

doi:10.11835/j.issn.1008-5831.2015.03.013

欢迎按以下格式引用:殷之明,周春雷.基于h指数的专家引荐分析法与实例分析[J].重庆大学学报:社会科学版,2015(3):108-113.

Citation Format: YIN Zhiming, ZHOU Chunlei. Evaluation of state key laboratory in physics on citation recommendation analysis based on H-index[J]. Journal of Chongqing University: Social Science Edition, 2015(3):108-113.

基于h指数的专家引荐 分析法与实例分析

殷之明¹,周春雷²

(1.集美大学 学校办公室,福建 厦门 361021;2.郑州大学 信息管理系,河南 郑州 450001)

摘要:引荐分析法(CRA)通过分析高影响力作者的施引列表来寻找该作者眼中的学界同行,通过对同行h指数的分析,可以揭示引荐者本身的学术地位。文章通过对中国物理类国家重点实验的引荐分析,验证了其引荐水平与其领域研究地位一致。

关键词:引文分析法;引荐分析法;h指数;国家重点实验室评价

中图分类号:C3 **文献标志码:**A **文章编号:**1008-5831(2015)03-0108-06

美国著名情报学家尤金·加菲尔德(E. Garfield)认为引文是学者付给同行的硬币工资,引文测度能把一些有用的、客观的因素引入评价过程中,而只涉及少量的使用调查技术的费用,是一种有效的同行专家评议方式^[1]。叶继元也认为引文索引可以帮助确定知识或科学的结构,反映学科之间的渗透情况,可以从一个重要侧面评价被引用论文、刊载被引论文期刊以及学者、学者群体的影响程度和水平^[2]。由于引文具有如此重要的价值,引文分析法自Garfield创立以来一直是一种经典的情报学分析方法,同时引文分析法也被发展成一种间接评价法,用以评价个人、团队、组织、机构的学术贡献。以此类推,对某机构学术成果的频繁施引,意味对对机构学术贡献的认同。

一、引荐分析法

引荐分析法(Citation Recommendation Analysis, CRA)是一种新型的引文分析法,通过从施引作者入手,不断寻找其领域同行、同行的同行来绘制整个研究领域的全景图。它融合了引文分析法、知识图谱分析等技术,为研究领域内的学术网络关系和学科结构提供了新途径。所谓引荐(Citation Recommendation, CR)是指同一施引者对某人的多次引用行为^[3]。从施引者的引用行为可知,对同一作者多次引用,意味着施引者对被引者学术贡献的认可。为了消除引荐行为的偶然性,我们将引用的下限,即至少引用的次数,定义为引荐阈值(Threshold of Citation Recommendation, TCR)。只有超出引荐阈值的频繁施引才会被采信为一次有效引荐,而一旦引用次数超出引荐阈值则数值大小将是无差别的^[4]。

与传统引文分析法不同的是,引荐分析法对施引者给予足够的关注。实际上,对于同一学术成果的引用,一般作者与领域专家的重要程度也不一样。马凤、武夷山认为:“一篇文献被普通作者和知名作者引用,被粗略引用和详细引用是不一样的,高质量的著作常被知名作者多次引用,引用认同结合同行评议可以更全面客观地评价作者及其著作。”^[5]绝大多数低产作者对特定领域的学术兴趣很不稳定,其学术判断未必准确,而专家往往因在特定领域进行了长期深入研究,对相应问题的看法常具有超越普通作者的见识,其对某作者的频繁施引往往意味着对该作者学术贡献的重视。我们在研究h指数的过程中发现,高影响力作者的

修回日期:2015-01-07

基金项目:ISTIC-THOMSON 科学计量学联合实验室开放基金项目“国家重点实验室评价指标体系研究”(IT2009006);教育部人文社会科学
研究项目“面向服务的高校图书馆绩效评价研究”(11YJC870032)

作者简介:殷之明(1974-),男,湖南宁乡人,集美大学副研究馆员,博士,主要从事教育管理与评价研究。

施引列表(其独著或合著论文的参考文献列表的汇总)蕴藏着很有价值的信息。绝大多数高影响力作者或其所在团队引用最多的是高影响力作者自己,其次是一些研究高度相关的同行^[3]。

因此,我们在引荐分析中,通常通过 h 指数来遴选专家,故也称为基于 h 指数的引荐分析法(H-index Based Experts Citation Recommendation Analysis,简称 HBECRA),实质是借鉴选举过程,将引文视作投票,借助学科专家的学术洞察力和引荐及时发现学界新秀、其他研究领域值得关注的学者以及国际上有影响力的学者,弥补了传统引文分析法对施引者关注不足的缺憾,为研究学科结构开辟了新的途径。如果将专家选择尺度进一步放宽,还可以勾勒更精细的学科结构图^[3]。

同理,就一个机构而言,其研究成果经常引用普通作者和知名作者,其重要程度是不一样的。通常机构施引的作者越知名,次数越多,我们认为对应机构越紧跟国际研究前沿,其研究成果也越具有价值。

二、引荐分析法步骤及分析

基于 h 指数的专家引荐法用于评价科研人员或机构时,通常需要实施如下八个步骤。

第一步,选定学科或主题领域。当前,一方面学科、主题划分的方法种类繁多,且相互间缺乏通用性;另一方面,学科间的交叉、融合情况日益显著,这都给学科或主题领域的选择带来了困难。

第二步,领域成果最大限度的收集。难点:关于学科或领域主题的“全”和“准”。领域成果的收集关键在于当前的各种类别数据库。数据库收录回溯时间的长短,收录成果类型的多寡,都将对最终的评价结果产生一定影响。

第三步,学科或主题作者 h 指数的计算与确定,依据 h 指数高低,遴选出学科或主题“h 指数权威专家”。难点在于甄别出同名作者,对于中文数据而言,通常可以通过机构加作者的方式识别,对于外文数据库,则通常困难重重。

第四步,下载获得对评价对象(人员或机构)的全部研究成果施引记录。

第五步,利用自编程序分析施引情况。

第六步,分析“h 指数权威专家”对被评价对象的“提名”情况。

第七步,选定引荐阈值,获得被评价对象“引荐情况”。引荐阈值的设定需要经历多次测试与检验,阈值太高的结果是引荐次数大幅减少,可能出现大多数被评价对象没有获得引荐的情况,从而无法实现评价比较的目标;阈值太低的结果,比如阈值为 1 时,导致评价的实质成了“h 指数权威专家”引用评价法。

第八步,分析“引荐情况”,得出评价结论。分析的角度可以拓展,引荐次数的高低是一个方面;反过来,引荐专家的平均 h 指数高低是另外一个方面;同时,引荐分析法也可以与引用评价进行对比,并从二者的差异中发现问题。

三、实例分析

(一)评价对象选择与确定

依据《中华人民共和国科学技术进步法》,国家科学技术部在 2002 年 4 月公布的《国家重点实验室建设与管理暂行办法》中明确:“重点实验室是国家科技创新体系的重要组成部分,是国家组织高水平基础研究和应用基础研究、聚集和培养优秀科学家、开展学术交流的重要基地。”经过 20 多年的建设与发展,国家重点实验室作为中国基础性研究的“国家队”,已经成为中国国家科技创新体系的中坚力量。选择国家重点实验室,来实践和检验引荐分析法评价科研机构,具有较好的代表性。故我们的引荐分析法应用评价研究选择全部 11 个物理类国家重点实验(表 1)。

表 1 物理类国家重点实验室名单

序号	实验室名称	依托单位	主任	主管单位
1	半导体超晶格	中国科学院半导体研究所	李树深	中国科学院
2	波谱与原子分子物理	中国科学院武汉物理与数学研究所	刘买利	中国科学院
3	非线性力学	中国科学院力学研究所	何国威	中国科学院
4	固体微结构物理	南京大学	王牡	教育部
5	核物理与核技术	北京大学	叶沿林	教育部
6	精密光谱科学与技术	华东师范大学	曾和平	教育部
7	强场激光物理	中国科学院上海光学精密机械研究所	李儒新	中国科学院
8	人工微结构和介观物理	北京大学	龚旗煌	教育部
9	声场声信息	中国科学院声学研究所	郭良浩	中国科学院
10	湍流与复杂系统研究	北京大学	余振苏	教育部
11	应用表面物理	复旦大学	蒋最敏	教育部

(二)数据来源及其处理方法

1. 基础数据来源及处理

依据 JCR 选取全部物理类期刊,剔除出各大类别之间的重复期刊后,共计 338 种期刊,其中“PHYSICS, APPLIED”类 88 种,“PHYSICS, ATOMIC, MOLECULAR & CHEMICAL”类 30 种,“PHYSICS, CONDENSED MATTER”52 种,“PHYSICS, FLUIDS & PLASMAS”24 种,“PHYSICS, MATHEMATICAL”40 种,“PHYSICS, MULTIDISCIPLINARY”65 种,“PHYSICS, NUCLEAR”13 种,“PHYSICS, PARTICLES & FIELDS”26 种。下载上述 338 种自 1986 年以来发表的 SCI 收录论文数据,共计约 185 万条作为研究的基础数据。检索时间为 2010 年 6-7 月,其中“PHYSICAL REVIEW B”的检索结果为大于 10 万条仅下载其较新的 70 000 条。

2. 实验室相关数据检索及说明

对各实验室被收录及施引情况进行了初检索,修改校正检索及第三次检索,目的是通过对比分析,不断修正。2010 年 7 月 20 日,进行了研究评价对象被 SCI 收录论文数据的第一轮检索;2010 年 8 月 1 日进行了第二轮检索,并对检索表达式进行了调整、改进与优化,并分别选取数据收录起始年为 1997 和 1986 进行了两次检索,数据分析表明各实验室自两个起始年份收录的论文数量无明显差异,基于本次评价基础数据的一致性考虑,同时考虑到实验室建设年份的不一致性,特别是个别实验室 2000 年以后才批准建设,本评价最终评价数据选择 1997 年为论文收录起始年。2010 年 8 月 10 日,我们以 1997 年为起始年,对 11 个物理类国家重点实验室被 SCI 收录的论文和施引论文进行了最终检索,并全部下载。检索式及检索结果如表 2 所示。

表 2 第三轮检索表达式及检索结果

序号	实验室名称	检索表达式	收录数	施引数
1	半导体超晶格	AD = (Beijing same ((State Key Lab *) OR (Natl Lab)) same (superlati *) same (Microstruc *))	678	3 899
2	波谱与原子分子物理	AD = (Wuhan same ((State Key Lab *) OR (Natl Lab)) same (Magnet Reson *) same (Atom) same (Mol Phys *))	779	3 011
3	非线性力学	AD = (Beijing same ((State Key Lab *) OR (Natl Lab)) AND (nonli *) AND (mech *))	461	2 310
4	固体微结构物理	AD = (Nanjing same ((State Key Lab *) OR (Natl Lab)) same (Solid State Microstruct *))	4 328	23 638
5	核物理与核技术	AD = (Beijing same ((State Key Lab *) OR (Natl Lab)) same (Nucl * Phys * and Technol *))	398	674
6	精密光谱科学与技术	AD = (Shanghai same ((State Key Lab *) OR (Natl Lab)) same (Precis * Spect *))	287	379
7	强场激光物理	AD = (Shanghai same ((State Key Lab *) OR (Natl Lab)) same (High Field Laser Phys *))	441	1 181
8	人工微结构和介观物理	AD = (Beijing same ((State Key Lab *) OR (Natl Lab)) same (Artificial Microstruct *) same (Mesoscop *))	229	528
9	声场声信息	AD = (Beijing same ((State Key Lab *) OR (Natl Lab)) same (Acoust *))	56	274
10	湍流与复杂系统研究	AD = (Beijing same ((State Key Lab *) OR (Natl Lab)) same (Turbulen *) same (Complex Syst *))	453	1 318
11	应用表面物理	AD = (Shanghai same ((State Key Lab *) OR (Natl Lab)) same (Surface Phys *))	180	1 041

3. 其他检索结果

在第三轮检索的过程中,我们同时记录了 11 个物理类国家重点试验室的如下数据:收录起始年、施引论

文起始年、被引频次、评价被引频次以及 h-index, 详见表 3。

表 3 物理类国家重点室评价数据检索相关数据

序号	实验室名称	收录数	起始年	施引篇数	起始年	被引频次	平均被引频次	h-index
1	半导体超晶格	678	1996	3 899	1997	4 925	7.26	33
2	波谱与原子分子物理	779	1997	3 011	1998	3 938	5.06	24
3	非线性力学	461	2000	2 310	2000	3 062	6.64	25
4	固体微结构物理	4 328	1996	23 638	1996	32 383	7.48	62
5	核物理与核技术	398	1997	674	1997	905	2.27	13
6	精密光谱科学与技术	287	2007	379	2008	537	1.87	10
7	强场激光物理	441	2005	1 181	2005	1 687	3.83	18
8	人工微结构和介观物理	229	1998	528	1999	625	2.73	11
9	声场声信息	56	1996	274	1999	301	5.38	8
10	湍流与复杂系统研究	453	2001	1 318	2002	2 077	4.58	20
11	应用表面物理	180	1997	1041	1998	1 129	6.27	18

(三) 学科领域 h 指数

依据前述基础数据, 经过统计共有 228 732 位作者(不只包含第一作者), 并计量得出领域内作者的 h 指数值。统计结果表明, 在 228 732 名作者中, h 指数值大于或者等于 25 的作者仅占总数的 0.044%; 处于 20 ~ 24 区间的作者有 239 位, 占总数 0.104%; 处于 15 ~ 19 区间的作者有 892 位, 占总数 0.39%; 处于 10 ~ 14 区间的作者有 4 248 位, 占总数 1.857%; 处于 8 ~ 9 区间的作者有 5 202 位, 占总数 2.274%; 处于 6 ~ 7 区间的作者有 11 391 位, 占总数 4.98%; 为 5 的作者有 10 499 位, 占总数 4.590%; 为 4 的作者有 16 717 位, 占总数 7.309%; 为 3 的作者有 28 155 位, 占总数 12.309%; 为 2 的作者有 53 305 位, 占总数 23.305%; 为 1 的作者有 97 983 位, 占总数 42.837% (详细数据见表 4)。

表 4 物理领域内作者 h 指数值分布情况表

h 指数值	作者数	占总数的百分比	累积百分比
≥25	101	0.044	0.044
20 ~ 24	239	0.104	0.149
15 ~ 19	892	0.390	0.539
10 ~ 14	4 248	1.857	2.396
8 ~ 9	5 202	2.274	4.670
6 ~ 7	11 391	4.980	9.650
5	10 499	4.590	14.240
4	16 717	7.309	21.549
3	28 155	12.309	33.858
2	53 305	23.305	57.163
1	97 983	42.837	100.000

(四) 评价结果分析

1. 物理类国家重点实验室整体引荐水平高

以半导体超晶格国家重点实验室引用的 3 899 论文为基础, 分析其引荐情况。在此我们仅分析作为第一作者被引荐的情况, 设置引荐阈值为 5, 共有 113 位作者获得引荐, 作者 h 指数值分布情况如表 5 第一行

所示。同样,对全部 11 个国家重点实验室的施引情况进行汇总,分别计算国家重点实验室引荐的作者的 h 指数值分布情况。

表 5 物理类国家重点实验室引荐作者 h 指数值分布情况

序号	实验室名称	≥10		5~9		3~4		2		1		引荐合计
		数量	%	数量	%	数量	%	数量	%	数量	%	
1	半导体超晶格	47	42	43	38	17	15	5	4	1	0.9	113
2	波谱与原子分子物理	33	49	23	34	7	10	2	3	3	4.4	68
3	非线性力学	20	38	10	19	12	23	6	11	5	9.4	53
4	固体微结构物理	108	79	22	16	4	2.9	0	0	3	2.2	137
5	核物理与核技术	41	45	29	32	14	15	5	5	2	2.2	91
6	精密光谱科学与技术	16	34	13	28	8	17	3	6	7	15	47
7	强场激光物理	13	45	7	24	6	21	2	7	1	3.4	29
8	人工微结构和介观物理	30	42	18	25	8	11	11	15	4	5.6	71
9	声场声信息	8	42	5	26	2	11	2	11	2	11	19
10	湍流与复杂系统研究	43	43	32	32	9	9.1	9	9	6	6.1	99
11	应用表面物理	54	42	33	26	21	16	16	12	5	3.9	129

整体上看,11 个物理类国家重点实验室引荐的作者整体分布情况良好,其中 9 个实验室 40% 以上的引荐对象属于物理学领域内 h 指数值前 3% 的顶尖作者,对于物理学领域内 h 指数值前 15% 的优秀作者的引荐百分比分别达到:79.65、82.35、56.60、94.89、76.92、61.70、68.97、67.61、68.42、75.76、67.44。特别是南京大学固体微结构物理国家重点实验室对顶尖作者的引荐占其总引荐量的 79%,其对前 15% 的优秀作者的引荐百分比分别达到 94.89。

2. 引荐结果与被引频次、h-index 相关度不高

把 11 个国家重点实验室对于物理学领域内 h 指数值前 15% 的优秀作者的引荐百分比进行排序,对表 3 获得的检索结果中的评价被引频次以及 h 指数值进行排序,生成表 6。

表 6 物理类国家重点实验室三类排序情况比较表

序号	实验室名称	引用 h 指数值前 15% 序	平均被引频次序	h-index 序
1	半导体超晶格	3	2	2
2	波谱与原子分子物理	2	6	4
3	非线性力学	11	3	3
4	固体微结构物理	1	1	1
5	核物理与核技术	4	10	8
6	精密光谱科学与技术	10	11	10
7	强场激光物理	6	8	7
8	人工微结构和介观物理	8	9	9
9	声场声信息	7	5	11
10	湍流与复杂系统研究	5	7	5
11	应用表面物理	9	4	6

通常情况下被引频次、机构 h 指数值是科研机构评价的重要指标,下面分析引荐评价指标与这两个评价指标之间的关系。从表 6 可以看出,11 个实验室中,半导体超晶格、固体微结构物理、精密光谱科学与技术、人工微结构和介观物理共 5 个实验室的三类排序相对稳定,即相互间的排序差不超过 1。非线性力学、核物

理与核技术、声场声信息、应用表面物理共 4 个实验室三类排序差较大,差值达到 4 或者以上。通过对三个评价指标做简单相关分析,得出三者皮尔森相关系数表(表 7)。仅从相关系数的数值分析,引荐分析法用于评价与传统的科研机构评价结果相关度不高。可以初步认为,其用于机构评价的信度和效度有待从其他角度去研究和论证。

表 7 三大评价结果序皮尔森相关系数表

	引荐评价法序	平均被引频次序	h-index 序
引荐评价法序	1	0.291	0.491
平均被引频次序	0.291	1	0.755
h-index 序	0.491	0.755	1

四、小结

从引用到引荐,其本质都是希望通过对已有成果的利用,通过施引者与被引者,引荐者与引荐者的关系,来揭示他们的影响、认同、推荐关系。本应用研究的基本思想是:一流的科学家必然紧紧跟着学科前沿,其应用(引用)的文献也以其本身或者其他一流的科学家为主,并由此引申到一流的研究机构其研究成果也因该是对一流成果和顶尖作者的引用。研究的目的是为了从引荐的角度去验证中国物理学研究的顶尖机构——物理类国家重点实验室是否引荐了一流的作者,结果也证明了我们的假设。

本研究下载了大量的数据,进行了海量的计算,然而受时间和精力限制,本研究在数据的完整性和数据的精准性方面还可以大幅提高。例如,中国物理类国家重点实验室批准成立挂牌的时间本身就差异较大,缺少了对每一个实验室发展阶段的把握,对其研究成果的检索就存在一定的瑕疵。

参考文献:

- [1] 尤金·加菲尔德. 引文索引法的理论及应用[M]. 侯汉清,等,译. 北京:北京图书馆出版社,2004:233.
- [2] 叶继元. 引文法既是定量又是定性的评价法[J]. 图书馆,2005(1):43-45.
- [3] 周春雷. 引荐分析法:一种新的引文分析法[J]. 情报学报,2010(4):671-678.
- [4] 马凤,武夷山. 引用认同——一个值得注意的概念[J]. 图书情报工作,2009,53(16):27-30,115.

Evaluation of state key laboratory in physics on citation recommendation analysis based on H-index

YIN Zhiming¹, ZHOU Chunlei²

(1. Administration Office of Jimei University, Xiamen 361021, P. R. China;

2. Department of Information, Zhengzhou University, Zhengzhou 450001, P. R. China)

Abstract: Citation Recommendation Analysis searches the peers of the high influence authors through analyzing their citation identity, and can reveal Academic status on Analysis on H-index. The paper analyzes physics state key laboratory with CRA, and proves that its status fits its position.

Key words: citation analysis; citation recommendation; H-index; state key laboratory; evaluation

(责任编辑 傅旭东)