

# 科学家评价方法述评\*

刘 强<sup>1,2</sup> 陈云伟<sup>1,2</sup>

(1.中国科学院成都文献情报中心 科学计量与科技评价研究中心 成都 610041;

2.中国科学院大学经济与管理学院 图书情报与档案管理系 北京 100190)

**摘 要** [目的/意义]梳理并评述国内外有关科学家评价的方法,归纳好的做法和经验,为相关研究机构制定科学家评价方法提供参考。[方法/过程]从定性、定量和新型评价范式等角度进行梳理,其中定量研究方法主要包括:基于论文数和引文数、标准化指标、h 指数及其衍生指数等单指标定量评价方法,基于链接的方法,替代计量学以及多指标综合评价方法等。定性方法主要从同行评议角度展开,新型评价范式主要介绍了大数据时代基于人工智能和机器学习的科学家评价新范式。[结果/结论]科学家评价工作还有很大的空间有待探索,包括:评价理论基础、时间维度的考虑、强调多指标综合的方法、指标优化、评价实践等多个方面。

**关键词** 科学家 评价 链接 方法论

中图分类号 G350

文献标识码 A

文章编号 1002-1965(2019)03-0080-07

引用格式 刘 强,陈云伟.科学家评价方法述评[J].情报杂志,2019,38(3):80-86,60.

DOI 10.3969/j.issn.1002-1965.2019.03.013

## Review on the Evaluation Methods for Scientific and Technological Talents

Liu Qiang<sup>1,2</sup> Chen Yunwei<sup>1,2</sup>

(1.Scientometrics & Evaluation Center (SERC), Chengdu Library and Information

Center, Chinese Academy of Sciences, Chengdu 610041;

2.Department of Library, Information and Archives Management, School of Economics and Management,

University of Chinese Academy of Sciences, Beijing 100190)

**Abstract** [Purpose/Significance] This paper sorts out and comments on the methods used for evaluating scientists, summarizes good practices and experiences, and provides reference for relevant institutions to evaluate scientists. [Method/Process] The paper starts from the perspectives of qualitative, quantitative and new evaluation paradigms. The quantitative methods mainly include single-index quantitative evaluation methods, linking-based methods, altimetric methods and multi-index comprehensive evaluation methods. The single-index quantitative methods include the methods based on the number of papers and citations, standardized indicators, h-index and its derivative index. The qualitative methods mainly start from the perspective of peer review. The part of new evaluation paradigms mainly introduces the new paradigm of scientists' evaluation based on artificial intelligence and machine learning in the era of big data. [Result/Conclusion] There is still much work to be done for the evaluation of scientists, which includes the theoretical basis of evaluation, the consideration of time dimension, the emphasis on multi-index synthesis, index optimization, evaluation practice and many other aspects.

**Key words** scientists evaluation link methodology

## 0 引 言

2018年7月,中共中央办公厅、国务院办公厅印

发了《关于深化项目评审、人才评价、机构评估改革的意见》(简称“《意见》”),深入项目评审、人才评价、机构评估改革。可以预见,此次在国家层面系统地引导

收稿日期:2018-08-06

修回日期:2018-10-10

基金项目:本研究由“四川省委组织部西部之光人才专项”资助。

作者简介:刘 强(ORCID:0000-0003-1861-2833),男,1995年生,硕士研究生,研究方向:情报理论与方法;陈云伟(ORCID:0000-0002-6597-7416),男,1978年生,博士,研究员,研究方向:科学计量学、社会网络分析、生物技术情报研究。

通信作者:陈云伟

科技评价工作,必将对我国未来项目、人才和机构评价产生深远的影响,同时也对构建符合该意见、又符合科学规律的评价指标或方法提出了极为紧迫的需求,特别是对于作为决定国家竞争力的关键要素的人才而言。改革的关键举措之一是要科学设立人才评价指标:突出品德、能力、业绩导向,克服唯论文、唯职称、唯学历、唯奖项倾向,推行代表作评价制度,注重标志性成果的质量、贡献、影响。把学科领域活跃度和影响力、重要学术组织或期刊任职、研发成果原创性、成果转化效益、科技服务满意度等作为重要评价指标。在对社会公益性研究、应用技术开发等类型科研人才的评价中,SCI(科学引文索引)和核心期刊论文发表数量、论文引用榜单和影响因子排名等仅作为评价参考。注重个人评价与团队评价相结合,尊重和认可团队所有参与者的实际贡献<sup>[1]</sup>。

科学地评价人才是科研事业健康发展的基本前提之一,具有重要的意义。上述《意见》已经明确提出了人才评价工作的总体思想和原则,关键在于如何落实与有效实施。科学家是整个人才队伍的核心力量、是科技事业的中坚力量,是科技创新的根基,拥有高水平的、领军的科学家对于国家科技竞争力具有决定性的作用。因此,科学地评价科学家是推进上述《意见》有效实施的重要一环,对营造潜心研究、追求卓越、风清气正的科研环境起到重大的影响作用。为此,本文的研究目的是梳理并评述国内外有关科学家的评价方法,尝试对全球范围内现有的有关科学家评价的方法和实践进行梳理,归纳国际上用于开展科学家评价好的做法和经验,为科学地开展科学家评价提供新的视角和方法,为相关研究机构制定科学家评价办法提供参考。

## 1 定性研究方法

国际上有关科学家评价方法的研究重点聚焦于对科学家科研表现的评价,可分为定量方法和定性方法两大方面。

定性研究方法以同行评议为主,同行评议的概念起源于17世纪英国皇家学会杂志《哲学汇刊》<sup>[2]</sup>,是指特定领域的若干专家学者对所在专业领域的学术成果或人才的科研表现、能力和发展潜力的一种定性评价,广泛用于机构评估、项目评审、人才评价等<sup>[2]</sup>。美国在遴选科研项目 and 人才时广泛采用同行评议,不简单以论文数量或承担科研项目数量为指标来衡量科技人才的能力,更看重的是人才的发展潜力,注重鼓励创新<sup>[3]</sup>。专家同行评议制度也是德国科技评价方法的主要形式,例如,马普学会为每一个研究所聘请本领域著名学者组成学术咨询委员会,独立开展评估<sup>[4]</sup>。

不过,传统的同行评议制度难免会存在诸如马太效应、人情关系和利益关联、偏向保守等缺点<sup>[5-6]</sup>,在用于评价原创性或开创性创新研究和成果时的适用性下降。随着网络技术的发展和新技术的应用,传统的同行评议进行了创新,利用网络技术的便捷性形成了开放式同行评议(open peer review),开放式同行评议包括出版前的开放评议(pre-open peer review)、出版后的开放评议(post-open peer review)和公众开放评议(community peer review)<sup>[7]</sup>。开放式的同行评议不同于传统同行评议,特点在于效率高、周期短、不受限、公开程度高,将读者群纳入了评议队伍,可修正传统同行评议的人情关系、耗时长、马太效应等缺点。

2015年,Hicks等在Nature上发文指出,目前在全球范围内可获取并用于学术评价的工具越来越多,量化的评价指标已经出现了不当使用或滥用的现象,针对此现象该文章提出了《莱顿宣言》<sup>[8]</sup>。《莱顿宣言》主要有十条原则,其中第一条和第七条提出,“定量评价是为支撑定性的同行(专家)评议服务的,而不能取而代之”和“对个体学者的评价应以定性评价为基础,阅读、判断学者论著内容比使用一个指标数据更适宜,此外,还要对处于不同学术生涯阶段的学者区别对待,不能一刀切。”《莱顿宣言》的观点与2012年提出的《旧金山宣言》<sup>[9]</sup>表达了相同的愿景,都强调了同行评议的重要性。与我国最新发布的上述《意见》中提及的“推行代表作评价制度,注重标志性成果的质量、贡献、影响”的观点也高度一致,即在评价科学家时,要重视学术实质贡献和质量,而不是科研产出的数量。

## 2 定量研究方法

从上述学者对定性评价的作用的描述可见,学术界强调不能完全依赖定量指标开展科学家评价,但定性评价并非唯一评价手段,定性评价也离不开定量指标所提供的数据支撑,即我们需要采用合理分配权重的定性与定量相结合的评价方法。

在理论与方法方面,学术界对定量方法的研究相对较多,主要包括基于论文、专利的数量和引文量的量化统计指标、社会网络分析方法<sup>[10]</sup>、替代计量学方法(Altmetric)、机器学习方法等。定量研究又包括单指标评价方法和多指标综合评价方法:单指标评价方法有基于论文和引文数的评价指标、h指数、基于作者合作的指数、PageRank和HITS等基于链接结构的方法<sup>[11]</sup>等;多指标综合评价方法通常是将多个指标整合在一起,赋以不同权重值,得到一个相对更加全面的评价模型或算法。

### 2.1 单指标定量评价方法

#### 2.1.1 基于论文数和引文数的评价指标 借鉴

ISI 期刊影响因子<sup>[12]</sup>的概念,产生了多个类影响因子指标,主要包括: Matsas 在测量单个科学家在群体里的影响时,提出了归一化影响因子(NIF)指标<sup>[13]</sup>,该因子是指某科学家全部论文的引文数之和与参考文献数之和的比值; Pan 和 Fortunato<sup>[14]</sup>提出了作者影响因子(AIF)指标,该因子是指测定年的前两年内作者被引数与发文数的比值; Lippi 和 Mattiuzzi<sup>[15]</sup>提出了科学家影响因子(SIF)指标,该因子是指科学家在测定年的后两年内的被引数与测定年的发文数的比值。在这三个相近的指标中,SIF 与 h 指数无相关性,可能是一项作为现有指标有益补充的方法,特别是更加适用于对年轻科学家的评价工作。

虽然我们强调不能唯论文论,但并不代表基于论文和引文的评价指标就一无是处,特别是基于高被引论文的评价研究,仍然存在较大的利用价值。例如,科睿唯安(原汤森路透)发布的引文桂冠奖(Clarivate Citation Laureates)<sup>[16]</sup>,遴选出最可能摘取诺贝尔奖的全球最具影响力的科学家,截至 2017 年,引文桂冠奖已经成功预测了 43 位诺贝尔奖得主,可见其对高被引科学家的评价价值。而高被引科学家(Highly Cited Researchers)则是与引文桂冠奖伴生的一个概念,高被引科学家是指在过去近 10 年间(最短 10 年零 2 个月,最长 11 年)在相应学科领域发表的高被引论文数量最多的科学家。2017 年高被引科学家遴选采用的是 2005-2015 年的数据,其他年的数据依次类推,当前方法总计公布了 4 次,分别为 2014、2015、2016、2017 年。高被引科学家遴选的主要依据是科睿唯安 ESI 数据库中的高被引论文数,ESI 高被引论文是指 ESI 数据库中被引频次位于同学科、同出版年论文中前 1% 的论文。高被引科学家的人数、学科分布等从一定程度上反映了一个国家、一个机构高影响力论文的产出和高水平科学家人数及相应学科分布的客观状况。国内外也有较多学者对其进行研究,包括论文合作关系、人才基本特征、人才聚集特征和人才流动特征四个方面<sup>[17-22]</sup>。

2.1.2 克服领域差异的标准化指标 在开展基于论文数量和引文数量的统计分析时,已有大量研究发现了领域差异性,即不同领域的引用行为不同,领域内平均每篇论文被引的频次存在较大差异。例如:在数学和生物学两个领域,会出现同样优秀的科研人员其文章被引的绝对数差距悬殊的现象,生物学领域高影响力的文章通常会被引数千次,而数学领域的经典之作最多也仅被引数百次。早在 2005 年,Podlubnyzai 对美国国家自然科学基金 NSF 发布的六个年度(1992 年、1994 年、1996 年、1997 年、1999 年和 2001 年)的引用数据分布报告进行了统计,以数学为基准,统计其他学科领域科学家被引次数相对于数学领域的倍数关

系,从中就能够看出数学以外的其他所有领域的被引数都是多于数学领域的(比值都是大于 1 的),而生物医药(biomedical research)和临床医学(clinical medicine)的比值最高,达到了 78<sup>[23]</sup>。Radicchi 等于 2009 年在 PNAS 上的一篇文章对若干不同的学科进行了比较,同样发现了不同学科之间被引数绝对值的巨大差异,如发育生物学的文献里被引 100 次的文献的数目是航空航天科学领域里被引 100 次文献的 50 倍,其他的各个不同学科之间也存在这样巨大的差异<sup>[24]</sup>。因此,当前基于论文和引文数量的科学家表现情况尚不能进行跨领域比较,甚至在同领域的不同研究主题或研究方向间也会存在差异性,通过标准化去除这种领域差异性,对在学科间观察科学家的科研表现的差异具有重要的意义。

标准化的方法主要包括从被引端进行标准化和从施引端进行标准化两类。从被引端进行标准化主要是应对不同学科之间的差异性,例如,Radicchi 提出了 Cf 指数,即某篇文章的被引用次数与某领域文献的平均被引用次数之比<sup>[24]</sup>; Abramo 提出用  $All_{max} = \frac{c}{\max - \min}$  和  $A_m = \frac{c}{\text{mean}}$  的方法进行标准化<sup>[25]</sup>,其中  $c$  是文章  $i$  的引用次数, $\max$  为 WoS 的分类里文章  $i$  所属的学科类别的最大和最小被引次数, $\text{mean}$  为所属学科类别的平均被引次数; 2015 年,Abramo 又采用了之前的平均被引次数作为分母的标准化方法,进一步提出 FSS(Fractional Scientific Strength)对科学家进行评估<sup>[26]</sup>,  $FSS = \frac{1}{t} * \sum_{i=1}^N \frac{c_i}{c_i} f_i$ ,在这个公式里  $t$  是作者论文发表的时间, $N$  是作者的文章数, $c_i$  是文章  $i$  被引次数, $f_i$  是文章  $i$  所属学科类别的平均被引次数, $f$  则是分数计数的作者贡献率,计算方法是作者数的倒数。通过学科类别的平均被引数来计算每一篇文章的相对被引数,达到标准化的目的。

从施引端进行标准化主要是以施引文献的参考文献为基础的一些理论。例如,Zitt<sup>[27]</sup>提出活跃期刊的概念对期刊进行标准化,指在一段时间内某期刊所引用的参考文献里来自 JCR 的数目。Waltman<sup>[28]</sup>借鉴活跃期刊的概念,提出 SNCS 方法来评价科学家。 $SNCS^{(1)} = \sum_{i=1}^c \frac{1}{a_i}$ ,其中  $a_i$  为第  $i$  个来源于 JCR 的期刊的引用文献,假设一篇文献在 2016 年发表,2018 年被 WOS 数据库文献(也就是 JCR 期刊文献)  $i$  引用,则  $a_i$  为文献  $i$  (JCR 期刊文献)所在期刊论文 2018 年的平均参考文献数。 $SNCS^{(2)} = \sum_{i=1}^c \frac{1}{r_i}$ ,其中  $r_i$  则代表来自 JCR 期刊的引用文献的当年期刊论文参考文献总数。

SNCS<sup>(3)</sup> =  $\sum_{i=1}^c \frac{1}{p_i r_i}$   $p_i$  是文献  $i$  的参考文献里活跃期刊文献所占的比例  $p_i$  可以针对不同学科进行标准化计算。Pepe<sup>[29]</sup> 同样也提出了根据引用参考文献的数量进行科学家影响力评价的总影响力指数 (Total Research Impact, TORI), 计算公式是  $TORI = \sum_{i=1}^n \frac{1}{a_i r_i}$ , 其中  $n$  是科研人员的引用数 (排除了自引数)  $a_i$  为科学家被引的文献作者数  $r_i$  为作者的施引文献 (即将作者文章作为参考文献的文章) 的参考文献总数。在同一篇文章中, Pepe 还提出了学术生涯长短对 ROTI 指数的影响, 即 RIQ (Research Impact Quotient) 指数,  $RIQ = \frac{\sqrt{tori}}{y}$   $y$  是科研人员的学术生涯时间。

总体来看, 以上基于论文和引文数量的科学家评价指标原理简单、清晰, 应用方便, 易于理解。研究人员对其局限性也有清醒的认识, 并不断尝试各种优化方案, 不断改进和发展这一类指标, 包括构造标准化方法。

2.1.3  $h$  指数及其衍生指数  $h$  指数<sup>[30]</sup> 为科学计量学研究开辟了一个新的思路, 也有大量研究不断地对  $h$  指数加以补充和修订, 形成了  $h$  指数及其衍生指数这一庞大的计量指标体系。 $h$  指数的优点是兼顾了论文的数量和质量, 计算简单, 也容易在大型数据库中应用 (如 CNKI、WoS 等)。然而  $h$  指数的缺点也很明显, 即该指数只增不减、无法测量团队学术影响力和个人在团体中的影响、不利于年轻学者、不利于发文少的学者。

$h$  指数及其衍生指数总体上可划分为以下四大类<sup>[31]</sup>:

第一类, 无法脱离  $h$  指数的衍生指数:  $m$  熵指数<sup>[30]</sup>、 $hg$  指数<sup>[32]</sup>、 $A$  指数<sup>[33]</sup>、 $R$  指数<sup>[33]</sup>、 $e$  指数、 $EM$  指数及  $EM'$  指数<sup>[34-35]</sup> 等。其中  $m$  熵指数是将学者学术生涯考虑进来;  $hg$  指数是把  $h$  指数和  $g$  指数的优点集合起来并把缺点变小;  $A$  指数和  $R$  指数都是为了体现高被引论文价值而设立的,  $A$  指数指的是纳入  $h$  指数计算的论文的引文数的平均值;  $R$  指数指的是纳入  $h$  指数计算的论文的引文数之和的平方根;  $e$  指数 (excess citation) 是指考虑构成  $h$  核的文章的引用中超出  $h$  指数值的部分, 克服了  $h$  指数无法评价  $h$  核文章的  $h$  指数次以外的引用;  $EM$  指数是  $h$  指数和  $e$  指数的延伸, 引入了多维  $h$  指数的概念, 用所有  $h$  核文章的引用进行多层次的科学家定量评价, 克服了  $e$  指数仅仅考虑了超出的引用 (excess citation) 的不足;  $EM'$  指数是  $EM$  指数的多维度扩展, 克服了  $EM$  指数没有考虑所有引文的问题。 $EM'$  指数是当前相对更平衡和细致的方法。

第二类, 相对来说独立于  $h$  指数的衍生指数, 如  $g$  指数、 $f$  指数、 $t$  指数、 $\pi$  指数和  $p$  指数<sup>[36]</sup> 等。 $p$  指数指的是卓越指数, 计算方式是  $p = (c^* c/N)^{1/3} = (c2/N)^{1/3}$   $c$  是被引次数,  $c/N$  是篇均被引次数。 $p$  指数本来是用于评价期刊, 国内学者赵蓉英将它用于评价科学家, 通过图书情报领域的学者进行实证, 取得了较好的效果<sup>[37]</sup>。

第三类, 从学科角度考虑 (适用于不同学科) 的评价指数, 如  $h_f$  指数、Normalized  $h$  指数、 $n$  指数、DSI (Domain Specific Index) 指数。

第四类, 将合作者考虑进来的指数, 如  $h_{p-c}$  指数、 $H_p$  指数和  $H_m$  指数。

2.2 基于链接的方法 利用社会网络分析等基于链接的方法开展科学家评价是近年来新兴起来的方法, 其应用得益于相关研究发现了科学家的合作结构与其科研表现间存在正相关关系, 科学家合作水平的高低在一定程度上可以反映其科研表现的强弱<sup>[11]</sup>。例如, 早在 2008 年, Whitfield 就指出科研合作越来越多的趋势, 而这种合作行为对决定一个团队的成功起到正面促进作用<sup>[38]</sup>; Abbasi 等指出, 科学家的科研表现与度数中心度、效率和平均合作强度成正比<sup>[39]</sup>; McCarty 等发现, 科学家需要开展更多的合作以获得最高的  $h$  指数, 同时一些合作者的  $h$  指数也非常高<sup>[40]</sup>。基于此理论基础, 陈云伟等构建了基于作者合作的复合合作强度 CCS 指数, 将合作者的来源机构情况, 以及合作论文比例和合作机构在论文中的分布特征均作为影响科学家表现的因素<sup>[41]</sup>; Li 等人提出一种在有向加权网络中测量节点中心度的新方法, 称之为双向  $h$  指数 (Bi-directional  $h$ -index)<sup>[42]</sup>; Christian 等人提出一个基于网络的引文指标<sup>[43]</sup>, 通过科学家在合作网络中的位置来评估科学家的科学表现, 主要有两个维度即科学家个人因为其在网络中相对于其他所有作者的位置所带来的被引用潜力和科学家实际获得的引用数 (并且对合作网络中位置相似的作者的实际引用数进行比较), 称之为  $s$  指数 (social index), 并且认为该方法能够较好地应用于跨学科合作网络的评估和扩展到其他指标。

此外, 许多研究致力于利用诸如 PageRank 等新指标来对科学家进行排名, 以克服传统指数的局限性。例如, Yan 等用 PageRank 方法在作者合作网中对科学家进行排名<sup>[44]</sup>; Liu 等开发了 AuthorRank (PageRank 的改良) 算法以科学家合作网络中链接的权重来测量其影响力<sup>[45]</sup>; Radicchi 等基于作者-作者引用网络, 利用类似于 PageRank 的算法计算作者的排名<sup>[46]</sup>。

2.3 替代计量学方法 替代计量学 (Altmetrics) 是在近年来新兴的一种评价方法<sup>[47]</sup>, 具有多样性、即时

性和非传统性的特点,主要的数据来源是 ImpactStory、Plum Analytics、PLOS 和 Altmetric.com<sup>[48]</sup>。由于目前还未形成特别健全的理论体系,对于 Altmetrics 的应用也相对较少,代表性的工作包括:王贤文等利用 Springer 的实时数据平台数据和 Facebook、Twitter 等社交媒体,追踪论文数字足迹(即被使用情况)并且了解科学家在不同时间段的工作情况<sup>[49-51]</sup>。国外的应用有用牙科研究专家的评价<sup>[52]</sup>,Kousha 等提出用 Altmetrics 指标进行书籍以及非标准的学术产生的评价,特别提到了软件、视频、数据集等经常被当前科研评价忽略的学术产出<sup>[53]</sup>。Chen 等针对社会科学和人文科学进行评价时,指出替代评估框架或指标也许能更好地反映社会科学和人文科学的学术表现,通过 Altmetrics 指标与学者获取的基金数量之间的相关性,反映了 Altmetrics 指标对于学者学术地位的揭示作用<sup>[54]</sup>。

2.4 多指标综合评价方法 现有研究已经表明,单一指标并不是普适性指标,在开展科学家评价工作时需根据实际情况慎重采用<sup>[55]</sup>,例如,学术界对基于引文数量开展科研评价一直存在很大争议,认为基于引文量的评价指标仅依赖于外在特征的关联,并未考察与研究质量有关的内在特征的关联<sup>[56]</sup>。因此,为了克服单一性评价指标的不足,学术界最近针对科技人才科研评价的研究工作更多的是构建综合的评价指标体系,且通常将定量与定性指标综合在一起来构建综合指标体系。例如,Costas<sup>[57]</sup>在 2010 年提出了包含产出指数、影响指数和期刊质量指数三个指数的复合指数来进行科学家评价,其中产出指数涵盖发文量、引用次数和  $h$  指数,影响指数涵盖高被引论文数、篇均引用次数和引用次数/学科领域平均引用次数,期刊质量指数涵盖发文期刊影响因子中位数、发文期刊影响因子排名/学科领域期刊影响因子排名以及发文期刊影响因子平均数/学科领域期刊影响因子平均数;张春霞于 2013 年提出了  $h'$  指数,是与  $e$  指数  $t$  指数都相关的一个综合测度指数<sup>[58]</sup>,把  $h$ 、 $e$ 、 $t$  三个指标合为一个指标;叶鹰在 2014 年将发文和引文的测度进行整合,提出一类新的整体综合测度指标用于学术评价<sup>[59]</sup>;Abramo<sup>[26]</sup>在 2015 年提出了 FSS(Fractional Scientific Strength, FSS)指数来开展科学家评价, FSS 考虑了论文发表时间长度、文章数、文章被引次数、作者贡献率(文章作者数目的倒数)以及文章所属领域的 WoS 文献平均引用次数;Wu 在 2015 年开发了一个包含引文数量、引文动机和同行评议三方面的平均评价强度指标<sup>[60]</sup>;Abramo 等在 2015 年设计了一个多变量随机模型 SIMCA 用于评价科学家表现<sup>[61]</sup>;张永莉等构建了一套创新人才创新力评估体系,从基础修养、知识技能

和创新体现三个模块对人才创新力进行评价<sup>[62]</sup>。陈云伟等在 2016 和 2017 年分别构建了“科学家创新力评价六元模型”<sup>[63]</sup>和“青年科学家竞争力评价模型”<sup>[64]</sup>。

### 3 新型评价范式

随着大数据时代的到来,情报研究人员也尝试利用人工智能、机器学习等方法开展科技评价工作。从目前检索到的文献来看,最早进行这方面尝试的是 Bodea,他于 2007 年提出将人工智能技术和工具应用到研究项目和计划的评估工作,提倡基于数据库和机器学习技术制定项目和方案的评估技术<sup>[65]</sup>。2010 年, Fu 通过基于内容和文献计量指标进行机器学习算法的应用,进行引文预测<sup>[66]</sup>。2013 年, Fu 针对生物医学领域的工具性引用(如果引用工作的假设是由所引用的工作激发的,或者引用的工作在没有引用的工作的情况下不能被执行则该引用是工具性引用,简而言之就是这一引用是必不可少的)的识别建立了计算模型<sup>[67]</sup>。Ebadi 和 Schiffauerova 在 2015 年提出了一个自动评价研究人员表现的计算系统<sup>[68]</sup>,又在 2016 年提出了一种名为 iSEER 的智能自动系统<sup>[69]</sup>,该系统使用文献计量学指标以及网络结构特征,根据研究人员合作模式、研究绩效和效率对研究人员进行分类,同时还可以预测研究人员出版物的数量及其应该取得的研究经费,是第一个专注于研究表现和资金预测的系统,作者也将在未来加入更多的维度来进一步提高模型的准确率,同时进行自引、消极引用和互引的研究。Fortunato、Borner 等于 2018 年在 *Science* 杂志上联合发表了一篇题为 *Science of science* 的物理学综述,里面就提到了科学学的一个重要的未来研究领域就是将人工智能和机器学习以一种机器和人的想法共同工作的方式进行整合<sup>[70]</sup>。

### 4 总结与启示

科学家评价工作是一个永恒的主题,虽然目前定性、定量的方法很多,诸如  $h$  指数、高被引、社会网络、同行评议等方法虽然在得以不断改进,但依然没有哪个指标是绝对客观可靠的。为此,科学家评价指标的设计、优化和改进工作也将伴随着科技的发展而不断发展。综合近年来国内外针对科学家评价的研究来看,科学家评价方法至少在以下几方面还有很大的探索空间:

首先,在评价理论基础方面,还需不断讨论引文量、合作关系、网络链接关系等与科学家科研表现的相关性,特别是需要借助大数据方法深入分析,为开展相关定量评价提供理论依据。同时,在借用评价期刊和

机构的方法时,对这种借鉴的合理性还需深入讨论并证明,在没有充分论证之前,尚不建议盲目借用。事实上,适用于宏观尺度的评价指标未必适用于微观尺度的科技人才评价,由于每一个科技人才都具有不同的学术表现,随着评价对象从宏观到微观,定量指标的适用性也逐步下降<sup>[10]</sup>。

其次,对科研人员的评价还需加入时间维度,构建动态的评价指标体系,可揭示处于不同科研生涯阶段的科技人才的差异。

第三,评价方法的选用应更多地考虑综合性,单一的指标很难全面地反映一个科技人才的水平,需合理设计基于多个指标的综合方法来克服单个指标的不足,包括定量和定性指标的综合。

第四,在具体指标的优化方面,第一,基于高被引论文数量的科学家评价,一定要区分学科特异性,同时重视评判文章的原创性与引领性,被引数量仅作为参考;第二,基于链接的方法还需从全网角度来观察科学家的地位。第三,基于人工智能和机器学习的新型评价范式获奖对当前评价方法产生颠覆性的影响或带来变革,需要高度关注并展开研究。

最后,在评价实践方面,科研管理部门应该加强对科技评价工作的重视,以促进我国科技事业的良性发展:首先,构建层次化、多途径评价机制,强化人才评价制度的激励效用,需要建立针对不同类型人才的不同评价标准,例如,针对科学人才和工程人才的评价指标应有明显区别;针对理论研究与应用研究应有不同标准;针对不同梯度的研究人员(如研究员与博士毕业刚刚踏入工作岗位的年轻科研人员)也应区别对待;杜绝一刀切式的评价机制带来的不利影响。其次,创建学科领域广泛、权威专家数量庞大的评审专家库,鼓励在专家库中采取随机方式组建评审组,采用同行评议方式对科研人员进行客观、公平、科学、透明的评价。第三,弱化基于论文数量和影响因子的看似客观的评价方式,避免由此类急功近利式的评价机制所带来的不良后果。

#### 参考文献

- [1] 新华社.中共中央办公厅 国务院办公厅印发《关于深化项目评审、人才评价、机构评估改革的意见》[EB/OL]. [2018-07-24]. [http://www.gov.cn/zhengce/2018-07/03/content\\_5303251.htm](http://www.gov.cn/zhengce/2018-07/03/content_5303251.htm).
- [2] Harriet Zuckerman, Robert K. Merton. Patterns of evaluation in science: Institutionalization, structure and functions of the referee system [J]. *Minerva*, 1971, 9(1): 66-100.
- [3] 于珈,王兰英,李兵等.浅析美国科技人才评价的做法与启示[J]. *中国科技资源导刊*, 2015, 47(2): 68-72, 80.
- [4] Evaluation the procedures of the max planck society [EB/OL].

- [5] 张济洲.美国高校科研经费分配的同行评议本质、局限与改进——以美国国家科学基金会(NSF)资助为例[J]. *中国高教研究*, 2011(10): 40-42.
- [6] 吴锦雅.同行评议面临的问题与可行性措施[J]. *编辑学报*, 2011, 23(3): 238-240.
- [7] Hodgkinson M, Dunckley J. Open peer review & community peer review [EB/OL]. [2017-11-05]. <https://journalology.blogspot.com/2007/06/open-peer-review-community-peer-review.html>.
- [8] Hicks D, Wouters P, et al. The Leiden Manifesto for research metrics [J]. *Nature*, 2015, 520: 429-431.
- [9] The san francisco declaration on research assessment (DO-RA) [EB/OL]. [2017-11-05]. [https://en.wikipedia.org/wiki/San\\_Francisco\\_Declaration\\_on\\_Research\\_Assessment](https://en.wikipedia.org/wiki/San_Francisco_Declaration_on_Research_Assessment).
- [10] Moed HF, Halevi G. Multidimensional assessment of scholarly research impact [J]. *Journal of the Association for Information Science and Technology*, 2015, 66(10): 1988-2002.
- [11] Lin LL, Xu Z M, Ding Y, et al. Finding topic-level experts in scholarly networks [J]. *Scientometrics*, 2013, 97(3): 797-819.
- [12] JCR 报告(期刊影响因子 IF) [EB/OL]. [2017-11-05]. <http://lib.cqmu.edu.cn/jcrbaogao.htm>.
- [13] Matsas G E A. What are scientific leaders? The introduction of a normalized impactor factor [J]. *Brazilian Journal of Physics*, 2012, 42: 319-322.
- [14] Pan R K, Fortunato S. Author impact factor: tracking the dynamics of individual scientific impact [J]. *Scientific Reports*, 2014, 4: 4880.
- [15] Lippi G, Mattiuzzi C. Scientist impact factor (SIF): a new metric for improving scientists' evaluation? [J]. *Ann Transl Med*, 2017, 5(15): 303.
- [16] Wikipedia. Clarivate citation laureates. [EB/OL]. [https://en.wikipedia.org/wiki/Clarivate\\_Citation\\_Laureates](https://en.wikipedia.org/wiki/Clarivate_Citation_Laureates).
- [17] Li X. An analysis of knowledge mapping of high cited papers on library and information science in ESI [J]. *Information Studies Theory & Application*, 2017.
- [18] 刘云,白旭,杨芳娟等.高被引科学家论文合作关系研究[C]. *全国科技评价学术研讨会*, 2015.
- [19] Bornmann L, Bauer J, Schlagberger E M. Characteristics of highly cited researchers 2015 in Germany [J]. *Scientometrics*, 2017, 111(1): 1-3.
- [20] 尹志欣,谢荣艳.我国顶尖科技人才现状及特征研究——以汤森路透2015高被引科学家为例[J]. *科技进步与对策*, 2017, 34(1): 136-140.
- [21] 邓侨侨.高被引科学家国别迁移过程中的集聚分析[J]. *高教探索*, 2014(5): 17-20.
- [22] 刘俊婉.高被引科学家人才流动的计量分析[J]. *科学学研究*, 2011, 29(2): 192-197.
- [23] Podlubny I. Comparison of scientific impact expressed by the number of citations in different fields of science [J]. *Scientometrics*, 2005, 64(1): 95-99.
- [24] Radicchi F, Fortunato S, Castellano C. Universality of citation dis-

- tributions: Toward an objective measure of scientific impact [J]. Proceedings of the National Academy of Sciences, 2008, 105 (45): 17268-17272.
- [25] Abramo G, Cicero T, D'Angelo C A. Revisiting the scaling of citations for research assessment [J]. Journal of Informetrics, 2012, 6 (4): 470-479.
- [26] Abramo G, D'Angelo C A. A methodology to compute the territorial productivity of scientists: The case of Italy [J]. Journal of Informetrics, 2015 (9): 675-685.
- [27] Zitt, Small H. Modifying the journal impact factor by fractional citation weighting: The audience factor [J]. Journal of the American Society for Information Science & Technology, 2008, 59 (11): 1856-1860.
- [28] Waltman L, Eck N J V. A systematic empirical comparison of different approaches for normalizing citation impact indicator [J]. Journal of Informetrics, 2013, 7 (4): 833-849.
- [29] Pepe A, Kurtz M. A measure of total research impact independent of time and discipline [J]. Plos One, 2012, 7 (11): 3725-3744.
- [30] Hirsch J E. An index to quantify an individual's scientific research output [J]. Proceedings of the National Academy of Sciences of the USA, 2005, 102 (46): 16569-16572.
- [31] Wildgaard L, Schneider J W, Larsen B A. A review of the characteristics of 108 author-level bibliometric indicators [J]. Scientometrics, 2014, 101 (1): 125-158.
- [32] Alonso S, Cabrerizo F J, Herrera-Viedma E, et al. hg-index: a new index to characterize the scientific output of researchers based on the h- and g-indices [J]. Scientometrics, 2010, 82 (2): 391-400.
- [33] 金碧辉. Rousseau.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.
- [34] Bihari A, Tripathi S. EM-index: a new measure to evaluate the scientific impact of scientists [J]. Scientometrics, 2017, 112 (1): 659-677.
- [35] Bihari A, Tripathi S. Year based EM-index: a new approach to evaluate the scientific impact of scholars [J]. Scientometrics, 2018 (4): 1-31.
- [36] Prathap G. The 100 most prolific economists using the p-index [J]. Scientometrics, 2010, 84 (1): 167-172.
- [37] 赵蓉英, 魏明坤, 杨慧云. p 指数应用于学者学术影响力评价的相关性研究——以图书情报学领域为例 [J]. 情报理论与实践, 2017, 40 (4): 61-65.
- [38] Whitfield J. Collaboration: Group theory [J]. Nature, 2008, 455: 720-723.
- [39] Abbasi A, Altmann J, Hossain L. Identifying the effects of co-authorship networks on the performance of scholars: A correlation and regression analysis of performance measures and social network analysis measures [J]. Journal of Informetrics, 2011, 5 (4): 594-607.
- [40] McCarty C, Jawitz J, Hopkins A, et al. Predicting author h-index using characteristics of the co-author network [J]. Scientometrics, 2013, 96 (2): 467-483.
- [41] 陈云伟, 邓勇, 陈方, 等. 复合合作强度指数构建及应用研究 [J]. 图书情报工作, 2015, 59 (13): 96-103.
- [42] Zhai L, Yan X, Zhang G. Bi-directional h-index: A new measure of node centrality in weighted and directed networks [J]. Journal of Informetrics, 2018, 12 (1): 299-314.
- [43] Schulz C, Uzzi B, Helbing D, et al. A network-based citation indicator of scientific performance [J]. 2018.
- [44] Yan E J, Ding Y. Discovering author impact: A pagerank perspective [J]. Information Processing and Management, 2011, 47 (1): 125-134.
- [45] Liu X, Bollen J, Nelson M L, et al. Co-authorship networks in the digital library research community [J]. Information Processing and Management, 2005, 41 (6): 1462-1480.
- [46] Radicchi F, Fortunato S, Markines B, et al. Diffusion of scientific credits and the ranking of scientists [J]. Physical Review E, 2009, 80 (5): 056103.
- [47] 邱均平, 余厚强. 替代计量学的提出过程与研究进展 [J]. 图书情报工作, 2013, 57 (19): 5-12.
- [48] Scott Chamberlain. Consuming article-level metrics: Observations and lessons from comparing aggregator provider data [J]. Information Standards Quarterly, 2013, 25 (2): 5-13.
- [49] Wang X, Fang Z, Guo X. Tracking the digital footprints to scholarly articles from social media [J]. Scientometrics, 2016, 109 (2): 1365-1376.
- [50] Wang X, Wang Z, Mao W, et al. How far does scientific community look back? [J]. Journal of Informetrics, 2014, 8 (3): 562-568.
- [51] Wang X, Xu S, Peng L, et al. Exploring scientists' working timetable: Do scientists often work overtime? [J]. Journal of Informetrics, 2012, 6 (4): 655-660.
- [52] Kolahi J. Altmetrics: A new emerging issue for dental research scientists [J]. Dental Hypotheses, 2015, 6 (1): 1-2.
- [53] Kousha K, Thelwall M. Web indicators for research evaluation: Part 3: books and non standard outputs [J]. El Profesional De La Información, 2015, 24 (6): págs.724-736.
- [54] Chen K H, Tang M C, Wang C M, et al. Exploring alternative metrics of scholarly performance in the social sciences and humanities in Taiwan [J]. Scientometrics, 2015, 102 (1): 97-112.
- [55] Wainer J, Vieira P. Correlations between bibliometrics and peer evaluation for all disciplines: the evaluation of Brazilian scientists [J]. Scientometrics, 2013, 96 (2): 395-410.
- [56] Bridges D. Research quality assessment in education: Impossible science, possible art? [J]. British Educational Research Journal, 2009, 35: 497-517.
- [57] Costas R, Leeuwen T V, Bordons M A. A bibliometric classificatory approach for the study and assessment of search performance at the individual level: The effects of age on productivity and impact [J]. Journal of the American Society for Information Science & Technology, 2010, 61 (8): 1564-1581.
- [58] Zhang C T. The h'-index, effectively improving the h-index based on the citation distribution [J]. PloS One, 2013, 8 (4): e59912.
- [59] 叶鹰. 国际学术评价指标研究现状及发展综述 [J]. 情报学报, 2014, 33 (2): 219-220.

(下转第 60 页)

ses—validating new indicators by analyzing application rationales [J]. *Research Policy* ,2004 ,33( 6/7) : 939-957.

[7] Harhoff D ,Scherer F M ,Vopel K. Citations ,family size ,opposition and the value of patent rights [J]. *Research Policy* ,2003 ,32( 8) : 1343-1363.

[8] Lai Y H ,Che H C. Modeling patent legal value by extension neural network [J]. *Expert Systems with Applications* ,2009 ,36( 7) : 10520-10528.

[9] Wang B ,Hsieh C H. Measuring the value of patents with fuzzy multiple criteria decision making: Insight into the practices of the industrial technology research institute [J]. *Technological Forecasting & Social Change* ,2015 ,92: 263-275.

[10] Baron J ,Delcamp H. The private and social value of patents in discrete and cumulative innovation [M]. Springer-Verlag New York ,Inc.2012.

[11] 万小丽 ,朱雪忠. 专利价值的评估指标体系及模糊综合评价 [J]. *科研管理* ,2008 ,29( 2) : 185-191.

[12] 邱洪华 ,陆潘冰. 基于专利价值影响因素评价的企业专利管理策略研究 [J]. *图书情报工作* ,2016 ,60( 6) : 77-83.

[13] 刘勤 ,王少康 ,胡良龙 ,等. 农机装备专利价值评估研究 [J]. *中国农业科技导报* ,2017 ,19( 5) : 86-91.

[14] 谢萍 ,王秀红 ,卢章平. 企业专利价值评估方法及实证分析 [J]. *情报杂志* ,2015 ,34( 2) : 93-98.

[15] 李锋 ,吴洁 ,尹洁. 政府主导型产学研合作中专利价值分析体系研究 [J]. *科技管理研究* ,2013 ,33( 24) : 172-175.

[16] 资智洪 ,何燕玲 ,袁杰 ,等. 专利价值二元分类评估方法的构建及应用 [J]. *科技管理研究* ,2017 ,37( 11) : 129-135.

[17] 范月蕾 ,毛开云 ,于建荣. 核心专利指标效力研究评述 [J]. *图书情报工作* ,2014 ,58( 24) : 121-125.

[18] 张娴 ,方曙 ,肖国华 ,等. 专利文献价值评价模型构建及实证分析 [J]. *科技进步与对策* ,2011 ,28( 6) : 127-132.

[19] 国家知识产权局. 专利导航试点工程工作手册 [EB/OL]. [http://www.sipo.gov.cn/ztl/ywzt/zldhsdgc/zcwj/201311/t20131104\\_874653.html](http://www.sipo.gov.cn/ztl/ywzt/zldhsdgc/zcwj/201311/t20131104_874653.html).

[20] 李睿 ,赵安琪. 文献计量学视野下专利寿命与多重因素的相

关性解析——来自催化领域的实证 [J]. *情报学报* ,2017 ,36( 6) : 547-554.

[21] Barney J. Comparative patent quality analysis: A statistical approach for rating and valuing patents [EB/OL]. [2013-10-10]. <http://www.patentratings.com>.

[22] Burke P F ,Reitzig M. Measuring patent assessment quality—Analyzing the degree and kind of ( in) consistency in patent offices' decision making [J]. *Research Policy* ,2007 ,36( 9) : 1404-1430.

[23] Lerner J. The importance of patent scope: An empirical analysis [J]. *Rand Journal of Economics* ,1994 ,25( 2) : 319-333.

[24] Albert M B ,Avery D ,Narin F ,et al. Direct validation of citation counts as indicators of industrially important patents [J]. *Research Policy* ,1991 ,20( 3) : 251-259.

[25] Frietsch R ,Neuhäusler P ,Jung T ,et al. Patent indicators for macroeconomic growth—the value of patents estimated by export volume [J]. *Technovation* ,2014 ,34( 9) : 546-558.

[26] Narin F ,Hamilton K S ,Olivastro D. The increasing linkage between U.S. technology and public science [J]. *Research Policy* ,2004 ,26( 3) : 317-330.

[27] 李清海 ,刘洋 ,吴泗宗 ,等. 专利价值评价指标概述及层次分析 [J]. *科学学研究* ,2007 ,25( 2) : 1-9.

[28] 朱相丽 ,谭宗颖 ,万昊. 识别有技术转移潜力的专利方法综述 [J]. *图书情报工作* ,2016( 8) : 132-138.

[29] Dietmar Harhoff ,Markus Reitzig. Determinants of opposition against EPO patent grants—the case of biotechnology and pharmaceuticals [J]. *International Journal of Industrial Organization* ,2004 ,22( 4) : 443-480.

[30] 万小丽. 专利质量指标中“被引次数”的深度剖析 [J]. *情报科学* ,2014 ,32( 1) : 68-73.

[31] 李睿 ,赵安琪. 文献计量学视野下专利寿命与多重因素的相关性解析——来自催化领域的实证 [J]. *情报学报* ,2017 ,36( 6) : 547-554.

( 责编/校对: 王平军)

( 上接第 86 页)

[60] Wu Z Q. Average evaluation intensity: A quality-oriented indicator for the evaluation of research performance [J]. *Library & Information Science Research* ,2015 ,37( 1) : 51-60.

[61] Abramo G ,Costa C ,D'Angelo C A. A multivariate stochastic model to assess research performance [J]. *Scientometrics* ,2015 ,102( 2) : 1755-1772.

[62] 张永莉 ,邹勇. 创新人才创新力评估体系与激励制度研究 [J]. *科学管理研究* ,2012 ,30( 6) : 89-93.

[63] 陈云伟 ,邓勇 ,陈方 ,等. 科学家创新力评价模型及实证研究. *情报杂志* ,2016 ,35( 3) : 142-147.

[64] 陈云伟 ,吴林震 ,陈方 ,等. 青年科学家竞争力评价模型及实证研究 [J]. *武汉大学学报( 信息科学版)* ,2017 ,42( 专辑) : 219-222.

[65] Bodea C. Artificial intelligence techniques applied to the evaluation of the research and technology development projects and programmes [J]. *Economic Computation & Economic Cybernetics*

*Studies & Research* ,2007 ,41( 3/4) : 141-150.

[66] Fu L D ,Aliferis C F. Using content-based and bibliometric features for machine learning models to predict citation counts in the biomedical literature [J]. *Scientometrics* ,2010 ,85( 1) : 257-270.

[67] Fu L D ,Aphinyanaphongs Y ,Aliferis C F. Computer models for identifying instrumental citations in the biomedical literature [J]. *Scientometrics* ,2013 ,97( 3) : 871-882.

[68] Ebadi A ,Schiffauerova A. A computer system for automatic evaluation of researchers' performance [C]// *International Conference on Scientometrics and Informetrics*. 2015.

[69] Ebadi A ,Schiffauerova A. iSEER: An intelligent automatic computer system for scientific evaluation of researchers [J]. *Scientometrics* ,2016 ,107( 2) : 477-498.

[70] Fortunato S ,Bergstrom C T ,Borner K ,et al. Science of science. [J]. *Science* ,2018 ,359( 6379) : eaao0185.

( 责编: 王菊; 校对: 白燕琼)