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

Christian Servin[1], Olga Kosheleva[2], and Vladik Kreinovich[2]

[1]El Paso Community College, El Paso, TX 79915, USA,
cservin@gmail.com
[2]University of Texas at El Paso, El Paso, Texas 79968, USA,
olgak@utep.edu, vladik@utep.edu

1. 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.
- This will help other experts perform better.
- Ideally, the system will make expert-quality decisions on its own, without the need for the experts.

2. Need for Fuzzy Knowledge (cont-d)

- 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,
  - examples: "close", "fast", "small", etc., but
  - 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.
- Such technique was proposed in the 1960s by Lotfi Zadeh from Berkeley under the name of *fuzzy logic*.

3. Need for Fuzzy Knowledge (cont-d)

- In fuzzy logic, to represent each word like "small" in numerical terms, we assign:
  - to each possible value x 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.

4. 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.

5. Where Fuzzy Degrees Come From (cont-d)

- For example, when we describe the meaning of the word "small", then:
  - for values x which are really small, we will have a large number of such records;
  - on the other hand, for values x which are not too small, we will have a few such records;
  - indeed, few experts will consider these values to be small.
- We can estimate the frequency with which different values x appear in our records.
- This frequency can be described by a probability density function (pdf) $\rho(x)$.
- When x is really small, the value $\rho(x)$ is big.

6. Where Fuzzy Degrees Come From (cont-d)

- When x is not so small, fewer experts will consider this value to be small.
- Thus, the value $\rho(x)$ will be much smaller.
- Thus, in principle, we could use the values $\rho(x)$ as the desired degrees.
- However, we want values of the membership function – and these values should be from the interval $[0, 1]$.
- However, the pdf can take values larger than 1.
- To make all the values ≤ 1, we can normalize these values, i.e., divide by the largest of them:

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

- This is a well-known way to get membership functions (Coletti, Huynh, Lawry, et al.)

7. 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*.

8. Centroid Defuzzification: Description, Successes, and Limitations

- The most widely used defuzzification procedure is *centroid* defuzzification, in which we select the value

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

- It has led to many successful applications of fuzzy control.
- However, it has two related limitations.
- First, it is heuristic, it is not justified by a precise argument.
- Therefore, we are not sure whether it will always work well.
- Second, it sometimes leads to disastrous results.

9. Centroid Defuzzification (cont-d)

- 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 as a situation.
- However, these techniques are also heuristic – and thus, not guaranteed to produce good results.

10. Optimization under Fuzzy Constraints

- Another class of situations in which fuzzy knowledge is important is optimization.
- Traditional optimization techniques finds x for which the objective function $f(x)$ attains its optimal value.
- This value can be argest 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.

11. Fuzzy Optimization (cont-d)

- There are well-known techniques for dealing with such constraints – e.g., the Lagrange multiplier method:
  - the problem of optimizing an objective function $f(x)$ under constraint $g(x) = 0$ reduces to
  - the unconstrained optimization of an auxiliary objective function $f(x) + \lambda \cdot g(x)$, for some 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.
- One way to do it is to make sure that the noise level is not high.

12. Fuzzy Optimization (cont-d)

- 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 $\mu(x)$:
  - 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?

13. Fuzzy Optimization (cont-d)

- A well-known heuristic solution to this problem was proposed in a joint paper by:
  - Lotfi Zadeh and
  - Richard Bellman (the famous specialist in optimization).
- They proposed to maximize an auxiliary function

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

  - $f_\&(a,b)$ 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 $f(x)$:

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

14. Fuzzy Optimization (cont-d)

- The above formula is used when we maximize $f(x)$.
- Minimizing $f(x)$ is equivalent to maximizing an auxiliary function $f'(x) \stackrel{\text{def}}{=} -f(x)$, so:

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

- These heuristic formulas have led to many useful application.
- However, these formulas are heuristic – and thus, lack a convincing justification.
- This makes users often somewhat reluctant to use them.

15. What We Do in This Talk

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

16. Probability-Based Approach Explains Heuristic Formulas of Defuzzification

- Crudely speaking, the membership function $\mu(x)$ describes the degree to which x is an optimal control.
- We consider the case when 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.
- 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 $\bar{x}$.

17. Probability-Based Approach (cont-d)

- Because of the probabilistic character of available information:
  - no matter what value we select,
  - there is a probability that this value will 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.

18. Probability-Based Approach (cont-d)

- The loss can thus be viewed as a function of this difference $L(x, x') = F(x - x')$ for some function $F(z)$.
- It is reasonable to assume that the loss function $F(z)$ is continuous in z.
- Every continuous function on an interval can be approximated:
  - with any given accuracy,
  - by an analytical function – e.g., by a polynomial.
- Thus, it is safe to assume that the function $F(z)$ is analytical, i.e.:

$$F(z) = a_0 + a_1 \cdot z + a_2 \cdot z^2 + a_3 \cdot z^3 + \ldots$$

- The difference $z = x - x'$ is usually reasonable small.
- So, from the practical viewpoint, we can safely ignore higher order terms and keep only the first few terms.

19. Probability-Based Approach (cont-d)

- From the purely mathematical viewpoint, the simplest possible case is when we keep only the constant term

$$F(z) = a_0.$$

- But then the loss does not depend on how far the selected value x' is from the unknown optimal value x.
- This does not make sense.
- What if we take into account linear terms, i.e., consider the loss function $F(z) = a_0 + a_1 \cdot z$?
- The loss function should attains its smallest value $F(z) = 0$ when the selected value x' is optimal $z = x' - x = 0$.
- However, a linear function does not attains its minimum at 0.

20. Probability-Based Approach (cont-d)

- So, the simplest case that makes sense is when we take quadratic terms into account:

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

- When $x' = x$, there is no loss, so, $F(0) = 0$ and $a_0 = 0$.
- Also, when $z = 0$, the loss is the smallest.
- Thus, for $z = 0$, the derivative $F'(0)$ is equal to 0 (hence $a_1 = 0$) and $F''(0) \le 0$ (so $a_2 > 0$).
- So, $F(z) = a_2 \cdot z^2$, so $L(x, x') = a_2 \cdot (x - x')^2$, and the expected value of the loss is:

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

- We want to find the value x' that minimizes this loss.

21. Probability-Based Approach (cont-d)

- To find this value, we differentiate the above expression by x' and equate the resulting derivative to 0; thus:

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

- So, $\int x \cdot \rho(x)\, dx - x' \cdot \int \rho(x)\, dx = 0$.
- The second integral in this formula is simply the total probability, i.e., 1.
- So the optimal value $\bar{x}$ is equal to the mean

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

- The membership function $\mu(x)$ is $\mu(x) = c \cdot \rho(x)$, so $\rho(x) = \dfrac{\mu(x)}{c}$.
- To find the c, we integrate both sides of the equality $\mu(x) = c \cdot \rho(x)$: $\int \mu(x)\, dx = c \cdot \int \rho(x)\, dx = c$.

22. Probability-Based Approach (cont-d)

- Thus, $\rho(x) = \dfrac{\mu(x)}{\int \mu(y)\,dy}$.
- Let us substitute this expression into the formula

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

- As a result, we get exactly the usual formula for centroid defuzzification:

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

23. Let's Improve Defuzzification

- 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 $\mu(x)$.

24. Let's Improve Defuzzification (cont-d)

- 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: select $\bar{x} = x'$ that minimizes:

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

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

- To find m and M, we, correspondingly, minimize or maximize the expression $\int (x-x')^2 \cdot \mu(x)\,dx$.

25. Let's Improve Defuzzification (cont-d)

- If we open parentheses, we can conclude that this expression is quadratic in terms of x':

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

$$\text{where } M_i \stackrel{\text{def}}{=} \int x^i \cdot \mu(x)\,dx.$$

- We know that the minimum of this expression is attained at the centroid value, $x_0 = \dfrac{M_1}{M_0}$.
- Thus, $m = M_2 - 2M_1 \cdot \dfrac{M_1}{M_0} + M_0 \cdot \left(\dfrac{M_1}{M_0}\right)^2 = M_2 - \dfrac{M_1^2}{M_0}$.
- For the quadratic function which attains its minimum,
  - its maximum on any interval
  - is attained at one the interval's endpoints. Thus:

26. The Resulting Modification of Centroid Defuzzification

- We know the membership function $\mu(x)$ on an interval
$$[x_-, x_+].$$
- We want to find the best value $\bar{x}$.
- 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 $m = M_2 - \dfrac{M_1^2}{M_0}$ 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
$$f_\& \left(\mu(x'), \frac{M - (M_2 - 2M_1 \cdot x' + M_0 \cdot (x')^2)}{M - m} \right).$$

27. This Is Indeed Better Than Centroid

- The main problem of centroid defuzzification is that it sometimes leads to very bad decisions when $\mu(\bar{x}) = 0$.
- This is possible for centroid defuzzification – since its algorithm does not take the value $\mu(\bar{x})$ into account.
- 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 its smallest possible value 0.
- Thus, this bad value will never be selected under the new approach.

28. 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.

29. Two Solutions (cont-d)

- So what should we do?
- The only way to preserve symmetry is to make a *probabilistic* decision.
- In this case, we select either $\bar{x}$ or $-\bar{x}$ with equal probability $1/2$.
- Thus again, probabilistic ideas help: namely, they help to retain a natural symmetry of the situation.
- 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*.
- The fact that the optimal strategies are probabilistic has been known since the beginning of game theory.

30. Two Solutions (cont-d)

- 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, then we send the team to one of the two locations.
- Then, 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.

31. Remaining Problem

- To come up with an improved defuzzification method, we used Bellman-Zadeh formulas.
- However, 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.

32. Probability-Based Approach Explains Heuristic Formulas of Fuzzy Optimization

- We want to maximize the value objective function $f(x)$ under the fuzzy constraint described by $\mu(x)$.
- (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.

33. Fuzzy Optimization (cont-d)

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

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

- Adding m to all the values of an objective function does not change which values are larger.
- Thus, maximizing the above objective function is equivalent to maximizing $p(x) \cdot (f(x) - m)$.
- We consider the cases when the probabilities are proportional to 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 the objective function by a constant does not change which values are larger.

34. Fuzzy Optimization (cont-d)

- The same person is the richest in Mexico whether we count his net worth in US dollars or in Mexican pesos.
- Thus, maximizing the above expression is equivalent to maximizing the product $\mu(x) \cdot (f(x) - m)$.
- The difference $M - m$ is also a constant not depending on x. Thus, the above maximization is equivalent to maximizing the expression

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

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

35. Acknowledgments

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