Research Journal of Applied Sciences 9 (12): 1167-1170, 2014
ISSN: 1815-932X


# Construction IF-Scoring Rule Within The Framework of New Generation of Metric Citations

[1]Vladimir M. Moskovkin, [2]Nikolay A. Golikov, [1]Ilya F. Isaev and [1]Olesya V. Serkina
[1]Belgorod State University, Pobedy St. 85, 308015 Belgorod, Russia
[2]Independent Researcher, Kharkov, Ukraine

**Abstract:** On the basis of a scoring rule approach, a citation metrics has been developed in this study which takes into account not only the number of articles published by an author and their citations but also the impact factors of journals in which the articles were published and impact factors of journals with articles citing the researcher in question. We have developed a special algorithm and a program on the basis of the Python language to identify the titles of Scopus-journals (so that they could be searched for with the help of Google Scholar) in which an author's articles are published and to determine their IF using SCIMAGO platform. The algorithm and program has been tested on two Google Scholar profiles of the most cited scientists (physicists) of Belgorod State University (Russia) Ruslan Kaibyshev and Andrey Belyakov.

**Key words:** Belgorod State University, bibliometric rankings, citation metrics, Google scholar, h-index, impact factor, python language, scopus-journals, SCIMAGO platform, IF-scoring rule

## INTRODUCTION

Starting with the classical research of Hirsch (2005), there came a boom of modifying h-index and creating similar indicators. As shown by Waltman and van Eck (2012), both in 2010 and 2011 almost one out of four publication in Scientometrics and Journals of Informetrics cited above mentioned Hirsch's work. In the scientific literature, one can find a set of indices (m, g, e, w, $q^2$, hg, etc.) used for the assessment of research productivity in terms of the number of publications and citations. As further noted by Waltman and van Eck (2012), a large part of the literature building on Hirsch's work is concerned with introducing variants, extensions and generalizations of the h-index. Thus, by Bornmann *et al*. (2011) we find a list of no <37 variants of the h-index.

The essence of all research of this kind was well described by Marchant (2009): "many researchers, analyzing previously existing indices, find that they have some drawbacks and then propose an adapted version of the incriminated index or a brand new one, supposedly better than the older one. Unfortunately, the reasoning of the proponents of such new indices is often ad hoc: they propose a new index not suffering the same drawbacks as the older one that they analyzed but nothing guarantees that the new index does not have many other weaknesses".

The imperfection of h-index at a fundamental level is demonstrated in the work of Waltman and van Eck (2012). This index turns out to violate the following three properties:

- If two scientists achieve the same relative performance improvement, their ranking relative to each other should remain unchanged
- If two scientists achieve the same absolute performance improvement, their ranking relative to each other should remain unchanged
- If the scientist $X_1$ is ranked higher than scientist $Y_1$ and scientist $X_2$ is ranked higher than scientist $Y_2$, then a research group consisting of scientists $X_1$ and $X_2$ should be ranked higher than a research group consisting of scientists $Y_1$ and $Y_2$

## MATERIALS AND METHODS

Given the inconsistency problem of the h-index and many other indicators, one may wonder what kind of alternative indicators can be used that does not have a similar problem (Waltman and van Eck, 2012). This problem was solved by Marchant (2009), through introducing the scoring rules (score-based or summation-based rankings (indices)).

**Corresponding Author:** Vladimir M. Moskovkin, Belgorod State University, Pobedy St. 85, 308015 Belgorod, Russia

In the simplest case, to calculate a scoring rule for a set of publications, one first calculates a score for each individual publication in the set. The score of a publication is determined by the number of citations of the publication. After calculating a score for each individual publication, the scoring rule is obtained by calculating the sum of the individual publication scores. Somewhat more formally, given a set of N publications with $C_1$, $C_2$, ..., $C_n$ citations, a scoring rule is equal to:

$$I(C_1, C_2, ..., C_n) = \sum_{i=1}^{n} f(C_i) \quad (1)$$

where, f(C) is an increasing function that determines the score of a publication based on the number of time the publication has been cited (Waltman and van Eck, 2012).

It is proposed to use a slightly increasing convex functions like $f(C) = \sqrt{C_i}$ or $f(C_i) = In(C_{i+1})$ (Lundberg, 2007) in the form of $f(C_i)$. In the work of Levene *et al*. (2012) in Eq. 1, a relationship $f(C_i) = \sqrt{C_i}$ and a function $I(C_i) = \sqrt{\sum_{i=1}^{n} C_i}$ were also considered.

With a proper approximation, T. Marchant puts down the scoring rule in the following way:

$$U(f) = \sum_{j \in J} \sum_{x \in N} \sum_{a \in N} f(i, x, a)\, u(j, x, a) \quad (2)$$

where, f(i, x, a) is the number of publications of researcher f in a journal j with exactly x citations and a co-authors (the number of authors being a+1). u(j, x, a) is the value or the score of one publication in the journal j with x citations and a co-authors; J = {j, k, l, ...}dN: represents the set of journals and N is the set of integers.

The triple sum (Eq. 2) represents the total score of the author. As noticed by Marchant (2009), many popular bibliometric rankings are scoring rules. For example, if we choose U equal to a positive constant, we obtain the ranking based on the number of publications. If we define U by u(j, x, a) = x for all j0J, x, a0N, we obtain the ranking, based on the number of citations. If we define u by u = (j, x, a) = IF(j) for all j0J, x, a0N where IF(j) is the impact factor of the journal j, we obtain a ranking based on the sum of the impact factors, used by Fava and Ottolini (2000).

For the purpose of the current research, we consider the usage of the function u(j, x, a) as shown below as the most suitable:

$$u(j, x, a) = \frac{xIF(j)}{(a+1)} \quad (3)$$

In this case, we obtain a simple scoring rule ranking authors according to their number of citations weighted by the number of researchers and the impact factor (Marchant, 2009). In the research, we shall abstract away from the number of researchers a but take into account additional impact-factors of journals corresponding to those articles which have links to the articles of the researcher in question.

We suggested the idea of constructing IF-scoring rule by Moskovkin and Golikov (2013). Now, we are going to render a conception and a mathematical model of constructing IF-scoring rule. A flowchart of IF-scoring rule calculation we suggested is shown in Fig. 1.

In it ($P_1$, $P_2$, ..., $P_i$, ..., $P_n$) stand for a set of articles published by some author; ($J_1$, $J_2$, ..., $J_i$, ..., $J_n$) a set of journals in which the articles in questions are published; ($IF_1$, $IF_2$, ..., $IF_i$, ..., $IF_n$) a set of impact-factors of the above journals; ($P_{i1}$, $P_{i2}$, ..., $P_{ij}$, ..., $P_{ici}$) a set of the $C_i$ number of articles citing $P_i$ article; ($J_{i1}$, $J_{i2}$, ..., $J_{ij}$, ..., $J_{ici}$) a set of journals corresponding to the set of articles; ($P_{i1}$, $P_{i2}$, ..., $P_{ij}$, ..., $P_{ici}$); ($IF_{i1}$, $IF_{i2}$, ..., $IF_{ij}$, ..., $IF_{ici}$) a set of impact-factors of citing journals. It is worth noting that some journals from sets ($J_1$, $J_2$, ..., $J_i$, ..., $J_n$) and ($J_{i1}$, $J_{i2}$, ..., $J_{ij}$, ..., $J_{ici}$) can coincide.

In the simple case, the formula for calculating the score rule, corresponding to the above algorithm (Fig. 1), can be put down as follows:

$$\begin{aligned} U(P_1, P_2.., P_i, .., P_n) = {} & IF_1(IF_{11} + IF_{12} + ... + IF_{1j} + ... + IF_{1c_1}) + \\ & IF_2(IF_{21} + IF_{22} + ... + IF_{2j} + ... + IF_{2c_2}) + \\ & IF_i(IF_{i1} + IF_{i2} + ... + IF_{ij} + ... + IF_{ic_i}) + \\ & IF_n(IF_{n1} + IF_{n2} + ... + IF_{nj} + ... + IF_{nc_i}) \\ = {} & \sum_{i=1}^{n} \sum_{j=1}^{C_i} IF_i IF_{ij} \end{aligned} \quad (4)$$

It means that $U(P_1, P_2, ..., P_i, ..., P_n) = Q(IF_i, IF_{ij})$ is an increasing quadratic function of many variables. It has a characteristic feature common to all scoring rules:$U(IF_i + ) IF_i, IF_{ij} + ) IF_{ij})$ where $) IF_i > ) IF_{ij} > 0$ stand for small increments. Let us have a look at some special cases of the function (Eq. 4):

- If $IF_i = IF_{ij} = 1$ then $U = \sum_{i=1}^{n} C_i$
- If $If_i ... 1$, $If_{ij} = 1$ then $U = \sum_{i=1}^{n} IF_i C_i$
- If $\sum_{i=1}^{C_i} IF_{ij} = 1$ then $U = \sum_{i=1}^{n} IF_i$

All these cases have resulted from the application of Eq. 2 by Marchant (2009).

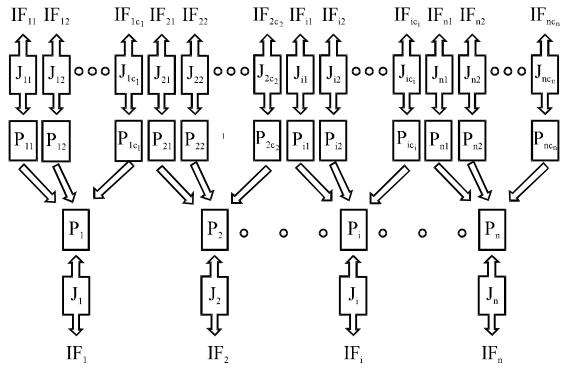

Fig. 1: A flowchart of IF-scoring rule calculation

Taking:

$$U = \overline{U} = \sum_{i=1}^{n} IF_i C_i$$

as a normalized function, one can see that when:

$$\sum_{j=1}^{C_i} IF_{ij} > C_i$$

and when the total impact factor of the journals citing the articles exceeds the number of citations, it will be true that $q = U/\overline{U} > 1$, otherwise q<1.

In addition to function (Eq. 4), we introduce five more functions of many variables:

$$U = \sum_{i=1}^{n} IF_i^{\frac{1}{2}} \left( \sum_{j=1}^{C_i} IF_{ij} \right)^{\frac{1}{2}} \quad (5)$$

$$U = \sum_{i=1}^{n} IF_i^{\frac{1}{2}} \sum_{j=1}^{C_i} IF_{ij}^{\frac{1}{2}} = \sum_{i=1}^{n} \sum_{j=1}^{C_i} IF_i^{\frac{1}{2}} IF_{ij}^{\frac{1}{2}} \quad (6)$$

$$U = \sum_{i=1}^{n} \sum_{j=1}^{C_i} (\delta + IF_i) IF_{ij} \quad (7)$$

$$U = \sum_{i=1}^{n} (\delta + IF_i)^{\frac{1}{2}} \left( \sum_{j=1}^{C_i} IF_{ij} \right)^{\frac{1}{2}} \quad (8)$$

$$U = \sum_{i=1}^{n} \left( \delta^{\frac{1}{2}} + IF_i^{\frac{1}{2}} \right) \sum_{j=1}^{C_i} IF_{ij}^{\frac{1}{2}} \quad (9)$$

where, * is some positive parameter. For Eq. 5, 6, 8 and 9, we take the square root in accordance with Lundberg (2007) to inhibit the growth of function U. Parameter * in Eq. 7-9 is introduced in order to attach value to the cited articles which were published in journals with a zero impact factor ($If_i = 0$).

## RESULTS AND DISCUSSION

For calculation by using Eq. 4-9, we developed a special algorithm and a Python-based program to identify titles of Scopus-journals (so that they could be searched for with the help of Google Scholar) in which an author's articles are published and to determine their IF using SCIMAGO platform.

The task of calculating IF-scoring rule for a particular author is divided into 2 sub-tasks: collection of information necessary for calculating, calculation proper. In the current research, the subtask (a) is much more time-consuming. It is, in turn, divided into 5 stages:

- Obtaining a list of journals with their IF
- Obtaining a list of author's articles with identifiers (titles) of journals in which the articles were published
- Obtaining a list of citing articles for every author's publication with identifiers (titles) of journals in which the articles were published

One should note that since the lists of articles referred to in points 2 and 3 were generated by scrapping and the following parsing of a search output of Google Scholar, this resulted in solving the most laborious problems of the entire program.

- Overcoming Google Scholar's protection from web crawlers
- Identifying and matching the titles of journals (or their fragments), obtained from the search output with the titles of journals from point 1

The algorithm and program were tested on the basis of two Google Scholar profiles of the most cited scientists (physicists) of Belgorod State University (Russia) Ruslan Kaibyshev and Andrey Belyakov (Table 1). The gathering of the initial information for calculating (Eq. 4-9) with the help of Google Scholar and SCIMAGO platform was done in August, 2013. $If_i$, $IF_{ij}$ in Eq. 4-9 were taken from SCIMAGO platform as IF = Cites/Doc (2 years).

In Table 1, the calculations by Eq. 7 when * = 0 conform with the calculations by Eq. 4; the calculations by Eq. 8 when * = 0 conform with the calculations by Eq. 5; the calculations by Eq. 9 when * = 0 conform with the calculations by Eq. 6. In Table 2, we show the values $\Delta U(*) = (U(* = 1)-U(* = 0)/U(* = 0))\times 100\%$ calculated on the basis of the data from Table 1.

Table 1: Calculation of IF-scoring rule for two most cited scientists of Belgorod State University (Russia), August 2013

| | Formula number/* | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 7 | | | | 8 | | | | 9 | | | |
| Author's Name | 0 | 0.01 | 0.1 | 1.0 | 0 | 0.01 | 0.1 | 1.0 | 0 | 0.01 | 0.1 | 1.0 |
| Rustam Kaibyshev Cited articles: 69 Citing articles from identified journals: 621 | 4014.9 | 4027.1 | 4137.3 | 5238.7 | 232.9 | 233.6 | 239.5 | 286.8 | 1389.9 | 1469.0 | 1639.9 | 2180.3 |
| Andrey Belyakov Cited articles: 40 Citing articles from identified journals: 292 | 1640.3 | 1646.5 | 1702.0 | 2257.0 | 173.2 | 173.6 | 177.5 | 210.6 | 592.6 | 630.2 | 711.5 | 968.7 |

Table 2: Values ) U(*), calculated on the basis of data from Table 1 (%)

| | Formula number | | |
|---|---|---|---|
| Author's Name | 7 | 8 | 9 |
| Rustam Kaibyshev | 30.5 | 23.1 | 56.9 |
| Andrey Belyakov | 37.6 | 21.6 | 63.5 |

From Table 2, we can see that the calculations by Eq. 8 are less sensitive to parameter variation * and this formula produces by-factor-of-ten smaller abcolute values of function U compared with those obtained through use of Eq. 7 and 9. This implies that for further calculations we recommend applying Eq. 5 as a particular case of Eq. 8.

## CONCLUSION

Basing on the scoring rule approach, there was designed a citation metrics, allowing for not only the number of author's articles published and their citations but also the impact factors of journals in which the articles were published as well as the impact factors of journals with articles citing the author. To calculate such a metrics, there were suggested six variants of formulas. For calculations by using these formulas, a special algorithm and a Python-based program identifying titles of Scopus-journals and determining their impact factors is developed.

The algorithm and program were tested on the two Google scholar profiles of the most cited scientists (physicists) of Belgorod State University (Russia). From the above mentioned computational formulas, we suggested applying the one providing by factor of ten smaller results.

## ACKNOWLEDGEMENT

Research Was Done According to The Government Task of The Ministry of Education and Science of The Russian Federation for 2014, Project Code-516).

## REFERENCES

Bornmann, L., R. Mutz, S.E. Hug and H.D. Daniel, 2011. A multilevel meta-analysis of studies reporting correlations between the h index and 37 different h index variants. J. Informetr., 5: 346-359.

Fava, G.A. and F. Ottolini, 2000. Impact factors versus actual citations. Psychother. Psychosom., 69: 285-286.

Hirsch, J.E., 2005. An index to quantify an individual's scientific research output. Proc. Natl. Acad. Sci. USA., 102: 16569-16572.

Levene, M., T. Fenner and J. Bar-Ilan, 2012. A bibliometric index based on the complete list of cited Publications. Cybermetrics, Vol. 16, No. 1.

Lundberg, J., 2007. Lifting the crown-citation z-score. J. Informetr., 1: 145-154.

Marchant, T., 2009. Score-based bibliometric rankings of authors. J. Am. Soc. Inform. Sci. Technol., 60: 1132-1137.

Moskovkin, V.M. and N.A. Golikov, 2013. [The new generation of citation metrics]. Proceedings of the International Conference on Construction of IF-Scoring Rules, October 10-12, 2013, Moscow, pp: 92-93.

Waltman, L. and N.J. van Eck, 2012. The inconsistency of the h-index. J. Am. Soc. Inform. Sci. Technol., 63: 406-415.