Using Entropy-Based Local Weighting to Improve Similarity Assessment

Héctor Núñez[1], Miquel Sànchez-Marrè[1], Ulises Cortés[1], Joaquim Comas[2], Ignasi Rodriguez-Roda[2], and Manel Poch[2]

[1]Knowledge Engineering & Machine Learning group, Computer Software Dept., Technical University of Catalonia, Campus Nord-Edifici C5, Jordi Girona 1-3, 08034 Barcelona, Catalonia, EU
{hnunez, miquel, ia}@lsi.upc.es
[2]Chemical and Environmental Engineering Laboratory, University of Girona, Campus de Montilivi, Lluís Santaló s/n,
17071 Girona, Catalonia, EU
{quim, ignasi, manel}@lequia.udg.es

Abstract. This paper enhances and analyses the power of local weighted similarity measures. The paper proposes a new entropy-based local weighting algorithm to be used in similarity assessment to improve the performance of the CBR retrieval task. It has been carried out a comparative analysis of the performance of unweighted similarity measures, global weighted similarity measures, and local weighting similarity measures. The testing has been done using several similarity measures, and some data sets from the UCI Machine Learning Database Repository and other environmental databases.

1 Introduction

Over the last decade an important progress has been made in the Case-based Reasoning (CBR) field. Specially, because problems are more clearly identified, and research results have led to real applications where CBR performs well. As noticed in [2], in this situation, it has also become clear that a particular strength of CBR over most other methods is its inherent combination of problem solving with sustained learning through problem solving experience. In CBR, similarity is used to decide which instance is closest to a new current case, and similarity measures have attracted the attention of many researchers in the field.

Theoretical frameworks for the systematic construction of similarity measures have been described in [16], [15], [17], [14] and [4]. Other research work introduced new measures for a practical use in CBR systems, such as Bayesian distance measures in [9] and some heterogeneous difference metrics in [21]. Also, a review of some used similarity measures was done in [12]. On the other hand, there are many works than propose discretization as a methodology to deal with continuous features [10], [6], [21].

This paper aims to analyse and to study the performance of several commonly used measures in addition with a discretization pre-process, global feature weighting, and local feature weighting. The measures are evaluated in terms of predictive accuracy on unseen cases, measured by a ten-fold cross-validation process. In this comparative analysis, we have selected two basic similarity measures (Euclidean and Manhattan), two unweighted similarity measures (Clark and Canberra), two heterogeneous similarity measures (Heterogeneous Value Difference Metric and Interpolated Value Difference Metric) and an exponential weighted similarity measure called *L'Eixample*. Although all these are distance measures, we can refer to similarity measures by means of the relation:

$$SIM(x,y) = 1 - DIST(x,y)$$

1.1 Related Work

One core task of CBR retrieval is similarity computation. Many researchers have been trying to improve similarity assessment. In later years, many of them are focusing on feature weighting. Feature weighting is a very important issue. It is intended to give more relevance to those features detected as important, and at the same time, it is intended to give lower importance to irrelevant features. Most general methods for feature weighting use a global scheme. It means to associate a weight to all the space of the feature (e.g., [13] and [8]). On the other hand, some local scheme methods have been proposed, such as assigning a different weight to each value of a feature ([18]), or allowing feature relevance to vary across the instances ([5]), or setting weight according to the class value ([7]), or using a different similarity metric for each test case ([3]), or combining both schemes ([1]). In this paper, it is argued that a feature may be irrelevant in some subspace of values, but in other subspaces of values it can correctly predict the class of an example. If a continuous attribute is present, a discretization pre-process is suggested and a specific weight is associated to each resulting range or interval. If the attribute is discrete, a specific weight is associated to each possible value. The importance of one specific value (or range) will be determined by the distribution of the class values for that feature value (or range). In other words, if the entropy of one value (or range) respect the class value is high, a low weight is set. On the other side, a high weight is associated to a value (or range) showing a low entropy.

[20] Presents a five-dimension framework for feature weighting: bias, weight space, representation, generality and knowledge. Taking into account that point of view, our approach can be classified as Bias: Preset, Weight Space: Continuous, Representation: Given, Generality: Local and Knowledge: Poor.

The paper is organised in the following way. Section 2 outlines main features about the new entropy-based local weighting algorithm. In Section 3, background information on selected similarity measures to carry out the study is provided. Section 4 presents the results comparing the performance of all distance measures using diverse criteria to handle continuous attributes and feature weights tested on ten randomly-selected databases from the UCI Machine Learning Repository and other databases. Finally, in Section 5 conclusions and future research directions are outlined.

2 Entropy-Based Local Weighting

A retrieval method should try to maximise the similarity between the current case and the retrieved one(s). And this task usually implies the use of general domain knowledge. *Selecting* the best similar case(s), it is usually performed in most CBR systems by means of some evaluation heuristic functions or distances, possibly domain dependent. Commonly, each attribute or dimension of a case has a determined importance value (weight), which is incorporated in the evaluation function. This weight could be static or dynamic depending on the CBR system purposes. Also, the evaluation function computes an absolute match score (a numeric value), although a relative match score between the set of retrieved cases and the new case can also be computed.

Most CBR systems represent cases as a plain structure composed by a set of attribute-value pairs, $x=\{x_1, x_2,\ldots, x_n, x_c\}$ where n is a set of (numeric or symbolic) features, and x_c is x's class value. In such situation, these systems use a generalised weighted distance function, which can be described as:

$$dist(C_i, C_j) = \frac{\sum_{k=1}^{n} w_k * atr\_dist(C_{i,k}, C_{j,k})}{\sum_{k=1}^{n} w_k}$$

In our approach, the continuous attributes have to be discretised, Afterwards, a specific weight is assigned, both to discretised continuous attributes or discrete attributes. Next, the discretis ation method used (CAIM algorithm), and the new entropy-based local weighting algorithm are described.

2.1 Discretisation

Some of the similarity measures have a good performance when the attributes are all continuous or all discrete. Others incorporate mechanisms to deal appropriately all the types of attributes. Our proposal is to make a discretization pre-process on the continuous attributes in such way that the general accuracy can be improved [6]. Discretization may serve to mark differences that are important in the problem domain. There exist many discretization algorithms in the literature, and had been compared among them to prove their general accuracy [6], [19]. To improve the retrieval accuracy, a *global* and *supervised* method to discretize all the continuous attributes, the CAIM algorithm proposed by Kurgan and Cios in [10], was selected. This algorithm tries to maximise the dependency relationship between the class label and the continuous-values attribute, and at the same time, to minimise the number of discrete intervals. In our approach all the continuous attributes were divided in a number of intervals equal to the number of present classes in the database, or in 5 intervals when the number or present classes are less than 5. The Class-Attribute Interdependency Maximisation (CAIM) criterion which measures the dependency between the class variable C and the discretization variable D for attribute F is defined as:

$$CAIM(C,D|F) = \frac{\sum_{i=1}^{n} \frac{\max_i^2}{M_{ir}}}{n}$$

where: n is the number of intervals

i iterates through all intervals, i.e. i=1,2,…n

max_i is the maximum value among all q_{ir} values (maximum value within the i^{th} column of the quanta matrix), r=1,2,…,S

M_{ir} is the total number of continuous values of attribute F that are within the interval $(d_{r-1},d_r]$

Table 1. Quanta matrix. Frequency matrix for attribute F and discretization scheme D

Class	Interval					Class
	$[d_0,d_1]$	…	$[d_{r-1},d_r]$	…	$[d_{n-1},d_n]$	Total
C_1	q_{11}	…	q_{1r}	…	q_{1n}	M_{1+}
:	:	:	:	:	:	:
C_i	q_{i1}	…	q_{ir}	…	q_{in}	M_{i+}
:	:	:	:	:	:	:
C_S	q_{S1}	…	q_{Sr}	…	q_{Sn}	M_{S+}
Interval	M_{+1}	…	M_{+r}	…		M

The CAIM criterion is a heuristic measure that quantifies the interdependence between classes and the number of unique values of the continuous attribute. For a complete details about the CAIM algorithm see [10].

2.2 Entropy-Based Local Weighting Algorithm

Many works has been done concerning assigning weights to features [Wettschereck and Aha], [Jarmulak et al.], [13], but most of them uses a *global* weight for the features. Frequently, a single feature may seem irrelevant if you take it in a global way, but perhaps, a range of this feature is a very good selector for a specific class. Our proposal is to assign a high weight to this range, and a low weight to the others. In the tests that have been carried out, entropy values have been used to assign weights to all the ranges. In the following paragraph we present our approach to calculate *local* weights for all the values (ranges) for all the attributes is detailed. The calculated values are in the range from 0 to 10 in ascending order of relevance. In our approach, a correlation matrix is filled for each attribute, represented the correlation between attribute's values and class value as show in Table 2.

Table 2. Correlation Matrix.

	C_1	C_2	…	C_n	Tot x Val
V_1	q_{11}	q_{12}	…	q_{1n}	q_{1+}
V_2	q_{21}	q_{22}	…	q_{2n}	q_{2+}
:	:	:	…	:	:
V_m	q_{m1}	q_{m2}	…	q_{mn}	q_{m+}
Tot x Cla	q_{+1}	q_{+2}	…	q_{+n}	q_{++}

Where:

- V_m is the *m* value of the attribute, when the attribute is continuous, V_m represents one interval.
- C_n is the class value *n*
- q_{mn} is the number of instances that have value *m* (or are in range *m* for a continuous attribute) and belong to class *n*.
- q_{+n} is the number of instances of class *n*.
- q_{m+} is the number of instances that have value *m* (or are in range *m* for a continuous attribute).
- q_{++} is the number of instance in the training set.

From this matrix, we can obtain the entropy from each value (range):

$$H_{ij} = -\frac{q_{j+}}{q_{++}} \sum_{k=1}^{n} \frac{q_{jk}}{q_{j+}} \log\left(\frac{q_{jk}}{q_{j+}}\right)$$

This entropy H_{ij} belonging to value (range) *j* from attribute *i* will be the basis to calculate the weight for the value *j* following this simple idea: if the value (or range) has a maximum possible entropy ("totally random"), then the weight must be 0. On the other hand, if the value (or range) has a minimum possible entropy ("perfectly classified") then the weight must be 10. The minimum possible entropy is 0 when all the instances with this value (range) belong to the same class, the maximum possible entropy occurs when the instances with this value (range) are equally distributed in all classes and can be calculated as follows:

$$H_{imax} = -\left(\frac{q_{j+}}{q_{++}}\right) \sum_{k=1}^{n} \left(\frac{\frac{q_{j+}}{n}}{q_{j+}}\right) \log\left(\frac{\frac{q_{j+}}{n}}{q_{j+}}\right)$$

This equation is equivalent to:

$$H_{i\max} = -\frac{q_{j+}}{q_{++}} \log\left(\frac{1}{n}\right)$$

From here, we can interpolate the weight for attribute *i* value (range) *j* between 0 and H_{imax} into the range from 0 to 10:

$$w_{ij} = 10 - \operatorname{int}\left(\frac{H_{ij}}{H_{i\max}}\right) * 10$$

3 Similarity measures

Currently, there are several similarity measures that have been used in CBR systems, and some comparison studies exist among these similarity measures (see [21] and [12]). The results obtained in these studies show that the different similarity measures have a performance strongly related to the type of attributes representing the case and to the importance of each attribute. Thus, is very different to deal with only lineal or quantitative data (continuous), with discrete or qualitative (entire) or nominal (qualitative not ordered). To give a greater distance contribution to a more important attribute than other less important attributes is necessary, too. In this study, the new proposed paradigm, discretization pre-process and local weights assignment, is performed to explore how some wide used similarity measure improves the retrieval task in CBR systems, comparing our approach against classical retrieval evaluation with global weights, with no weights, and in some cases with no discretization. The selected similarity measures were:

3.1 Measures Derived from Minkowski's Metric

$$d(x_i, x_j) = \left(\sum_{k=1}^{K} \left|x_{ik} - x_{jk}\right|^r\right)^{1/r} \quad r \geq 1$$

Where k is the number of input attributes. When r=1, *Manhattan* or *City-Block* distance function is obtained. If r=2, *Euclidean* distance is obtained. When including weights for all the attributes, the general formula becomes the following:

$$d(C_i, C_j) = \left(\frac{\sum_{k=1}^{K} weight_k^r * \left|d(A_{ik}, A_{jk})\right|^r}{\sum_{k=1}^{K} weight_k^r}\right)^{1/r}$$

Where for not ordered attributes, their contribution to the distance is,

$$d(A_{ik}, A_{j,k}) = 1 - d\left(qlv(A_{ik}), qlv(A_{jk})\right)$$

and δ is the δ of Kronecker.

3.2 Unweighted Similarity Measures

In this study two similarity measures that ignore attribute's weight were included:

Clark:

$$d(x_i, x_j) = \sum_{k=1}^{K} \frac{|x_{i,k} - x_{j,k}|^2}{|x_{i,k} + x_{j,k}|^2}$$

and *Canberra*:

$$d(x_i, x_j) = \sum_{k=1}^{K} \frac{|x_{i,k} - x_{j,k}|}{|x_{i,k} + x_{j,k}|}$$

3.3 Heterogeneous Similarity Measures

To obtain a broader study and results, other two distance measures that show very high values of efficiency have been included. These functions were proposed in [21] :

Heterogeneous Value Difference Metric (HVDM):

$$HVDM\,(i,\ j) = \sqrt{\sum_{a=1}^{m} d_a^{\,2}(x_a, y_a)}$$

Where m is the number of attributes. The function $d_a(x_a,y_a)$ returns a distance between the two values x and y for attribute a, and is defined as:

$$d_a^2(x_a, y_a) = \begin{cases} 1, & if\ x\ or\ y\ is\ unknown; otherwise \\ normalized\_vdm_a(x, y), & if\ a\ is\ nomina \\ normalized\_diff_a(x, y), & if\ a\ is\ linear \end{cases}$$

Where $normalized\_vdm_a(x,y)$, is defined as follows:

$$normlized\_vdm(x, y) = \sqrt{\sum_{c=1}^{C} \left| \frac{N_{a,x,c}}{N_{a,x}} - \frac{N_{a,y,c}}{N_{a,y}} \right|^2}$$

Where:

- $N_{a,x}$ is the number of instances that have value x for attribute a;
- $N_{a,x,c}$ is the number of instances that have value x for attribute a and output class c;
- C is the number of output classes in the problem domain

The function $normalized\_diff_a(x,y)$, is defined as showed below:

$$normalized\_diff_a(x,y) = \frac{|x-y|}{4\sigma_a}$$

where σ_a is the standard deviation of the numeric values of attribute a.

Interpolated Value Difference Metric (IVDM):

$$IVDM(x,y) = \sum_{a=1}^{m} ivdm_a(x_a, y_a)^2$$

Where $ivdm_a$ is defined as:

$$ivdm_a(x,y) = \begin{cases} vdm_a(x,y) & \text{if } a \text{ is discrete} \\ \sum_{c=1}^{C} \left| p_{a,c}(x) - p_{a,c}(y) \right|^2 & \text{otherwise} \end{cases}$$

where $vdm_a(x,y)$ is defined as follows:

$$vdm_a(x,y) = \sum_{c=1}^{C} \left| P_{a,x,c} - P_{a,y,c} \right|^2$$

C is the number of classes in the database. $P_{a,x,c}$ is the conditional probability that the output class is c given that attribute a has the value x. And:

$$P_{a,x,c} = \frac{N_{a,x,c}}{N_{a,x}}$$

Where $N_{a,x}$ is the number of instances that have value x for attribute a; $N_{a,x,c}$ is the number of instances that have value x for attribute a and output class c.

$P_{a,c}(x)$ is the interpolated probability value of a continuous value x for attribute a and class c, and is defined:

$$P_{a,c}(x) = P_{a,u,c} + \left(\frac{x - mid_{a,u}}{mid_{a,u+1} - mid_{a,u}} \right) * \left(P_{a,u+1,c} - P_{a,u,c} \right)$$

In this equation, $mid_{a,u}$ and $mid_{a,u+1}$ are midpoint of two consecutive discretized ranges such that $mid_{a,u} \leq x < mid_{a,u+1}$. $P_{a,u,c}$ is the probability value of the discretized range u, which is taken to be the probability value of the midpoint of range u. The value of u is found by first setting $u=discretize_a(x)$, and then subtracting 1 from u if $x<mid_{a,u}$. The value of $mid_{a,u}$ can be found as follows:

$$mid_{a,u}= min_a + width_a * (u+.5)$$

3.4 Exponential Weighted Similarity Measure

L'Eixample distance measure.

Some distance functions, sometimes, does not capture the significant differences among the attributes. They are a *lineal* weighted combination of one-dimensional distances that can lose significant differences as the number of attribute increases.
After a competence study, a *normalised exponential weight-sensitive distance function* was developed and was named as *L'Eixample* distance [17], [14]. It was thought that a non-lineal weighted multi-dimensional distance function would be required for a better matching performance. It takes into account the different nature of the quantitative or qualitative values of the lineal (ordered) attributes, and the modalities of categorical (not ordered) attributes.
L'Eixample distance is sensitive to weights. For the most important attributes, that is weight > α, the distance is computed based on their qualitative values, i.e. maintaining or amplifying the differences between cases. And for those less relevant ones, that is weight ≤ α, the distance is computed based on their quantitative values, i.e. reducing the differences between cases. *L'Eixample distance* used to rank the best cases is:

$$d(C_i,C_j)=\frac{\sum_{k=1}^{n} e^{w_k}\times d(A_{ki},A_{kj})}{\sum_{k=1}^{n} e^{w_k}}$$

where

$$d(A_{ki},A_{kj})=\begin{cases}\frac{|quantval(A_{ki})-quantval(A_{kj})|}{upperval(A_k)-lowerval(A_k)} & \text{if } A_k \text{ is an ordered attribute and } w_k \le a\\ \frac{|qualval(A_{ki})-qualval(A_{kj})|}{\#\mathrm{mod}(A_k)-1} & \text{if } A_k \text{ is an ordered attribute and } w_k > a\\ 1-d_{qualval(A_{ki}),qualval(A_{kj})} & \text{if } A_k \text{ is a non-ordered attribute}\end{cases}$$

and,

C_i is the case i; C_j is the case j; W_k is the weight of attribute k; A_{ki} is the value of the attribute k in the case i; A_{kj} is the value of the attribute k in the case j; qtv(A_{ki}) is the quantitative value of A_{ki}; qtv(A_{kj}) is the quantitative value of A_{kj}; A_k is the attribute k; upperval(A_k) is the upper quantitative value of A_k; lowerval(A_k) is the lower quantitative value of A_k; α is a cut point on the weight of the attributes; qlv(A_{ki}) is the qualitative value of A_{ki}; qlv(A_{kj}) is the qualitative value of A_{kj}; #mod(A_k) is the number of modalities (categories) of A_k; δqlv(A_{ki}),qlv(A_{kj}) is the δ of Kronecker.

4 Experimental Evaluation

To test the efficiency, all similarity measures were tested with no weights, with global weights and with the new entropy-based local-weighting approach. A nearest neighbour classifier was implemented, using each one of the 7 distance measures: HVDM, IVDM, Euclidean, Manhattan, Clark, Canberra and *L'Eixample*. Each distance measure was tested, with the three weighting schemes, in the 10 selected databases from the UCI database repository plus 2 other environmental databases. Detailed description of the databases is shown in table 3 where number of instances in each database (#Inst.), the number of continuous attributes (Cont), ordered discrete attributes (Disc), not ordered discrete attributes (NODisc), number of classes (#Class) and missing values percentage (%Mis.).

Table 3. Major properties of databases considered in the experimentation

Database	Short Name	#Inst	Cont	Disc	NODisc	#Class	%Miss
Breast Cancer	Br.Can.	699	0	9	0	2	0
Glass	Glass	214	9	0	0	7	0
Hepatitis	Hepat.	155	6	0	13	2	5.7
Ionosphere	Ionosp.	351	34	0	0	2	0
Iris	Iris	150	4	0	0	3	0
Liver Disorders	Liver D	345	6	0	0	2	0
Pima Indians Diabetes	Pima	768	8	0	0	2	0
Soybean (large)	Soyb. L	307	0	6	29	19	21.7
Votes	Votes	435	0	0	16	2	7.3
Zoo	Zoo	90	0	0	16	7	0
Air Pollution	Air P.	365	5	0	0	4	0
WasteWaterTreat.Pl.	WWTP	793	14	0	1	24	35.8

To verify the accuracy of the retrieval in a CBR system, a test by means of a 10-fold cross-validation process was implemented. The average accuracy over all 10 trials is reported for each data test, for each similarity measure, and for each weighting scheme. The highest accuracy achieved in each data set for the three weighting schemes is shown in boldface in tables 4a and 4b.

4.1 Normalisation

A weakness that most of the similarity measures show is that if one of the attributes has relatively large range of values, can hidden the meaning of the other attributes when the distance is computed. To avoid this effect, the contribution to the distance of each attribute is normalised, and the common way of doing it is to divide the distance of each attribute by the range (maximum value – minimum value) of the attribute. Thus, the contribution of each attribute to the total distance will be in the rank of 0..1. In the tests carried out in all the databases, the values were normalised for all the continuous attributes in the computation of the Euclidean, Manhattan and *L'Eixample* distance measures. Canberra and Clark distances make a type of normalisation avoiding that attributes influence into others. HVDM and IVDM make a normalisation by means of the standard deviation of the numeric values of the attributes.

4.2 Missing Values

In Euclidean, Manhattan, Clark, Canberra and *L'Eixample* distance measures, a pre-processing task was carried out to substitute the missing input values by the average value obtained of the instances with valid values. This was done for all the attributes. In the case of HVDM, a distance of 1 is given when one of the values compared or both are unknown. IVDM treats the unknown values as any another value. Thus, if the two values compared are both missing, the distance between them is 0.

4.3 Global Weighting Algorithm

Although there are some global weighting schemes in the literature, a new approach has been designed. To implement the global weighting algorithm we have used a new approach based on estimated probabilities and correlation. The information present in the correlation matrix (table 2) has been used looking for the maximum value at each column (q_{+i}). This value divided by the number of the instances belonging to class i, represent the best prediction of the class i in all the feature space. The main idea is to put together this information for all the class values, in such way that the global weight of the attribute will be higher in the same proportion that the prediction was higher. The computation of the prediction is as follows:

$$H_a = \frac{1}{n}\sum_{i=1}^{n}\frac{q_{max,i}}{q_{+i}}$$

where,

n is the number of classes
q_{+i} is the total of instances belonging to class i
$q_{max,i}$ is the maximum value of the column i

The number of different values of the feature biases this value, so that, when there are a few values, the lower limit of the prediction will be higher, such as 0.5 for two different values. In fact, the lowest limit will be $1/|a|$, where $|a|$ is the number of different feature values. With this in mind, H_a is escalated to obtain the global weight in the rank 0..10 for attribute a:

$$W_a = \text{int}\left(\frac{H_a - \frac{1}{|a|}}{1-\frac{1}{|a|}} * 10\right)$$

Table 4a. Generalisation accuracy results.

Similarity Measure		Databases					
Name	**Weight**	**Br.Can.**	**Glass**	**Hepat.**	**Ionosp.**	**Iris**	**Liver Dis.**
Clark	**N/A**	96.55	62.75	81.89	84.03	95.99	65.11
Canberra	**N/A**	96.33	67	80.11	89.46	94.70	61.76
Euclidean	**No**	95.46	67.59	79.59	83.21	95.28	61.19
	Global	96.35	72.05	78.40	84.34	95.32	61.74
	Local	**95.90**	**77.19**	**81.41**	**92.06**	**95.90**	**64.80**
Manhat-tan	**No**	**96.33**	70.43	79.59	90.61	**93.95**	61.78
	Global	95.91	76.23	78.59	91.27	**95.95**	61.14
	Local	**96.33**	**80.91**	**80.74**	**93.47**	95.90	**64.90**
L'Eixample	**No**	96.35	70.43	79.59	90.61	93.95	61.78
	Global	96.14	76.23	80.39	91.17	94.61	61.14
	Local	**96.76**	**79.46**	**83.93**	**95.16**	**96.57**	**64.01**
HVDM	**No**	94.99	72.36	76.67	86.32	94.67	62.92
	Global	95.48	73.95	76.67	89.77	**96.14**	63.73
	Local	**96.77**	**73.83**	**77.12**	**93.47**	93.94	**68.12**
IVDM	**No**	95.57	**70.54**	**82.58**	91.17	94.67	58.23
	Global	96.12	66.94	76.34	79	**95.23**	60.55
	Local	**96.59**	69	79.93	**93.16**	85.97	**66.05**

Table 4b. Generalisation accuracy results.

Similarity Measure		Databases						Average
Name	**Weight**	**Pima**	**Soyb Large**	**Votes**	**Zoo**	**Air P.**	**WWTP**	
Clark	**N/A**	67.31	92.55	93.15	96	91.05	36.31	80.22
Canberra	**N/A**	66.92	91.37	93.15	96	90.50	37.19	80.37
Euclidean	**No**	67.40	91.21	92.75	93.99	93.52	40.60	80.14
	Global	66.38	**91.51**	**95.51**	**96**	97.53	40.87	81.33
	Local	**71.72**	90.77	95.41	**96**	**100**	**42.61**	**83.64**
Manhat-tan	**No**	67.79	**91.50**	92.75	93.99	97.80	41.64	81.51
	Global	67.72	92.25	94.94	**96**	97.80	41.64	82.45
	Local	**71.96**	89.89	**96.47**	**96**	**99.72**	**45.52**	**84.31**
L'Eixample	**No**	67.79	91.50	92.75	93.99	91.05	41.97	80.98
	Global	69.28	92.25	94.41	**96**	98.61	41.76	82.66
	Local	**73.92**	**92.53**	**95.41**	**96**	**100**	**46.02**	**84.98**
HVDM	**No**	71.09	90.88	95.17	**98.89**	91.93	44.65	81.71
	Global	68.01	92.75	95.20	95.53	96.97	37.96	81.84
	Local	**71.20**	**92.85**	**95.40**	96	**97.09**	**46.16**	**83.49**
IVDM	**No**	**69.28**	92.18	95.17	**98.89**	92.74	**29.12**	**80.84**
	Global	65.85	**93.69**	95.36	95	91.78	26.36	78.51
	Local	68.71	92.07	**95.51**	96	**95.09**	28.50	80.54

5 Conclusions and Future Work

Main conclusions after the analyses of the performance among all three weighting schemes and all the similarity measures are that, in general, local weighting approach outperforms the global weighting schemes and the unweighted schemes. It can be argued from the whole table examination, and specifically, from the average accuracy of the local weighting schemes. They always are higher than the other schemes in all databases and in all measures, except one time. These results confirm the importance of weighting schemes in case-based similarity assessment.

A new entropy-based local weighting algorithm has been proposed. This local weighting approach seems to be better than the other weighting schemes with independence of the similarity measure or database used. Only IVDM measure seems to not to be very sensitive to the weighting schemes.

Also, it has been confirmed what was found out in a previous study [14]. *L'Eixample* distance similarity measure, which is very sensible to weighting schemes and discretization processes, seems to lightly outperform the other measures, in general, although it needs a very accurate weight selection and discretization processes, as pointed [11].

A first step has been done in the design of suitable weight selection techniques, with the proposed entropy-based local weighting approach. Future work will be focused on the design, study and analysis of other local weighting algorithms, as well as in new discretization algorithms. Comparison with other local weighting schemes is currently being done with promising results.

Acknowledgements

This work has been partially supported by the Spanish CICyT project TIC2000-1011, and EU project A-TEAM (IST 1999-10176).

References

1. D.W. Aha, R.L. Goldstone. Concept learning and flexible weighting. *Proceedings of the fourteenth Annual Conference of the Cognitive Science Society*. Bloomington, IN. The Cognitive Science Society, Lawrence Erlbaum Associates.
2. K. D. Althoff and A. Aamodt. Relating case-based problem solving and learning methods to task and domain characteristics: towards an analytic framework. AI Communications 9(3):109-116, 1996.
3. C.G. Atkeson, A.W. Moore, and S. Schaal. Locally weighted learning. *Artificial Intelligence Review*, Special Issue on lazy learning Algorithms, 1997.
4. D. Bridge. Defining and combining symmetric and asymmetric similarity measures. Procc. of *4th Eur. Work.. on Case-based Reasoning (EWCBR'98)*. LNAI-1488, pp. 52-63, 1998.

5. P. Domingos. Context-sensitive feature selection for lazy learners. *Artificial Intelligence Review*, 11, 227-253. 1997.
6. J. Dougherty, R. Kohavi and M. Sahami. Supervised and Unsupervised Discretization of continuous Features. *Procc. Of the 12th International Conference on Machine Learning*, pp. 194-202, 1995.
7. N. Howe, C. Cardie. Examining locally varying weights for nearest neighbour algorithms. *Proceedings of the Second International Conference on Case-Based Reasoning*. 1997. pp455-466. Berlin: Springer.
8. R. Kohavi, P. Langley, and Y. Yun. The utility of feature weighting in nearest-neighbour algorithms. http://citeseer.nj.nec.com/kohavi97utility.html
9. P. Kontkanen, J. Lathinen, P. Myllymäki and H. Tirri.. An unsupervised Bayesian distance measure. Procc. of *5th Eur. Work.. on Case-based Reasoning (EWCBR'2000)*. LNAI-1898, pp. 148-160, 2000.
10. L. Kurgan and K. J. Cios. Discretization Algorithm that Uses Class-Attribute Interdependence Maximisation, *Proceedings of the 2001 International Conference on Artificial Intelligence (IC-AI 2001)*, pp.980-987, Las Vegas, Nevada.
11. D.B. Leake, A. Kinley and D. Wilson. Case-based similarity assessment: estimating adaptability from experience. Procc. of *National Conference on Artificial Intelligence (AAAI'97)*. pp. 674-679, 1997.
12. T.W. Liao, and Z. Zhang. Similarity measures for retrieval in case-based reasoning systems, *Applied Artificial Intelligence*,12,267-288,1998.
13. T. Mohri and H. Tanaka. An Optimal Weighting Criterion of Case Indexing for Both Numeric and Symbolic Attributes, *Aha, D. W., editor, Case-Based Reasoning papers from the 1994 workshop*, AAAI Press, Menlo Park, CA.
14. H. Núñez, M. Sànchez-Marrè, U. Cortés, J. Comas, Ignasi R-Roda and M. Poch Improving Similarity Assessment in Case-Based Reasoning. Submitted to ECAI'2002, 2002.
15. H.R. Osborne and D. Bridge. Similarity metrics: a formal unification of cardinal and non-cardinal similarity measures. Procc. of *2nd Int. Conf. On Case-based Reasoning (ICCBR'97)*. LNAI-1266, pp. 235-244, 1997.
16. H.R. Osborne and D. Bridge. A case-based similarity framework. Procc. of *3rd Eur. Work.. on Case-based Reasoning (EWCBR'96)*. LNAI-1168, pp. 309-323, 1996.
17. M. Sànchez-Marrè, U. Cortés, I. R-Roda, and M. Poch. L'Eixample distance: a new similarity measure for case retrieval. *Procc. of 1st Catalan Conference on Artificial Intelligence (CCIA'98)*, ACIA bulletin 14-15 pp. 246-253.Tarragona, Catalonia, EU.
18. C. Stanfill, D. Waltz. Toward Memory-Based Reasoning, *Communications of the ACM*. 1986.
19. D. Ventura and T.R.Martínez. And Empirical Comparison of Discretization Methods. *Procc. Of the 10th International Symposium on Computer and Information Sciences*, pp. 443-450, 1995.
20. D. Wettschereck, D. W. Aha, and T. Mohri. A review and empirical evaluation of feature weighting methods for a class of lazy learning algorithms. *Artificial Intelligence Review*, Special Issue on lazy learning Algorithms, 1997.
21. D.R. Wilson and T.R.Martínez. Improved Heterogeneous Distance Functions, *Journal of Artificial Intelligence Research*, 6, 1-34, 1997.