Weighting methods for a Case-Based Classifier System

Maria Salamó, Elisabet Golobardes, David Vernet and Mireya Nieto

Abstract— **This paper describes two methods for weighting the feature relevance in a Case-Based Reasoning system. The first weighting method proposed inside the Case-Based Reasoning is based on Rough Sets theory. The second one is based on Sample Correlation. These weighting methods has been implemented into the platform called BASTIAN (case-BAsed SysTem In clAssificatioN), which is a Case-Based Classifier System. Experiments in different domains from the UCI repository show that these weighting methods improve accuracy rate.**

Keywords— **Case-Based Reasoning, Machine Learning, Diagnose, Knowledge Discovery.**

I. Introduction

OUR main goal is to develop, evaluate and improve classifier systems. Following this idea, we have been working on weighting methods to improve the accuracy rate in this kind of systems. This paper describes and analyses the Rough Sets theory as a weighting method in a Case-Based Classifier System. This hybrid system is compared to the Sample Correlation as weighting method to test its reliability.

The paper is structured as described. First we present an overview of Case-Based Reasoning and the main points of the platform used to test that experiments. Next, we explain both weighting methods analysed. Section III-A proposes the Rough Sets theory as a weighting method for a Case-Based Classifier system. Section III-C describes the Sample Correlation weighting method. Sections IV and V expose the testbed used and the results obtained respectively. Finally, the last section presents the conclusions and further work.

II. Case-Based Classifier System

Case-Based Reasoning integrates in one system two different characteristics: machine learning capabilities and problem solving capabilities. CBR uses a similar philosophy to that which humans sometimes use: it tries to solve new cases (examples) of a problem by using old previously solved cases [1], [2]. The process of solving new cases contributes with new information and new knowledge to the system. This new information can be used for solving other future cases. The basic method can be easily described in terms of its four phases [3].

The first phase *retrieves* old solved cases similar to the new one. In the second phase, the system tries to *reuse* the solutions of the previously retrieved cases for solving the new case. The third phase *revises* the proposed solution. Finally, the fourth phase *retains* the useful information obtained when solving the new case. In a Case-Based Classifier System, it is possible to simplify the reuse phase classifying the new case with the same class as the most similar retrieved case.

The retrieval phase is the kernel in a Case-Based Reasoning system. That phase retrieves the most similar case or cases to the new one. The most similar case is chosen using different similarity functions. The similarity functions used in that paper are based on distance concept, see section II-A.2. These similarity functions compute the similarity between two cases measuring the distance between features. If we assume an accurate weight setting of features, a Case-Based Classifier System can increase their prediction accuracy.

This paper is focused on weighting methods to compute the feature relevance. We compare 3 different ideas:

- **Not Weighting**, we do not weigh the features of our problems.
- **Rough Sets theory**, we propose the rough sets theory as a weighting method [4].
- **Sample Correlation**, we use the Sample Correlation as a weighting method [5]. This method has been proposed to compare the Rough Sets theory reliability.

A. Description of BASTIAN platform

BASTIAN (case **BA**sed **S**ys**T**em **I**n cl**A**ssificatio**N**) platform is a Case-Based Reasoning system used in classification. BASTIAN system is an extension of CaB-CS (**Ca**se-**B**ased **C**lassifier **S**ystem) system [5], [6], [7]. It allows the user to test several variants of CBR.

We present the main points of BASTIAN platform to explain in details how the Rough Sets theory is introduced in a Case-Based Reasoning system. The Sample Correlation has also been introduced into the BASTIAN system, but the original implementation was in CaB-CS system [8]. The platform developed using the JAVA programming language is explained in [4].

M. Salamó, E. Golobardes, D. Vernet and M. Nieto are with the Intelligent Systems Research Group, Enginyeria i Arquitectura La Salle (EALS), Ramon Llull University (URL), Barcelona, Spain. E-mail: {mariasal, elisabet, dave, mireyan}@salleURL.edu

A.1 General Structure of BASTIAN platform

The BASTIAN high level structure can be seen in figure 1. It maintains the four phases described in [3], [9]. The system adds a previous phase *StartupInterface*, not incorporate on the Case-Based Reasoning cycle, that prepares the initial start-up of the system.

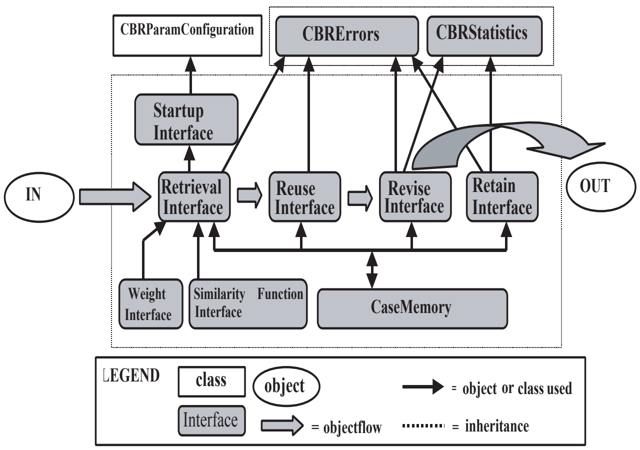

Fig. 1. General Structure in BASTIAN

The system functionalities are developed to work separately and independent in cooperation among the rest. The kernel of BASTIAN changes dinamically depending on the type of Case-Based Reasoner we want to develop. The main functionalities we focus on our paper are:

- *SimilarityFunctionInterface* concentrates all the characteristics related to similarity functions. Let us change the similarity function dynamically into the system during one execution. The similarity functions used in that paper are explained in section II-A.2.
- *WeightingInterface* contains the mecanisms to compute the feature relevance in a Case-Based Classifier System. It is related to the *RetrievalInterface*.
- {*Retrieval, Reuse, Revise, Retain*}*Interface* are the four phases of the CBR cycle. These interfaces describe the behaviour of each phase.

A.2 Similarity Functions

This paper uses the similarity functions based on the distance concept. The most used similarity function is the Nearest Neighbour algorithm [10], [11], which computes the similarity between two cases using a global similarity measure. The implementation used is based on the *Minkowsky's metric* [12], [7]. In this paper, we also use the *Clark's distance* and the *Cosine distance* [13].

Minkowsky's metric

The Minkowsky's metric is defined as:

$$Sim(Case\_x, Case\_y) = \sqrt[r]{\sum_{i=1}^{F} w_i \times |x_i - y_i|^r} \quad (1)$$

Where $Case\_x$ and $Case\_y$ are two cases, whose similarity is computed; F is the number of features that describes the case; x_i, y_i represent the value of the ith feature of cases $Case\_x$ and $Case\_y$ respectively; and w_i is the weight of the ith feature.

In this study we test the Minkowsky's metric for three different values of r: *Hamming distance* for $r = 1$, *Euclidean distance* for $r = 2$, and *Cubic distance* for $r = 3$.

Clark's distance

The Clark's distance is defined as:

$$Sim(Case\_x, Case\_y) = \sqrt[2]{\sum_{i=1}^{F} w_i \cdot \frac{|(x_i - y_i)|^2}{|(x_i + y_i)|^2}} \quad (2)$$

Where $Case\_x$ and $Case\_y$ are two cases, whose similarity is computed; F is the number of features that describes the case; and x_i, y_i represent the value of the ith feature of cases $Case\_x$ and $Case\_y$ respectively; and w_i is the weight of the ith feature.

Cosine distance

The Cosine distance is based on vector properties in an Euclidean space. It measures the Cosine angle in a n-dimensional vector space. This metric is defined as:

$$Sim(Case\_x, Case\_y) = \frac{\sum_{i=1}^{F} w_i \cdot (x_i \cdot y_i)}{\sqrt[2]{(\sum_{i=1}^{F} w_i \cdot x_i^2) \cdot (\sum_{i=1}^{F} w_i \cdot y_i^2)}} \quad (3)$$

Where F represents the number of features that describe the cases; and x_i, y_i represent the value of the ith feature of cases $Case\_x$ and $Case\_y$ respectively; and w_i is the weight of the ith feature.

III. Feature Relevance

Feature relevance is used to improve the accuracy rate of the Case-Based Classifier system [11], [14], [15]. The aim of this paper is to propose and evaluate the Rough Sets theory as a weigthing method. This approach is compared to the results obtained using the Sample Correlation [5], [16].

The section is divided in an introduction to the Rough Sets theory, the basis concepts of Rough Sets and the incorporation of Rough Sets into the Case-Based Classifier System. The last part shows the Sample Correlation as a weigthing method.

A. Rough Sets Theory

Zdzislaw Pawlak introduced Rough Sets theory in 1982 [17], [18], [19]. The idea of the Rough Sets consists of the approximation of a set by a pair of sets, called the lower and the upper approximation of this set. In fact, these approximations are inner and closure operations in a certain topology generated by the available data about elements of the set.

We use Rough Sets theory for reducing and extracting the dependencies in the knowledge. These dependencies are the basis for computing the relevance of each feature into the Case-Based Classifier System.

B. Rough Sets inside Case Based Reasoning System

We incorporate some concepts in this paper to explain how the dependencies we are looking for from the domain are obtained to select the best weighting.

B.1 Basic Concepts and Definitions

We compute from our **Universe (U)** (finite and not null set of objects that describes our problem, this is the case memory) the **concepts** (objects or cases) that form partitions of that Universe. The union of all the concepts make the entire Universe. Using all the concepts we can describe all the **equivalence relations (R)** over the universe (U). Let an equivalence relation be a set of features that describe a specific concept. U/R are the family of all equivalence classes of (R).

The universe and the relations form the **knowledge base (KB)**, defined as KB = < U, $\hat{R}$ >. Where $\hat{R}$ is the family of equivalence relations over U. Every relation over the universe is an elementary concept in the knowledge base.

All the concepts are formed by a set of equivalence relations that describe them. Thus, we search for the minimum set of equivalence relations that define the same concept as the initial set.

Definition 1 (Indiscernibility Relations)
It can be defined as IND($\hat{P}$)= $\bigcap \hat{R}$ where $\hat{P} \subseteq \hat{R}$. The indiscernibility relation is the intersection of properties over P. The indiscernibility shows the refined information over a concept and gives all the information about the equivalence relation that exists in $\hat{P}$.

Definition 2 (Basic Knowledge)
The basic knowledge is the family of ***all*** *equivalence classes of the equivalence relation IND($\hat{P}$). The basic knowledge shows all the knowledge associated with the family of equivalence relation P.*

Definition 3 (P-basic categories)
P-basic categories are those basic properties of the universe, which can be expressed using knowledge from P. They are the building blocks of the existing knowledge.

Let $K = (U, \hat{R})$ be a knowledge base.

IND(K) = (IND($\hat{P}$): $0 \neq \hat{P} \subseteq \hat{R}$) is the family of all equivalence relations defined in K.

B.2 Rough Sets

Let $X \subseteq U$ and R be an equivalence relation. We will say that:
- X is *R-definable* if X is the union of some *R-basic* categories; otherwise X is *R-undefinable.*
- The *R-definable* sets are those subsets of the universe which can be exactly defined in the knowledge base K, whereas the *R-undefinable* sets can not be defined in this knowledge base.
- The *R-undefinable* set will be also called *R-rough.*
- The set $X \subseteq U$ will be called *exact* in K if there exists $R \in IND(K)$ such that X is *R-exact*, and X is called to be *rough* in K, if X is *R-rough* for any $R \in IND(K)$.

Approximations of Set

This is the main idea of rough sets, approximate a set by other sets. The next definitions will explain this idea.

Suppose a given knowledge base $K =< U, \hat{R} >$. With each subset $X \subseteq U$ and an equivalence relation $R \subseteq IND(K)$ there are associate two subsets called:
- Lower approximation
- Upper approximation

Definition 4 (Lower approximation)
The lower approximation, defined as: $\underline{R}X = \bigcup \{ Y \in U/R : Y \subseteq X\}$. The lower approximation is the set of all elements of U which can be certainty classified as elements of X in the knowledge R.

Definition 5 (Upper approximation)
The upper approximation, $\overline{R}X = \bigcup \{ Y \in U/R : X \bigcap Y \neq \emptyset \}$. The upper approximation is the set of elements of U which can be possibly classified as elements of X, employing knowledge R.

Reduct and Core of knowledge

Intuitively, a **reduct** of knowledge is its essential part, which suffices to define all concepts occurring in the considered knowledge, whereas the *core* is the most important part of the knowledge.

Let $\hat{R}$ be a family of equivalence relations and let R $\in \hat{R}$. We will say that:
- R is *indispensable* if IND($\hat{R}$) $\neq$ IND($\hat{R}$ - R); otherwise it is *dispensable.*
- The family $\hat{R}$ is *independent* if each $R \in \hat{R}$ is *indispensable* in R; otherwise it is *dependent.*

Definition 6 (Reduct)
$\hat{Q} \in \hat{R}$ is a reduct of $\hat{R}$ if :
1. $\hat{Q}$ is independent.
2. $IND(\hat{Q}) = IND(\hat{R})$. Using Q it is possible approximate the same as using R.

Definition 7 (Core)
The set of all indispensable relations in R will be called the core of R, and will be denoted $CORE(R)$.

$$CORE(\hat{R}) = \bigcap RED(\hat{R}) \tag{4}$$

where $RED(\hat{R})$ is the family of all reducts of R.

Example III.1
If we consider a set of 8 objects in our Universe, $U = (x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8)$, using as a family of equivalence relations over U:$\hat{R}$ = (P, Q, S). Where P are colours (green, blue, red, yellow); Q are sizes (small, large, medium); and S are shapes (square, round, triangular, rectangular). In order to find the reducts and the core of the knowledge. Our equivalence classes are:

U/P = ((x_1, x_4, x_5), (x_2, x_8), (x_3),(x_6, x_7))
U/Q =((x_1, x_3, x_5), (x_6), (x_3, x_4, x_7, x_8))
U/S = ((x_1, x_5), (x_6), (x_2, x_7, x_8), (x_3, x_4))
Thus the relation $IND(R)$ has the equivalence classes:
$U/IND(\hat{R})$ = ((x_1, x_5), (x_2, x_8), (x_3), (x_4), (x_6), (x_7))
The relation P is indispensable in R, since:
$U/IND(\hat{R} - P)$ = ((x_1, x_5), (x_2, x_7, x_8), (x_3), (x_4), (x_6)) $\neq$ U/IND($\hat{R}$).
$U/IND(\hat{R} - Q)$ = ((x_1, x_5), (x_2, x_8), (x_3), (x_4), (x_6), (x_7)) = U/IND($\hat{R}$).

The information obtained is equal, so the relation Q is dispensable in R.
$U/IND(\hat{R} - S)$ = ((x_1, x_5), (x_2, x_8), (x_3), (x_4), (x_6), (x_7)) = U/IND($\hat{R}$).

Hence the relation S is also dispensable in R.

That means that the classification defined by the set of three equivalence relations P, Q and S is the same as the classification defined by relation P and Q or P and S.

So the reducts and the core are:

RED($\hat{R}$) = ((P,Q), (P,S))
CORE($\hat{R}$) = (P)

B.3 How introduce the RS theory in our CBR system?

We use the information of reducts and the core to weigh the relevance of each feature in the system. A feature that does not appear in the reducts has a weight value of 0.0, whereas a feature that appears in the core has a weight value of 1.0. The rest of features have a weight value depending on the proportional appearance in the reducts. This is the weight feature information used in BASTIAN.

Figure 2 shows the meta-level process when the Rough Sets theory are incorporated into BASTIAN. Rough Sets are divided in three steps: the first one discretises the examples, it is necessary to find the most relevant information using the Rough Sets theory; the second step searches the reducts and the core of knowledge using the Rough Sets theory; and finally, the third step uses the core and the reducts of knowledge to decide the feature relevance value.

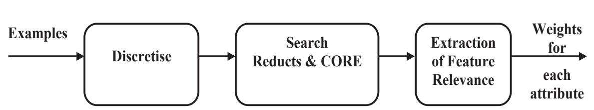

Fig. 2. High level process of *Rough Sets*

The RS theory has been introduced as weighting method in two phases of the CBR system: the first is the *start-up* phase and the second is in the *retain* phase. The start-up phase compute the weights from the initial case memory, these weights will be used by the retrieval phase later. The retain phase computes the weights from the case memory whether the new case is stored and the system works dynamically. This paper presents the results obtained when the system works statically. The feature relevance is computed in the initial case memory.

C. Sample Correlation

BASTIAN incorporates the Sample Correlation developed into CaB-CS system [8]. It uses *Sample Correlation* in order to compute the weights w_i that weigh the relevance of the features i. In other words, the weights are performed by the Sample Correlation which exists between each feature x_i and the class y ($corr(x_i, y)$). The $corr(x_i, y)$ is defined as:

$$Corr(x_i, y) \doteq \frac{1}{N-1} \sum_{j=1}^{N} \left(\frac{x_{ij} - \overline{x_i}}{S_{x_i}} \right) \left(\frac{y_j - \overline{y}}{S_y} \right) \qquad (5)$$

Where N is the number of cases; x_{ij} is the value of the ith feature for the case j; y_j is the class which belongs to the case j; $\overline{x_i}$ is the mean of the ith feature; $\overline{y}$ is the mean of the classes; S_{x_i} is the standard deviation of the feature x_i; and S_y is the standard deviation of the class y.

IV. Testbed

The experimentation has based on 4 data sets from the UCI repository (echocardiogram, iris, breast cancer Wisconsin, water-treatment), and one data set from our own repository (mammogram problem). See table I and table II which show their characteristics.

TABLE I
Data set used for these experiments

Domain	Reference
Echocardiogram	E
Iris	I
Breast cancer (Wisconsin)	BC
Water-treatment	WT
Mammogram problem	M

The mammogram problem consists of detecting breast cancer using the information found in a mammography [12], [16], [13]. A microcalcification (μCa) usually appears, in the mammographies, as small, bright, arbitrarily shaped regions on the large variety of breast texture background. Thus, their analysis and characterisation are performed throughout the extraction of features and visibility descriptors by means of several image processing techniques [20]. Each example contains the description of several μCa present in the image. For each of these microcalcifications there are 23 real valued features. In other words, the input information used is a set of $m \times 23$ real valued matrixes, where m is the number of μCa present on the image. The data set contains 216 examples.

The examples of each data set have been grouped in two sets: the training set and the test set. We use the first set to train the system, and the second one to test. The training set and the test set are generated using different proportions of the examples: 10% of the examples for the training set and the rest (90%) for the test set, 20% of the examples for the training set and the rest (80%) for the test set, ..., until 90% for the training set and 10% for the test set.

We have test each data set using different configurations of BASTIAN system, (like different similarity

TABLE II

Characteristics of the data set used in the experiments

Ref	Samples	Features	Classes	Missing Values	Inconsistent
E	132	9	2	132	Yes
I	150	4	3	0	No
BC	699	9	2	9	Yes
WT	527	38	13	591	Yes
M	216	23	2	0	Yes

functions, different retain policies, etc.), a total number of 2700 runs.

V. Results

We present in this section the main results obtained for each data set tested. Table III presents the maximum results obtained during the execution of the 90% proportion of training set and 10% test set. The ¬W column is the results obtained using BASTIAN without weighting the features, the RS-W column shows the results for the BASTIAN system using the Rough Sets theory as a weighting method, and the last one, Corr-W, shows the results for the Sample Correlation.

TABLE III

Maximum results obtained for each data set

Ref	¬W	RS-W	Corr-W
E	78.57	78.57	85.71
I	100.0	100.0	100.0
BC	98.71	100.0	98.71
WT	77.35	79.20	79.20
M	77.27	81.81	81.81

The results presented obtain a good accuracy rate. We want to outline that the maximum accuracy percentage obtained, using the Rough Sets as a weighting method, appear more frequently than the results obtained without weighting the features.

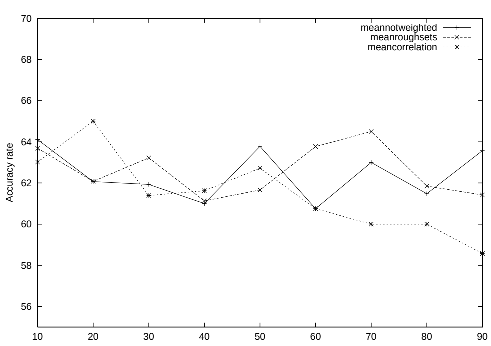

Fig. 3. Mean results in the echocardiogram problem

Figure 3 shows the mean results obtained for the echocardiogram problem in all the training set proportions. Figure 3 denotes how important is the number of cases into the case memory, and we can also notice that the results depend on the number of missing values.

TABLE IV

Results for the Iris problem

Prop Train	Max ¬W	Max RS-W	Max Corr-W	Mean ¬W	Mean RS-W	Mean Corr-W
40%	98.88	97.77	98.88	96.22	96.00	96.22
60%	97.77	97.77	98.33	95.33	95.50	96.16
70%	100.0	100.0	97.77	95.11	95.33	95.77
80%	100.0	100.0	100.0	97.00	97.00	97.33
90%	100.0	100.0	100.0	96.66	96.66	97.33

Table IV shows the results obtained in different training sets proportions for the Iris problem. The results presented are the maximum and the mean percentage values. As it can be seen there are few differences between the Rough Sets hybrid system and the original Case-Based Classifier System. The results denote also that it is very important the number of cases included into the case memory to achieve a good accuracy in the weighting method. That influence can be seen into the mean results for the Sample Correlation.

It is important to remark that the prediction accuracy depends on the case memory size. This fact can be seen in all the problems analysed.

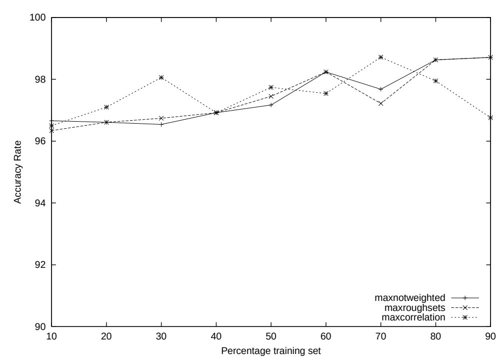

Fig. 4. Maximum results in Breast Cancer Wisconsin

The results obtained for the Breast Cancer Wisconsin problem can be found in figure 4. The case memory in that data set is bigger than the previous ones. That big case memory influences into the behaviour of the system. Weighting methods get better performance in early percentage of training sets than others data sets. But the system also is sensitive to the increasing number of samples when it arrives to the last percentages of training samples.

Figures 5 and 6 show the results obtained for all the training sets proportions in the Water Treatment and Mammogram problem respectively. As it can be seen, the weighting feature methods needs a huge amount of cases to develop a good weighting for the retrieval phase. However, the system accuracy rate increases when there are enough information in the system to develop a good weighting criterion. Also, the system decreases the standard deviation value if it uses the Rough Sets theory as a weighting method.

We can also notice that it is very important to select a representative initial case memory to achieve better results. Hence, most of the best results obtained have

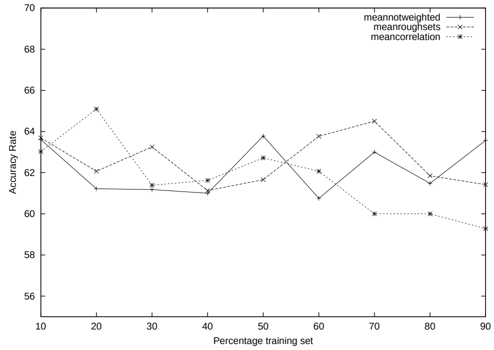

Fig. 5. Mean results in Water treatment

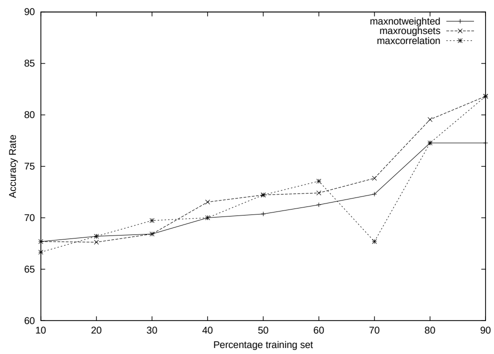

Fig. 6. Maximum results in Mammogram problem

been achieved using an initial training when the system load the initial case memory. The training set has been decreased following this method. In this way, the cases chosen were the more representatives to explain the problem.

Finally, it is important to denote that all the discretisation has been done using the same criterion. This criterion must be changed depending on the upper and lower bounds of each feature. This discretisation influences the results.

VI. Conclusions and Further Work

This paper has introduced two different approaches to weigh the feature relevance. The first one proposed is the introduction of Rough Sets theory into a Case-Based Classifier System. The second one is the Sample Correlation as a comparative system to evaluate the Rough Sets approach.

Both approaches has been tested using 4 data sets from the UCI repository and one from our own repository. We can conclude that: (1) both approaches need a large number of samples to be able to get accurate weighting values; (2) the Rough Sets approach help the system to balance its own results, there are not many differences in terms of deviation between all the versions tested.

Our further work in this area will be to achieve better performance using different criteria as weighting methods and analysing other methods reported at literature as [11].

Acknowledgements

This work was supported by the *Ministerio de Sanidad y Consumo, Instituto de salud Carlos III*, Fondo de Investigación Sanitaria of Spain, Grant No. 00/0033-02. The results of this work were obtained using the equipment co-funded by the *Direcció de Recerca de la Generalitat de Catalunya* (D.O.G.C 30/12/1997). We would thank EALS (URL) for their support to our AI Research Group.

References

[1] J. Kolodner, *Case-Based Reasoning*, Morgan Kaufmann Publishers, Inc., 1993.

[2] C.K. Riesbeck and R.C. Schank, *Inside Case-Based Reasoning*, Lawrence Erlbaum Associates, Hillsdale, NJ, US, 1989.

[3] A. Aamodt and E. Plaza, "Case-Based Reasoning: Foundations Issues, Methodological Variations, and System Approaches," in *AI Communications*, 1994, vol. 7, pp. 39–59.

[4] M. Salamó and E. Golobardes, "BASTIAN: Incorporating the Rough Sets theory into a Case-Based Classifier System," in *III Congrés Català d'Intelligència Artificial (CCIA'00)*, Barcelona, Spain.

[5] E. Golobardes and J.M. Garrel, "Avaluació d'un sistema classificador basat en casos per a la diagnosi de biòpsies de teixit de glàndules mamàries," in *Proceedings de les Jornades d'Inteligència Artifical (JIA'97)*, 1997.

[6] Garrell, J.M. and Golobardes , E, and Bernadó , E. and Llorà , X., "Automatic Classification of Mamary Biopsy Images with Machine Learning techniques," in *Proceedings of Engineering of Inteligent Systems (EIS'98)*, 1997.

[7] Garrell, J.M. and Golobardes, E. and Bernadó, E. and Llorà, X., "Automatic diagnosis with Genetic Algorithms and Case-Based Reasoning," vol. 13, no. 4, pp. 367–362, October 1999, Elsevier Science Ltd., ISSN 0954-1810.

[8] E. Golobardes, *Aportacions al raonament basat en casos per resoldre problemes de classificació*, Ph.D. thesis, Enginyeria La Salle, Universitat Ramon Llull, Juny 1998.

[9] I. Watson, *Applying Case-Based Reasoning: Techniques for Enterprise Systems*, Morgan Kaufmann Publishers Inc., 1997.

[10] D. Aha and P. Harrison, "Case-Based Sonogram Classification," Tech. Rep. AIC-93-041, Technical Report NRL/FR/5510-94-9707, Navy Center for Applied Research in AI, Washington, D.C., October 1994.

[11] D. Aha, "Feature weighting for lazy learning algorithms," Tech. Rep. AIC-98-003, Technical Report, Navy Center for Applied Research in AI, Washington, D.C., 1998.

[12] X. Llorà, E. Golobardes, and M. Salamó, "Diagnosis of microcalsification using Case-Based and Genetic Algorithms," in *Proceedings of Engineering of Inteligent Systems (EIS2000)*, 2000, vol. 1, pp. 254– 263.

[13] M. Salamó, E. Golobardes, J.M. Garrell, and J. Martí, "Clasificación de microcalcificaciones usando Razonamiento Basado en Casos," in *III Jornadas de Transferencia Tecnológica de Inteligencia Artificial (TTIA'99)*, Murcia, Spain, November 1999.

[14] P. Domingos, "Context-sensitive feature selection for lazy learners," 1997, vol. 11, pp. 227–253.

[15] N. Howe and C. Cardie, "Examining locally varying weights for nearest neighbor algorithms," in *Proceedings of the Second ICCBR Conference*.

[16] J. Martí, J. Español, E. Golobardes, J. Freixenet, R. García, and M. Salamó, "Classification of microcalcifications in digital mammograms using case-based reasonig," in *International Workshop on digital Mammography*, 2000.

[17] Z. Pawlak, "Rough Sets," in *International Journal of Information and Computer Science*, 1982, vol. 11.

[18] Z. Pawlak, *Rough Sets: Theoretical Aspects of Reasoning about Data*, Kluwer Academic Publishers, 1991.

[19] A. Skowron, "Synthesis of Adaptive Decision Systems from Experimental Data," in *Fith Scandinavian Conference on Artificial Intelligence*, A. Aamodt and J. Komorowski, Eds., Trondheim, Norway, May 1995, vol. 28, IOS Press.

[20] L. Shen, R.M. Rangayyan, and J.L. Dessautels, "Detection and classification of mammographic calcifications," in *In State of the Art in Digital Mammographic Image Analysis*, 1994, pp. 198–212.