The Relation between Inference and Interpolation in the Framework of Fuzzy Systems

Frank Klawonn

Dept. of Computer Science
University of Braunschweig
Bültenweg 74/75
D–38106 Braunschweig
Germany

Vilém Novák*)

University of Ostrava
Faculty of Natural Sciences
Dept. of Mathematics
Brafova 7
701 Ostrava, Czech Republic

and

Institute of Information and Automation
Theory AV ČR
Pod vodárenskou věží 4, 186 02 Praha 8,
Czech Republic

Abstract

This papers aims at clarifying the meaning of different interpretations of the Max-Min or, more generally, the Max-t-norm rule in fuzzy systems. It turns out that basically two distinct approaches play an important role in fuzzy logic and its applications: fuzzy interpolation on the basis of an imprecisely known function and logical inference in the presence of fuzzy information.

Keywords: Fuzzy logic; fuzzy control; Max-Min rule, fuzzy interpolation.

1 Introduction

This is a synthesizing paper which returns to the question, what is the role of the Max-Min (Max-t-norm) rule in fuzzy logic from the viewpoint of logical inference. We aim at demonstrating that two basic, more or less complementary approaches in fuzzy logic and its applications can be distinguished, namely: fuzzy interpolation of a fuzzily specified precise function and logical inference in the presence of fuzzy information.

The first task is solved using the Max-t-norm rule which essentially leads to search of a fuzzy set which is an image of some fuzzy relation. The whole procedure

*)Parts of this paper have been written during author's research stay at the University of Trento in Italy. The work has also been supported by the grants A1086501 of GA AV ČR and 101/94/1269 of GAČR.

is quite reasonable and gives good results. However, it has to be stressed that it is not a logical inference, i.e., a procedure aiming at the derivation of new facts from some other known ones using formal deduction rules. No logical implication is inside and thus, no modus ponens proceeds. This conclusion is based on the analysis of the Max-t-norm rule as a logical inference which meets unsurmountable problems (cf. [20]). We will return to this question later.

The second task leads to a set of logical deductions in many-valued logic. This can be demonstrated to fulfill all the intuitive expectations about approximate reasoning and opens an extensive field for further study. The logical inference mechanism makes it, moreover, possible to provide interpolation in a similar manner as Max-t-norm rule.

The paper is divided into 5 sections. In Section 2, the Max-Min interpolation rule is analyzed in detail and given reasons for calling it "interpolation". Section 3 focuses on the more general problem of how a function and a relation can be described. We demonstrate that two possible ways exist. The first one leads to the Max-t-norm (Max-Min) interpolation which described in Section 2. The second one leads to logical description using the concept of implication. This second possibility is further discussed in Section 4 where logical inference and a problem of chaining of rules are discussed. Furthermore, we consider also chaining of Max-t-norm rules and show that this brings some problems making it somewhat dubious. Note that chaining of rules is an important question for the design of (fuzzy) expert systems.

We will work with the following structure of truth values (membership degrees):

$$\mathcal{L} = \langle [0,1], \vee, \wedge, \otimes, \rightarrow, 1, 0 \rangle \tag{1}$$

where the operations $\vee$ and $\wedge$ are the operations of supremum (maximum) and infimum (minimum) respectively and $\otimes$, $\rightarrow$ are binary operations of *Łukasiewicz multiplication* (sometimes called also *bold multiplication*) and *residuation* (Łukasiewicz implication), respectively, given by

$$a \otimes b = 0 \vee (a + b - 1) \tag{2}$$

$$a \rightarrow b = 1 \wedge (1 - a + b), \qquad a, b \in [0,1]. \tag{3}$$

We also use the symbol $\sqcap$ for an arbitrary t-norm.

The reasons for choosing (1 – 3) are manyfold and they were explained in [23, 13, 14, 21]).

We will also deal with a formal language J which is the classical first-order language extended by symbols for truth values and some additional connectives (for details see the previously cited works). By F_J we denote a set of all well formed formulas in the language J.

We will use the notation based on that introduced in [13]. For fuzzy sets we use the greek symbols $\mu, \nu \ldots$. Fuzzy sets are functions $\mu : U \longrightarrow [0,1]$. Occasionally, we will write $\mu \underset{\sim}{\subset} U$ to stress that μ is a fuzzy set in the universe U. Explicitly, we will write the fuzzy set μ in the form

$$\{\, \mu(x)/x \mid x \in U\}$$

where U is the universe and $\mu(x) \in [0,1]$ is the membership degree of $x \in U$. By $\mathcal{F}(U) = [0,1]^U$ we denote the set of all the fuzzy sets on U. By dom and rng we denote the domain and range of the function in concern, respectively.

2 Max–Min Rule is Interpolation

Input-output behaviour of technical systems is described in terms of functions. A typical example is the table based control, where a function is constructed that assigns to measured inputs of actual process parameters a suitable control action as output. However, the definition of a function point by point is often an intractable task, especially since the precise values of the function are usually not known. Thus, there is a need for other methods for specification of a function.

Our situation can be described as follows. We intend to find a function $g : U \longrightarrow V$. However, for various reasons, we need not know this function completely but only an approximation of it which is a function, say G. The question arises, how exactly G is given and how well it fits the function g. A very usual situation, encountered in fuzzy control (and other applications of fuzzy logic) is, that G is a fuzzy function of type 2 (cf. [13]), sometimes called a *fuzzy graph*, i.e., it is a (classical, partial) mapping

$$G : \mathcal{F}(U) \longrightarrow \mathcal{F}(V).$$

Note that we explicitly allow G to be a partial mapping so that G may assign an output fuzzy set only to some input fuzzy sets. Of course, we should assume, that the fuzzy graph given by g covers g, i.e. $\mathrm{dom}(g) \subseteq \mathrm{Supp}\,(\bigcup \mathrm{dom}(G))$ and $\mathrm{rng}(g) \subseteq \mathrm{Supp}\,(\bigcup \mathrm{rng}(G))$ where dom and rng denote the domain and the codomain of a classical function. This situation is depicted on Fig. 1. The function G in the picture is outlined by only few circles. However, it follows from the condition on domain and range above that it covers the whole g. It is not intended that G generalizes the mapping $\hat{g}$ defined on the power set of U which assigns to each subset U_0 of U its image $\hat{g}(U_0) = g(U_0) = \{v \in V \mid (\exists u \in U_0)(g(u) = v)\}$ to fuzzy sets. G is understood as the extension of g to fuzzy points, i.e. to fuzzy sets of a certain type. What we understand by a fuzzy point will be explained later on in detail.

Our problem now is to fit the function g by means of G in some way and to be able to find a value $g(x_0) \in V$ to the given argument $x_0 \in U$ by means of G. From the mathematical point of view, interpolation is a possible approach. However, human experts do usually not think in terms of mathematical interpolation techniques, for example, those based on splines. It seems that two simple assumptions lead to a more appropriate model for the way how a human expert considers an input–output function of a technical system.

(i) For some representative inputs the output is at least approximately known. It is sufficient to specify these inputs approximately.

(ii) For more or less similar inputs, the outputs are also similar.

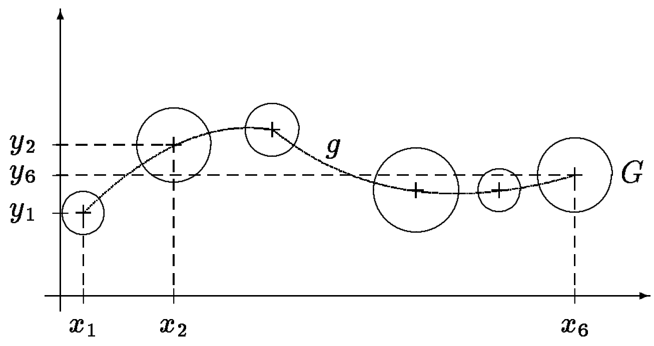

Figure 1: Schematic depiction of functions g and G.

Technically speaking, (i) means that G is described as a partial function assigning output values only to certain inputs. It is not necessary to know precisely either the input or the output values. It is sufficient to specify them only up to a certain precision.

(ii) states that a suitable interpolation in terms of some similarity has to be carried out among these given points.

In both cases, (i) and (ii), it is assumed that there is a notion of (maybe graded) distinguishability or indistinguishability of values. In (i), this indistinguishability is connected to the sufficient but limited precision needed. (ii) requires the distinguishability in order to decide, whether a certain output can be classified as appropriate or not.

It is clear that without additional information or assumptions, the partial function G given due to (i) cannot be extended to g defined for all inputs. To solve this problem, we have to exploit the similarity stated in (ii) which enables us to describe g imprecisely when the specified inputs are representative in the sense that they cover the input space sufficiently dense with respect to the similarity in (ii). In the following, we describe an intuitive approach to a model incorporating the ideas (i) and (ii). It turns out that this leads to the Max-Min (Max-t-norm) rule.

In order formalize the idea of similar inputs and outputs, we use the notion of a similarity relation. A similarity relation on the set U is a mapping $E : U \times U \longrightarrow [0,1]$ such that

(S1) $E(x,x) = 1$,

(S2) $E(x,x') = E(x',x)$, and

(S3) $E(x,x') \otimes E(x',x'') \leq E(x,x'')$, where $\otimes$ is the Łukasiewicz multiplication defined in equation (2), hold. If the Łukasiewicz multiplication is replaced by an arbitrary t-norm, the more general notion of an equality relation [6] or indistinguishability operator [25] is used. The choice of the Łukasiewicz multiplication in (S3) implies that the concept of a similarity relation is dual to pseudo-metrics. Thus a canonical interpretation of similarity in the sense of 'not being far away from each other' with respect to the corresponding pseudo-metric can be provided [7].

We stressed in the assumption (i) that the specified values do not have to be precise. An intuitive approach to reflect this notion is to describe these imprecise values by fuzzy sets, which are well suited for modelling of linguistic expressions like '*approximately* x_0'. This expression is represented by a fuzzy set $\mu \in \mathcal{F}(U)$ where U is the underlying domain for the variable X. We also assume that there is some $x_0 \in U$ such that $\mu(x_0) = 1$. Usually, U will be a subset of the real line, and in this case it is reasonable to assume that μ is non-decreasing for $x \leq x_0$ and non-increasing for $x \geq x_0$, implying that μ is a convex fuzzy set. We will return to this point later on in this section. At the moment, it is sufficient to understand the fuzzy set μ as a representation of the vaguely given value x_0. The membership degree $\mu(x)$ is interpreted as the degree to which x can be identified with the value x_0.

Let us now consider one simple statement

$$\mathcal{R} = \mathsf{IF}\ X\ \mathsf{is}\ \textit{approximately}\ x_0\ \mathsf{THEN}\ Y\ \mathsf{is}\ \textit{approximately}\ y_0. \tag{4}$$

that assigns the (imprecisely known) value y_0 to the (imprecisely known) value x_0, i.e., it determines the value of the function g in one (imprecisely known) point.

It will follow from the discussion below that (4) is an implication on the surface implication form which, however, is not necessarily understood as a logical implication. Later on we will discuss in more detail different ways to describe function in terms of logical implications and other statements.

The linguistic expressions '*approximately* x_0' and '*approximately* y_0' are represented by the fuzzy sets μ and ν, respectively. Now assume that we are given the input value x. What can we say about the output for x?

If $x = x_0$ then the output is '*approximately* y_0', which is represented by the fuzzy set ν. For $x \neq x_0$ we have to take into account to which degree x can be identified with x_0. This degree is given by the membership grade $\mu(x)$. What can be said about the output in this case?

Since (4) is the only information we have, we derive for a given possible output y that, if x is approximately x_0 then y can be accepted as output if it is approximately y_0. The *truth degree* of the statements '*x is approximately x_0*' and '*y is approximately y_0*' is determined by the fuzzy sets μ and ν, respectively. These two statements are connected by a conjunction. From the view point of fuzzy logic, a natural model of conjunction is the minimum or another suitable t-norm $\sqcap$. Therefore, the degree to which y can be considered as an appropriate output for x is obtained using

$$\mu(x) \sqcap \nu(y). \tag{5}$$

Formula (5) can be interpreted in the following sense.

$$f(x) \approx y \quad \longleftrightarrow \quad x \approx x_0 \wedge y \approx y_0 \tag{6}$$

where $\approx$ stands for (approximate) equality. In the crisp case, where $a \approx b$ is either true or false, (6) simply determines the partial function that is defined for the input x_0, yielding y_0 as output. But since we assumed that x_0 and y_0 are not precisely known, $a \approx b$ cannot be evaluated in terms of true and false alone, but in a graded approach leading to the idea of a fuzzy equality for which a possible axiomatization is given by the axioms (S1)–(S3). We will revisit this question again in the end of this section.

We now extend our considerations from one statement of the form (4) to a set of such statements

$$\mathcal{R}_i = \textsf{IF } X \textsf{ is } \textit{approximately } x_0^{(i)} \textsf{ THEN } Y \textsf{ is } \textit{approximately } y_0^{(i)}, \tag{7}$$

for $i \in I$. In the sense of the beginning of this section, (7) represents a fuzzy function G, i.e., it can be considered as a tabular form of giving G.

Analogously with the interpretation of (4) in the sense of (6), we associate with (7) the formula

$$f(x) \approx y \quad \longleftrightarrow \quad (\exists i \in I) \ (x \approx x_0 \wedge y \approx y_0). \tag{8}$$

Since each statement $\mathcal{R}_i$ determines the output (imprecisely) for one (vague) input value, we have to check for the given input x all statements $\mathcal{R}_i$ whether x fits to '*approximately* $x_0^{(i)}$'. If we can find some $i \in I$, for which x fits to '*approximately* $x_0^{(i)}$', then we know that '*approximately* $y_0^{(i)}$' is a suitable output for x. This justifies the use of the existential quantifier in (8).

It follows from the theory of (fuzzy) logic, that the existential quantifier should be interpreted by the supremum. Hence, given the input x, (8) leads to the output fuzzy set

$$\nu(y) = \bigvee_{i \in I} \left(\mu^{(i)}(x) \sqcap \nu^{(i)}(y) \right). \tag{9}$$

This is exactly the Max-t-norm inference and in the case of $\sqcap = \min$, the Max-Min rule.

The advantage of using the imprecisely known values being represented by fuzzy sets is the following. In the crisp case, (8) does determine only a partial function without giving an additional information how the partial function G can be extended to a function g. If we use the imprecise values, then the whole input space can be covered by a finite number of rules (8) (i.e., the table specifying the function G is finite) and thus, the extension to g is possible.

Concerning the form of (7), recall from [22] that a classical finite function can be equivalently described either by a conjunction of (classical) implications or disjunction of conjunctions. But G is a classical function (of course, with fuzzy values) and thus, the implication form of (7) is justified keeping in mind, however, that they are not treated as implications.

So far we have considered rules of the form (7) which admit only one input variable X. It is straightforward to generalize the proposed concepts to multi-input systems using rules of the form

$$\begin{aligned} \mathcal{R}_i \; = \; & \text{IF } X_1 \text{ is } \textit{approximately } x_1^{(i)} \text{ AND } \ldots \text{ AND } X_n \text{ is } \textit{approximately } x_n^{(i)} \\ & \text{THEN } Y \text{ is } \textit{approximately } y_0^{(i)}. \end{aligned} \tag{10}$$

The conjunction in the premise of (10) can be evaluated by a suitable t-norm, usually the minimum. The system of rules describes again a function in the same way as (8), i.e.,

$$f(x_1, \ldots, x_n) \approx y \;\longleftrightarrow\; (\exists i \in I)\; (x_1 \approx x_1^{(i)} \wedge \ldots \wedge x_n \approx x_n^{(i)} \wedge y \approx y_0^{(i)}). \tag{11}$$

The Max-t-norm (Max-Min) rule was motivated in terms of interpolation on a very intuitive basis. The fundamental assumptions for our considerations were that

- the fuzzy sets representing the linguistic terms appearing in the rules correspond to imprecisely known values,
- the membership grade of the value x to the fuzzy set μ that represents the imprecisely known value x_0 corresponds to the degree to which x can be identified with x_0.

Looking at common fuzzy systems, especially controllers, these two assumptions do not seem to be obvious, since expressions like '*positive big*' and '*negative small*' are usually used instead of the considered linguistic terms '*approximately* x_0'. In the following, we explain that in many typical applications it is really reasonable to reinterpret the system in the sense we have proposed above. As a first step, we will also provide a more rigorous approach to the interpretation of membership grades.

Consider the linguistic expression *approximately* x_0. If we are given a value x and have to decide whether x could be accepted as *approximately* x_0, we need some distance or similarity measure , in order to evaluate, whether x and x_0 can be identified. In the very common case where the domain U for the variable X is a subset of the real line, there exists a canonical distance measure, namely the standard metric on U, given by $\delta(x, x') = |x - x'|$. From this distance measure we obtain a similarity relation E on U defined by $E(x, x') = 1 - \min\{|x - x'|, 1\}$. Given this similarity relation, a value x_0 induces the fuzzy set

$$\mu_{x_0}(x) = E(x, x_0) \tag{12}$$

of values that are similar to x_0. Note that μ_{x_0} is a triangular membership function of the width two. Although such fuzzy sets are very common in fuzzy systems, they are definitely too restricted to cover a sufficient large number of applications. The similarity relation induced by the standard metric is generally not the best choice. The idea to obtain a more appropriate similarity relation is to adjust the standard metric to the problem to be considered. This adjustment is carried out on the basis of a suitable scaling. Two different intentions are inherent in this scaling. First of all, a normalization is one goal of scaling. Depending on the measurement unit the standard metric has to be adjusted by a constant scaling factor $c > 0$, i.e., we replace the metric $\delta(x, x') = |x - x'|$ by $\hat{\delta}(x, x') = |c \cdot x - c \cdot x'|$. For instance, the constant scaling factor transforming the unit hours to minutes is $c = 60$.

The second and more interesting point of our scaling concept is the following. Depending on the considered problem, there might be some ranges of the domain U of the variable X where it is not very important to know the precise value of X. This means that we do not have to distinguish very carefully between values in such ranges. To express this phenomenon in terms of a non-uniform scaling, we would choose a small scaling factor c near zero for these ranges. Thus two values x and x' in one of these ranges might have a great distance with regard to the standard metric, but have nevertheless a high degree of similarity given by $E(x, x') = 1 - \min\{|c \cdot x - c \cdot x'|, 1\}$.

On the other hand, for ranges where the considered system is very sensitive to small changes a greater scaling factor is more appropriate, leading to a low degree of similarity, even for values that are quite close to each other.

This idea of using different scaling factors can be generalized by assigning a scaling factor $c(x) \geq 0$ to each x in the domain U [7]. The value $c(x)$ is a measure for the distinguishability in the neighbourhood of x. In this case the modified distance is given by

$$\hat{\delta}(x, x') = \left| \int_{x}^{x'} c(s)ds \right|,$$

leading to the similarity relation

$$E(x, x') = 1 - \min\left\{ \left| \int_{x}^{x'} c(s)ds \right|, 1 \right\}. \tag{13}$$

From (13) we can derive the following lemma.

Lemma 1 *Let $c : R \longrightarrow [0, \infty)$ be an integrable function and let E be the similarity relation given by (13). Then for any imprecisely known value $x_0 \in R$ the corresponding fuzzy set $\mu_{x_0}(x) = E(x, x_0)$ is convex, piecewise differentiable, continuous, and satisfies $\mu_{x_0}(x_0) = 1$.*

Most applications use the fuzzy sets with the properties described in Lemma 1.

The motivation of fuzzy sets as representations of imprecisely known values in an environment whose indistinguishability is described in terms of scaling factors is intuitively appealing. Nevertheless, we still have to prove that these ideas are coherent with the use of fuzzy sets in applications.

For this, we recall the following theorem proved in [7]:

Theorem 1 *Let $(\mu_i)_{i \in I}$ be an at most countable family of fuzzy sets on R and let $(x_0^{(i)})_{i \in I}$ be a family of real numbers such that $\mu_i(x_0^{(i)}) = 1$ holds. Furthermore, let* *the fuzzy sets μ_i be convex, piecewise differentiable, and continuous. There exists a scaling function $c : R \longrightarrow [0, \infty)$ such that each fuzzy set μ_i for $i \in I$ coincides with the fuzzy set $\mu_{x_0^{(i)}}$ which is associated with the imprecisely known value $x_0^{(i)}$ with regard to the indistinguishability induced by c, if and only if*

$$\min\{\mu_i(x), \mu_j(x)\} > 0 \quad \Rightarrow \quad |\mu_i'(x)| = \left|\mu_j'(x)\right| \tag{14}$$

holds almost everywhere for all $i, j \in I$.

At first sight, condition (14) requiring that the absolute values of the derivatives of two fuzzy sets have to be equal on the intersection of their supports, might look technical. But it is, for example, implied by the very common requirement, that the sum of the membership degrees of neighbouring fuzzy sets of a fuzzy partition should add up to one. This can be guaranteed by taking a fuzzy partition which is obtained by choosing crisp values $x_1 < x_2 < \ldots < x_n$ and defining the fuzzy set μ_i for $1 < i < n$ by a triangular membership function which takes its maximum at x_i and reaches the value zero at x_{i-1} and x_{i+1}, respectively.

Thus we cannot only interpret typical fuzzy partitions in terms of a similarity relation induced by a scaling function; we provide also an explanation for reasonable conditions for fuzzy partitions.

Let us now consider how the problem of interpolation can be solved in the framework of indistinguishability induced by scaling functions. In order to describe the function which we want to interpolate in our framework, we need a set of imprecisely known values $x_0^{(i)} \in U$, $(i \in \{1, \ldots, k\})$, and the corresponding imprecisely known output values $y_0^{(i)} \in V$, where we assume that the domains U and V are real intervals. In addition, scaling functions on U and V, which induce the corresponding similarity relations, have to be specified. The scaling functions have to be chosen in such a way that low scaling factors are assigned to values where the function to be interpolated does not vary significantly. This reflects the idea that it is not necessary to distinguish strictly between input values for which the output is more or less identical. On the other hand, if the function to be interpolated is expected to change rapidly in a certain range, then for this range a greater scaling factor has to be specified, since it is important to distinguish well even between close input values, because their corresponding outputs might differ considerably. In this way, the similarity relations induced by the scaling functions characterize how precise the values should be in different ranges in order to obtain an acceptable interpolation.

To be able to carry out the interpolation, in addition to the scaling functions pairs of (imprecisely known) input-output values $(x_0^{(i)}, y_0^{(i)})$ are needed.

Before we turn to the problem of choosing appropriate points for interpolation, we justify the Max-Min rule in this stricter framework. Assume that we know the scaling functions c and d inducing the similarity relations E and F on the input domain U and the output domain V. In addition, we have the imprecisely known interpolation points $x_0^{(i)}$ with corresponding outputs $y_0^{(i)}$, i.e., a partial fuzzy function G is given. For each pair (x, y) we can determine to which degree it can be considered as belonging to G.

The pair (x, y) belongs to the fuzzy function G if and only if there exists an imprecisely known interpolation tuple $(x_0^{(i)}, y_0^{(i)})$ with which (x, y) can be identified, i.e., we have to look for the interpolation tuple $(x_0^{(i)}, y_0^{(i)})$ which fits best to (x, y). Since x should be identifiable with $x_0^{(i)}$ *and* y should be identifiable with $y_0^{(i)}$ it is clear that the degrees of similarity between x and $x_0^{(i)}$, and y and $y_0^{(i)}$ should be aggregated in a conjunctive manner, i.e., by a t-norm, for example the minimum. This means that we obtain a degree of similarity of $\min\{E(x, x_0^{(i)}), F(y, y_0^{(i)})\}$ between the pairs (x, y) and $(x_0^{(i)}, y_0^{(i)})$. Since we are looking for the best fitting interpolation point for the pair (x, y), we finally obtain

$$\max_i \left\{ \min\{E(x, x_0^{(i)}), F(y, y_0^{(i)})\} \right\} \tag{15}$$

as the degree to which (x, y) belongs to the imprecisely known partial function.

What can be said about the output for a certain input x? For each output y we can determine the degree to which (x, y) belongs to the imprecisely known partial function by equation (15). Thus, a description of the output fuzzy set is given by

$$\mu_{\text{out},x}(y) = \max_i \left\{ \min\{E(x, x_0^{(i)}), F(y, y_0^{(i)})\} \right\}. \tag{16}$$

Remembering that the fuzzy sets of values that can be identified with the value $x_0^{(i)}$ and $y_0^{(i)}$, respectively, are given by $\mu_i(x) = E(x, x_0^{(i)})$ and $\nu_i(y) = F(y, y_0^{(i)})$, equation (16) can be rewritten in the form

$$\mu_{\text{out},x}(y) = \max_i \{ \min\{\mu_i(x), \nu_i(y)\} \} \tag{17}$$

which is again the Max-Min rule.

In the above considerations, we gave lines how appropriate scaling functions can be chosen. The interpolation points were assumed to be given. Now, we propose a philosophy of selecting suitable interpolations points.

The case when a random sample of data is available will not be considered since it then is usually more appropriate to apply a regression technique. We concentrate on the case where a human expert has an idea of how the function should approximately look like. Of course, it might be reasonable to specify as many interpolation points as possible. However, we stick here to the philosophy that the expert tries to define as few interpolation points as are necessary for a satisfactory description of the function. This method frees the expert from specifying redundant knowledge and leads to a very information compressed representation of the function to be interpolated. Let us assume that the output $y_0^{(i)}$ for the imprecisely known input $x_0^{(i)}$ is given. The similarity relation E induced by the scaling function c on U enables us to get information about the output corresponding to the value x, as long as $E(x, x_0^{(i)}) > 0$ holds. Thus the next imprecisely known interpolation points $x_0^{(i-1)}$ and $x_0^{(i+1)}$ should be chosen such that $E(x_0^{(i-1)}, x_0^{(i)}) = 0 = E(x_0^{(i+1)}, x_0^{(i)})$ and $E(x, x_0^{(i)}) > 0$ for all $x_0^{(i-1)} < x < x_0^{(i+1)}$. If we follow this minimality philosophy, we obtain a fuzzy partition from the imprecisely known values $x_0^{(i)}$ that satisfies the condition $\mu_i(x) + \mu_{i+1}(x) = 1$ for all $x_0^{(i)} < x < x_0^{(i+1)}$. Thus we can provide an interpretation for such typical fuzzy partitions in terms of a 'lazy' expert who specifies as few interpolation points as necessary.

We have motivated the Max-Min rule as an interpolation technique based on imprecisely known interpolation points. The similarity relations provide additional information in the neighbourhood of the interpolations points. Therefore, it is possible to define a (fuzzy) output for any input, even if the input-output relation is known only in the form of a partial function. In this sense, tolerating a certain amount of imprecision provides a better framework for interpolation than insisting on exact values, for which other assumptions have to be made in order to define a reasonable interpolation function.

Let us return to idea of a fuzzy function $G : \mathcal{F}(U) \longrightarrow \mathcal{F}(V)$, stated in the beginning of this section and illustrated in figure 1. We emphasized that the ordinary partial function G should only be defined for fuzzy points. Taking into account the above considerations on similarity relations, it is clear that exactly those fuzzy sets can be interpreted as fuzzy points that induced by a single element (and the given similarity relation). In case, the similarity relation is not explicitly defined, Theorem 1 provides the answer to the question whether the fuzzy sets can be seen as fuzzy points.

We have restricted our considerations here to the simple case where the similarity relation is induced by the standard metric and a non-uniform scaling. The notion of a similarity relation is more general [24, 11]. Similar results about interpolation in the more general framework of equality relations can be formulated [8, 9]. However, the interpretation of similarity induced by a non-uniform scaling cannot be maintained.

3 General Problem of Description of a Function or a Relation

3.1 General problem

The previous section provided a method for interpolating or specifying a function on the basis of imprecisely known interpolation points and suitable similarity relations. This section deals with a more general question how a function or even a relation can be described.

We will discuss and examine different approaches to the problem. In order to clarify the concepts behind these approaches, we first restrict our considerations to the crisp case and discuss the fuzzy case later on.

Assume that we have an input domain U and an output domain V. We are looking for a (crisp) relation $R \subseteq U \times V$ containing the pairs $(x, y) \in U \times V$ such that y is a suitable output for the input x.

Two types of information about the unknown relation R can in principle be obtained. One possibility is that for some pairs $(x_i, y_i) \in U \times V$ where $i \in I$ we know that they surely belong to R, i.e.

$$\bigcup_{i \in I} \{(x_i, y_i)\} \;\subseteq\; R. \tag{18}$$

Note that we might even know a whole set of pairs that belong to R for sure. But this set can be split into its element and it is therefore sufficient to assume that we only specify single elements one by one as in equation (18).

The alternative to specification of pairs that surely belong to R is to find some sets R_i containing surely the relation R. Hence, we have

$$R \subseteq \bigcap_{i \in I} R_i. \tag{19}$$

A possible way how to obtain the relations R_i is to find some subsets $S_i \subseteq U \times V$, $i \in I$, which definitely contain no tuples of R. Hence, R is contained in the complement $R_i = (U \times V) - S_i$ of S_i.

The two approaches (18) and (19) could be called *lower* and *upper approximation.* These two approaches are also discussed by Dubois and Prade [4] and Yager and Filev [27] where they are the graph and the functional view or constructive and destructive models, respectively. Our aim is to examine these approaches in the view fuzzy sets induced by similarity relations.

The above considerations are very simple for the crisp case. But when we generalize the description of a relation to fuzzy sets, we have to be aware of the distinction between (18) and (19). Unfortunately, especially in the fuzzy control applications, confusion is caused by misinterpretations (or missing interpretation). The transfer function, or, in our terminology, the input-output relation is given in terms of linguistic rules of the form

$$\mathcal{R} = \; \mathsf{IF}\ X\ \mathsf{is}\ \mathcal{A}\ \mathsf{THEN}\ Y\ \mathsf{is}\ \mathcal{B} \tag{20}$$

where $\mathcal{A}$ and $\mathcal{B}$ stand for linguistic expressions like *positive big* or *negative small* that are represented by fuzzy sets $\mu_{\mathcal{A}}$ and $\nu_{\mathcal{B}}$, respectively. The question of how such rules should be interpreted is often not discussed in the applications. Recalling the above considerations about the description of a crisp input-output relation, there are two possible meanings for such rules, having consequences for the 'fuzzification' of the rule.

We now turn to the fuzzy case where we want to describe a fuzzy relation $\rho \underset{\sim}{\subseteq} U \times V$ instead of a crisp relation $R \subseteq U \times V$. The relation ρ might be interpreted as the representation of a crisp relation $R \subseteq U \times V$ taking some similarity relation H on $U \times V$ into account. In the same way a single element induces a fuzzy set with regard to a similarity relation (compare equation (12)), a crisp subset M of a domain U is associated with the fuzzy set

$$\mu_M(x) \;=\; \bigvee_{m \in M} E(x, m) \tag{21}$$

of elements that can be identified with at least one of the elements of U with regard to the similarity relation E. Thus we would have

$$\rho(x,y) \;=\; \bigvee_{(x_0,y_0)\in R} H((x,y),(x_0,y_0))$$

where $x \in U$, $y \in V$ and $R \subseteq U \times V$ is an unknown relation. Of course, also other interpretations of the fuzzy relation ρ are possible.

3.2 Lower Approximation

In this section, we can will discuss the relation between the rules of the form (20) and the (unknown) fuzzy relation ρ. The first interpretation of such rules is in the spirit of (18). The linguistics expressions $\mathcal{A}$ and $\mathcal{B}$ and their associated fuzzy sets $\mu_{\mathcal{A}}$ and $\nu_{\mathcal{B}}$ represent imprecisely known values x_0 and y_0 with regard to the similarity relations E on U and F on V, respectively. Therefore, the rule (20) states that the crisp tuple (x_0, y_0) belongs to the unknown crisp relation R. Since the x_0 and y_0 are only indirectly determined by the fuzzy sets $\mu_{\mathcal{A}}$ and $\nu_{\mathcal{B}}$, instead of the crisp tuple (x_0, y_0) we can obtain only its corresponding fuzzy set with respect to the similarity relation H on $U \times V$.

It is necessary to make some assumptions about the connection between the similarity relations E and F on U and V, respectively, and the similarity relation H on the product space $U \times V$.

First of all, we require that the similarity relations E and F satisfy a weak independence property, meaning that the similarity degree $H((x,y),(x',y'))$ of the tuples (x,y) and (x',y') depends only on the similarity degrees $E(x,x')$ between x and x' and $F(y,y')$ between y and y', but not on the specific choice of x, x', y, and y'. In other words,

(H1) there is a function $h : [0,1]^2 \longrightarrow [0,1]$ such that

$$H((x,y),(x',y')) \;=\; h(E(x,x'),F(y,y'))$$

holds.

The function h should fulfill at least the following three axioms.

(H2) $h(\alpha,\beta) = h(\beta,\alpha)$

(H3) $h(\alpha,1) = \alpha$

(H4) $\alpha \leq \gamma \;\Rightarrow\; h(\alpha,\beta) \leq h(\gamma,\beta)$

The commutativity of h in axiom (H2) assumes that the similarity relation on the product space is not affected by the order of sequence of the spaces U and V, i.e., E and F have the same influence on H.

(H3) is motivated by the assumption that

$$H((x,y),(x',y)) \;=\; E(x,x')$$

holds, stating that the similarity degree between the tuples (x,y) and (x',y) is equal to the similarity degree of x and x'. Assuming $E(x,x') = \alpha$, we therefore obtain

$$h(\alpha,1) \;=\; h(E(x,x'),F(y,y)) \;=\; H((x,y),(x',y)) \;=\; E(x,x') \;=\; \alpha.$$

The monotonicity condition (H4) is equivalent to the statement that the degree of similarity between (x,y) and (x',y') does not exceed the degree of similarity between (x'',y) and (x''',y') if the degree of similarity between x and x' is less than or equal to the similarity degree between x'' and x'''.

Proposition 1 *Let E, F, and H be similarity relations on U, V, and $U \times V$, respectively.*

(i)

$$H_L((x,y),(x',y')) \;=\; E(x,x') \otimes F(y,y')$$

is a similarity relation satisfying the axioms (H1) – (H4).

(ii)

$$H_U((x,y),(x',y')) \;=\; \min\{E(x,x'),F(y,y')\}$$

is a similarity relation satisfying the axioms (H1) – (H4).

(iii) If H fulfills (H1) – (H4), then

$$H_L \;\le\; H \;\le\; H_U$$

holds.

PROOF: (i) and (ii) are easily proved by deriving the properties (S1) – (S3) for H_L and H_U, respectively, by taking into account that E and F also have these properties.

Assume $H((x,y),(x',y')) \;=\; h(E(x,x'),F(y,y'))$. The left-hand part of (iii) is proved by

$$\begin{aligned} H((x,y),(x',y')) &\overset{\text{(S3)}}{\ge} H((x,y),(x',y)) \otimes H((x',y),(x',y')) \\ &\overset{\text{(H1)}}{=} h(E(x,x'),F(y,y)) \otimes h(E(x',x'),F(y,y')) \\ &\overset{\text{(H3)}}{=} E(x,x') \otimes F(y,y'). \end{aligned}$$

For the right hand part of (iii) we have to show that $h(\alpha,\beta) \leq \min\{\alpha,\beta\}$ holds. (H4) and (H3) imply $h(\alpha,\beta) \leq h(\alpha,1) = \alpha$. Making use of (H2) we obtain also $h(\alpha,\beta) \leq \beta$. □

Proposition 1 has direct consequences for the interpretation of rules of the form (20). The similarity relations E, F, and H are usually not explicitly given. We only assume that the fuzzy sets $\mu_{\mathcal{A}}$ and $\nu_{\mathcal{B}}$ represent the crisp values x_0 and y_0 with regard to E and F. We are interested in the fuzzy relation ρ which is induced by the crisp relation R on $U \times V$ with regard to H. The rule (20) states implicitly that $(x_0, y_0) \in R$ holds. Thus from (12) and (21) we can conclude that the fuzzy set $\rho_{\mathcal{R}} = \mu_{(x_0,y_0)}$ associated with (x_0, y_0) (with respect to H) is contained in ρ, i.e.

$$\rho_{\mathcal{R}} \leq \rho. \tag{22}$$

In order to determine $\rho_{\mathcal{R}}$ we would have to know the similarity relation H. Assuming that H satisfies the axioms (H1) – (H4), we obtain

$$\begin{aligned} \rho_{\mathcal{R}}(x,y) &= H((x,y),(x_0,y_0)) \\ &= h(E(x,x_0), F(y,y_0)) \\ &= h(\mu(x), \nu(y)). \end{aligned}$$

This means that $\rho_{\mathcal{R}}(x,y)$ depends only on $\mu(x)$ and $\nu(y)$. Furthermore, from Proposition 1 we derive

$$\mu(x) \otimes \nu(y) \;\leq\; \rho_{\mathcal{R}} \;\leq\; \min\{\mu(x), \nu(y)\}.$$

Thus, one reasonable solution is to assume that

$$\rho_{\mathcal{R}} \;=\; \min\{\mu(x), \nu(y)\} \tag{23}$$

since the similarity relation H_U in proposition 1, which leads to (23), seems to be a good choice, because it assumes that there is no interaction between the similarity relations E and F.

Until now we have considered only a single rule of the form (20). If a set of such $\mathcal{R}_i$, $i \in I$, is given, by (22) we obtain $\rho_{\mathcal{R}_i} \leq \rho$ for all $i \in I$ and therefore

$$\bigvee_{i \in I} \rho_{\mathcal{R}_i} \;\leq\; \rho.$$

Assuming that I is finite, (23) leads to

$$\max_{i \in I} \{\min\{\mu_i(x), \nu_i(y)\}\} \;\leq\; \rho \tag{24}$$

where μ_i and ν_i are the fuzzy sets associated with the linguistic expressions $\mathcal{A}_i$ and $\mathcal{B}_i$ occurring in the rule $\mathcal{R}_i$. One can see that (24) is again the Max-Min rule.

3.3 Upper Approximation

The second interpretation of rules of the form (20) is in the spirit of equation (19). Let us return, for a moment, to the crisp case and assume that the fuzzy sets $\mu_{\mathcal{A}}$ and $\mu_{\mathcal{B}}$ correspond to the crisp sets $M_{\mathcal{A}}$ and $N_{\mathcal{B}}$, respectively. Then the rule might be interpreted as follows. If the input belongs to M, then the output definitely has to be chosen from N; but N might still contain some elements which are not suitable. Therefore, the rule (20) states that the pairs in the set $S_{\mathcal{R}} = M \times \overline{N}$, where $\overline{N}$ denotes the complement of N, do not belong to the relation R. Equivalently, we can say that R is contained in the set

$$R_{\mathcal{R}} \;=\; (M \times N) \cup (\overline{M} \times V) \;=\; \{(x,y) \in U \times V \mid x \in M \Rightarrow y \in N\}. \tag{25}$$

Thus, the linguistic IF-THEN rule (20) is understood as a logical implication. Note, moreover, that this is the most general form of expressing that there is a dependence between some phenomena.

Returning to the fuzzy case, where the linguistic expressions are represented by the fuzzy sets $\mu_{\mathcal{A}}$ and $\nu_{\mathcal{B}}$, respectively, we come to the world of fuzzy logic and approximate reasoning discussed in the next section. Note that in this case, the rules are interpreted on the basis of fuzzy implications yielding the fuzzy relation

$$\rho_{\mathcal{R}}(x,y) = \mu_{\mathcal{A}}(x) \to \nu_{\mathcal{B}}(y) = \min\{1 - \mu_{\mathcal{A}}(x) + \nu_{\mathcal{B}}(y), 1\} \tag{26}$$

which contains the unknown fuzzy relation ρ, i.e., $\rho \le \rho_{\mathcal{R}}$.

If a set of rules $\mathcal{R}_i$, $i \in I$, is given, then $\rho \le \rho_{\mathcal{R}_i}$ holds for each $i \in I$ so that we obtain

$$\rho \;\le\; \bigwedge_{i \in I} \rho_{\mathcal{R}_i}. \tag{27}$$

This formula differs very much from the Max-Min rule. We will elaborate this case more in details in the next section when dealing with the logical approach.

3.4 Defuzzification

The above discussed approaches lead to the output which is a fuzzy set. For real applications it is necessary to determine one unique output value from the output fuzzy set. This step is called **defuzzification**. A lot of defuzzification strategies are available. Most of them are heuristic techniques leading, nevertheless, to satisfactory results.

A problem connected with the heuristic defuzzification methods is that the assumptions behind the respective heuristics are often not made explicit.

In most fuzzy control applications the idea is to determine the transfer function which gives a single output to each input. Thus the rules are indeed supposed to describe a (fuzzy) function. The Max-Min rule, most often applied in fuzzy control, models, as shown above, a lower approximation. Therefore, a lower approximation of a (fuzzy) function can be seen as the basic concept behind many fuzzy controllers.

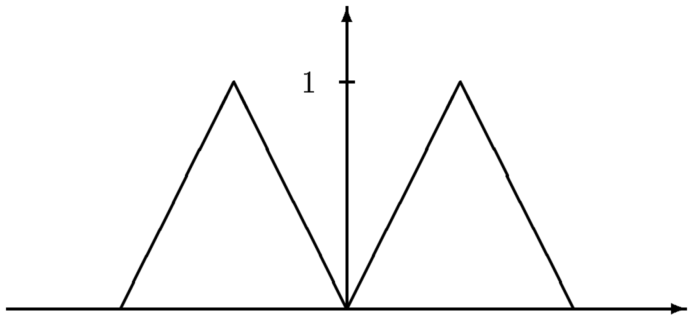

Figure 2: A fuzzy set causing difficulties for defuzzification.

Thus, a very common hidden assumption behind defuzzification strategies is that the output fuzzy set represents at most one (fuzzy) output value. But then one should be aware of the fact that it does not make sense to have an output fuzzy set as it is shown in figure 2 which obviously stands for two (fuzzy) values. It is not the responsibility of the defuzzification strategy to solve this problem by determining a single output from this fuzzy set. In the strict interpretation of the lower approximation of a fuzzy function, such an output indicates a contradiction in the rule base.

Thus, either the rule base has to be changed, or the idea of a lower approximation of a fuzzy function has to be given up, either by assuming a lower approximation of a fuzzy relation or by considering an upper approximation. However, for an upper approximation the Max-Min rule seems not to be appropriate.

If the aim of the rule base is the description of a fuzzy relation instead of a fuzzy function, it should be taken into account for the defuzzification strategy. Then defuzzification has two fulfill two tasks simultaneously:

- turning a fuzzy set into a crisp set,
- choosing one out of many (fuzzy) values.

If a lower approximation of a function is the intention, then the second task is superfluous. The basic defuzzification strategies are Center of Gravity method (COG), Mean of Maxima (MOM) and Fuzzy Mean. All of them demonstrate that they are indeed based on the assumption that the fuzzy set to be defuzzified represents only one (fuzzy) value. Thus, the problem is, how to find a strategy that would best fit the function g discussed in the beginning of Section 2.

Let us also remark, that the hidden assumption that the rules are intended to describe a fuzzy function raised also confusion of the interpretation of the Max-Min rule. Although the rules are stated as (a conjunction of) implications, they are treated as a disjunction of conjunctions according to the use of the maximum and the minimum. Recall once again that in classical logic, the description of a function (not of an arbitrary relation) using a conjunction of implications is equivalent with the description using a disjunction of conjunctions.

A different situation is faced with the upper approximation leading to a logical description — see the next section. A solution should be based on the assumption that we are dealing with logical reasoning and, possibly, also that the output fuzzy sets to be defuzzified represent linguistic expressions. Such an approach has been presented, e.g., in [22]. A more formal possible approach has been outlined in [20]. As we will see in the next section, the approximate reasoning can be explained as formal reasoning in a fuzzy theory ${}^{AR}T$ given by fuzzy sets of special axioms that are derived from the linguistic form of the rules $\mathcal{R}_i$, $i \in I$. Then, the defuzzification may be understood as finding a conclusion in ${}^{AR}T$ extended by additional fuzzy set K of special axioms. Their formulation should be based on the properties of the fuzzy theory ${}^{AR}T$ and kinds of formulas in concern.

4 Logical Inference, Chaining of Rules

In the previous sections, we have discussed the way how a function can be described and we demonstrated that the Max-t-norm procedure is a reasonable way of interpolation when searching a functional value of an imprecisely specified function. In this section, we will deal with a more general situation when making a logical inference. Recall that this concerns the upper approximation of the fuzzy relation discussed in Section 3.3.

4.1 Logical inference and inference in fuzzy logic

A logical inference, in general, is a procedure how to obtain new facts from some other previously given facts. Such procedures are inherent to human mind and are studied in logic since Aristoteles. Various formal methods have been developed, especially in this century. Two things are common to all of them: they introduce the concept of truth and falsity and the inference rule of modus ponens. What is, however, mostly neglected, is the vagueness of facts that are dealt with. Truth or falsity are always full, i.e., nothing between them is accepted. But vagueness requires intermediate truth values.

A "fact", which may be a property of objects φ, leads to a question whether an object x (taken from some universe of discourse) has, or has not, φ. However, if φ is vague then $\varphi(x)$ cannot be determined exactly. We thus naturally come to degrees of truth and, finally, to fuzzy logic[†)].

Recall that logical inference in logic is formally a sequence of formulas $B_1, \dots, B_m$,

[†)]One should realize, however, that this graded (fuzzy) approach (cf. [15]) is a possible but not the only one mathematical model of vagueness. Though we identify fuzziness with vagueness in this paper, it is only a working assumption. We by no means think that these two notions are indeed equivalent.

each of which is a logical or special axiom, or it is derived from some previous formulas using a logical inference rule.

Logical inference rules can schematically be written as

$$\frac{A_1, \ldots, A_n}{B}$$

where $A_1, \ldots, A_n$ are known facts and B is a new, derived fact. In modern logical systems, $A_1, \ldots, A_n, B \in F_J$ are formulas taken from a set F_J of well formed formulas in some formal language J. A typical and basic inference rule inherent to human mind is the rule of *modus ponens*

$$\frac{A, A \Rightarrow B}{B}$$

where $A \Rightarrow B$ is an implication between two facts. It is a formal, most general expression of the dependence of the fact B on the fact A. It may be given by our experience, knowledge, or stem from other source and, of course, it does not necessarily mean a causal relation. In presence of vagueness, the situation is complicated by the truth degrees.

Let $A(x) \in F_J$ be a formula which represents the vague fact φ. Obviously, $A(x)$ itself is not sufficient to characterize also the vagueness of φ. To solve this problem, we introduce the concept of evaluated formula.

A couple

$$[A;\ a]$$

where $A \in F_J$ and $a \in [0,1]$ is a (syntactic) truth degree is called the *evaluated formula*. Let M_V be the set of all terms without variables of the language J. Then the vagueness of φ is formally characterized by a set of evaluated formulas

$$\underline{A} = \{[A_x[t];\ a_t] \mid t \in M_V\}$$

where $A_x[t]$ is a formula obtained from A by replacing all free occurrences of x by the term t. Thus, we naturally get from vagueness of logical facts to many-valued logic which provides us with means for manipulation with evaluated formulas. Let us remark, that the set of evaluated formulas $\{[A_i;\ a_i] \mid i \in I\}$ can be understood also as a fuzzy set of formulas

$$\{\ {}^{a_i}/A_i \mid i \in I\}.$$

A many-valued inference rule is a scheme

$$\frac{[A_1;\ a_1], \ldots, [A_n;\ a_n]}{[B;\ b]} \tag{28}$$

where $B = r^{syn}(A_1, \ldots, A_n)$ is a formula syntactically derived from $A_1, \ldots, A_n$ and $b = r^{sem}(a_1, \ldots, a_n)$ is its resulting evaluation. The functions r^{syn}, r^{sem} must fulfil reasonable conditions (cf. [13, 14, 23]). For example, a *many-valued rule of modus ponens* has the form

$$\frac{[A;\ a], [A \Rightarrow B;\ b]}{[B;\ a \otimes b]}. \tag{29}$$

The properties of inference rules have been extensively discussed in [14, 23].

Since in approximate resoning, we work with vague facts, an inference rule has the following form

$$R : \frac{\underline{A_1}, \ldots, \underline{A_n}}{\underline{B}} \tag{30}$$

where $\underline{A_i} = \{[A_{x_i}[t_i];\ a_{t_i}] \mid t_i \in M_V\}$ are sets of evaluated formulas. The $\underline{B}$ is a resulting set of evaluated formulas

$$\underline{B} = \{[B_y[s];\ \bigvee_{t_1,\ldots,t_n \in M_V} (r^{sem}(a_{t_1}, \ldots, a_{t_n})] \mid s \in M_V\}$$

which is derived on the basis of an underlying many-valued inference rule (28).

For example, the rule of modus ponens has the form

$$R_{MP} : \frac{\underline{A}, \underline{A \Rightarrow B}}{\underline{B}} = \frac{\{[A_x[t];\ a_t] \mid t \in M_V\}, \{[A_x[t] \Rightarrow B_y[s];\ c_{ts}] \mid t, s \in M_V\}}{\{[B_y[s];\ \bigvee_{t \in M_V}(a_t \otimes c_{ts})] \mid s \in M_V\}}. \tag{31}$$

Let us stress that the rule of modus ponens is not the only one inference rule in fuzzy logic and approximate reasoning. Furthermore, unlike classical logic, there may be various modifications of modus ponens. A very important rule having practical applications in decision making and fuzzy control is the following: Let $\rhd(\cdot)$ be a special unary connective for the linguistic hedges with narrowing effect, for example *very, highly, extremely* and $\lhd(\cdot)$ that with widening effect, for example *more or less, roughly, very roughly*, etc. The effect of both connectives concerns truth values. Without going into details, we may introduce their basic properties in the form of additional schemes of logical axioms:

LHn $\vdash \rhd(A) \Rightarrow A$

LHw $\vdash A \Rightarrow \lhd(A)$

for every $A \in F_J$. It can be immediately seen that the $\rhd$ connective decreases truth values and $\lhd$ increases them. More specific properties should be stated explicitly for every concrete modifier.

Now we may introduce the rule of *modus ponens with hedges*

$$r_{MPH} : \frac{[\rhd A;\ a], [A \Rightarrow B;\ b]}{[\lhd B;\ a \otimes b]}. \tag{32}$$

Various other kinds of inference rules can be introduced in the logical system of approximate reasoning. A list of some sound inference rules is given below:

(a) *Modus ponens with conjunction of implications*

$$r_{CMP}: \frac{[A_k;\ a], [\bigwedge_{j=1}^m (A_j \Rightarrow B_j);\ b]}{[B_k;\ a \otimes b]} \qquad 1 \leq k \leq m.$$

(b) *Modus ponens with conjunction of implications and hedges*

$$r_{MPCH}: \frac{[\rhd A_k;\ a], [\bigwedge_i (A_i \Rightarrow B_i);\ b]}{[\lhd B_k;\ a \otimes b]} \qquad 1 \leq k \leq m.$$

(c) The following rule is hidden behind the proposal of L. A. Zadeh to do inference in the approximate reasoning:

$$r_C: \frac{[A;\ a], [A \wedge B;\ b]}{[B;\ a \wedge b]}.$$

(d) *Modus tollens*

$$r_{MT}: \frac{[\neg B;\ b], [A \Rightarrow B;\ c]}{[\neg A;\ b \otimes c]}.$$

Recall from [14, 13] that a *fuzzy theory* T is a triple

$$T = \langle A_L, A_S, R \rangle$$

where $A_L \underset{\sim}{\subset} F_J$ is a fuzzy set of logical axioms, $A_S \underset{\sim}{\subset} F_J$ is a fuzzy set of special axioms. Equivalently, A_L and A_S are sets of evaluated logical and special axioms, respectively. The R a set of sound inference rules.

Now, let us turn to the approximate reasoning. The basic situation is defined by specifying a linguistic description

$$\mathcal{R} = \{\mathcal{R}_1, \mathcal{R}_2, \ldots, \mathcal{R}_m\} \tag{33}$$

where each $\mathcal{R}_i$, $i = 1, \ldots, m$ usually has the form

$$\mathcal{R}_i = \mathsf{IF}\ X\ \mathsf{is}\ \mathcal{A}_i\ \mathsf{THEN}\ Y\ \mathsf{is}\ \mathcal{B}_i \tag{34}$$

and $\mathcal{A}_i, \mathcal{B}_i$, are certain linguistic expressions. We will often use the term *syntagm* which is a well formed linguistic expression. In our case, we mostly mean syntagms of the form

$$\langle \text{adverb} \rangle \langle \text{adjective} \rangle.$$

A linguistic expression (syntagm) is in general a name of some fact (property) and thus, via the above approach, each of $\mathcal{R}_i$, $i = 1, \ldots, m$, is interpreted by a set of evaluated formulas

$$\underline{R}_i = \{[A_{i,x}[t] \Rightarrow B_{i,y}[s];\ c_{ts}] \mid t, s \in M_V\}, \qquad i = 1, \ldots, m. \tag{35}$$

The linguistic description (33) represents the expert knowledge and from our point of view, it can be considered as a set of *linguistically expressed special axioms* being the basis of approximate reasoning at the given moment. In other words, (33) is a *theory* of approximate reasoning. This theory is adjoined a certain fuzzy theory ${}^{AR}T$ of fuzzy logic in narrow sense given by a fuzzy set of special axioms

$$ {}^{AR}A_S = \underset{\sim}{R}_1 \cup \cdots \cup \underset{\sim}{R}_m \tag{36} $$

Then the basic task of approximate reasoning which is finding a conclusion from (33) given a premise

$$ x \text{ is } \mathcal{A}' $$

translates to a set of deductions based on some of the inference rules listed above (for example, the rule r_{CMP}).

To end this section, we present a theorem demonstrating that fuzzy logical inference can in principle give similar results as a fuzzy interpolation. Several papers contain a proof that fuzzy controllers are universal approximators (cf. [1, 2, 10, 3]). We prove analogous theorem (inspired by [3]) about the logical inference. Of course, as all the cited theorems, it is only an existential result.

From now, we will work in a certain model of approximate reasoning whose support is the set of real numbers $\Re$. Let $U \subseteq \Re$ be a measurable subset of the real line and $a, b \in U$. Put

$$ F_U(a,b) = \{\mu : U \longrightarrow [0,1] \mid \mu(x) > 0 \quad \text{iff} \quad x \in (a,b)\}. $$

Let $\mu \underset{\sim}{\subseteq} U$ and put

$$ D(\mu) = \frac{\int_{U_\mu} \mu(x) \cdot x \, dx}{\int_{U_\mu} \mu(x) \, dx} \tag{37} $$

where

$$ U_\mu = \{x \in U \mid \mu(x) > \bigwedge_{y \in U} \mu(y)\}. \tag{38} $$

If $U_\mu = \emptyset$ then (37) is not defined. The function D is a defuzzification function on μ. The formula (38) enables to filter uninteresting values obtained due to the use of fuzzy implication (Łukasiewicz one).

Lemma 2 *Let $\mu \in F_U(a,b)$, $\nu \in F_V(c,d)$ for some $a < b \in U \subseteq \Re$, $c < d \in V \subseteq \Re$ where U, V are measurable and put $\nu_x = \{\, \mu(x) \rightarrow \nu(y)/y \mid y \in V\}$ for $x \in U$. Then*

$$ x \in (a,b) \quad \textit{implies} \quad D(\nu_x) \in (c,d) $$

and $D(\nu_x)$ is not defined for $x \notin (a,b)$.

PROOF: Let $x \notin (a,b)$. Then $\mu(x) = 0$ and so $\nu_x(y) = 1$ for every $y \in V$, i.e., $U_\mu = \emptyset$. Let $x \in U$, $y \in (c,d)$ and $y' \notin (c,d)$. Then $\nu(y) > 0$ and $\nu(y') = 0$, i.e., $\nu_x(y') < \nu_x(y)$ and we obtain $U_{\nu_x} = (c,d)$ which gives the lemma. □

To simplify the notation and explanation, we omit unnecessary formal definitions of terms and formulas in this restricted language (we work in a real line) and define logical inference as a procedure of finding a fuzzy set ν_x for the given x (taken as a unit fuzzy singleton) from the set of implications

$$\mu_i(x) \rightarrow \nu_i(y), \qquad i = 1, \ldots, m, \quad x \in U, y \in V \tag{39}$$

using the formula

$$\nu_x(y) = \bigwedge_{j \in K_x} (\mu_j(x) \rightarrow \nu_j(y)) \tag{40}$$

where $K_x = \{j \mid \mu_j(x) = \bigvee_{i=1}^m \mu_i(x)\}$. This definition is based on the inference rule r_{CMP} in which the input element x (i.e., a term t_0) represents a formula $A_{k,x}[t_0]$ for some $1 \leq k \leq m$.

Theorem 2 *Let $U \subseteq \Re$ be compact. Then to every bounded continuous function $f : U \longrightarrow \Re$ and $\varepsilon > 0$ there is a set of implications (39) such that*

$$|f(x) - D(\nu_x)| < \varepsilon$$

for every $x \in U$ where ν_x is a fuzzy set (40) and D is a function (37).

PROOF: Let $a \in U$. Denote $O_a = \{x \in U \mid |x - a| < \delta_a\}$ and $O_f(a) = \{y \in R \mid |y - f(a)| < \varepsilon\}$ where δ_a depends on ε. The continuity of f can be written as

$$x \in O_a \quad \text{implies} \quad f(x) \in O_{f(a)}$$

for every $a, x \in U$. Then $\bigcup_{a \in U} O_a$ is a covering of U and since U is compact, there is a finite subcovering $\{O_{a^i}\}_{1 \leq i \leq m}$. Let the rules be given by

$$\mu_i(x) \rightarrow \nu_i(y)$$

where $\mu_i \in F_U(O_{a^i})$ and $\nu_i \in F_R(O_{f(a^i)})$, $i = 1, \ldots, m$.

Let $x \in O_{a^j}$ for some (and therefore all) $j \in K_x$. We have to prove that then $D(\nu_x) \in O_{f(a^j)}$.

Let $y \in R$ and $\nu_j(y) = 0$ for some $j \in K_x$. Then $\nu_x(y) = \neg\mu_j(x)$ for every $j \in K_x$ (all $\mu_j(x)$ are equal). Otherwise $\nu_j(y) > 0$, i.e., $\nu_x(y) > \neg\mu_j(x)$ (for every $j \in K_x$). Hence, $\bigwedge_{y \in R} \nu_x(y) = \neg\mu_j(x)$ and

$$U_{\nu_x} = \{y \in R \mid \nu_x(y) > \bigwedge_{y \in R} \nu_x(y)\} = \bigcap_{j \in K_x} O_{f(a^j)}$$

by (40). Hence, $D(\nu_x) \in O_{f(a^j)}$ by Lemma 2. □

4.2 Chaining of rules

In the control of complex systems, in expert systems, in practical everyday human reasoning and in many other occasions, we may hardly manage with one inference only. A realististic approximate reasoning deals with many vague statements, many conditions and thus, chaining of rules is necessary. From the logical point of view, it is the realization of a proof. Formally, the proof w of a formula A_n is a sequence

$$w := A_1, \ldots, A_n$$

where each A_i is an axiom (logical or special), or is derived from some previous formulas in the proof using an inference rule. In fuzzy logic, we deal with evaluated formulas and, hence, the proof is also evaluated. An *evaluated proof* (or shortly, a proof) of a formula A from a fuzzy set A_S of formulas is a sequence of evaluated formulas

$$w := [A_1;\ a_1], \ldots, [A_n;\ a_n]$$

such that A_n is A and each evaluated formula consists of A_i and $a_i = A_L(A_i)$ if A_i is a logical axiom, A_i and $a_i = A_S(A_i)$ if A_i is a special axiom, or

$$[A_i;\ a_i] := [r^{syn}(A_{i_1}, \ldots, A_{i_n});\ r^{sem}(a_{i_1}, \ldots, a_{i_n})], \quad i_1, \ldots, i_n < i$$

where r is an n-ary sound rule of inference. The a_n is the *value* of the proof w. We usually write

$$a_n = \mathrm{Val}(w).$$

In the approximate reasoning, the concept of proof is analogous but, as we deal with linguistic syntagms interpreted by sets of evaluated formulas, a *proof in approximate reasoning* is, in fact, a sequence

$$w := \mathcal{A}_1[\underline{A_1}], \ldots, \mathcal{A}_n[\underline{A_n}]. \tag{41}$$

The linguistic terms $\mathcal{A}_i$, $i = 1, \ldots, n$ are not necessary during the proof and thus, we will simplify (41) to a sequence of sets of evaluated formulas

$$w := \underline{A}_1, \ldots, \underline{A}_n. \tag{42}$$

Each set of evaluated formulas

$$\underline{A}_i = \{[A_{i,x}[t];\ a_{it}] \mid t \in M_V\}, \qquad i = 1, \ldots, n$$

in (42) consists either of

(a) $[A_{i,x}[t];\ a_{it}]$ where $A_{i,x}[t]$ is an instance of a logical or a special axiom $A(x)$, and $a_{it} = A_L(A_{i,x}[t])$ or $a_{it} = A_S(A_{i,x}[t])$ respectively for all $t \in M_V$.

(b) $[A_{i,x}[t];\ a_{it}]$ where $A_{i,x}[t] = r^{syn}(A_{j_1,x_1}[t_1], \ldots, A_{j_n,x_n}[t_n])$, $1 \leq j_1, \ldots, j_n < i \leq n$ is a result of an n-ary inference rule applied on formulas from $\underline{A}_{j_1}, \ldots, \underline{A}_{j_n}$ preceding $\underline{A}_i$, and

$$a_{it} = \bigvee\{r^{sem}(a_{j_1 t_1}, \ldots, a_{j_n t_n}) \mid t_1, \ldots, t_n \in M_V\}.$$

Let us illustrate this formula on the simple proof in approximate reasoning consisting of two implications:

$$\underline{A}, \underline{A \Rightarrow B}, \underline{B}, \underline{B \Rightarrow C}, \underline{C}. \tag{43}$$

After rewriting, we obtain

$$\begin{aligned}\{[A_x[t];\ a_t] \mid t \in M_V\}, \{[A_x[t] \Rightarrow B_y[s];\ b_{ts}] \mid t, s \in M_V\},\\ \{[B_y[s];\ b_s = \bigvee_{t \in M_V} (a_t \otimes b_{ts}) \mid s \in M_V\}],\\ \{[B_y[s] \Rightarrow C_z[u];\ c_{su}] \mid s, u \in M_V\},\\ \{[C_z[u];\ \bigvee_{s \in M_V} (b_s \otimes c_{su})] \mid u \in M_V\},\end{aligned} \tag{44}$$

Note that $\underline{C}$ can be obtained also by means of a proof

$$\underline{A}, \underline{A \Rightarrow B}, \underline{B \Rightarrow C}, \underline{(A \Rightarrow B) \Rightarrow ((B \Rightarrow C) \Rightarrow (A \Rightarrow C))}, \underline{(B \Rightarrow C) \Rightarrow (A \Rightarrow C)}, \underline{C}. \tag{45}$$

It can be seen that proving in approximate reasoning is much more complicated than in many-valued logic since we deal with sets of evaluated formulas. Hence, fuzzy expert systems dealing with vague statements and making vague conclusions based on sets of linguistic statements are more complicated than the standard ones where only one weight interpreted as truth or uncertainty degree[†)] is considered.

Realize that the inference considered in most applications of approximate reasoning till now (e.g., in the fuzzy control) is a simple proof

$$\underline{A}, \underline{A \Rightarrow B}, \underline{B}.$$

As we have stated in other papers [14, 20], one formula may be proved by many kinds of proofs which, in general, have different values. The supremum of all of them is the *provability degree* in the given fuzzy theory (in our case, it is ${}^{AR}T$). Given a fuzzy theory T, we formally write

$$T \vdash_a A$$

where a is a provability degree of A. Hence, finding a proof of A gives only a lower estimation of the provability degree. It is our goal to reach the provability degree since then we know that nothing better can be obtained. Formally, in approximate reasoning we aim at obtaining a set

$$\{[B_y[s];\ b_s] \mid {}^{AR}T \vdash_{b_s} B_y[s], s \in M_V\}.$$

In [20, 21] we have demonstrated that if we confine ourselves to the linguistic syntagms that are commonly used in approximate reasoning then the formula of

[†)]This depends on the nature of the statements. Realistic expert systems should deal both with vagueness as well as with uncertainty. This task, however, seems to be far from the satisfatory solution.

approximate reasoning for IF–THEN statements considered as logical implications gives us the provability degrees. We give analogous theorem for the case of chaining of rules.

We say that two formulas A and B are *independent* if no variant of one is a subformula of the other.

Let F_0 be a set of evaluated formulas, which are mutually independent. We say that F_0 is *directed*, if:

(a) If $[(\forall x)A;\ a] \in F_0$ and $[A_x[t];\ b] \in F_0$, then $a \leq b$, where $t \in M_J$.

(b) If A is a logical axiom then $[A;\ a] \in F_0$ implies $a = A_L(A)$.

Lemma 3 *Let F_0 be a directed set of independent evaluated formulas and let $T = \{\,{}^{a}/A \mid [A;\ a] \in F_0\}$. Then there is a model $\mathcal{D} \models T$ such that*

$$\mathcal{D}(A) = a$$

holds for all $[A;\ a] \in F_0$.

Lemma 4 *Let a theory*

$$T = \{[A_1;\ a_1], [A_1 \Rightarrow A_2;\ a_2], \ldots, [A_{n-1} \Rightarrow A_n;\ a_n]$$

be given where $\{[A_1;\ a_1], \ldots, [A_n;\ a_n]\}$ is a directed set of independent evaluated formulas. Then

$$T \vdash_c A_n$$

where $c = a_1 \otimes \cdots \otimes a_n$.

PROOF: By assumption, the formulas $A_1, \ldots, A_n$ are independent and so, there is a truth valuation $\mathcal{D}$ such that $\mathcal{D}(A_1) = a_1$ and $\mathcal{D}(A_j) = a_1 \otimes \cdots \otimes a_j$, $j = 2, \ldots, n$. Then

$$a_i \leq \mathcal{D}(A_{i-1} \Rightarrow A_i) = a_1 \otimes \cdots \otimes a_{i-1} \rightarrow a_1 \otimes \cdots \otimes a_i,$$

i.e., $\mathcal{D} \models T$[†)]. On the other hand, there is a proof w of A_n with the value $\mathrm{Val}(w) = a_1 \otimes \cdots \otimes a_n$. The lemma then follows from the completeness theorem. □

Lemma 4 is the basis of the theorem below. Let $\mathcal{S}_i$, $i = 1, \ldots, n$ be disjoint sets of syntagms of the form

$$[\langle\text{linguistic modifier}\rangle]\langle\text{adjective}\rangle\langle\text{noun}\rangle,$$

where $\mathcal{A}_i \in \mathcal{S}_i$ be assigned the set $\underline{A_i} = \{[A_{xi}[t];\ a_{it}] \mid t \in M_V\}$. Note that $A_i(x)$, $i = 1, \ldots, n$ are independent formulas and so are also all their instances.

†) This symbol means that $\mathcal{D}$ is a *model* of the theory T. For the precise definition see [14].

Theorem 3 *Let a theory of approximate reasoning be given by a linguistic description*

$$\mathcal{T} = \{\mathcal{A}_1, \mathsf{IF}\ \mathcal{A}_1\ \mathsf{THEN}\ \mathcal{A}_2, \ldots, \mathsf{IF}\ \mathcal{A}_{n-1}\ \mathsf{THEN}\ \mathcal{A}_n\}$$

where $\mathcal{A}_i \in \mathcal{S}_i$, $i = 1, \ldots, n$ are the above defined syntagms. Then the theory $\mathcal{T}$ is assigned a fuzzy theory ${}^{AR}T$ given by the fuzzy set of special axioms

$$\begin{aligned} {}^{AR}A_S &= \{ {}^{a_{t_1}}/A_{1,x}[t_1] \mid t_1 \in M_V\} \cup \\ & \bigcup_{i=2}^{n} \{ {}^{a_{t_{i-1}t_i}}/(A_{i-1} \Rightarrow A_i)_{x_{i-1},x_i}[t_{i-1}, t_i] \mid t_i \in M_V, i = 1, \ldots, n\} \end{aligned}$$

Then the conclusion $\mathcal{A} = \mathcal{A}_n$ has the interpretation

$$\underline{A} = \{ {}^{a_t}/A_x[t] \mid {}^{AR}T \vdash_{a_t} A_x[t], t \in M_V\}$$

where

$$a_t = \bigvee_{t_1,\ldots,t_{n-1} \in M_V} (a_{t_1} \otimes a_{t_1 t_2} \otimes a_{t_{n-1} t}).$$

This theorem can be generalized also to the case when sets of linguistic implications are considered at each i-th step.

When chaining the rules, the computational complexity significantly increases. If we deal with modus ponens then our situation may be simplified by the properties of the Łukasiewicz product $\otimes$ since it is nilpotent (it pushes small values to zero). Another possibility is to defuzzify at each step. This means that each $\underline{A_i}$ occuring in an inference rule in the proof is replaced by a simple evaluated formula $[A_{ix}[t_{i0}]; a_{t_{i0}}]$ for some term t_{i0} before further inference steps. Thus, instead of $\mathrm{Card}(M_V)^2$ inferences at each step we have to make only $\mathrm{Card}(M_V)$ inferences.

4.3 Chaining of interpolations

Due to Sections 2 and 3, the fuzzy interpolation of an unknown function $g : X \longrightarrow Y$ is to find a point $y_0 \approx g(x)$ via the fuzzy function $G : A \longrightarrow B$ where $A \subseteq \mathcal{F}(X)$ and $B \subseteq \mathcal{F}(Y)$. This is given by

$$y_0 = D(\bigcup_{\mu \in A} (\{ {}^{G(\mu)(y) \wedge \mu(x)}/y \mid y \in Y\})$$

where $G(\mu) \in B \subseteq \mathcal{F}(Y)$ and D is a defuzzification function.

Let now a sequence of functions $g_1 : X_0 \longrightarrow X_1$, …, $g_n : X_{n-1} \longrightarrow X_n$ be given. Our task is to approximate an unknown value

$$y = h(x_0) = g_n(\cdots g_1(x_0) \cdots). \tag{46}$$

Each function g_i is approximated by a fuzzy function $G_i : A_{i-1} \longrightarrow A_i$ where

$$A_i \subseteq \mathcal{F}(X_i), \qquad i = 1, \ldots, n.$$

Let us denote

$$Q_i(x_{i-1}) = \bigcup\{\{ {}^{G_i(\mu_{i-1})(x_i) \wedge \mu_{i-1}(x_{i-1})}/x_i \mid x_i \in X_i\} \mid \mu_{i-1} \in A_{i-1}\} \tag{47}$$

where $x_{i-1} \in X_{i-1}$. In analogy with Sections 2 and 3 it is natural to put

$$y_0 = D(\bigcup\{Q_n(x_{n-1}) \mid x_i \in \text{Supp}(Q_i(x_{i-1})), i = 1, \ldots, n-1\}). \tag{48}$$

For example, given $g_1 : X_0 \longrightarrow X_1$, $g_2 : X_1 \longrightarrow X_2$, $G_1 : A_0 \longrightarrow A_1$, $G_2 : A_1 \longrightarrow A_2$, and $A_i \subseteq \mathcal{F}(X_i)$, $i = 0, 1, 2$. Then

$$y_0 = D(\bigcup\{Q_2(x_1) \mid x_1 \in \text{Supp}(Q_1(x_0)\}). \tag{49}$$

In (49), x_1 are taken from the support of a fuzzy set

$$Q_1(x_0) = \bigcup\{\{ {}^{G_1(\mu_0)(x_1) \wedge \mu_0(x_0)}/x_1 \mid x_1 \in X_1\} \mid \mu_0 \in A_0\}$$

and all of them determine the resulting fuzzy set to be defuzzified.

Proposition 2 *Let A_i, $i = 0, \ldots, n$ be a fuzzy partition of X_i and $\mu_i \cap \mu_i' \neq \emptyset$ for every $\mu_i, \mu_i' \in A_i$. Furthermore, let every G_i be an injection. Then the number of different fuzzy sets occuring in Q_n is at least $n+1$.*

PROOF: Let there be just one $\mu_0 \in A_0$ such that $\mu_0(x_0) > 0$. Then $\text{Supp}(Q_0) = \text{Supp}(\mu_1)$ for some $\mu_1 = G_1(\mu_0) \in A_1$ and there is at least one $\mu_1' \neq \mu_1$ such that $\mu_1'(x_1) > 0$ for some $x_1 \in \text{Supp}(\mu_1)$ due to the assumption. Repeating the same argument and taking the assumption that G_i is an injection we obtain the proposition. □

Corollary 1 *Let* $\text{Card}(A_n) \leq n+1$. *Then* $y_0 = D(\bigcup A_n)$.

The assumption $\mu_i \cap \mu_i' \neq \emptyset$ in Proposition 2 is natural to have covering of the space X_i. If some G_i are not injections then this proposition in general is not true and in specific cases, Q_n may consist of significantly smaller number of fuzzy sets. However, the danger of rapid increase of $\text{Supp}(Q_n)$ may drastically disqualify chaining of interpolations since y_0 in (48) might be derived from a very wide fuzzy set.

It seems more reasonable to defuzzify each fuzzy set $Q_i(x_{i-1})$ before following step. Then $x_i \approx g_i(\cdots g_1(x_0)\cdots)$ is a defuzzified value obtained in the previous step. Validity of the resulting y_0, however, decreases with the increase of n. We may somewhat improve it when increasing cardinality of the sets A_i, i.e., to make the fuzzy partitions more dense. But then the design and all the computations are more complicated.

Note that the situation in the chaining of interpolations is different in comparison with the logical inference. For example, if we compare (43) with (49) then the computational complexity is the same (without optimization) but the validity of the result in (43) is much higher since $\underline{C}$ is deduced using a sequence of precisely defined inferences — evaluated proofs. In (49), a fuzzy set of elements with a high degree of fuzziness is obtained carrying the information that the correct (searched) $h(x_0)$ is somewhere in its support and we have to find y_0 as close to it as possible. Since we propagate a mistake and increase vagueness, chaining of interpolations seems to be somewhat dubious.

4.4 Interpolation and logical inference for fuzzy input

In fuzzy control and various other applications, the input x_0 is crisp, interpreted as a fuzzy singleton $\{\,^1/x_0\}$ or, in logical inference, as a unit evaluated formula $[A_x[t_0]; 1]$. However, many other applications require one or more inputs to be fuzzy as well. In this case, the difference between both approaches becomes more apparent.

Let the linguistic description (33) and (34) be given. Then the inference in fuzzy logic is given by the formula

$$\frac{\{[A_{k,x_k}[t];\ a_t] \mid t \in M_V\}, \{[\bigwedge_{j=1}^m (A_{j,x}[t] \Rightarrow B_{j,y}[s]);\ c_{ts}] \mid t, s \in M_V\}}{\{[B_{y,k}[s];\ \bigvee_{t \in M_V}(a_t \otimes c_{ts})] \mid s \in M_V\}}. \tag{50}$$

Fuzzy interpolation leads to a formula

$$\nu' = \{\,\bigvee_{x\in U}(\mu'(x) \wedge \bigvee_{\mu\in A}(G(\mu)(y) \wedge \mu(x)))/y \mid y \in V\}. \tag{51}$$

To compare (50) and (51) we rewrite (50) as follows:

$$\nu' = \{\,\bigvee_{x\in U}(\mu'(x) \otimes \bigwedge_{j=1}^m(\mu_j(x) \to \nu_j(y)))/y \mid y \in V\}. \tag{52}$$

Let $\mu' = \mu_0$ for some $\mu_0 \in A$. If (51) were a logical inference, we would naturally expect that $\nu' = G(\mu_0)$.

Proposition 3 *Let $\mu_1, \mu_2 \in A$ be fuzzy sets and the height of $\mu_1 \cap \mu_2 = c$. Let $\mu' = \mu_1$ and $\mu_1 \cap \mu_2 = \emptyset$ for $j = 1, 2$.*

(a) Using (51) we obtain

$$\nu' = G(\mu_1) \cup G(\mu_2)_c \not\subseteq \nu_1 \tag{53}$$

where $G(\mu_2)_c = \{\,\mu_2(y) \wedge c/y \mid y \in V\}$.

(b) Using (52) we obtain

$$\nu' \subseteq \nu_1. \tag{54}$$

PROOF: (a) is obtained immediately after rewriting.
(b)

$$\bigvee_{x\in U}(\mu_1(x) \otimes \bigwedge_{j=1}^{2}(\mu_j(x) \to \nu_j(y))) = \nu_1(y) \wedge \bigvee_{x\in U}(\mu_1(x) \otimes (\mu_2(x) \to \nu_2(y))) \leq \nu_1(y)$$

holds for every $y \in V$. □

On the basis of this proposition we can derive the following. Logical inference is safe, i.e., it is not possible to get a fuzzy set ν' which would contain elements with non-zero membership degree not belonging at the same time to the fuzzy set ν_1. If μ_1 is normal and there is an x such that $\mu_1(x) = 1$ and $\mu_2(x) = 0$ then $\nu' = \nu_1$. The proper inclusion may in general occur due to the rule r_{CMP}. However, since more proofs can be found, this result can be improved. For example, if we use the rule r_{MP} with the implication

IF X is $\mathcal{A}_1$ THEN Y is $\mathcal{B}_1$

then $\nu' = \nu_1$. This is natural deduction when a set of special axioms (35) is considered.

In the case of fuzzy interpolation (51), some "additional" elements always belong to ν'. Note that this is correct when we do not consider (51) as the logical inference. The procedure simply takes into account some environmental elements that might also be function values of the (unknown) interpolated function g due to the width of the input fuzzy set. In the logical rule of modus ponens, however, this is not acceptable since it cannot give elements outside ν_1 (i.e., elements x with membership degree greater than $\nu_1(x)$).

In the end, let us remark that if the input fuzzy sets $\mu \in A$ are mutually disjoint then the fuzzy interpolation turnes to be a sound inference rule based on r_C (cf. [19]). However, this case is not very interesting.

5 Conclusions

In this paper, we have discussed the relation between fuzzy logical inference and fuzzy interpolation. This topic has already been elaborated from various sides. Our aim was to synthesize the results and state explicitly that there are two different kinds of IF–THEN rules in fuzzy logic each of which having its justification and place. However, they should not be interchanged since in general, they give different conclusions (though sometimes with similar effect). This concerns especially the fuzzy interpolation represented by Max-t-norm rule which is widely used. By misunderstanding, erroneous terms such as "Mamdani's implication", "Larsen's implication", etc. appear in the literature. We hope, that now it is clear to the reader that these are not implications, but they are justified from a different point of view and can be used for the special but important task of interpolation of a function.

Let us stress that neither of the two methods is supreme. In principle, logical inference is more general but also more complicated. It is more suitable for decision situations but can be successfully used also for approximation of a function (interpolation). It is safer but, hence, requires in general more rules since it does not interpolate from so wide fuzzy sets. Let us remark, however, that logical inference has already been successfully used in the practice for control of plants. The results are satisfactory and the work with this approach appeared to be quite effective and transparent.

Acknowledgements. We would like to express our thanks for the constructive and helpful comments of an anonymous referee.

References

[1] Buckley, J. J., *Universal fuzzy controllers*, Automatica **28**(1992), 1245–1248.

[2] Buckley, J. J. and Y. Hayashi, *Fuzzy input-output controllers are universal approximators*, Fuzzy Sets and Systems **58**(1993), 273–278.

[3] Castro, J., *Fuzzy Logic Controllers are Universal Approximators*, IEEE Trans. on Systems, Man, and Cybernetics **25**(1995), 629–635.

[4] Dubois, D., Prade, H., *Basic Issues on Fuzzy Rules and their Application to Fuzzy Control*, Proc. IJCAI–91 Workshop on Fuzzy Control, Sydney (1991), 5–17.

[5] Gottwald, S., **Fuzzy Sets and Fuzzy Logic**, Vieweg, Wiesbaden, 1993.

[6] Höhle, U., *M–Valued Sets and Sheaves over Integral Commutative CL-Monoids*, in Rodabaugh, S. E, Klement, E. P., Höhle, U., eds., **Applications of Category to Fuzzy Subsets**, Kluwer, Dordrecht, 1992, 33–72.

[7] Klawonn, F., *Fuzzy sets and vague environments*, Fuzzy Sets and Systems **66**(1994), 207–221.

[8] Klawonn, F., Kruse, R., *Equality Relations as a Basis for Fuzzy Control*, Fuzzy Sets and Systems **54**(1993), 147–156.

[9] Klawonn, F., Kruse, R., *Fuzzy Control as Interpolation on the Basis of Equality Relations*, Proc. IEEE Intern. Conf. on Fuzzy Systems, San Francisco, 1993, 1125–1130.

[10] Kosko, B., *Fuzzy Systems as Universal Approximators*, Proc. IEEE Conf. on Fuzzy Systems, San Diego 1992, 1153–1162.

[11] Kruse, R., Gebhardt, J., Klawonn, F., **Foundations of Fuzzy Systems**, Wiley, Chichester, 1994.

[12] Mamdani, E. H., Gaines, B. R., eds., **Fuzzy Reasoning and its Applications**, Academic Press, London 1981.

[13] Novák, V., **Fuzzy Sets and Their Applications**, Adam–Hilger, Bristol, 1989.

[14] Novák, V., *On the Syntactico-Semantical Completeness of First–Order Fuzzy Logic. Part I — Syntactical Aspects; Part II — Main Results.* Kybernetika **26**(1990), 47–66; 134–154.

[15] Novák, V., **The Alternative Mathematical Model of Linguistic Semantics and Pragmatics**, Plenum, New York, 1992.

[16] Novák, V., *On the logical basis of approximate reasoning*, in V. Novák, J. Ramík, M. Mareš, M. Černý and J. Nekola, eds., **Fuzzy Approach to Reasoning and Decision Making**. Kluwer, Dordrecht 1992.

[17] Novák, V., *Fuzzy logic as a basis of approximate reasoning*, In: Zadeh, L. A. and J. Kacprzyk, eds., **Fuzzy Logic for the Management of Uncertainty**, J. Wiley, New York 1992.

[18] Novák, V.: **LFLC-edu 1.3 — Linguistic Fuzzy Logic Controller for education**. User and programmer's guide. Academy of Sciences of Czech Republic, Institute of Geonics, Ostrava 1993.

[19] Novák, V.: *Logical Analysis of Max-Min Rule of Inference.* EUFIT'93, Aachen 1993.

[20] Novák, V.: *Fuzzy Control from the Point of View of Fuzzy Logic.* Fuzzy Sets and Systems **66**(1994), 159–173.

[21] Novák, V.: *Paradigm, Formal Properties and Limits of Fuzzy Logic.* Int. J. of General Systems (to appear).

[22] Novák, V.: *Linguistically Oriented Fuzzy Logic Control and Its Design.* Int. J. of Approximate Reasoning (to appear).

[23] Pavelka, J., *On fuzzy logic I, II, III*, Zeit. Math. Logic. Grundl. Math. **25**(1979), 45–52, 119–134, 447–464.

[24] Ruspini, E. H., *On the Semantics of Fuzzy Logic.* Int. J. Approximate Reasoning **5**(1991), 45–88.

[25] Trillas, E., Valverde, L., *On the Implication of Indistinguishability in the Setting of Fuzzy Logic*, in: Kacprzyk, J., Yager, R. R., eds., **Management Decision Support Systems Using Fuzzy Sets and Possibility Theory**, Verlag TÜV Rheinland, Köln, 1985, 198–212.

[26] Sugeno, M., ed., **Industrial Applications of Fuzzy Control**, North-Holland, Amsterdam 1985.

[27] Yager, R. R., Filev, P., **Essentials of Fuzzy Modeling and Control**, Wiley, New York 1994.

[28] Yager, R. R., L. A. Zadeh, eds., **An Introduction to Fuzzy Logic Applications in Intelligent Systems**, Kluwer, Dordrecht 1992.

[29] Zadeh, L. A., *Quantitative Fuzzy Semantics*, Inf. Sci.,**3**(1973), 159-176.

[30] Zadeh, L. A., *The concept of a linguistic variable and its application to approximate reasoning I, II, III*, Inf. Sci. **8**(1975), 199–257, 301–357;**9**(1975), 43–80.

[31] Zadeh, L. A., *PRUF — a Meaning Representation Language for Natural Languages*, Int. J. Man–Mach. Stud. **10**(1978), 395–460.

[32] Zadeh, L. A., *A computational approach to fuzzy quantifiers in natural languages*, Comp. Math. with Applic. **9**(1983), 149–184.