Learning-Based Similarity Measurement for Fuzzy Sets

Athena Tocatlidou*
A. I. Group, Department of Engineering Mathematics, University of Bristol, BS8 1TR, Bristol, England

The work described in this paper proposes a method for the measurement of similarity, viewed from the decision maker’s perspective. At first, an algorithm is presented that generalizes a discrete fuzzy set $F$, representing a model, given another discrete fuzzy set $G$ representing new evidence. The algorithm proceeds by expressing the fuzzy sets as possibility distributions, and then by extending the focal elements of $F$ with elements from the focal elements of $G$, constructs a generalized fuzzy set. If the initial fuzzy model is repetitively updated with the same evidence, a convergence state will be reached, the number of repetitions depending on the relative position of the two sets. This number is considered to give an indication of the conceptual proximity of the two entities, and therefore of the two fuzzy sets, and can be used as a similarity index. This proposal for similarity seems to be in better agreement with experimental findings in human similarity judgment than approaches based on pointwise distance metrics. © 1998 John Wiley & Sons, Inc.

I. INTRODUCTION

The role that similarity plays in many theories of perception, categorization, decision making, and analogical and case-based reasoning is fundamental. Still similarity exhibits a multifaceted character, changing with the experience of the decision maker, the context, the goal of the comparison task, the focus on structural or surface aspects, etc. Research on modelling similarity judgments in the behavioral and social sciences has moved from multidimensional distance metric models, to featural models concerned with aspects of human similarity statements such as the violation of the symmetry and triangular equality axioms. However, an open question remains, the appropriate selection of the relevant features, and a definition of similarity as a function of a specific comparison task could provide a more consistent way of determining similarity measurements.

*The author’s present address is Department of Mathematics, University of the Aegean, Karlovassi 83200, Samos, Greece; E-mail: atocat@kerkis.math.aegean.gr.

© 1998 John Wiley & Sons, Inc. CCC 0884-8173/98/020193-28
The present work consists essentially of five parts. First, the flexible and multifaceted nature and role of similarity are discussed, starting with the negative view of N. Goodman, regarding the possibility of definition of an “all-purpose” similarity measure, suitable for all contexts and tasks. Subsequently, the trends for similarity measurement within the social and behavioral sciences are outlined. The motivation for this is twofold: First it will highlight the main findings from these research fields that should be respected in any modeling of human similarity judgment (e.g., in decision making, diagnosis, reasoning, categorization, etc.). Second, the above findings and results can potentially provide useful guidelines and ideas for mathematical models of similarity measurement. This section is followed by a review of the modeling of similarity within fuzzy sets theory. All of the above sections have another purpose: to set the scene for the proposal and evaluation of a similarity ordering of fuzzy sets, defined in Section IV, in which similarity is viewed from another perspective: as a task-dependent notion. More specifically, the similarity between two fuzzy concepts is defined as the amount of effort that is necessary to learn (generalize) from one concept to another. This seems to be a more generic definition, and it is in agreement with psychological findings. Finally, in the fifth section, a small class of fuzzy-set pairs is used for testing and evaluating some representative similarity measures, and the new proposed indexes.

II. ON THE ROLE AND NATURE OF SIMILARITY

In 1972 Nelson Goodman accused similarity—emphasising its flexibility and multiaspect nature—of being too flexible and not robustly defined, going so far as to call it a pretender and false friend, “often found where it doesn’t belong, and professing powers it doesn’t possess.” It is a concept “clear enough when closely confined by context and circumstance, but it is hopelessly ambiguous when torn loose.” It needs a frame of reference, much like the notion of motion; otherwise any similarity statement is incomplete and unable to provide a stable foundation on which to ground categorization and recognition, since it is relative, variable, and context dependent.

In predictive and inductive reasoning, the aspects of similarity agreed to form the important criteria play a decisive role, their selection often being a goal-driven procedure. They therefore behave more like a dependent than an independent variable, a realization causing Goodman to state that “rather than similarity providing any guidelines for inductive practice, inductive practice may provide the basis for some canons of similarity.”

Another important concern expressed by Goodman is that often similarity judgments are based on partial matching, and a consideration of the whole structure of the compared objects is missing. In his words, “Similarity between particulars does not suffice to define qualities.” As an example, suppose we have three discs, the first one half red and half blue, the second one half blue and half yellow, and the third half yellow and half red. Although they share in pairs of color quality, there is no color in common to all three. In other words, dyadic likeness is not sufficient to define classes of entities. Even when important and
relevant aspects are explicitly considered in the comparison process, they do not
guarantee a robust definition of similarity, since importance and relevance are
volatile concepts, changing in any modification in context and interest.

A very important factor in setting the aspects in similarity judgments is the
role of the decision maker. In an airport check-in station, baggage is judged in
terms of its shape, size, and make by a spectator, in terms of its weight by the
pilot, and in terms of its ownership by a passenger. In that respect, laboratory
results, ignoring the decision maker in many cases, create rather than reflect
similarity among entities, often ordering them along a predefined scale, but
“they may be ordered, with good reason, in many different ways.”

Despite the previous criticism, the fundamental place that similarity holds
in many theories of perception, knowledge, and reasoning cannot be denied. It is
very difficult—if it is even possible at all—to define any procedure of catego-
rization or classification, explanation and retrieval, concept formation, inductive
reasoning, decision making, recognition, metaphor understanding, etc., without
using some sort of similarity measure. Moreover, in many cases, the membership
of an entity in a category itself is decided on the basis of its typicality for that
category. As Tversky and Kahneman\(^2\) have experimentally demonstrated, an
often used heuristic for estimating the probability that object \(x\) has property \(P\),
is in terms of how representative \(x\) is of \(P\).

It is clear, then, that “the human mind has a considerable investment in
similarity,”\(^3\) and similarity judgment is a very powerful and flexible capability of
human beings. Its role is so important that the performance of a model in
categorization, clustering, reasoning, and decision making depends to a great
extent on the success of the similarity measure used.

According to Goodman’s criticism, even when similarity exhibits a very
flexible character to ground categorization, and two entities can be considered
similar in too many ways to extract any useful information for this, constraints
can be imposed in different ways. One important mechanism generating con-
straints is the comparison process itself: When objects enter into a comparison,
the relevant common properties become relatively fixed.\(^4\) This process intro-
duces some kind of alignment between compared entities, and similarity in-
creases as the number of common features increases and the number of
distinctive features decreases.

From the experimental and theoretical literature in the behavioral and
social sciences, it is apparent that similarity is indeed characterized by a high
degree of flexibility, and is related to the task it is involved in. More specifically,

(i) Similarity changes with experience. Experiments\(^3\) show that expertise is a factor
affecting the feature selection for the comparison. Novices usually select
surface features, whereas experts search for deeper underlying structures and
principles.

(ii) Similarity changes with the context. Similarity judgments vary with the context
provided, the basic explanation being that different context characteristics
make salient or activate different aspects of the compared objects.\(^3, 4\) There are
reported cases where even the linguistic differences in phrasing a query may
result in different similarity evaluations.
Similarity is a function of the structural analogies between compared entities, and this constitutes the core element in analogical reasoning. For example, in the statement, “An atom is like the solar system,” the focus is on structural relations like “resolves around” or “attracts,” rather than on attribute values like “yellow” or “hot.”

Similarity in many cases is task or goal driven. To determine whether an unknown species of snake is poisonous, one compares it with other, familiar species of snakes, retrieving from the list of features associated with the object the relevant ones (e.g., color, shape, etc.), and ignoring others, like “snake rhymes with flake.” This notion of similarity is particularly useful in cased-based reasoning and retrieval, and ensures a better performance, since the set of features is smaller (containing only the useful ones), and the selected features are highly relevant to the final goal. It also serves as a filtering mechanism that imposes constraints on the otherwise infinite number of possible features associated with the compared objects.

Two main methods characterize the measurement of uncertainty in this area: geometric and featural. For the first, the main assumption is that the observed measure of similarity between entities is a monotonic decreasing metric function $f$ of the interpoint distance in an appropriate Euclidean space. In 1977 Tversky theoretically and experimentally questioned the metric proposals for similarity measurement. More specifically, he criticized the three metric axioms:

(a) The minimality axiom. Experimental and theoretical evidence has suggested that not all objects in the stimulus space considered share the same status. Certain objects are considered more prototypical than others, or are in some way more central to the domain, or are particularly salient and distinctive.

These properties affect the judgment of self-similarity for these objects, resulting in a higher measure for objects that are either simple or well distinguished. Thus the evidence that the measure of self-similarity of an object varies according to the special characteristics or the prototypicality of the object, and therefore is not the same for all objects, suggests a violation of the minimality axiom.

(b) The symmetry axiom. Similarity statements (and therefore similarity judgments) are directional. They have a referent, usually the most prototypical or salient, and a subject (we say “an ellipse is like a circle” and not “a circle is like an ellipse”). The converse similarity statement is not usually equivalent. The statement “This butcher is like a surgeon” reveals a completely different meaning than the statement “This surgeon is like a butcher.” Rosch also reported experimental findings in which subjects tend to place more salient objects in the referent position of similarity statements of the form “is

\[\text{is}^\dagger\] this is especially true in similes and metaphors as in “life is like a play” implying that people play roles, and “a play is like life” meaning that it captures essential elements of human life.
essentially "" or "" is virtually ."" This is further confirmed by the nonsymmetrical confusion matrices resulting from experimental similarity evaluations.\textsuperscript{3,10,11}

(c) The triangular inequality. The triangular inequality practically implies that if $a$ is similar to $b$, and $b$ is similar to $c$, then $a$ and $c$ cannot be very dissimilar. But according to W. James,\textsuperscript{11} flame is similar to the moon because they are both bright, and the moon is similar to a ball because they are both round. However, contrary to the triangular axiom, a flame and a ball are very dissimilar. This happens because the distance metric-based triangle inequality fails to capture the fact that, when objects are compared, only a subset of their features, the relevant ones, are involved in the task.

Following these considerations, Tversky proposed a feature-based model (known as the contrast model),\textsuperscript{8} according to which for two entities $a$ and $b$, their similarity $S(a,b)$ is given by the formula

$$S(a,b) = \theta f(A \cap B) - \alpha f(A - B) - \beta f(B - A)$$

where $A$ and $B$ are the set of features describing objects $a$ and $b$, respectively; $f$ is an additive function (usually the cardinality of $A$ and $B$); and $\theta, \alpha, \beta \geq 0$ are weight parameters (to emphasize common or a distinctive set of features).

### III. MODELLING OF THE NOTION OF SIMILARITY WITHIN FUZZY SET THEORY

The proposals for the measurement of similarity for the case of fuzzy sets are in their majority distance based, and they usually ignore asymmetric and nontransitive experimental similarity data. A possible explanation for this is that the compared fuzzy sets are considered as distributions from a mathematical approximation point of view, rather than from a decision-making perspective. This is reflected in the common characteristics found in most proposals for similarity measurement, such as pointwise calculations and min or max comparison operators, whereas a more holistic similarity judgment for the compared entities is missing.

The initial connection of similarity and the theory of fuzzy sets was initiated by Zadeh, who introduced similarity relations as a generalization of equivalence relations. Thus, if $(E_\alpha) \in [0,1]$ is a nested sequence of equivalence relations in $X$ (i.e., for $\alpha_1 > \alpha_2 \Rightarrow E_{\alpha_1} \subset E_{\alpha_2}$), $E_1$ nonempty and $\text{dom}E_\alpha = \text{dom}E_1$, then for any choice of $\alpha \in [0,1]$, a similarity relation $S$ is a fuzzy relation in $X$, defined as

$$S = \sum_\alpha \alpha E_\alpha \quad \text{where } 0 < \alpha \leq 1$$

where the nonfuzzy set $E_\alpha \in X \times X$ is defined by $E_\alpha = \{(x,y) | \mu s(x,y) \geq \alpha\}.\textsuperscript{12}$
As an example, the similarity relation defined by the fuzzy relation matrix

\[
\begin{array}{|c|c|c|c|c|c|}
\hline
 & x_1 & x_2 & x_3 & x_4 & x_5 & x_6 \\
\hline
x_1 & 1 & 0.2 & 1 & 0.6 & 0.2 & 0.6 \\
x_2 & 0.2 & 1 & 0.2 & 0.2 & 0.8 & 0.2 \\
x_3 & 1 & 0.2 & 1 & 0.6 & 0.2 & 0.6 \\
x_4 & 0.6 & 0.2 & 0.6 & 1 & 0.2 & 0.8 \\
x_5 & 0.2 & 0.8 & 0.2 & 0.2 & 1 & 0.2 \\
x_6 & 0.6 & 0.2 & 0.6 & 0.8 & 0.2 & 1 \\
\hline
\end{array}
\]

is analyzed as \( S = \sum_{\alpha=0.2}^{1} \alpha E_{\alpha} \) or

\[
S = 0.2\{(x_1, x_2, x_3, x_4, x_5, x_6)\} + 0.6\{(x_1, x_3, x_4, x_6), (x_2, x_5)\} \\
+ 0.8\{(x_1, x_3), (x_4, x_6), (x_2, x_5)\} + 1\{(x_1, x_3), (x_4), (x_6), (x_2), (x_5)\}
\]

Conversely, the above formula represents the resolution of the similarity relation \( S \) in \( X \). As a generalization of the concept of an equivalence relation, for a universe of discourse \( X \), and \( x, y, z \) in the domain of \( S \), the following axioms are satisfied:

(a) \( \mu_s(x, x) = 1 \).
(b) \( \mu_s(x, y) = \mu_s(y, x) \).
(c) \( \mu_s(x, z) \geq \max_y \{(\min \mu_s(x, y), \mu_s(y, z))\} \).

defining \( S \) as a reflexive, symmetric, and maxmin transitive relation.

Although possible uses of the above concepts are foreseen in this paper, the definitions are not motivated by any specific need to model human decision making.

Later on, the similarity problem evolved further in connection with the reverse problem, i.e., that of finding a fuzzy set \( F \), given another fuzzy set \( G \) and a relation \( R(F, G) \) linking them. Common characteristics in the reported solutions include the selection of one of the classical fuzzy set theoretic operators (i.e., \( \min, \max \), etc.), and proposals for families of fuzzy sets rather than a single solution, often determined by rather ad hoc requirements for the precision of a pointwise match between the compared fuzzy sets.

In the work of Pappis and Karacapilidis,\textsuperscript{13,14} the general philosophy is to express the similarity of two fuzzy sets as a function of their proximity. That is, given a small nonnegative number \( \varepsilon \in [0, 1] \), two fuzzy sets \( F \) and \( G \) are considered approximately equal iff for some similarity measure \( S \), \( S(F, G) \leq \varepsilon \) holds. In this framework, three such measures are proposed. The first, based on the operators of union and intersection for fuzzy sets, is expressed as the ratio of the cardinality of their intersection on the cardinality of their union. Formally,
for two sets $F$ and $G$ and their membership function vectors $\{\mu_F\}$ and $\{\mu_G\}$ ($i = 1, \ldots, n$), their similarity $M(F, G)$ is given by

$$M(F, G) = \frac{|F \cap G|}{|F \cup G|} = \frac{\sum_{i=1}^{n} (\mu_F^i \land \mu_G^i)}{\sum_{i=1}^{n} (\mu_F^i \lor \mu_G^i)}$$

This measure is symmetric ($M(F, G) = M(G, F)$), takes into account the common area between the fuzzy sets, and the self similarity for a fuzzy set equals one ($M(F, F) = 1$).

The second similarity measure $L$ is based on the maximum difference between the values of the membership functions of the two fuzzy sets, namely,

$$L(F, G) = 1 - \max_i (|\mu_F^i - \mu_G^i|)$$

This practically implies that the degree of similarity of the two sets is determined by the maximum difference between their membership functions. This measure is also symmetric ($L(F, G) = L(G, F)$), assigns 1 to the similarity of a fuzzy set to itself ($L(F, F) = 1$), and takes into account the common area between the compared fuzzy sets in an implicit way.

The third measure is based on the differences and sums of the membership functions:

$$S(F, G) = 1 - \frac{\sum_{i=1}^{n} |\mu_F^i - \mu_G^i|}{\sum_{i=1}^{n} (\mu_F^i + \mu_G^i)}$$

is also symmetric, and assigns one to self similarity.

These measures offer different alternatives to the pointwise equality approximation for two fuzzy sets in the sense that, on the basis of the properties they exhibit (e.g., symmetry, assignment of full similarity to a fuzzy set and itself, monotonic behavior for fuzzy set intersection or union, etc.), they can be used for the solution of different requirements and classes of inverse problems.

Later, Chen et al.\textsuperscript{15} proposed another set of similarity measures in the same spirit as Pappis and Karacapilidis, offering further choices in the selection of an appropriate measure for different problems.

The first of these is a distance metric similarity measure $W(F, G)$ in which, for two fuzzy sets $F$ and $G$,

$$W(F, G) = 1 - \frac{\sum_{i=1}^{n} |\mu_F^i - \mu_G^i|}{n}$$

The second is a set-theoretic measure $T(F, G)$ based on the supremum of the membership function of the intersection area of the two fuzzy sets:

$$T(F, G) = \sup_{x \in U} \mu_F \cap G(x)$$
And the third measure $P(F, G)$ is based on a matching function and is defined as

$$P(F, G) = \frac{\sum_{i=1}^{n} \mu_{F_i} \mu_{G_i}}{\operatorname{max}(\sum_{i=1}^{n} \mu_{F_i}, \sum_{i=1}^{n} \mu_{G_i})}$$

All of these similarity measures are symmetric, and except for the T measure, they assign 1 to the self-similarity of a fuzzy set.

These measures were criticized, though, by Wang et al. as counterintuitive. More specifically, for the measure $M$, and for two fuzzy sets $F$ and $G$,

$$M(F, G) = \begin{cases} 1 & \text{if } F \cap G = \emptyset, \text{ then } M(F, G) = 0 \\ \frac{\sum_{i=1}^{n} \min(\mu_{F_i}, \mu_{G_i})}{\sum_{i=1}^{n} \max(\mu_{F_i}, \mu_{G_i})} & \text{otherwise} \end{cases}$$

which means that for every $\epsilon \in [0, 1]$ $M(F, G) \leq \epsilon$, and therefore $F$ and $G$ are considered approximately equal. Additionally, in the case where

$$F = G, \quad M(F, G) = 1,$$

but for all $\epsilon \in [0, 1)$, the inequality $M(F, G) \leq \epsilon$ will not hold, leading to counterintuitive conclusions. To remedy this, the authors proposed a reformulated version of the previous measures, in which two fuzzy sets are considered equal to degree $a$ with respect to $M$ (denoted $F \approx_{a} G$):

$$M(F, G) \geq a \text{ for } a \in [0, 1],$$

with the assumption that the higher the $a$, the more similar the fuzzy sets are. In an analogous way, formulas are defined with respect to the other measures $L$ and $S$.

In the same paper, a new family of similarity measures is also introduced, viewing the notion of similarity as the degree of “sameness” for the two fuzzy sets involved. This notion is based on the work of Bandler and Kohout, in which the fuzzy set subsethood is considered through an extension of the equivalence of sets in the crisp case. In classic set theory, for two crisp sets $A$ and $B$, the following holds:

$$A = B \iff A \subseteq B \quad \text{and} \quad B \subseteq A$$

This notion is then extended to fuzzy sets through an implication operator. More specifically, fuzzy set $F$ is a subset of $G$ to the degree given by

$$\inf_{x \in X} I(\mu F(x), \mu G(x))$$

where $I$ is an implication operator. Thus the degree to which $F$ is equal to $G$ is given by

$$F \approx_{a} G = \min\left\{ \inf_{x \in X} I(\mu F(x), \mu G(x)), \inf_{x \in X} I(\mu G(x), \mu F(x)) \right\}$$
In this way families of similarity measures can be generated, corresponding to different implication operators, and generally to different t-norms.

In the framework of fuzzy expert systems and case-based reasoning, Bonissone and Ayub\textsuperscript{18} determine the similarity between two objects as an aggregation of the degrees of similarity between the objects’ features, expressed as fuzzy sets. Each fuzzy set $F$ is represented by a set of four parameters $F = (a, b, \alpha, \beta)$ over the domain of its definition. The first two points $a, b$ indicate the part of the support set in which the membership function is 1, and the third and fourth parameters $(\alpha, \beta)$ denote the left and right spread of the distribution; the membership function itself is assumed to be piecewise linear. The similarity between two fuzzy sets is computed as the complement of their distance, calculated from the four points that characterize each fuzzy set.

For instance, the degree of match of fuzzy sets $X = (a, b, \alpha, \beta)$ and $Y = (c, d, \gamma, \delta)$ is represented by the fuzzy set

$$M_{X,Y} = 1 - |X - Y| = 1 - (|a - d|, |b - c|, \alpha + \beta - \gamma - \delta)$$

This is further matched against a label, from a predetermined labels list, describing different degrees of match (e.g., low, very low, etc.). After this process, the overall similarity for the two fuzzy objects results from a hierarchical aggregation of their features’ similarities, according to a semantic taxonomy based on t-norms.

In the work of Gasos and Ralescu,\textsuperscript{19} the motivation is the comparison between data models stored in a knowledge base, and user-defined descriptions. In this context similarity is defined in relation to the closeness of meanings revealed by the compared fuzzy sets, which in turn is a function of their degree of match. Formally the authors propose the following functional format for the match $M(F, G)$ of two fuzzy sets and therefore for their similarity:

$$M(F, G) = \frac{A_F + A_G}{2A_F A_G} \left(1 - K(FZ_F)^2\right) \left(1 - K(FZ_G)^2\right)$$

where $F, G$ are the compared fuzzy sets; $A_F$ and $A_G$ denote the areas covered by fuzzy sets $F$ and $G$, respectively; $A_{FG}$ is the area of their intersection; $K$ is a constant, indicating the importance assigned to the fuzziness; and $FZ_F$ and $FZ_G$ are the fuzziness of fuzzy sets $F$ and $G$, expressed as the ratio of the spread of the support of the fuzzy set over the whole universe of discourse. In Ref. 20 an analogical reasoning method is proposed as an alternative to the generalized modus ponens rule of inference, initially proposed by Zadeh.

According to this approach, a rule in a reasoning process will be activated if the similarity between the (fuzzy) predicate values in the rule premises and the facts exceeds a certain threshold. In this case the same similarity measure will be used to construct a modification function will deduce the conclusion by modifying the consequent part of the rule. More specifically, given the rule $R: P \to Q$ and the observation (fact) $P'$, the basic idea is to modify the


consequent $Q$ according to the closeness of $P$ to $P'$. If this closeness (similarity) exceeds a predetermined threshold $\tau_0$, the rule will be fired, and the consequent will be adapted through a modification technique, to yield the final conclusion $Q'$. The similarity measure used is distance based, and for two fuzzy sets $F$ and $G$ is defined as

$$S(F, G) = \frac{1}{1 + D(F, G)}$$

where $D(F, G)$ is, in this case, the Euclidean distance generalized for fuzzy sets as

$$D(F, G) = \left( \frac{1}{n} \sum_{i=1}^{n} \left[ \mu_F(x_i) - \mu_G(x_i) \right]^2 \right)^{1/2}$$

A model that attempts to account for asymmetric similarity data is described in Ref. 21 as representing a fuzzified version of Tversky’s contrast model, and the comparison of two objects is performed in terms of their features, considered to be fuzzy sets. Therefore the similarity of object $S_k_i$ to object $S_k_j$ is calculated as

$$S(S_i, S_j) = \sum_k \omega_k \left( \alpha \min(\mu_i(k), \mu_j(k)) - \beta \max(\mu_i(k) - \mu_j(k), 0) - \gamma \max(\mu_j(k) - \mu_i(k), 0) \right)$$

where $k = 1, \ldots, t$ is the set of features and $\omega_k$ is a weighting factor to model the different effect that different features might have for each stimulus.

### IV. A SIMILARITY ORDERING OF FUZZY SETS
#### BASED ON A GENERALIZATION PROCESS

#### A. Motivation

The present proposal for similarity measurement has been motivated by the following considerations:

- A fuzzy set should be considered as a construct more complex than a mere distribution. The values attributed to its main characteristics (e.g., position of the apex, coverage of its support set, fuzziness, etc.) play a critical role in its semantic interpretation. This is especially true in the case where fuzzy set theory is used to model or support the decision-making process.
- In this context, a pointwise distance metric-based comparison of fuzzy sets may only incidentally agree with human-like processes of similarity judgment. Even then, its intuitive interpretation is not immediately clear.
- Similarity—as becomes apparent from the previous exposition—is a very flexible and multifaceted concept. That could, perhaps, support the claim that there should be more than one kind of similarity. It seems correct, though, to assume

\[\text{\textsuperscript{‡}}\text{The threshold } \tau_0 \in [0, 1], \text{ and the closer it is to } 1, \text{ the higher the similarity between the two fuzzy values.}\]
that, within the broad category of learning, updating, generalizing, etc. processes, a notion of similarity as a conceptual proximity among concepts is justifiable. If, for example, it is easier for an English native speaker to learn an unseen French word than an unseen Greek word (because of the common Latin alphabet), then it is reasonable to claim that these two new words are situated in a different conceptual distance from the existing knowledge of the learner. One way of measuring this distance could be through determining the effort needed until the new concept is fully learned. This effort can be represented by the necessary number of repetitions of the new piece of evidence.

In the sequel, a new algorithm will be presented, which expands a discrete fuzzy set in a model or rule, given new evidence in the form of another fuzzy set. After a brief justification for its introduction, a mathematical formalism will be described that models a notion of similarity for fuzzy sets, along the lines discussed above. An important part of every learning and generally extending procedure is the generalization process by which existing information is augmented because of new data. Within fuzzy sets theory, the process of extension is usually treated as part of a learning or aggregation algorithm. The operators most frequently used are max and min or the algebraic product and sum, in the case where the process is described in terms of logical connectives. It is apparent, though, that the max operator represents an overoptimistic attitude and subsequently introduces unnecessary uncertainty by generating wide fuzzy sets. Furthermore, as Zimmermann and Zysno have experimentally demonstrated, in general the above operators

(a) Do not demonstrate a balanced compensatory behavior, being either fully compensatory (max) or not compensatory at all (min).
(b) More importantly—an aspect that is central in the present work—they do not model the human processes involved in merging concepts. In these, humans use nonverbal (latent) connectives that, when asked to verbalize, they map into the words and or, but these mathematical operators represent only approximations of the real merging operators used.

Zimmermann and Zysno subsequently proposed a new parametric aggregation operator \( \theta \), which, for two (or more) fuzzy sets \( F \) and \( G \), will produce an aggregate fuzzy set \( F \theta G \) defined by the membership function

\[
\mu_{F \theta G} = \mu_{1-\gamma} F \cap G \cdot \mu_{\gamma} F \cup G
\]

or analytically,

\[
\mu_{F \theta G} = \left( \prod_{i=1}^{m} \mu_{i} \right)^{1-\gamma} \left( 1 - \prod_{i=1}^{m} (1 - \mu_{i}) \right)^{\gamma}
\]

where the algebraic product and sum are used as intersection and union operators, respectively, \( i = 1, \ldots, m \) is the number of sets to be connected, and \( 0 \leq \gamma \leq 1 \) is the compensation parameter. In the limit case where \( \gamma = 0 \) (no compensation), the aggregation operator becomes the product, and when \( \gamma = 1 \) (full compensation), it becomes the algebraic sum.
This operator offers a more sophisticated way to aggregate concepts, but the selection of the appropriate compensation parameter \( \gamma \) is not always straightforward, and it is highly dependent on the data sample. More importantly, it cannot model generalization, since if an element has zero membership in any of the aggregated fuzzy sets, i.e.,

\[
\text{if any } \mu_i = 0, \text{ then } \prod_{i=1}^{m} \mu_i = 0 \text{ and therefore } \mu_{FG} = 0
\]

This results in assigning a zero membership for that element to the aggregate fuzzy set, irrespective of its membership values in the remaining sets.

In the present work, generalization is considered to be a process by which existing knowledge structures are expanded in the light of new information. More formally, consider a space \( X \), and a knowledge representation scheme that defines a class \( C \) based on a property \( s \), i.e.,

\[
C = \{ x : x \in X \text{ and } s(x) \text{ is true} \}
\]

A generalization process \( g \) will extend \( C \) to \( C_g \), so that the domain of property \( s \) will be extended to include new elements, that is,

\[
\exists y \in X : y \notin C \text{ and } y \in C_g
\]

For the fuzzy sets case, let \( X \) be the space of definition of a fuzzy set \( F \) in a model or rule, with support set \( S_F = \{ x \in X : \mu_F(x) \neq 0 \} \). Suppose further that \( G \) represents a data fuzzy set with a support set \( S_G = \{ x \in X : \mu_G(x) \neq 0 \} \). Let \( T \) denote the set of elements in \( X \) that belongs to the support set of \( G \), but not to the support set of \( F \), and therefore have zero membership in the latter, i.e.,

\[
T = \{ x \in X : x \in S_G \text{ and } x \notin S_F \}
\]

or equivalently,

\[
T = \{ x \in X : \mu_G(x) \neq 0 \text{ and } \mu_F(x) = 0 \}
\]

The generalization algorithm will then extend \( F \) to a new fuzzy set \( F_1 \) containing all of the elements in \( S_F \), plus some elements from \( T \), i.e., a nonempty subset of \( T \).

Formally the support for the extended fuzzy set \( F_1 \) will be

\[
S_{F_1} = S_F \cup T' \quad \text{where } T' \subseteq T \text{ and } T' \neq \emptyset
\]

The proposed generalization algorithm is described next.

**B. The Generalization Algorithm**

Given a space \( M \) and a concept represented by the discrete fuzzy set \( m \in M \), a generalization process \( ff : M \times M \rightarrow M \) is defined, so that, given the
concept \( m \) and a new fuzzy datum \( d \in M \), it will generate a new concept \( m_i = f(m, d), m_i \in M \). This new concept will be a discrete fuzzy set \( m_i \) more general than \( m \) in the sense described earlier.§

The generalization is performed through the following steps:

(a) Given two discrete fuzzy sets \( F \) (model) and \( G \) (data), let \( F_i, i = 1, \ldots, n \) and \( G_j, j = 1, \ldots, m \) be their respective decompositions into \( \alpha \)-cuts, in order of increasing cardinality (i.e., \( F_1 \subset F_2 \subset \cdots \subset F_n \) and \( G_1 \subset G_2 \subset \cdots \subset G_m \)).

These \( \alpha \)-cuts can also be considered as the focal elements of a possibility distribution induced by the fuzzy set \( F \), accompanied by the mass assigned to each of them. 23, 25

(b) These focal elements of the two fuzzy sets are then aligned, so that the smallest focal element of \( G_j \), \( G_j \), will correspond to that focal element \( F_i \) of the model, as in 1 above and generally \( G_{m_k} \) will be aligned with \( F_{k+m-1} \).

(c) The resulting generalized fuzzy set will have the \( F_1, F_2, \ldots, F_{k} \) focal elements of the old fuzzy model, plus the union of the aligned focal elements, i.e.,

\[
F \oplus G = \{F_1, \ldots, F_{k-1}, F_k \cup G_1, F_{k+1} \cup G_2, \ldots, F_n \cup G_m\}
\]

while the masses associated with each focal element will be those of the initial fuzzy model \( F \).

In the case where \( m > n \), the last focal element of \( F \) will be replaced by \( m - n \) equal focal elements, but with the initial mass equally distributed among them. This results in a conservative extension of the model.

A formal description of the algorithm follows.

ALGORITHM (For Fuzzy Model Generalization). Let \( F \) be a discrete fuzzy set representing a model or prior knowledge, and \( G \) a discrete fuzzy set representing new evidence (data). These fuzzy sets induce a possibility distribution, where \( F_1, F_2, \ldots, F_n \) denote the focal elements of \( F \), and denote \( G_1, G_2, \ldots, G_m \) the focal elements of \( G \), and both are ordered by increasing cardinality (i.e., \( F_1 \subset F_2 \subset \cdots \subset F_n \) and \( G_1 \subset G_2 \subset \cdots \subset G_m \)).

(i) Generalization of the set

1. Find index \( k, 1 \leq k \leq n \), such that \( G_1 \subset F_{k+1} \) and \( G_1 \subset F_k \), where \( G_1 \) is the focal element corresponding to the 1-cut of \( G \).

2. Align the two lists of focal elements so that \( G_1 \) will be aligned with \( F_k \) (as in 1 above) and generally \( G_{m_k} \) will be aligned with \( F_{k+m-1} \).

3. The generalized fuzzy set will be described in terms of its focal elements as

\[
F \oplus G = \{F_1, \ldots, F_{k-1}, F_k \cup G_1, F_{k+1} \cup G_2, \ldots, F_n \cup G_m\}
\]

where the masses associated with them are the ones corresponding to the \( F_1 \) part (see Examples).

(ii) Membership of an element \( x \in X \) in the generalized \( F \oplus G \) fuzzy set

\( iia \) \( x \in F_i \) for \( 1 \leq i \leq n \) and \( x \not\in G_j \) for all \( j \). Then \( \mu_{F \oplus G}(x) = \mu_x(x) \).

§Fuzzy set \( m \) is also a subset of \( m_i \), according to Zadeh’s proposal, i.e., \( \forall x \in X, \mu_m(x) \leq \mu_{m_i}(x) \).
(ii) \( x \in F_i \) for all \( i \) and \( x \in G_{i+1}, \ldots, G_m \) for some \( p \). This \( G_p \) focal element corresponds to the model focal element \( F_q \), where \( q = k + p - 1 \). Then

\[
\mu_{F \oplus G}(x) = \sum_{i=q}^{n} m_{F_i}
\]

(iii) \( x \in F_i \) for \( i \geq k \), and \( x \in G_m \) for some \( m \) (which means that \( x \in F_k \), for some \( k' \in \{k, k+1, k+2, \ldots, n\} \)). Then find \( m^* = \min m \) such that \( x \in G_{m^*} \). This \( G_{m^*} \) focal element corresponds to the model focal element \( F_{k'} \), where \( k_{m^*} = k + m^* - 1 \). Therefore, \( x \in (F \oplus G)_{k'f} \), where \( kf = \min(k', k_{m^*}) \). Then

\[
\mu_{F \oplus G}(x) = \sum_{i=kf}^{n} M_{F_i}
\]

where \( m_{F_i} = m(F \oplus G)i \) is the mass associated with the focal element \( F_i \).

C. Examples

Let

\[ Model = \{10/0.1, 20/0.4, 30/0.8, 40/1, 50/0.7, 60/0.3, 70/0.2\} \]

be a fuzzy set in a rule (representing the existing value for a concept model or prior knowledge), and let

\[ Data1 = \{10/0.1, 20/0.2, 30/0.6, 40/0.8, 50/1, 60/0.6, 70/0.3\} \]

be a value in the data, representing new evidence. The aim is to update (generalize) the set \( Model \), in the face of the new evidence represented by \( Data1 \). The two fuzzy sets can be decomposed into their nested focal elements and the corresponding masses and subsequently aligned (according to the previously described algorithm) as follows:

\[ M1 = \{40\} : 0.2 \]
\[ M2 = \{40, 30\} : 0.1 \]
\[ M3 = \{40, 30, 50\} : 0.3 \]
\[ M4 = \{40, 30, 50, 20\} : 0.1 \]
\[ M5 = \{40, 30, 50, 20, 60\} : 0.1 \]
\[ M6 = \{40, 30, 50, 20, 60, 70\} : 0.1 \]
\[ M7 = \{40, 30, 50, 20, 60, 70, 10\} : 0.1 \]
\[ D_1 = \{ 50 \} : 0.2 \]
\[ D_2 = \{ 50, 40 \} : 0.2 \]
\[ D_3 = \{ 50, 40, 30, 60 \} : 0.3 \]
\[ D_4 = \{ 50, 40, 30, 60, 70 \} : 0.1 \]
\[ D_5 = \{ 50, 40, 30, 60, 70, 20 \} : 0.1 \]
\[ D_6 = \{ 50, 40, 30, 60, 70, 20, 10 \} : 0.1 \]

\[ \text{Data1} = \]

Data where the index \( k = 2 \), so that \( D_1 \subseteq M_3 \) and \( D_1 \not\subseteq M_2 \) (where \( D_1 = \{ 50 \} \) is the focal elements of Data1 for \( \alpha = 1 \)). Therefore the generalized fuzzy set Model_11 will have the following series of focal elements and masses assigned to it:

\[ \{ 40 \} : 0.2 \]
\[ \{ 40, 30 \} : 0.1 \]
\[ \{ 40, 30, 50 \} : 0.3 \]

\[ \text{Model}_1 = \{ 40, 30, 50, 20 \} : 0.1 \]
\[ \{ 40, 30, 50, 20, 60 \} : 0.1 \]
\[ \{ 40, 30, 50, 20, 60, 70 \} : 0.1 \]
\[ \{ 40, 30, 50, 20, 60, 70, 10 \} : 0.1 \]

or after merging equal focal elements,

\[ \{ 40 \} : 0.2 \]
\[ \{ 40, 30, 50 \} : 0.4 \]

\[ \text{Model}_1 = \{ 40, 30, 50, 60, 20 \} : 0.1 \]
\[ \{ 40, 30, 50, 60, 20, 70 \} : 0.1 \]
\[ \{ 40, 30, 50, 60, 20, 70 \} : 0.1 \]

which corresponds to the fuzzy set

\[ \text{Model}_1 = \{ 10/0.1, 20/0.4, 30/0.8, 40/1, 50/0.8, 60/0.4, 70/0.3 \} \text{ (Fig. 1)} \]
If now the modified model $Model_{11}$ is presented again with the same evidence (fuzzy set $Data_1$), it will be updated to a new model set $Model_{12}$, having the following nested set of $\alpha$-cuts (or focal elements):

$$
\begin{align*}
\{40, 50\} & : 0.2 \\
\{40, 30, 50\} & : 0.4 \\
Model_{12} = \{40, 30, 50, 60, 20\} & : 0.1 \\
\{40, 30, 50, 60, 20, 70\} & : 0.2 \\
\{40, 30, 50, 60, 70, 20, 10\} & : 0.1
\end{align*}
$$

corresponding to

$Model_{12} = \{10/0.1, 20/0.4, 30/0.8, 40/1, 50/1, 60/0.4, 70/0.3\}$ (Fig. 1)

If again the same piece of evidence ($Data_1$) is presented to $Model_{12}$, no further modification will occur, indicating that the generalization process has
converged. The number of repetitions until convergence state is reached, is equal to the index \( k \) (in this particular case \( k = 2 \)), and depends on the relative position of the two fuzzy sets; that is, the closer they are, the smaller the number of repetitions will be.

If, for example, the initial model fuzzy set \( Model \) is presented with the new data set

\[
Data_2 = \{40/0.1, 50/0.2, 60/0.6, 70/1, 80/0.8, 90/0.1\}
\]

then five repetitions will be needed, until convergence is reached, corresponding to the final fuzzy set

\[
Model_{25} = \{10/0.1, 20/0.4, 30/0.8, 40/1, 50/0.7, 60/0.7, 70/1, 80/0.8, 90/0.1\}
\]

The intermediate and final stages of the generalization in that case are shown in Figure 2.

### D. Some Characteristics of the Generalization Algorithm

- The algorithm provides a gradual updating mechanism that is not dependent on a parameter, therefore avoiding changing the problem of generalizing into the problem of selecting an appropriate parameter.
- It preserves the normality of fuzzy sets and the nested structure for the focal elements of the associated possibility distribution.
- If a model (referent) fuzzy set is presented with an identical fuzzy set as new evidence, no generalization will occur.
- The algorithm takes into account the relative position of the two fuzzy sets, therefore in extending the model fuzzy set, it will not introduce unjustified optimism and uncertainty (as in the case of the max operator).
- It has a set theoretic basis rather than a pointwise one, since it proceeds by taking the union of focal elements. Moreover, there is an intuitive aspect to it, since it expands the model by expanding its constituent parts, and the speed of generalizing is inversely related to the distance between the prior knowledge model fuzzy set and the new evidence (data fuzzy set).
- It converges after a (usually small) number of repetitions, simulating a “déjà vu” behavior. The resulting fuzzy set in the convergence state is very close to the one resulting by applying the max operator to the initial sets.

### E. A Similarity Ordering of Fuzzy Sets

In this section a new similarity ordering of fuzzy sets is described, according to which, given a referent fuzzy set (in a rule or a model), and a subject fuzzy set, their similarity is proportional to the effort needed until the referent fuzzy set is fully generalized, given the subject fuzzy set as new evidence.

More specifically, given an initial discrete fuzzy set \( M \), and a data fuzzy set \( D \), the generalization algorithm described in the previous section will generate a series of expanded models \( M_i \), each \( M_i \) being the result of the \( i \)th repetition of the data set \( D \) in a learning process. At some step \( k \), this learning will converge,
Figure 2. Generalization of fuzzy set Model, using fuzzy set Data2 as new evidence.
i.e.,

\[ \exists k: M \neq M_1 \neq M_2 \neq \cdots \neq M_k = M_{k+1} = M_{k+2} = \cdots \]

This number \( k \) can be considered a similarity index for the concepts represented by fuzzy sets \( M \) and \( D \), measuring in a way the “amount of effort” or the number of necessary lessons (repetitions, iterations) needed for an agent with \( M \) as prior knowledge to fully learn the new concept \( D \).

Furthermore, for concepts \( M, D_1, D_2 \), if by \( S(M, D_1) \) and \( S(M, D_2) \) we denote their similarity, then

\[ S(M, D_1) \geq S(M, D_2) \quad \text{iff} \quad k_1 \leq k_2 \]

where \( k_1, k_2 \) represent the number of times \( M \) has to be expanded, given \( D_1 \) and \( D_2 \), respectively, until convergence (final generalization) is reached. The meaning of this is that the similarity of the two concepts is inversely proportional to the number of repetitions.
A related concept for the measurement of similarity is proposed in the Perceptron model, in which for two entities $Y$ and $X$, their similarity is defined as the shortest path by which $Y$ is transformed into $X$. This similarity ordering, though, is sensitive only to the relative position of the apexes of the two fuzzy sets, and cannot capture the changes in the spread of the support set that occur because of the expansion process. For example, fuzzy sets $Data_a$ and $Data_b$ in Figure 3 will be considered equally similar to fuzzy set $Model$, based on the previous index.

This modification in the support set of the referent fuzzy set can be detected by a second similarity index described in the sequel.

Given a fuzzy set

$$F = x_1/\mu_F(x_1), x_2/\mu_F(x_2), \ldots, x_n/\mu_F(x_n)$$

defined on a discrete space $X = \{x_1, x_2, \ldots, x_n\}$, a possibility distribution can be induced, namely, $\pi(x_i) = \mu_F(x_i), i = 1, \ldots, n$. Over the corresponding series of the nested focal elements $F_1 \subset F_2 \subset \cdots \subset F_n$, a probability mass allocation $m(F_i)$ can be generated from the relations

$$\mu_F(x_i) = \pi(\{x_i\}) = \sum_{F_i \cap \{x_i\} \neq \emptyset} m(F_i)$$

for all $x_i$ in the support set of $F$ or after solving the set of equations

$$m(F_i) = \mu_F(x_i) - \mu_F(x_i + 1)$$

for $i = 1, \ldots, n$ and $\mu_F(x_n + 1) = 0$ by convention. Subsequently, a probability distribution on the support set of $F$ can be defined, by assuming a uniform distribution of the mass associated with each focal element among its members. Specifically, for a fuzzy set $F$ and for each $x$ in its support set, the probability

![Figure 3](https://via.placeholder.com/150)  
Figure 3. Fuzzy sets with different support sets.
associated with it is

$$Pr(F(x)) = \sum_{x \in F_i} \frac{mF_i}{|F_i|}$$

where $F_i$, $i = 1, \ldots, n$ are the focal elements of the fuzzy set $F$, $mF_i$ is the mass assigned to the $i$th focal element $F_i$, and $|F_i|$ is its cardinality.

Given now a fuzzy set $M$ in a model, a data fuzzy set $D$, and the series of their focal elements $\{M_1, M_2, \ldots, M_m\}$, and $\{D_1, D_2, \ldots, D_m\}$, respectively, let $x_1$ be the element with full membership in $M$, i.e., such that $\mu_M(x_1) = 1$. The probability associated with that element, before any generalization takes place, is

$$Pr_M(x_1)_{\text{Before}} = \sum_{i=1}^{n} \frac{mM_i}{|M_i|}$$

If $M \oplus D$ is the generalized fuzzy model at the convergence state, the probability assigned to the same element $x_1$ will now be

$$Pr_{M \oplus D}(x_1)_{\text{After}} = \sum_{p=1}^{t} \frac{m_{M \oplus D_p}}{|M \oplus D_p|}$$

where $M \oplus D_p$, $p = 1, \ldots, t$ are the focal elements (or equivalently $\alpha$-cuts) of the generalized fuzzy set.

The amount

$$Pr_M(x_1)_{\text{Before}} - Pr_{M \oplus D}(x_1)_{\text{After}}$$

representing the reduction of the probability of the peak element $x_1$, provides a second index for a similarity ordering of the two sets, which complements the previous index, in the sense that it “measures” the changes in the spread of the model’s support set, after the generalization process.

For the two cases described in the Examples section, the reduction of the probability of each element in each expanded version of the model is graphically represented in Figures 4 and 5.

V. COMPARATIVE ANALYSIS

One of the few attempts at a comparative analysis of different distance-based similarity measures for fuzzy sets is reported in Ref. 28. The validation has been performed through a behavioral experiment, in which subjects assessed the similarity of fuzzy concepts (in a linguistic context), and their ratings were subsequently compared to the degrees of similarity provided by the distance metrics.

In the present work, a somehow different approach is adopted for the validation of the different measures, and this is their evaluation from a decision maker’s perspective. The central assumption is that if a method for comparison of fuzzy sets is used to model or support the decision-making process of humans, then the similarity measurement has to take into account important characteristics of the involved sets, such as the relative position of their apexes, the
Figure 4. Probability distributions for initial and generalized fuzzy set (first case).

proportion of their overlapping parts, etc. In this context, the treatment of similarity of fuzzy sets through geometric distance metrics, evaluated pointwise, cannot consistently satisfy these intuitive requirements. This can be attributed mainly to the fact that these measures are motivated by the solution of the inverse problem related to fuzzy relations, and are therefore primarily concerned with the satisfaction of the necessary mathematical constraints and properties (e.g., preservation of the similarity orderings of the sets after the operations of union and intersection).

As an illustration of the above analysis, the similarity of five fuzzy sets pairs is evaluated, using seven metric distance-based similarity measures reported in the literature.13, 15, 16 The first set in all five pairs is kept the same, to demonstrate the way changes in the second set affect the degree of similarity between the members of the pair. All of the measures are chosen to belong to the geometric distance metrics, since these represent the dominating tendency.
There exist other recent proposals, though, that handle similarity in a more sophisticated and global way, as in Refs. 19, 21, and 29, and the second family of measures proposed in Ref. 16. The validation of these measures, though, is less straightforward, since additional information about the compared objects is required, such as the definition of an implication operator, or the set of features defining the fuzzy concepts (for the measures proposed in Refs. 21 and 29). The same family of fuzzy set pairs is subsequently evaluated through the two similarity indexes proposed in Section IV, and the results are shown in Tables I and II.

In the work of Pappis and Karacapilidis, the concept of similarity of two fuzzy sets is related to the concept of approximate equality of the sets. In that sense, two fuzzy sets $F$ and $G$ are approximately equal, and therefore similar, if for some distance metric $D$ and a threshold $\varepsilon \in [0, 1]$ $D(F, G) \leq \varepsilon$. Three such metrics are proposed (see Section III). The first of them, $M$, assigns the same similarity to the pairs $(Model, Data1)$ and $(Model, Data1a)$ in the sense that
Table I. Similarity measures for fuzzy sets (distance metric based). $F = Model, G = \{Data, Data_{1a}, Data_{2}, Data_{5}\}$.

<table>
<thead>
<tr>
<th>Proposed by</th>
<th>$M(F, G)$</th>
<th>$L(F, G)$</th>
<th>$S(F, G)$</th>
<th>$W(F, G)$</th>
<th>$P(F, G)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pappis and Karacapilidis</td>
<td>1</td>
<td>0.69</td>
<td>0.69</td>
<td>0.145</td>
<td>0.140</td>
</tr>
<tr>
<td>Pappis and Karacapilidis</td>
<td>1</td>
<td>0.7</td>
<td>0.3</td>
<td>0.1</td>
<td>0</td>
</tr>
<tr>
<td>Pappis and Karacapilidis</td>
<td>1</td>
<td>0.187</td>
<td>0.187</td>
<td>0.254</td>
<td>0.247</td>
</tr>
<tr>
<td>Wang, DeBaets and Kerre</td>
<td>1</td>
<td>0.69</td>
<td>0.69</td>
<td>0.145</td>
<td>0.140</td>
</tr>
<tr>
<td>Chen, Yeh and Hsiao</td>
<td>1</td>
<td>0.815</td>
<td>0.815</td>
<td>0.478</td>
<td>0.455</td>
</tr>
<tr>
<td>Chen, Yeh and Hsiao</td>
<td>1</td>
<td>0.8</td>
<td>0.8</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>Chen, Yeh and Hsiao</td>
<td>1</td>
<td>0.924</td>
<td>0.793</td>
<td>0.255</td>
<td>0.238</td>
</tr>
</tbody>
</table>
Table II. Similarity ordering of fuzzy sets based on a generalization algorithm.

<table>
<thead>
<tr>
<th>Model</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model = [10/0.1, 20/0.4, 30/0.8, 40/1, 50/0.7, 60/0.3, 70/0.2]</td>
<td>Data = [10/0.1, 20/0.4, 30/0.8, 40/1, 50/0.7, 60/0.3, 70/0.2]</td>
</tr>
<tr>
<td>Proposed Method</td>
<td>Number of repetitions</td>
</tr>
<tr>
<td>Model = [10/0.1, 20/0.4, 30/0.8, 40/1, 50/0.7, 60/0.3, 70/0.2]</td>
<td>0</td>
</tr>
<tr>
<td>Data = [10/0.1, 20/0.4, 30/0.8, 40/1, 50/0.7, 60/0.3, 70/0.2]</td>
<td>2</td>
</tr>
<tr>
<td>Model = [10/0.1, 20/0.4, 30/0.8, 40/1, 50/0.7, 60/0.3, 70/0.2]</td>
<td>4</td>
</tr>
<tr>
<td>Data = [10/0.1, 20/0.4, 30/0.8, 40/1, 50/0.7, 60/0.3, 70/0.2]</td>
<td>5</td>
</tr>
<tr>
<td>Model = [40/0.1, 50/0.2, 60/0.6, 70/0.8, 80/1, 90/0.3]</td>
<td>7</td>
</tr>
</tbody>
</table>
for every $\varepsilon \leq 0.69$ fuzzy sets, \textit{Data1} and \textit{Data\_1a} are considered approximately equal. These sets are different, though, in more than one aspect.\footnote{For instance, \textit{Data1} could represent the fuzzy number \textit{about 50} and \textit{Data\_1a} the number \textit{about 60}. Moreover, \textit{Data1} is convex, whereas \textit{Data\_1a} is nonconvex, etc.} Furthermore (as noted also in Ref. 16), in the case of identical fuzzy sets as in the (\textit{Model, Data}) pair, the value returned by $M$ is equal to 1, which means that for every $\varepsilon < 1$, the two sets are dissimilar. Also for two disjoint nonempty fuzzy sets $F, G$ $M(F, G) = 0$, implying that $\forall \varepsilon \in [0, 1]$ and they are approximately equal and thus completely similar.

A problem with the second measure, $L$, proposed in the same paper, is illustrated by the cases (\textit{Model, Data2}) and (\textit{Model, Data5}). In these, a small modification in the second fuzzy set (namely the assignment of a higher membership equal to one to the element 80) will result in qualifying the two fuzzy sets as totally dissimilar, despite the fact that many other characteristics (e.g., the overlapping area) remain the same. This phenomenon will appear in any fuzzy sets pair $F, G$, in which for some element $x$, $\mu_F(x) = 1$ and $\mu_G(x) = 0$.

For example, for the sets
\[
\begin{align*}
F &= 10/0.1 + 20/0.2 + 30/0.6 + 40/0.9 + 50/1 + 60/0.8 + 70/0.3 + 80/0.1 \\
G &= 10/0.1 + 20/0.2 + 30/0.6 + 40/0.9 + 50/1 + 60/1 + 70/1 + 80/1 \\
    &\quad + 90/1 + 100/0.1
\end{align*}
\]

their similarity $S(F, G) = 0$, despite the fact that $F$ is included in $G$ (Fig. 6).

In all of the previous measures, the underlying consensus is that each entity, described by a fuzzy set, can be adequately represented in one or more dimensions, which would allow for a pointwise calculation of their similarity. This may be true for some groups (like colors), but for some others (e.g., faces or countries), qualitative features are more appropriate for their description. An additional general criticism of the pure geometric treatment of similarity is...
concerned with the satisfaction, by all of the above measures, of the metric axioms (especially symmetry). This is implicitly assumed in all distance-based metrics, although there exists theoretical and experimental evidence against it.

Concerning the two proposed indexes in Section IV, as can be seen from Table II, they capture the intuitive similarity requirements for the ordering of the sets in the present family. These similarity indexes are not symmetric, and they assign a higher importance to the set to be generalized (considered to be the model or referent), which is in agreement with findings in similarity judgments.3,8,9,30,31

Regarding the first pair in Table II, the assignment to the index of a value equal to 0 has the following meaning. Since similarity is defined through the number of repetitions in the learning process, and the probability reduction of the peak element, for the case of identical fuzzy sets, no learning (generalization) will occur, and therefore no repetition will be required, leaving the probability of the peak element in the model set unchanged.

VI. CONCLUSIONS

The assumptions underlying any similarity measurement play an important role in the validity and performance of models for a variety of processes, ranging from updating and learning to cased-based reasoning and retrieval and induction in general. If these schemes aim to model human beings, in either an explicit or implicit way, then these assumptions should have a degree of psychological relevance, along with the compliance with mathematical requirements.

In the present work the problem of similarity for two fuzzy sets is approached at a higher level, considered to be process dependent. The process is that of updating prior knowledge, which is a form of monotonic learning. In this context a fuzzy set representing a model and a fuzzy set in new evidence have a degree of similarity that is defined by the number of necessary repetitions of the new knowledge until the learning process converges.

A second index of probabilistic character complements the first one and represents the reduction of the probability that the element has full membership in the model fuzzy set before the learning procedure and at the convergence state.

The author is grateful to Prof. J.F. Baldwin for insightful comments and discussions on earlier versions of this work. Thanks are also due to R. Ribeiro for corrections and suggestions on a previous draft of this paper.

References