The Interpretation of Fuzziness

Pei Wang
Center for Research on Concepts and Cognition
Indiana University
pwang@cogsci.indiana.edu

September 22, 1993

Abstract

Without a clear interpretation of fuzziness, it is hard for fuzzy logic to justify its rules, to get initial data from users, or to make its results understandable.

It is possible to interpret grade of membership, at least in some cases, as the proportion of positive evidence. In this way, fuzziness and randomness can be uniformly treated.

1 Zadeh on Fuzziness

Zadeh's idea of "fuzzy set" came from an observation: classes of objects in everyday thinking usually have no well-defined boundary ([29]).

More concretely, he made the following claims:

1. For these classes, no two-valued membership function can be defined on instances, and there are always instances that standing on the boundary, such as (his example) "animal", "beautiful women", "tall men", and "real numbers which are much greater than 1".

2. The above fact doesn't mean there is nothing we can say about the membership relation between a class (or a set, a concept) and an object (or an instance). On the contrary, such a relation can be compared, and even measured. It is a continuum of grades.

3. Since "the source of imprecision is the absence of sharply defined criteria of class membership rather than the presence of random variables", and "the notion of a fuzzy set is completely nonstatistical in nature", probability theory cannot be applied here. A new theory is needed (also see [3]).

Based on these intuitions, he defined the concept of *fuzzy set*, the relations between fuzzy sets (*equal* and *containment*), and operations on fuzzy sets (*complement*, *union*, and *intersection*), which become the kernel of the "fuzzy family" (*fuzzy set theory, fuzzy logic, fuzzy control system*, and so on).

What provided by these definitions are how to get the membership function of a compound set from the membership functions of its components (they are fuzzy sets themselves). For example, according to Zadeh, the membership function of *red flower* can be calculated from the membership functions of *red* and *flower* by applying the intersection operation, which is defined as *min* ([31]).

Therefore, a question will be risen naturally, that is, how to determine these functions at the first place, that is, for *red* and *flower*?

Zadeh suggested two ways to define a membership function ([31]):

1. By enumeratingly assigning membership values to objects in a domain. For example, the fuzzy concept *long-river* can be defined in the domain {*Nile, Hudson, Danube, Rhine, Mississippi*} as ([11]):

$$long\text{-}river = 1/Nile + 0.2/Hudson + 0.7/Danube + 0.4/Rhine + 0.8/Mississippi$$

2. By being a continuous and differentiable function of a numeral variable. For example, the membership of the fuzzy concept *old* is a function of the variable *age* ([31]):

$$old = \int_{50}^{100} (1 + (\frac{age - 50}{5})^{-2})^{-1}/age.$$

It is easy to see that both methods have preconditions: for the former, the domain of objects must be finite, and for the latter, there must be a measurable property that can serve as the variable of the function. Even when these preconditions are satisfied, there is still a problem: Where are these values come from? Anyway, people usually think and communicate without these numbers.

To answer the above question, we need to *interpret* fuzziness, that is, to answer the following questions:

- Why many (if not all) concepts are fuzzy?
- Why some instances have higher grades of membership than the others?
- What is measured by a grade of membership?

Here are Zadeh's opinions:

1. Fuzziness comes from the description of complex systems. He proposed the "Principle of Incompatibility" ([31]), which says that as the complexity of a system increases, our ability to make precise and yet significant statements about its behavior diminishes.

2. A membership function usually maps a continuous numeral variable to a distributed linguistic variable, so that the information can be summarized approximately. For example, "John is young" is an approximative way to say "John is 28". Since the underlying numeral variable changes continuously, there is no meaningful way to cut the boundary between the values of a linguistic variable. But, by membership function, we can describe the *compatibility* between a linguistic label and a numerical value, such as to say the compatibility between the label *young* and the age 28 to be 0.7 ([32, 34]).

3. Such a compatibility have no frequency interpretation. By "The membership of John's age to 'young' is 0.7", we don't mean that John's age is a random number, which takes the value "young" in 70% of the times. In [34], Zadeh proposed a "Possibility/probability consistency principle": a lessening of the possibility of an event tends to lessen its probability — but not vice versa.

4. Membership functions of primary terms are subjective and context-dependent, so there is no general method to determine them. "Their specification is a matter of definition, rather than objective experimentation or analysis". The task of fuzzy logic is to provide rules to compute the meaning of composite terms, once the meaning of the primary terms is specified in a given context ([30, 35]).

As a result, many totally different methods are used to get membership functions when fuzzy logic is applied to practical domains ([5]), which are chosen according to the designer's preference and experience.

2 Why to Interpret Fuzziness

Do we really need to further interpret the meaning (and origin) of membership values?

Yes, we do. From the standing point of artificial intelligence and cognitive science, at least we have the following reasons to require an interpretation ([5, 20]):

- Without a clear interpretation, it is hard for a computer system to generate the memberships automatically or to get them either from users or from sensory device. By "hard", I mean some values can be easily assigned, but they look quite arbitrary and artificial. In such a case, in what sense the system's results, which are determined by these initial assignments, are better than random choices?

- It is obvious that memberships are context dependent, and may be influenced by new knowledge. For example, "If 'Mary is young' is uttered in a kindergarten or in a retirement home situation, the effect on the expected age of Mary will be very different" ([4]). However, without a clear interpretation, there is no reasonable way to modify the memberships by new evidence, so they cannot be self-adjusted or be context-sensitive. On the other hand, it is unimaginable if the designer have to provide a system with a membership function for every concept (for instance, *young*) in every possible context (kindergarten, elementary school, ..., retirement home, even basketball team or cabinet) that the system may meet.

- The *max* and *min* operations, which are the most distinguish components of fuzzy theory, are not strongly supported by experimental evidence or theoretical consideration. They sometimes obviously lead to counter-intuitive results. For example, in [33] Zadeh define $big = long \wedge wide \wedge high$, but $\mu_{long} \times \mu_{wide} \times \mu_{high}$ looks much more reasonable than $\min\{\mu_{long}, \mu_{wide}, \mu_{high}\}$ as *big*'s membership function. Some psychological results are also inconsistent with the results predicted by the *min rule* ([16, 21]). Though there are some works show that *max* and *min* can be deduced from certain axioms ([2, 7]), it is still unclear that whether human cognitions really follows these axioms or why should we follow them.

- In his later papers ([3, 32]), Zadeh admitted that in some contexts the union/intersection operators should be *algebraic sum/product*, rather then *max/min*, but he didn't indicate *how* to determine with pair to used when facing a new context.

- The relationship between fuzziness and other types of uncertainty (such as probability and ignorance) is far from clearly explained. However, in practical problem solving, multiple types of uncertainty usually co-exist and merge with each other, as shown by the mixing of fuzziness (representativeness) and randomness (probability) in human judgments ([23, 20]).

In summary, fuzzy logic is not proposed as a pure formal system that only have some interesting mathematical properties, but as a formal model of fuzziness that happening all the time in human cognition, and as a tool that can handle this fuzziness for practical purposes. Why should we accept such a claim? The popular arguments are: (a) there are fuzziness in human cognition, (b) no frequency interpretation of the fuzziness has been found, and (c) some practical problems has been solved successfully by fuzzy logic ([36, 11]). Without a clear analysis of fuzziness in human cognition, these arguments are not enough for fuzzy logic to be accepted as a general cognitive model ([20]).

By an interpretation of fuzziness, I mean a mechanism by which membership can be explained, evaluated and adjusted. Such a mechanism should be able to relate, at least in principle, grade of membership to some more primary quantities in a psychologically plausible way, and concrete and formal enough to be implemented in a computer system. This doesn't mean to set up a universal and objective membership function for all concepts and all instances — membership is still subjective and context-dependent. What we need is an explanation about *how* it is influenced by the experience and the context of a system (human or computer).

Let's compare this issue with the case of probability. It's well known that there are various interpretations of probability, such as logical, empirical, and personal ([8, 10]). However, one idea is shared by the community: probability do need an interpretation, and all the operations carried on probabilities should be justified according to it. We cannot simply call it "degree of confirmation (or chance, belief)", then use any intuitively reasonable operations on it.

It is amazing that the same problem has not attracted enough attention in the fuzzy community, where much more efforts are spent on fuzzifying various mathematical tools and applying them to various practical domains.

From a theoretical point of view, the laking of an interpretation means that fuzziness is accepted as a matter of fact, but not clearly analyzed, so the operations on it seems quite arbitrary.

From a practical point of view, when fuzzy set theory and its variations are applied, the system designers are given enough freedom to choose membership functions and operators, usually in a try-and-error way. After a hand-tuning process, the system can do pretty well. However, the same methodology is hardly applicable when the system is general purpose, and the context is dynamic changed, not completely predictable by the designer. This suggests another explanation for why the most successful applications of fuzzy logic happen in some *control systems* ([11]), rather than in natural language processing, knowledge base management, general purpose reasoning, and machine learning ([6]), though in the latter domains fuzziness are more notable and more closely related to the initial idea of fuzzy set.

To solve the problem, we need to start by analyzing fuzziness.

3 Fuzziness from Relativity

At the very beginning, we need to distinguish two types of fuzziness: that mainly happens with adjectives and adverbs, and that mainly happens with nouns and verbs. Let's call them "type 1" and "type 2", respectively.

The difference between the two types is: though a concept of the type 2 can be treated as a fuzzy set with a relatively stable membership function, the same is not true for a concept of the type 1 — its membership function usually depends on the noun or verb it describing.

For example, if we treat "big" as a fuzzy set, just like "flea" and "animal", then "big flea" can be represented as $big \wedge flea$. From $\mu_{big \wedge flea}(A) = 0.9$, $\mu_{flea}(A) = 1$, and $flea \subset animal$, we get $\mu_{big \wedge animal}(A) = 0.9$, which is counter-intuitive ("A big flea is a big animal"). In other words, many adjectives are not *predicative* ([9]).

In AI community, this problem is usually explained by saying "the membership function of 'big' (as well as 'young', 'far', etc.) is context dependent". Of course it is, but *why* and *how*?

Now let's analyze and compare the following sentences:

1. "A is big."
2. "A is bigger than B."
3. "A is a big flea."
4. "A is a big animal."

Obviously, if there is no default or assumption about the context, "A is big" provides no information about A's size.

"A is bigger than B" does provide information about A's size, but in a relative way. The "bigger than" relation may become uncertain, due to incomplete information or imprecise measurement, but usually there is no fuzziness, since the relation is well-defined.

"A is a big flea" can be rephrased as "A is a flea, and it is big comparing with the other fleas". Similarly, "A is a big animal" can be rephrased as "A is an animal, and it is big comparing with the other animals". Now we can see that the information about A's size is also given in a *relative* way in this type of sentences.

The type 1 fuzziness appears exactly in this situation. "Bigger than" is a well-defined binary relation between two objects. When it is used between an object and a class of objects, uncertainty emerges. If we are told that "A is a big flea, and B is also a flea", then it is more plausible to assume that A is bigger than B than the reverse. However, there is uncertainty about whether the assumption will be conformed, since the concept "big flea" is fuzzy. Only when A is the biggest flea, can the uncertainty disappear, since we are sure that all other fleas are smaller.

Generally speaking, the fuzziness of type 1 appears in sentences with the pattern "A is a R C", where C is a class of objects, A is an object in C, and R is an adjective those comparative form "R-er than" is a binary relation on C, which is asymmetrical, transitive, and non-fuzzy. In such a case, "R C" is a fuzzy concept (such as "big flea", "tall men", and so on), because the information is given by comparing an object to a reference class.

Under such a situation, it is not a surprise to see that membership is a matter of degree, since "R C" means "R-er than the other Cs", whose truth value can be measured by an object's relative ranking in C with respect to the relation "R-er than".

There are many ways to represent information about relative ranking, but the most natural way is by a *ratio*

$$\mu_{R\_C}(A) = \frac{|(\{A\} \times C) \cap Rer\_than|}{|C - \{A\}|}$$

In the case of "big flea", A's membership is the ratio

the number of fleas that are smaller than A : the number of fleas minus 1 (it's not necessary to compare A to itself). Now $\mu_{R\_C}(A) = 1$ means that A is the biggest flea; $\mu_{R\_C}(A) = 0$ means that A is the smallest flea.

If the probability distribution of the size of fleas is given as $P(x)$, we can get a direct relation between the size of a flea, $S(A)$, and its grade of membership to "big flea", $\mu_{R\_C}(A)$:

$$\mu_{R\_C}(A) = \int_0^{S(A)} P(x)dx$$

Actually, this function identifies $\mu_{R\_C}(A)$ with the *percentage* of fleas that are smaller than A.

This equation can be generalized to all fuzzy concepts of the type 1 by considering $S(y) : C \rightarrow (-\infty, \infty)$ as a measurement corresponding to the relation "R-er than", and $P(x) : (-\infty, \infty) \rightarrow [0, 1]$ as the probability distribution of objects in C with respect to $S(y)$.

In this way, we get a function that calculates the membership of an object from a *fundamental argument*, as Zadeh did (see section 1). However, there is a basic difference. According to Zadeh, "The label *young* may be regarded as a *linguistic value* of the variable *age*, with the understanding that it plays the same role as the numerical value 25 but is less precise and hence less informative" ([32]). But here, it is interpreted as an approximate way to tell someone's relative youngfulness, with respect to a reference class. Only with a corresponding probability distribution, can the relative measurement be related to the absolute measurement. This is obvious in the above formula, where the reference class and a probability distribution are explicitly taken into consideration, while in Zadeh's formulas the context is implicit.

I say "A is a R C" provides information in a relative way, rather than in a absolute way, for the following reasons:

- As discussed previously, a word like "big" and "young" cannot be represented as a predicate or attribute that can be possessed by an object, but should be treated as a relation between objects.

- If "John is tall" is an approximate way to tell John's height, than it follows that this type of sentences is always less informative than the sentence like "John is 6 feet high". However, it is not always the case. For example, the sentence "To play basketball, tall players usually take advantage" cannot be rewritten by replacing "tall" by an accurate height, without lossing its generality. The sentence make the same sense in many contexts (from elementary school to MBA), where how "high" is mapped to height is drastically different.

- To say "A is a big flea", what one need to know it not A's size, but how it compares with other fleas. If A is the only known flea, we cannot say if it is a "big flea", even when we know its size exactly. On the contrary, if we always observe fleas through a magnifying glass, whose magnifying power is unknown, then we may have little idea about A's size, but "A is a big flea" still make sense. Actually, the sentence make the *same* sense, no matter how the sizes of fleas are distributed.

In practical usage, the context is often omitted in sentences. As a result, we only say "John is tall" or "A is big". Such omissions will cause problems in communication. If the default context of the speaker and that of the listener are different, misunderstandings will happen; if the listener doesn't sure what is the speaker's intended context, a guess has to be made, maybe according to the using frequency of various related contexts. Even when the reference class is explicitly in the sentence, as in "A is a big flea", it is still possible for the speaker and the listener to make different estimation about A's size, since due to personal experience, they may have different objects in mind when "flea" is mentioned. These factors cause uncertainty in communication, and they are closely related to fuzziness, but should not be confused with fuzziness, which (here I only mean the type 1) happens when an object is compared with a *class* of objects.

4 Fuzziness from Similarity

Now we'll turn to "type 2 fuzziness". This type happens mainly in nouns and verbs (such as "animal", "furniture", "to play", "to exist", and so on).

Psychologists have demonstrated the existence of fuzziness by well-documented experiments. It has been shown that people judge some instances to be better examples of a concept than some other instances are, and can answer category membership questions more rapidly for good examples than for poor examples ([18, 19, 15]).

Several theories are proposed by psychologists to explain the phenomena.

One explanation, *prototype theory*, suggests that from given members of a category, people abstract out the central tendency or *prototype* that becomes the summary mental representation for the category, then membership of a novel instance is measured by how similar it is to the prototype ([18, 19]).

Another explanation, *exemplar theory*, assume that membership of a novel instance is evaluated by directly comparing it with given members of the category ([12, 14]).

Generally speaking, the basic cause of the type 2 fuzziness is: the concept is not defined by sufficient/necessary conditions, but is exemplified by many objects/actions/events, which share common properties.

These results are often quoted as evidence in favor of fuzzy logic ([11]). However, exactly speaking, they only support the existence of fuzziness, rather than Zadeh's interpretation and suggested operations on it.

To psychologists, fuzziness, or grade of membership, is not a primary attribute of a concept that cannot be further analyzed. Rather, it is usually treated as a result that determined by some (more primary) factors, and there are rules that determine the membership evaluations ([18, 5, 12]).

More concretely, there is a consensus that grade of membership is determined by the degree of *similarity* between an instance to be judged and a prototype or a known instance, so (at least at the simplest cases) membership measurement is reduced to similarity measurement ([22, 14]).

Two kinds of similarities can be distinguished: those are symmetric and those are asymmetric ([22]). To avoid confusion, I'll define "inheritance relation" as an asymmetric similarity relation, and reserve the name "similarity relation" for the symmetric one. By "A inherit B's property", we mean that A has all the properties that B has, but not necessarily vice verse. By "A and B are similar", we mean that A has all the properties that B has, and vice verse.

How to measure the degree of inheritance and similarity? In the simplest case, let's assume that whether a object A has a property P is a matter of "all-or-none", and all properties of a object are equally weighted. Then the most natural measurement for the uncertainty in "A inherit B's property" is the "inheritance ratio" $\frac{|S_A \cap S_B|}{|S_B|}$, and the most natural measurement for the uncertainty in "A is similar to B" is $\frac{|S_A \cap S_B|}{|S_A \cup S_B|}$, where S_A and S_B are the set of properties for A and B, respectively (both formulas are special cases of *ratio model* of similarity, defined by Tversky in [22]).

Now, if we replace the A in the above two formulas by a variable X, the formulas become B's membership function, corresponding to the two interpretations of similarity (symmetric and asymmetric).

Various measurements of similarity (and inheritance, defined as above) have been suggested for different purposes ([22, 1, 14, 28, 13]). However, as long as they are defined in [0,1] (with 1 for "identical", and 0 for "completely different"), and are functions of weight of evidence, these measurements share the common form $\frac{w^+}{w^+ + w^-}$, that is, as the proportion (when all evidence is provided at the same time), or frequency (when evidence comes in a stream), of positive evidence. They only differ in the way that (positive and negative) evidence are defined and weighed. [1]

Such a membership evaluation depends on the system's experience and context. To a fuzzy concept, different (human or computer) systems may assign different properties to it, according to how the concept relate to the system's experience. On the other hand, when a system evaluate the degree of similarity of two concepts, usually only some of the properties, which are "activated" by the current situation, are taken into consideration. However, as in the previous section, these experience/context influences can be represented and processed explicitly, under the given interpretation.

5 A Unified Measurement of Uncertainty

In the previous sections, two interpretations are provided for the fuzziness of type 1 and 2, respectively. As a result, the membership functions of such fuzzy concepts are no longer come from intuition, preference, or experience of the system designer, but determined by the relevant evidence. Concretely, all the interpretations can be generalized into the following form:

$$\mu_c(x) = \frac{w^+}{w}$$

where C is the fuzzy concept, w is the weight of all relevant (to the membership relation) evidence, and w^+ is the weight of all positive evidence.

For a proposition like "A is a big flea", all fleas (except A) are relevant evidence, where fleas smaller than A are positive evidence of the proposition, and fleas bigger than A are negative evidence. The weight of evidence can be simply defined as the number of fleas under consideration.

For a proposition like "Penguins are birds", all properties of bird are relevant evidence, where those properties that shared by Penguin are positive evidence of the proposition, and whose properties that not shared by Penguin are negative evidence. The weight of evidence can be simply defined as the number of properties.

For a proposition like "A penguin and a robin are similar to each other", all properties of a penguin *or* a robin are relevant evidence, where those properties that shared by both are positive evidence of the proposition, and the properties that not shared are negative evidence. The weight of evidence can be simply defined as the number of properties.

Now we not only propose an interpretation for fuzziness, but also propose a *ratio*, or *frequency*, interpretation for it, which has be claimed as impossible by Zadeh. What follows naturally is the relation between fuzziness and randomness, where the latter is usually handled by probability

[1]Tversky's *contrast model* cannot be written into a ratio, but it is not a counter example of the above observation, since it is not defined on [0, 1].

theory. Probability is closely related to the ratio (or frequency, proportion) of positive evidence among all relevant evidence, so can also be represented as $\frac{w^+}{w}$, or its limits. Therefore, we can get an unified representation and interpretation for probability and membership: both are real numbers in $[0, 1]$, and both are indicating the ratio of positive evidence among all relevant evidence, that is, $\frac{w^+}{w}$. [2]

Based on such an interpretation, I'm building an intelligent reasoning system, Non-Axiomatic reasoning System, or NARS for short ([27, 26]), where fuzziness and randomness are uniformly processed as (part of) a judgment's *truth value*, and referred as the *frequency* of the judgment.

However, this doesn't mean that fuzziness and randomness cannot be distinguished. For a proposition "S is P", randomness always comes from the variety among the instances (or extension) of S, while fuzziness always comes from the variety among the properties (or intension) of P. When S has many instances, and some of them are P, while the others are not, "S is P" is a matter of degree, and the uncertainty is randomness; when P has many properties (or "intended meaning"), and some of them are possessed by S, while the others are not, "S is P" is also a matter of degree, but the uncertainty is fuzziness. In NARS, fuzziness and randomness are processed in a symmetrical way (see [26] for detail), and they are different in how the evidence is collected.

If these two types of uncertainty are different, why bother to treat them in an uniform way? The basic reason is: in many practical problems, they are involved with each other. Smets stressed the importance of this issue, and provided some examples, in which randomness and fuzziness are encountered in the same sentence ([20]). It is also true for inferences. Let's take medical diagnosis as an example. When a doctor want to determine whether a patient A is suffering from disease D, (at least) two types of information need to be taken into account: (1) whether A has D's symptoms, and (2) whether D is a common illness. Here (1) is evaluated by comparing A's symptoms with D's typical symptoms, so the result is usually fuzzy, and (2) is determined by previous statistics. After the total certainty of "A is suffering from D" is evaluated, it should be combined with the certainty of "T is a proper treatment to D" (which is usually a statistic statement, too) to get the doctor's "degree of belief" for "T should be applied to A". In such a situation (which is the usual case, rather than an exception), even if randomness and fuzziness can be distinguished in the premises, they are mixed in the middle and final conclusions.

Without an unified interpretation, it is still possible to set up rules for above operations, but such rules are not based on a consistent semantic foundation, therefore hard to be justified.

With a frequency interpretation of truth values, it is not surprise to see that NARS' truth value functions are the same for both randomness and fuzziness (as well as their "mixtures"), and the functions are defined more similar to probability theory than to fuzzy logic. For instance, the operations used for disjunction and conjunction are *algebraic sum/product*, rather than *max/min* ([26]). As a result, the truth value of the result is sensitive to the truth values of all the premises, so to avoid some counter-intuitive results of fuzzy logic ([16]).

This doesn't mean that a probability distribution on the proposition space is sufficient for representing the uncertainty in the system's knowledge base. To represent ignorance and to revise the system's belief in a general sense, a second number, *confidence*, is used in NARS, which is a real number in (0,1). Intuitively speaking, confidence indicates the stability of the judgment's frequency when challenged by new evidence, and is a function of the weight of total available evidence. For

[2] For small sample size, it is more reasonable to use a "squashed frequency" (in the form of $\frac{w^+ + k}{w + 2k}$), rather then the observed frequency itself, as the estimation of the probability ([8, 26]). However, the same is also true for fuzziness ([26]), so randomness and fuzziness can still be similarly processed.

detailed discussions on this issue, see [26] and [25].

6 Discussions

In this paper, a frequency interpretation of fuzziness is suggested, which has the following advantages:

- Different types of uncertainty, such as fuzziness, randomness, ignorance, and so on, can be processed by an unified mechanism;
- The membership function can be generated and modified by the system itself, according to its experience and the interpretation;
- The operators on uncertainty are no longer intuitively chosen, but can be justified according to the interpretation;
- Given a clear interpretation to the numbers, it is easier for the user to provide them when put new knowledge into a system, as well as to understand them when get results from a system.

Some people may argue that in this way, we'll lose one of the advantages of fuzzy logic, that is, the freedom for the system designer to determine the membership functions and operators. I disagree. At one hand, without an interpretation is a disadvantage for a theory, since it provide less guide for its users. On the other hand, with the frequency interpretation, a system (like NARS) can still be flexible. Because what the interpretation does is not to provide for each fuzzy set an "objective" membership function, but to indicate how such a function can be established and modified by the system according to its experience and the current context. In this sense, the interpretation takes some "freedom" from a human designer, and give it to the system itself.

There have been some attempts to interpret fuzziness in terms of probability. Let's mention two of them, and compare them with the approach of NARS.

1. Some people using polls to get membership function by identifying it with the percentage of people who agree to the membership relation ([5]). However, what this approach measures are actually the *degree of consensus* among a group of people. Though this type of uncertainty is related to fuzziness, as discussed before, it's not fuzziness.
2. Some proposers of Bayesian theory hope to handle fuzziness with probability theory by replace the frequency interpretation of probability by a personal one ([4]). They claim that fuzziness is just a type of "degree of belief". A problem of this approach is how to further explain the "degree of belief". On the other hand, since "Mary is young" is still treated as an approximate way to tell Mary's age in this approach, it is still unable to explain how the context influence the membership function.

Therefore, what proposed in this paper is not only that fuzziness can be interpreted as a frequency, but also that what kind to frequency it is.

Of course, the previous analysis of fuzziness is still far from complete. The type 1 and 2 are the simplest forms of fuzziness. Here are some more complicated cases:

1. For a fuzzy concept characterized by a set of properties, whether a object has a properties is usually a matter of degree, and the properties have different weights in determining the membership of an object. Therefore, the actual formula used (to determine membership) in NARS is a "weighted sum", rather than the simple "counting", as described above. The same is true for the case of type 1: whether two objects have a certain relation is often uncertain.

2. These two types of fuzziness are often co-exist in concepts, so should be combined. For example, a concept may be characterized by a set of properties shared by most of its instances (so it is "type 2 fuzzy"), and each property is a "type 1 fuzzy" concept itself.

3. Not all fuzzy concepts can be clearly classified as one of the two types defines above. For examples, perceptual categories (such as "red", "warm", and "soft") are neither defined in a completely relative way, nor defined purely by similarities, but depend heavily on human physiology ([18]). Without a physical sensory mechanism, these concepts are hard to handle for an AI system. For another example, I haven't found a natural way to manage fuzzy concepts in mathematics, like "real numbers which are much greater than 1", where the reference class is infinite.

4. Another interesting idea of Zadeh is *linguistic variables* ([32]). In NARS, linguistic variables can be used in the *interface language* by which the system communicates with users, but are not used in the *internal language* by which the system's knowledge is represented. Therefore, some translation rules are necessary for the mapping between these two languages. Such a mapping is possible due to the interpretation. For example, "John is young in C" can be translated into "John is younger than at least 2/3 of the others (in C)", but cannot be translated into something like "John's age is 18", since the mapping from (relative) youngfulness to (absolute) age depends on the reference class C. The concrete mapping function can be established by psychological experiments ([17, 24]).

Even with these problems in mind, we can still see the possibility to extend the interpretation of fuzziness proposed in this paper to those more complicated situations, so as to provide a frequency interpretation for fuzziness in general, and to process it consistently with other type of uncertainty.

Can NARS be referred as "a fuzzy logic"? Well, it is a matter of degree. The two approaches share some properties, but it seems that their differences weighs more than their similarities. Of course, it depends on the reference class of properties to be compared $\cdots$.

Acknowledgement

This work is supported by a research assistantship from Center for Research on Concepts and Cognition, Indiana University.

References

[1] G. Ashby and N. Perrin. Toward a unified theory of similarity and recognition. *Psychological Review*, 95:124–150, 1988.

[2] R. Bellman and M. Giertz. On the analytic formalism of the theory of fuzzy sets. *Information Science*, 5:149–157, 1973.

[3] R. Bellman and L. Zadeh. Decision-making in a fuzzy environment. *Management Science*, 17:141–164, 1970.

[4] P. Cheeseman. Probabilistic versus fuzzy reasoning. In L. Kanal and J. Lemmer, editors, *Uncertainty in Artificial Intelligence*, pages 85–102. North-Holland, Amsterdam, 1986.

[5] D. Dubois and H. Prade. *Fuzzy Sets and Systems*. Academic Press, New York, 1980.

[6] C. Elkan. The paradoxical success of fuzzy logic. In *Proceedings of the National Conference on Artificial Intelligence*, pages 698–703, 1993.

[7] B. Gaines. Fuzzy and probability uncertainty logics. *Information and Control*, 38:154–169, 1978.

[8] I. Good. *The Estimation of Probabilities*. The MIT Press, Cambridge, Massachusetts, 1965.

[9] J. Kamp. Two theories about adjectives. In E. Keenan, editor, *Formal Semantics of Natural Language*, pages 123–155. Cambridge University Press, Cambridge, 1975.

[10] H. Kyburg. *The Logical Foundations of Statistical Inference*. D. Reidel Publishing Company, Boston, 1974.

[11] D. McNeill and P. Freiberger. *Fuzzy Logic*. Simon & Schuster, New York, 1993.

[12] D. Medin and M. Schaffer. A context theory of classification learning. *Psychological Review*, 85:207–328, 1978.

[13] I. Niiniluoto. Analogy and similarity in scientific reasoning. In D. Helman, editor, *Analogical Reasoning*, pages 271–298. Kluwer Academic Publishers, Boston, 1988.

[14] R. Nosofsky. Typicality in logically defined categories: exemplar-similarity versus rule instantiation. *Memory and Cognition*, 17:444–458, 1991.

[15] G. Oden. Fuzziness in semantic memory: choosing exemplars of subjective categories. *Memory and cognition*, 5:198–204, 1977.

[16] G. Oden. Integration of fuzzy logical information. *Journal of Experimental Psychology: Human Perception and Performance*, 3:565–575, 1977.

[17] R. Reagan, F. Mosteller, and C. Youtz. Quantitative meanings of verbal probability expressions. *Journal of Applied Psychology*, 74:433–442, 1989.

[18] E. Rosch. On the internal structure of perceptual and semantic categories. In T. Moore, editor, *Cognitive Development and the Acquisition of Language*, pages 111–144. Academic Press, New York, 1973.

[19] E. Rosch and C. Mervis. Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7:573–605, 1975.

[20] P. Smets. Varieties of ignorance and the need for well-founded theories. *Information Sciences*, 57-58:135–144, 1991.

[21] E. Smith and D. Osherson. Conceptual combination with prototype concepts. *Cognitive Science*, 8:337–361, 1984.

[22] A. Tversky. Features of similarity. *Psychological Review*, 84:327–352, 1977.

[23] A. Tversky and D. Kahneman. Judgment under uncertainty: Heuristics and biases. *Science*, 185:1124–1131, 1974.

[24] T. Wallsten, D. Budescu, and R. Zwick. Comparing the calibration and coherence of numerical and verbal probability judgments. *Management Science*, 39:176–190, 1993.

[25] P. Wang. Belief revision in probability theory. In *Proceedings of the Ninth Conference on Uncertainty in Artificial Intelligence*, pages 519–526, Washington, DC, 1993. Morgan Kaufmann Publishers, San Mateo, California.

[26] P. Wang. From inheritance relation to non-axiomatic logic. Technical Report 84, Center for Research on Concepts and Cognition, Indiana University, Bloomington, Indiana, 1993.

[27] P. Wang. Non-axiomatic reasoning system (version 2.2). Technical Report 75, Center for Research on Concepts and Cognition, Indiana University, Bloomington, Indiana, 1993.

[28] P. Winston. Learning structural descriptions from examples. In P. Winston, editor, *The Psychology of Computer Vision*, pages 157–209. McGraw-Hill Book Company, New York, 1975.

[29] L. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

[30] L. Zadeh. A fussy-set-theoretic interpretation of linguistic hedges. *Journal of Cybernetics*, 2:4–34, 1972.

[31] L. Zadeh. Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Transactions on Systems, Man, and Cybernetics*, 3:28–44, 1973.

[32] L. Zadeh. The concept of a linguistic variable and its application to approximate reasoning. *Information Science*, pages 8:199–249, 8:301–357, 9:43–80, 1975.

[33] L. Zadeh. A fuzzy-algorithmic approach to the definition of complex or imprecise concepts. *International Journal of Man-Machine Studies*, 8:249–291, 1976.

[34] L. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and System*, 1:3–28, 1978.

[35] L. Zadeh. A theory of approximate reasoning. In J. Hayes, D. Michie, and L. Mikulich, editors, *Machine Intelligence*, volume 9, pages 149–194. Halstead Press, New York, 1979.

[36] L. Zadeh. Is probability theory sufficient for dealing with uncertainty in ai: a negative view. In L. Kanal and J. Lemmer, editors, *Uncertainty in Artificial Intelligence*, pages 103–116. North-Holland, Amsterdam, 1986.