

基于DEA的政府开放数据网站效率评价*

马海群¹, 王今²

(1.黑龙江大学信息资源管理研究中心, 哈尔滨 150080; 2.黑龙江大学信息管理学院, 哈尔滨 150080)

摘要: 基于对已有网站效率评价和政府开放数据评价的研究, 以我国8个具有代表性的政府开放数据网站为例, 通过因子分析得出适当的输入和输出指标进行DEA分析, 从而对政府开放数据网站总体效率、纯技术效率和规模效率进行评价。结果表明, 各地政府开放数据网站的效率存在差异, 相关部门应当对症下药, 全面提升我国政府开放数据网站的建设水平。

关键词: 开放数据; DEA分析法; 网站评价; 效率评价

中图分类号: G350

DOI: 10.3772/j.issn.1673-2286.2016.6.001

1 引言

政府部门在履行行政职能、管理社会公共事务的过程中采集和储存了大量数据。政府数据网站的开放, 不但能帮助政府提高自身管理透明度, 还能够提升治理能力和效率, 这也是促进经济发展和社会创新的重要推动力。自2014年, 上海市政府数据资源向社会开放工作计划, 建立第一个政府开放数据网站以来, 我国政府开放数据网站如雨后春笋般萌芽, 并得到政府的重视。如何对政府开放数据网站进行评价, 是发展政府开放数据必须要解决的问题, 对此国内外的政府部门、研究机构和大学等均进行了有益的尝试。

黄思棉等以上海市和北京市为研究对象, 通过人工的方式收集了这两个网站的数据项个数、数据下载量、网页更新时间和频率等数据集^[1]。通过分析得出, 政府开放数据网站存在更新时间较长、数据的价值性小且涵盖的范围比较窄等问题, 并据此提出改进意见。徐慧娜等以纽约市和上海市的开放政府数据平台为例, 引用技术接受模型, 收集了13个指标, 比较了两大城市政府开放数据网站的有用性、易用性和用户数据利用效果, 分析了影响开放政府数据效果的关键因素^[2]。卫军朝等全面分析了上海市政府开放数据建设的基础背景、数据开放情况, 并对比了国际政府开放数据建设

的现状^[3]。陈涛等以武汉市政府数据开放平台建设为例, 采取案例研究法, 对武汉市正在筹建的政府数据开放平台开展实证研究, 分析了武汉市政府数据开放平台建设的政策环境、建设目标、内容以及面临的挑战^[4]。已有的研究主要集中于对开放数据网站的案例研究, 一般会选取北京市、上海市这种成立时间较早, 运营时间较长的网站作分析, 少数研究者会关注到所有已经开放的政府开放数据网站, 但仅抓取主要几项网站数据进行简单的描述性统计分析。本文基于前人网站评价体系和开放数据评价体系的研究, 将DEA分析方法应用于开放数据网站效率的评价。通过因子分析进行指标的整合、确立, 并利用相关效率模型进行分析, 从而构成系统性的网站效率评价。

2 样本与指标选取

2.1 DEA简介

数据包络法(Data Envelopment Analysis, DEA)是著名运筹学家Charnes和Copper等学者以“相对效率”概念为基础, 根据多指标投入和多指标产出对相同类型的单位(部门)进行相对有效性或效益评价的一种新的系统分析方法^[5]。它是处理多目标决策问题的

* 本研究得到国家社科基金重点项目“开放数据与数据安全的政策协同研究”(编号: 15ATQ002)资助。

方法, 主要应用于处理多决策目标的情况, 通过数据规划模型计算并比较决策单元间的效率, 从而对决策对象作出评价。

2.2 样本选取

我国的数据开放平台主要包括大型公共政府数据库的开放数据网站(如天气数据、路程导航数据、普查人口数据、证券交易信息数据、医疗数据等)和各级地方政府的公共数据开放网站。考虑到各级地方政府数据开放网站往往采用专门的数据网站形式, 网站内容相对统一, 便于社会公众对数据进行访问和下载。本文将研究范围限定为地方政府的开放数据网站, 通过搜索引擎以“数据+gov.cn”和“data+gov.cn”为关键词进行查询, 以政府数据信息公开内容为主, 选取8个符合研究条件的政府开放数据网站作为研究样本, 结果见表1。

表1 样本网站

名称	网址
北京市政务数据资源网	www.bjdata.gov.cn
上海市政府数据服务网	www.datashanghai.gov.cn
无锡市政府数据服务网	http://opendata.wuxi.gov.cn/
青岛政务网	http://data.qingdao.gov.cn/data/
武汉市政府公开数据服务网	www.wuhandata.gov.cn
湛江数据服务网	data.zhanjiang.gov.cn
浙江政务服务网	http://data.zjzfw.gov.cn/
贵阳大数据开放服务平台	www.datagy.cn/

2.3 指标选取

从效率的角度评价网站, 很大程度上依赖于评价指标的选取是否有效、合理。万维网基金会所组织开展的“开放数据晴雨表”(Open Data Barometer)基于“准备度”“执行度”和“影响力”层面进行评价^[6]。郑磊等基于“开放数据晴雨表”建立的评价指标包含组织和管理、数据数量、开放性、元数据、时效性、易用性、数据导引、数据获取、数据应用、互动交流及界面体验^[7]。闫现洋等在评价网站效率的研究中总结了4个维度: 网站内容、网站易用性、网站影响度、网站结构, 共包含9个可测的指标: 内容丰裕度、网页的更新率、网站的连通率、下载速度、网站的总页面数、网站页面的平均大小、平均日点击访问量、外部网站指向链接的数及网络外链

接数^[8]。已有研究鲜有学者对政府开放数据网站给出较为完善的评价指标体系, 只是从政府开放数据的角度进行研究。因此, 本文基于以上政府开放数据评价体系的特征, 并结合网站效率评价的特点从内容、技术、网站结构等角度选取相应的指标, 构建评估政府开放数据网站效率的9个测度指标, 结果见表2。

其中数据量、数据内容、连通率、下载速度、网站的总页面数、网站的不重复链接数为输入指标, 下载量、访问量、网站外链接量为输出指标。网站内容指标主要考虑网站数据内容涵盖范围是否广泛; 数据格式、编排是否清晰等因素, 采用专家调查法和层次分析法进行定量和定性相结合的研究, 最终得出网站的排名和分数。网站的下载速度利用站长工具测量。数据量、下载量和访问量通过火车采集器进行网站后台的数据抓取。网站连通率、网站的总页面数、网站的不重复链接数量通过Maxamine web analyst软件获取。网站外链接量通过搜索引擎bing抓取。本文于2016年5月3日13:00点, 同时抓取上述所需要的分析数据。由于所抓取的数据量纲不同, 不具有可比性, 因此要进行标准化处理。

3 数据处理

首先应用spss统计软件, 对上文采用Maxamine web analyst 软件、搜索引擎bing、火车采集器、站长工具和打分方法采集的DEA输入输出指标数据进行标准化处理。本文采用spss中最常用的“Z”方法, 该方法是将每一变量值与其平均值之差除以该变量的标准差。通过这种标准化处理可以消除量纲和数量级的影响, 从而去除数据的单位限制, 将其转化为无量纲的纯数值。政府开放数据网站DEA输入输出指标数据标准化结果见表3。

DEA法在分析过程中, 输入指标和输出指标的乘积最好低于决策个数, 因此通过因子分析对输入输出指标进行降维处理, 使原有的评价指标聚类为有代表性的体系指标。首先采用KMO(Kaiser-Meyer-Olkin)和Bartlett对是否适合进行因子分析进行检验。KMO检验用于研究变量间的简单相关系数平方和与这些变量间偏相关系数平方和的差值。KMO值越接近于1, 越适于因子分析; KMO值过小, 则不适合。本研究中的KMO检验输入变量值为0.638、输出变量值为0.756, 二者都比较适合进行因子分析。

表2 政府开放数据网站效率评价指标

逻辑维度	指标名称	指标内容
输入指标	数据量	数据量主要包括网站的网页、图像及文档，其数量多少反映了网站的丰富度及投入量，对象越多投入量越大
	数据内容	数据内容主要包括：涵盖范围是否广泛，数据可达性，数据格式、编排是否清晰，界面友好程度。专家打分的结果反应一个网站浏览的舒适度和网站运营效果
	连通率	网站连通率反映了网站所有页面的可达性及网站导航设计的完整性，即从主页出发，通过各层链接能够到达的页面数占所有页面数的比率，连通率越高则网站越完整，相应的投入会越大
	下载速度	下载速度是指用户通过搜索引擎或外部指向网站的链接等访问网站的平均响应时间。访问速度越快，则网站的体验越好
	网站的总页面数	网站的总页面数主要反映网站的总体规模大小和内容丰富程度，页面越多，内容越丰富
	网站的不重复链接量	链接量反映页面内容的丰富程度及形式的多样化程度。其数值越大，说明内容越丰富，形式越多样化
输出指标	下载量	下载量越大，则说明页面吸引力越强，越受到用户的亲睐和认可，网站的影响力越强
	访问量	访问量即平均每天访问该网站的用户数。访问量越大，说明网站的用户数越多，网站的影响力越大
	网站外链链接量	网站总链接数包含网站的内部链接和网站的外部链接，链接数越多反映网站的规模越大

表3 政府开放数据网站DEA输入输出指标数据标准化结果

标准化结果	数据量	数据内容	连通率	下载速度	网站的总页面数	网站页面的平均大小	下载量	访问量	网站外链链接量
北京市政务数据资源网	-0.229 59	1.797 43	-0.656 18	-0.959 84	1.913 12	1.573 44	0.625 82	0.764 49	0.893 91
上海市政府数据服务网	1.163 07	-0.898 72	0.327 40	1.683 66	-0.839 57	-0.791 38	2.273 87	2.187 27	2.133 61
无锡市政府数据服务网	1.018 42	-0.599 14	0.539 51	-0.078 68	-0.589 33	-0.913 53	-0.332 05	-0.534 82	-0.530 62
青岛政务网	-0.044 57	0.299 57	0.539 51	-0.975 58	0.351 59	1.112 01	-0.519 98	-0.376 60	-0.382 04
武汉市政府公开数据服务网	1.159 70	0.299 57	-2.261 95	0.849 70	0.211 46	0.518 26	-0.564 11	-0.725 64	-0.701 58
湛江数据服务网	-1.211 84	0.898 72	0.539 51	-1.054 25	0.731 97	0.233 26	-0.487 43	-0.693 85	-0.672 49
浙江政务服务网	-1.030 19	-0.898 72	0.484 52	0.660 88	-0.909 64	-0.910 13	-0.446 66	-0.272 93	-0.336 83
贵阳大数据开放服务平台	-0.825 00	-0.898 72	0.487 66	-0.125 88	-0.869 60	-0.821 92	-0.549 46	-0.347 92	-0.403 96

巴特莱特球体检验 (Bartlett test of Sphericity) 的目的在于检测相关矩阵是否为单位矩阵，若显著性概率Sig小于0.05，即否定相关矩阵为单位矩阵的零假设，表明适合因子分析^[9]。经过分析，研究中的Bartlett

球型检验输入变量和输出变量统计量的显著性概率是0.001、0.000，都小于0.05，表明适合进行因子分析，因此可以提取公共因子。确定公因子的一般原则是选取spss因子分析结果中，解释的总方差中特征值大于1的

值。从表4的分析结果可以看出, 输入变量的前两个因子特征值均大于1, 累积贡献率达到88.676%, 较为充分地保留了原始变量信息, 因此提取这两个主因子。输出变量只有第一个因子的贡献率大于1, 因此提取这个主

因子。

使用DEA方法进行效率评价要求全部投入产出的数据非负, 因此对新指标数据全部进行取e的指数处理, 得到输入输出指标数据的非负处理结果, 如表5所示。

表4 解释的总方差

成分		原始特征值			提取平方和载入		
		合计	方差/%	累积/%	合计	方差/%	累积/%
输入变量	1	3.548	59.138	59.138	3.548	59.138	59.138
	2	1.772	29.538	88.676	1.772	29.827	88.676
	3	0.422	7.026	95.702	-	-	-
	4	0.149	2.490	98.192			
	5	0.106	1.764	99.955			
	6	0.003	0.045	100.000			
输出变量	1	2.977	99.219	99.219	2.977	99.219	99.219
	2	0.021	99.934	99.934	-	-	-
	3	0.002	0.066	100.000			

表5 政府开放数据网站主成分数据及非负化处理结果

网站	输出指标	输入指标		E指标处理结果		
	F1	E1	E2	F1	E1	E2
北京市政务数据资源网	0.765	1.740	0.059	2.148	5.695	1.061
上海市政府数据服务网	2.207	-1.074	0.880	9.086	0.342	2.410
无锡市政府数据服务网	-0.468	-0.666	0.119	0.626	0.514	1.126
青岛政务网	-0.428	0.585	-0.501	0.652	1.795	0.606
武汉市政府公开数据服务网	-0.667	0.434	1.933	0.513	1.544	6.912
湛江数据服务网	-0.621	0.655	-1.136	0.538	1.926	0.321
浙江政务服务网	-0.353	-0.923	-0.596	0.702	0.397	0.551
贵阳大数据开放服务平台	-0.435	-0.751	-0.758	0.647	0.472	0.469

4 效率模型应用分析

目前学术界应用最为广泛成熟的, 也是最基本的两个DEA模型是CCR模型和BCC模型, 其中CCR模型可以评估综合技术效率, 也称综合效率。综合技术效率=纯技术效率×规模效率。BCC模型可以评估纯技术效率。纯技术效率是指在给定投入的基础上, 由于管理和技术等因素影响的生产效率, 规模效率是由于规模因素影响的生产效率^[10]。本文采用CCR模型和BCC模型对网站的综合效率、纯技术效率和规模效率进行评价, 由此可发现网站存在的问题以及与其他相同类别网站

间的差距, 从而提出改进意见, 提高效率。

将8个样本的各输出、输入指标数据导入DEAP2.1软件, 进行政府开放数据网站相对效率分析。对开放政府数据网站指标进行输入、输出的CCR模型和BCC模型分析, 得到综合效率、纯技术效率和规模效率的输出结果见表6。

从表6可见, 这8个政府开放数据网站发展得非常不均衡, 总体来看综合效率偏低。但其中上海市的综合效率达到满分值, 可见发展水平较高。作为起步最早的数据开放平台, 上海市政府数据服务网领跑所有的政府数据开放网站, 值得学习和借鉴。另外, 北京市和湛

表6 政府开放数据网站效率 %

网站名称	综合效率	纯技术效率	规模效率
北京市政务数据资源网	0.537	0.674	0.797
上海市政府数据服务网	1.000	1.000	1.000
无锡市政府数据服务网	0.148	0.756	0.195
青岛政务网	0.285	0.673	0.424
武汉市政府公开数据服务网	0.020	0.236	0.083
湛江数据服务网	0.444	1.000	0.444
浙江政务服务网	0.338	1.000	0.338
贵阳大数据开放服务平台	0.366	1.000	0.366
平均值	0.392	0.792	0.456

江市的数据开放网站效率也超过平均水平。北京市作为我国的首都，其网站建设应该具有雄厚的实力，但是网站综合效率来看，并不是非常令人满意。造成这一现象的原因可能是北京市政府开设的相关部门政府网站过于冗杂，在技术的应用和拓展方面并不能做到均衡，因此对于开放时间较晚的北京政府开放数据网站没有足够的重视和投入，使得纯技术效率低于平均水平，从而整体综合效率受到影响。相对而言，湛江市作为一个地级市政府的网站，它的开放数据网站在所有开放数据网站的排名比较靠前值得关注。不难发现，湛江市数据服务网数据下载流程方便快捷，网站所发布的数据范围广且数据量大等，这些都可以看出湛江市政府在开放数据网站方面的重视程度。从纯技术的角度出发，抛开上海市政府数据服务网来看，湛江数据服务网、浙江政务服务网、贵阳大数据开放服务平台的纯技术效率也达到1.000，说明在投入不变的情况下，它们的管理和技术水平都较为突出。若想在今后全面地提高网站建设效率，需要注重规模效率的改进。最后从规模效率的角度入手，北京市政务数据资源网、上海市政府数据服务网位于前列，其他网站规模效率值均低于平均水平。因此，对于其他政府开放数据网站而言，要提高综合效率，需要着手整体网站的投入资源配置问题，进一步优化整合已有的资源。

5 结论及建议

本文采用DEA法对我国为数不多的8个具有代表性的政府开放数据网站进行效率评价和比较。较之以往单纯运用描述性统计来横向比较网站的发展而言，

避免了衡量角度单一化、片面的问题，提出了一个全新的视角。这样不仅能够对网站的效率水平进行综合比较，还能从技术效率和规模效率入手，分析现存状况的不足和待改进的地方，对其今后提高运行效率，改善运行效果提出合理化的建议，从而不断地推进开放数据的信息公开，使我国政府更好地为民服务。因此结合以上分析，认为可以从如下方面提高政府开放数据网站的效率。

首先，政府相关部门应该加强管理，注重网站技术水平的提高，使网站用户能够切实感受到整体网站的服务水平。例如，当网站的连通率提高，用户可以在第一时间进入网站；当网站的下载速度加快，用户可以及时获取所需的开放数据资料等。因此，为确保政府开放数据网站的效率，要提高相关技术人员的素质，及时进行网络技术的更新，从而提升网站的整体建设。

其次，提高政府开放数据网站的资源配置效率。政府开放数据网站作为近几年才得以发展的网站，在经费、技术和控制信息等方面都有所局限，因此要提高网站的整体效率，需要尽量实现资源的最佳配置。要在逐渐扩大网站规模的基础上，注重整体的规模效率，使政府开放数据网站的效率影响因素之间优势互补，产生系统效应，实现网站的总体效率提高。

最后，要针对不同网站建设出现的问题，对症下药。应依据各地区经济、文化和政府开放程度等状况的不同，制定相应的调整方案。例如，对于某些发展水平较低的地级市，要注重整体网站建设的资源配置，注重规模效率。而像北京市、上海市等经济发展水平较高的城市，则更需要通过增加财力投入，全面提高技术水平和扩大网站规模。

参考文献

- [1] 黄思棉,张燕华.当前中国政府数据开放平台建设存在的问题与对策研究——以北京、上海政府数据开放网站为例[J].中国管理信息化,2015,18(14): 175-177.
- [2] 徐慧娜,郑磊.面向用户利用的开放政府数据平台:纽约与上海比较研究[J].电子政务,2015(7): 37-45.
- [3] 卫军朝,蔚海燕.上海推进政府开放数据建设的路径及对策[J].科学发展,2014(11): 80-88.
- [4] 陈涛,李明阳.数据开放平台建设策略研究——以武汉市政府数据开放平台建设为例[J].电子政务,2015(7): 46-52.
- [5] 杨选辉,葛伟.基于DEA和外部认可的中文军事信息网站效率评价[J].情

- 报杂志, 2014(10): 184-189.
- [6] DAVIES T, SHARIF R, ALONSO J. Open data barometer-global report 2th ed[R/OL]. [2016-06-12]. <http://opendatabarometer.org/assets/downloads/Open%20Data%20Barometer%20-%20Global%20Report%20-%202nd%20Edition%20-%20PRINT.pdf>.
- [7] 郑磊, 高丰. 中国开放政府数据平台研究: 框架、现状与建议[J]. 电子政务, 2015(7): 8-16.
- [8] 闫现洋, 余小萍, 吴江. 基于DEA方法的高校图书馆网站效率评价研究——基于6所部属师范高校的实证分析[J]. 图书情报工作, 2011(23): 68-71.
- [9] 陈胜可. SPSS 统计分析从入门到精通[M]. 北京: 清华大学出版社, 2010: 317.
- [10] 汤志伟, 韩啸, 龚泽鹏. 基于数据包络分析的县级政府网站效率评价[J]. 党政研究, 2015(3): 83-86.

作者简介

马海群, 男, 1964年生, 黑龙江大学信息资源管理研究中心, 教授, 博士生导师, 研究方向: 信息政策与法律, E-mail: mahaiqun@sina.com.cn。
王今, 女, 1992年生, 黑龙江大学信息管理专业, 在读研究生, 研究方向: 信息管理, E-mail: 2914965387@qq.com。

Efficiency Evaluation about Government Open Data Website Based on DEA

MA HaiQun¹, WANG Jin²

(1. Research Center of Information Resources Management, Heilongjiang University, Harbin 150080, China; 2. College of Information Management, Heilongjiang University, Harbin 150080, China)

Abstract: Based on the analysis of previous studies, this paper, taking 8 government open data websites as a sample, obtains appropriate input and output parameters through factor analysis to make DEA analysis, there for to evaluate the overall efficiency, pure technical efficiency and scale efficiency of the government open data website. Result shows that there are some differences in the efficiency of those government open data sites. The relevant departments should find specific ways so that government open data websites will be improved.

Keywords: Open Data; Data Envelopment Analysis; Website Evaluation; Efficiency Evaluation

(收稿日期: 2016-05-30)

■ 书 讯 ■

《汉语主题词表》(工程技术卷)

《汉语主题词表》自1980年问世以后, 经1991年进行自然科学版修订, 在我国图书情报界发挥了应有的作用, 曾经获得了国家科学技术进步二等奖。为了适应网络环境下知识组织与数据处理的需要, 2009年由中国科学技术信息研究所主持, 并联合全国图书情报界相关机构, 完成《汉语主题词表(工程技术卷)》的重新编制工作。

全书共收录优选词19.6万条, 非优选词16.4万条, 等同率0.84。在体系结构、词汇术语、词间关系等方面进行改进创新。为了方便工程技术领域不同专业用户使用, 《汉语主题词表》(工程技术卷)按专业分13个分册出版, 同时建立《汉语主题词表》服务系统, 提供在线概念检索和辅助标引服务, 通过可视化技术展示各类概念关系, 是图书馆、档案馆、出版社、期刊杂志社、文献信息中心等专业工作者及科研、教育及工程技术领域人员必备的参考书。

《汉语主题词表(工程技术卷)》已于2014年由科学技术文献出版社出版, 全书2 300余万字, 总定价3 880元, 可分册购买。