What Is the Most Adequate Fuzzy Methodology?

Noah Velasco[1], Olga Kosheleva[2], and Vladik Kreinovich[1]
[1]Department of Computer Science
[2]Department of Teacher Education
University of Texas at El Paso
500 W. University
El Paso, TX 79968, USA
rnpadilla2@miners.utep.edu, vladik@utep.edu

Abstract

In practice, there is often a need to describe the relation $y = f(x)$ between two quantities in algorithmic form: e.g., we want to describe the control value y corresponding to the given input x, or we want to predict the future value y based on the current value x. In many such cases, we have expert knowledge about the desired dependence, but experts can only describe their knowledge by using imprecise ("fuzzy") words from a natural language. Methodologies for transforming such knowledge into an algorithm $y = f(x)$ are known as fuzzy methodologies. There exist several fuzzy methodologies, a natural question is: which of them is the most adequate? In this paper, we formulate the natural notion of adequacy: that if the expert rules are formulated based on some function $y = f(x)$, then the methodology should reconstruct this function as accurately as possible. We show that none of the existing fuzzy methodologies is the most adequate in this sense, and we describe a new fuzzy methodology that is the most adequate.

1 Outline

Question. In many practical control situations, we do not have the exact model of a system that we need to control but we have the experience of successful expert human controllers. Human controllers often formulate their experience by using imprecise ("fuzzy") words from natural language like "small". How can we translate this expert knowledge into a precise control strategy for an automatic controller?

A similar problem emerges when we want to use expert rules to predict the future state of the worlds.

To translate imprecise expert statements into precise form, Lotfi Zadeh invented a special methodology that he called *fuzzy*; see, e.g., [1, 2, 4, 5, 6, 8].

In this methodology, we start by describing each natural-language term A (e.g., "small") by a function that assigns:

- to each possible value x of the corresponding quantity,
- a degree $\mu_A(x)$ from the interval $[0,1]$ to which, in the controller's opinion, this value satisfies the corresponding property (e.g., the degree to which the value x is small).

This function is known as a *membership function* or, alternatively, as a *fuzzy set*.

Once we have fuzzy sets corresponding to all relevant natural-language terms, and we have all natural-language if-then rules provided by the human controllers, we need to transform this information into a precise control strategy. There are several different methods for generating such a strategy. A natural question is: which method should we select? In other words, which method is the most adequate?

What we do in this paper. To answer the above question, a natural requirement is that if the expert's if-then rules describe – in fuzzy terms – an actual control strategy $y = f(x)$, then the fuzzy methodology should return exactly this strategy. Somewhat surprisingly, it turns out that the existing fuzzy methodologies – including the very popular Mamdani approach – do not satisfy this requirement. In this paper, we show that this requirement actually leads to a new methodology, a methodology that we describe and analyze.

Structure of this paper. In Section 2, we briefly recall what is a fuzzy methodology and which fuzzy methodologies are typically used in practical applications. In Section 3, we describe a natural criterion for deciding which fuzzy methodology is the most adequate, and we show that from the viewpoint of this criterion, none of the current methodologies are perfect. In Section 4, we describe a methodology which is the most adequate according to our natural criterion – and analyze some properties of this methodology.

2 What Is Fuzzy Methodology: A Brief Reminder

Need for expert knowledge. In many practical situations, we want to make a decision; for example:

- we want to decide what control to apply to a system,
- we want to decide what is the patient's disease and what dose of what medicine should be the best for this patient,
- we want to predict tomorrow's weather, etc.

In many such situations, we do not have an accurate model of the system and thus, we cannot formulate this problem in precise terms. What we usually do have is the experience of experts:

- we have the experience of human expert controllers who control a plant,
- we have the experience of expert medical doctors who are good in diagnosing and treating the patient,
- we have the experience of expert meteorologists who can predict tomorrow's weather in their region with high accuracy, etc.

It is therefore desirable to use this expert knowledge to design an automatic controller and/or an automatic expert system.

Using expert knowledge is not easy. Most experts are willing to share their expertise, but the problem is that experts often cannot describe their knowledge in precise terms. Instead, they formulate this knowledge in terms of if-then rules that use imprecise ("fuzzy") words from natural language.

For example, many people know how to drive. So, at first glance, it may seem to be an easy task to design a self-driving car: just use the experience of good human drivers. However, this is not so easy. An automatic controller would need to know what control to apply in each situation. For example, if a car is going on a freeway with the speed of 100 km per hour, and car car in front – which is 10 meters ahead – slows down to 95 km per hour, what should we do? A natural human answer is "break a little bit", but what the automatic controller needs is with how many Newtons of force to push the break pedal and for how many milliseconds – and most human drivers cannot provide these numbers.

Fuzzy methodology: first step. To perform this challenging task, i.e., to extract precise knowledge from the imprecise expert knowledge, Lotfi Zadeh invented a new methodology that he called fuzzy. This methodology starts with providing a precise description of all natural-language words used by experts.

For this purpose:

- for each such word A and for each possible value x of the corresponding quantity,
- we ask the expert to mark, on a scale from 0 to 1, to what extent the value x has the corresponding property (e.g., to what extent x is small).

The intent is that:

- mark 1 correspond to the case when the expert is absolutely sure that x satisfies this property,
- mark 0 means that the expert is absolutely sure that x does not satisfy this property, and
- marks between 0 and 1 correspond to intermediate cases.

The resulting function $A(x)$ that assigns the degree to each value x is called a *membership function* or a *fuzzy set.*

Comment. Of course, there are infinitely many real numbers x, and we can only ask finitely many questions to the expert. So, in practice:

- we ask the expert a finite number of questions, about finitely many values $x_1, \ldots, x_n$, and then
- we use interpolation/extrapolation to estimate the values $A(x)$ for all other values x.

In particular, if we ask the expert to provide:

- the value M for which this user is absolutely sure that this property is satisfied (i.e., that $A(M) = 1$), and
- the values $\underline{m}$ and $\overline{m}$ such that outside the interval $[\underline{m}, \overline{m}]$, the property is *not* satisfied (i.e., $A(x) = 0$),

and use linear interpolation, then we get a frequently used *triangular* membership function.

If instead of a single value M, we get the whole interval $\left[\underline{M}, \overline{M}\,\right]$ on which the property A is satisfied, i.e., for which $A(M) = 1$ for all values M from this interval, and we use linear interpolation, then we get trapezoid membership functions.

Fuzzy methodology beyond the first step: what we have. After the first step, to determine the desired dependence $y = f(x)$, we have several expert if-then rules

$$\begin{array}{l} \text{If } x \text{ is } A_1 \text{ then } y \text{ is } B_1. \\ \text{If } x \text{ is } A_2 \text{ than } y \text{ is } B_2. \\ \qquad\qquad \ldots \\ \text{If } x \text{ is } A_k \text{ then } y \text{ is } B_k. \end{array}$$

where A_i and B_i are natural-language terms that are described by membership functions $A_i(x)$ and $B_i(y)$. Based on this information, we want to generate a function $y = f(x)$ that adequately describes these rules.

Example. To illustrate our ideas, let us consider a simple example of controlling a thermostat by turning a knob.

- If we turn the knob to the right, the temperature increases.
- If we turn it to the left, the temperature decreases.

In this example:

- the desired control variable y is the angle on which we turn the knob, and
- the input x is the difference $x \stackrel{\text{def}}{=} T - T_0$ between the actual temperature T and the desired temperature T_0.

If the temperature is close to the desired one, i.e., if the difference x is close to 0, then we should not change anything, i.e., the control y should be negligible. So, we arrive at the first rule:

If x is negligible, then y should be negligible.

If the temperature is slightly higher than desired, then we should turn the knob to the left a little bit. So, we arrive at the second tule:

If x is small positive, then y should be small negative.

Similarly, if the temperature is slightly lower than desired, then we should turn the knob to the right a little bit. So, we arrive at the second tule:

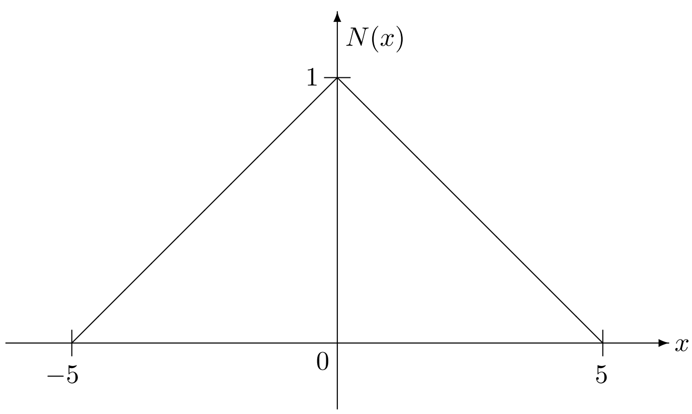
If x is small negative, then y should be small positive.

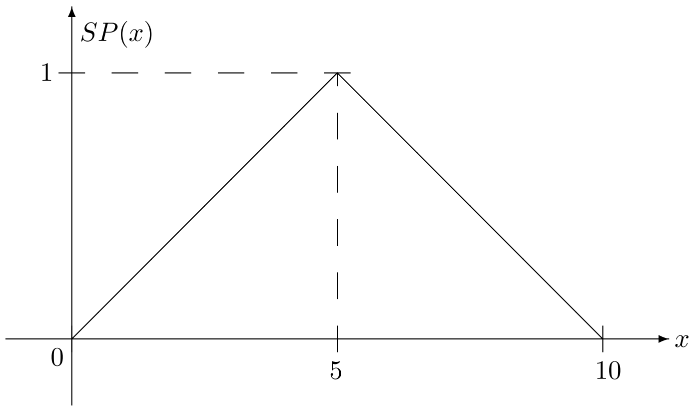
We can add more rules, but for simplicity, let us only consider these three rules. The restriction to these three rules makes sense in situations when the control is almost perfect, and we experience only small deviations from the desired temperature.

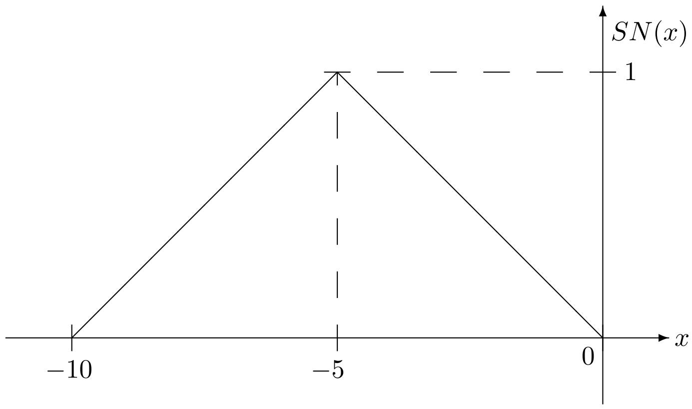
Also, for simplicity, let us consider simple triangular membership functions corresponding to "negligible", "small positive", and "small negative". We will denote them, correspondingly, by $N(x)$, $SP(x)$, and $SN(x)$. Based on our experience, we assume that:

- for "negligible": the value $M = 0$ is definitely negligible, and values outside the interval $[-5, 5]$ are definitely not negligible;
- for "small positive": the value $M = 5$ is definitely small positive, and values outside the interval $[0, 10]$ are definitely not small positive: value smaller than 0 are not positive, and values larger than 10 are not small;
- for "small negative": the value $M = -5$ is definitely small negative, and values outside the interval $[-10, 0]$ are definitely not small negative: value smaller than -10 are not small, and values larger than 0 are not negative.

In this case, linear interpolation leads to the following triangular membership functions:







Fuzzy methodologies beyond the first step: examples. Let us list the most frequently used fuzzy methodologies, i.e., methodologies of transforming fuzzy rules into a precise function $y = f(x)$.

Fuzzy methodology beyond the first step: Mamdani approach. One of the most widely used approaches was originally proposed by Mamdani and is, thus, known as Mamdani approach. In this approach, we first take into account that for a given value x, the value y is reasonable (R) if:

- either the first rule is applicable, i.e., x is A_1 *and* y is B_1,
- *or* the second rule is applicable, i.e., x is A_2 *and* y is B_2, etc.

We can symbolically describe it as follows:

$$R(y) \Leftrightarrow (A_1(x) \,\&\, B_1(y)) \vee (A_2(x) \,\&\, B_2(y)) \vee \ldots$$

To give this formula a numerical meaning, we need to provide the numerical meaning to the "and"- and "or"-operations, i.e., in effect, to extend the "and"-

and "or"-operations of the usual 2-valued logic (with the values "false" (0) and "true" (1)) to the whole interval $[0,1]$. From the computational viewpoint, the simplest such extensions and min and max. Thus, we arrive at the following membership function for "reasonable":

$$R(y) = \max(\min(A_1(x), B_1(y)), \min(A_2(x), B_2(y)), \ldots)$$

Our ultimate objective is to come up with a single value $\overline{y}$. A reasonable way to come up with this value is to minimize the weighted squared difference between this value and possible values y, weighted by the degree to which y is possible, i.e., to minimize the following expression:

$$\int R(y) \cdot (\overline{y} - y)^2 \, dy.$$

To find the minimizing value $\overline{y}$, we can differentiate this expression with respect to $\overline{y}$ and equate the derivative to 0. As a result, we get the following expression:

$$\overline{y} = \frac{\int y \cdot R(y) \, dy}{\int R(y) \, dy}.$$

This expression is known as *centroid defuzzification.*

Fuzzy methodology beyond the first step: Takagi-Sugeno approach. An alternative approach is that we replace each y-membership function $B_i(y)$ by the result of its defuzzification, for example, by the centroid value

$$y_i = \frac{\int y \cdot B_i(y) \, dy}{\int B_i(y) \, dy}.$$

In effect, we ignore the fuzziness of y in the rules and consider the following simplified rules:

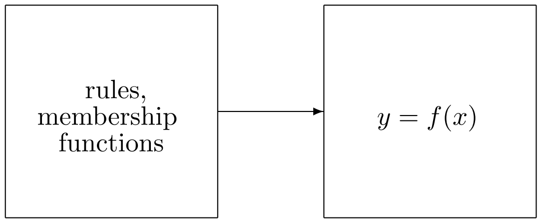
If x is A_1 then $y = y_1$.
If x is A_2 than $y = y_2$.
...
If x is A_k then $y = y_k$.

These rules can be treated the same way as in the previous approach, the only difference is that now the conclusions of each rule are not fuzzy. In this case, the value $R(y)$ is only different from 0 when y coincides with each of the points y_i, and for each of these values, we have $R(y_i) = A_i(x)$. Thus, the centroid formula leads to

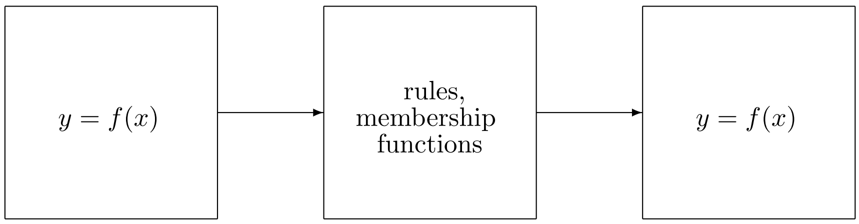
$$\overline{y} = \frac{\sum\limits_{i=1}^{k} A_i(x) \cdot y_i}{\sum\limits_{i=1}^{k} A_i(x)}.$$

3 How to Decide Which Fuzzy Methodology Is the Most Adequate

Idea. Fuzzy methodology transforms rules and membership functions into an exact control strategy $f(x)$:



Suppose now that we start with the actual function $y = f(x)$. As we have mentioned, fuzzy techniques deal with situations when the experts cannot explicitly describe this function. Instead, they formulate rules based on this function. In this case, a natural requirement is that once we process these rules, we should get back the original function $y = f(x)$. This is what we should have in the ideal case:



The closer the reconstructed function to the original function, the more adequate the fuzzy methodology – this is a natural idea of gauging adequacy of different methodologies.

What do we mean by rules generated by a function? Suppose that we know the function $y = f(x)$, and that we have fuzzy information about x: namely, that x is A_i for some property A_i which is described by a membership function $A_i(x)$. What can we then say about y? How can we describe the corresponding membership function $B_i(y)$?

The answer to this question is well-known in fuzzy research: it is provided by the so-called *Zadeh's extension principle*. This answer can be easily explained. Indeed, in this case, a real number Y is a possible value of the quantity y is there exists a value X which is a possible value of the quantity x and for which $f(X) = Y$. The degree to which X is a possible value of the quantity X is determined by the corresponding membership function $A_i(x)$ and is, thus, equal to $A_i(X)$. If there is only one X for which $f(X) = Y$ – this value X is then denoted by $X = f^{-1}(Y)$ – then $A_i(X) = A_i(f^{-1}(Y))$ is exactly the degree $B_i(Y)$ to which Y is a possible value of y. So, in this case, we have

$$B_i(y) = A_i(f^{-1}(x)). \tag{1}$$

What if there are several different values X for which $f(X) = Y$? This happens, e.g., when $f(x) = x^2$, then for each Y, there are two such values X: $X = \sqrt{Y}$ and $X = -\sqrt{Y}$. In this case, Y is possible if either we have the first of these values X or the second of these values X. The simplest way to estimate the degree to which an "or"-statement $A \vee B$ is true based in the degrees a and b to which individual statements A and B are true is to use maximum $\max(a, b)$. Thus, we get

$$B_i(y) = \max\{A_i(x) : f(x) = y\}. \tag{2}$$

This is exactly the formula that was first produced by Zadeh himself and is, thus, called Zadeh's extension principle. This membership function will be denoted as $B_i = f(A_i)$.

In these terms, the fuzzy methodology is most adequate if, based on the rules

$$\text{if } x \text{ is } A_i \text{ then } y \text{ is } B_i, \text{ where } B_i = f(A_i),$$

we should be able to reconstruct the original function $f(x)$.

Important comment. In the following text, we will use the known fact that for reasonable membership functions $A_i(x)$ – namely, for all the functions that first continuously increase from 0 to 1 and then continuously decrease from 1 to 0 – Zadeh's extension principle can be reformulated in terms of α-cuts, i.e., sets $\mathbf{A}_i(\alpha) \stackrel{\text{def}}{=} \{x : A_i(x) \geq \alpha\}$ and $\mathbf{B}_i(\alpha) \stackrel{\text{def}}{=} \{y : B_i(x) \geq \alpha\}$ for all $\alpha \in (0, 1]$. Namely, we have

$$\mathbf{B}_i(\alpha) = f(\mathbf{A}_i(\alpha)),$$

where for each set S, by $f(S)$, we mean

$$f(S) \stackrel{\text{def}}{=} \{f(x) : x \in S\}.$$

Are existing fuzzy methodologies most adequate? A natural question is: are the existing fuzzy methodologies – e.g., the ones described above – most adequate in this natural sense? Our answer is No. Let us explain this answer.

Mamdani methodology is not the most adequate (in the above sense). Let us explain, on a simple example, that Mamdani methodology is not the most adequate, i.e., that it does not reconstruct the original function $y = f(x)$.

Let us consider the above membership functions $N(x)$, $SP(x)$, and $SN(x)$, and a simple function $f(x) = -x$. In this case, as one can easily check, we have $f(N) = N$, $f(SP) = SN$, and $f(SN) = SP$. Thus, the rules generates by this function take exactly the form described in the previous section:

If x is N then y is N.
If x is SP then y is SN.
If x is SN, then y is SP.

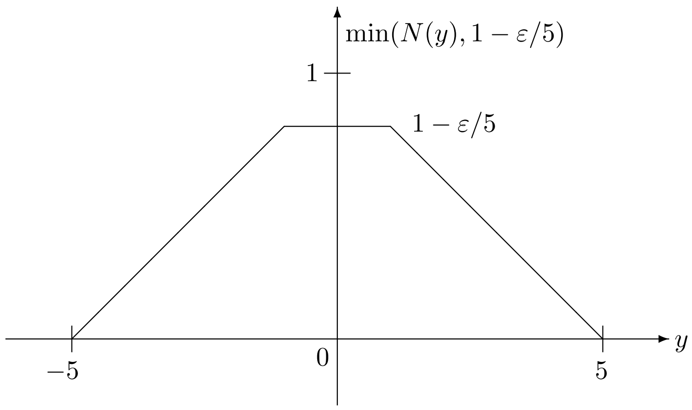
Let us consider a small negattve value $x = -\varepsilon$, where $\varepsilon > 0$. In this case,

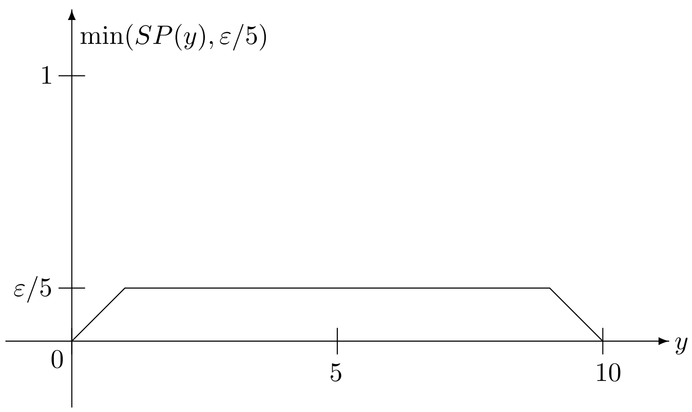
$$N(x) = 1 - \frac{\varepsilon}{5}, P(x) = \frac{\varepsilon}{5}, \text{ and } SN(x) = 0.$$

Thus, the reasonable value $R(y)$ is described by the formula

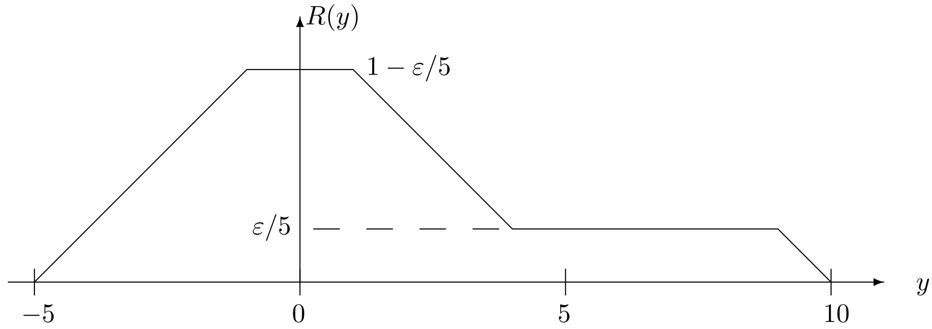
$$R(y) = \max\left(\min\left(N(y), 1 - \frac{\varepsilon}{5}\right), \min\left(SP(y), \frac{\varepsilon}{5}\right)\right).$$

The functions $\min(N(y), 1 - \varepsilon)$ and $\min(SP(y), \varepsilon/5)$ can be represented as follows:





Thus, the desired function $R(y)$ – which is the maximum of these functions – takes the following form



The result of the centroid defuzzification is the ratio of two integrals, so let us estimate these integrals. Let us first estimate the denominator $\int R(y)\,dy$. When ε tends to 0, the function $R(y)$ tends to $N(y)$, for which $\int N(y)\,dy$ is the area of the corresponding triangle with height 1 and base $5-(-5)=10$, i.e.,

$$\frac{1}{2}\cdot 10\cdot 1=5.$$

Thus, the denominator is equal to $5+O(\varepsilon)$.

The integral in the numerator can be represented as the sum of the parts: the symmetric part $R_{\rm sym}(y)=R_{\rm sym}(-y)$ corresponding to values from $y=-5$ to $y=5$, and the remaining part $r(y)\stackrel{\rm def}{=}R(y)-R_{\rm sym}(y)$. For the symmetric part $R_{\rm sym}(y)$, the integral $\int y\cdot R_{\rm sym}(y)\,dy$ is 0 – since for each $y>0$, contributions of the terms corresponding to y and to $-y$ cancel each other. Thus, the numerator is equal to $\int y\cdot r(y)\,dy$. For almost all the values y from $y=5$ to $y=10$, we have $r(y)=\varepsilon/5$, thus in the first approximation

$$\int y\cdot r(y)\,dy=\int_5^{10} y\cdot\frac{\varepsilon}{5}\,dy+o(\varepsilon)=\frac{\varepsilon}{5}\cdot\frac{1}{2}\cdot y^2\Big|_5^{10}+o(\varepsilon)=$$

$$\frac{\varepsilon}{5}\cdot\frac{1}{2}\cdot(10^2-5^2)+o(\varepsilon)=7.5\cdot\varepsilon+o(\varepsilon).$$

Thus, the desired ratio is equal to

$$\overline{y}=\frac{\int y\cdot R(y)\,dy}{\int R(y)\,dy}=\frac{7.5\cdot\varepsilon+o(\varepsilon)}{5+O(\varepsilon)}=1.5\cdot\varepsilon+o(\varepsilon).$$

This is clearly different from the original value

$$f(x)=f(-\varepsilon)=\varepsilon.$$

Takago-Sugeno approach is not the most adequate. It so happens that for the above example when $f(x)=-x$ and we have $N(x)$, $SP(x)$, and $SN(x)$, Takagi-Sugeno approach reconstructs the original function. However, for any nonlinear function $f(x)$, e.g., for $f(x)=-x+x^3$, this approach won't reconstruct the original function.

Indeed, the function reconstructed by this methodology is a linear combination of the membership functions corresponding to x. On the interval $[0, 5]$, all the membership functions are linear, so their linear combination is also linear – and thus, cannot be equal to any nonlinear function.

Remaining problem. Since none of the existing methodologies is the most adequate, we need to come up with a new most adequate fuzzy methodology.

4 Towards the Most Adequate Fuzzy Methodology

What is given: reminder. We are given fuzzy rules of the type

$$\text{If } x \text{ is } A_i \text{ then } y \text{ is } B_i,$$

for $i = 1, \ldots, k$, and we know the membership functions $A_i(x)$ and $B_i(y)$ describing these rules.

What we want: reminder. We want to make sure that when, for some function $f(x)$, we have $B_i = f(A_i)$ for all i, i.e., we have $\mathbf{B}_i(\alpha) = f(\mathbf{A}_i(\alpha))$ for all i and for all α, then this methodology should reconstruct the function $f(x)$. This prompts the following seemingly natural definition.

A seemingly natural idea. Let us return a function $f(x)$ for which, for all i and for all α, we have

$$\mathbf{B}_i(\alpha) = f(\mathbf{A}_i(\alpha)).$$

A problem with this idea. Expert knowledge is usually approximate. As a result, the membership function B_i may be slightly different from $f(A_i)$. In this case, we may not have a function $f(x)$ for which, in the above equation, we have exact equality.

A natural solution to this problem and the resulting description of the new fuzzy methodology. In view of the approximate character of expert knowledge, let us look for a function $f(x)$ for which

$$\mathbf{B}_i(\alpha) \approx f(\mathbf{A}_i(\alpha)).$$

We can interpret these approximate equalities, e.g., by using the usual least squares approach (see, e.g., [7]):

$$\sum_{i,\alpha} d^2(\mathbf{B}_i(\alpha), f(\mathbf{A}_i(\alpha))) \to \min,$$

where the distance between the two intervals $[\underline{a}, \overline{a}]$ and $[\underline{b}, \overline{b}]$ can be defined, e.g., as the Euclidean distance between the corresponding 2-D points $(\underline{a}, \overline{a})$ and $(\underline{b}, \overline{b})$:

$$d^2([\underline{a}, \overline{a}], [\underline{b}, \overline{b}]) = (\underline{a} - \underline{b})^2 + (\overline{a} - \overline{b})^2.$$

Case of monotonicity. In the control situation that we used as an example, the desired function $f(x)$ is decreasing. In general, situations in which the function $f(x)$ is increasing or decreasing are ubiquitous. In such situation, the above minimization problem can be simplified.

To describe this simplification, let us denote the endpoint of the interval $\mathbf{A}_i(\alpha)$ by $\underline{A}_i(\alpha)$ and $\overline{A}_i(\alpha)$, so that

$$\mathbf{A}_i(\alpha) = [\underline{A}_i(\alpha), \overline{A}_i(\alpha)].$$

Similarly, let us denote the endpoint of the interval $\mathbf{B}_i(\alpha)$ by $\underline{B}_i(\alpha)$ and $\overline{B}_i(\alpha)$, so that

$$\mathbf{B}_i(\alpha) = [\underline{B}_i(\alpha), \overline{B}_i(\alpha)].$$

In these terms, we can explicitly describe the expression for the range $f(\mathbf{A}_i(\alpha))$:

- If the function $f(x)$ is increasing, then

$$f(\mathbf{A}_i(\alpha)) = f([\underline{A}_i(\alpha), \overline{A}_i(\alpha)]) = [f(\underline{A}_i(\alpha)), f(\overline{A}_i(\alpha))].$$

- If the function $f(x)$ is decreasing, then

$$f(\mathbf{A}_i(\alpha)) = f([\underline{A}_i(\alpha), \overline{A}_i(\alpha)]) = [f(\overline{A}_i(\alpha)), f(\underline{A}_i(\alpha))].$$

In this case, the minimized expression becomes simpler:

- If we know that the function $f(x)$ is increasing, then, according to the proposed methodology, we should select the function $f(x)$ that minimizes the expression

$$\sum_{i=1}^{k}(\underline{B}_i(\alpha) - f(\underline{A}_i(\alpha)))^2 + \sum_{i=1}^{k}(\overline{B}_i(\alpha) - f(\overline{A}_i(\alpha)))^2.$$

- If we know that the function $f(x)$ is decreasing, then, according to the proposed methodology, we should select the function $f(x)$ that minimizes the expression

$$\sum_{i=1}^{k}(\underline{B}_i(\alpha) - f(\overline{A}_i(\alpha)))^2 + \sum_{i=1}^{k}(\overline{B}_i(\alpha) - f(\underline{A}_i(\alpha)))^2.$$

Often, we look for a function $f(x)$ as a linear combination of functions from the given basis, i.e., as an expression

$$f(x) = C_1 \cdot e_1(x) + \ldots + C_m \cdot e_m(x),$$

where the functions $e_j(x)$ are given and the coefficients C_j need to be determined. For example, we cna take $e_1(x) = 1$, $e_2(x) = x$, and $e_j(x) = x^{j-1}$; in this case, we are looking for a polynomial function $f(x)$. In this case, the above minimized expression become quadratic in terms of the unknown coefficients C_j. Thus, differentiating with respect to each of these coefficient and equating the derivatives to 0, we get an easy-to-solve system of linear equations for finding C_j.

Case of several inputs. Sometimes, we have rules whose conditions involve several inputs $x_1, \ldots, x_n$, i.e., rules of the type

$$\text{If } x_1 \text{ is } A_{i1} \text{ and } \ldots \text{and } x_n \text{ is } A_{in} \text{ then } y \text{ is } B_i.$$

Based on these rules, we need to find an appropriate function $y = f(x_1, \ldots, x_n)$.

In this case, Zadeh's extension principle takes the following form:

$$\mathbf{B}_i(\alpha) = f(\mathbf{A}_{i1}(\alpha), \ldots, \mathbf{A}_{in}(\alpha)),$$

where for every tuple of sets $S_1, \ldots, S_n$, the range $f(S_1, \ldots, S_n)$ means

$$f(S_1, \ldots, S_n) \stackrel{\text{def}}{=} \{f(x_1, \ldots, x_n) : x_1 \in S_1, \ldots, \text{ and } x_n \in S_n\}.$$

In this case, according to the proposed methodology, we should select the function $f(x_1, \ldots, x_n)$ for which

$$\mathbf{B}_i(\alpha) \approx f(\mathbf{A}_{i1}(\alpha), \ldots, \mathbf{A}_{in}(\alpha))$$

for all i and for all α, i.e., for example, for which the following expression attains the smallest possible value:

$$\sum_{i,\alpha} d^2(\mathbf{B}_i(\alpha), f(\mathbf{A}_{i1}(\alpha), \ldots, \mathbf{A}_{in}(\alpha))) \to \min.$$

Is this new methodology indeed the most adequate? Of course, by definition, if $f(A_i) = B_i$ for all i, then $f(x)$ is *one* of the functions satisfying the above condition.

Is this the only function with this property? Not necessarily: if all membership functions are constant on some interval $[\underline{x}, \overline{x}]$ – e.g., if we consider trapezoid functions – then all we can extract from the given information is the range of the function $f(x)$ on this interval, but we cannot uniquely determine how exactly the function $f(x)$ behaves on this interval:

- this function can be linear on this interval,
- it can be non-linear on this interval,

the membership functions $B_i(x)$ will be the same.

However, if take into account that in control situations similar to the one described above, the function $f(x)$ is either strictly increasing or strictly decreasing, then we can prove that the above exception is the only case when we cannot uniquely reconstruct the original function $f(x)$: in all other cases, the function $f(x)$ can be uniquely reconstructed.

Proposition 1. *Let $A_1(x)$, ..., $A_n(x)$ be continuous membership functions on an interval $[\underline{X}, \overline{X}]$ such that for every value x – except maybe a finite set of values – one of these membership functions is either strictly increasing or strictly decreasing in some neighborhood of this point. If for some increasing continuous functions $f(x)$ and $g(x)$, we have $f(A_i) = g(A_i)$ for all i, then for all $x \in [\underline{X}, \overline{X}]$, we have $f(x) = g(x)$.*

Proposition 2. *Let $A_1(x)$, ..., $A_n(x)$ be continuous membership functions on an interval $[\underline{X}, \overline{X}]$ such that for every value x – except maybe a finite set of values – one of these membership functions is either strictly increasing or strictly decreasing in some neighborhood of this point. If for some decreasing continuous functions $f(x)$ and $g(x)$, we have $f(A_i) = g(A_i)$ for all i, then for all $x \in [\underline{X}, \overline{X}]$, we have $f(x) = g(x)$.*

Proof. Let us show how to prove Proposition 1; the proof of Proposition 2 is similar. Let us take a point x from the given interval, and let us prove that $f(x) = g(x)$. Let $A_i(x)$ be the membership function which is either strictly increasing or strictly decreasing in the vicinity of the point x. As before, let us denote the endpoints of the interval $\mathbf{A}_i(\alpha)$ by $\underline{A}_i(\alpha)$ and $\overline{A}_i(\alpha)$, so that

$$\mathbf{A}_i(\alpha) = [\underline{A}_i(\alpha), \overline{A}_i(\alpha)].$$

Since the function $f(x)$ is increasing, we have

$$f(\mathbf{A}_i(\alpha)) = f([\underline{A}_i(\alpha), \overline{A}_i(\alpha)]) = [f(\underline{A}_i(\alpha)), f(\overline{A}_i(\alpha))].$$

Again, without losing generality, we can assume that x belongs to the increasing part of $A_i(x)$. In this case, the values $f(\underline{A}_i(\alpha))$ strictly increase with α, so there exists a value α for which $\underline{A}_i(\alpha)) = x$. For this value α, we have

$$f(\mathbf{A}_i(\alpha)) = [f(x), f(\overline{A}_i(\alpha))].$$

Similarly, we have

$$g(\mathbf{A}_i(\alpha)) = [g(x), g(\overline{A}_i(\alpha))].$$

Since we have $f(A_i) = g(A_i)$, we thus have

$$f(\mathbf{A}_i(\alpha)) = g(\mathbf{A}_i(\alpha))$$

for all α, therefore

$$[f(x), f(\overline{A}_i(\alpha))] = [g(x), g(\overline{A}_i(\alpha))]$$

and hence, $f(x) = g(x)$.

The equality $f(x) = g(x)$ is thus proven for all points x with the exception of finite many points. For each remaining point, this equality can be proved by continuity – since each of these points is a limit of nearby points which are not in this finite list. The proposition is proven.

Discussion. For analytical functions – i.e., functions that can be expanded in Taylor series in the neighborhood of each point – we can have even stronger results.

Proposition 3. *Let $A_1(x)$, ..., $A_n(x)$ be continuous membership functions such that on an interval $[\underline{x}, \overline{x}]$ one of these membership functions is either strictly increasing or strictly decreasing. If for some increasing analytical functions $f(x)$ and $g(x)$, we have $f(A_i) = g(A_i)$ for all i, then we have $f(x) = g(x)$ for all x.*

Proposition 4. *Let $A_1(x)$, ..., $A_n(x)$ be continuous membership functions such that on an interval $[\underline{x}, \overline{x}]$ one of these membership functions is either strictly increasing or strictly decreasing. If for some decreasing analytical functions $f(x)$ and $g(x)$, we have $f(A_i) = g(A_i)$ for all i, then we have $f(x) = g(x)$ for all x.*

Proof. Similarly to the proof of Propositions 1 and 2, we can conclude that the functions $f(x)$ and $g(x)$ coincide on the interval $[\underline{x}, \overline{x}]$. It is known that if two analytical functions coincide on some interval, then they are equal everywhere. The proposition is proven.

Discussion: we should be cautious when trying to extend this result to functions of several variables. For functions of two or more variables, the new methodology leads to reasonable results if we restrict ourselves to a finite-parametric family of functions – e.g., to linear combinations of known functions

$$f(x_1, \ldots, x_n) = C_1 \cdot e_1(x_1, \ldots, x_n) + \ldots + C_m \cdot e_m(x_1, \ldots, x_n),$$

where

$$e_1(x_1, \ldots, x_n), \ldots, e_1(x_1, \ldots, x_n)$$

are given functions and $C_1, \ldots, C)m$ are the coefficients that need to be determined.

However, it should be mentioned that, in contrast to the 1-D case, if we do not impose any such restriction, then, in general. the proposed minimization does not determine a unique function $f(x_1, \ldots, x_n)$. Indeed, the desired criterion only described the ranges $[\underline{y}_i(\alpha), \overline{y}_i(\alpha)]$ of the function $f(x_1, x_2, \ldots)$ on all α-cuts for all rules $i = 1, \ldots, k$. So, all we have is $2k$ functions of one variable $\underline{y}_i(\alpha)$ and $\overline{y}_i(\alpha)$, and this information is not sufficient to uniquely determine a function of two or more variables.

What about type-2? Up to now, we only considered what is usually called *type-1* fuzzy sets, when for each property A and for each value x, the degree to which the value x satisfies this property is described by a real number. In practice, just like experts cannot describe the exact values of the corresponding physical quantities, they cannot meaningfully describe their degree of confidence by a single number. It is more realistic to ask the experts to express each of their degrees of confidence by an interval of possible values, or even by a fuzzy subset of the interval $[0,1]$. the function assigning an interval or a fuzzy set to each value x are known as, correspondingly, interval-valued fuzzy sets and type-2 fuzzy sets; see, e.g., [4].

For rules in which properties A_i and B_i are described by such sets, it is also possible to formulate a similar criterion – since both Zadeh's extension principle and its α-cut reformulation can also be naturally extended to the interval-valued and type-2 fuzzy cases; see, e.g., [3].

Acknowledgments

This work was supported in part by the National Science Foundation grants 1623190 (A Model of Change for Preparing a New Generation for Professional Practice in Computer Science), and HRD-1834620 and HRD-2034030 (CAHSI Includes), and by the AT&T Fellowship in Information Technology.

It was also supported by the program of the development of the Scientific-Educational Mathematical Center of Volga Federal District No. 075-02-2020-1478, and by a grant from the Hungarian National Research, Development and Innovation Office (NRDI).

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

[3] V. Kreinovich, "From processing interval-valued fuzzy data to general type-2: towards fast algorithms", *Proceedings of the IEEE Symposium on Advances in Type-2 Fuzzy Logic Systems T2FUZZ'2011*, part of the IEEE Symposium Series on Computational Intelligence, Paris, France, April 11–15, 2011, pp. ix–xii.

[4] J. M. Mendel, *Uncertain Rule-Based Fuzzy Systems: Introduction and New Directions*, Springer, Cham, Switzerland, 2017.

[5] H. T. Nguyen, C. L. Walker, and E. A. Walker, *A First Course in Fuzzy Logic*, Chapman and Hall/CRC, Boca Raton, Florida, 2019.

[6] V. Novák, I. Perfilieva, and J. Močkoř, *Mathematical Principles of Fuzzy Logic*, Kluwer, Boston, Dordrecht, 1999.

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

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