When Should We Switch from Interval-Valued Fuzzy to Full Type-2 Fuzzy (e.g., Gaussian)?

Vladik Kreinovich[1] and Chrysostomos D. Stylios[2]
[1]Department of Computer Science
University of Texas at El Paso
500 W. University
El Paso, Texas 79968, USA
vladik@utep.edu
[2]Laboratory of Knowledge and Intelligent Computing
Department of Computer Engineering
Technological Educational Institute of Epirus
47100 Kostakioi, Arta, Greece
stylios@teiep.gr

Abstract

Full type-2 fuzzy techniques provide a more adequate representation of expert knowledge. However, such techniques also require additional computational efforts, so we should only use them if we expect a reasonable improvement in the result of the corresponding data processing. It is therefore important to come up with a practically useful criterion for deciding when we should stay with interval-valued fuzzy and when we should use full type-2 fuzzy techniques. Such a criterion is proposed in this paper. We also analyze how many experts we need to ask to come up with a reasonable description of expert uncertainty.

1 Formulation of the Problem

Need for fuzzy logic. In many application areas, we have expert knowledge formulated by using imprecise ("fuzzy") words from natural language, such as "small", "weak", etc. To use this knowledge in automated systems, it is necessary to reformulate it in precise computer-understandable terms. The need for such a reformulation was one of the motivations behind *fuzzy logic* (see, e.g., [3, 11, 15]). Fuzzy logic uses the fact that in a computer, "absolutely true" is usually represented as 1, and "absolutely false" is represented as 0. Thus, to describe expert's intermediate degrees of confidence, it makes sense to use real numbers intermediate between 0 and 1.

In this case, to represent an imprecise word like "small", we describe, for each real number x, the degree $\mu_{\text{small}}(x) \in [0,1]$ to which the expert considers this value to be small. The corresponding function from the set of possible value to the interval $[0,1]$ is known as a *membership function.*

Need to go beyond $[0,1]$-valued fuzzy logic. In most practical problems, we have several experts, and while their imprecise rules may coincide, their understanding of the meaning of the corresponding words may be slightly different. As a result, when we ask different experts, we get, in general, different membership functions corresponding to the same term – i.e., for each possible value x, we get, in general, different degrees $\mu(x)$ (describing the expert's opinion to what extent this value x satisfies the given property).

To adequately represent expert knowledge, it is desirable to capture this difference, i.e., to go beyond the original $[0,1]$-valued fuzzy logic – which was oriented towards capturing the opinion of a single expert.

Interval-valued fuzzy techniques. If for the same property P and for same value x, two different degrees of confidence, e.g., 0.6 and 0.8, are both possible – according to two experts – then it makes sense to assume that for other experts, intermediate viewpoints will also be possible. In other words, if two real numbers from the interval $[0,1]$ are possible degrees, then all intermediate numbers should also be possible degrees. In this case, for each property P and for each value x, the set of all possible degree that x satisfies the property P is an interval. This interval can be denoted by $\left[\underline{\mu}(x), \overline{\mu}(x)\right]$.

Interval-valued fuzzy techniques have indeed been successfully used in many applications; see, e.g., [7, 8, 10].

General type-2 fuzzy techniques. The interval-valued techniques do not fully capture the uncertainty of the experts' opinion: these techniques just describe the interval, but they do not take into account that some values from this interval are shared by many experts, while other values are "outliers", opinions of a few unorthodox experts. To capture this difference, a reasonable idea is to describe, for each value μ from the corresponding interval $\left[\underline{\mu}, \overline{\mu}\right]$, a degree to which this value is common.

In other words, for each possible value x of the original quantity, instead of single numerical degree $\mu(x)$, we now have a *fuzzy set* (membership function) describing this degree. Such situation in which, for every possible value x of the original quantity, the experts' degree of confidence that x satisfies the given property P is itself a fuzzy number is known as *type-2 fuzzy set.*

Of course, each interval-valued fuzzy set is a trivial particular case of the general type-2 fuzzy set, corresponding to the case when the degree is 1 inside the interval $\left[\underline{\mu}, \overline{\mu}\right]$ and 0 outside this interval.

The most commonly used non-trivial type-2 fuzzy sets are the *Gaussian* ones, in which, for each x, the corresponding membership function of the set of all possible values μ is Gaussian: $d(\mu) = \exp\left(-\dfrac{(\mu-\mu_0)^2}{2\sigma^2}\right)$ for some values μ_0 and σ. Such Gaussian-valued fuzzy sets are also used in applications [7, 8].

Comment. In addition to empirical success, there are also theoretical reasons why namely Gaussian membership functions are successfully used; see, e.g., [4].

Formulation of the problem.

- On the one hand, the transition from interval-valued to general type-2 fuzzy sets leads to a more adequate representation of the expert's knowledge. From this viewpoint, it may sound as if it is always beneficial to use general type-2 fuzzy sets.
- However, on the other hand, this transition requires that we store and process additional information about the secondary membership functions. So, we should only perform this switch if we expect a reasonable advantage.

It is therefore desirable to come up with a criterion for deciding when we should switch from interval-valued fuzzy to general type-2 fuzzy. The main objective of this paper is to come up with such a criterion.

Comment. A similar problem occurs in describing measurement uncertainty: we can simply store and use the interval of possible values of measurement error, or we may want to supplement this interval withe the information about the probability of different values within this interval – i.e., with a probability distribution. Here also, we face a similar problem of deciding when it is beneficial to switch form a simpler interval description to a more complex (but more adequate) probabilistic description. A possible solution to this problem – based on information theory – is presented in [1].

2 Analysis of the Problem

One more reason why Gaussian membership functions provide a good description of the expert diversity. There are many different factors that influence the expert's degree of confidence. The actual degree produced by an individual expert is a result of the joint effect of all these factors.

Such situations, when a quantity is influenced by many different factors, are ubiquitous. There is a known result – the Central Limit Theorem (see, e.g., [13]) – that helps to describe such situations, by proving that, under reasonable assumptions, the probability distribution of the joint effect of many independent factors is close to Gaussian. This is a well-known fact explaining the ubiquity of bell-shaped Gaussian (normal) distributions: they describe the distribution of people by height, by weight, by IQ, they describe the distribution of different animals and plants, they describe the measurement errors, etc.

It is therefore reasonable to assume that when we consider many experts providing their degrees of confidence, the resulting probability distribution of these degrees is also close to Gaussian (= normal), with some mean μ_0 and standard deviation σ.

For normally distributed expert estimates, what is the corresponding interval? Let us assume that for the same statement, different expert degrees of confidence are normally distributed with mean μ_0 and standard deviation σ. Let N denote the number of experts whose opinions we ask, and let $\mu_1, \ldots, \mu_N$ are degrees indicated by these experts.

If we use an interval approach, then, as the interval-valued degree of confidence $\left[\underline{\mu}, \overline{\mu}\right]$, we take the interval formed by these degrees μ_i, i.e., the interval $\left[\min\limits_i \mu_i, \max\limits_i \mu_i\right]$.

On average, when we have a sample of N random values, then one of the ways to approximate the original distribution is to build a histogram, i.e., sort the observed values μ_i in increasing order into a sequence

$$\mu_{(1)} < \mu_{(2)} < \ldots < \mu_{(N)},$$

and then take a distribution that has each of the values $\mu_{(i)}$ with the same probability $\dfrac{1}{N}$. It is known that in the limit $N \to \infty$, this histogram distribution distribution converges to the actual distribution (i.e., becomes closer and closer as N increases).

Thus, as a good approximation to the smallest possible value $\mu_{(1)} = \min\limits_i \mu_i$, it is reasonable to take the value $\underline{v}$ for which the probability $\text{Prob}(v \le \underline{v}) = \dfrac{1}{N}$. Similarly, as a good approximation to the largest possible value $\mu_{N)} = \max\limits_i \mu_i$, we can take the value $\overline{v}$ for which $\text{Prob}(v \ge \overline{v}) = \dfrac{1}{N}$, i.e., for which

$$\text{Prob}(v \le \overline{v}) = 1 - \frac{1}{N}.$$

For a normal distribution with mean μ_0 and standard deviation σ, the corresponding values $\underline{v}$ and $\overline{v}$ can be obtained as follows (see, e.g., [13]):

$$\underline{v} = \mu_0 - k(N) \cdot \sigma; \quad \overline{v} = \mu_0 + k(N) \cdot \sigma,$$

where

$$k(N) \stackrel{\text{def}}{=} \sqrt{2} \cdot \text{erf}^{-1}\left(1 - \frac{2}{N}\right),$$

and the *error function* $\text{erf}(x)$ is defined as

$$\text{erf}(x) \stackrel{\text{def}}{=} \int_{-x}^{x} \exp\left(-\frac{t^2}{2}\right) dt.$$

So, when is interval representation better? The value $k(N)$ increases with N and tends to ∞ when N increases. Thus, when the number of experts N is large, the lower endpoint of the interval

$$\left[\underline{\mu}, \overline{\mu}\right] = [\mu_0 - k(N) \cdot \sigma, \mu_0 + k(N) \cdot \sigma]$$

becomes negative, while its upper bound becomes larger than 1. Since the values μ_i are always located within the interval $[0,1]$, in this case, the interval-valued description of uncertainty is useless: the smallest value is 0 (or close to 0), the largest value is 1 (or close to 1). In such situations, we cannot use the interval-valued approach, so we need to use a more computationally complex Gaussian approach.

On the other hand, if we have

$$0 < \underline{\mu} = \mu_0 - k(N)\cdot\sigma \text{ and } \overline{\mu} = \mu_0 + k(N)\cdot\sigma < 1,$$

then, once we know the bounds $\underline{\mu}$ and $\overline{\mu}$, we can uniquely reconstruct both parameters μ_0 and σ as follows:

$$\mu_0 = \frac{\underline{\mu}+\overline{\mu}}{2}; \quad \sigma = \frac{\overline{\mu}-\underline{\mu}}{2k(N)}.$$

In this case, if we use the interval-valued approach, we do not lose any information in comparison with the Gaussian-based approach. Since the interval-valued approach is computationally easier than the Gaussian-based approach, it therefore makes sense to use the interval-based approach.

But are these expert estimates meaningful at all? What if the expert do not real have any knowledge, and their degrees are all over the map? In this case, processing these ignorance-based degrees does not make any sense. How can we detect such a situation?

In the cases when experts have no meaningful knowledge, their degree are simply uniformly distributed on the interval $[0,1]$. In this case, the variance is equal to $\sigma^2 = \dfrac{1}{12}$, in which case $\sigma \approx 0.3$. So, we can conclude that if the empirical standard deviation is greater than or equal to 0.3, then we should simply ignore the experts' degrees – since the experts' opinions disagree too much to be useful.

Thus, we arrive at the following recommendation.

3 Recommendation: When to Use Interval-Valued Approach and When to Use Gaussian Approach

What is given. For each property and for each possible value x, we have N experts that provide us with their degrees of confidence $\mu_1,\ldots,\mu_N$ that this value x satisfies the given imprecise property (e.g., that this value x is small).

Resulting algorithm. First, we use the standard formulas to estimate the mean μ_0 and standard deviation σ of the expert's degrees μ_i:

$$\mu_0 = \frac{\mu_1+\ldots+\mu_N}{N};$$

$$\sigma = \sqrt{\frac{1}{N-1} \cdot \sum_{i=1}^{N} (\mu_i - \mu_0)^2}.$$

If $\sigma \geq 0.3$, then we conclude that the experts' opinion disagree too much to be useful.

If $\sigma < 0.3$, then, based on the number of experts N, we estimate $k(N)$ as

$$k(N) = \sqrt{2} \cdot \operatorname{erf}^{-1}\left(1 - \frac{2}{N}\right).$$

Based on this value $k(N)$, we compute the values

$$\underline{\mu} = \mu_0 - k(N) \cdot \sigma \text{ and } \underline{\mu} = \mu_0 + k(N) \cdot \sigma.$$

Then:

- if $0 < \underline{\mu}$ and $\overline{\mu} < 1$, we use interval-valued approach, with interval-valued degree $[\underline{\mu}, \overline{\mu}]$;
- otherwise, if $\underline{\mu} \leq 0$ or $\overline{\mu} \geq 1$, we use a Gaussian approach, with the type-2 Gaussian degree of confidence

$$d(\mu) = \exp\left(-\frac{(\mu - \mu_0)^2}{2\sigma^2}\right).$$

4 Auxiliary Question: How Many Experts We Should Ask?

How many experts we should ask? For a general random variable, the larger the sample, we more accurate the estimates. For example, if we perform measurements, then we can decrease the random component of the measurement error if we repeat the measurement many times and take the average of the measurement results. This fact follows from the Large Numbers Theorem, according to which, when the sample size increases, the sample average tends to the mean of the corresponding random variable.

This makes sense if we deal with measurements of physical quantities, where more and more accurate description of this quantity makes perfect sense – and is desirable. For degree, however, the situation is different. A person can only provide his or her degree of confidence only with a low accuracy: e.g., an expert may distinguish between marks 6 and 7 on a scale from 0 to 10, but, when describing their degree of confidence, experts cannot meaningfully distinguish between, e.g., values 61 and 62 on a scale from 0 to 100.

Comment. Issues related to decision making in fuzzy context are handled, e.g., in [2, 5, 6].

Our idea. Psychologists have found out that we usually divide each quantity into 7 plus plus minus 2 categories – this is the largest number of categories whose meaning we can immediately grasp; see, e.g., [9, 12] (see also [14]). For some people, this "magical number" is $7 + 2 = 9$, for some it is $7 - 2 = 5$. This rule is in good accordance with the fact that in fuzzy logic, to describe the expert's opinion on each quantity, we usually use 7 ± 2 different categories (such as "small", "medium", etc.).

Since on the interval $[0, 1]$, we can only have 7 ± 2 meaningfully different degrees of confidence, the accuracy of these degrees ranges is, at best, $1/9$. When we estimate the mean μ_0 based on N values, the accuracy is of order $\dfrac{\sigma}{\sqrt{N}}$. It does not make sense to bring this accuracy below $1/9$, so it makes sense to limit the number of experts N to a value for which $\dfrac{\sigma}{\sqrt{N}} \approx \dfrac{1}{9}$, i.e., to the value $N \approx (9 \cdot \sigma)^2$.

Resulting recommendation. To estimate how many experts we need to ask, we ask a small number n of experts, and, based on their degrees μ_i, estimate σ as

$$\sigma = \sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^{n} (\mu_i - \mu_{\rm av})^2}, \text{ where } \mu_{\rm av} = \frac{1}{n} \cdot \sum_{i=1}^{n} \mu_i.$$

Then, we estimate the number N of experts to ask as $N = (9 \cdot \sigma)^2$.

Comment. Of course, if $N \le n$, this means that we do not have to ask any more experts, whatever information we have from n experts is enough.

Examples. If all experts perfectly agree with each other, i.e., if $\mu_i = \mu_j$ for all i and j, then $\sigma = 0$ and $N = 0$. In this cases, there is no need to ask any more experts.

Similarly, if all experts more or less agree with each other and $\sigma = 0.1$, then $N < 1$, meaning also that there is no need to ask more experts.

If $\sigma = 0.2$, then $N = 3.61$, meaning that we should ask at leat 4 experts to get a good estimate. For $\sigma = 0.3$, we get $N = 7.29$, meaning that we need to ask at least 7 experts.

This is about as bad as we can get: as we have mentioned, even when the expert's degrees are all over the map, i.e., uniformly distributed on the interval $[0, 1]$, then the variance is equal to $\sigma^2 = \dfrac{1}{12}$, in which case $\sigma \approx 0.3$, and we get $N = 9^2 \cdot \sigma^2 = \dfrac{81}{12} = 6.75$, meaning that we need to ask at most 7 experts.

Conclusions. In all cases, we need to ask at most seven experts to get a meaningful estimate (and sometimes, when the experts agree with each other, a smaller number of experts is sufficient).

Acknowledgment

This work was supported in part by the National Science Foundation grants HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence) and

DUE-0926721. This work was performed when C. Stylios was a Visiting Researcher at the University of Texas at El Paso.

References

[1] M. Beer and V. Kreinovich, "Interval or moments: which carry more information?", *Soft Computing*, 2013, Vol. 17, No. 8, pp. 1319–1327.

[2] R. E. Bellman and L. A. Zadeh, "Decision-making in a fuzzy environment", *Management Science*, 1970, Vol. 17, No. 4., pp. 141–164.

[3] G. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic*, Prentice Hall, Upper Saddle River, New Jersey, 1995.

[4] V. Kreinovich, C. Quintana, and L. Reznik, "Gaussian membership functions are most adequate in representing uncertainty in measurements", *Proceedings of NAFIPS'92: North American Fuzzy Information Processing Society Conference*, Puerto Vallarta, Mexico, December 15–17, 1992, NASA Johnson Space Center, Houston, TX, 1992, pp. 618–625.

[5] F. Mata, L. Martnez, and E. Herrera-Viedma, "An adaptive consensus support model for group decision-making problems in a multigranular fuzzy linguistic context", *IEEE Transactions on Fuzzy Systems*, 2009, Vol. 17, No. 2, pp. 279–290.

[6] L. Medsker, M. Tan, and E. Turban, "Knowledge acquisition from multiple experts: Problems and issues", *Expert Systems with Applications*, 1995, Vol. 9, No. 1, pp. 35–40.

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

[8] J. M. Mendel and D. Wu, *Perceptual Computing: Aiding People in Making Subjective Judgments*, IEEE Press and Wiley, New York, 2010.

[9] G. A. Miller, "The magical number seven, plus or minus two: some limits on our capacity for processing information", *Psychological Review*, 1956, Vol. 63, No. 2, pp. 81–97.

[10] 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.

[11] H. T. Nguyen and E. A. Walker, *A First Course in Fuzzy Logic*, Chapman and Hall/CRC, Boca Raton, Florida, 2006.

[12] S. K. Reed, *Cognition: Theories and Application*, Wadsworth Cengage Learning, Belmont, California, 2010.

[13] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman & Hall/CRC, Boca Raton, Florida, 2011.

[14] R. Trejo, V. Kreinovich, I. R. Goodman, J. Martinez, and R. Gonzalez, "A realistic (non-associative) logic and a possible explanations of 7±2 law, *International Journal of Approximate Reasoning*, 2002, Vol. 29, pp. 235–266.

[15] L. A. Zadeh, "Fuzzy sets", *Information and Control*, 1965, Vol. 8, pp. 338–353.