

Doi:10.11835/j.issn.1008-5831.pj.2021.08.001

欢迎按以下格式引用:许恩平,贾娜,李敏,等.一种时间维度的论文影响力评价研究——基于 ammaa 算法的改进[J].重庆大学学报(社会科学版),2021(6):111-124. Doi: 10.11835/j.issn.1008-5831.pj.2021.08.001.



Citation Format: XU Enping, JIA Na, LI Min, et al. A time-dimension paper influence evaluation research: Improvement based on ammaa algorithm[J]. Journal of Chongqing University (Social Science Edition), 2021(6): 111-124. Doi: 10.11835/j.issn.1008-5831.pj.2021.08.001.

一种时间维度的论文影响力评价研究 ——基于 ammaa 算法的改进

许恩平^a, 贾娜^b, 李敏^b, 余以胜^b

(华南师范大学 a. 科技处; b. 经济与管理学院, 广东 广州 510000)

摘要:针对目前 h 指数存在的不足,学界尚缺乏一种综合有效的评价指标,为此,笔者期待介绍一种论文评价的 ammaa 算法,并提出一种融入时间维度的优化算法,即 t_ammaa 算法,通过对论文影响力的评价来反映学者个人的影响力评价。研究过程中,以 Web of Science 作为数据源,聚焦国内图情领域作者发文,计算论文 ammaa 值及 t_ammaa 值,进而得出学者的 ammaa 值及 t_ammaa 值,并将两种算法结果排名与学者 H 值排名通过归一化处理,进行实证对比分析。结果表明:t_ammaa 算法综合考虑发文被引次数、被引次数的阈值限制、合著者人数及论文被引时间的异质性,既可以对独著和合著论文影响力进行综合性评价,也可以消除时间因素带来的影响,是一种更为合理的学者和论文影响力评价计量方法。

关键词: ammaa 算法; 时间异质性; 多作者论文影响力; 作者影响力; 评价指标; 破“五唯”

中图分类号: G350; G644.4 **文献标志码:** A **文章编号:** 1008-5831(2021)06-0111-14

众所周知,学者所进行的科学研究及成果传播在科学进步及社会发展中发挥着重要作用^[1],其成果形式多以学术论文、著作和专利为主。事实上,高校和科研单位的人才引进、职称评定也愈加倚重学术论文在个人学术影响力评价中的作用。2018年教育部联合多个部门,开展清理“唯论文、唯帽子、唯职称、唯学历、唯奖项”(以下简称“五唯”)专项行动。需要注意的是,在破“五唯”过程中,尤其是破“唯论文”现象时,尚应充分考虑我国科研发展的不同阶段及其具体情况,“一步到位”既不现实也无必要,笔者以为应着力于扭转不科学的教育评价导向,进而对学者进行合理的学术影响力评价。

目前,国际上广泛应用的学术影响力评价指标,如被引次数、h 指数、g 指数等均是用于评估单

基金项目: 国家社会科学基金项目“基于用户行为动机的 Altmetrics 评价模型构建与实证研究”(18BTQ075)

作者简介: 许恩平,华南师范大学科技处,Email:3553209@qq.com。

通信作者: 余以胜,华南师范大学经济与管理学院教授,Email:359239030@qq.com。

作者影响力的方法。但随着信息技术的发展,科学合作与知识交流普遍常态化,多著者合作的形式可以显著提高学术研究水平^[2],科学研究的进步需要多领域研究人员之间的合作^[3]。在理工科领域,物理学论文的合作者可达上百人,目前共同作者数量有增加的趋势,文章由不止一位作者撰写成为越来越正常的现象^[4]。但近年来,客座作者^[5]、礼物作者^[6]、非学术合作者^[7]等署名资格问题^[8]已经成为影响科研诚信的重要议题。相关研究也发现,合著者人数的增多,却未必能提高研究成果的影响力,如被引频次、下载量等^[9]。因此,学术期刊大多要求每位合著者都要作出相应的努力和贡献^[10-12],基于此,D. Gnana Bharathi 提出了一种多作者聚合分析算法指标,即 aggregating metrics for multiple authors' analysis,简称 ammaa 算法^[13]。

一、相关研究文献分析

2005年,Hirsch^[14]提出综合考虑学者发文数量和发文质量的h指数影响力评价指标,并得到学术界的认可和推广应用。同年,即被Nature文章^[15]评论道,h指数的魅力在于它能凸显那些有持久重大贡献却未得到与其声望相称荣誉的研究者。自h指数提出以来,国内外学者对其做了很多研究。随着研究的不断深入,h指数的固有缺点也不断暴露,如:(1)虚假合作,不能区分作者贡献^[16];(2)对论文数量不敏感,只考虑h核内的论文数量和被引频次,忽略h核以外的论文;(3)无法跨领域评价作者影响力^[17];(4)无法对学者近期影响力进行评价^[18]等。至今,对h指数的修正研究扩展指数已不下30余种^[19]。2006年,L. Egghe从论文被引频次的累计贡献角度,提出g指数^[14],同样侧重论文被引频次角度的还有R指数及Hr指数^[20]。2010年,Prathap G.提出p指数,优化了h指数的灵敏度和区分度^[21]。从作者合作角度进行修正的改进则包括均等共享和差异分享论文影响力两种。差异分享论文影响力研究包括考虑主要贡献作者和计算合作者权重。其中,考虑主要贡献作者是在计算h指数时,只纳入该作者作为主要贡献作者的论文,如 h_{maj} 指数^[22];计算合作者权重则基于作者排序和合著者数量的荣誉分配法^[23]。均等共享论文影响力研究包括直接对h指数进行平均和对论文数或被引次数进行平均,前者如 h_1 指数^[24],后者如 h_m 指数^[25]、 p_r 指数^[22]等。2016年,学者Shaon Sahoo针对合著论文个人贡献率问题提出I指数,使得各合著者随着论文作者数量的增加而减小,在一定程度上降低了没有实质贡献的“论文挂名”的影响^[26]。2019年,学者D. Gnana Bharathi提出多作者聚合分析算法指标,即 ammaa 算法^[13],该算法实现了通过对论文影响力的评价进而对学者影响力进行评价的目的。

综合以上研究,虽然都在一定程度上对h指数的缺点有所弥补或改进,但依然没有一种综合有效的方法应用于学者影响力评价中。为此,笔者期待在D. Gnana Bharathi提出的 ammaa 算法基础上考虑到时间异质性,提出一种融入时间因素的改进算法,即 t_ammaa 算法,并尝试通过数据实证,与h指数进行对比和相关性分析,证明 t_ammaa 算法对学者影响力评价结果更全面、公平和精细。

二、ammaa 算法介绍

(一) ammaa 算法原理

ammaa 算法引入一种阈值变量T,也就是合著者数量的乘数。对任一论文而言,从以下公式(1)(2)(3)不难看出,若论文为单作者独著成果,则该文 ammaa 值就是其被引频次;当作者人数大于1时,其 ammaa 值是总引文量的平方除以阈值T和作者数量,再加上按著者数平分总引文量的值,使

得每篇论文的影响力都随着引文数的增加而增加,直至该值达到由合著者数量和阈值设定的引文限制(即 $aT-T$),其 ammaa 值才等于文章的总引文量,如以下公式(2)所示,这时,每个著者才可以算是获得了该文章所有的引文影响力。这种所有作者平均共享被引次数,每增加 1 次引用,其影响力的值就会同等份额逐渐增加的度量方法,原理类似于为工人生产率的每一次增长都提供同等的绩效激励,直至达到目标要求,团队的所有成员都可以得到全额奖励。这样既不会由所有作者平分引文总数那样弱化实际作者贡献值,降低合著者的积极性,也不会夸大没有实际贡献合著者的影响力。

一篇论文的 ammaa 值计算公式如下:

对于单作者的文章,ammaa 值等于总被引次数,即:

$$\text{当 } a=1 \text{ 时, } AMMAA=c \quad (1)$$

对于其他任何形式多作者合著文章,即:

$$\text{当 } a>1, \text{ 且 } c<aT-T \text{ 时, } AMMAA=(c+c^2/T)/a \quad (2)$$

$$\text{当 } a>1, \text{ 且 } c\geq aT-T \text{ 时, } AMMAA=c \quad (3)$$

式中, c 为被引次数, a 为作者人数, T 为阈值变量。

在衡量同一学科的多作者论文影响力时, T 值是相同的,所有合著者都享有的全部引文量阈值($aT-T$)就会随着合著者人数的增多而变大,如图 1 所示。在达到阈值前,引文量相同的情况下,ammaa 值会随着合著者人数 a 的增多而变小,如图 2 所示。由此,ammaa 算法同时考虑了作者人数及论文被引频次两个因素,实现了既可以对单作者论文评价,也可以对多作者论文进行评价的创新融合。

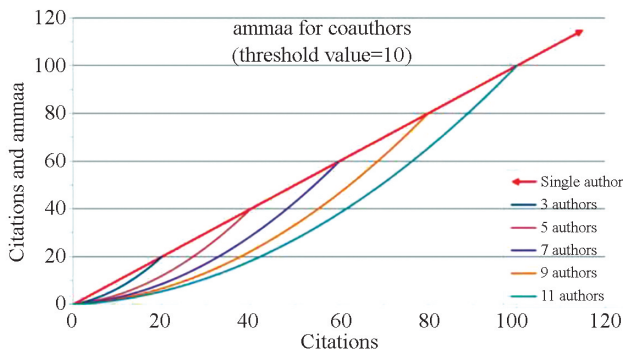


图 1 当 T 值固定时(此处等于 10),不同数量合著者的 ammaa 值,以及与单个作者 ammaa 值合并时的引文数

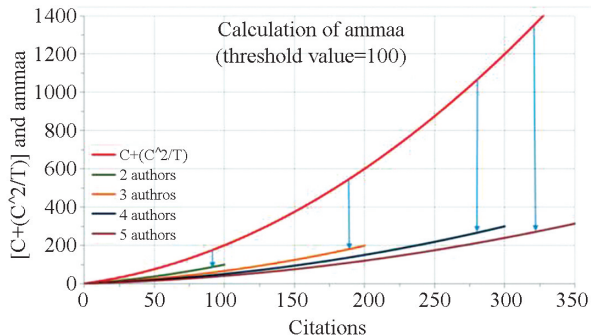


图 2 当 T 值固定时(此处等于 100),合著者人数越多,每个作者的 ammaa 值与 $(c+c^2/T)$ 的距离越大

T 的赋值可通过普遍共识或由特定国家、团体或组织来确定。传统上被引用次数较少的学科,如地质学、数学,其 T 值较低。被引用次数较多的学科,如生物医学、生物化学,其 T 值较高。 T 的其他任何值也可以在科学合理的基础上设置,通用情况下,作者建议将 T 值暂设为 100,一旦设置了 T ,就可以为每篇合著文章计算其 ammaa 值,也就是衡量每篇论文为合著者中的每位学者带来的影响力。 T 值的设定是根据学科性质确定的,所以,当出现普遍低被引学科的作者与普遍高被引学科的作者跨学科合作时, T 值的设定还可以平衡由学科差异带来的影响。

(二) ammaa 算法的不足

在论文影响力评价过程中,统计评价指标时常常忽略时间轴不统一的问题^[27],从而降低评价的有效性。ammaa 算法综合考虑了论文被引次数、作者人数两方面因素,却忽视了时间因素对评价效

果的影响。论文被引时间的异质性体现论文成果传播的时间差异性,被引时刻才是知识传播和交流、发挥效用的时间,发文与被引的时间差反映了论文的知识流动性和影响力的大小^[28]。两篇不同论文的发表时间不同,总被引频次相同时,发表时间更短的论文有更大的影响力;发表时间相同,总被引频次相同,引文量逐年上升和引文量逐年下降的论文学术价值也是不同的^[29]。因此,笔者在分析 ammaa 算法忽略时间维度的情况下,力求基于发文时间和被引时间的异质性对 ammaa 算法进行优化。

三、t_ammaa 算法的提出

加权是为平衡某一要素在整个要素体系中,因重要程度不同而分别赋予该要素不同权值的过程^[30]。本优化方法对论文发表后每年的 ammaa 值赋予不同的权重,计算论文的加权 ammaa。这样,一篇论文的 t_ammaa 值就是对它每年的 ammaa 值赋予不同权重后之和,作者的 t_ammaa 值就等于其所有论文 t_ammaa 值之和。

对于每篇论文而言:

$$t_ammaa = (1 * ammaa_1 + 2 * ammaa_2 + 3 * ammaa_3 + \dots + n * ammaa_n) / n \quad (4)$$

式中, $n = (\text{数据采集时间} - \text{论文发表时间}) / \text{年}$; $ammaa_1, ammaa_2, \dots, ammaa_n$ 表示论文发表后第一年、第二年……第 n 年的 ammaa 值,每年的 ammaa 值利用作者所发文章在这一年的被引频次 c 计算。

对于每个作者而言:

$$t_ammaa = t_ammaa_1 + t_ammaa_2 + t_ammaa_3 + \dots + t_ammaa_n \quad (5)$$

式中 n 为某作者的第 n 篇论文; $t_ammaa_1, t_ammaa_2, t_ammaa_3, \dots, t_ammaa_n$ 为该作者的第一篇、第二篇……第 n 篇论文的 t_ammaa 值。

四、t_ammaa 算法实证研究

(一) 数据来源及处理

本文以国内图情领域学者为研究对象,选取 Web of Science 核心合集(以下简称为 WOS)作为实验数据来源,检索策略为:SU=(Information science and library science);检索时间跨度为 2010 年至 2019 年,限定语言类型为英文;精炼限定:国家/地区为(“PEOPLES R China”) OR (“China”),文献类型为 ARTICLE,结果显示共有 21 279 条文献记录。

为保证所得数据的可靠性,剔除标注为“被撤回的出版物”以及无作者记录 12 条,考虑数据的可操作性,删除年均被引频次小于 1 的论文记录 6 528 条,剩余 14 739 条可用记录。通过 Python 程序分词统计得到 60 342 位作者,及每位作者的全部发文和被引详情。并采取前文所述中的作者建议,暂把 T 值设为 100。根据公式(1)(2)(3)和公式(4)(5),分别计算每位作者的 ammaa 值和 t_ammaa 值,并排序。

(二) 实验过程

本文尝试从两个角度进行数据梳理,然后再综合进行分析讨论。

第一个角度,从发文量和引用量出发,比较分析 ammaa 算法及 t_ammaa 算法和 h 指数的区别和相关性。为保证实验可操作性,设定(1)年均被引频次 40 次以上,在 2010—2019 年所发论文总被引

频次不少于 400 次,和(2)在(1)的基础上,年均发文 3 篇以上,在 2010—2019 年期间总发文量不少于 30 篇的作者两个筛选条件,最终筛选出 52 位符合以上条件的作者,并对 52 名作者进行核查,排除同名同姓的情况。因篇幅有限,摘录发文总数前 25 位作者发文及引用情况记录(表 1)。

h 指数为现行国际上使用的评价指标,故本文先通过 ammaa 算法与作者 h 值排名比较(即排名 1—排名 2),证明 ammaa 算法的调节力和区分度,再通过作者 ammaa 值排名减去 t_ammma 值排名比较(即排名 2—排名 3),对比分析 ammaa 算法和 t_ammma 算法。各位学者的 h 值是在 WOS 核心库中通过设定与数据源同一限定条件加上作者姓名检索获得。两排名比较中均采用正数表示作者排名上升,负数相反,数值绝对值表示作者排名变化量。因篇幅有限,现摘录部分学者结果,如表 2 所示。

第二个角度,从 ammaa 值出发,分析讨论 ammaa 值排名靠前的学者发文及被引频次情况。参考前文数据,故保留前 52 位作者并进行核查,排除同名同姓情况。现摘录部分数据如表 3 所示。

(三) 结果与分析

综合全部导出数据及表 1 可见,被调查的 14 739 篇论文中独著论文 157 篇,仅占 1.07%。其中,发文数大于 30 篇,总被引频次大于 400 次的 52 位学者中,仅有 7 位学者在 2010—2019 年各发表了 1 篇独著论文,表明论文合著的形式已被普遍接受,并可通过科学合作以提高科研成果影响力^[2]。该结果与国家科学技术评估中心—科睿唯安分析结果及国内外学者的研究结论吻合^[31-33]。

如表 3 所示,经计算 ammaa 值排名前 52 的学者中,发文数小于 10 篇的有 21 位,发文数介于 10 到 30 篇之间的学者有 26 位,占比 50%,大于 30 篇的有 5 位,占比不到 10%。学者 Law Rob 独著数 1 篇,学者 Zhou Tao 独著数 10 篇。其中,发文数与总被引频次均相同的学者有三组,分别为 Csordas Attila 和 Ternent Tobias;del-Toro Noemi 和 Dianas Jose A;Tang Zefang 和 Li Chenwei。经调查详细数据发现,Csordas Attila 和 Ternent Tobias、del-Toro Noemi 和 Dianas Jose A 参与合著 1 篇论文被引频次高达 2 023 次,Tang Zefang 和 Li Chenwei 参与合著 1 篇论文被引频次为 1 720 次,且三组学者第一组仅有 2 篇论文发表,后两组仅有 1 篇。假若综合考虑发文数,排除发文数少于 10 篇的学者记录,且以年均发文至少 1 篇为限制条件,则 ammaa 算法既可以突出发文能力较强如独著较多的学者,也能筛选出文章质量较高的学者。

本研究中,我们保留 ammaa 值排名前 52 且发文至少 10 篇的学者各指标值,用与前文相同的条件查询各位学者的 H 值并排名(表 4),综合第一角度一并做以下分析与讨论。

1. ammaa 算法与 t_ammma 算法调节能力结果与讨论

由表 2 和表 4 可知,两种算法的计算结果中,均未出现 ammaa 值、t_ammma 值相同的情况,而表 2 中 H 值查询结果中相同的学者有 Law,Rob 等 12 组,共 48 位学者,占比 92.3%;表 4 中 H 值相同的有 Zhang,Lei 等 6 组,共 19 位学者,占比 86.4%。分别对 H 值、ammma 值和 t_ammma 值三者排名进行归一化处理,并按照 H 值排序,如图 3、图 4 所示,ammma 算法和 t_ammma 算法排名较 H 指数排名总体趋势一致,二者均围绕 H 值排名上下浮动,且幅度较大;加入时间权重因素后,与 ammaa 值相比,表 2 中 52 位作者有 40 位学者的 t_ammma 值名次发生变化,名次变化学者占比 76.9%,表 4 中 22 位学者有 21 位学者名次发生变化,名次变化学者占比 95.5%。说明 ammaa 算法和 t_ammma 算法识别效果都较 H 指数好,且 t_ammma 算法较 ammaa 对学者影响力评价更敏感,调节作用更明显。

表1 发文数前25位作者总被引及每年被引频次详情

序号	作者姓名	发文总数	独著数量	总被引频次	2010年	2011年	2012年	2013年	2014年	2015年	2016年	2017年	2018年	2019年
1	Wang Wei	77	0	1 910	0	13	44	77	106	161	216	237	275	423
2	Liu Yang	59	0	2 469	2	11	18	30	49	74	107	188	453	745
3	Zhang Wei	59	0	1 670	3	8	13	36	50	92	104	199	320	432
4	Li Jing	59	0	1 245	1	24	44	58	100	128	134	158	166	231
5	Shu Hua	54	0	1 570	2	19	60	97	134	163	198	203	211	255
6	Wang Jian	53	0	1 568	3	19	27	47	79	147	151	204	220	343
7	Wang Yan	53	0	1 002	0	4	10	21	34	61	78	98	187	268
8	Liu Li	52	0	1 162	0	8	12	57	79	82	141	179	157	234
9	Li Xia	50	0	2 015	1	13	34	58	81	129	185	290	351	480
10	Wang Jun	50	0	1 655	4	20	57	115	146	170	162	181	211	266
11	Law Rob	49	1	2 716	3	23	41	85	135	227	295	368	399	625
12	Zhang Li	49	0	1 157	1	3	32	57	91	82	134	140	172	217
13	Wang Jing	48	0	785	0	5	25	39	52	59	66	97	117	161
14	Chen Wei	47	0	1 534	0	0	4	34	69	80	109	148	276	422
15	Wang Li	45	0	935	0	2	43	45	78	85	112	126	140	143
16	Li Wei	44	0	816	1	2	13	27	39	63	74	107	151	178
17	Li Yan	42	1	820	1	8	35	31	46	54	58	81	114	177
18	Wang Lei	42	0	709	0	4	16	26	35	45	54	82	92	161
19	Tao Dacheng	41	0	3 559	2	24	130	210	286	385	471	512	554	568
20	Li Xuelong	41	0	2 998	3	26	58	91	185	348	497	488	523	490
21	Liu Xiaoping	41	0	2 255	1	6	23	50	88	149	238	342	383	497
22	Zhang Yu	41	1	1 172	0	0	0	18	43	74	118	145	228	260
23	Yang Yang	41	0	1 014	0	9	9	25	38	49	87	137	170	254
24	Zhang Jing	41	0	777	3	9	11	23	37	73	65	117	134	159
25	Rousseau R	41	0	754	0	20	36	52	46	74	76	130	112	121

表2 部分作者各指标排名及排名变化表

作者姓名	h值	排名1	ammaa值	排名2	t_ammaa值	排名3	排名1-排名2	排名2-排名3
Li Xuelong	30	1	1 657.51	4	526.75	4	-3	0
Tao Dacheng	29	2	2 223.68	1	641.44	3	1	-2
Liu Xiaoping	26	3	920.04	6	309.27	6	-3	0
Law Rob	25	4	2 003.92	2	660.25	2	2	0
Wang Jian	24	5	499.08	19	160.01	21	-14	-2
Chen Wei	23	6	560.61	15	210.09	12	-9	3
Wang Jun	22	7	786.83	9	288.64	8	-2	1
Zhang Yan	21	8	152.59	44	62.38	45	-36	-1
Liu Yang	20	9	990.74	5	250.48	10	4	-5
Wang Yi	19	10	752.44	10	204.93	14	0	-4
Li Yan	19	10	268.19	28	102.03	28	-18	0
Wang Fang	19	10	135.18	46	63.04	44	-36	2
Zhao Jing	18	11	121.84	48	58.66	47	-37	1
Zhang Lei	18	11	1 684.34	3	790.55	1	8	2
Zhang Jian	17	12	179.38	40	71.02	40	-28	0
Zhang Lin	17	12	879.82	7	295.51	7	5	0
Wang Yu	16	13	81.77	51	36.49	51	-38	0
Wang Ying	16	13	117.46	49	50.51	49	-36	0
Li Li	15	14	102.50	50	45.44	50	-36	0
Zhang Jie	15	14	588.94	14	190.16	16	0	-2
Liu Xin	15	14	212.77	32	64.73	43	-18	-11
Lin Hongfei	15	14	212.28	33	91.01	33	-19	0
Wang Lei	14	15	177.33	41	69.72	41	-26	0
Wang Tao	12	16	69.79	52	30.36	52	-36	0

表 3 ammaa 值排名前 30 学者各指标情况

序号	作者姓名	ammaa 值	ammaa 值 排名	t_ammaa 值	t_ammaa 值排名	发文数	独著数量	合计引用 次数
1	Xu Xin	3 000.49	1	1 518.59	3	22	0	3 325
2	Thong James Y. L.	2 954.82	2	1 535.74	1	11	0	3 290
3	Venkatesh Viswanath	2 868.69	3	1 523.98	2	13	0	3 174
4	Ma Yi	2 339.76	4	718.14	12	4	0	2 485
5	Tao Dacheng	2 223.68	5	641.44	24	41	0	3 559
6	Xu Wei	2 222.12	6	1 313.58	4	20	0	2 691
7	Yang Ming	2 147.19	7	1 286.03	5	10	0	2 317
8	Hermjakob Henning	2 133.85	8	659.31	18	9	0	2 628
9	Perez-Riverol Yasset	2 121.47	9	653.99	20	4	0	2 412
10	Vizcaino Juan Antonio	2 119.24	10	653.14	21	3	0	2 390
11	Csordas Attila	2 118.01	11	652.85	23	2	0	2 378
12	Ternent Tobias	2 118.01	11	652.85	23	2	0	2 378
13	Wang Rui	2 081.56	12	652.97	22	18	0	2 353
14	Ji Shuiwang	2 076.06	13	1 255.60	6	4	0	2 099
15	Yu Kai	2 075.63	14	1 255.10	7	3	0	2 098
16	del-Toro Noemi	2 023.00	15	626.07	25	1	0	2 023
17	Dianes Jose A.	2 023.00	15	626.07	25	1	0	2 023
18	Law Rob	2 003.92	16	660.25	15	49	1	2 716
19	Gao Ge	1 832.83	17	681.56	13	6	0	1 993
20	Yan Shuicheng	1 775.87	18	620.60	26	17	0	2 218
21	Li Cheng	1 728.08	19	663.30	14	4	0	1 756
22	Zhang Zemin	1 720.41	20	660.07	16	2	0	1 722
23	Tang Zefang	1 720.00	21	659.87	17	1	0	1 720
24	Li Chenwei	1 720.00	21	659.87	17	1	0	1 720
25	Zhang Lei	1 684.34	22	790.55	10	31	0	2 273
26	Li Xuelong	1 657.51	23	526.75	32	41	0	2 998
27	Liu Guangcan	1 601.61	24	556.15	27	6	0	1 709
28	Zhang David	1 548.42	25	767.95	11	9	0	1 872
29	Lin Zhouchen	1 543.41	26	536.63	28	5	0	1 591
30	Yu Yong	1 533.89	27	532.75	29	4	0	1 568
31	Sun Ju	1 521.52	28	527.18	31	2	0	1 524
32	Tan Xiaoyang	1 520.81	29	942.45	8	2	0	1 538
33	Zhang Huaguang	1 517.36	30	387.48	43	15	0	1 995

表 4 ammaa 值排名前 52 位发文数大于等于 10 的学者各指标详情

序号	作者姓名	ammaa 值	ammaa 值 排名	t_ammaa 值	t_ammaa 值 排名	H 值	H 值 排名	发文数	独著数	合计引用 次数
1	Xu Xin	3 000.49	1	1 518.59	3	15	6	22	0	3 325
2	Thong James Y. L.	2 954.82	2	1 535.74	1	10	9	11	0	3 290
3	Venkatesh Viswanath	2 868.69	3	1 523.98	2	12	8	13	0	3 174
4	Tao Dacheng	2 223.68	4	641.44	9	26	2	41	0	2 485
5	Xu Wei	2 222.12	5	1 313.58	4	14	7	20	0	3 559
6	Yang Ming	2 147.19	6	1 286.03	5	8	11	10	0	2 691
7	Wang Rui	2 081.56	7	652.97	8	14	7	18	0	2 317
8	Law Rob	2 003.92	8	660.25	7	25	3	49	1	2 628
9	Yan Shuicheng	1 775.87	9	620.60	10	12	8	17	0	2 412
10	Zhang Lei	1 684.34	10	790.55	6	18	5	31	0	2 390
11	Li Xuelong	1 657.51	11	526.75	12	30	1	41	0	2 378
12	Zhang Huaguang	1 517.36	12	387.48	16	15	6	15	0	2 378
13	Huang Lei	1 307.41	13	530.79	11	9	10	10	0	2 353
14	Ge Shuzhi Sam	1 303.80	14	424.68	13	10	9	10	0	2 099
15	Xu Dong	1 139.74	15	357.92	17	12	8	18	0	2 098

续表 4

序号	作者姓名	ammaa 值	ammaa 值排名	t_ammaa 值	t_ammaa 值排名	H 值	H 值排名	发文数	独著数	合计引用次数
16	Cao Jinde	1 121. 82	16	399. 73	15	18	5	21	0	2 023
17	Zhou Tao	1 073. 24	17	412. 65	14	14	7	22	10	2 023
18	Chen Jie	1 028. 08	18	216. 27	22	14	7	25	0	2 716
19	Lu Yaobin	1 001. 91	19	277. 42	19	14	7	19	0	1 993
20	Liu Yang	990. 74	20	250. 48	21	20	4	59	0	2 218
21	Xu Ying	969. 01	21	265. 76	20	9	10	11	0	1 756
22	Sun Jian	950. 75	22	337. 78	18	10	9	14	0	1 722

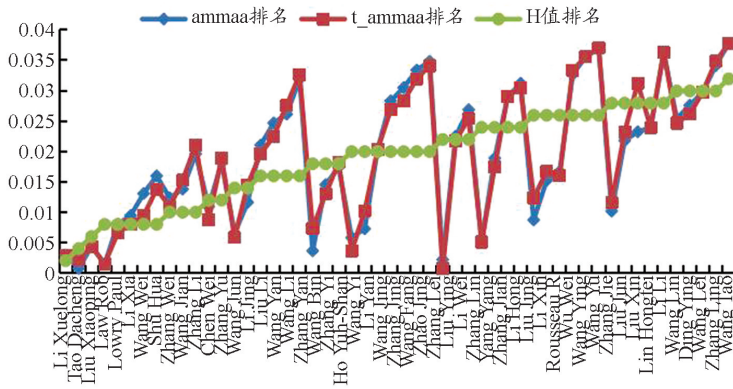


图 3 52 位作者 ammaa 值排名、t_ammaa 值排名和 H 值排名归一化比较

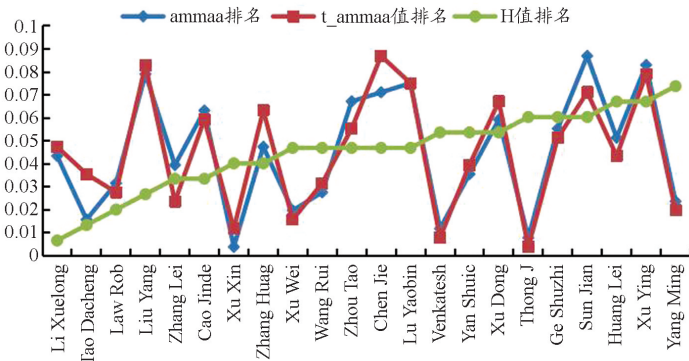


图 4 22 位作者 ammaa 值排名、t_ammaa 值排名和 H 值排名归一化比较

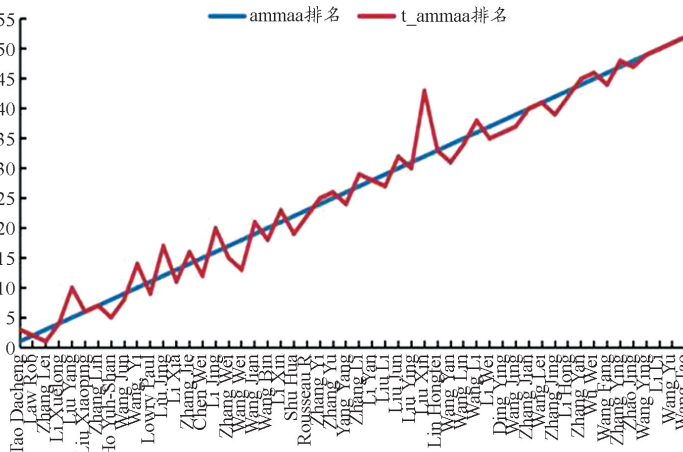


图 5 52 位作者 ammaa 值排名和 t_ammaa 值排名变化

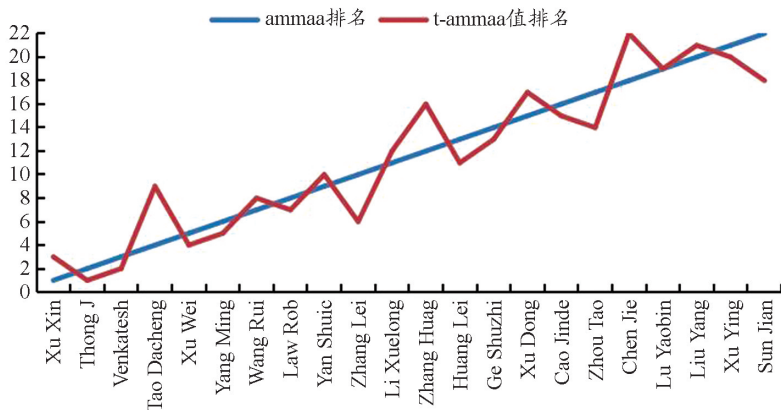


图6 22位作者 ammaa 值排名和 t_ammaa 值排名变化

需要说明的是,图3、图4、图5、图6因版面所限,图面文字不够清晰,有兴趣者可与笔者线上交流。

首先分析比较 ammaa 算法与 h 指数。在表2中,两种评价方法下除 Wang Yi 和 Zhang Jie 两位学者名次未发生变化外,其余学者排名均产生变动,变化最多的是学者 Wang Yu 和 Zhao Jing,变化值均为38。表4中,除学者 Wang Rui 名次未发生变化外,其余21位学者名次均发生变化,其中变化最大的是学者 Liu Yang,变化值为16。

下面就表2中名次变化最大的学者 Wang Yu 分析原因,除各位学者原本 h 值排名并列的情况下,结合每位作者的所有发文和被引频次及合著者人数,对比分析下降名次最多的 Wang Yu 和与其 h 值相同排名却未发生变化的 Zhang Lin,及排名上升最多的 Zhang Lei(上升8个名次)三位作者,均取其被引频次大于10的论文,发现学者 Wang Yu 平均每篇论文的合著者人数为8.62,最高被引频次为41次,且该篇合著者人数为9人。而学者 Zhang Lei 和 Zhang Lin,平均每篇论文的合著者人数分别为6.39和6.04,单篇最高被引频次分别为1219次和389次,且两位学者该篇合著者人数均为3人,故学者 Wang Yu 的 ammaa 值排名明显下降。以上对比分析说明,考虑合著者人数的 ammaa 算法对合著论文的评价调节作用明显,且对高被引频次文章具有敏感的识别力。

2. t_ammaa 算法与 ammaa 比较

观察表2的计算结果,有三组作者的 ammaa 值几乎相同,分别是 Li Xia(588.96)和 Zhang Jie(588.94),Liu Xin(212.77)和 Lin Hongfei(212.28),Wang Jing(179.90)和 Zhang Jian(179.38),进一步观察三组作者的 t_ammaa 值分别为236.17和190.16,64.73和91.01,78.38和71.02,其差值分别为46.01,26.28和7.36。观察表4的计算结果,Tao Dacheng 和 Xu Wei、Ge Shuzhi 和 Huang Lei 两组学者的 ammaa 值分别为2223.68和2222.12,1303.80和1307.4,差值分别为1.56和3.6,进一步观察两组作者的 t_ammaa 值分别为641.44和1313.58,424.68和530.79,差值分别为672.13和106.11。可见,相比 ammaa 算法,t_ammaa 算法的区分效果更加显著。

由表2和图5可知,加入时间权重后,t_ammaa 值名次下降最多的是作者 Liu Xin,下降11个名次,居第43位;而总被引频次和 ammaa 值都比其小的学者 Lin Hongfei 却没有变化,仍居第33位。观察表4和图6可知,考虑时间因素后,t_ammaa 值名次下降最多的是学者 Tao Dacheng,下降5个名次,上升最多的是学者 Zhang Lei 和 Sun Jian,均上升4个名次。这是在作者数量较少的情况下,如果作者数量较多,排名变化可能更大。

下面就学者 Liu Xin 和 Lin Hongfei 名次变化分析其原因。由表 1 及两位学者发文、每年被引次数可知:(1)在 2010—2019 年,两位学者均发文 30 篇,总被引次数分别为 829 和 722 次。(2)学者 Liu Xin 年被引频次在 2015 年较前一年下降,其余年份均较前一年有所增加,而学者 Lin Hongfei 年被引频次均保持逐年增加状态(图 7)。(3)两位学者的 ammaa 值总体呈上升状态,学者 Liu Xin 仅在 2013 和 2016 年 ammaa 值略高于 Lin Hongfei,其余年份与后者持平或远低于后者。以 2015 年为界,学者 Liu Xin 的 ammaa 值从 2015 年的 8 上升至 2019 年的 22,而 Lin Hongfei 则从 12 上升到 42,后者增幅较大(图 8)。说明学者 Lin Hongfei 影响力在 2015—2019 年间已逐渐超过学者 Liu Xin。故引入时间权值,越靠近统计时间的因素占比越重,学者 Liu Xin 的 t_ammaa 值排名较 Lin Hongfei 下降越明显。

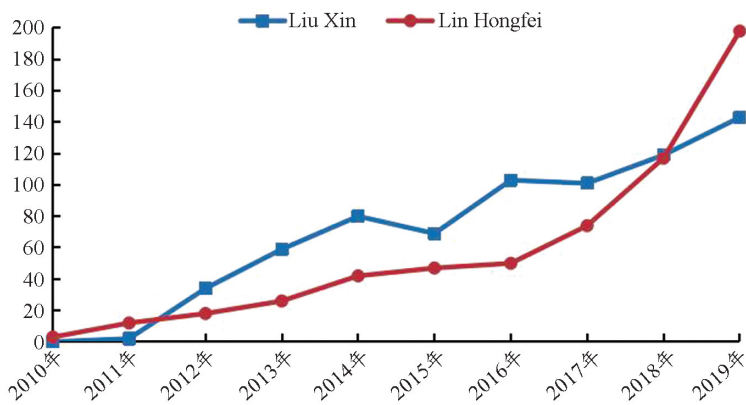


图 7 Liu Xin 和 Lin Hongfei 每年被引频次走势图

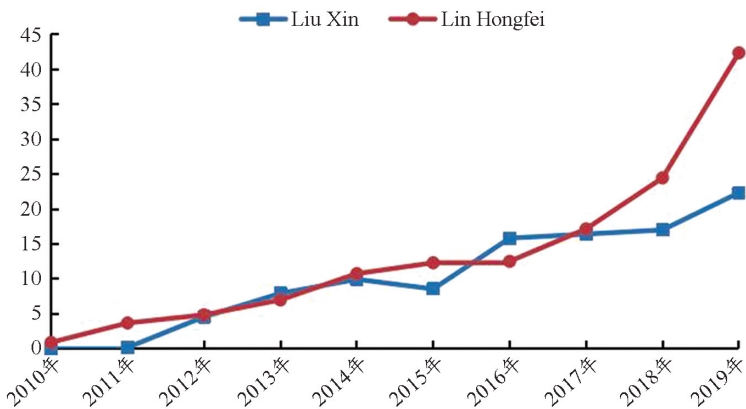


图 8 Liu Xin 和 Lin Hongfei 每年 ammaa 值曲线图

另外,结合表 2 和图 5 可以看出,分别由 ammaa 算法和 t_ammaa 算法得出的排名中,前三名均是 Tao Dacheng、Law Rob 和 Zhang Lei 三位学者,后三名均是 Li Li、Wang Yu 和 Wang Tao 三位学者。结合表 4 和图 6 可以看出,ammaa 算法和 t_ammaa 算法前三名均是 Xu Xin、Thong J 和 Venkatesh Viswanath,说明两种算法在描述作者学术影响力上的趋势是同步的,均能较准确识别高影响力学者。

3. 各指标相关性分析

如果某一算法改进后与原算法的计算结果相差太大,则认为对原算法的改进是不合理的^[34]。引入时间因素的 t_ammaa 算法对学者影响力有一定的调节作用,但这种调节作用对 ammaa 算法的

改善并不是颠覆性的。由此,我们将表 2 中 52 位作者的 t_ammaa 值、 $ammaa$ 值、 h 值、总被引次数和年均被引次数,进行斯皮尔曼相关性分析,证明 t_ammaa 算法更具合理性(表 5)。由分析数据可见,作者的 t_ammaa 值与其他指标均呈极显著相关关系,其中与 $ammaa$ 值、 h 值、总被引次数和年均被引次数的相关系数分别为 0.956、0.552、0.907 及 0.762。 h 值、总被引次数和年均被引频次均可用于对作者影响力进行评价^[14,35-36],说明与 h 指数呈极显著正相关的 $ammaa$ 算法和 t_ammaa 算法也适用于对作者影响力进行评价,且后者对前者的优化是合理的。

表 5 各指标相关性检验

	ammaa 值	t_ammaa 值	H 值	总被引频次	年均被引频次	
t_ammaa 值	皮尔逊相关性	1				
	Sig. (双尾)					
	个案数	52				
ammaa 值	皮尔逊相关性	0.956**	1			
	Sig. (双尾)	0.000				
	个案数	52	52			
H 值	皮尔逊相关性	0.552**	0.662**	1		
	Sig. (双尾)	0.000	0.000			
	个案数	52	52	52		
总被引频次	皮尔逊相关性	0.907**	0.966**	0.780**	1	
	Sig. (双尾)	0.000	0.000	0.000		
	个案数	52	52	52	52	
年均被引频次	皮尔逊相关性	0.762**	0.857**	0.734**	0.922**	1
	Sig. (双尾)	0.000	0.000	0.000	0.000	
	个案数	52	52	52	52	25

注:**在 0.01 级别(双尾)呈极显著相关性; **在 0.05 级别(双尾)呈显著相关性。

五、结论和建议

相对于 h 指数,笔者在 $ammaa$ 算法的基础上,提出一种基于时间维度的 $ammaa$ 优化方法—— t_ammaa 算法,通过对国内图情领域学者在 WOS 核心数据库中发文情况分析发现:(1)针对 h 指数无法识别高影响力论文,对合著论文评价不灵敏,以及无法涉及作者全部论文的缺陷, $ammaa$ 算法和 t_ammaa 算法对 h 值排名相同的情况有明显调节作用,通过设置阈值变量 T 和被引次数的平均共享,使合著论文的所有作者均等增加影响力的激励方法,不仅能更好地挖掘出高影响力和引用价值的科研文献,还能有效避免客座作者^[5]、礼物作者^[6]、非学术合作者^[7]较多出现在论文署名中,从而影响评价结果的公平性,使得作者影响力评价方法更为合理和客观。(2)与 $ammaa$ 算法相比, t_ammaa 算法考虑时间因素,能有效识别领域内长时间保持高活跃度,影响力持续发挥或不断上升的学者。由此, t_ammaa 算法综合考虑了作者发文数量、作者人数、被引频次、被引频次的阈值限制及被引时间异质性,解决了作者影响力评价方法中多作者署名问题,无法识别高影响力论文,被引

次数受论文发表时间长短影响,以及跨领域合作学者影响力评价等需要解决的问题。(3)在图情人文社会科学领域中,合著已成为普遍现象,独著论文占比极少,且存在学者一篇论文被引频次极高、总发文数极少,但 t_{ammaa} 值和 t_{ammaa} 值排名居前的现象。建议在人文社科类评价过程中,可考虑设置学者必须在近些年,比如近3年或5年内必须有若干论文或独著成果产出,以削弱虚假合作者在评价中的影响力。

由以上研究结论可以看出,通过论文影响力来间接揭示作者学术影响力,并非简单通过影响因子来评价,论文引用数据等仅可作为评价参考,尤其在2021年人社部、教育部印发《关于深化高等学校教师职称制度改革的指导意见》中特别强调,不得以SCI或SSCI等论文相关指标作为前置条件和判断的直接依据,必须考虑学术论文的实际内容,特别是真正具有高影响力和引用价值的代表性论文,本研究通过时间要素的引入,可以有效解决仅凭论文“数量”和引用次数评价的局限,在对有持续学术贡献力学者的甄别中,可以设置近1年或3年内发表论文的数量作为学者持续影响力的重要评价手段,解决了部分学者突击或集中于某个阶段发文,或在获得一定职称或荣誉后长期处于“学术休眠”状态的问题;最后,对于多作者的署名问题,本文提出的被引次数平均共享和T值,可以有效解决挂名作者、客座作者或通讯作者等问题,通过优化算法,使得真正有持续贡献力的学者可以被有效识别出来,提高论文或学者贡献度评价的客观性。

本研究也存在一些不足之处,一方面,本文仅选择图书情报领域学者在WOS中的发文数据对 t_{ammaa} 算法进行实证分析,样本涉及领域单一;另一方面,本文阈值T的设置采取原作者对通用领域的取值建议,今后可进一步研究科学设置T值的学科和条件。

参考文献:

- [1] 马瑞敏,张慧. 加权引用视角下的作者学术影响评价研究[J]. 情报学报, 2017(8): 790-797.
- [2] POLYAKOV M, POLYAKOV S, IFTEKHAR M S. Does academic collaboration equally benefit impact of research across topics? The case of agricultural, resource, environmental and ecological economics[J]. *Scientometrics*, 2017, 113(3): 1385-1405.
- [3] GAZNI A, SUGIMOTO C R, DIDEGAH F. Mapping world scientific collaboration: Authors, institutions, and countries[J]. *Journal of the American Society for Information Science and Technology*, 2012, 63(2): 323-335.
- [4] SHABAN S. Multiple authorship trends in prestigious journals from 1950 to 2005[J]. *Saudi Medical Journal*, 2007, 28(6): 927-932.
- [5] WAGER E, SINGHVI S, KLEINERT S. Too much of a good thing? An observational study of prolific authors[J]. *PeerJ*, 2015(3): e1154. doi.org/10.7717/peerj.1154.
- [6] JACK G. Is mass authorship destroying the credibility of papers[ED/OL]. [2020-11-09]. <https://www.timeshighereducation.com/news/mass-authorship-destroying-credibilitypapers#node-comments>.
- [7] SARNA-WOJCICKI D, PERRET M, EITZEL M V, et al. Where are the missing coauthors? Authorship practices in participatory research[J]. *Rural Sociology*, 2017, 82(4): 713-746.
- [8] 高晓培, 潘云涛, 马峥. 科技期刊论文署名规范化探讨与实践研究[J]. 编辑学报, 2012(1): 30-33.
- [9] IRIBARREN-MAESTRO I, LASCURAIN-SÁNCHEZ M L, SANZ-CASADO E. Are multi-authorship and visibility related? Study of ten research areas at Carlos III University of Madrid[J]. *Scientometrics*, 2009, 79(1): 191-200.
- [10] TARNOW E. Coauthorship in physics[J]. *Science and Engineering Ethics*, 2002, 8(1): 175-190.

- [11] ILAKOVAC V, FISTER K, MARUSIC M, et al. Reliability of disclosure forms of authors' contributions[J]. CMAJ, 2007, 176(1): 41-46.
- [12] YAGER K. Each co-author should sign to reduce risk of fraud[J]. Nature, 2007, 450(7170): 610.
- [13] BHARATHI D G. Measuring the impact of an author of multi-authored articles—Aggregating metrics for multiple authors' analysis[J]. ISSI, 2019, 2(5): 448-458.
- [14] HIRSCH J E. An index to quantify an individual's scientific research output[J]. Proceedings of the National Academy of Sciences of the United States of America, 2005, 102(46): 16569-16572.
- [15] BALL P. Index aims for fair ranking of scientists[J]. Nature, 2005, 436(7053): 900.
- [16] 周春雷. h 指数合作式注水缺陷与对策[J]. 图书情报知识, 2009(3): 109-112.
- [17] BORNMANN L, DANIEL H D. Does the h-index for ranking of scientists really work? [J]. Scientometrics, 2005, 65(3): 391-392.
- [18] COSTAS R, BORDONS M. The h-index: Advantages, limitations and its relation with other bibliometric indicators at the micro level[J]. Journal of Informetrics, 2007, 1(3): 193-203.
- [19] WILDGAARD L, SCHNEIDER J W, LARSEN B. A review of the characteristics of 108 author-level bibliometric indicators [J]. Scientometrics, 2014, 101(1): 125-158.
- [20] 金碧辉, Rousseau Ronald. R 指数、AR 指数: h 指数功能扩展的补充指标[J]. 科学观察, 2007(3): 1-8.
- [21] PRATHAP G. The 100 most prolific economists using the p-index[J]. Scientometrics, 2010, 84(1): 167-172.
- [22] HU X J, ROUSSEAU R, CHEN J. In those fields where multiple authorship is the rule, the h-index should be supplemented by role-based h-indices[J]. Journal of Information Science, 2010, 36(1): 73-85.
- [23] SHEN H W, BARABASI A L. Collective credit allocation in science[J]. PNAS, 2014, 111(34): 12325-12330.
- [24] BATISTA P D, CAMPITELI M G, KINOUCI O. Is it possible to compare researchers with different scientific interests? [J]. Scientometrics, 2006, 68(1): 179-189.
- [25] SCHREIBER M. To share the fame in a fair way, hm modifies h for multi-authored manuscripts[J]. New Journal of Physics, 2008, 10(4).
- [26] SAHOO S. Analyzing research performance: Proposition of a new complementary index[J]. Scientometrics, 2016, 108(2): 489-504.
- [27] 俞立平. 期刊影响力指标的时间异质性及其重构研究: 基于多属性评价的视角[J]. 图书情报工作, 2016(12): 109-114.
- [28] 谢瑞霞. 基于“时间异质性”和“合著网络”的作者影响力评价研究[D]. 曲阜: 曲阜师范大学, 2019.
- [29] 舒予, 张黎俐. 时序动态视角下的学术影响力评价方法及实证研究[J]. 现代情报, 2017(11): 74-80, 86.
- [30] 科普中国. 加权[EB/PL]. [2020-10-09]. [https://baike.baidu.com/item/加权/91816? fr=Aladdin](https://baike.baidu.com/item/加权/91816?fr=Aladdin).
- [31] National Center for Science and Technology Evaluation, Clarivate Analytics. China's international scientific research collaboration—A bibliometric analysis[R]. 2018: 1-69.
- [32] 张雪, 张志强, 陈秀娟. 基于期刊论文的作者合作特征及其对科研产出的影响: 以国际医学信息学领域高产作者为例 [J]. 情报学报, 2019(1): 29-37.
- [33] THELWALL M, SUD P. National, disciplinary and temporal variations in the extent to which articles with more authors have more impact: Evidence from a geometric field normalised citation indicator[J]. Journal of Informetrics, 2016, 10(1): 48-61.
- [34] 张欣, 马瑞敏. 基于改进 Page Rank 算法的核心专利发现研究[J]. 图书情报工作, 2018(10): 106-115.
- [35] 方卿, 王珏. 开放存取期刊学术质量评价(一): 基于总被引频次视角的分析[J]. 出版学, 2011(6): 67-70.

- [36] PASTERKAMP G, ROTMANS J I, KLEIJN D V P, et al. Citation frequency: A biased measure of research impact significantly influenced by the geographical origin of research articles[J]. *Scientometrics*, 2007, 70(1): 153–165.

A time–dimension paper influence evaluation research: Improvement based on ammaa algorithm

XU Enping^a, JIA Na^b, LI Min^b, YU Yisheng^b

(*a. Science and Technology Department; b. School of Economics and Management, South China Normal University, Guangzhou 510000, P. R. China*)

Abstract: Aiming at the deficiency of h index and the lack of a comprehensive and effective evaluation index, this paper introduces an ammaa algorithm for paper evaluation, and proposes an optimization algorithm integrating time dimension: t-ammaa algorithm, which reflects the influence evaluation of individual scholars through the evaluation of paper influence. Using Web of Science as the data source and focusing on the papers published by domestic authors in the field of library and information science, the ammaa value and t-ammaa value of these papers are calculated, and then the ammaa value and t-ammaa value of the scholars are obtained. The result ranking of the two algorithms and the scholars' H-value ranking are normalized for empirical comparison and analysis. The results show that t-ammaa algorithm considers the cited times, the cited threshold limit, co-author number and the temporal heterogeneity of the cited papers. It can not only comprehensively evaluate the influence of single-author and co-authored paper, but also eliminate the influence brought by time factor. It is a more reasonable measurement method for evaluating the influence of scholars and papers.

Key words: ammaa algorithm; temporal heterogeneity; multi-author paper influence; author influence; evaluation index; break the “five-only”

(责任编辑 彭建国)