

基于 TOPSIS 的信息资源间接测度研究¹

俞立平 潘云涛 武夷山

中国科学技术信息研究所 北京 100038

摘要：为了解决用指标体系和替代变量进行信息资源测度存在的问题，本文选取人口数、社会活跃程度、教育水平、传播水平四个信息资源的产生及影响因素，采取加权 TOPSIS 进行信息资源的测度，提出了一种新的信息资源间接测度方法，并讨论了本方法的应用范围和进一步改进的方向。

关键词：信息资源 间接测度 TOPSIS

中图分类号 G203

1 引言

信息资源是用来创造社会财富，增进人类福利的，经过加工处理的有用信息的集合。对信息资源进行测度，不仅可以定量考察信息化建设的成果，促进信息资源的开发与利用，优化信息资源配置，比较国家和地区之间的差距，从而给政府相关部门提供参考，而且对于进一步研究信息资源的要素贡献具有十分重要的作用，因此对信息资源测度的研究成为近年来学术界研究的热点。

对信息资源定量测度最早来源于国外有关信息社会和信息化的测度，模型与方法有数十种，其中影响较大的有马克卢普测度法、波拉特测度法、信息化指数法等。马克卢普 (F. Machlup, 1962)^[1]主要是从宏观上测算信息产业在国民生产总值中所占的比重、信息部门就业人数的比例以及信息部门的收入占国民总收入的比重大小，间接描述信息资源的作用与贡献。马克·波拉特 (M. Porat, 1977)^[2]宏观测度方法就是用信息活动的产值占国民生产总值 (GNP) 或国内生产总值 (GDP) 的比例大小、信息劳动者人数占就业人口的比例大小和信息部门就业者收入占国民收入的比例大小来衡量社会信息化程度。日本学者小松崎清介 (1965)^[3]提出信息化指数法并用其测度社会信息化程度，此方法能够从纵向上比较一个国家或地区不同时期的信息化程度以及从横向上比较不同国家或地区之间的信息化程度。国家信息化评测中心历时 8 年，研究提出了 20 项反映国家信息化水平的统计指标体系，在 2001 年确定了国家信息化指标构成方案，以《关于印发〈国家信息化指标构成方案〉的通知》(信部信[2001]434 号) 发布，成为官方标准。

在国内信息资源测度领域，谢康、肖静华 (1997)^[4]提出了信息资源丰裕度系数，选取若干指标，测度信息资源量，权重采取相等权重法。信息资源的生产能力指标有数据库数量、专利商标数量、图书报刊出版发行量、视听产品生产数量。信息资源储备能力指标有计算机拥有量、文化设施、新闻设施、娱乐设施、邮电设施、通信设施等。信息资源的处理潜力指标主要有识字率、在校生数、科研人员数、政府公务员数、咨询人员等。俞立平 (2005)^[5]提出了采用邮电业务量作为信息资源的替代变量。此外，一些学者建立了以图书馆信息资源为核心的信息资源测度指标体系，还有许多学者建立了网络信息资源评价指标体系，网络计量学方法也得到了越来越多的利用。总体上，在信息资源测度领域，还存在如下问题，至今没有解决。

¹国家十一五支撑计划项目 (2006BAH03B05) 基于海量信息的科技发展与科技评价的科学计量学研究；国家自然科学基金资助 (70673019) 强竞争与弱竞争态势下的科技发展与科技评价的科学计量学研究

作者简介：俞立平 (1967-) 男，江苏泰州人，博士，中国科学技术信息研究所博士后，扬州职业大学教授，主要从事信息经济、科学计量领域的教学科研工作。Email: chinayangzhou@yahoo.com.cn

潘云涛 (1967-)，女，中国科学技术信息研究所研究员，主要从事科学计量学研究。

武夷山 (1958-)，男，中国科学技术信息研究所总工程师，研究员，主要从事科技管理领域的研究。

第一，不同学者对信息资源的界定不同，有的特指信息资源本身，有的还包括人员、设备、载体等等。信息资源测度指标体系复杂而庞大，不同学者研究侧重点不同，选取指标也不尽相同。此外不同学者对同一指标的权重赋值不同，测度方法也难以一致，导致信息资源测度结果相差很大。

第二，基于信息资源本身的测度指标难以囊括人类社会生活中无法记录的大量信息，这些信息量是非常巨大的，有时甚至并不少于可以记录的信息量。人均报纸订阅份数、人均广播电视节目套数、人均数据库个数、人均网页浏览数等指标只能反映可以记录的信息，但这些指标无法反映那些人们日常生活中产生的大量但难以记录的信息。

第三，指标选取存在一系列问题。比如，有人用人均图书量作为地区信息资源的指标之一，并根据统计年鉴中列出的某地区每年出版图书总量来确定人均图书量，可是图书发行并不局限于该地区。将专利授权量作为信息资源的指标之一也是有问题的，专利更多地属于知识的范畴。用电脑数量作为信息资源的指标之一也不合适，用上网电脑更合适一些。人均文化设施如文化馆、档案馆、博物馆的数量更不能反映，因为这些场馆的大小是完全不一样的，以上类似的指标问题很多。

第四，指标体系庞大，必然带来数据量大、处理复杂、评价成本高、误差大等一系列问题。

第五，采用替代变量如邮电业务额虽然简化了计算，但仍然存在较大的偏差，如忽视人类日常生活中的大量信息，包裹邮递严格意义上不能表示信息，而且现在许多快递公司邮寄信件资料难以统计。

从现有的研究看，信息资源的测度主要有两大类方法，一是直接测度法，主要是采用指标体系和多属性评价方法进行测度。二是采用信息资源的替代变量。但两者均存在着不少问题，如果信息资源的定量研究继续因循以上两种思路的话，必然走入死胡同。本文从分析信息资源的产生和影响因素入手，提出了第三种的信息资源的测度方法——间接测度法，试图从根本上解决这个问题，

2 信息资源间接测度模型

2.1 信息资源的产生与影响因素分析

现实生活中存在着大量冗余、错误、无用的信息，对其进行剔除是十分困难的，但从宏观上讲，不同地区这些信息所占的比例大致差不多的，为了研究方便，本文假定一个前提条件，即不同地区这些信息的比例是相等的。

信息资源测度虽然非常复杂，但其产生和影响因素是可以观测的，这些因素类似于信息资源的“影子”，信息资源的产生及影响的主要因素有人口（Population）、社会活跃程度（Activity）、信息传播水平（Communication）、教育水平（Education）等。

人类是信息资源产生的最重要因素。人是地球上最为重要的高等动物，自从有了人类社会就有了信息，每个人类个体和组织都是极其重要的信息源，自然界与人类活动无关的环境虽然也会产生大量的信息，但这是没有意义的，信息资源首先必须与人类相关才有意义。一个国家或地区的人口越多，所产生的信息必然越多。

社会活跃程度也是信息资源产生的重要因素。农业社会节奏缓慢，信息资源量相对较少，而工业社会和信息社会人们的工作和生活节奏加快，社会活跃程度高，必然产生大量信息。

教育水平既是信息资源的产生因素、也是影响因素。教育水平影响着人们的信息处理水平和信息敏感性，教育水平越高，对周围事物的敏感性增强，加工处理信息的水平越高，从而带来社会信息资源总量的增加。教育水平越高，拥有和掌握信息传播工具的程度越高，信息传播能力越强。

信息传播水平对信息的影响是至关重要的，如果信息传播水平落后，那么信息一般只能

局限在小范围内共享, 整个社会信息资源量很少, 社会节奏缓慢, 信息资源的作用受到限制。只有当信息传播水平高度发达时, 信息才能自由地传播到需要它的地方。信息传播并不能产生新的信息 (有时会扭曲信息, 但本文已经做了有关垃圾信息的假定), 但可以使更多的人拥有信息, 从而发挥信息资源的作用。当然, 如果将信息资源总量界定为所有组织和个人拥有信息的总和的话, 由于信息传播会使更多的人了解信息, 因此, 信息传播也能增加信息资源总量, 也是信息资源的产生因素。

2.2 基于 TOPSIS 的信息资源测度方法

TOPSIS 法 (Technique for Order Preference by Similarity to Ideal Solution) 根据各被评估对象与理想解和负理想解之间的距离来排列对象的优劣次序。所谓理想解是设想的最好对象, 它的各属性值达到所有被评对象中的最优值; 而负理想解则是所设想的最差对象, 它的各属性值都是所有被评对象中的最差值, 用欧几里德范数作为距离测度, 计算各被评对象到理想解及到负理想解的距离, 距理想解愈近且距负理想解愈远的对象越优。与简单的加权平均汇总相比, TOPSIS 更加强调评价指标体系的系统性, 强调指标体系总体水平的协调, 因此本文选用 TOPSIS 法对信息资源进行评价。TOPSIS 的计算过程如下:

1、建立指标矩阵并将数据标准化

设方案集 $A_i (i=1, 2, \dots, m)$ 在指标 $S_j (j=1, 2, \dots, n)$ 下取值为 a_{ij} , 得到指标矩阵 $A=(a_{ij})_{m \times n}$ 。由于方案的指标较多, 众多指标之间存在错综复杂的关系, 有正向指标 (指标值愈高, 能力愈强) 和逆向指标 (指标值愈高, 能力愈差) 之分, 且各指标的量纲不同, 为便于比较, 对指标矩阵进行标准化处理。

2、确定方案集的最优点和最劣点

令:

$$x_j^+ = \max \{x_{ij}\} \quad x_j^- = \min \{x_{ij}\} \quad (1)$$

(j=1, 2, \dots, n)

则:

最优点集为: $A^+ = (x_1^+, x_2^+, \dots, x_n^+)$

最劣点集为: $A^- = (x_1^-, x_2^-, \dots, x_n^-)$

满意方案就是在决策点集中找出离最优点集最近离最劣点集最远的决策点。

3、计算各方案的密切值

方案 A_i 在的密切值为:

$$C_i = d_i^+ / d^+ - d_i^- / d^- \quad (2)$$

其中:

$$d_i^+ = \left[\sum_{j=1}^n \omega_j (x_{ij} - x_j^+)^2 \right]^{1/2} \quad (3)$$

$$d_i^- = \left[\sum_{j=1}^n \omega_j (x_{ij} - x_j^-)^2 \right]^{1/2} \quad (4)$$

$$d^+ = \min \{d_i^+\}, \quad d^- = \max \{d_i^-\} \quad (5)$$

d_i^+ 、 d_i^- 分别表示方案 A_i 与最优方案 A^+ 、最劣方案 A^- 之间的欧氏距离, d^+ 、 d^- 分别表示 m 个最优点距的最小值和 m 个最劣点距的最大值。 ω_j 表示各指标的权重。

C_i 的大小反映了方案集偏离最优点的程度, 当 $C_i > 0$ 时, A_i 偏离最优点, 其值越大, 偏离越远; 当 $C_i = 0$ 时, A 最接近最优点。以 C_i 的大小作为决策准则, C_i 最小的方案就是满意方

案。

3 变量与数据

在信息资源的 4 个产生和影响因素中,人的因素可以用一个国家或地区的人口表示。教育水平可以采取平均受教育年限表示。社会活跃程度采用地区国民生产总值 (GDP) 表示,因为通常情况下,社会活跃程度越高,经济越发达。

信息传播水平主要有电话 (固定电话、移动电话) 和互联网 (网民、网站等) 指标,虽然在统计年鉴上可以查到相关数据,但采用电信营业额表示信息传播水平更加合理。因为电话和互联网在某种程度上可以互相替代,在信息传播过程中,不同传播手段之间的替代是非常普遍的现象,比如某人发电子邮件后一般就不需要再电话联系。近年来不少地区信函的投递量是下降的,根本原因就在于电话和互联网对信函的替代,根据经济学中经济人假设,一般人们不会为了同一信息而使用不同的信息传播工具。之所以不选用邮电业务额作为信息传播的反映指标,主要原因是邮政数据中的包裹邮递不属于信息传播,另外许多快递公司的信函数据也无法统计。

本文所有数据均来自于 2007 年中国统计年鉴,实际上所有数据为 2006 年的数据,其中平均受教育年限是根据统计年鉴的数据加权计算而来的。所有数据都进行了标准化,方法是将指标值除以最大值再乘以 100。

指标权重采取专家打分法确定,传播水平的权重最高,为 0.5,因为当代信息社会赖以存在的基石就是传播水平,方便快捷的传播水平彻底改变了人们的工作与生活方式,社会活跃程度的权重为 0.3,人口与教育水平的权重均为 0.1。

4 实证结果

表 1 信息资源评价结果

地 区	人口	平均受 教育年限	电信 营业额	国内生产 总值	TOPSIS 得分×100	排序
广 东	9304	8.44	2587.66	26204.47	93.13	1
山 东	9309	8.09	917.36	22077.36	58.10	2
江 苏	7550	8.25	934.85	21645.08	56.90	3
浙 江	4980	8.06	997.09	15742.51	48.85	4
河 南	9392	8.05	680.82	12495.97	44.05	5
河 北	6898	8.13	608.15	11660.43	38.92	6
四 川	8169	7.24	563.51	8637.81	35.76	7
上 海	1815	10.44	509.89	10366.37	32.12	8
辽 宁	4271	8.92	532.00	9251.15	31.75	9
湖 南	6342	8.17	464.48	7568.89	30.86	10
湖 北	5693	8.26	429.39	7581.32	29.44	11
福 建	3558	7.73	604.02	7614.55	28.67	12
北 京	1581	10.95	468.39	7870.28	28.63	13
安 徽	6110	7.34	331.98	6148.73	26.41	14
黑 龙 江	3823	8.53	380.25	6188.90	24.43	15
广 西	4719	8.03	372.19	4828.51	23.54	16
陕 西	3735	8.30	404.12	4523.74	22.32	17
江 西	4339	7.71	296.34	4670.53	21.36	18
山 西	3375	8.70	306.83	4752.54	21.27	19

云 南	4483	6.66	317.78	4006.72	20.07	20
吉 林	2723	8.66	321.16	4275.12	19.94	21
天 津	1075	9.73	211.81	4359.15	19.10	22
内 蒙 古	2397	8.19	244.35	4791.48	18.53	23
重 庆	2808	7.57	263.77	3491.57	16.78	24
新 疆	2050	8.30	219.39	3045.26	15.91	25
贵 州	3757	6.59	215.44	2282.00	15.50	26
甘 肃	2606	6.78	159.13	2276.70	12.72	27
海 南	836	8.17	93.04	1052.85	11.76	28
宁 夏	604	7.63	57.09	710.76	9.96	29
青 海	548	6.99	37.94	641.58	8.23	30
西 藏	281	4.16	20.39	291.01	0.70	31

评价结果如表 1 所示, 信息资源较高的地区有广东、山东、江苏、浙江、河南等省, 较低的有西藏、青海、宁夏、海南、甘肃等省区。与人们的印象不太一致的是北京作为首都, 其信息资源量仅排在全国的第 13 位, 主要原因是首都作为全国的政治中心, 难以用指标来进行反映, 将首都与其他普通省市进行比较, 确实存在信息资源低估的可能性, 如何进行修正还需要进一步研究, 当然如果是进行国家间的比较则肯定没有这个问题。

一些中西部省份如河南、河北、四川信息资源量也较高, 主要有两个原因, 第一是由本文对信息资源的界定所引起的, 本文所指的信息资源主要是与人们生活和工作密切相关的信息资源, 大量图书文献信息资源应用范围主要集中在少数专业人员, 对大众而言这些信息资源更多地是一种“矿藏”, 公共图书馆借阅率较低是普遍现象; 第二是这些地区人口较多, 经济总量较大, 随着国家信息化的发展, 数字鸿沟不断缩小, 传播水平总体也较高, 此外, 地区间的教育水平也相差不大。

5 结论与讨论

5.1 信息资源间接测度法是一种较好的方法

采用反映信息资源量的直接指标进行测度, 由于指标众多, 指标选择分歧较大。此外数据不易获取, 必然增加了评价的成本, 而且难以测度大量无法统计和记录的人类活动信息。采用替代变量进行测度虽然简单方便, 但失去许多信息, 容易导致较大的误差。因此信息资源间接测度法是较好方法, 克服了以上两种方法的缺点。

5.2 信息资源间接测度法尚需进一步优化

采用本文的测度方法, 对作为政治中心首都的测度偏低, 需要采取一定的方法加以修正, 具体的方法尚需进一步研究, 当然, 如果是进行国家间的比较则没有这个问题。此外, 建议统计部门加强邮电业务额的进一步统计, 这样就可以用用邮电业务额减去包裹营业额再加上快递公司信函营业额来代替电信营业额, 从而可以进一步优化本文的测度模型。

5.3 注意总量指标与人均指标测度的不同

在一些信息化指标体系的相关测度中, 不少学者将总量指标与人均指标混用, 如国家信息化评测中心选取的指标大部分是人均指标, 但也用到长途光纤长度、电子商务交易额等绝对指标, 这是不恰当的。如果是采用信息资源间接测度法, 选取人均指标进行测度, 那么可以选用人均 GDP、平均受教育年限、人均电信营业额 3 个指标进行测度即可, 很显然, 与总量指标测度相比, 北京这 3 项指标的排名将会靠前, 当然最终评价结果排名也将大大靠前, 或许可以解释采用信息资源间接测度法北京排名靠后的主要原因, 因为我们看到的许多其它测度方法大部分都是采用的人均指标。究竟是采用总量指标还是人均指标进行测度关键看研究目的。

5.4 信息资源间接测度法是一种启发式方法 (heuristic)

管理者碰到的问题,大多数都很复杂,如果要进行严密精确的分析会步履艰难,很难用数学的方法来求解,但又不得不寻求答案,这时,凭借经验法则形成的逻辑依据就不失为可用的最佳方法,或许它不一定是绝对精确的,但这种简单的法则是适用的,信息资源的间接测度就是启发式方法的运用。

[1]Machlup F. The production and distribution of Knowledge in the United States. Princeton University Press, New Jersey, 1962

[2]Porat M. U. C. The information Economy. Washington, DC, Government Printing Office, 1977

[3]小松崎清介等著,李京文等译. 信息化与经济发展[M]. 北京: 社会科学文献出版社, 1994: 39.

[4]谢康, 肖静华. 信息资源测度、国际比较与中国的战略选择[J]. 情报杂志, 1997(3):202-211

[5]俞立平. 基于 Panel data 的信息资源对经济增长贡献的研究[J]. 现代图书情报技术, 2005(12): 74-77