An Algebraic Approximation to the Classification With Fuzzy Attributes

Antonio Gisolfi
Dipartimento di Informatica
Università di Salerno
Italy

Gustavo Nunez
Departament de llenguatges
i sistemes informatics
Facultat d'informàtica de Barcelona
Spain

ABSTRACT

This paper proposes an algebraic approximation to the problem of generating appropriate classifications in a universe of discourse in which the characteristics of the elements are defined through attributes whose values are linguistic labels. Classifications are generated by means of operations between ordered strings induced by the fuzzy attributes. Moreover, a heuristic measure is used, based on the concept of statistical variance, in order to recognize the potential relevance of each attribute with respect to the generated classifications.

KEYWORDS: ***fuzzy, classification, fuzzy attributes, algebraic structure, approximation, relevance.***

1. INTRODUCTION

Understanding the intimate nature of classification problems is a central issue having relevant practical implications in all cognitive sciences, because classifying is a basic step in the processes concerning knowledge

Address correspondence to Antonio Gisolfi, Dipartimento di Informatica, Università di Salerno, I 84081 Baronissi (SA), Italy.

This work was partially supported by Integrated Action Italy–Spain (1991).
Received January 1, 1992; accepted December 29, 1992.

International Journal of Approximate Reasoning 1993; 9:75–95

655 Avenue of the Americas, New York, NY 10010 0888-613X/93/$6.00

acquisition in whatever scientific field. In particular, machine learning is concerned with computer-based classification methods, and two classes of classification problems are currently being investigated:

1. Given a classification of the universe of discourse, define rules in order to decide to which class each element of the universe belongs.
2. Given a universe of discourse, construct appropriate classification of the universe.

Most efforts of the researchers in this area have focused on the first class of problems. We remember that most first-generation expert systems, e.g., Buchanan and Shortliffe [1] and Michalsky et al. [2] are based on classification systems that use suitable conjunctions of rules, implemented by decision trees, to determine, according to its characteristics, to what class a specific element of the universe is to be assigned. Usually this problem is tackled by letting the human expert make the ultimate decision about the most relevant attributes for the classification.

This approach presents several disadvantages—see for example [3], [4]—and, as a consequence, nowadays there is a trend towards the development of systems able to decide automatically, by means of methods based on inductive inference, what are the most relevant attributes that can be used to build the classification rules.

From the point of view of machine learning, the problem of constructing appropriate classifications starting from a conjunction of empirical data is a basic one because when one starts to investigate a particular knowledge area there are no classifications of the elements it consists of and the latter are to be obtained beginning from the results at disposal and by applying suitable methods that analyze and compare the available data. In this way the human expert has to tackle the hard, and often boring, task of studying huge collections of data to recognize similarities that enable definition of adequate initial classifications. It would be worth freeing the human expert from this burden, which does not require creativity, and consequently several methodologies are currently being developed to generate automatic classifications beginning from empirical data that can be used as a starting point to investigate a specific knowledge area. Traditionally this problem is coped with by considering appropriate measures defined in the universe of discourse.

In this paper a suitable algebraic approximation to the problem of generating classifications is illustrated; we suppose that the characteristics of the elements of the universe of discourse are defined by fuzzy attributes whose values are fuzzy numbers. The classifications are generated through operations that are carried out onto ordered strings that are induced by the fuzzy attributes. Moreover, a method is proposed whose goal is to evaluate the potential usefulness of the attributes that are present in a classification; the method relies on a heuristic measure based on the statistical variance that permits construction of a decision tree for each classification.

This paper is organized as follows. The algebraic structure induced by the fuzzy attributes is defined in Section 2 and some relevant properties are proved. Section 3 shows in what way the structure can generate different classifications in the same universe of discourse. Then, in Section 4, the problem concerning the nature of the relevance is discussed, and a heuristic measure is defined to evaluate the potential usefulness of a conjunction of attributes as regards a specific classification. Finally the concluding remarks and the work that is underway is briefly discussed in Section 5.

2. ALGEBRAIC CLASSIFICATION WITH FUZZY ATTRIBUTES

In this section we define the algebraic structure that is induced by a conjunction of fuzzy attributes. Some basic definitions are given and, in particular, the operation that distinguishes the structure is introduced. The properties that must be satisfied by the operation are discussed, and moreover, we prove that these properties completely characterize the operation. In paragraph 2.5 a method of linguistic approximation is introduced that can be used to avoid the combinatorial explosion of the classes when the method is applied to specific problems.

2.1. The Algebraic Structure

In the following we shall briefly, recall some of the basic features of the proposed algebraic structure [5], [6].

DEFINITION 2.1 *Let U be a universe of discourse. Suppose that the elements of U can be represented by ordered k-uples* $(A_1(u), \ldots, A_k(u))$ *where the* A_i *are fuzzy measures of the elements of U whose values are ordered sets of fuzzy numbers.*

Then we choose to represent each attribute A_i *of the set* $\{A_1, \ldots, A_k\}$ *by the ordered string*

$$a_n^{\alpha n} a_{n-1}^{\alpha n-1} \ldots a_2^{\alpha 2} a_1^{\alpha 1}$$

where $a_j = A_j^{-1}(\alpha_j)$ *is the subset formed by the elements of U for which the attribute* A_j *takes the value* α_j *(here* $\alpha_1 < \alpha_2 < \cdots < \alpha_n$*).*

We note that the collection of sets

$$\{a_n, a_{n-1}, \ldots, a_1\}$$

is an ordered partition of the set U. In this way one gets that each fuzzy attribute induces a fuzzy partition, i.e., a classification of the elements of U.

EXAMPLE 2.1. Consider, for instance, the set U of all Italian coins {10, 50, 100, 200, and 500 lira} and the following fuzzy measure, of second order, defined on U:

> *Value* = *false*/10 + *almost true*/(50 + 100) + *true*/200 + *completely true* /500
> where *false*, *almost true*, *true*, and *completely true* are linguistic labels of second order, which fulfill the previous relations, defined, in turn, by the following fuzzy numbers:
> *false* = 1/0 + 0.8/0.1 + 0.5/0.2 + 0.3/0.3
> *almost true* = 0.5/0.3 + 0.7/0.4 + 1/0.5 + 0.7/0.6 + 0.5/0.7
> *true* = 0.3/0.5 + 0.7/0.6 + 0.9/0.7 + 1/0.8 + 0.9/0.9 + 0.7/1
> *completely true* = 0.3/0.6 + 0.5/0.7 + 0.7/0.8 + 0.9/0.9 + 1/1

The fuzzy measure *Value* induces the ordered string $a_4^{ct} a_3^{t} a_2^{at} a_1^{f}$ where

$$a_4 = \{100\} = Value^{-1}(completely\ true);\ a_3 = \{200\} = Value^{-1}(true)$$
$$a_2 = \{50{,}100\} = Value^{-1}(almost\ true);\ a_1 = \{10\} = Value^{-1}(false)$$

Then, as each fuzzy measure induces a partition of the universe of discourse, one can suppose, without losing generality, that the fuzzy values of the attributes are linguistic labels. In fact, if we consider the fuzzy set:

> $Value_1$ = 0.1/10 + 0.3/50 + 0.5/100 + 0.7/200 + 0.8/500
> then it follows that *Value* is the composition of the functions *truth–value* and $Value_1$, where *truth–value* is the fuzzy number of second order defined as follows:
> *truth-value* = *false*/(0 + 0.1 + 0.2) + *almost true*/(0.3 + 0.4 + 0.5) + *true*/(0.6 + 0.7) + *completely true*/(0.8 + 0.9 + 1).

This fact shows that, in general, one can suppose that the range of the attributes is a set of linguistic labels. We note that, by means of the ordered strings, one can code, in some sense, the fuzzy information contained in each attribute in order to obtain homogenous elements that can be compared.

The basic idea to obtain the algebraic structure for classification is to define an operation between the ordered strings induced by the attributes, so that the resulting ordered string represents a finer classification of the universe in comparison with those furnished by the original strings. The definition of this operation can be justified intuitively by considering the analogy between the ordered strings and the positional system for natural numbers: in both systems each figure plays two distinct roles, i.e., value and position. For example, in the number "4678," the figure "7" denotes that its contribution to the number is obtained by multiplying 7 (value) by ten to the power of two (position). In the same way, in the string:

$$a_4^{ct} a_3^{t} a_2^{at} a_1^{f}$$

the absolute value of the symbol a_2 is the set of all 50 lira coins, whereas its relative value is the linguistic label "almost true." We note that:

$$a_4^{ct'}a_3^{t}a_2^{at'}a_1^{f} \neq a_4^{ct'}a_3^{t}a_1^{at'}a_2^{f}$$

Thus, the algebraic manipulation of the ordered strings can be seen as a slight modification of the well-known algorithm for the multiplication of natural numbers. However, we emphasize that in our case the elements that constitute the strings are not homogenous, because the bases are ordinary sets and the exponents are fuzzy sets. In fact, the operation among strings consists of two different operations: the first operation involves ordinary sets (bases) and the other fuzzy ones (exponents). In the above-mentioned example, one has:

first part: a_4 a_3 a_2 a_1 (bases)
second part: *ct t at f* (exponents)

2.2 Defining the Operation for the First Parts

Let us consider the strings of length 4
$A = a_4^{ct'}a_3^{t}a_2^{at'}a_1^{f}$ and $B = b_4^{ct'}b_3^{t}b_2^{at'}b_1^{f}$, then the algorithm for string multiplication concerning the first parts of A and B is defined as follows:

			a_4	a_3	a_2	a_1	$\otimes$
			b_4	b_3	b_2	b_1	
			$a_4 \otimes b_1$	$a_3 \otimes b_1$	$a_2 \otimes b_1$	$a_1 \otimes b_1$	$\oplus$
		$a_4 \otimes b_2$	$a_3 \otimes b_2$	$a_2 \otimes b_2$	$a_1 \otimes b_2$	—	$\oplus$
	$a_4 \otimes b_3$	$a_3 \otimes b_3$	$a_2 \otimes b_3$	$a_1 \oplus b_3$	—	—	$\oplus$
$a_4 \otimes b_4$	$a_3 \otimes b_4$	$a_2 \otimes b_4$	$a_1 \otimes b_4$	—	—	—	=
c_7	c_6	c_5	c_4	c_3	c_2	c_1	

where:

$$c_7 = [(a_4 \otimes b_4)]$$

$$c_6 = [((a_4 \otimes b_3) \oplus (a_3 \otimes b_4))]$$

$$c_5 = [((a_4 \otimes b_2) \oplus (a_3 \otimes b_3) \oplus (a_2 \otimes b_4))]$$

$$c_4 = [((a_4 \otimes b_1) \oplus (a_3 \otimes b_2) \oplus (a_2 \otimes b_3) \oplus (a_1 \otimes b_4))]$$

$$c_3 = [((a_3 \otimes b_1) \oplus (a_2 \otimes b_2) \oplus (a_1 \otimes b_3))]$$

$$c_2 = [((a_2 \otimes b_1) \oplus (a_1 \otimes b_2))]$$

$$c_1 = [(a_1 \otimes b_1)]$$

The symbols $\oplus$ and $\otimes$ are operations defined in the power set $P(U)$ and correspond to the ordinary operations of sum and multiplication of natural numbers, respectively.

In general, we can give the following

DEFINITION 2.2 *If* $A = a_n^{\alpha n} a_{n-1}^{\alpha n-1} \ldots a_1^{\alpha 2}$ *and* $B = b_m^{\beta m} b_{m-1}^{\beta m-1} \ldots b_1^{\beta 1}$ *are strings of length n and m, respectively, then the operation * for the first parts of the strings is defined as follows:*

$$(a_n a_{n-1} \ldots a_2 a_1)^* (b_m b_{m-1} \ldots b_2 b_1) = c_{m+n-1} c_{m+n-2} \ldots c_2 c_1$$

where, for $n \geq m$:

$$c_i = \begin{cases} \oplus_{j=1,\ldots,i} & a_{i-j+1} \otimes b_j & \text{if } 1 \leq i \leq m-1 \\ \oplus_{j=1,\ldots,m-(i-n)} & a_{n-j+1} \otimes b_{j+(1-n)} & \text{if } n+1 \leq i \leq m+n-1 \\ \oplus_{j=1,\ldots,m} & a_{i-j+1} \otimes b_j & \text{if } m \leq i \leq n \end{cases} \tag{1}$$

and for $n < m$:

$$c_i = \begin{cases} \oplus_{j=1,\ldots,i} & a_{i-j+1} \otimes b_j & \text{if } 1 \leq i \leq n \\ \oplus_{j=1,\ldots,n-(i-m)} & a_{n-j+1} \otimes b_{j+(1-n)} & \text{if } m+1 \leq i \leq n+m-1 \\ \oplus_{j=1,\ldots,n} & a_{n-j+1} \otimes b_{j+1(i-n)} & \text{if } n+1 \leq i \leq m \end{cases}$$

We emphasize that the operations $\oplus$ and $\otimes$ are fully determined if the set $c_1, \ldots, c_{m+n-1}$ has to represent a finer classification in comparison with those induced by A and B. These properties are the following:

1. Closure

   In order that the result, C, of the operation between A and B is a string, for each i different from j, the relation $c_i \cap c_j = \emptyset$ must hold true.
2. Commutativity

   This property must hold because the order the attributes are considered should not affect the final classification and is equivalent to the commutativity of the operations $\oplus$ and $\otimes$:

   $$a \oplus b = b \oplus a \qquad a \otimes b = b \otimes a$$

   for each couple of elements a and b belonging to $P(U)$.
3. Associativity

   This property derives from the requirement that the classification should be independent from the way the attributes are associated with respect to the operation defined. Requiring the associativity of the operation concerning the first parts of the strings is clearly equivalent to require the associativity of the operations $\oplus$ and $\otimes$:

$$a \oplus (b \oplus c) = (a \oplus b) \oplus c \qquad a \otimes (b \otimes c) = (a \otimes b) \otimes c$$

for each triple of elements belonging to $P(U)$.

4. Idempotence
It is quite obvious that by combining a classification with itself, the same classification is gotten. This property is equivalent to require that for the operations $\oplus$ and $\otimes$ the following properties hold true:

$$a \oplus a = a \qquad a \otimes a = a$$

5. Existence of the zero element
We require that by combining $a_n^{\alpha n} a_{n-1}^{\alpha n-1} \ldots a_1^{\alpha 1}$ with the trivial classification, the same classification is still achieved, i.e., this property is equivalent to the following ones:

$$a \oplus U = U \qquad a \otimes U = a$$
$$a \oplus \emptyset = a \qquad a \otimes \emptyset = \emptyset$$

It is apparent that the set theoretic union and intersection are operations that fulfill the properties 1, 2, 3, 4, and 5. Moreover these properties fully characterize the operations $\oplus$ and $\otimes$. Thus, in the following we suppose that these operations are the union and the intersection, respectively.

EXAMPLE 2.2. Consider the set $U = \{a, b, c, d, e\}$ and suppose that one has the following strings:

$$A = [a, b]^{ct}[c]^{t}[c]^{at}[e]^{f}$$
$$B = [a]^{ct}[c]^{t}[e, d]^{at}[b]^{f}$$

By multiplying the first parts of A and B, one gets:

			$[a, b]$	$[c]$	$[d]$	$[e]$	*
			$[a]$	$[c]$	$[e, d]$	$[b]$	
			$[b]$	$[-]$	$[-]$	$[-]$	
		$[-]$	$[-]$	$[d]$	$[e]$	$-$	
	$[-]$	$[c]$	$[-]$	$[-]$	$-$	$-$	
$[a]$	$[-]$	$[-]$	$[-]$	$-$	$-$	$-$	
$[a]$	$[-]$	$[c]$	$[b]$	$[d]$	$[e]$	$[-]$	

We note that the resulting string is a classification of the universe U utterly different from those induced by the strings A and B.

It is worth emphasizing that the operation * among the first parts of the strings is carried out in linear time. In fact, for instance, the element "c" is taken into account thirdly because it occupies the third position in the multiplier and thus in the third cycle "c" occupies the same position as in the multiplier. This fact implies a double shift as regards the resulting string. The final position of "c" within the string (the fifth, in our example) is the sum of its original position and of its left shift. In general, if an element occupies the position p in the multiplier and its left shift equals $p - 1$, then its final position in the resulting string equals $2p - 1$.

2.3. Defining the Operations for the Second Parts

To define the operations concerning the second parts, namely the strings of fuzzy numbers, a similar process has to be carried out. In general, this operation is defined as follows:

DEFINITION 2.3.1

$$(\alpha_n \dots \alpha_2 \alpha_1) \blacklozenge (\beta_m \dots \beta_2 \beta_1) = \gamma_{m+n-1} \dots \gamma_2 \gamma_1$$

where for $n \geq m$:

$$\gamma_i = \begin{cases} \oplus_{j=1,\dots,i} & \alpha_{i-j+1} \otimes \beta_j & \text{if } 1 \leq i \leq m-1 \\ \oplus_{j=1,\dots,m-(i-n)} & \alpha_{n-j+1} \otimes \beta_{j+(i-n)} & \text{if } n+1 \leq i \leq m+n-1 \\ \oplus_{j=1,\dots,m} & \alpha_{i-j+1} \otimes \beta_j & \text{if} m \leq i \leq n \end{cases} \tag{2}$$

In this case $\oplus$ and $\otimes$ are operations defined for couples of fuzzy numbers.

For example, if one assumes that $\oplus$ = max and $\otimes$ = min the structure (Fuzzy Numbers (0,1), $\oplus$, $\otimes$) is a distributive and semicomplemented lattice. If we consider the set of linguistic labels {f, at, t, ct} of example 2.2, then one gets:

			ct	*t*	*at*	*f*	♦
			ct	*t*	*at*	*f*	
			f	*f*	*f*	*f*	
		at	*at*	*at*	*f*	–	
	t	*t*	*at*	*f*	–	–	
ct	*t*	*at*	*f*	–	–	–	
ct	*t*	*t*	*at*	*at*	*f*	*f*	

If one combines the two parts, the resulting string is as follows:

$$[a]^{ct}[-]^{t}[c]^{t}[b]^{at}[d]^{at}[e]^{f}[-]^{f}$$

and, moreover, one gets the union of the sets sharing the same label:

$$[a]^{ct}[c]^{t}[b,d]^{at}[e]^{f}$$

Let us consider these results: the element b has the labels "ct" and "f" in A and B, respectively. The element d, in turn, has the label "at" in both strings. Comparing the first parts, one has two positions for these elements, the fourth for b, and the third for d. We note that an intermediate value between "ct" and "f" should not necessarily be the double "at " for d as occurs in this example. This fact occurs because the operations carried out on the second parts are not able to generate intermediate labels. This feature of the operation is a disadvantage when one aims at generating classifications because the resulting string is not necessarily a finer classification.

Thus the operation for ordered strings is defined as follows:

DEFINITION 2.3.2 *If* $a_n^{\alpha n} \ldots a_2^{\alpha 2} a_1^{\alpha 1}$ *and* $b_n^{\beta n} \ldots b_2^{\beta 2} b_1^{\beta 1}$ *are ordered strings then*

$$(a_n^{\alpha n} \ldots a_2^{\alpha 2} a_1^{\alpha 1}) \diamond (b_n^{\beta n} \ldots b_2^{\beta 2} b_1^{\beta 1}) = c_{m+n-1}^{\gamma m+n-1} \ldots c_2^{\gamma 2} c_1^{\gamma 1}$$

where c_j *and* γ_j *are obtained by the formulas* (1) *and* (2) *respectively. In the following we shall cope with the problem of characterizing the operation for the second parts.*

2.4. Characterizing the Operation Concerning the Second Parts

In general, in order that the operations concerning the two parts of the strings produce a string that represents a finer partition in comparison to those furnished by the original strings, the operation for the second parts should satisfy the following properties:

1. Closure
   The operation must furnish a fuzzy number in [0, 1], namely $\lambda_{2m-1}, \ldots, \gamma_2, \gamma_1$ must be fuzzy numbers defined in the interval [0, 1].
2. Commutativity
3. The operation must preserve the ordering among the fuzzy numbers This property is expressed formally as follows:

$$\gamma_1 \leq \gamma_2 \leq \ldots \leq \gamma_{m+n-1}$$

For the sake of simplicity, we have only considered a subclass of the fuzzy numbers in [0, 1], i.e., the so called L-L Fuzzy Numbers [7], [8] that include the triangular and trapezoidal fuzzy numbers that are, in turn, a subclass of the flat fuzzy numbers.

The work with triangular fuzzy numbers offers some advantages (for example their representation is simple) in comparison to the representation with trapezoidal numbers, thus we have used them in the following; however, this is not a limiting feature.

In particular, we recognize four triangular fuzzy numbers that can be depicted graphically as shown in Figure 1.

These fuzzy numbers have been used to represent the linguistic labels *completely true* (ct), *true* (t), *almost true* (at) and *false* (f).

By applying the algorithm developed by Dubois-Prade [9], the extended sum and product of the labels at and t can be depicted, respectively, as shown in Figures 2 and 3.

Here, $La[at, t]$ denotes the linguistic approximation resulting from the combination of the values "at" and "t". We note that we just have applied the normal operations of sum and product to the extrema of the fuzzy numbers. Another operation that is commonly carried out upon triangular numbers is defined as follows:

DEFINITION 2.4 *The "extended mean" is defined, according to the principle of extension of the binary operations, as follows*:

$$\mu_{A \circledR B}(z) = \sup_{z=(x+y)/2} \min(\mu_A(x), \mu_B(y))$$

The result of this operation, when the latter is applied to the labels "at" and "t" is depicted in Figure 4.

Also in this case, the extended mean is found by calculating the usual mean of the extrema of the fuzzy numbers.

The extended sum, product, and mean are the operations selected by us to tackle the problem of recognizing adequate definitions for ⊕ and ⊗ in order that the operations for the second parts have the above-mentioned properties.

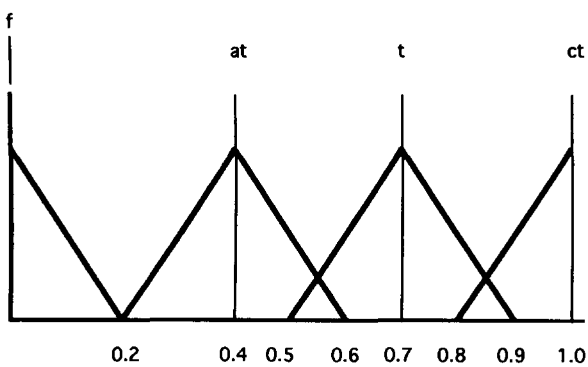

Figure 1. Triangular fuzzy numbers.

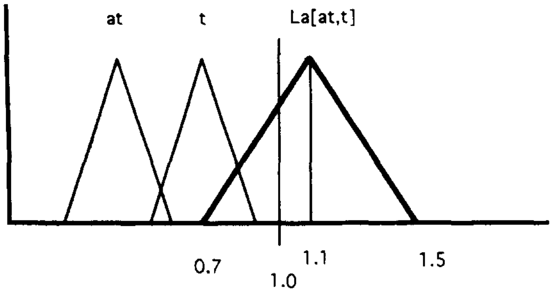

Figure 2. Extended sum of the labels "at" and "t".

Thus, if we consider ⊗ = product and ⊕ = sum, then the result can be depicted graphically as shown in Figure 5.

We note that labels ranging outside the interval [0, 1] are obtained. Moreover, the ordering is not preserved and, consequently, these operations cannot be used. For any choice of ⊕ and ⊗ in the set {product, sum} the same negative remarks can be easily drawn.

Let us consider now the extended mean operation previously defined, which will be represented by the symbol ®, in combination with the extended product and sum.

In particular if we take ⊗ = ® and ⊕ = product the results are depicted graphically in Figure 6.

In this case, all the linguistic labels correctly range in the interval [0, 1], but still the ordering among the labels is not preserved. Moreover, the values cluster around zero, and this derives from the fact that the product is a decreasing operation. If we consider the opposite choice, i.e., ⊗ = product and ⊕ = ®, the result is similar to that depicted above.

It is easily shown that the choices ⊕ = ® ⊗ = sum do not give the suitable properties. Let us consider what happens when one chooses only

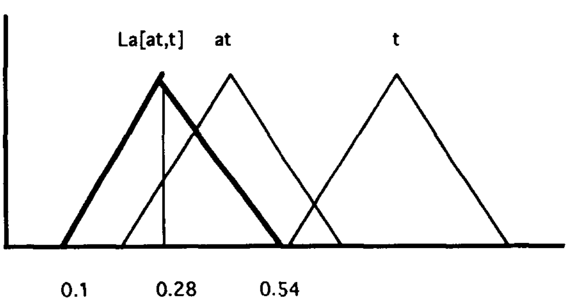

Figure 3. Extended product of the labels "at" and "t".

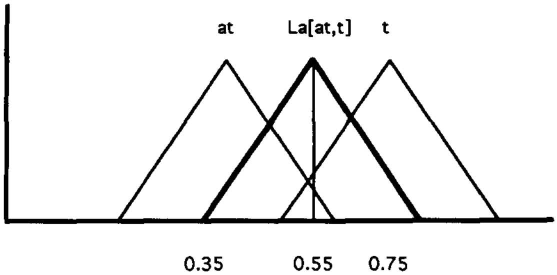

Figure 4. Extended mean applied to the labels "at" and "t".

the extended mean operation, i.e., suppose that ⊗ = ⊕ = ®. In this case, Figure 7 depicts the result.

We note that the ordering is preserved, the range is [0, 1], and the intermediate fuzzy numbers are appropriately arranged. If we consider again the strings A and B, as in the example 2.1., we have:

$$a\gamma_7\ c\gamma_5\ b\gamma_4\ d\gamma_3\ e\gamma_2$$

where each γ_i denotes a linguistic label whose value has to be suitably approximated. We note that, in some examples, 11 truth values have been considered that have been symbolically denoted by the labels $\text{tv}_0, \text{tv}_1, \ldots, \text{tv}_{10}$ (Figure 8).

2.5. The Linguistic Approximation

Because of the clustering of the elements, it is worth defining appropriate rules of approximation instead of new linguistic labels. Given two fuzzy numbers, we can calculate the least distance between the mean value of the fuzzy number to be approximated with respect to the 11 above-mentioned fuzzy values.

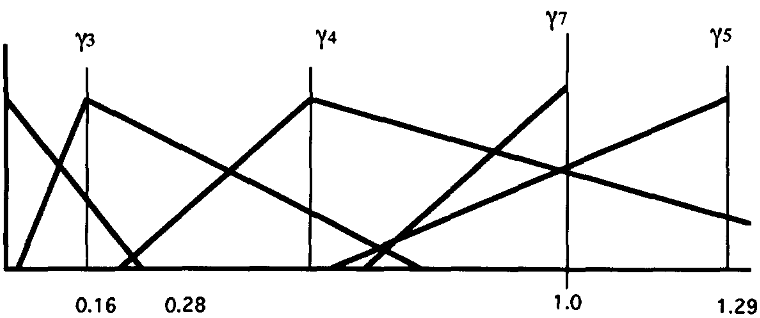

Figure 5. Case: ⊗ = product, ⊕ = sum.

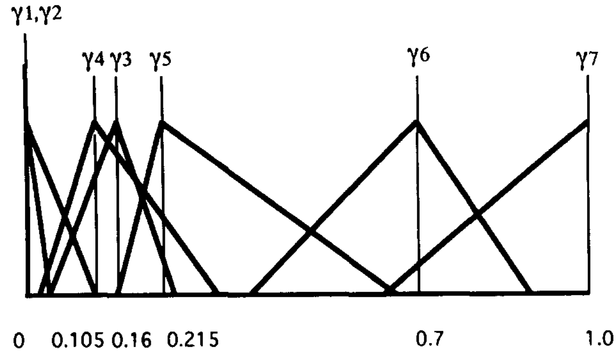

Figure 6. Case: ⊗ = ®, ⊕ = product.

Suppose that the mean value, m_a, of the fuzzy number to be approximated lies in the interval $(m_{tv_i}, m_{tv_{i+1}})$ whose extrema are the mean values of the fuzzy numbers tv_i and tv_{i+1}, and consider $d = m_{tv_{i+1}} - m_{tv_i}$. We say that the label corresponding to the fuzzy number to be approximated is tv_i if m_a lies in the interval $[mt_{v_i}, mt_{v_i} + d/10)$.

In case that $m_a \in \{m_{tv_i} + d/10, m_{tv_i} + 3d/10\}$ then we say that the label corresponding to the fuzzy number is called "next to" the label tv_i and it is represented by $nt[tv_i]$.

If $m_a \in \{m_{tv_i} + 3d/10, m_{tv_i} + 7d/10\}$ we are dealing with a number whose value is intermediate between the linguistic labels and, therefore, the label "included between" the labels tv_i and tv_{i+1} is attached to the fuzzy number. When $m_a \in \{m_{tv_i} + 7d/10, m_{tv_{i+1}} + 9d/10\}$ we say that the label of the fuzzy number is called "before" the label tv_{i+1} and is denoted by $bb[tv_{i+1}]$. Finally, if $m_a \in \{m_{tv_i} + 9d/10, t_{v_{i+1}}\}$ we say that the label corresponding is $t_{v_{i+1}}$.

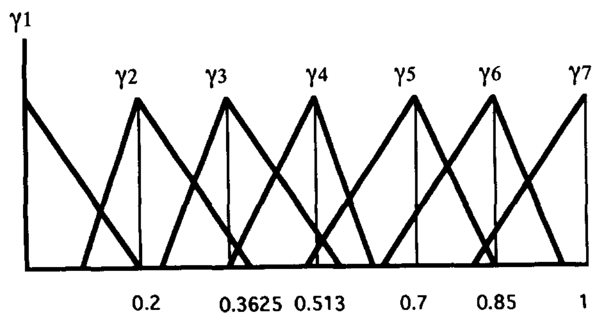

Figure 7. Case: ⊗ = ⊕ = ®.

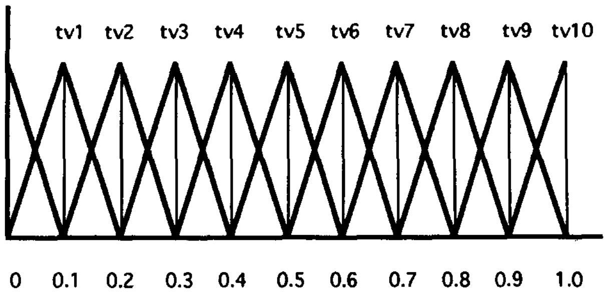

Figure 8. Depicting eleven truth values.

The above-mentioned linguistic approximation is calculated after implementing the operation ♦ concerning the second parts.

We note that there is a semantic problem with the linguistic approximations. In fact, in some cases, some labels whose semantic interpretation is very difficult have been generated.

We note that our process of linguistic approximation furnishes an upper bound to the number of obtainable labels—in fact their number cannot exceed the value $4n$-3 where n denotes the original number of linguistic labels that are taken for reference.

3. AN EXAMPLE CONCERNING MEDICAL DIAGNOSES

We aim at studying in what measure a disease can be inferred as a likely consequence of some quoted symptoms. For instance, consider a population whose elements suffer, to different extents, from, say, hypertension, diabetes, and cholesterol. Moreover, we have to consider also the social and psychological behavior of the individuals, because these features are as important as the previous symptoms in order to give rise to the disease *coronaropathy*; such forerunners might be stress, social status, anxiety, etc. We have to construct as many ordered strings as the distinguishing features. Each ordered string contains the individuals that suffer from a symptom S and a fuzzy number, i.e., a linguistic label, is associated with each of them. In such way we get statements having the following structure: (x has S) is τ.

Of course the labels have to be the same for all the ordered strings so that the operation can be carried out coherently. For the sake of simplicity we consider only the symptoms Cholesterol (C), Stress (S), Anxiety (A), and four grades for τ: *completely true, true, almost true, false*. In the following we shall write, for short, *ct*, *t*, *at*, *f*.

Let $U = \{a, b, c, d, e, f, g\}$ be a population of individuals and suppose that:

$$C = [a, b, c]^{ct}[d]^{t}[e, f, g]^{f}$$

$$S = [a, b,]^{ct}[c, d]^{t}[e]^{at}[f, g]^{f}$$

$$A = [a, b, c, d]^{ct}[e]^{at}[f, g]^{f} \tag{3}$$

In the following we will calculate the ordered string $C \diamond S \diamond A$ which describes how likely is the disease coronaropathy

First parts operations:

C:					$[a, b, c]$	$[d]$	–	$[e, f, g]$	*
S:					$[a, b,]$	$[c, d]$	$[e]$	$[f, g]$	
					–	–	–	$[f, g]$	
					–	–	–	$[e]$	
			$[c]$	$[d]$	–	–			
		$[a, b]$	–	–	–				
$C \diamond S$:		$[a, b]$	$[c]$	$[d]$	–	–	$[e]$	$[f, g]$	*
A:					$[a, b, c, d]$	–	$[e]$	$[f, g]$	
		–	–	–	–	–	–	$[f, g]$	
	–	–	–	–	–	$[e]$	–		
	$[a, b]$ $[c]$ $[d]$	–	–	–	–	–			
	$[a, b]$ $[c]$ $[d]$	–	–	–	–	$[e]$	–	$[f, g]$	

Second parts operations:

			ct	t	at	f	♦
			ct	t	at	f	
			$ct®f$	$t®f$	$at®f$	$f®f$	
		$ct®at$	$t®at$	$at®at$	$f®at$		
	$ct®t$	$t®t$	$at®t$	$f®t$			
$ct®ct$	$t®ct$	$at®ct$	$f®ct$				
γ_7	γ_6	γ_5	γ_4	γ_3	γ_2	γ_1	

Here, the γ_i are triangular fuzzy numbers that could be represented by:

$$\gamma_1 = [0,0,0.2] \quad \gamma_2 = [0.1,0.2,0.4] \quad \gamma_3 = [0.225,0.375,0.575]$$
$$\gamma_4 = [0.375,0.525,0.675] \quad \gamma_5 = [0.5,0.7,0.8]$$
$$\gamma_6 = [0.65,0.85,0.95] \quad \gamma 7 = [0.8,1,1]$$

			γ_7	γ_6	γ_5 ct	γ_4 t	γ_3 at	γ_2 f	γ_1	*
			$\gamma_7 \circledR f$	$\gamma_6 \circledR f$	$\gamma_5 \circledR f$	$\gamma_4 \circledR f$	$\gamma_3 \circledR f$	$\gamma_2 \circledR f$	$\gamma_1 \circledR f$	
		$\gamma_7 \circledR at$	$\gamma_6 \circledR at$	$\gamma_5 \circledR at$	$\gamma_4 \circledR at$	$\gamma_3 \circledR at$	$\gamma_2 \circledR at$	$\gamma_1 \circledR at$		
	$\gamma_7 \circledR t$	$\gamma_6 \circledR t$	$\gamma_5 \circledR t$	$\gamma_4 \circledR t$	$\gamma_3 \circledR t$	$\gamma_2 \circledR t$	$\gamma_1 \circledR t$			
$\gamma_7 \circledR ct$	$\gamma_6 \circledR ct$	$\gamma_5 \circledR ct$	$\gamma_4 \circledR ct$	$\gamma_3 \circledR ct$	$\gamma_2 \circledR ct$	$\gamma_1 \circledR ct$				
δ_{10}	δ_9	δ_8	δ_7	δ_6	δ_5	δ_4	δ_3	δ_2	δ_1	

REPRESENTATION	LINGUISTIC APPROXIMATION
$\delta_1 = [0,0,0.2]$	f
$\delta_2 = [0.075,0.15,0.3]$	$ib[f,at]$
$\delta_3 = [0.1406,0.2969,0.4967]$	$b[at]$
$\delta_4 = [0.275,0.4,0.5688]$	at
$\delta_5 = [0.3375,0.4869,0.6438]$	$nt[at]$
$\delta_6 = [0.4063,0.5688,0.7125]$	$ib[at,t]$
$\delta_7 = [0.4781,0.6469,0.7656]$	$b[t]$
$\delta_8 = [0.5938,0.7938,0.8813]$	$ib[ct,t]$
$\delta_9 = [0.6865,0.8875,0.9625]$	$ib[ct,t]$
$\delta_{10} = [0.8,1,1]$	ct

Finally, the ordered string, $C\Diamond S\Diamond A$, associated with the disease *coronaropathy*, is gotten:

$$C\Diamond S\Diamond A = [a,b]^{ct}[c,d]^{ib[ct,t]}[e]^{b[at]}[f,g]^{f}$$

We note that the partition induced by the result, $C\Diamond S\Diamond A$, is the same as that induced by the symptom S, however, the linguistic labels of each class are different.

Moreover, the algebraic structure can be used to generate other classifications in the same population corresponding to another set of diseases. For instance, the ordered string $S\Diamond A$ could represent, in a simplified way, the disease "depression".

4. THE RELEVANCE PROBLEM

In the previous section we have shown how the algebraic structure can be used to generate different classifications in a universe of discourse. Now we are going to investigate the related problem concerning the weight of the attributes in the classification. In fact, consider the example of medical diagnosis introduced in Section 3. We have seen that a suitable string manipulation allows inferring how likely is the event that an individual suffers from a certain disease. However, there are symptoms whose presence is more relevant than others for the diagnosis. The question that arises is: what does it mean that a symptom is more (or less) relevant than another one?

Intuitively, the more *information* supplied by a symptom, the greater its relevance. However, the relevance of a symptom (and, in general, of an attribute) is not an absolute property, but is relative to a diagnosis (classification or interpretation). For instance, there are symptoms (attributes) whose presence is relevant to a kind of diagnosis (classification) but, in turn, it is completely *irrelevant* to others. In general, the natural way to measure the information supplied by an attribute in a classification requires that its capacity for *discriminating* the elements of the universe is taken into account.

Of course, the problem arises of characterizing precisely the concept of attribute-discriminating capacity. In order to answer this question, we use some ideas about the formalization of the notion of relevance as introduced in Nunez [10].

If we assume that the relevance of an attribute depends upon its discriminating capacity, then it is necessary to determine upon what this discriminating capacity depends. The relevance of an attribute depends not only on the percentage of objects that this attribute discriminates among, but moreover on the way it interacts with the other attributes. The basic idea is that the attribute A is more relevant than the attribute B if *it supplies more information*, that is, *if it requires less additional information than B* to discriminate among all the elements of the universe.

Thus, given a certain attribute, if the distribution among the classes of the elements of the universe corresponding to each value is less random, the attribute is potentially more useful as regards the classification.

In machine learning, several *heuristic* methods have been developed to evaluate the attributes in terms of their potential utility (that has been identified, incorrectly, with their relevance), almost all of them are based upon the classical information theory (e.g., [3], [10], [11], [12] [14]). However, they do not distinguish clearly between values discriminating among a lot of objects from those that just discriminate among a few. Here, we use a heuristic trick not having this problem [10]. It is defined as follows:

Definition 4.1. *The function μ that evaluates the potential relevance of an attribute*

$$A = a_n^{\alpha n} \dots a_2^{\alpha 2} a_1^{\alpha 1},$$

concerning a classification $C_1, C_2, \ldots, C_m$, *say from* $C\diamond S\diamond A$, *is defined as*

$$\mu(A) = \sum_{j=1,\ldots,n} \sum_{k=1,\ldots,m} (p_{jk} - p_j)^2$$

where $p_{jk} = \#[A^{-1}(\alpha_j) \cap C_k]$ and $p_j = \#[A^{-1}(\alpha_j)]/m$ and # denotes the cardinality of the set.

The function μ takes into account the distribution of objects with respect to the values of the attributes and, what is more, reflects the proportion of objects that are assigned to each class for each value.

In the following, we calculate the potential relevance, with respect to each classification, of the attributes that were considered in the example of Section 3 and we refer to their describing strings (3).

SYMPTOM CHOLESTEROL As regards this symptom, one has

$$C_1 = \{a,b\}, \quad C_2 = \{c,d\}, \quad C_3 = \{e\}, \quad C_4 = \{f,g\}.$$

Thus, the values concerning the labels are as follows:

label "*ct*":
$p_{11} = \#[A^{-1}(ct) \cap C_1] = 2$
$p_{12} = \#[A^{-1}(ct) \cap C_2] = 1$
$p_{13} = \#[A^{-1}(ct) \cap C_3] = 0$
$p_{14} = \#[A^{-1}(ct) \cap C_4] = 0$
and $p_1 = 3/4$

label "*t*":
$p_{21} = \#[A^{-1}(t) \cap C_1] = 0$
$p_{22} = \#[A^{-1}(t) \cap C_2] = 1$
$p_{23} = \#[A^{-1}(t) \cap C_3] = 0$
$p_{24} = \#[A^{-1}(t) \cap C_4] = 0$
and $p_2 = 1/4$

label "*at*":
$p_{31} = \#[A^{-1}(at) \cap C_1] = 0$
$p_{32} = \#[A^{-1}(at) \cap C_2] = 0$
$p_{33} = \#[A^{-1}(at) \cap C_3] = 0$
$p_{34} = \#[A^{-1}(at) \cap C_4] = 0$
and $p_3 = 0$

label "*f*":
$p_{41} = \#[A^{-1}(f) \cap C_1] = 0$
$p_{42} = \#[A^{-1}(f) \cap C_2] = 0$
$p_{43} = \#[A^{-1}(f) \cap C_3] = 1$
$p_{44} = \#[A^{-1}(f) \cap C_4] = 2$
and $p_4 = 3/4$

Thus, by applying the formula (4) one gets:

$$\begin{aligned}\mu(C) = {} & (2 - 3/4)^2 + (1 - 3/4)^2 + (0 - 3/4)^2 + (0 - 3/4)^2 \\ & +(0 - 1/4)^2 + (1 - 1/4)^2 + (0 - 1/4)^2 + (0 - 1/4)^2 \\ & +(0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 \\ & +(0 - 3/4)^2 + (0 - 3/4)^2 + (1 - 3/4)^2 + (2 - 3/4)^2 \\ = {} & 100/16\end{aligned}$$

Simple calculations yield, in a similar way, the values of μ for the other two symptoms.

SYMPTOM STRESS

$$\begin{aligned}\mu(S) &= (2-1/2)^2 + (0-1/2)^2 + (0-1/2)^2 + (0-1/2)^2 \\ &\quad + (0-1/2)^2 + (2-1/2)^2 + (0-1/2)^2 + (0-1/2)^2 \\ &\quad + (0-1/4)^2 + (0-1/4)^2 + (1-1/4)^2 + (0-1/4)^2 \\ &\quad + (0-1/2)^2 + (0-1/2)^2 + (0-1/2)^2 + (2-1/2)^2 \\ &= 156/16\end{aligned}$$

SYMPTOM ANXIETY

$$\begin{aligned}\mu(A) &= (2-1)^2 + (2-1)^2 + (0-1)^2 + (0-1)^2 \\ &\quad + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 \\ &\quad + (0-1/4)^2 + (0-1/4)^2 + (1-1/4)^2 + (1-1/4)^2 \\ &\quad + (0-2/4)^2 + (0-2/4)^2 + (0-2/4)^2 + (2-2/4)^2 \\ &= 124/16\end{aligned}$$

Thus, with respect to the disease *coronaropathy*, stress is more relevant than the symptom anxiety, and this is more relevant than the symptom cholesterol.

5. CONCLUDING REMARKS

The algebraic structure illustrated in this paper represents a classification and inference tool that could be used in several applications.

In particular, the example given in Section 3 can be used to generate several classifications in domains for which only qualitative information is available. In our opinion, the proposed classification schema can be successfully applied to those classification problems that are affected by inherently diffused features (pattern, size, color, quality, etc.); instances of this class of problems are plankton, sponges, archaeological remains. Thus, by manipulating the strings one could deduce how likely is the event that a specific element belongs to a certain class. In fact, one could first carry out a screening of the values to get only the values above or below a certain threshold. Of course, this opinion should be confirmed or refuted by appropriate field tests, similar to the above-mentioned one.

It is worth emphasizing that the proposed method offers a relevant efficiency because inherent parallelism of the structure permits generation of classifications in linear time.

In turn, the heuristic discussed in Section 4 can be used to recognize, in a simple way, the features whose presence are more relevant to the classifications. With a slight modification, this heuristic can also be used to evaluate the potential relevance of sets of attributes [10]. Moreover, it can be utilized to obtain decision trees by selecting the attribute having the greatest potential relevance and in such way the attributes "nought" and "don't care" can be discarded.

For example, if one refers to the case study shown in Section 3, the symptom selected during the first step of the process would be that concerning "stress" because it possesses the greatest potential relevance. In this case, the symptom plays a role of the uttermost importance regarding the classification associated with the ordered string $C\diamond S\diamond A$.

References

1. Buchanan, B. G., and Shortliffe, E. H., Rule-based Expert Systems, in *The MYCIN Experiments of Stanford Heuristic Programming Project*, Addison Wesley, Reading, MA, 1984.
2. Michalsky, R. S., Davis, J. H., Bisht, V. S., and Sincler, J. B., PLANT/ds: An expert consulting system for the diagnosis of soybean diseases, *Proc. Eur. Conf. AI.* Orsay, France, 1982.
3. Baim, P. W., A method for attribute selection in inductive learning systems, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 10(6), 888–896, 1988.
4. Clancey, W. J., The epistemology of a rule-based expert system: A framework for explanation. Dep. Comput. Sci., Stanford Univ., Rep. STAN-CS-81-896, 1981.
5. Gisolfi, A., A Prolog implementation of a fuzzy algebraic structure for the uncertainty modelling, *ISIJMA 90. Uncertainty Modeling and Analysis* (B. M. Ayyub Ed.), Maryland, USA, 1–6, 1990.
6. Gisolfi, A., An algebraic fuzzy structure to the approximate reasoning, *Fuzzy Sets and Systems*, 45, 37–43, 1992.
7. Zadeh, L. A., A fuzzy set theoretic interpretation of linguistic hedges. *J. Cybern.* 2(3), 4–34, 1972.
8. Zadeh, L. A., The concept of linguistic variables and its application to approximate reasoning, *Inf. Sci.* 8, 199–249, 1975.
9. Dubois, D., and Prade, H., Fuzzy real algebra: some results, *Fuzzy Sets and Systems*, 3, 327–348, 1978.
10. Nunez, G., Alvarado, M., Cortes, U., and Belanche, L., About the attribute relevance's nature, *Proc. TEC-COM 91: Computacion no convencional: Hacia los sistemas inteligentes*, Mexico D.F, 1991.

11. Schlimner, J. C, and Fisher, D., A case study of incremental concept induction, *Proc. 5th National Conf. AI*. Morgan Kaufmann, (Ed.), 496–501. 1986.

12. Lopez de Mantaras, R., A Distance-Based Attribute Selection Measure for Decision Tree Induction, *Machine Learning* 6, 81–92, 1991.

13. Quinlan, J. R., Discovering rules by induction from large collections of examples, in *Experts Systems in the Micro-electronics Age*. (D. Michie, Ed.), Edinburgh University Press, 1979.

14. Quinlan, J. R., Induction of decision trees, *Machine Learning* 1, 81–106, 1986.