

基于引用期刊质量的影响因子修正法

郑美莺¹⁾ 梁飞豹²⁾

1)《福州大学学报(自然科学版)》编辑部; 2)福州大学数学与计算机科学学院:350116,福州

摘要 长期以来学术期刊影响因子的计算仅仅统计被引用文章的篇数而没有考虑引用期刊的学术质量,这种计算方法有不合理之处。认为在计算影响因子时应充分考虑引用期刊的质量,建议以引用文献的期刊质量为权重对原影响因子公式进行修正。经统计学模拟计算表明,修正后的公式能更客观地对期刊进行评价,这将有利于建立起更公正、合理的期刊评价体系。

关键词 期刊质量;影响因子;评价;修正

Impact factor modification based on the journal quality of citations//ZHENG Meiyang,LIANG Feibao

Abstract The impact factor of an academic journal has long been based on the arithmetic mean of citation counts, regardless of the quality of citations. However, the validity of the current algorithm is still open to question. We propose an alternative way of calculating the impact factor by the weighted mean, and the weights are based on the journal quality evaluation of citations. The simulation study indicates that the modified formula evaluates the academic journals more appropriately, and it can be used to build up a more justified system for scientific quality evaluation.

Keywords scientific journal quality; impact factor; evaluation; modification

First-author's address Editorial Department of Journal of Fuzhou University(Natural Science) ,350116, Fuzhou, China

长期以来,期刊界许多人士不断探讨科学评价学术期刊的理论和方法^[1-4]。现有的评价方法多以总被引频次、影响因子、即年指标、论文地区分布数、平均引用率、基金论文比等数据作为评价指标,其中,影响因子是最受认可和重视的指标。

影响因子于1972年由E·加菲尔德提出,它是一个相对统计量,现已成为国际上通用的期刊评价指标。由于影响因子的高低决定着期刊的品质和论文作者的评级,近年来甚至出现影响因子崇拜现象^[5]。影响因子在20世纪80年代引入中国,应用于核心期刊的遴选^[6]。由中国科学文献计量评价研究中心和清华大学图书馆共同研制,由《中国学术期刊(光盘版)》电子杂志社出版的2011年版《中国学术期刊影响因子年报》首次提出一套创新的期刊影响因子指标体系,从不同计量评价角度提出复合影响因子、综合影响因子、学科复合影响因子、基础研究类影响因子和技术研究类影响因子等一系列指标^[7],从而构建了多角度计量评价期刊的影响因子指标体系。2013年版《中国学术

期刊影响因子年报》还增加了5年影响因子指标^[8],这将更全面地反映期刊的影响力;然而,这一年报并没有考虑引用者产出论文的质量,即未考虑到引用者的学术能力,这将影响到期刊评价的客观性。

为此,本文提出基于引用期刊学术质量的影响因子修正法,即追踪引用者产出论文所在刊物的学术质量(影响因子),设计以权重修正的影响因子计算方法,以期更科学、客观地评价期刊的影响力。

1 原计算公式的缺陷

根据加菲尔德的定义,影响因子是指某种期刊在一定的统计源期刊范围内,该期刊前2年论文在这一范围内被引用的数量与该期刊前2年刊载论文数量之比。为研究第 t 年期刊的影响因子,需要如下数据:1)总的期刊(共 n 种): X_1, X_2, \dots, X_n ;2)每种期刊 X_i 在第 $t-1, t-2$ 年的文章总数: $N_{i,t-1}, N_{i,t-2}$;3)每种期刊 X_i 在第 t 年引用第 $t-1, t-2$ 年的期刊 X_j 的文章数: $a_{i,j,t-1}, a_{i,j,t-2} (i, j = 1, 2, \dots, n)$ 。

按照原影响因子计算方法,第 t 年某期刊 X_k 的影响因子为

$$I_{Fk,t} = \frac{\sum_{j=1}^n (a_{j,k,t-1} + a_{j,k,t-2})}{N_{k,t-1} + N_{k,t-2}} \quad (1)$$

式中: $\sum_{j=1}^n (a_{j,k,t-1} + a_{j,k,t-2})$ 是该刊 $t-1, t-2$ 年中所有文章在 t 年被引用的总次数; $N_{k,t-1} + N_{k,t-2}$ 是该刊在 $t-1, t-2$ 年的所有文章数。

这个计算公式已经沿用了几十年,在一定程度上确实为期刊的评价起到了重要的作用;然而,这种计算方法只考虑引用文章的总数,而没有考虑到引用期刊的质量,这对评价期刊的社会影响力不够全面。仅仅以某期刊被引用论文的次数和载文量来计算分析期刊的影响力及其在科学交流中的地位和作用显然不够科学严谨^[9]。

现在的期刊群类似于橄榄的形状,高端的期刊毕竟是少数,而处于橄榄中间部位的是为数众多的中等水平的期刊,正是许多年轻学者或初涉研究领域的学生所经常浏览阅读的对象。这部分期刊往往有很高的引用和下载量,引用者研究层次的差别势必带来引用质量上的区别。一些顶级的研究成果属于曲高和寡,

对于初级研究者可能暂时看不懂一时也用不上,所以未必会受到金字塔底部普通研究人员的追捧,这些高质量的文章就有可能得不到高的引用次数,如爱因斯坦的“相对论”^[10];相反,一些大众级别的期刊可能更适合初级研究者尤其是为数众多的在校硕士生的青睐;所以,仅仅凭借被引用文章数来衡量期刊的学术影响力显然不够全面和客观。应该考虑到引用者产出的论文质量即追踪引用者产出论文所在刊物的学术质量,设计一个权重因子来修正原有的计算公式,从而能更科学、客观地评价期刊的影响力。

2 对影响因子公式的修正

针对原公式存在的缺陷,笔者提出以引用文献的期刊质量作为权重进行修正。所谓权重,是指在某种求和公式中,为了体现各求和项所具有的不同重要性,分别给予不同的比例系数,称为加权,加权所赋予的比例系数,称为权重。为此,以期刊质量为权重把式(1)修正为

$$I_{Fk,t}^* = \frac{\sum_{j=1}^n w_j (a_{j,k,t-1} + a_{j,k,t-2})}{N_{k,t-1} + N_{k,t-2}} \quad (2)$$

其中 w_j 为权重。

权重有多种确定方法,本文采用一种直观的确定方法。衡量期刊的质量指标有很多,而最重要指标是影响因子,所以我们直接利用期刊影响因子占有所有期刊影响因子的比重,作为该刊质量的权重,即

$$w_j = \frac{n(I_{Fj,t-1} + I_{Fj,t-2})}{\sum_{j=1}^n (I_{Fj,t-1} + I_{Fj,t-2})} \quad (j=1,2,\dots,n) \quad (3)$$

其中: $I_{Fj,t-1}$ 、 $I_{Fj,t-2}$ ($j=1,2,\dots,n$) 分别表示 $t-1$ 年和 $t-2$ 年期刊的影响因子。

为了直观地说明修正后公式的合理性,以统计学模拟方法进行计算说明。

例 假设目前共有 10 种期刊(X_1, X_2, \dots, X_{10}),其 $t-1$ 、 $t-2$ 年的影响因子及文章总数以及由此得出的权重见表 1; 10 种期刊在 $t-1$ 年和 $t-2$ 年的相互引用情况见表 2 和表 3。由表 1 知, X_1 、 X_2 、 X_3 期刊的影响因子依次序递减,即 X_1 期刊的质量最高, X_2 期刊次之,而 X_{10} 期刊的质量最差。

表 1 10 种期刊在某 2 年的影响因子及文章总数

项目	期刊										
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	
影响因子	$t-1$	3.52	3.43	2.81	2.54	2.12	2.04	1.22	0.83	0.63	0.45
	$t-2$	3.56	3.41	2.98	2.76	2.07	2.12	1.02	0.91	0.55	0.42
文章总数	$t-1$	186	192	178	186	198	195	189	164	176	165
	$t-2$	188	190	165	184	200	199	184	166	172	162
权重 w_j		1.79	1.74	1.47	1.35	1.06	1.06	0.57	0.44	0.30	0.22

表 2 10 种期刊在 $t-1$ 年的相互引用 ($a_{j,k,t-1}$) 情况

引用期刊	被引用期刊									
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
X_1	80	70	50	50	55	25	15	13	5	3
X_2	65	65	55	45	58	23	25	17	10	7
X_3	62	64	70	48	50	35	18	15	8	5
X_4	70	72	55	65	47	35	25	20	10	4
X_5	54	65	34	34	40	42	29	11	16	3
X_6	63	74	63	53	53	56	38	21	17	7
X_7	85	65	53	61	33	45	51	26	6	6
X_8	50	54	43	43	33	48	20	23	10	5
X_9	65	70	55	48	25	55	30	10	25	10
X_{10}	42	45	30	31	22	57	21	23	21	13
引用总数	636	644	508	478	416	421	272	179	128	63

表 3 10 种期刊在 $t-2$ 年的相互引用 ($a_{j,k,t-2}$) 情况

引用期刊	被引用期刊									
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
X_1	85	75	45	45	57	22	18	8	8	2
X_2	55	63	50	45	56	25	25	21	6	1
X_3	68	64	65	50	52	38	22	17	7	7
X_4	65	72	65	62	46	37	25	20	6	5
X_5	54	60	34	34	48	46	26	6	11	8
X_6	69	61	63	53	53	46	38	18	15	8
X_7	89	75	61	65	30	48	45	26	11	5
X_8	50	60	33	43	20	43	20	23	13	8
X_9	75	75	55	38	25	54	24	10	22	10
X_{10}	45	42	40	51	22	57	21	14	18	10
引用总数	655	647	511	486	409	416	264	163	117	64

从表 2 和表 3 中选取 X_5 和 X_6 这 2 种期刊数据来分析(以黑体字表示),由表中可以直观地看出: X_6 期刊在 $t-1$ 年和 $t-2$ 年的被引用总数均高于 X_5 期刊,然而,影响因子排名靠前的期刊引用 X_6 期刊较少,而引用 X_5 期刊较多;相反,影响因子排名靠后的期刊引用 X_6 期刊的比较多。由此可直观看出 X_5 期刊的学术质量比 X_6 期刊高。

然而,按照原来的计算方法, X_6 期刊的影响因子高于 X_5 期刊(见表 4),这样显然不合理。

表 4 10 种期刊修正前后的影响因子比较

影响因子	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
修正前	3.45	3.38	2.97	2.61	2.07	2.12	1.44	1.04	0.70	0.39
修正后	3.57	3.49	3.11	2.65	2.42	1.84	1.36	0.99	0.57	0.32

采用修正后的计算公式对表 2 和表 3 的引用情况进行计算,结果见表 4。可以看出,经过修正后 X_6 期刊的影响因子低于 X_5 期刊。这是因为 X_5 期刊更多的是被影响因子较高的期刊所引用,而 X_6 期刊却更多的是被影响因子较低的期刊所引用,虽然被引文章总数 X_6 期刊高于 X_5 期刊,但引入权重修正后,却出现了相反的结果。这进一步说明对原计算公式修正的必要性。当然,修正公式中所用到的期刊质量权重还是原