Available online at www.sciencedirect.com
ScienceDirect
IERI Procedia 10 (2014) 57 – 62
Procedia IERI
www.elsevier.com/locate/procedia

`2014 International Conference on Future Information Engineering

Ranking of Journals in Science and Technology Domain: a Novel and Computationally Lightweight Approach

Neelam Jangid, Snehanshu Saha, Siddhant Gupta, Mukunda Rao J*

Dept of CSE and CBIMMC,PESIT Bangalore South Campus-560100, Library Information Science*
neelu.jangid88@gmail.com , snehanshusaha@pes.edu , sidpro.pesit@gmail.com , pesselibrarian@pes.edu

Abstract

In this paper, a regression analysis based method is proposed to calculate the Journal Influence Score. This Influence Score is used to measure the scientific influence of scholarly journals. Journal Influence Score is calculated by using various factors in a weighted manner. The Score is then compared with the SCImago Journal Score. The results show that the error is small between the existing and proposed methods, proving that the model is a feasible and effective way of calculating scientific impact of journals.

Selection and peer review under responsibility of Information Engineering Research Institute

Keywords: Multiple linear regression, scientific journal, Journal influence score, SCImago journal score, Quartile matching

Selection and peer review under responsibility of Information Engineering Research Institute
doi:10.1016/j.ieri.2014.09.091

Introduction

Ranking of journals has been one of the core and frequent topics of recent research due to academic reward system. Many ranking methods have been undertaken by organizations using different parameters and we propose an alternate algorithmic method for ranking of journals using "multiple linear regression model". We then compare our approach with the SCImago approach of journal ranking and find the match of the samples that lie in the same quarter.

Journal citation report (JCR) draws its data from Web of Science and is mainly based on citations and impact factor. But now in addition to that there is Scopus SJR, Google Scholar etc. Both use different methods for evaluation of journals [1]. A Journal is a central object in library science as well as bibliometrics and its visibility through impact is a major concern [2].

It is quite possible that a non-international journal may publish very advanced and innovative research [3]. Buela-Casal [4] concludes that a combination of suitably weighted criteria is the only way to quantify the degree of internationality. Our definition of "internationality" is based on the influence of a journal. We define and measure an "influence score" based on several scholastic parameters and the journals are ranked based on their calculated influence scores.

Different organizations use different parameters such as impact factor, citations, self-citations, percentage international collaboration, cites per doc etc. to estimate rank of journals. However there is not much correlation among different journal ranking methods. For calculation of impact factor/journal influence score, JCR uses citations of two preceding years where as Scimago uses the citations of past three years. Most of the journal ranking approaches are based on single measure of either citations or perceptions. If better results are obtained, one should combine these approaches for the ranking process. There is no guarantee that a top journal is a leading journal and hence application of multi-item measures is required for ranking of journals [5].Gaby Haddow and Paul Genoni [6], performed citation analysis of Australian Social Science journals and found that Scopus database provides higher citations for many of the journals.

The lack of a standard classification scheme used for the journals indexed in WoS and Scopus makes it difficult to compare the performance of all journal titles covered by both databases as the indices used to rank journals are different.

Nomenclature

SJR	SCImago Journal Rank ;	JCR	Journal Citation Report

1. Related Work

The Scimago Journal Rank (SJR) involves transfer of "prestige" (impact) from one journal to another. This is executed through the references journals share. The final prestige of a journal is computed iteratively, where the prestige in stage 'i' of a journal depends on the prestige of journal set in stage 'i-1'.

$$SJR_i = \frac{(1-d-e)}{N} + e.\frac{Art_i}{\sum_{j=1}^{N} Art_j} + d.\sum_{j=1}^{N}\frac{C_{ji}.SJR_j}{C_j}.\frac{1-\left(\sum_{k\in\{Dangling-\ nodes\}} SJR_k\right)}{\sum_{h=1}^{N}\sum_{k=1}^{N}\frac{C_{kh}SJR_k}{C_k}} + d.\left[\sum_{k\in\{Dangling-\ nodes\}} SJR_k\right].\frac{Art_i}{\sum_{j=1}^{N} Art_j} \quad (1)$$

$$\mathrm{SJRQ_i} = \frac{\mathrm{SJR_i}}{\mathrm{Art_i}} \quad (2)$$

where,
SJR_i -SCImago Journal Rank of Journal i.
C_{ji} - Citation from journal j to journal i.
C_j - Number of References of journal j.
d - Constant, normally 0.85.
e - Constant, normally 0.10.
N - Number of Journals
Art_j - Number of Articles of journal j
The calculation of the SJR involves three stages:

1) Initialization of SJR: in this stage a default prestige is assigned to every journal.
2) Iterative process of calculation: starting from stage 1, the computation is repeated to calculate the prestige of each journal based on the prestige transferred by the rest. The process ends when the variation of the SJR between two iterations is less than a threshold set initially. The final result is the SJR of each journal.
3) Computation of SJRQ: SJR value is divided by the number of articles published in the citation window to obtain the SJRQ indicator. The result is the average prestige per article.

SJR uses Google Page Rank Algorithm [7] to implement this approach.

Limitations of SJR:

- SJR is a computationally heavy process.
- More factors are needed.
- Data availability and storage.

Impact of this is:

- Evaluation cost will increase.
- Process is iterative so if the initial choice is wrong, it will take more time to calculate the SJR.

2. Our work

Our work is motivated by the thought of replacing the iterative process of journal classification with an alternate model that assigns weight to each factor based on the percentage of influence the factor might have in the calculation of the influence score. These weights can then be used directly to compute the score of a new journal without using the prestige of any other journal. This is a computationally lightweight scheme that does not require any data storage.

2.1 Model structure

A multiple linear regression model is used to predict a response variable **y**, as a function of k predictor variables $x_1,x_2,.....,x_k$. This is of the following form:

$$y = b_0 + b_1x_1 + b_2x_2 + \cdots + b_kx_k + e \qquad (3)$$

Where $b_0,b_1,...,b_k$ are fixed parameters that signify the weight of factors and e is the residual error.
Given a number of observations, the model consists of the following n equations:

$$y_1 = b_0 + b_1x_{11} + b_2x_{21} + \cdots + b_kx_{k1} + e_1 \qquad (4)$$

$$y_2 = b_0 + b_1x_{12} + b_2x_{22} + \cdots + b_kx_{k2} + e_2 \qquad (5)$$

$$\vdots$$

$$y_n = b_0 + b_1x_{1n} + b_2x_{2n} + \cdots + b_kx_{kn} + e_n \qquad (6)$$

This leads to a matrix system:

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} \cdots & x_{k1} \\ \vdots & \vdots & \vdots \\ 1 & x_{1n} \cdots & x_{kn} \end{bmatrix} \begin{bmatrix} b_0 \\ \vdots \\ b_k \end{bmatrix} + \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix} \qquad (7)$$

$$\text{i.e. } \mathbf{y}=\mathbf{Xb}+\mathbf{e} \quad (8)$$

Where $\boldsymbol{b} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}(\boldsymbol{X}^T\boldsymbol{y})$

Our source data for this study were SCImago Journal and Country Rank's portal which contained journals in Elsevier's Scopus.
http://www.scimagojr.com/journalrank.php?area=0&category=1706&country=all&year=2012&order=sjr&min=0&min_type=cd

The input parameters are: Quarter, H-Index, Total Docs 2012, Total Docs 3yrs, Total Cites 3yrs, Citable Docs 3yrs, Ref/Doc, Cites/Doc 3yrs and Total Ref , where

$$\text{Quarter(Probability of Influence)} = \frac{Q_i}{\sum Q_i}\text{ ; where i = 1, … ,4} \quad (9)$$

The probability of influence in our sample data validates the use of quarters in our model. Any journal to be evaluated in the first Quarter of the year has more probability of having greater influence, considering the number of publications is mostly limited. Hence the "quarter" of publication should be statistically significant. The model depends on a procedure of weighted influence score calculation, where the weights are assigned to every factor. This helps us to bypass the iterative procedure of SJR impact factors, hence reducing the time complexity of the model, while maintaining "reasonable" levels of accuracy and precision.

Table 1 Comparison between SJR model & proposed model

Basis of Comparison	SJR Model	Proposed Model
No. of Factors	13	9
Procedure	Iterative	Weighted
Expected Time Complexity	More	Less
Algorithm	Google Page Rank Algo	Weight Based on Regression
Database Size	Huge	Insignificant
Historical data	3-5 years	Not an absolute must

2.2 Algorithm

Step 1: Extract data from SJR database.
Step 2: Compute the corresponding quarter values.
Step 3: Compute the weight vector $\boldsymbol{b}$ for factors

$$\boldsymbol{b} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}(\boldsymbol{X}^T\boldsymbol{y}) \quad (10)$$

Step 4: Calculate the quartile match
4.1: Check the samples in each quarter.
4.2: Compare these samples with same number of samples from the results of our model for each quartile.
4.3: Calculate the percentage of match.

3. Results

The fitting model assumes the form:

*Journal Influence Score = 0.763897 - (0.192 * Quarter) + (0.001772 * H index) + (0.000643 * Total Documents For Current Year) - (9.4E-05 * Total Doc. in previous 3years) - (2.7E-05 * Total References) + (0.000194 * Total Cites in previous 3 years) - (0.00042 * Citable Documents in previous 3 years) + (0.260029 * Cites / Doc. in previous 2 years) + (Ref. / Doc. * 0.000262)* (11)

The above equation shows that the journal base score is 0.763897 which explains the initial score of each journal, affected further by the factor values of that journal. The other values in the equation signify the weight to their corresponding factors.

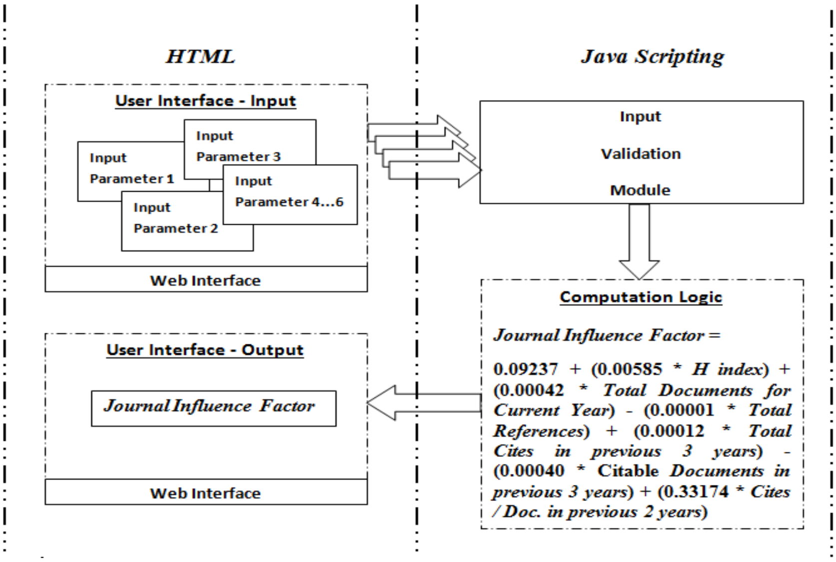

Fig 1.a - High level design of Journal Influence Score calculator

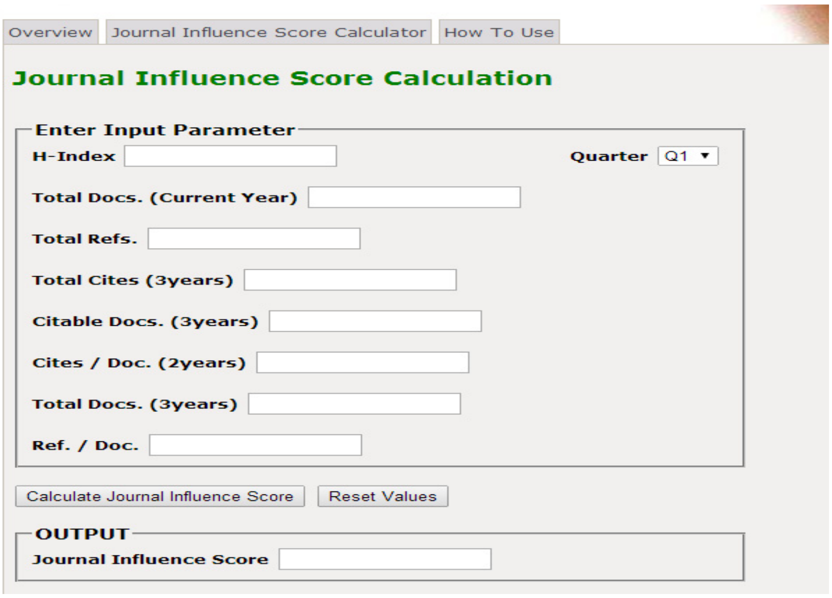

Fig 1.b - Snapshot of Web based Tool

Link to Web Based Tool

http://www.4shared.com/folder/sj31vMfu/_online.html

Table 2 Percentage matching between SJR model & proposed model

Quartile	Percentage of matching
Q1	89
Q2	87.8
Q3	91.23
Q4	94.6

The computed errors in the proposed model reflect reasonable level of accuracy in all the quarters possible. This comparison is made with the SJR data. Hence the model performs well in all possible cases and quarters.

Conclusion

The results show that error is minimal. The model performs reasonably well and can act as a substitute to the SJR method. This approach may be extended to a universal weight vector for all subject categories rather than evaluating separate weights for each. This might increase the percentage of errors in the model as the weights vary from categorically, but the approach would produce generic results.

SJR is currently the benchmark for ranking journals. The algorithm presented in the paper replicates and validates the ranking structure of SJR. The models portray significantly high accuracy levels as each quartile match varies from 87% - 95%. This is achieved without any iterative approach and/or requirements of data storage. Considering the fact that SJR is one of the most used standards but not the absolute ranking scheme, the regression approach could serve as a replacement for instantaneous journal ranking, a problem which SCOPUS can't handle.

References

[1] Abrizah; A.N. Zainab; K. Kiran; R.G. Raj LIS journals scientific impact and subject categorization: a comparison between Web of Science and Scopus Scientometrics (2013) 94:721–740 DOI 10.1007/s11192-012-0813-7
[2] Garfield E, Citation analysis as a tool in journal evaluation Science, 178 (4060) (1972) 471-479
[3] Chiang Kao, The authorship and internationality of industrial engineering journals, Scientometrics, 80 (3) (2009) 123-136.
[4]Gualberto Buela-Casal, Pandelis Perkakis, Michael Taylor and Purificacion Checha, Measuring internationality: reflections and perspectives on academic journals, Scientometrics, 67 (1) (2006) 45-65.
[5] Goran Svensson, Scholarly journal ranking(s) in marketing: single- or multi-item measures?, Marketing intelligence and planning, 26 (4) (2008) 340-352.
[6] Gaby Haddow; Paul Genoni, Citation analysis and peer ranking of Australian social science journals, Scientometrics, 85 (2) (2010) 471–487.
[7] Langville, A. N., & Meyer, C. D. (2006). Google's PageRank and Beyond: The Science of Search Engine Rankings. New Jersey: Princeton University Press
[8]Raj Jain,The art of Computer System Performance Analysis, Edition 2009 , Wiley-India