Some Distances as New Information Measures

Angel Garrido

Facultad de Ciencias de la UNED

Abstract

Our paper analyzes some new lines to advance on metric concepts, as can be the so-called Information Distance, on sets, graphs and networks in general. It will be very necessary to analyze the inner relationships with some other fuzzy measures, giving place to very interesting applications.

Keywords: Fuzzy Metric Spaces, Measure Theory, Fuzzy Measures, Similarity, Symmetry, Measures of Information and Entropy.

Mathematics Subject Classification: 68R10, 68R05, 05C78, 78M35, 94A17.

1 Introduction

Many aspects of Information Theory [3] are quickly evolving, and very interesting new open problems appear. One of them is to provide our theoretical construct with an efficient distance measure, to be applied in different fields of Computer Science; in particular, when working on current research lines of Artificial Intelligence. As Joel Ratsaby recalls in its recent paper [13], this would be essential in some fields, such as Pattern Recognition, to find a numerical value that represents the distance (or dissimilarity) between any two input patterns of the domain. Ratsaby also insists in the necessity of a consistent distance, which requires good information about the domain.

He introduces [13-15], after to a preliminary review on fundamental concepts, such as Kolmogorov Complexity, Entropy and so on, a definition of distance: for A and B sets,

$$\delta\left(A,B\right)=\log_{10}\left(\left|\left(A\cup B\right)\setminus A\right|\quad\left|A\right|\right)$$

where $|.|$ represents the cardinality of the corresponding set, which we prefer to denote as *card*, or simply as *c*. Other more awkward symbols might be used, like $\sharp$.

[1] **AMO - Advanced Modeling and Optimization. ISSN: 1841-4311**

Note that we use base two logarithm, because it is more adequate for the binary codification. But in this case, it appears as decimal logarithm, modifying the results according to the well known formula for transforming between different logarithmic bases.

And the cardinality is applied on the set of elements of the union, $A \cup B$, not included in A, i.e. $B \cap A^c$.

Joel Ratsaby [13-15] comments that the value $|(A \cup B) \setminus A|$ measures the additional description length (in bits) of an element in B, given knowledge of the set A. He also says that A acts as "a partial dictionary", and that the part of B that is not included in A requires $\log (|(A \cup B) \setminus A|)$ bits of description.

From here, the aforementioned author propose as a new distance measure the so-called *set-information distance*,

$$d (A,\ B) = \max \{\delta (A,\ B)\,,\ \delta (B,\ A)\,, 0\}$$

But because the cardinality $(A \cup B) \setminus A$ and A will be a natural number, supposing that they are non empty sets, it holds

$$\log_{10} (|(A \cup B) \setminus A| \ \ |A|) \geq \log\ 1 = 0$$

This makes for us unnecessary to include a comparative zero defining the precedent

$$d (A,\ B)$$

in the sense of Ratsaby paper.

Also we see that according the previous definitions, it may be

$$\delta (A,\ B) < 0$$
$$or$$
$$\delta (B,\ A) < 0$$

Is it really possible? For instance, we have

$$log_{10} (c (A \cap B^c) \ \ c (A)) = 10 \Leftrightarrow c (A \cap B^c) \ \ c (A) = 10$$

and

$$log_{10} (c (A \cap B^c) \ \ c (A)) = 0 \Leftrightarrow c (A \cap B^c) \ \ c (A) = 1$$

where both will be of the same cardinality, equal to one.

But

$$log_{10} (c (A \cap B^c) \ \ c (A)) < 0,\ for\ some\ A\ and\ B?$$

Being impossible to take negative values for the function

$$y = log_{10}\ x$$

because its graph is asymptotical respect to the straight line $y = 0$, never crossing the abscissa axis.

At most, it can take values that belongs to the real closed unit interval, $[0, 1]$.

As e.g.

$$log_{10}\left(c\left(A \cap B^c\right)\ c\left(A\right)\right) = \frac{1}{n},\ with\ n \in \mathbf{N}$$

i.e.

$$c\left(A \cap B^c\right)\ c\left(A\right) = 10^{\frac{1}{n}} > 0$$

In the limit, when $n \to \infty$, it holds

$$log_{10}\left(c\left(A \cap B^c\right)\ c\left(A\right)\right) \to 0^+$$

For these reasons, the null option in the bracket is unnecessary, when we define the final set-information distance.

And furthermore, it is more coherent and usual to suppose a logarithmic base equal to 2, because the binary strings are codified by sequences of 0's and 1's.

2 A new distance

Being interesting the introduction of such new measure proposed by Ratsaby [13], we consider the possible change of some essential aspects. So, it appears as convenient to introduce the following new distance measure.

Let A and B be two fuzzy sets. Then, the function defined by

$$\delta^*\left(A,\ B\right) = \log_2\left(c\left[A \triangle B\right]\ c\left[A \cap B\right]\right)$$

or

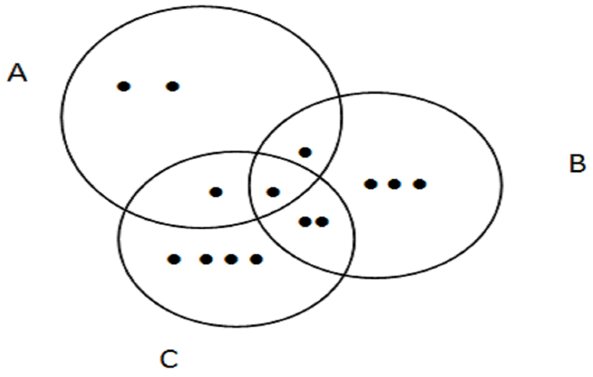
$$\delta^*\left(B,\ A\right) = \log_2\left(c\left[B \triangle A\right]\ c\left[B \cap A\right]\right)$$

But note that is will be equivalent to

$$\begin{aligned} \delta^*\left(A,\ B\right) = \log_2\left\{c\left[\left(A \setminus B\right) \cup \left(B \setminus A\right)\right]\ c\left[A \cap B\right]\right\} = \\ = \log_2\left\{\left[c\left(A \setminus B\right) + c\left(B \setminus A\right)\right]\ c\left[A \cap B\right]\right\} \end{aligned}$$

Because the sets $A \setminus B$ and $B \setminus A$ are mutually disjoint, or if instead are considered as events, they will be incompatible.

It is clear that the δ function in the sense of Ratsaby is not symmetric, i.e.



$$\delta(A,\ B) \neq \delta(B,\ A)$$

But this does not not occur in our case, in which

$$\delta^{*}(A,\ B) = \delta^{*}(B,\ A)$$

To show the precedent character non symmetric, we propose the following example

Here, we have

$$c\ \left(A \cap B^{c}\right) = 2$$
$$c\ (A) = 4$$

which implies

$$\delta(A,\ B) = \log_{2} 8$$

Whereas

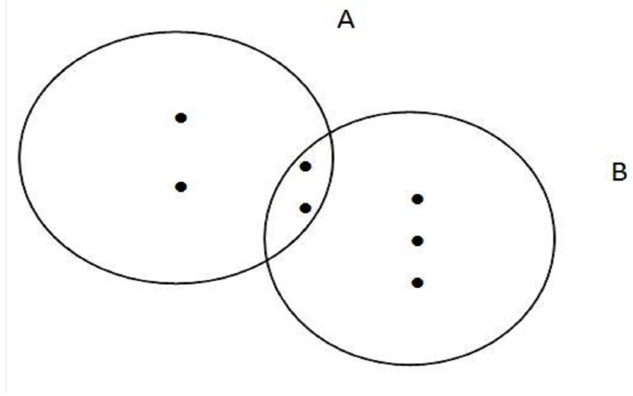
$$c\,(B \cap A) = 2$$
$$c\,(B) = 5$$

Thus,

$$\delta(B,\ A) = \log_{2} 10$$

Therefore, in general, δ is a non symmetrical function

$$\delta(A,\ B) \neq \delta(B,\ A)$$



Note that the symmetric difference is indeed symmetrical,

$$A \triangle B = B \triangle A$$

and so

$$c\left(A \triangle B\right) = c\left(B \triangle A\right)$$

But in our definition $\delta^*\left(A,\ B\right)$ and $\delta^*\left(B,\ A\right)$ does not differs in the last factor, because

$$c\left(B \cap A\right) = c\left(A \cap B\right)$$

For these reasons, it will be reasonable to introduce the new distance as

$$d\left(A,\ B\right) = \max\left\{\delta^*\left(A,\ B\right),\ \delta^*\left(B,\ A\right)\right\} = \delta^*\left(A,\ B\right)$$

The following theorem holds:

Theorem. The function δ^* acting on a pair of sets, A and B, holds the triangle inequality over the space of sets whose cardinality is at least two.

Previously to prove this theorem, we see an example.

Let $A,\ B,\ C$ be sets such that

We have

$$\begin{aligned} c\left(A \setminus B\right) &= 3 \\ c\left(B \setminus A\right) &= 5 \\ c\left(A \setminus C\right) &= 3 \\ c\left(C \setminus A\right) &= 6 \end{aligned}$$

$$\begin{gathered}c\,(B\setminus C)=4\\ c\,(C\setminus B)=5\\ c\,(A\cap B)=2\\ c\,(A\cap C)=2\\ c\,(B\cap C)=3\\ c\,(A\cap B\cap C)=1\end{gathered}$$

So, we have

$$\begin{gathered}log_2\ \ (\{3+6\}\ \ 2)=\log_2 18\leq \log\ \ (\{4+5\}\ \ 3)+\log_2\ \ (\{3+5\}\ \ 2)=\\ =\log_2\ 27+\log_2\ 16\end{gathered}$$

Proof.
As we know,

$$c\,[(A\setminus C)]\leq c\,[(A\setminus B)]+c\,[(B\setminus C)]$$

And so,

$$c\,[(A\setminus C)\cup(C\setminus A)]\leq c\,[(A\setminus B)\cup(B\setminus A)]+c\,[(B\setminus C)\cup(C\setminus B)]$$

Hence,

$$\begin{gathered}c\,[(A\setminus C)\cup(C\setminus A)]\ \ c\,[A\cap C]\leq\\ \leq\{c\,[(A\setminus B)\cup(B\setminus A)]\ \ c\,[A\cap B]\}\,\{c\,[(B\setminus C)\cup(C\setminus B)]\ \ c\,[B\cap C]\}\end{gathered}$$

Therefore,

$$\begin{gathered}\max\{\delta^*(A,\ C),\ \delta^*(C,\ A)\}\leq\max\{\delta^*(A,\ B),\ \delta^*(B,\ A)\}\\ +\max\{\delta^*(B,\ C),\ \delta^*(C,\ B)\}\end{gathered}$$

based in which

$$\delta^*(A,\ C)\leq\delta^*(A,\ B)+\delta^*(B,\ C)$$

for every triple of sets, *A, B* and *C.*

Theorem. The aforementioned function, *d,* is a metric over the space of the sets of cardinality at least two.

Proof.

Symmetry.

$d\,(A,B)=\max\{\delta^*(A,\ B),\ \delta^*(B,\ A)\}=\max\{\delta^*(B,\ A),\ \delta^*(A,\ B)\}=d\,(B,\ A)$

Non-negativity.

$$\delta^* (A,\ B) \geq 0$$
$$\text{and}$$
$$\delta^* (B,\ A) \geq 0$$

implies that

$$d\,(A,\ B) \geq 0$$

In particular,

$$d\,(A,\ B) = 0 \Leftrightarrow A = B$$

When $A \neq B$, it holds

$$A \triangle B \neq \varnothing$$

and being disjoint sets, or incompatible events,

$$A \cap B \neq \varnothing$$

Thus, either

$$\delta^* (A,\ B) > 0$$

or equivalently,

$$\delta^* (B,\ A) > 0$$

From either case, we obtain

$$d\,(A,\ B) > 0$$

I is very easy to show that the triangle inequality holds, in a similar way to [8], with minor modifications.

3 Another different “distance" measure

Alternatively, if we define the function through an expression as

$$\delta' (A, B) \equiv \log_2 (c\,[A \triangle B]\ \ c\,[A])$$

and

$$\delta' (B, A) \equiv \log_2 (c\,[B \triangle A]\ \ c\,[B])$$

both will be very different between them,

$$\delta' (B, A)$$

Note that in many cases, as they have distinct cardinality

$$c(A) \neq c(B)$$

the precedent values will be also different.

And not necessarily must consider non coincident sets to holds such equality.

Because taking two equipotent sets gives the same value for δ'through both ways; in this case,

$$\delta' (B, A)$$

But in general,

$$\delta' (B, A)$$

giving that δ' is not a metric.

It is not the case for the previously defined distances, as δ^*, or instead d.

4 Kolmogorov Complexity

This concept, of *Kolmogorov Complexity* (*KC,* in acronym), is also called *Turing Complexity, Kolmogorov-Chaitin Complexity,* or *Algorithmic Complexity*, among other names [9].

It was introduced and developed with different motivation, and independently, by Ray Solomonoff [11, 12], Andrei N. Kolmogorov [8], and also by Gregory Chaitin [4, 5].

Let s be a finite binary string of arbitrary length, i.e. an element of the set $\{0,1\}^*$. I.e. the function

$$\begin{array}{c} K : \{0,1\}^* \to \mathbf{N} \\ s \mapsto K(s) \end{array}$$

is defined on objects represented by binary strings. The subsequent definition will be extended to different types of objects, such as sets, numbers, functions or distributions.

We denote the Kolmogorov Complexity of s by

$$K(s)$$

It will be defined as the length of the shortest computer program that can produce this string on the Universal Turing Machine (UTM), and then halt. Or equivalently, it will be defined as the number of bits needed to encode s. Such UTM is not a real computer, but an imaginary reference machine. But because every Turing Machine may be implemented on every other one, the minimal length of the program only depends of an additive constant, being independent of the string considered.

An important result is that *KC is not computable,* because we cannot compute the output of every program. And it is due to the impossibility to create an algorithm which permits us predict of every program, if it will ever halt.

The *KC* can be also defined as the length of the string's shortest description in some fixed universal description language. It is equivalent to the previous interpretation. I.e. the *KC* will be thought as the length of the shortest program that print *s,* and then halts. This program may be in Java, LISP, or any other different universalprogramming language. The *Invariance Theorem* indicates that it does not matter which program we pick.

Therefore, *the KC of any string cannot be too much larger than the length of the string itself.*

Another trascendental result says that *among algorithms that decode strings from their descriptions, there exists an optimal one.*

5 Interpretation of such measures

Some distances are defined with the purpose to reach a measure of the dissimilarity between any two strings, s and s'. We shown two of them. Both measures are based on the Conditional Kolmogorov Complexity, $K\left(s/s'\right)$, which to amount to the length of the minimum size program that is needed to describe s, given s'. And they are also based on Algorithmic Complexity.

Joel Ratsaby also developed a distance between strings based on Combinatorial Complexity [13]. So, in the first place, we have

$$E\left(s, s'\right) = \max\left\{K\left(s/s'\right),\ K\left(s'/s\right)\right\}$$

and in second place, a normalized version, by

$$D\left(s, s'\right) = \max\left\{K\left(s/s',\ K\left(s'\right)\right),\ K\left(s'/s,\ K\left(s\right)\right)\right\}$$

As observed in [13-15], “the quality of clustering of data using the normalized compression distance depends on certain heuristic choices". It would be a very interesting remark to advance on the future research in the case of algoritmic information distance.

About another aspect of our distance, we will consider that as it is based on

$$A\triangle B \equiv (A\diagdown B)\cup(B\diagdown A) = (A\cap B^c)\cup(B\cap A^c) =$$
$$= (A\cup B)\diagdown(A\cap B)$$

So, it may be interpreted as set-conditional entropy of A given $B,$ and alternatively, of B given A. Hence,

$$\log_2(A\diagdown B) = \log_2(A\cap B^c)$$
$$\text{and}$$
$$\log_2(B\diagdown A) = \log_2(B\cap A^c)$$

represents the additional description length (measured in bits) of an element in B given knowledge of the set $A,$ and the same respective concept in A given the knowledge of B.

By a vision from an Information theoretical perspective, A will acts, in the former case, as a *"partial dictionary"*.

The part of B that is not included in A requires

$$\log_2(A\diagdown B)$$

bits of description, whereas the second value

$$\log_2(B\diagdown A)$$

reflects the number of bits needed to descript the part of A not included in B.

Finally,we can conclude that two alternative definitions of information measure distincts are possible.

More concretely,

$$d(A,\ B) = \max\ \left[\log_2(c[A\triangle B]\ \ c[A]),\ \log_2(c[B\triangle A]\ \ c[B])\right] =$$
$$= \max\ \left[\log_2(c[A\triangle B]) + \log_2(c[A]),\ \log_2(c[B\triangle A]) + \log_2\ c[B])\right]$$

and

$$d(A,\ B) = \min\ \left[\log_2(c[A\triangle B]\ \ c[A]),\ \log_2(c[B\triangle A]\ \ c[B])\right] =$$
$$= \min\ \left[\log_2(c[A\triangle B]) + \log_2(c[A]),\ \log_2(c[B\triangle A]) + \log_2\ c[B])\right]$$

may be interesting to be explored from a theoretical viewpoint.

6 Conclusions

Expressing by adequate formulae Information and Complexity in terms of program size will be currently a very useful idea. In fact, the applications extend to many different and promising fields, as may be Logic, Probability Theory, Physics, Theoretical Computer Science, and so on.

Acknowledgements

In memory of Professors Ray Solomonoff, José Mira, Miguel de Guzmán, and Dan Butnariu, great mathematicians and computer scientists from whom I have learned so much.

In this, my published paper number 116, I wish to dedicate my work to Neculai Andrei, Daniel and Nicoleta Breaz, Barna Lászlo Iántovics, Ioan Ileana, Hari Srivastava, Shigeyoshi Owa, Afrodita Iorgulescu, Shu-Kun Lin, Joel Ratsaby, Eugen Mandrescu, Adelina Georgescu, Antonio Almeida, Remus Joldes, Emil Olteanu, Cabiria Andreian Cazacu, Mircea Cazacu, Joe Rosen, Viorel Barbu, Solomon Marcus, Marcelina Mocanu, Boldur Barbat, Horia Pop, Florin Nichita, Themistocles Rassias, Adrian Constantinescu, Dorin Andrica, Petru Mocanu, Bogdan Patrut, Gheorge Paun, Michel Petitjean, Bazil Parv, Guillermo Garrido, Mirela Stefanescu, Paul Dan Cristea, Janos Fodor, Mircea Risteiu, Veselin Drensky, and many others. And also to my family. Without their support, this could not have been possible.

References

[1] C. H. Bennet et al. (1998). Information distance. *IEEE Trans. Info. Theory,* Volume 44.

[2] M. Burgin (1982). Generalized Kolmogorov Complexity in theory of computation. *Notices of RAS* (Russian Academy of Sciences). Volume 25, Issue No. 3, pp. 19-23.

[3] T. M. Cover, and J. A. Thomas (1991). *Elements of Information Theory.* Wiley - Intersciences, New York.

[4] G. Chaitin (1974). Information-Theoretic computational complexity. *IEEE Trans. on Info. Theory,* IT-20, pp. 10-15.

[5] G. Chaitin (1987). *Algorithmic Information Theory.* Cambridge University Press (CUP).

[6] N. C. Debneth, and M. Burgin (2003). Software Metrics from the Algorithmic Perspective. *Proc. ISCA Conference,* pp. 279-282, Honolulu, Haway.

[7] A. Garrido (2009). Information and Entropy Measures. *Acta Universitatis Apulensis (AUA journal),* Vol. 19, special issue, pp. 911-920.

[8] A. N. Kolmogorov (1965). Three approaches to the quantitative definition of information. *Problems of Information Transmission,* Volume 1, pp. 1-17.

[9] M. Li, and P. Vitányi (1997). *An Introduction to Kolmogorov Complexity and Its Applications.* Springer-Verlag.

[10] Y. Manin (1977). *A Course in Mathematical Logic.* Springer-Verlag.

[11] J. Ratsaby (2009). Combinatorial Information Distance. Technical Report No. arXiv: 0905.2386v1.

[12] J. Ratsaby (2007). Information efficiency. *33rd. Int. Conf.on Current Trends in Theory and Practice of Computer Science.* LNCS Volume 4362, pp. 475-487. Springer-Verlag, New York.

[13] J. Ratsaby (2006). On the combinatorial representation of information. *COCOON ´06 Conference.* LNCS Volume 4112, pp. 479-488, Springer-Verlag.

[14] R. Solomonoff (1960). *A Preliminary Report on a General Theory of Inductive Inference.* Report V-131, Zator Co., Cambridge (MA).

[15] R. Solomonoff (1964). A Formal Theory of Inductive Inference. *Information and Control.* Part I, Volume 7, Issue No. 1, pp. 1-22. And Part II, Vol. 7, Issue No. 2, pp. 225-254.

[16] Z. Wang, and G. J. Klir (1992). *Fuzzy Measure Theory.* Springer, Dordrecht, and Plenum Press, New York.

[17] Z. Wang, and G. J. Klir (2009). *Generalized Measure Theory.* IFSR International Series on Systems. Springer, New York.

[18] M. J. Wierman, and G. J. Klir. *Uncertainty-Based Information:Elements of Generalized Information Theory.* Second edition. Physica-Verlag. Springer, New York.

[19] B. Yuan, and G. J. Klir. *Fuzzy Sets and Fuzzy Logic: Theory and Applications.* Prentice Hall PTR, USA.

[20] L. A. Zadeh. *Fuzzy Sets, Fuzzy Logic and Fuzzy Systems: Selected Papers by Lofti A. Zadeh.* Edited by George J. Klir, and Bo Yuan. Advances in Fuzzy Systems - Applications and Theory - Volume 6, World Scientific.