# NOTICE TO THE READER

**This article, "The h-Index and the Number of Citations: Two Fuzzy Integrals, V. Torra, and Y. Narukawa, *IEEE Transactions on Fuzzy Systems*, Volume 16, Number 3, pp. 795-797, June 2008, DOI: 10.1109/TFUZZ.2007.896327 was submitted, processed and accepted as a LETTER. This article incorrectly appears under the PAPERS category and should be considered a LETTER for use in citation.**

**Editor-in-Chief, IEEE Transactions on Fuzzy Systems**

# The $h$-Index and the Number of Citations: Two Fuzzy Integrals

Vicenç Torra, *Senior Member, IEEE*, and Yasuo Narukawa

***Abstract*—In this paper, we review two of the most well-known citation indexes and establish their connections with the Choquet and Sugeno integrals. In particular, we show that the recently established $h$-index is a particular case of the Sugeno integral, and that the number of citations corresponds to the Choquet integral. In both cases, they use the same fuzzy measure. The results presented here permit one to envision new indexes defined in terms of fuzzy integrals using other types of fuzzy measures. A few considerations in this respect are also included in this paper. Indexes for taking into account recent research and the publisher credibility are outlined.**

***Index Terms*—Choquet integral, citation indexes, fuzzy integrals, $h$-index, Sugeno integral.**

## I. Introduction

ALTHOUGH citation indexes and impact factor analysis have been studied for some time, there is an increasing interest in this field that seems to be due to the increasing number of digital libraries as well as the need for simple decision making procedures. See, e.g., [1].

Until recently, one of the most common approaches for evaluating the impact of publications of a researcher was to consider the total number of citations of his/her papers. Recently, Hirsch proposed an alternative index, the so-called $h$-index [5], to overcome some of the problems of previous approaches. Roughly speaking, the index is to reduce the overall impact of those researchers with only a few papers but with a large number of citations, e.g., researchers with a few highly cited review papers. Instead, the $h$-index underlines those researchers with several papers with a much more relevant impact. This is at the expense of highlighting researchers with a lesser number of citations.

In this paper, we show that these two kinds of indexes correspond to two well-known fuzzy integrals. The first case corresponds to a Choquet integral and the second one to a Sugeno integral. As will be shown below, they correspond to fuzzy integrals with respect to the same fuzzy measure.

Establishing these two indexes as fuzzy integrals, we underline the fact that the indexes are aggregation operators and that other indexes can be developed considering other assumptions. In particular, using the same fuzzy integrals but other fuzzy measures will define indexes with other properties.

The structure of this paper is as follows. In Section II, we briefly review the fuzzy integrals. Then, in Section III, we review the indexes and establish the relationship between these indexes and the integrals. This paper finishes with some conclusions and a few remarks on how alternative indexes might be defined.

Manuscript received November 16, 2006; revised February 2, 2007. This work was supported in part by Generalitat de Catalunya under Grant AGAUR 2004XT 00004, by Spanish MEC Project PROPRIETAS under Grant SEG2004-04352-C04-02, and CONSOLIDER INGENIO 2010 CSD2007–00004.

V. Torra is with the Artificial Intelligence Research Institute, Spanish Scientific Research Council (IIIA-CSIC), Campus UAB, 08193 Bellaterra, Catalonia, Spain (e-mail: vtorra@iiia.csic.es)

Y. Narukawa is with Toho Gakuen, Kunitachi, Tokyo 186-0004, Japan, and also with the Institute for Research in Business Administration, Waseda University, Shinjuku-ku, Tokyo 169-8050, Japan (e-mail: narukawa@d4.dion.ne.jp).

Digital Object Identifier 10.1109/TFUZZ.2007.896327

## II. Preliminaries

This section reviews some elements of fuzzy measures and fuzzy integrals. For more details, see, e.g., [9], [2], or [4]. The Choquet integral was first defined in [3] and the Sugeno integral in [8].

We will consider measures and integrals with respect to a finite reference set $X = \{x_1, \ldots, x_N\}$. In the following section, we will consider $X$ as a set of published works.

*Definition 1:* A set function $\mu : 2^X \to \mathbb{R}_+$ is a fuzzy measure if it satisfies the following axioms:

i) $\mu(\emptyset) = 0$ (boundary conditions);

ii) $A \subseteq B$ implies $\mu(A) \leq \mu(B)$ (monotonicity).

Although it is also quite usual in the field of aggregation operators to consider the boundary condition $\mu(X) = 1$, we will not consider it here. This condition corresponds to a normalization on the measure.

There exist a large number of families of fuzzy measures. Among them, we should distinguish additive ones. A measure is additive when $\mu(A \cup B) = \mu(A) + \mu(B)$ for $A \cap B = \emptyset$. For example, probability measures are examples of additive measures. Another example of additive fuzzy measure is the one defined below, which is of special relevance in this paper and is used in Proposition 1.

*Definition 2:* Let $X$ be a reference set. Then the additive fuzzy measure $\mu : 2^X \to \mathbb{R}_+$ defined as $\mu(A) = |A|$ is called the counting measure.

Note that fuzzy measures generalize the concept of probability measure. In fact, the set of normalized additive fuzzy measures correspond to the set of probability measures.

Now we turn into the fuzzy integrals. We start with the Choquet integral and then turn to the Sugeno one. Fuzzy integrals are defined over a function on the reference set $X$. In our case, where $X$ is the set of publications, we have that the function will refer to the number of citations of such publications. So, given $x$ in $X$, we have that $f(x)$ corresponds to the number of citations of $x$.

*Definition 3:* Let $\mu$ be a fuzzy measure on $X$. Then the Choquet integral of a function $f : X \to \mathbb{R}_+$ with respect to $\mu$ corresponds to

$$\sum_{i=1}^{N} f(x_{\sigma(i)})[\mu(A_{\sigma(i)}) - \mu(A_{\sigma(i-1)})] \tag{1}$$

where $\{\sigma(1),\ldots,\sigma(N)\}$ is a permutation of $\{1,\ldots,N\}$ such that $f(x_{\sigma(1)}) \geq f(x_{\sigma(2)}) \geq \cdots \geq f(x_{\sigma(N)})$, $A_{\sigma(k)} = \{x_{\sigma(j)} \,|\, j \leq k\}$ (or, equivalently, $A_{\sigma(k)} = \{x_{\sigma(1)},\ldots,x_{\sigma(k)}\}$ when $k \geq 1$ and $A_{\sigma(0)} := \emptyset$)).

The Sugeno integral has a similar expression.

*Definition 4:* The Sugeno integral of a function $f : X \to \mathbb{R}_+$ with respect to a fuzzy measure $\mu$ corresponds to

$$\max_i \min(f(x_{\sigma(i)}), \mu(A_{\sigma(i)}))$$

where $A_{\sigma(k)} = \{x_{\sigma(j)} \,|\, j \leq k\}$ (or, equivalently, $A_{\sigma(k)} = \{x_{\sigma(1)},\ldots,x_{\sigma(k)}\}$ when $k \geq 1$ and $A_{\sigma(0)} := \emptyset$), and when $\sigma$ is a permutation such that $f(x_{\sigma(i)}) \geq f(x_{\sigma(i+1)})$ for $i \geq 1$.

## III. Citation Indexes

We start reviewing in this section two citation indexes. First, we consider the index corresponding to the number of citations.

*Definition 5:* The number of citations (NC)-index is defined as the total number of citations that papers from researcher $r$ have received so far. Formally speaking, let $X_r$ be the set of papers published by a given researcher $r$ and let $f(x)$ be the number of citations of paper $x \in X_r$; then we have that

$$\text{NC}_r = \sum_{x \in X_r} f(x). \tag{2}$$

Now, we turn to the $h$-index.

*Definition 6:* A researcher $r$ has an $h$-index $h$ if $h$ of his papers have received at least $h$ citations and the rest fewer than $h$ citations.

Using $X_r$ and $f$ as above, this can be expressed as: the $h$-index of researcher $r$ corresponds to

$$h_r = \max_i \min(f(x_{\sigma(i)}), i)$$

where $\{\sigma(1),\ldots,\sigma(N)\}$ is a permutation of $\{1,\ldots,N\}$ such that $f(x_{\sigma(1)}) \geq f(x_{\sigma(2)}) \geq \cdots \geq f(x_{\sigma(N)})$.

Now we show that the definitions above for NC and $h$ indexes correspond to Choquet and Sugeno integrals with respect to the counting measure $\mu$ given in Definition 2.

*Proposition 1:* The following holds.

- The NC index corresponds to the Choquet integral of $f$ with respect to the counting measure $\mu$.
- The $h$-index corresponds to the Sugeno integral of $f$ with respect to the counting measure $\mu$.

*Proof:* To prove this proposition, let us first consider the measure of the sets $A_{\sigma(k)} = \{x_{\sigma(1)},\ldots,x_{\sigma(k)}\}$. According to Definition 2, the measure of a set equals to its cardinality. Therefore, in this case, $\mu(A_{\sigma(k)}) = k$.

Now, we consider the two following two cases.

1) This case is trivial as $\mu(A_{\sigma(i)}) - \mu(A_{\sigma(i-1)}) = 1$ and, thus, (1) equals (2).
2) Let us consider the Sugeno integral

$$\max_i \min(f(x_{\sigma(i)}), \mu(A_{\sigma(i)})).$$

Taking into account that $\mu(A_{\sigma(i)}) = i$, we have that

$$\max_i \min(f(x_{\sigma(i)}), i)$$

which corresponds to the expression given above for the $h$-index.

■

## IV. Conclusion

In this paper, we have shown that the number of citations (NC-index) and the $h$-index correspond to two well-known methods for information aggregation: the Choquet and Sugeno integrals, respectively. Choquet and Sugeno integrals are two particular fuzzy integrals that have been extensively studied. This fact permits one to apply all results regarding fuzzy integrals in this setting. For example, the monotonicity of these integrals with respect to the function we are integrating implies that, for two functions $f$ and $g$ such that $f \leq g$, we have that the NC-index and the $h$-index of $f$ are smaller than the same indexes for $g$. Therefore, for both indexes, self-references increase the value of the index.

Additionally, we have shown that for both indexes, the integrals are computed with respect to the same fuzzy measure (the counting measure that corresponds to the cardinality of the set). This further stresses the similarities of both indexes when observed from the fuzzy setting.

The equalities presented here permit one to apply those results obtained in the field of fuzzy integrals and aggregation to analyze current indexes as well as to define new ones. As an example, we give below a few definitions of alternative indexes.

1) Index to stress the importance of *active researchers* (with respect to those that are no longer active): This index, inspired by the *contemporary h-index* in [7], is defined in terms of the following fuzzy measure:

$$\mu(A) = \sum_{x \in A} \rho_\alpha(\text{year}(x))$$

where $\rho_\alpha$ is defined as one when the year of a paper $x$ is greater than a certain year $\alpha$ and zero otherwise. For example, we can consider only papers published after year $\alpha = 2000$. If such an index is used to compare candidates, both Choquet and Sugeno integrals with respect to this measure will encourage the selection of those researchers that are still active (i.e., that have published papers with citations after the year 2000).
2) Index that takes into account *publisher credibility:* In a way similar to the index on active researchers, we define the measure in terms of a function over the publications. In this case, the function corresponds to publisher credibility. To make matters simple, we define *publisher credibility* as

the impact factor of the journal for the given year. So, the measure is defined as follows:

$$\mu(A) = \sum_{x \in A} \text{if}(\text{journal}(x), \text{year}(x))$$

where $if(j, y)$ is the impact factor of journal $j$ in year $y$ According to this definition, the larger the impact factor, the larger the overall NC and $h$-indexes for the researcher.

3) Index that takes into account the impact of related papers: Let $x_1$ and $x_2$ be two papers by the same author that are almost always cited together. Then, defining a measure $\mu$ that satisfies $\mu(A \cup \{x_1\}) = \mu(A \cup \{x_2\}) = \mu(A \cup \{x_1, x_2\})$, we have that only $x_i$ with a maximal $f$ (i.e., $x_1$ or $x_2$ with largest $f$) will be considered in the NC-index or $h$-index. Nevertheless, the definition of appropriate fuzzy measures to cope with this kind of situation is not an easy task. Note that a measure $\mu$ satisfying the equalities above would penalize an author when papers only cite one of the two papers ($x_1$ or $x_2$ but not both). Interaction indexes [6] might be considered for easing this process.

Other extensions are also foreseeable. We might consider the number of citations $f$ not only depending on the works by an author but also as a function of time. In this sense, multidimensional (fuzzy) integrals can also play a role.

## ACKNOWLEDGMENT

The comments of the referees are acknowledged.

## REFERENCES

[1] P. Ball, "Index aims for fair ranking of scientists," *Nature*, vol. 436, p. 900, 2005.

[2] T. Calvo, A. Kolesárová, M. Komorníková, and R. Mesiar, , T. Calvo, G. Mayor, and R. Mesiar, Eds., "Aggregation operators: Properties, classes and construction methods," in *Aggregation Operators*. Berlin, Germany: Physica-Verlag, 2002, pp. 3–104.

[3] G. Choquet, "Theory of capacities," *Ann. Inst. Fourier*, vol. 5, pp. 131–295, 1953/1954.

[4] M. Grabisch, T. Murofushi, and M. Sugeno, Eds., *Fuzzy Measures and Integrals: Theory and Applications*. Berlin, Germany: Physica-Verlag, 2000.

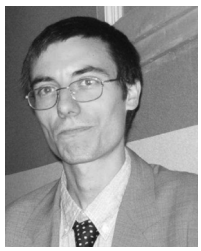
[5] J. E. Hirsch, "An index to quantify an individual's scientific research output," in *Proc. Nat. Acad. Sci.*, 2005, vol. 102, no. 45, pp. 16569–16572.

[6] T. Murofushi and S. Soneda, "Techniques for reading fuzzy measures (iii): interaction index," in *Proc. 9th Fuzzy Syst. Symp.*, Sapporo, Japan, 1993, pp. 693–696.

[7] A. Sidiropoulos, D. Katsaros, and Y. Manolopoulos, "Generalized $h$-index for disclosing latent facts in citation networks," 2006 [Online]. Available: http://arXiv.org, Cornell Univ. Library

[8] M. Sugeno, "Theory of fuzzy integrals and its applications," Ph.D. dissertation, Tokyo Inst. of Technol., Tokyo, Japan, 1974.

[9] V. Torra and Y. Narukawa, *Modeling Decisions: Aggregation Operators and Information Fusion*. Berlin, Germany: Springer, 2007.



**Vicenç Torra** (M'96–SM'03) received the B.Sc., M.Sc., and Ph.D. degrees all in computer science from the Universitat Politècnica de Catalunya (UPC), Spain, in 1991, 1992, in 1994, respectively.

He is an Associate Research Professor at the Artificial Intelligence Research Institute, Spanish Scientific Research Council (IIIA-CSIC) near Barcelona, Catalonia, Spain. His fields of interests are decision making, information fusion and soft computing tools and their applications to privacy issues. He has led several research projects funded by Catalan, Spanish, European and U.S. sources. His research has been published in specialized journals (more than 60 articles in ISI JCR journals and LNCS) and major conferences. He has written a book *Fonaments d'Intel·ligència Artificial* (in Catalan) and coauthored *Information Fusion and Aggregation Operators* (Springer, 2007).

Dr. Torra is a member of the editorial boards of *Fuzzy Sets and Systems*, *Journal of Privacy Technology*, and *Mathware and Soft Computing*. In 2004, he was co-founder of the conference series Modeling Decisions for Artificial Intelligence (MDAI), whose successive editions he has co-chaired ever since. He was also a Co-chairman of the Privacy in Statistical Databases conference (PSD 2004).

**Yasuo Narukawa** was born in Saitama, Japan, in October 1964. He received the B.Sc. degree in mathematics from Tokyo University of Science, Tokyo, Japan, in 1988, and the M.Sc. and Ph.D. degrees in systems science from the Tokyo Institute of Technology, in 1990 and 1998, respectively.

Previously, he was a Visiting Researcher at the Artificial Intelligence Research Institute, Spanish Scientific Research Council (IIIA-CSIC) near Barcelona, Catalonia, Spain, in 2005 and 2006. He is currently a Teacher at the Toho Gakuen (Kunitachi campus, Japan) and a Research Fellow at the Institute for Research in Business Administration, Waseda University, Tokyo. His fields of interests are decision theory, utility theory under risk, and data analysis and their applications. His research has been published in major journals and conference proceedings in the fields of mathematics and its applications in decision and mathematical education. He coauthored *Information Fusion and Aggregation Operators* (Springer, 2007).

Dr. Narukawa was co-founder of the 2004 conference series Modeling Decisions for Artificial Intelligence (MDAI), whose successive editions he has co-chaired ever since.