示两个高频词之

表 5 部分相关矩阵列表

	1	2	3	4	5	6	7	8	9	…	29	30
科普	1.000	0.063	0.072	0.083	0.113	0.097	0.099	0.087	0.097	…	0.115	0.213
科学技术	0.063	1.000	0.071	0.092	0.119	0.091	0.093	0.094	0.108	…	0.118	0.225
科普资源	0.072	0.071	1.000	0.092	0.119	0.089	0.099	0.088	0.088	…	0.103	0.146
科普教育	0.083	0.092	0.092	1.000	0.136	0.127	0.117	0.117	0.119	…	0.133	0.258
科普活动	0.113	0.119	0.119	0.136	1.000	0.171	0.163	0.146	0.156	…	0.178	0.295
社会	0.097	0.091	0.089	0.127	0.171	1.000	0.120	0.131	0.134	…	0.152	0.249
设施	0.099	0.093	0.099	0.117	0.163	0.120	1.000	0.108	0.137	…	0.133	0.258
青少年	0.087	0.094	0.088	0.117	0.146	0.131	0.108	1.000	0.130	…	0.164	0.239
农村	0.097	0.108	0.088	0.119	0.156	0.134	0.137	0.130	1.000	…	0.163	0.237
公众	0.137	0.119	0.103	0.130	0.179	0.173	0.136	0.142	0.159	…	0.160	0.307
科学素质	0.172	0.162	0.142	0.182	0.227	0.207	0.171	0.177	0.232	…	0.199	0.336
服务	0.055	0.066	0.052	0.082	0.070	0.067	0.067	0.052	0.080	…	0.076	0.132
经费	0.121	0.119	0.093	0.133	0.159	0.128	0.136	0.128	0.145	…	0.168	0.253
培训	0.098	0.111	0.098	0.114	0.155	0.155	0.136	0.143	0.168	…	0.153	0.294
科普宣传	0.144	0.157	0.135	0.160	0.218	0.178	0.176	0.175	0.210	…	0.233	0.384
规划	0.219	0.190	0.172	0.218	0.280	0.251	0.230	0.228	0.267	…	0.251	0.458
项目	0.112	0.120	0.135	0.124	0.172	0.132	0.139	0.123	0.141	…	0.119	0.249
社区	0.096	0.081	0.089	0.118	0.125	0.126	0.123	0.130	0.156	…	0.113	0.239
科普基地	0.150	0.153	0.116	0.149	0.190	0.203	0.159	0.167	0.170	…	0.192	0.332
能力	0.188	0.169	0.181	0.180	0.259	0.222	0.219	0.179	0.201	…	0.207	0.292
环境	0.137	0.125	0.103	0.130	0.188	0.165	0.152	0.142	0.142	…	0.172	0.277
申报	0.096	0.075	0.072	0.078	0.123	0.093	0.102	0.081	0.098	…	0.101	0.139
参与	0.151	0.169	0.169	0.216	0.255	0.245	0.195	0.212	0.188	…	0.233	0.401
共享	0.066	0.075	0.096	0.091	0.083	0.078	0.065	0.075	0.090	…	0.062	0.106
基层	0.121	0.131	0.117	0.137	0.180	0.155	0.150	0.139	0.200	…	0.208	0.266
社区科普	0.052	0.053	0.050	0.064	0.072	0.060	0.064	0.085	0.090	…	0.088	0.116
指导	0.156	0.151	0.165	0.132	0.234	0.225	0.144	0.153	0.175	…	0.184	0.316
科技馆	0.084	0.094	0.071	0.098	0.119	0.133	0.080	0.113	0.119	…	0.096	0.166
农民	0.115	0.118	0.103	0.133	0.178	0.152	0.133	0.164	0.163	…	1.000	0.290
人员	0.213	0.225	0.146	0.258	0.295	0.249	0.258	0.239	0.237	…	0.290	1.000

资料来源：根据“北大法律信息网”的“中国法律法规规章司法解释全库”下载数据整理统计而成。

间的距离，数值越大，说明高频词之间距离越近，关联度越高；数值越小，说明高频词之间距离越远，关联度越小。”^[3] 在相关矩阵中，主对角线上的数值表示某词与自身相关程度，值都为 1，说明其自身相关系数最大。譬如，关键词“科普”自身相关系数为

1。反之，矩阵中数值越小，说明两两关键词之间距离越远，相似度越低。例如，关键词“社区科普”与关键词“科普资源”的相关系数为 0.050，说明这两个关键词距离远，相似度低。为了有利于后续的聚类分析和多维尺度分析的需求，进一步消除误差，用 1 与

上述相关矩阵的各个数据相减,进而得到反映两个关键词间差异程度的相异矩阵,见表6(为方便显示,将横向表头“科普”至“人员”等关键词字样分别用序号1—30表示)。“与相关矩阵相反,相异矩阵中数值越大,表

明对应两词汇距离越远,相似度越差;数值越小,则相似度越好。”^[4]在相异矩阵中,主对角线上的数字为0,说明关键词自身不相异。两两关键词间在相异矩阵和相关矩阵中的数值大小,反映了关键词之间距离的远近,

表6 部分相异矩阵列表

	1	2	3	4	5	6	7	8	9	...	29	30
科普	0.000	0.937	0.928	0.917	0.887	0.903	0.901	0.913	0.903	...	0.885	0.787
科学技术	0.937	0.000	0.929	0.908	0.881	0.909	0.907	0.906	0.892	...	0.882	0.775
科普资源	0.928	0.929	0.000	0.908	0.881	0.911	0.901	0.912	0.912	...	0.897	0.854
科普教育	0.917	0.908	0.908	0.000	0.864	0.873	0.883	0.883	0.881	...	0.867	0.742
科普活动	0.887	0.881	0.881	0.864	0.000	0.829	0.837	0.854	0.844	...	0.822	0.705
社会	0.903	0.909	0.911	0.873	0.829	0.000	0.880	0.869	0.866	...	0.848	0.751
设施	0.901	0.907	0.901	0.883	0.837	0.880	0.000	0.892	0.863	...	0.867	0.742
青少年	0.913	0.906	0.912	0.883	0.854	0.869	0.892	0.000	0.870	...	0.836	0.761
农村	0.903	0.892	0.912	0.881	0.844	0.866	0.863	0.870	0.000	...	0.837	0.763
公众	0.863	0.881	0.897	0.870	0.821	0.827	0.864	0.858	0.841	...	0.840	0.693
科学素质	0.828	0.838	0.858	0.818	0.773	0.793	0.829	0.823	0.768	...	0.801	0.664
服务	0.945	0.934	0.948	0.918	0.930	0.933	0.933	0.948	0.920	...	0.924	0.868
经费	0.879	0.881	0.907	0.867	0.841	0.872	0.864	0.872	0.855	...	0.832	0.747
培训	0.902	0.889	0.902	0.886	0.845	0.845	0.864	0.857	0.832	...	0.847	0.706
科普宣传	0.856	0.843	0.865	0.840	0.782	0.822	0.824	0.825	0.790	...	0.767	0.616
规划	0.781	0.810	0.828	0.782	0.720	0.749	0.770	0.772	0.733	...	0.749	0.542
项目	0.888	0.880	0.865	0.876	0.828	0.868	0.861	0.877	0.859	...	0.881	0.751
社区	0.904	0.919	0.911	0.882	0.875	0.874	0.877	0.870	0.844	...	0.887	0.761
科普基地	0.850	0.847	0.884	0.851	0.810	0.797	0.841	0.833	0.830	...	0.808	0.668
能力	0.812	0.831	0.819	0.820	0.741	0.778	0.781	0.821	0.799	...	0.793	0.708
环境	0.863	0.875	0.897	0.870	0.812	0.835	0.848	0.858	0.858	...	0.828	0.723
申报	0.904	0.925	0.928	0.922	0.877	0.907	0.898	0.919	0.902	...	0.899	0.861
参与	0.849	0.831	0.831	0.784	0.745	0.755	0.805	0.788	0.812	...	0.767	0.599
共享	0.934	0.925	0.904	0.909	0.917	0.922	0.935	0.925	0.910	...	0.938	0.894
基层	0.879	0.869	0.883	0.863	0.820	0.845	0.850	0.861	0.800	...	0.792	0.734
社区科普	0.948	0.947	0.950	0.936	0.928	0.940	0.936	0.915	0.910	...	0.912	0.884
指导	0.844	0.849	0.835	0.868	0.766	0.775	0.856	0.847	0.825	...	0.816	0.684
科技馆	0.916	0.906	0.929	0.902	0.881	0.867	0.920	0.887	0.881	...	0.904	0.834
农民	0.885	0.882	0.897	0.867	0.822	0.848	0.867	0.836	0.837	...	0.000	0.710
人员	0.787	0.775	0.854	0.742	0.705	0.751	0.742	0.761	0.763	...	0.710	0.000

资料来源:根据“北大法律信息网”的“中国法律法规规章司法解释全库”下载数据整理统计而成。

也表明了关键词间的内部结构联系和亲疏关系。

4 科普政策内容多元统计分析

4.1 聚类分析

“聚类分析又称为群分析,它是研究分类问题的一种多元统计方法。”^[5]聚类分析依据关键词之间的共现强度,将共现强度较大的关键词进行聚集而形成一个个聚类。聚类分

析包括两种主要方法,一种是“快速聚类分析方法”(K — Means Cluster Analysis),另一种是“层次聚类分析方法”(Hierarchical Cluster Analysis),运用不同的聚类分析方法会得到不同的结论。本研究采用层次聚类分析法。将高频关键词的相异矩阵表(表6)导入SPSS22.0中进行层次聚类分析,利用“Classify”的“hierarchical clustering”功能,在度量标准中选择区间(Interval)中的平方

欧式距离 (Squared Euclidean distance) 度量, 选择“离差平方和 (Ward)”作为聚类方法, 最终以树形图 (Dendrogram) 的形式输出, 见图 1。

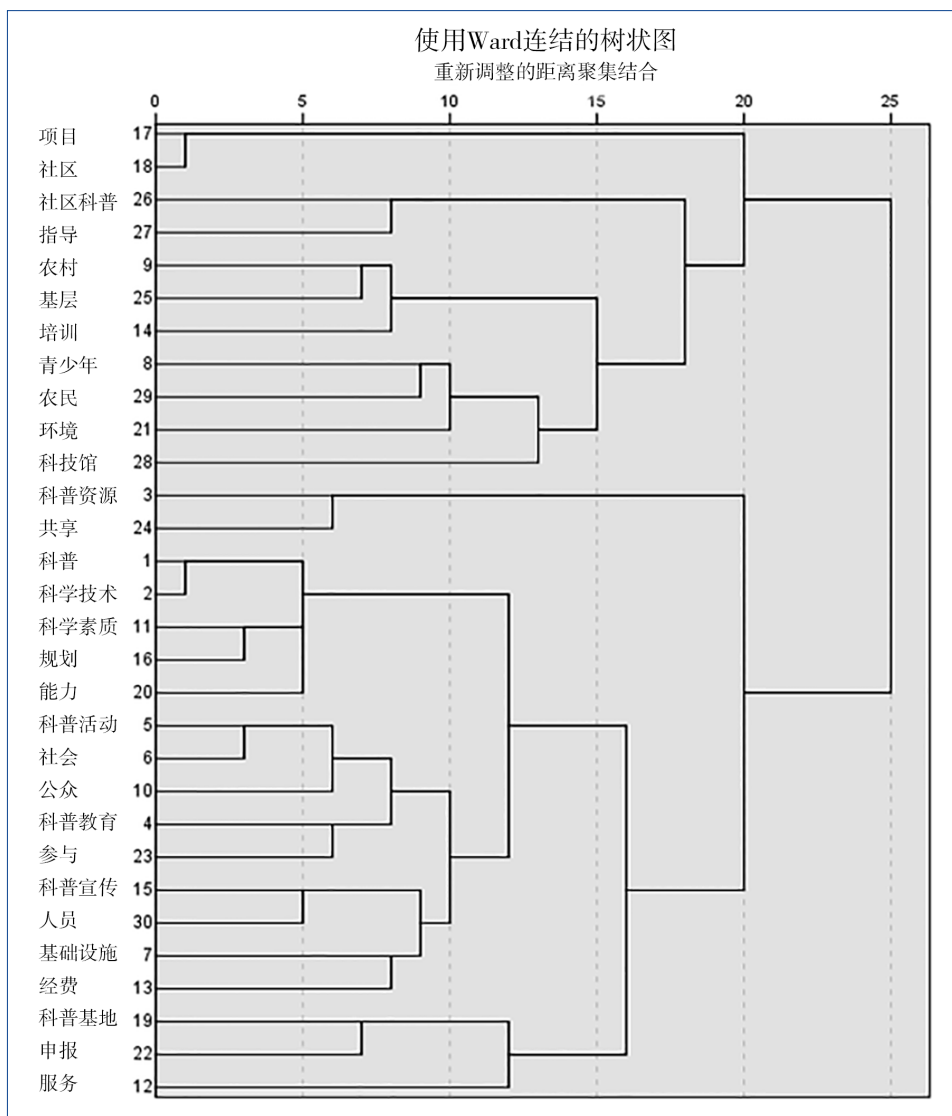


图1 聚类分析树状图

资料来源: 根据“北大法律信息网”的“中国法律法规规章司法解释全库”下载数据整理, 并通过SPSS软件生成。

“聚类分析树状图能够清晰地反映聚类过程的进度, 能够显示出每一次聚类过程中各个样本进行合并的情况。从聚类分析的树状图可以清晰地看出各个关键词间的聚类情况, 能够较为明确地说明各个关键词之间的关系, 根据高频关键词层次聚类分析的树状图能够直观地表明各个关键词之间聚类的先

后顺序。”^[6]由高频关键词层次聚类分析的树状图可知, 首先是关键词17和关键词18聚成一类, 其次是关键词26和关键词27聚成一类, 然后是关键词9和关键词25聚成的类和

关键词14聚合, 之后以此类推可得出其他聚类。总体上看, 聚类结果并不十分明显, 但仍然可将高频关键词大致分为8个类别, 即社区科普项目、基层和农村建设、青少年和农民的培训和指导、科普资源共享、提升科学素质和科普能力、科普活动和教育中的社会 and 公众参与、基础设施建设与经费投入、科普基地申报与服务。

4.2 多维尺度分析

“多维尺度分析是市场研究的一种有力手段, 它可以通过低维空间 (通常是二维空间) 展示多个研究对象之间的关系, 利用平面距离来反映研究对象之间的相似

程度。关键词越趋近于中心点且分布越集中, 表明这些关键词越是研究的热点, 越受到关注。相反, 越靠近边缘位置, 表明这些关键词越是研究冷点。”^[7]本研究采取多维尺度分析科普政策文本内容类目, 使用SPSS22.0中Scale菜单的多维尺度分析 (multidimensional scale, ALSCAL), 用序数数值 (ordinal) 作

为数据测度水平的指标,将度量模型设置为个别差异 Euclidean 距离 (D)。利用二维尺度分析对科普政策文本内容高频关键词的相异矩阵 (表 6) 进行分析,得到相关多维尺度分析的可视化结果,见图 2。

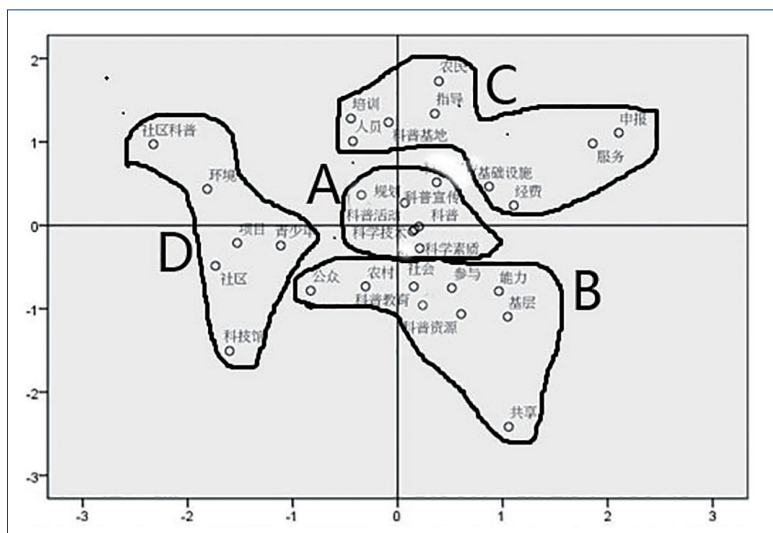


图 2 多维尺度分析结果

资料来源:根据“北大法律信息网”的“中国法律法规规章司法解释全库”下载数据整理,并通过 SPSS 软件生成。

从图 2 中可看出四大词团,即提升科学技术普及过程中的科普活动与科普宣传 (A)、社会公众参与和科普资源共享 (B)、科普人员培训及基础设施建设 (C)、社区科普项目及环境 (D)。其中词团 A 中包含“科普”“科学技术”“科普活动”“科学素质”“科普宣传”等关键词;词团 B 包含“社会”“公众”“参与”“科普教育”“科普资源”“共享”等关键词;词团 C 包含“基础设施”“经费”“科普基地”“人员”“培训”“农民”等关键词;词团 D 包含“社区科普”“项目”“环境”等关键词。这四大词团在一定程度上与聚类分析的结果大致吻合。此外,从图 2 中还

可知,词团 A 处于图的中心区域,并且词团中关键词之间联系紧密,说明该词团为科普政策的核心内容。其他词团围绕在词团 A 的周围,词团由 B 至 D 各内部关键词联系紧密程度呈递减趋势,属于科普政策次核心内容。

4.3 社会网络分析

4.3.1 关键词共现网络图谱

为了更为直观明确的得出关键词之间的相关关系,将“母文本”导入到 ROSTCM 分析软件中,利用嵌入在该软件中的 NetDraw,对行特征词共词矩阵进行网络构建,其中关键词共现数值是“母文本”中两两关键词同一行中出现的频数,这一过程是利用 ROSTCM 中的语义网络和社会网络生成工具来完成。包括提取高频词、过滤无意义词、提取行特征、构建网络、构

建矩阵等多个操作步骤,见图 3。最终得到

图 3 ROSTCM 分析软件中社会网络与语义网络分析

资料来源:根据“北大法律信息网”的“中国法律法规规章司法解释全库”下载数据整理,并通过 Ucinet 软件生成。

我国科普政策文本的关键词网络图谱,见图 4。其中每一个节点都代表政策文本中的关键词,并且节点间的连线代表其存在相互关联性。从图 4 中可以清晰看出,“科普”“科普资源”“社会”“科普活动”“服务”“科普教育”“科学素质”“农村”“培训”等关键词处

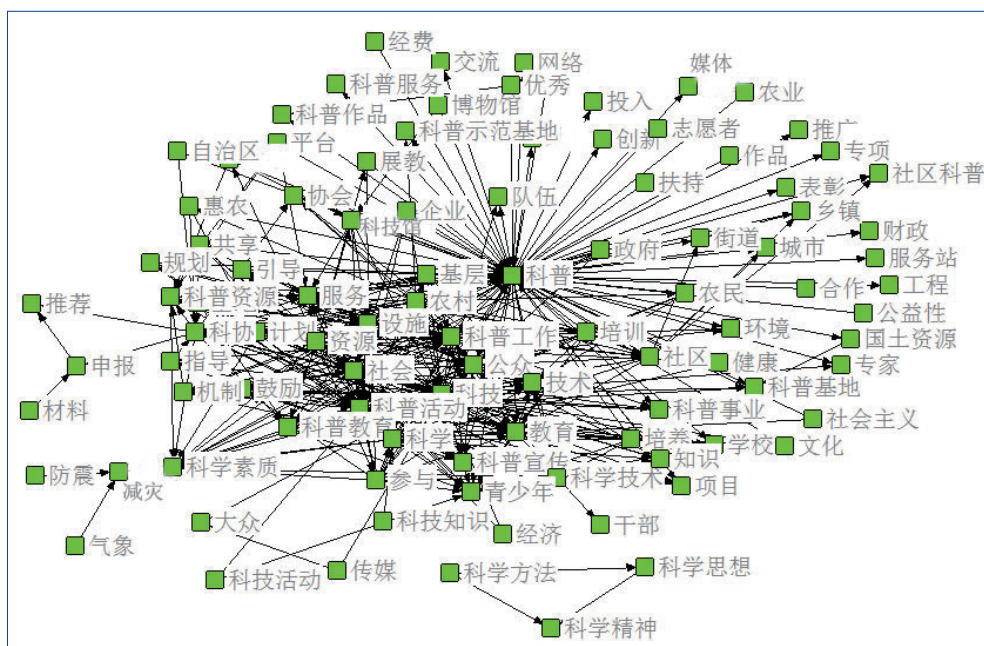


图 4 NetDraw 社会网络语义分析

资料来源：根据“北大法律信息网”的“中国法律法规规章司法解释全库”下载数据整理，通过 Ucinet 软件生成。

于图谱的中心位置, 这些关键词是科普政策文本的核心内容, 而在整个网络的边缘, 也存在一些零散的关键词, 这些边缘关键词虽然游离在核心位置之外, 但有些却是一些新兴的关键词, 有的也代表未来的发展趋势, 同样具有重要的研究价值, 如“科普服务”和“科普产品”等。

从 NetDraw 关键词网络分析结果中，不但能对政策文本中的关键词有着更直观的认识，还可以看出词组之间的结构关系。在该关键词网络中，“科普”这一节点位于网络中核心的位置，对其他关键词共现的影响能力也是最强的，也可以说“科普”与其他词组关联程度最大，在这一网络中它与许多节点都存在联系，这

```
BLOCK DENSITIES OR AVERAGE
-----
Input dataset:
|
Relation: 1-行特征词--共词
Density (matrix average) =
Standard deviation = 1.268
Use MATRIX>TRANSFORM>DICHOTOMY
Density table(s) saved as
Standard deviations saved
Actor-by-actor pre-image m
-----
Running time: 00:00:01
Output generated: 22 七月
Copyright (c) 1999-2008 And
```

资料来源：根据“北大法律信息网”的

些节点都需通过“科普”一词实现关键词共现。也可这样理解,哪些关键词在网络中与“科普”关系最为紧密,也可以将其视为核心词组。通过ROSTCM软件看出关键词之间联系的紧密程度,strength值越大说明两者联系越为密切,“资源”与

“科普”的 strength 值为 121, 说明二者结合的最为紧密, “科普”和“资源”是政策文本中的核心内容。另外, “科学技术”“社会”“科普活动”“基础设施”“服务”与“科普”的联系也十分紧密, strength 值分别为 195、179、160、143、125, 这也验证了上述关键词是该领域的核心内容这一分析。

4.3.2 网络密度分析

Ucinet 对网络密度的测量依据有向图密度公式进行, 在 Ucinet 中选择 “Network—

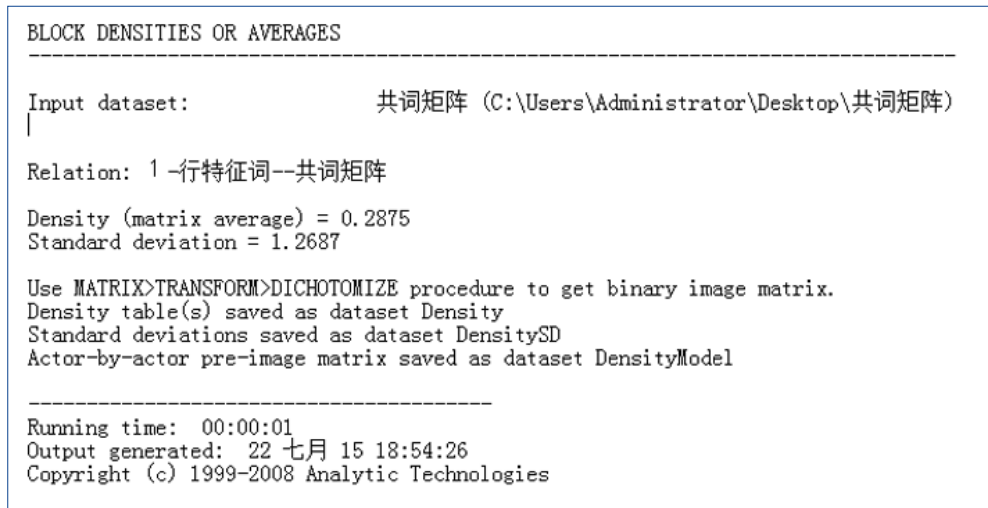


图 5 网络密度测量结果

资料来源：根据“北大法律信息网”的“中国法律法规规章司法解释全库”下载数据整理，通过 Ucinet 软件生成。

Cohesion—Density”项目，结果如图5所示。从图5可以看出，此关键词网络密度为0.2875，说明网络节点之间连接不紧密，关键词之间的联系程度不高。均方差为1.2687，说明网络离散程度较低，整个网络节点间连接较为稀疏，关键词之间存在小范围聚集现象，即只有少数关键词之间联系比较紧密，大部分关键词之间缺少有效的联系。以上分析结果表明，关键词网络中可同时存在密度小和离散程度也小的情况，这也反映出我国科普政策文本内容主要集中在少数几方面，彼此之间缺少联系，我国科普政策文本内容设计不对称的现状客观存在。

5 研究结论

5.1 科普政策内容聚焦点清晰

通过聚类分析和多维尺度分析表明科普政策内容涉及的主题广泛。从聚类分析得出，科普政策内容关键词可分为八大类，分别为社区科普项目、基层和农村建设、青少年和农民的培训和指导、科普资源共享、提升科学素质和科普能力、科普活动和教育中的社会 and 公众参与、基础设施建设与经费投入、科普基地申报与服务。从多维尺度分析可将我国科普政策内容分为四大词团，分别为提升科学技术普及过程中的科普活动与科普宣传、社会公众参与和科普资源共享、科普人员培训及基础设施建设、社区科普项目及环境。这表明科普政策内容领域有诸多主题亟待探讨，需以分散的研究成果来实现科普政策内容的有效整合和融会贯通，以提升科普政策质量，使其实际效应得以最大化发挥。

5.2 科普政策内容设计存在缺陷

为了更为直观明确的得出关键词之间的相关关系，利用嵌入在ROSTCM词频分析软件中的NetDraw，生成了我国科普政策文本的关键词网络图谱。利用Ucinet软件对关键词

进行了网络密度分析，研究得出，整个关键词网络表现出较低的中心度趋势，关键词网络密度为0.2875，说明整个网络节点之间连接不紧密，关键词之间的联系程度不高。均方差为1.2687，说明网络离散程度较低，关键词网络只在小范围内有一定的集中趋势，科普政策文本内容主要集中在少数几方面。在政策内容中过于强调科普硬实力的建设，而忽视了科普软实力的发展；在提高公民科学素质过程中，过于关注青少年群体而疏忽了其他年龄阶段群体，等等。这些问题反映出科普政策内容设计存在缺陷，而提高科普政策内容设计的科学性及实用性不失为解决这一问题的“灵丹妙药”。

6 完善科普政策内容的路径实现

6.1 强化科普政策内容的系统性

通过对国家层面科普政策文本内容的挖掘，发现我国科普政策内容设计呈现出明显的非对称性以及缺乏系统性，科普政策内容设计的不对称和缺乏系统性严重影响了科普政策实施效果，不利于科普事业的稳定发展。例如，在科普政策内容中过于关注青少年群体而疏忽了其他年龄阶段群体科学素质的提升；过于强调基础设施、科普基地等科普硬实力的建设，而忽视了科普文化、公民科学素质等科普软实力的发展。应对科普政策内容所缺失的重要方面进行填补，以增强科普政策内容的系统性。如加强科普软实力的建设力度，达到科普硬实力和软实力的均衡发展；进一步注重科普产业建设，实现公益性科普事业和经营性科普产业的兼顾发展；加大企业、学校等社会力量的参与程度，提高其积极性，形成政府引导、社会全方位参与的科普事业发展的格局。

6.2 加强科普政策内容的衔接性

全面提升公民的科学素质是我国科普政

策的基本目标,面对公民科学素质参差不齐的情况,政策主体应对科普政策内容进行优化,增强政策内容的衔接性,以完善科普政策内容。例如,经分析得出,青少年群体是我国科普工作重点培养对象,而对幼儿、中年人和老年人群体关注度明显不足,这说明在提升不同年龄阶段的公民科学素质过程中,科普工作链条发生了断裂,也表明科普政策内容在衔接性上出现了问题。因此,科普政策内容应该从整体上进行设计,政策内容应具有衔接性和层次性,根据不同的年龄阶段、接受知识的能力以及实际需要来规定具体的内容,保证科普政策内容设计的科学性和实效性,以此追求科普政策内容效果的最优化,全面满足公民素质提升的需要。

6.3 重视评估反馈工作以补充与修正科普政策内容

研究表明,我国科普政策对象范围存在一定的局限性,一些重要的关键因素未能在科普政策内容中得到应有的体现,如科普文化、科普公共服务等非物质因素的缺失等。

针对这一问题,首先应该重视对科普政策内容的补充,将一些容易被忽略的科普对象纳入到科普政策对象范畴中。其次,需要对颁布实行的科普政策效果进行定期的评估与反馈,及时搜集有关具体科普政策效果的资料与信息,并进行科学有效的分析,根据信息反馈和分析结果对科普政策进行必要的补充或修正,以完善科普政策内容。以《关于鼓励科普事业发展税收政策问题的通知》为例,根据全国科普工作统计调查数据,在387个国家级科普教育基地中,享受过税收优惠政策的只有78个,占总数的20.2%;在1629个省级科普教育基地中,享受过税收优惠政策的为176个,占总数的10.8%。可见,从这项政策受益的科普教育基地为数尚少。如能对科普政策效果进行及时的信息反馈,完全能避免此类事情的发生,因此,应加强对评估反馈工作的重视程度,只有这样才能及时、有效地根据实际情况对科普政策效果做出判断,从而对科普政策内容作出进一步调整与完善。

参考文献

- [1] 潘佳宝,喻国明.新闻传播学视域下中国舆论研究的知识图谱(1986—2015)——基于文献计量学的研究[J].新闻学与传播学,2017(9):1-11.
- [2] 吴进.基于文本分析的我国产业共性技术创新政策研究[D].广州:华南理工大学,2013.
- [3] 田丽,余雪丽.基于共词分析的我国阅读推广研究热点探析[J].情报科学,2016(12):127-132.
- [4] 王志强,杨庆梅.我国创业教育研究的知识图谱——2000—2016年教育学CSSCI期刊的文献计量学分析[J].教育研究,2017(6):58-64.
- [5] 冯楚建,谢其军.国内外光伏产业政策绩效对比研究[J].中国科技论坛,2017(2):58-65.
- [6] 顾洪涛.我国高校图书馆研究热点探析[D].大连:辽宁师范大学,2014.
- [7] 罗敏,朱雪忠.基于共词分析的我国低碳政策构成研究[J].管理学报,2014(11):1681-1685.

(编辑 张南茜)

A Study on the Development of Science Popularization Industry in China

Wang Kangyou Zheng Nian Wang Lihui

(China Research Institute for Science Popularization, Beijing 100081)

Abstract: Science popularization industry is a collection of all kinds of operating entities, which provides resources, products and services for the science popularization system. This study analyzes the dominant formats related to science and technology communication abroad. Based on statistical data and a survey by the research team, this paper gives a brief overview of the new development trend of the science popularization industry in China, and summarizes the problems in China's policy, institutional mechanism and market demand during the development of science popularization industry, and puts forward suggestions for the further development of China's science popularization industry.

Keywords: science popularization industry; business; public

CLC Numbers: N4 **Document Code:** A **DOI:** 10.19293/j.cnki.1673-8357.2018.03.001

The Content Mining of Science Popularization Policies in China from a Perspective of Visualization

Kong Deyi

(School of Humanities and Social Sciences, Henan University of Engineering, Zhengzhou 451191)

Abstract: In order to excavate the core content and its problems of science popularization policy, ROSTCM, a kind of word frequency analysis software, is used to extract keywords, and co-word matrix of keywords and correlation and dissimilarity matrix are constructed; then Ucinet and SPSS are used to make cluster analysis, multi-dimensional scaling analysis and social network analysis of these keywords. In cluster analysis, the keywords of the popular science policy content are classified into 8 classes, which is respectively conducting training for youngsters and peasants, popular science resource sharing, infrastructure and fund input, etc. It shows that popular science policy covers a wide range of topics. In multidimensional scaling, popular science policy content is classified into 4 classes words, content of science popularization policy possess distinct focus points. By analyzing the keywords of the popular science policy content in network density, the study shows that there exist flaws in the content design of those policies. In a nutshell, the paper puts forward some suggestions which strengthen the hierarchy of policy content, and make policy content be more systematically conducted, etc.

Keywords: visualization; popular science policy; content mining; co-word analysis

CLC Numbers: G322.0 **Document Code:** A **DOI:** 10.19293/j.cnki.1673-8357.2018.03.002