Fuzzy Modus Ponens as a Calculus of Logical Modifiers: Towards Zadeh's Vision of Implication Calculus

Bernadette Bouchon-Meunier[1] and Vladik Kreinovich[2,1]

[1]LAFORIA-IBP
University Paris VI, Case 169
4, place Jussieu
75252 Paris Cedex 05, France
email bouchon@laforia.ibp.fr

[2]Department of Computer Science
University of Texas at El Paso
El Paso, TX 79968, USA
email vladik@cs.utep.edu

Abstract

If we know that an implication $A \to B$ is to some extent true, and if we know that a statement A is true "to some extent", i.e., if we know that the statement $m(A)$ is true, for some logical modifier m ("approximately", "slightly", etc.), then we can conclude that B is also true to some extent, i.e., that $m'(B)$ is true for some logical modifier m'.

Thus, every fuzzy implication defines a mapping from modifiers to modifiers. This mapping describes the meaning of the fuzzy implication, so, it is desirable to find out what mapping are possible. This desire is in line with Zadeh's suggestion that the future of fuzzy logic lies with treating fuzzy implications as a calculus, rather than as analytical formulas.

In this paper, we formally define implication as a mapping from modifiers to modifiers that satisfy some reasonable properties. For an important particular case of *invertible* mappings, we get a complete description of all modifier-to-modifier mappings that characterize fuzzy implication.

Our main mathematical result can be also used in the foundations of knowledge elicitation.

1 Fuzzy Modus Ponens as a Calculus of Logical Modifiers: Main Idea

In classical (2-valued) logic, every statement is either true or false. For such statements, *modus ponens* means that:

- if we know that A is true, and
- if we know that A implies B (i.e., that the implication $A \to B$ is true),
- then, we can conclude that B is also true.

Fuzzy logic is an extension of classical logic that enables us to consider not only statements that are true or false, but also statements that are true "to a certain extent". For such statements, it is possible that A is only true to a certain extent, and that the implication $A \to B$ is only true to a certain extent. The question is: what can we then say about B?

In particular:

- if we know that (to some extent) A implies B, e.g., that "young" implies "healthy", and
- if we know that somebody is *approximately* young, or *slightly* young,
- then what can we conclude about that person's health?

Clearly, we can conclude that that person is to some extent healthy, the question is: to what exactly extent?

In other words, "fuzzy implication" $A \to B$ means that if we know that the statement $m(A)$ is true for a certain logical modifier m (like "approximately", "slightly", etc.), then we will be able to conclude that $m'(B)$ is true for some other modifier m'. So, from mathematical viewpoint, we get a *mapping* that maps each modifier m into some other modifier m'.

There are two ways to describe this mapping:

- We can describe this mapping *indirectly*, i.e., we can use the known formulas for fuzzy implication (see, e.g., [2, 3, 4]), and find the mapping that these formulas lead to. For several known implication operations, such mappings were (partially) described in [1].
- We can also try to describe this mapping *directly*. In other words, we can try to describe fuzzy modus ponens as a calculus of logical modifiers. This idea is in line with the idea of L. Zadeh who suggested that the future of fuzzy logic lies with the development of implication *calculus*, i.e., a *logical* approach to fuzzy implication, as opposed to the currently prevailing analytical (formula-based) approach.

In this paper, we will start the direct description of such mappings by describing all possible mappings from modifiers to modifiers that satisfy certain reasonable properties.

2 Desirable Properties of the Calculus of Logical Modifiers: Motivations

To every fuzzy implication $A \to B$, there corresponds a mapping F that transforms a modifier m characterizing the assumption A of this implication into a modifier $m' = F_m$ that characterizes its conclusion B. Let us describe the natural properties of this mapping F.

2.1 First property: composition turns into composition

Usually, the fact that an implication $A \to B$ is true only to a certain extent, means that the resulting degree of belief in the conclusion B is "weaker" that the degree of belief in the assumption A. In other words, if we know that "approximately" A is true, we will only conclude that, say, "to some extent" B is true. If this implication turns $m =$"approximately" into $m' =$"to some extent", then it is natural to assume that "approximately approximately"A ($m(m(A))$ is transformed into "to some extent to some extent"B. In other words, it is natural to assume that in any combination of modifiers, "approximately" for A is replaced by "to some extent" for B.

In mathematical terms, a composition $m_1(m_2(x))$ of modifiers m_1 and m_2 means a *composition* of the corresponding functions m_1 and m_2 (this composition is usually denoted by $m_1 \circ m_2$). In these terms, we can conclude that *composition turns into a composition*, i.e., if we denote the mapping by F, we get $F_{m_1 \circ m_2} = F_{m_1} \circ F_{m_2}$.

2.2 Every modifier can be obtained as F_m for some m

For an implication, "approximately"A only leads to a "to a certain extent"B. Intuitively, this does not mean that we cannot conclude "approximately"B: it just means that we have to make a very strong assumption about A so that, even after the weakening caused by the implication, we will still have "approximately"B.

In other words, it is natural to assume that whatever modifier m' we take for B, there exists a (strong enough) modifier m for which $m(A)$ implies $m'(B)$, i.e., for which $F_m = m'$.

2.3 Additional property: Different modifiers transform into different ones

The problem of describing all implication operations that satisfy the above two properties is still an open problem. In this paper, we will give a complete classification of all implication operations that satisfying the following *additional* property: *If we start with two different modifiers $m_1 \neq m_2$ that describe slightly* *different degrees of belief in A, we should end up with different conclusions about B, i.e., $m'_1 \neq m'_2$.*

If an implication operation satisfies this property, then not only a modifier at A uniquely determines the modifier at B, but the inverse is also true: if we know the modifier m' at the conclusion B, then we can uniquely determine the modifier at A that has caused this m'. In view of this remark, we will call implication operations that satisfy this property *invertible*.

3 Definitions and the Main Result

Definition 1.

- *By a logical modifier, we mean a function $m : [0,1] \to [0,1]$. The set of all logical modifiers will be denoted by M.*
- *We say that a mapping $F : M \to M$ describes implication iff this mapping F satisfies the following three properties:*
  - *The mapping F turns composition into composition, i.e., $F_{m_1 \circ m_2} = F_{m_1} \circ F_{m_2}$ for all m_1 and m_2.*
  - *The mapping F is "onto", i.e., for every $m' \in M$, there exists an $m \in M$ for which $F_m = m'$.*

Comment. The complete description of all mappings that describe implication is an open problem. In this paper, we will describe all the mappings that satisfy the following additional property:

Definition 2. *The mapping $F : M \to M$ that describes implication is called invertible if it is a one-to-one mapping, i.e., if $m \neq m'$ implies $F_m \neq F_{m'}$.*

The following theorem characterizes all such mappings F:

THEOREM 1. *For every mapping $F : M \to M$, the following two conditions are equivalent to each other:*

- *F describes an invertible implication.*
- *There exists a one-to-one function $\psi(x) : [0,1] \to [0,1]$ for which the mapping F has the form $m(x) \to F_m(x) = \psi(m(\psi^{-1}(x)))$.*

Comment. For reader's convenience, the proofs of this and other theorems are placed in the last section.

4 Possible Applications of Our Main Result to the Foundations of Knowledge Elicitation

Another possible application of this mathematical result is to *knowledge elicitation.*

4.1 It is necessary to elicit the value of logical modifiers

For every expert, the logical modifiers have different meanings; so, if we want to represent the expert's knowledge accurately, we must find out what exactly modifiers like "very", "slightly", etc., mean for a given expert.

How can we do that?

4.2 It seems natural to ask an expert to order his modifiers, but for an expert, it is often difficult

A natural idea is to ask the expert to *order* his modifiers, but it is often very difficult to *order* the modifiers.

For example, if we view "possible" and "probable" as modifiers, it is very difficult to tell which of them represents the greater extent of belief.

4.3 Alternative information: which combinations of modifiers coincide

However, it is often easy for an expert to tell which combinations of modifiers mean the same thing. For example, an expert may say, that "very slightly" A means the same thing to him as simply A. In this case, we can conclude that the composition $m_{\text{very}} \circ m_{\text{slightly}}$ of these two modifiers is equal to the degenerate modifier $m(x) = x$.

4.4 Main problem: informal formulation

With this additional information, we have the following problem:

- We have the set T of all possible terms (i.e., words and combinations of words) that an expert uses to describe his modifiers.
- We know (after asking experts) which pairs of terms correspond to one and the same modifier.
- We would like to reconstruct the functions $m(x) : [0,1] \to [0,1]$ that correspond to different modifiers.

To find out how uniquely we can reconstruct the desired modifiers functions from this information, let us formulate this problem in precise mathematical terms.

Comment. In this formulation, we will consider the ideal case when we can ask an expert about an arbitrary combination of modifiers. In reality, of course, there are infinitely many modifiers, and we can only ask finitely many questions.

4.5 Main problem: formalization

Definition 3.

- *Let a set T be given. Elements t of the set T will be called terms. On this set T, we have an equivalence relation $\sim$. Let us denote the corresponding set of equivalence classes ("factor-set") by $T/\sim$.*

- *On the set T, an operation $\circ$ is defined. This operation will be called* composition *of terms. The operation $\circ$ is consistent with $\sim$ in the sense that if $t_1 \sim t_1'$ and $t_2 \sim t_2'$, then $t_1 \circ t_2 = t_1' \circ t_2'$. Thus, we can define this operation on the factor-set $T/\sim$.*

- *By an* expert's description of modifiers, *we mean a one-to-one mapping $e : t \to e_t$ from the set $T/\sim$ onto the set M of all possible modifiers for which, for all terms t, t', and t'', $t'' = t \circ t'$ iff $e_{t''} = e_t \circ e_{t'}$.*

- *By a* reconstruction *of expert's description of modifiers, we mean a one-to-one mapping r from the set $T/\sim$ to the set M of of all possible modifiers for which, for all terms t, t', and t'', $t'' = t \circ t'$ iff $r_{t''} = r_t \circ r_{t'}$.*

The question is: *how close is the reconstruction r_t of the expert's description of modifiers to the original expert's description e_t of them?* The following theorem provides an answer to this question:

THEOREM 2.

- *For every reconstruction r, there exists a one-to-one function $\psi(x) : [0,1] \to [0,1]$ for which, for every term t, $r_t(x) = \psi(r_t(\psi^{-1}(x)))$.*

- *For every one-to-one function $\psi : [0,1] \to [0,1]$, the formula $r_t(x) = \psi(r_t(\psi^{-1}(x)))$ defines a reconstruction.*

Thus, we can reconstruct the logical modifiers uniquely modulo a 1-to-1 mapping from $[0,1] \to [0,1]$, i.e., modulo some re-scaling of the truth values from the interval $[0,1]$.

5 Proofs

5.1 Proof of Theorem 1

Comment. In this comment, for the convenience of mathematician readers, we will reformulate our result in purely mathematical terms. This reformulation is *not* needed for the understanding of the proof, so non-mathematical readers can simply skip this comment. From the mathematical viewpoint:

- the set M of functions with a composition operation forms a *semigroup*, and
- the mapping with the property $F_{m_1 \circ m_2} = F_{m_1} \circ F_{m_2}$ is a *homomorphism.*

So, in mathematical terms, the problem is: *to describe all homomorphisms $F : M \to M$ of the semigroup $(M, \circ)$ for which the image F_m of the set M coincides with M: $F_M = M$. End of Comment.*

It is easy to show that the expression $m(x) \to F_m(x) = \psi(m(\psi^{-1}(x)))$ actually describes invertible implication in the sense of Definitions 1 and 2. Let us now show that if F describes an invertible implication (in this sense), then it has this form for some function $\psi(x)$.

1. Let us first describe how F acts on the simplest possible modifiers: constant functions.

1.1. Let us denote a constant function $m(x) = a$ with a value a by c_a. Clearly, $c_a(m(x)) = a$ for all possible modifiers m. Therefore,

$$c_a \circ m = c_a \tag{1}$$

for all m.

Since F describes implication, we have $F_{m_1 \circ m_2} = F_{m_1} \circ F_{m_2}$ for arbitrary m_1 and m_2. In particular, for $m_1 = c_a$ and $m_2 = m$, we conclude that $F_{c_a \circ m} = F_{c_a} \circ F_m$. From (1), we can now conclude that

$$F_{c_a} = F_{c_a} \circ F_m \tag{2}$$

for all possible modifiers $m \in M$.

1.2. Let us use this property (2) to prove that F_{c_a} is also a constant function, i.e., that $F_{c_a} = c_b$ for some $b \in [0, 1]$. Indeed, let us show that for every two numbers x_1 and x_2, we have $F_{c_a}(x_1) = F_{c_a}(x_2)$. To show it, let us take a modifier m' that transforms x_1 into x_2 (e.g., a function that exchanges x_1 and x_2 and leaves all other numbers from the interval $[0, 1]$ unchanged). Since F describes implication, there exists a modifier m for which $F_m = m'$. For this m, the equality (2) turns into $F_{c_a} = F_{c_a} \circ m'$. The fact that the two functions coincide means that their values coincide for all possible values x. In particular, for $x = x_1$, we get $F_{c_a}(x_1) = F_{c_a}(m'(x_1))$. But due to our choice of m', we have $m'(x_1) = x_2$, and so $F_{c_a}(x_1) = F_{c_a}(x_2)$.

Hence, $F_{c_a}(x)$ is indeed a constant function, i.e., its value does not depend on x. If denote this value by $\psi(a)$, then we can conclude that for every $a \in [0, 1]$,

$$F_{c_a} = c_{\psi(a)}. \tag{3}$$

2. The fact that the mapping F is invertible means that different modifiers get transformed into different ones. Thus, if $a \neq a'$, we have $c_a \neq c_{a'}$, $F_{c_a} = c_{\psi(a)} \neq F_{c_{a'}} = c_{\psi(a')}$ and therefore, $\psi(a) \neq \psi(a')$.

3. Let us now prove that for every $a' \in [0, 1]$, there exists an $a \in [0, 1]$ for which $a' = \psi(a)$.

Indeed, from the second property of the mapping F, we conclude that $c_{a'} = F_{m_0}$ for some modifier m_0. Since $c_{a'} = F_{m_0}$ is a constant function, we have $F_{m_0} \circ m' = F_{m_0}$ for all possible modifiers m'. In particular, as m', we can take F_m for an arbitrary modifier m. Then, we get $F_{m_0} \circ F_m = F_{m_0}$. Since F transforms composition into composition, we conclude that $F_{m_0 \circ m} = F_{m_0}$ for all possible modifiers m.

From the fact that F is invertible (i.e., maps different modifiers into different ones), we can now conclude that $m_0 \circ m = m_0$ for all possible modifiers m. Similarly to part 1 of this proof, we can conclude that m_0 is a constant function, i.e., that $m_0 = c_a$ for some $a \in [0, 1]$.

From $c_{a'} = F_{m_0} = F_{c_a}$, we can now conclude that $a' = \psi(a)$.

4. Let us now describe the value of F_m for every other modifier m.

For every a and x, we have $m \circ c_a(x) = m(c_a(x)) = m(a)$. Thus, $m \circ c_a = c_{m(a)}$. Hence, $F_m \circ F_{c_a} = F_{c_{m(a)}}$. From (3), we can conclude that $F_m \circ c_{\psi(a)} = c_{\psi(m(a))}$. Both sides of this equality are constant functions:

- the left-hand side has the value $F_m(\psi(a))$, and
- the right-hand side has the value $\psi(m(a))$.

The fact that these two values coincide means that

$$F_m(\psi(a)) = \psi(m(a)). \tag{4}$$

From this formula, we can find the value $F_m(x)$ of the function F_m at every real number x: it is sufficient to take $a = \psi^{-1}(x)$; then $\psi(a) = x$, and from (4), we deduce the desired form for F.

The theorem is proven.

5.2 Proof of Theorem 2

Let e and r satisfy the conditions of Theorem 2. Since e is 1-to-1 and onto, there exists the inverse mapping $i : m \to i_m$ that maps each modifier into the corresponding term. Then, the composition $F_m = r_{i_m}$ of this inverse mapping $m \to i_m$ and the reconstruction mapping r describes what function will be reconstructed from the term t that, to the expert, meant m.

Since both mappings i_m and r preserve composition and are 1-to-1 and onto, the mapping F also preserves composition and is 1-to-t and onto. Thus, the mapping F satisfies the conditions of Theorem 1. From Theorem 1, we get the desired relation between the original modifier e_t and the reconstructed modifier r_t.

The theorem is proven.

Acknowledgments. This work was partially supported by NASA Grant No. NAG 9-757. This work was done when V.K. was an invited professor at Laforia, University of Paris VI, Summer 1996.

References

[1] B. Bouchon, "Stability of linguistic modifiers compatible with a fuzzy logic", In: B. Bouchon, L. Saitta, and R.R. Yager (eds.), *Uncertainty and Intelligent Systems*, Springer Lecture Notes in Computer Science, Vol. 313, Springer-Verlag, Berlin, 1988, pp. 63–70.

[2] D. Dubois and H. Prade, "Fuzzy sets in approximate reasoning, Part I: inference with possibility distributions", *Fuzzy Sets and Systems*, 1991, Vol. 40, pp. 143–202.

[3] G. Klir and B. Yuan, *Fuzzy sets and fuzzy logic: theory and applications*, Prentice Hall, Upper Saddle River, NJ, 1995.

[4] H. T. Nguyen and E. A. Walker, *A First Course in Fuzzy Logic*, CRC Press, Boca Raton, Florida, 1996 (to appear).