Similarity-Based Approximate Reasoning: Methodology and Application

Swapan Raha, Nikhil Ranjan Pal, and Kumar Sankar Ray

***Abstract*—This correspondence elucidates the importance of similarity, as modeled by a measure of similarity, in approximate reasoning. A set of axioms is proposed to compute a reasonable measure of similarity between two imprecise concepts represented as fuzzy sets. For approximate reasoning, a similarity index between the fact and the antecedent of a rule is computed and is used in the reasoning mechanism. Accordingly, the existing reasoning mechanism is modified. A new similarity-based approximate reasoning methodology is proposed. As an illustration of its effectiveness, the proposed mechanism is used to develop a rule-based pattern classifier.**

***Index Terms*—Approximate reasoning, fuzzy set, pattern classifier, similarity index.**

I. INTRODUCTION

The similarity between two objects suggests the degree to which properties of one may be inferred from those of the other. Distance functions may be used to define similarity between sets. For our investigation, we generalize the concept of distance function defined for crisp sets to the noncrisp sets. A similarity matching degree s may be defined from the distance function d (with range set in $[0, 1]$) according to the following:

$$s(\cdot, \cdot) = 1 - d(\cdot, \cdot). \tag{1}$$

The concepts of similarity and proximity of fuzzy sets play a fundamental role in reasoning with vague knowledge [1]–[3]. Turksen and Zhong [4] commented and showed that the notion of a similarity measure between two fuzzy sets may be successfully applied in fuzzy reasoning. Recently, similarity-based approximate reasoning mechanisms are being applied to pattern classification [5]. We assume that the universe of discourse is a finite set. Let $A = \sum_{u \in U} \{\mu_A(u)/u\}$ and $B = \sum_{u \in U} \{\mu_B(u)/u\}$ be two fuzzy sets defined over the universe of discourse U. A similarity index between the pair $\{A, B\}$ is denoted as $S(A, B; U)$ or simply $S(A, B)$. We propose two measures of similarity between fuzzy sets and discuss their properties. Given a similarity matching degree, there are different methods of inference based on the same. In [4], the authors proposed a similarity-based method called *approximate analogical reasoning schema*. It was shown that the method is applicable to both point-valued and interval-valued fuzzy sets. In [6], Chen proposed two similar methods for medical diagnosis problems. Several methods based on different modification procedures have been proposed in [7] and [8].

In all these works, similarity-based fuzzy reasoning does not require the construction of a fuzzy relation. It is based on the computation of the degree of similarity between the fact and the antecedent of a rule. Then, based on the similarity value, the membership value of each element of the consequent fuzzy set of the rule is modified to obtain a conclusion.

Manuscript received October 29, 1999; revised November 19, 2000 and January 26, 2002. This paper was recommended by Associate Editor M. Shahidehpour.

S. Raha is with the Department of Mathematics, Visva-Bharati, Santiniketan 731235, grIndia (e-mail: raha@vbharat.ernet.in).

N. R. Pal and K. S. Ray are with the Electronics and Communication Sciences Unit, Indian Statistical Institute, Calcutta 700108, India (e-mail: nikhil@isical.ac.in; ksray@isical.ac.in).

Digital Object Identifier 10.1109/TSMCA.2002.804787

In this paper, we propose two similarity-based approximate reasoning methods. Our first method is a modification of the method presented in [4]. The second method is an integration of similarity-based reasoning and Zadeh's compositional rule of inference. With different results we show that the proposed similarity-based approximate reasoning methods are reasonable.

II. A BRIEF REVIEW

A substantial amount of work has already been done in defining similarity degree between two fuzzy sets. Independent of these developments, researchers engaged in approximate reasoning used similarity measures in reasoning methodology. Accordingly, different similarity-based reasoning techniques have been developed. For clarity, we divide our discussion into two subsections: 1) similarity indexes and 2) similarity-based reasoning methods.

A. Similarity Indexes

A similarity measure, besides being symmetric, should satisfy the following properties [9]: $S(A, B) = 1$ when $A \nabla B = \Phi$; where $\mu_{A \nabla B}(u) = D(\mu_A(u), \mu_B(u))$ and $D(a, b) = \min[\max(a, b), \max(1 - a, 1 - b)] = \max[\min(1 - a, b), \min(a, 1 - b)]; 0 \leq a, b \leq 1$.

The following dissimilarity indexes are proposed in [9]:

$$d_1(A, B) = 1 - \frac{|A \cap B|}{|A \cup B|}$$

$$d_2(A, B) = |A \diamond B|$$

$$d_3(A, B) = \sup_{u \in U} \mu_{A \diamond B}(u)$$

where $\forall u \in U, \mu_{A \diamond B} = \max[\min(\mu_A, 1 - \mu_B), \min(1 - \mu_A, \mu_B)]$.

$S(A, B)$ can also be defined as [6], [10]:

$$S(A, B) = \frac{\sum_{u \in U} \{\mu_A(u)\mu_B(u)\}}{\max\left\{\sum_{u \in U} \mu_A^2(u), \sum_{u \in U} \mu_B^2(u)\right\}}$$

$$S(A, B) = \frac{|A||B| \cos(\theta)}{(\max\{|A|^2, |B|^2\})}$$

where $|A|$ is the length of the vector A and $\cos(\theta)$ is the cosine of the angle between the two vectors.

A family of measures of similarity of fuzzy sets having a strong logical background may be given as follows:

$$S(A, B) = \frac{1}{2}[(A \leftrightarrow B) + (A \leftrightarrow B)]$$

where $(A \leftrightarrow B) = (A \rightarrow B) \wedge (B \rightarrow A)$, $\wedge$ is a conjunction operator and $\rightarrow$ is an implication operator. Different interpretations of the operators will result in different measures of similarity between fuzzy sets. More work on these measures may be found in [11] and [12].

Pappis and Karacapilidis [13] proposed several measures of similarity between fuzzy sets, such as

$$S(A, B) = 1 - \max_{u \in U}(|\mu_A(u) - \mu_B(u)|)$$

and

$$S(A, B) = \frac{\sum_{u \in U} \min(\mu_A(u), \mu_B(u))}{\sum_{u \in U} \max(\mu_A(u), \mu_B(u))}.$$

Wang [14] modified the last measure to produce a new measure as

$$S(A, B) = \frac{1}{n} \sum_{i=1}^{n} \frac{\min(\mu_A(u), \mu_B(u))}{\max(\mu_A(u), \mu_B(u))},$$

Kwang *et al.* proposed another similarity measure, between two finite fuzzy sets as

$$S(A, B) = \max_{u \in U} [\min(\mu_A(u), \mu_B(u))].$$

Properties like reflexivity, symmetry, normalization, boundedness, and dissimilarity between fuzzy sets based on this measure have been studied [15]. Once an index is defined, the following questions arise: "How can we compare this with other existing indexes?" "How should we judge the goodness of such an index?" Questions of this nature carry immense importance for all practical purposes.

All similarity measures considered here satisfy the reflexivity, symmetry, and boundedness properties. These three properties are necessary for any similarity measure. In this regard, all measures work equally well. Besides these three properties, similarity measures should also satisfy properties like computational simplicity, monotonicity, and nondissimilarity.

Similarity measures based on the computation of overall sup-operation give more importance to a particular value and ignore the presence of others. Thus, two widely different fuzzy sets would be measured similar when they have the same supremum.

We should consider those indexes that can play important roles in fuzzy reasoning. This demands similarity measures based on a comparison of membership degrees of elements of each fuzzy set. Thus, the measure defined in [6] is useful in practical situations. A major drawback of the same is that $S(A, B)$ does not, in general, satisfy any monotonicity criterion. Moreover, this measure of similarity does not indicate the measure of nondissimilarity.

It is almost impossible to single out one similarity measure that works well for all purposes. This correspondence intends to provide the user with certain measures of similarity, each of which satisfies certain basic needs of being a measure of similarity.

B. Similarity-Based Reasoning

Many fuzzy systems are based on Zadeh's compositional rule of inference [16]. Despite its success in various systems, researchers have indicated certain drawbacks in the mechanism [4]. This motivated the introduction of similarity-based reasoning mechanisms [4], [6]–[8], [10], [17].

In such similarity-based reasoning schemes, from a given fact, the desired conclusion is derived using only a measure of similarity between the fact and the antecedent, in a rule-based system. In some cases, a threshold value τ is associated with a rule. If the degree of matching between the antecedent of the rule and the given fact exceeds τ, then only that rule is assumed to be fired. The conclusion is derived using some modification procedure. As an illustration, let us consider the two premises as in Table I. Here, A and A' are fuzzy sets defined over the same universe of discourse $U = \{u_1, u_2, \ldots, u_m\}$ and B, B' are defined over the universe of discourse $V = \{v_1, v_2, \ldots, v_n\}$. Let $S(A, A')$ denote some measure of similarity between two fuzzy sets A, A'. If $S(A, A') > \tau$ only then the rule is fired and the consequent of the rule is modified using $s = s'(S(A, A'), \tau)$ to produce the desired conclusion. Different authors used different functions s' [4], [6], [10].

Two types of modification procedures are proposed in [4]

$$\text{Expansion form: } \mu_{B'}(v_i) = \min[1, \mu_B(v_i)/s] \quad (2)$$

$$\text{Reduction form: } \mu_{B'}(v_i) = (\mu_B(v_i) \cdot s). \quad (3)$$

Chen [6], [10] uses a threshold value and the reduction form of inference with s as a measure of similarity. He does not provide any argument regarding the choice of the modification procedure.

Yeung and Tsang [7] use a certainty factor associated with each rule in the modification procedure. The inference is based on the number of propositions in the antecedent of the rule(s) as well as the operator(s) connecting them. In each case, the inference is of expansion type. In [8], Yeung and Tsang presented two more modification procedures and claimed two new fuzzy reasoning methods. One modification is based on Zadeh's inclusion and cardinality measure and the other one is based on equality and cardinality measure. Other operations remain almost identical.

III. Proposed Similarity Measure—Definition and Properties

In order to provide a definition for similarity index, a number of factors must be considered. A primary consideration is that, whatever way we choose to define such an index, it must satisfy the properties already mentioned. We expect that a similarity measure $S(A, B)$ should satisfy the following properties.

For all fuzzy sets A, B:

P1 $S(B, A) = S(A, B)$.

P2 $S(A^c, B^c) = S(A, B)$, A^c being some negation of A.

P3 $0 \le S(A, B) \le 1$.

P4 $A = B$ if and only if $S(A, B) = 1$.

P5 If $S(A, B) = 0$ then either $A \cap B = \Phi$ (null); or $A^c \cap B^c = \Phi$; or $B = 1 - A$.

For $0 \le \epsilon \le 1$, if $S(A, B) \ge \epsilon$, we say that the two fuzzy sets A and B are ϵ-similar. Thus, $\epsilon = 1$ corresponds to equality of fuzzy sets. There may be many functions satisfying properties **P1–P5**. One such measure of similarity satisfying properties **P1–P5** is given in *Definition 1*.

Definition 1: Let A and B be two fuzzy sets defined over the universe of discourse U. The similarity index of the pair $\{A, B\}$ is defined as

$$S(A, B) = \min\{\alpha(A, B), \alpha(A^c, B^c)\}$$

where

$$\alpha(A, B) = \left\{ \frac{\sum_{u \in U} \{\mu_A(u) \cdot \mu_B(u)\}}{\sum_{u \in U} [\max(\mu_A(u), \mu_B(u))]^2} \right\}^{\frac{1}{2}}$$

and $\mu_{A^c}(u) = 1 - \mu_A(u)$.

In the case of $\sum_{u \in U} \{\max(\mu_A(u), \mu_B(u))\}^2 = 0$, we find that A and B are null fuzzy sets, and we set $\alpha(A, B) = 1 = S(A, B)$.

It is easy to see that properties **P1–P5** are satisfied by *Definition 1*.

Example 1: Let $U = \{u_1, u_2, u_3, u_4, u_5\}$ and $A = \{.1, .3, .5, .75, 1.0\}$; $B = \{.01, .09, .25, .5625, 1.0\}$; $C = \{.316, .55, .707, .866, 1.0\}$. Then, by Definition 1, $S(A, B) = \underline{0.785}$ and $S(A, C) = \underline{0.689}$.

Although the last property **P5** is a plausible and an intuitively appealing one, it is possible to argue in favor of a stricter condition for which $S(A, B)$ should be zero. Two crisp sets A and B are completely dissimilar only when $A \cap B = \Phi$. If $A \cap B \ne \Phi$, then they have some similarity as A and B have some elements in common. The similarity between the two increases as the number of elements by which the two sets differ decreases. The similarity becomes maximum (the maximum value may be thought of as 1) when the two sets are identical, i.e., $|A \cap B| = |A| = |B|$. Here, we consider a direct extension of this concept in defining the similarity between fuzzy sets. For two fuzzy sets, it is reasonable to assume that the similarity should be zero if and only if $A \cap B = \Phi$. Property **P5** may now be reformulated as **P5'**. For all fuzzy sets A, B; $S(A, B) = 0$ iff $A \cap B = \Phi$.

Thus, the need arises to find measures of similarity satisfying properties **P1–P4** and **P5'**. There could be several such measures. A family of such simple measures is given in *Definition 2*.

Definition 2: Let A and B be two fuzzy sets defined over the same universe of discourse U. The similarity index $S(A,B)$ of the pair $\{A,B\}$ is defined by

$$S(A,B) = 1 - \left[(1/n)\sum_{u} |\mu_A(u) - \mu_B(u)|^q\right]^{1/q}$$

where n is the cardinality of the universe of discourse and $q \geq 1$ is the family parameter.

Similarity indexes given by *Definition 2* satisfy properties **P1**–**P4** and **P5'**.

Example 2: Let U, A, B, C be as in Example 1. With $q = 2$ it is found that $S(A,B) = \underline{0.923}$ and $S(A,C) = \underline{0.919}$.

$S(A,B) \geq S(A,C)$ will imply that "B is at least as close to A as C is close to A." $S(A,B)$, as given in *Definition 2*, is quite sensitive—every change in A or B will be reflected in $S(A,B)$. Next, some more properties of S are discussed.

Theorem 1: If $S(A,B) = 1$ and $S(B,C) = 1$ then $S(A,C) = 1$.

Proof is straightforward.

Of course, in general, for all fuzzy sets A, B, and C the numbers $S(A,B)$ and $S(B,C)$ cannot always determine $S(A,C)$. If some structural arrangement between the sets is prescribed then, we may find an estimate for $S(A,C)$ as in Theorem 2.

Theorem 2: For all fuzzy sets A, B, C if either $A \subset B \subset C$ or $A \supseteq B \supseteq C$ then $S(A,C) \leq \min\{S(A,B), S(B,C)\}$.

Theorem 2 motivates us to consider a monotonicity property of similarity measures between fuzzy sets. Therefore, we are now in a position to state the axioms for similarity measure. For all fuzzy sets A, B

- A1 $S(B,A) = S(A,B)$.
- A2 $S(A^c, B^c) = S(A,B)$, A^c being some negation of A.
- A3 $0 \leq S(A,B) \leq 1$.
- A4 $A = B$ if and only if $S(A,B) = 1$.
- A5 $S(A,B) = 0$ if and only if $A \cap B = \Phi$.
- A6 If $A \supset B \supset C$, then $S(A,B) \geq S(A,C)$.

Throughout the paper, we used Zadeh's complement for A^c. On the basis of the above axioms, it is easy to see that the family of similarity measures defined in *Definition 2* is a valid choice. A general characterization of similarity index satisfying the set of axioms is not in the scope of the present paper.

IV. Proposed Similarity-Based Approximate Reasoning

In the previous section, we developed the concept of similarity index for measuring the likeness of fuzzy sets over a given universe of discourse and proposed two measures for the same. Here, we restrict ourselves to the similarity measure in Definition 2.

Let X, Y be two linguistic variables and let U, V, respectively, denote the universes of discourse. Two typical propositions p and q are given and we derive a conclusion according to similarity-based inference. The scheme is described in Table I. Let fuzzy sets A, A', B, and B' in Table I be defined as

$$A = \sum_{i=1}^{l}\{\mu_A(u_i)/u_i\}; \quad A' = \sum_{i=1}^{l}\{\mu_{A'}(u_i)/u_i\}$$
$$B = \sum_{i=1}^{m}\{\mu_B(v_i)/v_i\}; \quad B' = \sum_{i=1}^{m}\{\mu_{B'}(v_i)/v_i\}.$$

Existing methods use the similarity measure for a direct computation of inference without considering the induced relation. In the proposed method, we translate the conditional statement into a fuzzy relation. Then, the similarity between the fact and the antecedent of the rule is used to modify the said relation. With this, every change in the conditional premise and in the fact, may be incorporated into the induced fuzzy relation. Then, a conclusion can be drawn using the sup-operation. Thus, the conclusion is influenced by the change in the fact and the antecedent of the rule fired. Our proposed inference scheme is such that a significant difference between A and A' makes the conclusion B' less specific. This is done by choosing an expansion type of inference scheme. Here, the "UNKNOWN" case, i.e., the fuzzy inference $B' = V$, is taken as the limit of nonspecificity. Explicitly, with decrease in similarity value, which occurs when A and A' differ significantly, the inferred fuzzy set should be close to V. For $A' = A$, we expect $B' = B$ and for all other A', the relation $B' \supseteq B$ should hold. This, in turn, implies that nothing better than what the rule says should be allowed as a valid conclusion.

TABLE I
Ordinary Approximate Reasoning

p :	if X is A	then	Y is B
q :	X is A'		
r :			Y is B'.

A. Schema

In view of the above observations, we propose the following algorithm for reasoning.

Algorithm SAR (similarity-based approximate reasoning):

Step 1) Translate premise p and compute the relation $R(A,B)$ using some suitable translating rule (possibly, a T-norm operator).

Step 2) Compute $s = S(A, A')$ using *Definition 2*.

Step 3) Modify $R(A,B)$ with s to obtain the modified conditional relation $R(A,B\,|\,A')$ using some **Scheme C.**

Step 4) Use sup-projection operation on $R(A,B\,|\,A')$ to obtain B' as

$$\mu_{B'}(v) = \sup_{u}\{\mu_{R(A,B\,|\,A')}(u,v)\}. \tag{4}$$

For a given fact q:X is A' and from the condition p: if X is A then Y is B, we propose **Scheme C1** and **Scheme C2** for computation of $R(A,B\,|\,A')$ needed in **Step 3**.

Scheme C1: The first **Scheme C1** is based on a concept similar (but NOT identical) to the method proposed in [4]. We may recall here that the authors computed the conclusion $B' = \min(1, B/s)$, where s is the measure of similarity between fuzzy sets A and A' without considering the information suggested by the conditional rule. Here, we propose to modify the conditional relation according to (5)

$$[r'_{u,v}]_{l\times m} = \begin{bmatrix} r'_{u,v} = \min(1, r_{u,v}/s) & \text{if } s > 0 \\ = 1 & \text{otherwise} \end{bmatrix}. \tag{5}$$

The difference between the proposed scheme and the one presented in [4] may be easily noted. The proposed scheme, unlike the scheme in [4] and [17], does not produce the same conclusion when A and A' are interchanged. It is not difficult to see that in (5), if $s < r_{u,v}$ for some $v \subset V$ then $r'_{u,v}$ becomes equal to one. Thus, making the membership of that v in the resultant fuzzy set equal to one.

Example 3: Let us consider a problem as posed in Table I. Let $U = \{u_1, u_2, u_3, u_4\}$; $V = \{v_1, v_2, v_3, v_4\}$

$$A = 1.00/u_1 + 0.75/u_2 + 0.50/u_3 + 0.25/u_4$$
$$A' = 1.00/u_1 + 0.80/u_2 + 0.40/u_3 + 0.10/u_4$$
$$B = 0.25/v_1 + 0.50/v_2 + 0.75/v_3 + 1.00/v_4.$$

Using $\max(0, a+b-1)$ for the translation of the conditional statement and (4), the consequence becomes

$$B' = 0.3028/v_1 + 0.6056/v_2 + 0.9084/v_3 + 1/v_4.$$

In this case, the method in [4] will produce the conclusion

$$B'' = 0.2623/v_1 + 0.5245/v_2 + 0.7868/v_3 + 1/v_4.$$

Remark: In our scheme, if we use Mamdani's min rule for the computation of R, then our scheme will also produce the same B'' as the conclusion.

This scheme, although a heuristic one, is intuitively a plausible one. Our next scheme for the computation of modified relation $R(A, B \mid A')$ is based on a set of axioms.

Scheme C2: We believe that, in a similarity-based reasoning methodology, a scheme for computation of the induced relation, when a fact and a conditional statement are given, should satisfy the following axioms:

- AC1 If $S(A, A') = 1$, i.e., if $A' = A$, then $\mu_{R(A,B \mid A')}(u, v) = \mu_{R(A,B)}(u, v); (\forall (u, v)) \in U \times V$.
- AC2 If $S(A, A') = 0$, i.e., if $A' \cap A = \Phi$, then $\mu_{R(A,B \mid A')} = 1 \forall (u, v) \in U \times V$.
- AC3 As $S(A, A')$ increase from 0 to 1, $\mu_{R(A,B \mid A')}(u, v)$ decreases uniformly from 1 to $\mu_{R(A,B)}(u, v); \forall (u, v) \in U \times V$.

The first axiom asserts that we should not modify the conditional relation when A' and A remain equal. The second axiom asserts that when A' is completely dissimilar to A, i.e., A' and A have disjoint support, we should not conclude specifically. In such a situation, anything is possible. The third axiom says that as the fact A' changes from the most dissimilar case (similarity value zero) to the most similar one (similarity value one), the inferred conclusion should change from the most nonspecific case i.e., the UNKNOWN case $(B' = V)$ to the most specific case, i.e., $B' = B$ smoothly. This, in turn, means that whatever A' is, $R(A, B \mid A') \supseteq R(A, B)$, i.e., the induced relation should not be more specific than what is given as a condition. For notational simplicity, let us denote $S(A, A')$ by s and $R(A, B \mid A')$ by r'. Now, the third axiom *uniquely* suggests a function of the form

$$\frac{dr'}{ds} = k(\text{a constant}). \Rightarrow r' = ks + c. \quad c \text{ is a constant.}$$

These two constants can be determined from the conditions prescribed in the first and the second axioms. More explicitly, when $s = 1$ we know that $r' = r$ and when $s = 0$ we know that $r' = 1$. Thus $r' = 1 - (1 - r) \cdot s$ is our new scheme for the modification of the conditional relational. Therefore, the axioms *uniquely* suggest **Scheme C2** as

$$\mu_{R(A,B \mid A')}(u, v) = 1 - (1 - \mu_{R(A,B)}(u, v)) \cdot S(A, A'). \quad (6)$$

From (5) and (6) it is easy to see that when $S(A, A') = 0$, we have $B' = V$. In other words, it is impossible to conclude anything when A and A' are completely dissimilar. It is also easy to see that when $S(A, A')$ is close to unity, then $R(A, B \mid A')$ is close to $R(A, B)$ and hence the inferred fuzzy set B' will be close to B, i.e., $S(B, B')$ is close to unity. The third axiom also ensures that a small change in the input produces a small change in the output and hence, in this sense the above mechanism of inference is stable. As in the previous case, in (6) if either $S(A, A') = 0$ or $\mu_{R(A,B)}(u, v) = 1$ then $r'_{u,v}$ becomes equal to one.

Example 4: Let us consider the same problem as in *Example 3*. Using Mamdani's min rule and *Definition 2*, we find the consequence as $B' = 0.2851/v_1 + 0.5234/v_2 + 0.7617/v_3 + 1/v_4$.

Instead of using similarity-based approximate reasoning methodology in deriving a consequence, if we consider the existing max-min compositional rule of inference then, the result would be

$$B'_1 = 0.25/v_1 + 0.5/v_2 + 0.75/v_3 + 1/v_4.$$

There is no change in the output (i.e., $B'_1 = B$) although the inputs differ significantly. Also, it may be shown that the same happens for a large class of fuzzy sets each different from the other. This is a drawback in executing max-min compositional rule of inference in its present form. If Mamdani's min-rule is used for the translation of the implication statement and only normal fuzzy sets are considered then, $A' = A$ will imply that $B' = B$. This is because, in this case, $S(A, A') = 1$ and hence $R(A, B \mid A')$ will be equal to $R(A, B)$.

Let A be a normal fuzzy set. If we assume that the translating rule used in generating the conditional relation is one of T-norm type then a basic and desirable property of the inferred proposition is: *nothing better than what the rule says can be concluded.* We present this in Theorem 3. Consider the model shown in Table I.

Theorem 3: For all $A, A', B' \supseteq B$.

Proof is straightforward.

B. Application to Different Models

A rule base hardly contains rules with only one clause in the conditional statement. For rule-bases with multiple clauses in the conditional statement, we can apply the proposed scheme in the following manner.

Let $X_1, X_2, \ldots, X_k, Y$ be $k + 1$-linguistic variables defined, respectively, over the universes of discourse $U_1, U_2, \ldots, U_k, V$ and let $U_i = \{u_i^j\}; j = 1, 2, \ldots, j_i, V = \{v_i\}; i = 1, 2, \ldots, l$. Let

p: if X_1 is A_1 & X_2 is A_2 & … & X_k is A_k then Y is B
q: X_1 is A'_1 & X_2 is A'_2 & … & X_k is A'_k

———————————

r: $\leftarrow Y$ is B'.

The consequence r may be derived according to the following basic steps. Let fuzzy sets A_i, A'_i, and B be given as

$$A_i = \sum_{j=1}^{j_i} \mu_{A_i}\left(u_i^j\right) / u_i^j$$

$$A'_i = \sum_{j=1}^{j_i} \mu_{A'_i}\left(u_i^j\right) / u_i^j; \quad i = 1, 2, \ldots, k$$

$$B = \sum_{i=1}^{l} \mu_B(v_i)/v_i.$$

Here, the conditional proposition p is first translated into a fuzzy relation R on the product space $U_1 \times U_2 \times \cdots U_k \times V$. Now R may be computed using any suitable translating rule, possibly a T-norm operator. Then we compute $S(A_i, A'_i)$ for $i = 1, 2, \ldots, k$ and set $s = \min\{S(A_1, A'_1), S(A_2, A'_2), \ldots, S(A_k, A'_k)\}$. The conditional relation is modified using either **Scheme C1** or **Scheme C2**. If **Scheme C1** is used, then we have $R' = R(A_1, A_2, \ldots, A_k, B \mid A'_1, A'_2, \ldots, A'_k)$ according to

$$\mu_{R'}(u_1, u_2, \ldots, u_k, v) = \min\left\{1, \frac{1}{s}\mu_R(u_1, u_2, \ldots, u_k, v)\right\}.$$

If, instead, **Scheme C2** is used, then we find $\mu_{R'}(u_1, u_2, \ldots, u_k, v) = 1 - s.(1 - \mu_R(u_1, u_2, \ldots, u_k, v))$.

In both cases, the conclusion B' will be given as

$$\mu_{B'}(v) = \sup_{u_1, \ldots, u_k} \left\{\mu_{R(A_1, \ldots, A_k, B \mid A'_1, \ldots, A'_k)}(u_1, \ldots, u_k, v)\right\}.$$

If $s = 0$ then, at least one pair (A_i, A'_i) will be complementary (disjoint support) and it would become impossible to conclude anything in particular. This is represented by the fact that anything follows as conclusion. In this case, **Scheme C2** produces $B' = V$ (UNKNOWN), whereas for **Scheme C1** we set $B' = V$. The algorithm is schematized as follows.

Algorithm A1:

Step 1) Compute $S(A_i, A'_i)$ for $i = 1, 2, \ldots, k$ and set $s = \min\{S(A_1, A'_1), S(A_2, A'_2), \ldots, S(A_k, A'_k)\}$.

Step 2) Translate premise p and compute $R(A_1, A_2, \ldots, A_k, B)$ using any suitable translating rule possibly, a T-norm operator.
Step 3) Modify $R(A, B)$ with s to obtain the modified conditional relation $R' = R(A_1, A_2, \ldots, A_k, B \mid A'_1, A'_2, \ldots, A'_k)$.
Step 4) Use sup-projection operation on R' to obtain B'.

So far, we have considered only one rule, but in a real life application we would encounter multiple rules as follows:

if X_1 is A_{11} & ... & X_k is A_{1k} then Y is B_1
else if X_1 is A_{21} & ... & X_k is A_{2k} then Y is B_2
⋮
else if X_1 is A_{M1} & ... & X_k is A_{Mk} then Y is B_M
X_1 is A_1 & ... & X_k is A_k.

———————

Conclusion: Y is B.

The problem is to find the linguistic value of the variable Y as suggested by the rules, when specific values of the k-variables are given. Under the conventional mechanism, for each rule, the consequent fuzzy set is calculated according to existing method of inference and then the union of all consequent fuzzy sets is taken as the conclusion which is defuzzified, if necessary, using some defuzzification scheme.

In the present case of similarity-based reasoning we cannot do this, as the membership values computed from the modified induced relation becomes less and less specific, the similarity between the facts and antecedent of a rule decreases. In Mamdani-type approximate reasoning, with the reduction of the firing strength, the membership values of various elements become equal to the firing strength, making it an ambiguous one (more alternatives with similar membership values), but the membership values at which the ambiguity occurs becomes less than one. For example, if the firing strength of a rule is, say 0.3, then all alternatives which have membership values greater than or equal to 0.3 take membership values equal to 0.3. On the other hand, in the present case, if the similarity value is 0.3, then the membership values of elements in the inferred fuzzy set will be at least 0.3. Moreover, the elements having membership value greater than or equal to 0.3 in the consequent of the rule will be equal to "1" in the consequent fuzzy set. This means that, with decrease in similarity the computed membership values increase and ultimately move close to the least specific case (with membership values of 1 for all alternatives). The above discussion is illustrated with the help of a diagram. In Fig. 1, let us suppose, the symmetric triangular fuzzy set represents the consequent of a rule. When the firing strength of the rule is 0.3, the derived conclusion from the rule is given by the trapezoid with height 0.3. Clearly, every value in $[a, b]$ in Fig. 1 has the same membership grade 0.3. On the other hand, if the similarity between the fact and the antecedent of the rule is 0.3, then the conclusion derived using similarity-based mechanism is given by the trapezoidal fuzzy set with height $\alpha \geq 0.3$ (shown with dotted line in Fig. 1). In this case, every alternative in $[c, d]$ in Fig. 1 could be a solution with membership value of α. Here, not only more alternatives have been offered (since $[a, b] \subseteq [c, d]$) with the same membership value than the previous case but, also the conclusion becomes more close to the least specific case. For this reason, we propose a new scheme, for computing the final conclusion, based on a measure of similarity. Our method is based on rule-selection and then rule-execution. In both cases, we use the concept of similarity between fuzzy sets as a basis of the task. First of all, we compute $s_{ij} = S(A_i, A_{ji})$; $i = 1, 2, \ldots, k$ and $j = 1, 2, \ldots, M$. Next, we compute the overall rule matching index as

$$s^j = \min_i s_{ij}. \tag{7}$$

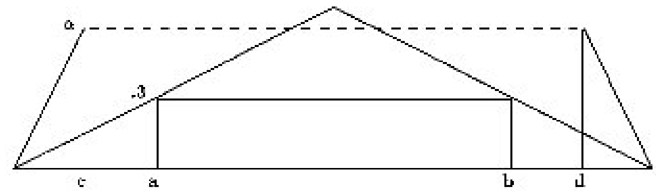

Fig. 1. Comparison of firing strength based and similarity-based reasoning scheme.

From among the M distinct rules, we choose those rules for which $s^j > \epsilon$. ϵ may be interpreted as a threshold in our case. Then, we apply **Algorithm A1** to generate a conclusion from each rule conformal to firing. The overall output may be generated using the intersection of conclusion fuzzy sets resulted from different fired rules. It is important to note that, the intersection operation is chosen in order to justify the rule-selection procedure. Here, fewer rules are fired and the output of each rule is significant. The algorithm is schematized as follows.

Algorithm A2:

Step 1) Compute s_{ij} for $i = 1, 2, \ldots, k$; $j = 1, 2, \ldots, M$ and then s^j according to (7).
Step 2) Define ϵ and find the rules conformal to firing.
Step 3) Translate the ith rule, provided $s^i > \epsilon$ and compute the relation R_i using any suitable translating rule possibly, a T-norm operator.
Step 4) Modify R_i with s^i to obtain the modified conditional relation R'_i according to either (5) or (6).
Step 5) Use sup-projection operation on R'_i to obtain B'_i

$$\mu_{B'_i}(c) = \sup_{u_1, \ldots, u_k} \mu_{R'_i}(u_1, \ldots, u_k, c). \tag{8}$$

Step 6) Compute the output $B = \bigcap_i B'_i$.

Examples and relevant issues are considered in Section V.

V. Pattern Classification

Designing a classifier is an easy task as long as, the objects concerned are well-defined and the boundaries of the groups or classes are nonoverlapping. In most practical instances, the classes are overlapped making the classification difficult. In such cases, fuzzy set theory may be used [18]–[21]. Here, we present a classification algorithm using a similarity-based inference mechanism. Let there be c classes and $X = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$, $\mathbf{x}_i \in R^p$ be the training data. Let us assume that, for each point in the training data set, the actual class it has come from, is known. The problem is to design a rule-based system, using the proposed similarity-based reasoning, so that unknown points may be classified. The classifier is designed, based on a set of rules of the form *If* X_1 *is* A^i_1 *and* X_2 *is* A^i_2 *and* $\cdots$ X_p *is* A^i_p *then* C *is* B^i where, X_j; $j = 1, 2, \ldots, p$ are p linguistic variables corresponding to the p feature values. A^i_j; $j = 1, 2, \ldots, p$ are the linguistic values of the respective variables in the i-th rule and B^i is a fuzzy set defined on the set of classes. Note that all A^i_j are not distinct. Rules are generated either using expert operator's knowledge or by exploratory data analysis.

At the learning stage, we impose a fuzzy partition on the feature space using fuzzy sets defined over appropriate universes. Here, triangular membership functions are used but other choices are possible also. Next, we generate *fuzzy if-then* rules as precondition for classification. These rules are tested against the classification of the training data. If found satisfactory, we proceed further; otherwise, the rules may be modified either by trial-and-error method or by some systematic methods like Genetic Algorithms, gradient search [18], [19].

For classifying an unknown pattern, each feature is fuzzified using triangle shaped fuzzy sets. For a particular feature, we compute the similarity between the fuzzified feature value and the different linguistic values of that feature, as used in the rule-base. From them, we find the maximum similarity value and the corresponding fuzzy set. Next, we choose all rules with the said feature having the corresponding fuzzy value. This process is continued for all features, and for all rules for which the previous features have fuzzy values with the maximum similarity. Ultimately, a single rule will be found satisfying the maximum similarity value criterion. Otherwise, the system has rules with identical rule-antecedent. Here lies the novelty of similarity-based approach to pattern classification. The best rule has been chosen from a class of rules for possible firing. Then the similarity-based approximate reasoning methodology is applied (with similarity value equal to the minimum of all the maximum similarity values computed earlier for each feature) to obtain a fuzzy set representation of different classes, using **Algorithm A1**. At the time of making a nonfuzzy decision, the class with the maximum membership value is selected. Ties (for patterns lying in the overlapped zone), if arises, may be broken arbitrarily or we may use first-of-maxima or the last-of-maxima. Multiple classification is a product of fuzzy algorithm. We summarize the preceding discussion in the following algorithm:

Algorithm PC: Pattern Classification:

- Step 1) Take an input vector $\mathbf{x} = (x_1, x_2, \ldots, x_p) \in R^p$.
- Step 2) Fuzzify each real value x_i using triangular membership function. Let these fuzzified values be $\{A_1{}', A_2{}', \ldots, A_p{}'\}$.
- Step 3) Compute $S(A_1{}', A_1^i)$ where the index i ranges over all fuzzy sets for the first feature F_1. Set $k_1 = \text{Argmax}_i\{S(A_1{}', A_1^i)\}$. Find $k_j = \text{Argmax}_i\{S(A_j{}', A_j^i)\}$ where i varies over those rules for which F_1 is A^{k_1} and F_2 is $A_2^{k_2}$ and … and F_{j-1} is $A_{j-1}^{k_{j-1}}$; $j = 2, 3, \ldots, p$.
- Step 4) Apply **Algorithm A1**, with fuzzy sets $\{A_1{}', A_2{}', \ldots, A_p{}'\}$, for firing the rule as found in Step 3, using the minimum of all the maximum similarity values as obtained in Step 3 and obtain a consequence B' on the class of patterns.
- Step 5) The class with the maximum membership value in B' is taken as the class of the input vector $\mathbf{x}$. Ties, if they arise, may be broken arbitrarily.

A. An Application to Telugu Vowel Classification

Telugu is one of the Indian languages spoken in the southern part of the country. The data set consists of 710 discrete phonetically balanced speech samples for the Telugu vowels in consonant-vowel nucleus-consonant (CNC) form. Let X_1 and X_2 be two linguistic variables used to represent the vague description about the two given formant frequencies. Reported results [22] suggest that inclusion of feature 3 does not add much discriminating power to the data set. So we use only the first two formant features. The initial rule set is so designed that it covers the entire input space. Each rule is of the form *If* X_1 *is* A_1^i *and* X_2 *is* A_2^i *then* C *is* B^i.

Here, A_j^i are the linguistic values of the linguistic variable X_j; $j = 1, 2$, that appear in the body of the ith rule. Since our intention is to show the efficiency of the proposed similarity-based reasoning methodology, we use a rule set similar to the one used in [22]. The rule base may also be assigned by experts or may be learnt. For X_1 and X_2 only five and seven linguistic values are used resulting in 35 rules, as shown in Tables II and III. The definition of the fuzzy sets for the two features used in the antecedent of the rules are presented in Tables IV and V. For an unknown input we apply **Algorithm PC** and compute the class represented by the input.

TABLE II
RULE BASE

A	ZE : Zero	BL : Below Low
BZ : Below Zero	.5/e+1/u	.1/e+.5/o+1/u
ZE : Zero	.5/e+1/i	.5/e+1/i
BL : Below Low	1/e+.1/o+.1/u+.5/ə	1/e+.5/i+.5/ə
LO : Low	.5/a+.1/e+1/o+.1/u+1/ə	.5/a+.5/e+1/o+.1/u+1/ə
ME : Medium	1/e+.5/i	.1/a+.5/o
HI : High	1/e+1/ə	1/e+.1/ə
AH : Above High	1/a+1/o+1/ə	1/a+1/ə

TABLE III
RULE BASE (*Continued.*)

A	LO : Low	ME : Medium	HI : High
BZ	.5/e+.5/o+1/u	.5/e+.1/o+1/u	.5/e+.5/i
ZE	1/o+1/u	.1/e+1/o+.5/u+.1/ə	1/e+1/o+.5/u+.5/ə
BL	1/a+1/i+.1/ə	1/e+.5/i	.1/a+1/o+.1/u
LO	.5/a+1/e+.1/o+1/ə	1/e+1/i+1/ə	1/e+.5/i+.1/ə
ME	1/a+1/o+.1/ə	1/a+.1/e+.5/o+1/ə	1/a+.5/e+.1/o+1/ə
HI	.5/a	.1/a+.1/o	1/a+.1/o+.1/ə
AH	.5/e+.5/ə	.5/e+1/ə	.1/a

TABLE IV
FUZZY SETS OF THE LINGUISTIC VALUES FOR F_1

ZE	0.1	0.5	1.0	0.5	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
BL	0.0	0.0	0.1	0.5	1.0	0.5	0.1	0.0	0.0	0.0	0.0	0.0	0.0
LO	0.0	0.0	0.0	0.0	0.1	0.5	1.0	0.5	0.1	0.0	0.0	0.0	0.0
ME	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.5	1.0	0.5	0.1	0.0	0.0
HI	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.5	1.0	0.5	0.1

TABLE V
FUZZY SETS OF THE LINGUISTIC VALUES FOR F_2

BZ	0.25 0.5 1.0 0.5 0.25 0.10 0 0 0 0 0 0 0 0 0 0 0 0
ZO	0 0.1 0.25 0.5 1.0 0.5 0.25 0.1 0 0 0 0 0 0 0 0 0 0 0
BL	0 0 0 0.1 0.25 0.5 1.0 0.5 0.25 0.1 0 0 0 0 0 0 0 0 0
LO	0 0 0 0 0 0.1 0.25 0.5 1.0 0.5 0.25 0.1 0 0 0 0 0 0 0
ME	0 0 0 0 0 0 0 0 0.1 0.25 0.5 1.0 0.5 0.25 0.1 0 0 0 0 0
HI	0 0 0 0 0 0 0 0 0 0 0 0.1 0.25 0.5 1.0 0.5 0.25 0.1 0 0
AH	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0.1 0.25 0.5 1.0 0.5 0.25 0.1

TABLE VI
RECOGNITION SCORE (%) FOR TELUGU VOWELS DATA

	first of maxima	last of maxima	middle of maxima	at random selection	suggested choice include correct class
a	72.29	67.47	72.29	49.40	80.75
e	95.00	56.00	95.00	65.50	92.00
i	59.40	76.95	59.40	75.19	93.98
o	89.66	54.31	89.66	61.21	96.55
u	57.14	71.43	57.14	70.54	82.14
ə	60.00	71.21	60.00	53.03	96.97
Overall	70.00	66.56	70.00	54.37	91.41

1) Results and Discussion: Table VI summarizes the results. Cell entries correspond to percentage of correct classification while the column label states the tie breaking strategy used. While generating column 2 of Table VI ties were broken by choosing the first class with maximum grade in the resultant fuzzy set. Table VI shows that the average recognition score is 70%. If, instead, we choose the last class with the maximum grade, to break a tie, the average recognition score is 66.56% (column 3). Column 4 presents the result when a middle class is chosen, in a tied situation and the average recognition score is found to be 70.00%. The result of random choice is provided in column 5. Here, the result of average recognition score is found to be 64.37%. A scatterplot of the data [22] reveals significant overlap between different pairs of classes and consequently the performance of the classifier is quite satisfactory. Some improvement in performance may be realized through tuning of the membership functions as well as through change of the rule base. The scatterplot in [22] suggests that, for any classifier which uses only two features, some misclassifications are bound to result. If the proposed scheme is a consistent one, then any fuzzy logic-based classifier is likely to suggest more than one choice with the highest membership value for points lying in the overlapping regions. In order to establish that it is indeed the case, let us assume that the system output is correct if the alternatives suggested by the rule-based system include the correct class. The last column in Table VI is generated keeping this in mind. If the rule base suggests only one class then the recognition score for that class is increased or else, if the system suggests more than one class containing the correct class then the recognition score for the correct class is increased. This results in a significant improvement in the recognition score (91.41%) as shown in Table VI. On scrutiny, we find that, when more than one alternative satisfies our criterion for selecting a single class, the corresponding input data point is found to come from some overlapped region containing the correct class.

VI. Conclusion

We have discussed the concept of similarity measure between fuzzy sets based on a pairwise comparison of elements and proposed a set of axioms for the choice of such measures. A family of such measures have been proposed based on these axioms and their properties were discussed. The use of such measures in deriving a fuzzy consequence from given condition(s) and fact(s) has been extensively discussed. We also developed a powerful mechanism, similarity-based approximate reasoning, through the integration of existing similarity-based reasoning mechanism and Zadeh's compositional rule of inference. We applied the said mechanism in designing rule-based fuzzy systems. The effectiveness of the scheme is demonstrated in pattern classification. We have established that similarity-based reasoning and conventional approximate reasoning can be integrated into a single framework for reasoning with vague concepts. We have also established that the concept of similarity between fuzzy sets is useful not only in deriving a consequence but also in selecting rules from the rule base to be fired, depending on a particular input specification.

References

[1] S. Raha and K. S. Ray, "Reasoning with vague default," *Fuzzy Sets Syst.*, vol. 91, no. 3, pp. 327–338, 1997.

[2] ——, "Approximate reasoning with time," *Fuzzy Sets Syst.*, vol. 107, no. 1, pp. 59–79, 2000.

[3] R. Zwick, E. Carlstein, and D. V. Budescu, "Measures of similarity among fuzzy concepts: A comparative analysis," *Int. J. Approx. Reas.*, vol. 1, pp. 221–242, 1987.

[4] I. B. Turksen and Z. Zhong, "An approximate analogical reasoning approach based on similarity measures," *IEEE Trans. Syst., Man, Cybern.*, vol. 18, no. 6, pp. 1049–1056, 1988.

[5] S. Raha, "Similarity-based approximate reasoning," in *Proc. IIZUKA'96, Methodologies for the Conception, Design, and Application of Intelligent Systems, 4th Int. Conf. Soft Computing*, Iizuka, Japan, pp. 414–417.

[6] S. M. Chen, "A new approach to handling fuzzy decision making problems," *IEEE Trans. Syst., Man, Cybern.*, vol. 18, no. 6, pp. 1012–1016, 1988.

[7] D. S. Yeung and E. C. C. Tsang, "Improved fuzzy knowledge representation and rule evaluation using fuzzy petri nets and degree of subsethood," *Intell. Syst.*, vol. 9, no. 12, pp. 1083–1100, 1994.

[8] ——, "A comparative study on similarity-based fuzzy reasoning methods," *IEEE Trans. Syst., Man, Cybern. B*, vol. 27, pp. 216–227, Apr. 1997.

[9] D. Dubois and H. Prade, "A unifying view of comparison indexes in a fuzzy set theoretic framework," in *Fuzzy Set and Possibility Theory—Recent Developments*, R. R. Yager, Ed. New York: Pergamon, pt. I, pp. 3–13.

[10] S. M. Chen, "A weighted fuzzy reasoning algorithm for medical diagnosis," *Dec. Support Syst.*, vol. 11, pp. 37–43, 1994.

[11] Z. Bien and M. G. Chun, "An inference network for bidirectional approximate reasoning based on an equality measure," *IEEE Trans. Fuzzy Syst.*, vol. 2, no. 2, pp. 177–180, 1994.

[12] W. Pedrycz, "Neurocomputations in relational systems," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, no. 3, pp. 289–297, 1991.

[13] C. P. Pappis and N. I. Karacapilidis, "A comparative assessment of measures of similarity of fuzzy values," *Fuzzy Sets Syst.*, vol. 56, pp. 171–174, 1993.

[14] W. J. Wang, "New similarity measures on fuzzy sets and on elements," *Fuzzy Sets Syst.*, vol. 85, pp. 305–309, 1997.

[15] H. L. Kwang, Y. S. Song, and K. M. Lee, "Similarity measure between fuzzy sets and between elements," *Fuzzy Sets Syst.*, vol. 62, pp. 291–293, 1994.

[16] L. A. Zadeh, "A theory of approximate reasoning," in *Machine Intelligence*, J. E. Hayes, D. Mitchie, and L. I. Mikulich, Eds. New York: Wiley, 1979, vol. 9, pp. 149–194.

[17] F. C. C. Tsang, J. W. T. Lee, and D. S. Yeung, "Similarity-based fuzzy reasoning methods for fuzzy production rules," in *Proc. IFSA, World Congress*, Sao Paolo, Brazil, 1995, pp. 157–160.

[18] S. L. Chiu, "Extracting fuzzy rules for pattern classification by cluster estimation," in *Proc. 6th IFSA, World Congress*, Sao Paulo, Brazil, July 1–4, 1995.

[19] ——, "Extracting fuzzy rules from data for function approximation and pattern classification," in *Fuzzy Information Engineering*, D. Dubois, H. Prade, and R. R. Yager, Eds. New York: Wiley, pp. 149–162.

[20] W. Pedrycz, "Fuzzy sets in pattern recognition: Methodology and methods," *Pattern Recognit.*, vol. 23, no. 1/2, pp. 121–146, 1993.

[21] K. S. Ray and J. Ghoshal, "Approximate reasoning approach to pattern recognition," *Fuzzy Sets Syst.*, vol. 77, pp. 125–150, 1996.

[22] ——, "Neuro-fuzzy approach to pattern recognition," *Neural Networks*, vol. 10, no. 1, pp. 161–182, 1977.