

非正常引用期刊引文结构、特征及其成因

——基于 CSSCI 的比较分析*

鞠秀芳¹ 吴金红² 田 紫³

(1. 南京大学中国社会科学评价中心 南京 210093;

2. 武汉纺织大学管理学院 武汉 430200 ;3. 昆明物理研究所红外技术编辑部 昆明 650223)

摘 要 利用 CSSCI 引文数据对非正常引用期刊的引文结构和分布特征进行了比较分析。分析认为,非正常引用期刊在统计年前 2、3 年发表的文章被引显著;非正常引用的期刊被引极其不均衡,显示出高度集中的特性,且自引率与本学科期刊相比偏高。非正常引用期刊从引文时间分布、被引均衡程度、引用主体几个方面与正常引用的期刊相比具有显著差异。

关键词 非正常引用 引文分析 引文时间分布 引用均衡度 自引率

中图分类号 G353

文献标识码 A

文章编号 1002-1965(2014)02-0095-04

DOI 10.3969/j.issn.1002-1965.2014.02.018

Citation Structure, Features and Reasons of Non-normal Citing Journals: an Analysis Based on CSSCI

Ju Xiufang¹ Wu Jinhong² Tian Ying³

(1. Chinese Social Sciences Research Evaluation Center, Nanjing University, Nanjing 210093;

2. School of Management, Wuhan Textile University, Wuhan 430200;

3. Kunming Institute of Physics, Kunming 650223)

Abstract In this paper, we analyzed the citation structure and distribution characteristics of non-normal citation journals using the CSSCI citation data. The result showed that papers published two or three years ago by non-normal cited journals in the statistic were cited more often; non-normal cited journals were cited highly imbalanced and showed a concentrated feature, and they had a higher self-citation rate compared with other journals in the same discipline. There were significant differences between non-normal citation journals and other journals on citing years distribution, citing equilibrium degree and self-citation rate.

Key words non-normal citing behavior citation analysis citing years distribution citing equilibrium degree self-citation rate

0 引 言

期刊评价及机构、个人成果评价是引文分析法的一个重要应用领域。随着广大科研机构、大学管理部门把《SCI》《CSSCI》的相关数据用于个人成果评价,并作为学者个人职称晋升、学术奖励的相关依据,《SCI》《CSSCI》等引文数据库的来源刊成为了广大学者关注的重中之重,能否成为某知名引文库的来源刊同时对期刊的稿源、经济效益、甚至生存状况都产生了

深刻的影响。为了能使自身的引文指标在引文数据库遴选来源刊时脱颖而出,部分期刊编辑有意识地运用各种方式人为改变期刊的引文,从而提高自身的引文指标数值。这种行为一方面干扰了文献之间的正常学术交流,另一方面污染了引文分析研究的基础数据源,从而损伤了利用引文数据进行评价的公平、公正性。

国外从 20 世纪 60 年代开始开展对引用行为的理论和实证研究,伴随着引文分析法在评价领域的应用推广,从 20 世纪 90 年代开始形成了一个小的热点。

收稿日期:2013-10-22

修回日期:2013-11-27

基金项目:国家社会科学基金项目“基于引文分析的期刊非正常引用行为研究”(编号:13CTQ034)研究成果之一;南京大学中央高校基本科研业务费专项资金资助。

作者简介:鞠秀芳(1976-),女,博士,副研究员,研究方向:期刊评价、引文分析、文献计量研究;吴金红(1974-),男,博士,副教授,研究方向:竞争情报、商业智能与智能信息系统;田 紫(1976-),女,本科,工程师,研究方向:期刊评价。

总的来说,对引用行为的研究经历了从理论到应用的循序渐进的研究过程,但是主要集中于对引用主体、引用动机的研究和在追逐引文指标的过程中引发的对影响因子、自引进行操控的研究。对于期刊的操控引用行为的特征以及成因分析没有全面展开。我国也有部分学者关注了期刊非正常引用现象,但主要还集中在定性的探讨上,缺乏定量的分析^[1]。本文试图从期刊引文时间分布、被引均衡程度、引用主体几个方面对期刊在正常引用和非正常引用时的引文数据进行比较分析,从而进一步分析、揭示期刊非正常引用行为下的数据表现特征。

1 期刊非正常引用行为及其引文分布特征

1.1 期刊非正常引用行为 科学文献的引证与被引证,是科学发展规律的表现,体现了科学知识和情报内容的积累性、连续性和继承性。然而,并不是所有引用都是合适的。既有由于作者不太重视参考文献的作用,经常出现参考文献的错引、漏引等而产生的不良引用现象^[2];也有在利益的驱动下,期刊部编辑采取的提高自引、联合引用、强制引用等限制作者引用的非正常行为^[3]。这种由期刊主导的非正常引用的行为方式由于具有一定的规模性,一方面引导了学术界的引文不正之风,另一方面对引文数据源造成了不可逆转的污染,改变了引文结构特征,影响了使用引文开展评价的公平公正,是本文研究的对象。

Matthew E. Falagas 等人在 2008 年对期刊非正常引文的方式进行了总结,提出了排名前十的期刊非正常引用的方式^[4],分别为:

a. 编辑要求文章修改后的稿件中参考文献要包括编辑所在期刊或相关期刊的文献;b. 发表与文章相关的引文的摘要;c. 通过编辑和作者对文章评论,扩增自引量;d. 发表一些题名但不被列为可引的文章来增加引文量;e. 发表较多综述性文章,包括原始研究,特别是案例报告;f. 拒绝消极研究,而不管他们的质量如何;g. 拒绝隐藏摘要的研究;h. 喜欢合著者多的文章,易接受起源于比较大的团队和科学性积极研究团队的文章;i. 喜欢著名科学家和研究带头人的成果,而不考察其真正质量;j. 主要发表有关热门话题的科学文章。

这 10 种方式仅是期刊非正常引用常用的方法,当然,现实操作中还存在其他的“技巧”,如同一集团、同一单位的两个或几个刊物结成“互引同盟”,授意该刊作者多引用“互引同盟”期刊的论文^[5]。更有甚者,有的编辑部人员越俎代庖,干脆给所投该刊的论文添加在该刊发表论文的文献;或者故意在论文中留下错误,制造悬念,引起争议;它们都是利用期刊引证指标设计的缺陷,从而达到人为提高自身引文指标数值的目的

的^[6-7]。

1.2 非正常引用的期刊引文分布特征及其成因分析 期刊引文由来自不同来源期刊或文献的引用组成,在很大程度上受到人为因素的影响,具有较大的随意性。但是,文献计量学的研究表明,科学引文具有一定的分布结构特征和规律。本文以 CSSCI 作为数据来源,在其中选取期刊界业内反映有非正常引用嫌疑的 10 本 CSSCI 人文社科综合类期刊的在 2006-2010 期间的引文数据做为研究对象(为了防止争执,我们以 m1 ~ m10 来表示这 10 本期刊),并以排除此 10 本期刊引文数据后的 CSSCI 引文数据作为标准数据源,从而对期刊引文在不同情况下的分布特征进行比较研究。

1.2.1 期刊引文时间分布特征比较分析 从时间的角度对引文分布规律进行统计分析是引文分析的主要内容之一。早在 1963 年,普赖斯就对西方的科技引文进行了大量统计工作,并指出:文章被引证的峰值是该文章发表以后的第 2 年,也就是说,当年发表的文章,引用的参考文献大量来自论文发表前 2 年^[8]。

为了分析期刊的引文年代分布规律,我们对 2006-2010 年的 CSSCI 期刊引文年代进行了统计,结果显示:CSSCI 引文年代分布同样表现出了明显的规律性特征,即被引论文年代的高峰为统计年前 2 年或前 3 年发表的论文,其后随发文年代向前推移,其被引次数占总被引的比率逐渐降低。从 2006-2010 年的引文年代分布来看,文章发表后的第 2、3 年被引用的次数之和占总被引次数的比率稳定在 26% ~ 34% 之间。

表 1 2006-2010 引文年代分布

发表年	引用年				
	2006(%)	2007(%)	2008(%)	2009(%)	2010(%)
2010	0	0	0	0.003	2.410
2009	0	0	0.004	2.470	13.422
2008	0	0.004	2.505	13.124	13.854
2007	0.001	2.577	15.232	15.044	12.563
2006	2.099	16.206	16.445	13.563	10.845

注:CSSCI 数据显示,时有作者引用尚未刊发的文章,由此造成部分论文被引提前,如 2008 年发表的论文会在 2007 年被引。

作为比较,我们对非正常引用期刊的引文年代分布特征进行分析,对这 10 种期刊在 2006-2010 年被引数据进行分析,统计结果如表 2 所示。

表 2 非正常引用期刊被引年代分布

发表年	引用年				
	2006(%)	2007(%)	2008(%)	2009(%)	2010(%)
2010	0	0	0	0	5.35
2009	0	0	0	3.731	26.51
2008	0	0	7.653	23.117	25.12
2007	0	7.84	34.589	19.805	14.84
2006	6.75	36.7	18.560	14.854	7.92

由表 2 可以看出,虽然非正常引用的期刊其引文

年代分布同样集中在论文发表后的第 2 年。但和表 1 中 CSSCI 来源刊总体相比,被引年代分布特征有着明显不同:非正常引用期刊在统计年前 2、3 年发表的文章被引总和占到了所有引文年代的 42% ~ 52% 之多,随后年份的引用比例急剧降落,和表 1 中 CSSCI 来源刊总体 26% ~ 34% 相比,具有显著的差异。分析其原因,是因为在影响因子数值的计算中,在分母(期刊载文量)一定的情况之下,分子(期刊近两年所发文章的被引次数)大则期刊影响因子的数值大,分子小则期刊影响因子的数值小。因此,期刊为了提高自身的影响因子,大幅引用统计年前 2 年所发文章,从而造成统计年前 2、3 年发表的文章被引显著增高。

1.2.2 被引集中度比较分析 引文年度分布从纵向上反应了期刊的引文分布特征,被引广度则从横向上反映了期刊之间学术交流程度。期刊被引广度是指按年度统计引用该期刊论文的期刊种数。一般说来,引用某一期刊的期刊数量越多,该期刊的影响度就越大,它反映了期刊对其他期刊的影响覆盖面。表 3 对 CSSCI 各学科的期刊在 2006-2010 的被引广度进行了统计。由表 3 可以看出 CSSCI 各学科的期刊被引广度差异显著,被引广度最低的体育学学科被引广度只有 20 多,而社会科学总论类的期刊由于是各地社会科学院,社科联办的综合性人文社会科学期刊,每本刊的研究范围均覆盖了 3 ~ 6 个学科,研究范围相比较一般的专业期刊相对广泛,因而其对期刊影响的覆盖面尤其大,位居各类期刊之首,每年的被引广度值均在 100 以上,且年增长率在 10% 左右,至 2010 年达到了 151.52 之多。各学科的被引广度差异极大,这一方面是由于小学科研究内容相对集中,关注面窄,另一方面是由于大学科的多学科融合趋势日趋明显,关注度相对广。与此同时,从表 3 来看虽然 CSSCI 各学科的被引广度各有不同,但在 2006-2010 年之间,CSSCI 各学科的被引广度呈现逐年上升的趋势,表明 CSSCI 期刊的影响覆盖面正在逐年显著增大。

做为对照,我们对非正常引用期刊的被引广度同样进行了统计。由表 4 可以看出,虽然非正常引用的期刊每年均有不少期刊被引广度 低于学科的被引广度值,其影响覆盖面相比较所在学科总体相对集中,亦有不少非正常引用期刊的被引广度高于学科平均值。从被引广度的发展趋势来看,非正常引用的期刊被引广度同样呈现出逐年上升的趋势,且上升的幅度和学科上升幅度持平。由此可见,在反映期刊非正常引用方面,被引广度无法对期刊的引用行为加以区分。分析期刊被引广度的定义,我们可以发现被引广度在统计数值时只考虑期刊引文覆盖的期刊种数,而不考虑期刊的引用次数,也就是说不管期刊施引多少次,在计

算被引广度时均被视为 1,因而无法有效甄别期刊被引是相对均衡还是高度集中在某几本期刊上。

表 3 CSSCI 学科引用广度统计

序号	学科	2006	2007	2008	2009	2010
1	管理学	92.15	102.23	116.70	125.50	132.35
2	马克思主义	50.25	62.17	74.27	80.18	88.17
3	哲学	64.08	71.50	78.83	82.50	86.62
4	宗教学	25.75	26.25	32.50	34.00	46.00
5	语言学	39.20	42.14	51.90	56.67	56.04
6	外国文学	31.33	38.50	36.17	41.17	47.67
7	中国文学	41.00	47.13	52.86	56.57	64.47
8	艺术学	23.15	29.69	28.75	31.29	29.60
9	历史学	53.64	53.92	60.27	63.12	66.46
10	考古学	67.00	58.86	62.57	63.43	69.57
11	经济学	90.23	98.86	117.61	124.48	126.29
12	政治学	49.31	53.56	62.26	66.82	72.45
13	法学	76.58	80.86	101.33	109.76	113.62
14	社会学	83.57	96.29	94.75	101.44	115.44
15	民族学	43.08	39.92	55.46	66.00	68.93
16	新闻学与传播学	27.57	31.47	38.21	48.71	47.33
17	图书馆、情报与文献学	37.72	41.06	44.67	48.10	46.60
18	教育学	57.19	55.61	66.32	70.24	68.16
19	体育学	20.44	24.11	32.80	27.50	26.80
20	统计学	76.50	93.00	118.75	136.50	139.00
21	心理学	68.00	68.71	76.00	83.71	93.43
22	社会科学总论	107.95	122.34	133.80	147.36	151.52
23	人文、经济地理	67.29	70.00	74.29	77.00	87.60
24	环境科学	52.40	57.20	65.40	75.60	89.86
25	高校人文社会科学报	90.31	96.85	95.75	106.60	105.72
26	年平均值	67.48	73.18	83.36	90.86	81.59

表 4 非正常引用期刊被引广度统计

序号	期刊	2006	2007	2008	2009	2010
1	m1	78	94	115	127	145
2	m2	95	99	85	75	145
3	m3	93	142	163	163	184
4	m4	52	60	89	100	116
5	m5	128	150	178	210	207
6	m6	129	145	185	197	214
7	m7	130	132	148	151	170
8	m8	87	89	95	99	103
9	m9	102	104	111	115	125
10	m10	96	99	107	115	135
11	社科总论	107.95	122.34	133.80	147.36	151.52

为了进一步挖掘期刊非正常引用时引用分布的特性,我们对施引期刊按引用次数进行排序,在期刊被引达 50% 时,对施引排名位于前列的期刊进行计数。由表 5 可以看出,非正常引用期刊被引的 50% 往往集中在少部分期刊之上,几本期刊的引用往往对某本期刊的被引做了相当于其他几十本期刊的贡献,2 ~ 3 个刊的引用有时可以达到该刊年被引用次数的 50% ,有时甚至可达 80% ,非正常引用的期刊被引极其不均衡,显示出高度集中的特性。分析其原因,非正常引用期刊为了追求高他引,往往与某几本期刊结成“互引同盟”,其大量的被引往往集中于某几本期刊。利用这个特征,我们不仅可以识别出非正常引用的期刊,还可

以顺藤摸瓜,寻找出与其交换引用的期刊。

表 5 非正常引用期刊被引广度(2006-2010)

序 号	期 刊	被引 50% 被引广度 (2006)	被引 50% 被引广度 (2007)	被引 50% 被引广度 (2008)	被引 50% 被引广度 (2009)	被引 50% 被引广度 (2010)
1	m1	12	8	5	3	3
2	m2	12	10	8	5	7
3	m3	15	11	9	5	5
4	m4	18	12	8	6	8
5	m5	12	15	13	6	1
6	m6	21	15	18	16	17
7	m7	29	22	18	17	17
8	m 8	10	8	5	4	5
9	m 9	10	8	12	8	6
10	m10	15	13	8	9	8

1.2.3 引用主体特征比较分析 根据引用主体的不同,引用可以分为自引和他引。期刊自引有合理的部分,但不少学者反映,一些期刊会诱导或硬性要求投稿者增加对投稿期刊的引用从而提高自身的总被引频次和影响因子,自引因此受到了国内外众多期刊评价机构的置疑。在推出的 JCR(2009)中,有 26 种期刊被“请出”^[9],这是由于这些期刊的影响因子值更多的是来自其对自身的引用,部分期刊的自引率甚至高达 90%,不能真实反映出期刊在某领域的实际知识传播、分享的影响与价值。我们对 CSSCI 2006-2010 的期刊自引率进行了统计(见表 6)。统计显示,CSSCI 中自引率低于 20% 的期刊占据主体,约占 75% 左右;自引

表 6 CSSCI 学科自引率统计

学科	2006 (%)	2007 (%)	2008 (%)	2009 (%)	2010 (%)
管理学	12.90	11.00	10.00	11.70	11.28
马克思主义	6.20	7.60	6.00	7.60	7.52
哲学	15.70	14.90	14.50	14.20	12.37
宗教学	15.60	9.80	12.50	13.70	9.90
语言学	23.20	22.30	23.30	20.00	20.89
外国文学	19.70	20.00	23.00	22.40	15.74
中国文学	24.80	22.60	24.10	23.90	16.93
艺术学	24.90	23.80	17.50	23.80	18.58
历史学	18.80	18.20	16.10	17.90	18.28
考古学	19.60	28.50	28.30	23.80	23.20
经济学	14.60	14.60	11.30	10.80	12.28
政治学	18.00	16.70	16.90	17.70	12.24
法学	9.10	9.00	7.30	8.10	8.44
社会学	17.70	15.90	26.20	21.90	19.21
民族学	21.60	21.90	17.10	19.20	16.96
新闻学与传播学	23.00	23.70	24.80	21.20	23.57
图书馆、情报与文献学	21.70	20.90	18.30	16.60	17.39
教育学	20.70	23.70	19.90	20.50	22.66
体育学	27.70	26.90	21.80	24.80	23.00
统计学	21.10	16.00	18.30	20.50	21.23
心理学	21.90	19.40	23.60	22.00	22.45
人文、经济地理	25.30	22.80	23.30	27.70	37.70
环境科学	41.00	38.60	47.40	42.30	24.57
高校学报	9.00	9.50	9.40	10.20	11.27
社科总论	7.30	7.90	7.70	6.70	7.12

率大于 50% 的期刊虽然数量不多,但每年都有出现。从学科分布来看,CSSCI 各学科的自引率差异较大,大学科由于其研究人员众多,因此其学科自引率相对较小,比如法学,学科自引率只有 10% 不到;而小学科由于其研究领域相对狭小,研究方向相对集中,因此其学科自引率相对较高,大多在 20% 以上,其中环境科学的学科自引率甚至达到了 40% 以上;综合性社科期刊的自引率最低,年平均自引率只有 7% 左右。

做为对照,我们同样对非正常引用的 10 种期刊自引率进行了统计(见表 7)。从表 7 的统计结果来看,除了 3 本期刊每年的自引率均低于综合性社会科学期刊的平均自引率,其余期刊的年自引率均高于学科平均值,其中三本期刊每年的自引率都高出学科平均值 50%。一般来说,综合性社会科学期刊的研究范围广泛,同样的研究方向不少专业期刊亦在研究、关注,不太容易出现高自引现象,这些期刊出现的长时间的、高强度的自引显然与期刊编辑的导向不无关系。

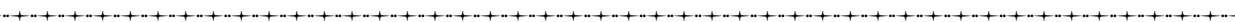
表 7 非正常引用期刊自引率统计

期刊	2006(%)	2007(%)	2008(%)	2009(%)	2010(%)
m1	8.80	6.90	16.40	11.40	19.29
m2	4.50	6.10	20.30	9.70	5.81
m3	3.00	4.80	4.00	5.30	7.49
m4	6.20	10.40	8.00	7.90	10.95
m5	35.30	30.70	39.30	20.20	35.21
m6	11.23	11.00	13.00	10.00	10.76
m7	6.56	6.00	5.00	7.00	6.93
m8	6.76	12.00	6.00	14.00	12.33
m9	5.37	4.00	6.00	6.00	5.95
m10	11.24	15.00	13.00	10.00	9.52

2 结 论

上述分析显示,非正常引用期刊从引文时间分布、被引均衡程度、引用主体几方面与正常引用的期刊引文相比表现出明显的特性:非正常引用期刊在统计年前二、三年发表的文章被引显著;非正常引用的期刊被引极其不均衡,显示出高度集中的特性,且自引率与本学科期刊相比偏高。

非正常引用期刊的引文特征可以帮助我们建立起一个识别期刊非正常引用的指标体系,从而为净化期刊的学术环境、保证选刊指标的公正性提供一些帮助。然而,期刊引文庞大复杂,从海量的引文数据中识别出期刊引用是否正常,还需要进一步建立结构清晰、算法准确的数据模型,才能确保对非正常引用期刊的识别更为准确、有效。再者,杜绝非正常引用行为,用识别指标加以监督是不得已而为之的消极办法,有效的杜绝办法依靠的是评价体制的有效改革和积极的职业伦理教育。



(上接第 98 页)

参 考 文 献

[1] 鞠秀芳,郑彦宁,潘云涛. 期刊引用操纵行为研究综述[J]. 西南民族大学学报:人文社会科学版,2013(4):224-228.

[2] 吕亚平. 学术论文参考文献不良引用行为分析及防范措施浅议[J]. 图书馆工作与研究,2012(11):53-56.

[3] 叶继元. 引文的本质及其学术评价功能辨析[J]. 中国图书馆学报,2010(1):64-68.

[4] Matthew E Falagas, Vangelis G Alexiou. The Top-ten in Journal Impact Factor Manipulation[J]. Ethics in Science, 2008(56): 223-226.

[5] 潘云涛,武夷山. 自引、他引:说不尽的故事[J]. 科技导报, 2007(25):35-40.

[6] Alper Sevinc. Manipulating Impact Factor: Anunethical Issue O-ran Editor's Choice? [J]. Swiss Medical Weekly, 2009(134): 410.

[7] 董建军. 人为因素对科技期刊影响因子评价指标的影响[J]. 编辑学报,2008(8):365-366.

[8] 邱均平. 信息计量学[M]. 武汉:武汉大学出版社,2007:89-92.

[9] Journal Citation Reports® Notices[EB/OL]. [2011-02-12]. http://admin-apps.isiknowledge.com/JCR/static_html/notices/notices.htm.

(责编:刘武英)