

基于数据正态化处理的期刊学术影响力研究*

刘贵伟 彭彩红 赵福荣

(《大连工业大学学报》编辑部,116034,辽宁大连)

摘要 科技期刊学术影响力评价的重要指标——总被引频次和影响因子都是以某一时间段期刊发表的单篇论文被引次数为基础数据进行测算的,由于基础数据的非正态分布性,必然造成测算结果失真。该文提出一种基于基础数据正态化处理的科技期刊学术影响力构想,以期能够全面、客观、动态地反映科技期刊的学术影响力。

关键词 科技期刊;学术影响力;正态分布

Study of sci-tech journals academic influence power based on normal distribution//LIU Guiwei,PENG Caihong,ZHAO Furong

Abstract Cited index and impact factor are the vital indication evaluating the impact power of scientific journal, both of which are calculated based on the cited number of the single paper published in the journal in a period of time. However, the non-normal distribution of the basic data can causes inevitably the distortion in calculated results. Therefore, an impact evaluation of scientific journal based on the data modification via normal distribution was discussed in this paper, which can respond the impact power of the journal more overall, objectively and dynamically.

Key words scientific journal; impact power; normal distribution

Author's address Editorial Office of Journal of Dalian Polytechnic University,116034,Dalian,Liaoning,China

现行的国家科技评价体系和科研量化考核体系中,把研究论文的产出量和刊载论文期刊的影响因子作为重要指标,加之国家对硕士研究生、博士研究生培养质量要求的提高,作为刊载研究成果平台的科技期刊发展也很迅速,然而,科技期刊学术质量与学术影响力的大小却有很大差别。如何能够客观评估科技期刊的学术影响力,同人做了大量的研究工作,取得了许多值得借鉴和参考的成果^[1-4]。笔者也对此进行了研究,提出

了对指标基础数据进行科学处理,形成基于数据正态化处理后的期刊学术影响力测算构想。

1 问题的提出

在现行的科技期刊评价体系和科研绩效考核体系中,存在着这样2种现象:一是科研绩效评价中关于论文的考核往往以科技期刊的影响因子作为论文水平的高低来衡量作者的学术水平^[5]。极端情况下,可能该作者的论文根本无人关注,因没有被引用的价值,根本没有被引用。二是科技期刊评价中更多地依赖少数高被引论文的引文频次,来提升整个科技期刊的影响因子,或不恰当引用来人为提高影响因子。极端情况下,可能是极少数论文引用频次高,而绝大多数论文根本没有被引用。正像学术界普遍关注的不能以科技期刊的影响因子来考查论文、评价作者学术水平一样,科技期刊学术影响力的评价也要弱化极少数论文的高引文频次对整个期刊的影响。这样,对引文频次数据进行科学处理就显得十分必要。

2 数据处理的理论依据

从数理统计学角度讲,只有当被研究对象是正态分布时,其平均数才有充分的意义。科技期刊发表论文的引用频次随机性很大,根本不能呈现正态分布^[6],因此,对这样一组不能呈正态分布的数据进行绝对的平均值计算和绝对的累加都会失去其真正的现实意义。如何能让一组非正态分布的数据变成正态分布,数学上有一种科学的处理方法就是对其取对数。这样处理后,再进行计算就能弱化个别极端离散的数据对结果造成的影响,而使结果更具现实意义。

[2] Garfield E. Citation indexing: its theory and application in science, technology, and humanities [M]. New York: John Wiley and Sons, 1979

[3] Buela-Casal G. Assessing the quality of articles and scientific journals: proposal for weighted impact factor [J]. Psychology in Spain, 2004, 8(1) : 60-76

[4] Amin M, Mabe M. Impact factor: use and abuse [J]. Perspectives in Publishing, 2000, (1) : 1-6

[5] Bollen J, Marko A. Rodriguez, Herbert V. de Sompel. Journal Status [EB/OL]. [2008-10-18]. <http://arxiv.org/abs/cs.DL/0601030>

[6] Kleinberg JM. Authoritative sources in a hyperlinked environment [C]//Proceedings of the 9th ACM SIAM Symposium on Discrete Algorithms, Baltimore, 1998: 668-677

[7] 中国科学技术信息研究所. 2007年版中国科技期刊引证报告:核心版[M]. 北京:科学技术文献出版社, 2007

[8] Langville A, Meyer C. Google's pagerank and beyond: the science of search engine rankings[M]. Princeton: Princeton University Press, 2006

(2009-04-16 收稿;2009-06-08 修回)

* 辽宁省教育厅科技研究资助项目(2008076)

3 期刊影响因子的处理

为了既方便研究又能直观地说明问题,我们选取2种样刊作为研究对象。样刊A、样刊B在统计年前2年发表的10篇论文和统计当年被引用情况如表1所示。按照影响因子(F_i)的定义,从表1可以直接得出2种样刊的影响因子 $F_{iA} = F_{iB} = 33.6$ 。

表1 样刊引用频次和经正态化处理后的数据表

论文编号	引用频次		正态化处理数据	
	n_A	n_B	$\lg n_A$	$\lg n_B$
1	8	40	0.903	1.602
2	1	60	0.000	1.778
3	25	25	1.398	1.398
4	1	50	0	1.699
5	40	40	1.602	1.602
6	2	9	0.301	0.954
7	1	35	0	1.544
8	246	32	2.391	1.505
9	3	7	0.477	0.845
10	9	38	0.954	1.580

对单篇论文引用频次进行正态化处理后,2种样刊的数据见表1。这里不妨根据影响因子的概念,将经处理后测算出的影响因子定义为“正态影响因子”,记为 F_{ii} 。从表1可以得出, $F_{iIA} = 0.803$, $F_{iIB} = 1.451$ 。

从以上分析可以看出,2种样刊尽管影响因子相同,但学术影响力的差别是显而易见的;然而,经过正态化处理后的正态影响因子却不相同,两者的正态影响因子值相差较大,这与2种样刊的实际情况相吻合,这也验证了数据处理的科学性和方法的可靠性。

这一问题处理的结果,与美国科学家 J. E. Hirsch 于2005年8月提出的一项新的文献计量指标即 h 指数指标的思想有相近之处。某一期刊的 h 指数是指该刊在一定的时间窗口内发表的论文至少有 h 篇的被引频次不低于 h 次,其余论文的被引频次不大于 h 次^[7]。一般而言, h 指数越高,其学术影响力就越大。

文献[8]显示,一般被引半衰期平均在5年以上,因此,建议在应用正态影响因子时可以直接采用5年的数据,测算5年正态影响因子 F_{ii} 值。这样,既可以满足统计学上的大样本的可靠性要求,同时一篇论文的被引用情况随着时间的推移,被引的可能性会下降,其对数值也趋于稳定,从而使科技期刊学术影响力的

评价结果能够更客观、更真实。

4 总被引频次的处理

目前,总被引频次是指某期刊自创刊以来所登载的全部论文在统计当年的统计刊源中被引用的总频次。这是文献计量中一个基础性的指标,能够在一定程度上反映该刊在学术交流中总体被使用和受重视的程度。实际上该指标是某期刊自创刊年开始,到统计当年各年度论文被引用频次的累加。为了尽可能地消除或弱化个别年份或非正常情况引用对期刊学术影响力评价的干扰,也可将各年总被引频次数据分别取对数,然后再进行累加,得出正态总被引频次。

5 结语

科技期刊学术影响力研究,对客观评价期刊、正确引导期刊发展会起到十分重要的作用;数据选取和处理的科学性、方法的可靠性是整个研究工作的基础,也是能够得出全面、客观评价结果的重要前提。笔者的工作正是基于这种想法对这一问题进行了初步探讨,更深入、系统的工作还有待进一步研究。

6 参考文献

- [1] 何学锋,彭超群. 科技期刊学术影响力的动态评估模型[J]. 编辑学报,2002,14(4):238-240
- [2] 何荣利. 期刊被引频次和影响因子与载文量的关系趋势分析[J]. 中国科技期刊研究,2005,16(4):500-503
- [3] 刘明寿. 采用德尔菲法评价高校学报学术影响力[J]. 贵州大学学报:自然科学版,2004,21(4):437-440
- [4] 董秀玥. 进一步完善期刊影响因子值计算的方法探讨[J]. 中国科技期刊研究,2006,17(2):290-292
- [5] 金碧辉,汪寿阳,任胜利,等. 论期刊影响因子与论文学术质量的关系[J]. 中国科技期刊研究,2000,11(4):202-205
- [6] 任霄鹏. 科学家提出期刊评价新标准[EB/OL]. [2009-01-26]. <http://www.sciencenet.cn/htmlpaper/200837940-475951359.html>
- [7] 万锦堃,薛芳渝. 中国学术期刊综合引证报告[M]. 2008版. 北京:科学出版社,2008
- [8] 中国科学技术信息研究所. 2006年版中国科技期刊引证报告[M]. 北京:科学技术文献出版社,2006
(2009-02-06 收稿;2009-06-08 修回)