Information Sciences 180 (2010) 2156–2175

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

A non-linear index to evaluate a journal's scientific impact

Sozon Papavlasopoulos [a], Marios Poulos [a,*], Nikolaos Korfiatis [b], George Bokos [a]

[a] Information Technology Laboratory, Dept. of Archives and Library Sciences, Ionian University, Ioannou Theotoki 9, GR-49100, Corfu, Greece
[b] Department of Economics, University of Copenhagen, Øster Farimagsgade 5, Copenhagen, Denmark

ARTICLE INFO

Article history:
Received 4 December 2008
Received in revised form 29 November 2009
Accepted 14 January 2010

Keywords:
Bibliometrics
Semantic classification
Elman neural network
Impact factor

ABSTRACT

The purpose of this study is to define a bibliometric indicator of the scientific impact of a journal, which combines objectivity with the ability to bridge many different bibliometric factors and in particular the side factors presented along with celebrated ISI impact factor. The particular goal is to determine a standard threshold value in which an independent self-organizing system will decide the correlation between this value and the impact factor of a journal. We name this factor "Cited Distance Factor (CDF)" and it is extracted via a well-fitted, recurrent Elman neural network. For a case study of this implementation we used a dataset of all journals of cell biology, ranking them according to the impact factor from the Web of Science Database and then comparing the rank according to the cited distance. For clarity reasons we also compare the cited distance factor with already known measures and especially with the recently introduced eigenfactor of the institute of scientific information (ISI).

1. Introduction

1.1. Background and motivation

Ever since the initial celebrated work by Garfield [13], Garfield and Merton [14], Pinski and Narin [33] on the evaluation of a scientific impact of a scientific journal, a great body of research has emerged on the application of information processing methods for evaluating scientific publication venues, extending it also to the evaluation of an individuals' research output [20].

Nonetheless, the issue concerning the evaluation of a journals' scientific impact still remains the essential priority of the scientometrics field [4,37,22] due to the fact that is often used as a yardstick to provide an indication for the allocation of scientific budgets, the direction and future of research, as well as organizational decisions such as the employment of the researchers, the effectiveness of the research policy pursued and the subscription policy of academic libraries. The journal citation reports (JCR) provided by the Institute for Scientific Information (ISI), instituted by the work of Garfield [12] are often the main source for these indicators to academic and research evaluation committees. Undoubtedly, in current academic practice, JCR is one of the most used sources for facilitating a researcher's access to high-quality, latter-day research.[1]

* Corresponding author.
E-mail addresses: sozon@ionio.gr (S. Papavlasopoulos), mpoulos@ionio.gr (M. Poulos), nikolaos.korfiatis@econ.ku.dk (N. Korfiatis), gbokos@ionio.gr (G. Bokos).

[1] With Elsevier's SCOPUS and other products such as NCBI Pubmed to receive also significant attention from researchers. Several disciplines such as Physics have advanced the evaluation of publications to an individual level such as for example the SPIRES research database hosted at Stanford Linear Accelerator Center (SLAC).

0020-0255/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ins.2010.01.018

Apart from the impact factor, ISI provides a set of journal performance indicators supplied along with the impact factor of the journal. These performance indicators are categorized as follows: (a) Impact Factor (IF), a measure of the frequency with which the average article in a journal has been cited in a given period of time which is referred to, in a two year time span after publication in other words this cites in year x to items published in: $x-1$ and $x-2$; (b) Immediacy index (I.I) which concerns the average number of times an article is cited in the year it is published; (c) Cited half-life (Cd. H-L), the number of years, going back from the current year, that account for half the total citations received by the cited journal in the current year; (d) Citing half-life (Cg. H-L), the number of years from the current year that account for 50% of the cited references from articles published by a journal in the current year.[2]

1.2. Problems with the impact factor

While undoubtedly the impact factor is accepted as a key indicator for scientific quality there are criticism as to its misuse since it cannot be directly related to an individuals' research output and thus is not always a reliable instrument for measuring the quality of publication venues [38]. Also several attempts made by different authors to count the number of citations, no matter how prestigious the citing journal is [18], or to introduce more sophisticated journal citation measures and the reasons why many indicators aiming at a correction of methodological limitations of the Impact Factor have also made a point to that direction. In particular the shortcomings of the slow citation window and subject biases [16] cannot be tackled by the overall assessment of the impact factor and any attempt to evaluate a journal should take into account these factors and that is in fact an intuition in this study. It is broadly argued that its use for purposes for which it was not intended, causes even greater unfairness [1,21]. Related research also suggests that research evaluation should also be adjusted to account for variables such as domain specialty, citation density, and half-life [42,27]. Furthermore, apart from being non-representative, the journal impact factor is encumbered with several shortcomings of a technical and more fundamental nature such as the intention to cite [28].

To this end, the focus of this study is to provide an index that results as a combination of the different performance indicators that cover a publication period and citation window. These indicators are designed to measure ageing characteristics of subject fields and journal literature, or to help to distinguish between slow and fast reception of scientific information [16]. Therefore the problem which originates from the separation slow and fast reception sciences journal, leads to calculate unequal values for Impact Factor in different scientific areas while they have the same impact in their category. We summarize the aim and scopes of this study in the section that follows.

1.3. Aim and scopes of this study

The major goal of this study, is to provide empirical evidence to support that journal performance indicators, and in particular the indexes provided by Journal Citation Report (JCR): Impact Factor (IF), Immediacy Index (I.I), Cited half-life (Cd-h.l) and Citing half-life (Cg-h.l), play an important role in differentiating the original ranking by the IF when taking also these factors into consideration. To achieve this objective we create an ideal factor namely the Cited Distance factor, which measures the difference between a predetermined value and a representative value for each journal. Thus, the smallest the aforementioned distance is, the higher the ranking of the journal will be. The determination of these values is given in Sections 2.2 and 2.3.

All the journal performance indicators mentioned above have a common property: they evaluate the citations in a different statistical way. The next issue that we address in this study is to identify the statistical procedure that could solve the problem of the combination of these indicators in a single non-linear index for measuring and evaluating publication venues.

We chose to address this issue with a well-defined neural network (NN) in order to provide an appropriate weight in the process of learning for input vectors. This can be judged by the fact that each indicator refers to a different characteristic of the performance of the journal (e.g. citation window) and a simple linear combination of the above cannot provide an accurate picture. Some similar attempts such as the Eigenfactor move to that direction as well [2]. Furthermore the recent introduction by ISI of the Eigenfactor as a measure for evaluating the scientific impact for a journal, justifies the need for a more sophisticated way of combining the different journal performance indicators into a more accurate and holistic measure. The intuition of the Eigenfactor can to some extend be considered similar with other celebrated measures that rely on the calculation of eigenvalues [3] such as for example the celebrated page rank algorithm (where the page rank values are computed from the stationary vector rather than the eigenvector) and to that extend we also examine the relation of the proposed CD factor with other measures and in particular with the Eigenfactor and journal adapted h-index.

The problems which are handled with these vectors are the variability which they present periodically for which they depend in turn on the differentiation of the aforementioned indicators. Furthermore, this NN could be trained to recognize and produce both spatial and temporal patterns which solve the problem of the factors variability using a *threshold* value in order to support rule-based decision process. Thus, the proposed **cited distance factor,**which is extracted via a neural network processing, returns a number which may be considered as an ideal combination of the JCR side factors. In this way, we

[2] http://www.isi.org/ (Accessed 6th June 2009).