1st Order, 2nd Order, What Next? We Do Not Really Need Third-Order Descriptions: A View From A Realistic (Granular) Viewpoint

Vladik Kreinovich
Comp. Science, Univ. of Texas at El Paso
El Paso, TX 79968, USA
vladik@cs.utep.edu

Hung T. Nguyen
Mathem. Sciences, New Mexico State Univ.
Las Cruces, NM 88003, USA
hunguyen@nmsu.edu

Abstract

To describe experts' uncertainty in a knowledge-based system, we usually use numbers from the interval $[0, 1]$ (*subjective probabilities*, *degrees of certainty*, etc.). The most direct way to get these numbers is to ask an expert; however, the expert may not be 100% certain what exactly number describes his uncertainty; so, we end up with a *second-order* uncertainty – a degree of certainty describing to what extent a given number d adequately describes the expert's uncertainty about a given statement A.

At first glance, it looks like we should not stop at this second order: the expert is probably as uncertain about his second-order degree as about his first-order one, so we need third order, fourth order descriptions, etc. In this paper, we show that from a realistic (granular) viewpoint, taking into consideration that in reality, an expert would best describe his degrees of certainty by a word from a finite set of words, it is sufficient to have a second-order description; from this viewpoint, higher order descriptions can be uniquely reconstructed from the second-order one, and in this sense, the second-order description is sufficient.

1. Second Order Descriptions: The Main Idea

Experts are often not 100% certain in the statements they make; therefore, in the design of knowledge-based systems, it is desirable to take this uncertainty into consideration. Usually, this uncertainty is described by a number from the interval $[0, 1]$; this number is called *subjective probability*, *degree of certainty*, etc. (see, e.g., [8]).

One of the main problems with this approach is that we must use *exact* numbers from the interval $[0, 1]$ to represent experts' degrees of certainty; an expert may be able to tell whether his degree of certainty is closer to 0.9 or to 0.5, but it is hardly possible that an expert would be able to meaningfully distinguish between degrees of certainty, say, 0.7 and 0.701. If you ask the expert whether his degree of certainty about a certain statement A can be described by a certain number d (e.g., $d = 0.701$), the expert will, sometimes, not be able to give a definite answer, she will be uncertain about it. This uncertainty can be, in its turn, described by a number from the interval $[0, 1]$. It is, therefore, natural to represent our degree of certainty in a statement A not by a *single* (crisp) number $d(A) \in [0, 1]$ (as in the $[0, 1]$-based description), but rather by a *function* $\mu_{\mathbf{d}(A)}$ which assigns, to each possible real number $d \in [0, 1]$, a *degree* $\mu_{\mathbf{d}(A)}(d)$ with which this number d can be the (desired) degree of certainty of A. This is called a *second-order* description of uncertainty.

2. Third and Higher Order Descriptions

In second-order description, to describe a degree with which a given number $d \in [0, 1]$ can be a degree of certainty of a statement A, we use a *real number* $\mu_{\mathbf{d}(A)}(d)$. As we have already mentioned, it is difficult to describe our degree of certainty by a single number. Therefore, to make this description even more realistic, we can represent each degree of certainty $d(P(x))$ not by a (more traditional) $[0, 1]$-based description, but by a *second order* description. As a result, we get the *third order* description.

Similarly, to make our description even more realistic, we can use the third order descriptions to describe degrees of certainty; then, we get *fourth order* uncertainty, etc.

3. Are Third Order Descriptions Really Necessary? Towards Formalization of the Problem

3.1. Third Order Descriptions Are Not Used: Why?

Theoretically, we can define third, fourth order, etc., descriptions, but in practical applications, only second order descriptions were used so far (see, e.g., [2, 3, 4, 6]). Based on this empirical fact, it is natural to conclude that third and higher order descriptions are not really necessary. We will show that this conclusion can be theoretically justified.

3.2. First Step in Describing Uncertainty: Set of Uncertainty-Describing Words

Let us first describe the problem formally. An expert uses words from a natural language to describe his degrees of certainty. In every language, there are only finitely many words, so we have a finite set of words that needs to be interpreted. We will denote this set of words by W.

3.3. Second Step: A Fuzzy Property Described by a Word-Valued "Membership Function"

If we have any property P on a universe of discourse U, an expert can describe, for each element $x \in U$, his degree of certainty $d(x) \in W$ that the element x has the property P.

3.4. Traditional Fuzzy Logic as a First Approximation: Numbers Assigned to Words Describing Uncertainty

Our ultimate goal is to provide a computer representation for each word $w \in W$. In the traditional $[0, 1]$-based description, this computer representation assigns, to every word, a *real number* from the interval $[0, 1]$; in general, we may have some other computer representations (examples will be given later). Let us denote the set of all possible computer representations by S.

In the first approximation, i.e., in the first order description, we represent each word $w \in W$, which describes a degree of uncertainty, by an element $s \in S$ (e.g., by a real number from the interval $[0, 1]$). In this section, we will denote this first-approximation computer representation of a word w by $s = \|w\|$.

If the set S is too small, then it may not contain enough elements to distinguish between different expert's degree of belief: this was exactly the problem with classical $\{0, 1\}$-based description, in which we only have two possible computer representations – "true" and "false" – that are not enough to adequately describe the different degrees of certainty. We will therefore assume that the set S is rich enough to represent different degrees of certainty.

In particular, the set $[0, 1]$ contains infinitely many points, so it should be sufficient; even if we only consider computer-representable real numbers, there are still much more of them (millions and billions) than words in a language (which is usually in hundreds of thousands at most), so we can safely make this "richness" assumption. In mathematical terms, it means that two different degrees of belief are represented by different computer terms, i.e., that if $w_1 \neq w_2$, then $\|w_1\| \neq \|w_2\|$.

3.5. First Approximation Is Not Absolutely Adequate

The problem with the first-order representation is that the relation between words $w \in W$ and computer representation $s \in S$ is, in reality, also imprecise. Typically, when we have a word $w \in W$, we cannot pick a single corresponding representative $s \in S$; instead, we may have *several* possible representatives, with different degrees of adequacy.

3.6. Actual Description of Expert Uncertainty: Word-Valued Degree to Which a Word Describes Uncertainty

In other words, instead of a *single* value $s = \|w\|$ assigned to a word w, we have *several* values $s \in S$, each with its own degree of adequacy; this degree of adequacy can also be described by an expert, who uses an appropriate word $w \in W$ from the natural language.

In other words, for every word $w \in W$ and for ever representation $s \in S$, we have a degree $w' \in W$ describing to what extent s is adequate in representing w. Let us represent this degree of adequacy by $a(w, s)$; the symbol a represents a function $a : W \times S \to W$, i.e., a function that maps every pair (w, s) into a new word $a(w, s)$.

3.7. Second-Order Description of Uncertainty as a Second Approximation to Actual Uncertainty

So, the meaning of a word $w \in W$ is represented by a *function* a which assigns, to every element $s \in S$, a degree of adequacy $a(w, s) \in W$. We want to represent this degree of adequacy in a computer; therefore, instead of using the word $a(w, s)$ itself, we will use the computer representation $\|a(w, s)\|$ of this word. Hence, we get a *second-order* representation, in which a degree of certainty corresponding to a word $w \in W$ is represented not by a *single* element $\|w\| \in S$, but by a *function* $\mu_w : S \to S$, a function which is defined as $\mu_w(s) = \|a(w, s)\|$.

3.8. Second-Order Description Is Not 100% Adequate Either; Third-, Fourth-Order Descriptions, etc.

The second-order representation is also not absolutely adequate, because, to represent the degree $a(w, s)$, we used a single number $\|a(w, s)\|$. To get a more adequate representation, instead of this single value, we can use, for each element $s' \in S$, a degree of adequacy with which the element s' represents the word $a(w, s)$. This degree of adequacy is also a word $a(a(w, s), s')$, so we can represent it by an appropriate element $\|a(a(w, s), s')\|$. Thus, we get a *third-order* representation, in which to every element s, we assign a second-order representation. To get an even more adequate representation, we can use fourth- and higher order representations.

Let us express this scheme formally.

4. Definitions and the Main Results

Definition 1.

- *Let W be a finite set; element of this set will be called* words.
- *Let U be set called a* universe of discourse. *By a* fuzzy property *P, we mean a mapping which maps each element $x \in U$ into a word $P(x) \in W$; we say that this word* described the degree of certainty *that x satisfies the property P.*
- *By a* first-approximation uncertainty representation, *we mean a pair $\langle S, \|.\|\rangle$, where:*
  - *S is a set; elements of this set will be called* computer representations; *and*
  - *$\|.\|$ is a function from W to S; we say that an element $\|w\| \in S$* represents *the word w.*
- *We say that an uncertainty representation is* sufficiently rich *if for every two words $w_1, w_2 \in W$, $w_1 \neq w_2$ implies $\|w_1\| \neq \|w_2\|$.*

Definition 2. *Let W be a set of words, and let S be a set of computer representations. By an* adequacy function, *we mean a function $a : W \times S \to W$; for each word $w \in W$, and for each representation $s \in S$, we say that $a(w, s)$ describes the degree to which the element s* adequately describes *the word w.*

Definition 3. *Let U be a universe of discourse, and let S be a set of computer representations. For each $n = 1, 2, \ldots$, we define the notions of* n-th order degree of certainty *and of a* n-th order fuzzy set, *by the following induction over n:*

- *By a* first-order degree of certainty, *we mean an element $s \in S$ (i.e., the set S_1 of all first-order degrees of certainty is exactly S).*
- *For every n, by a* n-th order fuzzy set, *we mean a function $\mu : U \to S_n$ from the universe of discourse U to the set S_n of all n-th order degrees of certainty.*
- *For every $n > 1$, by a* n-th order degree of certainty, *we mean a function s_n which maps every value $s \in S$ into an $(n-1)$-th order degree of certainty (i.e., a function $s_n : S \to S_{n-1}$).*

Definition 4. *Let W be a set of words, let $\langle S, \|.\|\rangle$ be an uncertainty representation, and let a be an adequacy function. For every $n > 1$, and for every word $w \in W$, we define the* n-th order degree of uncertainty *$\|w\|_{a,n} \in S_n$ corresponding to the word w as follows:*

- *As a first order degree of uncertainty $\|w\|_{a,1}$ corresponding to the word w, we simply take $\|w\|_{a,1} = \|w\|$.*
- *If we have already defined degrees of orders $1, \ldots, n-1$, then, as an n-th order degree of uncertainty $\|w\|_{a,n} \in S_n$ corresponding to the word w, we take a function s_n which maps every value $s \in S$ into a $(n-1)$-th order degree $\|a(w, s)\|_{a,n-1}$.*

Definition 5. *Let W be a set of words, let $\langle S, \|.\|\rangle$ be an uncertainty representation, let a be an adequacy function, and let P be a fuzzy property on a universe of discourse P. Then, by a* n-th order fuzzy set *(or a* n-th order membership function*) $\mu^{(n)}_{P,a}(x)$ corresponding to P, we mean a function which maps every value $x \in U$ into an n-th order degree of certainty $\|P(x)\|_{a,n}$ which corresponds to the word $P(x) \in W$.*

We will prove that for properties which are *non-degenerate* in some reasonable sense, it is sufficient to know the *first* and *second* order membership functions, and then the others can be uniquely reconstructed. Moreover, if we know the membership functions of first two orders for a non-degenerate *class* of fuzzy properties, then we will be able to reconstruct the higher order membership functions for *all* fuzzy properties from this class.

Definition 6.

- *We say that a fuzzy property P on a universe of discourse U is non-degenerate if for every $w \in W$, there exists an element $x \in U$ for which $P(x) = w$.*

- *We say that a class $\mathcal{P}$ of fuzzy properties P on a universe of discourse U is non-degenerate if for every $w \in W$, there exists a property $P \in \mathcal{P}$ and an element $x \in U$ for which $P(x) = w$.*

Comment. For example, if $W \neq \{0, 1\}$, then every crisp property, i.e., every property for which $P(x) \in \{0, 1\}$ for all x, is *not* non-degenerate (i.e., degenerate).

Proposition 1. *Let W be a set of words, let $\langle S, \|.\|\rangle$ be a sufficiently rich uncertainty representation, let U be a universe of discourse. Let P and P' be fuzzy properties, so that P is non-degenerate, and let a and a' be adequacy functions. Then, from $\mu^{(1)}_{P,a} = \mu^{(1)}_{P',a'}$ and $\mu^{(2)}_{P,a} = \mu^{(2)}_{P',a'}$, we can conclude that $\mu^{(n)}_{P,a} = \mu^{(n)}_{P',a'}$ for all n.*

Comments.

- In other words, under reasonable assumptions, for each property, the information contained in the first and second order fuzzy sets is sufficient to reconstruct all higher order fuzzy sets as well; therefore, in a computer representation, it is sufficient to keep only first and second order fuzzy sets.

- This result is somewhat similar to the well-known result that a Gaussian distribution can be uniquely determined by its moments of first and second orders, and all higher order moments can be uniquely reconstructed from the moments of the first two orders.

- It is possible to show that the non-degeneracy condition is needed, because if a property P is not non-degenerate, then there exist adequacy functions $a \neq a'$ for which $\mu^{(1)}_{P,a} = \mu^{(1)}_{P,a'}$ and $\mu^{(2)}_{P,a} = \mu^{(2)}_{P,a'}$, but $\mu^{(3)}_{P,a} \neq \mu^{(3)}_{P,a'}$ already for $n = 3$.

- For reader's convenience, all the proofs are placed in the last section of the paper.

- This result was first mentioned in our survey [5].

Proposition 2. *Let W be a set of words, let $\langle S, \|.\|\rangle$ be a sufficiently rich uncertainty representation, let U be a universe of discourse. Let $\mathcal{P}$ and $\mathcal{P}'$ be classes of fuzzy properties, so that the class $\mathcal{P}$ is non-degenerate, and let $\varphi : \mathcal{P} \to \mathcal{P}'$ be a 1-1-transformation, and let a and a' be adequacy functions. Then, if for every $P \in \mathcal{P}$, we have $\mu^{(1)}_{P,a} = \mu^{(1)}_{\varphi(P),a'}$ and $\mu^{(2)}_{P,a} = \mu^{(2)}_{\varphi(P),a'}$, we can conclude that $\mu^{(n)}_{P,a} = \mu^{(n)}_{\varphi(P),a'}$ for all n.*

Comment. So, even if we do not know the adequacy function (and we do not know the corresponding fuzzy properties $P \in \mathcal{P}$), we can still uniquely reconstruct fuzzy sets of all orders which correspond to all fuzzy properties P.

5. Proofs

5.1. Proof of Propositions 1 and 2

Proposition 1 can be viewed as a particular case of Proposition 2, when $\mathcal{P} = \{P\}$, $\mathcal{P}' = \{P'\}$, and φ maps P onto P'. Therefore, to prove both Propositions 1 and 2, it is sufficient to prove Proposition 2.

We will show that under the conditions of Proposition 2, from $\mu^{(1)}_{P,a} = \mu^{(1)}_{\varphi(P),a'}$ and $\mu^{(2)}_{P,a} = \mu^{(2)}_{\varphi(P),a'}$, we will be able to conclude that $\varphi(P) = P$ for all $P \in \mathcal{P}$, and that $a = a'$; therefore, we will easily conclude that $\mu^{(n)}_{P,a} = \mu^{(n)}_{\varphi(P),a'}$ for all n.

Indeed, by definition of the first membership function, for every $x \in U$, we have $\mu^{(1)}_{P,a}(x) = \|P(x)\|$. Thus, from the equality $\mu^{(1)}_{P,a} = \mu^{(1)}_{\varphi(P),a'}$, we conclude that for every $P \in \mathcal{P}$, we have $\|P(x)\| = \|\varphi(P)(x)\|$ for all $x \in U$. Since the uncertainty representation is assumed to be sufficiently rich, we can conclude that $\varphi(P)(x) = P(x)$ for all $x \in U$, i.e., that $\varphi(P) = P$ for every $P \in \mathcal{P}$.

Let us now show that $a = a'$, i.e., that for every $w \in W$ and for every $s \in S$, we have $a(w, s) = a'(w, s)$. Indeed, since $\mathcal{P}$ is a non-degenerate class, there exists a value $x \in U$ and a property $P \in \mathcal{P}$ for which $P(x) = w$. Let us consider the equality of the second order membership functions for this very P. Since $\varphi(P) = P$, the given equality $\mu^{(2)}_{P,a} = \mu^{(2)}_{\varphi(P),a'}$ can be simplified into the following form: $\mu^{(2)}_{P,a} = \mu^{(2)}_{P,a'}$. Let us consider this equality for the above-chosen value x (for which $P(x) = w$). For this x, by definition of the second-order membership function, $\mu^{(2)}_{P,a}(x) = \|P(x)\|_{a,2} = \|w\|_{a,2}$; and similarly, $\mu^{(2)}_{P,a'}(x) = \|P(x)\|_{a,2} = \|w\|_{a',2}$; thus, $\|w\|_{a,2} = \|w\|_{a',2}$.

By definition, $\|w\|_{a,2}$ is a function which maps every value $s \in S$ into a 1-st order degree $\|a(w,s)\|_{a,1} = \|a(w,s)\|$. Thus, from the equality of the functions $\|w\|_{a,2}$ and $\|w\|_{a',2}$, we can conclude that their values at a given s are also equal, i.e., that $\|a(w,s)\| = \|a'(w,s)\|$. Since the uncertainty structure is sufficiently rich, we conclude that $a(w,s) = a'(w,s)$. The proposition is proven.

5.2. Proof of a Comment After Proposition 1

Since P is *not* non-degenerate, there exists a value $w_0 \in W$ which cannot be represented as $P(x)$ for any $x \in U$. Let us pick arbitrary elements $x_0 \in U$ and $s_0 \in S$, and define $a(w,s)$ and $a'(w,s)$ as follows:

- first, we define $a(w,s) = a'(w,s)$ for all words w of the type $w = P(x)$: namely, we take $a(P(x_0),s_0) = a'(P(x_0),s_0) = w_0$ and take arbitrary other values for different pairs (w,s) with $w = P(x)$;
- then, we define $a(w,s)$ and $a'(w,s)$ for the remaining pairs (w,s): namely, we take $a(w_0,s_0) = w_0$, $a'(w_0,s_0) = P(x_0) \neq w_0$, and we define a and a' arbitrarily for all other pairs (w,s).

Let us show that for thus chosen adequacy functions, the membership functions of first and second order coincide, but the membership functions of the third order differ. Indeed:

- For the *first* order, we have, for every x, $\mu^{(1)}_{P,a}(x) = \|P(x)\|$ and similarly, $\mu^{(1)}_{P,a'}(x) = \|P(x)\|$; therefore, $\mu^{(1)}_{P,a}(x) = \mu^{(1)}_{P,a'}(x)$ for all x. Hence, $\mu^{(1)}_{P,a} = \mu^{(1)}_{P,a'}$.
- For the *second* order, for every x, $\mu^{(2)}_{P,a}(x)$ is a function which maps $s \in S$ into a value $\|a(P(x),s)\|_{a,1} = \|a(P(x),s)\|$. Similarly, $\mu^{(2)}_{P,a'}(x)$ is a function which maps $s \in S$ into a value $\|a'(P(x),s)\|_{a',1} = \|a'(P(x),s)\|$. For words w of the type $P(x)$, we have defined a and a' in such a way that $a(w,s) = a'(w,s)$; therefore, $\|a(P(x),s)\| = \|a'(P(x),s)\|$ for all x and s. Thus, $\mu^{(2)}_{P,a} = \mu^{(2)}_{P,a'}$.
- Finally, let us show that the *third* order membership functions differ. We will show that the values of the functions $\mu^{(3)}_{P,a}$ and $\mu^{(3)}_{P,a'}$ differ for $x = x_0$. Indeed, by definition of the third order membership function,
  - $\mu^{(3)}_{P,a}(x_0)$ is a function which maps every s into the value $\|a(P(x_0),s)\|_{a,2}$, and
  - $\mu^{(3)}_{P,a'}(x_0)$ is a function which maps every s into the value $\|a'(P(x_0),s)\|_{a',2}$.

  To prove that these function are different, it is sufficient to show that their values differ for *some* values s; we will show that they differ for $s = s_0$, i.e., that $\|a(P(x_0),s_0)\|_{a,2} \neq \|a'(P(x_0),s_0)\|_{a',2}$. By our construction of a, we have $a(P(x_0),s_0) = a'(P(x_0),s_0) = w_0$, so the inequality that we need to prove takes the form $\|w_0\|_{a,2} \neq \|w_0\|_{a',2}$.

  By definition, $\|w_0\|_{a,2}$ is a function which maps every value $s \in S$ into $\|a(w_0,s)\|_{a,1} = \|a(w_0,s)\|$. Similarly, $\|w_0\|_{a',2}$ is a function which maps every value $s \in S$ into $\|a'(w_0,s)\|_{a,1} = \|a'(w_0,s)\|$. For s_0, according to our construction of a and a', we have $a(w_0,s_0) = w_0 \neq P(x_0) = a'(w_0,s_0)$. Thus, since the uncertainty representation is sufficiently rich, we conclude that $\|a(w_0,s_0)\| \neq \|a'(w_0,s_0)\|$, and therefore, that $\|w_0\|_{a,2} \neq \|w_0\|_{a',2}$ and $\mu^{(3)}_{P,a} = \mu^{(3)}_{P,a'}$.

The statement is proven.

Acknowledgments

This work was supported in part by NASA under cooperative agreement NCC5-209, by Future Aerospace Science and Technology Program (FAST) Center for Structural Integrity of Aerospace Systems, effort sponsored by the Air Force Office of Scientific Research, Air Force Materiel Command, USAF, under grant number F49620-00-1-0365, and by Grant No. W-00016 from the U.S.-Czech Science and Technology Joint Fund.

References

[1] G. Klir and B. Yuan, *Fuzzy sets and fuzzy logic: theory and applications,* Prentice Hall, Upper Saddle River, NJ, 1995.

[2] J. Mendel, *Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions*, Prentice-Hall, Upper Saddle River, NJ, 2001.

[3] *NAFIPS/IFIS/NASA'94, Proceedings of the First International Joint Conference of The North American Fuzzy Information Processing Society Biannual Conference, The Industrial Fuzzy Control and Intelligent Systems Conference, and The NASA Joint Technology Workshop on Neural Networks and Fuzzy Logic, San Antonio, December 18–21, 1994*, IEEE, Piscataway, NJ, 1994.

[4] H.T. Nguyen and V. Kreinovich, "Towards theoretical foundations of soft computing applications", *Int'l J. on Uncertainty, Fuzziness, and Knowledge-Based Systems,* 1995, Vol. 3, No. 3, pp. 341–373.

[5] H.T. Nguyen and V. Kreinovich, "Possible new directions in mathematical foundations of fuzzy technology: a contribution to the mathematics of fuzzy theory", In: Nguyen Hoang Phuong and A. Ohsato (eds.), *Proceedings of the Vietnam-Japan Bilateral Symposium on Fuzzy Systems and Applications VJFUZZY'98*, HaLong Bay, Vietnam, 30th September–2nd October, 1998, pp. 9–32.

[6] H.T. Nguyen, V. Kreinovich, and Q. Zuo, "Interval-valued degrees of belief: applications of interval computations to expert systems and intelligent control", *International Journal of Uncertainty, Fuzziness, and Knowledge-Based Systems (IJUFKS)*, 1997, Vol. 5, No. 3, pp. 317–358.

[7] H.T. Nguyen and E.A. Walker, *A first course in fuzzy logic,* CRC Press, Boca Raton, Florida, 1999.

[8] G. Shafer and J. Pearl (eds.) *Readings in Uncertain Reasoning*, M. Kaufmann, San Mateo, CA, 1990.