



特征因子算法与 SJR 算法的比较分析及实证研究

□庄纪林*

摘要 给出一定条件,实现了 SJR 算法向特征因子算法的转化。选取 CSSCI 收录的图书情报学期刊 2007 年互引矩阵,分别计算被引量、EF、SJR、IF、AI、SJRQ,并分成两类进行相关度分析。结果表明,在一定条件下,SJR 算法可转化成特征因子算法,并且期刊评价指标 EF 与 SJR 指标显著相关,它们完全可以作为被引频次与 IF 的参考、补充甚至替代。

关键词 特征因子 文章影响分值 SJR SJRQ 影响因子 期刊评价 引文分析

分类号 G35

DOI 10.16603/j.issn1002-1027.2019.02.015

1 引言

期刊评价是文献计量学的重要应用领域,主要用于反映学术期刊的质量和水平。期刊评价的理论和方法经过长期的研究与发展,已经取得了众多成果,在实际应用方面也取得了长足进展。文献计量指标从传统的载文量、被引量、影响因子(Impact Factor,以下简称 IF)^[1,2]、到后来提出的 h 指数^[3]、g 指数^[4]、A 指数、r 指数等类 h 指数^[5,6]。随着谷歌的页面排名(PageRank)算法出世,2008 年 Scopus 推出了基于自身数据库的评价指标 SCIMAGO 期刊排名(SCImago journal rank,以下简称 SJR)^[7]与每篇论文的参考源标准影响因子(Source Normalized Impact per Paper,以下简称 SNIP)^[8]。作为回应,汤姆森路透科技集团也推出了期刊评价指标特征因子(Eigenfactor,EF)^[9]作为 JCR 的增强功能,并于 2009 年初正式采用。2012 年 SJR 和 SNIP 对各自的算法进行修正与改进,发表了修正指标 SJR2^[10]与 SNIP2^[11]。2015 年汤姆森路透科技集团也发布了两个新的文献计量学指标,一个是标准特征因子(Normalized Eigenfactor,以下简称 NEI)^[12],一个是期刊影响因子百分位(Journal Impact Factor Percentile,以下简称 JIFP)^[13]。而作为新型期刊评价指标的代表 SJR 和 EF,两者都采用了类 PageRank 算法模型来评价科技期刊的质量与影响力,是当前期刊评价指标的研究热点。

几十年来,IF 得到了广泛的认可和应用,然而近年来学者们把更多注意力投入到 SJR 和 EF 等类 PageRank 算法中,究其原因,是因为 IF 存在着以下几点问题^[14]:

1)IF 在计算过程中并没有考虑引用质量,只有对引文数量的统计。一篇学术权威发表的文章中引用的参考文献显然应该具有更高的价值。

2)在计算过程中未能排除期刊的自引行为,这使得期刊可以通过增加自引来提高影响因子。

3)IF 计算方法采用的时间段为两年,由于研究领域差异性导致不同领域的论文的被引频次有较大差异。那些科研周期较长、发表周期也相对较长的领域,有可能出现引用行为因超出两年的时限不能被计入被引量,进而影响 IF 的计算结果。

4)SCI 收录的不同学科的期刊数目差别很大,这种差别不仅反映在论文数量上,而且反映在引文数量上。而相同或相近研究领域的论文倾向于相互引证,这又反过来加大了不同学科间刊物的影响因子的差异。

5)基于少量期刊刊载大量有重大意义和影响的科研成果的研究结果,SCIE 收录期刊只占全世界期刊总量的 3.6%,而且语种大多为英文,非常有利于以英语为母语的国家的刊物获得高影响因子。

自从 SJR 和 EF 算法问世以来,由于这类算法不仅考虑了引文的数量,而且考虑了引文的质量,能

* 庄纪林,ORCID:0000-0002-7783-0964,邮箱:zhuangjl@pku.edu.cn.



够很容易地去除自引,其统计年限窗口也分别扩大为3年和5年,因此受到了广大学者的逐步关注与认可,国内也有很多学者对其进行了研究。《特征因子原理及实证研究》^[15]一文以18种CSSCI收录图书情报学期刊为例,计算出它们的特征因子分值和论文影响分值,在此基础上对这两项指标同其他期刊评价指标的关系进行了探讨,得出了特征因子分值、论文影响分值和期刊综合指数、h指数、影响因子之间存在较强的皮尔逊相关性的结论;《SJR指数研究及其与影响因子的比较分析》^[14]一文将SJR指数与影响因子各方面的特征进行了比较,并通过2007年的实际数据对这两项指标的实际效果进行了对比分析,总结了两者在期刊评价中的优势以及不足之处;《评价期刊影响力的三项指标比较研究》^[16]一文以JCR中信息科学与图书馆学中影响因子排名前20的期刊为样本,对影响因子、5年影响因子、SJR以及特征因子的排名进行比较分析,得出SJR与其他三者之间没有发现显著相关性的结论。以上文章或者是实现了某种期刊评价算法并试图对评价结果进行分析,或者是对研究机构发布的多个期刊评价指标直接分析相关性。

SJR算法和EF算法虽然都借鉴了PageRank算法的思想,但这两种算法对期刊进行评价的结果又不尽相同,这固然是由于其统计源期刊和统计年限窗口的不同而引起的,但如果去除了这两个因素,只从算法中去看,SJR算法和特征因子算法之间又存在什么样的区别呢?在什么情况下两者是等价的?这就是本文所要探讨的问题。笔者将通过算法的比较分析以及实证研究来说明SJR算法和EF算法这两者之间的异同。

2 理论分析

2.1 EF算法推导

EF指标的计算考虑到了期刊的引用权重,通过引文构建起文献的引用网络,对期刊的影响力进行评价^[17]。其计算公式如下:

$$\pi^{(k+1)} = \alpha H \pi^{(k)} + a(\alpha d \pi^{(k)} + 1 - \alpha) \quad (1)$$

$$EF = 100 \frac{H \pi^*}{\sum_i [H \pi^*]_i} \quad (2)$$

$$AI_i = 0.01 \frac{EF_i}{a_i} \quad (3)$$

其中:

1) α 是一个常数,一般取 0.85。

2) a 是期刊的文章向量,其第 i 个分量 a_i 就是统计年限窗口内期刊 i 的文章总和除以所有统计源期刊的文章总和,即统计年限窗口内期刊 i 在所有统计源期刊中的发文占比。 a 为列向量。

3) H 是规范化期刊互引矩阵,定义为

$$H = \begin{bmatrix} z_{11}/z_1 & \cdots & z_{1N}/z_N \\ \vdots & \ddots & \vdots \\ z_{N1}/z_1 & \cdots & z_{NN}/z_N \end{bmatrix} \quad (4)$$

其中 $H_{ij} = \frac{z_{ij}}{z_j}$, z_{ij} 表示期刊 j 对期刊 i 的引用次数, $z_j = \sum_k z_{kj}$ 表示期刊 j 的参考文献总数。由于排除了自引,所以当 $i=j$ 时, $z_{ij}=0$ 。

4) N 为统计源期刊总数。

5) d 为悬点向量,如果期刊 i 不是悬点,则 $d_i=0$,反之, $d_i=1$ 。 $d=(d_1, d_2, \dots, d_N)$ 是一个行向量。

6) $\pi^{(k)}$ 为迭代计算式(1)的第 k 次迭代结果。 $\pi^{(k)}$ 为列向量。

7) π^* : 当 $|\pi^{(k+1)} - \pi^{(k)}|$ 中的每个分量都小于计算之前设置的某个阈值时,迭代终止,此时 $\pi^* = \pi^{(k+1)}$ 。

8) EF_i 为期刊 i 的特征因子。 EF 为列向量。

9) AI_i 为期刊 i 的论文影响分值。 AI 为列向量。

2.2 SJR算法推导

SJR指标的计算基于期刊之间的声望传递,这种声望是通过一个期刊对其他期刊和自身的引用来实现的^[18]。期刊的最终声望的计算也是一个迭代过程,期刊第 i 阶段的声望取决于第 $i-1$ 阶段的期刊的声望。其计算公式如下:

$$SJR_i = \frac{(1-d-e)}{N} + e \frac{Art_i}{\sum_{j=1}^N Art_j} + d \sum_{j=1}^N \frac{c_{ij} \cdot SJR_j}{c_j} \frac{1 - (\sum_{k \in \{Dangling\ nodes\}} SJR_k)}{\sum_{h=1}^N \sum_{k=1}^N \frac{C_{kh} \cdot SJR_k}{C_k}} + d \left(\sum_{k \in \{dangling\ nodes\}} SJR_k \right) \frac{Art_i}{\sum_{j=1}^N Art_j} \quad (5)$$

$$SJRQ_i = \frac{SJR_i}{Art_i} \quad (6)$$

其中:

1) d 为常数,一般取 0.85。



2) e 为常数,一般取 0.10。

3) N 为统计源期刊总数。

4) ART_j : 期刊 j 的论文总数。

5) c_{ji} : 期刊 j 引用期刊 i 的次数。

6) c_j : 期刊 j 的参考文献总数。

7) Dangling-nodes: 期刊引用网中的孤立节点(悬点), 它们与其他期刊没有任何引用关系。

8) SJR_i : 期刊 i 的 SJR 值。

9) $SJRQ_i$: 期刊 i 的 SJRQ 值。

2.3 EF 算法与 SJR 算法的比较分析

通过以上对 EF 算法和 SJR 算法的分析,不难发现,类似于 PageRank 算法,期刊的 EF 指数与 SJR 指数都与该期刊由被引所累积的声望相关,声望越大,研究者访问该期刊的概率越大。特征因子算法计算出 EF 与 AI 两个指标, SJR 算法计算出 SJR 与 SJRQ 两个指标。从对应关系上来看, EF 对应 SJR, 代表了期刊的声望或影响力, 决定了研究者再次访问该期刊的概率; AI 与 SJRQ 代表了该期刊中文章的篇均影响力, 消除了该期刊发文章大小的影响。如果再考虑影响因子算法中的指标, 显然被引量对应 EF 与 SJR, IF 对应 AI 与 SJRQ。这里, 将被引量、EF、SJR 看作第一类指标, 将 IF、AI、SJRQ 看作第二类指标, 将在后面的小节中分析它们之间的相关性。必须指出的是, 如果将 EF 值或 SJR 值与影响因子 IF 进行相关性分析, 这是不恰当的。正如将人的体重与人的体重身高比(代表肥胖程度)进行相关性分析、试图得出体重和肥胖程度之间存在较强相关性一样, 这是不恰当的。

EF 算法和 SJR 算法都采用了类似 PageRank 的算法, 但在数据和算法上还是存在着以下区别:

表 1 EF 算法和 SJR 算法的主要特征对比

特征	EF 算法	SJR 算法
发布时间	2009 年	2007 年
开发者	Thomson 科技	SCImago
数据来源	ISI Web of Science	Scopus
期刊数量	7000 多种	16000 多种
更新频率	每周	每日
统计窗口年限	5 年	3 年
期刊自引	排除	包括

期刊引用数	只包含对统计源期刊的引用	所有引用的总和
迭代计算式项数	互引矩阵和文章向量	互引矩阵、文章向量和期刊数的倒数(1/N)
最终值	对迭代过程中满足条件的收敛值多了一步处理, 即(2)式	直接采用迭代过程中满足条件的收敛值, 即(5)式

通过表 1, 可以很清楚地发现 EF 算法和 SJR 算法不仅在统计源期刊和统计年限窗口上不一样, 而且在算法上也存在一定的区别, 这就导致了采用了 EF 算法与采用 SJR 算法进行期刊评价结果之间的差异。如果不考虑统计源期刊和统计年限窗口因素的影响, 单从算法而论, 这两者之间又存在着多大的区别呢? 通过笔者的推导, 发现在一定条件下可实现 SJR 算法向 EF 算法的转化。

2.4 SJR 算法向 EF 算法的转化

为了清晰地说明 SJR 算法向 EF 算法转化的过程并进行比较, 先仔细观察一下(5)式中第 3 项加数

$$\sum_{j=1}^N \frac{c_{ji} \cdot SJR_j}{c_j} \quad (7)$$

这个部分有一个较易令人迷惑的地方就是在 SJR 算法中以 c_{ji} 表示期刊 j 对期刊 i 的引用数量, 而 EF 算法中以 Z_{ij} 表示期刊 j 对期刊 i 的引用数量, c_{ji} 与 Z_{ij} 的下标顺序正好相反。

还是按照(5)式中 SJR 算法的标记法, 即以 c_{ji} 表示期刊 j 对期刊 i 的引用数量。现在重点分析(7)

$$\text{式, 令 } R_i = \sum_{j=1}^N \frac{c_{ji} \cdot SJR_j}{c_j} \quad (8)$$

把(8)式改写成向量的点积形式为

$$R_i = \begin{pmatrix} \frac{c_{1i}}{c_1} & \frac{c_{2i}}{c_2} & \dots & \frac{c_{Ni}}{c_N} \end{pmatrix} \begin{pmatrix} SJR_1 \\ SJR_2 \\ \vdots \\ SJR_N \end{pmatrix} \quad (9)$$

以矩阵形式表示(9)式为

$$\begin{pmatrix} R_1 \\ R_2 \\ \vdots \\ R_N \end{pmatrix} = \begin{pmatrix} \frac{c_{11}}{c_1} & \frac{c_{21}}{c_2} & \dots & \frac{c_{N1}}{c_N} \\ \frac{c_{12}}{c_1} & \frac{c_{22}}{c_2} & \dots & \frac{c_{N2}}{c_N} \\ \dots & \dots & \dots & \dots \\ \frac{c_{1N}}{c_1} & \frac{c_{2N}}{c_2} & \dots & \frac{c_{NN}}{c_N} \end{pmatrix} \begin{pmatrix} SJR_1 \\ SJR_2 \\ \vdots \\ SJR_N \end{pmatrix} \quad (10)$$



(10)式矩阵中 $\frac{c_{ji}}{c_j}$ 出现在第 i 行第 j 列。按照 SJR 算法对 c_{ji} 的定义,不难得知:矩阵行方向代表施引文献,列方向代表被引文献,这与 EF 算法中矩阵 H 在行方向与列方向代表的意义是一样的。在对矩阵元素进行表示时,通常用下标的左半部分表示该元素的位置在矩阵中的行数,下标的右半部分表示该元素的位置在矩阵中的列数。而(10)式矩阵中 $\frac{c_{ji}}{c_j}$ 下标的表示法正好与对矩阵元素位置的传统表示法相反,不妨,可设 $M_{ij} = c_{ji}$ 、 $M_j = c_j$,则(10)式可改写为

$$\begin{pmatrix} R_1 \\ R_2 \\ \vdots \\ R_N \end{pmatrix} = \begin{pmatrix} \frac{M_{11}}{M_1} & \frac{M_{12}}{M_2} & \dots & \frac{M_{1N}}{M_N} \\ \frac{M_{21}}{M_1} & \frac{M_{22}}{M_2} & \dots & \frac{M_{2N}}{M_N} \\ \dots & \dots & \dots & \dots \\ \frac{M_{N1}}{M_1} & \frac{M_{N2}}{M_2} & \dots & \frac{M_{NN}}{M_N} \end{pmatrix} \begin{pmatrix} \sum_{j=1}^N \frac{M_{1j} \cdot SJR_j}{M_j} \\ \sum_{j=1}^N \frac{M_{2j} \cdot SJR_j}{M_j} \\ \vdots \\ \sum_{j=1}^N \frac{M_{Nj} \cdot SJR_j}{M_j} \end{pmatrix} \quad (11)$$

所以(5)式可改写为

$$SJR_i = \frac{(1-d-e)}{N} + e \frac{Art_i}{\sum_{j=1}^N Art_j} + d \sum_{j=1}^N \frac{M_{ij} \cdot SJR_j}{M_j} \frac{1 - (\sum_{k \in \{Dangling-nodes\}} SJR_k)}{\sum_{h=1}^N \sum_{k=1}^N \frac{M_{hk} \cdot SJR_k}{M_k}} + d \left(\sum_{k \in \{Dangling-nodes\}} SJR_k \right) \frac{Art_i}{\sum_{j=1}^N Art_j} \quad (12)$$

其中: M_{ij} 表示期刊 j 对期刊 i 的引用数量, M_j 表示期刊 j 的参考文献总数,其余符号的定义同 SJR 原算法中的一样。

下面,从(12)式出发,对(12)式做一些形式上的转换。为了实现 SJR 算法向 EF 算法的转化,把 SJR 算法中的参数尽量写成 EF 算法中同类参数的形式:

1)统一 SJR 算法中常数的符号表示:令 $\alpha=d, \beta=e$

2)在 SJR 算法中构造文章向量 a : a 为列向量,其第 i 个分量为

$$a_i = \frac{Art_i}{\sum_{j=1}^N Art_j}$$

可以看到,这里定义的文章向量 a_i 同 EF 算法中的文章向量 a_i 是一样的。

3)在 SJR 算法中构造悬点向量:用一个行向量 $d=(d_1, d_2, \dots, d_N)$ 来表示悬点向量,如果期刊 i 不是悬点,则 $d_i=0$,反之, $d_i=1$ 。这里定义的悬点向量 d 同 EF 算法中的悬点向量 d 是一样的

$$4) \text{在 SJR 算法中引入列向量 } e = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}, \text{行向量}$$

$$e^T = (1 \quad 1 \quad \dots \quad 1)。$$

5)在 SJR 算法中构造规范化互引矩阵 H :令 $H_{ij} = \frac{M_{ij}}{M_j}$,即

$$H = \begin{pmatrix} \frac{M_{11}}{M_1} & \frac{M_{12}}{M_2} & \dots & \frac{M_{1N}}{M_N} \\ \frac{M_{21}}{M_1} & \frac{M_{22}}{M_2} & \dots & \frac{M_{2N}}{M_N} \\ \dots & \dots & \dots & \dots \\ \frac{M_{N1}}{M_1} & \frac{M_{N2}}{M_2} & \dots & \frac{M_{NN}}{M_N} \end{pmatrix} \quad (13)$$

其中, M_{ij} 表示期刊 j 对期刊 i 的引用数; M_j 定义为期刊 j 的参考文献数, M_j 不仅包含期刊 j 对统计源期刊的引用,而且包含其对统计源期刊之外的引用。所以,除悬点以外,总有 $\sum_{i=1}^N H_{ij} \leq 1$ 。矩阵 H 中悬点列仍是全为 0 的列向量。另外,由于 SJR 算法中并没有提到去除自引,所以 H 矩阵中对角线元素不为 0,这是和 EF 算法不同的地方。

6)在 SJR 算法中引入 N 维列向量 $\pi^{(k+1)}$, $\pi_i^{(k+1)}$ 为第 $k+1$ 次迭代计算后得到的 SJR_i 。

这样,(12)式改写成

$$\begin{aligned} \pi^{(k+1)} &= \frac{1-\alpha-\beta}{N} e + \beta \alpha + \alpha H \pi^{(k)} \frac{1-d\pi^{(k)}}{e^T H \pi^{(k)}} + \alpha d \pi^{(k)} \alpha \\ &= \frac{1-\alpha-\beta}{N} e + \alpha H \pi^{(k)} \frac{1-d\pi^{(k)}}{e^T H \pi^{(k)}} + \alpha (ad \pi^{(k)} + \beta) \end{aligned} \quad (14)$$

(14)式中 e 为元素全为 1 的 N 维列向量。

经过以上的分析及转换,(5)式转化成(12)式继而转化成(14)式,这三式是完全相同的。为了便于



比较 SJR 算法与 EF 算法,这里更换了 SJR 算法中常数的符号表示并引入了向量与矩阵的运算。下面要探讨的是在什么条件下 SJR 算法才能转化成 EF 算法。在(14)式基础上,如果对 SJR 算法做以下调整:

1)令(14)式中的 α 等于 EF 算法中的参数 α ,并令 $\beta=1-\alpha$,此时 $\alpha=0.85, \beta=0.15$ 且 $\frac{1-\alpha-\beta}{N}=0$

也就是说,在 SJR 算法中把第一项 $\frac{1}{N}$ 的权重由 0.05 调整为 0,第三项文章向量的权重由 0.1 调整为 0.15,第二项规范化互引矩阵所在项的权重保持不变仍为 0.85。此时(14)式改写为

$$\pi^{(k+1)} = \alpha H \pi^{(k)} \frac{1-d\pi^{(k)}}{e^T H \pi^{(k)}} + a(ad\pi^{(k)} + 1 - \alpha) \quad (15)$$

2)令非悬点期刊 j 的参考文献数只包含对统计源期刊的引用数,而不是其所有的引用数。亦即(13)式矩阵 H 中非悬点列中的 $M_j = \sum_{i=1}^N M_{ij}$, 则有

$$e^T H \pi^{(k)} = (1 \quad 1 \quad \dots \quad 1) \begin{pmatrix} M_{11} & M_{12} & \dots & M_{1N} \\ M_1 & M_2 & \dots & M_N \\ M_{21} & M_{22} & \dots & M_{2N} \\ M_1 & M_2 & \dots & M_N \\ \dots & \dots & \dots & \dots \\ M_{N1} & M_{N2} & \dots & M_{NN} \\ M_1 & M_2 & \dots & M_N \end{pmatrix} \begin{pmatrix} \pi_1^{(k)} \\ \pi_2^{(k)} \\ \vdots \\ \pi_N^{(k)} \end{pmatrix} = \left(\sum_{i=1}^N \frac{M_{i1}}{M_1} \quad \sum_{i=1}^N \frac{M_{i2}}{M_2} \quad \dots \quad \sum_{i=1}^N \frac{M_{iN}}{M_N} \right) \begin{pmatrix} \pi_1^{(k)} \\ \pi_2^{(k)} \\ \vdots \\ \pi_N^{(k)} \end{pmatrix} \quad (16)$$

如果期刊 j 为悬点,则向量 $(\sum_{i=1}^N \frac{M_{i1}}{M_1} \quad \sum_{i=1}^N \frac{M_{i2}}{M_2} \quad \dots \quad \sum_{i=1}^N \frac{M_{iN}}{M_N})$ 中第 j 个分量为 0; 如果期刊 j 不是悬点,则向量 $(\sum_{i=1}^N \frac{M_{i1}}{M_1} \quad \sum_{i=1}^N \frac{M_{i2}}{M_2} \quad \dots \quad \sum_{i=1}^N \frac{M_{iN}}{M_N})$ 中第 j 个分量为 $\sum_{i=1}^N \frac{M_{ij}}{M_i}$ 。由于满足 $M_j = \sum_{i=1}^N M_{ij}$, 所以

$\sum_{i=1}^N \frac{M_{ij}}{M_j} = 1$ 。又因为 $\sum_{i=1}^N \pi_1^{(k)} = 1$, 所以(16)式可转化为

$$e^T H \pi^{(k)} = 1 - \left(\sum_{j \in \{\text{Dangling nodes}\}} \pi_j \right) = 1 - d\pi^{(k)} \quad (17)$$

其中, d 为悬点向量。在满足以上两个条件的情况下,(15)式可转化为:

$$\pi^{(k+1)} = \alpha H \pi^{(k)} + a(ad\pi^{(k)} + 1 - \alpha) \quad (18)$$

不难发现:计算 SJR 值时,适当调整一下权重,并且令期刊 j (非悬点)的参考文献数只包含对统计源期刊的引用数,计算 EF 指标的迭代表达式(1)与计算 SJR 指标的迭代表达式(18)是完全一样的。当然,这只是形式上的一样,实际上(1)式和(18)式中的矩阵 H 是不一样的,因为计算 EF 时去除了自引,而在计算 SJR 时没有去除自引。如果在 SJR 的计算过程中也去除自引,则(1)式和(18)式中的矩阵 H 完全相同,其迭代表达式也完全一样。

在经过以上调整的同时,如果 SJR 算法与 EF 算法又选取了相同的来源数据与统计年限窗口,则这两种算法的迭代表达式和迭代结果都是相同的。

3 实证检验

从以上分析可以看出,EF 算法与 SJR 算法存在着高度的相似性,而且在满足一定条件下,两者的迭代过程是完全一样的。可以预见 EF 算法与 SJR 算法在进行期刊评价时其结果具有高度的相关性。为此,笔者选取 2008—2009 年 CSSCI 收录图书情报学期刊 2007 年度互引矩阵^[15],并实现了 EF 和 SJR 算法,在相同统计年限窗口内分别统计被引量、IF、EF、AI、SJR、以及 SJRQ 指标,并分类验证指标间的相关度。从第 4 小节分析可知,从指标间的对应关系上来看,被引量与 EF、SJR 相对应,IF 与 AI、SJRQ 相对应,下面分别对这两组指标考察其相关性。

表 2 共包含 18 种期刊的互引数据,由于 SJR 算法中计算某期刊的引用总数时不仅包含对统计源期刊的引用数,还包含对统计源期刊外的引用数,所以这里把表 2 中前 12 种期刊作为统计源期刊,后 6 种期刊作为统计源外期刊。经过整理可得以下表。



表 2 2008—2009 年 CSSCI 收录图书情报学期刊 2007 年度互引矩阵(排除自引)

被引期刊	引用期刊														文章总数 ①				
	中国图书馆学报	大学图书馆学报	情报学报	图书情报工作	情报理论与实践	图书情报知识	图书馆	现代图书情报技术	情报资料工作	图书馆建设	情报科学	图书馆论坛	图书馆杂志	国家图书馆学刊		图书馆工作与研究	情报杂志	图书馆学研究	图书与情报
中国图书馆学报	0	26	13	71	25	14	51	8	46	18	35	53	34	15	15	50	69	13	799
大学图书馆学报	13	0	5	51	25	7	35	9	29	36	30	50	31	11	15	51	75	24	854
情报学报	26	10	0	63	46	17	4	37	27	1	73	11	3	0	5	80	16	10	768
图书情报工作	18	16	19	0	28	27	15	23	61	25	41	52	24	11	21	78	56	15	2438
情报理论与实践	11	4	30	38	0	3	4	15	25	5	48	23	5	2	14	62	37	5	949
图书情报知识	7	10	4	36	16	0	14	5	26	24	19	26	15	2	13	28	30	13	997
图书馆	18	13	1	23	6	14	0	1	22	21	8	33	24	10	13	11	43	14	1055
现代图书情报技术	8	3	18	35	24	3	6	0	19	4	27	25	13	2	7	39	35	7	1630
情报资料工作	10	5	4	21	13	6	10	1	0	8	33	20	11	5	10	23	22	2	1465
图书馆建设	7	11	1	30	11	8	28	2	21	0	16	47	21	9	18	17	47	6	1469
情报科学	14	11	23	53	35	11	13	10	25	11	0	17	3	2	15	81	43	14	2056
图书馆论坛	17	9	7	32	21	17	46	5	28	60	28	0	27	9	35	32	72	14	2250
图书馆杂志	5	8	2	24	5	6	16	2	20	24	8	28	0	4	13	19	30	13	1688
国家图书馆学刊	4	3	0	5	3	4	5	5	7	10	3	19	6	0	7	5	13	8	515
图书馆工作与研究	3	10	1	17	8	7	18	1	8	18	12	17	15	9	0	9	21	4	1044
情报杂志	8	10	15	53	34	12	15	13	24	12	61	22	11	3	15	0	72	12	3127
图书馆学研究	7	5	0	19	5	2	21	2	16	23	20	30	15	7	14	28	0	8	2013
图书与情报	5	4	0	10	3	5	15	0	9	7	10	12	5	6	5	7	21	0	853

表 3 经过整理的互引矩阵

		引用期刊												被引总数	文章总数
		中国图书馆学报	大学图书馆学报	情报学报	图书情报工作	情报理论与实践	图书情报知识	图书馆	现代图书情报技术	情报资料工作	图书馆建设	情报科学	图书馆论坛		
被引期刊	中国图书馆学报	0	26	13	71	25	14	51	8	46	18	35	53	360	799
	大学图书馆学报	13	0	5	51	25	7	35	9	29	36	30	50	290	854
	情报学报	26	10	0	63	46	17	4	37	27	1	73	11	315	768
	图书情报工作	18	16	19	0	28	27	15	23	61	25	41	52	325	2438
	情报理论与实践	11	4	30	38	0	3	4	15	25	5	48	23	206	949
	图书情报知识	7	10	4	36	16	0	14	5	26	24	19	26	187	997
	图书馆	18	13	1	23	6	14	0	1	22	21	8	33	160	1055
	现代图书情报技术	8	3	18	35	24	3	6	0	19	4	27	25	172	1630
	情报资料工作	10	5	4	21	13	6	10	1	0	8	33	20	131	1465
	图书馆建设	7	11	1	30	11	8	28	2	21	0	16	47	182	1469
	情报科学	14	11	23	53	35	11	13	10	25	11	0	17	223	2056
	图书馆论坛	17	9	7	32	21	17	46	5	28	60	28	0	270	2250
	图书馆杂志	5	8	2	24	5	6	16	2	20	24	8	28		
	国家图书馆学刊	4	3	0	5	3	4	5	5	7	10	3	19		
	图书馆工作与研究	3	10	1	17	8	7	18	1	8	18	12	17		
情报杂志	8	10	15	53	34	12	15	13	24	12	61	22			
图书馆学研究	7	5	0	19	5	2	21	2	16	23	20	30			
图书与情报	5	4	0	10	3	5	15	0	9	7	10	12			
引用总数 1		149	118	125	453	250	127	226	116	329	213	358	357		
引用总数 2		181	158	143	581	308	163	316	139	413	307	472	485		



其中,“引用总数 1”只统计了表 2 中引用期刊对前 12 种被引期刊的引用总数,供计算 EF 使用;“引用总数 2”统计了表 2 中引用期刊对所有 18 种被引期刊的引用总数,供计算 SJR 使用。

根据表 3 的数据,计算第一类期刊评价指标:被引量、 π^* 、EF、SJR 及其排名。这里之所以把计算 EF 值时的 π^* 也考虑进去,是因为 π^* 不仅与 SJR 值一样是迭代计算的直接结果,而且 EF 值也是由 π^* 经过线性变换而得,因此 π^* 也可作为被引量、EF 与 SJR 的同一类指标。经过计算可得。

表 4 被引量、 π^* 、EF、SJR 计算结果及排名

序号	期刊名	被引量	被引量排名	π^*	π^* 排名	EF	EF排名	SJR	SJR排名
1	中国图书馆学报	360	1	0.108127	2	11.878	2	0.107951	3
2	大学图书馆学报	290	4	0.084734	6	9.0679	5	0.084607	6
3	情报学报	315	3	0.105082	3	11.5525	3	0.108801	2
4	图书情报工作	325	2	0.125585	1	12.2031	1	0.123493	1
5	情报理论与实践	206	7	0.078521	7	8.2367	7	0.082670	7
6	图书情报知识	187	8	0.060622	10	6.0803	9	0.060699	10
7	图书馆	160	11	0.060303	11	5.9816	10	0.060054	11
8	现代图书情报技术	172	10	0.067058	8	6.1699	8	0.067875	8
9	情报资料工作	131	12	0.052981	12	4.6877	12	0.052427	12
10	图书馆建设	182	9	0.063084	9	5.8721	11	0.061242	9
11	情报科学	223	6	0.094588	5	8.9593	6	0.095131	4
12	图书馆论坛	270	5	0.099315	4	9.3107	4	0.095049	5

根据表 3 的数据,计算第二类期刊评价指标:IF、AI、SJRQ 及其排名,可得。

表 5 IF、AI、SJRQ 计算结果及排名

序号	期刊名	IF	IF排名	AI	AI排名	SJRQ	SJRQ排名
1	中国图书馆学报	0.450563	1	2.4871	2	0.000135	2
2	大学图书馆学报	0.339578	3	1.7764	3	0.000099	3
3	情报学报	0.410156	2	2.5166	1	0.000142	1
4	图书情报工作	0.133306	7	0.8374	7	0.000051	7
5	情报理论与实践	0.217071	4	1.4521	4	0.000087	4
6	图书情报知识	0.187563	5	1.0203	5	0.000061	5
7	图书馆	0.151659	6	0.9486	6	0.000057	6
8	现代图书情报技术	0.105521	11	0.6333	11	0.000042	11
9	情报资料工作	0.089420	12	0.5353	12	0.000036	12
10	图书馆建设	0.123894	8	0.6688	10	0.000042	10
11	情报科学	0.108463	10	0.729	8	0.000046	8
12	图书馆论坛	0.120000	9	0.6923	9	0.000042	9

需要说明的是,这里的各指标值都是在相同的统计源期刊和相同的统计年限窗口下计算的,实际上它们之间的统计源期刊和统计年限窗口不尽相同,本文统一了统计源期刊和统计年限窗口的目的是为了尽可能探索这些期刊评价指标之间的相关度。

通过表 4 与表 5 的排名来看,被引量、 π^* 、EF 及 SJR 之间具有极强的相关性,IF、AI 及 SJRQ 之间也具有极强的相关性。为了进一步说明它们之间的相关性,笔者利用 SPSS 统计分析软件,进行相关度分析(Spearman 相关系数),其结果如表 6、7、8、9 所示。

表 6 被引量、 π^* 、EF 及 SJR 值之间的相关性分析表

		相关系数			
		被引量	π^*	EF	SJR
Spearman 的 rho	被引量	1.000	.944**	.968**	.923**
	相关系数		.000	.000	.000
	Sig. (双侧)		.12	.12	.12
	N	12	12	12	12
π^*	被引量	.944**	1.000	.968**	.986**
	相关系数			.000	.000
	Sig. (双侧)			.12	.12
	N	12	12	12	12
EF	被引量	.968**	.968**	1.000	.968**
	相关系数				.000
	Sig. (双侧)				.12
	N	12	12	12	12
SJR	被引量	.923**	.986**	.968**	1.000
	相关系数				.000
	Sig. (双侧)				.12
	N	12	12	12	12

**在置信度(双侧)为 0.01 时,相关性是显著的。

表 7 被引量、 π^* 、EF 及 SJR 排名之间的相关性分析表

		相关系数			
		被引量排名	π^* 排名	EF排名	SJR排名
Spearman 的 rho	被引量排名	1.000	.944**	.951**	.923**
	相关系数		.000	.000	.000
	Sig. (双侧)		.12	.12	.12
	N	12	12	12	12
π^* 排名	被引量排名	.944**	1.000	.972**	.986**
	相关系数			.000	.000
	Sig. (双侧)			.12	.12
	N	12	12	12	12
EF排名	被引量排名	.951**	.972**	1.000	.951**
	相关系数				.000
	Sig. (双侧)				.12
	N	12	12	12	12
SJR排名	被引量排名	.923**	.986**	.951**	1.000
	相关系数				.000
	Sig. (双侧)				.12
	N	12	12	12	12

**在置信度(双侧)为 0.01 时,相关性是显著的。

表 8 IF、AI 及 SJRQ 值之间的相关性分析表

		相关系数		
		IF	AI	SJRQ
Spearman 的 rho	IF	1.000	.965**	.958**
	相关系数		.000	.000
	Sig. (双侧)		.12	.12
	N	12	12	12
AI	IF	.965**	1.000	.993**
	相关系数			.000
	Sig. (双侧)			.12
	N	12	12	12
SJRQ	IF	.958**	.993**	1.000
	相关系数			.000
	Sig. (双侧)			.12
	N	12	12	12

**在置信度(双侧)为 0.01 时,相关性是显著的。



表 9 IF、AI 及 SJRQ 排名之间的相关性分析表

		相关系数			
Spearman 的 rho	IF 排名	AI 排名	SJRQ 排名		
Spearman 的 rho	IF 排名	1.000	.965**	.965**	
	相关系数		.000	.000	
	Sig. (双侧)		.000	.000	
	N	12	12	12	
AI 排名	AI 排名	.965**	1.000	1.000**	
	相关系数				
	Sig. (双侧)				
	N	12	12	12	
SJRQ 排名	SJRQ 排名	.965**	1.000**	1.000	
	相关系数				
	Sig. (双侧)				
	N	12	12	12	

** 在置信度 (双侧) 为 0.01 时, 相关性是显著的。

从表 6、7、8、9 可以清楚地看出被引量、 π^* 、EF 及 SJR 两两之间无论是数值还是排名的相关度都在 0.9 以上, 是显著相关的; IF、AI 与 SJRQ 两两之间无论是数值还是排名的相关度也都在 0.9 以上, 也是显著相关的。通过相关度分析, 从另一侧面说明了基于 PageRank 算法的 EF 算法与 SJR 算法大同小异, 存在着高度的相似性。

4 讨论与小结

经由前面的理论分析及算法推导, 如果在计算 SJR 指标时:

- 1) 调整权重令 $\alpha + \beta = d + e = 1$ 。
- 2) 期刊 j 的参考文献数只包含对统计源期刊的引用数, 而不是其所有的引用数。
- 3) 去除自引因素的影响。

那么, 计算 EF 指标的迭代表达式 (1) 和计算 SJR 指标的迭代表达式 (18) 从形式和意义上是完全一样的。

同时, 根据表 1 所列 EF 算法与 SJR 算法之间的差别, 如果再满足以下两个条件:

- 4) 统计年限窗口一样。
- 5) 统计源期刊一样。

那么计算 EF 和计算 SJR 的迭代过程和结果是完全一样的。当然, 只是“迭代”的过程和结果完全一样, 并不是最终的 EF 和 SJR 值完全一样。从 SJR 算法的定义得 $SJR = \pi^*$, 而 EF 定义成 (2) 式

$$EF = 100 \frac{H\pi^*}{\sum_i [H\pi^*]_i}$$

那么, 为什么 EF 算法不像 SJR 一样直接采用 π^* 呢? 应该是出于以下几点考虑:

- 1) EF 算法的迭代过程中用文章向量替代 H' 矩阵悬点列放大了悬点的作用, H 是替代前的包含悬点的规范化互引矩阵, 让 H 点乘 π^* 可部分抵消这

种替代的影响。

2) 同样, (2) 式分子中没有加上文章向量贡献的权重 $(1 - \alpha)a$ 也是出于部分抵消迭代过程中文章向量所给予的初始值的影响。

3) (2) 式中分母乘以 100 是为了换算成百分值; (2) 式中分母的引入是出于规范化的目的: 由于 $\sum_i [H\pi^*]_i = 1 - d\pi^*$, d 为悬点向量, 所以, 如果存在悬点, 则 $\sum_i [H\pi^*]_i < 1$ 。分母的引入保证了 $\sum_i EF = 100$ 。

当然, EF 算法中在得出 π^* 后从 π^* 到 EF 转换相较于 SJR 算法直接采用 π^* , 这两者究竟哪个更合理, 则超出了本文的讨论范围, 有兴趣的研究者可以继续研究下去。

因此, 虽然期刊评价指标 EF 和 SJR 之间存在着一些区别, 这些区别主要表现在计算方法与来源数据两个方面: 从计算方法上看, 主要涉及引用关系的权重、处理自引的方式以及引用论文的统计方式; 从来源数据上看, SJR 指标以 Scopus 数据库的数据为基础, EF 以 Web of Science 的数据为基础, SJR 指标的统计窗口年限为 3 年, EF 指标的统计窗口年限为 5 年。但是, 由于 EF 和 SJR 都是基于 PageRank 算法, 两者是一脉相承的, 在满足一定的条件下, SJR 算法的迭代式与 EF 算法的迭代式从形式和意义上完全一样的。

通过实证研究, 发现被引量、 π^* 、EF 及 SJR 两两之间是显著相关的, IF、AI 与 SJRQ 两两之间也是显著相关的。传统期刊指标被引量与影响因子虽然存在着一些缺陷, 但在实践上仍存在着相当的合理性。EF 算法、SJR 算法作为新兴算法, 克服了被引量与影响因子只考虑引文数量忽略引文质量的缺陷, 也完全可以作为被引量与影响因子的参考、补充甚至替代。

参考文献

- 1 Garfield E. Citation analysis as a tool in journal evaluation[J]. Science, 1972, 178(4060): 471-479.
- 2 Henk F, Moed. Citation analysis of scientific journals and journal impact measures[J]. Current Science, 2005, 98(12): 1990-1996.
- 3 Hirsch J E. An index to quantify an individual's scientific output[J]. Proceedings of the National Academy of Sciences of the United States of America, 2005, 102(46): 16569-16572.
- 4 Egghe L. How to improve the h-index[J]. Scientist, 2006, 20(3): 15-15.
- 5 Jin B H et al. The R- and AR-indices: Complementing the h-index[J]. Chinese Science Bulletin, 2007, 52(6): 855-863.



- 6 叶鹰. H 指数和类 h 指数的机理分析与实证研究导引[J]. 大学图书馆学报, 2007(5): 2-5.
- 7 Butler D. Free Journal-ranking Tool Enters Citation Market[J]. Nature, 2008, 451(7174): 6.
- 8 Henk F Moed. Measuring Contextual Citation Impact of Scientific Journals[J]. Journal of Informetrics, 2010(4): 265-277.
- 9 Bergstrom C T., West J D. et al. The Eigenfactor Metrics[J]. The Journal of Neuroscience, 2008, 28(45): 11433-11433.
- 10 Guerrero-Bote V P, Moya-Anegón F. A further step forward in measuring journals' scientific prestige: The SJR2 indicator[J]. Journal of Informetrics, 2012, 6(4): 674-688.
- 11 Ludo Waltman, Nees Jan van Eck, Thed N. van Leeuwen, et al. Some modifications to the SNIP journal impact indicator[J]. Journal of Informetrics, 2013, 7(2): 272-285.
- 12 Thomson Reuters. Normalized eigenfactor © score[EB/OL]. [2017-11-03] <http://ipsience-help.thomsonreuters.com/in-citesLiveJCR/glossaryAZgroup/g9/9709-TRS.html>.
- 13 Thomson Reuters. Journal impact factor percentile[EB/OL]. [2017-11-03] <http://ipsience-help.thomsonreuters.com/in-citesLiveJCR/glossaryAZgroup/g8/9586-TRS.html>.
- 14 杨康, 刘明政, 张旭. SJR 指数研究及其与影响因子的比较分析[J]. 情报杂志, 2009, 28(11): 27-30.
- 15 米佳, 濮德敏. 特征因子原理及实证研究[J]. 大学图书馆学报, 2009, 27(6): 63-68.
- 16 田质兵. 评价期刊影响力的三项指标比较研究[J]. 现代情报, 2010, 30(9): 141-143, 146.
- 17 EIGENFACTOR. org. Eigenfactor™ score and article influence™ score; detailed methods[EB/OL]. [2017-11-03] <http://www.eigenfactor.org/methods.pdf>.
- 18 SCImago Research Group. Description of scimago journal rank indicator[EB/OL]. [2017-11-03] <http://www.scimagojr.com/SCImagoJournalRank.pdf>.

作者单位: 北京大学图书馆, 北京, 100871

收稿日期: 2017 年 12 月 18 日

Comparison and Empirical Research between the Algorithms of Eigenfactor and SJR Indicator

Zhuang Jilin

Abstract: Given certain conditions, the transformation from SJR algorithm to Eigenfactor algorithm is realized. The mutual citation matrix of Library and information science journals in CSSCI in 2007 is selected to calculate the citation frequency, EF, SJR, IF, AI and SJRQ respectively, and the correlation analysis is divided into two categories. The results show that, under certain conditions, the SJR algorithm can be transformed into Eigenfactor algorithm, and there is a significant correlation between the indicator of EF and SJR, and they can be used as reference, supplement, and even replacement of the indicator of the citation frequency and IF.

Keywords: Eigenfactor; Article Influence Score; SJR; SJRQ; Impact Factor; Journal Assessment; Citation Analysis

(接第 61 页)

Thoughts on the Application of Electronic Resource Management System

—Taking Xiamen University Libraries as an Example

Chen Juan Xiao Dehong

Abstract: Electronic resource management (ERM) is a challenging task of libraries. The implementation and application of an Electronic resource Management Systems (ERMS) that meets requirements of libraries will improve ERM efficiencies. ERMS development is relatively slow in China as both Integrated Library System (ILS) vendors and digital content providers in China haven't released ERMS products based on digital resource lifecycles. While the high costs of subscription and maintenance of imported ERMS products cause difficulties and problems. This article introduces the functions and features of ERMS developed by Xiamen University Libraries (XMUL), and aims to offer references for academic libraries, ILS vendors, as well as digital content providers in China. It provides a brief introduction of the functions and development of ERMS in China and at abroad before illustrates function modules, implementation and application situations of ERMS. ERMS developed by XMUL fundamentally changes previous way of digital resource management in Xiamen University to a way of streamlined, standardized, electronic, automatic and systematic. As next generation library service platforms development is emerging and needs time to become mature, it is meaningful to have standalone ERMS before it is replaced by new generation systems in China.

Keywords: Electronic Resource Management; Electronic Resources Management Systems; Digital Resources; Xiamen University Libraries