Similarity Classifier Using Similarities Based on Modified Probabilistic Equivalence Relations

Pasi Luukka

Laboratory of Applied Mathematics
Lappeenranta University of Technology
P.O. Box 20, Lappeenranta, Finland
Email: pasi.luukka@lut.fi

Abstract

This paper examines a classifier based on similarity measures originating from probabilistic equivalence relations with a generalized mean. Equivalences are weighted and weight optimization is carried out with differential evolution algorithms. In the classifier, a similarity measure based on the Łukasiewicz structure has previously been used, but this paper concentrates on measures which can be considered to be weighted similarity measures defined in a probabilistic framework, applied variable by variable and aggregated along the features using a generalized mean. The weights for these measures are determined using a differential evolution process. The classification accuracy with these measures are tested on different data sets. Classification results are obtained with medical data sets, and the results are compared to other classifiers, which gives quite good results. The result presented in this paper are promising, and in several cases better results were achieved.

Key words: Similarity, Generalized mean, Classification, Probabilistic equivalence relation, Differential evolution

1 Introduction

Because similarity is an equivalence relation that can be used to classify multi-valued objects, it is suitable for classifying problems that can be classified based on clustering by finding similarities in objects. In instance-based classification methods, the selection of a similarity measure is critical. Some times poor results obtained from instance-based methods may actually originate in the underlying similarity measure, not the method itself.

The motivation for this paper is to study similarities based on probabilistic equivalence relations introduced in [4] in the task of classification. This paper introduces
a bit more general way of applying these measures in classification by using a gen-
eralized mean. Also a systematic study on parameter sensitivity is carried out and
the method is applied to several data sets. Quite often only an arithmetic mean is
used, but obviously also other means can be applied. Many different means which
could be used instead of the arithmetic one, can be found in literature, and here it
is shown that in some cases they give better results. The mean in this study is a
generalised mean holding the arithmetic, harmonic and geometric means as spe-
cial cases even though they are the most common means. Previously we have used
in the classifier [6], [7], [8], [14], [15] fuzzy similarity based on the Łukasiewicz
structure [5], which is a popular similarity measure. However, here it is shown that
these similarities originating from probabilistic equivalence relations also give very
promising results.

The data sets were chosen so that their dimensions varied and they were as diverse
as possible in order for the properties of the classifiers to be apparent. The data sets
were taken from a UCI Repository of Machine Learning Database archive (avail-
able in [10]) in order for them to be differently distributed and their dimensions
varied. The classifiers were implemented with the MATLAB$^\text{TM}$ software.

2 Mathematical Background

In the classifier, the main measures are a number of connectivities which are based,
in a certain sense, on probabilistic considerations. Detailed information about these
measures can be found in [4]. These measures can be considered to be weighted
similarity measures defined in a probabilistic framework, applied variable by vari-
able and aggregated along the features using a generalized mean. The weights are
determined using a differential evolution process. The measures are tested on dif-
ferent data sets.

Before going into detail on how the classifier is built, let us first consider the math-
ematical background of the measures used starting from fuzzy relations.

Fuzzy relations

**Definition 1** Let $X$ and $Y$ be nonempty sets. A fuzzy relation $R$ is a fuzzy subset of
the Cartesian product $X \times Y$.

If $X = Y$, we say that $R$ is a binary fuzzy relation on $X$. Let $R$ be a binary fuzzy
relation in $\mathbb{R}$, then $R(x, y)$ is interpreted as the degree of membership $\mu_R(x, y)$ of the
ordered pair $(x, y)$ in $R$. 

2
Fuzzy relations are very important because they can describe the strength of interactions between variables.

**Similarity relation**

**Definition 2** A fuzzy relation \( S \) on \( X \) is called a similarity relation on \( X \) if it fulfills the following conditions [9]:

1. **reflexivity**: \( \forall x_1 \in X : \mu_S(x_1, x_1) = 1 \)
2. **symmetry**: \( \forall x_1, x_2 \in X : \mu_S(x_1, x_2) = \mu_S(x_2, x_1) \)
3. **transitivity**: \( \forall x_1, x_2, x_3 \in X : \mu_S(x_1, x_3) \geq \max_{x_2} \{\min\{\mu_S(x_1, x_2), \mu_S(x_2, x_3)\}\} \)

A fuzzy binary relation that is reflexive, symmetric and transitive is known as a fuzzy equivalence relation or similarity relation [3].

While an equivalence relation clearly groups elements that are equivalent under the relation into disjoint classes, the interpretation of a similarity relation can be approached in two different ways. First, it can be considered to effectively group elements into crisp sets whose members are 'similar' to each other to some specified degree. Obviously, when this degree is equal to 1, the grouping is an equivalence class. Alternatively, however, we may wish to consider the degree of similarity that the elements of \( X \) have to some specifier element \( x \in X \). Thus, for each \( x \in X \), a similarity class can be defined as a fuzzy set in which the membership grade of any particular element represents the similarity of that element to the element \( x \). If all the elements in the class are similar to \( x \) to the degree of 1 and similar to all elements outside the set to the degree of 0, then the grouping again becomes an equivalence class. See [11] and [3] for more information.

Just as equivalence classes are defined by an equivalence relation, similarity classes are defined by a similarity relation. The similarity class for each \( x_1 \in X \) is a fuzzy set in which the membership grade of each element \( x_2 \in X \) is simply the strength of that element's relation to \( x_1 \), or \( R(x_1, x_2) \). Thus, the similarity class for an element \( x_1 \) represents the degree to which all the other members of \( X \) are similar to \( x_1 \). Except in the restricted case of equivalence classes themselves, similarity classes are fuzzy [3].

As an example, we can construct a lattice called the normal Łukasiewicz structure [5] or more formally simply the Łukasiewicz structure:
Definition 3  Łukasiewicz structure: $x_1 \odot x_2 = \max \{x_1 + x_2 - 1, 0\}$, $x_1 \Rightarrow x_2 = \min\{1, 1 - x_1 + x_2\}$.

Where $\odot$ and $\Rightarrow$ are conjunction and implication. In the Łukasiewicz structure, the similarity relation $x_1 \Leftrightarrow x_2$ is defined as

$$x_1 \Leftrightarrow x_2 = 1 - |x_1 - x_2|.$$  \hspace{1cm} (1)

Good sources of information on the Łukasiewicz structure are [5], [16] and [17].

In the classifier, a choice situation where the features of different objects can be expressed in values between $[0,1]$ is examined. Let $X$ be a set of $m$ objects. If we know the similarity value of the features $x_1, ..., x_n$ between the objects, we can choose the object that has the highest similarity value. The problem is finding a similar object $x_i$, for object $x_j$ where $1 \leq i, j \leq m$ and $i \neq j$. By choosing the Łukasiewicz structure for the features of the objects we get $n$ similarities for comparing two objects $(x_1, x_2)$:

$$S(x_1, x_2) = \mu_1(x_i) \Leftrightarrow \mu_2(x_i),$$  \hspace{1cm} (2)

where $x_1, x_2 \in X$ and $i \in \{1, ..., n\}$.

If we examine similarities, $S_i, i = 1, ..., n$ in a set $X$, we can define a binary relation in $L$ by stipulating $S(x_1, x_2) = \frac{1}{n} \sum_{i=1}^{n} S_i(x_1, x_2)$ for all $x_1, x_2 \in X$. Reference [16] states that this is still a Łukasiewicz-valued similarity. There is also a close relation between the notion of similarity and that of distance (see for example [2] and [18]).

As the Łukasiewicz structure is chosen for the membership of objects, we can define the similarity as follows:

$$S(x_1, x_2) = \frac{1}{n} \sum_{i=1}^{n} (\mu_1(x_i) \Leftrightarrow \mu_2(x_i)).$$  \hspace{1cm} (3)

Moreover, we can give different non-zero weights $(w_1, \ldots, w_n)$ to the different features and obtain the following equation:

$$S(x_1, x_2) = \sum_{i=1}^{n} w_i (\mu_1(x_i) \Leftrightarrow \mu_2(x_i)).$$  \hspace{1cm} (4)
In the generalized Łukasiewicz structure, the equivalence relation can be defined as [6] [3]:

\[ x_1 \Leftrightarrow x_2 = (1 - |x_1^p - x_2^p|)^{1/p}. \]  

(5)

Also weights can be set to the similarity measure, and it can now be defined as follows [6]:

\[ S(x_1, x_2) = \sum_{i=1}^{n} w_i \sqrt[1]{1 - |(\mu_1(x_i))^p - (\mu_2(x_i))^p|}. \]  

(6)

### 2.1 Weighted similarities based on probabilistic equivalence relations

In this part, a number of similarities are introduced. The similarities are weighted versions of the connectivities introduced in [4]. The connectivities are based, in a certain sense, on probabilistic considerations. The connectivities are canonically derived from the connectives of classical logic by means of a simple probabilistic "black box": Weighted versions of equivalences are introduced and called similarities based on the mentioned equivalence. For information about connectivities, see [4]. Here we simply introduce the formulae which are used in classification.

P-probabilistic extension can be stated as [4]:

\[ x_1 \Leftrightarrow x_2 = 1 - x_1 - x_2 + 2x_1x_2. \]  

(7)

From here, simply by taking an arithmetic mean and adding weights, we get the following equation:

Similarity based on P - probabilistic extension of equivalence:

\[ S_{pe}(x_1, x_2) = \frac{1}{n} \sum_{i=1}^{n} w_i (1 - \mu_1(x_i) - \mu_2(x_i) + 2\mu_1(x_i)\mu_2(x_i)). \]  

(8)

One can also again take the generalized mean

\[ M_m(a_1, a_2, \ldots, a_n) = \left( \frac{1}{n} \sum_{i=1}^{n} a_k^m \right)^{1/m} \]  

(9)
after which the equation gets the following form:

\[ S_{P_{pe}}(x_1, x_2) = \left( \frac{1}{n} \sum_{i=1}^{n} w_i (1 - \mu_1(x_i) - \mu_2(x_i) + 2\mu_1(x_i)\mu_2(x_i))^m \right)^{1/m} \]  \hspace{1cm} (10)

The next three equations are created in a same manner.

**Similarity based on Y - probabilistic extension of equivalence(Y2pe):**

\[ S_{Y2_{pe}}(x_1, x_2) = \max(A_1(x_1, x_2), B_1(x_1, x_2)) \]  \hspace{1cm} (11)

where

\[ A_1(x_1, x_2) = \left( \frac{1}{n} \sum_{i=1}^{n} w_i (1 - \mu_1(x_i) - \mu_2(x_i))^m \right)^{1/m} \] and

\[ B_1(x_1, x_2) = \left( \frac{1}{n} \sum_{i=1}^{n} w_i (3 - \mu_1(x_i) - \mu_2(x_i) - 2\sqrt{(1 - \mu_1(x_i))^2 + (1 - \mu_2(x_i))^2} \right)^{1/m} \]

**Similarity based on DP1/2 - probabilistic connective of equivalence (DP1/2pc):**

\[ S_{DP1_{2pc}}(x_1, x_2) = \frac{A_2(x_1, x_2)B_2(x_1, x_2)}{\max(A_2(x_1, x_2), B_2(x_1, x_2), \frac{1}{2})} \]  \hspace{1cm} (12)

where

\[ A_2(x_1, x_2) = \left( \frac{1}{n} \sum_{i=1}^{n} w_i (1 - \mu_1(x_i) + \mu_2(x_i) - \frac{(1 - \mu_1(x_i))\mu_2(x_i)}{\max(1 - \mu_1(x_i), \mu_2(x_i), \frac{1}{2})} \right)^{1/m} \] and

\[ B_2(x_1, x_2) = \left( \frac{1}{n} \sum_{i=1}^{n} w_i (1 - \mu_2(x_i) + \mu_1(x_i) - \frac{(1 - \mu_2(x_i))\mu_1(x_i)}{\max(1 - \mu_2(x_i), \mu_1(x_i), \frac{1}{2})} \right)^{1/m} \]

**Similarity based on E - probabilistic connectives for equivalence (Epc):**

\[ S_{E_{pc}}(x_1, x_2) = \frac{A_3(x_1, x_2)B_3(x_1, x_2)}{C_1(x_1, x_2)} \]  \hspace{1cm} (13)
In future, we simply use the short names $P_{pe}$, $DP_{1/2pc}$, $Epc$ and $Y2_{pe}$ for these measures.

3 The Classifier

Next, we describe the general type of classification problem we are dealing with and the structure of the classifier.

We would like to classify set $X$ of objects to the $N$ different classes $C_1, \ldots, C_N$ by their features. We suppose that $D$ is the number of different kinds of features $f_1, \ldots, f_D$ that we can measure from objects. We suppose that the values for the magnitude of each feature are normalized so that they can be presented as a value between $[0, 1]$. So, the objects we want to classify are basically vectors that belong to $[0, 1]^D$.

First, we determine for each class the ideal vector $v_i = (v_i(1), \ldots, v_i(D))$ that presents class $i$ as well as possible. This vector can be user defined or calculated from some sample set $X_i$ of vectors $x = (x(1), \ldots, x(D))$ which are known to belong to class $C_i$. A generalized mean is used in this article for calculating $v_i$, which is

$$v_i(d) = \left( \frac{1}{|X_i|} \sum_{x \in X_i} x(d)^m \right)^{1/m}, \quad \forall d = 1, \ldots, D$$

(14)

where power value $m$ is fixed for all $i, d$. Once the ideal vectors have been determined, the decision to which class the arbitrarily chosen $x \in X$ belongs is made by comparing it to each ideal vector. We have made that comparison by using the
previously introduced measures. Similarity measure can be defined as e.g.

\[
S(\mathbf{x}, \mathbf{y}) = \left( \frac{1}{D} \sum_{d=1}^{D} w_d (1 - x(d) - y(d) + 2x(d)y(d))^m \right)^{1/m} \tag{15}
\]

for \( \mathbf{x}, \mathbf{y} \in [0, 1]^d \). The parameter value \( m \) and weights \( w_d \in [0, 1] \) for each dimension are supposed to be fixed. We decide that \( \mathbf{x} \in C_m \) if

\[
S(\mathbf{x}, \mathbf{v}_m) = \max_{i=1,...,N} S(\mathbf{x}, \mathbf{v}_i) . \tag{16}
\]

In the example, the similarity relation based on the \( P \)-probabilistic extension is used. However, it can be replaced with any other similarity measure in a very straightforward way. Weights for these measures can be optimized for example by using the differential evolution algorithm in optimization. The differential evolution algorithm has been introduced in [12].

3.1 Differential evolution

The basic idea of evolutionary algorithms is that we create a population \( V_0 \) of trial solutions (vectors) for the optimization problem. Next, we combine the members of \( V_0 \) in a certain way and check if the combined solutions are better than the original trial solutions in \( V_0 \). The best solutions continue to population \( V_1 \), the next step of the evolution, and then the whole procedure starts again.

Differential evolution (DE) can be considered as quite a new algorithm, since it was developed around 1995 by Storn and Price [12]. Unlike genetic algorithms, DE works directly with continuous variables without the encoding and decoding of vectors. We denote the population at evolution step \( n \) by set \( V_n \). Unlike in GA, our DE algorithm is written for minimizing problems, so fitness function \( f : \mathbb{R}^D \rightarrow [0, 1] \) is now set so that if \( f(\mathbf{v}) < f(\mathbf{u}) \) then \( \mathbf{v} \) is a better solution for our problem than \( \mathbf{u} \).

A basic DE contains also crossover operation, but it has no real mutation operation. Instead of mutation, DE demonstrates differential variation operation. In one evolution step, we carry out the following two basic operations for all \( \mathbf{v}_n = \{v_1, \ldots, v_D\} \in V_n \):

1. **Differential variation.** The basic idea of differential variation is to add "noise vector" \( \mathbf{n} \) to vector \( \mathbf{w} \in V_n \). Both \( \mathbf{n} \) and \( \mathbf{w} \) can be chosen in many ways. Usually, \( \mathbf{n} \) is either the difference vector between two vectors in \( V_n \) or a linear combination of these kinds of difference vectors. From \( \mathbf{n} \) and \( \mathbf{w} \) we form a new vector

\[
\mathbf{u} = \mathbf{w} + F\mathbf{n} , \tag{17}
\]
where usually parameter $F$ is constant and between 0 and 1.

2. Crossover. From $u = \{u_1, \ldots, u_D\}$ and $v_n$ we form a trial vector $t = \{t_1, \ldots, t_D\}$ by setting

$$t_i = \begin{cases} u_i & \text{if } x_i < CR \\ v_i & \text{otherwise} \end{cases}$$

where $x_i \in [0, 1]$ is a uniformly distributed random variable and $CR$ is the crossover probability. Usually values $F = 0.9$ and $CR = 0.8$ work quite well. Finally the vector $v_{n+1}$ which is chosen to the next generation $V_{n+1}$ is

$$v_{n+1} = \begin{cases} t & \text{if } f(t) < f(v) \\ v_n & \text{otherwise} \end{cases}$$

and $w$ was chosen as the best member of $V_n$,

$$f(w) = \min_{v \in V} \{f(v)\},$$

and $n$ was the difference vector

$$n = y_1 - y_2,$$

where $y_1 \in V_n$ and $y_2 \in V_n$ were randomly chosen vectors.

3.2 Data Sets

Next, we will briefly describe the different data sets used in this study. The data sets are all freely available in [10]. Descriptions are written using information available from the data sets. The fundamental properties of the data sets are shown in Table 1.

Liver disorders:
This data was contributed by R. S. Forsyth to [10]. The problem is to predict whether or not a male patient has a liver disorder based on blood tests and alcohol consumption. The attribute information for liver disorder data is the following 1) mean corpuscular volume 2) alkaline phosphotase 3) alamine aminotransferase 4) aspartate aminotransferase 5) gamma-glutamyl transpeptidase 6) number of half-pint equivalents of alcoholic beverages drunk per day. The first five variables are all blood tests which are thought to be sensitive to liver disorders that might arise from excessive alcohol consumption.

Echocardiogram:
In this data set, one sample consists of results from echocardiogram measurements
of one patient who has recently suffered an acute heart attack. Measurements are taken from echocardiograms, which are ultrasound measurements of the heart itself. The goal of physicians using these measurements is to predict a patient’s chances of survival. Experimental work is being performed to determine if an echocardiogram, in conjunction with other measures, could be used to predict whether or not a patient would survive for longer than a certain time period. If this data can be classified accurately enough, it would give means for predicting future heart attacks in former heart patients.

**Hepatitis:**
The purpose of this data set is to predict whether a patient will survive from the decease or not. There are 19 attributes measured from the patient, and the number of instances is altogether 155.

**PIMA Indians:**
The Pima Indian data set concerns the presence or absence of diabetes among Pima Indian women living near Phoenix, Arizona. There are eight covariates: number of pregnancies; plasma glucose concentration; diastolic blood pressure (mmHg); triceps skin fold thickness (mm); serum insulin (µU/ml); body mass index (kg/m²); diabetes pedigree function and age in years.

**Breast cancer data set:**
The breast cancer data set based on image analysis work (see [19]) was done in early 1990. The goal was to diagnose the sample based on a digital image of a small section of the FNA slide. Nuclear size, shape and texture features are used to distinguish benign from malignant breast cytology. The benign and malignant cell samples were obtained by fine needle aspiration from consecutive series of 569 patients: 212 with cancer and 357 with fibrocystic breast masses [20].

4 Classification results

The results of the classification can be seen in Table 2, which reports mean classification accuracies, variances and also the optimal $m$ value where results were found for these four measures. Data was split in half; one half for the training set and the other half for the testing set. The data was split in half randomly 30 times, and mean accuracies from these computations are reported. A sensitivity test for parameter $m$ was also carried out to see how results changed with respect to changes with the $m$
value. In Figures 1 – 2, mean classification accuracies are plotted with respect to changes in the $m$ value.

As can be seen from Table 2, the highest mean accuracy with the liver-disorder data set was gained with $Y_{2pe}$, the mean accuracy being 68.96% and the variance 0.001. With the Pima-indian diabetes data set, the highest mean accuracy was 76.28%, this time with the measure $E_{pc}$. The variance in this case was 0.0003. With the hepatitis data set measure $P_{pe}$ had the highest mean accuracy of 84.44%, the variance being 0.0008. With the echocardiogram data, the highest mean accuracy was 87.40% and again this was gained using the measure $Y_{2pe}$. The variance in this case was 0.0026. With the breast cancer data set, again measure $Y_{2pe}$ managed to classify with the highest mean accuracy, this time 96.56, the variance being 0.0001.

As can be seen from the results, $DP_{1/2pc}$ usually performed the most poorly with these four measures and $Y_{2pe}$ seemed often to give the highest mean classification accuracy. Although $Y_{2pe}$ seems to do well with these data sets, one should also notice that sometimes the best mean accuracy was gained with $E_{pc}$ or $P_{pe}$. Which measure works best seems to be data dependent, and even though $Y_{2pe}$ did well with these particular data sets, one should not disregard the other measures in trying them on a new data set. As for choosing the best $m$ value; for these data sets the optimal $m$ value for each measure is reported in Table 2. In Figures 1 – 2, classification results are also plotted with respect to $m$ value changes. As can be seen from the figures and table, the best $m$ value changes quite considerably when the data set is changed. Therefore $m$ values are also data dependent. In general, it seems that positive $m$ values usually give higher mean accuracies, but sometimes, as can be seen from Table 2, the best results are gained with negative $m$ values. What is notable in the $m$ values is that they are almost never $-1$, 0 or 1 when the generalized mean equals the harmonic mean, approaches the geometric mean or equals the arithmetic mean. This indicates that the generalized mean should be used instead of e.g. the arithmetic mean and proper parameter value $m$ should be found.

Classification results are compared to other results in Table 3. In Table 3, the results are compared to those reported by [1]. As can be seen, higher results are gained using these probabilistic extensions in four cases out of five when compared to DIMLP, MLP and CN2 classifiers.

Next, these measures are compared to similarity based on the Łukasiewicz structure
which is a measure that has previously been used with this classifier [8]. As can be seen from Table 4, $Y_{2pe}$ had the highest accuracy in three data sets out of five, and $P_{pe}$ and $E_{pc}$ had the highest mean accuracy in one data set each.

[Table 4 about here.]

5 Conclusions

In this study, we have shown that similarities based on probabilistic equivalence relations and the generalized mean are very useful in classification. When applied to the above-mentioned medical data, quite good results were achieved. The major advantage of the method is that it provides semantic information about the classification task by allowing partial membership in the class.

We also demonstrated that optimal results were almost never achieved with the common arithmetic mean, but the use of the generalized mean enhanced the classification results. Usually, negative means did not work as well as positive means, but at times negative means gave better results.

The measure that seemed to give the best results was similarity based on the $Y_2$ - probabilistic extension of equivalence. This was the case with the data sets on liver disorder, echocardiograms and breast cancer. Similarity based on $E$ probabilistic connectives for equivalence performed the best with Pima Indians diabetes data set, and similarity based on the $P$ probabilistic extension of equivalence classified with the highest accuracy in the hepatitis data set.

Classification results were compared to results obtained with classifiers DIMLP, MLP and CN2, and in four cases out of five a higher accuracy was achieved. When compared to the similarity measure based on the Łukasiewicz structure measures, $Y_{2pe}$, $E_{pc}$ and $P_{pe}$ also compare well.

References


List of Figures

1 Classification results with respect to the parameter $m$ value mean classification accuracies and variances reported a) & b) with liver disorders data c) & d) with Pima Indian diabetes data 16

2 Classification results with respect to the parameter $m$ value mean classification accuracies and variances reported a) & b) with hepatitis data c) & d) with echocardiogram data 17
Fig. 1. Classification results with respect to the parameter $m$ value mean classification accuracies and variances reported a) & b) with liver disorders data c) & d) with Pima Indian diabetes data
Mean classification accuracies with hepatitis data

<table>
<thead>
<tr>
<th>m-value</th>
<th>Classification accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>-15</td>
<td>0.50</td>
</tr>
<tr>
<td>-10</td>
<td>0.55</td>
</tr>
<tr>
<td>-5</td>
<td>0.60</td>
</tr>
<tr>
<td>0</td>
<td>0.65</td>
</tr>
<tr>
<td>5</td>
<td>0.70</td>
</tr>
<tr>
<td>10</td>
<td>0.75</td>
</tr>
</tbody>
</table>

Variances with hepatitis data

<table>
<thead>
<tr>
<th>m-value</th>
<th>Classification accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>-15</td>
<td>0.005</td>
</tr>
<tr>
<td>-10</td>
<td>0.01</td>
</tr>
<tr>
<td>-5</td>
<td>0.015</td>
</tr>
<tr>
<td>0</td>
<td>0.02</td>
</tr>
<tr>
<td>5</td>
<td>0.025</td>
</tr>
<tr>
<td>10</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Mean classification accuracies with echocardiogram data

<table>
<thead>
<tr>
<th>m-value</th>
<th>Classification accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>-15</td>
<td>0.50</td>
</tr>
<tr>
<td>-10</td>
<td>0.55</td>
</tr>
<tr>
<td>-5</td>
<td>0.60</td>
</tr>
<tr>
<td>0</td>
<td>0.65</td>
</tr>
<tr>
<td>5</td>
<td>0.70</td>
</tr>
<tr>
<td>10</td>
<td>0.75</td>
</tr>
</tbody>
</table>

Variances with echocardiogram data

<table>
<thead>
<tr>
<th>m-value</th>
<th>Classification accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>-15</td>
<td>0.005</td>
</tr>
<tr>
<td>-10</td>
<td>0.01</td>
</tr>
<tr>
<td>-5</td>
<td>0.015</td>
</tr>
<tr>
<td>0</td>
<td>0.02</td>
</tr>
<tr>
<td>5</td>
<td>0.025</td>
</tr>
<tr>
<td>10</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Fig. 2. Classification results with respect to the parameter m value mean classification accuracies and variances reported a) & b) with hepatitis data c) & d) with echocardiogram data
**List of Tables**

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Test data sets and their properties.</td>
<td>19</td>
</tr>
<tr>
<td>2</td>
<td>Classifier results with measures $DP_{1/2pc}$, $Ep_c$, $Y_{2pe}$ and $P_{pe}$. Mean classification accuracies, variances and the best $m$ value are reported.</td>
<td>20</td>
</tr>
<tr>
<td>3</td>
<td>Classifier models</td>
<td>21</td>
</tr>
<tr>
<td>4</td>
<td>Classification result comparison to the similarity measure based on the Łukasiewicz structure and similarity derived from probabilistic equivalence relations.</td>
<td>22</td>
</tr>
<tr>
<td>Name</td>
<td>Nb. classes</td>
<td>Dim</td>
</tr>
<tr>
<td>-------------------------------------</td>
<td>-------------</td>
<td>-----</td>
</tr>
<tr>
<td>PIMA data</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>Liver disorders data</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>Breast cancer Wisconsin diagnostic data</td>
<td>2</td>
<td>30</td>
</tr>
<tr>
<td>Hepatitis</td>
<td>2</td>
<td>19</td>
</tr>
<tr>
<td>Echocardiogram</td>
<td>2</td>
<td>7</td>
</tr>
</tbody>
</table>
Table 2
Classifier results with measures $DP_{1/2pc}$, $E_{pc}$, $Y_{2pe}$ and $P_{pe}$. Mean classification accuracies, variances and the best $m$ value are reported.

<table>
<thead>
<tr>
<th>Data Set</th>
<th>$DP_{1/2pc}$</th>
<th>$E_{pc}$</th>
<th>$Y_{2pe}$</th>
<th>$P_{pe}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liver-disorder</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean accuracy</td>
<td>59.09%</td>
<td>68.50%</td>
<td>68.96%</td>
<td>67.26%</td>
</tr>
<tr>
<td>Variance</td>
<td>0.0007</td>
<td>0.0008</td>
<td>0.0010</td>
<td>0.0007</td>
</tr>
<tr>
<td>$m$ value</td>
<td>−0.1</td>
<td>−4.1</td>
<td>3.9</td>
<td>9.9</td>
</tr>
<tr>
<td>Pima</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean accuracy</td>
<td>69.82%</td>
<td>76.28%</td>
<td>74.18%</td>
<td>76.22%</td>
</tr>
<tr>
<td>Variance</td>
<td>0.0012</td>
<td>0.0003</td>
<td>0.0005</td>
<td>0.0003</td>
</tr>
<tr>
<td>$m$ value</td>
<td>0.9</td>
<td>0.9</td>
<td>1.9</td>
<td>0.9</td>
</tr>
<tr>
<td>Hepatitis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean accuracy</td>
<td>81.84%</td>
<td>83.38%</td>
<td>82.95%</td>
<td>84.44%</td>
</tr>
<tr>
<td>Variance</td>
<td>0.0420</td>
<td>0.0021</td>
<td>0.0011</td>
<td>0.0008</td>
</tr>
<tr>
<td>$m$ value</td>
<td>0.9</td>
<td>−2.1</td>
<td>−0.1</td>
<td>2.0</td>
</tr>
<tr>
<td>Echocardiogram</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean accuracy</td>
<td>77.46%</td>
<td>85.76%</td>
<td>87.40%</td>
<td>85.31%</td>
</tr>
<tr>
<td>Variance</td>
<td>0.0087</td>
<td>0.0015</td>
<td>0.0026</td>
<td>0.0030</td>
</tr>
<tr>
<td>$m$ value</td>
<td>−0.1</td>
<td>7.9</td>
<td>2.9</td>
<td>0.9</td>
</tr>
<tr>
<td>Breast cancer</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean accuracy</td>
<td>93.30%</td>
<td>96.14%</td>
<td>96.56%</td>
<td>96.40%</td>
</tr>
<tr>
<td>Variance</td>
<td>0.0003</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.1451</td>
</tr>
<tr>
<td>$m$ value</td>
<td>2.9</td>
<td>1.0</td>
<td>4.9</td>
<td>2.0</td>
</tr>
</tbody>
</table>
Table 3  
Classifier models

<table>
<thead>
<tr>
<th>Data set</th>
<th>DIMLP</th>
<th>MLP</th>
<th>CN2</th>
<th>(E_{pc})</th>
<th>(Y_{2_{pe}})</th>
<th>(P_{pe})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liver disorders</td>
<td>70.1</td>
<td><strong>70.2</strong></td>
<td>65.5</td>
<td>68.5</td>
<td>69.0</td>
<td>67.3</td>
</tr>
<tr>
<td>Pima Indians</td>
<td>75.4</td>
<td>75.5</td>
<td>74.2</td>
<td><strong>76.3</strong></td>
<td>74.2</td>
<td>76.2</td>
</tr>
<tr>
<td>Hepatitis</td>
<td>79.1</td>
<td>79.7</td>
<td>81.8</td>
<td>83.4</td>
<td>83.0</td>
<td><strong>84.4</strong></td>
</tr>
<tr>
<td>Echocardiogram</td>
<td>68.5</td>
<td>65.7</td>
<td>68.6</td>
<td>85.8</td>
<td><strong>87.4</strong></td>
<td>85.3</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>96.5</td>
<td>96.5</td>
<td>93.8</td>
<td>96.1</td>
<td><strong>96.6</strong></td>
<td>96.4</td>
</tr>
</tbody>
</table>
Table 4
Classification result comparison to the similarity measure based on the Łukasiewicz structure and similarity derived from probabilistic equivalence relations.

<table>
<thead>
<tr>
<th>Data set</th>
<th>Łukasiewicz sim.</th>
<th>$D_{P_{1/2}PC}$</th>
<th>$E_{PC}$</th>
<th>$Y_{2_{PE}}$</th>
<th>$P_{PE}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liver disorders</td>
<td>66.6</td>
<td>59.1</td>
<td>68.5</td>
<td><strong>69.0</strong></td>
<td>67.3</td>
</tr>
<tr>
<td>Pima Indians</td>
<td>75.3</td>
<td>69.8</td>
<td><strong>76.3</strong></td>
<td>74.2</td>
<td>76.2</td>
</tr>
<tr>
<td>Hepatitis</td>
<td>83.6</td>
<td>81.8</td>
<td>83.4</td>
<td>83.0</td>
<td><strong>84.4</strong></td>
</tr>
<tr>
<td>Echocardiogram</td>
<td>87.2</td>
<td>77.5</td>
<td>85.8</td>
<td><strong>87.4</strong></td>
<td>85.3</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>95.1</td>
<td>93.3</td>
<td>96.1</td>
<td><strong>96.6</strong></td>
<td>96.4</td>
</tr>
</tbody>
</table>