Weighted Fuzzy Similarity Classifier in the Łukasiewicz-Structure

KALLE SAASTAMOINEN[1], PASI LUUKKA[1], VILLE KÖNÖNEN[2]
[1]Department of Information Technology
Lappeenranta University of Technology
P.O. Box 20, FIN-53851
FINLAND

[2]Neural Networks Research Centre
Helsinki University of Technology
P.O. Box 5400, FIN-02015 HUT
FINLAND

Abstract: - The aim of this paper is to introduce improvements made to a classifier based on the fuzzy similarity [1]. Improvements are based on the use of generalized Łukasiewicz-structure and weight optimization. We are presenting some new results and a more detailed description of the theoretical background and fixing some terminology compared in to our previous work [2]. The main benefits of the classifier are its computational efficiency and its strong mathematical background. It is based on many-valued logic and it provides semantic information about classification results. We will show that if we choose the power value in appropriate manner in the generalized Łukasiewicz-structure and the optimal weights for different features, we will see significant enhancements in classification results.

Key-Words: - Łukasiewicz-structure, Fuzzy logic, Fuzzy classifier, Similarity measures, Genetic algorithms

1 Introduction

Same way as notion of fuzzy subset generalizes that of the classical subset, the concept of similarity can be considered as a many-valued generalization of the classical notion of equivalence [3]. Equivalence relation is a familiar way to classify similar mathematical objects. Fuzzy similarity is an equivalence relation that can be used to classify multi-valued objects. Because of this, it is suitable for classifying problems that are possible to classify based on clustering by finding similarities in objects. It has been said by Dubois and Prade [4] that the evaluation of similarity between two multi-feature descriptions of objects may be specially of interest in analogical reasoning. If we assume that each feature is associated with an attribute domain equipped with similarity relation modelling approximate equality on this domain, the problem is then to aggregate the degrees of similarity between the objects pertaining to each feature into a global similarity index. This means that the resulting index should still have properties like reflexivity, symmetry and max-$\odot$-transitivity. Moreover, we may think of weighted aggregation if we consider that we are dealing with a fuzzy set of features having different levels of importance.

This article handles the suitability of similarity measures in Łukasiewicz-structure to the pattern recognition problem. A classifier based on the fuzzy similarity can be categorized as a supervised, non-parametric learning method [5]. In this study we have derived a classifier based on the generalized Łukasiewicz-structure, which is more suitable than normal Łukasiewicz-structure for classification tasks. Right choice of power value in a generalized Łukasiewicz-structure seemed to improve classification results remarkable. Also weighting similarity measure, as done in the fuzzy similarity principle, and finding the optimal weights improved classification. Results improved when these two classifiers were combined in a right manner. The data sets were chosen as diverse as possible so that properties of the classifiers would be apparent. Data sets were taken from a UCI-Repository of Machine Learning Database [6] archive so that they were differently distributed and their dimensions varied. Classifiers were implemented with *MATLAB™* -software.

2 Fuzzy Set Theory in Pattern Recognition

Fuzzy set theory is an active research area, highly mathematical in its nature. It can provide a robust and consistent foundation for information processing, including pattern-formatted information processing. It plays at least two roles in the pattern recognition. In one role, it serves as an interface between the linguistic variables seemingly preferred by humans and the quantitative characterizations appropriate for machines. In this role, it might also serve as a bridge between symbolic processing of artificial intelligence and the parallel distributed processing approaches favored by adaptive pattern recognition. In another role, it emphasizes the possibility-distribution interpretation of the concept of fuzziness. The value of this role is that it legitimizes and provides a meaningful interpretation for some distributions that we believe to be useful, but that might be difficult to justify on the basis of the objective probabilities. The two roles are not distinct, but the differences are interesting and worth [7].

3 Use of Fuzzy Set Theory in Pattern Recognition

There are four well-known circumstances in which the concepts and techniques of fuzzy set theory are uniquely helpful in the practice of pattern recognition. One is where a fuzzy set serves as an interface between a linguistically formatted feature (that is, a nonnumeric, symbolic feature) and quantitative measurements. This role of fuzzy sets is well understood, and evidence of its use is widespread, including in the medical diagnostic system, MYCIN [7]. In pattern recognition, there is an additional question of how to aggregate the evidence represented by an array of membership-function values. Different approaches are described in the literature. We call this interface the *first circumstance of use* of fuzzy set theory. Then the *second circumstance of use* is at the class-membership level, rather than at the feature level. In the crisp case, classification consists of relegating a pattern to membership in one of the many possible classes. In the fuzzy set approach, the class membership of a pattern itself is a fuzzy set, and different class indices constitute the support for that fuzzy set. A pattern does not necessarily belong to just one class. There is a certain degree of possibility that the pattern might belong to each one of the classes, and membership functions supply values for these various possibilities. Nothing much is gained if information processing stops at that first step of classification, because ultimately one would have to decide in favor of one specific class, perhaps the one with the largest membership- function value. The different possibilities are of value, however, if the import of the decision propagates into a network of other related decisions. When we have knowledge of the other possibilities, we need not discard or forget meaningful options and alternatives. Fuzzy clustering or the fuzzy ISODATA procedure constitutes an instructive example of this *second circumtance of use*. The *third circumstance of use* is in hand in a case where the membership-function values are used to help provide an estimate of missing or incomplete knowledge. The *fourth circumstance of use* is similar to that of the second, but the context is that of structural rather than geometric pattern recognition. In the parsing of a structural such as sentence, the fuzzy set approach yields values for different possibilities of that structure being due to the action of various production rules. This circumstance of use has been cumbersome and are not generally extensible [7].

4 Łukasiewicz-Structure in Pattern Recognition

We have chosen to use Łukasiewicz-structure in defining memberships of objects. There are two good reasons why we have chosen to use it in defining memberships of objects. One reason is that Łukasiewicz-structure holds the fact that the mean of many fuzzy similarities is still a fuzzy similarity [8]. Secondly it also has a strong connection to the first-order fuzzy logic [9], which is a well studied area in the modern mathematics. Next we will introduce mathematical background concerning fuzzy similarity and generalized Łukasiewicz-structure. After that we can construct algorithms that uses fuzzy similarity in generalized Łukasiewicz-structure as a base of a fuzzy classification method.

4.1 Mathematical Background

Definition 1: (Cartesian Product Space). Let (x, y) be an *ordered pair*, where $x \in X$ and $y \in Y$, the *Cartesian product* is defined as the set:

$$X \times Y = \{(x, y) : x \in X, y \in Y\}. \quad (1)$$

Definition 2: (Binary relation). Any subset $R \subseteq X \times Y$ defines a binary relation between the elements of X and Y:

$$R = \{(x, y) \in X \times Y : R(x, y) \quad holds\} \quad (2)$$

A relation is a multi-valued correspondence:

$$R: X \times Y \to \{0,1\} \quad (x,y) \to R(x,y). \tag{3}$$

Definition 3: A binary relation is a quasi-order if it is reflexive and transitive. If it is also anti-symmetric then binary relation is partial order. If quasi-order is symmetric it is an equivalence relation.

Definition 4: A partially ordered set or poset is a set L on which an order relation $\leq$ has been defined. Of course, on a set L various order relations can be defined. If in a poset L either $x \leq y$ or $y \leq x$ for each $x, y \in L$, then L is linear and is called chain. In such case the order $\leq$ is total order and is called linearly ordered.

Definition 5: A **lattice** is a poset L such that for any $x, y \in L$, $x \wedge y$ and $x \vee y$ exist in L. $x \wedge y$ is called conjuction (meet, infinum) $x \vee y$ is called disjunction (join, supremum) of x and y. A lattice L is a (countable) complete lattice if $\vee \{x \mid x \in X\}$ and $\wedge\{x \mid x \in X\}$ exist in L for any (countable) subset $X \subseteq L$. A lattice is often denoted by $\langle L, \leq, \wedge, \vee \rangle$.

Remark 1: By setting $X = L$ we see that any complete lattice contains the least element **0** and the greatest element **1**.

Example 1: The unit interval I, for example, is a complete lattice under the usual order of $x \vee y = max\{x,y\}$ and $x \wedge y = min\{x,y\}$.

Definition 6: A lattice is called resituated if it contains the greatest element **1**, and binary operations $\odot$ (called multiplication) and $\to$ (called residuum) such that following conditions hold

1. $\odot$ is associative, commutative and isotone
2. a$\odot$ **1** = **a** for all elements a $\in L$ and
3. for all elements a,b,c $\in L$, a $\odot$ b $\leq$ b if and only if $a \leq b \to c$.

Definition 7: Let L be a resituated lattice and X is a non-empty set. L – valued binary relation S defined in X is a fuzzy similarity if it fulfills the following conditions: [3].

1. $\forall x \in X : S\langle x,x \rangle = 1$
2. $\forall x,y \in X : S\langle x,y \rangle = S\langle y,x \rangle$
3. $\forall x,y,z \in X : S\langle x,y \rangle \odot S\langle y,z \rangle \leq S\langle x,z \rangle$

Depending on the choice of the operation $\odot$ (sometimes marked as $*$), S is also called a fuzzy equivalence relation [10], indistinguishability operator [11], fuzzy equality (relation) [12] or proximity relation [4].

It is easy to see that letting L be the two element set $\{0,1\}$, fuzzy similarity coincides with the usual equivalence relation.

Definition 8: **Łukasiewics norm** or **Łukasiewics conjuction**: a $\odot$ b=max{a+b-1,0}.

This is the *t – norm*, which means that it preserves transitivity w.r.t. the triangular inequality. It is also important to realize that for practical applications with the unit interval as the underlying lattice L, considering only MV-algebras is very restrictive, since $a \odot b = \max\{a+b-1,0\}$ is the only choice for the operation $\odot$ up to isomorphism.

We can construct a lattice called **normal Łukasiewicz-structure** or more formally just Łukasiewicz-structure:

Definition 9: Łukasiewics-structure:
$a \odot b = \max\{a+b-1,0\}, \quad a \to b = \min\{1, 1-a+b\}$

If we examine Łukasiewicz-valued fuzzy similarities $S_i \quad i = 1,...,n$ in a set X we can define a binary relation in L by stipulating $S\langle x,y \rangle = \frac{1}{n}\sum_{i=1}^{n} S_i \langle x,y \rangle$ for all $x, y \in X$. It is easy to prove that this is still a Łukasiewicz-valued fuzzy similarity [7].

Definition 10: Letting L be the real unit interval [0,1] endowed with the usual order relation, we may construct the following usual residuated lattice: Generalized Łukasiewicz-structure:
$a \odot b = \sqrt[p]{\max\{a^p + b^p - 1, 0\}}$,
$a \to b = \min\left\{1, \sqrt[p]{1 - a^p + b^p}\right\}$, where p is a fixed natural number.

Now we are ready to study how to use fuzzy similarity for finding similar pairs. We are examining a choice situation where features of different objects can be expressed in values between [0,1]. Let X be the set of m objects. If we know the similarity value of the features $f_1,...,f_n$ between objects, we can choose the object that has the highest total similarity value. The problem is to find for object x_i similar object x_j, where $1 \le i, j \le m$ and $i \ne j$. By choosing Łukasiewicz-structure for features of the objects we get n fuzzy similarities for comparing two objects (x_1, x_2):

$$S_{f_i}\langle x_1, x_2\rangle = x_1(f_i) \leftrightarrow x_2(f_i), \tag{4}$$

where, x_1 , x_2 $\in$ X and i $\in$ {1,…,n}. Because Łukasiewicz-structure is chosen for membership of objects, we can define the fuzzy similarity as follows:

$$S\langle x_1, x_2\rangle = \frac{1}{n}\sum_{i=1}^{n}(x_1(f_i) \leftrightarrow x_2(f_i)). \tag{5}$$

It is a very important to realize that this holds only in the Łukasiewicz-structure. Moreover, in the Łukasiewicz-structure we can give different non-zero weights ($W_1,..,W_n$) to the different features and we get the following formula which again meets the definition of the fuzzy similarity:

$$S\langle x_1, x_2\rangle = \frac{\sum_{i=1}^{n} W_i\left(x_1(f_i) \leftrightarrow x_2(f_i)\right)}{\sum_{i=1}^{n} W_i} \tag{6}$$

In Łukasiewicz-structure equivalence relation $a \leftrightarrow b$ is defined as $1 - \max\{a,b\} + \min\{a,b\}$ or equivalently $1-|a-b|$. The formula of fuzzy similarity in such a case is:

$$S\langle x_1, x_2\rangle = 1 - \frac{1}{n}\sum_{i=1}^{n}|\, x_1(f_i) - x_2(f_i)\,| \tag{7}$$

or

$$S\langle x_1, x_2\rangle = 1 - \frac{\sum_{i=1}^{n} W_i\,|\, x_1(f_i) - x_2(f_i)\,|}{\sum_{i=1}^{n} W_i} \tag{8}$$

We have applied a general form of an equivalence relation $a \leftrightarrow b$ to the Łukasiewicz-structure in the following way.

Proposition 1: In the generalized Łukasiewicz-structure equivalence relation can be defined as $\min\{\sqrt[p]{1-a^p+b^p}, \sqrt[p]{1-b^p+a^p}\}$.

Proof: In the generalized Łukasiewicz-structure holds for implication that $a \rightarrow b = \min\left\{1, \sqrt[p]{1-a^p+b^p}\right\}$. Using this equivalence gets the following form $a \leftrightarrow b = \min\left\{1, \sqrt[p]{1-a^p+b^p}\right\} \wedge \min\left\{1, \sqrt[p]{1-b^p+a^p}\right\}$ and because always $a \le b$ or $a > b$. We get the form $a \leftrightarrow b = \min\{\sqrt[p]{1-a^p+b^p}, \sqrt[p]{1-b^p+a^p}\}$.

The formula of the fuzzy similarity gets now the following form:

$$S\langle x_1, x_2\rangle = \frac{1}{n}\sum_{i=1}^{n}\sqrt[p]{1-\left|x_1^p(f_i) - x_2^p(f_i)\right|} \tag{9}$$

or

$$S\langle x_1, x_2\rangle = \frac{\sum_{i=1}^{n} W_i \sqrt[p]{1-\left|x_1^p(f_i) - x_2^p(f_i)\right|}}{\sum_{i=1}^{n} W_i}. \tag{10}$$

4.2 Description of Pattern Recognition Method in the Algorithmic Form

Method starts with fuzzification of a data set, which means, in practice, scaling the data between zero and one. After this an ideal vector (e.g. mean vector) is calculated for every class. Samples are classified so that the fuzzy similarity value between ideal and test vector is calculated. The sample can be classified to the class with the highest similarity value using generalized Łukasiewicz-structure. The method gets test element *test* (dimension *dim*), learning set *learn* (dimension *dim*, *n* different classes) and dimension (*dim*) of the data as its parameters. In addition, different weights for features can be set in *weights* and value *p* can be set for generalized Łukasiewicz-structure. The method is presented in the pseudo-code form below:

Require: *test,learn[1…n],weights,dim*
 Scale *test* between [0,1]
 Scale *learn* between [0,1]
 For i = 1 to n **do**
 Idealvec[i]=IDEAL[*learn[i]*]

$$\max\ sim[i] = \frac{\sum_{j=1}^{\dim} weights\ [j]\sqrt[p]{1-\left|idealvec\ [i][j]^p - test\ [j]^p\right|}}{\sum_{j=1}^{\dim} weights\ [j]}$$

 end for
Class = arg max$_i$, maxsim[i]

In the algorithm, IDEAL[i] is the vector that best characterizes the class *i*. In this paper, the IDEAL operator is the mean vector of the class. For this algorithm, weights can be optimized for example by using genetic algorithms in optimization.

In Fig. 1 there is a short description in a form of a flowchart for how optimal weights can be found.

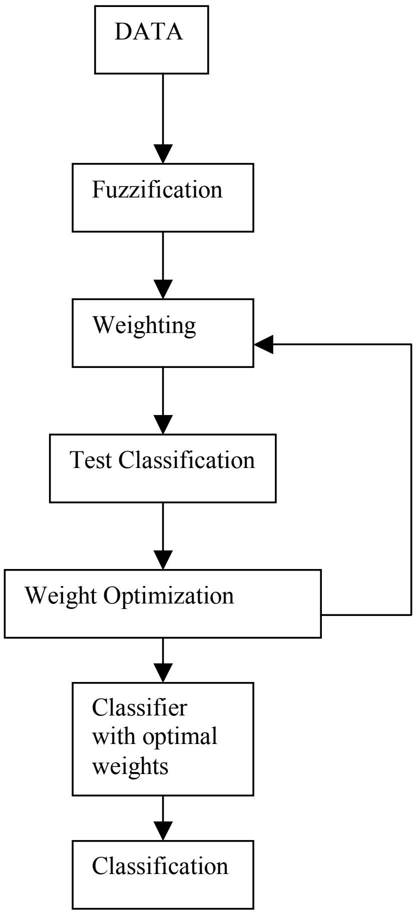

Fig. 1. Optimization of weights using genetic algorithms

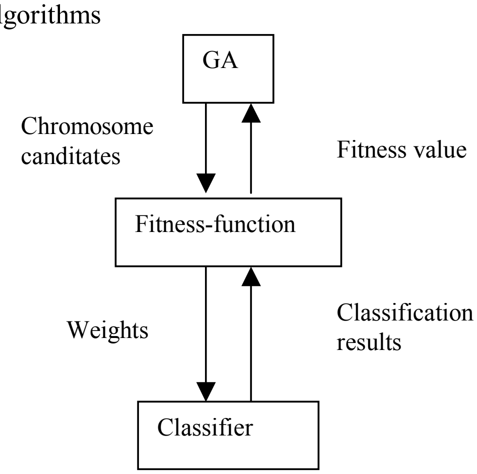

Fig. 2. A more detailed flowchart of how weights are being optimized with GA.

In Fig. 2 there is a closer description for how it would be done by using genetic algorithm.

When the generalized Łukasiewicz-structure is used, there is one additional free parameter present, the power value of terms used in the similarity measure. Genetic algorithm is also used in the optimization of this parameter. This process is outviewed in the Fig. 3.

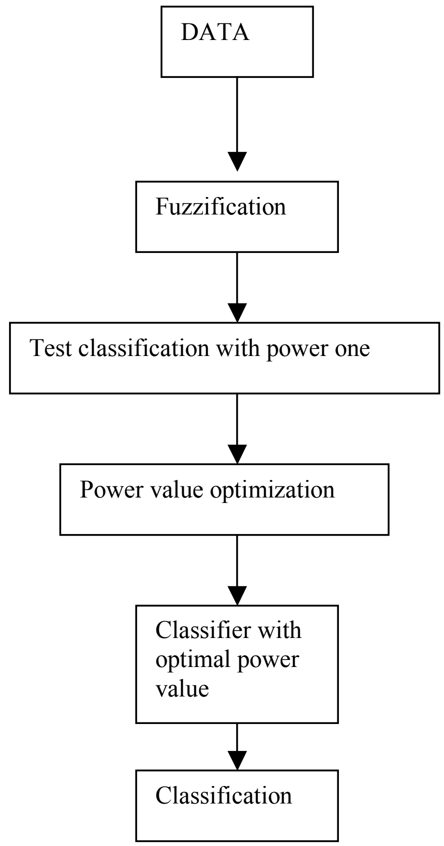

Fig. 3. Power value optimization in the generalized Łukasiewicz-structure

5 Empirical Results

Classification is a good way to test how weighted similarity measure works in practice. We tested our classifier in three phases. At first, only weights in normal Łukasiewicz-structure were optimized. At the next phase, normal Łukasiewicz-structure was extended to the generalized Łukasiewicz-structure and the power value was optimized. At the third phase both weights and power value were optimized and results were compared to well known KNN-classifier.

5.1 Fuzzy Similarity with Weight Optimization in the Task of Classification

A priori information for determining weights was not available for data sets. We used a genetic algorithms, because of their robustness, to find the optimal weights for our fuzzy similarity classifier. Of course other optimizers can be used as well. We used three different data sets that were splitted in half. One half was used for learning and the other half for testing.

Iris-data Set Data set consists of three classes and it is four-dimensional. Classification results were better with the optimally weighted average similarity measure than with the un-weighted similarity measure. The same results can be achieved either with weight optimization or with finding optimal power value.

Thyroid gland-data Set Data consist also of three classes and it is five dimensional. Classification results got much better with optimal weights, but not as good as the classifier in the generalized Łukasiewicz-structure with the optimal power value.

Wine recognition-data Set Data consist of three classes and it is 13 dimensional. Here our optimizer did not find better weight values than the simple average of the features. The reason is that the wine recognition data is linear in its nature. In this case, no additional advantage can be get by using classifier in generalized Łukasiewicz -structure.

Table 1
Classification results with three different data sets and comparison of classification results with fuzzy similarity and normal similarity measure

Class	1	2	3
Wine data unweighted	100%	89%	100%
Wine data weighted	100%	89%	100%
Iris data unweighted	100%	92%	92%
Iris data weighted	100%	96%	92%
Thyroid gland unweighted	100%	67%	67%
Thyroid gland weighted	100%	82%	80%

5.2 Generalized Łukasiewicz-Structure in Classification

We tested the classifier with different power values to three different data sets. Different power values had different effect in classification. One very interesting point was that always power value could be found, that managed to classify classes with the same or better accuracy than by using just power one.

Iris-data Set From results, it can be seen that the best results were achieved with the power of three. Conclusion is that Iris-data is a nonlinear data set and new properties can be extracted by using generalized Łukasiewicz-structure, instead of normal Łukasiewicz -structure.

Thyroid gland-data Set Data set was very interesting because classification results were much better by using generalized Łukasiewicz -structure than by normal Łukasiewicz -structure. Power value which gave the best results was near 0.5.

Wine recognition-data Set Here two optimal power values was found. With powers 1 and 0.5 very similar results was gained. From classification results can be seen that both power values managed to classify two classes 100% correctly, and difference with one class was only one per cent.

Table 2
Comparison for normal Łukasiewicz -structure and generalized Łukasiewicz -structure with optimal power

Class	1	2	3
Wine data (power 1)	100%	89%	100%
Wine data (power 0.5)	100%	90%	100%
Iris data (power 1)	100%	92%	92%
Iris data (power 3)	100%	96%	92%
Thyroid gland (power 1)	100%	67%	67%
Thyroid gland (power 0.5)	100%	94%	80%

5.3 Fuzzy Similarity with Generalized Łukasiewicz -Structure in Classification

After previous experiments we decided to test how our fuzzy similarity classifier would work in generalized Łukasiewicz-structure. The same data sets were used. Results from normal similarity classifier was also included in table 3 for comparison purposes. Here also results were compared to nearest neighbor classifier.

Iris-data Set New information was not found by classifying with fuzzy similarity in generalized Łukasiewicz-structure compared to normal similarity. It managed to get the same results as weighted classifier or classifier with the optimal power value. Three samples from the data set was not classified correctly. In table 3 1-NN classifiers results are also included. As seen from table, third class was classified here better with similarity classifier than nearest neighbor classifier. In first and second class, results were same.

Thyroid gland-data Set Here results got better. We managed to classify two classes 100 per cent correctly. Also results from the third class improved. This data set clearly shows, that there are cases when combining these two methods brings better results compared to using only one of them. Also here the results with similarity classifier compared to 1-NN classifier were much better.

Wine recognition-data Set Here we managed to get results that one of previous methods had already found. Weighting and choosing a correct power value, both managed to find the same new properties in data. New properties was not found by combining these two methods. Compared to nearest neighbor classifier results were much better.

Table 3
Classification results of fuzzy similarity in general Łukasiewicz -structure

Class	1	2	3
Wine data (normal)	100%	89%	100%
Wine data (generalized)	100%	90%	100%
Wine data (1-NN)	83%	85%	46%
Iris data (normal)	100%	92%	92%
Iris data (generalized)	100%	96%	92%
Iris data (1-NN)	100%	96%	88%
Thyroid gland (normal)	100%	67%	67%
Thyroid gland (generalized)	100%	100%	83%
Thyroid gland (1-NN)	97%	89%	69%

6 Conclusions

Study of fuzzy similarity classification has been made. Also comparison to well-known nearest neighbor classifier was done and conclusions can be drawn. In this study we have shown that the fuzzy similarity is very useful in classification and it better results can be achieved than with traditional nearest neighbor classifier. Fuzzy similarity has clearly something new to offer in classification compared to traditional similarity measure. Another interesting result was that normal Łukasiewicz-structure is not ideal for similarity measure in classification but better results can be achieved by using generalized Łukasiewicz-structure. Generalized Łukasiewicz -structure should be used instead of normal Łukasiewicz-structure, if the boundaries are not simple and clear. Fuzzy similarity classifier managed to achieve best results by using generalized Łukasiewicz-structure. It always managed to get at least the same results as in normal Łukasiewicz-structure. The major advantages of the method is that it provides semantic information about the classification task by allowing partial membership of the class.

Interesting point to study next is to develop an algorithm that calculated similarities to its neighborhood. Class of the sample is based on this information. Very interesting from mathematical point of view is to study relations between different metrics and similarity measures. In that way it might be possible to study connections between fuzzy similarity classifier and self-organizing maps.

Acknowledgment

This work was partially supported by East Finland Universities Graduate School in Computer Science and Engineering (ECSE).

References:

[1] Könönen, V., Luukka, P., Saastamoinen, K.: *New Classifier Based on Maximal Fuzzy Similarity.* Paper sent for review to Fuzzy Sets and Systems

[2] Luukka, P., Saastamoinen, K., Könönen, V.: *A Classifier Based on the Maximal Fuzzy Similarity in the Generalized Łukasiewicz -structure*
Paper published in the FUZZ-IEEE 2001 Conference.

[3] Zadeh, L.*: Similarity Relations and Fuzzy Ordering.* Inform Sci, 3, 1971

[4] Dubois, D., Prade, H.*: Similarity-Based Approximate Reasoning.* in: 'Computational Intelligence Imitating Life.' Zurada, J.M., Marks II, R.J., Robinson, C.J., eds., IEEE Press, New York, 1994

[5] Duda, R., Hart, P.*: Pattern Classification and Scene Analysis* John Wiley \& Sons, 1973

[6] UCI Repository of Machine Learning Databases network document. Referenced 4.11.2000. Available: *ftp://ftp.ics.uci.edu/pub/machine-learning-databases*

[7] *Pao, Y.-H.: The Pattern Recognition and Neural Networks* Addison-Wesley Publishing Company, Inc, 1989

[8] Turunen, E.: *Mathematics behind Fuzzy Logic.* Advances in Soft Computing, Physica-Verlag, Heidelberg, 1999

[9] Novak, V.*: On the Syntactico-semantical Completeness of First-Order Fuzzy Logic.* Kybernetika 26, 1990

[10] Klawonn, F., Castro J.L.: *Similarity in Fuzzy Reasoning.* Math Soft Comp, 2, 1995

[11] Trillas, E., Valverde L.: *An Inquiry into Indistinguishability Operators.* in: 'Aspects of Vaguenes.' Skala, H.J., Termini, S., Trillas, E., eds., Reidel, Dordrecht, 1984

[12] Höhle, U., Stout, L.N.: *Foundations of Fuzzy Sets.* Fuzzy Sets and Systems, 40, 1991