Information Sciences 263 (2014) 166–174

Contents lists available at ScienceDirect

Information Sciences

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

Monotone measures and universal integrals in a uniform framework for the scientific impact assessment problem

Marek Gagolewski [a,b,*], Radko Mesiar [c,d]

[a] Systems Research Institute, Polish Academy of Sciences, ul. Newelska 6, 01-447 Warsaw, Poland
[b] Faculty of Mathematics and Information Science, Warsaw University of Technology, ul. Koszykowa 75, 00-662 Warsaw, Poland
[c] Faculty of Civil Engineering, Department of Mathematics, Slovak University of Technology, 81 368 Bratislava, Slovakia
[d] National Supercomputing Center IT4Innovations, Division University of Ostrava IRAFM, 30. Dubna 22, 701 03 Ostrava 1, Czech Republic

ARTICLE INFO

Article history:
Received 12 June 2013
Received in revised form 21 October 2013
Accepted 7 December 2013
Available online 12 December 2013

Keywords:
Universal integral
Monotone measures
Aggregation
Indices of scientific impact
Bibliometrics
h-index

ABSTRACT

The Choquet, Sugeno, and Shilkret integrals with respect to monotone measures, as well as their generalization – the universal integral, stand for a useful tool in decision support systems. In this paper we propose a general construction method for aggregation operators that may be used in assessing output of scientists. We show that the most often currently used indices of bibliometric impact, like Hirsch's *h*, Woeginger's *w*, Egghe's *g*, Kosmulski's MAXPROD, and similar constructions, may be obtained by means of our framework. Moreover, the model easily leads to some new, very interesting functions.

1. Introduction

Many practical situations, especially in decision making, face us with the problem of aggregating sequences having not necessarily equal lengths. In such cases not only the aggregated elements have impact on the overall evaluation, but also the length of the sequence. This is the case of the bibliometric impact assessment problem, which concerns aggregating the number of citations received by articles published by authors having different productiveness. Of course, "raw" citations are not the only way to measure the quality of a paper: we can use other indicators, like impact factors of their journals, or citations which are normalized by the scientific domain and the number of authors, cf. [3,11,20]. Other instances of this issue include, e.g. manufacturing, quality engineering, webometrics, evaluation of open source software packages, see e.g. [7,10].

Let us assume that the whole information on the "producer's" (e.g. the author's) performance is represented by a vector $\mathbf{x} \in \mathbb{I}^{1,2,\ldots} = \bigcup_{n=1}^{\infty} \mathbb{I}^n$, where $x_i \in \mathbb{I} = [0, \infty]$ denotes the quality of his/her ith "product" (e.g. paper; of course, how to measure its quality is a problem on its own). Our interest here lies in finding a *method* that may be used to synthesize $\mathbf{x}$ so that his/her performance may be *described* with a single numeric value. Bibliometricians generally agree, see [8,10,17–19,24–26], that such an aggregation function should be (a) nondecreasing with respect to the quality of individual papers, e.g. after increasing the number of citations of a single article one should not get lower overall evaluation; (b) nondecreasing

* Corresponding author at: Systems Research Institute, Polish Academy of Sciences, ul. Newelska 6, 01-447 Warsaw, Poland. Tel.: +48 22 38 10 393; fax: +48 22 38 10 105.

E-mail addresses: gagolews@ibspan.waw.pl (M. Gagolewski), mesiar@math.sk (R. Mesiar).

0020-0255/$ - see front matter
http://dx.doi.org/10.1016/j.ins.2013.12.004

with respect to the number of papers, e.g. an author cannot be penalized if he/she publishes yet another paper, even if it has 0 citations at the beginning; and (c) symmetric, i.e. not depending on the order in which the aggregated papers are being presented.

One example of such an aggregation method is the famous *h*-index [12]. Torra and Narukawa showed that the *h*-index is equivalent to the discrete Sugeno integral with respect to the counting measure, see [23]. In fact, we will see that many of the indices of scientific impact may be expressed as monotone (fuzzy, not necessarily additive) integrals, like Choquet [5], Sugeno [22], or Shilkret [21] ones which have been known for over forty years. However, to the best of our knowledge, except for [23], there are only two other papers which describe the connection between the monotone integrals and scientometrics. Beliakov and James [2] consider applications of the Choquet integral-based classifiers to the problem of ranking of scientific journals. Additionally, Bras-Amorós et al. [4] mention that their bibliometric index based on the collaboration distance between cited and citing authors corresponds to a Sugeno integral w.r.t. some fuzzy measure.

The aim of this paper is to present a uniform model for the scientific impact assessment problem (as well as other similar domains) via monotone measures and integrals. Such a framework not only is very flexible, provides intuitive graphical interpretations for the aggregation process, and allows for constructing many new and interesting classes that may be used to describe the scientific record of a scientist. It also stands for another successful application of the fuzzy measure theory.

The paper is structured as follows. In the next section we present an axiomatization of the aggregation operators that are most often used in the post-Hirsch bibliometric impact assessment of individuals. After reviewing the most prominent impact functions, we recall the notion of a monotone measure and universal integral. In Section 3 we propose the uniform model that is based on universal integrals, and in Section 4 we show how to obtain the indices currently applied in practice, and also how to generate and compute very interesting new ones. Section 5 concludes the paper, indicating some important issues concerning the impact assessment task.

2. Preliminaries

One of the problems with the aggregation of vectors in $\mathbb{I}^{1,2,\ldots}$ is that in fact we are required to introduce a *family* of functions, each operating on fixed-length vectors. This is because if $\mathsf{F}: \mathbb{I}^{1,2,\ldots} \to \mathbb{I}$, then F may be written as $\mathsf{F} = (\mathsf{F}^{(1)}, \mathsf{F}^{(2)}, \ldots)$, where $\mathsf{F}^{(n)}: \mathbb{I}^n \to \mathbb{I}$. Thus, to achieve the main aim of the paper, we would have to consider *families* of monotone measures and *families* of integrals. The resulting model, although being interesting from the theoretical viewpoint, would be far too complex for practitioners.

We will therefore focus on the so-called zero-insensitive aggregation operators, cf. e.g. [10]. In such case, each uncited paper is treated as non-existing. Even though this setting may seem quite limiting, in fact most of the currently used bibliometric impact indices do obey this property, see [1] and also Section 2.1.

Let us consider the space $\mathcal{S}$ of infinite nonincreasing sequences with elements in $\mathbb{I}$. Let $\tilde{\cdot}: \mathbb{I}^{1,2,\ldots} \to \mathcal{S}$ be an operator such that $\tilde{\mathbf{x}} = (x_{\{1\}}, x_{\{2\}}, \ldots, x_{\{n\}}, 0, 0, \ldots)$, where $x_{\{i\}}$ denotes the *i*th greatest value in $\mathbf{x}$.

Proposition 1. *Let $\mathsf{F}: \mathbb{I}^{1,2,\ldots} \to \mathbb{I}$ be an aggregation function. Then F is a zero-insensitive impact function, i.e. it fulfills the following properties:*

1. $\mathsf{F}(0) = 0$ *(lower bound);*
2. $(\forall n)\ (\forall \mathbf{x}, \mathbf{y} \in \mathbb{I}^n)\ (\forall i)\ x_i \leqslant y_i \Rightarrow \mathsf{F}(\mathbf{x}) \leqslant \mathsf{F}(\mathbf{y})$ *(nondecreasingness);*
3. $(\forall \mathbf{x} \in \mathbb{I}^{1,2,\ldots})\ (\forall y \in \mathbb{I})\ \mathsf{F}(\mathbf{x}) \leqslant \mathsf{F}(\mathbf{x}, y)$ *(arity-monotonicity);*
4. $(\forall n)\ (\forall \mathbf{x} \in \mathbb{I}^n)\ (\forall \sigma \in \mathfrak{S}_n)\ \mathsf{F}(\mathbf{x}) = \mathsf{F}(x_{\sigma(1)}, \ldots, x_{\sigma(n)})$, *where $\mathfrak{S}_n$ denotes the set of all permutations of $\{1, \ldots, n\}$ (symmetry);*
5. $(\forall \mathbf{x} \in \mathbb{I}^{1,2,\ldots})\ \mathsf{F}(\mathbf{x}, 0) = \mathsf{F}(\mathbf{x})$ *(zero-insensitivity);*

if and only if there exists a nondecreasing function $\mathsf{E}: \mathcal{S} \to \mathbb{I}$, $\mathsf{E}(0, 0, \ldots) = 0$, such that for all $\mathbf{x} \in \mathbb{I}^{1,2,\ldots}$ we have $\mathsf{F}(\mathbf{x}) = \mathsf{E}(\tilde{\mathbf{x}})$.

The proof is straightforward and therefore omitted.

2.1. A review of impact functions

From now on we consider only vectors in $\mathcal{S}$. Some of the notable examples of zero-insensitive impact functions are listed below.

- Total number of citations:

$$\mathsf{S}(\mathbf{x}) = \sum_{i=1}^{\infty} x_i, \tag{1}$$

or, more generally, a weighted sum of elements of $\mathbf{x} \in \mathcal{S}$. This includes, e.g. "the total number of citations of 5 most cited papers".