

科技期刊文献引用分布规律的探讨

董建军^{1,2)}

收稿日期:2012-04-11
修回日期:2012-09-05

1) 新乡医学院期刊社《眼科新进展》编辑部,453003 河南新乡,E-mail:djj1919@163.com
2) 河南省科技期刊研究中心,453003 河南省新乡市新乡医学院

摘要 通过随机选取学科、学科子类以及期刊来研究科技期刊文献引用分布规律,按照被引频次由高到低的顺序统计不同被引频次的文献量。研究发现无论是从学科层次来分析还是期刊层次来分析论文的引用在发表后的2~6年间均呈现偏态分布,随着被引频次的增高被引文献量逐渐降低,占期刊已被引文献数30%的论文提供50%以上的引用,随着论文发表时间的延长这种引用的集中趋势有所增强,但是变化幅度不大。

关键词 科技期刊 引用分布 引用贡献率

目前,评价期刊的方法很多,影响因子、总被引频次、h指数以及特征因子等等,都受到人们广泛地使用和关注^[1-3]。这些指标有的是基于期刊的引用绝对值,有的是基于期刊引用相对值,有的是基于期刊引用的权重来评价期刊的。大多数是脱离了期刊的引用分布来评价期刊的。我们前期研究发现我国科技期刊所发表的论文中零被引和低被引所占的比例较大,这种特殊的引用分布以及人为的操作引用行为使得我们很多的期刊评价指标偏离了原来的设定意义,如基金论文比等等。此外这些基于引文的评价指标受到很多因素的影响,比如影响因子就存在自引问题、时滞问题、不同学科的比较问题等等局限性^[4]。期刊论文发表后不可能达到均衡的被引,论文写作中的引用行为是一种半随机的行为,这种行为具有盲目性的一面也具有选择性的一面,因此要评价期刊就一定不能脱离期刊引用的分布规律来进行评价。那么,我国科技期刊引用的分布有何规律呢?这就是本次研究的核心问题。

1 数据来源和研究设计

1.1 数据来源与检索

本研究的研究数据来源于中国知网学术文献总库的中国学术期刊网络出版总库所收录的论文。通过专业检索入口,按照被引频次组配不同的期刊、专业、年份等来检索某学科的文章以及某期刊中的文章的被引用情况。

1.2 研究的设计

我们随机选择该数据库中的五个学科军事学、眼科学、天文学、生物学、气象学、数学进行研究,首先整理出这五个学科按照被引频次从高到低的引用分布情况,将各学科中已经被引用的论文分成两部分,前30%为高被引区,剩余的70%为低被引区(注意此种划分是在借鉴了帕累托定律以及在前期预分析的基础上进行的),计算出高被引区论文总的被引频次占该学科总被引频次的比例,也即该学科被引论文

中前30%论文的被引频次占总被引频次的比例(高被引区论文引用的贡献率)。分析的年份为2008年、2009年、2010年各学科所发表的论文,自发表后到2012年被引用次数分布情况。此外在生物学中随机选取6个子类,进行和以上相同的数据整理和分析,分析的年份为2010年各子类所发表的论文的引用分布情况。

分别从生物学、数学以及眼科学中各选取5种杂志,进行与以上各学科相同的数据整理和分析。

2 论文的引用分布规律

2.1 学科中论文的引用分布规律

五种学科不同年份的论文引用分布情况见表1~3。从历年来各学科的引用分布来看学科中论文的引用呈现一种偏态的分布,依据被引频次从高到低的顺序排列,论文篇数逐渐增多。我们将被引频次大于等于10的论文的篇数及其被引频次总和和计算在一起是因为这部分论文在各学科论文总数中所占比例极少,由于被引频次较高此部分论文对学科论文引用的贡献率却很高,不容忽视。高被引区的论文占已发生被引论文的比例为30%,然而此部分论文的引用贡献率却均在50%以上,且各个学科历年的变化有逐渐上升的趋势,但是这种上升趋势较为稳定,而且变化的幅度不大。可能是由于论文在发表初期的零被引率和低被引率较高,随着发表时间的推移有很多论文逐渐获得引用,导致这种占已发生引用的30%的论文贡献50%以上的引用规律随着时间不会发生太大改变。

此外,学科研究越热这种引用分布集中的趋势越明显,比如天文学、生物学、气象学和数学。学科细分之后论文的引用分布又是如何呢?我们从生物学中随机抽出6个子类,整理分析其2010年发表的论文的引用分布情况结果见表4。从表4中可以看出生物学各子类引用分布与5大学科的引用分布基本相似,呈现偏态分布,高被引区论文的引用贡献

率基本上变化不大,以生物学的引用贡献率为核心上下波动,这种引用分布的集中趋势基本上相对稳定。

表1 五种学科2010年发表的论文的引用分布情况

(篇)

	军事学	眼科学	天文学	生物学	气象学	数学
论文总数	10281	6490	39048	22065	8141	25124
被引频次大于等于10次文章篇数(被引频次和)	0	0	169(2386)	72(1017)	73(906)	60(945)
被引次数						
9	1	0	79	32	32	22
8	1	0	111	44	47	38
7	2	1	181	84	62	61
6	4	3	276	121	84	75
5	13	11	416	245	105	179
4	14	31	779	423	236	330
3	60	108	1589	902	444	657
2	221	328	3275	1961	778	1628
1	839	1221	7627	4344	1526	4487
高被引区论文引用贡献率(%)	50.64	50.41	58.54	57.05	60.87	57.21

表2 五种学科2009年发表的论文的引用分布情况

(篇)

	军事学	眼科学	天文学	生物学	气象学	数学
论文总数	10518	6742	37203	22445	7156	27895
被引频次大于等于10次文章篇数(被引频次总和)	12(139)	8(95)	1104(16600)	523(7340)	340(4865)	557(8824)
被引次数						
9	4	6	287	176	94	157
8	8	10	419	244	121	216
7	7	17	564	348	149	319
6	22	32	793	516	206	434
5	46	64	1072	736	255	619
4	76	133	1676	1070	377	1021
3	198	263	2547	1657	534	1712
2	411	663	4083	2678	747	2834
1	1104	1503	7094	4229	1156	5650
高被引区论文引用贡献率(%)	55.00	54.58	65.27	62.91	64.70	64.74

表3 五种学科2008年发表的论文的引用分布情况

(篇)

	军事学	眼科学	天文学	生物学	气象学	数学
论文总数	9128	6857	35257	23563	6647	25531
被引频次大于等于10次文章篇数(被引频次总和)	27(120)	34(489)	2550(43125)	1466(22692)	821(14281)	839(14200)
被引次数						
9	18	22	541	284	147	177
8	24	34	678	485	163	268
7	39	56	833	615	194	337
6	56	71	1102	839	232	408
5	79	167	1456	1080	268	760
4	124	302	2033	1457	379	1097
3	225	411	2913	2087	492	1711
2	467	803	4049	2905	634	2807
1	1087	1503	5970	3993	962	5181
高被引区论文引用贡献率(%)	58.94	59.12	67.76	65.14	68.67	67.34

表4 生物学各子类2010年发表的论文的引用分布情况

(篇)

	普通生物学	细胞学	遗传学	生物化学	古生物学	微生物学理论 及其应用
论文总数	478	438	671	1389	618	4129
被引频次大于等于10次文章篇数(被引频次总和)	4(56)	0	2(22)	3(32)	0	15(238)
被引次数						
9	1	1	0	1	0	6
8	0	0	2	3	0	6
7	5	1	2	3	1	17
6	3	2	3	9	4	24
5	10	6	3	23	4	43
4	12	4	10	25	6	69
3	25	9	18	59	11	192
2	28	36	65	125	36	387
1	81	90	130	296	97	920
高被引区论文引用贡献率(%)	61.68	54.90	52.39	56.42	54.30	57.44

2.2 期刊论文引用的分布规律

我们从眼科学、生物学以及数学各学科中随机各抽取5种期刊,分析不同学科的不同期刊2010~2008年所发表论文的引用分布情况,结果见表5~13。各表中的数据显示期刊的引用分布和各学科的基本一致,只是由于期刊的论文数较少,有些期刊可能会受到个别异常引用的影响,但是大多

数期刊的高被引区论文引用贡献率基本上处于50%~65%之间。随着论文发表时间的延长这种占期刊已引用论文30%的文章提供50%以上的引用贡献率的引用分布规律有略微集中的趋势,但其变化幅度不大。期刊论文在发表初期并未得到充分的被引用,大多数杂志的零被引率很高,随着发表时间的推移期刊论文的被引逐渐达到引用高峰。

表5 眼科学中5种期刊2010年发表的论文的引用分布情况

(篇)

	眼科新进展	眼科研究	国际眼科 杂志	中国中医 眼科杂志	眼外伤职业 眼病杂志
论文总数	339	323	993	154	376
被引频次大于等于10次文章篇数(被引频次总和)	0	0	0	0	0
被引次数					
9	0	1	0	0	0
8	0	0	0	0	0
7	1	0	0	1	0
6	0	1	3	0	0
5	3	1	5	0	1
4	3	1	11	0	0
3	12	5	27	8	5
2	30	17	87	12	14
1	89	82	237	26	68
高被引区论文引用贡献率(%)	52.42	51.23	51.47	50.86	46.90

表6 眼科学中5种期刊2009年发表的论文的引用分布情况

(篇)

	眼科新进展	眼科研究	国际眼科 杂志	中国中医 眼科杂志	眼外伤职业 眼病杂志
论文总数	358	445	1001	181	444
被引频次大于等于10次文章篇数(被引频次总和)	1(11)	1(10)	3(41)	1(12)	1(10)
被引次数					
9	0	1	8	2	0
8	0	3	2	2	0
7	4	2	6	0	0
6	3	5	12	5	0
5	5	6	22	2	4
4	11	10	29	6	8
3	23	15	76	14	11
2	38	50	146	16	34
1	91	96	215	39	103
高被引区论文引用贡献率(%)	57.70	57.64	57.51	60.22	61.39

表7 眼科学中5种期刊2008年发表的论文的引用分布情况

(篇)

	眼科新进展	眼科研究	国际眼科 杂志	中国中医 眼科杂志	眼外伤职业 眼病杂志
论文总数	389	382	985	189	431
被引频次大于等于10次文章篇数(被引频次总和)	6(71)	5(102)	24(287)	0	1(13)
被引次数					
9	6	3	9	4	0
8	3	2	17	1	1
7	4	3	23	2	7
6	10	6	27	3	7
5	19	14	51	3	16
4	35	18	61	12	13
3	36	21	105	10	35
2	69	50	114	19	66
1	80	78	199	38	105
高被引区论文引用贡献率(%)	52.83	63.42	59.19	59.92	57.00

表8 数学中5种期刊2010年发表的论文的引用分布情况

(篇)

	工程数学学报	计算数学	模糊系统 与数学	应用数学 学报	应用数学
论文总数	122	39	132	115	156
被引频次大于等于10次文章篇数(被引频次总和)	0	0	0	0	0
被引次数					
9	0	0	1	0	0
8	0	0	0	1	0
7	0	0	0	0	0
6	2	0	1	0	1
5	0	1	1	1	0
4	2	3	4	1	3
3	1	2	4	6	3
2	6	3	11	11	7
1	40	5	30	23	30
高被引区论文引用贡献率(%)	52.40	51.76	57.20	53.50	55.49

表9 数学中5种期刊2009年发表的论文的引用分布情况

(篇)

	工程数学学报	计算数学	模糊系统与数学	应用数学学报	应用数学
论文总数	139	32	147	102	137
被引频次大于等于10次文章篇数(被引频次总和)	2(25)	0	1(12)	2(34)	0
被引次数					
9	0	0	2	1	2
8	1	0	0	2	0
7	1	0	1	1	2
6	2	0	2	1	4
5	2	1	6	2	1
4	3	3	4	4	6
3	11	3	3	10	5
2	18	4	20	21	12
1	27	7	33	15	26
高被引区论文引用贡献率(%)	58.59	51.71	61.69	60.59	61.47

表 10 数学中 5 种期刊 2008 年发表的论文的引用分布情况

(篇)

	工程数学期刊	计算数学	模糊系统与数学	应用数学期刊	应用数学
论文总数	153	40	151	112	144
被引频次大于等于 10 次文章篇数(被引频次总和)	2(22)	2(26)	10(112)	4(46)	6(77)
被引次数					
9	3	0	1	2	0
8	0	2	3	1	0
7	0	1	5	1	2
6	10	3	2	1	3
5	5	5	7	5	4
4	6	1	14	10	8
3	13	6	9	14	0
2	28	4	13	15	19
1	34	8	37	18	26
高被引区论文引用贡献率(%)	59.55	57.69	63.32	58.00	66.93

表 11 生物学中 5 种期刊 2010 年发表的论文的引用分布情况

(篇)

	病毒学报	昆虫学报	生命科学	生态学报	生物学杂志
论文总数	65	95	204	270	206
被引频次大于等于 10 次文章篇数(被引频次总和)	1(18)	0	2(29)	9(110)	1(10)
被引次数					
9	1	0	0	4	0
8	0	0	0	4	0
7	0	0	0	4	1
6	0	0	2	8	1
5	2	1	2	12	2
4	0	3	2	26	8
3	3	3	8	23	7
2	5	14	15	49	34
1	12	21	30	69	51
高被引区论文引用贡献率(%)	74.12	51.47	61.59	61.80	53.17

表 12 生物学中 5 种期刊 2009 年发表的论文的引用分布情况

(篇)

	病毒学报	昆虫学报	生命科学	生态学报	生物学杂志
论文总数	75	90	175	332	182
被引频次大于等于 10 次文章篇数(被引频次总和)	1(12)	1(11)	4(55)	74(1101)	6(84)
被引次数					
9	0	1	0	13	0
8	0	4	3	14	3
7	0	1	2	16	5
6	4	3	3	22	5
5	2	8	3	21	6
4	4	5	6	35	16
3	2	8	10	29	15
2	12	14	22	48	22
1	11	10	33	25	47
高被引区论文引用贡献率(%)	59.42	55.03	63.97	60.92	62.78

	病毒学报	昆虫学报	生命科学	生态学报	生物学杂志
论文总数	68	106	206	320	150
被引频次大于等于 10 次文章篇数(被引频次总和)	8(144)	8(100)	11(174)	134(2205)	7(93)
被引次数					
9	0	5	0	10	4
8	0	5	0	24	5
7	2	5	4	21	6
6	5	5	10	17	6
5	3	9	12	19	11
4	4	7	13	27	12
3	6	12	18	14	14
2	9	17	31	16	19
1	10	13	31	11	36
高被引区论文引用贡献率(%)	59.85	58.33	63.34	56.38	61.59

3 结语

期刊论文引用分布真实反映了期刊论文被利用的过程。同时,期刊论文引用的分布也反映出期刊发表后存在着核心论文的事实。本研究中我们通过对大的学科、学科子类一直到期刊论文引用分布情况进行分析,观察了相当于不同级别抽样的论文引用分布情况。期刊论文引用分布基本上呈现一种偏态分布,由于我国期刊的零被引率较高^[5],期刊论文的引用能够达到引用高峰的时间较长,通常计算影响因子时认为论文在发表两年之后可以达到引用高峰,然而我们发现期刊论文历年的论文被引用率在发表后 2~6 年间呈现逐年上升的趋势,而且高被引论文率也是逐年增高的。国内很多学者研究认为国内期刊论文引用的峰值并不是发表后的 2 年^[6-8],而是相对滞后一些,我们本次通过论文引用分布分析也发现此现象。

本次我们研究发现期刊论文发表后存在着核心区域,30% 已发生被引的论文提供 50% 以上的引用频次。由于国内期刊发表后两年内零被引率较高,所以这里我们只研究已经发生被引的论文中的规律。上面提到的核心区的引用规律随着论文发表时间的推移有集中的趋势,但是变化不太大。随着时间的推移一些未被引用的论文也会获得引用,一些低被引的论文会获得较高的被引。但是这种核心区引用规律对于大多数期刊来说有集中的趋势,而不会出现大的分散现象。如果我们把期刊论文的引用看做一个随机事件,当然存在着事件概率问题,高质量的文章被引的概率当然要高于低质量的论文。此外由于文献引用行为的一些偏差和参考文献著录的一些问题^[9-13],使得低被引很难反映论文的质量。因此本研究提示期刊论文引用核心区的论文对期刊被引的贡献率较大,也能充分反映出期刊质量。但是期刊论文发表后何时才能达到引用增长的高峰期,是值得我们进一

步去研究的问题。我们推测在增长高峰期期刊论文的核心区论文才是质量较高的论文。

参考文献

- 1 王群英,林耀明.影响因子、总被引频次与期刊载文量的关系研究——以资源、生态、地理方面的 8 个期刊为例.中国科技期刊研究,2012,23(1):76-79
- 2 毛定娟,王鲁燕.影响因子与 h 指数在期刊影响力评价中的比较分析.科技情报开发与经济,2011,21(28):1-3
- 3 刘艳华,华薇娜.期刊评价新指标——特征因子.情报杂志,2010,29(7):122-126
- 4 高巍.影响因子在期刊评价中的应用及其局限性.山西财经大学学报,2011,33(2):227.
- 5 刘雪立,方红玲,周志新,董建军,盛丽娜,苗媛,付中静.科技期刊反向评价指标——零被引论文率及其与其他文献计量学指标的关系.中国科技期刊研究,2011,(4):525-528
- 6 方红玲.2003~2008 年眼科学高被引论文统计分析.中国科技期刊研究,2010,21(2):197-200
- 7 姜春霞.《郑州大学学报(医学版)》2006~2008 年载文被引情况分析.中国科技期刊研究,2010,21(3):313-315
- 8 张坤,赵芬侠,曹龙.我国林业类核心期刊高被引论文统计分析.中国科技期刊研究,2011,22(4):549-554
- 9 陶范.学术论著应杜绝无效引用.编辑学报,2007,19(3):173-175
- 10 彭桃英.学术论文参考文献的隐性错误分析.中国科技期刊研究,2010,21(3):368-371
- 11 马诚.参考文献引用及其研究的盲点与误区.编辑学报,2007,19(2):87-89
- 12 陶范.参考文献的转引现象探析.编辑学报,2006,18(3):199-200
- 13 朱红梅,张大志,孙宇航,任红.高影响力医学期刊参考文献引用错误分析.中国科技期刊研究,2012,23(2):243-247