But, What Is It Actually a Fuzzy Set?

Enric Trillas

Abstract Supported in a new view of meaning as a quantity in whatever universe of discourse, and for its possible use concerning plain language and ordinary reasoning in 'Computing with Words', the paper deals with the basic concept of a fuzzy set. That is, not only with the collective a linguistic label generates in language, but also with what membership functions reflect on it once ideally seen as measures of meaning.

1 Introduction

More than fifty years after its introduction [1], the idea of a fuzzy set is not yet clear enough, and although no 'if and only if' definition of it exists, too often fuzzy sets are seen as if they were mathematical entities in themselves instead of linguistic entities.

The identification of a fuzzy set with a single one of its possible membership functions is something very bizarre, since the question that should be immediately posed is, 'But, which one of them?'. Paraphrasing Quine's words [2], it does not seem possible to describe an entity without some criteria of identity; and the correspondence between fuzzy sets and membership functions is one-to-many.

In addition, identifying the fuzzy sets in a universe of discourse X with all the functions in $[0, 1]^X$ is still another oddity, since no general criteria are known for assigning to each one of these this enormous amount of functions a linguistic label of which it can be a membership function. In front of the unknown multiplicity of predicative words acting on X, it is the aleph-one cardinality of the functions in $[0, 1]^X$. Furthermore, an important characteristic usually required of membership

To Professor José Luis Verdegay, with deep affection for his academic work. Curro, thanks for the large friendship we jointly kept from so long ago!

E. Trillas (✉)
University of Oviedo, Asturias, Spain
e-mail: etrillasetrillas@gmail.com

D.A. Pelta and C. Cruz Corona (eds.), *Soft Computing Based Optimization and Decision Models*, Studies in Fuzziness and Soft Computing 360,
DOI 10.1007/978-3-319-64286-4_12

functions is their continuity, for which a topology in X is necessary; when X can be seen as a subset of the real line such topology is automatically given, but, in general its existence is unknown. In plain language, words do act in whatsoever universe of discourse.

Nevertheless, most books on fuzzy set theory and fuzzy logic usually begin under these presumptions that, although sometimes not explicit but implicitly present in them, not even can be sustained as simple ‘working hypothesis’. Hence both theories usually appear as based on moving grounds; trying to find, at least, a more solid and suitable basement for the idea of fuzzy set is the only goal of this paper.

2 Language and Fuzzy Sets

2.1 What does not seem to be debatable is that a fuzzy set in a universe of discourse X comes from a linguistic label, that is, a predicative word P whose behavior in X is manifested through the elemental statements ′x is P′, and generates the ‘fuzzy set labeled P′. Each one of these statements reflects that a property p, with name P, not only holds up to some extent for the elements in X, but is exhibited by them and is externally recognized.

For instance, in the universe of the London inhabitants, the word ‘young’ is so well anchored in plain language that it allows to speak on the ‘young Londoners’; something that is understood by everybody. In the universe of the positive integers, the word ‘odd’ is also so well anchored in the language of Arithmetic, that it allows to speak on the ‘odd numbers’. In a universe of trains, the word ‘large’ is so well anchored in plain language that it allows to speak on the ‘large trains’. At each part of language, its speakers perfectly understand these statements. Often, in plain language, recognizing p in the elements in X has an empirical character.

There are predicative words generating ‘linguistic collectives’, that is, words that ‘collectivize’ in language [3]. When such language is an artificial one, like it is that of Arithmetic, those words whose behavior is defined by ‘if and only if’ conditions, are the precisely used words that, of course, also exist in plain language. For instance, and in a universe of discourse constituted by two people, John and Sarah, the word ‘young’ can be undoubtedly applied to Sarah provided it is known that she is 18 years old, but John is 67. Notwithstanding, the same word ‘young’ does not admit a precise use in a universe with a very big number of elements like it is that of London inhabitants; for instance, provided it were accepted that ′x is young’ if and only if x is no more than thirty years old, is that a Londoner of thirty years and a few days old, should cease to be qualified as young? Obviously, it is not in plain language where its speakers see linguistic collectives as a single entity.

It can be said that the use of precisely used words is rigid, but that of the not precisely used is flexible. Hence, the first are those words that collectivize in just a subset of the universe of discourse; they perfectly classify the universe in those elements fulfilling the property p, and those not fulfilling p. It just corresponds to the ‘axiom of specification’ in the Naïve Set Theory [4], stating that a bi-valued property p generates in X a perfect classification of its elements in two subsets, that with those x fully verifying p, and that with those not verifying p at all. A representation of a precisely used word P in X, is done by the set **P** stated by the axiom of specification; its use is rigid, and the Boolean algebra of the subsets in X is the non debatable structure for representing these linguistic labels; it is a mathematical model perfectly mimicking how precise words are used when all the necessary information on their use is, at least, potentially known.

The words not precisely used, and as it is shown by a ‘Sorites’ type argument [5] applied to each one of them, cannot collectivize in a subset. The, well anchored in plain language, linguistic collectives they generate are like gaseous, or cloudy, entities; but, anyway, gas volumes and clouds do actually exist, and are scientifically studied. Indeed, linguistic collectives are rooted in plain language, and founding a mathematical model for them is imperative; at least for symbolically representing the ordinary reasoning that is greatly permeated by imprecision.

2.2 It seems that the study of the collectives generated by not precisely used words deserve to be approached by not considering them as pure mathematical entities characterized by an ‘if and only if’ definition, but in a scientific-like style. Such entities should be seen in a different way than sets, but, in any case, they can, and will, be identified with the fuzzy sets; that is, naïvely renaming them as fuzzy sets. The fuzzy set in X labeled P is just the linguistic collective P generates in X, the collective named ‘the Ps of X′, and it allows to see a fuzzy set as a single, although not precise, entity. In this way, the linguistic collective generated by a precisely used word is just a subset of X; subsets represent but degenerate, bi-valued, collectives; a non-degenerated collective is a purely linguistic concept.

In this line of thought, the fuzzy set labeled P can be seen as a single, although cloudy, linguistic entity rooted in X. But it still lacks to answer the question: How a fuzzy set can be specified in such a way that the axiom of specification for the precisely used words can follow from it? A possible answer for this question lies in ‘meaning’; it is a semantic topic.

3 Words with Measurable Meaning

3.1 The meaning of a word is not independent of the context of its use; for instance, ‘odd’ does not mean the same in a universe of positive integer numbers than in one of people. Sometimes the meaning of words can even change depending on the purpose for its use, as it is with ‘odd’ when used to qualify people either with a joking, or a hilarious, or an insulting purpose. It suffices to look for words in a dictionary to check all this by seeing how their uses are described.

Hence, no realistic theory of meaning can assign to words a single and universal meaning, since it depends on the universe of discourse, and on the particular context of its use; the meaning of words is context-dependent and purpose-driven. Thus, and in addition, for arriving at a theory of meaning it should be considered how P 'behaves', or 'acts' in X. Language is not static, but dynamic; almost always, time intervenes in language.

Once a pair (X, P) is given, how the 'behavior' of the word P in the universe X can be described? It only can come from considering the action of P for the x in X; that is, from the elemental statements ′x is P′, that constitute a different entity X [P] than X; the second can be physical or virtual, but the first is always virtual. Nevertheless, to capture the 'behavior' of P in X, it should be recognized not only the action of P on the elements x, but the context on which the statements ′x is P′ are used. To capture how P behaves, or acts, in X, it not suffices to statically capture what indicate the statements ′x is P′, it is also necessary to know how such action varies along X; its internal dynamism. That is, recognizing the linguistic relationship [6],

$$'\text{x is less P than y}',\ \text{equivalent to}\ '\text{y is more P than x}';$$

in other words, that x verifies p, less than y does. Such recognition is, in the case of plain language, often of an empirical character not immediately allowing the assignation of a degree to the verification of p.

Let's, symbolically, denote it by $x <_P y$, and by $x =_P y$ the case in which both $x <_P y$ and $y <_P x$, or $x <_P^{-1} y$, hold; obviously, $=_P = <_P \cap <_P^{-1}$.

To avoid the possibility $<_P = \emptyset$, let's state that

$$x <_P x \text{ holds for all x in X, that } <_P \text{ is a reflexive relation,}$$

although no reason for stating other properties like symmetry, or transitivity, etc., can really, and generally, exist. In general, $<_P$ is not a partial order in X; less again it is a total, or linear, relation since usually non comparable elements will exist, that is, pairs x, y for which it is neither $x <_P y$, nor $y <_P x$.

In the case the word P is precisely used in X, the relation $<_P$ collapses in the relation $=_P$; that is, $<_P = <_P \cap <_P^{-1} \Longleftrightarrow <_P = <_P^{-1}$. For instance, there is no way to directly stating that '7 is less odd than 517′; all odd numbers are equally odd; analogously, no way of directly establishing that '7 is less prime than 17′ exists, all prime numbers are equally prime. The word 'directly' refers to doing it without a previous new definition of 'less than' for the corresponding word, but only under the old definitions of 'odd' and 'prime'. Numbers are perfectly classified in 'odd' and 'not odd', 'prime' and 'not prime', etc.

Once the relation $<_P$ is known, the graph $(X, <_P)$ reflects the 'semantic organization' the use of P introduces in X; the graph is a *qualitative meaning* of P in X. It should be pointed out that the universe X cannot be always supposed to be mathematically structured; in plain language, words act in universes whatsoever. For instance, the Kolmogorov's theory of probability concerns 'probable' events in a Boolean algebra, but, in plain language, the same word 'probable' is not only applied to such kind of rigid events. A theory of linguistic meaning cannot presume that the universe of discourse is directly endowed with a particular mathematical structure.

Notice that such a definition remembers the intuitive idea that, when (intelligently) talking on some subject, some 'ordering' between the statements that are uttered, gestured, or written at the respect and for the corresponding reasoning's argumentation, is tried to be introduced among them. Additionally, the former definition also seems to be in agreement with the famous Ludwig Wittgenstein's statement [7], 'The meaning of a word is its use in language'.

3.2. Once a graph $(X, <_P)$ is known, the possibility of measuring to which extent x verifies p is open. A measure on such graph is a mapping $m_P: X \to [0, 1]$, such that [8]:

(1) If $x <_P y$, then $m_P(x) \leq m_P(y)$,
(2) If x is maximal in the graph, that is, no other y verifying $x <_P y$ does exist, then $m_P(x) = 1$,
(3) If y is minimal in the graph, that is, no other x verifying $x <_P y$ does exist, then $m_P(x) = 0$.

Notice that the closed unit interval could be substituted by any closed interval [a, b] in the real line, by just a playing the character of 0, and b that of 1 in the former definition. Notice also that no additive law is presumed for m_P; its definition is free from considering 'and', 'or', and either the concepts of incompatibility or contradiction that are only indistinguishable in the framework of Boolean algebras.

The additive law is deeply involved with a 'rigid form' of classifying elements, and, in plain language contradiction is independent of incompatibility, contrarily, for instance to the case of Ortho-lattices where the first implies the second ($p \leq q' \Rightarrow p \cdot q = 0$), with the reciprocal only holding provided the Ortho-lattice is a Boolean algebra, that is, it is distributive and consequently verifies the law of 'perfect repartition', $p = p \cdot q + p \cdot q'$, for all pair p, q. In this case, $p \cdot q = 0$ implies $p = p \cdot q' \Leftrightarrow p \leq q'$, and the reciprocal also holds. A lot of structural laws is necessary for the equivalence between contradiction and incompatibility; something that cannot be presumed in plain language.

The former general definition of a measure is inspired on that of a 'fuzzy measure' introduced by Michio Sugeno [9], but liberated from the constraints imposed by just measuring subsets; it is free from any mathematical structure in X further than that of graph. An antecedent of it can be found in the concept of a 'fuzzy entropy', introduced by Aldo de Luca, and Settimo Termini [10], where P = fuzzy and $<_{fuzzy}$ coincides with the so called 'sharpened order' between functions in $[0, 1]^X$.

Like, for instance, in the case with probabilities, with Sugeno's λ-measures, or with de Luca-Termini fuzzy entropies, all of them measures, the three axioms of a measure are not sufficient for individuating a single m_P, and additional suppositions should be added for each one of them. Anyway, each triplet $(X, <_P, m_P)$ facilitates a quantity reflecting a *quantitative meaning* of P in X, and, in this way, each 'full meaning' can be seen as a quantity.

By paraphrasing Lord Kelvin's words [11], 'If you cannot measure it, it is not Science', viewing meaning as a quantity can open the door towards a scientific-like study of collectives. Let's show a toy-example to illustrate what has been said.

Consider X = [0, 10], and P = big, generating the linguistic collective, or fuzzy set 'big numbers between 0 and 10′. This collective is not a set, provided 'big' is not rigidly but flexibly used; for instance, provided it can be stated that '8 is big', also '7.99 is big' can be stated. A 'Sorites' argument [5] shows that, not being the 'collective big' empty since 10 is always considered big, it is not a subset of [0, 10]. The qualitative use of big in [0, 10] can be described by:

$$x <_{big} y \Leftrightarrow x \leq y,$$

that is, by the qualitative meaning ([0, 10], $\leq$), a graph that is but a linearly ordered interval with maximum 10 (the only maximal), and minimum 0 (the only minimal). Hence, the measures of big in [0, 10] are the mappings m_{big}: [0, 10] $\rightarrow$ [0, 1], such that,

(1) $x \leq y => m_{big}(x) \leq m_{big}(y)$
(2) $m_{big}(10) = 1$
(3) $m_{big}(0) = 0$,

to which, adding the condition of 'usual flexibility',

(4) If x can be qualified as big, it exists ε (x) > 0 such that those y in the interval (x – ε (x), x] can be also qualified as big,

that could be translated into m_{big}, as
(4*) m_{big} is continuous in [0, 10].

Hence, the measures of big are the mappings between [0, 10] and [0, 1] that are strictly non-decreasing, and verify the border conditions (2) and (3). There is an enormous amount of them.

Consequently, to specify a measure for the meaning of big, it is necessary to add some additional information on the behavior of big, like it can be on its shape. For instance, provided it is known, or can be reasonably presumed, that the measure should be linear, $m_{big}(x) = ax + b$, it follows that the only possible linear measure is x/10, but provided the information on its shape were that it is quadratic, $m_{big}(x) = ax^2 + bx + c$, several possibilities for the values of a and b are possible, since by (3) it follows c = 0, and from (2) that 100 a + 10 b = 1; for instance, a quadratic measure is $x^2/100$ (with a = 1/100 and b = 0), but, obviously, it is not the only quadratic measure that is possible for 'big'.

In conclusion, the graphs ([0, 10], $\leq$, m_{big}), with m_{big} verifying the axioms (1)–(4), plus some additional information or additional reasonable hypotheses, are the quantities that specify a full meaning of 'big' in [0, 10]. These quantities require the 'design' of the corresponding measures.

Notice that axioms (4) and (4*) exclude to interpret the use of big as a precise one, since it will be specified by a subset of [0, 10] whose measure cannot be continuous but with jumps; notwithstanding, such rigid interpretation is possible by avoiding (4), by renouncing to flexibility. That is by ‘making precise’ the meaning of ‘big’; something that means, indeed, changing the ordinary and often usual, use of ‘big’ in plain language.

In some cases, the designer should add to the axioms, some hypotheses, reasonable for the current situation he is faced with, and like it is the former linear hypothesis. In this way, the linguistic collective P generates is qualitatively described by the graph $(X, <_P)$ once it is known, and shows different ‘informational states’ each one given once a quantity $(X, <_P, m_P)$ is specified.

3.3. The former interpretation of meaning as a quantity, actually preserves what has been said for the precisely used words, and it can be proven as follows.

If P is precisely used in X, the graph is $(X, =_P)$, and then if $x =_P y \Leftrightarrow x <_P y$ & $y <_P x$. It implies $m_P(x) \le m_P(y)$ & $m_P(y) \le m_P(x)$, or $m_P(x) = m_P(y)$. Thus, only the values 0 and 1 can be taken by the measure, since there are no other elements than the maximal (those verifying P), and the minimal (those not verifying P). Hence, the subset $m_P^{-1}(1)$ consists in the maximal (or prototypical) elements, and is the set **P** specified by P in X; obviously, it is $m_P^{-1}(0) = \mathbf{P}^c$ that contains the minimal or anti-prototypical elements.

4 Membership Functions

Is there any difference between the former measures of the linguistic label ‘big’, and the membership functions assigned to it in any book on Fuzzy Logic? There is no one. Thus, it can be stated that a membership function is, ideally, but a measure of the qualitative meaning of its linguistic label; is a quantitative ‘informational state’ of the fuzzy set.

Nevertheless, as the word ‘ideally’ tries to remark, this statement should be submitted to caution since a membership function is designed with the information available to its designer; an information not always consisting in all the relation ‘less than’, nor in knowing all its maximal and minimal elements but only some of them. There is some similarity with what happens when saying that the probability of obtaining each one of the six faces in throwing a die is 1/6, but without knowing if the die has some imperfection, or it is a tricky one, or the landing surface is not perfectly plane. An ‘ideal’ die is supposed to have a probability of 1/6 for each one of its faces; but often, ideal dice, those who throw them, and the landing surfaces, are not perfect. To say nothing when the die is tricky, for instance, having inside and in front of a face, a very small piece of plumb, or when it is thrown into sand. Throwing a die is a real situation, and often real situations are not ‘ideal’; analogously, the design of a membership function departs from some real situation in plain language. No doubt that the measures are membership functions, but, are membership functions always measures?

In addition, once the designer arrives at a membership function μ_P, a new relation is automatically established in X:

$$x \leq_{\mu} y \Leftrightarrow \mu_P(x) \leq \mu_P(y),$$

that is a linear one and, consequently, not always coincidental with $<_P$. Thus, provided $\leq_{\mu}$ substitutes $<_P$, the new 'qualitative meaning after design' is not already the original one. Provided μ_P were a measure, since

$$x <_P y => \mu_P(x) \leq \mu_P(y) \Leftrightarrow x \leq_{\mu} y, \text{ or } <_P \subseteq \leq_{\mu},$$

the design enlarges the qualitative meaning. Hence, the original qualitative meaning could be changed by the larger one coming from design; design could modify meaning. This is certainly risky, since practitioners usually look at the behavior of P in X after counting with a membership function and just through its shape; with it, they can easily appreciate a larger qualitative meaning than the real one.

In conclusion, in most practical cases, the membership function cannot be exactly a measure, but an often unknown approximation to one of them. Thus, to well representing the meaning of a linguistic label, it is important to know when a membership function can be seen as a good enough approximation to a measure. Working with a membership function not well reflecting the meaning of its linguistic label, can conduct to wrong results coming, for instance, from the fact that the membership function actually represents a different linguistic label.

5 Searching for a Definition and a Theorem

Fuzzy Control counts with a theorem stating on which conditions the computed output universally approaches the real one [12], but it lacks a similar theorem for assuming that a membership function truly approaches a measure. That is, a theorem from whose antecedent it can follow that, given a measure m_P of a qualitative meaning $(X, <_P)$ of P in X, a designed membership function μ_P is 'good enough' provided it exists a measure m_P such that either

(1) For all $\varepsilon > 0$ is $|m_P(x) - \mu_P(x)| \leq \varepsilon$, for all x in X,

or

(2) It is minimized the function $\text{Sup}_{x \in X} |m_P(x) - \mu_P(x)|$.

Instead of a system of rules and a defuzzification's method, what here is initially known are the relation $<_P$, or a part of it (both can be seen as a system of rules), and the axioms m_P should verify.

Both (1) and (2) could be considered as suitable definitions for stating that μ_P approaches m_P. Nevertheless, to prove them as a theorem's conclusion, some reasonable hypotheses, or some additional contextual information on the behavior of P in X, should be taken into account. But it depends on each particular case, and, for example, on how the designer did build up the membership function, that is, on the information available to s/he on the behavior of P in X. It remains an open question that, perhaps, should be posed from a different point of view as it is, for instance, beginning by modifying up to some limits the relation $<_P$, or the measure, or by just considering some type of measures.

I mention such possibility under the conviction that a generally accepted definition concerning the relation a designed membership function should keep with measures, as well as to proving on which conditions it can hold, is an important topic. Of course, were μ_P a measure, both the definition and the theorem are unnecessary. Anyway, and at least, some sufficient condition for knowing if the membership function approaches a measure will be interesting. Up to when something similar will be found, the design of linguistically described systems will continue being done in a not standardized and blind form.

6 Conclusion

This paper is just a first trial to penetrate on what the idea of a fuzzy set is, and on what its description by membership functions means. That the topic is just open, but not fully achieved, is manifested by the final call towards clarifying which membership functions can be actually considered as a good enough approximation to the fuzzy sets informational states.

What does not seem dubious is that fuzzy sets are linguistic, not mathematical, entities, that rooted in plain language belong to its domain, and that their membership functions should come from a process of design. In themselves, fuzzy sets seem to need a scientific-like study instead of a purely mathematical one. In addition, it yet lacks to count with a standardization of the corresponding design's processes for what concerns the approximate character of membership functions.

Since plain language is full of imprecise words, to mathematically mimicking ordinary reasoning, that is, to establish mathematical models for the not fully deductive types of reasoning, it is strictly necessary to consider imprecision, and, hence managing fuzzy sets for its symbolic representation. Thus, clarifying the idea of 'linguistic collectivization' is relevant.

There are still several problems that remain open for counting with good enough theoretical foundations of Zadeh's 'Computing with Words and Perceptions', whose ground lies in plain language and ordinary reasoning. For instance, ambiguity also permeates plain language and no mathematical model for scientifically managing ambiguity is currently known.

Notwithstanding, what can be asserted is that only one kind of fuzzy set actually exists, and that names like 'type-two fuzzy set' only can refer to the range in which the membership function takes its values.

References

1. Zadeh, L.A.: Fuzzy Sets. Inf. Control **8**, 338–353 (1965)
2. Quine, W.V.O.: Ontological Relativity and Other Essays. Columbia University Press, New York (1969)
3. Trillas, E.: On a model for the meaning of Predicates. In: Seising, R. (ed.) Views on Fuzzy Sets and Systems from Different Perspectives, pp. 175–205. Springer, Berlin (2009)
4. Halmos, P.R.: Naive Set Theory. Van Nostrand, New York (1960)
5. Trillas, E., Urtubey, L.A.: Towards the dissolution of the sorites paradox. J. Appl. Soft Comput. **11**(2), 1506–1510 (2011)
6. Trillas, E., Eciolaza, L.: Fuzzy logic. An Introductory Course for Engineers. Springer, Berlin (2015)
7. Wittgenstein, L.: Philosophical Investigations. Basil Blackwell, London (1958)
8. Trillas, E.: On a Model for the meaning of predicates. In: Seising, R. (ed.) Views of Fuzzy Sets and Systems from Different Perspectives pp. 175–2005. Springer, Berlin (2009)
9. Sugeno, M.: Theory of fuzzy integrals and its applications. Ph.D. Tokyo Institute of Technology, Tokyo (1974)
10. de Luca, A., Termini, S.: A definition of a non-probabilistic entropy in the setting of fuzzy sets theory. Inf. Control **20**, 301–312 (1972)
11. Lord Kelvin, W.T.: Lecture on electrical units of measurement; popular lectures and addresses, I/79. Mc Millan, London (1889)
12. Castro, J.L., Delgado, M.: Fuzzy systems with defuzzification are universal approximators. IEEE Trans. Syst. Man Cybern. **26**, 149–152 (1996)