

Reasoning with Non-Numeric Linguistic Variables

Jon Williams, Nigel Steele and Helen Robinson

Computational Intelligence Group, School of Mathematical and Information Sciences, Coventry University

Where decisions are based on imprecise numeric data and linguistic variables, the development of automated decision aids presents particular difficulties. In such applications, linguistic variables often take their values from a pre-ordered set of vaguely defined linguistic terms. The mathematical structures that arise from the assumption that sets of linguistic terms are pair-wise tolerant are considered. A homomorphism between tolerance spaces, filter bases and fuzzy numbers is shown. A proposal for modeling linguistic terms with an ordered set of fuzzy numbers is introduced. A procedure for structured knowledge acquisition based on the topology of the term sets and the cognitive theory of prototypes is shown to give rise to sparse rule bases. Similarity as a function of “distance” between fuzzy numbers treated as tolerance mappings is used as an inference mechanism in sparse rule bases to give linguistically valued outputs. Measuring the “distance” between fuzzy sets to correspond to intuitive notions of nearness is not straightforward, since the usual metric axioms are not adequate. An alternative way of measuring “distance” between fuzzy numbers is introduced, which reduces to the usual one when applied to crisp numbers.

Keywords: linguistic variables, tolerance spaces, fuzzy numbers, similarity based reasoning.

1. Introduction

There is a particular class of problems which have a small set of mutually exclusive decision outcomes where the development of automated decision aids presents particular difficulties. In these domains decisions are based on both imprecise numeric data and linguistic variables which have no underlying numeric scale. (A very simple example is deciding whether a student should pass or fail a course based on their mark and performance in seminars. The mark is imprecise because of marker variation and performance in a seminar takes a linguistic value

such as **competent**). These decisions tend not to have well defined rule sets, making traditional expert systems difficult to develop.

There is a substantial body of work dealing with linguistic variables and their modeling with fuzzy sets. It is Zadeh’s contention that Fuzzy logic = Computing with words [Zadeh 1996]. However, in applications, fuzzy sets and fuzzy logic are most often applied to variables which have an underlying numeric scale. Modeling non numeric linguistic variables is acknowledged to be less well developed. (For example by Cox on who wrote in the `comp.ai.fuzzy` newsgroup on 28 March 1999 “*The application of fuzzy logic and fuzzy metrics to non-numeric objects (and events) has long been a difficult task*”.)

Most work relating fuzzy logic and fuzzy set theory to linguistic variables starts from the premise that Zadeh’s contention is correct. We intend to take a different approach by making assumptions about the mathematical structure of sets of pre-ordered linguistic terms from which a linguistic variable takes values. It is also usual to define linguistic variables and linguistic terms using Zadeh’s definition [Zadeh 1973]. Here we will start with modified definition(s) and show that a fuzzy interpretation is one way of making these initial definitions operational. Fuzzy Numbers are a device which has been used to model non-numeric linguistic variables [Bojadziev 1995, Zimmermann 1990], however Mareš [Mares 1999] draws attention to the fact that operations on fuzzy numbers using the extension principle or interval arithmetic [Klir 1995, Moore 1996, Zimmerman 1987] appear not to give results which mirror the intuitive way humans handle vague quantities. In

this paper, we show how alternative, but essentially equivalent mathematical structures of *filter bases* and *tolerance spaces*, provide a possible solution to this problem. The approach taken here is designed to deal with problems where the only variables are non-numeric and also to combine non-numeric variables with numeric variables. In both cases it is possible to give imprecise linguistic outputs which convey the uncertainty of outcomes to a user in an intuitive way.

Definition 1. (Non-Numeric Linguistic Variable) *A non-numeric linguistic variable is characterised by a quintuple $\langle v, L, X, g, m \rangle$; in which v is the name of the variable; L is a finite set of linguistic terms $\{l_0, \dots, l_n\}$ which describe v whose states range over a universal set X of states of v ; g is a grammar for generating linguistic terms, and m is a semantic rule which assigns to each term $l \in L$ its meaning $m(l)$ on X ; that is $m : L \rightarrow A \subset X$.*

Example 1. Consider the performance of a student in a seminar. Then $v = \mathbf{seminar}$. The set $X_{\mathbf{seminar}}$ of possible states of a student's performance ranges from **no contribution** to **superb**. These states are described by a pre-ordered set $L_{\mathbf{seminar}}$ with a grammar (g) of at least three terms, for example **{poor, competent, good}**. The grammar (g) might also specify how many additional terms may be added. So for example the terms **very poor** and **very good** might be added. (The number of terms used in applications is discussed in section 6.2.). Then $m(\mathbf{poor})$ maps to the lower portion of the seminar performance range so that $m(\mathbf{poor}) \leq m(\mathbf{competent})$.

Definition 2. (Numeric Linguistic Variable) *A numeric linguistic variable is characterised by a quintuple $\langle v, L, X, g, m \rangle$; in which v names the variable; L is a finite set of linguistic terms $\{l_0, \dots, l_n\}$ of the base variable v whose values range over a universal set X of values of v ; g is a grammar for generating linguistic terms; and m is a semantic rule which assigns to each term $l \in L$ its meaning $m(l)$ on X , (i.e. $m(l) : L \rightarrow A \subset X$).*

Example 2. Consider the marks gained by a student on a course. Then $v = \mathbf{mark}$. The set $X_{\mathbf{mark}}$ of possible values of a student's mark

range from 0 to 100. These states are described by a pre-ordered set $L_{\mathbf{mark}}$ with a grammar (g) of at least two terms, for example **{fail, pass}**. Again the grammar might also specify how additional terms were to be added. (So, for example, the term **marginal pass** may be added). Then we could have $m(\mathbf{fail}) = [0, 40)$ and $m(\mathbf{pass}) = [40, 100]$. We might also have $m(\mathbf{marginal pass}) = (35, 45]$ if it is possible to pass with a mark less than 40, but not with a mark less than 35.

2. Sets of Linguistic Terms — Mathematical Structure

In this section assumptions will be made about the mathematical structure of sets of linguistic terms. The following definitions are required.

2.1. Orderings

Definition 3. (Preorders) *Let A be a set, then whenever there is a defined concept of less than or equal to then this establishes a pre-order on A [Nachbin 1965]. More formally, a preordered set is a set A equipped with a reflexive transitive operation \leq such that if $a, b, c \in A$ then $a \leq a$ (reflexivity) and in addition $a \leq b$ and $b \leq c$ implies $a \leq c$ (transitivity).*

Definition 4. (Partial Orders and Chains) *A partial order is a preorder where in addition \leq satisfies $a, b \in A$ then $a \leq b$ and $b \leq a$ implies $a = b$. A partial order relation is linearly ordered or a chain if*

$$\forall a, b \in A \text{ either } a \leq b \text{ or } b \leq a.$$

Definition 5. (Up sets) *Let P be a partially ordered set with $Q \subseteq P$. Then Q is an up set whenever $x \in Q, y \in P$ and $y \geq x$ then $y \in Q$. If x is the least element in Q then the set is often referred to as the up set of x . Down sets are defined by duality. The up set of x is denoted $\uparrow x$ and the down set of x by $\downarrow x$.*

Example 3. The set of integers less than or equal to 10, with the usual order, is a chain. The down set of 3, $\downarrow 3 = \{3, 2, 1\}$; for the same set, $\uparrow 8 = \{8, 9, 10\}$.

Example 4. The set of pairs $\{(m, n) : (m, n) \in \mathbb{N} \times \mathbb{N}\}$ with $(m_1, n_1) \leq (m_2, n_2) \Leftrightarrow m_1 \leq m_2 \ \& \ n_1 \leq n_2$ is a partial order, but not a chain, since $(4, 3)$ is not comparable with $(3, 4)$.

2.2. Vague Linguistic Terms

The remainder of this paper is based on the following assumptions about the sets of vague linguistic terms which are used to evaluate variables in decision making environments. Let a finite set of mutually exclusive decisions $D := d_1, \dots, d_n$ have a finite set $p \in \mathbb{N}$ of attributes described by a finite set of linguistic variables $V := v_1, \dots, v_p$.

Preorder

Each v_k takes a value or range of values from a finite pre-ordered set of linguistic terms value: $V \rightarrow L$ to give value(v_k) = l_i , ($0 \leq i \leq \text{card}(L)$) and the pair $\langle l_i, v_k \rangle$ which we will denote l_{i,v_k} .

Example 5. For the variable **mark** we have the linguistic variables **pass_{mark}** and **fail_{mark}**.

Kernels

Assume for each l_{i,v_k} there is at least one state of v_k to which only that term applies, the *kernel* $\ker(l_{i,v_k})$ of the term, so that

$$m(\ker(l_i)) \cap m(\ker(l_{i+1})) = \emptyset$$

for each $l_i \in L$.

Example 6. For the term set

$$\{\text{poor}, \text{competent}, \text{good}\}_{\text{seminar}}$$

there exist states which are only mapped to one term so that the strict precursor relation

$$\ker(\text{poor}) \prec \ker(\text{competent}) \prec \ker(\text{good})$$

holds and the meanings m (see definition 1) are disjoint

$$m(\ker(\text{competent})) \cap m(\ker(\text{good})) = \emptyset.$$

Pairwise tolerance

We assume that vague linguistic terms are pairwise tolerant so that

$$l_{i,v_k} \cap l_{i+1,v_k} = \tau_R(l_{i,v_k}) = \tau_L(l_{i+1,v_k}) \neq \emptyset$$

the meaning $m : L_{v_k} \rightarrow X$ gives $m(\tau_R(l_i)) = m(\tau_L(l_{i+1}))$. Here τ_R is the *right tolerance* and τ_L the *left tolerance*. The least element of L_{v_k} is assumed to have no left tolerance and the greatest element no right tolerance. Note that $\tau_L(l_i) < \ker(l_i) < \tau_R(l_i)$.

Example 7. The term **poor** \sim **competent** $\implies m(\text{poor}) \cap m(\text{competent}) \neq \emptyset$.

Support

Taking the previous two assumptions together, assume that each linguistic term has as a *support* set the points $\text{supp}(l_i) = \{\tau_L(l_i), \ker(l_i), \tau_R(l_i)\}$ so that $m(\text{supp}(l_i)) = m(l_i)$.

Pairwise concatenation

The operation of concatenation \oplus of the linguistic terms $l_i, l_m \in L_{v_k}$ is defined pairwise. So that

$$l_{i \oplus m} = l_i \oplus \dots \oplus l_m = \bigcup_i^m l_j \quad (i \leq j \leq m)$$

which is the same as

$$l_{i \oplus m} = \{\uparrow l_i \cap \downarrow l_m : l_i < l_m\}.$$

This is equivalent to the logical *OR* ' \vee ' operation. Note that

$$\ker(l_{i \oplus m}) = \{\ker(l_i), \tau_R(l_i), \dots, \tau_L(l_m), \ker(l_m)\}.$$

Example 8. For a variable v_0 described by the term set

$$\{\text{very bad} \lesssim \text{bad} \lesssim \text{medium} \\ \lesssim \text{good} \lesssim \text{very good}\}_{v_0}$$

then

$$\text{very bad}_{v_0} \oplus \text{medium}_{v_0} = \\ \{\text{very bad}, \text{bad}, \text{medium}\}_{v_0}$$

which may be read

$$\{\text{very bad}_{v_0} \vee \text{bad}_{v_0} \vee \text{medium}_{v_0}\}.$$

Products

The product \prod of two sets of linguistic terms is defined on the product space of their respective variables $v_k \times v_j$ to give the pair l_{i,v_k}, l_{i,v_j} and naturally extended to $\prod_1^p v_k$ to give $(l_{i,v_1}, \dots, l_{i,v_p})$.

Example 9. For two variables $v_1 = \mathbf{mark}$ and $v_2 = \mathbf{seminar}$ we have the product space $\mathbf{mark} \times \mathbf{seminar}$. A range of numerical values [35, 45] of \mathbf{mark} may be described by the term **marginal pass**, and range of states of $\mathbf{seminar}$ by the term **competent**. Then the pair of terms ($\mathbf{marginal\ pass}_{\mathbf{mark}}, \mathbf{competent}_{\mathbf{seminar}}$), can be read

{ $\mathbf{marginal\ pass}_{\mathbf{mark}}$ and $\mathbf{competent}_{\mathbf{seminar}}$ }

3. Mathematical Structures and Linguistic Variables

Albrecht [Albrecht 1998] draws attention to the fact that many elements of knowledge-based systems and uncertain reasoning can be captured by traditional mathematics and, in particular, by topology. Hoveseppian [Hoveseppian 1992] shows that a tolerance space is sufficient to model the sorites paradox using a three valued logic and Stout [Stout 1992] considers the relationship between tolerance spaces and fuzzy sets in a category theoretic setting.

Definition 6. (Topological Space – Open Sets)

A topology on set X is a collection \mathcal{T} of subsets of X with the following properties:

- O1** $X \in \mathcal{T}$ and $\emptyset \in \mathcal{T}$.
- O2** The union of any number of sets in \mathcal{T} belongs to \mathcal{T} .
- O3** The intersection of any two sets in \mathcal{T} belongs to \mathcal{T} .

The pair (X, \mathcal{T}) is called a topological space.

Definition 7. (Open sets) Members of \mathcal{T} are called open sets. The complement of an open set is closed. An alternative, equivalent, definition of a topology may be phrased in terms of closed sets [Kelley 1955].

Definition 8. (Bases) Let (X, \mathcal{T}) be a topological space. A base for \mathcal{T} is a collection $\mathcal{B} \subset \mathcal{T}$ such that every set in \mathcal{T} is a union of sets from \mathcal{B} .

Definition 9. (Sub-bases) A sub-base for a topology \mathcal{T} is a collection $\mathcal{S} \subset \mathcal{T}$ such that any set in \mathcal{T} is a union of finite intersections of sets from \mathcal{S} .

It will be shown later that if the supports of linguistic terms (i.e. $\{\tau_L(l_i), \ker(l_i), \tau_R(l_i)\}$) are taken as a sub-base for a topology, then the kernels $\ker(l_i)$ are closed.

Different topologies have different separation properties [Kelley 1955] described by the T_i axioms two of which we will require.

Definition 10. A topology on (X, \mathcal{T}) is:

T_2 or Hausdorff If and only if, given a pair of distinct points x, y of X there exist open sets \mathcal{O} and \mathcal{O}' with $x \in \mathcal{O}$ and $y \in \mathcal{O}'$, such that $\mathcal{O} \cap \mathcal{O}' = \emptyset$.

T_0 If for each distinct x and y in X , there exists an open $\mathcal{O} \in \mathcal{T}$ such that either $x \in \mathcal{O}$ and $y \notin \mathcal{O}$, or there exists $\mathcal{O}' \in \mathcal{T}$, such that $y \in \mathcal{O}'$ and $x \notin \mathcal{O}'$.

Equivalently a topology is T_0 if and only if, given a pair of distinct points x, y of X there exists an open set \mathcal{O} , such that $\mathcal{O} \cap \{x, y\}$ is a singleton.

Definition 11. (Tolerance Relation) A tolerance relation is a symmetric reflexive relation $\xi \subset X \times X$, such that $x\xi x$ for every $x \in X$ and $x\xi y \implies y\xi x$ for all $x, y \in X$.

Definition 12. (Tolerance Space – [Zeeman 1962]) A tolerance space, $\langle X, \xi \rangle$ is a space X with a tolerance relation $\xi \in X \times X$, the tolerance on X . If $\langle x, y \rangle \in \xi$ denoted $x\xi y$, then x is said to be within ξ tolerance of y . A space may be equipped with more than one tolerance. A bi-tolerant space is a space $\langle X, \langle \xi, \zeta \rangle \rangle$ with $\xi \supseteq \zeta \supseteq \delta$, where δ is the discrete tolerance and $x\delta y \iff x = y$. In a tolerance space $\langle X, \xi \rangle$ the set $N(x)\{y:y\xi x\}$, is called a t -neighborhood of $x \in X$. The t -neighborhood of a subset A of X is the union of the t -neighborhoods of all its points, that is

$$N(A) = \bigcup_{a \in A} N(a).$$

Example 10. Take $r_1, r_2 \in \mathbb{R}$ and let $r_1 \xi r_2 \iff |r_1 - r_2| < 1$; then $N(r_1) = (r_1 - 1, r_1 + 1)$. Now take $[1, 2] \subset \mathbb{R}$; then $N([1, 2]) = (0, 3)$.

A topological space (X, \mathcal{T}) is *tolerable* [Hovesepian 1992] if there exists a ξ on X , such that the set $\{N(x) : x \in X\}$ of t-neighborhoods serves as a sub-basis for a topology \mathcal{T} on X .

Example 11. Consider the relation described in example 10; the set $\{N(r) : r \in \mathbb{R}\}$ is a sub-basis for the usual topology.

Example 12. Consider the set of neighborhoods $\{N(1), N(2)\}$ with the tolerance relation defined in example 10. Then they generate a topology on $N([1, 2])$ given by

$$\{(0, 3), (0, 2), (1, 3), (1, 2), \emptyset\}.$$

Definition 13. (Tolerance Mapping) Let $f : \langle X, \xi \rangle \rightarrow \langle Y, \zeta \rangle$ be a map between two tolerance spaces then f is a tolerance map, if

$$x \xi y \in \langle X, \xi \rangle \implies f(x) \zeta f(y) \in \langle Y, \zeta \rangle.$$

If the converse holds, then f is a tolerance embedding.

3.1. Topologies of Linguistic Terms

Different approaches can be taken to topologies of the term set of a linguistic variable. The supports of the linguistic terms can be taken as a sub-base for a topology. On the other hand, the kernels and tolerances may be taken as a base for a topology which gives a different topology with different, but related separation properties. It is also possible to define a topology by taking the (concatenated) kernels alone as a basis for the closed sets. The difference in the nature of these spaces is important; since empirical evidence [Zwicky 1988, Yoshikawa 1996] suggests that people judge the similarity of linguistic terms on the basis of the separation of the kernels of the terms.

Theorem 1. Let $S = \bigcup_{i \in I} \text{supp}(l_i)$ be the set of tolerances and kernels for the k^{th} linguistic variable. Then the topology \mathcal{L}_S on S obtained by taking $\{\text{supp}(l_i)\}_{i \in I}$ as a sub-base is T_0 but not T_2 .

Proof. Sets of the form $\{\tau_L(l_i), \dots, \tau_R(l_i)\}$ are open in (S, \mathcal{L}_S) .

Every open set which contains $\ker(l_i)$ also contains $\tau_L(l_i) = \tau_R(l_{i-1}) \in l_{i-1}$ hence (S, \mathcal{L}_S) is not T_2 .

Take $\ker(l_i), \ker(l_j)$ or $\tau(l_i), \tau(l_j)$ as distinct points, then there are open sets l_i and $\tau(l_i)$ such that $l_i \cap \{\ker(l_i), \ker(l_j)\}$ and $\tau(l_i) \cap \{\tau(l_i), \tau(l_j)\}$ are singletons. Taking $\ker(l_i), \tau(l_k)$ then $\tau(l_k) \cap \{\ker(l_j), \tau(l_k)\} = \tau(l_k)$ is a singleton and (S, \mathcal{L}_S) is T_0 . \square

Theorem 2. Let $S = \bigcup_{i \in I} \text{supp}(l_i)$. Then the topology (S, \mathcal{L}_k) obtained by taking the concatenated kernels $K = \{\ker(l_{i \oplus j}) : i \in I\}$ as a base is T_0 but not T_2 .

Proof. $\ker(l_i), \tau_R(l_i)$ are distinct points not in disjoint sets since $\tau_R(l_i) \in \ker(l_{i \oplus i+1})$ and $\ker(l_i) \cap \ker(l_{i \oplus i+1}) \neq \emptyset$ hence (S, \mathcal{L}_k) is not T_2 . Every member of \mathcal{L}_k which contains $\ker(l_i)$ also contains $\tau_L(l_i)$ or $\tau_R(l_{i+1})$ hence (S, \mathcal{L}_k) is T_0 . \square

Theorem 3. Take S as in theorems 1 and $\mathcal{L}_p = \{\{\tau_L(l_i)\}, \{\ker(l_i)\}, \{\tau_R(l_i)\} : i \in I\}$ as a basis for a topology on S then (S, \mathcal{L}_p) is Hausdorff.

Proof. The $\tau(l_i)$ are open and disjoint, both from each other and the $\ker(l_i)$. Similarly the $\ker(l_i)$ are closed and disjoint, both from each other and the $\tau(l_i)$. \square

The topology (S, \mathcal{L}_p) is also T_0 since all T_2 spaces are T_0 [Kelley 1995], however the converse does not hold.

It follows that the topology on the product space $(\mathfrak{S}, \mathfrak{L}_S)$ obtained by taking the partially ordered set

$$\mathfrak{L}_S = \prod_{k=1}^n L_{v_k} = \prod_{k=1}^n \{l_{i, v_k}\}$$

as a sub-base is not Hausdorff. However the topology given by $(\mathfrak{S}, \mathfrak{L}_p)$ obtained by taking as a basis the partially ordered set

$$\mathfrak{L}_p = \prod_{k=1}^n P_{v_k} = \prod_{k=1}^n \{p_{i, v_k}\}$$

where p_i is the i^{th} element in the chain

$$\ker(l_0) < \tau_R(l_0) < \dots < \tau_R(l_{n-1}) < \ker(l_n)$$

is Hausdorff. Since all metric spaces are Hausdorff these results suggest that a metric space may be an inappropriate model for the term sets of linguistic variables. There is, however, a continuous mapping $\phi : (\mathfrak{S}, \mathfrak{L}_p) \rightarrow (\mathfrak{S}, \mathfrak{L}_s)$ between the two spaces such that

$$\phi(\ker(l_{i+j,v_1}) \times \dots \times \ker(l_{i+j,v_k})) \mapsto (l_{i+j,v_1} \times \dots \times l_{i+j,v_k}), \quad (0 \leq i, j \leq |L| - 1)$$

which means that it is possible to work in the Hausdorff space $(\mathfrak{S}, \mathfrak{L}_p)$ and map into the T_0 space $(\mathfrak{S}, \mathfrak{L}_s)$ or to work with the bi-topological space $(\mathfrak{S}, \langle \mathfrak{L}_s, \mathfrak{L}_p \rangle)$ and so move between the strictly T_0 spaces of the linguistic terms and the T_2 space of the disjoint kernels and tolerances.

3.2. Filters and Filter Bases

It is convenient to introduce here the concepts of *Filters* and *Filter Bases*.

Definition 14. (Filter Base) A non empty system \mathcal{B} of subsets of the set X is called a filter basis on X if the following conditions are satisfied.

FB1 the intersection of any two sets from \mathcal{B} contains a set from \mathcal{B} .

FB2 the empty set does not belong to \mathcal{B} .

Example 13. For any $a > 0, x \in \mathbb{R}$; $\mathcal{B} = \{\{x\}, [x - a, x + a]\}$ is a filter base.

Given this definition, we prove the following theorem

Theorem 4. Let $l = \{l_i, \ker(l_i)\}$ with $l_i = \{\tau_L(l_i), \ker(l_i), \tau_L(l_{i+1})\}$ then l is a filter-base.

Proof. Both FB1 and FB2 are satisfied since: $l_i \cap \ker(l_i) = \ker(l_i) \in l$ and $\emptyset \notin l$. \square

If we add to every set of a filter basis all of its supersets, then FB1 is sharpened and we get a *filter*.

Definition 15. (Filter) A non empty system \mathcal{F} of subsets of the set X is called a filter (on X) if the following conditions are satisfied.

F1 Every superset of a set from \mathcal{F} belongs to \mathcal{F} .

F2 The intersection of any two sets from \mathcal{F} belongs to \mathcal{F} .

F3 The empty set does not belong to \mathcal{F} .

Note that while every filter is a filter base the converse is not true.

Example 14. For any $a_i, b_i, x \in \mathbb{R}, i \in \mathbb{N}$ and $a_i < b_i$ with $a_1 > a_2 > \dots > a_n$ and $b_1 > b_2 > \dots > b_n$ then $\mathcal{B} = \{\{x\}, [x - a_i, x + b_i]\}$ is a filter.

Any unconcatenated linguistic term l_i can be transformed into a filter by introducing the nested sets $\tau_L(l_i) \supseteq \tau_{L_j}(l_i) \supset \dots \supset \tau_{L_{j'}}(l_i)$ with $\tau_{L_{j'}}(l_i) < \ker(l_i)$ and similarly for $\{\tau_{R_j}\} > \ker(l_i)$ $j \in J$, an indexing set. Then we can prove

Theorem 5. If $l' = \{l_i, \ker(l_i)\}$ with $l_i = \{\{\tau_{L_j}(l_i)\}, \ker(l_i), \{\tau_{R_j}(l_i)\} : i \in [0, |L| - 1] j \in J(\text{an indexing set})\}$ then l' is a filter.

Proof. F1, F2 and F3 are satisfied since:

$$\begin{aligned} & \{\tau_{L_j}(l_i), \ker(l_i), \tau_{R_j}(l_i)\} \\ & \supset \{\tau_{L_{j'}}(l_i), \ker(l_i), \tau_{R_{j'}}(l_i)\} \\ & \supset \ker(l_i) \end{aligned}$$

also

$$\begin{aligned} & \{\tau_{L_j}(l_i), \ker(l_i), \tau_{R_j}(l_i)\} \\ & \cap \{\tau_{L_{j'}}(l_i), \ker(l_i), \tau_{R_{j'}}(l_i)\} \\ & = \{\tau_{L_{j'}}(l_i), \ker(l_i), \tau_{R_{j'}}(l_i)\} \in l' \end{aligned}$$

and

$$\ker(l_i) \cap \{\tau_{L_j}(l_i), \ker(l_i), \tau_{R_j}(l_i)\} = \ker(l_i) \in l'$$

and finally $\emptyset \notin l'$. \square

Theorem 6. Let $S = \bigcup_{i \in I} \text{supp}(l_i)$ then the topology (S, \mathcal{L}_f) obtained by taking

$$F = \{ \{ \tau_{L_j}(l_i), \ker(l_i), \tau_{R_j}(l_i) : j \in J \} : i \in I \}$$

as a sub-base is T_0 but not T_2 .

Proof. Take $p \in \tau_{R_j}(l_i)$ then there is no $j \in J$ such that p and $\ker(l_i)$ are in disjoint open sets; hence (S, \mathcal{L}_f) is not T_2 .

Consider $W_1 = \{p_1, p_2 : p_1 \neq p_2\}$ such that $p_1, p_2 \in \tau_L(l_i)$ then $W_1 \cap \tau_{L_j}(l_i)$ is a singleton for some $j \in J$, a similar argument applies to $p_1, p_2 \in \tau_R(l_i)$. Next consider $W_2 = \{p_1, \ker(l_i) : p_1 \neq \ker(l_i)\}$ such that $p_1 \in \tau_L(l_i)$ or $p_1 \in \tau_R(l_i)$, then $W_2 \cap \tau_L(l_i)$ or $W_2 \cap \tau_R(l_i)$ are a singletons. Now take $W_3 = \{p_1, p_2 : p_1 \neq p_2\}$ such that $p_1 \in \tau_L(l_i)$ and $p_2 \in \tau_R(l_i)$ then $W_3 \cap \tau_L(l_i)$ is a singleton. Finally take $W_4 = \{\ker(l_i), \ker(l_{i'})\}$ then $W_4 \cap \text{supp}(l_i)$ is a singleton. Hence given any pair of distinct points p_1, p_2 of S there exists an open set \mathcal{O} , such that $\mathcal{O} \cap \{p_1, p_2\}$ is a singleton and (S, \mathcal{L}_f) is T_0 . \square

4. Filter Bases and Fuzzy Sets

In this section we show how filter bases are related to fuzzy sets. Before doing so, two preliminary definitions are needed.

Definition 16. (Proper Filter Base) A filter base is proper if $\bigcap \mathcal{B} \neq \emptyset$.

Definition 17. Let $f : X \rightarrow Y$ then f is a \leq -homomorphism if $\forall a, b \in X$ $a \leq b$ implies $f(a) \leq f(b)$. If f preserves meets and joins so that $f(a \vee b) = f(a) \vee f(b)$ and $f(a \wedge b) = f(a) \wedge f(b)$ then f is a lattice homomorphism [Davey and Priestley 1990].

Albrecht [Albrecht 1998] shows that given a proper filter base such that $\mathcal{B} = \{B_j : j \in J\}$ on $\langle \mathcal{P}(B), \subseteq, \cup, \cap \rangle$ a non-empty lattice $\langle C, \leq, \sqcap, \sqcup \rangle$ and a \leq -homomorphism $\phi : \mathcal{P}(B) \rightarrow C$ with $c_j = \phi(B_j)$ then $\{c_j\} = \phi(\mathcal{B})$ is a filter base on C if all $c_j \neq \emptyset$ and $\phi(\lim \mathcal{B}) \leq \lim \phi(\mathcal{B})$. Where $\lim \mathcal{B} = \bigcap_{j \in J} B_j$ and $\lim \phi(\mathcal{B}) = \bigcap_{j \in J} \phi(B_j)$.

Now consider $\phi^{-1} : C \rightarrow \mathcal{P}(B)$ defined for all $c \in C$ by $\phi^{-1}(c) := \bigcup_{\phi(U)=c} U$. Then ϕ^{-1} is a lattice homomorphism [Bourbaki 1967]. (Requiring \mathcal{B} to be a proper filter base ensures ϕ^{-1} is well defined since $\phi^{-1}(c) \neq \emptyset, \forall c \in C$.)

If $\mathcal{C} = \{c_l : l \in L\}$ is a filter base on C and for all $c \in \mathcal{C}$, $\phi^{-1}(c) \neq \emptyset$ then $\phi^{-1}(\mathcal{C})$ is a filter base on $\mathcal{P}(B)$. In the function $(B_j, c_j)_{j \in J}$ the elements B_j of a filter base \mathcal{B} with support $B = \bigcup \mathcal{B}$ are evaluated by c_j with $c_j = \phi(B_j)$.

Definition 18. (Fuzzy Set) Zadeh defines [Zadeh 1965] a fuzzy subset A of a set X as a nonempty subset

$$A = \{(x, \mu_A(x)) : x \in X\} \text{ of } X \times [0, 1]$$

for some function $\mu_A : X \rightarrow [0, 1]$.

The common practice of referring to fuzzy subsets as fuzzy sets will be adopted from now on. The alpha-level set of a fuzzy set A_α on X is defined as

$$A_\alpha = \{x \in X : \mu_A(x) \geq \alpha \text{ for each } \alpha \in (0, 1]\}$$

The support of a fuzzy set is given by

$$A_{0+} = \{x \in X : \mu_A(x) > 0\}$$

and the kernel of a fuzzy set by

$$A_1 = \{x \in X : \mu_A(x) = 1\}$$

Fuzzy subsets of \mathbb{R} are often referred to as fuzzy numbers or fuzzy intervals and are defined as follows:

Definition 19. [Goetschel 1983] A fuzzy number or fuzzy interval is a fuzzy set with membership function

$$\mu : \mathbb{R} \rightarrow I = [0, 1] \text{ with the properties:}$$

1. μ is upper semi-continuous.
2. $\mu(x) = 0$ outside some interval $(a, d) \in \mathbb{R}$.
3. there are b and $c \in \mathbb{R}$ with $a \leq b \leq c \leq d$, such that μ is increasing on $(a, b]$; decreasing on $[c, d)$ and $\mu(x) = 1 \forall x \in [b, c]$.

Two representations which are particularly computationally efficient are triangular fuzzy numbers and trapezoidal fuzzy intervals defined as follows:

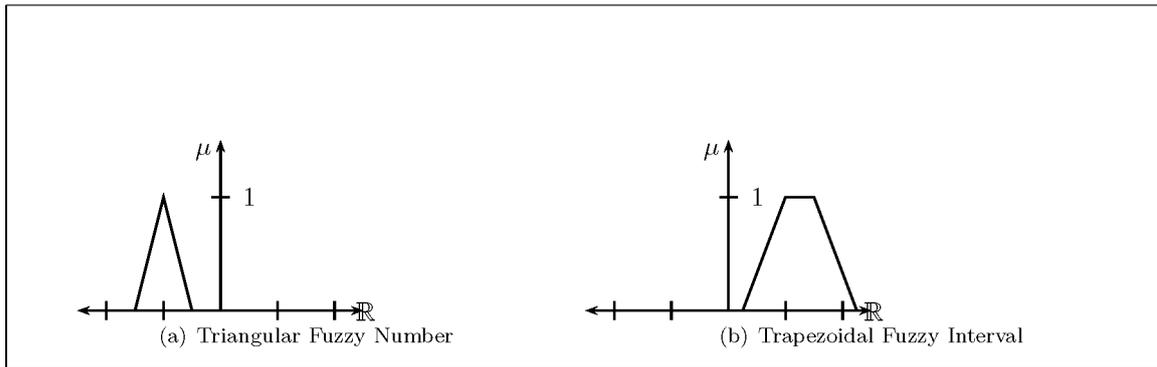


Fig. 1. Fuzzy Numbers.

Definition 20. A Triangular Fuzzy Number (Figure 1(a)) \tilde{n} is characterised by an ordered triple $\langle a_1, a_2, a_3 \rangle$ with $a_1 \leq a_2 \leq a_3$ such that $\tilde{n}_{0+} = (a_1, a_3)$ and $\tilde{n}_1 = \{a_2\}$ or for symmetrical triangular fuzzy numbers as the tolerance space $\langle a_2, \langle \pm \frac{(a_3 - a_1)}{2} \rangle \rangle$.

Definition 21. A Trapezoidal Fuzzy Interval \tilde{I} is characterised by an ordered quadruple $\langle a_1, a_2, a_3, a_4 \rangle$ with $a_1 \leq a_2 \leq a_3 \leq a_4$ such that $\tilde{I}_{0+} = (a_1, a_4)$ and $\tilde{I}_1 = [a_2, a_3]$. For symmetrical trapezoidal fuzzy intervals they can also be represented as the bi-tolerance space $\langle \frac{(a_2 - a_1)}{2}, \langle \pm \frac{(a_4 - a_1)}{2}, \pm \frac{(a_3 - a_2)}{2} \rangle \rangle$. The membership function describes a trapezoid (Figure 1(b)) since the boundaries of the α -level sets

$$\tilde{I}_\alpha = [a_2 - (1 - \alpha)(a_2 - a_1), a_3 + (1 - \alpha)(a_4 - a_3)]$$

are the straight lines joining successive members of the quadruple. A triangular fuzzy number is a special case of a trapezoidal fuzzy interval when $a_2 = a_3$. Note that allowing the membership function to be semi-continuous permits $a_1 = a_2 = a_3$ for Triangular Fuzzy Numbers and $a_1 = a_2 = a_3 = a_4$ for Trapezoidal Fuzzy Intervals.

Having introduced the concept of fuzzy sets, we can now show how they are related to filter bases.

Theorem 7. ([Albrecht 1998]) A fuzzy set, fuzzy interval or fuzzy number is a valuated filter base.

Proof. Take $B = [a, b] \subset \mathbb{R}$, with $a < b$, and $C = (0, 1] \subset \mathbb{R}$ and a function $\mu : B \rightarrow C$ with

the requirement that $\exists x \in B : \mu(x) = 1$. Then define a filter base $\mathcal{C} = \{[\alpha, 1] : \alpha \in (0, 1]\}$ and $B_\alpha := \mu^{-1}([\alpha, 1])$ as an alpha-cut. Then the filter base $\mathcal{B} = \{B_\alpha : \alpha \in (0, 1]\}$ describes a fuzzy set, B with membership function μ . Since fuzzy intervals and fuzzy numbers are fuzzy sets this result applies to them equally. \square

The support (A_{0+}) and kernel (A_1) of a trapezoidal fuzzy interval are a trivial filter base since $A_1 \cap A_{0+} = A_1$ and neither A_1 nor A_{0+} is empty.

Theorem 8. A fuzzy set is a sequence of valuated neighborhoods in a sequence of tolerance spaces.

Proof. Let $\langle X, \xi \rangle$ be a tolerance space, then a tolerance ζ is said to be finer than ξ if $\xi \supset \zeta$, the finest tolerance is the discrete tolerance δ and the space $\langle X, \delta \rangle$ is the discrete tolerance space where every point is only in tolerance of itself. A neighborhood of $a \in X$, $N(a) \supset N'(a)$ if $N(a) \in \langle X, \xi \rangle$ and $N'(a) \in \langle X, \xi' \rangle$ with $\xi \supset \xi'$. So $\{N_i(a)\} := \{N_i(a) \in \langle X, \xi_i \rangle : \xi_i \supset \dots \supset \xi_n \supset \delta\}$ is a filter base and the result follows from theorem 7. \square

Taking these two results together with theorem 5 gives the following theorem.

Theorem 9. If $l = \{l_j, \ker(l_j)\}$ with $l_j = \{\{\tau_{L_i}(l_j)\}, \ker(l_j), \{\tau_{R_i}(l_j)\}\}$ then l can be represented by a fuzzy set with $\mu_l(\ker(l_j)) = 1$ and $0 < \mu_l(\{\tau_{L_i}(l_j)\}), \mu_l(\{\tau_{R_i}(l_j)\}) \leq 1$.

Proof. Apply theorems 7 and 8. \square

In section 5 a way of representing a non numeric linguistic term with a fuzzy number or fuzzy interval is discussed in detail. The following traditional definition of linguistic variable [Zadeh 1973] is now justified.

Definition 22. (Fuzzy Linguistic Variable) A fuzzy linguistic variable is characterised by a quintuple $\langle v, L, X, g, m \rangle$; in which v is the name of the variable, L is a finite set of linguistic terms $\{l_0, \dots, l_n\}$ which describe v whose states range over a universal set X of states of v ; g is a grammar for generating linguistic terms, and m is a semantic rule which assigns to each term $l \in L$ its meaning $m(l)$ which is a fuzzy set on X ; (i.e. $m(l) : L \rightarrow \mathcal{F}X$).

An example of student marks shows how a tolerance mapping may be a better model than a fuzzy one based on the extension principle [Zimmerman 1987] or interval arithmetic [Moore 1966].

Example 15. Let q be a mark in $[0, 100] \in \mathbb{Q}$ and let there be a variation of ± 5 in the marks awarded because of marker variation. Then a mark q can be described as a tolerance mapping to the space $\langle [0, 100], \pm 5 \rangle$ or as a triangular fuzzy number $\tilde{q} = \langle q - 5, q, q + 5 \rangle$.

If we wanted to see the effect of adding an additional marks 10 to a student's score of 30 then adding the marks $30 + 10$ gives 40 ± 5 . This is a mapping from $\langle [0, 100], \pm 5 \rangle \times \langle [0, 100], \pm 5 \rangle \rightarrow \langle [0, 100], \pm 5 \rangle$ and is equivalent to a fuzzy number $\langle 35, 40, 45 \rangle$.

However $30 + 10 = \langle 40 - 10, 40, 40 + 10 \rangle = \langle 30, 40, 50 \rangle$ mapping to the triangular fuzzy numbers using the extension principle.

Since the tolerance is known for any mark there is no reason to suppose that it will double if additional marks are added. This example shows how the tolerance model may capture the intuitive notions better than the fuzzy one based on the extension principle.

5. Numeric Representations of Non-Numeric Linguistic Variables

In applications we need to represent non-numeric linguistic terms for which there is no underlying

numeric scale other than an ordinal one. This section looks at a principled way of arriving at that representation.

Shepard [Shepard 1987] and Nofosky [Nofosky 1984] suggest that the similarity η of two non-identical point stimuli x and y is given by a Universal Law of Generalization.

$$\eta(x, y) = e^{-\alpha d(x, y)} \quad (1)$$

where $d(x, y)$ is a the Hamming or Euclidean metric and α is a constant.

Now fix a finite set $P = \{p \in \mathbb{N}\}$ of points in the state space of a non-numeric linguistic variable. The probability that a state $y \in [0, |P|]$, will be generalised to $a \in P$ can be found as follows.

The similarity η of any point y to a fixed point $a \in P$ is given by $\eta(a, y) = e^{-\alpha d(a, y)}$. The probability that a point $y \in [0, p]$ is generalised to $a \in p$ is given by

$$\mathbf{P}(a|y) = \frac{\eta(a, y)}{\sum_{n=1}^p \eta(n, y)} \quad (2)$$

If $\eta(k, x) = 0$ for $k \in P, a - 1 \leq k \leq a + 1$ with $b = a - 1$ and $c = a + 1$, then the probability that a point $x \in [b, c]$ will generalise to a is given by

$$\mathbf{P}(a|x) = \frac{\eta(a, x)}{\eta(b, x) + \eta(a, x) + \eta(c, x)} \quad (3)$$

for $x \in [b, c]$

For example if $d(x, y) = 1$ and $\alpha = 7$ then $\eta(x, y) = e^{-\alpha d(x, y)} \approx 0$ as can be seen in Figure 2.

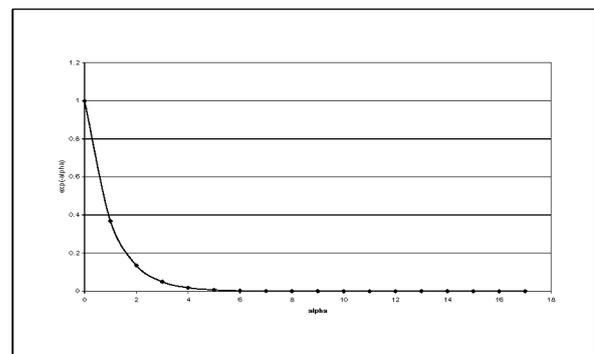


Fig. 2. $e^{-\alpha}$ for $\alpha \in [0, 17]$.

Define

$$\eta'(x, y) = \begin{cases} e^{-7d(x,y)} & d(x, y) < 1 \\ 0 & d(x, y) \geq 1. \end{cases} \quad (4)$$

Then by (3) the probability \mathbf{P} that a point $y \in \mathbb{R}$ generalises to the ordinal 1 is approximated by:

$$\mathbf{P}(1|y) = \frac{\eta'(1, y)}{\eta'(0, y) + \eta'(1, y) + \eta'(2, y)}$$

for $y \in [0, 2]$.

Theorem 7 can be applied by defining tolerances ξ and ζ such that $1 \xi a$ if $\mathbf{P}(1|a) \leq 0.95$ and $1 \zeta b$ if $\mathbf{P}(1|b) \geq 0.05$ then $1 \xi a$ if $|a - 1| \lesssim 0.25$ and $1 \zeta b$ if $|b - 1| \lesssim 0.75$. Hence the set of neighborhoods $\{N(1), N'(1)\}$ with $N(1) \in \xi$ and $N'(1) \in \zeta$ is a filter base and the distribution can be approximated by a fuzzy set with $\mu_1(a) = 1$ for $a \in 1 \pm 0.25$ and $\mu_1(a) = 0$ for $a \notin 1 \pm 0.75$ to give the trapezoidal fuzzy interval $\langle 0.25, 0.75, 1.25, 1.75 \rangle$ as illustrated in Figure 3. This process can be generalised to any $i \in \mathbb{N}$, $0 \leq i \leq |L|$ so that l_i is represented by the fuzzy interval $\langle i - 0.75, i - 0.25, i + 0.25, i + 0.75 \rangle$

Linguistic terms modeled this way fulfill the assumptions in section 2.2. In the present section a fuzzy integer is denoted \tilde{l} .

Preorder

The set of fuzzy integers is pre-ordered with $\tilde{0} < \dots < \tilde{n}$ mirroring the natural order on the integers which is a chain.

Kernels

For each \tilde{l} there is a value $i \in \mathbb{N}$ to which only that term applies. Similarly for each $\tilde{i} \oplus \tilde{j}$ there is an interval $[i, j]$ to which only that term applies.

Pairwise tolerance

Adjacent fuzzy integers $\tilde{l}, \tilde{l} + 1$ are pairwise tolerant since $\tilde{l} \cap \tilde{l} + 1 \neq \emptyset$.

Pairwise concatenation

The operation of join \oplus on adjacent fuzzy integers is defined pairwise to give fuzzy intervals. So that $\tilde{a} \oplus \tilde{a} + i$ is represented by the fuzzy interval

$$\langle a - 0.75, a - 0.25, (a+i) + 0.25, (a+i) + 0.75 \rangle.$$

Support

The support of a fuzzy integer \tilde{a} is given by $\text{supp}(\tilde{a}) = [a - 0.75, a + 0.75]$ and for a fuzzy interval $\tilde{a} \oplus \tilde{a} + i$ by $[a - 0.75, a + i + 0.75]$.

Products

The product \times of two sets of fuzzy integers terms is defined on their product space to give the pair $(\tilde{l}_1, \tilde{l}_2)$ and naturally extended to \prod_1^p to give $(\tilde{l}_1, \dots, \tilde{l}_p)$. The product space is partially ordered.

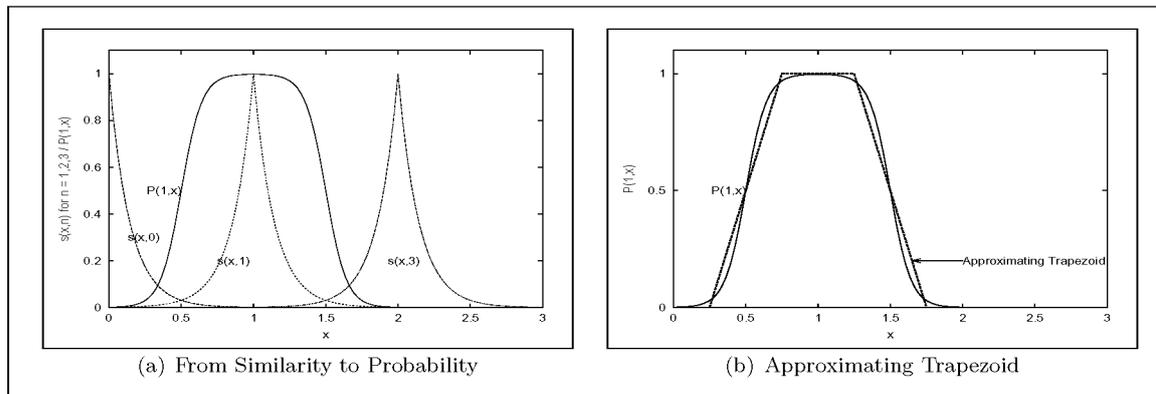


Fig. 3. Modeling Non Numeric Linguistic Variables.

It has been shown that, given a reasonable set of assumptions about the structure of vague non-numeric linguistic variables, they can be modeled with fuzzy sets, filter bases or tolerance spaces. We have also shown an equivalence between these representations.

5.1. Modeling Numeric Linguistic Variables

The modeling of numeric linguistic variables is well established and they can be acquired using parametric method outlined by Kuz'min [Kuz'min 1981] or outlined in any of the standard texts [Klir 1995, Zimmermann 1990]. The techniques outlined above can also be adapted and applied in a numeric context.

5.2. Weighted Linguistic Variables

In applications numeric representations of linguistic variables are usually normalised since there is no common measurement scale between variables [Koczy 1993]; there is also often a measure of relative importance attached to different linguistic variables. Relative importance can be handled [Esboque 1980, Esboque 1983, Zimmermann 1990, Yager 1978] by applying a weighting to each variable, using extended scalar multiplication. The relative weight of variables can be obtained by techniques such as Saaty's [Saaty 1980] Analytic Hierarchy Process. It is possible to use this process to produce fuzzy weights [Laarhoven 1983], however, in this paper only crisp weights will be used.

The tolerance spaces used in this work are all bounded and normalised, in addition all weightings found by the AHP are in the interval $[0, 1]$ hence applying a weighting is equivalent to applying a scaling factor to the whole space. Hence, $w \in [0, 1]$ can be applied to a tolerance space using a scaling $(*)$ function

$$\begin{aligned} * : [0, 1] \times \langle [0, 1], \xi \rangle &\rightarrow \langle [0, w], w * \xi \rangle \\ \text{by } * : (w, x) &\mapsto wx \end{aligned}$$

this is equivalent to extended scalar multiplication [Klir 1995] of a fuzzy number or fuzzy interval where $\tilde{\mathbb{R}}_{\square}$ denotes the set of trapezoidal fuzzy intervals.

$$\begin{aligned} * : \mathbb{R} \times \tilde{\mathbb{R}}_{\square} &\rightarrow \tilde{\mathbb{R}}_{\square} \\ * : (w, \tilde{x}) &= \langle w * x_1, w * x_2, w * x_3, w * x_4 \rangle \end{aligned}$$

where \tilde{x} is represented by the trapezoidal fuzzy set $\langle x_1, x_2, x_3, x_4 \rangle$.

6. Sparse Rule Bases

In decision support the aim is to associate an input case c_m taking a set of values $(\{l_{i,v_k}\}, c_m)$ which is an (intent, extent) pair (see section 6.1.) with a particular d_j with values $(\{l_{i,v_k}\}, d_j)$. If $\text{intent}(\{l_{i,v_k}\}, c_m) \cap \text{intent}(\{l_{i,v_k}\}, d_j) = \text{intent}(\{l_{i,v_k}\}, c_m)$ then $c_m \rightarrow d_j$. If all possible values taken by a variable can be assigned to a decision type so that:

$$\bigcup_{j=1}^n (\{l_{i,v_k}\}, d_j) = (L_V, D) \quad (5)$$

and the same values are not taken by different decision types that is:

$$\bigcap_{j=1}^n \text{intent}(\{l_{i,v_k}\}, d_j) = \emptyset \quad (6)$$

the set valued function $\text{Dec} : (L_V, D) \rightarrow D$, $\text{Dec}(\{\text{intent}(l_{i,v_k}, d_j)\}) \mapsto d_j$ is a bijection and the "rule base" is complete. If

$$\bigcup_{j=1}^n (\{l_{i,v_k}\}, d_j) \neq (L_V, D)$$

and still 6 holds then Dec is not a bijection and the rule base is incomplete. The smaller the number of values for which exact partitions exists the more sparse the rule base.

In the absence of a complete rule base Albrecht [Albrecht 1998] suggests that uniform topologies may be used to find partial or incomplete mappings between inputs and outputs of a rule base. Uniform topologies are a generalisation of metric spaces and generalise the notion of distance between objects. A pre-ordered set has a uniform topology [Page 1978] and it is possible therefore to use generalised notions of distance within such sets.

Before considering the nature of distance measures in more detail, an approach to structured knowledge acquisition for sparse rule bases will be introduced, based on the cognitive theory of prototypes [Rosch 1988] and the topology of the support sets of linguistic terms (S, \mathcal{L}) introduced in section 3.1.

6.1. Prototype Theory

Prototype theory [Rosch 1988] seeks to model aspects of human cognition, according to Hampton [Hampton 1993] the standard prototype model

... assumes that concepts are defined by a set of intensional [sic] properties which determine the reference of the concept term to sets of objects in the world as extensions.

To make this clear, the *extent* consists of all objects belonging to a concept and the *intent* is the collection of attributes shared by the objects. A prototype of one class is highly dissimilar to prototypes of another class [Rosch 1988] and does not generally consist of a single exemplar. This aspect of prototype theory is particularly useful in structuring knowledge acquisition where non numeric linguistic variables are used since it is reasonable to assume that if P_1 and P_2 are prototypes then $\text{intent}(P_1) \cap \text{intent}(P_2) = \emptyset$ since then $P_1 \approx P_2$. It also suggests that in acquiring prototypes from experts a range of exemplars is required.

6.2. Structured Knowledge Acquisition in Vague Environments

In this section we introduce an algorithm for structured knowledge acquisition, the following are definitions required both here and in section 9.

Definition 23. (Maximum (Minimum) Term)

A maximum (minimum) term is the greatest (least) point of a concatenated linguistic term. A maximum (minimum) term represented by a trapezoidal fuzzy interval is a maximum (minimum) point.

Definition 24. (Anchor Term)

An anchor term is the most representative point of a concatenated linguistic term, in general it is the minimum (maximum) member of a concatenated term which includes the least (greatest) member of the whole term set. Otherwise it is the central term(s). An anchor term represented by a trapezoidal fuzzy interval is an anchor point. The anchor point is analogous to Zeleny's [Zeleny 1991] notion of an ideal point in multi criteria decision making.

The algorithm is as follows:

1. Acquire a set of decision types D (see section 2.2.) from user
2. Acquire a set of linguistic variables V which are used to distinguish decision types from user.
3. Associate each v_k with one or more d_j to give the pairs (v_k, d_j) .
4. Acquire prototypical minimum and maximum values for (l_{i,v_k}, d_j) from $l_{0,v_k}, \dots, l_{n,v_k}$ and form the meet as the concatenated term $(l_{\min \oplus \max, v_k}, d_j)$ select an anchor term which must be present (usually the max or min of the term set).

The following example based on one that appears in [Wolkenhauer 1999] illustrates this process.

Example 16. (Applying the Algorithm) Suppose we have a decision about whether to pass or fail a student following an assessment. Then $D = \{\mathbf{pass}, \mathbf{fail}\}$. For the purposes of illustration, suppose there are two linguistic variables $V = \{\mathbf{mark}, \mathbf{seminar}\}$ because where a student is borderline the exam board considers evidence about their seminar performance (seminar) in reaching a decision. Each variable is associated with a decision to give the pairs **(mark, pass)**, **(seminar, pass)**, **(mark, fail)** and **(seminar, fail)**. The union of pairs with the same extent that is $\cup_j(v_k, e_j)$ gives (variable decision) pairs, for example **({mark, seminar}, fail)**.

The pass mark is 40 but it is known that this may reflect a mark in the range [35, 45] because of marker variability, representing marks as triangular fuzzy sets using the theory linking tolerance spaces and fuzzy sets, the **pass mark** is a fuzzy set $\langle 35, 40, 45 \rangle$ and any mark m is represented by the fuzzy number $\tilde{m} = \langle m-5, m, m+5 \rangle$. The variable **seminar** takes values from the pre-ordered set

performance = {**none, very poor, poor, mediocre, competent, excellent, superb**}.

The cardinality of this set is seven. This is the optimum cardinality for a term set based on the semantic differential technique [Osgood 1957], the number of pieces of information a person

can handle at any time [Miller 1956] and that used in comparable fuzzy systems [Godo 1989].

The prototypical minimum and maximum values for (**performance**_{seminar}, **pass**) are {**competent**, **superb**} and the concatenated term set is {**competent**, **excellent**, **superb**}. The prototypical term set for (**performance**_{seminar}, **fail**) is {**none**, **very poor**, **poor**}. For the numeric linguistic variable a similar process gives (\tilde{m}_{mark} , **pass**) = 40 so (\tilde{m}_{mark} , **fail**) = 35. The rule base is sparse because pairs such as (35_{mark} , **excellent**_{seminar}) do not have a known extent and are not included within the knowledge base of cases which have a known decision type.

7. Reasoning in Sparse Rule Bases

In the absence of a complete rule base some other means of inference is required. The approach taken here takes Hume's view that:

All kinds of reasoning consist in nothing but a comparison, and a discovery of those relations, either constant or inconstant, which two or more objects bear to each other.

(D.Hume, *The Treatise*, Book I, Part III, Section II)

Suppose for a given inference we have a body of evidence \mathcal{E} and a hypothesis \mathcal{H} but no probabilistic data, function or relation mapping \mathcal{E} to \mathcal{H} . In the decision making context this is equivalent to having a known decision d_j and a case to be classified c_j and no firm data linking d_j to c_j . Then, one way forward is to use similarity or possibility-based methods.

Ruspini [Ruspini 1996] characterizes possibilistic reasoning as follows:

Possibilistic reasoning methods, . . . , determine if there is a modified version of \mathcal{H} (or relaxation) that is consistent with the evidence \mathcal{E} . We may say that we are trying to find out how far we need to stretch [or relax] the truth to fit the hypothesis to the evidence. Possibilistic methods exploit relations of similarity and of relative preference between alternative explanations of the evidence.

Figures 4 and 5 illustrate this process and show how it may be applied to decision making.

Similarity is reflexive and symmetric but not transitive and the process can therefore also be seen as determining how far a hypothesis must be stretched (relaxed) to tolerate the evidence.

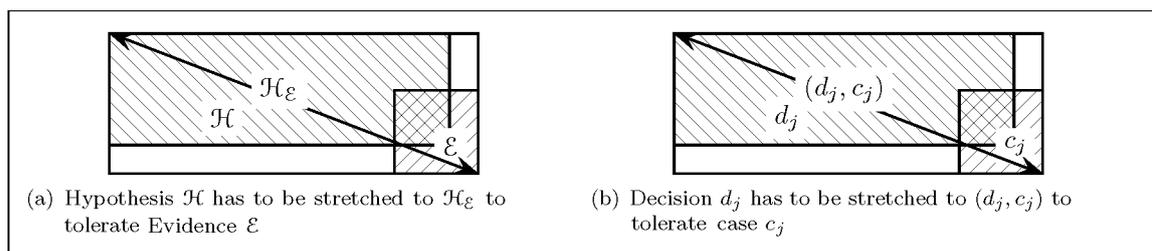


Fig. 4. Possibilistic Reasoning — stretching the hypothesis to tolerate the evidence.

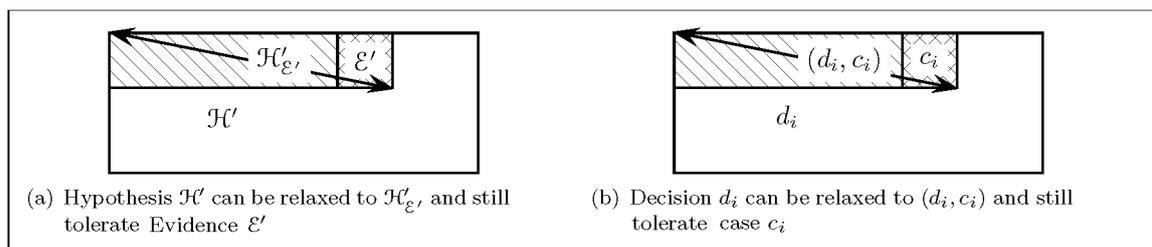


Fig. 5. Possibilistic Reasoning — relaxing the hypothesis to tolerate the evidence.

This notion of stretching or relaxing a hypothesis is closely related to the idea of transformational distance proposed by both Hahn and Chater [Hahn 1997] and Imai [Imai 1977] in the cognitive literature. The tolerances used in section 3. are an example of this kind of stretching. The tolerance can be seen as the maximum amount a hypothesis about the application of a linguistic variable can be stretched and still hold. So $N(1)$ retains a degree of “*oneness*” on all of $(0, 2)$ but not at 2 and 0.

8. Distances between Vague Points

In section 4. we showed that linguistic terms could be modeled with fuzzy numbers, tolerance spaces, and filter bases and that these representations are homomorphic. This gives three possible approaches to finding the distance between vague points. Before doing so, we firstly introduce some distance measures which may be useful in developing the arguments.

8.1. Metrics, Pseudo-metrics, Separations and T_0 -metrics

Definition 25. (Metrics and Pseudo-metrics)

A metric is a function $d : X \times X \rightarrow \mathbb{R}$ such that the following conditions hold:

- M1** $d(x, y) \geq 0, d(x, y) = 0 \Leftrightarrow x = y$
- M2** $d(x, y) = d(y, x); \forall x, y \in X.$
- M3** $d(x, y) + d(y, z) \geq d(x, z) \forall x, y, z \in X.$

A pseudo-metric is a function $d_p : X \times X \rightarrow \mathbb{R}$ such that **M1** is replaced by:

- PM1** $d_p(x, x) = 0.$

These definitions produce the well known Hausdorff or T_2 spaces. A function can also be specified to find the separation between intervals as follows:

Definition 26. The Hausdorff separation is a non-symmetric function [Diamond 1994] (Figure 6) on two sets given by:

$$\mathbf{S} \quad d_H^s(A, B) = \sup\{d(a, B) : a \in A\}.$$

This function is not a metric since $d_H^s(A, B) = 0$ but $A \neq B$ is possible.

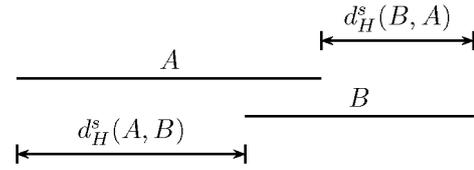


Fig. 6. Hausdorff separation of A and B.

The Hausdorff distance is a metric [Diamond 1994] on the sets A and B given by $d_H(A, B) = \max\{d_H^s(A, B), d_H^s(B, A)\}$

Whilst the Hausdorff separation **S** shows how to construct a non symmetric distance measure a T_0 metric (definition 27) shows how minimality may also be dispensed with.

Definition 27. (T_0 -Metric) A T_0 metric [O’Neil 1998] is a set X, with a function $t : X \times X \rightarrow \mathbf{R}$ such that the following axioms hold:

- T1** $t(x, x) = t(x, y) = t(y, y) \Rightarrow x = y.$
- T2** $t(x, y) \leq t(x, z) + t(z, y) - t(z, z)$
 $\forall x, y, z \in X.$

A space with this set of axioms has a T_0 topology as follows.

Definition 28. Let (S, t) be a space equipped with a T_0 metric. Then for $x \in S$ and $0 < \varepsilon \in \mathbf{R}$ define an open ball

$$B_\varepsilon(x) = \{y \in S : t(x, y) < t(x, x) + \varepsilon\}$$

Lemma 1. If (S, t) is a space equipped with a T_0 metric then the open balls of S are a basis for a T_0 topology on S.

Proof. Suppose that $B_{\varepsilon_x}(x)$ and $B_{\varepsilon_y}(y) \in (S, t)$ and that $z \in B_{\varepsilon_x}(x) \cap B_{\varepsilon_y}(y)$, next define

$$\delta = \min(t(x, x) + \varepsilon_x - t(x, z), t(y, y) + \varepsilon_y - t(y, z)) > 0$$

now show that $B_\delta(z) \subseteq B_{\varepsilon_x}(x) \cap B_{\varepsilon_y}(y)$. Suppose $z' \in B_\delta(z)$ then by **T2**

$$\begin{aligned} t(x, z') &\leq t(x, x) + t(z, z') - t(z, z) \\ &< t(x, z) + \delta \\ &\leq t(x, x) + \varepsilon_x \end{aligned}$$

and $z' \in B_{\varepsilon_x}(x)$. It can be shown in a similar way that $z' \in B_{\varepsilon_y}(y)$. Since $S = \cup_{x \in S} B_1(x)$ the open balls form a basis for a topology on S . Suppose $x, y \in S$ and $t(x, x) < t(x, y)$; let $\varepsilon = t(x, y) - t(x, x) > 0$ then $x \in B_{\varepsilon_x}(x)$ but $y \notin B_{\varepsilon_x}(x)$. Hence the topology of (S, t) is T_0 . \square

Note that **T1** establishes identity, but does not require symmetry or minimality. **T2** is the triangle inequality modified to allow non zero self distances.

Example 17. All metrics are trivially T_0 metrics since $d(x, x) = d(x, y) = d(y, y) = 0$ implies $x = y$ and $d(x, y) \leq d(x, z) + d(z, y) - d(z, z)$ since $d(z, z) = 0$.

Example 18. As a more substantive example, let X be the set of closed intervals $[a, b] \in \mathbb{R}$ then

$$t : X \times X \rightarrow \mathbb{R}$$

$$t([a_1, b_1], [a_2, b_2]) = \max[b_1, b_2] - \min[a_1, a_2]$$

is a T_0 metric but not a metric [Matthews 1997].

In section 3.1. it was established that the topology of a vague set of linguistic terms was T_0 , but not necessarily $T2$ the use of a T_0 metric reflects this fact. The T_0 metric in example 18 can be applied to the separation of a point p and interval $I = [a, b]$ so that

$$D(p, I) = \max(t(p, a), t(p, b)) - t(I, I)$$

which gives

$$D(p, I) \begin{cases} > 0, & p \notin I \\ \leq 0, & p \in I \end{cases}$$

a property which will be useful in section 9.

8.2. Distance Measures and Cognitive Similarity

The metric basis for similarity is proposed by Shepard in the *Universal law of Generalisation*:

A psychological space is established for any set of stimuli by determining metric distances between the stimuli

such that the probability that a response level to any stimulus will generalise to any other is an invariant monotonic function of the distance between them.

Tversky [Tversky 1997] on the other hand raises two major objections to the metric basis of similarity.

Minimality – M2, PM1 is questioned because the probability of judging two stimuli as different is not constant for all stimuli. In recognition experiments an object may be identified with another object more often than it is with itself.

Symmetry – M2 is questioned because many statements of similarity appear to be directional so a is like b rather than b is like a . For instance, an ellipse is judged more similar to a circle than a circle to an ellipse. Note that this is an asymmetry in the judged degree of similarity, not a denial of the reflexivity of similarity.

These objections can be overcome fairly straightforwardly when generalizing the metric space axioms to apply to intervals rather than points: the minimality objection by having a mapping into \mathbb{R} (as in a T_0 metric) instead of \mathbb{R}^+ the symmetry objection by using non-symmetrical directional separation functions reflecting the fact that a hypothesis must be stretched or relaxed from some fixed point beyond which it no longer holds.

If a metric-based approach is accepted, the question then arises of what specific metric best models the intuitive approach which people have. Attneave [Attneave 1950] and Shepard [Shepard 1980] both suggest the d_1 or city block metric given by equation 7

$$d_1(x_i, y_i) = \sum_{i=1}^n |x_i - y_i| \quad (7)$$

is the most appropriate. Shepard makes the particular point that

... there is a fundamental connection between the sharp-cornered form of the isosimilarity contour and the city block metric and some of the significant phenomena of discontinuous or insightful learning and of instabilities of choice ...

The isometric curves for the city block metric are continuous but not differentiable. The compression of an axis by weighting can therefore mean minor shifts in adjudged distance cause major shifts in judged similarity [Everitt 1997].

In section 9, a distance measure is defined which whilst a metric for points is not for intervals. This measure also meets Tversky's objections. Summation will be applied in a way which gives a d_1 metric when it is applied to multi-dimensional points.

8.3. Fuzzy Approaches

There are number of approaches to finding the distance between fuzzy sets in the literature [Klir 1995, Diamond 1994, Goetschel 1993, Koczy 1993, Kaleva1987]. Most are based on the Hausdorff distance between the alpha-level cuts or the extension principle. However these approaches are not guaranteed to produce output sets of the same nature as the input sets [Diamond 1984, Hsiao 1996]. In empirical studies it has been found that the similarity measures on fuzzy sets which correlate best with the similarity between the verbal descriptions of those sets are those which are based on the kernels (\tilde{A}_1), or the centre of gravity of the fuzzy sets [Yoshikava 1996, Zwick1988]. These are important findings which lend further weight to the tolerance space model of fuzzy sets discussed earlier.

Another approach to fuzzy numbers is to treat them as tolerance spaces. We have already shown in section 4. how we can derive fuzzy sets from a tolerance space. Distances in a tolerance space can be treated as tolerance mappings (definition 13) for example

$$d : \langle \mathbb{R}, \xi \rangle \times \langle \mathbb{R}, \xi \rangle \rightarrow \langle \mathbb{R}, \xi \rangle$$

$$d(x, y) = |x - y|$$

This is easily translated to a fuzzy set using the techniques outlined earlier.

9. Distances between Cases and Prototypes

In this section, a proposal for a distance measure, which meets the criteria outlined for it in the preceding sections, will be made. Since

prototypes are composed of concatenations of linguistic terms, they are represented by intervals rather than points. The notion of an anchor point introduced in section 6.2. will be needed again here. The considerations discussed in sections 4. and 7. will also be taken into account.

Example 19. The anchor term of { **competent** \oplus **excellent** \oplus **superb** } taken from the pre-ordered set

$$\text{performance} = \{ \text{none, very poor, poor, mediocre, competent, excellent, superb} \}$$

is **superb**. The anchor term of { **poor** \oplus **mediocre** \oplus **competent** } is **mediocre**. In applications these terms are represented by a fuzzy number which may be referred to as the *anchor point*.

The distance between two terms described by fuzzy numbers can be found as follows. Let $\tilde{A} = \langle a_1, a_2, a_3, a_4 \rangle$ then its *centre* is $c(\tilde{A}_1)$ is $(a_2 + a_3)/2$ and the alpha-cut at 1 gives a tolerance $\tau(\tilde{A}_1) = (a_3 - c)$ the alpha-cut at 0 gives the tolerances $\tau_L(\tilde{A}_0) = c - a_1$ and $\tau_R(\tilde{A}_0) = a_4 - c$ so $\tilde{A} = \langle c - \tau_L, c - \tau, c + \tau, c + \tau_R \rangle$ also denoted $\langle c; (\tau, \tau_L, \tau_R) \rangle$.

For the prototype (intent, extent) pair

$$\mathbf{P} = \left(\prod_{i=1}^k W_i \tilde{P}_i, P \right)$$

and the case pair

$$\mathbf{C} = \left(\prod_{i=1}^k W_i \tilde{C}_i, C \right)$$

where W_i is the weight for the i^{th} variable the weighted distance between the centres of the intent terms $c(D)$ is:

$$c(D) : \tilde{\mathbb{R}}_{\square}^n \times \tilde{\mathbb{R}}_{\square}^n \rightarrow \mathbb{R}$$

$$c(D(\mathbf{C}, \mathbf{P})) = \sum_i^n W_i \begin{cases} c(C_i) - c(\max(P_i)), & \text{if } c(\text{anchor}(P_i)) \leq c(C_i), \\ c(\min(P_i)) - c(C_i), & \text{if } c(\text{anchor}(P_i)) > c(C_i), \end{cases} \quad (8)$$

where $\tilde{\mathbb{R}}_{\square}^n$ is the set of trapezoidal fuzzy sets $\langle x_1, x_2, x_3, x_4 \rangle$ defined on \mathbb{R}^n .

Then the expression for the distance between case and prototype becomes

$$D : \tilde{\mathbb{R}}_{\square}^n \times \tilde{\mathbb{R}}_{\square}^n \rightarrow \tilde{\mathbb{R}}_{\square}^n$$

$$D(\mathbf{C}, \mathbf{P}) = \left\langle \begin{array}{l} c(D); (\max\{W_i\tau_i\}, \max\{W_i\tau_{i_L}\}, \\ \max\{W_i\tau_{i_R}\}) \end{array} \right\rangle,$$

which is the trapezoidal fuzzy interval

$$\left\langle \begin{array}{l} c(D) - \max\{W_i\tau_{i_L}\}, c(D) - \max\{W_i\tau\}, \\ c(D) + \max\{W_i\tau\}, c(D) + \max\{W_i\tau_{i_R}\} \end{array} \right\rangle.$$

Similarity can then be calculated as a function of distance as in section 7. Similarity measures based on D meet all of Tversky's objections to distance based similarity. D does not imply minimality, since a case may be closer to a prototype than to itself as self distance is around 0 but distance from a prototype may be negative.

A negative distance implies that, rather than being similar to the prototype, the case is to some extent a *typical* example of that prototype [Osherson 1997], unlike similarity which is usually expressed in the interval $[0, 1]$ typicality is not necessarily bounded above [Osherson 1997, pp 190]. In decision making this is the type of case whose outcome is immediately obvious to a human decision maker. In the student domain this could be a student with a mark in excess of say 80 whose performance is excellent; such a student is typical of the kind of student we would wish to pass, and more typical than a student with a mark of 45 who has performed competently in seminars. However both are examples of the pass prototype.

The measured extent of similarity/typicality based on D is also neither symmetric nor transitive. However, when applied to the crisp numbers, D is the d_1 (city block) metric, as suggested in section 3.1., and symmetry and transitivity are restored.

Example 20. Having acquired prototypes in example 16, the student example is developed further. Suppose that student obtains a mark of 38 and their performance in seminars is excellent. Should that student be passed or failed? Marks have been found to be 10 times more important than seminar performance in making this decision.

The fail prototype is represented by the (intent, extent) pair

$$\mathbf{P}_{\text{fail}} = (\langle \{\text{none, very poor, poor}\} \times [0, 35] \rangle, \text{fail})$$

the anchor terms are **none** and 0; the maximum terms are **poor** and 35. This gives the following giving the numerical representation:

$$\mathbf{P}_{\text{fail}} = \left(\langle [0, 3]; \langle \pm 0.25, -0.75, +0.75 \rangle \rangle \times \langle [0, 35], \langle \pm 0, \pm 5, \pm 5 \rangle \rangle, \text{fail} \right)$$

with the centres of the anchor and maximum points given by

$$c(\text{anchor}([0, 3])) = 0, c(\text{anchor}([0, 35])) = 0;$$

and

$$c(\text{max}([0, 3])) = 3, c(\text{max}([0, 35])) = 35,$$

which is normalised to

$$\mathbf{P}_{\text{fail}} \left(\langle [0, 0.5], \langle \pm 0.04, -0.13, +0.13 \rangle \rangle \times \langle [0, 0.35], \langle \pm 0, -0.05, +0.05 \rangle \rangle, \text{fail} \right)$$

Similarly the pass prototype is represented by the (intent, extent) pair

$$\mathbf{P}_{\text{pass}} = \left(\langle \{\text{competent, excellent, superb}\} \rangle \times [40, 100] \rangle, \text{pass} \right)$$

giving the normalised numerical representation of the pass prototype as

$$\mathbf{P}_{\text{pass}} \left(\langle [0.75, 1], \langle \pm 0.04, -0.13, +0.13 \rangle \rangle \times \langle [0.4, 1], \langle \pm 0, -0.05, +0.05 \rangle \rangle, \text{pass} \right)$$

applying the weighting to the mark variable gives

$$\mathbf{P}_{\text{fail}} \left(\langle [0, 0.5], \langle \pm 0.04, -0.13, +0.13 \rangle \rangle \times \langle [0, 3.5], \langle \pm 0, -5, +5 \rangle \rangle, \text{fail} \right)$$

and

$$\mathbf{P}_{\text{pass}} \left(\langle [0.75, 1], \langle \pm 0.04, -0.13, +0.13 \rangle \rangle \times \langle [4, 10], \langle \pm 0, .5, +5 \rangle \rangle, \text{pass} \right)$$

a similar exercise for the input case

$$\mathbf{C} = (\langle \{\text{excellent} \times 38\} \rangle, \text{case})$$

gives

$$\left(\langle 0.83, \langle \pm 0.04, -0.13, +0.13 \rangle \rangle \times \langle 3.5, \langle \pm 0, -5, +5 \rangle \rangle, \text{case} \right)$$

Applying equations 8 and 9 gives

$$D(C, \mathbf{P}_{\text{fail}}) = \langle 0.13, 0.59, 0.67, 1.13 \rangle$$

and

$$D(C, \mathbf{P}_{\text{pass}}) = \langle -0.22, 0.24, 0.32, 0.78 \rangle$$

Similarity (η) can be calculated by applying a continuous monotone function; an exponential function $\eta = e^{-D}$ is used here.

$$\eta(C, \mathbf{P}_{\text{fail}}) = \langle 0.32, 0.51, 0.55, 0.88 \rangle$$

and

$$\eta(C, \mathbf{P}_{\text{pass}}) = \langle 0.46, 0.73, 0.79, 1.25 \rangle$$

10. Outputs and Outcomes

There are five possible outcomes, four of which are illustrated in Figures 7 and 8.

In these figures dec_A is the set representing $\eta(C, \text{dec}_A)$ (the similarity of an input case to

decision prototype A) and dec_B the set representing $\eta(C, \text{dec}_B)$ where dec_X is a decision prototype and C is a case. The support and the kernel of the fuzzy set are denoted supp and ker respectively and it is assumed in cases 1 – 4 that $\text{ker}(\eta(C, \text{dec}_A)) \gtrsim \text{ker}(\eta(C, \text{dec}_B))$

Case 1 $\text{supp}(\eta(C, \text{dec}_A)) \not\sim (\eta(C, \text{dec}_B))$, Figure 7(a). Where decision prototype dec_B is strongly preferred. A person would not usually hesitate to make this decision and might describe it as self evident.

Case 2 $\text{supp}(\eta(C, \text{dec}_A)) \sim \text{supp}(\eta(C, \text{dec}_B))$, Figure 7(b). Where decision prototype dec_B is preferred. A person would usually make this decision and without difficulty but it would not be self evident.

Case 3 $\text{ker}(\eta(C, \text{dec}_A)) \sim \text{supp}(\eta(C, \text{dec}_B))$, Figure 8(a). Where decision prototype dec_B is weakly preferred. A person would perhaps be hesitant in making this decision but would usually be content to make it on the basis of the evidence.

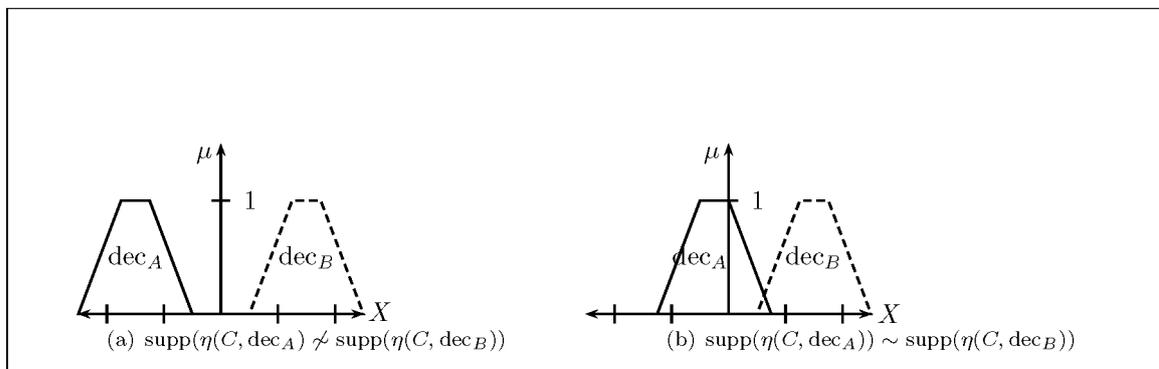


Fig. 7. Cases 1 and 2.

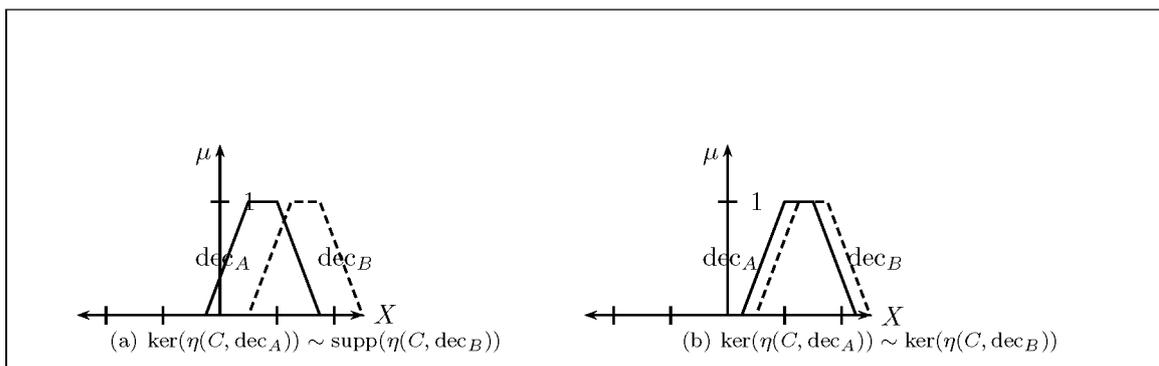


Fig. 8. Cases 3 and 4.

Case 4 $\ker(\eta(C, \text{dec}_A)) \sim \ker(\eta(C, \text{dec}_B))$, Figure 8(b). Where decision prototype dec_B is very weakly preferred. These are the cases where a person would usually consider it advisable to apply another test (assuming one is available) before making a decision.

Case 5 $\ker(\eta(C, \text{dec}_A)) = \ker(\eta(C, \text{dec}_B))$. Where it is undecidable which decision prototype dec_A or dec_B is preferred. These are the cases where a person would want to apply another other test (or flip a coin!) before making a decision.

This process allows a mapping from linguistic inputs to linguistic outputs.

Example 21. Returning again to the examples 16 and 20 we have the output sets shown in Figure 9 which shows $\text{supp}(\eta(C, \mathbf{P}_{\text{fail}})) \sim \ker(\eta(C, \mathbf{P}_{\text{pass}}))$. So a decision to pass is weakly preferred. This reflects the kind of decision that might be made in reality given these circumstances. A student with a mark of 38, but an excellent seminar record, would usually be passed – but only just. For this example we could have the following pre-ordered set of linguistic outputs

poor fail \succsim **fail** \succsim **just fail** \succsim **viva voce** \succsim
just pass \succsim **pass** \succsim **good pass**.

Case 4 and case 5 have been assigned the same linguistic output on this scale.

If the two distances do not overlap, then the course of action should be clear. If they do overlap, then it may be that an alternative way of distinguishing between the alternatives should

be considered depending on the degree of uncertainty. In this domain it indicates if there is a case for giving the student a *viva voce*; in other domains it might trigger the use of some other additional selection test. A measure of the degree of overlap is given by intersection of the membership functions. So in this example the case for giving the student a *viva voce* is stronger than the case for not doing so since $\mu_{D(C, \mathbf{P}_{\text{fail}})}(x) \cap \mu_{D(C, \mathbf{P}_{\text{pass}})}(x) \approx 0.7$. This gives a decision maker an alternative way of resolving the case should they wish to do so.

11. Conclusion

By starting with an intuitive set of assumptions about the mathematical properties of a set of non numeric linguistic terms it has been shown that pre-ordered sets of linguistic terms can be modeled with fuzzy numbers, filter bases and tolerance spaces. Using filter bases and tolerance spaces allows the distance between fuzzy numbers to be found in a way which is considered more intuitive than the usual approaches based on the extension principle for fuzzy numbers. The use of a measure which is not a metric to find distances between cases and prototypes overcomes objections [Tversky 1977] to a purely geometric approach to similarity. The relative distance of input cases from different decision prototypes gives fuzzy numbers which can then be used for similarity-based reasoning in sparse, linguistically valued rule bases. The relative positions of the output fuzzy sets also make it possible to devise pre-ordered sets of

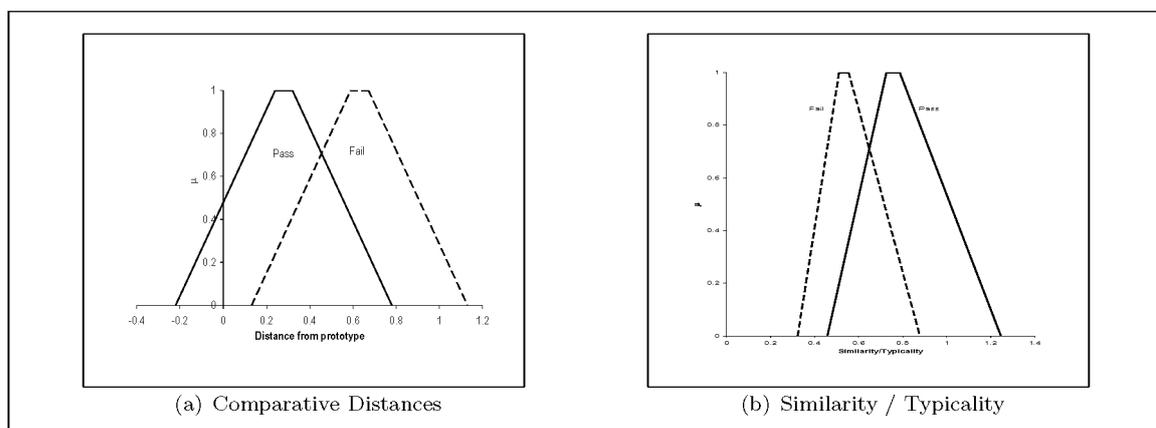


Fig. 9. Outputs for Student Example.

linguistic outputs which offer a decision maker guidance on how they may wish to proceed. In developing systems using these techniques the aim is not to replace the human decision maker but rather to offer them the chance to take “advice” from a source not directly involved with the decision.

References

- [1] R.F. ALBRECHT, On Mathematical Systems Theory. In *Systems: Theory and Practice* (R.F. Albrecht, Ed.), (1998) pp. 33–86. Springer.
- [2] F. ATTNEAVE, Dimensions of similarity. *American Journal of Psychology*, **63** 1950, pp. 516–556.
- [3] G. BOJADZIEV, M. BOJADZIEV, *Fuzzy Sets, Fuzzy Logic, Applications*, World Scientific, 1995.
- [4] N. BOURBAKI, *General Topology*, Addison-Wesley, 1975.
- [5] B.A. DAVEY, H.A. PRIESTLEY, *Introduction to Lattices and Order*, Cambridge University Press, 1990.
- [6] P. DIAMOND, P. KLOEDEN, *Metric Spaces of Fuzzy Sets: Theory and Application*, World Scientific, 1994.
- [7] A.O. ESBOGUE, R.C. ELDER, Fuzzy sets and the modeling of the physician decision processes: Part II: Fuzzy diagnosis decision models, *Fuzzy Sets and Systems*, **3** 1980, pp. 1–9.
- [8] A.O. ESBOGUE, R.C. ELDER, Measurement and valuation of a fuzzy mathematical model for medical diagnosis, *Fuzzy Sets and Systems*, **10** 1983, pp. 223–242.
- [9] B.S. EVERITT, S. RABE-HESKETH, *The Analysis of Proximity Data*, Arnold, 1997.
- [10] L. GODO ET AL., MILORD: The Architecture and the Management of Linguistically based Uncertainty, *International Journal of Intelligent Systems*, **4** 1989, pp. 471–501.
- [11] R. GOETSCHEL, W. VOXMAN, Topological Properties of Fuzzy Sets, *Fuzzy Sets and Systems*, **10** 1983, pp. 87–99.
- [12] U. HAHN, N. CHATER, Concepts and similarity, In *Knowledge, Concepts and Categories* (L. Lambart and D. Shanks Eds.) 1992, pp. 43–91, Psychology Press.
- [13] J. HAMPTON, Conceptual combination, In *Categories and Concepts* (Mechelen, I.V. et al. Eds.), 1993, pp. 67–96, Academic Press.
- [14] W-H. HSIAO, ET AL., A new interpolative reasoning method in sparse rule-based systems, *Fuzzy Sets and Systems*, **93** 1998, pp. 17–22.
- [15] F. HOVESEPIAN, *A Metalogical Analysis of Vagueness: An Exploratory Study into the Geometry of Logic*, Ph.D. Thesis, University of Warwick, 1992.
- [16] S. IMAI, Pattern similarity and cognitive transformations *Acta Psychologica*, **41** 1977, pp. 433–447.
- [17] O. KALEVA, S. SIEKKALA, On fuzzy metric spaces, *Fuzzy Sets and Systems*, **12** 1987, pp. 301–317.
- [18] J.L. KELLEY, *General Topology*, Von Nostrand, 1955.
- [19] G.J. KLIR, B. YUAN, *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Prentice Hall, 1995.
- [20] L.T. KOCZY, K. HIROTA, Ordering and closeness of fuzzy sets, *Fuzzy Sets and Systems* **59** 1993, pp. 282–293.
- [21] V.B. KUZ'MIN, A parametric approach to the description of linguistic variables and hedges, *Fuzzy Sets and Systems* **6** 1981, pp. 27–41.
- [22] P.J.M. VAN LAARHOVEN, W. PEDRYCZ, A fuzzy extension to Saaty's priority theory, *Fuzzy Sets and Systems*, **11** 1983, pp. 229–241.
- [23] M. MAREŠ, R. MESIAR, Calculation over verbal quantities. In: *Computing with Words in Information/Intelligent Systems 1–Foundations* (L.A. Zadeh, J. Kacprzyk Eds.) (1999) pp. 409–427. Physica-Verlag.
- [24] S. MATTHEWS, *Three Approaches to Studying Vagueness* (1997)
<http://www.dcs.warwick.ac.uk/~sgm/pmetric/index.html>
- [25] R. MOORE, *Interval Analysis*, Prentice Hall, 1966.
- [26] G.A. MILLER, The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Information Processing, *Psychological Review* **63** 1956, pp. 81–97.
- [27] J.R. MUNKRES, *Topology*, Prentice Hall, 1975.
- [28] L. NACHBIN, *Topology and Order*, Von Nostrand Mathematical Studies, 1965.
- [29] R.M. NOFOSKY, Choice, Similarity, and the Context Theory of Classification, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **10** 1984, pp. 104–114.
- [30] S.J. O'NIEL, *A Fundamental Study into the Theory and Application of the Partial Metric Spaces*, Ph.D. Thesis, University of Warwick, 1998.
- [31] C.E. OSGOOD, ET AL. *The Measurement of Meaning*, University of Illinois Press, 1957.
- [32] D. OSHERSON, E.E. SMITH, On typicality and vagueness, *Cognition*, **64** 1997, pp. 189–206.
- [33] W. PAGE, *Topological Uniform Structures*, Dover, 1978.
- [34] E. ROSCH, Principles of Categorisation In *Readings in Cognitive Science*, (A. Collins, E.E. Smith, Eds.), 1988, pp. 312–322, Morgan Kaufmann.

- [35] E. RUSPINI, *Approximate Reasoning Foundations*, <http://www.ai.sri.com/~ruspini/ar.html>.
- [36] T.L. SAATY, *The Analytic Hierarchy Process*. New York, 1980.
- [37] R.N. SHEPARD, Multidimensional scaling, tree-fitting and clustering *Science*, **210** 1980, pp. 340–398.
- [38] R.N. SHEPARD, Towards a Universal Law of Generalisation for Psychological Science *Science*, **237** 1987, pp. 1317–1323.
- [39] L.N. STOUT, The Logic of Unbalanced Subobjects in a Category with Two Closed Structures, In *Applications of Category Theory to Fuzzy Subsets* (E. Rodabaugh et al., Eds.), 1992, pp. 73–106, Kluwer Academic Publishers.
- [40] A. TVERSKY, Features of similarity *Psychological Review*, **84** 1977, pp. 327–352.
- [41] O. WOLKENHAUER, *Data Engineering*, 1999, <http://www.csc.umist.ac.uk/fstb/dsu.pdf>
- [42] R.R. YAGER, Fuzzy decision making including unequal objectives *Fuzzy Sets and Systems*, **1** 1978, pp. 87–95.
- [43] A. YOSHIKAWA, T. NISHIMURA, Relationship between subjective degree of similarity and some similarity indices of fuzzy sets, Methodologies for the Conception, Design, and Application of Intelligent Systems (Presented at *Proceedings of IIZUKA'96*, 1996, pp. 818–821.
- [44] L.A. ZADEH, Fuzzy Sets *Information and Control*, **8** 1965, pp. 856–865.
- [45] L.A. ZADEH, The concept of a linguistic variable and its application to approximate reasoning., Memorandum ERL-M 411, Berkley, 1973.
- [46] L.A. ZADEH, Fuzzy Logic = Computing with Words, *IEEE Transactions on Fuzzy Systems*, **4** 1996, pp. 103–111.
- [47] E.C. ZEEMAN, The topology of the brain and visual perception In *The Topology of 3-Manifolds* (M. K. FORT, Ed.), 1962, pp. 240–256. Prentice Hall.
- [48] M. ZELENY, Cognitive equilibrium: A knowledge-based theory of fuzziness and fuzzy sets, *International Journal of General Systems*, **19**, 1991, pp. 359–381.
- [49] H-J. ZIMMERMANN, *Fuzzy Sets, Decision Making, and Expert Systems*, Kluwer Academic Publishers, 1987.
- [50] H-J. ZIMMERMANN, *Fuzzy Set Theory – and Its Applications, Second Edition*, Kluwer Academic Publishers, 1990.
- [51] R. ZWICK, ET AL., Measures of Similarity Among Fuzzy Concepts: A Comparative Analysis *International Journal of Approximate Reasoning*, **1** 1988, pp. 221–241.

Received: January, 2001
Accepted: September, 2002

Contact address:

Jon Williams,
2 Pershore Road
Upton Snodsbury
Worcester, UK
WR7 4NR

e-mail: jdw47@tutor.open.ac.uk

JON WILLIAMS is an Associate Lecturer with the Open University and was formerly Head of Health and Housing at Wychavon District Council. He holds first degrees from the University of Aston and the Open University, and a Masters from Oxford Brookes University. He is currently a part-time post graduate Ph.D student at the Coventry University.

PROFESSOR NIGEL STEELE was educated at King Henry VIII School Coventry and Southampton University and is currently Head of Mathematics at Coventry University. He is co-author of a text book on 'Mathematics in Signal Processing', a contributing author to a major text book on modern engineering mathematics. He has written a number of papers and articles in such fields as fluid mechanics, applications of control theory and lately, neural networks and fuzzy systems. He has a particular interest in the mathematical education of Engineers and was a co-author of the SEFI core curriculum report, which has influenced developments in this field.

He is a Fellow of the Institute of Mathematics and Its Applications and a Chartered Mathematician. He is currently an Hon. Secretary of the Institute, with responsibility for Education.

DR. HELEN ROBINSON has a first degree from Cambridge University and a topology Ph.D from Liverpool University. She has worked at Coventry University, where she is a Senior Lecturer, since 1970 and has been carrying out work related to fuzzy sets for around 4 years. She spent a short time early in her career seconded to Marconi, and some time as a research assistant at Warwick University in 1991. She is Chairman of the London Mathematical Society Women in Mathematics Committee.
