A PATTERN RECOGNITION APPROACH TO THE PROBLEM OF LINGUISTIC APPROXIMATION IN SYSTEM ANALYSIS

by

P. P. Bonissone

Memorandum No. UCB/ERL M78/57

11 August 1978

A PATTERN RECOGNITION APPROACH TO THE PROBLEM OF LINGUISTIC APPROXIMATION IN SYSTEM ANALYSIS*

Piero P. Bonissone

Computer Science Division
Department of Electrical Engineering and Computer Sciences
University of California, Berkeley

Abstract

The problem of the linguistic approximation is defined on the basis of semantic equivalence. The process consists of interpreting the meaning of any given membership distribution and attaching to it a linguistic label. Some problems of existing implementations of the linguisitc approximation are pointed out. A new approach to the problem, based on feature selection and pattern recognition, is introduced. The sentences used as linguistic labels belong to a language generated by a context free grammar. A membership distribution is associated to each corresponding lable. A prescreening process is performed among the distributions, using the information coded and ordered in the parameter space. The result is a non-fuzzy subset of sentences which are proposed as possible candidates to solve the linguistic approximation problem. Finally, different metrics are applied to these preselected labels and the best candidate is chosen to be the linguistic approximation to the given distribution. Possible modifications to this method are suggested. Some remarks are made about the flexibility and efficiency of this approach. In the conclusions some notes on the necessary trade-off to be considered in the implementation are given. An illustrative example is shown in the appendix.

*This research was supported by the Naval Electronic Systems Command Contract N00039-78-C-0013.

1. Index

2. Introduction

The theory of Fuzzy Sets, introduced in 1965 [1], has proven to be an appropriate tool in dealing with the inherent imprecision of the concepts involved in human reasoning and natural language [2,3,4,5,9,20]. The classical two-valued logic is just too rigid to efficiently represent this kind of information.

It is believed that the human ability to make rational decisions in complex or fuzzy environments, based only on imprecise information, is an approximate rather than a precise process [6].

The information, which constitutes the premises for the deductive process, generally has a qualitative nature, its form being, for instance, an expression in natural language. By using fuzzy sets we are able to transform it to a quantitative representation: the same concept now is represented by a linguistic variable with a value which is a fuzzy subset of the universe of discourse whose label is a sentence in a natural or synthetic language.

Therefore, a language can be seen [7] as a fuzzy relation L from a set of terms S to a universe of discourse U, which assigns to each pair (s,u) element of $S \times U$ a grade of membership

$$\mu_L(s,u) .$$

If we fix s, the membership function $\mu_L(s,u)$ determines a fuzzy subset $A(s)$ of U whose membership function is:

$$\mu_{A(s)}(u) = \mu_L(s,u) , \quad u \in U, \quad s \in S .$$

The fuzzy subset $A(s)$ of U is the meaning of s. The term s is the label of $A(s)$.

Among the terminals of the grammar which generates the language we find primary terms, hedges, relations, conjunctions and disjunctions. While the first are labels of primary fuzzy sets, the rest can be seen as labels of different kinds of operators* which act on the primary fuzzy sets, modifying their original membership function values (membership distributions).

A problem arises when, as a result of a fuzzy reasoning process, we obtain a fuzzy subset of U whose membership distribution does not have a corresponding label.

Then the process of the linguistic approximation to this fuzzy subset consists in finding a linguistic value (label), element of S, whose meaning is the same or the closest (according to some metrics) to the meaning of the unlabelled fuzzy subset whose distribution we just obtained.

3. Definition of Linguistic Approximation

An informal definition of the linguistic approximation was already given in the previous section. The idea of "...finding a label...whose meaning is the same or the closest to the meaning of the unlabelled fuzzy subset..." is nothing else but looking for the term s element of S which has the highest semantic similarity to the unlabelled fuzzy subset.

The meaning of s, which we defined previously to be $A(s) \subset U$, is completely characterized by its membership distribution**

$$\mu_{A(s)}(u) .$$

*Refer to Section 12.2 for more details or to [8,9,20] for a more complete treatment.

**In order to simplify the notation, from now on we will just write $\mu_s(u)$ instead of $\mu_{A(s)}(u)$.

Therefore the semantic comparison between two fuzzy subsets of the same universe of discourse can be done by comparing their membership distributions.

According to the criterion used in performing this comparison we may have different criteria of semantic similarity: we can already see that the linguistic approximation to a fuzzy subset is not unique. A discussion on the different metrics (distances) used to evaluate this similarity is given in Section 3.2 and more extensively in Section 9.

Because of the lack of a precisely defined criterion of similarity, the approximation has to be found by "ad hoc" procedures. A remark is given in [2]: Zadeh notes that "...the standard of precision in computations involving linguistic values are, in general, rather low. This however is entirely consistent with the imprecise nature of fuzzy logic and its role in approximate reasoning."

We will try to formalize, within the limits dictated by implementation restraints, the informal definition previously given by saying:

$$LA[A] = s_A ,$$

if $$d(\mu_A(u),\mu_{A'}(u)) = \min_{\forall B \in L1} d(\mu_A(u),\mu_B(u))$$

where: A and A' are fuzzy subsets of the same universe U;

A is the unlabelled fuzzy subset;

$L1$ is a finite term set language (the linguistic labels);

$s_{A'}$ is the linguistic label (sentence) associated to A', which belongs to $L1$;

$d(A,B)$ is any distance which satisfies the axioms of a metric.*

*These axioms are [10]: $d(A,B) > 0$ if $A \neq B$; $d(A,B) = 0$ if $A = B$; $d(A,B) = d(B,A)$, $d(A,C) \leq d(A,B) + d(B,C)$.

3.1 Existing Implementations

The previous definition of linguistic approximation is based on a best-fit method and two existing implementations follow this method:

Wenstop [11] implemented it using

$$d(A,B) = d1(A,B) = \max_i \left(\mu_A(u_i)-\mu_B(u_i)\right)^2 \quad \text{for all } u_i \text{ elements of } U .$$

Kacprzyk [7,12] used the following distances (for the case in which the universe is partitioned into D points):

$$d(A,B) = d2(A,B) = \sum_{i=1}^{D} \left(\mu_A(u_i)-\mu_B(u_i)\right)^2 .$$

Kaufmann [13] proposes two normalized distances as a measurement between two fuzzy sets (he introduces them as measurements between two fuzzy languages):

$$d(A,B) = d3(A,B) = \frac{1}{D} \sum_{i=1}^{D} |\mu_A(u_i)-\mu_B(u_i)|$$

and

$$d(A,B) = d4(A,B) = \{\frac{1}{D} \sum_{i=1}^{D} \left(\mu_A(u_i)-\mu_B(u_i)\right)^2\}^{0.5} .$$

Procyk [14] instead of using the best-fit method, prefers to express the linguistic approximation to a fuzzy subset of U as a linear combination of rule output sets (the terms in the last column of a decision table; refer to Section 12.2).

Since the unlabelled fuzzy set is obtained by:

$$A = \max_i \left(t_i \min B_i\right)$$

where t_i's are scalars in [0,1] which are then interpreted as truth values and B_i is the i^{th} element of the last column of the decision table.

Then the linguistic approximation is given in terms of the linguistic values of each B_i with non-zero t_i, preceded by the confidence adverb associated to the corresponding t_i.

3.2 Comments on These Implementations

In order to evaluate the different distances let us partition the universe of discourse U into D discrete points, i.e. $|U| = D$ and let us use a finite language L1 which contains M sentences, i.e. $|L1| = M$. G1 is the grammar which generates L1 (for more details refer to Section 5) and contains the following terminals:

LOW, MIDDLE, HIGH, ALL, NONE (primary fuzzy sets)

VERY, MOREORLESS, SORT OF (hedges)

OR, AND (connectors)

NOT (negation)

(,) (markers)

The membership distribution associated with the sentences of L1 are obtained by modifying the membership distributions corresponding to the primary fuzzy sets with a combination of operators (hedges, connectors, negation). The combination is determined by the production rules of grammar G1 (refer to Section 5) and the function of each operator is described in Section 12.2, while its implementation is shown in Section 12.3. A simple example is given, for the case of $D = 21$.

Distance d1 (introduced by Wenstop) has the advantage that it is easy to program and, in the comparison of two distributions, it may not be necessary to perform D differences on the elements of U, since if any of the already computed differences are equal to 1 (= max value of the distance), then the value of the final distance will be equal to 1.

LOW

1 0.98 0.92 0.82 0.68 0.5 0.32 0.18 0.08 0.02 0 0 0 0 0 0 0 0 0 0 0

MIDDLE

0 0 0 0 0 0.06 0.22 0.5 0.78 0.94 1 0.94 0.78 0.5 0.22 0.06 0 0 0 0 0

HIGH

0 0 0 0 0 0 0 0 0 0 0 0.02 0.08 0.18 0.32 0.5 0.68 0.82 0.92 0.98 1

ALL

1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

NONE

0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

VERY LOW

1 0.96 0.85 0.67 0.46 0.25 0.1 0.03 0.01 0 0 0 0 0 0 0 0 0 0 0 0

MOREORLESS LOW

1 0.99 0.96 0.91 0.82 0.71 0.57 0.42 0.28 0.14 0 0 0 0 0 0 0 0 0 0 0

SORTOF LOW

0 0.1 0.35 0.7 1 0.9 0.72 0.53 0.35 0.18 0 0 0 0 0 0 0 0 0 0 0

NOT LOW

0 0.02 0.08 0.18 0.32 0.5 0.68 0.82 0.92 0.98 1 1 1 1 1 1 1 1 1 1 1

LOW OR MIDDLE

1 0.98 0.92 0.82 0.68 0.5 0.32 0.5 0.78 0.94 1 0.94 0.78 0.5 0.22 0.06 0 0 0 0 0

LOW AND MIDDLE

0 0 0 0 0 0.06 0.22 0.18 0.08 0.02 0 0 0 0 0 0 0 0 0 0 0

However, it has two serious disadvantages:

1) It penalizes too much the difference in just one point. This is illustrated in Fig. 1; if we have an unlabelled fuzzy set A and we have to select between B and C, C would be chosen (instead of B, which intuitively should be the result).

2) If we are comparing all the M distributions with the unlabelled one, distance d1 would come out equal to 1 (which is its maximum value) for most distributions. We may refer to the example given in this section, where

$$d1(\text{LOW, MOREORLESS LOW}) = 0.06$$

but

$$d1(\text{LOW, SORTOF LOW}) = 1 \ ,$$

as well as

$$d1(\text{LOW, MIDDLE}) = 1 \ ,$$

$$d1(\text{LOW, HIGH}) = 1 \ , \quad \text{etc.}$$

This means that most distributions are considered too distinct and are rejected. Because of this kind of strict tolerance, we need a large value of M (equivalent to say a large language) if we want to assure a decent fit of the obtained linguistic approximation. However, this implies a considerable increase in the number of comparisons since we need to perform an exhaustive comparison among all the M distributions.

Distances d2 and d4 are conceptually equivalent. So we will analyze only one of them, namely d4. d4 has the advantage of being less strict than d1, since it avoids the problem of overpenalizing the difference in just one point. (In the example of Fig. 1 the result is, in fact, that d2(A,C) > d2(A,B)). It has three disadvantages:

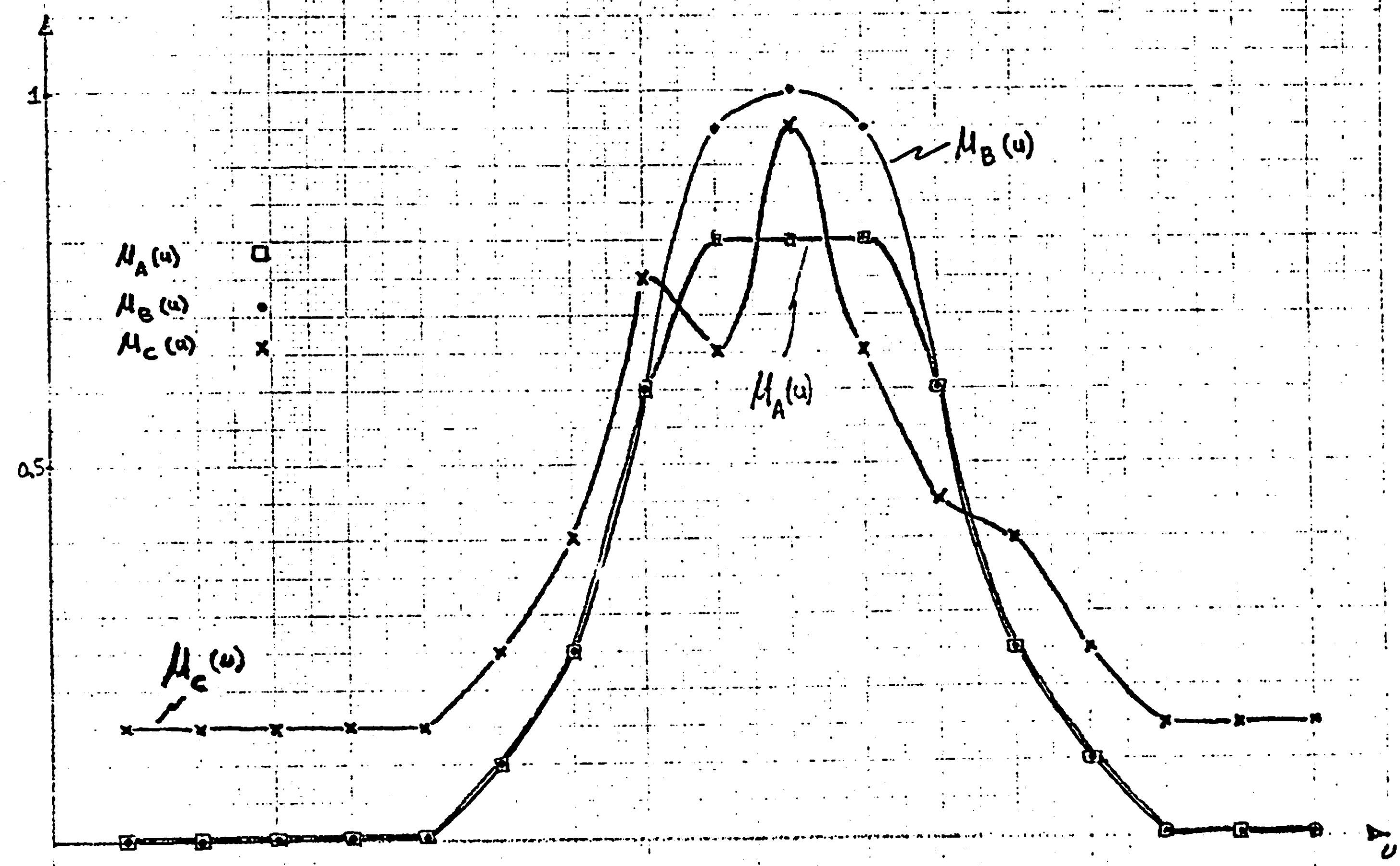


Figure 1

$$d_1(\mu_A(u), \mu_C(u)) < d_1(\mu_A(u), \mu_B(u))$$

$$d_4(\mu_A(u), \mu_C(u)) > d_4(\mu_A(u), \mu_B(u))$$

```
      (LOW SQUERR MOREORLESS LOW)[2]
0.0625

      (LOW SQUERR SORTOF LOW)[2]
1

      (LOW SQUERR MIDDLE)[2]
1

      (LOW SQUERR HIGH)[2]
1

      (LOW SQUERR VERY HIGH)[2]
1
```

Evaluation of distance d1

```
      (LOW SQUERR MIDDLE)[1]
0.6316267735

      (LOW SQUERR HIGH)[1]
0.642406191

      (LOW SQUERR VERY HIGH)[1]
0.6059742254

      (LOW SQUERR MOREORLESS MIDDLE)[1]
0.667368578

      ((VERY LOW) SQUERR VERY HIGH)[1]
0.5672070253

      (ALL SQUERR NONE)[1]
1
```

Evaluation of distance d2

1) It requires an exhaustive comparison over all the D elements of each of the M distributions.

2) It does not uniformily take values over the interval [0,1] (unless we compute d4(ALL,NONE)) and, consequently, after it reaches a certain value it becomes very insensitive to any longer distances between two distributions. We may refer to the example included in this section.

3) It is not very congruent in the sense that the value of the distance between two distributions may be larger than the value of the distance between two less similar distributions. We may refer to the same example, where:

$$\text{d4(LOW, MOREORLESS MIDDLE)} > \text{d4(LOW, HIGH)}$$

and

$$\text{d4(LOW, HIGH)} > \text{d4(VERY LOW, VERY HIGH)} .$$

Distance d3 has exactly the same advantage and disadvantages that d2 and d4 have.

Procyk's method for finding the linguistic approximation has the advantage of an easy computation but it is severely restricted by a very reduced language (the only sentences allowed are the ones which appear in the output rule set, combined with few confidence adverbs). Another limitation is the fact that the method applies only to the result of a combination of an entry of the table with the relation which represents the decision table itself. The implementation is then based on a particular interpretation of the implication, being in this case the cartesian product. However we could have different ways to interpret the implication [15], in which case the problem of the linguistic approximation should be redefined (for this approach). It must also be noted that the unlabelled fuzzy set may not have been generated by a composition with a relation, but by the application of some new relation or new hedge on the primary fuzzy set. In this case, as well as in other cases (e.g., when we want to find the linguistic value to be assigned to the truth value of a fuzzy proposition of which we are doing the truth-qualification [16]), this approach can not be applied.

In general we can say that even if it were good for very simple cases, it lacks a metric which would give an estimate of the 'goodness' of the approximation.

3.3 Requirements for a Good Linguistic Approximation

The requirements of linguistic approximation confront us with at least four problems, not all of which are adequately dealt with by the above implementations.

1) There is a need for the use of a large language, based on a rich vocabulary, in order to have a good approximation of any fuzzy subset of U.

2) This is equivalent to having a large number M of sentences in the available language. Then we need a grammar (preferably a context free grammar) to generate such a language.

3) We must be able to cope with the increase of the number of operations on a data set consisting of $M \times D$ elements, and try, it it is possible, to avoid an exhaustive search among the M sentences.

4) We need an adequate metric which has to be more flexible than d1 and more congruent than d2, d3, d4.

We have to make an important note, which will be considered in the proposal of the grammar (refer to Section 5): the expansion of the vocabulary does not mean that we need more levels of recursion in the productions of the grammar. In fact this would lead to having more nested hedges and it could create sentences which would be completely incomprehensible. Increasing the size of the language has to be done in a horizontal rather than vertical sense (looking at the production tree). This means that without changing the levels of recursion we can increase the number of terminals (primary fuzzy sets, hedges, connectors, relations, etc.). In other words we enrich the vocabulary. The restriction on the number of recursions, in order to have an intuitively understandable answer, justifies the use of a grammar whose productions don't have cyclic nonterminals (i.e. it produces a finite language [24]).

4. Pattern Space Approach

We want to solve the above mentioned problems by trying to reduce the dimensionality of the data and consequently the complexity of the search. In order to do this we introduce the concept of a pattern space which, roughly speaking, has to satisfy four criteria:

a) low dimensionality

b) retention of sufficient information

c) enhancement of distance in pattern space as a measure of the similarity of physical parameters

d) comparability of features.

Let us assume for the sake of simplicity (and since it is a requirement of the implementation) that the universe of discourse U is finite and discrete (i.e. $|U| = D$).

Let A be a non-fuzzy set of M fuzzy subsets of U

$$A = \{A_1(s), A_2(s), \ldots, A_M(s)\}$$

and let S be the set of labels corresponding to A

$$S = \{s_1, s_2, \ldots, s_M\} .$$

Let us define a function F

$$F: [0,1]^D \to \mathbb{R}^p$$

and let us call P the pattern space, which is $\mathbb{R}^p$, in which the mapped membership distribution is represented by a point.

If we apply F to the membership distribution $\mu_{A_i}(u)$ of a fuzzy subset $A_i \subset U$, we obtain as a result the p-tuple

$$\underline{P}^i = (P_1^i, P_2^i, \ldots, P_p^i)$$

which is formed by the coordinates of the points in space P, i.e.

$$F(\mu_{A_i}(u)) = \underline{P}^i .$$

We have to make an important remark: we have said that $\underline{P}^i$ is the representation of A_i in the p-dimensional space P, as $\mu_{A_i}(u)$ is the representation of A_i in the D-dimensional space U. However, the one we just obtained is not a complete representation.*

In fact, our purpose is simply to find a short-cut in the comparison process: we will try to compare the meaning of the fuzzy sets in the pattern space P, rather than in U, taking advantage of the fact that $p \ll D$ $(|P| \ll |U|)$.

*We would have a complete representation only if the mapping function F is injective (one to one). Since we cannot guarantee it, it may happen that

for some pair

$$A_i \neq A_j \Leftrightarrow (\mu_{A_i}(u) \neq \mu_{A_j}(u)) , \quad u \in U$$

and

$$\underline{P}^i = \underline{P}^j .$$

A very important point will be choosing the components of the function F, namely

$$F(\ ) = \begin{bmatrix} f1(\ ) \\ f2(\ ) \\ \vdots \\ fp(\ ) \end{bmatrix}$$

In the selection of these real valued functions we will try to have the minimum amount of redundant information such that with few patterns we can represent most of the information (we then have to reach a trade-off between the number of parameters, which we want to be small, and the amount of information, which we want to be high (even if it is not complete)).

4.1 New Definition of Linguistic Approximation in This Space

We have to choose a criterion to evaluate the semantic similarity between two fuzzy sets A, A', represented as two points in P.

Since we are in a euclidean space $\mathbb{R}^p$, a reasonable distance between two points of this space is the weighted euclidean distance d5:

$$d5(A,A') = \left(\sum_{i=1}^{p} W_i^2 (P_i^A - P_i^{A'})^2\right)^{.5} .$$

Then we can define the linguistic approximations of the fuzzy set A to be:

$$LA[A] = \{s_{A'}\}$$

such that $s_{A'}$ is the label corresponding to fuzzy set A'; A and A' are fuzzy subsets of the same universe U; $d5(A,A') < E$ where E is a parameter which defines our tolerance in judging the similarity.

In Section 8.2 we will see how to determine the weighting factors w_i's and the tolerance parameter E. We can, however, justify at this point the need for W_i's and E. The W_i can be interpreted as the relevance that a difference in parameter P_i has in the evaluation of the semantic similarity between the two fuzzy sets. The W_i's also have the function of normalizing the difference in distinct parameters such that it will be meaningful to add them together. The reason for introducing a tolerance parameter E is based on the fact that the p-tuple $\underline{P}^A$ is not a complete representation of the fuzzy set A. Then, as it can be noted in the definition of linguistic approximation used for this approach, we cannot expect a unique label s_A, to be its linguistic approximation. We rather prefer to find a small non-fuzzy subset of S such that all its elements are within the tolerance E.

The size (cardinality) of this non-fuzzy subset of S is clearly determined by the value of E.

Given $A \subset U$, this approach does not allow us to claim (on an analytical basis) that we can find and pinpoint the label s_i corresponding to the fuzzy subset of U which minimizes, over the set term S, the distance

$$d\left(\mu_{s_i}(u),\mu_A(u)\right)$$

based on the membership distributions (like d1, d3 or d4). We cannot claim either that such a label is always included in LA[A].

However if we use an 'efficient' representation of A in P, we can say, based on several experiments, that s_i corresponds to a point $\underline{P}^i$ in P which is generally fairly close to the point $\underline{P}^A$ representing A. Therefore a relatively small value of E (refer to Section 8.2) will enable us to include s_i in LA[A] almost always. And in any case the computed labels are always quite reasonable.

The advantage of this approach is that the complexity of the search does not increase with an increment in the dimension of U, which may be useful if we require a better resolution.

It is also interesting to note that we need to perform the same kind of operations on a data set of size $(p \times M)$, which is much smaller than $(D \times M)$. Moreover we can increase the length M of the language which forms S and not have to worry very much about the complexity of the search. We can anticipate that this point is of vital importance when we try to implement it in a computer. (Refer to Section 8.3). At this point, if we want a unique answer, we have to apply some metric between each element of the set of linguistic approximations that we have obtained and the unlabelled fuzzy set A. This however is not a serious problem, since, by using a small tolerance parameter E, $|LA[A]|$ is small.

5. Proposal of a Context Free Grammar

In order to be able to deal with a language with several sentences we need a grammar to generate it. A very convenient one is the so-called context free grammar [17]. We will limit the level of recursions by not creating cyclic nonterminals in order to assure that the generated language will be of finite length.

Let us propose the following grammar G1:

$$G1 = (V_N, V_T, S, P)$$

where $\{V_N\} = \{A,B,V,H,U,T\}$ is the set of non-terminals

$\{V_T\} = \{$NOT, OR, AND, VERY, MOREORLESS, SORTOF, ALL, NONE, LOW, MIDDLE, HIGH$\}$ is the set of terminals

S is the starting symbol

P is the set of productions:

$S \rightarrow A$

$S \rightarrow ACA$

$S \rightarrow U$

$A \rightarrow B$

$A \rightarrow NB$

$B \rightarrow T$

$B \rightarrow HT$

$B \rightarrow VVT$

$N \rightarrow NOT$

$C \rightarrow AND$

$C \rightarrow OR$

$H \rightarrow VERY$

$H \rightarrow MOREORLESS$

$H \rightarrow SORTOF$

$V \rightarrow VERY$

$T \rightarrow LOW$

$T \rightarrow MIDDLE$

$T \rightarrow HIGH$

$U \rightarrow ALL$

$U \rightarrow NONE$

5.1 Finite Language Generated by the Grammar

In order to compute the number of sentences of the language let us call:

c = number of connectors ($|C|$)

u = number of universal terms ($|U|$)

t = number of primary fuzzy sets ($|T|$)

h = number of hedges ($|H|$)

v = number of hedges which can be nested once

Then if we consider that we will not apply a connector to two identical expressions, i.e., $(A_i C A_i)$, since this is semantically equal to A_i, we end up with a total number of sentences:*

*Refer to Section 12.4 for its derivation.

$$|L(G1)| = u + t(h+1+v^2)[2 + 2ct(h+1+v^2) - c] .$$

If we evaluate this expression for the particular values of 'u, t, h, c, v' that we have in our proposed G1 (namely $u = 2$, $t = 3$, $h = 3$, $c = 2$, $v = 1$) we find that

$$|L(G1)| = 902 .$$

The previous expression, however, will give us the number of sentences in the language that we obtain if we modify the vocabulary (corresponding to change the values of 'u, t, h, c, v').

Some sample sentences of this language are given on the next page, while the implementation of the grammar is shown in Section 12.3.

5.2 Association of a Membership Distribution to Each Sentence

The membership distribution of each label (sentence of L(G1)) is computed by generating first the membership distribution of the primary fuzzy sets and then by applying to them the operators corresponding to the labels of connectors and hedges.

We have implemented our ideas in APL. In the implementation, each sentence generated by G1 is a combination of compatible functions which operate on the vectors containing the values of the membership distributions of the primary fuzzy sets. The latter are generated by a starting function in which it is possible to specify the number of partitions, allowing the user to arbitrarily change this number. This is a very flexible feature of the system. For more details refer to Section 12.3.

On the next page we show the membership distributions associated with the labels used to show a sample of L(G1).

```
      S[1;]
 (LOW                          )AND MIDDLE

      S[10;]
 (LOW                          )AND MOREORLESS MIDDLE

      S[210;]
 (VERY        HIGH             )AND SORTOF     HIGH

      S[645;]
 (VERY        HIGH             )OR  SORTOF     HIGH

      S[500;]
 (HIGH                         )OR  MOREORLESS MIDDLE

      S[700;]
 (MOREORLESS HIGH              )OR  SORTOF     LOW


      S[835;]
(NOT VERY         LOW     )OR  NOT VERY        MIDDLE


      S[570;]
(VERY VERY MIDDLE       )OR  NOT SORTOF     HIGH


      A[1;]
LOW

      A[12;]
MOREORLESS HIGH

      A[29;]
NOT SORTOF      MIDDLE

      U[1;]
ALL

      U[2;]
NONE
```

LOW AND MIDDLE

0 0 0 0 0.06 0.22 0.18 0.08 0.02 0 0 0 0 0 0 0 0 0 0 0

LOW AND MOREORLESS MIDDLE

0 0 0 0 0.24 0.32 0.18 0.08 0.02 0 0 0 0 0 0 0 0 0 0 0

(VERY HIGH) AND SORTOF HIGH

0 0 0 0 0 0 0 0 0 0 0 0.01 0.03 0.1 0.25 0.46 0.67 0.35 0.1 0

(VERY HIGH) OR SORTOF HIGH

0 0 0 0 0 0 0 0 0 0 0.18 0.35 0.53 0.72 0.9 1 0.7 0.85 0.96 1

HIGH OR MOREORLESS MIDDLE

0 0 0 0 0.24 0.47 0.71 0.88 0.97 1 0.97 0.88 0.71 0.47 0.5 0.68 0.82 0.92 0.98 1

(MOREORLESS HIGH) OR SORTOF LOW

0.1 0.35 0.7 1 0.9 0.72 0.53 0.35 0.18 0 0.14 0.28 0.42 0.57 0.71 0.82 0.91 0.96 0.99
1

(NOT VERY LOW) OR NOT VERY MIDDLE

1 1 1 1 1 0.95 0.97 0.99 1 1 1 1 1 1 1 1 1 1 1 1

(VERY VERY MIDDLE) OR SORTOF HIGH

0 0 0 0 0 0 0.06 0.37 0.77 1 0.77 0.37 0.53 0.72 0.9 1 0.7 0.35 0.1 0

LOW

0.98 0.92 0.82 0.68 0.5 0.32 0.18 0.08 0.02 0 0 0 0 0 0 0 0 0 0 0

MOREORLESS HIGH

0 0 0 0 0 0 0 0 0 0 0.14 0.28 0.42 0.57 0.71 0.82 0.91 0.96 0.99 1

NOT SORTOF MIDDLE

1 1 1 1 0.66 0.34 0 0.11 0.68 1 0.68 0.11 0 0.34 0.66 1 1 1 1 1

ALL

1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

6. Feature Selection

This is a very crucial point, since the right selection of features determines the success or failure of any pattern recognition process [18]. After several tries and experiments looking for an 'efficient'* representation, we ended up with the following four parameters:

1) Power
2) Entropy
3) First moment
4) Skewness

6.1 Power

The power of a fuzzy set is defined [19] as the summation of the membership value of each element of the support** of the fuzzy set, i.e.

$$\mathrm{Power}(A) = \sum_{i=1}^{D} \mu_A(u_i) \ .$$

This definition is for the assumed case of having a finite discrete universe of discourse U, such that $|U| = D$.

This concept may also be interpreted as a numerical summary of the fuzzy cardinality of a fuzzy set [6].

In the implemented APL function this value is normalized such that it will take values in $[0,1]$.

---

*By efficient we mean having a small set of parameters containing a lot of information on the distribution.

**The support of a fuzzy set $A \subset U$ is a non-fuzzy subset of U such that all its elements have a non-zero membership value in A. Then the summation over the support is identical to the summation over the entire universe.

6.2 Entropy

The fuzzy entropy of a fuzzy set $A \subset U$ is defined as [19]:

$$\text{Entropy}(A) = \sum_{i=1}^{D} S(\mu_A(u_i))$$

(for $\mu_A(u_i) \neq 0$ and $\mu_A(u_i) \neq 1$) where $S(\ )$ is the Shannon function:

$$S(x) = -x \ln x - (1-x) \ln (1-x) \ .$$

The entropy is a measure of the degree of fuzziness [19] and its existence (non-divergence) has been proven for the case of a finite support [19], as well as for the continuous support [21,22].

In the APL function this value has been normalized with respect to the maximum value that the entropy can take over all the subsets of U. This happens to be the entropy value of a fuzzy subset whose membership value is equal to 0.5 for each element of U.

6.3 First Moment

This parameter indicates the 'center of gravity' of the membership distribution, just as the well-known mean indicates the center of a probability distribution.

The parameter is simply obtained by:

$$\text{First moment}(A) = En_A(u)$$

and we define $En_A(\ )$ (which stands for ensamble average of A) as:

$$En_A(f(u)) = \sum_{i=1}^{D} f(u_i) \frac{\mu_A(u_i)}{\text{power}(A)} \ .$$

In the implementation, for the sake of simplicity, we scale the points

$u_i \in U$ such that they take values only on the integers, starging from 0, i.e.

$$u = (i-1) \ , \quad i = 1,\ldots,D \ .$$

Then $En(u)$ becomes

$$En_A(u) = En_A(i-1) = \sum_{i=1}^{D} (i-1) \frac{\mu_A(u_i)}{power(A)} \ .$$

The shift $(i-1)$ occurs simply because we are starting from 0 in the assignment of the values of u_i. The membership distribution is normalized such that its area is equal to one. This is useful for the computation of this parameter (first moment) and the parameter discussed in the next section (skewness). However, the distribution will not be normalized during the inference process (which is part of the approximate reasoning) nor during the linguistic approximation process.

6.4 Skewness

This parameter is interpreted as a measure of asymmetry of the distribution with respect to its center of gravity. It is simply the third moment. It is defined as:

$$\begin{aligned} \text{Skewness}(A) &= \sum_{i=1}^{D} [(i-1)-(\text{first moment}(A))]^3 [\frac{\mu_A(u_i)}{power(A)}] \\ &= En_A[(i-1)^3] \ . \end{aligned}$$

A distribution skewed to the left can be shown to have a negative third moment, while the one skewed to the right will be a positive one.

The way this parameter is obtained, in the implementation, is:

$$\text{Skewness}(A) = En_A[(i-1)^3] - 3 \times \text{First moment}(A) \times \text{Var}(A) - (\text{First moment}(A)^3)$$

where $$\mathrm{Var}(A) = \mathrm{En}_A[(i-1)^2] - (\mathrm{En}_A[i-1])^2 \, .$$

7. Analysis of the Correlation Among Features

Since it is very important to have a very small number of parameters, in order to have a small-dimensional space P, and in this way reduce considerably the complexity of the search for similarity, we try to avoid having two or more parameters with a high cross-correlation (which would imply redundancy of information).

An experimental analysis was performed in the following way: a set of seven different parameters was chosen and a non-fuzzy subset of distributions, which was considered a good representative sample of the set of distributions in U, was mapped in this 7-dimensional pattern space. Then projections of the points (representing the distributions) were taken over all the possible 21 planes. Each projection was then analyzed, looking for a possible function of the two coordinates of the corresponding plane.

A high correlation was found between the parameters:

Power and Bandwidth*

Power and Spread**

The maximum of the distribution (analogous to the mode), which together with the parameters discussed in Section 6 completed the group (certainly not exhaustive) of parameters used in this correlation analysis, did not contribute a relevant amount of information. Hence only the previously discussed parameters were used in the representation of the set of distributions on U discussed in Section 6.

---

*Bandwidth is defined as the size (cardinality) of the subset of U such that the membership value of each point of this subset is bigger than 0.5. Refer to Section 12.2 for more details.

**Spread is a measure of the dispersion of the distribution. It is calculated (analogously to the variance in a probability distribution) by computing the second moment with respect to the center of gravity.

A sample of the more interesting projections of the points of this 7-dimensional space on some of its planes is shown in the appendix (Section 12.5). We can verify the strong correlation between power, bandwidth and spread, as mentioned before, by observing the almost straight line that the projections form on the corresponding planes. In an analogous way we can see how other parameters present a very low cross-correlation: the projection on the plane that they form is a very well-spread cluster of points.

We would like to remark that the parameters we have chosen do not form the only possible set of patterns useful for the purpose of evaluating similarity among the distributions. It is also clear that the addition of some other parameter could have 'sharpened' the discrimination since we would have a more complete representation. However, we consider the group of parameters described in Section 6 to be a good answer to the trade-off between complexity and completeness of representation.

8. Prescreening Process

We will extract the information from the set of distributions corresponding to the set of labels (sentences of L(G1)), thus obtaining, for each fuzzy set, a 4-tuple which characterizes it in P.

Then we can form a tableau with five columns: the first four for the parameters and the fifth for an index which indicates the corresponding label in the language. We could look at this language as if it were an alphanumerical array of finite dimensions such that each row of the array would correspond to a sentence. Then the index (number of the row) identifies the sentence.

Since we are going to search in this tableau each time we want to find the linguistic approximation to an unlabelled fuzzy subset of U, it is very convenient to order the tableau according to some kind of structure which simplifies the search.

8.1 Parameters Data Structure

Since a tree is a more efficient representation than a tableau, we will order the latter in such a way that it will be very easy to transform it into a tree. Perhaps this is best illustrated with an example as shown in Section 12.5. It is very easy to come up with an algorithm which does this in a recursive fashion. An example of such an algorithm is given in Section 12.3.

In the tree corresponding to the tableau (also shown in the example), a node at each level is equivalent to a group of elements with the same parameter value in the corresponding column.

The searching problem can be speeded up by the use of this ordering.

A possible searching algorithm could be the following: At the first level we compute, for each different node, the square of the first component of the distance d5 from the leaf to the point which represents the unlabelled fuzzy set. Then a first comparison against the tolerance parameter E^2 is made and if for some node we are already exceeding this value (i.e. $(W_1^2(p_1^A-p_1^s)^2) > E^2)$), then we drop all the successors of that node. We pass to the second level and compute, for the surviving nodes, the square of the second component of d5, add it to the previously computed square of the first component and compare again with E^2. We repeat the process until we arrive at the leaves of the last 'survivor' nodes. Of course some other algorithm could be found to perform this search. In particular, if we slightly modify the criterion of similarity in this space, a very fast algorithm could be implemented.

In fact, a geometrical interpretation of such criterion is the following: We want all the points s we select to satisfy

$$d5(A,s) < E \ ;$$

then if we take the square of the previous inequality

$$\sum_{i=1}^{4} W_i^2 (P_i^A - P_i^s)^2 < E^2$$

this is equivalent to having a hypersphere of radius E in a space of dimensions

$$[W_1 e_1, W_2 e_2, W_3 e_3, W_4 e_4]$$

where

$$e_i = (P_i^A - P_i^s) \ .$$

Then the criterion is reduced to check if the point is contained in the hypersphere.

If we change the shape of the region of tolerance from a hypersphere to a hypercube, whose center is in the origin and whose edge length is E, we have the criterion of similarity changed into:

$$(4 \times (W_i^2 e_i^2)) < E^2$$

for $i = 1,\ldots,4$ which in fact can be rewritten as:

$$|W_i e_i| < (E/2)$$

for $i = 1,\ldots,4$.

This metric is called 'city-block distance'.

This of course defines a tougher criterion of similarity since the hypercurbe is inscribed in the hypersphere. If we assume the points to have a uniform density in this space $[w_1e_1,\dots,w_4e_4]$, then the reduction in volume would be proportional to the reduction in the number of points which satisfy the criterion.

If we want to maintain the same volume we have to come up with an edge length E', between E and $2E$, such that a hypercube with this edge would have the same volume as the hypersphere of radius E. This is a simple problem of advanced calculus.*

What interests us, however, is the fact that now we can check for the criterion in this new way:

$$(W_iP_i^A - \frac{E'}{2}) < W_iP_i^S < (W_iP_i^A + \frac{E'}{2})$$

for $i = 1,\dots,4$.

It is quite clear that the check for this new criterion in the ordered tableau (or in the tree) may be performed in a very fast and efficient way.

8.2 Determination of Weights and Tolerance

The weights W can be expressed as:

$$W_i = \frac{I_i}{R_i}$$

where I_i is a factor which measures the relative importance of the parameter P_i with respect to the other ones. R_i is the length of the range of values that parameter P_i takes over all the points that constitute

*It can be proven (refer to Appendix 12.4) that

$$E' = (\frac{\pi}{1.4142})^{.5}E .$$

our data, i.e.

$$R_i = \max_j \{P_i^j\} - \min_j \{P_i^j\}$$

for $j = 1,\ldots,M$, $i = 1,\ldots,p$. We want to remind the reader that P_i^j is the i^{th} parameter of the fuzzy set whose label is s_j. We could impose the restriction that the I's have to be normalized such that:

$$\prod_{i=1}^{4} I_i = 1 \ .$$

Clearly if all the parameters are equally important, then $I_i = 1$, for all i. This, however, gives the user the flexibility of tailoring his own definition of similarity, by assigning different values to the importance of each parameter. By defining the weights in this way, we obtain (for the case of $I_i = 1$ for all i) that the value of the weighted difference on parameter P_i, i.e.

$$W_i^2(P_i^A-P_i^s)^2 = W_i^2 e_i^2$$

is in the interval $[0,1]$, provided that the value of the coordinates of the point representing A in P falls within the range R. This should be achieved if we have enough data (points in P, representing sentences in L(G1)).

Because of this normalization, the different dispersion that each pattern had is now compensated. Then the summation over distinct dimensions of P (different parameters) is now a meaningful operation.

In pattern recognition literature [10,23] other possible ways of finding the weights have been suggested, such that the maximum distance between members of the same set is minimized. Such proposed weights are:

$$W_i = \frac{K}{\mathrm{Sigma}_i^2}$$

where $Sigma^2$ is the sample variance of the set of values of parameter P. K is a constant which is chosen in order to satisfy a normalizing restriction on W_i, like:

$$K = \frac{1}{\sum_{i=1}^{p} \left(\frac{1}{Sigma_i^2}\right)} \quad \text{if} \quad \sum_{i=1}^{p} W_i = 1$$

or

$$K = \prod_{i=1}^{p} (Sigma_i^2)^{1/p} \quad \text{if} \quad \prod_{i=1}^{p} W_i = 1 \ .$$

We can see that these expressions (and the idea involved in them) are very similar to the one we are using.

An important remark has to be made: the weights are calculated after the parameters of the set of membership distributions corresponding to the labels in S have been computed. Thus they, as well as the tree parameters, are independent from the values of the coordinates of the point corresponding to A. Because of this, they are calculated only once.

We have already seen, in Section 8.1, the geometrical interpretation of the tolerance parameter E. If we know that $W_i e_i$ is in [0,1] for all i, then we can see that to fix a certain E implies that an average percentage of points would be inside the hypersphere.

E = 1 → 50%
E = .5 → 25%
E = .4 → 20%
E = .3 → 15%
etc.

If we use the 'city block distance', corresponding to the hypercube instead of the hypersphere, as our tolerance region, then E (or E') is interpreted as the maximum (symmetric) range of error for each parameter.

Then we have:

$$E = 1 \rightarrow \max_i |W_i e_i| < .5$$

$$E = .5 \rightarrow \max_i |W_i e_i| < .25$$

etc.

8.3 Notes on the Implementation

Since the list of programs is given in Section 12.3 and the results are illustrated with an example in Section 12.1, we will limit ourselves in this section to some general but useful remarks on the implementation.

We only have to store the $p \times M$ real values corresponding to the parameter representation of A.

Since we need to keep track of the elements of the original set and, after the prescreening, we will want to know the distribution of some of them, we keep an alphanumerical representation of each set. In other words we store the labels (and its corresponding index in the language), instead of storing the distributions.*

When we have selected the elements of the non-fuzzy subset of linguistic approximations LA[A] (from the prescreening process), we use an APL feature called 'execute' which we apply to the alphanumerical label transforming it into an executable function which gives as a result the corresponding membership distribution.

Another important remark is the following: it may be that the context free grammar G1, which generates the language, will produce a set of sentences which have exactly the same meaning (exact semantical equivalence

*It is important to note that while an alphanumerical character needs only a 1 byte representation, a real number requires 8 bytes to represent it.

implies that the associated membership distributions are identical), even if they are syntactically different. An example is given on the next page.

In this case we want to perform a 'compression' on the sentences of the language and suppress the sentences which are semantically identical, representing them with the one belonging to the same class of equivalence which has the shortest and therefore more intuitively understandable label.

Another observation could be made on the need of writing the programs which deal with the parameters in such a way that they take as input the columns of the tableau rather than the whole table. This is done in order to avoid the possibility that, for a large M, we could have a table too big to be fully represented in the available resident memory. In fact, some time-sharing systems in which APL is generally available, are blessed with severe memory restrictions. An example of this is UCLA CCN IBM 360-91, where the size of any active workspace is limited to 48K.

9. Label Selection

We have already obtained a small non-fuzzy subset of S, LA[A], as a result of the prescreening process. Now we want to compare the similarity between the unlabelled fuzzy set A and the membership distribution of each member of LA[A], the set of linguistic approximations, in order to select the 'best' one.

In order to do this we need to define an appropriate measure of the similarity (or dissimilarity) of two fuzzy sets.

9.1 Euclidean and Hamming Distance

In Section 3.2 we have already examined the Euclidean distance (d4) and the Hamming distance (d3). However, we will show the result of applying

```
(LOW                    )AND VERY VERY LOW

      A[4;]
VERY VERY LOW


      S[473;]
(MIDDLE                 )OR  MOREORLESS MIDDLE

      A[11;]
MOREORLESS MIDDLE


      ⍝  THESE ARE SENTENCES OF THE SAME LANGU GE
      ⍝  L1  GENERATED BY GRAMMAR G1.
      ⍝  EVIDENTLY THEY ARE SYNTACTICALLY DIFFERENT.


      LOW AND VERY VERY LOW
1  0.92  0.72  0.45  0.21  0.06  0.01  0  0  0  0  0  0  0  0  0  0  0  0  0

      VERY VERY LOW
1  0.92  0.72  0.45  0.21  0.06  0.01  0  0  0  0  0  0  0  0  0  0  0  0  0


      MIDDLE OR MOREORLESS MIDDLE
0  0  0  0  0  0.24  0.47  0.71  0.88  0.97  1  0.97  0.88  0.71  0.47  0.24  0  0  0  0  0

      MOREORLESS MIDDLE
0  0  0  0  0  0.24  0.47  0.71  0.88  0.97  1  0.97  0.88  0.71  0.47  0.24  0  0  0  0  0
```

those distances, as well as d1, to a sample of fuzzy subsets of U, in order to illustrate the incongruences pointed out in Section 3.2.

9.2 Bhattacharyya Distance

Recently, possibilistic interpretations have been given to membership distributions [25,26]. This justifies an attempt in using, after proper extension, some probabilistic measures to compare two possibility (membership) distributions.

A good measure of such a comparison is the Bhattachryya Distance, which is defined [27] as:

$$d6(p1(u),p2(u)) = -\ln R$$

where R is called the Bhattacharyya coefficient and, for the discrete case ($|U| = D$), is defined [27] as:

$$R(p1,p2) = \sum_{i=1}^{D} [(p1(u_i) \times p2(u_i)]^{.5} .$$

Note that in this definition we are using membership distributions which have been normalized, such that:

$$\sum_{i=1}^{D} p_j(u_i) = 1 \quad \text{for all} \quad j .$$

This is simply obtained by:

$$p_j(u_i) = \frac{\mu_{A_j}(u_i)}{\text{power}(A_j)} .$$

This measure, however, does not satisfy the triangle inequality, and therefore is not a metric (according to the axioms of footnote on p. 6).

(Note : The two distance values are respectively d4 and d1)

LOW SQUERR VERY LOW

0.1011599394 0.0625

LOW SQUERR LOW

0 0

LOW SQUERR MOREORLESS LOW

0.108671633 0.0625

LOW SQUERR SORTOF MIDDLE

0.615165483 1

LOW SQUERR MOREORLESS MIDDLE

0.667368678 1

LOW SQUERR MIDDLE

0.6316267735 1

LOW SQUERR VERY MIDDLE

0.5976700794 1

LOW SQUERR MOREORLESS HIGH

0.6844810339 1

LOW SQUERR HIGH

0.642406191 1

LOW SQUERR VERY HIGH

0.6059742254 1

(VERY LOW) SQUERR (VERY HIGH)

0.5672070253 1

(MOREORLESS LOW) SQUERR (VERY HIGH)

0.650410127 1

(MOREORLESS LOW) SQUERR MOREORLESS HIGH

0.7241152237 1

(MOREORLESS LOW) SQUERR SORTOF HIGH

0.6507248705 1

(MOREORLESS LOW) SQUERR MIDDLE

0.6413378873 1

(NOT LOW) SQUERR HIGH

0.6042074699 1

LOW HAMMINGDISTANCE VERY LOW

0.05571428571

LOW HAMMINGDISTANCE LOW

0

LOW HAMMINGDISTANCE SORTOF MIDDLE

0.4780952381

LOW HAMMINGDISTANCE MOREORLESS MIDDLE

0.540952381

LOW HAMMINGDISTANCE MIDDLE

0.4942857143

LOW HAMMINGDISTANCE VERY MIDDLE

0.4485714286

LOW HAMMINGDISTANCE MOREORLESS HIGH

0.5857142857

LOW HAMMINGDISTANCE HIGH

0.5238095238

LOW HAMMINGDISTANCE VERY HIGH

0.4680952381

LOW HAMMINGDISTANCE NOT HIGH

0.4761904762

(VERY LOW) HAMMINGDISTANCE VERY HIGH

0.4123809524

(MOREORLESS LOW) HAMMINGDISTANCE VERY HIGH

0.53

(MOREORLESS LOW) HAMMINGDISTANCE MOREORLESS HIGH

0.6476190476

(MOREORLESS LOW) HAMMINGDISTANCE MIDDLE

0.5028571429

ALL HAMMINGDISTANCE NONE

1

We can observe that R takes values in [0,1], being equal to 1 when both distributions are identical. Then we can define a distance d7 [29]:

$$d7(p1,p2) = [1 - R(p1,p2)]^{.5} .$$

It can be easily proved that this measure does satisfy all the axioms of a metric.

By applying this measure to the same sample of fuzzy subsets of U, used to illustrate the incongruences of d3 and d4, we can see that this new distance reflects very well the semantic distance among fuzzy sets. It does not show any of the incongruencies mentioned in Section 3.2 for d3 and d4 and it is also less strict than d1. This example is shown on the next page.

This distance has been applied in the implementation and has provided very good results.

A geometric interpretation of the Bhattacharyya coefficient R is the following: if we regard the numbers

$$[p_j(u)]^{.5} , \quad j = 1,2 \quad \text{for all allowable} \quad u$$

as the direction-cosine of the two vectors in the space U (we can in fact visualize a fuzzy subset p(u) in a discrete U ($|U| = D$) as a D-dimensional vector), then the coefficient R is the cosine of the angle between these two vectors.

9.3 Other Possible Distances

Some other distances, actually used in probabilistic contexts, could be used to measure the dissimilarity of two possibility distributions.

LOW BHATTADISTANCE VERY LOW

0.1366279857

LOW BHATTADISTANCE LOW

0

LOW BHATTADISTANCE MOREORLESS LOW

0.1142214296

LOW BHATTADISTANCE SORTOF MIDDLE

0.8505931171

LOW BHATTADISTANCE MOREORLESS MIDDLE

0.8761510068

LOW BHATTADISTANCE MIDDLE

0.8967076653

LOW BHATTADISTANCE VERY MIDDLE

0.928484797

LOW BHATTADISTANCE SORTOF HIGH

1

LOW BHATTADISTANCE HIGH

1

(MOREORLESS LOW) BHATTADISTANCE MIDDLE

0.8429398365

(NOT LOW) BHATTADISTANCE HIGH

0.5132778802

(NOT LOW) BHATTADISTANCE LOW

0.8200559701

(VERY LOW) BHATTADISTANCE MOREORLESS LOW

0.2457860041

(MOREORLESS LOW) BHATTADISTANCE MOREORLESS MIDDLE

0.820374781

ALL BHATTADISTANCE NONE

1

(NOT LOW) BHATTADISTANCE MIDDLE

0.4949688234

For example, we could try to use:

$$d8(p1,p2) = En_{p1}[\ln \frac{p1(u)}{p2(u)}] - En_{p2}[\ln \frac{p1(u)}{p2(u)}]$$

where

$$En_{pj}[f(u)] = \sum_{i=1}^{D} f(u)_i \frac{\mu_{pj}(u_i)}{power(pj)}$$

as it was also defined in Section 6.3.

This distance is called Divergence and it is was introduced by Jeffreys [29].

Note that the Divergence is not a metric, since it does not satisfy the triangle inequality.

Several other distance measures could be adapted for this application. They are extensively treated (for the probabilistic context) in [28,30].

10. Remarks, Applications and Conclusions

Remarks

1) The described approach shows a great flexibility, allowing the user to define his own concept of similarity by changing the relevance value I_i which multiplies the corresponding parameter P_i. In fact, it may be the case that for some particular application he wants the membership distribution of the linguistic approximation to have the same area under the curve that the unlabelled distribution has. This can be easily achieved by selecting a proper high value for the I corresponding to the parameter 'power'.

2) Since the result of the prescreening process gives us a non-fuzzy subset of S with a small cardinality (compared with $|S| = |L(G1)|$), we could apply different metrics to evaluate the similarity of each member of the set of linguistic approximations and the unlabelled one. Then, since

each metric would measure a different feature, we could use a majority rule (2 out of 3) to arrive at a deterministic choice of the label.

3) The parameter 'entropy' has shown to be very sharp in the discrimination of points but it is also very sensitive. This may cause the presence of a 'disturbance' (like a plateau with a small membership value (0.2) widely spread over the universe of discourse) which would provoke a big 'jump' in the value of the parameter.

Since all the points are well-distributed (almost uniformly distributed $[0,1]^4$ in the weighted pattern space), a large deviation of point $\underline{P}^A$, representing the unlabelled fuzzy set A even on only one of the dimensions, will take it completely off the region where all the points are located.

Three different ways of handling this problem are suggested. The first one has been implemented successfully.

a) We can calculate the minimum distance of point $\underline{P}^A$ to the set of points and assign to the tolerance parameter E a value of the same order of magnitude.

b) We can reduce the weight corresponding to the parameter entropy, in order to compensate for its sensitivity.

c) We can use a smoothing filter, applying it to the unlabelled distribution before we extract its parameters. This can be done with some kind of deterministic threshold or fuzzy threshold or a combination of both.*

*A deterministic threshold Td(A) is:

$$\mu_{Td(A)}(u_i) = \begin{cases} \mu_A(u_i) & \text{if } \mu_A(u_i) \geq \mu_0 \\ 0 & \text{if } \mu_A(u_i) < \mu_0 \end{cases}$$

while a fuzzy threshold Tf(A) could be:

$$\mu_{Tf(A)}(u_i) = \begin{cases} \mu_A(u_i) & \text{if } \mu_A(u_i) \geq \mu_0 \\ K\mu_A^2(u_i) & \text{if } \mu_A(u_i) < \mu_0 \end{cases}$$

where $(1/\mu_0) > K > 0$. μ_0 represents the level of the threshold.

4) If we could derive a distance from the concept of a dissimilarity relation,* then, if $d(s_i, A)$ is minimum over all the s_i members of LA[A], we could interpret it by saying that A and s_i are the only two elements of the same equivalence class of a partition with a threshold equal to $(1 - d(s_i, A))$.

5) If we want to add or replace some parameters, looking for any possible improvement, we have to keep in mind the four criteria which any 'good' pattern space has to satisfy. (We listed them in Section 4).

Applications

The linguistic approximation is practically the final step of the process of approximate reasoning, in which the result is interpreted semantically and a qualitative answer is given back to the user.

We can refer to three situations in which this step is necessary:

1) We can approximate an infinite language $L0 = L(G0)$ (where G0 contains some cyclic nonterminals) with a finite language $L1 = L(G1)$. In fact, the fuzzy set associated to any sentence of L0 is treated as an unlabelled fuzzy set, and its corresponding 'closest' label is searched in L1. The 'goodness' of this approximation depends, of course, on the proper choice of G1, given G0.

*Dissimilarity relation d is defined [31] as:

$$\mu_d(x,y) = 1 - \mu_S(x,y)$$

where $\mu_S(x,y)$ is the similarity relation between x and y. The similarity relation is a generalization of the concept of equivalence relation and must satisfy:

$$\mu_S(x,x) = 1$$

$$\mu_S(x,y) = \mu_S(y,x)$$

$$\mu_S(x,z) \geq \bigvee_y \mu_S(x,y) \wedge \mu_S(y,z)$$

2) We can find the label for the result of performing a composition of an entry with a relation which represents the decision table. Such is the case of the example given in Section 12.1. This is the classical inference process, present in most applications of fuzzy sets.

3) We can interpret the results of a truth-qualification:

$$(X \text{ is } F) \rightarrow \Pi_X = F$$
$$(X \text{ is } F) \text{ is } L \rightarrow \Pi_X = F^*$$

where $\mu_{F^*}(u) = \mu_L(\mu_F(u))$. Then we are able to find $LA[F^*]$. Note: Π_X represents the induced possibility distribution of X.

Conclusions

The described approach does not pretend to be the panacea for all the problems which may arise when we try to solve the non-trivial task of finding the linguistic approximation to any unlabelled fuzzy subset of some universe U.

Since the tendency is moving toward a more flexible language (hence a more complicated grammar), this method attempts to cope with the increasing complexity of an exhaustive search by offering a short-cut (the reduction of the dimensionality of the data).

It is insensitive (as far as increasing of complexity is concerned) to increasing the number of elements of the universe of discourse, which sometimes may be useful in achieving a better resolution of membership distribution curves.

It also shows a solution to the classical trade-off between space (for storage) and time (for processing) that we have to face in the implementation. By keeping only the alphanumerical representation of the distributions and a reduced number of real-valued data, we save memory space; by using the pattern space representation, we save processing time.

Acknowledgment

This research was supported by the Naval Electronic Systems Command Contract N00039-78-C-0013.

I wish to thank Professor L.A. Zadeh for his guidance and suggestions.

I would also like to thank Mucio Dias, Dennis Allard and Christian Freksa for the stimulating discussions on issues related to those considered in this report.

I want to express my gratitude to Judith Pinson for her valuable contribution in correcting and editing the final version of this report.

11. Bibliography

[1] ZADEH, L.A. (1965). Fuzzy sets. Information and Control 8, 338-353.

[2] ZADEH, L.A. (1975). Fuzzy logic and approximate reasoning. Synthese 30, 407-428.

[3] ZADEH, L.A. (1975). The concept of a linguistic variable and its application to approximate reasoning. Part I. Information Sciences 8, 199-249; Part II. Information Sciences 8, 301-357; Part III. Information Sciences 9, 43-80.

[4] ZADEH, L.A. (1976). A fuzzy-algorithmic approach to the definition of complex or imprecise concepts. Int. J. Man-Machine Studies 8, 249-291.

[5] GAINES, B.R. (1976). Foundations of fuzzy reasoning. Int. J. Man-Machine Studies 6, 623-668.

[6] ZADEH, L.A. (1977). A theory of approximate reasoning. Memorandum No. UCB/ERL/M77/58. Presented at 9th International Machine Intelligence Symposium, Leningrad.

[7] KACPRZYK, J. (1976). Fuzzy-set-theoretic approach to the optimal assignment of work places. Proc. IFAC Symp. on Large Scale Systems Theory and Applic., Udine, Italy, 123-132.

[8] ZADEH, L.A. (1972). A fuzzy-set-theoretic interpretation of linguistic hedges. J. Cybernetics 2, 4-34.

[9] LAKOFF, G. (1973). Hedges: a study in meaning criteria and the logic of fuzzy concepts. Journal of Philosophical Logic 2, 458-408.

[10] SEBESTYEN, G.S. (1962). Decision-Making Process in Pattern Recognition. Macmillan Co., New York.

[11] WENSTOP, F. (1975). Application of linguistic variables in the analysis of organizations. Ph.D. Dissertation, School of Business Administration, University of California, Berkeley.

[12] KACPRZYK, J. (1976). Application of linguistic variables and fuzzy conditional statements for the description of complex systems. Proc. 3rd National Conference on Information Sciences, Katowice.

[13] KAUFMANN, A. (1975). Introduction to the Theory of Fuzzy Subsets, Vol. II. Applications to Linguistics, Logic and Semantics. Masson and Co., Paris.

[14] PROCYK, T.J. (1976). Linguistic representation of fuzzy variables. Fuzzy Logic Working Group, Queen Mary College, London.

[15] BELLMAN, R.E. and ZADEH, L.A. (1976). Local and fuzzy logics. Memorandum No. ERL-M584, Electronics Research Laboratory, University of California, Berkeley.

[16] ZADEH, L.A. (1977). Theory of fuzzy sets. Memorandum No. UCB/ERL/M77/1, University of California, Berkeley.

[17] KNUTH, D. (1968). Semantic of context free language. Math. System Theory 2, 127-147.

[18] KANAL, L. (1974). Patterns in pattern recognition: 1968-1974. IEEE Trans. on Inf. Theory IT-20, 6, 697-722.

[19] DE LUCA, A. and TERMINI, S. (1971). A definition of a nonprobabilistic entropy in the setting of fuzzy sets theory. Information and Control 20, 301-312.

[20] HERSCH, H.M. and CARAMAZZA, A. (1975). A fuzzy set approach to modifiers and vagueness in natural language. J. Experimental Psychology 105, 3, 254-276.

[21] KNOPFMACHER, J. (1975). On measures of fuzziness. J. Math. Anal. Appl. 49, 529-534.

[22] DE LUCA, A. and TERMINI, S. (1977). On the convergence of entropy measures of a fuzzy set. Kybernetes 6, 219-227.

[23] ANDREWS, H.C. (1972). Mathematical Techniques in Pattern Recognition. J. Wiley and Sons, New York.

[24] AHO, A. and ULLMAN, J.D. (1973). The Theory of Parsing, Translation and Compiling. Prentice-Hall, Englewood Cliffs, New Jersey.

[25] ZADEH, L.A. (1977). Fuzzy sets as a basis for a theory of possibility. Memorandum No. UCB/ERL/M77/12, Electronics Research Laboratory, University of California, Berkeley.

[26] ZADEH, L.A. (1977). PRUF - a meaning representation language for natural language. Memorandum No. UCB/ERL/M77/61, Electronics Research Laboratory, University of California, Berkeley.

[27] BHATTACHARRYA, A. (1943). On a measure of divergence between two statistical populations defined by their probability distributions. Bull. Calcutta Math. Soc. 35, 99-109.

[28] KAILATH, T. (1967). The civergence and Bhattacharrya distance measure in signal selection. IEEE Trans. Comm. Tech. 15, 1, 52-60.

[29] JEFFREYS, H. (1946). An invariant form for the prior probability in estimation problems. Proc. Royal Soc. A 186, 453-461.

[30] KITTLER, J. (1974). Mathematical methods of feature selection in pattern recognition. Int. J. Man-Machine Studies 7, 609-637.

[31] ZADEH, L.A. (1971). Similarity relations and fuzzy orderings. Information Sciences 3, 177-200.

12. Appendices

12.1 Illustrative Example

In order to illustrate this method, a couple of examples are given. In these examples a small language (containing 32 sentences) was chosen. This reduced language is a non-fuzzy subset of L(G1) and it is listed on the next page. In the first example an unlabelled fuzzy set was created and (for a tolerance $E = 0.4$) three labels were selected in the prescreening process: HIGH, MOREORLESS HIGH, SORTOF HIGH. Finally HIGH was selected among the three labels to be the linguistic approximation to the unlabelled one.

In the second example we created a 3 by 2 decision table TOT:

TOT =

X	Y
LOW	HIGH
MIDDLE	MIDDLE
SLIGHTLY HIGH	MOREORLESS HIGH

which can be interpreted as:

if X is LOW then Y is HIGH or

if X is MIDDLE then Y is MIDDLE or

if X is SLIGHTLY HIGH then Y is MOREORLESS LOW.

Then the following compositions were performed:

$$T_1 = \text{MIDDLE} \circ \text{TOT}$$

$$T_2 = (\text{SLIGHTLY HIGH}) \circ \text{TOT}$$

$$T_3 = (\text{VERY HIGH}) \circ \text{TOT}$$

The linguistic approximation for T1, T2, T3 was found using the above described method, yielding these results:

LA[T1] = MIDDLE

LA[T2] = MOREORLESS LOW

LA[T3] = MOREORLESS LOW

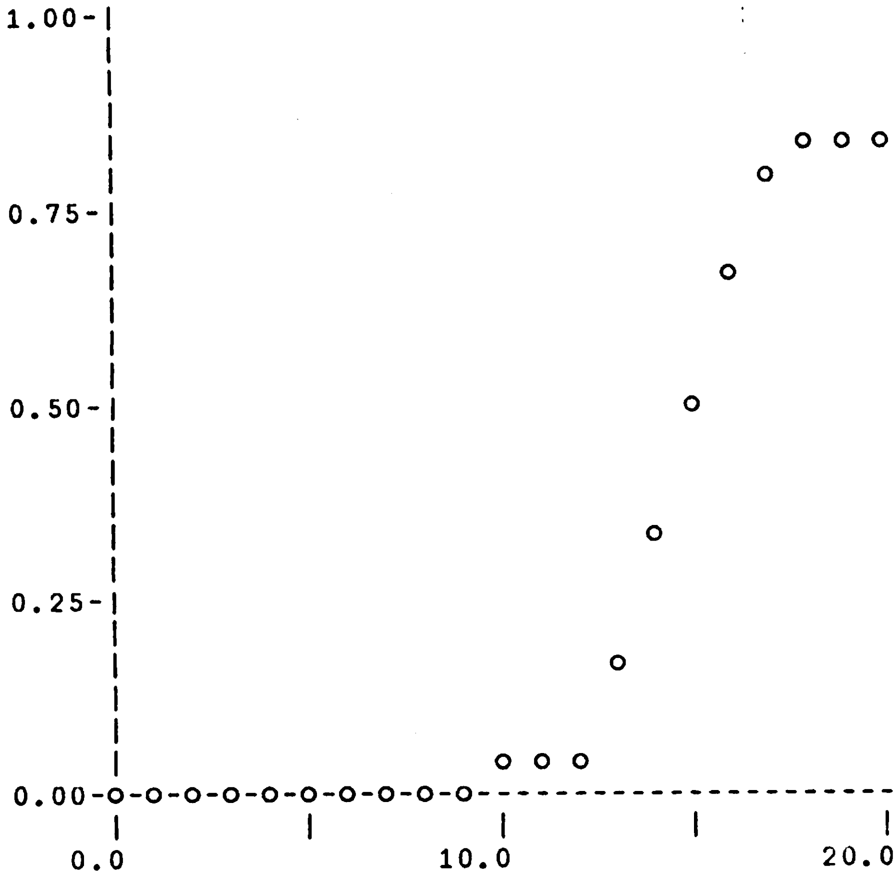
It is important to note that in our very limited language we did not have any not normalized fuzzy set.

This is the reason why the linguistic approximation to T3, which is sub-normal, is a normal fuzzy set.

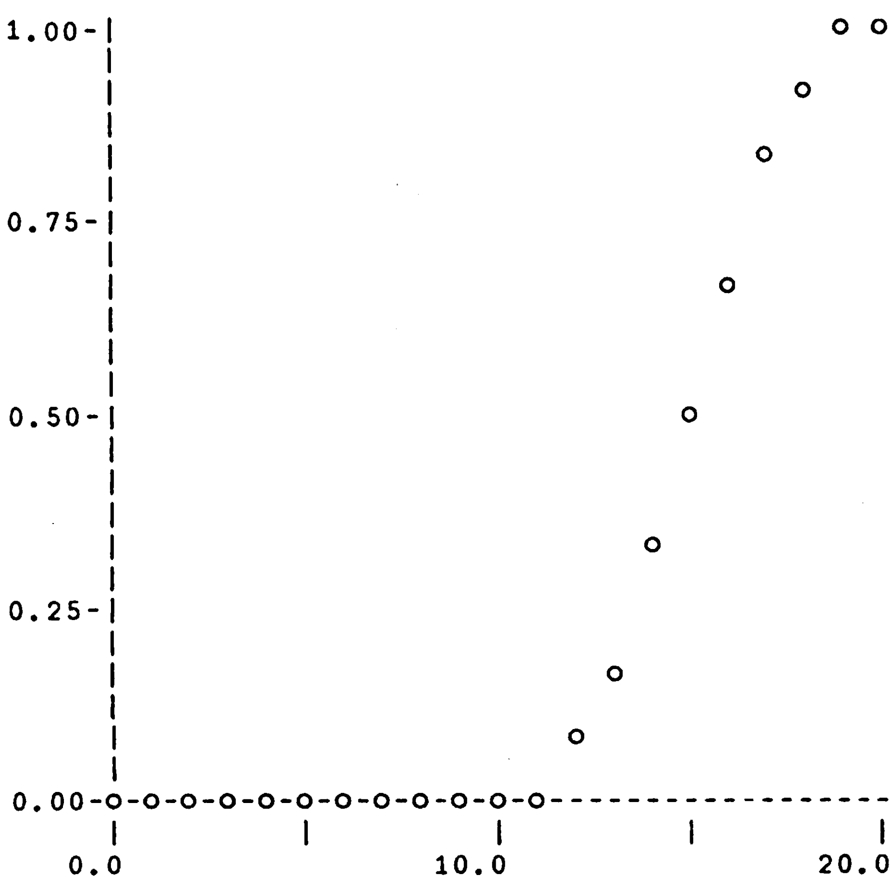
This can be solved by using a larger language which includes connectors like AND whose application creates sub-normal fuzzy sets.

The results are shown on the following pages.

```
      OBS←OBS,' '
      OBS←OBS,A2LAST
      OBS←OBS,A1LAST

      OBS

LOW                        1
MIDDLE                     2
HIGH                       3
VERY VERY LOW              4
VERY VERY MIDDLE           5
VERY VERY HIGH             6
VERY       LOW             7
VERY       MIDDLE          8
VERY       HIGH            9
MOREORLESS LOW            10
MOREORLESS MIDDLE         11
MOREORLESS HIGH           12
SORTOF     LOW            13
SORTOF     MIDDLE         14
SORTOF     HIGH           15
NOT LOW                   16
NOT MIDDLE                17
NOT HIGH                  18
NOT VERY VERY LOW         19
NOT VERY VERY MIDDLE      20
NOT VERY VERY HIGH        21
NOT VERY       LOW        22
NOT VERY       MIDDLE     23
NOT VERY       HIGH       24
NOT MOREORLESS LOW        25
NOT MOREORLESS MIDDLE     26
NOT MOREORLESS HIGH       27
NOT SORTOF     LOW        28
NOT SORTOF     MIDDLE     29
NOT SORTOF     HIGH       30
ALL                       31
NONE                      32
```

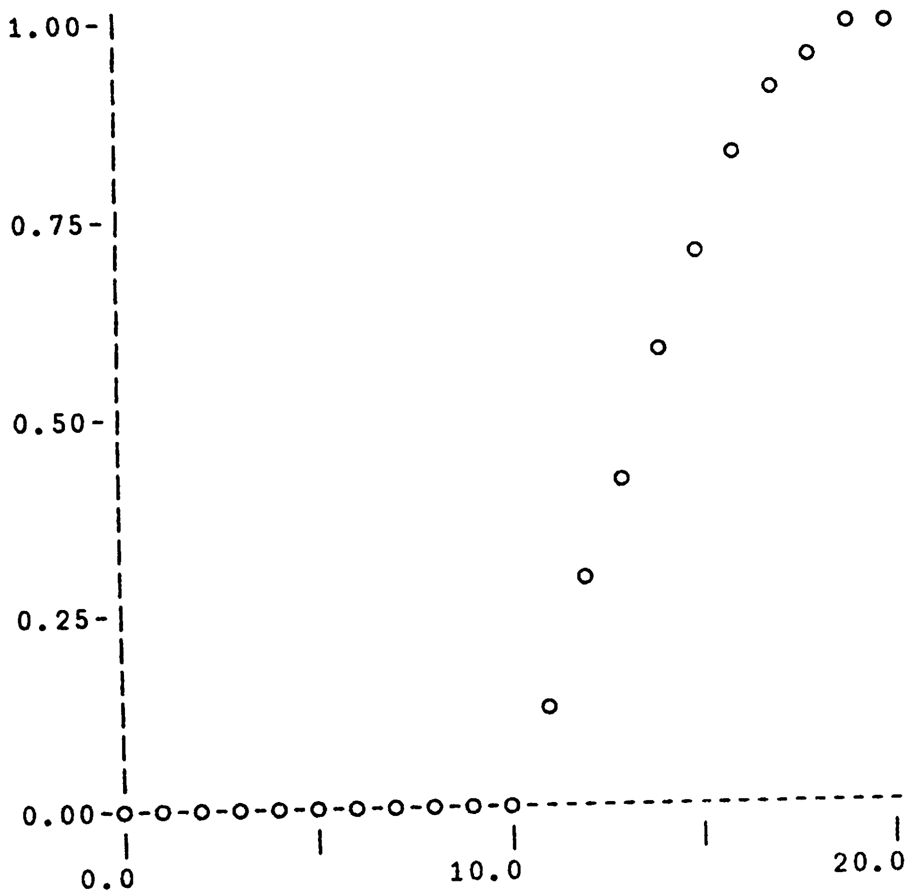


A *THIS IS THE UNLABELLED FUZZY SET*

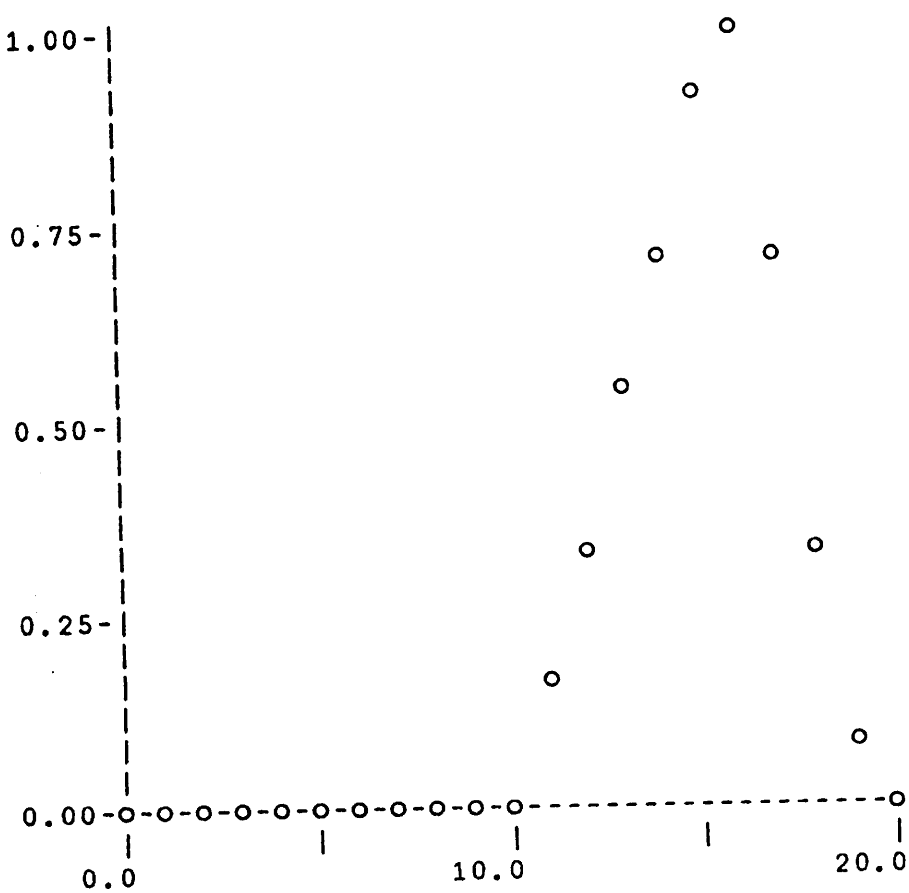


A *THIS IS THE MEMBERSHIP DISTRIBUTØON OF THE SELECTED LABEL 'HIGH'*

A *THESE ARE THE MEMBERSHIP DISTRIBUTIONS OF THE OTHER TWO POSSIBLE CANDIDATES*
A *TO THE LINGUISTIC APPROXIMATION OF THE UNLABELLED FUZZY SET.*



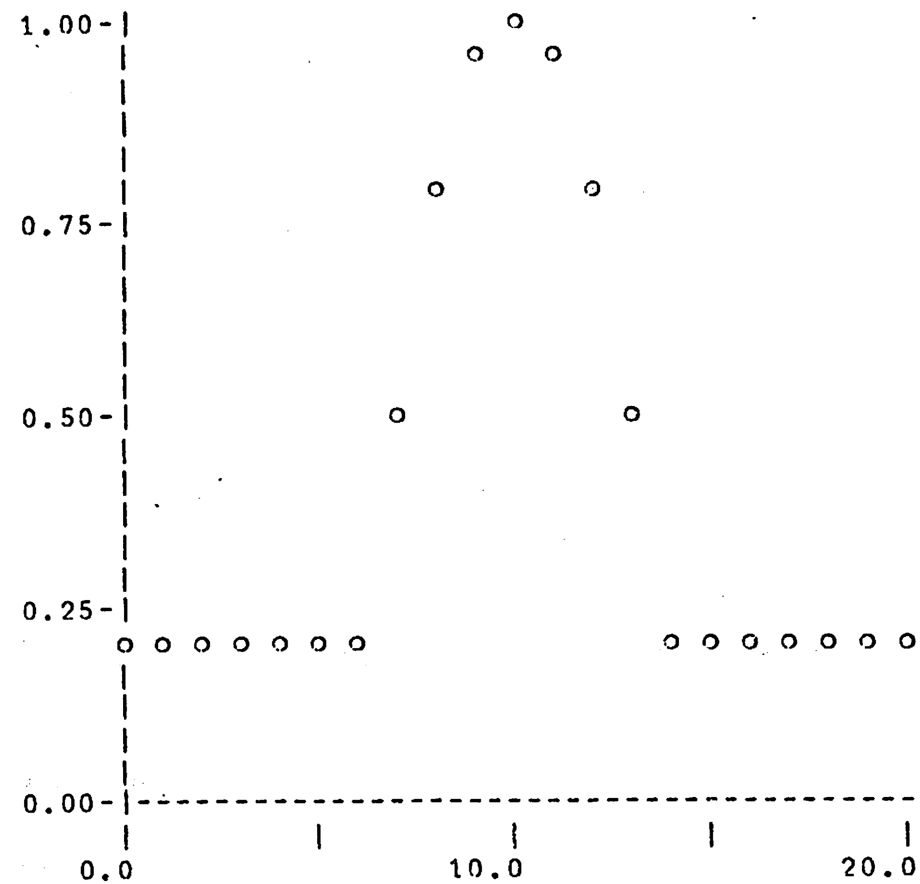
A *THIS CORRESPONDS TO 'MOREORLESS HIGH'*



A *THIS CORRESPONDS TO 'SORTOF HIGH'*

```
      VEC1 ← LOW , HIGH
      VEC2 ← MIDDLE , MIDDLE
      VEC3 ← (SLIGHTLY HIGH) , (MOREORLESS LOW)

      TOT ← TABLE2 VEC1
      TOT ← TOT OR TABLE2 VEC2
      TOT ← TOT OR TABLE2 VEC3

      T1← MIDDLE COMPOSITION2 TOT

      T2← (SLIGHTLY HIGH) COMPOSITION2 TOT

      T3← (VERY HIGH) COMPOSITION2 TOT

      T1
0.22  0.22  0.22  0.22  0.22  0.22  0.22  0.5  0.78  0.94  1  0.94  0.78  0.5  0.22  0.22  0.22  0.22  0.22
      0.22  0.22

      T2
1  0.99  0.96  0.91  0.82  0.71  0.57  0.42  0.28  0.22  0.22  0.22  0.22  0.22  0.22  0.06  0  0  0  0
      0

      T3
0.67  0.67  0.67  0.67  0.67  0.67  0.57  0.42  0.28  0.14  0.1  0.1  0.1  0.1  0.1  0.06  0  0  0  0  0
```

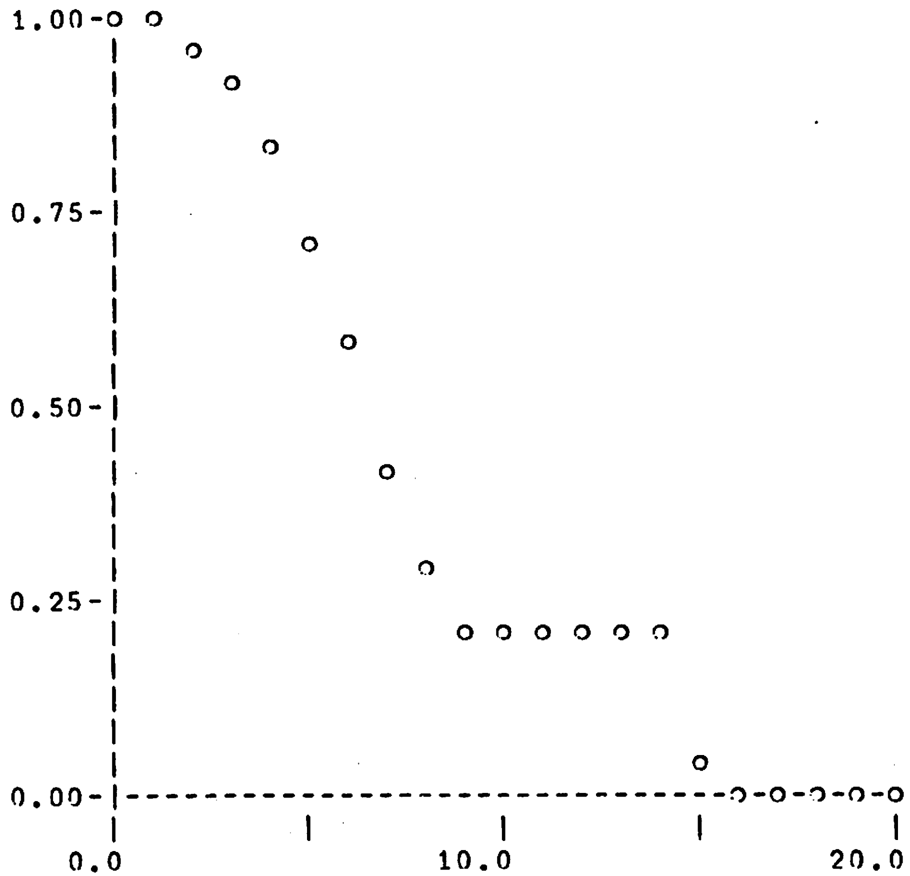
```
X←X1←¯1+⍳D
Y←T1
X←⍉2 21⍴X,Y

PLOT X
```




```
T1
0.22  0.22  0.22  0.22  0.22  0.22  0.22  0.5  0.78  0.94  1  0.94  0.78  0.5  0.22  0.22  0.22  0.22  0.22
      0.22  0.22
```

```
X←X1
Y←T2
Z←⍉2 21⍴X,Y
```

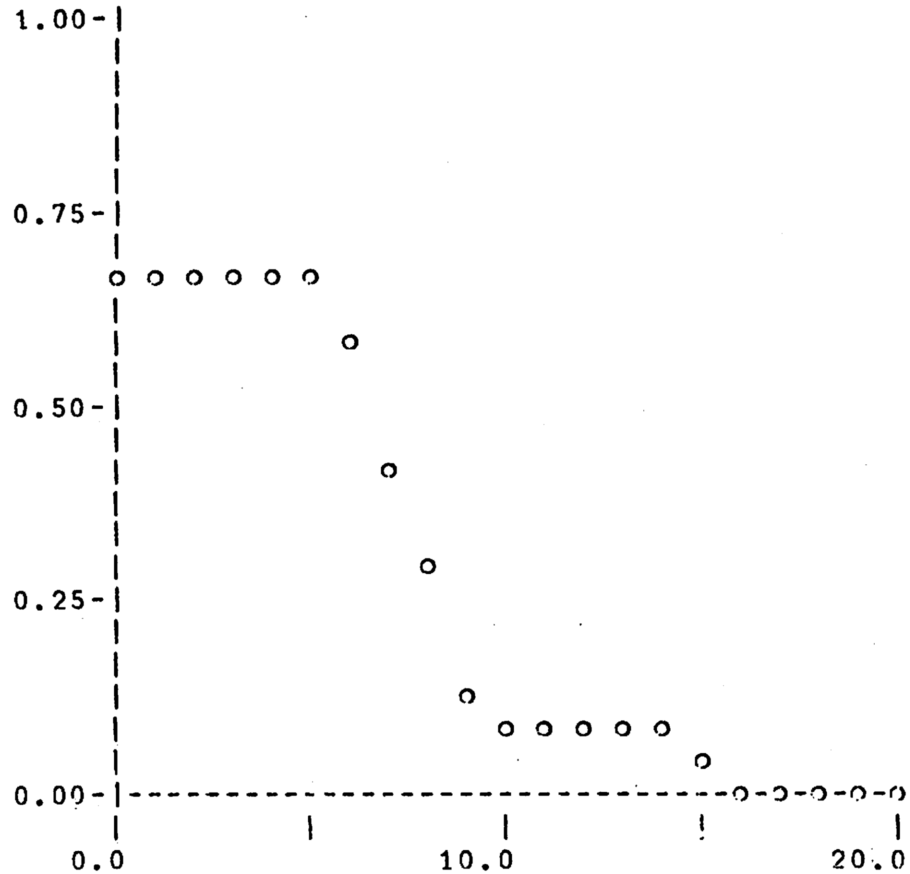
PLOT Z



T2

```
1  0.99  0.96  0.91  0.82  0.71  0.57  0.42  0.28  0.22  0.22  0.22  0.22  0.22  0.22  0.06  0  0  0  0
      0
```

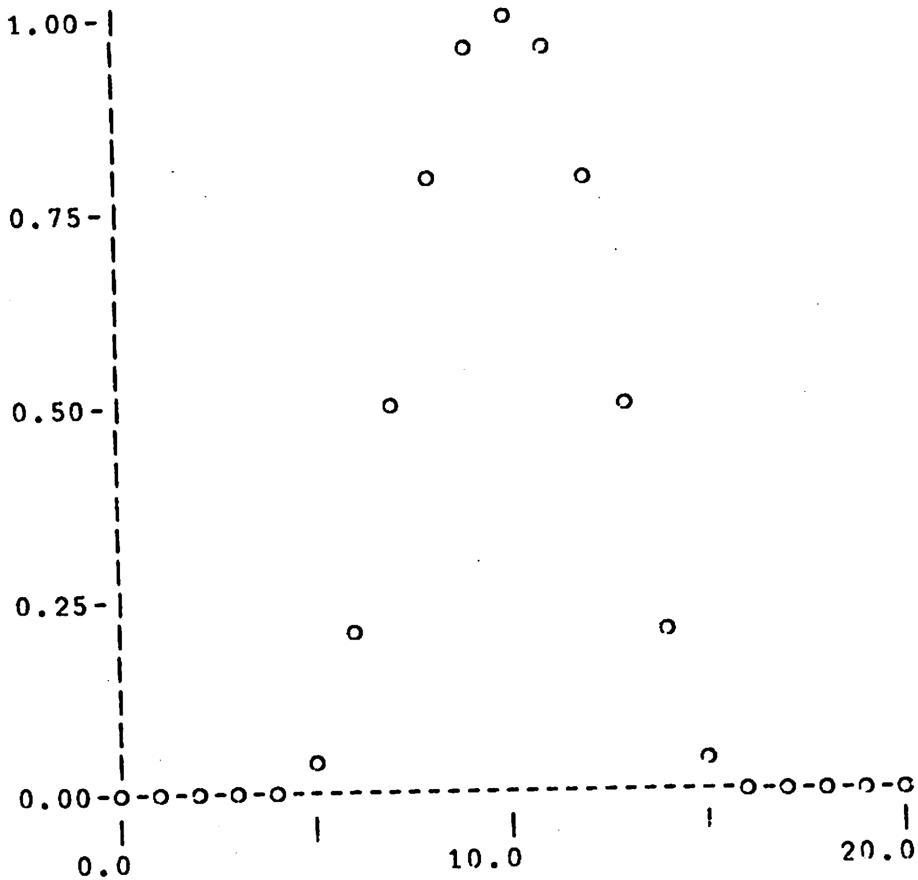
```
X←X1
Y←T3
Z←⍉2 21ρX,Y

PLOT Z
```




```
T3
0.67  0.67  0.57  0.57  0.67  0.67  0.57  0.42  0.28  0.14  0.1  0.1  0.1  0.1  0.1  0.05  0  0  0  0  0
```

```
X←X1
Y←MIDDLE
Z←⍉2 21⍴X,Y
```

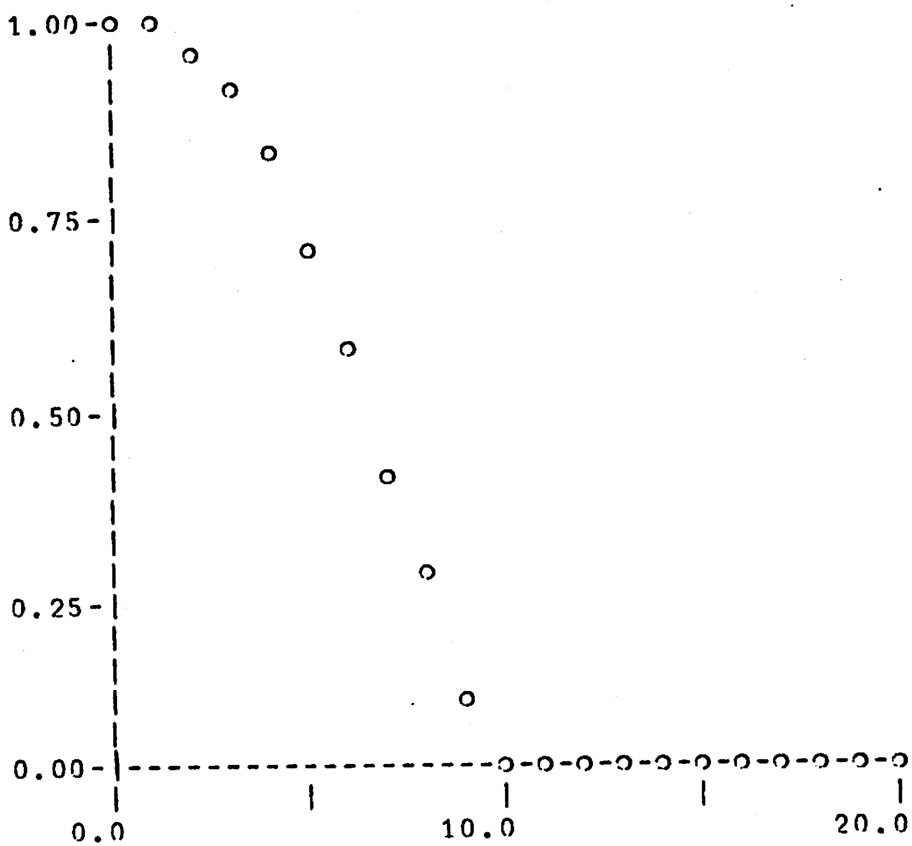
```
PLOT Z
```




```
MIDDLE
0 0 0 0 0 0.05 0.22 0.5 0.78 0.94 1 0.94 0.79 0.5 0.22 0.05 0 0 0 0 0
```

```
X←X1
Y←MOREORLESS.LOW
Z←⍉2 21⍴X,Y
```

PLOT Z



MOREORLESS LOW

1 0.99 0.96 0.91 0.82 0.71 0.57 0.42 0.28 0.14 0 0 0 0 0 0 0 0 0 0 0

12.2 Basic Concepts of Fuzzy Sets

The purpose of this appendix is to give a very brief introduction to the basic concepts of Fuzzy Set Theory. Its scope is limited. For a much more complete treatment please refer to [1,15,16].

Let U be a collection of objects or concepts $\{u\}$ and let us refer to U as the 'universe of discourse'. Let A, B, C be three fuzzy subsets of U. Let $\mu_A(u)$ be a function which maps from U into $[0,1]$ and let us refer to $\mu_A(u)$ as the 'membership function of A'. We can represent A as

$$A = \bigcup_{\forall u \in U} \mu_A(u) \ .$$

Given a universe U, we say that $A \subset U$ is fully characterized by its membership function $\mu_A(u)$. We define the following operations on fuzzy sets, based on their corresponding membership function.

Equality (=)

$$A = B \quad \text{iff} \quad \mu_A(u) = \mu_B(u) \quad \text{for all} \quad u \in U$$

Containment ($\subset$)

$$A \subset B \quad \text{iff} \quad \mu_A(u) < \mu_B(u) \quad \text{for all} \quad u \in U$$

Union (OR)

$$C = A \text{ OR } B \quad \text{iff} \quad \mu_C(u) = \max_{\substack{\text{for all} \\ u \in U}} [\mu_A(u), \mu_B(u)]$$

Intersection (AND)

$$C = A \text{ AND } B \quad \text{iff} \quad \mu_C(u) = \min_{\substack{\text{for all} \\ u \in U}} [\mu_A(u), \mu_B(u)]$$

Complementation (NOT)

$$B = \text{NOT } A \quad \text{iff} \quad \mu_B(u) = 1 - \mu_A(u)$$

Union, intersection and complementation satisfy De Morgan's laws, as well as the associative and distributive properties.

Definition of an 'alpha-level-set'. An alpha level of a fuzzy subset A is

$$A_{alpha} = \{u \mid \mu_A(u) \geq alpha\} \quad .$$

Then A satisfies the resolution identity

$$A = \bigcup_{\forall\, alpha \in [0,1]} [alpha\ A_{alpha}] \;.$$

The 'bandwidth' of A is the 0.5 level-set of A. The 'support' of A is defined as

$$\text{Support}(A) = \{u \mid \mu_A(u) > 0\} \;.$$

Other useful operations on fuzzy sets are:

Bounded Sum ($\oplus$)

$$C = A \oplus B \rightarrow \mu_C(u) = \bigcup_{\forall u \in U} [1 \text{ AND } (\mu_A(u) \oplus \mu_B(u))]$$

Bounded Difference ($\ominus$)

$$C = A \ominus B \rightarrow \mu_C(u) = \bigcup_{\forall u \in U} [0 \text{ OR } (\mu_A(u) - \mu_B(u)]$$

Raising a fuzzy set A to a real number power 'beta',

$$B = A^{beta} \rightarrow \mu_B(u) = [\mu_A(u)]^{beta} \;,$$

The linguistic hedges used in the grammar proposed in Section 5, when applied to the label of a fuzzy set A, are the operators defined as:

$$\text{VERY } A = A^2$$

$$\text{MOREORLESS } A = A^{1/2}$$

Implication ($\Rightarrow$)

Several ways of defining implication could be listed. We will limit ourselves to only two of them.

$$A \Rightarrow B = [A \times B] \text{ OR } [(\text{NOT } A) \times V]$$

or

$$A \Rightarrow B = [(\text{NOT } (A \times V)) \oplus (U \times B)]$$

where U and V are the universes of discourse of A and B respectively. $\times$ denotes the cartesian product which is defined as:

$$C = A \times B \rightarrow \mu_C(u_i, u_j) = \bigcup_{\substack{\text{for all} \\ u \in U}} [\mu_A(u_i) \text{ AND } \mu_B(u_j)] \quad .$$

12.3 <u>List of APL Programs</u>

A list of APL programs follows.

```
      )LOAD FUZZY1
SAVED  17.13.48 06/19/78

      ∇START[⎕]∇

    ∇ START;M
[1]   'ENTER DIMENSION'
[2]   D←⎕
[3]   HIGH←(⌊((D-1)÷2)) SFN(D-1)
[4]   LOW←⌽HIGH
[5]   M←⌈((D-1)÷4)
[6]   MIDDLE←(M+1×(⌊(M÷2))≠(M÷2)) PFNCT(⌈((D-1)÷2))
[7]   ALL←Dρ1
[8]   NONE←Dρ0
    ∇


      ∇SFN[⎕]∇

    ∇ S←A SFN C;B;X1;X2;X3;X4
[1]   B←⌈(A+C)÷2
[2]   S←X1←(A+1)ρ0
[3]   S←S,X2←2×(((A+⍳(B-A))-A)÷(C-A))*2
[4]   S←S,X3←1-(2×(((B+⍳(C-B))-C)÷(C-A))*2)
[5]   S←S,X4←(D-(C+1))ρ1
[6]   S←0.01×⌊0.5+100×S
    ∇


      ∇PFNCT[⎕]∇

    ∇ P←B PFNCT C;X1;X2
[1]   P←X1←((¯1+⍳D)≤C)/((C-B) SFN C)
[2]   P←P,X2←((¯1+⍳D)>C)/(1-(C SFN(C+B)))
[3]   P←0.01×⌊0.5+100×P
    ∇


      ∇TABLE2[⎕]∇

    ∇ R←TABLE2 Z;X1;X2
[1]   X1←Z[⍳D]
[2]   X2←DρZ[D+⍳D]
[3]   R←X1 RELATION X2
    ∇


      ∇RELATION[⎕]∇

    ∇ Y←A RELATION B
[1]   Y←A∘.⌊B
    ∇


      ∇COMPOSITION2[⎕]∇

    ∇ Y←A COMPOSITION2 B
[1]   Y←A⌈.⌊B
    ∇
```

```
∇MOREORLESS[□]∇

    ∇ Y←MOREORLESS A
[1]   Y←A*0.5
[2]   Y←0.01×⌊0.5+100×Y
    ∇


∇SORTOF[□]∇

    ∇ Y←SORTOF A
[1]   Y←NORM((NOT(VERY VERY A)) AND(DIL A))
    ∇


∇NORM[□]∇

    ∇ Y←NORM A
[1]   Y←A÷(⌈/A)
[2]   Y←0.01×⌊0.5+100×Y
    ∇


∇VERY[□]∇

    ∇ Y←VERY A
[1]   Y←A×A
[2]   Y←0.01×⌊0.5+100×Y
    ∇


∇DIL[□]∇

    ∇ Y←DIL A
[1]   Y←MOREORLESS A
    ∇


∇OR[□]∇

    ∇ Y←A OR B
[1]   Y←A⌈B
    ∇


∇AND[□]∇

    ∇ Y←A AND B
[1]   Y←A⌊B
    ∇


∇SLIGHTLY[□]∇

    ∇ Y←SLIGHTLY A
[1]   Y←INT(NORM((PLUS A) AND(NOT(VERY A))))
[2]   Y←0.01×⌊0.5+100×Y
    ∇


∇INT[□]∇

    ∇ Y←INT A
[1]   Y←NONE
[2]   Y[(A≤0.5)/⍳D]←2×(A[(A≤0.5)/⍳D])*2
[3]   Y[(A≥0.5)/⍳D]←1-(2×(1-A[(A≥0.5)/⍳D])*2)
[4]   Y←0.01×⌊0.5+100×Y
    ∇
```

```
      ∇PLUS[⎕]∇

    ∇ Y←PLUS A
[1]   Y←A*(¯1+(5)*0.5)
[2]   Y←0.01×⌊0.5+100×Y
    ∇
```

```
      ∇HAMMINGDISTANCE[⎕]∇

    ∇ R←A HAMMINGDISTANCE B
[1]   R←(+/(|(A-B)))÷D
    ∇
```

```
      ∇SQUERR[⎕]∇

    ∇ Y←A SQUERR B;Y1
[1]   Y←((+/(A-B)*2)*0.5)÷(D*0.5)
[2]   Y←Y,Y1←(⌈/((A-B)*2))
    ∇
```

```
      ∇BHATTADISTANCE[⎕]∇

    ∇ R←A BHATTADISTANCE B;R1
[1]  ⍝ IT COMPUTES THE BHATTACHARYYA DISTANCE
[2]   A1←A÷(1×((+/A)=0)+(+/A))
[3]   B1←B÷(1×((+/B)=0)+(+/B))
[4]   R1←(A1×B1)*(0.5)
[5]   R←(1-(+/R1))*0.5
    ∇
```

```
      ∇POWER[⎕]∇

    ∇ Y←POWER A;Y1
[1]   Y←(+/A)÷D
[2]   Y←Y,Y1←+/(A>0)
    ∇


      ∇ENTROPY[⎕]∇

    ∇ Y←ENTROPY A;B;C;V;V1
[1]   B←A[(A>0)/⍳D]
[2]   C←(1-A)[((1-A)>0)/⍳D]
[3]   V←NONE
[4]   V1←NONE
[5]   V[(A>0)/⍳D]←-(B×⍟B)
[6]   V1[((1-A)>0)/⍳D]←-(C×⍟C)
[7]   Y←(+/(V+V1))÷ENTROPNORM
    ∇


      ∇ENTROPNORM[⎕]∇

    ∇ Y←ENTROPNORM;A
[1]   A←0.5×ALL
[2]   Y←2×(+/-(A×⍟A))
    ∇


      ∇FIRSTMOMENT[⎕]∇

    ∇ R←FIRSTMOMENT A;A1
[1]   A1←A÷(1×((+/A)=0)+(+/A))
[2]   R←+/((¯1+⍳D)×A1)
    ∇


      ∇SKULNESS[⎕]∇

    ∇ R←SKULNESS A
[1]   R←(THIRDMOMENT A)-((3×(FIRSTMOMENT A)×(SPREAD A))+(FIRSTMOMENT A)*3)
    ∇


      ∇SPREAD[⎕]∇

    ∇ R←SPREAD A
[1]   R←(SECONDMOMENT A)-(FIRSTMOMENT A)*2
    ∇


      ∇SECONDMOMENT[⎕]∇

    ∇ R←SECONDMOMENT A;A1
[1]   A1←A÷(1×((+/A)=0)+(+/A))
[2]   R←+/(((¯1+⍳D)*2)×A1)
    ∇


      ∇THIRDMOMENT[⎕]∇

    ∇ R←THIRDMOMENT A;A1
[1]   A1←A÷(1×((+/A)=0)+(+/A))
[2]   R←+/(((¯1+⍳D)*3)×A1)
    ∇
```

```
    ∇GRAMMAR[⎕]∇
  ∇ GRAMMAR
[1]   T← 3 6 ρ'LOW   ','MIDDLE','HIGH  '
[2]   H← 3 10 ρ'VERY      ','MOREORLESS','SORTOF    '
[3]   C← 2 3 ρ'AND','OR '
[4]   U← 2 4 ρ'ALL ','NONE'
[5]   V←4ρ'VERY'
[6]   OP←'('
[7]   CP←')'
[8]   SP←' '
[9]   N←3ρ'NOT'
[10]  B1←B1,SP,(B1←((ρT)[1],(ρV))ρV),SP,T
[11]  B2←(((ρH)[1]×(ρT)[1]),(ρT)[2])ρ,T
[12]  B3←((ρT)[1],((ρH)[1]×(ρH)[2]))ρ,H
[13]  I←0
[14]  B4←(((ρH)[1]×(ρT)[1]),(ρH)[2])ρSP
[15] L:B4[(I×(ρT)[1])+ι((ρT)[1]);]←B3[;(I×(ρH)[2])+ι((ρH)[2])]
[16]  I←I+1
[17]  →(I<(ρH)[1])/L
[18]  B4←B4,SP,B2
[19]  B←(T,((ρT)[1],(1+(ρH)[2]))ρSP),[1](B1,((ρB1)[1],((ρH)[2]-1+2×(ρV)))ρSP),[1] B4
[20]  A←(B,((ρB)[1],(ρN)+1)ρSP),[1](A←(((ρB)[1],(ρN))ρN),SP,B)
[21]  IX←0
[22]  J←1
[23]  S←(((ρC)[1]×((ρA)[1])×(¯1+(ρA)[1])÷2),(3+((ρC)[2])+2×((ρA)[2])))ρSP
[24] L1:IN←1
[25] L2:K←1
[26] L3:IX←IX+1
[27]  S[IX;]←OP,A[IN;],CP,C[J;],SP,A[(IN+K);]
[28]  K←K+1
[29]  →((IN+K)≤((ρA)[1]))/L3
[30]  IN←IN+1
[31]  →(IN<((ρA)[1]))/L2
[32]  J←J+1
[33]  →(J≤((ρC)[1]))/L1
  ∇
```

```
      )LOAD COLUMNS
VED  13.42.10 06/07/78

      ∇ORDER[⎕]∇
    ∇ B←ORDER A
      B←⍋A
    ∇


      ∇LENGTH[⎕]∇
    ∇ LE←LENGTH V
      LE←((V=1)/⍳(⍴V)),((⍴V)+1)
      LE←(1↓LE)-(¯1↓LE)
    ∇


      ∇BOOLVECT[⎕]∇
    ∇ VD←BOOLVECT C
      VD←1,(C[⍳(¯1+(⍴C))]≠C[1+⍳(¯1+(⍴C))])
    ∇


      ∇ORDERNEXT[⎕]∇
    ∇ IN←LE ORDERNEXT C;TEMP;I;ST;LEN
      I←1
      ST←0
      LEN←LE[1]
      IN←⍳0
     LOOP:TEMP←C[ST+⍳LEN]
      IN←IN,(ST+(⍋TEMP))
      I←I+1
      →(I>(⍴LE))/END
      ST←LEN+ST
      LEN←LE[I]
      →LOOP
     END:
    ∇


      ∇SEARCHTREE[⎕]∇
    ∇ R←SEARCHTREE A;BOLA;BOLB;BOLC;BOLD;ERRA;ERRB;ERRC
      PB←((BOLA←((SUM←ERRA←LK[2]×WE[2]×((A[2]-K2)*2))≤(EPSI*2)))/⍳(⍴K2))
      →((+/BOLA)=0)/NIHIL
      PC←((BOLB←((SUM[PB]←SUM[PB]+(ERRB←LK[1]×WE[1]×((A[1]-K1[PB])*2)))≤(EPSI*2)))/PB)
      →((+/BOLB)=0)/NIHIL
      PD←((BOLC←((SUM[PC]←SUM[PC]+(ERRC←LK[3]×WE[3]×((A[3]-K3[PC])*2)))≤(EPSI*2)))/PC)
      →((+/BOLC)=0)/NIHIL
      R←((BOLD←((SUM[PD]←SUM[PD]+(ERRD←LK[4]×WE[4]×((A[4]-K4[PD])*2)))≤(EPSI*2)))/PD)
      →((+/BOLD)≠0)/END
     NIHIL:R←⍳0
     END:
    ∇
```

```
      )LOAD PARAM
SAVED  13.34.42 06/07/78

      ∇LA[⎕]∇

    ∇ Y←SUR LA DIST
[1]    PNTR←K5[SUR]
[2]    ALRES←OBS[PNTR;]
[3]    CATRES
[4]    BD←DIST BHAD B
[5]    ED←DIST EUCD B
[6]    Y1←((⌊/BD)=BD)/⍳(⍴BD)
[7]    ED1←ED[¯1+2×⍳((⍴ED)÷2)]
[8]    Y2←((⌊/ED1)=ED1)/⍳(⍴ED1)
[9]    →(Y1≠Y2)/TWO
[10]   Y←ALRES[Y1;]
[11]   →END
[12]  TWO:EITH← 1 7 ⍴'EITHER '
[13]   ORALPH← 1 4 ⍴' OR '
[14]   Y←EITH,ALRES[Y1;],ORALPH,ALRES[Y2;]
[15]  END:
    ∇

      ∇CATRES[⎕]∇

    ∇ CATRES;AD;AU;STRING;IND
[1]     STRING←LITERAL(⍴PNTR)
[2]     AD←STRING[⍳(⍴PNTR)]
[3]     AU←STRING[(⍴PNTR)+⍳(⍴PNTR)]
[4]     AR←'←',ALRES
[5]     AR←']',AR
[6]     AR←';',AR
[7]     AR←AU,AR
[8]     AR←AD,AR
[9]     AR←'[',AR
[10]    AR←'B',AR
[11]    B←((⍴PNTR),D)⍴0
[12]    IND←1
[13]  LOOP:XEQ AR[IND;]
[14]    IND←IND+1
[15]    →(IND≤(⍴PNTR))/LOOP
    ∇

      ∇LITERAL[⎕]∇

    ∇ R←LITERAL N;AU;AD
[1]     R←AD←N⍴(9⍴' '),(10⍴'1'),(10⍴'2'),(10⍴'3'),(10⍴'4'),(10⍴'5'),(10⍴'6'),(10⍴'7'),(10⍴'8'),(10⍴'9')
[2]     R←R,AU←N⍴'1234567890'
    ∇
```

```
      ∇SEARCHEX3[⎕]∇
    ∇ Y←SEARCHEX3 A
[1]    WE←WEIGHT×LK
[2]    ER←((⍴K1),4)⍴0
[3]    ER[;1]←WE[1]×LK[1]×((A[1]-K1)*2)
[4]    ER[;2]←WE[2]×LK[2]×((A[2]-K2)*2)
[5]    ER[;3]←WE[3]×LK[3]×((A[3]-K3)*2)
[6]    ER[;4]←WE[4]×LK[4]×((A[4]-K4)*2)
[7]    R3←+/ER
[8]   HERE:Y←(BOLY←(R3<(EPSI*2)))/⍳⍴R3
[9]    →((+/BOLY)=0)/WARNING
[10]   →END
[11]  WARNING:' NO POINTS WITHIN THE GIVEN TOLERANCE EPSI = ';EPSI
[12]   ' '
[13]   ' THE CLOSEST POINT IS : '
[14]   Y←((⌊/R3)=R3)/⍳(⍴R3)
[15]   Y
[16]   ' '
[17]   ' ITS DISTANCE FROM THE GIVEN POINT IS : '
[18]   EP←(R3[Y])*0.5
[19]   EP
[20]   ' '
[21]   ' THE SQUARE OF THE COMPONENTS OF THE DISTANCE ARE : ='
[22]   ER[Y;]
[23]   ' '
[24]   EP←1.15×EP
[25]   'SUGGESTION: REPEAT THE PROCESS USING A NEW VALUE FOR EPSI'
[26]   '             THIS VALUE IS OBTAIN BY ASSION  EPSI ← EP     '
[27]   '             THIS NEW VALUE IS 15 PERCENT HIGHER THAN THE MINIMUM DISTANCE'
[28]   '             IN THIS CASE  EP = ';EP
[29]   ' '
[30]   'THIS WILL BE DONE AUTOMATICALLY BY THE PROGRAM'
[31]   EPSI←EP
[32]   →HERE
[33]  END:
    ∇


      ∇EXTRACT[⎕]∇
    ∇ Y←EXTRACT A;Y1;Y2;Y3
[1]    Y←(POWER A)[1]
[2]    Y←Y,Y1←ENTROPY A
[3]    Y←Y,Y2←FIRSTMOMENT A
[4]    Y←Y,Y3←SKULNESS A
    ∇
```

12.4 Formulae Derivations

Derivation for the expression for $|L(G1)|$:

$$b = t + ht + v^2t = t(h+1+v^2)$$

$$a = 2v \qquad = 2t(h+1+v^2)$$

$$s = a + u + c\sum_{i=1}^{a-1}(a-i)$$

$$= a + u + \frac{(a-1)a}{2}c$$

$$= u + t[2(h+1+v^2)] + t[2th^2+4th+4thv^2+4tv^2+2tv^4-h+2t-1-v^2]c$$

$$= u + t[2(h+1+v^2)] + t[2tc(h+1+v^2)^2 - c(h+1+v^2)]$$

$$= u + t(h+1+v^2)[2 + 2tc(h+1+v^2) - c]$$

Derivation for the expression for the volume of a hypersphere of radius e:

$$V = \int_{-e}^{e} dw \iiint_{R^3} dv = \int_{-e}^{e} dw\left(\tfrac{4}{3}\pi\rho^2\right) = 2\int_0^e \tfrac{4}{3}\pi\rho^3(w)dw$$

$$[\tfrac{4}{3}\pi\rho^3 = \text{volume of a sphere of radius } \rho \text{ in } \mathbb{R}^3]$$

Conditions: $x^2 + y^2 + z^2\ (= \rho^2) + w^2 = e^2 \Rightarrow \rho^2 + w^2 = e^2$

Then

$$2\rho d\rho = -2wdw \qquad w = \sqrt{e^2-\rho^2} \quad \Rightarrow \quad dw = -\frac{\rho}{\sqrt{e^2-\rho^2}}\, d\rho$$

$$w \to +e \to \rho = 0$$

$$w \to 0 \ \to \rho = e$$

$$V = 2\int_e^0 \left(\tfrac{4}{3}\pi\rho^3\right)\left(-\frac{\rho}{\sqrt{e^2-\rho^2}}\right)d\rho = \frac{8}{3}\pi \int_0^e \frac{\rho^4}{\sqrt{e^2-\rho^2}}\, d\rho$$

From tables: $\displaystyle\int \frac{x^4}{\sqrt{e^2-x^2}}\, dx = -\tfrac{1}{4}x^3\sqrt{e^2-x^2} - \tfrac{3}{8}e^2x\sqrt{e^2-x^2} + \tfrac{3}{8}e^4\sin^{-1}\frac{x}{|e|} + c = f(x)$

implies $$V = \frac{8}{3}\pi\big(f(e)-f(0)\big) = \frac{8}{3}\pi\left[\frac{3}{8}e^4\frac{\pi}{2}\right] .$$

$$\boxed{V = \frac{\pi^2}{2}\, e^4}$$

but $V = e'^4$ implies

$$\boxed{e' = \left(\sqrt{\frac{\pi}{\sqrt{2}}}\right)e}$$

12.5 Graphs and Tables

Pertinent graphs and tables follow.

2.6190476E¯1	2.5029317E¯1	2.6363636E0	5.4806912E0	1
2.8571429E¯1	3.0240237E¯1	1.0000000E1	¯1.7053026E¯13	2
2.6190476E¯1	2.5029317E¯1	1.7363636E1	¯5.4806912E0	3
1.6047619E¯1	1.6191073E¯1	1.4569733E0	1.6758724E0	4
1.6190476E¯1	1.9582174E¯1	1.0000000E1	¯5.6843419E¯14	5
1.6047619E¯1	1.6191073E¯1	1.8543027E1	¯1.6758724E0	6
2.0619048E¯1	2.0561410E¯1	1.9976905E0	3.3570716E0	7
2.1809524E¯1	2.4684148E¯1	1.0000000E1	1.7053026E¯13	8
2.0619048E¯1	2.0561410E¯1	1.8002309E1	¯3.3570716E0	9
3.2380952E¯1	2.7215771E¯1	3.2838235E0	6.3427147E0	10
3.5904762E¯1	3.2237294E¯1	1.0000000E1	¯5.6843419E¯14	11
3.2380952E¯1	2.7215771E¯1	1.6716176E1	¯6.3427147E0	12
2.3000000E¯1	2.9620703E¯1	4.9378882E0	1.3920596E0	13
3.0571429E¯1	3.0989855E¯1	1.0000000E1	5.6843419E¯14	14
2.3000000E¯1	2.9620703E¯1	1.5062112E1	¯1.3920596E0	15
7.3809524E¯1	2.5029317E¯1	1.2612903E1	¯1.6304146E1	16
7.1428571E¯1	3.0240237E¯1	1.0000000E1	¯5.6843419E¯14	17
7.3809524E¯1	2.5029317E¯1	7.3870968E0	1.6304146E1	18
8.3952381E¯1	1.6191073E¯1	1.1633012E1	¯7.8058532E0	19
8.3809524E¯1	1.9582174E¯1	1.0000000E1	0.0000000E0	20
8.3952381E¯1	1.6191073E¯1	8.3669881E0	7.8058532E0	21
7.9380952E¯1	2.0561410E¯1	1.2078584E1	¯1.1556179E1	22
7.8190476E¯1	2.4684148E¯1	1.0000000E1	2.8421709E¯13	23
7.9380952E¯1	2.0561410E¯1	7.9214157E0	1.1556179E1	24
6.7619048E¯1	2.7215771E¯1	1.3216197E1	¯2.0340746E1	25
6.4095238E¯1	3.2237294E¯1	1.0000000E1	2.2737368E¯13	26
6.7619048E¯1	2.7215771E¯1	6.7838028E0	2.0340746E1	27
7.7000000E¯1	2.9620703E¯1	1.1512059E1	¯1.1481841E2	28
6.9428571E¯1	3.0989855E¯1	1.0000000E1	3.4106051E¯13	29
7.7000000E¯1	2.9620703E¯1	8.4879406E0	1.1481841E2	30
1.0000000E0	0.0000000E0	1.0000000E1	1.1368684E¯13	31
0.0000000E0	0.0000000E0	0.0000000E0	0.0000000E0	32

A *THIS IS A 32 X 5 MATRIX THAT WE WILL USE AS AN EXAMPLE*

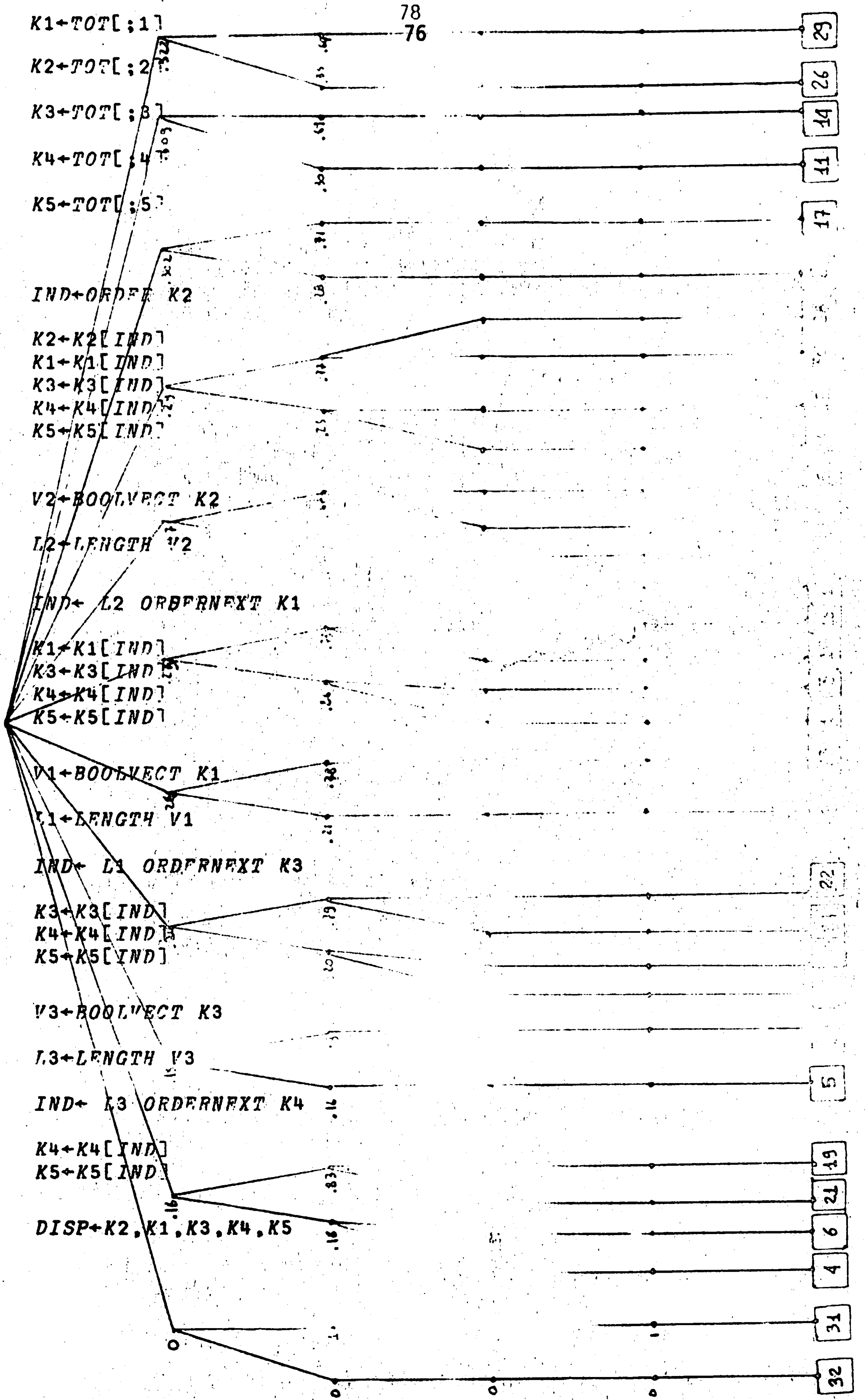
A *FOR THE ORDERING PROCEDURE*

```
      )COPY TREE TOT
SAVED  15.18.38 05/02/78

      ⍴TOT
32  4

      I1←ORDER  TOT[;1]
      K1←TOT[;1]
      K1←K1[I1]
      V1←BOOLVECT K1
      N1←+/V1

      N1
22

      K2←TOT[;2]
      I2←ORDER K2
      K2←K2[I2]
      V2←BOOLVECT K2
      N2←+/V2

      N2
11

      K3←TOT[;3]
      I3←ORDER K3
      K3←K3[I3]
      V3←BOOLVECT K3
      N3←+/V3

      N3
22

      K4←TOT[;4]
      I4←ORDER K4
      K4←K4[I4]
      V4←BOOLVECT K4
      N4←+/V4

      N4
29




      TOT←TOT,(⍳32)

      ⍴TOT
32  5

      '4E14.8,I4' ΔFMT TOT
```

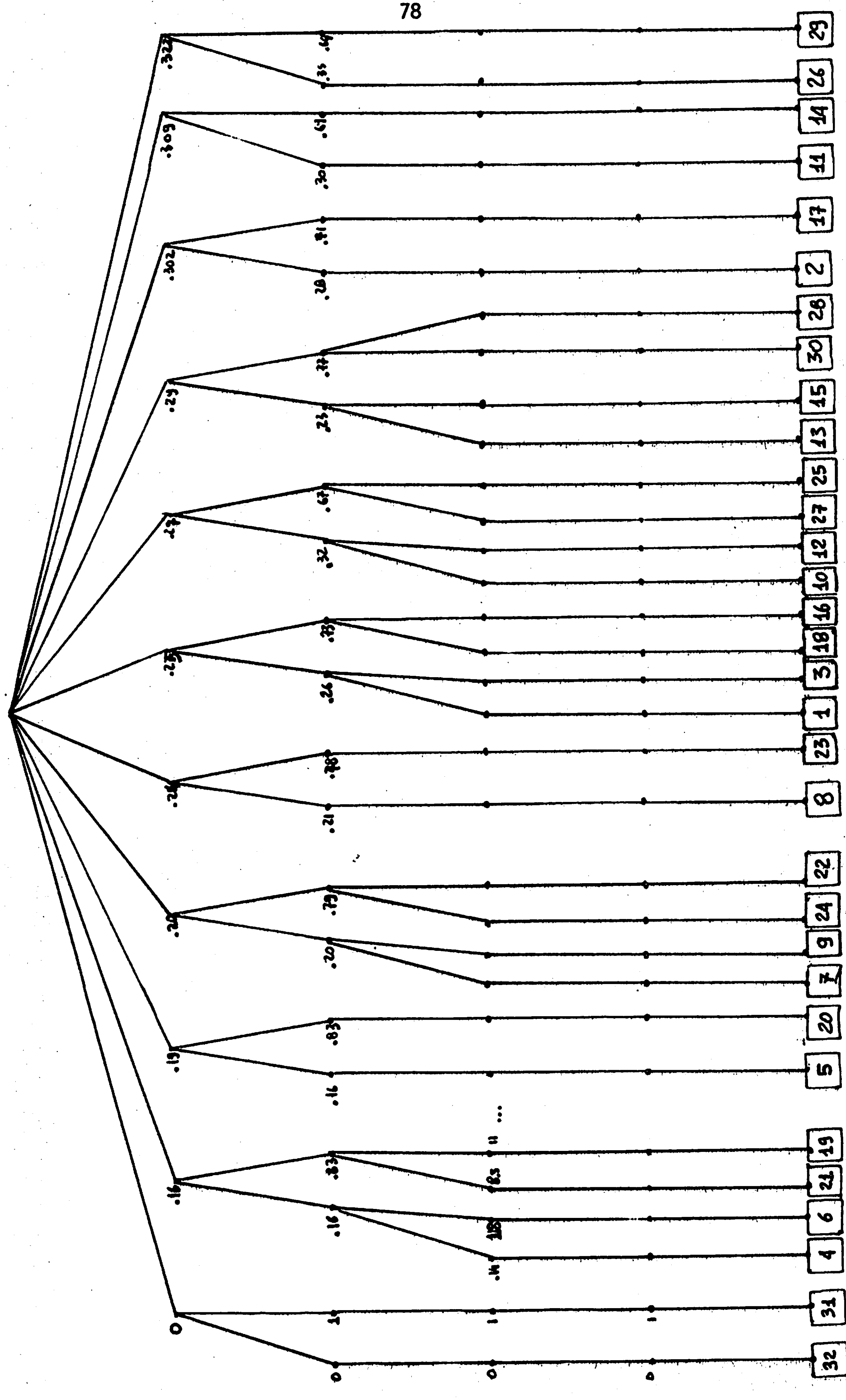
```
PERMTOT←⍉(K2,[.5]K1)
PERMTOT←PERMTOT,K3
PERMTOT←PERMTOT,K4
PERMTOT←PERMTOT,K5

'4E14.8,I4' ΔFMT PERMTOT
```

```
2.5029317E¯1   2.6190476E¯1   2.6363636E0    5.4806912E0      1
3.0240237E¯1   2.8571429E¯1   1.0000000E1   ¯1.7053026E¯13    2
2.5029317E¯1   2.6190476E¯1   1.7363636E1   ¯5.4806912E0      3
1.6191073E¯1   1.6047619E¯1   1.4569733E0    1.6758724E0      4
1.9582174E¯1   1.6190476E¯1   1.0000000E1   ¯5.6843419E¯14    5
1.6191073E¯1   1.6047619E¯1   1.8543027E1   ¯1.6758724E0      6
2.0561410E¯1   2.0619048E¯1   1.9976905E0    3.3570716E0      7
2.4684148E¯1   2.1809524E¯1   1.0000000E1    1.7053026E¯13    8
2.0561410E¯1   2.0619048E¯1   1.8002309E1   ¯3.3570716E0      9
2.7215771E¯1   3.2380952E¯1   3.2838235E0    6.3427147E0     10
3.2237294E¯1   3.5904762E¯1   1.0000000E1   ¯5.6843419E¯14   11
2.7215771E¯1   3.2380952E¯1   1.6716176E1   ¯6.3427147E0     12
2.9620703E¯1   2.3000000E¯1   4.9378882E0    1.3920596E0     13
3.0989855E¯1   3.0571429E¯1   1.0000000E1    5.6843419E¯14   14
2.9620703E¯1   2.3000000E¯1   1.5062112E1   ¯1.3920596E0     15
2.5029317E¯1   7.3809524E¯1   1.2612903E1   ¯1.6304146E1     16
3.0240237E¯1   7.1428571E¯1   1.0000000E1   ¯5.6843419E¯14   17
2.5029317E¯1   7.3809524E¯1   7.3870968E0    1.6304146E1     18
1.6191073E¯1   8.3952381E¯1   1.1633012E1   ¯7.8058532E0     19
1.9582174E¯1   8.3809524E¯1   1.0000000E1    0.0000000E0     20
1.6191073E¯1   8.3952381E¯1   8.3669881E0    7.8058532E0     21
2.0561410E¯1   7.9380952E¯1   1.2078584E1   ¯1.1566179E1     22
2.4684148E¯1   7.8190476E¯1   1.0000000E1    2.8421709E¯13   23
2.0561410E¯1   7.9380952E¯1   7.9214157E0    1.1566179E1     24
2.7215771E¯1   6.7619048E¯1   1.3216197E1   ¯2.0340746E1     25
3.2237294E¯1   6.4095238E¯1   1.0000000E1    2.2737368E¯13   26
2.7215771E¯1   6.7619048E¯1   6.7838028E0    2.0340746E1     27
2.9620703E¯1   7.7000000E¯1   1.1512059E1   ¯1.1481841E2     28
3.0989855E¯1   6.9428571E¯1   1.0000000E1    3.4106051E¯13   29
2.9620703E¯1   7.7000000E¯1   8.4879406E0    1.1481841E2     30
0.0000000E0    1.0000000E0    1.0000000E1    1.1368684E¯13   31
0.0000000E0    0.0000000E0    0.0000000E0    0.0000000E0     32
```

```
K1←TOT[;1]
K2←TOT[;2]
K3←TOT[;3]
K4←TOT[;4]
K5←TOT[;5]

IND←ORDER K2

K2←K2[IND]
K1←K1[IND]
K3←K3[IND]
K4←K4[IND]
K5←K5[IND]

V2←BOOLVECT K2

L2←LENGTH V2

IND← L2 ORDERNEXT K1

K1←K1[IND]
K3←K3[IND]
K4←K4[IND]
K5←K5[IND]

V1←BOOLVECT K1

L1←LENGTH V1

IND← L1 ORDERNEXT K3

K3←K3[IND]
K4←K4[IND]
K5←K5[IND]

V3←BOOLVECT K3

L3←LENGTH V3

IND← L3 ORDERNEXT K4

K4←K4[IND]
K5←K5[IND]

DISP←K2,K1,K3,K4,K5
```



```
DISP←⍉(K2,[.5]K1)
DISP←DISP,K3
DISP←DISP,K4
DISP←DISP,K5

'4E14.8,I4' ΔFMT DISP

 0.0000000E0    0.0000000E0    0.0000000E0    0.0000000E0     32
 0.0000000E0    1.0000000E0    1.0000000E1    1.1368684E¯13   31
 1.6191073E¯1   1.6047619E¯1   1.4569733E0    1.6758724E0      4
 1.6191073E¯1   1.6047619E¯1   1.8543027E1   ¯1.6758724E0      6
 1.6191073E¯1   8.3952381E¯1   8.3669881E0    7.8058532E0     21
 1.6191073E¯1   8.3952381E¯1   1.1633012E1   ¯7.8058532E0     19
 1.9582174E¯1   1.6190476E¯1   1.0000000E1   ¯5.6843419E¯14    5
 1.9582174E¯1   8.3809524E¯1   1.0000000E1    0.0000000E0     20
 2.0561410E¯1   2.0619048E¯1   1.9976905E0    3.3570716E0      7
 2.0561410E¯1   2.0619048E¯1   1.8002309E1   ¯3.3570716E0      9
 2.0561410E¯1   7.9380952E¯1   7.9214157E0    1.1566179E1     24
 2.0561410E¯1   7.9380952E¯1   1.2078584E1   ¯1.1566179E1     22
 2.4684148E¯1   2.1809524E¯1   1.0000000E1    1.7053026E¯13    8
 2.4684148E¯1   7.8190476E¯1   1.0000000E1    2.8421709E¯13   23
 2.5029317E¯1   2.6190476E¯1   2.6363636E0    5.4806912E0      1
 2.5029317E¯1   2.6190476E¯1   1.7363636E1   ¯5.4806912E0      3
 2.5029317E¯1   7.3809524E¯1   7.3870968E0    1.6304146E1     18
 2.5029317E¯1   7.3809524E¯1   1.2612903E1   ¯1.6304146E1     16
 2.7215771E¯1   3.2380952E¯1   3.2838235E0    6.3427147E0     10
 2.7215771E¯1   3.2380952E¯1   1.6716176E1   ¯6.3427147E0     12
 2.7215771E¯1   6.7619048E¯1   6.7838028E0    2.0340746E1     27
 2.7215771E¯1   6.7619048E¯1   1.3216197E1   ¯2.0340746E1     25
 2.9620703E¯1   2.3000000E¯1   4.9378882E0    1.3920596E0     13
 2.9620703E¯1   2.3000000E¯1   1.5062112E1   ¯1.3920596E0     15
 2.9620703E¯1   7.7000000E¯1   8.4879405E0    1.1481841E2     30
 2.9620703E¯1   7.7000000E¯1   1.1512059E1   ¯1.1481841E2     28
 3.0240237E¯1   2.8571429E¯1   1.0000000E1   ¯1.7053026E¯13    2
 3.0240237E¯1   7.1428571E¯1   1.0000000E1   ¯5.6843419E¯14   17
 3.0989855E¯1   3.0571429E¯1   1.0000000E1   ¯5.6843419E¯14   11
 3.0989855E¯1   6.9428571E¯1   1.0000000E1    5.6843419E¯14   14
 3.2237294E¯1   3.5904762E¯1   1.0000000E1    2.2737368E¯13   26
 3.2237294E¯1   6.4095238E¯1   1.0000000E1    3.4106051E¯13   29
```


29
26
14
11
17
2
28
30
15
13
25
27
12
10
16
18
3
1
23
8
22
24
9
7
20
5
19
21
6
4
31
32