Logic-Motivated Choice of Fuzzy Logic Operators

Pratit Santiprabhob, Assumption U., Thailand, pratit@s-t.au.ac.th

Hung T. Nguyen

Mathematics, New Mexico State University, Las Cruces, NM 88003, USA, hunguyen@nmsu.edu

Witold Pedrycz, U. of Alberta, Canada, pedrycz@ee.ualberta.ca

Vladik Kreinovich, U. Texas at El Paso, USA, vladik@cs.utep.edu

***Abstract*—Many different "and"- and "or"-operations have been proposed for use in fuzzy logic; ; see, e.g., [4], [13]. It is therefore important to select, for each particular application, the operations which are the best for this particular application. Several papers discuss the optimal choice of "and"- and "or"-operations for fuzzy control, when the main criterion is to get the stablest control (or the smoothest or the most robust or the fastest-to-compute). In reasoning applications, however, it is more appropriate to select operations which are the best in reflecting human reasoning, i.e., operations which are "the most logical". In this paper, we explain how we can use logic motivations to select fuzzy logic operations, and show the consequences of this choice. As one of the unexpected consequences, we get a surprising relation with the entropy techniques, well known in probabilistic approach to uncertainty.**

Main Idea. One of the main ideas behind fuzzy logic is that often, we do not have a 100% confidence in a certain statement; to describe the different degrees of such confidence, we can, e.g., use numbers from the interval $[0,1]$.

The more arguments we have in favor of a certain statement A, the larger our degree of confidence in this statement. It is therefore natural to take, as our degree of confidence $d(A)$ in the statement A, the (relative) number of arguments in favor of this statement.

Let us use this natural interpretation of fuzzy degrees to come up with natural logical operations of these degrees.

Selecting an "and"-Operation. Let us start with an "and"-operation (t-norm) $f_\&(a,b)$. The purpose of an "and"-operation is, given our degrees of certainty $a=d(A)$ and $b=d(B)$ in statements A and B, to estimate our degree of certainty $d(A\,\&\,B)$ in a composite statement $A\,\&\,B$. If the values $d(A)$ and $d(B)$ are the only information that we have about A and B, then our estimate for $d(A\,\&\,B)$ can depend only on this available information, i.e., in must be a function of these two numbers. Thus, an estimate must take the form $d(A\,\&\,B)\approx f_\&(d(A),d(B))$ for some function $f_\&:[0,1]\times[0,1]\to[0,1]$. This function $f_\&(a,b)$ is called an *"and"-operation* or a *t-norm.*

From the viewpoint of the above logic-motivated idea, the fact that our degree of certainty in the statement A is equal to $d(A)$ means that we have $d(A)$ arguments in favor of A. Similarly, the fact that our degree of certainty in the statement B is equal to $d(B)$ means that we have $d(B)$ arguments in favor of B. We are in the situation when we have no other information about the statements A and B, in particular, we have no information about the possible relation between these two statements.

In this situation, if we have an argument in favor of "A and B", then (since we have no information about the relation between A and B) we must have an argument in favor of A *and* an argument in favor of B. Vice versa, a pair of arguments, one in favor of A and one in favor of B, forms an argument in favor of $A\,\&\,B$.

Thus, the number of arguments in favor of $A\,\&\,B$ coincides with the numbers of pairs

$$\langle \text{argument in favor of } A,\ \text{argument in favor of } B\rangle$$

The number of such pairs is equal to the product:

$$(\#\text{ of arguments in favor of } A)\times(\#\text{ of arguments in favor of } B),$$

i.e., to $d(A)\cdot d(B)$.

Thus, logic motivates the use of an *algebraic product $f_\&(a,b)=a\cdot b$ as an "and"-operation (t-norm).*

Comments.

- It is worth mentioning that algebraic product is one of the two "and"-operations introduced in the pioneer paper by L. Zadeh [16].
- An alternative justification of the use of this operation comes, e.g., from requiring that it should be, on average, the least sensitive to possible uncertainty in determining the exact values of $d(A)$ and $d(B)$; see [12].
- The interpretation of a composite statement $A\,\&\,B$ as the set of pairs (Cartesian product) $S(A)\times S(B)$ of the sets $S(A)$ and $S(B)$ corresponding to the original statements A and B is not new in mathematical logic: it has been actively used in topos theory (see, e.g., [1], [9]) and in linear logic (see, e.g., [3], [14]). The fact that the same mathematical interpretation appears in linear logic and in fuzzy logic is not surprising – there is a natural relationship between fuzzy logic and linear logic; see, e.g., [7], [10] and references therein.

Selecting an Implication Operation. Let us now continue with an implication operation $f_\to(a,b)$. The purpose of an implication operation is, given our degrees of certainty $a=d(A)$ and $b=d(B)$ in statements A and B, to estimate our degree of certainty $d(A\to B)$ in a composite statement "A implies B".

From the viewpoint of the above logic-motivated idea, the fact that our degree of certainty in the statement A is equal to $d(A)$ means that we have $d(A)$ arguments in favor of A. Similarly, the fact that our degree of certainty in the statement B is equal to $d(B)$ means that we have $d(B)$ arguments in favor of B.

If we have an argument in favor of the implication "A implies B", then, by combining each argument in favor of A with the argument in favor of the implication, we get an argument in favor of the conclusion B. Thus, whenever we have an argument in favor of the implication, we thus have a transformation which transforms each argument in favor of A into an argument in favor of B. In mathematical terms, we thus have a *function* which maps the set $S(A)$ of arguments in favor of a statement A into the set $S(B)$ of arguments in favor of the statement B.

Vice versa, if we have a function which converts every argument in favor of the statement A into an argument in favor of a statement B, then this function can be viewed as an argument in favor of the implication $A \to B$.

Thus, the number of arguments in favor of $A \to B$ coincides with the numbers of functions from the set $S(A)$ to the set $S(B)$. The number of such functions is known to be equal to $d(B)^{d(A)}$.

Thus, logic motivates the use of $f_\to(a,b) = b^a$ *as an implication operation.*

Comments.

• The implication operation $f_\to(a,b) = b^a$ was first introduced by R. Yager and is called *Yager's implication.*

• An alternative justification of the use of Yager's implication comes from requiring that several natural properties of classical implication, such as

$$(A \to B) \,\&\, (A \to C) \equiv (A \to (B \,\&\, C)) \quad \text{and}$$

$$(A \to (B \to C) \equiv (A \,\&\, B) \to C,$$

hold for fuzzy implication operation as well; for details, see [15].

• Strictly speaking, the value b^a is not well defined when $a = b = 0$, and the above definition has to be specifically supplemented by explicitly defining what 0^0 should stand for. At first glance, this may sound like an inconvenience, but in reality, the ambiguity of an implication $A \to B$ for the case when both A and B are false (i.e., when $d(A) = d(B) = 0$) is in good accordance with common sense. In traditional (mathematical) logic, the implication is assumed to be true if both the condition and the conclusion are false. However, from the commonsense viewpoint, phrases like "if the Moon is made of green cheese then $2 + 2 = 5$" do not seem like convincingly true.

• Similarly to the interpretation of a composite statement $A \,\&\, B$ as the set of pairs (Cartesian product) $S(A) \times S(B)$, the interpretation of the implication $A \to B$ as the set of all the functions from the set $S(A)$ to the set $S(B)$ has also been been actively used in topos theory and in linear logic.

We Can Now Interpret If–Then Rules. Once we have selected a fuzzy "and"-operation $f_\&(a,b)$ and a fuzzy implication operation $f_\to(a,b)$, we are able to transform an arbitrary set of fuzzy if–then rules connecting inputs $x_1, \ldots, x_n$ and the output y into a crisp function $y = f(x_1, \ldots, x_n)$. Indeed, let us assume that the relation between the inputs $x_1, \ldots, x_n$ and the output y can be characterized by several if–then rules:

$$(A_{11}(x_1) \,\&\, \ldots \,\&\, A_{1n}(x_n)) \to B_1(y);$$

$$\ldots$$

$$(A_{i1}(x_1) \,\&\, \ldots \,\&\, A_{in}(x_n)) \to B_i(y);$$

$$\ldots$$

$$(A_{m1}(x_1) \,\&\, \ldots \,\&\, A_{mn}(x_n)) \to B_m(y),$$

where $A_{ij}(x_j)$ and $B_i(y)$ are properties expressed by words from natural language. This interpretation consists of the following steps (see, e.g., [11]):

- First, we can use one of the known elicitation techniques to determine the membership functions $\mu^A_{ij}(x_j)$ and $\mu^B_i(y)$ corresponding to the words $A_{ij}(x_j)$ and $B_i(y)$.

- Then, we can use the fuzzy "and" operation $f_\&(a,b) = a \cdot b$ to determine, for each rule i and for given input $x_1, \ldots, x_n$, the degree c_i to which the given input satisfies the conditions $A_{i1}(x_1) \,\&\, \ldots \,\&\, A_{in}(x_n)$ of the given rule. This value is equal to $c_i = \mu^A_{i1}(x_1) \cdot \ldots \cdot \mu^A_{in}(x_n)$.

- Next, we use the fuzzy implication operation $f_\to(a,b) = b^a$ to determine, for each rule i, for given input $x_1, \ldots, x_n$, and for an arbitrary value y, the degree d_i to which this value y satisfies this rule. This value is equal to

$$d_i(y) = \left(\mu^B_i(y)\right)^{c_i} = \left(\mu^B_i(y)\right)^{\mu^A_{i1}(x_1)\cdot\ldots\cdot\mu^A_{in}(x_n)}.$$

- Next, we use the fuzzy "and"-operation $f_\&(a,b) = a \cdot b$ to determine, for the given input and for an arbitrary value y, the degree $d(y)$ to which the value y satisfies all m rules, i.e., Rule 1 *and* Rule 2 … *and* Rule m. Since we already know the degrees $d_1(y), \ldots, d_m(y)$ to which each of these rules is satisfied, we can thus determine the desired degree $d(y)$ as $d(y) = d_1(y) \cdot \ldots \cdot d_m(y)$.

- Finally, for the given input $x_1, \ldots, x_n$, we find the desired value y as the value for which the degree $d(y)$ is the largest possible: $d(y) \to \max\limits_y$. It is worth mentioning that for the frequently used Mamdani approach, the selection of the largest degree does not lead to a very good control, so more sophisticated defuzzification methods – like centroid defuzzification [4] – have to be used. We will see that in our case, maximum works just fine, and more sophisticated defuzzification techniques are not necessary.

In the following sections, we will show that for several important rule bases, the logic-motivated operations $f_\&(a,b) = a \cdot b$ and $f_\to(a,b) = b^a$ indeed lead to a natural function $y = f(x_1, \ldots, x_n)$.

Logic-Motivated Fuzzy Logic Operations Lead to a Natural Interpolation. Let us start with a simplest example of if–then rules, in which we only have one input variable $x = x_1$, and all the properties of the output variable y are exactly the same as for the input variable x_1. For such examples, since all the properties of the output y are exactly the same as the properties of the input x, it is natural to expect that the defuzzification procedure would lead to $y = x$. We will show, on a simple example, that for Mamdani approach with a centroid defuzzification, we do not get $y = x$, but for our approach, with logic-motivated fuzzy logic operations, we indeed get the expected function $y = x$.

For simplicity, let us assume that both x and y take values in the interval $[0, 1]$ and we have two rules:

If x is small, then y is small.
If x is large, then y is large.

Here, $n = 1$, $m = 2$, $A_{11} = B_1$ ="small" and $A_{21} = B_2$ ="large".

Since we only consider values from the interval $[0, 1]$, the largest value ($= 1$) from this interval should be considered large, while the smallest value ($= 0$) from this interval should be absolutely not large. Thus, the membership function $\mu^A_{21}(x)$ for "large" should be equal to 0 for $x = 0$ and to 1 for $x = 1$. the simplest such function is $\mu^A_{21}(x) = x$.

Similarly, the membership function $\mu^A_{11}(x)$ for "small" should be equal to 1 for $x = 0$ and to 0 for $x = 1$. the simplest such function is $\mu^A_{11}(x) = 1 - x$.

One can easily check that for Mamdani approach with centroid defuzzification (a standard approach in fuzzy control), we *do not* get $y = x$. However, for the logic-based fuzzy operations, we do get the desired function:

Proposition 1. *For the above rules, logic-based fuzzy logic operations lead to $y = x$.*

(For reader's convenience, all the proofs are placed at the end of the paper.)

A Natural Derivation of the Standard Fuzzy Negation. We can apply the same approach to the determination of the fuzzy "negation" operation. In classical logic, there are only two truth values: "true" and "false". Therefore, we can describe the classical negation $y = \neg x$ by the following two if–then rules:

If x is false, then y is true.
If x is true, then y is false.

We can use these same rules to describe fuzzy negation. For a truth value $x \in [0, 1]$ which is different from 0 and 1, the statement "x is true" becomes fuzzy. the value x itself describes to what extent it is true: 1 means absolutely true, 0 means not true at all, intermediate values mean "true to some extent". Thus, as the truth value $\mu^A_{21}(x)$ that x is true, it is natural to take this same value x.

Similarly, as a truth value $\mu^A_{11}(x)$ that x is false, it is natural to take $1-x$. Correspondingly, we get $\mu^B_1(y) = y$ and $\mu^B_2(y) = 1 - y$.

It turns out that for this natural choice, the above scheme leads to the standard negation operation:

Proposition 2. *For the above rules, logic-based fuzzy logic operations lead to $y = 1 - x$.*

The Resulting Fuzzy "Or"-Operation: "Algebraic Sum". A similar approach can select the "or" operation (a t-conorm). Specifically, the classical "or" can be described by the following four if–then rules:

If x_1 is false and x_2 is false, then y is false.
If x_1 is false and x_2 is true, then y is true.
If x_1 is false and x_2 is true, then y is true.
If x_1 is true and x_2 is true, then y is true.

With the same membership functions as for negation, we get the "algebraic sum" t-conorm as a result:

Proposition 3. *For the above rules, logic-based fuzzy logic operations lead to $y = x_1 + x_2 - x_1 \cdot x_2$.*

Is This Approach Consistent? Checking That It Returns the Original Fuzzy "And"-Operation. What if we apply this same approach to reconstruct the "and" operation? The classical "and" can be described by the following four if–then rules:

If x_1 is false and x_2 is false, then y is false.
If x_1 is false and x_2 is true, then y is false.
If x_1 is false and x_2 is true, then y is false.
If x_1 is true and x_2 is true, then y is true.

If we use the same membership functions as for negation and for "or", then Mamdani's approach with defuzzification leads to a function which is *different* from the original t-norm. The above logic-motivated approach is *consistent* in the sense that we get the exact same "and"-operation $f_\&(a, b) = a \cdot b$ back:

Proposition 4. *For the above rules, logic-based fuzzy logic operations lead to $y = x_1 \cdot x_2$.*

Comment. It would be interesting to find out what other fuzzy logic operations are "consistence" in this sense.

Relation to Entropy. Since we started talking about consistency, let us go back to the example of a simple interpolation. In this example, we argued, in fact, that from the commonsense viewpoint, the rules like $A_1(x) \rightarrow A_1(y), \ldots, A_m(x) \rightarrow A_m(y)$, should be true for $y = x$. When the properties A_i are fuzzy, these rules still hold, but only to a certain degree.

The degree d to which the above rules hold for all values of x is equal to

$$d = \prod_x \left(\left(\mu^A_1(x)\right)^{\mu^A_1(x)} \cdot \ldots \cdot \left(\mu^A_m(x)\right)^{\mu_m(x)} \right).$$

Maximizing d is equivalent to maximizing its logarithm, i.e., the value

$$L = \sum_x \left(\mu^A_1(x) \cdot \ln\left(\mu^A_1(x)\right) + \ldots + \mu^A_m(x) \cdot \ln\left(\mu^A_m(x)\right)\right).$$

This expression is similar to the expression for *entropy* S in probability theory and information theory (see, e.g., [2]); namely, the entropy of a probability distribution characterized by probabilities $p_1, \ldots, p_n$ is equal to:

$$S = -\left(p_1 \cdot \ln(p_1) + \ldots + p_m \cdot \ln(p_m)\right).$$

This relation is not just a coincidental similarity between the two formulas: it can be shown, e.g., that if we use the maximum entropy approach to select the most appropriate "and" and "or" operations, we get exactly the same operations $f_\&(a,b) = a \cdot b$ and $f_\vee(a,b) = a + b - a \cdot b$ as our logic-motivated approach [6].

There may also be a relation between the fact that the Maximum Entropy principle in statistics is often used to justify Gaussian distribution, and the fact that Gaussian membership functions are often used in fuzzy logic methodology (see, e.g., [4] and references therein; please note that there are alternative explanations of Gaussian membership functions; see, e.g., [8]).

A similar relation with entropy techniques can be made for the case when conclusions differ from the conditions, i.e., for the rules of the type $A_i(x) \to B_i(x)$, $1 \le i \le m$. In this case, the degree to which all these rules are satisfied is equal to:

$$d' = \prod_x \left(\left(\mu_1^B(x)\right)^{\mu_1^A(x)} \cdot \ldots \cdot \left(\mu_m^B(x)\right)^{\mu_m^A(x)} \right).$$

Since $B_i \neq A_i$, this degree is smaller than the degree d corresponding to $B_i = A_i$. The decrease can be characterize by the ratio d'/d. The logarithm L' of this ratio is equal to

$$\sum_x \left(\mu_1^A(x) \cdot \ln\left(\frac{\mu_1^B(x)}{\mu_1^A(x)}\right) + \ldots + \mu_m^A(x) \cdot \ln\left(\frac{\mu_m^B(x)}{\mu_m^A(x)}\right) \right).$$

This expression is similar to the known expression for a *relative entropy*:

$$S' = -\left(p_1 \cdot \ln\left(\frac{q_1}{p_1}\right) + \ldots + p_m \cdot \ln\left(\frac{q_m}{p_m}\right) \right).$$

Proof of Proposition 1. In accordance with the above description, for a given x, and for an arbitrary y, the degree $d(y)$ is equal to $y^x \cdot (1-y)^{1-x}$. Maximizing this degree is equivalent to maximizing its logarithm, i.e., the value $L(y) \stackrel{\text{def}}{=} x \cdot \ln(y) + (1-x) \cdot \ln(1-y)$. Differentiating $L(y)$ w.r.t. y and equating the derivative to 0, we conclude that $x/y - (1-x)/(1-y) = 0$; subtracting the fractions, we get $y = x$.

Proof of Proposition 2. Similarly to the proof of Proposition 1, we get $d(y) = y^{1-x} \cdot (1-y)^x$. Differentiating the logarithm of this expression w.r.t. y and equating the resulting derivative to 0, we get the desired value of y.

Proof of Proposition 3. Here,

$$d(y) = (1-y)^{(1-x_1)\cdot(1-x_2)} \cdot y^{(1-x_1)\cdot x_2} \cdot y^{x_1 \cdot (1-x_2)} \cdot \cdot y^{x_1 \cdot x_2}.$$

Combining powers of y together, we conclude that $d(y) = (1-y)^{(1-x_1)\cdot(1-x_2)} \cdot y^{x_1 + x_2 - x_1 \cdot x_2}$. Differentiating the logarithm of this expression w.r.t. y and equating the resulting derivative to 0, we get the desired value of y.

Proof of Proposition 4. This case is similar to the proof of Proposition 3, with the only difference that $d(y)$ is equl to:

$$(1-y)^{(1-x_1)\cdot(1-x_2)} \cdot (1-y)^{(1-x_1)\cdot x_2} \cdot (1-y)^{x_1 \cdot (1-x_2)} \cdot \cdot y^{x_1 \cdot x_2}.$$

Combining powers of $1-y$ together, we conclude that $d(y) = (1-y)^{1-x_1 \cdot x_2} \cdot y^{x_1 \cdot x_2}$. Differentiating the logarithm of this expression w.r.t. y and equating the resulting derivative to 0, we get the desired value of y.

Acknowledgments

This work was supported in part by NASA grants NCC5-209 and NCC 2-1232, NSF grants CDA-9522207 and 9710940 Mexico/Conacyt, Air Force Office of Scientific Research grant F49620-00-1-0365, and by Grant No. W-00016 from the U.S.-Czech Science and Technology Joint Fund.

The authors are thankful to anonymous referees for their valuable comments.

References

[1] J. Bell, *Toposes and Local Set Theories: An Introduction*, Oxford University Press, Oxford, 1988.

[2] G. J. Erickson, J. T. Rychert, and C. R. Smith (eds.), *Maximum Entropy and Bayesian Methods*, Kluwer, Dordrecht, 1998.

[3] J.-Y. Girard, "Linear logic", *Theoretical Computer Science*, 1987, Vol. 50, pp. 1–102.

[4] G. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Prentice Hall, Upper Saddle River, NJ, 1995.

[5] V. Kreinovich, G. C. Mouzouris, and H. T. Nguyen, "Fuzzy rule based modeling as a universal control tool", In: H. T. Nguyen and M. Sugeno (eds.), *Fuzzy Systems: Modeling and Control*, Kluwer, Boston, MA, 1998, pp. 135–195.

[6] V. Kreinovich, H. T. Nguyen, and E. A. Walker, "Maximum entropy (MaxEnt) method in expert systems and intelligent control: new possibilities and limitations", In: K. M. Hanson and R. N. Silver (eds.), *Maximum Entropy and Bayesian Methods*, Kluwer, Dordrecht, 1996, pp. 93–100.

[7] V. Kreinovich, H. T. Nguyen, and P. Wojciechowski, "Fuzzy Logic as Applied Linear Logic", *BUSEFAL*, No. 67, July 1996, p. 4–13.

[8] 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.

[9] F. W. Lawvere (ed.), *Toposes, Algebraic Geometry and Logic*, Springer Lecture Notes in Mathematics, 1972, Vol. 274, pp. 1–12.

[10] H. T. Nguyen and V. Kreinovich, "Fuzzy Logic, Logic Programming, and Linear Logic: Towards a New Understanding of Common Sense", *Proceedings of NAFIPS'96, Biennial Conference of the North American Fuzzy Information Processing Society*, Berkeley, CA, June 20-22, 1996, pp. 546–550.

[11] H. T. Nguyen and V. Kreinovich, "Methodology of fuzzy control: an introduction", In: H. T. Nguyen and M. Sugeno (eds.), *Fuzzy Systems: Modeling and Control*, Kluwer, Boston, MA, 1998, pp. 19–62.

[12] H. T. Nguyen, V. Kreinovich, and D. Tolbert, "A measure of average sensitivity for fuzzy logics", *International Journal on Uncertainty, Fuzziness, and Knowledge-Based Systems*, 1994, Vol. 2, No. 4, pp. 361–375.

[13] H. T. Nguyen and E. A. Walker, *First Course in Fuzzy Logic*, CRC Press, Boca Raton, FL, 1999.

[14] A. S. Troelstra, *Lectures on linear logic*, CSLI, Stanford, 1992.

[15] I. B. Türkşen, V. Kreinovich, and R. R. Yager, "A new class of fuzzy implications (axioms of fuzzy implication revisited)", *Fuzzy Sets and Systems*, 1998, Vol. 100, pp. 267–272.

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