International Journal of Computer and Information Sciences, Vol. 10, No. 2, 1981

Approximate Reasoning and Possibilistic Models in Classification

Ronald R. Yager[1]

Received July 1980; revised February 1981

We introduce some ideal from the theory of approximate reasoning and from possibility theory based on fuzzy sets. We shown how these ideas can form the basis for building classification models which enable one to use imprecise information in their construction.

KEY WORDS: Fuzzy sets; implication; classification; model building; approximate reasoning.

1. INTRODUCTION

The problem of classification objects, of which pattern recognition and multicriteria decision-making are special applications, is one which permeates much of the current technical literature.

Much of the information which can be of use in classification techniques may not be used because of its soft or imprecise nature. A significant improvement to classification procedures would be the ability to include all the available information whether hard or soft.

In a number of articles (concluding with Ref. 21) Zadeh has developed a theory of approximate reasoning based upon fuzzy subset theory (see Ref. 22 for a complete listing of Zadeh's work). This theory has the ability to handle both soft and hard data as well as various types of uncertainty. Many aspects of this development can be incorporated into a general classifier which has the ability to include all types of information which may be of use to a person trying to classify some objects. Sanchez[(8)] has used this methodology for medical diagnosis which is essentially a pattern recognition problem. Yager[(10),(16)] has used aspects of this theory in formulating solutions to multiple-criteria decision problems.

[1] Machine Intelligence Institute, Iona College, New Rochelle, New York.

0091-7036/81/0400-0141$03.00/0

In this manuscript, which is somewhat tutorial in nature, we shall use Zadeh's theory of approximate reasoning to supply a general framework for classification problems. This framework will be broad enough to include imprecise as well as precise granules of information which may be of assistance in the classification of objects.

We shall first begin with a discussion of aspects of fuzzy set theory and approximate reasoning which will be of use in the construction of our general classifier.

2. FUZZY SUBSETS, LINGUISTIC, VARIABLES AND POSSIBILITY THEORY

Assume U is a set of elements a fuzzy subset A is a subset of U in which the membership grades lie in the unit interval instead of the usual binary set $\{0, 1\}$. Thus, a fuzzy subset A can be characterized as a mapping

$$A: U \to [0, 1]$$

in which for each $u \in U$, $A(u)$ indicates the degree of membership of u in the subset A. The larger the value $A(u)$, the stronger the membership. A calculus has been developed by Zadeh for manipulating fuzzy subsets.[6] Assume A and B are two fuzzy subsets of U, then

(1) the intersection $C = A \cap B$ is also a fuzzy subset of U such that $C(u) = A(u) \wedge B(u)$, where $\wedge = \min$

(2) the union $D = A \cup B$ is also a fuzzy subset of U such that $D(u) = A(u) \vee B(u)$, where $V = \max$

(3) the complement A' is also a fuzzy subset of U where $A'(u) = 1 - A(u)$.

(4) A^n, where $n \geqslant 0$, is also a fuzzy subset of U where $A^n(u) = (A(u))^n$.

A significant use of fuzzy subsets is its ability to represent concepts in which there exists some imprecision with respect to its definition. For example, if U is a set of heights and A is the concept of tall, than fuzzy subsets allows us to represent this idea in a more natural manner than ordinary set theory.

A primary application of fuzzy subsets is in the representation of linguistic variables. Assume M is a variable, such as the gas mileage on a car. Let U be the set of values which M can assume. U is referred to as universe of discourse of M. In many instances, an exact or precise value for M cannot be obtained, but instead some imprecise or inexact value for M is available. For example, the miles per gallong for a particular type car may be stated as "high" or "low" or "about 20 miles per gallon." Using fuzzy subsets we can express this information in an analytic manner which will enable us to manipulate these imprecise values. In particular, if we know that the miles per gallon is to "about 20," we can represent this information by assigning the value F to M, where F is the fuzzy subset of U representing the value "about 20." More generally, if M is a variable taking values in the universe of discourse U, using linguistic values for M allows us to indicate either precise[2] or imprecise values for M by setting

$$M = F,$$

where F is a fuzzy subset of U.

Following a suggestion by Zadeh,[(20)] the statement that $M = F$ has the effect of acting as a restriction on the values which M can assume. In particular, the information that $M = F$ associates with the variable M is a possibility distribution over U, such that for each $u \in U$, $\Pi_M(u)$, the possibility that $M = u$, is equal to $F(u)$. In the simplest case the statement $M = u_i$ implies that the possibility of M assuming any value other than that of u_i is zero, while the possibility of assuming the value u_i is one. Thus, by using fuzzy subsets to represent linguistic values for variables we can convert information into a possibility distribution for that variable over its universe of discourse U.

The possibility distribution is a generalization of the concept of ease with which M can assume a particular value in U given the known information about M.

3. CYLINDRICAL EXTENSIONS AND FUZZY SUBSETS OVER DIFFERENT SETS

In the previous section we defined the operations of union and intersection of fuzzy subsets in the case when both subsets are defined over the same set. In this section, we shall extend these operations to the case when the fuzzy subsets are defined on different sets.

Definition. Assume U and V are two sets, let A be a fuzzy subset of U, the cylindrical extension of A to $U \times V$, denoted $\bar{A}$ is defined as the fuzzy subset of $U \times V$, such that

$$\bar{A}(u, v) = A(u).$$

[2] A precise value for M, for example, $M = 50$ where $50 \in U$ can be expressed as the fuzzy subset $\{1/50\}$.

Example. Assume $U = \{1, 2, 3\}$ and $V = \{a, b\}$, let

$$A = \left\{\frac{.5}{1}, \frac{.9}{2}, \frac{1}{3}\right\}$$

then the cylindrical extension of A to $U \times V$ is

$$\bar{A} = \left\{\frac{.5}{(1, \mathrm{a})}, \frac{.5}{(1, \mathrm{b})}, \frac{.9}{(2, \mathrm{a})}, \frac{.9}{(2, \mathrm{b})}, \frac{1}{(3, \mathrm{a})}, \frac{1}{(3, \mathrm{b})}\right\}$$

Again, assuming U and V are two sets, let A and B be fuzzy subsets of U and V, respectively, we define the intersection and union of A and B as follows.

(1) Intersection

$C = A \cap B$, where C is a fuzzy subset of $U \times V$

such that,

$C(u, v) = A(u) \wedge B(v)$.

Using the notation of cylindrical extensions

$C = \bar{A} \cap \bar{B}$ where

$\bar{A}$ and $\bar{B}$ are the cylindrical extensions of A and B to $U \times V$.

(2) Union

$D = A \cap B$, where D is a fuzzy subset of $U \times V$ such that,

$D(u, v) = A(u) \vee B(v)$.

Using the notation of cylindrical extension

$D = \bar{A} \cup \bar{B}$

The opposite of the operation of cylindrical extension is the concept of the projection of a fuzzy subset.

Definition. Assume G is a fuzzy subset of $U \times V$, the projection of G on U, denoted $\underset{\text{on } U}{\text{Proj}}\, G$, is a fuzzy subset of U defined as follows:

$$\underset{\text{on } U}{\text{Proj}}\, G(u) = \underset{v \in V}{\text{Sup}}\, G(u, v)$$

Example. $U = \{1, 2, 3\}$ $\quad V = \{a, b\}$

let

$$G = \left\{\frac{.7}{(1, a)}, \frac{.6}{(2, a)}, \frac{.9}{(3, a)}, \frac{.5}{(1, b)}, \frac{.8}{(2, b)}, \frac{.1}{(3, b)}\right\}$$

$$F = \underset{\text{on } u}{\text{Proj}} \quad G = \left\{\frac{.7}{1}, \frac{.8}{2}, \frac{.9}{3}\right\}$$

4. FUZZY LOGIC

The fuzzy logic developed by Zadeh[21] is the logic of approximate reasoning. Its constituents are a set of translation rules and a set of rules of inference. The translation rules consists of a set of procedures for modifying or forming composite propositions out of basic propositions. The rules of inference are procedures for making logical deductions from fuzzy propositions.

We shall first briefly define some of the translation rules that will be of interest to us.

Assume X, Y, and Z are variables taking values in the universe of discourses U, V, and W respectively. Let F, G, and H be fuzzy subsets of U, V, and W, respectively.

(1) The proposition "X is F," $(X = F)$, which assigns a linguistic value to X, translates into a possibility distribution Π_X for X on U, such that for each $u \in U$

$$\Pi_X(u) = F(u).$$

(2) The proposition

$$X \text{ is not } F,$$

which again assigns a linguistic value to X, translates into a possibility distribution for X of Π_X on U such that for each $u \in U$

$$\Pi_X(u) = 1 - F(u).$$

(3) The proposition

$$X \text{ is very } F,$$

which again assigns a linguistic value to X, translates according to Zadeh, into a possibility distribution for X of Π_X on U such that

$$\Pi_X(u) = F^2(u).$$

(4) The proposition

$$X \text{ is sort of } F,$$

translates into the possibility distribution defined by

$$\Pi_X(u) = F^{1/2}(u).$$

Among other uses these translation rules enable us to define our information in terms of certain primary concepts which can be defined by a user. In a sense, these primary concepts from a linguistic scale which the describer uses to measure his other linguistic values. In the case of a variable corresponding to age, the primary terms can be, for example, young, old, middle age, etc.

(5) The proposition

$$X \text{ is } F \text{ and } Y \text{ is } G$$

translates into the joint possibility distribution $\Pi_{X,Y}$ of the binary variable (X, Y) on $U \times V$ such that for each $u \in U$ and $v \in V$, $\Pi_{X,Y}(u, v) = F(u) \wedge G(v)$, when X and Y are possibilistic independent.[3] The operation can be easily extended to n variables, X_1 is F_1 and X_2 is F_2 and X_3 is $F_3 \cdots$ and Xn is Fn translates into

$$\Pi_{X_1, X_2, \ldots, X_n}(u_1, u_2, \ldots, u_n) = F_1(u_1) \wedge F_2(u_2) \wedge \cdots \wedge F_n(u_n)$$

(6) The proposition

$$X \text{ is } F \text{ or } Y \text{ is } G$$

translates into the joint possibility distribution $\Pi_{X,Y}$ of the binary variable (X, Y) on $U \times V$ such that for each $u \in U$ and $v \in V$

$$\Pi_{XY}(u, v) = F(u) \vee G(v).$$

This operation can also be easily extended to n variables.

[3] When X and Y are possibilitically dependent the marginal possibility $G(v)$ must be replaced by the conditional possibility of v given u, $G(v/u)$.

(7) The proposition

$$\text{If } X \text{ is } F \text{ than } Y \text{ is } G$$

translates into the conditional possibility distribution $\Pi_{Y/X}$ on the set $U \times V$ such that

(1) $\Pi_{Y/Z}(v/u) = (1 - F(u) + G(v)) \wedge 1.$

There are a number of alternative possible translations of this proposition, among those finding primary applications are:

(2) $\Pi_{Y/X}(v/u) = (1 - F(u)) \vee G(v)$

(3) $\Pi_{Y/X}(v/u) = (F(u) \wedge G(v)) \vee (1 - F(u))$

(4) $\Pi_{Y/X}(v/u) = 1 \quad \text{if} \quad G(v) \geqslant F(u)$

$= \dfrac{G(v)}{F(u)}$ otherwise

(5) $\Pi_{Y/X}(v/u) = 1 \quad \text{if} \quad G(v) \geqslant F(u)$

$= G(v)$ otherwise.

A number of authors have investigated the various properties of these differing interpretations of this proposition.[2,3,7,9]

Note 1. Combinations of the above propositions are of course possible. For example,

$(X$ is very F or Y is $G)$ and Z is H translates into:

$$\Pi_{X,Y,Z}(u, v, w) = (F^2(u) \vee G(v)) \wedge H(z)$$

The rules of fuzzy inference govern the deductions which can be made in approximate reasoning.[21]

(I). *The Entailment Principle.* Assume F and G are fuzzy subsets of U such that for all $u \in U$, $F(u) \leqslant G(u)$, that is $F \subset G$, than the entailment principle states that

$$\Pi_X = F \Rightarrow \Pi_X = G.$$

The entailment principle is a generalization of the ability to go from a very specific statement of fact to a less specific statement of fact while keeping within the truth. For example, the knowledge that X equals 7 implies that X is less than 10 is also true.

The entailment principle is also related to the idea that possibilistic information is a reflection of putting restrictions on our possibilities. Thus, by using the entailment principle, we are in a sense not using all our information.

(II). *The Projection Principle.* Let $X_1, X_2, \ldots, X_n$ be a set of variables whose universe of discourse are U_2, $U_2, \ldots, U_n$, respectively. Assume P is a fuzzy proposition which implies that:

$$\Pi_{X_1, X_2, \ldots, X_n} = F,$$

where F is a fuzzy subset of $U_1 \times U_2 \times \cdots \times U_n = U$.

Let $S = (i_1, i_2, i_3, \ldots, i_k)$ be a subsequence of $(1, 2, \ldots, n)$, let us denote $X = (X_1, X_2, \ldots, X_n)$ and $(X_{i_1}, X_{i_2}, \ldots, X_{i_k}) = X_s$.

We shall denote the projection of F on $U_s = U_{i_1} \times U_{i_2} \times \cdots \times U_{i_k}$, as $\text{Proj}_{u_s} F$. $\text{Proj}_{u_s} F$ is a fuzzy subset of U_s such that for any

$$(u_{i_1}, u_{i_2}, \ldots, u_{i_k}) \in U_s$$

$\text{Proj}_{u_s} F(u_{i_1}, u_{i_2}, \ldots, u_{i_k}) = \text{SUP } F$
over all points in U such that in the subsequence S the elements have the argument $(u_{i_1}, u_{i_2}, \ldots, u_{i_k})$

The projection principle asserts that

$$\Pi_X = F \Rightarrow \Pi_{X_s} = \text{Proj}_{u_s} F.$$

Note. The projection principle is in reality a special application of the entailment principle.

Consider the variables X and Y defined on the base sets U and V. Let

$$\Pi_{X,Y} = F$$

where F is a fuzzy subset of $U \times V$. Consider the fuzzy subset G of $U \times V$ where

$$G(u, v) = \underset{v \in V}{\text{Max}} \, F(u, v).$$

we observe

(1) $F \subset G$, thus the entailment principle implies $\Pi_{X,Y} = G$.

(2) for a given value of u, $G(u, v)$ is fixed. That is, G is just solely determined by its X value, Y is a dummy variable in G. Thus, we can say that if $\Pi_{X,Y} = G$ this really implies a unique distribution over U s.t. $\Pi_x(u) = G(u, v)$ for any v.

(3) $G(u, v)$ is the projection on U.

Before proceeding with our rules for inference in approximate reasoning we must express some results from the calculus of possibility theory.

Assume X and Y are two variables taking values in the sets U and V. Let $\Pi_{Y/X}$ be the conditional possibility distribution associated with these variables, i.e., $\Pi_{Y/X}(v/u)$ indicates the possibility that $Y=v$ given $X=u$, thus it is defined for all elements in the set $U \times V$. Let Π_X be a marginal possibility distribution associated with the variable X, i.e., $\Pi_X(u)$ indicates the possibility that $X=u$, thus Π_x is defined for all elements in U. Furthermore, let $\Pi_{X,Y}$ be the joint possibility distribution associated with X and Y, i.e. $\Pi_{X,Y}(u, v)$ indicates the possibility that $X=u$ and $Y=v$, thus $\Pi_{X,Y}$ is defined for all elements in $U \times V$. As suggested by Zadeh[20] and Hisdale[4]

$$\Pi_{X,Y}(u, v) = \Pi_X(u) \wedge \Pi_{Y/X}(v/u).$$

More generally this can be expressed as

$$\Pi_{X,Y} = \Pi_{Y/X} \cap \bar{\Pi}_X$$

where $\bar{\Pi}_X$ is the cylindrical extention of Π_X to $U \times V$.

A special law of inference in fuzzy logic is the rule of fuzzy compositional inference,[18] which is the application of the above relationship followed by the application of the projection principle.

Assume we have two propositions P_1 and P_2, such that P_1 translates into a possibility distribution on the set U associated with the variable X, i.e. $X=g$ where g is a linguistic value representable as a fuzzy set G of U and therefore $\Pi_X(u)=G(u)$. Assume that P_2 translates into a possibility distribution on $U \times V$ associated with the conditional variable Y/X, i.e., if $X=h$ then $Y=k$, where h and k are linguistic values representable as the fuzzy subsets H and K over U and V, respectively, from our previous discussion of translation rules a choice of interpretation for implication leads to $\Pi_{Y/X}(v/u)=D(u, v)$, where for example $D(u, v)=(1-H(u)) \vee K(v)$. From the previous relationship we can infer that

$$\Pi_{X,Y}(u, v) = \Pi_X(u) \wedge \Pi_{Y/X}(v/u)$$

or

$$\Pi_{X,Y} = \bar{\Pi}_X \cap \Pi_{Y/X}.$$

The projection principle allows us to infer a marginal possibility distribution $\Pi_Y(v)$ of Y as

$$\Pi_Y(v) = \operatorname*{Max}_{u \in U} \Pi_{X,Y}(u, v).$$

We note we can then retranslate $\Pi_Y(v)$ into a fuzzy subset F over V such that $F(v) = \Pi_F(v)$, where F is representative of some linguistic value f.

Symbolically we can represent the law of compositional inference as

$P_3 = P_1 \circ P_2$

or

$$Y = F = G \circ D$$

where

$$F(v) = \Pi_Y(v) = \underset{u \in U}{\mathrm{Max}}[G(u) \wedge D(u, v] = \underset{u \in U}{\mathrm{Max}}[\Pi_X(u) \wedge \Pi_{Y/X}(v/u)].$$

One further rule of inference must be supplied. Assume P_1 and P_2 are two propositions or statements such that they both supply some information about the same variable, i.e.,

$$P_1: X \text{ is } A$$

$$P_2: X \text{ is } B$$

where A and B are two fuzzy subsets of U, then these two statements jointly translate into the possibility distribution

$$\Pi_X(u) = A(u) \wedge B(u).$$

A special case of this occurs when

$$P_1: \text{If } X \text{ is } A_1 \text{ then } Y = C_1$$

$$P_2: \text{If } X \text{ is } A_2 \text{ then } Y = C_2$$

then this translates into $\Pi_{Y/X}(v/u) = D_1(u, v) \wedge D_2(u, v)$ where D_1 and D_2 are the translations of each of the statements.

5. GENERAL CLASSIFICATION PROBLEM

The general classification problem, which we shall describe will be shown to be amenable to solution techniques based upon fuzzy set theory.

Assume we have a set of categories $W = \{w_1, w_2, ..., w_n\}$. Let X be $X_1, X_2, ..., X_p$ be a class of properties or characteristics useful in describing, defining or distinguishing between these categories. Assume that these characteristics are each measurable on the respective universes of discourse $U_1, U_2, ..., U_p$. Furthermore, assume we have a set of rules, R, which based upon the characteristics $X_1, X_2,..., X_p$ describes the possibility of membership in the categories of W. The general classification problem involves the classification of some object, A, with respect to the categories of W, based upon the values of the object A with respect to the characteristics $X_1, X_2,..., X_p$.

In medical diagnosis, for example,[8] W could be a set of diseases under consideration, the X's would be a class of characteristics associated with human well being, such as, temperature, blood pressure, blood count, etc., the sets of U 's would be scales on which these characteristics are measured, R would be the information describing the various elements of W in terms of values for the X's, that is, R is the experts information, in this case the doctors and finally A, would be a patient described in terms of his readings for the characteristics $X_1, X_2,..., X_p$. In this case the problem would be to determine the possibility of A having the various diseases based upon the readings of the X_i's for A.

As we shall shortly see an advantage of the fuzzy set approach will be our ability to represent a vast variety of forms of expert information in the construction of R.

Assume V is some unobservable variable, which can take values in some set W. Let $X_1, X_2,..., X_p$ be some set of observable values which can be related though R to our unobservable variable V. Then using this general classification approach, given a set of readings for the X's, we could determine possible values for the unobservable variable V.

The general classification problem can be applied to problems involved in drilling for oil. In this case W would be a set of possible conditions a site could have with respect to the availability of underground oil. The X's would be a set of measurable characteristics which are related through R to the availability of oil. Then given the readings for a site in terms of its values for X we could then determine the possible structure of our site.

Control systems can also be subsumed under this general problem (see Mamdani[6]). Decision problems can also be shown to be describable under this model.

6. AN EXAMPLE OF THE FORM OF R

The heart of our classification procedure consists of the set of rules R which relates the variable V to be classified to the observed variables $X_1, X_2,..., X_p$. The rules which compose R are a reflection of the information available about the relationship between the X's and V. In many cases this information takes the form of imprecise observations made by an expert. The structure which we shall present affords us the opportunity to include this type of information. It allows us to handle imprecise or precise valuations for the objects to be classified.

In the case we shall study in this section we shall assume R is made up of a collection of descriptors each of which is a granule of information[5,19] which describes the variable to be classified in terms of the characteristic variables, the X's. Mathematically these granules are mappings from fuzzy subsets of $U_1 \times U_2 \times \cdots \times U_p$ into fuzzy subsets of W.

As an illustrative example we shall consider the problem of an administrator whose task is to categorize candidates as to their successful performance on a job. Let the categories considered W, be

$$W = (\text{highly successful, marginally successful, failure}).$$

Assume the characteristics which the administrator considers as significant in the performance of the job to be college average and age.

Based upon his experience our administrator can make some imprecise inferential observations on the success of a candidate. These observations are what we shall call information granules.

An information granule in this case could be d_1: If a person has a high college average and is not old he will be highly successful. Using Zadeh's theory of approximate reasoning we can express this information as a fuzzy relationship on the appropriate sets.

Let X_1 be the variable corresponding to college average. For simplicity, we can assume it is measured on the set,

$$U_1 = \{0, 1, 2, 3, 4\}.$$

Let X_2 be the variable corresponding to the age of the candidate, let it be measured on the universe of discourse

$$U_2 = \{20, 25, 30, 35, 40\}.$$

We can express our descriptor or granule in propositional form as

$$d_1: \text{If } X_1 = \text{high and } X_2 = \text{not old then } V = \text{highly successful.}$$

We can then express each of the linguistic values in the above descriptor as a fuzzy subset of the appropriate base set.

For example,

$$G_1 = \text{high college average} = \left\{\frac{0}{0}, \frac{0}{1}, \frac{0}{2}, \frac{.5}{3}, \frac{1}{4}\right\}$$

and if

$$\text{old} = \left\{\frac{0}{20}, \frac{0}{25}, \frac{.3}{30}, \frac{.7}{35}, \frac{1}{40}\right\} = F$$

then

$$F' = \text{not old} = \left\{\frac{1}{20}, \frac{1}{25}, \frac{.7}{30}, \frac{.3}{35}, \frac{0}{40}\right\}$$

and finally

$$K_1 = \text{very successful} = \left\{\frac{1}{\text{very successful}}, \frac{0}{\text{marginal}}, \frac{0}{\text{failure}}\right\}.$$

Let $X = \{X_1, X_2\}$ and $U = U_1 \times U_2$ then $X_1 = \text{high}$ and $X_2 = \text{not old}$ can be expressed as a fuzzy subset M_1 of U such that

$$M_1(x, y) = G(x) \wedge F'(y) \quad \text{for } x \in U_1 \text{ and } y \in U_2.$$

Thus our descriptor becomes, "if $X = M_1$ then $V = K_1$"

In our case we get as the possibility distribution for M_1: $(u_i \in U)$. See Table I.

Selecting an appropriate definition for the "if–then" translation we can express our descriptor as a possibility distribution S_1 over $U \times W$.

Table I. Possibility Distribution for M_1

$(0, 20) = u_1$	$M_1(u_1) = 0$	$(1, 20) = u_6$	$M_1(u_6) = 0$
$(0, 25) = u_2$	$M_1(u_2) = 0$	$(1, 25) = u_7$	$M_1(u_7) = 0$
$(0, 30 = u_3$	$M_1(u_3) = 0$	$(1, 30) = u_8$	$M_1(u_8) = 0$
$(0, 35) = u_4$	$M_1(u_4) = 0$	$(1, 35) = u_9$	$M_1(u_{10}) = 0$
$(0, 40) = u_5$	$M_1(u_5) = 0$	$(1, 40) = u_{10}$	$M_1(u_{10}) = 0$
$(2, 20) = u_{11}$	$M_1(u_{11}) = 0$	$(2, 20) = u_{16}$	$M_1(u_{17}) = .5$
$(2, 25) = u_{12}$	$M_1(u_{12}) = 0$	$(3, 25) = u_{17}$	$M_1(u_{17}) = .5$
$(2, 30) = u_{13}$	$M_1(u_{13}) = 0$	$(3, 30) = u_{18}$	$M_1(u_{18}) = .5$
$(2, 35) = u_{14}$	$M_1(u_{14}) = 0$	$(3, 35) = u_{19}$	$M_1(u_{19}) = .3$
$(2, 40) = u_{15}$	$M_1(u_{15}) = 0$	$(3, 40) = u_{20}$	$M_1(u_{20}) = 0$
$(4, 20) = u_{21}$	$M_1(u_{21}) = 1$		
$(4, 25) = u_{22}$	$M_1(u_{22}) = 1$		
$(4, 30) = u_{23}$	$M_1(u_{23}) = .7$		
$(4, 35) = u_{24}$	$M_1(u_{24}) = .3$		
$(4, 40) = u_{25}$	$M_1(u_{25}) = 0$		

Table II. Possibility Distribution for S_1

	W_1 = Very successful	W_2 = Marginal	W_3 = Failure
u_1	1	1	1
u_2	1	1	1
u_3	1	1	1
u_4	1	1	1
u_5	1	1	1
u_6	1	1	1
u_7	1	1	1
u_8	1	1	1
u_9	1	1	1
u_9	1	1	1
u_{10}	1	1	1
u_{11}	1	1	1
u_{12}	1	1	1
u_{13}	1	1	1
u_{14}	1	1	1
u_{15}	1	1	1
u_{16}	1	.5	.5
u_{17}	1	.5	.5
u_{18}	1	.5	.5
u_{19}	1	.7	.7
u_{20}	1	1	1
u_{21}	1	0	0
u_{22}	1	0	0
u_{23}	1	.3	.3
u_{24}	1	.7	.7
u_{25}	1	1	1

"If X is M_1 then V is K_1 translates into $\Pi_{X,V} = S_1$, where S_1 is a fuzzy subset or possibility distribution of $U \times W$ such that

$$S_1(w/u) = 1 \wedge (1 - M_1(u) + K_1(w))$$

For possibility distribution for S_1 see Table II.

Thus we have shown our descriptor can be expressed as a possibility distribution over $U \times W = U_1 \times U_2 \times W$.

Assume our expert has a second distinct granule of information, d_2, denoting some further information with respect to the relationship between our characteristic variables and our variable to be classified.

For example

$$d_2: \text{If } X_2 = \text{very old then } V = \text{failure}.$$

First we note that a descriptor of the form,

If $X_2 = F$ then $V = K$ is equivalent to

If $X_1 = U_1$ and $X_2 = F$ the $V = K$.

Following the same procedure as above we translate this descriptor into a fuzzy subset S_2 over $U \times W$ shown in Table III.

Our classifier R can now be considered as being made up on the two propositions

$$R = d_1 \text{ and } d_2.$$

Thus

$$R = S_1 \cap S_2$$

Table III. Fuzzy Subset S_2 over $U \times W$

	S_2		
	w_1	w_2	w_3
u_1	1	1	1
u_2	1	1	1
u_3	.91	.91	1
u_4	.51	.51	1
u_5	0	0	1
u_6	1	1	1
u_7	1	1	1
u_8	.91	.91	1
u_9	.51	.51	1
u_{10}	0	0	1
u_{11}	1	1	1
u_{12}	1	1	1
u_{13}	.91	.91	1
u_{14}	.51	.51	1
u_{15}	0	0	1
u_{16}	1	1	1
u_{17}	1	1	1
u_{18}	.51	.51	1
u_{20}	0	0	1
u_{21}	1	1	1
u_{22}	1	1	1
u_{23}	.91	.91	1
u_{24}	.51	.51	1
u_{25}	0	0	1

where using the intersection of two fuzzy sets we get

$$R(w/u) = \text{Min}[S_1(w/u), S_2(w/u)].$$

If we had more descriptors then we would just form the intersection of all these descriptors. Thus we see that R is a intersection of fuzzy subset each reflecting the information supplied by a granule or descriptor.

For the case we are studying we get R as shown in Table IV.

We can now generalize the procedure for the formulation of the relationship R. Assume we have n descriptors $d_1, d_2, d_3, \ldots, d_n$, p variables $X_1, X_2, \ldots, X_p$ each measured on $U_1, U_2, \ldots, U_p$ respectively, where a descriptor is of the form

$$di\text{: If } (X_1 = Ai_1 \text{ and } X_2 = Ai_2 \cdots \text{ and } X_p = A_{ip}) \text{ then } V = K_i.$$

Table IV. ***R* as an Intersection of Fuzzy Subset**

	R		
	w_1	w_2	w_3
u_1	1	1	1
u_2	1	1	1
u_3	.91	.91	1
u_4	.51	.51	1
u_5	0	0	1
u_6	1	1	1
u_7	1	1	1
u_8	.91	.91	1
u_9	.51	.51	1
u_{10}	0	0	1
u_{11}	1	1	1
u_{12}	1	1	1
u_{13}	.91	.91	1
u_{14}	.51	.51	1
u_{15}	0	0	1
u_{16}	1	.5	.5
u_{17}	1	.5	.5
u_{18}	.91	.5	.3
u_{19}	.51	.51	.7
u_{20}	0	0	1
u_{21}	1	0	0
u_{22}	1	0	0
u_{23}	.91	.3	.3
u_{24}	.51	.51	.7
u_{25}	0	0	1

The A_{i_j}'s are fuzzy subsets of the U_j's and K_i is a fuzzy subset of W. Each descriptor reflects into a fuzzy subset S_i on the variable $(W_1, X_2, ..., X_p, V)$ over the set $U \times W$ $(U = U_1 \times U_2 \times \cdots \times U_p)$ such that $S_i(y_i, y_2, y_3, ..., y_p, w) = 1 \wedge ((1 - M_i(y_1, y_2, ..., y_p) + K_i(w))$ for each $y_i \in U_i$ and $w \in W$, where

$$M_i(y_1, y_2, ..., y_p) = \underset{j=1,2,...,p}{\text{Min}} [A_{i_j}(y_j)].$$

Finally

$$R = S_1 \cap S_2 \cap \cdots \cap S_n$$

where R is a fuzzy subset of $U \times W$ where

$$R(w/y) = \underset{i=1,2,...,n}{\text{Min}} [Si(w/y)]$$

for each $y \in U$ and $w \in W$.

7. CLASSIFYING OBJECTS

Having developed a classifier relationship R the next problem involves the classification of a particular object. In our general case an arbitrary object which is to be classified is characterized by its values for the X's. Assume

$$\begin{aligned} X_1 &= A_1^* \\ X_2 &= A_2^* \\ &\vdots \\ X_p &= A_p^*, \end{aligned}$$

where A_i^* is a fuzzy subset of U_i indicating the value of X_i for our object.

Thus our object is defined by the *Pary* variable $X = (X_1, X_2, ..., X_p)$, where $X = M^*$. M^* is a fuzzy subset of $U_1 \times U_2 \times \cdots \times U_p$ defined as $M^* = \bar{A}_1^* \cap \bar{A}_2^* \cap \bar{A}_3^* \cap \bar{A}_p^*$ where

$$M^*(y_1, y_2, y_p) = \underset{i=1,...,p}{\text{Min}} [A_i(y_i)]$$

for $y_i \in U_i$. Again denoting $U = U_1 \times U_2 \times \cdots \times U_p$ we have the following information:

$$P_1 : \Pi_X = M^* \qquad \text{and} \qquad P_2 : \Pi_{V/X} = R$$

where M^* is a fuzzy subset of U and R is a fuzzy subset of $U \times W$.

We can now apply the rule of fuzzy compositional inference to get V^* the possibility distribution associated with our object:

$$V^* = K^*,$$

where K^* is a fuzzy of W s.t.

$$K^*(w) = \operatorname*{Max}_{u \in U}(M^*(u) \wedge R(w/u))$$

We interpret K^* as a the possibility distribution on the variable V over the set W. This distribution associates with every $w \in W$ the possibility of it being the classification of the object being studied.

For the example we were studying in the previous section let us assume

$$X_1^* = 3 = \left\{\frac{0}{0}, \frac{0}{1}, \frac{0}{2}, \frac{1}{3}, \frac{0}{4}\right\}$$

$$X_2^* = 30 = \left\{\frac{0}{20}, \frac{0}{25}, \frac{1}{30}, \frac{0}{35}, \frac{0}{40}\right\}$$

Thus we can calculate M^* as

$$\begin{aligned} &M^*(3, 30) = M^*(u_{18}) = 1 \\ &M^*(ui) = 0 \qquad i \neq 18. \end{aligned} \qquad \text{i.e.,} \quad M^* = \left\{\frac{1}{(3, 30)}\right\}.$$

Applying this set to our fuzzy classifier R we obtain the following possibility distribution for our object:

$$K^*(w_1) = .91$$

$$K^*(w_2) = .5$$

$$K^*(w_3) = .5.$$

Thus, the possibility our candidate will be very successful is .91, while the possibility of him being marginally successful or a failure is .5.

The output of our fuzzy classifier is a possibility distribution Π over the set W where for any $w \in W$, $\Pi(w)$ indicates the possibility of the classified object being in category w. Assume L is a subset of W, for some purposes we may be interested in expressing the possibility of our object comming from the set L. The calculus of possibility theory developed by Zadeh[20] suggests a manner of doing this. In particular, if σ indicates the object to be classified and if Π_σ is the possibility distribution over W obtained from our classifier then

$$\text{Poss}\{\sigma \in L\} = \operatorname*{Max}_{w \in L}[\Pi_\sigma(w)].$$

More generally, if L is a fuzzy subset of W, with membership function $L(w)$, then

$$\text{Poss}\{\sigma \in L\} = \underset{w \in W}{\text{Max}}[\Pi_\sigma(w) \wedge L(w)].$$

For other purposes we may be interested in determining the certainty with which we can say $\sigma \in L$, where L may be a fuzzy subset of W. Zadeh[5] suggests a means of measuring this certainty,

$$\text{CERT}\{\sigma \in L\} = \underset{W}{\text{INF}}[(1 - \Pi_\sigma(w)) \vee L(\omega)]$$

For the special case when $L = \{w\cdot\}$ we can obtain the certainty that σ is w_i as

$$\text{CERT}\{\sigma \text{ is } w_i\} = \underset{w \in W - w_i}{\text{INF}}[1 - \Pi_\sigma(w)] = 1 - \underset{w \in W - w_i}{\text{Max}}[\Pi_Q(w)].$$

Thus in our example the certainty the candidate will be very successful is .5.

8. CONSIDERATIONS IN THE CONSTRUCTION OF *R*

In constructing the classifier R there are a number of factors to be considered, particularly with respect to the determination of the translation rule for the "if–then" type proposition. We noted in an earlier section that there are a number of possible ways of translating "if–then." Among the factors useful in determinning which rule to use to translate this type of proposition are scale information, degree of restriction, implicit meaning, default evaluation, and ease of computation. In this section we shall investigate these considerations.

The construction of a classifier of the type we have described requires the user to supply information on the membership values for the linguistic variables which make up the information granules which compose our classifier R. We have tacitly assumed that these pieces of information are supplied from an absolute scale having values in the unit interval. In many cases it may be difficult to get the expert to supply information in this detail. We shall investigate the possibilities of supplying information in less detail.

We note that $R = \bigcap_{i=1}^{k} Di$, where Di is the translation of the i^{th} descriptor. Furthermore, if B is the combined characteristic of the object than

$$V = A = B \circ R$$

where $A(w) = \text{Max}_{u \in u}[B(u) \wedge R(w/u)]$ is the possibility distribution of the classification of the object.

Fact 1—if B and R are drawn from the same ordinal scale than we can perform the operations necessary to obtain A.

Thus the compositional inference operation just requires an ordinal scale (actually all we need is a lattice).

Furthermore, since

$$R = \bigcap_{i=1}^{k} D_i$$

where $R(w/u) = \text{Min}\, D_i(w/u)$ we can conclude:

Fact 2—If each of the Di have membership grades drawn from the same ordinal scale we can perform the operations necessary to calculate R from these granules.

Thus the construction of R from the Di's just requires an ordinal scale (actually again all we need is lattice).

We shall now look into the scale requirements necessary for the formulation of the various definitions of implication from the basic constituents.

Assuming our implication statement is

"if $X = F$ then $V = G$,"

we see the following:

Fact 3—if F and G have membership grades drawn from the same ordinal scale with a maximum element, we can perform the operations necessary to calculate the implication 5, defined by

$$1.\quad D(w/u) = 1 \qquad \text{if} \quad G(w) \geqslant F(w)$$
$$= G(w) \qquad \text{if} \quad F(u) > G(w).$$

Fact 4—If F and G have membership grades drawn from the same scale which is ordinal and has a negation defined on it we can perform the necessary operations to calculate implications 2 and 3 defined by

2. $D \times F' \cup G$

   $D(w/u) = \text{Max}[1 - F(u), G(w)]$

3. $D = (F \cap G) \cup F'$

   $D(w/u) = \text{Max}([F(u) \wedge G(w)], 1 - F(u)])$

Fact 5—implications 4 and 1 require a ratio and absolute scale respectively to perform the necessary operations.

Fact 6—the construction of the m-ary variable $X=(X_1=F_1)$ and $(X_2=F_2)\cdots$ and $(X_n=F_n)$ just requires the membership grades of each of the F's to be drawn from the same ordinal scale.

Based upon these observations we can conclude: (i) If we use implication 5 we can construct and apply our classifier using a finite linear ordered scale to measure the membership grades for our basic elements, (ii) for implications 2 and 3 we need a finite a finite linear ordered scale with negation defined on it; (iii) for implications 4 we need a ratio scale (iv). For implication 1 we need an absolute scale.

It should be noted that while it is easier to obtain an ordinal scale than a ratio scale, and still more difficult to obtain an absolute scale, having an absolute scale enables us to perform linguistic hedges such as "very" to modify the values of our variables.

In addition various other procedures we may want to perform on our model are more readily performable if our information is absolute.

Unless otherwise specified we shall assume we have our basic information drawn from an absolute scale.

We note that the effect of an "if–then" granule is to supply some restriction on the possible values of the consequent based upon the value of the antecedent. In making an "if–then" statement there is some implicit idea of tightness of connection between consequent and antecedent.(14) The following relationships have been shown to hold between the various forms of implication.(2,14)

Theorem. Assume A and B are two fuzzy subsets of U and W, respectively, consider the fuzzy implication, "if $X=A$ then $Y=B$," let Ri be the resulting possibility distribution of this proposition via translation rule i, then

1. $R_1 \supset R_2 \supset R_3$
2. $R_4 \supset R_5$
3. $R_1 \supset R_5$

The effect of this theorem is that R_1 is a least-restrictive type of implication. Thus, in situations in which the expert is not very strong in the statement of his descriptor we may want to use R_1. On the other hand if the implication is meant to infer a very tight connection between antecedent and consequent we may want to use R_3. A fuller discussion of these considerations can be found in Ref. 14. The main emphasis of that discussion is that the selection of the translation rule must take into consideration the tightness between antecedent and consequent implied by the expert.

Another factor of significance in the construction of the classifier relationship R, from the individual descriptors, di, is the evaluation of the consequent for default of the antecedent. That is, given a descriptor if $X = A$ then $Y = B$, what do we mean to happen if $X = \text{not } A$.

Theorem. The translation of the proposition

$$\text{"if } X = A \text{ than } Y = B\text{"}$$

is equivalent to the translation of

$$\text{if } X = A \text{ then } Y = B \text{ and if } X = \text{not } A \text{ than } Y = W,$$

where A and B are fuzzy subsets of U and W, for all five of our translation rules.

Proof:

1. Rule 1—if A then $B \rightarrow R(\omega/u) = 1 \wedge (1 - A(u) + B(w))$

$$\text{if } \bar{A} \text{ then } W \rightarrow R^*(w/u) = 1 \wedge (1 + A(u)) = 1$$

$R(w/u)$ and $R^*(w/u) = R(w/u) \wedge 1 = R(w/u)$

2. Rule 2—if A then $B \rightarrow R(\omega/u) = (1 - A(u)) \vee B(w)$

$$\text{if } \bar{A} \text{ then } W \rightarrow R^*(\omega/u) = (A(u) + 1) \wedge 1 = 1$$

$R(w/u)$ and $R^*(w/u) = R(w/u) \wedge 1 = R(\omega/u)$

The cases for rules 3, 4, and 5 can be similarly shown.

In some cases we may want to overide this default consideration by explicitly inserting the evaluation of Y given not A for X. Thus in the construction of R we must give consideration to what effect we want if not A is the value of X.

Baldwin[2] raises some questions on the computational considerations associated with the compositional inference.

Consider an information granule "if $X_1 = A_1$ and $X_2 = A_2 \cdots$ and $X_p = A_p$ then $V = B$", where $A_1, A_2, \ldots, A_p$ are fuzzy subsets of $U_1, U_2, \ldots, U_p$, respectively, and B is a fuzzy subset of W. If $n_1, n_2, \ldots, n_p$ are the respective cardinalities of $U_1, U_2, \ldots, U_n$, then $U = U_1 \times U_2 \times U_p$ is a set of cardinality $n_1 \times n_2 \times \cdots \times n_n = n$. In many cases n could get very large and cause computational problems. Baldwin[2] suggests some interesting alternatives in trying to solve this problem.

9. ANOTHER METHOD OF CONSTRUCTING R

A different method of constructing R has been suggested by Mamdani[6] and studied by Baldwin.[2]

In this method, if A_i and B_i are fuzzy subsets of U and W, respectively, then the descriptors

$$d_1\colon \text{ if } X=A_1 \text{ then } Y=B_1$$
$$d_2\colon \text{ if } X=A_2 \text{ then } Y=B_2$$
$$d_n \quad \text{if } X=A_n \text{ then } Y=B_n$$

are combined to form the classification relationship R by

$$R=\bigcup_{i=1}^{n} Ai\times Bi \qquad \text{where} \quad [A_i\times B_i]=\bar{A}_i\cap\bar{B}_i.$$

We shall call the translation of "if $X=A$ then $Y=B$" via this rule Mamdani's implication.

In many cases of interest to us the antecedent of the descriptor di, $X=A_i$, is represented by the conjunction of the form $X_1=A_{i1}$ and $X_2=A_{i2}\cdots$ and $X_k=A_{ik}$ where Aij is a fuzzy subset of U_j, then $X=[X_1,X_2,\ldots,X_k]$ and $A_i=A_{i1}\cap A_{i2}\cap\cdots\cap A_{ik}$ is a fuzzy subset of $U=U_1\times U_2\times U_k$. Baldwin[2] has shown certain practical computational advantages of Mamdani's approach to the construction of R.

Theorem. (From Baldwin[2]). Assume R is constructed as above, assume an object has readings, $X_1=P_1$, $X_2=P_2\cdots X_k=P_k$, where P_i is a fuzzy subset of U_i, let $P=\bar{P}_1\cap\bar{P}_2\cap\cdots\cap\bar{P}_n$ then the fuzzy compositional inference

$$Q=P\circ R$$

is equivalent to

$$Q=\bigcup_{i=1,2,\ldots,n}[\cap(P_j\circ D_{ij})]$$

where

$$D_{ij}=A_{ij}\times B_i.$$

The computational advantages shown by Baldwin raises the question under what conditions is Mamdani's approach to the construction of R valid in the light of Zadeh's rules of inference in approximate reasoning.

Consider again the set of n descriptors $d_1, d_2, \ldots, d_n$, each of which is an implication of the form" if $X = A_i$ the $Y = B_i$," after translating each of these propositions into a possibility distribution R_i, using one of the five rules for translation suggested by Zadeh, we can obtain R as

$$R = \bigcap_{i=1}^{n} R_i$$

The question of interest is under what consideration is Mamdani's formulation

$$R = \bigcap_{i=1}^{n} A_i \times B_i$$

a good approximation to the above. In order to study this problem we shall use as our translation rule for

"if $X = A$ then $Y = B$"

the rule suggested by Zadeh,[18] that the implication is translated as $(A \times B) \cup A' = (\bar{A} \cap \bar{B}) \cup A'$, where the overbar indicates the cyclindrical extension and prime the negation.

We shall consider first the case when we have three descriptors than generalize.

Assume A_1, A_2, A_3 are fuzzy subsets of U and B_1, B_2, B_3 are fuzzy subsets of W and let A'_1, A'_2, A'_3 be the negation of A_1, A_3, and A_3, respectively.

Consider three descriptor

$$d_1\colon \text{ if } X = A_1 \text{ then } Y = B_1$$
$$d_2\colon \text{ if } X = A_2 \text{ then } Y = B_2$$
$$d_3\colon \text{ if } X = A_3 \text{ then } Y = B_3$$

in this case

$$R = ((A_1 \cap B_1) \cup A'_1) \cap ((A_2 \cap B_2) \cup A'_2) \cap ((A_3 \cap B_3) \cup A'_3)$$

$$\begin{aligned} R = {} & (A_1 \cap B_1 \cap A_2 \cap B_2 \cap A_3 \cap B_3) \cup (A_1 \cap B_1 \cap A_2 \cap B_2 \cap A'_3) \\ & \cup (A_1 \cap B_1 \cap A'_2 \cap A_3 \cap B_3) \cup (A_1 \cap B_1 \cap A'_2 \cap A'_3) \\ & \cup (A'_1 \cap A_2 \cap B_2 \cap A_3 \cap B_3) \cup (A'_1 \cap A_2 \cap B_2 \cap A'_3) \\ & \cup (A'_1 \cap A'_2 \cap A_3 \cap B_3) \cup (A'_1 \cap A'_2 \cap A'_3). \end{aligned}$$

$$R = T_1 \cup T_2 \cup T_3 \cup T_4 \cup T_5 \cup T_6 \cup T_7 \cup T_8.$$

We make the following assumptions

$$\text{I.}\quad A_1 \cup A_2 \cup A_3 = U$$

$$\text{II.}\quad A_1 \subset (A_2' \cap A_3')$$

$$A_2 \subset (A_1' \cap A_3')$$

$$A_3 \subset (A_1' \cap A_2')$$

From assumption I, $T_8 = \Phi$
From assumption II

$$T_4 = A_1 \cap B_1 \cap A_2' \cap A_3' = A_1 \cap B_1$$

$$T_7 = A_1' \cap A_2' \cap A_3 \cap B_3 = A_3 \cap B_3$$

$$T_6 = A_1' \cap A_2 \cap B_2 \cap A_3' = A_2 \cap B_2$$

Furthermore

$$T_1 \subset T_4 \text{ hence } T_1 \cup T_4 = T_4$$

$$T_2 \subset T_4 \text{ hence } T_4 \cup T_2 = T_4$$

$$T_3 \subset T_4 \text{ hence } T_4 \cup T_3 = T_4$$

$$T_5 \subset T_6 \text{ hence } T_5 \cup T_6 = T_6$$

thus

$$R = T_4 \cup T_6 \cup T_7 = \bigcup_{i=1}^{3} (Ai \cap Bi) = \bigcup_{i=1}^{3} (Ai \times Bi).$$

We can generalize this result:

Theorem. Assume $A_1, A_2, \ldots, A_k$ are fuzzy subsets of U and $B_1, B_2, \ldots, B_k$ are fuzzy subsets of W, if we have K fuzzy implications of the form

"if $X = A_i$ then $Y = B_i$,"

the conjunction of these implications can be represented in the form

$$\bigcup_{i=1}^{K} [A_i \times B_i]$$

if the conditions

$$(1) \quad \bigcup_{i=1}^{K} A_i = U$$

$$(2) \quad \text{For each } i, i = 1, 2, \ldots, K$$

$$A_i \subset \bigcap_{j=1, j \neq i}^{K} A_j'.$$

are satisfied.

Corollary. Assume U is a finite dimensional set. If we know the consequent for each $u \in U$ than we can use the above techniques.

Furthermore if $\bigcup_{i=1}^{K} A_i \neq U$, but the second condition is satisfied our model becomes

$$R = \bigcup_{i=1}^{K} (A_i \times B_i) \cup \bigcup_{i=1}^{K} (A_i').$$

Thus, if we can obtain a good covering of the set U in the sense of satisfying conditions 1 and 2 then Mamdani's approach is acceptable.

10. ALTERNATIVE APPROACHES TO CLASSIFICATION

The exist alternative approaches to the classification problem using fuzzy subset theory. One approach is due to Baldwin(1) another has been suggested by Yager.(12) Before discussing these approaches we must present some further operations involving fuzzy sets.

Assume A and B are two fuzzy subsets of U. The compatibility of B with A is defined as

$$\text{Comp}[B/A] = \left\{ \frac{A(u)}{B(u)} \right\} \qquad \text{for} \quad u \in U,$$

which is a fuzzy subset of the unit interval. The above compability can be interpreted as the truth of the statement $X = B$ given the proposition $X = A$.

Assume we have a proposition

$$(X \text{ is } Q) \text{ is } T$$

where Q is a fuzzy subset of U and T is a linguistic truth value, a fuzzy subset of the unit interval, the translation of this proposition(21) is

$$X \text{ is } Q^+$$

where Q^+ is also a fuzzy subset of U having membership

$$Q^+(u) = T(Q(u)) \qquad \text{for} \quad u \in U$$

Assume T_1 and T_2 are two linguistic truth values, fuzzy subsets of the unit interval, corresponding to the truth of two propositions P_1 and P_2; $V(P_1) = T_1$ is the truth of proposition P_1 similarly the truth of P_2 is $v(P_2) = T_2$. The truth of the combined proposition "P_1 and P_2," denoted $v(P_1 \text{ and } P_2) = T_3$, is also a linguistic truth value defined such that

$$T_2(y) = \text{Max}[T_1(l) \wedge T_2(k)]$$

overall $l, k \in [0, 1]$ such that

$$l \wedge k = y.$$

The truth of the disjunction, "P_1 or P_2," denoted $\vee (P_1 \text{ or } P_2) = T_4$ is defined as

$$T_4(y) = \text{Max}[T_1(l) \wedge T_2(K)].$$

over all $l, k \in [0, 1]$ such that

$$l \vee k = y.$$

With this background we can now define Baldwin's[1] approach to fuzzy inference.

Assume we have proposition P_1 describing some form of fuzzy implication,

$$\text{"if } X = A \text{ then } Y = B,\text{"}$$

where A and B are fuzzy subsets of U and V, respectively. Assume we have another fuzzy proposition, $X = C$, indicating the evaluation of X.

Baldwin[1] suggests the following procedure for finding the restriction on Y based upon these two propositions.

(1) Find $T = \{C(w)/A(w)\}$, the truth of $X = A$ given $X = C$.

(2) Let H be the fuzzy relationship on $[0, 1] \times [0, 1]$ indicating the Lukasiewicz implication,

$$H(a, b) = 1 \wedge (1 - a + b) \qquad a, b \in [0, 1].$$

Calculate the fuzzy subset W of the unit interval,

$$W = T \circ H,$$

where

$$W(b) = \operatorname*{Max}_{a \in [0,1]} [T(a) \wedge H(a, b)]$$

(3) Since W is the truth of the proposition $Y = B$, hence using truth function modification we get $Y = B^+$, as our restriction on the variable Y, where

$$B^+(v) = W[B(v)] \qquad \text{for each} \quad v \in V.$$

If we have K fuzzy implication statements each of the form "if $X = Ai$ then $Y = Bi$," as we would have if we building a fuzzy classifier, then we proceed as follows:

(1) For each implication statement calculate

$$T_i = \text{Comp}[Bi/C] = \left\{ \frac{C(w)}{Bi(w)} \right\}$$

(2) For each implication calculate the value Wi,

$$Wi = T_i \circ H$$

(3) Calculate the restriction due to each implication

$$Bi^+(v) = Wi[Bi(v)].$$

(4) The restriction on Y, Q, is equal to

$$Y = Q = B_1^+ \cap B_2^+ \cap \qquad \cap B_K^+$$

where

$$Q(v) = \operatorname*{Min}_i [Bi^+(v)].$$

If in our fuzzy classifiers the antecedent, $X = A$, is a compound statement of many properties, Baldwin's procedure handles this as follows:

Assume: P_1: If $X_1 = A_1$ and $X_2 = A_2 \cdots$ and $X_p = A_p$ then $Y = B$, with $X_1 = C_1$, $X_2 = C_2$, $X_p = C_p$, where A_1 and C_1, A_2 and C_2,..., A_p and C_p, are fuzzy subsets of U_1, U_2,..., U_p, respectively.

The procedure suggested by Baldwin is a follows:

(1) Let

$$Mi = \text{Comp}[Ai/Ci] = \left\{ \frac{Ci(u)}{Ai(u)} \right\} \qquad u \in U_i$$

(2) Calculate the conjunction of these truth values to give us the truth value associated with this implication

$$T(y) = \underset{\substack{\text{over all } a_j \in [0,1] \\ \text{Min } a_j = y}}{\text{Max}} \quad \underset{i=1,\ldots,k}{[\text{Min}[Mi(aj)]]}$$

(3) Proceed as above using T.

The approach suggested by Yager[12] though somewhat similar to that suggested by Baldwin assumes that the effect of the compatibility between A and C is to determine the importance of the constraint imposed upon the consequent.

Again given

$$\text{if } X = A \text{ then } Y = B$$

where $X = C$, Yager's method proceeds as follows

(1) Calculate

$$T = \text{Comp}[A/C],$$

however here T is interpreted as the importance associated with satisfying the consequent.

(2) Calculate the fuzzy subset Q of V, defined by

$$Q = B^T,$$

then

$Y = Q$ is the restriction imposed on Y by this information.

11. SPECIFICITY OF OUR CLASSIFICATION

By whatever method we use to model a problem, as a result of our classification procedure we obtain a possibility distribution Q over the set W of classification categories such that $Q(w)$ indicates the possibility that $w \in W$ is the classification of the object.

If Q has membership grade of one for one element and zero for all the other elements this indicates that our model is very specifically pointing to one value of W as the classification of the object we are studying. Yager[11,15] has suggested a measure of specificity to be associated with a possibility distribution.

Assume Q is a possibility distribution defined over the finite set W, let Q_α be the α level set of Q,

$$Q_\alpha = \{w \mid w \in W, Q(w) \geqslant \alpha\}$$

then the specificity of Q is defined as

$$S(Q, W) = \int_0^{\alpha \max} \frac{1}{\text{card } Q_\alpha}$$

where $\alpha \max$ is the maximal membership grade in Q and card $|Q_\alpha|$ is the number of elements in Q_α. We note that $S(Q, W)$ assumes its largest value of one when $Q = \{1/wi\}$ for some $wi \in W$ and it assumes its minimal value of zero when $Q = \Phi$. We note that if Q is normal, has possibility for one for at least one element, then if $Q \subset B$, $S(Q, S) \geqslant S(B, S)$. Furthermore, $Q = W$, has specificity of $1/n$, where n is the number of elements in W.

In a certain sense the specificity is a measure of the amount of information we have about the classification of our object with respect to the set W. The larger $S(Q, W)$ the more we know about which category in W our object belongs. We note that if the specificity is low, less than $1/n$ our model is telling us less information about which category of W our object belongs to than if we did not even use the model. This situation would indicate that our model needs to be revised to include a descriptor to give some information about objects of the type we are just trying to classify.

In some cases we must make some decision as to the most possible category in which our object lies. In general the selection is to choose the element with the largest grade of membership in Q. The specificity would measure the confidence we have in selecting this element.

12. VERIFICATION OF THE RELATIONS

Having constructed a classifier based upon an expert's descriptor granules we may be interested in determining how good this model is working. In order to do this we must again introduce the concept of compatibility between two fuzzy sets. Assume we have a fuzzy subset A of W which we call our reference set. Let B also be a fuzzy subset of U. The compability of B with the reference A is defined[21] to be

$$T = \text{Comp}[B/A] = \left\{ \frac{A(u)}{B(u)} \right\}$$

which is a fuzzy subset of the unit interval indicating the truth of the statement $X = B$ given that $X = A$ is true.

Example

$$A = \left\{\frac{.1}{u_1}, \frac{.6}{u_2}, \frac{1}{u_3}\right\}$$

$$B = \left\{\frac{.5}{u_1}, \frac{1}{u_2}, \frac{.8}{u_3}\right\}$$

$$T = \left\{\frac{.7}{.5}, \frac{.6}{1}, \frac{1}{.8}\right\}$$

Assume we have a classifier R, then given an object σ_1 to be classified if we apply the rules of fuzzy logic

$$\sigma_1 \circ R = B_1$$

where B_1 is the fuzzy subset of W indicating the category to which this object is to be classified. Assume that further investigation leads to the knowledge that the true classification of the object σ_1 is A_1. We can use the compability to measure the degree to which the model was correct. Therefore

$$T_1 = \text{Comp}[B_1/A_1]$$

measure the performance of the relation R. If we have a sample of n objects each generating measures of $T_1, T_2, \ldots, T_n$, respectively, we would like to find the mean performance of our model.

Consider a fuzzy subset T of the unit interval Yager[13] has suggested a real number measure of the value of this fuzzy number. Let $T_{[\alpha]}$ be the α level set associated with T,

$$T_{[\alpha]} = \{t/T(t) \geqslant \alpha, t \in I\},$$

let $M[T_{[\alpha]}]$ be the mean value of $T_{[\alpha]}$ and let a max be the maximum grade of membership of any element in T. Then

$$F[T] = \frac{1}{a \max} \int_0^{a \max} m[T_{[\alpha]}]\, d\alpha$$

acts as a representative value for T.

Furthermore Yager[13] has shown that if $T = T_1 + T_2 + \cdots + T_n/n$, then

$$F(T) = \frac{F(T_1) + F(T_2) + \cdots + F(T_n)}{n}.$$

Considering our situation in which we have n samples each of which has a truth value of Ti, we can use the above procedure to obtain an overall average measure of the performance of our classifier.

13. PROBABILISTIC INFORMATION

In some cases the information granules supplied by the expert may involve probabilistic information. For example, a doctor may say that if the blood pressure is high and the temperature is low then the disease w_1 is very likely. We shall describe a procedure for handling this type of information in our classifier. Our procedure will be based upon some ideas on probabilistic information granules developed by Zadeh.[19]

Assume V is a variable taking values in the base set W. Let F be a fuzzy subset of W. Furthermore assume λ is a linguistic probability, for example, likely, unlikely, about .7. λ can be represented as a fuzzy subset of the unit interval.

Consider a statement

$$V \text{ is } F \text{ is } \lambda,$$

for example the candidate is successful is likely. Zadeh[X] has suggested that such a statement has the effect of inducing a possibility distribution over the set of probabilities that can be associated with W. Let S be the set of all probability distributions which can be associated with the elements of W. The above information granule induces a possibility distribution Π over the set S such that for each $p \in S$

$$\Pi(p) = \lambda\left(\sum_{wi \in W} (P(wi) \cdot F(wi))\right)$$

Thus V is F is λ induces the possibility distribution Π over S.

Consider the granule

$$V \text{ is } G,$$

we can consider this as a special case of the above in the following manner. Let $Pi \in S$ be the probability distribution in which the probability of w_i is one. Then V is G can be said to induce the possibility distribution Π_G over S in which

$$\Pi_G(Pi) = G(wi) \qquad \text{for } wi \in W \text{ and } Pi \in S \text{ s.t. Prob}(wi) = 1$$

and

$$\Pi_G(P) = 0 \qquad \text{for all other } p \in S$$

We can now apply these ideas to the formulation of classification models.

Considering W to be our set of categories used for classification and $X = (X_1, X_2, \ldots, X_q)$ our classifying variables measured on the set $U = U_1 \times U_2 \times \cdots \times U_q$, assume we have as one of our descriptors a statement

$$d_1: \text{If } X = M \text{ then } (V = F \text{ is } \lambda),$$

where M is a fuzzy subset of U, F is a fuzzy subset of W and λ is a linguistic probability. From the above discussion we note that V is F is λ induces a possibility distribution Π_F over the set S of probabilities associated with W. Then the above descriptor is translated into a fuzzy relationship, D_1 on $U \times S$ which is defined by

$$D_1(u, p) = 1 \wedge (1 - M(u) + \Pi_F(p)).$$

Consider a second descriptor of the form

$$d_2: \text{If } X = L \text{ then } V = G,$$

where L is a fuzzy subset of U and G is a fuzzy subset of W. As we suggested above, $V = G$ can also be represented as a fuzzy subset of the set S of probability distributions on W, Π_G as defined above. Thus again we can then represent D_2 as a fuzzy subset of $U \times S$.

Then our classification relation

$$R = D_1 \cap D_2$$

is also a fuzzy subset of $U \times S$.

An object to be classified can be represented by a fuzzy subset H of U which indicates the evaluation of its properties. Then we can apply the law of fuzzy compositional inference which says

$$V = B = H \circ D,$$

where B is a fuzzy subset of S such that

$$B(p) = \operatorname*{Max}_{u \in U} [H(u) \wedge D(u, p)].$$

Thus as a result of applying the classifier to our object we obtain as the classification a possibility distribution over the set S of probability distribution associated with the elements of W. Using linguistic approximation techniques we may be able to retranslate the statement $V = B$ into a statement $V = E$ is λ_B, where E is a fuzzy subset of W and λ_B is a linguistic probability.

14. CONCLUSION

We have tried to develop the necessary structure for a general classification procedure based upon fuzzy set. This procedure enables one to use imprecise information in the construction of the model. It also enables one to handle imprecise evaluations for the characteristics to be determined. Since it is relationship based, the characteristic variables can be drawn from a nonnumeric scale. It should be noted that another form of classification based on discriminant analysis requires numeric scales. The output of our model is a possibility distribution associated with the categories an object can be a member.

REFERENCES

1. J. F. Baldwin, "A new approach to approximate reasoning using fuzzy logic," *Fuzzy Sets Syst.* **2**:309–325 (1979).
2. J. F. Baldwin and B. W. Pilsworth, "A model for fuzzy reasoning through multivalued logic and set theory," *Int. J. Man-Mach. Stud.* **11**:351–380 (1979).
3. W. Bandler and L. J. Kohout, "Semantics of implication operators and fuzzy relational products," Report FRP-8, University of Essex, England (1979).
4. E. Hisdale, "Generalized fuzzy systems and particularization," *Fuzzy Sets Syst.* **4**:275–291 (1980).
5. A. Kandel and R. R. Yager, "A 1979 bibliography on fuzzy sets, their applications and related topics," in *Advances in Fuzzy Set Theory and Applications* , M. M. Gupta, R. K. Ragade, and R. R. Yager, eds. (North Holland, Amsterdam, 1979), pp. 621–744.
6. E. H. Mamdani, "Advances in the linguistic synthesis of fuzzy controllers," *Int. J. Man-Mach. Stud.* **8**:669–678 (1976).
7. M. Mizumoto, S. Fukamo, and K. Tanaka, "Some methods of fuzzy reasoning," in *Advances in Fuzzy Set Theory and Applications*, M. M. Gupta, R. K. Ragade, and R. R. Yager, (eds.) (North Holland, Amsterdam, 1979), pp. 117–136.
8. E. Sanchez, J. Gouvernet, and H. Joly, "Medical diagnosis oriented linguistic approach to pattern classification," *Proc. Medinfo*, Tokyo, 1980.
9. B. S. Sembi and E. H. Mamdani, "On the nature of implication in fuzzy logic," *Proceedings of the 9th International Symposium on Multiple Valued Logic*, (1979), 143–151.
10. R. R. Yager, "Multiple objective decision making using fuzzy subsets," *Int. J. Man-Mach. Stud.* **9**:375–382 (1977).
11. R. R. Yager, "Measuring tranquility and anxiety in decision making: An application of fuzzy sets," Tech. Report #RRY 79-03, Iona College, New Rochelle (1979).
12. R. R. Yager, "An approach to inference in approximate reasoning," *Int. J. Man-Mach. Stud.* **13**: 323–338 (1980).
13. R. R. Yager, "A procedure for ordering fuzzy subsets of the unit interval," *Inf. Science* (to appear).

14. R. R. Yager, "On the implication operator in fuzzy logic," Tech. Report #RRY 80-03, Iona College, New Rochelle (1980).
15. R. R. Yager, "A measure of information on possibility distributions," *Proceedings of Round Table Conference on Fuzzy Sets*, Lyon, France (1980).
16. R. R. Yager, "Fuzzy subsets of type II in decisions," *J. Cybern.* **10**:137–159 (1980).
17. L. A. Zadeh, "Fuzzy sets," *Inf. Control* **8**:338–353 (1965).
18. L. A. Zadeh, "Ooutline of a new approach to the analysis of complex systems and decision processes," *IEEE Trans. Syst. Man Cybern.* **2**:28–44 (1973).
19. L. A. Zadeh, "A theory of approximate reasoning" Memo #UCB/ERL M77/58, University of California, Berkeley (1977).
20. L. A. Zadeh, "Fuzzy sets as a basis for a theory of possibility," *Fuzzy Sets Syst.* **1**:3–18 (1978).
21. L. A. Zadeh, "Possibility theory and soft data analysis," Memo #UCB/ERL M79/66, University of California, Berkeley (1979).
22. L. A. Zadeh, "Fuzzy sets and information granularity," in *Advances in Fuzzy Set Theory and Applications*, M. M. Gupta, R. K. Ragade, and R. R. Yager, (eds.) (North Holland, Amsterdam, 1979), pp. 3–18.

Printed in Belgium