

改进的 p 指数测度单篇论文学术质量的探讨*

刘运梅 李长玲 冯志刚 刘小慧

山东理工大学科技信息研究所 淄博 255049

摘要: [目的/意义]继承 p 指数的思想,提出基于引证文献的单篇论文评价指标 p_q 指数、基于参考文献和引证文献的综合性单篇论文评价指标 p_c 指数,以综合评价单篇论文的学术质量。[方法/过程]选择图书情报学 2013 年的载文做样本数据,对 p_q 指数、 p_c 指数进行验证。[结果/结论]发现在测度单篇论文学术质量时, p_q 指数不仅继承了 h 指数、学术迹基于引证文献评价单篇论文的优势,而且分别在数值区分度、计算过程复杂度方面优于 h 指数和学术迹。 p_c 指数将单篇论文的引证文献、参考文献综合在一个评价体系中,不仅具有 p_q 指数、h 指数、学术迹的性能,而且为单篇论文评价方法提供了一个全新的视角,具有全面反映学术论文内容质量和学术影响力的综合优势。

关键词: p 指数 单篇论文评价 p_q 指数 p_c 指数 h 指数 学术迹

分类号: G250

DOI: 10.13266/j.issn.0252-3116.2017.21.013

1 引言

长期以来,对各种学术评价对象的科学计量评价方法已经形成一定的规模,并相对比较成熟。其中,影响较大的影响因子^[1]、h 指数^[2]以及新兴的学术迹^[3-4]、p 指数^[5]、替代计量学^[6]等,被广泛应用于评价人才、科研机构、期刊乃至国家的科研能力与水平。然而,任何伟大的发明创造,都必然要落实到具体的论著中^[7]。学术论著是学术评价的终极对象,尤其是对单篇论文学术价值的判定奠定了学术评价的基石,只有在合理评价单篇学术论文的基础上,评价机构、人才等的科研贡献,其评价方法才会有较强的说服力。因此,针对单篇论文学术质量的有效评价意义重大。

起初,单篇论文评价指标只是用被引频次、他引频次、发文期刊影响因子等来衡量;2005年,J. E. Hrish 提出了 h 指数^[2],很快被应用于评价不同层次的学术对象,随后 A. Schubert 通过引证文献的施引 h 指数评价单篇论文的学术影响力^[8-9];吴勤于 2007 年提出用引证强度评价单篇论文^[10];另外,钟文一提出了基于引文系数的论文影响力评价^[11];2014 年,叶鹰将测度

引文偏态分布的科研影响力评价指标 I3 与 h 指数相结合,提出了一个新的学术评价指标——学术迹^[3],学术迹将所有优劣纳入测度,实现了定量、多维度测量学术成果的评价;王术等提出将影响矩^[12]作为测度单篇论文影响力的评价指标,兼顾了学术评价的理论继承性和计算简洁性,为单篇论文的评价增添了新维度^[7]。

上述单篇论文评价指标或多或少存在一些局限性,具体概括为以下三点:①只能反映单篇论文的一方面性质:被引频次、他引频次、发文期刊影响因子等单指标仅反映了节点文献一方面的特性,在一定程度上具有片面性。②只能评价高被引论文质量:h 指数在评价单篇论文时,通过考察引证文献中 h 核的数量作为评价参数,只能反映高被引和优质论著的特征^[9],并不能综合考察引证文献总体的数量和质量分布。③以单篇论文的被引情况片面涵盖论文的学术质量:以学术迹、h 指数等为代表的复合评价指标只考虑了单篇论文的引证文献和引证文献的被引情况,反映的是单篇论文的学术影响力,具有一定片面性。

因此,本文改进综合性科技评价指标——p 指数^[5],从单篇论文的引证文献和参考文献两方面综合

* 本文系国家自然科学基金项目“基于量化与质性数据的跨学科合作行为研究”(项目编号:16BTQ078)研究成果之一。

作者简介:刘运梅(ORCID:0000-0002-8921-0826),硕士研究生;李长玲(ORCID:0000-0001-6266-4820),教授,硕士生导师,硕士,通讯作者,E-mail:lichl69@163.com;冯志刚(ORCID:0000-0002-1166-6698),硕士研究生;刘小慧(ORCID:0000-0002-7233-7820),硕士研究生。

收稿日期:2017-06-16 修回日期:2017-08-23 本文起止页码:106-113 本文责任编辑:王善军

测度单篇论文的学术质量,以解决单指标评价单篇论文的片面性,弥补 h 指数只能评价高影响力文献的局限性,打破以学术迹、h 指数等为代表的复合评价指标过分依赖被引用的制约性。

2 单篇论文评价指标的计算方法

2.1 评价单篇论文的 h 指数与学术迹

2.1.1 单篇论文的 h 指数计算方法 2005 年 J. E. Hrisch 提出 h 指数综合评价作者、机构的发文数量与被引用情况^[2]。2009 年 A. Schubert 将 h 指数延伸,用于定量评价单篇论文的学术质量。其计算方法为^[8]:如果引用某论文的所有论著中有 n 篇论文每篇至少被引用了 n 次,那么该论文的 h 指数就是 n。

A. Schubert 的 h 指数综合考虑了单篇论文被引的文献数量以及所有引用该论文的其他论著的被引频次两项指标,不仅能体现论文的直接影响,也能反映论文的间接影响。但是,由于单篇论文的 h 指数选取引证文献中 h 核的数量作为评价参数,因此,其只适用于评价高被引论文的学术质量^[13]。

2.1.2 单篇论文学术迹的计算方法 2014 年 F. Y. Ye 和 L. Leydesdorff 提出了学术迹,将其定义为学术矩阵的迹。其中,学术矩阵是引文曲线按发文的被引量排布,由 h 指数将引文曲线分为 h 核、h 尾和零被引三区段,并由这三个区段的发文和引文分布构成的三阶矩阵^[3]。在学术迹被提出后,唐继瑞将其应用于定量评价单篇论文的学术质量。单篇论文的学术迹 T 计算公式如下^[7]:

$$T = \frac{h^4(C_h - h^2)^2}{C} + \frac{(P - h - P_z)^2 - P_z^2}{P} \quad (\text{公式 1})$$

其中 P 表示单篇论文的被引频次; h 表示单篇论文的 h 指数; C 表示单篇论文所有引证文献的总被引频次; P_z 表示引证文献中零被引论文数; C_h 表示引证文献中被引频次大于或等于 h 的引证文献数量。

单篇论文学术迹综合考察了单篇论文的引证文献数量与质量分布,将引证文献分成三个部分: h 核、h 尾和零被引部分,并结合 I3 指数思想构建学术矩阵,最后通过矩阵的迹来度量论文的学术质量^[14]。学术迹测算时考虑了引文曲线的整体分布,涉及 P、h、C、 C_h 、 P_z 五个参数,其计算过程相对较复杂。

2.2 基于 p 指数的评价单篇论文的 p_q 指数

2010 年 G. Prathap 教授提出了一个综合性人才评价指标——p 指数^[5]。如果将某学者的论文发表数量记作 N,总被引次数记作 C,则 p 指数计算公式

为^[15]:

$$p = \left(\frac{C^2}{N} \right)^{\frac{1}{3}} \quad (\text{公式 2})$$

G. Prathap 认为 $(C^2/N)^{1/3}$ 的内涵丰富,可以被看作一个综合指标,具有模拟 h 指数的功能^[5]。他运用 p 指数计算 100 位经济学家的科研水平,并与其它传统评价指标进行相关性分析,验证了在进行人才评价时 p 指数较 h 指数具有更好的适用性^[15];他还运用 iCE 地图对全球 20 个科研强国及化学领域 100 种顶尖期刊的科研影响力进行了对比分析^[16-17];以影响因子排名前 10 的期刊为样本,对 p 指数、载文量、总被引量、影响因子和特征因子、论文影响分值等指标的相关性进行了探讨^[18]。这一系列的研究成果验证了 p 指数在应用于评价国家、科研机构、期刊和人才方面具有较好的通用性。但目前,国内外并没有 p 指数在评价单篇论文学术质量方面的研究。

本文继承上述 p 指数的思想,从单篇论文的被引角度,基于引证文献的数量及其被引频次提出测度单篇论文学术影响力的 p_q 指数。其定义如下:

定义 1 基于引证文献评价单篇论文的 p_q 指数:节点文献的引证文献总被引频次与其篇均被引频次乘积的三次开方。令某篇论文的被引频次为 QN,即该论文被 QN 篇论文引用,其中,第 j 篇引证文献的被引频次为 QC_j ,则基于引证文献的表征单篇论文学术影响力的 p_q 指数计算公式为:

$$p_q = \left(\sum_{j=1}^{QN} QC_j \times \frac{\sum_{j=1}^{QN} QC_j}{QN} \right)^{\frac{1}{3}} \quad (\text{公式 3})$$

2.3 综合评价单篇论文学术质量的 p_c 指数

被引频次是进行论文评价的重要指标之一。目前大部分的科学计量评价研究以单篇论文的被引频次以及引用该论文的其他论著的被引频次两项指标为重,很少有学者注意到参考文献对论文学术质量的影响。但是,参考文献往往是决定论文学术水平和学术质量的重要依据,具有极其重要的学术价值。阅读并且借鉴高质量的参考文献,是从事科学研究的前提,参考文献代表着论文作者前期工作的深度和广度。同时,有研究表明,参考文献可作为论文评价体系中独立于被引频次的评价指标^[19]。

因此,本文基于上述 p 指数的思想,从节点文献的参考文献、引证文献两个角度,根据引证文献数量、二级引证频次(即引证文献的被引频次)、参考文献数量以及参考-引证频次(即参考文献的被引频次),计算

相应的 p 指数,并给予合理的权重,提出一个新的综合性单篇论文学术质量测度指标 p_c 指数,同时反映学术论文的内容质量和学术影响力。 p_c 指数的构建模型如图 1 所示:

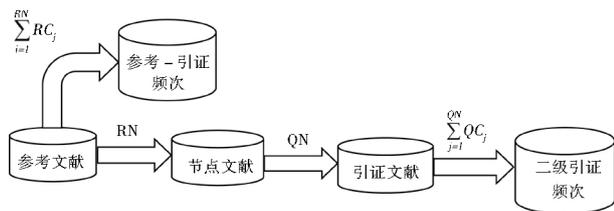


图 1 综合性单篇论文测度指标 p_c 指数模型

参考基于引证文献的 p_q 指数构建思想,将节点文献的参考文献数量及参考-引证频次作为测度的基本指标,提出基于参考文献的 p_r 指数。其定义如下:

定义 2 基于参考文献评价单篇论文的 p_r 指数:节点文献的参考文献总被引频次与其篇均被引频次乘积的三次开方。假设某篇论文有 RN 篇参考文献,其中,第 i 篇参考文献的被引频次为 RC_i ,则基于参考文献的表征单篇论文质量的 p_r 指数计算公式为:

$$p_r = \left(\sum_{i=1}^{RN} RC_i \times \frac{\sum_{i=1}^{RN} RC_i}{RN} \right)^{\frac{1}{3}} \quad (\text{公式 4})$$

为了综合考虑单篇论文的学术质量,本文结合节点文献的参考文献和引证文献,为公式 3 的 p_q 指数和公式 4 的 p_r 指数分配合理的权重,将分配权重后的两个指标加和,得到一个综合测度单篇论文学术质量的评价指标 p_c 指数。其定义如下:

定义 3 综合测度单篇论文学术质量的 p_c 指数:综合基于引证文献的 p_q 指数和基于参考文献的 p_r 指数,为其分配相应的权重 w_1 和 w_2 ,全面反映单篇论文的学术质量。 p_c 指数计算公式为:

$$p_c = w_1 \times p_q + w_2 \times p_r$$

$$= w_1 \times \left[\frac{(\sum_{j=1}^{QN} QC_j)^2}{QN} \right]^{\frac{1}{3}} + w_2 \times \left[\frac{(\sum_{i=1}^{RN} RC_i)^2}{RN} \right]^{\frac{1}{3}} \quad (\text{公式 5})$$

$$\text{且 } 0 \leq w_1, w_2 \leq 1, w_1 + w_2 = 1$$

3 实例验证

3.1 数据来源及统计计算

本文选择中国学术期刊网络出版总库(CNKI)作为来源数据库,在期刊检索页面的文献分类目录选择“信息科技”大类下的“图书情报与数字图书馆”分类,以限定数据样本为图书情报学领域。普赖斯提出,科

学论文一般在发表 2 年后,达到被引用峰值,由于本文研究数据涉及二级引证文献,因此选择具有 4 年时间差的 2013 年发表论文为样本数据。具体的时间推算过程如图 2 所示:



图 2 时间窗推算过程

选择图书情报学领域的期刊论文,时间限定 2013 年,检索共得到 22 674 条结果。选择论文被引频次最高的前 40 篇论文,由于部分外文文献数据的缺失,去掉其中 5 篇,最终得到 35 篇论文作为数据分析样本。借助 Python 爬虫工具,统计这 35 篇论文的被引频次 QN ,按被引频次由高到低排序,并给论文编号 1-35,相关数据见表 1 第一、二列。由于网上电子公告[EB/OL]、报纸[N]、图书[M]、联机网上数据库[DB/OL]等文献类型的参考文献没有被引情况的统计数据,因此,本文将这些文献类型的参考文献忽略不计,只统计可查被引情况的有效参考文献。其中,中文期刊论文和学位论文的被引情况通过知网数据库获得,外文期刊论文的被引频次通过 WoS 数据库获得。统计可查被引的参考文献数量 RN (见表 1 第四列)、每篇引证文献的被引频次(QC_j)、每篇可查被引参考文献的被引频次(RC_i)。然后依据公式 3 和公式 4,分别计算 35 篇论文所有引证文献的总被引频次(QC)、所有可查被引的参考文献的总被引频次(RC)、 p_q 指数和 p_r 指数,分别见表 1 第三、五、六、七列。

3.2 参数确定

此时,通过 35 篇高被引论文的被引频次(QN)、二级引证频次(QC)、参考文献数量(RN)、参考-引证频次(RC)四个指标得到了相应的 p_q 指数和 p_r 指数。由于计算 p_c 指数的 w_1 、 w_2 未知,需要通过数据验证得到 p_q 指数、 p_r 指数最佳的权重分配。

p_c 指数是结合了引证文献和参考文献两个评价角度的综合性指标,本文认为 p_c 指数期望的最佳取值应是 p_q 指数和 p_r 指数对 p_c 指数占同等重要的贡献度。由于参考文献与引证文献的功能与性质不同,两者的数量及其被引情况差别很大(参照表 1 第二列到第五列数据可知)。因而, p_q 指数与 p_r 指数所占权重就不能一分为二各占 0.5。那么,这就需要平衡两者对评

表 1 2013 年图书情报学领域高被引论文各指标量统计计算结果

论文序号	被引频次 QN	二级被引频次 $QC = \sum_{j=1}^{QN} QC_j$	可查被引的参考文献数量 RN	参考 - 被引频次 $RC = \sum_{i=1}^{RN} RC_i$	p_q 指数	p_r 指数
1	233	1 876	2	239	24.719	30.567
2	149	473	3	412	11.451	38.391
3	148	624	2	232	13.805	29.967
4	133	79	0	0	3.607	0
5	132	293	17	809	8.664	33.766
6	130	714	2	254	15.769	31.833
7	115	454	4	128	12.147	16.000
8	114	333	4	814	9.908	54.920
9	108	452	3	134	12.368	18.156
10	105	645	2	295	15.824	35.173
11	100	375	2	1133	11.204	86.260
12	97	288	6	822	9.492	48.291
13	96	401	2	48	11.876	10.483
14	90	405	0	0	12.215	0
15	89	475	7	1012	13.635	52.693
16	87	207	7	1001	7.897	52.311
17	87	485	9	714	13.931	38.405
18	84	469	3	112	13.783	6.110
19	83	445	0	0	13.362	0
20	81	383	7	854	12.189	47.055
21	77	191	9	199	7.796	16.387
22	67	108	11	954	5.584	43.575
23	67	272	5	297	10.336	26.032
24	62	107	3	27	5.695	6.240
25	62	349	3	376	12.524	36.120
26	61	328	15	1488	12.082	52.849
27	61	141	0	0	6.882	0
28	58	109	2	190	5.895	26.232
29	56	95	3	59	5.442	10.508
30	55	154	3	44	7.555	8.642
31	53	214	0	0	9.525	0
32	52	59	9	266	4.060	19.884
33	52	165	3	108	8.060	15.724
34	52	160	14	795	7.896	35.607
35	49	83	3	126	5.200	17.426

注: 数据采集时间为 2017.05.01 - 05.05。表 1 第四列可查参考文献数量较少, 是因为本文选择了 2013 年的 35 篇高被引论文作为数据样本, 其中大部分论文的主题是关于微信、大数据等一些新的研究热点, 参考文献中网络文献数量较多, 期刊论文、学位论文等可查被引的参考文献数量较少。

价结果的影响程度。本文认为可以对 w_1 和 w_2 赋予不同的值, 分别计算 p_q 指数、 p_r 指数在不同赋值权重下, 与 p_c 指数的 Pearson 相关系数, 用以表示 p_q 指数与 p_r 指数对评价结果的影响程度。用 $r(p_q-p_c)$ 表示 p_q 指数与 p_c 指数的相关系数, 用 $r(p_r-p_c)$ 表示 p_r 指数与 p_c 指数的相关系数。当 $r(p_q-p_c) = r(p_r-p_c)$ 时, 对应的 w_1 、 w_2 即是期望中的最佳取值。 w_1 、 w_2 取值的判定过程如下:

为了使计算过程和计算结果方便简练, 本文将 w_1 、 w_2 的取值精确到 0.1 位, 由于 $0 \leq w_1, w_2 \leq 1, w_1 + w_2 = 1$, 所以共有 11 种 w_1 、 w_2 的赋值方案。将每一种方案的 w_1 、 w_2 取值代入到公式 5 中, 得到相应的 p_c 指数, 并利用 SPSS 18.0 统计分析软件分别计算 p_q 指数、 p_r 指数与 p_c 指数的 Pearson 相关系数。实验结果见图 3。

其中 横坐标表示 w_2 取值由 0 到 1 变化, 纵坐标

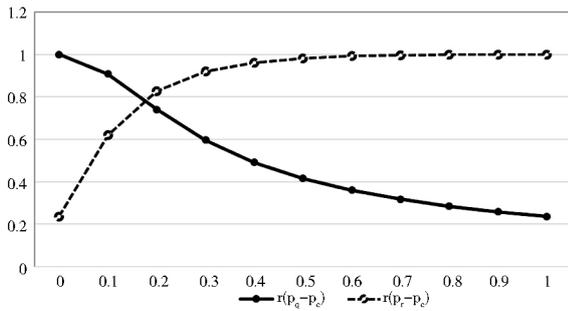


图 3 不同权重的 pearson 相关系数变化趋势图

数值代表 Pearson 相关系数; 实线表示随着 w_1 取值的逐渐减小、 w_2 取值的逐渐增大, p_q 指数与 p_c 指数的 Pearson 相关系数变化趋势; 虚线表示 p_r 指数与 p_c 指数的 Pearson 相关系数变化趋势。

根据图 3 可以发现, 当 w_2 取值在 0.2 附近时, 两曲线相交。也就是, 大约当 $w_1=0.8$, $w_2=0.2$ 时, p_q 指数与 p_c 指数的相关系数近似等于 p_r 指数与 p_c 指数的

相关系数。此时, 引证文献与参考文献对 p_c 指数的影响差别最小, 近似于相等。因此, 本文公式 5 中的 w_1 、 w_2 取值分别为 0.8、0.2。

事实上, w_1 、 w_2 是两个比例常数, 必须从拟合数据的经验中确定。因此, 其数值会根据数据库来源不同或学科分类不同而发生变化。

3.3 单篇论文 p_c 指数的计算与评价结果分析

根据 w_1 、 w_2 的最佳取值, 依据公式 5 计算数据样本中 35 篇高被引论文的 p_c 指数, 结果见表 2 第九列。为了对比 p_q 指数、 p_c 指数与其他单篇论文评价指标的评价效果, 本文依据 h 指数的计算方法和公式 1, 计算得到 35 篇高被引论文的 h 指数和学术迹, 计算结果分别见表 2 第三列和第五列。分别按照 h 指数、学术迹、 p_q 指数、 p_c 指数的大小对 35 篇高被引论文进行排序, 结果分别见表 2 第四列、第六列、第八列、第十列。

表 2 2013 年图书情报学领域高被引论文的各评价指标计算结果

论文序号	被引频次	h 指数	R(h 指数)	学术迹	R(学术迹)	p_q 指数	R(p_q 指数)	p_c 指数	R(p_c 指数)
1	233	17	1	114.04	3	24.72	1	25.89	2
2	149	10	11	50.39	18	11.45	16	16.84	13
3	148	11	9	66.46	12	13.80	5	17.04	12
4	133	4	34	-64.60	35	3.61	35	2.89	35
5	132	9	18	9.34	28	8.66	22	13.68	16
6	130	15	2	154.78	1	15.77	3	18.98	7
7	115	11	9	77.84	8	12.15	13	12.92	21
8	114	9	18	28.39	22	9.91	19	18.91	8
9	108	12	5	74.38	9	12.37	10	13.53	17
10	105	13	4	90.77	4	15.82	2	19.69	5
11	100	10	11	57.27	14	11.20	17	26.21	1
12	97	10	11	67.33	11	9.49	21	17.25	10
13	96	10	11	52.77	17	11.88	15	11.10	22
14	90	12	5	84.91	6	12.21	11	9.77	25
15	89	12	5	88.12	5	13.64	7	21.45	3
16	87	9	18	54.95	15	7.90	24	16.78	14
17	87	12	5	79.54	7	13.93	4	18.83	9
18	84	14	3	145.10	2	13.78	6	14.25	15
19	83	10	11	43.88	19	13.36	8	10.69	23
20	81	8	22	12.00	27	12.19	12	19.16	6
21	77	7	24	24.17	25	7.80	26	9.51	27
22	67	5	30	-2.32	33	5.58	31	13.18	20
23	67	10	11	74.20	10	10.34	18	13.48	18
24	62	6	29	7.33	29	5.69	30	5.80	33
25	62	8	22	25.01	24	12.52	9	17.24	11
26	61	10	11	63.29	13	12.08	14	20.24	4
27	61	7	24	31.31	21	6.88	28	5.51	34
28	58	5	30	1.18	32	5.89	29	9.96	24

(续表 2)

论文序号	被引频次	h 指数	R(h 指数)	学术迹	R(学术迹)	p_q 指数	R(p_q 指数)	p_c 指数	R(p_c 指数)
29	56	5	30	2.59	31	5.44	32	6.46	32
30	55	7	24	23.55	26	7.55	27	7.77	28
31	53	9	18	54.21	16	9.52	20	7.62	30
32	52	4	34	-6.14	34	4.06	34	7.23	31
33	52	7	24	27.84	23	8.06	23	9.59	26
34	52	7	24	31.57	20	7.90	24	13.44	19
35	49	5	30	6.96	30	5.20	33	7.65	29

为了进一步比较和量化 h 指数、学术迹、 p_q 指数、 p_c 指数这四个评价指标的评价结果,本文利用 SPSS 18.0 统计分析软件对表 2 中这四项评价指标的排序结果进行 Spearman 秩相关分析,结果见表 3。根据相关系数 r 判断两变量相关的密切程度, $|r| \geq 0.7$ 表示两变量高度相关, $0.4 \leq |r| < 0.7$ 表示两变量中度相关, $|r| < 0.4$ 表示两变量低度相关^[20]。

表 3 四个评价指标排序结果的 Spearman 相关性分析

指标	h 指数	学术迹	p_q 指数	p_c 指数
h 指数	Spearman 相关系数 r	0.956 **	0.922 **	0.681 **
	显著性(双尾)	0.000	0.000	0.000
	df	35	35	35
学术迹	Spearman 相关系数 r	0.956 **	0.840 **	0.610 **
	显著性(双尾)	0.000	0.000	0.000
	df	35	35	35
p_q 指数	Spearman 相关系数 r	0.922 **	0.840 **	0.730 **
	显著性(双尾)	0.000	0.000	0.000
	df	35	35	35
p_c 指数	Spearman 相关系数 r	0.681 **	0.610 **	0.730 **
	显著性(双尾)	0.000	0.000	0.000
	df	35	35	35

注: **在 0.01 水平(双尾)上显著相关

通过以上数据计算与相关性分析,可以发现:

(1) p_q 指数、 p_c 指数继承了 h 指数、学术迹的优势,有识别被引质量较低的高被引论文的能力。h 指数、学术迹、 p_q 指数、 p_c 指数这四项评价指标都综合考虑了单篇论文被引的文献数量以及所有引用该论文的其他论著的被引情况,既能反映单篇论文的直接影响,也能反映单篇论文的间接影响。从表 2 中可以看到,虽然被引频次最高的前十篇论文(即论文序号 1~10)具有较高的被引频次,但有一半的论文 h 指数、学术迹、 p_q 指数、 p_c 指数偏低,四项指标的排名在第十位之外。例如,论文 4 的被引频次居 35 篇论文的第四位,但其 h 指数却排倒数第二位,学术迹、 p_q 指数、 p_c 指数甚至排到了最后一位。这是由于 h 指数、学术迹、 p_q 指

数不仅考虑了单篇论文被引的文献数量,还将引证文献的被引频次分布情况考虑在内。除此之外, p_c 指数还将单篇论文的参考文献及其被引情况考虑在内。因而 p_q 指数、 p_c 指数像 h 指数、学术迹一样,都能够同时反映单篇论文的直接学术影响力和间接学术影响力,改进被引频次这类单指标评价的片面性,有效避免了故意自引、假引用以及马太效应对论文质量判断的不良影响,可以有效地筛选出一批高质量的学术论文。

(2) p_q 指数、 p_c 指数、学术迹比 h 指数具有更好的区分性。表 2 中 35 篇论文的学术迹分布区间为 $[-64.60, 154.78]$, p_q 指数分布区间为 $[3.61, 24.72]$, p_c 指数分布区间为 $[2.89, 26.21]$ 。而 h 指数的分布区间仅在 $[4, 17]$,不具有较好的数值区分度。另外,35 篇论文的 h 指数值重复率很高,约有 85.7% 的论文 h 指数值有重复, h 指数为 10 的论文就有 7 篇。因此, h 指数在应用于评价单篇论文时,不能有效区分论文的学术质量,具有一定局限性。相比之下, p_q 指数、 p_c 指数、学术迹明显能够以数学运算的优势,使单篇论文学术质量的评价结果产生差异,具有区分性。

(3) p_q 指数、 p_c 指数与学术迹都能从引证文献的多角度评价单篇论文学术质量,且测算过程比学术迹简练。根据表 3 的相关性分析结果, p_q 指数与 h 指数、学术迹在 0.01 水平上都表现出高度相关, Spearman 相关系数都已经超过 0.9。说明这三项评价指标的评价效果相差不大,可以认为 p_q 指数表现了 h 指数、学术迹的大部分特征,可将 p_q 指数、h 指数、学术迹作为同类评价指标。

另外,由表 2 第五列数据可以看出,35 篇高被引论文的学术迹分布范围较广,甚至出现了负值。相比于其他三个评价指标,学术迹综合考虑了更多的因素,其测算模型用五个参数 P, h, C, C_h, P_z 反映引文曲线的整体分布,这就需要在测算过程中统计引证文献的被引频次分布情况,并计算五个参数值。因此,学术迹的测算过程相对较复杂。相比之下, p_q 指数、 p_c 指数的

计算方法较简便。 p_c 指数只需要基于引文数据库获取四个基本参数值(被引频次 Q_N 、二级引证频次 Q_C 、参考文献数量 R_N 、参考-引证频次 RC)即可。 p_q 指数仅需要两个参数值,无需测算引文的整体分布。

(4) p_c 指数在继承 h 指数、学术迹、 p_q 指数被引文献分析的基础上,利用基于参考文献的 p_c 指数对其他评价指标进行补充,具有全面反映学术论文内容质量和学术影响力的优势,可作为一个独立指标存在。由表3可以看出, p_c 指数与 h 指数、学术迹的Spearman相关系数分别为0.681、0.610,在0.01水平上均表现为中度相关。表明 p_c 指数与 h 指数、学术迹具有一定的相关性,但相关程度不是非常高。相比于 p_q 指数、 h 指数、学术迹三个指标之间的高度相关关系, p_c 指数则具有一定的相对独立性,可以作为一个独立的评价指标。

另外,从 p_c 指数的构建模型看, p_c 指数以参考文献和引证文献两个角度来评价单篇论文的学术质量。在基于 h 指数、学术迹和 p_q 指数的引文分析基础上,部分脱离了被引文献分析的范畴,通过参考文献及其被引情况反映论文研究工作和文章内容的深度、广度,是一个同时注重学术论文内容质量和学术影响力的综合性评价指标。

(5) 研究的局限性。本文以国内图书情报学领域的高被引论文为例,对 p_q 指数和 p_c 指数的评价性能进行了实证研究,得出以上结论。但在一般情况下,计量分析研究得到的结论仅仅是基于数据样本得出的。因此,本文的结论是针对图书情报领域高被引论文的一般性规律,并非一定适应于所有学科领域。若要验证本文的方法是否同样适应于其他学科领域的不同样本文献,还应基于其他学科的样本数据进行实证研究和深入分析。

4 结语

本文继承学术评价指标 p 指数的思想,提出了基于引证文献的单篇论文评价指标 p_q 指数、基于参考文献和引证文献的综合性单篇论文评价指标 p_c 指数。同时,参考应用较广的被引频次、 h 指数、学术迹,探讨不同评价方法和评价效果的差异及相关性。在测度单篇论文质量时, h 指数、学术迹、 p_q 指数、 p_c 指数都能从各自的角度取得一定的评价效果。其中, p_q 指数不仅继承 h 指数、学术迹评价单篇论文的优势,还在数值区分度、计算过程复杂度方面优于 h 指数和学术迹。 p_c 指数不仅具有 p_q 指数的性能,还基于参考文献对其他评价指标有所补充,具有全面综合反映学术论文内容质量和学术影响力的优势,为现有的单篇论文定量评

价方法提供了一个独特的参考视角,并期望能在应用中进一步发展完善。

随着科学技术和计算机网络的不断发展,各种学术评价指标也需要不断地发展和完善,无论何种评价指标都有其局限性,没有完美无瑕的测量指标,同样也不外乎于单篇论文学术质量的评价指标。 p_c 指数虽然综合考察了单篇论文的参考文献和引证文献,可作为一个全新的独立性评价指标,但是,仍未能摆脱引文的限制。本文认为在后续的评价工作中,可以深入到单篇论文具体的内容结构中,进行语义挖掘,发掘论文的潜在价值,进一步解决引文分析的局限性。

参考文献:

- [1] GARFIELD E. Citation indexing: its theory and applications in science, technology and humanities[M]. New York: Wiley, 1979.
- [2] HIRSCH J E. An index to quantify an individual's scientific research output that takes into account the effect of multiple coauthorship[J]. *Scientometrics*, 2010, 85(3): 741-754.
- [3] YE F Y, LEYDESDORFF L. The "academic trace" of the performance matrix: a mathematical synthesis of the h-index and the Integrated Impact Indicator(I3) [J]. *Journal of the American Society for Information Science and Technology*, 2014, 65(4): 742-750.
- [4] 薛菲,叶鹰. 大学的学术矩阵和学术迹探讨[J]. *大学图书馆学报*, 2014(1): 25-29.
- [5] PRATHAP G. Is there a place for a mock h-index? [J]. *Scientometrics*, 2010, 84(1): 153-165.
- [6] HAUSTEIN S, PETERS I, BAR-ILAN J, et al. Coverage and adoption of altmetrics sources in the bibliometric community [J]. *Scientometrics*, 2014, 101(2): 1145-1163.
- [7] 唐继瑞,叶鹰. 单篇论著学术迹与影响矩比较研究[J]. *中国图书馆学报*, 2015(2): 4-16.
- [8] SCHUBERT A. Using the h-index for assessing single publications [J]. *Scientometrics*, 2009, 78(3): 559-565.
- [9] 杨建林,严明. 单篇期刊论文 h 指数的实证研究——以图书情报学为例[J]. *图书情报工作*, 2010, 54(12): 145-148.
- [10] 吴勤. 基于引证强度的学术论文质量评价方法研究[J]. *情报学报*, 2007, 26(4): 522-526.
- [11] 钟文一,陈云鹏. 基于引证系数的论文影响力评价方法研究[J]. *情报科学*, 2011(5): 706-712.
- [12] 王术,叶鹰. 影响矩作为测度单篇论著影响力的评价指标探讨[J]. *大学图书馆学报*, 2014(5): 12-18.
- [13] EGGHE L. On the relation between Schubert's h-index of a single paper and its total number of received citations [J]. *Scientometrics*, 2010, 84(1): 115-117.
- [14] 陈小清,刘丽,邢美园. 单篇论著影响力评价指标比较分析——学术迹与 altmetrics 评分、F1000 评分、Comment 的比较[J]. *情报理论与实践*, 2017(3): 114-118.
- [15] PRATHAP G. The 100 most prolific economists using the p-index [J]. *Scientometrics*, 2010, 84(1): 167-172.

[16] PRATHAP G. An iCE map approach to evaluate performance and efficiency of scientific production of countries[J]. *Scientometrics*, 2010 85(1): 185-191.

[17] PRATHAP G. The iCE approach for journal evaluation[J]. *Scientometrics*, 2010 85(2): 561-565.

[18] PRATHAP G. Evaluating journal performance metrics[J]. *Scientometrics*, 2012 92(2): 403-408.

[19] 姜磊, 林德明. 参考文献对论文被引频次的影响研究[J]. *科研管理*, 2015(1): 121-126.

[20] 罗家洪, 薛茜. 医学统计学: 案例版[M]. 北京: 科学出版社, 2008.

作者贡献说明:

刘运梅: 负责数据搜集与处理、论文撰写;

李长玲: 负责研究设计、论文指导与修改;

冯志刚: 负责论文修改与校对;

刘小慧: 负责论文修改与校对。

Discussion of Improved p-index Evaluating the Academic Quality of Single Paper

Liu Yunmei Li Changling Feng Zhigang Liu Xiaohui

Institute of Scientific and Technical Information, Shandong University of Technology, Zibo 255049

Abstract: [Purpose/significance] In order to comprehensively evaluate the academic quality of single paper, this paper inherits the idea of p-index, and puts forward the p_q -index of single paper based on the citation literature, and the p_c -index of single paper based on the reference literature and citation literature. [Method/process] Taking the published paper data of 2013 in the field of library and information science as an example, this paper verified the evaluation effect of p_q -index and p_c -index. [Result/conclusion] It is found that in the measurement of the academic quality of single paper, the p_q -index not only inherits the advantage of h-index and the academic trace, but also superior to h-index and the academic trace in terms of numerical differentiation and computational complexity. The p_c -index also combines the reference literature and citation literature in a rating system. It not only has the performance of p_q -index, h-index, the academic trace, but also provides a new perspective for single paper evaluation methods, which has a comprehensive advantage of evaluating the content quality and the academic influence of academic papers.

Keywords: p-index evaluation of single paper p_q -index p_c -index h-index academic trace

“图书馆与数字人文”国际研讨会会议通知

随着数字人文研究的不断深入,国内外越来越多的图书馆开始进行相关的研究及实践。数字人文未来如何发展?图书馆在数字人文发展中应该扮演什么角色?图书馆又到底能为数字人文做些什么?基于这些问题,同时也为了进一步提升国内学术界对数字人文的认知,助力国内学术界及图书馆界把握住数字人文发展所带来的机遇,深圳大学城图书馆拟于2017年12月6日至8日举办“图书馆与数字人文”国际研讨会。

届时来自斯坦福大学、加州大学洛杉矶分校、伦敦大学学院、伦敦大学国王学院、澳大利亚国立大学、北京大学、武汉大学等全球数字人文领域及图书馆界著名专家、学者将出席本次会议。

一、会议主题:

- 数字人文中的关键技术
- 用于数字人文的数据、资源、网络、平台构建
- 高等教育与数字人文
- 数字人文与多学科应用研究

二、会议时间: 2017年12月6日至12月8日(含报到时间)

会议地点: 深圳大学城图书馆(深圳市科技图书馆)

报名和汇款截止日期: 2017年11月30日

在线报名: 在线提交注册信息。

三、联络方式:

联系人: 陈丹婷 王霞

电话: 0755-26032359 0755-26032470

邮箱: chendt@utsz.edu.cn

详情请登录本次会议主页 <http://dhl.utsz.edu.cn/>。