

知识管理论文著者科学生产率统计分析

伍玉成¹, 刘小乐^{2,3}, 马捷^{2,3}

(1. 长春理工大学图书馆, 吉林 长春 130022;

2. 吉林大学管理学院, 吉林 长春 130022; 3. 吉林大学信息资源研究中心, 吉林 长春 130022)

摘要: 对论文著者的科学生产率进行简单的概述, 以知识管理专题论文为例研究论文著者科学生产率状况, 并将研究的结果与论文著者的洛特卡定律进行比较分析, 得出知识管理专题论文著者的科学生产率情况, 进而得出"知识管理"领域的研究状况。

关键词: 科学生产率; 洛特卡定律; 知识管理; K-S检验; 论文著者

中图分类号: G350 文献标识码: A 文章编号: 1007-7634(2012)01-21-04

Statistical Analysis on Scientific Productivity of Knowledge Management Thesis Authors

WU Yu-cheng¹, LIU Xiao-le^{2,3}, MA Jie^{2,3}

(1. Library of Changchun University of Science and Technology, Changchun 130022, China;

2. School of Management, Jilin University, Changchun, 130022, China;

3. Center for Studies of Information Resources of Jilin University, Changchun, 130022, China)

Abstract: This article mainly summarized scientific productivity of thesis authors and researched scientific productivity situation. At the same time, the thesis carried on comparative analysis between the results of the research and Lotka's law of thesis authors. Through the analysis we can get scientific productivity situation of knowledge management papers and research status of knowledge management subject.

Keywords: scientific productivity, Lotka's law, knowledge management, K-S test, thesis author

1 引言

数字图书馆的不断发展带动了学术研究的热潮。伴随着知识经济时代的到来, 学科知识不断更新, 新的学科不断涌现, 学科之间的交叉性不断增强, 学者对于新学科的研究越来越多。现代研究表明, 科学文献的数量与质量是科学研究的重要表现。衡量一门科学的发展的重要指标主要有两个: 一是发表的这门学科文献数量; 二是发表这些文献的科学家即论文著者的科学生产率^[1]。可以从文献计量角度分析论文著者与发表论文数量的关系, 分析论文著者科学生产率模型, 从而得出学科的洛

特卡分布。本文首先对科学生产率进行了简单概述, 构造了知识管理专题论文著者的科学生产率模型, 验证洛特卡定律, 最终得出知识管理学科的研究状况。

2 科学生产率

2.1 科学生产率概念

1926年, 科学生产率(Scientific Productivity)概念由美国情报学家洛特卡首次提出, 它主要是指科研人员在科学上所表现出的能力和工作效果, 与劳动生产率有共通之处, 但是本文所讲的科学生产率

收稿日期: 2011-09-22

作者简介: 伍玉成(1958-), 男, 吉林长春人, 馆员。

通常由著者生产的科学文献的数量也就是文章作者发表论文篇数的多少来衡量^[2]。科学生产率概念与经济领域里的劳动生产率概念相比有着明显的区别。科学生产率主要针对的是论文著者所著的科学文献的数量,并且科学生产率具有一定的科学性和合理性,能够形成定律来揭示科学文献作者与所著文献数量之间的关系。然而劳动生产率针对的是劳动者生产单位产品所消耗的劳动量。

2.2 研究科学生产率的意义

首先,洛特卡提出的科学生产率概念与洛特卡定律紧密相关。运用洛特卡定律可分析出某一学科领域论文著者的科学生产状况。

其次,通过研究论文著者的产出状况,验证学科领域是否符合洛特卡定律,并且根据洛特卡定律得出的参数值可以揭示出论文著者与所著文献数量的关系,从而更加清晰的区分本学科专题领域的高产作者以及低产作者,从而能够进一步描述出某一学科领域的科学生产率分布状况。

最后,研究科学生产率可以认识该学科或主题的科学生产规律以及学科的学术层次^[3],根据研究结果可以有针对性的与其他学科专题进行合作或者是进行交叉性的研究。

总之,研究科学生产率不仅能够扩展和丰富学科专题论文的科学生产率分布定律的内涵和应用范围^[4],得出专题论文的科学生产率状况。同时能够分析科学文献著者的科学生产率分布规律,有助于我们全面了解和掌握知识规律。

3 论文著者科学生产率分析

随着科学技术的进步及学科的发展,学科的分支逐渐细化,并且出现学科交叉现象。知识管理是信息管理的发展和延伸,随着知识经济时代的到来,信息技术使得信息转化为知识的速度得到了提高和进步,越来越多的学者开始研究知识管理。知识管理运用到高校、企业、个人生活管理的多个方面。知识管理论文著者的科学生产率也逐步提高。谭观音、左泽平研究虚拟企业的知识管理模型;陈跃飞研究了知识管理在企业管理中应用;郎庆华在基于知识管理的机构知识库服务体系构建分析中,研究了知识服务系统的知识管理现状;黄忠、闫兰兰基于Web2.0对大学生个人知识管理研究。由此可见,许多学者投入到知识管理的研究中,越来越多的学者

由信息管理转向知识管理研究,同时知识管理专题覆盖企业、高校、个人等,覆盖范围较广。研究其科学生产状况,对于完善学科发展有重大的意义。

3.1 研究思路

本文在研究知识管理专题论文著者的科学生产状况时主要是依据中国期刊全文数据库的数据信息,首先了解论文著者所著论文的数量情况,其次在收集的数据基础上进行分析,验证洛特卡定律,根据洛特卡定律的函数参数值揭示出这一专题论文著者与所著文献数量之间的关系。

研究知识管理专题论文著者的科学生产状况的方法主要是收集知识管理论文,通过 excel 表格分析论文著者所著论文的数量,统计出知识管理论文著者的科学文献生产数量,得出论文的著者分布规律。

3.2 数据收集

对知识管理论文数据的收集是进行科学生产率研究的前提。本文主要是利用中国知网中的中国学术期刊网络出版总库作为论文知识库即源数据库。检索时间为2011年1月。由于在企业管理、图情档、信息科学等学科领域都涉及到了知识管理专题,因此学科领域不再进一步进行限制选择全部学科领域。在检索过程中利用知识管理检索词进行关键词检索。年限限制在2006-2010年近五年来的数据。图1是检索结果第一页面的显示。

序号	题名	作者	刊名	年期	被引 频次	下载次数
1	知识管理在高等院校管理中的应用性研究	韩文德, 胡西厚	现代情报	2010/12	45	
2	奇瑞企业知识管理应用探讨	杨斌	现代情报	2010/12	101	
3	知识管理的内涵研究	刘耀	价值工程	2010/06	65	
4	知识管理课程内容的需求调查分析	廖小平	图书情报论坛	2010/06	58	
5	知识管理标准的探讨	魏书旺, 李永, 冯士清	图书情报工作	2010/20	86	
6	国内外知识管理流程研究述评	魏书旺, 殷春侠, 陈英	情报理论与实证	2007/06	16	739
7	基于知识管理理念的图书馆服务工作	金凤	农业图书情报学刊	2010/12	18	
8	论知识管理研究重心的转移	李兴霖, 李爱华, 张玲玲	当代经济管理	2010/12	76	
9	基于组织文化的知识管理标准化	张艳理, 岳源峰, 顾立强	标准科学	2010/11		
10	从知识管理思想的起源探讨知识管理的概念	刘丽斌	图书情报杂志	2008/06	4	443
11	图书馆学视野下的知识管理研究	王军, 王强, 周丽霞	情报资料工作	2010/03	1	101
12	国外图书馆情报学领域知识管理的可视化分析	何南洋, 金毅	图书情报工作	2010/22		129
13	有效实施知识管理的基础研究	曹芳莲	情报学刊	2008/02		123

图1 知识管理论文检索结果首页

3.3 数据筛选

由于研究的是知识管理专题论文著者的科学生产率分布,所以对于检索出的导言、书评、会议综述、团体作者所著文献等不作为学科论文统计。由于核心期刊与一般期刊的不同,依照洛特卡定律的最初原则,本次论文数量只统计第一作者,将数据检索结果数据导入到Excel表中,利用电子表格的函数,从

而统计出了2088名作者所著的2642篇论文的分布情况。由于洛特卡定律较符合低产作者,因此的十篇及十篇以上的知识管理论文文献以及论文著者一起统计。

3.4 数据统计

通过数据收集及数据筛选的结果对2006-2010年知识管理专题论文的著者分布进行数据统计分析,如表1所示

表1 2006-2010年知识管理专题论文的著者分布

论文量x	作者量y _x	作者论文量	作者所占比例
1	1767	1767	0.8463
2	212	424	0.1015
3	60	180	0.0287
4	26	104	0.0125
5	9	45	0.0043
6	5	30	0.0024
7	3	21	0.0014
8	1	8	0.0005
9	1	9	0.0005
10及以上	4	54	0.0019

3.5 结果分析

由图表进行分析,2006-2010年间知识管理专题论文的科研人员发表一篇论文的作者占总数的84.63%,越来越多的学者从事知识管理研究。发表两篇到四篇的作者所占比例也相对较高。而四篇以上的占少数,因此高产作者较少,比例只占1.1%。发表一篇或两篇知识管理专题论文的著者占所有著者总数的94%以上。

4 研究结果与洛特卡定律比较分析

本章节主要是对数据统计分析的结果即实际的科学生产情况与洛特卡定律进行比较分析,验证知识管理论文著者的科学生产分布状况是否符合洛特卡定律从而得出最终结论。通过近年来国内学者在不同学科领域对洛特卡定律的验证中,我们发现学科特征、统计条件、研究方法不同得出的结论不同。有的学科文献著者完全符合洛特卡定律,而有的不支持,甚至与洛特卡定律相悖^[5]。本文将数据结果与洛特卡定律比较分析时,首先明确洛特卡定律的基本内容,并选择符合知识管理学科领域的统计方法对洛特卡定律进行验证,得出知识管理论文著者的洛特卡分布,分析学科特征。

4.1 洛特卡定律的基本内容

(1)洛特卡定律是描述作者与论文数量之间关系的经验规律,这一定律是由美国情报学家洛特卡于1926年提出。洛特卡通过选择美国《化学文摘》和

德国奥尔巴赫《物理学史一览表》为数据源研究科技工作的论著分布^[6],从而得出著名的论文著者百分比分布应符合如下公式:即著者分布的平方反比公式^[2]: $f(n)=c/n^2$,c约等于0.6079。

(2)通过国内外学者对洛特卡定律的适用性研究得出其一般表达式为 $f(x)=C/x^n$ (1) 其中,x设为论文数量,f(x)为写x篇论文的作者占作者总数的比例^[7];C,n为参数。由于学科领域、学科水平的不同,参数值C与n值也有所不同。因此所有的学科领域并不一定符合平方反比公式。根据实际论文统计数可以计算出其参数值。从而验证是否符合洛特卡定律。

洛特卡定律实用性强,通过洛特卡定律可以较为全面清晰的描绘出专业领域的科学生产率状况。20世纪80年代开始,国内学者开始对洛特卡定律进行验证,对于洛特卡定律方面的研究就一直连续不断。本文主要是通过知识管理领域论文著者科学生产率结果与洛特卡定律相比较分析,验证其是否符合洛特卡定律,根据验证结果得出知识管理领域的科学发展情况,为知识管理领域学者的研究提供了指导。但是洛特卡定律分布是一种静态的表述,而文献发展是动态的,随着科学的发展,洛特卡定律逐渐的失去了其全面性,不能够对动态知识进行描述,因此在国外提供的某些数据上有时会显示一篇论文作者分布比率偏高^[8]。但是这并不会影响分析得出论文著者科学生产的一般性规律。

4.2 洛特卡定律验证

将统计分析的结果与洛特卡定律进行对比分析,验证知识管理论文著者的科学生产率与洛特卡定律的关系。

首先,估算非线性方程 $f(x)=C/x^n$ (1)中参数C与n的值。主要运用统计学里的最小二乘法。

步骤:第一,估计非线性方程 $f(x)=C/x^n$ (1)中参数n的数值,对(1)式方程的两端取对数,得出 $\lg y_x = -n \lg x + C$ 。最小二乘估计值:

$$\bar{N} = \frac{-N[\sum(\lg x)(\lg y_x)] + [\sum(\lg x)][\sum(\lg y_x)]}{N[\sum(\lg x)^2] - [\sum(\lg x)]^2} \quad (2)$$

根据数据统计的结果,利用(2)式得出公式中的相关数值(如表2)。

其中N=10,根据公式(2)以及表2统计计算的数据得出n的最小二乘估计值 $n=3.209$ 。

由 $n=3.209$ 以及Pao.M.L给出的下列公式^[7]:

$$c = \left(\sum_{x=1}^{\infty} \frac{1}{x^n} \right)^{-1} \sum_{x=1}^{\infty} \frac{1}{x^n}$$

$$= \sum_{x=1}^{20} \left[\frac{1}{x^n} + [(n-1)(20^{n-1})]^{-1} + \frac{1}{2 \times 20^n} + \frac{n}{24 \times 19^{n+1}} \right]$$

得出 $C=0.8471$ 从而得出知识管理论文的作者分布规律为 $f(x)=0.8471/x^{3.209}$, 其中 $f(x)$ 为写 x 篇论文的作者占作者总数的比例。

其次, 得出知识管理论文作者分布规律后, 利用 K-S 检验所统计的知识管理专题论文在理论上是否符合洛特卡定律。为了验证我国知识管理专题的作者科学生产率实际分布与洛特卡定律理论分布的一致性, 对得出的式子进行 K-S 检验。

表2 论文量与作者量相关数据统计分析

x	1	2	3	4	5	6	7	8	9	10	合计
lgx	0	0.3010	0.4771	0.6021	0.6990	0.7782	0.8451	0.9031	0.9542	1	6.5598
lgy	3.2472	2.3263	1.7782	1.4150	0.9542	0.6990	0.4771	0	0	0.6021	11.4991
(lgx) ²	0	0.0906	0.2276	0.3625	0.4888	0.6055	0.7142	0.8156	0.9105	1	5.2151
lgx ² lg y	0	0.7002	0.8484	0.8520	0.6670	0.5440	0.4032	0	0	0.6021	4.6169

4.3 K-S 检验

在对统计数据进行 K-S 检验前, 首先应该明确 K-S 检验的基本原理, 其次计算 K-S 检验数据, 在数据分析基础上进行比较研究。

4.3.1 K-S 检验的基本原理

K-S 检验的基本原理计算出统计结果中值的累积频率分布与理论累积频率分布之间的差异, 选取最大差值记为 D_{max} , 若最大差值 D_{max} 小于临界值, 则不能认为有显著差异, 即所统计的科学生产率分布状况理论上符合洛特卡分布, 反之实际统计结果不符合洛特卡分布规律^[7]。

4.3.2 K-S 检验数据

计算 K-S 检验数据时, 主要在知识管理论文的作者分布规律 $f(x)=0.8471/x^{3.209}$ 基础上进行进一步的运算, 计算出论文的实际累计频率、理论累计频率和频率差(如表3)。

表3 K-S 检验数据

x	1	2	3	4	5
实际累计频率	0.8463	0.9478	0.9765	0.9890	0.9933
理论累计频率	0.8471	0.9387	0.9636	0.9735	0.9783
频率差D	0.0008	0.0091	0.0129	0.0155	0.0150
x	6	7	8	9	10
实际累计频率	0.9957	0.9971	0.9976	0.9981	1.0000
理论累计频率	0.9809	0.9825	0.9837	0.9845	1.0000
频率差D	0.0148	0.0146	0.0139	0.0136	0.0000

由表3可直观得出 $D_{max}=0.0155$, 当显著性水平 $\alpha=0.01$ 时, 临界值 K-S 值为 $1.63/\sqrt{\sum y_x} = 1.63/\sqrt{\sum 2088} = 0.0357$ 。

根据 K-S 检验思想 $0.0357 > D_{max}=0.0155$, 因此

在显著性水平 $\alpha=0.01$ 时, 认为知识管理领域论文作者的科学生产率理论上符合洛特卡定律。知识管理论文著者的洛特卡分布为 $f(x)=0.8471/x^{3.209}$

5 结果分析与结论

通过分析知识管理专题论文著者的科学生产率得出其符合洛特卡定律, 由于在统计中我们对于合著作品只取其第一作者, 所以会有一定的误差^[9], 但是忽略误差不计, 论文著者的洛特卡分布创造性地揭示了科学论文作者与论文数量之间的数量关系。普赖斯在洛特卡定律的基础上对科学著述中的高产作者及其论文贡献率得出过如下推论: “大约 10% 的杰出科学工作者发表的论文量占全部论文量的一半^[10]”。通过对洛特卡分布的 C 与 n 值得出知识管理专题领域的学科难度较低, 较多的人对这一专题进行研究。同时 n 值较大, 则认为作者较为分散, 许多著者发表一篇文献, 高产作者较少。从而得出知识管理专题领域的学科渗透水平较高。研究出了这一学科专题的学术水平及学科层次同时也清晰的展现了这一专题的论文著者科学生产分布状况, 为以后的学科研究奠定了基础。

参考文献

- 尹方屏, 张大为, 刘静伟. 中国服装学科论文著者的洛特卡分布研究[J]. 农业图书情报学刊, 2007, 19(2): 141-144.
- Lotka A. J. The Frequency Distribution of Scientific Productivity[C]. Washington: Journal of the Washington Academy of Sciences, 1926.
- 林秀梅, 刘玉玲, 周松青. 《当代经济研究》论文作者科学生产率分布的统计分析[J]. 当代经济研究, 2000, (10): 70-71.
- 谢钰荣, 罗 芳. 从科技论文产出看高校的科研产出能力—以南昌大学 2005-2007 年论文为例[J]. 科技广场, 2009, (6): 218-219.
- 李小霞. 近年来国内洛特卡定律研究综述[J]. 科技情报开发与经济, 2005, 15(13): 27-28.
- 靖继鹏, 马费成, 张向先. 情报科学理论[M]. 北京: 科学出版社, 2009: 102-133.
- 夏 鸣. 图书情报领域作者生产率研究[J]. 河南图书馆学刊, 2008, 28(6): 28-31.
- 王 静, 王宏鑫. 关于 Lotka 定律的研究—纪念洛特卡定律创立 80 周年[J]. 情报杂志, 2007, (4): 94-96.
- 邱均平, 王明芝. 1999~2008 年国内数字图书馆研究论文的计量分析[J]. 情报杂志, 2010, 29(2): 1-5.
- 邱均平. 文献计量学[M]. 北京: 科学技术文献出版社, 1987: 6.

(责任编辑: 赵立军)