

Probability-Based Approach Explains (and Even Improves) Heuristic Formulas of Defuzzification*

Christian Servin¹[0000–0001–9965–4386], Olga Kosheleva²[0000–0002–1244–1650],
and
Vladik Kreinovich²[0000–0002–1244–1650]

¹ Computer Science and Information Technology Systems Department, El Paso Community College, 919 Hunter, El Paso, TX 79915, USA cservin@gmail.com

² University of Texas at El Paso, El Paso, TX 79968, USA
olgak@utep.edu, vladik@utep.edu

Abstract. Fuzzy techniques have been successfully used in many applications. However, often, formulas for processing fuzzy information are often heuristic: they lack a convincing justification, and thus, users are sometimes reluctant to use them. In this paper, we show that we can justify (and sometimes even improve) these methods if we use a probability-based approach.

Keywords: Defuzzification · Probabilistic approach · Heuristic algorithms · Optimality

1 Formulation of the Problem

Need for fuzzy knowledge. In many practical situations, ranging from medicine to driving, we rely on expert knowledge of how to cure diseases, how to drive in a complex city environment, etc. Some medical doctors are more qualified than others, some drivers are more skilled than others. It is therefore desirable to incorporate their skills and their knowledge in a computer-based system that will help other experts perform better – and ideally, make expert-quality decisions on its own, without the need for the experts.

One of the main obstacles to designing such a system is the fact that experts usually formulate their knowledge by using imprecise (“fuzzy”) words from natural language like “close”, “fast”, “small”, etc., and computers are not efficient in processing words, they are much more efficient in processing numbers. It is therefore desirable to represent the natural-language fuzzy knowledge in numerical terms.

A technique for such a representation was proposed in the 1960s by Lotfi Zadeh from Berkeley under the name of *fuzzy logic*. In fuzzy logic, to represent each word like “small” in numerical terms, we assign, to each possible value x

* This work was supported in part by the US National Science Foundation grant HRD-1242122 (Cyber-ShARE Center of Excellence).

of the corresponding quantity, a degree $\mu(x) \in [0, 1]$ to which, in the expert's option, the value x can be described by this word (e.g., to what extent x is small); see, e.g., [1, 4, 7, 9, 10, 13].

Where fuzzy degrees come from. There are many different ways to elicit the desired degrees.

If we are just starting the analysis and we do not have any records, then we can ask an expert to mark, on a scale, say, from 0 to 10, to what extent x is small. If the expert marks 7, we take 7/10 as the desired degree.

Usually, however, we already have a reasonably large database of records in which the experts used the corresponding word to describe different values of the corresponding quantity x . For values x which are really small, we will have a large number of such records; for values x which are not too small, we will have a few such records. Based on the available records, we can estimate the probability density function (pdf) $\rho(x)$ that describes the frequency with which different values x appear in our records.

When x is really small, the value $\rho(x)$ is big; when x is not so small – so that fewer experts will consider this value to be small – the value $\rho(x)$ is much smaller. Thus, in principle, we could use the values $\rho(x)$ as the desired degrees. However, we want values from the interval $[0, 1]$, while the pdf can take values larger than 1. To make all the value smaller than or equal to 1, we can normalize these values, i.e., divide by the largest of them. As a result, we get

$$\mu(x) = \frac{\rho(x)}{\max_y \rho(y)}.$$

Need for defuzzification. By using expert knowledge transformed into the numerical form, we can determine, for each possible value u of the control, the degree $\mu(u)$ to which this value is reasonable.

These degrees can help an expert make better decisions. However, if we want to make an automatic system, we must select a single value u that the system will apply. Selecting such a value is known as *defuzzification*.

Centroid defuzzification: description, successes, and limitations. The most widely used defuzzification procedure is *centroid* defuzzification, in which we select the value

$$\bar{x} = \frac{\int x \cdot \mu(x) dx}{\int \mu(x) dx}.$$

It has led to many successful applications of fuzzy control; see, e.g., [1, 4, 7, 9, 10]. However, it has two related limitations:

- First, it is heuristic, it is not justified by a precise argument and therefore, we are not sure whether it will always work well.
- Second, it sometimes leads to disastrous results. For example, when a car encounters an obstacle on an empty road, it can go around it by veering to the left or by veering to the right. The situation is completely symmetric

with respect to the direction to the obstacle. As a result, the centroid will lead exactly to the center – i.e., smack into the obstacle. The actual fuzzy control algorithms use some techniques to avoid such a situation, but these techniques are also heuristic – and thus, not guaranteed to produce good results.

Optimization under fuzzy constraints. Another class of situations in which fuzzy knowledge is important is optimization.

Traditional optimization techniques allow us to find the values of the parameters x for which the objective function attains its optimal value – largest or smallest depending on the problem. These techniques assume – explicitly or implicitly – that all possible combinations x are possible.

In practice, there are usually *constraints* restricting possible combinations. In some cases, constraints are formulated in precise terms – for example, there are regulations limiting noise level and pollution level from a plant. There are well-known techniques for dealing with such constraints – e.g., the Lagrange multiplier method that reduced the problem of optimizing an objective function $f(x)$ under constraint $g(x) = 0$ to the unconstrained optimization problem of optimizing the auxiliary objective function $f(x) + \lambda \cdot g(x)$, for an appropriate parameter λ .

Often, however, we also have imprecise (fuzzy) constraints. For example, a company that designs a plant in a city usually wants not just to satisfy all the legal requirements, but also to keep good relation with the city, and one way to do it is to make sure that the noise level is not high. This “not high” is clearly an example of an imprecise constraint.

Another case when fuzzy constraints are important is when one of the objectives is to make customers happy. For example, an elevator must be reasonable fast but also reasonably smooth.

We can describe the fuzzy constraint by a membership function $\mu(x)$, so that for each possible combination x of the corresponding parameters, $\mu(x)$ is a degree to which the alternative corresponding to these parameter values satisfies the constraint. How can we optimize an objective function $f(x)$ under such fuzzy constraints?

A well-known heuristic solution to this problem was proposed in a joint paper [2] by Lotfi Zadeh (the father of fuzzy logic) and Richard Bellman (the famous specialist in optimization): namely, to maximize an objective function $f(x)$ under fuzzy constraints, they proposed to maximize an auxiliary function

$$f_{\&}\left(\mu(x), \frac{f(x) - m}{M - m}\right),$$

where $f_{\&}(a, b)$ is usually either the minimum $\min(a, b)$ or the product $a \cdot b$, and m and M are, correspondingly, the minimum and the maximum of $f(x)$ over the set X of all theoretically possible combinations x :

$$m \stackrel{\text{def}}{=} \min_{x \in X} f(x), \quad M \stackrel{\text{def}}{=} \max_{x \in X} f(x).$$

The above formula is used when we want to maximize $f(x)$; when we want to minimize $f(x)$, then we maximize a slightly different auxiliary expression – which comes from the fact that minimizing $f(x)$ is equivalent to maximizing an auxiliary function $f'(x) \stackrel{\text{def}}{=} -f(x)$:

$$f_{\&} \left(\mu(x), \frac{M - f(x)}{M - m} \right).$$

Similarly to defuzzification:

- on the one hand, these heuristic formulas have led to many useful application, but
- on the other hand, the fact that these formulas are heuristic – and thus, lack a convincing justification – makes users often somewhat reluctant to use them.

What we do in this paper. In this paper, we show that if we take into account the widely spread probability-based origin of fuzzy techniques, then many heuristic techniques – including techniques related to defuzzification and optimization – become *justified*. Moreover, this use of probabilistic ideas sometimes enables us to *improve* the existing heuristic fuzzy techniques.

Comment. In our opinion, the above justification is a good example of the need for integrated uncertainty: in a situation where pure probabilistic methods are not natural, and where pure fuzzy techniques lack a convincing justification, a combination of probabilistic and fuzzy approaches helps.

2 Probability-Based Approach Explains Heuristic Formulas of Defuzzification

Let us start by showing that probability-based approach explains the main formula of centroid justification.

Crudely speaking, the membership function $\mu(x)$ describes the degree to which x corresponds to the optimal control. If the membership function comes from a probability distribution $\rho(x)$, this means that we do not know exactly which value x is optimal: different values x may turn out to be optimal, and the corresponding values $\rho(x)$ describes the probability of different values to be optimal.

Based on this information, we want to select a single value \bar{x} . Because of the probabilistic character of available information, no matter what value we select, there is a probability that this value which be not optimal. So, no matter what value we select, there will be a loss caused by this non-optimality. It is reasonable to select the value \bar{x} for which the expected value of this loss is the smallest.

The loss happens if the optimal value x is different from the selected value x' . In other words, the loss $L(x, x')$ is caused by the fact that difference $x - x'$

is different from 0. The loss can thus be viewed as a function of this difference $L(x, x') = F(x - x')$ for some function $F(z)$.

The difference $z = x - x'$ is usually reasonable small, so we can expand the function $F(z)$ in Taylor series and keep only the first few terms in this expansion:

$$F(z) = a_0 + a_1 \cdot z + a_2 \cdot z^2.$$

When the selected value x' is exactly the optimal value x , i.e., when $z = 0$, then there is no loss: $F(0) = 0$. Substituting $z = 0$ into the above quadratic formula, we conclude that $a_0 = 0$. Also, when $z = 0$, the loss is the smallest; thus, for $z = 0$, the derivative $F'(0)$ is equal to 0 – which implies that $a_1 = 0$ – and the second derivative is non-negative – so $a_2 > 0$.

So, $F(z) = a_2 \cdot z^2$, so the loss is equal to $L(x, x') = a_2 \cdot (x - x')^2$, and the expected value of this loss is equal to

$$\int L(x, x') \cdot \rho(x) dx = \int a_2 \cdot (x - x')^2 \cdot \rho(x) dx.$$

We want to find the value x' that minimizes this loss. To find this value, we differentiate the above expression by x' and equate the resulting derivative to 0. As a result, we get

$$\int 2 \cdot a_2 \cdot (x - x') \cdot \rho(x) dx = 0.$$

Dividing both sides by $2a_2$ and representing the integral of the difference as the difference between the two integrals, we conclude that

$$\int x \cdot \rho(x) dx - x' \cdot \int \rho(x) dx = 0.$$

The second integral in this formula is simply the total probability, i.e., 1, so the optimal value \bar{x} of the control x is equal to the mean

$$\bar{x} = \int x \cdot \rho(x) dx.$$

As we have mentioned earlier, the membership function $\mu(x)$ and the corresponding probability density function $\rho(x)$ differ only by a multiplicative constant:

$$\mu(x) = c \cdot \rho(x)$$

for an appropriate constant c – so that $\rho(x) = \frac{\mu(x)}{c}$. To find the constant c , we can integrate both sides of the equality $\mu(x) = c \cdot \rho(x)$. We thus get $\int \mu(x) dx = c \cdot \int \rho(x) dx$. If we take into account that, as we have recently mentioned, $\int \rho(x) dx = 1$, we conclude that $c = \int \mu(x) dx$. Thus, $\rho(x) = \frac{\mu(x)}{\int \mu(y) dy}$.

Substituting this expression for $\rho(x)$ into the above formula for \bar{x} , we get exactly the usual formula for centroid defuzzification.

3 Let Us Use the Probability-Based Justification to Improve the Heuristic Formulas for Defuzzification

Formulas for defuzzification need improvement: reminder. As we have mentioned earlier, the problem with the existing defuzzification formulas is not only that they are heuristic and thus need justification, but also that these formulas sometimes lead to disastrous (or at least suboptimal) results.

It is therefore desirable to improve these formulas.

Idea. A reasonable idea is to take into account that we are not just interested in finding the values x that minimize the total loss; ideally, the selected value x should also be optimal in relation to the original control problem. The corresponding degree of optimality is described by the membership function $\mu(x)$. Thus, in effect, we have a problem of optimization under fuzzy constraint: minimize the expression

$$\int (x - \bar{x})^2 \cdot \rho(x) dx = \frac{\int (x - \bar{x})^2 \cdot \mu(x) dx}{\int \mu(x) dx}$$

under the fuzzy constraint described by the original membership function $\mu(x)$.

The denominator of the minimized expression does not depend on the selection of the control parameter \bar{x} ; so, minimizing the above ratio is equivalent to minimizing the numerator $\int (x - \bar{x})^2 \cdot \mu(x) dx$.

To solve this problem, we can therefore use the Bellman-Zadeh approach. Let us see what we get.

Towards the resulting modification. Instead of the centroid defuzzification, we should select the value $\bar{x} = x'$ for which the following expression attains the smallest possible value

$$f_{\&} \left(\mu(x'), \frac{M - \int (x - x')^2 \cdot \mu(x) dx}{M - m} \right),$$

where

$$m \stackrel{\text{def}}{=} \min_{x'} \int (x - x')^2 \cdot \mu(x) dx$$

and

$$M \stackrel{\text{def}}{=} \max_{x'} \int (x - x')^2 \cdot \mu(x) dx.$$

Let us show how this formula can be simplified. To find m and M , we, correspondingly, minimize or maximize the expression $\int (x - x')^2 \cdot \mu(x) dx$. If we open parentheses, we can conclude that this expression is quadratic in terms of x' :

$$\int (x - x')^2 \cdot \mu(x) dx = M_2 - 2M_1 \cdot x' + M_0 \cdot (x')^2,$$

where we denoted $M_2 \stackrel{\text{def}}{=} \int x^2 \cdot \mu(x) dx$, $M_1 \stackrel{\text{def}}{=} \int x \cdot \mu(x) dx$, and $M_0 \stackrel{\text{def}}{=} \int \mu(x) dx$. We know that the minimum of this expression is attained at the centroid value,

$x_0 = \frac{M_1}{M_0}$. Substituting this value x_0 into the above expression for the integral $\int (x - x')^2 \cdot \mu(x) dx$, we conclude that

$$m = M_2 - 2M_1 \cdot \frac{M_1}{M_0} + M_0 \cdot \left(\frac{M_1}{M_0} \right)^2 = M_2 - \frac{M_1^2}{M_0}.$$

For the quadratic function which attains its minimum, its maximum on any interval is attained at one the interval's endpoints.

Thus, we arrive at the following modified version of the centroid defuzzification.

The resulting modification of centroid defuzzification. Once we know the membership function $\mu(x)$ on an interval $[x_-, x_+]$, then, to find the best value \bar{x} , we do the following:

- first, we compute the values $M_0 = \int \mu(x) dx$, $M_1 = \int x \cdot \mu(x) dx$, and $M_2 = \int x^2 \cdot \mu(x) dx$;
- then, we compute the values $m = M_2 - \frac{M_1^2}{M_0}$ and

$$M = \max(M_2 - 2M_1 \cdot x_- + M_0 \cdot x_-^2, M_2 - 2M_1 \cdot x_+ + M_0 \cdot x_+^2);$$

- finally, we find the value $\bar{x} = x'$ that maximizes the expression

$$f_{\&} \left(\mu(x'), \frac{M - (M_2 - 2M_1 \cdot x' + M_0 \cdot (x')^2)}{M - m} \right).$$

This is indeed better than centroid. As we have mentioned earlier, the main problem of centroid defuzzification is that it sometimes leads to very bad decisions, i.e., decisions \bar{x} for which the value $\mu(\bar{x})$ is 0 (or close to 0). This is possible for centroid defuzzification – since its algorithm does not take the value $\mu(\bar{x})$ into account at all.

However, for our new method, this is not possible. Indeed, for both $f_{\&}(a, b) = \min(a, b)$ and $f_{\&}(a, b) = a \cdot b$, we have $f_{\&}(0, a) = 0$ for all $a \in [0, 1]$. Thus, if $\mu(\bar{x}) = 0$, then the corresponding objective function is equal to 0 – i.e., to its smallest possible value – and thus, will never be selected under the new approach.

What if we have two equally possible solutions? In the case of a symmetric obstacle, we will no longer go straight into this obstacle, so the corresponding angle $x = 0$ is not possible. Hence we select a value $\bar{x} \neq 0$.

Due to symmetry, if $\bar{x} \neq 0$ is a solution, then $-\bar{x}$ is a solution as well. Thus, we have at least two different solutions. Which one should we choose?

The situation is symmetric, so our decision should be symmetric as well. However, if we select one of the two possible solutions \bar{x} or $-\bar{x}$, we violate $x \leftrightarrow -x$ symmetry. So what should we do?

The only way to preserve symmetry is to make a *probabilistic* decision, i.e., in this case, to select either \bar{x} or $-\bar{x}$ with equal probability $1/2$.

Comment. Thus again, probabilistic ideas help: namely, they help to retain a natural symmetry of the situation.

Discussion. In fuzzy control, this may be a new idea, but in general, that symmetry sometimes naturally leads to randomness is a known fact.

The first such example is *game theory*; see, e.g., [6, 8, 12]. The fact that the optimal strategies are probabilistic has been known since the beginning of game theory. Indeed, suppose that we want to protect two equally valuable locations from a terrorist attack, but we only have resources for a single protection team. If we select a deterministic decision – i.e., send the team to one of the two locations – we will lose no matter which location we select, since the terrorists will successfully attack the remaining location. The best strategy is to each time send a team to one of the locations at random.

A more relaxed example is the rock-paper-scissors game, in which each of the two players selects either rock, or paper, or scissors. Paper beats rock, rock beats scissors, and scissors beat paper. If one side selects a deterministic strategy – i.e., selects the same choice every time – the opponent will always win by selecting the choice that beats this selection. The only way to avoid this defeat is to select each of the three choices with equal probability.

Another example is *physics*, for example, the radioactivity phenomenon, when some atoms spontaneously decay; see, e.g., [3, 11]. Let us show that radioactivity cannot be deterministic, i.e., at the present moment t_0 , we cannot predict the moment of time t at which the atom will decay (i.e., equivalently, the time period $t - t_0$ until the decay). Indeed, the laws of physics do not change if we simply change the starting point for measuring time. Thus, if the decay process was deterministic, then we will be able to conclude that when we observe the not-yet-decayed atom at a moment $t_0 + \varepsilon$ for some $\varepsilon > 0$, then we should also predict decay $t - t_0$ seconds in the future – but this cannot be, since there is only $t - t_0 - \varepsilon$ second to the deterministic decay. Thus, the decay cannot be deterministic, it has to be probabilistic.

Comment. This example shows, by the way, that the probabilistic character of quantum physics is not some counter-intuitive feature, it is a natural consequence of simple symmetries.

Remaining problem. Our idea seems reasonable. However, there is still a problem: to come up with an improved defuzzification method, we used Bellman-Zadeh formulas – and, as we have mentioned earlier, these formulas are heuristic. It is thus desirable to come up with a justification for these formulas. Let us show that the probability-based approach provides exactly such a justification.

Comment. The main ideas behind this justification first appeared in [5].

4 Probability-Based Approach Explains Heuristic Formulas of Optimization under Fuzzy Constraints

Let us consider the maximization case: we want to maximize the value objective function $f(x)$ under the fuzzy constraint described by the membership function $\mu(x)$. (The minimization case can be treated similarly.)

If we select a value x , and this value is possible, then we get the gain $f(x)$. On the other hand, if we select x , and this value x is *not* possible, then we will have to go back to the worst-case scenario m . Let us denote the probability of the value x to be possible by $p(x)$. Then:

- with probability $p(x)$, we get $f(x)$, and
- with the remaining probability $1 - p(x)$ we get m .

The expected gain is this equal to $p(x) \cdot f(x) + (1 - p(x)) \cdot m$. This expression can be reformulated as

$$p(x) \cdot f(x) + m - p(x) \cdot m = m + p(x) \cdot (f(x) - m).$$

Adding a constant to all the values of an objective function (in this case, the constant m) does not change which values are larger and which values are smaller. Thus, maximizing the above objective function is equivalent to maximizing a simpler expression $p(x) \cdot (f(x) - m)$.

We consider the cases when the probabilities are proportional to the corresponding values of the membership function: $p(x) = c \cdot \mu(x)$. In this case, the above maximized expression takes the form $c \cdot \mu(x) \cdot (f(x) - m)$. Multiplying all the values of an objective function by the same constant does not change which values are larger and which are smaller – e.g., the richest person in Mexico remains the richest whether we count his net worth in US dollars or in Euros or in Mexican pesos. Thus, maximizing the above expression is equivalent to maximizing the product $\mu(x) \cdot (f(x) - m)$. Similarly, since the difference $M - m$ is also a constant not depending on x , the above maximization is equivalent to maximizing the expression

$$\mu(x) \cdot \frac{f(x) - m}{M - m}.$$

This is exactly Bellman-Zadeh formula for $f_{\&}(a.b) = a \cdot b$. Thus, the probability-based approach indeed explains this heuristic formula.

References

1. R. Belohlavek, J. W. Dauben, and G. J. Klir, *Fuzzy Logic and Mathematics: A Historical Perspective*, Oxford University Press, New York, 2017.
2. R. E. Bellman and L. A. Zadeh, “Decision making in a fuzzy environment”, *Management Science*, 1970, Vol. 17, No. 4, pp. B 141–B 164.
3. R. Feynman, R. Leighton, and M. Sands, *The Feynman Lectures on Physics*, Addison Wesley, Boston, Massachusetts, 2005.

4. G. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic*, Prentice Hall, Upper Saddle River, New Jersey, 1995.
5. O. Kosheleva and V. Kreinovich, “Why Bellman-Zadeh Approach to Fuzzy Optimization”, *Applied Mathematical Sciences*, 2018, Vol. 12, No. 11, pp. 517–522.
6. R. D. Luce and H. Raiffa, *Games and Decisions: Introduction and Critical Survey*, Dover, New York, 1989.
7. J. M. Mendel, *Uncertain Rule-Based Fuzzy Systems: Introduction and New Directions*, Springer, Cham, Switzerland, 2017.
8. R. B. Myerson, *Game Theory: Analysis of Conflict*, Harvard University Press, Harvard, Massachusetts, 1997.
9. H. T. Nguyen and E. A. Walker, *A First Course in Fuzzy Logic*, Chapman and Hall/CRC, Boca Raton, Florida, 2006.
10. V. Novák, I. Perfilieva, and J. Močkoř, *Mathematical Principles of Fuzzy Logic*, Kluwer, Boston, Dordrecht, 1999.
11. K. S. Thorne and R. D. Blandford, *Modern Classical Physics: Optics, Fluids, Plasmas, Elasticity, Relativity, and Statistical Physics*, Princeton University Press, Princeton, New Jersey, 2017.
12. J. von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior*, Princeton University Press, Princeton, New Jersey, 1944.
13. L. A. Zadeh, “Fuzzy sets”, *Information and Control*, 1965, Vol. 8, pp. 338–353.