Using Fuzzy Sets To Define the Quantity Of Information

Marcel Stoica

Economic Cybernetics Department
Academy of Economic Studies
15-17 Calea Dorobantilor,
71131 Bucharest
ROMANIA

Abstract: Information is considered as a change of a fuzzy set expressing a state.
In this case, a change of state is equivalent to a change in the grade of membership $\Delta\mu_A(x)$ of an element x belonging to a fuzzy set A. The quantity of information can be computed using certain functions, called informational functions. Four axioms are presented, which must be satisfied by any informational function. In the case of information accumulation, they will organise themselves in a pattern.
The intensity of the relation between two elements $x \in A$ and $y \in A$ at time t_0 is measured by a grade of membership.
In this case, the information gain is expressed by the change of the grade of membership $\Delta\mu_A(x)$.
Some examples are given from the field of figures of speech used by writers. In the case of metaphor, the concept of author-reader communication is defined as the possibility an educated reader has of "guessing" the author's intentions. In the algorithm used by the reader to decipher the message, the grade of membership to the author's message should increase at every step, ideally converging to one.

1. Introduction

In order to define the quantity of information we can use a method similar to the one used in fuzzy sets theory [1]. Considering a fuzzy set defined on a set A and a level characterized by a real number $a \in [0, 1]$, we can define the set of those elements x with the property that their grades of membership satisfy: $\mu_A(x) \geq a$.

In this way we obtain a set: $N_a^A \subseteq A$ of the following form:

$$N_a^A = \{ x \mid \mu_A(x) \geq a, a \in [0, 1], x \in A \} \qquad (1)$$

If we consider another level b, $b \neq a$, then we obtain another subset

$$N_b^A = \{ x \mid \mu_A(x) \geq b, b \in [0, 1], x \in A \} \qquad (2)$$

If $b \geq a$ then

$$N_b^A \subseteq N_a^A \qquad (3)$$

Supposing that a system S called the observing system studies a fuzzy set A then it is required that two distinct elements $x \in A$ and $y \in A$ are also perceived as distinct. Let Θ be a difference operation between two elements x and y. Let Δx be the measure of this difference, that is:

$$\Delta x = x \Theta y \qquad (4)$$

having the property of symmetry, i.e.:

$$x \Theta y = y \Theta x \qquad (5)$$

For example, if $x, y \in R$, in order to ensure this property we take the measure to be given by the absolute value of the difference between x and y.

We assume that the observing system S can perceive the difference between x and y only if the measure of this difference is above a certain tolerance limit (threshold) denoted by ε, i.e.:

$$\Delta x = x \Theta y \geq \varepsilon \qquad (6)$$

If $\Delta x < \varepsilon$ then system S considers that the elements x and y are equal. In the same way, we consider a threshold η for the grades of membership that correspond to the change Δx, with the property that

$$\mu_A(x) - \mu_A(y) \geq \eta \qquad (7)$$

if system S perceives x and y as distinct elements, and

$$\mu_A(x) - \mu_A(y) < \eta \qquad (8)$$

if system S considers $x = y$.

In this case, η is called discernability threshold.

The grade of membership $\mu_A(x)$ is determined by the observing system S (with the aid of a function, by simple estimation, or following various computations). The quantity of information the system has depends on the grade of membership $\mu_A(x)$. At a given moment t_0 we assume that this level is given by $\mu_A^0(x)$. From the point of view of system S, receiving information from the fuzzy set A means the carrying out of various observations which lead to the conclusion that the grade of membership of element x has changed, becoming $\mu_A^1(x)$, so that

$$\Delta\mu_A(x) = |\mu_A^0(x) - \mu_A^1(x)| > \eta \quad (9)$$

which means it remarks that the element has changed its grade of membership. The number of levels that element x has "jumped" over is:

$$n = \text{Round}[\Delta\mu_A(x) / \eta], \quad n \in N \quad (10)$$

where: Round is a round-off operator (in most cases the rule of accounting can be applied). This indicator expresses a measure of change. If there is no round-off we obtain:

$$z = \Delta\mu_A(x) / \eta, \quad z \in N \quad (11)$$

If the observator remarks that certain relations (which may be stronger or weaker) exist between pairs of elements, then he represents his knowledge as a pattern. In order to do so he builds an oriented graph in which at time t_0 he attaches a grade of membership to each edge (x, y) denoted by $\mu_A^0(x, y)$. It is obvious that at a later moment t_1 the grade of membership changes and becomes $\mu_A^1(x, y)$. As a consequence there is an information gain

$$n_{xy} = \text{Round}(\Delta\mu_A(x, y) / \eta_{xy}) =$$

$$\text{Round}(|\mu_A^0(x, y) - \mu_A^1(x, y)| / \eta_{xy}) \quad (12)$$

respectively

$$z = \Delta\mu_A(x, y) / \eta_{xy} \quad (13)$$

A special case is when instead of grades of membership one uses probabilities (which sum up to 1).

Relations (10) and (11) are obviously not affected by substituting grades of membership $\mu_A(x)$ with probabilities $p_A(x)$. Still, there are differences between the fuzzy and probabilistic representations which need be considered:

1) If the grades of membership are not normalized, $\Sigma\mu_A(x) \neq 1$, whereas $\Sigma p_A(x) = 1$;

2) Logical operations on grades of membership differ from those on probabilities. In the absence of information, the observer considers all grades of membership to be equal (or all probabilities equal, respectively). If the information refers to n states, then in using probabilities we have $p_A(x) = 1/n$, whereas we have $\mu_A(x)$.

In what follows, we shall assume that this assumption holds.

2. Informational Functions

If we interpret information as a change in the initial knowledge of observer S, then there is a function taking as argument a natural number n or a real number z which, if satisfying certain axioms, allows the computation of the quantity of information. Such a function f, called an informational function, must satisfy the following axioms:

1) $f: N \rightarrow R^+$ or $f: R^+ \rightarrow R^+$

2) $f(0) = 0$

3) $f(1) = 1$ (14)

4) if $z < w$ then $f(z) < f(w)$ i.e. it has to strictly increase.

The simplest informational function is the identical function

$$f(n) = n \text{ and } f(z) = z, \text{ respectively.} \quad (15)$$

Non-linear functions can also be used, e.g.:

$$f(n) = nm \text{ or } f(z) = zm, \ m > 0. \quad (16)$$

Another informational function could be logarithmic [2]:

$$f(n) = \text{logb}(b-1+n) \quad (17)$$

and

$$f(z) = \text{logb}(b-1+z)$$

respectively, where $b > 1$, which ensures that

$$b - 1 + n > 0 \quad (18)$$

and

$$b - 1 + z > 0$$

respectively.

The quantity of information QI is equal to the level given by an informational function.

A special case is when $b = 2$. The quantity of fuzzy information (QIf) now becomes:

$$Qif = \log2(1 + n)$$

and

$$Qif = \log2(1+z) \quad (19)$$

respectively.

For any informational function we can define a conventional unit of information that expresses the quantity of information received by the observing system S when the change in the grade of membership of the element of the observed fuzzy set is equal to the discernability threshold.

The total quantity of information is given by the relation:

$$Qit = \Sigma f(n) \text{ or } Qit = \Sigma f(z).$$

3. Examples From the Field of Figures of Speech

Literary forms of expression contain several kinds of imprecise statements. The observer - in this case the reader - successively receives several pieces of information that the author sends with the aid of carefully chosen words. Writers usually use a real arsenal of figures of speech, such as: elaborate epithets, similes, metaphors, allegories, etc. For instance, suppose that a literary work begins with the hero entering a forest (moment t_0). The reader, having registered this information, is uncertain as to the context in which this event took place. Anticipating this uncertainty, the author successively sends a good deal of information on the context of the action. For this purpose, the author can attach an epithet to the word "forest" (for instance: green, brown, red, etc.). The reader will register an informational gain, from the initial information ("the hero entered a forest") to the more complete information ("the hero entered a green forest"). Since at time t_0 the state of the forest was unspecified, the fuzzy set A attached to it had the following form:

$$A = \{ (\text{Green}, 1/n), (\text{Reddish}, 1/n), \ldots, (\text{Dry}, 1/n) \}$$

where n is the number of states the reader could have imagined (given his training). In this case, the grade of membership for x = green is

$$\mu_A{}^0(x) = 1/4$$

When the reader finds out that the forest is green, the same grade of membership becomes

$$\mu_A{}^1(x) = 1.$$

It follows that

$$\Delta\mu_A{}^1(x) = \mu_A{}^1(x) - \mu_A{}^0(x) = 3/4$$

Supposing that the observer is capable of discerning at least four shades for each state it follows that in this case we have:

$$n_1 = z_1 = \Delta\mu_A{}^1(x) \;/\; \eta = (3/4) \;/\; (1/16) = 12 \text{ levels.}$$

This means that the observer has managed to alter his knowledge by "jumping" over 12 informational levels. If we admit that the observer knows four shades of green, then the new fuzzy set is:

$$B = \{ (1, 1/4), (2, 1/4), (3, 1/4), (4, 1/4) \}$$

The author can tell the reader what shade of green the forest was via a simile. In this case, he may compare the color of the forest with the color of an unripe apple. The message sent becomes:

"The hero entered a forest the color of a green apple."

Since the reader can discern only four shades of green it follows that:

$$\eta_2 = 1/4$$

The information gain brought by the simile is

$$\Delta\mu_B{}^2(x) = 1 - 1/4 = 3/4$$

The number of levels the observer has "jumped" is

$$n_2 = z_2 = \Delta\mu_A{}^2(x) \;/\; \eta_2 = (3/4) \;/\; (1/4) = 3 \text{ levels.}$$

Therefore with the aid of the simile the reader is helped by the author to "jump" other three levels on the informational scale. On a logarithmical (base 2) scale, the same simile would allow to the reader an informational gain of:

$\log_2 (1+ n_2) = \log_2 (1+ 3) = 2$

The effects of the metaphor are much more difficult to analyse since there may be cases when the reader is unable to correctly interpret them.

For instance, the author could have chosen a metaphorical form of expression:

"The hero entered an emerald forest." The advantage is two-fold if the reader has enough training, viz.

- the message is shorter than in the case of the simile;
- the reader is more satisfied from an artistical point of view (he cooperates, in fact, with the author, with whom he finds himself in perfect dialogue).

The disadvantage of a metaphor is that it can determine a lack of communication between the author and the reader, in which case the metaphor gets misinterpreted or not interpreted at all. Let p be the probability that a metaphor is correctly interpreted. This probability is a measure of the degree of communication between the author and the reader.

The quantity of information sent by a metaphor can be computed by considering that initially the fuzzy set has, for instance, 16 kinds of states and that the probability of deciphering the message is p, i.e.:

$QI_{met} = (1 - 1/16)p / (1/16) = 15\ p$

In order to understand messages sent in a metaphorical language, the reader needs a real training. At first, when he lacks training $p \to 0$. But as he continues to read and instruct himself to understand such messages he becomes better at deciphering metaphors and obviously $p \to 1$. From the point of view of the informational scale it can be said that through new reading and reinterpreting of metaphorical messages, the grade of membership $\mu_A(x)$ estimated by the reader goes to the one intended by the author. This interpretation allows us to give a fuzzy definition of communication.

Two systems SE = the sender and S = the observer, together with a set A (which may be fuzzy or crisp) which SE sends to S, communicate if S can establish, on the basis of a reasonably complex algorithm, a set ... that is very close to A. As the decryption algorithm runs, the grades of membership grow, i.e. there is an information gain at every step.

REFERENCES

1. SULARIA, M., **Fuzzy Sets Representation: Useful Expressions in Algebraic Operations Definition**, REVISTA ROMANA DE INFORMATICA SI AUTOMATICA, Vol. 7, No. 2, 1997, pp.11-16 (in Romanian).

2. STOICA, M., **Information Measurement Coefficients**, National Conference on Present Trends in Cybernetics and Science Philosophy. The "Stefan Odobleja" Cybernetics Academy , Baile Felix, Romania, October 17-19, 1996 (in Romanian).