FUZZY SETS AND THEIR APPLICATION

TO PATTERN CLASSIFICATION AND CLUSTER ANALYSIS

by

L. A. Zadeh

Memorandum No. ERL-M607

28 October 1976

FUZZY SETS AND THEIR APPLICATION TO PATTERN CLASSIFICATION AND CLUSTER ANALYSIS

by

L.A. Zadeh

Memorandum No. ERL-M607

28 October 1976

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

FUZZY SETS AND THEIR APPLICATION TO PATTERN CLASSIFICATION AND CLUSTER ANALYSIS

L.A. Zadeh*

Abstract

Most of the realistic problems in pattern classification and cluster analysis do not lend themselves to a precise mathematical formulation. For this reason, the theory of fuzzy sets and, in particular, the linguistic approach may provide a more natural setting for the formulation and solution of problems in pattern recognition than the conventional approaches based on classical set theory, probability theory and two-valued logic.

In the present paper, the problem of pattern classification is formulated as that of converting an opaque fuzzy recognition algorithm acting on a collection of objects into a transparent fuzzy recognition algorithm defined on an associated space of mathematical objects. A fuzzy subset of the space of objects (or mathematical objects) is assumed to be characterized by a relational tableau in which the entries are, in general, linguistic rather than numerical. A translation rule for relational tableaus is described and an approach to the interpolation of such tableaus is outlined.

The problem of cluster analysis is formulated as a conjunction of two subproblems. Problem a is that of converting an opaque recognition algorithm which defines a fuzzy similarity relation on a collection of pairs of objects into a transparent recognition algorithm defined on an associated space of pairs of mathematical objects; and Problem b is that of deducing from the fuzzy similarity relation a collection of fuzzy subsets (clusters) of

*Computer Science Division, Department of Electrical Engineering and Computer Sciences and the Electronics Research Laboratory, University of California, Berkeley, CA 94720. Research supported by U.S. Army Research Office Contract DAHC04-75-G0056.

mathematical objects which possess what is termed the property of fuzzy affinity. Such clusters may be obtained by applying a Dunn-Bezdek type of clustering algorithm to a fuzzy level-set of the fuzzy similarity relation.

FUZZY SETS AND THEIR APPLICATION

TO PATTERN CLASSIFICATION AND CLUSTER ANALYSIS

L.A. Zadeh*

1. Introduction

The development of the theory of fuzzy sets in the early sixties drew much of its initial inspiration from problems relating to pattern classification -- especially the analysis of proximity relations and the separation of subsets of R^n by hyperplanes. In a more fundamental way, however, the intimate connection between the theory of fuzzy sets and pattern classification rests on the fact that most real-world classes are fuzzy in nature -- in the sense that the transition from membership to nonmembership in such classes is gradual rather than abrupt. Thus, given an object x and a class F, the real question in most cases is not whether x is or is not a member of F, but the degree to which x belongs to F or, equivalently, the grade of membership of x in F.

There is, however, still another and as yet little explored connection between the theory of fuzzy sets and pattern classification. What we have in mind is the possibility of applying fuzzy logic and the so-called linguistic approach [1]-[4] to the definition of the basic concepts in pattern analysis as well as to the formulation of fuzzy algorithms for pattern recognition. The principal motivation for this approach is that most of the practical problems in pattern classification do not lend themselves to a precise mathematical formulation, with the consequence that the less precise methods based on the linguistic approach may well prove to be better matched to the imprecision which is intrinsic in such problems.

---

*Computer Science Division, Department of Electrical Engineering and Computer Sciences and the Electronics Research Laboratory, University of California, Berkeley, CA 94720. Research supported by U.S. Army Research Office Contract DAHC04-75-G0056.

Although the literature of the theory of fuzzy sets contains a substantial number of papers dealing with various aspects of pattern classification,[1] we do not, as yet, have a unified theory of pattern classification based on the theory of fuzzy sets. It is reasonable to assume that such a theory will eventually be developed, but its construction is likely to be a long-drawn task because it will require a complete reworking of the conceptual structure of the theory of pattern classification and radical changes in our formulation and implementation of pattern recognition algorithms.

In this perspective, the limited objective of the present paper is to outline a conceptual framework for pattern classification and cluster analysis based on the theory of fuzzy sets, and draw attention to some of the significant contributions by other investigators in which concrete pattern reconition and cluster analysis algorithms are described. For convenience of the reader, a brief exposition of the relevant aspects of the theory of fuzzy sets is presented in the Appendix.

2. Pattern Classification in a Fuzzy-Set-Theoretic Framework

To place the application of the theory of fuzzy sets to pattern classification in a proper perspective, we shall begin with informal definitions of some of the basic terms which we shall employ in later analysis.

To begin with, it will be necessary for our purposes to differentiate between an <u>object</u> which is pointed to (or labeled) by a pointer (identifier) p, and a <u>mathematical object</u>, x, which may be characterized precisely by specifying the values of a finite (or, more generally, a countable) set of parameters. For example, in the proposition "<u>Susan</u> is very intelligent," <u>Susan</u> is a pointer to a person named Susan. The person in question, however,

[1] Some of the representative papers bearing on the application of fuzzy sets to pattern classification and cluster analysis are listed in the bibliography.

is not a mathematical object until a set of measurement procedures $\{M_1,\ldots,M_n\}$ is defined such that the application of $\{M_1,\ldots,M_n\}$ to the object p (or, more precisely, the object pointed to by p) yields an n-tuple of constants $(x_1,\ldots,x_n)$ which represent the attribute-values (or feature-values) of the object in question. The n-tuple $x \triangleq (x_1,\ldots,x_n)$, then, characterizes a mathematical object associated with p, expressed symbolically as[2]

$$x \triangleq M(p) \tag{2.1}$$

where $M = (M_1,\ldots,M_n)$. For example, M_1, M_2, M_3, M_4 could be, respectively, the procedures for measuring the height, weight and temperature, and determining the sex of the object in question. In this case, a 4-tuple of the form (5'7", 125, 98.6, F) would be a mathematical object associated with the person named Susan.

An important point that needs to be noted is that there are many -- indeed an infinity -- of mathematical objects that may be associated with p. In the first place, different combinations of attributes may be measured. And second, different mathematical objects result when the precision of measurement -- or, equivalently, the resolution level -- of an attribute is varied. Thus, to associate a mathematical object x with an object p it is necessary to specify, explicitly or implicitly, the resolution levels of the attributes of p. Usually this is done implicitly rather than explicitly, which is the reason why the concept of a resolution level -- although important in principle -- does not play an overt role in pattern recognition.

Let U^o be a universe of objects, let U be the universe of associated mathematical objects, and let F be a fuzzy subset of U^o (or U).

---

[2] The symbol $\triangleq$ stands for "denotes" or "is equal to by definition."

There are three distinct ways in which F may be characterized:

(a) Listing. If the support[3] of F is a finite set, then F may be defined by a listing of its elements together with their respective grades of membership in F. For example, if U^o is the set of persons pointed to by the labels John, Luise, Sarah and David, and F is the fuzzy subset labeled tall, then F may be characterized as the collection of ordered pairs {(John, 0.9), (Luise, 0.8), (David, 0.7) and (Sarah, 0.8)}, which may be expressed more conveniently as the linear form (see A2)

$$\underline{\text{tall}} = 0.9\text{ John} + 0.8\text{ Luise} + 0.7\text{ David} + 0.8\text{ Sarah} \tag{2.2}$$

where + denotes the union rather than the arithmetic sum.

(b) Recognition algorithm. Such an algorithm, when applied to an object p, yields the grade of membership of p in F. For example, if someone were to point to Luise and ask "What is the degree to which Luise is tall?" then a recognition algorithm applied to the object Luise would yield the answer 0.8.

(c) Generation algorithm. In this case, an algorithm generates those elements of U^o which belong to the support of F and associates with each such element its grade of membership in F. As a simple illustration, the recurrence relation

$$x_n = x_{n-1} + x_{n-2} \tag{2.3}$$

with $x_0 = 0$, $x_1 = 1$ may be viewed as a nonfuzzy generation algorithm which defines the set of Fibonacci numbers {1,2,3,5,8,13,...}.[4] As an example of a generation algorithm which defines a fuzzy set, let U be the set of

[3] The support of a fuzzy set F is the set of elements of the universe of discourse whose grades of membership in F are positive.

[4] Many examples of nonfuzzy pattern generation algorithms may be found in the books by U. Grenander [5] and K.S. Fu [6].

strings over a finite alphabet, say {a,b}, and let G be a fuzzy context-free grammar whose production system is given by

$$\begin{array}{ll} S \xrightarrow{0.8} bA & B \xrightarrow{0.4} b \\ S \xrightarrow{0.6} aB & A \xrightarrow{0.3} bSA \\ A \xrightarrow{0.2} a & B \xrightarrow{0.5} aSB \end{array} \qquad (2.4)$$

in which S, A, B are nonterminals and the number above a production indicates its "strength." The fuzzy language, L(G), generated by this grammar may be defined as follows. Let x be a terminal string derived from S by a sequence of substitutions in which the left-hand side of a production in G is replaced by its right-hand side member, e.g.,

$$S \xrightarrow{0.8} bA \xrightarrow{0.3} bbSA \xrightarrow{0.2} bbSa \xrightarrow{0.6} bbaBa \xrightarrow{0.4} bbaba \qquad (2.5)$$

The strength of the derivation chain from S to x is defined to be the minimum of the strengths of constitutent productions in the chain, e.g., in the case of (2.5), the strength of the chain is $0.8 \wedge 0.3 \wedge 0.2 \wedge 0.6 \wedge 0.4 = 0.2$ (where $\wedge$ is the infix symbol for min). The grade of membership of x in L(G) is then defined as the strength of the strongest leftmost derivation[5] chain from S to x [7]. In the case of $x \triangleq$ bbaba, there is just one leftmost derivation, namely,

$$S \xrightarrow{0.8} bA \xrightarrow{0.3} bbSA \xrightarrow{0.6} bbaBA \xrightarrow{0.4} bbabA \xrightarrow{0.2} bbaba \qquad (2.6)$$

whose strength is 0.2. Consequently, the grade of membership of the string $x \triangleq$ bbaba in the fuzzy set L(G) is 0.2. In this way, one can associate a grade of membership in L(G) with every string that may be generated by G, and thus the production system (2.4) together with the rule for computing

---

[5]In leftmost derivation, the leftmost nonterminal is replaced by the right-hand member of the corresponding production.

the grade of membership of any string in U in $L(G)$, constitutes a generation algorithm which characterizes the fuzzy subset, $L(G)$, of U.

Opaque vs. Transparent Algorithms

For the purposes of our analysis, it is necessary to differentiate between recognition algorithms which are opaque and those which are transparent. Informally, by an opaque recognition algorithm we mean an algorithm whose description is not known. For example, the user of a hand calculator may not know the algorithm which is employed in the calculator to perform exponentiation. Or, a person may not be able to articulate the algorithm which he/she uses to assign a grade of membership to a painting in the fuzzy set of beautiful paintings.

As its designation implies, a recognition algorithm is transparent if its description is known. For example, a parsing algorithm which parses a string generated by a context-free grammar and thereby yields the grade of membership of the string in the fuzzy language generated by the grammar would be classified as a transparent algorithm.

Pattern Classification

Within the framework of the theory of fuzzy sets, the problem of pattern classification may be viewed -- in its essential form -- as that of conversion of an opaque recognition algorithm into a transparent recognition algorithm. More specifically, let U^o be a universe of objects and let R_{op} be an opaque recognition algorithm which defines a fuzzy subset F of U^o. Then, pattern classification -- or, equivalently, pattern recognition -- may be defined as the process of converting an opaque recognition algorithm R_{op} into a transparent recognition algorithm R_{tr}.[6]

---

[6] We assume for simplicity that only one fuzzy subset of U^o is defined by R_{op}. More generally, there may be a number of such subsets, say $F_1,\ldots,F_N$, with R_{op} yielding the grade of membership of p in each of these subsets.

As an illustration of this formulation, consider the following typical problem. Suppose that U^o is the universe of handwritten letters and that when a letter, p, is presented to a person, P, that person -- by employing an opaque recognition algorithm R_{op} -- can specify the grade of membership, $\mu_F(p)$, of p in, say, the fuzzy set, F, of handwritten A's. Thus, in symbols,

$$\mu_F(p) = R_{op}(p) \ , \quad \text{for } p \text{ in } U^o \ . \tag{2.7}$$

Usually, P is presented with a finite set of sample letters $p_1,\ldots,p_m$, so that the result of application of R_{op} to $p_1,\ldots,p_m$ is a set of ordered pairs $\{(p_1,\mu_F(p_1)),\ldots,(p_m,\mu_F(p_m))\}$ which in the notation of fuzzy sets may be expressed as the linear form

$$S_F = \mu_F(p_1)p_1 + \cdots + \mu_F(p_m)p_m \tag{2.8}$$

where S_F stands for a fuzzy set of samples from F, and a term of the form $\mu_F(p_i)p_i$, $i = 1,\ldots,m$, signifies that $\mu_F(p_i)$ is the grade of membership of p_i in F.

If, based on the knowledge of S_F, we could convert the opaque recognition algorithm R_{op} into a transparent recognition algorithm R_{tr}, then given any p we could deduce $\mu_F(p)$ by applying R_{tr} to p. Equivalently, we may view this as the process of interpolation of the membership function of F from the knowledge of the values which it takes at the points $p_1,\ldots,p_m$. It should be remarked that this is the way in which the problem of pattern classification was defined in [8], but the present formulation based on the conversion of R_{op} to R_{tr} appears to be more natural.

An important implicit assumption in pattern classification is that the recognition process must be <u>automatic</u>, in the sense that it must be performed by a machine rather than a human. This requires that the transparent recognition algorithm R_{tr} act on a mathematical object, $M(p)$, rather than on p itself, since an object must be well-defined in order to be capable of manipulation by a machine.

In more concrete terms, let U^o be a universe of objects and let M be a measurement procedure which associates with each object p in U^o a mathematical object $M(p)$ in U. Let F be a fuzzy subset of U^o which is defined by an opaque recognition algorithm R_{op} in the sense that

$$\mu_F(p) = R_{op}(p) \ , \quad p \in U^o \ .$$

Denote by R_{tr} a transparent recognition algorithm which acting on the mathematical object $M(p)$ yields $\mu_F(p)$. Then, the problem of <u>automatic</u> (or <u>machine</u>) <u>pattern recognition</u> may be expressed in symbols as that of determining M and R_{tr} such that

$$\mu_F(p) = R_{op}(p) \tag{2.9}$$

$$R_{tr}(M(p)) = R_{op}(p) \ , \quad p \in U^o \ . \tag{2.10}$$

Thus, the problem of automatic pattern recognition involves two distinct subproblems: (a) conversion of the object p into a mathematical object $M(p)$; and (b) conversion of the opaque recognition algorithm R_{op} which acts on p's into a transparent recognition algorithm which acts on $M(p)$'s. Of these, problem (a) is by far the more difficult. In the conventional nonfuzzy approach to pattern classification, it is closely related to the problem of feature analysis -- a problem which falls into the least well-defined and least well-developed area in pattern recognition [35]-[46].

It is important to observe that, from a practical point of view, it is desirable that (i) $M(p)$ be defined by a small number of attributes, and (ii) that the measurement of these attributes be relatively simple. With these added considerations, then, the problem of pattern classification may be reformulated in the following terms.

Given an opaque recognition algorithm R_{op} which defines a fuzzy subset of objects p in U^o.

Problem I. Specify a preferably small set of preferably simple measurement procedures which convert an object p in U^o into a mathematical object $M(p) = \{M_1(p),\ldots,M_n(p)\}$ in U.

Problem II. Convert R_{op} into a transparent recognition algorithm R_{tr} which acts on $M(p)$ and yields the grade of membership of p in F as defined by R_{op}.

In the above formulation, the problem of pattern classification is not mathematically well-defined. In part, this is due to the fact that, as pointed out earlier, the notion of an object does not admit of precise definition and hence the functions $M_1,\ldots,M_n$ cannot be regarded as functions in the accepted mathematical sense. In addition, since the desired equality

$$R_{tr}(M(p)) = R_{op}(p) \ , \quad p \in U^o \tag{2.11}$$

cannot be realized precisely, the problem of pattern classification does not admit of exact solution. Furthermore, an added source of imprecision in pattern classification problems relates to the difficulty of assessing the goodness of a transparent recognition algorithm which may be offered as a solution to a given problem.

The main thrust of the above comments is that the problem of pattern classification is intrinsically incapable of precise mathematical formulation. For this reason, the conceptual structure of the theory of fuzzy sets may well provide a more natural setting for the formulation and approximate solution of problems in pattern classification than the more traditional approaches based on classical set theory, probability theory and two-valued logic [35]-[46].

3. The Linguistic Approach to Pattern Classification

Most of the conventional approaches to pattern recognition are based on the tacit assumption that the mapping from the object space U^o to the feature space U has the property that if two mathematical objects $M(p)$ are "close" to one another in terms of some metric defined on U, then p and q are likely to be in the same class in U^o.[7] When F is a fuzzy set, this assumption may be expressed more concretely but not very precisely as the property of <u>μ-continuity</u> of M, namely: If p and q are objects in U^o and for almost all p and q $M(p)$ is close to $M(q)$ in terms of a metric defined on U, then the grade of membership of p in F, $\mu_F(p)$, is close to that of q, $\mu_F(q)$.

The importance of μ-continuity derives from the fact that it provides a basis for reducing Problem II to the interpolation of a "well-behaved" (i.e., smooth, slowly-varying) membership function. More significantly for our purposes, it makes it possible to employ the linguistic approach for describing the dependence of μ_F on the linguistic values of the attributes of an object.

More specifically, suppose that $M(p)$ has n components $x_1 \triangleq M_1(p),\ldots,x_n \triangleq M_n(p)$, with x_i, $i = 1,\ldots,n$, taking values in U_i. Let $\mu_F(p)$ denote the grade of membership of p in F. We assume that the dependence of $\mu_F(p)$ on $x_1,\ldots,x_n$ is expressible as an $(n+1)$-ary fuzzy relation R in $U_1 \times \cdots \times U_n \times V$, where $V \triangleq [0,1]$. In what follows, R will be referred to as the <u>relational tableau</u> defining $\mu_F(p)$.

[7] This assumption is implicit in perceptron-type approaches and is related to the notion of compactness in the potential function method of Aizerman, Braverman and Rozonoer [9]-[12].

An essential assumption which motivates the linguistic approach is that our perception of the dependence of $\mu_F(p)$ on $x_1,\ldots,x_n$ is generally not sufficiently precise or well-defined to enable us to tabulate $\mu_F(p)$ as a function of the numerical values of $x_1,\ldots,x_n$. As a coarser and hence less precise characterization of this dependence, we allow the tabulated values of $x_1,\ldots,x_n$ and $\mu_F(p)$ to be linguistic rather than numerical, employing the techniques of the linguistic approach to enable us to interpolate R for the untabulated values of $x_1,\ldots,x_n$.

To be more specific, it is helpful to assume, as in [86], that a linguistic value of x_i, $i = 1,\ldots,n$, is an answer to the question Q_i: "What is the value of x_i?" and that the corresponding linguistic value of $\mu_F(p)$ is the answer to the question Q: "If the answers to $Q_1,\ldots,Q_n$ are $r_1,\ldots,r_n$, respectively, then what is the value of $\mu_F(p)$?" A purpose of this interpretation of the values of $x_1,\ldots,x_n,\mu_F(p)$ is to express the recognition algorithm R_{tr} as a branching questionnaire, that is, a questionnaire in which the questions are asked sequentially, with the question asked at stage j depending on the answers to the previous questions. The conversion of a relational tableau to a branching questionnaire is discussed in greater detail in [86].

Typically, the entries in a relational tableau are of the form shown in Table 1, in which the rows correspond to different objects, with the entry under Q_i representing a linguistic value of x_i for a particular object. (For simplicity, we shall speak interchangeably of the values of x_i and Q_i.) The questions $Q_1,\ldots,Q_n$ will be referred to as the <u>constituent questions</u> of R (or Q).

Q_1	Q_2	Q_3	Q
true	small	wide	high
very true	very small	not wide	very high
not very true	medium	NA	not very high
borderline	very large	not wide	low
not true	not very small	not very wide	more or less low
true or not very true	small	not very wide	very low

Table 1. A relational tableau defining the dependence of Q on Q_1, Q_2, Q_3.

In this table, the entries in the column labeled Q_i constitute a subset of the term-set of Q_i (see A66), that is, the possible linguistic values that may be assigned to Q_i. For example, the term-set of Q_1 might be: {true, very true, not very true, borderline, very (not true), not true, not borderline, very very true, ...}. The elements of the term-set of Q_i are assumed to be generated by a context-free grammar. For instance, the elements of the term-set of Q_1 can be generated by the grammar

$$
\begin{array}{ll}
S \rightarrow A & C \rightarrow D \\
S \rightarrow S \text{ or } A & C \rightarrow E \\
A \rightarrow B & D \rightarrow \text{very } D \\
A \rightarrow A \text{ and } B & E \rightarrow \text{very } E \\
B \rightarrow C & D \rightarrow \text{true} \\
B \rightarrow \text{not } C & E \rightarrow \text{borderline}
\end{array} \tag{3.1}
$$

in which S, A, B, C, D, E are nonterminals and "or," "and," "not," "very," "true" and "borderline" are terminals. Using the production system of this grammar, the linguistic value "true or not very true" may be derived from S by the chain of substitutions

$$
\begin{aligned}
&S \rightarrow S \text{ or } A \rightarrow A \text{ or } A \rightarrow B \text{ or } A \rightarrow C \text{ or } A \rightarrow D \text{ or } A \rightarrow \text{true or } A \rightarrow \\
&\text{true or } B \rightarrow \text{true or not } C \rightarrow \text{true or not } D \rightarrow \text{true or not very } D \rightarrow \\
&\text{true or not very true}
\end{aligned} \tag{3.2}
$$

The linguistic values of Q_i play the role of labels of fuzzy subsets of a universe of discourse which is associated with Q_i. For example, in the case of Q_1 the universe U_1 is the unit interval $[0,1]$, and "true" is a fuzzy subset of U_1 whose membership function might be defined in terms of the S-function (see A17) by

$$\mu_{true}(v) = S(v;0.6,0.75,0.9)\ , \quad v \in [0,1] \tag{3.3}$$

where $S(v;\alpha,\beta,\gamma)$ is an S-shaped function which vanishes to the left of α, is unity to the right of γ and takes the value 0.5 at $\beta = \frac{\alpha+\gamma}{2}$. Similarly, the membership function of the fuzzy subset labeled "borderline" may be defined in terms of the π-function (see A18) by

$$\mu_{borderline}(v) = \pi(v;0.3,0.5) \tag{3.4}$$

where $\pi(v;\beta,\gamma)$ is a bell-shaped function whose bandwidth is β and which achieves the value 1 at γ.

By the use of a semantic technique which is described in [2], it is possible to compute in a relatively straightforward fashion the membership function of the fuzzy set which plays the role of the meaning of a linguistic value in the term-set of Q_i. For example, the membership functions of "not true," "very true," "not very true" and "true or not very true" are related to that of "true" by the equations (in which the argument v is suppressed for simplicity)

$$\mu_{not\ true} = 1 - \mu_{true} \tag{3.5}$$

$$\mu_{very\ true} = (\mu_{true})^2 \tag{3.6}$$

$$\mu_{not\ very\ true} = 1 - (\mu_{true})^2 \tag{3.7}$$

$$\mu_{true\ or\ not\ very\ true} = \mu_{true} \vee (1 - (\mu_{true})^2) \tag{3.8}$$

where $(\mu_{true})^2$ denotes the square of the membership function of true and $\vee$ stands for the infix form of max.

A fuzzy set (or fuzzy sets) in terms of which the meaning of all other linguistic values in the term-set of Q_i may be computed is termed a primary fuzzy set (or sets). Thus, in the case of Q_1 the primary fuzzy set is labeled "true;" in the case of Q_2 the primary fuzzy sets are "small," "medium" and "large;" and in the case of Q the primary fuzzy sets are "high" and "medium," with "low" defined in terms of "high" by

$$\mu_{low}(v) = \mu_{high}(1-v) \ , \quad v \in [0,1] \ . \tag{3.9}$$

In effect, a primary fuzzy set plays a role akin to that of a unit whose meaning is context-dependent and hence must be defined a priori. The important point is that once the meaning of the primary terms is specified, the meaning of non-primary terms in the term-set of each Q_i may be computed by the application of the semantic rule which is associated with that Q_i.

The entry NA in Q_3 stands for "not applicable." What this means is that if the answer to Q_1 is, say, "not very true" and the answer to Q_2 is "medium," then Q_3 is not applicable to the object corresponding to the third row in the table. As a simple illustration of non-applicability, if the answer to the question "Is p a prime number?" is "true," then the question "What is the largest divisor of p other than 1?" is not applicable to p.

In the representation of R in the form of a relational tableau, it is helpful to divide the constituent questions into two categories: attributional and classificational. As its name implies, an attributional question is one which asks for the value of an attribute of p, e.g., Q_2 and Q_3 in Table 1 are attributional questions. A classificational question, on the other hand, relates to the degree to which a specified property is possessed by the object in question. Thus, the answer to a classificational question is either a truth-value, as in Q_1, or the grade of membership, as in Q. In both cases, the universe of discourse associated with a classificational question is assumed to be the interval [0,1]. Generally, we shall assume that "high" is equivalent to "true;" "medium" to "borderline;" and "low" to "false," where, by analogy with (3.9), "false" is defined by

$$\mu_{false}(v) = \mu_{true}(1-v) , \quad v \in [0,1] . \tag{3.10}$$

As an illustration of the above approach, assume that we wish to characterize the concept of an oval[8] contour, with U being the space of curved, smooth, simply-connected and non-self-intersecting contours in a plane.[9] To simplify the example, we assume that the constituent questions are limited to the following.

Classificational: $Q_1 \triangleq$ Does p have an axis of symmetry?

Classificational: $Q_2 \triangleq$ Does p have a second axis of symmetry?

Classificational: $Q_3 \triangleq$ Are the two axes of symmetry orthogonal?

Classificational: $Q_4 \triangleq$ Does p have more than two axes of symmetry?

Attributional: $Q_5 \triangleq$ What is the ratio of the lengths of the major and minor axes?

Classificational: $Q_6 \triangleq$ Is p convex?

For simplicity, the answers to the classificational questions are allowed to be only <u>true</u>, <u>borderline</u> and <u>false</u>, abbreviated to t, b and f, respectively, with the membership functions of t, b and f expressed in terms of the S and π functions by (3.3), (3.4) and[10]

---

[8]For purposes of this example, by oval we mean a shape resembling that of an egg.

[9]Note that the point of departure in this example is U rather than U^o because we assume that a contour is a mathematical object.

[10]It should be understood that "true" and "false" in the present context do not have the same meaning as they do in classical logic. Rather, as in fuzzy logic [3], "true" in the sense of (3.3) means "approximately true," and likewise for "false."

$$\mu_f(v) = \mu_t(1-v) \qquad (3.11)$$
$$= 1 - S(v;0.1,0.25,0.4)\ .$$

Similarly, the term-set for Q_5 is assumed to be

$$T(Q_5) = \{\text{about } 1, \text{ about } 1.5, \text{ about } 2, \text{ about } 2.5, \text{ about } 3, \text{ about } 4, \text{ about } 5, > \text{about } 5\}$$

where about α, $\alpha = 2,\ldots,5$, is defined by (with the arguments of π and S suppressed for simplicity)

$$\text{about } \alpha = \pi(0.4,\alpha) \qquad (3.12)$$

and

$$\text{about } 1 = 1 - S(1,0.2,0.4)\ . \qquad (3.13)$$

The answer to Q_6 is assumed to be provided by a subquestionnaire with an unspecified number of classificational constituent questions $Q_{61}, Q_{62}, \ldots$ which are intended to check on whether the slope of the tangent to the contour is a monotone function of the distance traversed along the contour by an observer. Thus, if an observer begins to traverse the contour in, say, the counterclockwise direction starting at a point a_0, and $a_1,\ldots,a_m$ are regularly spaced points on the contour, with $a_{m+1} = a_0$, then Q_{6i} would be the question

$Q_{6i} \triangleq$ Is the slope of the tangent at a_i greater than that at a_{i-1}, $i = 1,2,\ldots,m+1$?

The answer to Q_6 is assumed to be true if and only if the answers to all of the constituent questions $Q_{61}, Q_{62}, \ldots$ are true.

In terms of the constituent questions defined above, the relational tableau characterizing an oval object may be expressed in a form such as shown in Table 2. For simplicity, only a few of the possible combinations of answers to these questions are exhibited in the table (NA stands for <u>not applicable</u>).

Q_1	Q_2	Q_3	Q_4	Q_5	Q_6	Q
t	t	t	f	about 1	t	b
t	t	t	f	about 1.5	t	t
f	f	t	f	about 1	t	f
t	f	NA	f	about 1	t	f
t	b	NA	f	about 1	t	b
t	b	NA	f	about 1.5	t	b

Table 2. Relational tableau characterizing an oval object

The first row in this table signifies that if the answer to Q_1 is t (i.e., p has one axis of symmetry); the answer to Q_2 is t (i.e., p has a second axis of symmetry; the answer to Q_3 is t (i.e., the two axes of symmetry are orthogonal); the answer to Q_4 is f (i.e., p has two and only two axes of symmetry); the answer to Q_5 is about 1 (i.e., the major and minor axes are about equal in length); and the answer to Q_6 is t (i.e., Q_6 is convex), with the answer to Q_6 provided by the subquestionnaire; then the answer to Q is b (i.e., p is an oval object to a degree which is approximately equal to 0.5, with "approximately equal to 0.5" defined by (3.4)).

Similarly, the fifth row in the table signifies that if the answer to Q_1 is t; the answer to Q_2 is b; the answer to Q_3 is NA; the answer to Q_4 is f; the answer to Q_5 is about 1 and the answer to Q_6 is t; then the answer to Q is b. Comparing the entries in row 5 with those of row 6, we note the answer to Q remains b when we change the answer to Q_5 from about 1 to about 1.5.

4. Translation Rules and the Interpolation of a Relational Tableau

Assuming that we have a characterization of $M(p)$ in the form of a relational tableau R, the question that arises is: How can we deduce from R the grade of membership of an object p in F?

As a preliminary to arriving at an approximate answer to this question, we have to develop a way of converting R into an $(n+1)$-ary fuzzy relation in $U_1 \times \cdots \times U_n \times V$. To this end, we shall employ the translation rules of fuzzy logic -- rules which provide a basis for translating a composite fuzzy proposition into a system of so-called relational assignment equations [14].

More specifically, let p be a pointer to an object and let q be a proposition of the form

$$q \triangleq p \text{ is } F \tag{4.1}$$

where F is a fuzzy subset of U. For example, q may be

$$q \triangleq \text{Pamela is tall} \, . \tag{4.2}$$

Translation rule of Type I asserts that q translates into

$$p \text{ is } F \rightarrow R(A(p)) = F \tag{4.3}$$

where $A(p)$ is an implied attribute of p and $R(A(p))$ is a <u>fuzzy restriction</u>[11] on the variable $A(p)$. Thus, (4.3) constitutes a relational assignment equation in the sense that the fuzzy set F -- viewed as a unary fuzzy relation in U -- is assigned to the restriction on $A(p)$. For example, in the case of (4.2), the rule in question yields

$$\text{Pamela is tall} \rightarrow R(\text{Height}(\text{Pamela})) = \text{tall}$$

[11] A fuzzy restriction is a fuzzy relation which acts as an elastic constraint on the values that may be assigned to a variable [2], [14].

where R(Height(Pamela)) is a fuzzy restriction on the values that may be assigned to the variable Height(Pamela).

Now let us consider two propositions, say

$$q_1 \triangleq p_1 \text{ is } F_1 \tag{4.4}$$

and

$$q_2 \triangleq p_2 \text{ is } F_2 \tag{4.5}$$

where p_1 and p_2 are possibly distinct objects, and F_1 and F_2 are fuzzy subsets of U_1 and U_2, respectively. For example, q_1 and q_2 might be $q_1 \triangleq X$ is large and $q_2 \triangleq Y$ is small.

By (4.3), the translations of q_1 and q_2 are given by

$$p_1 \text{ is } F_1 \rightarrow R(A_1(p_1)) = F_1 \tag{4.6}$$

$$p_2 \text{ is } F_2 \rightarrow R(A_2(p_2)) = F_2 \tag{4.7}$$

where $A_1(p_1)$ and $A_2(p_2)$ are implied attributes of p_1 and p_2.

By the rule of conjunctive composition [4], the translation of the composite proposition q_1 and q_2 is given by

$$q_1 \text{ and } q_2 \rightarrow R(A_1(p_1),A_2(p_2)) = F_1 \times F_2 \tag{4.8}$$

where $F_1 \times F_2$ denotes the cartesian product of F_1 and F_2 (see A56) which is assigned to the restriction on $A_1(p_1)$ and $A_2(p_2)$. Dually, by the rule of disjunctive composition, the translation of the composite composition q_1 or q_2 is given by

$$q_1 \text{ or } q_2 \rightarrow R(A_1(p_1),A_2(p_2)) = \bar{F}_1 + \bar{F}_2 \tag{4.9}$$

where $\bar{F}_1$ and $\bar{F}_2$ are the cylindrical extensions of F_1 and F_2 (see A59) and + denotes the union.

As we shall see presently, these two rules provide a basis for constructing a translation rule for relational tableaus. More specifically, consider a tableau of the form shown in Table 3

A_1	A_2	$\cdots$	A_n
r_{11}	r_{12}		r_{1n}
r_{21}	r_{22}	$\cdot$	r_{2n}
$\vdots$	$\vdots$	$\vdots$	$\vdots$
r_{m1}	r_{m2}		r_{mn}

Table 3. A relational tableau

in which $A_1,\ldots,A_n$ are variables taking values in $U_1,\ldots,U_n$, and the r_{ij} are linguistic labels of fuzzy subsets of U_j. (In relation to Table 1, the A_j play the roles of Q_j and Q.)

Expressed in words, the meaning of the tableau in question may be stated as:

$$\begin{array}{ll} & A_1 \text{ is } r_{11} \text{ and } A_2 \text{ is } r_{12} \text{ and } \cdots \text{ and } A_n \text{ is } r_{1n} \\ \text{or} & \\ & A_1 \text{ is } r_{21} \text{ and } A_2 \text{ is } r_{22} \text{ and } \cdots \text{ and } A_n \text{ is } r_{2n} \\ \text{or} & \\ \vdots & \quad \cdots\cdots \\ \text{or} & \\ & A_1 \text{ is } r_{m1} \text{ and } A_2 \text{ is } r_{m2} \text{ and } \cdots \text{ and } A_n \text{ is } r_{mn} \end{array} \tag{4.10}$$

Regarding (4.9) as a composite proposition and applying (4.8) and (4.9) to (4.10), we arrive at the tableau translation rule which is expressed by

$$\begin{array}{|c|c|c|} A_1 & \cdots & A_n \\ \hline r_{11} & \cdots & r_{1n} \\ \vdots & \cdots & \vdots \\ r_{m1} & \cdots & r_{mn} \end{array} \longrightarrow R(A_1,\ldots,A_n) = r_{11}\times\cdots\times r_{1n} + \cdots + r_{m1}\times\cdots\times r_{mn} \tag{4.11}$$

where $r_{11}\times\cdots\times r_{1n} + \cdots + r_{m1}\times\cdots\times r_{mn}$ is an n-ary fuzzy relation in

$U_1 \times \cdots \times U_n$ which is assigned to the restriction $R(A_1,\ldots,A_n)$ on the values of the variables $A_1,\ldots,A_n$.

As a very simple illustration of the tableau translation rule, assume that the tableau of R is given by [86]

$$\begin{array}{c|c||c} Q_1 & Q_2 & Q \\ \hline t & t & vf \\ f & f & t \end{array} \tag{4.12}$$

where t, f and vf are abbreviations for true, false and very false, respectively, and

$$U_1 = U_2 = V = 0 + 0.2 + 0.4 + 0.6 + 0.8 + 1 \tag{4.13}$$

$$t = 0.6/0.8 + 1/1 \tag{4.14}$$

$$f = 1/0 + 0.6/0.2 \tag{4.15}$$

and by (3.6)

$$vf = 1/0 + 0.36/0.2 \quad . \tag{4.16}$$

Applying the translation rule (4.11) to the table in question, we obtain the ternary fuzzy relation in $V \times V \times V$:

$$\begin{aligned} R(Q_1,Q_2,Q) &= t \times t \times vf + f \times f \times t \\ &= (0.6/0.8 + 1/1) \times (0.6/0.8 + 1/1) \times (1/0 + 0.36/0.2) \\ &\quad + (1/0 + 0.6/0.2) \times (1/0 + 0.6/0.2) \times (0.6/0.8 + 1/1) \\ &= 0.36/\big((0.8,0.8,0.2) + (0.8,1,0.2) + (1,0.8,0.2) + (1,1,0.2)\big) \\ &\quad + 0.6/\big((0,0,0.8) + (0,0.2,0.8) + (0.2,0,0.8) + (0.2,0.2,0.8) \\ &\qquad + (0,0.2,1) + (0.2,0.2,1)\big) \\ &\quad + 1/\big((0,0,1) + (1,1,0)\big) \end{aligned} \tag{4.17}$$

as the expression for the meaning of the relational tableau (4.12).

The Mapping Rule

The translation rule expressed by (4.11) provides a basis for an interpolation of a relational tableau, yielding an approximate value for the answer to Q given the answers to $Q_1,\ldots,Q_n$ which do not appear in R.

Specifically, let $(r_{i1},\ldots,r_{in},r_i)$ denote the i^{th} (n+1)-tuple in R and let $\tilde{R}$ denote the (n+1)-ary fuzzy relation in $U_1 \times \cdots \times U_n \times V$, $V \triangleq [0,1]$, expressed by

$$\tilde{R} = r_{11} \times \cdots \times r_{1n} \times r_1 + \cdots + r_{m1} \times \cdots \times r_{mn} \times r_m \tag{4.13}$$

where, as in (4.11), $\times$ and $+$ denote the cartesian product and union, respectively.

Now suppose that $g_1,\ldots,g_n$ are given fuzzy subsets of $U_1,\ldots,U_n$, respectively, and that we wish to compute the value of Q given that the values of $Q_1,\ldots,Q_n$ are $g_1,\ldots,g_n$.

Let $R(g_1,\ldots,g_n)$ denote the result of the substitution and hence the desired value of Q, and let G denote the cartesian product

$$G = g_1 \times \cdots \times g_n \ . \tag{4.14}$$

Then, the <u>mapping rule</u> may be expressed compactly as[12]

$$R(g_1,\ldots,g_n) = \tilde{R} \circ G \tag{4.15}$$

where $\circ$ denotes the composition (see A60) of the (n+1)-ary fuzzy relation $\tilde{R}$ with the n-ary fuzzy relation G.

In more explicit terms, the right-hand member of (4.15) is a fuzzy subset of U which may be computed as follows.

[12] This mapping rule may be viewed as an extension to a fuzzy relation of the mapping rule employed in such query languages as SQUARE and SEQUEL [15], [16].

Assume for simplicity that $U_1,\ldots,U_n,V$ are finite sets which may be expressed in the form (+ denotes the union)

$$\begin{aligned} U_1 &= u_1^1+\cdots+u_{k_1}^1 \\ U_2 &= u_1^2+\cdots+u_{k_2}^2 \\ &\cdots\cdots\cdots \\ U_n &= u_1^n+\cdots+u_{k_n}^n \\ V &= v_1+\cdots+v_k \end{aligned} \tag{4.16}$$

Now suppose that the g_i are expressed as fuzzy subsets of the U_i by (see A6)

$$\begin{aligned} g_1 &= \gamma_1^1 u_1^1+\cdots+\gamma_{k_1}^1 u_{k_1}^1 \\ &\cdots\cdots\cdots\cdots \\ g_n &= \gamma_1^n u_1^n+\cdots+\gamma_{k_n}^n u_{k_n}^n \end{aligned} \tag{4.17}$$

so that

$$G = \sum_I \gamma_{i_1}^1 \wedge \gamma_{i_2}^2 \wedge\cdots\wedge \gamma_{i_n}^n / u_{i_1}^1 u_{i_2}^2 \cdots u_{i_n}^n \tag{4.18}$$

where I denotes the index sequence $(i_1,\ldots,i_n)$, with $1\le i_1\le k_1$, $1\le i_2\le k_2,\ldots,1\le i_n\le k_n$; $u_{i_1}^1 u_{i_2}^2\cdots u_{i_n}^n$ is an abbreviation for the n-tuple $(u_{i_1}^1,u_{i_2}^2,\ldots,u_{i_n}^n)$, and $\gamma_{i_1}^1\wedge\gamma_{i_2}^2\wedge\cdots\wedge\gamma_{i_n}^n$ is the grade of membership of the n-tuple $u_{i_1}^1 u_{i_2}^2\cdots u_{i_n}^n$ in the n-ary fuzzy relation G.

By the definition of composition, the composition of $\tilde{R}$ with G may be expressed as the projection on $U_1\times\cdots\times U_n$ of the intersection of $\tilde{R}$ with the cylindrical extension of G. Thus,

$$\tilde{R}\circ G = \operatorname*{Proj}_{U_1\times\cdots\times U_n} (\tilde{R}\cap\bar{G}) \tag{4.19}$$

where $\bar{G}$ is given by

$$\bar{G} = \sum_{(I,i)} \gamma^1_{i_1} \wedge \gamma^2_{i_2} \wedge \cdots \wedge \gamma^n_{i_n} / u^1_{i_1} u^2_{i_2} \cdots u^n_{i_n} v_i \quad . \tag{4.20}$$

In this expression, (I,i) denotes the index sequence $(i_1,\ldots,i_n,i)$, with $1 \le i \le k$, and $u^1_{i_1} u^2_{i_2} \cdots u^n_{i_n} v_i$ is an abbreviation for the $(n+1)$-tuple $(u^1_{i_1}, u^2_{i_2}, \ldots, u^n_{i_n}, v_i)$.

Now suppose that the computation of the right-hand member of (4.13) yields $\tilde{R}$ in the form

$$\tilde{R} = \sum_{(I,i)} \mu_{(I,i)} / u^1_{i_1} u^2_{i_2} \cdots u^n_{i_n} v_i \quad . \tag{4.21}$$

Then, the intersection of $\tilde{R}$ with $\bar{G}$ is given by

$$\tilde{R} \cap \bar{G} = \sum_{(I,i)} \gamma^1_{i_1} \wedge \gamma^2_{i_2} \wedge \cdots \wedge \gamma^n_{i_n} \wedge \mu_{(I,i)} / u^1_{i_1} \cdots u^n_{i_n} v_i \tag{4.22}$$

and the projection[13] of $\tilde{R} \cap \bar{G}$ on $U_1 \times \cdots \times U_n$ -- and hence the composition of $\tilde{R}$ and G -- is expressed by

$$\tilde{R} \circ G = \sum_{(I,i)} \gamma^1_{i_1} \wedge \gamma^2_{i_2} \wedge \cdots \wedge \gamma^n_{i_n} \wedge \mu_{(I,i)} / v_i \tag{4.23}$$

where, to recapitulate:

$R(g_1,\ldots,g_n) = \tilde{R} \circ G$

= result of substitution of g_i for Q_i, $i = 1,\ldots,n$, in R;

$G = g_1 \times \cdots \times g_n$;

$\gamma^\lambda_{i_\lambda}$ = grade of membership of u_{i_λ} in g_λ, $\lambda = 1,\ldots,n$;

[13] A convenient way of obtaining the projection is to set $u^1_{i_1} = \cdots = u^n_{i_n} = 1$ in the right-hand member of (4.22) and treat the $(n+1)$-tuple $(u^1_{i_1},\ldots,u^n_{i_n},v_i)$ as if it were an algebraic product of $u^1_{i_1},\ldots,u^n_{i_n},v_i$.

$$I \triangleq (i_1,\ldots,i_n)$$
$$(I,i) \triangleq (i_1,\ldots,i_n,i)$$
$$\mu_{(I,i)} \triangleq \text{grade of membership of } (u^1_{i_1},u^2_{i_2},\ldots,u^n_{i_n},v_i) \text{ in } \tilde{R}$$
$$\tilde{R} = r_{11}\times\cdots\times r_{1n}\times r_1 + \cdots + r_{m1}\times\cdots\times r_{mn}\times r_m$$

It should be noted that we would obtain the same result by assigning $g_1,\ldots,g_n$ to $Q_1,\ldots,Q_n$ in sequence rather than simultaneously. This is a consequence of the identity

$$\tilde{R}\circ G = (\cdots((\tilde{R}\circ g_1)\circ g_2)\cdots\circ g_n) \tag{4.24}$$

which in turn follows from the identity

$$\sum_{(I,i)} \gamma^1_{i_1} \wedge \gamma^2_{i_2} \wedge \cdots \wedge \gamma^n_{i_n} \wedge \mu_{(I,i)} / v_i \tag{4.25}$$
$$= \sum_{(I,i)} [[[\gamma^1_{i_1} \wedge \mu_{(I,i)}/u^1_{i_1} u^2_{i_2} \cdots u^n_{i_n} v_i]_{u^1_{i_1}=1} \wedge \gamma^2_{i_2}]_{u^2_{i_2}=1} \wedge \cdots \wedge \gamma^n_{i_n}]_{u^n_{i_n}=1} \, .$$

As a very simple illustration of the mapping operation, assume that

$n = 2$;

$U_1 = U_2 = V = 0 + 0.2 + 0.4 + 0.6 + 0.8 + 1$;

$\tilde{R}$ is given by

$$\begin{aligned}\tilde{R} = {} & 1/(0,0,0) + 0.8/(0,0,0.2) + 0.7/(0.2,0.2,0) \\ & + 0.6/(0.2,0,0) + 0.8/(0.4,0.6,0.4) + 0.8/(0.4,0.2,0) \\ & + 0.5/(0.4,0.2,0.4) + 0.6/(0.2,0.6,0.8) + 0.8(0.8,0.8,0.2) \\ & + 0.9/(0.8,0.8,1) + 0.8/(0.8,1,0.8) + 0.6/(0.2,0.8,1) \\ & + 0.8/(0.6,0.8,1)\end{aligned} \tag{4.26}$$

and

$$g_1 = 0.6/0.4 + 1/0.2 \tag{4.27}$$

$$g_2 = 1/0.6 + 0.8/0.2 \quad . \tag{4.28}$$

Then by (4.18)

$$\begin{aligned} g &= g_1 \times g_2 \qquad (4.29) \\ &= 0.6/(0.4,0.6) + 0.6/(0.4,0.2) + 1/(0.2,0.6) + 0.8/(0.2,0.2) \end{aligned}$$

and thus

$$\begin{aligned} R(g_1,g_2) &= \tilde{R} \circ g \qquad (4.30) \\ &= 0.6 \wedge 0.8/0.4 + 0.8 \wedge 0.6/0 + 0.5 \wedge 0.6/0.4 + 0.6 \wedge 1/0.8 \\ &\quad + 0.7 \wedge 0.8/0 \\ &= 0.6/0.4 + 0.7/0 + 0.6/0.8 \ . \end{aligned}$$

There are two points related to the computation of $\tilde{R} \circ g$ that are in need of comment. First, if $\tilde{R}$ is sparsely tabulated in the sense that many of the possible n-tuples of values of $Q_1,\ldots,Q_n$ are not in the table, then the interpolation of R by the use of (4.23) may not yield a valid approximation to the answer to Q. And second, the result of substitution of

$$g = r_{i1} \times \cdots \times r_{in}$$

in $\tilde{R}$ would not, in general, be exactly equal to r_i -- as one might expect to be the case. As pointed out in [14], the cause of this phenomenon is the interference between the rows of $\tilde{R}$, which in turn is due to the fact that the fuzzy sets which constitute a column of R are not, in general, disjoint, that is, do not have an empty intersection.

An important assumption that underlies the procedure described in this section is that one has or can obtain a relational tableau which characterizes the dependence of the grade of membership of an object on the linguistic values of its attributes and/or the degree to which it possesses specified properties. The main contribution of the linguistic approach is that it makes it possible to describe this dependence in an approximate manner, using words rather than numbers as values of the relevant variables.

5. Cluster Analysis

Theory of fuzzy sets was first applied to cluster analysis by E. Ruspini [17]-[19]. More recently, J. Dunn and J. Bezdek have made a number of important contributions to this subject and have described effective algorithms for deriving optimal fuzzy partitions of a given set of sample points [20]-[32].

Viewed within the framework described in Section 2, cluster analysis differs from pattern classification in three essential respects.

First, the point of departure in cluster analysis is not -- as in pattern classification -- an opaque recognition algorithm in U^o which defines a fuzzy subset F of U^o, but a fuzzy similarity relation S^o which is a fuzzy subset of $U^o \times U^o$ and which is characterized by an opaque recognition algorithm R_{op}. Thus, when presented with two objects p and q in U^o, R_{op} yields the degree, $\mu_{S^o}(p,q)$, to which p and q are similar. The function $\mu_{S^o}: U^o \times U^o \to [0,1]$ is the membership function of the fuzzy relation S^o in U^o.

Second, the problem of cluster analysis includes as a subproblem the following problem in pattern classification.

Let p and q be objects in U^o and let $x \triangleq M(p)$ and $y \triangleq M(q)$ be their correspondents in the space of mathematical objects $U = \{M(p)\}$. The problem is to convert the opaque recognition algorithm R_{op} which acting on p and q yields

$$R_{op}(p,q) = \mu_{S^o}(p,q) \ , \tag{5.1}$$

into a transparent recognition algorithm R_{tr} which acting on x and y yields the same result as R_{op}, i.e.,

$$\begin{aligned} R_{tr}(x,y) &= R_{op}(p,q) \\ &= \mu_{S^o}(p,q) \ . \end{aligned} \tag{5.2}$$

It should be noted that this problem is of the same type as that formulated in Section 2, with the fuzzy subset S^o of $U^o \times U^o$ playing the role of F.

Third, assuming that we have R_{tr} -- which acts on elements of $U \times U$ -- the objective of cluster analysis is to derive from R_{tr} a number, say k, of transparent recognition algorithms $R_{tr_1},\ldots,R_{tr_k}$ -- acting on elements of U -- such that the fuzzy subsets (fuzzy clusters) $F_1,\ldots,F_k$ in U defined by $R_{tr_1},\ldots,R_{tr_k}$, have a property which may be stated as follows.

<u>Fuzzy Affinity Property</u>

Let $x = M(p)$ and $y = M(q)$ be mathematical objects in U corresponding to the objects p and q in U^o. Let $\{F_1,\ldots,F_k\}$ be a collection of well-separated[14] fuzzy subsets of U with membership functions $\mu_1,\ldots,\mu_n$, respectively. Then the F_i are <u>fuzzy clusters</u> induced by S^o if they have the <u>fuzzy affinity property</u> defined below.

(a) Both x and y have high grades of membership in some F_r, $r = 1,\ldots,k \Leftrightarrow (x,y)$ has a high grade of membership in S (the similarity relation induced in U by S^o).

[14] By well-separated we mean that if F_r and F_t are distinct fuzzy sets in $\{F_1,\ldots,F_k\}$, then every point of U has a low grade of membership in $F_r \cap F_t$.

(b) x has a high grade of membership in some F_r, $r = 1,\ldots,k$ and y has a high grade of membership in F_t, $t \neq r \Rightarrow (x,y)$ does not have a high grade of membership in S.

Stated less formally, the fuzzy affinity property implies that (a) if x and y have a high degree of similarity then they have a high grade of membership in some cluster, and vice-versa; and (b) if x and y have high grades of membership in different clusters then they do not have a high degree of similarity. It should be noted that this property of fuzzy clusters is more demanding than that implicit in the conventional definitions in which the degree of similarity of objects which belong to the same cluster is merely required to be greater than the degree of similarity between objects which belong to different clusters. Another point that should be noted is that, if we assumed that the only alternative to the consequent of (b) is "(x,y) has a high grade of membership in S," then (b) would be implied by (a) since the latter consequent would imply that x and y have a high grade of membership in some F_r -- which contradicts the antecedent of (b). Thus, by stating (b) we are tacitly assuming that (x,y) is not restricted to having either "high" or "not high" grades of membership in S. For example, the grade of membership of (x,y) in S could be "not high and not low."

An important implication of the fuzzy affinity property is the following. Suppose that x and y have high grades of membership in some fuzzy cluster F_r, and that z has a high grade of membership in a different fuzzy cluster, say F_t. Then, by (a) and (b), we have

similarity of x and y is high (5.3)

similarity of y and z is not high

similarity of x and z is not high

which implies that we could not have

similarity of x and y is high (5.4)

similarity of y and z is high

similarity of x and z is not high .

The inconsistency of the assertions in (5.4) is ruled out by the <u>fuzzy transitivity</u> of the similarity relation S which may be stated as[15]

similarity of x and z is at least as great as the (5.5)

similarity of x and y or the similarity of y and z .

Thus, if S has the fuzzy transitivity property and the similarities of both x and y and y and z are high, then the similarity of x and z must also be high.

Another point that should be noted is that the fuzzy affinity property does not require that the fuzzy clusters $\{F_1,\dots,F_k\}$ form a fuzzy partition in the sense of Ruspini. However, the stronger assumption that the F_r form a fuzzy partition makes it possible for Dunn and Bezdek to construct an effective algorithm for deriving from a fuzzy similarity relation a family of fuzzy clusters which form a fuzzy partition.

As described in [26], the Dunn-Bezdek fuzzy ISODATA algorithm may be stated as follows.

Let $\mu_1,\dots,\mu_k$ denote the membership functions of $F_1,\dots,F_k$, where the F_i, $i = 1,\dots,k$, are fuzzy subsets (clusters) of a finite subset, X,

---

[15] In more precise terms, the transitivity of a fuzzy relation R in U is defined by (see [13])

$$\mu_R(u,v) \geq \vee_w\left(\mu_R(u,w) \wedge \mu_R(w,v)\right) , \quad (u,v) \in U \times U$$

where $\mu_R(u,v)$ is the grade of membership of (u,v) in R, and $\vee_w$ is the supremum over $w \in U$.

of points in U. The fuzzy clusters $F_1,\dots,F_k$ form a <u>fuzzy k-partition of X</u> if and only if

$$\mu_1(x)+\cdots+\mu_k(x) = 1 , \quad x \in X \tag{5.6}$$

where + denotes the arithmetic sum. The goodness of a fuzzy partition is assumed to be assessed by the criterion functional

$$J(\mu) = \min_{v} \sum_{i=1}^{k} \sum_{x\in X} (\mu_i(x))^2 \|x-v_i\|^2 \tag{5.7}$$

where $\mu \triangleq (\mu_1,\dots,\mu_k)$, $v = (v_1,\dots,v_k)$, $v_i \in L$, and $L \triangleq$ vector space with inner product induced norm $\| \|$. Intuitively, the v_i represent the "centers" of $F_1,\dots,F_k$ and $J(\mu)$ provides a measure of the weighted dispersion of points in X in the relation to the optimal locations of the centers $v_1,\dots,v_k$.

<u>Step 1</u>: Choose a fuzzy partition $F_1,\dots,F_k$ characterized by k nonempty membership functions $\mu = (\mu_1,\dots,\mu_k)$, with $2 \le k \le n$.

<u>Step 2</u>: Compute the k weighted means (centers)

$$v_i = \frac{\sum_{x\in X} (\mu_i(x))^2 x}{\sum_{x\in X} (\mu_i(x))^2} , \quad 1 \le i \le k \tag{5.8}$$

where $x \in X \subset L$.

<u>Step 3</u>: Construct a new partition, $\hat{F}_1,\dots,\hat{F}_k$, characterized by $\hat{\mu} = (\hat{\mu}_1,\dots,\hat{\mu}_k)$, according to the following rule.

Let $I(x) \triangleq \{1 \le i \le k \mid v_i = x\}$. If $I(x)$ is not empty let $\hat{i}$ be the least integer $I(x)$ and put

$$\hat{\mu}_i(x) = 1 \quad \text{if} \quad i = \hat{i} \qquad (5.9)$$
$$= 0 \quad \text{if} \quad i \neq \hat{i}$$

for $1 \leq i \leq k$. Otherwise, if $I(x)$ is empty (the usual case), set

$$\hat{\mu}_i(x) = \frac{\dfrac{1}{\|x-v_i\|^2}}{\sum_{j=1}^{k} \left(\dfrac{1}{\|x-v_j\|^2}\right)} . \qquad (5.10)$$

Step 4: Compute some convenient measure, δ, of the defect between μ and $\hat{\mu}$. If $\delta \leq \varepsilon \triangleq$ a specified threshold, then stop. Otherwise go to Step 2.

In a number of papers [20]-[32], Bezdek and Dunn have studied the behavior of this and related algorithms and have established their convergence and other properties. Clearly, the work of Bezdek and Dunn on fuzzy clustering constitutes an important contribution to both the theory of cluster analysis and its practical applications.

Fuzzy Level-Sets

As was pointed out in [13], the conventional hierarchical clustering schemes [33] may be viewed as the resolution of a fuzzy similarity relation into a nested collection of nonfuzzy equivalence relations. To relate this result to the fuzzy affinity property, it is necessary to extend the notion of a level-set as defined in [13] to that of a fuzzy level-set. More specifically, let F be a fuzzy subset of U and let F_α, $0 < \alpha \leq 1$, be the α-level subset of U defined by

$$F_\alpha \triangleq \{x \mid \mu_F(x) \geq \alpha\} \qquad (5.11)$$

where μ_F is the membership function of F. We note that F_α -- which is a nonfuzzy set -- may be expressed equivalently as

$$F_\alpha = \mu_F^{-1}([\alpha,1]) \tag{5.12}$$

where μ_F^{-1} is the relation from [0,1] to U which is converse to μ_F, and F_α is the image of the interval $[\alpha,1]$ under this relation -- or, equivalently, multi-valued mapping -- μ_F^{-1}. It is easy to verify that in terms of the membership functions of F_α, F and $[\alpha,1]$, (5.12) translates into

$$\mu_{F_\alpha}(x) = \mu_{[\alpha,1]}(\mu_F(x)) \ , \quad x \in U \tag{5.13}$$

where μ_{F_α} and $\mu_{[\alpha,1]}$ denote the membership (characteristic) functions of the nonfuzzy sets F_α and $[\alpha,1]$, respectively.

Now suppose that α is a fuzzy subset of [0,1] labeled, say, high, with μ_{high} defined by (see A17)

$$\mu_{high}(v) = S(v;0.6,0.7,0.8) \ , \quad 0 \le v \le 1 \ . \tag{5.14}$$

When α is a fuzzy subset of [0,1], the fuzzy set $\ge \alpha$ may be expressed as the composition of the nonfuzzy binary relation $\ge$ with the unary fuzzy relation α. Thus, if $\alpha \triangleq$ high, then

$$\begin{aligned} \ge\alpha &= \ \ge \circ\, \alpha \\ &= \ \ge \circ\, high \\ &= high \end{aligned} \tag{5.15}$$

since the membership function of high is monotone nondecreasing in v. Correspondingly, the expression for the membership function of the fuzzy level set F_{high} becomes (see A73)

$$\mu_{F_{high}}(x) = \mu_{high}(\mu_F(x)) \ . \tag{5.16}$$

To relate this result to the fuzzy affinity property, we note that if the objects x, y in U have a high degree of similarity, then the ordered pair (x,y) has a high grade of membership in the fuzzy similarity relation S. Thus, by analogy with (5.12), the set of pairs (x,y) in $U \times U$ which have a high grade of membership in S form a fuzzy level-set of S defined by

$$S_{high} = \mu_S^{-1}(\mu_{high}) \tag{5.17}$$

or, equivalently,

$$\mu_{S_{high}}(x,y) = \mu_{high}(\mu_S(x,y)) \ . \tag{5.18}$$

This expression makes it possible to derive in a straight-forward fashion the fuzzy level-set S_{high} from the similarity relation S.

An important property of S_{high} may be stated as the

<u>Proposition</u>. If S is a transitive fuzzy relation, so is S_{high}.

The validity of this proposition is readily established by observing that the transitivity of S means that (see (5.5))

$$\mu_S(x,y) \geq \vee_z \mu_S(x,z) \wedge \mu_S(z,y) \ , \quad x,\ y,\ z \in U \ . \tag{5.19}$$

Now, (5.19) implies and is implied by

$$\forall z \ \left(\mu_S(x,y) \geq \mu_S(x,z) \wedge \mu_S(z,y)\right) \tag{5.20}$$

which in turn implies and is implied by

$$\forall z \ \left(\mu_S(x,y) \geq \mu_S(x,z) \text{ or } \mu_S(x,y) \geq \mu_S(z,y)\right) \ . \tag{5.21}$$

Since μ_{high} is a monotone nondecreasing function, we have

$$\mu_S(x,y) \geq \mu_S(x,z) \Rightarrow \mu_{high}(\mu_S(x,y)) \geq \mu_{high}(\mu_S(x,z)) \quad (5.22)$$

and

$$\mu_S(x,y) \geq \mu_S(z,y) \Rightarrow \mu_{high}(\mu_S(x,y)) \geq \mu_{high}(\mu_S(z,y)) \quad (5.23)$$

and hence

$$\forall z \ (\mu_{S_{high}}(x,y) \geq \mu_{S_{high}}(x,z) \text{ or } \mu_{S_{high}}(x,y) \geq \mu_{S_{high}}(z,y)) \quad (5.24)$$

which by (5.21) and (5.20) leads to the conclusion that S_{high} is transitive.

Basically, the employment of fuzzy level-sets for purposes of clustering may be viewed as an application of a form of contrast intensification [34] to a fuzzy similarity relation which defines the degrees of similarity of mathematical objects in U. Thus, given a collection of such objects, we can derive S_{high} from S by the use of (5.18) and then apply a Dunn-Bezdek type of fuzzy clustering algorithm to group the given collection of objects into a set of fuzzy clusters $\{F_1,\ldots,F_k\}$.

6. Concluding Remarks

In the foregoing discussion, we have touched upon only a few of the many basic issues which arise in the application of the theory of fuzzy sets to pattern classification and cluster analysis. Although this is not yet the case at present, it is very likely that in the years ahead it will be widely recognized that most of the problems in pattern classification and cluster analysis are intrinsically fuzzy in nature and that the conceptual framework of the theory of fuzzy sets provides a natural setting both for the formulation of such problems and their solution by fuzzy-algorithmic techniques.

Appendix

Fuzzy Sets -- Notation, Terminology and Basic Properties

The symbols $U, V, W, \ldots,$ with or without subscripts, are generally used to denote specific universes of discourse, which may be arbitrary collections of objects, concepts or mathematical constructs. For example, U may denote the set of all real numbers; the set of all residents in a city; the set of all sentences in a book; the set of all colors that can be perceived by the human eye, etc.

Conventionally, if A is a fuzzy subset of U whose elements are $u_1, \ldots, u_n$, then A is expressed as

$$A = \{u_1, \ldots, u_n\} \ . \tag{A1}$$

For our purposes, however, it is more convenient to express A as

$$A = u_1 + \cdots + u_n \tag{A2}$$

or

$$A = \sum_{i=1}^{n} u_i \tag{A3}$$

with the understanding that, for all i, j,

$$u_i + u_j = u_j + u_i \tag{A4}$$

and

$$u_i + u_i = u_i \ . \tag{A5}$$

As an extension of this notation, a finite <u>fuzzy</u> subset of U is expressed as

$$F = \mu_1 u_1 + \cdots + \mu_n u_n \tag{A6}$$

or, equivalently, as

$$F = \mu_1/u_1 + \cdots + \mu_n/u_n \tag{A7}$$

where the μ_i, $i = 1,\ldots,n$, represent the grades of membership of the u_i in F. Unless stated to the contrary, the μ_i are assumed to lie in the interval [0,1], with 0 and 1 denoting no membership and full membership, respectively.

Consistent with the representation of a finite fuzzy set as a linear form in the u_i, an arbitrary fuzzy subset of U may be expressed in the form of an integral

$$F = \int_U \mu_F(u)/u \tag{A8}$$

in which $\mu_F: U \to [0,1]$ is the membership or, equivalently, the compatibility function of F; and the integral $\int_U$ denotes the union (defined by (A28)) of fuzzy singletons $\mu_F(u)/u$ over the universe of discourse U.

The points in U at which $\mu_F(u) > 0$ constitute the support of F. The points at which $\mu_F(u) = 0.5$ are the crossover points of F.

Example A9. Assume

$$U = a + b + c + d\ . \tag{A10}$$

Then, we may have

$$A = a + b + d \tag{A11}$$

and

$$F = 0.3a + 0.9b + d \tag{A12}$$

as nonfuzzy and fuzzy subsets of U, respectively.

If

$$U = 0 + 0.1 + 0.2 + \cdots + 1 \tag{A13}$$

then a fuzzy subset of U would be expressed as, say,

$$F = 0.3/0.5 + 0.6/0.7 + 0.8/0.9 + 1/1\ . \tag{A14}$$

If $U = [0,1]$, then F might be expressed as

$$F = \int_0^1 \frac{1}{1+u^2}/u \tag{A15}$$

which means that F is a fuzzy subset of the unit interval $[0,1]$ whose membership function is defined by

$$\mu_F(u) = \frac{1}{1+u^2} . \tag{A16}$$

In many cases, it is convenient to express the membership function of a fuzzy subset of the real line in terms of a standard function whose parameters may be adjusted to fit a specified membership function in an approximate fashion. Two such functions are defined below.

$$\begin{aligned} S(u;\alpha,\beta,\gamma) &= 0 && \text{for } u \leq \alpha \\ &= 2\left(\frac{u-\alpha}{\gamma-\alpha}\right)^2 && \text{for } \alpha \leq u \leq \beta \\ &= 1 - 2\left(\frac{u-\gamma}{\gamma-\alpha}\right)^2 && \text{for } \beta \leq u \leq \gamma \\ &= 1 && \text{for } u \geq \gamma \end{aligned} \tag{A17}$$

$$\begin{aligned} \pi(u;\beta,\gamma) &= S(u;\gamma-\beta,\gamma-\tfrac{\beta}{2},\gamma) && \text{for } u \leq \gamma \\ &= 1 - S(u;\gamma,\gamma+\tfrac{\beta}{2},\gamma+\beta) && \text{for } u \geq \gamma . \end{aligned} \tag{A18}$$

In $S(u;\alpha,\beta,\gamma)$, the parameter β, $\beta = \frac{a+\gamma}{2}$, is the crossover point. In $\pi(u;\beta,\gamma)$, β is the bandwidth, that is the separation between the crossover points of π, while γ is the point at which π is unity.

In some cases, the assumption that μ_F is a mapping from U to $[0,1]$ may be too restrictive, and it may be desirable to allow μ_F to take values in a lattice or, more particularly, in a Boolean algebra. For most purposes, however, it is sufficient to deal with the first two of the following hierarchy of fuzzy sets.

Definition A19. A fuzzy subset, F, of U is of type 1 if its membership function, μ_F, is a mapping from U to [0,1]; and F is of type n, n = 2,3,..., if μ_F is a mapping from U to the set of fuzzy subsets of type n-1. For simplicity, it will always be understood that F is of type 1 if it is not specified to be of a higher type.

Example A20. Suppose that U is the set of all nonnegative integers and F is a fuzzy subset of U labeled small integers. Then F is of type 1 if the grade of membership of a generic element u in F is a number in the interval [0,1], e.g.,

$$\mu_{\underline{small}\ \underline{integers}}(u) = \left(1 + \left(\frac{u}{5}\right)^2\right)^{-1} , \quad u = 0,1,2,\ldots . \tag{A21}$$

On the other hand, F is of type 2 if for each u in U, $\mu_F(u)$ is a fuzzy subset of [0,1] of type 1, e.g., for u = 10,

$$\mu_{\underline{small}\ \underline{integers}}(10) = \underline{low} \tag{A22}$$

where low is a fuzzy subset of [0,1] whose membership function is defined by, say,

$$\mu_{\underline{low}}(v) = 1 - S(v;0,0.25,0.5) , \quad v \in [0,1] \tag{A23}$$

which implies that

$$\underline{low} = \int_0^1 (1 - S(v;0,0.25,0.5))/v . \tag{A24}$$

If F is a fuzzy subset of U, then its α-level-set, F_α, is a nonfuzzy subset of U defined by

$$F_\alpha = \{u \mid \mu_F(u) \geq \alpha\} \tag{A25}$$

for $0 < \alpha \leq 1$.

If U is a linear vector space, the F is <u>convex</u> if and only if for all $\lambda \in [0,1]$ and all u_1, u_2 in U,

$$\mu_F(\lambda u_1 + (1-\mu)u_2) \geq \min(\mu_F(u_1), \mu_F(u_2)) \ . \tag{A26}$$

In terms of the level-sets of F, F is convex if and only if the F_α are convex for all $\alpha \in (0,1]$.[26]

The relation of containment for fuzzy subsets F and G of U is defined by

$$F \subset G \Leftrightarrow \mu_F(u) \leq \mu_G(u) \ , \quad u \in U \ . \tag{A27}$$

Thus, F is a fuzzy subset of G if (A27) holds for all u in U.

<u>Operations on Fuzzy Sets</u>

If F and G are fuzzy subsets of U, their <u>union</u>, $F \cup G$, <u>intersection</u>, $F \cap G$, <u>bounded-sum</u>, $F \oplus G$, and <u>bounded-difference</u>, $F \ominus G$, are fuzzy subsets of U defined by

$$F \cup G \triangleq \int_U \mu_F(u) \vee \mu_G(u)/u \tag{A28}$$

$$F \cap G \triangleq \int_U \mu_F(u) \wedge \mu_G(u)/u \tag{A29}$$

$$F \oplus G \triangleq \int_U 1 \wedge (\mu_F(u) + \mu_G(u))/u \tag{A30}$$

$$F \ominus G \triangleq \int_U 0 \vee (\mu_F(u) - \mu_G(u))/u \tag{A31}$$

---

[26] This definition of convexity can readily be extended to fuzzy sets of type 2 by applying the extension principle (see (A70)) to (A26).

where $\vee$ and $\wedge$ denote max and min, respectively. The complement of F is defined by

$$F' = \int_U (1-\mu_F(u))/u \tag{A32}$$

or, equivalently,

$$F' = U \ominus F . \tag{A33}$$

It can readily be shown that F and G satisfy the identities

$$(F \cap G)' = F' \cup G' \tag{A34}$$

$$(F \cup G)' = F' \cap G' \tag{A35}$$

$$(F \oplus G)' = F' \ominus G \tag{A36}$$

$$(F \ominus G)' = F' \oplus G \tag{A37}$$

and that F satisfies the resolution identity

$$F = \int_0^1 \alpha F_\alpha \tag{A38}$$

where F_α is the α-level-set of F; αF_α is a set whose membership function is $\mu_{\alpha F_\alpha} = \alpha\mu_{F_\alpha}$, and $\int_0^1$ denotes the union of the αF, with $\alpha \in (0,1]$.

Although it is traditional to use the symbol $\cup$ to denote the union of nonfuzzy sets, in the case of fuzzy sets it is advantageous to use the symbol $+$ in place of $\cup$ where no confusion with the arithmetic sum can result. This convention is employed in the following example, which is intended to illustrate (A28), (A29), (A30), (A31) and (A32).

Example A39. For U defined by (A10) and F and G expressed by

$$F = 0.4a + 0.9b + d \tag{A40}$$

$$G = 0.6a + 0.5b \tag{A41}$$

we have

$$F + G = 0.6a + 0.9b + d \tag{A42}$$

$$F \cap G = 0.4a + 0.5b \tag{A43}$$

$$F \oplus G = a + b + d \tag{A44}$$

$$F \ominus G = 0.4b + d \tag{A45}$$

$$F' = 0.6a + 0.1b + c \tag{A46}$$

The linguistic connectives and (conjunction) and or (disjunction) are identified with $\cap$ and $+$, respectively. Thus,

$$F \text{ and } G \triangleq F \cap G \tag{A47}$$

and

$$F \text{ or } G \triangleq F + G \ . \tag{A48}$$

As defined by (A47) and (A48), and and or are implied to be noninteractive in the sense that there is no "trade-off" between their operands. When this is not the case, and and or are denoted by and* and or* respectively, and are defined in a way that reflects the nature of the trade-off. For example, we may have

$$F \ \underline{\text{and}}^* \ G \triangleq \int_U \mu_F(u)\mu_G(u)/u \tag{A49}$$

$$F \ \underline{\text{or}}^* \ G \triangleq \int_U (\mu_F(u) + \mu_G(u) - \mu_F(u)\mu_G(u))/u \tag{A50}$$

whose $+$ denotes the arithmetic sum. In general, the interactive versions of and and or do not possess the simplifying properties of the connectives defined by (A47) and (A48), e.g., associativity, distributivity, etc.

If α is a real number, then F^{α} is defined by

$$F^{\alpha} \triangleq \int_{U} (\mu_F(n))^{\alpha}/u \ . \tag{A51}$$

For example, for the fuzzy set defined by (A40), we have

$$F^2 = 0.16a + 0.81b + d \tag{A52}$$

and

$$F^{1/2} = 0.63a + 0.95b + d \ . \tag{A53}$$

These operations may be used to approximate, very roughly, the effect of the linguistic modifiers <u>very</u> and <u>more or less</u>. Thus,

$$\underline{\text{very}}\ F \triangleq F^2 \tag{A54}$$

and

$$\underline{\text{more or less}}\ F \triangleq F^{1/2} \ . \tag{A55}$$

If $F_1,\ldots,F_n$ are fuzzy subsets of $U_1,\ldots,U_n$, then the <u>cartesian product</u> of $F_1,\ldots,F_n$ is a fuzzy subset of $U_1 \times \cdots \times U_n$ defined by

$$F_1 \times \cdots \times F_n = \int_{U_1 \times \cdots \times U_n} (\mu_{F_1}(u_1) \wedge \cdots \wedge \mu_{F_n}(u_n))/(u_1,\ldots,u_n) \ . \tag{A56}$$

As an illustration, for the fuzzy sets defined by (A40) and (A41), we have

$$\begin{aligned} F \times G &= (0.4a + 0.9b + d) \times (0.6a + 0.5b) \\ &= 0.4/(a,a) + 0.4/(a,b) + 0.6/(b,a) \\ &\quad + 0.5/(b,b) + 0.6/(d,a) + 0.5/(d,b) \end{aligned} \tag{A57}$$

which is a fuzzy subset of $(a+b+c+d) \times (a+b+c+d)$.

Fuzzy Relations

An n-ary fuzzy relation R in $U_1 \times \cdots \times U_n$ is a fuzzy subset of $U_1 \times \cdots \times U_n$. The projection of R on $U_{i_1} \times \cdots \times U_{i_k}$, where $(i_1,\ldots,i_k)$ is a subsequence of $(1,\ldots,n)$, is a relation in $U_{i_1} \times \cdots \times U_{i_k}$ defined by

$$\text{Proj } R \text{ on } U_{i_1} \times \cdots \times U_{i_k} \triangleq \int_{U_{i_1} \times \cdots \times U_{i_k}} \vee_{u_{j_1},\ldots,u_{j_\ell}} \mu_R(u_1,\ldots,u_n)/(u_1,\ldots,u_n) \quad \text{(A58)}$$

where $(j_1,\ldots,j_\ell)$ is the sequence complementary to $(i_1,\ldots,i_k)$ (e.g., if $n = 6$ then $(1,3,6)$ is complementary to $(2,4,5)$), and $\vee_{u_{j_1},\ldots,u_{j_\ell}}$ denotes the supremum over $U_{j_1} \times \cdots \times U_{j_\ell}$.

If R is a fuzzy subset of $U_{i_1},\ldots,U_{i_k}$, then its cylindrical extension in $U_1 \times \cdots \times U_n$ is a fuzzy subset of $U_1 \times \cdots \times U_n$ defined by

$$\bar{R} = \int_{U_1 \times \cdots \times U_n} \mu_R(U_{i_1},\ldots,U_{i_k})/(u_1,\ldots,u_n) \; . \quad \text{(A59)}$$

In terms of their cylindrical extensions, the composition of two binary relations R and S (in $U_1 \times U_2$ and $U_2 \times U_3$, respectively) is expressed by

$$R \circ S = \text{Proj } \bar{R} \cap \bar{S} \text{ on } U_1 \times U_3 \quad \text{(A60)}$$

where $\bar{R}$ and $\bar{S}$ are the cylindrical extensions of R and S in $U_1 \times U_2 \times U_3$. Similarly, if R is a binary relation in $U_1 \times U_2$ and S is a unary relation in U_2, their composition is given by

$$R \circ S = \text{Proj } R \cap \bar{S} \text{ on } U_1 \; . \quad \text{(A61)}$$

Example A62. Let R be defined by the right-hand member of (A57) and

$$S = 0.4a + b + 0.8d \ . \tag{A63}$$

Then

$$\text{Proj } R \text{ on } U_1 \ (\triangleq a + b + c + d) = 0.4a + 0.6b + 0.6d \tag{A64}$$

and

$$R \circ S = 0.4a + 0.5b + 0.5d \ . \tag{A65}$$

Linguistic Variables

Informally, a linguistic variable, χ, is a variable whose values are words or sentences in a natural or artificial language. For example, if *age* is interpreted as a linguistic variable, then its *term-set*, $T(\chi)$, that is, the set of linguistic values, might be

$$\begin{aligned} T(\text{age}) = {} & \text{young} + \text{old} + \text{very young} + \text{not young} \\ & + \text{very old} + \text{very very young} \\ & + \text{rather young} + \text{more or less young} + \cdots \end{aligned} \tag{A66}$$

where each of the terms in $T(\text{age})$ is a label of a fuzzy subset of a universe of discourse, say $U = [0,100]$.

A linguistic variable is associated with two rules: (a) a *syntactic rule*, which defines the well-formed sentences in $T(\chi)$; and (b) a *semantic rule*, by which the meaning of the terms in $T(\chi)$ may be determined. If X is a term in $T(\chi)$, then its *meaning* (in a denotational sense) is a subset of U. A *primary term* in $T(\chi)$ is a term whose meaning is a *primary fuzzy set*, that is, a term whose meaning must be defined a priori, and which serves as a basis for the computation of the meaning of the non-primary terms in $T(\chi)$. For example, the primary terms in (A66) are *young* and *old*, whose meaning might be defined by their respective compatibility functions μ_{young} and μ_{old}. From these, then, the meaning -- or, equivalently, the compatibility functions -- of the non-primary terms in (A66) may be computed by the application of a semantic rule. For example, employing (A54) and (A55) we have

$$\mu_{\text{very young}} = (\mu_{young})^2 \tag{A67}$$

$$\mu_{\text{more or less old}} = (\mu_{old})^{1/2} \tag{A68}$$

$$\mu_{\text{not very young}} = 1 - (\mu_{young})^2 \ . \tag{A69}$$

The Extension Principle

Let g be a mapping from U to V. Thus,

$$v = g(u) \tag{A70}$$

where u and v are generic elements of U and V, respectively.

Let F be a fuzzy subset of U expressed as

$$F = \mu_1 u_1 + \cdots + \mu_n u_n \tag{A71}$$

or, more generally,

$$F = \int_U \mu_F(u)/u \ . \tag{A72}$$

By the extension principle, the image of F under g is given by

$$g(F) = \mu_1 g(u_1) + \cdots + \mu_n g(u_n) \tag{A73}$$

or, more generally,

$$g(F) = \int_U \mu_F(u)/g(u) \ . \tag{A74}$$

Similarly, if g is a mapping from $U \times V$ to W, and F and G are fuzzy subsets of U and V, respectively, then

$$g(F,G) = \int_W (\mu_F(u) \wedge \mu_G(v))/g(u,v) \; . \tag{A75}$$

Example A76. Assume that g is the operation of squaring. Then, for the set defined by (A14), we have

$$\begin{aligned} & g(0.3/0.5 + 0.6/0.7 + 0.8/0.9 + 1/1) \\ & \quad = 0.3/0.25 + 0.6/0.49 + 0.8/0.81 + 1/1 \; . \end{aligned} \tag{A77}$$

Similarly, for the binary operation $\vee$ ($\triangleq$ max), we have

$$\begin{aligned} & (0.9/0.1 + 0.2/0.5 + 1/1) \vee (0.3/0.2 + 0.8/0.6) \\ & \quad = 0.3/0.2 + 0.2/0.5 + 0.8/1 + 0.8/0.6 + 0.2/0.6 \; . \end{aligned} \tag{A78}$$

It should be noted that the operation of squaring in (A77) is different from that of (A51) and (A52).

References

[1] L.A. Zadeh, "Outline of a New Approach to the Analysis of Complex Systems and Decision Processes," IEEE Trans. on Systems, Man and Cybernetics *SMC-3* (1973), 28-44.

[2] L.A. Zadeh, "The Concept of a Linguistic Variable and Its Application to Approximate Reasoning," Information Sciences, Part I, *8* (1975), 199-249; Part II, *8* (1975), 301-357; Part III, *9* (1975), 43-80.

[3] L.A. Zadeh, "Fuzzy Logic and Approximate Reasoning (In Memory of Grigore Moisil)," Synthese *30* (1975), 407-428.

[4] R.E. Bellman and L.A. Zadeh, "Local and Fuzzy Logics," Memorandum No. ERL-M584, Electronics Research Laboratory, University of California, Berkeley, 1976. (To appear in the Proceedings of the International Symposium on Multi-Valued Logic, University of Indiana, 1975).

[5] U. Grenander, Pattern Synthesis: Lectures in Pattern Theory (vol. 1), Springer-Verlag, New York, 1976.

[6] K.S. Fu, Syntactic Methods in Pattern Recognition, Academic Press, New York, 1974.

[7] E.T. Lee and L.A. Zadeh, "Note on Fuzzy Languages," Information Sciences *1* (1969), 421-434.

[8] R.E. Bellman, R. Kalaba and L.A. Zadeh, "Abstraction and Pattern Classification," J. Math. Anal. and Appl. *13* (1966), 1-7.

[9] M. Minsky and S. Papert, Perceptrons, M.I.T. Press, Cambridge, 1969.

[10] M.A. Aizerman, E.M. Braverman and L.I. Rozonoer, Method of Potential Functions in the Theory of Learning Machines, Science Press, Moscow, 1974.

[11] A.G. Arkad'ev and E.M. Braverman, Computers and Pattern Recognition, Thompson Book Co., Washington, D.C., 1967.

[12] V.H. Vapnik and A. Ya. Chervonenkis, Theory of Pattern Recognition, Science Press, Moscow, 1974.

[13] L.A. Zadeh, "Similarity Relations and Fuzzy Orderings," Information Sciences *3* (1971), 177-200.

[14] L.A. Zadeh, "Calculus of Fuzzy Restrictions," Proc. U.S.-Japan Seminar on Fuzzy Sets and Their Applications, L.A. Zadeh, K.S. Fu, K. Tanaka, M. Shimura (eds.), Academic Press, New York, 1975, 1-39.

[15] R.F. Boyce, D.D. Chamberlin, W.F. King III and M.M. Hammer, "Specifying Queries as Relational Expressions: SQUARE," IBM Research Report RJ1291, 1973.

[16] D.D. Chamberlin and R.F. Boyce, "SEQUEL: A Structured English Query Language," Proc. ACM SIGFIDET Workshop on Data Description, Access and Control, 1975. (Also see IBM Research Report RJ1318, 1973.)

[17] E.R. Ruspini, "A New Approach to Clustering," Information and Control *15* (1969), 22-32.

[18] E.R. Ruspini, "Numerical Methods for Fuzzy Clustering," Information Sciences *2* (1970), 319-350.

[19] E.R. Ruspini, "New Experimental Results in Fuzzy Clustering," Information Sciences *6* (1973), 273-284.

[20] J.C. Bezdek, "Fuzzy Mathematics in Pattern Classification," Ph.D. Dissertation, Center for Applied Mathematics, Cornell University, Ithaca, New York, 1973.

[21] J.C. Bezdek, "Numerical Taxonomy with Fuzzy Sets," J. Math. Biology *1* (1974), 57-71.

[22] J.C. Dunn, "Some Recent Investigations of a Fuzzy Partitioning Algorithm and Its Application to Pattern Classification Problems," Center for Applied Mathematics, Cornell University, Ithaca, New York, 1974.

[23] J.C. Dunn, "A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well Separated Clusters," J. of Cybernetics *3* (1974), 32-57.

[24] J.C. Dunn, "Well Separated Clusters and Optimal Fuzzy Partitions," Center for Applied Mathematics, Cornell University, Ithaca, New York, 1974.

[25] J.C. Dunn, "A Graph-Theoretic Analysis of Pattern Classification via Tamura's Fuzzy Relation," IEEE Trans. on Systems, Man and Cybernetics *SMC-4* (1974), 310-313.

[26] J.C. Bezdek and J.C. Dunn, "Optimal Fuzzy Partitions: A Heuristic for Estimating the Parameters in a Mixture of Normal Distributions," IEEE Trans. on Computers *C-24* (1975), 835-838.

[27] J.C. Bezdek, "Cluster Validity with Fuzzy Sets," J. of Cybernetics *3* (1974), 58-73.

[28] J.C. Bezdek and J.D. Harris, "Convex Decompositions of Fuzzy Partitions," Departments of Mathematics, Utah State University, Logan, Utah and Marquette University, Milwaukee, Wisconsin, 1976.

[29] J.C. Bezdek, "A Physical Interpretation of Fuzzy ISODATA," IEEE Trans. on Systems, Man and Cybernetics *SMC-6* (1976), 387-390.

[30] J.C. Bezdek, "Mathematical Models for Systematics and Taxonomy," Proc. 8th International Conference on Numerical Taxonomy, G. Estabrook (ed.), Freeman Co., San Francisco, 1975, 143-164.

[31] J.C. Bezdek, "Feature Selection for Binary Data: Medical Diagnosis with Fuzzy Sets," Proc. National Computer Conference, S. Winkler (ed.), AFIPS Press, Montvale, N.J., 1976, 1057-1058.

[32] J.C. Bezdek and P.F. Castelaz, "Prototype Classification and Feature Selection with Fuzzy Sets," Department of Mathematics, Utah State University, Logan, Utah and Department of Electrical Engineering, Marquette University, Milwaukee, Wisconsin, 1976.

[33] S.C. Johnson, "Hierarchical Clustering Schemes," Psychometrica *32* (1967), 241-254.

[34] L.A. Zadeh, "A Fuzzy-Set-Theoretic Interpretation of Linguistic Hedges," J. of Cybernetics *2* (1972), 4-34.

[35] R. Sokal and P. Sneath, Principles of Numerical Taxonomy, Freeman Co., San Francisco, 1963.

[36] Methodologies of Pattern Recognition, S. Watanabe (ed.), Academic Press, New York, 1969.

[37] R. Jardine and R. Sibson, Mathematical Taxonomy, Wiley, New York, 1971.

[38] K. Fukunaga, Introduction to Statistical Pattern Recognition, Academic Press, New York, 1972.

[39] R. Duda and P. Hart, Pattern Classification and Scene Analysis, Wiley-Interscience, New York, 1973.

[40] J. Tou and R. Gonzales, Pattern Recognition Principles, Addison-Wesley, Reading, Mass., 1974.

[41] E.A. Patrick, Fundamentals of Pattern Recognition, Prentice-Hall, Englewood Cliffs, N.J., 1972.

[42] L.M. Uhr, Pattern Recognition, Learning and Thought, Prentice-Hall, Englewood Cliffs, N.J., 1973.

[43] W. Meisel, Computer-Oriented Approaches to Pattern Recognition, Academic Press, New York, 1972.

[44] G. Nagy, "State of the Art in Pattern Recognition," Proc. IEEE *56* (1969), 836-862.

[45] Y.C. Ho and A.K. Agrawala, "On Pattern Classification Algorithms: Introduction and Survey," Proc. IEEE *56* (1968), 2101-2114.

[46] L. Kanal, "Patterns in Pattern Recognition: 1968-1974," IEEE Trans. on Information Theory *IT-20* (1974), 697-722.

[47] A. Schroeder, "Recognition of Components of a Mixture," Ph.D. Dissertation, University of Paris VI, 1974.

[48] E. Diday, "New Methods and New Concepts in Automatic Classification and Pattern Recognition," Ph.D. Dissertation, University of Paris VI, 1972.

[49] V.I. Loginov, "Probability Treatment of Zadeh's Membership Functions and Their Use in Pattern Recognition," Engineering Cybernetics (1966), 68-69.

[50] C.L. Chang, "Fuzzy Sets and Pattern Recognition," Ph.D. Dissertation, Department of Electrical Engineering, University of California, Berkeley, 1967.

[51] W.G. Wee, "On a Generalization of Adaptive Algorithms and Applications of the Fuzzy Set Concept to Pattern Classification," Technical Report 67-7, Department of Electrical Engineering, 1967.

[52] R.H. Flake and B.L. Turner, "Numerical Classification for Taxonomic Problems," J. Theo. Biol. *20* (1968), 260-270.

[53] A.N. Borisòv and E.A. Kokle, "Recognition of Fuzzy Patterns by Feature Analysis," Cybernetics and Diagnostics, No. 4, Riga, U.S.S.R., 1970.

[54] I. Gitman and M.D. Levine, "An Algorithm for Detecting Unimodal Fuzzy Sets and Its Applications as a Clustering Technique," IEEE Trans. on Computers *C-19* (1970), 583-593.

[55] S. Otsuki, "A Model for Learning and Recognizing Machine," Information Processing *11* (1970), 664-671.

[56] S.K. Chang, "Automated Interpretation and Editing of Fuzzy Line Drawings," Proc. Spring Joint Conf. (1971), 393-399.

[57] S. Tamura, S. Niguchi and K. Tanaka, "Pattern Classification Based on Fuzzy Relations," IEEE Trans. on Systems, Man and Cybernetics *SMC-1* (1971), 937-944.

[58] S. Tamura, "Fuzzy Pattern Classification," Proc. Symp. on Fuzziness in Systems and Its Processing (1971), Tokyo.

[59] E.T. Lee, "Proximity Measures for the Classification of Geometric Figures, J. of Cybernetics *2* (1972), 43-59.

[60] M. Shimura, "Application of Fuzzy Functions to Pattern Classification," Trans. IECE *55-d* (1972), 218-225.

[61] M. Sugeno, "Evaluation of Similarity of Patterns by Fuzzy Integrals," Ann. Conf. Records of SICE, Tokyo, 1972.

[62] K. Kotoh and K. Hiramatsu, "A Representation of Pattern Classes Using the Fuzzy Sets," Trans. IECE *56-d* (1973), 275-282.

[63] L.E. Larsen, E. Ruspini, J.J. McNew, D.O. Walter and W.R. Adey, "A Test of Sleep Staging Systems in the Unrestrained Chimpanzee," Brain Research *40* (1972), 319-343.

[64] P. Siy, "Fuzzy Logic and Hard-Written Character Recognition," Ph.D. Dissertation, Department of Electrical Engineering, University of Akron, Ohio, 1972.

[65] R.K. Ragade, "A Multiattribute Perception and Classification of Visual Similarities," Systems Res. and Planning Papers, S-001-73, Bell Northern Research, Canada, 1973.

[66] C.C. Negoita, "On the Application of the Fuzzy Sets Separation Theorem for Automatic Classification in Information Retrieval Systems," Information Sciences *5* (1973), 279-286.

[67] M. Sugeno, "Constructing Fuzzy Measure and Grading Similarity of Patterns by Fuzzy Integrals," Trans. SICE *9* (1973), 359-367.

[68] Y. Noguchi, "Pattern Recognition Systems Based on the Feature Extraction Technique," Report No. 739, Electrotechnical Laboratory, Tokyo, 1973.

[69] B. Conche, "A Method of Classification Based on the Use of a Fuzzy Automaton," University of Paris - Dauphine, 1973.

[70] N. Malvache, "Analysis and Identification of Visual Systems in Humans," Ph.D. Dissertation, University of Lille, Lille, 1973.

[71] N. Okada and T. Tamachi, "Automated Editing of Fuzzy Line Drawings for Picture Description," Trans. IECE *57-a* (1974), 216-223.

[72] M. Woodbury and J. Clive, "Clinical Pure Types vs. a Fuzzy Partition," J. of Cybernetics *4* (1974), 111-120.

[73] P. Siy and C.S. Cheu, "Fuzzy Logic for Handwirtten Numerical Character Recognition," IEEE Trans. on Systems, Man and Cybernetics *SMC-4* (1974), 570-575.

[74] M.G. Thomason, "Finite Fuzzy Automata, Regular Fuzzy Languages, and Pattern Recognition," Department of Electrical Engineering, Duke University, Durham, N.C., 1974.

[75] A. Kaufmann, Introduction to the Theory of Fuzzy Subsets (vol. III): Applications to Classification, Pattern Recognition, Automata and Systems, Masson, Paris, 1975.

[76] C.V. Negoita and D.A. Ralescu, Theory of Fuzzy Sets and Its Applications, Wiley, New York, 1975.

[77] T. Pavlidis, "Fuzzy Representations as Means of Overcoming the Overcommitment of Segmentation," Proc. Conf. on Computer Graphics, Pattern Recognition and Data Structures, Los Angeles, Calif., 1975.

[78] A. Rosenfeld, "Fuzzy Graphs," Proc. U.S.-Japan Seminar on Fuzzy Sets and Their Applications, L.A. Zadeh, K.S. Fu, K. Tanaka and M. Shimura (eds.), Academic Press, New York, 1975, 77-95.

[79] E.T. Lee, "Shape-Oriented Chromosome Identification," IEEE Trans. on Systems, Man and Cybernetics *SMC-5* (1975), 629-632.

[80] R.T. Yeh and S.Y. Bang, "Fuzzy Relations, Fuzzy Graphs, and Their Applications to Clustering Analysis," Proc. U.S.-Japan Seminar on Fuzzy Sets and Their Applications, L.A. Zadeh, K.S. Fu, K. Tanaka and M. Shimura (eds.), Academic Press, New York, 1975, 125-149.

[81] G.F. DePalma and S.S. Yau, "Fractionally Fuzzy Grammars with Application to Pattern Recognition," Proc. U.S.-Japan Seminar on Fuzzy Sets and Their Applications, L.A. Zadeh, K.S. Fu, K. Tanaka and M. Shimura (eds.), Academic Press, New York, 1975, 329-351.

[82] M. Shimura, "An Approach to Pattern Recognition and Associative Memories Using Fuzzy Logic," Proc. U.S.-Japan Seminar on Fuzzy Sets and Their Applications, L.A. Zadeh, K.S. Fu, K. Tanaka and M. Shimura (eds.), Academic Press, New York, 1975, 449-476.

[83] R.L. Chang and T. Pavlidis, "Fuzzy Decision Trees," Technical Report No. 203, Department of Electrical Engineering and Computer Science, Princeton University, Princeton, 1976.

[84] S. Sugeno, "Fuzzy Systems and Pattern Recognition," Workshop on Discrete Systems and Fuzzy Reasoning, Queen Mary College, University of London, 1976.

[85] H. Bremermann, "Pattern Recognition by Deformable Prototypes," in Structural Stability, the Theory of Catastrophes and Applications in the Sciences, Springer Notes in Mathematics *25* (1976), 15-57.

[86] L.A. Zadeh, "A Fuzzy-Algorithmic Approach to the Definition of Complex or Imprecise Concepts," Int. Jour. Man-Machine Studies *8* (1976), 249-291.