

基于Web of Science数据库预测SCI期刊影响因子的方法*

刘雪立

河南省科技期刊研究中心,新乡医学院期刊社《眼科新进展》编辑部,453003,河南新乡

摘要 分析了科技期刊影响因子预测的意义。根据影响因子的概念和计算公式,详细介绍了利用Web of Science数据库的引证分析功能,提前计算SCI来源期刊2012年影响因子的方法。对10种国际权威科技期刊和我国20种权威期刊2012年影响因子提前进行了计算,并与2013年版JCR报告的影响因子进行比较,多数期刊影响因子误差在5%左右。因此认为,该方法是提前预知SCI期刊影响因子的可靠方法。

关键词 SCI数据库;期刊引证报告;科技期刊;影响因子预测

* 河南省哲学社会科学规划项目(编号:2102BZH004)。

1955年, Garfield博士^[1]首次提出影响因子的概念,1963年影响因子正式被确定为期刊的评价指标^[2-3]。影响因子是一个定量的评价方法^[1],对于评价专业性和综合性期刊的总体水平是很有价值的^[4]。随着时间的推移,影响因子的应用延伸到了论文质量评价和研究者业绩评价等领域,甚至应用于学术职务任命、研究计划的基金资助和其他研究计划的资金支持^[5]。科研人员的职称晋升也受到高影响因子期刊的影响^[6]。Kirchhof等^[7]、刘雪立等^[8]认为,在欧美等国家,影响因子也广泛用于科研绩效评价,逐渐形成了全球性影响因子崇拜^[8],被印度科学家Bachhawat^[9]、Lakhotia等^[10]称为影响因子综合征,被美国迈阿密大学的Elsaie

源。通过丰富的专业信息,才能有效地吸引作者和读者,并进而提升学报的影响,打造品牌效应。

第三,加强互动功能的建设。互动是网络传播得天独厚的优势,从调研的网站来看,学报期刊的互动功能还存在很大的不足,设置了互动栏目的网站的比例较低,而有些互动则反馈迟缓,在用户提交问题之后十几天、甚至一两个月都没有任何回复,互动功能形同虚设。学报网站应完善互动功能,增加互动的形式,并提高反馈的效率。

第四,拓展网站服务功能。除了互动功能之外,网站的拓展功能也是吸引用户的重要因素。就目前的期刊网站建设来看,网站的服务功能还不够完善。如期刊的专题服务较少,个性化信息服务不够、作者服务功能不强等。学报网站应在服务功能上继续加强,以更好地服务作者与读者。

4 结语

经过近几年科协期刊主管部门和学报主办部

门的不懈努力,学报网站已初具规模,数字化进程取得了一定的成效,但同时也还存在较多的问题与不足。期刊网站的建设并非一劳永逸,已经建立了网站的期刊还应不断充实网站信息,完善网站功能,提高对作者、读者的服务水平能力,利用自建网站促进期刊产品的数字化生产、传播与营销,从而提升期刊的竞争力,促进期刊的可持续发展。

参考文献

- [1] 陈文静. 浅析数字期刊网站的发展[J]. 长江大学学报:社会科学版, 2012, 35(6):161-162.
- [2] 梁天天. 科技期刊自建网站版块设置初探[J]. 中国科技信息, 2012(11):187-187.
- [3] 王媛. 我国科技期刊自建网站现状与分析[J]. 科技与出版, 2010(7):32-37.

等^[11]称为影响因子神话综合征。英国《睡眠研究杂志》(Journal of Sleep Research)主编Lavie^[12]把期刊竞相提高影响因子的现象称为“影响因子竞赛”;巴西科学家Pêgo-Fernandel等^[13]发出质疑,影响因子是维生素还是毒药?著名的法国文献计量学家Zitt^[14]提出,影响因子究竟是天使、魔鬼还是替罪的羔羊?所有这些都说明,科技期刊的影响因子已经应用到全球科学评价的各个领域。

美国Thomson Reuters科技信息集团通过其Web of Knowledge数据平台的《期刊引证报告》(Journal Citation Reports, JCR),每年发布上一年度SCI和SSCI数据库收录期刊的影响因子,一般是每年6月中、下旬更新,如2010年6月18日发布了2009年的JCR^[15],2011年6月28日发布了2010年JCR^[16],2012年6月29日发布了2011年JCR^[17]。2013年6月20日公布了2012年JCR。但是,很多情况下我们需要提前知道SCI数据库来源期刊的影响因子。本文将重点探讨利用Web of Science(WoS)数据库提供的数据,提前计算SCI来源期刊影响因子的方法,对国内外较有影响的期刊2012年的影响因子进行预测,并与2012年JCR报告的影响因子进行比较。

1 预测科技期刊影响因子的意义

预测科技期刊的影响因子具有以下几方面的意义:

(1)对于科技期刊编辑从业人员来讲,了解科技期刊影响因子预测方法、提前精确计算科技期刊影响因子能够提前了解目标期刊新一年的影响因子,正确把握国际科技期刊影响力变化趋势;对于情报学研究工作者,能够提高情报分析与预测的科学性,提升科技信息咨询和情报服务的时效性、针对性和有效性;对于科研管理和科学评价研究人员,能够科学合理地进行科研绩效评价,在一定程度上提高了科研绩效评价的时效性。

(2)对于SCI来源期刊的管理者来讲,了解精确预测和计算影响因子的方法,可以核对JCR公布的各自期刊的影响因子是否正确,有无较大出入。如《加拿大医学会杂志》(Canadian Medical Association Journal, CMAJ)的Joseph等^[18]就曾经发现1999年JCR公布的CMAJ的影响因子有误,

希望作者一如既往地关注该刊。

(3)对于科研人员来讲,了解精确预测和计算影响因子的方法,能够极大提高投稿的针对性、期刊选择的正确性。目前,国内外许多大学和科研机构出台了形形色色的科研绩效评价和奖励政策,对本单位科研人员发表论文期刊的影响因子提出了各种要求。所以,很多科研人员投稿时非常关注期刊的影响因子。但让他们困惑的是,每年6月份前瞄准影响因子 m 的期刊投稿,结果论文录用或发表后,新一年JCR公布的该刊影响因子有明显下降。提前计算科技期刊的影响因子对指导作者投稿就显得很有必要。

2 预测科技期刊影响因子的条件

影响因子的计算公式^[19-20]如下:

$$JIF(Y) = \frac{\text{某期刊Y1和Y2年发表论文在Y年的被引频次}}{\text{某期刊Y1和Y2年发表的可被引文献数}}$$

式中, $JIF(Y)$ 指某期刊Y年的影响因子, $Y1=Y-1$,指Y年的上一年, $Y2=Y-2$,指Y1年的上一年。经过观察2011年版JCR中Nature杂志2010年影响因子的计算过程,可被引文献是指期刊发表的论文(Article)和综述(Review)等2类文献类型。由此公式可知,预测科技期刊影响因子必须具备下列条件:

- (1)具有Web of Science数据库访问权限;
- (2)精确获取某期刊前2年发表的论文在当年的被引频次;
- (3)精确获取某期刊前2年发表的论文和综述等两类文献的数量;
- (4)每年2月份以后、6月份以前(JCR发布新一年数据前)获取数据。

3 预测影响因子的方法

3.1 单一期刊影响因子预测方法

现以2012年科技期刊影响因子预测为例进行介绍。

被引频次获取方法。登录Web of Science数据库,检索该期刊2010年和2011年发表的全部文献。我们选择通过期刊的ISSN号进行高级检索,检索式:“IS=XXXX-XXXX AND PY=2010-2011”。点击检索结果(带有超级链接的检索到的文献条数),显示出检索结果列表,再点击页面上

角的“创建引文报告”,该期刊前2年发表的所有文献在2012年的被引频次总和就清晰地显示出来。

可被引文献数的获取方法。回到上述检索结果列表页面,点开左边工具栏中的“文献类型”,查看“Article”和“Review”等两类文献的总数,即为该刊可被引文献数。

影响因子计算。带入影响因子计算公式,精确计算该刊2012年的影响因子。

3.2 多种期刊影响因子预测方法

同时预测多种期刊影响因子和单一期刊影响因子预测的原理是相同的,只是文献检索和数据处理的方法有所区别,现介绍如下。

(1)各期刊被引频次获取方法。登录Web of Science数据库,同时检索多种期刊2010年和2011年发表的全部文献(检索结果条目数控制在1万以内,若超过1万条,无法创建引文报告)。我们选择通过各期刊的ISSN号进行高级检索,检索式:“IS=(XXXX-XXXX OR XXXX-XXXX OR ……)AND PY=2010-2011”。点击检索结果(带有超级链接的检索到的文献条数),显示出检索结果列表,再点击页面右上角的“创建引文报告”,在完整页面左下角“输出记录”区域选择最后选项,在文本框内填写相应数字,如1-500,501-1000,1001-1500……,每次输出记录不得超过500条(数据库限制);“输出文本格式”选择Excel文件。文献记录输出完成后,把所有Excel文件汇总为一个完整的Excel文件,仅保留出版物名称和2012年被引频次等数据,其他数据均可删除。然后按照出版物名称排序,通过Excel简单操作即可获取各期刊2010年和2011年论文在2012年的被引频次。

(2)各期刊可被引文献数的获取方法。回到上述检索结果列表页面,点开左边工具栏中的“文献类型”,勾选“Article”和“Review”等两类文献的总数,点击“精炼”按钮,这样就把这2类文献保留下来。然后点击右上角的“分析检索结果”,选择按“来源出版物”进行分析。这样就把各期刊2010年和2011年发表的“Article”和“Review”等两类文献的数量清晰地呈现在页面上。

(3)各期刊影响因子计算。把各期刊2010年和2011年发表的所有文献在2012年的被引频次

和该2年发表的“Article”和“Review”等两类文献的数量汇总到Excel文件,通过批处理计算出各期刊2012年的影响因子。

4 部分期刊2012年影响因子预测结果与误差

4.1 国外部分期刊2012年影响因子预测结果与误差

采用本文介绍的方法,对国际权威性综合性自然科学期刊Nature、Science,2011年影响因子最高的期刊Cancer Journal for Clinicians及部分权威医学期刊2012年的影响因子进行了预测。预测的Cancer Journal for Clinicians影响因子为149.459(2012年JCR报告的实际影响因子为153.459,预测值低了2.6%),比2011年影响因子有大幅度增加,依然是SCI数据库中影响因子最高的期刊。如此高的影响因子主要由Jemal等^[21-22]的2篇文章“2010年癌症统计”和“全球癌症统计”贡献。Nature的影响因子预测值为37.642(2012年JCR报告的实际影响因子为38.597,预测值低了2.5%)。国外10种权威期刊2012年影响因子预测结果及预测误差见表1。除British Medical Journal和PLoS Medicine影响因子预测误差较大外,其他期刊影响因子预测误差均在5%左右,预测值均小于实际值。

表1 国外10种权威期刊2012年影响因子预测结果及预测误差

期刊名	2010-2011 论文数/篇	2012年 被引频/次	2012年 IF预测值	2012年 IF实际值	预测误差/ %
Cancer Journal for Clinicians	37	5530	149.459	153.459	0.026
New England journal of Medicine	694	34764	50.092	51.658	0.030
Nature	1703	64105	37.642	38.597	0.025
Lancet	547	20237	36.996	39.060	0.053
Cell	658	20905	31.770	31.957	0.006
Science	1734	52141	30.070	31.027	0.031
Journal of the American Medical Association	452	13104	28.991	29.978	0.033
British Medical Journal	575	7620	13.252	17.215	0.230
PLoS Medicine	225	2891	12.849	15.253	0.158
Annals of Internal Medicine	338	4382	12.964	13.976	0.072

4.2 国内部分期刊2012年影响因子预测结果及预测误差

对2011年我国影响因子最高的20种期刊2012年的影响因子进行了预测。预测结果及预测误差见表2。

由表2可知,2011年我国影响因子较高的前20种期刊有1种在JCR数据库中已检索不到,其余19种SCI来源期刊影响因子预测误差最大的期刊是Communications in Computational Physics(预测误差为29.3%),其次是Acta Pharmacologica Sinica和CNS Neuroscience Therapeutics,误差超过了10%。15种期刊预测误差在5%左右,无一例外均为预测值小于实际值,与国外SCI期刊影响因子预测误差基本相同。说明该预测和计算影响因子的方法是可信的,也得到了该数据库技术服务人员的认可。

表2 国内部分期刊2012年影响因子预测结果

期刊名	2010-2011 论文数/篇	2012年 被引频 次/次	2012年 IF预测值	2012年 IF实际值	预测 误差
Cell Research	210	2088	9.943	10.526	0.055
Journal of Molecular Cell Biology	65	457	7.031	7.308	0.038
Nano Research	222	1553	6.995	7.392	0.054
Molecular Plant	175	1069	6.109	6.126	0.003
Fungal Diversity	119	601	5.050	5.319	0.051
CNS Neuroscience Therapeutics	120	465	3.925	4.458	0.120
Journal of Integrative Plant Biology	180	661	3.672	3.750	0.021
Cellular Molecular Immunology	129	416	3.225	3.419	0.057
Science China Life Sciences	303	866	2.858	-	-
International Journal of Oral Science	57	149	2.614	2.719	0.039
World Journal of Gastroenterology	1490	3714	2.493	2.547	0.021
Journal of Genetics and Genomics	144	298	2.069	2.076	0.003
Asian Journal of Andrology	222	458	2.063	2.140	0.036
Acta Pharmacologica Sinica	396	793	2.003	2.354	0.149
Journal of Digestive Diseases	129	239	1.853	1.853	0.000
Journal of Systematics and Evolution	101	177	1.752	1.851	0.053
Journal of Environmental Sciences China	565	938	1.660	1.773	0.064
Communications in Computational Physics	233	307	1.318	1.863	0.293
Particology	191	248	1.298	1.419	0.085
Journal of Plant Ecology	62	79	1.274	1.355	0.060

4.3 影响因子预测误差产生的原因

几乎各期刊影响因子预测值与实际值相比均存在一定误差,预测误差产生的原因主要有3个方面:作者引文错误。就预测期刊影响因子的误差,专门咨询了数据库技术服务人员,认为这样预测和计算影响因子是科学的,误差产生的原因主要是作者引文错误。通过“创建引文报告”的引文信息必须与原文完全匹配才会被计为1次被引用,只要有任意的错误,如起止页码不一致、人名拼写错误、文题不完整或拼写错误等,都将导致数据库检索的论文被引频次小于实际被引频次。而计算影响因子所用的论文被引频次只要期刊名和论文发表年代匹配,期刊的被引用就是有效的。所以,通过“创建引文报告”获取的各论文被引频次一般小于计算影响因子的被引频次。因此出现了所有期刊影响因子预测值均小于实际值的情况。数据加工和处理错误。仅2012年1年,WoS处理的文献条目就超过182万条(数据库检索结果),期刊引证数据很难避免出现差错,这也将导致影响因子预测误差。时间因素。由于部分期刊出版延迟和数据库人力资源限制,导致个别期刊录入数据库延迟时间较长,到次年5月份还有上年度论文收入数据库,预测影响因子过早的话,有部分被引用尚未能在数据库显示,一般到4月底预测期刊影响因子的误差基本可以忽略不计。

5 需要注意的几个问题

5.1 准确界定和计量论文被引频次和论文数

本文介绍的是2012年期刊影响因子的预测方法,那么影响因子计算公式中的分母是某期刊2010年和2011年发表的论文在2012年(1年内)的被引频次,不得累计2010年和2011年的被引频次,更不能计算2013年的被引频次。其分母仅计算2010年和2011年2年内发表的论文(Article)和综述(Review)等2类文献类型的数量。准确界定和计量论文被引频次和论文数是精确预测影响因子的关键。

5.2 同时预测多种期刊影响因子的问题

同时预测多种期刊影响因子时,必须保证这些期刊2年内发表的所有文献类型条目(包括Article、

Review、Letter、Editorial、Correspondence、Notes、Correction等)不超过1万条。因为,如果超过1万条,系统“创建引文报告”功能不可用,将无法准确计量2012年各期刊被引频次。

5.3 预测新一年期刊影响因子的时机

JCR发布新一年数据后就不存在预测影响因子的问题(但可以验证目标期刊影响因子计算是否正确和准确),2月份前获取数据,前两年论文的被引频次数据还不全。所以预测科技期刊影响因子最佳时机应在每年的3月初。

5.4 注意与上一年影响因子对比

如果预测某期刊的影响因子与上一年该刊影响因子差异很大,一定要反复核对数据。如果数据获取和计算无误,应查找导致影响因子大幅度增加的高被引论文。如果影响因子下降明显,则应查看上一个影响因子计算时间区间内是否有极高被引论文。这样做的目的,是确认预测的影响因子的正确性和准确性。

参考文献

- [1] Garfield E. Citation indexes for science. A new dimension in documentation through association of ideas[J]. Science, 1955, 122(3159):108-111.
- [2] 任胜利,王宝庆,郭志明,等.应慎重使用期刊的影响因子评价科研成果[J]. 科学通报, 2000, 45(2): 218-222.
- [3] 黄劲松.期刊影响因子的偏差分析[J]. 编辑学报, 2004, 16(1):77-78.
- [4] Garfield E. Citation analysis as a tool in journal evaluation. Journals can be ranked by frequency and impact of citations for science policy studies[J]. Science, 1972, 178(60):471-479.
- [5] 刘雪立,秦小川.解读欧洲科学编辑学会关于影响因子不当使用的声明[J]. 中国科技期刊研究, 2009, 20(1):98-100.
- [6] Deng P, Loh M, Mondry A. The “ impact factor ” revisited[J]. Biomedical Digital Libraries, 2005, 2 :7.
- [7] Kirchhof B, Bornfeld N, Grehn F. The delicate topic of the impact factor[J]. Graefe's Arch Exp Ophthalmol, 2007, 245 :925-927.
- [8] 刘雪立. 全球性SCI现象与影响因子崇拜[J]. 中国科技期刊研究, 2012, 23(2):185-190.
- [9] Bachhawat A K. The impact factor syndrome[J]. Current Science, 2002, 82(11):1307.
- [10] Lakhota SC. “Impact factor” and “we also ran” syndrome[J]. Current Science, 2010, 99(4):410.
- [11] Elsaie ML, Kammer J. Impactitis: The impact factor myth syndrome[J]. Indian J Dermatol, 2009, 54(1):83-85.
- [12] Lavie P. The race for the impact factor[J]. J Sleep Res, 2009(18):283-284.
- [13] Pêgo-Fernandel P M, Marianill A W. Impact factor: vitamin or poison[J]. Sao Paulo Med J, 2010(128): 185-186.
- [14] Zitt, M. The journal impact factor: angel, devil, or scapegoat[J]? Scientometrics, 2012(92):485-503.
- [15] 刘雪立,方红玲,王梅英,等.2009年7347种SCIE期刊文献计量学指标统计分析[J]. 中国科技期刊研究, 2011, 22(1):44-52.
- [16] 康晓东,齐颖. 美国SCI检索收录期刊目录(分类目录)2010年影响因子公布中国大陆医学影像学类杂志榜上无名[J]. 磁共振成像, 2011(4):320.
- [17] 《催化学报》编辑部.《催化学报》2011年SCI影响因子首次突破1.0[J]. 催化学报, 2012, 33(7): 1145.
- [18] Joseph K S, Hoey J. CMAJ's impact factor: Room for recalculation[J]. CMAJ, 1999, 161(8):977-978.
- [19] Campanario JM, Molina A. Surviving bad times: The role of citations, self-citations and numbers of citable items in recovery of the journal impact factor after at least four years of continuous decreases[J]. Scientometrics, 2009, 81(3):859-864.
- [20] Andrade A, González-Jonte R, Campanario J M. Journals that increase their impact factor at least fourfold in a few years: The role of journal self-citations[J]. Scientometrics, 2009, 80(2):517-530.
- [21] Jemal A, Siegel R, Xu J Q, et al. Cancer Statistics, 2010[J]. Cancer Journal for Clinicians, 2010, 60(5):277-300.
- [22] Jemal A, Bray F, Center M M, et al. Global Cancer Statistics[J]. Cancer Journal for Clinicians, 2011, 61(2):69-90.