Mathware & Soft Computing 8 (2001) 113-128

Rule-Based Fuzzy Object Similarity

H. Bunke[1], X. Fabregas[1] and A. Kandel[2]
[1]Department of Computer Science
University of Bern, Switzerland
[2]Department of Computer Science and Engineering
University of South Florida, Tampa, FL

Abstract

A new similarity measure for objects that are represented by feature vectors of fixed dimension is introduced. It can simultaneously deal with numeric and symbolic features. Also, it can tolerate missing feature values. The similarity measure between two objects is described in terms of the similarity of their features. IF-THEN rules are being used to model the individual contribution of each feature to the global similarity measure between a pair of objects. The proposed similarity measure is based on fuzzy sets and this allows us to deal with vague, uncertain and distorted information in a natural way. Several formal properties of the proposed similarity measure are derived; in particular, we show that the measure can be used to model the Euclidean distance as well as other, non-Euclidean distance functions. Also, an application of the proposed similarity measure to nearest-neighbor classification in a medical expert system is described.

Keywords: Similarity, Euclidean distance, nearest-neighbor classifier, fuzzy linguistic variable, fuzzy inference, case-based reasoning, medical expert systems, thyroid gland diagnosis

1 Introduction

Similarity is a key concept in intelligent information processing. For example, in information retrieval pieces of information are tagged with labels or indexes and the relevance of some entry in the database with respect to a query is determined based on the similarity of the indexes of the query and the database entry [1]. Case-based reasoning (CBR) attempts to solve problems by reusing past experience [2]. An actual problem is compared to cases that have been previously analyzed by the CBR system and those previous solutions are transformed in order to solve the actual problem. Clearly, the determination of the similarity of the actual problem and a previous case is one of the most important steps in case-based reasoning. In pattern recognition, models or prototypes of known objects are stored in a database and the identity of an unknown pattern is determined based on its similarity to the prototypes [3]. This approach, which is also known as nearest-neighbor classification, has been applied in various instances to numerous problems [4].

One of the most widely used approaches to similarity detection is Euclidean distance. It is applicable if objects are represented by N-dimensional vectors of numerical features. Given two objects $\boldsymbol{x} = (x_1, ..., x_N)$ and $\boldsymbol{y} = (y_1, ..., y_N)$ where x_i and y_i represent the feature values of $\boldsymbol{x}$ and $\boldsymbol{y}$, respectively, the Euclidean distance between $\boldsymbol{x}$ and $\boldsymbol{y}$ is given by

$$d_E(\boldsymbol{x}, \boldsymbol{y}) = \sqrt{\sum_{i=1}^{N} (x_i - y_i)^2} \tag{1.1}$$

or, equivalently,

$$d'_E(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{N} |x_i - y_i| \tag{1.2}$$

Clearly, the more similar $\boldsymbol{x}$ and $\boldsymbol{y}$ are, the smaller is their Euclidean distance, and vice versa. If objects are described by symbolic rather than numeric features then Eq. (1.1) or (1.2) can still be used as a similarity measure if the distance, or dissimilarity, of all pairs of feature values x_i and y_i is appropriately defined. More general similarity measures are string, tree, and graph edit distance [5, 6, 7]. Here the similarity of two structures is defined in terms of the minimum number of edit operations that are required to transform one of the structures into the other. While Euclidean distance as well as edit distance based similarity measures are application independent, many other application specific similarity measures have been proposed, for example, in the domain of case-based reasoning [8, 9, 10].

In this paper, a new application independent similarity measure is introduced. It is very general in the sense that it allows to deal with numeric as well as symbolic features at the same time. The values of particular object features may be missing. Rather than treating each feature in a uniform manner as it is done, for example, in Eqs. (1.1) and (1.2), the proposed similarity measure is based on rules that deal with features on an individual basis. Thus not only the individual importance of a feature can be easily modeled, but also interdependencies between features can be taken into account. The proposed similarity measure is based on fuzzy set theory [11, 12], which provides us with proper capabilities to deal with vague, uncertain, and distorted information in a natural way. The similarity of objects can be expressed in qualitative and quantitative terms; in particular, the new similarity measure can be used to model both the Euclidean distance and other, non-Euclidean distance functions. Thus it is more powerful than just Euclidean distance.

In Section 2, the new similarity measure is introduced. Some formal properties are derived in Section 3. An application of the proposed similarity measure to nearest-neighbor classification in a medical expert system is described in Section 4. Finally, a discussion and conclusions is presented in Section 5.

2 Rule-Based Fuzzy Similarity

Let each object $\boldsymbol{x}$ in our problem domain be represented by an N–dimensional vector, i.e.

$$\boldsymbol{x} = (x_1, \ldots, x_N) \tag{2.1}$$

such that x_i the i–th feature of $\boldsymbol{x}$. A feature can be from any domain $\mathcal{D}_\rangle$, i.e. $x_i \in \mathcal{D}_\rangle$. For example, features can be integers, real numbers, or symbols from some discrete set of values. The domain of feature i can be different from the domain of feature j. It is also admissible that the value of a feature is undefined or missing.

Let X denote the set of objects under consideration, such that objects $\boldsymbol{x} = (x_1, \ldots, x_N)$, $\boldsymbol{y} = (y_1, \ldots, y_N) \in X$. For example, $\boldsymbol{x}$ may represent some unknown object, whose identity is to be determined, while $\boldsymbol{y}$ belongs to a database of known objects. In this section we develop a similarity measure which is based on the distance of objects $\boldsymbol{x}$ and $\boldsymbol{y}$, and object features x_i and y_i. The smaller the objects' distance is, the greater is their similarity and vice versa. In order to measure the distance of a pair of features, x_i and y_i, $1 \leq i \leq N$, we introduce the fuzzy linguistic variable *distance of the i–th feature*, or $F(x_i, y_i)$ for short. This fuzzy linguistic variable consists of a number of fuzzy sets $F_1, \ldots, F_n$ that are all defined over $\mathcal{D}_\rangle \times \mathcal{D}_\rangle$. Whenever x_i or y_i is undefined, all of $F_1, \ldots, F_n$ are undefined as well. The membership function of fuzzy set F_j is given by

$$\mu_{F_j} \ : \ \mathcal{D}_\rangle \times \mathcal{D}_\rangle \rightarrow [/, \infty] \ , \infty \leq | \leq \backslash. \tag{2.2}$$

For example, for $x_i, y_i \in \Re$, $F(x_i, y_i)$ may consist of the following three fuzzy sets (i.e. $n = 3$):

$$\begin{aligned} F_1 &= \text{identical}, \\ F_2 &= \text{similar}, \\ F_3 &= \text{different}, \end{aligned} \tag{2.3}$$

with the following membership functions (for a graphical representation see Fig. 1):

$$\mu_{F_1}(x_i, y_i) = \begin{cases} \frac{1}{10}(10 - |x_i - y_i|) & \text{if } |x_i - y_i| \leq 10 \\ 0 & \text{otherwise} \end{cases}$$

$$\mu_{F_2}(x_i, y_i) = \begin{cases} 0 & \text{if } |x_i - y_i| \geq 20 \\ \frac{1}{10}|x_i - y_i| & \text{if } |x_i - y_i| \leq 10 \\ 1 - \frac{1}{10}(|x_i - y_i| - 10) & \text{otherwise} \end{cases} \tag{2.4}$$

$$\mu_{F_3}(x_i, y_i) = \begin{cases} 0 & \text{if } |x_i - y_i| \leq 10 \\ 1 & \text{if } |x_i - y_i| \geq 20 \\ \frac{1}{10}(|x_i - y_i| - 10) & \text{otherwise} \end{cases}$$

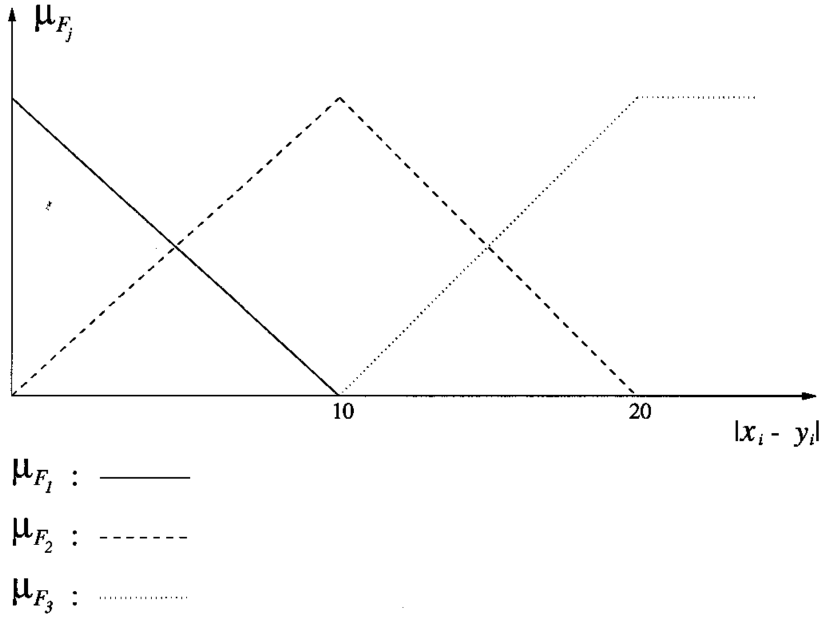

Fig. 1. An example of the linguistic variable *distance of the i-th feature*, consisting of three fuzzy sets F_1, F_2, F_3

Fuzzy sets such as those given in Fig. 1 have to be defined for all features $i = 1, \ldots, N$. In general, the fuzzy sets $F_1, \ldots, F_n$ that represent the distance of the i–th feature will be different from the fuzzy sets that represent the distance of the j–th feature, depending on the nature of the features. In other words, $F_1 = F_1(x_i, y_i), \ldots, F_n = F_{n_i}(x_i, y_i)$. In order to keep our notation simple, however, we don't explicitly express this dependency in the formulas.

By definition, for each $i = 1, \ldots, N$, we assume an ordering

$$(F_1, F_2, \ldots, F_n) \tag{2.5}$$

on the fuzzy sets that represent the distance of feature i such that fuzzy set F_{j+1} corresponds to a larger degree of distance than fuzzy set F_j, $1 \leq j \leq n - 1$. Formally, for each pair of fuzzy sets F_j and F_k, $j < k$, we require the existence of a partition of $\mathcal{D}_\rangle \times \mathcal{D}_\rangle$ into two disjoint subsets $S \subseteq \mathcal{D}_\rangle \times \mathcal{D}_\rangle$ and $\overline{S} \subseteq \mathcal{D}_\rangle \times \mathcal{D}_\rangle$ such that the following conditions are satisfied:

- any pair $(x, y) \in S$ has a smaller distance than any pair from $\overline{S}$, and
- $\mu_{F_j}(x_i, y_i) \geq \mu_{F_k}(x_i, y_i) \Leftrightarrow (x_i, y_i) \in S$ and (2.6)
- $\mu_{F_j}(x_i, y_i) \leq \mu_{F_k}(x_i, y_i) \Leftrightarrow (x_i, y_i) \in \overline{S}$

Moreover, we require that $\mu_{F_1}(x_i, y_i)$ takes on its maximum value if x_i and y_i are identical, i.e.

$$\mu_{F_1}(x_i, x_i) = 1 \text{ for } 1 \leq i \leq N \tag{2.7}$$

For some applications it may be useful to furthermore require, for $1 \leq i \leq N$,

$$\mu_{F_j}(x_i, y_i) = \mu_{F_j}(y_i, x_i) \text{ for } 1 \leq j \leq n \tag{2.8}$$

As can be seen later this is a necessary and sufficient condition for our distance measure to be symmetric.

In order to model the global similarity of objects, we introduce the fuzzy linguistic variable *distance of objects* $\boldsymbol{x}$ *and* $\boldsymbol{y}$, or $D = D(\boldsymbol{x}, \boldsymbol{y})$ for short. D consists of a number of fuzzy sets $D_1, \ldots, D_m$ that are defined over the interval $[0, 1]$. Let $u \in [0, 1]$ denote the base variable of linguistic variable D. The membership functions of fuzzy sets D_k are given by

$$\mu_{D_k} \ : \ [0,1] \rightarrow [0,1] \ , \ 1 \leq k \leq m \tag{2.9}$$

For example, for $m = 5$, D may consist of the following five fuzzy sets:

$$\begin{aligned} D_1 &= \text{identical}, \\ D_2 &= \text{similar}, \\ D_3 &= \text{similar to a small degree}, \\ D_4 &= \text{rather different}, \\ D_5 &= \text{completely different}, \end{aligned} \tag{2.10}$$

with membership functions shown in Fig. 2.

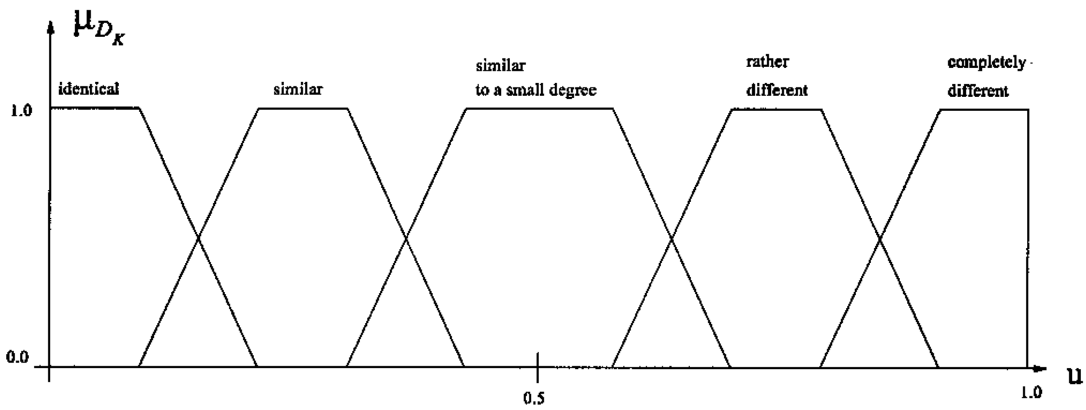

Fig. 2. An example of the linguistic variable *distance of objects* $\boldsymbol{x}$ *and* $\boldsymbol{y}$ consisting of five fuzzy sets $D_1, \ldots, D_5$.

Similarly to the F_j's we assume an ordering on the D_k's

$$(D_1, D_2, \ldots D_m) \tag{2.11}$$

such that D_{k+1} represents a higher degree of distance between $\boldsymbol{x}$ and $\boldsymbol{y}$ than D_k, $1 \leq k \leq m-1$. Formally, we require for each pair of fuzzy sets D_k and D_l, $k < l$, the existence of a real number $0 \leq I \leq 1$ that divides the range $[0,1]$ of the base variable u into two intervals $[0,I]$ and $[I,1]$ such that

$$\begin{aligned} &\mu_{D_k}(u) \geq \mu_{D_l}(u) \text{ if } u \in [0,I], \text{and} \\ &\mu_{D_k}(u) \leq \mu_{D_l}(u) \text{ if } u \in [I,1] \end{aligned} \tag{2.12}$$

The smaller the value u is, the smaller is the distance between $\boldsymbol{x}$ and $\boldsymbol{y}$ (see Fig.2). The cases $u = 0$ and $u = 1$ represent the smallest and largest possible distance (i.e. the largest and smallest degree of similarity), respectively. Similarly to (2.7) we require

$$\mu_{D_1}(u) = 1 \text{ if } u = 0 \tag{2.13}$$

Given the linguistic variables $F(x_1,y_1),\ldots,F(x_n,y_n)$ and $D(\boldsymbol{x},\boldsymbol{y})$, the actual distance between vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ is defined by means of rules. The general format of a rule is

if the similarity of the i_1–th feature is X_1 and/or
the similarity of the i_2–th feature is X_2 and/or
⋮
the similarity of the i_M–th feature is X_M
then the similarity of $\boldsymbol{x}$ and $\boldsymbol{y}$ is Y

or, more formally,

$$\begin{aligned} &\textbf{if } F(x_{i1},y_{i1}) = F_{j1} \oplus_1 \\ &F(x_{i2},y_{i2}) = F_{j2} \oplus_2 \\ &\vdots \\ &F(x_{iM},y_{iM}) = F_{jM} \\ &\textbf{then } D = D_k \end{aligned} \tag{2.14}$$

where each $\oplus_i$ denotes either an 'and' or an 'or' connector. Examples of rules are the following (see also (2.3) and (2.10)):

if the similarity of the 1st feature is *different*
then the objects are *completely different*
if the similarity of the 2nd feature is *identical* and
(the similarity of the 3rd feature is *identical* or
the similarity of the 3rd feature is *similar*)
then the objects are *identical*.

These rules describe a situation where one feature, x_1, is suffcient to classify two objects as completely different, while the identity of the objects can be concluded based on features x_2 and x_3. Rules of this kind provide a very flexible and natural way for human experts to express their knowledge about a certain problem domain.

Once the similarity of both features and cases has been appropriately defined, standard inference techniques can be used to draw conclusions about the similarity of two given objects, $\boldsymbol{x}$ and $\boldsymbol{y}$. In the application described in Section 4, the well-known max-min inference procedure has been used [12]. For a rule of the form given in (2.14) we first determine the degree of membership of all pairs of features $(x_{i1}, y_{i1}), \ldots, (x_{iM}, y_{iM})$ with respect to fuzzy sets $F_{j1}, \ldots, F_{jM}$, respectively, and combine these membership degrees according to the logical connectors $\oplus_1, \ldots, \oplus_{M-1}$. The result is a numerical value $L \in [0,1]$ that represents the degree to which the complete left-hand side of the rule is satisfied. By definition, if any of the features involved in the left-hand side of the rule is undefined, the rule is not applicable. Similarly, we may define a threshold T and declare a rule not applicable if $L < T$. For example, if the minimum-operator is used for 'and' and the maximum-operator for 'or' then the rule in (2.14) evaluates to degree

$$L = OP_{j_{M-1}}(\ldots OP_{j2}(OP_{j1}(\mu_{F_{j1}}, \mu_{F_{j2}}), \mu_{F_{j3}}) \ldots) \tag{2.15}$$

where

$$OP_i = \left\{ \begin{array}{l} \min \text{ if } \oplus_i = \text{and} \\ \max \text{ if } \oplus_i = \text{or} \end{array} \right\}, \; 1 \le i \le j_{M-1}$$

The degree L to which the left-hand side of the rule is satisfied is transferred to the right-hand side in the next step. Under max-min inference, for example, the membership function μ_{D_k} (see (2.14)) is modified into $\tilde{\mu}_{D_k}$ according to the following operation:

$$\tilde{\mu}_{D_k}(u) = \min\{\mu_{D_k}(u), L\}, \; u \in [0,1] \tag{2.16}$$

If fuzzy set D_k is affected by more than one rule, i.e. if it occurs in the right-hand side of several rules, then the corresponding membership functions are combined by the maximum-operator. Formally, if the application of rule j results in the modified membership function $\tilde{\mu}^{(j)}_{D_k}$, and there are totally R rules, i.e., $j = 1, \ldots, R$ then

$$\mu^{total}_{D_k} = \max\{\tilde{\mu}^{(j)}_{D_k} | j = 1, \ldots, R\} \tag{2.17}$$

Furthermore, under max-min inference all membership functions $\mu^{total}_{D_k} : [0,1] \rightarrow [0,1]$ are combined into a single membership function $\mu_D : [0,1] \rightarrow [0,1]$ by means of the maximum operator. Formally,

$$\mu_D(u) = \max\{\mu^{total}_{D_1}(u), \ldots, \mu^{total}_{D_k}(u)\}, \; u \in [0,1] \tag{2.18}$$

The final step of the inference procedure consists of defuzzification. The purpose of this step is to transform the membership function μ_D that is defined over the interval $[0,1]$ into one specific number from that interval corresponding to the distance of objects $\boldsymbol{x}$ and $\boldsymbol{y}$. Several approaches to defuzzification have been proposed in the literature [13]. In the application described in Section 4, defuzzification is achieved by computing

$$\delta(\boldsymbol{x}, \boldsymbol{y}) = \frac{\int_0^1 \mu_D(u) u \, du}{\int_0^1 \mu_D(u) du} \tag{2.19}$$

Obviously, $\delta(\boldsymbol{x}, \boldsymbol{y})$ is the center of gravity of the function $\mu_D(u)$ over the interval $[0, 1]$. Let

$$\delta_{\min} = \delta(\boldsymbol{x}, \boldsymbol{x}) \tag{2.20}$$

be the value that is taken on by $\delta(\boldsymbol{x}, \boldsymbol{y})$ for identical arguments, i.e. $\boldsymbol{x} = \boldsymbol{y}$. Eventually we define the distance between $\boldsymbol{x}$ and $\boldsymbol{y}$ as $d(\boldsymbol{x}, \boldsymbol{y})$ where

$$d(\boldsymbol{x}, \boldsymbol{y}) = \delta(\boldsymbol{x}, \boldsymbol{y}) - \delta_{\min} \tag{2.21}$$

In the next section, some formal properties of the distance measure are studied.

3 Formal Properties

Intuitively, $d(\boldsymbol{x}, \boldsymbol{y})$ can be understood as a distance and a similarity measure at the same time. Small(large) values of the distance $d(\boldsymbol{x}, \boldsymbol{y})$ correspond to a large(small) degree of similarity between $\boldsymbol{x}$ and $\boldsymbol{y}$. From (2.19)-(2.21) we can immediately conclude that

$$0 \leq d(\boldsymbol{x}, \boldsymbol{y}) \leq 1 \tag{3.1}$$

for any pair of objects $\boldsymbol{x}, \boldsymbol{y} \in X$. If (2.8) is satisfied then

$$d(\boldsymbol{x}, \boldsymbol{y}) = d(\boldsymbol{y}, \boldsymbol{x}) \tag{3.2}$$

Moreover, if $\boldsymbol{x} = \boldsymbol{y}$ then

$$d(\boldsymbol{x}, \boldsymbol{y}) = 0 \tag{3.3}$$

On the other hand if $d(\boldsymbol{x}, \boldsymbol{y}) = 0$ then it is not necessarily true that $\boldsymbol{x} = \boldsymbol{y}$. Whether $\boldsymbol{x} = \boldsymbol{y}$ or $\boldsymbol{x} \neq \boldsymbol{y}$ depends on the fuzzy membership functions that define feature and object similarity, the inference technique, and the defuzzification procedure. Clearly, in some applications it may be meaningful to treat the situations $\boldsymbol{x} = \boldsymbol{y}$ and '$\boldsymbol{x}$ similar to $\boldsymbol{y}$' in the same manner, i.e. to have $d(\boldsymbol{x}, \boldsymbol{y}) = 0$ not only for $\boldsymbol{x} = \boldsymbol{y}$ but also if $\boldsymbol{x}$ is similar to $\boldsymbol{y}$.

Next, we show that the distance measure introduced in Section 2 can be used to model the Euclidean distance. By "modeling" we mean the existence of a monotonous function f such that $d_E(\boldsymbol{x}, \boldsymbol{y}) = f(d(\boldsymbol{x}, \ \boldsymbol{y}))$ where $d_E(\boldsymbol{x}, \boldsymbol{y})$ denotes the Euclidean distance, and $d(\boldsymbol{x}, \boldsymbol{y})$ the distance introduced in Section 2. For $\boldsymbol{x}, \boldsymbol{y} \in X$ assume that $x_i, y_i \in \Re$, $1 \leq i \leq N$. Let the linguistic variable $F(x_i, y_i)$ consist of just one fuzzy set F_1 with membership function

$$\mu_{F_1}(x_i, y_i) = \frac{(x_i - y_i)^2}{\triangle \cdot N} \tag{3.4}$$

where

$$\triangle = \max\{(x_i - y_i)^2 | i = 1, \ldots, N; \ \boldsymbol{x}, \boldsymbol{y} \in X\} \tag{3.5}$$

Similarly, let the linguistic variable $D(\boldsymbol{x}, \boldsymbol{y})$ consist of just one fuzzy set D_1 with membership function

$$\mu_{D_1}(u) = 1 \ , \ u \in [0, 1] \tag{3.6}$$

Under this function, the degree of membership is constant over the complete range of u. For the definition of the distance of any two objects $\boldsymbol{x}$ and $\boldsymbol{y}$ we use just one rule :

$$\begin{aligned} &\textbf{if}\, F(x_1, y_1) = F_1(x_1, y_1) \text{ and} \\ &\quad F(x_2, y_2) = F_1(x_2, y_2) \text{ and} \\ &\qquad \vdots \\ &\quad F(x_N, y_N) = F_1(x_N, y_N) \\ &\textbf{then}\, D(\boldsymbol{x}, \boldsymbol{y}) = D_1(\boldsymbol{x}, \boldsymbol{y}) \end{aligned} \tag{3.7}$$

In the inference procedure we do not apply the minimum-operator for the and-connection in rule (3.7), but rather summation. Namely, for any pair of membership functions $\mu_A(u)$ and $\mu_B(u)$ of fuzzy sets A and B defined over the interval [0,1], we let

$$\mu_{A \text{ and } B}(u) = \mu_A(u) + \mu_B(u) \text{ for } u \in [0,1] \tag{3.8}$$

It is easy to see that in the present application always $\mu_{A \text{ and } B}(u) \leq 1$ for all $u \in [0,1]$. Using (3.8), Eq. (2.15) evaluates to

$$L = \sum_{i=1}^{N} \mu_{F_1}(x_i, y_i) = \sum_{i=1}^{N} \frac{(x_i - y_i)^2}{\triangle \cdot N} \tag{3.9}$$

From (2.16), we get

$$\tilde{\mu}_{D_1}(u) = \min\left\{1, \sum_{i=1}^{N} \frac{(x_i - y_i)^2}{\triangle \cdot N}\right\} = \sum_{i=1}^{N} \frac{(x_i - y_i)^2}{\triangle \cdot N} \quad ; \quad u \in [0,1] \tag{3.10}$$

Notice that this function is constant over the complete range of u. As there is only one rule, (2.17) yields

$$\mu_{D_1}^{total}(u) = \tilde{\mu}_{D_1}(u), \tag{3.11}$$

and as the fuzzy linguistic variable D consists of only one membership function, D_1, we conclude from (2.18)

$$\mu_D(u) = \mu_{D_1}^{total}(u)\ ,\ u \in [0,1] \tag{3.12}$$

In contrast with (2.19) we now integrate $\mu_D(u)$ over the interval [0,1] for the purpose of defuzzification. Thus

$$\delta(\boldsymbol{x}, \boldsymbol{y}) = \int_0^1 \sum_{i=1}^{N} \frac{(x_i - y_i)^2}{\triangle \cdot N} du = \sum_{i=1}^{N} \frac{(x_i - y_i)^2}{\triangle \cdot N} \tag{3.13}$$

Because of

$$\delta_{\min} = 0 \tag{3.14}$$

we finally get

$$d(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{\triangle N} \sum_{i=1}^{N} (x_i - y_i)^2 \tag{3.15}$$

Clearly, this function can be used to model the Euclidean distance. Notice that a distance function modeling the weighted Euclidean distance

$$d_{WE}(\boldsymbol{x}, \boldsymbol{y}) = \sqrt{\sum_{i=1}^{N} \alpha_i (x_i - y_i)^2} \tag{3.16}$$

with

$$\sum_{i=1}^{N} \alpha_i = 1 \tag{3.17}$$

can be constructed similarly.

Next, we illustrate, by means of a simple example, that rule-based fuzzy similarity can be used to construct distance measures that do not fulfill the triangular inequality. For simplicity, let $N = 1, \boldsymbol{x} = x, \boldsymbol{y} = y$. Consider some interval $[\alpha, \beta]$ on the real axis with $\beta - \alpha = K,\ K > 0$. Define a distance measure as follows :

$$\gamma(x, y) = \begin{cases} K, & \text{if } \begin{cases} x \in [\alpha, \beta] \text{ and } y \notin [\alpha, \beta], \text{ or} \\ x \notin [\alpha, \beta] \text{ and } y \in [\alpha, \beta] \end{cases} \\ |x - y|, & \text{otherwise} \end{cases} \tag{3.18}$$

Now choose numbers $x, y, z \in \Re$ with $x < \alpha < y < \beta < z$ and $|x - z| > 2K$. Clearly,

$$\gamma(x, y) + \gamma(y, z) < \gamma(x, z) \tag{3.19}$$

Thus γ doesn't fulfill the triangular inequality. Therefore, it is certainly a non-Euclidean distance measure. On the other hand, we can easily construct, by means of rule-based fuzzy similarity, a distance measure that is equivalent to γ. For this purpose let

$$F_1(x, y) = \frac{\gamma(x, y)}{\triangle} \tag{3.20}$$

where

$$\triangle = \max\{\gamma(x, y) \mid x, y \in X\} \tag{3.21}$$

Hence, we can infer similarly to the above derivation that

$$d(x, y) = \frac{1}{\triangle} \cdot \gamma(x, y) \tag{3.22}$$

Thus the class of distance measures that can be represented by rule-based fuzzy similarity properly includes the Euclidean distance.

4 An Application to Thyroid Gland Diagnosis

The similarity measure introduced in Section 2 has been implemented and tested as part of a knowledge based system for the diagnosis of thyroid gland diseases [14]. Other expert systems for this particular application described in the literature include [15], where a rule based approach was taken. In [16], the considered diagnoses were hierarchically structured and the inference procedure was based on Dempster-Shafer theory [17]. A system where diagnostic inferences are made from scintigraphic images of the thyroid gland has been described in [18].

In the inference of a thyroid gland diagnosis, a human expert considers many different sources of information. These include the symptoms reported by the patient, the patient's history, palpation findings, results of laboratory tests, results from image analysis, and others. In contrast with other expert systems where only part of this information is taken into account, the system described in this section tries to expoit all available information, thus 'simulating' the reasoning of a human expert as closely as possible.

The inference of a diagnosis by a human expert usually consists of two phases. First, a few hypotheses, i.e. possible diagnoses, are established based on initial input data. The initial input data, which is easy to acquire, include the patient's symptoms, the patient's history, and palpation findings. Depending on the hypotheses established in the first phase, further information, for example through more expensive laboratory tests and images, is gathered in order to derive the final diagnosis in the second phase.

In the system described in this section, an attempt modeling this two stage approach has been implemented. In order to find plausible hypotheses, given the initial input data, a nearest-neighbor classification schema using the similarity measure introduced in Section 2 has been adopted. For the second phase of the diagnostic inference, other reasoning techniques have been used [14]. The nearest-neighbor approach adopted here can be interpreted as a special instance of case-based reasoning [2]. The basic idea is to store a number of prototypical cases with known diagnoses in a case library. Given an actual case the identity of which is to be determined, we compare it with the case library and retrieve all cases that are similar to it. This approach seems very natural to establishing the set of initial hypotheses. Moreover, the system can acquire additional knowledge by incorporating new cases in the library, and thus it can learn. In the following, we describe this nearest-neighbor classification procedure in more detail.

A case $\boldsymbol{x}$ is represented in the case library as a vector of 27 features. These features are of different nature. Some are numeric, such as the age of a person or the number of nodules detected during palpation. Others are binary symbolic features representing, for example, the patient's sex, or the presence or absence of certain symptoms. The third group of features is symbolic with multiple values. For example, the time elapsed since the first occurrence of a symptom. This feature can take on values from the set {*days, weeks, months, years*}. Some features may be undefined, i.e. unknown.

The similarity of numerical features is defined by means of the presentation in Section 2; for an example see Fig.1. For binary features the linguistic variable *feature similarity* F consists of two fuzzy sets, F_1 = equal and F_2 = different, with membership functions

$$\mu_{F_1}(x,y) = \left\{ \begin{array}{ll} 1 & \text{if } x = y \\ 0 & \text{otherwise} \end{array} \right\}$$

$$\mu_{F_2}(x,y) = \left\{ \begin{array}{ll} 1 & \text{if } x \neq y \\ 0 & \text{otherwise} \end{array} \right\}$$

The feature similarity of non-binary features is dependent on the meaning of the particular feature. For example, for the feature that represents the time elapsed since the first occurrence of a symptom, whose possible values are from the set {*days, weeks, months, years*}, the linguistic variable *distance of feature* consists of five fuzzy sets, F_1 = identical, F_2 = very similar, F_3 = similar, F_4 = rather different, F_5 = different. The membership functions of these fuzzy sets are defined as follows:

μ_{F_1}(day, day) = 1.0
μ_{F_1}(day, week) = 0.2
μ_{F_1}(day, month) = μ_{F_1}(day, year) = 0.0
μ_{F_2}(day, day) = 0.0
μ_{F_2}(day, week) = 0.8
μ_{F_2}(day, month) = μ_{F_2}(day, year) = 0.0
μ_{F_3}(day, day) = 0.0
μ_{F_3}(day, week) = 0.3
μ_{F_3}(day, month) = 0.3
μ_{F_3}(day, year) = 0.0
μ_{F_4}(day, day) = μ_{F_4}(day, week) = 0.0
μ_{F_4}(day, month) = 0.9
μ_{F_4}(day, year) = 0.0
μ_{F_5}(day, day) = μ_{F_5}(day, week) = 0.0
μ_{F_5}(day, month) = 0.1
μ_{F_5}(day, year) = 1.0
μ_{F_1}(week, day) = 0.2
μ_{F_1}(week, week) = 1.0
μ_{F_1}(week, month) = 0.1
μ_{F_1}(week, year) = 0.0
a.s.o.

The linguistic variable *distance of object* - actually *distance of cases* in the present application - consists of five fuzzy sets, D_1 = indentical, D_2 = very similar, D_3 = similar, D_4 = rather different, D_5 = different. Notice that, although these fuzzy sets have the same names as the F_i's that represent feature distance, they are different because they are defined over the interval [0,1].

The rules for the definition of case distance can be grouped into two different categories. The first category is called 'context-free' rules. These rules are valid independent of the diagnostic class of the actual case from the case library that is selected for comparison. An example of such a rule is:

> **if** the function is different and the morphology is different
> **then** the cases are different

The rules in the second category, which are called 'context-dependent', take the diagnostic class of the considered case from the case library into account. These rules are of the general format:

if the diagnostic category of the case from the case library is ... and the distance of feature i_1 is ...

$\vdots$

and the distance of feature i_M is ...
then the distance of the cases is ...

By means of these rules it is possible to directly model the fact that one particular feature may be of high saliency for one particular disease, but may be of no relevance with respect to another disease.

The nearest-neighbor classifier was implemented and run on a Unix- workstation. A case base was collected consisting of 35 cases representing 15 different diagnoses. (According to findings by medical doctors, these 15 diagnoses cover about 90% of all actual cases.) In the first experiment, 28 new cases were classified using the 35 prototypes in the case base. As the primary aim of the nearest-neighbor classification consists of finding a set of possible hypotheses rather than coming up with just one diagnosis, we considered not only the most similar case from the case base, but the n most similar ones for $n = 1, ..., 4$. The result is shown in Table 1 (first row).

Experiment	1st rank	1st ... 2nd rank	1st ... 3rd rank	1st ... 4th rank	below 4th rank
1st	71%	82%	89%	93%	7%
2nd	54%	89%	93%	100%	0%

Table 1: Results of nearest-neighbour classification

In order to study the learning ability of the system, we performed a second experiment. A new set of 18 cases that were not used in the first experiment - neither in the case base nor in the test set - were gathered and classified by the system. Out of these 18 cases, four cases that were not correctly classified (i.e. the correct class was not among the top four ranks) were selected and added to the case base. Thus an enlarged case base of cardinality 39 resulted. Then the original test set used in the first experiment was classified again, but this time based on the enlarged case base. The result is shown in Table 1 (second row). Although the recognition rate in the first rank dropped, we notice an improvement if more than just the first rank is taken into account. Regarding the first four ranks, all 28 test cases were correctly classified.

5 Discussion and Conclusions

A new approach to measuring object similarity is proposed in this paper. It has a high degree of flexibility from both the theoretical and practical point of view. Formally, the proposed measure includes, as a special case, the well-known Euclidean distance. However, it can be used to model other, non-Euclidean distance functions as well. Thus the new similarity measure properly includes the Euclidean distance. From the practical point of view, the new similarity measure can deal with arbitrary mixtures of numerical and symbolic features. It is admissible that some feature values are missing. Thus the measure includes partial similarity as a special case. For each particular feature, an individual similarity function can be defined. The similarity of the individual features is combined into a global object similarity measure by means of rules. Thus the importance of each individual feature can be represented. In particular it is possible to model the situation where the importance of a feature depends on the object's class. Also, potential interdependencies of features can be taken into account. For example, if the degree of similarity of the i-th feature of two objects $\boldsymbol{x}$ and $\boldsymbol{y}$ is known to be S, then feature j may not yield any additional information any longer. However, if the similarity of the i-th feature has value S', then the j-th feature may be of crucial importance to the similarity of $\boldsymbol{x}$ and $\boldsymbol{y}$.

Due to the rule-based approach our proposed similarity measure 'inherits' a number of benefits that are commonly found in rule-based systems. First, it is often natural and easy for human experts to express knowledge in terms of rules. Next, a rule-based representation of object similarity is easy to understand, debug and modify. Moreover, there are well-understood inference techniques available in order to draw conclusions, given a set of rules and some specific input data. Finally, the implementation of rule-based inference is supported by a number of software tools [19].

In the similarity measure investigated here, the similarity of both features and objects is modeled by means of fuzzy linguistic variables and fuzzy sets. There are many application domains where a fuzzy approach is meaningful in order to deal with uncertain information and noisy data. The proposed measure, however, can be used not only in an uncertain (fuzzy) environment, but also in order to represent crisp functions such as Euclidean or other distance functions. It can be used to model both feature and object similarity in qualitative and quantitative terms.

Rule-based fuzzy object similarity is of generic nature with respect to the underlying techniques that deal with fuzzy sets. There are almost no restrictions regarding the fuzzy linguistic variables that define feature and object similarity. Moreover, practically any known technique can be used for the implementation of the fuzzy logical operators that evaluate the degree of satisfaction of a rule. Similarly, any technique for rule inference and defuzzification may be adopted.

Our similarity measure has been applied as part of a case-based reasoning component in a medical expert system for thyroid gland diagnosis. The medical experts involved in the project confirmed the hypothesis that the rule-based formalism is very suitable to express their knowledge of the problem domain, i.e. knowledge of the role of the various features with repect to case similarity. A nearest-neighbor classifier has been implemented using the proposed similarity measure. The task to be solved by this nearest-neighbor classifier was the preselection of possible diagnoses by retrieving a number of candidates from a case library that are similar to the actual case. In an experiment with this classifier, 26 out of 28 cases were correctly ranked among the top four candidates. Enlarging the database from 35 to 39 cases, all of the 28 test cases were among the top four choices. Thus the performance of the nearest-neighbor classifier can be regarded very good in this experiment. Clearly, the data used for training and testing are not sufficient to draw general conclusions about the performance of the proposed approach in other applications. Nevertheless, the result is very encouraging and leads to the conjecture that rule-based fuzzy object similarity can be successfully applied to other problems as well.

References

[1] G. Salton, M.E. Lesk, Computer evaluation of indexing and text processing, Journal of ACM, 15(1), 8-36, 1968

[2] J. Kolodner: Case-Based Reasoning, Morgan Kaufmann Publ. 1993

[3] J. Schuermann: Pattern Classification: A Unified View of Statistical and Neural Approaches, J. Wiley, 1996

[4] B.V. Dasarathy (ed.): Nearest Neighbor Pattern Classification Techniques, IEEE Comp. Soc. Press, 1991

[5] R.A. Wagner, M.J. Fischer: The string-to-string correction problem, Journal of the ACM, 21(1), 168 - 173, 1974

[6] H. Bunke, G. Allermann: Inexact graph matching for structural pattern recognition, Pattern Recognition Letters 1, 245 - 253, 1983

[7] L.G. Shapiro, R.M. Haralick: Structural descriptions and inexact matching, IEEE Trans. Pattern Analysis and Machine Intelligence PAMI 3, 504 - 519, 1981

[8] R. Bareiss, J.A. King: Similarity assessment in case-based reasoning, Proc. Case-Based Reasoning Workshop, Morgan Kaufmann Publ., 1989, 67-71.

[9] M.M. Veloso: Prodigy/Analogy: Analogical reasoning in General Problem Solving, in S.Wess, K.D. Althoff, M. Richter (eds.): Topics in Case-Based Reasoning, Lecture Notes in Art. Intelligence 837, Springer Verlag, 1994, 33-50.

[10] C. Bento, E. Costu: A similarity metric for retrieval of cases imperfectly explained, in S.Wess, K.D. Althoff, M. Richter (eds.): Topics in Case-Based Reasoning, Lecture Notes in Art. Intelligence 837, Springer Verlag, 1994, 92-105.

[11] L.A. Zadeh, Fuzzy sets, Inf. and Control 8, 1965, 338-353

[12] A. Kandel, Fuzzy Mathematical Techniques with Applications, Addison Wesley, 1986

[13] R.R. Yager, D.P. Filev: Essentials of Fuzzy Modeling and Control, Wiley Interscience, New York, 1994

[14] X. Fabregas: An Expert System for Thyroid Gland Diagnosis Including Knowledge-Based Image Analysis of Scinitigraphic Images of Thyroid Gland, PhD Thesis, Computer Science Department, University of Bern, Switzerland, 1994 (in German)

[15] C.A. Kulikowski and J.H. Ostroff. Constructing an expert knowledge base for thyroid consultation using generalized ai techniques, in IEEE Proceedings 4th Annual Symposium on Computer Application, pages 175-180, 1980.

[16] F.L. Degner and R. Santen. DES: A domain independent expert system running on a microcomputer: diagnostic of thyroid disorders. In P.L. Reichertz and D.A. Lindberg, editors, Lecture Notes in Medical Informatics, volume 36, pages 334-337, Berlin, 1988 Springer-Verlag.

[17] G. Shafer: A Mathematical Theory of Evidence, Princeton University Press, 1976

[18] V. Ellam, N. Maisey. A knowledge based system to assist in the diagnosis of thyroid disease from a radioisotope scan. In D. P. Pretschner and B. Urrutia, editors, Knowledge-based Systems to Aid Medical Image Analysis, Volume 1, pages 113-132, Brussels, 1990. Commission of the European Communities.

[19] M. Schneider, A. Kandel, G. Langholz, G. Chew: Fuzzy Expert System Tools, J. Wiley, 1996