Information Complexity and Fuzzy Control

Arthur Ramer
School of Computer Science, University of New South Wales
P.O. Box 1, Kensington, N.S.W. 2033, Australia
ramer@cse.unsw.edu.au

Vladik Kreinovich
Computer Science Department, University of Texas at El Paso
El Paso, TX 79968, USA
vladik@cs.ep.utexas.edu

1 Preamble

Our article analyzes complexity of extracting information from fuzzy data. Such process can be properly viewed as contributing information, hence also alleviating uncertainty about the result of the process itself.

Notion of information has been utilized frequently, both formally and informally. In the formal vein, Shannon entropy of probabilistic distributions is foremost, followed by host of less perfect variants. In the theory of possibility [3], based on fuzzy sets, there is U-uncertainty and several related functions [7, 16]. Both these concepts can be introduced axiomatically, using general rules of combining *information* [7]. Such rules deal with combining independent events, restricting ranges of possible outcomes, and other operations on distributions. Specialized to either probability or possibility theories, they become axioms, which can be shown sufficient to characterize the respective information measures uniquely [16].

Independent of those methods, a question of *complexity* of computing numerical answers to analytical queries has been studied [18, 19]. Here the query is associated with a function about which we have only a partial information. A typical example is finding a root or the integral of a function, for which available are only the values at a discrete set of points. A particularly important example, which also serve as the point of departure in [18], is the *binary search*.

The very same problem arose in analysis of complexity of securing information from fuzzy membership functions. This problem was introduced in [10, 11], where the resulting information gain was termed *entropy*. However, it is clearly most closely related to the information-based complexity [18, 19], and we shall use a neutral term *information* throughout this article. This usage obviates the need of axiomatic characterization, which is not germaine to this concept. At the same time there are certain ties to Shannon entropy, which we discuss in the closing section [14].

Other ties, perhaps more significant, are to the possibilistic information on continuous domains. This relationship can be made formal as we point out in the closing section. In particular, the main information expression is virtually identical both in our information complexity study, and in possibilistic uncertainty.

Here we begin by studying the complexity of element identification through binary search, based on progressively more complex fuzzy domains. (Obviously, for crisp domains our results parallel those in [18].) Using the expedient of function *rearrangement* we can apply it to fuzzy sets based on arbitrary measurable functions. The second part deals with relationship of complexity to the choice of fuzzy operators. Such operators correspond to binary connectives of fuzzy logic; their choice plays significant role in performance of fuzzy control processes. We show how to select the operators corresponding both to the maximum and to the minimum values of information. In terms of fuzzy control they represent either the most stable, or the smoothest control strategies.

Numerical quantification of *information* can be always viewed equally as quantification of *uncertainty*. Any analysis, experiment, or observation which *gains* information, conveys *removing* uncertainty. In all the frameworks discussed analytically, rather than epistemically, these notions can be used interchangeably. We shall do so here, mostly for stylistic reasons.

In our notation we follow the possibilistic interpretation of fuzzy sets and membership grades. In particular we use $\pi(x)$ to denote the membership on domain X, thus making $\pi(x)$ the *possibility* of $x \in X$. As an added convenience we can use $\mu(x)$ to mean the Lebesgue measure, which we assume defined on all 'continuous' domains of discourse.

2 Introduction

Traditional control theory is applicable in the situations when the analytical behavior of the system to be controlled, is known. However, in many situations, e.g., in space exploration, we do not have this knowledge, but we still have to make control decisions. In such situations, a reasonable idea is to find a human operator who is good in this kind of control, and translate his control experience into a precise formula. This control experience is usually formulated in terms of the natural-language rules like *if x is small, then control must also be small.* The methodology of translating these rules into an actual control strategy, due to its relationship to fuzzy sets and logic, became known under the name of *fuzzy control* [1, 13, 12].

Three main steps are necessary to specify this translation. Firstly, we must determine membership functions that correspond to natural language terms (like *small* or *big*) that appear in the rules. Secondly, we must choose operations that correspond to & and $\vee$. As a result we get a membership function $\pi_C(u)$ for a control; then we need a method to transform this function $\pi_C(u)$ into a single control value (*defuzzification method.*)

Different choices may lead to dramatically different control strategies, making such choice very important [10, 11]. We begin by considering the problem of interpreting $\&-$ and $\vee-$operations. As making a choice restricts the set of possible control strategies, and wrong choice may lead to a poor control strategy, it is reasonable to preserve as many possibilities as feasible. In other words, we should choose $\&-$ and $\vee-$operations in such a way that the uncertainty corresponding to $\pi_C(u)$ is the greatest possible. This methodology is well known in probability theory under the name of *maximum entropy formalism*, and has been applied widely to a variety of problems, ranging from pattern recognition to processing uncertainties in expert systems [5, 8, 2, 9].

Just like in the probabilistic case, we want to evaluate the uncertainty of a membership function as the average number of binary questions that one needs to ask in order to determine the value. In the present paper, we propose the formulas that compute this uncertainty and determine operations for which this uncertainty is maximized. We prove that the desired maximum uncertainty is attained when we use $min(a+b,1)$ for $\vee$, and $min(a,b)$ for &.

In control theory terms, maximum entropy leads to the maximally stable strategy. This result is an intuitive one—we minimized the lost opportunities, and hopefully ended up with the best possible control.

The above arguments are reasonable only if we are ready to apply various defuzzification techniques to extract the best control from $\pi_C(u)$. However, in industrial applications, a defuzzification rule is usually fixed. Since this rule is not necessarily the most appropriate [20, 11], it is reasonable to try to depend on it to the least extent. It suggests choosing & and $\vee-$ operations based on condition that the uncertainty related to $\pi_C(u)$ is the *least* possible. We find that the operations for which the resulting uncertainty is the least are $max(a,b)$ for $\vee$ and ab for &. In control terms, minimum entropy leads to the maximally smooth strategy. This result is also quite intuitive—since we are extremely cautious, we end up with a very smooth control.

3 Uncertainty of membership functions

3.1 Motivation

Let us recall where the values $\pi(x)$ of a membership function come from. If $\pi(x)$ corresponds to, say, *small*, then $\pi(x)$ is our degree of belief that x is small. One of the most natural ways to evaluate numerically this degree of belief is to poll several experts, asking whether they consider x small or not, and after M out of N answer "yes", take M/N as $\pi(x)$ [7]. This approach permits interpreting the value $\pi(x)$ as either frequency or subjective probability that x is small.

With this interpretation in mind, let us estimate uncertainty $U(\pi)$ that corresponds to function $\pi(x)$, and suppose that a certain notion like *small* is being described by the function $\pi(x)$. If the only thing we know about some real value x is that it satisfies this property, then how many binary questions do we have to ask to determine x? As it will be become apparent, we do not just want to compute the simple complexity of answering such query, but also to recognize the maximum values that related possibilities may assume.

The analysis is best organized as a sequence of progressively more general cases. We begin with finitely many alternatives, when it is easy to estimate uncertainty, and proceed towards a general membership function. The discussion places stress on the background and motivation, leaving some mathematical details to the last part of the article.

3.2 Information of crisp assignments

We want to evaluate uncertainty according to the number of binary questions that we have to ask to get the complete knowledge. This number is easy to dfine and to understand in the discrete case, when we have finitely many alternatives $x_1, \ldots, x_n$. In this case finitely many questions suffice to determine the actual alternative and our uncertainty can be estimated as the smallest number of binary *yes-no* questions that we have to ask to determine x_i. As x_i come from a numerical domain—they represent possibility values—they can be considered as ordered, and binary search employed. For n alternatives we need $Q = \lceil \log_2(n) \rceil$ questions, thus proving

PROPOSITION 1 *The uncertainty value of n alternatives is* $\log_2(n)$.

Now, let us start describing uncertainty for membership functions, starting with the simplest one, when the fuzzy set is crisp. First, let us recall basic definitions.

By a *membership function* we mean a function $\pi : R \to [0, 1]$ that is not identically 0. If for every x, $\pi(x) \in \{0, 1\}$, the function is called *crisp*. Such functions are *characteristic* functions of sets $S \subset R$, permitting us to term such standard sets as *crisp*.

We are facing the problem of estimating uncertainty when membership function $\pi(x)$ is the characteristic function of an interval $[a, b]$, expressing the property that $x \in [a, b]$. Now the set of alternatives coincides with the set of all points on an interval and is infinite. If we ask finitely many questions, then we will have only finitely many possible answers. Therefore, to find the precise value of x, we need to ask infinitely many questions.

The natural solution of this problem is related to the fact that in real life, we never know precise values of physical quantities. We can measure them, but all the measurements have finite precision. The only information that we can get from a measurement procedure is a value $\tilde{x}$ such that the actual value x satisfies the inequality $|x - \tilde{x}| \leq \varepsilon$, where $\varepsilon > 0$ is the precision of this measurement. In other words, the actual value x belongs to the interval

$[\tilde{x}-\varepsilon, \tilde{x}+\varepsilon]$. After we know such an $\tilde{x}$, we say that we know x with precision ε.

Therefore, to estimate of the information that $x \in [a, b]$, we fix some ε, and estimate the number of binary questions that have to be asked in order to find x with precision ε.

DEFINITION 1 *Let S be a set of real numbers, and $\varepsilon > 0$. $\varepsilon-$alternatives are a set $\{x_1, \ldots, x_n\}$ of real numbers such that*

$$S \subset \bigcup_{i=1}^{n} [x_i - \varepsilon, x_i + \varepsilon].$$

They are called $\varepsilon-$alternatives because if we know which of the values x_i is $\varepsilon-$close to the actual x, then we know x with precision ε. They are considered *alternatives*, because according to the above definition, for each possible $x \in S$ there exists an x_i, for which $|x - x_i| \leq \varepsilon$. To express the *uncertainty* of this set of alternatives we take $\log_2(n)$.

For a given set S we can have several different sets of $\varepsilon-$alternatives. Different sets of $\varepsilon-$alternatives may have different number of elements, and thus lead to different values of uncertainty. It is natural to consider the smallest of these values as the measure of uncertainty of S. We introduce a convenient notation.

DEFINITION 2 *Let S be a set of real numbers, and $\varepsilon > 0$. By $\varepsilon-$uncertainty $Q(\pi_S, \varepsilon)$ of S, we mean the uncertainty of the smallest possible set of ε-alternatives.*

This notation may look somewhat clumsy: why not simply $Q(S, \varepsilon)$? The reason is that later on, we will define $Q(\pi, \varepsilon)$ for membership functions π that do not necessarily describe crisp sets.

A straightforward argument establishes the next result.

PROPOSITION 2 $Q(\pi_{[a,b]}, \varepsilon) = \log_2(\lceil (b-a)/(2\varepsilon) \rceil)$.

To prove it we observe that the smallest family of 2ε-length intervals, whose union covers $[a, b]$ has $\lceil (b-a)/(2\varepsilon) \rceil$ elements. Taking its logarithm gives the proposition.

By explicitly mentioning a precision ε when we described uncertainty, we ended up with a finite value of uncertainty. The drawback is that instead of one value that described uncertainty in a finite case, we now have a function that correspond to different $\varepsilon > 0$. It would be desirable to isolate the components contributed by the interval $[a, b]$ itself, and by the precision factor ε.

As we are interested primarily in behavior of $Q(\pi, \varepsilon)$ for small $\varepsilon > 0$ the following definition is in order.

DEFINITION 3 *Two functions $f(\varepsilon)$ and $g(\varepsilon)$ are asymptotically equivalent, denoted $f(\varepsilon) \sim g(\varepsilon)$ if $\lim_{\varepsilon \to 0} f(\varepsilon) = \infty$, and $\lim_{\varepsilon \to 0}(f(\varepsilon) - g(\varepsilon)) = 0$.*

PROPOSITION 3 *For $\varepsilon \to 0$*

$$Q(\pi_{[a,b]}, \varepsilon) \sim \log_2(b - a) - \log_2(2\varepsilon).$$

¿From it follows that when we consider intervals, the behavior of the function $Q(\pi_{[a,b]}, \varepsilon)$ for small $\varepsilon > 0$ is uniquely determined by a single number $\log_2(b - a)$. We will see that the same is true for more complicated sets. However, for fuzzy (non-crisp) membership functions π, we will need two numbers to describe the behavior of $Q(\pi, \varepsilon)$: the coefficients at 1 and $\log_2(2\varepsilon)$. Therefore, we consider these coefficients as the measure of uncertainty that corresponds to a membership function π.

DEFINITION 4 *Let π be a membership function. We say that for π, uncertainty is dfined if $Q(\pi, \varepsilon) \sim u - m \log_2(2\varepsilon)$ for some real numbers u and m. The pair (u, m) is called the uncertainty of a fuzzy set π. The value of u will be denoted by $U(\pi)$, and the value of m by $m(\pi)$.*

The following is immediate.

PROPOSITION 4 *For the interval $[a, b]$, $U(\pi_{[a,b]} = \log_2(b - a)$ and $m(\pi_{[a,b]}) = 1$.*

To avoid possible misunderstanding, let us note that the previously defined uncertainties stated the number (or an average number) of binary questions needed to to determine the correct alternative (uniquely or with the precision ε.) As the number of questions is always non-negative, all previously defined measures of uncertainty were themselves non-negative.

However, it can be easily seen, that the expression $U(\pi_{[a,b]}) = \log_2(b-a)$ becomes negative if $b - a < 1$. There is not contradiction in there, because $U(\pi_{[a,b]})$ is not equal to the (non-negative) number of questions. To find the number of questions $Q(\pi_{[a,b]}, \varepsilon)$, one has to select ε, and combine $U(\pi_{[a,b]})$ with the term $-\log_2(2\varepsilon)$. For a small ε, this second term is positive and tends to $+\infty$ as $\varepsilon \to 0$. Therefore, for sufficiently small ε the resulting sum is positive.

There is an analogy between this situation and the entropy of continuous distributions. For discrete variables, that take only finitely many values x_i with probabilities p_i, the entropy is defined as $-\sum_i p_i \log_2(p_i)$. This sum is always non-negative. For continuous variable with density $\rho(x)$ the usual analog of entropy is the integral $S = -\int \rho(x) \log_2(\rho(x))\, dx$. Unlike discrete entropy, this integral can be negative: for example, if we take a uniform distribution on an interval $[a, b]$, i.e., a function $f(x) = 1/(b-a)$ for $x \in [a, b]$ and 0 outside this interval, then this entropy is equal to $\log_2(b - a)$, and this value is negative for $b - a < 1$. The reason is the same as for our non-statistical dfinition: Shannon's entropy $-\sum_i p_i \log_2(p_i)$ can be interpreted as an average number of questions, and therefore, is always non-negative. But its continuous analog is only indirectly related to the number of questions: namely, $S - \log_2(2\varepsilon)$ is the average number of questions that we have to ask to determine x with precision ε.

3.3 Information of fuzzy sets

We start with a crisp set, whose nontrivial part $\{x : \pi(x) > 0\}$ is a collection of intervals.

PROPOSITION 5 *For an arbitrary set S which is a union of disjoint intervals, uncertainty is defined for π_S, $U(\pi_S) = \log_2(\mu(S))$, and $m(\pi_S) = 1$.*

We use $\mu(S)$ to denote Lebesgue measure, here simply the sum of lengths of the intervals. A much more important is the case of $\pi(x)$ piecewise constant. First, a couple of definitios.

DEFINITION 5 *Membership function $\pi(x)$ is normalized if $\sup_x \pi(x) = 1$.*

DEFINITION 6 *Membership function is piecewise constant if there exist values $x_1 < x_2 < \cdots < x_n$ such that $\pi(x) = 0$ for $x < x_1$ and $x > x_n$, $\pi(x) = const$*

on each of the intervals (x_i, x_{i+1})*, and for each* i*,* $\pi(x_i)$ *coincides either with the value of* $\pi(x)$ *for* $x < x_i$*, or with the values of* $\pi(x)$ *for* $x > x_i$*.*

The estimate of the number of binary questions is complicated by the need to recognize the relative weight of each possibility value. The function being piecewise constant, it takes only finitely many different values. Let us order them $h_0 = 0 < h_1 < h_2 < ... < h_k$ and put $\sup \pi(x) = h_k$. In this section, we consider only normalized functions, thus $h_k = 1$.

Let us first consider an example. Suppose that a membership function $\pi(x)$ is equal to 1 for $x \in [-a, a]$, is equal to 0.6 for $x \in [-2a, -a]$ and $x \in [a, 2a]$, and $\pi(x) = 0$ for $x \notin [-2a, 2a]$. In this case, 60% of the experts believe that the area of possible values of x is the interval $[-a, a]$ of length $2a$, and the remaining 40% believe that x is in the interval $[-2a, 2a]$ (of length $4a$). If the experts from this 60% majority are right, then we need $\sim \log_2(2a) - \log_2 \varepsilon$ binary questions. If the minority experts are right, then we need $\sim \log_2(4a) - \log_2 \varepsilon$ questions. Since all the experts are considered equally good, it is reasonable to assume that in general, in 60% of the cases the majority is right, and in 40% of cases, the minority is right. Therefore, the average number of binary questions that we have to ask in order to locate x in an interval of length ε, is $\sim 0.6(\log_2(2a) - \log_2 \varepsilon) + 0.4(\log_2(4a) - \log_2 \varepsilon) = (0.6 \log_2(2a) + 0.4 \log_2(4a)) - \log_2 \varepsilon$.

In the general case, the h_1-th part of all the experts believe that x belongs to the set $\{x : \pi(x) \geq h_1\}$, $(h_2 - h_1)$ of them believe that $x \in \{x : \pi(x) \geq h_2\}$, $(h_3 - h_2)$ of them believe that $x \in \{x : \pi(x) \geq h_3\}$, ..., and $h_k - h_{k-1}$ of them believe that $x \in \{x : \pi(x) \geq h_k\}$. If $x \in \{x : \pi(x) \geq h_1\}$, then we need $Q(\{x : \pi(x) \geq h_1\}, \varepsilon)$ questions to determine x with the precision ε. If $x \in \{x : \pi(x) \geq h_2\}$, then we need $Q(\{x : \pi(x) \geq h_2\}, \varepsilon)$ questions, etc.

Therefore, according to the opinion of h_1 of experts, we need $Q(\{x : \pi(x) \geq h_1\}, \varepsilon)$ questions. Next, according to the opinion of $(h_2 - h_1)$ of the experts, we need to ask $Q(\{x : \pi(x) \geq h_2\}, \varepsilon)$ questions, and similarly for the remaining data.

It becomes natural to define the expected number of questions as

$$Q(\pi, \varepsilon) = \sum_{i=0}^{k-1} (h_{i+1} - h_i) Q(\pi_{\{x:\pi(x) \geq h_{i+1}\}}, \varepsilon).$$

The case of non-normalized membership function can be handled along the same lines. Suppose, for example, that some term from the natural language is represented by a (fuzzy) membership function $\pi(x)$ which is equal to 0.6 for all x from $[a, b]$, and to 0 for all other x. In view of the above interpretation of $\pi(x)$, it means that only 60% of all the experts had any opinion about what the initial natural language term means, and the others simply gave no answers. In this case, we can take into consideration only the opinions of those who made an actual statement.

As we have already argued, it is reasonable to assume that in 60% of cases the majority is right. Therefore, the only thing that we know about the average number of questions is that it is $\geq 0.6Q(\pi_{[a,b]}, \varepsilon) \sim 0.6(\log_2(b - a) - \log_2(2\varepsilon))$. When the remaining 40% of the experts make their decisions, there may be more questions, but right now, when we are given the above-described membership function $\pi(x)$, this number $0.6Q(\pi_{[a,b]}, \varepsilon)$ is the only estimate that we can get. Again, it is reasonable to take it as the description of uncertainty of this membership function $\pi(x)$.

It justifies the following definition:

DEFINITION 7 *Let $\pi(x)$ be a piecewise-constant membership function, that takes only the values $h_0 = 0 < h_1 < \cdots < h_k$. Its $\varepsilon-$uncertainty $Q(\pi, \varepsilon)$ is*

$$Q(\pi, \varepsilon) = \sum_{i=0}^{k-1} (h_{i+1} - h_i) Q(\pi_{\{x:\pi(x) \geq h_{i+1}\}}, \varepsilon).$$

PROPOSITION 6 *For an arbitrary piecewise constant normalized membership function $\pi(x)$, uncertainty $(U(\pi), m(\pi))$ is dfined, $m(\pi) = 1$, and*

$$U(\pi) = \sum_{i=0}^{k-1} (h_{i+1} - h_i) \log_2(\mu\{x : \pi(x) \geq h_{i+1}\}).$$

It brings us to the general case of an arbitrary membership function $\pi(x)$ that can appear in control problems. In our earlier work we postulated some regularity conditions to assure existence and useful asymptotic properties of uncertainty expressions. All such restrictions can be removed through a judicial application of *function rearrangements.* It is a technique originated in the theory of inequalities [4], esp. in applied mathematics, and used recently for continuous fuzzy uncertainty [15].

Here, however, we aim at application to typical fuzzy control problems. And membership functions that are usually considered in fuzzy control [12] are all rather regular. Namely, their domain can be divided into finitely many intervals of monotonicity motivating the following definition.

DEFINITION 8 *Membership function* $\pi(x)$ *is regular if* $\pi(x) \to 0$ *as* $|x| \to \infty$, *and if there exist values* $x_1 < x_2 < ... < x_n$ *such that on each of the intervals* $(-\infty, x_1)$, (x_i, x_{i+1}), $1 \le i \le n-1$, *and* (x_n, ∞), *the function* $\pi(x)$ *is either monotonic increasing or monotonic decreasing, and for each* i, *the value* $\pi(x_i)$ *coincide either with the left-side limit* $\lim_{\varepsilon>0,\varepsilon\to 0} \pi(x-\varepsilon)$, *or with the right-side limit* $\lim_{\varepsilon>0,\varepsilon\to 0} \pi(x+\varepsilon)$.

Although not needed in the future, we remark that $\pi(x) \to 0$ for $x \to -\infty$ and $\pi(x) \ge 0$ for all x. Therefore, a regular function $\pi(x)$ can only be increasing on $(-\infty, x_1)$ nad only decreasing on $[x_n, \infty)$.

When we define a membership function in mathematical terms, we say that its values $\pi(x)$ are real numbers. To describe a real number with infinite precision, one needs infinitely many bits. Therefore, a procedure that allows to describe the experts' degree of belief with ever increasing precision would, after any finite number of questions, only approximate the value of $\pi(x)$.

There is a similarity between this argument and an earlier one. There we knew x with some precision, while here we can know $\pi(x)$ only with some precision.

At each stage of determining $\pi(x)$, it is known with some precision δ. At this stage there are only finitely many distinguishable degrees of belief $h_0 = 0 < h_1 < h_2 < ... < h_k = 1$ such that $|h_{i+1} - h_i| \le \delta$. In other words, all the values of $\pi(x)$ from $h_0 = 0$ to h_1 are indistinguishable from 0, all the values from h_1 to h_2 are indistinguishable from h_1, etc. Changing the values $\pi(x)$ to the corresponding values $h_i$resultsin a piecewise constant function $\bar{\pi}(x)$ that is (at this stage) quite possible. Therefore, its uncertainty $U(\bar{m})$ is a possible value of uncertainty $U(\pi)$.

We get an entire interval of the possible values of $U(\bar{m})$. Each additional measurement diminishes the set of possible functions $\bar{m}$, making the resulting interval smaller. For $\delta \to 0$ this interval shrinks to a point, making natural to consider the limit of uncertainties of membership function π. Let us now state formal definitions.

DEFINITION 9 *A monotonic sequence* $h_0 = 0 < h_1 < h_2 < \cdots < h_k = 1$ *is* $\delta-$*precise,* $\delta > 0$, *if* $h_{i+1} - h_i \le \delta$ *for all* i.

DEFINITION 10 *For arbitrary monotonic* $\vec{h}$ *its projection* $pr_{\vec{h}} : [0,1] \to [0,1]$ *is* $pr_{\vec{h}}(y) = 0$ *for* $y < h_1$ *and* $pr_{\vec{h}}(y) = h_i$ *for* $h_i \le y < h_{i+1}$.

They have an immediate conclusion.

PROPOSITION 7 *If $\pi(x)$ is a membership function, and $\vec{h}$ is any sequence, then $\bar{\pi}(x) = pr_{\vec{v}}(\pi(x))$ is a piecewise constant membership function.*

It follows that for each of such projected membership functions $\bar{\pi}(x) = pr_{\vec{v}}(\pi(x))$, the uncertainty $U(\bar{\pi})$ is well defined. We need two more definitions.

DEFINITION 11 *Let $\delta > 0$, and $\pi(x)$ be a regular membership function. We say that a pair of real numbers (u, m) is a $\delta-$possible value of uncertainty for π, if $u = U(pr_{\vec{h}}(\pi(x))$ and $m = m(pr_{\vec{h}}(\pi(x))$ for some $\delta-$precise sequence $\vec{h}$.*

DEFINITION 12 *Let $\pi(x)$ be a regular membership function. We say that its uncertainty is defined and equal to (u, m) if whenever (u_n, m_n) is a δ_n-possible value of uncertainty for π, then $(u_n, m_n) \to (u, m)$. These values u and m will be denoted by $U(\pi)$ and $m(\pi)$.*

We can now state the main results.

THEOREM 1 *For an arbitrary regular membership function $\pi(x)$ its uncertainty is defined, $m(\pi) = \sup_x \pi(x)$, and*

$$U(\pi) = \int_0^{\sup_x \pi(x)} \log_2(\mu(\{x : \pi(x) \geq h\}))\, dh.$$

An immediate consequence is that for a normalized regular function $\pi(x)$, $m(\pi) = 1$ and

$$U(\pi) = \int_0^1 \log_2(\mu(\{x : \pi(x) \geq h\}))\, dh.$$

For a monotonic normalized function $\pi(x)$, this expression can be simplified further.

THEOREM 2 *Let $x_0 > 0$ be a positive real number. If $\pi(x) = 0$ for $x < 0$, $\pi(0) = 1$, $\pi(x) = 0$ for $x \geq x_0$, and for $x \in [0, x_0)$, $\pi(x)$ is continuous and decreasing, then*

$$U(\pi) = \log_2(x_0) - \frac{1}{\ln(2)} \int_0^{x_0} \frac{1 - \pi(x)}{x}\, dx.$$

The first term in this formula is the uncertainty contributed just just belonging to $[0, x_0]$. It implies that if we know only that $x \in [0, x_0]$, than we estimate the uncertainty in the value of x as $\log_2(x_0)$. If we include the additional information that x belongs to a fuzzy set described by a membership function $\pi(x)$, then we lower the certainty to the value $U(\pi) = \log_2(x_0) - (1/\ln(2)) \int_0^{x_0} (1 - \pi(x))/x\, dx$. This decrease in uncertainty $U(\pi_{[0,x_0]}) - U(\pi)$ measures the information that is brought on by this additional knowledge. For numerical representation of this information we can take this difference in uncertainties

$$I(\pi) = (1/\ln(2)) \int_0^{x_0} (1 - \pi(x))/x\, dx.$$

It is interesting to mention that the resulting expression for the information practically coincides with the expression $I(\pi) = \int_0^{x_0} (1 - \pi(x))/x\, dx$ that was introduced (from different assumptions) in [15, 17]. The factor $1/\ln(2)$ that makes them different is simply due to the fact that we use $\log_2$ for entropy, while some other authors use natural logarithms. This correspondence can be demonstrated formally by a suitable rearrangement of the membership functions.

4 Comparison of uncertainties

We will look now for the operations that lead, respectively, to the greatest and the least uncertainty. To formalize that, we must learn to compare uncertainties of different membership functions.

When we use entropy $S(p)$ as a definition of uncertainty of a probability distribution p, we just say that the uncertainty of a distribution p_1 is greater than the uncertainty of p_2 if $S(p_1) \geq S(p_2)$.

To be able to define uncertainty as a pair of numbers we introduce a relevant ordering.

DEFINITION 13 *Uncertainty* (u_1, m_1) *is greater than* (u_2, m_2)*, denoted* $(u_1, m_1) \succ (u_2, m_2)$ *if there exists an* ε_0 *such that for all* $\varepsilon < \varepsilon_0$*,*

$$u_1 - m_1 \log_2(2\varepsilon) > u_2 - m_2 \log_2(2\varepsilon).$$

The resulting order structure is lexicographic.

PROPOSITION 8 $(u_1, m_1) \succ (u_2, m_2)$ *if and only if either* $m_1 > m_2$*, or* $m_1 = m_2$ *and* $u_1 > u_2$*.*

4.1 Maximum uncertainty operations

Let us consider a typical fuzzy control setting. Assume given a set $S \subset R^n$, whose elements $\vec{x} = (x_1, \ldots, x_n) \in S$ ar the states of the system. Informally, the values $x_1, \ldots, x_n$ describe everything that we need to know to make a control decision. For example, if we control a heater/cooler, then $n = 1$, and the only variable we need to know is the difference $x_1 = t - t_0$ between the actual and the desired temperature. If we are controlling a spaceship, then we need to know its coordinates x_1, x_2, x_3, its current velocity vector (three more variables $x_4 = \dot{x}_1, x_5 = \dot{x}_2, x_6 = \dot{x}_3$), and two angles that describe the orientation. So, for a spaceship, $n = 8$.

Now let us fix a finite set $\mathcal{P}$ of continuous membership functions and view its elements as *fuzzy properties.* Examples are offered by such concepts as *big, medium* etc.

DEFINITION 14 *Elementary formula E is an expression of the form $P_i(x_i)$, where P_i is a fuzzy property. A rule is an expression of the form $E_1, \ldots, E_m \rightarrow P(u)$, where E_i are elementary formulae, P is a fuzzy property, and u is a special variable reserved for control. Formulae E_i are called conditions, while $P(u)$ is the conclusion of the rule. By a knowledge base we understand a finite set of rules.*

As an example one can consider rule $N(x_1) \rightarrow N(u)$, stating that if the difference $t - t_0$ between the actual and the desired temperatures is negligible, then the control should be negligible. Another possible rule is $SP(x_1) \rightarrow SN(u)$, meaning that if the difference $t - t_0$ is small positive, then we need to apply a small negative control (switch on the cooler for a little bit.) A similar rule $SN(x_1) \rightarrow SP(u)$ tells that if it becomes a little bit cold, it is necessary to switch on the heater for a while.

If we have a set of rules, then we can say that a control u is appropriate if and only if one the rules is applicable, and u appropriate according to this rule. Let us denote the statement *control u is appropriate* by $C(u)$. Then, for the three rules that describe the cooler/heater, we have the informal *formula* that describes when a control u is appropriate:

$$C(u) \equiv (N(x)\&N(u)) \vee (SP(x)\&SN(u)) \vee (SN(x)\&SP(u)).$$

Since $N(x)$, $N(u)$, $\ldots$, are fuzzy statements, we can get only fuzzy conclusions about the control, whereby $C(u)$ also becomes a fuzzy statement.

To extract the precise values, we need to choose some operations that would describe & and $\vee$ for fuzzy values. We can make a strong case for selecting control that corresponds to the biggest uncertainty.

When we make a choice of $\&-$ and $\vee-$operations, we restrict the set of possible control strategies. Since a wrong choice can lead to a low quality control, it sounds reasonable to try to loose as few possibilities as possible. In other words, it sounds reasonable to choose $\&-$ and $\vee-$operations in such a way that the uncertainty corresponding to $\pi_C(u)$ is the biggest possible.

This methodology is well known in the case when the uncertainty is probabilistic; it is called a *maximum entropy approach*, and it is widely applied to various problems ranging from processing physical data to processing uncertainties in expert systems [5, 8, 2, 9].

Initially, Zadeh [21] proposed using min and max, though stressing that these operations "are not the only operations in terms of which the union and intersection can be defined," and "which of these ... definitions is more appropriate depends on the context." Accordingly, we propose using the maximally general operations. Choice of operations that correspond to & and $\vee$, permits using the above formula to describe, for each u, what is the reasonable degree of belief that this value u is an appropriate control. In other words, we will be able to generate a membership function $\pi_C(u)$ that corresponds to a control rule. Following that we need some *defuzzification* procedure to transform this membership function into a single recommended control value.

To establish precise definitions, we need to consider what $\&-$ and $\vee-$operations could be appropriate. For the $\&-$operation (we will denote it by $f_\&(a, b)$) let us suppose that we have two statements A and B. Our degree of belief in A is equal to a, and belief in B is equal to b. If we have no other information about A and B, what must the reasonable degree of belief in $A\&B$ equal to? This reasonable degree of belief will be denoted by $f_\&(a, b)$. In the same situation, a reasonable degree of believe in $A \vee B$ will be denoted by $f_\vee(a, b)$, and $f_\vee$ will be called an $\vee-$operation.

In describing uncertainty of a membership function, we used the interpretation of membership values $\pi(x)$ as frequencies. Namely, we assumed that as a truth value $t(A)$ of an uncertain statement A, we take the ratio $t(A) = N(A)/N$, where $N(A)$ is the number of experts who believe in A, and N is the total number of experts that were questioned. In this interpretation, the following inequalities are true

$$\begin{aligned} N(A \vee B) &\leq N(A) + N(B) \\ N(A \vee B) &\leq N \\ N(A \vee B) &\geq N(A) \\ N(A \vee B) &\geq N(B) \end{aligned}$$

If we divide both sides of these inequalities by N, and combine them into one, we get

$$\max(t(A), t(B)) \leq (A \vee B) \leq \min(t(A) + t(B), 1),$$

hence

$$\max(a, b) \leq f_\vee(a, b) \leq \min(a + b, 1).$$

Likewise, from $N(A\&B) \leq N(A)$ and $N(A\&B) \leq N(B)$ we conclude that

$$t(A\&B) \leq \min(t(A), t(B)),$$

and thus $f_\&(a, b) \leq \min(a, b)$.

If beliefs in A and in B were independent, then we would have $t(A\&B) = t(A)t(B)$. In real life situations beliefs are not independent—if an expert has strong beliefs in several statements that later turn out to be true, then this means that he is really a good expert, and therefore it is reasonable to expect that his degree of belief in other statements that are true is greater. If A and B are complicated statements, then most of those experts who believe in A are presumably quite accomplished, and therefore they believe in B as well (and hence in $A\&B$). Therefore, the total number $N(A\&B)$ of experts who believe in $A\&B$ must be greater than the same number in the case when beliefs in A and B were uncorrelated random events. We conclude that the following inequality is reasonable to postulate

$$t(A\&B) \geq t(A)t(B),$$

hence, $f_\&(a, b) \geq ab$. In statistical terms, we can interpret it by saying that A and B are non-negatively correlated. We arrive at the following definitions

DEFINITION 15 *The and-or pair of continuous functions are* $f_\&, f_\vee : [0, 1] \times [0, 1] \to [0, 1]$, *which are non-decreasing in both variables, and satisfy*

- $\max(a,b) \le f_{\&}(a,b) \le \min(a+b,1)$
- $f_{\vee}(a,b) \le \min(a,b)$
- $f_{\&}(0,a)=0, f_{\&}(1,a)=a, f_{\vee}(0,a)=a, f_{\vee}(1,a)=1$
- $f_{\vee}(a,b)=f_{\vee}(b,a), f_{\&}(a,b)=f_{\&}(b,a)$

An *and-or* pair is called *correlated* if $f_{\&}(a,b) \ge ab$ for all a and b. The considerations behind these definition are quite simple:.

- If A is false, then $A\&B$ is also false, and $f_{\&}(0,a)=0$ for all a.
- If A is true, then $A\&B$ is true if and only if B is true. In this case $t(A\&B)=t(B)$, hence $f_{\&}(a,1)=a$ for all a.
- The statements that A and B are both true or that B and A are both true are equivalent. Therefore, $t(A\&B)$ must be always equal to $t(B\&A)$, or $f_{\&}(a,b)=f_{\&}(b,a)$ for all a,b.
- If our degree of belief in A increases, then our degree of belief in $A\&B$ cannot become less. The function $f_{\&}$ must be non-decreasing in both variables.
- If our degrees of belief in A and B change a slightly, then our degree of belief in $A\&B$ cannot change substantially. The smaller is the change in $t(A)$, $t(B)$, the smaller must be the change in $t(A\&B)$. In other words, the function $f_{\&}$ must be continuous.

Similar arguments justify the conditions on $f_{\vee}$.

It is convenient to extend our notation to more than two arguments. For three numbers a,b,c, we write $f_{\&}(a,b,c)=f_{\&}(f_{\&}(a,b),c)$, and in general, for $a,b,\ldots,c$

$$f_{\&}(a,b,\ldots,c)=f_{\&}(\ldots(f_{\&}(f_{\&}(a,b),\ldots)c).$$

Likewise

$$f_{\vee}(a,b,\ldots,c)=f_{\vee}(\ldots(f_{\vee}(f_{\vee}(a,b),\ldots)c).$$

The notation reflects the common understanding that $A\&B\&C=(A\&B)\&C$, and $A\vee B\vee C=(A\vee B)\vee C$. Now, we can define fuzzy control.

DEFINITION 16 *Assume given a knowledge base* $K = \{R_1, R_2, \ldots\}$*, an and-or pair* $(f_\&, f_\vee)$*, and a state* $\vec{x} \in S$*. The membership function corresponding to a rule* $P_1(x_{i_1}), \ldots, P_m(x_{i_m}) \to P(u)$*, is given by a function* $\pi_R = f_\&(P_1(x_{i_1}), \ldots, P_m(x_{i_m}, P(u))$*. Control membership function which corresponds to the knowledge base and the state* $\vec{x}$ *is*

$$\pi_C(u) = f_\vee(\pi_{R_1}, \pi_{R_2}, \ldots),$$

where $R_1, R_2, \ldots$ *are all the rules from* K*.*

We are now in a position to establish the operators that offer maximum uncertainty for the ensuing control rules.

THEOREM 3 *Suppose that* K *is a knowledge base, and* $\vec{x}$ *is a state. Let us denote by* $\pi_C(u)$ *the control membership function, that corresponds to an arbitrary* and-or *pair* $(f_\&(a,b), f_\vee(a,b))$*, and by* $\tilde{\pi}_C(u)$ *the control membership function that corresponds to the* and-or *pair* $(\min(a,b), \min(a+b,1))$*. Then,* $(U(\tilde{\pi}_C), m(\tilde{\pi}_C)) \succeq (U(\pi_C), m(\pi_C))$*.*

In other words, the greatest uncertainty is attained when we use $\min(a,b)$ for &, and $\min(a+b,1)$ for $\vee$. The proof is an immediate consequence of the definitions and the earlier remarks about inequalities among the uncertainty values. An interesting interpretation is provided by [11] that these operations lead to maximally stable controls. This agrees with a common-sense perception—we minimized the lost opportunities, and therefore ended up with the best possible control.

4.2 Minimization of uncertainty

Some of the earlier arguments are reasonable only if we are ready to apply various defuzzification techniques to extract the best control from $\pi_C(u)$. However, in industrial applications, a defuzzification rule is usually fixed. Since this rule may not necessarily be the most appropriate [20, 11], it is reasonable to try to rely on it only to a smallest extent feasible. In other words, in such cases, it is reasonable to choose & and $\vee-$ operations on the premise that the uncertainty related to $\pi_C(u)$ is the *least* possible. The following result is again immediate.

THEOREM 4 *Let* $\pi_C(u)$ *be the control membership function corresponding to an arbitrary correlated* and-or *pair* $(f_\&(a,b), f_\vee(a,b))$, *and let* $\tilde{\pi}_C(u)$ *be the control membership function for the pair* $(ab, \max(a,b)$. *Then*

$$(U(\pi_C), m(\pi_C)) \succeq (U(\tilde{\pi}_C), m(\tilde{\pi}_C)).$$

In other words, uncertainty is the least when we use ab for &, and $\max(a,b)$ for $\vee$. It has been shown in [11] that these very and-or operations lead to maximally smooth controls. This result also has a natural interpretation: since we are extremely cautious, we end up with a very smooth control.

5 Closing remarks

We already noted some analogies between our information and probabilistic entropy. It is, of course, the case of general complexity based informations. In a recent article [14] this is analyzed in considerable detail.

Similar correspondence can be established with fuzzy *continuous* information measures. We remarked on how the use of rearrangements helps to consolidate the results. Rearrangements other than decreasing are also of interest; in particular, those symmetric about the midpoint of the domain are likely to be most useful [6]. Further analysis is a subject of the on-going work.

ACKNOWLEDGEMENTS

The second author was supported by NSF Grant No. CDA–9015006, NASA Research Grant No. 9-482 and the Institute for Manufacturing and Materials Management grant. The authors are thankful to Bob Lea, Ron Yager, and John Yen for the stimulating discussions.

References

[1] S. S. L. Chang and L. A. Zadeh. *On fuzzy mapping and control*, IEEE Transactions on Systems, Man and Cybernetics, 1972, Vol. SMC-2, pp. 30–34.

[2] P. Cheeseman. *In defense of probability* in Proceedings of the 8-th International Joint Conference on AI, Los Angeles, 1972, pp. 1002–1009.

[3] D. Dubois and H. Prade. *Possibility Theory.* Plenum Press, New York 1988.

[4] G. H. Hardy, J. E. Littlewood, and G. Polya. *Inequalities.* Cambridge University Press, Cambridge 1934.

[5] E. T. Jaynes. *Where do we stand on maximum entropy?* in R. D. Levine and M. Tribus (Eds.) *The maximum entropy formalism*, MIT Press, Cambridge, MA, 1979.

[6] B. Kawohl. *Rearrangements and Convexity of Level Sets in PDE.* Lecture Notes in Mathematics 1150, Springer-Verlag 1985.

[7] G. J. Klir and T. A. Folger. *Fuzzy sets, uncertainty and information.* Prentice Hall, Englewood Cliffs, NJ, 1988.

[8] O. M. Kosheleva, V. Kreinovich. *A letter on maximum entropy method.* Nature, 1979, Vol. 281, pp. 708–709.

[9] V. Kreinovich. *Entropy approach for the description of uncertainty in knowledge bases*, Tech. Rep., Center for the New Informational Technology "Informatika", Leningrad, 1989 (in Russian).

[10] V. Kreinovich, C. Quintana, and R. Lea. *What procedure to choose while designing a fuzzy control? Towards mathematical foundations of fuzzy control*, Working Notes, 1st Int. Workshop on Industrial Appl. Fuzzy Control and Intelligent Systems, College Station, TX, 1991, pp. 123–130.

[11] V. Kreinovich, C. Quintana, R. Lea, O. Fuentes, A. Lokshin, S. Kumar, I. Boricheva, and L. Reznik. *What non-linearity to choose? Mathematical foundations of fuzzy control.* Proc. 1992 Int. Fuzzy Systems and Intelligent Control Conference, Louisville, KY, 1992, pp. 349–412.

[12] C. C. Lee. *Fuzzy logic in control systems: fuzzy logic controller.* IEEE Trans. Systems, Man and Cybernetics, 1990, Vol. 20, No. 2, pp. 404–435.

[13] E. H. Mamdani. *Application of fuzzy algorithms for control of simple dynamic plant*, Proceedings of the IEE, 1974, Vol. 121, No. 12, pp. 1585–1588.

[14] E. W. Packel, J. F. Traub, H. Wozniakowski. *Measures of uncertainty and information in computation*, Information Sciences, Vol. 65, pp. 253–273.

[15] A. Ramer. *Concepts of fuzzy information measures on continuous domains*, Int. J. General Systems, 1990, Vol. 17, pp. 241–248.

[16] A. Ramer, J. Hiller, C. Padet. *Principle of Maximizing Fuzzy Uncertainty*, Int. J. General Systems, 1993 (to appear.)

[17] A. Ramer, R. Yager. *Analysis of specificity of fuzzy sets*, Proceedings of the IEEE International Conference on Fuzzy Systems, San Diego, CA, 1992, pp. 1097–1104.

[18] J. F. Traub, H. Wozniakowski. *A General Theory of Optimal Algorithms.* Academic Press, New York 1980.

[19] J. F. Traub, G. W. Wasilkowski, H. Wozniakowski. *Information-Based Complexity.* Academic Press, New York 1988.

[20] J. Yen, N. Pfluger, and R. Langari. *A defuzzification strategy for a fuzzy logic controller employing prohibitive information in command formulation*, Proc. IEEE Int. Conference on Fuzzy Systems, San Diego, CA, March 1992.

[21] L. A. Zadeh. *The concept of a linguistic variable and its application to approximate reasoning*, Part 1, Information Sciences, 1975, Vol. 8, pp. 199–249.